Testing Common Nonlinear Features in Vector Nonlinear Autoregressive Models

Dao Li  
Changli He  
Editor: Hasan Fleyeh

Nr: 2012:04
Testing Common Nonlinear Features in Vector Nonlinear Autoregressive Models

Dao Li, Changli He

August 12, 2013

Abstract

This paper studies a special class of vector smooth-transition autoregressive (VSTAR) models, which contains common nonlinear features (CNFs). We proposed a triangular representation for these models and developed a procedure for testing CNFs in a VSTAR model. We first test a unit root against a stationary STAR process for each individual time series and subsequently examine whether CNFs exist in the system with the Lagrange Multiplier (LM) test if unit root is rejected in the first step. The LM test has a standard Chi-square asymptotic distribution. The critical values of our unit root tests and finite-sample properties of the $F$ form of our LM test are studied by Monte Carlo simulations. We illustrate how to test and model CNFs using the monthly growth of consumption and income data for the United States (1985:1 to 2011:11).

KEYWORDS: Common features, Lagrange Multiplier test, Vector STAR models.

1. Introduction

A family of STAR models has been comprehensively studied for modeling nonlinear structures, such as data breaks and regime switching, in time series data since Teräsvirta (1994) introduced many theoretical issues and applications for the STAR models. The extensions of the univariate STAR models have been developed for exploring the nonlinearity of time series data. Escribano (1987) and Granger and Swanson (1999) introduced smooth-transition error-correction model

This paper studies a special class of VSTAR models to allow for CNFs. A group of time series is said to have CNFs if each of the time series is nonlinear, but their combination does not contain nonlinearity. Modeling vector nonlinear time series can be improved by imposing CNFs. The previous studies of common features are mostly related to the generalized method of moments estimations and canonical correlation in a parametric framework. For example see, Engle and Kozicki (1993) and Anderson and Vahid (1998). Engle and Kozicki (1993) provided a general method for detecting common features, similar to those used in cointegration by Engle and Granger (1987). The idea is straightforward when applied to common nonlinear features, but estimation and testing procedures must be developed for nonlinear models. Anderson and Vahid (1998) approached common nonlinearities in a similar method in multivariate nonlinear models. Their test of common nonlinearity is applicable to different nonlinear models, but their test replies on a polynomial approximation of the testing regression in which the nonlinear regressors come from the approximation of the corresponding individual testing regressions for nonlinearity. The possibility of some of the variables having common nonlinearity follows when the null of linearity is rejected in each variable. However, the approximation under the null of linearity is later used for testing common nonlinearity when the null is rejected. Consequently, the transition function suggested in Anderson and Vahid (1998) does not behave as a third degree polynomial expansion of the transition function when each series is nonlinear. Therefore, the proposed approximation for testing CNFs is not the approximation under the condition that each series is nonlinear, although it is an approximation for testing
the linearity of each series. Using an approximately specified model under both the null and the alternative hypotheses for testing common nonlinearity makes it difficult to interpret the test. However, their test seems promising for determining the number of CNFs. Further, as their procedure is based on using several stages to conclude how many nonlinear factors are present, the empirical size and power of each stage were not reported because authors think the interpretation is very difficult. Their work motivated this paper to seek another method to examine CNFs. It begins with the simplest case of \( n - 1 \) common nonlinearities with one linearly independent linear combination. Although the issue of how to determine the number of common nonlinearities is definitely worthy of consideration, this paper is the first step for developing another possible test for CNFs.

In this paper, studying one combination for CNFs, we propose an \( LM \) test to investigate CNFs in a VSTAR model based on a triangular representation of the VSTAR model containing CNFs. This avoids approximating the transition function with a Taylor expansion and nonlinear estimation is necessary. In the triangular representation, CNFs are tested in a VSTAR model to determine whether the population residual in the regression of CNF relations is linear. This is related to the \( LM \) test for the adequacy of a nonlinear specification against the alternative of additional STAR nonlinearity in \cite{litthoeim1996}, in which several tests for misspecification were derived, following \cite{godfrey1988}. However, the object of the testing for remaining nonlinearity in this paper is a combination of several series where the relation is unknown. We derive the asymptotic distribution of the test and examine its small sample properties in a simulation study. A convenient method for performing the \( LM \) test in the VSTAR models with CNFs is given in detail.

The empirical application in this paper concerns the monthly growth of the consumption expenditures and disposable income of the United States, dating from 1985:1 until 2011:11. The permanent-income hypothesis and the study in \cite{campbell1990} imply that consumption and income are linearly cointegrated in the levels (see \cite{engle1987}, for more discussions).
However, are there any other common features in those two dynamic time series? Removing the possibility of cointegration by looking at the growth, we investigate whether there are CNFs between consumption and income when each variable presents nonlinearity.

The remaining paper is outlined as follows. Section 2 studies VSTAR models. Section 3 discusses CNFs in a VSTAR model. The procedure for testing CNFs is provided in Section 4. The asymptotic distribution of our LM test is derived in Section 5. Simulation studies are presented in Section 6. An empirical application is illustrated in Section 7. and conclusions can be found in Section 8.

2. Vector Smooth-transition Autoregressive Models

Consider a vector smooth-transition autoregressive model of order \( p \) \( \text{VSTAR}(p) \),

\[
y_t = \mu + \sum_{k=1}^{p} \Phi_k y_{t-k} + \left( \tilde{\mu} + \sum_{k=1}^{p} \Gamma_k y_{t-k} \right) G (s_t; \gamma, c) + v_t \tag{2.1}
\]

where \( y_t = (y_{1t}, y_{2t}, ..., y_{nt})' \) denotes an \((n \times 1)\) vector time series, \( \mu \) and \( \tilde{\mu} \) are two \((n \times 1)\) constant vectors, \( \Phi_k \) and \( \Gamma_k \) are \((n \times n)\) coefficient matrices, \( G (s_t; \gamma, c) \) is a smooth-transition (ST) function, and the \((n \times 1)\) vector \( v_t \) is the error term, independent and identically distributed. The process \( y_t \) defined in
\( G(s_t; \gamma, c) \) is stationary and ergodic (see Assumption 2 in Section 5). When \( n = 1 \), (2.1) is a simple univariate STAR(p) model, such that
\[
y_t = \mu + \sum_{k=1}^{p} \Phi_k y_{t-k} + \left( \tilde{\mu} + \sum_{k=1}^{p} \Gamma_k y_{t-k} \right) G(s_t; \gamma, c) + v_t \tag{2.2}
\]
where \( \mu, \tilde{\mu}, \Phi_k \) and \( \Gamma_k \) are scaler parameters.

In this paper, we only consider logistic VSTAR models with the ST function
\( G(s_t; \gamma, c) \) defined as follows,
\[
G(s_t; \gamma, c) = \left( 1 + \exp \left\{ -\gamma \prod_{j=1}^{k} (s_{jt} - c_j) \right\} \right)^{-1}, \quad \gamma > 0, \tag{2.3}
\]
where \( \gamma \) is a slope parameter, \( c_j \) are location parameters and \( s_{jt} \) are transition variables. We assume \( s_{jt} = y_{ut-d} \) (\( d \leq p \)), where \( d \) is called the delay parameter. The ST function \( G(s_t; \gamma, c) \) is a bounded function on \([0, 1]\) and at least fourth-order differentiable with respect to \( \gamma \) in a neighborhood of \( \gamma = 0 \). Given any value of \( G(s_t; \gamma, c) \) between zero and one, all the roots of \( |I_n - \sum_{k=1}^{p} (\Phi_k + \Gamma_k G) z^k| = 0 \) lie outside the unit circle.

Figure 2 and Figure 5-8 depict the featured statistical properties of (2.3) (Figure 5-8 are in Appendix). When \( k = 1 \), (2.3) is a non-decreasing nonlinear
function. When \( k > 1 \), (2.3) is flexibly changeable depending on the different transition variables \( s_t \) and location parameters \( c \). For example, Figure 1 illustrates (2.3) with \( k > 1 \) and \( c_j \) that are the same for different \( j \) \((j = 1, ..., k)\); Figure 2 illustrates (2.3) with \( k > 1 \) and \( c_j \) that are different for different \( j \). Due to the statistical properties of the ST function (2.3), (2.1) not only captures the data break, dynamic structure or nonlinear shift in a vector of time series well but also nests the mostly applied vector linear autoregressions and threshold autoregressions. When \( \gamma = 0 \), (2.1) is reduced to a vector autoregression. When \( \gamma \) goes to infinity, (2.1) becomes a vector threshold autoregression.

3. Common Nonlinear Features in Vector STAR Models

Introduced by Engle and Kozicki (1993), the common features are some that features are present in each of several variables, and a linear combination of those variables has no such features. These common features are used for analyzing data properties in linear frameworks. The cointegrated system by Engle and Granger (1987) is a particular case that contains common stochastic trends.

In this paper, we consider a group of nonlinear time series \( y_t \) as defined in (2.1) and investigate whether there are CNFs in \( y_t \). An \((n \times 1)\) nonlinear vector \( y_t \) is said to contain CNFs if each of the series \( y_{it} \) is nonlinear and there is a nonzero vector \( \alpha \) such that \( \alpha' y_t \) is linear. The vector \( \alpha \) can be extended to an \( n \times (n - r) \) matrix whose rows are linearly independent indicating that there exist \( n - r \) different relations of CNFs (\( r \) CNFs). However, this paper only considers a simple \( n \times 1 \) vector \( \alpha \).

Suppose \( y_t \) contains CNFs. Then, we can find a nonzero \((n \times 1)\) vector \( \alpha \) such that the nonlinearity vanishes in the linear combination \( \alpha' y_t \). Now, multiplying both sides of (2.1) by the vector \( \alpha \) results in

\[
\alpha' y_t = \alpha' \mu + \sum_{k=1}^{p} \alpha' \Phi_k y_{t-k} + \left( \alpha' \hat{\mu} + \sum_{k=1}^{p} \alpha' \Gamma_k y_{t-k} \right) G(s_t; \gamma, c) + \alpha' v_t \quad (3.1)
\]

where \( y_t, \Phi_k, \Gamma_k, \mu, \hat{\mu} \) and \( v_t \) are the same as defined in (2.1) and \( G(s_t; \gamma, c) \) as
The existence of CNFs in $y_t$ implies that $\alpha' \tilde{\mu} = 0'$ and $\alpha' \Gamma_k = 0'$ for all $k$. Therefore, the linear combination (3.1) is equal to

$$\alpha' y_t = \alpha' \mu + \sum_{k=1}^{p} \alpha' \Phi_k y_{t-k} + \alpha' v_t. \quad (3.2)$$

This simple consideration is mostly practicable in such applications, and it is a reasonable first step for determining the number of CNFs.

