

TESTING FOR A UNIT ROOT AGAINST TRANSITIONAL AUTOREGRESSIVE MODELS

by

Joon Y. Park and Mototsugu Shintani



Working Paper No. 05-W10

April 2005

DEPARTMENT OF ECONOMICS
VANDERBILT UNIVERSITY
NASHVILLE, TN 37235

www.vanderbilt.edu/econ

Testing for a Unit Root against Transitional Autoregressive Models ¹

Joon Y. Park
Department of Economics
Rice University and SKKU

Mototsugu Shintani
Department of Economics
Vanderbilt University

Abstract

This paper considers the test of a unit root in transitional autoregressive models. In particular, we develop the asymptotic theory of the inf- t test for the null hypothesis of a unit root in a wide class of nonlinear autoregressive models having parameters that are identified only under the alternative of stationarity. Our framework is very general and allows for virtually all potentially interesting models with the threshold, discrete and smooth transition functions. The specifications of shortrun dynamics used in the paper are also fully general, and comparable to those used in the linear unit root models. Most importantly, our asymptotics take it into consideration that the parameter space has a random limit. This is an essential feature of the unit root test in transitional autoregressive models, which has been ignored in the literature. For this very general class of transitional autoregressive models, we show that the inf- t test has well-defined limit distribution depending only upon the transition function and the limit parameter space. The critical values of the test are provided for some of the commonly used models under the conventional specification of the parameter space. Our simulation study shows that the test has good size with the power that is significantly higher than the usual ADF test even for samples of relatively small sizes. We apply the test to various economic time series and find strong evidence for the rejection of random walks in favor of stationary transitional autoregressive models.

First Draft: September 2004

This version: April 2005

JEL Classification: C12, C16, C22.

Key words and phrases: unit root test, threshold autoregressive models (TAR), logistic and exponential smooth transition autoregressive models (LSTAR and ESTAR).

¹We thank Yoosoon Chang for many useful discussions, and to seminar participants at Rice University for their helpful comments. This paper is prepared for the presentation at The 9th World Congress of the Econometric Society on August 18-24, 2005, London, UK.

1. Introduction

In many economic models, the economic agents face some types of costs that prevent an instantaneous adjustment of variables toward their long-run equilibrium levels. As a result of comparing the cost and benefit by the agents, the speed of adjustment naturally depends on the size of deviation from the equilibrium. Empirically, such a partial adjustment dynamics can be described by a nonlinear autoregressive (AR) model that allows transition from one regime, with a mean-reverting behavior, to the other regime, lacking any force toward the long-run level [see, for example, Balke and Fomby (1997)]. Since the behavior of the latter regime is often characterized by a unit root, the partial adjustment model may be viewed as a nonlinear AR model that contains a partial unit root. This nonlinear structure, however, is known to be difficult to be discriminated from the full unit root case, i.e., the model with no long-run equilibrium. In particular, the poor power performance of the standard unit root test has been reported by many studies including Balke and Fomby (1997), Taylor (2001), and Taylor, Peel and Sarno (2001) in various nonlinear transitional AR models.

In this paper, we consider the test for the null hypothesis of a unit root in general transitional AR models having parameters that are identified only under the alternative of stationarity. Distinctive features of our approach are as follows. First, our framework is truly general, permitting virtually all potentially interesting models with the threshold, discrete and smooth transition functions. It not only unifies into a single framework all the transitional models studied previously in the literature, but also provides a new class of more flexible and realistic models. Second, we only require very mild assumptions on the short-run dynamics. In particular, for the unit root test in models with smooth transition functions, we specify the underlying time series to be generated as a linear process driven by the martingale difference innovations with conditional heteroskedasticities. Third, we propose the one-sided $\inf-t$ test, in place of the two-sided tests that have been used in the existing literature introduced below. Our testing problem here is one-directional, i.e., the test of the null of a unit root against the alternative of a stationary transitional model, which can be more effectively dealt with by the one-sided test. Fourth, an arbitrary lag delay is also permitted in the transition function.

Lastly and most importantly, we explicitly specify the parameter space to have a random limit, and fully account its variation in deriving our null asymptotics. To test in transitional AR models including parameters that are only identified under the alternative hypothesis, the parameter space is routinely set as a function of the data, such as an interval with end points given by some quantiles of the data or a fixed interval normalized by the sample standard deviation. Under the null of a unit root, however, such a parameter space has a random limit even under appropriate normalization. Therefore, it is well expected that the null asymptotics are dependent upon the limit of the random parameter space. This dependency on the limit parameter space of the asymptotic critical values has never been properly addressed in the literature. The random parameter space poses an important new technical problem that we have to deal with in obtaining the limit distribution of an extremum test such as our $\inf-t$ test. In particular, the usual stochastic equicontinuity result that is valid only on a fixed compact parameter space is no longer sufficient to derive the required asymptotics.

Under this very general and realistic setup, we derive the full asymptotics for the transitional AR models under the unit root hypothesis. The asymptotic results developed in the paper can be used to analyze the limit behaviors of various statistics under the null hypothesis of a unit root. In particular, we show that the proposed $\text{inf-}t$ test has well-defined limit null distribution, which is free of any nuisance parameters and depends only on the type of transition function and the limit parameter space. The critical values of the test can therefore be computed for each type of transition function and the specification of the parameter space. In the paper, we provide the critical values of the test for the models that are commonly employed in practice, including the threshold autoregression (TAR), logistic smooth transition autoregression (LSTAR), exponential smooth transition autoregression (ESTAR), and their extensions such as double TAR (D-TAR) and double LSTAR (D-LSTAR) that are often referred to as the three-regime TAR and LSTAR, respectively, under the usual specification of the parameter space. The test for a unit root against the transitional AR models can therefore be very easily implemented in practical applications. Neither simulation nor bootstrap is necessary to calculate the critical values.

The test for a unit root in transitional AR models has recently been investigated by many authors. Nevertheless, none of their tests and asymptotic theories is directly comparable to ours. Caner and Hansen (2001) look at the test for a unit root against the TAR model. However, their model has the stationary transition variables, and consequently, their asymptotics differ vastly from ours. Kapetanios, Shin and Snell (2003) consider the unit root test against the ESTAR model, but their methodology is based on the Taylor approximation of the transition function which requires totally different asymptotic analysis. Recent studies on the D-TAR model by Kapetanios and Shin (2003), Seo (2003), Bec, Guay and Guerre (2004) and Bec, Ben Salem and Carrasco (2004a) are more closely related to ours. Their null asymptotics are, however, derived under more stringent conditions and thus have limited practical relevancy. The former two obtain the null asymptotics under the assumption of a fixed compact parameter space, resulting in a null distribution that is degenerate with respect to the threshold parameter. The latter two, though they allow the parameter space to expand as the sample size grows, assume that the limit of the normalized parameter space is nonrandom and compact.² More recent work by Bec, Ben Salem and Carrasco (2004b) on the D-LSTAR model uses the same framework as Bec, Ben Salem and Carrasco (2004a), and has the same limitation.³

The performance of the $\text{inf-}t$ test looks very promising. Through simulations, we find that the test has good finite sample performance. The actual rejection probabilities are indeed quite close to the nominal size for the samples of relatively small sizes in all the transitional AR models considered in the simulations. Moreover, the test yields considerably higher powers over the usual ADF test for a wide range of transitional AR models. The power gain from the $\text{inf-}t$ test is often quite substantial, relative to the usual ADF test. As empirical illustrations, we conduct the test for a unit root in several economic and financial time series using the $\text{inf-}t$ test. The test results are generally in favor of the stationarity

²Their asymptotics rely on the stochastic equicontinuity in Bec, Guay and Guerre (2004, Lemma D.3), which assumes that the parameter space is nonrandom and compact.

³This paper came to our attention after the first version of this paper was written. For the required stochastic equicontinuity, they simply state that it can be established as in Bec, Guay and Guerre (2004).

for all the time series we consider in the paper. This is much more so, if compared with the results from the ADF test. In particular, our empirical results make it clear that, for many economic and financial time series, the stationary transitional autoregressions can be much more plausible alternatives to the random walk models than the usual stationary linear autoregressions.

The rest of the paper is organized as follows. Section 2 provides some preliminaries. It introduces the model and the hypotheses to be tested, which are followed by some important discussions on the normalization and parameter space. The assumptions on the transition functions and the preliminary asymptotic results are also given here. In Section 3, the test statistic is presented with the relevant limit theories for the prototypical transitional AR models. In particular, the asymptotic null distribution is obtained and its nuisance parameter dependency is analyzed. The test consistency is also established. Section 4 extends the results in Section 3 to the models with general transition functions and data generating processes. More specifically, we look at the models with the intercept term, multiple transition functions and errors of unknown form. Section 5 summarizes the findings from our Monte Carlo experiments. In Section 6, we present the empirical analyses of the U.S. unemployment rate, purchasing power parity (PPP) and target zone model. Section 7 concludes the paper. All mathematical proofs are gathered in Appendix.

A word of notation. We use \mathbb{R} (\mathbb{R}_+) to denote the set of (nonnegative) real numbers, and \mathbb{Z} the set of integers. As usual, $\rightarrow_{a.s.}$, \rightarrow_p , \rightarrow_d signify convergences almost surely, in probability and in distribution, respectively, and $=_d$ the equality in distribution. The standard Brownian motion is designated by W throughout the paper. Finally, $\|\cdot\|$ denotes the usual Euclidean norm if applied to a vector, and the standard operator norm if applied to a matrix. Therefore, we have $\|x\| = (x_1^2 + \dots + x_n^2)^{1/2}$ for an n -vector $x = (x_i)$ and $\|A\| = \sup_{x \in \mathbb{R}^n} \|Ax\|/\|x\|$ for an $n \times n$ matrix $A = (a_{ij})$.

2. Preliminaries

2.1 The Model and Hypothesis

We first look at the prototypical version of the transitional AR models that will be considered in the paper. Suppose that we are interested in modelling the transition between two regimes given by the unit root regime

$$y_t = y_{t-1} + u_t \tag{1}$$

and the mean-reverting regime

$$\Delta y_t = \lambda y_{t-1} + u_t \tag{2}$$

with $\lambda < 0$, where (u_t) is the zero mean sequence of errors that will be more precisely specified later. Typically, the unit root regime represents no adjustment movement, while the mean-reverting regime represents linear adjustment towards long-run equilibrium. However, as will be shown below, the model can also be viewed as a nonlinear AR model of nonlinear mean-reversion with an appropriate choice of transition function.

Now we introduce the transition function $\pi(y_{t-d}, \theta)$ as a weight on the latter regime. Here, y_{t-d} is the transition variable with lag delay $d \geq 1$, θ is an m -dimensional parameter, and π is a real-valued function on $\mathbb{R} \times \mathbb{R}^m$ that usually takes values between zero and unity.⁴ The resulting model is then given by

$$y_t = [1 - \pi(y_{t-d}, \theta)]y_{t-1} + \pi(y_{t-d}, \theta)(1 + \lambda)y_{t-1} + u_t,$$

which reduces to

$$\Delta y_t = \lambda y_{t-1} \pi(y_{t-d}, \theta) + u_t \quad (3)$$

with $\lambda < 0$. Throughout the paper, we assume that d is arbitrary, but fixed.

With various choices of the transition function π , the model (3) can represent a very wide class of nonlinear partial adjustment AR models with a state dependent speed of adjustment. However, if $\lambda = 0$, there will be only a single regime with a unit root that represents no adjustment towards long-run equilibrium. For this reason, in the model (3), it is of interest to test the null hypothesis

$$H_0: \lambda = 0 \quad (4)$$

or no long-run equilibrium. Under the alternative hypothesis

$$H_1: \lambda < 0, \quad (5)$$

our model (3) encompasses a variety of nonlinear AR models with a long-run equilibrium.⁵ Some prototypical examples of the transition function π are given below.

Model	Transition Function	Parameter
TAR	$\pi(x, \theta) = 1\{x \leq \mu\}$	$\theta = \mu$
LSTAR	$\pi(x, \theta) = [1 + e^{\kappa(x-\mu)}]^{-1}$	$\theta = (\mu, \kappa)$
ESTAR	$\pi(x, \theta) = 1 - e^{-\kappa^2 x^2}$	$\theta = \kappa$
D-TAR	$\pi(x, \theta) = 1\{x \leq \mu_1\} + 1\{x \geq \mu_2\}$	$\theta = (\mu_1, \mu_2)$
D-LSTAR	$\pi(x, \theta) = [1 + e^{\kappa_1(x-\mu_1)}]^{-1} + [1 + e^{-\kappa_2(x-\mu_2)}]^{-1}$	$\theta = (\mu_1, \mu_2, \kappa_1, \kappa_2)$

The TAR model can be used to consider an abrupt transition at a certain level, say μ . Several variations of this class of transitional autoregressive model were considered by Caner and Hansen (2001), Kapetanios and Shin (2003) and Seo (2003). To introduce a smooth transition, it can be replaced by the LSTAR model with a logistic transition function. The TAR model is often considered as a limit case of the LSTAR model with $\kappa \rightarrow \infty$, where $\kappa (> 0)$ is the scale parameter. The ESTAR model is often used when the economic agent can

⁴The range of this function, however, is not restricted to $[0,1]$ in our subsequent theoretical development.

⁵The alternative hypothesis $H_1: \lambda < 0$ is neither sufficient nor necessary for the stationarity of (y_t) . See, e.g., Chan, Petrucci, Tong and Woolford (1985). In general, we need extra conditions to ensure that (y_t) is stationary. On the other hand, (y_t) can be stationary even when $\lambda = 0$. However, we will mainly consider in the paper stationary (y_t) under the alternative hypothesis.

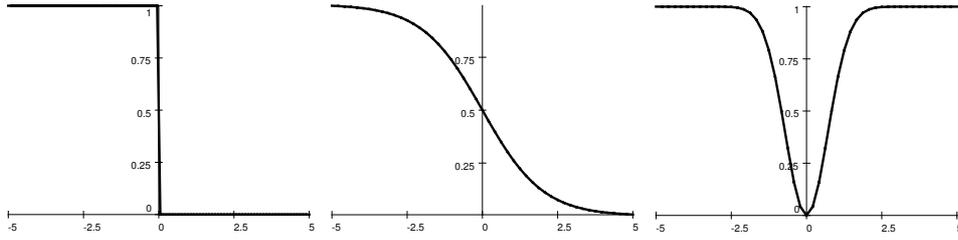


Figure 1: Transition Functions for TAR, LSTAR and ESTAR Models

have arbitrage opportunities by facing deviation (x) from the long-run equilibrium in either directions. In such a case, unit root regime becomes an inner regime, and mean-reverting regime becomes two outer regimes. The leading example is the aggregate real exchange rate dynamics under the purchasing power parity in the long-run [see, e.g., Michael, Nobay, and Peel (1997), and Taylor, Peel, and Sarno (2001)]. The transition functions for the TAR, LSTAR and ESTAR models are shown in Figure 1.⁶

If we are interested in the international difference in the price of a particular individual good rather than the aggregate price index, it may be more appropriate to employ a trade cost model with fixed trade cost, μ . In such a case, adjustment process of deviation (x) from the long-run law of one price level is often estimated by a model with a transition function given by $\pi(x, \theta) = 1\{|x| \geq \mu\}$, $\theta = \mu > 0$ [see, for example, Obstfeld and Taylor (1997)]. Similar discrete adjustment process can be also applied to the model of monetary policy intervention, including exchange rate target zones, and spread between Fed Funds rate and Discount rate (Balke and Fomby, 1997, p.628). Multiple threshold values can be also incorporated by the transition function $\pi(x, \theta) = 1\{x \leq \mu_1\} + 1\{x \geq \mu_2\}$, $\theta = (\mu_1, \mu_2)$, where $\mu_1 < \mu_2$. The D-TAR model with this type of transition function is sometimes referred to as a three-regime TAR model.

Using the analogy to the relationship between the TAR and LSTAR models, a smoothed version of the D-TAR model, the D-LSTAR model, can be also introduced. Such a model is recently considered by van Dijk, Teräsvirta and Franses (2002) and Bec, Ben Salem and Carrasco (2004b).⁷ Both transition functions for the D-TAR and D-LSTAR models are shown in Figure 2.⁸ Finally, we note that the D-LSTAR model can be simplified by introducing restrictions such as $\mu_1 = -\mu_2$ and $\kappa_1 = \kappa_2$, or by considering the limit case in one of the two scale parameters, say, $\kappa_2 \rightarrow \infty$. In the latter case, we have a combination of discrete and smooth transition with $\pi(x, \theta) = [1 + \exp\{\kappa_1(x - \mu_1)\}]^{-1} + 1\{x \geq \mu_2\}$. It should be noted that our theory given in the next section will also allow such a hybrid-type transition function.

⁶For the models in Figure 1, we set the location parameter μ to zero and the scale parameter κ to unity.

⁷An alternative approach to allow smooth transition in D-TAR model is to consider the second-order LSTAR model with transition function $\pi(x, \theta) = [1 + \exp\{-\kappa^2(x - \mu_1)(x - \mu_2)\}]^{-1}$, $\theta = (\mu_1, \mu_2, \kappa)$, where $\mu_1 < \mu_2$ and $\kappa > 0$.

⁸For the models in Figure 2, the location and scale parameters are set to $\mu_1 = -3$, $\mu_2 = 3$ and $\kappa_1 = \kappa_2 = 5$.

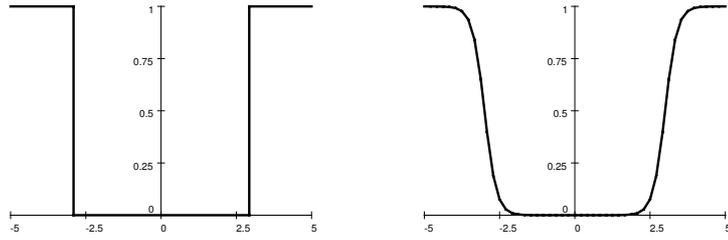


Figure 2: Transition Functions for D-TAR and D-LSTAR Models

2.2 Normalization and Parameter Space

Of course, the parameter θ in the transition function is not identified under the null hypothesis (4). This would naturally lead us to use as the test statistic the infimum of the t -ratios given for each of the possible values of the parameter θ , which we call the *inf-t test* in the paper. In formulating the test, it is extremely important to properly select the parameter space for θ . As we show below, it is not simply an empirical matter. An inappropriately specified parameter space yields improper limit null distributions with incorrect asymptotic critical values, and hence leads us to an invalid inference. This is mainly because the transition variable (y_t) behaves rather distinctively under the null and alternative hypotheses. Under the null of a unit root, (y_t) has a stochastic trend and explodes as the sample size increases. Under the alternative of stationarity, on the other hand, (y_t) does not show such a trending and explosive behavior. Needless to say, this differing behavior of the transition variable should be well taken into consideration, if the parameter space is set to be data-dependent and given as a function of the data (y_1, \dots, y_n) for each n .

It has indeed been the usual convention in the related literature dealing with nonlinear transition functions such as ours to set the parameter space to be randomly given by the functions of the data. For the location parameter μ (or μ_1 and μ_2) in the transitional AR models introduced earlier, it is customary to set the parameter space given by some fixed percentiles in the range of (y_1, \dots, y_n) [see, e.g., Caner and Hansen (2001)]. Moreover, the scale parameter κ (or κ_1 and κ_2) in those models is usually searched over some fixed interval normalized by the sample standard deviation $(\sum_{t=1}^n y_t^2/n)^{1/2}$ of the data (y_1, \dots, y_n) [see, e.g., van Dijk, Teräsvirta and Franses (2002)]. When (y_t) is stationary, they yield parameter spaces that are well defined in the limit. This is not so, however, if (y_t) has a unit root. The space for the location parameter expands at the rate of \sqrt{n} , while for the scale parameter it shrinks towards the origin at the rate of \sqrt{n} . Therefore, the usual assumption on the compactness of the parameter space is not appropriate in this case. The assumption simply cannot be met for the location parameter, and yields degenerate asymptotics with respect to the scale parameter. It is obvious that under the null hypothesis of a unit root we need to redefine the parameter, or equivalently, normalize the transition variable before we require the parameter space to be compact.

