

Inference for Unit Roots in a Panel Smooth Transition Autoregressive Model where the Time Dimension is Fixed

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. 581

January 2005

Abstract

In this paper we derive a unit root test against a Panel Logistic Smooth Transition Autoregressive (PLSTAR) model. The analysis is concentrated on the case where the time dimension is fixed and the cross section dimension tends to infinity. Under the null hypothesis of a unit root, we show that the LSDV estimator of the autoregressive parameter in the linear component of the model is inconsistent due to the inclusion of fixed effects. The test statistic, adjusted for the inconsistency, has an asymptotic normal distribution whose first two moments are calculated analytically. To complete the analysis, finite sample properties of the test are examined. We highlight scenarios under which the traditional panel unit root tests by Harris and Tzavalis have inferior or reasonable power compared to our test.

JEL classification: C12; C23; C52

Key words: Dynamic Nonlinear Panel; Smooth Transitions; Structural Breaks; Unit roots; LSDV estimation; Central limit theorem

*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. The paper has been presented at the 11th Biennial Panel Data Conference at College Station, Texas 2004, and at the Ibmecc Business School, São Paulo, Brazil 2004.

1 Introduction

One can expect the traditional panel data unit root tests by Quah (1994), Harris and Tzavalis (1999), and Im, Pesaran, and Shin (2003), to have low power if the time series for the cross sections exhibit structural shifts in levels and/or trends. One explanation to this is that the authors quoted above base their unit root tests on panels where each cross section is modelled as a linear autoregressive process considered in e.g. Dickey and Fuller (1979), Phillips (1987), and Phillips and Perron (1988). As such, it has been pointed out by Perron (1990) that the classical univariate unit root tests are biased towards nonrejection in time series with structural changes, and "adding" up the nonlinearities in single time series into a panel framework will most likely lead to a bias towards nonrejection for the traditional panel unit root tests as well. Considering this fact in addition to that evidence of nonlinearities (such as structural breaks) in many single, say, macroeconomic time series are found, it seems that the traditional panel data unit root tests in such cases are based on too restrictive panels. In particular, this will have serious implications for applied work because the shocks to each cross section will be treated as if they have a permanent effect.

Panel data unit root tests allowing for structural breaks can for instance be found in Im and Lee (1999), Silvestre, Barrio-Castro, and Lopez-Bazo (2001), Tzavalis (2002), and Bai and Silvestre (2003). They derive unit root tests in a panel where each cross section has an abrupt structural shift in the level and/or the time trend. However, in many cases a gradual or more smooth change between two regimes seems preferable, see e.g. Ripatti and Saikkonen (2001). In this paper we generalize the idea with an instant shift in levels by introducing a nonlinear dynamic panel accommodating a smooth cross section specific change in levels and a homogeneous smooth shift in dynamics, in which we test the null hypothesis of a common unit root.

Our test of a common unit root is based on the normalized LSDV estimator of the autoregressive coefficient in an auxiliary regression equation. The time dimension is fixed and the cross section dimension tends to infinity. The analytical limiting distribution of the test is the standard normal where the two first moments are calculated analytically. Our approach is similar to the one in Harris and Tzavalis (1999) whose results are obtained as special cases of ours. We choose therefore to compare the power of our test to the power of their tests. This gives the opportunity to demonstrate when the tests in Harris and Tzavalis (1999) actually have substantial power when in fact a nonlinear panel is considered, as well as scenarios when the traditional tests are heavily biased towards nonrejection.

The rest of the paper is organized as follows. In Section 2 we present the nonlinear dynamic panel. In Section 3 we present the procedure for testing a unit root and derive the test statistic. Section 4 contains simulation experiments to examine the finite-sample properties of the test. Concluding remarks are given in Section 5. Thereafter an Appendix follows where proofs can be found.

2 The model

Consider a first-order panel smooth transition autoregressive (PSTAR(1)) model,

$$y_{it} = \pi_{i,10} + \pi_{11}y_{i,t-1} + (\pi_{i,20} + \pi_{21}y_{i,t-1})F(t; \gamma, c) + u_{it}, \quad i = 1, \dots, n, \quad t = 1, \dots, T, \quad (1)$$

where the index i represents the i 'th cross sectional unit, t indexes time series observations, and u_{it} is the error term. In (1), $F(t; \gamma, c)$ is a transition function satisfying the conditions: (i) $F(t; \gamma, c)$ is a bounded and continuous function for all t , γ , and c . (ii) $F(t; 0, c) = 0$. (iii) In an open interval $(-\varepsilon, \varepsilon)$, for $\varepsilon > 0$, $\partial F(t; \gamma, c)/\partial \gamma$ is non-zero and $\partial^2 F(t; \gamma, c)/\partial \gamma^2$ exists. (iv) For fixed γ and c , $F(t; \gamma, c)$ is monotonic in t . A suitable choice of a transition function in (1) that meets the conditions in (i)-(iv) is the logistic cumulative distribution function (after a downward shift)

$$F(t; \gamma, c) = \frac{1}{1 + \exp\{-\gamma(t - c)\}} - \frac{1}{2}. \quad (2)$$

In (2), $\gamma > 0$ is a slope parameter indicating how rapid the transition is, and $c \in (0, T)$ is a location parameter around which the transition (symmetrically) takes place. Restriction $\gamma > 0$ is an identifying restriction, and implies that $F(t)$ is increasing in t . Model (1) with (2) defines the panel logistic smooth transition autoregressive model of order one, called the PLSTAR(1) model for short. For convenience we only consider a first-order polynomial of t in (2). For a discussion about higher-order polynomials in t , see e.g. Lin and Teräsvirta (1994), or Teräsvirta (1998).

The PLSTAR(1) model contains nonlinear heterogeneous fixed effects, nonlinear homogeneous autoregressive coefficients, and homogeneous slope and location parameters.¹ For each individual equation i , the function $F(t; \gamma, c)$ in (2) allows for a smooth change between regimes in intercepts and dynamics.

It is evident that the PLSTAR(1) model specification nests many panel models studied in the literature. In particular, when $\gamma \rightarrow \infty$, $F(t; \gamma, c)$ in (2) becomes an indicator function, i.e. $F(t; \infty, c) = -0.5$ if $t \in [0, c)$ and $F(t; \infty, c) = 0.5$ if $t \in [c, T]$, and the PLSTAR(1) model displays a panel threshold AR(1) (PTAR(1)) model with a single structural break at $t = c$. At the other end, by letting $\gamma = 0$ in (2) the PLSTAR(1) model collapses into a linear panel AR(1) (PAR(1)) model. It may be mentioned that González, Teräsvirta, and van Dijk (2004) recently introduced a different panel STAR model by generalizing the panel threshold model of Hansen (1999). In their model, $y_{i,t-1}$, is replaced by a vector of exogenous variables \mathbf{x}_{it} so the model is not dynamic and unit roots are not an issue. Furthermore, the transition variable in the transition function (2) is a stochastic variable which can be an element of \mathbf{x}_{it} . Even there, when $\gamma \rightarrow 0$ in (2), the model becomes a linear homogeneous panel model with exogenous variables.

¹The homogeneity assumption imposed on $\pi_{i,10}$ and $\pi_{i,11}$ is needed for the coming testing procedure.

3 Test statistic

We now consider a test statistic for testing the hypothesis of a panel unit root in the PLSTAR(1) model. Under this hypothesis, $H_0 : \pi_{i,10} \in \mathbb{R}$ for all i , $\pi_{11} = 1$, and $\gamma = 0$, in (1) and (2).² It is tested against a stable PLSTAR(1) model with $\gamma > 0$. The stability conditions are given by $\pi_{11} - 0.5\pi_{21} \in (-1, 1)$ and $\pi_{11} + 0.5\pi_{21} \in (-1, 1)$ to rule out non-stationary or explosive trajectories. Note, however, that the PLSTAR(1) model also becomes linear by for any i setting $\pi_{i,20} = \pi_{21} = 0$ in (1). This shows that there is an identification problem in the PLSTAR(1) model under the null hypothesis $\gamma = 0$ because then the parameters $\pi_{i,20}$, π_{21} , and c are not identified. We circumvent this difficulty by an approximation of $F(t; \gamma, c)$, as suggested by Luukkonen, Saikkonen, and Teräsvirta (1988). An obvious candidate is the first-order Taylor expansion of $F(t; \gamma, c)$ around $\gamma = 0$. Applying this approximation to (2) and merging terms and reparameterizing, we obtain the following version of the PLSTAR(1) model

$$y_{it} = \alpha_i + \rho y_{i,t-1} + \delta_i t + \phi t y_{i,t-1} + u_{it}^*, \quad (3)$$

where u_{it}^* is an adjusted error term such that $u_{it}^* = u_{it}$ holds under the null hypothesis, i.e. the distributional properties of the error process are preserved under the null hypothesis and are not affected by the Taylor approximation. The parameters α_i, ρ, δ and ϕ are all functions of the originally defined parameters such that the originally stated null hypothesis is transformed into

$$H_0^{aux} : \alpha_i \in \mathbb{R} \quad \text{for all } i, \quad \rho = 1, \quad \delta_i = 0, \quad \phi = 0. \quad (4)$$

Note that the linear models in Harris and Tzavalis (1999) are nested in the auxiliary regression (3), and we thereby find two plausible competing tests in their Theorems 2 and 3, which will be referred to as the HT2 and HT3 tests respectively. This will also give the opportunity to examine the expectation that traditional panel data unit root test are biased towards nonrejection under models with a shift in levels. Furthermore, the HT2 and HT3 tests are based on models letting $(\delta_i, \phi) = (0, 0)$ and $\phi = 0$, respectively, in (3). To proceed we impose the following assumptions on the PLSTAR(1) model.

Assumption 1 (A1) Let $\{u_{it}\}_{i,t \in \mathbb{N}}$ be an i.i.d. sequence of random variables such that $E(u_{it}) = 0$ and $E(u_{it}^2) = \sigma_u^2$ hold for all i and t . (A2) The individual effect α_i equals 0 for all i . (A3) $E(u_{it}^4) = \mu_4 < \infty$ for all i and t .

3.1 Bias estimator

The panel unit root test statistic is constructed from the normalized coefficient statistic based directly on the LSDV estimator of the coefficients of the auxiliary fixed effect model (3). Under the null hypothesis (4), the deviation form of the LSDV estimator of

²That is, a joint test of parameter constancy (linearity) and a unit root.

ρ in the model (3) is given by

$$\hat{\rho} - 1 = \frac{\left[\left(\sum_{i=1}^n W_{i,2t} \right) \left(\sum_{i=1}^n W_{i,4t} \right) - \left(\sum_{i=1}^n W_{i,3t} \right) \left(\sum_{i=1}^n W_{i,5t} \right) \right]}{\left[\left(\sum_{i=1}^n W_{i,1t} \right) \left(\sum_{i=1}^n W_{i,2t} \right) - \left(\sum_{i=1}^n W_{i,3t} \right)^2 \right]}. \quad (5)$$

In (5), $W_{i,1t} = \mathbf{y}'_{i,-1} \mathbf{Q}_T \mathbf{y}_{i,-1}$, $W_{i,2t} = \mathbf{y}'_{i,-1} \mathbf{D}_T \mathbf{Q}_T \mathbf{D}_T \mathbf{y}_{i,-1}$, $W_{i,3t} = \mathbf{y}'_{i,-1} \mathbf{Q}_T \mathbf{D}_T \mathbf{y}_{i,-1}$, $W_{i,4t} = \mathbf{y}'_{i,-1} \mathbf{Q}_T \mathbf{u}_i$, $W_{i,5t} = \mathbf{y}'_{i,-1} \mathbf{D}_T \mathbf{Q}_T \mathbf{u}_i$ with vectors $\mathbf{y}_{i,-1} = (y_{i0}, \dots, y_{i,T-1})'$, $\mathbf{u}_i = (u_{i1}, \dots, u_{iT})'$. Furthermore, \mathbf{Q}_T is the $(T \times T)$ within transformation matrix defined by $\mathbf{Q}_T = \mathbf{I}_T - \mathbf{M}_T$ where $\mathbf{M}_T = \mathbf{X}_T (\mathbf{X}'_T \mathbf{X}_T)^{-1} \mathbf{X}'_T$ with $\mathbf{X}_T = (\boldsymbol{\iota}_T, \boldsymbol{\tau}_T)$ and $\boldsymbol{\iota}_T$ is the unit column vector of length T , $\boldsymbol{\tau}_T = (1, 2, \dots, T)'$, and $\mathbf{D}_T = \text{diag}\{1, 2, \dots, T\}$. Under the null hypothesis (4) the LSDV estimator in (5) is inconsistent for fixed T as $n \rightarrow \infty$. This result is stated in the following theorem.

Theorem 1 *Consider model (3) when (4) and (A1)-(A3) in Assumption 1 hold. Then, for any fixed $T > 2$, the LSDV estimator $(\hat{\rho} - 1)$ in (5) satisfies*

$$\text{plim}_{n \rightarrow \infty} (\hat{\rho} - 1) = B_1(T), \quad (6)$$

where

$$B_1(T) \equiv -\frac{1}{4} \frac{23T^2 - 21T - 74}{(T^2 - 2)(T + 2)}.$$

Proof. See Appendix A. ■

Theorem 1 states that when n tends infinity and T is fixed, the LSDV estimator $\hat{\rho}$ in (3) is inconsistent under the null hypothesis (4). The degree of inconsistency only depends upon T , and it is order equals $\mathcal{O}(T^{-1})$. Thus, $T \rightarrow \infty$ is required for $\text{plim}_{n, T \rightarrow \infty} \hat{\rho} = 1$ to hold. The inconsistency arises because of the elimination of the fixed effects α_i and the time trend δ_i by the \mathbf{Q}_T matrix from each observation of the panel, see Nickell (1981). This makes the explanatory variables correlated with the error term as Hsiao (1986) pointed out. An interesting feature is that the bias is negative. To see this note that under (4), $\mathbf{y}_{i,-1} = y_{i0} \boldsymbol{\iota}_T + \mathbf{C}_T \mathbf{u}_i$ holds where \mathbf{C}_T is the strictly lower triangular $(T \times T)$ matrix

$$\mathbf{C}_T = \begin{bmatrix} 0 & 0 & \cdots & 0 \\ 1 & 0 & \cdots & 0 \\ \vdots & \ddots & \ddots & \vdots \\ 1 & \cdots & 1 & 0 \end{bmatrix}, \quad (7)$$

which implies that the inequalities

$$E(\mathbf{y}'_{i,-1} \mathbf{Q}_T \mathbf{u}_i) = -E(\mathbf{u}'_i \mathbf{C}'_T \mathbf{M}_T \mathbf{u}_i) < 0$$

and

$$E(\mathbf{y}'_{i,-1} \mathbf{D}_T \mathbf{Q}_T \mathbf{u}_i) = -E(\mathbf{u}'_i \mathbf{C}'_T \mathbf{D}_T \mathbf{M}_T \mathbf{u}_i) < 0$$

