A nonlinear alternative to the unit root hypothesis

Bruno Eklund*
Dept. of Economic Statistics

Stockholm School of Economics
SSE/EFI Working Paper Series in Economics and Finance
No. 547

November 2003

Abstract

This paper considers testing the unit root hypothesis against a smooth transition autoregressive model as the alternative. The model specification makes it possible to discriminate between nonstationary random walk and stationary nonlinear processes. Some new limit results are presented, extending earlier work, and two F type tests are proposed. Small sample simulations show some size distortions, why a bootstrap method for estimating p-values to the tests are considered. Power simulations show some gain in power, compared to the common Augmented Dickey-Fuller tests. Finally, the two proposed F type tests are applied on a number of real exchange rates. For several of the exchange rates considered the linear unit root is rejected in favor of the stationary nonlinear model, supporting the purchasing power parity hypothesis.

Keywords: Smooth transition autoregressive model, nonlinearity, unit root, Brownian motion, bootstrap, critical values, Monte Carlo simulations, real exchange rates.
JEL codes: C12, C22, F31.

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 valuable comments.

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

*Bruno.Eklund@hhs.se

1 Introduction

There are situations in which the standard Dickey-Fuller tests do not function as well as one might desire. Difficulties in detecting structural change, shifts in mean or growth rate, or nonlinear behavior have been noticed in several studies. Nelson and Plosser (1982) showed for a number of financial and macroeconomical time series that the Dickey-Fuller tests are unable to reject the unit root hypothesis. Pippenger and Goering (1993) argued that examining long-run economic relationships using the unit root tests is questionable in the presence of transaction costs or hysteresis thresholds. Perron (1989) argued that low power against structural breaks in level and growth rate can result in overstating the evidence in favor of unit roots. However, the standard Dickey-Fuller tests are also affected by size distortions in a number of situations. Size distortions, in the form of too frequent rejections of the null, have been observed when there is a single structural break in trend or variance under the null hypothesis, see Leybourne, Mills and Newbold (1998) and Hamori and Tokihisa (1997). Similar size distortion, when the true null model contains Markov regime switching in trend growth rate, was demonstrated by Nelson, Piger and Zivot (2001). They also showed power loss when the unit root hypothesis was tested against a true alternative process with a Markov-switching trend.

The emphasis in these studies, and in a major part of the early literature, has been on the linear model which today is increasingly viewed to be somewhat inadequate. Any possible nonlinear characteristics of the time series have thus been ignored. The increasing empirical evidence on nonlinear relationships and features in economical time series during the last few years has resulted in efforts to incorporate nonlinear models and techniques into the existing econometric framework. The literature on testing the unit root hypothesis against nonlinear models, or vice versa, has recently grown rapidly in this direction.

One of the more recent studies in this area is Caner and Hansen (2001) who analyzed and provided tests of the unit root hypothesis against the threshold autoregressive (TAR) model. The authors also proposed a bootstrap method to approximate the limit distribution of the test under the null, and showed that the unit root hypothesis can be rejected for the U.S. unemployment rate in favor of the nonlinear TAR model. Also considering the TAR model as an alternative to the unit root, Enders and Granger (1998) found that movements toward long-run equilibrium relationship of an interest rate are best described as an asymmetric process. Modifying the test in Enders and Granger (1998), Berben and van Dijk (1999) found evidence of asymmetric adjustments towards long-run equilibrium for a number of forward premium time series. Kapetanios and Shin (2000) developed and analyzed the unit root test with the self-exciting threshold autoregressive (SETAR) model as the alternative process. Designed to take into account the threshold nature under the alternative, they reported some gain in power compared to the Dickey-Fuller test. Other studies have integrated the unit root test with nonlinear models with a smooth transition between the regimes. Kapetanios, Shin and Snell (2003) proposed a test of the joint unit root and linearity hypothesis against a very simple exponential smooth transition autoregressive (ESTAR) model that only allows a regime shift in the slope parameter. They were able to reject the unit root for a number of the real interest rates in favor of the ESTAR model. Bec, Salem and Carrasco (2002) tested the unit root hypothesis against a nonlinear logistic STAR model with three regimes. Rejecting the unit root for eleven out of twenty eight real exchange rates considered, their empirical results lent support to the so called purchasing power parity (PPP) hypothesis and indicated also strong mean reversion for large departures from PPP. Eklund (2003) proposed and analyzed the unit root test against the logistic smooth transition autoregressive (LSTAR) model, allowing for regime shift in both intercept and growth rate, and showed that the U.S. monthly unemployment rate is better described by a STAR model rather than a random walk.

In this paper, the recent work in Eklund (2003) is extended. Two tests are constructed for the joint linearity and unit root hypothesis against the second-order logistic STAR model. Compared to the earlier work in Eklund (2003), the alternative STAR model in this paper allows for regime shifts in intercept, growth rate and in level. The nonlinear model considered allow the adjustment towards long-run equilibrium to be sudden as well as smooth.

The paper is outlined as follows. In Section 2 the model is specified. Limit results and critical values are found in Section 3, while Section 4 includes a Monte Carlo study of the size and the power properties. Section 5 includes a small introduction to the so called purchasing power parity problem, and an empirical application on real exchange rates. Concluding remarks are given in Section 6, and mathematical proofs are presented in the appendix.

2 Model and joint unit root and linearity hypothesis

Consider the univariate smooth transition autoregressive (STAR) model

$$\Delta y_t = \theta_0 + \theta_1 \Delta y_{t-1} + \psi_1 y_{t-1} + \left(\varphi_0 + \varphi_1 \Delta y_{t-1} + \psi_2 y_{t-1}\right) F\left(\gamma, c_1, c_2, \triangle y_{t-1}\right) + \varepsilon_t, \quad (1)$$

where Δy_t, and errors, ε_t, are assumed to be stationary, satisfying $E\varepsilon_t = 0$, $E\left|\varepsilon_t\right|^{6+r} < \infty$ for some $r > 0$, and $t = 1, \ldots, T$. The nonlinearity is introduced via the transition function $F(\cdot)$ which is a bounded continuous function such that $F(\cdot) \in [-\frac{1}{2}, \frac{1}{2}]$. This allows the dynamic behavior of Δy_t to change smoothly and nonlinearly with the transition variable $\triangle y_{t-1}$ between the two regimes, $F(\cdot) = -\frac{1}{2}$ and $F(\cdot) = 1/2$. Several possibilities exists for the choice of the function $F(\cdot)$, see Granger and Teräsvirta (1993) and Teräsvirta (1998) for a detailed presentation and discussion. This paper will focus on (1) with a second-order logistic function

$$F\left(\gamma, c_1, c_2, \triangle y_{t-1}\right) = \left(1 + \exp\left[-\gamma\left(\triangle y_{t-1} - c_1\right)\left(\triangle y_{t-1} - c_2\right)\right]\right)^{-1} - \frac{1}{2}, \quad (2)$$

where the parameters c_1 and c_2 are the threshold parameters and γ is the speed of transition between the regimes, $\gamma > 0$ for identification reasons. Note that the function $F(\cdot)$ is constant for $\gamma = 0$, in which case model (1) is linear. This fact can be used when testing linearity. However, testing the hypothesis $H_0 : \gamma = 0$ in model (1) cannot be performed directly, since this restriction involves an identification problem, see Luukkonen, Saikkonen and Teräsvirta (1988), Teräsvirta (1994a,b), Lin and Teräsvirta (1994) for details.

Applying the idea by Luukkonen et al. (1988), the identification problem can be circumvented by a first-order Taylor approximation of $F\left(\gamma, c_1, c_2, \triangle y_{t-1}\right)$ around $\gamma = 0$.

Inserting the approximation in (1) results in the following auxiliary model:

$$
\begin{aligned}
\Delta y_t &= \theta_0 + \theta_1 \Delta y_{t-1} + \psi_1 y_{t-1} \\
&\quad + \left(\varphi_0 + \varphi_1 \Delta y_{t-1} + \psi_2 y_{t-1}\right) \frac{\gamma}{4} \left(\triangle y_{t-1} - c_1\right)\left(\triangle y_{t-1} - c_2\right) + \varepsilon_t^* \\
&= \left(\theta_1 - \frac{\varphi_0 \gamma \left(c_1 + c_2\right)}{4} + \frac{\varphi_1 \gamma c_1 c_2}{4}\right) \Delta y_{t-1} + \frac{\varphi_0 \gamma - \varphi_1 \gamma \left(c_1 + c_2\right)}{4} \left(\triangle y_{t-1}\right)^2 + \\
&\quad + \frac{\varphi_1 \gamma}{4} \left(\triangle y_{t-1}\right)^3 - \frac{\psi_2 \gamma \left(c_1 + c_2\right)}{4} y_{t-1} \triangle y_{t-1} + \frac{\psi_2 \gamma}{4} y_{t-1} \left(\triangle y_{t-1}\right)^2 + \\
&\quad + \left(\theta_0 + \frac{\varphi_0 \gamma c_1 c_2}{4}\right) + \left(\psi_1 + \frac{\psi_2 \gamma c_1 c_2}{4}\right) y_{t-1} + \varepsilon_t^* \\
&= \delta_1 \Delta y_{t-1} + \delta_2 \left(\Delta y_{t-1}\right)^2 + \delta_3 \left(\Delta y_{t-1}\right)^3 + \phi_1 y_{t-1} \Delta y_{t-1} + \\
&\quad + \phi_2 y_{t-1} \left(\Delta y_{t-1}\right)^2 + \alpha + \zeta y_{t-1} + \varepsilon_t^* .
\end{aligned} \tag{3}
$$

where $\varepsilon_t^* = \varepsilon_t + \left(\varphi_0 + \varphi_1 \Delta y_{t-1} + \psi_2 y_{t-1}\right) R_1\left(\gamma, \Delta y_{t-1}\right)$, and R_1 is the remainder. Moving y_{t-1} from the left-hand to the right-hand side in (3) yields the form

$$
\begin{aligned}
y_t &= \delta_1 \Delta y_{t-1} + \delta_2 \left(\Delta y_{t-1}\right)^2 + \delta_3 \left(\Delta y_{t-1}\right)^3 + \phi_1 y_{t-1} \Delta y_{t-1} + \\
&\quad + \phi_2 y_{t-1} \left(\Delta y_{t-1}\right)^2 + \alpha + \rho y_{t-1} + \varepsilon_t^* ,
\end{aligned} \tag{4}
$$

