Testing the unit root hypothesis against the logistic smooth transition autoregressive model

Bruno Eklund*
Dept. of Economic Statistics

Stockholm School of Economics
SSE/EFI Working Paper Series in Economics and Finance
No. 546

November 2003

Abstract

In this paper two simple tests to distinguish between unit root processes and stationary nonlinear processes are proposed. New limit distribution results are provided, together with two F type test statistics for the joint unit root and linearity hypothesis against a specific nonlinear alternative. Nonlinearity is defined through the smooth transition autoregressive model. Due to occasional size distortion in small samples, a simple bootstrap method is proposed for estimating the p-values of the tests. Power simulations show that the two tests, F_{nd} and F_d, have at least the same or higher power than the corresponding Dickey-Fuller tests. Finally, as an example, the tests are applied on the seasonally adjusted U.S. monthly unemployment rate. The linear unit root hypothesis is strongly rejected, showing considerable evidence that the series is better described by a stationary smooth transition autoregressive process than a random walk.

Keywords: Smooth transition autoregressive model, nonlinearity, unit root, Brownian motion, critical values, bootstrap, Monte Carlo simulations, unemployment rate.
JEL codes: C12, C22.

Acknowledgement: Financial support from Jan Wallander's and Tom Hedelius's Foundation is gratefully acknowledged. I also wish to thank Pentti Saikkonen and Svend Hylleberg for helpful comments.

Stockholm School of Economics, P.O. Box 6501, SE-113 83 Stockholm, Sweden

*Bruno.Eklund@hhs.se

1 Introduction

The research and development of unit root tests have been considerable for the past twenty years. A main motivation has been to analyze and explain the allegedly poor properties of the standard unit root tests and to propose alternative tests, often in an attempt to increase the applicability and power of the tests. As Nelson and Plosser (1982) showed, standard Dickey-Fuller tests are unable to reject the hypothesis of a unit root in several macroeconomic and financial time series. The poor size and power properties of the standard tests when the time series contain structural changes, shifts in mean or growth rate, or nonlinear behavior, have been noticed in several studies. Pippenger and Goering (1993) showed how the power of the standard Dickey-Fuller tests falls considerably when the true alternative is a threshold autoregressive (TAR) model. They pointed out that in the presence of transaction costs or hysteresis thresholds the usefulness of standard unit root tests in examining long-run economic relationships is suspect. Diebold and Rudebusch (1990) showed similar loss in power when the true alternative is a fractionally integrated process. Perron (1989) argued that the low power against structural breaks in level and growth rate can result in overstating the evidence in favor of unit roots. The converse problem does, however, also exist, that standard unit root tests reject too often when there is a single structural break in trend or variance under the null hypothesis, as demonstrated by Leybourne, Mills and Newbold (1998), and Hamori and Tokihisa (1997). Nelson, Piger and Zivot (2001) showed similar results of size distortion when the true null model contains Markov regime switching in trend growth rate. They also showed low power testing the unit root hypothesis against an alternative process with Markov-switching trend.

Since the main focus of these studies was on analyzing the linear model, any possible nonlinear properties or features of the time series were ignored. On the other hand, there exists empirical evidence indicating that many features of macroeconomic and financial time series cannot be adequately described and analyzed using linear techniques. As a result, nonlinear models have become an active area of research in econometrics. Among other things, interest has been devoted to the problem of testing the joint hypothesis of linearity and unit root of a time series against specific nonlinear and stationary alternatives. The literature in this area has been growing rapidly.

An example of a recent study of this kind is Enders and Granger (1998). The authors analyzed and provided a test of the unit root hypothesis against the threshold autoregressive (TAR) model. They found that movements toward long-run equilibrium relationship of an interest rate are best estimated as an asymmetric process. Berben and van Dijk (1999), who applied a modified version of the Enders and Granger (1998) test, found asymmetric adjustments in several forward premium series. Caner and Hansen (2001), who also considered the TAR model as an alternative to the unit root hypothesis, proposed a bootstrap procedure to approximate the sampling distribution of the test statistic under the null. They reported strong evidence that U.S. male unemployment is better described by a stationary TAR process than a unit root process. Further examples are Kapetanios and Shin (2000), who developed and analyzed a test against the self-exciting threshold autoregressive (SETAR) model. Their test was more powerful than the Dickey-Fuller test that ignores the threshold nature under the alternative. Kapetanios, Shin and Snell (2003) considered a simple exponential smooth transition autoregressive model, only allowing for a regime shift in the slope parameter, as the alternative to the joint linearity and unit root hypothesis. As an illustration they provided an application to real interest rates, and rejected the null hypothesis for several interest rates considered, whereas the standard Augmented Dickey-Fuller tests failed to do that.

This paper will consider testing the joint linearity and unit root hypothesis against a smooth transition autoregressive (STAR) model. Standard STAR models has two extreme regimes, and the transition between them is smooth; see Teräsvirta (1994) for more discussion. Furthermore, the two-regime TAR model is included in the STAR model as a special case.

The paper will be organized as follows. Section 2 contains the model specification. Asymptotic results, limiting distributions for the two resulting test statistics and critical values are provided in section 3. Section 4 describes the bootstrap method to estimate p-values, and reports results of Monte Carlo simulations of the small sample properties, size and power, of the proposed tests. Section 5 contains an empirical application, and Section 6 concludes. The appendix includes proofs of the two theorems.

A few words on the notation. All limits are taken as $T \longrightarrow \infty$, and weak convergence is denoted as $\Rightarrow$.

2 Model, null hypothesis and auxiliary regression

Consider the following univariate smooth transition autoregressive (STAR) model

$$\Delta y_t = \theta_0 + \theta_1 \Delta y_{t-1} + \psi y_{t-1} + (\varphi_0 + \varphi_1 \Delta y_{t-1}) F(\gamma, c, y_{t-1}) + \varepsilon_t, \tag{1}$$

where $t = 1, \ldots, T$, and the dependent variable, y_t, is included in the model both as differences, Δy_t, and levels, y_t, but in the function $F(\cdot)$ only in levels. The differences, Δy_t, and errors, ε_t, are assumed to be stationary, and the errors satisfy $E\varepsilon_t = 0$. Furthermore, the transition function $F(\cdot)$ is a bounded continuous function, $F(\cdot) \in [-\frac{1}{2}, \frac{1}{2}]$. This allows the dynamic behavior of the model to change between two regimes corresponding to the cases when $F(\cdot) = -\frac{1}{2}$ and $F(\cdot) = \frac{1}{2}$, smoothly with the transition variable y_{t-1}. A number of different possibilities exist for the choice of the function $F(\cdot)$, see Granger and Teräsvirta (1993) or Teräsvirta (1998) for a presentation and discussion of the most common functional forms. This paper will focus on the logistic function

$$F(\gamma, c, y_{t-1}) = (1 + \exp(-\gamma (y_{t-1} - c)))^{-1} - \frac{1}{2}, \tag{2}$$

where the parameter c is the transition midpoint parameter, and γ the speed of transition from one extreme regime to the other. Note that as $\gamma \longrightarrow \infty$ the function $F(\cdot)$ approaches a step function, so the model ultimately becomes a threshold autoregressive (TAR) model, see Tong (1983). On the other hand, when $\gamma = 0$ the function $F(\cdot)$ is constant for all values of y_{t-1}, implying that model (1) is linear. The linearity hypothesis can thus be defined as $\gamma = 0$.

Testing $H_0 : \gamma = 0$ in (1) and (2) is not straightforward, however. The reason for this is the identification problem as (1) is only identified for $\gamma > 0$ but not for $\gamma = 0$, see Luukkonen, Saikkonen and Teräsvirta (1988), Teräsvirta (1994a,b), Lin and Teräsvirta (1994) for details. Following the idea in Luukkonen et al. (1988), this problem can be circumvented by a first-order Taylor approximation around $\gamma = 0$ in $F(\gamma, c, y_{t-1})$. This results in the following approximate model:

$$\begin{aligned}\Delta y_t &= \theta_0 + \theta_1 \Delta y_{t-1} + \psi y_{t-1} + (\varphi_0 + \varphi_1 \Delta y_{t-1}) \frac{\gamma}{4} (y_{t-1} - c) + \varepsilon_t^* \\ &= \left(\theta_1 - \frac{\varphi_1 \gamma c}{4}\right) \Delta y_{t-1} + \frac{\varphi_1 \gamma}{4} y_{t-1} \Delta y_{t-1} + \left(\theta_0 - \frac{\varphi_0 \gamma c}{4}\right) \\ &+ \left(\psi + \frac{\varphi_0 \gamma}{4}\right) y_{t-1} + \varepsilon_t^* \\ &= \delta \Delta y_{t-1} + \phi y_{t-1} \Delta y_{t-1} + \alpha + \zeta y_{t-1} + \varepsilon_t^*, \end{aligned} \tag{3}$$

where $\varepsilon_t^* = \varepsilon_t + R_1(\gamma, y_{t-1})(\varphi_0 + \varphi_1 \Delta y_{t-1})$, R_1 being the remainder. In equation (3), the linearity hypothesis now corresponds to $\phi = 0$. Also note that under the null hypothesis $\varepsilon_t^* = \varepsilon_t$ since the remainder $R_1 \equiv 0$. Moving y_{t-1} to the right hand side results in the following model:

$$y_t = \delta \Delta y_{t-1} + \phi y_{t-1} \Delta y_{t-1} + \alpha + \rho y_{t-1} + \varepsilon_t^*, \tag{4}$$

where $\rho = \zeta + 1$. When the additional term $y_{t-1} \Delta y_{t-1}$ is excluded, this auxiliary autoregression is the model used in the Augmented Dickey-Fuller (ADF) test with a constant and one lag of Δy_t as regressors. As a consequence, if $\varphi_1 = 0$ in the STAR model (1), the resulting auxiliary and ADF models are indistinguishable from each other, even for $\varphi_0 \neq 0$. No additional power, compared to the ADF tests, can therefore be expected of the two tests to be proposed below if there is a regime shift only in the intercept. A remedy to this problem would be to base the tests on a third-order Taylor approximation to (2), as in Teräsvirta (1994a,b). This case will not, however, be considered any further here.

In Kapetanios et al. (2003), the corresponding auxiliary autoregression had the form

$$\Delta y_t = \delta y_{t-1}^3 + \varepsilon_t, \tag{5}$$

indicating that they have a random walk without drift or time trend under their null hypothesis, $\delta = 0$. This was also implied by their alternative model, but by using de-meaned and de-trended variables in a two-step procedure they were able to allow for a random walk with drift and a random walk with time-trend under the null.

