Journal of the Korean Statistical Society (2007), **36***: 2, pp 257–274*

TESTING FOR SMOOTH TRANSITION NONLINEARITY IN PARTIALLY NONSTATIONARY VECTOR AUTOREGRESSIONS†

BYEONGSEON SEO[1]

ABSTRACT

This paper considers the tests for the presence of smooth transition nonlinearity in the partially nonstationary vector autoregressive model. The transition parameters cannot be identified under the null hypothesis of linearity, and therefore this paper develops the tests for smooth transition nonlinearity, the associated asymptotic theory and the bootstrap inference. The Monte Carlo simulation evidence shows that the bootstrap inference generates moderate size and power performances.

AMS 2000 subject classifications. Primary 62M10; Secondary 62H15.
Keywords. Nonlinear adjustment, smooth transition.

1. INTRODUCTION

The smooth transition autoregressive (STAR) model was proposed by Chan and Tong (1986) as a generalization of the threshold autoregressive (TAR) model, and since then it has attracted wide attention in the recent literature on the business cycles and the equilibrium parity relationships of commodity prices, exchange rates and equity prices. Economic behavior is affected by asymmetric transaction costs and institutional rigidities, and thus a large number of studies (for example, Neftci, 1984; Teräsvirta and Anderson, 1992; Michael *et al.*, 1997) have shown that many economic variables and relations display asymmetry and nonlinear adjustment.

One of the most crucial issues in models of this kind is testing for the presence of nonlinear adjustment with the null hypothesis of linearity. The smooth transition model entails transition parameters, which cannot be identified under

Received September 2006; accepted January 2007.
†This work was supported by the Korea Research Foundation Grant (KRF-2003-041-B00069).
[1]Department of Economics, Soongsil University, Seoul 156-743, Korea (e-mail: seo@ssu.ac.kr)

the null hypothesis, and as a result the likelihood ratio statistic has the nonstandard distribution. The problem of unidentified parameters has been discussed by Davies (1987), Andrews (1993) and Hansen (1996). However, this issue in the smooth transition model has been neglected, and thus this paper develops the tests for smooth transition nonlinearity and the associated distribution theory.

Many empirical studies have found evidence on the presence of stochastic nonlinear adjustment in equilibrium relations such as the purchasing power parity. For example, Michael *et al.* (1997), considering the equilibrium model of real exchange rate in the presence of transaction costs, found strong evidence of nonlinear adjustment, which conforms to the exponential smooth transition model. There exists a huge literature and it is growing in this area. This paper develops the tests for nonlinear adjustment in the partially nonstationary vector autoregressive model, and thereby contributes to the literature.

One technical difficulty is to estimate the smooth transition model. As noted by Haggan and Ozaki (1981) and Teräsvirta (1994), it is difficult to estimate the smooth transition parameters jointly with the other slope parameters. The gradient of the transition parameter forces its estimate to blow up to infinity; thus, we cannot depend on the standard estimation algorithm. Our tests are based on the Lagrange multiplier statistic, which can be calculated under the null hypothesis.

This paper finds that our tests have the asymptotic distribution, which is based on the Gaussian process. However, the asymptotic distribution depends on the nuisance parameters and the covariances are data-dependent, therefore the tabulation of asymptotic distribution is not feasible. This paper suggests the bootstrap inference to approximate the sampling distribution of the test statistics. Simulation evidence shows that the bootstrap inference generates moderate size and power performances.

We denote $\Rightarrow$ as weak convergence with respect to the uniform metric and $\xrightarrow{p}$ as convergence in probability. The expression $|\cdot|$ represents the matrix norm, that is, $|A| = (\text{tr}A'A)^{1/2}$ and $||X||_p = (E|X|^p)^{1/p}$. Also, vec$(\cdot)$ is the column-stacking operator.

This paper is organized as follows. Section 2 introduces the smooth transition vector error correction model and develops the tests for smooth transition nonlinearity. Section 3 explores the asymptotic distribution of the proposed tests. Section 4 provides the Monte Carlo simulation evidence on the size and power of the tests.

2. MODEL

Consider a p-dimensional nonstationary time series x_t generated by a smooth transition vector error correction model as follows:

$$\Delta x_t = A' z_t(\beta) + D' z_t(\beta) F(q_t; \lambda, \gamma) + u_t, \tag{2.1}$$

where $z_t(\beta) = (1, w_{t-1}(\beta), \Delta x'_{t-1}, \Delta x'_{t-2}, \dots, \Delta x'_{t-l})'$.

We assume that the cointegration space is known and equals 1. Thus, the normalized cointegrating relationship $w_t(\beta) = x_{1t} + \beta' x_{2t}$ is stationary, where the cointegrating coefficient β and the corresponding variable x_{2t} are $(p-1)$-dimensional. The regressor z_t is a k-dimensional vector, where $k = pl + 2$, and the coefficient matrices A and D are $k \times p$.

We define the σ-field $\mathcal{F}_t$ generated by x_{t-i} for $i = 1, 2, \dots$. We assume that the error u_t is a vector-valued Martingale difference sequence with a finite variance $\Sigma = E(u_t u_t') < \infty$.

The transition function $F(q_t; \lambda, \gamma)$ depends on the transition variable q_t and the associated parameters. The functional form can be specified in several ways, depending on the characteristics of nonlinear adjustment. As in Teräsvirta (1994), this paper considers the exponential and logistic transition functions as follows:

$$F(q_t; \lambda, \gamma) = 1 - \exp\left[-\lambda (q_t - \gamma)^2\right], \ \lambda > 0 \tag{2.2}$$

and

$$F(q_t; \lambda, \gamma) = \frac{1}{1 + \exp[-\lambda (q_t - \gamma)]}, \ \lambda > 0. \tag{2.3}$$

The exponential specification (2.2) allows for a smooth transition based on the inverted normal density function, while the logistic specification (2.3) indicates a smooth transition based on the cumulative logistic distribution. The exponential specification (2.2) implies a symmetric three-regime transition, where the short-run dynamics are explained by the coefficient matrix A in the mid regime, and the coefficient $A + D$ corresponds to the tail regimes. The transition rate parameter λ determines the speed of transition. As λ increases, the transition from the mid regime to both tail regimes, and its reverse, can be made quickly. If $\lambda = 0$, there is no transition and only the mid regime is prevalent. If λ approaches ∞, the mid regime disappears and our model reduces to the linear error correction model. Both cases lead to the linear error correction model. Thus, smooth transition has meaning only if $0 < \lambda < \infty$. The location parameter γ determines the average location of transition. We assume that γ lies inside the support of the transition variable.