where the original linearity condition, $\gamma = 0$, now corresponds to $\delta_2 = \delta_3 = \phi_1 = \phi_2 = 0$. Note that $\varepsilon_t^* = \varepsilon_t$ under the linearity hypothesis, since $R_1 = 0$ when $\gamma = 0$. As the regression model for the Augmented Dickey-Fuller (ADF) test, with a constant and one lag of Δy_t, is nested in this auxiliary model, a joint test of the linearity and the unit root hypothesis amounts to testing the hypothesis $H_{01} : \delta_2 = \delta_3 = \phi_1 = \phi_2 = \alpha = 0$, $\rho = 1$ in (4). Under this hypothesis equation (4) becomes

$$
y_t = \delta_1 \Delta y_{t-1} + y_{t-1} + \varepsilon_t . \tag{5}
$$

In reduced form equation (5) equals

$$
\Delta y_t = \frac{\varepsilon_t}{1 - \delta_1 L} = \sum_{i=0}^{\infty} \omega_i \varepsilon_{t-i} = \omega\left(L\right) \varepsilon_t = u_t , \tag{6}
$$

where L is the lag operator, i.e. $L y_t = y_{t-1}$. Under H_{01}, $\{y_t\}$ is thus a unit root process without drift. Note that if H_{01} is rejected and the alternative is accepted as a basis for further modelling, the parameters θ_0 and ψ_1 should be included in the alternative model. This follows from the fact that testing $\alpha = 0$ also implies a test of $\theta_0 = 0$ in equation (3) under the original linearity condition $\gamma = 0$. The same reason holds for ψ_1. Excluding $\alpha = 0$ from H_{01} results in another null hypothesis H_{02} that allows for a unit root with a drift component. Assuming that Δy_t is stationary implies that δ_1 is restricted to $|\delta_1| < 1$ in equation (5). For $\delta_1 = 1$, Δy_t is $I\left(1\right)$ so that y_t is $I\left(2\right)$. Furthermore, when $\delta_1 = -1$, y_t has a negative unit root, and values $|\delta_1| > 1$ implies a nonstationary Δy_t process. Thus, as a consequence, problems can arise in practice if the value of δ_1 is near -1 or 1, as such values may distort the size of the tests.

3 Limit results and critical values

In this section, the necessary results for the asymptotic theory for testing H_{01} and H_{02} is derived. The least squares estimator $b_T = \left(\widehat{\delta}_1, \widehat{\delta}_2, \widehat{\delta}_3, \widehat{\phi}_1, \widehat{\phi}_2, \widehat{\alpha}, \widehat{\rho}\right)'$ of the parameters in the auxiliary model (4) has the form

$$b_T - \beta = \left(\sum_{t=1}^{T} x_t x_t'\right)^{-1} \sum_{t=1}^{T} x_t \varepsilon_t, \tag{7}$$

where $x_t = \left(\Delta y_{t-1}, (\Delta y_{t-1})^2, (\Delta y_{t-1})^3, y_{t-1}\Delta y_{t-1}, y_{t-1}(\Delta y_{t-1})^2, 1, y_{t-1}\right)'$. In order to derive the asymptotic limit distributions for the elements of (7) not previously considered in the literature, the following assumption, see 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,*

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

Allowing for weakly dependent heterogeneous data, this assumption and the theorems by Hansen (1992) are applicable to a number of different processes that typically arise in econometric applications. Assumption 1 will be used in the following theorem that contains a set of new convergence results needed in this work.

Theorem 1 *Let u_t, defined in (6), satisfy Assumption 1, and let $\{\varepsilon_t\}$ be an i.i.d. sequence with mean zero, variance σ^2, and $E|\varepsilon_t|^{6+r} < \infty$ for some $r > 0$. Define*

$$\begin{aligned}
\gamma_j &= E\left(u_t u_{t-j}\right) = \sigma^2 \sum_{s=0}^{\infty} \omega_s \omega_{s+j} &&, j = 0, 1, \ldots \\
\mu_j &= E u_t^j &&, j = 3, 4, \ldots \\
\lambda &= \sigma \sum_{j=0}^{\infty} \omega_j = \sigma\omega(1) && \\
\xi_t &= \sum_{i=0}^{t} u_i &&, t = 1, 2, \ldots, T,
\end{aligned}$$

with $\xi_0 = 0$. Then the following sums converge jointly

$$\begin{aligned}
&(a) \quad T^{-1/2} \sum_{t=1}^{T} u_{t-1}^2 \varepsilon_t \quad \Rightarrow \quad \sigma\sqrt{\mu_4} W(1) \\
&(b) \quad T^{-1/2} \sum_{t=1}^{T} u_{t-1}^3 \varepsilon_t \quad \Rightarrow \quad \sigma\sqrt{\mu_6} W(1) \\
&(c) \quad T^{-1} \sum_{t=1}^{T} \xi_{t-1} u_{t-1}^2 \varepsilon_t \quad \Rightarrow \quad \sigma\sqrt{\mu_4}\lambda \int_0^1 W(r)\, dB(r)
\end{aligned}$$

$$(d) \quad T^{-3/2}\sum_{t=1}^{T}\xi_t u_t^4 \;\Rightarrow\; \mu_4\lambda\int_0^1 W(r)\,dr$$

$$(e) \quad T^{-3/2}\sum_{t=1}^{T}\xi_t u_t^5 \;\Rightarrow\; \mu_5\lambda\int_0^1 W(r)\,dr$$

$$(f) \quad T^{-2}\sum_{t=1}^{T}\xi_t^2 u_t^3 \;\Rightarrow\; \mu_3\lambda^2\int_0^1 W^2(r)\,dr$$

$$(g) \quad T^{-2}\sum_{t=1}^{T}\xi_t^2 u_t^4 \;\Rightarrow\; \mu_4\lambda^2\int_0^1 W^2(r)\,dr$$

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

Proof See the appendix.

Consulting the rates of convergence in Theorem 1, some recent limit results from Eklund (2003), and other known results one can define the following scaling matrix

$$\Upsilon_T = diag\left(\; T^{1/2}, \;\; T^{1/2}, \;\; T^{1/2}, \;\; T, \;\; T, \;\; T^{1/2}, \;\; T \;\right). \tag{8}$$

Pre-multiplying both sides of (7) by the scaling matrix Υ_T, finite limits to the rescaled ordinary least squares estimates are given by

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

Now write $H_{01} : R\beta = r$, where $R = \begin{bmatrix} \mathbf{0} & I_6 \end{bmatrix}$, $\mathbf{0}$ is a (6×1) column vector, $\beta = (\delta_1, \delta_2, \delta_3, \phi_1, \phi_2, \alpha, \rho)'$, and $r = \begin{pmatrix} 0 & 0 & 0 & 0 & 0 & 1 \end{pmatrix}'$. An F test statistic can then be defined in the usual way as

$$F = \left(b_T - \beta\right)'\left(R\Upsilon_T\right)'\left\{s_T^2 R\Upsilon_T\left(\sum_{t=1}^{T} x_t x_t'\right)^{-1}\Upsilon_T R'\right\}^{-1} R\Upsilon_T\left(b_T - \beta\right)/k, \tag{10}$$

where $k = 6$ equals the number of restrictions, and s_T^2 is a consistent estimator of the residual variance in (4). The test statistic for hypothesis H_{02} that allows for the presence of a drift component can be defined similarly setting $k = 5$ and excluding row 5 of R and r. The two resulting test statistics will be called F_{nd} and F_d, where nd and d corresponds to '*no drift*' and '*drift*' respectively. Clearly, from Theorem 1, these test statistics do not have standard asymptotic distributions, as would be the case for stationary processes.

As the analytical limit expressions to the F test in (10) only can be computed with considerably difficulty, explicit expressions for F_{nd} and F_d are not given. The main reason for this difficulty is the problem of first obtaining and then simplifying the expressions of the inverses of the relatively large matrices that appear in equation (10): their dimensions are (5×5), (6×6) and (7×7).

Critical values for F_{nd} and F_n can, however, easily be obtained by a Monte Carlo simulation. Generating data from the null model (5) for $\delta_1 = 0$, Table 1 contains the critical values based on 1000000 replications for these statistics corresponding to sample sizes $T = 25, 50, 100, 250, 500, 5000$. Since the explicit limit expressions of the statistics are not known, critical values cannot be calculated for the asymptotic null distributions.

Table 1. Critical values for the test statistics F_{nd} and F_d, when $\delta_1 = 0$.

	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	2.49	3.06	3.65	4.49	7.00	2.70	3.36	4.05	5.02	7.85
50	2.28	2.71	3.12	3.68	5.18	2.50	3.00	3.50	4.15	5.84
100	2.22	2.60	2.97	3.43	4.59	2.45	2.90	3.33	3.90	5.28
250	2.20	2.56	2.90	3.34	4.37	2.44	2.86	3.27	3.78	5.05
500	2.20	2.55	2.89	3.31	4.33	2.44	2.86	3.27	3.77	4.99
5000	2.20	2.55	2.88	3.30	4.29	2.44	2.86	3.27	3.77	4.96

4 Small sample properties of the tests

In this section the size and the power of F_{nd}, F_d and the two corresponding ADF tests are compared. The latter are called ADF_{nd} and ADF_d respectively. A simple method to estimate p-values of the tests is also proposed in order to adjust for size distortion that is present in the tests for values of δ close to -1 or 1.

