

Li, Dao; He, Changli

**Working Paper**

## Testing Common Nonlinear Features in Nonlinear Vector Autoregressive Models

Working Paper, No. 7/2012

**Provided in Cooperation with:**

Örebro University School of Business

*Suggested Citation:* Li, Dao; He, Changli (2013) : Testing Common Nonlinear Features in Nonlinear Vector Autoregressive Models, Working Paper, No. 7/2012, Örebro University School of Business, Örebro

This Version is available at:

<https://hdl.handle.net/10419/244481>

**Standard-Nutzungsbedingungen:**

Die Dokumente auf EconStor dürfen zu eigenen wissenschaftlichen Zwecken und zum Privatgebrauch gespeichert und kopiert werden.

Sie dürfen die Dokumente nicht für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, öffentlich zugänglich machen, vertreiben oder anderweitig nutzen.

Sofern die Verfasser die Dokumente unter Open-Content-Lizenzen (insbesondere CC-Lizenzen) zur Verfügung gestellt haben sollten, gelten abweichend von diesen Nutzungsbedingungen die in der dort genannten Lizenz gewährten Nutzungsrechte.

**Terms of use:**

*Documents in EconStor may be saved and copied for your personal and scholarly purposes.*

*You are not to copy documents for public or commercial purposes, to exhibit the documents publicly, to make them publicly available on the internet, or to distribute or otherwise use the documents in public.*

*If the documents have been made available under an Open Content Licence (especially Creative Commons Licences), you may exercise further usage rights as specified in the indicated licence.*



WORKING PAPER

7/2012

TESTING COMMON NONLINEAR FEATURES IN  
VECTOR NONLINEAR AUTOREGRESSIVE MODELS

DAO LI, CHANGLI HE  
STATISTICS

ISSN 1403-0586

# Testing Common Nonlinear Features in Vector Nonlinear Autoregressive Models

Dao Li\*, Changli He†

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

*This version: 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

---

\*Correspondence to: Dao Li, Örebro University School of Business, Örebro, Sweden. The earlier versions of this paper have been presented at GRAPES workshop, Örebro, August 2010, Econometric Society Australasian Meeting, Melbourne, July 2012, the 66th European Meeting of the Econometric Society, Málaga, August 2012, and organized to the 6th International Conference on CFE, Ovideo, December 2012.

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

Granger and Swanson (1996) introduced smooth-transition error-correction model for nonlinear adjustment. Johansen (1995) incorporated the smooth-transition mechanism into an error-correction model. Moreover, van Dijk and Franses (1999) contributed that multiple-regime STAR models could accommodate more than two regimes, which could not be performed with the STAR models in Teräsvirta (1994). Lin and Teräsvirta (1994) and He and Sandberg (2006) studied time-varying STAR models which allow for structural instability. He et al. (2009) discussed parameter constancy in a VSTAR 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 relies 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 Eitrheim and Teräsvirta (1996), in which several tests for misspecification were derived, following Godfrey (1988). 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 Campbell and Mankiw (1990) imply that consumption and income are linearly

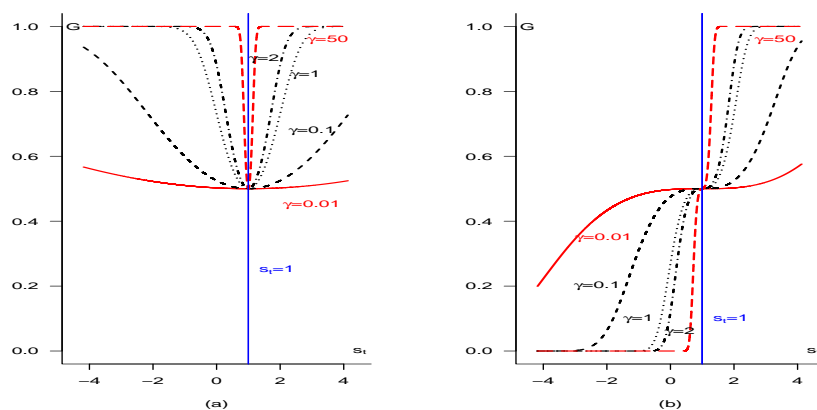


Figure 1: The ST function (2.3) with  $c_j = 1$  ( $j = 1, \dots, k$ ) for both  $k = 2$  in panel (a) and  $k = 3$  in panel (b) and  $s_{jt} \sim N(0, 1)$ .

cointegrated in the levels (see Engle and Granger; 1987, 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  $p$  (VSTAR( $p$ )),

$$\mathbf{y}_t = \boldsymbol{\mu} + \sum_{k=1}^p \boldsymbol{\Phi}_k \mathbf{y}_{t-k} + \left( \tilde{\boldsymbol{\mu}} + \sum_{k=1}^p \boldsymbol{\Gamma}_k \mathbf{y}_{t-k} \right) G(\mathbf{s}_t; \boldsymbol{\gamma}, \mathbf{c}) + \mathbf{v}_t \quad (2.1)$$

where  $\mathbf{y}_t = (y_{1t}, y_{2t}, \dots, y_{nt})' = (y_{1t}, \mathbf{y}'_{2t})'$  denotes an  $(n \times 1)$  vector time series,  $\boldsymbol{\mu}$  and  $\tilde{\boldsymbol{\mu}}$  are two  $(n \times 1)$  constant vectors,  $\boldsymbol{\Phi}_k$  and  $\boldsymbol{\Gamma}_k$  are  $(n \times n)$  coefficient matrices,  $G(\mathbf{s}_t; \boldsymbol{\gamma}, \mathbf{c})$  is a smooth-transition (ST) function, and the  $(n \times 1)$  vector  $\mathbf{v}_t$  is the

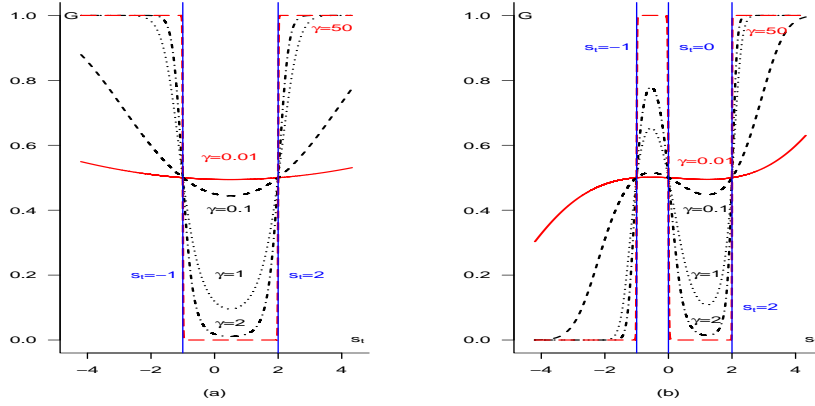


Figure 2: The ST function (2.3) with  $c_j = \{-1, 2\}$  in panel (a) ( $k = 2$ ) and  $c_j = \{-1, 0, 2\}$  in panel (b) ( $k = 3$ ) and  $s_{jt} \sim N(0, 1)$ .

error term, independent and identically distributed. The process  $\mathbf{y}_t$  defined in (2.1) is stationary and ergodic (see Assumption 2 in Section 5). When  $n = 1$ , (2.1) is a simple univariate STAR(p) model, such that

$$y_t = \mu + \sum_{k=1}^p \Phi_k y_{t-k} + \left( \tilde{\mu} + \sum_{k=1}^p \Gamma_k y_{t-k} \right) G(\mathbf{s}_t; \gamma, \mathbf{c}) + v_t \quad (2.2)$$

where  $\mu$ ,  $\tilde{\mu}$ ,  $\Phi_k$  and  $\Gamma_k$  are scalar parameters.

In this paper, we only consider logistic VSTAR models with the ST function  $G(\mathbf{s}_t; \gamma, \mathbf{c})$  defined as follows,

$$G(\mathbf{s}_t; \gamma, \mathbf{c}) = \left( 1 + \exp \left\{ -\gamma \prod_{j=1}^k (s_{jt} - c_j) \right\} \right)^{-1}, \quad \gamma > 0, \quad (2.3)$$

where  $\gamma$  is a slope parameter,  $c_j$  are location parameters and  $s_{jt}$  are transition variables. We assume  $s_{jt} = y_{it-d}$  ( $d \leq p$ ), where  $d$  is called the delay parameter. The ST function  $G(\mathbf{s}_t; \gamma, \mathbf{c})$  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  $G(\mathbf{s}_t; \gamma, \mathbf{c})$  between zero and one, all the roots of  $|\mathbf{I}_n - \sum_{k=1}^p (\Phi_k + \Gamma_k G) z^k| = 0$  lie outside the unit circle.

