Testing Parameter Constancy in Unit Root Autoregressive Models against Continuous Change

Changli He* and Rickard Sandberg†

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

SSE/EFI Working Paper Series in Economics and Finance No. 579

January 2005

Abstract

In this paper we derive tests for parameter constancy when the data generating process is non-stationary against the hypothesis that the parameters of the model change smoothly over time. To obtain the asymptotic distributions of the tests we generalize many theoretical results, as well as new are introduced, in the area of unit roots . The results are derived under the assumption that the error term is a strong mixing. Small sample properties of the tests are investigated, and in particular, the power performances are satisfactory.

JEL classification: C12; C22; C52

Key words: Parameter constancy; LSTAR; unit root; Brownian motion; strong mixing

---

*e-mail:changli.he@hhs.se

†e-mail:rickard.sandberg@hhs.se

Acknowledgements: This research has been supported by the Jan Wallander's and Tom Hedelius' Foundation, Grant No. J02-35. A first version of the paper was presented at a seminar at Stockholm School of Economics 2003 and the authors thank the Department of Economic Statistics for helpful discussions and comments. This version of the paper has been presented at the workshop on Econometric Time Series Analysis - Methods and Applications, Linz, Austria 2003, and at the Econometrics and Empirical Economics (EEE) Programme of the 59th European Meeting of the Econometric Society (ESEM) Madrid, Spain 2004.

1 Introduction

The issue of parameter constancy in time series models is important and arises since it appears that data from many e.g. economic time series rather support models whose parameters are likely to be affected by an economical environment, see for instance Stock and Watson (1996). These time series exhibit properties that could be captured by models allowing for parameters that vary, and this most likely in a nonlinear way. Testing parameter constancy in nonlinear models becomes therefore important in attempts to detect misspecified models and avoid invalid inference.

There is vast literature on how to proceed, and the general approach is to have parameter constancy as the null hypothesis and test it against a parametric alternative. The parametric alternative and its properties are therefore in focus, and could for instance be characterized by parameters varying over time or that they change (mostly a one-time change) with some other endogenous/exogenous threshold variable, see for instance Chu and White (1991), Hansen (1992), and Andrews (1993). Time-varying parameters could be divided into two broad categories, i.e. parameters being stochastic (e.g. a random walk) or that they change according to a certain nonlinear deterministic function over time. For the former, see for instance Nyblom (1989). The latter will be in focus in this paper.

Considering nonlinear deterministic functions, the theory has evolved from testing against abrupt changes (an instantaneous structural break) to testing against more general and flexible functions where the parameters are allowed to have multiple smooth changes over time. As such, Lin and Teräsvirta (1994) discuss testing the null hypothesis of a linear autoregressive model against the logistic smooth transition autoregression (LSTAR) model. They assume that the data generating process (DGP) is stationary under the null hypothesis and asymptotic normality holds in their framework.

However, for many time series it may be difficult just by visual inspection to distinguish between data that have been generated by a random walk or data generated by a model that is nonlinear in parameters. It is therefore natural to discuss the same issue as in Lin and Teräsvirta (1994) when the stationary assumption is relaxed, and instead consider a random walk (with or without a drift) as the DGP when the modeller tests parameter constancy in the LSTAR model. Furthermore, another appealing feature of our approach is that our test is a direct test in the sense that it is based on "raw" data, and the first step is not to make the data stationary by taking first differences.

Imposing the assumption of a random walk means that we abandon the asymptotic normality of the OLS estimates for a non-standard limiting distribution characterized by various functionals of Brownian motions on the unit interval. The derived test statistic is robust for a wide class of error terms. In the sequel of finding the limiting distribution of our test, we generalize many asymptotic theoretical results in the area of unit roots that for instance could be found in Phillips (1987), Phillips and Perron (1988), and Hamilton (1994) among others, and new results are introduced.

The rest of this paper is organized as follows. In Section 2 we present the parametric alternative, i.e. the LSTAR model. Joint tests of linearity and non-stationarity in an LSTAR model as well as the distributional theory of the tests and their properties are described in Section 3. In Section 4 the Monte Carlo experiments are reported. Concluding remarks are found in Section 5. A Lemma and proofs are given in the Mathematical Appendix.

2 The model

Consider the logistic smooth transition autoregression model of order p, henceforth abbreviated the LSTAR(p) model,

$$y_t = \mathbf{x}_t'\boldsymbol{\pi}_1 + \mathbf{x}_t'\boldsymbol{\pi}_2 F(t;\boldsymbol{\theta}) + u_t, \quad t = 1, ..., T, \tag{1}$$

where $\mathbf{x}_t' = (1, y_{t-1}, ..., y_{t-p})$ is a $(p+1)\times 1$ vector and $p \in \mathbb{N}$, $\boldsymbol{\pi}_1 = (\pi_{10}, ..., \pi_{1,p})' \in \mathbb{R}^{p+1}$, $\boldsymbol{\pi}_2 = (\pi_{20}, ..., \pi_{2,p})' \in \mathbb{R}^{p+1}$, u_t is an error term with properties discussed in detail in the next subsection, $F(t, \boldsymbol{\theta})$ is a logistic smooth transition function. Following Lin and Teräsvirta (1994) a full parametrization of $F(t, \boldsymbol{\theta})$ in (1) is given by[1]

$$F(t;\boldsymbol{\theta}) = \frac{1}{(1 + \exp\{-\gamma(t^k + \alpha_1 t^{k-1} + ... + \alpha_{k-1}t + \alpha_k)\})}, \quad k = 1, 2, 3, \tag{2}$$

where $\boldsymbol{\theta} = (\gamma, \boldsymbol{\alpha}')' \in [0, \infty) \times \mathbb{R}^k$, $\boldsymbol{\alpha} = (\alpha_1, ..., \alpha_k)'$ and $\boldsymbol{\alpha} \in \mathbb{R}^k$ such that the roots of the polynomial in (2) are real. For any fixed $\boldsymbol{\theta}$, $F(t)$ defines a bounded function since $F(t) : \mathbb{R}_+ \to [0, 1]$. In the special case $k = 1$ and $\gamma \in (0, \infty)$ the transition is increasing in t allowing the model in (1) with (2) to change from $E[y_t|\mathcal{F}_{t-1}] \approx \mathbf{x}_t'\boldsymbol{\pi}_1$ to $E[y_t|\mathcal{F}_{t-1}] \approx \mathbf{x}_t'(\boldsymbol{\pi}_1 + \boldsymbol{\pi}_2)$ with t. The parameters γ and α_1 have a clear interpretation. The latter is a location parameter which indicates where the symmetric transition takes place, and the former determines the speed of transition. Letting $\gamma \to \infty$ implies that the transition takes place instantaneously at α_1, and F becomes the indicator function: $\mathbf{1}_{(\alpha_1,0)}(t) = 1$ if $t \in (\alpha_1, 0)$ and $\mathbf{1}_{(\alpha_1,0)}(t) = 0$ if $t \notin (\alpha_1, 0)$ and the model in (1) becomes a threshold autoregressive (TAR) model of order p. On the other hand, letting $\gamma \to 0$ implies that $F \to 1/2$ and the resulting model in (1) is a linear AR(p) model with parameter vector equal to $(\boldsymbol{\pi}_1 + \boldsymbol{\pi}_2)/2$. Moreover, with $k \geq 2$ the transition function F exhibits two or more transitions and the transition paths are nonmonotonic. Finally note that for any k and $\gamma \in [0, \infty)$ we have that F is differentiable and for $\gamma = \infty$, F exhibits point(s) of discontinuity.

3 Testing procedures

Our aim is to test for parameter constancy in (1), under the assumption that the true process is a random walk. To proceed and ease up exposition, we focus on the LSTAR(1) model with one transition, i.e. we let $p = k = 1$ in (1) and (2).

[1] For $k = 2$, Lin and Teräsvirta (1994) define the transition function $F(t) = 1 - \exp\{-\gamma(t - \alpha)^2\}$.

3.1 Testing parameter constancy in the LSTAR(1) model

We will first exploit the implications of linearity. We can, without loss of generality, replace F in (2) by a downward shift, say $\tilde{F}$, defined by

$$\tilde{F}(t, \boldsymbol{\theta}) \equiv F(t, \boldsymbol{\theta}) - 0.5. \tag{3}$$

It is now evident that we could test for parameter constancy in (1) by letting $\gamma = 0$, since $\tilde{F}(t, \boldsymbol{\theta}) = 0$ for all t and $\boldsymbol{\alpha}$. It is therefore natural to define the hypothesis $H_0 : \gamma = 0$ against $H_1 : \gamma > 0$.[2] We have, however, an identification problem under the null hypothesis, since the vectors $\boldsymbol{\pi}_2$ and $\boldsymbol{\alpha}$ will be undetermined, see e.g. Luukkonen, Saikkonen, and Teräsvirta (1988). This problem could be solved by approximating $\tilde{F}$ in a neighborhood of $\gamma = 0$. An obvious candidate is a first-order Taylor approximation of $\tilde{F}$ around $\gamma = 0$, and yields

$$A_1(t; \gamma, \alpha_1) = 0.25\gamma(t + \alpha_1) + r_1, \tag{4}$$

and r_1 is the remainder. Substituting (4) into (1) gives the linear approximation of the LSTAR model,

$$y_t = \pi_{10} + \pi_{11} y_{t-1} + (\pi_{20} + \pi_{21} y_{t-1})\, c\gamma(t + \alpha_1) + u_t^*, \tag{5}$$

and u_t^* is the error term adjusted with respect to the Taylor expansion where $u_t^* = u_t$ holds under the null hypothesis. Collecting the terms in (5) yields the reparametrized auxiliary regression model

$$y_t = \mathbf{s}_t' \boldsymbol{\lambda} + (y_{t-1}\mathbf{s}_t)' \boldsymbol{\varphi} + u_t^*, \tag{6}$$

where $\mathbf{s}_t = (1, t)'$, $\boldsymbol{\lambda} = (\lambda_0, \lambda_1)'$, $\boldsymbol{\varphi} = (\varphi_{01}, \varphi_{11})'$. The auxiliary null hypothesis of parameter constancy becomes

$$H_0^{aux}: \quad \lambda_0 \in \mathbb{R}, \quad \lambda_1 = 0, \quad \varphi_{01} \in \mathbb{R}, \quad \varphi_{11} = 0. \tag{7}$$

The null hypothesis in (7) implies that the regression equation in (6) is reduced to an AR(1) process with an intercept since the terms involving a time trend are equal to zero.

3.2 Testing unit roots in the LSTAR(1) model

Lin and Teräsvirta (1994) assume that the model (6) is stationary under the null hypothesis, i.e. $\varphi_{01} \in (-1, 1)$. We will in contrast consider a random walk as the true DGP implying that $\varphi_{01} = 1$, and assume stability under the alternative hypothesis, i.e. $H_1 : \varphi_{01} < 1$. Therefore, the joint auxiliary hypothesis of parameter constancy and non-stationarity is given by

$$H_0^{aux}: \quad \lambda_0 \in \mathbb{R}, \quad \lambda_1 = 0, \quad \varphi_{01} = 1, \quad \varphi_{11} = 0. \tag{8}$$

[2] To identify (1) under the alternative, one has to assume either $\gamma > 0$ or $\gamma < 0$. We choose *a priori* to rule out $\gamma < 0$ because in the defintion of F in (2) it is assumed that $\gamma \in [0, \infty)$.

