Testing Linearity against Nonlinearity and Detecting Common Nonlinear Components for Industrial Production of Sweden and Finland

Feng Li

Supervisor: Changli He

HÖGSKOLAN DALARNA

Master thesis in statistics
School of Economics and Social Sciences
Dalarna University, Sweden
June, 2008

Contents

Abstract

The paper discusses a nonlinear parametric time series model for industrial production. The model is based on a standard logistic smooth transition autoregressive (LSTAR) model for the first difference of industrial production. We introduce two test statistics testing the null hypothesis of a random walk (with and without drift) against LSTAR models that accommodate a smooth nonlinear shift in the level. We derive theoretical limiting distributions and finite sample results for all the tests. Linearity tests are performed for a number of quarterly, unadjusted, industrial production series of Sweden and Finland, and linearity are rejected for them. Nonlinearity found by testing can be modeled satisfactorily by use of our LSTAR model. Furthermore, we also introduce the concept of "common nonlinear components" into the LSTAR model and provide a method for detecting common nonlinear components in multiple LSTAR time series. In the industrial production of Sweden and Finland a common nonlinear component is detected that finds linear combination of the two LSTAR models with nonlinearity which implies potential for forecasting.

Key words: Smooth Transition Autoregressive Model; Nonlinear Dickey-Fuller Type Test; Linearity Test; Common Nonlinear Components.

1. Introduction

In most OECD countries, the industry production index have increased markedly since the 1950s. The tendency of the industry production to remain on a level it has reached is often called hysteresis[1]. Furthermore a lot of economic systems show signs of hysteresis and it is the key issue for economic prediction. To illustrate this, consider a deterministic system with no hysteresis and no dynamics. In that case, we can predict the output of the system at some instant in time, given only the input to the system at that instant. If the system has hysteresis, then this is not the case; we can't predict the output without looking at the history of the input. In order to predict the output, we must look at the path that the input followed before it reached its current value.

Consequently, a popular way of investigating hysteresis or full persistence in industry production has been to test the null hypothesis that the industrial production has a unit root. It is well known that classical unit-root tests based on linear models, such as those by Dickey and Fuller (1979), Phillips (1987) and Phillips and Perron (1988) among others, lack power when the model specification under the alternative hypothesis is nonlinear. Smooth Transition Autoregressive (STAR) models provide useful starting points for analyzing nonlinear phenomena in macroeconomics. The LSTAR models allowing nonlinear dynamic structures with smooth "regime–transition" have been wildly applied on economic data. Thus, testing linearity against non-linearity is an essential strategy in the context of nonlinear modeling.

Furthermore, many theoretical models of the economy imply common factors in systems of interrelated variables, e.g. Cox *et al* (1985) have showed the instantaneous interest rate is common to yields on bills of all maturities. Anderson and Vahid (1998) have pointed out the importance of common components saying the existence of common components enables parsimony, which is especially important when estimaling nonlinear models, knowledge about these components can also help us understand

[1]A system with hysteresis can be summarized as a system that may be in any number of states, independent of the inputs to the system. To be exact, a system with hysteresis exhibits path-dependence, or "rate-independent memory" .

economic linkages between variables. In such cases it will be interesting to study the common nonlinear components in our LSTAR models.

In this article, we propose to investigate such characteristics in multivariate data sets. The interesting questing is whether the STAR nonlinearity features that are detected in single data series are actually shared in common. Engle and Kozicki (1993) gave a more special case saying a feature will be said to be common if a linear combination of the series fails to have the feature even though each of the series individually has the feature.

The plan of the paper is as follows. We introduce the LSTAR models in Section 2 and in Section 3 we introduce the procedure for testing unit root against the stable LSTAR model and test for common nonlinear components. In Section 4 we present the asymptotic properties of the Dickey-Fuller type F test statistic, and we also investigate finite sample properties of the F test by Monte Carlo experiment. In Section 6 an empirical example about the OECD country industrial production of Sweden and Finland from 1958(i) to 2007(iv) is illustrated by applying the proposed testing procedure of the Nonlinear Dickey-Fuller type F test and a simulation based study of properties of nonlinear common components . Concluding remarks can be found in the final section. All proofs of theorems in this paper are given in the Appendix .

2. The Basic Model

Consider the following LSTAR model of order one introduced by Teräsvirta (1994)

$$y_t = \pi_{10} + \pi_{11} y_{t-1} + (\pi_{20} + \pi_{21} y_{t-1}) \boldsymbol{G}(z_t; \gamma, c) + \mu_t \tag{2.1}$$

where $\mu_t \sim nid(0, \sigma^2)$ and $\boldsymbol{G}$ is a transition function[2] which by convention is bounded by -1/2 and 1/2 and $z_t = \Delta y_{t-1}$.

$$\boldsymbol{G}(z_t; \gamma, c) = \frac{1}{1 + \exp(-\gamma(\Delta y_{t-1} - c))} - \frac{1}{2}, \quad \gamma > 0 \tag{2.2}$$

[2]In this section, we focus on the first order delay in the transition function, but from our later discussions and theoretical work in Appendix, we can expand this case to a more general situation where the transition variable can be n-th order delay ($n < T$), and the inference of the asymptotic distribution can be easily obtained.

It is clear that (2.1) is an autoregressive model with order one if $\gamma = 0$. Notice that when $\gamma \to \infty$ in (2.2), $\boldsymbol{G}(z_t; \gamma, c) = 1/2$, if $\Delta y_{t-1} \leqslant c$, $\boldsymbol{G}(z_t; \gamma, c) = -1/2$, if $\Delta y_{t-1} > c$. The transition function (2.2) is a monotonically increasing function of z_t. The slope parameter γ indicates how rapid the transition from -1/2 to 1/2 is as a function of z_t and the location parameter c determines where the transition occurs.

Applied to the modeling of business cycles indicators the model describes a situation where the contraction and expansion phases of an economy may have rather different dynamics, and a transition (change in dynamics) from one to the other may be smooth(For more discussions about LSTAR model see e.g. Teräsvirta and Anderson (1992)). The specification in (2.2) allows for one transition[3] for each parameter, where the slope parameter γ determines the speed of transition from one regime to another, and the location parameter c determines where the transition occurs.

3. Testing Methodology

In this section, we will show our testing procedure in the unit-root test. Later we want introduce a method to detect non linear common factors in multivariate time series if all the single time series was reject of the null of unit-root.

3.1 The Unit-root Tests

Assuming that we want to test the null hypothesis of a random walk without drift against the LSTAR(1) model and we also want to test the null hypothesis of a random walk with drift against the LSTAR(1). If the null of linearity is accepted, the conclusion is that the business cycle indicator can be adequately described by a linear auto-regression model. If it is rejected, the specification of the nonlinear model becomes a matter of concern.

[3]In this paper we focus on a smooth transition function with a single transition between regimes. However, He and Sandberg (2006) discussed a universal case where the transition functions with an "arbitrary" number of transitions with the transition variable $z_t = t$. Similarly, we could generalize the discussion with transition variable $z_t = \Delta y_{t-d}$.

Testing the unit-root hypothesis in the LSTAR(1) models is formalized as follows:

$$\begin{aligned} &H_{01}: \; y_t = y_{t-1} + \mu_t, \\ &H_{02}: \; y_t = \pi_{10} + y_{t-1} + \mu_t, \; \pi_{10} \neq 0, \\ &H_{am}: \; y_t = \pi_{10} + \pi_{11} y_{t-1} + (\pi_{20} + \pi_{21} y_{t-1}) \boldsymbol{G}(z_t; \gamma, c) + \mu_t, \; m = 1, 2. \end{aligned} \tag{3.1}$$

The implementation of these three tests is straightforward in the sense that for $m = 1,\ 2$, the models under the alternative hypothesis, H_{am} nest the models under the null hypothesis H_{0m}.

