NONLINEAR COINTEGRATION IN NONLINEAR VECTOR AUTOREGRESSIVE MODELS

Changli He, Dao Li

2010-06 (3)

Abstract

Existing cointegration studies mainly concern with integrated time series. This is not often applicable while the time series processes are stable process. In this paper, we propose a definition of smooth-transition (ST) type nonlinear cointegration for a group of individually stable time series. We study smooth-transition vector autoregressive (STVAR) model to consider the proposed nonlinear cointegration. Our model is also suitable to study common nonlinear factors in such an econmic system. We study the properties of STVAR models and test for common nonlinear factors and nonlinear cointegration. Simulation studies have been carried out to show the asymptotic characteristic of the tests. Finally, we apply our work to consumption and income data (United States, monthly from 1959:1 to 2010:3), and compare the forecasting results with linear AR model.

Keywords: Smooth transition vector autoregressive model; Nonlinear cointegration; Common features; Forecasts.

1. Introduction

Cointegration becomes a popular economic tool to analyze related integrated time series as a group, since Engle and Granger(1987) formalized application issues for the cointegration approach. However, linear cointegration is not sufficient to describe a lot of nonlinear behaviors in a complex economic system, see He and Li (2009) for a detailed discussion. Granger and Teräsvirta (1993) also discusses many modeling issues for nonlinear economic relations. However, there are still lack of cointegration analyses for individual nonlinear time series, such as stable process. He and Li (2009) proposes two definitions of smooth-transition (ST) type cointegration which allows for regime switching structures in the cointegrated system. It is a more general discussion of nonlinear cointegration which is defined in two different ways for two different cases, i) each individual time series is I(1) and their nonlinear combination is I(0); ii) each individual time series is a nonlinear process (stable) and their linear combination is I(0). The difference of the two definitions is the assumption of individual time series, one is linear $I(1)$, while the other one is stable. The first definition is an extending of linear cointegration in Engle and Granger(1987), which also nests threshold cointegration by Balke and Fomby (1997). He and Li (2009) has studied the ST cointegration and has shown how it works for a group of nonlinear behaviors when linear cointegration cannot describe the supposed economic cointegration relation. In this paper, we focus on the second definition of nonlinear cointegration in He and Li (2009) and apply it to the U.S consumption and income data for forecasting. As we say, this is another situation of nonlinear cointegration, which is not the extending of Engle and Granger's linear cointegration anymore. Actually, it is more like a compliment of such a linear/nonlinear cointegration whose basement is Engle and

Granger (1987)'s idea for a group of individual linear integrated time series.

Consumer and income price indexes are primary measures of inflation in modern economic system. Therefore it is crucial to study the relationship between the two dynamic time series. The permanent-income hypothesis (see e.g. Campbell and Mankiw, 1990 among others) and the model in Engle and Granger (1987) imply that consumption and income are (linear) cointegrated so that savings is stationary. Our definition of nonlinear cointegration studied in this paper is a few considered case with subject to the different assumption for individual time series. Although the first definition in He and Li (2009) nests the most popular cointegration since Engle and Grander (1987), both linear cointegration and nonlinear cointegration start from the assumption of individual linear $I(1)$ processes. Obviously, a doubt has raised up. Once each process is not a linear $I(1)$, isn't there cointegration type relationship? But many economic hypothesis support such kinds of relationship in the groups of macroeconomic variables. Therefore the definition in this paper is necessarily considered, which starts from individual stable process instead of linear unit root. To our knowledge, few literature has studied whether or not such a nonlinear cointegration feature or common nonlinear factor presents in consumer and income indexes especially in the proposed ST vector autoregressive (STVAR) models. We aim to forecast consumption by analyzing the relation between consumption and income by STVAR model. The different starting point gives the chance to study smooth-transition vector autoregressive (STVAR) models with its statistical properties. However, the parameter estimation of the STVAR model can be a big challenge. Bad estimation may make us to give up such a nonlinear model, but we do not want to accept less bad results from a incorrect linear model. We also show some helpful suggestions for improving the parameters' stability and consistency based on the data, and develop a residual-based test for common nonlinear factors and nonlinear cointegration for our consumption and income data.

The paper is organized as follows. We introduce the definition of nonlinear cointegration in Section 2. Section 3 studies the STVAR model and its properties. A residual-based test is proposed in Section 4 and we study the finite-sample distribution and small sample properties of the test in section 5. In the final Section 6, we analyze the data of consumption expenditures and disposable income from the United States, monthly 1959-2010 and forecast four horizons of 1-month, 4-month, 8-month and 12-month, compared with linear model.

2. Definition of Nonlinear Cointegration

The definition in this section is a continuity studying of ST type nonlinear cointegration in He and Li (2009). We focus on the second definition in that paper which is now illustrated as follows. But due to the different assumption for individual time series, we introduce the smooth-transition autoregressive model of order p ($STAR(p)$ model) firstly to describe a nonlinear and global stationary time series y_t such that,

$$y_t = \pi_{10} + \boldsymbol{\pi}_1'\mathbf{w}_t + \boldsymbol{\pi}_2'\mathbf{w}_t \cdot G(\mathbf{s}_t; \gamma, \mathbf{c}) + v_t \tag{2.1}$$

where $\mathbf{w}_t = (y_{t-1}, y_{t-2}, ..., y_{t-p})'$, $\boldsymbol{\pi}_i'$ are vector parameters and π_{10} is a scalar parameter, v_t is the error term; and $G(\mathbf{s}_t; \gamma, \mathbf{c})$ is a smooth-transition (ST) function

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

where γ is the slope parameter, c_j are location parameters and s_{jt} are transition variables. For detailed model interpretation see Teräsvirta (1994) and ST function property see Section 3.

Definition. *Let* $\mathbf{y}_t = (y_{1t}, y_{2t}, ..., y_{nt})'$ *be an* $(n \times 1)$ *vector time series where each of the series taken individually is a nonlinearly stationary process, i.e.* $y_{it} \sim STAR(p)$ *in (2.1). The vector time series* $\mathbf{y}_t$ *is said to be nonlinearly cointegrated if there exists a nonzero* $(n \times 1)$ *constant vector* $\boldsymbol{\alpha} = (\alpha_1, \alpha_2, ..., \alpha_n)'$ *such that the combination of the series* $\mathbf{y}_t$ *is linear stationary, that is,* $\boldsymbol{\alpha}'\mathbf{y}_t \sim I(0)$.

Furthermore, we also consider the STAR model with deterministic time trend in this paper,

$$y_t = \pi_{10} + \boldsymbol{\pi}_1'\mathbf{w}_t + \boldsymbol{\pi}_2'\mathbf{w}_t \cdot G\left(\mathbf{s}_t; \gamma, \mathbf{c}\right) + \delta \cdot t + v_t \tag{2.3}$$

for individual time series assumption, which is called as trend-stable process. In this case, *if there exists a nonzero* $(n \times 1)$ *constant vector* $\boldsymbol{\alpha} = (\alpha_1, \alpha_2, ..., \alpha_n)'$ *such that the combination of the series* $\mathbf{y}_t$ *is linearly trend-stationary, the system has common nonlinear factors*. Actually, a cointegrated system is one of the situations with common factors, see Engle and Kozicki (1993). *If there exists a nonzero vector* $\boldsymbol{\alpha}$ *to make the combination of* $\mathbf{y}_t$ *to be a linear stationary process, the system will be nonlinearly cointegrated*, which nests common nonlinear factors based on model (2.3), while is equivalent to common nonlinear factors based on model (2.1).

The point in this definition of nonlinear cointegration is that we are concerned whether or not the nonlinear factors disappear. Therefore, two different basements for individual time series in a group, with or without trend, are ready for the true processes depending on the behavior of data. Testing nonlinear cointegration or common nonlinear factors has similar guide in this paper. We will achieve this by analyzing our STVAR models in Section 3.

3. Smooth-transition Vector Autoregressive Models

This smooth-transition vector autoregressive model of order p (STVAR(p)) in (3.1) is a general vector case of STAR(p) models in (2.1) for a system $\mathbf{y}_t = (y_{1t}, y_{2t}, ..., y_{nt})'$, which regresses on its own lagged values $\mathbf{y}_{t-1}, \mathbf{y}_{t-2}, ..., \mathbf{y}_{t-p}$ both in linear and nonlinear components,

$$\mathbf{y}_t = \boldsymbol{\mu} + \sum_{k=1}^{p} \boldsymbol{\Phi}_k \mathbf{y}_{t-k} + \sum_{k=1}^{p} \boldsymbol{\Gamma}_k \mathbf{y}_{t-k} G\left(s_t; \gamma, \mathbf{c}\right) + \mathbf{v}_t \tag{3.1}$$

where $\boldsymbol{\mu}_t = (\mu_{1t}, ..., \mu_{nt})'$ is the constant term, $\boldsymbol{\Phi}_k$ and $\boldsymbol{\Gamma}_k$ are $(n \times n)$ coefficient matrices, $G\left(s_t; \gamma, \mathbf{c}\right)$ is a ST function in (2.2), and $\mathbf{v}_t = (v_{1t}, ..., v_{nt})'$ is the error term.