Since the model in (6) contains an intercept, y_{t-1} is under the null hypothesis asymptotically equivalent to $\lambda_0(t-1)$. Thereby we will have a problem with collinearity in large samples because a time trend is already included in (6), see e.g. Sims, Stock, and Watson (1990) and Hamilton (1994). To avoid this problem, we first note that the process under the null hypothesis is a random walk with drift implying that $E[y_t] = \lambda_0 t$. We define therefore a new explanatory variable according to $\xi_{t-1} = y_{t-1} - E_{H_0^{aux}}[y_{t-1}]$. This transformation implies the regression model

$$y_t = \mathbf{s}_t^{*\prime}\boldsymbol{\lambda}^* + \left(\xi_{t-1}\mathbf{s}_t\right)' \boldsymbol{\varphi}^* + u_t^*, \tag{9}$$

where $\mathbf{s}_t^* = (\mathbf{s}_t', \, t^2)'$, $\boldsymbol{\lambda}^* = (\lambda_0^*, ..., \, \lambda_2^*)'$, and $\boldsymbol{\varphi}^* = \boldsymbol{\varphi}$. The null hypothesis of (8) is now equivalent to

$$H_0^{aux}: \quad \lambda_1^* \in \mathbb{R}, \quad \lambda_0^* = \lambda_2^* = 0, \quad \varphi_{01}^* = 1, \quad \varphi_{11}^* = 0. \tag{10}$$

All necessary transformations are now accomplished in the sense that the process under the null hypothesis of (10) is a random walk without drift given by $\xi_t = \xi_{t-1} + u_t$.

3.3 A joint test of parameter constancy and non-stationarity

We make the following general assumptions on the error term in (1) under the null hypothesis (10).

Assumption 1 *Let $\{u_t\}_{t=1}^{\infty}$ be sequence of random variables defined on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$, satisfying*

(A.1.1) $Eu_t = 0$ *for all* t.

(A.1.2) $\sup_{t\in\mathbf{N}} E\left|u_t\right|^{\beta} < \infty$ *for some* $\beta \in (2, \infty]$.

(A.1.3) $\sigma^2 = \lim_{T\to\infty} T^{-1} E\left(\sum_{t=1}^{T} u_t\right)^2$ *exists and* $\sigma^2 \in (0, \infty)$.

(A.1.4) $\{u_t\}_{t=1}^{\infty}$ is strong mixing with mixing coefficients $\alpha(m)$ satisfying $\sum_{m=1}^{\infty} \alpha(m)^{1-2/\beta} < \infty$.

Under Assumption 1 the limiting distribution for the OLS estimators $(\hat{\boldsymbol{\lambda}}^{*\prime}, \hat{\boldsymbol{\varphi}}^{*\prime})$ in (9) under the null of (10) can be derived. The result is given in the following theorem.

Theorem 1 *Consider model (9) when (10) holds. Furthermore, assume that $\{u_t\}_{t=1}^{\infty}$ satisfies Assumption 1. Then the least square estimator $\hat{\boldsymbol{\psi}} = \left(\sum_{t=1}^{T} \mathbf{h}_t\mathbf{h}_t'\right)^{-1} \left(\sum_{t=1}^{T} \mathbf{h}_t y_t\right)$ of $\boldsymbol{\psi} = (\boldsymbol{\lambda}^{*\prime}, \boldsymbol{\varphi}^{*\prime})'$ in (9), where $\hat{\boldsymbol{\psi}} = (\hat{\boldsymbol{\lambda}}^{*\prime}, \hat{\boldsymbol{\varphi}}^{*\prime})'$ and $\mathbf{h}_t = (\mathbf{s}_t^{*\prime}, (\xi_{t-1}\mathbf{s}_t)')'$, has the following asymptotic properties*

$$\boldsymbol{\gamma}_T(\hat{\boldsymbol{\psi}} - \boldsymbol{\psi}) \xrightarrow{d} \boldsymbol{\Psi}^{-1}\boldsymbol{\Pi}, \tag{11}$$

$$\hat{\boldsymbol{\psi}} - \boldsymbol{\psi} \xrightarrow{p} \mathbf{0}, \tag{12}$$

where $\boldsymbol{\gamma}_T = diag\left\{ \begin{array}{ccccc} T^{1/2} & T^{3/2} & T^{5/2} & T & T^2 \end{array} \right\}$, *and*

$$\boldsymbol{\Psi} = \left[\begin{array}{cc} \mathbf{M}_1 & \sigma\mathbf{M}_2 \\ \sigma\mathbf{M}_2' & \sigma^2\mathbf{M}_3 \end{array}\right], \quad \boldsymbol{\Pi} = \left[\begin{array}{c} \sigma\boldsymbol{\Pi}_1 \\ \left(\sigma^2/2\right)\boldsymbol{\Pi}_2 \end{array}\right],$$

where the sub-matrices are given by

$$\mathbf{M}_1 = \begin{bmatrix} 1 & 1/2 & 1/3 \\ 1/2 & 1/3 & 1/4 \\ 1/3 & 1/4 & 1/5 \end{bmatrix},$$

$$\mathbf{M}_2 = \begin{bmatrix} \int_0^1 W(r)dr & \int_0^1 rW(r)dr \\ \int_0^1 rW(r)dr & \int_0^1 r^2W(r)dr \\ \int_0^1 r^2W(r)dr & \int_0^1 r^3W(r)dr \end{bmatrix},$$

$$\mathbf{M}_3 = \begin{bmatrix} \int_0^1 W^2(r)dr & \int_0^1 rW^2(r)dr \\ \int_0^1 rW^2(r)dr & \int_0^1 r^2W^2(r)dr \end{bmatrix},$$

$$\mathbf{\Pi}_1 = \begin{bmatrix} W(1) \\ W(1) - \int_0^1 W(r)dr \\ W(1) - 2\int_0^1 rW(r)dr \end{bmatrix},$$

$$\mathbf{\Pi}_2 = \begin{bmatrix} W(1)^2 - \bar{\sigma}_u^2/\sigma^2 \\ W(1)^2 - \int_0^1 W^2(r)dr - \bar{\sigma}_u^2/(2\sigma^2) \end{bmatrix},$$

where $\xrightarrow{p}$, *and* $\xrightarrow{d}$ *denotes convergence in probability and distribution, respectively,* $W(r)$ *abbreviates a standard Brownian motion on* $[0,1]$, *and* $\bar{\sigma}_u^2 = \lim_{T\to\infty} T^{-1}\sum_{t=1}^T Eu_t^2$ *such that* $\bar{\sigma}_u^2 \in (0,\infty)$.

Proof. See Appendix A. ■

Note that under the null hypothesis (10), $\varphi_{01}^* = \pi_{11}$. It follows from Theorem 1 that the limiting distribution of $T(\hat{\varphi}_{01}^* - 1)$ is given as follows.

Corollary 2 *Suppose that the conditions of Theorem 1 hold. Define the test statistic*

$$T_n \equiv T(\hat{\varphi}_{01}^* - 1).$$

Then

$$T_n \xrightarrow{d} Q_1(W(r)) + Q_2\left(W(r);\sigma^2,\bar{\sigma}_u^2\right), \tag{13}$$

where Q_1 *and* Q_2 *are functions of Brownian motions defined in the Appendix A. Remark that* Q_2 *depends upon the nuisance parameters* σ^2 *and* $\bar{\sigma}_u^2$. *Hence, define the adjusted restricted test statistic*

$$T_a \equiv T_n - \hat{Q}_2,$$

where

$$\hat{Q}_2 = T^2\left(\sigma^2 - \bar{\sigma}_u^2\right)\Lambda_{44}/\left(2s_T^2\right) + T^3\left(\sigma^2 - \bar{\sigma}_u^2\right)\Lambda_{45}/\left(4s_T^2\right),$$

with

$$\Lambda_{44} = s_T^2\mathbf{r}_1\left[\sum\nolimits_{t=1}^T \mathbf{h}_t\mathbf{h}_t'\right]^{-1}\mathbf{r}_1', \quad \mathbf{r}_1 = \begin{bmatrix} 0 & 0 & 0 & 1 & 0 \end{bmatrix},$$

$$\Lambda_{45} = s_T^2\mathbf{r}_1\left[\sum\nolimits_{t=1}^T \mathbf{h}_t\mathbf{h}_t'\right]^{-1}\mathbf{r}_2', \quad \mathbf{r}_2 = \begin{bmatrix} 0 & 0 & 0 & 0 & 1 \end{bmatrix},$$

$$s_T^2 = \sum_{t=1}^T (y_t - \mathbf{h}_t'\hat{\boldsymbol{\psi}}_1)^2/(T-5).$$

Then,

$$T_a \xrightarrow{d} Q_1\left(W(r)\right). \tag{14}$$

Proof. See Appendix A. ■

There are several things to note about the results in (13) and (14). First, both tests are invariant with respect to λ_0 in (8), meaning that both tests are invariant with respect to if the true model is a random walk with or without drift. Second, it is only the limiting distribution for T_a that is invariant with respect to all nuisance parameters $\{\lambda_0, \sigma^2, \bar{\sigma}_u^2\}$, whereas, as shown in appendix, the limiting distribution for T_n is a function of σ^2 and $\bar{\sigma}_u^2$. However, if $\{u_t\}_{t=1}^{\infty}$ is an i.i.d. sequence, $\sigma^2 = \bar{\sigma}_u^2$ holds, and it follows that $T_n \xrightarrow{d} Q_1$. Third, even if the limiting distribution Q_1 is nuisance parameter free, the test statistic T_a itself contains nuisance parameters. This means that we must replace σ^2 and $\bar{\sigma}_u^2$ with some consistent estimates to operationalize the test, see for instance Phillips (1987). So far only the properties under the null hypothesis are mentioned, but under the alternative hypothesis and especially that $\varphi_{01}^* \in (-1, 1)$, the estimates $\hat{\boldsymbol{\psi}}$ from the auxiliary regression equation in (9) are normally distributed, see Lin and Teräsvirta (1994). Particularly, in our case it follows that $\hat{\varphi}_{01}^*$ is $\mathcal{O}_p(T)$ rather than $\mathcal{O}_p(T^{1/2})$.

Furthermore, it is natural to consider the F_{OLS}-test of the joint hypothesis in (10) which can be expressed as $\mathbf{R}\boldsymbol{\psi} = \mathbf{r}$, where $\mathbf{R} = \begin{bmatrix} \mathbf{0} & \mathbf{I}_4 \end{bmatrix}$ and $\mathbf{0}$ is the 4×1 the null vector and $\mathbf{r} = \begin{bmatrix} 0 & 0 & 1 & 0 \end{bmatrix}'$. The F_{OLS}-test statistic and its limiting distribution are given in the following corollary.