Since model (4) nests the ADF model, this specification makes it possible to set up a joint test of the unit root and linearity hypotheses, allowing y_t to follow a stationary nonlinear process under the alternative. A joint test of the unit root and linearity hypotheses against nonlinearity amounts to testing the hypothesis $H_{01} : \phi = \alpha = 0$, $\rho = 1$ in (4). It is easily seen that y_t in fact has a unit root under this null hypothesis, since model (4) then equals:

$$y_t = \delta \Delta y_{t-1} + y_{t-1} + \varepsilon_t. \tag{6}$$

Equation (6) can also be written as

$$\Delta y_t = \delta \Delta y_{t-1} + \varepsilon_t, \tag{7}$$

or, equivalently, as an infinite-order moving average model

$$\Delta y_t = (1 - \delta L)^{-1} \varepsilon_t = \sum_{i=0}^{\infty} \omega_i \varepsilon_{t-i} = \omega(L) \varepsilon_t = u_t, \tag{8}$$

where L is the lag operator, that is, $L y_t = y_{t-1}$. Thus, under H_{01}, $\{y_t\}$ is a random walk without drift. Under the maintained hypothesis of a unit root under the null, the hypothesis $H_{02} : \phi = 0$, $\rho = 1$ corresponds to the case of a random walk with drift.

From equation (3) it is seen that testing the hypothesis $\alpha = 0$ implies a test of $\theta_0 = 0$ in the STAR model (1) under the original linearity condition $\gamma = 0$. Also, by the same token, restriction $\rho = 1$ implies a test of $\psi = 0$. Thus, when the null hypothesis H_{01} is rejected, parameters θ_0 and ψ must be included in the alternative model.

Since the Δy_t process is assumed to be stationary, equation (7) implies that the parameter $|\delta| < 1$. Note that if $\delta = 1$, Δy_t has a unit root so that y_t is an $I(2)$ process. On the other hand, if $\delta = -1$, y_t has a negative unit root. From this follows that problems can arise in practice if δ is close to -1 or 1. This problem will be analyzed and discussed in section 4.

3 Limit results and the asymptotic tests

Let $b_T = \left(\widehat{\delta}, \widehat{\phi}, \widehat{\alpha}, \widehat{\rho}\right)'$ be the ordinary least squares estimator of $\beta = (\delta, \phi, \alpha, \rho)'$ in (4), so that

$$b_T - \beta = \left(\sum_{t=1}^{T} x_t x_t'\right)^{-1} \sum_{t=1}^{T} x_t \varepsilon_t, \tag{9}$$

where $x_t = (\Delta y_{t-1}, y_{t-1}\Delta y_{t-1}, 1, y_{t-1})'$. The convergence rates and limit distributions for most of the elements in the matrix and vector in equation (9) are known from previous studies. However, probability limits to some of the terms have not been considered before and are given in Theorem 1. First, the following assumption, employed by Hansen (1992), is assumed to be satisfied.

Assumption 1 *For some $q > \eta > 2$, $\{\upsilon_t\}$ is a zero mean, strong mixing sequence with mixing coefficients of size $-q\eta/(q-\eta)$, and $\sup_{t\geq 1} \|\upsilon_t\| = C < \infty$. In addition,*

$$T^{-1}E\left(V_T V_T'\right) \longrightarrow \Omega < \infty \text{ as } T \longrightarrow \infty, \text{ where } V_T = \sum_{t=1}^{T} \upsilon_t. \tag{10}$$

This assumption allows for a wide variety of different mixing processes, and in particular processes with weakly dependent heterogeneous observations that are common in econometric applications. The following result can now be stated:

Theorem 1 *Assume that $\{u_t\}$ in equation (8) satisfies Assumption 1, and let $\{\varepsilon_t\}$ be an i.i.d. sequence with mean zero, variance σ^2, and a finite fourth moment. Define*

$$\begin{aligned}
\gamma_j &= E\left(u_t u_{t-j}\right) = \sigma^2 \sum_{s=0}^{\infty} \omega_s \omega_{s+j} \quad , j = 0, 1, \ldots \\
\lambda &= \sigma \sum_{j=0}^{\infty} \omega_j = \sigma\omega(1) \\
\xi_t &= \sum_{i=0}^{t} u_i \quad , t = 1, 2, \ldots, T,
\end{aligned}$$

with $\xi_0 = 0$. *Then the following expressions converge jointly:*

$$(a) \quad T^{-3/2} \sum_{t=1}^{T} \xi_t u_t^2 \;\Rightarrow\; \gamma_0 \lambda \int_0^1 W(r)\,dr$$

$$(b) \quad T^{-3/2} \sum_{t=1}^{T} \xi_t u_t^3 \;\Rightarrow\; E\left(u_t^3\right) \lambda \int_0^1 W(r)\,dr$$

$$(c) \quad T^{-2} \sum_{t=1}^{T} \xi_t^2 u_t^2 \;\Rightarrow\; \gamma_0 \lambda^2 \int_0^1 W^2(r)\,dr$$

$$(d) \quad T^{-1} \sum_{t=1}^{T} \xi_{t-1} u_{t-1} \varepsilon_t \;\Rightarrow\; \sigma \sqrt{\gamma_0} \lambda \int_0^1 W(r)\,dB(r)$$

$$(e) \quad T^{-2} \sum_{t=1}^{T} \xi_t^2 u_t = o_p(1)$$

where $W(r)$ *and* $B(r)$ *are two independent standard Brownian motions defined for* $r \in [0,1]$.

Proof. See the Appendix.

Observing the rates of convergence in Theorem 1 and other known limit results allows one to define the scaling matrix

$$\Upsilon_T = diag\left(\; T^{1/2}, \;\; T, \;\; T^{1/2}, \;\; T \;\right). \tag{11}$$

Then, pre-multiplying $b_T - \beta$ in equation (9) by Υ_T, finite limits to the ordinary least squares estimates are given by

$$\Upsilon_T\left(b_T - \beta\right) = \left\{\Upsilon_T^{-1}\left(\sum_{t=1}^{T} x_t x_t'\right)\Upsilon_T^{-1}\right\}^{-1}\left\{\Upsilon_T^{-1}\left(\sum_{t=1}^{T} x_t \varepsilon_t\right)\right\}. \tag{12}$$

The null hypothesis, $H_{01}: \phi = \alpha = 0$, $\rho = 1$, has the alternative representation $H_{01}: R\beta = r$, where $R = \left[\; 0_3 \;\; I_3 \;\right]$, $0_3 = (0,0,0)'$, $\beta = (\delta, \phi, \alpha, \rho)'$ as before, and $r = \left(\; 0 \;\; 0 \;\; 1 \;\right)'$. The F test statistic is then defined in the usual way as

$$F = \left(b_T - \beta\right)'\left(R\Upsilon_T\right)'\left\{s_T^2 R\Upsilon_T\left(\sum_{t=1}^{T} x_t x_t'\right)^{-1}\Upsilon_T R'\right\}^{-1} R\Upsilon_T\left(b_T - \beta\right)/k, \tag{13}$$

where

$$s_T^2 = \frac{1}{T-4}\sum_{t=1}^{T}\left(y_t - \widehat{\delta}\Delta y_{t-1} - \widehat{\phi} y_{t-1}\Delta y_{t-1} - \widehat{\alpha} - \widehat{\rho} y_{t-1}\right)^2 \tag{14}$$

is a consistent estimator of σ^2, and k equals the number of restrictions. In the present case, $k = 3$. The corresponding statistic for testing the hypothesis H_{02} that allows a drift term is obtained by setting

$$R = \begin{bmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}, \; r = \begin{bmatrix} 0 \\ 1 \end{bmatrix} \tag{15}$$

and $k = 2$. The test statistics will be called F_{nd} and F_d, where nd and d stand for '*no drift*' and '*drift*', respectively. Obviously, they do not have standard asymptotic distributions as is the case in testing linearity in stationary STAR processes. Their asymptotic distribution theory under the two null hypotheses is given in the following theorem:

Theorem 2 *Let Assumption 1 and the results of Theorem 1 hold. Then the test statistics, F_{nd} and F_d, will have the following asymptotic distributions under the two null hypotheses;*

(*i*) *Under* $H_{01}: \phi = \alpha = 0, \ \rho = 1,$

$$F_{nd} \Rightarrow \frac{1}{3}W^2(1) + \frac{1}{3}\frac{\left(W(1)\int_0^1 W(r)\,dr - \int_0^1 W(r)\,dB(r)\right)^2}{\int_0^1 W^2(r)\,dr - \left(\int_0^1 W(r)\,dr\right)^2} + \frac{1}{3}\frac{\left(W(1)\int_0^1 W(r)\,dr - \frac{1}{2}\left\{W^2(1)-1\right\}\right)^2}{\int_0^1 W^2(r)\,dr - \left(\int_0^1 W(r)\,dr\right)^2}. \tag{16}$$

(*ii*) *Under* $H_{02}: \phi = 0, \ \rho = 1,$

$$F_d \Rightarrow \frac{1}{2}\frac{\left(W(1)\int_0^1 W(r)\,dr - \int_0^1 W(r)\,dB(r)\right)^2}{\int_0^1 W^2(r)\,dr - \left(\int_0^1 W(r)\,dr\right)^2} + \frac{1}{2}\frac{\left(W(1)\int_0^1 W(r)\,dr - \frac{1}{2}\left\{W^2(1)-1\right\}\right)^2}{\int_0^1 W^2(r)\,dr - \left(\int_0^1 W(r)\,dr\right)^2}. \tag{17}$$

Proof. See the appendix.

Asymptotic critical values for the F_{nd} and F_d statistics have been generated by simulation. This has been done by estimating the two F-type test statistics from observations generated from the null model (6) with $\delta = 0$ at sample sizes $T = 25, 50, 100, 250, 500$ with 1000000 replications. These values can be found in Table 1 where quantiles for the asymptotic distributions of the tests, (16) and (17), are also included. They have been estimated using the same number of replications as before with $T = 10000$ observations.

Table 1. Critical values for the test statistics F_{nd} and F_d, $\delta = 0$.

