Testing for Nonlinear Adjustment in Smooth Transition Vector Error Correction Models

Byeongseon Seo *

Texas A&M University †

and

Soongsil University

March 2004

*I would like to thank Thomas Fomby, Bruce Hansen, Ingmar Prucha, and Pentti Saikkonen for useful comments and suggestions. All errors are my responsibility.

†Department of Economics, College Station, TX 77843-4228, USA. E-mail: seo@econmail.tamu.edu. Phone: 979-845-7307. Fax: 979-847-8757.

Abstract

This paper considers testing for the presence of nonlinear adjustment in the smooth transition vector error correction model. The direct tests for smooth transition nonlinear adjustment, based on the exact specification of smooth transition and calculated under the linear error correction model, are proposed. This paper particularly focuses on the optimality issue in smooth transition models, which has not been explicitly explored. The transition parameters cannot be identified under the null hypothesis, and therefore this paper develops the optimal tests for smooth transition nonlinearity, the associated asymptotic theory, and the bootstrap inference. Simulation evidence shows that the bootstrap inference generates moderate size and power of the tests.

Key words: Nonlinear Adjustment; Optimal Tests; Smooth Transition

JEL classification: C12; C32

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), Terasvirta and Anderson (1992), and Michael, Nobay, and Peel (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 of linearity. Luukkonen, Saikkonen, and Terasvirta (1988) expanded the transition function and proposed the variable addition tests as the tests of linearity against smooth transition nonlinearity, and the tests have been used in many empirical studies. However, the test statistics are based on the polynomial approximation, and the approximation errors may affect statistical inference depending on the parameter values of transition rate and location. Furthermore, the tests are not directly related to the smooth transition model, and thus we cannot retrace what causes the rejection of linearity. This paper considers the direct tests for nonlinear adjustment, which are based on the exact specification of smooth transition.

The smooth transition model entails transition parameters, which cannot be identified under the null hypothesis. However, the optimality issue in the smooth transition model has not been treated extensively. The optimality issue regarding unidentified parameters has been developed by Davies (1987), Andrews (1993), and Hansen (1996). Hansen (1996) particularly considered the optimality issue in threshold models. The threshold parameter cannot be identified under the null hypothesis, and as a result the likelihood ratio statistic has the nonstandard distribution. The smooth transition model generalizes the threshold model, and thus this paper develops the appropriate tests and the associated distribution theory based on the optimality argument.

Many empirical studies have found evidence on the presence of stochastic nonlinear dependence in equilibrium relations such as purchasing power parity. For example, Michael, Nobay, and Peel (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. However, the econometric methods and the formal theory have been limited. This paper proposes the tests for nonlinear adjustment in the smooth transition vector error correction models, and thereby fills the deficiency in the literature.

One technical difficulty is to estimate the smooth transition model. As noted by Haggan and Ozaki (1981) and Terasvirta (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. Therefore, our tests are easy to implement and thus useful.

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; thus, 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.

There are other related papers by Caner and Hansen (2001), Kapetanios, Shin, and Snell (2003), Saikkonen and Choi (2004), and Hansen and Seo (2002). Caner and Hansen (2001) considered the unit root tests in the TAR model. Kapetanios, Shin, and Snell (2003) considered the unit root tests in the STAR model. Saikkonen and Choi (2004) considered cointegration tests in the STAR model. This paper does not consider the tests of unit root or cointegration because it is difficult to deal directly with the nonstationary variable in the transition function. Hansen and Seo (2002) considered the tests for threshold nonlinearity in vector error correction model. This paper extends Hansen and Seo (2002) to the smooth transition vector error correction model.

We denote $\Rightarrow$ as weak convergence with respect to the uniform metric and $\rightarrow^p$ 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 vectorization operator.

This paper is organized as follows. Section 2 introduces the smooth transition vector error correction model and develops the optimal tests for nonlinear stochastic dependence. Section 3 explores the asymptotic distribution of the proposed tests. Section 4 provides the simulation evidence on the size and power of the tests. An economic application on the S&P index future arbitrage is illustrated in Section 5.

2 Model

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

$$\Delta x_t = A^{'} z_t(\beta) + D^{'} z_t(\beta) F(q_t; \lambda) + u_t, \tag{1}$$

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

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 $i = 1, 2, \ldots$ 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 $F(q_t; \lambda)$ depends on the transition variable q_t and the associated parameter vector λ. The functional form can be specified in several ways, depending on the characteristics of nonlinear adjustment. As in Luukkonen and Terasvirta (1991), this paper considers the exponential and logistic transition functions as follows:

$$F(q_t; \lambda_1, \lambda_2) = 1 - \exp[-\lambda_1 (q_t - \lambda_2)^2], \ \ \lambda_1 > 0, \tag{2}$$

and

$$F(q_t; \lambda_1, \lambda_2) = \frac{1}{1 + \exp[-\lambda_1 (q_t - \lambda_2)]}, \ \ \lambda_1 > 0. \tag{3}$$

The exponential specification (2) allows for a smooth transition based on the inverted normal density function, while the logistic specification (3) models a smooth transition based on the cumulative logistic distribution. The exponential specification (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_1 determines the speed of transition. As λ_1 increases, the transition from the mid regime to both tail regimes, and its reverse, can be made quickly. If $\lambda_1 = 0$, there is no transition and only the mid regime is prevalent. If λ_1 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 $0 < \lambda_1 < \infty$. The location parameter λ_2 determines the average location of transition. We assume that λ_2 lies inside the support of the transition variable. That is, $\min(q_t) < \lambda_2 < \max(q_t)$.