Corollary 3 *Define the OLS F statistic*

$$F_{OLS} \equiv \left(\mathbf{R}\hat{\boldsymbol{\psi}} - \mathbf{r}\right)' \left\{ s_T^2 \mathbf{R} \left[\sum_{t=1}^{T} \mathbf{h}_t \mathbf{h}_t' \right]^{-1} \mathbf{R}' \right\}^{-1} \left(\mathbf{R}\hat{\boldsymbol{\psi}} - \mathbf{r}\right)/4. \tag{15}$$

Consider model (9) when (10) holds. Assume that $\{u_t\}_{t=1}^{\infty}$ satisfies Assumption 1. Then

$$F_{OLS} \xrightarrow{d} \boldsymbol{\Pi}' \boldsymbol{\Psi}^{-1} \mathbf{R}' \left\{ \bar{\sigma}_u^2 \mathbf{R} \boldsymbol{\Psi}^{-1} \mathbf{R}' \right\}^{-1} \mathbf{R} \boldsymbol{\Psi}^{-1} \boldsymbol{\Pi}/4. \tag{16}$$

Furthermore, if $\{u_t\}_{t=1}^{\infty}$ is an i.i.d. sequence, then,

$$\begin{aligned} F_{OLS} \xrightarrow{d} & \begin{bmatrix} \boldsymbol{\Pi}_1 \\ \boldsymbol{\Pi}_2/2 \end{bmatrix}' \begin{bmatrix} \mathbf{M}_1 & \mathbf{M}_2' \\ \mathbf{M}_2 & \mathbf{M}_3 \end{bmatrix}^{-1} \mathbf{R}' \left\{ \mathbf{R} \begin{bmatrix} \mathbf{M}_1 & \mathbf{M}_2' \\ \mathbf{M}_2 & \mathbf{M}_3 \end{bmatrix}^{-1} \mathbf{R}' \right\}^{-1} \\ & \times \mathbf{R} \begin{bmatrix} \mathbf{M}_1 & \mathbf{M}_2' \\ \mathbf{M}_2 & \mathbf{M}_3 \end{bmatrix}^{-\mathbf{1}} \begin{bmatrix} \boldsymbol{\Pi}_1 \\ \boldsymbol{\Pi}_2/2 \end{bmatrix}/4. \end{aligned} \tag{17}$$

Proof. See Appendix A. ■

Without the i.i.d. assumption for the error term, the limiting distribution in (16) of F_{OLS} would contain the nuisance parameters σ^2 and $\bar{\sigma}_u^2$. However, adding the i.i.d. assumption gives that $\sigma^2 = \bar{\sigma}_u^2$ holds and it is immediately seen that (17) is nuisance parameter free.

4 Monte Carlo experiments

4.1 Asymptotic critical values and empirical size of the parameter constancy tests

In this subsection we present the critical values for the adjusted restricted test statistic, T_a, in Corollary 2. For illustration and comparison we also present the critical values for the corresponding test statistic based on a third-order Taylor expansion denoted T_3.[3] The critical values for the F_{OLS} test statistic in Corollary 3 are presented as well. When generating the asymptotic distributions, we let $T = 1\ 000\ 000$ to simulate a Brownian motion $W(r)$ on $[0,1]$, and the number of replications are set to 1 000 000, yielding the asymptotic critical values. The finite-sample critical values are obtained by simulating data from the model $y_t = y_{t-1} + u_t$ where $u_t \sim nid(0,1)$ with desired sample sizes, and thereafter the test statistics in (14) and (15) are calculated. This procedure is repeated 1 000 000 times, yielding the finite-sample distributions of the tests. The results are shown in Table 1.

Table 1: Critical values for the parameter constancy tests when the DGP is a random walk with i.i.d. increments.

	Probability that T_a is less than entry			Probability that T_3 is less than entry			Probability that F_{OLS} is less than entry		
T	0.01	0.05	0.10	0.01	0.05	0.10	0.90	0.95	0.99
50	-44.28	-32.82	-27.05	-119.64	-90.06	-75.56	4.61	5.38	7.12
100	-49.33	-35.70	-29.18	-138.48	-103.72	-86.31	4.42	5.10	6.51
250	-52.98	-37.69	-30.59	-159.44	-116.44	-95.79	4.30	4.91	6.24
500	-54.83	-38.72	-31.34	-168.48	-121.80	-99.730	4.28	4.88	6.16
1000	-54.83	-38.72	-31.34	-173.38	-124.70	-101.80	4.26	4.87	6.13
∞	-55.62	-39.35	-31.87	-175.31	-126.50	-103.11	4.26	4.85	6.09

Notes: The probability shown at the head of each column is the area in the left-hand tail. The results are based on 1 000 000 replications.

In Table 1 we see that the rejection of the null hypothesis requires large absolute values when using the T_a or T_3 tests, and the empirical distributions are heavily skewed to the left. We also see that the critical values for the T_a test converge faster to the asymptotic critical values than for the T_3 test, but both tests and their critical values at small sample sizes provide rather poor approximations of the asymptotic critical values. For the F_{OLS} test we see that it requires substantially higher observed values to reject the null hypothesis than compared to the standard F_{OLS} distribution.

[3] The critical values for T_3 in Table **??** refer to the test statistic $T_3 = T(\hat{\varphi}^*_{01} - 1)$, based on the auxiliary regression $y_t = \mathbf{s}_t^{*\prime}\boldsymbol{\lambda}^* + (\xi_t \mathbf{s}_t)' \boldsymbol{\varphi}^* + u_t^*$, where $\mathbf{s}_t^* = (s_t', t^4)'$, $\boldsymbol{\lambda}^* = (\lambda_1^*, ..., \lambda_5^*)$, $\boldsymbol{\varphi}^* = (\varphi_{01}^*, ..., \varphi_{04}^*)'$, and u_t^* is an error term with the same properties as the error term in (1).

In the case $u_t \sim nid(0,1)$ we can ignore the estimation of the nuisance parameters σ^2 and $\bar{\sigma}_u^2$ and we have that $\hat{Q}_2 = 0$ holds for all sample sizes. However, allowing for some more complex structure for the error term, we have to estimate $\hat{Q}_2$. It is well known that classical unit root tests suffer from rather large size distortions when the error term follows an MA(1) process with moving average coefficient close to plus/minus 1, see e.g. Hall (1989) and Schwert (1989). We choose therefore to study the empirical size of the T_a test in (14) when the error term is given by the MA(1) process $u_t = \varepsilon_t + \theta\varepsilon_{t-1}$ with $\varepsilon_t \sim nid(0,1)$ and $\theta \in \{-0.8, -0.5, 0, 0.5, 0.8\}$. To see how this affects the size of the T_a test, we calculate the empirical size (estimated size) for $T = 50, 100, 250$ and 1 000, based on 1 000 000 replications. The empirical size, for a fixed sample size and significance level, is measured by calculating $T_n - \hat{Q}_2$ and seeing how many times it is less than the critical value given in Table 1. To operationalize $\hat{Q}_2$ we use the Newey-West estimator with lag truncation parameter l equal to 4 to estimate σ^2, and $\bar{\sigma}_u^2$ is estimated, consistently, with $T^{-1}\sum \hat{u}_t^2$.[4] The results are presented in Table 2.

Table 2: The Empirical size of T_1 when the DGP is a random walk with MA(1) increments.

	T=50		T=100		T=250		T=1000	
	UAT	AT	UAT	AT	UAT	AT	UAT	AT
θ								
-0.8	0.845	0.695	0.982	0.976	0.994	0.992	0.981	0.946
-0.5	0.463	0.444	0.597	0.551	0.672	0.439	0.695	0.050
0.0	0.050	0.053	0.050	0.051	0.050	0.051	0.050	0.050
0.5	0.003	0.013	0.003	0.014	0.003	0.015	0.003	0.014
0.8	0.002	0.006	0.001	0.007	0.001	0.007	0.001	0.007

Notes: The nominal size is 5% and the results are based on 10 000 replications. UAT refers to (14) without the adjusting term. AT abbreviates the test in (14) with the adjusting term.

In Table 2 the empirical sizes reported in the first columns are based on the unadjusted test (UAT) statistic, i.e. we don't subtract $\hat{Q}_2$ from Q_1. The results in the second columns are based on $T_n - \hat{Q}_2$, i.e. the adjusted test (AT) statistic. From this table we see that we reject the null hypothesis far too often for negative values of θ, and also that the test is undersized for positive values of θ. Of course one hopes that AT statistic would perform significantly better than the UAT statistic, and with an estimated size that is close to the nominal size. This is not the case, however, and the improvement upon the empirical size is rather modest. It is clear that the adjustment is far from satisfactory. These results are in line with the findings for the classical unit root tests examined in Schwert (1989).

[4] The lag-truncation parameter in the Newey-West estimator is in our case chosen somewhat arbitrarily. Phillips (1987) shows that the Newey-West estimator can provide a consistent estimate of σ^2 in an MA(∞) process when l, $T \to \infty$ such that $l/T^{1/4} \to 0$ holds.

The poor size properties for the adjusted test statistic are explained by the fact that the adjustment factor $\hat{Q}_2$ is hardly correct in small samples. However, even for $T = 1000$ the estimates of σ^2 and $\bar{\sigma}^2$ can be quite far away from their true values. This problem could be solved, at least partially, by considering the approach of instrumental variables as in Hall (1989) or by adding lags of the error term. In our case, the latter would be equivalent to considering an LSTAR(p) model instead of the LSTAR(1) model. Schwert (1989) reports that for the classical augmented Dickey-Fuller unit root tests with 12 lags included and the corresponding t-type of test statistic, the size distortions are eliminated. Even though the deficit performance of the adjustment factor $\hat{Q}_2$, Campbell and Perron (1991) argue that a false rejection of the kind reported above is not necessarily bad, and in fact when $\theta \approx -1$, we essentially reject for a white noise process. To this end, we also studied the size properties of the T_a test, however not reported here, under the assumption that the error term was an AR(1) process. Also in this case it is found that the finite-sample performance of the adjustment quantity $\hat{Q}_2$ is non-satisfactory, and the test becomes undersized. In general it seems that the Newey-West estimator of σ^2 is hardly precise enough, and even though we in theory allow for possibly weakly dependent heterogeneously distributed errors, we may in practice end up in situations where the estimated size is either very high or low.

