

Dickey-Fuller type of Tests against Nonlinear Dynamic Models

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. 580

January 2005

Abstract

In this paper we introduce several test statistics of testing the null hypotheses of a random walk (with or without drift) against models that accommodate a smooth nonlinear shift in the level, the dynamic structure, and the trend. We derive analytical limiting distributions for all tests. Finite sample properties are examined. The performance of the tests is compared to that of the classical unit root tests by Dickey-Fuller and Phillips and Perron, and is found to be superior in terms of power.

JEL classification: C12; C22; C52

Key words: Dickey-Fuller test; LSTAR(p); LSTART(p); Nonlinear Trends; Parameter Constancy; Unit Root, Brownian motion

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

†e-mail: strs@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 2002 and the authors thank the Department of Economic Statistics for helpful discussions and comments. A second version of the paper was presented at the Latin American Meeting of the Econometric Society (LAMES), Panama City, Panama 2003, and at the workshop Econometrics and Computational Economics, Helsinki, Finland 2003 were the authors are especially thankful to Pentti Saikkonen for helpfull comments. This version of the paper has been presented at a contributed session of the 19th Meeting of the European Economic Association (EEA), Madrid, Spain 2004, and at Uppsala University, Sweden, 2004.

1 Introduction

It is well known that classical unit root tests based on linear models such as those by Dickey and Fuller (1979), Phillips (1987), and Phillips and Perron (1988) among others, lack power when the model specification under the alternative hypothesis is nonlinear. In particular, nonlinear models with structural changes in levels and trends bias the classical tests towards nonrejection, as pointed out in Perron (1990). In the light of that many time series exhibit jumps or more smooth structural changes in levels and trends over time, the bias in the classical tests is particularly a non-desirable property. As such, this has serious implications for applied work because shocks will be treated as if they have a permanent effect, when they in fact are transitory.

Nowadays there are many ways of how to test the unit root hypothesis in a nonlinear set-up. One approach is to test unit roots in time series models with a change in levels and/or trends where the break point is known as in Perron (1989), Perron (1990), and Lanne and Lütkepohl (2002) among others, or unknown as in Perron and Vogelsang (1992), Zivot and Andrews (1992) and Banerjee, Lumsdaine, and Stock (1992) as examples.

Most of the approaches in above quoted literature deal with a single break in threshold type of models. In many cases this can be seen a too restrictive and models with more smooth and multiple breaks in level and/or trends are called for. Example of more general models can be found in Leybourne, Newbold, and Vougas (1998) and Harvey and Mills (2002), where the changes are determined by a logistic smooth transition function, or as in Saikkonen and Lütkepohl (2002) and Lanne, Lütkepohl, and Saikkonen (2003) where the functional form of the changes is set very flexible.

The approach that we are taking is similar to the one in Leybourne, Newbold, and Vougas (1998) and Harvey and Mills (2002), i.e. the null hypothesis of a unit root is tested against a model with a logistic smooth transition in the intercept and in the time trend. We extend, however, their discussion by explicitly allowing for a smooth transition in the dynamics. This extension is reasoned by e.g. that the amplitude of the fluctuations after a trend shift may not be the same as before the trend shift. As such, the models that we consider are the p th order logistic smooth transition autoregressive model (LSTAR(p)) in Lin and Teräsvirta (1994), and also an LSTAR(p) model where a time trend is included (LSTART(p)). We assume that the transitions take place over time and that the speed of transition between regimes may not be the same for all parameters.

We derive several unit root tests in the aforementioned LSTAR type of models and analytical limiting distributions for the tests are presented. This is in contrary to Leybourne, Newbold, and Vougas (1998) and Harvey and Mills (2002) because the limiting distributions of their unit root tests are found by simulations. Their testing methodology is also different and is based on NLS estimation, whereas our tests are based on a Taylor approximation of the transition function which implies that inference for unit roots is easily obtained by simple regressions.

It should be mentioned that there already exists several unit root tests in the smooth transition autoregressive (STAR) framework, see for instance Enders and Granger (1998), Bec, Salem, and Carrasco (2002), Eklund (2003a), Eklund (2003b),

and Kapetanios, Shin, and Snell (2003), but they differ, however fundamentally since they do *not* model the transition as a function of time, it is modelled as a function of lagged dependent variables.

The rest of the paper is organized as follows. In Section 2 the models are presented. The procedure for testing a unit root against these nonlinear models is described in Section 3. In Section 4 theoretical asymptotic properties of the tests are presented. Finite-sample properties of the tests are investigated in Section 5. Concluding remarks are given in Section 6. Thereafter two appendices follow where proofs and additional tables can be found.

2 The models

In this section we introduce a new family of nonlinear models by adding a time trend with its own individual transition function to the LSTAR model defined in Lin and Teräsvirta (1994). Such a model could be defined as

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

where $\mathbf{x}_t = (1, y_{t-1}, y_{t-2}, \dots, y_{t-p}, t)'$ is a $(p+2) \times 1$ vector and p is an integer such that $p \geq 1$, $\boldsymbol{\pi}_1 = (\pi_{10}, \dots, \pi_{1,p+1})' \in \mathbb{R}^{p+2}$, $\mathbf{F} = \text{diag}\{F_0, \dots, F_{p+1}\}$ is a $(p+2) \times (p+2)$ matrix with transition functions on its diagonal, $\boldsymbol{\pi}_2 = (\pi_{20}, \dots, \pi_{2,p+1})' \in \mathbb{R}^{p+2}$, and u_t is an error term whose properties are discussed later on. Suitable transition functions could be defined according to

$$F_i(t; \boldsymbol{\phi}_i) = \frac{1}{1 + \exp\{-\gamma_i(t - c_i)\}} - \frac{1}{2}, \quad i = 0, \dots, p+1, \quad (2)$$

where $\boldsymbol{\phi}_i = (\gamma_i, c_i) \in \mathbb{R}_+ \times (0, T)$. Viewing $F_i(t; \boldsymbol{\phi}_i)$ as a function of t , with $\boldsymbol{\phi}_i$ fixed, it is a bounded continuous non-decreasing transition function in t such that $F_i(t) : \mathbb{R}_+ \rightarrow [-1/2, 1/2]$. The specification in (2) allows for one transition over time for each parameter, where γ_i determines the speed of transitions from one regime to another, and the parameter c_i indicates which point in time the transition is symmetric about.¹ Viewing $F_i(t; \gamma_i, c_i)$ as a function of γ_i , with t and c_i fixed, $\lim_{\gamma_i \rightarrow \infty} F_i(\gamma_i) = -0.5$ if $t \in (c_i, T]$, and $\lim_{\gamma_i \rightarrow \infty} F_i = 0.5$ if $t \in [0, c_i]$, meaning that the transition functions make a jump (point of discontinuities) at c_i . On the other hand, $\lim_{\gamma_i \rightarrow 0} F_i(\gamma_i) = 0$, and the resulting model in (1) is linear. Finally we note that $F_i(t; \gamma_i, c_i) \in C^n(\mathbb{R}_+, \mathbb{R}_+ \times (0, T))$ where n is an integer such that $n \geq 1$.²

It is evident that the model in (1) with (2) is able to capture the properties of quite complex nonlinear time series where e.g. the level, trend and dynamics are initially at an equilibrium and as time evolves a nonlinear adjustment towards a new long-run equilibrium takes place. Notice also that the model specification in (1) with (2) nests models such as the LSTAR(p) model introduced in Lin and Teräsvirta (1994), the classical linear autoregressive models in Dickey and Fuller (1979) and

¹The concern in this chapter is a smooth transition function with a single transition between regimes. However, in the Appendix A it becomes clear that we could generalize the discussion to transition functions with an arbitrary number of transitions, i.e. let $F(t; \gamma, c_1, \dots, c_m) = 1/(1 + e^{-\gamma \prod_{i=1}^m (t - c_i)}) - 1/2$, where $\gamma > 0$, and $c_1 < \dots < c_m$ for $m \geq 1$.

² $C^n(\mathbb{R}_+, \mathbb{R}_+ \times (0, T))$ denotes the family of continuous functions that are differentiable of order n with respect to $t \in \mathbb{R}_+$, $\gamma_i \in \mathbb{R}_+$, and $c_i \in (0, T)$.

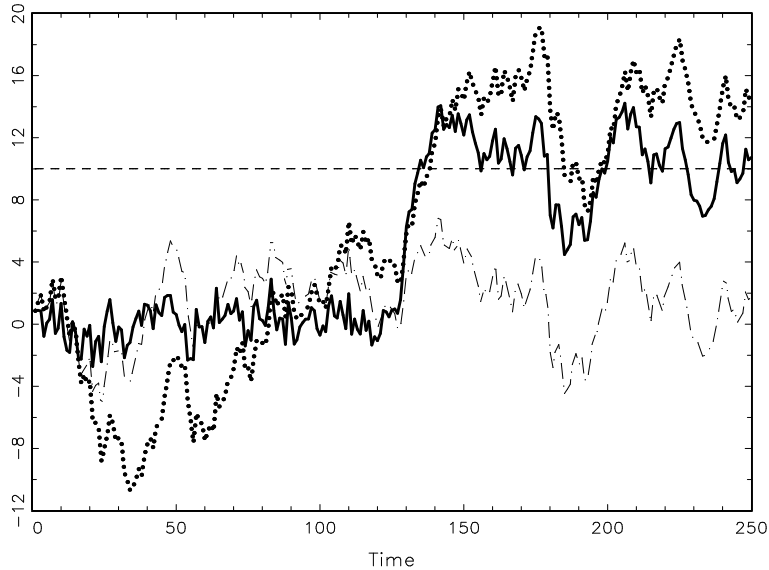


Figure 1: Typical realizations of an LSTAR(1) model (solid line) with a long-run equilibrium (dashed line), a stationary AR(1) process (dash-dotted line), and a random walk (dotted line).

Phillips and Perron (1988), many of the nonlinear models in Leybourne, Newbold, and Vougas (1998) and Harvey and Mills (2002), and the threshold autoregressive (TAR) model discussed in Chan and Tong (1986).

2.1 The LSTAR(p), LSTART(p), and LSTD(p) models

The discussion in this paper will focus, without loss of generality due to our testing methodology in the next section, on a sub-class of models in (1) with (2), where all transition functions are set equal. The generic transition function will hereafter be denoted F and is defined as F_i in (2) but with all indices dropped. Within the sub-class of models characterized by equal transition functions, there are three models of particular interest. First, the LSTAR(p) model obtained by letting $\pi_{1,p+1} = \pi_{2,p+1} \equiv 0$ in (1). Second, a new model that we call the LSTART(p) model which is characterized by the same nonlinearities as the LSTAR(p) model but in addition accommodates a nonlinear trend. To obtain the LSTART(p) model, no restrictions in (1) are imposed. Finally, another new model with a smooth transition only in the deterministic part, i.e. in the intercept and the time trend, is considered. This model is obtained by letting $\pi_{21} = \dots = \pi_{2,p} \equiv 0$ in (1), and is referred to as the LSTD(p) model. Typical realizations of an LSTAR and LSTART model together with their linear counterparts are depicted in Figures 1 and 2, respectively.

In the coming discussion it is shown that the power of conventional unit root tests such as those by Dickey and Fuller (1979) and Phillips and Perron (1988) have power close to zero under the nonlinear models illustrated in Figures 1 and 2. In these figures it is also seen that a random walk is a relevant null hypothesis in LSTAR type of models, i.e. the trajectories of a random walk in small samples may appear similarly to the trajectories generated by an LSTAR type of model.

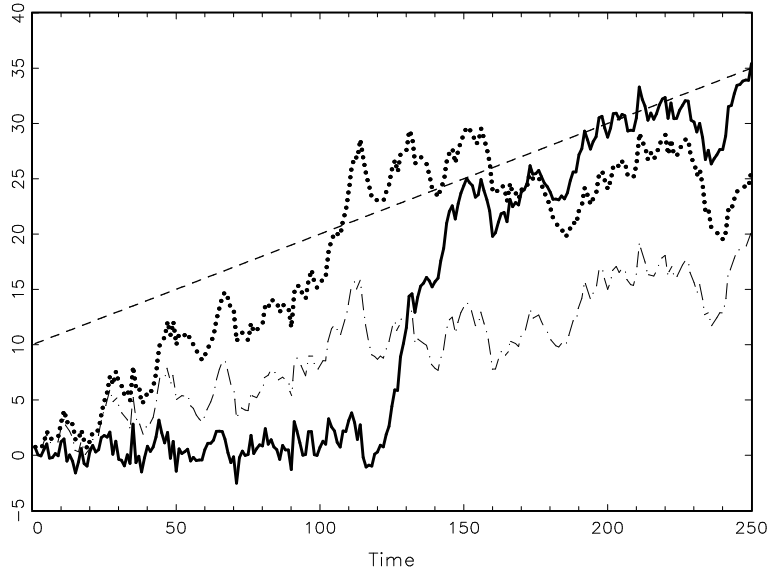


Figure 2: Typical realizations of an LSTART(1) model (solid line) with a long-run equilibrium (dashed line), a trend stationary AR(1) process (dash-dotted line), and a random walk with drift (dotted line).

3 Testing methodology

In this section we derive several testing equations that serve as devices for the inference about unit roots in the LSTAR, LSTART and LSTD models proposed above. To proceed it is convenient to separate the cases $p = 1$ and $p \geq 2$. When $p = 1$, the testing equations are called nonlinear Dickey-Fuller (NDF) testing equations. Letting $p \geq 2$, the resulting testing equations are abbreviated to the nonlinear augmented Dickey-Fuller (NADF) testing equations, as well as nonlinear partially augmented Dickey-Fuller (NPADF) testing equations.

3.1 The NDF testing equations

Assuming that $p = 1$ we want to test the null hypothesis of a random walk without drift against the LSTAR(1) model. This is reasoned by that both models lack the property of a time trend. Furthermore, we also want to test the null hypothesis of a random walk with drift against the LSTART(1) and the LSTD(1) models. As such, both the null and the alternative hypotheses display a trend, but are fundamentally different since in the former case the trend is stochastic and in the latter case it is nonlinear and deterministic. Testing the unit root hypothesis in the LSTAR(1), LSTART(1), and LSTD(1) models is formalized as follows

$$\begin{aligned}
 H_{01} & : y_t = y_{t-1} + u_t, \\
 H_{a1} & : y_t = \pi_{10} + \pi_{11}y_{t-1} + (\pi_{20} + \pi_{21}y_{t-1})F(t, \phi) + u_t, \\
 H_{0m} & : y_t = \pi_{10} + y_{t-1} + u_t, \quad m = 2, 3, \\
 H_{a2} & : y_t = \pi_{10} + \pi_{11}y_{t-1} + \pi_{12}t + (\pi_{20} + \pi_{21}y_{t-1} + \pi_{22}t)F(t, \phi) + u_t, \\
 H_{a3} & : y_t = \pi_{10} + \pi_{11}y_{t-1} + \pi_{12}t + (\pi_{20} + \pi_{22}t)F(t, \phi) + u_t.
 \end{aligned}$$

