System-Equation ADL Tests for Threshold Cointegration

Jing Li *

Department of Economics
South Dakota State University.

Abstract

In this paper we develop new system-equation tests for threshold cointegration based on a threshold vector autoregressive distributed lag (ADL) model. The proposed tests do not require weak exogeneity, and can be applied when cointegrating vector is unknown. The asymptotic null distributions of the tests are expressed as functionals of two-parameter Brownian motion. The distributions are free of nuisance parameters, and critical values are tabulated. Monte Carlo simulations show good finite-sample performances. The new tests are illustrated with long term and short term interest rates. We show that the system-equation model can accommodate more types of asymmetry than the single-equation model.

Keywords: Threshold Cointegration; Cointegration; Autoregressive Distributed Lag Model.

JEL Classification: C22, C12, C13

*Department of Economics, Scobey Hall, Box 504, South Dakota State University, Brookings, SD 57007. Phone: 605.688.4848, Fax: 605.688.6386, Email: Jing.Li@sdstate.edu.
Introduction

Some economic models imply that variables are cointegrated only within certain regime. For instance, according to the modified law of one price, spatial price linkage may occur only in the regime where price differential exceeds transaction cost. Outside that regime arbitrage becomes non-profitable, so the price linkage breaks\(^1\). This type of discontinuous adjustment toward equilibrium is the key idea of the threshold cointegration introduced by Balke and Fomby (1997). In practice interests are often on testing threshold cointegration, namely, the existence of long run relationship with regime-dependent adjustment speeds. This paper makes contributions to the literature by proposing a new system-equation ADL test for threshold cointegration that improves the existing tests as follows.

Enders and Siklos (2001) (ES) first propose a single-equation Engle-Granger type test for threshold cointegration. The ES test assumes the number of cointegration relationships is at most one. This assumption is questionable whenever more than two variables are in question. By contrast the proposed test is based on a system-equation model, so can be used when multiple cointegration relationships possibly exist. We will show that the system-equation model is more informative than the single-equation framework by accommodating more types of asymmetry.

Next Seo (2006) suggests a system-equation test using a threshold vector error correction model (ECM). The ECM test requires the cointegrating vector is known. In fact the cointegrating vector is typically unknown since many economic theories are not numerically specific. The ADL test, nevertheless, can be applied with unknown cointegrating vectors. The ADL test is more general because it is constructed in a generalized autoregressive distributed lag model in which the lagged regressand, rather than the error correction term, appears on the right hand side.

The ADL test employs a more efficient grid search for the threshold parameter than the ECM test. The ECM test searches the nonstationary (under the null) error correction term over a range of fixed values. As a result the threshold parameter is absent in the limit distribution reported by Theorem 2 of Seo (2006). This fact implies that the information contained in grid search is asymptotically unexploited by the ECM test. The ADL test instead grid searches the difference of the error correction term, which is stationary under the null. This approach is initially introduced by Enders and Granger (1998), and is called the momentum model therein. Intuitively it makes more sense to grid search a mean-reverting

\(^1\)More discussions about the transaction cost and regime-switching can be found in Jang et al. (2007), Sephton (2003) and Giovanini (1988), among others.
stationary series rather than a diverging nonstationary series. Eventually we can express the limit null distribution of the ADL test as the functional of two-parameter Brownian motion; the threshold parameter is one of them.

The ADL test allows that under the null hypothesis a nonzero drift component is present in the series, an issue largely ignored by previous studies. By adding the trend in the testing regression, the ADL test is able to distinguish difference stationarity from trend stationarity; the ECM test is not. This improvement is important as many economic series are trending. Following Sims et al. (1990), we develop the asymptotic theory based on the canonical form of the trend-augmented model.

A single-equation ADL test is considered in Li and Lee (2010). The single-equation test is easier to compute than the system-equation test. However the validity of the single-equation test relies on the assumption of weak exogeneity, i.e., only one variable is error-correcting; others are not. See Boswijk (1994) for more discussion about weak exogeneity. Weak exogeneity becomes disputable for a high-dimensional economic system. The proposed system-equation ADL test is more robust than the single-equation test by assuming away weak exogeneity.