Figure 1-2 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  $\mathbf{s}_t$  and location parameters  $\mathbf{c}$ . For example, Figure 1 illustrates (2.3) with  $k > 1$  and  $c_j$  that are the same for different  $j$  ( $j = 1, \dots, k$ ); 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  $\mathbf{y}_t$  as defined in (2.1) and investigate whether there are CNFs in  $\mathbf{y}_t$ . An  $(n \times 1)$  nonlinear vector  $\mathbf{y}_t$  is said to contain CNFs if each of the series  $y_{it}$  is nonlinear and there is a nonzero vector  $\boldsymbol{\alpha}$  such that  $\boldsymbol{\alpha}'\mathbf{y}_t$  is linear. The vector  $\boldsymbol{\alpha}$  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  $\boldsymbol{\alpha}$ .

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

$$\boldsymbol{\alpha}'\mathbf{y}_t = \boldsymbol{\alpha}'\boldsymbol{\mu} + \sum_{k=1}^p \boldsymbol{\alpha}'\boldsymbol{\Phi}_k\mathbf{y}_{t-k} + \left( \boldsymbol{\alpha}'\tilde{\boldsymbol{\mu}} + \sum_{k=1}^p \boldsymbol{\alpha}'\boldsymbol{\Gamma}_k\mathbf{y}_{t-k} \right) G(\mathbf{s}_t; \gamma, \mathbf{c}) + \boldsymbol{\alpha}'\mathbf{v}_t \quad (3.1)$$



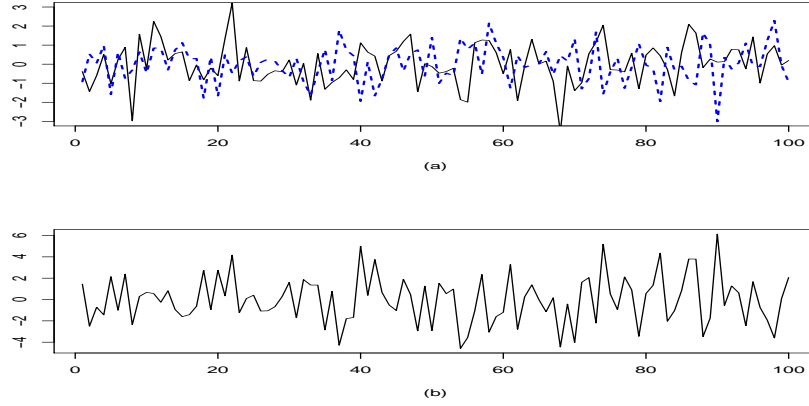


Figure 3: Two nonlinear time series (solid, dashed) with CNFs in panel (a); their linear combination in panel (b).

where  $\mathbf{y}_t$ ,  $\Phi_k$ ,  $\Gamma_k$ ,  $\boldsymbol{\mu}$ ,  $\tilde{\boldsymbol{\mu}}$  and  $\mathbf{v}_t$  are the same as defined in (2.1) and  $G(\mathbf{s}_t; \gamma, \mathbf{c})$  as in (2.3). The existence of CNFs in  $\mathbf{y}_t$  implies that  $\boldsymbol{\alpha}'\tilde{\boldsymbol{\mu}} = \mathbf{0}'$  and  $\boldsymbol{\alpha}'\Gamma_k = \mathbf{0}'$  for all  $k$ . Therefore, the linear combination (3.1) is equal to

$$\boldsymbol{\alpha}'\mathbf{y}_t = \boldsymbol{\alpha}'\boldsymbol{\mu} + \sum_{k=1}^p \boldsymbol{\alpha}'\Phi_k \mathbf{y}_{t-k} + \boldsymbol{\alpha}'\mathbf{v}_t. \quad (3.2)$$

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 ( $\boldsymbol{\mu} = \mathbf{0}$ ,  $\tilde{\boldsymbol{\mu}} = \mathbf{0}$ ,  $n = 2$ ,  $t = 1, 2, \dots, 200$ ):

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

where  $G(\mathbf{s}_t; \gamma, \mathbf{c}) = (1 + \exp\{-2(y_{2t-1} - 1)\})^{-1}$  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  $(0.12y_{1t-1} + 0.07y_{2t-1}) G(\mathbf{s}_t; \gamma, \mathbf{c})$ , which

denotes a CNF in this bivariate STAR(1) model. Figure 3 depicts the two nonlinear 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,

$$y_t = \mu + \phi_1 y_{t-1} + (\tilde{\mu} + \Gamma_1 y_{t-1}) G(\mathbf{s}_t; \gamma, \mathbf{c}) + v_t. \quad (4.1)$$

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  $G(\mathbf{s}_t; \gamma, \mathbf{c})$  with its first-order Taylor expansion round  $\gamma = 0$  as a linear approximation. An auxiliary regression is then given

in (4.2) after ignoring the remainder in the Taylor expansion of  $G(\mathbf{s}_t; \gamma, \mathbf{c})$  and reparameterization,

$$y_t = \mu^* + \phi_1^* y_{t-1} + \phi_2^* s_t y_{t-1} + \phi_3^* s_t + v_t^*, \quad (4.2)$$

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

$$H_0 : \mu^* = 0, \phi_1^* = 1, \phi_2^* = 0, \phi_3^* = 0. \quad (4.3)$$

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

$$y_t = \mu^* + \tilde{\phi}_1^* y_{t-1} + \phi_2^* y_{t-1}^2 + v_t^* \quad (4.4)$$

and correspondingly (4.3) will be updated to

$$H_0 : \mu^* = 0, \tilde{\phi}_1^* = 1, \phi_2^* = 0. \quad (4.5)$$

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  $\mathbf{y}_t$  in (2.1) when the null of a unit root is rejected in the first step, we propose a triangular representa-

tion of a system containing CNFs.

As described in Section 3, we can find a nonzero vector  $\boldsymbol{\alpha}$  such that  $\boldsymbol{\alpha}'\mathbf{y}_t$  is linear formulated in (3.2) when CNFs exist. We normalize one of the nonzero elements in  $\boldsymbol{\alpha}$  to unity and partition  $\boldsymbol{\alpha}$  as  $(1, -\boldsymbol{\alpha}^*)'$ . We then partition  $\mathbf{y}_t$  as  $(y_{1t}, \mathbf{y}'_{2t})'$  and  $\mathbf{v}_t$  as  $(v_{1t}, \mathbf{v}'_{2t})'$ , correspondingly. Model (3.2) is therefore equivalent to

$$y_{1t} = \mu_1 + \boldsymbol{\alpha}^* \mathbf{y}_{2t} + \sum_{k=1}^p \phi'_{1k} \mathbf{y}_{t-k} + u_t \quad (4.6)$$

where  $\mu_1 = \boldsymbol{\alpha}'\boldsymbol{\mu}$ ,  $\phi'_{1k} = \boldsymbol{\alpha}'\boldsymbol{\Phi}_k$ , and  $u_t = \boldsymbol{\alpha}'\mathbf{v}_t = v_{1t} - \boldsymbol{\alpha}^* \mathbf{v}_{2t}$ .

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

$$\begin{aligned} y_{1t} &= \mu_1 + \boldsymbol{\alpha}^* \mathbf{y}_{2t} + \sum_{k=1}^p \phi'_{1k} \mathbf{y}_{t-k} + u_t \\ \mathbf{y}_{2t} &= \boldsymbol{\mu}_2 + \sum_{k=1}^p \boldsymbol{\Phi}_{2k} \mathbf{y}_{t-k} + \left( \tilde{\boldsymbol{\mu}}_2 + \sum_{k=1}^p \boldsymbol{\Gamma}_{2k} \mathbf{y}_{t-k} \right) G(\mathbf{s}_t; \boldsymbol{\gamma}, \mathbf{c}) + \mathbf{v}_{2t} \end{aligned} \quad (4.7)$$

where the first equation is (4.6) containing the CNF relation in the system of  $\mathbf{y}_t$  and the second equation is the last  $n-1$  lines in (2.1) where  $\boldsymbol{\Phi}_{2k} = (\mathbf{0}, \mathbf{I}_{n-1})\boldsymbol{\Phi}_k$  and  $\boldsymbol{\Gamma}_{2k} = (\mathbf{0}, \mathbf{I}_{n-1})\boldsymbol{\Gamma}_k$  are  $(n-1) \times n$  matrices.

When a system  $\mathbf{y}_t$  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  $\mathbf{y}_t$  is constructed by the hypothesis

$$\begin{aligned} H_0 : u_t \text{ is linear} \\ H_a : u_t \text{ is nonlinear} \end{aligned} \quad (4.8)$$

in (4.6)-(4.7).

