Residual-based tests for cointegration in three-regime TAR models

Daiki Maki*
Ryukoku University

Shin-ichi Kitasaka†
Doshisha University

Abstract

This paper proposes residual-based tests for cointegration in three-regime threshold autoregressive (TAR) models. We propose Wald-type and t-type tests that have the null hypothesis of no cointegration and the alternative of cointegration with three-regime TAR adjustment, and also derive the asymptotic distributions. Monte Carlo simulations show that the proposed tests perform better than the Engle-Granger cointegration test and the cointegration test in a two-regime TAR model introduced by Enders and Siklos (2001), under cointegration with three-regime TAR adjustment, particularly when the threshold and sample size increase. When we apply these tests to the money demand of the U.S., the proposed tests reject the null of no cointegration whereas other tests do not.

Keywords: cointegration; three-regime TAR model; money demand
JEL Classification: C12; C22

*Corresponding author: Faculty of Economics, Ryukoku University, 67 Tsukamoto-cho, Fukakusa Fushimi-ku, Kyoto 612-8577 Japan (E-mail: maki@econ.ryukoku.ac.jp

†Faculty of Economics, Doshisha University, E-mail:skitasak@mail.doshisha.ac.jp

1 Introduction

Cointegration tests, which are important for investigating equilibrium relationship among economic variables, have already been used as standard tools for time series analysis. While these tests, including those in Engle and Granger (1987) and Johansen (1991), have been standard tests, they assume linear or constant adjustment toward equilibrium. Linear adjustment toward long-run equilibrium implies that the equilibrium error is continuously adjusted in all periods. However, the presence of transaction and trade costs causes discontinuous adjustment toward equilibrium (e.g., Balke and Fomby, 1997). Such adjustment is often modeled by the three-regime threshold autoregressive (TAR) process that has a unit root process in the middle regime but a stationary AR process in the outer regimes. Accordingly, the use of three-regime TAR models is valid in order to accurately analyze economic systems. In fact, the three-regime TAR model has been popularly used in economic applications, including those of purchasing power parity (Taylor, 2001), the term structure of interest rates (Clements and Galvão, 2003), and the law of one price (Lo and Zivot, 2001).

Bec, Ben Salem, and Carrasco (2004), Park and Shintani (2005), Kapetanios and Snell (2006), Seo (2008), and Bec, Guay, and Guerre (2008) have recently developed unit root tests, i.e., univariate cointegration tests with known cointegrating parameters, in three-regime TAR models[1)]. This is because, as discussed in Pippenger and Gorening (1993) and Taylor (2001), standard unit root tests have low power against three-regime TAR processes. Despite these significant studies, when applied researchers test for three-regime TAR cointegration with unknown cointegrating parameters, they employ a two-step procedure. The first step confirms the presence of cointegration relationship among variables using standard tests. If the cointegration relationship is obtained, the second step uses linearity tests to ascertain whether a threshold behavior exists. However, this procedure does not possess sufficient power against threshold cointegration because of the low power of standard cointegration tests under three-regime TAR cointegration similar to the standard unit root tests (e.g., Pippenger and Gorening, 2001). Since it is possible that the low power of standard cointegration tests leads to unreliable results, a direct test that has both the null hypothesis of no cointegration and the alternative hypothesis of cointegration with three-regime TAR adjustment is required.

This paper proposes residual-based tests for cointegration in three-regime TAR models. Although recent studies, including those of Enders and Siklos (2001), Hansen and Seo (2002), and Kapetanios, Shin, and Snell (2006), have also developed cointegration tests in nonlinear frameworks, residual-based tests for the null hypothesis of no cointegration against the alternative hypothesis of cointegration with three-regime TAR adjustment have not been established yet[2)]. Seo (2006) developed cointegration tests based on the three-regime TAR vector error correction model (VECM) with a known cointegrating vector. We use the residual-based approach because this approach not only makes it possible to estimate the cointegrating vector but also is simple and convenient for practitioners. We propose Wald-type and t-type tests and derive the asymptotic distributions. Unlike the limit distributions in Kapetanios and Shin (2006) and Seo (2006, 2008), our approach, in particular, does not degenerate with respect to the threshold parameters in the limit because we appropriately specify the parameter space of the thresholds. The point accurately shown by Park and Shintani (2005) is important in the case of testing for both cointegration and linearity. The proposed tests do not require bootstrap to calculate the critical values and improve over-rejection and are as convenient and practical for applied researchers as are standard cointegration tests.

Monte Carlo simulations demonstrate that under cointegration with three-regime TAR adjustment, the proposed tests perform better than the Engle-Granger cointegration test and the cointegration test in a two-regime TAR model introduced by Enders and Siklos (2001), particularly when the threshold and sample size increase. To substantiate the usefulness of our tests for empirical applications, we apply some tests to the money demand of the U.S. and find that only the proposed test rejects the null of no cointegration whereas other tests do not. This implies that the money demand of the U.S. has three-regime TAR adjustment toward equilibrium. Thus, Monte Carlo simulations and applications clearly prove the advantages of the proposed tests.

The organization of this paper is as follows. Section 2 presents the test statistics and asymptotic distributions. Section 3 investigates the size and power of the tests introduced in Section 2, by using Monte Carlo simulations. In Section 4, the empirical applications in U.S. money demand are presented. Finally, Section 5 summarizes the paper. Proofs of theorems are gathered in the appendix.

2 Testing for cointegration in three-regime TAR models

2.1 Test statistics

As an assumption of the tests for threshold cointegration, let y_t and $\mathbf{x}_t'$ denote observable $I(1)$ variables, where y_t is a scalar and $\mathbf{x}_t = (x_{1t}, \cdots, x_{mt})'$ is an $(m \times 1)$ vector. The long-run equilibrium relationship is given by

$$y_t = \beta' \mathbf{x}_t + u_t, \qquad t = 1, 2, \cdots, T, \tag{1}$$

where $\beta' = (\beta_1, \ldots, \beta_m)$ are estimated parameters and u_t is the equilibrium error. u_t follows the three-regime TAR process:

$$u_t = \begin{cases} \phi_1 u_{t-1} + e_t & \text{if } u_{t-1} \leq \lambda_1 \\ u_{t-1} + e_t & \text{if } \lambda_1 < u_{t-1} \leq \lambda_2, \\ \phi_2 u_{t-1} + e_t & \text{if } u_{t-1} > \lambda_2 \end{cases} \tag{2}$$

where e_t is a zero mean error, λ_1 and λ_2 are thresholds, and we assume $\lambda_2 > \lambda_1$[3)]. The existence of the long-run equilibrium relationship involves the stationarity of u_t. The stationarity conditions of (2) require that $-1 < \phi_1 < 1$ and $-1 < \phi_2 < 1$[4)]. For (2), while u_t has a unit root process in the range of $\lambda_1 < u_{t-1} \leq \lambda_2$ and does not revert to long-run equilibrium, i.e., 0, u_t reverts toward 0 if $u_{t-1} \leq \lambda_1$ or $u_{t-1} > \lambda_2$. u_t modeled by (2) is related to various economic phenomena where relatively small deviations from long-run equilibrium do not adjust the equilibrium error, while relatively large deviations do. Model (2) includes various restricted models. The Engle-Granger test is obtained by imposing $\phi_1 = \phi_2$ and $\lambda_1 = \lambda_2 = 0$. The two-regime threshold cointegration test introduced by Enders and Siklos (2001) is obtained by imposing $\lambda_1 = \lambda_2 = 0$. To test for two-regime threshold cointegration, Enders and Siklos (2001) proposed an F statistic.

To test for cointegration in (1) and (2), we consider the following regression model using the residual expressed by $\hat{u}_t = y_t - \hat{\beta}'\mathbf{x}_t$:

$$\Delta\hat{u}_t = \rho_1\hat{u}_{t-1}\mathbf{1}\{\hat{u}_{t-1} \leq \lambda_1\} + \rho_2\hat{u}_{t-1}\mathbf{1}\{\hat{u}_{t-1} > \lambda_2\} + \sum_{j=1}^{p}\alpha_j\Delta\hat{u}_{t-j} + \epsilon_t, \tag{3}$$

where ϵ_t is a stationary error with zero mean; $\mathbf{1}\{\cdot\}$ is the indicator function such that $\mathbf{1}\{\cdot\}$ is 1 if $\{\cdot\}$ is true and 0, otherwise[5)]. $\mathbf{1}\{\hat{u}_{t-1} \leq \lambda_1\}$ and $\mathbf{1}\{\hat{u}_{t-1} > \lambda_2\}$ are orthogonal to each other. The null hypothesis of no cointegration and the alternative hypothesis of threshold cointegration for (3) are

$$H_0 : \rho_1 = \rho_2 = 0, \qquad H_1 : \rho_1 < 0, \rho_2 < 0. \tag{4}$$

Denoting $\alpha = (\alpha_1, \ldots, \alpha_p)'$ and $\Delta\hat{u}_{t-p}^{p} = (\Delta\hat{u}_{t-1}, \ldots, \Delta\hat{u}_{t-p})$, (3) is rewritten as

$$\Delta\hat{u}_t = h_t'\theta + \epsilon_t, \tag{5}$$

where $h_t = (\hat{u}_{t-1}\mathbf{1}\{\hat{u}_{t-1} \leq \lambda_1\}, \hat{u}_{t-1}\mathbf{1}\{\hat{u}_{t-1} > \lambda_2\}, \Delta\hat{u}_{t-p}^{p})'$, and $\theta = (\rho_1, \rho_2, \alpha')'$.

We first consider a fixed $\lambda = (\lambda_1, \lambda_2)$. Let $\hat{\theta}$ be the OLS estimator of θ, $\hat{\epsilon}_t = \Delta\hat{u}_t - h_t'\hat{\theta}$, and $\hat{\sigma}^2 = \sum_{t=1}^{T}\hat{\epsilon}_t^2/(T-2-p)$. When $\lambda = (\lambda_1, \lambda_2)$ is given, the Wald statistic required to test for (4) is given by

$$W_T(\lambda) = \frac{1}{\hat{\sigma}^2}\hat{\rho}'\left[R\left(\sum_{t=1}^{T}h_th_t'\right)^{-1}R'\right]^{-1}\hat{\rho}, \tag{6}$$

where $\hat{\rho} = (\hat{\rho}_1, \hat{\rho}_2)'$ is the OLS estimator of ρ and R is the $2 \times (p+2)$ matrix such that $R\hat{\theta} = \hat{\rho}$. For an unknown $\lambda = (\lambda_1, \lambda_2)$, we compute (6) for each possible threshold and take the largest value across all possible thresholds. Then, the test statistic using a supremum-type statistic is defined as

$$\sup_{\lambda \in \Lambda_T} W_T(\lambda) \equiv \sup_{\lambda \in [\lambda_{min}, \lambda_{max}]} W_T(\lambda), \tag{7}$$

where Λ_T is a random sequence of the parameter space of thresholds given by the functions of $(\hat{u}_1, \cdots, \hat{u}_T)$. In order to utilize the sup statistic, it is required that $\lambda \in [\lambda_{min}, \lambda_{max}]$. First, we arrange the values of $\hat{u}_t$ in the ascending order, i.e., $\hat{u}_{(1)} < \hat{u}_{(2)} < \ldots < \hat{u}_{(T)}$; second, we select, for example, $\lambda_{min} = \hat{u}_{([5T/100])}$ and $\lambda_{max} = \hat{u}_{([95T/100])}$, where $[\cdot]$ is the integer part. Furthermore, when assuming that $0.1 \leq P(\lambda_1 \leq \hat{u}_t \leq \lambda_2) \leq 0.9$, this selection guarantees the existence of at least 10% of the observations for the inside and outside regimes. In this assumption, the threshold λ_1 includes equally spaced 100 points between the 5% and 45% quantiles of the arranged values, and the upper threshold λ_2 includes equally spaced 100 points between the 55% and 95% quantiles. Although the selections of λ_{min} and λ_{max} are rather arbitrary, it is important to guarantee sufficient observations to identify the regression parameters. When the cointegration relationship has a constant term, the demeaned residual $\hat{u}_t = y_t - \hat{\delta}_0 - \hat{\beta}'\mathbf{x}_t$ is employed. When the cointegration relationship has a constant and trend term, the demeaned and detrended residual $\hat{u}_t = y_t - \hat{\delta}_0 - \hat{\delta}_1 t - \hat{\beta}'\mathbf{x}_t$ is employed.

It should be noted that the Wald statistic cannot clarify the difference between H_1 of (4) and the intermediate case of the threshold no cointegration

$$H_2 : \rho_1 = 0,\ \rho_2 < 0 \ \text{ or } \ \rho_1 < 0,\ \rho_2 = 0 \tag{8}$$

because the Wald statistic increases even in the case of H_2. For a two-regime TAR model, Enders and Siklos (2001) and Caner and Hansen (2001) pointed out the problem and introduced a t-statistic. In order to accurately distinguish between H_0, H_1, and H_2 under three-regime TAR cointegration, we propose a t-statistic[6)]. Clearly, threshold cointegration requires $\rho_1 < 0$ and $\rho_2 < 0$. This implies that the condition of threshold cointegration is satisfied if the largest of the individual t-statistic is significant. For example, when the t-statistics of ρ_1 and ρ_2 are -2.5 and -1.7, respectively, investigating whether the

t-statistic of ρ_2 is significant will be sufficient for the condition of threshold cointegration. For a fixed $\lambda = (\lambda_1, \lambda_2)$, we denote the largest t-statistic between ρ_1 and ρ_2 as

$$t_T(\lambda)_{\max} = \max[t_1, t_2], \tag{9}$$

where t_1 and t_2 are t-statistics of ρ_1 and ρ_2, respectively. For an unknown $\lambda = (\lambda_1, \lambda_2)$, we compute $t_T(\lambda)_{\max}$ for each possible threshold and take the smallest value across all possible thresholds. Then, the test statistic using an infimum-type statistic is given by

$$\inf_{\lambda \in \Lambda_T} t_T(\lambda)_{\max} \equiv \inf_{\lambda \in [\lambda_{min}, \lambda_{max}]} t_T(\lambda)_{\max}. \tag{10}$$

We also introduce the BAND-TAR model given by

$$\Delta \hat{u}_t = (\mu_1 + \rho_1 \hat{u}_{t-1}) \mathbf{1}\{\hat{u}_{t-1} \leq \lambda_1\} + (\mu_2 + \rho_2 \hat{u}_{t-1}) \mathbf{1}\{\hat{u}_{t-1} > \lambda_2\} + \sum_{j=1}^{p} \alpha_j \Delta \hat{u}_{t-j} + \epsilon_t. \tag{11}$$

The main difference between (3) and (11) is the existence of each intercept parameter in the outer regimes. Equation (11) has a regime specific mean, but (3) does not. For (11), the equilibrium error adjusts to the edge of the band $[\lambda_1, \lambda_2]$. Unlike the linear model, the intercept parameters in (11) contribute to the persistence of the process. The BAND-TAR model (11) is motivated by situations such that a policy intervention attempts to control an equilibrium error within a target band rather than toward an equilibrium point (zero) for (3). The null hypothesis and the alternative hypothesis for (11) are[7)]

$$H_0 : \rho_1 = \rho_2 = 0, \qquad H_1 : \rho_1 < 0, \rho_2 < 0. \tag{12}$$

(11) is rewritten as

$$\Delta \hat{u}_t = g_t' \theta_B + \epsilon_t, \tag{13}$$

where $g_t = (\mathbf{1}\{\hat{u}_{t-1} \leq \lambda_1\}, \hat{u}_{t-1}\mathbf{1}\{\hat{u}_{t-1} \leq \lambda_1\}, \mathbf{1}\{\hat{u}_{t-1} > \lambda_2\}, \hat{u}_{t-1}\mathbf{1}\{\hat{u}_{t-1} > \lambda_2\}, \Delta \hat{u}_{t-p}^{p})'$ and $\theta_B = (\mu_1, \rho_1, \mu_2, \rho_2, \alpha')'$. For a fixed $\lambda = (\lambda_1, \lambda_2)$, the Wald statistic to test for (12) is

$$W_T^B(\lambda) = \frac{1}{\hat{\sigma}_B^2} \hat{\rho}' \left[R_B \left(\sum_{t=1}^{T} g_t g_t' \right)^{-1} R_B' \right]^{-1} \hat{\rho}, \tag{14}$$

where $\hat{\sigma}_B^2 = \sum_{t=1}^{T} \hat{\epsilon}_t^2/(T-4-p)$, $\hat{\epsilon}_t$ are the residuals obtained from (11), and R_B is the $2 \times (p+4)$ matrix such that $R_B\hat{\theta}_B = \hat{\rho}$. For an unknown $\lambda = (\lambda_1, \lambda_2)$, the test statistic is defined as

$$\sup_{\lambda\in\Lambda_T} W_T^B(\lambda) \equiv \sup_{\lambda\in[\lambda_{min},\lambda_{max}]} W_T^B(\lambda), \tag{15}$$

where λ is selected in a manner that is similar to (7). The test statistic using the t-type statistic such as (10) is given by

$$\inf_{\lambda\in\Lambda_T} t_T^B(\lambda)_{\max} \equiv \inf_{\lambda\in[\lambda_{min},\lambda_{max}]} t_T^B(\lambda)_{\max}, \tag{16}$$

where $t_T^B(\lambda)_{\max} = \max[t_1^B, t_2^B]$; t_1^B and t_2^B are t-statistics of ρ_1 and ρ_2 in (11).