	F_{nd}					F_d				
T	0.10	0.05	0.025	0.01	0.001	0.10	0.05	0.025	0.01	0.001
25	3.24	4.04	4.87	6.02	9.34	4.23	5.38	6.58	8.25	12.99
50	3.09	3.76	4.43	5.33	7.72	4.08	5.06	6.05	7.36	10.81
100	3.04	3.66	4.27	5.07	7.05	4.04	4.96	5.85	7.03	10.01
250	3.02	3.62	4.20	4.95	6.79	4.03	4.92	5.78	6.90	9.63
500	3.02	3.61	4.18	4.91	6.72	4.03	4.92	5.77	6.86	9.55
∞	3.00	3.58	4.14	4.86	6.62	4.03	4.90	5.74	6.83	9.43

Note that even if the limit distributions do not depend on any nuisance parameters, the critical values for small sample sizes do. Under the null hypothesis, y_t is a function of δ, as is seen from equation (6). Thus, under perfect conditions, with δ known, critical values can be easily estimated. This is of course not normally the case in practice. As noted above, special attention is needed for values of δ close to -1 or 1. The time series Δy_t is then close to having a unit root or becoming nonstationary. In these situations the test may reject the null hypothesis too often. This property of the tests will be investigated in detail in the next section.

4 Small sample properties of the tests

In this section both the size and the power properties of these statistics are examined. For comparison, this is also done for the corresponding ADF tests, here called ADF_{nd} and ADF_d. Furthermore, a bootstrap method for adjusting the size of the tests in small samples is proposed.

4.1 Size simulations

To investigate the size properties the rejection probabilities of the true null hypothesis are computed, using Monte Carlo simulations and corresponding critical values from Table 1, for sample sizes $T = 25, 50, 100, 250, 500, 1000$. The nominal size considered for each test is 5%, and the number of replications equals 1000000. The data are generated from the null model (6) assuming $\{\varepsilon_t\} \sim nid\,(0, 1)$. Since the empirical size in small samples is affected by δ, it has been computed for a number of different values of δ ranging from minus one to one. Figures 1 to 4 display the estimated sizes of the F_{nd}, ADF_{nd}, F_d and ADF_d test statistics as functions of δ for the different sample sizes. In the two ADF tests, the regression equation contains the first lag of Δy_t.

The test statistics whose empirical sizes are shown in Figures 1 to 4 all share the characteristic that their size properties become poor for values of δ close to one. Furthermore, the test statistics F_{nd} and F_d are oversized for values of δ near -1 as well. As noted above, $\{y_t\}$ is $I(2)$ for $\delta = 1$, which manifests itself in high rejection frequencies of the true null hypothesis when δ is near unity. For $|\delta| = 1$ the stationarity assumption of Δy_t is violated, which, as can be expected, results in more frequent rejections of the null hypothesis than the asymptotic theory prescribes. In the figures, the smaller the sample size, the larger the deviation from the nominal 5% size level. The size tends towards the nominal 5% when the sample size is increased. Another notable fact is that the deviations from the nominal size for the two tests allowing for a drift term, F_d and ADF_d, are not as large as the ones for the F_{nd} and ADF_{nd}. As the critical values in Table 1 are estimated for $\delta = 0$, calculating new critical values or using bootstrapped p-values in the tests is recommended whenever δ is suspected to be close to -1 or 1 and the sample size is small.

The empirical size of F_{nd} and F_d have also been estimated for two cases of non-normal errors. Considering errors drawn from $t\,(6)$ and $\chi^2\,(1) - 1$ the Monte Carlo simulation setup is repeated. The simulation results indicate that the F_d is robust against both types of non-normal errors. No difference in estimated size can be detected between the two non-normal cases and the normal case. The same result holds for F_{nd} when considering $t\,(6)$-distributed errors. On the other hand, errors from the $\chi^2\,(1)-1$ distribution result in slightly higher rejection frequencies compared to the normal case at values of δ between -1 and about 0.6. The difference is the largest for the three smallest sample sizes, $T = 25, 50, 100$. For $T = 25$, the estimated size distortion is up to 3 percentage points higher, highest for values of δ close to -1. When $T = 100$ the size is only up to 1 percentage points higher. At other sample sizes, and for all sample sizes at values of δ

close to 1, no effect on the size can be detected compared to the normal case.

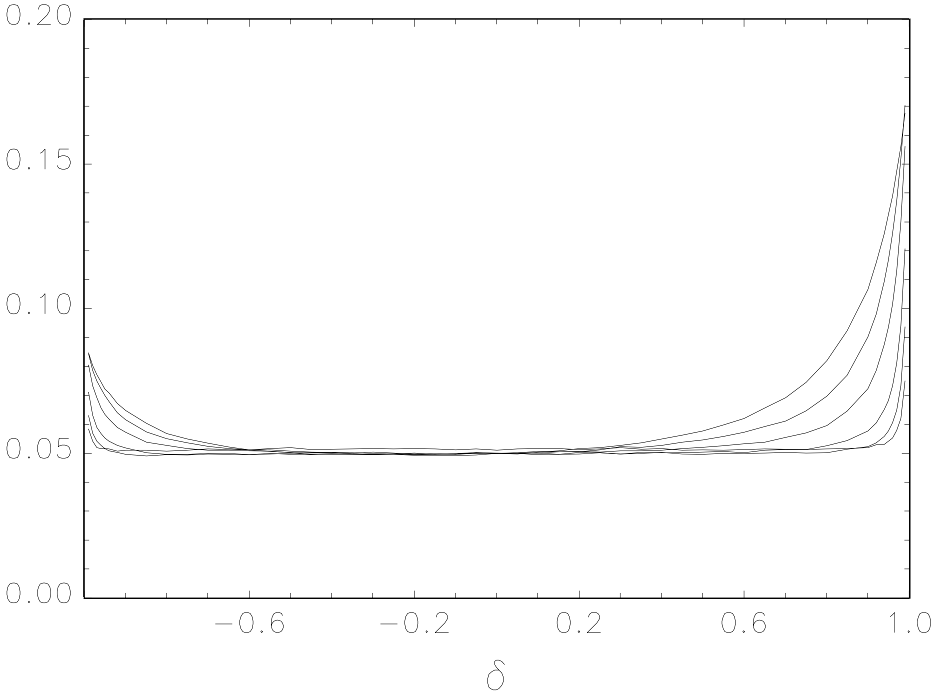

Figure 1. Size of the F_{nd} statistic for $T = 25$, 50, 100, 250, 500, 1000. The deviations from the nominal 5% size level, decrease with the sample size.

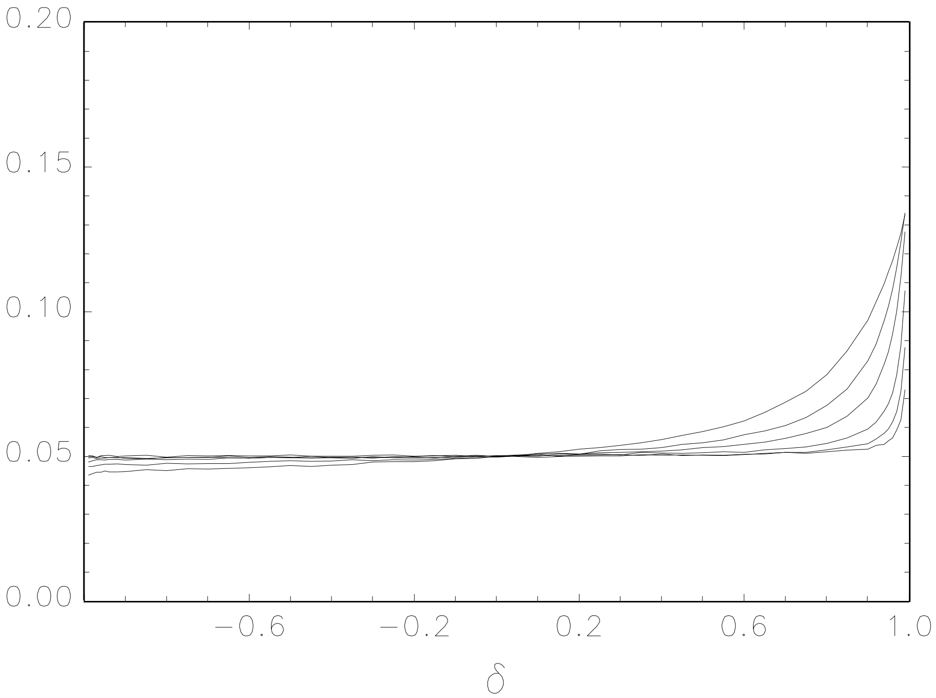

Figure 2. Size of the ADF_{nd} statistic for $T = 25$, 50, 100, 250, 500, 1000. The deviations from the nominal 5% size level, decrease with the sample size.

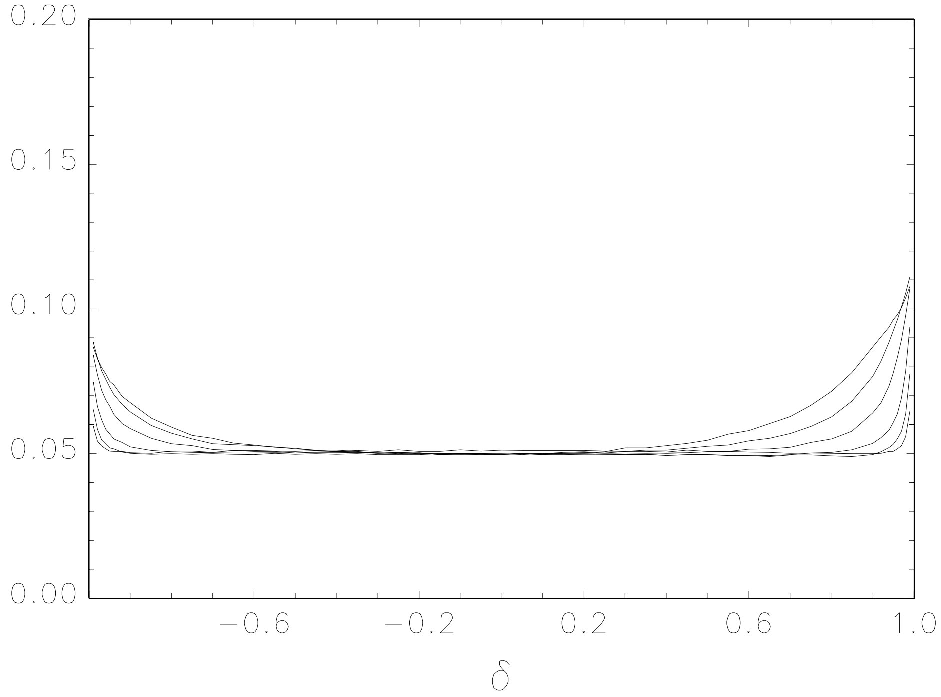

Figure 3. Size of the F_d statistic for $T = 25$, 50, 100, 250, 500, 1000. The deviations from the nominal 5% size level, decrease with the sample size.