Finally, the proposed test extends the linear ADL cointegration test of Banerjee et al. (1986). The linear test assumes constant error-correcting speed. Therefore the linear test may lose power in the presence of regime switching, c.f., Pippenger and Goering (1993). By contrast, allowing for regime-varying adjustment speeds enables the proposed test to detect cointegration among threshold series.

Testing for threshold cointegration is nonstandard because the unknown threshold parameter cannot be identified under the null hypothesis. Following Hansen (1996) we adopt the sup Wald statistic to address the so called Davies’ problem, c.f., Davies (1977, 1987). The limit null distributions of the proposed tests are expressed as functions of two-parameter Brownian motion. These distributions generalize those in Horvath and Watson (1995) to threshold processes, and generalize Caner and Hansen (2001) to multivariate cases. We conduct Monte Carlo simulations to compare the finite-sample performances of our tests and existing tests. Particular interest is on the size distortion as it determines whether the bootstrap procedure is needed.

Throughout the paper 1(.) denotes the indicator function, \(|.|\) the Euclidean norm, \([x]\) the integer part of \(x\), \(\Rightarrow\) the weak convergence with respect to the uniform metric on \([0,1]^2\). All mathematical proofs are in the appendix.
Threshold ADL Model

Consider a two-regime threshold vector autoregressive distributed lag model (TVADLM)

$$\alpha(L)\Delta y_t = \delta_1 y_{t-1} 1_{1t-1} + \delta_2 y_{t-1} 1_{2t-1} + d_t + e_t,$$

where $y_t$ is an $(n \times 1)$ data vector from a sample of size $T$, $d_t$ represents the deterministic term, and $e_t$ is an i.i.d $(n \times 1)$ vector of error terms. $\alpha(L) \equiv I_n - \sum_{j=1}^{p} \alpha_j L^j$ denotes a $p$-th order matrix polynomial in the lag operator $L$. Economic factors such as transaction cost, policy intervention and inertia can result in regime switching. To account for that, the indicators are defined as

$$1_{1t-1} \equiv 1(z_{t-1} < \tau), \quad 1_{2t-1} \equiv 1(z_{t-1} \geq \tau).$$

In words, the value of the lagged threshold variable, $z_{t-1}$, is less than the threshold parameter, $\tau$, in regime one, while greater than or equal to $\tau$ in regime two. We assume

**Assumption 1**  
(a) $z_{t-1}$ has a marginal uniform distribution, $z_{t-1} \sim U(0,1)$.  
(b) $e_t$ is i.i.d, $Ee_t = 0$, $Ee_t e_t' = \Omega$, and $E|e_t|^4 < \infty$.  
(c) All roots of $|\alpha(L)| = 0$ are outside the unit circle.

Assumption 1-(c) is typically imposed for linear models. Following Caner and Hansen (2001) Assumption 1-(a) and (b) are made so that the following weak convergence holds. Under Assumption 1, as $T \to \infty$,

$$T^{-1/2} \sum_{t=1}^{[Tr]} 1_{1t-1} e_t \Rightarrow PW(r, \tau),$$

where $W(r, \tau)$ denotes the multivariate two-parameter Brownian motion (TPBM), and $P$ is the Cholesky factor of $\Omega$, i.e., $\Omega = PP'$. Basically (3) extends Theorem 1 of Caner and Hansen (2001) to the multivariate case; see the proof in the appendix. Notice that TPBM involves the threshold parameter $\tau$. Hence $\tau$ is asymptotically relevant for our theory, which is not the case for Seo (2006). Let $\gamma$ denote the $(n \times q)$ matrix of cointegrating vectors and

$$x_{t-1} \equiv \gamma' y_{t-1}$$

Alternative specifications are possible. For example, $z_{t-1}$ can be replaced with $z_{t-d}$. A three-regime model can be specified with $1_{1t-1} = 1(z_{t-1} < \tau_1), 1_{2t-1} = 1(\tau_2 \leq z_{t-1} \leq \tau_2)$ and $1_{3t-1} = 1(z_{t-1} > \tau_2)$. This paper focuses on the basic two-regime model (1).
the lagged error correction term. The threshold cointegration literature usually lets the regime-switching be governed by \( x_{t-1} \), i.e., the previous deviation from equilibrium. In light of Assumption 1-(a) we define the threshold variable as

