WORKING PAPER

6/2012

TESTING LINEAR COINTEGRATION AGAINST SMOOTH-TRANSITION COINTEGRATION

DAO LI, CHANGLI HE
STATISTICS

ISSN 1403-0586

http://www.oru.se/Institutioner/Handelshogskolan-vid-Orebro-universitet/Forskning/Publikationer/Working-papers/
Örebro University School of Business
701 82 Örebro
SWEDEN

Testing Linear Cointegration against Smooth-transition Cointegration

Dao Li*, Changli He†

First version: March 2010, Second version: September 2011,

This version: August 12, 2013

Abstract

This study studies a smooth-transition (ST) type of cointegration. The proposed ST cointegration allows for regime switching structure in a cointegrated system and incorporates the linear cointegration developed by Engle and Granger (1987) and the threshold cointegration studied by Balke and Fomby (1997). We developed F-type tests to examine linear cointegration against ST cointegration in a class of vector ST cointegrating regression models. The null asymptotic distributions of the tests are derived when stationary transition variables are involved. Finite-sample distributions and the small-sample performances of these tests are studied using Monte Carlo simulations. Our F-type tests have better power when the system contains ST cointegration than when the system is linearly cointegrated. The testing procedure in this study is applied to purchasing power parity (PPP) data as an example, where we observe that there is no linear cointegration, but an ST cointegration exists in the system.

KEYWORDS: nonlinear cointegration; threshold cointegration; smooth transition; F test.

1. INTRODUCTION

Cointegration analysis in a vector nonstationary time series has received considerable attention over the last two decades since Engle and Granger (1987) developed linear cointegration. It has been successfully applied to analyzing economic systems in linear frameworks when individual economic variables are tied together by

*Correspondence to: Dao Li, Örebro University School of Business, Örebro, Sweden. This paper has participated in the 66th European Meeting of the Econometric Society, Málaga, August 2012.

†Changli He: School of Technology and Business Studies, Dalarna University, Borlänge, Sweden.

some long-run equilibrium relation. Examples of such applications can be found in Baillie and Selover (1987) and Corbae and Ouliaris (1988) for analyzing purchasing power parity data, in Kremers (1989) for studying a system of GNP and debt, and in Clarida (1994) for investigating imports.

Moreover, many empirical studies of economic systems indicate that most time series data display nonlinear features. In practice, economic data usually display dramatic fluctuations and obvious changes in regimes. Those regime shifts are important for the inference of econometric phenomena (see, e.g., Teräsvirta 1994). As such, the development of cointegration in nonlinear frameworks is necessary. The long-run equilibrium in nonlinear frameworks is represented by a nonlinear combination of the nonstationary variables. Two schemes primarily developed for nonlinear cointegration are threshold cointegration and smooth-transition (ST) cointegration. Granger and Teräsvirta (1993) and Granger (1995) extended linear cointegration in Engle and Granger (1987) to nonlinear frameworks. Balke and Fomby (1997) used a two-step approach to examine threshold cointegration. Park and Phillips (1999, 2001) and Chang et al. (2001) studied parameter estimation in nonlinear cointegrating regressions. Hansen and Seo (2002) tested threshold effects in a two-regime vector error-correction model. Furthermore, Choi and Saikkonen (2004) investigated Lagrange Multiplier (LM) tests to examine linearity in an ST cointegrating regression as discussed in Saikkonen and Choi (2004). There are fruitful applications of nonlinear cointegration in the literature.

In this study, we first propose a definition of ST cointegration that allows for regime switching structure in a cointegrated system. Our definition is flexible in that it covers a wide range of nonlinear cointegration. Moreover, it nests the linear cointegration developed by Engle and Granger (1987) and the threshold cointegration developed by Balke and Fomby (1997). The main purpose of this study is to test linear cointegration against ST cointegration in dependent time series by applying ST cointegrating regression models and F-type tests. We include two ST cointegrating regression models. In one of the models, regressors are $I(1)$ processes whose first-order differences have zero means and regression does not include a constant term. In the other model, regressors are $I(1)$ processes whose first-order differences have nonzero means and regression includes both a constant term and a time trend. The models in this study are similar to the cointegrating regression in Saikkonen and Choi (2004), but we include stationary variables as transition variables rather than $I(1)$ variables. For example, the transition variables could be past values of the first-order differences of the $I(1)$ regressors or certain stationary exogenous variables. This is an important complement of this class of models, because there is no reason to exclude the possibility of the transition variables given by certain stationary variables. In addition, Saikkonen and Choi (2004) only considered the current value of the nonstationary regressors as the transition variables, but lagged values of the differences of the levels or the levels are meaningful when they appear as the transition variables. When choosing nonstationary transition variables, the LM tests in Saikkonen and Choi (2004) have standard asymptotic distributions, whereas our F-type tests have nonstandard distributions when the transition variables are replaced by stationary variables. This has important implications for empirical work. It is crucial to distinguish between the cases with stationary and nonstationary transition variables so that the correct test can be used.

In practice, an additional step of testing unit roots is necessary before detecting ST cointegration to fulfill the assumption in the definition of cointegration that the time series individually is an $I(1)$ process. The unit root tests used for nonlinear cointegration should not be the standard unit root tests, because the possible nonlinearity of each time series may misinform the results of the standard unit root tests. When the individual time series are global stationary time series with nonlinearity, incorrect non-rejection of a unit root in the standard unit root tests could occur. In this study, we test the null of a unit root against nonlinear stationary process denoted by the LSTAR model rather than against a linear stationary process. After linear cointegration is rejected, a diagnosis step needs to be performed, in case cointegration does not exist at all. The stationary assumption of the residuals should hold for the existence of cointegration. In an empirical study involving PPP data (United States-Italy), nonlinear cointegration relation in the system that does not contain linear cointegration is reported.

The outline of the remainder of the manuscript is as follows: Section 2 presents the definition of ST cointegration and the properties of ST function; Section 3 discusses two ST cointegrating regression models. Section 4 develops F-type tests for testing linear cointegration against ST cointegration. Section 5 details the asymptotic distributions of the F-type tests. Section 6 provides finite-sample simulation studies. Section 7 applies our approach to an empirical PPP dataset. The conclusion is presented in Section 8. The proofs are detailed in Appendix A and the simulations of nonstandard unit root tests are provided in Appendix B.

2. DEFINITION OF SMOOTH-TRANSITION COINTEGRATION

In this section, we begin by introducing a definition of ST cointegration that nests the linear cointegration proposed by Engle and Granger (1987) and the threshold cointegration by Balke and Fomby (1997).

First, we describe two processes, $I(0)$ and $I(1)$. A time series y_t is said to be integrated of order 1, denoted by $I(1)$, if its first difference, Δy_t, is $I(0)$, in which $I(0)$ is a strictly stationary process with a long-run variance that is finite and positive (see Hayashi 2000, pages 558-560 for additional details).

Let $\mathbf{y}_t = (y_{1t}, y_{2t}, ..., y_{nt})'$ be an $(n \times 1)$ vector time series in which each of the series taken individually is $I(1)$. As defined in Engle and Granger (1987), if there is a nonzero constant vector $\boldsymbol{\alpha}$, such that $\boldsymbol{\alpha}'\mathbf{y}_t \sim I(0)$, we say that the vector $\mathbf{y}_t$ contains linear cointegration, and the vector $\boldsymbol{\alpha}$ is referred to as the linear cointegrating vector.

However, many empirical studies suggest that the cointegrating vector $\boldsymbol{\alpha}$ could not be constant. For example, it can depend on time t or be a vector of random variables. Threshold cointegration is a special case in which linear cointegration is applicable inside a given regime, but requires different cointegration in other regimes (see Balke and Fomby 1997). Our ST cointegration approach nests both linear and threshold cointegration. It is defined such that the cointegration vector is time-varying or a vector of random variables.

Definition. *Let* $\mathbf{y}_t = (y_{1t}, y_{2t}, ..., y_{nt})'$ *be an* $(n \times 1)$ *vector time series in which each of the series is I(1), i.e.,* $y_{it} \sim I(1)$ *for each* $i = 1, 2, ..., n$*. The vector time series* $\mathbf{y}_t$ *is said to contain ST cointegration if there is an* $(n \times 1)$ *time-varying or random vector* $\boldsymbol{\alpha}_t = (\alpha_{1t}, \alpha_{2t}, ..., \alpha_{nt})'$ *such that the nonlinear combination of the series* $\mathbf{y}_t$ *is I(0). That is,*

$$\boldsymbol{\alpha}_t'\mathbf{y}_t \sim I(0). \tag{2.1}$$

In the vector $\boldsymbol{\alpha}_t$*,* α_{it} *is assumed to be* $\alpha_{it} = \alpha_i G_i(\mathbf{s}_t; \gamma, \mathbf{c})$ *for each* $i = 1, 2, ..., n$*, in which* $G_i(\mathbf{s}_t; \gamma, \mathbf{c})$ *for each* i *is an ST function such that*

$$G(\mathbf{s}_t; \gamma, \mathbf{c}) = \left(1 + exp\left\{-\gamma \cdot \prod_{j=1}^{k} (s_{jt} - c_j)\right\}\right)^{-1}, \quad \gamma > 0 \tag{2.2}$$

where γ *is a slope parameter,* c_j *are location parameters, and* s_{jt} *are time-varying or random transition variables.*

For a different index i, $\mathbf{s}_t$, γ and $\mathbf{c}$ in $G_i(\mathbf{s}_t; \gamma, \mathbf{c})$ are possibly different. We omit the index i for $\mathbf{s}_t$, γ and $\mathbf{c}$ here for simplification of notation. Consider a simple smooth-transition function when $k = 1$ in (2.2):

$$G(s_t; \gamma, c) = (1 + \exp\{-\gamma(s_t - c)\})^{-1}, \quad \gamma > 0. \tag{2.3}$$

$G(s_t; \gamma, c)$ in (2.3) is a non-decreasing nonlinear function of s_t with the restriction $\gamma > 0$. The slope parameter γ measures the speed of transition over time from one regime to another, and the location parameter c gives the point in time in which the transition is symmetric around. $G(s_t; \gamma, c)$ in (2.3) is reduced to the constant 0.5 when γ equals zero, but jumps sharply from one regime ($G(s_t; \gamma, c) = 0$) to another regime ($G(s_t; \gamma, c) = 1$) when γ goes to infinity. When γ is between those two extreme cases, $G(s_t; \gamma, c)$ is a smoothly increasing function of s_t bounded between 0 and 1. The plot of $G(s_t; \gamma, c)$ in Figure 1 as a function of s_t, given γ and c, graphically illustrates the above properties of $G(s_t; \gamma, c)$, in which s_t is a

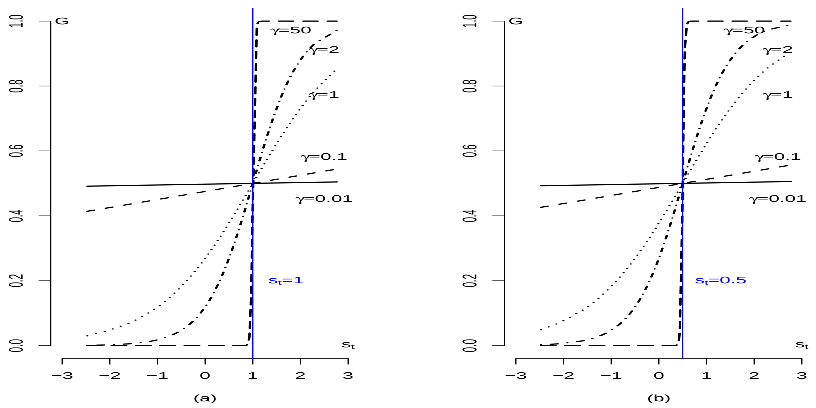

Figure 1: Smooth transition function $G(s_t;\gamma,c)$ of s_t, in which s_t is a standard normal random variable; $\gamma = 0.01, 0.1, 1, 2, 50$, $c = 1$ in panel (a), and $c = 0.5$ in panel (b).

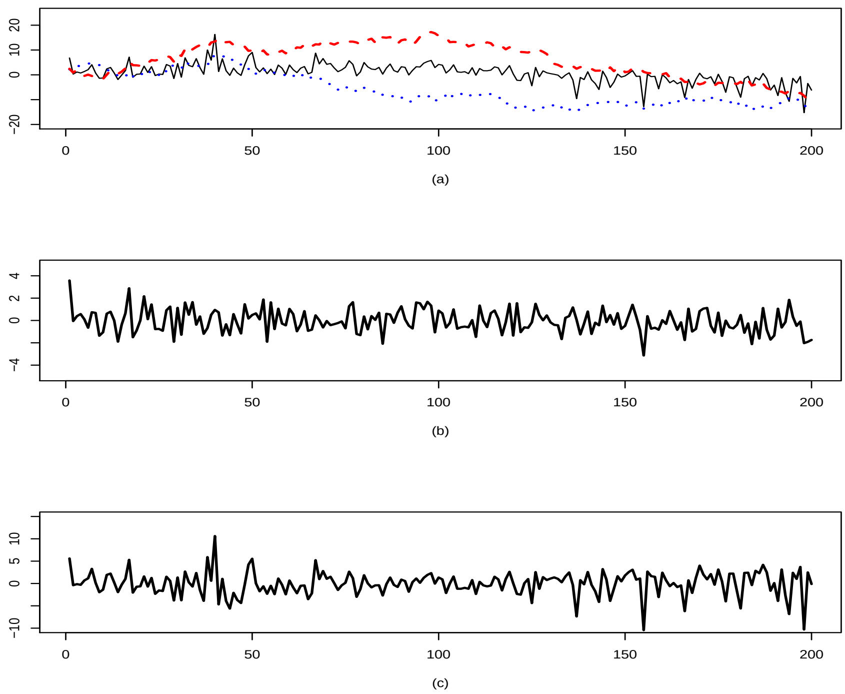

Figure 2: Panel (a) presents three $I(1)$ time series (solid, dotted, and dashed); panel (b) depicts their nonlinear combination (stationary); and panel (c) depicts their linear combination (nonstationary).

standard normal random variable, with $c = 1$ in panel (a), and $c = 0.5$ in panel (b). In each panel, we compare $G(s_t;\gamma,c)$ by varying γ from 0.01 to 50.

Engle and Granger (1987) mentioned that the notion of cointegration can be extended to a vector of series with first-order difference, $\Delta\mathbf{y}_t$ that, has nonzero mean. In that case, the $I(1)$ system $\mathbf{y}_t$ contains both stochastic trends and deterministic trends. If the cointegrating vector defined above eliminates the stochastic trends but not necessarily the deterministic trends, the combination of a nonstationary system $\mathbf{y}_t$ is then trend-stationary rather than $I(0)$ in (2.1). This case has been discussed in the linear framework in Hayashi (2000, pages 629-631) and is referred to as stochastic cointegration. The two cointegrating regression models in the following section are designed for the two cases.

Panel (a) in Figure 2 illustrates a numerical example of the ST cointegration defined in (2.1), in which y_{it} $(i = 1, 2, 3)$ are generated from the following system:

$$\begin{aligned} y_{1t} &= \alpha_{2t}y_{2t} + \alpha_{3t}y_{3t} + \varepsilon_{1t} \\ y_{2t} &= y_{2,t-1} + \varepsilon_{2t} \\ y_{3t} &= y_{3,t-1} + \varepsilon_{3t}, \end{aligned} \tag{2.4}$$

in which $\alpha_{2t} = 0.8\,(1 + \exp\{-3(s_t - 0.8)\})^{-1}$, $\alpha_{3t} = (1 + \exp\{-(s_t - 1)\})^{-1}$, the error term $\boldsymbol{\varepsilon}_t = (\varepsilon_{1t}, \varepsilon_{2t}, \varepsilon_{3t})'$ is a vector *n.i.d.* sequence, and s_t is here simulated from the standard normal distribution. Panel (b) presents their stationary nonlinear combination, whereas their linear combination in panel (c) is not stationary (the test for linear cointegration is achieved by the Phillips-Ouliaris Z_ρ test in Phillips 1987).

3. SMOOTH-TRANSITION COINTEGRATING REGRESSION MODELS

Let an $(n \times 1)$ vector time series $\mathbf{y}_t = (y_{1t}, y_{2t}, ..., y_{nt})'$ be nonlinearly cointegrated according to the definition of ST cointegration in Section 2. We propose two vector ST cointegrating regression models as follows:

$$y_{1t} = \boldsymbol{\alpha}_t'\mathbf{y}_{2t} + u_t \tag{3.1a}$$

$$\mathbf{y}_{2t} = \mathbf{y}_{2,t-1} + \mathbf{v}_t, \tag{3.1b}$$

and

$$y_{1t} = a_1 + \delta t + \boldsymbol{\alpha}_t' \mathbf{y}_{2t} + u_t \tag{3.2a}$$

$$\mathbf{y}_{2t} = \mathbf{a}_2 + \mathbf{y}_{2,t-1} + \mathbf{v}_t, \tag{3.2b}$$

in which $\mathbf{y}_{2t} = (y_{2t}, y_{3t}, ..., y_{nt})'$, $\mathbf{v}_t$ is an $(n-1)$ vector stationary process, each α_{it} in $\boldsymbol{\alpha}_t$ is defined in (2.1), $\mathbf{a}_2$ is an $(n-1) \times 1$ constant vector, a_1 and δ are scalar parameters, the error term $(u_t, \mathbf{v}_t')'$ is serially correlated, and u_t and $\mathbf{v}_t$ are allowed to be correlated.