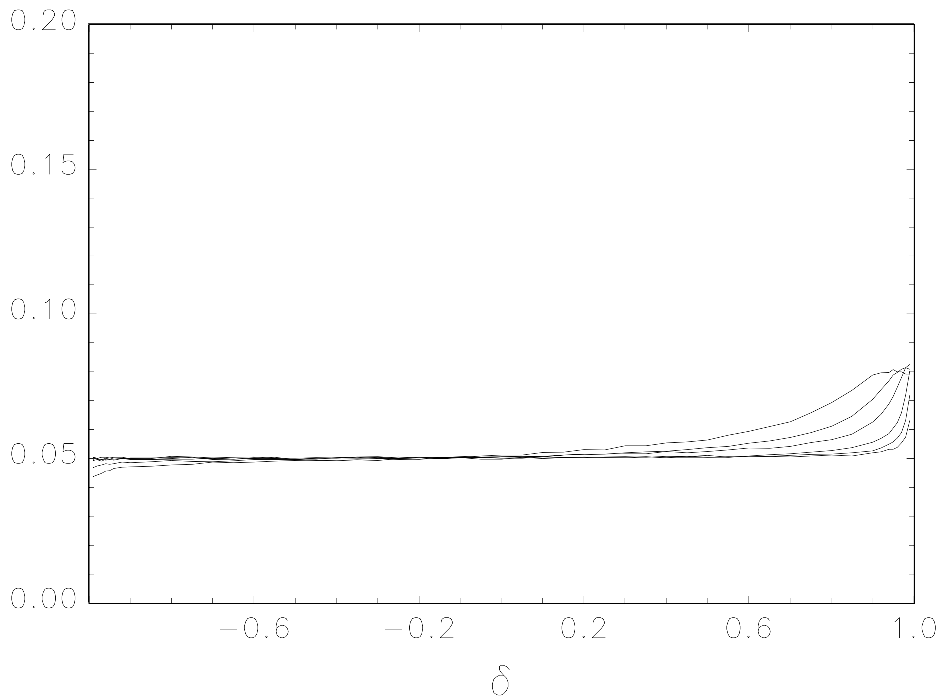

Figure 4. Size of the ADF_d statistic for $T = 25$, 50, 100, 250, 500, 1000. The deviations from the nominal 5% size level, decrease with the sample size.

4.2 Bootstrapping the critical values

In this section, a simple bootstrap method is suggested in order to correct the size distortion affecting the tests. Basing inference on a bootstrap distribution can substantially improve the finite sample properties of various test statistics, since the bootstrap p-value converges to the true p-value of the test as the number of bootstrap replications increases. In practice, the bootstrap p-value is estimated by simulation. For a survey on bootstrapping time series see Li and Maddala (1996). Caner and Hansen (2001) suggested, when testing the unit root hypothesis against the TAR model, basing the inference on a bootstrap approximation of the limit distribution of the test statistic under the null.

As is seen from the previous results, the small sample distributions of the two F tests, F_{nd} and F_d, depend on the parameter δ. Since the model simplifies to $y_t = \delta\Delta y_{t-1}+y_{t-1}+\varepsilon_t$ under H_{01}, and to $y_t = \delta\Delta y_{t-1} + \alpha + y_{t-1} + \varepsilon_t$ under H_{02}, a model-based bootstrap can be used for estimating p-values for both tests. Consider first statistic F_{nd}, let $\widehat{\delta}$ and $\widehat{D}$ be estimates of δ and the distribution D of the errors ε_t. Let ε_t^b be a random draw from $\widehat{D}$, and generate the bootstrap time series

$$y_t^b = \widehat{\delta}\Delta y_{t-1}^b + y_{t-1}^b + \varepsilon_t^b, \quad t = 1, \ldots, T. \tag{18}$$

Initial values for the resampling can be set to sample values of the demeaned series. The distribution of the series y_t^b is called the bootstrap distribution of the data. The test statistic, now called F_{nd}^b, is calculated from the resampled series y_t^b in the usual way. Repeating this resampling operation B times yields the empirical distribution of F_{nd}^b, which is the bootstrap distribution of F_{nd}, completely determined by $\widehat{\delta}$ and $\widehat{D}$. For a large number of independent F_{nd}^b tests, estimated from the B resampled series, the bootstrap p-value, defined by $p_T = P\left(F_{nd}^b > F_{nd}\right)$, can then be approximated by the frequency of simulated F_{nd}^b that exceeds the observed value of F_{nd}.

The resampling scheme is easily modified to fit statistic F_d. In order to obtain the bootstrap distribution of F_d, model (18) is augmented as follows:

$$y_t^b = \widehat{\delta}\Delta y_{t-1}^b + \widehat{\alpha} + y_{t-1}^b + \varepsilon_t^b, \quad t = 1, \ldots, T, \tag{19}$$

where $\widehat{\alpha}$ is the least squares estimator of α. The corresponding bootstrap distribution and the p-value are obtained as before.

4.3 Power simulations

The power study involves generating data from a stationary STAR model under the alternative hypothesis. However, there is no analytical answer to the question of which parameter combinations correspond to a stationary model under the alternative hypothesis. Only a guideline can be accomplished by setting the transition function $F = 0$ or 1, in order to obtain reasonable boundaries to the model parameters. On the other hand, simulations can show where the model is nonstationary, at least in the sense that a realization of $\{y_t\}$ cross a predetermined boundary. Such a crossing is taken to mean that the model is nonstationary for that specific choice of parameters. Using this idea in the Monte Carlo setup, the time series y_t is said to be nonstationary if $|y_t| > \sigma t$ for $t > 1000000$, where σ equals the standard error of the errors ε_t in (1). This is of course just a rough indication on nonstationarity, especially for parameter choices very close to the boundary between the stationary and nonstationary regions.

Data is generated from the STAR model (1) using $T = 50$, assuming $\{\varepsilon_t\} \sim nid\,(0, 1)$. Since no extra power, compared to the ADF tests, can be expected if a regime shift only involves the intercept, the two constants θ_0 and ϕ_0 are set to zero. The size of the regime shift is then determined by the parameters θ_1 and φ_1. Let $\theta_1 = -\varphi_1$ for simplicity, and let the parameters in the logistic function be $\gamma = 10$ and $c = 0$. The model is then determined by only two unknown parameters, φ_1 and ψ.

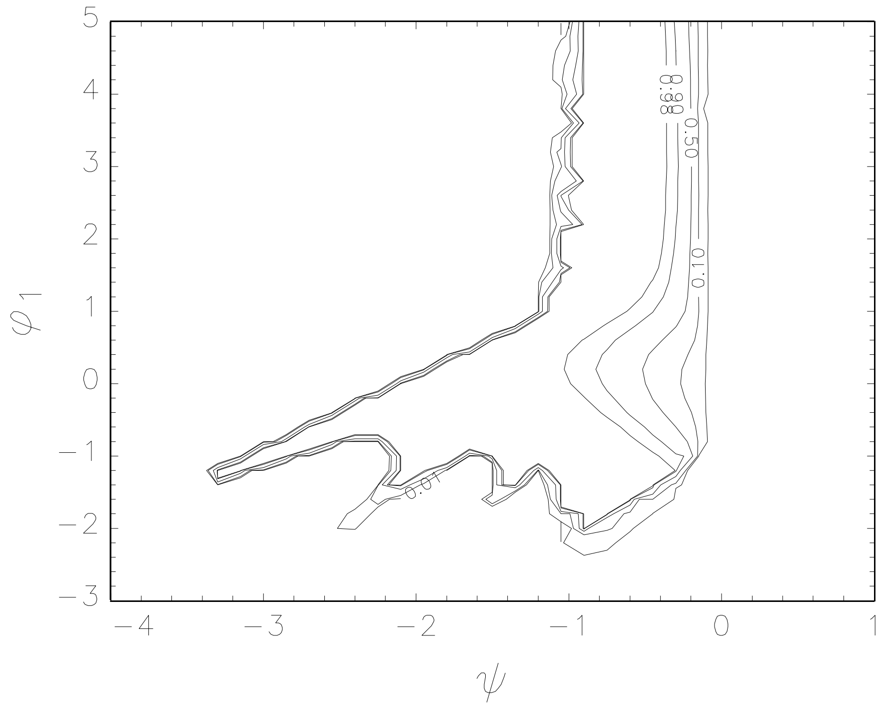

Figure 5. Power of the F_{nd} statistic for $T = 50$ observations.

Figures 5, 6, 8 and 9 shows contour plots of the power to the four tests, F_{nd}, ADF_{nd}, F_d and ADF_d, for a grid of φ_1 and ψ values. The difference in power between F_{nd} and ADF_{nd}, expressed as power of F_{nd} minus power of ADF_{nd}, can be found in Figure 7. The difference in power between F_d and ADF_d, expressed in the similar fashion, is shown in Figure 10. Due to substantial computational costs, the number of replications only equals 10 000, and 500 bootstrap replications are used to estimate the p-values.

The simulations show that the parameter combinations (φ_1, ψ) in the area outside the contour lines result in processes whose realizations grow without a bound with the number of observations. The largest gain in power from using F instead of ADFoccurs at parameter combinations of φ_1 and ψ such that $-1 < \psi < 0$. Other combinations indicate negligible or small differences in power between the two pairs of tests. The single largest gain for the F_{nd} test is as much as 56.7 percentage points whereas the smallest one equals -6.1 percentage points. The number of pairs (φ_1, ψ) with a negative gain corresponds to about 2.2% of the total number of pairs. The single largest and smallest gain for the F_d test are 52.2 and -6.9 percentage points, respectively. The gain is even here negative for about 2.2% of the pairs.