Following Luukkonen, Saikkonen and Teräsvirta (1988), we remedy this problem by a first-order Taylor expansion of γ around 0 in $\boldsymbol{G}$. A first-order Taylor approximation of the transition function is given by

$$T(\Delta y_{t-1}; \gamma, c) = \frac{\gamma(\Delta y_{t-1} - c)}{4} + r(\gamma) \tag{3.2}$$

where $r(\gamma)$ is a remainder such that $r(0) = 0$.

Substituting (3.2) into the models in H_{a1}, H_{a2} and merging terms yields the auxiliary regression equations

$$H_{am}: \; y_t^{aux} = \pi_{10}^* + \pi_{11}^* y_{t-1} + \pi_{20}^* y_{t-2} + \pi_{21}^* y_{t-1} \Delta y_{t-1} + \mu_t^*, \quad m = 1,\ 2 \tag{3.3}$$

where μ_t^* is an error term adjusted with respect to the Taylor expansions such that $\mu_t^* = \mu_t$ whenever $\gamma = 0$. The corresponding auxiliary null hypotheses are given by

$$H_{01}: \; \pi_{10}^* = 0, \; \pi_{11}^* = 1, \; \pi_{20}^* = 0, \; \pi_{21}^* = 0, \tag{3.4}$$

$$H_{02}: \; \pi_{10}^* \neq 0, \; \pi_{11}^* = 1, \; \pi_{20}^* = 0, \; \pi_{21}^* = 0, \tag{3.5}$$

3.2 Testing for Common Nonlinear Components

In the single LSTAR tests, if the tests reject the null hypothesis in each case. The evidence then suggests that the variables y_{1t} and y_{2t} have LSTAR nonlinearities with the same transition variable Δy_{t-1}. If the LSTAR nonlinearity is common, then there will be at least one $\boldsymbol{\alpha}$ making $\boldsymbol{\alpha' y_t}$ linear in mean (Anderson & Vahid (1998)).

Consider a situation in which the unit-root test we derived in Section 3.1 rejects the null hypothesis of unit-root for each of the elements of the two vectors $(y_{1t}, y_{2t})'$. That is

$$\begin{pmatrix} y_{1t} \\ y_{2t} \end{pmatrix} = \begin{pmatrix} \pi_1 \\ \pi_2 \end{pmatrix} + \begin{pmatrix} \alpha_{11} & \alpha_{12} \\ \alpha_{21} & \alpha_{22} \end{pmatrix} \begin{pmatrix} y_{1,t-1} \\ y_{2,t-1} \end{pmatrix} + \begin{pmatrix} \beta_{11} & \beta_{12} \\ 0 & 0 \end{pmatrix} \begin{pmatrix} y_{1,t-1} \\ y_{2,t-1} \end{pmatrix} G\left(z_t; \gamma, c\right) + \begin{pmatrix} \mu_{1t} \\ \mu_{2t} \end{pmatrix} \tag{3.6}$$

where the transition functions $G(z_t; \gamma, c)$ is same as we mentioned in (2.2) allows the parameters of the model to change smoothly from one extreme regime to the other as a function of the transition variable z_t. Moreover,

$$\begin{pmatrix} \mu_{1t} \\ \mu_{2t} \end{pmatrix} \sim NID\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} \sigma_{11} & \sigma_{12} \\ \sigma_{21} & \sigma_{22} \end{bmatrix} \right) \tag{3.7}$$

where σ_{11} and σ_{22} are the variance of μ_{1t} and μ_{2t} respectively and σ_{12} (is equal to σ_{21}) is the correlation between μ_{1t} and μ_{2t} .

Hence a vector smooth transition autoregressive (VSTAR) model in (3.6) is introduced in the analysis of "common nonlinear components". Noticing that the matrix in (3.6) with the elements of β_{11} and β_{12} is not full ranked which we call the model is a reduced-rank autoregressive model (Ahn, & Reinsel (1988)).

In (3.6), if $G(z_t; \gamma, c)$ is equal to zero, i.e. γ is equal to zero, there is a linear combination exit for y_{1t} and y_{2t}. In this sense, we can generalize the linear combination of y_{1t} and y_{2t} as the following way

$$y_{2t} = \alpha_0 + \sum_{i=1}^{p} \alpha_i L^i y_{1t} + \sum_{j=0}^{q} \beta_j L^j y_{2t} + \mu_t \tag{3.8}$$

where L^i and L^j are denoted as the lag operator of y_{1t} and y_{2t} and α_i $(i = 0, 1, ..., p)$ and β_j $(j = 0, 1, ..., q)$ are parameters such that the residuals series μ_t does not contain nonlinear components.

Commonly used method to test for such "common nonlinearities" will seek to determine if such combinations exist. Varies of methods can be used in this test[4]. We will consider it in another practical way, that whether we can find a p and q such that μ_t fulfills the Gauss-Markov assumptions (i) $E(\mu_i) = 0$; (ii) $Var(\mu_i) = \sigma^2 < \infty$, and (iii) $Cov(\mu_i, \mu_j) = 0$ for $i \neq j$. If this is successfully done, it means there exits a combination of y_{1t} and y_{2t} linear. That is to say nonlinear component was detected.

4. The Asymptotic and Finite-sample Results

In this section, we will give the theoretical result of the asymptotic distribution of the Nonlinear Dicky-Fuller type F statistic. In order to validate the theoretical work, we make computer based experiments give the finite-sample size results.

4.1 The Nonlinear Dickey-Fuller Type *F* Tests

We will derive two unit-root tests based on the assumption in Section 2. As a remark on notation in the following theorems and corollaries, $\xrightarrow{d}$ and $\xrightarrow{p}$ denote convergence in distribution and probability, respectively, and $W(r)$ denotes a standard Brownian motion defined on [0, 1].

Assumption 1. *Let* $(\mu)_{t=1}^{\infty}$ *be an i.i.d sequence of random variables defined on a probability space* $(\Omega, F, \mathbb{P})$*, such that* $E(\mu_t) = 0,\ E(\mu_t^2) = \sigma_\mu^2$ *and* $E(\mu_t^4) < \infty$ *hold for all t.*

Theorem 1. *Consider model (3.3) when (3.4) holds. furthermore, assume that Assumption 1 is fulfilled, then,*

$$\boldsymbol{\gamma_{T1}}\left(\boldsymbol{b_{T1}} - \boldsymbol{\beta_1}\right) \xrightarrow{d} \boldsymbol{Q^{-1}h_1},\ \boldsymbol{b_{T1}} - \boldsymbol{\beta_1} \xrightarrow{p} \boldsymbol{0} \tag{4.1}$$

where $\boldsymbol{b_{T1}} = (\hat{\pi}_{10}^*, \hat{\pi}_{11}^*, \hat{\pi}_{20}^*, \hat{\pi}_{21}^*)'$, $\boldsymbol{\beta_1} = (\pi_{10}^*, \pi_{11}^*, \pi_{20}^*, \pi_{21}^*)' = (0, 1, 0, 0)'$ *and* $\boldsymbol{\gamma_{T1}} = diag\left(T^{1/2}, T, T, T^{1/2}\right)$*. Moreover, in (4.1),*

$$\boldsymbol{Q} = (q_{ij})_{4\times 4}\,, \boldsymbol{h_1} = (h_{1i})_{4\times 1} \tag{4.2}$$