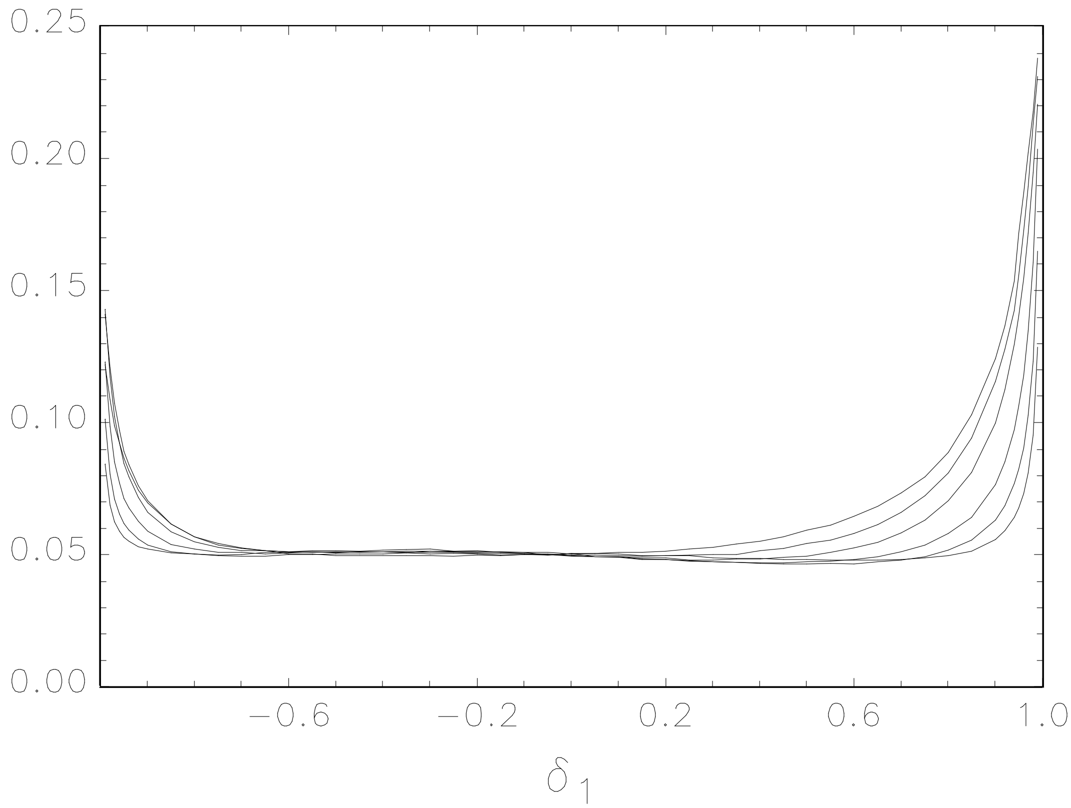
4.1 Size simulations

In order to consider the size of the tests, data are generated from model (5) under the null hypothesis, assuming $\{\varepsilon_t\} \sim nid\,(0,1)$. Using Monte Carlo simulations with 1000000 replications and critical values from Table 1, the rejection frequencies are calculated for the sample sizes $T = 25, 50, 100, 250, 500, 1000$. The nominal size for each test equals 5%. Since the null model depends on the parameter δ_1 in small samples, the size have been calculated for a number of different values of δ_1 ranging from -1 to 1. In the ADF tests, the correct number of lags (one) is assumed known. Figures 1 to 4 shows the estimated sizes of F_{nd} and F_d and the corresponding ADF tests for different sample sizes. The deviation from the nominal 5% size level decrease with the increasing sample size.

It is clear from the figures that the size of the tests is distorted when the value of δ_1 is close to -1 or 1. The reason is that the stationarity assumption of $\triangle y_t$ is violated for $|\delta_1| \geq 1$. All four tests have poor size characteristics for values of δ_1 close to 1. Furthermore, F_{nd} and F_d are also oversized, albeit less strongly, when δ_1 is close to -1. On the other hand, the ADF tests are not affected by small values of δ_1 in the same way. It is also worth noting that the two ADF tests are less distorted than their corresponding F tests, and that the two tests with a drift term, F_d and ADF_d, are not distorted as much as F_{nd} and ADF_{nd}. Furthermore, as may be expected, the deviation from the nominal 5% size level decreases with an increasing sample size. As the critical values in Table 1 are estimated for $\delta_1 = 0$, it is obvious that special attention is needed in practice if the sample size is small and the value of δ_1 is believed to be close to either -1 or 1.

In order to investigate how robust the tests are against non-normal errors, the empirical size of F_{nd} and F_d have been reestimated for errors drawn from the $t\,(6)$ and the $\chi^2\,(1) - 1$ distribution. The simulation results indicate that both F_{nd} and F_d are affected by the non-normal errors. For $t\,(6)$-distributed errors both F_{nd} and F_d show, for $T = 25$, about 1 to 2 percentage points higher size distortion at all values of δ_1 than the normal case. As the sample size increases the difference in size between the non-normal and normal size decreases. The size distortion is considerably larger for $\chi^2\,(1) - 1$ errors than it is for the $t\,(6)$-case. When $T = 25$, F_{nd} has at all values of δ_1 about 5 to 6 percentage points

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

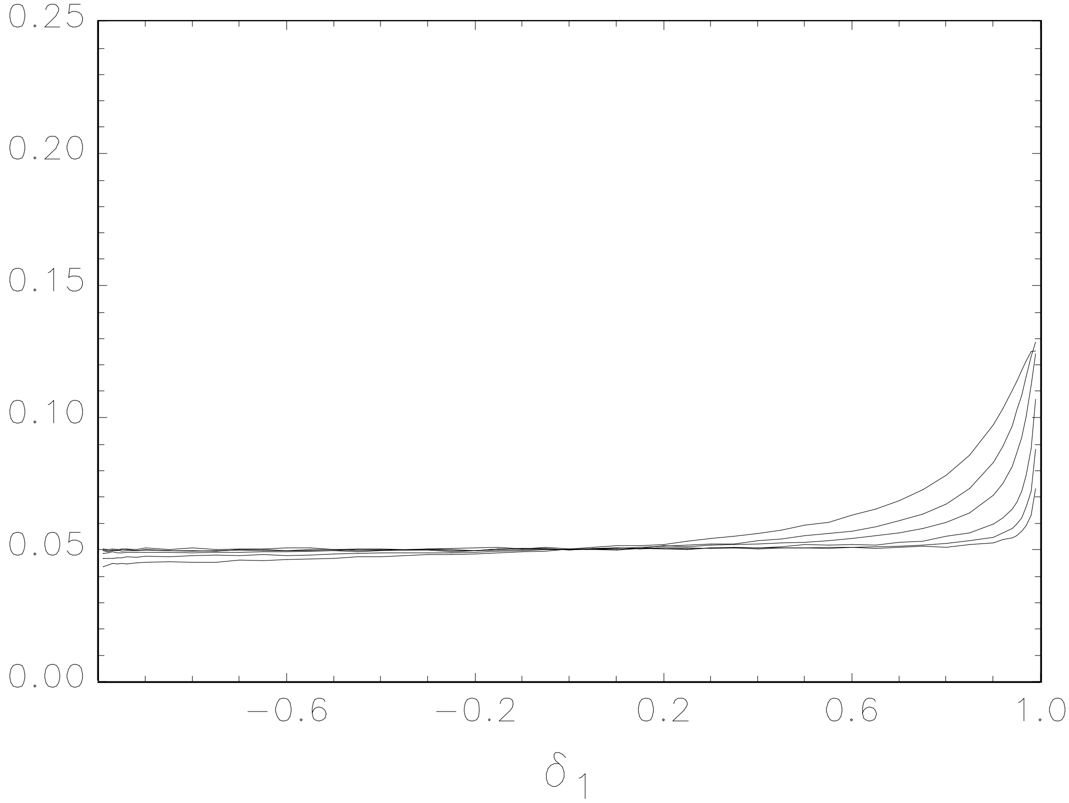


higher estimated size than it has for normal errors. At the same time F_d shows a slightly smaller increase in size, about 3 to 4 percentage points. This diference diminishes with increasing sample size and becomes negligible when $T = 500$.

4.2 Bootstrapping the p-values

As the size is distorted for values of δ_1 close to 1 for all four tests, and also close to -1 for F_{nd} and F_d, an appropriate method of adjusting the size would be desirable. One way to prevent size distortion is to calculate new critical values for some particular value of δ_1, call it $\widehat{\delta}_1$. Another method would be to base the inference on bootstrap distributions of the tests. Size distortion can be diminished by obtaining the p-values by a bootstrap. Small sample properties of the tests are then considerably improved; see the survey by Li and Maddala (1996) for more information and details on bootstrapping time series, and Caner and Hansen (2001) who, when testing the unit root against the TAR model, based the inference on a bootstrap approximation to the asymptotic null distribution of the test statistic. The same bootstrap method was also used by Eklund (2003) when testing the unit root hypothesis against the STAR model. Using the bootstrap method to estimate p-values in stationary models requires that the test statistic is pivotal. Since the analytical limit expressions of F_{nd} and F_d are not given, it is not clear whether or not the test statistics are pivotal. However, results by Li and Maddala (1996) indicate that considering pivotal statistics may not be as important in the context of unit root models as it is in stationary models.

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



From Figures 1 and 3 it has already been possible to see how δ_1 influences the size of F_{nd} and F_d in small samples. Under H_{01}, the auxiliary model (4) simplifies to $y_t = \delta_1 \Delta y_{t-1} + y_{t-1} + \varepsilon_t$. Under H_{02}, the null model has the form $y_t = \delta_1 \Delta y_{t-1} + \alpha + y_{t-1} + \varepsilon_t$. The necessary p-values can then be obtained by a model-based bootstrap. This is also the case for the ADF tests. Consider first the resampling procedure for F_{nd}. Let $\widehat{\delta}_1$ and $\widehat{D}$ be the estimates of δ_1 and the distribution D of the error ε_t. Generate the bootstrap time series

$$y_t^b = \widehat{\delta}_1 \Delta y_{t-1}^b + y_{t-1}^b + \varepsilon_t^b,\ t = 1, \ldots, T, \tag{11}$$

where ε_t^b is a random draw from $\widehat{D}$, and the time series y_t^b, $t = 1, \ldots, T$, is the resampled bootstrap series. Initial values needed for resampling can be set to sample values of the demeaned series y_t. The distribution of y_t^b is called the bootstrap distribution of the data. The value F_{nd}^b of F_{nd} is now obtained from the resampled series y_t^b. Repeating this B times yields B values of F_{nd} that constitute a realization from the distribution of F_{nd}, completely determined by $\widehat{\delta}_1$ and $\widehat{D}$. Defined by $p_T = P\left(F_{nd}^b > F_{nd}\right)$, the p-value p_T is in practice approximated by the frequency of the obtained F_{nd}^b that exceeds the observed value F_{nd}.

The resampling scheme can easily be modified to fit F_d. Including the estimate of the parameter α in the resampling model (11):

$$y_t^b = \widehat{\delta}_1 \Delta y_{t-1}^b + \widehat{\alpha} + y_{t-1}^b + \varepsilon_t^b,\ t = 1, \ldots, T. \tag{12}$$

The corresponding bootstrap distribution and the p-value p_T are then obtained in the same way as before.

