

THRESHOLD AUTOREGRESSION WITH A UNIT ROOT

BY MEHMET CANER AND BRUCE E. HANSEN¹

This paper develops an asymptotic theory of inference for an unrestricted two-regime threshold autoregressive (TAR) model with an autoregressive unit root. We find that the asymptotic null distribution of Wald tests for a threshold are nonstandard and different from the stationary case, and suggest basing inference on a bootstrap approximation. We also study the asymptotic null distributions of tests for an autoregressive unit root, and find that they are nonstandard and dependent on the presence of a threshold effect. We propose both asymptotic and bootstrap-based tests. These tests and distribution theory allow for the joint consideration of nonlinearity (thresholds) and nonstationary (unit roots).

Our limit theory is based on a new set of tools that combine unit root asymptotics with empirical process methods. We work with a particular two-parameter empirical process that 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 the threshold effect is in the short-run dynamics, rather than in the dominant root. While the conventional ADF test for a unit root is insignificant, our TAR unit root tests are arguably significant. The evidence is quite strong that the unemployment rate is not a unit root process, and there is considerable evidence that the series is a stationary TAR process.

KEYWORDS: Bootstrap, nonlinear time series, identification, nonstationary, Brownian motion, unemployment rate.

1. INTRODUCTION

THE THRESHOLD AUTOREGRESSIVE (TAR) MODEL was introduced by Tong (1978) and has since become quite popular in nonlinear time series. See Tong (1983, 1990) for reviews. A sampling theory of inference has been quite slow to develop, however. Among the more important contributions, Chan (1991) and Hansen (1996) describe the asymptotic distribution of the likelihood ratio test for a threshold, Chan (1993) showed that the least squares estimate of the threshold is super-consistent and found its asymptotic distribution, Hansen (1997b, 2000) developed an alternative approximation to the asymptotic distribution, and Chan and Tsay (1998) 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, ergodic, and have no unit roots. This makes it impossible to discriminate nonstationarity from nonlinearity. To aid in the analysis of possibly nonstationary and/or nonlinear time series, we provide the first rigor-

¹Caner thanks TUBITAK and Hansen thanks the National Science Foundation and the Alfred P. Sloan Foundation for research support. We thank Frank Diebold, Peter Pedroni, Pierre Perron, Simon Potter, four referees and the co-editor for stimulating comments on earlier drafts.

ous treatment of statistical tests that simultaneously allow for both effects. Specifically, we examine a two-regime TAR(k) with an autoregressive unit root. Within this model, we study Wald tests for a threshold effect (for nonlinearity) and Wald and t tests for unit roots (for nonstationarity). We allow for general autoregressive orders, and do not artificially restrict the coefficients across regimes.

We find that the Wald test for a threshold has a nonstandard asymptotic null distribution. This is partially due to the presence of a parameter that is not identified under the null (see Davies (1987), Andrews and Ploberger (1994), and Hansen (1996)), and partially due to the assumption of a nonstationary autoregression. The asymptotic null distribution has two components, one that reflects the unit root and deterministic trends but is otherwise free of nuisance parameters, and the other component that is identical to the empirical process found in the stationary case, and is nuisance-parameter dependent. Hence the asymptotic distribution is nonsimilar and cannot be tabulated. We propose bootstrap procedures to approximate the sampling distribution.

We find that Wald tests for a unit root have asymptotic null distributions that depend on whether or not there is a true threshold effect, and construct bounds that are free of nuisance parameters. Our simulations suggest that these asymptotic approximations are inferior to bootstrap methods, which we recommend for empirical practice. Using simulations, we show that our threshold unit root tests have better power than the conventional ADF unit root test (Said and Dickey (1984)) when the true process is nonlinear.

Our distribution theory is based on a new set of asymptotic tools utilizing a double-indexed empirical process that converges weakly to a two-parameter Brownian motion, and we establish weak convergence to a stochastic integral defined with respect to this two-parameter process. This theory may have applications beyond those presented here.

The results presented here relate to a growing literature on threshold autoregressions with unit roots. In a Monte Carlo experiment, Pippenger and Goering (1993) document that the power of the Dickey-Fuller (1979) unit root test falls dramatically within one class of TAR models. Balke and Fomby (1997) introduce a multivariate model of threshold cointegration, but offer no rigorous distribution theory. Tsay (1997) introduces a univariate unit root test when the innovations follow a threshold process. He finds the asymptotic distribution when the threshold is known, and provides simulations for the case of estimated threshold. His model requires the leading autoregressive lag to be constant across threshold regimes, and is a special case of the model we consider. Gonzalez and Gonzalo (1998) carefully examine a TAR(1) model allowing for a unit root. They provide conditions under which the process is stationary and geometrically ergodic, and discuss testing for a threshold in the TAR(1) model.

We illustrate our proposed techniques through an application to the monthly U.S. unemployment rate among adult males. There is a substantial literature documenting nonlinearities and threshold effects in the U.S. unemployment rate. A partial list includes Rothman (1991), Chen and Lee (1995), Montgomery,

Zarnowitz, Tsay, and Tiao (1998), Altissimo and Violante (1996), Chan and Tsay (1998), Hansen (1997b), and Tsay (1997). This literature is connected to a broader literature studying nonlinearities in the business cycle, which includes contributions by Neftci (1984), Hamilton (1989), Beaudry and Koop (1993), Potter (1995), and Galbraith (1996). Empirical researchers are faced with the fact that the conventional unit root tests are unable to reject the hypothesis that the post-war unemployment rate is nonstationary. Prior statistical methods cannot disentangle nonstationarity from nonlinearity because of the joint modeling problem of unit roots and thresholds. With our new methods, we are able to rigorously address these issues. In our application, we find very strong evidence that the unemployment rate has a threshold nonlinearity. Furthermore, we find strong evidence against the unit root hypothesis, and fairly strong (although not conclusive) evidence in favor of a stationarity threshold specification. Our methods point to the conclusion that the unemployment rate is a stationary nonlinear process.

This paper is organized as follows. Section 2 presents the TAR model. Section 3 introduces a new set of asymptotic tools that are useful for the study of threshold processes with possible unit roots. Section 4 presents the distribution theory for the threshold test, including a Monte Carlo study of size and power. 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 that replicates the empirical work is available from the webpage www.ssc.wisc.edu/~bhansen.

2. TAR MODEL

The model is the following threshold autoregression (TAR):

$$(1) \quad \Delta y_t = \theta'_1 x_{t-1} 1_{\{Z_{t-1} < \lambda\}} + \theta'_2 x_{t-1} 1_{\{Z_{t-1} \geq \lambda\}} + e_t,$$

$t = 1, \dots, T$, where $x_{t-1} = (y_{t-1} r'_t \Delta y_{t-1} \cdots \Delta y_{t-k})'$, $1_{\{\cdot\}}$ is the indicator function, e_t is an iid error, $Z_t = y_t - y_{t-m}$ for some $m \geq 1$, and r_t is a vector of deterministic components including an intercept and possibly a linear time trend. The threshold λ is unknown. It takes on values in the interval $\lambda \in A = [\lambda_1, \lambda_2]$ where λ_1 and λ_2 are picked so that $P(Z_t \leq \lambda_1) = \pi_1 > 0$ and $P(Z_t \leq \lambda_2) = \pi_2 < 1$. It is typical to treat π_1 and π_2 symmetrically so that $\pi_2 = 1 - \pi_1$, which imposes the restriction that no "regime" has less than $\pi_1\%$ of the total sample. The particular choice for π_1 is somewhat arbitrary, and in practice must be guided by the consideration that each "regime" needs to have sufficient observations to adequately identify the regression parameters. This choice is discussed in more detail at the end of Section 4.2.

The particular specification for the threshold variable Z_{t-1} is not essential to the analysis. In general, what is necessary for our results is that Z_{t-1} be predetermined, strictly stationary, and ergodic with a continuous distribution

function. Our particular choice $Z_t = y_t - y_{t-m}$ is convenient because it is ensured to be stationary under the alternative assumptions that y_t is $I(1)$ and $I(0)$.

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

$$\theta_1 = \begin{pmatrix} \rho_1 \\ \beta_1 \\ \alpha_1 \end{pmatrix}, \quad \theta_2 = \begin{pmatrix} \rho_2 \\ \beta_2 \\ \alpha_2 \end{pmatrix},$$

where ρ_1 and ρ_2 are scalar, β_1 and β_2 have the same dimension as r_t , and α_1 and α_2 are k -vectors. Thus (ρ_1, ρ_2) are the slope coefficients on y_{t-1} , (β_1, β_2) are the slopes on the deterministic components, and (α_1, α_2) are the slope coefficients on $(\Delta y_{t-1}, \dots, \Delta y_{t-k})$ in the two regimes.

Our model (1) specifies that all the slope coefficients switch between the regimes, but in some applications it may be desirable for only a subset of the coefficients to depend on the regime. There is nothing essential in this choice and other parameterizations may be used in other contexts. For the theoretical presentation, we retain the general unrestricted model (1) for ease of exposition.

We impose the following maintained conditions on the model:

ASSUMPTION 1: e_t is an iid mean-zero sequence with a bounded density function, and $E|e_t|^{2\gamma} < \infty$ for some $\gamma > 2$. For some matrix δ_T and continuous vector function $r(s)$, $\delta_T r_{[Ts]} \Rightarrow r(s)$. The following parameter restrictions apply: $\rho_1 = \rho_2 = 0$; for constants μ_1 and μ_2 , $\beta_1' r_t = \mu_1$ and $\beta_2' r_t = \mu_2$; and $|\alpha_1' \iota| < 1$ and $|\alpha_2' \iota| < 1$, where ι is a k -vector of ones.

The assumption that e_t is an independent sequence is essential for our asymptotic distribution theory and for our bootstrap approximations, and appears to be a meaningful restriction on the model. The parameter restrictions ensure that the time-series Δy_t is stationary and ergodic, so that y_t is integrated of order one and can be described as a unit root process. The restriction that $\beta_1' r_t = \mu_1$ and $\beta_2' r_t = \mu_2$ implies that the only “trend” component that enters the true process is the intercept. This restriction is standard in the unit root testing literature, and guarantees that there are no quadratic trends in y_t .

An important question in applications is how to specify the deterministic component r_t . If the series y_t is nontrended it would seem natural to set $r_t = 1$, while if the series is highly trended then a natural option is to set $r_t = (1 \ t)'$. The inclusion of the linear trend will be necessary to ensure that the unit root tests we discuss in Section 5 have power against trend stationary alternatives.

The coefficient restrictions on α_1 and α_2 given in Assumption 1 are sufficient to ensure that the series Δy_t is stationary and ergodic (see Pham and Tran (1985)), which is the only role of these restrictions. While these are a known set of sufficient conditions, they are not necessary. The region of ergodicity is larger than these assumptions, which is what is essential for our results.

The TAR model (1) is estimated by least squares (LS). To implement LS estimation, it is convenient to use concentration. For each $\lambda \in \Lambda$, (1) is esti-

mated by ordinary least squares (OLS):

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

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 $\hat{\sigma}^2(\lambda)$:

$$\hat{\lambda} = \underset{\lambda \in A}{\operatorname{argmin}} \hat{\sigma}^2(\lambda).$$

The LS estimates of the other parameters are then found by plugging in the point estimate $\hat{\lambda}$, vis. $\hat{\theta}_1 = \hat{\theta}_1(\hat{\lambda})$, and $\hat{\theta}_2 = \hat{\theta}_2(\hat{\lambda})$. We write the estimated model as

$$(3) \quad \Delta y_t = \hat{\theta}'_1 x_{t-1} 1_{\{z_{t-1} < \hat{\lambda}\}} + \hat{\theta}'_2 x_{t-1} 1_{\{z_{t-1} \geq \hat{\lambda}\}} + \hat{e}_t,$$

which also defines the LS residuals \hat{e}_t . Let $\hat{\sigma}^2 = T^{-1} \sum_{t=1}^T \hat{e}_t^2$ denote the residual variance from the LS estimation.

The estimates (3) can be used to conduct inference concerning the parameters of (1) using standard Wald and t statistics. While the statistics are standard, their sampling distributions are nonstandard, due to the presence of possible unidentified parameters and nonstationarity. We explore large-sample approximations in the following sections.

3. UNIT ROOT ASYMPTOTICS FOR THRESHOLD PROCESSES

The sampling distributions for our proposed statistics will require some new asymptotic tools. Rather than develop these tools for our specific model, we first develop the needed results under a set of more general conditions. Let “ \Rightarrow ” denote weak convergence as $T \rightarrow \infty$ with respect to the uniform metric on $[0, 1]^2$.

ASSUMPTION 2: For the sequence $\{U_t, e_t, X_t, w_t\}$, let \mathfrak{F}_t denote the natural filtration.

1. $\{U_t, e_t, w_t\}$ is strictly stationary and ergodic and strong mixing with mixing coefficients α_m satisfying $\sum_{m=1}^\infty \alpha_m^{1/2-1/r} < \infty$ for some $r > 2$;
2. U_t has a marginal $U[0, 1]$ distribution;
3. e_t is independent of \mathfrak{F}_{t-1} , $E(e_t) = 0$, and $E|e_t|^4 = \kappa < \infty$;
4. there exists a nonrandom matrix δ_T such that the array $X_{Tt} = \delta_T X_t$ satisfies $X_{T\lceil Ts \rceil} \Rightarrow X(s)$ on $s \in [0, 1]$, where $X(s)$ is continuous almost surely;
5. $E|w_t|^{2+\eta} < \infty$ for some $\eta > 0$.

The two most natural examples of processes X_t that satisfy condition 4 are integrated processes and polynomials in time. First, if X_t is an $I(1)$ process, then $\delta_T = T^{-1/2}$ and $X(s)$ is a scaled Brownian motion. Second, if $X_t = (1 \ t)'$ (a constant and linear trend), then

$$\delta_T = \begin{pmatrix} 1 & 0 \\ 0 & T^{-1} \end{pmatrix} \quad \text{and} \quad X(s) = (1 \ s)'$$

Other polynomials in time, or higher-order integrated processes, can be handled similarly.

Define $1_t(u) = 1_{\{u_t < u\}}$, the partial-sum process

$$W_i(u) = \sum_{t=1}^i 1_{t-1}(u)e_t$$

and scaled array

$$\begin{aligned} W_T(s, u) &= \frac{1}{\sigma\sqrt{T}} W_{[Ts]}(u) \\ &= \frac{1}{\sigma\sqrt{T}} \sum_{t=1}^{[Ts]} 1_{t-1}(u)e_t, \end{aligned}$$

where $\sigma^2 = Ee_t^2 < \infty$.

DEFINITION 1: $W(s, u)$ is a two-parameter Brownian motion on $(s, u) \in [0, 1]^2$ if $W(s, u) \sim N(0, su)$ and

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

THEOREM 1: Under Assumption 2,

$$(4) \quad W_T(s, u) \Rightarrow W(s, u)$$

on $(s, u) \in [0, 1]^2$ as $T \rightarrow \infty$, where $W(s, u)$ is a two-parameter Brownian motion.

It may be helpful to think of Theorem 1 as a two-parameter generalization of the usual functional limit theorem. We now define stochastic integration with respect to the two-parameter process $W(s, u)$. Let

$$\begin{aligned} J(u) &= \int_0^1 X(s) dW(s, u) \\ &\equiv \text{plim}_{N \rightarrow \infty} \sum_{j=1}^N X\left(\frac{j-1}{N}\right) \left(W\left(\frac{j}{N}, u\right) - W\left(\frac{j-1}{N}, u\right) \right), \end{aligned}$$

where plim denotes convergence in probability. The integration is over the first argument of $W(s, u)$, holding the second argument constant. We will call the process $J(u)$ a stochastic integral process.