Consider the following numerical example, a bivariate STAR(1) model with a CNF ($\mu = 0$, $\tilde{\mu} = 0$, $n = 2$, $t = 1, 2, ..., 200$):

$$\begin{pmatrix} y_{1t} \\ y_{2t} \end{pmatrix} = \begin{pmatrix} 0.03 & 0.05 \\ 0.02 & 0.1 \end{pmatrix} \begin{pmatrix} y_{1t-1} \\ y_{2t-1} \end{pmatrix} + \begin{pmatrix} 2 \\ 1 \end{pmatrix} (0.12y_{1t-1} + 0.07y_{2t-1}) G (s_t; \gamma, c) + \begin{pmatrix} v_{1t} \\ v_{2t} \end{pmatrix},$$

where $G (s_t; \gamma, c) = (1 + \exp \{-2(y_{2t-1} - 1)\})^{-1}$ and $(v_{1t}, v_{2t})'$ is generated from a $(2 \times 1)$ vector n.i.d$(0, 1)$ sequence. In the example, two processes, $y_{1t}$ and $y_{2t}$, share a nonlinear structure expressed as $(0.12y_{1t-1} + 0.07y_{2t-1}) G (s_t; \gamma, c)$, which denotes a CNF in this bivariate STAR(1) model. Figure 3 depicts the two nonlinear time series (solid, dashed) with CNFs in panel (a); their linear combination in panel (b).
ear time series that have a CNF in panel (a) and their linear combination in panel (b).

4. Testing Procedure

We propose a procedure that consists of two steps for testing CNFs in the VSTAR model in \((2.1)\). The first step is to test the null of a unit root process against a stationary STAR process for the individual time series. If the null is rejected which indicates that each of the series is nonlinear, the second step will examine whether there are CNFs between those series. However, testing a unit root is not a necessary first step before testing CNFs. When there is no reason to suspect a unit root, starting with testing linearity against nonlinearity is more realistic; we refer to Luukkonen et al. (1988) for the possibility. This paper starts from a non-standard unit root test for the emphasis of the likely misleading in the standard one because of nonlinearity.

4.1 Testing a unit root against a stationary STAR process

In the first step of the testing procedure, our non-standard unit root test is built up in a nonlinear framework \((2.2)\). The reason for not using the standard unit root tests is that nonlinear structures bias the classical unit root tests towards non-rejection (see Perron, 1990).

Take an STAR model \((2.2)\) with \(p = 1\) for convenience of discussion,

\[
y_t = \mu + \phi_1 y_{t-1} + (\tilde{\mu} + \Gamma_{1} y_{t-1}) G(s_t; \gamma, \mathbf{c}) + v_t. \tag{4.1}
\]

Under the null hypothesis, \(y_t\) is a unit root process. Therefore, the linearity restriction, \(\gamma = 0\), is imposed under the null. However, the restriction \(\gamma = 0\) causes a parameter identification problem (see Luukkonen et al. 1988, for details). The solution is that we replace \(G(s_t; \gamma, \mathbf{c})\) with its first-order Taylor expansion round \(\gamma = 0\) as a linear approximation. An auxiliary regression is then given in \((1.2)\) after ignoring the remainder in the Taylor expansion of \(G(s_t; \gamma, \mathbf{c})\) and
reparameterization,

\[ y_t = \mu^* + \phi_1^* y_{t-1} + \phi_2^* y_{t-1} + \phi_3^* s_t + v_t^*, \]  

(4.2)

where \( v_t^* \) is assumed \( i.i.d \) and \( s_t \) is usually taken as \( y_{t-1} \) or \( \Delta y_{t-1} \). Consequently, the null hypothesis for testing unit root in (4.1) is launched by

\[ H_0 : \mu^* = 0, \phi_1^* = 1, \phi_2^* = 0, \phi_3^* = 0. \]  

(4.3)

In particular, if \( s_t \) is specified as \( y_{t-1} \) in (4.2), then (4.2) will be updated to

\[ y_t = \mu^* + \tilde{\phi}_1^* y_{t-1} + \phi_2^* y_{t-1} + v_t^* \]  

(4.4)

and correspondingly (4.3) will be updated to

\[ H_0 : \mu^* = 0, \tilde{\phi}_1^* = 1, \phi_2^* = 0. \]  

(4.5)

We further consider (4.2) along with the null hypothesis of \( \phi_1^* = 1 \) and (4.4) with \( \tilde{\phi}_1^* = 1 \).

Although there are discussions of testing the unit root against nonlinear dynamic models in the literature (for example, He and Sandberg (2006) and Kapetanios and Shin (2006)), there are no distribution theory and critical values available for our cases. Table 1-2 in Section 6.1 provide the critical values for our non-standard Dickey-Fuller \( F \)-type and \( t \)-type tests for the empirical application in this paper. However, the distribution theory underlying the critical values is not reported here to concentrate on the primary aim, the test for CNFs.

4.2 Testing CNFs

In the second step of the testing procedure, testing CNFs in \( y_t \) in (2.1) when the null of a unit root is rejected in the first step, we propose a triangular representation of a system containing CNFs.
As described in Section 3, we can find a nonzero vector \( \alpha \) such that \( \alpha'y_t \) is linear formulated in (3.2) when CNFs exist. We normalize one of the nonzero elements in \( \alpha \) to unity and partition \( \alpha \) as \((1, -\alpha^*)'\). We then partition \( y_t \) as \((y_{1t}, y_{2t})'\) and \( v_t \) as \((v_{1t}, v_{2t})'\), correspondingly. Model (3.2) is therefore equivalent to

\[
y_{1t} = \mu_1 + \alpha^*y_{2t} + \sum_{k=1}^{p} \phi_{1k}^t y_{t-k} + u_t
\]

(4.6)

where \( \mu_1 = \alpha'\mu, \phi_{1k}^t = \alpha'\Phi_k, \) and \( u_t = \alpha'v_t = v_{1t} - \alpha^*v_{2t}. \)

Let \( y_t = (y_{1t}, y_{2t})' \) in (2.1) contain CNFs. It is straightforward to propose a triangular representation from (4.6) for the system \( y_t \), such that

\[
y_{1t} = \mu_1 + \alpha^*y_{2t} + \sum_{k=1}^{p} \phi_{1k}^t y_{t-k} + u_t
\]

\[
y_{2t} = \mu_2 + \sum_{k=1}^{p} \Phi_{2k} y_{t-k} + \left( \tilde{\mu}_2 + \sum_{k=1}^{p} \Gamma_{2k} y_{t-k} \right) G(s_t; \gamma, c) + v_{2t}
\]

(4.7)

where the first equation is (4.6) containing the CNF relation in the system of \( y_t \) and the second equation is the last \( n - 1 \) lines in (2.1) where \( \Phi_{2k} = (0, I_{n-1}) \Phi_k \) and \( \Gamma_{2k} = (0, I_{n-1}) \Gamma_k \) are \((n-1) \times n\) matrices.

When a system \( y_t \) contains CNFs, the population residual \( u_t \) in (4.6) does not appear nonlinear. Otherwise, \( u_t \) is a nonlinear process. Thus, our test for CNFs in \( y_t \) is constructed by the hypothesis

\[
H_0 : u_t \text{ is linear}
\]

\[
H_a : u_t \text{ is nonlinear}
\]

(4.8)

in (4.6)-(4.7).

In this section, we consider the following triangular representation with \( p = 1 \) for notation simplification,

\[
y_{1t} = \mu_1 + \alpha^*y_{2t} + \phi_{11}^t y_{t-1} + u_t
\]

(4.9a)

\[
y_{2t} = \mu_2 + \Phi_{21} y_{t-1} + \left( \tilde{\mu}_2 + \Gamma_{21} y_{t-1} \right) G(s_t; \gamma, c) + v_{2t}
\]

(4.9b)
where $\phi_{11}$ is an $(n \times 1)$ vector, $\mu_2$ and $\tilde{\mu}_2$ are $(n - 1) \times 1$ vectors, and $\Phi_{21}$ and $\Gamma_{21}$ are $(n - 1) \times n$ matrices. We illustrate an LM test for testing CNFs based on (4.9a)-(4.9b).

In principle, there are different ways to test (4.8). To test the linearity of $u_t$, we consider the population residual $u_t$ in (4.9a), which has the form:

$$u_t = \beta^*_0 + \beta^*_1 u_{t-1} + \beta^*_2 u_{t-1}^2 + e_t \quad (4.10)$$

under the alternative hypothesis, as an example. $e_t$ is independently and identically distributed. When a general VSTAR model with $p > 1$ is necessary, the order of the autoregression in (4.10) needs to increase because the null that $u_t$ is linear can be rejected by reason of that the error process contains relatively few lags.