A further extending from (3.1) by adding a time trend is that

$$\mathbf{y}_t = \boldsymbol{\mu} + \sum_{k=1}^{p} \boldsymbol{\Phi}_k \mathbf{y}_{t-k} + \sum_{k=1}^{p} \boldsymbol{\Gamma}_k \mathbf{y}_{t-k} G\left(s_t; \gamma, \mathbf{c}\right) + \boldsymbol{\delta} \cdot t + \mathbf{v}_t \tag{3.2}$$

where $\boldsymbol{\delta}$ is a $(n \times 1)$ coefficient vector in addition. Consider a nonzero combination of $\mathbf{y}_t$ as

$$\begin{aligned} \boldsymbol{\alpha}'\mathbf{y}_t &= \boldsymbol{\alpha}'\boldsymbol{\mu} + \sum_{k=1}^{p} \boldsymbol{\alpha}'\boldsymbol{\Phi}_k \mathbf{y}_{t-k} + \sum_{k=1}^{p} \boldsymbol{\alpha}'\boldsymbol{\Gamma}_k \mathbf{y}_{t-k} G\left(s_t; \gamma, \mathbf{c}\right) + \boldsymbol{\alpha}'\boldsymbol{\delta} \cdot t + \boldsymbol{\alpha}'\mathbf{v}_t \\ &= \mu^* + \sum_{k=1}^{p} \boldsymbol{\Phi}_k^* \mathbf{y}_{t-k} + \sum_{k=1}^{p} \boldsymbol{\Gamma}_k^* \mathbf{y}_{t-k} G\left(s_t; \gamma, \mathbf{c}\right) + \delta^* \cdot t + v_t^* \end{aligned} \tag{3.3}$$

by an $(n \times 1)$ nonzero vector $\boldsymbol{\alpha}$, where $\boldsymbol{\Phi}_k^* = \boldsymbol{\alpha}'\boldsymbol{\Phi}_k$ and $\boldsymbol{\Gamma}_k^* = \boldsymbol{\alpha}'\boldsymbol{\Gamma}_k$ are $1 \times n$ vector parameters, $\mu^* = \boldsymbol{\alpha}'\boldsymbol{\mu}$ and $\delta^* = \boldsymbol{\alpha}'\boldsymbol{\delta}$ are scalar parameters, and $v_t^* = \boldsymbol{\alpha}'\mathbf{v}_t$ is the error term. In (3.3), we have following possible cases,

- When $\delta^* = 0$, the system $\mathbf{y}_t$ contains common trend, actually, there must be such a nonzero $\boldsymbol{\alpha}$ to make $\delta^* = 0$ if $\boldsymbol{\delta}$ is a real-value vector;

- When $\mathbf{\Gamma}_k$ has reduced rank for all $k = 1, 2, ., p$, we can surely find out a nonzero $\boldsymbol{\alpha}$ to make all $\mathbf{\Gamma}_k^* = \mathbf{0}'$ workable, then the system $\mathbf{y}_t$ contains common nonlinear factors; when it also makes $\delta^* = 0$ at the same time, $\mathbf{y}_t$ has nonlinear cointegration, nesting common nonlinear factors;

- When $\mathbf{\Phi}_k$ has reduced rank, $\mathbf{\Phi}_k^* = \mathbf{0}'$ is then workable, the system $\mathbf{y}_t$ contains common linear factors, but we are not interested in this case.

By the same reason, when all $\mathbf{\Gamma}_k$ in model (3.1) has reduced rank, the system $\mathbf{y}_t$ is nonlinearly cointegrated based on our definition in this paper. When $\mathbf{\Phi}_1 = \mathbf{I}_n$, the model (3.1) can present a combination proceess by unit root plus nonlinear component in (3.4),

$$\mathbf{y}_t = \boldsymbol{\mu} + \mathbf{I}_n \mathbf{y}_{t-1} + \sum_{k=2}^{p} \mathbf{\Phi}_k \Delta \mathbf{y}_{t-k} + \sum_{k=1}^{p} \mathbf{\Gamma}_k \Delta \mathbf{y}_{t-k} G\left(s_t; \gamma, \mathbf{c}\right) + \mathbf{v}_t \tag{3.4}$$

where $\mathbf{I}_n$ is the $(n \times n)$ identity matrix. It may be meaningfully called nonlinear unit root process.

The smooth transition function $G(\mathbf{s}_t; \gamma, \mathbf{c})$ has featured statistical properties. When $k = 1$, $G(\mathbf{s}_t; \gamma, \mathbf{c})$ in (2.2) is a non-decreasing nonlinear function bounded between zero and one, see discussion in He and Li (2009). But in this paper, we discuss more for choices of ST functional forms which should be available for wider applications. When the number of values of k is more than one, we need to consider for further different transition variables s_{jt} and location parameters c_j for different index j $(j = 1, 2, ..., k)$. The simplest case is that they are the same for different j. In this situation, when k is an odd, the behavior of function $G(\mathbf{s}_t; \gamma, \mathbf{c})$ is similar with the case when $k = 1$ but changed more steeply; when k is a even, the value of $G(\mathbf{s}_t;\gamma,\mathbf{c})$ is bounded between 0.5 and 1. Suppose s_{jt} $(j=1,2)$ is a random variable from the standard normal distribution and $c_1=c_2=1$, such that,

$$\begin{aligned} G\left(\mathbf{s}_t;\gamma,\mathbf{c}\right) &= \left(1+exp\left\{-\gamma\cdot(s_t-c)^2\right\}\right)^{-1}, \quad \gamma>0 \\ G\left(\mathbf{s}_t;\gamma,\mathbf{c}\right) &= \left(1+exp\left\{-\gamma\cdot(s_t-c)^3\right\}\right)^{-1}, \quad \gamma>0. \end{aligned} \tag{3.5}$$

The property of $G(\mathbf{s}_t;\gamma,\mathbf{c})$ when $k=2$ and $k=3$ are presented in Figure 1. The panel (a) and (b) in Figure 1 has the extreme case of a point regime in the middle when γ is sufficient large.

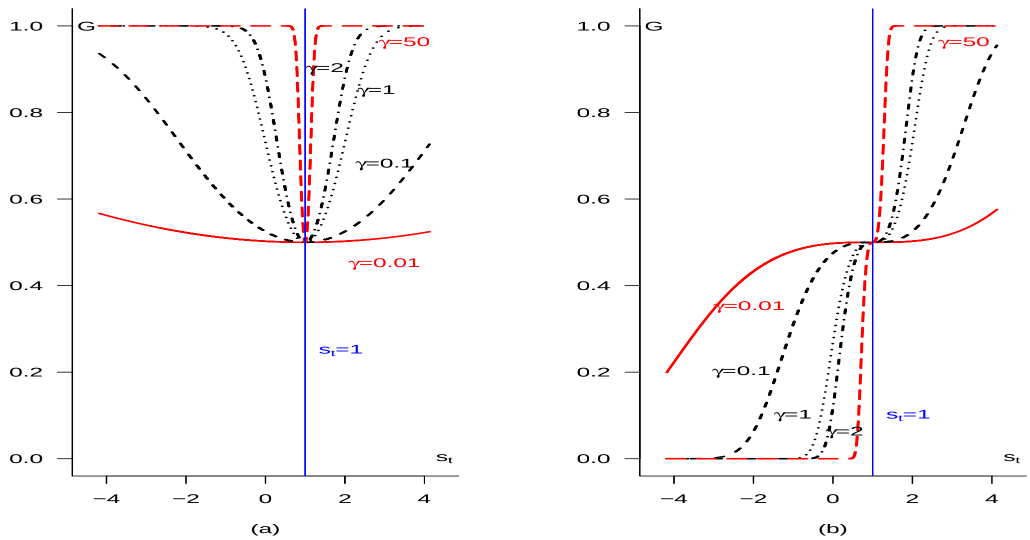

Figure 1: Smooth-transition function G when $k=2$ in the panel (a) and $k=3$ in the panel (b), with same $s_{jt}\sim N(0,1)$ and $c_j=1$.

If the location parameters in $G(\mathbf{s}_t;\gamma,\mathbf{c})$ are not same when $k>1$, such that,

$$\begin{aligned} G\left(\mathbf{s}_t;\gamma,\mathbf{c}\right) &= \left(1+exp\left\{-\gamma\cdot(s_t-c_1)(s_t-c_2)\right\}\right)^{-1}, \quad \gamma>0 \\ G\left(\mathbf{s}_t;\gamma,\mathbf{c}\right) &= \left(1+exp\left\{-\gamma\cdot(s_t-c_1)(s_t-c_2)(s_t-c_3)\right\}\right)^{-1}, \quad \gamma>0 \end{aligned} \tag{3.6}$$

Although the situation becomes more complicated but easy to understand. We show that in Figure 2. The number of regimes when γ goes to infinity equals to $k+1$. When $k>1$, $G(\mathbf{s}_t;\gamma,\mathbf{c})$ is still reduced to a constant when γ is equal to zero.