[4]See e.g. Anderson & Vahid (1998) for a discussion

where

$$q_{ij} = q_{ji}, q_{11} = 1, q_{21} = q_{31} = \sigma \int_0^1 W(r)\, dr,$$
$$q_{22} = q_{33} = q_{32} = \sigma^2 \int_0^1 (W(r))^2\, dr,$$
$$q_{41} = \frac{1}{2}\sigma^2 \left((W(r))^2 - 1\right),$$
$$q_{42} = q_{43} = c_s\sigma^2, q_{44} = \gamma_0\sigma^2,$$
$$h_{11} = \sigma W(1), h_{12} = h_{13} = \frac{1}{2}\sigma^2 \left((W(r))^2 - 1\right)$$
$$h_{14} = \sigma^3 W(1)$$

Proof of Theorem 1. See Appendix A □

Theorem 1 applies to the nonlinear Dickey-Fuller F test statistic of the joint hypothesis $\boldsymbol{R\beta} = \boldsymbol{r}$ given by (3.4) for $\boldsymbol{R}$ a known $[m \times (p+1)]$ matrix where m is the number of restrictions. The F test is then

$$F_{T1} = \frac{1}{4}(\boldsymbol{b}_{T1} - \boldsymbol{\beta}_1)' \boldsymbol{R}' \left(S_T^2 \boldsymbol{R} \left(\sum_{t=1}^{T} \boldsymbol{xx}' \right)^{-1} \boldsymbol{R}' \right)^{-1} \boldsymbol{R}(\boldsymbol{b}_{T1} - \boldsymbol{\beta}) \tag{4.3}$$

where

$$S_T^2 = \frac{1}{T-4} \sum_{t=1}^{T} \hat{\mu}_t^2 \tag{4.4}$$

Noticing that (4.3) can be written as

$$\begin{aligned} F_{T1} &= \frac{1}{4}(\boldsymbol{b}_{T1} - \boldsymbol{\beta}_1)' \boldsymbol{R}'\boldsymbol{\gamma}_{T1}' \left(S_T^2 \boldsymbol{\gamma}_{T1} \boldsymbol{R} \left(\sum_{t=1}^{T} xx' \right)^{-1} \boldsymbol{R}'\boldsymbol{\gamma}_{T1}' \right)^{-1} \boldsymbol{\gamma}_{T1}\boldsymbol{R}(\boldsymbol{b}_{T1} - \boldsymbol{\beta}_1) \\ &\xrightarrow{L} \frac{1}{4}\left(\boldsymbol{Q}^{-1}\boldsymbol{h}_1\right)' \left(\sigma^2\boldsymbol{Q}^{-1}\right)^{-1} \left(\boldsymbol{Q}^{-1}\boldsymbol{h}_1\right) \\ &= \frac{\boldsymbol{h}_1'\boldsymbol{Q}^{-1}\boldsymbol{h}_1}{4\sigma^2} \end{aligned} \tag{4.5}$$

Proof of (4.5). See Appendix A. □

Theorem 2. *Consider model (3.3) when (3.5) holds. furthermore, assume that Assumption 1 is fulfilled, then,*

$$\gamma_{T2}\left(\boldsymbol{b_{T2}}-\boldsymbol{\beta_2}\right)\xrightarrow{L}\boldsymbol{N}\left(\mathbf{0},\sigma^2\boldsymbol{\Omega}^{-1}\right) \tag{4.6}$$

where $\boldsymbol{b_{T2}}=(\hat{\pi}^*_{10},\hat{\pi}^*_{11},\hat{\pi}^*_{20},\hat{\pi}^*_{21})'$, $\boldsymbol{\beta_2}=(\pi^*_{10},\pi^*_{11},\pi^*_{20},\pi^*_{21})'=(\alpha,1,0,0)'$ *and* $\boldsymbol{\gamma_{T1}}=diag\left(T^{1/2},T^{3/2},T^{3/2},T^{3/2}\right)$*. Moreover, in (4.6),*

$$\boldsymbol{\Omega}=(\omega_{ij})_{4\times 4} \tag{4.7}$$

Proof of Theorem 2. See Appendix A. □

Thus, for the unit-root with drift case, both estimated coefficients are asymptotically Gaussian.[5] Similarly, Theorem 2 applies to the nonlinear Dickey-Fuller F test statistic of the joint hypothesis $\boldsymbol{R\beta}=\boldsymbol{r}$ given by (3.5) for $\boldsymbol{R}$ a known $[m\times(p+1)]$ matrix where m is the number of restrictions. The F test is then

$$\begin{aligned}F_{T2}&=\frac{1}{4}\left(\boldsymbol{b}_{T2}-\boldsymbol{\beta}_2\right)'\boldsymbol{R}\left(S_T^2\boldsymbol{R}\left(\sum_{t=1}^{T}\boldsymbol{x}\boldsymbol{x}'\right)^{-1}\boldsymbol{R}'\right)^{-1}\boldsymbol{R}\left(\boldsymbol{b}_{T2}-\boldsymbol{\beta}_2\right)\\&\xrightarrow{p}\frac{1}{4}\left(\boldsymbol{\gamma}_{T2}\left(\boldsymbol{b}_{T2}-\boldsymbol{\beta}_2\right)\right)'\left(\sigma^2\boldsymbol{\Omega}^{-1}\right)^{-1}\boldsymbol{\gamma}_{T2}\left(\boldsymbol{b}_{T2}-\boldsymbol{\beta}_2\right)\end{aligned} \tag{4.8}$$

Furthermore

$$4F_{T2}\xrightarrow{L}\chi^2\left(4\right) \tag{4.9}$$

Proof of (4.8). It is similar to the Proof of (4.5). See Appendix A for more details. □

4.2 Monte Carlo Experiments

To find the asymptotic and finite-sample critical values, we let $T=1,000,000$, to simulate a Brownian Motion $W(r)$ on [0, 1], and the number replications are set to 1,000,000. The finite-sample critical values for the same tests are obtained by simulating data from the model $y_t=y_{t-1}+\mu_t$ where $\mu_t\sim$ n.i.d. (0, 1) with desired sample sizes. Finite sample critical values are reported in Table 1. Use the same method, we can generate in the other

[5]This is easy to explain because y_{t-1} is asymptotically dominated by the time trend $\alpha(t-1)$. In large samples, it is as if the explanatory variable y_{t-1} were replaced by the time trend $\alpha(t-1)$. For more details please check the proof in the Appendix.

Table 1: Asymptotic and finite-sample critical values (without drift)

Sample	$P(X > x) = 1 - \alpha$							
Size (T)	.99	.975	.95	.90	.10	.05	.025	.01
25	.29	.37	.47	.61	3.52	4.58	5.77	7.40
50	.28	.38	.48	.61	3.06	3.80	4.56	5.71
100	.31	.39	.48	.61	2.81	3.38	3.91	4.68
200	.28	.38	.47	.61	2.69	3.24	3.73	4.39
500	.30	.39	.48	.62	2.66	3.11	3.61	4.15
∞	.29	.38	.49	.62	2.63	3.08	3.55	4.13

case[6] . The finite-sample critical values for the same tests are obtained by simulating data from the model $y_t = 1 + y_{t-1} + \mu_t$ where $\mu_t \sim$ n.i.d. (0, 1) with desired sample sizes. Finite sample critical values are reported in Table 2.