2.2 Asymptotic distribution

To derive the asymptotic distribution, we denote an $(n \times 1)$ vector of $I(1)$ variables as $z_t = (y_t, \mathbf{x}_t')'$, where y_t is a scalar and $\mathbf{x}_t$ is an $m(= n-1)$ vector. z_t is generated by

$$z_t = z_{t-1} + \xi_t, \tag{17}$$

where ξ_t is assumed to be a stationary ARMA process with zero mean and a finite variance matrix. We make the following assumptions.

Assumption 1. (a) $\xi_t = \sum_{j=0}^{\infty} C_j \upsilon_{t-j}$, where $C_0 = I_n$, $\sum_{j=0}^{\infty} j\|C_j\| < \infty$, $\upsilon_t \sim$i.i.d.$(0, \sum)$ with $\sum > 0$, $E|\upsilon_t|^r < \infty$ for some $r \geq 4$, and I_n is the $(n \times n)$ identity matrix. (b) z_0 is a random vector with $\|z_0\| < \infty$.

From Assumption 1, $T^{-1/2}\sum_{t=1}^{[Tr]} \xi_t$ weakly converges to $(n \times 1)$ vector Brownian motion $[0, 1]$ with covariance matrix Ω. We denote $(n \times 1)$-vector Brownian motion as $B(r) = (B_y(r), B_x(r)')'$, where $B_x(r)$ is $(m \times 1)$-vector Brownian motion. Covariance matrix Ω is defined as

$$\Omega = \begin{pmatrix} \omega_{11} & \omega_{12}' \\ \omega_{21} & \Omega_{22} \end{pmatrix}, \tag{18}$$

where we assume $\Omega_{22} > 0$. Ω is decomposed as $\Omega = L'L$. L is given by

$$L = \begin{pmatrix} \ell_{11} & 0 \\ \ell_{21} & L_{22} \end{pmatrix}, \tag{19}$$

where $\ell_{11} = (\omega_{11} - \omega_{21}'\Omega_{22}^{-1}\omega_{21})^{1/2}$, $\ell_{21} = \Omega^{-1/2}\omega_{21}$, and $L_{22} = \Omega_{22}^{1/2}$. From (19), we have $B(r) = L'W(r)$, where $W(r) = (W_y(r), W_x(r)')'$ is $(n \times 1)$-vector standard Brownian motion.

When we set $\hat{\eta} = (1, -\hat{\beta}')'$ in regression (3), it follows that

$$\begin{aligned} \hat{\eta} &= \left(1, -(T^{-2}\textstyle\sum_{t=1}^{T} y_t \mathbf{x}_t')(T^{-2}\sum_{t=1}^{T} \mathbf{x}_t \mathbf{x}_t')^{-1}\right)' \\ &\Rightarrow \left(1, -(\textstyle\int_0^1 B_y B_x' \int_0^1 B_x B_x')^{-1}\right)' \equiv \eta. \end{aligned} \tag{20}$$

Using Lemma 2.2 of Phillips and Ouliaris (1990), it can be shown that

$$\begin{aligned} T^{-1/2}\hat{u}_t &\Rightarrow \eta' B(r) \\ &= \ell_{11} k' W(r) = \ell_{11} W^*(r), \end{aligned} \tag{21}$$

where $k = \left(1, -(\int_0^1 W_y W_x' \int_0^1 W_x W_x')^{-1}\right)'$ and $W^*(r) = W_y(r) - \int_0^1 W_y W_x' (\int_0^1 W_x W_x')^{-1} W_x(r)$. In addition, $\Delta\hat{u}_t$ is denoted as $\Delta\hat{u}_t = \hat{\eta}'\xi_t \Rightarrow \eta' dB(r)$. Since ξ_t is a stationary ARMA process under Assumption 1, $\Delta\hat{u}_t$ is also a stationary ARMA process. We represent it as $\epsilon_t = \sum_{j=0}^{\infty} D_j \Delta\hat{u}_{t-j} = D(L)\Delta\hat{u}_t$, where L is the lag operator.

For (3), the threshold parameter $\lambda = (\lambda_1, \lambda_2)$ has an identification problem: it is not identified under the null hypothesis of no cointegration, but it is identified only under the alternative hypothesis of cointegration with three-regime TAR adjustment. The problem wherein a nuisance parameter is identified only under the alternative hypothesis is known as the Davies problem. Davies (1987), Andrews and Ploberger (1994), and Hansen (1996) introduced approaches to overcome the problem. It should be noted that the transition variable $\hat{u}_{t-1}$ behaves differently under the null and alternative hypotheses. $\hat{u}_{t-1}$ explodes under H_0, but not under H_1. This implies that it is important to select an appropriate parameter space of thresholds in order to derive correct asymptotic distributions. If $\hat{u}_t$ is stationary under both H_0 and H_1, we can deal with similar parameter spaces of thresholds in the limit. However, the test that has the null hypothesis of no cointegration and the alternative hypothesis of cointegration with three-regime TAR adjustment is not such a case.

We specify the indicator function under the null hypothesis as

$$\mathbf{1}\{T^{-1/2}\hat{u}_{t-1} \leq T^{-1/2}\lambda_1\} \text{ and } \mathbf{1}\{T^{-1/2}\hat{u}_{t-1} > T^{-1/2}\lambda_2\} \text{ with } \lambda = (\lambda_1, \lambda_2) \in \Lambda_T \quad (22)$$

and under the alternative hypothesis as

$$\mathbf{1}\{\hat{u}_{t-1} \leq \lambda_1\} \text{ and } \mathbf{1}\{\hat{u}_{t-1} > \lambda_2\} \text{ with } \lambda = (\lambda_1, \lambda_2) \in \Lambda_T, \quad (23)$$

where Λ_T is a random sequence of the parameter space of thresholds given by functions of $(\hat{u}_1, \cdots, \hat{u}_T)$. Equations (22) and (23) have normalized and unnormalized sets of threshold parameters under H_0 and H_1, respectively. The setting is important for deriving the asymptotic distribution of the test statistics. We follow Park and Shintani (2005) and make the following assumption.

Assumption 2. $\Lambda_T \Rightarrow \Lambda$, where Λ is a compact subset of the real line.

Assumption 2 implies that limit parameter space Λ is a random subset of the real line and makes it possible to allow the presence of the middle regime under the null hypothesis. In other words, the probability of belonging to the middle regime with $\lambda_1 < \hat{u}_t \leq \lambda_2$ is positive at all times, even when the sample size increases. This probability is expressed as

$$\begin{aligned} P(\lambda_1 < \hat{u}_t \leq \lambda_2) &= P(T^{-1/2}\lambda_1 < T^{-1/2}\hat{u}_t \leq T^{-1/2}\lambda_2) \\ &\Rightarrow P(\lambda_{1T} < \eta' B(r) \leq \lambda_{2T}) \\ &= P(\tilde{\lambda}_1 < W(r)^* \leq \tilde{\lambda}_2) > 0, \end{aligned} \quad (24)$$

where $(\lambda_{1T}, \lambda_{2T}) = (T^{-1/2}\lambda_1, T^{-1/2}\lambda_2)$ and $(\tilde{\lambda}_1, \tilde{\lambda}_2) = (\ell_{11}^{-1}\lambda_{1T}, \ell_{11}^{-1}\lambda_{2T})$, respectively. Statistics derived from this assumption do not degenerate with respect to the threshold parameters in the limit. In contrast, if Assumption 2 does not hold, the probability of being in the middle regime with $\lambda_1 < \hat{u}_t \leq \lambda_2$ becomes 0 when the sample size grows. Thus,

$$P(\lambda_1 < \hat{u}_t \leq \lambda_2) = P(T^{-1/2}\lambda_1 < T^{-1/2}\hat{u}_t \leq T^{-1/2}\lambda_2) \to 0.$$

The difference in the probability of being in the middle regime is crucial for the asymptotic distribution, and as a result, it appears to cause the difference in the size and power performances.

The following theorem is required to derive the asymptotic distributions of the test statistics.

Theorem 1. *If Assumptions 1 and 2 and H_0 hold, then we obtain*

$$(1a) \qquad T^{-1}\sum_{t=1}^{T} \hat{u}_{t-1}\mathbf{1}\{\hat{u}_{t-1} \leq \lambda_1\}\epsilon_t \Rightarrow D(1)\ell_{11}^2 \int_0^1 \mathbf{1}\{W^* \leq \tilde{\lambda}_1\}W^* dW^*,$$

$$(1b) \qquad T^{-2}\sum_{t=1}^{T} \hat{u}_{t-1}^2\mathbf{1}\{\hat{u}_{t-1} \leq \lambda_1\} \Rightarrow \ell_{11}^2 \int_0^1 \mathbf{1}\{W^* \leq \tilde{\lambda}_1\}W^{*2},$$

$$(1c) \qquad T^{-1}\sum_{t=1}^{T} \hat{u}_{t-1}\mathbf{1}\{\hat{u}_{t-1} > \lambda_2\}\epsilon_t \Rightarrow D(1)\ell_{11}^2 \int_0^1 \mathbf{1}\{W^* > \tilde{\lambda}_2\}W^* dW^*,$$

$$(1d) \qquad T^{-2}\sum_{t=1}^{T} \hat{u}_{t-1}^2\mathbf{1}\{\hat{u}_{t-1} > \lambda_2\} \Rightarrow \ell_{11}^2 \int_0^1 \mathbf{1}\{W^* > \tilde{\lambda}_2\}W^{*2},$$

where W^* is a shorthand notation for $W^*(r)$ and $D(1) = \sum_0^\infty D_j$.

If the parameter space of thresholds is a fixed compact set, as assumed by Kapetanios and Shin (2006) and Seo (2006, 2008), the threshold parameters degenerate under a unit root process or no cointegration. Accordingly, the indicator functions from $(1a)$ to $(1d)$ in Theorem 1 are replaced by $\mathbf{1}\{W^* \leq 0\}$ or $\mathbf{1}\{W^* > 0\}$. As demonstrated by Seo (2006) and Kapetanios and Shin (2006), the use of sup-type tests in the parameter space of thresholds with a fixed compact set causes severe over-rejection under H_0, and as a result, the sup-type tests using asymptotic critical values will lead to unreliable results[8]. However, the test statistic proposed in the paper has a better size performance even when we employ asymptotic critical values; this is because we appropriately deal with the parameter space of thresholds. The following theorems present the asymptotic distributions of the test statistics.

Theorem 2. *If Assumptions 1 and 2 and* $p = o(T^{1/3})$ *hold, statistics (7) and (10) have the following asymptotic distributions.*

$$(2a) \qquad \sup_{\lambda \in \Lambda_T} W_T(\lambda) \Rightarrow \sup_{\lambda \in \Lambda} W(\lambda),$$

$$(2b) \qquad \inf_{\lambda \in \Lambda_T} t_T(\lambda)_{\max} \Rightarrow \inf_{\lambda \in \Lambda} t(\lambda)_{\max}.$$

$W(\lambda)$ *and* $t(\lambda)_{\max}$ *are defined as follows:*

$$W(\lambda) = \frac{\left(\int_0^1 1\{W^* \le \tilde{\lambda}_1\} W^* dW^*\right)^2}{(k'k)\int_0^1 1\{W^* \le \tilde{\lambda}_1\} W^{*2}} + \frac{\left(\int_0^1 1\{W^* > \tilde{\lambda}_2\} W^* dW^*\right)^2}{(k'k)\int_0^1 1\{W^* > \tilde{\lambda}_2\} W^{*2}}$$

and

$$t(\lambda)_{\max} = \max\left[\frac{\int_0^1 1\{W^* \le \tilde{\lambda}_1\} W^* dW^*}{\left(k'k\int_0^1 1\{W^* \le \tilde{\lambda}_1\} W^{*2}\right)^{1/2}}, \frac{\int_0^1 1\{W^* > \tilde{\lambda}_2\} W^* dW^*}{\left(k'k\int_0^1 1\{W^* > \tilde{\lambda}_2\} W^{*2}\right)^{1/2}}\right].$$

Under the alternative hypothesis that has cointegration with three-regime TAR adjustment, (2a) and (2b) diverge to infinity and minus infinity, respectively, as $T \to \infty$.

The test statistics of BAND-TAR model (11) have the following asymptotic distributions.