The two models (3.1a)-(3.1b) and (3.2a)-(3.2b) that contain the defined ST cointegration have triangular representations. The triangular representation, introduced by Phillips (1991), is useful for describing the dynamic structures in a system. In particular, the triangular representation is applied to model the cointegration relation with fewer parameters in a reduced form of vector autoregressions. See Phillips (1991) for more discussions of the triangular representation. In (3.1a), the regression model has no constant term, and the regressor $\mathbf{y}_{2t}$ in (3.1b) is a vector $I(1)$ process whose first-order difference has a zero mean. Furthermore, we consider the case in (3.2a) that the regression model includes both a constant term and a time trend, and the regressor $\mathbf{y}_{2t}$ in (3.2b) is a vector $I(1)$ process whose first-order difference has a nonzero mean.

Considering the properties of $G(\mathbf{s}_t; \gamma, c)$ in (2.3), we discover that the nonlinear cointegrating regression in (3.1a)-(3.1b) or (3.2a)-(3.2b) has the following properties: (a) it is reduced to a linear cointegrating regression when all γ in $\boldsymbol{\alpha}_t$ are equal to zero. (b) It becomes a model containing threshold cointegration when at least one γ in $\boldsymbol{\alpha}_t$ goes to infinity. (c) It allows for regime-switching structures when all γ in $\boldsymbol{\alpha}_t$ are finite and larger than zero, which is proposed as the general form of ST cointegrating regression models in this study.

4. TESTING PROCEDURE

In this section, we develop F-type tests to examine linear cointegration against nonlinear cointegration in the proposed ST cointegrating regression models (3.1a)-(3.1b) and (3.2a)-(3.2b) in Section 3.

However, in practice, an additional step of testing unit roots is necessary before detecting ST cointegration as introduced. This is due to the assumption in the definition of cointegration that the time series individually is an $I(1)$ process. The standard unit root tests are not robust to the possible nonlinearity in the series, because the possible nonlinearity may mislead the results of the standard unit root tests. When individual time series are stationary time series with nonlinearity, incorrect non-rejection of a unit root in the standard unit root tests could occur. In this study, we test the null of a unit root against the nonlinear stationary process denoted by the LSTAR model rather than against a linear stationary process. Finite-sample distributions of the test are provided by simulations for the application, but the asymptotic ones are available upon request to save space by concentrating on the testing for nonlinear cointegration. After linear cointegration is rejected, another diagnosis step needs to be performed, in case cointegration does not exist at all. The KPSS test by Kwiatkowski et al. (1992) can be used to examine the stationary assumption of the residuals.

The hypothesis is introduced by imposing the following parameter restrictions:

$$H_0 : \boldsymbol{\gamma} = 0 \quad \text{against} \quad H_1 : \boldsymbol{\gamma} > 0, \tag{4.1}$$

in which $\boldsymbol{\gamma}$ is an $(n-1) \times 1$ vector with each element γ_i as the slope parameter in (2.3). As pointed out by Luukkonen et al. (1988), the restriction of $\gamma = 0$ leads to the identification problem that the parameter c in (2.3) is not identified under the null of $\gamma = 0$. Therefore, we replace $G(s_t; \gamma, c)$ with its first-order Taylor expansion around $\gamma = 0$ as its linear approximation (ignoring the remainder) if $G(s_t; \gamma, c)$ has satisfied Assumption 1 below.

Assumption 1. *The function $G(s_t;\gamma,c)$ in (2.3) and its derivatives are continuous in an open interval $(-\varepsilon,\varepsilon)$ for $\varepsilon > 0$. $G(s_t;\gamma,c)$ is at least fourth-order differentiable in a neighbourhood of $\gamma = 0$, and $G(s_t;\gamma,c)$ is bounded for all s_t.*

The corresponding auxiliary regressions of (3.1a) and (3.2a) are given in (4.2) and (4.3), respectively, by substituting the first-order expansion for $G(s_t;\gamma,c)$:

$$y_{1t} = \boldsymbol{\beta}'\mathbf{x}_t + u_t^*, \tag{4.2}$$

and

$$y_{1t} = a_1 + \delta t + \boldsymbol{\beta}'\mathbf{x}_t + u_t^*, \tag{4.3}$$

where $\boldsymbol{\beta} = (\boldsymbol{\beta}_2{}', \boldsymbol{\beta}_3{}')'$ and $\mathbf{x}_t = (\mathbf{y}_{2t}', (\mathbf{s}_t\mathbf{y}_{2t})')'$. $\boldsymbol{\beta}_2$ and $\boldsymbol{\beta}_3$ are $(n-1)\times 1$ parameter vectors with each element $\beta_{2i} = \alpha_i/2 - c_i\alpha_i\gamma_i/4$ and $\beta_{3i} = \alpha_i\gamma_i/4$ $(i = 2, 3, ..., n)$. $\mathbf{s}_t$ is an $(n-1)\times(n-1)$ diagonal matrix with diagonal elements that are s_{it} $(i = 2, 3, ..., n)$. In this study, we consider the stationary process as s_{it} in $G_i(s_t;\gamma,c)$, for example, the lagged first difference of a regressor, $\Delta y_{i,t-d}$, or an exogenous variable z_{it}. Under the null of $\boldsymbol{\gamma} = 0$, the error term u_t^* is equivalent to u_t in (3.1a) and (3.2a). In the auxiliary regressions in (4.2) and (4.3), the hypothesis of (4.1) is then rewritten to be

$$H_0 : \boldsymbol{\beta}_3 = \mathbf{0} \quad \text{against} \quad H_1 : \boldsymbol{\beta}_3 \neq 0. \tag{4.4}$$

The estimators of the parameters in (4.2) or (4.3), which minimize the sum of the squared residuals, are consistent when the $I(1)$ regressor $\mathbf{y}_{2t}$ is not cointegrated, as shown in Phillips and Durlauf (1986), Stock (1987), and Hansen (1992a). Thus, the simultaneity bias is not a problem in large samples, even though the error term u_t^* and the $I(1)$ regressors are correlated. In finite samples, however, the bias of the estimators can be very large. For that reason, we provide a more efficient estimator by correcting (4.2) and (4.3) by adding leads and lags of $\Delta\mathbf{y}_{2t}$

(see Dickey and Fuller 1979) as follows:

$$y_{1t} = \boldsymbol{\beta}'\mathbf{x}_t + \sum_{s=-p}^{p} \boldsymbol{\zeta}_s' \Delta \mathbf{y}_{2,t-s} + \tilde{u}_t = \boldsymbol{\beta}'\mathbf{x}_t + \boldsymbol{\zeta}'\mathbf{v}_t^* + \tilde{u}_t = \boldsymbol{\theta}'\tilde{\mathbf{x}}_t + \tilde{u}_t \tag{4.5}$$

and

$$y_{1t} = a_1 + \delta t + \boldsymbol{\beta}'\mathbf{x}_t + \boldsymbol{\zeta}'\mathbf{v}_t^* + \tilde{u}_t \tag{4.6}$$

in which $\tilde{\mathbf{x}}_t = (\mathbf{v}_t^{*'}, \mathbf{x}_t')'$, $\boldsymbol{\theta} = (\boldsymbol{\zeta}', \boldsymbol{\beta}')'$, $\boldsymbol{\zeta} = (\boldsymbol{\zeta}_p', \boldsymbol{\zeta}_{p-1}', ..., \boldsymbol{\zeta}_1', \boldsymbol{\zeta}_0', \boldsymbol{\zeta}_{-1}', ..., \boldsymbol{\zeta}_{-p}')'$, $\mathbf{v}_t^* = (\Delta\mathbf{y}_{2,t-p}', \Delta\mathbf{y}_{2,t-p+1}', ..., \Delta\mathbf{y}_{2,t-1}', \Delta\mathbf{y}_{2t}', \Delta\mathbf{y}_{2,t+1}', ..., \Delta\mathbf{y}_{2,t+p-1}', \Delta\mathbf{y}_{2,t+p}')'$. After abstracting $\mathbf{v}_t^*$ from u_t^*, $\tilde{u}_t$ is now uncorrelated with $\mathbf{v}_\tau$ for all t and τ.

In contrast to $\mathbf{y}_{2t}$ in (4.2), the first-order difference of $\mathbf{y}_{2t}$ in (4.3) is not zero. Thus, $\mathbf{y}_{2t}$ in (4.3) is asymptotically equivalent to a time trend. That makes the regressors in (4.3) collinear in a large sample because a time trend is already included as a separate variable. Note that

$$\mathbf{y}_{2t} = \mathbf{a}_2 + \mathbf{y}_{2,t-1} + \mathbf{v}_t = \mathbf{y}_{20} + \mathbf{a}_2 t + \boldsymbol{\xi}_{2,t-1} + \mathbf{v}_t,$$

where $\boldsymbol{\xi}_{2t} = \boldsymbol{\xi}_{2,t-1} + \mathbf{v}_t = \mathbf{y}_{2t} - \mathbf{a}_2 t$ is an $(n-1)$ vector random walk without drift and $\mathbf{y}_{20}$ is a zero $(n-1)$ vector. Describing the limiting distribution of the estimates therefore further requires a rotation of a detrending transformation. Under the null, (4.6) can be transformed as

$$y_{1t} = a_1 + \delta^* t + \boldsymbol{\beta}_2'\boldsymbol{\xi}_{2t} + \boldsymbol{\beta}_3'\mathbf{s}_t\boldsymbol{\xi}_{2t} + \boldsymbol{\zeta}'\mathbf{v}_t^* + u_t^* = \boldsymbol{\theta}^{*'}\tilde{\mathbf{x}}_t^* + \tilde{u}_t \tag{4.7}$$

in which $\delta^* = \delta + \boldsymbol{\beta}_2'\mathbf{a}_2$, $\tilde{\mathbf{x}}_t^* = (\mathbf{v}_t^{*'}, \boldsymbol{\xi}_{2t}', (\mathbf{s}_t\boldsymbol{\xi}_{2t})', t, 1)'$ and $\boldsymbol{\theta}^* = (\boldsymbol{\zeta}', \boldsymbol{\beta}_2', \boldsymbol{\beta}_3', \delta^*, a_1)'$, and the error term in (4.6) and (4.7) are identical under the null. We now consider the auxillary regression models (4.5) and (4.7). The null hypotheses (4.4) are updated in (4.8) and (4.9), respectively, for (4.5) and (4.7), such that

$$H_0 : \tilde{\mathbf{R}}\boldsymbol{\theta} = \mathbf{r} \tag{4.8}$$

and

$$H_0 : \tilde{\mathbf{R}}^* \boldsymbol{\theta}^* = \mathbf{r} \tag{4.9}$$

where $\tilde{\mathbf{R}} = \left(\mathbf{0}_{(n-1)\times(2p+2)\times(n-1)}, \mathbf{I}_{(n-1)}\right)$, $\tilde{\mathbf{R}}^* = \left(\mathbf{0}_{(n-1)\times(2p+2)\times(n-1)}, \mathbf{I}_{(n-1)}, \mathbf{0}_{(n-1)\times 2}\right)$, $\mathbf{r} = \mathbf{0}_{n-1}$.

To test linear cointegration against ST cointegration in (3.1a)-(3.1b) under the null of (4.1), an F-type statistic is calculated as

$$\tilde{F}_T = \left(\tilde{\mathbf{R}}\tilde{\boldsymbol{\Upsilon}}_T\left(\hat{\boldsymbol{\theta}}_T - \boldsymbol{\theta}_0\right)\right)' \left\{\tilde{s}_T^2 \tilde{\mathbf{R}}\tilde{\boldsymbol{\Upsilon}}_T \left(\sum \tilde{\mathbf{x}}_t \tilde{\mathbf{x}}_t'\right)^{-1} \tilde{\boldsymbol{\Upsilon}}_T \tilde{\mathbf{R}}' m\right\}^{-1} \tilde{\mathbf{R}}\tilde{\boldsymbol{\Upsilon}}_T\left(\hat{\boldsymbol{\theta}}_T - \boldsymbol{\theta}_0\right). \tag{4.10}$$

Similarly, to test the linear cointegration against ST cointegration in (3.2a)-(3.2b) under the null of (4.1), an F-type statistic is calculated as

$$\tilde{F}_T^* = \left(\tilde{\mathbf{R}}^*\tilde{\boldsymbol{\Upsilon}}_T^*\left(\hat{\boldsymbol{\theta}}_T^* - \boldsymbol{\theta}_0^*\right)\right)' \left\{\tilde{s}_T^{*2} \tilde{\mathbf{R}}^*\tilde{\boldsymbol{\Upsilon}}_T^* \left(\sum \tilde{\mathbf{x}}_t^* \tilde{\mathbf{x}}_t^{*\prime}\right)^{-1} \tilde{\boldsymbol{\Upsilon}}_T^* \tilde{\mathbf{R}}^{*\prime} m\right\}^{-1} \tilde{\mathbf{R}}^*\tilde{\boldsymbol{\Upsilon}}_T^*\left(\hat{\boldsymbol{\theta}}_T^* - \boldsymbol{\theta}_0^*\right). \tag{4.11}$$

In (4.10) and (4.11), $\tilde{\boldsymbol{\Upsilon}}_T$ an $\tilde{\boldsymbol{\Upsilon}}_T^*$ are $(2p+4)\times(n-1)\times(2p+4)\times(n-1)$ diagonal matrices of convergence rates, $\tilde{\mathbf{R}}$ and $\tilde{\mathbf{R}}^*$ are $((n-1)\times(2p+4)\times(n-1))$ matrices restricted under the null hypothesis (4.8) and (4.9), $\hat{\boldsymbol{\theta}}_T$ and $\hat{\boldsymbol{\theta}}_T^*$ are the consistent estimator of the true parameter $\boldsymbol{\theta}_0$ and $\boldsymbol{\theta}_0^*$ under the null, $\tilde{s}_T^2$ and $\tilde{s}_T^{*2}$ are consistent estimators of Λ_1^2, respectively, and m is the number of restrictions under the null (4.8) and (4.9) which is $(n-1)$ here.

5. TESTS OF LINEAR COINTEGRATION

In this section, we derive the null asymptotic distributions of the F-type tests in (4.10)-(4.11) to test linear cointegration against nonlinear cointegration in two ST cointegrating regressions (4.5) and (4.7). First, the following assumptions are required for considering these tests.

Assumption 2. *The time series $\mathbf{y}_t$ presented in (3.1a)-(3.1b) and (3.2a)-(3.2b) are linear $I(1)$. Assume $\mathbf{y}_{2t}$ has the form*

$$\mathbf{y}_{2t} = \mathbf{a}_2 + \mathbf{y}_{2,t-1} + \mathbf{v}_t, \ \ t = 1, 2, ..., T \tag{5.1}$$

in which the initial values of $\mathbf{y}_{2t}$, $\mathbf{y}_{20}$, are equal to zero, $\mathbf{a}_2 = \mathbf{0}$ in (3.1b) and $\mathbf{a}_2 \neq \mathbf{0}$ in (3.2b), and $\mathbf{v}_t$ is a zero-mean vector stationary process allowing for serial correlations.

Assumption 3. *The parameter space for (4.5) and (4.6) is a compact subset of Euclidean space.*

Assumption 4. *Let* $\mathbf{u}_t$ *be an* $(n \times 1)$ *vector* $\mathbf{u}_t = (\tilde{u}_t, \mathbf{v}_t')'$ *such that*

$$\mathbf{u}_t = \sum_{s=0}^{\infty} \mathbf{\Psi}_s \boldsymbol{\varepsilon}_{t-s},$$

in which $\{s \cdot \mathbf{\Psi}_s\}_{s=0}^{*}$ *is absolutely summable; that is,* $\sum_{s=0}^{\infty} s \cdot \left|\psi_{ij}^{(s)}\right| < \infty$ *for each* $i, j =$ *1,2,...,n for the row i, column j element of* $\mathbf{\Psi}_s$, $\psi_{ij}^{(s)}$. $\{\boldsymbol{\varepsilon}_t\}$ *is an i.i.d. sequence with mean zero, variance* $\mathbf{PP}'$*, and finite fourth-order moments.* $\mathbf{\Psi}(1)\mathbf{P}$ *is nonsingular. Assume that* $\tilde{u}_t$ *is uncorrelated with* $\mathbf{v}_\tau$ *for all* t *and* τ*. Then,* $\mathbf{P}$ *and* $\mathbf{\Psi}(L)$ *can have block-diagonal forms,*

$$\underset{(n\times n)}{\mathbf{P}} = \begin{pmatrix} \sigma_1 & \mathbf{0}' \\ \mathbf{0} & \underset{(n-1)\times(n-1)}{\mathbf{P}_{22}} \end{pmatrix} \quad \text{and} \quad \underset{(n\times n)}{\mathbf{\Psi}(L)} = \begin{pmatrix} \Psi_{11}(L) & \mathbf{0}' \\ \mathbf{0} & \underset{(n-1)\times(n-1)}{\mathbf{\Psi}_{22}(L)} \end{pmatrix},$$

such that

$$\underset{(n\times n)}{\mathbf{\Lambda}} = \mathbf{\Psi}(1) \cdot \mathbf{P} = \begin{pmatrix} \Lambda_1 & \mathbf{0}' \\ \mathbf{0} & \underset{(n-1)\times(n-1)}{\mathbf{\Lambda}_2} \end{pmatrix}$$

is also block-diagonal, in which $\Lambda_1 = \sigma_1 \cdot \Psi_{11}(1)$ *and* $\mathbf{\Lambda}_2 = \mathbf{\Psi}_{22}(1) \cdot \mathbf{P}_{22}$.