THEOREM 2: *Under Assumption 2,*

$$\begin{aligned} \frac{1}{\sigma\sqrt{T}} \sum_{t=1}^T X_{Tt-1}1_{t-1}(u)e_t &= \int_0^1 X_T(s) dW_T(s, u) \\ \Rightarrow J(u) &= \int_0^1 X(s) dW(s, u) \end{aligned}$$

on $u \in [0, 1]$ as $T \rightarrow \infty$, and $J(u)$ is almost surely continuous.

This result is a natural extension of the theory of weak convergence to stochastic integrals (see Hansen (1992)).

Finally, we need to describe the asymptotic covariances between stationary processes and the nonstationary process X_t when interacted with the indicator function $1_{t-1}(u)$. Define the moment functionals $h(u) = E(1_{t-1}(u)w_{t-1})$ and $H(u) = E(1_{t-1}(u)w_{t-1}w'_{t-1})$.

THEOREM 3: *Under Assumption 2, on $u \in [0, 1]$ as $T \rightarrow \infty$,*

1. $(1/T)\sum_{t=1}^T 1_{t-1}(u)w_{t-1}X'_{Tt-1} \Rightarrow h(u)\int_0^1 X(s)' ds$;
2. $(1/T)\sum_{t=1}^T 1_{t-1}(u)w_{t-1}w'_{t-1} \Rightarrow H(u)$;
3. $(1/T)\sum_{t=1}^T 1_{t-1}(u)X_{Tt-1}X'_{Tt-1} \Rightarrow u\int_0^1 X(s)X(s)' ds$.

Theorems 1–3 will serve as the building blocks for the subsequent theory developed in this paper.

4. TESTING FOR A THRESHOLD EFFECT

4.1. Wald Test Statistic

In model (1) a question of particular interest is whether or not there is a threshold effect. The threshold effect disappears under the joint hypothesis

$$(5) \quad H_0: \theta_1 = \theta_2.$$

Our test² of (5) is the standard Wald statistic W_T for this restriction. This statistic can be written as

$$W_T = T \left(\frac{\hat{\sigma}_0^2}{\hat{\sigma}^2} - 1 \right)$$

where $\hat{\sigma}^2$ is defined above as the residual variance from (3), and $\hat{\sigma}_0^2$ is the residual variance from OLS estimation of the null linear model.

The following relationship may be of some interest. Let

$$W_T(\lambda) = T \left(\frac{\hat{\sigma}_0^2}{\hat{\sigma}^2(\lambda)} - 1 \right)$$

²In applications it may also be useful to consider statistics that focus on subvectors of θ_1 and θ_2 . See Section 4.5.

denote the Wald statistic of the hypothesis (5) for fixed λ from regression (2). Then since $W_T(\lambda)$ is a decreasing function of $\hat{\sigma}^2(\lambda)$ we see that

$$(6) \quad W_T = W_T(\hat{\lambda}) = \sup_{\lambda \in A} W_T(\lambda).$$

Thus the Wald statistic for H_0 is often called the ‘‘Sup-Wald’’ statistic.

4.2. Asymptotic Distribution

Under the null hypothesis (5) of no threshold effect the parameter λ is not identified, rendering the testing problem nonstandard. The asymptotic distribution of W_T for stationary data has been investigated by Davies (1987), Chan (1991), Andrews and Ploberger (1994), and Hansen (1996). Our concern is with the case of a unit root, which has not been studied previously.

Let $G(\cdot)$ denote the marginal distribution function of Z_t , set $\pi_1 = G(\lambda_1)$ and $\pi_2 = G(\lambda_2)$, and define $1_t(u) = 1_{\{G(Z_t) < u\}}$ and $w_{t-1} = (\Delta y_{t-1}, \dots, \Delta y_{t-k})$.

THEOREM 4: *Under Assumption 1, $H_0: \theta_1 = \theta_2$,*

$$W_T \Rightarrow T = \sup_{\pi_1 \leq u \leq \pi_2} T(u),$$

where

$$(7) \quad T(u) = Q_1(u) + Q_2(u),$$

and $Q_1(u)$ and $Q_2(u)$ are the independent stochastic processes defined in (8) and (9) below.

1. Let $W(s, u)$ be a two-parameter Brownian motion; set $W(s) = W(s, 1)$. Set $X(s) = (W(s) \ r(s)')'$, $J_1(u) = \int_0^1 X(s) dW(s, u)'$, a stochastic integral process as defined in Section 3, and set $J_1^*(u) = J_1(u) - uJ_1(1)$. Then

$$(8) \quad Q_1(u) = J_1^*(u)' \left(u(1-u) \int_0^1 X(s)X(s)' ds \right)^{-1} J_1^*(u).$$

2. Let $J_2(u)$ be a zero-mean Gaussian process, independent of $W(s, u)$, with covariance kernel $E(J_2(u_1)J_2(u_2)') = \Omega(u_1 \wedge u_2)$, where $\Omega(u) = H(u) - u^{-1}h(u)h(u)'$, $H(u) = E(1_{t-1}(u)w_{t-1}w_{t-1}')$, and $h(u) = E(1_{t-1}(u)w_{t-1})$. Then

$$(9) \quad Q_2(u) = J_2^*(u)' (\Omega(u) - \Omega(u)\Omega(1)^{-1}\Omega(u))^{-1} J_2^*(u).$$

Theorem 4 gives the large sample distribution of the conventional Wald statistic for a threshold for the nonstationary autoregression (1) under the unit root restriction $\rho_1 = \rho_2 = 0$. Notice that the distribution T can be written as the supremum of the sum of two independent processes $Q_1(u)$ and $Q_2(u)$. The process $Q_2(u)$ is a chi-square process, taking the same form as found by Hansen (1996) for threshold tests applied to stationary data. The process $Q_1(u)$ takes a very different form, and is a reflection of the nonstationary regressors. We see

that the presence of nonstationarity in the data changes the asymptotic distribution of the threshold test, and this will need to be taken into consideration for correct asymptotic inference.

The case of stationary data can be deduced from Theorem 4 by removing $W(s)$ from the definition of $X(s)$, which is the result of omitting y_{t-1} from the regression model (1). The asymptotic distribution corresponds to that found by Hansen (1996).

In general, the asymptotic distribution T is nonpivotal and depends upon the nuisance parameter function $\Omega(u)$. The dependence on the data structure is quite complicated, so as a result, critical values cannot be tabulated. In the following section, we discuss a bootstrap method to approximate the null distribution of W_T .

It is also helpful to observe that Theorem 4 shows that the critical values of T will increase as π_1 decreases and/or π_2 decreases. This means that larger values of W_T will be needed to reject the null of stationarity when extreme values of π_1 and/or π_2 are used. In analogy to the discussion in Andrews (1993) concerning the choice of trimming in tests for structural change, the distribution of T diverges to positive infinity as $\pi_1 \rightarrow 0$ or $\pi_2 \rightarrow 1$. Thus setting $\pi_1 = 0$ or $\pi_2 = 1$ renders the test inconsistent. It follows that it is necessary to select values of π_1 and π_2 in the interior of $(0, 1)$, and values too close to the endpoints reduce the power of the test. On the other hand, it is desirable to select π_1 and π_2 so that the true value of $G(\lambda_0)$ lies in the interval $[\pi_1, \pi_2]$ (under the alternative hypothesis); otherwise the test may have difficulty in detecting the presence of the threshold effect. Andrews (1993) suggests that setting $\pi_1 = .15$ and $\pi_2 = .85$ provides a reasonable trade-off between these considerations, and these are the values we select in our simulations and applications. Since the particular choice is somewhat arbitrary, it appears sensible in practical applications to explore the robustness of the results to this choice.

4.3. Bootstrap

In this section, we discuss two bootstrap approximations to the asymptotic distribution of W_T , one based on the unrestricted estimates, and the other enforcing the restriction of a unit root. These bootstrap approximations can be used to calculate critical values and p -values.

Under the null hypothesis, $\theta_1 = \theta_2 = \theta$, say, so for simplicity we omit subscripts on the coefficients for the remainder of this section. Under H_0 and the assumption that the only deterministic component is the intercept μ (see Assumption 1) the model simplifies to $\Delta y_t = \rho y_{t-1} + \mu + \alpha' \tilde{\Delta} y_{t-1} + e_t$, where $\tilde{\Delta} y_{t-1} = (\Delta y_{t-1} \cdots \Delta y_{t-k})'$. As the distribution of the test is invariant to level shifts, we can set $\mu = 0$ so the model simplifies to $\Delta y_t = \rho y_{t-1} + \alpha' \tilde{\Delta} y_{t-1} + e_t$. Since this is entirely determined by ρ , α , and the distribution F of the error e_t , we can use a model-based bootstrap.

We first describe the unrestricted bootstrap estimate. Let $(\tilde{\rho}, \tilde{\alpha}, \tilde{F})$ be estimates of (ρ, α, F) . The bootstrap distribution W_T^b is a conditional distribution determined by the random inputs $(\tilde{\rho}, \tilde{\alpha}, \tilde{F})$. It is determined as follows. Let e_t^b be a random draw from \tilde{F} , and let y_t^b be generated as $\Delta y_t^b = \tilde{\rho} y_{t-1}^b + \tilde{\alpha}' \tilde{\Delta} y_{t-1}^b + e_t^b$ where $\tilde{\Delta} y_{t-1}^b = (\Delta y_{t-1}^b \cdots \Delta y_{t-k}^b)'$. Initial values for the recursion can be set to sample values of the demeaned series. The distribution of y_t^b is the bootstrap distribution of the data. Let W_T^b be the threshold Wald test calculated from the series y_t^b . The distribution of W_T^b is the bootstrap distribution of the Wald test. Its bootstrap p -value is $p_T = P(W_T^b > W_T | \mathfrak{S}_T)$, where conditioning on \mathfrak{S}_T denotes that this probability is conditional on the observed data. Typically, the bootstrap p -value is calculated by simulation, where a large number of independent Wald tests W_T^b are simulated, and the p -value p_T is approximated by the frequency of simulated W_T^b that exceed W_T .

To implement the bootstrap we need estimates $(\tilde{\rho}, \tilde{\alpha}, \tilde{F})$. For (ρ, α) we need an estimate that imposes the null hypothesis; an obvious choice is to use the estimate $(\tilde{\rho}, \tilde{\alpha})$ obtained by regressing y_t on x_t . An estimator for F is the empirical distribution of the OLS residuals \tilde{e}_t . In typical statistical contexts (when the asymptotic distribution is a smooth function of the model parameters and the parameter estimates are consistent) bootstrap distributions will converge in probability to the correct asymptotic distribution (denoted $W_T^b \Rightarrow_p T$),³ implying that the bootstrap p -value will be first-order asymptotically correct. In our model, this convergence depends on the true value of ρ . If the time-series is stationary, then the bootstrap will achieve the correct first-order asymptotic distribution, since the model parameters are consistently estimated and the asymptotic distribution is a smooth function of these parameters (a similar formal argument is presented in Hansen (1996)). If the time-series has a unit root, however, this will not be the case. The asymptotic distribution is discontinuous in the parameters at the boundary $\rho = 0$, so the bootstrap distribution will not be consistent for the correct sampling distribution.

We can achieve the correct asymptotic distribution by imposing the true unit root. This is done by imposing the constraint $\rho = 0$. This can be done by setting the estimates of (ρ, α, F) to be $(0, \tilde{\alpha}, \tilde{F})$, where $(\tilde{\alpha}, \tilde{F})$ are defined above. Then generate random samples y_t^b from the model $\Delta y_t^b = \tilde{\alpha}' \tilde{\Delta} y_{t-1}^b + e_t^b$ with e_t^b drawn randomly from \tilde{F} . These samples are unit root processes. For each sample y_t^b , calculate the test statistic W_T^b . The estimated bootstrap p -value is the percentage of simulated W_T^b that exceed the observed W_T .

This constrained bootstrap is first-order correct under H_0 if the true parameter values satisfy $\rho = 0$. If the true process is stationary, however, the constrained bootstrap will be incorrect. We see that we have introduced two bootstrap methods, one appropriate for the stationary case, and the other appropriate for the unit root case. If the true order of integration is unknown (as is likely in applications), then it appears prudent to calculate the bootstrap

³The symbol " \Rightarrow_p " denotes "weak convergence in probability" as defined in Gine and Zinn (1990), which is the appropriate convergence definition for bootstrap distributions.

p -values both ways, and base inference on the more conservative (the larger) p -value.

4.4. *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 $k = 1$, a linear time trend, and $z_{t-1} = \Delta y_{t-1}$:

$$(10) \quad \Delta y_t = (\rho_1 y_{t-1} + \beta_1 t + \mu_1 + \alpha_1 \Delta y_{t-1}) 1_{\{\Delta y_{t-1} < \lambda\}} + (\rho_2 y_{t-1} + \beta_2 t + \mu_2 + \alpha_2 \Delta y_{t-1}) 1_{\{\Delta y_{t-1} \geq \lambda\}} + e_t,$$

and e_t iid $N(0, 1)$. The sample size we use is $T = 100$. We examine nominal 5% size tests based on estimation of model (10) using bootstrap critical values, the latter calculated using 500 bootstrap replications. All calculations are empirical rejection frequencies from 10,000 Monte Carlo replications, and in all experiments the tests are based on least-squares estimation of the unrestricted model (10).

We first examined the size of the bootstrap tests (used on the unconstrained estimates and the unit-root-constrained estimates). Under the null hypothesis of no threshold, data are generated by the AR(1) process $\Delta y_t = \rho y_{t-1} + \alpha \Delta y_{t-1} + e_t$. We explored how the size is affected by the parameters ρ and α . The results are presented in Table I.

For all cases considered, the size of both tests is excellent. Interestingly, the two bootstrap procedures have near identical size in our simulations, with the unit-root-constrained bootstrap being slightly liberal in some cases, and the unconstrained bootstrap being slightly more conservative in some cases. This evidence suggests that it might not matter much which procedure is used; however, our recommendation is to compute both procedures and take the more conservative results.

Next, we explore the power of the test against local alternatives. Because of the minor differences between the two bootstrap procedures, we calculate the power using the unconstrained bootstrap method. We consider three alternatives allowing $\mu_1 \neq \mu_2$, $\rho_1 \neq \rho_2$, and $\alpha_1 \neq \alpha_2$ separately. The first alternative allows a switching intercept:

$$\Delta y_t = \rho y_{t-1} + \mu_1 1_{\{\Delta y_{t-1} < \lambda\}} + \mu_2 1_{\{\Delta y_{t-1} \geq \lambda\}} + \alpha \Delta y_{t-1} + e_t,$$

TABLE I
SIZE OF 5% BOOTSTRAP THRESHOLD TESTS

	Unconstrained Bootstrap				Constrained Bootstrap			
	$\rho = -.25$	$\rho = -.15$	$\rho = -.05$	$\rho = 0$	$\rho = -.25$	$\rho = -.15$	$\rho = -.05$	$\rho = 0$
$\alpha = -.5$.038	.055	.051	.048	.060	.054	.041	.059
$\alpha = .5$.040	.050	.049	.042	.043	.047	.044	.058

Note: $T = 100$. Nominal size 5%. Rejection frequencies from 10,000 replications.