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

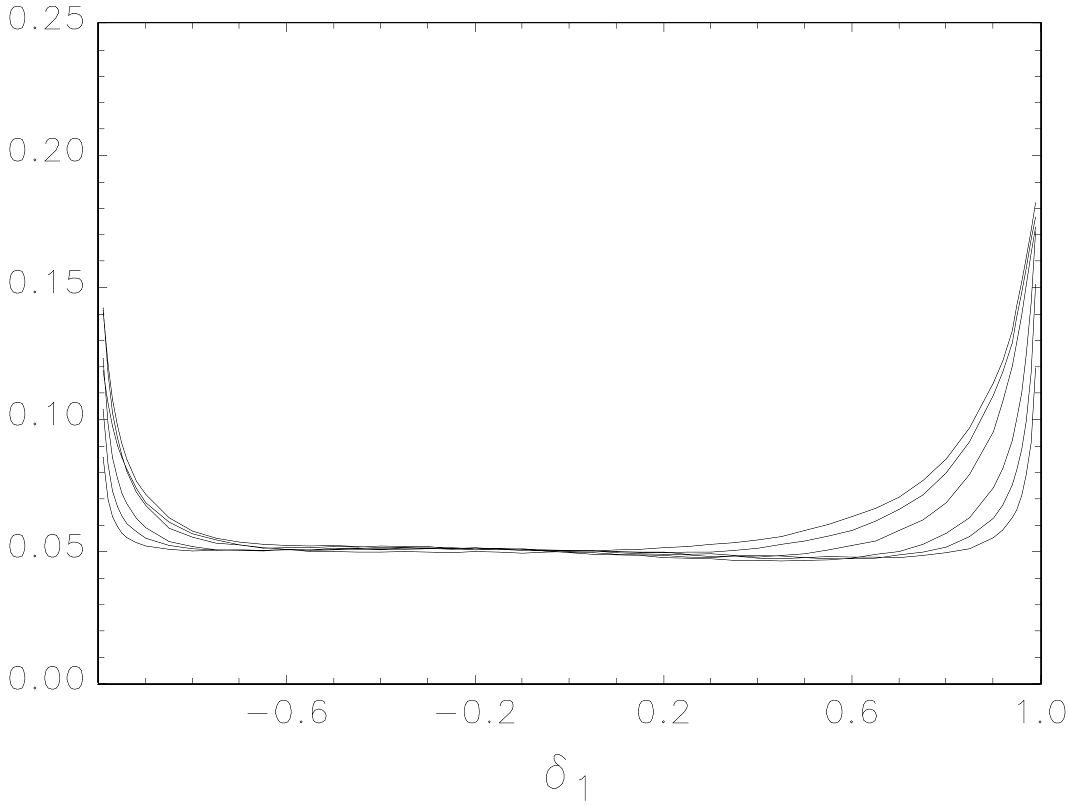
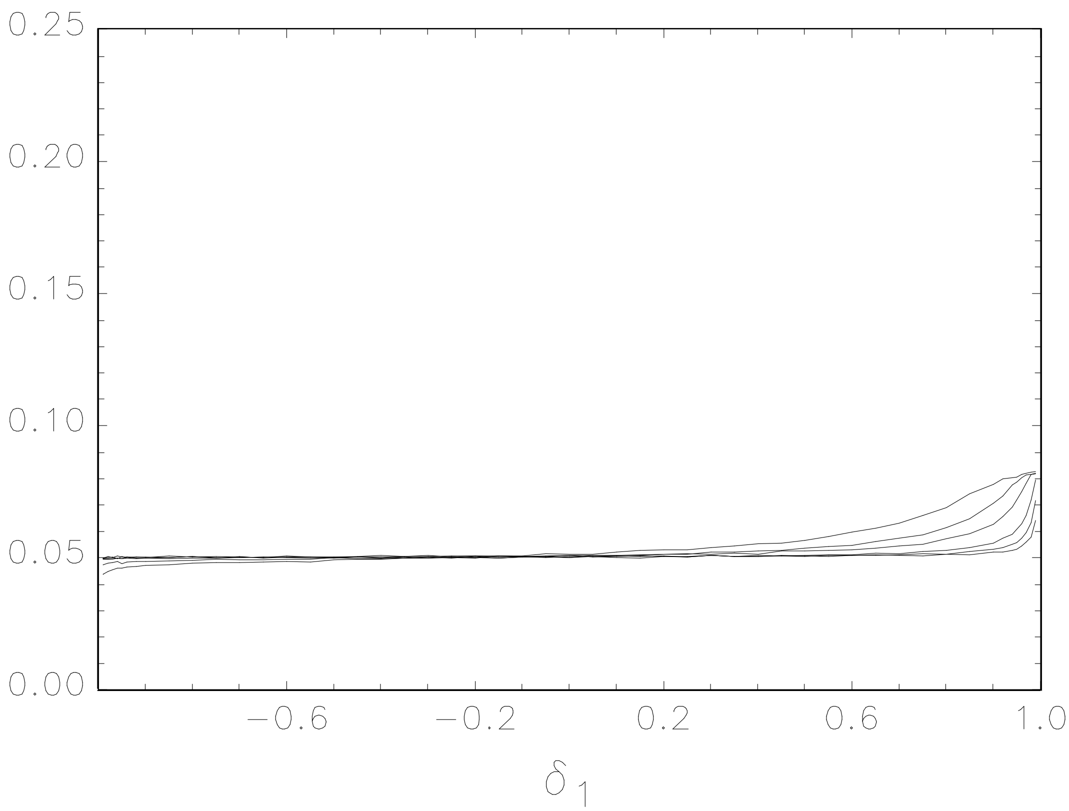


Figure 4 Size of statistic ADF_d for $T = 25, 50, 100, 250, 500, 1000$. The deviations from the nominal 5% size level, decrease with the sample size.



4.3 Power simulations

In order to investigate the power properties of the tests, data is generated from the STAR model (1), i.e. the stationary alternative model. The fact that no analytical results are available for determining the parameter combinations yielding a stationary model constitutes a problem in a simulation study. By setting the transition function F to 0 or 1, some simple guidelines can be reached about stationarity, but no general conclusions can be drawn. Therefore, an approximative method is applied to determine when the model is nonstationary. By simulation, the model is taken to be nonstationary for a specific choice of parameters, if a realization exceeds a preset boundary with t. In this study, a realization of the alternative model, y_t, is said to originate from a nonstationary process 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. For parameter choices on, or close to, the boundary between the stationary and the nonstationary regions, the approximation work less well, but probably good enough as a mean to compare the F_{nd} and F_d tests with the ADF tests.

In the data generating process (1), $\theta_0 = 0$, $\varphi_0 = 0$, and the parameters in the nonlinear function $F(\cdot)$ are $\gamma = 10$, $c_1 = 1$ and $c_2 = -0.5$. For simplicity, $\theta_1 = -\varphi_1$, and in the same way, $\psi_1 = -\psi_2$. The alternative model then equals

$$\Delta y_t = -\varphi_1 \Delta y_{t-1} - \psi_2 y_{t-1} + (\varphi_1 \Delta y_{t-1} + \psi_2 y_{t-1}) F(10, 1, -0.5, \triangle y_{t-1}) + \varepsilon_t, \qquad (13)$$

where $\varepsilon_t \sim nid(0,1)$. In (13), the magnitude of the regime shift is only determined by the two parameters φ_1 and ψ_2. Using a Monte Carlo simulation, with 50 observations, 10000 replications, and 500 bootstrap replications for estimating the p-values, the power of the tests is estimated for a number of combinations of φ_1 and ψ_2. Contour plots of the power of F_{nd} and ADF_{nd} are depicted in Figures 5 and 6. Figure 7 shows for each combination of φ_1 and ψ_2 the difference in power between the tests, expressed as power of F_{nd} minus power of ADF_{nd}. Figures 8, 9, and 10 shows the power and difference for the two other tests, F_d and ADF_d, in the same way.

From Figures 5, 6, 8 and 9, it is seen that the STAR model is stationary in a vertical band of combinations of φ_1 and ψ_2. The value of ψ_2 range from about 0.5 to 2.7, whereas φ_1 does not seem to be restricted. It appears that it can take any value between -5 and 5. Outside this vertical band the model is nonstationary. All four tests have the strongest power for values of ψ_2 in the interval between about 1.6 and 2.6. But then, the gain from using F_{nd} or F_d instead of their Dickey Fuller counterparts is smallest in this specific interval, as seen in Figures 7 and 10. It is in fact negligible there and even slightly negative for some values of $\varphi_1 > 4$. Negative gain is also found for the F_d test for values of $\varphi_1 < -4$ observed in Figure 10. The strongest gains for both F tests are found for values of ψ_2 between 0.8 and 1.0. In this interval there are two separate regions with relatively large gains in power. The single largest gain for F_{nd} compared to ADF_{nd} is 18.9% percentage points more rejections of the true alternative hypothesis, the smallest is -5.6% percentage points. The corresponding largest and smallest gain for F_d are 16.6% and -6.4% percentage points. For about 10.6% of the combinations of φ_1 and ψ_2 the gain is negative for F_{nd}. The same figure for F_d is 12.7%.

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

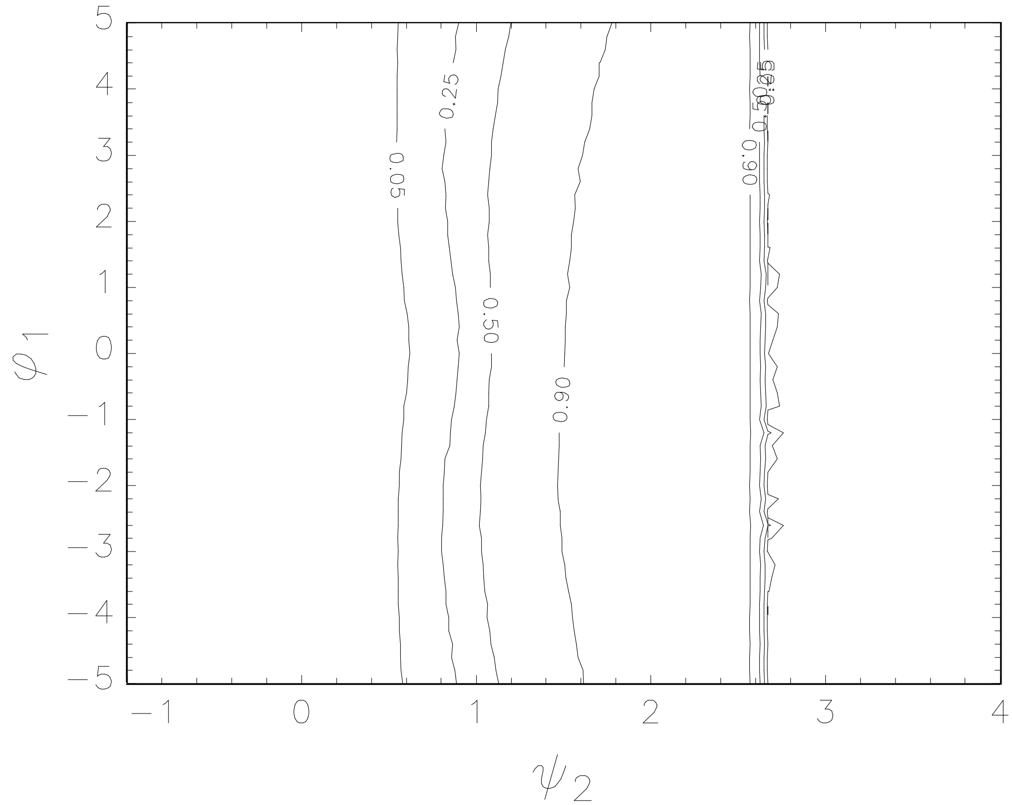


Figure 6 Power of the ADF_{nd} statistic for $T = 50$ observations.