The assumption that $\tilde{u}_t$ is uncorrelated with $\mathbf{v}_\tau$ for all t and τ in Assumption 4 is not desired for the errors in (4.2) and (4.3) but in (4.5) and (4.6) after the correlations between u_t^* and $\mathbf{y}_{2t}$ are eliminated.

We next derive the asymptotic distributions of the two F-type statistics, $\tilde{F}_T$ and $\tilde{F}_T^*$, formulated in (4.10)-(4.11).

Let an $(n \times 1)$ vector $\mathbf{W}(r) = (W_1(r), W_2(r), ..., W_n(r))'$ be n-dimensional standard Brownian motion and be partitioned as $\mathbf{W}(r) = (W_1(r), \mathbf{W}_2(r))'$, in which $W_i(r)$ are independent processes for $i = 1, 2, ..., n$ and $r \in [0, 1]$.

If the transition variables are specified, then the asymptotic distribution of the F-type statistic $\tilde{F}_T$ in (4.10) can be formulated in terms of standard Brownian motion $W_i(r)$ as in the following Theorem 1.

Theorem 1. *Let $\hat{\boldsymbol{\theta}}_T$ be the OLS estimator of the parameter $\boldsymbol{\theta}$ in (4.5). $\boldsymbol{\theta}_0$ is the true parameter in (4.5) under the null (4.8). Assume that Assumptions 1-4 hold. Assume further that the transition variables in $\mathbf{s}_t$ in (4.5) are stationary processes. We then have the following results:*

(a) Under the null hypothesis (4.8), the asymptotic distribution of the estimator $\hat{\boldsymbol{\theta}}_T$ satisfies

$$\tilde{\boldsymbol{\Upsilon}}_T\left(\hat{\boldsymbol{\theta}}_T-\boldsymbol{\theta}_0\right)\xrightarrow{L}\tilde{\mathbf{Q}}^{-1}\tilde{\mathbf{h}}, \tag{5.2}$$

in which

$$\tilde{\boldsymbol{\Upsilon}}_T=\begin{pmatrix} T^{1/2}\mathbf{I}_{(2p+1)} & \mathbf{0} \\ \mathbf{0}' & T\mathbf{I}_{(2n-2)} \end{pmatrix},\quad \tilde{\mathbf{Q}}=\begin{pmatrix} \tilde{\mathbf{Q}}_{11} & \mathbf{0} & \mathbf{0} \\ \mathbf{0}' & \tilde{\mathbf{Q}}_{22} & \mathbf{0} \\ \mathbf{0}' & \mathbf{0} & \tilde{\mathbf{Q}}_{33} \end{pmatrix},\quad \tilde{\mathbf{h}}=\begin{pmatrix} \tilde{\mathbf{h}}_1 \\ \tilde{\mathbf{h}}_2 \\ \tilde{\mathbf{h}}_3 \end{pmatrix},$$

whereas matrix $\tilde{\mathbf{Q}}$ and vector $\tilde{\mathbf{h}}$ are stochastic processes expressed in terms of standard Brownian motion as follows:

1) $\tilde{\mathbf{Q}}_{11}=E(\mathbf{v}_t^*\mathbf{v}_t^{*'})$;

2) $\tilde{\mathbf{Q}}_{22}=\boldsymbol{\Lambda}_2\left\{\int_0^1[\mathbf{W}_2(r)][\mathbf{W}_2(r)]'dr\right\}\boldsymbol{\Lambda}_2'$;

3) The $(i-1,j-1)$th element of $\tilde{\mathbf{Q}}_{33}$ is $E(s_{it}s_{jt})\{\tilde{\mathbf{Q}}_{22}\}_{i-1,j-1}$, in which $\{\tilde{\mathbf{Q}}_{22}\}_{i-1,j-1}$ is the $(i-1,j-1)$th element of $\tilde{\mathbf{Q}}_{22}$, (i,j=2,3,...,n);

4) $\tilde{\mathbf{h}}_1=N(\mathbf{0},\boldsymbol{\Sigma})$ in which $\boldsymbol{\Sigma}=\sum_{s=-\infty}^{\infty}E(\tilde{u}_t\tilde{u}_{t-s}\mathbf{v}_t^*\mathbf{v}_{t-s}^{*'})$;

5) $\tilde{\mathbf{h}}_2=\Lambda_1\boldsymbol{\Lambda}_2\left\{\int_0^1\mathbf{W}_2(r)\,dW_1(r)\right\}$;

6) The $(i-1)$th element of $\tilde{\mathbf{h}}_3$ is the $(i-1)$th diagonal element of $\boldsymbol{\Lambda}_2\int_0^1\mathbf{W}_2(r)d\mathbf{W}_2^{*'}(r)\boldsymbol{\Lambda}_4^{*'}$, in which $\boldsymbol{\Lambda}_4^*$ are defined in Lemma 2 (b),

whereas Λ_1 and $\boldsymbol{\Lambda}_2$ in 2), 4), 5) above are defined in Assumption 4.

(b) Consider the F-type statistic $\tilde{F}_T$ defined in (4.10). Under Assumptions 1-4, the asymptotic distribution of the $\tilde{F}_T$ statistic has a form

$$\tilde{F}_T\xrightarrow{L}\tilde{\mathbf{h}}_3'\tilde{\mathbf{Q}}_{33}^{'-1}\tilde{\mathbf{h}}_3/(m\Lambda_1^2), \tag{5.3}$$

in which $\tilde{\mathbf{Q}}_{33}$ and $\tilde{\mathbf{h}}_3$ are given in (a) above.

Proof of Theorem 1. See Appendix A. □

Note that the asymptotic distribution of the F-type statistic in (4.10) is expressed in terms of standard Brownian motion. The $\tilde{F}_T$ statistic does not have a standard asymptotic distribution because the vector $\tilde{\mathbf{h}}_3$ contained in vector $\tilde{\mathbf{h}}$ in Theorem 1 are non-Gaussian processes. Taking a simple case as an example, we observe that $\int_0^1 W_i(r)dW_j(r) + \int_0^1 W_j(r)dW_i(r) = W_i(1)W_j(1)$ when $i \neq j$, in which $W_i(1)W_j(1)$ has a so-called "product-normal distribution†" and $\int_0^1 W_i(r)dW_i(r) = \frac{1}{2}\{[W_i(1)]^2 - 1\}$ when $i = j$, which follows a Chi-squared distribution. Obviously, $\int_0^1 W_i(r)dW_j(r)$ indicates a non-Gaussian process. Similarly, in multivariate cases, $\mathbf{\Lambda}^* \int_0^1 \mathbf{W}^*(r)d\mathbf{W}^{*'}(r)\mathbf{\Lambda}^{*'}$ in Lemma 2 (b) follows a combination of Chi-square and product-normal distributions. Then $\mathbf{\Lambda}_2 \int_0^1 \mathbf{W}_2(r)d\mathbf{W}_2^{*'}(r)\mathbf{\Lambda}_4^{*'}$ in Lemma 2 (b) has a combination of product-normal distributions. Therefore, each element in vector $\tilde{\mathbf{h}}_3$, which is the diagonal element of matrix $\mathbf{\Lambda}_2 \int_0^1 \mathbf{W}_2(r)d\mathbf{W}_2^{*'}(r)\mathbf{\Lambda}_4^{*'}$, is a non-Gaussian process. As such, the standard results in Choi and Saikkonen (2004) cannot apply when we consider stationary transition variables rather than $I(1)$ transition variables in the cointegrating regression (4.5).

Now, consider (4.7) and (3.2b). The asymptotic results for the statistic in (4.11) are given in the following Theorem 2.

Theorem 2. *Let $\hat{\boldsymbol{\theta}}_T^*$ be the OLS estimator of the parameter $\boldsymbol{\theta}^*$ in (4.6). $\boldsymbol{\theta}_0^*$ is the true parameter in (4.6) under the null (4.9). When the Assumptions 1-4 hold and the transition variables in $\mathbf{s}_t$ in (4.6) are assumed to be stationary processes, the following results (a) and (b) are established:*

(a) Under the null hypothesis (4.9), the limiting distribution of the estimator $\hat{\boldsymbol{\theta}}_T^*$ satisfies,

$$\tilde{\mathbf{\Upsilon}}_T^* \left(\hat{\boldsymbol{\theta}}_T^* - \boldsymbol{\theta}_0^*\right) \xrightarrow{L} \tilde{\mathbf{Q}}^{*-1}\tilde{\mathbf{h}}^*, \tag{5.4}$$

in which

$$\tilde{\mathbf{\Upsilon}}_T^* = \begin{pmatrix} T^{1/2}\mathbf{I}_{(2p+1)\times(n-1)} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0}' & T\mathbf{I}_{(2n-2)} & \mathbf{0} & \mathbf{0} \\ \mathbf{0}' & \mathbf{0}' & T^{3/2} & \mathbf{0} \\ \mathbf{0}' & \mathbf{0}' & \mathbf{0}' & T^{1/2} \end{pmatrix},$$

†If X_1 and X_2 are two independent standard normal random variables, then the product $Z = X_1X_2$ follows the product-normal distribution with density function $f_Z(z) = \pi^{-1}K_0(|z|)$, in which K_0 is the modified Bessel function of the second kind. See Weisstein (n.d.).

$$\tilde{\mathbf{Q}}^* = \begin{pmatrix} \tilde{\mathbf{Q}}^*_{11} & \tilde{\mathbf{Q}}^*_{12} & \tilde{\mathbf{Q}}^*_{13} & \tilde{\mathbf{Q}}^*_{14} & \tilde{\mathbf{Q}}^*_{15} \\ \tilde{\mathbf{Q}}^*_{21} & \tilde{\mathbf{Q}}^*_{22} & \tilde{\mathbf{Q}}^*_{23} & \tilde{\mathbf{Q}}^*_{24} & \tilde{\mathbf{Q}}^*_{25} \\ \tilde{\mathbf{Q}}^*_{31} & \tilde{\mathbf{Q}}^*_{32} & \tilde{\mathbf{Q}}^*_{33} & \tilde{\mathbf{Q}}^*_{34} & \tilde{\mathbf{Q}}^*_{35} \\ \tilde{\mathbf{Q}}^*_{41} & \tilde{\mathbf{Q}}^*_{42} & \tilde{\mathbf{Q}}^*_{43} & \tilde{Q}^*_{44} & \tilde{Q}^*_{45} \\ \tilde{\mathbf{Q}}^*_{51} & \tilde{\mathbf{Q}}^*_{52} & \tilde{\mathbf{Q}}^*_{53} & \tilde{Q}^*_{54} & \tilde{Q}^*_{55} \end{pmatrix}, \quad \tilde{\mathbf{h}}^* = \begin{pmatrix} \tilde{\mathbf{h}}^*_1 \\ \tilde{\mathbf{h}}^*_2 \\ \tilde{\mathbf{h}}^*_3 \\ \tilde{h}^*_4 \\ \tilde{h}^*_5 \end{pmatrix}$$

whereas matrix $\tilde{\mathbf{Q}}^*$ *and vector* $\tilde{\mathbf{h}}^*$ *are stochastic processes expressed in terms of standard Brownian motion as follows:*

1) $\tilde{\mathbf{Q}}^*_{11} = E(\mathbf{v}^*_t \mathbf{v}^{*'}_t)$; $\tilde{\mathbf{Q}}^*_{22} = \tilde{\mathbf{Q}}_{22}$;

The $(i-1, j-1)$*th element of* $\tilde{\mathbf{Q}}^*_{33}$ *is* $E(s_{it}s_{jt})\{\tilde{\mathbf{Q}}_{22}\}_{i-1,j-1}$ *(i,j=2,3,...,n);*

3) $\tilde{\mathbf{Q}}^*_{14} = E(\mathbf{v}^*_t/2)$; $\tilde{\mathbf{Q}}^*_{41} = \tilde{\mathbf{Q}}^{*'}_{14}$; $\tilde{\mathbf{Q}}^*_{15} = E(\mathbf{v}^*_t)$; $\tilde{\mathbf{Q}}^*_{51} = \tilde{\mathbf{Q}}^{*'}_{15}$;

4) $\tilde{\mathbf{Q}}^*_{21} = \left(\mathbf{\Lambda}_2 \int_0^1 \mathbf{W}_2(r)dr\right) E(\mathbf{v}^{*'}_t)$; $\tilde{\mathbf{Q}}^*_{12} = \tilde{\mathbf{Q}}^{*'}_{21}$;

5) $\tilde{\mathbf{Q}}^*_{24} = \mathbf{\Lambda}_2 \cdot \int_0^1 r\mathbf{W}_2(r)dr$; $\tilde{\mathbf{Q}}^*_{42} = \tilde{\mathbf{Q}}^{*'}_{24}$;

6) $\tilde{\mathbf{Q}}^*_{25} = \mathbf{\Lambda}_2 \cdot \int_0^1 \mathbf{W}_2(r)dr$; $\tilde{\mathbf{Q}}^*_{52} = \tilde{\mathbf{Q}}^{*'}_{25}$;

7) $\tilde{\mathbf{Q}}^*_{31} = \left(E(v^*_{1t}\mathbf{s}_t)\mathbf{\Lambda}_2 \int_0^1 \mathbf{W}_2(r)dr, ..., E(v^*_{(2p+1)(n-1),t}\mathbf{s}_t)\mathbf{\Lambda}_2 \int_0^1 \mathbf{W}_2(r)dr\right)$

and $\tilde{\mathbf{Q}}^*_{13} = \tilde{\mathbf{Q}}^{*'}_{31}$;

8) $\tilde{\mathbf{Q}}^*_{32} = E(\mathbf{s}_t)\mathbf{\Lambda}_2 \int_0^1 \mathbf{W}_2(r)\mathbf{W}'_2(r)dr\mathbf{\Lambda}'_2$ *and* $\tilde{\mathbf{Q}}^*_{23} = \tilde{\mathbf{Q}}^{*'}_{32}$;

9) $\tilde{\mathbf{Q}}^*_{34} = E(\mathbf{s}_t)\mathbf{\Lambda}_2 \int_0^1 r\mathbf{W}_2(r)dr$; $\tilde{\mathbf{Q}}^*_{43} = \tilde{\mathbf{Q}}^{*'}_{34}$;

10) $\tilde{\mathbf{Q}}^*_{35} = E(\mathbf{s}_t)\mathbf{\Lambda}_2 \int_0^1 \mathbf{W}_2(r)dr$; $\tilde{\mathbf{Q}}^*_{53} = \tilde{\mathbf{Q}}^{*'}_{35}$;

11) $\tilde{Q}^*_{44} = 1/3$; $\tilde{Q}^*_{55} = 1$; $\tilde{Q}^*_{45} = \tilde{Q}^*_{54} = 1/2$;

12) $\tilde{\mathbf{h}}_1 = N(\mathbf{0}, \mathbf{\Sigma})$ *in which* $\mathbf{\Sigma} = \sum_{s=-\infty}^{\infty} E(\tilde{u}_t\tilde{u}_{t-s}\mathbf{v}^*_t\mathbf{v}^{*'}_{t-s})$; $\tilde{\mathbf{h}}^*_2 = \tilde{\mathbf{h}}_2$; *The* $(i-1)$*th element of* $\tilde{\mathbf{h}}_3$ *is the* $(i-1)$*th diagonal element of* $\mathbf{\Lambda}_2 \int_0^1 \mathbf{W}_2(r)d\mathbf{W}^{*'}_2(r)\mathbf{\Lambda}^{*'}_4$, *in which* $\mathbf{\Lambda}^*_4$ *are defined in Lemma 2 (b);*

14) $\tilde{h}^*_4 = \Lambda_1 \cdot \left\{W_1(1) - \int_0^1 W_1(r)dr\right\}$;

15) $\tilde{h}^*_5 = \Lambda_1 W_1(1)$.

(b) Consider the F-type statistic $\tilde{F}_T^*$ defined in (4.11). Under Assumptions 1-4, the asymptotic distribution of the $\tilde{F}_T^*$ statistic has a form,

$$\tilde{F}_T^* \xrightarrow{L} \left(\tilde{\mathbf{R}}^*\tilde{\mathbf{Q}}^{*-1}\tilde{\mathbf{h}}^*\right)' \left\{\Lambda_1^2 \cdot \tilde{\mathbf{R}}^*\tilde{\mathbf{Q}}^{*-1}\tilde{\mathbf{R}}^* m\right\}^{-1} \left(\tilde{\mathbf{R}}^*\tilde{\mathbf{Q}}^{*-1}\tilde{\mathbf{h}}^*\right), \tag{5.5}$$

in which $\tilde{\mathbf{Q}}^*$ and $\tilde{\mathbf{h}}^*$ are given in (a) above.