In the subsequent development of our asymptotic theories, we specify the transition term in our model as

$$\pi\left(\frac{y_{t-d}}{\sqrt{n}}, \theta\right) \quad \text{with } \theta \in \Theta_n \quad (6)$$

and

$$\pi(y_{t-d}, \theta) \quad \text{with} \quad \theta \in \Theta_n \quad (7)$$

respectively under the null and alternative hypotheses, and assume

Assumption 2.1 $\Theta_n \rightarrow_d \Theta$, where Θ is a compact subset of \mathbb{R}^m a.s.

In general, (Θ_n) is a random sequence of parameter spaces given for each n as functions of the sample (y_1, \dots, y_n) . Naturally, $\Theta_n \rightarrow_d \Theta$ implies that the limit parameter space Θ is a random subset of \mathbb{R}^m whose distribution is given by the distributional limit of such functions. Our specification of the transition term in (6) and (7), of course, does not imply that we have different transition functions under the null and alternative hypotheses. They are meant to be the same function. We simply require that the condition in Assumption 2.1 apply to two different sets of parameters, one normalized and the other unnormalized, depending upon whether we analyze our model under the null or alternative hypothesis.

Throughout the paper, we will denote more explicitly by θ_n the parameter in the normalized transition function (6), whenever it is necessary to directly relate it to the original unnormalized parameter θ in (7). For the location parameter μ (μ_1 and μ_2) in the TAR (D-TAR) model, we have the relationship $\mu_n = \mu/\sqrt{n}$ ($\mu_{1n} = \mu_1/\sqrt{n}$ and $\mu_{2n} = \mu_2/\sqrt{n}$). Note that μ (μ_1 and μ_2) lying in an interval given by certain percentiles of (y_1, \dots, y_n) is equivalent to μ_n (μ_{1n} and μ_{2n}) belonging to an interval given by the corresponding percentiles of $(y_1/\sqrt{n}, \dots, y_n/\sqrt{n})$. For the scale parameter κ (κ_1 and κ_2) in the LSTAR and ESTAR (D-LSTAR), on the other hand, we have $\kappa_n = \sqrt{n}\kappa$ ($\kappa_{1n} = \sqrt{n}\kappa_1$ and $\kappa_{2n} = \sqrt{n}\kappa_2$). Therefore, requiring κ (κ_1 and κ_2) to be in a fixed interval normalized by $(\sum_{t=1}^n y_t^2/n)^{1/2}$ amounts to assuming κ_n (κ_{1n} and κ_{2n}) to be in the same interval normalized by $(\sum_{t=1}^n y_t^2/n^2)^{1/2}$. As we may easily see, the parameter space given by the percentiles of $(y_1/\sqrt{n}, \dots, y_n/\sqrt{n})$, and any fixed interval normalized by $(\sum_{t=1}^n y_t^2/n^2)^{1/2}$ satisfy our condition in Assumption 2.1 for a general unit root process (y_t) . With our convention in (6) and (7), Assumption 2.1 is therefore met for a wide class of transitional AR models, including all the models introduced in the previous section, under the null hypothesis of a unit root as well as under the alternative hypothesis of stationarity.

Our convention in (6) and (7) is extremely important in deriving proper and nondegenerate asymptotics under the null hypothesis of a unit root. In the previous literature, none of the limit null distributions is obtained under appropriate assumptions on the parameter space. Kapetanios and Shin (2003) and Seo (2003) derive the limit null distribution of the sup-Wald test in a D-TAR model under the assumption of the compact parameter space, i.e., the condition comparable to that in Assumption 2.1 imposed on the unnormalized transition function. Therefore, it is not applicable if the usual convention is followed to set the parameter space. On the other hand, Bec, Ben Salem and Carrasco (2004a) allow the location parameter space to expand in investigating the sup-Wald, LM and LR tests for their D-TAR model. However, their derivations of the limit null distributions rely critically on the stochastic equicontinuity result established in Bec, Guay and Guerre (2004, Lemma D.3), which is valid only for a nonrandom and compact parameter space. As we have discussed earlier, the usual parameter space employed in the literature remains to be random

in the limit and has the support on the entire real line under the null hypothesis, even if appropriately normalized. The limit theory of the D-LSTAR model in Bec, Ben Salem and Carrasco (2004b) is also based on Bec, Guay and Guerre (2004) for the stochastic equicontinuity, and has a similar problem.⁹

2.3 Assumptions on Transition Functions

We now introduce the precise conditions for the transition function. To present the required conditions for π , it will be convenient to introduce some preliminary regularity conditions. These we will provide in Definitions 2.1 and 2.2 below.

Definition 2.1 A transformation v on \mathbb{R} is said to be *regular* if and only if, for any compact subset K of \mathbb{R} given, there exists a δ -sequence of continuous functions (\underline{v}_δ) and (\overline{v}_δ) such that $\underline{v}_\delta(x) \leq v(x) \leq \overline{v}_\delta(x)$ for all $|x-y| < \delta$ on K , and such that $\int_K (\overline{v}_\delta - \underline{v}_\delta)(x) dx \rightarrow 0$ as $\delta \rightarrow 0$.

The regularity condition in Definition 2.1 is satisfied by a large class of functions including all continuous and piecewise continuous functions as noted by Park and Phillips (2001), who introduced it earlier in their study of nonlinear regressions with integrated time series.¹⁰

Definition 2.2 A functional ϖ on $\mathbb{R} \times \mathbb{R}^m$ is said to be *regular* if and only if

- (a) for all $\theta_0 \in \mathbb{R}^m$, $\varpi(\cdot, \theta_0)$ is regular, and
- (b) for each $\theta_0 \in \mathbb{R}^m$ given, there exists δ -sequence of regular functions $(\underline{\varpi}_\delta)$ and $(\overline{\varpi}_\delta)$ such that $\underline{\varpi}_\delta(x) \leq \varpi(x, \theta) \leq \overline{\varpi}_\delta(x)$ for all $\|\theta - \theta_0\| < \delta$, and such that $\overline{\varpi}_\delta(x) - \underline{\varpi}_\delta(x) \rightarrow 0$ as $\delta \rightarrow 0$ a.e. $x \in \mathbb{R}$.

The regularity conditions in Definition 2.2 are comparable to those in Park and Phillips (2001). Our conditions are, however, weaker than theirs, especially in that ours allow for the functionals that are discontinuous in $\theta \in \mathbb{R}^m$. This is necessary to accommodate the indicator type transition functions in our model.

Now we are ready to introduce the conditions required for π .

Assumption 2.2 We assume that

- (a) π is bounded on $\mathbb{R} \times \mathbb{R}^m$,
- (b) π is regular on $\mathbb{R} \times \mathbb{R}^m$,
- (c) for any $\theta_1, \theta_2 \in \mathbb{R}^m$, we have

$$\int_{\mathbb{R}} [\pi(x, \theta_1) - \pi(x, \theta_2)]^2 e^{-x^2/2\omega^2} dx < c \|\theta_1 - \theta_2\|$$

⁹Bec, Ben Salem and Carrasco (2004b) simply refer for the proof of their stochastic equicontinuity to Bec, Guay and Guerre (2004), which does not explicitly deal with the D-LSTAR model. However, we believe that the required extension is possible as long as the normalized parameter space has a nonrandom limit.

¹⁰They also require that the regular function be continuous at infinity. This condition, however, is unnecessary and not invoked here.

for some constant $c > 0$ and the constant ω that will be introduced later in (12).

The conditions in Assumption 2.2 are not stringent and satisfied for virtually all transition functions that can possibly be used in practical applications.

They hold widely for π that is differentiable with respect to θ . Indeed, the conditions (b) and (c) are satisfied for a large class of π 's that are differentiable with respect to θ . To show this, we let $\dot{\pi}(x, \theta) = (\partial/\partial\theta)\pi(x, \theta)$, and we assume there exists a regular function $\bar{\pi}$ on \mathbb{R} such that

$$\sup_{\theta \in \mathbb{R}^m} |\dot{\pi}(x, \theta)| \leq \bar{\pi}(x) \quad (8)$$

for all $x \in \mathbb{R}$, and such that

$$\int_{\mathbb{R}} \bar{\pi}^2(x) e^{-x^2/2\omega^2} dx < \infty. \quad (9)$$

Now it follows immediately that the existence of a regular function $\bar{\pi}$ satisfying (8) and (9) imply the regularity condition (b) of Definition 2.2. To see this, note that we have

$$|\pi(x, \theta_1) - \pi(x, \theta_2)| \leq \bar{\pi}(x) \|\theta_1 - \theta_2\|$$

for all $\theta_1, \theta_2 \in \mathbb{R}^m$. Therefore, if we choose $\theta_0 \in \mathbb{R}^m$ arbitrarily and let

$$\underline{\pi}_\delta(x) = \pi(x, \theta_0) - \delta \bar{\pi}(x) \quad \text{and} \quad \bar{\pi}_\delta(x) = \pi(x, \theta_0) + \delta \bar{\pi}(x),$$

then we have $\underline{\pi}_\delta(x) \leq \pi(x, \theta) \leq \bar{\pi}_\delta(x)$ for all $\|\theta - \theta_0\| < \delta$ and $\bar{\pi}_\delta(x) - \underline{\pi}_\delta(x) \rightarrow 0$ for all $x \in \mathbb{R}$, as required to satisfy condition (b) of Definition 2.2. Clearly, (8) and (9) together imply condition (c) of Assumption 2.2.

The conditions in Assumption 2.2 also hold for a variety of π that is discontinuous with respect to θ . In particular, they allow for indicator type functions, including the ones used in the TAR and D-TAR models. For example, it is easy to see that they are satisfied for $\pi(x, \theta) = 1\{x \leq \theta\}$. Clearly, such function π is bounded (by unity) for all $x, \theta \in \mathbb{R}$. Also, the function $1\{\cdot \leq \theta_0\}$ is piecewise continuous for all $\theta_0 \in \mathbb{R}$ and therefore regular. Moreover, if we fix $\theta_0 \in \mathbb{R}$ and define

$$\begin{aligned} \underline{\pi}_\delta(x) &= \inf_{|\theta - \theta_0| < \delta} 1\{x \leq \theta\} = 1\{x \leq \theta_0 - \delta\} \\ \bar{\pi}_\delta(x) &= \sup_{|\theta - \theta_0| < \delta} 1\{x \leq \theta\} = 1\{x < \theta_0 + \delta\} \end{aligned}$$

then we have $\underline{\pi}_\delta(x) \leq \pi(x, \theta) \leq \bar{\pi}_\delta(x)$ for all $\|\theta - \theta_0\| < \delta$ and $\bar{\pi}_\delta(x) - \underline{\pi}_\delta(x) \rightarrow 0$ for all $x \in \mathbb{R}$ except $x = \theta_0$. Therefore, π is regular. Finally, we have for any $\theta_1, \theta_2 \in \mathbb{R}$

$$\begin{aligned} &\int_{\mathbb{R}} [1\{x \leq \theta_1\} - 1\{x \leq \theta_2\}]^2 e^{-x^2/2\omega^2} dx \\ &\leq \int_{\mathbb{R}} [1\{x \leq \theta_1\} - 1\{x \leq \theta_2\}]^2 dx \leq |\theta_1 - \theta_2|. \end{aligned}$$

Our model can therefore be seen as generalizations of the models with indicator transition functions considered earlier by many authors.

2.4 Preliminary Asymptotic Results

For the subsequent development of our asymptotic theories, we make the following assumptions for the innovations (ε_t) generating the errors (u_t) in (3).

Assumption 2.3 $(\varepsilon_t, \mathcal{F}_t)$ is a martingale difference sequence, with some filtration (\mathcal{F}_t) , such that

- (a) $\mathbb{E} \varepsilon_t^2 = \sigma^2$ for all $t \in \mathbb{Z}$,
- (b) $\sup_{t \in \mathbb{Z}} \mathbb{E} |\varepsilon_t|^r < \infty$ for some $r \geq 4$, and
- (c) $\sup_{1 \leq i \leq n} \frac{1}{n} \sum_{t=1}^i [\mathbb{E}(\varepsilon_t^2 | \mathcal{F}_{t-1}) - \sigma^2] \rightarrow_p 0$ as $n \rightarrow \infty$.

Assumption 2.3 is quite general, and allows for the conditional heteroskedasticity in the innovation sequence (ε_t) . In the unit root literature, it is routinely assumed that the fourth conditional moment of the innovation sequence is bounded. See, e.g., Stock (1994) and Park and Phillips (2001). We do not impose this uniform boundedness of the conditional fourth moment, which is not satisfied for instance by the usual ARCH processes. Our assumption here holds for a wide class of ARCH-type processes.

As is well expected, the usual (infeasible) variance estimator $\sigma_n^2 = (1/n) \sum_{t=1}^n \varepsilon_t^2$ is consistent under Assumption 2.3, as we show in the following lemma.

Lemma 2.1 Under Assumption 2.3, we have $\sigma_n^2 \rightarrow_p \sigma^2$ as $n \rightarrow \infty$.

The result in Lemma 2.1 is essential for the consistent estimation of the error variance.

Moreover, conditions (a) and (b) of Assumption 2.3 are sufficient to ensure that (ε_t) satisfies an invariance principle, as shown in, e.g., Hall and Heyde (1980, Theorem 4.1, p99). A properly constructed partial sum process of (ε_t) would therefore weakly converge to a Brownian motion. For our subsequent theory, however, it would be more convenient to use a more direct method of embedding a distributionally equivalent copy of the partial sum process of (ε_t) into a Brownian motion. We achieve this using the so-called Skorokhod embedding, which we state below as a lemma.

Lemma 2.2 Under Assumption 2.3, there exists a probability space supporting a Brownian motion U with variance σ^2 and a time change (i.e., a nondecreasing sequence of stopping times) τ_i such that

$$\frac{1}{\sqrt{n}} \sum_{t=1}^i \varepsilon_t =_d U\left(\frac{\tau_i}{n}\right)$$

for $i = 1, \dots, n$, and that

$$\sup_{1 \leq i \leq n} \left| \frac{\tau_i}{n} - \frac{i}{n} \right| \rightarrow_p 0 \tag{10}$$

as $n \rightarrow \infty$.

The reader is referred to Hall and Heyde (1980, Appendix) for the Skorokhod embedding introduced in Lemma 2.1, and to Park and Phillips (1999, 2001) for its application to the analysis of nonlinear models with integrated processes. In what follows, we set $\tau_0 \equiv 0$.

We set $u_t = \Delta y_t$. Under the null hypothesis (4), we have

$$\alpha(L)u_t = \varepsilon_t, \quad (11)$$

where $\alpha(z) = 1 - \sum_{k=1}^p \alpha_k z^k$. Assume

Assumption 2.4 $\alpha(z)$ has roots outside the unit circle.

Then it follows that

$$\frac{1}{\sqrt{n}} \sum_{t=1}^i u_t = \frac{1}{\sqrt{n}\alpha(1)} \sum_{t=1}^i \varepsilon_t + o_p(1)$$

uniformly in $1 \leq i \leq n$, as shown in, e.g., Phillips and Solo (1992). Consequently, if we define a Brownian motion V by

$$V(r) = \frac{1}{\alpha(1)} U(r),$$

i.e., V is a Brownian motion with variance ω^2 , where

$$\omega^2 = \frac{\sigma^2}{\alpha(1)^2}, \quad (12)$$

then we have

$$\frac{1}{\sqrt{n}} \sum_{t=1}^i u_t =_d V\left(\frac{\tau_i}{n}\right) + o_p(1)$$

uniformly in $1 \leq i \leq n$. In what follows, we denote by V_n the process defined as

$$V_n(r) = \sum_{i=1}^n V\left(\frac{\tau_{i-1}}{n}\right) 1\left\{\frac{\tau_{i-1}}{n} \leq r < \frac{\tau_i}{n}\right\} + V\left(\frac{\tau_n}{n}\right) 1\left\{r \geq \frac{\tau_n}{n}\right\} \quad (13)$$

with the convention $\tau_0 = 0$ a.s. Clearly, $V_n \rightarrow_d V$ on $[0, 1]$ as $n \rightarrow \infty$, due in particular to our result in (10).

We now establish a functional central limit theory for the nonlinear transformation of integrated time series that can be applied to a class of transition functions introduced in this section. The main results of this section are given below. In what follows, we let

$$M_n(r, \theta) = \int_0^r \pi(V_n(s), \theta) dU(s)$$

$$M(r, \theta) = \int_0^r \pi(V(s), \theta) dU(s)$$

be the stochastic processes defined on $(r, \theta) \in [0, 1] \times \mathbb{R}^m$.

Theorem 2.3 Under Assumptions 2.1, 2.2, 2.3 and 2.4, we have

$$M_n \rightarrow_d M$$

jointly with $V_n \rightarrow_d V$, as $n \rightarrow \infty$.

Proposition 2.4 M has a modification which is almost surely continuous.

The functional central limit theory established in Theorem 2.3 is crucial in developing our subsequent asymptotic theories. It implies, in particular, that

$$f(M_n, V_n) \rightarrow_d f(M, V) \quad (14)$$

for any continuous functional f , as follows from the continuous mapping theorem (CMT). Due to Proposition 2.4, we may assume without loss of generality that the limit process M has a.s. continuous sample paths on $[0, 1] \times \mathbb{R}^m$, by taking such a modification if necessary. This convention will be made throughout the paper. As a consequence, the weak convergence in (14) holds for any functional f that is continuous on $\mathcal{C}([0, 1]) \times \mathcal{C}([0, 1] \times \mathbb{R}^m)$, where $\mathcal{C}(D)$ denotes the set of continuous functions defined on D endowed with the uniform norm. It is very important to note that the space for the parameter θ in the processes M_n and M is not restricted to be a compact space. This is crucial for our asymptotic theories, which allow for the parameter space that remains to be random in the limit.