The logistic specification (3) models 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_1 determines the speed of transition, and the location parameter λ_2 determines the average location of transition. As λ_1 increases, the transition from the first regime to the second regime, and its reverse, can be made quickly. If $\lambda_1 = 0$, there is no transition and only the first regime remains. This case leads to the linear error correction model, and thus we assume that $\lambda_1 > 0$. As λ_1 approaches ∞, the logistic transition converges to the threshold transition such that $lim_{\lambda_1 \to \infty} F(q_t; \lambda_1, \lambda_2) = 1(q_t \geq \lambda_2)$, 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 λ_2 lies inside the support of the transition variable. That is, $\min(q_t) < \lambda_2 < \max(q_t)$.

The transition variable q_t is a stationary transformation of the predetermined variables; for example, $q_t = w_{t-1}$ and $q_t = |w_{t-1}|$. To simplify analysis, we focus on $q_t = w_{t-1}$. Main results do not change when other predetermined variables are used as the transition variable if the variable is stationary.

The smooth transition error correction model has continuously varying coefficients depending on the current state w_{t-1}. The nonlinear dynamics can be explained by the coefficient matrices A and D. Our model (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.

Our model can be reduced to linear error correction model (4) when the coefficient matrix D is zero.

$$\Delta x_t = A^{'} z_t(\beta) + u_t = \mu + \alpha w_{t-1}(\beta) + \sum_{i=1}^{l} \Gamma_i \Delta x_{t-i} + u_t. \tag{4}$$

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

$$\mathcal{H}_0 : D = 0 \text{ against } \mathcal{H}_1 : D \neq 0.$$

We define the parameter vector

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

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

$$\mathcal{L}_n(\theta, \lambda) = -\frac{1}{2} \sum_{t=1}^{n} [log|\Sigma| + u_t^{'}(\lambda, \theta)\Sigma^{-1} u_t(\lambda, \theta)], \tag{5}$$

where $u_t(\lambda, \theta)$ is defined in (1).

We denote $\hat{\theta}(\lambda)$ as the maximum likelihood estimator (MLE) of θ for known λ. As noted by Haggan and Ozaki (1981) and Terasvirta (1994), technical difficulty arises when the transition parameters λ 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 λ as fixed until we define optimal tests for unknown λ.

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 (1) is linear in parameters A and D for fixed λ. We denote $\hat{A}(\beta,\lambda)$ and $\hat{D}(\beta,\lambda)$ as the MLE for given β and λ. Thus, the MLE $\hat{D}(\beta,\lambda)$ is given by

$$\hat{D}(\beta,\lambda) = (\sum_{t=1}^{n} z_{2t}^{*}(\beta,\lambda) z_{2t}^{*\prime}(\beta,\lambda))^{-1} \sum_{t=1}^{n} z_{2t}^{*}(\beta,\lambda)\Delta x_t^{\prime}, \tag{6}$$

where

$$\begin{aligned} z_{2t}^{*}(\beta,\lambda) &= z_{2t}(\beta,\lambda) - \sum_{t=1}^{n} z_{2t}(\beta,\lambda) z_t^{\prime}(\beta)(\sum_{t=1}^{n} z_t(\beta) z_t^{\prime}(\beta))^{-1} z_t(\beta), \text{ and} \\ z_{2t}(\beta,\lambda) &= z_t(\beta) F(w_{t-1}(\beta);\lambda). \end{aligned}$$

For the null hypothesis $\mathcal{H}_0 : D = 0$, we define the score function $g_n(\lambda)$ as follows:

$$g_n(\lambda) = \frac{1}{\sqrt{n}} \text{vec}(\sum_{t=1}^{n} z_{2t}^{*}(\tilde{\beta},\lambda)\Delta x_t^{\prime}) \tag{7}$$

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

We define $\tilde{u}_t = \Delta x_t - \tilde{A}^{\prime} z_t(\tilde{\beta})$ and $\tilde{v}_t(\lambda) = \tilde{u}_t F(w_{t-1}(\tilde{\beta});\lambda)$. To allow for time-varying conditional variances, we define the heteroskedasticity-robust covariance estimator of the score function as follows:

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

where

$$\Omega_{11n} = \frac{1}{n} \sum_{t=1}^{n} (\tilde{u}_t \tilde{u}_t^{\prime} \otimes z_t(\tilde{\beta}) z_t^{\prime}(\tilde{\beta}))$$

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

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

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

$$
LM_n(\lambda) = g_n(\lambda)'V_n^{-1}(\lambda)g_n(\lambda). \tag{8}
$$

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

Compared to the threshold transition, the exponential and logistic transition models entail the transition rate as well as the location parameters. Because the support of the transition rate parameter λ_1 is unbounded, we assume a monotonic transformation $h(\cdot)$, which leads to $\lambda_1 = h^{-1}(\nu_1)$ for $\nu_1 \subset (0,1)$.