The implementation of these three tests is straightforward in the sense that for $m = 1, 2, 3$ the models under the alternative hypotheses, H_{am} , nest the models under the null hypotheses, H_{0m} . It is evident that these tests can be conducted by imposing the parameter restrictions

$$\gamma = 0, \quad \pi_{10} = 0, \quad \pi_{11} = 1, \quad (3)$$

on the LSTAR(1) model represented by H_{a1} to obtain the null hypothesis H_{01} , and imposing the restrictions

$$\gamma = 0, \quad \pi_{10} \in \mathbb{R}, \quad \pi_{11} = 1, \quad \pi_{12} = 0, \quad (4)$$

on the LSTAR(1) and the LSTD(1) models represented by H_{a2} and H_{a3} , respectively, to deduce the null hypotheses H_{02} and H_{03} . However, it is known that letting $\gamma = 0$ leads to an identification problem under the null hypotheses. Following Luukkonen, Saikkonen, and Teräsvirta (1988) we remedy this problem by a third-order Taylor expansion of γ around 0 in F . This is feasible because $F \in C^4[\mathbb{R}_+, \mathbb{R}_+ \times (0, T)]$, and since F is an odd function we especially note that $\partial F / \partial \gamma|_{\gamma=0} \neq 0$ and $\partial^3 F / \partial \gamma^3|_{\gamma=0} \neq 0$ in an open interval $(-\varepsilon, \varepsilon)$, for $\varepsilon > 0$. The motivation of a third-order Taylor approximation, rather than a first, is mainly the robustness in power to a wider class of DGP's under the alternative hypotheses. As far as the LSTAR model is concerned, it is well known that a first-order Taylor approximation results in low power if the transition only takes place in the intercept, see Luukkonen, Saikkonen, and Teräsvirta (1988) and the discussion in He and Sandberg (2005a). Similar arguments hold for the LSTAR and LSTD models and a third-order approximation is in many cases strongly preferable. Thus, a third-order Taylor approximation of the transition function is given by

$$T_3(t; \gamma, c) = \gamma(t - c)/4 + \gamma^3(t - c)^3/48 + r(\gamma), \quad (5)$$

where $r(\gamma)$ is a remainder such that $r(0) = 0$. Substituting (5) into the models in H_{a1} , H_{a2} , and H_{a3} , and merging terms yields the auxiliary regression equations³

$$\begin{aligned} H_{a1}^{aux} &: y_t = \mathbf{s}'_{1t} \boldsymbol{\lambda}_1 + (y_{t-1} \mathbf{s}_{1t})' \boldsymbol{\varphi}_1 + u_t^*, \\ H_{a2}^{aux} &: y_t = \mathbf{s}'_{2t} \boldsymbol{\lambda}_2 + (y_{t-1} \mathbf{s}_{1t})' \boldsymbol{\varphi}_2 + u_t^*, \\ H_{a3}^{aux} &: y_t = \mathbf{s}'_{2t} \boldsymbol{\lambda}_3 + y_{t-1} \varphi_{30} + u_t^*, \end{aligned} \quad (6)$$

where $\mathbf{s}_{1t} = (1, t, t^2, t^3)'$, $\boldsymbol{\lambda}_1 = (\lambda_{10}, \dots, \lambda_{13})'$, $\boldsymbol{\varphi}_1 = (\varphi_{10}, \dots, \varphi_{13})'$, u_t^* is an error term adjusted with respect to the Taylor expansions such that $u_t^* = u_t$ holds whenever $\gamma = 0$, $\mathbf{s}_{2t} = (1, t, \dots, t^4)'$, $\boldsymbol{\lambda}_2 = (\lambda_{20}, \dots, \lambda_{24})'$, $\boldsymbol{\varphi}_2 = (\varphi_{20}, \dots, \varphi_{23})'$, and $\boldsymbol{\lambda}_3 = (\lambda_{30}, \dots, \lambda_{34})'$. The corresponding auxiliary null hypotheses are given by

$$\begin{aligned} H_{01}^{aux} &: \lambda_{1i} = 0 \quad \forall i, \quad \varphi_{10} = 1, \quad \varphi_{1j} = 0, \quad j \geq 1, \\ H_{02}^{aux} &: \lambda_{20} \in \mathbb{R}, \quad \lambda_{2i} = 0, \quad i \geq 1, \quad \varphi_{20} = 1, \quad \varphi_{2j} = 0 \quad j \geq 1, \\ H_{03}^{aux} &: \lambda_{30} \in \mathbb{R}, \quad \lambda_{3i} = 0, \quad i \geq 1, \quad \varphi_{30} = 1. \end{aligned} \quad (7)$$

³Relaxing the assumption that all transitions are set equal in the transition function (2) will not change the set of explanatory variables in H_{a1}^{aux} , H_{a2}^{aux} , and H_{a3}^{aux} . However, the induced parameters in the auxiliary regression will not be the same in terms of the original parameters in the models in H_{a1} , H_{a2} , and H_{a3} .

It should be pointed out that the auxiliary regression model for the LSTD(1) model given by H_{a3}^{aux} is the same type of regression model as in Ouliaris, Park, and Phillips (1989), but in addition they allow for a trend of arbitrary order. Applying the same order of the time trends, (i.e. trends up to the fourth-order) we note that their null hypothesis would correspond to our null hypothesis H_{03}^{aux} but where the restrictions on $\lambda_{31}, \dots, \lambda_{33}$ are relaxed. In this context the null hypothesis is a random walk where the drift is characterized by a third order polynomial in t , and with implications that are to be stressed later on.

To proceed, the auxiliary null hypotheses H_{02}^{aux} and H_{03}^{aux} imply that the data generating processes (DGP's) are random walks with drift. This leads to problems with multicollinearity in large samples since y_{t-1} is then asymptotically equivalent to $t-1$, an explanatory variable already included in the regression equations in H_{a2}^{aux} and H_{a3}^{aux} . We avoid this problem by the transformation $\xi_t \equiv y_t - \lambda_{m0}t$ for $m = 2, 3$ and where $\lambda_{m0}t = E_{H_{0m}^{aux}}[y_t] = \pi_{10}t$, see e.g. Sims, Stock, and Watson (1990), Hamilton (1994), and He and Sandberg (2005a) for details. These transformations give rise to the new set of auxiliary regression equations

$$\tilde{H}_{a2}^{aux} : y_t = \mathbf{s}'_{2t} \boldsymbol{\lambda}_2^* + (\xi_{t-1} \mathbf{s}_{1t})' \boldsymbol{\varphi}_2 + u_t^*, \quad (8)$$

$$\tilde{H}_{a3}^{aux} : y_t = \mathbf{s}'_{2t} \boldsymbol{\lambda}_3^* + \xi_{t-1} \varphi_{30} + u_t^*, \quad (9)$$

where $\boldsymbol{\lambda}_2^* = (\lambda_{20}^*, \dots, \lambda_{24}^*)'$, $\xi_{t-1} = y_{t-1} - \pi_{10}(t-1)$, and $\boldsymbol{\lambda}_3^* = (\lambda_{30}^*, \dots, \lambda_{34}^*)'$. The null hypotheses H_{a2}^{aux} and H_{a3}^{aux} transform further and yield

$$\tilde{H}_{02}^{aux} : \lambda_{21}^* \in \mathbb{R}, \quad \lambda_{2i}^* = 0, \quad i \neq 1, \quad \varphi_{20} = 1, \quad \varphi_{2j} = 0, \quad j \geq 1, \quad (10)$$

$$\tilde{H}_{03}^{aux} : \lambda_{31}^* \in \mathbb{R}, \quad \lambda_{3i}^* = 0, \quad i \neq 1, \quad \varphi_{30} = 1. \quad (11)$$

It is now clear that imposing the parameter restrictions given in (7), (10), and (11) on the corresponding auxiliary regression equations (6), (8) and (9), respectively, the DGP's $y_t = y_{t-1} + u_t$ for $m = 1$, and $\xi_t = \xi_{t-1} + u_t$ for $m = 2, 3$, are obtained. In other words, we end up with a random walk without a drift for *all* cases with the same error term as in (1). The inference about unit roots in the LSTAR(1), LSTART(1), and LSTD(1) models is now based on the NDF testing equations given in H_{a1}^{aux} , \tilde{H}_{a2}^{aux} , and \tilde{H}_{a3}^{aux} , respectively.

3.2 The NADF testing equations

Assuming that $p \geq 2$ means that the LSTAR(p), LSTART(p) and LSTD(p) models are in focus. In these augmented models we shall test for a single unit root. The desired testing situations is summarized as follows

$$\begin{aligned} H_{01}^a : & \quad (1 - \sum_{i=1}^p \pi_{1i} L^i) y_t = u_t, \\ H_{a1}^a : & \quad y_t = \pi_{10} + \pi_{11} y_{t-1} + \dots + \pi_{1,p} y_{t-p} \\ & \quad + (\pi_{20} + \pi_{21} y_{t-1} + \dots + \pi_{2,p} y_{t-p}) F(t, \boldsymbol{\phi}) + u_t, \end{aligned}$$

$$\begin{aligned}
H_{0m}^a &: (1 - \sum_{i=1}^p \pi_{1i} L^i) y_t = \pi_{10} + u_t, \quad m = 2, 3, \\
H_{a2}^a &: y_t = \pi_{10} + \pi_{11} y_{t-1} + \dots + \pi_{1,p} y_{t-p} + \pi_{1,p+1} t \\
&\quad + (\pi_{20} + \pi_{21} y_{t-1} + \dots + \pi_{2,p} y_{t-p} + \pi_{2,p+1} t) F(t, \phi) + u_t, \\
H_{a3}^a &: y_t = \pi_{10} + \pi_{11} y_{t-1} + \dots + \pi_{1,p} y_{t-p} + \pi_{1,p+1} t \\
&\quad + (\pi_{20} + \pi_{2,p+1} t) F(t, \phi) + u_t,
\end{aligned}$$

where L denotes the lag operator. As before, the models under the alternative nest the models under the null hypotheses and the desirable tests could be carried out by imposing the restrictions $\gamma = 0$ and $\pi_{10} = 0$ on H_{a1}^a , and $\gamma = 0$ and $\pi_{10} \in \mathbb{R}$ on H_{a2}^a and H_{a3}^a , respectively. In addition, we assume, for all m , that $\{y_t\}_{t=1}^{\infty}$ contains a single unit root under the null hypothesis meaning that the characteristic polynomial $c(z) = 1 - \sum_{i=1}^p \pi_{1i} z^i = 0$ has a single unit root and all other roots lie outside the unit circle. Furthermore, the restriction $\gamma = 0$ implies the same kind of identification problem that was mentioned in Section 3.1, and is circumvented in the same way. In other words, applying (5) to the models represented in H_{a1}^a , H_{a2}^a , and H_{a3}^a , respectively, yields the augmented auxiliary regression equations

$$\begin{aligned}
H_{a1}^{a,aux} &: y_t = \mathbf{s}'_{1t} \boldsymbol{\lambda}_1^a + (\mathbf{s}_{1t} \otimes \mathbf{y}_{-p})' \boldsymbol{\varphi}_1^a + u_t^*, \\
H_{a2}^{a,aux} &: y_t = \mathbf{s}'_{2t} \boldsymbol{\lambda}_2^a + (\mathbf{s}_{1t} \otimes \mathbf{y}_{-p})' \boldsymbol{\varphi}_2^a + u_t^*, \\
H_{a3}^{a,aux} &: y_t = \mathbf{s}'_{2t} \boldsymbol{\lambda}_3^a + \mathbf{y}_{-p} \boldsymbol{\varphi}_3^a + u_t^*,
\end{aligned}$$

where $\boldsymbol{\lambda}_1^a = (\lambda_{10}^a, \dots, \lambda_{13}^a)'$, $\mathbf{y}_{-p} = (y_{t-1}, \dots, y_{t-p})'$ is a $p \times 1$ vector, $\boldsymbol{\varphi}_1^a = (\varphi_{10}^a, \dots, \varphi_{13}^a)'$ and $\boldsymbol{\varphi}_{1n}^a = (\varphi_{1n1}^a, \dots, \varphi_{1np}^a)'$ where $n = 0, \dots, 3$, u_t^* is an error term adjusted with respect to the Taylor expansions such that $u_t^* = u_t$ holds whenever $\gamma = 0$, $\boldsymbol{\lambda}_2^a = (\lambda_{20}^a, \dots, \lambda_{24}^a)'$, $\boldsymbol{\varphi}_2^a = (\varphi_{20}^a, \dots, \varphi_{23}^a)'$ and $\boldsymbol{\varphi}_{2n}^a = (\varphi_{2n1}^a, \dots, \varphi_{2np}^a)'$ where $n = 0, \dots, 3$, $\boldsymbol{\lambda}_3^a = (\lambda_{30}^a, \dots, \lambda_{34}^a)'$, and $\boldsymbol{\varphi}_3^a = (\varphi_{301}^a, \dots, \varphi_{30p}^a)'$.

Letting $\Delta y_t = y_t - y_{t-1}$, more convenient representations of the auxiliary regression equations above are found as follows.

Lemma 1 (i) *Assuming $p \geq 2$, then the models in $H_{a1}^{a,aux}$ and $H_{a2}^{a,aux}$ can be represented by*

$$y_t = \mathbf{s}'_{1t} \boldsymbol{\lambda}_1^a + (y_{t-1} \mathbf{s}_{1t})' \boldsymbol{\rho}_1^a + (\mathbf{s}_{1t} \otimes \Delta \mathbf{y}_t)' \boldsymbol{\zeta}_1^a + u_t^*, \quad (12)$$

$$y_t = \mathbf{s}'_{2t} \boldsymbol{\lambda}_2^a + (y_{t-1} \mathbf{s}_{2t})' \boldsymbol{\rho}_2^a + (\mathbf{s}_{1t} \otimes \Delta \mathbf{y}_t)' \boldsymbol{\zeta}_2^a + u_t^*, \quad (13)$$

where $\boldsymbol{\rho}_m^a = (\rho_{m0}^a, \dots, \rho_{m3}^a)'$, $\Delta \mathbf{y}_t = (\Delta y_{t-1}, \dots, \Delta y_{t-p+1})'$ is a $(p-1) \times 1$ vector, $\boldsymbol{\zeta}_m^a = (\zeta_{m0}^a, \dots, \zeta_{m3}^a)'$ with $\boldsymbol{\zeta}_{mn}^a = (\zeta_{mn1}^a, \zeta_{mn2}^a, \dots, \zeta_{mn,p-1}^a)'$ where $n = 0, \dots, 3$. Furthermore, the parameter vectors $\boldsymbol{\rho}_m^a$ and $\boldsymbol{\zeta}_m^a$ are defined through the parameter vector $\boldsymbol{\varphi}_m^a$ in

$H_{a1}^{a,aux}$ and $H_{a2}^{a,aux}$ by

$$\begin{aligned}\rho_{mn}^a &= \sum_{q=1}^p \varphi_{mnq}^a, \\ \zeta_{mn1}^a &= -\sum_{q=2}^p \varphi_{mnq}^a, \\ \zeta_{mn2}^a &= -\sum_{q=3}^p \varphi_{mnq}^a, \\ &\vdots \\ \zeta_{mn,p-1}^a &= -\varphi_{mnp}^a.\end{aligned}$$

(ii) Assuming $p \geq 2$, then the model in $H_{a3}^{a,aux}$ can be represented by

$$y_t = \mathbf{s}'_{2t} \boldsymbol{\lambda}_3^a + y_{t-1} \rho_{30}^a + \Delta \mathbf{y}'_t \boldsymbol{\zeta}_3^a + u_t^*, \quad (14)$$

where $\boldsymbol{\zeta}_3^a = (\zeta_{301}^a, \zeta_{302}^a, \dots, \zeta_{30,p-1}^a)$. Moreover, ρ_{30}^a and the parameter vector $\boldsymbol{\zeta}_3^a$ are defined through the parameter vector $\boldsymbol{\varphi}_3^a$ in $H_{a3}^{a,aux}$ by

$$\begin{aligned}\rho_{30}^a &= \sum_{q=1}^p \varphi_{30q}^a, \\ \zeta_{301}^a &= -\sum_{q=2}^p \varphi_{30q}^a, \\ \zeta_{302}^a &= -\sum_{q=3}^p \varphi_{30q}^a, \\ &\vdots \\ \zeta_{30,p-1}^a &= -\varphi_{30p}^a.\end{aligned}$$