3. Test Statistics and Asymptotic Theory

The hypothesis (4) can be tested in the regression

$$\Delta y_t = \lambda w_t(\theta) + \sum_{i=1}^p \alpha_i \Delta y_{t-i} + \varepsilon_t, \quad (15)$$

where $w_t(\theta)$ is defined as

$$w_t(\theta) = y_{t-1} \pi(y_{t-d}, \theta), \quad (16)$$

due in particular to our assumption on the error sequence (u_t) in (11).¹¹ To define the test more explicitly, we first consider the least squares regression

$$\Delta y_t = \hat{\lambda}_n(\theta) w_t(\theta) + \sum_{i=1}^p \hat{\alpha}_{ni}(\theta) \Delta y_{t-i} + \hat{\varepsilon}_t(\theta) \quad (17)$$

fitted for each value of $\theta \in \Theta_n$. The usual t -ratio for λ in regression (17) is then defined as

$$T_n(\theta) = \frac{\hat{\lambda}_n(\theta)}{s(\hat{\lambda}_n(\theta))}, \quad (18)$$

¹¹We may consider the model $\Delta y_t = \lambda [y_{t-1} + \sum_{i=1}^p \alpha_i \Delta y_{t-i}] \pi(y_{t-d}, \theta) + \varepsilon_t$, similarly as in Bec, Ben Salem and Carrasco (2004a). Indeed, our subsequent theory can be easily modified to allow for such a model. This, however, will not be pursued in the paper.

where $s(\hat{\lambda}_n(\theta))$ is the standard error of the estimate $\hat{\lambda}_n(\theta)$. If we define

$$A_n(\theta) = \sum_{t=1}^n w_t(\theta) \Delta y_t - \sum_{t=1}^n w_t(\theta) x_t' \left(\sum_{t=1}^n x_t x_t' \right)^{-1} \sum_{t=1}^n x_t \Delta y_t \quad (19)$$

$$B_n(\theta) = \sum_{t=1}^n w_t^2(\theta) - \sum_{t=1}^n w_t(\theta) x_t' \left(\sum_{t=1}^n x_t x_t' \right)^{-1} \sum_{t=1}^n x_t w_t(\theta), \quad (20)$$

where

$$x_t = (\Delta y_{t-1}, \dots, \Delta y_{t-p})',$$

then we have $\hat{\lambda}_n(\theta) = A_n(\theta)/B_n(\theta)$ and $s(\hat{\lambda}_n(\theta)) = \hat{\sigma}_n(\theta)/\sqrt{\overline{B_n(\theta)}}$. Here and elsewhere in the paper, we denote by $\hat{\sigma}_n^2(\theta)$ the error variance estimate given by $\hat{\sigma}_n^2(\theta) = (1/n) \sum_{t=1}^n \hat{\varepsilon}_t^2(\theta)$.

The inf- t test may now simply be defined as

$$T_n = \inf_{\theta \in \Theta_n} T_n(\theta), \quad (21)$$

i.e., the infimum of $T_n(\theta)$ in (18) taken over all values of $\theta \in \Theta_n$. Naturally, the parameter θ may be estimated by

$$\hat{\theta}_n = \operatorname{argmin}_{\theta \in \Theta_n} T_n(\theta),$$

and we may also define the inf- t test as $T_n = T_n(\hat{\theta}_n)$. The estimator $\hat{\theta}_n$ is in general not identical to the nonlinear least squares estimator $\tilde{\theta}_n$, say, of θ . As is well known, $\tilde{\theta}_n$ is given by $\tilde{\theta}_n = \operatorname{argmax} \{ T_n^2(\theta) \mid \theta \in \Theta_n \}$. In contrast, we may define $\hat{\theta}_n$ by

$$\hat{\theta}_n = \operatorname{argmax} \left\{ T_n^2(\theta) \mid \hat{\lambda}_n(\theta) < 0, \theta \in \Theta_n \right\},$$

as long as the set of $\theta \in \Theta_n$ for which $\hat{\lambda}_n(\theta) < 0$ is nonempty. Under the alternative of stationarity and with suitable regularity conditions, it is well expected that both $\hat{\theta}_n$ and $\tilde{\theta}_n$ are consistent. The former, however, will be more efficient than the latter, if the true value of λ is negative.¹²

The test defined in (21) can be directly applied to the TAR, LSTAR and ESTAR models introduced in Section 2.1. For the location parameter μ in these models, we follow, e.g., Caner and Hansen (2001), and suggest to set the parameter space as

$$[Q_n(15), Q_n(85)], \quad (22)$$

where $Q_n(15)$ and $Q_n(85)$ are, respectively the 15th and 85th percentiles of (y_1, \dots, y_n) . On the other hand, we recommend the scale parameter κ to be searched over the interval given by

$$[10^{-1}P_n, 10^3P_n], \quad (23)$$

where $P_n = (\sum_{t=1}^n y_t^2/n)^{-1/2}$, as in, e.g., van Dijk, Teräsvirta and Franses (2002) for their study of stationary smooth transition AR models. In the subsequent development of our

¹²We will not further study the properties of these estimators in the paper, since our main purpose is to test the unit root hypothesis.

theory for the TAR, LSTAR and ESTAR models, we will assume that the specification of the parameter space is given by (22) and (23). The other specifications are, of course, also possible, and we may easily modify our results to accommodate them.

To develop the null asymptotics, we now let $w_t(\theta)$ in (16) be defined with the normalized transition function (6) in place of (7). This is according to our specification of the transition term under the null hypothesis of a unit root. To distinguish them from the unnormalized original parameters, we let μ_n and κ_n be respectively the location and scale parameters in the normalized transition function for the TAR, LSTAR and ESTAR models. As noted earlier in Section 2.2, we may easily deduce that $\mu_n = \mu/\sqrt{n}$ and $\kappa_n = \sqrt{n}\kappa$. Consequently, the parameter spaces for μ_n and κ_n corresponding to (22) and (23) become

$$n^{-1/2}[Q_n(15), Q_n(85)] \quad \text{and} \quad n^{1/2}[10^{-1}P_n, 10^3P_n].$$

Moreover, under the null hypothesis of a unit root, we have

$$n^{-1/2}[Q_n(15), Q_n(85)] \rightarrow_d [Q(15), Q(85)], \quad (24)$$

where Q is the quantile function for V over $r \in [0, 1]$, i.e., $\int_0^1 1\{V(r) \leq Q(100s)\} dr = s$ for $s \in [0, 1]$, and

$$n^{1/2}[10^{-1}P_n, 10^3P_n] \rightarrow_d [10^{-1}P, 10^3P], \quad (25)$$

where $P = (\int_0^1 V(r)^2 dr)^{-1/2}$. The limit parameter spaces for μ_n and κ_n are therefore given respectively by $[Q(15), Q(85)]$ and $[10^{-1}P, 10^3P]$. These are random intervals that are compact a.s., as we assume in Assumption 2.1.

To obtain the limit null distribution of our test statistic T_n , we first show

Lemma 3.1 Let Assumptions 2.1, 2.2, 2.3 and 2.4 hold. Then, under the null hypothesis (4), we have

$$\sup_{\theta \in K} \left| \frac{1}{n} \sum_{t=1}^n u_{t-i} \frac{y_{t-1}}{\sqrt{n}} \pi \left(\frac{y_{t-d}}{\sqrt{n}}, \theta \right) \right| \rightarrow_p 0 \quad (26)$$

as $n \rightarrow \infty$, $i = 1, \dots, p$, for any compact subset K of \mathbb{R}^m .

The result in Lemma 3.1 establishes the asymptotic orthogonality of the leading transition term and the lagged differences. This is not surprising and indeed well expected from the regression theory for the unit root models. Note that the orthogonality in Lemma 3.1 applies uniformly in $\theta \in K$ for any compact subset K of \mathbb{R}^m . The inclusion of the lagged differences would thus have no effect on the testing for the transition term over all possible values of the transition parameter $\theta \in \Theta_n$. It simply washes away the serial correlation in the innovations.

The limit distribution of our inf- t statistic in (21) may now be easily deduced from Theorem 2.3 and Lemma 3.1. As mentioned earlier, W signifies the standard Brownian motion here and elsewhere in the paper.

Theorem 3.2 Let Assumptions 2.1, 2.2, 2.3 and 2.4 hold, and let T_n be defined as in (21). Under the null hypothesis (4), we have

$$T_n = \inf_{\theta \in \Theta_n} T_n(\theta) \rightarrow_d T = \inf_{\theta \in \Theta} T(\theta)$$

as $n \rightarrow \infty$, where $T(\theta)$ is a stochastic process defined by

$$T(\theta) = \frac{\int_0^1 \Pi(\omega W(r), \theta) dW(r)}{\left(\int_0^1 \Pi^2(\omega W(r), \theta) dr \right)^{1/2}} \quad (27)$$

with $\Pi(x, \theta) = x\pi(x, \theta)$.

The distribution of T , in general, depends upon the nuisance parameter ω , which is the long-run variance of (u_t) introduced in (12), as well as the transition function π and the limit parameter space Θ . The dependency of the distribution of T on ω , however, is very simple to deal with and we may easily get rid of it for most of the transition functions used in practical applications. This is shown in the following corollary.

Corollary 3.3 Suppose that $\pi(\omega x, \theta) = \pi(x, \theta^*(\omega, \theta))$ for some θ^* , which does not depend upon $x \in \mathbb{R}$. If we let $\Theta^* = \{\theta^* | \theta^* = \theta^*(\omega, \theta), \theta \in \Theta\}$, then we have $T = \inf_{\theta^* \in \Theta^*} T^*(\theta^*)$, where

$$T^*(\theta^*) = \frac{\int_0^1 \Pi(W(r), \theta^*) dW(r)}{\left(\int_0^1 \Pi^2(W(r), \theta^*) dr \right)^{1/2}} \quad (28)$$

with Π defined as in Theorem 3.2.

If the conditions in Corollary 3.3 are met, the asymptotic critical values of our inf- t test based on T_n depend only upon the transition function π and the limit parameter space Θ . Especially, the dependency of the distribution of T on the long-run variance of (u_t) disappears. The conditions hold virtually all the transition functions used in practical applications. For the TAR, LSTAR and ESTAR models introduced in Section 2.1 with parameters $\theta = \mu$, $\theta = (\mu, \kappa)$ and $\theta = \kappa$, we may easily see that they are satisfied for $\theta^* = \mu^*$, $\theta^* = (\mu^*, \kappa^*)$ and $\theta^* = \kappa^*$, respectively, with $\mu^* = \mu/\omega$ and $\kappa^* = \omega\kappa$.

The actual implementation of our test in the TAR, LSTAR and ESTAR models is fairly simple, due in particular to the result in Corollary 3.3. Let W be the standard Brownian motion given by $V = \omega W$, and define Q^* and P^* from W similarly as Q and P introduced in (24) and (25), i.e., Q^* is the quantile function for W and $P^* = (\int_0^1 W(r)^2 dr)^{-1/2}$. Then we may easily deduce that

$$[Q(15), Q(85)] = \omega [Q^*(15), Q^*(85)]$$

and

$$[10^{-1}P, 10^3P] = (1/\omega)[10^{-1}P^*, 10^3P^*].$$

Therefore, if we denote by μ^* and κ^* the redefined location and scale parameters that are given by $\mu^* = \mu/\omega$ and $\kappa^* = \omega\kappa$, their parameter spaces are given by $[Q^*(15), Q^*(85)]$ and $[10^{-1}P^*, 10^3P^*]$, respectively, for the TAR, LSTAR and ESTAR models. Consequently, the limit null distribution of the inf- t test depends only upon the transition function and the limit parameter space. The critical values for the inf- t test with the parameter spaces in (22) and (23) are tabulated in Table 1 for the TAR, LSTAR and ESTAR models. It may be worth emphasizing once again that they are invariant with respect to the long-run variance of (u_t) , but dependent upon the limit parameter space in a very critical manner.

We now establish the consistency of our inf- t test, which is given in the following proposition.

Proposition 3.4 Suppose that (y_t) is a stationary process with finite second moment, and that there exists a sequence $\theta_n \in \Theta_n$ such that $\text{plim}_{n \rightarrow \infty} \hat{\lambda}_n(\theta_n) < 0$ and the conditional variance of $(y_{t-1}\pi(y_{t-d}, \theta_n))$ given $(\Delta y_{t-1}, \dots, \Delta y_{t-p})$ is nonzero. Moreover, let $\hat{\sigma}_n(\theta)$ be bounded away from zero a.s. uniformly in $\theta \in \Theta_n$ and all large n . Then we have

$$T_n \rightarrow_p -\infty$$

as $n \rightarrow \infty$.

The conditions in Proposition 3.4 are of course satisfied if (y_t) is indeed a stationary process generated by the transitional AR model (3) with the error sequence given by (11). In this case, we have in particular $\hat{\lambda}_n(\theta_0) \rightarrow_p \lambda_0 < 0$, where we denote by θ_0 and λ_0 the true values of θ and λ , respectively. Proposition 3.4 shows that the inf- t test is also consistent against other stationary models as long as for all large n there are parameter values $\theta_n \in \Theta_n$ to fit them with transitional AR models (3) with some $\hat{\lambda}(\theta_n) < 0$.

4. Extensions

Our theories developed in earlier sections can be used to analyze the transitional AR models with more complicated (and more realistic) transition dynamics. In this section, we will show how the necessary extensions can actually be made. The extended theories presented in this section are applicable, in particular, for a wide variety of the transitional AR models with the errors driven by a general linear process of unknown order.

4.1 Models with General Transition Functions

Here we consider the transitional AR model given by

$$\Delta y_t = \lambda \left[\sum_{i=1}^{\ell} \rho_i (y_{t-1} - \nu_i) \pi_i(y_{t-d_i}, \theta_i) \right] + u_t, \quad (29)$$

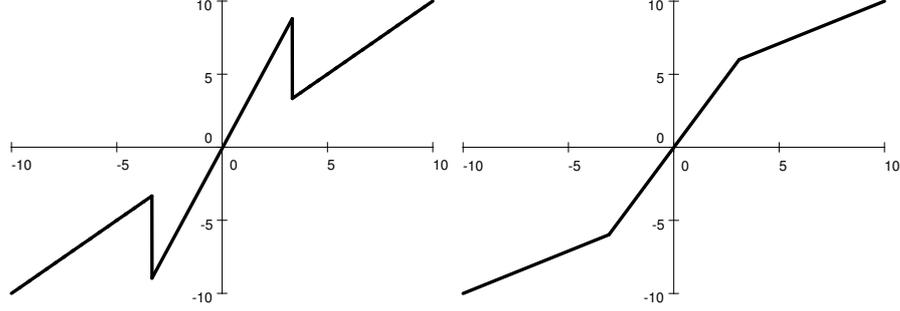


Figure 3: Transition Dynamics for D-TAR Models with No-intercept and Intercept

which is much more general than (3) and includes as special cases the D-TAR and D-LSTAR models introduced in Section 2. The model in (29) allows for nonzero intercepts, as well as multiple transition functions. As a result, it provides much more plausible and realistic transition dynamics for the AR models. The parameter ν_i can be interpreted as the intercept parameter. In many applications, continuous transition dynamics are often preferable to discontinuous ones. However, the models without the intercept parameters generally imply the transition dynamics that are discontinuous. See, for instance, Figure 3 for the transition dynamics for the D-TAR models without and with the intercept term.¹³ On the other hand, the parameter ρ_i allows us to consider asymmetric transition dynamics. It is indeed simple to note that we have asymmetry in the transition dynamics unless $\rho_i = 1$ for all $i = 1, \dots, \ell$. For the D-TAR and D-LSTAR models, we mainly consider them with the symmetry restriction $\rho_1 = \rho_2 = 1$ in the paper.

The null hypothesis (4) in model (29) can be tested based on the regression (15) with $w_t(\theta)$ replaced by

$$w_t(\rho, \nu, \theta) = \sum_{i=1}^{\ell} \rho_i (y_{t-1} - \nu_i) \pi_i(y_{t-d}, \theta_i). \quad (30)$$

We let $\rho = (\rho_1, \dots, \rho_\ell)'$, $\nu = (\nu_1, \dots, \nu_\ell)'$ and $\theta = (\theta'_1, \dots, \theta'_\ell)'$, and set $\rho_1 = 1$ for identification and $\rho_i \geq 0$ for all $i = 1, \dots, \ell$. Also, we denote by Φ_n , Λ_n and Θ_n the sequences of random parameter spaces given respectively for ρ, ν and θ as functions of (y_1, \dots, y_n) . Following our earlier convention made in (6) and (7), we define the transition term (30) as

$$\sum_{i=1}^{\ell} \rho_i \left(\frac{y_{t-1}}{\sqrt{n}} - \nu_i \right) \pi_i \left(\frac{y_{t-d}}{\sqrt{n}}, \theta_i \right) \quad \text{with } (\rho, \nu, \theta) \in (\Phi_n, \Lambda_n, \Theta_n) \quad (31)$$

and

$$\sum_{i=1}^{\ell} \rho_i (y_{t-1} - \nu_i) \pi_i(y_{t-d}, \theta_i) \quad \text{with } (\rho, \nu, \theta) \in (\Phi_n, \Lambda_n, \Theta_n), \quad (32)$$

respectively, under the null and alternative hypothesis. We assume

¹³In Figure 3, the transition dynamics are drawn for the D-TAR models given by $\Delta y_t = -0.7[y_{t-1}1\{y_{t-1} \leq -3\} + y_{t-1}1\{y_{t-1} \geq 3\}] + u_t$ and $\Delta y_t = -0.7[(y_{t-1} + 3)1\{y_{t-1} \leq -3\} + (y_{t-1} - 3)1\{y_{t-1} \geq 3\}] + u_t$.

Assumption 4.1 $(\Phi_n, \Lambda_n, \Theta_n) \rightarrow_d (\Phi, \Lambda, \Theta)$, where (Φ, Λ, Θ) is a compact subset of $\mathbb{R}_+^\ell \times \mathbb{R}^\ell \times \mathbb{R}^{m\ell}$ a.s.

Assumption 4.1 is entirely analogous to Assumption 2.1 with the same notation used for the distributional convergence of the sequence of random parameter spaces. The motivation for the definition of the transition term in (31) and (32) is precisely the same as the one given for (6) and (7). If we denote by ν_{in} and ρ_{in} the parameters normalized under the null hypothesis of a unit root as in (31), then they are given by $\nu_{in} = \nu_i/\sqrt{n}$ and $\rho_{in} = \sqrt{n}\rho_i$ in terms of the unnormalized original parameters ν_i and ρ_i , for $i = 1, \dots, \ell$.

Assumption 4.2 π_i satisfies the conditions in Assumption 2.2 for $i = 1, \dots, \ell$.

Like the parameter θ , the parameters ρ and ν are identified only under the alternative hypothesis. Therefore, the infimum of the t -ratio should now be taken with respect to ρ and ν , as well as θ . If we let $A_n(\rho, \nu, \theta)$ and $B_n(\rho, \nu, \theta)$ be defined similarly as in (19) and (20) with $w_t(\theta)$ replaced by $w_t(\rho, \nu, \theta)$, then the t -ratio is given for each value of (ρ, ν, θ) by

$$T_n(\rho, \nu, \theta) = \frac{\hat{\lambda}_n(\rho, \nu, \theta)}{s(\hat{\lambda}_n(\rho, \nu, \theta))},$$

where $\hat{\lambda}_n(\rho, \nu, \theta) = A_n(\rho, \nu, \theta)/B_n(\rho, \nu, \theta)$ and $s(\hat{\lambda}_n(\rho, \nu, \theta)) = \hat{\sigma}_n(\theta)/\sqrt{B_n(\rho, \nu, \theta)}$ with the usual error variance estimate $\hat{\sigma}_n^2(\theta)$ introduced earlier. The inf- t test may then be defined as

$$T_n = \inf_{(\rho, \nu, \theta) \in \Phi_n \times \Lambda_n \times \Theta_n} T_n(\rho, \nu, \theta) \quad (33)$$

in place of (21).

Our test here is applicable in particular to the D-TAR and D-LSTAR models with no intercept, and also to the TAR, LSTAR, ESTAR, D-TAR and D-LSTAR models with intercept. For the D-TAR and D-LSTAR models with no intercept, we have parameters $\theta = (\mu_1, \mu_2)$ and $\theta = (\mu_1, \mu_2, \kappa_1, \kappa_2)$ respectively. The parameter spaces for the location parameters μ_i 's and the scale parameters κ_i 's may be set as in (22) and (23) for these models. For the TAR, LSTAR, ESTAR, D-TAR and D-LSTAR models with intercept, we need to set the parameter spaces for the parameters ν_i 's and ρ_i 's. Unless there are reasons to specify otherwise, it seems natural to also set the parameter spaces for the parameters ν_i 's and ρ_i 's to be exactly as those for the parameters μ_i 's and κ_i 's, respectively, in (22) and (23). Note that ν_i 's and ρ_i 's also designate the location and scale of the transition functions. We use these parameter spaces in our simulations and empirical applications.