TABLE II
POWER OF 5% BOOTSTRAP THRESHOLD TEST

	Change in μ		
	$\Delta\mu = .2$	$\Delta\mu = 1.0$	$\Delta\mu = 2.0$
$\rho = -.05$.054	.389	.982
$\rho = 0$.052	.357	.979
	Change in ρ		
	$\Delta\rho = -.05$	$\Delta\rho = -.10$	$\Delta\rho = -.20$
$\rho_1 = -.05$.165	.430	.738
$\rho_1 = 0$.821	.931	.955
	Change in α		
	$\Delta\alpha = .5$	$\Delta\alpha = 1.0$	$\Delta\alpha = 1.9$
$\rho = -.05$.157	.575	.996
$\rho = 0$.047	.100	.344

Note: $T = 100$. Nominal size 5%. Rejection rates from 2000 replications.

setting $\alpha = .5$, and $\lambda = 0$, and varying ρ among 0 and $-.05$. We control the size of the threshold effect by varying $\Delta\mu = \mu_2 - \mu_1$ and set $\mu_1 = -\mu_2$ for simplicity. The power of the 5% bootstrap test is presented in the first two rows of Table II.

The second alternative allows a switching slope on y_{t-1} :

$$\Delta y_t = \rho_1 y_{t-1} 1_{\{\Delta y_{t-1} < \lambda\}} + \rho_2 y_{t-1} 1_{\{\Delta y_{t-1} \geq \lambda\}} + \mu + \alpha \Delta y_{t-1} + e_t,$$

setting $\mu = 1$, $\alpha = .5$, and $\lambda = 0$, and varying ρ_1 among 0 and $-.05$. The threshold effect is controlled by $\Delta\rho = \rho_2 - \rho_1$. The power of the 5% bootstrap test is presented in the second section of Table II.

The third alternative allows a switching slope on Δy_{t-1} :

$$\Delta y_t = \rho y_{t-1} + \mu + \alpha_1 \Delta y_{t-1} 1_{\{\Delta y_{t-1} < \lambda\}} + \alpha_2 \Delta y_{t-1} 1_{\{\Delta y_{t-1} \geq \lambda\}} + e_t,$$

with $\mu = 1$, and $\lambda = 0$, and varying ρ among 0 and $-.05$. The threshold effect is controlled by $\Delta\alpha = \alpha_2 - \alpha_1$, and we set $\alpha_1 = \Delta\alpha/2$. The power of the 5% bootstrap test is presented in the third section of Table II.

In all three alternatives, the power of the test is increasing in the size of the threshold effect. Even in the small sample setting of $T = 100$, the power of the test is quite large against moderate alternatives.

4.5. Subset Tests

It is possible that while an unconstrained model of the form (1) may have been estimated by (2), a researcher is interested in testing for the equality of only a subset of the coefficients of θ . We now briefly discuss inference in such cases. It turns out that the correct asymptotic distribution (and bootstrap method) depends on the unknown true properties of the coefficients.

If the goal is to test for the presence of a threshold effect, the relevant null hypothesis is that there is no threshold, in which case $\theta_1 = \theta_2$. This is the same null hypothesis as studied in Theorem 4, and it follows that a subset Wald test will have an asymptotic distribution with a similar form. The bootstrap methods of Section 4.3 can be directly applied to calculate critical values and p -values.

On the other hand, the goal may be to test the equality of some coefficients, taking for granted that some of the other coefficients indeed differ. For example, the goal may be to test the hypothesized equality $H_0: \alpha_1 = \alpha_2$, under the maintained assumption that $\rho_1 \neq \rho_2$. In this case, the asymptotic distribution is quite different than in Theorem 4. Since the truth is that there is a threshold effect, the threshold parameter is consistently estimated by the threshold estimate (Chan (1993)), and the Wald test will have the same asymptotic distribution as if the threshold parameter were known a priori. If the hypothesis concerns the equality $\alpha_1 = \alpha_2$, the asymptotic distribution of the Wald test is chi-square with degrees of freedom equal to the number of coefficients tested for equality. However, if the test concerns the equality $\rho_1 = \rho_2$, the asymptotic distribution will be nonstandard, due to the estimated unit root. The statistical significance of these tests, however, should never be taken as evidence in favor of the existence of the threshold effect, since the latter requires the rejection of the null hypothesis $H_0: \theta_1 = \theta_2$.

5. TESTING FOR UNIT ROOTS AND NONSTATIONARITY

5.1. Test Statistics

In model (1) under Assumption 1, the parameters ρ_1 and ρ_2 control the “stationarity” of the process y_t . A leading case of interest is

$$(11) \quad H_0: \rho_1 = \rho_2 = 0.$$

When (11) holds, then the model (1) can be rewritten as a stationary threshold autoregression in the variable Δy_t , so that y_t is $I(1)$ and can be described as having a “unit root.”

Another case of interest is when the series is stationary and ergodic. In general models of the form (1) the region of stationarity is not completely understood. However in the special case of $p = 1$ the model is stationary if $\rho_1 < 0$, $\rho_2 < 0$, and $(1 + \rho_1)(1 + \rho_2) < 1$. See Chan and Tong (1985). This suggests that the natural alternative to H_0 is

$$H_1: \rho_1 < 0 \quad \text{and} \quad \rho_2 < 0.$$

There is a third case of interest however. This is the intermediate case of a partial unit root:

$$H_2: \begin{cases} \rho_1 < 0 & \text{and} & \rho_2 = 0, \\ & \text{or} & \\ \rho_1 = 0 & \text{and} & \rho_2 < 0. \end{cases}$$

If H_2 holds, then the process y_t will behave like a unit root process in one regime, but will behave like a stationary process in the other. Under H_2 , the process is nonstationary, but it is not a classic unit root process. In applications, it will be interesting to distinguish between the cases H_0 , H_1 , and H_2 .

We now discuss possible tests to discriminate between these cases. The standard test for (11) against the unrestricted alternative $\rho_1 \neq 0$ or $\rho_2 \neq 0$ is the Wald statistic from (3). This statistic is

$$R_{2T} = t_1^2 + t_2^2$$

where t_1 and t_2 are the t ratios for $\hat{\rho}_1$ and $\hat{\rho}_2$ from the OLS regression (3). As the alternatives H_1 and H_2 are one-sided, however, this two-sided Wald statistic (hence the subscript “2”) is ill-focused and thus may have less power than a one-sided version. While it is unclear (in our context) how to form an optimal⁴ one-sided Wald test, it seems prudent for the test to focus on negative values of $\hat{\rho}_1$ and $\hat{\rho}_2$. Hence, we consider the simple one-sided Wald statistic

$$R_{1T} = t_1^2 1_{\{\hat{\rho}_1 < 0\}} + t_2^2 1_{\{\hat{\rho}_2 < 0\}},$$

which is testing H_0 against the one-sided alternative $\rho_1 < 0$ or $\rho_2 < 0$.

Both tests R_{1T} and R_{2T} will have power against both alternatives H_1 and H_2 . Thus while a “significant” test statistic can justify the rejection of the unit root hypothesis, it cannot discriminate between the stationary case H_1 and the partial unit root case H_2 . This calls for a test focused on the stationary alternative H_1 . We suggest examining the individual t statistics t_1 and t_2 . To retain the convention that the test rejects for large values of the statistic, we will actually consider the negative of the t statistics, vis., $-t_1$ and $-t_2$. If only one of $-t_1$ or $-t_2$ is statistically significant, this would be consistent with the partial unit root case H_2 , allowing us to distinguish between H_0 , H_1 , and H_2 .

All the above test statistics are continuous functions of the t ratios t_1 and t_2 . To unify the presentation, we therefore consider the class of all test statistics

$$R_T = R(t_1, t_2)$$

where $R(x_1, x_2)$ is a continuous function of x_1 and x_2 . We presume that $R(\cdot, \cdot)$ has been normalized so that H_0 is rejected for large values of R_T , as is true for the specific tests described above.

We have described a class of test statistics R_T for H_0 against H_1 and H_2 , and have suggested that H_0 should be rejected for significantly large values of R_T . To determine “significance” we need the sampling distribution of the test under H_0 . We develop appropriate approximations in the next sections.

⁴Andrews (1998) shows how to construct optimal one-sided tests in the context where the unrestricted estimators have asymptotic normal distributions. It is not clear if these results extend to the nonstandard case of unit root distributions, and we do not pursue this extension in this paper.

5.2. *Asymptotic Distribution*

We now derive large sample approximations to the distribution of the test R_T under the null hypothesis of a unit root (11). A difficulty arises in specifying the threshold effect, as the null of a unit root ($\rho_1 = \rho_2 = 0$) is compatible with either the existence ($\theta_1 \neq \theta_2$) or nonexistence ($\theta_1 = \theta_2$) of a threshold effect. It turns out that the asymptotic distributions are different in these two cases. Since the truth is typically unknown we consider both.

5.2.1. *Unidentified Threshold*

We first examine the case that there is no threshold effect.

THEOREM 5: *Under Assumption 1 and $\theta_1 = \theta_2$, then*

$$(t_1, t_2) \Rightarrow (t_1(u^*), t_2(u^*))$$

and

$$R_T \Rightarrow R(t_1(u^*), t_2(u^*)) \leq \sup_{u \in [\pi_1, \pi_2]} R(t_1(u), t_2(u))$$

where

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

$T(u)$ is defined as in (7),

$$t_1(u) = \frac{\int_0^1 W^*(s) dW(s, u)'}{\left(u \int_0^1 W^*(s)^2 ds\right)^{1/2}},$$

$$t_2(u) = \frac{\int_0^1 W^*(s)(dW(s, 1) - dW(s, u))}{\left((1-u) \int_0^1 W^*(s)^2 ds\right)^{1/2}},$$

$W(s, u)$ is a two-parameter Brownian motion, $W(s) = W(s, 1)$, and

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

Several facts about this limiting distribution are interesting. The distributions of the t statistics are the random functions $t_1(u)$ and $t_2(u)$ evaluated at the random argument u^* . The distributions of the random functions $t_1(u)$ and $t_2(u)$ do not depend on any nuisance parameters. By symmetry, we can see that the pointwise distribution of $t_1(u)$ is the same as that of $t_2(1-u)$. But since the random maximizer u^* depends on the nuisance parameter function $\Omega(u)$, so does the limiting distribution of the t statistics, and hence any test constructed from the t statistics. A bound, however, can be obtained by maximizing over the

argument u . This bound is free of nuisance parameters other than the trimming range $[\pi_1, \pi_2]$ and hence can be tabulated (although it depends on the particular functional $R(\cdot, \cdot)$). Critical values for the bound for several choices of $R(\cdot, \cdot)$ and trimming ranges are reported in Table III. Due to the symmetry of $t_1(u)$ and $t_2(1 - u)$ the asymptotic bounds for t_1 and t_2 are the same under symmetric trimming. The critical values were calculated by simulation from the asymptotic formula in Theorem 5. The simulated draws approximated the stochastic integrals using a grid with 10,000 steps over the argument s and 100 steps over the argument u . The critical values were computed as the empirical quantiles from 100,000 independent draws from these distributions.

Also reported in Table III are p -value functions based on chi-square approximation and computed using the methods of Hansen (1997a). The approximations take the form

$$c_0 + c_1 R_T + c_2 R_T^2 \sim \chi_q^2,$$

with (c_0, c_1, c_2, q) as free parameters. The approximations can be used to compute asymptotic p -values for the statistics R_T , by using the χ_q^2 distribution on $c_0 + c_1 R_T + c_2 R_T^2$. For the R_{1T} and R_{2T} asymptotic distributions, the

TABLE III
ASYMPTOTIC CRITICAL VALUE BOUNDS FOR UNIT ROOT TESTS UNIDENTIFIED CASE

		Demeaned Case, $r_t = 1$							
		Critical Values				p -Value Function			
	$[\pi_1, \pi_2]$	20%	10%	5%	1%	c_0	c_1	c_2	q
R_{1T}	[.15, .85]	8.78	10.84	12.75	16.97	1.113	1.130		8
	[.10, .90]	9.01	11.09	13.00	17.23	0.959	1.119		8
	[.05, .95]	9.26	11.35	13.29	17.51	0.784	1.107		8
R_{2T}	[.15, .85]	9.23	11.31	13.24	17.50	-0.011	1.064		7
	[.10, .90]	9.55	11.66	13.59	17.85	-0.262	1.054		7
	[.05, .95]	9.93	12.04	14.03	18.24	-0.572	1.044		7
$-t_1, -t_2$	[.15, .85]	2.61	2.97	3.26	3.82	1.476	-0.023	1.048	6
	[.10, .90]	2.66	3.01	3.31	3.85	1.212	-0.562	1.070	5
	[.05, .95]	2.71	3.05	3.34	3.89	1.044	1.636	1.040	11
		Detrended Case, $r_t = (1 - \rho)^t$							
		Critical Values				p -Value Function			
	$[\pi_1, \pi_2]$	20%	10%	5%	1%	c_0	c_1	c_2	q
R_{1T}	[.15, .85]	8.78	10.84	12.75	16.97	0.456	1.104		10
	[.10, .90]	9.01	11.09	13.00	17.23	0.282	1.098		10
	[.05, .95]	9.26	11.35	13.29	17.51	0.102	1.091		10
R_{2T}	[.15, .85]	9.23	11.31	13.24	17.50	-0.285	1.043		9
	[.10, .90]	9.55	11.66	13.59	17.85	-0.020	1.092		10
	[.05, .95]	9.93	12.04	14.03	18.24	-0.350	1.085		10
$-t_1, -t_2$	[.15, .85]	2.61	2.97	3.26	3.82	6.479	3.382	0.975	22
	[.10, .90]	2.66	3.01	3.31	3.85	5.930	3.742	1.006	22
	[.05, .95]	2.71	3.05	3.34	3.89	4.963	3.960	0.986	22

Note: Calculated from 100,000 simulations.

quadratic term was unnecessary for an accurate approximation (reported p -values within 0.003 of actual), so c_2 was set to zero, but the quadratic term was necessary for the $-t$ statistics.

5.2.2. *Identified Threshold*

We now assume that there is a threshold effect, or that $\theta_1 \neq \theta_2$, in which case λ_0 is identified. We also assume that $E\Delta y_t = 0$, which holds in model (1) under Assumption 1 if $\mu_1 P(Z_{t-1} < \lambda) + \mu_2 P(Z_{t-1} \geq \lambda) = 0$. (If $E\Delta y_t \neq 0$, then a time trend should be included in the model (1) and the following results still hold, with Δy_t replaced by $\Delta y_t - E\Delta y_t$.)

Under the unit root null, Δy_t is strictly stationary and geometrically ergodic. Let

$$(12) \quad \sigma_y^2 = \sum_{k=-\infty}^{\infty} E(\Delta y_t \Delta y_{t+k})$$

denote its long-run variance, and define the long-run correlations

$$(13) \quad \delta_1 = \frac{\sum_{k=-\infty}^{\infty} E(e_t 1_{\{Z_{t-1} < \lambda_0\}} \Delta y_{t+k})}{(E(e_t^2)G(\lambda_0)\sigma_y^2)^{1/2}},$$

$$(14) \quad \delta_2 = \frac{\sum_{k=-\infty}^{\infty} E(e_t 1_{\{Z_{t-1} \geq \lambda_0\}} \Delta y_{t+k})}{(E(e_t^2)(1 - G(\lambda_0))\sigma_y^2)^{1/2}},$$

which satisfy the inequality $\delta_1^2 + \delta_2^2 \leq 1$. Roughly, $\delta_1^2 + \delta_2^2$ is smaller when the threshold effect is stronger. To see this, note that in the limiting case of no threshold effect, Δy_t is a linear function of lagged values of $e_t 1_{\{Z_{t-1} < \lambda_0\}}$ and $e_t 1_{\{Z_{t-1} \geq \lambda_0\}}$, so we find that $\delta_1^2 + \delta_2^2 = 1$.