Proof. See Appendix A. ■

In Lemma 1 it is seen that the two first terms on the right-hand sides of (12)-(14) are the same set of explanatory variables as in the case $p = 1$. The augmented auxiliary regression equations are therefore obtained by adding lagged differences of the dependent variable multiplied by time trends up to order three to the auxiliary regression equations in H_{a1}^{aux} , H_{a2}^{aux} , and H_{a3}^{aux} , respectively. That the characteristic polynomial $c(z) = 1 - \sum_{i=1}^p \pi_{1i} z^i = 0$ only contains a single unit root corresponds now to $\rho_{m0}^a = 1$ and that the roots of $c(z) = 1 - \sum_{q=1}^{p-1} \zeta_{m0q}^a z^q = 0$ lie outside the unit circle. The null hypotheses can be expressed as

$$\begin{aligned}H_{01}^{a,aux} : & \lambda_{1i}^a = 0 \quad \forall i, \quad \rho_{10}^a = 1, \quad \rho_{1j}^a = 0, \quad j \geq 1, \\ & \zeta_{10q}^a \in \mathbb{R}, \quad \zeta_{1nq}^a = 0, \quad n \geq 1, \quad q = 1, \dots, p-1, \\ H_{02}^{a,aux} : & \lambda_{20}^a \in \mathbb{R}, \quad \lambda_{2i}^a = 0, \quad i \geq 1, \quad \rho_{20}^a = 1, \quad \rho_{2j}^a = 0, \quad j \geq 1, \\ & \zeta_{20q}^a \in \mathbb{R}, \quad \zeta_{2nq}^a = 0, \quad n \geq 1, \quad q = 1, \dots, p-1, \\ H_{03}^{a,aux} : & \lambda_{30}^a \in \mathbb{R}, \quad \lambda_{3i}^a = 0, \quad i \geq 1, \quad \rho_{30}^a = 1, \\ & \zeta_{30q}^a \in \mathbb{R}, \quad q = 1, \dots, p-1.\end{aligned} \quad (15)$$

Imposing the restrictions in $H_{01}^{a,aux}$, $H_{02}^{a,aux}$, and $H_{03}^{a,aux}$ on (12), 13 and (14), respectively, yields

$$\begin{aligned} \left(1 - \sum_{q=1}^{p-1} \zeta_{10q}^a L^q\right) \Delta y_t &= u_t, \\ \left(1 - \sum_{q=1}^{p-1} \zeta_{m0q}^a L^q\right) \Delta y_t &= \lambda_{m0}^a + u_t, \quad m = 2, 3. \end{aligned}$$

Because the roots of $c(z) = 1 - \sum_{q=1}^{p-1} \zeta_{m0q}^a z^q = 0$ fall outside the unit circle, it follows that

$$\begin{aligned} H_{01}^{a,aux} &: \Delta y_t = v_t, \\ H_{0m}^{a,aux} &: \Delta y_t = \mu_m + v_t, \quad m = 2, 3, \end{aligned}$$

where $v_t = (1 - \zeta_{m01}^a L - \zeta_{m02}^a L^2 - \dots - \zeta_{m0,p-1}^a L^{p-1})^{-1} u_t$ and $\mu_m = \lambda_{m0}^a / (1 - \zeta_{m01}^a - \zeta_{m02}^a - \dots - \zeta_{m0,p-1}^a)$. For $m = 2, 3$, the models under $H_{0m}^{a,aux}$ are random walks with drift, and again we encounter a problem with multicollinearity in the auxiliary regression equations (13) and (14). The problem is circumvented by defining $\xi_t \equiv y_t - \mu_m t$ and $\Delta \xi_t \equiv \Delta y_t - \mu_m$ where $\mu_m t = E_{H_{0m}^{a,aux}}[y_t]$ and $\mu_m = E_{H_{0m}^{a,aux}}[\Delta y_t]$. These transformations imply that (13) and (14) can be expressed as

$$\tilde{H}_{a2}^{a,aux} : y_t = \mathbf{s}'_{2t} \boldsymbol{\lambda}_2^{a*} + (\xi_{t-1} \mathbf{s}_{1t})' \boldsymbol{\rho}_2^a + (\mathbf{s}_{1t} \otimes \Delta \boldsymbol{\xi}_t)' \boldsymbol{\zeta}_2^a + u_t^*, \quad (16)$$

$$\tilde{H}_{a3}^{a,aux} : y_t = \mathbf{s}'_{2t} \boldsymbol{\lambda}_3^{a*} + \xi_{t-1} \rho_{30}^a + \Delta \boldsymbol{\xi}_t' \boldsymbol{\zeta}_3^a + u_t^*, \quad (17)$$

where $\boldsymbol{\lambda}_2^{a*} = (\lambda_{20}^{a*}, \dots, \lambda_{24}^{a*})'$, $\boldsymbol{\lambda}_3^{a*} = (\lambda_{30}^{a*}, \dots, \lambda_{34}^{a*})'$, and $\Delta \boldsymbol{\xi}_t = (\Delta \xi_{t-1}, \dots, \Delta \xi_{t-p+1})'$. The null hypotheses transform to

$$\begin{aligned} \tilde{H}_{02}^{a,aux} : \lambda_{21}^{a*} \in \mathbb{R}, \lambda_{2i}^{a*} = 0, i \neq 1, \rho_{20}^a = 1, \rho_{2j}^a = 0, j \geq 1, \\ \zeta_{20q}^a \in \mathbb{R}, \zeta_{2nq}^a = 0, n \geq 1, q = 1, \dots, p-1, \end{aligned} \quad (18)$$

$$\begin{aligned} \tilde{H}_{03}^{a,aux} : \lambda_{31}^{a*} \in \mathbb{R}, \lambda_{3i}^{a*} = 0, i \neq 0, \rho_{30}^a = 1, \\ \zeta_{30q}^a \in \mathbb{R}, q = 1, \dots, p-1. \end{aligned} \quad (19)$$

Applying the parameter restrictions (15), (18), and (19) to (12), (16), and (17), respectively, a random walk *without* drift for all m is obtained. Specifically, $m = 1$ implies $y_t = y_{t-1} + v_t$, and $m = 2, 3$ gives that $\xi_t = \xi_{t-1} + v_t$. The difference between the random walks derived in this section and the random walks in Section 3.1 is that $\{v_t\}$ defines a sequence of serially correlated errors, whereas $\{u_t\}$ presumably is an i.i.d. sequence. To this end, notice that the inference about unit roots for the models in H_{a1}^a , H_{a2}^a , and H_{a3}^a is now based on the NADF testing equations in $H_{a1}^{a,aux}$, $\tilde{H}_{a2}^{a,aux}$, and $\tilde{H}_{a3}^{a,aux}$, respectively.

3.3 The NPADF testing equations

The NADF testing equations presented above contain rather many regressors and increase rapidly, especially for $m = 1, 2$, with p . It is therefore interesting to consider partially augmented testing equations. For $m = 1, 2$, obvious choices are to consider the following NPADF testing equations

$$H_{a1}^{pa,aux} : y_t = \mathbf{s}'_{1t} \boldsymbol{\lambda}_1^{pa} + (y_{t-1} \mathbf{s}_{1t})' \boldsymbol{\rho}_1^{pa} + \Delta \mathbf{y}'_t \boldsymbol{\zeta}_1^{pa} + u_t^*, \quad (20)$$

$$\tilde{H}_{a2}^{pa,aux} : y_t = \mathbf{s}'_{2t} \boldsymbol{\lambda}_2^{pa*} + (\xi_{t-1} \mathbf{s}_{1t})' \boldsymbol{\rho}_2^{pa} + \Delta \boldsymbol{\xi}'_t \boldsymbol{\zeta}_2^{pa} + u_t^*, \quad (21)$$

where $\lambda_1^{pa} = (\lambda_{10}^{pa}, \dots, \lambda_{13}^{pa})'$, $\lambda_2^{pa*} = (\lambda_{20}^{pa*}, \dots, \lambda_{24}^{pa*})'$, $\rho_m^{pa} = (\rho_{m0}^{pa}, \dots, \rho_{m3}^{pa})'$, $\zeta_m^{pa} = (\zeta_{m01}^{pa}, \zeta_{m02}^{pa}, \dots, \zeta_{m0,p-1}^{pa})'$, and u_t^* is an error term where $u_t^* = u_t$ holds when $\gamma = 0$. By (20) and (21) we keep the first two parts on the right-hand side of (12) and (16) (i.e. keeping the same set of regressors that are obtained in the case $p = 1$), but all terms with lagged differences of the dependent variables that are multiplied by a time trend are excluded. Using this approach, the number of reduced regressors equals $3(p - 1)$. It is also evident that with this choice of NPADF testing equations, the equation in (17) qualifies both as a NADF and NPADF testing equation. Furthermore, the relevant null hypotheses are given by

$$H_{01}^{pa,aux} : \lambda_{1i}^{pa} = 0 \quad \forall i, \quad \rho_{10}^{pa} = 1, \quad \rho_{1j}^{pa} = 0, \quad j \geq 1, \\ \zeta_{10q}^{pa} \in \mathbb{R}, \quad q = 1, \dots, p - 1, \quad (22)$$

$$\tilde{H}_{02}^{pa,aux} : \lambda_{2i}^{pa*} \in \mathbb{R}, \quad \lambda_{2i}^{pa*} = 0, \quad i \neq 1, \quad \rho_{20}^{pa} = 1, \quad \rho_{2j}^{pa} = 0, \quad j \geq 1, \\ \zeta_{20q}^{pa} \in \mathbb{R}, \quad q = 1, \dots, p - 1. \quad (23)$$

Imposing the restrictions in (20) and (21) on (22) and (23), respectively, we obtain the same DGP's under the null hypotheses as in Section 3.2, i.e. $y_t = y_{t-1} + v_t$ and $\xi_t = \xi_{t-1} + v_t$, respectively.

4 The NDF, NADF, and NPADF unit root tests

In this section several unit root tests based on the NDF, NADF, and NPADF testing equations in the previous section are derived. The unit root tests will therefore be denoted as the NDF, NADF, and NPADF unit root tests.

All theoretical results presented below are derived under the assumption that the errors u_t in (1) are i.i.d. yielding limiting distributions for the test statistics that are nuisance parameter free (or that the nuisance parameters are trivially eliminated). Although, similar results can be established under the assumption that the error term in (1) is a strong mixing, the inference will be affected by nuisance parameters depending on the mixing assumption, see e.g. Phillips (1987), Phillips and Perron (1988), and He and Sandberg (2005a). As a remark on notation in the following theorems and corollaries, \xrightarrow{d} and \xrightarrow{p} denote convergence in distribution and probability, respectively, and $B(r)$ denotes a standard Brownian motion defined on $[0, 1]$.

Assumption 1 Let $\{u_t\}_{t=1}^{\infty}$ be an i.i.d. sequence of random variables defined on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ such that $E(u_t) = 0$, $E(u_t^2) = \sigma_u^2$, and $E(u_t^4) < \infty$ hold for all t .

Theorem 2 Consider models (6), (8), and (9) when (7), (10), and (11) hold, respectively. Furthermore, assume that $\{u_t\}_{t=1}^{\infty}$ fulfills Assumption 1. Then, for $m = 1, 2, 3$,

$$\gamma_{m1}(\hat{\psi}_m - \psi_m) \xrightarrow{d} \Psi_m^{-1} \mathbf{\Pi}_m, \quad (24)$$

and

$$\hat{\psi}_m - \psi_m \xrightarrow{p} \mathbf{0},$$

where

$$\gamma_{11} = \text{diag}\{\mathbf{T}_1\}, \quad \mathbf{T}_1 = \begin{bmatrix} T^{1/2} & T^{3/2} & T^{5/2} & T^{7/2} & T & T^2 & T^3 & T^4 \end{bmatrix}, \\ \gamma_{21} = \text{diag}\{\mathbf{T}_2\}, \quad \mathbf{T}_2 = \begin{bmatrix} T^{1/2} & T^{3/2} & T^{5/2} & T^{7/2} & T^{9/2} & T & T^2 & T^3 & T^4 \end{bmatrix}, \\ \gamma_{31} = \text{diag}\{\mathbf{T}_3\}, \quad \mathbf{T}_3 = \begin{bmatrix} T^{1/2} & T^{3/2} & T^{5/2} & T^{7/2} & T^{9/2} & T \end{bmatrix},$$

and $\hat{\boldsymbol{\psi}}_1 = (\hat{\boldsymbol{\lambda}}'_1, \hat{\boldsymbol{\varphi}}'_1)'$, $\hat{\boldsymbol{\psi}}_2 = (\hat{\boldsymbol{\lambda}}'^*_2, \hat{\boldsymbol{\varphi}}'_2)'$, and $\hat{\boldsymbol{\psi}}_3 = (\hat{\boldsymbol{\lambda}}'^*_3, \hat{\boldsymbol{\varphi}}_{30})'$ are the least square estimators of $\boldsymbol{\psi}_1 = (\boldsymbol{\lambda}'_1, \boldsymbol{\varphi}'_1)'$, $\boldsymbol{\psi}_2 = (\boldsymbol{\lambda}'^*_2, \boldsymbol{\varphi}'_2)'$, and $\boldsymbol{\psi}_3 = (\boldsymbol{\lambda}'^*_3, \boldsymbol{\varphi}_{30})'$ in (6), (8), and (9), respectively. Moreover, in (24),

$$\boldsymbol{\Psi}_m = \begin{bmatrix} \mathbf{A}_m & \sigma_u \mathbf{B}_m \\ \sigma_u \mathbf{B}'_m & \sigma_u^2 \mathbf{C}_m \end{bmatrix}, \quad \boldsymbol{\Pi}_m = \begin{bmatrix} \sigma_u \mathbf{D}_m \\ \sigma_u^2 \mathbf{E}_m \end{bmatrix},$$

where

$$\begin{aligned} \mathbf{A}_1 &= [a_{ij}]_{4 \times 4}, & \mathbf{B}_1 &= [b_{ij}]_{4 \times 4}, & \mathbf{C}_1 &= [c_{ij}]_{4 \times 4}, & \mathbf{D}_1 &= [d_i]_{4 \times 1}, & \mathbf{E}_1 &= [e_i]_{4 \times 1}, \\ \mathbf{A}_2 &= [a_{ij}]_{5 \times 5}, & \mathbf{B}_2 &= [b_{ij}]_{5 \times 4}, & \mathbf{C}_2 &= \mathbf{C}_1, & \mathbf{D}_2 &= [d_i]_{5 \times 1}, & \mathbf{E}_2 &= \mathbf{E}_1, \\ \mathbf{A}_3 &= \mathbf{A}_2, & \mathbf{B}_3 &= [b_{ij}]_{5 \times 1}, & \mathbf{C}_3 &= [c_{ij}]_{1 \times 1}, & \mathbf{D}_3 &= \mathbf{D}_2, & \mathbf{E}_3 &= [e_i]_{1 \times 1}, \end{aligned}$$

with

$$\begin{aligned} a_{ij} &= 1/(i+j-1), \\ b_{ij} &= \int_0^1 r^{i+j-2} B(r) dr, \\ c_{ij} &= \int_0^1 r^{i+j-2} B(r)^2 dr, \\ d_i &= B(1) - (i-1) \int_0^1 r^{i-2} B(r) dr, \\ e_i &= \left(B(1)^2 - (i-1) \int_0^1 r^{i-2} B(r)^2 dr - 1/i \right) / 2. \end{aligned}$$

Proof. See Appendix A. ■

The joint limiting distributions in (24) contain the nuisance parameter σ_u . However, it is clear that the sub-matrices \mathbf{A}_m , \mathbf{B}_m , \mathbf{C}_m , \mathbf{D}_m , and \mathbf{E}_m are nuisance parameter free.