The negative gain for some of the parameter combinations can not be explained by sampling error alone. The main explanation is that the alternative STAR model at these

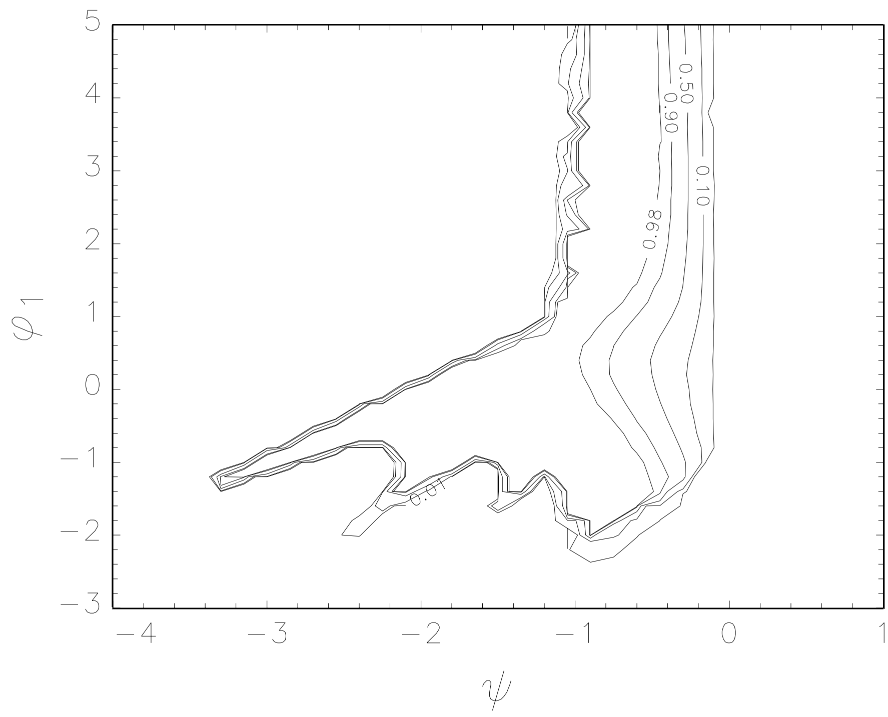

Figure 6. Power of the ADF_{nd} statistic for $T = 50$ observations.

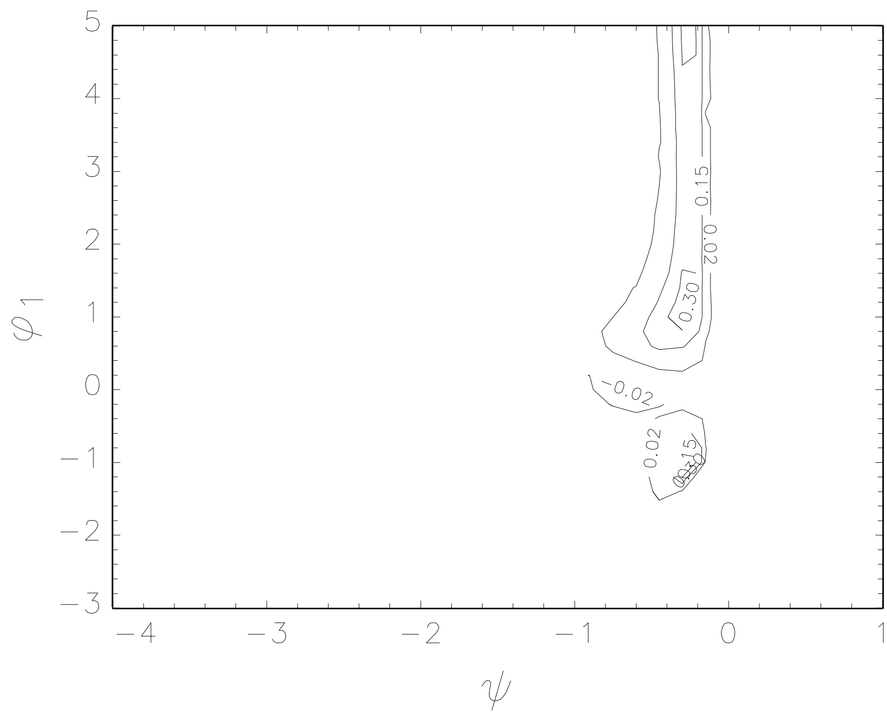

Figure 7. Difference in power, $F_{nd} - ADF_{nd}$.

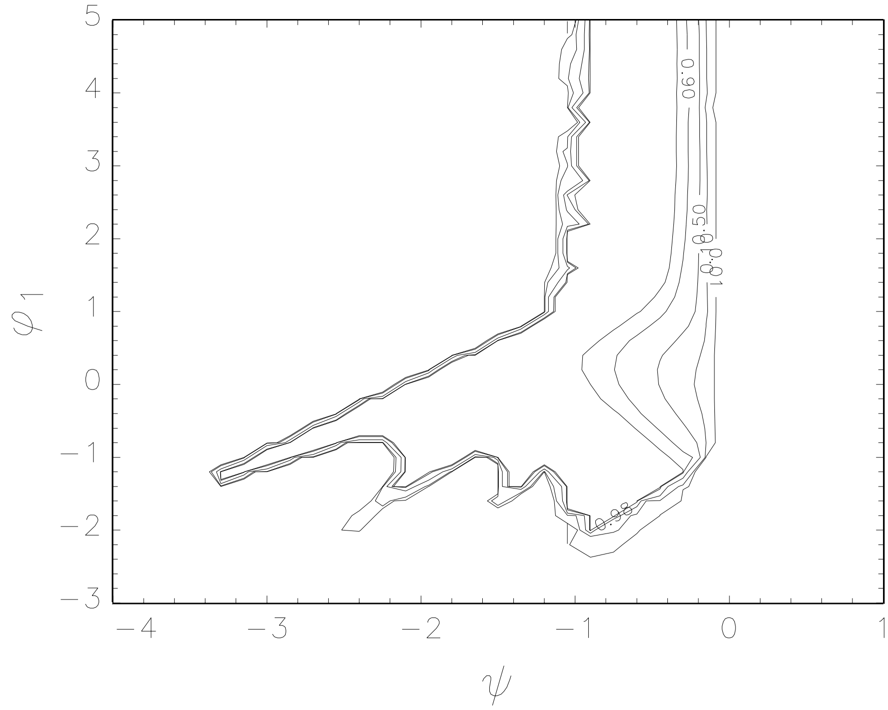

Figure 8. Power of the F_d statistic for $T = 50$ observations.

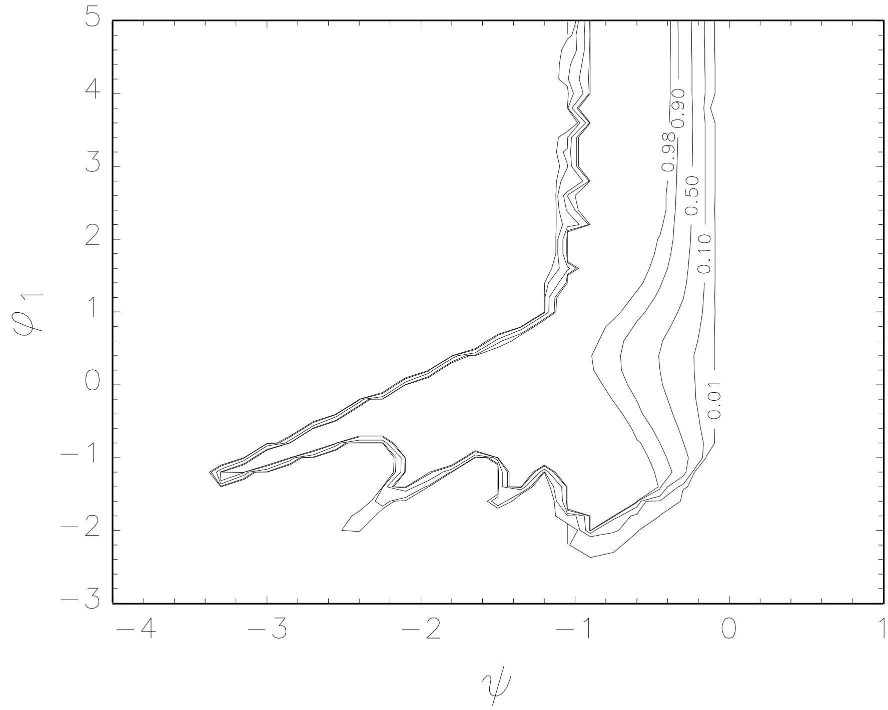

Figure 9. Power of the ADF_d statistic for $T = 50$ observations.

parameter combinations is very close to the linear alternative model considered in the ADF test. The auxiliary model (4) is then very close to or indistinguishable from the ADF model, and the power of the tests is reduced because of the extra parameter to be tested. As a whole, the simulation results show, however, that the F_{nd} and F_d tests have about the same or considerable higher power than the corresponding standard ADF tests when the alternative exhibit nonlinear behavior.

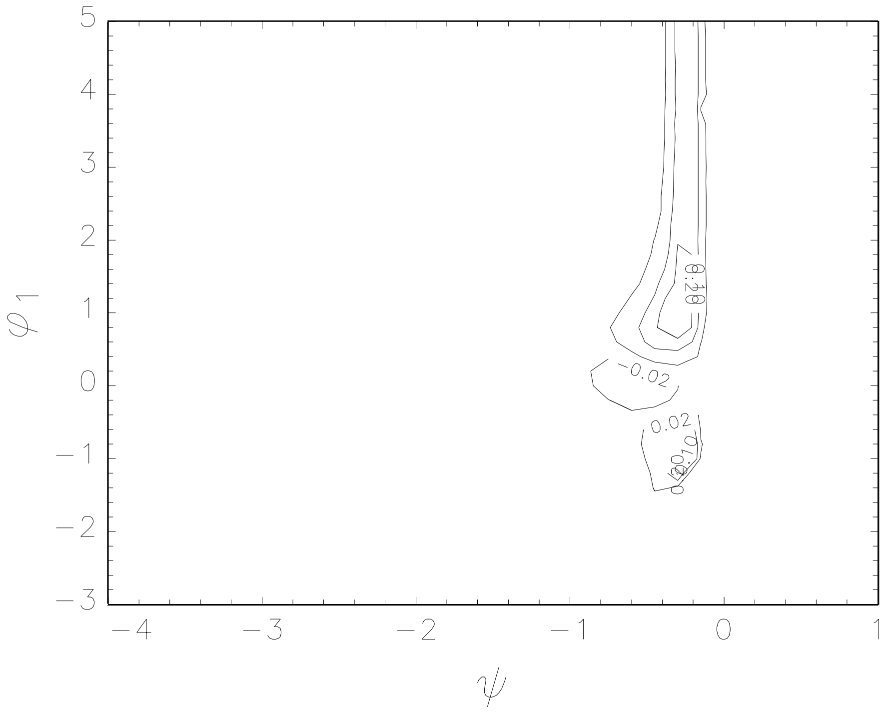

Figure 10. Difference in power, $F_d - ADF_d$.