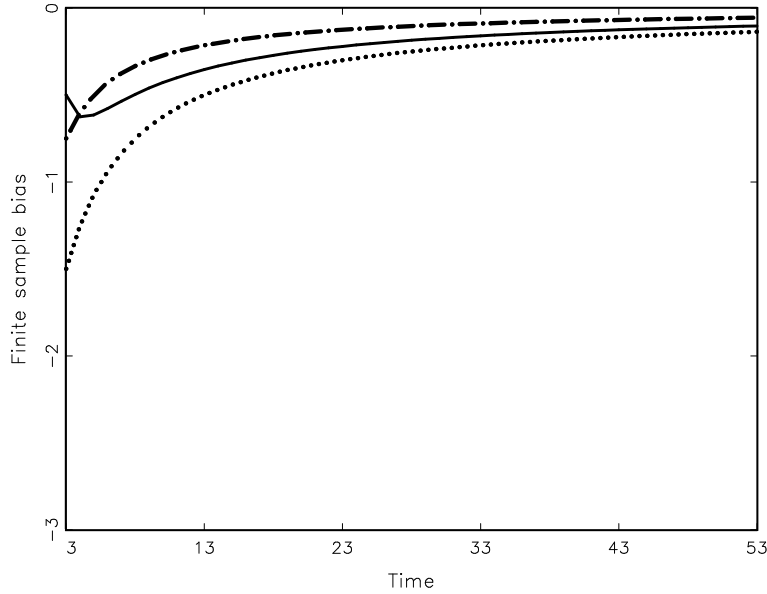
hold. By these inequalities it follows that induced correlations between the explanatory variables and the error disturbances are always negative. This feature plays a key role in yielding the negative value of $B_1(T)$. Furthermore, the biases for the HT2 and HT3 test statistics are obtainable by using (5), and equal

$$\text{plim}_{n \rightarrow \infty}(\hat{\rho} - 1) = B_{HT2}(T) \equiv -3(T + 1)^{-1}, \quad (8)$$

$$\text{plim}_{n \rightarrow \infty}(\hat{\rho} - 1) = B_{HT3}(T) \equiv -\frac{15}{2}(T + 2)^{-1}. \quad (9)$$

For $T > 3$, one can show that $|B_{HT2}(T)| < |B_1(T)| < |B_{HT3}(T)|$. In fact, for small T , $|B_1(T)| - |B_{HT3}(T)|$ could be rather substantial. This is illustrated in Figure 1.

Figure 1: The finite sample bias of the LSDV estimator in Theorem 1 (solid line), and the corresponding biases HT2 (dash-dotted line) and HT3 (dotted line) in Harris and Tzavalis (1999).



3.2 Asymptotic distribution

Because the bias of $(\hat{\rho} - 1)$ is known, it is possible to derive the limiting distribution for the bias corrected normalized statistic for the model (3) under the null hypothesis (4). The result is given in the following theorem.

Theorem 2 Consider model (3) when (4) and (A1)-(A3) in Assumption 1 hold. Then, for any fixed $T > 2$, the limiting distribution of the LSDV estimator $(\hat{\rho} - 1)$ in (5), adjusted by $B_1(T)$ in (6), is given by

$$\sqrt{n}(\hat{\rho} - 1 - B_1(T)) \xrightarrow{d} N\left(0, \sigma_{\hat{\rho}}^2(T, \kappa_4)\right), \quad (10)$$

where

$$\sigma_{\hat{\rho}}^2(T, \kappa_4) \equiv 5\kappa_4 \frac{n_1(T)}{n_2(T)} + \frac{n_3(T)}{n_4(T)},$$

with

$$\begin{aligned} n_1(T) = & 8428\,767T^{11} - 13\,614\,689T^{10} - 120\,059\,496T^9 \\ & + 186\,771\,124T^8 + 721\,928\,310T^7 - 948\,544\,018T^6 \\ & - 2393\,879\,224T^5 + 2116\,570\,904T^4 + 5166\,454\,483T^3 \\ & + 615\,163\,035T^2 - 1914\,301\,704T - 461\,936\,628, \end{aligned}$$

$$n_2(T) = 512\,512\,(T^2 - 2)^4\,(T + 2)^3\,(T^2 - 1)\,(T - 2)\,T\,(T - 3)^{-1},$$

$$\begin{aligned} n_3(T) = & 686\,450\,089T^{13} - 2714\,666\,460T^{12} + 5972\,242\,321T^{11} \\ & + 22\,845\,456\,210T^{10} - 149\,532\,661\,418T^9 - 51\,654\,581\,616T^8 \\ & + 893\,153\,037\,170T^7 - 96\,760\,187\,484T^6 - 2612\,622\,746\,635T^5 \\ & + 322\,041\,658\,116T^4 + 4127\,083\,405\,469T^3 + 994\,368\,662\,874T^2 \\ & - 1478\,687\,733\,396T - 374\,168\,668\,680, \end{aligned}$$

$$n_4(T) = 9225\,216\,(T^2 - 2)^4\,(T + 2)^3\,(T - 2)\,(T - 1)\,(T + 1)\,T,$$

and

$$\kappa_4 = \mu_4 / \sigma_u^4.$$

Proof. See Appendix A. ■

Theorem 2 states that the test statistic defined in (10), corrected by $B_1(T)$ in (6) for the inconsistency of $\hat{\rho}$ in (5), is normally distributed with mean zero and variance $\sigma_{\hat{\rho}}^2(T, \kappa_4)$ as $n \rightarrow \infty$. The asymptotic variance $\sigma_{\hat{\rho}}^2(T, \kappa_4)$ is a function of T and the nuisance parameter κ_4 . The dependence on κ_4 can be eliminated by imposing the normality assumption on the disturbances $\{u_{it}\}$, and the simplified form of the variance $\sigma_{\hat{\rho}}^2(T, \kappa_4)$ in (10) appears in the following corollary.

Corollary 3 Consider model (3) when (4) and (A1)-(A3) in Assumption 1 hold. Furthermore, assume that u_{it} are normally distributed. Then, for any fixed $T > 2$, the limiting distribution of the LSDV estimator $(\hat{\rho} - 1)$ in (5), adjusted by $B_1(T)$ in (6), is given by

$$\sqrt{n}(\hat{\rho} - 1 - B_1(T)) \xrightarrow{d} N(0, \sigma_{\hat{\rho}}^2(T)), \quad (11)$$

where

$$\sigma_{\hat{\rho}}^2(T) \equiv \frac{n_5(T)}{n_6(T)}, \quad (12)$$

with

$$\begin{aligned} n_5(T) = & 52\,803\,853T^{10} - 33\,761\,490T^9 - 295\,736\,530T^8 \\ & + 78\,337\,770T^7 - 438\,526\,236T^6 - 538\,473\,642T^5 \\ & + 3583\,336\,934T^4 + 1400\,993\,790T^3 - 4271\,003\,921T^2 \\ & + 1598\,065\,812T + 4063\,557\,132, \end{aligned}$$

and

$$n_6(T) = 709\,632\,(T^2 - 2)^4\,(T + 2)^3\,(T - 2).$$

Proof. See Appendix A. ■

From Corollary 3 we see that the asymptotic normality of $\sqrt{n}(\hat{\rho} - 1 - B_1(T))$ yields a test statistic which depends only on the estimated parameter $\hat{\rho}$ and known values of n and T . Hence, $\sqrt{n}(\hat{\rho} - 1 - B_1(T))/\sigma_{\hat{\rho}}(T)$ can be readily used for statistical inference, and the critical values of the standard normal distribution apply. Moreover, by similar manipulations we also find the results

$$\sqrt{n}(\hat{\rho} - 1 - B_{HT2}(T)) \xrightarrow{d} N(0, \sigma_{\hat{\rho}_{HT2}}^2(T)), \quad (13)$$

$$\sqrt{n}(\hat{\rho} - 1 - B_{HT3}(T)) \xrightarrow{d} N(0, \sigma_{\hat{\rho}_{HT3}}^2(T)), \quad (14)$$

where

$$\sigma_{\hat{\rho}_{HT2}}^2(T) \equiv \frac{3}{5} \frac{(17T^2 - 20T + 17)}{(T - 1)(T + 1)^3},$$

and

$$\sigma_{\hat{\rho}_{HT3}}^2(T) \equiv \frac{15}{112} \frac{(193T^2 - 728T + 1147)}{(T - 2)(T + 2)^3},$$

are the variances Harris and Tzavalis (1999) obtained under the normality assumption. Note that although $|B_{HT2}(T)| < |B_1(T)| < |B_{HT3}(T)|$ for $T > 3$, one can show that $\sigma_{\hat{\rho}}^2(T) > \sigma_{\hat{\rho}_{HT3}}^2(T) > \sigma_{\hat{\rho}_{HT2}}^2(T)$ holds for $T > 4$. The situation is illustrated in the left-hand panel of Figure 2 where it is seen that $\sigma_{\hat{\rho}}^2(T)$ is a decreasing function in T for $T \geq 4$. In addition we note that $\max_{T \in (2, \infty)} \{\sigma_{\hat{\rho}}^2(T)\} = 2.29$ occurs at $T = 4$. The reason for the ordering between the variances is that under the auxiliary null hypothesis (4), the LSDV estimators $\hat{\rho}$ and $\hat{\phi}$ are correlated.

Note that the result in Corollary 3 is specific to the case that $\alpha_i = 0$ for all i . If (A2) in Assumption 1 is relaxed the limiting distribution of Corollary 3 is no longer invariant with respect to α_i and σ_u^2 . By including cross section specific trends in the regression model (1c) in Harris and Tzavalis (1999) it is shown that their limiting distribution of the test statistic in (14) is invariant with respect to α_i and σ_u^2 without assuming $\alpha_i = 0$. Although our regression in (3) contains a time trend, the limiting distribution in (11) requires that $\alpha_i = 0$ for any i due to the nonlinear feature of (3).

Letting both n and T tend to infinity in (11) results in a degenerate limiting distribution. Because $\sigma_{\hat{\rho}}^2(T)$ in (12) is $\mathcal{O}(T^{-2})$. For this reason it is necessary to re-scale the test statistic (11) in Corollary 3. We have the following result.

Corollary 4 *Suppose that the conditions in Corollary 3 hold for model (3). Then, as $T \rightarrow \infty$ and $n \rightarrow \infty$*

$$\sqrt{n}T(\hat{\rho} - 1) + \frac{23}{4}\sqrt{n} \xrightarrow{d} N\left(0, \frac{52\,803\,853}{709\,632}\right). \quad (15)$$

Proof. The proof of (15) follows immediately from Corollary 3. ■

As noted by Levin, Lin, and Chu (2002), in contrast to the case of stationary panel data, the presence of a unit root causes the fixed effects to influence the asymptotic distribution of the panel autoregressive estimator by factor $23\sqrt{n}/4$, even as both n and T become large. Also, Corollary 4 implies that $\hat{\rho}$ in (5) converges at the rate $\sqrt{n}T$, which is higher than the convergence rate of the LSDV estimator in the stationary case. Comparing the test statistic (15) with (11), we see that the term adjusting for inconsistency of $\hat{\rho}$ in Corollary 4 when $T \rightarrow \infty$ is greater than $\sqrt{n}|B_1(T)|$ in Corollary 3 when T is fixed. Furthermore, the asymptotic variance $\lim_{T \rightarrow \infty} T^2 \sigma_{\hat{\rho}}^2(T) = 52\,803\,853/709\,632 \approx 74.41$ is always greater than $T^2 \sigma_{\hat{\rho}}^2(T)$ for any fixed T , see the right-hand panel of Figure 2.

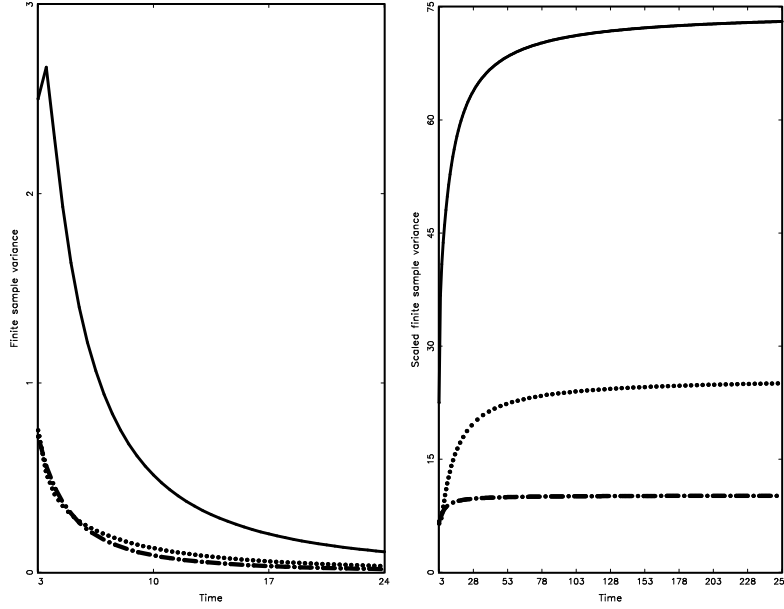


Figure 2: The finite sample variance $\sigma_{\hat{\rho}}^2(T)$ and the scaled finite sample variance $T^2 \sigma_{\hat{\rho}}^2(T)$ (solid lines), and the corresponding variances for HT2 (dash-dotted lines) and HT3 (dotted lines) in Harris and Tzavalis (1999).

The results in Corollaries 3 and 4 may be applied to consider the consequences, as suggested by Harris and Tzavalis (1999), of assuming that T is asymptotic rather than fixed. In detail, Corollary 4 implies that we would use

$$\frac{T\sqrt{n}(\hat{\rho} - 1)}{\sqrt{74.41}} + \frac{23}{4} \frac{\sqrt{n}}{\sqrt{74.41}} \xrightarrow[n, T \rightarrow \infty]{d} N(0, 1), \quad (16)$$

for inference when $T, n \rightarrow \infty$, whereas the true distribution for $T < \infty$ and $n \rightarrow \infty$ is

$$\frac{\sqrt{n}T(\hat{\rho} - 1)}{\sqrt{74.41}} + \frac{23}{4} \frac{\sqrt{n}}{\sqrt{74.41}} C_1(T) \xrightarrow[n \rightarrow \infty]{d} N(0, C_2(T)), \quad (17)$$

where $C_1(T) = -TB_1(T)/(23/4)$ and $C_2(T) = T^2\sigma_\rho^2(T)/74.41$ such that $C_1(T), C_2(T) \in (0, 1)$ holds for $4 < T < \infty$, and $\lim_{T \rightarrow \infty} C_1(T) = \lim_{T \rightarrow \infty} C_2(T) = 1$. From (17) it is clear that we have two possible effects if we erroneously use (16) in finite-samples. The first effect is the mean shift effect, $C_1 < 1$ and therefore the asymptotic distribution in (16) is located to the right of the finite-sample distribution in (17). This implies an increase in the size over the nominal level. The second effect is the variance effect, $C_2 < 1$, which implies that the asymptotic variance in (16) is too large so the tails of the asymptotic distribution contain excess probability mass. This leads to a decrease in the size of the standardized test statistic over the nominal level. As a conclusion, if the mean shift effect dominates the variance effect, the test will be oversized and the power is increased. The relative importance of these effects are investigated in the next section using Monte Carlo simulations.