Corollary 3 (NDF tests) Suppose that the conditions of Theorem 2 hold. Define the test statistic

$$\psi_m \equiv T(\hat{\varphi}_{m0} - 1), \quad m = 1, 2, 3.$$

Then,

$$\psi_m \xrightarrow{d} \mathbf{r}_m \tilde{\boldsymbol{\Psi}}_m^{-1} \tilde{\boldsymbol{\Pi}}_m, \quad (25)$$

where $\mathbf{r}_1 = [r_{1i}]_{1 \times 8}$, $\mathbf{r}_2 = [r_{2i}]_{1 \times 9}$, and $\mathbf{r}_3 = [r_{3i}]_{1 \times 6}$, with elements defined as $r_{15} = r_{26} = r_{36} = 1$ and zero otherwise. Moreover, in (25),

$$\tilde{\boldsymbol{\Psi}}_m = \begin{bmatrix} \mathbf{A}_m & \mathbf{B}_m \\ \mathbf{B}'_m & \mathbf{C}_m \end{bmatrix}, \quad \tilde{\boldsymbol{\Pi}}_m = \begin{bmatrix} \mathbf{D}_m \\ \mathbf{E}_m \end{bmatrix}.$$

Proof. See Appendix A. ■

Corollary 4 (NDF t-type of tests) Suppose that the conditions of Theorem 2 hold. Define the t-test statistic

$$t_m \equiv (\hat{\varphi}_{m0} - 1) / \hat{\sigma}_{\hat{\varphi}_{m0}}, \quad m = 1, 2, 3,$$

where $\hat{\sigma}_{\hat{\varphi}_{m0}}$ is the estimated standard deviation of $\hat{\varphi}_{m0}$. Then

$$t_m \xrightarrow{d} \frac{\mathbf{r}_m \tilde{\boldsymbol{\Psi}}_m^{-1} \tilde{\boldsymbol{\Pi}}_m}{\left(\mathbf{r}_m \tilde{\boldsymbol{\Psi}}_m^{-1} \mathbf{r}'_m \right)^{1/2}}. \quad (26)$$

Proof. See Appendix A. ■

It is evident from (25) and (26) in Corollary 3 and 4, respectively, that both the NDF and the NDF t -type of tests have asymptotic distributions that are nuisance parameter free. Another important property is that for $m = 2, 3$, the tests in (25) and (26) are invariant with respect to if the random walk under the null hypothesis is with or without drift, i.e. the tests are invariant with respect to the parameters λ_{20} and λ_{30} , respectively. For $m = 1$ this is not the case. Moreover, it is only the test based on the LSTD (1) model that are invariant with respect to maintained trends up to order three under the null hypothesis, as noted in Ouliaris, Park, and Phillips (1989). In the present context this means that for $m = 3$, the NDF and the NDF t -type of tests are invariant with respect to $\lambda_{30}, \dots, \lambda_{33}$. However, in our case the restrictions $\lambda_{31} = \dots = \lambda_{33} = 0$ in H_{03}^{aux} are implied restrictions from those that are initially imposed on the LSTD(1) model. Even though the restrictions in this case are not needed in the testing equation in H_{a3}^{aux} , it seems rather *ad hoc* to relax them in terms of the originally state null hypothesis in (4).

Assumption 2 Let $\{u_t\}_{t=1}^{\infty}$ be an i.i.d. sequence of random variables satisfying Assumption 1. Furthermore, define $v_t \equiv c(L)u_t = \sum_{j=0}^{\infty} c_j u_{t-j}$, where $c(L) = \sum_{j=0}^{\infty} c_j L^j$ is a one-sided moving average polynomial in the lag operator such that (A.2.1) $c(1) \neq 0$ (no unit roots), and (A.2.2) $\sum_{j=0}^{\infty} j|c_j| < \infty$ (one-summability).

Due to the earlier assumptions of a single unit root, (A.2.1) in Assumption 2 is satisfied under $H_{0m}^{a,aux}$. The only assumption we have to impose is (A.2.2), i.e. we have to control for the amount of serial dependence in $\{v_t\}_{t=1}^{\infty}$.

Theorem 5 Consider models (12), (16), and (17) when (15), (18), and (19) hold. Furthermore, assume that $\{u_t\}_{t=1}^{\infty}$ and $\{v_t\}_{t=1}^{\infty}$ fulfill Assumption 2. Then, for $m = 1, 2, 3$,

$$\gamma_{m2}^a (\hat{\psi}_m^a - \psi_m^a) \xrightarrow{d} (\Psi_m^a)^{-1} \Pi_m^a, \quad (27)$$

and

$$\hat{\psi}_m^a - \psi_m^a \xrightarrow{p} \mathbf{0},$$

where

$$\begin{aligned} \gamma_{12}^a &= \text{diag} \left\{ \mathbf{T}_1, \left(\left[T^{1/2} \ T^{3/2} \ T^{5/2} \ T^{7/2} \right]' \otimes \mathbf{1} \right)' \right\}, \\ \gamma_{22}^a &= \text{diag} \left\{ \mathbf{T}_2, \left(\left[T^{1/2} \ T^{3/2} \ T^{5/2} \ T^{7/2} \right]' \otimes \mathbf{1} \right)' \right\}, \\ \gamma_{32}^a &= \text{diag} \left\{ \mathbf{T}_3, \left(T^{1/2} \mathbf{1} \right)' \right\}, \end{aligned}$$

and $\mathbf{1}$ is a $(p-1) \times 1$ vector of ones, $\hat{\psi}_1^a = (\hat{\lambda}_1^{a'}, \hat{\rho}_1^{a'}, \hat{\zeta}_1^{a'})'$, $\hat{\psi}_2^a = (\hat{\lambda}_2^{a*'}, \hat{\rho}_2^{a'}, \hat{\zeta}_2^{a'})'$, and $\hat{\psi}_3^a = (\hat{\lambda}_3^{a*'}, \hat{\rho}_{30}^a, \hat{\zeta}_3^{a'})'$ are the least square estimators of $\psi_1^a = (\lambda_1^{a'}, \rho_1^{a'}, \zeta_1^{a'})'$, $\psi_2^a = (\lambda_2^{a*'}, \rho_2^{a'}, \zeta_2^{a'})'$, and $\psi_3^a = (\lambda_3^{a*'}, \rho_{30}^a, \zeta_3^{a'})'$ in (12), (16), and (17), respectively. Moreover, in (27),

$$(\Psi_m^a)^{-1} \Pi_m^a = \begin{bmatrix} \mathbf{P}_m^a \\ \mathbf{Q}_m^a \end{bmatrix},$$

with sub-matrices given by

$$\mathbf{P}_m^a = \begin{bmatrix} \mathbf{A}_m & \lambda \mathbf{B}_m \\ \lambda \mathbf{B}_m' & \lambda^2 \mathbf{C}_m \end{bmatrix}^{-1} \begin{bmatrix} \sigma_u \mathbf{D}_m \\ \sigma_u \lambda \mathbf{E}_m \end{bmatrix}, \quad \mathbf{Q}_m^a \sim N(\mathbf{0}, \sigma_u^2 \mathbf{V}_m^{-1}),$$

where $\lambda = c(1)\sigma_u$, and

$$\mathbf{V}_1 = \mathbf{A}_1 \otimes \Sigma_1, \quad \mathbf{V}_2 = \mathbf{A}_1 \otimes \Sigma_2, \quad \mathbf{V}_3 = \Sigma_3,$$

with

$$\Sigma_m = \begin{bmatrix} \gamma_{m,0} & \gamma_{m,1} & \cdots & \gamma_{m,p-2} \\ \gamma_{m,1} & \gamma_{m,0} & \cdots & \gamma_{m,p-3} \\ \vdots & \vdots & \ddots & \vdots \\ \gamma_{m,p-2} & \gamma_{m,p-3} & \cdots & \gamma_{m,0} \end{bmatrix},$$

where $\gamma_{m,j}$ equals $E[\Delta y_t \Delta y_{t-j}]$ for $m = 1$ and $E[\Delta \xi_t \Delta \xi_{t-j}]$ for $m = 2, 3$.

Proof. See Appendix A. ■

Theorem 6 Consider models (20) and (21) when (22) and (23) hold, respectively. Furthermore, assume that $\{u_t\}_{t=1}^\infty$ and $\{v_t\}_{t=1}^\infty$ fulfill Assumption 2. Then, for $m = 1, 2, 3$,

$$\gamma_{m2}^{pa} (\hat{\psi}_m^{pa} - \psi_m^{pa}) \xrightarrow{d} (\Psi_m^{pa})^{-1} \Pi_m^{pa}, \quad (28)$$

and

$$\hat{\psi}_m^{pa} - \psi_m^{pa} \xrightarrow{p} \mathbf{0},$$

where

$$\gamma_{m2}^{pa} = \text{diag} \left\{ \mathbf{T}_m, \left(T^{1/2} \mathbf{1} \right)' \right\},$$

and $\hat{\psi}_1^{pa} = (\hat{\lambda}_1^{pa}, \hat{\rho}_1^{pa}, \hat{\zeta}_1^{pa})'$ and $\hat{\psi}_2^{pa} = (\hat{\lambda}_2^{pa}, \hat{\rho}_2^{pa}, \hat{\zeta}_2^{pa})'$ are the least square estimators of $\psi_1^{pa} = (\lambda_1^{pa}, \rho_1^{pa}, \zeta_1^{pa})'$ and $\psi_2^{pa} = (\lambda_2^{pa}, \rho_2^{pa}, \zeta_2^{pa})'$ in (20) and (21), respectively. Moreover, in (28),

$$(\Psi_m^{pa})^{-1} \Pi_m^{pa} = \begin{bmatrix} \mathbf{P}_m^{pa} \\ \mathbf{Q}_m^{pa} \end{bmatrix},$$

where

$$\mathbf{P}_m^{pa} = \mathbf{P}_m^a, \quad \mathbf{Q}_m^{pa} \sim N(\mathbf{0}, \sigma_u^2 \Sigma_m^{-1}).$$

Proof. See Appendix A. ■

Note first that we allow for $m = 3$ in Theorem 6. This is only for notational convenience due to the classification about regression models made in Section 3.3, and thus it follows that $\gamma_{32}^{pa} (\hat{\psi}_3^{pa} - \psi_3^{pa}) = \gamma_{32}^a (\hat{\psi}_3^a - \psi_3^a)$, and $(\Psi_3^a)^{-1} \Pi_3^a$ in Theorem 5 equals $(\Psi_3^{pa})^{-1} \Pi_3^{pa}$ in Theorem 6. Furthermore, both the asymptotic distributions (27) and (28) in Theorems 5 and 6 contain the nuisance parameters σ_u and λ .

Corollary 7 (NADF and NPADF type of tests) Suppose that the conditions of Theorems 5 and 6 hold. Define the test statistics,

$$\psi_m^a \equiv T(\hat{\rho}_{m0}^a - 1), \quad \psi_m^{pa} \equiv T(\hat{\rho}_{m0}^{pa} - 1), \quad m = 1, 2, 3.$$

Then,

$$c^* \psi_m^a \xrightarrow{d} \mathbf{r}_m \tilde{\Psi}_m^{-1} \tilde{\Pi}, \quad (29)$$

and

$$c^* \psi_m^{pa} \xrightarrow{d} \mathbf{r}_m \tilde{\Psi}_m^{-1} \tilde{\Pi}, \quad (30)$$

where $c^* = \lambda/\sigma_u$.

Proof. See Appendix A. ■

Especially three interesting properties about the NADF and NPADF type of tests are revealed in Corollary 7. First, although the tests ψ_m^a and ψ_m^{pa} allow for serially correlated errors, their limiting distributions are trivially adjusted to become nuisance parameter free. As such, we instead obtain test statistics depending on nuisance parameters through the quantity c^* . Second, the tests $c^* \psi_m^a$ based on the NADF testing equations have the same asymptotic distribution as the $c^* \psi_m^{pa}$ tests based on the NPADF testing equations. Even though $c^* \psi_m^a$ is a test involving the estimation of a number of $3(p-1)$ more regressors than the $c^* \psi_m^{pa}$ test, the extra set of regressors do not affect the asymptotic distributions. Third, the asymptotic distributions for $c^* \psi_m^a$ and $c^* \psi_m^{pa}$ collapse into the same limiting distributions as for the ψ_m tests with $p=1$ in Corollary 3. This means that the tests $c^* \psi_m^a$ and $c^* \psi_m^{pa}$ are asymptotically invariant with respect to the number of lags included in the NADF and NPADF testing equations.

Corollary 8 (*NADF and NPADF t-type of tests*) Suppose that the conditions of Theorem 5 and 6 hold. Define the t-test statistics,

$$t_m^a \equiv (\hat{\rho}_{m0}^a - 1)/\hat{\sigma}_{\hat{\rho}_{m0}^a}, \quad t_m^{pa} \equiv (\hat{\rho}_{m0}^{pa} - 1)/\hat{\sigma}_{\hat{\rho}_{m0}^{pa}}, \quad m = 1, 2, 3,$$

where $\hat{\sigma}_{\hat{\rho}_{m0}^a}$ and $\hat{\sigma}_{\hat{\rho}_{m0}^{pa}}$ are the estimated standard deviations of $\hat{\rho}_{m0}^a$ and $\hat{\rho}_{m0}^{pa}$, respectively. Then,

$$t_m^a \xrightarrow{d} \frac{\mathbf{r}_m \tilde{\Psi}_m^{-1} \tilde{\Pi}_m}{\left(\mathbf{r}_m \tilde{\Psi}_m^{-1} \mathbf{r}_m'\right)^{1/2}}, \quad (31)$$

and

$$t_m^{pa} \xrightarrow{d} \frac{\mathbf{r}_m \tilde{\Psi}_m^{-1} \tilde{\Pi}_m}{\left(\mathbf{r}_m \tilde{\Psi}_m^{-1} \mathbf{r}_m'\right)^{1/2}}.$$

Proof. See Appendix A. ■

Corollary 8 reveals that the NADF and NPADF t-type of tests mainly have the same asymptotic properties as the NADF and NPADF type of tests in Corollary 7. However, in contrast to the tests in Corollary 7, the limiting distributions for the tests statistics t_m^a and t_m^{pa} need no further manipulations to become nuisances parameter free.

5 Asymptotic and finite-sample properties of the non-linear Dickey-Fuller type of tests

In this section we present the asymptotic as well as finite-sample critical values for some of the tests presented in Section 4. Due to the findings in Corollaries 7 and 8,

the asymptotic critical values for the tests considering partially or fully augmented models are the same. However, from a finite-sample point of view, we choose only to present the critical values for the tests based on partially augmented models since they require estimation of less parameters. Moreover, a size study is performed for a NPADF t -type of test. Finally, the power properties of the tests based on the LSTAR and the LSTART models are examined.

5.1 Asymptotic and finite-sample critical values

When generating the asymptotic distribution of the tests ψ_m , t_m , $c^*\psi_m^{pa}$, and t_m^{pa} , we let $T = 1\,000\,000$ to simulate a Brownian motion $B(r)$ on $[0, 1]$, and the number of replications are set to $1\,000\,000$. The finite-sample critical values for the tests ψ_m and t_m are obtained by simulating data from the model $y_t = y_{t-1} + u_t$ where $u_t \sim \text{nid}(0, 1)$ with desired sample sizes, and thereafter the test statistics in Corollaries 3 and 4 are calculated. This procedure is repeated $1\,000\,000$ times, yielding the finite-sample distributions of the tests. Furthermore, when generating the finite-sample critical values for the test statistics $c^*\psi_m^{pa}$ and t_m^{pa} , we proceed as above but we simulate data from the model $\left(1 - L\zeta_{m01}^{pa} - \dots - L^{p-1}\zeta_{m,0,p-1}^{pa}\right)\Delta y_t = u_t$. Asymptotically this causes no problems, which is shown in Corollaries 7 and 8, but all finite-sample distributions depend on the nuisance parameters $\zeta_{m01}^{pa}, \dots, \zeta_{m,0,p-1}^{pa}$. To proceed we let $\zeta_{m0q}^{pa} = 0$ for all m and q under the null hypothesis, and in the next subsection it is shown that the tests and their finite-sample distributions are rather robust against the values of ζ_{m0q}^{pa} , and the size distortions of the tests become modest. To operationalize $c^*\psi_m^{pa}$ we note that $c^* = 1/\left(1 - \zeta_{m01}^{pa} - \dots - \zeta_{m,0,p-1}^{pa}\right)$, which is consistently estimated by $1/\left(1 - \hat{\zeta}_{m01}^{pa} - \dots - \hat{\zeta}_{m,0,p-1}^{pa}\right)$.⁴

From Table 1 it is seen that the critical values for ψ_m are large in absolute values and negative. The same conclusions are drawn for t_m , however, with critical values being standardized. In fact, the empirical distributions turn out to be heavily skewed to the left with negative means. The reason for this is the inclusion of time trends of high orders creating a downward bias in $\hat{\varphi}_{m0}$. In addition, for $m = 1, 2$ and the ψ_m tests, there are big differences for the critical values among different sample sizes. Furthermore, the critical values for the ψ_2 test are the same as for the T_3 test in He and Sandberg (2005a) and they are there presented in the Table 1.1. In addition, the critical values for the ψ_3 test and the sample size $T = 500$, are the same critical values as for the $K_4(\hat{\alpha})$ test in Ouliaris, Park, and Phillips (1989) and are reported in their Appendix 2.