The logistic specification (2.3) posits a two-regime transition, where the short-run dynamics are explained by the coefficient matrix A in the first regime, and the coefficient $A + D$ corresponds to the second regime. As in the exponential specification, the transition rate parameter λ determines the speed of transition, and the location parameter γ determines the average location of transition. As λ increases, the transition from the first regime to the second regime, and its reverse, can be made quickly. If $\lambda = 0$, there is no transition and only the first regime remains. This case leads to the linear error correction model, and thus we assume that $\lambda > 0$. As λ approaches ∞, the logistic transition converges to the threshold transition such that $\lim_{\lambda\to\infty} F(q_t; \lambda, \gamma) = 1(q_t \geq \gamma)$, where $1(\cdot)$ is the indicator function. Then, our model is the same as the threshold vector error correction model, which was considered in Hansen and Seo (2002). As in the exponential transition, we assume that γ lies inside the support of the transition variable.

The transition variable q_t is a stationary transformation of the predetermined variables, for example, $q_t = w_{t-1}(\tilde{\beta})$ and $q_t = |w_{t-1}(\tilde{\beta})|$, where $\tilde{\beta}$ is the cointegrating vector estimate computed from the linear model, which will be defined later. Here, we set $q_t = w_{t-1}(\tilde{\beta})$. Main results do not change when other predetermined variables are used as the transition variable if the variable is stationary.

Although the smooth transition specification allows for the linear and threshold models, identification may fail when the transition rate approaches 0 or ∞. Thus, we impose this restriction by assuming that $\lambda \in [\lambda_L, \lambda_U] \subset R^+$. For a monotonic transformation $h(\cdot) : R^{(0,\infty)} \to R^{(0,1)}$, we consider a reparametrization satisfying $\lambda = h^{-1}(\nu_1)$ for $\nu_1 \subset (0,1)$. Thus, $\nu_1 \in [\nu_{1L}, \nu_{1U}] \subset (0,1)$.

Also, the smooth transition has meaning only if $0 < P(q_t \leq \gamma) < 1$. Using the monotonic transformation $\nu_2 = P(q_t \leq \gamma)$, we impose this constraint by assuming that $\nu_2 \in [\nu_{2L}, \nu_{2U}] \subset (0,1)$ or $\gamma \in [\gamma_L, \gamma_U]$, where $\nu_{2L} = P(q_t \leq \gamma_L)$ and $\nu_{2U} = P(q_t \leq \gamma_U)$.

Let $\nu = (\nu_1, \nu_2)$ for $\nu_1 = h(\lambda)$ and $\nu_2 = P(q_t \leq \gamma)$. In the simulation and empirical results of this paper, we set $\nu_{1L} = 1 - \nu_{1U} = 0.05$ and $\nu_{2L} = 1 - \nu_{2U} = 0.10$.

The smooth transition model has smoothly varying coefficients depending on the current state q_t. The nonlinear dynamics can be explained by the coefficient matrices A and D. Our model (2.1) allows all short-run coefficients to vary. However, the parsimonious specification may relieve computational cost if it does not affect the validity condition. In this respect, we may allow the coefficient on the error correction term or the coefficients on the error correction term and intercept to vary while setting the other coefficients to be constant.

When the coefficient matrix D is zero, our model can be reduced as follows,

$$\Delta x_t = A' z_t(\beta) + u_t = \mu + \alpha w_{t-1}(\beta) + \sum_{i=1}^{l} \Gamma_i \Delta x_{t-i} + u_t. \tag{2.4}$$

Hence, the null and alternative hypotheses for testing linearity in adjustment dynamics can be postulated as follows:

$$H_0 : D = 0 \text{ against } H_1 : D \neq 0.$$

We define the parameter vector

$$\theta = \text{vec}(D, A, \beta, \Sigma) \in \Theta.$$

The true parameter value is denoted as θ_0. The log-likelihood function, with the auxiliary condition that u_t is normally distributed, is given by

$$\mathcal{L}_n(\theta, \nu) = -\frac{1}{2} \sum_{t=1}^{n} \left[\log |\Sigma| + u_t'(\theta, \nu) \Sigma^{-1} u_t(\theta, \nu) \right], \tag{2.5}$$

where $u_t(\theta, \nu)$ is defined in (2.1).

We denote $\hat{\theta}(\nu)$ as the maximum likelihood estimator (MLE) of θ for fixed ν. As noted by Haggan and Ozaki (1981) and Teräsvirta (1994), technical difficulty arises when the transition parameters ν are jointly estimated with the other slope parameters θ. Particularly, the estimate of the transition rate parameter tends to be inflated and the convergence cannot be made easily. In a practical sense, the estimation of the transition rate requires a large number of observations because, depending on the parameter values of the transition rate, the convergence becomes slower. To estimate the transition rate, Haggan and Ozaki (1981) suggested a conditional least squares with a grid search on the transition rate. However, our tests do not require the estimation of the smooth transition parameters, and therefore we treat ν as fixed until we define the test statistic for unknown ν.

Under the null hypothesis of linearity, the cointegrating vector can be estimated by reduced rank regression, and the short-run parameters can be estimated by least squares. We denote $\tilde{\beta}$ and $\tilde{A}$ as the linear estimates of the cointegrating vector and short-run parameters, respectively.