Proof of Theorem 2. See Appendix A. □

The nuisance parameters in (5.3) and (5.5) need to be estimated in practice. For example, from samples simulated from DGP_1 in Section 6, $E(s_{it}s_{jt})$ in $\tilde{\mathbf{Q}}_{33}$ can be estimated by the sample mean $\sum_{t=1}^T s_{it}s_{jt}/T$, and $\mathbf{\Lambda}_2$ in $\tilde{\mathbf{Q}}_{33}$ can be estimated as the Cholesky factorization of $\sum_{s=-1}^1 \left(\sum_{t=1}^T (\Delta\mathbf{y}_{2t} - \hat{\mathbf{a}}_2)(\Delta\mathbf{y}_{2t-s} - \hat{\mathbf{a}}_2))'/T\right)$. Similarly, we can estimate other nuisance parameters.

6. SIMULATION STUDY

This section exmines the finite-sample distributions and small-sample properties of the F-type statistics in (4.10) and (4.11) using Monte Carlo experiments. All simulations are designed for $\mathbf{y}_t = (y_{1t}, y_{2t}, y_{3t})'$ and $n = 3$. Two options for the transition variable, $s_{it} = \Delta y_{i,t-1}$ and $s_{it} = z_{it}$, are considered in each experiment. A total of 100,000 replications were performed for each experiment.

6.1 Critical values of the F-type test

To assess the distribution approximation of the F-type statistic in (4.10), the data-generating process (DGP_1) is designed under the null hypothesis (4.8). The estimation is achieved from (4.5). Then, the F-type statistic is derived from (4.10) for the finite sample sizes $T = 50, 100, 200, 500$, starting from DGP_1. The distribution of F is directly calculated from (5.3) for the large sample size $T = 1000$ by approximating the Brownian motion functions described in DGP_{BM} below.

(i) DGP_1: Considering (4.5) under the null (4.8), DPG_1 is given by

$$y_{1t} = 0.4y_{2t} + 0.6y_{3t} + 0.18\Delta y_{2,t-1} + 0.21\Delta y_{3,t-1} + \tilde{u}_t$$
$$y_{2t} = y_{2,t-1} + v_{1t}$$
$$y_{3t} = y_{3,t-1} + v_{2t},$$

$t = 1, 2, ..., T$, in which the error term $\mathbf{u}_t = (\tilde{u}_t, v_{1t}, v_{2t})'$ satisfies

$$\mathbf{u}_t = \boldsymbol{\varepsilon}_t + \mathbf{B}\boldsymbol{\varepsilon}_{\mathbf{t-1}},$$

where the $(n \times 1)$ vector $\boldsymbol{\varepsilon}_t$ is an $n.i.d$ process with zero mean and the covariance matrix

$$\begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0.5 \\ 0 & 0.5 & 1 \end{pmatrix},$$

and $\mathbf{B}$ is a block-diagonal matrix

$$\begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & \tilde{\omega} \\ 0 & \tilde{\omega} & 1 \end{pmatrix}$$

with $\tilde{\omega} = 0.2$ and 0.8 in our simulations.

To assess the distribution approximation of the F-type statistic in (4.11), DGP_2 is designed under the null hypothesis (4.9). The estimation is achieved from (4.6). Then, the F-type statistic is calculated as (4.11) for the finite sample sizes $T = 50, 100, 200, 500$, starting from DGP_2. The distribution of F is approximated based on model (4.7) for the large sample size $T = 1000$.

(ii) DGP_2: Considering model (4.6) under the null of (4.9), DPG_2 is given by

$$y_{1t} = 0.3 + 0.26t + 0.4y_{2t} + 0.6y_{3t} + 0.18\Delta y_{2,t-1} + 0.21\Delta y_{3,t-1} + \tilde{u}_t$$
$$y_{2t} = 0.11 + y_{2,t-1} + v_{1t}$$
$$y_{3t} = 0.19 + y_{3,t-1} + v_{2t}$$

$t = 1, 2, ..., T$, where $(\tilde{u}_t, v_{1t}, v_{2t})'$ are assumed the same as in DGP_1.

Table 1: Critical values for the F-type test under the null (4.8), based on (4.5) with $s_{it} = \Delta y_{i,t-1}$

			Probability that F-type test is greater than entry							
s_{it}	$\tilde{\omega}$	T	0.99	0.975	0.95	0.90	0.10	0.05	0.025	0.01
$\Delta y_{i,t-1}$	0.2	50	0.01	0.04	0.08	0.16	3.71	4.96	6.23	7.95
	0.2	100	0.02	0.04	0.08	0.16	3.56	4.67	5.85	7.34
	0.2	200	0.01	0.04	0.08	0.16	3.49	4.57	5.67	7.15
	0.2	500	0.01	0.04	0.08	0.16	3.44	4.49	5.53	6.91
	0.2	1000	0.01	0.04	0.07	0.15	3.44	4.48	5.52	6.88
	0.8	50	0.01	0.03	0.07	0.14	3.34	4.43	5.57	7.11
	0.8	100	0.01	0.03	0.07	0.14	3.20	4.18	5.20	6.54
	0.8	200	0.01	0.03	0.07	0.14	3.16	4.14	5.12	6.41
	0.8	500	0.01	0.03	0.07	0.14	3.13	4.11	5.06	6.29
	0.8	1000	0.01	0.03	0.07	0.14	3.09	4.03	4.98	6.20

The probability shown at the head of the column is the area in the right-hand tail.

Table 2: Critical values for the F-type test under the null (4.8), based on (4.5) with $s_{it} = z_{it}$

			Probability that F-type test is greater than entry							
s_{it}	$\tilde{\omega}$	T	0.99	0.975	0.95	0.90	0.10	0.05	0.025	0.01
z_{it}	0.2	50	0.01	0.03	0.05	0.10	2.43	3.21	4.03	5.15
	0.2	100	0.01	0.02	0.05	0.10	2.33	3.09	3.81	4.80
	0.2	200	0.01	0.02	0.05	0.10	2.32	3.05	3.78	4.71
	0.2	500	0.01	0.03	0.05	0.11	2.31	3.01	3.71	4.63
	0.2	1000	0.01	0.02	0.05	0.10	2.32	3.01	3.70	4.59
	0.8	50	0.01	0.03	0.05	0.11	2.42	3.21	4.06	5.20
	0.8	100	0.01	0.02	0.05	0.10	2.35	3.10	3.87	4.92
	0.8	200	0.01	0.03	0.05	0.11	2.33	3.05	3.79	4.76
	0.8	500	0.01	0.02	0.05	0.10	2.32	3.01	3.74	4.68
	0.8	1000	0.01	0.02	0.05	0.10	2.29	2.97	3.68	4.61

The probability shown at the head of the column is the area in the right-hand tail.

Table 3: Critical values for the F-type test under the null (4.9), based on auxiliary model (4.6) with $s_{it} = \Delta y_{i,t-1}$

			Probability that F type test is greater than entry							
s_{it}	$\tilde{\omega}$	T	0.99	0.975	0.95	0.90	0.10	0.05	0.025	0.01
$\Delta y_{i,t-1}$	0.2	50	0.01	0.02	0.05	0.10	2.44	3.33	4.29	5.71
	0.2	100	0.01	0.02	0.04	0.09	1.99	2.56	3.32	4.23
	0.2	200	0.01	0.02	0.04	0.08	1.84	2.40	2.98	3.74
	0.2	500	0.01	0.02	0.04	0.08	1.76	2.29	2.83	3.57
	0.2	1000	0.01	0.02	0.04	0.08	1.74	2.28	2.81	3.50
	0.8	50	0.01	0.02	0.04	0.08	2.14	2.93	3.80	4.98
	0.8	100	0.01	0.02	0.04	0.08	1.79	2.38	2.98	3.82
	0.8	200	0.01	0.02	0.04	0.07	1.66	2.18	2.72	3.46
	0.8	500	0.01	0.02	0.04	0.07	1.60	2.08	2.57	3.21
	0.8	1000	0.01	0.02	0.03	0.07	1.57	2.04	2.54	3.18

The probability shown at the head of the column is the area in the right-hand tail.

Table 4: Critical values for the F-type test under the null (4.9), based on auxiliary model (4.6) with $s_{it} = z_{it}$

			Probability that F type test is greater than entry							
s_{it}	$\tilde{\omega}$	T	0.99	0.975	0.95	0.90	0.10	0.05	0.025	0.01
z_{it}	0.2	50	0.01	0.01	0.03	0.06	1.58	2.15	2.81	3.73
	0.2	100	0.01	0.01	0.03	0.06	1.32	1.75	2.18	2.82
	0.2	200	0.01	0.01	0.03	0.05	1.23	1.62	2.02	2.55
	0.2	500	0.01	0.01	0.03	0.05	1.19	1.55	1.91	2.39
	0.2	1000	0.00	0.01	0.03	0.05	1.16	1.50	1.84	2.30
	0.8	50	0.01	0.01	0.03	0.06	1.56	2.15	2.76	3.67
	0.8	100	0.01	0.01	0.03	0.06	1.31	1.74	2.19	2.78
	0.8	200	0.01	0.01	0.03	0.05	1.23	1.62	2.02	2.56
	0.8	500	0.01	0.01	0.03	0.05	1.18	1.54	1.91	2.41
	0.8	1000	0.00	0.01	0.03	0.05	1.16	1.51	1.88	2.35

The probability shown at the head of the column is the area in the right-hand tail.

(iii) DGP_{BM}: To calculate quantiles of the asymptotic distribution of the F-type statistic in (5.3) and (5.5), we first approximate the Brownian motion functions by simulating the partial sums of the normal random variables. For example, $\int_0^1 \mathbf{W}_2(r)d\mathbf{W}_2^{*\prime}(r)$ in $\tilde{\mathbf{h}}_3$ is approximated as

$$T^{-1}\sum_{t=1}^{T}\mathbf{x}_{t-1}\mathbf{e}_t',$$

in which $\mathbf{x}_t = \sum_{j=1}^{t}\mathbf{e}_j^*$ and both $\mathbf{e}_t$ and $\mathbf{e}_t^*$ are independent, standard normal random variables with $E(\mathbf{e}_t\mathbf{e}_t') = E(\mathbf{e}_t^*\mathbf{e}_t^{*\prime}) = \mathbf{I}$. Furthermore, $\int_0^1 [\mathbf{W}_2(r)][\mathbf{W}_2(r)]'dr$ in $\tilde{\mathbf{Q}}_{33}$ can be approximated as

$$T^{-2}\sum_{t=1}^{T}\mathbf{x}_{t-1}\mathbf{x}_{t-1}'.$$

Tables 1-4 report the distribution approximations of the F-type statistic $\tilde{F}_T$ in (4.10) and (4.11) considering the two transition variables $s_{it} = \Delta y_{i,t-1}$ and $s_{it} = z_{it}$ for each case, where z_{it} are stationary exogenous variables.

6.2 A size study

In this subsection, the considered sample sizes in our small-sample experiments are $T = 50, 100, 200$. To examine the size properties of the F-type statistics, we generate data under the null following DPG_1 and DGP_2 with $\tilde{\omega} = 0.2$ given in subsection 6.1.

The critical values used are the asymptotic version at sample size 1000 in Tables 1-4. The empirical size estimates at the significance levels $10\%, 5\%$ and 1% are reported in Table 5 for $\tilde{F}_T$ and reported in Table 6 for $\tilde{F}_T^*$.

The results in Tables 5-6 illustrate that there is not much of a size distortion, although the tests are oversized when the sample size is small. The effect of size distortion decreases when the sample size increases. When the sample size is greater than 200, the approximation to the asymptotic distribution appears reasonable.

Table 5: Empirical size estimates of the F-type test (4.10) at given significance levels based on (4.5) with $s_{it} = \Delta y_{i,t-1}$ and $s_{it} = z_{it}$

		given significance levels					given significance levels		
s_{it}	T	10%	5%	1%	s_{it}	T	10%	5%	1%
$\Delta y_{i,t-1}$	50	0.118	0.065	0.017	z_{it}	50	0.110	0.060	0.016
	100	0.108	0.057	0.013		100	0.103	0.054	0.013
	200	0.105	0.055	0.011		200	0.100	0.050	0.011

Table 6: Empirical size estimates of the F-type test (4.11) given significance levels based on (4.6) with $s_{it} = \Delta y_{i,t-1}$ and $s_{it} = z_{it}$

		given significance levels					given significance levels		
s_{it}	T	10%	5%	1%	s_{it}	T	10%	5%	1%
$\Delta y_{i,t-1}$	50	0.180	0.114	0.044	z_{it}	50	0.172	0.111	0.043
	100	0.132	0.073	0.021		100	0.128	0.074	0.021
	200	0.112	0.057	0.013		200	0.113	0.062	0.015
	500	0.102	0.050	0.011		500	0.106	0.055	0.012

Table 7: Empirical power estimates of the F-type test at given significance levels based on (4.5) with $s_{it} = \Delta y_{i,t-1}$ and $s_{it} = z_{it}$

Model (4.5) with $s_{it} = \Delta y_{i,t-1}$					Model (4.5) with $s_{it} = z_{it}$				
		given significance levels					given significance levels		
γ	T	10%	5%	1%	γ	T	10%	5%	1%
0.01	50	0.103	0.051	0.010	0.01	50	0.104	0.053	0.011
0.01	100	0.107	0.055	0.011	0.01	100	0.114	0.059	0.013
0.01	200	0.128	0.069	0.016	0.01	200	0.161	0.093	0.025
0.1	50	0.245	0.155	0.054	0.1	50	0.401	0.296	0.141
0.1	100	0.585	0.478	0.284	0.1	100	0.778	0.703	0.547
0.1	200	0.946	0.914	0.819	0.1	200	0.980	0.968	0.928
0.5	50	0.938	0.902	0.800	0.5	50	0.985	0.974	0.939
0.5	100	1	0.999	0.998	0.5	100	1	1	1
0.5	200	1	1	1	0.5	200	1	1	1
1	50	0.992	0.985	0.956	1	50	1	0.999	0.996
1	100	1	1	1	1	100	1	1	1
1	200	1	1	1	1	200	1	1	1
5	50	0.956	0.927	0.838	5	50	1	1	1
5	100	0.999	0.997	0.990	5	100	1	1	1
5	200	1	1	1	5	200	1	1	1
10	50	0.930	0.887	0.764	10	50	1	1	0.999
10	100	0.997	0.994	0.979	10	100	1	1	1
10	200	1	1	1	10	200	1	1	1
50	50	0.902	0.844	0.688	50	50	1	1	0.999
50	100	0.994	0.987	0.960	50	100	1	1	1
50	200	1	1	0.999	50	200	1	1	1

Table 8: Empirical power estimates of the F-type test at given significance levels based on (4.6) with $s_{it} = \Delta y_{i,t-1}$ and $s_{it} = z_{it}$

Model (4.6) with $s_{it} = \Delta y_{i,t-1}$					Model (4.6) with $s_{it} = z_{it}$				
		given significance levels					given significance levels		
γ	T	10%	5%	1%	γ	T	10%	5%	1%
0.01	50	0.182	0.116	0.046	0.01	50	0.177	0.114	0.044
0.01	100	0.145	0.081	0.025	0.01	100	0.154	0.092	0.029
0.01	200	0.187	0.112	0.037	0.01	200	0.236	0.156	0.061
0.1	50	0.421	0.327	0.192	0.1	50	0.524	0.438	0.293
0.1	100	0.767	0.694	0.547	0.1	100	0.843	0.790	0.678
0.1	200	0.985	0.974	0.943	0.1	200	0.989	0.983	0.963
0.5	50	0.970	0.954	0.911	0.5	50	0.989	0.982	0.962
0.5	100	1	1	0.999	0.5	100	1	1	1
0.5	200	1	1	1	0.5	200	1	1	1
1	50	0.996	0.993	0.983	1	50	0.999	0.999	0.997
1	100	1	1	1	1	100	1	1	1
1	200	1	1	1	1	200	1	1	1
5	50	0.984	0.977	0.959	5	50	0.998	0.998	0.998
5	100	0.999	0.999	0.998	5	100	1	1	1
5	200	1	1	1	5	200	1	1	1
10	50	0.978	0.968	0.943	10	50	0.997	0.997	0.997
10	100	0.999	0.999	0.997	10	100	1	1	1
10	200	1	1	1	10	200	1	1	1
50	50	0.971	0.958	0.923	50	50	0.996	0.996	0.996
50	100	0.999	0.998	0.995	50	100	1	1	1
50	200	1	1	1	50	200	1	1	1

6.3 A power study

For the power studies of the F-type statistics in (4.10) and (4.11), we generate data under the alternative by the system

$$\begin{aligned} y_{1t} =& \frac{0.8}{1 + exp(-\gamma(s_{2t} - 0.5))} y_{2t} + \frac{1.2}{1 + exp(-\gamma(s_{3t} - 0.5))} y_{3t} \\ &+ 0.18 \Delta y_{2,t-1} + 0.21 \Delta y_{3,t-1} + u_t \\ y_{2t} =& y_{2,t-1} + v_{1t} \\ y_{3t} =& y_{3,t-1} + v_{2t} \end{aligned}$$

for the first case, and by

$$\begin{aligned} y_{1t} =& 0.3 + 0.26t + \frac{0.8}{1 + exp(-\gamma(s_{2t} - 0.5))} y_{2t} + \frac{1.2}{1 + exp(-\gamma(s_{3t} - 0.5))} y_{3t} \\ &+ 0.18 \Delta y_{2,t-1} + 0.21 \Delta y_{3,t-1} + u_t \\ y_{2t} =& 0.11 + y_{2,t-1} + v_{1t} \\ y_{3t} =& 0.19 + y_{3,t-1} + v_{2t} \end{aligned}$$

for the second case, in which $(u_t, v_{1t}, v_{2t})'$ are generated as in DPG_1 with $\tilde{\omega} = 0.2$ and values of $\gamma = 0.01$, 0.1, 0.5, 1, 5, 10, or 50 for the comparisons. Simulation results for the power studies are given in Tables 7-8. Our tests demonstrate good power. We observed that the power of our tests increases reasonably as γ increases, however, when γ is greater than 5, the power decreases slightly. This observation is not surprising because it indicates another case known as threshold cointegration when γ is sufficient large (see the properties of the function $G(s_t; \gamma, c)$ in Section 2). When γ is very small, our tests have low power because the model is nearly linear when γ is close to zero.