Theorem 3. *If Assumptions 1 and 2 and* $p = o(T^{1/3})$ *hold, statistics (15) and (16) have the following asymptotic distributions.*

(3a) $$\sup_{\lambda\in\Lambda_T} W_T^B(\lambda) \Rightarrow \sup_{\lambda\in\Lambda} W^B(\lambda),$$

(3b) $$\inf_{\lambda\in\Lambda_T} t_T^B(\lambda)_{\max} \Rightarrow \inf_{\lambda\in\Lambda} t^B(\lambda)_{\max}.$$

$W^B(\lambda)$ and $t^B(\lambda)_{\max}$ are defined as follows:

$$W^B(\lambda) = \frac{\left(\int_0^1 I_1 \int_0^1 I_1 W^* dW^* - \int_0^1 W^* I_1 \int_0^1 I_1 dW^*\right)^2}{k'k\{\int_0^1 I_1 \int_0^1 W^{*2} I_1 - (\int_0^1 W^{*2} I_1)^2\}} + \frac{\left(\int_0^1 I_2 \int_0^1 I_2 W^* dW^* - \int_0^1 W^* I_2 \int_0^1 I_2 dW^*\right)^2}{k'k\{\int_0^1 I_2 \int_0^1 W^{*2} I_2 - (\int_0^1 W^{*2} I_2)^2\}}$$

and

$$t^B(\lambda)_{\max} = \max\left[\frac{\int_0^1 I_1 \int_0^1 I_1 W^* dW^* - \int_0^1 W^* I_1 \int_0^1 I_1 dW^*}{\left(k'k\{\int_0^1 I_1 \int_0^1 W^{*2} I_1 - (\int_0^1 W^{*2} I_1)^2\}\right)^{1/2}}, \frac{\int_0^1 I_2 \int_0^1 I_2 W^* dW^* - \int_0^1 W^* I_2 \int_0^1 I_2 dW^*}{\left(k'k\{\int_0^1 I_2 \int_0^1 W^{*2} I_2 - (\int_0^1 W^{*2} I_2)^2\}\right)^{1/2}}\right]$$

where $I_1 = 1\{W^* \leq \tilde{\lambda}_1\}$ and $I_2 = 1\{W^* > \tilde{\lambda}_2\}$. Under the alternative hypothesis that has cointegration with BAND-TAR adjustment, (3a) and (3b) diverge to infinity and minus infinity, respectively, as $T \to \infty$.

For theorems 2 and 3, when the first regression has a constant term, W^* is replaced by the demeaned Brownian motion[9)]. Similarly, when the first regression has a constant and trend, W^* is replaced by the demeaned and detrended Brownian motion.

It should be noted that the critical values for the test statistics depend on λ_{min}, λ_{max}, and the number of regressors. However, since bootstrap methods are not necessary to calculate the critical values, the test is convenient and practical for applied researchers. Tables 1 and 2 show critical values of the tests from $m = 1$ to $m = 5$. The asymptotic critical values approximated by $T = 1,000$ are obtained from 10,000 replications. We present three models: Model 0 contains no deterministic terms; Model 1 contains an intercept in the first regression; Model 2 contains both intercept and trend in the first regression. We calculate these critical values for three types of grid space $(\lambda_{min}, \lambda_{max}) =$ Gp$(\hat{u}_{[\gamma T]}, \hat{u}_{[(1-\gamma)T]})$ with $0 < \gamma < 0.5$: G1 with $\gamma = 0.05$, G2 with $\gamma = 0.10$, and G3 with $\gamma = 0.15$. Each model searches from the $100\gamma\%$ to the $50(1-2\gamma)\%$ quantiles of the arranged sample $(\hat{u}_{(1)}, \cdots, \hat{u}_{(T)})$ to determine the lower threshold λ_1, and from the $50(1+2\gamma)\%$ to the $100(1-\gamma)\%$ quantiles to determine the upper threshold λ_2 such that at least $(100 \times 2\gamma)\%$ of the sample is in the middle regime. For each model, the threshold λ_1 includes equally spaced 100 points between the $100\gamma\%$ and $50(1-2\gamma)\%$ quantiles of the arranged values, and the upper threshold λ_2 includes equally spaced 100 points between the $50(1+2\gamma)\%$ and $100(1-\gamma)\%$ quantiles. Although the selection of γ is rather arbitrary, it is more important that it guarantees sufficient observations to identify the regression parameters.

3 Monte Carlo simulations

In this section, the size and power properties of the tests introduced in Section 2 are examined and compared to the properties of the tests in Engle and Granger (1987) and Enders and Siklos (2001). The nominal size of the test is 0.05, and we examine the sample sizes of $T = 100$, 200, and 400. For all experiments, 100 initial observations are discarded in order to avoid the effect of the initial conditions (the initial value is set to zero); that is, data with $T + 100$ are generated. The number of simulations is 10,000. In this section, we denote each test as follows: the t-type test of Engle and Granger (1987) as EG-t; the Φ test in a two-regime TAR model of Enders and Siklos (2001) as ES-Φ; (7) with the grid space G1 as $W1$ and G3 as $W3$; (10) with the grid space G1 as $t1$ and G3 as $t3$; (15) with the grid space G1 as W^B1 and G3 as W^B3; and (16) with the grid space G1 as t^B1 and G3 as t^B3. Note that the differences between G1 and G3 depend on the manner in which the grids for λ_1 and λ_2 are selected. It is important to investigate whether the size and power properties depend on the grid space of the thresholds. We use the demeaned data $\hat{u}_t = y_t - \hat{\delta}_0 - \hat{\beta}_1 x_{1t}$, where $\hat{\delta}_0$ and $\hat{\beta}_1$ are the OLS estimators.

3.1 Size

We generate the following data in order to examine the size performance:

$$y_t = \delta_0 + \beta_1 x_t + u_t, \tag{25}$$

$$\Delta u_t = e_t, \tag{26}$$

$$e_t = \phi e_{t-1} + \epsilon_{1t}, \tag{27}$$

$$\Delta x_t = \epsilon_{2t}, \tag{28}$$

$$\begin{pmatrix} \epsilon_{1t} \\ \epsilon_{2t} \end{pmatrix} \sim \text{i.i.d.N}\left(\mathbf{0} \begin{bmatrix} \sigma_1^2 & 0 \\ 0 & \sigma_2^2 \end{bmatrix}\right), \tag{29}$$

where $\delta_0 = 1$ and $\beta_1 = 2$, and we consider $\phi = (0.5, 0, -0.5)$, $\sigma_1^2 = 1$, and $\sigma_2^2 = (1, 4)$. For all tests, we use the regression with an augmentation term $\Delta\hat{u}_{t-1}$, except for $\phi = 0$. A lag order $\Delta\hat{u}_{t-1}$ is added to the models when $\phi \neq 0$, whereas no lag order is added when $\phi = 0$. Table 3 reports the rejection frequencies of the tests. The sizes of EG-t, ES-Φ, $W1$, $t1$, $W3$, and $t3$ are close to the nominal level of 5% and exhibit reasonable and acceptable size properties when $\phi = 0$ and $\sigma_2^2 = 1$, regardless of the sample size. The tests in BAND-TAR model slightly under-reject the null hypothesis in small samples. It should be noted that these tests require the estimation of additional parameters. This indicates that the estimation of additional parameters leads to under-rejection in a small sample. Although these tests tend to slightly under-reject the null hypothesis in a small sample, the under-rejection appears to become less frequent as the sample size increases. The use of the size-adjusted critical values based on a finite sample may be recommended for a strict analysis.

All the tests for $\phi = 0$ and $\sigma_2^2 = 4$ have properties that are similar to those when $\phi = 0$ and $\sigma_2^2 = 1$. The size does not depend on the degree of the variance of ϵ_{2t}. The EG-t and ES-Φ tests have an appropriate size even in the presence of a serially correlated error. The proposed tests slightly under-reject or over-reject the null hypothesis in the presence of a serially correlated error, whereas the under-rejections appear to become less frequent when the sample sizes are large. In addition, the under- or over-rejections of $W3$ are less significant than those of $W1$. Although sup-type tests using asymptotic critical values as demonstrated by Kapetanios and Shin (2006) and Seo (2006) exhibit severe over-rejections even in a relatively large sample, our proposed statistics have more reasonable and acceptable sizes even in a small sample and have no severe over-rejections. It is noteworthy that our proposed tests do not degenerate with respect to the threshold parameters under no cointegration, and the properties allow a positive probability of being in the middle regime in the limit. The results indicate that it is necessary to (a) have the appropriate threshold parameter space that does not degenerate under no cointegration and (b) guarantee the probability of being in the middle regime in the limit, in order to acquire a good size performance in a small sample when we use asymptotic critical values.

3.2 Power

Next, we focus on the power comparison under threshold cointegration. To avoid the effects of slight under-rejection, as reported in Table 3, and accurately examine the power performance, we use size-adjusted critical values based on finite samples[10)]. The data to examine the power is generated as follows:

$$y_t = \delta_0 + \beta_1 x_t + u_t, \tag{30}$$

$$\Delta u_t = \rho_1 u_{t-1} \mathbf{1}_{\{u_{t-1} \leq -\lambda\}} + \rho_2 u_{t-1} \mathbf{1}_{\{u_{t-1} > \lambda\}} + \epsilon_{1t}, \tag{31}$$

$$\Delta x_t = \epsilon_{2t}, \tag{32}$$

$$\begin{pmatrix} \epsilon_{1t} \\ \epsilon_{2t} \end{pmatrix} \sim \text{i.i.d.N} \left(\mathbf{0} \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \right), \tag{33}$$

where $\delta_0 = 1$, $\beta_1 = 2$, $\rho_1 = \rho_2 = -0.05$, and $\lambda = (0, 2, 4, 8)$. Table 4 presents the results of power performance and the percentage of data in the middle regime. The increase in λ expands the no-adjustment region toward equilibrium. Although the power of the tests increases with the increase in the sample size, a general finding obtained from Table 4 is that the powers of all the tests change drastically with an increase in λ, i.e., with an increase in the percentage of the middle regime. This implies that the threshold has a clear impact on the power of each test, particularly as the threshold increases.

When $\lambda = 0$, (31) reduces to an AR(1) model. The EG-t test performs best for this alternative since this alternative is designed for this test. In addition, the EG-t test also has a better performance even when the alternative is threshold cointegration with $\lambda = 2$. It is noteworthy that in the case where the EG-t test performs better than the tests based on three-regime TAR models, the percentage of the middle regime is small; for example, it is approximately 43% for $\lambda = 2$. However, the power of the EG-t test decreases rapidly as the no-adjustment region expands and the sample size increases. In other words, the larger the percentage of the middle regime, the more difficult it is for the EG-t test to detect the reversion toward long-run equilibrium.

However, the decrease in the powers of the proposed tests is clearly different from that of the EG-t test. When the threshold increases, the power of the sup tests is superior to that of the EG-t test, and the decrease in the powers of the Wald-type tests are much slower than that of the EG-t test, which ignores the threshold behavior λ. We notice that the middle regime % employed by these statistics slowly approaches the true value when the sample size increases. The ES-Φ test, which is designed for two-regime TAR models, also performs better than the EG-t test when the threshold and sample size increase. For example, from Table 4 with $\lambda = 8$ and $T = 400$, we observe that the powers of EG-t, ES-Φ, $W1$, $W3$, W^B1, and W^B3 are 0.156, 0.172, 0.256, 0.224, 0.183, and 0.180, respectively. A threshold does not have a clear impact on the powers of Wald-type tests as compared with the EG-t and ES-Φ tests. The results show that the Wald-type tests perform better than EG-t and ES-Φ when the threshold and sample size increase.

A comparison between Wald-type and t-type tests show that Wald-type tests are dominant to t-type tests when the threshold and sample size increase. Practitioners are advised to use the Wald-type tests if economic theories predict threshold cointegration and the

EG-t test cannot reject the null hypothesis of no cointegration. The t-type tests would be used in a supplementary test to ascertain the robustness of stationarity. In addition, $W3$ and $t3$ outperform $W1$ and $t1$, respectively, except for $\lambda = 8$ and $T = 400$. $W1$, which is the statistic with the largest grid space of thresholds, performs poorly under the three-regime TAR process with a small threshold, whereas it performs better in the case of an increase in the percentage of the middle regime for a large sample size. The results imply that the grid space of the threshold parameters influences the power performance of the tests.

The rejection rates of the BAND-TAR models are significantly low for a small threshold and/or small samples. It is possible that the estimation of additional parameters and the test for joint significance lead to a low power in a small threshold and/or small samples. In fact, W^B1 and W^B3 give less rejection rates than do $W1$ and $W3$. From the underlying model, it is expected that the test based on the BAND-TAR model outperforms EG-t when the true process is a three-regime TAR process. Contrary to expectation, the test based on the BAND-TAR model has much lower power than the EG-t test, particularly when the threshold and sample size are relatively small. The BAND-TAR model may be useless in uncovering long-run equilibrium as long as the threshold and/or sample size are small. For example, for $\lambda = 2$ and $T = 200$, the powers of EG-t, ES-Φ, $W1$, $W3$, W^B1, and W^B3 are 0.194, 0.179, 0.148, 0.166, 0.071, and 0.066, respectively.

Table 5 reports the results for $\rho_1 = \rho_2 = -0.3$. We observe from Tables 4 and 5 that the no-adjustment region for $\rho_1 = \rho_2 = -0.3$ is larger than that for $\rho_1 = \rho_2 = -0.05$. For the case when $\lambda = 4$, the three-regime TAR models with $\rho_1 = \rho_2 = -0.05$ have approximately 68% observations in the middle regime, whereas the three-regime TAR models with $\rho_1 = \rho_2 = -0.3$ have approximately 90% observations. The power gain of the Wald-type tests over the EG-t and ES-Φ tests become more substantial than $\rho_1 = \rho_2 = -0.05$ when the threshold and sample size increase. For example, when $T = 400$ and $\lambda = 8$, the powers of EG-t, ES-Φ, $W1$, $W3$, W^B1, and W^B3 are 0.292, 0.350, 0.786, 0.541, 0.856, and 0.791, respectively. The use of the Wald-type tests is considerably more valid than that of the

EG-t and ES-Φ tests. It should be noted that W^B1 and W^B3 also have a high power superiority for a large threshold. This indicates that the BAND-TAR model performs much better as the threshold and sample size increase, and the outer regimes are less persistent.

We also examine power performance under the three-regime TAR process with an asymmetric adjustment $(\rho_1, \rho_2) = (-0.15, -0.05)$. It would be more informative to assess the power for $\rho_1 \neq \rho_2$ because some applications supported asymmetric adjustment (e.g., Bec, Ben Salem, and Carrasco, 2004; Kapetanios and Shin, 2006). The test results presented in Table 6 indicate that all the tests have a higher power than those given in Table 4, but a lower power than those given in Table 5. The percentage of observations in the middle regime is greater than that in Table 4 but lower than that in Table 5. Thus, the results in Table 6 indicate performances between Tables 4 and 5.

Table 7 presents power performance under the BAND-TAR model given by

$$\Delta u_t = (\mu_1 + \rho_1 u_{t-1})\mathbf{1}_{\{u_{t-1} \leq -\lambda\}} + (\mu_2 + \rho_2 u_{t-1})\mathbf{1}_{\{u_{t-1} > \lambda\}} + \epsilon_t, \tag{34}$$

where $\mu_1 = \lambda\rho_1$, $\mu_2 = -\lambda\rho_2$, $\rho_1 = \rho_2 = -0.3$, and $\lambda = (2, 4, 8)$. The adjustment speed in the outer regimes is similar to that of Table 5. A comparison between the results of Tables 5 and 7 provides evidence that the rejection frequencies of the tests are affected by a constant of each regime. The reason is that BAND-TAR model (34) has a larger persistence than equilibrium TAR model (31) when both models have the same adjustment speed in outer regimes[11)]. However, we do not observe much difference between the results of Tables 4 and 7 because the percentage of observations in the middle regime for Table 7 is similar to that of Table 4, where the adjustment speed of the outer regime of (31) is $\rho_1 = \rho_2 = -0.05$. Although it is expected that W^B1 and W^B3 designed for the alternative hypothesis of the BAND-TAR model exhibit the best power among the tests, their power performances are not markedly different from those of $W1$ and $W3$ even for a large threshold. Employing the BAND-TAR model may not significantly influence the power even if the true process is a BAND-TAR one.