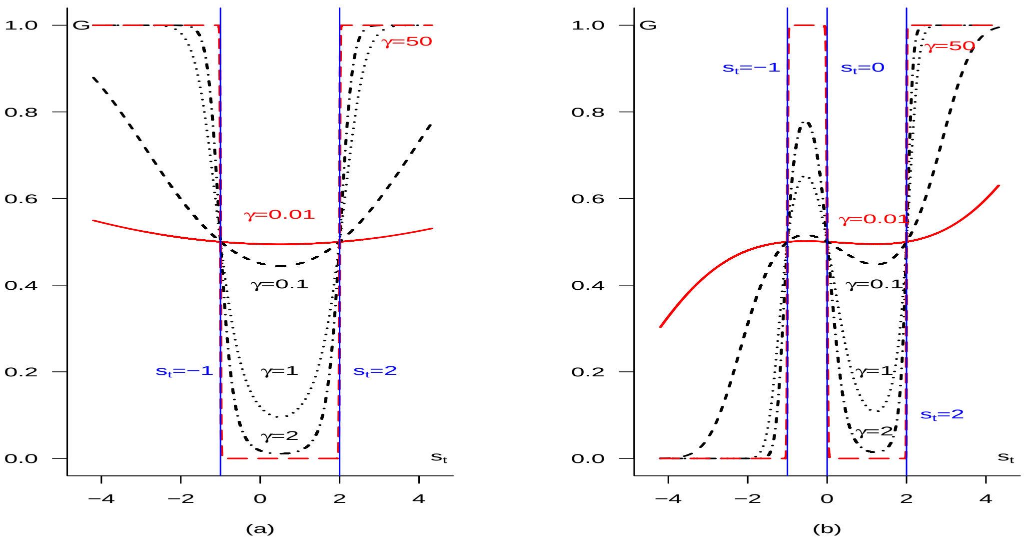

Figure 2: Smooth-transition function G when $k=2$ in the panel (a) and $k=3$ in the panel (b), with same $s_{jt}\sim N(0,1)$ but different $c_j=\{-1,2\}$ and $c_j=\{-1,0,2\}$.

And it also jumps sharply from one regime to another when γ goes to infinity, for example, jumping down/up from $s_t<c_1$ to $c_1<s_t<c_2$ then jumping up/down to $s_t>c_2$ when $k=2$.

Suppose $s_{1t}\sim N(0,1)$ and $s_{2t}\sim N(0,0.5)$ in another case with different transition variables when $k>1$. Taking $k=2$ as an example,

$$G\left(\mathbf{s}_t;\gamma,\mathbf{c}\right)=\left(1+exp\left\{-\gamma\cdot\left(s_{1t}-c_1\right)\left(s_{2t}-c_2\right)\right\}\right)^{-1},\quad \gamma>0 \tag{3.7}$$

in Figure 3, not all the location parameter c_j give the point in time where the transition is symmetric around c_j. Since it is hard to say how weighted the different transition variables should be to show each c_j give the transition point in time.

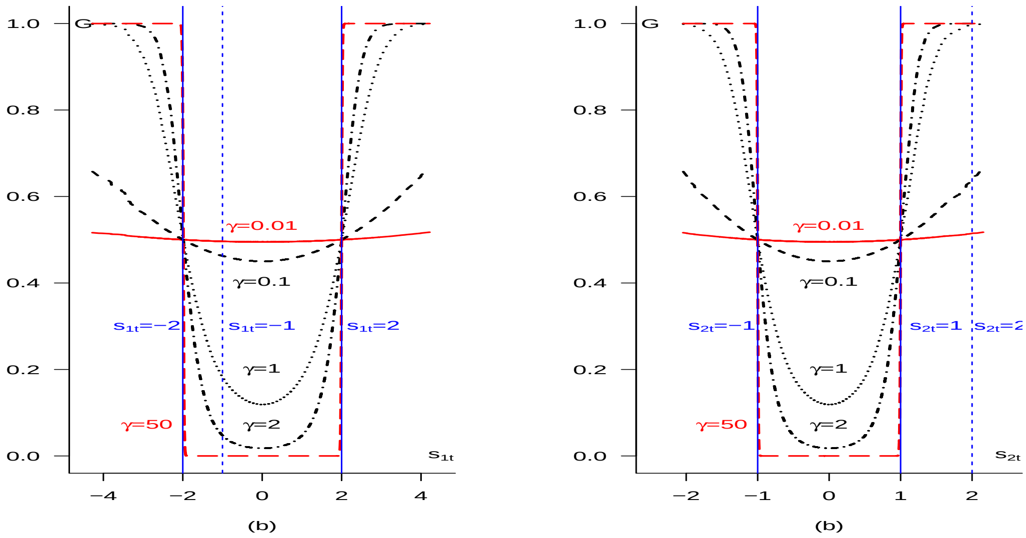

Figure 3: Smooth-transition function G in $s_{1t} \sim N(0,1)$ in the panel (a) and $s_{2t} \sim N(0,0.5)$ in the panel (b), when $k=2$ with different both s_{jt} and $c_j = \{-1, 2\}$.

Under the same assumption of transition variables, the situation does not become easier when $c_1 = c_2 = c$. The reason is that the value of c and the distribution of those different transition variables are highly effected by each other. When $k = 2$,

$$G(\mathbf{s}_t; \gamma, \mathbf{c}) = \left(1 + exp\left\{-\gamma \cdot (s_{1t} - c)(s_{2t} - c)\right\}\right)^{-1}, \quad \gamma > 0 \tag{3.8}$$

we usually have the behavior in Figure 4 when the value of c is in the range of smallest s_{jt}.

But if the value of c is outside some ranges of s_{jt}, Figure 5 and Figure 6 show the changes. There is no $k+1$ regimes anymore when γ goes to infinity. The regimes are reduced. The values of $G(\mathbf{s}_t; \gamma, \mathbf{c})$ in Figure 5 and Figure 6 are symmetric acoording to $G = 0.5$.

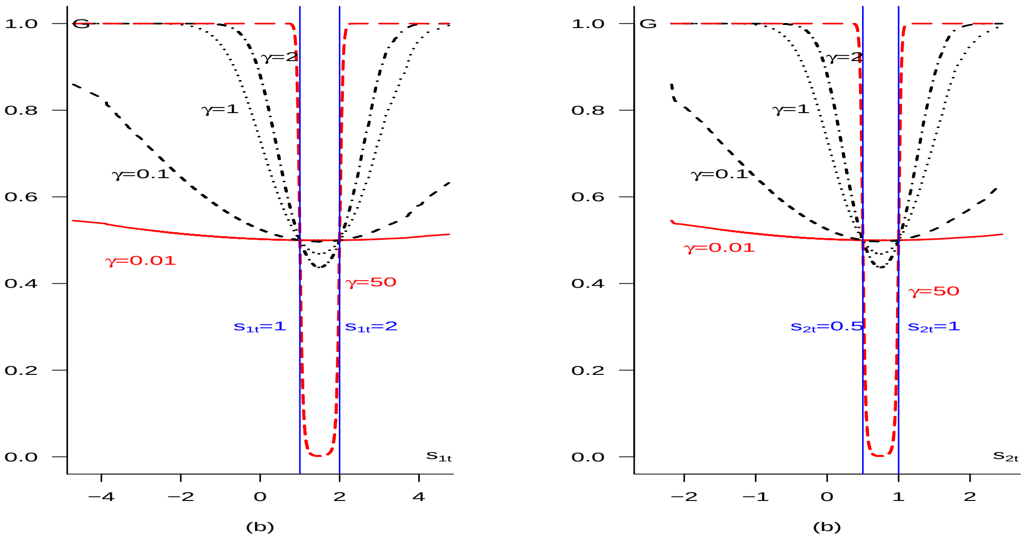

Figure 4: Smooth-transition function G in $s_{1t} \sim N(0,1)$ in the panel (a) and $s_{2t} \sim N(0,0.5)$ in the panel (b), when $k=2$ with different both s_{jt} but same $c_j = 1$.

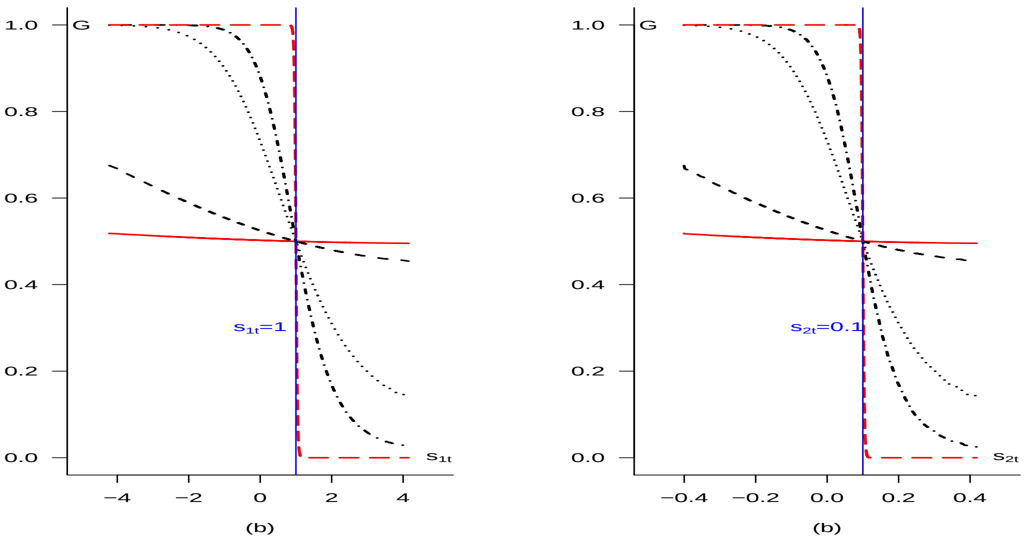

Figure 5: Smooth-transition function G in $s_{1t} \sim N(0,1)$ in the panel (a) and $s_{2t} \sim N(0,0.1)$ in the panel (b), when $k=2$ with different both s_{jt} but same $c_j = 1$.