In this section, we consider the following triangular representation with  $p = 1$

for notation simplification,

$$y_{1t} = \mu_1 + \boldsymbol{\alpha}' \mathbf{y}_{2t} + \boldsymbol{\phi}'_{11} \mathbf{y}_{t-1} + u_t \quad (4.9a)$$

$$\mathbf{y}_{2t} = \boldsymbol{\mu}_2 + \boldsymbol{\Phi}_{21} \mathbf{y}_{t-1} + (\tilde{\boldsymbol{\mu}}_2 + \boldsymbol{\Gamma}_{21} \mathbf{y}_{t-1}) G(\mathbf{s}_t; \boldsymbol{\gamma}, \mathbf{c}) + \mathbf{v}_{2t} \quad (4.9b)$$

where  $\boldsymbol{\phi}_{11}$  is an  $(n \times 1)$  vector,  $\boldsymbol{\mu}_2$  and  $\tilde{\boldsymbol{\mu}}_2$  are  $(n-1) \times 1$  vectors, and  $\boldsymbol{\Phi}_{21}$  and  $\boldsymbol{\Gamma}_{21}$  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  $u_{t-1} = y_{1t-1} - \mu_1 - \boldsymbol{\alpha}' \mathbf{y}_{2t-1} - \boldsymbol{\phi}'_{11} \mathbf{y}_{t-2}$ . Then, (4.9a) takes the form of:

$$\begin{aligned} y_{1t} = & \mu_1 + \boldsymbol{\alpha}' \mathbf{y}_{2t} + \boldsymbol{\phi}'_{11} \mathbf{y}_{t-1} + \beta_0^* + \beta_1^* \left( y_{1t-1} - \mu_1 - \boldsymbol{\alpha}' \mathbf{y}_{2t-1} - \boldsymbol{\phi}'_{11} \mathbf{y}_{t-2} \right) \\ & + \beta_2^* \left( y_{1t-1} - \mu_1 - \boldsymbol{\alpha}' \mathbf{y}_{2t-1} - \boldsymbol{\phi}'_{11} \mathbf{y}_{t-2} \right)^2 + e_t. \end{aligned} \quad (4.11)$$

Reorganizing the parameters in (4.11) yields:

$$\begin{aligned} y_{1t} &= \mu_1^* + \boldsymbol{\alpha}^* \mathbf{y}_{2t} + \boldsymbol{\phi}_{11}^* \mathbf{y}_{t-1} + \boldsymbol{\phi}_{12}^* \mathbf{y}_{t-2} + \beta_2^* (\boldsymbol{\alpha}' \mathbf{y}_{t-1} - \boldsymbol{\phi}_{11}' \mathbf{y}_{t-2})^2 + e_t \\ &= \mu_1^* + \boldsymbol{\alpha}^* \mathbf{y}_{2t} + \boldsymbol{\phi}_{11}^* \mathbf{y}_{t-1} + \boldsymbol{\phi}_{12}^* \mathbf{y}_{t-2} + \boldsymbol{\theta}' \mathbf{w}_t + e_t \end{aligned} \quad (4.12)$$

where  $\mathbf{w}_t$  and  $\boldsymbol{\theta}$  are  $((2n^2 + n) \times 1)$  vectors, such that

$$\begin{aligned} \mathbf{w}_t &= (y_{1t-1}y_{1t-1}, \dots, y_{1t-1}y_{nt-1}, y_{2t-1}y_{2t-1}, \dots, y_{2t-1}y_{nt-1}, \dots, y_{nt-1}y_{nt-1}; \\ & y_{1t-2}y_{1t-2}, \dots, y_{1t-2}y_{nt-2}, y_{2t-2}y_{2t-2}, \dots, y_{2t-2}y_{nt-2}, \dots, y_{nt-2}y_{nt-2}; \\ & y_{1t-1}y_{1t-2}, \dots, y_{1t-1}y_{nt-2}, y_{2t-1}y_{1t-2}, \dots, y_{2t-1}y_{nt-2}, \dots, y_{nt-1}y_{1t-2}, \dots, y_{nt-1}y_{1t-2})' \end{aligned}$$

and

$$\begin{aligned} \boldsymbol{\theta} &= (\theta_{111}, \dots, \theta_{11n}, \theta_{122}, \dots, \theta_{12n}, \dots, \theta_{1nn}; \theta_{211}, \dots, \theta_{21n}, \theta_{222}, \dots, \theta_{22n}, \dots, \theta_{2nn}; \\ & \theta_{311}, \dots, \theta_{31n}, \theta_{321}, \dots, \theta_{32n}, \dots, \theta_{3n1}, \dots, \theta_{3nn})'. \end{aligned}$$

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

$$H_0 : \boldsymbol{\theta} = \mathbf{0}. \quad (4.13)$$

However,  $\mathbf{y}_{2t}$  in model (4.12) are endogenous regressors that are correlated with  $e_t$ . One way to correct the correlations between  $\mathbf{y}_{2t}$  and  $e_t$  is to add the leads and lags of  $\mathbf{v}_{2t} = \mathbf{y}_{2t} - \boldsymbol{\mu}_2 - \boldsymbol{\Phi}_{21} \mathbf{y}_{t-1} - (\tilde{\boldsymbol{\mu}}_2 + \boldsymbol{\Gamma}_{21} \mathbf{y}_{t-1}) G(\mathbf{s}_t; \boldsymbol{\gamma}, \mathbf{c})$ , which was suggested in the literature by Saikkonen (1991) and Stock and Waston (1993). Specifically, we can define:

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

where  $\{\tilde{e}_t\}$  is an *i.i.d* sequence with mean zero, variance  $\sigma^2$ , and  $\tilde{e}_t$  is uncorrelated with  $\mathbf{v}_{2t-s}$  for  $s = -l, -l + 1, \dots, l$ . When the model is correctly specified, it is true that  $l = 0$  is enough to correct the correlations between  $\mathbf{y}_{2t}$  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:

$$\begin{aligned}
y_{1t} &= \mu_1^* + \boldsymbol{\alpha}^{*'} \mathbf{y}_{2t} + \boldsymbol{\phi}_{11}^{*'} \mathbf{y}_{t-1} + \boldsymbol{\phi}_{12}^{*'} \mathbf{y}_{t-2} + \boldsymbol{\theta}' \mathbf{w}_t + \sum_{s=-l}^l \boldsymbol{\delta}'_s \mathbf{v}_{2t-s} + \tilde{e}_t \\
&= \tilde{\mu} + \boldsymbol{\theta}' \mathbf{w}_t + \boldsymbol{\delta}'_1 \mathbf{y}_{2t+l} + \sum_{s=-l+1}^{\max(2,l+1)} \boldsymbol{\delta}'_{2s} \mathbf{y}_{t-s} \\
&\quad + \sum_{s=-l}^l \boldsymbol{\delta}'_{3s} \mathbf{y}_{t-1-s} G(\mathbf{s}_{t-s}; \gamma, \mathbf{c}) + \sum_{s=-l}^l \delta_{4s} G(\mathbf{s}_{t-s}; \gamma, \mathbf{c}) + \tilde{e}_t \quad (4.15)
\end{aligned}$$

where  $\tilde{e}_t$  is uncorrelated with all regressors in (4.15),  $\tilde{\mu}$  and  $\delta_{4s}$  are scalar parameters,  $\boldsymbol{\delta}'_1$  is an  $(n-1) \times 1$  vector parameter, and  $\boldsymbol{\delta}'_{2s}$  and  $\boldsymbol{\delta}'_{3s}$  are  $(n \times 1)$  vector parameters. For notation simplification, we rewrite (4.15) as:

$$y_{1t} = \boldsymbol{\theta}' \mathbf{w}_t + f(\mathbf{y}_{t-s}; \boldsymbol{\psi}) + \tilde{e}_t \quad (4.16)$$

by defining

$$\begin{aligned}
f(\mathbf{y}_{t-s}; \boldsymbol{\psi}) &= \tilde{\mu} + \boldsymbol{\delta}'_1 \mathbf{y}_{2t+l} + \sum_{s=-l+1}^{\max(2,l+1)} \boldsymbol{\delta}'_{2s} \mathbf{y}_{t-s} \\
&\quad + \sum_{s=-l}^l \boldsymbol{\delta}'_{3s} \mathbf{y}_{t-1-s} G(\mathbf{s}_{t-s}; \gamma, \mathbf{c}) + \sum_{s=-l}^l \delta_{4s} G(\mathbf{s}_{t-s}; \gamma, \mathbf{c}),
\end{aligned}$$