Theorem 4.1 Let Assumptions 2.3, 2.4, 4.1 and 4.2 hold, and let T_n be defined as in (33). Under the null hypothesis (4), we have

$$T_n = \inf_{(\rho, \nu, \theta) \in \Phi_n \times \Lambda_n \times \Theta_n} T_n(\rho, \nu, \theta) \rightarrow_d T = \inf_{(\rho, \nu, \theta) \in \Phi \times \Lambda \times \Theta} T(\rho, \nu, \theta)$$

as $n \rightarrow \infty$, where $T(\rho, \nu, \theta)$ is a stochastic process defined similarly as in (27) with $\Pi(\omega W(r), \theta)$ replaced by $\Pi(\omega W(r), \rho, \nu, \theta)$ for $\Pi(x, \rho, \nu, \theta) = \sum_{i=1}^{\ell} \rho_i(x - \nu_i)\pi(x, \theta_i)$.

The results in Theorem 4.1 are comparable to those in Theorem 3.2. The distribution of T given in Theorem 4.1 depends upon the nuisance parameter ω , just as that of T in Theorem 3.2, as well as the transition function π and the limit parameter space $\Phi \times \Lambda \times \Theta$. The dependency on ω of the limit distribution of T , however, can easily be dealt with similarly as in Corollary 3.3. The following corollary extends our results in Corollary 3.3.

Corollary 4.2 Suppose that $\pi(\omega x, \theta) = \pi(x, \theta^*(\omega, \theta))$ for some θ^* , which does not depend upon $x \in \mathbb{R}$. If we let $\Theta^* = \{\theta^* | \theta^* = \theta^*(\omega, \theta), \theta \in \Theta\}$, $\Lambda^* = \{\nu^* | \nu^* = \nu/\omega, \nu \in \Lambda\}$ and $\Phi^* = \{\rho^* | \rho^* = \omega\rho, \rho \in \Phi\}$, then we have $T = \inf_{(\rho^*, \nu^*, \theta^*) \in \Phi^* \times \Lambda^* \times \Theta^*} T^*(\rho^*, \nu^*, \theta^*)$, where $T^*(\rho^*, \nu^*, \theta^*)$ is a stochastic process defined similarly as in (28) with $\Pi(W(r), \theta^*)$ replaced by $\Pi(W(r), \rho^*, \nu^*, \theta^*)$ for Π defined as in Theorem 4.1.

Corollary 4.2 implies that the critical values of our inf- t test are only dependent upon the transition function and the limit parameter space for a wide class of transitional AR models. This is exactly identical to what is implied by Corollary 3.3.

For the D-TAR and D-LSTAR models with the parameters given by $\theta = (\mu_1, \mu_2)$ and $\theta = (\mu_1, \mu_2, \kappa_1, \kappa_2)$, the conditions in Corollary 4.2 are satisfied for $\theta^* = (\mu_1^*, \mu_2^*)$ and $\theta^* = (\mu_1^*, \mu_2^*, \kappa_1^*, \kappa_2^*)$, respectively, with $\mu_i^* = \mu_i/\omega$ and $\kappa_i^* = \omega\kappa_i$, $i = 1, 2$. For the TAR, LSTAR, ESTAR, D-TAR and D-LSTAR models with intercept, we have additional parameters ν (ν_i 's) and ρ (ρ_i 's), for which the conditions in Corollary 4.2 hold with the redefined parameters ν^* (ν_i^* 's) and ρ^* (ρ_i^* 's) given by $\nu^* = \nu/\omega$ ($\nu_i^* = \nu_i/\omega$) and $\rho^* = \omega\rho$ ($\rho_i^* = \omega\rho_i$), similarly as μ^* (μ_i^* 's) and κ^* (κ_i^* 's), respectively. The critical values of the inf- t test in all these models are therefore invariant with respect to the long-run variance of (u_t) . The critical values of the inf- t test are tabulated in Table 1 for the D-TAR and D-LSTAR models, and in Table 2 for the TAR, LSTAR, ESTAR, D-TAR and D-LSTAR models with intercept. All the critical values are computed using (22) and (23) as the parameter spaces respectively for the location and scale parameters in these models.

Finally, the consistency of the inf- t test can be established for the general models exactly as in Proposition 3.4. It is, however, trivial, and we do not state it as a separate proposition to save the space.

4.2 Models with Errors of Unknown Form

It is possible to further accommodate a broader class of transitional autoregressive models in many other directions. In particular, we may show that our tests are valid for the models with (u_t) driven by a general linear process. Now we specify (u_t) as

$$u_t = \varphi(L)\varepsilon_t = \sum_{i=1}^{\infty} \varphi_i \varepsilon_{t-i}, \quad (34)$$

where we assume that

Assumption 4.3 Let $\varphi(z) \neq 0$ for all $|z| \leq 1$, and $\varphi_k = O(k^{-s-\delta})$ for some $s \geq 2$ and $\delta > 0$.

With the coefficient summability condition in Assumption 4.3, condition (a) of Assumption 2.3 implies that (u_t) is weakly stationary. Moreover, condition (b) of Assumption 2.3 guarantees that the r -th moment of (u_t) exists and is uniformly bounded for all $t \in \mathbb{Z}$, due to the Marcinkiewicz-Zygmund inequality in, e.g., Sout (1974, Theorem 3.3.6).

Due to the so-called the Beveridge-Nelson decomposition, we may write (u_t) given in (34) as

$$u_t = \varphi(1)\varepsilon_t + (\tilde{u}_{t-1} - \tilde{u}_t),$$

where

$$\tilde{u}_t = \sum_{i=0}^{\infty} \tilde{\varphi}_i \varepsilon_{t-i}, \quad \tilde{\varphi}_i = \sum_{j=i+1}^{\infty} \varphi_j.$$

Under Assumption 4.3, we have $\sum_{i=0}^{\infty} |\tilde{\varphi}_i| < \infty$ as shown in Phillips and Solo (1992), and therefore (\tilde{u}_t) is well defined both in a.s. and L^r sense [see Brockwell and Davis (1991)]. Moreover, we have

$$\frac{1}{\sqrt{n}} \sum_{t=1}^i u_t = \varphi(1) \frac{1}{\sqrt{n}} \sum_{t=1}^i \varepsilon_t + o_p(1)$$

uniformly in $1 \leq i \leq n$. Therefore, the asymptotics in Theorem 2.3 hold also in this case with the definition of V by $V(r) = \varphi(1)U(r)$ and $\omega^2 = \sigma^2 \varphi(1)^2$.¹⁴

Under Assumptions 2.3 and 4.3, we may let $\alpha(L)u_t = \varepsilon_t$ with $\alpha(z) = 1 - \sum_{i=1}^{\infty} \alpha_i z^i$, and approximate (u_t) in r -th mean by a finite order AR process

$$u_t = \alpha_1 u_{t-1} + \cdots + \alpha_p u_{t-p} + \varepsilon_{t,p}$$

with

$$\varepsilon_{t,p} = \varepsilon_t + \sum_{i=p+1}^{\infty} \alpha_i u_{t-i}.$$

As is well known [see, e.g., Brillinger (1975)], the condition in Assumption 4.3 implies that $\alpha_k = O(k^{-s-\delta})$, and we have $\sum_{i=p+1}^{\infty} |\alpha_i| = o(p^{-(s-1)})$. Given the existence of the r -th moment of (u_t) implied by Assumptions 2.3 and 4.1, we therefore have

$$\mathbb{E}|\varepsilon_{p,t} - \varepsilon_t|^r \leq \mathbb{E}|u_t|^r \left(\sum_{i=p+1}^{\infty} |\alpha_i| \right)^r = o(p^{-r(s-1)}).$$

The approximation error thus becomes small as p gets large.

It is now well expected that our tests are valid for the models driven by general linear processes of unknown order under suitable conditions, if we let the order of fitted autoregression increase as the sample size gets large. This is what we will show below. We start by introducing assumptions on the order of fitted autoregression and the transition function. For the order of fitted autoregression, we write $p = p_n$ to make it explicit that p is a function of the sample size n and assume

¹⁴Now we require that part (c) of Assumption 2.2 holds with this newly defined ω^2 in place of the one given in (12).

Assumption 4.4 Let $p_n \rightarrow \infty$ and $p_n = o(n^{1/2})$ as $n \rightarrow \infty$.

The condition in Assumption 4.4 is very mild and the same as the one used in Chang and Park (2002) to derive the asymptotics for the ADF unit root t -test.

We now consider the model (29) with the transition function satisfying the conditions in

Assumption 4.5 $\pi_i(\cdot, \theta)$ is differentiable for all $\theta \in \mathbb{R}^m$ with derivatives bounded uniformly in θ on any compact subset K of \mathbb{R}^m , for $i = 1, \dots, \ell$.

The conditions in Assumption 4.5 are stronger and less general than those introduced originally in Assumption 2.2. In particular, they do not hold for threshold autoregressive models like TAR. However, they are satisfied for all smooth transition autoregressive models such as LSTAR and ESTAR.

Lemma 4.3 Let Assumptions 2.1, 2.3, 4.3, 4.4 and 4.5 hold. Then, under the null hypothesis (4), we have

$$\sup_{\theta \in K} \left| \frac{1}{\sqrt{n}} \sum_{t=1}^n u_{t-i} \frac{y_{t-1}}{\sqrt{n}} \pi_j \left(\frac{y_{t-d}}{\sqrt{n}}, \theta \right) \right| = O_p(1) \quad (35)$$

as $n \rightarrow \infty$, $i \geq 1$ and $j = 1, \dots, \ell$, for any compact subset K of \mathbb{R}^m .

The result in Lemma 4.3 ensures the asymptotic orthogonality between the transition term and all of the lagged differences, the number of which increases as the sample size gets large. It is comparable to the result in Lemma 3.1. However, the obtained bound for the cross product of the transition function and the differenced lag terms is tighter in Lemma 4.3, compared with the one given in Lemma 3.1. This is to allow for the number of the lagged differences to increase. The tighter bound obtained in Lemma 4.3 requires the differentiability of the transition function introduced in Assumption 4.5.

Theorem 4.4 Under Assumptions 2.3, 4.1, 4.3, 4.4 and 4.5, Theorem 4.1 holds.

Under appropriate conditions, our previous results on the inf- t tests are therefore applicable for the models driven by general linear processes of unknown order.

5. Monte Carlo Experiment

In this section, we conduct simulation experiments to investigate the finite sample performance of our tests in comparison with that of the conventional ADF test. In particular, the inf- t tests are examined in three transition functions, namely, (a) TAR, (b) LSTAR, and (c) ESTAR models, using the asymptotic critical values provided in Table 1. Our simulations are based on the AR model given by

$$\Delta y_t = \lambda y_{t-1} \pi(y_{t-1}, \theta) + \alpha \Delta y_{t-1} + \varepsilon_t, \quad (36)$$

where (ε_t) is iid $N(0, 1)$. We consider the samples of the sizes $n = 50, 100$ and 200 . All the results reported in this section are based on 10,000 replications and for the nominal 5% tests. The order p of the AR model is assumed to be known and set to 1. The parameter spaces are set according to (22) and (23) respectively for the location and scale parameters.

5.1 Sizes

To investigate the sizes of the *inf-t* tests, the time series are generated from the model (36) with $\lambda = 0$ and $\alpha \in \{-0.5, -0.2, 0, 0.2, 0.5\}$. Table 3 reports the actual rejection frequencies of the *inf-t* tests and compares them with those of the ADF test. In general, the *inf-t* tests have reasonable size properties and are comparable to the ADF test. In particular, they seem to work reasonably well and have acceptable size properties for the sample sizes that are typical in practice. When the sample size is very small, however, the size distortions can be substantial. In this case, it seems desirable to make the finite sample adjustments and use the size-corrected critical values, based on the fitted AR models. For some of the models that are used in our empirical applications, we find that the size-corrected critical values can be moderately different from the asymptotic critical values provided in Tables 1 and 2.

Of the three transitional AR models that we consider in our simulations, the TAR model yields the most noticeable small sample size distortions for the *inf-t* test. The distortions of the *inf-t* tests are less significant in the LSTAR model, and much less so in the ESTAR model. In both the TAR and LSTAR models, the *inf-t* tests tend to under-reject the null hypothesis of a unit root, especially when the AR coefficient is negative and the generated time series have negative autocorrelations. The under-rejections, however, appear to become less frequent as the sample size increases. In contrast, the *inf-t* test somewhat over-rejects the null hypothesis of a unit root in the ESTAR models for all values of AR coefficients, though the over-rejection is fairly mild at all sample sizes.

5.2 Powers

To evaluate the powers of the *inf-t* tests, we set $\lambda < 0$ in the same model (36) to generate the time series. We fix $\alpha = 0$, but vary the parameters λ and θ . To avoid the effects of the size distortions on our evaluations of the powers, we compute the size-corrected critical values and calculate the size-adjusted powers. Our simulation results for the size-adjusted powers of the *inf-t* tests are given in Tables 4 and 5.

Table 4 reports the results for the TAR model with the ranges of parameter values $\lambda \in \{-1, -0.5, -0.3, -0.1\}$ and $\mu \in \{-1, -0.5, 0, 0.5, 1\}$. In general, the powers of all the tests become larger when the location parameter μ becomes larger in absolute value. When the stationary regime has less persistent roots, including $\lambda = -1$ and -0.5 , the *inf-t* test outperforms the ADF test substantially regardless of the sample size and the location parameter. When the root becomes close to unity with $\lambda = -0.1$, the power of the *inf-t* test in the TAR model is less than that of the ADF test with a small sample size. However, even in such a case, the power of the *inf-t* test eventually becomes larger as the sample size increases.

Table 5 reports the results for the LSTAR and ESTAR models with $\lambda \in \{-1, -0.5, -0.3, -0.1\}$. The upper panel provides the results for the LSTAR model with $(\mu, \kappa) = (0, 1), (1, 1), (0, 20)$, and $(1, 20)$. For both the inf- t test and ADF test, the powers are generally higher when the location parameter μ is large (i.e., $\mu = 1$) and when the scale parameter κ is small (i.e., $\kappa = 1$). The comparison between the inf- t test and the ADF test in these cases is very similar to that in the TAR case in the sense that the inf- t test eventually outperforms the ADF test as the sample size increases, even if the stationary regime contains a root near unity. When the scale parameter becomes as large as $\kappa = 20$, the results for the LSTAR model are almost identical to those for the TAR model. This is reasonable and well expected, since the transition function of the LSTAR model approaches that of the TAR model as the scale parameter gets large.

The lower panel of Table 5 presents the results for the ESTAR model with the scale parameter $\kappa \in \{0.10, 0.15, 0.20\}$. These are the values of κ , for which the difference in the power properties of the inf- t test and the ADF test are most apparent. The finite sample powers of both tests depend crucially on the value of κ . When the parameter κ takes small values such as those chosen above for our simulations, the inf- t test clearly outperforms the ADF test. However, when κ becomes large, the powers of both tests quickly approach 100% as the sample size increases. In this sense, the advantage of the inf- t test in the ESTAR model is not as obvious as in the TAR and LSTAR models, and can vary widely across the models with different parameter values. For this reason, it would be more informative to assess the powers of the tests in each empirical application, using the models with estimated parameters or the parameter values that seem to be relevant.

6. Empirical Results

For the purpose of illustration, we apply in this section the inf- t tests to several economic time series based on the transitional AR models. In our empirical applications, the parameter spaces are set according to (22) and (23), respectively for the location and scale parameters, precisely as in our simulations. Throughout this section, the lag length p is selected based either on the previous studies or on the selection procedures such as AIC and BIC under the null hypothesis. The lag delay d is not identified under the null, and therefore, we follow Caner and Hansen (2001) and select d that minimizes the inf- t test statistic.

6.1 Unemployment Rates

Our first application is to the U.S. unemployment rate, which has been often investigated using transitional AR models in the literature. Among others, Rothman (1998) and Caner and Hansen (2001) considered the TAR model and van Dijk, Terasvirta and Franses (2002) estimated the LSTAR model. The data we use in our application here are seasonally adjusted monthly U.S. unemployment rates among males of 20 years and over, from January 1948 to December 2001 ($n = 636$), and obtained from the Bureau of Labor Statistics.

In particular, we employ the *inf-t* test for the TAR model given by

$$\Delta y_t = \lambda(y_{t-1} - \nu)1\{y_{t-d} \leq \mu\} + \sum_{i=1}^p \alpha_i \Delta y_{t-i} + \varepsilon_t.$$

One possible interpretation of this model is to consider the transition between a traditional ‘natural rate’ hypothesis and the ‘hysteresis’ hypothesis of Blanchard and Summers (1987), since the former dynamics are often described by reversion to the long-run level or the natural rate of unemployment, while the latter predict high persistence or random walk-like behavior. In the current context, the unemployment rate reverts to the natural rate ν as long as the lagged unemployment rate is below the threshold point μ , but it is characterized by the unit root process if it exceeds the threshold level. In addition to the TAR model, we also consider the LSTAR model

$$\Delta y_t = \lambda(y_{t-1} - \nu) [1 + \exp\{\kappa(y_{t-d} - \mu)\}]^{-1} + \sum_{i=1}^p \alpha_i \Delta y_{t-i} + \varepsilon_t,$$

which implies a smooth transition. In both the TAR and LSTAR models, we set $p = 12$. This is to make our results directly comparable to those in Caner and Hansen (2001).¹⁵

The results based on both the ADF tests and *inf-t* tests are reported in Table 6. The values of the ADF tests are -2.74 and -2.82, respectively, with a constant (ADF- μ) and with a trend (ADF- τ). Since the 5% asymptotic critical values of the two tests are -2.86 and -3.41, respectively, neither of them rejects the null hypothesis. The unit root hypothesis is therefore supported unambiguously. The conclusion, however, is sharply reversed when the *inf-t* tests are used. The *inf-t* test yields the value -3.72 if based on the TAR model, and the value -3.72 if based on the LSTAR. Since the 5% asymptotic critical values of the two tests are -3.39 and -3.45, respectively, both tests significantly reject the unit root hypothesis. In particular, it seems clear from our results that the transitional AR models are much more plausible alternatives to the random walk model, compared with the linear stationary AR model.

Taking into account the possible small sample biases in sizes and powers that we observed in the simulations, we also consider the tests more explicitly based on the estimated models. The time series are therefore generated directly from the estimated models, i.e., the models given by the estimated parameters and the actual sample size, under both the null and alternative hypotheses, and subsequently use them to obtain the size-corrected critical values and calculate the finite sample powers. In each case, we replicate the computation 2,000 times to get the reported results. In sum, the use of the size-corrected critical values does not change the results of the ADF and *inf-t* tests. Only the *inf-t* tests reject the null hypothesis of a unit root. The comparison of the size-adjusted powers reported in the lower half of Table 6, however, shows a clear advantage of the *inf-t* test over the ADF test at least for this particular example.

¹⁵We also get this choice of lag order if AIC is used. On the other hand, BIC selects $p = 3$.