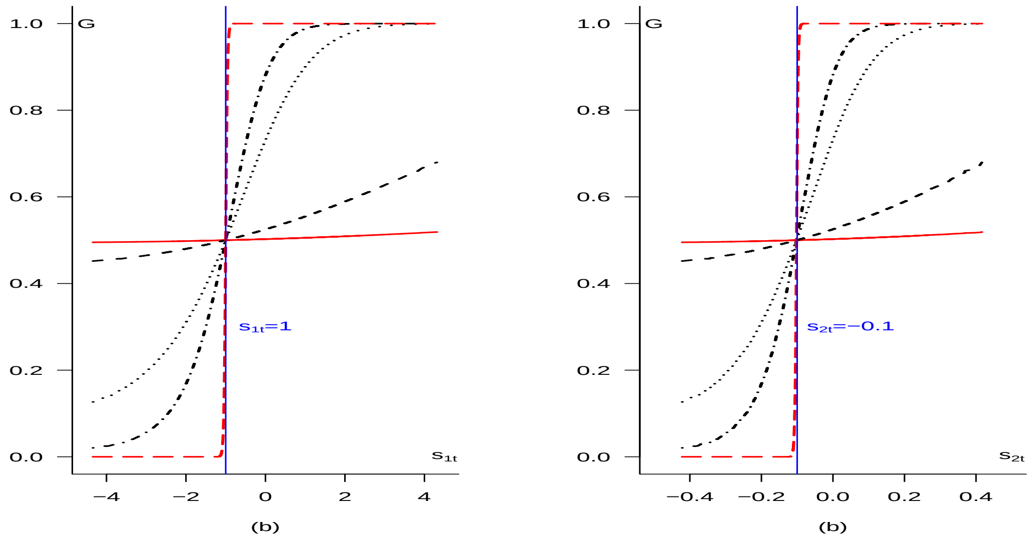

Figure 6: Smooth-transition function G in $s_{1t} \sim N(0,1)$ in the panel (a) and $s_{2t} \sim N(0,0.1)$ in the panel (b), when $k=2$ with different both s_{jt} but same $c_j = -1$.

4. Residual-based Tests for Common Nonlinear Factors

When a system $\mathbf{y}_t$ in (3.2) contains common nonlinear factors, the linear combination of $\mathbf{y}_t$ does not include nonlinear behaviors. A residual-based test for common nonlinear factors focuses on the estimated residuals $\hat{u}_t = y_{1t} - \hat{\boldsymbol{\alpha}}'\mathbf{y}_{2t}$ from regressing y_{1t} on $\mathbf{y}_{2t} = \{y_{2t}, ..., y_{nt}\}$. The big advantage of the method is easy to study one process $\hat{u}_t$ instead of multivariate case in (3.3). The null hypothesis is that there exists common nonlinear factors in a group of random variables $\mathbf{y}_t$. In other words, there are no nonlinear components in $\hat{u}_t$.

We reasonablly consider similar nonlinear component in the model of estimated $\hat{u}_t$, such that

$$\hat{u}_t = \beta_0 + \sum_{k=1}^{p} \beta_{1p}\hat{u}_{t-p} + \sum_{k=1}^{p} \beta_{2p} G^*(s_t;\gamma,\mathbf{c})\hat{u}_{t-p} + \beta_3 t + e_t. \tag{4.1}$$

Under the null hypothesis of linearity with $\gamma = 0$, the parameter $\mathbf{c}$ are not identified. As recommended by Luukkonen, Saikkonen and Teräsvirta (1988), it is an efficient way to approximately represent the model (4.1) to be the auxiliary one,

$$\hat{u}_t = \beta_0 + \sum_{k=1}^{p} \beta_{1k}^{*} \hat{u}_{t-k} + \sum_{k=1}^{p} \beta_{2k}^{*} \hat{u}_{t-q} \hat{u}_{t-k} + \beta_3 t + e_t \tag{4.2}$$

by first-order Taylor expansion around $\gamma = 0$, where q probably is one of the elements in set $\{1, 2, ..., p\}$. Sequentially, the parameter restriction is rewritten to be

$$H_{01} : \beta_{2p}^{*} = 0 \text{ for all } p \tag{4.3}$$

imposing into the model (4.2). This is equivalent to

$$H_{01} : \mathbf{R}\boldsymbol{\beta}^{*} = \mathbf{r} \tag{4.4}$$

where $\mathbf{R} = \left(\mathbf{0}_{p\times(p+1)}, \mathbf{I}_{p\times p}, \mathbf{0}_{p\times 1}\right)$, $\boldsymbol{\beta}^{*} = (\beta_0, \beta_{11}^{*}, ..., \beta_{1p}^{*}, \beta_{21}^{*}, ..., \beta_{2p}^{*}, \beta_3)'$, and $\mathbf{r} = \mathbf{0}_{p\times 1}$.

Testing for nonlinear cointegration is achieved by testing those null hypothesis together with $\beta_3 = 0$ as a joint null, such that,

$$H_{02} : \beta_{2p}^{*} = 0 \text{ for all } p \text{ and } \beta_3 = 0 \tag{4.5}$$

which is equivalent to

$$H_{02} : \mathbf{R}\boldsymbol{\beta}^{*} = \mathbf{r} \tag{4.6}$$

where $\mathbf{R} = \left(\mathbf{0}_{p\times(p+1)}, \mathbf{I}_{p\times(p+1)}\right)$, $\boldsymbol{\beta}^{*} = (\beta_0, \beta_{11}^{*}, ..., \beta_{1p}^{*}, \beta_{21}^{*}, ..., \beta_{2p}^{*}, \beta_3)'$, and $\mathbf{r} = \mathbf{0}_{(p+1)\times 1}$.

Both for those two hypothesis, a Wald-type test is given by

$$W_T = \left(\mathbf{R}\boldsymbol{\Upsilon}_T\left(\hat{\boldsymbol{\beta}}^*_T - \boldsymbol{\beta}^*\right)\right)'\left\{s_T^2\mathbf{R}\boldsymbol{\Upsilon}_T\left(\sum \mathbf{x}_t\mathbf{x}_t'\right)^{-1}\boldsymbol{\Upsilon}_T\mathbf{R}'\right\}^{-1}\mathbf{R}\boldsymbol{\Upsilon}_T\left(\hat{\boldsymbol{\beta}}^*_T - \boldsymbol{\beta}^*\right) \tag{4.7}$$

where $\hat{\boldsymbol{\beta}}^*_T$, which is carried out by minimizing the sum of squared residuals in the model (4.2), is the consistent estimator of true parameter $\boldsymbol{\beta}^*$ under the null hypothesis, $\boldsymbol{\Upsilon}_T$ is a known square matrix of convergence rates, $\mathbf{R}$ is a $(p \times (2p+2))$ matrix restricted by the null hypothesis. The two tests are probably asymptotic χ^2 distributions. See related theorems for (4.2) below.

Theorem 1. *Consider model (4.2). The tests W_T under the null H_{01} converges, since*

$$T^{\frac{1}{2}}\left(\hat{\boldsymbol{\beta}}^*_T - \boldsymbol{\beta}^*\right) \xrightarrow{L} \mathbf{Q}_1^{-1}\mathbf{h}_1 \tag{4.8}$$

holds under H_{01}, which is a multivariate normal distribution, where

$$\boldsymbol{\Upsilon}_T = diag(T^{1/2}\mathbf{I}_{(2p+1)\times(2p+1)}, T^{3/2}),$$

$\mathbf{h}_1$ *is normally distributed and $\mathbf{Q}_1^{-1}$ exists. But $\mathbf{Q}_1^{-1}\mathbf{h}_1$ is dependent on transition variable $s_t = \hat{u}_{t-q}$ in (4.2).*

Proof of Theorem 1. See Appendix.

□

Theorem 2. *Consider model (4.2). The tests W_T under the null H_{02} converges, since*

$$T^{\frac{1}{2}}\left(\hat{\boldsymbol{\beta}}^*_T - \boldsymbol{\beta}^*\right) \xrightarrow{L} \mathbf{Q}_2^{-1}\mathbf{h}_2 \tag{4.9}$$

holds under H_{02}, which is a multivariate normal distribution, where

$$\boldsymbol{\Upsilon}_T = diag(T^{1/2}\mathbf{I}_{(2p+1)\times(2p+1)}, T^{3/2}),$$

$\mathbf{h}_2$ *is normally distributed and* $\mathbf{Q}_2^{-1}$ *exists. But* $\mathbf{Q}_2^{-1}\mathbf{h}_2$ *is dependent on transition variable* $s_t = \hat{u}_{t-q}$ *in (4.2).*

Proof of Theorem 2. See Appendix.