Once β is known, the smooth transition error correction model (2.1) is linear in parameters A and D for fixed ν. We denote $\hat{A}(\beta,\nu)$ and $\hat{D}(\beta,\nu)$ as the MLE for given β and ν. Thus, the MLE $\hat{D}(\beta,\nu)$ is given by

$$\hat{D}(\beta,\nu) = \left(\sum_{t=1}^{n} z_{2t}^{*}(\beta,\nu) z_{2t}^{*'}(\beta,\nu)\right)^{-1} \sum_{t=1}^{n} z_{2t}^{*}(\beta,\nu)\Delta x_t', \tag{2.6}$$

where

$$z_{2t}^{*}(\beta,\nu) = z_{2t}(\beta,\nu) - \sum_{t=1}^{n} z_{2t}(\beta,\nu) z_t'(\beta)\left(\sum_{t=1}^{n} z_t(\beta) z_t'(\beta)\right)^{-1} z_t(\beta)$$

and

$$z_{2t}(\beta,\nu) = z_t(\beta)F(q_t;\nu).$$

For the null hypothesis $H_0 : D = 0$, we define the score function $g_n(\nu)$ as follows:

$$g_n(\nu) = \frac{1}{\sqrt{n}}\text{vec}\left(\sum_{t=1}^{n} z_{2t}^{*}(\tilde{\beta},\nu)\Delta x_t'\right), \tag{2.7}$$

where $\tilde{\beta}$ is the linear cointegrating vector estimator.

We define $\tilde{u}_t = \Delta x_t - \tilde{A}' z_t(\tilde{\beta})$ and $\tilde{v}_t(\nu) = \tilde{u}_t F(q_t;\nu)$. To allow for time-varying conditional variances, we define the heteroskedasticity-robust covariance estimator of the score function as follows:

$$\begin{aligned} V_n(\nu) &= \Omega_{22n}(\nu) - Q_{21n}(\nu) Q_{11n}^{-1}\Omega_{12n}(\nu) - \Omega_{21n}(\nu) Q_{11n}^{-1} Q_{12n}(\nu) \\ &\quad + Q_{21n}(\nu) Q_{11n}^{-1}\Omega_{11n} Q_{11n}^{-1} Q_{12n}(\nu), \end{aligned}$$

where

$$\begin{aligned} \Omega_{11n} &= \frac{1}{n}\sum_{t=1}^{n}\left(\tilde{u}_t\tilde{u}_t' \otimes z_t(\tilde{\beta}) z_t'(\tilde{\beta})\right), \\ \Omega_{12n}(\nu) &= \frac{1}{n}\sum_{t=1}^{n}\left(\tilde{u}_t\tilde{v}_t'(\nu) \otimes z_t(\tilde{\beta}) z_t'(\tilde{\beta})\right), \\ \Omega_{22n}(\nu) &= \frac{1}{n}\sum_{t=1}^{n}\left(\tilde{v}_t(\nu)\tilde{v}_t'(\nu) \otimes z_t(\tilde{\beta}) z_t'(\tilde{\beta})\right), \\ Q_{11n} &= \frac{1}{n}\sum_{t=1}^{n}\left(I \otimes z_t(\tilde{\beta}) z_t'(\tilde{\beta})\right), \\ Q_{12n}(\nu) &= \frac{1}{n}\sum_{t=1}^{n}\left(I \otimes z_t(\tilde{\beta}) z_{2t}'(\tilde{\beta},\nu)\right) \end{aligned}$$

and $\Omega_{21n}(\nu) = \Omega_{12n}'(\nu)$ and $Q_{21n}(\nu) = Q_{12n}'(\nu)$.

Thus, the tests for nonlinear adjustment can be based on the following LM statistic:

$$\mathrm{LM}_n(\nu) = g_n(\nu)' V_n^{-1}(\nu) g_n(\nu). \tag{2.8}$$

The LM statistic can be calculated if we have the linear cointegrating vector estimator $\tilde{\beta}$, the residual $\tilde{u}_t$ and the data. We do not need to estimate the smooth transition error correction model, and we can avoid the difficulty of estimating the transition parameters.

The transition parameter ν cannot be identified under the null hypothesis. The argument regarding the unidentified parameter has been raised by Davies (1987), Andrews (1993) and Hansen (1996). Hansen (1996) particularly considered this issue in threshold models.

The LM statistic has been defined for fixed ν. This is appropriate only when ν is known. If ν is unknown, the testing procedure is nonstandard because the nuisance parameter appears only under the alternative hypothesis, and the likelihood function is flat under the null hypothesis.

If we assume that ν lies in $\mathcal{N} = [\nu_{1L}, \nu_{1U}] \times [\nu_{2L}, \nu_{2U}] \subset R^{(0,1)} \times R^{(0,1)}$, then the test statistic can be defined as follows:

$$\mathrm{SupLM}_n = \sup_{\nu \in \mathcal{N}} \mathrm{LM}_n(\nu). \tag{2.9}$$

3. MAIN RESULTS

First, we use the representation theorem by Engle and Granger (1987). The linear error correction model (2.4) has the following representation:

$$\Delta x_t = C(L) u_t, \tag{3.1}$$

$$x_t = C(1) \sum_{i=1}^{t} u_i + C^*(L) u_t, \tag{3.2}$$

$$w_t = B' C^*(L) u_t, \tag{3.3}$$

where $C^*(L) = \{C(L) - C(1)\}/(1 - L)$ and $B = (1, \beta')'$.

Therefore, x_t can be decomposed into stochastic trends and a stationary component. The cointegrating vector eliminates the stochastic trends, and thus the cointegrating relationship $w_t(\beta) = (1, \beta')x_t$ is stationary as defined in Engle and Granger (1987).

Let θ_0 be the true parameter value. We denote $w_t = w_t(\beta_0)$ and $z_t =$

$z_t(\beta_0)$. By reparametrization, we define $\nu = (\nu_1, \nu_2)$, where $\nu_1 = h(\lambda)$ and $\nu_2 = P(q_t \leq \gamma)$. We denote $u_t(\nu) = u_t(\theta_0, \nu)$, $F_t(\nu) = F(q_{0t}; \nu)$, $q_{0t} = w_{t-1}(\beta_0)$ and $v_t(\nu) = F_t(\nu)u_t(\nu)$. Note that the error u_t does not depend on ν under the null hypothesis.

ASSUMPTION 3.1.

(i) $\nu \in \mathcal{N} \subset R^{(0,1)} \times R^{(0,1)}$.

(ii) $\{u_t, \mathcal{F}_t\}$ *is a vector-valued Martingale difference sequence with* $\sup_t ||u_t||_4 < \infty$.