4 Applications to the money demand of the U.S.

In this section, we apply the tests introduced in Section 2 to the money demand of the U.S. We consider the money demand function as follows:

$$\text{Double-log model:} \quad M_t/P_t = \delta_0 + \beta_Y Y_t + \beta_R R_t + u_t, \tag{35}$$

$$\text{Semi-log model:} \quad M_t/P_t = \delta_0 + \beta_Y Y_t + \beta_r r_t + u_t, \tag{36}$$

where M_t, P_t, Y_t, R_t, and r_t denote nominal money, prices, real income, interest rate in logarithm, and interest rate in level, respectively. u_t denotes the equilibrium error of the money demand function. For the underlying theoretical backgrounds of the TAR model in money demand, Milbourne (1987) pointed out that money demand was characterized by the Buffer-Stock model. This shows that agents in economies do not act to adjust their money balances when the deviation from equilibrium is within adjustment costs but they do so when the deviation is relatively large, i.e., over the thresholds. Sarno, Taylor, and Peel (2003), who proposed the nonlinear error correction model in the presence of transaction costs, supported the smooth transition adjustment of the money demand of the U.S. (see also Cuthbertson and Taylor, 1987; Mizen, 1997; Sarno, 1999). Maki and Kitasaka (2006) provided empirical evidence to prove that the money demand in Japan is characterized by a two-regime TAR process. These studies provide findings supporting nonlinear adjustment of money demand.

We use M1 as nominal money, the consumer price index as prices, and the index of industrial production as real income. These variables are data that are seasonally adjusted. For the interest rate, we use the three-month Treasury Bill interest rate. The monthly data obtained from the Federal Reserve Bank of St.Louis consist of 577 periods from 1960:1 to 2008:1. We consider two sample periods -1960:1-2008:1(577 periods) and 1979:10-2008:1(337 periods)- corresponding to changes in Federal Reserve operating procedures. The tests have a constant term in the first regression. Although we do not present the results of the unit root tests of variables, the tests provide evidence of $I(1)$. We determined the appropriate lag length for $p = 12$ and then excluded the insignificant augmentation terms. Since Enders and Siklos (2001) did not tabulate the critical values of the Φ test for regressors with $m = 2$, we calculated the values[12)].

Table 8 presents the results of cointegration tests. We find clear differences among the tests. The results of the double-log model for the full sample show that EG-t, ES-Φ, $W1$, $W3$, and $t3$ reject the null hypothesis, but the BAND-TAR model does not. This implies that the equilibrium error is a three-regime TAR model with a small percentage of the middle regime because Monte Carlo simulations in the previous section demonstrate that the BAND-TAR model does not perform well when the true process is a three-regime TAR model with a small percentage of the middle regime. On the other hand, in the case of the semi-log model, the EG-t and ES-Φ tests do not reject the null hypothesis of no cointegration in both the sample periods even at the 10% significance level, whereas tests based on the three-regime TAR models reject the null hypothesis at the 1% or 5% significance level and provide strong evidence of a cointegration relationship. Monte Carlo simulations and empirical results indicate that the equilibrium error of money demand in the semi-log model is a three-regime TAR process with a large threshold because tests based on the three-regime TAR models reject the null hypothesis of no cointegration, whereas the EG-t and ES-Φ tests fail to reject the null hypothesis. Thus, we obtained more stable and reliable results for threshold cointegration.

Figures 1 and 2 draw the time path of the residuals for two samples. Both figures denote a strong tendency to move back toward equilibrium, although the processes appear to be highly persistent in the sub sample. In order to confirm these findings and the results of Table 8, we estimate the following model:

$$\Delta \hat{u}_t = \rho_1 \hat{u}_{t-1} \mathbf{1}_{\{\hat{u}_{t-1} \leq \lambda_1\}} + \rho_2 \hat{u}_{t-1} \mathbf{1}_{\{\hat{u}_{t-1} > \lambda_2\}} + \sum_{j=1}^{p} \alpha_j \Delta \hat{u}_{t-j} + \epsilon_t, \tag{37}$$

where $\hat{u}_t$ denotes the demeaned residual such that $\hat{u}_t = M_t/P_t - \hat{\delta}_0 - \hat{\beta}_Y Y_t - \hat{\beta}_R R_t$ or $\hat{u}_t = M_t/P_t - \hat{\delta}_0 - \hat{\beta}_Y Y_t - \hat{\beta}_r r_t$. We determined the appropriate lag length for $p = 12$ and then excluded the insignificant augmentation terms. The threshold parameter $\lambda = (\lambda_1, \lambda_2)$

is estimated by minimizing the sum of the squared residuals over $[\lambda_{min}, \lambda_{max}]$, where λ_{min} and λ_{max} are set to 5% and 95% quantiles, respectively, of the ordered sample $\hat{u}_{(1)} < \hat{u}_{(2)} < \ldots < \hat{u}_{(T)}$.

From the estimations presented in Table 9, the middle regime for the double-log model in the full sample is approximately 15%. This would be the reason why EG-t, ES-Φ, $W1$, and $W3$ in Table 8 rejected the null hypothesis but the tests in the BAND-TAR model did not. In the case of the semi-log model, we notice that the percentage of the middle regime for the sub sample is approximately 90%. Furthermore, the adjustment process toward equilibrium is persistent below the estimated threshold parameter $\hat{\lambda}_1$, whereas the deviations from the equilibrium are quickly eliminated above the estimated threshold parameter $\hat{\lambda}_2$. We could interpret the results of Tables 8 and 9 from the Monte Carlo simulations presented in Table 6, where Wald-type tests have a higher power when the threshold is large, i.e., when the middle regime percentage increases. The reason why EG-t and ES-Φ tests fail to reject the null hypothesis of no cointegration for the semi-log model, as demonstrated in the previous section, is that these tests have a low power when the equilibrium error is the three-regime TAR process with a relatively large percentage of observations in the middle regime. These findings strongly support the stability of the long-run money demand characterized by the three-regime TAR model. Accordingly, the use of the cointegration test in three-regime TAR models provides a better alternative to standard cointegration tests in the cases where economic theories indicate three-regime TAR adjustment.

5 Summary

This paper has proposed residual-based tests for cointegration in three-regime TAR models. We have proposed the Wald-type and t-type statistics, which are the tests for the null hypothesis of no cointegration against the alternative of cointegration with the three-regime TAR adjustment, and derived its asymptotic distributions. Since our approach does not degenerate with respect to the threshold parameters in the limit, the proposed tests using supremum statistics have better size properties and do not require bootstrap to calculate the critical values and, like standard cointegration tests, are convenient and practical for applied researchers.

Monte Carlo simulations demonstrated that the proposed tests perform better than the Engle-Granger cointegration test and the two-regime cointegration test introduced by Enders and Siklos (2001), under cointegration with three-regime TAR adjustment, particularly when the threshold and sample size increase. When we applied these tests to the money demand of the U.S., the proposed tests rejected the null of no cointegration whereas the other tests did not; that is, by allowing for three-regime TAR adjustment, we obtained results showing the stability of the money demand function. This implies that the money demand of the U.S. has three-regime TAR adjustment toward equilibrium. The results obtained from Monte Carlo simulations and the applications to the money demand of the U.S. underline the usefulness of the proposed tests for applied investigators in cases where cointegration tests are used in three-regime TAR frameworks.

Footnotes

1) Park and Shintani (2005) developed unit root tests in various transitional AR models, including the three-regime TAR model. Maki (2009) reviewed unit root tests in three-regime TAR models and investigated the power of these models. Other unit root tests in nonlinear frameworks have also been proposed by Enders and Siklos (1998), Caner and Hansen (2001), and Kapetanios, Shin, and Snell (2003). Choi and Moh (2007) investigated the power of some unit root tests in nonlinear frameworks.

2) Enders and Siklos (2001) proposed cointegration tests in two-regime TAR models. Hansen and Seo (2002) developed the test in two-regime TAR vector error correction models, and Kapetanios, Shin, and Snell (2006) proposed cointegration tests in smooth transition autoregressive models.

3) If (2) is a three-regime TAR model with a symmetric adjustment, the model under the alternative is $-1 < \phi_1 = \phi_2 < 1$. In addition, a special case of a symmetric three-regime TAR model is given by $u_t = \phi u_{t-1}\mathbf{1}_{\{|u_{t-1}|>\lambda\}} + e_t$. Since (2) also includes these restricted models, we only consider general model (2).

4) See Bec et al. (2004) and Kapetanios and Shin (2006) for details on the stationarity condition of three-regime TAR models.

5) Although it is possible that $\Delta\hat{u}_{t-j}$ also follows a TAR process, for the sake of simplicity, we do not consider this case as in Enders and Siklos (2001). Even if $\Delta\hat{u}_{t-j}$ is a TAR process, the asymptotic distribution of the test statistic does not change.

6) It is possible to propose a test that has the null hypothesis of ($\rho_1 < 0$ and $\rho_2 = 0$) or ($\rho_1 = 0$, $\rho_2 < 0$) and the alternative hypothesis of ($\rho_1 < 0$ and $\rho_2 < 0$). However, as pointed out by Seo (2006), the parameter space for the test is complicated under the null hypothesis. Therefore, we introduce the test in the present paper.

7) Although the intercept parameter in the BAND-TAR model also plays a role in determining the stationarity of u_t, we focus on the AR parameters because the stationarity of the AR parameters is important for general tests for unit root and cointegration. See Balke and Fomby (1997) for the returning drift model, where the intercept parameter in the BAND-TAR model determines the existence of cointegration.

8) Seo (2006) proposed a bootstrap method to improve size distortions.

9) If $\tilde{\lambda}_1 = \tilde{\lambda}_2 = 0$ or $\tilde{\lambda}_1 = \tilde{\lambda}_2$, the test statistic is found to be similar to the Φ statistic by using an F statistic of Enders and Siklos (2001), although they did not show the asymptotic distribution.

10) Owing to space constraints, we have not tabulated finite size critical values. These are available with the author and will be provided on request.

11) See also Balke and Fomby (1997) and Lo and Zivot (2001).

12) Critical values of the ES-Φ test are 6.208, 7.272, and 9.640 at 90%, 95%, and 99%, respectively.

Appendix

Proof of Theorem 1

From (21), we obtain that $T^{-1/2}\ell_{11}^{-1}\hat{u}_t \Rightarrow W^*(r)$. Given that the function $\mathbf{1}\{\hat{u}_{t-1} \leq \lambda_1\}$ is continuous in $\hat{u}_{t-1}$, using Lemma A2 of Park and Phillips (2001), we have

$$\begin{aligned} \mathbf{1}\{\hat{u}_{t-1} \leq \lambda_1\} &= \mathbf{1}\{T^{-1/2}\hat{\eta}' z_{t-1} \leq T^{-1/2}\lambda_1\} \\ &\Rightarrow \mathbf{1}\{\eta' B(r) \leq \lambda_{1T}\} = \mathbf{1}\{W^*(r) \leq \tilde{\lambda}_1\}, \end{aligned} \tag{A.1}$$

where $\lambda_{1T} = T^{-1/2}\lambda_1$ and $\tilde{\lambda}_1 = \ell_{11}^{-1}\lambda_{1T}$. From $\epsilon_t = \sum_{j=0}^{\infty} D_j \xi'_{t-j}\eta = D(L)\xi'_t\eta$ and Lemma 2.1 of Phillips and Ouliaris (1990), we have

$$T^{-1/2}\sum_{t=1}^{[Tr]} \epsilon_t \Rightarrow D(1)\eta' B(r), \tag{A.2}$$

where $D(1) = \sum_{j=0}^{\infty} D_j$. (A.1), (A.2), and the continuous mapping theorem (CMT) yield

$$\begin{aligned} T^{-1}\sum_{t=1}^{T} \hat{u}_{t-1}\mathbf{1}\{\hat{u}_{t-1} \leq \lambda_1\}\epsilon_t &= T^{-1/2}\hat{\eta}' \sum_{t=1}^{T} T^{-1/2} z_{t-1}\mathbf{1}\{T^{-1/2}\hat{\eta}' z_{t-1} \leq T^{-1/2}\lambda_1\} D(L)\xi'_t\hat{\eta} \\ &\Rightarrow D(1)\eta' \int_0^1 BdB'\eta\mathbf{1}\{\eta' B \leq \lambda_{1T}\} \\ &= D(1)\ell_{11}^2 \int_0^1 \mathbf{1}\{W^* \leq \tilde{\lambda}_1\} W^* dW^* \end{aligned} \tag{A.3}$$

and

$$\begin{aligned} T^{-2}\sum_{t=1}^{T} \hat{u}_{t-1}^2\mathbf{1}\{\hat{u}_{t-1} \leq \lambda_1\} &= T^{-1}\sum_{t=1}^{T}(T^{-1/2}z'_{t-1}\eta)'(T^{-1/2}z'_{t-1}\eta)\mathbf{1}\{T^{-1/2}\hat{\eta}' z_{t-1} \leq T^{-1/2}\lambda_1\} \\ &\Rightarrow \eta' \int_0^1 BB'\eta\mathbf{1}\{\eta' B \leq \lambda_{1T}\} \\ &= \ell_{11}^2 \int_0^1 \mathbf{1}\{W^* \leq \tilde{\lambda}_1\} W^{*2}. \end{aligned} \tag{A.4}$$

A similar type of analysis can be applied to the term for $\mathbf{1}\{\hat{u}_{t-1} > \lambda_2\}$. Therefore, it can be also shown that

$$T^{-1}\sum_{t=1}^{T} \hat{u}_{t-1}\mathbf{1}\{\hat{u}_{t-1} > \lambda_2\}\epsilon_t \Rightarrow D(1)\ell_{11}^2 \int_0^1 \mathbf{1}\{W^* > \tilde{\lambda}_2\} W^* dW^* \tag{A.5}$$

and

$$T^{-2}\sum_{t=1}^{T}\hat{u}_{t-1}^{2}\mathbf{1}\{\hat{u}_{t-1}>\lambda_2\}\Rightarrow \ell_{11}^{2}\int_0^1 \mathbf{1}\{W^*>\tilde{\lambda}_2\}W^{*2}. \tag{A.6}$$

Proof of Theorem 2

Part of (2a). The statistic $W_T(\lambda)$ can be written as

$$W_T(\lambda)=\frac{1}{\hat{\sigma}^2}\hat{\rho}'(\mathbf{U}'\mathbf{Q}_p\mathbf{U})\hat{\rho}, \tag{A.7}$$

where

$$\mathbf{U}=\begin{pmatrix}\hat{u}_0\mathbf{1}\{\hat{u}_0\le\lambda_1\} & \hat{u}_0\mathbf{1}\{\hat{u}_0>\lambda_2\}\\ \hat{u}_1\mathbf{1}\{\hat{u}_1\le\lambda_1\} & \hat{u}_1\mathbf{1}\{\hat{u}_1>\lambda_2\}\\ \vdots & \vdots\\ \hat{u}_{T-1}\mathbf{1}\{\hat{u}_{T-1}\le\lambda_1\} & \hat{u}_{T-1}\mathbf{1}\{\hat{u}_{T-1}>\lambda_2\}\end{pmatrix}$$

and $\mathbf{Q}_p=\mathbf{I}-\mathbf{M}_p(\mathbf{M}_p'\mathbf{M}_p)^{-1}\mathbf{M}_p'$ is a $T\times T$ idempotent matrix. $\mathbf{M}_p$ is the matrix of observations on $\Delta\hat{u}_{t-p}^{p}$. Since under H_0 with $\rho_1=\rho_2=0$, $\hat{\rho}$ is given by $\hat{\rho}=(\mathbf{U}'\mathbf{Q}_p\mathbf{U})^{-1}\mathbf{U}'\mathbf{Q}_p\boldsymbol{\epsilon}$, where $\boldsymbol{\epsilon}=(\epsilon_1,\cdots,\epsilon_T)'$, (A.7) is expressed by

$$W_T(\lambda)=\frac{1}{\hat{\sigma}^2}\boldsymbol{\epsilon}'\mathbf{Q}_p\mathbf{U}(\mathbf{U}'\mathbf{Q}_p\mathbf{U})^{-1}\mathbf{U}'\mathbf{Q}_p\boldsymbol{\epsilon}. \tag{A.8}$$