3.3 Heterogeneous errors

The errors u_{it} in Assumption 1 are assumed to be i.i.d. such that $E(u_{it}) = 0$ and $E(u_{it}^2) = \sigma_u^2$ hold for all i and t , but this is easily relaxed to allow for heterogeneous errors.

Assumption 2 (B1) Let $\{u_{it}\}_{i,t \in \mathbb{N}}$ be a sequence of independently distributed random variables for all i and t with $E(u_{it}) = 0$ and $E(u_{it}^2) = \sigma_i^2 < \infty$, and $\lim_{n \rightarrow \infty} n^{-1} \sum_{i=1}^n \sigma_i^2 = \sigma_u^2 < \infty$. (B2) The individual effect α_i equals 0 for all i . (B3) $E|u_{it}|^{4+\delta} < \infty$ holds for $\delta > 0$, and $\lim_{n \rightarrow \infty} n^{-1} \sum_{i=1}^n \mu_{4i} = \mu_4 < \infty$ where $E u_{it}^4 = \mu_{4i}$.

Assumption 2 allows us to derive the same results as in Theorems 1 and 2 and Corollaries 3 and 4 by applying the Markov Law of Large Numbers (LLN) and the Liapounov Central Limit Theorem (CLT).

4 Simulation experiments

In this section we conduct several Monte Carlo experiments to explore finite-sample and asymptotic properties of our test statistics defined in Corollaries 3 and 4, denoted T_1 and T_2 , respectively. In addition, the properties of the HT2 and HT3 tests by Harris and Tzavalis (1999) will be investigated as well.

4.1 Size simulations

The aim of the first experiment is to assess the size properties of the test in Corollary 3. The DGP under the null hypothesis is given by

$$y_{it} = y_{i,t-1} + u_{it}, \quad i = 1, \dots, n, \quad t = 1, \dots, T, \quad (18)$$

where $u_{it} \sim \text{nid}(0, 1)$ for all i and t . The empirical size and its finite-sample accuracy are reported in Table 1.

As can be seen from this table, the empirical distribution of the test statistic in Corollary 3 approximates fairly well the standard normal distribution for almost all n and T . When n is small relatively to T , there is a slight size distortion because the time dimension dominates the cross section dimension. In this case we expect the finite-sample distribution to be a less satisfactory approximation to the asymptotic distribution. However, increasing n to match the time dimensions, we see that the size discrepancy vanishes. For comparison, the bottom line in Table 1 reports the quantiles for the standard normal distribution.

4.2 Power simulations

4.2.1 A homogeneous nonlinear panel

We examine the empirical power under a modified PLSTAR(1) model because the transition function in (2) is replaced with $\tilde{F}(t; \gamma, c) = F(t; \gamma, c) + 0.5$. It is clear that $\tilde{F}(t) : \mathbb{R}_+ \rightarrow [0, 1]$. Furthermore, the error term is assumed to be the same as in (18). The parameters in the modified PLSTAR model are assigned the following values

$$\pi_{i,10} = 0 \quad \forall i, \quad \pi_{11} = 0.4, \quad \pi_{i,20} = 1 \quad \forall i,$$

$$\pi_{21} \in \{0.4, 0.5, 0.55\}, \quad \gamma \in \{0.01, 1, 100\}, \quad c = T/2.$$

which generates a completely homogeneous panel. We first examine the power properties under an almost linear PLSTAR(1) model with $\gamma = 0.01$. Second, the power when the speed of transition in the PLSTAR(1) model may be characterized as intermediate with $\gamma = 1.00$ is investigated. Finally, we study the power when the transition takes place almost instantaneously with $\gamma = 100$, so the model practically contains a single structural break. Furthermore, within these three experiments the change in the intercept is set modest and equals 1, and the stationary root increases from 0.4 to 0.95 (assuming that a complete transition takes place). The design of this experiment implies panels with clear shift in levels and dynamics as long as $\gamma \geq 1$, see Figure 4. The results are presented in Tables 2-4.

When the DGP is an almost linear PLSTAR(1) model, we see from Table 2 that HT2 has the highest power, which is due to the near-linearity of the process. Test statistic HT2 is designed to have high power against linear models, and for small n and T it actually exhibits substantially higher power than in the case of a completely linear panel where the autoregressive coefficients range from 0.8 to 0.95, cf. Table 2b in Harris and Tzavalis (1999). This is natural because in our case the autoregressive parameter

Table 1: Empirical quantiles of the test statistic in Corollary 3.

T	n	1%	5%	10%	50%	90%	95%	99%	Size
5	5	-2.60	-1.79	-1.38	0.03	1.47	1.98	2.93	0.06
5	10	-2.44	-1.71	-1.31	0.02	1.37	1.79	2.65	0.06
5	25	-2.33	-1.63	-1.27	0.02	1.29	1.66	2.51	0.05
5	50	-2.34	-1.63	-1.27	0.01	1.31	1.70	2.43	0.05
5	100	-2.29	-1.66	-1.26	0.01	1.32	1.71	2.43	0.05
10	5	-2.55	-1.80	-1.40	-0.03	1.30	1.70	2.44	0.07
10	10	-2.50	-1.78	-1.39	-0.03	1.27	1.63	2.32	0.06
10	25	-2.48	-1.75	-1.35	-0.02	1.27	1.64	2.33	0.06
10	50	-2.41	-1.70	-1.33	-0.01	1.27	1.65	2.27	0.06
10	100	-2.43	-1.69	-1.32	-0.01	1.28	1.62	2.34	0.05
25	5	-2.90	-1.98	-1.53	-0.12	1.21	1.57	2.17	0.08
25	10	-2.64	-1.85	-1.43	-0.04	1.23	1.58	2.23	0.07
25	25	-2.49	-1.78	-1.41	-0.05	1.22	1.58	2.21	0.06
25	50	-2.44	-1.71	-1.33	-0.03	1.24	1.58	2.19	0.06
25	100	-2.43	-1.68	-1.32	-0.02	1.24	1.61	2.22	0.05
50	5	-3.10	-2.05	-1.59	-0.10	1.18	1.50	2.17	0.09
50	10	-2.69	-1.91	-1.48	-0.09	1.16	1.50	2.15	0.07
50	25	-2.67	-1.84	-1.42	-0.05	1.19	1.53	2.17	0.07
50	50	-2.50	-1.75	-1.39	-0.04	1.26	1.60	2.19	0.06
50	100	-2.46	-1.71	-1.33	-0.01	1.23	1.57	2.28	0.06
100	5	-3.11	-2.07	-1.63	-0.10	1.19	1.51	2.09	0.09
100	10	-2.80	-1.87	-1.45	-0.06	1.18	1.51	2.11	0.07
100	25	-2.72	-1.87	-1.43	-0.05	1.21	1.53	2.16	0.07
100	50	-2.52	-1.77	-1.37	-0.04	1.20	1.57	2.22	0.06
100	100	-2.48	-1.74	-1.34	-0.03	1.23	1.57	2.16	0.06
$N(0, 1)$		-2.33	-1.65	-1.28	0.00	1.28	1.65	2.33	

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

Table 2: Empirical power of the test statistic in Corollary 3 and the tests in Harris and Tzavalis (1999). The DGP is a PAR(1) model.

$\gamma = 0.01$										
$\pi_{12} = 0.40$			$\pi_{21} = 0.40$			$\pi_{21} = 0.50$			$\pi_{21} = 0.55$	
T	n		T_1	HT2	HT3	T_1	HT2	HT3	T_1	HT2 HT3
5	5		0.05	0.35	0.08	0.05	0.27	0.07	0.04	0.23 0.06
5	10		0.06	0.54	0.10	0.05	0.41	0.09	0.05	0.34 0.08
5	25		0.06	0.86	0.16	0.05	0.72	0.12	0.05	0.63 0.11
5	50		0.06	0.99	0.23	0.06	0.93	0.17	0.05	0.87 0.15
5	100		0.07	1.00	0.38	0.07	1.00	0.24	0.06	0.99 0.21
10	5		0.14	0.80	0.28	0.12	0.69	0.21	0.10	0.62 0.19
10	10		0.18	0.97	0.41	0.14	0.91	0.30	0.11	0.85 0.26
10	25		0.28	1.00	0.72	0.20	1.00	0.56	0.16	0.99 0.47
10	50		0.41	1.00	0.93	0.27	1.00	0.81	0.21	1.00 0.70
10	100		0.61	1.00	1.00	0.39	1.00	0.97	0.30	1.00 0.93
25	5		0.68	1.00	0.97	0.58	1.00	0.89	0.52	1.00 0.83
25	10		0.89	1.00	1.00	0.79	1.00	0.99	0.74	1.00 0.98
25	25		1.00	1.00	1.00	0.98	1.00	1.00	0.96	1.00 1.00
25	50		1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00 1.00
25	100		1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00 1.00

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

slowly grows e.g. from 0.4 to 0.67 over time (the case $T = 25$, $\gamma = 0.01$, and $\pi_{21} = 0.5$), rather than taking on e.g. the value 0.95 throughout the whole sample in a linear panel.³ Statistic HT3 also performs better than our test. The reason is that HT3 is based on a more parsimonious alternative than our test which is penalized when the actual DGP is relatively simple.

For $T \geq 25$ and $n \geq 25$ our test performs satisfactorily and for $T \geq 50$ and all n (not reported here) all tests have unit power. In fact, increasing π_{21} has only a moderate impact on power because the transition is very slow, see Figure 3.

Consider next the PLSTAR(1) model with $\gamma = 1$. In Table 3 we see that T_1 outperforms the other tests. It now becomes evident that our less parsimonious model is justified and is in fact necessary if one wants to capture the nonlinear behavior characterized by the PLSTAR(1) model. Our test actually has substantial power when the time dimension is as small as $T = 5, 10$, for almost all n and π_{21} . This is in contrast to

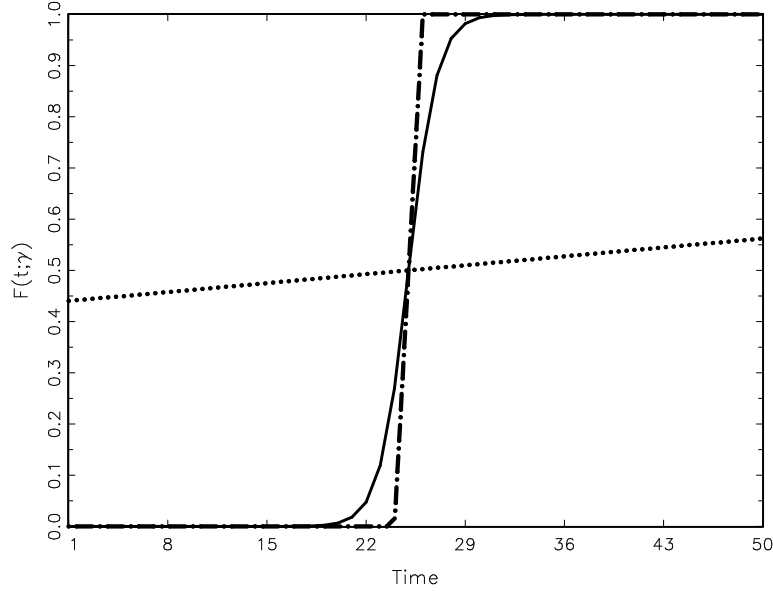
³For small T and $\gamma = 0.01$, a full transition from zero to one does not take place, see Figure 3. For example, $F(t = T = 25; \gamma = 0.01, c = 12.5) \approx 0.53$, implying that the value of the autoregressive parameter at the end of the period equals $\pi_{11} + 0.53\pi_{21} \in [0.61, 70]$.

Table 3: Empirical power of the test statistic in Corollary 3 and the tests in Harris and Tzavalis (1999). The DGP is a PLSTAR(1) model.

$\gamma = 1.00$										
$\pi_{11} = 0.40$			$\pi_{21} = 0.40$			$\pi_{21} = 0.50$			$\pi_{21} = 0.55$	
T	n	T_1	HT2	HT3	T_1	HT2	HT3	T_1	HT2	HT3
5	5	0.20	0.06	0.08	0.23	0.03	0.06	0.26	0.02	0.05
5	10	0.27	0.04	0.10	0.32	0.01	0.08	0.37	0.01	0.07
5	25	0.46	0.01	0.16	0.58	0.00	0.10	0.65	0.00	0.08
5	50	0.69	0.00	0.24	0.83	0.00	0.14	0.89	0.00	0.10
5	100	0.90	0.00	0.40	0.98	0.00	0.21	0.99	0.00	0.14
10	5	0.47	0.00	0.17	0.55	0.00	0.08	0.58	0.00	0.05
10	10	0.68	0.00	0.23	0.77	0.00	0.09	0.81	0.00	0.03
10	25	0.94	0.00	0.38	0.98	0.00	0.09	0.99	0.00	0.01
10	50	1.00	0.00	0.59	1.00	0.00	0.10	1.00	0.00	0.00
10	100	1.00	0.00	0.85	1.00	0.00	0.13	1.00	0.00	0.00
25	5	0.79	0.01	0.52	0.76	0.00	0.09	0.74	0.00	0.00
25	10	0.95	0.00	0.77	0.93	0.00	0.09	0.92	0.00	0.00
25	25	1.00	0.00	0.98	1.00	0.00	0.10	0.99	0.00	0.00
25	50	1.00	0.00	1.00	1.00	0.00	0.11	1.00	0.00	0.00
25	100	1.00	0.00	1.00	1.00	0.00	0.14	1.00	0.00	0.00
50	5	0.97	0.21	0.96	0.77	0.00	0.17	0.57	0.00	0.00
50	10	1.00	0.46	1.00	0.94	0.00	0.30	0.77	0.00	0.00
50	25	1.00	0.92	1.00	1.00	0.00	0.67	0.97	0.00	0.00
50	50	1.00	1.00	1.00	1.00	0.00	0.95	1.00	0.00	0.00
50	100	1.00	1.00	1.00	1.00	0.00	1.00	1.00	0.00	0.00
100	5	1.00	1.00	1.00	0.79	0.00	0.76	0.26	0.00	0.00
100	10	1.00	1.00	1.00	1.00	0.00	0.97	0.35	0.00	0.00
100	25	1.00	1.00	1.00	1.00	0.00	1.00	0.56	0.00	0.00
100	50	1.00	1.00	1.00	1.00	0.00	1.00	0.79	0.00	0.00
100	100	1.00	1.00	1.00	1.00	0.00	1.00	0.96	0.00	0.00

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

Figure 3: The logistic transition function with different values of the speed of transition parameter: $\gamma = 0.01$ (dotted line), $\gamma = 1.00$ (solid line), and $\gamma = 100$ (dash-dotted line).