4.2 Power of the parameter constancy tests

We continue with investigation of the power properties of the parameter constancy tests. The DGP is chosen to be an LSTAR(1) model with one transition, i.e. $k = 1$. The experiments are conducted for the sample sizes $T = 100$ and $T = 250$, and we let $u_t \sim nid(0, 1)$. The stability restrictions $\pi_{11} \in (0, 1)$ and $\pi_{11} + \pi_{21} \in (0, 1)$ are imposed on the skeleton of the LSTAR(1) to rule out non-stable or explosive trajectories under the alternative hypothesis. Specifically, the autoregressive coefficients are assigned values according to the two scenarios

$$\text{Scenario I:} \quad \pi_{11} \in \{0.65, 0.70, 0.75, 0.80, 0.85\}, \quad \pi_{21} = 0.1;$$

$$\text{Scenario II:} \quad \pi_{11} = 0.1, \quad \pi_{21} \in \{0.65, 0.70, 0.75, 0.80, 0.85\}.$$

The remaining parameters are chosen to be the same in the two scenarios and are given by

$$\pi_{10} = 0, \quad \pi_{20} = 1, \quad \gamma \in \{0.01,\ 1,\ 100\}, \quad \alpha_1 = -T/2.$$

In Scenario I the autoregressive parameter varies in the linear part while it is kept fixed in the nonlinear part. Vice versa is exemplified in Scenario II. Specifically, the time series realized by Scenario I displays autoregressive parameters that at the beginning of the period are in a range of relatively high values. As time evolves, a nonlinear adjustment towards a new level with an autoregressive parameter almost equal to the autoregressive parameter at the beginning of the period takes place. That is, the difference between the two autoregressive parameters at the beginning and at the end of the period is about 0.1, depending on if a complete transition takes place or not. By Scenario II we realize time series with a low value of the autoregressive parameter at the beginning of the period, and here a nonlinear adjustment towards a new level implies that the autoregressive parameter lie in a range of high values at the end of the period. The differences in the autoregressive parameters between the beginning and the end of the period range now from about 0.55 to 0.75 (depending on if a complete transition takes place or not). Scenario I represents less nonlinear time series because the impact of a nonlinear change in dynamics is suppressed. Consequently, Scenario II reveals more nonlinear time series since the nonlinear changes in the dynamics are more pronounced. The difference between the intercept in the linear and nonlinear part in the LSTAR(1) model is set modest and equals 1. Furthermore, the influence of the speed of transition, occurring at time $T/2$, between regimes by varying γ is studied. With $\gamma = 0.01$ the LSTAR model is almost linear, and this is the only case when a complete transition does not take place over the sample periods because $F(t = T = 100; \gamma = 0.01, \alpha_1 = -50) \approx 0.62$ and $F(t = T = 250; \gamma = 0.01, \alpha_1 = -125) \approx 0.78$. Letting $\gamma = 1$ implies a smooth transition from one regime to another, and finally letting $\gamma = 100$ means that an instant single structural break between two extreme regimes takes place, and the result is a threshold autoregressive model.

One aim with Scenarios I and II is to illustrate the well-known fact that the tests based on first-order Taylor approximation suffer from low power when only a change in the intercept is considered, see Luukkonen, Saikkonen, and Teräsvirta (1988). We expect therefore that Scenario I will generate quite poor power results for the T_a test since the main source of power for this test would come from a nonlinear change in the dynamics – a feature that is suppressed in Scenario I. The opposite is expected in Scenario II – here a more evident nonlinear change in the dynamics takes place and we would expect more convincing power results for T_a. Furthermore, we note that by the design of the experiments above we realize time series that are initially stable with autoregressive parameter π_{11} and having an attractor at $\pi_{10} = 0$. A nonlinear adjustment towards a new long-run equilibrium, i.e. the attractor $\pi_{20}/(1 - \pi_{11} - \pi_{21}) \in [4, 20]$,[5] characterized by the autoregressive parameter $\pi_{11} + \pi_{21}$, takes place. Time series generated under Scenarios I and II have the same attractors, but in the way the trajectories travel from the one attractor to the other is different. Note also that the level of the long-run equilibrium does not only depend on π_{20} but also on π_{11} and π_{21}, a property that we choose to call the level leverage effect. We expect the leverage effect to have negative influence on the power for a test based on a first order approximation, because when $\pi_{11} + \pi_{12}$ is high, the difference in levels between the beginning and the end of the sample becomes large even for small values of π_{20}. In addition, when $\pi_{11} + \pi_{12}$ is close to unity, we encounter a process that is close to a unit root process at the end of the period, which also causes a reduction in power.

The power analysis is restricted to the tests T_a and T_3 in order to illuminate the difference between a first and third-order Taylor approximation. Moreover, to our knowledge, there are no similar parameter constancy tests (with a non-stationary null hypothesis), but we note that the auxiliary regression equation in (9) nests the models $y_t = a_0 + a_1 y_{t-1} + u_t$ and $y_t = a_0 + a_1 y_{t-1} + a_2 t + u_t$, considered e.g. in Dickey and

[5] Assuming that a complete transition takes place.

Fuller (1979) and Phillips and Perron (1988), and that their null hypotheses coincide with ours. The unit root test based on the former model will be denoted the DF test, and the test based on the latter model will henceforth be abbreviated to the PP test. The power results for the tests under Scenario I are found in Tables 3 and 4, and for Scenario II in Tables 5 and 6.

Table 3: Empirical power of the parameter constancy tests, the Dickey-Fuller test, and the Phillips-Perron test. The DGP is an LSTAR(1) model under Scenario I.

$T = 100$		T_a		DF		PP		T_3	
		5%	10%	5%	10%	5%	10%	5%	10%
	π_{11}								
	0.65	0.534	0.697	0.999	0.999	0.981	0.997	0.108	0.203
	0.70	0.414	0.583	0.987	0.998	0.912	0.975	0.087	0.159
$\gamma = 0.01$	0.75	0.297	0.454	0.903	0.974	0.734	0.884	0.067	0.137
	0.80	0.192	0.324	0.632	0.822	0.465	0.664	0.057	0.120
	0.85	0.114	0.207	0.239	0.420	0.222	0.380	0.052	0.099
	0.65	0.337	0.494	0.190	0.454	0.722	0.877	0.229	0.355
	0.70	0.196	0.327	0.024	0.097	0.425	0.637	0.207	0.331
$\gamma = 1.00$	0.75	0.084	0.164	0.001	0.005	0.158	0.303	0.200	0.320
	0.80	0.020	0.048	0.000	0.000	0.028	0.071	0.182	0.310
	0.85	0.003	0.008	0.000	0.000	0.001	0.002	0.163	0.275
	0.65	0.323	0.483	0.185	0.447	0.702	0.871	0.301	0.368
	0.70	0.187	0.317	0.023	0.090	0.409	0.622	0.223	0.340
$\gamma = 100$	0.75	0.079	0.157	0.001	0.005	0.146	0.293	0.205	0.322
	0.80	0.019	0.046	0.000	0.000	0.025	0.065	0.187	0.315
	0.85	0.002	0.007	0.000	0.000	0.000	0.002	0.172	0.293

Notes: The nominal sizes of the tests are 5% and 10%. The results are based on 10 000 replications.

It is seen in Table 3 that the DF test performs uniformly the best among all tests when $\gamma = 0.01$. This is to be expected since the LSTAR model appears linear without any evidence of a time trend or other pronounced nonlinearities, see panel (a) in Figure 1 (an LSTAR(1) model with γ= 0.01, π_{11}= 0.7, and π_{21}= 0.1). The power of the T_a and T_3 test statistics are modest because when the LSTAR model is nearly linear, the many regressors come to no efficient use and the tests are instead penalized. By this reasoning it is clear why the power for T_3 is lower than that for T_a which in turn is lower than that for PP.

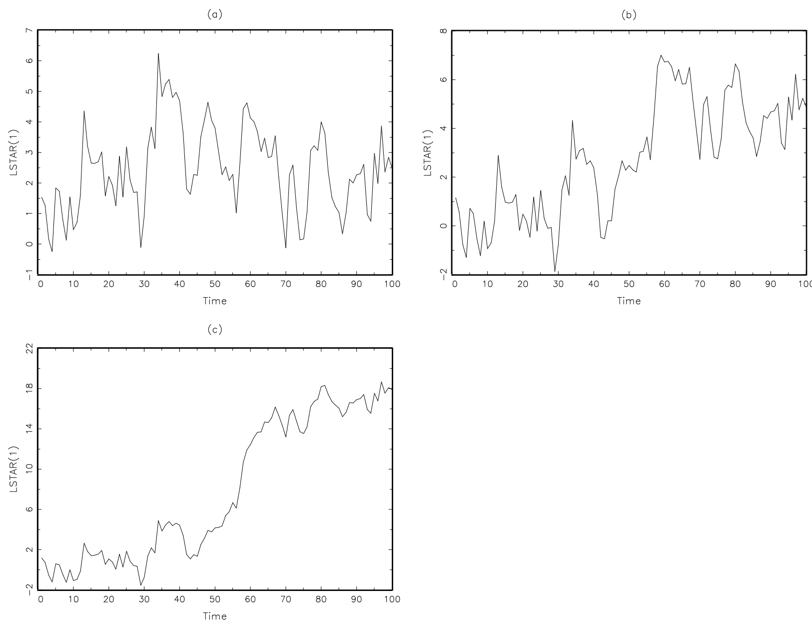

Figure 1: Realizations of the LSTAR(1) models under Scenario I.

Table 4: Empirical power of the parameter constancy tests, the Dickey-Fuller test, and the Phillips-Perron test. The DGP is an LSTAR(1) model under Scenario I.

$T = 250$		T_a		DF		PP		T_3	
		5%	10%	5%	10%	5%	10%	5%	10%
	π_{11}								
	0.65	0.992	0.998	1.000	1.000	1.000	1.000	0.410	0.568
	0.70	0.971	0.990	1.000	1.000	1.000	1.000	0.324	0.473
$\gamma = 0.01$	0.75	0.912	0.961	1.000	1.000	1.000	1.000	0.248	0.378
	0.80	0.766	0.877	0.976	0.997	0.998	1.000	0.171	0.295
	0.85	0.512	0.680	0.291	0.531	0.862	0.949	0.119	0.210
	0.65	0.933	0.971	1.000	1.000	1.000	1.000	0.652	0.788
	0.70	0.793	0.894	0.875	0.991	0.999	1.000	0.596	0.731
$\gamma = 1.00$	0.75	0.477	0.656	0.072	0.356	0.899	0.977	0.526	0.675
	0.80	0.100	0.208	0.000	0.000	0.222	0.444	0.488	0.648
	0.85	0.007	0.001	0.000	0.000	0.000	0.001	0.433	0.600
	0.65	0.930	0.970	1.000	1.000	1.000	1.000	0.654	0.780
	0.70	0.787	0.890	0.874	0.991	0.999	1.000	0.593	0.737
$\gamma = 100$	0.75	0.472	0.648	0.070	0.356	0.896	0.976	0.529	0.695
	0.80	0.099	0.203	0.000	0.000	0.217	0.440	0.485	0.649
	0.85	0.006	0.001	0.000	0.000	0.000	0.002	0.429	0.610

Note: The nominal sizes of the tests are 5% and 10%. The results are based on 10 000 replications.

Increasing the speed of transition to $\gamma = 1.00$, we see a turn in ranking among the test. The DF test is now inferior and its power is close to zero or equal to zero for high values of π_{11}. We see that the PP test still performs better than the T_a test. In fact, the PP test has substantial power for low values of π_{11} which may be explained by that the trajectories display rather linear time trends, see panel (b) in Figure 1 (an LSTAR(1) model with γ= 1, $\pi_{11} = 0.7$, and $\pi_{21} = 0.1$). The presence of a possible time trend and/or shift in levels also explains the remarkable drop in power for the DF test. The non-satisfactory performance of the T_a test just confirms the above stated expectations under Scenario I. If a complete transition takes place the impact of the level leverage effect on the power for T_a, by increasing π_{11}, without a clear nonlinear change in dynamics is revealed. The T_3 test is robust in the sense that it is quite independent of the values of π_{11}, and is in fact the only test that has power for high values of π_{11}. This can also be understood by inspection of panel (c) in Figure 1 (an LSTAR(1) model with $\gamma = 1.00$, $\pi_{11} = 0.85$, and $\pi_{21} = 0.1$), where the change in level between the regimes at the beginning and the end of the sample period is about 20 units. To this end we note that T_3 is the only test whose power is increasing in γ.