To derive the LM test, we follow a procedure in Godfrey (1988, pages 114-115) that is used to test a statistical property for the population residual. The technique is to substitute (4.10) for the population residual $u_t$ in (4.9a) and write $u_{t-1}$ in terms of $u_{t-1} = y_{t-1} - \mu_1 - \alpha^r y_{2t-1} - \phi'_{11} y_{t-2}$. Then, (4.9a) takes the form of:

$$y_{t} = \mu_1 + \alpha^r y_{2t} + \phi'_{11} y_{t-1} + \beta^*_0 + \beta^*_1 \left(y_{t-1} - \mu_1 - \alpha^r y_{2t-1} - \phi'_{11} y_{t-2}\right)$$
$$\quad \quad + \beta^*_2 \left(y_{t-1} - \mu_1 - \alpha^r y_{2t-1} - \phi'_{11} y_{t-2}\right)^2 + e_t. \quad (4.11)$$

Reorganizing the parameters in (4.11) yields:

$$y_{t} = \mu_1 + \alpha^r y_{2t} + \phi'_{11} y_{t-1} + \phi'_{12} y_{t-2} + \beta^*_2 \left(\alpha^r y_{t-1} - \phi'_{11} y_{t-2}\right)^2 + e_t$$
$$\quad = \mu_1 + \alpha^r y_{2t} + \phi'_{11} y_{t-1} + \phi'_{12} y_{t-2} + \theta^r w_t + e_t \quad (4.12)$$
where \( \mathbf{w}_t \) and \( \boldsymbol{\theta} \) are \( ((2n^2 + n) \times 1) \) vectors, such that

\[
\mathbf{w}_t = (y_{1t-1}y_{1t-1}, \ldots, y_{1t-1}y_{nt}; y_{2t-1}y_{2t-1}, \ldots, y_{2t-1}y_{nt-1}, \ldots, y_{nt-1}y_{nt-1}; y_{1t-2}y_{1t-2}, \ldots, y_{1t-2}y_{nt}; y_{2t-2}y_{2t-2}, \ldots, y_{2t-2}y_{nt-2}, \ldots, y_{nt-2}y_{nt-2}; y_{1t-3}y_{1t-3}, \ldots, y_{1t-3}y_{nt}; y_{2t-3}y_{2t-3}, \ldots, y_{2t-3}y_{nt-2}, \ldots, y_{nt-3}y_{nt-2}; \ldots, \ldots, \ldots, \ldots, \ldots, \ldots, \ldots, \ldots, y_{1t-2}y_{1t-2}, \ldots, y_{1t-2}y_{nt}; y_{2t-2}y_{2t-2}, \ldots, y_{2t-2}y_{nt-2}, \ldots, y_{nt-2}y_{nt-2} ) \]

and

\[
\boldsymbol{\theta} = (\theta_{111}, \ldots, \theta_{11n}, \theta_{122}, \ldots, \theta_{12n}, \ldots, \theta_{1nn}; \theta_{211}, \ldots, \theta_{21n}, \theta_{222}, \ldots, \theta_{22n}, \ldots, \theta_{2nn}; \theta_{311}, \ldots, \theta_{31n}, \theta_{321}, \ldots, \theta_{32n}, \ldots, \theta_{3nn})' .
\]

Thus, the null hypothesis in (4.8) is rewritten to be

\[
H_0 : \boldsymbol{\theta} = \mathbf{0}.
\] (4.13)

However, \( y_{2t} \) in model (4.12) are endogenous regressors that are correlated with \( e_t \). One way to correct the correlations between \( y_{2t} \) and \( e_t \) is to add the leads and lags of

\[
\mathbf{v}_{2t} = y_{2t} - \mu_2 - \Phi_{21}y_{t-1} - (\tilde{\mu}_2 + \Gamma_{21}y_{t-1}) G(s_t; \gamma, \mathbf{c}),
\]

which was suggested in the literature by Saikonen (1991) and Stock and Watson (1993). Specifically, we can define:

\[
e_t = \sum_{s=-l}^{l} \delta'_s \mathbf{v}_{2t-s} + \tilde{e}_t
\] (4.14)

where \( \{ \tilde{e}_t \} \) is an \textit{i.i.d} sequence with mean zero, variance \( \sigma^2 \), and \( \tilde{e}_t \) is uncorrelated with \( \mathbf{v}_{2t-s} \) for \( s = -l, -l + 1, \ldots, l \). When the model is correctly specified, it is true that \( l = 0 \) is enough to correct the correlations between \( y_{2t} \) and \( e_t \) because \( \mathbf{v}_t \) is not serially correlated. However, additionally including leads and lags is more robust in the case where the model is not easy to specify exactly. Although the order \( p \) can be moderately large, this solution does not require a large number of observations, depending on \( l \), because \( l \) is not necessarily far away from zero. In the case where the model is not specified correctly and \( p \) is not small, a small \( l \) would not cause many difficulties because serial correlation usually shows up at
short lags.

Thereafter, (4.12) can be rewritten as:

\[
y_{1t} = \mu_1' + \alpha' y_{2t} + \phi_{11}' y_{t-1} + \phi_{12}' y_{t-2} + \theta' w_t + \sum_{s=-l}^{\max(2,l+1)} \delta'_s y_{2t-s} + \tilde{\epsilon}_t
\]

\[
= \tilde{\mu} + \theta' w_t + \delta'_1 y_{2t+1} + \sum_{s=-l+1}^{\max(2,l+1)} \delta'_s y_{1-s}
\]

\[
+ \sum_{s=-l}^{l} \delta'_{3a} y_{t-1-s} G(s_{t-s}; \gamma, c) + \sum_{s=-l}^{l} \delta'_{4a} G(s_{t-s}; \gamma, c) + \tilde{\epsilon}_t
\]

(4.15)

where \( \tilde{\epsilon}_t \) is uncorrelated with all regressors in (4.13), \( \tilde{\mu} \) and \( \delta_{4s} \) are scaler parameters, \( \delta_1 \) is an \( (n-1) \times 1 \) vector parameter, and \( \delta_{2s} \) and \( \delta_{3s} \) are \( (n \times 1) \) vector parameters. For notation simplification, we rewrite (4.13) as:

\[
y_{1t} = \theta' w_t + f(y_{t-s}; \psi) + \tilde{\epsilon}_t
\]

(4.16)

by defining

\[
f(y_{t-s}; \psi) = \tilde{\mu} + \delta'_1 y_{2t+1} + \sum_{s=-l+1}^{\max(2,l+1)} \delta'_s y_{1-s}
\]

\[
+ \sum_{s=-l}^{l} \delta'_{3a} y_{t-1-s} G(s_{t-s}; \gamma, c) + \sum_{s=-l}^{l} \delta'_{4a} G(s_{t-s}; \gamma, c),
\]

where \( \psi = (\tilde{\mu}, \delta'_1, \delta'_{2,-l+1}, \ldots, \delta'_{2,\max(2,l+1)}, \delta'_{3,-l}, \ldots, \delta'_{3,l}, \delta_{4,-l}, \ldots, \delta_{4,l}, \gamma, c)' \).

Assume the necessary starting values \( y_{-p}; y_{-p+1}; \ldots; y_{-1}; y_0 \) are fixed. Given a realization of \( y_t, t = 1, 2, \ldots, T \), the conditional nonlinear least square estimators of \( (\theta', \psi)' \) are carried out by solving the problem of minimizing:

\[
L(\theta, \psi) = \sum_{t=1}^{T} (y_{1t} - \theta' w_t - f(y_{t-s}; \psi))^2
\]

(4.17)
with respect to \((\theta', \psi')'\). The first-order partial derivatives of function \(L\) are

\[
\frac{\partial L}{\partial \theta} = \sum_{t=1}^{T} 2(y_{1t} - \theta' w_t - f(y_{t-s}; \psi)) w_t, \quad (4.18a)
\]

\[
\frac{\partial L}{\partial \psi} = \sum_{t=1}^{T} 2 \left( y_{1t} - \theta' w_t - f(y_{t-s}; \psi) \right) \frac{\partial f(y_{t-s}; \psi)}{\partial \psi}. \quad (4.18b)
\]

Under the null of \(\theta = 0\) in (4.19), (4.18a) and (4.18b) are reduced to

\[
\frac{\partial L}{\partial \theta} \bigg|_{H_0} = \sum_{t=1}^{T} 2 z_{1t} w_t \quad (4.19a)
\]

\[
\frac{\partial L}{\partial \psi} \bigg|_{H_0} = \sum_{t=1}^{T} 2 z_{1t} z_{2t} \quad (4.19b)
\]

where \(z_{1t} = y_{1t} - f(y_{t-s}; \psi)\) and \(z_{2t} = \partial f(y_{t-s}; \psi) / \partial \psi\). Given the consistent nonlinear least square estimators \(\hat{\psi}\) of the true parameter \(\psi\), (4.19a) and (4.19b) are calculated by \(\sum_{t=1}^{T} 2 \hat{z}_{1t} w_t\) and \(\sum_{t=1}^{T} 2 \hat{z}_{1t} \hat{z}_{2t}\) where \(\hat{z}_{1t} = y_{1t} - f(y_{t-s}; \hat{\psi})\) and \(\hat{z}_{2t} = \partial f(y_{t-s}; \hat{\psi}) / \partial \psi\). Our LM statistic for testing CNFs in (4.19) under the null (4.13) is calculated by the following formula:

\[
\hat{\sigma}^{-2} \left( \sum_{t=1}^{T} \hat{z}_{1t} w_t \right) \left( \sum_{t=1}^{T} w_t w_t' - \sum_{t=1}^{T} w_t \hat{z}_{2t} \sum_{t=1}^{T} \hat{z}_{2t} \hat{z}_{2t}' \left( \sum_{t=1}^{T} \hat{z}_{2t} \hat{z}_{2t}' \right)^{-1} \sum_{t=1}^{T} \hat{z}_{2t} w_t' \right)^{-1} \left( \sum_{t=1}^{T} \hat{z}_{1t} w_t' \right)
\]

where \(\hat{\sigma}^2 = \sum_{t=1}^{T} \hat{z}_{1t}^2 / T\) is a consistent estimator of \(\sigma^2\) and \(\hat{z}_{1t}\) and \(\hat{z}_{2t}\) are defined when the form of transition function \(G(s_t; \gamma, c)\) in (4.10) is specified. Taking \(G(s_t; \gamma, c) = (1 + e^{-\gamma(s_t-c)})^{-1}\) as an example, \(\hat{z}_{1t}\) is easily obtained and we further have

\[
\frac{\partial G(s_t; \hat{\gamma}, \hat{c})}{\partial \gamma} = \hat{g}_\gamma(t) = (1 + e^{-\hat{\gamma}(s_t-\hat{c})})^{-2} (s_t - \hat{c}) e^{-\hat{\gamma}(s_t-\hat{c})}, \quad (4.21a)
\]

\[
\frac{\partial G(s_t; \hat{\gamma}, \hat{c})}{\partial c} = \hat{g}_c(t) = -\hat{\gamma} (1 + e^{-\hat{\gamma}(s_t-\hat{c})})^{-2} e^{-\hat{\gamma}(s_t-\hat{c})}. \quad (4.21b)
\]
It follows that the estimated vector $\hat{z}_{2t}$ in this case is

$$\hat{z}_{2t} = \frac{\partial f(y_{t-s}; \hat{\psi})}{\partial \hat{\psi}} = (1, y'_{2t+l}, y'_{t-\text{max}(2l+1)}, \ldots, y'_{t+l-1},$$

$$y'_{t-1-l}G(s_{t-l}; \hat{\gamma}, \hat{\delta}), \ldots, y'_{t+l}G(s_{t+l}; \hat{\gamma}, \hat{\delta}),$$

$$\sum_{s=-l}^{l} (\hat{\delta}'_{3s}y_{t-1-s} + \hat{\delta}'_{4s}) \hat{g}_{\theta}(t-s),$$

$$\sum_{s=-l}^{l} (\hat{\delta}'_{3s}y_{t-1-s} + \hat{\delta}'_{4s}) \hat{g}_{c}(t-s))'.$$  \quad (4.22)