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 $\nu_1 \in [\nu_{1L}, \nu_{1U}] \subset (0,1)$ or $\lambda_1 \in [\lambda_{1L}, \lambda_{1U}] \subset R^+$.

Also, the smooth transition has meaning only if $0 < P(w_t \leq \lambda_2) < 1$. Using the monotonic transformation $\nu_2 = P(w_t \leq \lambda_2)$, we impose this constraint by assuming that $\nu_2 \in [\nu_{2L}, \nu_{2U}] \subset (0,1)$ or $\lambda_2 \in [\lambda_{2L}, \lambda_{2U}]$, where $\nu_{2L} = P(w_t \leq \lambda_{2L})$ and $\nu_{2U} = P(w_t \leq \lambda_{2U})$.

The simulation and empirical results are based on $\nu_{1L} = 1 - \nu_{1U} = 0.05$ and $\nu_{2L} = 1 - \nu_{2U} = 0.10$.

The LM statistic has been defined for fixed λ. This is appropriate only when λ is known. If λ 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. This paper extends the optimality treatment of Hansen (1996) to the tests for nonlinear adjustment in smooth transition error correction models.

If we assume that λ lies in $\Lambda = [\lambda_{1L}, \lambda_{1U}] \times [\lambda_{2L}, \lambda_{2U}] \subset R^2$, then the optimal test statistic can be defined as follows:

$$\text{SupLM} = \text{Sup}_{\lambda \in \Lambda} \text{LM}_n(\lambda). \tag{9}$$

3 Main Results

First, we use the representation theorem by Engle and Granger (1987). The linear error correction model (4) has the following representation:

$$\Delta x_t = C(L) u_t \tag{10}$$

$$x_t = C(1) \sum_{i=1}^{t} u_i + C^*(L) u_t \tag{11}$$

$$w_t = B^{'} C^*(L) u_t, \tag{12}$$

where $C^*(L) = \frac{C(L)-C(1)}{1-L}$ 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 $\nu_1 = h(\lambda_1)$ and $\nu_2 = P(w_t \leq \lambda_2)$. We denote $u_t(\nu) = u_t(\theta_0, \nu)$, $F_t(\nu) = F(w_{t-1}(\beta_0); \nu)$, 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 1

1. $\nu \in \mathcal{N} \subset (0,1)^2$.

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

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

4. $sup_{\theta\in\Theta} \ |\theta| < \infty$.

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

We use Assumption 1.5 to show the stochastic equicontinuity of the sum $\frac{1}{\sqrt{n}}\sum_{t=1}^{n} v_t(\nu)$, where $v_t(\nu) = u_t F_t(\nu)$. The exponential and logistic transition functions satisfy this condition if Assumptions 1.1-1.4 hold.

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

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

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

Lemma 1 *Under Assumption 1,*

$$\begin{pmatrix} \frac{1}{\sqrt{n}}\sum_{t=1}^{n} u_t \\ \frac{1}{\sqrt{n}}\sum_{t=1}^{n} v_t(\nu) \end{pmatrix} \Rightarrow \begin{pmatrix} U_1 \\ U_2(\nu) \end{pmatrix} \sim N(0, \begin{pmatrix} \Sigma_{11} & \Sigma_{12}(\nu) \\ \Sigma_{21}(\nu) & \Sigma_{22}(\nu) \end{pmatrix}), \quad (13)$$

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 2 *Under the null hypothesis and Assumption 1,*

$$\begin{pmatrix} \frac{1}{\sqrt{n}}\sum_{t=1}^{n}(u_t \otimes z_t) \\ \frac{1}{\sqrt{n}}\sum_{t=1}^{n}(v_t(\nu) \otimes z_t) \end{pmatrix} \Rightarrow \begin{pmatrix} W_1 \\ W_2(\nu) \end{pmatrix} \sim N(0, \begin{pmatrix} \Omega_{11} & \Omega_{12}(\nu) \\ \Omega_{21}(\nu) & \Omega_{22}(\nu) \end{pmatrix}), \quad (14)$$

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 *Under the null hypothesis and Assumption 1,*

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

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 1 *Under the null hypothesis and Assumption 1,*

$$LM_n(\lambda) \Rightarrow B^b(\nu)' B^b(\nu) \equiv LM(\nu), \tag{15}$$

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

Therefore,

$$Sup_{\lambda\in\Lambda}\ LM_n(\lambda) \ \Rightarrow\ Sup_{\nu\in\mathcal{N}}\ LM(\nu). \tag{16}$$

Note that $B^b(\nu)$ is a Gaussian process for each ν. The LM statistic has the chi-squared distribution for each known ν. However, the process depends on unknown parameters 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 (2002).

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. There are many bootstrap algorithms, and it is hard to tell which algorithm performs suitably in our model in terms of consistency and refinement. 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 $SupLM^b$ 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 optimal 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 ($l = 1$).

$$\Delta x_t = \mu + \begin{pmatrix} \alpha_1 \\ \alpha_2 \end{pmatrix} \begin{pmatrix} 1 \\ \beta \end{pmatrix}' x_{t-1} + \Gamma \Delta x_{t-1} + u_t, \tag{17}$$

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 w_{t-1} and its weight $F(w_{t-1}; \lambda)$. We consider the exponential and logistic transitions as defined in (2) and (3), respectively.