The same conclusions as with $\gamma = 1.00$ hold for $\gamma = 100$, and the power results for all tests are hardly affected by a more instantaneous transition.

We continue examining Scenario I with a larger sample size $T = 250$. From Table 4 we see that for $\pi_{11} \leq 0.75$ the power of the T_a increases to quite satisfactory levels, and that the PP test has the highest power among all tests. Even though the sample size is increased, we see that the impact of the level leverage effect on the T_a test still dominates, and as soon as a complete transition takes places the power for T_a is rapidly decreasing in π_{11}. The power for the T_3 test for $\gamma \geq 1.00$ is substantial and quite robust against variations in π_{11}. Once again it is the only test having reasonable power with high values of π_{11}.

Table 5: Empirical power of the parameter constancy tests, the Dickey-Fuller test, and the Phillips-Perron test. The DGP is an LSTAR(1) model under Scenario II.

$T = 100$		T_a		DF		PP		T_3	
		5%	10%	5%	10%	5%	10%	5%	10%
	π_{21}								
	0.65	0.970	0.989	1.000	1.000	1.000	1.000	0.307	0.450
	0.70	0.965	0.988	1.000	1.000	1.000	1.000	0.285	0.442
$\gamma = 0.01$	0.75	0.956	0.983	1.000	1.000	1.000	1.000	0.267	0.416
	0.80	0.950	0.977	1.000	1.000	1.000	1.000	0.254	0.411
	0.85	0.938	0.976	1.000	1.000	1.000	1.000	0.240	0.390
	0.65	0.976	0.989	0.538	0.825	0.959	0.988	0.661	0.789
	0.70	0.938	0.975	0.134	0.453	0.854	0.939	0.715	0.839
$\gamma = 1.00$	0.75	0.833	0.912	0.007	0.034	0.559	0.755	0.785	0.883
	0.80	0.489	0.636	0.000	0.000	0.133	0.290	0.865	0.936
	0.85	0.071	0.125	0.000	0.000	0.001	0.005	0.917	0.964
	0.65	0.970	0.983	0.519	0.817	0.959	0.988	0.667	0.807
	0.70	0.932	0.967	0.123	0.341	0.828	0.932	0.729	0.851
$\gamma = 100$	0.75	0.814	0.893	0.006	0.027	0.536	0.722	0.806	0.900
	0.80	0.462	0.614	0.000	0.000	0.123	0.278	0.884	0.943
	0.85	0.056	0.106	0.000	0.000	0.001	0.005	0.932	0.971

Note: The nominal sizes of the tests are 5% and 10%. The results are based on 10 000 replications.

From Table 5, illustrating Scenario II with $T = 100$, it is observed that for $\gamma = 0.01$ the classical unit root tests still have better power due to the linear properties of the LSTAR model. What is interesting is the positive change in power for the T_a test compared to Table 3. It is obvious that the presence of more pronounced nonlinear changes in dynamics result in an increase in power. For an example, the power in the case $(\gamma = 1.00, \pi_{11} = 0.7, \pi_{21} = 0.1)$ of Table 3 equals 0.196, whereas the power in and the case $(\gamma = 1.00, \pi_{11} = 0.1, \pi_{21} = 0.7)$ of Table 5 equals 0.938. The increase also confirms the expectations for Scenario II. In fact, it is the explanatory variable $t\xi_{t-1}$ in the auxiliary regression (9) that accounts for the increase in power compared to Scenario I. The power for the T_3 test has also increased, but suffers still from loss in power due to its ambiguity in modelling a relatively simple DGP. The power results of unity for the DF and PP tests are also rather expected since the autoregressive coefficient never exceeds 0.63 during the whole sample (which should be compared to that the autoregressive coefficient Scenario I never exceeds 0.92, explaining why the power for the DF and PP tests is lower in Table 3 than in Table 5).

Table 6: Empirical power of the parameter constancy tests, the Dickey-Fuller test, and the Phillips-Perron test. The DGP is an LSTAR(1) model under Scenario II.

$T = 250$		T_a		DF		PP		T_3	
		5%	10%	5%	10%	5%	10%	5%	10%
	π_{21}								
	0.65	1.000	1.000	1.000	1.000	1.000	1.000	0.918	0.964
	0.70	1.000	1.000	1.000	1.000	1.000	1.000	0.913	0.957
$\gamma = 0.01$	0.75	1.000	1.000	1.000	1.000	1.000	1.000	0.911	0.952
	0.80	1.000	1.000	1.000	1.000	1.000	1.000	0.898	0.950
	0.85	1.000	1.000	1.000	1.000	1.000	1.000	0.897	0.950
	0.65	1.000	1.000	1.000	1.000	1.000	1.000	0.997	0.999
	0.70	1.000	1.000	0.998	1.000	1.000	1.000	0.999	1.000
$\gamma = 1.00$	0.75	1.000	1.000	0.527	0.896	0.998	1.000	1.000	1.000
	0.80	0.983	0.996	0.000	0.001	0.754	0.906	1.000	1.000
	0.85	0.142	0.285	0.000	0.000	0.002	0.016	1.000	1.000
	0.65	1.000	1.000	1.000	1.000	1.000	1.000	0.998	0.999
	0.70	1.000	1.000	0.999	1.000	1.000	1.000	1.000	1.000
$\gamma = 100$	0.75	1.000	1.000	0.529	0.897	0.998	1.000	1.000	1.000
	0.80	0.980	0.994	0.000	0.009	0.740	0.890	1.000	1.000
	0.85	0.130	0.265	0.000	0.000	0.002	0.018	1.000	1.000

Note: The nominal sizes of the tests are 5% and 10%. The results are based on 10 000 replications.

The benefits from including nonlinear regressors become clearer when studying the cases $\gamma \geq 1.00$. We see that the T_a test now performs uniformly better than the PP and DF tests at all values of π_{21}. This is plausible since none of the DF and PP tests contains the regressor $t\xi_{t-1}$ combined with the fact that a full transition takes place. Furthermore, the power for T_a is decreasing with γ and π_{21} which indicates the trade-off between the fact that a distinct difference in π_{11} and π_{21} is needed to gain power through the term $t\xi_{t-1}$, and the fact that the larger the difference in π_{11} and π_{21} is, the more pronounced the level leverage effect will be. The domination of the latter is, however, less evident than in Scenario I. Moreover, the opposite holds for the T_3 test where the power actually increases with γ and π_{21}. That is, with evident changes in levels and dynamics, the advantages with a third-order Taylor expansion are illuminated.

The results when increasing the sample size under Scenario II are presented in Table 6. We see that the same conclusions can be drawn as from Table 5. However, we note that the power for T_a is close to unity, except for models that are almost non-stable at the end of the period, which holds independently of the value of γ. Moreover, the T_3

test exhibits a rather extraordinary robustness in power close to unity regardless of how γ and π_{21} are varied. Especially in a larger sample, the many regressors come into their own and a test based on third-order Taylor approximation seems highly preferable.

5 Conclusions

In this paper we derive tests for parameter constancy in an first order LSTAR model when the null hypothesis is a random walk (possibly with a drift). This means that we relax the stationarity assumption made by Lin and Teräsvirta (1994) under the null hypothesis in the LSTAR model. We argue that in many cases it can be difficult to distinguish if data have been generated by a nonlinear model with a smooth structural break in parameters or a random walk. A non-stationary process, rather than a stationary autoregressive process, might in many cases be a more plausible null hypothesis in an LSTAR(1) model.

To obtain our tests, and considerably simplify the testing procedures, we make a Taylor approximation of the smooth transition function yielding an auxiliary model. The inference about unit roots is then based on the OLS estimates from the auxiliary regression model. Analytical expressions for the asymptotic distributions of the tests are enabled by a very applicable Lemma, given and proved in the Mathematical appendix, which generalizes many asymptotic results derived in Phillips (1987), Phillips and Perron (1988), and Hamilton (1994) among others. The asymptotic results hold under a wide class of error terms, and nuisance parameter free tests are presented. All tests are invariant with respect to a possible drift in the random walk.

Despite the fact that the asymptotic results are robust against a wide class of errors, the tests suffer from rather large size distortions, as many of the conventional unit root tests do, in finite-samples. The size distortions are especially severe in the presence of a moving average error structure.

The power of our test, T_a, based on a first-order approximation is compared to the power of the Dickey and Fuller (DF) and Phillips and Perron (PP) tests, when the DGP is an LSTAR(1) model. We conclude that, in general, the DF test is the best test when a nearly linear LSTAR(1) model is considered. Studying LSTAR(1) models with more pronounced nonlinearities, both the T_a and PP tests perform better than the DF test. The main reason for this is that the DF test is based on a linear model without a time trend. Moreover, under modest nonlinear changes in the intercept and in the dynamics in the LSTAR(1) model, the power for the PP test is higher than for our T_a test. However, with a modest change in the intercept and a more pronounced nonlinear change in the dynamics, the T_a test performs better than the PP test. It is also revealed that with a dynamic root that is close to unity at the end of the sample period the impact of the level leverage effects is large and both the T_a and PP tests have modest power. As such, we also introduced a parameter constancy test based on a third-order Taylor approximation which in contrary to the T_a and PP tests performs very satisfactorily in the cases of clear nonlinear changes both in the intercept and in the dynamics.

Mathematical Appendix A

To be able to prove Theorem 1, we first introduce the following lemma.

Lemma 4 *If $\{u_t\}_{t=1}^{\infty}$ satisfies Assumption 1 and $\xi_t = \xi_{t-1} + u_t$, with $P(\xi_0 = 0) = 1$, then as $T \to \infty$*

$$T^{-(p+q/2+1)} \sum_{t=1}^{T} t^p \xi_{t-1}^q \xrightarrow{d} \sigma^q \int_0^1 r^p W(r)^q dr, \tag{18}$$

$$T^{-(p+1)} \sum_{t=1}^{T} t^p u_t^2 \xrightarrow{a.s.} \bar{\sigma}_u^2/(p+1), \tag{19}$$

$$T^{-(v+1/2)} \sum_{t=1}^{T} t^v u_t \xrightarrow{d} \sigma W(1) - v\sigma \int_0^1 r^{v-1} W(r) dr, \tag{20}$$

$$T^{-(p+1)} \sum_{t=1}^{T} t^p \xi_{t-1} u_t$$

$$\xrightarrow{d} \tfrac{1}{2} \left(\sigma^2 W(1)^2 - \sigma^2 p \int_0^1 r^{p-1} W(r)^2 dr - \frac{\bar{\sigma}_u^2}{(p+1)} \right), \tag{21}$$

where $p, v \geq 0$, $q \geq 1$, and $\xrightarrow{a.s.}$ denotes convergence almost surely.

Note that Lemma 4 gives more general results than needed to prove Theorem 1. In fact, it enables us to derive the limiting distribution for a parameter constancy test in a first order LSTAR model based on any order of approximation of the logistic transition function with an arbitrary number of transitions.