In Table 2 the critical values for the tests $c^*\psi_m^{pa}$ and t_m^{pa} with $p = 2$ are presented. Critical values for $p = 3, 4$ can be found in Tables 10 and 11 in the Appendix B. Notable is that in small samples the estimates of $\hat{\zeta}_{m01}$ on average are quite inaccurate. This, of course, affects the finite-sample distributions because c^* is biased upwards and thus the critical values become even more negative than corresponding critical values in Table 1. However, as the sample size tends to infinity, the critical values at all significance levels for $c^*\psi_m^{pa}$ and t_m^{pa} equal the critical values for ψ_m and t_m in Table 1 and $T = \infty$, which also confirms the findings in Corollaries 7 and 8.

⁴In our set-up this means that as $T \rightarrow \infty$, $\hat{\zeta}_{m01}^{pa} \rightarrow 0, \dots, \hat{\zeta}_{m,0,p-1}^{pa} \rightarrow 0$ hold for all m , and $p \lim \hat{c}^* = 1$.

Table 1: Critical values for the NDF tests ψ_m and t_m in Corollaries 3 and 4.

T	ψ_1			t_1		
	0.01	0.05	0.10	0.01	0.05	0.10
50	-109.61	-81.81	-68.15	-3.54	-2.82	-2.41
100	-125.97	-92.91	-76.83	-3.64	-2.93	-2.54
250	-142.22	-102.54	-83.87	-3.75	-3.00	-2.61
500	-149.16	-106.44	-86.57	-3.76	-3.04	-2.65
1000	-152.84	-108.44	-87.93	-3.76	-3.04	-2.65
∞	-156.45	-109.40	-89.33	-3.77	-3.05	-2.66
T	ψ_2			t_2		
	0.01	0.05	0.10	0.01	0.05	0.10
50	-119.64	-90.06	-75.56	-3.40	-2.65	-2.26
100	-138.48	-103.72	-86.31	-3.53	-2.80	-2.41
250	-159.44	-116.44	-95.79	-3.58	-2.89	-2.52
500	-168.48	-121.80	-99.73	-3.63	-2.91	-2.54
1000	-173.38	-124.70	-101.80	-3.65	-2.94	-2.56
∞	-175.31	-126.50	-103.11	-3.66	-2.95	-2.57
T	ψ_3			t_3		
	0.01	0.05	0.10	0.01	0.05	0.10
50	-38.94	-32.69	-29.50	-5.38	-4.70	-4.36
100	-43.77	-35.90	-32.07	-5.20	-4.61	-4.30
250	-47.11	-38.13	-33.80	-5.11	-4.55	-4.26
500	-48.27	-38.90	-34.42	-5.08	-4.53	-4.26
1000	-48.31	-39.33	-34.77	-5.07	-4.52	-4.25
∞	-48.50	-39.83	-34.91	-5.05	-4.51	-4.24

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.

Table 2: Critical values for the NPADF tests $c^* \psi_m^{pa}$ and t_m^{pa} in Corollaries 7 and 8 with $p = 2$.

T	$\hat{\zeta}_{101}$	$c^* \psi_1^{pa}$			t_1^{pa}		
		0.01	0.05	0.10	0.01	0.05	0.10
50	0.160	-184.30	-127.99	-103.68	-3.90	-3.11	-2.70
100	0.081	-161.97	-115.78	-94.54	-3.82	-3.08	-2.68
250	0.033	-157.32	-112.16	-91.29	-3.80	-3.07	-2.67
500	0.017	-156.66	-111.33	-90.41	-3.78	-3.06	-2.67
1000	0.008	-156.50	-110.40	-90.01	-3.77	-3.05	-2.66
∞	0.000	-156.46	-109.51	-89.89	-3.77	-3.05	-2.66
T	$\hat{\zeta}_{201}$	$c^* \psi_2^{pa}$			t_2^{pa}		
		0.01	0.05	0.10	0.01	0.05	0.10
50	0.180	-211.89	-148.73	-121.31	-3.80	-3.01	-2.60
100	0.095	-185.66	-133.45	-109.68	-3.72	-2.99	-2.59
250	0.039	-178.53	-128.80	-105.62	-3.69	-2.97	-2.58
500	0.020	-178.00	-128.11	-104.74	-3.68	-2.97	-2.58
1000	0.010	-177.90	-127.63	-103.90	-3.67	-2.96	-2.57
∞	0.000	-175.60	-126.55	-103.15	-3.66	-2.95	-2.57
T	$\hat{\zeta}_{301}$	$c^* \psi_3^{pa}$			t_3^{pa}		
		0.01	0.05	0.10	0.01	0.05	0.10
50	0.127	-66.05	-50.19	-43.21	-5.42	-4.72	-4.38
100	0.064	-56.91	-44.37	-38.71	-5.22	-4.61	-4.31
250	0.025	-52.30	-41.50	-36.40	-5.11	-4.55	-4.26
500	0.013	-50.88	-40.67	-35.81	-5.08	-4.54	-4.25
1000	0.006	-50.19	-40.16	-35.41	-5.07	-4.53	-4.25
∞	0.000	-48.70	-39.87	-34.99	-5.06	-4.51	-4.24

Notes: The probability shown at the head of each column is the area in the left-hand tail. The reported estimates of the nuisance parameters are average OLS estimates from 1 000 000 replications when the true values are set to zero.

As a final remark, the critical values presented in Tables 1 and 2 can be seen as the critical values for unit root tests in LSTAR(p), LSTART(p), and LSTD(p) models with a transition function allowing for three transitions, i.e.

$$F(t) = \frac{1}{1 + \exp\{-\gamma(t - c_1)(t - c_2)(t - c_3)\}} - \frac{1}{2},$$

and where we instead apply a first-order Taylor approximation. The reason for this is that in the present case the same set of explanatory variables as in the cases $H_{a1}^{aux} - H_{a3}^{aux}$, $H_{a1}^{a,aux} - H_{a3}^{a,aux}$, and $H_{a1}^{pa,aux} - H_{a3}^{pa,aux}$, are obtained, and the same null hypotheses will be tested.

5.2 A size study

We examine the size properties of the t_1^{pa} test statistic with $p = 2$ and the influence of the nuisance parameter ζ_{101} . The results are reported in Table 3. The size properties of the tests t_2^{pa} and t_3^{pa} are similar and therefore excluded. Moreover, in small samples the t_m^{pa} tests are preferable to the ψ_m^{pa} type of tests, and the main reason for this is, though not reported here, that in general the size distortions are smaller for the t_m^{pa} tests.⁵

In Table 3 we see that the size distortions are substantial, and this especially for $|\zeta_{101}| \geq 0.8$. The size distortions are mitigated by increasing the sample size, and become negligible for $T \geq 500$ except in the two extreme cases $\zeta_{101} = \pm 0.99$. When the estimated size is larger than 6%, we choose to report size-adjusted critical values, given within the parentheses in Table 3.

Another source of size distortion could arise if we relax our assumption on the error term v_t and allow for some other error structure. It is well known that the classical Dickey and Fuller type of tests suffer from large size distortion (up to 90%) when the error term has a unit root moving average structure, see e.g. Schwert (1989) and He and Sandberg (2005a). A natural approach to this problem is to allow for a wider class of error terms, e.g. by assuming that the error term is a strong mixing. However, as pointed out in Schwert (1989) and He and Sandberg (2005a), this approach does not solve the problem satisfactorily, and the size distortions are not eliminated. It seems better to adopt the instrumental variable approach suggested by Hall (1989), or adding lags of the error term. The latter is an approach that is at hand. In fact, at a 5% nominal level the estimated size becomes 12% for the t_1^{pa} test when $T = 250$ and the error term is an MA(1) process with moving average coefficient equal to -0.8 (not reported here). This should be compared to 99% in a similar testing situation in He and Sandberg (2005a). In general, increasing the order of augmentation further mitigates the size distortions.

5.3 Power studies

The power properties of the t_1 and t_2 tests in Corollary 4 are investigated. The findings for the t_3 test are similar and therefore omitted. It should be pointed out that the power of the ψ_m tests are generally higher than for the t_m tests. However, as argued in the previous subsection, the ψ_m tests may suffer from large size distortions,

⁵The ranking of the tests assumes that one would like to control for the significance levels of the tests. If power would be the criteria, the ψ_m^{pa} tests should, in general, be used.

Table 3: Empirical size of the NPADF t_1^{pa} test in Corollary 8 with $p = 2$.

ζ_{101}	$T = 50$		$T = 100$		$T = 250$	
	$\hat{\zeta}_{101}$	Size	$\hat{\zeta}_{101}$	Size	$\hat{\zeta}_{101}$	Size
0.99	0.59	0.115 (-3.64)	0.75	0.111 (-3.66)	0.88	0.109 (-3.61)
0.8	0.57	0.090 (-3.49)	0.68	0.079 (-3.36)	0.76	0.061 (-3.19)
0.6	0.50	0.066 (-3.30)	0.55	0.063 (-3.23)	0.58	0.053
0.4	0.40	0.056	0.40	0.055	0.40	0.054
0.2	0.28	0.052	0.24	0.052	0.22	0.052
0.0	0.16	0.051	0.08	0.049	0.03	0.050
-0.2	-0.03	0.048	-0.08	0.049	-0.15	0.050
-0.4	-0.10	0.050	-0.25	0.046	-0.33	0.047
-0.6	-0.24	0.048	-0.41	0.050	-0.52	0.049
-0.8	-0.39	0.052	-0.58	0.053	-0.70	0.051
-0.99	-0.57	0.086 (-3.42)	-0.76	0.082 (-3.39)	-0.89	0.078 (-3.36)

ζ_{101}	$T = 500$		$T = 1000$	
	$\hat{\zeta}_{101}$	Size	$\hat{\zeta}_{101}$	Size
0.99	0.94	0.103 (-3.53)	0.97	0.09 (-3.47)
0.8	0.78	0.058	0.79	0.055
0.6	0.59	0.051	0.60	0.047
0.4	0.40	0.048	0.40	0.051
0.2	0.21	0.054	0.20	0.052
0.0	0.02	0.049	0.01	0.051
-0.2	-0.18	0.049	-0.19	0.049
-0.4	-0.36	0.049	-0.38	0.050
-0.6	-0.56	0.050	-0.58	0.053
-0.8	-0.75	0.048	-0.78	0.048
-0.99	-0.94	0.076 (-3.33)	-0.96	0.070 (-3.22)

Notes: The nominal size of the test is 5%. The reported estimates of the nuisance parameters are the average OLS estimates from 1 000 000 replications where the true values are shown in the column to the left. Values in parentheses correspond to size-adjusted critical values.

and consequently, the power results can be misleading. Furthermore, we choose not to report the power results for the augmented tests, but one can expect the main findings to coincide with the results for the t_m tests.

5.3.1 Empirical power for the NDF t_1 type of test when the DGP is an LSTAR(1) model

The power of the t_1 test is examined when the LSTAR(1) model in H_{a1} is the DGP. Without loss of generality the transition function is replaced, as a matter of convenience, by $\tilde{F} \equiv F + 1/2$. It is clear that $\tilde{F}(t) : \mathbb{R}_+ \rightarrow [0, 1]$. In all coming experiments, we impose the Lagrange stability condition,

$$\pi_{11} \in (0, 1), \quad \pi_{11} + \pi_{21} \in (0, 1), \quad (32)$$

on the skeleton of the LSTAR(1) model to rule out unrealistic trajectories (unstable or explosive trajectories), see Tong (1990). The parameters of the LSTAR(1) model are assigned to the following values

$$\begin{aligned} \pi_{10} &= 0, \quad \pi_{20} \in \{1, 2\}, \quad \gamma \in \{0.01, 0.1, 1\}, \\ c &\in \{0.5T, 0.25T, 0.75T\}, \quad T = 250, \end{aligned} \quad (33)$$

and the autoregressive coefficients in the linear and nonlinear part are assigned values according to the scenarios

$$\begin{aligned} \text{Low-High:} \quad & \pi_{11} = 0.10, \quad \pi_{21} \in \{0.75, 0.80, 0.85\}, \\ \text{Medium-High:} \quad & \pi_{11} = 0.40, \quad \pi_{21} \in \{0.45, 0.50, 0.55\}, \\ \text{High-High:} \quad & \pi_{11} = 0.75, \quad \pi_{21} \in \{0.10, 0.15, 0.20\}. \end{aligned} \quad (34)$$

By this design we note that all trajectories start at zero and head towards the long-run attractor $\pi_{20}/(1 - \pi_{11} - \pi_{21})$, as time evolves. Now varying π_{20} , as well as π_{11} and π_{21} , we see that the mean of the LSTAR(1) model at the end of the period ranges from 6.67 to 40 (assuming that a complete transition takes place). How fast the trajectories reach the long-run attractor is determined through γ . When $\gamma = 0.01$ the transition is slow and we encounter an almost linear LSTAR(1) model.⁶ Letting $\gamma = 0.1$, the transition is intermediate, although a sample size of 250 is enough for a complete transition to take place. When $\gamma = 1.00$ the transition is more abrupt and the LSTAR(1) model behaves almost like a TAR(1) process. We also vary the point in time around which the transition will take place. For instance, when $c = 0.25T$, it is illustrative to think that we have an TAR(1) model where the first 25% of the sample is an AR(1) process characterized by π_{11} , and in the remaining 75% of the sample the AR(1) process is characterized by the parameter vector $(\pi_{20}, \pi_{11} + \pi_{21})'$.

Finally, the three scenarios reflect the impact of the nonlinear part in the LSTAR model in terms of dynamics. The Low-High scenario describes an LSTAR model with

⁶When $\gamma = 0.01$ a full transition does not take place during the sample period. Notice that $\tilde{F}(t = T = 250; \gamma = 0.01, c = 62.5) = 0.87$, $\tilde{F}(t = T = 250; \gamma = 0.01, c = 125) = 0.78$, and $\tilde{F}(t = T = 250; \gamma = 0.01, c = 187.5) = 0.66$.

Table 4: (Low-High) Empirical power of the t_1 test and the corresponding Dickey-Fuller t -type of test.

$T = 250$		$c = 0.5T$				$c = 0.25T$				$c = 0.75T$			
		t_1		DF		t_1		DF		t_1		DF	
γ	π_{21}/π_{20}	1	2	1	2	1	2	1	2	1	2	1	2
0.01	0.75	0.54	0.64	1.00	1.00	0.47	0.63	1.00	1.00	0.58	0.64	1.00	1.00
	0.80	0.53	0.64	1.00	1.00	0.47	0.63	1.00	0.97	0.57	0.63	1.00	1.00
	0.85	0.53	0.64	1.00	1.00	0.48	0.64	1.00	0.47	0.57	0.64	1.00	1.00
0.10	0.75	0.97	1.00	0.45	0.00	1.00	1.00	0.47	0.00	0.31	0.24	0.98	0.00
	0.80	0.99	1.00	0.00	0.00	1.00	0.98	0.00	0.00	0.30	0.25	0.47	0.00
	0.85	1.00	1.00	0.00	0.00	0.99	0.95	0.00	0.00	0.30	0.28	0.01	0.00
1.00	0.75	1.00	1.00	0.23	0.00	0.96	0.54	0.00	0.00	0.19	0.33	0.81	0.00
	0.80	1.00	1.00	0.00	0.00	0.87	0.29	0.00	0.00	0.21	0.51	0.05	0.00
	0.85	1.00	1.00	0.00	0.00	0.63	0.14	0.00	0.00	0.34	0.86	0.00	0.00

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

a low value of the autoregressive root in the beginning of the period and reaches a closer to unstable state, $\pi_{11} + \pi_{21} = 0.95$, at the end of the period. The nonlinear impact of dynamics is viewed as high. The Medium-High scenario presents an LSTAR model where the autoregressive root is intermediate in the beginning of the period and reaches a closer to unstable state at the end of the period. The magnitude of nonlinear dynamics is seen as medium. In the High-High scenario the LSTAR model displays an autoregressive root with a high value in the beginning of the period and reaches a state closer to unstableness at the end of the period. The nonlinear change in dynamics is perceived as low.

We choose to compare the power of the t_1 test with the power of the Dickey-Fuller (DF) t -type of test based on the model $y_t = a_0 + a_1 y_{t-1} + u_t$. This will give the opportunity to investigate the bias towards nonrejection for the DF tests when the DGP has a smooth change in levels and dynamics. The results are presented in Tables 4-6.