Let $\mathbf{U}_1=(\hat{u}_0\mathbf{1}\{\hat{u}_0\le\lambda_1\},\cdots,\hat{u}_{T-1}\mathbf{1}\{\hat{u}_{T-1}\le\lambda_1\})'$ and $\mathbf{U}_2=(\hat{u}_0\mathbf{1}\{\hat{u}_0>\lambda_2\},\cdots,\hat{u}_{T-1}\mathbf{1}\{\hat{u}_{T-1}>\lambda_2\})'$. From the orthogonality between $\mathbf{1}\{\hat{u}_{t-1}\le\lambda_1\}$ and $\mathbf{1}\{\hat{u}_{t-1}>\lambda_2\}$, we can decompose (A.8) as

$$\begin{aligned}W_T(\lambda)&=(\boldsymbol{\epsilon}'\mathbf{Q}_p\mathbf{U}_1,\boldsymbol{\epsilon}'\mathbf{Q}_p\mathbf{U}_2)\begin{pmatrix}\mathbf{U}_1'\mathbf{Q}_p\mathbf{U}_1 & 0\\ 0 & \mathbf{U}_2'\mathbf{Q}_p\mathbf{U}_2\end{pmatrix}^{-1}\begin{pmatrix}\mathbf{U}_1'\mathbf{Q}_p\boldsymbol{\epsilon}\\ \mathbf{U}_2'\mathbf{Q}_p\boldsymbol{\epsilon}\end{pmatrix}\\ &=\frac{1}{\hat{\sigma}^2}\{\boldsymbol{\epsilon}'\mathbf{Q}_p\mathbf{U}_1(\mathbf{U}_1'\mathbf{Q}_p\mathbf{U}_1)^{-1}\mathbf{U}_1'\mathbf{Q}_p\boldsymbol{\epsilon}+\boldsymbol{\epsilon}'\mathbf{Q}_p\mathbf{U}_2(\mathbf{U}_2'\mathbf{Q}_p\mathbf{U}_2)^{-1}\mathbf{U}_2'\mathbf{Q}_p\boldsymbol{\epsilon}\}.\end{aligned} \tag{A.9}$$

Consider $\boldsymbol{\epsilon}'\mathbf{Q}_p\mathbf{U}_1(\mathbf{U}_1'\mathbf{Q}_p\mathbf{U}_1)^{-1}\mathbf{U}_1'\mathbf{Q}_p\boldsymbol{\epsilon}$. It follows from CMT that $T^{-1}\mathbf{U}_1'\mathbf{M}_p=O_p(1)$, $T^{-1}\mathbf{M}_p'\mathbf{M}_p=O_p(1)$, and $T^{-1/2}\mathbf{M}_p'\boldsymbol{\epsilon}=O_p(1)$. Combining these results and Theorem 1, we have

$$\begin{aligned}T^{-1}\mathbf{U}_1'\mathbf{Q}_p\boldsymbol{\epsilon}&=T^{-1}\mathbf{U}_1'\boldsymbol{\epsilon}-T^{-1/2}\cdot T^{-1}\mathbf{U}_1'\mathbf{M}_p(T^{-1}\mathbf{M}_p'\mathbf{M}_p)^{-1}T^{-1/2}\mathbf{M}_p'\boldsymbol{\epsilon}\\ &=T^{-1}\mathbf{U}_1'\boldsymbol{\epsilon}+o_p(1)\Rightarrow D(1)\ell_{11}^{2}\int_0^1 1\{W^*\le\tilde{\lambda}_1\}W^*dW^*\end{aligned} \tag{A.10}$$

and

$$T^{-2}\mathbf{U}_1'\mathbf{Q}_p\mathbf{U}_1 = T^{-2}\mathbf{U}_1'\mathbf{U}_1 - T^{-1}\cdot T^{-1}\mathbf{U}_1'\mathbf{M}_p(T^{-1}\mathbf{M}_p'\mathbf{M}_p)^{-1}T^{-1}\mathbf{M}_p'\mathbf{U}_1$$
$$= T^{-2}\mathbf{U}_1'\mathbf{U}_1 + o_p(1) \Rightarrow \ell_{11}^2\int_0^1 \mathbf{1}\{W^* \leq \tilde{\lambda}_1\}W^{*2}. \tag{A.11}$$

Similarly, it can be shown that

$$T^{-1}\mathbf{U}_2'\mathbf{Q}_p\boldsymbol{\epsilon} = T^{-1}\mathbf{U}_2'\boldsymbol{\epsilon} + o_p(1) \Rightarrow D(1)\ell_{11}^2\int_0^1 \mathbf{1}\{W^* > \tilde{\lambda}_2\}W^*dW^* \tag{A.12}$$

and

$$T^{-2}\mathbf{U}_2'\mathbf{Q}_p\mathbf{U}_2 = T^{-2}\mathbf{U}_2'\mathbf{U}_2 + o_p(1) \Rightarrow \ell_{11}^2\int_0^1 \mathbf{1}\{W^* > \tilde{\lambda}_2\}W^{*2}. \tag{A.13}$$

Next, we consider the variance $\hat{\sigma}^2$. Note that $\hat{\rho}_1 = O_p(T^{-1})$, $\hat{\rho}_2 = O_p(T^{-1})$, and $(\hat{\alpha}_j - \alpha_j) = O_p(T^{-1/2})$. Then, using Lemma 2.2 of Phillips and Ouliaris (1990), we obtain

$$\begin{aligned}\hat{\sigma}^2 &= T^{-1}\sum_{t=1}^{T}\left(\Delta\hat{u}_t - \hat{\rho}_1\hat{u}_{t-1}\mathbf{1}\{\hat{u}_{t-1}\leq\lambda_1\} - \hat{\rho}_2\hat{u}_{t-1}\mathbf{1}\{\hat{u}_{t-1}>\lambda_2\} - \textstyle\sum_{j=1}^{p}(\hat{\alpha}_j-\alpha_j)\Delta\hat{u}_{t-j}\right)^2\\ &= T^{-1}\sum_{t=1}^{T}\epsilon_t^2 + o_p(1) = T^{-1}\sum_{t=1}^{T}D(L)^2\hat{\eta}'\xi_t'\xi_t\hat{\eta}'\\ &\Rightarrow D(1)^2\eta'\Omega\eta = D(1)^2\ell_{11}^2k'k.\end{aligned} \tag{A.14}$$

Therefore, by (A.10) − (A.14),

$$\begin{aligned}W_T(\lambda) &= \frac{1}{\hat{\sigma}^2}\{T^{-1}\boldsymbol{\epsilon}'\mathbf{U}_1(T^{-2}\mathbf{U}_1'\mathbf{U}_1)^{-1}T^{-1}\mathbf{U}_1'\boldsymbol{\epsilon} + T^{-1}\boldsymbol{\epsilon}'\mathbf{U}_2(T^{-2}\mathbf{U}_2'\mathbf{U}_2)^{-1}T^{-1}\mathbf{U}_2'\boldsymbol{\epsilon}\} + o_p(1)\\ &\Rightarrow \frac{1}{D(1)^2\ell_{11}^2k'k}\left[\frac{\left(D(1)\ell_{11}^2\int_0^1\mathbf{1}\{W^*\leq\tilde{\lambda}_1\}W^*dW^*\right)^2}{\ell_{11}^2\int_0^1\mathbf{1}\{W^*\leq\tilde{\lambda}_1\}W^{*2}} + \frac{\left(D(1)\ell_{11}^2\int_0^1\mathbf{1}\{W^*>\tilde{\lambda}_2\}W^*dW^*\right)^2}{\ell_{11}^2\int_0^1\mathbf{1}\{W^*>\tilde{\lambda}_2\}W^{*2}}\right]\\ &= \frac{\left(\int_0^1\mathbf{1}\{W^*\leq\tilde{\lambda}_1\}W^*dW^*\right)^2}{(k'k)\int_0^1\mathbf{1}\{W^*\leq\tilde{\lambda}_1\}W^{*2}} + \frac{\left(\int_0^1\mathbf{1}\{W^*>\tilde{\lambda}_2\}W^*dW^*\right)^2}{(k'k)\int_0^1\mathbf{1}\{W^*>\tilde{\lambda}_2\}W^{*2}}.\end{aligned} \tag{A.15}$$

We can deduce the required results from the asymptotic distribution of $W_T(\lambda)$ and Assumption 2.

Under the alternative hypothesis, $W_T(\lambda)$ can be written as

$$W_T(\lambda) = \frac{1}{\hat{\sigma}^2}\rho'(\mathbf{U}'\mathbf{Q}_p\mathbf{U})\rho + \frac{1}{\hat{\sigma}^2}\boldsymbol{\epsilon}'\mathbf{Q}_p\mathbf{U}(\mathbf{U}'\mathbf{Q}_p\mathbf{U})^{-1}\mathbf{U}'\mathbf{Q}_p\boldsymbol{\epsilon}. \tag{A.16}$$

Note that under the alternative hypothesis, $T^{-1}\mathbf{U}'\mathbf{Q}_p\mathbf{U} = O_p(1)$, $T^{-1/2}\mathbf{U}'\mathbf{Q}_p\boldsymbol{\epsilon} = O_p(1)$, and $\hat{\sigma}^2 = O_p(1)$. We can see that

$$\begin{aligned} W_T(\lambda) &= T\frac{1}{\hat{\sigma}^2}\rho'(T^{-1}\mathbf{U}'\mathbf{Q}_p\mathbf{U})\rho + \frac{1}{\hat{\sigma}^2}T^{-1/2}\boldsymbol{\epsilon}'\mathbf{Q}_p\mathbf{U}(T^{-1}\mathbf{U}'\mathbf{Q}_p\mathbf{U})^{-1}T^{-1/2}\mathbf{U}'\mathbf{Q}_p\boldsymbol{\epsilon} \\ &= T\frac{1}{\hat{\sigma}^2}\rho'(T^{-1}\mathbf{U}'\mathbf{Q}_p\mathbf{U})\rho + O_p(1) \\ &= O_p(T). \end{aligned} \tag{A.17}$$

Therefore, $W_T(\lambda)$ diverges to infinity as $T \to \infty$. This also implies that the test statistic diverges to infinity as $T \to \infty$.

Part of (2b). From the proof of part (2a), the statistic $t_T(\lambda)_{\max}$ can be written as

$$t_T(\lambda)_{\max} = \max\left[\frac{\hat{\rho}_1}{\{\hat{\sigma}^2(\mathbf{U}_1'\mathbf{Q}_p\mathbf{U}_1)^{-1}\}^{1/2}}, \frac{\hat{\rho}_2}{\{\hat{\sigma}^2(\mathbf{U}_2'\mathbf{Q}_p\mathbf{U}_2)^{-1}\}^{1/2}}\right]. \tag{A.18}$$

Since $\hat{\rho}_i$ $(i = 1, 2)$ is given by $\hat{\rho}_i = (\mathbf{U}_i'\mathbf{Q}_p\mathbf{U}_i)^{-1}\mathbf{U}_i'\mathbf{Q}_p\boldsymbol{\epsilon}$ under H_0 with $\rho_1 = \rho_2 = 0$, $t_T(\lambda)_{\max}$ is expressed as

$$t_T(\lambda)_{\max} = \max\left[\frac{\mathbf{U}_1'\mathbf{Q}_p\boldsymbol{\epsilon}}{\{\hat{\sigma}^2(\mathbf{U}_1'\mathbf{Q}_p\mathbf{U}_1)\}^{1/2}}, \frac{\mathbf{U}_2'\mathbf{Q}_p\boldsymbol{\epsilon}}{\{\hat{\sigma}^2(\mathbf{U}_2'\mathbf{Q}_p\mathbf{U}_2)\}^{1/2}}\right]. \tag{A.19}$$

Using from (A.9) to (A.13), we have

$$\begin{aligned} t(\lambda)_{\max} \Rightarrow \max&\left[\frac{D(1)\ell_{11}^2\int_0^1 1\{W^* \le \tilde{\lambda}_1\}W^*dW^*}{\left(D(1)^2\ell_{11}^2 k'k\ell_{11}^2\int_0^1 1\{W^* \le \tilde{\lambda}_1\}W^{*2}\right)^{1/2}}, \frac{D(1)\ell_{11}^2\int_0^1 1\{W^* > \tilde{\lambda}_2\}W^*dW^*}{\left(D(1)^2\ell_{11}^2 k'k\ell_{11}^2\int_0^1 1\{W^* > \tilde{\lambda}_2\}W^{*2}\right)^{1/2}}\right] \\ = \max&\left[\frac{\int_0^1 1\{W^* \le \tilde{\lambda}_1\}W^*dW^*}{\left(k'k\int_0^1 1\{W^* \le \tilde{\lambda}_1\}W^{*2}\right)^{1/2}}, \frac{\int_0^1 1\{W^* > \tilde{\lambda}_2\}W^*dW^*}{\left(k'k\int_0^1 1\{W^* > \tilde{\lambda}_2\}W^{*2}\right)^{1/2}}\right]. \end{aligned} \tag{A.20}$$

We can deduce the required results from the asymptotic distribution of $t_T(\lambda)_{\max}$ and Assumption 2. Next, we consider the properties of the test statistic under the alternative hypothesis. Under the alternative hypothesis, $t_T(\lambda)_{\max}$ can be written as

$$\begin{aligned} t_T(\lambda)_{\max} = \max&\left[\frac{\rho_1}{\{\hat{\sigma}^2(\mathbf{U}_1'\mathbf{Q}_p\mathbf{U}_1)^{-1}\}^{1/2}} + \frac{1}{\hat{\sigma}}\mathbf{U}_1'\mathbf{Q}_p\boldsymbol{\epsilon}(\mathbf{U}_1'\mathbf{Q}_p\mathbf{U}_1)^{-1/2},\right. \\ &\left.\frac{\rho_2}{\{\hat{\sigma}^2(\mathbf{U}_2'\mathbf{Q}_p\mathbf{U}_2)^{-1}\}^{1/2}} + \frac{1}{\hat{\sigma}}\mathbf{U}_2'\mathbf{Q}_p\boldsymbol{\epsilon}(\mathbf{U}_2'\mathbf{Q}_p\mathbf{U}_2)^{-1/2}\right]. \end{aligned} \tag{A.21}$$

Then, we have

$$
\begin{aligned}
t_T(\lambda)_{\max} =& \max\left[\frac{T^{1/2}\rho_1}{\{\hat{\sigma}^2(T^{-1}\mathbf{U}_1'\mathbf{Q}_p\mathbf{U}_1)^{-1}\}^{1/2}} + \frac{1}{\hat{\sigma}}T^{-1/2}\mathbf{U}_1'\mathbf{Q}_p\boldsymbol{\epsilon}(T^{-1}\mathbf{U}_1'\mathbf{Q}_p\mathbf{U}_1)^{-1/2},\right. \\
&\left.\frac{T^{1/2}\rho_2}{\{\hat{\sigma}^2(T^{-1}\mathbf{U}_2'\mathbf{Q}_p\mathbf{U}_2)^{-1}\}^{1/2}} + \frac{1}{\hat{\sigma}}T^{-1/2}\mathbf{U}_2'\mathbf{Q}_p\boldsymbol{\epsilon}(T^{-1}\mathbf{U}_2'\mathbf{Q}_p\mathbf{U}_2)^{-1/2}\right] \\
=& \max[O_p(T^{1/2}), O_p(T^{1/2})].
\end{aligned} \tag{A.22}
$$

from $T^{-1}\mathbf{U}_1'\mathbf{Q}_p\mathbf{U}_1 = O_p(1)$, $T^{-1/2}\mathbf{U}_1'\mathbf{Q}_p\boldsymbol{\epsilon} = O_p(1)$, $T^{-1}\mathbf{U}_2'\mathbf{Q}_p\mathbf{U}_2 = O_p(1)$, $T^{-1/2}\mathbf{U}_2'\mathbf{Q}_p\boldsymbol{\epsilon} = O_p(1)$, and $\hat{\sigma}^2 = O_p(1)$. Therefore, $t_T(\lambda)_{\max}$ diverges to minus infinity as $T \to \infty$. This also implies that the test statistic diverges to minus infinity under the alternative hypothesis as $T \to \infty$.