In the experiment, our tests are based on (17), 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 optimal test statistics are calculated using $\lambda_1 = \frac{\nu_1}{1-\nu_1}$, $\nu_{1L} = 1 - \nu_{1U} = 0.05$ and $\nu_{2L} = \nu_{2U} = 0.10$, and using 50 grid points on each $[\lambda_{1L}, \lambda_{1U}]$ and $[\lambda_{2L}, \lambda_{2U}]$.

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 1 reports the rejection frequencies of the optimal 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 1 shows the percentage of the simulated p-values which are smaller than the nominal size.

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 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.

Table 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

Next, we consider the experiment on the power of the optimal 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:

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

where $w_t(\beta) = x_{1t} + \beta x_{2t}$, and $F(w_{t-1}(\beta);\lambda)$ is defined as (2) or (3).

We fix $\mu_1 = \mu_2 = 0$, $\alpha_1 = -0.2$, $\Gamma = 0$, and $\beta = -1$. The transition parameter λ_2 is set at zero for both exponential and logistic transitions. We vary the parameter $\lambda_1 = \frac{\nu_1}{1-\nu_1}$ 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 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 and 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 $[\lambda_{1L}, \lambda_{1U}]$ and $[\lambda_{2L}, \lambda_{2U}]$ to reduce the computational costs.

Table 2. Power of SupLM Tests

		SupLM Test				$LM(\lambda_0)$ Test			
	$\nu_1 \backslash \delta$	0.2	0.4	0.6	0.8	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 2 shows, the rejection frequency of the tests increases as the shift parameter δ deviates from the null hypothesis. Table 2 also shows the standard LM test for nonlinearity, which assumes that the 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 parameters, 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 a sufficient sample size.

5 Application: Index Futures Arbitrage

The optimal tests are applied to the index futures arbitrage. The arbitrage relationship between stock-index futures price and spot price can be formulated by the cost of carry model.

$$H_t = S_t \exp((r_t - q_t)(T - t)),$$

where H_t and S_t are the theoretical futures price and the stock-index spot price, respectively. Also, r_t is the risk-free interest rate, q_t is the dividend yield on the stock index, and $T - t$ is the time to maturity of the futures contract.

We allow for pricing error w_t, which is the deviation of the actual futures price in logarithm f_t from the theoretical price in logarithm h_t as follows:

$$f_t = h_t + w_t.$$

If the pricing error is stationary, then the arbitrage relation forms a cointegrating relationship. The empirical work uses the intraday S&P 500 index and futures market data for the month of May 1993.[1] The sample size used in the application is 7060.

Let $x_t = (f_t, h_t)$. All variables are written in logarithms and multiplied by 100. The ADF unit root test shows that the futures prices and theoretical prices are integrated of order 1. Johansen's cointegration test shows that f_t and h_t are cointegrated. That is, the pricing error w_t is stationary. The VAR lag length picked by BIC is 5; that is, $l = 5$. These results are based on the linear error correction model as follows:

$$\Delta x_t = \mu + \alpha w_{t-1}(\beta) + \sum_{i=1}^{l} \Gamma_i \Delta x_{t-i} + u_t,$$

where $w_{t-1}(\beta) = f_{t-1} + \beta h_{t-1}$.

In the financial market, there exist transaction costs such as brokerage fees, bid-ask spread, price impact, and regulations, which affect the volume and frequency of trading. The transaction costs prevent the arbitrage opportunity from being realized as long as the arbitrage does not produce a net gain. There are many indirect costs such as index tracking error and execution risk, which are not easily measurable. In particular, the financial market is composed of heterogenous agents, and the actual transaction costs are different between investors.

Thus, we consider smooth transition nonlinear adjustment as a possibility of better empirical description.

$$\Delta x_t = A' z_t + D' z_t^* F(w_{t-1}(\beta); \lambda) + u_t,$$

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

[1] The dataset was provided by Forbes, et al. (1999), and it can be extracted from the data archive: www.econ.queensu.ca/jae/1998-v13.3.

We define three models depending on the variables included in smooth transition z_t^*. Model 1 includes the error correction term only, and Model 2 includes the error correction term and the constant. Model 3 includes all short-run variables. In the application, we consider two transition specifications: exponential and logistic transitions.

$$\begin{aligned} F^E(w_{t-1}(\beta); \lambda) &= 1 - \exp[-\lambda_1(w_{t-1} - \lambda_2)^2] \\ F^L(w_{t-1}(\beta); \lambda) &= \frac{1}{1 + \exp[-\lambda_1(w_{t-1} - \lambda_2)]} \end{aligned}$$

Table 3 shows the SupLM statistics, the associated p-values, and the 5% critical values. We consider three specifications: the coefficients on error correction only, the error correction and constant, and all short-run variables to switch. In each case, we find a strong evidence of smooth transition nonlinear adjustment toward the arbitrage relation. For example, the SupLM statistics for exponential transition have p-values less than 0.001, and thus the null hypothesis of linearity can be rejected in favor of nonlinear adjustment at the 5% size. The p-values are calculated from 1,000 bootstrap replications of the SupLM statistic under the null hypothesis.