An $F$ form of our $LM$ test is recommended in practice in (4.23) because its empirical size is usually close to the nominal one, and the power is good when the sample size is small. Usually, the $F$ form is preferable to $LM$ statistics in small sample in the literature, for example, Eitrheim and Teräsvirta (1996) and Brännäs et al. (1998). The $F$ form of our $LM$ test can be performed as follows:

(i) Estimate the parameter $\psi$ in model (4.16) by nonlinear least square estimation under the assumption of $\theta = 0$ and compute the residual sum of squares $SSR_0 = \sum_{t=1}^{T} \hat{z}_{1t}^2$;

(ii) Regress $\hat{z}_{1t}$ on $w_t$ and $\hat{z}_{2t}$ and compute the residual sum of squares $SSR$;

(iii) Calculate the test statistic

$$F_{LM} = \frac{(SSR_0 - SSR)/(2n^2 + n)}{SSR/(T - 2n^2 - n - k)}$$

where $k$ is the dimension of vector $\hat{z}_{2t}$.

When $T$ goes to infinity, $(2n^2 + n)F_{LM}$ has the same asymptotic distribution as the $LM$ statistic.
5. Asymptotic distribution of the LM test

In this section, the asymptotic distribution of our LM test in (4.20) is derived under the null of CNFs. The following assumptions are required.

**Assumption 1.** In (4.16), \( \{ \tilde{e}_t \} \) is an i.i.d sequence with mean zero and variance \( \sigma^2 \), and \( \tilde{e}_t \) is uncorrelated with all leads and lags of \( \epsilon_t \).

**Assumption 2.** \( y_t \) in (2.1) is an \((n \times 1)\) vector stationary process in that

1. Given any value of the transition function \( G \) between zero and one,
   \[
   \left| I_n - \sum_{k=1}^{p} (\Phi_k + \Gamma_k G) z_k \right| \neq 0 \text{ if } |z| \leq 1;
   \]

2. The second-order moments of \( y_t \) depend on \( t \) defined by
   \[
   E(y_t y'_t) = \begin{pmatrix}
   E(y_{1t}y_{1t}) & E(y_{1t}y'_{2t}) \\
   E(y_{2t}y'_{1t}) & E(y_{2t}y'_{2t})
   \end{pmatrix} = \Omega_t = \begin{pmatrix}
   \Lambda_{11t} & \Lambda_{12t} \\
   \Lambda_{21t} & \Lambda_{22t}
   \end{pmatrix}
   \]
   which is a positive definite matrix with
   \[
   \frac{1}{T} \sum_{t=1}^{T} \Omega_t \rightarrow \Omega = \begin{pmatrix}
   \Lambda_{11} & \Lambda_{12} \\
   \Lambda_{21} & \Lambda_{22}
   \end{pmatrix}
   \]
   as \( T \rightarrow \infty \), in which \( \Omega \) is a positive definite matrix;

3. \( T^{-1} \sum_{t=1}^{T} y_t y'_t \xrightarrow{p} \Omega \) as \( T \rightarrow \infty \).

4. The process \( y_t \) is ergodic for the fourth-order moments.

Given \( G(s_t; \gamma, \mathbf{c}) \) in (2.3), \( \theta' \mathbf{w}_t + f(y_{t-s}; \psi) \) is continuously differentiable twice with respect to \( (\theta', \psi')' \) in some neighborhood of \( (\theta_0', \psi_0')' \) under the null. Assumption 2 implies the regularity conditions in Klimko and Nelson (1978, page 634) holds. Hence, the nonlinear least square estimator of \( (\theta', \psi')' \) in (4.16) is
consistent according to Theorem 3.1 in [Klinek and Nelson (1978)]. Under Assumptions [H],[2], the asymptotic distribution of

\[ T^{-1/2} \left( \sum_{t=1}^{T} z_{1t} z_{2t} + \sum_{t=1}^{T} \sum_{i=1}^{n} w_{i} z_{1t} \right) \]

is normality as shown in Lemma in Appendix [A]. It follows that our LM test in (4.20) has an asymptotic Chi-square distribution with \((2n^2 + n)\) degrees of freedom as \(T \to \infty\). See Theorem in Appendix [A].

6. Simulation

6.1 Critical values of non-standard unit root tests

The Monte Carlo experiments in this subsection study the finite-sample distributions of the non-standard Dickey-Fuller \(F\)-type and \(t\)-type tests illustrated in Section 4.1. The unit root tests in a nonlinear framework are not standard. Under the null of unit root in an STAR model, the simulation is based on the following data-generating process (DGP):

\[ y_t = y_{t-1} + v_t \]

where \(v_t \sim n.i.d(0,1)\) with desired sample sizes \(T = 50, 100, 200, 500, 1000\). The estimation is based on an auxiliary regression (4.2) with \(s_t = \Delta y_{t-1}\) or (4.4) with \(s_t = y_{t-1}\) and then we perform the Dickey-Fuller \(F\)-type and \(t\)-type tests in Section 4.1. In each experiment, \(T + 500\) observations were generated and the first 500 observations were discarded to minimize the initial effects. 100,000 replications were performed for each experiment. The finite-sample distributions of the tests are given in Table 1 when \(s_t = \Delta y_{t-1}\) and Table 2 when \(s_t = y_{t-1}\).

6.2 Finite-sample properties of the \(F_{LM}\) test

The Monte Carlo experiments in this subsection study the finite-sample properties of the \(F\) form of our \(LM\) test in Section 4.2. The simulations are designed for \(n =\)
2, as there are two variables in the system $y_t$ for our application in consumption and income data. The desired sample sizes are $T = 100, 200, 500, 1000$. We discarded the first 500 out of $T + 500$ observations to minimize the initial effects from the simulated time series data in each experiment as we did in Section 5.2, and each experiment was performed 10,000 times.

The size study is based on the DGP under the null hypothesis of existing CNFs such that:

$$
\begin{pmatrix}
  y_{1t} \\
  y_{2t}
\end{pmatrix} = \begin{pmatrix}
  0.4 \\
  0.2
\end{pmatrix} + \begin{pmatrix}
  0.3 & 0.05 \\
  0.2 & -0.1
\end{pmatrix} \begin{pmatrix}
  y_{1t-1} \\
  y_{2t-1}
\end{pmatrix} + \begin{pmatrix}
  2 \\
  1
\end{pmatrix} \begin{pmatrix}
  0.05 \\
  -0.18
\end{pmatrix} \begin{pmatrix}
  y_{1t-1} \\
  y_{2t-1}
\end{pmatrix} G(s_t; \gamma, c) + \begin{pmatrix}
  v_{1t} \\
  v_{2t}
\end{pmatrix}
$$

where $G(s_t; \gamma, c) = (1 + exp \{-5(y_{2t-1} - 0.5)\})^{-1}$ and $(v_{1t}, v_{2t})'$ are simulated from $n.i.d(0, \Sigma)$ with

$$
\Sigma = \begin{pmatrix}
  1 & 0.5 \\
  0.5 & 1
\end{pmatrix}.
$$

The estimation of the parameter $\psi$ is from (12.10) under the null of $\theta = 0$. In this case, $f(y_{t-s}; \psi)$ has a form

$$
f(y_{t-s}; \psi) = \tilde{\mu} + \delta_{11} y_{2t} + \delta_{21} y_{t-1} + \delta_{22} y_{t-2} + (\delta'_{30} y_{t-1} + \delta_{30}) G(s_t; \gamma, c).
$$

To calculate our $F_{LM}$ statistic, $w_t$ is given in detail such that:

$$
w_t = (y_{1t-1}^2, y_{1t-1} y_{2t-1}, y_{2t-1}^2, y_{1t-2}^2, y_{1t-2} y_{2t-2}, y_{2t-2}^2, y_{1t-3} y_{1t-2}, y_{1t-1} y_{2t-2}, y_{2t-1} y_{1t-2}, y_{2t-1} y_{2t-2})'.
$$

The DGP under the alternative for the power study follows a VSTAR model (without CNFs) such that

$$
\begin{pmatrix}
  y_{1t} \\
  y_{2t}
\end{pmatrix} = \begin{pmatrix}
  0.4 \\
  0.2
\end{pmatrix} + \begin{pmatrix}
  0.3 & 0.05 \\
  0.2 & -0.1
\end{pmatrix} \begin{pmatrix}
  y_{1t-1} \\
  y_{2t-1}
\end{pmatrix} + \begin{pmatrix}
  0.1 \\
  -0.16
\end{pmatrix} + \begin{pmatrix}
  0.24 & -0.18 \\
  0.12 & -0.36
\end{pmatrix} \begin{pmatrix}
  y_{1t-1} \\
  y_{2t-1}
\end{pmatrix} G(s_t; \gamma, c) + \begin{pmatrix}
  v_{1t} \\
  v_{2t}
\end{pmatrix}
$$
Table 1: Critical values for our Dickey-Fuller $F$-type and $t$-type tests under the null \((4.3)\) with $s_t = \Delta y_{t-1}$.