For $\gamma = 0.01$ it is seen in Tables 4-6 that the DF test, in general, is superior to the t_1 test. This is to be expected since the LSTAR(1) model appears almost linear. The modest power for the t_1 test is explained by the fact that it is penalized due to an ambiguous set of explanatory variables.

When $\gamma = 0.10$ we notify the collapse in power of the DF test under all scenarios. The reason for this is that a full transition actually takes place for all values of c , and that the speed of transition is fast enough to generate truly nonlinear models with a clear shift in levels. The power of the t_1 test is increased and is close to unity under the Low-High and Medium-High scenarios with $c = 0.5T$, $\pi_{20} = 1$, and all values of π_{21} . This illuminates how the power of the t_1 test responds to the timing of the transition and the magnitude of nonlinearities. It also confirms that the linearity part in the LSTAR(1) model dominates both when $c = 0.75T$ or the High-High scenario is concerned, and consequently the t_1 test performs less satisfactorily. Even

Table 5: (Medium-High) Empirical power of the t_1 test and the corresponding Dickey-Fuller t -type of test.

$T = 250$		$c = 0.5T$				$c = 0.25T$				$c = 0.75T$			
		t_1		DF		t_1		DF		t_1		DF	
γ	π_{21}/π_{20}	1	2	1	2	1	2	1	2	1	2	1	2
0.01	0.45	0.27	0.36	1.00	1.00	0.24	0.39	1.00	0.99	0.27	0.34	1.00	1.00
	0.50	0.26	0.37	1.00	0.99	0.25	0.39	1.00	0.60	0.28	0.34	1.00	1.00
	0.55	0.26	0.38	1.00	0.80	0.27	0.40	0.94	0.01	0.27	0.35	1.00	1.00
0.10	0.45	0.77	0.98	0.00	0.00	0.95	0.93	0.00	0.00	0.11	0.08	0.85	0.00
	0.50	0.87	0.99	0.00	0.00	0.95	0.87	0.00	0.00	0.11	0.09	0.16	0.00
	0.55	0.94	0.99	0.00	0.00	0.90	0.78	0.00	0.00	0.11	0.10	0.01	0.00
1.00	0.45	0.95	1.00	0.00	0.00	0.76	0.25	0.00	0.00	0.07	0.12	0.51	0.00
	0.50	0.99	1.00	0.00	0.00	0.61	0.11	0.00	0.00	0.08	0.22	0.01	0.00
	0.55	1.00	1.00	0.00	0.00	0.40	0.06	0.00	0.00	0.13	0.55	0.00	0.00

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

Table 6: (High-High) Empirical power of the t_1 test and the corresponding Dickey-Fuller t -type of test.

$T = 250$		$c = 0.5T$				$c = 0.25T$				$c = 0.75T$			
		t_1		DF		t_1		DF		t_1		DF	
γ	π_{21}/π_{20}	1	2	1	2	1	2	1	2	1	2	1	2
0.01	0.10	0.07	0.13	1.00	0.83	0.08	0.17	1.00	0.86	0.06	0.09	1.00	0.96
	0.15	0.07	0.13	0.96	0.19	0.09	0.17	0.90	0.12	0.06	0.10	1.00	0.69
	0.20	0.09	0.15	0.76	0.00	0.10	0.17	0.20	0.00	0.07	0.11	0.96	0.18
0.10	0.10	0.22	0.53	0.05	0.00	0.35	0.37	0.35	0.00	0.03	0.02	0.40	0.00
	0.15	0.32	0.62	0.00	0.00	0.41	0.35	0.00	0.00	0.02	0.02	0.01	0.00
	0.20	0.46	0.68	0.00	0.00	0.44	0.35	0.00	0.00	0.02	0.02	0.00	0.00
1.00	0.10	0.37	0.67	0.01	0.00	0.21	0.05	0.00	0.00	0.01	0.02	0.17	0.00
	0.15	0.52	0.73	0.00	0.00	0.18	0.03	0.00	0.00	0.01	0.04	0.00	0.00
	0.20	0.65	0.77	0.00	0.00	0.15	0.02	0.00	0.00	0.01	0.10	0.00	0.00

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

though in the present context the nonlinear part and its dynamics are suppressed, the generated trajectories appear nonlinear enough and the power for the DF test is modest. It is also interesting to note that the t_1 test is very robust against changes in the intercept and levels, in contrast to tests based on first-order Taylor approximation, see He and Sandberg (2005a) for the discussion about the level leverage effect.

For $\gamma = 1.00$ the main findings are the same as for $\gamma = 0.10$. Especially note that for $c = 0.25T$ and $\pi_{20} = 2$, the power of the test t_1 decreases to rather low levels due to that (approximately) 75% of the sample is characterized by a linear AR process.

To this end, we conclude that if we are right about the timing of the transition in the sense that it takes place around the middle of the sample ($c = 0.5T$), the t_1 test has very satisfactory power and *increases* in π_{20} and π_{21} , as well as with γ , in all scenarios. An opposite relationship for the power holds for the DF test, and the power is in general close to zero or equal to zero.

5.3.2 Empirical power for the NDF t_2 type of test when the DGP is an LSTART(1) model

The power of the t_2 test presented in Corollary 4 is explored when the DGP is the LSTART(1) model in H_{a2} . The analysis is conducted under the same prerequisites as in Subsection 5.3.1, but the analysis of varying the timing of the transition is excluded and only the case $c = 0.5T$ is in focus. In addition to these prerequisites, the coefficients of the time trend are assigned the following values

$$\pi_{12} = 0.005, \quad \pi_{22} \in \{0.005, 0.0075, 0.01\}.$$

The (linearly time-dependent) long-run attractors implied by this set-up are given by $\pi_{20}/(1 - \pi_{11} - \pi_{21}) + (\pi_{12} + \pi_{22})/(1 - \pi_{11} - \pi_{21})t$, and the slope coefficient ranges from 0.0667 to 0.3 at the end of the sample period (assuming that a complete transition takes place). The power of the t_2 test is compared to the classical t -type of test in Phillips and Perron (1988) denoted PP, and is based on the model $y_t = a_0 + a_1y_{t-1} + a_2t + u_t$. This means that we will investigate the bias towards nonrejection for a test based on a linear trend-stationary model when the DGP accommodates a smooth shift in levels, dynamics, and trends. The power results are presented in Tables 7-9.

For $\gamma = 0.01$ it appears in Tables 7-9 that the PP test outperforms, in general, the t_2 test. This is because the LSTART(1) model acts like a linear trend stationary process. More interesting is the extraordinary robustness in the power of the t_2 test under the Low-High and Medium-High scenarios whenever $\gamma \geq 0.1$ and regardless of how the values of the other parameters are varied. That is, when the nonlinear impact is medium or high, it appears that the design of the t_2 test is very satisfactory, and the probability of rejecting the null hypothesis when it is wrong is unity. The opposite is true for the PP test and the lack of power is evident. Moreover, in the High-High scenario in Table 9, we see that the power is substantial but, however, lower than for the two other scenarios. This emphasizes that the power for the t_2 test is reduced due to a less evident impact of nonlinearity in the High-High scenario.

To this end, we note that the collapse in power for the PP test is even more pronounced than for the DF test in the previous section. It is also seen that the

Table 7: (Low-High) Empirical power of the t_2 test and the corresponding Phillips-Perron t -type of test.

$T = 250$		$\pi_{22} = 0.005$				$\pi_{22} = 0.0075$				$\pi_{22} = 0.01$			
		t_2		PP		t_2		PP		t_2		PP	
γ	π_{21}/π_{20}	1	2	1	2	1	2	1	2	1	2	1	2
0.01	0.75	0.59	0.62	1.00	1.00	0.61	0.63	1.00	1.00	0.63	0.65	1.00	1.00
	0.80	0.59	0.61	1.00	1.00	0.61	0.65	1.00	1.00	0.64	0.65	1.00	1.00
	0.85	0.60	0.63	1.00	1.00	0.64	0.67	1.00	1.00	0.66	0.69	1.00	1.00
0.10	0.75	1.00	1.00	0.00	0.00	1.00	1.00	0.00	0.00	1.00	1.00	0.00	0.00
	0.80	1.00	1.00	0.00	0.00	1.00	1.00	0.00	0.00	1.00	1.00	0.00	0.00
	0.85	1.00	1.00	0.00	0.00	1.00	1.00	0.00	0.00	1.00	1.00	0.00	0.00
1.00	0.75	1.00	1.00	0.00	0.00	1.00	1.00	0.00	0.00	1.00	1.00	0.00	0.00
	0.80	1.00	1.00	0.00	0.00	1.00	1.00	0.00	0.00	1.00	1.00	0.00	0.00
	0.85	1.00	1.00	0.00	0.00	1.00	1.00	0.00	0.00	1.00	1.00	0.00	0.00

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

Table 8: (Medium-High) Empirical power of the t_2 test and the corresponding Phillips-Perron t -type of test.

$T = 250$		$\pi_{22} = 0.005$				$\pi_{22} = 0.0075$				$\pi_{22} = 0.01$			
		t_2		PP		t_2		PP		t_2		PP	
γ	π_{21}/π_{20}	1	2	1	2	1	2	1	2	1	2	1	2
0.01	0.45	0.30	0.33	1.00	1.00	0.32	0.33	1.00	1.00	0.34	0.36	1.00	1.00
	0.50	0.32	0.34	1.00	1.00	0.33	0.37	1.00	0.99	0.37	0.39	1.00	0.97
	0.55	0.33	0.37	0.99	0.92	0.37	0.40	0.93	0.71	0.39	0.42	0.73	0.38
0.10	0.45	1.00	1.00	0.01	0.00	1.00	1.00	0.00	0.00	1.00	1.00	0.00	0.00
	0.50	1.00	1.00	0.01	0.00	1.00	1.00	0.00	0.00	1.00	1.00	0.00	0.00
	0.55	1.00	1.00	0.00	0.00	1.00	1.00	0.00	0.00	1.00	1.00	0.00	0.00
1.00	0.45	1.00	1.00	0.01	0.00	1.00	1.00	0.00	0.00	1.00	1.00	0.00	0.00
	0.50	1.00	1.00	0.00	0.00	1.00	1.00	0.00	0.00	1.00	1.00	0.00	0.00
	0.55	1.00	0.98	0.00	0.00	1.00	0.96	0.00	0.00	0.99	0.01	0.00	0.00

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

Table 9: (High-High) Empirical power of the t_2 test and the corresponding Phillips-Perron t -type of test.

$T = 250$		$\pi_{22} = 0.005$				$\pi_{22} = 0.0075$				$\pi_{22} = 0.01$			
		t_2		PP		t_2		PP		t_2		PP	
γ	π_{21}/π_{20}	1	2	1	2	1	2	1	2	1	2	1	2
0.01	0.10	0.07	0.07	1.00	1.00	0.08	0.08	1.00	1.00	0.09	0.09	1.00	0.98
	0.15	0.08	0.09	0.94	0.87	0.09	0.10	0.82	0.70	0.11	0.11	1.00	0.46
	0.20	0.11	0.11	0.26	0.07	0.13	0.13	0.07	0.01	0.14	0.15	0.73	0.00
0.10	0.10	0.58	0.76	0.14	0.00	0.69	0.81	0.01	0.00	0.77	0.85	0.00	0.00
	0.15	0.76	0.85	0.00	0.00	0.82	0.87	0.00	0.00	0.86	0.89	0.00	0.00
	0.20	0.82	0.83	0.00	0.00	0.83	0.84	0.00	0.00	0.87	0.85	0.00	0.00
1.00	0.10	0.72	0.87	0.00	0.00	0.81	0.88	0.00	0.00	0.86	0.90	0.00	0.00
	0.15	0.86	0.90	0.00	0.00	0.84	0.88	0.00	0.00	0.84	0.86	0.00	0.00
	0.20	0.84	0.70	0.00	0.00	0.80	0.59	0.00	0.00	0.71	0.49	0.00	0.00

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

t_2 test indicates an extraordinary robustness in power, and even though the t_2 test is based on auxiliary regression equations, it handles smooth or abrupt changes in mean, dynamics, and the time trends comfortably.

6 Concluding remarks

In this paper we propose a new model that accommodate a smooth change in mean, trend, and dynamics. The model is a generalization of the LSTAR(p) model introduced in Lin and Teräsvirta (1994) because a nonlinear trend is added. In particular this model should be an appealing alternative when testing the null hypothesis of a random walk with drift.

Several unit root tests are derived in both the LSTAR(p) and the LSTART(p) models. The tests are enhanced by a third-order Taylor approximation, and inference about unit roots is based on the LS estimators obtained from the auxiliary testing equations. The distributional asymptotic theory for the LS estimators is non-standard because of the unit root assumption. In order to find these asymptotic distributions, we generalize many theoretical results that are derived in the econometric unit root literature. In comparison to the related approaches by Leybourne, Newbold, and Vougas (1998) and Harvey and Mills (2002), we note that the nonlinear models under the alternative are similar, but it should be pointed out that their testing methodology is different and no analytical limiting distributions for their unit root tests are provided.

Finite-sample properties are investigated and our tests indicate modest size distortions. The empirical power is compared to the power of the conventional unit root tests by Dickey and Fuller (1979), Phillips (1987), and Phillips and Perron (1988) when the DGP is an LSTAR and LSTART model, respectively. It turns out that the canonical tests only have satisfactorily power when the LSTAR and LSTART

models appear linear, and under these circumstances our tests are penalized due to the many explanatory variables. Assuming more nonlinear LSTAR and LSTART models, the robustness in power of our tests in a wide range of parameter values, due to a third-order Taylor approximation, is rather extraordinary and equal unity in most of the situations. The opposite holds for the conventional tests, and the power is close to or equal to zero.

Appendix A

The proofs given in this Appendix are only for $m = 1$ (i.e. the LSTAR(p) model). The proofs for $m = 2, 3$ are similar and therefore omitted.

Proof of Lemma 1. The proof is straightforward and follows by induction. ■

Proof of Theorem 2. Write

$$\gamma_{11} \left(\hat{\psi}_1 - \psi_1 \right) = \begin{bmatrix} \tilde{\mathbf{A}}_1 & \tilde{\mathbf{B}}_1 \\ \tilde{\mathbf{B}}_1' & \tilde{\mathbf{C}}_1 \end{bmatrix}^{-1} \begin{bmatrix} \tilde{\mathbf{D}}_1 \\ \tilde{\mathbf{E}}_1 \end{bmatrix},$$

where

$$\begin{aligned} \tilde{\mathbf{A}}_1 &= [\tilde{a}_{ij}]_{4 \times 4}, \text{ and } \tilde{a}_{ij} = T^{-(i+j-1)} \sum_{t=1}^T t^{(i+j-2)}, \\ \tilde{\mathbf{B}}_1 &= [\tilde{b}_{ij}]_{4 \times 4}, \text{ and } \tilde{b}_{ij} = T^{-(i+j-1/2)} \sum_{t=1}^T t^{(i+j-2)} y_{t-1}, \\ \tilde{\mathbf{C}}_1 &= [\tilde{c}_{ij}]_{4 \times 4}, \text{ and } \tilde{c}_{ij} = T^{-(i+j)} \sum_{t=1}^T t^{(i+j-2)} y_{t-1}^2, \\ \tilde{\mathbf{D}}_1 &= [\tilde{d}_i]_{4 \times 1}, \text{ and } \tilde{d}_i = T^{-(i-1/2)} \sum_{t=1}^T t^{i-1} u_t, \\ \tilde{\mathbf{E}}_1 &= [\tilde{e}_i]_{4 \times 1}, \text{ and } \tilde{e}_i = T^{-i} \sum_{t=1}^T t^{i-1} y_{t-1} u_t. \end{aligned}$$

By using Lemma 4 in He and Sandberg (2005a) we conclude that the following expressions converge jointly for all i and j :

$$\begin{aligned} \tilde{a}_{ij} &\xrightarrow{d} a_{ij}, & \tilde{b}_{ij} &\xrightarrow{d} \sigma_u b_{ij}, & \tilde{c}_{ij} &\xrightarrow{d} \sigma_u^2 c_{ij}, \\ \tilde{d}_i &\xrightarrow{d} \sigma_u d_i, & \tilde{e}_i &\xrightarrow{d} \sigma_u^2 e_i. \end{aligned}$$

Thus, (24) holds.