□

5. Simulation

The Monto Carlo experiments in this section study the finite-sample distributions and small-sample properties of the Wald-type tests in (4.7). The shown simulation results are designed when $n = 2$ in the system $\mathbf{y}_t$ for our application of consumption and income. Of course, the resutls can be easily and surely extended into general cases in the same way. The number of the replication of each experiment is 100,000.

The data generating process (DGP) under the null hypothesis H_{01} is as follows in (5.1),

$$\begin{aligned}
\begin{pmatrix} y_{1t} \\ y_{2t} \end{pmatrix} &= \begin{pmatrix} 0.4 \\ 0.6 \end{pmatrix} + \begin{pmatrix} 0.16 \\ 0.08 \end{pmatrix} t + \begin{pmatrix} 0.3 & 0.05 \\ 0.2 & 0.1 \end{pmatrix} \begin{pmatrix} y_{1t-1} \\ y_{2t-1} \end{pmatrix} \\
&+ \begin{pmatrix} 0.3 & 0.05 \\ 0.2 & 0.1 \end{pmatrix} \begin{pmatrix} y_{1t-2} \\ y_{2t-2} \end{pmatrix} + \begin{pmatrix} 0.3 & 0.05 \\ 0.2 & 0.1 \end{pmatrix} \begin{pmatrix} y_{1t-3} \\ y_{2t-3} \end{pmatrix} + \begin{pmatrix} 0.3 & 0.05 \\ 0.2 & 0.1 \end{pmatrix} \begin{pmatrix} y_{1t-4} \\ y_{2t-4} \end{pmatrix} \\
&+ \begin{pmatrix} 2 \\ 1 \end{pmatrix} \begin{pmatrix} 0.12 & 0.18 \end{pmatrix} \begin{pmatrix} y_{1t-1} \\ y_{2t-1} \end{pmatrix} \frac{1}{1 + \exp\left(-20\left(s_t - 1\right)\right)} + \begin{pmatrix} v_{1t} \\ v_{2t} \end{pmatrix}
\end{aligned} \tag{5.1}$$

with the restriction of $\mathbf{\Gamma}_k$ having reduced rank, where $\mathbf{y}_t = (y_{1t}, y_{2t})'$ has common nonlinear factor under the null. But under H_{02}, the DGP has slight change of time trend.

Then we estimate the combination vector to obtain the residuals $\hat{u}_t$ from $y_{1t} - \hat{\alpha} y_{2t}$. Under the null hypothesis, we construct the Wald-type tests W in (4.7) from the auxiliary regression,

$$\hat{u}_t = \beta_0 + \beta_{11}^* \hat{u}_{t-1} + \beta_{12}^* \hat{u}_{t-2} + \beta_{13}^* \hat{u}_{t-3} + \beta_{14}^* \hat{u}_{t-4} + \beta_2^* \hat{u}_{t-1} s_t^* + \beta_3^* t + e_t. \tag{5.2}$$

The approximate critical values of the Wald-type tests are presented in the following Table 1 and Table 2, respectively for test H_{01} and H_{02}. For different s_t, we show corresponded results in Appendix.

Table 1: Critical values for non-standard unit root tests under H_{01}.

		Probability that F is greater than entry							
s_t	T	99%	97.5%	95%	90%	10%	5%	2.5%	1%
u_{t-1}	50	0.00	0.00	0.00	0.02	3.00	4.29	5.62	7.49
	100	0.00	0.00	0.00	0.02	3.13	4.44	5.83	7.72
	200	0.00	0.00	0.00	0.02	3.04	4.33	5.62	7.43
	500	0.00	0.00	0.00	0.02	2.88	4.09	5.30	7.02
	1000	0.00	0.00	0.00	0.02	2.79	3.94	5.13	6.79

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

Table 2: Critical values for non-standard unit root tests under H_{02}.

		Probability that F is greater than entry							
s_t	T	99%	97.5%	95%	90%	10%	5%	2.5%	1%
u_{t-1}	50	0.00	0.00	0.00	0.01	2.63	3.76	5.01	6.80
	100	0.00	0.00	0.00	0.02	2.65	3.80	5.00	6.58
	200	0.00	0.00	0.00	0.02	2.65	3.76	4.93	6.53
	500	0.00	0.00	0.00		2.70	3.84	5.03	6.67
	1000	0.00	0.00	0.00	0.01	2.69	3.83	5.03	6.60

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

6. Application

6.1 Data

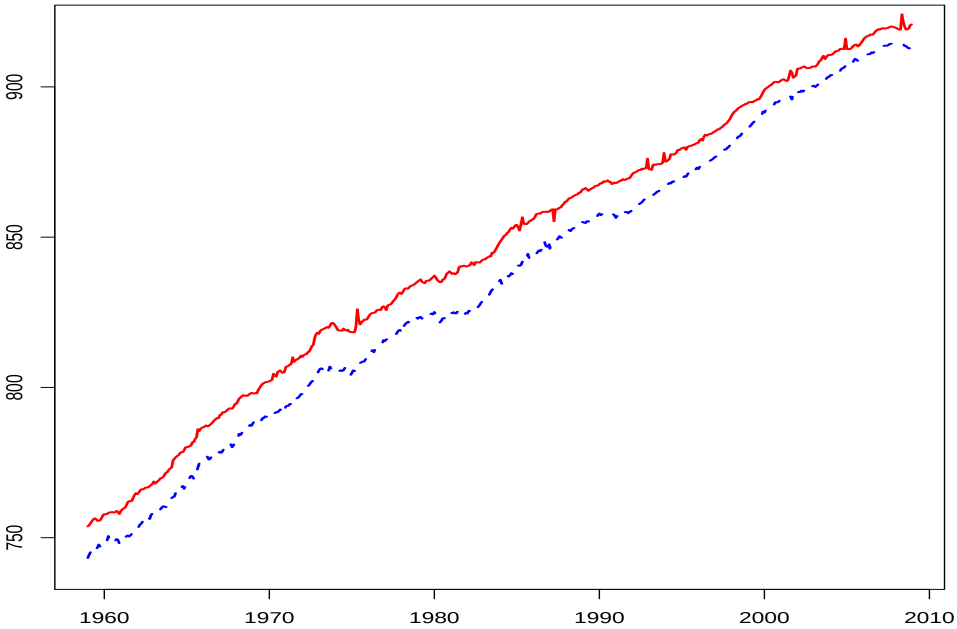

Figure 7: One hundred times the log of personal consumption expenditures (y_{1t}, blue) and personal disposable income (y_{2t}, red) for the United States in billions of 2005 dollars, monthly, 1959:1-2010:3.

The personal consumption expenditures data and and personal disposable income is downloaded from the website of U.S. bureau of economic analysis. We have two monthly time series: real personal consumption expenditures by major type of product (y_1) and the real disposable personal income (y_2) from 1959:1 to 2010:3 (615 observations). Figure 7 presents the date that has been seasonally adjusted at annual rates.

To specify the data to a correct class of time series process is definitely an important start point in model specification before forecasting. Taking standard tests of unit root with alternative regression containing both drift and trend is a reasonable choice for the data. Standing on the linear stationarity point of view, it seems that both y_{1t} and y_{2t} are $I(1)$ processes by the results in Table 3.

Table 3: Standard unit root tests.

Time series	Tests	p-value	Suggestion
y_{1t} Consumption	ADF[1]	**0.293**	unit root
	KPSS[2]	**0.01**	unit root
Δy_{1t}	ADF	**0.01**	stationary
	KPSS	**0.10**	stationary
y_{2t} Income	ADF	**0.72**	unit root
	KPSS	**0.01**	unit root
Δy_{2t}	ADF	**0.01**	stationary
	KPSS	**0.10**	stationary

1 Augmented Dickey-Fuller test of unit root against trend-stationary, see Dickey and Fuller (1979);
2 Kwiatkowski et al. unit root test of stationarity against unit root, see Kwiatkowski et al (1992).

But is that the only character for the data behavior? Here should we remind a crucial issue that it can mislead to wrong test result in the standard unit root test if the true process is actually a stable process, see He and Li (2009). We now consider the global stationary STAR process but time trend is added as the alternative regression for the non-standard unit root test. Under the null of $y_t = \theta_0^* + y_{t-1} + v_t^*$, the test is set up from an auxiliary regression of

$$y_t = \theta_0^* + \theta_1^* y_{t-1} + \theta_2^* s_t y_{t-1} + \theta_3^* t + v_t^* \tag{6.1}$$

with the joint null hypothesis $H_0 : \theta_1^* = 1, \theta_2^* = 0, \theta_3^* = 0$. We also add lag values of Δy_t in (6.1) when doing the estimation to avoid the possible serial correlations in v_t^*. We set both Δy_{t-k} and y_{t-k} as the options of transition variable s_t under the null hypothesis. The results are shown in Table 4. The null is significantly rejected at 5% significant level. Both consumption and income series perform more likely as a stable process with trend instead of a unit root. Therefore it is easy to discovery that the previous results in standard unit root test become no longer reliable.