7. EMPIRICAL APPLICATION OF PPP SYSTEM

PPP refers to the power purchasing parity theory in economics, which was developed by Cassel (1918). PPP data are often used to study cointegration relation. Figure 3 presents normalized monthly PPP data from 1973:1 to 1989:10 (202 observations) of the price level of United States (solid line) and the price level of Italy (dashed line) along with the dollar-lira exchange rate (dot-dashed line), denoted by y_{1t}, y_{2t} and y_{3t}. Natural

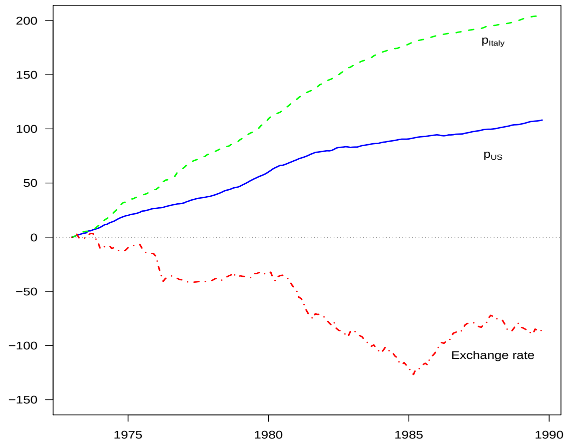

Figure 3: PPP data: the price level of United States (solid line, blue) and Italy (dashed line, green) and the dollar-lira exchange rate (dot-dashed line, red)

logs of the raw data were taken, multiplied by 100, and then normalized by subtracting the initial value for 1973:1 for ease of reading.

7.1 Testing for unit roots

According to our definition of ST cointegration, we first examine whether each of the series in the system is individually $I(1)$. Notice that if nonlinear structures exist in the time series, the classical unit root tests may be invalid because of incorrect non-rejection of a unit root by nonlinearity. This question was first posed by Perron (1990) and studied by several others, e.g., He and Sandberg (2005). Therefore, we set up the null hypothesis of a unit root against the alternative of a stable process in a nonlinear framework.

Consider a nonlinear model under the alternative hypothesis

$$y_t = \theta_1 + \theta_2 y_{t-1} + (\theta_3 + \theta_4 y_{t-1})G(s_t : \gamma, c) + \theta_5 t + u_t, \tag{7.1}$$

with an auxiliary regression given in (7.2) for the parameter identification problem with restriction $\gamma = 0$ under the null

$$y_t = \theta_1^* + \theta_2^* y_{t-1} + \theta_3^* s_t y_{t-1} + \theta_4^* s_t + \theta_5^* t + u_t^*. \tag{7.2}$$

However, when considering $s_t = y_{t-1}$, the alternative model (7.2) should be updated to

$$y_t = \theta_1^* + \tilde{\theta}_2^* y_{t-1} + \theta_3^* s_t y_{t-1} + \theta_5^* t + u_t^*. \tag{7.3}$$

The test for unit roots here is derived by imposing the null hypothesis $H_o : \theta_2^* = 1$ and the following joint null hypothesis:

$$H_o : \theta_2^* = 1, \theta_3^* = 0, \theta_4^* = 0, \theta_5^* = 0 \tag{7.4}$$

in (7.2), or by imposing the null $H_o : \tilde{\theta}_2^* = 1$ and the joint null hypothesis

$$H_o : \tilde{\theta}_2^* = 1, \theta_3^* = 0, \theta_5^* = 0 \tag{7.5}$$

in (7.3).

We calculate the Dickey-Fuller-type t statistic for testing the null of $\theta_2^* = 1$ or $\tilde{\theta}_2^* = 1$ and the F statistic for the joint null (7.4) or (7.5). We include twelve lags $\sum_{p=1}^{12} \delta_p \Delta y_{t-p}$ in the regression (7.2) for estimation to correct serial correlations. The critical values of those two statistics are not standard and are obtained by simulations that are summarized in Tables 11-14 in Appendix B.

Data y_{it} cannot reject the null of a unit root in either the linear or nonlinear framework, and first-order differenced data Δy_{it} reject the null of a unit root, as displayed in Table 9. Therefore, it is possible to state that each time series is a unit root process.

7.2 Cointegration Analysis

The study of linear cointegration for the same data set has been performed by Hamilton (1994, Chapter 19). Given the cointegrating vector $(1, -1, -1)'$, the augmented Dickey-Fuller (ADF) t test is -2.04, which is larger than the 5% critical value of -2.88. The ADF F test is 2.19, which is less than the 5% critical value of 4.66. Alternatively, when considering an estimated cointegrating vector, the Phillips-Ouliaris Z_ρ test of whether $\hat{u}_t$ is stationary is -7.54, which is larger than the 5% critical value of -27.1. By the Phillips-Perron Z_ρ and Z_t tests, Hamilton (1994, Chapter 19) concludes that there is no linear cointegration relation in the system of the PPP data.

Following the testing procedure in Section 4, the results of testing linear cointegration

Table 9: The unit root tests.

			DF statistics		critical values	
framework	time series	s_t	t	F	t	F
linear	y_{1t}	-	-1.9547	2.4129	-3.43	6.34
	y_{2t}	-	-0.1320	4.2500	-3.43	6.34
	y_{3t}	-	-1.5844	1.4897	-3.43	6.34
nonlinear	y_{1t}	Δy_{1t-1}	-1.8059	1.6189	-3.43	4.77
		y_{1t-1}	-0.4415	3.5068	-3.73	5.64
	y_{2t}	Δy_{2t-1}	-0.4336	2.9254	-3.43	4.77
		y_{2t-1}	0.2838	3.3264	-3.73	5.64
	y_{3t}	Δy_{3t-1}	-1.5082	1.4250	-3.43	4.77
		y_{3t-1}	0.2281	1.2047	-3.73	5.64
linear	Δy_{1t}	-	-4.9350	12.179	-3.43	6.34
	Δy_{2t}	-	-4.5360	10.508	-3.43	6.34
	Δy_{3t}	-	-8.0722	32.613	-3.43	6.34

Table 10: $m\tilde{F}_T^*$ tests

s_{2t}	$\Delta y_{2,t-1}$	$\Delta y_{3,t-1}$	$\Delta y_{2,t-1}$	$\Delta y_{3,t-1}$
s_{3t}	$\Delta y_{3,t-1}$	$\Delta y_{2,t-1}$	$\Delta y_{2,t-1}$	$\Delta y_{3,t-1}$
$m\tilde{F}_T^*$	12.12	5.25	**16.93**	2.04
s_{2t}	$\Delta y_{2,t-1}$	$\Delta y_{3,t-1}$	$\Delta y_{1,t-1}$	$\Delta y_{1,t-1}$
s_{3t}	$\Delta y_{1,t-1}$	$\Delta y_{1,t-1}$	$\Delta y_{2,t-1}$	$\Delta y_{3,t-1}$
$m\tilde{F}_T^*$	16.04	6.71	9.88	6.55

*Critical value at 5% significance level is 2.40.

against ST cointegration are presented in Table 10. The null of linear cointegration is significantly rejected, and we observe ST cointegration in the PPP system. The test of nonlinearity is also used to specify the model whereby the one that presents the most nonlinear relation in the system is chosen. Table 10 gives an example. Among the choices in Table 10, the transition variables $s_{2t} = \Delta y_{2,t-1}$ and $s_{3t} = \Delta y_{2,t-1}$ are most optimal because they contribute to the most significant rejection of the null. An estimated cointegrating vector is suggested as follows, and the linear and nonlinear combination of the PPP system is depicted in Figure 4.

$$\begin{aligned}
\hat{y}_{1t} = &- \underset{(0.330)}{1.07} - \underset{(0.011)}{0.05}\, t + \underset{(0.013)}{0.59}\, y_{2t} + \underset{(0.009)}{0.06}\, y_{3t} \\
&+ \underset{(0.008)}{0.05}\, y_{2t} \left(1 + \exp\{- \underset{(0.927)}{3.88}\, (\Delta y_{2,t-1} - \underset{(0.051)}{0.84})/\hat{\sigma}_{\Delta y_{2,t-1}}\}\right)^{-1} \\
&+ \underset{(0.012)}{0.06}\, y_{3t} \left(1 + \exp\{- \underset{(1.714)}{5.69}\, (\Delta y_{2,t-1} - \underset{(0.052)}{0.83})/\hat{\sigma}_{\Delta y_{2,t-1}}\}\right)^{-1} \\
&+ \underset{(0.114)}{0.48}\, \Delta y_{2,t+1} + \underset{(0.116)}{0.43}\, \Delta y_{2,t+2} + \underset{(0.113)}{0.41}\, \Delta y_{2,t+3} + \underset{(0.108)}{0.51}\, \Delta y_{2,t+4} \\
&- \underset{(0.152)}{0.79}\, \Delta y_{2,t-1} - \underset{(0.023)}{0.07}\, \Delta y_{3,t+1} - \underset{(0.024)}{0.16}\, \Delta y_{3t} - \underset{(0.024)}{0.14}\, \Delta y_{3,t-1} \\
&- \underset{(0.025)}{0.16}\, \Delta y_{3,t-2} - \underset{(0.024)}{0.14}\, \Delta y_{3,t-3} - \underset{(0.024)}{0.13}\, \Delta y_{3,t-4},
\end{aligned} \tag{7.6}$$

in which $\hat{\sigma}_{\Delta y_{2,t-1}} = 0.643$.

8. CONCLUDING REMARKS

This study proposes a definition of nonlinear cointegration in the sense that nonstationary variables are tied within a nonlinear combination. The definition is given when the nonlinear combination is restricted as a ST combination for ease of illustrating the models and the testing procedure in this study, but it is straightforward to apply our approach to other combinations or extend it to a general form. Our definition further nests two popular schemes of primarily developed cointegration, Engle and Granger's linear cointegration and Balke and Fomby's threshold cointegration, as two special cases. The proposed two cointegrating regression models cover both situations when the first-order differences of individual series have either zero means or nonzero means. Current models of nonlinear cointegration are extended by considering stationary (instead of nonstationary) transition variables. A main result and important message of this paper is that this

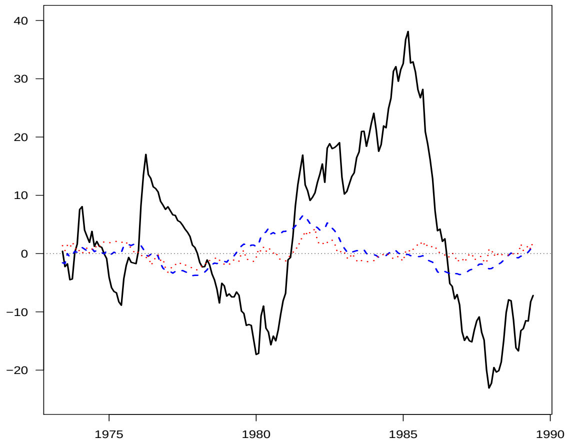

Figure 4: Linear combination of the PPP system, with given cointegrating vector $(1, -1, -1)'$ (solid, black) and with estimated cointegrating vector (dashed, blue); Nonlinear combination (dotted, red).

leads to a nonstandard testing situation and previously derived tests for the nonstationary case are not valid with stationary transition variables. Our F-type tests perform quite well when the sample size is small. They have a higher power when the slope parameter gamma is large (more nonlinearity in the true model) than when gamma is small (less nonlinearity in the true model). However, when gamma is large enough, the power slightly decreases because the true model represents a threshold cointegration.

In the application to monthly PPP data from 1973:1 to 1989:10 (202 observations) of the price levels of United States and Italy along with the dollar-lira exchange rate, linear cointegration relation is not likely to exist. This outcome can be puzzling, because apart from transportation costs, the lack of the relative price having long-run equilibrium contradicts the economic PPP theory. However, the observed nonlinear cointegration relation explains the reason. The PPP theory just expresses in a more complicated way that individual components are tied together represented by nonlinear combinations rather than linear combinations. This data has been studied for cointegration analysis in Hamilton (1994), which corroborated that linear cointegration has not been found. Applying our procedure further, a ST cointegration is then observed. This is quite reasonable and truely exciting. However, the increased difficulty with correctly specifying, estimating and interpreting the richer model remains an issue.

Acknowledgments

Li is grateful to Sune Karlsson, Kenneth Carling and Johan Lyhagen for invaluable advice. Li also thanks the Riksbankens Jubileumsfond Foundation (Grant No. P2005-1247) and School of Technology and Business Studies at Högskolan Dalarna for partial financial support. Special thanks to James D. Hamilton for his help for my understanding the detail of his book.

References

Baillie, R. T. and Selover, D. D. (1987). Cointegration and models of exchange rate determination, *International Journal of Forecasting* **3**: 43–51.

Balke, N. S. and Fomby, T. B. (1997). Threshold cointegration, *International Economic Review* **38**: 627–645.

Cassel, G. (1918). Abnormal deviations in international exchanges, *The Economic Journal* **28**(112): 413–415.

Chang, Y., Park, J. Y. and Phillips, P. C. B. (2001). Nonlinear econometric models with cointegrated and deterministically trending regressors, *Econometrics Journal* **4**: 1–36.

Choi, I. and Saikkonen, P. (2004). Testing linearity in cointegrating smooth transition regressions, *Econometrics Journal* **7**: 341–365.

Clarida, R. H. (1994). Cointegration, aggregate consumption, and the demand for imports: A structural econometric investigation, *The American Economic Review* **84**(1): 298–308.

Corbae, D. and Ouliaris, S. (1988). Cointegration and tests of purchasing power parity, *Review of Economics and Statistics* **70**: 508–11.

Dickey, D. A. and Fuller, W. A. (1979). Distribution of the estimators for autoregressive time series with a unit root, *Journal of the American Statistical Association* **74**: 427–431.

Engle, R. F. and Granger, C. W. J. (1987). Co-integration and error correction: Representation, estimation and testing, *Econometrica* **55**(2): 251–276.

Granger, C. W. J. (1995). Modelling nonlinear relationships between extended-memory variables, *Econometrica* **63**(2): 265–279.

Granger, C. W. J. and Teräsvirta, T. (1993). *Modelling Nonlinear Economic Relationships*, Oxford University Press, New York.

Hamilton, J. D. (1994). *Time Series Analysis*, Princeton University Press, Princeton, NJ.

Hansen, B. E. (1992a). Efficient estimation and testing of cointegrating vectors in the presence of deterministic trends, *Journal of Econometrics* **53**: 87–121.

Hansen, B. E. (1992b). Convergence to stochastic integrals for dependent heterogeneous processes, *Econometric Theory* **8**(4): 489–500.

Hansen, B. E. and Seo, B. (2002). Testing for two-regime threshold cointegration in vector error correction models, *Journal of Econometrics* **110**: 293–318.

Hayashi, F. (2000). *Econometrics*, Princeton University Press, Princeton, NJ.

He, C. and Sandberg, R. (2005). Testing for unit roots in nonlinear dynamic heterogeneous panels. SSE/EFI Working Paper Series in Economics and Finance No. 582.

Kremers, J. J. M. (1989). U.s. federal indebtedness and the conduct of fiscal policy, *Journal of Monetary Economics* **23**: 219–38.

Kwiatkowski, D., Phillips, P. C., Schmidt, P. and Shin, Y. (1992). Testing the null hypothesis of stationarity against the alternative of a unit root: How sure are we that economic time series have a unit root?, *Journal of Econometrics* **54**: 159–178.

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

Park, J. and Phillips, P. C. B. (1999). Asymptotics for nonlinear transformations of time series, *Econometric Theory* **15**: 269–298.

Park, J. and Phillips, P. C. B. (2001). Nonlinear regressions with integrated time series, *Econometrica* **69**: 117–161.

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. (1991). Optimal inference in cointegrated systems, *Econometrica* **59**: 283–306.

Phillips, P. C. B. and Durlauf, S. N. (1986). Multiple time series regression with integrated processes, *Review od Economic Studies* **53**: 473–495.

Saikkonen, P. and Choi, I. (2004). Cointegrating smooth transition regressions, *Econometric Theory* **20**(2): 301–340.

Stock, J. H. (1987). Asymptotic properties of least square estimators of cointegrating vectors, *Econometrica* **55**: 1035–56.

Teräsvirta, T. (1994). Specification, estimation, and evaluation of smooth transition autoregressive models, *Journal of the American Statistical Association* **88**(425): 208–218.