5 Application

In order to demonstrate the use of the tests in practice they are applied to the seasonally adjusted monthly U.S. unemployment rate, from January 1961 to February 2000, obtained from SourceOECD. The sample period contains 468 observations of the differenced series. A plot of the time series can be found in Figure 11. A typical feature of the series is that there are periods of rapid increase of unemployment. An interesting feature is the asymmetry around the peaks, that is, the increase in unemployment is indeed more rapid than in the subsequent decrease. Such asymmetric behavior cannot be described properly with a linear model. Whether or not this unemployment rate can be assumed stationary is not quite clear from the figure, although a visual inspection may suggest mean reversion.

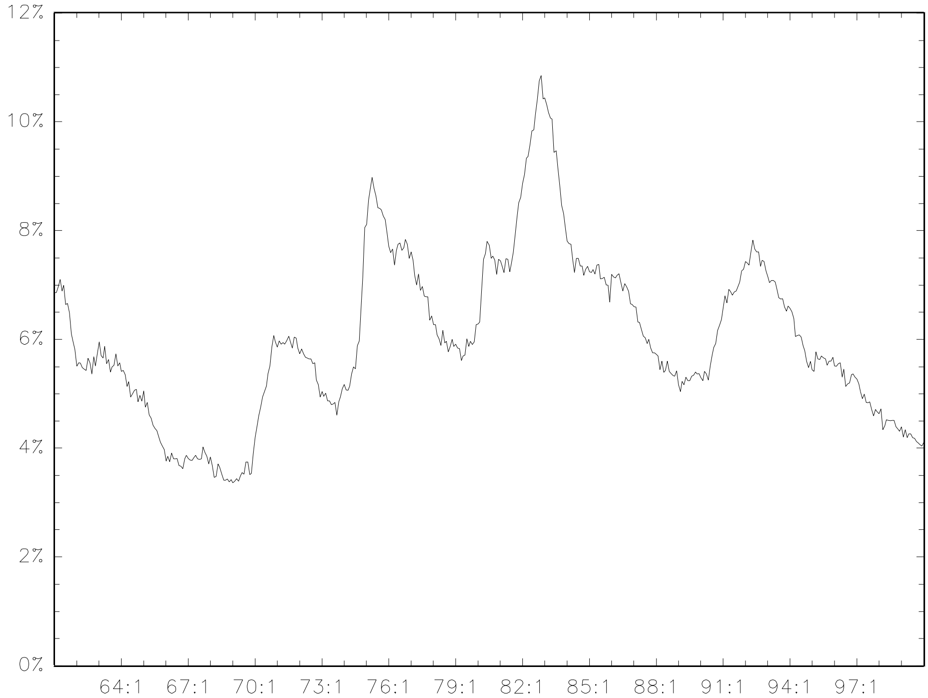

Figure 11. Seasonally adjusted unemployment for U.S.A. in %.

Carrying out the tests, using ordinary least squares, the auxiliary model (4) is estimated under the two null hypotheses H_{01} and H_{02}, and under the alternative hypothesis. The estimated equations and sums of squared residuals are as follows:

Under $H_{01} : \phi = \alpha = 0, \ \rho = 1,$

$$y_t = 0.1380\Delta y_{t-1} + y_{t-1} + \widehat{\varepsilon}_t, \qquad SSR_{01} = 13.49$$

Under $H_{02} : \phi = 0, \ \rho = 1,$

$$y_t = 0.1370\Delta y_{t-1} - 0.0051 + y_{t-1} + \widehat{\varepsilon}_t, \quad SSR_{02} = 13.48$$

Under the alternative hypothesis,

$$\begin{aligned} y_t = {} & -0.5301\Delta y_{t-1} + 0.0999 y_{t-1}\Delta y_{t-1} \\ & +0.0302 + 0.9938 y_{t-1} + \widehat{\varepsilon}_t, \qquad SSR = 13.09 \end{aligned}$$

Table 2 contains the observed values of the four test statistics F_{nd}, ADF_{nd}, F_d and ADF_d. The p-value for each test statistic has been estimated using the model-based bootstrap method with $B = 10\ 000$.

Table 2. Estimated test statistics and p-value.

	F_{nd}	ADF_{nd}	F_d	ADF_d
statistic	4.7554	0.9209	6.9160	1.4181
p-value	0.0174	0.8294	0.0095	0.6332

Since the p-values of F_{nd} and F_d are about 1.7% and 0.9%, the null hypotheses can be rejected at the 5% significance level. For 468 observations, the corresponding critical values from Table 1 are about 3.6 and 4.9 for the 5% level tests. The null hypotheses are thus also rejected when the inference is based on the critical values in Table 1. Using these critical values appears justified because of the negligible size distortion for values of δ close to the consistently estimated parameter $\widehat{\delta} = 0.1380$. The actual size for $\widehat{\delta}$ is very close to 5% as seen in Figures 1 and 3. These rejections support the conclusion that the U.S. seasonally adjusted monthly unemployment rate can be better characterized by a stationary nonlinear model than by a random walk. The use of the two ADF tests, however, leads to a different conclusion as the null hypothesis of a unit root is not rejected at any customary level of significance. The p-values are large, 83% and 63%, and corresponding critical values are 4.7 and 8.3 for ADF_{nd} and ADF_d respectively. As for the F tests, the actual size is very close to 5%, see Figures 2 and 4, and basing the ADF tests on the critical values in Table 1 would result in no or negligible size distortions.

6 Conclusions

This paper contains statistical theory for testing the unit root hypothesis against the smooth transition autoregressive (STAR) model. Some new limit results, two F-type test statistics and critical values for them are presented. The joint hypothesis of unit root and linearity allows one to distinguish between random walk processes, with or without drift, and stationary nonlinear processes of the smooth transition autoregressive type. This is important in applications because steps taken in modelling the series are likely to be drastically different depending on whether or not the unit root hypothesis is rejected. For illustration, the tests are applied to the seasonally adjusted U.S. monthly unemployment rate. The unit root hypothesis is strongly rejected, indicating that the unemployment series is better described by a STAR model rather than a random walk. The test result is of interest when the possible presence of hysteresis in the U.S. unemployment is considered.

References

[1] Berben, R.-P., and van Dijk, D., (1999), "Unit root tests and asymmetric adjustment: A reassessment", Working paper, Erasmus University Rotterdam.

[2] Caner, M., and Hansen, B. E., (2001), "Threshold autoregression with a unit root", *Econometrica*, 69, 1555-1596.

[3] Diebold, F. X., and Rudebusch, G. D., (1990), "On the power of Dickey-Fuller tests against fractional alternatives", *Economics Letters*, 35, 155-160.

[4] Enders, W., and Granger, C. W. J., (1998), "Unit-root tests and asymmetric adjustment with an example using the term structure of interest rate", *Journal of Business and Economic Statistics*, 16, 304-311.

[5] Granger, C. W. J., and Teräsvirta, T., (1993), *Modelling nonlinear economic relationships*, Oxford, Oxford University Press.

[6] Hamori, S., and Tokihisa, A., (1997), "Testing for a unit root in the presence of a variance shift", *Economics Letters*, 57, 245-253.

[7] Hansen, B. E., (1992), "Convergence to stochastic integrals for dependent heterogeneous processes", *Econometric Theory*, 8, 489-500.

[8] Kapetanios, G., and Shin, Y. (2000), "Testing for a linear unit root against nonlinear threshold stationarity", Discussion paper, http://www.econ.ed.ac.uk/research/discussion_papers.html, University of Edingburgh.

[9] Kapetanios, G., Shin, Y., and Snell, A., (2003), "Testing for a unit root in the nonlinear STAR framework", *Journal of Econometrics*, 112, 359-379.

[10] Leybourne, S. J., Mills, T. C., and Newbold, P., (1998), "Spurious rejections by Dickey-Fuller tests in the presence of a break under the null", *Journal of Econometrics*, 87, 191-203.

[11] Li, H. and Maddala, G. S., (1996), "Bootstrapping time series models", *Econometric Reviews*, 15, 115-158.

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

[13] Nelson, C. R., Piger, J., and Zivot, E., (2001), "Markov regime switching and unit root tests", *Journal of Business and Economic Statistics*, 19, 404-415.

[14] Nelson, C. R., and Plosser, C. I., (1982), "Trends and random walks in macroeconomic time series", *Journal of Monetary Economics*, 10, 139-162.

[15] Perron, P., (1989), "The great crash, the oil price shock, and the unit root hypothesis", *Econometrica*, 57, 1361-1401.

[16] Pippenger, M. K., and Goering, G. E., (1993), "A note on the empirical power of unit root tests under threshold processes", *Oxford Bulletin of Economics and Statistics*, 55, 473-481.

[17] Teräsvirta, T., (1994a), "Specification, estimation and evaluation of smooth transition autoregressive models", *Journal of the American Statistical Association*, 89, 208-218.

[18] Teräsvirta, T., (1994b), "Testing linearity and modelling nonlinear time series", *Kybernetika*, 30, 319-330.

[19] Teräsvirta, T., (1998), "Modeling economic relationships with smooth transition regressions", in A. Ullah and D. E. A. Giles (eds.), *Handbook of applied economic statistics*, Dekker, New York, 507-552.

[20] Tong, H., (1983), *Threshold models in non-linear time series analysis*, New York: Springer-Verlag.

Appendix A

Proof of Theorem 1

Let $\Longrightarrow$ denote weak convergence.