Proof of Theorem 3

Part of (3a).

Under H_0, $\hat{\rho}$ is given by $\hat{\rho} = (\mathbf{U}'\mathbf{S}_p\mathbf{U})^{-1}\mathbf{U}'\mathbf{S}_p\boldsymbol{\epsilon}$, where $\mathbf{S}_p = \mathbf{I} - \mathbf{N}_p(\mathbf{N}_p'\mathbf{N}_p)^{-1}\mathbf{N}_p'$ is a $T \times T$ idempotent matrix and

$$
\mathbf{N}_p = \begin{pmatrix} \mathbf{1}\{\hat{u}_0 \leq \lambda_1\} & \mathbf{1}\{\hat{u}_0 > \lambda_2\} & \Delta\hat{u}_{1-p}^p \\ \mathbf{1}\{\hat{u}_1 \leq \lambda_1\} & \mathbf{1}\{\hat{u}_1 > \lambda_2\} & \Delta\hat{u}_{2-p}^p \\ \vdots & \vdots & \vdots \\ \mathbf{1}\{\hat{u}_{T-1} \leq \lambda_1\} & \mathbf{1}\{\hat{u}_{T-1} > \lambda_2\} & \Delta\hat{u}_{T-p}^p \end{pmatrix}.
$$

Similar to (A.9), we can rewrite $W_T^B(\lambda)$ as

$$
W_T^B(\lambda) = \frac{1}{\hat{\sigma}^2}\{\boldsymbol{\epsilon}'\mathbf{S}_p\mathbf{U}_1(\mathbf{U}_1'\mathbf{S}_p\mathbf{U}_1)^{-1}\mathbf{U}_1'\mathbf{S}_p\boldsymbol{\epsilon} + \boldsymbol{\epsilon}'\mathbf{S}_p\mathbf{U}_2(\mathbf{U}_2'S_p\mathbf{U}_2)^{-1}\mathbf{U}_2'\mathbf{S}_p\boldsymbol{\epsilon}\}. \tag{A.23}
$$

Note that

$$
T^{-1}\mathbf{U}_1'\mathbf{S}_p\boldsymbol{\epsilon} = T^{-1}\mathbf{U}_1'\boldsymbol{\epsilon} - T^{-3/2}\mathbf{U}_1'\mathbf{N}_p(T^{-1}\mathbf{N}_p'\mathbf{N}_p)^{-1}T^{-1/2}\mathbf{N}_p'\boldsymbol{\epsilon}. \tag{A.24}
$$

Consider $(T^{-1}N_pN_p)^{-1}$. Since we obtain $T^{-1}\sum_{t=1}^{T}\mathbf{1}\{\hat{u}_{t-1} \leq \lambda_1\} \Rightarrow \int_0^1 \mathbf{1}\{W(r)^* \leq \tilde{\lambda}_1\}dr$ from Theorem 3.1 of Park and Phillips (2001), it can be shown that

$$
T^{-1}\mathbf{N}_p'\mathbf{N}_p \Rightarrow \begin{pmatrix} \int_0^1 I_1 dr & 0 & \mathbf{0} \\ 0 & \int_0^1 I_2 dr & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \Gamma \end{pmatrix},
$$

where $I_1 = \mathbf{1}\{W(r)^* \leq \tilde{\lambda}_1\}$, $I_2 = \mathbf{1}\{W(r)^* > \tilde{\lambda}_2\}$, and $\Gamma = \lim_{T\to\infty} T^{-1}\sum E(\Delta\hat{u}_{t-p}^{p\prime}\Delta\hat{u}_{t-p}^{p})$. Therefore, we have

$$(T^{-1}\mathbf{N}_p'\mathbf{N}_p)^{-1} \Rightarrow \begin{pmatrix} (\int_0^1 I_1 dr)^{-1} & 0 & \mathbf{0} \\ 0 & (\int_0^1 I_2 dr)^{-1} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \Gamma^{-1} \end{pmatrix}. \tag{A.25}$$

We also have

$$\begin{aligned} T^{-3/2}\mathbf{U}_1'\mathbf{N}_p &= T^{-3/2}(\sum \hat{u}_{t-1}\mathbf{1}\{\hat{u}_{t-1} \leq \lambda_1\}, 0, \sum \hat{u}_{t-1}\mathbf{1}\{\hat{u}_{t-1} \leq \lambda_1\}\Delta\hat{u}_{t-p}^{p}) \\ &\Rightarrow (\ell_{11}\textstyle\int_0^1 W^* I_1, 0, 0), \end{aligned} \tag{A.26}$$

and

$$\begin{aligned} T^{-1/2}\mathbf{N}_p'\boldsymbol{\epsilon} &= T^{-1/2}(\sum \mathbf{1}\{\hat{u}_{t-1} \leq \lambda_1\}\epsilon_t, \sum \mathbf{1}\{\hat{u}_{t-1} > \lambda_2\}\epsilon_t, \sum \Delta\hat{u}_{t-p}^{p}\epsilon_t)' \\ &\Rightarrow (D(1)\ell_{11}\textstyle\int_0^1 I_1 dW^*, D(1)\ell_{11}\int_0^1 I_2 dW^*, O_p(1))'. \end{aligned} \tag{A.27}$$

It follows from Theorem 1, (A.25), (A.26), and (A.27) that

$$\begin{aligned} T^{-1}\mathbf{U}_1'\mathbf{S}_p\boldsymbol{\epsilon} &\Rightarrow D(1)\ell_{11}^2\textstyle\int_0^1 I_1 W^* dW^* - D(1)\ell_{11}^2\int_0^1 W^* I_1(\int_0^1 I_1)^{-1}\int_0^1 I_1 dW^* \\ &= D(1)\ell_{11}^2\{\textstyle\int_0^1 I_1\int_0^1 I_1 W^* dW^* - \int_0^1 W^* I_1\int_0^1 I_1 dW^*\}. \end{aligned} \tag{A.28}$$

Next, consider $\mathbf{U}_1'\mathbf{S}_p\mathbf{U}_1$. It is easily seen that

$$T^{-2}\mathbf{U}_1'\mathbf{S}_p\mathbf{U}_1 = T^{-2}\mathbf{U}_1'\mathbf{U}_1 - T^{-3/2}\mathbf{U}_1'\mathbf{N}_p(T^{-1}\mathbf{N}_p'\mathbf{N}_p)^{-1}T^{-3/2}\mathbf{N}_p'\mathbf{U}_1. \tag{A.29}$$

Using (A.25) and (A.26), we obtain

$$-T^{-3/2}\mathbf{U}_1'\mathbf{N}_p(T^{-1}\mathbf{N}_p'\mathbf{N}_p)^{-1}T^{-3/2}\mathbf{N}_p'\mathbf{U}_1 \Rightarrow \ell_{11}^2(\textstyle\int_0^1 W^* I_1)^2(\int_0^1 I_1)^{-1}. \tag{A.30}$$

Therefore, we have

$$\begin{aligned} T^{-2}\mathbf{U}_1'\mathbf{S}_p\mathbf{U}_1 &\Rightarrow \ell_{11}^2\textstyle\int_0^1 W^{*2} I_1 - \ell_{11}^2(\int_0^1 W^* I_1)^2(\int_0^1 I_1)^{-1} \\ &= \ell_{11}^2\{\textstyle\int_0^1 I_1\int_0^1 W^{*2} I_1 - (\int_0^1 W^* I_1)^2\}. \end{aligned} \tag{A.31}$$

Since similar analysis can be applied to $\boldsymbol{\epsilon}'\mathbf{S}_p\mathbf{U}_2(\mathbf{U}_2' S_p\mathbf{U}_2)^{-1}\mathbf{U}_2'\mathbf{S}_p\boldsymbol{\epsilon}$, it can also be shown that

$$T^{-1}\mathbf{U}_2'\mathbf{S}_p\boldsymbol{\epsilon} \Rightarrow D(1)\ell_{11}^2\{\textstyle\int_0^1 I_2\int_0^1 I_2 W^* dW^* - \int_0^1 W^* I_2\int_0^1 I_2 dW^*\} \tag{A.32}$$

and

$$T^{-2}\mathbf{U}_2'\mathbf{S}_p\mathbf{U}_2 \Rightarrow \ell_{11}^2\{\int_0^1 I_2 \int_0^1 W^{*2} I_2 - (\int_0^1 W^* I_2)^2\}. \tag{A.33}$$

Noting that $(\hat{\mu}_1 - \mu_1) = O_p(T^{-1/2})$ and $(\hat{\mu}_2 - \mu_2) = O_p(T^{-1/2})$, similar to (A.14), we obtain

$$\begin{aligned}\hat{\sigma}^2 &= T^{-1}\sum_{t=1}^{T}\Big\{\Delta\hat{u}_t - (\hat{\mu}_1 - \mu_1 + \hat{\rho}_1)\hat{u}_{t-1}\mathbf{1}\{\hat{u}_{t-1} \le \lambda_1\} - (\hat{\mu}_2 - \mu_2 + \hat{\rho}_2)\hat{u}_{t-1}\mathbf{1}\{\hat{u}_{t-1} > \lambda_2\} \\ &\qquad - \textstyle\sum_{j=1}^{p}(\hat{\alpha}_j - \alpha_j)\Delta\hat{u}_{t-j}\Big\}^2 \\ &= T^{-1}\sum_{t=1}^{T}\epsilon_t^2 + o_p(1) \Rightarrow D(1)^2\ell_{11}^2 k'k.\end{aligned} \tag{A.34}$$

Therefore, by (A.28) − (A.34),

$$\begin{aligned}W_T^B(\lambda) \Rightarrow &\frac{1}{D(1)^2\ell_{11}^2 k'k}\left[\frac{\Big\{D(1)\ell_{11}^2(\int_0^1 I_1 \int_0^1 I_1 W^* dW^* - \int_0^1 W^* I_1 \int_0^1 I_1 dW^*)\Big\}^2}{\ell_{11}^2\{\int_0^1 I_1 \int_0^1 W^{*2} I_1 - (\int_0^1 W^* I_1)^2\}}\right. \\ &\left. + \frac{\Big\{D(1)\ell_{11}^2(\int_0^1 I_2 \int_0^1 I_2 W^* dW^* - \int_0^1 W^* I_2 \int_0^1 I_2 dW^*)\Big\}^2}{\ell_{11}^2\{\int_0^1 I_2 \int_0^1 I_2 W^{*2} dW^* - (\int_0^1 W^* I_2)^2\}}\right] \\ &= \frac{1}{k'k}\frac{\Big(\int_0^1 I_1 \int_0^1 W^* I_1 - \int_0^1 W^* I_1 \int_0^1 I_1 dW^*\Big)^2}{\int_0^1 I_1 \int_0^1 W^{*2} I_1 - (\int_0^1 W^* I_1)^2} \\ &+ \frac{1}{k'k}\frac{\Big(\int_0^1 I_2 \int_0^1 W^* I_2 - \int_0^1 W^* I_2 \int_0^1 I_2 dW^*\Big)^2}{\int_0^1 I_2 \int_0^1 W^{*2} I_2 - (\int_0^1 W^* I_2)^2}.\end{aligned} \tag{A.35}$$

We can deduce the required results from the asymptotic distribution of $W_B^T(\lambda)$ and Assumption 2.

Under the alternative hypothesis, we can write $W_T^B(\lambda)$ as

$$W_T^B(\lambda) = \frac{1}{\hat{\sigma}^2}\rho'(\mathbf{U}'\mathbf{S}_p\mathbf{U})\rho + \frac{1}{\hat{\sigma}^2}\boldsymbol{\epsilon}'\mathbf{S}_p\mathbf{U}(\mathbf{U}'\mathbf{S}_p\mathbf{U})^{-1}\mathbf{U}'\mathbf{S}_p\boldsymbol{\epsilon}. \tag{A.36}$$

Since under the alternative hypothesis, we have $T^{-1}\mathbf{U}'\mathbf{S}_p\mathbf{U} = O_p(1)$, $T^{-1/2}\mathbf{U}'\mathbf{S}_p\boldsymbol{\epsilon} = O_p(1)$, and $\hat{\sigma}^2 = O_p(1)$, it can be shown that

$$\begin{aligned}W_T^B(\lambda) &= T\frac{1}{\hat{\sigma}^2}\rho'(T^{-1}\mathbf{U}'\mathbf{S}_p\mathbf{U})\rho + \frac{1}{\hat{\sigma}^2}T^{-1/2}\boldsymbol{\epsilon}'\mathbf{S}_p\mathbf{U}(T^{-1}\mathbf{U}'\mathbf{S}_p\mathbf{U})^{-1}T^{-1/2}\mathbf{U}'\mathbf{S}_p\boldsymbol{\epsilon} \\ &= T\frac{1}{\hat{\sigma}^2}\rho'(T^{-1}\mathbf{U}'\mathbf{S}_p\mathbf{U})\rho + O_p(1) \\ &= O_p(T).\end{aligned} \tag{A.37}$$

Therefore, $W_T^B(\lambda)$ diverges to infinity as $T \to \infty$. This also implies that the test statistic diverges to infinity as $T \to \infty$.