<table>
<thead>
<tr>
<th>$s_t$</th>
<th>$T$</th>
<th>99%</th>
<th>97.5%</th>
<th>95%</th>
<th>90%</th>
<th>10%</th>
<th>5%</th>
<th>2.5%</th>
<th>1%</th>
</tr>
</thead>
<tbody>
<tr>
<td>\Delta y_{t-1}</td>
<td>50</td>
<td>0.2791</td>
<td>0.3616</td>
<td>0.4517</td>
<td>0.5865</td>
<td>2.7163</td>
<td>3.2606</td>
<td>3.8141</td>
<td>4.5300</td>
</tr>
<tr>
<td></td>
<td>100</td>
<td>0.2870</td>
<td>0.3734</td>
<td>0.4691</td>
<td>0.6006</td>
<td>2.6671</td>
<td>3.1710</td>
<td>3.6714</td>
<td>4.3042</td>
</tr>
<tr>
<td></td>
<td>200</td>
<td>0.2895</td>
<td>0.3794</td>
<td>0.4736</td>
<td>0.6106</td>
<td>2.6423</td>
<td>3.1243</td>
<td>3.5941</td>
<td>4.2172</td>
</tr>
<tr>
<td></td>
<td>500</td>
<td>0.2902</td>
<td>0.3755</td>
<td>0.4732</td>
<td>0.6077</td>
<td>2.6139</td>
<td>3.0910</td>
<td>3.5613</td>
<td>4.1128</td>
</tr>
<tr>
<td></td>
<td>1000</td>
<td>0.2893</td>
<td>0.3806</td>
<td>0.4774</td>
<td>0.6142</td>
<td>2.6284</td>
<td>3.1001</td>
<td>3.5578</td>
<td>4.1509</td>
</tr>
</tbody>
</table>

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

<table>
<thead>
<tr>
<th>$s_t$</th>
<th>$T$</th>
<th>-3.5351</th>
<th>-3.1853</th>
<th>-2.8986</th>
<th>-2.5813</th>
<th>-0.3758</th>
<th>-0.0057</th>
<th>0.3228</th>
<th>0.7151</th>
</tr>
</thead>
<tbody>
<tr>
<td>\Delta y_{t-1}</td>
<td>50</td>
<td>-3.4738</td>
<td>-3.1449</td>
<td>-2.8786</td>
<td>-2.5716</td>
<td>-0.4096</td>
<td>-0.0491</td>
<td>0.2749</td>
<td>0.6591</td>
</tr>
<tr>
<td></td>
<td>100</td>
<td>-3.4593</td>
<td>-3.1346</td>
<td>-2.8736</td>
<td>-2.5685</td>
<td>-0.4244</td>
<td>-0.0562</td>
<td>0.2595</td>
<td>0.6399</td>
</tr>
<tr>
<td></td>
<td>200</td>
<td>-3.4273</td>
<td>-3.1274</td>
<td>-2.8660</td>
<td>-2.5707</td>
<td>-0.4381</td>
<td>-0.0718</td>
<td>0.2564</td>
<td>0.6242</td>
</tr>
<tr>
<td></td>
<td>500</td>
<td>-3.4251</td>
<td>-3.1117</td>
<td>-2.8523</td>
<td>-2.5593</td>
<td>-0.4425</td>
<td>-0.0778</td>
<td>0.2507</td>
<td>0.6198</td>
</tr>
</tbody>
</table>

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

Table 2: Critical values for our Dickey-Fuller $F$-type and $t$-type tests under the null \((4.5)\) with $s_t = y_{t-1}$.

<table>
<thead>
<tr>
<th>$s_t$</th>
<th>$T$</th>
<th>99%</th>
<th>97.5%</th>
<th>95%</th>
<th>90%</th>
<th>10%</th>
<th>5%</th>
<th>2.5%</th>
<th>1%</th>
</tr>
</thead>
<tbody>
<tr>
<td>(y_{t-1})</td>
<td>50</td>
<td>0.2638</td>
<td>0.3437</td>
<td>0.4281</td>
<td>0.5606</td>
<td>3.0599</td>
<td>3.7372</td>
<td>4.4263</td>
<td>5.3806</td>
</tr>
<tr>
<td></td>
<td>100</td>
<td>0.2654</td>
<td>0.3467</td>
<td>0.4319</td>
<td>0.5619</td>
<td>2.9649</td>
<td>3.5937</td>
<td>4.2322</td>
<td>5.0352</td>
</tr>
<tr>
<td></td>
<td>200</td>
<td>0.2725</td>
<td>0.3500</td>
<td>0.4348</td>
<td>0.5639</td>
<td>2.9164</td>
<td>3.5180</td>
<td>4.1144</td>
<td>4.8323</td>
</tr>
<tr>
<td></td>
<td>500</td>
<td>0.2735</td>
<td>0.3511</td>
<td>0.4353</td>
<td>0.5639</td>
<td>2.9108</td>
<td>3.5004</td>
<td>4.0708</td>
<td>4.7926</td>
</tr>
<tr>
<td></td>
<td>1000</td>
<td>0.2704</td>
<td>0.3501</td>
<td>0.4343</td>
<td>0.5642</td>
<td>2.8923</td>
<td>3.4874</td>
<td>4.0520</td>
<td>4.7917</td>
</tr>
</tbody>
</table>

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

<table>
<thead>
<tr>
<th>$s_t$</th>
<th>$T$</th>
<th>-2.7063</th>
<th>-2.3056</th>
<th>-1.9857</th>
<th>-1.6188</th>
<th>0.8153</th>
<th>1.1617</th>
<th>1.4753</th>
<th>1.8644</th>
</tr>
</thead>
<tbody>
<tr>
<td>(y_{t-1})</td>
<td>50</td>
<td>-2.6606</td>
<td>-2.2822</td>
<td>-1.9621</td>
<td>-1.6027</td>
<td>0.7821</td>
<td>1.1224</td>
<td>1.4300</td>
<td>1.7939</td>
</tr>
<tr>
<td></td>
<td>100</td>
<td>-2.6591</td>
<td>-2.2723</td>
<td>-1.9623</td>
<td>-1.6025</td>
<td>0.7708</td>
<td>1.1163</td>
<td>1.4164</td>
<td>1.7675</td>
</tr>
<tr>
<td></td>
<td>200</td>
<td>-2.6414</td>
<td>-2.2629</td>
<td>-1.9461</td>
<td>-1.5908</td>
<td>0.7603</td>
<td>1.0917</td>
<td>1.3794</td>
<td>1.7126</td>
</tr>
<tr>
<td></td>
<td>500</td>
<td>-2.6394</td>
<td>-2.2716</td>
<td>-1.9490</td>
<td>-1.5943</td>
<td>0.7502</td>
<td>1.0865</td>
<td>1.3687</td>
<td>1.7021</td>
</tr>
</tbody>
</table>

The probability shown at the head of the column is the area in the left-hand tail.
### Table 3: Empirical size and power of the $F_{LM}$ test.

<table>
<thead>
<tr>
<th>$T$</th>
<th>Size</th>
<th>Size-adjusted Power</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>given significant levels</td>
<td>given significant levels</td>
</tr>
<tr>
<td>100</td>
<td>0.3536 0.2053 0.1179 0.0324</td>
<td>0.1394 0.0652 0.0272 0.0043</td>
</tr>
<tr>
<td>200</td>
<td>0.3471 0.2055 0.1202 0.0305</td>
<td>0.1411 0.0649 0.0292 0.0051</td>
</tr>
<tr>
<td>500</td>
<td>0.2896 0.1605 0.0892 0.0226</td>
<td>0.1743 0.0889 0.0428 0.0077</td>
</tr>
<tr>
<td>1000</td>
<td>0.2864 0.1556 0.0851 0.0189</td>
<td>0.2815 0.1698 0.1066 0.0368</td>
</tr>
</tbody>
</table>

### Table 4: Empirical size and power of the $F_{LM}$ test.

<table>
<thead>
<tr>
<th>$T$</th>
<th>Size</th>
<th>Power</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>given significant levels</td>
<td>given significant levels</td>
</tr>
<tr>
<td>100</td>
<td>0.3536 0.2053 0.1179 0.0324</td>
<td>0.2688 0.1458 0.0777 0.0161</td>
</tr>
<tr>
<td>200</td>
<td>0.3471 0.2055 0.1202 0.0305</td>
<td>0.2703 0.1469 0.0814 0.0180</td>
</tr>
<tr>
<td>500</td>
<td>0.2896 0.1605 0.0892 0.0226</td>
<td>0.2603 0.1394 0.0790 0.0179</td>
</tr>
<tr>
<td>1000</td>
<td>0.2864 0.1556 0.0851 0.0189</td>
<td>0.3618 0.2375 0.1506 0.0564</td>
</tr>
</tbody>
</table>

### Table 5: Empirical size and power of the $F_{LM}$ test, DGP2.

<table>
<thead>
<tr>
<th>$T$</th>
<th>Size</th>
<th>Size-adjusted Power</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>given significant levels</td>
<td>given significant levels</td>
</tr>
<tr>
<td>100</td>
<td>0.3536 0.2053 0.1179 0.0324</td>
<td>0.1878 0.0906 0.0427 0.0080</td>
</tr>
<tr>
<td>200</td>
<td>0.3471 0.2055 0.1202 0.0305</td>
<td>0.2133 0.1095 0.0566 0.0109</td>
</tr>
<tr>
<td>500</td>
<td>0.2896 0.1605 0.0892 0.0226</td>
<td>0.3936 0.2463 0.1445 0.0410</td>
</tr>
<tr>
<td>1000</td>
<td>0.2864 0.1556 0.0851 0.0189</td>
<td>0.5296 0.3628 0.2474 0.0976</td>
</tr>
<tr>
<td>2000</td>
<td>0.2940 0.1684 0.0906 0.0207</td>
<td>0.7520 0.6190 0.4910 0.2500</td>
</tr>
</tbody>
</table>
where similarly \( G(s_t; \gamma, c) = (1 + \exp\{-5(y_{2t-1} - 0.5)\})^{-1} \) and \((v_{1t}, v_{2t})'\) are simulated from \( n.i.d(0, \Sigma) \). This method for performing our \( F_{LM} \) test is similar to the method in the previous size study. Further, another DGP for power experiments is included, which is actually an extension of the VSTAR model with different nonlinearity for each series, indicating that CNFs is not likely to occur in the system. In the experiments, we set \( s_t^* = y_{1t-1} \) and \( s_t = y_{2t-1} \).