Table 3. Tests for Nonlinear Adjustment

	Model	SupLM_n	P-value	5% C-value
Exponential	$1: z_t^* = w_{t-1}$	18.630	0.001	6.500
	2: $z_t^* = (w_{t-1}, 1)$	44.455	0.000	10.082
	3: $z_t^* = z_t$	64.918	0.000	39.781
Logistic	1: $z_t^* = w_{t-1}$	52.369	0.000	8.684
	$2: z_t^* = (w_{t-1}, 1)$	52.881	0.000	10.624
	$3: z_t^* = z_t$	64.618	0.000	38.454

Table 4 provides the estimates of the smooth transition error correction model. As noted by Haggan and Ozaki (1981) and Terasvirta (1994), it is difficult to estimate the transition parameters jointly with the other slope parameters. The joint MLE tends to produce the gradients which force the transition rate estimate to deviate from the true value to infinity. Haggan and Ozaki (1981) have suggested the conditional least squares method with a grid on the transition parameter. Because our model contains the cointegrating vector, we propose using an algorithm of conditional maximum likelihood with a grid search of transition parameters as follows:

$$Min_{\lambda \in \Lambda} Min_{\theta \in \Theta} - \mathcal{L}_n(\theta, \lambda),$$

where $\mathcal{L}_n(\theta, \lambda)$ is defined in (5).

First, we set a grid on $\lambda \in \Lambda$, where λ corresponds to $\nu \in \mathcal{N} \subset (0,1)^2$. Second, for each fixed λ, we estimate the model parameters and the likelihood. Then, we find the MLE $(\hat{\lambda}, \hat{\theta})$ that maximizes the likelihood function.

Table 4 shows the estimates of the smooth transition models. We allow the coefficients on the constant and the error correction term to switch between regimes. The cointegrating coefficient is set at -1. The estimation results with an unknown cointegrating vector, which are not reported here, are similar to Table 4.

The exponential transition specification shows that the actual future prices do not respond to the pricing error in the mid regime. The future prices respond to the pricing error negatively in tail regimes, but its response is not significant. On the other hand, the theoretical prices respond to the pricing error significantly in the mid regime, and the response gets stronger in the direction of both tail regimes. Thus, the transition effect is significant.

The logistic specification also reveals regime-dependent smooth transition. The theoretical futures prices respond significantly to the pricing error, and this response gets stronger in the regime of backwardation.

Figure 1 shows the short-run dynamics of actual futures price, and Figure 2 shows nonlinear dynamics of theoretical futures price. In the mid regime, the arbitrage opportunity may not be realized because the opportunity is dominated by the transaction costs, and thus the pricing error is persistent. However, in the tail regimes, the arbitrage is profitable and this stimulates mean-reverting behavior. However, the linear model cannot explain nonlinear stochastic dependence, and its likelihood is lower than that of the exponential and logistic smooth transition models.

Table 4. Estimation of Smooth Transition Error Correction Model

		Equation f_t		Equation h_t	
		coefficient	s.e.	coefficient	s.e.
Exponential	w_{t-1}	0.00046892	0.01024871	0.02694313	0.00625216
	1	0.00012835	0.00049261	0.00062831	0.00028422
	$w_{t-1}F_t$	-0.01370542	0.06637785	0.32411154	0.04050376
	F_t	0.00433195	0.00829190	-0.03503646	0.00554685
		$\hat{\lambda}_1 = 10.579$		$\hat{\lambda}_2 = -0.025$	
		Log-likelihood = 53154.483			
Logistic	w_{t-1}	-0.04654683	0.14360252	0.79515054	0.09090497
	1	-0.01014776	0.03780484	0.18793993	0.02350037
	$w_{t-1}F_t$	0.02191674	0.04256329	-0.17173913	0.02887034
	F_t	0.01962562	0.07187855	-0.35673112	0.04465265
		$\hat{\lambda}_1 = 7.800$		$\hat{\lambda}_2 = -0.013$	
		Log-likelihood = 53154.204			
Linear	w_{t-1}	0.00143875	0.00650160	0.05266889	0.00471494
	1	0.00032464	0.00037670	-0.00038843	0.00021882
		Log-likelihood = 52942.736			

6 Concluding Remarks

This paper develops the optimal tests for nonlinear adjustment in smooth transition error correction models. Our tests do not use polynomial approximation and there is no missing link between the model and the tests. This paper particularly focuses on the optimality issue in the smooth transition model; therefore, this paper is necessary and required.

One of the most important extensions of this paper would be the analysis of cointegration with smooth transition. We expect to develop a formal test and the associated distribution theory, which does not depend on approximation. 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 1: Since u_t is a square integrable Martingale difference sequence (MDS), the central limit theorem can be applied to show $U_{1n} = \frac{1}{\sqrt{n}} \sum_{t=1}^{n} u_t \Rightarrow U_1$.

We need to show that $U_{2n}(\nu) = \frac{1}{\sqrt{n}} \sum_{t=1}^{n} v_t(\nu) \Rightarrow U_2(\nu)$, 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 1.2 implies the finite dimensional distributional convergence.