(iii) $\sum_{k=1}^{\infty} k|C_k| < \infty$, *where* $\Delta x_t = C(L)u_t = \sum_{k=0}^{\infty} C_k u_{t-k}$.

(iv) The parameter space Θ *is compact.*

(v) $F_t(\nu)$ *is continuously differentiable and* $\sup_t || \sup_{\nu \in \mathcal{N}} |F_t'(\nu)| \,||_2 < \infty$, *where* $F_t'(\nu) = \partial F_t(\nu)/\partial \nu$.

We use Assumption 3.1(v) to show the stochastic equicontinuity of the sum $(1/\sqrt{n})\sum_{t=1}^{n} v_t(\nu)$, where $v_t(\nu) = u_t F_t(\nu)$. The exponential and logistic transition functions satisfy this condition if Assumptions 3.1(i)–(iv) hold.

We need to define weak convergence of the sum $(1/\sqrt{n})\sum_{t=1}^{n} v_t(\nu)$. Thus, we denote $\Rightarrow$ as weak convergence on $\mathcal{N}$ with respect to the uniform metric $\rho(\cdot)$,

$$\rho(g, h) = \sup_{\nu \in \mathcal{N}} |g(\nu) - h(\nu)|,$$

where $|\cdot|$ is the matrix norm.

LEMMA 3.1. *Under Assumption 3.1,*

$$\begin{pmatrix} \frac{1}{\sqrt{n}} \sum_{t=1}^{n} u_t \\ \frac{1}{\sqrt{n}} \sum_{t=1}^{n} v_t(\nu) \end{pmatrix} \Rightarrow \begin{pmatrix} U_1 \\ U_2(\nu) \end{pmatrix} \sim N\left(0, \begin{pmatrix} \Sigma_{11} & \Sigma_{12}(\nu) \\ \Sigma_{21}(\nu) & \Sigma_{22}(\nu) \end{pmatrix}\right), \tag{3.4}$$

where U_1 *and* $U_2(\nu)$ *are Gaussian processes* , $\Sigma_{11} = E(u_t u_t')$, $\Sigma_{12}(\nu) = E(u_t v_t'(\nu))$, $\Sigma_{21}(\nu) = \Sigma_{12}'(\nu)$ *and* $\Sigma_{22}(\nu) = E(v_t(\nu)v_t'(\nu))$.

We use the following lemmas to show the main results.

LEMMA 3.2. *Under the null hypothesis and Assumption 3.1,*

$$\begin{pmatrix} \frac{1}{\sqrt{n}}\sum_{t=1}^{n}(u_t \otimes z_t) \\ \frac{1}{\sqrt{n}}\sum_{t=1}^{n}(v_t(\nu) \otimes z_t) \end{pmatrix} \Rightarrow \begin{pmatrix} W_1 \\ W_2(\nu) \end{pmatrix} \sim N\left(0, \begin{pmatrix} \Omega_{11} & \Omega_{12}(\nu) \\ \Omega_{21}(\nu) & \Omega_{22}(\nu) \end{pmatrix}\right), \tag{3.5}$$

where $\Omega_{11} = E(u_t u_t' \otimes z_t z_t')$, $\Omega_{12}(\nu) = E(u_t v_t'(\nu) \otimes z_t z_t')$, $\Omega_{21}(\nu) = \Omega_{12}'(\nu)$ *and* $\Omega_{22}(\nu) = E(v_t(\nu) v_t'(\nu) \otimes z_t z_t')$.

LEMMA 3.3. *Under the null hypothesis and Assumption 3.1,*

$$\begin{aligned} g_n(\nu) &\Rightarrow W_2(\nu) - Q_{21}(\nu)Q_{11}^{-1}W_1, \\ V_n(\nu) &\xrightarrow{p} \Omega_{22}(\nu) - Q_{21}(\nu)Q_{11}^{-1}\Omega_{12}(\nu) - \Omega_{21}(\nu)Q_{11}^{-1}Q_{12}(\nu) \\ &\qquad + Q_{21}(\nu)Q_{11}^{-1}\Omega_{11}Q_{11}^{-1}Q_{12}(\nu) \equiv V(\nu), \end{aligned}$$

where $Q_{11} = E(I \otimes z_t z_t')$, $Q_{12}(\nu) = E(I \otimes z_t v_t'(\nu))$ *and* $Q_{21}(\nu) = Q_{12}'(\nu)$.

THEOREM 3.1. *Under the null hypothesis and Assumption 3.1,*

$$\mathrm{LM}_n(\nu) \Rightarrow B^b(\nu)' B^b(\nu) \equiv \mathrm{LM}(\nu), \tag{3.6}$$

where $B^b(\nu) = V^{-1/2}(\nu)[W_2(\nu) - Q_{21}(\nu)Q_{11}^{-1}W_1]$.

Therefore,

$$\sup_{\nu \in \mathcal{N}} \mathrm{LM}_n(\nu) \Rightarrow \sup_{\nu \in \mathcal{N}} \mathrm{LM}(\nu). \tag{3.7}$$

Note that $B^b(\nu)$ is a Gaussian process for each ν. The LM statistic has the chi-squared distribution for each fixed ν. However, the parameters are unknown and the covariances are data-dependent, which prevent the tabulation of the asymptotic distribution. Davies (1987) suggested calculating the upper bound of the distribution, but this method inevitably generates approximation errors, as noted by Caner and Hansen (2001).

The asymptotic distribution is similar to that of Hansen and Seo (2002), especially when the parameter ν_1 approaches 1 for the case of the logistic transition. In Hansen and Seo (2002), uniform convergence hinges on the known cointegrating vector because the threshold transition function is not continuous. However, uniform convergence follows directly because this paper assumes smooth transition.

As in Hansen and Seo (2002), this paper suggests the bootstrap inference as the asymptotic theory is nonstandard and the tabulation is not feasible. This paper considers the standard residual bootstrap algorithm. We assume the error u_t is independent. The residual bootstrap approximates the sampling distribution of the test statistic using the null model and the parameter estimates obtained under the null hypothesis.