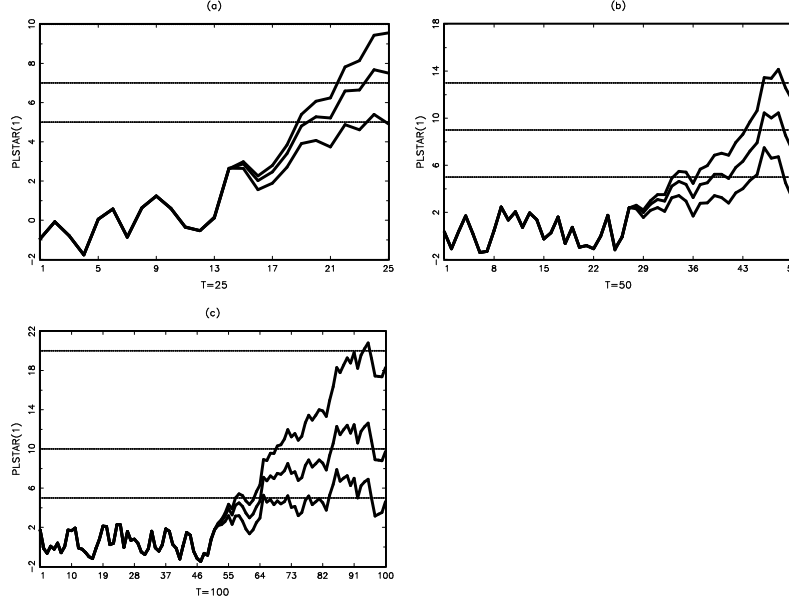


HT2 and HT3, and especially the power of HT2 is very low.

For $T \geq 25$ there are two points worth stressing. The first one is the successive break-down in power of the tests based on linear models by increasing π_{21} . To take an example, in the case $T = n = 50$ and $\pi_{21} = 0.40$ all tests have a power of unity. When π_{21} is increased to 0.50, a remarkable drop in power occurs for HT2 (from 1 to 0) whereas the power for T_1 and HT3 basically remains unchanged. Increasing π_{21} further to 0.55, there is a similar drop in power for HT3 (from 0.95 to 0) whereas our test still has power of unity. This emphasizes the relevance of the inclusion of the set of extra explanatory variables $\{t, ty_{i,t-1}\}$, missing from HT2, and $\{ty_{i,t-1}\}$ missing from HT3, in our auxiliary regression equation (3). The break-down in power could be explained by investigating the shape of the trajectories from the LSTAR(1) model as time evolves. In the cases $\pi_{21} = 0.40, 0.50$, and 0.55 , the autoregressive parameter changes smoothly over time from 0.4 to 0.80, 0.90, and 0.95, and typical realizations of the LSTAR model start at zero and end up at levels around 5, 9, and 13 respectively, see the panel (b) in Figure 4.⁴ Thus, increasing π_{21} does not only imply that the LSTAR model is closer to being non-stable at the end of the period, it also implies a strive towards higher levels, see He and Sandberg (2005a) for the discussion about the level leverage effect. These two effects contribute to the successive break-down in power of the HT2 and HT3 tests. The bias towards nonrejection for HT2 and HT3 tests when the time series in the panel have a pronounced shift in levels is demonstrated.

⁴With $T = 50$ and $\gamma = 1$, a complete transition takes place, see Figure 3.

Figure 4: Typical realizations for a cross section unit in the PLSTAR(1) model where the sample sizes and the autoregressive roots in the nonlinear parts are varied.



The second phenomenon is that the power of T_1 actually decreases in T when $\pi_{21} = 0.55$ and $T \geq 25$, although it still increases monotonically with n . To study this, consider the case $n = 5$ while varying T and π_{21} . For $T = 25$ and $\pi_{21} = 0.40, 0.50$, and 0.55 , the T_1 test show about the same power (≈ 0.77). In panel (a) in Figure 4, we can see that the trajectories at the end of the period reaches levels approximately equal to 5, 7, and 9 respectively. The differences in levels at the end of the period are modest. For $T = 50$ there are two obvious tendencies. First there is an evident drop in power for T_1 from 0.97 to 0.77 to 0.57 when π_{21} ranges from 0.40 to 0.55. The reason for this is that the levels of the trajectories at the end of the period now equal about 5, 9 and 13 respectively, see panel (b) in Figure 4, and more clear differences in the levels are encountered. Second, and perhaps more interesting, is that the power when $\pi_{21} = 0.55$ and $T = 50$ is lower than that when $\pi_{21} = 0.55$ and $T = 25$, i.e. 0.57 compared to 0.74. This drop in power is explained by a larger jump in level for $T = 50$ than for $T = 25$ (13 compared to 9) and that increasing the length of the time-series reveals the complexity of the process (cf. the (a) and (b) panels in Figure 4). It becomes clear that the term ty_{t-1} is not able to capture distinct changes in level and the autoregressive coefficient at the same time⁵. The same two tendencies, even more pronounced, can be observed when T is increased further. Specifically, when $T = 100$, the power decreases from 1.00 to 0.79 to 0.26 by increasing π_{21} , and in panel (c) in Figure 4, we see that

⁵Despite a panel set-up and the large amount of information available, these results indicate that a test with a third-order Taylor approximation of the transition function in (2) might be preferable. However, this will ruin the analytical tractability of the results derived in Section 3.

the trajectories have now reached their long-run equilibriums, i.e. the levels at 5, 10, and 20 respectively.⁶ Furthermore, the reduction in power from 0.57 (the case $T = 50$ and $\pi_{21} = 0.55$) to 0.26 (the case $T = 100$ and $\pi_{21} = 0.55$) is larger than the reduction in power from 0.74 (the case $T = 25$ and $\pi_{21} = 0.55$) to 0.57 (the case $T = 50$ and $\pi_{21} = 0.55$). Larger reductions in power by increasing T can be explained by the fact that the levels at the end of the sample periods for $T = 25, 50$, and 100 , equal 9, 13, and 20 respectively.

Moreover, it should also be mentioned that increasing T further results in yet another reduction in power. This reduction in power continues actually until $T = 250$ (not reported here), from where the power rapidly increases and reaches unity. This can then be seen as a measure of that we need T (sufficiently) larger than 250 for the term ty_{t-1} to adequately capture the nonlinear structure of an LSTAR(1) model allowing for a modest shift in the intercept and an almost non-stable root at the end of the sample period. We conclude that despite a panel data model with an increased information set, there may still be a need for a third-order approximation of the DGP (1) with (2) to achieve acceptable power when the cross section dimension of the panel is small.

From Table 4 we can see that when the DGP is a PLSTAR(1) model behaving almost like a PTAR(1) model, the empirical powers of the tests are lower than in the previous case. Differences in power compared to what is reported in Table 3 are modest, however, and the response of the tests is robust against the change in γ from 1 to 100 in the PLSTAR model.

4.2.2 A heterogeneous nonlinear panel

A less restricted approach is adopted in the next two experiments because a heterogeneous panel is considered. In the first of these experiments, this is achieved by specifying the following parameter values in the PLSTAR model

$$\begin{aligned} \pi_{i,10} &= 0 \quad \forall i, \quad \pi_{12} = 0.4, \quad \pi_{i,20} \sim U[0.5, 1.5], \\ \pi_{21} &= 0.5, \quad c = T/2, \quad \gamma = 1. \end{aligned} \tag{19}$$

The cross section specific parameter $\pi_{i,20}$ is drawn once from the uniform distribution and thereafter held fixed throughout the replications. By doing this we allow for cross section specific long-run attractors.⁷

In the second experiment we increase the heterogeneity of our PLSTAR model allowing π_{21} to be individual-specific as well, denoted $\pi_{i,21}$. We choose the same design as in (19) but let $\pi_{i,21} \sim U[0.5, 0.55]$.⁸ The empirical powers for these two experiments are reported in Table 5.

In Table 5 we can see that the power for T_1 is satisfactory for all combinations of T and n , whereas HT3 requires $T, n \geq 50$ or $T = 100$ and $n \geq 5$ to achieve a

⁶These long-run equilibriums are given by $(\pi_{10} + \pi_{20})/(1 - \pi_{11} - \pi_{21})$.

⁷The long-run attractor for individual i is given by $(\pi_{10} + \pi_{i,20})/(1 - \pi_{11} - \pi_{21}) \in [5, 15]$.

⁸The long-run attractor for individual i is now given by $(\pi_{10} + \pi_{i,20})/(1 - \pi_{11} - \pi_{i,21}) \in [5, 30]$. Furthermore, notice that the model specification in (1) with (2) only supports heterogeneity in $\pi_{i,10}$ and $\pi_{i,20}$.

Table 4: Empirical power of the test statistic in Corollary 3 and the tests in Harris and Tzavalis (1999). The DGP is a PTAR(1) model.

$\gamma = 100$										
$\pi_{11} = 0.40$		$\pi_{21} = 0.40$			$\pi_{21} = 0.50$			$\pi_{21} = 0.55$		
T	n	T_1	HT2	HT3	T_1	HT2	HT3	T_1	HT2	HT3
5	5	0.21	0.01	0.04	0.25	0.01	0.02	0.28	0.00	0.02
5	10	0.28	0.00	0.04	0.35	0.01	0.02	0.39	0.01	0.01
5	25	0.45	0.00	0.03	0.58	0.00	0.01	0.65	0.00	0.00
5	50	0.65	0.00	0.02	0.81	0.00	0.01	0.86	0.00	0.00
5	100	0.87	0.00	0.01	0.97	0.00	0.00	0.98	0.00	0.00
10	5	0.36	0.00	0.10	0.42	0.00	0.04	0.44	0.00	0.01
10	10	0.51	0.00	0.12	0.59	0.00	0.03	0.65	0.00	0.03
10	25	0.81	0.00	0.15	0.89	0.00	0.01	0.93	0.00	0.01
10	50	0.97	0.00	0.21	0.99	0.00	0.00	1.00	0.00	0.00
10	100	1.00	0.00	0.34	1.00	0.00	0.00	1.00	0.00	0.00
25	5	0.67	0.00	0.41	0.59	0.00	0.04	0.57	0.00	0.00
25	10	0.86	0.00	0.65	0.80	0.00	0.03	0.76	0.00	0.00
25	25	1.00	0.00	0.95	0.98	0.00	0.02	0.97	0.00	0.00
25	50	1.00	0.00	1.00	1.00	0.00	0.02	1.00	0.00	0.00
25	100	1.00	0.00	1.00	1.00	0.00	0.02	1.00	0.00	0.00
50	5	0.96	0.20	0.95	0.65	0.00	0.13	0.44	0.00	0.00
50	10	1.00	0.42	1.00	0.86	0.00	0.24	0.62	0.00	0.00
50	25	1.00	0.91	1.00	0.99	0.00	0.56	0.90	0.00	0.00
50	50	1.00	1.00	1.00	1.00	0.00	0.89	0.99	0.00	0.00
50	100	1.00	1.00	1.00	1.00	0.00	0.99	1.00	0.00	0.00
100	5	1.00	0.00	1.00	0.81	0.00	0.74	0.21	0.00	0.00
100	10	1.00	0.00	1.00	0.97	0.00	0.92	0.26	0.00	0.00
100	25	1.00	0.00	1.00	1.00	0.00	1.00	0.41	0.00	0.00
100	50	1.00	0.00	1.00	1.00	0.00	1.00	0.63	0.00	0.00
100	100	1.00	0.00	1.00	1.00	0.00	1.00	0.85	0.00	0.00

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

Table 5: Empirical power of the test statistic in Corollary 3 and the tests in Harris and Tzavalis (1999). Heterogeneous panels.

T	n	T_1		HT2		HT3	
25	5	0.73	(0.73)	0.00	(0.00)	0.07	(0.04)
25	10	0.92	(0.91)	0.00	(0.00)	0.07	(0.02)
25	25	1.00	(1.00)	0.00	(0.00)	0.06	(0.00)
25	50	1.00	(1.00)	0.00	(0.00)	0.04	(0.00)
25	100	1.00	(1.00)	0.00	(0.00)	0.02	(0.00)
50	5	0.71	(0.62)	0.00	(0.00)	0.17	(0.05)
50	10	0.87	(0.84)	0.00	(0.00)	0.25	(0.06)
50	25	1.00	(0.98)	0.00	(0.00)	0.49	(0.08)
50	50	1.00	(1.00)	0.00	(0.00)	0.81	(0.10)
50	100	1.00	(1.00)	0.00	(0.00)	0.98	(0.11)
100	5	0.77	(0.64)	0.00	(0.00)	0.65	(0.34)
100	10	0.93	(0.75)	0.00	(0.00)	0.89	(0.39)
100	25	1.00	(0.97)	0.00	(0.00)	1.00	(0.74)
100	50	1.00	(1.00)	0.00	(0.00)	1.00	(0.93)
100	100	1.00	(1.00)	0.00	(0.00)	1.00	(1.00)

Notes: The nominal size is 5%, and the results are based on 10 000 replications. Values in parentheses correspond to the power from a panel with a higher degree of heterogeneity.

substantial power. In addition, the power for HT3 (the power for HT2 remains zero) is clearly reduced compared to the results in Table 3 with $\pi_{21} = 0.5$. For instance, when $T = n = 50$ the power is reduced by 15% (from 0.95 to 0.81) in the first experiment and by 89% (from 0.95 to 0.10) in the second experiment where the heterogeneity is increased, whereas the power of our test is still unity. In general the reduction is much smaller for our test. This is of course an important aspect for practitioners, and we see that the T_1 test appears more robust than the HT3 test when each individual is allowed for its own long-run attractor, even though the design of the DGP's is not supported by the model specification in (1) with (2).

4.2.3 A heterogeneous linear panel

In the next experiment we examine the power when the DGP is either a stationary (S) or trend stationary (TS) process defined according to the DGP's under the alternative in Harris and Tzavalis (1999). Thus,

$$y_{it} = \alpha_i + \varphi y_{i,t-1} + u_{it}, \quad (\text{S}),$$

Table 6: Empirical power of the test statistic in Corollary 3 and the tests in Harris and Tzavalis (1999). The DGP's are stationary and trend stationary.