Next, we show stochastic equicontinuity.

$$
\begin{aligned}
P(\sup_{|\nu-\nu'|\le\delta} |U_{2n}(\nu) - U_{2n}(\nu')| > \epsilon) &\le \frac{1}{\epsilon} E \sup_{|\nu-\nu'|\le\delta} |U_{2n}(\nu) - U_{2n}(\nu')| \\
&= \frac{1}{\epsilon} E \sup_{|\nu-\nu'|\le\delta} |\frac{1}{\sqrt{n}} \sum_{t=1}^{n} u_t (F_t(\nu) - F_t(\nu'))| \\
&= \frac{1}{\epsilon} E \sup_{|\nu-\nu'|\le\delta} |\frac{1}{\sqrt{n}} \sum_{t=1}^{n} u_t F_t'(\nu^*)(\nu - \nu')| \\
&\le \frac{\delta}{\epsilon} E \sup_{\nu\in\mathcal{N}} \frac{1}{\sqrt{n}} \sum_{t=1}^{n} |u_t||F_t'(\nu)| \\
&\le \frac{\delta}{\epsilon} \sup_t || \sup_{\nu\in\mathcal{N}} |F_t'(\nu)| ||_2 \frac{1}{\sqrt{n}} \sum_{t=1}^{n} ||u_t||_2,
\end{aligned}
$$

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

Using Burkholder's inequality, we can show that $\frac{1}{\sqrt{n}} \sum_{t=1}^{n} ||u_t||_2 \le c_1 \sup_t ||u_t||_2$, where $c_1 = 36\sqrt{2}$. Therefore, $P(\sup_{|\nu-\nu'|\le\delta} |V_n(\nu) - V_n(\nu')| > \epsilon) \to 0$ 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 2: Since $(u_t \otimes z_t)$ is a square integrable MDS, we can show $W_{1n} = \frac{1}{\sqrt{n}} \sum_{t=1}^{n} (u_t \otimes z_t) \Rightarrow W_1$.

Now, we need to show that $W_{2n}(\nu) = \frac{1}{\sqrt{n}} \sum_{t=1}^{n} (v_t(\nu) \otimes z_t) \Rightarrow W_2(\nu)$. 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 1.2 implies the finite dimensional distributional convergence.

We use the following to show stochastic equicontinuity.

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

Using Burkholder's inequality, we can show that $\frac{1}{\sqrt{n}} \sum_{t=1}^{n} ||u_t||_4 \le c_2 \sup_t ||u_t||_4$, where $c_2 = 144/\sqrt{3}$. Therefore, $P(\sup_{|\nu-\nu'|\le\delta} |W_{2n}(\nu) - W_{2n}(\nu')| > \epsilon) \to 0$ 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: We want to show $g_n(\lambda) \Rightarrow W_2(\nu) - Q_{21}(\nu)Q_{11}^{-1}W_1$, where

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

We note that under the null hypothesis $\mathcal{H}_0 : D = 0$

$$\begin{aligned}\frac{1}{\sqrt{n}}\sum_{t=1}^{n} z_{2t}^{*}(\beta,\lambda)\Delta x_t^{'} &= \frac{1}{\sqrt{n}}\sum_{t=1}^{n} z_{2t}^{*}(\beta,\lambda)u_t^{'} \\ &= \frac{1}{\sqrt{n}}\sum_{t=1}^{n} z_{2t}(\beta,\lambda)u_t^{'} - \frac{1}{n}\sum_{t=1}^{n} z_{2t}(\beta,\lambda)z_t^{'}(\beta)(\frac{1}{n}\sum_{t=1}^{n} z_t(\beta)z_t^{'}(\beta))^{-1}\frac{1}{\sqrt{n}}\sum_{t=1}^{n} z_t(\beta)u_t^{'}.\end{aligned}$$

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

To prove uniform convergence, we first show stochastic equicontinuity.

$$\begin{aligned}P(\sup_{|\nu-\nu^{'}|\le\delta}|\frac{1}{n}\sum_{t=1}^{n}(z_t z_t^{'}F_t(\nu) - z_t z_t^{'}F_t(\nu^{'}))| > \epsilon) &\le \frac{1}{\epsilon}E\sup_{|\nu-\nu^{'}|\le\delta}|\frac{1}{n}\sum_{t=1}^{n} z_t z_t^{'}(F_t(\nu) - F_t(\nu^{'}))| \\ &\le \frac{1}{\epsilon}E\sup_{|\nu-\nu^{'}|\le\delta}\frac{1}{n}\sum_{t=1}^{n}|z_t z_t^{'}||F_t^{'}(\nu^{*})||\nu-\nu^{'}| \\ &\le \frac{\delta}{\epsilon}\frac{1}{n}\sum_{t=1}^{n}||z_t z_t^{'}||_2||\sup_{\nu\in\mathcal{N}}|F_t^{'}(\nu)|||_2 \\ &\le \frac{\delta}{\epsilon}\sup_t||\sup_{\nu\in\mathcal{N}}|F_t^{'}(\nu)|||_2\sup_t||z_t||_4^2,\end{aligned}$$

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

Assumptions 1.2-1.3 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 $\frac{1}{n}\sum_{t=1}^{n} z_{2t}(\nu)z_t^{'} \to^{p} E(z_{2t}(\nu)z_t^{'})$ uniformly in $\nu \in \mathcal{N}$.