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

Assume the necessary starting values  $\mathbf{y}_{-p}, \mathbf{y}_{-p+1}, \dots, \mathbf{y}_{-1}, \mathbf{y}_0$  are fixed. Given a realization of  $\mathbf{y}_t, t = 1, 2, \dots, T$ , the conditional nonlinear least square estimators

of  $(\boldsymbol{\theta}', \boldsymbol{\psi}')$  are carried out by solving the problem of minimizing:

$$\mathcal{L}(\boldsymbol{\theta}, \boldsymbol{\psi}) = \sum_{t=1}^T (y_{1t} - \boldsymbol{\theta}' \mathbf{w}_t - f(\mathbf{y}_{t-s}; \boldsymbol{\psi}))^2 \quad (4.17)$$

with respect to  $(\boldsymbol{\theta}', \boldsymbol{\psi}')$ . The first-order partial derivatives of function  $\mathcal{L}$  are

$$\frac{\partial \mathcal{L}}{\partial \boldsymbol{\theta}} = \sum_{t=1}^T 2(y_{1t} - \boldsymbol{\theta}' \mathbf{w}_t - f(\mathbf{y}_{t-s}; \boldsymbol{\psi})) \mathbf{w}_t, \quad (4.18a)$$

$$\frac{\partial \mathcal{L}}{\partial \boldsymbol{\psi}} = \sum_{t=1}^T 2 \left( y_{1t} - \boldsymbol{\theta}' \mathbf{w}_t - f(\mathbf{y}_{t-s}; \boldsymbol{\psi}) \right) \frac{\partial f(\mathbf{y}_{t-s}; \boldsymbol{\psi})}{\partial \boldsymbol{\psi}}. \quad (4.18b)$$

Under the null of  $\boldsymbol{\theta} = \mathbf{0}$  in (4.13), (4.18a) and (4.18b) are reduced to

$$\left. \frac{\partial \mathcal{L}}{\partial \boldsymbol{\theta}} \right|_{H_0} = \sum_{t=1}^T 2z_{1t} \mathbf{w}_t \quad (4.19a)$$

$$\left. \frac{\partial \mathcal{L}}{\partial \boldsymbol{\psi}} \right|_{H_0} = \sum_{t=1}^T 2z_{1t} \mathbf{z}_{2t} \quad (4.19b)$$

where  $z_{1t} = y_{1t} - f(\mathbf{y}_{t-s}; \boldsymbol{\psi})$  and  $\mathbf{z}_{2t} = \partial f(\mathbf{y}_{t-s}; \boldsymbol{\psi}) / \partial \boldsymbol{\psi}$ . Given the consistent nonlinear least square estimators  $\hat{\boldsymbol{\psi}}$  of the true parameter  $\boldsymbol{\psi}$ , (4.19a) and (4.19b) are calculated by  $\sum_{t=1}^T 2\hat{z}_{1t} \mathbf{w}_t$  and  $\sum_{t=1}^T 2\hat{z}_{1t} \hat{\mathbf{z}}_{2t}$  where  $\hat{z}_{1t} = y_{1t} - f(\mathbf{y}_{t-s}; \hat{\boldsymbol{\psi}})$  and  $\hat{\mathbf{z}}_{2t} = \partial f(\mathbf{y}_{t-s}; \hat{\boldsymbol{\psi}}) / \partial \boldsymbol{\psi}$ . Our *LM* statistic for testing CNFs in (4.16) under the null (4.13) is calculated by the following formula:

$$\hat{\sigma}^{-2} \left( \sum_{t=1}^T \hat{z}_{1t} \mathbf{w}_t \right) \left( \sum_{t=1}^T \mathbf{w}_t \mathbf{w}_t' - \sum_{t=1}^T \mathbf{w}_t \hat{\mathbf{z}}_{2t}' \left( \sum_{t=1}^T \hat{\mathbf{z}}_{2t} \hat{\mathbf{z}}_{2t}' \right)^{-1} \sum_{t=1}^T \hat{\mathbf{z}}_{2t} \mathbf{w}_t' \right)^{-1} \left( \sum_{t=1}^T \hat{z}_{1t} \mathbf{w}_t' \right) \quad (4.20)$$

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  $\hat{\mathbf{z}}_{2t}$  are defined when the form of transition function  $G(\mathbf{s}_t; \boldsymbol{\gamma}, \mathbf{c})$  in (4.16) is specified. Taking  $G(\mathbf{s}_t; \boldsymbol{\gamma}, \mathbf{c}) = (1 + e^{-\gamma(s_t - c)})^{-1}$  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  $\hat{\mathbf{z}}_{2t}$  in this case is

$$\begin{aligned} \hat{\mathbf{z}}_{2t} = \partial f(\mathbf{y}_{t-s}; \hat{\boldsymbol{\psi}}) / \partial \boldsymbol{\psi} = & (1, \mathbf{y}'_{2t+l}, \mathbf{y}'_{t-\max(2, l+1)}, \dots, \mathbf{y}'_{t+l-1}, \\ & \mathbf{y}'_{t-1-l} G(s_{t-l}; \hat{\gamma}, \hat{c}), \dots, \mathbf{y}'_{t-1+l} G(s_{t+l}; \hat{\gamma}, \hat{c}), \\ & G(s_{t-l}; \hat{\gamma}, \hat{c}), \dots, G(s_{t+l}; \hat{\gamma}, \hat{c}), \\ & \sum_{s=-l}^l (\hat{\boldsymbol{\delta}}'_{3s} \mathbf{y}_{t-1-s} + \hat{\delta}_{4s}) \hat{g}_\gamma(t-s), \\ & \sum_{s=-l}^l (\hat{\boldsymbol{\delta}}'_{3s} \mathbf{y}_{t-1-s} + \hat{\delta}_{4s}) \hat{g}_c(t-s))'. \end{aligned} \quad (4.22)$$

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  $\boldsymbol{\psi}$  in model (4.16) by nonlinear least square estimation under the assumption of  $\boldsymbol{\theta} = \mathbf{0}$  and compute the residual sum of squares  $SSR_0 = \sum_{t=1}^T \hat{z}_{1t}^2$ ;
- (ii) Regress  $\hat{z}_{1t}$  on  $\mathbf{w}_t$  and  $\hat{\mathbf{z}}_{2t}$  and compute the residual sum of squares  $SSR$ ;
- (iii) Calculate the test statistic

$$F_{LM} = \frac{(SSR_0 - SSR)/(2n^2 + n)}{SSR/(T - 2n^2 - n - k)} \quad (4.23)$$

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

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),  $\{\tilde{\varepsilon}_t\}$  is an i.i.d sequence with mean zero and variance  $\sigma^2$ , and  $\tilde{\varepsilon}_t$  is uncorrelated with all leads and lags of  $\varepsilon_t$ .

**Assumption 2.**  $\mathbf{y}_t$  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,

$$\left| \mathbf{I}_n - \sum_{k=1}^p (\Phi_k + \Gamma_k G) z^k \right| \neq 0 \quad \text{if } |z| \leq 1;$$

(2) The second-order moments of  $\mathbf{y}_t$  depend on  $t$  defined by

$$E(\mathbf{y}_t \mathbf{y}'_t) = \begin{pmatrix} E(y_{1t} y_{1t}) & E(y_{1t} \mathbf{y}'_{2t}) \\ E(\mathbf{y}_{2t} y'_{1t}) & E(\mathbf{y}_{2t} \mathbf{y}'_{2t}) \end{pmatrix} = \Omega_t = \begin{pmatrix} \Lambda_{11t} & \Lambda_{12t} \\ \Lambda_{21t} & \Lambda_{22t} \end{pmatrix}$$

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

(4) The process  $\mathbf{y}_t$  is ergodic for the fourth-order moments.