\[
\begin{pmatrix}
y_{1t} \\
y_{2t}
\end{pmatrix} = 
\begin{pmatrix}
0.4 \\
0.2
\end{pmatrix} + 
\begin{pmatrix}
0.3 & 0.05 \\
0.2 & -0.1
\end{pmatrix}
\begin{pmatrix}
y_{1t-1} \\
y_{2t-1}
\end{pmatrix} + 
\begin{pmatrix}
0.1G(s_t^*; \gamma, c) \\
-0.16G(s_t; \gamma, c)
\end{pmatrix}
+ 
\begin{pmatrix}
0.24 & -0.18 \\
0.12 & -0.36
\end{pmatrix}
\begin{pmatrix}
y_{1t-1}G(s_t^*; \gamma, c) \\
y_{2t-1}G(s_t; \gamma, c)
\end{pmatrix} + 
\begin{pmatrix}
v_{1t} \\
v_{2t}
\end{pmatrix}
\]  

(6.1)

We compare the finite-sample distributions of our \( F_{LM} \) test to the \( F \)-distribution and present the estimated size and (size-adjusted) power in Tables \ref{tab:1}–\ref{tab:4}. Our \( F_{LM} \) test is oversized, but the empirical size decreases as sample size increases. The power is acceptable before size-corrected. The size-adjusted power increases as the sample size increases, although it is not better. This issue could be due to the numerical problem; the residual vector may not be orthogonal to the gradient matrix. In the simulations, it is difficult to make good nonlinear estimators for each replication, when only one set of initial values for the parameters is used across all replications, as is typical. Sometimes, the iterations cannot converge with inappropriate initial values. However, when the system is less likely to have CNFs, for example, in (6.1), the power is much better.

In the cointegration literature, Elliott et al. (2005) stated that unknown cointegration vector produces the loss of power. Not imposing the cointegrating vector in the testing procedure would make the test to lose power. With this idea in mind, we further examine the finite-sample properties, assuming that the combination vector \( \alpha \) for CNFs in our case is known. This makes model (6.10) simpler because it is not required to correct the correlations between \( y_{2t} \) and \( e_t \). Therefore, \( f(y_{t-s}; \psi) \) has a linear form. In this specific experiment, \( f(y_{t-s}; \psi) \) has the form
Table 6: Empirical size and power of the $F_{LM}$ test, DGP2.

<table>
<thead>
<tr>
<th>$T$</th>
<th>Size</th>
<th>Power</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>given significant levels</td>
<td>given significant levels</td>
</tr>
<tr>
<td>100</td>
<td>0.3536 0.2053 0.1179 0.0324</td>
<td>0.3350 0.1932 0.1079 0.0266</td>
</tr>
<tr>
<td>200</td>
<td>0.3471 0.2055 0.1202 0.0305</td>
<td>0.3809 0.2208 0.1287 0.0345</td>
</tr>
<tr>
<td>500</td>
<td>0.2896 0.1605 0.0892 0.0226</td>
<td>0.5094 0.3401 0.2230 0.0781</td>
</tr>
<tr>
<td>1000</td>
<td>0.2864 0.1556 0.0851 0.0189</td>
<td>0.6372 0.4674 0.3304 0.1424</td>
</tr>
<tr>
<td>2000</td>
<td>0.2940 0.1684 0.0906 0.0207</td>
<td>0.8540 0.7140 0.6050 0.3440</td>
</tr>
</tbody>
</table>

Table 7: Empirical size and power of the $F_{LM}$ test when $\alpha$ is known.

<table>
<thead>
<tr>
<th>$T$</th>
<th>Size</th>
<th>Size-adjusted Power</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>given significant levels</td>
<td>given significant levels</td>
</tr>
<tr>
<td>100</td>
<td>0.1830 0.0877 0.0423 0.0085</td>
<td>0.2998 0.1753 0.0994 0.0266</td>
</tr>
<tr>
<td>200</td>
<td>0.1954 0.0972 0.0473 0.0088</td>
<td>0.3968 0.2502 0.1581 0.0512</td>
</tr>
<tr>
<td>500</td>
<td>0.1986 0.1013 0.0492 0.0088</td>
<td>0.6716 0.5172 0.3986 0.2054</td>
</tr>
<tr>
<td>1000</td>
<td>0.1976 0.0991 0.0481 0.0092</td>
<td>0.9158 0.8410 0.7543 0.5340</td>
</tr>
</tbody>
</table>

as follows ($\alpha^*$ is known),

$$f(y_t-s; \Psi) = \mu_t^* + \alpha^* y_{2t} + \sum_{s=1}^{2} \phi_{1s}^* y_{t-s}.$$  

The empirical size and (size-adjusted) power of this simple case in Tables 7-8 are substantially better than the one in Tables 3-4.

Table 8: Empirical size and power of the $F_{LM}$ test when $\alpha$ is known.

<table>
<thead>
<tr>
<th>$T$</th>
<th>Size</th>
<th>Power</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>given significant levels</td>
<td>given significant levels</td>
</tr>
<tr>
<td>100</td>
<td>0.1830 0.0877 0.0423 0.0085</td>
<td>0.2791 0.1541 0.0886 0.0218</td>
</tr>
<tr>
<td>200</td>
<td>0.1954 0.0972 0.0473 0.0088</td>
<td>0.3752 0.2232 0.1274 0.0325</td>
</tr>
<tr>
<td>500</td>
<td>0.1986 0.1013 0.0492 0.0088</td>
<td>0.6485 0.4824 0.3482 0.1372</td>
</tr>
<tr>
<td>1000</td>
<td>0.1976 0.0991 0.0481 0.0092</td>
<td>0.0931 0.8131 0.6992 0.4250</td>
</tr>
</tbody>
</table>
7. Application

7.1 Data

The data set, monthly personal consumption expenditures and personal disposable income in United States dating from 1985:1 until 2011:11 (323 observations), is download from U.S. Bureau of Economic Analysis to analyze CNFs between consumption and income. Both of them are seasonally adjusted at annual rates. The natural logs of the raw data were taken and then multiplied by 100. Figure 3 presents the percentage growth ($\Delta y_{1t}$ and $\Delta y_{2t}$). We use the observations from 1985:1 to 2006:12 (264 observations) for modeling and save the remaining ones for out-of-sample forecasting (future work).

The data in Figure 3 have obvious fluctuations in the late 1980s, the early 1990s and 2000s. A well-known credit crunch of U.S. savings and loans occurred during 1980s and was a major factor that led to the U.S. recession of 1990-1991. Another well-known financial crisis was the crash of the dot-com bubble in 2000-2001. In late 2008 and 2009, the subprime mortgage crisis and the bursting of other real estate bubbles around the world has led to recession in the U.S. and a number of other countries, and there is a now-deflating United States housing bubble. Moreover, the international financial crises, when many Latin American countries defaulted on their debt in the early 1980s, may cause the large data break in the middle of 1980s. The data behaviors suggest that a nonlinear model could be a good consideration.

7.2 Testing procedure

Applying our procedure to consumption and income growth, our unit root tests in Section 4.1 are imposed first. Table 9 presents the results of Dickey-Fuller $t$ and $F$ statistics for those unit root tests. From Table 9, both $\Delta y_{1t}$ and $\Delta y_{2t}$ likely perform as nonlinearly stabilized processes.

In the following analysis, we are interested in whether $\Delta y_{1t}$ and $\Delta y_{2t}$ have CNFs according to the testing procedure in Section 4.2. The test regression follows (4.16).
Figure 4: Percentage change in personal consumption expenditures (upper) and disposable income (lower), United States, billions of 2005 dollars, monthly, 1985:1-2011:11.

Table 9: The unit root tests.

<table>
<thead>
<tr>
<th>time series</th>
<th>$s_t$</th>
<th>$t$</th>
<th>$F$</th>
<th>$t$</th>
<th>$F$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\Delta y_{1t}$</td>
<td>$\Delta^2 y_{1t-1}$</td>
<td>-2.7476</td>
<td>4.3017</td>
<td>-2.87</td>
<td>3.12</td>
</tr>
<tr>
<td>$\Delta y_{1t}$</td>
<td>$\Delta^2 y_{2t-1}$</td>
<td>-2.6744</td>
<td>4.8862</td>
<td>-1.96</td>
<td>3.52</td>
</tr>
<tr>
<td>$\Delta y_{2t}$</td>
<td>$\Delta^2 y_{2t-1}$</td>
<td>-5.7069</td>
<td>18.029</td>
<td>-2.87</td>
<td>3.12</td>
</tr>
<tr>
<td>$\Delta y_{2t}$</td>
<td>$\Delta^2 y_{2t-1}$</td>
<td>-5.6537</td>
<td>17.810</td>
<td>-1.96</td>
<td>3.52</td>
</tr>
</tbody>
</table>
choosing $p = 1$ and $l = 2$. Based on the nonlinear least square estimation, the $F_{LM}$ test statistic in (123) is calculated. $F_{LM} = 1.7526$ and the critical value is 2.11. The data lead to a non-rejection of the null of CNFs in our test. It is reasonable to believe that there are CNFs in the system of consumption and income.

Table 10: Choose transition variable.