Table 2: Asymptotic and finite-sample critical values (with drift = 1)

Sample	$P(X > x) = 1 - \alpha$							
Size (T)	.99	.975	.95	.90	.10	.05	.025	.01
25	.08	.12	.19	.29	2.63	3.41	4.21	5.76
50	.07	.12	.18	.27	2.21	2.75	3.28	3.92
100	.08	.13	.19	.28	2.09	2.59	3.03	3.59
200	.08	.12	.18	.27	2.03	2.45	2.89	3.45
500	.08	.13	.18	.27	1.94	2.36	2.79	3.34
∞	.08	.13	.19	.27	1.94	2.33	2.78	3.36

Figure 1 shows the density of asymptotic distribution plot of the Dickey-Fuller type F test statistic. We can compare them with $\chi^2(4)$ distribution. It is clear that both of the distributions with and without drift cases have a similar distribution compared with chi-square distribution with degree

[6]For more universal case with different sample size and drift, a computer program written with R is available. See Appendix B for more information.

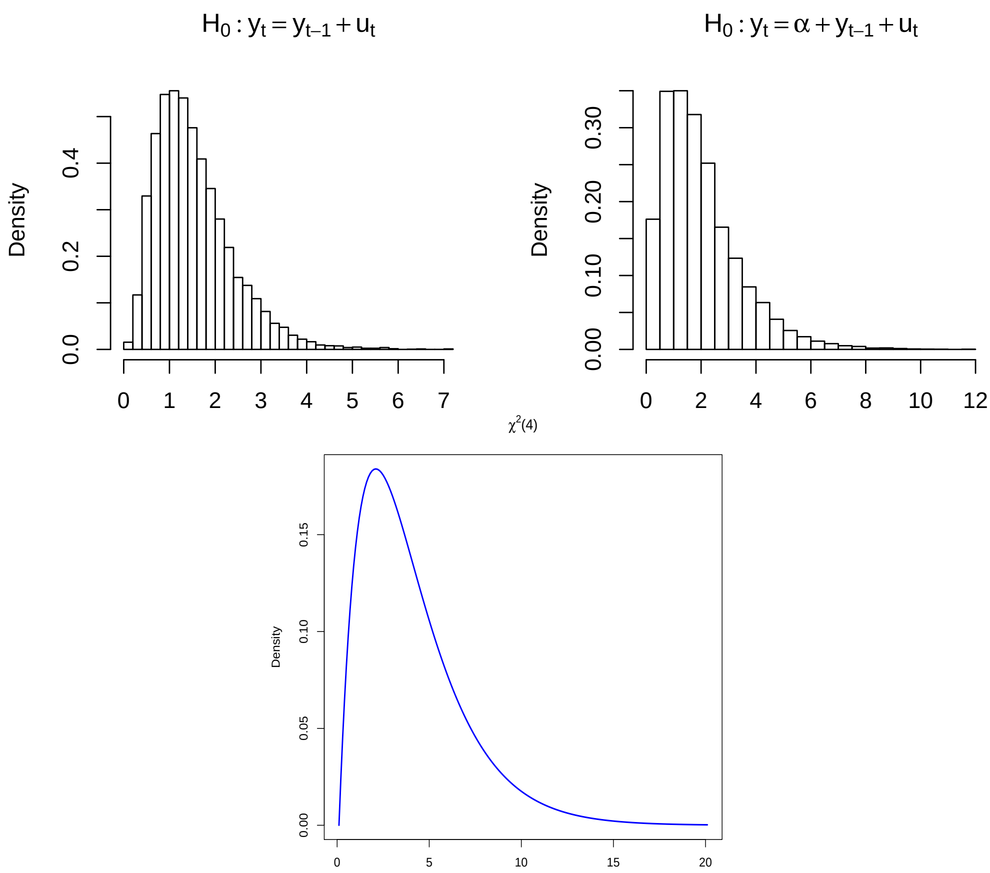

Figure 1: Density of asymptotic distribution of Dickey-Fuller type F statistic compared with $\chi^2(4)$ distribution

of freedom 4 which can support our theoretical work in Section 4.1

To study of the properties of the nonlinear common factors of two LSTAR models, following the experimental DGPs provided in Appendix B, we can learn how this procedure would work in a situation mentioned above from (3.6) to (3.8). Table 3 shows that any nonlinearities which might be present in the system are LSTAR nonlinearities with the same known transition variable.

The test are able to distinguish whether or not a nonlinear factor is

Table 3: Performance of the common nonlinearities test

Sample Size	LSTAR nonlinearities found(‰)
100	71
200	45
500	16

Notice: relative frequencies in 1000 replication.

present. The ability to discriminate improves quite sharply with sample size.

5. Empirical Applications

The index of Industrial Production is a fixed-weight measure of the physical output of the nation's factories, mines, and utilities. Although these sectors contribute only a small portion of GDP (Gross Domestic Product), they are highly sensitive to interest rates and consumer demand. This makes Industrial Production an important tool for forecasting future GDP and economic performance. Industrial Production figures are also used by central banks to measure inflation, as high levels of industrial production can lead to uncontrolled levels of consumption and rapid inflation.

Furthermore, when studying the possible nonlinearity of business cycles it is useful to choose a business cycle indicator that shows as much cyclical variation as possible(Teräsvirta & Anderson (1992)). Based on the above reasons, we use industrial production data for our analysis because this series shows more cyclical variables then GDP. The source is the *OECD Main Economic Indicators*[7].

5.1 The Swedish and Finnish Industrial Production

We choose the industrial production data in Sweden and Finland from 1958(i) to 2007(iv) for forecasting. From Figure 2, it is apparent that all

[7]See Appendix C for more data information.