Given  $G(\mathbf{s}_t; \gamma, \mathbf{c})$  in (2.3),  $\boldsymbol{\theta}' \mathbf{w}_t + f(\mathbf{y}_{t-s}; \boldsymbol{\psi})$  is continuously differentiable twice with respect to  $(\boldsymbol{\theta}', \boldsymbol{\psi}')'$  in some neighborhood of  $(\boldsymbol{\theta}', \boldsymbol{\psi}'_0)'$  under the null. Assumption 2 implies the regularity conditions in Klimko and Nelson (1978, page

634) holds. Hence, the nonlinear least square estimator of  $(\boldsymbol{\theta}', \boldsymbol{\psi}')$  in (4.16) is consistent according to Theorem 3.1 in Klimko and Nelson (1978). Under Assumptions 1-2, the asymptotic distribution of

$$T^{-1/2} \left( \sum_{t=1}^T z_{1t} \mathbf{z}'_{2t}, \sum_{t=1}^T z_{1t} \mathbf{w}'_t \right)'$$

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  $T \rightarrow \infty$ . 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 Monto 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  $\mathbf{y}_t$  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 6.1, 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:

$$\begin{pmatrix} y_{1t} \\ y_{2t} \end{pmatrix} = \begin{pmatrix} 0.4 \\ 0.2 \end{pmatrix} + \begin{pmatrix} 0.3 & 0.05 \\ 0.2 & -0.1 \end{pmatrix} \begin{pmatrix} y_{1t-1} \\ y_{2t-1} \end{pmatrix} \\ + \begin{pmatrix} 2 \\ 1 \end{pmatrix} \left( 0.05 + \begin{pmatrix} 0.12 & -0.18 \end{pmatrix} \begin{pmatrix} y_{1t-1} \\ y_{2t-1} \end{pmatrix} \right) G(\mathbf{s}_t; \gamma, \mathbf{c}) + \begin{pmatrix} v_{1t} \\ v_{2t} \end{pmatrix}$$

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

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

The estimation of the parameter  $\boldsymbol{\psi}$  is from (4.16) under the null of  $\boldsymbol{\theta} = \mathbf{0}$ . In this case,  $f(\mathbf{y}_{t-s}; \boldsymbol{\psi})$  has a form

$$f(\mathbf{y}_{t-s}; \boldsymbol{\psi}) = \tilde{\mu} + \delta_1 y_{2t} + \boldsymbol{\delta}'_{21} \mathbf{y}_{t-1} + \boldsymbol{\delta}'_{22} \mathbf{y}_{t-2} + (\boldsymbol{\delta}'_{30} \mathbf{y}_{t-1} + \delta_{40}) G(\mathbf{s}_t; \gamma, \mathbf{c}).$$

To calculate our  $F_{LM}$  statistic,  $\mathbf{w}_t$  is given in detail such that:

$$\mathbf{w}_t = (y_{1t-1}^2, y_{1t-1}y_{2t-1}, y_{2t-1}^2, y_{1t-2}^2, y_{1t-2}y_{2t-2}, y_{2t-2}^2, y_{1t-1}y_{1t-2}, y_{1t-1}y_{2t-2}, \\ y_{2t-1}y_{1t-2}, y_{2t-1}y_{2t-2})'.$$

The  $DGP$  under the alternative for the power study follows a VSTAR model

Table 1: Critical values for our Dickey-Fuller  $F$ -type and  $t$ -type tests under the null (4.3) with  $s_t = \Delta y_{t-1}$ .

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

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

$s_t$	$T$	Probability that $t$ is less than entry							
		99%	97.5%	95%	90%	10%	5%	2.5%	1%
$\Delta y_{t-1}$	50	-3.5351	-3.1853	-2.8986	-2.5813	-0.3758	-0.0057	0.3228	0.7151
	100	-3.4738	-3.1449	-2.8786	-2.5716	-0.4096	-0.0491	0.2749	0.6591
	200	-3.4593	-3.1346	-2.8736	-2.5685	-0.4244	-0.0562	0.2595	0.6399
	500	-3.4273	-3.1274	-2.8660	-2.5707	-0.4381	-0.0718	0.2564	0.6242
	1000	-3.4251	-3.1117	-2.8523	-2.5593	-0.4425	-0.0778	0.2507	0.6198

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  $s_t = y_{t-1}$ .

$s_t$	$T$	Probability that $F$ is greater than entry							
		99%	97.5%	95%	90%	10%	5%	2.5%	1%
$y_{t-1}$	50	0.2638	0.3437	0.4281	0.5606	3.0599	3.7372	4.4263	5.3806
	100	0.2654	0.3467	0.4319	0.5619	2.9649	3.5937	4.2322	5.0352
	200	0.2725	0.3500	0.4348	0.5639	2.9164	3.5180	4.1144	4.8323
	500	0.2735	0.3511	0.4353	0.5639	2.9108	3.5004	4.0708	4.7926
	1000	0.2704	0.3501	0.4343	0.5642	2.8923	3.4874	4.0520	4.7917

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

$s_t$	$T$	Probability that $t$ is less than entry							
		99%	97.5%	95%	90%	10%	5%	2.5%	1%
$y_{t-1}$	50	-2.7063	-2.3056	-1.9857	-1.6188	0.8153	1.1617	1.4753	1.8644
	100	-2.6606	-2.2822	-1.9621	-1.6027	0.7821	1.1224	1.4300	1.7939
	200	-2.6591	-2.2723	-1.9623	-1.6025	0.7708	1.1163	1.4164	1.7675
	500	-2.6414	-2.2629	-1.9461	-1.5908	0.7603	1.0917	1.3794	1.7126
	1000	-2.6394	-2.2716	-1.9490	-1.5943	0.7502	1.0865	1.3687	1.7021

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.

$T$	Size				Size-adjusted Power			
	given significant levels				given significant levels			
	0.2000	0.1000	0.0500	0.0100	0.2000	0.1000	0.0500	0.0100
100	0.3536	0.2053	0.1179	0.0324	0.1394	0.0652	0.0272	0.0043
200	0.3471	0.2055	0.1202	0.0305	0.1411	0.0649	0.0292	0.0051
500	0.2896	0.1605	0.0892	0.0226	0.1743	0.0889	0.0428	0.0077
1000	0.2864	0.1556	0.0851	0.0189	0.2815	0.1698	0.1066	0.0368

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

$T$	Size				Power			
	given significant levels				given significant levels			
	0.2000	0.1000	0.0500	0.0100	0.2000	0.1000	0.0500	0.0100
100	0.3536	0.2053	0.1179	0.0324	0.2688	0.1458	0.0777	0.0161
200	0.3471	0.2055	0.1202	0.0305	0.2703	0.1469	0.0814	0.0180
500	0.2896	0.1605	0.0892	0.0226	0.2603	0.1394	0.0790	0.0179
1000	0.2864	0.1556	0.0851	0.0189	0.3618	0.2375	0.1506	0.0564

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

$T$	Size				Size-adjusted Power			
	given significant levels				given significant levels			
	0.2000	0.1000	0.0500	0.0100	0.2000	0.1000	0.0500	0.0100
100	0.3536	0.2053	0.1179	0.0324	0.1878	0.0906	0.0427	0.0080
200	0.3471	0.2055	0.1202	0.0305	0.2133	0.1095	0.0566	0.0109
500	0.2896	0.1605	0.0892	0.0226	0.3936	0.2463	0.1445	0.0410
1000	0.2864	0.1556	0.0851	0.0189	0.5296	0.3628	0.2474	0.0976
2000	0.2940	0.1684	0.0906	0.0207	0.7520	0.6190	0.4910	0.2500

(without CNFs) such that

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

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

$$\begin{aligned} \begin{pmatrix} y_{1t} \\ y_{2t} \end{pmatrix} &= \begin{pmatrix} 0.4 \\ 0.2 \end{pmatrix} + \begin{pmatrix} 0.3 & 0.05 \\ 0.2 & -0.1 \end{pmatrix} \begin{pmatrix} y_{1t-1} \\ y_{2t-1} \end{pmatrix} + \begin{pmatrix} 0.1G(s_t^*; \gamma, c) \\ -0.16G(s_t; \gamma, c) \end{pmatrix} \\ &+ \begin{pmatrix} 0.24 & -0.18 \\ 0.12 & -0.36 \end{pmatrix} \begin{pmatrix} y_{1t-1}G(s_t^*; \gamma, c) \\ y_{2t-1}G(s_t; \gamma, c) \end{pmatrix} + \begin{pmatrix} v_{1t} \\ v_{2t} \end{pmatrix} \end{aligned} \quad (6.1)$$

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 3-6. 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

Table 6: Empirical size and power of the  $F_{LM}$  test, DGP2.

$T$	Size				Power			
	given significant levels				given significant levels			
	0.2000	0.1000	0.0500	0.0100	0.2000	0.1000	0.0500	0.0100
100	0.3536	0.2053	0.1179	0.0324	0.3350	0.1932	0.1079	0.0266
200	0.3471	0.2055	0.1202	0.0305	0.3809	0.2208	0.1287	0.0345
500	0.2896	0.1605	0.0892	0.0226	0.5094	0.3401	0.2230	0.0781
1000	0.2864	0.1556	0.0851	0.0189	0.6372	0.4674	0.3304	0.1424
2000	0.2940	0.1684	0.0906	0.0207	0.8540	0.7140	0.6050	0.3440

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

$T$	Size				Size-adjusted Power			
	given significant levels				given significant levels			
	0.2000	0.1000	0.0500	0.0100	0.2000	0.1000	0.0500	0.0100
100	0.1830	0.0877	0.0423	0.0085	0.2998	0.1753	0.0994	0.0266
200	0.1954	0.0972	0.0473	0.0088	0.3968	0.2502	0.1581	0.0512
500	0.1986	0.1013	0.0492	0.0088	0.6716	0.5172	0.3986	0.2054
1000	0.1976	0.0991	0.0481	0.0092	0.9158	0.8410	0.7543	0.5340

mind, we further examine the finite-sample properties, assuming that the combination vector  $\alpha$  for CNFs in our case is known. This makes model (4.16) simpler because it is not required to correct the correlations between  $\mathbf{y}_{2t}$  and  $e_t$ . Therefore,  $f(\mathbf{y}_{t-s}; \boldsymbol{\psi})$  has a linear form. In this specific experiment,  $f(\mathbf{y}_{t-s}; \boldsymbol{\psi})$  has the form as follows ( $\alpha^*$  is known),

$$f(\mathbf{y}_{t-s}; \boldsymbol{\psi}) = \mu_1^* + \alpha^* y_{2t} + \sum_{s=1}^2 \phi_{1s}^* \mathbf{y}_{t-s}.$$

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.