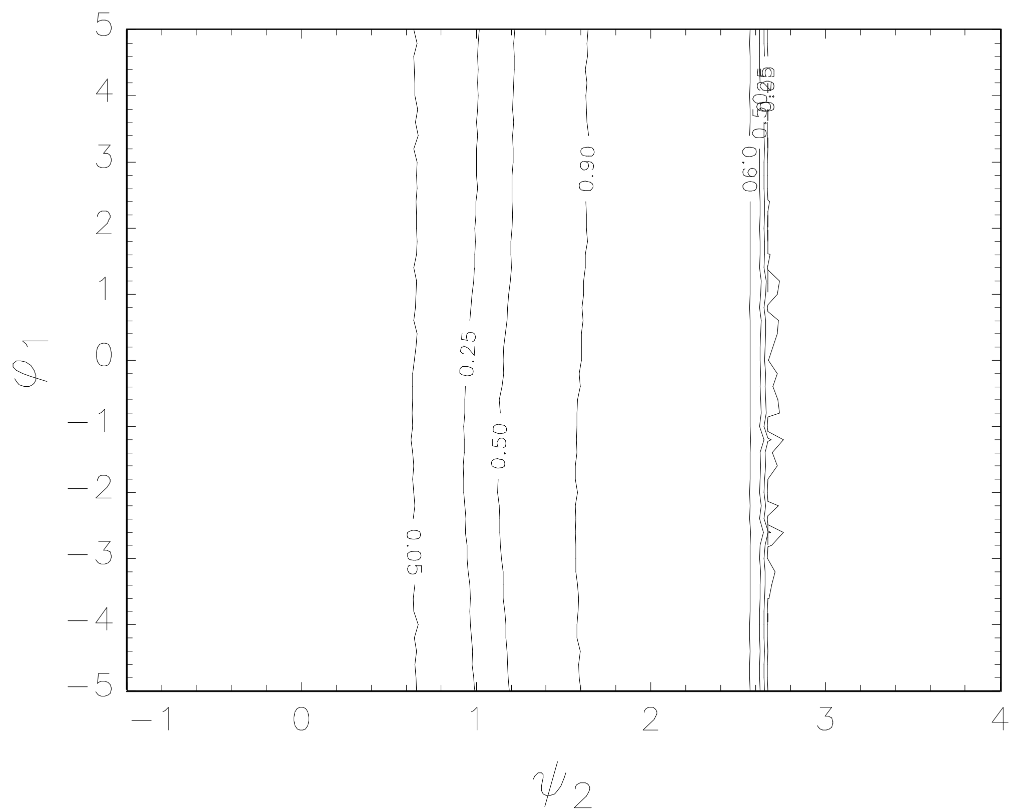


Figure 7 Difference in power, $F_{nd} - ADF_{nd}$.

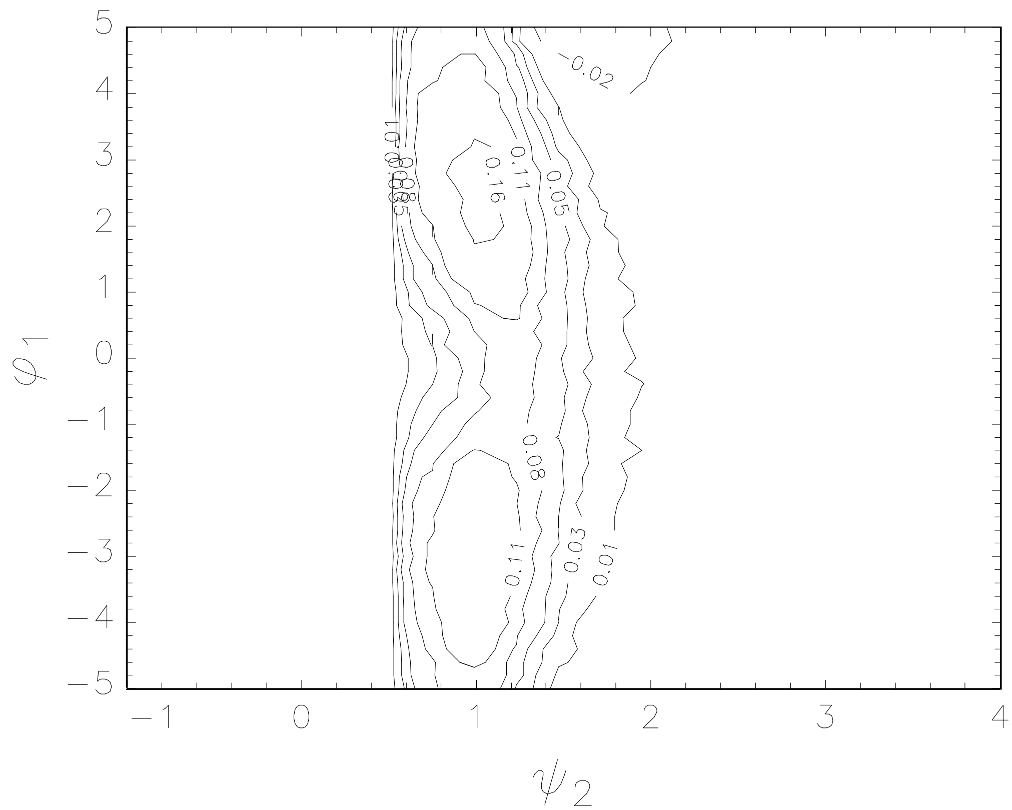


Figure 8 Power of the F_d statistic for $T = 50$ observations.

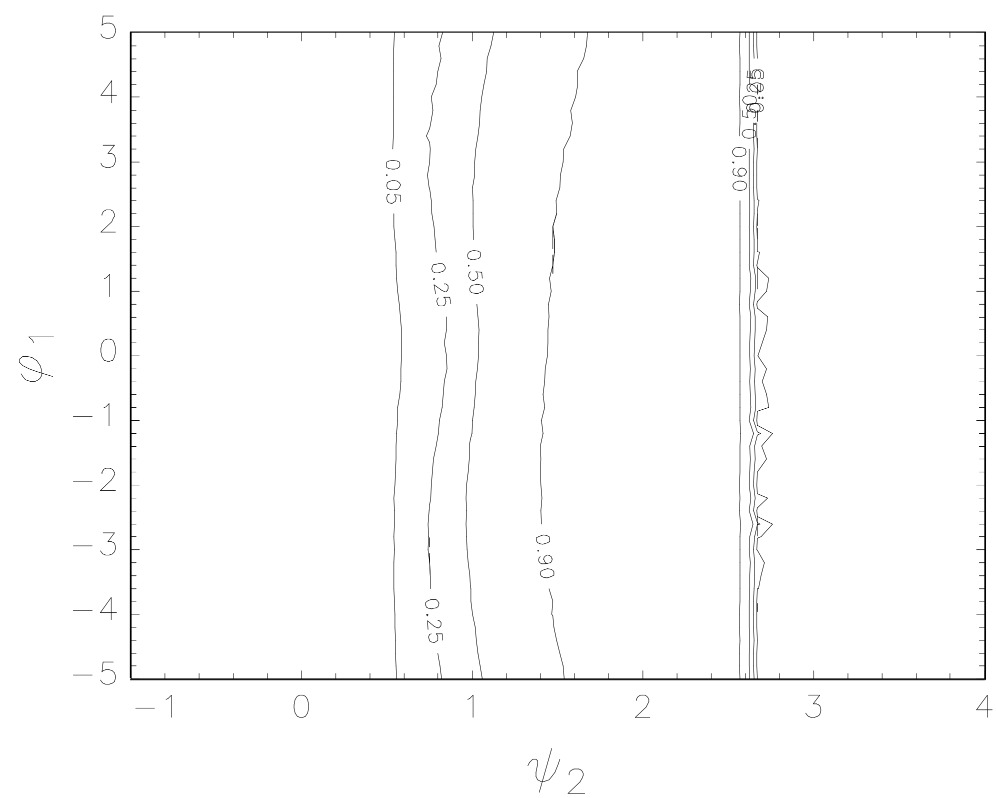
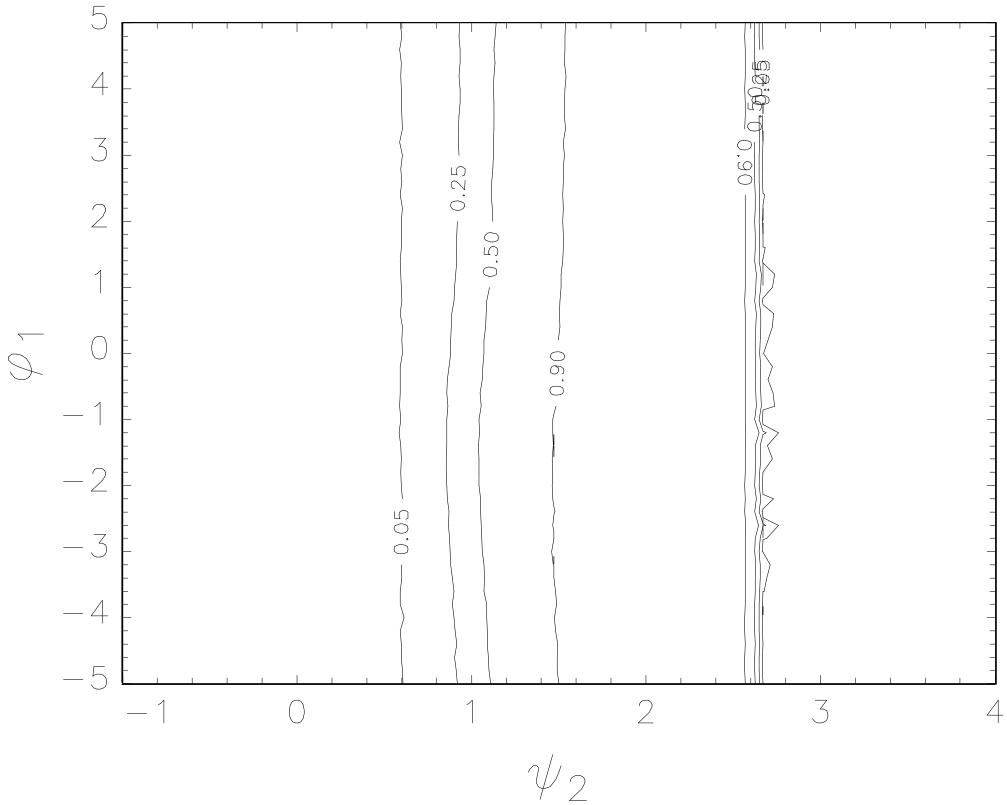


Figure 9 Power of the ADF_d statistic for $T = 50$ observations.