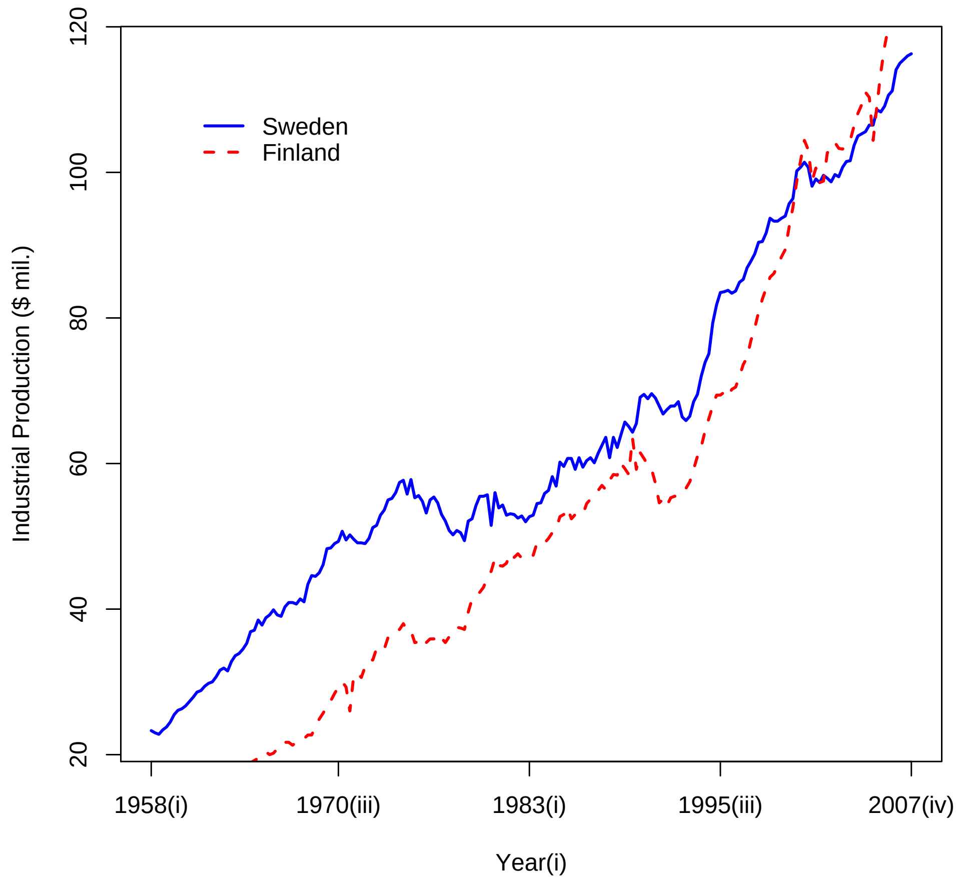

Figure 2: Industrial Production in Sweden and Finland

two countries listed have a total increase trend in the industrial production from the year 1958 to 2007. But when we take a close look at the trends we may discern that both of Sweden and Finland have at least one jump point around the years 1975–1980 and 1990–1995 respectively. If we use a linear method to make a forecasting, it may lead us astray. It is necessary to take a linearity test.

5.2 Linearity Tests

To test the linearity, we calculate the Nonlinear Dickey-Fuller type F statistics we have derived in (4.3) using the industrial production data of Sweden and Finland. The Nonlinear Dickey-fuller type F statistic is 9.98 in Sweden and 13.17 in Finland. The computer program give the critical value $P(X > 3.24) = 1 - \alpha$, where $\alpha = .05$ (See Table 1). The result implies that in both cases we reject the null hypothesis of unit-root, which means the industrial production data in Sweden and Finland can be specified as LSTAR models.

5.3 Testing for Common Nonlinearity

Following the method we proposed in Section 3.2, we make a linear regression with the data of Sweden and Finland in (5.1) for all $p, q < T$. Our purpose bases a criterion that it is the aim to choose a model with give restriction residuals and with as few parameters as possible. We use stepwise regression method to identify a good subset model, with considerably less computing than that is required for all possible regressions.

$$y_t^{sv} = \underset{(2.35)}{1.07} + \underset{(48.46)}{.96}\, y_{t-1}^{sv} + \underset{(4.59)}{.28}\, y_t^{fi} - \underset{(-4.01)}{.25}\, y_{t-1}^{fi} + \hat{\mu}_t \tag{5.1}$$

where y_t^{sv} denotes the industrial production of Sweden and y_t^{fi} is for Finnish industrial production.

The t statistic in (5.1) tell us the model fits well. Later we can get the estimated residuals. It is shown in Figure 3 that no time trend exists in the residuals. Then we can see that the residual contains no nonlinear components which means that common nonlinear components have been detected.

5.4 Model Specification

Since we have rejected the linearity for both industrial production time series in Sweden and Finland, we discuss the specification of a LSTAR model for the series. Newton-type algorithm is used in the nonlinear regression to get the parameters. [8]

One should mention that there are sometimes numerical problems in the estimation of LSTAR models and that they are related to estimating

[8]See e.g. Teräsvirta (1994) for a discussion.

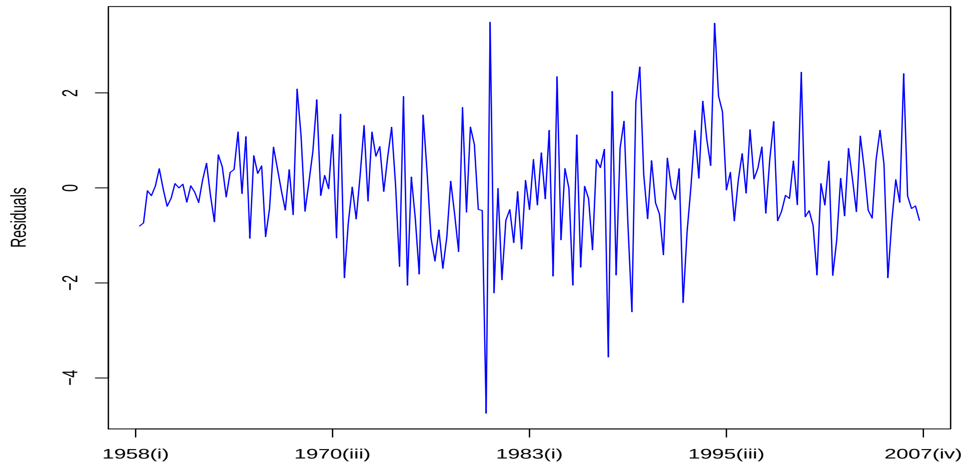

Figure 3: The estimated residuals plot for the linear combination

the slope parameter γ of the transition function. For γ is not a scale-free parameter as its value depends on the magnitude of the values of the transition variable z_t (Teräsvirta (1998)). In order to reduce this dependence it is advisable to standardize the exponent of the transition function by dividing it by the sample standard deviation.

For the Swedish industrial production data, we get the estimated LSTAR model in (5.2) that γ is equal to 3.919 and locational parameter c is equal to 10.416

$$y_t = .220 + 1.003y_{t-1} + (.300 + .700y_{t-1}) \times (1 + \exp(-3.919(\Delta y_{t-1} - 10.416)))^{-1} + \hat{\mu}_t \tag{5.2}$$

Similarly, we obtain the LSTAR model for the Finnish industrial production data in (5.3) for which γ is 45.589 and c is 12.019.

$$y_t = .099 + 1.009y_{t-1} + (.800 + 1.000y_{t-1}) \times (1 + \exp(-45.589(\Delta y_{t-1} - 12.019)))^{-1} + \hat{\mu}_t \tag{5.3}$$

It is clear that the slope parameter in (5.3) is much larger than it in (5.2) which means the regime transition from one to the other is much faster in the Finnish industrial production data compared with the Swedish industrial production. In order to validate this, one should check Figure 2 according to the location parameter. In fact, Figure 2 shows there is a very sharp data jump in Finnish industrial production around the year 1990 compared with the data break in Swedish industrial production around the year 1978 which produces evidence in support of our estimation result.