(a) Consider first

$$T^{-1}\sum_{t=1}^{T}\xi_t u_t^2 = T^{-1}\sum_{t=1}^{T}\xi_t\left(u_t^2 - Eu_t^2\right) + T^{-1}\sum_{t=1}^{T}\xi_t Eu_t^2. \tag{A.1}$$

Now, since $\xi_t = \xi_{t-1} - u_t$, the first sum on the right-hand side equals

$$T^{-1}\sum_{t=1}^{T}\xi_t\left(u_t^2 - Eu_t^2\right) = T^{-1}\sum_{t=1}^{T}\xi_{t-1}\left(u_t^2 - Eu_t^2\right) + T^{-1}\sum_{t=1}^{T}u_t\left(u_t^2 - Eu_t^2\right) \tag{A.2}$$

where the last term is $O_p(1)$, or $o_p(1)$ if $Eu_t^3 = 0$ as in the Gaussian case. Now let $\upsilon_t = \left(u_t, u_t^2 - Eu_t^2\right)'$, $V_t = \sum_{i=1}^{t}\upsilon_i$ and $V_0 = 0$. Then, from Hansen (1992), Theorem 4.1, it is known that the sum $T^{-1}\sum_{t=1}^{T}V_{t-1}\upsilon_t'$ converges weakly to a stochastic integral. Thus $T^{-1}\sum_{t=1}^{T}\xi_{t-1}\left(u_t^2 - Eu_t^2\right) = O_p(1)$ and

$$T^{-3/2}\sum_{t=1}^{T}\xi_t u_t^2 = T^{-3/2}\sum_{t=1}^{T}\xi_t Eu_t^2 + o_p(1) \Rightarrow \gamma_0\lambda\int_0^1 W(r)\,dr. \tag{A.3}$$

(b) Using the same idea as in (a),

$$T^{-1}\sum_{t=1}^{T}\xi_t u_t^3 = T^{-1}\sum_{t=1}^{T}\xi_t\left(u_t^3 - Eu_t^3\right) + T^{-1}\sum_{t=1}^{T}\xi_t Eu_t^3. \tag{A.4}$$

The first sum on the right-hand side then equals

$$T^{-1}\sum_{t=1}^{T}\xi_t\left(u_t^3 - Eu_t^3\right) = T^{-1}\sum_{t=1}^{T}\xi_{t-1}\left(u_t^3 - Eu_t^3\right) + T^{-1}\sum_{t=1}^{T}u_t\left(u_t^3 - Eu_t^3\right) \tag{A.5}$$

where the last term is $O_p(1)$. Let $\upsilon_t = (u_t, u_t^3 - Eu_t^3)'$, $V_t = \sum_{i=1}^{t} \upsilon_i$ and $V_0 = 0$. Then, again from Hansen (1992), Theorem 4.1, the sum $T^{-1} \sum_{t=1}^{T} V_{t-1}\upsilon_t'$ converges weakly to a stochastic integral. We have $T^{-1} \sum_{t=1}^{T} \xi_{t-1}(u_t^3 - Eu_t^3) = O_p(1)$ and

$$T^{-3/2} \sum_{t=1}^{T} \xi_t u_t^3 = T^{-3/2} Eu_t^3 \sum_{t=1}^{T} \xi_t + o_p(1) \Rightarrow Eu_t^3 \lambda \int_0^1 W(r)\, dr. \tag{A.6}$$

(c) As a starting-point, consider the sum

$$T^{-3/2} \sum_{t=1}^{T} \xi_{t-1}^2 u_t^2 = T^{-3/2} \sum_{t=1}^{T} \xi_{t-1}^2 \left(u_t^2 - Eu_t^2\right) + T^{-3/2} \sum_{t=1}^{T} \xi_{t-1}^2 Eu_t^2 \tag{A.7}$$

and let υ_t, V_t and V_0 be as in (a). Let Assumption 1 hold with $\eta = 3$. It then follows from Hansen (1992), Theorem 4.2, that the sum $T^{-3/2} \sum_{t=1}^{T} (V_{t-1} \otimes V_{t-1})\, \upsilon_t'$ converges weakly to a stochastic integral. This implies that

$$T^{-3/2} \sum_{t=1}^{T} \xi_{t-1}^2 \left(u_t^2 - Eu_t^2\right) = O_p(1) \tag{A.8}$$

so that

$$T^{-2} \sum_{t=1}^{T} \xi_{t-1}^2 u_t^2 = T^{-2} \sum_{t=1}^{T} \xi_{t-1}^2 Eu_t^2 + o_p(1) \Rightarrow \gamma_0 \lambda^2 \int_0^1 W^2(r)\, dr. \tag{A.9}$$

It remains to be shown that

$$T^{-2} \sum_{t=1}^{T} \xi_t^2 u_t^2 = T^{-2} \sum_{t=1}^{T} \xi_{t-1}^2 u_t^2 \text{ as } T \longrightarrow \infty. \tag{A.10}$$

Since $\xi_t = \xi_{t-1} + u_t$,

$$\sum_{t=1}^{T} \xi_t^2 u_t^2 = \sum_{t=1}^{T} (\xi_{t-1} + u_t)^2 u_t^2 = \sum_{t=1}^{T} \xi_{t-1}^2 u_t^2 + 2 \sum_{t=1}^{T} \xi_{t-1} u_t^3 + \sum_{t=1}^{T} u_t^4. \tag{A.11}$$

It follows directly from (A.11) that

$$\sum_{t=1}^{T} \left(\xi_t^2 - \xi_{t-1}^2\right) u_t^2 = 2 \sum_{t=1}^{T} \xi_{t-1} u_t^3 + \sum_{t=1}^{T} u_t^4, \tag{A.12}$$

where the two sums on the right-hand side are $O_p\left(T^{3/2}\right)$ and $O_p(T)$ respectively. The difference is thus $o_p(T^2)$, implying that (A.9) holds.

(*d*) Let $\upsilon_t = (u_t, u_{t-1}\varepsilon_t)'$, $V_t = \sum_{i=1}^{t} \upsilon_i$ and $V_0 = 0$. Then, since $T^{-1/2}V_T \Longrightarrow \left(\lambda W(1), \sigma\sqrt{\gamma_0}W(1)\right)'$, from Hansen (1992), Theorem 4.1, it follows that the elements of the sum

$$T^{-1}\sum_{t=1}^{T} V_{t-1}\upsilon_t' = T^{-1}\sum_{t=1}^{T}\begin{bmatrix} \xi_{t-1} \\ \sum_{i=1}^{t-1} u_{i-1}\varepsilon_i \end{bmatrix}\begin{bmatrix} u_t & u_{t-1}\varepsilon_t \end{bmatrix}$$

$$= T^{-1}\sum_{t=1}^{T}\begin{bmatrix} \xi_{t-1}u_t & \xi_{t-1}u_{t-1}\varepsilon_t \\ u_t\sum_{i=1}^{t-1} u_{i-1}\varepsilon_i & u_{t-1}\varepsilon_t\sum_{i=1}^{t-1} u_{i-1}\varepsilon_i \end{bmatrix} \quad \text{(A.13)}$$

converge weakly to some stochastic integrals. In particular,

$$T^{-1}\sum_{t=1}^{T} \xi_{t-1}u_{t-1}\varepsilon_t \Rightarrow \sigma\sqrt{\gamma_0}\lambda\int_0^1 W(r)\,dB(r) + \Lambda_{1,2} \quad \text{(A.14)}$$

where $\Lambda_{1,2}$ is element $(1,2)$ in the matrix

$$\Lambda = \lim_{T\longrightarrow\infty}\frac{1}{T}\sum_{i=1}^{T}\sum_{j=i+1}^{\infty} E\begin{bmatrix} u_i \\ u_{i-1}\varepsilon_i \end{bmatrix}\begin{bmatrix} u_j & u_{j-1}\varepsilon_j \end{bmatrix}$$

$$= \lim_{T\longrightarrow\infty}\frac{1}{T}\sum_{i=1}^{T}\sum_{j=i+1}^{\infty} E\begin{bmatrix} u_iu_j & u_iu_{j-1}\varepsilon_j \\ u_ju_{i-1}\varepsilon_i & u_{i-1}u_{j-1}\varepsilon_j\varepsilon_i \end{bmatrix} \quad \text{(A.15)}$$

Then, since u_iu_{j-1} and ε_j are independent for all $j \geq i+1$,

$$\Lambda_{1,2} = \lim_{T\longrightarrow\infty}\frac{1}{T}\sum_{i=1}^{T}\sum_{j=i+1}^{\infty} E\left(u_iu_{j-1}\varepsilon_j\right) = \lim_{T\longrightarrow\infty}\frac{1}{T}\sum_{i=1}^{T}\sum_{j=i+1}^{\infty} E\left(u_iu_{j-1}\right)E\left(\varepsilon_j\right) = 0. \quad \text{(A.16)}$$

Thus the two Brownian motions $W(r)$ and $B(r)$ are independent, and the result follows.

(*e*) As a starting-point, consider the sum $T^{-3/2}\sum_{t=1}^{T}\xi_{t-1}^2u_t$, and suppose that Assumption 1 holds for $\eta = 3$. Now let $\upsilon_t = u_t$, $V_t = \sum_{i=1}^{t}\upsilon_i$ and $V_0 = 0$. It follows from Hansen (1992), Theorem 4.2, that $T^{-3/2}\sum_{t=1}^{T}\left(V_{t-1}\otimes V_{t-1}\right)\upsilon_t$ converges weakly to a stochastic integral. Thus

$$T^{-3/2}\sum_{t=1}^{T}\xi_{t-1}^2u_t \Rightarrow \lambda^3\int_0^1 W^2(r)\,dW(r) + 2\lambda\Lambda\int_0^1 W(r)\,dr \quad \text{(A.17)}$$

where $\Lambda = \lim_{T\longrightarrow\infty}\frac{1}{T}\sum_{i=1}^{T}\sum_{j=i+1}^{\infty}Eu_iu_j$. The difference between the sum $\sum_{t=1}^{T}\xi_{t-1}^2u_t$ and $\sum_{t=1}^{T}\xi_t^2u_t$ is given by

$$\sum_{t=1}^{T}\left(\xi_t^2 - \xi_{t-1}^2\right)u_t = 2\sum_{t=1}^{T}\xi_{t-1}u_t^2 + \sum_{t=1}^{T}u_t^3 \quad \text{(A.18)}$$

where the two sums on the right hand side are $O_p\left(T^{3/2}\right)$ and $O_p\left(T\right)$ respectively. This implies that

$$\sum_{t=1}^{T}\xi_t^2 u_t = O_p\left(T^{3/2}\right) = o_p\left(T^2\right), \tag{A.19}$$

which proves (e) and completes the proof of Theorem 1. ■

Proof of Theorem 2

(i) Since $u_t = \Delta y_t$ under the null hypothesis, the elements in the matrix equation (12) are

$$\Upsilon_T^{-1}\left(\sum_{t=1}^{T} x_t x_t'\right)\Upsilon_T^{-1} = \tag{A.20}$$