We can also show $\frac{1}{n}\sum_{t=1}^{n} z_t z_t^{'} \to^{p} E(z_t z_t^{'})$ 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 Johansen (1988) and Seo (1998).

Therefore,

$$\begin{aligned}g_n(\lambda) &= \text{vec}(\frac{1}{\sqrt{n}}\sum_{t=1}^{n} z_{2t}^{*}(\tilde{\beta},\lambda)\Delta x_t^{'}) \\ &= \text{vec}(\frac{1}{\sqrt{n}}\sum_{t=1}^{n} z_{2t}^{*}(\lambda)\Delta x_t^{'}) + o_p(1) \\ &= \frac{1}{\sqrt{n}}\sum_{t=1}^{n}(v_t(\nu)\otimes z_t) - \frac{1}{n}\sum_{t=1}^{n}(I\otimes z_{2t}(\nu)z_t^{'})(\frac{1}{n}\sum_{t=1}^{n}(I\otimes z_t z_t^{'}))^{-1}\frac{1}{\sqrt{n}}\sum_{t=1}^{n}(u_t\otimes z_t) + o_p(1) \\ &\Rightarrow W_2(\nu) - Q_{21}(\nu)Q_{11}^{-1}W_1.\end{aligned}$$

Proof of Theorem 1: Using Lemma 3, we can show that

$$\begin{aligned} LM_n(\lambda) &\Rightarrow (W_2(\nu) - Q_{21}(\nu)Q_{11}^{-1}W_1)^{'}V^{-1}(W_2(\nu) - Q_{21}(\nu)Q_{11}^{-1}W_1) \\ &= B^b(\nu)^{'}B^b(\nu), \end{aligned}$$

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

$$Sup_{\lambda\in\Lambda}\ LM_n(\lambda) \ \Rightarrow\ Sup_{\nu\in\mathcal{N}}\ LM(\nu),$$

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

References

[1] Anderson, H., 1997. Transaction Costs and Non-linear Adjustment towards Equilibrium in the US Treasury Bill Market. Oxford Bulletin of Economics and Statistics 59, 465-484.

[2] Andrews, D.W.K., 1993. Tests for parameter instability and structural change with unknown change point. Econometrica 61, 821-856.

[3] Andrews, D.W.K., Ploberger, W., 1994. Optimal tests when a nuisance parameter is present only under the alternative. Econometrica 62, 1383-1414.

[4] Balke, N. S., Fomby, T. B., 1997. Threshold cointegration. International Economic Review 38, 627-645.

[5] Caner, M., Hansen, B. E., 2001. Threshold autoregression with a unit root. Econometrica 69, 1555-1596.

[6] Chan, K.S., 1993. Consistency and limiting distribution of the least squares estimator of a threshold autoregressive model. Annals of Statistics 21, 520-533.

[7] Chan, K.S., Tong, H., 1986. On estimating thresholds in autoregressive models. Journal of Time Series Analysis 7, 179-190.

[8] Chang, Y., Park, J., 2003. Index Models with Integrated Time Series. Journal of Econometrics 114, 73-106.

[9] Dahl, C., Gonzalez-Rivera, G., 2003. Testing for Neglected Nonlinearity in Regression Models Based on the Theory of Random Fields. Journal of Econometrics 114, 141-164.

[10] Davies, R.B., 1987. Hypothesis testing when a nuisance parameter is present only under the alternative. Biometrika 74, 33-43.

[11] Enders, W., Granger, C., 1998. Unit-root tests and asymmetric adjustment with an example using the term structure of interest rates. Journal of Business and Economic Statistics 16, 304-311.

[12] Enders, W., Siklos, P., 2001. Cointegration and Threshold Adjustment. Journal of Business and Economic Statistics 19, 166-176.

[13] Engle, R.F., Granger, C.W.J., 1987. Cointegration and error correction representation, estimation, and testing. Econometrica 55, 251-276.

[14] Forbes, C., Kalb, G., Kofman, P., 1999. Bayesian arbitrage threshold analysis. Journal of Applied Econometrics 17, 364-372.

[15] Gine, E., Zinn, J., 1990. Bootstrapping general empirical measures. Annals of Probability 18, 851-869.

[16] Haggan, V., Ozaki, T., 1981. Modelling Nonlinear Random Vibrations Using an Amplitude-Dependent Autoregressive Time Series Model. Biometrika 68, 189-196.

[17] Hansen, B. E., 1996. Inference when a nuisance parameter is not identified under the null hypothesis. Econometrica 64, 413-430.

[18] Hansen, B.E., 2000. Sample splitting and threshold estimation. Econometrica 68, 575-603.

[19] Hansen, B., Seo, B., 2002. Testing for two-regime threshold cointegration in vector error-correction models. Journal of Econometrics 110, 293-318.

[20] Johansen, S., 1988. Statistical analysis of cointgerating vectors. Journal of Economic Dynamics and Control 12, 231-254.

[21] Kapetanios, G., Shin, Y., Snell, A., 2003. Testing for a Unit Root in the Nonlinear STAR Framework. Journal of Econometrics 112, 359-379.

[22] Luukkonen, R., Saikkonen, P., Terasvirta, T., 1988. Testing linearity against smooth transition autoregressive models. Biometrika 75, 491-499.