Let " $X \ll Y$ " denote the X is first-order stochastically dominated by Y , meaning that for all x , $P(X \leq x) \geq P(Y \leq x)$.

THEOREM 6: *Under Assumption 1 and $\theta_1 \neq \theta_2$, if $E\Delta y_t = 0$ and $\sigma_y^2 > 0$, then*

$$-t_1 \Rightarrow (1 - \delta_1^2)^{1/2} Z_1 + \delta_1 DF \ll DF,$$

and

$$-t_2 \Rightarrow (1 - \delta_2^2)^{1/2} Z_2 + \delta_2 DF \ll DF$$

where

$$(15) \quad \begin{pmatrix} Z_1 \\ Z_2 \end{pmatrix} \sim N\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 & \sigma_{21} \\ \sigma_{21} & 1 \end{pmatrix}\right)$$

is independent of

$$DF = - \frac{\int_0^1 W^* dW}{\int_0^1 W^{*2}}$$

(the negative of the conventional detrended Dickey-Full t distribution), and

$$(16) \quad \sigma_{21} = \frac{-\delta_1 \delta_2}{\sqrt{(1 - \delta_1^2)(1 - \delta_2^2)}}.$$

Also,

$$R_{2T} \Rightarrow \chi_1^2 + \left((1 - a^2)^{1/2} Z + aDF \right)^2 \ll \chi_1^2 + DF^2$$

where χ_1^2 is chi-square with one degree of freedom and is independent of $Z \sim N(0, 1)$, and

$$(17) \quad a = \sqrt{\delta_1^2 + \delta_2^2} \in (0, 1].$$

Theorem 6 shows that if the threshold is identified, the two t ratios are asymptotically linear combinations of normal and Dickey-Fuller variates. The distributions depend on the unknown mixing parameters δ_1 and δ_2 , but the result provides a useful bound on the asymptotic distribution, as the mixture distribution is stochastically dominated by the standard Dickey-Fuller t distribution, and so the Dickey-Fuller provides a conservative bound.

Since $R_T = R(t_1, t_2)$ is a continuous function of the arguments (t_1, t_2) , Theorem 6 can be used in principle to give an expression for the asymptotic distribution of the test R_T . For some functions, such as the one-sided Wald test R_{1T} , these expressions do not appear to be very useful. For the two-sided Wald statistic R_{2T} , however, we have found a useful expression and bound, which is reported in Theorem 6. The limiting distribution takes a mixture form that can be bounded by the sum of the squared Dickey-Fuller and chi-square distributions. This bound is free of nuisance parameters, and can be calculated numerically. We found the 10%, 5%, and 1% critical values to be 11.17, 13.12, and 17.29, respectively.

Theorems 5 and 6 together give asymptotic approximations to the null distribution of the TAR unit root tests R_T under differing assumptions concerning the threshold. The source of the difference 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 samples, and R_T inherits the randomness from $\hat{\lambda}$. In contrast, when there is a threshold effect, then λ is identified and $\hat{\lambda}$ will be close to the true value λ_0 in large samples. In this case the asymptotic distribution of R_T is equivalent to the case where λ_0 is known.

5.3. Bootstrap

While the distributions of R_T have asymptotic approximations, improved finite sample inference may be conducted using a bootstrap distribution. One has to be careful, however, as there is not a unique bootstrap distribution. Most importantly, it is possible to construct a bootstrap distribution that imposes an identified threshold effect or imposes an unidentified threshold effect. Theorems 5 and 6 show that the asymptotic distribution of R_T is different in these two cases, implying that the bootstrap distribution will likely differ substantially as well. In this section, we discuss how to calculate these two bootstrap distributions, and in the next section compare their performance using Monte Carlo methods.

The unidentified threshold bootstrap imposes the restrictions $\theta = \theta_1 = \theta_2$ (no threshold) and $\rho = 0$ (unit root). This can be done using the constrained bootstrap method introduced in Section 4.3, since the null hypothesis is identical. To repeat, let $(\tilde{\rho}, \tilde{\alpha}, \tilde{F})$ be estimates of (ρ, α, F) discussed in Section 4.3 ($\tilde{\rho}, \tilde{\alpha}$ are obtained by a linear autoregression, and \tilde{F} is the empirical distribution of the OLS residuals). Then generate random samples from the model $\Delta y_t^b = \tilde{\alpha}' \tilde{\Delta y}_{t-1}^b + e_t^b$ with e_t^b drawn randomly from \tilde{F} , and for each sample y_t^b calculate the test statistic R_T^b . The estimated bootstrap p -value is the percentage of simulated R_T^b that exceed the observed R_T .

The identified threshold bootstrap, on the other hand, requires simulation from a unit root TAR process. Let $(\tilde{\rho}_1, \tilde{\beta}_1, \tilde{\alpha}_1, \tilde{\rho}_2, \tilde{\beta}_2, \tilde{\alpha}_2, \hat{\lambda}, \hat{F})$ be the estimates obtained from the unrestricted model. Then generate samples y_t^b from model (1) using the restricted estimates $(0, \tilde{\beta}_1, \tilde{\alpha}_1, 0, \tilde{\beta}_2, \tilde{\alpha}_2, \hat{\lambda}, \hat{F})$. Again, for each sample calculate R_T^b , and estimate the bootstrap p -value by the percentage of R_T^b that exceed R_T .

5.4. Monte Carlo Experiment

5.4.1. Size

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

We first study the size of nominal 5% tests. The data are simulated under the null from model (1) with $k = 1$, $m = 1$, setting $\rho_1 = \rho_2 = 0$. For simplicity,⁵ we allow for a threshold effect in the intercept μ , but not the AR lag α . Thus the null model is

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

We use samples of size $T = 100$ and generate e_t as iid $N(0, 1)$. For simplicity, we set $\mu_1 = -\mu_2$, and denote the size of the threshold effect as $\Delta\mu = \mu_2 - \mu_1$. We

⁵We performed a limited set of experiments with a switching AR slope α and with samples of size 200, and found similar results.

set $\lambda = 0$ and vary $\Delta\mu$ among $\{0, 1, 2, 3\}$. Note that when $\Delta\mu = 0$, the threshold λ is not identified. We vary α among $\{-.5, -.2, 0, .2, .5\}$.

The tests compared are the two-sided Wald test R_{2T} , the one-sided Wald test R_{1T} , the individual t ratio t_1 (the size for t_2 is similar and omitted), and the conventional ADF test based on a fitted AR(2) with linear time trend. The asymptotic critical values are taken from Table III for the R_{2T} , R_{1T} , and t_1 tests. Also the asymptotic critical values for R_{2T} from Theorem 6 (the ‘‘Identified Case’’) are included for comparison.

Table IV reports rejection frequencies from 1000 Monte Carlo replications. In this experiment, we can see that both the t_1 and ADF tests have reasonable size, at least for smaller threshold effects (smaller $\Delta\mu$). The R_{1T} and R_{2T} tests, however, substantially over-reject for some parameter configurations. For the R_{2T} test, the critical values from Table III (the ‘‘Unidentified Case’’) perform much better than those from Theorem 6 (the ‘‘Identified Case’’), and based on these results we recommend that practitioners use the critical values from Table III for asymptotic inference.

Due to the substantial size distortions, we explored the performance of the two bootstrap methods described in Section 5.3. Due to the substantial computational costs, we restricted α to $\{-.5, 0, .5\}$. The results are reported in Table V. From these results, we can see that both bootstrap procedures have meaningful size distortions, but are substantially reduced relative to the size distortions of the asymptotic tests. The rejection rates using the unidentified threshold model appear to be less sensitive to the nuisance parameters, and are our preferred choice.

Based on this information, our recommendation is to calculate p -values using the unidentified threshold bootstrap. For a quick calculation, an asymptotic p -value may be calculated from Theorem 3, but is not as reliable.

TABLE IV
SIZE OF ASYMPTOTIC UNIT ROOT TESTS

		R_{2T} Test								R_{1T} Test			
		Unidentified Case				Identified Case							
α	$\Delta\mu$:	0	1	2	3	0	1	2	3	0	1	2	3
.5		8.4	9.1	20.5	22.7	21.2	22.1	35.6	34.3	8.5	10.8	24.7	21.8
.2		7.8	6.5	8.5	24.4	18.7	19.0	21.7	39.1	7.2	7.9	10.4	26.0
0		7.4	7.5	7.3	13.9	17.0	16.3	17.7	26.1	7.0	5.6	8.2	13.3
-.2		7.9	7.3	6.1	7.2	19.9	15.5	16.6	17.7	5.7	7.0	6.9	9.6
-.5		7.6	5.2	6.2	7.3	17.5	18.9	18.8	15.4	5.9	6.7	7.9	8.7
		t_1 test				ADF test							
α	$\Delta\mu$:	0	1	2	3	0	1	2	3				
.5		4.2	7.0	7.0	0.4	6.9	5.2	7.6	11.2				
.2		5.4	2.7	1.4	4.2	5.5	5.2	3.0	7.8				
0		5.2	4.7	3.2	2.4	6.1	5.9	4.0	1.6				
-.2		4.5	5.3	3.7	2.0	7.5	7.7	4.8	1.8				
-.5		3.1	4.6	4.5	4.1	6.1	5.8	6.1	2.5				

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

TABLE V
SIZE OF BOOTSTRAP UNIT ROOT TESTS

$\Delta\mu$:	R_{2T} Test								R_{1T} Test			
	Unidentified Case				Identified Case				Unidentified Case			
	0	1	2	3	0	1	2	3	0	1	2	3
$\alpha = .5$	6.2	4.0	8.2	11.0	9.0	7.1	13.9	14.1	5.6	3.9	7.1	9.2
$\alpha = 0$	4.4	3.8	1.1	2.2	6.6	4.6	1.5	3.2	4.8	2.0	1.8	1.6
$\alpha = -.5$	6.0	3.3	1.4	0.7	5.3	3.3	0.7	0.8	4.4	2.5	1.7	0.5

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

5.4.2. Power

We next explore the power of the tests. The model and tests are the same as before except that we fix the serial correlation parameter at $\alpha = 0$ and do not impose $\rho_1 = \rho_2 = 0$. We report size-adjusted power for the R_{2T} and R_{1T} tests (rejection rates based on finite sample critical values) to control for the size distortions reported in Table IV.

We consider three experiments. In the first, we restrict $\rho_1 = \rho_2$, and vary ρ_2 among $\{-.05, -.10, -.15\}$ and $\Delta\mu$ among $\{0, 1, 2, 3\}$. This is the setting that is 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 the first section of Table VI. When $\Delta\mu = 0$ there is no threshold effect and the ADF test has considerably more power than the threshold unit root tests. As $\Delta\mu$ is increased, however, the R_{2T} and R_{1T} tests gain more power than the ADF test, and the relative ranking switches. The R_{1T} test has slightly more power than the R_{2T} test, and the t tests have significantly less power.

For our second power experiment, we allow for a threshold effect in ρ , setting $\rho_1 = 0$ and letting ρ_2 vary among $\{-.05, -.10, -.15\}$ and $\Delta\mu$ as above. This is a partial unit root model. The results are presented in the second section of Table VI. In this experiment, the R_{2T} and R_{1T} tests have substantially greater power than the ADF test in most parameterizations. In particular, the ADF suffers when $\Delta\mu$ is large. The t_2 test has even better power than R_{1T} and R_{2T} . The rejection rate of t_1 is close to the nominal size, which means that the individual t ratio tests can help discriminate between the pure unit root, partial unit root, and stationary cases.

For our third power experiment, we set $\rho_1 = -.05$ and vary ρ_2 and $\Delta\mu_2$ as above. This is a stationary model. Across most parameterizations, R_{1T} has the best power, with R_{2T} a close second.

As expected, these calculations show that in the presence of threshold effects, the threshold unit root tests have good power relative to the ADF test. In most cases, the one-sided Wald test R_{1T} has somewhat better power than two-sided version R_{2T} . The individual t ratio tests are able to successfully distinguish between the pure unit root, partial unit root, and stationary cases.

TABLE VI
POWER OF UNIT ROOT TESTS

ρ_2 :	R_{2T}			R_{1T}			t_1			t_2			ADF		
	-.05	-.1	-.15	-.05	-.1	-.15	-.05	-.1	-.15	-.05	-.1	-.15	-.05	-.1	-.15
$\rho_1 = \rho_2$															
$\Delta\mu$															
0	6	11	21	8	15	28	5	12	20	6	13	22	8	18	35
1	12	33	62	14	38	62	9	26	26	10	24	47	14	36	71
2	27	73	96	29	76	95	15	50	50	14	53	83	12	69	89
3	53	96	99	54	96	99	17	65	65	16	70	96	22	81	99
$\rho_1 = 0$															
0	6	11	17	6	12	23	3	4	3	9	17	30	5	7	26
1	11	25	53	11	28	54	3	5	7	9	32	58	6	11	45
2	20	69	92	17	64	92	3	6	9	17	67	93	4	16	39
3	44	88	97	41	90	97	3	3	3	32	87	96	3	9	15
$\rho_1 = -.05$															
0	6	10	20	8	12	20	5	6	7	6	15	25	8	12	17
1	12	23	42	14	28	47	9	13	14	10	25	49	14	33	36
2	27	58	85	29	58	87	15	21	32	14	49	83	12	41	63
3	53	87	99	54	89	99	17	21	30	16	79	98	22	67	85

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

6. U.S. UNEMPLOYMENT RATE

Our application is to the U.S. unemployment rate⁶ among adult males, monthly from January, 1956 through August, 1999. A plot is given in Figure 1.

To establish a baseline, we first fit by OLS a linear model with $k = 12$ lagged differences. The point estimate for ρ is $\hat{\rho} = -0.014$. Its t statistic (the ADF test for a unit root) is insignificant at -2.40 (the 5% asymptotic critical value is -2.86 .) This leads to the standard conclusion that the linear representation for the unemployment rate has a unit root.

Our first question is to ask if there is any statistical evidence to reject the linear AR model in favor of a threshold model. An appropriate test statistic for this question is the Wald test W_T of Section 4.1. In Table VII we report the Wald tests W_T , 1% bootstrap critical values, and bootstrap p -values for threshold variables of the form $Z_t = y_t - y_{t-m}$ for delay parameters m from 1 to 12. Each statistic is highly significant and easily rejects the null hypothesis of linearity in favor of the threshold model.

Since the W_T test rejects the null of no threshold for practically any choice of m , it appears certain 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

⁶The series is created by dividing the Citibase file LHMU (Adult Male Unemployment) by LHMC (Adult Male Labor Force), and is scaled to range from 0 to 100.