T	n	$\varphi = 0.80$				$\varphi = 0.90$				$\varphi = 0.95$			
		S		T.S.		S		T.S.		S		T.S.	
		T_1	HT2	T_1	HT3	T_1	HT2	T_1	HT3	T_1	HT2	T_1	HT3
5	5	0.05	0.29	0.08	0.06	0.05	0.18	0.08	0.05	0.05	0.13	0.09	0.05
5	10	0.03	0.41	0.13	0.07	0.04	0.22	0.11	0.05	0.04	0.14	0.14	0.05
5	25	0.03	0.69	0.42	0.08	0.03	0.39	0.32	0.06	0.04	0.17	0.41	0.05
5	50	0.01	0.91	0.77	0.09	0.02	0.51	0.61	0.06	0.03	0.23	0.70	0.06
5	100	0.00	1.00	0.98	0.11	0.01	0.76	0.95	0.06	0.02	0.35	0.93	0.06
10	5	0.09	0.50	0.11	0.11	0.06	0.30	0.07	0.08	0.06	0.19	0.07	0.07
10	10	0.08	0.70	0.17	0.14	0.05	0.40	0.13	0.08	0.05	0.23	0.11	0.07
10	25	0.08	0.94	0.44	0.20	0.04	0.64	0.29	0.09	0.04	0.34	0.27	0.07
10	50	0.08	0.99	0.82	0.30	0.03	0.87	0.62	0.10	0.02	0.48	0.52	0.07
10	100	0.08	1.00	0.99	0.47	0.02	0.98	0.94	0.13	0.02	0.71	0.89	0.07
25	5	0.29	0.93	0.34	0.43	0.13	0.59	0.15	0.17	0.09	0.36	0.11	0.11
25	10	0.37	0.99	0.62	0.61	0.14	0.79	0.24	0.21	0.08	0.47	0.13	0.11
25	25	0.59	1.00	0.98	0.92	0.17	0.97	0.55	0.31	0.06	0.73	0.31	0.11
25	50	0.82	1.00	1.00	0.99	0.19	1.00	0.91	0.48	0.06	1.00	0.63	0.14
25	100	0.97	1.00	1.00	1.00	0.25	1.00	1.00	0.72	0.04	1.00	0.94	0.19
50	5	0.86	1.00	0.91	0.95	0.29	0.92	0.36	0.43	0.14	0.58	0.16	0.18
50	10	1.00	1.00	1.00	0.99	0.38	0.99	0.64	0.63	0.15	0.78	0.25	0.20
50	25	1.00	1.00	1.00	1.00	0.60	1.00	0.98	0.91	0.18	0.97	0.58	0.32
50	50	1.00	1.00	1.00	1.00	0.81	1.00	1.00	0.99	0.20	1.00	0.90	0.46
50	100	1.00	1.00	1.00	1.00	0.97	1.00	1.00	1.00	0.27	1.00	1.00	0.72
100	5	0.98	1.00	1.00	1.00	0.67	1.00	0.91	0.94	0.28	0.93	0.38	0.42
100	10	1.00	1.00	1.00	1.00	0.87	1.00	1.00	1.00	0.38	0.99	0.65	0.62
100	25	1.00	1.00	1.00	1.00	0.99	1.00	1.00	1.00	0.59	1.00	0.98	0.91
100	50	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	0.81	1.00	1.00	1.00
100	100	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	0.96	1.00	1.00	1.00

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

or

$$y_{it} = \alpha_i + \alpha_i(1 - \varphi)t + \varphi y_{i,t-1} + u_{it}, \quad (\text{T.S.}),$$

where α_i is drawn once from the standard normal distribution and thereafter held fixed throughout the replications and $\varphi \in \{0.80, 0.90, 0.95\}$. The null hypothesis is the DGP in (18).

The results in Table 6 show, as may be expected, that T_1 is clearly inferior to HT2 for all n and T . However, our test is reasonably powerful for $\varphi = 0.80, 0.90$, and $T = 50$ and $n \geq 25$, or $T = 100$ and all n . However, for $\varphi = 0.95$ we see that our test has power close to the nominal size. The situation is different when we instead consider a trend stationary DGP's and compare the power of T_1 to HT3. We see that for $n = 5$ and $T \geq 5$ the differences in power between T_1 and HT3 are marginal. In fact, studying the same situation but assuming that $n > 5$ and $T \geq 5$, our test T_1 actually has higher power than the HT3 test, and in particular, for $\varphi = 0.95$, T_1 performs substantially better than the latter test statistic. We may conclude that our test seems to have reasonable power properties against stationary/trend stationary alternatives, and that the power of T_1 approaches unity at a faster rate with n than the power of HT3.

4.2.4 Power when viewing T and n as asymptotic when the true T is finite

The last experiment concerns the empirical power of the test in Corollary 4 (denoted T_2) when we treat both n and T as asymptotic but T is actually finite. This also demonstrates the mean-shift effect and variance effect addressed in Sub-section 3.2. For comparison the same effects are investigated for the corresponding tests by Harris and Tzavalis (1999), denoted HT2* and HT3*. For this study we choose the same set-up as for the experiment with a homogeneous panel and $\gamma = 1.00$. These findings are presented in Table 7

In this table we see that treating both T and n as asymptotic leads to rather severe size distortions and the statistic T_2 becomes undersized for almost all n and $T \leq 25$. This indicates that the variance effect discussed in Section 3 dominates the mean shift effect, resulting in a net reduction in the size of the test, which agree with the findings reported in Harris and Tzavalis (1999). As a result we see a clear drop in power (cf. Table 3). For $T \geq 50$ and all n the power reduction is marginal. It seems that in order to maintain a test with correct size when letting n and T tend to infinity, a diagonal convergence criterion should be imposed to control for the mean-shift and variance effect.

Table 7: Empirical power and size of the test statistic in Corollary 4 and the tests in Harris and Tzavalis (1999), assuming that T and n are large.

$\gamma = 1.00$									
$\pi_{11} = 0.40$		$\pi_{21} = 0.40$				$\pi_{21} = 0.50$			$\pi_{21} = 0.55$
T	n	Size	Power			Power			Power
		T_2	T_2	HT2*	HT3*	T_2	HT2*	HT3*	T_2
5	5	0.00	0.02	0.01	0.00	0.02	0.00	0.00	0.02
5	10	0.00	0.01	0.00	0.00	0.01	0.00	0.00	0.01
5	25	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.01
5	50	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.01
5	100	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.01
10	5	0.01	0.21	0.00	0.01	0.26	0.00	0.00	0.27
10	10	0.01	0.34	0.00	0.00	0.76	0.00	0.00	0.47
10	25	0.00	0.63	0.00	0.00	0.96	0.00	0.00	0.83
10	50	0.00	0.88	0.00	0.00	1.00	0.00	0.00	0.99
10	100	0.00	1.00	0.00	0.00	1.00	0.00	0.00	1.00
25	5	0.05	0.72	0.01	0.32	0.66	0.00	0.02	0.64
25	10	0.04	0.92	0.00	0.51	0.88	0.00	0.02	0.86
25	25	0.02	1.00	0.00	0.89	1.00	0.00	0.01	0.99
25	50	0.02	1.00	0.00	1.00	1.00	0.00	0.00	1.00
25	100	0.00	1.00	0.0	01.00	1.00	0.00	0.00	1.00
50	5	0.08	0.96	0.18	0.95	0.73	0.00	0.11	0.53
50	10	0.06	1.00	0.38	1.00	0.91	0.00	0.18	0.72
50	25	0.04	1.00	0.87	1.00	1.00	0.00	0.43	0.95
50	50	0.03	1.00	1.00	1.00	1.00	0.00	0.79	1.00
50	100	0.02	1.00	1.00	1.00	1.00	0.00	0.99	1.00
100	5	0.08	1.00	1.00	1.00	0.85	0.00	0.72	0.23
100	10	0.07	1.00	1.00	1.00	0.97	0.00	0.95	0.31
100	25	0.05	1.00	1.00	1.00	1.00	0.00	0.00	0.51
100	50	0.04	1.00	1.00	1.00	1.00	0.00	0.00	0.74
100	100	0.03	1.00	1.00	1.00	1.00	0.00	0.00	0.94

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

5 Conclusions

In this paper we argue that many of the traditional panel data unit root tests are based on too restrictive panels because it is likely that e.g. macroeconomic panels involve cross sections with time series that exhibit structural changes in levels. We also emphasize the importance of testing unit roots in a nonlinear panel which accommodates a smooth shift in levels and the dynamic structure (the PLSTAR model) because here the conventional unit root tests, such as the test in Harris and Tzavalis (1999), are biased towards nonrejection of the null hypothesis.

The unit root test that we derive in the PLSTAR model is based on an auxiliary regression, and inference is based on the LSDV estimator under the assumption that the disturbances are independently and identically distributed under the null hypothesis. It is shown that the test statistic is normally distributed where the first two moments are calculated analytically. Due to the fact that these moments are known for a fixed sample size, we are able to analyze how this will affect the inference when T is infinite.

Finite-sample properties of the test are explored through Monte Carlo simulations and show satisfactory results. The size distortion is modest and the power is generally superior to the power of the tests in Harris and Tzavalis (1999). We especially demonstrate that the traditional tests in Harris and Tzavalis (1999) lack power if the change in levels is too evident, but with more modest shifts the test with a linear trend has reasonable power. There are situations, however, in which our test may be modified in order to increase the power. This can be done by applying a higher-order Taylor approximation to the PLSTAR model, but it will make the analytical results less tractable, and the question is left for further research.

Appendix A

Lemma 5 Let $\mathbf{M}_T = [m_{ij}]_{T \times T}$ be the matrix defined in (5). Then the (i, j) -th element of \mathbf{M}_T can be expressed

$$m_{ij} = \frac{2}{T(T^2 - 1)} ((T + 1)((2T + 1) - 3i - 3j) + 6ij), \quad (20)$$

for any $i, j = 1, \dots, T$.

Proof. Note that the inverse $(\mathbf{X}'_T \mathbf{X}_T)^{-1} = \frac{2}{T(T-1)} \begin{bmatrix} (2T+1) & -3 \\ -3 & 6(T+1)^{-1} \end{bmatrix}$. Thus, the formula for m_{ij} in (20) holds by computing \mathbf{M}_T . ■

Lemma 6 Let \mathbf{C}_T be the $(T \times T)$ matrix defined in (7), and \mathbf{D}_T and \mathbf{M}_T be the $(T \times T)$ matrices defined in (5). Then

$$\begin{aligned} & \text{the } (i, j)\text{-th element of } \mathbf{C}'_T \mathbf{M}_T \mathbf{C}_T \\ = & \begin{cases} \sum_{t=i+1}^T \sum_{s=j+1}^T m_{ts}, & i, j = 1, \dots, T-1. \\ 0, & i = T \text{ or } j = T. \end{cases} \end{aligned} \quad (21)$$

$$\begin{aligned} & \text{the } (i, j)\text{-th element of } \mathbf{C}'_T \mathbf{D}_T \mathbf{M}_T \mathbf{D}_T \mathbf{C}_T \\ = & \begin{cases} \sum_{t=i+1}^T \sum_{s=j+1}^T t s m_{ts}, & i, j = 1, \dots, T-1. \\ 0, & i = T \text{ or } j = T. \end{cases} \end{aligned} \quad (22)$$

$$\begin{aligned} & \text{the } (i, j)\text{-th element of } \mathbf{C}'_T \mathbf{M}_T \mathbf{D}_T \mathbf{C}_T \\ = & \begin{cases} \sum_{t=i+1}^T \sum_{s=j+1}^T s m_{ts}, & i, j = 1, \dots, T-1. \\ 0, & i = T \text{ or } j = T. \end{cases} \end{aligned} \quad (23)$$

$$\begin{aligned} & \text{the } (i, j)\text{-th element of } \mathbf{C}'_T \mathbf{M}_T \\ = & \begin{cases} \sum_{t=i+1}^T m_{tj}, & i = 1, \dots, T-1, j = 1, \dots, T. \\ 0, & i = T. \end{cases} \end{aligned} \quad (24)$$

$$\begin{aligned} & \text{the } (i, j)\text{-th element of } \mathbf{C}'_T \mathbf{D}_T \mathbf{M}_T \\ = & \begin{cases} \sum_{t=i+1}^T t m_{tj}, & i = 1, \dots, T-1, j = 1, \dots, T. \\ 0, & i = T. \end{cases} \end{aligned} \quad (25)$$

Proof. Applying (20) to $\mathbf{C}'_T \mathbf{M}_T \mathbf{C}_T$, $\mathbf{C}'_T \mathbf{D}_T \mathbf{M}_T \mathbf{D}_T \mathbf{C}_T$, $\mathbf{C}'_T \mathbf{M}_T \mathbf{D}_T \mathbf{C}_T$, $\mathbf{C}'_T \mathbf{M}_T$, and $\mathbf{C}'_T \mathbf{D}_T \mathbf{M}_T$, respectively, gives the formulas in (21)-(25). ■

Lemma 7 . Consider model (3) when (4) and (A1)-(A3) in Assumption 1 hold. Furthermore, let $W_{i,jt}$, $j = 1, \dots, 5$, be defined in (5). Then, for any fixed T ,

$$p \lim_{n \rightarrow \infty} n^{-1} \sum_{i=1}^n W_{i,1t} = \frac{1}{15}(T^2 - 4)\sigma_u^2, \quad (26)$$

$$p \lim_{n \rightarrow \infty} n^{-1} \sum_{i=1}^n W_{i,2t} = \frac{1}{420}(T^2 - 4)(11T^2 + 14T - 1)\sigma_u^2, \quad (27)$$

$$p \lim_{n \rightarrow \infty} n^{-1} \sum_{i=1}^n W_{i,3t} = \frac{1}{30}(T^2 - 4)(T + 1)\sigma_u^2, \quad (28)$$

$$p \lim_{n \rightarrow \infty} n^{-1} \sum_{i=1}^n W_{i,4t} = -\frac{1}{2}(T - 2)\sigma_u^2, \quad (29)$$

$$p \lim_{n \rightarrow \infty} n^{-1} \sum_{i=1}^n W_{i,5t} = -\frac{1}{60}(17T + 19)(T - 2)\sigma_u^2. \quad (30)$$

Proof. (i) Under (4), express the model in (3) as

$$\mathbf{y}_{i,-1} = y_{i0}\boldsymbol{\iota}_T + \mathbf{C}_T \mathbf{u}_i. \quad (31)$$

(ii) Note that \mathbf{Q}_T defined in (5) is orthogonal to $\boldsymbol{\iota}_T$ so pre-multiplying \mathbf{Q}_T on both sides of $\mathbf{y}_{i,-1}$ in (31) yields

$$\mathbf{Q}_T \mathbf{y}_{i,-1} = \mathbf{Q}_T \mathbf{C}_T \mathbf{u}_i. \quad (32)$$

Since \mathbf{Q}_T is idempotent it follows from (32) that

$$\begin{aligned} W_{i,1t} &= \mathbf{y}'_{i,-1} \mathbf{Q}_T \mathbf{y}_{i,-1} \\ &= \mathbf{u}'_i \mathbf{C}'_T \mathbf{C}_T \mathbf{u}_i - \mathbf{u}'_i \mathbf{C}'_T \mathbf{M}_T \mathbf{C}_T \mathbf{u}_i. \end{aligned} \quad (33)$$

In (33), we have

$$\mathbf{u}'_i \mathbf{C}'_T \mathbf{C}_T \mathbf{u}_i = \sum_{t=1}^T (T-t)u_{it}^2 + 2 \sum_{t=1}^{T-1} \sum_{s=t+1}^T (T-s)u_{it}u_{is},$$

and

$$\mathbf{u}'_i \mathbf{C}'_T \mathbf{M}_T \mathbf{C}_T \mathbf{u}_i = \sum_{t=2}^T \left(\sum_{i=t}^T \sum_{j=t}^T m_{ij} \right) u_{i,t-1}^2 + 2 \sum_{t=2}^T \sum_{s=t+1}^T \left(\sum_{i=t}^T u_{i,t-1} \sum_{j=s}^T m_{ij} \right) u_{i,s-1}.$$

By assumption (A1), $\{u_{it}\}$ is a sequence of i.i.d. random variables across i and t so that

$$\begin{aligned} E(W_{i,1t}) &= E(\mathbf{u}'_i \mathbf{C}'_T \mathbf{C}_T \mathbf{u}_i) - E(\mathbf{u}'_i \mathbf{C}'_T \mathbf{M}_T \mathbf{C}_T \mathbf{u}_i) \\ &= \sum_{t=1}^T (T-t)\sigma_u^2 - \sum_{t=2}^T \left(\sum_{i=t}^T \sum_{j=t}^T m_{ij} \right) \sigma_u^2, \end{aligned}$$

and it follows now from (21) that $E(W_{i,1t}) = \frac{1}{15}(T^2 - 4)\sigma_u^2$. Finally, by the LLN, (26) holds. Similarly, we can show that (27)-(30) hold. ■

Proof of Theorem 1. (i) It follows from Anderson and Hsiao (1981) that the LSDV estimators of ρ and ϕ for model (3), under the null hypothesis (4), equal

$$\begin{pmatrix} \hat{\rho} - 1 \\ \hat{\phi} \end{pmatrix} = \left[\sum_{i=1}^n \begin{pmatrix} \mathbf{y}'_{i,-1} \mathbf{Q}_T \mathbf{y}_{i,-1} & \mathbf{y}'_{i,-1} \mathbf{Q}_T \mathbf{D}_T \mathbf{y}_{i,-1} \\ \mathbf{y}'_{i,-1} \mathbf{D}_T \mathbf{Q}_T \mathbf{y}_{i,-1} & \mathbf{y}'_{i,-1} \mathbf{D}_T \mathbf{Q}_T \mathbf{D}_T \mathbf{y}_{i,-1} \end{pmatrix} \right]^{-1} \\ \times \left[\sum_{i=1}^n \begin{pmatrix} \mathbf{y}'_{i,-1} \mathbf{Q}_T \mathbf{u}_i \\ \mathbf{y}'_{i,-1} \mathbf{D}_T \mathbf{Q}_T \mathbf{u}_i \end{pmatrix} \right], \quad (34)$$

where \mathbf{Q}_T and \mathbf{D}_T are defined in (5). It follows immediately from (34) that the expression for $(\hat{\rho} - 1)$ in (5) holds.