The resampled residuals u_t^b are randomly drawn from the sample residuals, and then x_t^b can be constructed using the parameter estimates and the resampled residuals. The SupLM_n^b statistic can be calculated for each resampled data, and then we obtain the bootstrap p-value, which is the probability that the simulated statistic exceeds the sample SupLM statistic. If the p-value is less than the size chosen, then we reject the null hypothesis in favor of the alternative of nonlinear stochastic dependence.

Typically, the standard residual bootstrap assumes *i.i.d.* condition. However, the actual data in general show volatile movement and time-varying conditional variances. For a complete specification we should consider conditional heteroskedasticity, but it is difficult to specify the volatility structure each time we have a different dataset. Instead, we allow for conditional heteroskedasticity and make the tests robust to heteroskedasticity by using the White heteroskedasticity-consistent covariance estimator.

4. Simulation Evidence

We have shown that the tests for nonlinear adjustment have nonstandard distributions. Because the asymptotic distributions are data-dependent, this paper suggests the bootstrap inference. In this section, we examine the finite sample performance of the optimal tests using the Monte Carlo simulation study.

First, we design the experiments on the null distribution using a bivariate error correction model with one lagged variable ($l = 1$).

$$\Delta x_t = \mu + \begin{pmatrix} \alpha_1 \\ \alpha_2 \end{pmatrix} \begin{pmatrix} 1 \\ \beta \end{pmatrix}' x_{t-1} + \Gamma \Delta x_{t-1} + u_t, \tag{4.1}$$

where $x_t = (x_{1t}, x_{2t})'$ and $u_t = (u_{1t}, u_{2t})'$.

The alternative hypothesis allows for smooth transition, and hence the short-run coefficients vary smoothly depending on the transition variable $q_t = w_{t-1}(\tilde{\beta})$ and its weight $F(q_t; \nu)$. We consider the exponential and logistic transitions as defined in (2.2) and (2.3), respectively. In the experiment, our tests are based on

(4.1), allowing the coefficients on the intercept and the error correction to switch smoothly.

The experiments on size are based on a sample size of 250 and 1000 simulation replications, and for each replication 200 bootstrap replications are made to calculate the bootstrap p-values. The test statistics are calculated using $\lambda = \nu_1/(1-\nu_1)$, $\nu_{1L} = 1 - \nu_{1U} = 0.05$ and $\nu_{2L} = 1 - \nu_{2U} = 0.10$ and using 50 grid points on each $[\nu_{1L}, \nu_{1U}]$ and $[\nu_{2L}, \nu_{2U}]$.

For simplicity, we fix $\mu = 0$, $\beta = -1$ and $\alpha_1 = -1$. We vary α_2 among $(0, -0.5, 0.5)$ and Γ among

$$\Gamma_0 = \begin{pmatrix} 0 & 0 \\ 0 & 0 \end{pmatrix}, \quad \Gamma_1 = \begin{pmatrix} -0.2 & 0 \\ -0.1 & -0.2 \end{pmatrix} \text{ and } \Gamma_2 = \begin{pmatrix} -0.2 & -0.1 \\ -0.1 & -0.2 \end{pmatrix}.$$

The errors u_{1t} and u_{2t} are generated under homoskedastic and conditional heteroskedastic specifications. The homoskedastic case assumes that the errors are independently $N(0,1)$-distributed. The heteroskedastic case assumes that the errors u_{1t} and u_{2t} follow independent GARCH(1,1) processes, with $u_{it} \sim N(0, \sigma_{it}^2)$ and $\sigma_{it}^2 = 1 + 0.2u_{it-1}^2 + \phi\sigma_{it-1}^2$ for $i = 1, 2$.

Table 4.1 reports the rejection frequencies of the tests with exponential and logistic transitions at the nominal sizes 5%, 10%, 25% and 50%. The random sample is simulated from a linear error correction model, which is consistent with the null hypothesis. For each simulated data, the SupLM statistics and the bootstrap p-values are calculated. Table 4.1 shows the percentage of the simulated p-values which are smaller than the nominal size.

TABLE 4.1 *Size of* SupLM *tests*

Parameters			*Exponential*				*Logistic*			
α_2	Γ	ϕ	5%	10%	25%	50%	5%	10%	25%	50%
Homoskedastic										
0	Γ_0	0	0.046	0.092	0.249	0.517	0.059	0.103	0.231	0.479
−0.5	Γ_0	0	0.051	0.098	0.271	0.514	0.058	0.107	0.261	0.505
0.5	Γ_0	0	0.047	0.097	0.213	0.460	0.036	0.082	0.231	0.513
0	Γ_1	0	0.049	0.109	0.253	0.521	0.051	0.106	0.234	0.495
0	Γ_2	0	0.047	0.093	0.250	0.514	0.060	0.097	0.243	0.492
Heteroskedastic										
0	Γ_0	0.25	0.039	0.090	0.251	0.504	0.057	0.100	0.235	0.496
0	Γ_0	0.50	0.048	0.095	0.236	0.501	0.051	0.109	0.254	0.495
0	Γ_0	0.75	0.060	0.109	0.229	0.479	0.048	0.097	0.260	0.492

For the homoskedastic case, the errors u_{1t} and u_{2t} are generated from the independent $N(0,1)$ distribution. The rejection frequencies are calculated with different parameters of α_2 and Γ. The simulated null distribution appears to be close to the nominal size and similar across the various parameter specifications, as in Table 4.1. The results do not vary greatly between the exponential and logistic transition specifications.

For the heteroskedastic case, u_{1t} and u_{2t} are generated from the independent GARCH(1,1) processes. The other parameters are the same as the baseline specification. Our test statistics use the heteroskedasticity-robust covariance, and the simulated null distribution does not appear to be affected seriously by conditional heteroskedasticity. However, if the standard covariance estimator is used, the rejection rates tend to be affected seriously by heteroskedasticity as the magnitude of heteroskedasticity increases. Hence, we do not report the size performance of the tests with standard covariance estimator.