[23] Michael, M., Nobay, R., Peel, D. A., 1997. Transactions costs and nonlinear adjustment in real exchange rates: an empirical investigation. Journal of Political Economy 105, 862-879.

[24] Neftci, S., 1984. Are Economic Time Series Asymmetric over the Business Cycle? Journal of Political Economy 92, 307-329.

[25] Saikkonen, P., Choi, I., 2004. Cointegrating smooth transition regressions. Econometric Theory, forthcoming.

[26] Seo, B., 1998. Tests for structural change in cointegrated systems. Econometric Theory 14, 222-259.

[27] Terasvirta, T., 1994. Specification, Estimation, and Evaluation of Smooth Transition Autoregressive Models. Journal of the American Statistical Association 89, 208-218.

[28] Terasvirta, T., Anderson, H., 1992. Characterizing Nonlinearities in Business Cycles Using Smooth Transition Autoregressive Models. Journal of Applied Econometrics 7, S119-36.

[29] Tse, Y., 2001. Index Arbitrage with Heterogeneous Investors: A Smooth Transition Error Correction Analysis. Journal of Banking and Finance, 25, 1829-1855.

[30] White, H., 1980. A heteroskedasticity-consistent covariance matrix estimator and a direct test for heteroskedasticity. Econometrica 48, 817-838.

Figure 1. Dynamic Response of S&P Future Index (Δf_t)

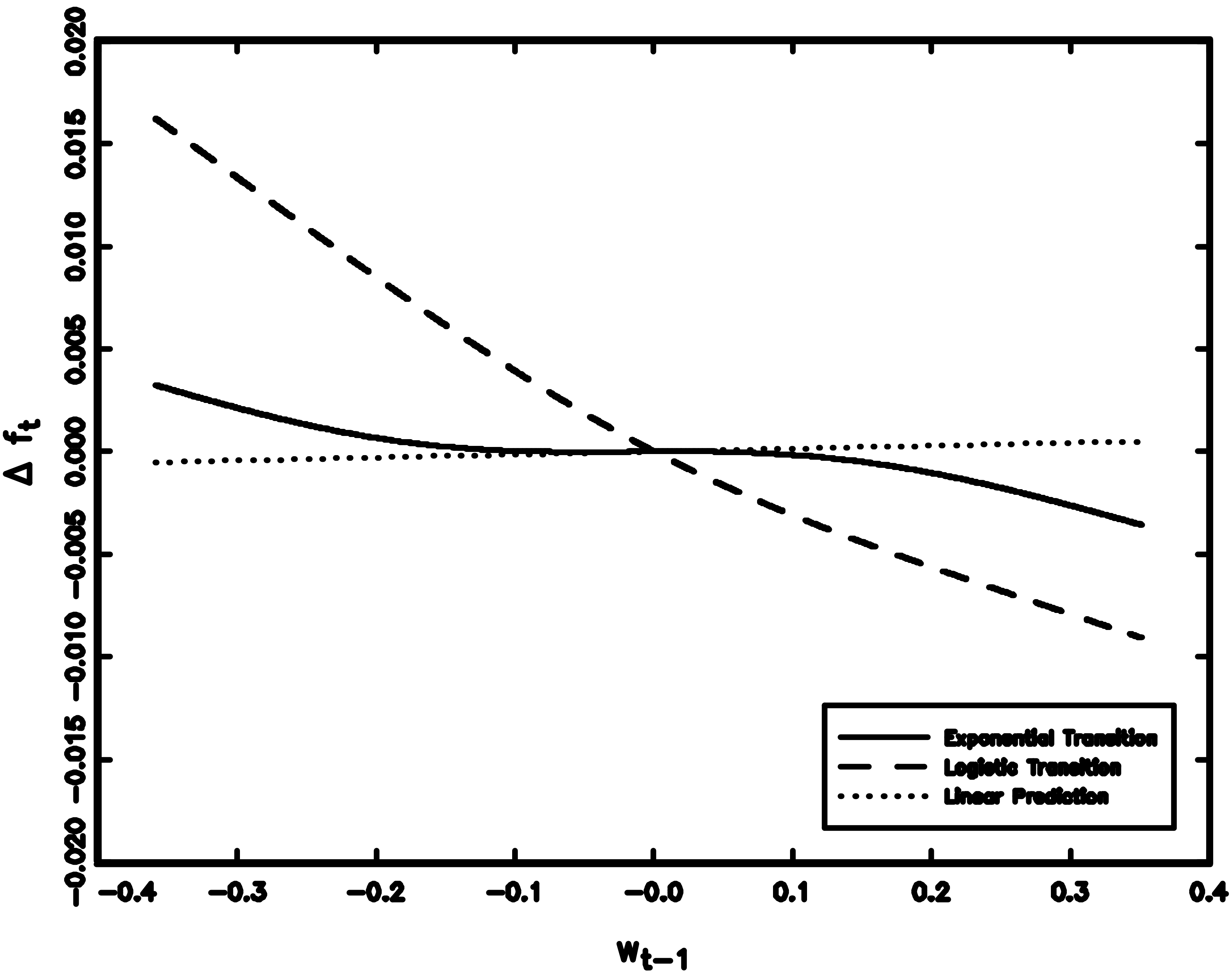

Figure 2. Dynamic Response of Theoretical Index (Δh_t)