## 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),



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

$T$	Size				Power			
	given significant levels				given significant levels			
	0.2000	0.1000	0.0500	0.0100	0.2000	0.1000	0.0500	0.0100
100	0.1830	0.0877	0.0423	0.0085	0.2791	0.1541	0.0886	0.0218
200	0.1954	0.0972	0.0473	0.0088	0.3752	0.2232	0.1274	0.0325
500	0.1986	0.1013	0.0492	0.0088	0.6485	0.4824	0.3482	0.1372
1000	0.1976	0.0991	0.0481	0.0092	0.0931	0.8131	0.6992	0.4250

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 4 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 4 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.

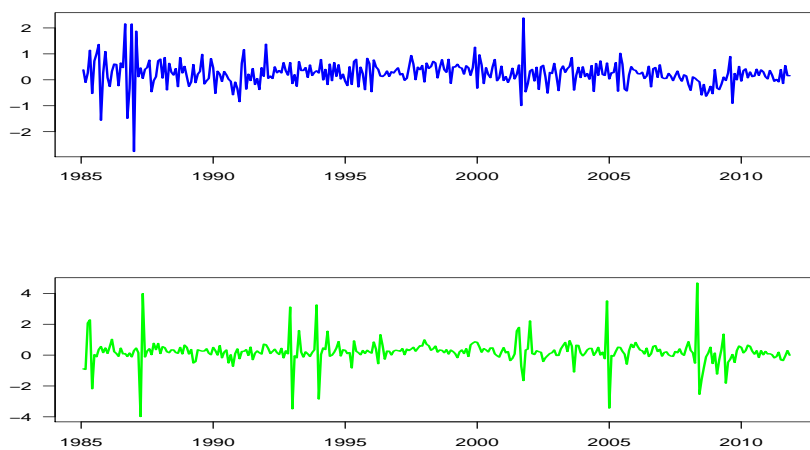


Figure 4: Percentage change in personal consumption expenditures (upper) and disposable income (lower), United States, billions of 2005 dollars, monthly, 1985:1-2011:11.

Table 9: The unit root tests.

time series	$s_t$	DF statistics		critical values	
		$t$	$F$	$t$	$F$
$\Delta y_{1t}$	$\Delta^2 y_{1t-1}$	-2.7476	4.3017	-2.87	3.12
	$\Delta y_{1t-1}$	-2.6744	4.8862	-1.96	3.52
$\Delta y_{2t}$	$\Delta^2 y_{2t-1}$	-5.7069	18.029	-2.87	3.12
	$\Delta y_{2t-1}$	-5.6537	17.810	-1.96	3.52

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) choosing  $p = 1$  and  $l = 2$ . Based on the nonlinear least square estimation, the  $F_{LM}$  test statistic in (4.23) 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.

data	test	$s_t$					
		$x_{1,t-1}$	$x_{1,t-2}$	$x_{1,t-3}$	$x_{1,t-4}$	$x_{1,t-5}$	$x_{1,t-6}$
$x_{1t}$	$t$	-2.6744	-2.7522	-3.1715	-2.3702	-2.4773	-2.5426
	$F$	4.8862	3.7963	<b>8.7313</b>	5.9177	<b>6.1487</b>	3.8504
$x_{2t}$	$t$	-4.6066	-5.3846	-4.5765	-5.0576	-4.7381	-5.2491
	$F$	<b>19.504</b>	14.806	15.725	15.762	15.604	14.271
data	test	$s_t$					
		$x_{1,t-7}$	$x_{1,t-8}$	$x_{1,t-9}$	$x_{1,t-10}$	$x_{1,t-11}$	$x_{1,t-12}$
$x_{1t}$	$t$	-3.1982	-2.5386	-2.4964	-2.0108	-3.4666	-2.9038
	$F$	<b>6.1462</b>	3.7753	5.1453	<b>6.4307</b>	<b>9.1321</b>	<b>6.0254</b>
$x_{2t}$	$t$	-5.2798	-4.2107	-5.3020	-5.3532	-5.3216	-4.4308
	$F$	<b>24.610</b>	<b>20.225</b>	14.169	14.402	16.495	<b>19.319</b>
data	test	$s_t$					
		$x_{2,t-1}$	$x_{2,t-2}$	$x_{2,t-3}$	$x_{2,t-4}$	$x_{2,t-5}$	$x_{2,t-6}$
$x_{1t}$	$t$	-2.8004	-2.5715	-2.6882	-2.7418	-2.8424	-2.6694
	$F$	4.1458	4.2530	3.7920	3.9519	4.0765	4.1619
$x_{2t}$	$t$	-5.6537	-5.6306	-5.2734	-5.3828	-5.1479	-5.4606
	$F$	17.810	16.894	14.196	14.506	15.249	15.322
data	test	$s_t$					
		$x_{2,t-7}$	$x_{2,t-8}$	$x_{2,t-9}$	$x_{2,t-10}$	$x_{2,t-11}$	$x_{2,t-12}$
$x_{1t}$	$t$	-2.8831	-2.7102	-2.6630	-2.4212	-2.5734	-2.7697
	$F$	4.4869	3.7515	3.9776	4.8118	4.4453	3.9163
$x_{2t}$	$t$	-5.6110	-4.5739	-5.0931	-5.4967	-5.2708	-5.4674
	$F$	17.485	17.585	16.843	15.108	14.160	16.565

\*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  $G(\mathbf{s}_t; \gamma, \mathbf{c})$ . 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:

$$\begin{aligned} \begin{pmatrix} \hat{x}_{1t} \\ \hat{x}_{2t} \end{pmatrix} &= \begin{pmatrix} 0.3682 \\ 0.3328 \end{pmatrix} + \begin{pmatrix} -0.3778 & 0.0291 \\ 0.0087 & -0.3228 \end{pmatrix} \begin{pmatrix} x_{1t-1} \\ x_{2t-1} \end{pmatrix} \\ &+ \begin{pmatrix} -0.0378 \\ -0.0545 \end{pmatrix} (0.0493 + 0.0100x_{1t-1} - 0.0850x_{2t-1}) G(\mathbf{s}_t; \hat{\gamma}, \hat{\mathbf{c}}) \end{aligned}$$

where  $G(\mathbf{s}_t; \hat{\gamma}, \hat{\mathbf{c}}) = (1 + \exp \{-47.5(x_{1t-7} - 0.98)/\hat{\sigma}_{x_{1t-7}}\})^{-1}$  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

- Anderson, H. M. and Vahid, F. (1998). Testing multiple equation systems for common nonlinear components, *Journal of Econometrics* **84**(1): 1–36.
- Brännäs, K., De Gooijer, J. G. and Teräsvirta, T. (1998). Testing linearity against

- nonlinear moving average models, *Communications in Statistics - Theory and Methods* **27**(8): 2025–2035.
- Campbell, J. Y. and Mankiw, N. G. (1990). Permanent income, current income, and consumption, *Journal of Business & Economic Statistics* **8**(3): 265–79.
- Eitrheim, O. and Teräsvirta, T. (1996). Testing the adequacy of smooth transition autoregressive models, *Journal of Econometrics* **74**(1): 59–57.
- Elliott, G., Jansson, M. and Pesavento, E. (2005). Optimal power for testing potential cointegrating vector with known parameters for nonstationarity, *Journal of Business & Economic Statistics* **23**(1): 34–48.
- Engle, R. F. and Granger, C. W. J. (1987). Co-integration and error correction: Representation, estimation and testing, *Econometrica* **55**(2): 251–276.
- Engle, R. and Kozicki, S. (1993). Testing for common features, *Journal of Business and Economic Statistics* **11**(4): 369–380.
- Escribano, A. (1987). *Nonlinear Error-Correction: The Case of Money Demand in the UK (1878-1970)*, Chapter iv, dissertation, University of California: San Diego.
- Godfrey, L. G. (1988). *Misspecification tests in econometrics*, Cambridge University Press, Cambridge.
- Granger, C. W. J. and Swanson, N. R. (1996). Future developments in the study of cointegrated variables, *Oxford Bulletin of Economics and Statistics* **58**: 537–553.
- He, C. and Sandberg, R. (2006). Dickey-fuller type of tests against nonlinear dynamic models, *Oxford Bulletin of Economics and Statistics* **68**(s1): 835–861.
- He, C., Teräsvirta, T. and González, A. (2009). Testing parameter constancy in stationary vector autoregressive models against continuous change, *Econometric Reviews* **28**(1-3): 1–21.