(ii) It follows from Lemma 7, (31), and (32) that the probability of the numerator of (34) is

$$p \lim_{n \rightarrow \infty} n^{-1} \sum_{i=1}^n \begin{pmatrix} \mathbf{y}'_{i,-1} \mathbf{Q}_T \mathbf{u}_i \\ \mathbf{y}'_{i,-1} \mathbf{D}_T \mathbf{Q}_T \mathbf{u}_i \end{pmatrix} = -\sigma_u^2 \begin{pmatrix} \frac{1}{2}(T-2) \\ \frac{1}{60}(17T+19)(T-2) \end{pmatrix},$$

whereas the probability limit of the denominator in (34) equals

$$p \lim_{n \rightarrow \infty} n^{-1} \sum_{i=1}^n \begin{pmatrix} \mathbf{y}'_{i,-1} \mathbf{Q}_T \mathbf{y}_{i,-1} & \mathbf{y}'_{i,-1} \mathbf{Q}_T \mathbf{D}_T \mathbf{y}_{i,-1} \\ \mathbf{y}'_{i,-1} \mathbf{D}_T \mathbf{Q}_T \mathbf{y}_{i,-1} & \mathbf{y}'_{i,-1} \mathbf{D}_T \mathbf{Q}_T \mathbf{D}_T \mathbf{y}_{i,-1} \end{pmatrix} \\ = \sigma_u^2 \begin{pmatrix} \frac{1}{15}(T^2-4) & \frac{1}{30}(T^2-4)(T+1) \\ \frac{1}{30}(T^2-4)(T+1) & \frac{1}{420}(T^2-4)(11T^2+14T-1) \end{pmatrix}.$$

(iii) By the Slutsky Theorem (see, e.g. page 286 of Davidson (1994)) we obtain

$$p \lim_{n \rightarrow \infty} \begin{pmatrix} \hat{\rho} - 1 \\ \hat{\phi} \end{pmatrix} \\ = \left[p \lim_{n \rightarrow \infty} n^{-1} \sum_{i=1}^n \begin{pmatrix} \mathbf{y}'_{i,-1} \mathbf{Q}_T \mathbf{y}_{i,-1} & \mathbf{y}'_{i,-1} \mathbf{Q}_T \mathbf{D}_T \mathbf{y}_{i,-1} \\ \mathbf{y}'_{i,-1} \mathbf{D}_T \mathbf{Q}_T \mathbf{y}_{i,-1} & \mathbf{y}'_{i,-1} \mathbf{D}_T \mathbf{Q}_T \mathbf{D}_T \mathbf{y}_{i,-1} \end{pmatrix} \right]^{-1} \\ \times \left[p \lim_{n \rightarrow \infty} n^{-1} \begin{pmatrix} \sum_{i=1}^n \mathbf{y}'_{i,-1} \mathbf{Q}_T \mathbf{u}_i \\ \sum_{i=1}^n \mathbf{y}'_{i,-1} \mathbf{D}_T \mathbf{Q}_T \mathbf{u}_i \end{pmatrix} \right] \\ = \begin{bmatrix} -0.25(23T^2 - 21T - 74) / ((T^2 - 2)(T + 2)) \\ -3.5 / (T^2 - 2) \end{bmatrix}. \quad (35)$$

provided that $T \geq 2$. The inconsistency of $(\hat{\rho} - 1)$ in (6) is the upper element of the limit vector on the right-hand side of (35). ■

Lemma 8 Let $\mathbf{M} = [m_{ij}]_{T \times T}$ and $\mathbf{N} = [n_{ij}]_{T \times T}$ be any constant matrices. Assume that (A1) and (A3) in Assumption 1 hold. Then,

$$\begin{aligned}
& E(\mathbf{u}_i' \mathbf{M} \mathbf{u}_i)(\mathbf{u}_i' \mathbf{N} \mathbf{u}_i) \\
= & \mu_4 \sum_{t=1}^T m_{tt} n_{tt} + \sigma_u^4 \left\{ \sum_{t=1}^T [m_{tt} \left(\sum_{s=t+1}^T n_{ss} \right)] + \sum_{t=1}^T [n_{tt} \left(\sum_{s=t+1}^T m_{ss} \right)] \right. \\
& + \sum_{t=1}^{T-1} \sum_{s=t+1}^T (m_{ts} n_{ts}) + \sum_{s=1}^{T-1} \sum_{t=s+1}^T (m_{ts} n_{ts}) \\
& \left. + \sum_{t=1}^{T-1} \left[\sum_{s=t+1}^T (m_{ts} n_{st}) \right] + \sum_{s=1}^{T-1} \left[\sum_{t=s+1}^T (m_{ts} n_{st}) \right] \right\}.
\end{aligned}$$

Proof. Because u_{it} is i.i.d., $\mathbf{u}_i' \mathbf{M} \mathbf{u}_i = \sum_{s=1}^T \sum_{t=1}^T m_{st} u_{is} u_{it}$, and $\mathbf{u}_i' \mathbf{N} \mathbf{u}_i = \sum_{s=1}^T \sum_{t=1}^T n_{st} u_{is} u_{it}$, it follows by collecting all the terms of $m_{tt} n_{tt}$ for $t = 1, \dots, T$ in $(\mathbf{u}_i' \mathbf{M} \mathbf{u}_i)(\mathbf{u}_i' \mathbf{N} \mathbf{u}_i)$ and now taking the expectations we obtain the coefficient of μ_4 . For the coefficients of σ_u^4 , considering all the terms of $m_{ts} n_{ts}$, $m_{ts} n_{st}$ and $m_{tt} n_{ss}$, for $t \neq s$, in $(\mathbf{u}_i' \mathbf{M} \mathbf{u}_i)(\mathbf{u}_i' \mathbf{N} \mathbf{u}_i)$ and computing all the expectations for those terms yield the coefficients for σ_u^4 . ■

Lemma 9 Consider model (3) when (4) and (A1)-(A3) in Assumption 1 hold. Let $W_{i,jt}$, $j = 1, \dots, 5$, be defined in (5). Then, for any fixed $T > 1$,

$$\begin{aligned}
E(W_{i,1t})^2 &= \frac{1}{210} \left\{ (T^2 - 4) / (T(T^2 - 1)) \right\} \times \{ \mu_4 (T^4 - 25) \\
&\quad + \sigma_u^4 \left[\frac{1}{10} (13T^5 - 30T^4 + 10T^3 - 23T + 750) \right] \}, \tag{36}
\end{aligned}$$

$$\begin{aligned}
E(W_{i,2t})^2 &= \frac{1}{1081080} \{ (T^2 - 4) / (T(T^2 - 1)) \} \times \{ \mu_4(1382T^8 + 3185T^7 \\
&\quad + 4141T^6 + 3770T^5 - 80204T^4 - 153907T^3 + 163997T^2 \\
&\quad + 343512T - 2484) + \sigma_u^4 [\frac{1}{210}(244231T^9 - 282072T^8 \\
&\quad - 1261143T^7 - 1695918T^6 - 7526883T^5 + 40810812T^4 \\
&\quad + 99837023T^3 - 95101902T^2 - 215126028T \\
&\quad + 1564920)] \}, \tag{37}
\end{aligned}$$

$$\begin{aligned}
E(W_{i,3t})^2 &= \frac{1}{13860} \{ (T^2 - 4) / (T(T^2 - 1)) \} \times \{ \mu_4(24T^6 + 33T^5 + 32T^4 \\
&\quad - 926T^2 - 825T + 1182) + \sigma_u^4 [\frac{1}{30}(715T^7 - 873T^6 - 902T^5 \\
&\quad - 1890T^4 - 7205T^3 + 81063T^2 + 78672T - 106380)] \}, \tag{38}
\end{aligned}$$

$$\begin{aligned}
E(W_{i,4t})^2 &= \frac{1}{105} \{ (T - 2) / (T(T^2 - 1)) \} \times \{ \mu_4(29T^3 - 47T^2 - 23T + 59) \\
&\quad + \sigma_u^4 [\frac{1}{2}(56T^4 - 167T^3 + 226T^2 + 131T - 354)] \}, \tag{39}
\end{aligned}$$

$$\begin{aligned}
E(W_{i,5t})^2 &= \frac{1}{630} \{ (T - 2) / (T(T^2 - 1)) \} \times \{ \mu_4(64T^5 + 29T^4 - 37T^3 + 16T^2 \\
&\quad - 357T - 435) + \sigma_u^4 [\frac{1}{20}(1153T^6 - 904T^5 - 2840T^4 - 3130T^3 \\
&\quad - 1013T^2 + 23834T + 26100)] \}. \tag{40}
\end{aligned}$$

Proof. For instance, by the definition of $W_{i,3t}$ we can write

$$\begin{aligned}
E(W_{i,3t})^2 &= E(\mathbf{u}_i' \mathbf{C}_T' \mathbf{D}_T \mathbf{C}_T \mathbf{u}_i)^2 - 2E(\mathbf{u}_i' \mathbf{C}_T' \mathbf{D}_T \mathbf{C}_T \mathbf{u}_i)(\mathbf{u}_i' \mathbf{C}_T' \mathbf{M}_T \mathbf{D}_T \mathbf{C}_T \mathbf{u}_i) \\
&\quad + E(\mathbf{u}_i' \mathbf{C}_T' \mathbf{M}_T \mathbf{D}_T \mathbf{C}_T \mathbf{u}_i)^2. \tag{41}
\end{aligned}$$

From the assumptions on u_{it} and Lemmas 6 and 8 we are able to compute the first term of the right-hand side of (41),

$$\begin{aligned}
&E(\mathbf{u}_i' \mathbf{C}_T' \mathbf{D}_T \mathbf{C}_T \mathbf{u}_i)^2 \\
&= \mu_4 \left[\sum_{t=1}^T \left(\sum_{j=1}^T j - \sum_{i=1}^t i \right)^2 \right] + \sigma_u^4 \left\{ 2 \sum_{t=1}^T \left[\left(\sum_{j=1}^T j - \sum_{i=1}^t i \right) \left(\sum_{s=t+1}^T \left(\sum_{j=1}^T j - \sum_{i=1}^s i \right) \right) \right] \right. \\
&\quad \left. + 4 \sum_{t=1}^{T-1} \sum_{s=t+1}^T \left[\left(\sum_{j=1}^T j - \sum_{i=1}^s i \right)^2 \right] \right\} \\
&= \mu_4 \left[\frac{1}{60} T (T^2 - 1) (8T^2 + 5T - 2) \right] \\
&\quad + \sigma_u^4 \left[\frac{1}{180} T (T^2 - 1) (T - 2) (50T^2 + 22T - 21) \right], \tag{42}
\end{aligned}$$

for the second term of the right-hand side of (41), we have

$$\begin{aligned}
& E(\mathbf{u}_i' \mathbf{C}_T' \mathbf{D}_T \mathbf{C}_T \mathbf{u}_i)(\mathbf{u}_i' \mathbf{C}_T' \mathbf{M}_T \mathbf{D}_T \mathbf{C}_T \mathbf{u}_i) \\
&= \mu_4 \left[\sum_{t=1}^T \left(\sum_{j=1}^T j - \sum_{k=1}^t k \right) \left(\sum_{k=t+1}^T \sum_{j=t+1}^T j m_{kj} \right) \right] \\
&\quad + \sigma_u^4 \left\{ \sum_{t=1}^T \left[\left(\sum_{j=1}^T j - \sum_{k=1}^t k \right) \left(\sum_{s=t+2}^T \left(\sum_{k=s}^T \sum_{j=s}^T j m_{kj} \right) \right) \right] \right. \\
&\quad + \sum_{t=1}^T \left[\left(\sum_{k=t+1}^T \sum_{j=t+1}^T j m_{kj} \right) \left(\sum_{s=t+1}^T \left(\sum_{j=1}^T j - \sum_{k=1}^s k \right) \right) \right] \\
&\quad + 2 \sum_{t=1}^{T-1} \sum_{s=t+1}^T \left[\left(\sum_{j=1}^T j - \sum_{k=1}^s k \right) \left(\sum_{k=t+1}^T \sum_{j=s+1}^T j m_{kj} \right) \right] \\
&\quad \left. + 2 \sum_{s=1}^{T-1} \sum_{t=s+1}^T \left[\left(\sum_{j=1}^T j - \sum_{k=1}^t k \right) \left(\sum_{k=t+1}^T \sum_{j=s+1}^T j m_{kj} \right) \right] \right\} \\
&= \mu_4 \left[\frac{1}{120} (T+1) (15T^4 - 7T^3 - 12T^2 + 8T + 8) \right] \\
&\quad + \sigma_u^4 \left[\frac{1}{15 \cdot 120} (T-2) (4037T^5 + 1717T^4 - 4811T^3 \right. \\
&\quad \left. + 779T^2 + 4758T + 1584) \right], \tag{43}
\end{aligned}$$

and for the third term we find that,

$$\begin{aligned}
& E(\mathbf{u}_i' \mathbf{C}_T' \mathbf{M}_T \mathbf{D}_T \mathbf{C}_T \mathbf{u}_i)^2 \\
&= \mu_4 \left[\sum_{t=2}^T \left(\sum_{k=t}^T \sum_{j=t}^T j m_{kj} \right)^2 \right] \\
&\quad + \sigma_u^4 \left\{ 2 \sum_{t=2}^T \left[\left(\sum_{k=t}^T \sum_{j=t}^T j m_{kj} \right) \left(\sum_{s=t+1}^T \left(\sum_{k=s}^T \sum_{j=s}^T j m_{kj} \right) \right) \right] + \sum_{t=2}^{T-1} \sum_{s=t+1}^T \left(\sum_{k=t}^T \sum_{j=s}^T j m_{kj} \right)^2 \right. \\
&\quad \left. + \sum_{t=2}^{T-1} \sum_{s=t+1}^T \left(\sum_{k=s}^T \sum_{j=t}^T j m_{kj} \right)^2 + 2 \sum_{t=2}^{T-1} \left[\sum_{s=t+1}^T \left(\sum_{k=t}^T \sum_{j=s}^T j m_{kj} \right) \left(\sum_{k=s}^T \sum_{j=t}^T j m_{kj} \right) \right] \right\} \\
&= \mu_4 \left[\frac{1}{13860} (1641T^8 + 726T^7 - 3760T^6 - 594T^5 + 4259T^4 \right. \\
&\quad \left. + 792T^3 + 1652T^2 + 1452T - 4728) \right] / (T(T^2 - 1)) \\
&\quad + \sigma_u^4 \left[\frac{1}{69300} (T-2) (17875T^8 + 7362T^7 - 34963T^6 \right. \\
&\quad \left. + 7726T^5 + 50465T^4 + 10018T^3 + 4603T^2 - 6026T \right. \\
&\quad \left. - 35460) \right] / (T(T^2 - 1)). \tag{44}
\end{aligned}$$