6.2 Purchasing Power Parity

We next consider the ESTAR model for the real exchange rate. The presence of various types of transaction costs generally implies slower adjustment for a larger deviation of the current aggregate price from the equilibrium level predicted by the purchasing power parity. To describe such a continuous nonlinear speed of adjustment in the aggregate real exchange rate in two directions, the ESTAR model has been popularly used in many studies including Michael, Nobay and Peel (1997) and Taylor, Peel and Sarno(2001). Here we follow Taylor, Peel and Sarno (2001) and consider the ESTAR model given by

$$\Delta y_t = \lambda(y_{t-1} - \nu)[1 - \exp\{-\kappa^2(y_{t-d} - \nu)^2\}] + \sum_{k=1}^p \alpha_k \Delta y_{t-k} + \varepsilon_t.$$

The lag order is chosen by BIC and set at $p = 1$.¹⁶ The seasonally adjusted monthly U.K./U.S. (log) real exchange rate series is constructed from the U.K. sterling pound and the U.S. dollar nominal exchange rates, as well as the consumer price indexes of the U.K. and U.S. Each series is obtained from the International Monetary Fund's International Financial Statistics, and the sample period is from January 1973 to July 2004 ($n = 379$).

The results are reported in Table 7. As in the case of the unemployment rate, the null hypothesis of a unit root is not rejected based on the ADF tests at the 5% significance level. The test value is -2.61 or -2.87 depending upon whether we include a constant or linear time trend. The unit root hypothesis is, however, rejected against the ESTAR model, if based on our inf- t test. The test yields the value -3.35, while the asymptotic critical value is given by -3.30. The conclusion does not change even if we employ the size-adjusted critical values instead of the asymptotic critical values. Moreover, the power comparison based on the estimated models shows the superior power property of the inf- t test compared to the ADF test. Once again, it seems that the nonlinear transitional AR model is a very plausible and convincing alternative to the random walk model.

6.3 Target Zone Model

Our third empirical example is an application of the D-TAR and D-LSTAR-based inf- t tests to the target zone exchange rate model. During the 1980s and 1990s, the exchange rates within the European Monetary System (EMS) were subject to intervention by central banks based on the bands set at $\pm 2.25\%$ around the central parity in most of the member countries. As pointed out by Balke and Fomby (1997, p.628), the target zone systems may well be characterized by the D-TAR model with a partial unit root in the inner regime, since the exchange rates are allowed to fluctuate freely within the band. While the official band is set at $\pm 2.25\%$ around the central parity, it is likely that intervention begins before the rate actually hits $\pm 2.25\%$ points. Therefore, it seems reasonable to consider unknown threshold values μ_1 and μ_2 in the transitional AR models. In particular, we consider the

¹⁶It also corresponds to the lag order selected by AIC.

D-TAR and D-LSTAR models specified as

$$\Delta y_t = \lambda [(y_{t-1} - \nu_1)1\{y_{t-d} \leq \mu_1\} + (y_{t-1} - \nu_2)1\{y_{t-d} \geq \mu_2\}] + \sum_{k=1}^p \alpha_k \Delta y_{t-k} + \varepsilon_t$$

$$\Delta y_t = \lambda \left[\frac{y_{t-1} - \nu_1}{1 + e^{\kappa(y_{t-d} - \mu_1)}} + \frac{y_{t-1} - \nu_2}{1 + e^{-\kappa(y_{t-d} - \mu_2)}} \right] + \sum_{k=1}^p \alpha_k \Delta y_{t-k} + \varepsilon_t.$$

The lag order is chosen by BIC and set at $p = 5$.¹⁷ For the empirical analysis here, we use the daily French franc/Deutschemark exchange rates measured in log deviations from the central parity. The sample period is from 1979/3/13 to 1993/7/30 ($n = 3645$), during which the central rate was realigned six times.

The results are reported in Table 8. For this series, both the ADF tests with a constant and with a trend reject the unit root hypothesis at the 5% significance level. At the same time, the *inf-t* tests reject the null hypothesis against both the D-TAR and D-LSTAR models. The tests yield the values -5.94 and -5.75, which are far below their asymptotic 5% critical values that are given by -3.04 and -3.13, respectively for the D-TAR and D-LSTAR models. The test results do not change even if we use the size-adjusted critical values. We draw the same conclusions that are in favor of the transitional AR models. It appears that our conclusions here are strong and unambiguous. The size-adjusted powers of our tests are almost perfect. They are indeed 100% in all cases. However, these results are not very surprising given the fact that the series is by definition confined within a band, as well as the availability of the series with a large sample size.

Interestingly, the conclusion can differ drastically from the above if we consider subsamples of smaller sizes. To demonstrate such a possibility, we consider a subsample based on the series only up to the time of the third realignment (in 1982) in the total of six realignments. The results from this experiment are also provided in Table 8. For this subsample ($n = 819$), the ADF tests fail to reject the null hypothesis of a unit root. On the other hand, both the *inf-t* tests applied to the D-TAR and D-LSTAR models yield the values -4.26 and -4.53, respectively, and reject the unit root hypothesis rather strongly. The higher size-adjusted powers of the *inf-t* tests, in comparison with those of the ADF tests, are also consistent with the fact that only the *inf-t* tests can reject the unit root in this subsample case.

7. Conclusion

In this paper, we consider the test of a unit root against the transitional AR models. Our framework is truly general, allowing for a wide range of AR models with threshold, discrete and smooth transition dynamics. The models that we study in the paper include all the transitional AR models considered previously in the literature. Moreover, we only impose very mild conditions on the innovation sequence, which are minimal in the sense that they are also required for the validity of standard unit root tests in the linear AR models. Under

¹⁷If AIC is used, $p = 6$ is selected.

this very general and flexible specification, we investigate the $\text{inf-}t$ test that can be used to effectively discriminate the unit root model from the stationary transitional AR models. Such a test is motivated by the fact that the parameters in the transitional AR models are not identified under the null hypothesis of a unit root. The full asymptotic theory for the $\text{inf-}t$ test is developed in the paper. In particular, we show that the test has well defined limit distribution that is free of any nuisance parameters and depends only on the transition function and the limit parameter space. The critical values of the transitional AR models that are used most frequently in practical applications are tabulated in the paper.

The small sample performance of the $\text{inf-}t$ test is very encouraging. The test has a reasonably good size property with sample sizes typically available in practice. It also has a rather satisfactory power property in finite samples. While the power performance somewhat depends on the range of parameters when the sample size is small, the $\text{inf-}t$ test generally has significantly better power than the usual ADF test. The power gain is often very substantial, for some of the empirically relevant cases. For illustrations, we examine several economic time series and test of a unit root against some popular transitional AR models. The results are quite surprising. The unit root model is unambiguously rejected for all the series that we investigated, in favor of the stationary transitional AR models. The evidence is either reversed or becomes much weaker, if the standard unit root test is used. It seems apparent that the transitional AR models are much more promising than the linear AR models, in explaining the data generating process of many economic time series. Most of all, the ubiquitous unit root in many economic time series disappears, as we allow for the nonlinear transition dynamics.

Appendix: Mathematical Proofs

The proofs of some of our theorems and lemmas utilize the local time of Brownian motion V , which is defined by

$$\ell(r, x) = \lim_{\delta \rightarrow 0} \frac{1}{2\delta} \int_0^r 1_{\{|V(s) - x| \leq \delta\}} ds \quad (37)$$

The reader is referred to, e.g., Chung and Williams (1990) for an elementary introduction to the local time, and to Park and Phillips (1999, 2001) for its application in the study of nonlinear models with integrated time series. The local time L yields the occupation times formula

$$\int_0^r v(V(r)) dr = \int_{\mathbb{R}} v(x) \ell(r, x) dx, \quad (38)$$

which holds for any nonnegative measurable function v on \mathbb{R} .

Useful Lemmas and Their Proofs

Lemma A1 (a) If π is regular on $\mathbb{R} \times \mathbb{R}^m$, then so are $|\pi|$ and π^2 .
 (b) If π is regular on $\mathbb{R} \times \mathbb{R}^m$ and v is regular on \mathbb{R} , then the functional $v\pi$, defined by $(v\pi)(x, \theta) = v(x)\pi(x, \theta)$, is regular on $\mathbb{R} \times \mathbb{R}^m$.

Proof of Lemma A1 Fix $\theta_0 \in \mathbb{R}^m$ arbitrarily and let $\underline{\pi}_\delta$ and $\overline{\pi}_\delta$ be the regular functions on \mathbb{R} such that $\underline{\pi}_\delta \leq \pi(x, \theta) \leq \overline{\pi}_\delta(x)$ for all $\|\theta - \theta_0\| < \delta$, and such that $\overline{\pi}_\delta(x) - \underline{\pi}_\delta(x) \rightarrow 0$ a.e. $x \in \mathbb{R}$ as $\delta \rightarrow 0$. For the proof of part (a), we may assume $\pi \geq 0$ without loss of generality. As in the proof of Lemma A1 in Park and Phillips (2001), we may write $\pi = \pi^+ - \pi^-$, where π^+ and π^- are respectively the positive and negative parts of π , and consider them separately. Under this convention, the stated result for $|\pi|$ is trivial. For π^2 , we consider $\underline{\pi}_\delta^2$ and $\overline{\pi}_\delta^2$. It is obvious that $\underline{\pi}_\delta^2 \leq \pi^2(x, \theta) \leq \overline{\pi}_\delta^2$ for all $\|\theta - \theta_0\|$. Also, they are regular, as shown in Lemma A1 of Park and Phillips (2001). Moreover, we have

$$\overline{\pi}_\delta^2(x) - \underline{\pi}_\delta^2(x) = (\overline{\pi}_\delta(x) + \underline{\pi}_\delta(x)) (\overline{\pi}_\delta(x) - \underline{\pi}_\delta(x)) \rightarrow 0$$

as $\delta \rightarrow 0$ a.e. $x \in \mathbb{R}$, since being regular $\underline{\pi}_\delta$ and $\overline{\pi}_\delta$ are bounded uniformly in $\delta > 0$ for any $x \in \mathbb{R}$. The proof of (a) is therefore complete. To prove part (b), we look at $v\underline{\pi}_\delta$ and $v\overline{\pi}_\delta^2$, for which the rest of the proof goes entirely analogously as for part (a). \square

Lemma A2 Let Assumptions 2.1, 2.3 and 4.1 hold. If ϖ is regular on $\mathbb{R} \times \mathbb{R}^m$, then

$$\int_0^1 \varpi(V_n(r), \theta) dr \xrightarrow{a.s.} \int_0^1 \varpi(V(r), \theta) dr$$

uniformly in $\theta \in K$, as $n \rightarrow \infty$, where K is any compact subset of \mathbb{R}^m .

Proof of Lemma A2 Choose $\theta_0 \in K$ arbitrarily. Since ϖ is regular on $\mathbb{R} \times \mathbb{R}^m$, there exists a δ -sequence of regular functions $(\underline{\varpi}_\delta)$ and $(\overline{\varpi}_\delta)$ such that

$$\underline{\varpi}_\delta(x) \leq \varpi(x, \theta) \leq \overline{\varpi}_\delta(x) \quad (39)$$

for all $\|\theta - \theta_0\| < \delta$, and

$$\overline{\varpi}_\delta(x) - \underline{\varpi}_\delta(x) \rightarrow 0 \quad (40)$$

a.e. $x \in \mathbb{R}$.

Since $\underline{\varpi}_\delta$ and $\overline{\varpi}_\delta$ are regular, we have

$$\int_0^1 \underline{\varpi}_\delta(V_n(r)) dr \rightarrow_{a.s.} \int_0^1 \underline{\varpi}_\delta(V(r)) dr \quad (41)$$

$$\int_0^1 \overline{\varpi}_\delta(V_n(r)) dr \rightarrow_{a.s.} \int_0^1 \overline{\varpi}_\delta(V(r)) dr \quad (42)$$

as shown in Park and Phillips (2001). Moreover,

$$\begin{aligned} \int_0^1 \overline{\varpi}_\delta(V(r)) dr - \int_0^1 \underline{\varpi}_\delta(V(r)) dr &= \int_0^1 [\overline{\varpi}_\delta(V(r)) - \underline{\varpi}_\delta(V(r))] dr \\ &= \int_{\mathbb{R}} [\overline{\varpi}_\delta(x) - \underline{\varpi}_\delta(x)] \ell(1, x) dr \rightarrow_{a.s.} 0 \end{aligned} \quad (43)$$

as $\delta \rightarrow 0$, due to (40) and dominated convergence. Now the stated result follows from (39), (41) and (42) as in Park and Phillips (2001). \square

Lemma A3 Let Assumptions 2.1, 2.2, 2.3 and 4.1 hold, and let (v_t) be a time series satisfying $(1/n) \sum_{t=1}^n v_t^2 = O_p(1)$ for all $n \in \mathbb{N}$. Then we have

$$\sup_{\theta \in K} \left| \frac{1}{n} \sum_{t=1}^n v_t \left[\frac{y_t}{\sqrt{n}} \pi \left(\frac{y_t}{\sqrt{n}}, \theta \right) - \frac{y_{t-1}}{\sqrt{n}} \pi \left(\frac{y_{t-1}}{\sqrt{n}}, \theta \right) \right] \right| \rightarrow_p 0$$

as $n \rightarrow \infty$, for any compact subset K of \mathbb{R}^m .

Proof of Lemma A3 Since

$$\begin{aligned} &\frac{y_t}{\sqrt{n}} \pi \left(\frac{y_t}{\sqrt{n}}, \theta \right) - \frac{y_{t-1}}{\sqrt{n}} \pi \left(\frac{y_{t-1}}{\sqrt{n}}, \theta \right) \\ &= \frac{u_t}{\sqrt{n}} \pi \left(\frac{y_t}{\sqrt{n}}, \theta \right) + \frac{y_{t-1}}{\sqrt{n}} \left[\pi \left(\frac{y_t}{\sqrt{n}}, \theta \right) - \pi \left(\frac{y_{t-1}}{\sqrt{n}}, \theta \right) \right], \end{aligned}$$

it suffices to show that

$$\sup_{\theta \in K} \left| \frac{1}{n} \sum_{t=1}^n v_t \frac{u_t}{\sqrt{n}} \pi \left(\frac{y_t}{\sqrt{n}}, \theta \right) \right| \rightarrow_p 0 \quad (44)$$

$$\sup_{\theta \in K} \left| \frac{1}{n} \sum_{t=1}^n v_t \frac{y_{t-1}}{\sqrt{n}} \left[\pi \left(\frac{y_t}{\sqrt{n}}, \theta \right) - \pi \left(\frac{y_{t-1}}{\sqrt{n}}, \theta \right) \right] \right| \rightarrow_p 0 \quad (45)$$

as $n \rightarrow \infty$, for any compact subset K of \mathbb{R}^m .

It is easy to show (44), since we have by Cauchy-Schwarz inequality and the boundedness of π

$$\left| \frac{1}{n} \sum_{t=1}^n v_t \frac{u_t}{\sqrt{n}} \pi \left(\frac{y_t}{\sqrt{n}}, \theta \right) \right| \leq \frac{c}{\sqrt{n}} \left(\frac{1}{n} \sum_{t=1}^n v_t^2 \right) \left(\frac{1}{n} \sum_{t=1}^n u_t^2 \right)$$

for all $\theta \in \mathbb{R}^m$, with some constant $c > 0$. To prove (45), we use Cauchy-Schwarz inequality to obtain

$$\begin{aligned} & \left| \frac{1}{n} \sum_{t=1}^n v_t \frac{y_{t-1}}{\sqrt{n}} \left[\pi \left(\frac{y_t}{\sqrt{n}}, \theta \right) - \pi \left(\frac{y_{t-1}}{\sqrt{n}}, \theta \right) \right] \right| \\ & \leq \left(\frac{1}{n} \sum_{t=1}^n v_t^2 \right)^{1/2} \left(\frac{1}{n} \sum_{t=1}^n \left(\frac{y_{t-1}}{\sqrt{n}} \right)^2 \left[\pi \left(\frac{y_t}{\sqrt{n}}, \theta \right) - \pi \left(\frac{y_{t-1}}{\sqrt{n}}, \theta \right) \right]^2 \right)^{1/2} \end{aligned} \quad (46)$$

and show that

$$\sup_{\theta \in K} \frac{1}{n} \sum_{t=1}^n \left(\frac{y_{t-1}}{\sqrt{n}} \right)^2 \left[\pi \left(\frac{y_t}{\sqrt{n}}, \theta \right) - \pi \left(\frac{y_{t-1}}{\sqrt{n}}, \theta \right) \right]^2 \rightarrow_p 0 \quad (47)$$

as $n \rightarrow \infty$, for any compact subset K of \mathbb{R}^m .

We now set out to prove (47). As in the proof of Lemma A1, we may assume without loss of generality that $\pi \geq 0$. For $\theta_0 \in \mathbb{R}^m$ chosen arbitrarily, we let $\underline{\pi}_\delta$ and $\overline{\pi}_\delta$ be the regular functions such that $\underline{\pi}_\delta(x) \leq \pi(x, \theta) \leq \overline{\pi}_\delta(x)$ for all $\|\theta - \theta_0\| < \delta$, and

$$\overline{\pi}_\delta(x) - \underline{\pi}_\delta(x) \rightarrow 0 \quad (48)$$

as $\delta \rightarrow 0$ a.e. $x \in \mathbb{R}$. The existence of such functions is warranted by the regularity of π . Then we have

$$\sup_{\|\theta - \theta_0\| < \delta} \left| \pi \left(\frac{y_t}{\sqrt{n}}, \theta \right) - \pi \left(\frac{y_{t-1}}{\sqrt{n}}, \theta \right) \right| \leq \overline{\pi}_\delta \left(\frac{y_t}{\sqrt{n}} \right) - \underline{\pi}_\delta \left(\frac{y_{t-1}}{\sqrt{n}} \right)$$

and it follows that

$$\begin{aligned} & \sup_{\|\theta - \theta_0\| < \delta} \frac{1}{n} \sum_{t=1}^n \left(\frac{y_{t-1}}{\sqrt{n}} \right)^2 \left[\pi \left(\frac{y_t}{\sqrt{n}}, \theta \right) - \pi \left(\frac{y_{t-1}}{\sqrt{n}}, \theta \right) \right]^2 \\ & \leq \frac{1}{n} \sum_{t=1}^n \left(\frac{y_{t-1}}{\sqrt{n}} \right)^2 \left[\overline{\pi}_\delta \left(\frac{y_t}{\sqrt{n}} \right) - \underline{\pi}_\delta \left(\frac{y_{t-1}}{\sqrt{n}} \right) \right]^2 \rightarrow_d \int_0^1 V(r)^2 [\overline{\pi}_\delta(V(r)) - \underline{\pi}_\delta(V(r))]^2 dr \end{aligned}$$

as $n \rightarrow \infty$. However, we may deduce from the occupation times formula (38) and (48) that

$$\int_0^1 V(r)^2 [\overline{\pi}_\delta(V(r)) - \underline{\pi}_\delta(V(r))]^2 dr = \int_{\mathbb{R}} x^2 [\overline{\pi}_\delta(x) - \underline{\pi}_\delta(x)]^2 \ell(1, x) dx \rightarrow_{a.s.} 0$$

as $\delta \rightarrow 0$. The proof for (47) is therefore complete. \square

Lemma A4 Let Assumptions 2.1, 2.3, 4.1 and 4.4 hold, and let (v_t) be a time series satisfying $(1/n) \sum_{t=1}^n v_t^2 = O_p(1)$ for all $n \in \mathbb{N}$. Then, under the null hypothesis (4), we have

$$\sup_{\theta \in K} \left| \frac{1}{\sqrt{n}} \sum_{t=1}^n v_t \left[\frac{y_t}{\sqrt{n}} \pi \left(\frac{y_t}{\sqrt{n}}, \theta \right) - \frac{y_{t-1}}{\sqrt{n}} \pi \left(\frac{y_{t-1}}{\sqrt{n}}, \theta \right) \right] \right| = O_p(1)$$

as $n \rightarrow \infty$, for any compact subset K of \mathbb{R}^m .