Part of (3b). From the proof of part (2b) and (3a), under H_0, the statistic $t_T^B(\lambda)_{\max}$ can be written as

$$t_T^B(\lambda)_{\max} = \max\left[\frac{\mathbf{U}_1'\mathbf{S}_p\boldsymbol{\epsilon}}{\{\hat{\sigma}^2(\mathbf{U}_1'\mathbf{S}_p\mathbf{U}_1)\}^{1/2}}, \frac{\mathbf{U}_2'\mathbf{S}_p\boldsymbol{\epsilon}}{\{\hat{\sigma}^2(\mathbf{U}_2'\mathbf{S}_p\mathbf{U}_2)\}^{1/2}}\right]. \tag{A.38}$$

Using from (A.28) to (A.34), we have

$$\begin{aligned} t^B(\lambda)_{\max} = \max & \left[\frac{\int_0^1 I_1 \int_0^1 I_1 W^* dW^* - \int_0^1 W^* I_1 \int_0^1 I_1 dW^*}{\left(k'k\{\int_0^1 I_1 \int_0^1 W^{*2} I_1 - (\int_0^1 W^{*2} I_1)^2\}\right)^{1/2}}, \right. \\ & \left. \frac{\int_0^1 I_2 \int_0^1 I_2 W^* dW^* - \int_0^1 W^* I_2 \int_0^1 I_2 dW^*}{\left(k'k\{\int_0^1 I_2 \int_0^1 W^{*2} I_2 - (\int_0^1 W^{*2} I_2)^2\}\right)^{1/2}}\right] \end{aligned} \tag{A.39}$$

We can deduce the required results from the asymptotic distribution of $t_T^B(\lambda)_{\max}$ and Assumption 2. Next, we consider the properties of the test statistic under the alternative hypothesis. Under the alternative hypothesis, $t_T^B(\lambda)_{\max}$ can be expressed as

$$\begin{aligned} t_T^B(\lambda)_{\max} = \max & \left[\frac{\rho_1}{\{\hat{\sigma}^2(\mathbf{U}_1'\mathbf{S}_p\mathbf{U}_1)^{-1}\}^{1/2}} + \frac{1}{\hat{\sigma}}\mathbf{U}_1'\mathbf{S}_p\boldsymbol{\epsilon}(\mathbf{U}_1'\mathbf{S}_p\mathbf{U}_1)^{-1/2}, \right. \\ & \left. \frac{\rho_2}{\{\hat{\sigma}^2(\mathbf{U}_2'\mathbf{S}_p\mathbf{U}_2)^{-1}\}^{1/2}} + \frac{1}{\hat{\sigma}}\mathbf{U}_2'\mathbf{S}_p\boldsymbol{\epsilon}(\mathbf{U}_2'\mathbf{S}_p\mathbf{U}_2)^{-1/2}\right]. \end{aligned} \tag{A.40}$$

Using $T^{-1}\mathbf{U}_1'\mathbf{S}_p\mathbf{U}_1 = O_p(1)$, $T^{-1/2}\mathbf{U}_1'\mathbf{S}_p\boldsymbol{\epsilon} = O_p(1)$, $T^{-1}\mathbf{U}_2'\mathbf{S}_p\mathbf{U}_2 = O_p(1)$, $T^{-1/2}\mathbf{U}_2'\mathbf{S}_p\boldsymbol{\epsilon} = O_p 1$, and $\hat{\sigma}^2 = O_p(1)$, we obtain

$$\begin{aligned} t_T^B(\lambda)_{\max} = \max & \left[\frac{T^{1/2}\rho_1}{\{\hat{\sigma}^2(T^{-1}\mathbf{U}_1'\mathbf{S}_p\mathbf{U}_1)^{-1}\}^{1/2}} + \frac{1}{\hat{\sigma}}T^{-1/2}\mathbf{U}_1'\mathbf{S}_p\boldsymbol{\epsilon}(T^{-1}\mathbf{U}_1'\mathbf{S}_p\mathbf{U}_1)^{-1/2}, \right. \\ & \left. \frac{T^{1/2}\rho_2}{\{\hat{\sigma}^2(T^{-1}\mathbf{U}_2'\mathbf{S}_p\mathbf{U}_2)^{-1}\}^{1/2}} + \frac{1}{\hat{\sigma}}T^{-1/2}\mathbf{U}_2'\mathbf{S}_p\boldsymbol{\epsilon}(T^{-1}\mathbf{U}_2'\mathbf{S}_p\mathbf{U}_2)^{-1/2}\right] \\ = \max & [O_p(T^{1/2}), O_p(T^{1/2})], \end{aligned} \tag{A.41}$$

which shows that $t_T^B(\lambda)_{\max}$ diverges to minus infinity as $T \to \infty$. This also implies that the test statistic diverges to minus infinity under the alternative hypothesis as $T \to \infty$.

References

Andrews, D. W. K., Ploberger, H., (1994) Optimal tests when a nuisance parameter is present only under the alternative, *Econometrica* 62, 1383-1414.

Balke, N. S., Fomby, T., (1997) Threshold cointegration, *International Economic Review* 38, 627-643.

Bec, F., Ben Salem, M., Carrasco, M., (2004) Tests for unit-root versus threshold specification with an application to the purchasing power parity relationship, *Journal of Business and Economic Statistics* 22, 382-395.

Bec, F., Guay, A., Guerre, E., (2008) Adaptive consistent unit-root tests based on autoregressive threshold model, *Journal of Econometrics* 142, 94-133.

Caner, M., Hansen, B. E., (2001) Threshold autoregression with a unit root, *Econometrica* 69, 1555-1596.

Choi, C. Y., Moh, Y. K., (2007) How useful are tests for unit-root in distinguishing unit-root processes from stationary but nonlinear processes? *Econometrics Journal* 10, 82-112.

Clements, M. P., Galvãro, A. B., (2003) Testing the expectations theory of the term structure of interest rates in threshold models, *Macroeconomic Dynamics* 7, 567-585.

Cuthbertson, K., Taylor, M.P., (1987) The demand for money: A dynamic rational expectation model, *Economic Journal* 97, 65-76.

Davies, R. B., (1987) Hypothesis testing when a nuisance parameter is present under the alternative, *Biometrika*, 74, 33-43.

Enders, W., Granger, C. W. J., (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.

Enders, W., Siklos, P. L. (2001) Cointegration and threshold adjustment, *Journal of Business and Economic Statistics* 19, 166-176.

Engle, R. F., Granger, C. W. J., (1987) Cointegration and error correction: representation, estimation, and testing, *Econometrica* 55, 251-276.

Hansen, B. E., (1996) Inference when a nuisance parameter is not identified under the null hypothesis, *Econometrica* 64, 414-430.

Hansen, B. E., Seo, B., (2002) Testing for two-regime threshold cointegraion in vector error correction models, *Journal of Econometrics* 110, 293-318.

Johansen, S., (1991) Estimation and hypothesis testing of cointegration vectors in Gaussian vector autoregressive models, *Econometrica* 59, 1551-1580.

Kapetanios, G., Shin, Y., (2006) Unit root tests in three-regime SETAR models, *Econometrics Journal* 9, 252-278.

Kapetanios, G., Shin, Y., Snell, A., (2003) Testing for a unit root in the nonlinear STAR framework, *Journal of Econometrics* 112, 359-379.

Kapetanios, G., Shin, Y., Snell, A., (2006) Testing for cointegration in nonlinear smooth transition error correction models, *Econometric Theory* 22, 279-303.

Lo, M. C., Zivot, E., (2001) Threshold cointegration and nonlinear adjustment to the low of one price, *Macroeconomic Dynamics* 5, 533-576.

Maki, D., (2009) Tests for a unit root using three-regime TAR models: Power comparison and some applications, *Econometric Reviews* 28, 335-363.

Maki, D., Kitasaka, S., (2006) The equilibrium relationship among money, income, prices, and interest rates: Evidence from a threshold cointegration test, *Applied Economics* 38, 1585-1592.

Milbourne, R., (1987) Re-examining the Buffer-Stock model of money, *Economic Journal* 97, 130-142.

Mizen, P., (1997) Microfoundations for a stable demand for money function, *Economic Journal* 107, 1202-1212.

Park, J. Y., Phillips, P. C. B., (2001) Nonlinear regressions with integrated time series, *Econometrica* 69, 117-161.

Park, J. Y., Shintani, M., (2005) Testing for a unit root test against transitional autoregressive models, Working paper, Vanderbilt University.

Pippenger, M. K., Goering, G. E., (1993) A note on the empirical power of unit root tests under threshold processes, *Oxford Bulletin of Economics and Statistics* 55, 473-481.

Pippenger, M. K., Goering, G. E., (2001) Additional results on the power of unit root and cointegration tests under threshold processes, *Applied Economics Letters* 7, 641-644.

Phillips, P. C. B., Ouliaris, S. (1990) Asymptotic properties of residual based tests for cointegration, *Econometrica* 58, 165-193.

Sarno, L., (1999) Adjustment costs and nonlinear dynamics in the demand for money: Italy, 1861-1991, *International Journal of Finance and Economics* 4, 155-177.

Sarno L., Taylor, M. P., Peel, D. A., (2003) Nonlinear equilibrium correction in U.S. real money balances, 1869-1997, *Journal of Money, Credit, and Banking* 35, 787-797.

Seo, M., (2006) Bootstrap testing for the null of no cointegration in a threshold vector error correction model, *Journal of Econometrics* 134, 129-150.

Seo, M., (2008) Unit root test in a threshold autoregression: Asymptotic theory and residual-based bootstrap, *Econometric Theory* 24, 1699-1716.

Taylor, A. M., (2001) Potential pitfalls for the purchasing-power-parity puzzle? Sampling and specification biases in mean-reversion tests of the law of one price, *Econometrica* 69, 473-498.

Table 1: Critical values of three-regime TAR models

			Model=0			Model=1			Model=2	
	90%	95%	99%	90%	95%	99%	90%	95%	99%	
$\gamma = 0.05$										
sup-Wald	m=1	14.64	16.78	21.58	16.34	18.54	23.18	18.86	21.31	26.45
	2	16.44	18.76	23.60	18.62	20.94	25.44	20.94	23.54	28.84
	3	18.48	21.04	26.31	20.78	23.44	29.10	23.45	26.44	32.10
	4	20.72	23.30	28.05	23.08	25.99	31.70	25.79	28.62	34.50
	5	23.20	26.02	32.19	25.40	28.21	34.16	28.02	30.99	37.37
max-t	m=1	-2.306	-2.503	-2.904	-2.486	-2.683	-3.076	-2.716	-2.930	-3.345
	2	-2.515	-2.707	-3.081	-2.684	-2.892	-3.269	-2.902	-3.111	-3.515
	3	-2.703	-2.897	-3.300	-2.866	-3.076	-3.483	-3.092	-3.301	-3.737
	4	-2.886	-3.090	-3.473	-3.072	-3.282	-3.692	-3.255	-3.464	-3.897
	5	-3.087	-3.289	-3.739	-3.226	-3.438	-3.857	-3.431	-3.632	-4.014
$\gamma = 0.1$										
sup-Wald	m=1	13.01	15.13	19.16	14.82	17.06	22.05	17.60	20.23	25.44
	2	15.03	17.27	21.87	17.27	19.57	24.60	19.98	22.56	28.21
	3	17.29	19.68	24.41	19.85	22.67	28.27	22.35	25.14	31.08
	4	19.78	22.37	27.69	22.20	24.90	30.97	24.97	27.83	33.92
	5	22.34	24.91	30.61	24.62	27.37	33.41	27.46	30.22	36.42
max-t	m=1	-2.137	-2.345	-2.730	-2.347	-2.546	-2.964	-2.618	-2.820	-3.243
	2	-2.373	-2.589	-2.992	-2.576	-2.790	-3.167	-2.810	-3.030	-3.444
	3	-2.610	-2.801	-3.179	-2.810	-3.031	-3.436	-3.018	-3.225	-3.619
	4	-2.815	-3.018	-3.468	-3.017	-3.215	-3.593	-3.225	-3.429	-3.810
	5	-3.041	-3.231	-3.590	-3.183	-3.384	-3.760	-3.388	-3.600	-3.968
$\gamma = 0.15$										
sup-Wald	m=1	11.58	13.62	17.95	13.95	16.20	20.92	16.80	18.97	23.67
	2	14.05	16.27	20.50	16.42	18.82	23.48	19.01	21.52	27.19
	3	16.38	18.81	23.68	19.02	21.45	26.63	21.80	24.63	29.95
	4	19.15	21.62	26.97	21.58	24.11	29.84	23.95	26.94	32.87
	5	21.47	24.34	29.45	24.02	26.81	32.86	26.64	29.69	36.23
max-t	m=1	-2.022	-2.230	-2.671	-2.274	-2.483	-2.877	-2.533	-2.740	-3.140
	2	-2.297	-2.514	-2.915	-2.518	-2.725	-3.116	-2.754	-2.972	-3.400
	3	-2.523	-2.750	-3.128	-2.743	-2.953	-3.329	-2.961	-3.179	-3.591
	4	-2.778	-2.982	-3.370	-2.956	-3.184	-3.572	-3.150	-3.360	-3.750
	5	-2.979	-3.183	-3.580	-3.138	-3.348	-3.753	-3.330	-3.540	-3.970

Table 2: Critical values of BAND-TAR models

			Model=0			Model=1			Model=2	
	90%	95%	99%	90%	95%	99%	90%	95%	99%	
$\gamma = 0.05$										
sup-Wald	m=1	15.99	18.29	23.09	16.08	18.33	23.00	16.16	18.31	23.02
	2	16.21	18.40	22.98	16.05	18.30	22.73	16.43	18.56	23.43
	3	16.35	18.61	23.18	16.48	18.61	23.40	16.79	19.17	24.16
	4	16.61	19.04	24.30	16.91	19.23	24.18	17.63	20.07	24.97
	5	17.10	19.35	23.86	17.50	19.88	24.67	17.90	20.49	25.31
max-t	m=1	-2.477	-2.667	-3.044	-2.450	-2.657	-3.030	-2.466	-2.670	-3.053
	2	-2.481	-2.688	-3.053	-2.460	-2.661	-3.030	-2.490	-2.694	-3.096
	3	-2.485	-2.682	-3.066	-2.505	-2.702	-3.082	-2.531	-2.712	-3.092
	4	-2.511	-2.716	-3.134	-2.544	-2.740	-3.145	-2.595	-2.807	-3.180
	5	-2.538	-2.747	-3.109	-2.594	-2.785	-3.188	-2.627	-2.819	-3.232
$\gamma = 0.1$										
sup-Wald	m=1	14.49	16.58	21.16	14.33	16.53	21.15	14.18	17.08	22.30
	2	14.54	16.44	21.37	14.53	16.57	21.45	15.19	17.47	21.84
	3	14.57	16.88	22.16	15.04	17.41	21.97	15.52	17.91	22.86
	4	15.33	17.70	22.31	15.51	17.81	22.60	16.00	18.23	22.68
	5	15.49	17.84	22.54	15.96	18.34	23.30	16.36	18.63	23.61
max-t	m=1	-2.282	-2.500	-2.904	-2.270	-2.474	-2.914	-2.317	-2.526	-2.936
	2	-2.296	-2.499	-2.915	-2.282	-2.491	-2.916	-2.356	-2.573	-2.949
	3	-2.304	-2.504	-2.961	-2.345	-2.551	-2.954	-2.372	-2.582	-3.018
	4	-2.352	-2.563	-2.963	-2.392	-2.604	-3.014	-2.419	-2.634	-3.001
	5	-2.384	-2.585	-2.994	-2.418	-2.628	-3.080	-2.474	-2.697	-3.055
$\gamma = 0.15$										
sup-Wald	m=1	13.08	15.38	20.14	12.89	15.08	19.80	13.35	15.39	20.08
	2	13.11	15.39	20.33	13.05	15.26	19.77	13.68	16.05	21.23
	3	13.31	15.50	19.83	13.61	15.80	20.45	13.96	16.29	21.13
	4	13.63	15.95	20.71	13.85	15.95	20.48	14.28	16.47	21.47
	5	13.97	16.20	21.09	14.21	16.64	20.95	15.04	17.50	22.67
max-t	m=1	-2.117	-2.337	-2.787	-2.119	-2.337	-2.758	-2.152	-2.390	-2.798
	2	-2.115	-2.344	-2.792	-2.126	-2.345	-2.760	-2.211	-2.422	-2.858
	3	-2.152	-2.375	-2.798	-2.181	-2.402	-2.866	-2.208	-2.453	-2.872
	4	-2.189	-2.418	-2.815	-2.216	-2.436	-2.833	-2.260	-2.483	-2.903
	5	-2.210	-2.456	-2.890	-2.239	-2.456	-2.895	-2.313	-2.543	-2.969