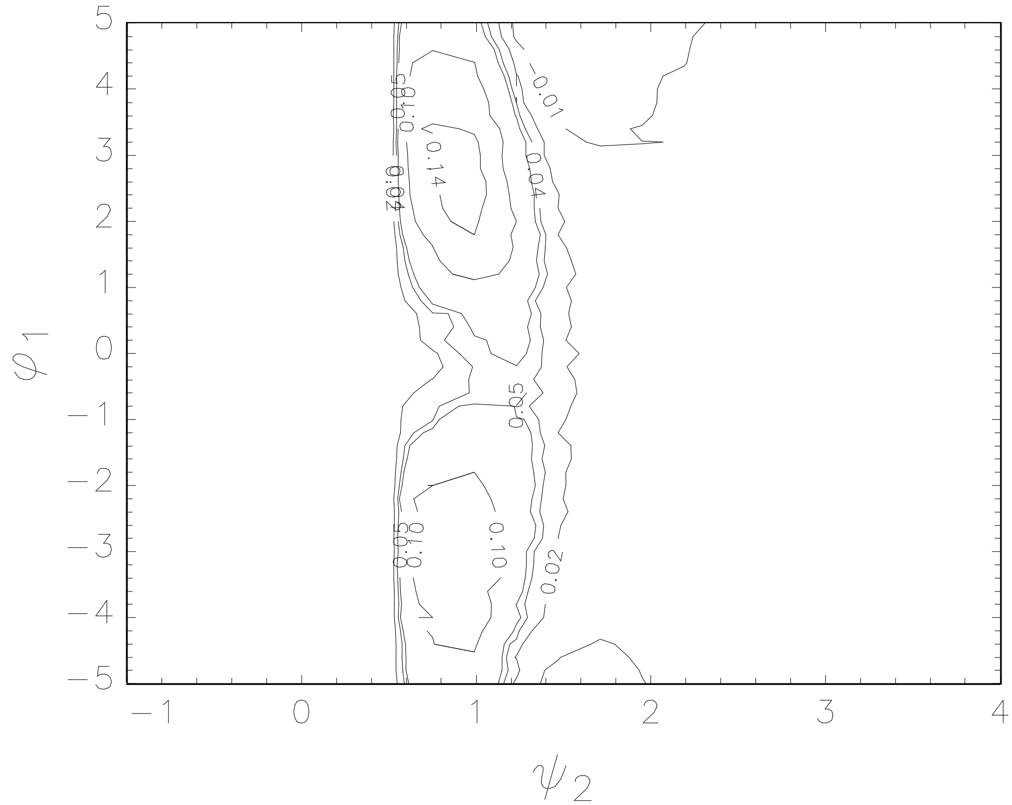


The main explanation why the gain is negative at some of the parameter combinations, is that the alternative STAR model is very close to the linear alternative model considered in the ADF test for these parameters. The auxiliary model (4) is then very close to or indistinguishable from the ADF model. This reduces the power of the tests compared to the ADF test because of the four extra parameters to be tested. As the area with positive gain dominates Figures 7 and 10 it appears safe to conclude that in general both F_{nd} and F_d have similar or higher power than the corresponding standard ADF tests when the alternative exhibit nonlinear behavior. Their use in situations where the STAR model is indeed an appropriate alternative can therefore be recommended.

5 Empirical application

The STAR model (1) is capable to characterizing asymmetric behavior in time series, and series with sudden upswings and downturns. When Δy_{t-1} is close to c_1 or c_2 in (2), the STAR model (1) behaves almost as if $\gamma = 0$, and possible as a unit root process depending on the values of θ_0 and ψ_1. On the other hand, if the difference between Δy_{t-1} and the parameter c_1, or c_2, is large, the model is nonlinear and stationary, implying a mean reverting behavior of the y_t process. These features of the second-order logistic STAR model make it attractive for modelling the real exchange rate for deviations above and below the equilibrium level. Determining the presence, or the absence, of a unit root in the real exchange rate is the main issue in testing the purchasing power parity (PPP) hypothesis. Before turning to the empirical application in this section, a brief discussion of the PPP literature will be given.

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



5.1 A short introduction to the PPP literature

In macroeconomic applications, and theory, the ability to be able to discriminate between stationary and nonstationary time series is of major importance. An area where this problem has received considerable attention is the one concerning the purchasing power parity (PPP) hypothesis, see Froot and Rogoff (1995) and Sarno and Taylor (2002) for two thorough surveys. The PPP corresponds to the idea that national price levels in different countries should tend to equal one another when expressed in a common currency. Price differences, or deviations from PPP, between two countries will be eliminated by arbitrage forces. As a consequence, for PPP to hold in the long-run, real exchange rates must be stationary. A theoretical insight into the deviations from the PPP, or long run equilibrium level, was given by Dumas (1992), who analyzed the dynamic process of the real exchange rate in spatially separated markets in the presence of proportional transactions costs. Dumas showed that deviations from PPP follow a nonlinear process that is mean reverting, and that the speed of adjustment varies directly on the magnitude of the deviation from PPP. This implies that the exchange rate will become increasingly mean reverting with the size of the deviation.

Stylized empirical findings indicate, on the other hand, high persistence in the deviations from PPP. O'Connell (1998) argued that it appears as if large deviations from PPP can be more persistent than small deviations, and that market frictions alone cannot account for the difficulty of detecting mean reversion in post-Bretton Woods real exchange rates. Lothian and Taylor (1996) argued that the high persistence of deviations from PPP, together with the low power of standard unit root tests, may account for the widespread failure of empirical tests to support long-run PPP. By considering observations over a sample period of two centuries they where able to reject the unit root hypothesis in favor of a mean reverting process for two real exchange rates, U.S. dollar-pound sterling and French franc-pound sterling. Michael, Nobay and Peel (1997) rejected the linear framework in favor of an exponential smooth transition autoregressive (ESTAR) process, thus providing evidence of mean reverting behavior for PPP deviations. Taylor, Peel and Sarno (2001) considered a multivariate linear unit root test and provided empirical evidence that four major real bilateral dollar exchange rates are well characterized by nonlinear mean reverting processes, based on the ESTAR model.

Sarno and Taylor (2002) concluded that, at the present time, the long-run PPP seem to have some validity, at least for the major exchange rates, even though a number of problems have to be analyzed and resolved.

5.2 Testing the PPP hypothesis in practice

The real exchange rate for country i versus country j is expressed in logarithmic form as

$$z_t = p_{it} - s_t - p_{jt}, \quad t = 1, \ldots, T, \tag{14}$$

where s_t is the logarithm of the nominal exchange rate between country i and j expressed in country i's currency per country j's currency, and p_{it} and p_{jt} denote the logarithms of the consumer price index (CPI) for country i and j, respectively. The series z_t can then be interpreted as a measure of the deviation from the long-run steady state or PPP.

The real exchange rates are constructed for sixteen countries from the CPI series and the exchange rates defined as the price of US dollars in the currency of each home country. The data, consumer price indices and nominal exchange rates, are obtained from EcoWin and covers the following sixteen countries; Austria, Belgium, Canada, Denmark, Finland, France, Germany, Italy, Japan, the Netherlands, Norway, Spain, Sweden, Switzerland, United Kingdom and USA. The sample consists of monthly observations from January 1960 to October 2002, except for five countries for which some early observations are missing. The sample for Germany only includes observations from 1968:1 to 2002:10, Japan from 1970:1 to 2002:10, the Netherlands from 1960:4 to 2002:10, Spain from 1961:1 to 2002:10, and, finally, Switzerland from 1974:1 to 2002:10. In total, 120 real exchange rates are constructed. For pairs of countries that both joined the common currency in January 1999, data only up to December 1998 is used. Otherwise the largest available sample size is used for all pairs, which means that the length of the series varies between 344 and 512.

Tables 2 and 3 show the estimated p-values of F_{nd} and F_d based on 10000 replications. P-values less than 0.05 are printed in boldface. The results in Table 2 indicate that the random walk without drift can be rejected for 31 out of the 120 real exchange rates considered. When including a drift term under the null, the number of rejections increase to 44, as seen in Table 3. The corresponding number of rejections for ADF_{nd} and ADF_d are 7 and 23, respectively. The exchange rates for which the unit root is rejected can thus be considered stationary, giving support to the PPP hypothesis.

The test results both support and disagree with earlier work. Results in line with earlier studies, as in Bec et al. (2002), show that the PPP hypothesis can be supported for the real exchange rates BEL/GER, FIN/GER, FRA/GER, ITA/GER and UK/GER,

Table 2 Estimated p-values to the F_{nd} test. Bold values corresponds to a significant test at the 5% level.

vs	BEL	CAN	DEN	FIN	FRA	GER	ITA	JAP	NL	NOR	SPA	SWE	SWI	UK	USA
AUS	0.23	0.091	0.054	0.34	0.60	**0.0006**	0.089	0.81	0.32	0.76	0.26	0.29	**0.022**	**0.033**	**0.0021**
BEL	-	**0.0034**	**0.040**	0.44	0.51	**0.020**	0.13	0.76	0.34	0.63	0.20	0.69	**0.0068**	0.27	**0.0018**
CAN		-	0.072	0.37	0.087	0.10	0.46	0.49	0.10	0.093	0.42	0.45	0.12	0.42	0.13
DEN			-	0.23	0.70	**0.048**	0.19	0.63	0.051	**0.0058**	0.59	0.18	**0.030**	0.062	**0.0001**
FIN				-	0.59	0.069	0.43	0.31	0.41	0.15	**0.013**	0.48	**0.012**	**0.0058**	0.49
FRA					-	**0.044**	0.30	0.48	0.48	0.38	0.39	0.96	**0.0001**	0.41	0.13
GER						-	**0.039**	0.88	0.091	0.33	**0.038**	0.52	**0.0056**	**0.033**	**0.0046**
ITA							-	0.26	0.11	0.24	0.15	0.16	**0.0004**	0.10	0.63
JAP								-	0.85	0.69	0.23	0.17	0.77	0.41	0.70
NL									-	0.074	0.44	0.39	**0.0065**	**0.043**	**0.0018**
NOR										-	0.30	0.38	0.15	0.052	**0.0014**
SPA											-	**0.017**	**0.0089**	0.10	0.52
SWE												-	0.10	0.24	0.24
SWI													-	**0.012**	0.15
UK														-	0.23