<table>
<thead>
<tr>
<th>data</th>
<th>test</th>
<th>$x_{1t}$</th>
<th>$x_{1,t-2}$</th>
<th>$x_{1,t-3}$</th>
<th>$x_{1,t-4}$</th>
<th>$x_{1,t-5}$</th>
<th>$x_{1,t-6}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$x_{1t}$</td>
<td>$t$</td>
<td>-2.6744</td>
<td>-2.7522</td>
<td>-3.1715</td>
<td>-2.3702</td>
<td>-2.4773</td>
<td>-2.5426</td>
</tr>
<tr>
<td>$x_{2t}$</td>
<td>$t$</td>
<td>-4.6066</td>
<td>-5.3846</td>
<td>-4.5765</td>
<td>-5.0576</td>
<td>-4.7381</td>
<td>-5.2491</td>
</tr>
<tr>
<td>$x_{1t}$</td>
<td>$F$</td>
<td>4.8862</td>
<td>3.7963</td>
<td>8.7313</td>
<td>5.9177</td>
<td>6.1487</td>
<td>3.8504</td>
</tr>
<tr>
<td>$x_{2t}$</td>
<td>$F$</td>
<td>19.504</td>
<td>14.806</td>
<td>15.725</td>
<td>15.762</td>
<td>15.604</td>
<td>14.271</td>
</tr>
<tr>
<td>$x_{1t}$</td>
<td>$t$</td>
<td>3.1982</td>
<td>2.5386</td>
<td>2.4964</td>
<td>2.0108</td>
<td>3.4666</td>
<td>2.9038</td>
</tr>
<tr>
<td>$x_{2t}$</td>
<td>$t$</td>
<td>5.2798</td>
<td>4.2107</td>
<td>5.3020</td>
<td>5.3532</td>
<td>5.3216</td>
<td>4.4308</td>
</tr>
<tr>
<td>$x_{1t}$</td>
<td>$F$</td>
<td>6.1462</td>
<td>3.7753</td>
<td>5.1453</td>
<td>6.4307</td>
<td>9.1321</td>
<td>6.0254</td>
</tr>
<tr>
<td>$x_{1t}$</td>
<td>$t$</td>
<td>2.8004</td>
<td>2.5715</td>
<td>2.6882</td>
<td>2.7418</td>
<td>2.8424</td>
<td>2.6694</td>
</tr>
<tr>
<td>$x_{2t}$</td>
<td>$t$</td>
<td>17.810</td>
<td>16.894</td>
<td>14.196</td>
<td>14.506</td>
<td>15.249</td>
<td>15.322</td>
</tr>
<tr>
<td>$x_{1t}$</td>
<td>$F$</td>
<td>4.1548</td>
<td>4.2530</td>
<td>3.7920</td>
<td>3.9519</td>
<td>4.0765</td>
<td>4.1619</td>
</tr>
<tr>
<td>$x_{2t}$</td>
<td>$F$</td>
<td>4.4869</td>
<td>3.7515</td>
<td>3.9776</td>
<td>4.8118</td>
<td>4.453</td>
<td>3.9163</td>
</tr>
<tr>
<td>$x_{1t}$</td>
<td>$t$</td>
<td>4.6110</td>
<td>4.5739</td>
<td>5.0931</td>
<td>5.4967</td>
<td>5.2708</td>
<td>5.4674</td>
</tr>
<tr>
<td>$x_{2t}$</td>
<td>$t$</td>
<td>17.485</td>
<td>17.585</td>
<td>16.843</td>
<td>15.108</td>
<td>14.160</td>
<td>16.565</td>
</tr>
</tbody>
</table>

*Critical values of Dickey-Fuller $t$ and $F$ tests at 5% significance level are -1.96 and 3.92, respectively.

7.3 Modeling

Based on the conclusion above, we want to model the first-differenced data $\Delta y_{1t}$ and $\Delta y_{2t}$ with a VSTAR model containing CNFs. Let $x_{it} = \Delta y_{it}$ denote the percentage change in consumption $(i = 1)$ and income $(i = 2)$. Before we perform estimation of a VSTAR model containing CNFs, the model needs to be specified by deciding how to choose the form of the ST function $G(s_t; \gamma, c)$. We take the
simplest one with \( k = 1 \). Then, to specify the transition variable \( s_t \) in an ST-type model, a technique in [Teräsvirta (1994)] can be applied by using the result of testing for nonlinearity such that the model which most significantly presents nonlinearity is the best choice. Following the principle, we found from testing the unit root against nonlinearity in Table 10 that \( x_{1,t-d} \) is better than \( x_{2,t-d} \) because \( s_t = x_{1,t-d} \) contributes more to the nonlinearity of \( x_{it} \) than \( s_t = x_{2,t-d} \). Further, \( s_t = x_{1,t-7} \) is recognized as a good consideration. \( s_t = \Delta x_{i,t-d} \) is not considered because \( \Delta x_{i,t-d} \) is not as meaningful when \( x_{it} \) is stationary. (However, it is reasonable to include it as an option for \( s_t \) under the null of the unit root in the tests.)

The following estimated model is suggested:

\[
\begin{pmatrix}
\hat{x}_{1t} \\
\hat{x}_{2t}
\end{pmatrix} = 
\begin{pmatrix}
0.3682 \\
0.3328
\end{pmatrix} + 
\begin{pmatrix}
-0.3778 & 0.0291 \\
0.0087 & -0.3228
\end{pmatrix} 
\begin{pmatrix}
x_{1t-1} \\
x_{2t-1}
\end{pmatrix} 
+ 
\begin{pmatrix}
-0.0378 \\
-0.0545
\end{pmatrix} 
(0.0493 + 0.0100x_{1t-1} - 0.0850x_{2t-1}) \cdot G(s_t; \hat{\gamma}, \hat{c})
\]

where \( G(s_t; \hat{\gamma}, \hat{c}) = (1 + \exp\{-47.5(x_{1t-7} - 0.98)/\hat{\sigma}_{x_{1t-7}}\})^{-1} \) and \( \hat{\sigma}_{x_{1t-7}} = 0.5146 \).

The Ljung-Box test is considered to examine the autocorrelation of the residual from the estimated model above. It is approximately distributed as Chi-square with degrees of freedom \( n^2(s - k) \) (see [Ljung and Box, 1978; Hosking, 1980]). For our data, \( n = 2, k = 1, \) and \( s = 60 \) (\( s \) is reasonably taken as \( T/4 \) following the literature, see for example [Juselius (2007)], Chapter 4.3.3), the degree of freedom is 236. The null of no autocorrelation cannot be rejected, and the residuals do not have serial correlation.

8. Conclusions

In this paper we propose a testing procedure to examine the existence of CNFs in a VSTAR model. A triangular representation for performing our LM test in the VSTAR model containing CNFs is developed. Using the triangular representation
avoids a regression approximation for testing CNFs, which has been involved in the existing literature, although nonlinear estimation has to be used as a consequence. Therefore, our test is easy to interpretate and the test properties are easily examined. The test under the null of existing CNFs in a VSTAR model has an asymptotic Chi-square distribution. As the model moves further from containing CNFs or when the combination vector is known, the test performs better in small samples. We illustrate our testing procedure by applying it to the U.S. consumption and income data, and we observe that the growth of consumption and income contain CNFs. Afterward, modeling and estimation are briefly discussed. To make this work more practicable, further development for more than one combination of removing nonlinearity, considering a matrix instead of a vector $\alpha$, should be considered.

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.

References


APPENDIX

A. Proofs

The following Lemma is obtained from the central limit theorem (CLT) for sequences of mixing random variables. See details of CLT in White (2001, Theorem 5.20).

Lemma. Consider (4.16) under the null (4.13). Under Assumptions 1 and 2, the first-order derivative of \( L \),

\[
T^{-\frac{1}{2}} \left( \sum_{t=1}^{T} z_{1t} w_t - \sum_{t=1}^{T} z_{1t} z_{2t} \right),
\]

has the multivariate normal distribution \( N(0, E(z_{1t}^2)V) \) in which \( \frac{1}{T} \sum_{t=1}^{T} V_t \overset{p}{\rightarrow} V \) as \( T \to \infty \) implied by Assumption 2, whereas

\[
V_t = \begin{pmatrix}
E(w_t w_t') & E(w_t z_{2t}') \\
E(z_{2t} w_t') & E(z_{2t} z_{2t}')
\end{pmatrix},
\]

and \( V_t \) and \( V \) are positive definite matrices.

Theorem. Consider (4.16) and (4.9b) under the null of CNFs. When Assumptions 1-2 hold, our LM test in (4.20) has an asymptotic Chi-squared distribution with \( (2n^2 + n) \) degrees of freedom as \( T \to \infty \).

Proof of Theorem. Consider (4.16) under the null (4.13). Under Assumption 2, we have stated in Section 4.2 that the nonlinear least square estimator of \( (\theta', \psi')' \) are consistent. Therefore, \( \hat{z}_{1t} \) and \( \hat{z}_{2t} \) are consistent estimators of \( z_{1t} \) and \( z_{2t} \) respectively. Then we have

\[
T^{-\frac{1}{2}} \left( \sum_{t=1}^{T} \hat{z}_{1t} w_t - \sum_{t=1}^{T} z_{1t} w_t \\
\sum_{t=1}^{T} \hat{z}_{1t} z_{2t} - \sum_{t=1}^{T} z_{1t} z_{2t} \right) \overset{p}{\rightarrow} 0
\]

(A.2)
as \( T \to \infty \) by consistency. Thus,

\[
T^{-\frac{1}{2}} \left( \sum_{t=1}^{T} \hat{z}_{1t} w_t \over \sum_{t=1}^{T} \hat{z}_{1t} \hat{z}_{2t} \right) \quad (A.3)
\]

has the same asymptotic distribution with

\[
T^{-\frac{1}{2}} \left( \sum_{t=1}^{T} z_{1t} w_t \over \sum_{t=1}^{T} z_{1t} z_{2t} \right) \xrightarrow{L} N(0, E(z_{1t}^2) \mathbf{V}) \quad (A.4)
\]

which is shown in Lemma. Then it follows from (A.4) that

\[
(\hat{\sigma}^2 \hat{\mathbf{V}})^{-1/2} T^{-1/2} \left( \sum_{t=1}^{T} \hat{z}_{1t} w_t \over \sum_{t=1}^{T} \hat{z}_{1t} \hat{z}_{2t} \right) \xrightarrow{L} N(0, \mathbf{I}) \quad (A.5)
\]

as \( T \to \infty \) in which \( \hat{\sigma}^2 = \sum_{t=1}^{T} \hat{z}_{1t}^2 / T \), \( \mathbf{I} \) is the identity matrix, and

\[
\hat{\mathbf{V}} = \begin{pmatrix}
\sum_{t=1}^{T} w_t w_t' / T & \sum_{t=1}^{T} w_t \hat{z}_{2t} / T \\
\sum_{t=1}^{T} \hat{z}_{2t} w_t' / T & \sum_{t=1}^{T} \hat{z}_{2t} \hat{z}_{2t} / T
\end{pmatrix} .
\]