Proof of Lemma A4 Similarly as in the proof of Lemma A3, it suffices to show that

$$\sup_{\theta \in K} \left| \frac{1}{\sqrt{n}} \sum_{t=1}^n v_t \frac{u_t}{\sqrt{n}} \pi \left(\frac{y_t}{\sqrt{n}}, \theta \right) \right| = O_p(1) \quad (49)$$

$$\sup_{\theta \in K} \left| \frac{1}{\sqrt{n}} \sum_{t=1}^n v_t \frac{y_{t-1}}{\sqrt{n}} \left[\pi \left(\frac{y_t}{\sqrt{n}}, \theta \right) - \pi \left(\frac{y_{t-1}}{\sqrt{n}}, \theta \right) \right] \right| = O_p(1) \quad (50)$$

as $n \rightarrow \infty$, for any compact subset K of \mathbb{R}^m . It is straightforward to establish (49), since it follows from Cauchy-Schwarz inequality and the boundedness of π that

$$\left| \frac{1}{\sqrt{n}} \sum_{t=1}^n v_t \frac{u_t}{\sqrt{n}} \pi \left(\frac{y_t}{\sqrt{n}}, \theta \right) \right| \leq c \left(\frac{1}{n} \sum_{t=1}^n v_t^2 \right) \left(\frac{1}{n} \sum_{t=1}^n u_t^2 \right)$$

for all $\theta \in \mathbb{R}^m$, with some constant $c > 0$.

To prove (50), we use Cauchy-Schwarz inequality to obtain

$$\begin{aligned} & \left| \frac{1}{\sqrt{n}} \sum_{t=1}^n v_t \frac{y_{t-1}}{\sqrt{n}} \left[\pi \left(\frac{y_t}{\sqrt{n}}, \theta \right) - \pi \left(\frac{y_{t-1}}{\sqrt{n}}, \theta \right) \right] \right| \\ & \leq \left(\frac{1}{n} \sum_{t=1}^n v_t^2 \right)^{1/2} \left(\frac{1}{n} \sum_{t=1}^n \left(\frac{y_{t-1}}{\sqrt{n}} \right)^2 n \left[\pi \left(\frac{y_t}{\sqrt{n}}, \theta \right) - \pi \left(\frac{y_{t-1}}{\sqrt{n}}, \theta \right) \right]^2 \right)^{1/2}, \end{aligned} \quad (51)$$

where

$$\frac{1}{n} \sum_{t=1}^n v_t^2, \frac{1}{n} \sum_{t=1}^n \left(\frac{y_{t-1}}{\sqrt{n}} \right)^2 = O_p(1).$$

The result in (50) can now be easily deduced from (51), since we have from Assumption 4.5

$$\sup_{\theta \in K} \left[\pi \left(\frac{y_t}{\sqrt{n}}, \theta \right) - \pi \left(\frac{y_{t-1}}{\sqrt{n}}, \theta \right) \right] \leq c \frac{|u_t|}{\sqrt{n}}$$

for some $c > 0$ and any compact subset K of \mathbb{R}^m . The proof is therefore complete. \square

Proofs of Theorems and Lemmas

Proof of Lemma 2.1 We write

$$\varepsilon_t^2 - \sigma^2 = [\varepsilon_t^2 - \mathbf{E}(\varepsilon_t^2 | \mathcal{F}_{t-1})] + [\mathbf{E}(\varepsilon_t^2 | \mathcal{F}_{t-1}) - \sigma^2].$$

As in the proof of Lemma 4.4 in Hall and Heyde (1980), we may show that

$$\sup_{1 \leq i \leq n} \frac{1}{n} \sum_{t=1}^i [\varepsilon_t^2 - \mathbf{E}(\varepsilon_t^2 | \mathcal{F}_{t-1})] \rightarrow_p 0$$

under condition (b) of Assumption 2.3. Therefore, it follows from condition (c) of Assumption 2.3 that

$$\sup_{1 \leq i \leq n} \frac{1}{n} \sum_{t=1}^i (\varepsilon_t^2 - \sigma^2) \rightarrow_p 0.$$

The stated result is now immediate from condition (a) of Assumption 2.3. \square

Proof of Lemma 2.2 The stated result follows directly from condition (c) of Assumption 2.3 and Lemma 4.4 of Hall and Heyde (1980, p106). \square

Proof of Theorem 2.3 We prove the convergence of finite dimensional distributions and the tightness. The former is trivial. We will establish the latter below. For the required tightness, it suffices to show that

$$\mathbb{E}[M_n(r_1, \theta_1) - M_n(r_2, \theta_2)]^4 \leq c [(r_1 - r_2)^2 + \|\theta_1 - \theta_2\|^2] \quad (52)$$

for all $(r_1, \theta_1), (r_2, \theta_2) \in [0, 1] \times \mathbb{R}^m$. This is due to Kolmogorov's criterion in, e.g., Revuz and Yor (1994, p489).

To prove (52), we first let (r_1, θ_1) and (r_2, θ_2) be chosen arbitrarily in $[0, 1] \times \mathbb{R}^m$ and define

$$C_n = M_n(r_1, \theta_1) - M_n(r_2, \theta_2).$$

Then it follows that

$$C_n = A_n + B_n,$$

where

$$A_n = \int_{r_2}^{r_1} \pi(V_n(s), \theta_1) dU(s)$$

$$B_n = \int_0^{r_2} [\pi(V_n(s), \theta_1) - \pi(V_n(s), \theta_2)] dU(s)$$

as one may easily check.

We have

$$\mathbb{E}C_n^4 \leq 8 (\mathbb{E}A_n^4 + \mathbb{E}B_n^4) \quad (53)$$

Moreover, it follows from the Burkholder-Davis-Gundy inequality [see, e.g., Revuz and Yor (1994, p153)]

$$\mathbb{E}A_n^4 \leq c \left(\mathbb{E} \int_{r_2}^{r_1} \pi^2(V_n(s), \theta_1) ds \right)^2 \quad (54)$$

and

$$\mathbb{E}B_n^4 \leq c \left(\mathbb{E} \int_0^{r_2} [\pi(V_n(s), \theta_1) - \pi(V_n(s), \theta_2)]^2 ds \right)^2 \quad (55)$$

for some absolute constant $c > 0$.

Due to (53) - (55), we now only need to show that

$$\mathbb{E} \int_{r_2}^{r_1} \pi^2(V_n(s), \theta_1) ds \leq c |r_1 - r_2| \quad (56)$$

$$\mathbb{E} \int_0^{r_2} [\pi(V_n(s), \theta_1) - \pi(V_n(s), \theta_2)]^2 ds \leq c \|\theta_1 - \theta_2\| \quad (57)$$

for some constant $c > 0$ to establish (52). It is straightforward to deduce (56), since π is bounded on $[0, 1] \times \mathbb{R}^m$. To establish (57), we first note that

$$\int_0^{r_2} [\pi(V_n(s), \theta_1) - \pi(V_n(s), \theta_2)]^2 ds$$

is bounded and

$$\int_0^{r_2} [\pi(V_n(s), \theta_1) - \pi(V_n(s), \theta_2)]^2 ds \rightarrow_d \int_0^{r_2} [\pi(V(s), \theta_1) - \pi(V(s), \theta_2)]^2 ds,$$

from which we have

$$\mathbb{E} \int_0^{r_2} [\pi(V_n(s), \theta_1) - \pi(V_n(s), \theta_2)]^2 ds \rightarrow \mathbb{E} \int_0^{r_2} [\pi(V(s), \theta_1) - \pi(V(s), \theta_2)]^2 ds. \quad (58)$$

Then we apply the occupation times formula introduced in (38) and the Fubini's theorem to deduce that

$$\begin{aligned} \mathbb{E} \int_0^{r_2} [\pi(V(s), \theta_1) - \pi(V(s), \theta_2)]^2 ds &= \mathbb{E} \int_{\mathbb{R}} [\pi(x, \theta_1) - \pi(x, \theta_2)]^2 \ell(r_2, x) dx \\ &= \int_{\mathbb{R}} [\pi(x, \theta_1) - \pi(x, \theta_2)]^2 \mathbb{E} \ell(r_2, x) dx \end{aligned} \quad (59)$$

However, it follows as shown in, e.g., Park (2004) that

$$\mathbb{E} \ell(r, x) = 2[\phi(x/\omega) - |x/\omega| \Phi(-|x/\omega|)],$$

where ϕ and Φ are respectively the standard normal density function and distribution function. We may now readily obtain (57) from (58), (59) and condition (d) of Assumption 2.2. The proof is therefore complete. \square

Proof of Proposition 2.4 Due to the Kolmogorov's criterion [see, e.g., Revuz and Yor (1994, p.18)], it suffices to show that

$$\mathbb{E}[M(r_1, \theta_1) - M(r_2, \theta_2)]^4 \leq c [(r_1 - r_2)^2 + \|\theta_1 - \theta_2\|^2] \quad (60)$$

for all $(r_1, \theta_1), (r_2, \theta_2) \in [0, 1] \times \mathbb{R}^m$. The proof of (60) is entirely analogous to that of (52) in the proof of Theorem 2.3. We let (r_1, θ_1) and (r_2, θ_2) be chosen arbitrarily in $[0, 1] \times \mathbb{R}^m$, and write $C = A + B$, where $C = M(r_1, \theta_1) - M(r_2, \theta_2)$, $A = \int_{r_2}^{r_1} \pi(V(s), \theta_1) dU(s)$ and $B = \int_0^{r_2} [\pi(V(s), \theta_1) - \pi(V(s), \theta_2)] dU(s)$. As in (52), we have $\mathbb{E}C^4 \leq 8(\mathbb{E}A^4 + \mathbb{E}B^4)$. Furthermore, due to the Burkholder-Davis-Gundy inequality, we have

$$\begin{aligned} \mathbb{E}A^4 &\leq c \left(\mathbb{E} \int_{r_2}^{r_1} \pi^2(V(s), \theta_1) ds \right)^2 \\ \mathbb{E}B^4 &\leq c \left(\mathbb{E} \int_0^{r_2} [\pi(V(s), \theta_1) - \pi(V(s), \theta_2)]^2 ds \right)^2 \end{aligned}$$

exactly as in (53) and (54), for some absolute constant $c > 0$. Now (60) follows immediately from the boundedness of π and (59), and the proof is complete. \square

Proof of Lemma 3.1 We may assume $d = 1$, without loss of generality. This is because

$$\begin{aligned} \frac{1}{n} \sum_{t=1}^n u_{t-i} \frac{y_{t-1}}{\sqrt{n}} \pi \left(\frac{y_{t-d}}{\sqrt{n}}, \theta \right) &= \frac{1}{n} \sum_{t=1}^n u_{t-i} \frac{y_{t-d}}{\sqrt{n}} \pi \left(\frac{y_{t-d}}{\sqrt{n}}, \theta \right) \\ &\quad + \frac{1}{n} \sum_{t=1}^n u_{t-i} \left(\frac{y_{t-1}}{\sqrt{n}} - \frac{y_{t-d}}{\sqrt{n}} \right) \pi \left(\frac{y_{t-d}}{\sqrt{n}}, \theta \right), \end{aligned}$$

and, due to Cauchy-Schwarz inequality and the boundedness of π , we have

$$\begin{aligned} &\left| \frac{1}{n} \sum_{t=1}^n u_{t-i} \left(\frac{y_{t-1}}{\sqrt{n}} - \frac{y_{t-d}}{\sqrt{n}} \right) \pi \left(\frac{y_{t-d}}{\sqrt{n}}, \theta \right) \right| \\ &\leq \frac{c}{\sqrt{n}} \left(\frac{1}{n} \sum_{t=1}^n u_{t-i}^2 \right)^{1/2} \left(\frac{1}{n} \sum_{t=1}^n (y_{t-1} - y_{t-d})^2 \right)^{1/2} \end{aligned}$$

uniformly in $\theta \in \mathbb{R}^m$, for some constant $c > 0$.

We will first look at the case $i = 1$. Write

$$\begin{aligned} \frac{1}{n} \sum_{t=1}^n u_{t-1} \frac{y_{t-1}}{\sqrt{n}} \pi \left(\frac{y_{t-1}}{\sqrt{n}}, \theta \right) &= \frac{1}{n} \sum_{t=1}^n u_{t-1} \frac{y_{t-2}}{\sqrt{n}} \pi \left(\frac{y_{t-2}}{\sqrt{n}}, \theta \right) \\ &\quad + \frac{1}{n} \sum_{t=1}^n u_{t-1} \left[\frac{y_{t-1}}{\sqrt{n}} \pi \left(\frac{y_{t-1}}{\sqrt{n}}, \theta \right) - \frac{y_{t-2}}{\sqrt{n}} \pi \left(\frac{y_{t-2}}{\sqrt{n}}, \theta \right) \right], \end{aligned}$$

and note that

$$\frac{1}{n} \sum_{t=1}^n u_{t-1} \left[\frac{y_{t-1}}{\sqrt{n}} \pi \left(\frac{y_{t-1}}{\sqrt{n}}, \theta \right) - \frac{y_{t-2}}{\sqrt{n}} \pi \left(\frac{y_{t-2}}{\sqrt{n}}, \theta \right) \right] = o_p(1)$$

uniformly in $\theta \in K$, for any compact subset K of \mathbb{R}^m , which follows from Lemma A3. We therefore have

$$\frac{1}{n} \sum_{t=1}^n u_{t-1} \frac{y_{t-1}}{\sqrt{n}} \pi \left(\frac{y_{t-1}}{\sqrt{n}}, \theta \right) = \frac{1}{n} \sum_{t=1}^n u_{t-1} \frac{y_{t-2}}{\sqrt{n}} \pi \left(\frac{y_{t-2}}{\sqrt{n}}, \theta \right) + o_p(1) \quad (61)$$

uniformly in $\theta \in K$.

Due to (61), it now suffices to show that

$$\frac{1}{n} \sum_{t=1}^n u_{t-1} \frac{y_{t-2}}{\sqrt{n}} \pi \left(\frac{y_{t-2}}{\sqrt{n}}, \theta \right) = o_p(1) \quad (62)$$

uniformly in $\theta \in K$. To show (62), we write

$$\begin{aligned} \frac{1}{n} \sum_{t=1}^n u_{t-1} \frac{y_{t-2}}{\sqrt{n}} \pi \left(\frac{y_{t-2}}{\sqrt{n}}, \theta \right) &= \frac{1}{n\alpha(1)} \sum_{t=1}^n \varepsilon_{t-1} \frac{y_{t-2}}{\sqrt{n}} \pi \left(\frac{y_{t-2}}{\sqrt{n}}, \theta \right) \\ &\quad + \frac{1}{n} \sum_{t=1}^n (\tilde{u}_{t-2} - \tilde{u}_{t-1}) \frac{y_{t-2}}{\sqrt{n}} \pi \left(\frac{y_{t-2}}{\sqrt{n}}, \theta \right), \end{aligned} \quad (63)$$

where

$$\tilde{u}_t = \frac{1}{\alpha(1)} \sum_{i=1}^p \left(\sum_{j=i}^p \alpha_j \right) u_{t-i+1}.$$

Note in particular that we have $(1/n) \sum_{t=1}^n \tilde{u}_t^2 = O_p(1)$ for the process (\tilde{u}_t) .

As shown in Theorem 2.3, we have

$$\frac{1}{\sqrt{n}} \sum_{t=1}^n \varepsilon_{t-1} \frac{y_{t-2}}{\sqrt{n}} \pi \left(\frac{y_{t-2}}{\sqrt{n}}, \theta \right) = O_p(1) \quad (64)$$

uniformly in $\theta \in K$. Moreover, it follows from summation by parts that

$$\begin{aligned} &\frac{1}{n} \sum_{t=1}^n (\tilde{u}_{t-2} - \tilde{u}_{t-1}) \frac{y_{t-2}}{\sqrt{n}} \pi \left(\frac{y_{t-2}}{\sqrt{n}}, \theta \right) \\ &= \frac{1}{n} \sum_{t=1}^n \tilde{u}_{t-1} \left[\frac{y_{t-1}}{\sqrt{n}} \pi \left(\frac{y_{t-1}}{\sqrt{n}}, \theta \right) - \frac{y_{t-2}}{\sqrt{n}} \pi \left(\frac{y_{t-2}}{\sqrt{n}}, \theta \right) \right] + o_p(1) = o_p(1) \end{aligned} \quad (65)$$

uniformly in $\theta \in K$, due once again to Lemma A3. We may thus readily deduce (62) from (63) - (65).

For the case $i = 2$, we write

$$\begin{aligned} \frac{1}{n} \sum_{t=1}^n u_{t-2} \frac{y_{t-1}}{\sqrt{n}} \pi \left(\frac{y_{t-1}}{\sqrt{n}}, \theta \right) &= \frac{1}{n} \sum_{t=1}^n u_{t-2} \frac{y_{t-2}}{\sqrt{n}} \pi \left(\frac{y_{t-2}}{\sqrt{n}}, \theta \right) \\ &\quad + \frac{1}{n} \sum_{t=1}^n u_{t-2} \left[\frac{y_{t-1}}{\sqrt{n}} \pi \left(\frac{y_{t-1}}{\sqrt{n}}, \theta \right) - \frac{y_{t-2}}{\sqrt{n}} \pi \left(\frac{y_{t-2}}{\sqrt{n}}, \theta \right) \right], \end{aligned}$$

and deduce from Lemma A3 that

$$\frac{1}{n} \sum_{t=1}^n u_{t-2} \frac{y_{t-1}}{\sqrt{n}} \pi \left(\frac{y_{t-1}}{\sqrt{n}}, \theta \right) = \frac{1}{n} \sum_{t=1}^n u_{t-2} \frac{y_{t-2}}{\sqrt{n}} \pi \left(\frac{y_{t-2}}{\sqrt{n}}, \theta \right) + o_p(1), \quad (66)$$

which holds uniformly in $\theta \in K$. We may therefore establish (26) using the proof for the case $i = 1$. Obviously, our result in (66) easily extends to the case with general i , and the proof is therefore complete. \square

Proof of Theorem 3.2 As in the proof of Lemma 3.1, we let $d = 1$ without loss of generality. Let

$$w_{nt}(\theta) = \frac{y_{t-1}}{\sqrt{n}} \pi \left(\frac{y_{t-1}}{\sqrt{n}}, \theta \right).$$

Under the null hypothesis (4), A_n and B_n in (19) and (20) reduce to

$$\begin{aligned} \frac{A_n(\theta)}{n} &= \frac{1}{\sqrt{n}} \sum_{t=1}^n w_{nt}(\theta) \varepsilon_t - \left(\frac{1}{n} \sum_{t=1}^n w_{nt}(\theta) x_t' \right) \left(\frac{1}{n} \sum_{t=1}^n x_t x_t' \right)^{-1} \left(\frac{1}{\sqrt{n}} \sum_{t=1}^n x_t \varepsilon_t \right) \\ \frac{B_n(\theta)}{n^2} &= \frac{1}{n} \sum_{t=1}^n w_{nt}^2(\theta) - \left(\frac{1}{n} \sum_{t=1}^n w_{nt}(\theta) x_t' \right) \left(\frac{1}{n} \sum_{t=1}^n x_t x_t' \right)^{-1} \left(\frac{1}{n} \sum_{t=1}^n x_t w_{nt}(\theta) \right). \end{aligned}$$

Furthermore, we have

$$\begin{aligned} \hat{\sigma}_n^2(\theta) &= \frac{1}{n} \sum_{t=1}^n \varepsilon_t^2 - \frac{1}{n} \left(\frac{1}{\sqrt{n}} \sum_{t=1}^n \varepsilon_t x_t' \right) \left(\frac{1}{n} \sum_{t=1}^n x_t x_t' \right)^{-1} \left(\frac{1}{\sqrt{n}} \sum_{t=1}^n x_t \varepsilon_t \right) \\ &\quad - \frac{1}{n} \left[\frac{1}{\sqrt{n}} \sum_{t=1}^n w_{nt}(\theta) \varepsilon_t - \left(\frac{1}{\sqrt{n}} \sum_{t=1}^n \varepsilon_t x_t' \right) \left(\frac{1}{n} \sum_{t=1}^n x_t x_t' \right)^{-1} \left(\frac{1}{n} \sum_{t=1}^n x_t w_{nt}(\theta) \right) \right]^2 \\ &\quad \cdot \left[\frac{1}{n} \sum_{t=1}^n w_{nt}^2(\theta) - \left(\frac{1}{n} \sum_{t=1}^n w_{nt}(\theta) x_t' \right) \left(\frac{1}{n} \sum_{t=1}^n x_t x_t' \right)^{-1} \left(\frac{1}{n} \sum_{t=1}^n x_t w_{nt}(\theta) \right) \right]^{-1} \end{aligned}$$

under the null hypothesis (4).