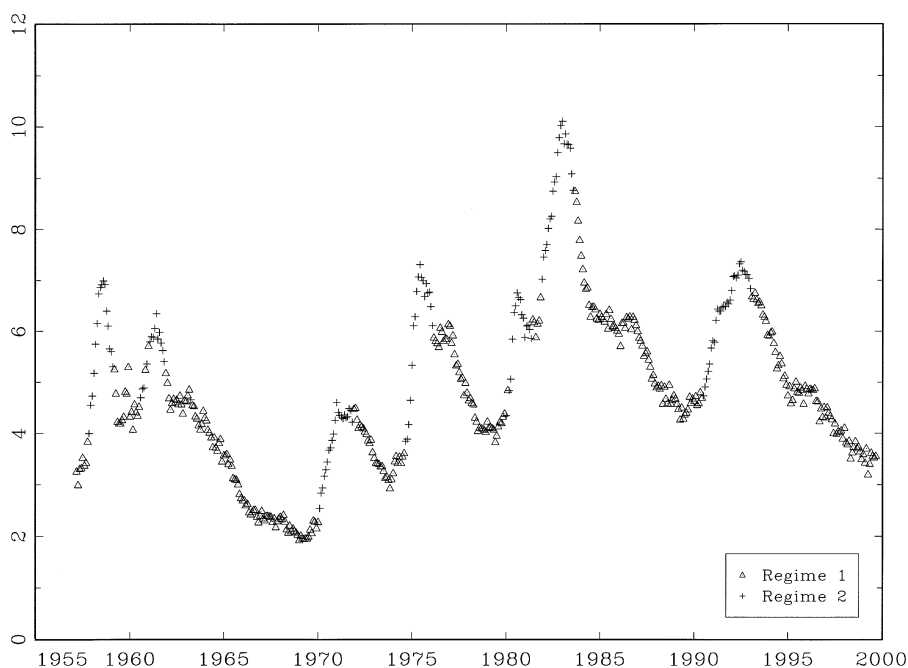


FIGURE 1.—U.S. adult male unemployment rate, classified by threshold regime.

squares estimate of m is the value that 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 that maximizes W_T . This estimate is $\hat{m} = 12$, corresponding to the threshold test statistic of $W_T = 80.4$. The reported bootstrap p -value of 0.000 in Table VII assumes that m is known and fixed. It is

TABLE VII
THRESHOLD AND UNIT ROOT TESTS UNCONSTRAINED MODEL

m	Bootstrap Threshold Test			Unit Root Tests, p -Values					
	W_T	1% C.V.	p -Value	R_{1T}		t_1		t_2	
				Asym.	Boot.	Asym.	Boot.	Asym.	Boot.
1	34.9	39.3	0.034	0.091	0.052	0.254	0.104	0.351	0.138
2	53.2	40.1	0.000	0.148	0.084	0.713	0.362	0.157	0.589
3	35.5	38.8	0.027	0.057	0.064	0.089	0.037	0.566	0.252
4	42.7	39.5	0.005	0.071	0.042	0.068	0.026	0.747	0.397
5	54.1	39.3	0.001	0.054	0.034	0.029	0.013	0.925	0.634
6	62.2	39.3	0.000	0.069	0.042	0.080	0.033	0.681	0.341
7	48.5	39.3	0.001	0.131	0.078	0.113	0.046	0.793	0.438
8	70.0	39.2	0.000	0.095	0.058	0.056	0.025	0.909	0.608
9	77.8	39.4	0.000	0.042	0.029	0.036	0.015	0.786	0.435
10	75.9	39.1	0.000	0.056	0.038	0.065	0.027	0.681	0.348
11	67.8	38.8	0.000	0.086	0.058	0.096	0.040	0.693	0.360
12	80.4	38.7	0.000	0.105	0.072	0.141	0.057	0.619	0.303

Note: Bootstrap p -values calculated from 10,000 replications.

easy, however, to incorporate estimation of m into the calculation of bootstrap p -values. We can recalculate the bootstrap p -value allowing for the estimation of m , and when we do so, we still calculate a bootstrap p -value of 0.000, implying that it is extremely unlikely that the linear AR model (1) could generate a test statistic this large. We conclude that there is very strong evidence for a TAR model.

While the LS point estimate for the delay parameter is $\hat{m} = 12$, the choice $m = 9$ yields a near-identical value for the residual sum-of-squares and hence test statistic W_T , as seen from Table VII. This means that $m = 9$ is an equivalently good statistical choice, and all else held equal, we prefer models with smaller delay parameters, leading us to take $m = 9$ as our preferred model specification.

Our second question concerns the presence of a unit root. We calculate the threshold unit root test statistics R_{1T} , R_{2T} , t_1 and t_2 for each delay parameter m from 1 to 12, and in Table IX report both asymptotic and bootstrap p -values for R_{1T} , t_1 and t_2 . (The R_{2T} test results are nearly identical to the R_{1T} test.) The asymptotic p -value bounds are calculated using the p -value functions reported in Table III (demeaned case), and the bootstrap p -values are calculated using the “unidentified threshold bootstrap” described in Section 5.3. For all cases, the asymptotic p -value bounds are more conservative than the bootstrap p -values, but not dramatically so. The bootstrap calculations suggest that all twelve R_{1T} statistics are significant at the 10% level, and four at the 5% level. The most relevant statistics are for the $m = 9$ and $m = 12$ cases, which have bootstrap p -values of 0.03 and 0.07, respectively.

Turning to the individual t ratios t_1 and t_2 , we see that the bootstrap p -values for t_1 are 0.015 and 0.057 for the $m = 9$ and $m = 12$ specifications, respectively, giving strong evidence that indeed we can reject the unit root hypothesis in favor of $\rho_1 < 0$. The t_2 statistic is statistically insignificant, so we are unable to reject that $\rho_2 = 0$.

For our preferred specification of $m = 9$, we present the LS parameter estimates in Table VIII. The point estimate of the threshold $\hat{\lambda}$ is 0.33. Thus the TAR splits the regression function depending on whether the variable $Z_{t-1} = y_{t-1} - y_{t-10}$ lies above or below 0.33. The first regime is when $Z_{t-1} < 0.33$, which occurs when the unemployment rate has fallen, remained constant, or has risen by less than .33 points (e.g. from 5.40 to 5.73) over a nine-month period. Approximately 73% of the observations fall in this category. The second regime is when $Z_{t-1} > 0.33$, which occurs when the unemployment rate has risen by more than .33 points over a nine month period. Approximately 27% of the observations fall in this regime.

In addition to the parameter estimates, we report in Table VIII tests for the pairwise equality of individual coefficients, and bootstrap p -values based on the null of no threshold, which is the procedure described in Section 4.5. (Conditional on the presence of a threshold effect, a χ_1^2 asymptotic approximation is also appropriate, which implies a 5% critical values of 3.84.) Looking at the point estimates and test results, it appears that the coefficients on Δy_{t-1} and

TABLE VIII
LEAST SQUARES ESTIMATES UNCONSTRAINED THRESHOLD MODEL

Regressor	Estimates $\hat{m} = 9, \hat{\lambda} = 0.33$				Tests for Equality of Individual Coefficients	
	$Z_{t-1} < \hat{\lambda}$		$Z_{t-1} \geq \hat{\lambda}$		Wald Statistics	Bootstrap <i>p</i> -Value
	Estimate	s.e.	Estimate	s.e.		
Constant	0.075	(0.032)	0.195	(0.060)	3.3	0.367
y_{t-1}	-0.024	(0.007)	-0.014	(0.011)	0.1	0.887
Δy_{t-1}	-0.163	(0.054)	0.109	(0.081)	21.5	0.000
Δy_{t-2}	0.036	(0.054)	0.346	(0.078)	7.1	0.068
Δy_{t-3}	0.046	(0.053)	0.146	(0.083)	0.7	0.572
Δy_{t-4}	0.090	(0.055)	0.012	(0.076)	1.86	0.359
Δy_{t-5}	0.030	(0.054)	-0.003	(0.084)	0.1	0.808
Δy_{t-6}	-0.002	(0.054)	-0.191	(0.089)	0.7	0.567
Δy_{t-7}	0.010	(0.055)	-0.189	(0.087)	3.4	0.206
Δy_{t-8}	-0.018	(0.054)	-0.201	(0.089)	0.3	0.692
Δy_{t-9}	-0.011	(0.052)	0.008	(0.090)	0.3	0.700
Δy_{t-10}	-0.021	(0.050)	0.164	(0.081)	3.5	0.214
Δy_{t-11}	0.091	(0.050)	0.015	(0.080)	3.4	0.217
Δy_{t-12}	-0.197	(0.050)	-0.231	(0.078)	3.5	0.218

Δy_{t-2} are driving the threshold model, with the other coefficients either less important or invariant across regimes. To verify this conjecture, we compute a joint Wald test for the equality of the coefficients on Δy_{t-3} through Δy_{t-12} , yielding a test statistic of 16.0, with a bootstrap *p*-value of 0.448, suggesting that this restriction is compatible with the data. Imposing this constraint, we re-estimate the model and report the results in Table IX. As expected, the estimates are qualitatively quite similar to those in Table VIII. In particular, the

TABLE IX
LEAST SQUARES ESTIMATES CONSTRAINED THRESHOLD MODEL

Regressor	$\hat{m} = 9, \hat{\lambda} = 0.33$			
	$Z_{t-1} < \hat{\lambda}$		$Z_{t-1} \geq \hat{\lambda}$	
	Estimate	s.e.	Estimate	s.e.
Constant	0.056	(0.033)	0.193	(0.059)
y_{t-1}	-0.022	(0.007)	-0.025	(0.010)
Δy_{t-1}	-0.200	(0.055)	0.274	(0.072)
Δy_{t-2}	0.052	(0.055)	0.271	(0.075)
Δy_{t-3}		0.063 (0.045)		
Δy_{t-4}		0.050 (0.044)		
Δy_{t-5}		0.018 (0.044)		
Δy_{t-6}		-0.056 (0.044)		
Δy_{t-7}		-0.022 (0.044)		
Δy_{t-8}		-0.024 (0.045)		
Δy_{t-9}		0.026 (0.045)		
Δy_{t-10}		-0.025 (0.045)		
Δy_{t-11}		0.029 (0.043)		
Δy_{t-12}		-0.238 (0.043)		

estimate of the threshold $\hat{\lambda}$ is identical, so the division of the data into regimes is the same as for the unconstrained model. The estimated division of the data into the two threshold regimes is shown in Figure 1.

What is quite striking about the point estimates from Table IX is that the coefficients on y_{t-1} in the two regimes are quite similar, about -0.02 , suggesting that the difference between the two regimes is probably not the “stationarity” of the regime. The major difference in coefficients is the coefficient on Δy_{t-1} , which switches from -0.20 to 0.27 , having a big impact on the first-order serial correlation properties of the series.

Since the constrained model of Table IX has many fewer parameters than the unconstrained model, the threshold and unit root tests applied in this context may have additional precision. These results are presented in Table X. As found in Table VII for the unconstrained model, the threshold tests overwhelmingly reject the null of no threshold effect, and the delay orders 9 and 12 are the best least-squares fits. Some differences are found in the unit root tests, where the results in Table X lend much greater support to the stationarity hypothesis. For $m = 9$, the R_{1T} , t_1 and t_2 tests have bootstrap p -values of 0.004, 0.010, and 0.172, respectively, strongly rejecting the null of a unit root, yet being unable to distinguish between the partial unit root and stationary cases. For $m = 12$ the results are similar, except that the bootstrap p -value for t_2 drops to 0.106, lending more support to stationarity. Combined with the point estimates from Table IX, which suggest that ρ_1 and ρ_2 do not differ greatly, the combined evidence points quite strongly to the hypothesis that both $\rho_1 < 0$ and $\rho_2 < 0$, which means that the unemployment rate series is a stationary threshold process.

To assess robustness with respect to subsamples, the constrained TAR model with $m = 9$ was re-estimated on the two subsamples obtained by splitting the

TABLE X
THRESHOLD AND UNIT ROOT TESTS CONSTRAINED MODEL

m	Bootstrap Threshold Test			Unit Root Tests, p -Values					
	W_T	1% C.V.	p -Value	R_{1T}		t_1		t_2	
				Asym.	Boot.	Asym.	Boot.	Asym.	Boot.
1	20.4	20.0	0.008	0.027	0.023	0.026	0.015	0.745	0.405
2	27.0	20.1	0.001	0.067	0.052	0.177	0.090	0.374	0.188
3	22.7	20.3	0.004	0.060	0.045	0.061	0.033	0.726	0.385
4	27.0	20.1	0.001	0.040	0.032	0.076	0.042	0.493	0.242
5	38.1	20.5	0.000	0.021	0.020	0.020	0.013	0.743	0.407
6	40.0	20.2	0.000	0.018	0.017	0.021	0.014	0.673	0.354
7	34.0	19.9	0.000	0.019	0.018	0.062	0.035	0.338	0.164
8	46.9	20.7	0.000	0.011	0.012	0.032	0.022	0.393	0.193
9	61.1	20.4	0.000	0.003	0.004	0.010	0.010	0.349	0.172
10	56.8	20.8	0.000	0.008	0.010	0.028	0.021	0.347	0.168
11	52.8	20.3	0.000	0.013	0.013	0.048	0.029	0.311	0.157
12	62.6	20.5	0.000	0.012	0.013	0.076	0.046	0.202	0.106

Note: Bootstrap p -values calculated from 10,000 replications.

TABLE XI
SUBSAMPLE COMPARISONS, CONSTRAINED MODEL, $m = 9$

	First Half Sample	Second Half Sample
$\hat{\lambda}$.479	.267
$\hat{\rho}_1$	-.029	-.049
$\hat{\rho}_2$	-.033	-.031
$\hat{\alpha}_{1(1)}$	-.090	-.309
$\hat{\alpha}_{2(1)}$.241	.165
W_T (p -value)	.000	.001
R_{1T} (p -value)	.167	.013
\hat{m}	12	8

Note: Bootstrap p -values calculated from 10,000 replications.

sample at its midpoint (December, 1978). The subsample estimates of the parameters λ , ρ_1 , ρ_2 , and the first elements of α_1 and α_2 , denoted as $\alpha_{1(1)}$ and $\alpha_{2(1)}$, respectively, are reported in Table XI. They appear⁷ to be remarkably stable across the two regimes. We also report the bootstrap p -values for the threshold test W_T and the unit root test R_{1T} . On each subsample, the threshold test W_T easily rejects the null of linearity in favor of threshold nonlinearity. The unit root tests are split, with the first subsample failing to reject the null hypothesis, while the null of a unit root is rejected in the second subsample.

We also assessed robustness with respect to alternative specifications of the dependent variable y_t . In our preferred model, the dependent variable is UR_t , the unemployment rate scaled to range from 0 to 100. By construction, this variable is bounded, and hence cannot strictly be a linear unit root process. It is tempting to think that the boundary effect may bias our results, as the estimated threshold effect may be merely to incorporate this boundary condition. To explore this issue, we experimented with four transformations of the dependent variable that are unbounded in either or both directions. The specific transformations and results are listed in Table XII. For each transformation (setting

TABLE XII
SUMMARY RESULTS FOR ALTERNATIVE SPECIFICATIONS
UNCONSTRAINED MODEL

Dependent Variable	ADF Statistic	Log-Likelihood	W_T p -Value	R_{1T} p -Value
UR_t	-2.40	174	0.000	0.029
$\log(UR_t/(1 - UR_t))$	-2.41	161	0.000	0.078
$\log(UR_t)$	-2.42	159	0.000	0.095
$-100 \log(1 - UR_t/100)$	-2.41	172	0.000	0.025
$100(\exp(UR_t/100) - 1)$	-2.41	172	0.000	0.026

Note: Bootstrap p -values calculated from 10,000 replications.

⁷While it is tempting to attempt a formal test for parameter stability by comparing the estimated parameters, we have no theory that is appropriate in this context and interpretation of the tests could be highly misleading.

$m = 9$), the linear ADF statistic, Gaussian log-likelihood,⁸ and the bootstrap p -values for the W_T and R_{1T} tests are reported. Two important results emerge. First, none of the results are sensitive to the transformation selected. The linear ADF statistic is virtually unchanged, and the p -values of the threshold tests are all overwhelmingly significant. The unit root tests yield minor differences, with small changes in the p -values across specifications. The differences, however, are not sufficient to alter our conclusions. Second, our preferred specification (which sets $y_t = UR_t$) has the highest Gaussian log-likelihood. While we have no formal test to compare the models, the Gaussian log-likelihood is still a valid model selection criteria, and its value certainly does not provide any evidence against our preferred specification.