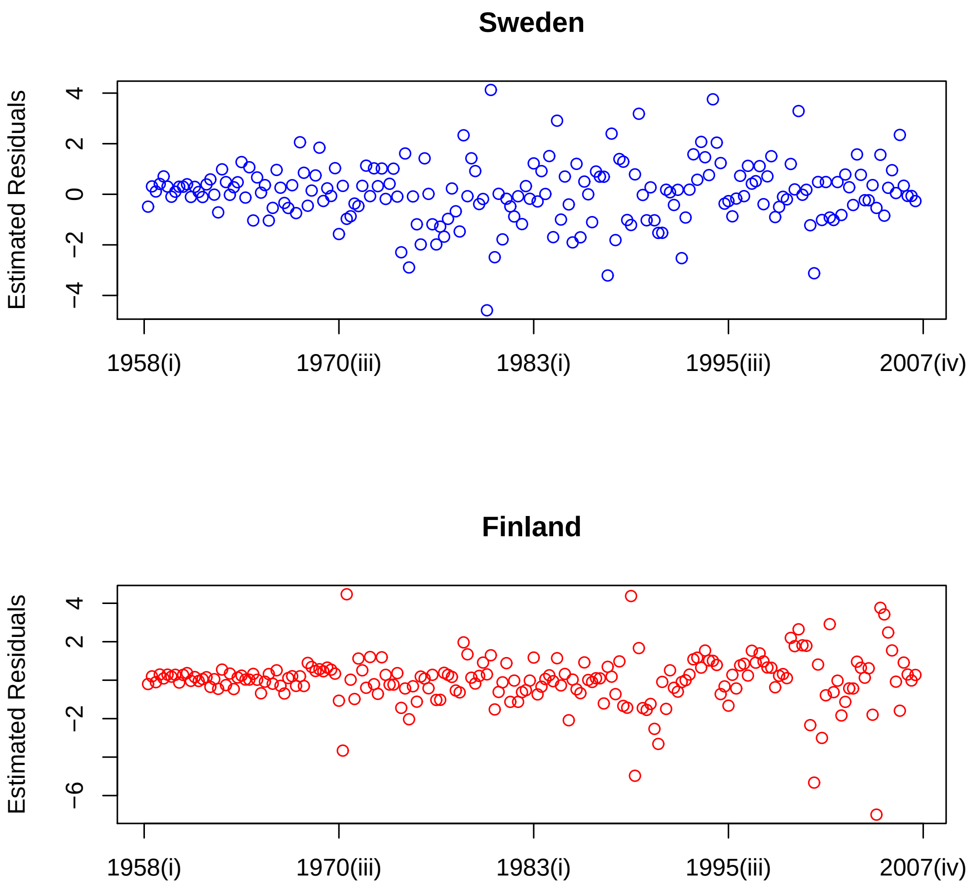

Figure 4: The estimated residuals in industrial production data

To check whether our models full fill the assumptions in (2.1) and (2.2) we have declared in Section 2, we have to check the estimated residuals for both models. Figure 4 shows the estimated residuals in (5.2) and (5.3), from which we know our models have vanished the trend in Figure 2 that maybe caused by the first ordered delay parameter of Δy_{t-1}. Furthermore, the residuals full fill the n.i.d assumption. It is worth to mention γ should not be to small for a good fitted model since small γ may lead to a linearity of the series.

6. Concluding Remarks

In this paper we propose a LSTAR model that accommodates a nonlinear change in dynamics with transition function containing first ordered delay parameter of Δy_{t-1}. Our inference about unit roots is based on the least square estimators obtained from auxiliary testing equations. In order to find the asymptotic distributions for the test, we generalize theoretical results that are derived in the unit-root literature. An important advantage of our tests is that they are computationally easy to carry out.

The LSTAR models in industrial production have practical implications. When industrial production are being forecast using these models, the estimated forecast densities are asymmetric (Skalin & Teräsvirta (2002)). This is important information for the policymakers. Furthermore, in some situations where the null hypothesis of linearity is rejected it may be of interest to try to detect the nonlinear components. If this is successfully done, the estimated model is likely to contain useful information about the nature of parameter change.

References

[1] Ahn, S. K. & Reinsel, G., 1988. Nested reduced-rank autoregressive models for multiple time series. *Journal of the American Statistical Association*, 83(403), p.849–56.

[2] Anderson, H. M. & Vahid, F., 1998. Testing multiple equation systems for common nonlinear components. *Journal of Econometrics*, 84(1), p.1–36.

[3] Cox, J. C. & Ingersoll, J. E. & Jr, & Ross, S. A., 1985. A theory of the term structure of interest rates. *The Econometric Society*, 53(2), p.385–407.

[4] Dickey, D. A. & Fuller, W. A., 1979. Distribution of the estimators for autoregressive time series with a unit root. *Journal of the American Statistical Association*, 74(366), p.427–31.

[5] Engle, R. F. & Kozicki, S., 1993. Testing for common features. *Journal of Business* & *Economic Statistics*, 11(4), p.369–80.

[6] Hamilton, J. D., 1994. *Time Series Analysis*. Princeton: Princeton University Press.

[7] He, C. & Sandberg, R., 2006. Dickey-fuller type of tests against nonlinear dynamic models. *Oxford Bulletin of Economics and Statistics*, 68(s1), p.835–61.

[8] Luukkonen, R. & Saikkonen, P. & Teräsvirta, T., 1988. Testing linearity against smooth transition autoregressive models. *Biometrika*, 75(3), p.491–99.

[9] Phillips, P. C. B., 1987. Time series regression with a unit root. *Econometrica*, 55(2), p.277–301.

[10] Phillips, P. C. B. & Perron, P., 1988. Testing for a unit root in time series regression. *Biometrika*, 75(2), p.335–46.

[11] Skalin, J. & Teräsvirta, T., 2002. Modeling asymmetries and moving equilibria in unemployment rates. *Macroeconomic Dynamics*, 6(262), p.202–41

[12] Teräsvirta, T., 1994. Specification, estimation, and evaluation of smooth transition autoregressive molders. *Journal of the American Statistical Association*, 89(425), p.208–18.

[13] Teräsvirta, T., 1998. Modeling economic relationships with smooth transition regressions. *Handbook of Applied Economic Statistics*, 155, p.507–52.

[14] Teräsvirta, T. & Anderson, H. M., 1992. Characterizing nonlinearities in business cycles using smooth transition autoregressive models. *Journal of Applied Econometrics*, 7(S), p.S119–36.

A. Proofs of Theorems

Proof of Theorem 1. If $(\mu_t)_{t=1}^{\infty}$ satisfy Assumption 1, define

$$\upsilon_t = \sum_{j=0}^{\infty} c_j L^j \mu_t \tag{A.1}$$

where L^j is the lag operator. Notice that

$$\begin{aligned} T^{-1}\sum_{t=2}^{T} \upsilon_{t-1}^2 \mu_t^2 =& T^{-1}\sum_{t=2}^{T} \left(\upsilon_{t-1}^2 \mu_t^2 - E\left(\upsilon_{t-1}^2 \mu_t^2\right)\right) \\ &+ T^{-1}\sum_{t=2}^{T} E\left(\upsilon_{t-1}^2 \mu_t^2\right) \xrightarrow{p} \gamma_0 \sigma^2 \end{aligned} \tag{A.2}$$

holds because

$$T^{-1}\sum_{t=2}^{T} \left(\upsilon_{t-1}^2 \mu_t^2 - E\left(\upsilon_{t-1}^2 \mu_t^2\right)\right) \xrightarrow{p} 0 \tag{A.3}$$

and

$$E\left(\upsilon_{t-1}^2 \mu_t^2\right) = \gamma_0 \sigma_\mu^2. \tag{A.4}$$

Similarly, we can know that

$$\begin{aligned} T^{-3/2}\sum_{t=2}^{T} \upsilon_{t-1}^2 \mu_t =& T^{-3/2}\sum_{t=2}^{T} \left(\upsilon_{t-1}^2 \mu_t - E\left(\upsilon_{t-1}^2 \mu_t\right)\right) \\ &+ T^{-3/2}\sum_{t=2}^{T} E\left(\upsilon_{t-1}^2 \mu_t\right) \xrightarrow{p} c_s \sigma^2 \end{aligned} \tag{A.5}$$

and

$$T^{-1/2}\sum_{t=2}^{T} y_{t-1} \mu_{t-1}^2 = \sigma^3 W(1) \tag{A.6}$$

Furthermore,

$$T^{-1}\sum_{t=1}^{T}1=1 \tag{A.7}$$

$$T^{-3/2}\sum_{t=2}^{T}y_{t-1}\xrightarrow{L}\sigma\int_0^1 W\left(r\right)dr \tag{A.8}$$

$$T^{-1}\sum_{t=2}^{T}y_{t-1}\Delta y_t=T^{-1}\sum_{t=1}^{T}y_{t-1}\mu_t\xrightarrow{L}\frac{1}{2}\sigma^2\left(\left(W\left(r\right)\right)^2-1\right) \tag{A.9}$$

$$T^{-2}\sum_{t=2}^{T}y_{t-1}^2\xrightarrow{L}\sigma^2\int_0^1\left(W\left(r\right)\right)^2dr \tag{A.10}$$

The proofs were give by Hamilton 1994 in pp. 479-484.