Note that

$$\left(\frac{1}{n} \sum_{t=1}^n x_t x_t' \right)^{-1}, \quad \frac{1}{\sqrt{n}} \sum_{t=1}^n x_t \varepsilon_t = O_p(1). \quad (67)$$

Moreover, we have

$$\frac{1}{\sqrt{n}} \sum_{t=1}^n w_{nt}(\theta) \varepsilon_t = O_p(1) \quad \text{and} \quad \frac{1}{n} \sum_{t=1}^n x_t w_{nt}(\theta) = o_p(1) \quad (68)$$

uniformly in $\theta \in \Theta$, due respectively to Theorem 2.3 and Lemma 3.1. Finally, as shown in Park and Phillips (2001), we have as $n \rightarrow \infty$

$$\frac{1}{n} \sum_{t=1}^n w_{nt}^2(\theta) =_d \int_0^1 V_n(r)^2 \pi^2(V_n(r), \theta) dr + o_p(1) \rightarrow_{a.s.} \int_0^1 V(r)^2 \pi^2(V(r), \theta) dr \quad (69)$$

uniformly in $\theta \in \Theta$.

Under the null hypothesis (4), it follows from (67) - (69) that

$$\frac{A_n(\theta)}{n} = \frac{1}{\sqrt{n}} \sum_{t=1}^n w_{nt}(\theta) \varepsilon_t + o_p(1) \quad (70)$$

$$\frac{B_n(\theta)}{n^2} = \frac{1}{n} \sum_{t=1}^n w_{nt}^2(\theta) + o_p(1) \quad (71)$$

uniformly in $\theta \in \Theta$, and that

$$\hat{\sigma}_n^2 = \frac{1}{n} \sum_{t=1}^n \varepsilon_t^2 + o_p(1) \quad (72)$$

uniformly in $\theta \in \Theta$. However, we have from Theorem 2.3, Proposition 2.4 and the CMT that

$$\frac{1}{\sqrt{n}} \sum_{t=1}^n w_{nt}(\theta) \varepsilon_t =_d \int_0^1 V_n(r) M_n(dr, \theta) + o_p(1) \rightarrow_d \int_0^1 V(r) M(dr, \theta) \quad (73)$$

as $n \rightarrow \infty$. We may now easily deduce the stated result from (70), (71), (72) and Assumption 2.1, using (69), (73) and Lemma 2.1. \square

Proof of Corollary 3.3 Obvious and omitted. \square

Proof of Proposition 3.4 Let $w_t(\theta)$ be defined as in (16). For the sequence $\theta_n \in \Theta$, we have $\text{plim}_{n \rightarrow \infty} \hat{\lambda}_n(\theta_n) < 0$, and

$$\text{plim}_{n \rightarrow \infty} \frac{B_n(\theta_n)}{n} > 0,$$

due to our assumption that the conditional variance of $(y_{t-1} \pi(y_{t-d}, \theta_n))$ given $(\Delta y_{t-1}, \dots, \Delta y_{t-p})$ is nonzero. Therefore, we have

$$T_n \leq T_n(\theta_n) = \sqrt{n} \frac{\hat{\lambda}_n(\theta_n)}{\hat{\sigma}_n(\theta_n)} \sqrt{\frac{B_n(\theta_n)}{n}} \rightarrow_p -\infty$$

as $n \rightarrow \infty$. This was to be shown. \square

Proof of Theorem 4.1 The proof is quite similar to that of Theorem 3.2. For instance, if we let

$$w_{nt}(\rho, \nu, \theta) = \sum_{i=1}^{\ell} \rho_i \left(\frac{y_{t-1}}{\sqrt{n}} - \nu_i \right) \pi_i \left(\frac{y_{t-1}}{\sqrt{n}}, \theta_i \right),$$

then we may show under the null hypothesis of a unit root that

$$\begin{aligned} \frac{A_n(\rho, \nu, \theta)}{n} &= \frac{1}{\sqrt{n}} \sum_{t=1}^n w_{nt}(\rho, \nu, \theta) \varepsilon_t + o_p(1) \\ \frac{B_n(\rho, \nu, \theta)}{n^2} &= \frac{1}{n} \sum_{t=1}^n w_{nt}^2(\rho, \nu, \theta) + o_p(1) \end{aligned}$$

uniformly in $(\rho, \nu, \theta) \in \Phi \times \Lambda \times \Theta$. Moreover, it can also be readily deduced that

$$\begin{aligned} \frac{1}{\sqrt{n}} \sum_{t=1}^n w_{nt}(\rho, \nu, \theta) \varepsilon_t &= \int_0^1 \left[\sum_{i=1}^{\ell} \rho_i (V_n - \nu_i) M_{ni}(dr, \theta_i) \right] + o_p(1) \\ &\rightarrow_d \int_0^1 \left[\sum_{i=1}^{\ell} \rho_i (V - \nu_i) M_i(dr, \theta_i) \right] \end{aligned}$$

where M_{in} and M_i are defined exactly as M_n and M , respectively, with π_i for $i = 1, \dots, \ell$, and

$$\begin{aligned} \frac{1}{n} \sum_{t=1}^n w_{nt}^2(\rho, \nu, \theta) &= \int_0^1 \left[\sum_{i=1}^{\ell} \rho_i (V_n(r) - \nu_i) \pi_i(V_n(r), \theta_i) \right]^2 dr + o_p(1) \\ &\rightarrow_d \int_0^1 \left[\sum_{i=1}^{\ell} \rho_i (V(r) - \nu_i) \pi_i(V(r), \theta_i) \right]^2 dr \end{aligned}$$

as $n \rightarrow \infty$. The further details of the proof are not provided, since they are essentially the same as those of Theorem 3.2. \square

Proof of Corollary 4.2 Obvious and omitted. \square

Proof of Lemma 4.3 The result can be obtained by using the similar argument used in the proof of Lemma 3.1 and by using the result from Lemma A4 instead of Lemma A3. \square

Proof of Theorem 4.4 Here we consider the simpler model in (3) for expositional convenience. The extension for the model in (29) is straightforward, and can be done as shown in the proof of Theorem 4.1. Let us again redefine A_n and B_n in (19) and (20) by

$$\begin{aligned} A_n(\theta) &= \sum_{t=1}^n y_{t-1} \pi_t(\theta) \varepsilon_{p,t} - \left(\sum_{t=1}^n y_{t-1} \pi_t(\theta) x'_{p,t} \right) \left(\sum_{t=1}^n x_{p,t} x'_{p,t} \right)^{-1} \left(\sum_{t=1}^n x_{p,t} \varepsilon_{p,t} \right) \\ B_n(\theta) &= \sum_{t=1}^n y_{t-1}^2 \pi_t(\theta)^2 - \left(\sum_{t=1}^n y_{t-1} \pi_t(\theta) x'_{p,t} \right) \left(\sum_{t=1}^n x_{p,t} x'_{p,t} \right)^{-1} \left(\sum_{t=1}^n x_{p,t} y_{t-1} \pi_t(\theta) \right), \end{aligned}$$

where

$$\pi_t(\theta) = \pi(y_{t-d}/\sqrt{n}, \theta) \quad \text{and} \quad x_{p,t} = (\Delta y_{t-1}, \dots, \Delta y_{t-p})'.$$

First, note that,

$$\frac{A_n(\theta)}{n} = \frac{1}{n} \sum_{t=1}^n y_{t-1} \pi_t(\theta) \varepsilon_{p,t} + o_p(1)$$

uniformly in $\theta \in \Theta$, since

$$\begin{aligned} & \left| \left(\sum_{t=1}^n y_{t-1} \pi_t(\theta) x'_{p,t} \right) \left(\sum_{t=1}^n x_{p,t} x'_{p,t} \right)^{-1} \left(\sum_{t=1}^n x_{p,t} \varepsilon_{p,t} \right) \right| \\ & \leq \left\| \sum_{t=1}^n x_{p,t} y_{t-1} \pi_t(\theta) \right\| \left\| \left(\sum_{t=1}^n x_{p,t} x'_{p,t} \right)^{-1} \right\| \left\| \sum_{t=1}^n x_{p,t} \varepsilon_{p,t} \right\| = o_p(n). \end{aligned}$$

uniformly in $\theta \in \Theta$. To show the last equality, $\left\| \left((1/n) \sum_{t=1}^n x_{p,t} x'_{p,t} \right)^{-1} \right\| = O_p(1)$ and $\left\| \sum_{t=1}^n x_{p,t} \varepsilon_{p,t} \right\| = o_p(np^{-1/2})$ directly follows from Lemma 3.2 (a) and (c), respectively, in Chang and Park (2002). $\left\| \sum_{t=1}^n x_{p,t} y_{t-1} \pi_t(\theta) \right\| = O_p(np^{1/2})$ uniformly in $\theta \in \Theta$ follows from Lemma 3.4.

In addition,

$$\sum_{t=1}^n y_{t-1} \pi_t(\theta) \varepsilon_{p,t} = \sum_{t=1}^n y_{t-1} \pi_t(\theta) \varepsilon_t + \sum_{t=1}^n y_{t-1} \pi_t(\theta) (\varepsilon_{p,t} - \varepsilon_t).$$

The second term can be decomposed as

$$\begin{aligned} & \sum_{t=1}^n y_{t-1} \pi_t(\theta) (\varepsilon_{p,t} - \varepsilon_t) \\ & = \varphi(1) \sum_{t=1}^n \pi_t(\theta) \sum_{k=1}^{t-1} \varepsilon_k (\varepsilon_{p,t} - \varepsilon_t) + \tilde{u}_0 \sum_{t=1}^n \pi_t(\theta) (\varepsilon_{p,t} - \varepsilon_t) - \sum_{t=1}^n \pi_t(\theta) \tilde{u}_{t-1} (\varepsilon_{p,t} - \varepsilon_t) \\ & = R_{1n}(\theta) + R_{2n}(\theta) + R_{3n}(\theta) = o_p(n) + o_p(n) + o_p(n) = o_p(n) \end{aligned}$$

uniformly in $\theta \in \Theta$. To show this, let us write

$$\varepsilon_{p,t} - \varepsilon_t = \sum_{k=p+1}^{\infty} \alpha_k u_{t-k} = \sum_{k=p+1}^{\infty} \varphi_{p,k} \varepsilon_{t-k}$$

where

$$\sum_{k=p+1}^{\infty} \varphi_{p,k}^2 \leq c \sum_{k=p+1}^{\infty} \alpha_k^2 = o(p^{-2s})$$

as in Berk (1974, Proof of Lemma 2, p492). Also, denote by δ_{ij} the usual Kronecker delta. To show that $R_{1n}(\theta) = o_p(n)$, we write

$$\begin{aligned}
& \sum_{t=1}^n \pi_t(\theta) \sum_{k=1}^{t-1} \varepsilon_k (\varepsilon_{p,t} - \varepsilon_t) \\
&= \sum_{t=1}^n \pi_t(\theta) \left(\sum_{i=1}^{t-1} \varepsilon_i \right) \sum_{j=p+1}^{\infty} \varphi_{p,j} \varepsilon_{t-j} \\
&= \sum_{j=p+1}^{\infty} \varphi_{p,j} \sum_{t=1}^n \pi_t(\theta) \sum_{i=1}^{t-1} \varepsilon_{t-i} \varepsilon_{t-j} \\
&= \sum_{j=p+1}^{\infty} \varphi_{p,j} \sum_{t=1}^n \pi_t(\theta) \varepsilon_{t-j}^2 + \sum_{j=p+1}^{\infty} \varphi_{p,j} \sum_{t=1}^n \pi_t(\theta) \sum_{i=1, i \neq j}^{t-1} \varepsilon_{t-i} \varepsilon_{t-j} \\
&= \sum_{j=p+1}^{\infty} \varphi_{p,j} \sum_{t=j}^n \pi_t(\theta) \sigma^2 + \sum_{j=p+1}^{\infty} \varphi_{p,j} \sum_{t=1}^n \pi_t(\theta) (\varepsilon_{t-j}^2 - \sigma^2) + \sum_{j=p+1}^{\infty} \varphi_{p,j} \sum_{t=1}^n \pi_t(\theta) \sum_{i=1, i \neq j}^{t-1} \varepsilon_{t-i} \varepsilon_{t-j} \\
&= \sigma^2 \sum_{j=p+1}^{\infty} \varphi_{p,j} \sum_{t=j}^n \pi_t(\theta) + \sum_{j=p+1}^{\infty} \varphi_{p,j} \sum_{t=1}^n \pi_t(\theta) \sum_{i=1}^{t-1} (\varepsilon_{t-i} \varepsilon_{t-j} - \sigma^2 \delta_{ij})
\end{aligned}$$

The first term is $o_p(np^{-s})$, since

$$\begin{aligned}
\left(\mathbb{E} \left[\sum_{j=p+1}^{\infty} \varphi_{p,j} \sum_{t=j}^n \pi_t(\theta) \right]^2 \right)^{1/2} &\leq \left(\mathbb{E} \left[\sum_{j=p+1}^{\infty} |\varphi_{p,j}| \sum_{t=j}^n |\pi_t(\theta)| \right]^2 \right)^{1/2} \\
&= nc \sum_{j=p+1}^{\infty} |\varphi_{p,j}| = o(np^{-s})
\end{aligned}$$

for some constant $c > 0$ where the last inequality follows from the boundedness of π . The second term is $o_p(np^{-s})$, since

$$\begin{aligned}
& \left(\mathbb{E} \left[\sum_{j=p+1}^{\infty} \varphi_{p,j} \sum_{t=1}^n \pi_t(\theta) \sum_{i=1}^{t-1} (\varepsilon_{t-i} \varepsilon_{t-j} - \sigma^2 \delta_{ij}) \right]^2 \right)^{1/2} \\
&\leq \sum_{j=p+1}^{\infty} |\varphi_{p,j}| \left(\mathbb{E} \left[\sum_{t=1}^n \pi_t(\theta) \sum_{i=1}^{t-1} (\varepsilon_{t-i} \varepsilon_{t-j} - \sigma^2 \delta_{ij}) \right]^2 \right)^{1/2} \\
&\leq c \sum_{j=p+1}^{\infty} |\varphi_{p,j}| \sum_{t=1}^n \left(\mathbb{E} \left[\sum_{i=1}^{t-1} (\varepsilon_{t-i} \varepsilon_{t-j} - \sigma^2 \delta_{ij}) \right]^2 \right)^{1/2} \\
&\leq c'n \left(\sum_{j=p+1}^{\infty} |\varphi_{p,j}| \right)
\end{aligned}$$

for some constant $c, c' > 0$ where the third inequality follows from the boundedness of π and the last inequality follows from the fourth moment condition in Assumption 2.3. It now follows immediately that $R_{1n}(\theta) = o_p(n)$.

The result $R_{2n}(\theta) = o_p(n)$ can be deduced from

$$\begin{aligned}
\left(\mathbb{E} \left[\sum_{t=1}^n \pi_t(\theta) (\varepsilon_{p,t} - \varepsilon_t) \right]^2 \right)^{1/2} &\leq \left(\mathbb{E} \left[\sum_{t=1}^n \pi_t(\theta)^2 \sum_{t=1}^n (\varepsilon_{p,t} - \varepsilon_t)^2 \right] \right)^{1/2} \\
&\leq \left(nc \mathbb{E} \left[\sum_{t=1}^n (\varepsilon_{p,t} - \varepsilon_t)^2 \right] \right)^{1/2} \\
&= \left(nc \mathbb{E} \left[\sum_{t=1}^n \left(\sum_{k=p+1}^{\infty} \varphi_{p,k} \varepsilon_{t-k} \right)^2 \right] \right)^{1/2} \\
&= \left(n^2 c \sigma^2 \sum_{k=p+1}^{\infty} \varphi_{p,k}^2 \right)^{1/2} = o(np^{-s})
\end{aligned}$$

for some constant $c > 0$ where the second inequality follows from the boundedness of π . Finally, to show that $R_{3n}(\theta) = o_p(n)$, we write

$$\begin{aligned}
\sum_{t=1}^n \pi_t(\theta) \tilde{u}_{t-1}(\varepsilon_{p,t} - \varepsilon_t) &= \sum_{t=1}^n \pi_t(\theta) \left(\sum_{i=0}^{\infty} \tilde{\varphi}_i \varepsilon_{t-i} \right) \left(\sum_{j=p+1}^{\infty} \varphi_{p,j} \varepsilon_{t-j} \right) \\
&= \sum_{i=0}^{\infty} \sum_{j=p+1}^{\infty} \tilde{\varphi}_i \varphi_{p,j} \sum_{t=1}^n \pi_t(\theta) \varepsilon_{t-i} \varepsilon_{t-j} \\
&= \sigma^2 \sum_{j=p+1}^{\infty} \tilde{\varphi}_{j-1} \varphi_{p,j} \sum_{t=1}^n \pi_t(\theta) \\
&\quad + \sum_{i=0}^{\infty} \sum_{j=p+1}^{\infty} \tilde{\varphi}_i \varphi_{p,j} \sum_{t=1}^n \pi_t(\theta) (\varepsilon_{t-i} \varepsilon_{t-j} - \sigma^2 \delta_{i,j}).
\end{aligned}$$

The first term is of order $o_p(np^{-s})$, since

$$\begin{aligned}
\left(\mathbb{E} \left[\sum_{j=p+1}^{\infty} \tilde{\varphi}_{j-1} \varphi_{p,j} \sum_{t=1}^n \pi_t(\theta) \right]^2 \right)^{1/2} &\leq \left(\mathbb{E} \left[\sum_{j=p+1}^{\infty} |\tilde{\varphi}_{j-1} \varphi_{p,j}| \sum_{t=j}^n |\pi_t(\theta)| \right]^2 \right)^{1/2} \\
&\leq nc \sum_{j=p+1}^{\infty} |\tilde{\varphi}_{j-1} \varphi_{p,j}| \\
&\leq nc \left(\sum_{k=p+1}^{\infty} \tilde{\varphi}_{k-1}^2 \right)^{1/2} \left(\sum_{k=p+1}^{\infty} \varphi_{p,k}^2 \right)^{1/2} = o(np^{-s})
\end{aligned}$$

for some constant $c > 0$ where the second inequality follows from the boundedness of π and the last equality follows from $\sum_{k=p+1}^{\infty} \varphi_{p,k}^2 = o(p^{-2s})$ and

$$\sum_{k=p+1}^{\infty} \tilde{\varphi}_{k-1}^2 \leq \sum_{k=1}^{\infty} \tilde{\varphi}_{k-1}^2 < \infty.$$

The second term is $o_p(np^{-s})$, since

$$\begin{aligned} & \left(\mathbb{E} \left[\sum_{i=0}^{\infty} \sum_{j=p+1}^{\infty} \tilde{\varphi}_i \varphi_{p,j} \sum_{t=1}^n \pi_t(\theta) (\varepsilon_{t-i} \varepsilon_{t-j} - \sigma^2 \delta_{ij}) \right]^2 \right)^{1/2} \\ & \leq \sum_{i=0}^{\infty} \sum_{j=p+1}^{\infty} |\tilde{\varphi}_i \varphi_{p,j}| \left(E \left[\sum_{t=1}^n \pi_t(\theta) (\varepsilon_{t-i} \varepsilon_{t-j} - \sigma^2 \delta_{ij}) \right]^2 \right)^{1/2} \\ & \leq c \sum_{i=0}^{\infty} \sum_{j=p+1}^{\infty} |\tilde{\varphi}_i \varphi_{p,j}| \sum_{t=1}^n \left(E [(\varepsilon_{t-i} \varepsilon_{t-j} - \sigma^2 \delta_{ij})^2] \right)^{1/2} \\ & \leq c'n \left(\sum_{i=0}^{\infty} \sum_{j=p+1}^{\infty} |\tilde{\varphi}_i \varphi_{p,j}| \right) \leq c'n \left(\sum_{k=p+1}^{\infty} \tilde{\varphi}_{k-1}^2 \right)^{1/2} \left(\sum_{k=p+1}^{\infty} \varphi_{p,k}^2 \right)^{1/2} = o(np^{-s}) \end{aligned}$$

for some constant $c, c' > 0$ where the second inequality follows from the boundedness of π and the third inequality follows from the fourth moment condition in Assumption 2.3.