Next, we consider the experiment on the power of the tests for smooth transition nonlinear adjustment. For simplicity, we allow the parameters on intercept and error correction to switch smoothly. We generate the data from the following model:

$$\begin{aligned}\Delta x_t &= \mu_1 + \begin{pmatrix}\alpha_1\\0\end{pmatrix} w_{t-1}(\beta)\\ &\quad + \left[\mu_2 + \begin{pmatrix}-\delta\\0\end{pmatrix} w_{t-1}(\beta)\right] F\left(q_t(\beta);\lambda,\gamma\right) + \Gamma\Delta x_{t-1} + u_t,\end{aligned}$$

where $w_t(\beta) = x_{1t} + \beta x_{2t}$, and $F(q_t(\beta);\lambda,\gamma)$ is defined as (2.2) or (2.3) with $q_t(\beta) = w_{t-1}(\beta)$.

We fix $\mu_1 = \mu_2 = 0$, $\alpha_1 = -0.2$, $\Gamma = 0$ and $\beta = -1$. The transition parameter γ is set at zero for both exponential and logistic transitions. We vary the parameter $\lambda = \nu_1/(1-\nu_1)$ to take on several values. If $\delta = 0$, then the null hypothesis is maintained and there is no transition effect in the error correction process. However, if $\delta > 0$, then the alternative hypothesis holds and nonlinear smooth transition appears.

Table 4.2 shows the rejection frequency of the SupLM tests for smooth transition at the 5% size. The experiments on power are based on the sample sizes 250, 500 and 1000 replications. Other parameters are set at the same values as in the experiments on size, but we use 25 grid points on each $[\nu_{1L},\nu_{1U}]$ and $[\nu_{2L},\nu_{2U}]$ to reduce the computational costs.

TABLE 4.2 *Power of* SupLM *tests*

	$\nu_1 \backslash \delta$	*SupLM Test* 0.2	0.4	0.6	0.8	*LM(λ_0) Test* 0.2	0.4	0.6	0.8
Exponential	0.10	0.110	0.293	0.561	0.783	0.192	0.526	0.799	0.938
$n = 250$	0.25	0.099	0.285	0.609	0.869	0.165	0.486	0.835	0.969
	0.50	0.069	0.126	0.284	0.459	0.086	0.188	0.406	0.692
	0.75	0.047	0.077	0.103	0.124	0.048	0.070	0.103	0.177
	0.90	0.047	0.057	0.049	0.057	0.043	0.068	0.079	0.115
Logistic	0.10	0.049	0.070	0.082	0.097	0.052	0.070	0.091	0.095
$n = 250$	0.25	0.094	0.212	0.358	0.520	0.116	0.256	0.454	0.631
	0.50	0.203	0.619	0.900	0.984	0.243	0.708	0.944	0.989
	0.75	0.203	0.596	0.884	0.966	0.261	0.730	0.940	0.992
	0.90	0.197	0.554	0.844	0.953	0.229	0.676	0.908	0.969
Exponential	0.10	0.188	0.663	0.931	0.992	0.375	0.865	0.986	1.000
$n = 500$	0.25	0.145	0.609	0.943	0.999	0.285	0.792	0.986	1.000
	0.50	0.080	0.226	0.521	0.837	0.145	0.389	0.749	0.959
	0.75	0.051	0.065	0.118	0.229	0.056	0.094	0.185	0.344
	0.90	0.044	0.045	0.058	0.053	0.049	0.057	0.061	0.085
Logistic	0.10	0.051	0.077	0.107	0.158	0.070	0.109	0.144	0.195
$n = 500$	0.25	0.165	0.466	0.718	0.889	0.227	0.551	0.837	0.950
	0.50	0.458	0.943	0.997	1.000	0.528	0.979	1.000	1.000
	0.75	0.454	0.941	0.999	1.000	0.532	0.981	1.000	1.000
	0.90	0.427	0.917	0.995	1.000	0.495	0.966	0.999	0.999

As Table 4.2 shows, the rejection frequency of the tests increases as the shift parameter δ deviates from the null hypothesis. Table 4.2 also shows the standard LM test for nonlinearity, which assumes that the true values of transition parameters (ν_1, ν_2) are known. For example, the SupLM test with logistic transition rejects 62% of the null hypothesis at $\delta = 0.4$, $\nu_1 = 0.50$ and $n = 250$. The LM test, which is based on the true transition parameter values, rejects 71% of the null hypothesis.

As the parameter ν_1 approaches 0 or 1, the smooth transition model reduces to the linear model, and we cannot identify the transition effect unless we have a sufficiently large sample size. For the exponential transition, the slope of transition becomes steep as the transition rate ν_1 increases, which requires a large number of observations to identify the smooth transition effect. Thus, the SupLM and the LM tests for exponential transition do not provide significant power performance when the parameter ν_1 is large. On the other hand, the logistic transition function becomes flat as the transition rate ν_1 decreases. In this respect, the tests for logistic transition do not provide significant power when the parameter ν_1 is small. The smooth transition effect is likely to be identified as the sample size increases, and therefore the power function depends on the parameter values of transition and the sample size.

5. Concluding Remarks

This paper develops the tests for smooth transition nonlinear adjustment in the partially nonstationary vector autoregressive model. As the transition parameters cannot be identified under the null hypothesis, this paper explores the appropriate test statistic and its distribution theory.

One of the most important extensions of this paper would be the analysis of cointegration with smooth transition. The estimation of the smooth transition error correction model and the distribution theory of the estimators are also left to future research.

Appendix: Mathematical Proofs

Proof of Lemma 3.1

Since u_t is a square integrable Martingale difference sequence (MDS), the central limit theorem can be applied to show $U_{1n} = (1/\sqrt{n})\sum_{t=1}^{n} u_t \Rightarrow U_1$.

We need to show that $U_{2n}(\nu) = (1/\sqrt{n})\sum_{t=1}^{n} v_t(\nu) \Rightarrow U_2(\nu)$, where $v_t(\nu) = u_t F_t(\nu)$. Since $v_t(\nu)$ is a square integrable Martingale difference sequence (MDS) for each $\nu \in \mathcal{N}$, the central limit theorem can be applied. Thus, Assumption 3.1(ii) implies the finite dimensional distributional convergence.