Weisstein, E. W. (n.d.). Normal product distribution. From MathWorld - A Wolfram Web Resource.
URL: *http://mathworld.wolfram.com/NormalProductDistribution.html*

APPENDIX

A . Proofs of Theorems

The following Lemma 1 is an extending of Proposition 18.1 in Hamilton (1994) and Lemma 2 is an application of Hansen (1992b). Let an $(n \times 1)$ vector $\mathbf{W}(r) = (W_1(r), \mathbf{W}_2'(r))'$ denote n-dimensional standard Brownian motion, in which a set of n independent processes, denoted by $W_1(r)$, $W_2(r)$,..., $W_n(r)$, are collected.

Lemma 1. *Consider the time series* $\mathbf{y}_t = (y_{1t}, \mathbf{y}_{2t})'$ *and* $\mathbf{u}_t = (\tilde{u}_t, \mathbf{v}_t')'$ *in (4.5) and (3.2b). Under Assumptions 1-4, the following results are established from Hamilton (1994):*

(a) $T^{-2} \sum_{t=1}^{T} \mathbf{y}_{2,t-1}\mathbf{y}_{2,t-1}' \xrightarrow{L} \mathbf{\Lambda}_2 \left\{ \int_0^1 [\mathbf{W}_2(r)] [\mathbf{W}_2(r)]' dr \right\} \mathbf{\Lambda}_2'$;

(b) $T^{-1} \sum_{t=1}^{T} \mathbf{y}_{2,t-1}\tilde{u}_t \xrightarrow{L} \Lambda_1 \mathbf{\Lambda}_2 \left\{ \int_0^1 [\mathbf{W}_2(r)] \, dW_1(r) \right\}$,

$T^{-1} \sum_{t=1}^{T} \mathbf{y}_{2,t-1}\mathbf{v}_t' \xrightarrow{L} \mathbf{\Lambda}_2 \left\{ \int_0^1 [\mathbf{W}_2(r)] \, d\mathbf{W}_2(r) \right\} \mathbf{\Lambda}_2' + \sum_{s=1}^{\infty} E(\mathbf{v}_{t-s}\mathbf{v}_t')$;

(c) $T^{-1/2} \sum_{t=1}^{T} \tilde{u}_t \xrightarrow{L} \Lambda_1 \cdot W_1(1)$,

$T^{-1/2} \sum_{t=1}^{T} \mathbf{v}_t \xrightarrow{L} \mathbf{\Lambda}_2 \cdot \mathbf{W}_2(1)$;

(d) $T^{-3/2} \sum_{t=1}^{T} t\tilde{u}_t \xrightarrow{L} \Lambda_1 \cdot \left\{ W_1(1) - \int_0^1 W_1(r) dr \right\}$,

$T^{-3/2} \sum_{t=1}^{T} t\mathbf{v}_t \xrightarrow{L} \mathbf{\Lambda}_2 \cdot \left\{ \mathbf{W}_2(1) - \int_0^1 \mathbf{W}_2(r) dr \right\}$;

(e) $T^{-3/2} \sum_{t=1}^{T} \mathbf{y}_{2,t-1} \xrightarrow{L} \mathbf{\Lambda}_2 \cdot \int_0^1 \mathbf{W}_2(r) dr$;

(f) $T^{-5/2} \sum_{t=1}^{T} t\mathbf{y}_{2,t-1} \xrightarrow{L} \mathbf{\Lambda}_2 \cdot \int_0^1 r\mathbf{W}_2(r) dr$;

(g) $T^{-3} \sum_{t=1}^{T} t\mathbf{y}_{2,t-1}\mathbf{y}_{2,t-1}' \xrightarrow{L} \mathbf{\Lambda}_2 \left\{ \int_0^1 r \, [\mathbf{W}_2(r)] [\mathbf{W}_2(r)]' dr \right\} \mathbf{\Lambda}_2'$;

(h) $T^{-(\nu+1)} \sum_{t=1}^{T} t^{\nu} \xrightarrow{p} 1/(\nu+1)$ *for* $\nu \geq 0$;

(i) $T^{-1} \sum_{t=1}^{T} \mathbf{u}_t\mathbf{u}_{t-s}' \xrightarrow{p} E(\mathbf{u}_t\mathbf{u}_{t-s}')$ *for* $s \geq 0$,

in which the definitions of Λ_1 *and* $\mathbf{\Lambda}_2$ *can be found in Assumption 4.*

Lemma 2. *Consider the time series* $\mathbf{y}_t = (y_{1t}, \mathbf{y}_{2t})'$ *and* $\mathbf{u}_t = (\tilde{u}_t, \mathbf{v}_t')'$ *in (4.5) and (3.2b). Assume* s_{it} *and* s_{jt} $(i, j = 2, 3, ..., n)$ *are zero-mean stationary processes. Under Assumptions 1-4, applying the results in Hansen (1992b) yields:*

(a) The $(i-1, j-1)$*th element of matrix* $T^{-2} \sum_{t=1}^{T} \mathbf{s}_t\mathbf{y}_{2t}(\mathbf{s}_t\mathbf{y}_{2t})'$,

$T^{-2} \sum_{t=1}^{T} s_{it}s_{jt}y_{i,t-1}y_{j,t-1} \xrightarrow{L} E(s_{it}s_{jt})\{\mathbf{D}\}_{i-1,j-1}$, *in which* $\{\mathbf{D}\}_{i-1,j-1}$ *is the* $(i-1, j-1)$*th element of matrix* $\mathbf{D} = \mathbf{\Lambda}_2 \left\{ \int_0^1 [\mathbf{W}_2(r)] [\mathbf{W}_2(r)]' dr \right\} \mathbf{\Lambda}_2'$;

(b) $T^{-1}\sum_{t=1}^{T}\mathbf{y}_{2t-1}\mathbf{e}_t^{*'}\xrightarrow{L}\mathbf{\Lambda}_2\int_0^1\mathbf{W}_2(r)d\mathbf{W}_2^{*'}(r)\mathbf{\Lambda}_4^{*'}$, *in which* $\mathbf{e}_t^{*}=(\tilde{u}_t s_{2t},\tilde{u}_t s_{3t},...,\tilde{u}_t s_{nt})'$ *and* $\mathbf{\Lambda}_4^{*}$ *is defined in the proof of Lemma 2 (b);*

(c) $\sum_{t=1}^{T}\mathbf{y}_{2t-1}\mathbf{y}_{2t-1}'\mathbf{s}_t'=O_p(T^{3/2})$.

Proof of Lemma 2. (a) Lemma 1 (a) gives $T^{-2}\sum_{t=1}^{T}y_{i,t-1}y_{j,t-1}\xrightarrow{L}\{\mathbf{D}\}_{i-1,j-1}$, in which $\{\mathbf{D}\}_{i-1,j-1}$ is the $(i-1,j-1)$th element of

$$\mathbf{D}=\mathbf{\Lambda}_2\left\{\int_0^1\left[\mathbf{W}_2\left(r\right)\right]\left[\mathbf{W}_2\left(r\right)\right]'dr\right\}\mathbf{\Lambda}_2'.$$

Theorem 3.3 in Hansen (1992b) implies that

$$\sup\left|\frac{1}{T}\sum_{t=1}^{T}\left(y_{i,t-1}y_{j,t-1}\right)\left(s_{it}s_{jt}-E(s_{it}s_{jt})\right)\right|\xrightarrow{p}0\ ,\qquad\forall i,j=2,\ldots,n,$$

from which $T^{-2}\sum_{t=1}^{T}\left(y_{i,t-1}y_{j,t-1}\right)\left(s_{it}s_{jt}-E(s_{it}s_{jt})\right)\xrightarrow{p}0$. It follows that

$$\begin{aligned}&T^{-2}\sum_{t=1}^{T}s_{it}s_{jt}y_{i,t-1}y_{j,t-1}\\&=T^{-2}\sum_{t=1}^{T}\left(s_{it}s_{jt}-E(s_{it}s_{jt})\right)y_{i,t-1}y_{j,t-1}+T^{-2}\sum_{t=1}^{T}E(s_{it}s_{jt})y_{i,t-1}y_{j,t-1}\\&\xrightarrow{L}E(s_{it}s_{jt})\{\mathbf{D}\}_{i-1,j-1}.\end{aligned}$$

(b) Let $\mathbf{e}_t^{*}=(\tilde{u}_t s_{2t},\tilde{u}_t s_{3t},...,\tilde{u}_t s_{nt})'$. We set $\mathbf{V}_t=\mathbf{V}_{t-1}+\mathbf{e}_t^{*}$ and $\mathbf{U}_t=\mathbf{y}_{2,t-1}$ and define $\mathbf{U}_t^{*}=(\mathbf{U}_t',\mathbf{V}_t')'$ and $\tilde{\mathbf{e}}_t^{*}=(\mathbf{v}_{t-1}',\mathbf{e}_{t-1}^{*'})'$. It follows from Theorem 4.1 in Hansen (1992b) that

$$T^{-1}\sum_{t=1}^{T}\mathbf{U}_t^{*}\tilde{\mathbf{e}}_{t+1}^{*'}\xrightarrow{L}\mathbf{\Lambda}^{*}\int_0^1\mathbf{W}^{*}(r)d\mathbf{W}^{*'}(r)\mathbf{\Lambda}^{*'}+\mathbf{\Sigma}_1$$

in which $\mathbf{\Sigma}_1=\sum_{s=1}^{\infty}E(\tilde{\mathbf{e}}_{t-s}^{*}\tilde{\mathbf{e}}_t^{*'})$ and $\mathbf{W}^{*}(r)=\left(\mathbf{W}_2'(r),\mathbf{W}_2^{*'}(r)\right)'$. $\mathbf{W}_2^{*}(r)$ denotes $(n-1)$-dimensional standard Brownian motion. Similar to $\mathbf{\Lambda}$, $\mathbf{\Lambda}^{*}$ is defined as

$$\mathbf{\Lambda}^{*}=\begin{pmatrix}\mathbf{\Lambda}_2 & \mathbf{0}'\\ \mathbf{0} & \mathbf{\Lambda}_4^{*}\end{pmatrix}$$

because $\tilde{u}_t s_{it}$ is uncorrelated with $v_{j\tau}$ for all τ and t for each $i,j=2,3,...,n$ as

$E(\tilde{u}_t s_{it} v_{j\mathcal{T}}) = 0$, where $\mathbf{\Lambda}_4^* \mathbf{\Lambda}_4^{*'} = \sum_{s=-\infty}^{\infty} E(\mathbf{e}_t^* \mathbf{e}_{t-s}^{*'})$. Hence,

$$T^{-1} \sum_{t=1}^{T} \mathbf{y}_{2t-1} \mathbf{e}_t^{*'} = T^{-1} \sum_{t=1}^{T} \begin{pmatrix} \mathbf{I}_{n-1} & \mathbf{0} \end{pmatrix} \mathbf{U}_t^* \tilde{\mathbf{e}}_t^{*'} \begin{pmatrix} \mathbf{0} \\ \mathbf{I}_{n-1} \end{pmatrix}$$

$$\xrightarrow{L} \mathbf{\Lambda}_2 \int_0^1 \mathbf{W}_2(r) d\mathbf{W}_2^{*'}(r) \mathbf{\Lambda}_4^{*'} + \mathbf{\Sigma}_1^* \xrightarrow{p} \mathbf{\Lambda}_2 \int_0^1 \mathbf{W}_2(r) d\mathbf{W}_2^{*'}(r) \mathbf{\Lambda}_4^{*'}$$

where $\mathbf{\Sigma}_1^* = \sum_{s=1}^{\infty} E(\mathbf{e}_{t-s}^* \mathbf{v}_t') = \mathbf{0}$.

(c) It follows from Theorem 4.2 in Hansen (1992b) that

$$T^{-\frac{3}{2}} \sum_{t=1}^{T} \mathbf{y}_{2t-1} \otimes \mathbf{y}_{2t-1} \mathbf{v}_t'$$

$$\xrightarrow{L} \int_0^1 \mathbf{B} \otimes \mathbf{B} d\mathbf{B}' + \mathbf{\Sigma}_2 \otimes \int_0^1 \mathbf{B} + \int_0^1 \mathbf{B} \otimes \mathbf{\Sigma}_2$$

in which $\mathbf{B} = \mathbf{L}\mathbf{W}_2(r)$ denotes $(n-1)$-dimensional Brownian motion with covariance matrix $(T-1)^{-1} E\left(\sum_{t=1}^{T-1} \mathbf{v}_t \sum_{t=1}^{T-1} \mathbf{v}_t'\right) \xrightarrow{p} \mathbf{L}\mathbf{L}'$ as $T \to \infty$, $T^{-1} \sum_{i=1}^{T} \sum_{j=i+1}^{\infty} E\left(\mathbf{v}_{i-1} \mathbf{v}_{j-1}'\right) \xrightarrow{p} \mathbf{\Sigma}_2$ as $T \to \infty$, and the symbol "$\otimes$" denotes Kronecker product of two matrices.

□

Proof of Theorem 1. (a) Consider (4.5) with stationary transition variables s_{it} ($i = 2, 3, ..., n$). The deviation of the *OLS* estimate $\hat{\boldsymbol{\theta}}_T$ from the true value $\boldsymbol{\theta}_0$ under the null (4.8) is given by

$$\hat{\boldsymbol{\theta}}_T - \boldsymbol{\theta}_0 = \begin{pmatrix} \hat{\boldsymbol{\zeta}_T} - \boldsymbol{\zeta}_0 \\ \hat{\boldsymbol{\beta}_T} - \boldsymbol{\beta}_0 \end{pmatrix} = \mathbf{A}^{-1} \mathbf{B}, \tag{A.1}$$

in which

$$\mathbf{A} = \begin{pmatrix} \sum \mathbf{v}_t^* \mathbf{v}_t^{*'} & \sum \mathbf{v}_t^* \mathbf{y}_{2t}' & \sum \mathbf{v}_t^* (\mathbf{s}_t \mathbf{y}_{2t})' \\ \sum \mathbf{y}_{2t} \mathbf{v}_t^{*'} & \sum \mathbf{y}_{2t} \mathbf{y}_{2t}' & \sum \mathbf{y}_{2t} \mathbf{y}_{2t}' \mathbf{s}_t' \\ \sum \mathbf{s}_t \mathbf{y}_{2t} \mathbf{v}_t^{*'} & \sum \mathbf{s}_t \mathbf{y}_{2t} \mathbf{y}_{2t}' & \sum \mathbf{s}_t \mathbf{y}_{2t} (\mathbf{s}_t \mathbf{y}_{2t})' \end{pmatrix} \text{ and } \mathbf{B} = \begin{pmatrix} \sum \mathbf{v}_t^* \tilde{u}_t \\ \sum \mathbf{y}_{2t} \tilde{u}_t \\ \sum \mathbf{s}_t \mathbf{y}_{2t} \tilde{u}_t \end{pmatrix}.$$

Define the scaling matrix in this case as

$$\tilde{\mathbf{\Upsilon}}_T = \begin{pmatrix} T^{1/2} \mathbf{I}_{(2p+1)} & \mathbf{0} \\ \mathbf{0}' & T \mathbf{I}_{(2n-2)} \end{pmatrix}.$$

Premultiplying (A.1) by $\tilde{\boldsymbol{\Upsilon}}_T$ results in

$$\tilde{\boldsymbol{\Upsilon}}_T\left(\hat{\boldsymbol{\theta}_T}-\boldsymbol{\theta}_0\right)=\left(\tilde{\boldsymbol{\Upsilon}}_T^{-1}\mathbf{A}\tilde{\boldsymbol{\Upsilon}}_T^{-1}\right)^{-1}\left(\tilde{\boldsymbol{\Upsilon}}_T^{-1}\mathbf{B}\right). \tag{A.2}$$

When Assumptions 1-4 hold, consider the first factor in the right-hand side of (A.2),

$$\tilde{\boldsymbol{\Upsilon}}_T^{-1}\mathbf{A}\tilde{\boldsymbol{\Upsilon}}_T^{-1}=\tilde{\mathbf{Q}}_T=\begin{pmatrix} T^{-1}\sum\mathbf{v}_t^*\mathbf{v}_t^{*'} & T^{-3/2}\sum\mathbf{v}_t^*\mathbf{y}_{2t}' & T^{-3/2}\sum\mathbf{v}_t^*(\mathbf{s}_t\mathbf{y}_{2t})' \\ T^{-3/2}\sum\mathbf{y}_{2t}\mathbf{v}_t^{*'} & T^{-2}\sum\mathbf{y}_{2t}\mathbf{y}_{2t}' & T^{-2}\sum\mathbf{y}_{2t}\mathbf{y}_{2t}'\mathbf{s}_t' \\ T^{-3/2}\sum\mathbf{s}_t\mathbf{y}_{2t}\mathbf{v}_t^{*'} & T^{-2}\sum\mathbf{s}_t\mathbf{y}_{2t}\mathbf{y}_{2t}' & T^{-2}\sum\mathbf{s}_t\mathbf{y}_{2t}(\mathbf{s}_t\mathbf{y}_{2t})'\end{pmatrix}.$$

We show that $\tilde{\mathbf{Q}}_T\xrightarrow{L}\tilde{\mathbf{Q}}$ holds where $\tilde{\mathbf{Q}}$ is defined in (5.2). This is because that:

1) $T^{-1}\sum_{t=1}^T\mathbf{v}_t^*\mathbf{v}_t^{*'}\xrightarrow{p}\tilde{\mathbf{Q}}_{11}$ by the law of large numbers (LLN) (see Hamilton 1994, p.191);

2) $T^{-2}\sum_{t=1}^T\mathbf{y}_{2t}\mathbf{y}_{2t}'\xrightarrow{L}\tilde{\mathbf{Q}}_{22}$ holds because

$$\sum_{t=1}^T\mathbf{y}_{2t}\mathbf{y}_{2t}'=\sum_{t=1}^T\mathbf{y}_{2,t-1}\mathbf{y}_{2,t-1}'+\sum_{t=1}^T\mathbf{y}_{2,t-1}\mathbf{v}_t'+\sum_{t=1}^T\mathbf{v}_t\mathbf{y}_{2,t-1}'+\sum_{t=1}^T\mathbf{v}_t\mathbf{v}_t',$$

in which $T^{-2}\sum_{t=1}^T\mathbf{y}_{2,t-1}\mathbf{y}_{2,t-1}'\xrightarrow{L}\boldsymbol{\Lambda}_2\left\{\int_0^1[\mathbf{W}_2(r)][\mathbf{W}_2(r)]'dr\right\}\boldsymbol{\Lambda}_2'$ from Lemma 1 (a), $\sum_{t=1}^T\mathbf{y}_{2,t-1}\mathbf{v}_t'=O_p(T)$ from Lemma 1 (b), and $\sum_{t=1}^T\mathbf{v}_t\mathbf{v}_t'=O_p(T)$ by the LLN;

3) $T^{-2}\sum_{t=1}^T\mathbf{s}_t\mathbf{y}_{2t}(\mathbf{s}_t\mathbf{y}_{2t})'\xrightarrow{L}\tilde{\mathbf{Q}}_{33}$ holds because

$$\begin{aligned}\sum_{t=1}^T\mathbf{s}_t\mathbf{y}_{2t}(\mathbf{s}_t\mathbf{y}_{2t})'=&\sum_{t=1}^T\mathbf{s}_t\mathbf{y}_{2,t-1}(\mathbf{s}_t\mathbf{y}_{2,t-1})'+\sum_{t=1}^T\mathbf{s}_t\mathbf{y}_{2,t-1}(\mathbf{s}_t\mathbf{v}_t)'+\sum_{t=1}^T\mathbf{s}_t\mathbf{v}_t(\mathbf{s}_t\mathbf{y}_{2,t-1})'\\&+\sum_{t=1}^T\mathbf{s}_t\mathbf{v}_t(\mathbf{s}_t\mathbf{v}_t)',\end{aligned}$$

in which $\sum_{t=1}^T\mathbf{s}_t\mathbf{v}_t(\mathbf{s}_t\mathbf{v}_t)'=O_p(T)$ by the LLN, $\sum_{t=1}^T\mathbf{s}_t\mathbf{y}_{2,t-1}(\mathbf{s}_t\mathbf{v}_t)'=O_p(T)$ from Lemma 2 (b), the $(i-1,j-1)$th element of $T^{-2}\sum_{t=1}^T\mathbf{s}_t\mathbf{y}_{2,t-1}(\mathbf{s}_t\mathbf{y}_{2,t-1})'$, $T^{-2}\sum_{t=1}^T s_{it}s_{jt}y_{i,t-1}y_{j,t-1}\xrightarrow{L}E(s_{it}s_{jt})\{\tilde{\mathbf{Q}}_{22}\}_{i-1,j-1}$ from Lemma 2 (a);

4) $T^{-3/2}\sum_{t=1}^T\mathbf{y}_{2t}\mathbf{v}_t^{*'}\xrightarrow{p}\mathbf{0}$ because $\sum_{t=1}^T\mathbf{y}_{2t}\mathbf{v}_t^{*'}=\sum_{t=1}^T\mathbf{y}_{2t-1}\mathbf{v}_t^{*'}+\sum_{t=1}^T\mathbf{v}_t\mathbf{v}_t^{*'}=O_p(T)$ by Lemma 1 (b) and the LLN; Similarly, it follows from Lemma 2 (b)-(c) and the

LLN that $T^{-3/2}\sum_{t=1}^{T}\mathbf{s}_t\mathbf{y}_{2t}\mathbf{v}_t^{*\prime}\xrightarrow{p}\mathbf{0}$ and $T^{-2}\sum_{t=1}^{T}\mathbf{s}_t\mathbf{y}_{2t}\mathbf{y}_{2t}^{\prime}\xrightarrow{p}\mathbf{0}$.

From continuous mapping theorem, it follows that

$$\left(\tilde{\mathbf{\Upsilon}}_T^{-1}\mathbf{A}\tilde{\mathbf{\Upsilon}}_T^{-1}\right)^{-1}=\tilde{\mathbf{Q}}_T^{-1}\xrightarrow{L}\tilde{\mathbf{Q}}^{-1}. \tag{A.3}$$

Next, consider the second factor in the right-hand side of (A.2),

$$\tilde{\mathbf{\Upsilon}}_T^{-1}\mathbf{B}=\begin{pmatrix}T^{-1/2}\sum\mathbf{v}_t^{*}\tilde{u}_t\\ T^{-1}\sum\mathbf{y}_{2t}\tilde{u}_t\\ T^{-1}\sum\mathbf{s}_t\mathbf{y}_{2t}\tilde{u}_t\end{pmatrix}.$$

We show that

$$\tilde{\mathbf{\Upsilon}}_T^{-1}\mathbf{B}\xrightarrow{L}\tilde{\mathbf{h}}, \tag{A.4}$$

in which $\tilde{\mathbf{h}}$ is defined in (5.2). This can be seen in the following:

5) $T^{-1/2}\sum_{t=1}^{T}\mathbf{v}_t^{*}\tilde{u}_t\xrightarrow{L}\tilde{\mathbf{h}}_1$ from Lemma 1 (c);

6) $T^{-1}\sum_{t=1}^{T}\mathbf{y}_{2t}\tilde{u}_t\xrightarrow{L}\tilde{\mathbf{h}}_2$ because

$$T^{-1}\sum_{t=1}^{T}\mathbf{y}_{2t}\tilde{u}_t=T^{-1}\sum_{t=1}^{T}\mathbf{y}_{2,t-1}\tilde{u}_t+T^{-1}\sum_{t=1}^{T}\mathbf{v}_t\tilde{u}_t,$$

in which $T^{-1}\sum_{t=1}^{T}\mathbf{y}_{2,t-1}\tilde{u}_t\xrightarrow{L}\Lambda_1\mathbf{\Lambda}_2\left\{\int_0^1\left[\mathbf{W}_2\left(r\right)\right]dW_1\left(r\right)\right\}$ from Lemma 1 (b) and $T^{-1}\sum_{t=1}^{T}\mathbf{v}_t\tilde{u}_t\xrightarrow{p}E(\mathbf{v}_t\tilde{u}_t)=\mathbf{0}$;

7) $T^{-1}\sum\limits_{t=1}^{T}\mathbf{s}_t\mathbf{y}_{2t}\tilde{u}_t\xrightarrow{L}\tilde{\mathbf{h}}_3$ because

$$T^{-1}\sum_{t=1}^{T}\mathbf{s}_t\mathbf{y}_{2t}\tilde{u}_t=T^{-1}\sum_{t=1}^{T}\tilde{u}_t\mathbf{s}_t\mathbf{y}_{2t-1}+T^{-1}\sum_{t=1}^{T}\tilde{u}_t\mathbf{s}_t\mathbf{v}_t,$$

in which $T^{-1}\sum_{t=1}^{T}\tilde{u}_t\mathbf{s}_t\mathbf{y}_{2t-1}\xrightarrow{L}\tilde{\mathbf{h}}_3$ from Lemma 2 (b), whereas the $(i-1)$th element of the $(n-1)$ vector $\tilde{\mathbf{h}}_3$, denoted by $\{\tilde{\mathbf{h}}_3\}_{i-1}$, is the $(i-1)$th diagonal element of the $(n-1)\times(n-1)$ matrix $\mathbf{\Lambda}_2\int_0^1\mathbf{W}_2(r)d\mathbf{W}_2^{*\prime}(r)\mathbf{\Lambda}_4^{*\prime}$ in Lemma 2 (b), and $T^{-1}\sum_{t=1}^{T}\tilde{u}_t\mathbf{s}_t\mathbf{v}_t\xrightarrow{p}E(\tilde{u}_t\mathbf{v}_{t-d}\mathbf{v}_t)=0$.

Substituting (A.3) through (A.4) into (A.2) gives

$$\tilde{\boldsymbol{\Upsilon}}_T\left(\hat{\boldsymbol{\theta}_T}-\boldsymbol{\theta}_0\right)\overset{L}{\to}\tilde{\mathbf{Q}}^{-1}\tilde{\mathbf{h}}. \tag{A.5}$$

(b) Substituting (A.5) and (A.3) into (4.10) results in

$$\begin{aligned}\tilde{F}_T &\overset{L}{\to}\left(\tilde{\mathbf{R}}\tilde{\mathbf{Q}}^{-1}\tilde{\mathbf{h}}\right)'\left\{E(\tilde{u}_t^2)\cdot\tilde{\mathbf{R}}\tilde{\mathbf{Q}}^{-1}\tilde{\mathbf{R}}m\right\}^{-1}\left(\tilde{\mathbf{R}}\tilde{\mathbf{Q}}^{-1}\tilde{\mathbf{h}}\right)\\ &=(\tilde{\mathbf{Q}}_{33}^{-1}\tilde{\mathbf{h}}_3)'\tilde{\mathbf{Q}}_{33}(\tilde{\mathbf{Q}}_{33}^{-1}\tilde{\mathbf{h}}_3)/(m\Lambda_1^2)\\ &=\tilde{\mathbf{h}}_3'\tilde{\mathbf{Q}}_{33}'^{-1}\tilde{\mathbf{h}}_3/(m\Lambda_1^2).\end{aligned}$$

□

Proof of Theorem 2. (a) Consider (4.6) with stationary transition variables s_{it} $(i = 2,3,...,n)$. The deviation of the *OLS* estimate $\hat{\boldsymbol{\theta}_T^*}$ from the true value $\boldsymbol{\theta}_0^*$ under the null (4.9) is given by

$$\hat{\boldsymbol{\theta}_T^*}-\boldsymbol{\theta}_0^*=\begin{pmatrix}\hat{\boldsymbol{\zeta}_T}-\boldsymbol{\zeta}_0\\ \hat{\boldsymbol{\beta}}_{2T}-\boldsymbol{\beta}_{20}\\ \hat{\boldsymbol{\beta}}_{3T}-\boldsymbol{\beta}_{30}\\ \hat{\delta}_T^*-\delta_0^*\\ \hat{a}_{1T}-a_{10}\end{pmatrix}=\mathbf{A}^{-1}\mathbf{B}, \tag{A.6}$$

in which

$$\mathbf{A}=\begin{pmatrix}\sum\mathbf{v}_t^*\mathbf{v}_t^{*\prime} & \sum\mathbf{v}_t^*\boldsymbol{\xi}_{2t}' & \sum\mathbf{v}_t^*(\mathbf{s}_t\boldsymbol{\xi}_{2t})' & \sum t\mathbf{v}_t^* & \sum\mathbf{v}_t^*\\ \sum\boldsymbol{\xi}_{2t}\mathbf{v}_t^{*\prime} & \sum\boldsymbol{\xi}_{2t}\boldsymbol{\xi}_{2t}' & \sum\boldsymbol{\xi}_{2t}(\mathbf{s}_t\boldsymbol{\xi}_{2t})' & \sum t\boldsymbol{\xi}_{2t} & \sum\boldsymbol{\xi}_{2t}\\ \sum\mathbf{s}_t\boldsymbol{\xi}_{2t}\mathbf{v}_t^{*\prime} & \sum\mathbf{s}_t\boldsymbol{\xi}_{2t}\boldsymbol{\xi}_{2t}' & \sum\mathbf{s}_t\boldsymbol{\xi}_{2t}(\mathbf{s}_t\boldsymbol{\xi}_{2t})' & \sum t\mathbf{s}_t\boldsymbol{\xi}_{2t} & \sum\mathbf{s}_t\boldsymbol{\xi}_{2t}\\ \sum t\mathbf{v}_t^{*\prime} & \sum t\boldsymbol{\xi}_{2t}' & \sum t(\mathbf{s}_t\boldsymbol{\xi}_{2t})' & \sum t^2 & \sum t\\ \sum\mathbf{v}_t^{*\prime} & \sum\boldsymbol{\xi}_{2t}' & \sum(\mathbf{s}_t\boldsymbol{\xi}_{2t})' & \sum t & \sum 1\end{pmatrix}$$

and $\mathbf{B}=\left(\sum\mathbf{v}_t^{*\prime}\tilde{u}_t,\sum\boldsymbol{\xi}_{2t}'\tilde{u}_t,\sum(\mathbf{s}_t\boldsymbol{\xi}_{2t})'\tilde{u}_t,\sum t\tilde{u}_t,\sum\tilde{u}_t\right)'$. In this case, define the scaling matrix as

$$\tilde{\boldsymbol{\Upsilon}}_T^*=\begin{pmatrix}T^{1/2}\mathbf{I}_{(2p+1)\times(n-1)} & \mathbf{0} & \mathbf{0} & \mathbf{0}\\ \mathbf{0}' & T\mathbf{I}_{(2n-2)} & \mathbf{0} & \mathbf{0}\\ \mathbf{0}' & \mathbf{0}' & T^{3/2} & \mathbf{0}\\ \mathbf{0}' & \mathbf{0}' & \mathbf{0}' & T^{1/2}\end{pmatrix}.$$

Premultiplying (A.6) by $\tilde{\boldsymbol{\Upsilon}}_T^*$ gives

$$\tilde{\boldsymbol{\Upsilon}}_T^* \left(\hat{\boldsymbol{\theta}}_T^* - \boldsymbol{\theta}_0^*\right) = (\tilde{\boldsymbol{\Upsilon}}_T^{*-1}\mathbf{A}\tilde{\boldsymbol{\Upsilon}}_T^{*-1})^{-1}(\tilde{\boldsymbol{\Upsilon}}_T^{*-1}\mathbf{B}). \tag{A.7}$$

When Assumptions 1-4 hold, consider the first factor in the right-hand side of (A.7),

$$\tilde{\boldsymbol{\Upsilon}}_T^{*-1}\mathbf{A}\tilde{\boldsymbol{\Upsilon}}_T^{*-1} = \tilde{\mathbf{Q}}_T^* =$$

$$\begin{pmatrix} T^{-1}\sum \mathbf{v}_t^*\mathbf{v}_t^{*'} & T^{-3/2}\sum \mathbf{v}_t^*\boldsymbol{\xi}_{2t}' & T^{-3/2}\sum \mathbf{v}_t^*(\mathbf{s}_t\boldsymbol{\xi}_{2t})' & T^{-2}\sum t\mathbf{v}_t^* & T^{-1}\sum \mathbf{v}_t^* \\ T^{-3/2}\sum \boldsymbol{\xi}_{2t}\mathbf{v}_t^{*'} & T^{-2}\sum \boldsymbol{\xi}_{2t}\boldsymbol{\xi}_{2t}' & T^{-2}\sum \boldsymbol{\xi}_{2t}(\mathbf{s}_t\boldsymbol{\xi}_{2t})' & T^{-5/2}\sum t\boldsymbol{\xi}_{2t} & T^{-3/2}\sum \boldsymbol{\xi}_{2t} \\ T^{-3/2}\sum \mathbf{s}_t\boldsymbol{\xi}_{2t}\mathbf{v}_t^{*'} & T^{-2}\sum \mathbf{s}_t\boldsymbol{\xi}_{2t}\boldsymbol{\xi}_{2t}' & T^{-2}\sum \mathbf{s}_t\boldsymbol{\xi}_{2t}(\mathbf{s}_t\boldsymbol{\xi}_{2t})' & T^{-5/2}\sum t\mathbf{s}_t\boldsymbol{\xi}_{2t} & T^{-3/2}\sum \mathbf{s}_t\boldsymbol{\xi}_{2t} \\ T^{-2}\sum t\mathbf{v}_t^{*'} & T^{-5/2}\sum t\boldsymbol{\xi}_{2t}' & T^{-5/2}\sum t(\mathbf{s}_t\boldsymbol{\xi}_{2t})' & T^{-3}\sum t^2 & T^{-2}\sum t \\ T^{-1}\sum \mathbf{v}_t^{*'} & T^{-3/2}\sum \boldsymbol{\xi}_{2t}' & T^{-3/2}\sum (\mathbf{s}_t\boldsymbol{\xi}_{2t})' & T^{-2}\sum t & 1 \end{pmatrix}$$

It is shown that $\tilde{\mathbf{Q}}_T^* \xrightarrow{p} \tilde{\mathbf{Q}}^*$, in which $\tilde{\mathbf{Q}}^*$ is defined in (5.4), as follows:

1) It is easy to see from Theorem 1 that $\tilde{\mathbf{Q}}_{22}^* = \tilde{\mathbf{Q}}_{22}$, and $\tilde{\mathbf{Q}}_{11}^*$ and $\tilde{\mathbf{Q}}_{33}^*$ have same formula of $\tilde{\mathbf{Q}}_{33}$ and $\tilde{\mathbf{Q}}_{11}$ but depend on different $\mathbf{v}_t^*$ and $\mathbf{s}_t$ from Theorem 1;