Table 3: Size performance

	EG-t	ES-Φ	$W1$	$t1$	$W3$	$t3$	W^B1	t^B1	W^B3	t^B3
$\sigma_2^2 = 1$, $\phi = 0$										
$T = 100$	0.051	0.052	0.050	0.046	0.046	0.043	0.031	0.028	0.037	0.031
200	0.050	0.049	0.046	0.046	0.043	0.046	0.035	0.036	0.046	0.042
400	0.046	0.047	0.050	0.051	0.045	0.050	0.041	0.041	0.046	0.044
$\sigma_2^2 = 1$, $\phi = 0.5$										
$T = 100$	0.049	0.050	0.058	0.056	0.052	0.052	0.048	0.044	0.047	0.044
200	0.047	0.046	0.054	0.055	0.049	0.051	0.044	0.047	0.052	0.048
400	0.045	0.045	0.052	0.058	0.050	0.051	0.052	0.051	0.052	0.049
$\sigma_2^2 = 1$, $\phi = -0.5$										
$T = 100$	0.046	0.043	0.039	0.054	0.046	0.064	0.031	0.032	0.035	0.036
200	0.044	0.046	0.041	0.054	0.043	0.052	0.032	0.037	0.041	0.043
400	0.041	0.042	0.045	0.053	0.049	0.052	0.039	0.044	0.049	0.051
$\sigma_2^2 = 4$, $\phi = 0$										
$T = 100$	0.052	0.051	0.051	0.046	0.045	0.042	0.032	0.030	0.036	0.032
200	0.050	0.048	0.045	0.047	0.044	0.043	0.033	0.031	0.044	0.042
400	0.047	0.048	0.047	0.049	0.044	0.046	0.041	0.043	0.046	0.047

Table 4: Power of cointegration tests, $\rho_1 = \rho_2 = -0.05$

	EG-t	ES-Φ	$W1$	$t1$	$W3$	$t3$	W^B1	t^B1	W^B3	t^B3	Middle regime%
$\lambda = 0$											
$T = 100$	0.098	0.089	0.071	0.084	0.094	0.097	0.055	0.058	0.054	0.054	0
			(63.75)		(53.19)		(52.52)		(51.11)		
200	0.218	0.212	0.147	0.158	0.182	0.187	0.067	0.061	0.062	0.063	0
			(61.23)		(51.83)		(51.25)		(49.75)		
400	0.660	0.655	0.444	0.466	0.553	0.559	0.092	0.095	0.086	0.085	0
			(56.16)		(50.24)		(46.90)		(48.26)		
$\lambda = 2$											
$T = 100$	0.095	0.088	0.073	0.090	0.084	0.092	0.056	0.058	0.053	0.056	43.85
			(63.60)		(53.42)		(52.48)		(50.81)		
200	0.194	0.179	0.148	0.161	0.166	0.176	0.071	0.063	0.066	0.065	43.94
			(61.80)		(52.47)		(51.18)		(49.55)		
400	0.608	0.614	0.447	0.452	0.540	0.542	0.105	0.106	0.107	0.101	43.67
			(58.24)		(51.27)		(45.12)		(47.57)		
$\lambda = 4$											
$T = 100$	0.088	0.083	0.073	0.082	0.087	0.088	0.059	0.063	0.060	0.061	68.75
			(64.15)		(53.76)		(52.76)		(50.76)		
200	0.139	0.131	0.127	0.133	0.138	0.133	0.083	0.070	0.082	0.079	68.67
			(64.36)		(53.62)		(51.60)		(49.75)		
400	0.363	0.371	0.393	0.367	0.435	0.397	0.162	0.148	0.155	0.157	68.80
			(63.49)		(53.88)		(45.65)		(46.97)		
$\lambda = 8$											
$T = 100$	0.092	0.085	0.074	0.080	0.087	0.085	0.067	0.065	0.067	0.061	87.98
			(65.47)		(54.03)		(53.01)		(50.91)		
200	0.118	0.116	0.119	0.103	0.134	0.115	0.098	0.081	0.098	0.084	87.69
			(66.66)		(54.45)		(52.62)		(49.85)		
400	0.156	0.172	0.256	0.189	0.224	0.168	0.183	0.148	0.180	0.137	87.86
			(67.92)		(54.96)		(51.50)		(49.36)		

Middle regime% denotes the percentage of observations in the middle regime. For sup tests, the percentage of the middle regime used in each test is given with in parentheses.

Table 5: Power of cointegration tests, $\rho_1 = \rho_2 = -0.3$

	EG-t	ES-Φ	$W1$	$t1$	$W3$	$t3$	W^B1	t^B1	W^B3	t^B3	Middle regime%
$\lambda = 0$											
$T = 100$	0.975	0.970	0.845	0.833	0.914	0.887	0.185	0.168	0.145	0.133	0
			(49.32)		(48.32)		(37.91)		(46.39)		
200	1	1	1	0.999	0.999	0.999	0.443	0.370	0.330	0.295	0
			(41.45)		(45.49)		(29.19)		(42.84)		
400	1	1	1	1	1	1	0.870	0.802	0.728	0.650	0
			(34.03)		(41.82)		(20.92)		(38.95)		
$\lambda = 2$											
$T = 100$	0.775	0.755	0.793	0.693	0.846	0.716	0.461	0.388	0.431	0.372	74.55
			(65.80)		(56.48)		(39.13)		(46.05)		
200	1	1	0.999	0.993	1	0.995	0.897	0.790	0.853	0.773	74.60
			(68.84)		(59.28)		(32.31)		(42.65)		
400	1	1	1	1	1	1	0.999	0.993	0.999	0.990	74.61
			(71.42)		(62.57)		(28.40)		(39.98)		
$\lambda = 4$											
$T = 100$	0.243	0.248	0.495	0.312	0.376	0.225	0.477	0.306	0.445	0.272	90.83
			(71.19)		(56.91)		(51.16)		(50.85)		
200	0.678	0.678	0.950	0.728	0.865	0.597	0.846	0.629	0.814	0.585	90.90
			(75.75)		(58.47)		(50.98)		(50.44)		
400	1	1	1	0.991	1	0.984	0.996	0.945	0.993	0.922	90.89
			(81.27)		(61.26)		(52.11)		(51.62)		
$\lambda = 8$											
$T = 100$	0.143	0.166	0.311	0.125	0.217	0.109	0.398	0.148	0.350	0.125	97.38
			(70.13)		(55.99)		(56.10)		(51.98)		
200	0.184	0.205	0.463	0.188	0.289	0.136	0.607	0.262	0.521	0.212	97.34
			(73.00)		(56.38)		(60.19)		(52.78)		
400	0.292	0.350	0.786	0.422	0.541	0.287	0.856	0.541	0.791	0.438	97.35
			(75.79)		(57.78)		(64.30)		(54.08)		

Table 6: Power of cointegration tests, (ρ_1,ρ_2)=(-0.05,-0.15)

	EG-t	ES-Φ	$W1$	$t1$	$W3$	$t3$	W^B1	t^B1	W^B3	t^B3	Middle regime%
$\lambda = 0$											
$T = 100$	0.193	0.175	0.132	0.148	0.159	0.165	0.075	0.075	0.065	0.063	0
			(62.14)		(53.39)		(48.57)		(49.36)		
200	0.488	0.475	0.356	0.332	0.428	0.394	0.112	0.096	0.096	0.089	0
			(59.54)		(52.34)		(41.58)		(46.84)		
400	0.959	0.970	0.891	0.767	0.939	0.829	0.228	0.174	0.200	0.166	0
			(56.79)		(51.55)		(32.78)		(43.23)		
$\lambda = 2$											
$T = 100$	0.151	0.148	0.124	0.137	0.149	0.147	0.087	0.080	0.077	0.072	51.85
			(63.72)		(54.03)		(47.60)		(49.45)		
200	0.393	0.379	0.353	0.318	0.393	0.352	0.161	0.117	0.140	0.120	51.62
			(64.82)		(54.84)		(41.97)		(46.70)		
400	0.932	0.940	0.889	0.739	0.925	0.805	0.378	0.244	0.340	0.214	51.67
			(66.12)		(56.49)		(34.74)		(43.26)		
$\lambda = 4$											
$T = 100$	0.126	0.121	0.124	0.113	0.134	0.117	0.111	0.092	0.097	0.084	75.75
			(66.50)		(54.82)		(50.01)		(50.18)		
200	0.202	0.206	0.289	0.211	0.272	0.210	0.223	0.145	0.204	0.147	75.67
			(69.25)		(55.92)		(48.24)		(49.07)		
400	0.651	0.669	0.787	0.611	0.767	0.629	0.486	0.314	0.463	0.293	75.62
			(73.28)		(57.85)		(45.76)		(48.26)		
$\lambda = 8$											
$T = 100$	0.111	0.109	0.128	0.095	0.124	0.100	0.131	0.086	0.121	0.079	91.40
			(67.78)		(55.37)		(52.73)		(51.05)		
200	0.140	0.134	0.208	0.124	0.172	0.122	0.212	0.111	0.192	0.110	91.36
			(70.09)		(55.76)		(55.50)		(50.79)		
400	0.184	0.201	0.406	0.241	0.286	0.185	0.390	0.217	0.354	0.193	91.36
			(72.78)		(56.51)		(57.27)		(51.31)		

Table 7: Power of cointegration tests, BAND-TAR model with $\rho_1 = \rho_2 = -0.3$

	EG-t	ES-Φ	$W1$	$t1$	$W3$	$t3$	W^B1	t^B1	W^B3	t^B3	Middle regime%
$\lambda = 2$											
$T = 100$	0.226	0.207	0.216	0.218	0.252	0.225	0.163	0.146	0.152	0.140	56.63
			(64.20)		(54.12)		(46.80)		(48.88)		
200	0.738	0.711	0.758	0.653	0.798	0.665	0.400	0.326	0.374	0.334	56.64
			(65.80)		(55.24)		(41.05)		(46.43)		
400	1	1	1	0.993	1	0.994	0.858	0.765	0.823	0.749	56.62
			(67.43)		(57.41)		(34.21)		(43.49)		
$\lambda = 4$											
$T = 100$	0.135	0.129	0.121	0.119	0.131	0.115	0.114	0.101	0.095	0.085	72.43
			(65.37)		(54.55)		(50.95)		(50.36)		
200	0.196	0.194	0.298	0.230	0.256	0.194	0.230	0.177	0.213	0.173	72.40
			(67.69)		(55.05)		(48.95)		(48.90)		
400	0.590	0.623	0.842	0.663	0.807	0.616	0.567	0.449	0.555	0.451	72.32
			(70.58)		(56.80)		(45.34)		(47.74)		
$\lambda = 8$											
$T = 100$	0.100	0.093	0.087	0.088	0.098	0.092	0.087	0.076	0.073	0.070	83.83
			(65.59)		(54.50)		(52.69)		(50.78)		
200	0.120	0.125	0.147	0.122	0.149	0.112	0.128	0.094	0.122	0.094	84.02
			(67.18)		(54.59)		(53.45)		(50.25)		
400	0.165	0.176	0.285	0.186	0.232	0.158	0.245	0.169	0.240	0.151	84.00
			(69.44)		(55.43)		(53.13)		(50.24)		

Table 8: Empirical results for money demand

	EG-t	ES-Φ	$W1$	$t1$	$W3$	$t3$	W^B1	t^B1	W^B3	t^B3
Double-log model										
1960:1-2008:1	5%	5%	5%	10%	5%	5%	-	-	-	-
1979:10-2008:1	10%	10%	10%	5%	-	-	10%	10%	-	-
Semi-log model										
1960:1-2008:1	-	-	-	-	-	-	1%	-	-	-
1979:10-2008:1	-	-	1%	-	-	-	1%	5%	1%	5%

1%, 5%, and 10% indicate that the null hypothesis of no cointegration is rejected at the 1%, 5%, and 10% significance levels, respectively.

Table 9: TAR estimates

	$\hat{\rho}_1$	$\hat{\rho}_2$	$\hat{\lambda}_1$	$\hat{\lambda}_2$	$\hat{\sigma}$	Middle regime%
Double-log model						
1960:1-2008:1	-0.077(0.019)	-0.036(0.012)	-0.012	0.009	0.012	14.55
1979:10-2008:1	-0.087(0.029)	-0.085(0.026)	-0.044	0.055	0.008	84.86
Semi-log model						
1960:1-2008:1	-0.042(0.027)	-0.099(0.026)	-0.103	0.122	0.016	90.12
1979:10-2008:1	-0.050(0.024)	-0.165(0.034)	-0.048	0.054	0.009	86.64

The standard errors are given within parentheses. $\hat{\sigma}$ shows the estimate of the standard deviation of the error term. Middle regime % denotes the percentage of observations in the middle regime.

Figure 1: Residuals of the double-log model($- - -$: the estimated thresholds)

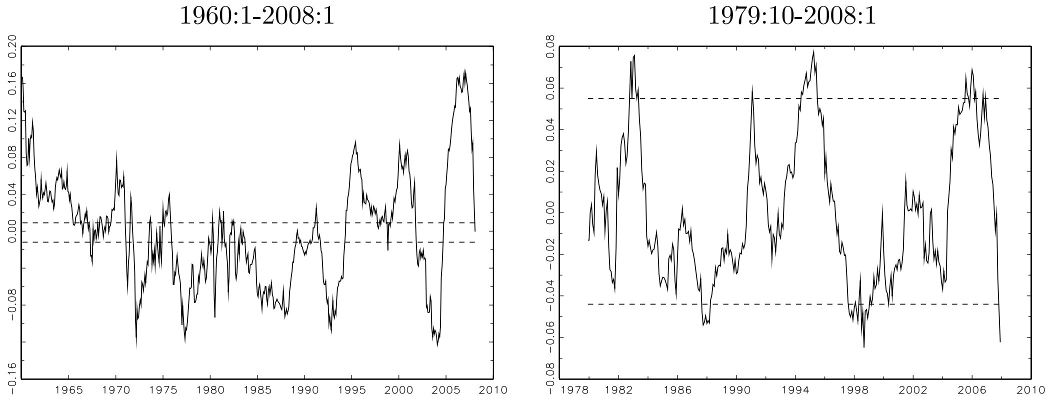

Figure 2: Residuals of the semi-log model($- - -$: the estimated thresholds)