Proof of (18). Define the following cadlag function on $D[0,1]$,[6]

$$W_t(r, \omega) = \frac{1}{\sigma\sqrt{T}} \xi_{[Tr]} = \frac{1}{\sigma\sqrt{T}} \xi_{t-1}, \qquad r \in [\tfrac{t-1}{T}, \tfrac{t}{T}), \qquad t = 1, ..., T,$$

where $[.]$ denotes the integer part of its argument. Using $([Tr]+1)/T = t/T$, we can conclude that

$$(([Tr]+1)/T)^p \, W_t(r)^q = (t/T)^p \left(\frac{1}{\sigma\sqrt{T}} \xi_{t-1} \right)^q$$

holds. The left-hand side now defines a continuous functional of $W_t(r)$ on $D[0,1]$. It follows from the Functional Central Limit Theorem (FCLT), the Continuous Mapping Theorem (CMT), and $\lim_{T\to\infty} (([Tr]+1)/T)^p = r^p$, that

$$\begin{aligned} & T^{-(p+q/2+1)} \sum_{t=1}^{T} t^p \xi_{t-1}^q \\ = \; & \sigma^q T^{-1} \sum_{t=1}^{T} (t/T)^p \left(\frac{1}{\sqrt{T}\sigma} \xi_{t-1} \right)^q \\ = \; & \sigma^q \int_0^1 (([Tr]+1)/T)^p \, W_t(r)^q dr \xrightarrow{d} \sigma^q \int_0^1 r^p W(r)^q dr. \end{aligned}$$

[6] See for instance Billingsley (1968) and Davidson (1994).

Thus, (18) holds. ■

Proof of (19). Note that the left-hand side in (19) can be written as

$$T^{-(p+1)}\sum_{t=1}^{T} t^p u_t^2 = T^{-1}\sum_{t=1}^{T}(t/T)^p\left(u_t^2-\sigma_t^2\right)+T^{-1}\sum_{t=1}^{T}(t/T)^p\sigma_t^2,$$

where $\sigma_t^2 = E(u_t^2)$. The condition $\sup_{t\in\mathbf{N}} E\left|u_t\right|^{\beta} < \infty$, where $\beta > 2$ and $(t/T)^p \in [0,1]$ implies that $\sum_{t=1}^{\infty}\left(E(t/T)^p\left|u_t^2-\sigma_t^2\right|^{\beta}/t^{\beta}\right)^{1/\beta} < \infty$ holds. It follows by the strong law of large numbers for α -mixings (see, e.g. McLeish (1975), Theorem 2.10, and Herrndorf (1984)), that $T^{-1}\sum_{t=1}^{T}(t/T)^p\left(u_t^2-\sigma_t^2\right) \stackrel{a.s.}{\to} 0$. The second term on the right-hand side converges to $\bar{\sigma}_u^2/(p+1)$ where $\bar{\sigma}_u^2 = \lim_{T\to\infty} T^{-1}\sum_{t=1}^{T}\sigma_t^2$. ■

Proof of (20). First, the case where $v = 0$ follows by Wooldridge and White (1988) which says that $T^{-1/2}\sum_{t=1}^{T} u_t \stackrel{d}{\to} \sigma W(1) \sim N(0,\sigma^2)$. Furthermore, for some integer $v \geq 1$ we use the result in (18) letting $p = v-1$ and $q = 1$, to deduce that

$$\begin{aligned}
& T^{-(v+1/2)}\sum_{t=1}^{T} t^{v-1}\xi_{t-1} \\
= \; & T^{-(v+1/2)}\sum_{t=1}^{T}\left(\sum_{i=1}^{T} i^{v-1}-\sum_{j=1}^{t} j^{v-1}\right)u_t \\
= \; & T^{-(v+1/2)}\sum_{t=1}^{T} u_t\sum_{i=1}^{T} i^{v-1} - T^{-(v+1/2)}\sum_{t=1}^{T}\left(\sum_{j=1}^{t} j^{v-1}\right)u_t.
\end{aligned}$$

Rewriting $\sum_{i=1}^{T} i^{v-1} = T^v/v+\mathcal{O}(T^{v-1})$ and $\sum_{j=1}^{t} j^{v-1} = t^v/v+\mathcal{O}(t^{v-1})$, gives

$$\begin{aligned}
& T^{-(v+1/2)}\sum_{t=1}^{T} u_t t^v \\
= \; & T^{-1/2}\sum_{t=1}^{T} u_t - T^{-(v+1/2)}v\sum_{t=1}^{T} t^{v-1}\xi_{t-1} + \mathcal{O}_p(T^{-1}) \\
& \stackrel{d}{\to} \sigma W(1) - \sigma v\int_0^1 r^{v-1}\, W(r)dr,
\end{aligned}$$

and (20) is proved. ■

Proof of (21). This is a bit more problematic since we cannot define any continuous functional with bounded variation, almost surely, corresponding to the expression in (21). We shall solve this problem by two different approaches.

Approach 1: Define a random polygon function on $C[0,1]$,

$$\begin{aligned}
W_t'(r,\omega) &= \frac{1}{\sigma\sqrt{T}}\xi_{[Tr]}(\omega) + \frac{Tr-[Tr]}{\sigma\sqrt{T}}u_{[Tr]+1}(\omega) \\
&= \frac{1}{\sigma\sqrt{T}}\xi_{t-1}(\omega) + \frac{1}{\sigma\sqrt{T}}\frac{u_t(\omega)}{1/T}r - \frac{t-1}{T},
\end{aligned}$$

which has continuous sample paths with bounded variation, almost surely. Note that $W_t'(r,\omega)$ is linear on each of the subintervals $r \in \left[\frac{t-1}{T}, \frac{t}{T}\right)$ and taking on the value $\xi_t/\sqrt{T}\sigma$ at the point t/T. Since $W_t'(r,\omega)$ has bounded variation, almost surely, we can define the Riemann Stieltjes integral

$$\begin{aligned}
&\int_0^1 r^p W_t'(r) dW_t'(r) \\
&= \sum_{t=1}^{T} \int_{(t-1)/T}^{t/T} r^p \left(\frac{1}{\sigma\sqrt{T}} \xi_{t-1} + \frac{1}{\sigma\sqrt{T}} \frac{u_t}{1/T} \left(r - \frac{t-1}{T} \right) \right) dW_t'(r) \\
&= \sum_{t=1}^{T} \int_{(t-1)/T}^{t/T} \left(r^p \left(\frac{\xi_{t-1} - (t-1){u_t}^2}{\sigma^2} \right) + r^{p+1} \frac{{u_t}^2 T}{\sigma^2} \right) dr \\
&= T^{-(p+1)} \sum_{t=1}^{T} \left(\frac{t^p \xi_{t-1} u_t}{\sigma^2} + \frac{t^p {u_t}^2}{2\sigma^2} \right) + \mathcal{O}_p(T^{-1}), \qquad (22)
\end{aligned}$$

where we at the second equality used that $dW_t'(r) = (\sqrt{T} u_t/\sigma) dr$, to obtain the result, by direct integration, at the last line. Moreover, because the integral in (22) is defined in a Riemann-Stieltjes sense, we use the rules for partial integration to obtain the following result for the left-hand side in (22)

$$\int_0^1 r^p W_t' dW_t'(r) = 0.5 \left(W_t'(1)^2 - p \int_0^1 r^{p-1} W_t'(r)^2 dr \right). \qquad (23)$$

Combining (22) and (23), making use of the FCLT, CMT, and the result in (19), we deduce that

$$\begin{aligned}
&T^{-(p+1)} \sum_{t=1}^{T} t^p \xi_{t-1} u_t \\
&= \sigma^2/2 \left(W_t'(1)^2 - p \int_0^1 r^{p-1} W_t'^2(r) dr \right) - T^{-(p+1)}/2 \sum_{t=1}^{T} t^p {u_t}^2 + \mathcal{O}_p(T^{-1}) \\
&\xrightarrow{d} 0.5 \left(\sigma^2 W(1)^2 - \sigma^2 p \int_0^1 r^{p-1} W(r)^2 dr - \bar{\sigma}_u^2/(p+1) \right),
\end{aligned}$$

and (21) follows.

Approach 2: Above result could also be established without using the theory of Riemann-Stieltjes integrals. To see this note that

$$\begin{aligned}
&T^{-(p+1)} \sum_{t=1}^{T} t^{p-1} \xi_{t-1}^2 \\
&= T^{-(p+1)} \sum_{t=1}^{T} \left(\sum_{i=1}^{T} i^{p-1} - \sum_{j=1}^{t} j^{p-1} \right) (u_t^2 + 2u_t \xi_{t-1}) \\
&= T^{-(p+1)} \sum_{t=1}^{T} (u_t^2 + 2u_t \xi_{t-1}) \sum_{i=1}^{T} i^{v-1} - T^{-(p+1)} \sum_{t=1}^{T} \left(\sum_{j=1}^{t} j^{v-1} \right) (u_t^2 + 2u_t \xi_{t-1}).
\end{aligned}$$

Rewriting $\sum_{i=1}^{T} i^{p-1} = T^p/p + \mathcal{O}(T^{p-1})$ and $\sum_{j=1}^{t} j^{v-1} = t^p/p + \mathcal{O}(t^{p-1})$ gives

$$
\begin{aligned}
& T^{-(p+1)} \sum_{t=1}^{T} t^p \xi_{t-1} u_t \\
= \; & T^{-1} \sum_{t=1}^{T} \xi_{t-1} u_t + 0.5T^{-1} \sum_{t=1}^{T} u_t^2 - 0.5T^{-(p+1)} p \sum_{t=1}^{T} t^{p-1} \xi_{t-1}^2 \\
& -0.5T^{-(p+1)} \sum_{t=1}^{T} t^p u_t^2 + \mathcal{O}_p(T^{-1}) \\
= \; & 0.5\sigma^2 W_t(1)^2 - 0.5T^{-(p+1)} p \sum_{t=1}^{T} t^{p-1} \xi_{t-1}^2 \\
& -0.5T^{-(p+1)} \sum_{t=1}^{T} t^p u_t^2 + \mathcal{O}_p(T^{-1}) \\
& \xrightarrow{d} 0.5 \left(\sigma^2 W(1)^2 - \sigma^2 p \int_0^1 r^{p-1} W(r)^2 dr - \bar{\sigma}_u^2/(p+1) \right),
\end{aligned}
$$

where we have used $T^{-1} \sum_{t=1}^{T} \xi_{t-1} u_t = 0.5\sigma^2 W_t(1)^2 - 0.5T^{-1} \sum_{t=1}^{T} u_t^2$ to obtain the last equality, and thereafter the weak convergence follows by using the results derived in (18) and (19). The claim in (21) follows once again. ■