Next, we show stochastic equicontinuity.

$$
\begin{aligned}
P\left(\sup_{|\nu-\nu'|\le\delta} |U_{2n}(\nu) - U_{2n}(\nu')| > \epsilon\right) &\le \frac{1}{\epsilon} E \sup_{|\nu-\nu'|\le\delta} |U_{2n}(\nu) - U_{2n}(\nu')| \\
&= \frac{1}{\epsilon} E \sup_{|\nu-\nu'|\le\delta} \left|\frac{1}{\sqrt{n}}\sum_{t=1}^{n} u_t \left(F_t(\nu) - F_t(\nu')\right)\right| \\
&= \frac{1}{\epsilon} E \sup_{|\nu-\nu'|\le\delta} \left|\frac{1}{\sqrt{n}}\sum_{t=1}^{n} u_t F_t'(\nu^*)(\nu - \nu')\right| \\
&\le \frac{\delta}{\epsilon} E \sup_{\nu\in\mathcal{N}} \frac{1}{\sqrt{n}}\sum_{t=1}^{n} |u_t||F_t'(\nu)|
\end{aligned}
$$

$$\leq \frac{\delta}{\epsilon} \sup_t || \sup_{\nu \in \mathcal{N}} |F_t^{'}(\nu)|||_2 \frac{1}{\sqrt{n}} \sum_{t=1}^{n} ||u_t||_2,$$

where $\nu^* \in [\nu, \nu^{'}]$.

Using Burkholder's inequality, we can show that $(1/\sqrt{n})\sum_{t=1}^{n} ||u_t||_2 \leq c_1 \sup_t ||u_t||_2$, where $c_1 = 36\sqrt{2}$. Therefore, $P(\sup_{|\nu-\nu^{'}|\leq\delta} |V_n(\nu) - V_n(\nu^{'})| > \epsilon) \to 0$ as $n \to \infty$ by picking δ sufficiently small.

Therefore, the pointwise central limit theorem and stochastic equicontinuity imply weak convergence $U_{2n}(\nu) \Rightarrow U_2(\nu)$.

Proof of Lemma 3.2

Since $(u_t \otimes z_t)$ is a square integrable MDS, we can show $W_{1n} = (1/\sqrt{n})\sum_{t=1}^{n}(u_t \otimes z_t) \Rightarrow W_1$.

Now, we need to show that $W_{2n}(\nu) = (1/\sqrt{n})\sum_{t=1}^{n}(v_t(\nu) \otimes z_t) \Rightarrow W_2(\nu)$. Since $(v_t(\nu) \otimes z_t)$ is a square integrable Martingale difference sequence (MDS) for each $\nu \in \mathcal{N}$, the central limit theorem can be applied. Assumption 3.1(ii) implies the finite dimensional distributional convergence.

We use the following to show stochastic equicontinuity.

$$P\left(\sup_{|\nu-\nu^{'}|\leq\delta} |W_{2n}(\nu) - W_{2n}(\nu^{'})| > \epsilon\right)$$
$$\leq \frac{1}{\epsilon} E \sup_{|\nu-\nu^{'}|\leq\delta} \left|\frac{1}{\sqrt{n}} \sum_{t=1}^{n} \left(u_t \left(F_t(\nu) - F_t(\nu^{'})\right) \otimes z_t\right)\right|$$
$$\leq \frac{\delta}{\epsilon} \sup_t || \sup_{\nu \in \mathcal{N}} |F_t^{'}(\nu)|||_2 \frac{1}{\sqrt{n}} \sum_{t=1}^{n} ||u_t||_4^2 \sup_t ||z_t||_4^2.$$

Using Burkholder's inequality, we can show that $(1/\sqrt{n})\sum_{t=1}^{n} ||u_t||_4 \leq c_2 \sup_t ||u_t||_4$, where $c_2 = 144/\sqrt{3}$. Therefore, $P(\sup_{|\nu-\nu^{'}|\leq\delta} |W_{2n}(\nu) - W_{2n}(\nu^{'})| > \epsilon) \to 0$ as $n \to \infty$ by picking δ sufficiently small.

Therefore, the pointwise central limit theorem and stochastic equicontinuity imply $W_{2n}(\nu) \Rightarrow W_2(\nu)$.

Proof of Lemma 3.3

We want to show $g_n(\nu) \Rightarrow W_2(\nu) - Q_{21}(\nu)Q_{11}^{-1}W_1$, where

$$g_n(\nu) = \text{vec}\left(\frac{1}{\sqrt{n}} \sum_{t=1}^{n} z_{2t}^{*}(\tilde{\beta}, \nu)\Delta x_t^{'}\right).$$

We note that under the null hypothesis $H_0 : D = 0$

$$\frac{1}{\sqrt{n}}\sum_{t=1}^{n} z_{2t}^{*}(\beta,\nu)\Delta x_t' = \frac{1}{\sqrt{n}}\sum_{t=1}^{n} z_{2t}^{*}(\beta,\nu)u_t'$$

$$= \frac{1}{\sqrt{n}}\sum_{t=1}^{n} z_{2t}(\beta,\nu)u_t' - \frac{1}{n}\sum_{t=1}^{n} z_{2t}(\beta,\nu)z_t'(\beta) \times \left(\frac{1}{n}\sum_{t=1}^{n} z_t(\beta)z_t'(\beta)\right)^{-1} \frac{1}{\sqrt{n}}\sum_{t=1}^{n} z_t(\beta)u_t'.$$

We show $(1/n)\sum_{t=1}^{n} z_{2t}(\nu)z_t' \xrightarrow{p} E(z_{2t}(\nu)z_t')$ uniformly in $\nu \in \mathcal{N}$.

To prove uniform convergence, we first show stochastic equicontinuity.

$$P\left(\sup_{|\nu-\nu'|\le\delta}\left|\frac{1}{n}\sum_{t=1}^{n}\left(z_t z_t' F_t(\nu) - z_t z_t' F_t(\nu')\right)\right| > \epsilon\right)$$

$$\le \frac{1}{\epsilon}E\sup_{|\nu-\nu'|\le\delta}\left|\frac{1}{n}\sum_{t=1}^{n} z_t z_t'\left(F_t(\nu) - F_t(\nu')\right)\right|$$

$$\le \frac{1}{\epsilon}E\sup_{|\nu-\nu'|\le\delta}\frac{1}{n}\sum_{t=1}^{n}|z_t z_t'||F_t'(\nu^*)||\nu-\nu'|$$

$$\le \frac{\delta}{\epsilon}\frac{1}{n}\sum_{t=1}^{n}||z_t z_t'||_2||\sup_{\nu\in\mathcal{N}}|F_t'(\nu)|||_2$$

$$\le \frac{\delta}{\epsilon}\sup_t||\sup_{\nu\in\mathcal{N}}|F_t'(\nu)|||_2\sup_t||z_t||_4^2,$$

where $\nu^* \in [\nu,\nu']$.