Let \( m_1 = 2n^2 + n \) denote the dimension of vector \( \mathbf{w}_t \) and \( m_2 = k \) denote the dimension of vector \( \hat{\mathbf{z}}_{2t} \). The null hypothesis (A.3) could be rewritten as

\[
H_0 : \mathbf{R} \begin{pmatrix}
\theta \\
\psi
\end{pmatrix} = 0 \quad (A.6)
\]

in which \( \mathbf{R} = (\mathbf{I}_{m_1}, \mathbf{0}) \). Then

\[
\mathbf{R}(\hat{\sigma}^2 \hat{\mathbf{V}})^{-1/2} \left( \sum_{t=1}^{T} \hat{z}_{1t} w_t \over \sum_{t=1}^{T} \hat{z}_{1t} \hat{z}_{2t} \right) \xrightarrow{L} N(0, \mathbf{R I}_{m_1 + m_2} \mathbf{R}') \equiv N(0, \mathbf{I}_{m_1}) \quad (A.7)
\]
Table 11: Critical values for non-standard unit root tests under the null (B.1) with $s_t = \Delta y_{t-1}$.

<table>
<thead>
<tr>
<th>$s_t$</th>
<th>$T$</th>
<th>Probability that $F$ is greater than entry</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>99%</td>
</tr>
<tr>
<td>$\Delta y_{t-1}$</td>
<td>50</td>
<td>0.0415</td>
</tr>
<tr>
<td></td>
<td>100</td>
<td>0.0421</td>
</tr>
<tr>
<td></td>
<td>200</td>
<td>0.0417</td>
</tr>
<tr>
<td></td>
<td>500</td>
<td>0.0426</td>
</tr>
<tr>
<td></td>
<td>1000</td>
<td>0.0431</td>
</tr>
</tbody>
</table>

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

<table>
<thead>
<tr>
<th>$s_t$</th>
<th>$T$</th>
<th>Probability that $t$ is less than entry</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>99%</td>
</tr>
<tr>
<td>$\Delta y_{t-1}$</td>
<td>50</td>
<td>-2.6570</td>
</tr>
<tr>
<td></td>
<td>100</td>
<td>-2.6014</td>
</tr>
<tr>
<td></td>
<td>200</td>
<td>-2.5903</td>
</tr>
<tr>
<td></td>
<td>500</td>
<td>-2.5836</td>
</tr>
<tr>
<td></td>
<td>1000</td>
<td>-2.5545</td>
</tr>
</tbody>
</table>

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

As $T \to \infty$. Thus,

$$
\hat{\sigma}^{-2} \left( \sum_{t=1}^{T} \hat{z}_{1t} w_t' \sum_{t=1}^{T} \hat{z}_{1t} \hat{z}_{2t}' \right) R' \left( R(T \hat{V})^{-1} R' \right) R \left( \sum_{t=1}^{T} \hat{z}_{1t} w_t \right) \left( \sum_{t=1}^{T} \hat{z}_{1t} \hat{z}_{2t}' \right) \rightarrow \chi^2(m_1)
$$

(A.8)

as $T \to \infty$ in which

$$
R(T \hat{V})^{-1} R' = \left( \sum_{t=1}^{T} w_t w_t' - \sum_{t=1}^{T} w_t \hat{z}_{2t}' \left( \sum_{t=1}^{T} \hat{z}_{2t}' \hat{z}_{2t}' \right)^{-1} \sum_{t=1}^{T} \hat{z}_{2t} w_t' \right)^{-1}.
$$

B. Simulation of unit root tests ($\mu^* = 0$)

Similar to the simulation procedure in Section 6.1, we give the finite-sample distributions of Dickey-Fuller $F$-type and $t$-type tests of another case in which $\mu^* = 0$.
Table 12: Critical values for non-standard unit root tests under the null \((B.3)\) with \(s_t = y_{t-1}\).

<table>
<thead>
<tr>
<th>(s_t)</th>
<th>(T)</th>
<th>99%</th>
<th>97.5%</th>
<th>95%</th>
<th>90%</th>
<th>10%</th>
<th>5%</th>
<th>2.5%</th>
<th>1%</th>
</tr>
</thead>
<tbody>
<tr>
<td>(y_{t-1})</td>
<td>50</td>
<td>0.0589</td>
<td>0.1369</td>
<td>0.2450</td>
<td>0.4006</td>
<td>3.5804</td>
<td>4.4741</td>
<td>5.3629</td>
<td>6.5900</td>
</tr>
<tr>
<td></td>
<td>100</td>
<td>0.0610</td>
<td>0.1443</td>
<td>0.2510</td>
<td>0.4112</td>
<td>3.4751</td>
<td>4.3232</td>
<td>5.1667</td>
<td>6.2926</td>
</tr>
<tr>
<td></td>
<td>200</td>
<td>0.0601</td>
<td>0.1447</td>
<td>0.2550</td>
<td>0.4086</td>
<td>3.4880</td>
<td>4.3158</td>
<td>5.1192</td>
<td>6.2031</td>
</tr>
<tr>
<td></td>
<td>500</td>
<td>0.0623</td>
<td>0.1472</td>
<td>0.2563</td>
<td>0.4094</td>
<td>3.4468</td>
<td>4.2620</td>
<td>5.0649</td>
<td>6.0750</td>
</tr>
<tr>
<td></td>
<td>1000</td>
<td>0.0606</td>
<td>0.1456</td>
<td>0.2562</td>
<td>0.4137</td>
<td>3.4496</td>
<td>4.2618</td>
<td>5.0486</td>
<td>6.0007</td>
</tr>
</tbody>
</table>

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

<table>
<thead>
<tr>
<th>(s_t)</th>
<th>(T)</th>
<th>99%</th>
<th>97.5%</th>
<th>95%</th>
<th>90%</th>
<th>10%</th>
<th>5%</th>
<th>2.5%</th>
<th>1%</th>
</tr>
</thead>
<tbody>
<tr>
<td>(y_{t-1})</td>
<td>50</td>
<td>-2.3973</td>
<td>-2.0115</td>
<td>-1.6624</td>
<td>-1.2137</td>
<td>2.4042</td>
<td>2.7741</td>
<td>3.0839</td>
<td>3.4591</td>
</tr>
<tr>
<td></td>
<td>100</td>
<td>-2.3652</td>
<td>-2.0041</td>
<td>-1.6580</td>
<td>-1.2083</td>
<td>2.3872</td>
<td>2.7292</td>
<td>3.0361</td>
<td>3.3724</td>
</tr>
<tr>
<td></td>
<td>200</td>
<td>-2.3936</td>
<td>-2.0031</td>
<td>-1.6508</td>
<td>-1.2129</td>
<td>2.3871</td>
<td>2.7338</td>
<td>3.0213</td>
<td>3.3567</td>
</tr>
<tr>
<td></td>
<td>500</td>
<td>-2.3839</td>
<td>-1.9835</td>
<td>-1.6379</td>
<td>-1.2100</td>
<td>2.3811</td>
<td>2.7128</td>
<td>3.0059</td>
<td>3.3445</td>
</tr>
<tr>
<td></td>
<td>1000</td>
<td>-2.3833</td>
<td>-2.0029</td>
<td>-1.6512</td>
<td>-1.2085</td>
<td>2.3817</td>
<td>2.7167</td>
<td>3.0060</td>
<td>3.3290</td>
</tr>
</tbody>
</table>

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

in \((4.2)\) in Section 4.1. Test

\[ H_0 : y_t = y_{t-1} + v_t^* \]  \hspace{1cm} (B.1)

against

\[ H_1^{aux} : y_t = \phi_1^* y_{t-1} + \phi_2^* s_t y_{t-1} + \phi_3^* s_t + v_t^* \]  \hspace{1cm} (B.2)

when we consider \(s_t = \Delta y_{t-1}\) \((H_0 : \phi_1^* = 1, \phi_2^* = 0, \phi_3^* = 0)\). However, when we consider \(s_t = y_{t-1}\), we test

\[ H_0 : y_t = y_{t-1} + v_t^* \]  \hspace{1cm} (B.3)

against

\[ H_1^{aux} : y_t = \tilde{\phi}_1^* y_{t-1} + \phi_2^* s_t y_{t-1} + v_t^* \]  \hspace{1cm} (B.4)

\((H_0 : \tilde{\phi}_1^* = 1, \phi_2^* = 0)\). Then we consider \((B.2)\) with the null hypothesis of \(\phi_1^* = 1\) and \((B.3)\) with \(\tilde{\phi}_1^* = 1\). Critical values are in Table 11-12.

C. Figures
Figure 5: Smooth-transition function $G$ in $s_{1t} \sim N(0, 1)$ in panel (a) and $s_{2t} \sim N(0, 0.5)$ in panel (b), when $k = 2$ with different both $s_{jt}$ and $c_j = \{-1, 2\}$.

Figure 6: Smooth-transition function $G$ in $s_{1t} \sim N(0, 1)$ in panel (a) and $s_{2t} \sim N(0, 0.5)$ in panel (b), when $k = 2$ with different both $s_{jt}$ but same $c_j = 1$. 
CNFs in VSTAR Models

Figure 7: Smooth-transition function $G$ in $s_{1t} \sim N(0, 1)$ in panel (a) and $s_{2t} \sim N(0, 0.1)$ in panel (b), when $k = 2$ with different both $s_{jt}$ but same $c_{j} = 1$.

Figure 8: Smooth-transition function $G$ in $s_{1t} \sim N(0, 1)$ in panel (a) and $s_{2t} \sim N(0, 0.1)$ in panel (b), when $k = 2$ with different both $s_{jt}$ but same $c_{j} = -1$. 