Combining these results yields

$$\frac{A_n(\theta)}{n} = \frac{1}{\sqrt{n}} \sum_{t=1}^n \frac{y_{t-1}}{\sqrt{n}} \pi_t(\theta) \varepsilon_t + o_p(1)$$

uniformly in $\theta \in \Theta$.

Using the similar argument, we have

$$\frac{B_n(\theta)}{n^2} = \frac{1}{n^2} \sum_{t=1}^n y_{t-1}^2 \pi_t(\theta)^2 + o_p(1),$$

since

$$\begin{aligned} & \left| \left(\sum_{t=1}^n y_{t-1} \pi_t(\theta) x'_{p,t} \right) \left(\sum_{t=1}^n x_{p,t} x'_{p,t} \right)^{-1} \left(\sum_{t=1}^n x_{p,t} y_{t-1} \pi_t(\theta) \right) \right| \\ & \leq \left\| \sum_{t=1}^n x_{p,t} y_{t-1} \pi_t(\theta) \right\| \left\| \left(\sum_{t=1}^n x_{p,t} x'_{p,t} \right)^{-1} \right\| \left\| \sum_{t=1}^n x_{p,t} y_{t-1} \pi_t(\theta) \right\| \\ & = O_p(np^{1/2}) O_p(n^{-1}) O_p(np^{1/2}) = o_p(np) = o_p(n^2) \end{aligned}$$

uniformly in $\theta \in \Theta$.

Finally,

$$\begin{aligned}
\hat{\sigma}_n^2(\theta) &= \frac{1}{n} \sum_{t=1}^n \varepsilon_{p,t}^2 - \frac{1}{n} \left(\frac{1}{\sqrt{n}} \sum_{t=1}^n \varepsilon_{p,t} x'_{p,t} \right) \left(\frac{1}{n} \sum_{t=1}^n x_{p,t} x'_{p,t} \right)^{-1} \left(\frac{1}{\sqrt{n}} \sum_{t=1}^n x_{p,t} \varepsilon_{p,t} \right) \\
&\quad - \frac{1}{n} (A_n(\theta)^2 B_n(\theta)^{-1}) \\
&= \frac{1}{n} \sum_{t=1}^n \varepsilon_{p,t}^2 + o_p(p^{-1}) + O_p(1/n) \\
&= \frac{1}{n} \sum_{t=1}^n \varepsilon_t^2 + o_p(1) \\
&= \sigma^2 + o_p(1)
\end{aligned}$$

uniformly in $\theta \in \Theta$ where the second equality follows from $\left\| \left((1/n) \sum_{t=1}^n x_{p,t} x'_{p,t} \right)^{-1} \right\| = O_p(1)$, $\left\| \sum_{t=1}^n x_{p,t} \varepsilon_{p,t} \right\| = o_p(np^{-1/2})$ and $A_n(\theta)^2 B_n(\theta)^{-1} = O_p(1)$ uniformly in $\theta \in \Theta$, the third equality follows from Lemma 3.1 (c) of Chang and Park (2002), and the last equality follows from Lemma 2.1. The desired results follows from combining the results for $A_n(\theta)$, $B_n(\theta)$ and $\hat{\sigma}_n^2(\theta)$ and using the limiting distribution in the proof of Theorem 3.2. \square

References

- [1] Balke, N. and T. Fomby (1997) "Threshold Cointegration," *International Economic Review* 38, 627-645.
- [2] Bec, F., M. Ben Salem and M. Carrasco (2004a) "Tests for Unit-Root Versus Threshold Specification With an Application to the Purchasing Power Parity Relationship," *Journal of Business and Economic Statistics* 22, 382-395.
- [3] Bec, F., M. Ben Salem and M. Carrasco (2004b) "Detecting Mean Reversion in Real Exchange Rates from a Multiple Regime STAR Model," mimeo, CREST-ENSAE.
- [4] Bec, F., A. Guay and E. Guerre (2004) "Adaptive Consistent Unit Root Tests Based on Autoregressive Threshold Model," mimeo, CREST-ENSAE.
- [5] Berk, K. N. (1974) "Consistent Autoregressive Spectral Estimates," *Annals of Statistics* 2, 489-502.
- [6] Blanchard, O. J. and L. H. Summers (1986) "Hysteresis and the European Unemployment Problem," in S. Fischer ed., *NBER Macroeconomics Annual*, Cambridge MA: MIT Press, 15-78.
- [7] Brillinger, D. R. (1975) *Time Series: Data Analysis and Theory*. Holt, Rinehart and Winston: New York.

- [8] Brockwell, P. J. and R. A. Davis (1991) *Time Series: Theory and Methods*, 2nd ed. Springer-Verlag.
- [9] Caner, M., and B. E. Hansen (2001) "Threshold Autoregression With a Unit Root," *Econometrica*, 69, 1555-1596.
- [10] Chan, K. S., J. Petrucci, H. Tong and S. Woolford (1985) "A Multiple-Threshold AR(1) Model," *Journal of Applied Probability*, 22, 267-279.
- [11] Chang, Y. and J. Y. Park (2002) "On the Asymptotics of ADF Tests for Unit Roots," *Econometric Reviews* 21(4), 431-447.
- [12] Chung, K. L. and R. J. Williams (1990) *Introduction to Stochastic Integration*, 2nd ed. Boston: Birkhäuser.
- [13] Hall, P. and C. C. Heyde (1980) *Martingale Limit Theory and Its Application*. New York: Academic Press.
- [14] Kapetanios, G. and Y. Shin (2003) "Unit Root Tests in Three-Regime SETAR Models," mimeo, University of Edinburgh.
- [15] Kapetanios, G., Y. Shin and A. Snell (2003) "Testing for a Unit Root in the Nonlinear STAR Framework," *Journal of Econometrics*, 112, 359-379.
- [16] Michael, P., A. R. Nobay and D. A. Peel (1997) "Transaction Costs and Nonlinear Adjustments In Real Exchange Rates: An Empirical Investigation," *Journal of Political Economy* 105, 862-879.
- [17] Obstfeld, M. and A. M. Taylor (1997) "Nonlinear Aspects of Goods-Market Arbitrage and Adjustment: Heckscher's Commodity Points Revisited," *Journal of the Japanese and International Economies* 11, 441-479.
- [18] Park, J. Y. and P. C. B. Phillips (1999) "Asymptotics for Nonlinear Transformations of Integrated Time Series," *Econometric Theory*, 15, 269-298.
- [19] Park, J. Y. and P. C. B. Phillips (2001) "Nonlinear Regression with Integrated Time Series," *Econometrica* 69, 117-161.
- [20] Phillips, P. C. B. and V. Solo (1992) "Asymptotics for Linear Process," *Annals of Statistics* 20, 971-1001.
- [21] Revuz, D. and M. Yor (1994) *Continuous Martingale and Brownian Motion*, 2nd ed. New York: Springer-Verlag.
- [22] Rothman, P. (1998) "Forecasting Asymmetric Unemployment Rates," *Review of Economics and Statistics* 80, 164-168.
- [23] Seo, M. (2003) "Unit Root Test in a Threshold Autoregression: Asymptotic Theory and Residual-based Block Bootstrap," mimeo, University of Wisconsin-Madison.

- [24] Stock, J. H. (1994) "Unit Roots, Structural Breaks and Trends," in R.F. Engle and D. McFadden eds., *Handbook of Econometrics*, Vol. 4., Amsterdam: North-Holland, 2740-2841.
- [25] Taylor, A. M. (2001) "Potential Pitfalls for the Purchasing-Power-Parity Puzzle? Sampling and Specification Biases In Mean-Reversion Tests of the Law of One Price," *Econometrica* 69, 473-98.
- [26] Taylor, M. P., D. A. Peel and L. Sarno (2001) "Nonlinear Mean-Reversion in Real Exchange Rates: Toward a Solution to the Purchasing Power Parity Puzzles," *International Economic Review* 42, 1015-1042.
- [27] van Dijk, D., T. Terasvirta and P. Franses (2002) "Smooth Transition Autoregressive Models: A Survey of Recent Developments," *Econometric Reviews* 21, 1-47.

Table 1. Asymptotic Critical Values of Unit Root Tests

Transition function	Probability of a smaller value								
	1%	5%	10%	25%	50%	75%	90%	95%	99%
(a) TAR	-3.23	-2.67	-2.38	-1.94	-1.44	-0.75	0.14	0.49	1.11
(b) LSTAR	-3.21	-2.66	-2.37	-1.93	-1.44	-0.81	-0.02	0.34	0.96
(c) ESTAR	-2.85	-2.27	-1.98	-1.51	-1.01	-0.38	0.33	0.76	1.52
(d) D-TAR	-3.30	-2.71	-2.40	-1.92	-1.35	-0.67	0.18	0.67	1.48
(e) D-LSTAR	-3.29	-2.70	-2.40	-1.92	-1.35	-0.67	0.17	0.65	1.47

Note: Based on discrete approximation to the Brownian motion by partial sums of standard normal random variable with 1,000 steps and 10,000 replications.

Table 2. Asymptotic Critical Values of Unit Root Tests with Intercepts

Transition function	Probability of a smaller value								
	1%	5%	10%	25%	50%	75%	90%	95%	99%
(a) TAR	-3.92	-3.39	-3.11	-2.65	-2.20	-1.77	-1.39	-1.10	-0.33
(b) LSTAR	-3.99	-3.45	-3.16	-2.72	-2.26	-1.85	-1.49	-1.28	-0.75
(c) ESTAR	-3.86	-3.30	-3.03	-2.59	-2.16	-1.76	-1.42	-1.25	-0.94
(d) D-TAR	-3.63	-3.04	-2.77	-2.28	-1.75	-1.16	-0.32	0.25	1.14
(e) D-LSTAR	-3.73	-3.13	-2.86	-2.38	-1.83	-1.24	-0.41	0.14	1.02

Note: Based on discrete approximation to the Brownian motion by partial sums of standard normal random variable with 1,000 steps and 10,000 replications.

Table 3. Size of Unit Root Tests

test	n	α				
		-0.5	-0.2	0	0.2	0.5
(a) TAR	50	3.4	3.5	3.5	4.1	4.1
	100	3.9	4.1	4.2	4.4	4.7
	200	3.8	4.0	4.3	4.5	4.7
(b) LSTAR	50	4.0	4.4	4.1	4.9	5.0
	100	4.5	4.7	4.8	5.2	5.1
	200	4.2	4.4	4.7	4.8	5.1
(c) ESTAR	50	5.2	5.1	5.1	5.2	5.2
	100	5.4	5.3	5.4	5.4	5.6
	200	5.0	5.0	5.3	5.1	5.1
(d) ADF	50	4.7	4.8	4.8	4.9	4.9
	100	5.3	5.4	5.4	5.4	5.4
	200	4.9	4.9	4.8	4.9	4.8

Note: Empirical rejection rate of 5% level tests based on asymptotic critical values. Results are based on 10,000 replications.

Table 4. Size-Adjusted Power of Unit Root Tests Against TAR Model

n	$\lambda = -1.0$		-0.5		-0.3		-0.1	
	inf- t	ADF	inf- t	ADF	inf- t	ADF	inf- t	ADF
(a) TAR $\mu = -1.0$								
50	51.9	19.3	37.9	22.7	25.9	23.7	10.6	12.5
100	57.2	15.3	46.9	18.6	36.8	20.7	15.5	20.3
200	57.3	12.9	54.3	15.1	47.6	17.0	26.0	21.9
(a) TAR $\mu = -0.5$								
50	49.6	16.5	35.5	20.3	24.8	22.1	10.8	12.6
100	53.7	12.9	44.8	16.2	35.1	18.1	15.4	19.8
200	52.6	11.0	51.7	13.5	45.5	15.7	25.0	21.1
(a) TAR $\mu = 0$								
50	46.3	15.2	34.0	19.0	24.0	21.4	10.7	12.5
100	50.5	11.7	42.4	15.3	34.0	18.1	15.1	19.6
200	49.1	10.2	49.3	12.8	43.9	15.0	24.7	20.6
(a) TAR $\mu = 0.5$								
50	54.2	18.1	38.5	21.1	26.0	22.9	10.9	12.8
100	58.5	14.1	47.4	17.0	36.6	19.3	15.4	20.0
200	56.7	11.9	54.2	14.1	47.2	16.0	25.3	21.1
(a) TAR $\mu = 1.0$								
50	73.9	31.7	50.8	29.5	31.5	27.8	11.7	13.8
100	77.5	25.7	61.9	24.1	45.4	23.9	16.6	21.4
200	75.1	20.7	67.7	19.4	57.3	19.6	27.7	22.6

Note: Empirical rejection rate of 5% level tests based on size-adjusted critical values. Results are based on 10,000 replications. The value of inf- t test is shown in bold font if it is greater than that of ADF test.

Table 5. Size-Adjusted Power of Unit Root Tests Against STAR Model

n	$\lambda = -1.0$		-0.5		-0.3		-0.1	
	inf- t	ADF	inf- t	ADF	inf- t	ADF	inf- t	ADF
(b) LSTAR $(\mu, \kappa) = (0.0, 1.0)$								
50	76.5	52.1	48.7	37.5	30.6	31.3	11.1	13.8
100	85.9	52.2	62.7	36.1	45.8	30.9	17.3	22.5
200	89.2	48.2	73.4	31.8	59.0	26.8	29.9	25.4
(b) LSTAR $(\mu, \kappa) = (1.0, 1.0)$								
50	96.1	84.7	67.1	57.4	40.2	42.5	12.3	16.0
100	98.4	87.4	83.0	59.6	60.6	44.8	20.1	26.4
200	98.6	86.1	90.2	57.1	74.8	41.1	35.5	29.5
(b) LSTAR $(\mu, \kappa) = (0.0, 20.0)$								
50	46.0	15.3	34.4	19.1	25.1	21.5	11.1	12.5
100	50.5	11.8	41.2	15.3	33.5	18.1	15.6	19.6
200	50.4	10.2	49.0	12.8	42.9	15.0	26.1	20.6
(b) LSTAR $(\mu, \kappa) = (1.0, 20.0)$								
50	74.1	32.3	50.7	29.7	33.0	27.9	11.9	13.9
100	78.1	26.6	60.7	24.3	44.5	24.0	17.3	21.5
200	76.4	21.4	67.5	19.6	56.4	19.6	29.9	22.6
(c) ESTAR $\kappa = 0.1$								
50	38.7	23.8	17.1	11.6	10.5	8.3	6.3	6.0
100	93.9	85.6	64.0	43.7	37.6	22.8	11.3	9.0
200	100.0	100.0	99.7	99.6	94.8	89.7	41.3	30.3
(c) ESTAR $\kappa = 0.15$								
50	74.7	59.2	38.9	25.1	20.9	14.1	8.1	7.2
100	99.9	99.7	93.7	86.7	71.8	55.1	21.6	14.9
200	100.0	100.0	100.0	100.0	99.8	99.9	72.8	63.1
(c) ESTAR $\kappa = 0.2$								
50	92.0	87.8	61.8	46.6	35.2	24.1	10.8	8.9
100	100.0	100.0	99.3	98.8	90.5	83.4	33.0	23.1
200	100.0	100.0	100.0	100.0	100.0	100.0	88.4	85.6

Note: Empirical rejection rate of 5% level tests based on size-adjusted critical values. Results are based on 10,000 replications. The value of inf- t test is shown in bold font if it is greater than that of ADF test.

Table 6. US Unemployment Rate

(1) Unit root test							
Transition function	n	p	d	$\hat{\lambda}$	inf- t	ADF- μ	ADF- τ
(a) TAR	636	12	9	-0.04	-3.72*	-2.74	-2.82
(b) LSTAR	636	12	9	-0.04	-3.72*	-2.74	-2.82
(2) Size-adjusted power							
	inf- t			ADF- μ		ADF- τ	
(a) TAR	85.5			51.9		46.1	
(b) LSTAR	84.3			51.8		46.0	

Note: (*)Significant at 5% level. Sample period is 1948:1-2001:12. Size-adjusted power results are based on 2,000 replications. The value of inf- t test is shown in bold font if it is greater than that of ADF test.

Table 7. UK Sterling/US Dollar Real Exchange Rate

(1) Unit root test							
Transition function	n	p	d	$\hat{\lambda}$	inf- t	ADF- μ	ADF- τ
(c) ESTAR	379	1	1	-0.80	-3.35*	-2.61	-2.87
(2) Size-adjusted power							
	inf- t			ADF- μ		ADF- τ	
(c) ESTAR	57.3			38.6		22.5	

Note: (*)Significant at 5% level. Sample period is 1973:1-2004:7. Size-adjusted power results are based on 2,000 replications. The value of inf- t test is shown in bold font if it is greater than that of ADF test.

Table 8. French Franc/Deutschemark Exchange Rate:
Deviation from Central Parity

(1) Unit root test								
Transition Function	Sample	n	p	d	$\hat{\lambda}$	inf- t	ADF- μ	ADF- τ
(d) D-TAR	Full Sample	3645	5	2	-0.03	-5.94*	-4.68*	-5.87*
	Subsample	819	5	2	-0.07	-4.26*	-2.26	-2.39
(e) D-LSTAR	Full Sample	3645	5	2	-0.03	-5.75*	-4.68*	-5.87*
	Subsample	819	5	2	-0.09	-4.53*	-2.26	2.39
(2) Size-adjusted power								
		inf- t			ADF- μ		ADF- τ	
(d) D-TAR	Full Sample	100.0			100.0		100.0	
	Subsample	76.3			71.4		45.9	
(e) D-LSTAR	Full Sample	100.0			100.0		100.0	
	Subsample	67.8			56.8		36.0	

Note: (*)Significant at 5% level. Sample periods are 1979/3/13-1993/7/30 for full sample and 1979/3/13-1982/6/11 for subsample. Size-adjusted power results are based on 2,000 replications. The value of inf- t test is shown in bold font if it is greater than that of ADF test.