$$= \begin{bmatrix} T^{-1}\sum_{t=1}^{T} u_{t-1}^2 & T^{-3/2}\sum_{t=1}^{T} y_{t-1}u_{t-1}^2 & T^{-1}\sum_{t=1}^{T} u_{t-1} & T^{-3/2}\sum_{t=1}^{T} y_{t-1}u_{t-1} \\ T^{-3/2}\sum_{t=1}^{T} y_{t-1}u_{t-1}^2 & T^{-2}\sum_{t=1}^{T} y_{t-1}^2 u_{t-1}^2 & T^{-3/2}\sum_{t=1}^{T} y_{t-1}u_{t-1} & T^{-2}\sum_{t=1}^{T} y_{t-1}^2 u_{t-1} \\ T^{-1}\sum_{t=1}^{T} u_{t-1} & T^{-3/2}\sum_{t=1}^{T} y_{t-1}u_{t-1} & 1 & T^{-3/2}\sum_{t=1}^{T} y_{t-1} \\ T^{-3/2}\sum_{t=1}^{T} y_{t-1}u_{t-1} & T^{-2}\sum_{t=1}^{T} y_{t-1}^2 u_{t-1} & T^{-3/2}\sum_{t=1}^{T} y_{t-1} & T^{-2}\sum_{t=1}^{T} y_{t-1}^2 \end{bmatrix}$$

and

$$\left\{\Upsilon_T^{-1}\left(\sum_{t=1}^{T} x_t \varepsilon_t\right)\right\}' = \tag{A.21}$$

$$= \left[\begin{array}{cccc} T^{-1/2}\sum_{t=1}^{T} u_{t-1}\varepsilon_t & T^{-1}\sum_{t=1}^{T} y_{t-1}u_{t-1}\varepsilon_t & T^{-1/2}\sum_{t=1}^{T}\varepsilon_t & T^{-1}\sum_{t=1}^{T} y_{t-1}\varepsilon_t \end{array}\right].$$

Given the limit distributions in Theorem 1 and other known limit results, the ordinary least squares estimator, (12), converges weakly as follows

$$\Upsilon_T\left(b_T - \beta\right) = \left\{\Upsilon_T^{-1}\left(\sum_{t=1}^{T} x_t x_t'\right)\Upsilon_T^{-1}\right\}^{-1}\left\{\Upsilon_T^{-1}\left(\sum_{t=1}^{T} x_t\varepsilon_t\right)\right\} \tag{A.22}$$

$$\Rightarrow \begin{bmatrix} \gamma_0 & \gamma_0\lambda\int_0^1 W(r)\,dr & 0 & 0 \\ \gamma_0\lambda\int_0^1 W(r)\,dr & \gamma_0\lambda^2\int_0^1 W^2(r)\,dr & 0 & 0 \\ 0 & 0 & 1 & \lambda\int_0^1 W(r)\,dr \\ 0 & 0 & \lambda\int_0^1 W(r)\,dr & \lambda^2\int_0^1 W^2(r)\,dr \end{bmatrix}^{-1}$$

$$\times \begin{bmatrix} \sigma\sqrt{\gamma_0}W(1) \\ \sigma\sqrt{\gamma_0}\lambda\int_0^1 W(r)\,dB(r) \\ \sigma W(1) \\ \frac{1}{2}\sigma\lambda\left\{W^2(1)-1\right\} \end{bmatrix}$$

$$= \frac{1}{\Sigma} \begin{bmatrix} \gamma_0^{-1}\int_0^1 W^2(r)\,dr & -(\gamma_0\lambda)^{-1}\int_0^1 W(r)\,dr & 0 & 0 \\ -(\gamma_0\lambda)^{-1}\int_0^1 W(r)\,dr & (\gamma_0\lambda^2)^{-1} & 0 & 0 \\ 0 & 0 & \int_0^1 W^2(r)\,dr & -\lambda^{-1}\int_0^1 W(r)\,dr \\ 0 & 0 & -\lambda^{-1}\int_0^1 W(r)\,dr & \lambda^{-2} \end{bmatrix}$$

$$\times \begin{bmatrix} \sigma\sqrt{\gamma_0}W(1) \\ \sigma\sqrt{\gamma_0}\lambda\int_0^1 W(r)\,dB(r) \\ \sigma W(1) \\ \frac{1}{2}\sigma\lambda\left\{W^2(1)-1\right\} \end{bmatrix}$$

where $\Sigma = \int_0^1 W^2(r)\,dr - \left(\int_0^1 W(r)\,dr\right)^2$. It then follows that, for the hypothesis $H_{01}: R\beta = r$ where $R = \begin{bmatrix} \mathbf{0} & I_3 \end{bmatrix}$, the F_{nd} test statistic defined by (13), equals

$$F_{nd} = (b_T - \beta)'(R\Upsilon_T)' \left\{ s_T^2 R\Upsilon_T \left(\sum_{t=1}^{T} x_t x_t'\right)^{-1} \Upsilon_T R' \right\}^{-1} R\Upsilon_T (b_T - \beta)/3 \quad \text{(A.23)}$$

$$\Rightarrow \frac{1}{3\sigma^2\Sigma} \begin{bmatrix} \sigma\sqrt{\gamma_0}W(1) & \sigma\sqrt{\gamma_0}\lambda\int_0^1 W(r)\,dB(r) & \sigma W(1) & \frac{1}{2}\sigma\lambda\{W^2(1)-1\} \end{bmatrix}$$

$$\times \begin{bmatrix} \gamma_0^{-1}\left(\int_0^1 W(r)\,dr\right)^2 & -(\gamma_0\lambda)^{-1}\int_0^1 W(r)\,dr & 0 & 0 \\ -(\gamma_0\lambda)^{-1}\int_0^1 W(r)\,dr & (\gamma_0\lambda^2)^{-1} & 0 & 0 \\ 0 & 0 & \int_0^1 W^2(r)\,dr & -\lambda^{-1}\int_0^1 W(r)\,dr \\ 0 & 0 & -\lambda^{-1}\int_0^1 W(r)\,dr & \lambda^{-2} \end{bmatrix}$$

$$\times \begin{bmatrix} \sigma\sqrt{\gamma_0}W(1) \\ \sigma\sqrt{\gamma_0}\lambda\int_0^1 W(r)\,dB(r) \\ \sigma W(1) \\ \frac{1}{2}\sigma\lambda\{W^2(1)-1\} \end{bmatrix}$$

$$= \frac{1}{3\Sigma}\left[W^2(1)\left(\int_0^1 W(r)\,dr\right)^2 - 2W(1)\int_0^1 W(r)\,dr\int_0^1 W(r)\,dB(r) + \left(\int_0^1 W(r)\,dB(r)\right)^2 \right.$$

$$\left. + W^2(1)\int_0^1 W^2(r)\,dr - W(1)\int_0^1 W(r)\,dr\{W^2(1)-1\} + \frac{1}{4}\{W^2(1)-1\}^2 \right]$$

$$= \frac{1}{3}W^2(1) + \frac{1}{3\Sigma}\left(W(1)\int_0^1 W(r)\,dr - \int_0^1 W(r)\,dB(r)\right)^2$$

$$+ \frac{1}{3\Sigma}\left(W(1)\int_0^1 W(r)\,dr - \frac{1}{2}\{W^2(1)-1\}\right)^2$$

which ends the proof of (i).

(ii) Hypothesis H_{02} has the alternative representation $H_{02}: R\beta = r$ where $R = \begin{bmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}$ and $r = \begin{bmatrix} 0 & 1 \end{bmatrix}'$. The F test statistic, defined by (13), equals

$$F_d = (b_T - \beta)'(R\Upsilon_T)' \left\{ s_T^2 R\Upsilon_T \left(\sum_{t=1}^{T} x_t x_t'\right)^{-1} \Upsilon_T R' \right\}^{-1} R\Upsilon_T (b_T - \beta)/2 \quad \text{(A.24)}$$

and converges weakly as follows:

$$F_d \Rightarrow \frac{1}{2\sigma^2\Sigma}\left[\begin{array}{cccc} \sigma\sqrt{\gamma_0}W(1) & \sigma\sqrt{\gamma_0}\lambda\int_0^1 W(r)\,dB(r) & \sigma W(1) & \frac{1}{2}\sigma\lambda\left\{W^2(1)-1\right\}\end{array}\right]$$

$$\times\left[\begin{array}{cccc} \gamma_0^{-1}\left(\int_0^1 W(r)\,dr\right)^2 & -(\gamma_0\lambda)^{-1}\int_0^1 W(r)\,dr & 0 & 0 \\ -(\gamma_0\lambda)^{-1}\int_0^1 W(r)\,dr & (\gamma_0\lambda^2)^{-1} & 0 & 0 \\ 0 & 0 & \left(\int_0^1 W(r)\,dr\right)^2 & -\lambda^{-1}\int_0^1 W(r)\,dr \\ 0 & 0 & -\lambda^{-1}\int_0^1 W(r)\,dr & \lambda^{-2}\end{array}\right]$$

$$\times\left[\begin{array}{c} \sigma\sqrt{\gamma_0}W(1) \\ \sigma\sqrt{\gamma_0}\lambda\int_0^1 W(r)\,dB(r) \\ \sigma W(1) \\ \frac{1}{2}\sigma\lambda\left\{W^2(1)-1\right\}\end{array}\right]$$

$$= \frac{1}{2\Sigma}\left(2W^2(1)\left(\int_0^1 W(r)\,dr\right)^2 - 2W(1)\int_0^1 W(r)\,dr\int_0^1 W(r)\,dB(r)\right.$$

$$\left. + \left(\int_0^1 W(r)\,dB(r)\right)^2 - W(1)\int_0^1 W(r)\,dr\left\{W^2(1)-1\right\} + \frac{1}{4}\left\{W^2(1)-1\right\}^2\right)$$

$$= \frac{1}{2\Sigma}\left(W(1)\int_0^1 W(r)\,dr - \int_0^1 W(r)\,dB(r)\right)^2$$

$$+ \frac{1}{2\Sigma}\left(W(1)\int_0^1 W(r)\,dr - \frac{1}{2}\left\{W^2(1)-1\right\}\right)^2$$

where $\Sigma = \int_0^1 W^2(r)\,dr - \left(\int_0^1 W(r)\,dr\right)^2$ as in (i). This completes the proof of (ii) and Theorem 2. ■