The above result implies

$$\begin{aligned}
&\sum_{t=3}^{T}\begin{pmatrix}1\\ y_{t-1}\\ y_{t-2}\\ y_{t-1}\Delta y_{t-1}\end{pmatrix}\begin{pmatrix}1 & y_{t-1} & y_{t-2} & y_{t-1}\Delta y_{t-1}\end{pmatrix}\\
&=\begin{pmatrix}
\sum_{t=3}^{T}1 & \sum_{t=3}^{T}y_{t-1} & \sum_{t=3}^{T}y_{t-2} & \sum_{t=3}^{T}y_{t-1}\Delta y_{t-1}\\
\sum_{t=3}^{T}y_{t-1} & \sum_{t=3}^{T}y_{t-1}^2 & \sum_{t=3}^{T}y_{t-1}y_{t-2} & \sum_{t=3}^{T}y_{t-1}^2\Delta y_{t-1}\\
\sum_{t=3}^{T}y_{t-2} & \sum_{t=3}^{T}y_{t-1}y_{t-2} & \sum_{t=3}^{T}y_{t-2}^2 & \sum_{t=3}^{T}y_{t-1}y_{t-2}\Delta y_{t-1}\\
\sum_{t=3}^{T}y_{t-1}\Delta y_{t-1} & \sum_{t=3}^{T}y_{t-1}^2\Delta y_{t-1} & \sum_{t=3}^{T}y_{t-1}y_{t-2}\Delta y_{t-1} & \sum_{t=3}^{T}y_{t-1}^2\Delta y_{t-1}^2
\end{pmatrix}\\
&=\begin{pmatrix}
O_p\left(T\right) & O_p\left(T^{3/2}\right) & O_p\left(T^{3/2}\right) & O_p\left(T\right)\\
O_p\left(T^{3/2}\right) & O_p\left(T^2\right) & O_p\left(T^2\right) & O_p\left(T^{3/2}\right)\\
O_p\left(T^{3/2}\right) & O_p\left(T^2\right) & O_p\left(T^2\right) & O_p\left(T^{3/2}\right)\\
O_p\left(T\right) & O_p\left(T^{3/2}\right) & O_p\left(T^{3/2}\right) & O_p\left(T\right)
\end{pmatrix}
\end{aligned} \tag{A.11}$$

and

$$\sum_{t=3}^{T} \begin{pmatrix} \mu_t \\ y_{t-1}\mu_t \\ y_{t-2}\mu_t \\ y_{t-1}\Delta y_{t-1}\mu_t \end{pmatrix} = \begin{pmatrix} O_p\left(T^{1/2}\right) \\ O_p\left(T\right) \\ O_p\left(T\right) \\ O_p\left(T^{1/2}\right) \end{pmatrix} \tag{A.12}$$

We have

$$\boldsymbol{\gamma}_T\left(\boldsymbol{b}_T - \boldsymbol{\beta}_1\right) \xrightarrow{d} \boldsymbol{Q}^{-1}\boldsymbol{h}_1 \tag{A.13}$$

□

Proof of (4.5). Since

$$\begin{aligned} \boldsymbol{\gamma}_{T1}\left(\boldsymbol{b}_{T1} - \boldsymbol{\beta}_1\right) &= \boldsymbol{\gamma}_{T1}\left(\sum_{t=1}^{T} x'x\right)^{-1} \boldsymbol{\gamma}_{T1}\boldsymbol{\gamma}_{T1}^{-1}\left(\sum_{t=1}^{T} x_t\mu_t\right) \\ &= \left(\boldsymbol{\gamma}_{T1}^{-1}\left(\sum_{t=1}^{T} x'x\right)^{-1}\boldsymbol{\gamma}_{T1}^{-1}\right)^{-1}\left(\boldsymbol{\gamma}_{T1}^{-1}\left(\sum_{t=1}^{T} x_t\mu_t\right)\right) \end{aligned} \tag{A.14}$$

and the matrix has the property that

$$\boldsymbol{\gamma}_T\boldsymbol{R} = \boldsymbol{R}\boldsymbol{\gamma}_T \tag{A.15}$$

Apply the above results, we can easily prove (4.5) □

Proof of Theorem 2. Suppose that the true process is a random walk with drift:

$$y_t = \alpha + y_{t-1} + \mu_t \tag{A.16}$$

which implies that

$$y_t = y_0 + \alpha t + (\mu_1 + \mu_2 + ... + u_t) = y_0 + \alpha t + \xi_t \tag{A.17}$$

where ξ_t is an unit-root without drift.

Thus we can apply the results in the proof section of Theorem (1). We have

$$\sum_{t=1}^{T} \alpha^2 = O_p(T) \tag{A.18}$$

$$\sum_{t=2}^{T} y_{t-1} = \underbrace{\sum_{t=2}^{T} y_0}_{O_p(T)} + \underbrace{\sum_{t=2}^{T} \alpha(t-1)}_{O_p(T^2)} + \underbrace{\sum_{t=2}^{T} \xi_{t-1}}_{O_p(T^{3/2})} = O_p(T^2) \tag{A.19}$$

$$\begin{aligned} \sum_{t=2}^{T} y_{t-1}\Delta y_t &= \sum_{t=2}^{T} \left(y_0 + \alpha(t-1) + \xi_{t-1}\right)(\alpha + \mu_t) \\ &= \underbrace{\alpha \sum_{t=2}^{T} y_{t-1}}_{O_p(T^2)} + \underbrace{\sum_{t=2}^{T} y_0\mu_t}_{O_p(T^{1/2})} + \underbrace{\sum_{t=2}^{T} \alpha(t-1)\mu_t}_{O_p(T^{3/2})} + \underbrace{\sum_{t=2}^{T} \xi_t\mu_t}_{O_p(T)} \\ &= O_p(T^2) \end{aligned} \tag{A.20}$$

From (A.20), we know that $y_{t-1}\Delta y_t$ is asymptotically dominated by the time trend y_{t-1}, while y_{t-1} is asymptotically dominated by the time trend $\alpha(t-1)$. In large samples, it is as if the explanatory variable $y_{t-1}\Delta y_t$ were replaced by the time trend $\alpha(t-1)$. Similar result can be obtained as follows.