Furthermore, $\text{plim}_{T \rightarrow \infty} (\hat{\psi}_1 - \psi_1) = \mathbf{0}$ holds because the element with the slowest rate of convergence in $\hat{\psi}_1$ equals $\mathcal{O}_p(T^{-1/2})$. ■

Proof of Corollary 3. It follows from Theorem 2 and the Slutsky Theorem that

$$\begin{aligned} \psi_1 &\equiv T(\hat{\varphi}_{10} - 1) \\ &= \mathbf{r}_1 \begin{bmatrix} \tilde{\mathbf{A}}_1 & \tilde{\mathbf{B}}_1 \\ \tilde{\mathbf{B}}_1' & \tilde{\mathbf{C}}_1 \end{bmatrix}^{-1} \begin{bmatrix} \tilde{\mathbf{D}}_1 \\ \tilde{\mathbf{E}}_1 \end{bmatrix} \xrightarrow{d} \mathbf{r}_1 \Psi_1^{-1} \Pi_1. \end{aligned}$$

Define the matrix $\mathbf{S}_1 = \text{diag} \left\{ 1 \ 1 \ 1 \ 1 \ \sigma_u \ \sigma_u \ \sigma_u \ \sigma_u \right\}$. Then

$$\begin{aligned} \mathbf{r}_1 \Psi_1^{-1} \Pi_1 &= \mathbf{r}_1 \left(\mathbf{S}_1^{-1} \begin{bmatrix} \mathbf{A}_1 & \mathbf{B}_1 \\ \mathbf{B}_1 & \mathbf{C}_1 \end{bmatrix}^{-1} \mathbf{S}_1^{-1} \right) \left(\sigma_u \mathbf{S}_1 \begin{bmatrix} \mathbf{D}_1 \\ \mathbf{E}_1 \end{bmatrix} \right) \\ &= \mathbf{r}_1 \begin{bmatrix} \mathbf{A}_1 & \mathbf{B}_1 \\ \mathbf{B}_1 & \mathbf{C}_1 \end{bmatrix}^{-1} \begin{bmatrix} \mathbf{D}_1 \\ \mathbf{E}_1 \end{bmatrix}, \end{aligned}$$

because $\sigma_u \mathbf{r}_1 \mathbf{S}_1^{-1} = \mathbf{r}_1$. By Theorem 2, (25) holds for $m = 1$. ■

Proof of Corollary 4. We first note that

$$\begin{aligned} T^2 \hat{\sigma}_{\hat{\varphi}_{10}}^2 &= S_T^2 \mathbf{r}_1 \begin{bmatrix} \tilde{\mathbf{A}}_1 & \tilde{\mathbf{B}}_1 \\ \tilde{\mathbf{B}}_1' & \tilde{\mathbf{C}}_1 \end{bmatrix}^{-1} \mathbf{r}_1' \\ &\xrightarrow{d} \mathbf{r}_1 \tilde{\Psi}_1^{-1} \mathbf{r}_1', \end{aligned}$$

where $S_T^2 = \left(y_t - (\mathbf{s}'_{1t}, (y_{t-1} \mathbf{s}_{1t})') \hat{\psi}_1 \right)^2 / (T - 8)$. From Theorem 2, $plim_{T \rightarrow \infty} S_T^2 = \sigma_u^2$ holds. By Corollary 3 and the Slutsky Theorem we conclude that

$$\begin{aligned} t_1 &\equiv (\hat{\varphi}_{10} - 1) / \hat{\sigma}_{\hat{\varphi}_{10}} \\ &= \frac{T(\hat{\varphi}_{10} - 1)}{(T^2 \hat{\sigma}_{\hat{\varphi}_{10}})^{1/2}} \xrightarrow{d} \frac{\mathbf{r}_1 \tilde{\psi}_1^{-1} \tilde{\Pi}_1}{\left(\mathbf{r}_1 \tilde{\psi}_1^{-1} \mathbf{r}_1' \right)^{1/2}}. \end{aligned}$$

Thus, (26) holds for $m = 1$. ■

To prove Theorem 5, we introduce the following lemma.

Lemma 9 *If $\{u_t\}_{t=1}^\infty$ and $\{v_t\}_{t=1}^\infty$ satisfy Assumptions 1 and 2, respectively, and $\xi_t = \xi_{t-1} + v_t$ with $P(\xi_0 = 0) = 1$, then as $T \rightarrow \infty$*

$$T^{-(p+q/2+1)} \sum_{t=1}^T t^p \xi_{t-1}^q \xrightarrow{d} \lambda^q \int_0^1 r^p W(r)^q dr, \quad (35)$$

$$T^{-(p+1)} \sum_{t=1}^T t^p v_t v_{t-h} \xrightarrow{a.s.} \gamma_h / (p+1), \quad (36)$$

$$T^{-(v+1/2)} \sum_{t=1}^T t^v v_{t-h} \xrightarrow{d} \lambda W(1) - v \lambda \int_0^1 r^{v-1} W(r) dr, \quad (37)$$

$$T^{-(v+1/2)} \sum_{t=1}^T t^v u_t \xrightarrow{d} \sigma_u W(1) - v \sigma_u \int_0^1 r^{v-1} W(r) dr, \quad (38)$$

$$\begin{aligned} &T^{-(p+1)} \sum_{t=1}^T t^p \xi_{t-1} u_t \xrightarrow{d} \\ &0.5 \lambda \sigma_u \left(W(1)^2 - p \int_0^1 r^{p-1} W(r)^2 dr - \frac{1}{(p+1)} \right), \end{aligned} \quad (39)$$

$$\begin{aligned} &T^{-(p+1)} \sum_{t=1}^T t^p \xi_{t-1} v_{t-h} \xrightarrow{d} \\ &(i) \begin{cases} 0.5 \left(\lambda^2 W(1)^2 - p \lambda^2 \int_0^1 r^{p-1} W(r)^2 dr - \frac{\gamma_0}{(p+1)} \right), \\ \text{for } h = 0, \end{cases} \end{aligned} \quad (40)$$

$$(ii) \begin{cases} 0.5 \left(\lambda^2 W(1)^2 - p \lambda^2 \int_0^1 W(r)^2 r^{p-1} dr - \frac{\gamma_0}{(p+1)} \right) + \frac{\sum_{s=0}^{h-1} \gamma_s}{p+1}, \\ \text{for } h > 0, \end{cases}$$

$$T^{-(p+1/2)} \sum_{t=1}^T t^p v_{t-h} u_t \xrightarrow{d} N \left(0, \frac{\gamma_0 \sigma_u^2}{2p+1} \right), \quad h > 0, \quad (41)$$

where $\lambda = c(1)\sigma_u$, $\gamma_h = \sigma_u^2 \sum_{j=0}^{\infty} c_j c_{j+h}$, $h, v, p \geq 0$, and $q \geq 1$.

Note that Lemma 9 gives more general results than needed to prove Theorem 5. In fact, it enables us to derive the limiting distribution for the NADF and NPADF tests in the LSTAR(p), LSTART(p), and the LSTD(p) models based on any order of approximation of a logistic transition function that accommodates an arbitrary number of transitions. Lemma 9 also generalizes many of the results derived in Hamilton (1994).

Proof of (35) in Lemma 9. Define the following cadlag function on $D[0, 1]$,

$$W_t(r, \omega) = \frac{1}{\sigma_u \sqrt{T}} \xi_{[Tr]} = \frac{1}{\sigma_u \sqrt{T}} \xi_{t-1}, \quad r \in [\frac{t-1}{T}, \frac{t}{T}), \quad t = 1, \dots, T,$$

where $\xi_t = \xi_{t-1} + v_t$. Using the Beveridge and Nelson (1981) decomposition we conclude that $W_t \xrightarrow{d} c(1)W$, where W is a standard Brownian motion on $[0, 1]$. We can rely upon the proof for (18) in Lemma 4 in He and Sandberg (2005a), where we replace σ with $c(1)\sigma_u$. ■

Proof of (36) in Lemma 9. By Assumption 2

$$T^{-1} \sum_{t=1}^T v_t v_{t-h} \xrightarrow{a.s.} E[v_t v_{t-h}] = \gamma_h = \sigma_u^2 \sum_{j=0}^{\infty} c_j c_{j+h},$$

where $\sigma_u^2 = E[u_t^2]$, see Phillips and Solo (1992) for a proof. Furthermore,

$$\begin{aligned} & T^{-(p+1)} \sum_{t=1}^T t^p v_t v_{t-h} \\ &= T^{-1} \sum_{t=1}^T (t/T)^p (v_t v_{t-h} - E v_t v_{t-h} + E v_t v_{t-h}) \\ &= T^{-1} \sum_{t=1}^T (t/T)^p (v_t v_{t-h} - E v_t v_{t-h}) + T^{-1} \sum_{t=1}^T (t/T)^p E[v_t v_{t-h}] \\ &\xrightarrow{a.s.} \sigma_u^2 \sum_{j=0}^{\infty} c_j c_{j+h} / (p+1), \end{aligned}$$

because $\left| T^{-1} \sum_{t=1}^T (t/T)^p (v_t v_{t-h} - E v_t v_{t-h}) \right| \leq T^{-1} \sum_{t=1}^T |(v_t v_{t-h} - E v_t v_{t-h})| \xrightarrow{a.s.} 0$, and

$$\begin{aligned} \lim_{T \rightarrow \infty} T^{-1} \sum_{t=1}^T (t/T)^p E[v_t v_{t-h}] &= \lim_{T \rightarrow \infty} \left\{ (\sigma_u^2 \sum_{j=0}^{\infty} c_j c_{j+h}) T^{-1} \sum_{t=1}^T (t/T)^p \right\} \\ &= (\sigma_u^2 \sum_{j=0}^{\infty} c_j c_{j+h}) / (p+1) = \gamma_h / (p+1) \end{aligned}$$

hold. ■

Proof of (37) in Lemma 9. This proof follows from (20) in Lemma 4 in He and Sandberg (2005a), because

$$\begin{aligned}
& T^{-(v+1/2)} \sum_{t=1}^T t^v v_{t-h} \\
= & T^{-1/2} \sum_{t=1}^T v_{t-h} - T^{-(v+1/2)} v \sum_{t=1}^T t^{v-1} \xi_{t-h-1} + O_p(T^{-1}) \\
& \xrightarrow{d} \sigma_u c(1) W(1) - \sigma_u c(1) v \int_0^1 r^{v-1} W(r) dr,
\end{aligned}$$

holds for $h \geq 0$, and it is assumed that u_{-h}, \dots, u_0 are additional starting values or random variables with finite second moments. ■

Proof of (38) and (39) in Lemma 9. These are special cases of (20) and (21), respectively, in Lemma 4 in He and Sandberg (2005a). ■

Proof of (40)(i) in Lemma 9. Consider first the case $h = 0$, we then have that

$$\begin{aligned}
& T^{-(p+1)} \sum_{t=1}^T t^p \xi_{t-1} v_t \\
& \xrightarrow{d} 0.5 \left(\lambda^2 W(1)^2 - p \lambda^2 \int_0^1 W(r)^2 r^{p-1} dr - \frac{\gamma_0}{(p+1)} \right) \quad (42)
\end{aligned}$$

follows from (21) in Lemma 4 in He and Sandberg (2005a), where we replace σ^2 and $\bar{\sigma}_u^2$ with λ^2 and γ_0 , respectively. ■

Proof of (40)(ii) in Lemma 9. For $h > 0$ notice that the following recursive relationship,

$$\begin{aligned}
\xi_{t-1} v_{t-h} &= \left(\xi_{t-(h+1)} + \underbrace{v_{t-h} + v_{t-h+1} + \dots + v_{t-1}}_{h \text{ - terms}} \right) v_{t-h} \\
&= \xi_{t-(h+1)} v_{t-h} + v_{t-h}^2 + \sum_{s=1}^{h-1} v_{t-h} v_{t-h+s},
\end{aligned}$$

holds, and implies

$$\begin{aligned}
& T^{-(p+1)} \sum_{t=1}^T t^p \xi_{t-1} v_{t-h} \\
= & T^{-1} \sum_{t=1}^T (t/T)^p \xi_{t-(h+1)} v_{t-h} + T^{-1} \sum_{t=1}^T (t/T)^p v_{t-h}^2 \\
& + T^{-1} \sum_{t=1}^T (t/T)^p \sum_{s=1}^{h-1} v_{t-h} v_{t-h+s} \\
& \xrightarrow{d} 0.5 \left(\lambda^2 W(1)^2 - p \lambda^2 \int_0^1 W(r)^2 r^{p-1} dr - \frac{\gamma_0}{(p+1)} \right) \\
& + \frac{\gamma_0}{(p+1)} + \frac{\sum_{s=1}^{h-1} \gamma_s}{(p+1)},
\end{aligned}$$

where the convergence follows from the results in (36) and (42). ■

Proof of (41) in Lemma 9. Write

$$\begin{aligned} & T^{-1} \sum_{t=1}^T (t/T)^{2p} v_{t-h}^2 u_t^2 \\ &= T^{-1} \sum_{t=1}^T (t/T)^{2p} (v_{t-h}^2 u_t^2 - E v_{t-h}^2 u_t^2) + T^{-1} \sum_{t=1}^T (t/T)^{2p} E v_{t-h}^2 u_t^2 \\ &\xrightarrow{a.s.} \frac{\gamma_0 \sigma_u^2}{(2p+1)} \end{aligned}$$

because $T^{-1} \sum_{t=1}^T (t/T)^{2p} (v_{t-h}^2 u_t^2 - E v_{t-h}^2 u_t^2) \xrightarrow{a.s.} 0$ and notice that $E v_{t-h}^2 u_t^2 = \gamma_0 \sigma_u^2 = \sigma_u^4 \sum_{s=0}^{\infty} c_s^2 < \infty$ combined with $\lim_{T \rightarrow \infty} T^{-1} \sum_{t=1}^T (t/T)^{2p} = 1/(2p+1)$ gives the convergence results. Using a Central Limit Theorem (CLT) for Martingale Difference Sequences (MDS) ($\{(t/T)^p v_{t-h} u_t\}$ defines a MDS) we conclude that

$$T^{-(p+1/2)} \sum_{t=1}^T t^p v_{t-h} u_t \xrightarrow{d} N\left(0, \frac{\gamma_0 \sigma_u^2}{2p+1}\right).$$

■

Proof of Theorem 5. Write

$$\gamma_{12}^a (\hat{\psi}_1^a - \psi_1^a) = \begin{bmatrix} \tilde{\mathbf{F}}_1 & \tilde{\mathbf{G}}_1 \\ \tilde{\mathbf{G}}_1' & \tilde{\mathbf{H}}_1 \end{bmatrix}^{-1} \begin{bmatrix} \tilde{\mathbf{K}}_1 \\ \tilde{\mathbf{L}}_1 \end{bmatrix},$$

where

$$\begin{aligned} \tilde{\mathbf{F}}_1 &= \begin{bmatrix} \tilde{\mathbf{A}}_1 & \tilde{\mathbf{B}}_1 \\ \tilde{\mathbf{B}}_1' & \tilde{\mathbf{C}}_1 \end{bmatrix}_{8 \times 8}, \\ \tilde{\mathbf{G}}_1 &= \mathbf{0}_{8 \times 4(p-1)} \text{ under } H_{01}^{a,aux}, \\ \tilde{\mathbf{H}}_1 &= \left[\mathbf{T}_{11}^{-1} \left(\sum_{t=1}^T (\mathbf{s}_{1t} \mathbf{s}_{1t}') \otimes [\tilde{h}_{ij}]_{(p-1) \times (p-1)} \right) \mathbf{T}_{11}^{-1} \right]_{4(p-1) \times 4(p-1)}, \text{ with} \\ \tilde{h}_{ij} &= v_{t-i} v_{t-j} \text{ for } i, j \geq 1, \\ \mathbf{T}_{11} &= \text{diag} \left\{ \left(\left[\begin{array}{cccc} T^{1/2} & T^{3/2} & T^{5/2} & T^{7/2} \end{array} \right]' \otimes \mathbf{1} \right)'_{1 \times 4(p-1)} \right\}, \\ \tilde{\mathbf{K}}_1 &= \begin{bmatrix} \tilde{\mathbf{D}}_1 \\ \tilde{\mathbf{E}}_1 \end{bmatrix}_{8 \times 1}, \\ \tilde{\mathbf{L}}_1 &= \left[\mathbf{T}_{11}^{-1} \left(\sum_{t=1}^T \mathbf{s}_{1t} \otimes [\tilde{l}_i]_{(p-1) \times 1} \right) \right]_{4(p-1) \times 1}, \text{ with } \tilde{l}_i = v_{t-i} u_t \text{ for } i \geq 1. \end{aligned}$$