Substituting (42)-(44) for $E(\mathbf{u}_i' \mathbf{C}_T' \mathbf{D}_T \mathbf{C}_T \mathbf{u}_i)^2$, $E(\mathbf{u}_i' \mathbf{C}_T' \mathbf{D}_T \mathbf{C}_T \mathbf{u}_i)$
 $(\mathbf{u}_i' \mathbf{C}_T' \mathbf{M}_T \mathbf{D}_T \mathbf{C}_T \mathbf{u}_i)$ and $E(\mathbf{u}_i' \mathbf{C}_T' \mathbf{M}_T \mathbf{D}_T \mathbf{C}_T \mathbf{u}_i)^2$, respectively, on the right-hand side of
(41) gives (38). Similarly we can show that (36), (37), (39), and (40) hold. ■

Lemma 10 Consider model (3) when (4) and (A1)-(A3) in Assumption 1 hold. Let
 $W_{i,jt}$, $j = 1, \dots, 5$, be defined in (5). Then, for any fixed $T > 1$,

$$\begin{aligned}
E(W_{i,1t}W_{i,2t}) &= \frac{1}{83160} \{ (T^2 - 4) / (T(T^2 - 1)) \} \times \{ \mu_4(155T^6 + 198T^5 \\
&\quad + 104T^4 - 5677T^2 - 4950T + 7290) + \sigma_u^4 [\frac{1}{5}(935T^7 \\
&\quad - 1038T^6 - 3157T^5 - 570T^4 - 4675T^3 + 82878T^2 \\
&\quad + 78177T - 109350)] \}, \\
E(W_{i,1t}W_{i,3t}) &= \frac{1}{420} \{ (T^2 - 4) / (T(T - 1)) \} \times \{ \mu_4(T^2 - 5)(T^2 + 5) \\
&\quad + \sigma_u^4 [\frac{1}{10}(13T^5 - 30T^4 + 10T^3 - 23T + 750)] \}, \\
E(W_{i,1t}W_{i,4t}) &= -\frac{1}{420} \{ (T^2 - 4) / (T(T^2 - 1)) \} \times \{ \mu_4(14T^3 - 15T^2 - 14T - 33) \\
&\quad + \sigma_u^4 (14T^4 - 42T^3 + 31T^2 + 42T + 99) \}, \\
E(W_{i,1t}W_{i,5t}) &= -\frac{1}{840} \{ (T^2 - 4) / (T(T^2 - 1)) \} \times \{ \mu_4(17T^4 - 15T^3 + 9T^2 \\
&\quad - 33T - 170) + \sigma_u^4 [\frac{1}{15}(253T^5 - 555T^4 + 175T^3 - 615T^2 \\
&\quad + 1732T + 7650)] \}, \\
E(W_{i,2t}W_{i,3t}) &= \frac{1}{166320} \{ (T^2 - 4) / (T(T - 1)) \} \times \{ \mu_4(245T^6 + 198T^5 \\
&\quad + 290T^4 - 11839T^2 - 4950T + 26424) \\
&\quad + \sigma_u^4 [\frac{5}{2}(539T^7 - 1194T^6 - 649T^5 - 1680T^4 \\
&\quad - 8899T^3 + 87654T^2 + 44649T - 198180)] \}, \\
E(W_{i,2t}W_{i,4t}) &= -\frac{1}{2520} \{ (T^2 - 4) / (T(T^2 - 1)) \} \times \{ \mu_4(30T^5 + 5T^4 + 27T^3 \\
&\quad - 29T^2 - 489T - 312) + \sigma_u^4 [\frac{1}{5}(163T^6 - 317T^5 - 95T^4 \\
&\quad - 545T^3 + 292T^2 + 7342T + 4680)] \},
\end{aligned}$$

$$\begin{aligned}
E(W_{i,2t}W_{i,5t}) &= -\frac{1}{166320} \{(T^2 - 4)/(T(T^2 - 1))\} \times \{\mu_4(1682T^6 + 1089T^5 \\
&\quad + 1502T^4 - 396T^3 - 34792T^2 - 29205T + 18648) \\
&\quad + \sigma_u^4[\frac{1}{5}(6490T^7 - 8598T^6 - 11825T^5 - 64110T^4 - 33770T^3 \\
&\quad + 546828T^2 + 466785T - 279720)]\}, \\
E(W_{i,3t}W_{i,4t}) &= -\frac{1}{840} \{(T^2 - 4)/(T(T^2 - 1))\} \times \{\mu_4(13T^4 - T^3 + 7T^2 \\
&\quad - 47T - 164) + \sigma_u^4[\frac{1}{15}(208T^5 - 375T^4 - 5T^3 - 525T^2 \\
&\quad + 1957T + 7380)]\}, \\
E(W_{i,3t}W_{i,5t}) &= -\frac{1}{5040} \{(T^2 - 4)/(T(T^2 - 1))\} \times \{\mu_4(63T^5 + 10T^4 \\
&\quad + 24T^3 - 58T^2 - 951T - 624) + \sigma_u^4[\frac{1}{5}(251T^6 - 319T^5 \\
&\quad - 280T^4 - 1720T^3 + 749T^2 + 14999T + 9360)]\}, \\
E(W_{i,4t}W_{i,5t}) &= \frac{1}{210} \{(T - 2)/(T(T^2 - 1))\} \times \{\mu_4(29T^4 - 3T^3 - 40T^2 \\
&\quad - 15T - 43) + \sigma_u^4[\frac{1}{2}(63T^5 - 104T^4 - 52T^3 + 170T^2 \\
&\quad + 97T + 258)]\}.
\end{aligned}$$

Proof. We only give the proof for $E(W_{i,2t}W_{i,3t})$. Thus, write

$$\begin{aligned}
&E(W_{i,2t}W_{i,3t}) \\
&= E(\mathbf{u}'_i \mathbf{C}'_T \mathbf{D}_T^2 \mathbf{C}_T \mathbf{u}_i)(\mathbf{u}'_i \mathbf{C}'_T \mathbf{D}_T \mathbf{C}_T \mathbf{u}_i) \\
&\quad - E(\mathbf{u}'_i \mathbf{C}'_T \mathbf{D}_T^2 \mathbf{C}_T \mathbf{u}_i)(\mathbf{u}'_i \mathbf{C}'_T \mathbf{M}_T \mathbf{D}_T \mathbf{C}_T \mathbf{u}_i) \\
&\quad - E(\mathbf{u}'_i \mathbf{C}'_T \mathbf{D}_T \mathbf{M}_T \mathbf{D}_T \mathbf{C}_T \mathbf{u}_i)(\mathbf{u}'_i \mathbf{C}'_T \mathbf{D}_T \mathbf{C}_T \mathbf{u}_i) \\
&\quad + E(\mathbf{u}'_i \mathbf{C}'_T \mathbf{D}_T \mathbf{M}_T \mathbf{D}_T \mathbf{C}_T \mathbf{u}_i)(\mathbf{u}'_i \mathbf{C}'_T \mathbf{M}_T \mathbf{D}_T \mathbf{C}_T \mathbf{u}_i). \tag{45}
\end{aligned}$$

Further manipulation for those terms on the right-hand side of (45), by applying Lemmas 6 and 8, yields,

$$\begin{aligned}
&E(\mathbf{u}'_i \mathbf{C}'_T \mathbf{D}_T^2 \mathbf{C}_T \mathbf{u}_i)(\mathbf{u}'_i \mathbf{C}'_T \mathbf{D}_T \mathbf{C}_T \mathbf{u}_i) \\
&= \mu_4 \left\{ \sum_{t=1}^{T-1} \left[\left(\sum_{j=1}^T j^2 - \sum_{k=1}^t k^2 \right) \left(\sum_{j=1}^T j - \sum_{k=1}^t k \right) \right] \right\} \\
&\quad + \sigma_u^4 \left\{ \sum_{t=1}^{T-2} \left[\left(\sum_{j=1}^T j^2 - \sum_{k=1}^t k^2 \right) \left(\sum_{j=t+1}^{T-1} \left(\sum_{s=1}^T s - \sum_{k=1}^j k \right) \right) \right] \right. \\
&\quad \left. + \sum_{t=1}^{T-2} \left[\left(\sum_{j=1}^T j - \sum_{k=1}^t k \right) \left(\sum_{j=t+1}^{T-1} \left(\sum_{s=1}^T s^2 - \sum_{k=1}^j k^2 \right) \right) \right] \right. \\
&\quad \left. + 4 \sum_{t=1}^{T-2} \sum_{s=t+1}^{T-1} \left[\left(\sum_{j=1}^T j^2 - \sum_{k=1}^s k^2 \right) \left(\sum_{j=1}^T j - \sum_{k=1}^s k \right) \right] \right\} \\
&= \mu_4 \left[\frac{1}{72} T^2 (7T + 2) (T - 1) (T + 1)^2 \right] \\
&\quad + \sigma_u^4 \left[\frac{1}{840} T (T^2 - 1) (T - 2) (178T^3 + 209T^2 + 15T - 12) \right], \tag{46}
\end{aligned}$$

$$\begin{aligned}
& E(\mathbf{u}_i' \mathbf{C}_T' \mathbf{D}_T^2 \mathbf{C}_T \mathbf{u}_i) (\mathbf{u}_i' \mathbf{C}_T' \mathbf{M}_T \mathbf{D}_T \mathbf{C}_T \mathbf{u}_i) \\
= & \mu_4 \left\{ \sum_{t=1}^{T-1} \left[\left(\sum_{k=t+1}^T \sum_{j=t+1}^T j m_{kj} \right) \left(\sum_{j=1}^T j^2 - \sum_{k=1}^t k^2 \right) \right] \right\} \\
& + \sigma_u^4 \left\{ \sum_{t=1}^{T-2} \left[\left(\sum_{k=t+1}^T \sum_{j=t+1}^T j m_{kj} \right) \left(\sum_{j=t+1}^{T-1} \left(\sum_{s=1}^T s^2 - \sum_{k=1}^j k^2 \right) \right) \right] \right. \\
& + 2 \sum_{t=1}^{T-2} \sum_{s=t+1}^{T-1} \left[\left(\sum_{k=t+1}^T \sum_{j=s+1}^T j m_{kj} \right) \left(\sum_{j=1}^T j^2 - \sum_{k=1}^s k^2 \right) \right] \\
& \left. + 2 \sum_{s=1}^{T-2} \sum_{t=s+1}^{T-1} \left[\left(\sum_{k=t+1}^T \sum_{j=s+1}^T j m_{kj} \right) \left(\sum_{j=1}^T j^2 - \sum_{k=1}^t k^2 \right) \right] \right\} \\
= & \mu_4 \left[\frac{1}{15 \cdot 120} (1367T^6 + 1635T^5 - 1051T^4 - 1263T^3 \right. \\
& \left. + 1196T^2 + 1572T + 144) \right] \\
& + \sigma_u^4 \left[\frac{1}{7560} (T-2) (1541T^6 + 1806T^5 - 1030T^4 \right. \\
& \left. - 678T^3 + 1793T^2 + 1608T + 288) \right], \tag{47}
\end{aligned}$$

$$\begin{aligned}
& E(\mathbf{u}_i' \mathbf{C}_T' \mathbf{D}_T \mathbf{M}_T \mathbf{D}_T \mathbf{C}_T \mathbf{u}_i) (\mathbf{u}_i' \mathbf{C}_T' \mathbf{D}_T \mathbf{C}_T \mathbf{u}_i) \\
= & \mu_4 \left\{ \sum_{t=1}^{T-1} \left[\left(\sum_{k=t+1}^T \sum_{j=t+1}^T k j m_{kj} \right) \left(\sum_{j=1}^T j - \sum_{k=1}^t k \right) \right] \right\} \\
& + \sigma_u^4 \left\{ \sum_{t=1}^{T-2} \left[\left(\sum_{k=t+1}^T \sum_{j=t+1}^T k j m_{kj} \right) \left(\sum_{j=t+1}^{T-1} \left(\sum_{s=1}^T s - \sum_{k=1}^j k \right) \right) \right] \right. \\
& + \sum_{t=1}^{T-2} \left[\left(\sum_{j=1}^T j - \sum_{k=1}^t k \right) \left(\sum_{j=t+1}^{T-1} \left(\sum_{k=j+1}^T \sum_{s=j+1}^T k s m_{ks} \right) \right) \right] \\
& \left. + 4 \sum_{t=1}^{T-2} \sum_{s=t+1}^{T-1} \left[\left(\sum_{k=t+1}^T \sum_{j=s+1}^T k j m_{kj} \right) \left(\sum_{j=1}^T j - \sum_{k=1}^s k \right) \right] \right\} \\
= & \mu_4 \left[\frac{1}{7560} (691T^6 + 855T^5 - 509T^4 - 783T^3 + 394T^2 + 792T + 144) \right] \\
& + \sigma_\mu^4 \left[\frac{1}{3780} (T-2) (760T^6 + 879T^5 - 554T^4 \right. \\
& \left. - 492T^3 + 694T^2 + 765T + 180) \right], \tag{48}
\end{aligned}$$