Proof of Theorem 1. The limiting distribution is obtained by applying the results in Lemma 4. To see this note that,

$$
\boldsymbol{\gamma}_T(\hat{\boldsymbol{\psi}} - \boldsymbol{\psi}) = \begin{bmatrix} \hat{\mathbf{M}}_1 & \hat{\mathbf{M}}_2 \\ \hat{\mathbf{M}}_2' & \hat{\mathbf{M}}_3 \end{bmatrix}^{-1} \begin{bmatrix} \hat{\mathbf{P}}_1 \\ \hat{\mathbf{P}}_2 \end{bmatrix}, \tag{24}
$$

where

$$
\begin{aligned}
\hat{\mathbf{M}}_1 &= [\hat{m}_{1ij}]_{3\times 3}, \text{ and } \hat{m}_{1ij} = T^{-(i+j-1)} \sum_{t=1}^{T} t^{(i+j-2)}, \\
\hat{\mathbf{M}}_2 &= [\hat{m}_{2ij}]_{3\times 2}, \text{ and } \hat{m}_{2ij} = T^{-(i+j-1/2)} \sum_{t=1}^{T} t^{(i+j-2)} \xi_{t-1}, \\
\hat{\mathbf{M}}_3 &= [\hat{m}_{3ij}]_{2\times 2}, \text{ and } \hat{m}_{3ij} = T^{-(i+j)} \sum_{t=1}^{T} t^{(i+j-2)} \xi_{t-1}^2, \\
\hat{\mathbf{P}}_1 &= [\hat{p}_{1i}]_{3\times 1}, \text{ and } \hat{p}_{1i} = T^{-(i-1/2)} \sum_{t=1}^{T} t^{i-1} u_t, \\
\hat{\mathbf{P}}_2 &= [\hat{p}_{2i}]_{2\times 1}, \text{ and } \hat{p}_{2i} = T^{-i} \sum_{t=1}^{T} t^{i-1} \xi_{t-1} u_t.
\end{aligned}
$$

It follows from Lemma 4 that

$$
\begin{aligned}
& \hat{\mathbf{M}}_1 \xrightarrow{d} \mathbf{M}_1, \quad \hat{\mathbf{M}}_2 \xrightarrow{d} \sigma \mathbf{M}_2, \quad \hat{\mathbf{M}}_3 \xrightarrow{d} \sigma^2 \mathbf{M}_3, \\
& \hat{\mathbf{P}}_1 \xrightarrow{d} \sigma \boldsymbol{\Pi}_1, \quad \hat{\mathbf{P}}_2 \xrightarrow{d} 0.5\sigma^2 \boldsymbol{\Pi}_2,
\end{aligned}
$$

hold, where $\mathbf{M}_1$, $\mathbf{M}_2$, $\mathbf{M}_3$, $\mathbf{\Pi}_1$, and $\mathbf{\Pi}_2$, are defined in Theorem 1. Because (24) defines a continuous function we conclude that (11) holds.

The consistency of the OLS estimators is an immediate consequence of (24). ■

Proof of (13) in Corollary 2. Define the matrices

$$\begin{aligned} \mathbf{D}_1 &= diag\left\{ \begin{array}{ccccc} 1 & 1 & 1 & \sigma & \sigma \end{array} \right\}, \\ \tilde{\mathbf{\Pi}}_{12} &= \left[\begin{array}{c} 0.5(W(1)^2 - 1) \\ 0.5(W(1)^2 - \int W^2 - 0.5) \end{array} \right]. \end{aligned}$$

We can now decompose $\mathbf{\Psi}^{-1}\mathbf{\Pi}$ as

$$\mathbf{\Psi}^{-1}\mathbf{\Pi} = \sigma \mathbf{D}_1^{-1} \left[\begin{array}{cc} \mathbf{M}_1 & \mathbf{M}_2 \\ \mathbf{M}_2' & \mathbf{M}_3 \end{array} \right]^{-1} \left(\left[\begin{array}{c} \mathbf{\Pi}_1 \\ \tilde{\mathbf{\Pi}}_2 \end{array} \right] + \left[\begin{array}{c} \mathbf{0} \\ \vartheta^2/2\sigma^2 \\ \vartheta^2/4\sigma^2 \end{array} \right] \right),$$

where $\vartheta^2 = \sigma^2 - \bar{\sigma}_u^2$ and $\mathbf{0}$ is a (3×1) vector of zeros. It follows from Theorem 1 that $T_n = \mathbf{r}_1 \boldsymbol{\gamma}_T (\hat{\boldsymbol{\psi}} - \boldsymbol{\psi}) \xrightarrow{d} \mathbf{r}_1 \mathbf{\Psi}^{-1}\mathbf{\Pi}$, and where we can write

$$\mathbf{r}_1 \mathbf{\Psi}^{-1}\mathbf{\Pi} = \mathbf{r}_1 \left[\begin{array}{cc} \mathbf{M}_1 & \mathbf{M}_2 \\ \mathbf{M}_2' & \mathbf{M}_3 \end{array} \right]^{-1} \left[\begin{array}{c} \mathbf{\Pi}_1 \\ \tilde{\mathbf{\Pi}}_2 \end{array} \right] + \frac{\vartheta^2}{\sigma^2} \mathbf{r}_1 \left[\begin{array}{cc} \mathbf{M}_1 & \mathbf{M}_2 \\ \mathbf{M}_2' & \mathbf{M}_3 \end{array} \right]^{-1} \left[\begin{array}{c} \mathbf{0} \\ 1/2 \\ 1/4 \end{array} \right], \qquad (25)$$

because $\mathbf{r}_1 \sigma \mathbf{D}_1^{-1} = \mathbf{r}_1$ holds. We see that (25) involves the inversion of a 5×5 matrix, but the pre-multiplication with $\mathbf{r}_1$ implies that it is only the elements at the 4th row of the inverse that are of interest. Hence, to proceed we use the rules for the inverse of partioned matrices

$$\left(\begin{array}{cc} \mathbf{M}_1 & \mathbf{M}_2 \\ \mathbf{M}_2' & \mathbf{M}_3 \end{array} \right)^{-1} = \left(\begin{array}{cc} \mathbf{M}_1^{-1} + \mathbf{M}_1^{-1}\mathbf{M}_2\mathbf{H}^{-1}\mathbf{M}_2'\mathbf{M}_1^{-1} & -\mathbf{M}_1^{-1}\mathbf{M}_2\mathbf{H}^{-1} \\ -\mathbf{H}^{-1}\mathbf{M}_2'\mathbf{M}_1^{-1} & \mathbf{H}^{-1} \end{array} \right), \qquad (26)$$

where $\mathbf{H} = \mathbf{M}_3 - \mathbf{M}_2'\mathbf{M}_1^{-1}\mathbf{M}_2$. Obviously we must find the expressions for $-\mathbf{H}^{-1}\mathbf{M}_2'\mathbf{M}_1^{-1}$ and $\mathbf{H}^{-1}$. Thus, we adopt the following notation

$$\mathbf{H} = \left[\begin{array}{cc} H_{11} & H_{12} \\ H_{21} & H_{22} \end{array} \right],$$

with elements

$$
\begin{aligned}
H_{11} &= \int W^2 + 72\int W\int rW - 60\int W\int r^2W + 360\int rW\int r^2W \\
&\quad -9\left(\int W\right)^2 - 192\left(\int rW\right)^2 - 180\left(\int r^2W\right)^2, \\
H_{12} &= \int rW^2 - 9\int W\int rW + 36\int W\int r^2W - 30\int W\int r^3W \\
&\quad -222\int rW\int r^2W + 180\int rW\int r^3W - 180\int r^2W\int r^3W \\
&\quad +36\left(\int rW\right)^2 + 180\left(\int r^2W\right)^2, \\
H_{21} &= H_{12}, \\
H_{22} &= \int r^2W^2 + 72\int rW\int r^2W - 60\int rW\int r^3W + 360\int r^2W\int r^3W \\
&\quad -9\left(\int rW\right)^2 - 192\left(\int r^2W\right)^2 - 180\left(\int r^3W\right)^2,
\end{aligned}
$$

where for instance $\int W^2$ is short for $\int_0^1 W(r)^2 dr$. This implies that

$$
\mathbf{H}^{-1} = \frac{1}{D}\begin{bmatrix} H_{22} & -H_{12} \\ -H_{12} & H_{11} \end{bmatrix} = \begin{bmatrix} H_{44}^* & -H_{45}^* \\ -H_{54}^* & H_{55}^* \end{bmatrix},
$$

where $D = H_{11}H_{22} - H_{12}^2$, $H_{44}^* = D^{-1}H_{22}$, $H_{45}^* = H_{54}^* = D^{-1}H_{12}$, and $H_{55}^* = D^{-1}H_{11}$. The sub-indices of the elements with asteriks denotes the actual position they have in (26). Due to $\mathbf{r}_1$ we are only interested in H_{44}^* and H_{45}^*. For the same reason considering the 2×3 matrix $-\mathbf{H}^{-1}\mathbf{M}_2'\mathbf{M}_1^{-1}$, we only need the result of the upper row (corresponding to the 4th row in (26)), denoted $(4,.)$. Hence, $(-\mathbf{H}^{-1}\mathbf{M}_2'\mathbf{M}_1^{-1})_{(4,.)} = \begin{bmatrix} H_{41}^* & H_{42}^* & H_{43}^* \end{bmatrix}$ where

$$
\begin{aligned}
H_{41}^* &= H_{44}^*\left(9\int W - 36\int rW + 30\int r^2W\right) \\
&\quad +H_{45}^*\left(-9\int rW + 36\int r^2W + 30\int r^3W\right) \\
H_{42}^* &= H_{44}^*\left(-36\int W + 192\int rW - 180\int r^2W\right) \\
&\quad +H_{45}^*\left(36\int rW - 192\int r^2W + 180\int r^3W\right) \\
H_{43}^* &= H_{44}^*\left(30\int W - 180\int rW + 180\int r^2W\right) \\
&\quad +H_{45}^*\left(-30\int rW + 180\int r^2W - 180\int r^3W\right).
\end{aligned}
$$

This implies that the first term on the right-hand side of (25) is given by

$$\begin{aligned}
&\mathbf{r}_1 \begin{bmatrix} \mathbf{M}_1 & \mathbf{M}_2' \\ \mathbf{M}_2 & \mathbf{M}_3 \end{bmatrix}^{-1} \begin{bmatrix} \mathbf{\Pi}_1 \\ \tilde{\mathbf{\Pi}}_2 \end{bmatrix} \\
= \; & H_{41}^* W(1) + H_{42}^* \left(W(1) - \int W \right) + H_{43}^* \left(W(1) - 2 \int rW \right) \\
& + H_{44}^* \left(0.5(W(1)^2 - 1) \right) - H_{45}^* \left(0.5(W(1)^2 - \int W^2 - 0.5) \right) \\
\equiv \; & Q_1(W(r)),
\end{aligned}$$

and that the second term on the right-hand side of (25) equals

$$\frac{\vartheta^2}{\sigma^2} \mathbf{r}_1 \begin{bmatrix} \mathbf{M}_1 & \mathbf{M}_2' \\ \mathbf{M}_2 & \mathbf{M}_3 \end{bmatrix}^{-1} \begin{bmatrix} \mathbf{0} \\ 0.5 \\ 0.25 \end{bmatrix} = \frac{\vartheta^2}{2\sigma^2} \left(H_{44}^* - 0.5 H_{45}^* \right) \equiv Q_2(W(r); \sigma^2, \bar{\sigma}_u^2).$$

Adding up $Q_1(W(r))$ and $Q_2(W(r); \sigma^2, \bar{\sigma}_u^2)$ we obtain the limiting distribution for T_n. It is evident that only Q_1 is nuisance parameter free, and the claim in (13) follows.