Table 4: Non-standard unit root tests.

Time series	s_t	F type statistic	Critical value* (5%)
y_{1t} Consumption	Δy_{1t-1}	14.45	**4.73**
	y_{1t-1}	12.25	**5.67**
y_{2t} Income	Δy_{2t-1}	11.44	**4.73**
	y_{2t-1}	13.19	**5.67**

* See Appendix.

6.2 Model Specification

If linear model is adequate to represent the time series by the data, we will not have to consider nonlinear model. In the ST autogressive model, they predetermine the order value of p first and then the form of transition function by testing linearity against nonlinearity. In our model (3.2), STAR with trend, we take this technique to predetermine the lag p and transition variable s_t. We capture the range of the order p by linear autoregression (with time trend) together with an order selection criterion based on AIC. In single autoregressive (AR) model of y_{it}, model (2.3), the consumption process preferably regresses on $y_{1,t-k}(k = 1, 2, 6, 10)$, the income process preferably regresses on $y_{2,t-k}(k = 1, 4)$. But considering vector autoregressive (VAR) model, the consumption process preferably regresses on $y_{1,t-k}(k = 1, 2, 6, 10)$, $y_{2,t-k}(k = 1, 5)$, see Table 5.

Based on the results in Table 5, further specification of selecting the lag p and the transition variable s_t depends on the above non-standard unit root test. From the results in Table 6, first of all, two series are significant nonlinear, although the critical values of different model forms are slightly different. Hereby we present an

Table 5: Selection of order p by linear autoregression.

Time series	AR	VAR
y_{1t} Consumption	$y_{1,t-1}$,$y_{1,t-2}$,$y_{1,t-6}$,$y_{1,t-10}$	$y_{1,t-1}$,$y_{1,t-2}$,$y_{1,t-6}$,$y_{1,t-10}$, $y_{2,t-1}$,$y_{2,t-5}$
y_{2t} Income	$y_{2,t-1}$,$y_{2,t-4}$	–

easy-to-use technique to choose p and s_t by a test for linearity: we choose the one who makes nonlinearity more significant, that is, the one who can reject the test more significantly. Although we include $\Delta y_{i,t-k}$ as an option of s_t, it is not very suitable to select the model when the process is a stable one with trend instead a unit root. Therefore, we find the best choice of transition variable s_t for y_{1t} is y_{1t-1} or y_{2t-1}, and y_{2t-1} for y_{2t}. If y_{1t} and y_{2t} have similar nonlinear component, lag one is a good choice.

6.3 Testing common nonlinear factors

We aim to forecast consumption and income series by ST type autoregressive models. Thus, testing the common nonlinear factors or nonlinear cointegration in the system can be ready for model evaluation, since forecasts in a vector ST type autoregressive model with some certain relations can be more accuracy as if the estimation has no much negative contributions.

When we do the residual-based test, the first step tells us that the estimated relation vector for common nonlinear factor in Section 2 is $\hat{\boldsymbol{\alpha}} = (1, -0.9877)'$. Sequentially, we start to analyze the residual $\hat{u}_t = y_{1t} - 0.9877 y_{2t}$ in Figure 8.

Although we have little information of the form of ST function a priori, there is no problem to just pick the simplest form of $G(\mathbf{s}_t; \gamma, \mathbf{c})$ in (2.2) for our residual-

Table 6: Selection of lag and transition variable by non-standard unit root tests.

Time series	Lags	s_t	F type statistic
y_{1t} Consumption	1	Δy_{1t-1}	14.45
	1	Δy_{1t-2}	10.12
	1	Δy_{2t-1}	16.25
	1	y_{1t-1}	12.25
	1	y_{1t-2}	12.15
	1	y_{2t-1}	**12.68**
	1,2	Δy_{1t-1}	17.39
	1,2	Δy_{1t-2}	11.61
	1,2	Δy_{2t-1}	16.34
	1,2	y_{1t-1}	**18.13**
	1,2	y_{1t-2}	17.90
	1,2	y_{2t-1}	16.72
	1,2,6,10	Δy_{1t-1}	28.66
	1,2,6,10	Δy_{1t-2}	18.39
	1,2,6,10	Δy_{2t-1}	18.02
	1,2,6,10	y_{1t-1}	25.71
	1,2,6,10	y_{1t-2}	25.70
	1,2,6,10	y_{2t-1}	**30.05**
y_{2t} Income	1	Δy_{2t-1}	11.44
	1	Δy_{2t-4}	9.25
	1	y_{2t-1}	**13.19**
	1	y_{2t-4}	12.72
	1,4	Δy_{2t-1}	25.97
	1,4	Δy_{2t-4}	13.35
	1,4	y_{2t-1}	**20.92**
	1,4	y_{2t-4}	19.25

based test, which is,

$$G_1\left(\mathbf{s}_t;\gamma,\mathbf{c}\right) = \left(1+exp\left\{-\gamma(\hat{u}_{t-1}-c)\right\}\right)^{-1}. \tag{6.2}$$

This is because when the nonlinear component is getting more complicated, the nonlinear behavior is easier to detect even with a very simple form. Here we show the results in Table 7 by taking $p \leq 10$ and $d = 1$ in the model (4.2). The reason is from the previous subsection of specification. Our data is unable to

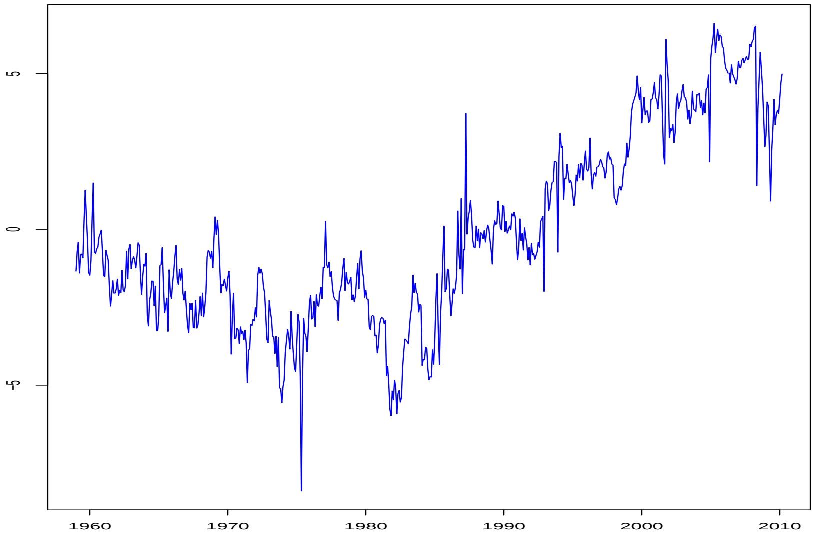

Figure 8: $\hat{u}_t = y_{1t} - 0.9877y_{2t}$.

reject the null of common nonlinear factors. But they reject the null of nonlinear cointegration. Here has a similar idea about the lag selection for the model. Obviously, lag 3,6,10 have negative contributions for common nonlinear factors. Thus, we consider $\hat{u}_{t-k}(k = 1)$ included in the model and $s_t = \hat{u}_{t-1}$ to forecast the residual $\hat{u}_t$.

Table 7: Order selection by the tests of common nonlinear factors, when $s_t = \hat{u}_{t-1}$ in (6.2).

lags	$W(H_{01})$	$W(H_{02})$
1	**0.09**	18.82
1,2	**0.13**	13.85
1,2,3	0.97	12.14
1,2,3,4	0.84	8.19
1,2,4	**0.58**	7.89
1,2,4,6	1.69	9.56
1,2,4,6,10	2.40	12.18

But before preceding that, we consider further another three types of the non-

Table 8: $G(\mathbf{s}_t;\gamma,\mathbf{c})$ form selection by the tests of common nonlinear factors.

$G(\mathbf{s}_t;\gamma,\mathbf{c})$	lags	$W(H_{01})$
$G_1(\mathbf{s}_t;\gamma,\mathbf{c})$	1	**0.09**
$G_2(\mathbf{s}_t;\gamma,\mathbf{c})$	1	5.91
$G_3(\mathbf{s}_t;\gamma,\mathbf{c})$	1	1.38
$G_4(\mathbf{s}_t;\gamma,\mathbf{c})$	1	0.14

linear component $G(\mathbf{s}_t;\gamma,\mathbf{c})$ that

$$
\begin{aligned}
G_2(\mathbf{s}_t;\gamma,\mathbf{c}) &= \left(1+exp\left\{-\gamma(\hat{u}_{t-1}-c)^2\right\}\right)^{-1} \\
G_3(\mathbf{s}_t;\gamma,\mathbf{c}) &= \left(1+exp\left\{-\gamma(\hat{u}_{t-1}-c_1)(\hat{u}_{t-2}-c_2)\right\}\right)^{-1} \\
G_4(\mathbf{s}_t;\gamma,\mathbf{c}) &= \left(1+exp\left\{-\gamma(\hat{u}_{t-1}-c_1)\right\}\right)^{-1}+\left(1+exp\left\{-\gamma(\hat{u}_{t-2}-c_2)\right\}\right)^{-1}.
\end{aligned} \tag{6.3}
$$

The reason why we include $\hat{u}_{t-2}$ as an option of the transition variable here is to see whether more complicated model can give significant positive contribution. Under those assumptions in (6.3), the results in Table 8 shows that we may consider $G_4(\mathbf{s}_t;\gamma,\mathbf{c})$ as another possible choice besides $G_1(\mathbf{s}_t;\gamma,\mathbf{c})$. But the simpler one $G_1(\mathbf{s}_t;\gamma,\mathbf{c})$ is slightly better based on the principle "simple is better". Besides the forecasting is far away from expectation by the simple one, we could consider to use $G_4(\mathbf{s}_t;\gamma,\mathbf{c})$.