3) By the LLN for martingale difference sequence, $T^{-2}\sum_{t=1}^T t\mathbf{v}_t^* \xrightarrow{p} \tilde{\mathbf{Q}}_{14}^*$ and $T^{-1}\sum_{t=1}^T \mathbf{v}_t^* \xrightarrow{L} \tilde{\mathbf{Q}}_{15}^*$; $\tilde{\mathbf{Q}}_{41}^* = \tilde{\mathbf{Q}}_{14}^{*'}$; $\tilde{\mathbf{Q}}_{51}^* = \tilde{\mathbf{Q}}_{15}^{*'}$;

4) $T^{-3/2}\sum_{t=1}^T \boldsymbol{\xi}_{2t}\mathbf{v}_t^{*'} \xrightarrow{L} \tilde{\mathbf{Q}}_{21}^*$ because

$$\begin{aligned} T^{-3/2}\sum_{t=1}^T \boldsymbol{\xi}_{2t}\mathbf{v}_t^{*'} &= T^{-3/2}\sum_{t=1}^T \boldsymbol{\xi}_{2t-1}\mathbf{v}_t^{*'} + T^{-3/2}\sum_{t=1}^T \mathbf{v}_t\mathbf{v}_t^{*'} \xrightarrow{L} T^{-3/2}\sum_{t=1}^T \boldsymbol{\xi}_{2t-1}E(\mathbf{v}_t^{*'}) \\ &\xrightarrow{L} \left(\boldsymbol{\Lambda}_2\int_0^1 \mathbf{W}_2(r)dr\right)E(\mathbf{v}_t^{*'}) \end{aligned}$$

since Lemma 1 (b) and $\sum_{t=1}^T \mathbf{v}_t\mathbf{v}_t^{*'} = O_p(T)$; $\tilde{\mathbf{Q}}_{12}^* = \tilde{\mathbf{Q}}_{21}^{*'}$;

5) $T^{-5/2}\sum_{t=1}^T t\boldsymbol{\xi}_{2t} \xrightarrow{L} \tilde{\mathbf{Q}}_{24}^*$, because

$$T^{-5/2}\sum_{t=1}^T t\boldsymbol{\xi}_{2t} = T^{-5/2}\sum_{t=1}^T t\boldsymbol{\xi}_{2t-1} + T^{-5/2}\sum_{t=1}^T t\mathbf{v}_t,$$

in which $T^{-5/2}\sum_{t=1}^T t\boldsymbol{\xi}_{2t-1} \xrightarrow{L} \boldsymbol{\Lambda}_2\cdot\int_0^1 r\mathbf{W}_2(r)dr$ from Lemma 1 (f) and $\sum_{t=1}^T t\mathbf{v}_t = O_p(T^{3/2})$ from Lemma 1 (d); $\tilde{\mathbf{Q}}_{42}^* = \tilde{\mathbf{Q}}_{24}^{*'}$;

6) $T^{-3/2}\sum_{t=1}^{T}\boldsymbol{\xi}_{2t}\xrightarrow{L}\tilde{\mathbf{Q}}_{25}^{*}$, because

$$T^{-3/2}\sum_{t=1}^{T}\boldsymbol{\xi}_{2t}=T^{-3/2}\sum_{t=1}^{T}\boldsymbol{\xi}_{2t-1}+T^{-3/2}\sum_{t=1}^{T}\mathbf{v}_{t},$$

in which $T^{-3/2}\sum_{t=1}^{T}\boldsymbol{\xi}_{2t-1}\xrightarrow{L}\boldsymbol{\Lambda}_2\cdot\int_0^1\mathbf{W}_2(r)dr$ from Lemma 1 (e) and $\sum_{t=1}^{T}\mathbf{v}_t=O_p(T^{1/2})$ from Lemma 1 (c); $\tilde{\mathbf{Q}}_{52}^{*}=\tilde{\mathbf{Q}}_{25}^{*'}$;

7) $T^{-3/2}\sum_{t=1}^{T}\mathbf{s}_t\boldsymbol{\xi}_{2t}\mathbf{v}_t^{*'}\xrightarrow{L}\tilde{\mathbf{Q}}_{31}^{*}$, because

$$T^{-3/2}\sum_{t=1}^{T}\mathbf{s}_t\boldsymbol{\xi}_{2t}\mathbf{v}_t^{*'}=T^{-3/2}\sum_{t=1}^{T}\mathbf{s}_t\boldsymbol{\xi}_{2t-1}\mathbf{v}_t^{*'}+T^{-3/2}\sum_{t=1}^{T}\mathbf{s}_t\mathbf{v}_t\mathbf{v}_t^{*'}$$

in which $\sum_{t=1}^{T}\mathbf{s}_t\mathbf{v}_t\mathbf{v}_t^{*'}=O_p(T)$ and the jth column of the first item would be

$$T^{-3/2}\sum_{t=1}^{T}v_{jt}^{*}\mathbf{s}_t\boldsymbol{\xi}_{2t-1}\xrightarrow{L}T^{-3/2}\sum_{t=1}^{T}E(v_{jt}^{*}\mathbf{s}_t)\boldsymbol{\xi}_{2t-1}\xrightarrow{L}E(v_{jt}^{*}\mathbf{s}_t)\tilde{\mathbf{Q}}_{25}^{*};$$

$\tilde{\mathbf{Q}}_{13}^{*}=\tilde{\mathbf{Q}}_{31}^{*'}$;

8) $T^{-2}\sum_{t=1}^{T}\mathbf{s}_t\boldsymbol{\xi}_{2t}\boldsymbol{\xi}_{2t}'\xrightarrow{L}\tilde{\mathbf{Q}}_{32}^{*}$, because

$$\begin{aligned}T^{-2}\sum_{t=1}^{T}\mathbf{s}_t\boldsymbol{\xi}_{2t}\boldsymbol{\xi}_{2t}'=&T^{-2}\sum_{t=1}^{T}\mathbf{s}_t\boldsymbol{\xi}_{2t-1}\boldsymbol{\xi}_{2t-1}'+T^{-2}\sum_{t=1}^{T}\mathbf{s}_t\boldsymbol{\xi}_{2t-1}\mathbf{v}_t'+T^{-2}\sum_{t=1}^{T}\mathbf{s}_t\mathbf{v}_t\boldsymbol{\xi}_{2t-1}'\\&+T^{-2}\sum_{t=1}^{T}\mathbf{s}_t\mathbf{v}_t\mathbf{v}_t'\end{aligned}$$

in which $\sum_{t=1}^{T}\mathbf{s}_t\boldsymbol{\xi}_{2t}\mathbf{v}_t'=O_p(T^{3/2})$, $\sum_{t=1}^{T}\mathbf{s}_t\mathbf{v}_t\mathbf{v}_t'=O_p(T)$, $\sum_{t=1}^{T}t\mathbf{s}_t\mathbf{v}_t=O_p(T^2)$, and

$$T^{-2}\sum_{t=1}^{T}\mathbf{s}_t\boldsymbol{\xi}_{2t-1}\boldsymbol{\xi}_{2t-1}'\xrightarrow{L}E(\mathbf{s}_t)T^{-2}\sum_{t=1}^{T}\boldsymbol{\xi}_{2t-1}\boldsymbol{\xi}_{2t-1}'\xrightarrow{L}E(\mathbf{s}_t)\boldsymbol{\Lambda}_2\cdot\int_0^1\mathbf{W}_2(r)\mathbf{W}_2'(r)dr\cdot\boldsymbol{\Lambda}_2'$$

since Lemma 2 (c); $\tilde{\mathbf{Q}}_{23}^{*}=\tilde{\mathbf{Q}}_{32}^{*'}$;

9) $T^{-5/2}\sum_{t=1}^{T}t\mathbf{s}_t\boldsymbol{\xi}_{2t}\xrightarrow{L}\tilde{\mathbf{Q}}_{34}^{*}$, because

$$T^{-5/2}\sum_{t=1}^{T}t\mathbf{s}_t\boldsymbol{\xi}_{2t}=T^{-5/2}\sum_{t=1}^{T}t\mathbf{s}_t\boldsymbol{\xi}_{2t-1}+T^{-5/2}\sum_{t=1}^{T}t\mathbf{s}_t\mathbf{v}_t$$

in which $T^{-5/2}\sum_{t=1}^{T} t\mathbf{s}_t\boldsymbol{\xi}_{2t-1} \xrightarrow{L} E(\mathbf{s}_t)\boldsymbol{\Lambda}_2\int_0^1 r\mathbf{W}_2(r)dr$ and $\sum_{t=1}^{T} t\mathbf{s}_t\mathbf{v}_t = O_p(T^2)$; $\tilde{\mathbf{Q}}_{43}^* = \tilde{\mathbf{Q}}_{34}^{*\prime}$;

10) $T^{-3/2}\sum_{t=1}^{T}\mathbf{s}_t\boldsymbol{\xi}_{2t} \xrightarrow{L} \tilde{\mathbf{Q}}_{35}^*$, because

$$T^{-3/2}\sum_{t=1}^{T}\mathbf{s}_t\boldsymbol{\xi}_{2t} = T^{-3/2}\sum_{t=1}^{T}\mathbf{s}_t\boldsymbol{\xi}_{2t-1} + T^{-3/2}\sum_{t=1}^{T}\mathbf{s}_t\mathbf{v}_t$$

in which $T^{-3/2}\sum_{t=1}^{T}\mathbf{s}_t\boldsymbol{\xi}_{2t} \xrightarrow{L} E(\mathbf{s}_t)\boldsymbol{\Lambda}_2\int_0^1\mathbf{W}_2(r)dr$, $\sum_{t=1}^{T}\mathbf{s}_t\mathbf{v}_t = O_p(T)$; $\tilde{\mathbf{Q}}_{53}^* = \tilde{\mathbf{Q}}_{35}^{*\prime}$;

11) $T^{-3}\sum_{t=1}^{T} t^2 \xrightarrow{p} \tilde{Q}_{44}^*$ and $T^{-2}\sum_{t=1}^{T} t \xrightarrow{p} \tilde{Q}_{45}^* = \tilde{Q}_{54}^*$ from Lemma 1 (h); $1 = \tilde{Q}_{55}^*$.

It follows that

$$\left(\tilde{\boldsymbol{\Upsilon}}_T^{*-1}\mathbf{A}\tilde{\boldsymbol{\Upsilon}}_T^{*-1}\right)^{-1} = \tilde{\mathbf{Q}}_T^{*-1} \xrightarrow{L} \tilde{\mathbf{Q}}^{*-1} \tag{A.8}$$

by continuous mapping theorem. Next, consider the second factor in the right-hand side of (A.7),

$$\tilde{\boldsymbol{\Upsilon}}_T^{*-1}\mathbf{B} = \begin{pmatrix} T^{-1/2}\sum\mathbf{v}_t^*\tilde{u}_t \\ T^{-1}\sum\boldsymbol{\xi}_{2t}\tilde{u}_t \\ T^{-1}\sum\mathbf{s}_t\boldsymbol{\xi}_{2t}\tilde{u}_t \\ T^{-3/2}\sum t\tilde{u}_t \\ T^{-1/2}\sum\tilde{u}_t \end{pmatrix}$$

We show that

$$\tilde{\boldsymbol{\Upsilon}}_T^{*-1}\mathbf{B} \xrightarrow{L} \tilde{\mathbf{h}}^*, \tag{A.9}$$

in which $\tilde{\mathbf{h}}^*$ is defined in (5.4). This can be seen in the following:

12) $\tilde{\mathbf{h}}_2^* = \tilde{\mathbf{h}}_2$, and $\tilde{\mathbf{h}}_1^*$ and $\tilde{\mathbf{h}}_3^*$ have same formula to $\tilde{\mathbf{h}}_1$ and $\tilde{\mathbf{h}}_3$ in Theorem 1 but depend on different $\mathbf{v}_t^*$ and $\mathbf{s}_t$;

14) $T^{-3/2}\sum t\tilde{u}_t \xrightarrow{L} \tilde{h}_5^*$ from Lemma 1 (d);

15) $T^{-1/2}\sum\tilde{u}_t \xrightarrow{L} \tilde{h}_6^*$ from Lemma 1 (c).

Substituting (A.8) and (A.9) into (A.7) results in

$$\tilde{\boldsymbol{\Upsilon}}_T^*\left(\hat{\boldsymbol{\theta}}_T^* - \boldsymbol{\theta}_0^*\right) \xrightarrow{L} \tilde{\mathbf{Q}}^{*-1}\tilde{\mathbf{h}}^*. \tag{A.10}$$

(b) Substituting (A.10) and (A.8) into (4.11) results in

$$\tilde{F}_T \xrightarrow{L} \left(\tilde{\mathbf{R}}^* \tilde{\mathbf{Q}}^{*-1} \tilde{\mathbf{h}}^*\right)' \left\{\Lambda_1^2 \cdot \tilde{\mathbf{R}}^* \tilde{\mathbf{Q}}^{*-1} \tilde{\mathbf{R}}^* m\right\}^{-1} \left(\tilde{\mathbf{R}}^* \tilde{\mathbf{Q}}^{*-1} \tilde{\mathbf{h}}^*\right).$$

□

B. Simulation of unit root tests

The Monto Carlo experiments in this section study the finite-sample distributions of the nonstandard Dickey-Fuller-type t and F tests illustrated in the testing procedure in Section 7.1. The unit root tests in a nonlinear framework are not standard. Under the null of unit root in a STAR model, the simulation is based on the following data-generating process:

$$y_t = 0.5 + y_{t-1} + v_t$$

in which $v_t \sim n.i.d(0,1)$ with desired sample sizes $T = 50, 100, 200, 500, 1000$. The estimation is based on the auxiliary regression (7.2) with $s_t = y_{t-1}$ or (7.3) with $s_t = \Delta y_{t-1}$, and then the nonstandard Dickey-Fuller-type t and F tests in Section 7.1 are carried out. Each time, $T + 500$ observations are generated and the first 500 observations are discarded for minimizing the initial effects. The number of the replications of each experiment is 100,000. The finite-sample distributions of the tests are given in Table 11-14.

Table 11: Critical values for the Dickey-Fuller-type t test with $s_t = \Delta y_{t-1}$.

		Probability that t-type test is less than entry							
s_t	T	0.01	0.025	0.05	0.10	0.90	0.95	0.975	0.99
Δy_{t-1}	50	-4.17	-3.80	-3.51	-3.18	-1.18	-0.84	-0.53	-0.16
	100	-4.04	-3.72	-3.45	-3.15	-1.22	-0.89	-0.60	-0.24
	200	-4.00	-3.70	-3.43	-3.13	-1.23	-0.92	-0.63	-0.28
	500	-3.99	-3.67	-3.42	-3.14	-1.24	-0.94	-0.66	-0.31
	1000	-3.96	-3.67	-3.41	-3.13	-1.24	-0.94	-0.66	-0.31

The probability shown at the head of the column is the area in the left-hand tail.

Table 12: Critical values for the Dickey-Fuller-type t test with $s_t = y_{t-1}$.

		Probability that t-type test is less than entry							
s_t	T	0.01	0.025	0.05	0.10	0.90	0.95	0.975	0.99
y_{t-1}	50	-4.18	-3.81	-3.49	-3.12	-0.05	0.53	1.02	1.53
	100	-4.26	-3.91	-3.62	-3.29	-0.61	-0.04	0.46	1.04
	200	-4.31	-4.00	-3.73	-3.43	-1.15	-0.65	-0.18	0.36
	500	-4.36	-4.05	-3.80	-3.51	-1.54	-1.16	-0.78	-0.34
	1000	-4.35	-4.06	-3.82	-3.53	-1.66	-1.34	-1.01	-0.61

The probability shown at the head of the column is the area in the left-hand tail.

Table 13: Critical values for the Dickey-Fuller-type F test with $s_t = \Delta y_{t-1}$.

		Probability that F-type test is greater than entry							
s_t	T	0.99	0.975	0.95	0.90	0.10	0.05	0.025	0.01
Δy_{t-1}	50	0.65	0.79	0.93	1.13	4.39	5.07	5.87	6.90
	100	0.64	0.78	0.93	1.14	4.13	4.84	5.54	6.45
	200	0.63	0.78	0.93	1.14	4.06	4.77	5.45	6.29
	500	0.63	0.77	0.93	1.13	4.06	4.72	5.36	6.17
	1000	0.62	0.77	0.93	1.14	4.03	4.68	5.30	6.09

The probability shown at the head of the column is the area in the right-hand tail.

Table 14: Critical values for the Dickey-Fuller-type F test with $s_t = y_{t-1}$.

		Probability that F-type test is greater than entry							
s_t	T	0.99	0.975	0.95	0.90	0.10	0.05	0.025	0.01
y_{t-1}	50	0.86	1.04	1.22	1.46	4.93	5.76	6.61	7.77
	100	0.92	1.11	1.29	1.54	4.91	5.67	6.44	7.43
	200	0.96	1.15	1.34	1.59	4.91	5.64	6.34	7.28
	500	0.97	1.16	1.35	1.61	4.93	5.65	6.33	7.22
	1000	0.98	1.17	1.37	1.62	4.94	5.67	6.34	7.19

The probability shown at the head of the column is the area in the right-hand tail.