Moreover, to prove (14) note first that

$$\begin{aligned}
T^2 \Lambda_{44} &= s_T^2 \mathbf{r}_1 \left(\boldsymbol{\gamma}_T \left[\sum\nolimits_{t=1}^{T} \mathbf{h}_t \mathbf{h}_t' \right]^{-1} \boldsymbol{\gamma}_T \right) \mathbf{r}_1' \xrightarrow{p} \frac{\bar{\sigma}_u^2}{\sigma^2} \mathbf{r}_1 \begin{bmatrix} \mathbf{M}_1 & \mathbf{M}_2' \\ \mathbf{M}_2 & \mathbf{M}_3 \end{bmatrix}^{-1} \mathbf{r}_1' = \frac{\bar{\sigma}_u^2}{\sigma^2} H_{44}^*, \\
T^3 \Lambda_{45} &= s_T^2 \mathbf{r}_1 \left(\boldsymbol{\gamma}_T \left[\sum\nolimits_{t=1}^{T} \mathbf{h}_t \mathbf{h}_t' \right]^{-1} \boldsymbol{\gamma}_T \right) \mathbf{r}_2' \xrightarrow{p} \frac{\bar{\sigma}_u^2}{\sigma^2} \mathbf{r}_1 \begin{bmatrix} \mathbf{M}_1 & \mathbf{M}_2' \\ \mathbf{M}_2 & \mathbf{M}_3 \end{bmatrix}^{-1} \mathbf{r}_2' = -\frac{\bar{\sigma}_u^2}{\sigma^2} H_{45}^*,
\end{aligned}$$

since $s_T^2 \xrightarrow{p} \bar{\sigma}_u^2$ holds under the null hypothesis and that the identities $T\mathbf{r}_1 = \mathbf{r}_1 \boldsymbol{\gamma}_T$, and $T^2 \mathbf{r}_2 = \mathbf{r}_2 \boldsymbol{\gamma}_T$ are valid. We conclude that,

$$\frac{T^2 \Lambda_{44} \vartheta^2}{2 s_T^2} + \frac{T^3 \Lambda_{45} \vartheta^2}{4 s_T^2} \xrightarrow{p} \frac{\vartheta^2}{2\sigma^2} H_{44}^* - \frac{\vartheta^2}{4\sigma^2} H_{45}^* = Q_2,$$

implying that

$$T_a = T_n - \left(\frac{T^2 \Lambda_{44} \vartheta^2}{2 s_T^2} + \frac{T^3 \Lambda_{45} \vartheta^2}{4 s_T^2} \right) \xrightarrow{d} Q_1.$$

Thus, (14) holds. ■

Proof of Corollary 3. Define $\tilde{\boldsymbol{\gamma}}_T = diag \left\{ \begin{array}{cccc} T^{1/2} & T^{5/2} & T & T^2 \end{array} \right\}$. Note that $\tilde{\boldsymbol{\gamma}}_T \mathbf{R} = \mathbf{R} \boldsymbol{\gamma}_T$ holds. It follows that

$$\begin{aligned}
F_{OLS} &\equiv \left(\mathbf{R}\hat{\boldsymbol{\psi}} - \mathbf{r} \right)' \left\{ s_T^2 \mathbf{R} \left[\sum_{t=1}^{T} \mathbf{h}_t \mathbf{h}_t' \right]^{-1} \mathbf{R}' \right\}^{-1} \left(\mathbf{R}\hat{\boldsymbol{\psi}} - \mathbf{r} \right) / 4 \\
&= \left(\hat{\boldsymbol{\psi}} - \boldsymbol{\psi} \right)' \mathbf{R}' \tilde{\boldsymbol{\gamma}}_T \left\{ s_T^2 \tilde{\boldsymbol{\gamma}}_T \mathbf{R} \left[\sum_{t=1}^{T} \mathbf{h}_t \mathbf{h}_t' \right]^{-1} \mathbf{R}' \tilde{\boldsymbol{\gamma}}_T \right\}^{-1} \tilde{\boldsymbol{\gamma}}_T \mathbf{R} \left(\hat{\boldsymbol{\psi}} - \boldsymbol{\psi} \right) / 4 \\
&= \left(\mathbf{R} \boldsymbol{\gamma}_T \left(\hat{\boldsymbol{\psi}} - \boldsymbol{\psi} \right) \right)' \left\{ s_T^2 \mathbf{R} \boldsymbol{\gamma}_T \left[\sum_{t=1}^{T} \mathbf{h}_t \mathbf{h}_t' \right]^{-1} \boldsymbol{\gamma}_T \mathbf{R}' \right\}^{-1} \left(\mathbf{R} \boldsymbol{\gamma}_T \left(\hat{\boldsymbol{\psi}} - \boldsymbol{\psi} \right) \right) / 4.
\end{aligned}$$

Moreover, $\boldsymbol{\gamma}_T(\hat{\boldsymbol{\psi}} - \boldsymbol{\psi}) \xrightarrow{d} \boldsymbol{\Psi}^{-1}\boldsymbol{\Pi}$ holds by Theorem 1, and s_T^2 is an consistent estimate to $\bar{\sigma}_u^2$, so the Slutsky Theorem yields

$$F_{OLS} \xrightarrow{d} \left(\mathbf{R}\boldsymbol{\Psi}^{-1}\boldsymbol{\Pi}\right)' \left\{\bar{\sigma}_u^2\mathbf{R}\boldsymbol{\Psi}^{-1}\mathbf{R}'\right\}^{-1} \mathbf{R}\boldsymbol{\Psi}^{-1}\boldsymbol{\Pi}/4,$$

and thus, (16) holds.

Furthermore, assuming that $\{u_t\}_{t=1}^{\infty}$ is an i.i.d. sequence implies that $\bar{\sigma}_u^2 = \sigma^2$. Define $\mathbf{D}_2 = diag\left\{ \begin{array}{cccc} \sigma & \sigma & 1 & 1 \end{array} \right\}$. We obtain

$$\mathbf{R}\boldsymbol{\Psi}^{-1}\boldsymbol{\Pi} = \mathbf{D_2R}\left[\begin{array}{cc} \mathbf{M}_1 & \mathbf{M}_2' \\ \mathbf{M}_2 & \mathbf{M}_3 \end{array}\right]^{-1}\left[\begin{array}{c} \boldsymbol{\Pi}_1 \\ \boldsymbol{\Pi}_2/2 \end{array}\right], \tag{27}$$

and

$$\sigma^2\mathbf{R}\boldsymbol{\Psi}^{-1}\mathbf{R}' = \mathbf{D}_2\mathbf{R}\left[\begin{array}{cc} \mathbf{M}_1 & \mathbf{M}_2' \\ \mathbf{M}_2 & \mathbf{M}_3 \end{array}\right]^{-1}\mathbf{R}'\mathbf{D}_2. \tag{28}$$

Using (27) and (28) yields

$$\begin{aligned} & F_{OLS} \xrightarrow{d} \left(\mathbf{R}\boldsymbol{\Psi}^{-1}\boldsymbol{\Pi}\right)' \left\{\sigma^2\mathbf{R}\boldsymbol{\Psi}^{-1}\mathbf{R}'\right\}^{-1} \mathbf{R}\boldsymbol{\Psi}^{-1}\boldsymbol{\Pi}/4 \\ = & \left[\begin{array}{c} \boldsymbol{\Pi}_1 \\ \boldsymbol{\Pi}_2/2 \end{array}\right]'\left[\begin{array}{cc} \mathbf{M}_1 & \mathbf{M}_2' \\ \mathbf{M}_2 & \mathbf{M}_3 \end{array}\right]^{-1} \times \mathbf{R}'\left\{\mathbf{R}\left[\begin{array}{cc} \mathbf{M}_1 & \mathbf{M}_2' \\ \mathbf{M}_2 & \mathbf{M}_3 \end{array}\right]^{-\mathbf{1}}{}' \mathbf{R}\right\} \\ & \times\mathbf{R}\left[\begin{array}{cc} \mathbf{M}_1 & \mathbf{M}_2' \\ \mathbf{M}_2 & \mathbf{M}_3 \end{array}\right]^{-1}\left[\begin{array}{c} \boldsymbol{\Pi}_1 \\ \boldsymbol{\Pi}_2/2 \end{array}\right]/4. \end{aligned}$$

Finally, since $\{u_t\}_{t=1}^{\infty}$ is an i.i.d. sequence, $\boldsymbol{\Pi}_2$ is nuisance parameter free. Thus, (17) holds. ■

References

Andrews, D. W. (1993): "Tests for parameter instability and structural change with unknown change point," *Econometrica*, (4), 821–856.

Billingsley, P. (1968): *Convergence of Probability Measures*. John Wiley.

Campbell, J. Y., and P. Perron (1991): "Pitfalls and Opportunities: What Macroeconomists Should Know About Unit Roots," *NBER Macroeconomics Annual*, 0, 141–200.

Chu, C. S., and H. White (1991): "Testing for structural change in some simple time series models," Department of Economics, University of California, San Diego, CA, Discussion paper no. 91-6.

Davidson, J. (1994): *Stochastic Limit Theory*. Oxford University Press.

Dickey, D. A., and W. A. Fuller (1979): "Distribution of the Estimators for Autoregressive Time Series," *Journal of the American Statistical Association*, 74, 427–431.

Hall, A. (1989): "Testing for a unit root in the presence of moving average errors," *Biometrika*, 76, 49–56.

Hamilton, J. D. (1994): *Time Series Analysis*. Princeton University Press, New Jersey.

Hansen, B. E. (1992): "Tests for Parameter Instability in Regressions with I(1) Processes," *Journal of Business and Economic Statistics*, 10, 321–335.

Herrndorf, N. (1984): "A Functional Central Limit Theorem for Weakly Dependent Sequences of Random Variables," *The Annals of Probability*, 12, 141–153.

Lin, C. F. J., and T. Teräsvirta (1994): "Testing the constancy of regression parameters against continuous structural change," *Journal of Econometrics*, 62, 211–228.

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

McLeish, D. L. (1975): "A Maximal Inequality and Dependent Strong Laws," *The Annals of Probability*, 3, 829–839.

Nyblom, J. (1989): "Testing for the constancy of parameters over time," *Journal of the American Statistical Association*, 84, 223–230.

Phillips, P. C. B. (1987): "Time Series Regression with a Unit Root," *Econometrica*, 55, 277–301.

Phillips, P. C. B., and P. Perron (1988): "Testing for Unit Root in Time Series Regression," *Biometrika*, 75, 335–346.

Schwert, G. W. (1989): "Tests for Unit Roots: A Monte Carlo Investigation," *Journal of Business and Economic Statistics*, 7, 147–159.

Sims, C. A., J. H. Stock, and M. W. Watson (1990): "Inference in Linear Time Series with Some Unit Roots," *Econometrica*, 58, 113–144.

Stock, J. H., and M. W. Watson (1996): "Evidence on structural Instability in Macroeconomic Time Series Relations," *Journal of Business and Economic Statistics*, 14, 11–30.

Wooldridge, J. M., and H. White (1988): "Some invariance principles and central limit theorems for dependent hetergeneous processes," *Econometric Theory*, 4, 210–230.