6.4 Estimation and evaluation

We take a subset between January 1959 to December 2008 from the data to carry out out-of-sample predictions. The remaining subset (observations form 2009:1 to 2010:3) of each y_{1t} are used to examine the accuracy of forecasting. Of course, we do not wish to forecast income. Based on the suggestion in previous subsections, we firstly consider individual STAR model with trend in (2.3) for each $y_{it}, i=1,2$

by the following selected model

$$
\begin{aligned}
y_{1t} =& \mu_1 + \delta_1 t + \phi_{11} y_{1t-1} + \phi_{12} y_{1t-2} + \phi_{13} y_{1t-6} + \phi_{14} y_{1t-10} + \Gamma_1 y_{1t-1} G\left(s_t; \gamma, \mathbf{c}\right) + v_{1t} \\
y_{2t} =& \mu_1 + \delta_2 t + \phi_{21} y_{2t-1} + \phi_{22} y_{2t-4} + \Gamma_2 y_{2t-1} G\left(s_t; \gamma, \mathbf{c}\right) + v_{2t}
\end{aligned}
\tag{6.4}
$$

The reason why we need to do individual estimation of each process is that single model is easier to get accurate estimation and easier to learn some information of important initial values of parameters for nonlinear iterative algorithm. Furthermore, we can also obtain some information about the range of slope parameter γ and location parameter c. In the estimating procedure of individual model, we have found that it is a good way to get empirical positive definite Hessian matrix by rescaling the slope parameter γ by multiplying it by 10 and location parameter c by multiplying it by 100. The rescaling of c is straightforward to consider, since the range of our y_{it} is too large compared to the estimated parameters. Also Teräesvirta (1994) suggests standardizing the exponent by dividing it by the standard deviation of y_{it} before doing estimation. At last, we suggest the estimated individual models

$$
\begin{aligned}
\hat{y}_{1t} =& \underset{(10.70)}{24.69} + \underset{(.0039)}{0.0086}\, t + \underset{(.0553)}{0.7763}\, y_{1t-1} + \underset{(.0594)}{0.2053}\, y_{1t-2} + \underset{(.0450)}{0.0983}\, y_{1t-6} - \underset{(.0326)}{0.1153}\, y_{1t-10} \\
& + \underset{(.0150)}{0.0031}\, y_{1t-1} \left(1 + exp\left\{ - \underset{(0.6758)}{0.0712} (y_{1t-1} - \underset{(1.2513)}{726.3})\right\}\right)^{-1} \\
\hat{y}_{2t} =& \underset{(13.63)}{45.40} + \underset{(.0044)}{0.0145}\, t + \underset{(.0305)}{0.7938}\, y_{2t-1} + \underset{(.0289)}{0.1444}\, y_{2t-4} \\
& + \underset{(.0085)}{0.0038}\, y_{2t-1} \left(1 + exp\left\{ - \underset{(.2865)}{0.0742} (y_{2t-1} - \underset{(.5256)}{753.9})\right\}\right)^{-1}
\end{aligned}
\tag{6.5}
$$

where we can see that they have obvious signal of common nonlinear factors.

Then we turn to the vector case. We forecast the residual series by previous analysis and then forecast consumption by using the common factor relation between consumption and income. The advantage is that the residual is a linear process (trend stationary) based on our results from specification. The standard unit root test for $\hat{u}_t$ shows it is stationary and lag selection based on AIC supports only lag one included, which is not any with our specification for the model. Thus, we suggest

$$\hat{\hat{u}}_t = -0.3539 + 0.0012 \cdot t + 0.9125\hat{u}_{t-1} \tag{6.6}$$

for $\hat{u}_t$ in (6.6).

6.5 Forecasting

The forecast horizons are 1-month, 4month, 8month and 12-month step ahead. We don't respecify model since we forecast in one year. But parameters are re-estimated based on the bootstrap forecast for each forecast horizon in linear models. For nonlinear model, we do not re-estimate due to estimation problem. The last column in Table 9 is the results from a linear benchmark model. This

Table 9: Out-of-sample forecasts of Consumption.

Date	Horizon(Month)	Observe	Model (6.5)	Model (6.6)	AR model
2009-01	1	912.8106	912.5761	912.5906	912.7307
2009-04	4	912.4874	912.9902	913.1073	913.4590
2009-08	8	913.8242	913.9054	913.1389	914.4468
2009-12	12	913.9105	915.2871	913.8887	914.4264
		RMSE	2.079	0.870	0.951
		MAPE	0.874	0.067	0.091

Notes: RMSE-Root Mean Square Error; MAPE-Mean Absolute Percentage Error.

is the AR model with lag 1,6,7,8 for Δy_{1t}, since consumption series is a unit root under linear assumption. It is clearly shown from Table 9 that both Model (6.5) and Model (6.6) give acceptable forecasting results. But RMSE and MAPE suggest that Model (6.6) outperforms Model (6.5). There are two reasons to explain that. First, nonlinear estimation produces uncertainty for sure, and in our case, to estimate vector nonlinear model with common factor is really difficult to carry out, thus, univariate nonlinear model here increases the inaccurate. Second, model (6.6) is actually considered from vector nonlinear model with estimated certain nonlinear relation. It is easy to estimate , but it still works with common nonlinear factor.

7. Conclusions

In this paper we have proposed a definition of smooth-transition type nonlinear cointegration for a group of individually nonlinear global stationary time series. Simulation studies have been carried out to show the asymptotic characteristic of the tests. Finally, we apply our work to United States consumption and income data. We have compared the forecasting results with linear AR model and have shown that our nonlinear vector model without nonlinear estimation outperforms the linear model. But nonlinear estimation obviously make things bad. Thus, this kind of models like (6.5) using the information from our residual-based test for common nonlinear factor works very well. The method can successfully avoid nonlinear estimation but use the information from vector nonlinear model in the data.

References

[1] Dickey, D. A. and Fuller, W. A., 1979. Distributions of the Estimators For Autoregressive Time Series with a Unit Root. *Journal of the American Statistical Association*, 75, pp.427-431.

[2] Engle, R., Kozicki, S., 1993. Testing dor Common Features. *Journal of Business & Economic Statistics*, 11(4), pp.369-380.

[3] Granger & Teräsvirta, 1993. *Modelling Nonlinear Economic Relationships.* USA: Oxford University Press.

[4] He, C., Li, D., 2009. Testing for linear cointegration against smooth transition cointegration. Working Paper.

[5] Hinich, M.J., 1982. Testing for Gaussianity and linearity of a stationary time series. *Journal of Time Series Analysis*, 3, pp.169-176.

[6] Kwiatkowski, D., Phillips, P.C.B., Schmidt, P. and Shin, Y., 1992. Testing the Null Hypothesis of Stationarity Against the Alternative of a Unit Root: How Sure Are We That Economic Time Series Have a Unit Root? *Journal of Econometrics*, 54, pp.159-178.

[7] Lee, T. H., White, H., and Granger, C. W. J., 1993. Testing for neglected nonlinearity in time series models. *Journal of Econometrics*, 56, pp.269-290.

[8] McLeod, A.I., Li, W.K., 1983. Diagnostic checking ARMA time series models using squared residuals autocorrelations. *Journal of Time Series Analysis*, 4, pp.269-273.

[9] Teräsvirta, T., 1994. Specification, Estimation, and Evaluation of Smooth Transition Autoregressive Models. *Journal of the American Statistical Association*, 89(425), pp.208-218.

[10] Teräesvirta, T., Lin, C. F. and Granger C. W. J., 1993. Power of the Neural Network Linearity Test. *Journal of Time Series Analysis*, 14, pp.209-220.

[11] Tsay, R.S., 1986. Nonlinearity tests for time series. *Biometrica*, 73, pp.461-466.

[12] Zivot, E. and Andrews, Donald W.K., 1992. Further Evidence on the Great Crash, the Oil-Price Shock, and the Unit-Root Hypothesis. *Journal of Business & Economic Statistics*, 10(3), pp.251-270.

APPENDIX

A. Proofs of Theorems

Let ξ_{2t} is a time series process following the autoregressive model

$$\xi_{2t} = \phi_0 + \sum_{k=1}^{p} \phi_k \xi_{2t-k} + \varepsilon_t \tag{A.1}$$

with roots of $(1 - \phi_1 z - \phi_2 z^2 - ... - \phi_p z^p) = 0$ outside the unit circle and with $\{\varepsilon_t\}$ an *i.i.d.* sequence with mean zero, variance σ^2, and finite eighth moment μ_8.