- Hosking, J. R. M. (1980). The multivariate portmanteau statistic, *Journal of the American Statistical Association* **75**: 6602–608.
- Johansen, S. (1995). *Likelihood-Based Inference in Cointegrated Vector Autoregressive Models*, Oxford University Press, Oxford.
- Juselius, K. (2007). *The Cointegrated VAR Model: Methodology and Applications*, Oxford University Press, New York.
- Kapetanios, G. and Shin, Y. (2006). Unit root tests in three-regime setar models, *Econometrics Journal* **9**: 252–278.
- Klimko, L. A. and Nelson, P. I. (1978). On conditional least square estimation for stochastic processes, *Annals of Statistics* **6**: 629–642.
- Lin, C. F. J. and Teräsvirta, T. (1994). Testing the constancy of regression parameters against continuous structural change, *Journal of Econometrics* **62**: 211–228.
- Ljung, G. and Box, G. (1978). On a measure of lack of fit in time series models, *Biometrika* **66**: 67–72.
- Luukkonen, R., Saikkonen, P. and Teräsvirta, T. (1988). Testing linearity against smooth transition autoregressive models, *Biometrika* **75**: 491–499.
- 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.
- Saikkonen, P. (1991). Asymptotic properties of residual based tests for cointegration, *Econometric Theory* **7**: 1–21.
- Stock, J. and Watson, M. (1993). A simple estimator of cointegrating vectors in higher order integrated systems, *Econometrica* **61**: 783–820.
- Teräsvirta, T. (1994). Specification, estimation, and evaluation of smooth transition autoregressive models, *Journal of the American Statistical Association* **88**(425): 208–218.

van Dijk, D. and Franses, P. H. (1999). Modeling multiple regimes in the business cycle, *Macroeconomic Dynamics* **3**: 311–340.

White, H. (2001). *Asymptotic Theory for Econometricians. Revised edition.*, Emerald Group Publishing Limited, UK.

## 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  $\mathcal{L}$ ,

$$T^{-\frac{1}{2}} \begin{pmatrix} \sum_{t=1}^T z_{1t} \mathbf{w}_t \\ \sum_{t=1}^T z_{1t} \mathbf{z}_{2t} \end{pmatrix}, \quad (\text{A.1})$$

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

$$\mathbf{V}_t = \begin{pmatrix} E(\mathbf{w}_t \mathbf{w}_t') & E(\mathbf{w}_t \mathbf{z}_{2t}') \\ E(\mathbf{z}_{2t} \mathbf{w}_t') & E(\mathbf{z}_{2t} \mathbf{z}_{2t}') \end{pmatrix},$$

and  $\mathbf{V}_t$  and  $\mathbf{V}$  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  $T \rightarrow \infty$ .

**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  $(\boldsymbol{\theta}', \boldsymbol{\psi}')$



are consistent. Therefore,  $\hat{z}_{1t}$  and  $\hat{\mathbf{z}}_{2t}$  are consistent estimators of  $z_{1t}$  and  $\mathbf{z}_{2t}$  respectively. Then we have

$$T^{-\frac{1}{2}} \begin{pmatrix} \sum_{t=1}^T \hat{z}_{1t} \mathbf{w}_t - \sum_{t=1}^T z_{1t} \mathbf{w}_t \\ \sum_{t=1}^T \hat{z}_{1t} \hat{\mathbf{z}}_{2t} - \sum_{t=1}^T z_{1t} \mathbf{z}_{2t} \end{pmatrix} \xrightarrow{p} \mathbf{0} \quad (\text{A.2})$$

as  $T \rightarrow \infty$  by consistency. Thus,

$$T^{-\frac{1}{2}} \begin{pmatrix} \sum_{t=1}^T \hat{z}_{1t} \mathbf{w}_t \\ \sum_{t=1}^T \hat{z}_{1t} \hat{\mathbf{z}}_{2t} \end{pmatrix} \quad (\text{A.3})$$

has the same asymptotic distribution with

$$T^{-\frac{1}{2}} \begin{pmatrix} \sum_{t=1}^T z_{1t} \mathbf{w}_t \\ \sum_{t=1}^T z_{1t} \mathbf{z}_{2t} \end{pmatrix} \xrightarrow{L} N(\mathbf{0}, E(z_{1t}^2) \mathbf{V}) \quad (\text{A.4})$$

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

$$(\hat{\sigma}^2 \hat{\mathbf{V}})^{-\frac{1}{2}} T^{-\frac{1}{2}} \begin{pmatrix} \sum_{t=1}^T \hat{z}_{1t} \mathbf{w}_t \\ \sum_{t=1}^T \hat{z}_{1t} \hat{\mathbf{z}}_{2t} \end{pmatrix} \xrightarrow{L} N(\mathbf{0}, \mathbf{I}) \quad (\text{A.5})$$

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

$$\hat{\mathbf{V}} = \begin{pmatrix} \sum_{t=1}^T \mathbf{w}_t \mathbf{w}_t' / T & \sum_{t=1}^T \mathbf{w}_t \hat{\mathbf{z}}_{2t}' / T \\ \sum_{t=1}^T \hat{\mathbf{z}}_{2t} \mathbf{w}_t' / T & \sum_{t=1}^T \hat{\mathbf{z}}_{2t} \hat{\mathbf{z}}_{2t}' / T \end{pmatrix}.$$

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 (4.13) could be rewritten as

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

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

$$\mathbf{R}(\hat{\sigma}^2 T \hat{\mathbf{V}})^{-1/2} \begin{pmatrix} \sum_{t=1}^T \hat{z}_{1t} \mathbf{w}_t \\ \sum_{t=1}^T \hat{z}_{1t} \hat{\mathbf{z}}_{2t} \end{pmatrix} \xrightarrow{L} N(\mathbf{0}, \mathbf{R} \mathbf{I}_{m_1+m_2} \mathbf{R}') \equiv N(\mathbf{0}, \mathbf{I}_{m_1}) \quad (\text{A.7})$$

as  $T \rightarrow \infty$ . Thus,

$$\begin{aligned} & \hat{\sigma}^{-2} \begin{pmatrix} \sum_{t=1}^T \hat{z}_{1t} \mathbf{w}_t' & \sum_{t=1}^T \hat{z}_{1t} \hat{\mathbf{z}}_{2t}' \end{pmatrix} \mathbf{R}' \left( \mathbf{R} (T \hat{\mathbf{V}})^{-1} \mathbf{R}' \right) \mathbf{R} \begin{pmatrix} \sum_{t=1}^T \hat{z}_{1t} \mathbf{w}_t \\ \sum_{t=1}^T \hat{z}_{1t} \hat{\mathbf{z}}_{2t} \end{pmatrix} \\ &= \hat{\sigma}^{-2} \left( \sum_{t=1}^T \hat{z}_{1t} \mathbf{w}_t' \right) \left( \mathbf{R} (T \hat{\mathbf{V}})^{-1} \mathbf{R}' \right) \left( \sum_{t=1}^T \hat{z}_{1t} \mathbf{w}_t' \right) \xrightarrow{L} \chi^2(m_1) \end{aligned} \quad (\text{A.8})$$

as  $T \rightarrow \infty$  in which

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

□

## 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  $\mu^* = 0$  in (4.2) in Section 4.1. Test

$$H_0 : y_t = y_{t-1} + v_t^* \quad (\text{B.1})$$

against

$$H_1^{aux} : y_t = \phi_1^* y_{t-1} + \phi_2^* s_t y_{t-1} + \phi_3^* s_t + v_t^*. \quad (\text{B.2})$$

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

$$H_0 : y_t = y_{t-1} + v_t^* \quad (\text{B.3})$$

Table 11: Critical values for non-standard unit root tests under the null (B.1) with  $s_t = \Delta y_{t-1}$ .