and

$$\begin{aligned}
& E(\mathbf{u}_i' \mathbf{C}_T' \mathbf{D}_T \mathbf{M}_T \mathbf{D}_T \mathbf{C}_T \mathbf{u}_i)(\mathbf{u}_i' \mathbf{C}_T' \mathbf{M}_T \mathbf{D}_T \mathbf{C}_T \mathbf{u}_i) \\
= & \mu_4 \left\{ \sum_{t=1}^{T-1} \left[\left(\sum_{k=t+1}^T \sum_{j=t+1}^T k j m_{kj} \right) \left(\sum_{j=t+1}^T \sum_{k=t+1}^T j m_{kj} \right) \right] \right\} \\
& + \sigma_u^4 \left\{ \sum_{t=1}^{T-2} \left[\left(\sum_{k=t+1}^T \sum_{j=t+1}^T k j m_{kj} \right) \left(\sum_{j=t+1}^{T-1} \left(\sum_{s=j+1}^T \sum_{k=j+1}^T s m_{ks} \right) \right) \right] \right. \\
& + \sum_{t=1}^{T-2} \left[\left(\sum_{k=t+1}^T \sum_{j=t+1}^T j m_{kj} \right) \left(\sum_{j=t+1}^{T-1} \left(\sum_{s=j+1}^T \sum_{k=s+1}^T k s m_{ks} \right) \right) \right] \\
& + 2 \sum_{t=1}^{T-2} \sum_{s=t+1}^{T-1} \left[\left(\sum_{k=t+1}^T \sum_{j=s+1}^T j m_{kj} \right) \left(\sum_{k=t+1}^T \sum_{j=s+1}^T k j m_{kj} \right) \right] \\
& \left. + 2 \sum_{s=1}^{T-2} \sum_{t=s+1}^{T-1} \left[\left(\sum_{k=t+1}^T \sum_{j=s+1}^T j m_{kj} \right) \left(\sum_{k=s+1}^T \sum_{j=t+1}^T k j m_{kj} \right) \right] \right\} \\
= & \mu_4 \left[\frac{1}{83160T(T-1)} (7157T^8 + 1067T^7 - 13952T^6 \right. \\
& + 44T^5 + 11887T^4 + 1661T^3 + 21908T^2 \\
& + 7524T - 52848)] \\
& + \sigma_\mu^4 \left[\frac{1}{46200T(T-1)} (T-2) (8976T^8 + 1356T^7 \right. \\
& - 14723T^6 + 5768T^5 + 16904T^4 + 6014T^3 \\
& \left. + 9883T^2 - 16058T - 44040) \right]. \tag{49}
\end{aligned}$$

By (46)-(49) further algebra gives (45). Similarly, we can show that the remaining formulas in Lemma 10 hold. ■

Lemma 11 *Consider model (3) when (4) and (A1)-(A3) in Assumption 1 hold. Then, for any fixed $T \geq 2$,*

$$\sqrt{n} \left[\begin{pmatrix} \hat{\rho} - 1 \\ \hat{\phi} \end{pmatrix} - \begin{pmatrix} B_1(T) \\ B_2(T) \end{pmatrix} \right] \xrightarrow{d} N \left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \mathbf{Q}_T^{-1} \mathbf{\Omega}_T \mathbf{Q}_T^{-1} \right), \tag{50}$$

where

$$\mathbf{Q}_T = \begin{pmatrix} E(W_{i,1t}) & E(W_{i,3t}) \\ E(W_{i,3t}) & E(W_{i,2t}) \end{pmatrix}, \tag{51}$$

and

$$\mathbf{\Omega}_T = E \begin{bmatrix} E(G_{1t}^2) & E(G_{1t}G_{2t}) \\ E(G_{1t}G_{2t}) & E(G_{2t}^2) \end{bmatrix}, \tag{52}$$

whereas

$$G_{i,1t} = W_{i,4t} - B_1(T)W_{i,1t} - B_2(T)W_{i,3t}, \tag{53}$$

$$G_{i,2t} = W_{i,5t} - B_1(T)W_{i,3t} - B_2(T)W_{i,2t}, \tag{54}$$

$$B_2(T) = -3.5/(T^2 - 2). \tag{55}$$

Proof. (i) The inconsistency of $\hat{\phi}$ given by $B_2(T)$ in (55) is the lower element of the limit vector on the right-hand side of (35).

(ii) The LSDV estimator of $(\rho, \phi)'$, under the null hypothesis (4), corrected by $[B_1(T), B_2(T)]'$, can be expressed as

$$\begin{aligned}
& \begin{bmatrix} \hat{\rho} - 1 \\ \hat{\phi} \end{bmatrix} - \begin{bmatrix} B_1(T) \\ B_2(T) \end{bmatrix} \\
&= \begin{bmatrix} \sum_{i=1}^n W_{i,1t} & \sum_{i=1}^n W_{i,3t} \\ \sum_{i=1}^n W_{i,3t} & \sum_{i=1}^n W_{i,2t} \end{bmatrix}^{-1} \\
&\quad \times \left(\begin{bmatrix} \sum_{i=1}^n W_{i,4t} \\ \sum_{i=1}^n W_{i,5t} \end{bmatrix} - \begin{bmatrix} \sum_{i=1}^n W_{i,1t} & \sum_{i=1}^n W_{i,3t} \\ \sum_{i=1}^n W_{i,3t} & \sum_{i=1}^n W_{i,2t} \end{bmatrix} \begin{bmatrix} B_1(T) \\ B_2(T) \end{bmatrix} \right) \\
&= \begin{bmatrix} \sum_{i=1}^n W_{i,1t} & \sum_{i=1}^n W_{i,3t} \\ \sum_{i=1}^n W_{i,3t} & \sum_{i=1}^n W_{i,2t} \end{bmatrix}^{-1} \begin{bmatrix} \sum_{i=1}^n G_{i,1t} \\ \sum_{i=1}^n G_{i,2t} \end{bmatrix}, \tag{56}
\end{aligned}$$

where $W_{i,jt}$, $j = 1, \dots, 5$, are defined in (5), $G_{i,1t}$ and $G_{i,2t}$ are given by (53) and (54), respectively. Note that for a fixed T , $(G_{i,1t}, G_{i,2t})'$ is by Assumption 1 an i.i.d. vector across i with zero mean and finite covariance matrix $\mathbf{\Omega}_T$ defined in (52). It follows from Lemmas 9 and 10 that

$$\begin{aligned}
E(G_{i,1t}^2) &= E(W_{i,4t}^2) + B_1^2(T)E(W_{i,1t}^2) + B_2^2(T)E(W_{i,3t}^2) \\
&\quad - 2B_1(T)E(W_{i,4t}W_{i,1t}) - 2B_2(T)E(W_{i,4t}W_{i,3t}) \\
&\quad + 2B_1(T)B_2(T)E(W_{i,3t}W_{i,1t}) \\
&= \mu_4 \frac{1}{110880} \{[(T-3)(T-2)/((T^2-2)^2(T^2-1)(T+2)T)] \\
&\quad \times (6543T^7 - 13599T^6 - 36602T^5 + 170566T^4 \\
&\quad - 25173T^3 - 707131T^2 - 29344T + 589812)\} \\
&\quad + \sigma_\mu^4 \{ \frac{1}{3326400} (T-2)/((T^2-2)^2(T^2-1)(T+2)T) \} \\
&\quad \times (392843T^9 - 1987498T^8 + 3181711T^7 + 5981726T^6 \\
&\quad - 34139267T^5 + 37748138T^4 + 74407229T^3 - 182674806T^2 \\
&\quad - 70246356T + 159249240),
\end{aligned}$$

$$\begin{aligned}
E(G_{i,1t}G_{i,2t}) &= E(W_{i,4t}W_{i,5t}) - B_1(T)E(W_{i,3t}W_{i,4t}) - B_2(T)E(W_{i,2t}W_{i,4t}) \\
&\quad - B_1(T)E(W_{i,1t}W_{i,5t}) + B_1^2(T)E(W_{i,1t}W_{i,3t}) \\
&\quad + B_1(T)B_2(T)E(W_{i,1t}W_{i,2t}) - B_2(T)E(W_{i,3t}W_{i,5t}) \\
&\quad + B_1(T)B_2(T)E(W_{i,3t}^2) + B_2^2(T)E(W_{i,2t}W_{i,3t}) \\
&= \mu_4 \frac{1}{221760} \{ (T^2-9)(T-2)/((T^2-2)^2(T^2-1)(T+2)T) \} \\
&\quad \times (3647T^7 + 1535T^6 + 3726T^5 - 10450T^4 - 179845T^3 \\
&\quad - 226905T^2 + 219080T + 348188) \\
&\quad + \sigma_u^4 \{ \frac{1}{6652800} (T-2)/((T^2-2)^2(T+1)(T+2)(T-1)T) \} \\
&\quad \times (382943T^{10} - 1478555T^9 - 1151547T^8 + 11553117T^7 \\
&\quad - 1552641T^6 - 12256869T^5 + 21309007T^4 - 120784417T^3 \\
&\quad - 220115082T^2 + 156524484T + 282032280),
\end{aligned}$$

and

$$\begin{aligned}
E(G_{i,2t}^2) &= E(W_{i,5t}^2) + B_1^2(T)E(W_{i,3t}^2) + B_2^2(T)E(W_{i,2t}^2) \\
&\quad - 2B_1(T)E(W_{i,5t}W_{i,3t}) - 2B_2(T)E(W_{i,2t}W_{i,5t}) \\
&\quad + 2B_1(T)B_2(T)E(W_{i,3t}W_{i,2t})
\end{aligned}$$

$$\begin{aligned}
= & \mu_4 \left\{ \left[\frac{1}{2882880} (T-3)(T-2) / \left((T^2-2)^2 (T^2-1)(T+2)T \right) \right] \right. \\
& \times (55489T^9 + 125238T^8 + 298513T^7 + 975550T^6 - 4756977T^5 \\
& - 18293686T^4 - 2616677T^3 + 34056714T^2 + 19476036T \\
& - 6731784) + \sigma_u^4 \left\{ \frac{1}{259459200} (T-2) / \right. \\
& \left. \left((T^2-2)^2 (T^2-1)(T+2)T \right) \right\} \times (11556935T^{11} - 40537573T^{10} \\
& - 55120525T^9 + 286295547T^8 + 63649665T^7 + 923928009T^6 \\
& + 843787177T^5 - 12232771967T^4 - 10684267460T^3 \\
& \left. + 21359553264T^2 + 17174416608T - 5452745040) \right\}.
\end{aligned}$$

As $n \rightarrow \infty$, with T fixed, the Linderberg-Lévy CLT implies that the numerator of (56) converges at the rate \sqrt{n} to a normal random variable,

$$n^{-1/2} \sum_{i=1}^n \begin{bmatrix} G_{i,1t} \\ G_{i,2t} \end{bmatrix} \xrightarrow{d} N \left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \mathbf{\Omega}_T \right). \quad (57)$$

On the other hand, the denominator of (56) converges in probability at the rate n such that

$$n^{-1} \sum_{i=1}^n \begin{bmatrix} W_{i,1t} & W_{i,3t} \\ W_{i,3t} & W_{i,2t} \end{bmatrix} \xrightarrow{p} \mathbf{Q}_T. \quad (58)$$

where \mathbf{Q}_T is defined in (51). It follows from (57) and (58) that as $n \rightarrow \infty$ with T fixed, (50) holds. ■

Proof of Theorem 2. For $T > 2$, \mathbf{Q}_T defined in Lemma 11 is positive definite. Thus, \mathbf{Q}_T^{-1} exists. Let $\mathbf{Q}_T^{-1} = \begin{pmatrix} q_{11} & q_{12} \\ q_{12} & q_{22} \end{pmatrix}$. It follows from Lemma 11 that $\sqrt{n}(\hat{\rho} - 1 - B_1(T)) \xrightarrow{d} N(0, \sigma_{\hat{\rho}}^2(T, \kappa_4))$ holds with

$$\sigma_{\hat{\rho}}^2(T, \kappa_4) = q_{11}^2 E(G_{i,1t}^2) + 2q_{11}q_{12}E(G_{i,1t}G_{i,2t}) + q_{12}^2 E(G_{i,2t}^2). \quad (59)$$

Further manipulation in (59) by using the results in Lemma 11 gives the expression of $\sigma_{\hat{\rho}}^2(T, \kappa_4)$ in Theorem 2. ■

Proof of Corollaries 3. This is an immediate consequence of Theorem 2. ■

References

- ANDERSON, T. W., AND C. HSIAO (1981): “Estimation of dynamic models with error components,” *Journal of the American Statistical Association*, 76, 598–606.
- BAI, J., AND J. SILVESTRE (2003): “Structural changes, common stochastic trends and unit roots in panel data,” Unpublished manuscript.
- DAVIDSON, J. (1994): *Stochastic Limit Theory*. Oxford University Press.
- DICKEY, D. A., AND W. A. FULLER (1979): “Distribution of the Estimators for Autoregressive Time Series,” *Journal of the American Statistical Association*, 74, 427–431.
- GONZÁLEZ, A. G., T. TERÄSVIRTA, AND D. VAN DIJK (2004): “Panel smooth transition regression and an application to investment under credit constraints,” Stockholm School of Economics, Working paper.
- HANSEN, B. E. (1999): “Threshold effects in non-dynamic panels: Estimating, testing, and inference,” *Journal of Econometrics*, 93, 345–368.
- HARRIS, R. D. F., AND E. TZAVALIS (1999): “Inference for unit roots in dynamic panels where the time dimension is fixed,” *Journal of Econometrics*, 91, 201–226.
- 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.
- HSIAO, C. (1986): *Analysis of panel data*. Cambridge University Press.
- IM, K. S., AND J. LEE (1999): “LM unit root test with panel data; a test robust to structural changes,” Wichita State University, Working paper.
- IM, K. S., M. H. PESARAN, AND Y. SHIN (2003): “Testing for unit roots in heterogeneous panels,” *Journal of Econometrics*, 115, 53–74.
- LEVIN, A., C. LIN, AND C. CHU (2002): “Unit root tests in panel data: Asymptotic and finite-sample properties,” *Journal of Econometrics*, 108, 1–24.
- 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.
- NICKELL, S. (1981): “Biases in dynamic models with fixed effects,” *Econometrica*, 49, 1417–1426.
- PERRON, P. (1990): “Testing for a unit root in a time series with a changing mean,” *Journal of Business and Economic Statistics*, 8, 153–162.

- 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.
- QUAH, D. (1994): “Exploiting cross section variation for unit root inference in dynamic data,” *Economic Letters*, 44, 9–19.
- RIPATTI, A., AND P. SAIKKONEN (2001): “Vector autoregressive process with nonlinear time trends in cointegrating relations,” *Macroeconomic Dynamics*, 5, 557–597.
- SILVESTRE, C. I., T. D. BARRIO-CASTRO, AND E. LOPEZ-BAZO (2001): “Level shifts in a panel data based unit root test. An application to the rate of unemployment,” Mimeo, Department of Econometrica, University of Barcelona, Spain.
- TERÄSVIRTA, T. (1998): “Modeling economic relationships with smooth transition regressions,” *Handbook of applied economic statistics*, 507–552. New York: Dekker.
- TZAVALIS, E. (2002): “Structural breaks and unit roots for short panels,” Queen Mary University of London, Working paper No. 459.