We start from Theorem 2 since the assumption of $\hat{u}_t$ under the null H_{02} makes all things easier.

Proof of Theorem 2. In model (4.2), under the null hypothesis H_{02}, $\hat{u}_t$ is a covariance-statioanry process as ξ_{2t} defined in (A.1). Let $\mathbf{x}_t = (1, \mathbf{x}_{1t}, \mathbf{x}_{2t}, t)'$ where $\mathbf{x}_{1t} = (\hat{u}_{t-1}, \hat{u}_{t-2}, ..., \hat{u}_{t-p})'$ and $\mathbf{x}_{2t} = (\hat{u}_{t-1}\hat{u}_{t-q}, \hat{u}_{t-2}\hat{u}_{t-q}, ..., \hat{u}_{t-p}\hat{u}_{t-q})'$. Define

$$\hat{\boldsymbol{\beta}}^*_T - \boldsymbol{\beta}^* \; = \mathbf{A}^{-1}\mathbf{B}$$

where

$$\mathbf{A} = \sum \mathbf{x}_t \mathbf{x}_t' = \begin{pmatrix} \sum 1 & \sum \mathbf{x}_{1t}' & \sum \mathbf{x}_{2t}' & \sum t \\ \sum \mathbf{x}_{1t} & \sum \mathbf{x}_{1t}\mathbf{x}_{1t}' & \sum \mathbf{x}_{1t}\mathbf{x}_{2t}' & \sum t\mathbf{x}_{1t} \\ \sum \mathbf{x}_{2t} & \sum \mathbf{x}_{2t}\mathbf{x}_{1t}' & \sum \mathbf{x}_{2t}\mathbf{x}_{2t}' & \sum t\mathbf{x}_{2t} \\ \sum t & \sum t\mathbf{x}_{1t} & \sum t\mathbf{x}_{2t} & \sum t^2 \end{pmatrix}$$

$$= \begin{pmatrix} O_p(T) & O_p(T) & O_p(T) & O_p(T^2) \\ O_p(T) & O_p(T) & O_p(T) & O_p(T^{3/2}) \\ O_p(T) & O_p(T) & O_p(T) & O_p(T^{3/2}) \\ O_p(T^2) & O_p(T^{3/2}) & O_p(T^{3/2}) & O_p(T^3) \end{pmatrix}$$

$$\mathbf{B} = \sum \mathbf{x}_t e_t = \begin{pmatrix} \sum e_t & \sum \mathbf{x}_{1t}' e_t & \sum \mathbf{x}_{2t}' e_t & \sum t e_t \end{pmatrix}',$$

and

$$\boldsymbol{\Upsilon}_T = \begin{pmatrix} T^{1/2} & 0 & 0 & 0 \\ 0 & T^{1/2}\mathbf{I}_{p\times p} & 0 & 0 \\ 0 & 0 & T^{1/2}\mathbf{I}_{p\times p} & 0 \\ 0 & 0 & 0 & T^{3/2} \end{pmatrix},$$

then we rewrite

$$\boldsymbol{\Upsilon}_T \left(\hat{\boldsymbol{\beta}}_T^* - \boldsymbol{\beta}^* \right) = \left(\boldsymbol{\Upsilon}_T^{-1}\mathbf{A}\boldsymbol{\Upsilon}_T^{-1}\right)^{-1}\left(\boldsymbol{\Upsilon}_T^{-1}\mathbf{B}\right)$$

where

$$\mathbf{Q}_{2T} = \boldsymbol{\Upsilon}_T^{-1}\mathbf{A}\tilde{\boldsymbol{\Upsilon}}_T^{-1} \xrightarrow{L} \begin{pmatrix} 1 & E(\mathbf{x}_{1t}') & E(\mathbf{x}_{2t}') & \frac{1}{2} \\ E(\mathbf{x}_{1t}) & E(\mathbf{x}_{1t}\mathbf{x}_{1t}') & E(\mathbf{x}_{1t}\mathbf{x}_{2t}') & \mathbf{0} \\ E(\mathbf{x}_{2t}) & E(\mathbf{x}_{2t}\mathbf{x}_{1t}') & E(\mathbf{x}_{2t}\mathbf{x}_{2t}') & \mathbf{0} \\ \frac{1}{2} & \mathbf{0} & \mathbf{0} & \frac{1}{3} \end{pmatrix} = \mathbf{Q}_2$$

$$\mathbf{h}_{2T} = \boldsymbol{\Upsilon}_T^{-1}\mathbf{B} = \begin{pmatrix} T^{-1/2}\sum e_t \\ T^{-1/2}\sum \mathbf{x}_{1t}e_t \\ T^{-1/2}\sum \mathbf{x}_{2t}e_t \\ T^{-3/2}te_t \end{pmatrix} \xrightarrow{L} \begin{pmatrix} N(0,\sigma^2) \\ N(0,\sigma^2 E(\mathbf{x}_{1t}\mathbf{x}_{1t}')) \\ N(0,\sigma^2 E(\mathbf{x}_{2t}\mathbf{x}_{2t}')) \\ N(0,\sigma^2/3) \end{pmatrix} = \mathbf{h}_2$$

who are not identical depending on the different options of transition variable $\hat{u}_{t-q}$. □

Let ξ_{1t} is a time series process following the autoregressive model

$$\xi_{1t} = \phi_0 + \sum_{k=1}^{p} \phi_k \xi_{1t-k} + \theta \cdot t + \varepsilon_t \tag{A.2}$$

with roots of $(1 - \phi_1 z - \phi_2 z^2 - ... - \phi_p z^p) = 0$ outside the unit circle and with $\{\varepsilon_t\}$ an *i.i.d.* sequence with mean zero, variance σ^2, and finite eighth moment μ_8.

Proof of Theorem 1. In model (4.2), under the null hypothesis H_{01}, $\hat{u}_t$ is a trend-statioanry process as ξ_{1t} defined in (A.2). We also define $\mathbf{x}_t = (1, \mathbf{x}_{1t}, \mathbf{x}_{2t}, t)'$ as in Proof of Theorem 2 above.

We consider a useful transformation of the regressors in model (A.2) as Hamilton (1994, p464) described. Then under the null of H_{01}, the converge rate $\boldsymbol{\Upsilon}_T$ is the same if we replace $\hat{u}_t$ as transformed $\hat{u}_{t-i}^*(i = 1, 2, ..., p)$ by the silimar way, and the distribution is still the normal distribution. But $\mathbf{Q}_1$ and $\mathbf{h}_1$ are not identical with $\mathbf{Q}_2$ and $\mathbf{h}_2$. □

B. Critical Values

Table 10: Critical values for non-standard unit root tests.

		Probability that F is greater than entry							
s_t	T	99%	97.5%	95%	90%	10%	5%	2.5%	1%
Δy_{t-1}	50	0.65	0.79	0.93	1.14	4.29	5.09	5.88	6.91
	100	0.64	0.78	0.93	1.14	4.15	4.86	5.54	6.45
	200	0.64	0.78	0.93	1.13	4.05	4.74	5.41	6.22
	500	0.63	0.78	0.92	1.13	4.05	**4.73**	5.36	6.12
	1000	0.63	0.78	0.92	1.13	4.04	4.68	5.33	6.17
y_{t-1}	50	0.86	1.04	1.21	1.46	4.94	5.78	6.63	7.74
	100	0.92	1.10	1.29	1.54	4.90	5.68	6.41	7.34
	200	0.96	1.14	1.33	1.59	4.92	5.65	6.36	7.28
	500	0.97	1.15	1.35	1.61	4.93	**5.67**	6.37	7.23
	1000	0.98	1.16	1.36	1.63	4.94	5.67	6.36	7.24

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

Table 11: Critical values for common factor tests under H_{01}.

		Probability that F is greater than entry							
s_t	T	99%	97.5%	95%	90%	10%	5%	2.5%	1%
u_{t-2}	100	0.00	0.00	0.00	0.02	3.35	4.78	6.27	8.31
	50	0.00	0.00	0.00	0.02	3.28	4.65	6.07	8.08
	500	0.00	0.00	0.00	0.02	3.07	4.38	5.70	7.41
	200	0.00	0.00	0.00	0.02	2.88	4.08	5.31	6.99
	1000	0.00	0.00	0.00	0.02	2.79	3.95	5.13	6.77

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

Table 12: Critical values for common factor tests under H_{02}.

		Probability that F is greater than entry							
s_t	T	99%	97.5%	95%	90%	10%	5%	2.5%	1%
u_{t-2}	50	0.00	0.00	0.00	0.01	2.47	3.54	4.68	6.28
	100	0.00	0.00	0.00	0.02	2.56	3.65	4.77	6.39
	200	0.00	0.00	0.00	0.02	2.64	3.73	4.93	6.47
	500	0.00	0.00	0.00	0.02	2.68	3.79	4.93	6.52
	1000	0.00	0.00	0.00	0.02	2.68	3.78	4.93	6.53

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