Table 3 Estimated p-values to the F_d test. Bold values corresponds to a significant test at the 5% level.

vs	BEL	CAN	DEN	FIN	FRA	GER	ITA	JAP	NL	NOR	SPA	SWE	SWI	UK	USA
AUS	0.12	**0.036**	**0.015**	0.21	0.89	**0.0003**	**0.037**	0.72	0.17	0.81	0.31	0.17	**0.0051**	**0.0050**	**0.0004**
BEL	-	**0.0012**	**0.027**	0.32	0.46	**0.011**	0.065	0.74	0.25	0.50	0.11	0.87	**0.0004**	0.12	**0.0011**
CAN		-	**0.045**	0.30	0.066	**0.031**	0.35	0.41	0.064	0.063	0.31	0.36	**0.036**	0.32	0.058
DEN			-	0.14	0.65	**0.014**	0.069	0.61	**0.035**	**0.0031**	0.83	0.096	**0.0028**	**0.011**	**0.0002**
FIN				-	0.49	**0.028**	0.25	0.17	0.30	0.077	**0.0084**	0.38	**0.0052**	**0.0043**	0.40
FRA					-	**0.023**	0.15	0.43	0.38	0.27	0.27	0.93	**0.0000**	0.31	0.084
GER						-	**0.033**	0.73	**0.020**	0.18	**0.027**	0.85	**0.0019**	**0.0023**	**0.0004**
ITA							-	0.20	**0.036**	0.11	0.12	0.064	**0.0001**	0.059	0.78
JAP								-	0.87	0.68	0.13	0.073	0.65	0.32	0.71
NL									-	0.059	0.50	0.28	**0.0004**	**0.0086**	**0.0010**
NOR										-	0.19	0.24	0.059	**0.0092**	**0.0007**
SPA											-	**0.014**	**0.0046**	0.062	0.54
SWE												-	**0.030**	0.17	0.16
SWI													-	**0.0005**	0.067
UK														-	0.12

and as in Taylor et al. (2001), for USA/GER. On the other hand, Bec et al. (2002) rejected the unit root hypothesis for another six exchange rates, Taylor et al. (2001) for another three. Conversely, for five of the significant real exchange rates considered in this paper, Bec et al. (2002) were unable to reject the unit root: CAN/BEL, USA/BEL, UK/FIN, USA/GER, and GER/CAN. They consider a smaller sample size, only from 1973:9 to 2000:9, which could explain some of the differences with this paper. Otherwise, as a whole, the proportion of rejected exchange rates is almost the same in this paper as in Bec et al. (2002).

The results in this paper can be viewed as a complement to earlier studies. Under the assumption that the PPP hypothesis holds, the two tests presented here, F_{nd} and F_d, have low power discriminating a random walk from a stationary nonlinear process at sample sizes available for the tests. The same fact holds for most univariate unit root tests. However, Taylor et al. (2001) noted that, somewhat paradoxically, the failure to reject a unit root may indicate that the real exchange rate has, on average, been relatively close to equilibrium, rather than implying that no such long-run equilibrium exists. In their estimated ESTAR model, the real exchange rate will be closer to a unit root process the closer it is to its long-run equilibrium.

6 Conclusions

In this paper, two F-type tests are proposed for the joint unit root and linearity hypothesis against a second-order logistic smooth transition autoregressive (STAR) model. The tests allows one to discriminate between nonstationary and stationary time series. This is important in statistical analysis and empirical applications. Some new limit results, extending earlier work, and critical values for the F-tests are presented. As the alternative model is well suited for modelling real exchange rates, the two tests are applied to a number of real exchange rates as an illustration. The test results complement earlier studies. Support to the purchasing power parity (PPP) hypothesis is at any rate provided for 44 out of 120 real exchange rates considered in this work.

References

[1] Bec, F., Salem, M. B., and Carrasco, M., (2002), "The purchasing power parity puzzle: Evidence from a LSTAR specification", Working Paper, Université de Paris I.

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

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

[4] Dumas, B., (1992), "Dynamic equilibrium and the real exchange rate in a spatially separated world", *Review of Financial Studies*, 5, 153-180.

[5] Eklund, B., (2003), "Testing the unit root hypothesis against the logistic smooth transition autoregressive model", Working Paper Series in Economics and Finance, No. 546, Stockholm School of Economics.

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

[7] Froot, K. A., and Rogoff, K., (1995), "Perspectives on PPP and long-run real exchange rates", in G. Grossman and K. Rogoff (eds.), *Handbook of International Economics*, North-Holland, Amsterdam, 1647-1688.

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

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

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

[11] 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.

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

[13] 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.

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

[15] Lothian, J. R., and Taylor, M. P., (1996), "Real exchange rate behavior: The recent float from the perspective of the past two centuries, *Journal of Political Economy*, 104, 488-509.

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

[17] Michael, P., Nobay, R. A., and Peel, D. A., (1997), "Transactions costs and nonlinear adjustment in real exchange rates: an empirical investigation", *Journal of Political Economy*, 105, 862-879.

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

[19] O'Connell, P. G. J., (1998), "Market frictions and real exchange rates", *Journal of International Money and Finance*, 17, 71-95.

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

[21] 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.

[22] Sarno, L., and Taylor, M. P., (2002), "Purchasing power parity and the real exchange rate", IMF Staff Papers, 49, No. 1.

[23] Taylor, M. P., Peel, D. A., and Sarno, L., (2001), ”Nonlinear mean-reversion in real exchange rates: toward a solution to the purchasing power parity puzzles”, *International Economic Review*, 42, 1015-1042.

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

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

[26] 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.

Appendix A

Proof of Theorem 1

(a) Let u_t, defined in (6), satisfy Assumption 1 and let $\sum_{j=0}^{\infty} |\omega_j| < \infty$. Set $X_t = u_{t-1}^2 \varepsilon_t$. X_t is then a martingale difference sequence. Furthermore,

$$EX_t^2 = \sigma^2 E u_t^4 \tag{A.1}$$
$$E\left|X_t\right|^{2+r} = E\left|u_{t-1}^2 \varepsilon_t\right|^{2+r} = E\left|u_t\right|^{4+2r} E\left|\varepsilon_t\right|^{2+r} < \infty.$$

so that $\{X_t\}$ is a uniformly integrable sequence. Now,

$$T^{-1}\sum_{t=1}^{T} X_t^2 = T^{-1}\sum_{t=1}^{T} u_{t-1}^4 \varepsilon_t^2 = T^{-1}\sum_{t=1}^{T} u_{t-1}^4 \left(\varepsilon_t^2 - \sigma^2\right) + T^{-1}\sigma^2 \sum_{t=1}^{T} u_{t-1}^4. \tag{A.2}$$

Set $Z_t = u_{t-1}^4 (\varepsilon_t^2 - \sigma^2)$. It follows that $\{Z_t\}$ is a martingale difference sequence. Furthermore,

$$E\left|Z_t\right|^{1+r} = E\left|u_{t-1}^4 \left(\varepsilon_t^2 - \sigma^2\right)\right|^{1+r} = E\left|u_t\right|^{4+4r} E\left|\varepsilon_t^2 - \sigma^2\right|^{1+r} < \infty \tag{A.3}$$

implies that

$$T^{-1}\sum_{t=1}^{T} Z_t = T^{-1}\sum_{t=1}^{T} u_{t-1}^4 \left(\varepsilon_t^2 - \sigma^2\right) \xrightarrow{p} 0. \tag{A.4}$$

Also,

$$T^{-1}\sigma^2 \sum_{t=1}^{T} u_{t-1}^4 \xrightarrow{p} \sigma^2 E u_t^4 \tag{A.5}$$

which gives the result

$$T^{-1/2}\sum_{t=1}^{T} u_{t-1}^2 \varepsilon_t \Rightarrow \sigma\sqrt{E u_t^4} W(1). \tag{A.6}$$

(*b*) Let u_t and ε_t be as in (*a*). Set $X_t = u_{t-1}^3 \varepsilon_t$. Then $\{X_t\}$ is a martingale difference sequence. Furthermore,

$$EX_t^2 = \sigma^2 E u_t^6 \tag{A.7}$$
$$E\left|X_t\right|^{2+r} = E\left|u_{t-1}^3 \varepsilon_t\right|^{2+r} = E\left|u_t\right|^{6+3r} E\left|\varepsilon_t\right|^{2+r} < \infty$$

so that $\{X_t\}$ is uniformly integrable. Now,

$$T^{-1}\sum_{t=1}^{T} X_t^2 = T^{-1}\sum_{t=1}^{T} u_{t-1}^6 \varepsilon_t^2 = T^{-1}\sum_{t=1}^{T} u_{t-1}^6\left(\varepsilon_t^2 - \sigma^2\right) + T^{-1}\sigma^2\sum_{t=1}^{T} u_{t-1}^6. \tag{A.8}$$

Set $Z_t = u_{t-1}^6\left(\varepsilon_t^2 - \sigma^2\right)$. Then $\{Z_t\}$ is a martingale difference sequence. For $r > 0$,

$$E\left|Z_t\right|^{1+r} = E\left|u_{t-1}^6\left(\varepsilon_t^2 - \sigma^2\right)\right|^{1+r} = E\left|u_t\right|^{6+6r} E\left|\varepsilon_t^2 - \sigma^2\right|^{1+r} < \infty \tag{A.9}$$

which implies

$$T^{-1}\sum_{t=1}^{T} Z_t = T^{-1}\sum_{t=1}^{T} u_{t-1}^6\left(\varepsilon_t^2 - \sigma^2\right) \xrightarrow{p} 0. \tag{A.10}$$

Furthermore,

$$T^{-1}\sigma^2\sum_{t=1}^{T} u_{t-1}^6 \xrightarrow{p} \sigma^2 E u_t^6 \tag{A.11}$$

which gives the desired result

$$T^{-1/2}\sum_{t=1}^{T} u_{t-1}^3 \varepsilon_t \Rightarrow \sigma\sqrt{E u_t^6}\, W(1). \tag{A.12}$$