Assumptions 3.1(ii)–(iii)imply that $\sup_t||z_t||_4 < \infty$. We also note that $\sup_t E|z_t z_t' F_t(\nu)|_r \le \sup_t E|z_t z_t'|_r \le \sup_t||z_t||_{2r}^2 < \infty$ for all $\nu \in \mathcal{N}$ and for some $r > 1$. Therefore, pointwise convergence and stochastic equicontinuity imply that $(1/n)\sum_{t=1}^{n} z_{2t}(\nu)z_t' \xrightarrow{p} E(z_{2t}(\nu)z_t')$ uniformly in $\nu \in \mathcal{N}$.

We can also show $(1/n)\sum_{t=1}^{n} z_t z_t' \xrightarrow{p} E(z_t z_t')$ because $\sup_t E|z_t z_t'|_r \le \sup_t||z_t||_{2r}^2 < \infty$ for some $r > 1$.

Next, we use the asymptotic result $n(\tilde{\beta} - \beta_0) = O_p(1)$. The proof is given in Seo (1998).

Therefore,

$$g_n(\nu) = \text{vec}\left(\frac{1}{\sqrt{n}}\sum_{t=1}^{n} z_{2t}^{*}(\tilde{\beta},\nu)\Delta x_t'\right)$$

$$
\begin{aligned}
&= \operatorname{vec}\left(\frac{1}{\sqrt{n}}\sum_{t=1}^{n} z_{2t}^{*}(\nu)\Delta x_t'\right) + o_p(1) \\
&= \frac{1}{\sqrt{n}}\sum_{t=1}^{n}(v_t(\nu)\otimes z_t) - \frac{1}{n}\sum_{t=1}^{n}\left(I\otimes z_{2t}(\nu)z_t'\right) \\
&\quad \times \left(\frac{1}{n}\sum_{t=1}^{n}\left(I\otimes z_t z_t'\right)\right)^{-1}\frac{1}{\sqrt{n}}\sum_{t=1}^{n}(u_t\otimes z_t) + o_p(1) \\
&\Rightarrow W_2(\nu) - Q_{21}(\nu)Q_{11}^{-1}W_1.
\end{aligned}
$$

Proof of Theorem 3.1

Using Lemma 3.3, we can show that

$$
\begin{aligned}
\mathrm{LM}_n(\nu) &\Rightarrow \left(W_2(\nu) - Q_{21}(\nu)Q_{11}^{-1}W_1\right)' V^{-1}\left(W_2(\nu) - Q_{21}(\nu)Q_{11}^{-1}W_1\right) \\
&= B^b(\nu)'B^b(\nu),
\end{aligned}
$$

where $B^b(\nu) = V^{-1/2}(\nu)[W_2(\nu) - Q_{21}(\nu)Q_{11}^{-1}W_1]$.

The continuous mapping theorem implies that

$$
\sup_{\nu\in\mathcal{N}} \mathrm{LM}_n(\nu) \Rightarrow \sup_{\nu\in\mathcal{N}} \mathrm{LM}(\nu),
$$

where $\mathrm{LM}(\nu) = B^b(\nu)'B^b(\nu)$.

ACKNOWLEDGEMENTS

I would like to thank David Drukker, Thomas Fomby, Bruce Hansen, Sanggyu Park, Ingmar Prucha, Pentti Saikkonen and the seminar participants at Texas Camp Econometrics and FEMES for useful comments and suggestions. The author acknowledges the research support from Soongsil University.

REFERENCES

ANDREWS, D. W. K. (1993). "Tests for parameter instability and structural change with unknown change point", *Econometrica*, **61**, 821–856.

CANER, M. AND HANSEN, B. E. (2001). "Threshold autoregression with a unit root", *Econometrica*, **69**, 1555–1596.

CHAN, K. S. AND TONG, H. (1986). "On estimating thresholds in autoregressive models", *Journal of Time Series Analysis*, **7**, 179–190.

DAVIES, R. B. (1987). "Hypothesis testing when a nuisance parameter is present only under the alternative", *Biometrika*, **74**, 33–43.

ENGLE, R. F. AND GRANGER, C. W. J. (1987). "Co-integration and error correction: representation, estimation, and testing", *Econometrica*, **55**, 251–276.

HAGGAN, V. AND OZAKI, T. (1981). "Modelling nonlinear random vibrations using an amplitude-dependent autoregressive time series model", *Biometrika*, **68**, 189–196.

HANSEN, B. E. (1996). "Inference when a nuisance parameter is not identified under the null hypothesis", *Econometrica*, **64**, 413–430.

HANSEN, B. E. AND SEO, B. (2002). "Testing for two-regime threshold cointegration in vector error-correction models", *Journal of Econometrics*, **110**, 293–318.

MICHAEL, P., NOBAY, A. R. AND PEEL, D. A. (1997). "Transactions costs and nonlinear adjustment in real exchange rates: an empirical investigation", *Journal of Political Economy*, **105**, 862–879.

NEFTCI, S. N. (1984). "Are economic time series asymmetric over the business cycle?", *Journal of Political Economy*, **92**, 307–328.

SEO, B. (1998). "Tests for structural change in cointegrated systems", *Econometric Theory*, **14**, 222–259.

TERÄSVIRTA, T. (1994). "Specification, estimation, and evaluation of smooth transition autoregressive models", *Journal of the American Statistical Association*, **89**, 208–218.

TERÄSVIRTA, T. AND ANDERSON, H. M. (1992). "Characterizing nonlinearities in business cycles using smooth transition autoregressive models", *Journal of Applied Econometrics*, **7**, S119–S136.