$s_t$	$T$	Probability that $F$ is greater than entry							
		99%	97.5%	95%	90%	10%	5%	2.5%	1%
$\Delta y_{t-1}$	50	0.0415	0.0774	0.1258	0.2062	2.1703	2.7473	3.3246	4.1092
	100	0.0421	0.0778	0.1267	0.2071	2.1378	2.6923	3.2481	3.9582
	200	0.0417	0.0783	0.1260	0.2101	2.1343	2.6766	3.1867	3.8797
	500	0.0426	0.0805	0.1297	0.2137	2.1549	2.6797	3.1830	3.8652
	1000	0.0431	0.0803	0.1290	0.2140	2.1448	2.6846	3.2070	3.8666

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

$s_t$	$T$	Probability that $t$ is less than entry							
		99%	97.5%	95%	90%	10%	5%	2.5%	1%
$\Delta y_{t-1}$	50	-2.6570	-2.2911	-1.9906	-1.6495	0.9819	1.3791	1.7118	2.1186
	100	-2.6014	-2.2417	-1.9558	-1.6287	0.9705	1.3715	1.7049	2.1020
	200	-2.5903	-2.2322	-1.9451	-1.6129	0.9589	1.3460	1.6789	2.0671
	500	-2.5836	-2.2485	-1.9589	-1.6248	0.9555	1.3457	1.6796	2.0693
	1000	-2.5545	-2.2183	-1.9306	-1.6091	0.9585	1.3506	1.6912	2.0912

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

Table 12: Critical values for non-standard unit root tests under the null (B.3) with  $s_t = y_{t-1}$ .

$s_t$	$T$	Probability that $F$ is greater than entry							
		99%	97.5%	95%	90%	10%	5%	2.5%	1%
$y_{t-1}$	50	0.0589	0.1369	0.2450	0.4006	3.5804	4.4741	5.3629	6.5900
	100	0.0610	0.1443	0.2510	0.4112	3.4751	4.3232	5.1667	6.2926
	200	0.0601	0.1447	0.2550	0.4086	3.4880	4.3158	5.1192	6.2031
	500	0.0623	0.1472	0.2563	0.4094	3.4468	4.2620	5.0649	6.0750
	1000	0.0606	0.1456	0.2562	0.4137	3.4496	4.2618	5.0486	6.0007

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

$s_t$	$T$	Probability that $t$ is less than entry							
		99%	97.5%	95%	90%	10%	5%	2.5%	1%
$y_{t-1}$	50	-2.3973	-2.0115	-1.6624	-1.2137	2.4042	2.7741	3.0839	3.4591
	100	-2.3652	-2.0041	-1.6580	-1.2083	2.3872	2.7292	3.0361	3.3724
	200	-2.3936	-2.0031	-1.6508	-1.2129	2.3871	2.7338	3.0213	3.3567
	500	-2.3839	-1.9835	-1.6379	-1.2100	2.3811	2.7128	3.0059	3.3445
	1000	-2.3833	-2.0029	-1.6512	-1.2085	2.3817	2.7167	3.0060	3.3290

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

against

$$H_1^{aux} : y_t = \tilde{\phi}_1^* y_{t-1} + \phi_2^* s_t y_{t-1} + v_t^* \quad (\text{B.4})$$

( $H_0 : \tilde{\phi}_1^* = 1, \phi_2^* = 0$ ). Then we consider (B.2) with the null hypothesis of  $\phi_1^* = 1$  and (B.4) with  $\tilde{\phi}_1^* = 1$ . Critical values are in Table 11-12.

### C. Figures

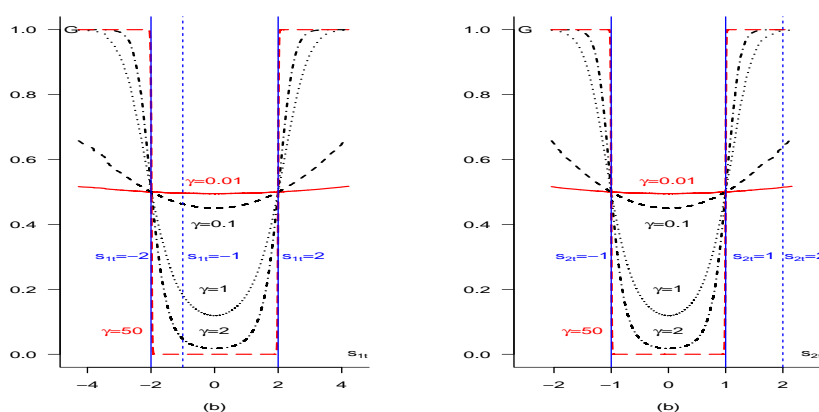


Figure 5: Smooth-transition function  $G$  in  $s_{1t} \sim N(0, 1)$  in panel (a) and  $s_{2t} \sim N(0, 0.5)$  in panel (b), when  $k = 2$  with different both  $s_{jt}$  and  $c_j = \{-1, 2\}$ .

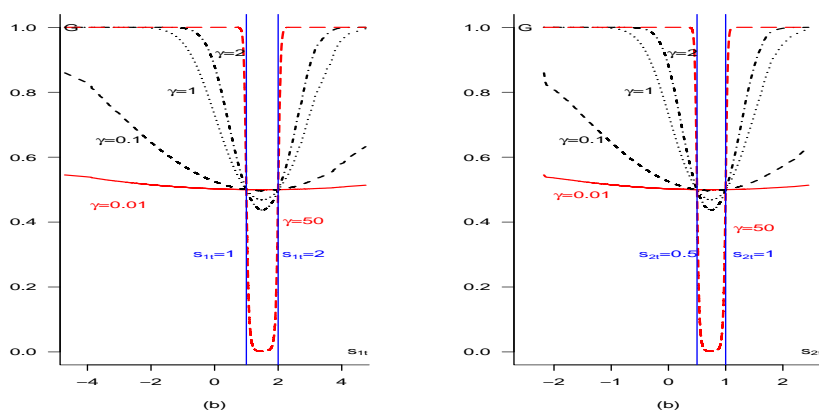


Figure 6: Smooth-transition function  $G$  in  $s_{1t} \sim N(0, 1)$  in panel (a) and  $s_{2t} \sim N(0, 0.5)$  in panel (b), when  $k = 2$  with different both  $s_{jt}$  but same  $c_j = 1$ .

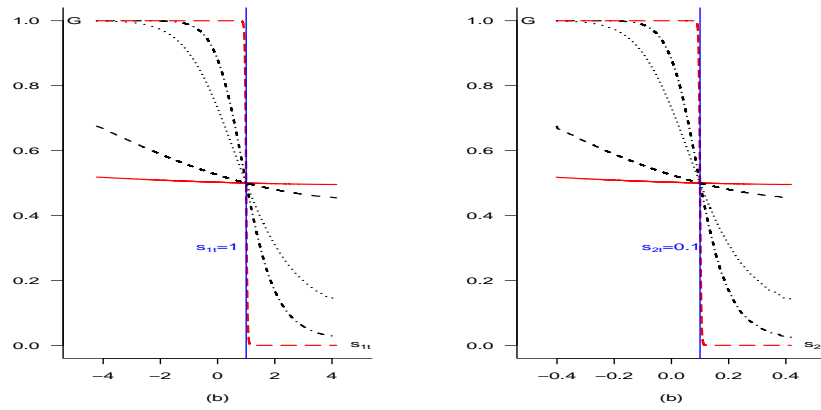


Figure 7: Smooth-transition function  $G$  in  $s_{1t} \sim N(0, 1)$  in panel (a) and  $s_{2t} \sim N(0, 0.1)$  in panel (b), when  $k = 2$  with different both  $s_{jt}$  but same  $c_j = 1$ .

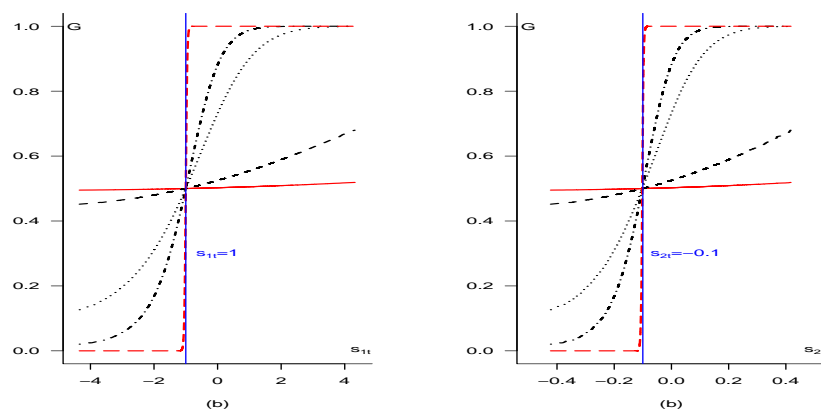


Figure 8: Smooth-transition function  $G$  in  $s_{1t} \sim N(0, 1)$  in panel (a) and  $s_{2t} \sim N(0, 0.1)$  in panel (b), when  $k = 2$  with different both  $s_{jt}$  but same  $c_j = -1$ .