$$\begin{aligned} \sum_{t=2}^{T} y_{t-1}^2 &= \sum_{t=2}^{T} \left(y_0 + \alpha(t-1) + \xi_{t-1}\right)^2 \\ &= \underbrace{\sum_{t=2}^{T} y_0^2}_{O_p(T)} + \underbrace{\sum_{t=2}^{T} (\alpha(t-1))^2}_{O_p(T^3)} + \underbrace{\sum_{t=2}^{T} \xi_{t-1}^2}_{O_p(T^2)} \\ &\quad + \underbrace{2\sum_{t=2}^{T} y_0\alpha(t-1)}_{O_p(T^2)} + \underbrace{2\sum_{t=2}^{T} \alpha(t-1)\xi_{t-1}}_{O_p(T^{5/2})} + \underbrace{2\sum_{t=2}^{T} y_0\xi_{t-1}}_{O_p(T^{3/2})} \\ &= O_p(T^3) \end{aligned} \tag{A.21}$$

$$\begin{aligned}
\sum_{t=2}^{T} y_{t-1}^2 \Delta y_t =& \sum_{t=2}^{T} \left(y_0 + \alpha(t-1) + \xi_{t-1}\right)^2 (\alpha + \mu_t) \\
=& \sum_{t=2}^{T} \alpha y_{t-1}^2 + \sum_{t=2}^{T} \left(y_0 + \alpha(t-1) + \xi_{t-1}\right)^2 \mu_t \\
=& \underbrace{\sum_{t=2}^{T} \alpha y_{t-1}^2}_{O_p(T^3)} + \underbrace{\sum_{t=2}^{T} y_0^2 \mu_t}_{O_p(T^{1/2})} + \underbrace{\sum_{t=2}^{T} \left(\alpha(t-1)\right)^2 \mu_t}_{O_p(T^3)} + \underbrace{\sum_{t=2}^{T} \xi_{t-1}^2 \mu_t}_{O_p(T^{3/2})} \\
&+ 2 \underbrace{\sum_{t=2}^{T} y_0 \alpha(t-1) \mu_t}_{O_p(T^2)} + 2 \underbrace{\sum_{t=2}^{T} \alpha(t-1) \xi_{t-1} \mu_t}_{O_p(T^2)} + 2 \underbrace{\sum_{t=2}^{T} y_0 \xi_{t-1} \mu_t}_{O_p(T)} \\
=& O_p(T^3)
\end{aligned} \tag{A.22}$$

holds because $\sum_{t=1}^{T} t^p \xi_{t-1} \mu_t = O_p(T^{p+1})$. See He and Sandberg 2006 for a proof.

Furthermore,

$$\begin{aligned}\sum_{t=2}^{T} y_{t-1}^2 \Delta y_t^2 =& \sum_{t=2}^{T} \left(y_0 + \alpha(t-1) + \xi_{t-1}\right)^2 \left(\alpha + \mu_t\right)^2 \\ =& \sum_{t=2}^{T} \left(y_0 + \alpha(t-1) + \xi_{t-1}\right)^2 \left(\alpha^2 + \mu_t^2 + 2\alpha\mu_t\right) \\ =& \underbrace{\sum_{t=2}^{T} \alpha^2 y_{t-1}^2}_{O_p(T^3)} + \underbrace{\sum_{t=2}^{T} 2\alpha y_{t-1}^2 \mu_t}_{O_p(T^3)} + \sum_{t=2}^{T} \left(y_0 + \alpha(t-1) + \xi_{t-1}\right)^2 \mu_t^2 \\ =& \underbrace{\sum_{t=2}^{T} \alpha^2 y_{t-1}^2}_{O_p(T^3)} + \underbrace{\sum_{t=2}^{T} 2\alpha y_{t-1}^2 \mu_t}_{O_p(T^3)} + \underbrace{\sum_{t=2}^{T} y_0^2 \mu_t^2}_{O_p(T)} \\ &+ \underbrace{\sum_{t=2}^{T} \left(\alpha(t-1)\right)^2 \mu_t^2}_{O_p(T^3)} + \underbrace{\sum_{t=2}^{T} \xi_{t-1}^2 \mu_t^2}_{O_p(T)} + \underbrace{2\sum_{t=2}^{T} y_0 \alpha(t-1) \mu_t^2}_{O_p(T^3)} \\ &+ \underbrace{2\sum_{t=2}^{T} \alpha(t-1)\xi_{t-1}\mu_t^2}_{O_p(T^{5/2})} + \underbrace{2\sum_{t=2}^{T} y_0 \xi_{t-1}\mu_t^2}_{O_p(T^{1/2})} \\ =& O_p(T^3)\end{aligned} \tag{A.23}$$

and

$$\begin{aligned}\sum_{t=2}^{T} y_{t-1} \Delta y_t \mu_t =& \sum_{t=2}^{T} \left(y_0 + \alpha(t-1) + \xi_{t-1}\right)\left(\alpha + \mu_t\right)\mu_t \\ =& \underbrace{\sum_{t=2}^{T} y_{t-1}\mu_t^2}_{O_p(T^{3/2})} + \underbrace{\sum_{t=2}^{T} y_{t-1}\mu_t}_{O_p(T^{3/2})} = O_p(T^{3/2}).\end{aligned} \tag{A.24}$$

Thus, we can get the asymptotic distribution that the true model is random walk with drift

$$\boldsymbol{\gamma}_{T2}^{-1}\left(\sum_{t=1}^{T}\boldsymbol{x}\boldsymbol{x}'\right)\boldsymbol{\gamma}_{T2}^{-1}\xrightarrow{p}\boldsymbol{\Omega}^{-1} \tag{A.25}$$

where $\boldsymbol{\gamma}_{T2} = diag\left(T^{1/2}, T^{3/2}, T^{3/2}, T^{3/2}\right)$ and $\boldsymbol{\Omega}$ is a 4×4 diagonal matrix and ω_{ij} is a function of α. And,

$$\boldsymbol{\gamma}_{T2}^{-1}\left(\sum_{t=1}^{T}\boldsymbol{x}\boldsymbol{\mu}'\right)\xrightarrow{L}\boldsymbol{N}\left(\boldsymbol{0}, \sigma^2\boldsymbol{\Omega}\right) \tag{A.26}$$

It follows that

$$\boldsymbol{\gamma}_{T2}\left(\boldsymbol{b_{T2}} - \boldsymbol{\beta_2}\right)\xrightarrow{L}\boldsymbol{N}\left(\boldsymbol{0}, \sigma^2\boldsymbol{\Omega}^{-1}\right) \tag{A.27}$$

□

B . DGPs in the Monte Carlo

These were based on two DGPs given by (5.2)

$$\begin{aligned} y_t =& .220 + 1.003 y_{t-1} + (.300 + .700 y_{t-1}) \\ & \times \left(1 + \exp\left(-3.919\left(\Delta y_{t-1} - 10.416\right)\right)\right)^{-1} + \hat{\mu}_t \end{aligned} \tag{B.1}$$

and (5.3)

$$\begin{aligned} y_t =& .099 + 1.009 y_{t-1} + (.800 + 1.000 y_{t-1}) \\ & \times \left(1 + \exp\left(-45.589\left(\Delta y_{t-1} - 12.019\right)\right)\right)^{-1} + \hat{\mu}_t \end{aligned} \tag{B.2}$$

where $\mu_{1t},\ \mu_{2t} \sim n.i.d(0,1)$. Later, a linear regression is carried, and the estimator of α can be obtained from

$$y_{1t} = \hat{\alpha}_0 + \hat{\alpha} y_{2t} + \hat{\mu}_t \tag{B.3}$$

For a universal case of DGPs discussions, see Anderson and Vahid (1998).

C. Data and Simulation Codes

The data sets are the industrial production data of Sweden and Finland which are available at OECD website
`http://stats.oecd.org/wbos/default.aspx?datasetcode=MEI_CLI`
For the code, please mail to the author: *feng@ilee.name*