Furthermore, we explored the sensitivity of our results to the trimming region $[\pi_1, \pi_2]$. Our reported results set $[\pi_1, \pi_2] = [.15, .85]$, but we also tried $[\pi_1, \pi_2] = [.10, .90]$ and $[\pi_1, \pi_2] = [.05, .95]$. The point estimates are unchanged, and the p -values for the threshold test statistics remain as reported. The only difference arises for the p -values for the unit root tests, which increase somewhat. For example, the bootstrap p -value for the R_{1T} statistic rises from 0.029 to 0.035 to 0.043, for the three respective trimming regions. The model was also re-estimated adding a fitted linear time trend to each regime. None of the point estimates or test statistics changed meaningfully, except that the unit root tests were reduced in statistical significance. For example the bootstrap p -value for the R_{1T} statistic rises to 0.103.

The stationarity of the post-war unemployment rate in a TAR model has also been recently explored by Tsay (1997). His conclusions are quite similar to ours, although his methods differ. His analysis is based on quarterly data over 1948–1993, and uses a lagged first difference for the threshold variable. His unit root test imposes the restriction $\rho_1 = \rho_2$, and he compares his unit root t statistic to the standard Dickey-Fuller distribution,⁹ while we base our inferences on a bootstrap distribution.

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—allows a researcher to distinguish nonlinear from nonstationary 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 nonlinear threshold autoregression.

⁸Properly adjusted for the Jacobian of the transformation of the dependent variable.

⁹This is justified in Tsay's paper only for the case of known threshold. An analog of our Theorem 7, however, shows that Tsay's test with an estimated threshold continues to have the standard Dickey-Fuller distribution under the auxiliary assumption that the threshold is identified.

It would be useful to generalize our analysis in several directions, including multivariate models, multiple thresholds, and smooth threshold (STAR) models. Such extensions may require different methods than those presented here.

Dept. of Economics, University of Pittsburgh, WW Posvar Hall 4S01, Pittsburgh, PA 15260, U.S.A.; caner@pitt.edu

and

Dept. of Economics, Social Science Bldg., University of Wisconsin, Madison, WI 53706-1393, U.S.A.; bhansen@ssc.wisc.edu; http://www.ssc.wisc.edu/~bhansen/

Manuscript received December, 1997; final revision received July, 2000.

APPENDIX

PROOF OF THEOREM 1: For simplicity, normalize $\sigma^2 = 1$. For all u , $\{1_{t-1}(u)e_t, \mathfrak{F}_t\}$ is a strictly stationary and ergodic MDS with variance $E(1_{t-1}(u)e_t^2) = E(1_{t-1}(u)) = u$. By the MDS central limit theorem, for any (s, u) ,

$$Y_T(s, u) = \frac{1}{\sqrt{T}} \sum_{t=1}^{[Ts]} 1_{t-1}(u)e_t \rightarrow_d N(0, su).$$

Furthermore, the asymptotic covariance kernel is determined by

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

Combined with the Cramer-Wold device, this can be used to yield the convergence of the finite dimensional distributions.

The stochastic equicontinuity of $Y_T(s, u)$ over Γ follows from that of

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

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

$$\begin{aligned} Y_T^{**} &= Y_T^*(s_2, u_2) - Y_T^*(s_1, u_2) - Y_T^*(s_2, u_1) + Y_T^*(s_1, u_1) \\ &= T^{-1/2} \sum_{t=[Ts_1]}^{[Ts_2]} e_t a_{t-1} \end{aligned}$$

where $a_t = 1_{\{u_1 \leq U_t < u_2\}} - (u_2 - u_1)$.

We show below that for any constant $\tau > 0$ such that

$$(A.1) \quad \frac{\tau}{T} \leq u_2 - u_1 \quad \text{and} \quad \frac{\tau}{T} \leq s_2 - s_1,$$

then

$$(A.2) \quad E|Y_T^{**}|^{2\gamma} \leq K[1 + \tau^{-1}](s_2 - s_1)^\alpha (u_2 - u_1)^\alpha$$

for $K < \infty$ and $\alpha = 1 + 1/r$. Bai (1996, Theorem A.1) established the stochastic equicontinuity of $Y_T^*(s, u)$ in a similar context. While he used a different set of dependence and moment bounds, a careful reading of his proof shows that these conditions are only used to prove inequality (A.2) under (A.1) (his equations (23) and (24)). Thus (A.1)–(A.2) are sufficient to establish that $Y_T(s, u)$ is stochastically equicontinuous.

We now prove (A.2) under (A.1). First, from (A.1) we deduce

$$(A.3) \quad \begin{aligned} T^{-1} &\leq \tau^{-1}(s_2 - s_1)^{1/r}(u_2 - u_1)^{(1-1/r)} \\ &= \tau^{-1}(s_2 - s_1)^{(\alpha-1)}(u_2 - u_1)^{(\alpha-2/r)}. \end{aligned}$$

Second, the uniform distribution of U_{t-1} (Assumption 2.2) implies

$$\begin{aligned} E|a_{t-1}|^2 &= E\left(1_{\{u_1 \leq U_t < u_2\}} - 21_{\{u_1 \leq U_t < u_2\}}(u_2 - u_1) + (u_2 - u_1)^2\right) \\ &= (u_2 - u_1) - 2(u_2 - u_1)(u_2 - u_1) + (u_2 - u_1)^2 \\ &\leq (u_2 - u_1). \end{aligned}$$

Since $|a_{t-1}|^2 \leq 1$, then for any $\psi \geq 2$,

$$(A.4) \quad E|a_{t-1}|^\psi \leq E|a_{t-1}|^2 \leq (u_2 - u_1).$$

Third, set $q_t = |a_{t-1}|^2 - E|a_{t-1}|^2$ and note that $E|q_t|^r \leq 2^{r-1}(u_2 - u_1)$ by the c_r inequality and (A.4). Using the expansion $|a_{t-1}|^2 = E|a_{t-1}|^2 + q_t$, the c_r inequality, and Corollary 3 of Hansen (1991) (which holds under Assumption 2.1 for some $K' < \infty$), we find

$$(A.5) \quad \begin{aligned} E\left(\frac{1}{T} \sum_{t=[Ts_1]}^{[Ts_2]} |a_{t-1}|^2\right)^2 &= E\left((s_2 - s_1)E|a_{t-1}|^2 + \frac{1}{T} \sum_{t=[Ts_1]}^{[Ts_2]} q_t\right)^2 \\ &\leq 2(s_2 - s_1)^2(u_2 - u_1)^2 + 2E\left|\frac{1}{T} \sum_{t=[Ts_1]}^{[Ts_2]} q_t\right|^2 \\ &\leq 2(s_2 - s_1)^2(u_2 - u_1)^2 + \frac{2K'}{T}(s_2 - s_1)(E|q_t|^r)^{2/r} \\ &\leq 2(s_2 - s_1)^2(u_2 - u_1)^2 + \frac{8K'}{T}(s_2 - s_1)(u_2 - u_1)^{2/r}. \end{aligned}$$

Finally, since $\{e_t a_{t-1}, \mathfrak{F}_t\}$ is an MDS, by Rosenthal's inequality (Hall and Heyde (1980, p. 23)) for some $C < \infty$, Assumption 2.3, (A.5), (A.4), and (A.3),

$$\begin{aligned} E|Y_T^{**}|^4 &= E\left|T^{-1/2} \sum_{t=[Ts_1]}^{[Ts_2]} e_t a_{t-1}\right|^4 \\ &\leq C \left[E\left(\frac{1}{T} \sum_{t=[Ts_1]}^{[Ts_2]} E(|e_t a_{t-1}|^2 | \mathfrak{F}_{t-1})\right)^2 + \frac{1}{T^2} \sum_{t=[Ts_1]}^{[Ts_2]} E|e_t a_{t-1}|^4 \right] \\ &= C \left[E\left(\frac{1}{T} \sum_{t=[Ts_1]}^{[Ts_2]} |a_{t-1}|^2\right)^2 + T^{-1}(s_2 - s_1)E|e_t|^4 E|a_{t-1}|^4 \right] \\ &\leq C[2(s_2 - s_1)^2(u_2 - u_1)^2 + 8T^{-1}K'(s_2 - s_1)(u_2 - u_1)^{2/r} \\ &\quad + T^{-1}(s_2 - s_1)\kappa(u_2 - u_1)] \\ &\leq K(1 + \tau^{-1})(s_2 - s_1)^\alpha (u_2 - u_1)^\alpha \end{aligned}$$

where $K = C \max(2, 8K' + \kappa)$, using the fact that $(s_2 - s_1)^2(u_2 - u_1)^2 \leq (s_2 - s_1)^\alpha(u_2 - u_1)^\alpha$. This is (A.2) and completes the proof. *Q.E.D.*

PROOF OF THEOREM 2: For all u , $Y_T(s, u)$ is a martingale with square integrable innovations $\sigma^{-1}T^{-1/2}1_{t-1}(u)e_t$. For fixed u , Theorem 1 above and Theorem 2.2 in Kurtz and Protter (1991) yield the stated result. Their result can be extended to the needed uniformity over $u \in [0, 1]$ as their equation (1.13) holds uniformly in u , which can be verified using the bounded convergence theorem. *Q.E.D.*

PROOF OF THEOREM 3: We prove part 1 under the weaker assumption that

$$(A.6) \quad E|w_t|^{1+\delta/2} < \infty.$$

Parts 2 and 3 follow directly by redefinition of w_t and X_{Tt} .

Let $v_t(u) = 1_t(u)w_t - h(u)$ so that $E v_t(u) = 0$. Since $(1/T)\sum_{t=1}^T X_{Tt} \Rightarrow \int_0^1 X(s) ds$ by Assumption 2.4 and the continuous mapping theorem (CMT), it is sufficient to show that

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

Fix $\varepsilon > 0$ and let $c = E|w_t| < \infty$. Observe that

$$(A.7) \quad \sup_{0 \leq u \leq 1} |v_t(u)| \leq |w_t| + c.$$

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

$$(A.8) \quad P\left(2c \sup_{|s-s'| \leq \delta} |X(s) - X(s')| \leq \varepsilon\right) \geq 1 - \varepsilon.$$

Set $N = \lceil 1/\delta \rceil$, and for $k = 0, \dots, N-1$ set $t_k = \lceil kT\delta \rceil + 1$ and $t_k^* = t_{k+1} - 1$. By (A.7), the Ergodic Theorem, Assumption 2.4, and the CMT,

$$(A.9) \quad \begin{aligned} \frac{1}{T} \sum_{k=0}^{N-1} \sum_{t=t_k}^{t_k^*} |X_{Tt} - X_{Tt_k}| \sup_{0 \leq u \leq 1} |v_t(u)| &\leq \sup_{|t-t'| \leq T\delta} |X_{Tt} - X_{Tt'}| \frac{1}{T} \sum_{t=1}^T \sup_{0 \leq u \leq 1} |v_t(u)| \\ &\leq \left[\sup_{|t-t'| \leq T\delta} |X_{Tt} - X_{Tt'}| \right] \left(\frac{1}{T} \sum_{t=1}^T (|w_t| + c) \right) \\ &\Rightarrow 2c \sup_{|s-s'| \leq \delta} |X(s) - X(s')|. \end{aligned}$$

Lemma 1 of Hansen (1996) states that

$$(A.10) \quad V_N = \sup_{0 \leq u \leq 1} \left| \frac{1}{N} \sum_{t=1}^N v_t(u) \right| \rightarrow_p 0.$$

Since (A.7) implies that $V_N \leq \frac{1}{N} \sum_{t=1}^N |w_t| + c$, and $E|w_t|^{1+\delta/2} < \infty$ by (A.6), it follows that V_N is uniformly integrable. Theorem 25.12 of Billingsley (1986) states that uniform integrability and (A.10) imply $EV_N \rightarrow 0$. This implies

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

as $T \rightarrow \infty$, which implies

$$(A.11) \quad \frac{1}{T} \sum_{k=0}^{N-1} \sup_{0 \leq u \leq 1} \left| \sum_{t=t_k}^{t_k^*} v_t(u) \right| \rightarrow_p 0$$

by Markov's inequality. (A.9), (A.11) and the fact that $X_{Tt} = O_p(1)$ uniformly in t imply

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

where the last inequality holds with probability exceeding $1 - \varepsilon$ by (A.8). This establishes the result. *Q.E.D.*

PROOF OF THEOREM 4: Set $U_t = G(Z_t) \sim U[0, 1]$. Due to the equality $1_{\{Z_{t-1} < \lambda\}} = 1_{\{U_{t-1} < G(\lambda)\}}$, the change-of-variables $u = G(\lambda)$ allows us to re-write the test statistic as

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

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

$$(A.12) \quad \Delta y_t = x'_{t-1} \theta_1 1_{\{U_{t-1} < u\}} + x'_{t-1} \theta_2 1_{\{U_{t-1} \geq u\}} + e_t.$$

Define

$$w_{1t-1}(u) = w_{t-1} - \frac{\sum_{t=1}^T 1_{\{U_{t-1} < u\}} w_{t-1}}{\sum_{t=1}^T 1_{\{U_{t-1} < u\}}}$$

and

$$w_{2t-1}(u) = w_{t-1} - \frac{\sum_{t=1}^T 1_{\{U_{t-1} \geq u\}} w_{t-1}}{\sum_{t=1}^T 1_{\{U_{t-1} \geq u\}}},$$

which is w_{t-1} projected orthogonal to $1_{\{U_{t-1} < u\}}$ and $1_{\{U_{t-1} \geq u\}}$, respectively. Since the intercepts $1_{\{U_{t-1} < u\}}$ and $1_{\{U_{t-1} \geq u\}}$ are included in the regression (A.12), we can replace $w_{t-1} 1_{\{U_{t-1} < u\}}$ and $w_{t-1} 1_{\{U_{t-1} \geq u\}}$ by $w_{1t-1}(u) 1_{\{U_{t-1} < u\}}$ and $w_{2t-1}(u) 1_{\{U_{t-1} \geq u\}}$, respectively, without affecting the test statistic $W_T^*(u)$. We can also rescale y_{t-1} and t without affecting the value of $W_T^*(u)$, so that the latter can be calculated as the Wald statistic from the regression

$$\Delta y_t = (X'_{Tt-1} \quad w_{1t-1}(u)') \theta_1 1_{\{U_{t-1} < u\}} + (X'_{Tt-1} \quad w_{2t-1}(u)') \theta_2 1_{\{U_{t-1} \geq u\}} + e_t,$$

where

$$X_{Tt-1} = \left(\frac{y_{t-1}}{(\sigma/a)T^{1/2}} \quad (\delta_T r_t)' \right)'$$

Now define $1_{t-1}(u) = 1_{\{u_{t-1} < u\}}$. Assumptions 2.1, 2.2, 2.3, and 2.4 are directly implied by Assumption 1, so Theorem 1 yields

$$(A.13) \quad \frac{1}{\sigma\sqrt{T}} \sum_{t=1}^{[Ts]} 1_{t-1}(u)e_t \Rightarrow W(s, u),$$

where $W(s, u)$ is a two-parameter Brownian motion. Furthermore,

$$(A.14) \quad \frac{1}{\sigma\sqrt{T}} \sum_{t=1}^T 1_{t-1}(u)w_{t-1}e_t \Rightarrow W_2(u),$$

where $W_2(u)$ is a mean-zero vector Gaussian process with kernel

$$E(W_2(u_1)W_2(u_2)') = H(u_1 \wedge u_2),$$

by Theorem 1 of Hansen (1996). Furthermore, equations (A.13)–(A.14) hold jointly. Namely, the processes $W(s, u)$ and $W_2(u)$ are jointly Gaussian with covariance kernel

$$E(W_2(u_1)W(s, u_2)) = h(u_1 \wedge u_2).$$

The convergence of the joint finite dimensional distributions may be obtained by the MDS central limit theorem and the Cramer-Wold device. Stochastic equicontinuity follows from (A.13) and (A.14) and the fact that our convergence is in respect to the uniform metric (since the limit processes are continuous).