(*c*) Let $v_t = \left(u_t, u_{t-1}^2 \varepsilon_t\right)'$, $V_t = \sum_{i=1}^{t} v_i$ and $V_0 = 0$. Then, since $T^{-1/2} V_T \Longrightarrow \left(\lambda W(1), \sigma\sqrt{E u_t^4}\, W(1)\right)'$, Hansen (1992), Theorem 4.1, states that the elements of the sum

$$T^{-1}\sum_{t=1}^{T} V_{t-1} v_t' = T^{-1}\sum_{t=1}^{T}\begin{bmatrix} \xi_{t-1} \\ \sum_{i=1}^{t-1} u_{i-1}^2 \varepsilon_i \end{bmatrix}\begin{bmatrix} u_t & u_{t-1}^2 \varepsilon_t \end{bmatrix} \tag{A.13}$$
$$= T^{-1}\sum_{t=1}^{T}\begin{bmatrix} \xi_{t-1} u_t & \xi_{t-1} u_{t-1}^2 \varepsilon_t \\ u_t \sum_{i=1}^{t-1} u_{i-1}^2 \varepsilon_i & u_{t-1}^2 \varepsilon_t \sum_{i=1}^{t-1} u_{i-1} \varepsilon_i \end{bmatrix}$$

will converge weakly to some stochastic integrals. In particular,

$$T^{-1}\sum_{t=1}^{T} \xi_{t-1} u_{t-1}^2 \varepsilon_t \Rightarrow \sigma\sqrt{E u_t^4}\,\lambda\int_0^1 W(r)\, dB(r) + \Lambda_{1,2} \tag{A.14}$$

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

$$\begin{aligned}\Lambda &= \lim_{T\longrightarrow\infty} T^{-1} \sum_{i=1}^{T} \sum_{j=i+1}^{\infty} E\upsilon_i \upsilon_j' \\ &= \lim_{T\longrightarrow\infty} T^{-1} \sum_{i=1}^{T} \sum_{j=i+1}^{\infty} E \begin{bmatrix} u_i \\ u_{i-1}^2 \varepsilon_i \end{bmatrix} \begin{bmatrix} u_j & u_{j-1}^2 \varepsilon_j \end{bmatrix} \\ &= \lim_{T\longrightarrow\infty} T^{-1} \sum_{i=1}^{T} \sum_{j=i+1}^{\infty} E \begin{bmatrix} u_i u_j & u_i u_{j-1}^2 \varepsilon_j \\ u_j u_{i-1}^2 \varepsilon_i & u_{i-1}^2 u_{j-1}^2 \varepsilon_j \varepsilon_i \end{bmatrix}.\end{aligned} \tag{A.15}$$

Then, since $u_i u_{j-1}^2$ and ε_j are independent for $j \geq i+1$,

$$\begin{aligned}\Lambda_{1,2} &= \lim_{T\longrightarrow\infty} T^{-1} \sum_{i=1}^{T} \sum_{j=i+1}^{\infty} E\left(u_i u_{j-1}^2 \varepsilon_j\right) \\ &= \lim_{T\longrightarrow\infty} T^{-1} \sum_{i=1}^{T} \sum_{j=i+1}^{\infty} E\left(u_i u_{j-1}^2\right) E\left(\varepsilon_j\right) = 0.\end{aligned} \tag{A.16}$$

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

(d) Consider

$$T^{-1} \sum_{t=1}^{T} \xi_t u_t^4 = T^{-1} \sum_{t=1}^{T} \xi_t \left(u_t^4 - E u_t^4\right) + T^{-1} \sum_{t=1}^{T} \xi_t E u_t^4. \tag{A.17}$$

Now, since $\xi_t = \xi_{t-1} - u_t$, the first sum on the right-hand side of (A.17) equals

$$T^{-1} \sum_{t=1}^{T} \xi_t \left(u_t^4 - E u_t^4\right) = T^{-1} \sum_{t=1}^{T} \xi_{t-1} \left(u_t^4 - E u_t^4\right) + T^{-1} \sum_{t=1}^{T} u_t \left(u_t^4 - E u_t^4\right) \tag{A.18}$$

where the last term is $O_p(1)$. Now let $\upsilon_t = (u_t, u_t^4 - E u_t^4)'$, $V_t = \sum_{i=1}^{t} \upsilon_i$ and $V_0 = 0$. Then, from Hansen (1992), Theorem 4.1, it follows that the sum $T^{-1} \sum_{t=1}^{T} V_{t-1} \upsilon_t'$ converges weakly to a stochastic integral. Therefore, as a consequence, $T^{-1} \sum_{t=1}^{T} \xi_{t-1} (u_t^4 - E u_t^4) = O_p(1)$ and

$$T^{-3/2} \sum_{t=1}^{T} \xi_t u_t^4 = T^{-3/2} E\left(u_t^4\right) \sum_{t=1}^{T} \xi_t + o_p(1) \Rightarrow E u_t^4 \lambda \int_0^1 W(r)\, dr. \tag{A.19}$$

(e) First consider

$$T^{-1} \sum_{t=1}^{T} \xi_t u_t^5 = T^{-1} \sum_{t=1}^{T} \xi_t \left(u_t^5 - E u_t^5\right) + T^{-1} \sum_{t=1}^{T} \xi_t E u_t^5. \tag{A.20}$$

Now, since $\xi_t = \xi_{t-1} - u_t$, the first sum on the right-hand side of (A.20) equals

$$T^{-1}\sum_{t=1}^{T}\xi_t\left(u_t^5 - Eu_t^5\right) = T^{-1}\sum_{t=1}^{T}\xi_{t-1}\left(u_t^5 - Eu_t^5\right) + T^{-1}\sum_{t=1}^{T}u_t\left(u_t^5 - Eu_t^5\right) \quad \text{(A.21)}$$

where the last term is $O_p(1)$ as before. Now let $\upsilon_t = \left(u_t, u_t^5 - Eu_t^5\right)'$, $V_t = \sum_{i=1}^{t}\upsilon_i$ and $V_0 = 0$. Then, again using Hansen (1992), Theorem 4.1, one can conclude that the sum $T^{-1}\sum_{t=1}^{T}V_{t-1}\upsilon_t'$ converges weakly to a stochastic integral. Thus, $T^{-1}\sum_{t=1}^{T}\xi_{t-1}\left(u_t^5 - Eu_t^5\right) = O_p(1)$ and

$$T^{-3/2}\sum_{t=1}^{T}\xi_t u_t^5 = T^{-3/2}E\left(u_t^5\right)\sum_{t=1}^{T}\xi_t + o_p(1) \Rightarrow Eu_t^5\lambda\int_0^1 W(r)\,dr. \quad \text{(A.22)}$$

(*f*) As a starting-point, consider the sum

$$T^{-3/2}\sum_{t=1}^{T}\xi_{t-1}^2 u_t^3 = T^{-3/2}\sum_{t=1}^{T}\xi_{t-1}^2\left(u_t^3 - Eu_t^3\right) + T^{-3/2}\sum_{t=1}^{T}\xi_{t-1}^2 Eu_t^3 \quad \text{(A.23)}$$

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

$$T^{-3/2}\sum_{t=1}^{T}\xi_{t-1}^2\left(u_t^3 - Eu_t^3\right) = O_p(1) \quad \text{(A.24)}$$

and

$$T^{-2}\sum_{t=1}^{T}\xi_{t-1}^2 u_t^3 = T^{-2}\sum_{t=1}^{T}\xi_{t-1}^2 Eu_t^3 + o_p(1) \Rightarrow E\left(u_t^3\right)\lambda^2\int_0^1 W^2(r)\,dr. \quad \text{(A.25)}$$

What remains to show is that $\underset{T\longrightarrow\infty}{plim}\, T^{-2}\sum_{t=1}^{T}\xi_t^2 u_t^3 = \underset{T\longrightarrow\infty}{plim}\, T^{-2}\sum_{t=1}^{T}\xi_{t-1}^2 u_t^3$. It is easily shown that

$$\xi_t^2 - \xi_{t-1}^2 = 2\xi_{t-1}u_t + u_t^2. \quad \text{(A.26)}$$

The difference between the two sums in (A.25) is given by

$$\sum_{t=1}^{T}\xi_t^2 u_t^3 - \sum_{t=1}^{T}\xi_{t-1}^2 u_t^3 = \sum_{t=1}^{T}\left(\xi_t^2 - \xi_{t-1}^2\right)u_t^3 = 2\sum_{t=1}^{T}\xi_{t-1}u_t^4 + \sum_{t=1}^{T}u_t^5, \quad \text{(A.27)}$$

where the first sum on the right-hand side is $O_p\left(T^{3/2}\right)$ from (*d*) above, and the second sum is $O_p(T)$. This implies that $T^{-2}\sum_{t=1}^{T}\left(\xi_t^2 - \xi_{t-1}^2\right)u_t^3 = o_p\left(T^2\right)$, and

$$T^{-2}\sum_{t=1}^{T}\xi_t^2 u_t^3 \Rightarrow E\left(u_t^3\right)\lambda^2\int_0^1 W^2(r)\,dr \quad \text{(A.28)}$$

as desired.

(g) Begin by considering the sum

$$T^{-3/2}\sum_{t=1}^{T}\xi_{t-1}^{2}u_t^4 = T^{-3/2}\sum_{t=1}^{T}\xi_{t-1}^{2}\left(u_t^4 - Eu_t^4\right) + T^{-3/2}\sum_{t=1}^{T}\xi_{t-1}^{2}Eu_t^4 \tag{A.29}$$

and let $\upsilon_t = \left(u_t, u_t^4 - Eu_t^4\right)'$, V_t and V_0 be as before. Let Assumption 1 hold with $\eta = 3$. It then follows from Hansen (1992), Theorem 4.2, that the sum $T^{-3/2}\sum_{t=1}^{T}\left(V_{t-1}\otimes V_{t-1}\right)\upsilon_t'$ converges weakly to a stochastic integral. This implies that

$$T^{-3/2}\sum_{t=1}^{T}\xi_{t-1}^{2}\left(u_t^4 - Eu_t^4\right) = O_p\left(1\right) \tag{A.30}$$

and

$$T^{-2}\sum_{t=1}^{T}\xi_{t-1}^{2}u_t^4 = T^{-2}\sum_{t=1}^{T}\xi_{t-1}^{2}Eu_t^4 + o_p\left(1\right) \Rightarrow Eu_t^4\lambda^2\int_0^1 W^2\left(r\right)dr. \tag{A.31}$$

Using the same idea as in the proof of (f),

$$\sum_{t=1}^{T}\left(\xi_t^2 - \xi_{t-1}^2\right)u_t^4 = 2\sum_{t=1}^{T}\xi_{t-1}u_t^5 + \sum_{t=1}^{T}u_t^6, \tag{A.32}$$

where the first sum on the right-hand side is $O_p\left(T^{3/2}\right)$ from (e) above, and the second one is $O_p\left(T\right)$. The result then follows since

$$T^{-2}\sum_{t=1}^{T}\xi_t^2u_t^4 = T^{-2}\sum_{t=1}^{T}\xi_{t-1}^{2}u_t^4 + o_p\left(T\right) \Rightarrow Eu_t^4\lambda^2\int_0^1 W^2\left(r\right)dr. \tag{A.33}$$

This concludes the proof of Theorem 1. ■