Noticing that $\tilde{\mathbf{L}}_1$ defines a vector MDS with covariance matrix $\sigma_u^2 \mathbf{A}_1 \otimes \boldsymbol{\Sigma}_1$ and obeys a CLT for MDS's, applying Lemma 4 in He and Sandberg (2005a), and Lemma 9, gives that the following expressions converge jointly under the null hypothesis:

$$\begin{aligned} \tilde{\mathbf{G}}_1 &\xrightarrow{p} \mathbf{0}, \quad \tilde{\mathbf{F}}_1 \xrightarrow{d} \boldsymbol{\Psi}_1^a, \quad \tilde{\mathbf{H}}_1 \xrightarrow{d} \mathbf{A}_1 \otimes \boldsymbol{\Sigma}_1, \\ \tilde{\mathbf{K}}_1 &\xrightarrow{d} \boldsymbol{\Pi}_1^a, \quad \tilde{\mathbf{L}}_1 \xrightarrow{d} \mathbf{R} \end{aligned}$$

where

$$\Psi_1^a = \begin{bmatrix} \mathbf{A}_1 & \lambda \mathbf{B}_1 \\ \lambda \mathbf{B}_1' & \lambda^2 \mathbf{C}_1 \end{bmatrix}, \quad \Pi_1^a = \begin{bmatrix} \sigma \mathbf{D}_1 \\ \sigma \lambda \mathbf{E}_1 \end{bmatrix}, \quad \mathbf{R} \sim N(\mathbf{0}, \sigma_u^2 \mathbf{A}_1 \otimes \Sigma_1).$$

By using the Slutsky Theorem,

$$\gamma_{12}^a (\hat{\psi}_1^a - \psi_1^a) \xrightarrow{d} \begin{bmatrix} \Psi_1^a & \mathbf{0} \\ \mathbf{0} & \mathbf{A}_1 \otimes \Sigma_1 \end{bmatrix}^{-1} \begin{bmatrix} \Pi_1^a \\ \mathbf{R} \end{bmatrix} = \begin{bmatrix} (\Psi_1^a)^{-1} \Pi_1^a \\ (\mathbf{A}_1 \otimes \Sigma_1)^{-1} \mathbf{R} \end{bmatrix}.$$

Thus, $\mathbf{Q}_1^a = \mathbf{V}_1^{-1} \mathbf{R} \sim N(\mathbf{0}, \sigma_u^2 \mathbf{V}_1^{-1})$. Furthermore, the fact that $\hat{\psi}_1^a$ is consistent is an immediate consequence of (27). ■

Proof of Theorem 6. This is similar to the proof of Theorem 5 and is therefore omitted. ■

Proof of Corollary 7. First, in the case ψ_1^a define $\mathbf{r}_1^* = \begin{bmatrix} \mathbf{r}_1 & \mathbf{0}' \end{bmatrix}$, where $\mathbf{0}$ is a $4(p-1) \times 1$ vector of zeros. It follows from Theorem 5 that

$$\psi_1^a \equiv T(\hat{\rho}_{10}^a - 1) = \mathbf{r}_1^* \gamma_{12}^a (\hat{\psi}_1^a - \psi_1^a) \xrightarrow{d} \mathbf{r}_1^* (\Psi_1^a)^{-1} \Pi_1^a = \mathbf{r}_1 \mathbf{P}_1^a. \quad (43)$$

Furthermore, define $\mathbf{S}_3 = \text{diag} \{ 1 \ 1 \ 1 \ 1 \ \lambda \ \lambda \ \lambda \ \lambda \}$. Then,

$$\begin{aligned} \mathbf{r}_1 \mathbf{P}_m^a &= \mathbf{r}_1 \left(\mathbf{S}_3^{-1} \begin{bmatrix} \mathbf{A}_1 & \mathbf{B}_1 \\ \mathbf{B}_1 & \mathbf{C}_1 \end{bmatrix}^{-1} \mathbf{S}_3^{-1} \right) \left(\sigma_u \mathbf{S}_3 \begin{bmatrix} \mathbf{D}_1 \\ \mathbf{E}_1 \end{bmatrix} \right) \\ &= \frac{\sigma_u}{\lambda} \mathbf{r}_1 \begin{bmatrix} \mathbf{A}_1 & \mathbf{B}_1 \\ \mathbf{B}_1 & \mathbf{C}_1 \end{bmatrix}^{-1} \begin{bmatrix} \mathbf{D}_1 \\ \mathbf{E}_1 \end{bmatrix}. \end{aligned}$$

and pre-multiplying with λ/σ_u in (43) ensures that (29) holds for $m = 1$.

Second, in the case ψ_1^{pa} , define $\tilde{\mathbf{r}}_1 = \begin{bmatrix} \mathbf{r}_1 & \mathbf{0}' \end{bmatrix}$, where $\mathbf{0}$ is a $(p-1) \times 1$ vector of zeros. It follows from Theorem 6 that

$$\psi_1^{pa} \equiv T(\hat{\rho}_{10}^{pa} - 1) = \tilde{\mathbf{r}}_1 \gamma_{12}^{pa} (\hat{\psi}_1^{pa} - \psi_1^{pa}) \xrightarrow{d} \tilde{\mathbf{r}}_1 (\Psi_1^{pa})^{-1} \Pi_1^{pa} = \mathbf{r}_1 \mathbf{P}_m^a,$$

and that (30) now holds follows from the first part of the proof. ■

Proof of Corollary 8. We only give the proof for (31), since the proof for partial augmentation is similar. Hence,

$$\begin{aligned} T^2 \hat{\sigma}_{\hat{\rho}_{10}^a}^2 &= (S_T^a)^2 \mathbf{r}_1^* \begin{bmatrix} \tilde{\mathbf{F}}_1 & \tilde{\mathbf{G}}_1 \\ \tilde{\mathbf{G}}_1' & \tilde{\mathbf{H}}_1 \end{bmatrix}^{-1} \mathbf{r}_1^{*'} \\ &\xrightarrow{d} \sigma_u^2 \mathbf{r}_1^* \begin{bmatrix} \Psi_1^a & \mathbf{0} \\ \mathbf{0} & \mathbf{A}_1 \otimes \Sigma_1 \end{bmatrix}^{-1} \mathbf{r}_1^{*'} \\ &= \sigma_u^2 \mathbf{r}_1 (\Psi_1^a)^{-1} \mathbf{r}_1' \\ &= \sigma_u^2 \mathbf{r}_1 \mathbf{S}_3^{-1} \tilde{\Psi}_1^{-1} \mathbf{S}_3^{-1} \mathbf{r}_1' \\ &= \sigma_u^2 / \lambda^2 \mathbf{r}_1 \tilde{\Psi}_1^{-1} \mathbf{r}_1', \end{aligned}$$

where $(S_T^a)^2 = \left(y_t - (\mathbf{s}'_{1t}, (y_{t-1}\mathbf{s}_{1t})', (\mathbf{s}_{1t} \otimes \Delta y_t)') \hat{\boldsymbol{\psi}}_1^a \right)^2 / (T - 8 - 4(p-1))$. It follows from Theorem 5 that $(S_T^a)^2 \xrightarrow{p} \sigma_u^2$. Using Corollary 7 and the Slutsky Theorem we conclude that

$$t_1^a \equiv (\hat{\rho}_{10}^a - 1) / \hat{\sigma}_{\hat{\rho}_{10}^a} = \frac{T(\hat{\rho}_{10}^a - 1)}{(T^2 \hat{\sigma}_{\hat{\rho}_{10}^a}^2)^{1/2}} \xrightarrow{d} \frac{\mathbf{r}_1 \tilde{\boldsymbol{\Psi}}_1^{-1} \tilde{\boldsymbol{\Pi}}_1}{(\mathbf{r}_1 \tilde{\boldsymbol{\Psi}}_1^{-1} \mathbf{r}'_1)^{1/2}}.$$

■

Appendix B

Table 10: Critical values for the NPADF tests $c^*\psi_m^{pa}$ and t_m^{pa} in Corollaries 7 and 8 with $p = 3$.

T	$c^*\psi_1^{pa}$			t_1^{pa}		
	0.01	0.05	0.10	0.01	0.05	0.10
50	-451.27	-230.45	-170.90	-3.99	-3.21	-2.81
100	-211.43	-143.54	-115.99	-3.89	-3.15	-2.77
250	-172.32	-122.21	-98.90	-3.83	-3.11	-2.71
500	-163.55	-115.41	-93.87	-3.80	-3.08	-2.70
1000	-160.35	-112.81	-91.20	-3.79	-3.06	-2.67
∞	-156.46	-109.5	-89.89	-3.77	-3.05	-2.66
T	$c^*\psi_2^{pa}$			t_2^{pa}		
	0.01	0.05	0.10	0.01	0.05	0.10
50	-655.89	-302.44	-219.37	-3.93	-3.16	-2.75
100	-247.81	-171.04	-138.84	-3.83	-3.09	-2.71
250	-197.13	-141.11	-115.12	-3.74	-3.02	-2.63
500	-188.72	-133.50	-109.12	-3.71	-3.00	-2.61
1000	-182.92	-130.21	-105.79	-3.69	-2.97	-2.59
∞	-175.60	-126.55	-103.15	-3.66	-2.95	-2.57
T	$c^*\psi_3^{pa}$			t_3^{pa}		
	0.01	0.05	0.10	0.01	0.05	0.10
50	-190.16	-98.82	-74.52	-5.32	-4.63	-4.30
100	-79.46	-56.64	-47.75	-5.18	-4.57	-4.27
250	-58.06	-45.00	-39.09	-5.08	-4.53	-4.25
500	-53.48	-42.22	-37.06	-5.07	-4.52	-4.25
1000	-51.45	-41.11	-36.03	-5.06	-4.52	-4.24
∞	-48.70	-39.87	-34.99	-5.06	-4.51	-4.24

Notes: The probability shown at the head of each column is the area in the left-hand tail. The results are based on 100 000 replications when the true values of the nuisance parameters are zero.

Table 11: Critical values for the NPADF tests $c^*\psi_m^{pa}$ and t_m^{pa} in Corollaries 7 and 8 with $p = 4$.

T	$c^*\psi_1^{pa}$			t_1^{pa}		
	0.01	0.05	0.10	0.01	0.05	0.10
50	-823.27	-663.55	-364.47	-4.09	-3.32	-2.92
100	-322.00	-197.03	-151.67	-3.97	-3.23	-2.84
250	-190.65	-133.64	-107.55	-3.89	-3.15	-2.76
500	-173.87	-121.52	-98.11	-3.83	-3.12	-2.73
1000	-163.29	-115.74	-93.81	-3.80	-3.09	-2.69
∞	-156.46	-109.5	-89.89	-3.77	-3.05	-2.66
T	$c^*\psi_2^{pa}$			t_2^{pa}		
	0.01	0.05	0.10	0.01	0.05	0.10
50	-1553.21	-722.89	-488.81	-4.05	-3.28	-2.88
100	-419.25	-247.02	-189.14	-3.91	-3.17	-2.79
250	-222.47	-156.82	-127.78	-3.81	-3.09	-2.70
500	-198.86	-141.57	-115.27	-3.75	-3.03	-2.65
1000	-188.85	-134.74	-109.63	-3.71	-2.99	-2.62
∞	-175.60	-126.55	-103.15	-3.66	-2.95	-2.57
T	$c^*\psi_3^{pa}$			t_3^{pa}		
	0.01	0.05	0.10	0.01	0.05	0.10
50	-1462.068	-332.69	-184.10	-5.34	-4.65	-4.32
100	-136.38	-80.78	-63.83	-5.17	-4.57	-4.27
250	-66.20	-49.74	-42.81	-5.09	-4.54	-4.26
500	-56.49	-44.02	-38.43	-5.06	-4.52	-4.25
1000	-52.79	-41.75	-36.67	-5.06	-4.51	-4.24
∞	-48.70	-39.87	-34.99	-5.06	-4.51	-4.24

Notes: The probability shown at the head of the column is the area in the left-hand tail. The results are based on 100 000 replications when the true values of the nuisance parameters are zero.

References

- BANERJEE, A., R. L. LUMSDAINE, AND H. J. STOCK (1992): “Recursive and sequential tests of the unit-root and trend break hypotheses: Theory and international evidence,” *Journal of Business and Economic Statistics*, 10, 271–287.
- BEC, F., M. B. SALEM, AND M. CARRASCO (2002): “The purchasing power parity puzzle: Evidence from a LSTAR specification,” Université de Paris I, Working paper.
- BEVERIDGE, S., AND C. R. NELSON (1981): “A new approach to decomposition of economic time series into permanent and transitory components with particular attention to measurement of the business cycle,” *Journal of Monetary Economics*, 7, 151–174.
- CHAN, K. S., AND H. TONG (1986): “On estimating thresholds in autoregressive models,” *Journal of Time Series Analysis*, 7, 179–191.
- 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.
- EKLUND, B. (2003a): “Testing the unit root hypothesis against the logistic smooth transition autoregressive model,” Stockholm School of Economics, Working paper.
- (2003b): “A nonlinear alternative to the unit root hypothesis,” Stockholm School of Economics, Working paper.
- ENDERS, W., AND C. GRANGER (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.
- 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.
- HARVEY, D. I., AND T. C. MILLS (2002): “Unit Roots and Double Smooth Transitions,” *Journal of Applied Statistics*, 29, 675–683.
- HE, C., AND R. SANDBERG (2005a): “Testing Parameter Constancy in Unit Root Autoregressive Models Against Continuous Change,” Stockholm School of Economics, SSE/EFI Working Paper Series in Economics and Finance No. 579.
- KAPETANIOS, G., Y. SHIN, AND A. SNELL (2003): “Testing for a unit root in the nonlinear STAR framework,” *Journal of Econometrics*, 112, 359–379.
- LANNE, M., AND H. LÜTKEPOHL (2002): “Unit root test for time series with level shifts: A comparison of different proposals,” *Economic Letters*, 75, 109–114.
- LANNE, M., H. LÜTKEPOHL, AND P. SAIKKONEN (2003): “Test procedures for unit roots in time series with level shifts at unknown time,” *Oxford Bulletin of Economics and Statistics*, 65, 91–115.

- LEYBOURNE, S., P. NEWBOLD, AND D. VOUGAS (1998): "Unit roots and smooth transitions," *Journal of Time Series Analysis*, 19, 83–97.
- 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.
- OULIARIS, S., J. PARK, AND P. C. B. PHILLIPS (1989): "Testing for a unit root in the presence of a maintained trend," *Kluwer*, pp. 7–28.
- PERRON, P. (1989): "The great crash, the oil price shock, and the unit root hypothesis," *Econometrica*, 57, 1361–1401.
- (1990): "Testing for a unit root in a time series with a changing mean," *Journal of Business and Economic Statistics*, 8, 153–162.
- PERRON, P., AND T. J. VOGELANG (1992): "Nonstationarity and level shifts with an application to purchasing power parity," *Journal of Business and Economic Statistics*, 10, 301–320.
- 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.
- PHILLIPS, P. C. B., AND V. SOLO (1992): "Asymptotics for Linear Processes," *Annals of Statistics*, 20, 971–1001.
- SAIKKONEN, P., AND H. LÜTKEPOHL (2002): "Testing for a unit root in a time series with level shift at unknown time," *Econometric Theory*, 18, 313–348.
- 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.
- TONG, H. (1990): *Non-linear time series*. Oxford Science publications.
- ZIVOT, E., AND D. W. K. ANDREWS (1992): "Further Evidence on the Great Crash, the Oil Price Shock, and the Unit Root Hypothesis," *Journal of Business and Economic Statistics*, 10, 251–270.