The inclusion of the intercept and trend variables in the regression means that the test statistic $W_T^*(u)$ is invariant to the actual value of μ . We can thus without loss of generality set $\mu = 0$. Under H_0 we see that y_t is generated by the stochastic process $a(L)\Delta y_t = e_t$ where $a(L)$ is a lag polynomial satisfying $a(1) = a = 1 - \alpha't$. As a consequence of (A.13),

$$\frac{1}{\sigma\sqrt{T}} \sum_{t=1}^{[Ts]} e_t \Rightarrow W(s, 1) \equiv W(s)$$

and

$$(A.15) \quad T^{-1/2}y_{[Ts]} \Rightarrow (\sigma/a)W(s).$$

It follows that $X_{T[Ts]} \Rightarrow X(s) = (W(s) \ r(s)')$, establishing Assumption 2.4.

Standard algebra shows that $W_T^*(u)$ can be written as

$$W_T^*(u) = \frac{S_T^*(u)'(M_T(u) - M_T(u)M_T(1)^{-1}M_T(u))^{-1}S_T^*(u)}{\hat{\sigma}^2(u)/\sigma^2},$$

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

$$S_T(u) = \begin{pmatrix} S_{1T}(u) \\ S_{2T}(u) \end{pmatrix} = \begin{pmatrix} \frac{1}{\sigma\sqrt{T}} \sum_{t=1}^T 1_{t-1}(u)X_{Tt-1}e_t \\ \frac{1}{\sigma\sqrt{T}} \sum_{t=1}^T 1_{t-1}(u)w_{1t-1}(u)e_t \end{pmatrix},$$

and

$$M_T(u) = \begin{pmatrix} M_{11T}(u) & M_{21T}(u)' \\ M_{21T}(u) & M_{22T}(u) \end{pmatrix} = \begin{pmatrix} \frac{1}{T} \sum_{t=1}^T 1_{t-1}(u) X_{Tt-1} X'_{Tt-1} & \frac{1}{T} \sum_{t=1}^T 1_{t-1}(u) X_{Tt-1} w_{1t-1}(u)' \\ \frac{1}{T} \sum_{t=1}^T 1_{t-1}(u) w_{1t-1}(u) X'_{Tt-1} & \frac{1}{T} \sum_{t=1}^T 1_{t-1}(u) w_{1t-1}(u) w_{1t-1}(u)' \end{pmatrix}.$$

Theorem 3 yields

$$M_{11T}(u) \Rightarrow u \int_0^1 X(s) X(s)' ds,$$

$$M_{21T}(u) = \frac{1}{T} \sum_{t=1}^T 1_{t-1}(u) w_{t-1} X'_{Tt-1} - \frac{1}{T} \frac{\sum_{t=1}^T 1_{t-1}(u) w_{t-1}}{\sum_{t=1}^T 1_{t-1}(u)} \sum_{t=1}^T 1_{t-1}(u) X'_{Tt-1}$$

$$\Rightarrow h(u) \int_0^1 X(s)' ds - \frac{h(u)}{u} u \int_0^1 X(s)' ds = 0,$$

and

$$M_{22T}(u) = \frac{1}{T} \sum_{t=1}^T 1_{t-1}(u) w_{t-1} w'_{t-1} - \frac{\sum_{t=1}^T 1_{t-1}(u) w_{t-1} \sum_{t=1}^T 1_{t-1}(u) w'_{t-1}}{\sum_{t=1}^T 1_{t-1}(u)}$$

$$\Rightarrow H(u) - \frac{h(u)h(u)'}{u} \equiv \Omega(u).$$

Hence $M_T(u)$ is asymptotically block-diagonal. It follows that

$$W_T^*(u) = S_{1T}^*(u)' (M_{11T}(u) - M_{11T}(u) M_{11T}(1)^{-1} M_{11T}(u))^{-1} S_{1T}^*(u) + S_{2T}^*(u)' (M_{22T}(u) - M_{22T}(u) M_{22T}(1)^{-1} M_{22T}(u))^{-1} S_{2T}^*(u) + o_p(1),$$

where

$$S_{1T}^*(u) = S_{1T}(u) - M_{11T}(u) M_{11T}(1)^{-1} S_{1T}(1)$$

and

$$S_{2T}^*(u) = S_{2T}(u) - M_{22T}(u) M_{22T}(1)^{-1} S_{2T}(1).$$

Theorem 2 yields

$$S_{1T}(u) \Rightarrow \int_0^1 X(s) dW(s, u)' \equiv J_1(u),$$

and

$$S_{1T}^*(u) \Rightarrow J_1(u) - u \left(\int_0^1 X(s) X(s)' ds \right) \left(\int_0^1 X(s) X(s)' ds \right)^{-1} J_1(1) = J_1^*(u).$$

(A.13) implies

$$S_{2T}(u) = \frac{1}{\sigma\sqrt{T}} \sum_{t=1}^T 1_{t-1}(u)w_{t-1}e_t - \frac{1}{\sigma\sqrt{T}} \frac{\sum_{t=1}^T 1_{t-1}(u)w_{t-1}}{\sum_{t=1}^T 1_{t-1}(u)} \sum_{t=1}^T 1_{t-1}(u)e_t$$

$$\Rightarrow W_2(u) - \frac{h(u)}{u}W(u) \equiv J_2(u),$$

and

$$S_{2T}^*(u) \Rightarrow J_2(u) - \Omega(u)\Omega(1)^{-1}J_1(1).$$

Observe that $J_2(u)$ is a Gaussian process with kernel $\Omega(u)$. To see this, note that if $u_1 \leq u_2$,

$$E(J_2(u_1)J_2(u_2)') = E\left(\left(W_2(u_1) - \frac{h(u_1)}{u_1}W(u_1)\right)\left(W_2(u_2) - \frac{h(u_2)}{u_2}W(u_2)\right)\right)$$

$$= H(u_1) - \frac{h(u_1)}{u_1}h(u_1)' - h(u_1)\frac{h(u_2)'}{u_2} + \frac{h(u_1)}{u_1}\frac{h(u_2)'}{u_2}u_1$$

$$= H(u_1) - \frac{h(u_1)h(u_1)'}{u_1} = \Omega(u_1).$$

Also, $J_2(u)$ is independent of $W(s, u)$, since they are Gaussian and orthogonal:

$$E(J_2(u)W(s, u)) = E(W_2(u)W(s, u)) - \frac{h(u)}{u}E(W_2(u)W(s, u))$$

$$= h(u)s - \frac{h(u)}{u}su = 0.$$

It follows that J_1 and J_2 are independent processes.

Together, we find that

$$W_T^*(u) \Rightarrow \frac{J_1^*(u)' \left(\int_0^1 X(s)X(s)'ds \right)^{-1} J_1^*(u)}{u(1-u)}$$

$$+ J_2^*(u)'(\Omega(u) - \Omega(u)\Omega(1)^{-1}\Omega(u))^{-1} J_2^*(u)$$

$$= Q_1(u) + Q_2(u) = T(u).$$

We conclude that

$$W_T = \sup_{u \in [\pi_1, \pi_2]} W_T^*(u) \Rightarrow \sup_{u \in [\pi_1, \pi_2]} T(u),$$

which is the stated result. Q.E.D.

PROOF OF THEOREM 5: Since the regression includes a trend, the test statistic is invariant to the intercept $\mu = \mu_1 = \mu_2$, so we set $\mu = 0$ to ease exposition. We reparameterize the model as in (A.12), letting $u = G(\lambda)$. Standard algebraic results for linear regression show that $t_1 = t_1(\hat{u})$ and $t_2 = t_2(\hat{u})$ where $t_1(u)$ and $t_2(u)$ are the t ratios for $\hat{\rho}_1(u)$ and $\hat{\rho}_2(u)$ in equation (A.12) with u fixed, and

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

where $W_T^*(u)$ is defined before (A.12). To find the limit distribution of R_T , we find the limit distributions of $t_1(u)$ and $t_2(u)$ considered as a function of u , find the limit distribution of \hat{u} , and combine these results with the continuous mapping theorem.

Let $r_{Tt} = (\delta_T r_t)'$ and set $w_{1t-1}(u)$ as in the proof of Theorem 4. We can write $t_1(u)$ explicitly as

$$(A.16) \quad t_1(u) = \frac{N_T^*(u)}{(D_T^*(u)\hat{\sigma}^2(u))^{1/2}},$$

where

$$N_T^*(u) = N_T(u) - B_T(u)'C_T(u)^{-1}G_T(u),$$

$$D_T^*(u) = D_T(u) - B_T(u)'C_T(u)^{-1}B_T(u),$$

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

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

$$B_T(u) = \begin{bmatrix} \frac{1}{T^{3/2}} \sum_{t=1}^T r_{Tt} y_{t-1} 1_{t-1}(u) \\ \frac{1}{T^{3/2}} \sum_{t=1}^T w_{1t-1}(u) y_{t-1} 1_{t-1}(u) \end{bmatrix},$$

$$C_T(u) = \begin{bmatrix} \frac{1}{T} \sum_{t=1}^T r_{Tt} r_{Tt}' 1_{t-1}(u) & \frac{1}{T} \sum_{t=1}^T r_{Tt} w_{1t-1}(u)' 1_{t-1}(u) \\ \frac{1}{T} \sum_{t=1}^T w_{1t-1}(u) r_{Tt}' 1_{t-1}(u) & \frac{1}{T} \sum_{t=1}^T w_{1t-1}(u) w_{1t-1}(u)' 1_{t-1}(u) \end{bmatrix},$$

and

$$G_T(u) = \begin{bmatrix} \frac{1}{T^{1/2}} \sum_{t=1}^T r_{Tt} e_t 1_{t-1}(u) \\ \frac{1}{T^{1/2}} \sum_{t=1}^T w_{1t-1}(u) e_t 1_{t-1}(u) \end{bmatrix}.$$

From (A.15) we see that $y_{[Tr]} \Rightarrow (\sigma/a)W(s)$. Also, note that $r_{T[Tr]} \Rightarrow r(s)$. Hence

$$N_T(u) \Rightarrow (\sigma^2/a) \int_0^1 W(s) dW(s, u),$$

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

$$B_T(u) \Rightarrow \begin{bmatrix} (\sigma/a)u \int_0^1 r(s)W(s) ds \\ 0 \end{bmatrix},$$

$$C_T(u) \Rightarrow \begin{bmatrix} u \int_0^1 r(s)r(s)' ds & 0 \\ 0 & \Omega(u) \end{bmatrix},$$

and

$$G_T(u) \Rightarrow \begin{bmatrix} \sigma \int_0^1 r(s) dW(s, u) \\ \sigma J_2(u) \end{bmatrix}.$$

Thus $N_T^*(u) \Rightarrow (\sigma^2/a) \int_0^1 W^*(s) dW(s, u)$, where $W^*(s)$ is detrended Brownian motion as defined in the statement of Theorem 5. Similarly, $D_T^*(u) \Rightarrow (\sigma^2/a^2) \int_0^1 W^*(s)^2 ds$. It follows that

$$(A.17) \quad t_1(u) \Rightarrow \frac{\int_0^1 W^*(s) dW(s, u)}{(u \int_0^1 W^*(s)^2 ds)^{1/2}} \equiv t_1(u).$$

Using similar arguments, we can show that

$$(A.18) \quad t_2(u) \Rightarrow \frac{[\int_0^1 W^*(s)(dW(s, 1) - dW(s, u))]}{((1-u) \int_0^1 W^*(s)^2 ds)^{1/2}} \equiv t_2(u),$$

and (A.17) and (A.18) hold jointly. Theorem 2 shows that $t_1(u)$ and $t_2(u)$ are continuous functions of u (almost surely).

Finally, we need the limit distribution of \hat{u} . In Theorem 4 we showed that $W_T^*(u) \Rightarrow T(u)$. This limit process $T(u)$ 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

$$(A.19) \quad \hat{u} = \operatorname{argmax}_{u \in [\pi_1, \pi_2]} W_T^*(u) \Rightarrow \operatorname{argmax}_{u \in [\pi_1, \pi_2]} T(u) = u^*.$$

Equation (A.17), (A.18), and (A.19) combine to yield the first statement of the theorem:

$$(t_1, t_2) \Rightarrow (t_1(u^*), t_2(u^*)).$$

The convergence of R_T follows from the continuity of $R(\cdot, \cdot)$ and the continuous mapping theorem. *Q.E.D.*

PROOF OF THEOREM 6: As in the previous proof we use the reparameterized model (A.12) where $u = G(\lambda)$, $\hat{u} = G(\hat{\lambda})$, and $u_0 = G(\lambda_0)$. Since the test statistic is invariant to σ^2 , we set $\sigma^2 = 1$.

As in the prior proof, $t_1 = t_1(\hat{u})$ and $t_2 = t_2(\hat{u})$. Consider $t_1(u)$. For $u < u_0$, $t_1(u)$ is given by (A.16). For $u > u_0$, note that

$$\begin{aligned} \Delta y_t &= \theta'_1 x_{t-1} 1_{\{U_{t-1} < u_0\}} + \theta'_2 x_{t-1} 1_{\{U_{t-1} \geq u_0\}} + e_t \\ &= \theta'_1 x_{t-1} 1_{\{U_{t-1} < u\}} + \theta'_2 x_{t-1} 1_{\{U_{t-1} \geq u\}} \\ &\quad + ((\mu_2 - \mu_1) + (\alpha_2 - \alpha_1)' w_{t-1}) 1_{\{u_0 \leq U_{t-1} < u\}} + e_t, \end{aligned}$$

so by linear projection,

$$(A.20) \quad t_1(u) = \frac{N_T^*(u) + A_T(u)}{(D_T^*(u) \hat{\sigma}^2(u))^{1/2}},$$

where $N_T^*(u)$ and $D_T^*(u)$ are defined below (A.16),

$$A_T(u) = \frac{1}{T} \sum_{t=1}^T y_{t-1}^{**}(u) [(\mu_2 - \mu_1) + w'_{t-1}(\alpha_2 - \alpha_1)] 1_{\{u_0 \leq U_{t-1} < u\}},$$

and $y_{t-1}^{**}(u)$ is $y_{t-1} 1_{\{U_{t-1} < u\}}$ projected orthogonal to the other regressors.

We now show that $A_T(\hat{u}) \rightarrow_p 0$. Let

$$\begin{aligned} \xi_t &= (\mu_2 - \mu_1) + w'_{t-1}(\alpha_2 - \alpha_1), \\ \xi(u) &= E(\xi_t 1_{\{u_0 \leq U_{t-1} < u\}}) \\ &= (\mu_2 - \mu_1)(u - u_0) + (\alpha_2 - \alpha_1)'(h(u) - h(u_0)), \end{aligned}$$

and

$$G_T(u) = \frac{1}{\sqrt{T}} \sum_{t=1}^T (\xi_t 1_{\{u_0 \leq U_{t-1} < u\}} - \xi(u)).$$

Then

$$\begin{aligned} A_T(u) &\leq \frac{1}{\sqrt{T}} \max_{t \leq T} |y_{t-1}^{**}(u)| \frac{1}{\sqrt{T}} \sum_{t=1}^T \xi_t 1_{\{u_0 \leq U_{t-1} < u\}} \\ &\leq O_p(1) \cdot [G_T(u) + \sqrt{T} \xi(u)], \end{aligned}$$

uniformly in u . We complete the claim by next showing that $G_T(\hat{u}) \rightarrow_p 0$ and $\sqrt{T} \xi(\hat{u}) \rightarrow_p 0$.

In the context of an identified ergodic TAR, Chan (1993) has shown that $T(\hat{u} - u_0) = O_p(1)$, and his proof extends¹⁰ to the present model. The empirical process $G_T(u)$ satisfies the conditions of Theorem 1, application 4, of Doukhan, Massart, and Rio (1995), so $G_T(u) \Rightarrow G(u)$, a Gaussian process with continuous sample paths. As $\hat{u} \rightarrow_p u_0$, it follows that $G_T(\hat{u}) \Rightarrow G(u_0) = 0$. Since e_t has a continuous distribution, $E(w_{t-1} | U_{t-1} = u)$ is continuous in u , so is bounded in the neighborhood of u_0 . Thus

$$\begin{aligned} \frac{d}{du} \xi(u) &= (\mu_2 - \mu_1) + (\alpha_2 - \alpha_1)' \frac{d}{du} h(u) \\ &= (\mu_2 - \mu_1) + (\alpha_2 - \alpha_1)' E(w_{t-1} | U_{t-1} = u) \end{aligned}$$

is bounded in this neighborhood, and hence

$$\sqrt{T} |\xi(\hat{u})| = \sqrt{T} |\xi(\hat{u}) - \xi(u_0)| \leq \left| \frac{d}{du} \xi(\hat{u}) \right| \sqrt{T} |\hat{u} - u_0| = o_p(1).$$

We conclude that $A_T(\hat{u}) \rightarrow_p 0$ as desired.

Combined with (A.20), it follows that

$$t_1(\hat{u}) = \frac{N_T^*(\hat{u})}{(D_T^*(\hat{u}) \hat{\sigma}^2(\hat{u}))^{1/2}} + o_p(1).$$

Let σ_y^2 be defined as (12), and as noted before the statement of Theorem 6, Δy_t is zero-mean, strictly stationary and geometrically ergodic. An extension of Theorem 1 is the joint convergence over $(s, u) \in [0, 1]^2$:

$$\begin{pmatrix} \frac{1}{\sqrt{T}} \sum_{t=1}^{[Ts]} e_t 1_{t-1}(u) \\ \frac{1}{\sigma_y \sqrt{T}} \sum_{t=1}^{[Ts]} \Delta y_t \end{pmatrix} \Rightarrow \begin{pmatrix} W(s, u) \\ W_3(s) \end{pmatrix}$$

¹⁰This extension is available on request from the authors.

(recall that we have set $Ee_t^2 = \sigma^2 = 1$), where $W(s, u)$ is a two-parameter Brownian motion and W_3 is a standard Brownian motion. This implies

$$\begin{pmatrix} \frac{1}{\sqrt{u_0} \sqrt{T}} \sum_{t=1}^{[Ts]} e_t 1_{\{u_{t-1} < u_0\}} \\ \frac{1}{\sqrt{1-u_0} \sqrt{T}} \sum_{t=1}^{[Ts]} e_t 1_{\{u_{t-1} \geq u_0\}} \\ \frac{1}{\sigma_y \sqrt{T}} \sum_{t=1}^{[Ts]} \Delta y_t \end{pmatrix} \Rightarrow \begin{pmatrix} W_1(s) \\ W_2(s) \\ W_3(s) \end{pmatrix} \equiv W(s)$$

where alternative expressions are

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

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

By construction, the vector Brownian motion $W(s)$ has covariance matrix

$$(A.21) \quad EW(1)W(1)' = \begin{pmatrix} 1 & 0 & \delta_1 \\ 0 & 1 & \delta_2 \\ \delta_1 & \delta_2 & 1 \end{pmatrix},$$

where δ_1 and δ_2 are defined in (13) and (14). These results imply that

$$\frac{1}{\sqrt{T}} y_{[Ts]} \Rightarrow \sigma_y W_3(s).$$

Using an arguments identical to that of the previous proof, we see that

$$D_T^*(u) \Rightarrow u \sigma_y^2 \int_0^1 W_3^*(s)^2 ds$$

and

$$N_T^*(u) \Rightarrow \sigma_y \int_0^1 W_3^*(s) dW(s, u)$$

where $W_3^*(s) = W_3(s) - \int_0^1 W_3(a)r(a)' da (\int_0^1 r(a)r(a)' da)^{-1} r(s)$.

Since these limits are continuous in u almost surely and $\hat{u} \rightarrow_p u_0$, it follows that

$$(A.22) \quad t_1 = t_1(\hat{u}) \Rightarrow \frac{\sigma_y \int_0^1 W_3^*(s) dW(s, u_0)}{\left(u_0 \sigma_y^2 \int_0^1 W_3^*(s)^2 ds\right)^{1/2}}$$

$$= \frac{\int_0^1 W_3^*(s) dW_1(s)}{\left(\int_0^1 W_3^*(s)^2 ds\right)^{1/2}}.$$

Similar arguments yield

$$(A.23) \quad t_2 \Rightarrow \frac{\int_0^1 W_3^*(s) dW_2(s)}{\left(\int_0^1 W_3^*(s)^2 ds\right)^{1/2}}.$$

Using (A.21) we can write¹¹

$$(A.24) \quad \begin{pmatrix} W_1(s) \\ W_2(s) \end{pmatrix} = \begin{pmatrix} (1 - \delta_1^2)^{1/2} W_{1,3}(s) \\ (1 - \delta_2^2)^{1/2} W_{2,3}(s) \end{pmatrix} + \begin{pmatrix} \delta_1 \\ \delta_2 \end{pmatrix} W_3(s),$$

where $(W_{1,3}, W_{2,3})'$ is independent of W_3 and has covariance matrix

$$\begin{pmatrix} 1 & \sigma_{21} \\ \sigma_{21} & 1 \end{pmatrix},$$

where σ_{21} is defined in (16). From (A.22)–(A.24) we find that

$$t_1 \Rightarrow (1 - \delta_1^2)^{1/2} \frac{\int_0^1 W_3^*(s) dW_{1,3}(s)}{\left(\int_0^1 W_3^*(s)^2 ds\right)^{1/2}} + \delta_1 \frac{\int_0^1 W_3^*(s) dW_3(s)}{\left(\int_0^1 W_3^*(s)^2 ds\right)^{1/2}}$$

and

$$t_2 \Rightarrow (1 - \delta_2^2)^{1/2} \frac{\int_0^1 W_3^*(s) dW_{2,3}(s)}{\left(\int_0^1 W_3^*(s)^2 ds\right)^{1/2}} + \delta_2 \frac{\int_0^1 W_3^*(s) dW_3(s)}{\left(\int_0^1 W_3^*(s)^2 ds\right)^{1/2}},$$

which can be written as

$$(A.25) \quad \hat{t} = \begin{pmatrix} t_1 \\ t_2 \end{pmatrix} \Rightarrow \begin{pmatrix} (1 - \delta_1^2)^{1/2} Z_1 \\ (1 - \delta_2^2)^{1/2} Z_2 \end{pmatrix} + \begin{pmatrix} \delta_1 \\ \delta_2 \end{pmatrix} DF$$

where (Z_1, Z_2) has distribution (15) and is independent of DF . This establishes the first stated result of the theorem.

Now set

$$H = \begin{pmatrix} \delta_1/a & -\delta_2/a \\ \delta_2/a & \delta_1/a \end{pmatrix}$$

where a is defined in (17), and note that $H'H = I$. Then some algebraic manipulations and (A.25) show that

$$\begin{aligned} H'\hat{t} &\Rightarrow H' \begin{pmatrix} (1 - \delta_1^2)^{1/2} Z_1 \\ (1 - \delta_2^2)^{1/2} Z_2 \end{pmatrix} + H' \begin{pmatrix} \delta_1 \\ \delta_2 \end{pmatrix} DF \\ &= \begin{pmatrix} N_1 \\ N_2 \end{pmatrix} + \begin{pmatrix} a \\ 0 \end{pmatrix} DF \end{aligned}$$

¹¹The derivation is straightforward, yet cumbersome, so is omitted; it can be obtained on request from the authors.

where $N_1 \sim N(0, 1 - a^2)$, $N_2 \sim N(0, 1)$, and DF are mutually independent. Hence since $HH' = I$,

$$\begin{aligned} R_{2T} &= t_1^2 + t_2^2 \\ &= \hat{t}'\hat{t} \\ &= \hat{t}'HH'\hat{t} \\ &\Rightarrow \left[\begin{pmatrix} N_1 \\ N_2 \end{pmatrix} + \begin{pmatrix} a \\ 0 \end{pmatrix} DF \right]' \left[\begin{pmatrix} N_1 \\ N_2 \end{pmatrix} + \begin{pmatrix} a \\ 0 \end{pmatrix} DF \right] \\ &= (N_1 + aDF)^2 + N_2^2, \end{aligned}$$

which is the final statement of the theorem, setting $Z = (1 - a^2)^{-1/2}N_1$ and $\chi_1^2 = N_2^2$. Q.E.D.

REFERENCES

- ALTISSIMO, F., AND G. VIOLANTE (1996): "Persistence and Nonlinearity in U.S. GNP and Unemployment: An Endogenous Delay Threshold VAR," University of Pennsylvania.
- ANDREWS, D. W. K. (1993): "Tests for Parameter Instability and Structural Change with Unknown Change Point," *Econometrica*, 61, 821-856.
- (1998): "Hypothesis Testing with a Restricted Parameter Space," *Journal of Econometrics*, 84, 155-199.
- ANDREWS, D. W. K., AND W. PLOBERGER (1994): "Optimal Tests when a Nuisance Parameter is Present only under the Alternative," *Econometrica*, 62, 1383-1414.
- BAI, J. (1996): "Testing for Parameter Constancy in Linear Regressions: An Empirical Distribution Function Approach," *Econometrica*, 64, 597-622.
- BALKE, N. S., AND T. B. FOMBY (1997): "Threshold Cointegration," *International Economic Review*, 38, 627-645.
- BEAUDRY, P., AND G. KOOP (1993): "Do Recessions Permanently Change Output?" *Journal of Monetary Economics*, 31, 149-164.
- BILLINGSLEY, P. (1986): *Probability and Measure*, 2nd Ed. New York: Wiley.
- CHAN, K. S. (1991): "Percentage Points of Likelihood Ratio Tests for Threshold Autoregression," *Journal of the Royal Statistical Society, Series B*, 53, 691-696.
- (1993): "Consistency and Limiting Distribution of the Least Squares Estimator of a Threshold Autoregressive Model," *The Annals of Statistics*, 21, 520-533.
- CHAN, K. S., AND H. TONG (1985): "On the Use of the Deterministic Lyapunov Function for the Ergodicity of Stochastic Difference Equations," *Advances in Applied Probability*, 17, 667-678.
- CHAN, K. S., AND R. S. TSAY (1998): "Limiting Properties of the Least Squares Estimator of a Continuous Threshold Autoregressive Model," *Biometrika*, 45, 413-426.
- CHEN, C. W. S., AND J. C. LEE (1995): "Bayesian Inference of Threshold Autoregressive Models," *Journal of Time Series Analysis*, 16, 483-492.
- DAVIES, R. B. (1987): "Hypothesis Testing when a Nuisance Parameter is Present only under the Alternative," *Biometrika*, 74, 33-43.
- 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.
- DOUKHAN, P., P. MASSART, AND E. RIO (1995): "Invariance Principles for Absolutely Regular Empirical Processes," *Annales de l'Institut Henri Poincaré*, 31, 393-427.
- GALBRAITH, J. W. (1996): "Credit Rationing and Threshold Effects in the Relation between Money and Output," *Journal of Applied Econometrics*, 11, 419-429.
- GINE, E., AND J. ZINN (1990): "Bootstrapping General Empirical Measures," *The Annals of Probability*, 18, 851-869.
- GONZALEZ, M., AND J. GONZALO (1998): "Threshold Unit Root Models," U. Carlos II de Madrid.
- HALL, P., AND C. C. HEYDE (1980): *Martingale Limit Theory and Its Application*. New York: Academic Press.

- HAMILTON, J. (1989): "A New Approach to the Economic Analysis of Nonstationary Time Series and the Business Cycle," *Econometrica*, 5, 357–384.
- HANSEN, B. E. (1991): "Strong Laws for Dependent Heterogeneous Processes," *Econometric Theory*, 7, 213–221.
- (1992): "Convergence to Stochastic Integrals for Dependent Heterogeneous Processes," *Econometric Theory*, 8, 489–500.
- (1996): "Inference when a Nuisance Parameter is not Identified under the Null Hypothesis," *Econometrica*, 64, 413–430.
- (1997a): "Approximate Asymptotic p -Values for Structural Change Tests," *Journal of Business and Economic Statistics*, 15, 60–67.
- (1997b): "Inference in TAR Models," *Studies in Nonlinear Dynamics and Econometrics*, 1, 119–131.
- (2000): "Sample Splitting and Threshold Estimation," *Econometrica*, 68, 575–603.
- KIM, J., AND D. POLLARD (1990): "Cube Root Asymptotics," *The Annals of Statistics*, 18, 191–219.
- KURTZ, T. G., AND P. PROTTER (1991): "Weak Limit Theorems for Stochastic Integrals and Stochastic Differential Equations," *The Annals of Probability*, 19, 1035–1070.
- MONTGOMERY, A. L., V. ZARNOWITZ, R. S. TSAY, AND G. C. TIAO (1998): "Forecasting the U.S. Unemployment Rate," *Journal of the American Statistical Association*, 93, 478–493.
- NEFTCI, S. (1984): "Are Economic Time Series Asymmetric over the Business Cycle?" *Journal of Political Economy*, 92, 307–328.
- PHAM, T. D., AND L. T. TRAN (1985): "Some Mixing Properties of Time Series Models," *Stochastic Processes and Their Applications*, 19, 297–303.
- PIPPENGER, M. K., AND G. E. GOERING (1993): "A Note on the Empirical Power of Unit Root Tests under Threshold Processes," *Oxford Bulletin of Economics and Statistics*, 55, 473–481.
- POTTER, S. M. (1995): "A Nonlinear Approach to U.S. GNP," *Journal of Applied Econometrics*, 2, 109–125.
- ROTHMAN, P. (1991): "Further Evidence on the Asymmetric Behavior of Unemployment Rates over the Business Cycle," *Journal of Macroeconomics*, 13, 291–298.
- SAID, S. E., AND D. A. DICKEY (1984): "Testing for Unit Roots in Autoregressive-Moving Average Models of Unknown Order," *Biometrika*, 71, 599–608.
- TONG, H. (1978): "On a Threshold Model," in *Pattern Recognition and Signal Processing*, ed. by C. H. Chen. Amsterdam: Sijhoff and Noordhoff.
- (1983): *Threshold Models in Non-linear Time Series Analysis. Lecture Notes in Statistics*, 21. Berlin: Springer.
- (1990): *Non-Linear Time Series: A Dynamical System Approach*. Oxford: Oxford University Press.
- TSAY, R. S. (1997): "Unit-root Tests with Threshold Innovations," University of Chicago.