\[
    z_{t-1} \equiv \text{cdf}(\gamma' \Delta x_{t-1})
\]  

where \( \text{cdf}(\cdot) \) denotes the empirical distribution function\(^3\) and \( \gamma \) is a \((q \times 1)\) vector. Notice that under the null hypothesis of no threshold cointegration \( x_{t-1} \) is integrated of order one, but \( \Delta x_{t-1} \) is stationary. It follows that \( \text{cdf}(\gamma' \Delta x_{t-1}) \sim U(0, 1) \), so Assumption 1-(a) holds. Grid searching \( \Delta x_{t-1} \) is first suggested in Enders and Granger (1998).

The reason \( \Delta x_{t-1} \) is pre-multiplied by \( \gamma' \) in (5) is dimension reduction. Note the number \( q \) (or rank of \( \gamma \)) determines the number of cointegration relationships. Given that model (1) is multivariate, this paper permits \( q > 1 \). But then it becomes unclear how to conduct grid search based on a multidimensional \( \Delta x_{t-1} \). Pre-multiplying \( \Delta x_{t-1} \) by \( \gamma' \), which can be seen as adding weights to the \( q \) error correction terms, is a straightforward way of dimension reduction\(^4\). For example, one may consider equal weights of \( \gamma' = (q^{-1}, \ldots, q^{-1}) \). Our theory does not hinge on the specific form of dimension reduction, as long as Assumption 1-(a) holds. We make the next assumption in order to highlight the link between model (1) and other models.

**Assumption 2** If \( y_t \) in model (1) is cointegrated, then

\[
    \text{rank}(\delta_1) = \text{rank}(\delta_2) = q < n.
\]

Assumption 2 is imposed in Seo (2006) as well, and has two implications. First, the number of cointegration relationships (if exist) is known\(^5\). Second, under Assumption 2 we can always write

\[
    \delta_1 = \beta_1 \gamma', \quad \delta_2 = \beta_2 \gamma',
\]

where \( \beta_1 \) and \( \beta_2 \) are \((n \times q)\) matrices. Factorization (7) generalizes the one used to derive the traditional linear ECM. By virtue of (4) and (7), model (1) can be rewritten as a two-regime

---

\(^3\)The empirical distribution function is non-decreasing, and is a step function that jumps for \( 1/N \) at each of the \( N \) data points.

\(^4\)Other methods such as extracting the first principal component of \( \Delta x_{t-1} \) can be used.

\(^5\)The issue of testing \( q \) against \( q + 1 \) threshold cointegration relationships is left to future research.
threshold vector error correction model (TVECM)

\[ \alpha(L) \Delta y_t = \beta_1 x_{t-1} 1_{1t-1} + \beta_2 x_{t-1} 1_{2t-1} + d_t + e_t. \]  

(8)

In model (8) the regime-dependent error-correcting speeds are quantified by \( \beta_1 \) and \( \beta_2 \). By allowing for \( \beta_1 \neq \beta_2 \), model (8) extends the traditional linear ECM; with \( \delta_1 \neq \delta_2 \) model (1) extends the linear ADL model. Seo (2006) considers testing \( \beta_1 = \beta_2 = 0 \) in (8), called the ECM test for threshold cointegration. The TVECM used in Seo (2006) is actually a band or three-regime model that specifies the indicator as \( 1_{1t-1} \equiv 1(x_{t-1} < \tau_1) \), \( 1_{2t-1} \equiv 1(x_{t-1} \geq \tau_2) \). Pay attention that Seo uses the nonstationary \( x_{t-1} \) as the threshold variable, and grid searches \( \tau \) over a range of fixed values. Equation (12) of Seo (2006) makes it clear that this way of conducting grid search is inefficient as \( \tau \) becomes asymptotically irrelevant. In fact it makes more sense to let a range of fixed values be the searching range for a stationary series, a motivation for (5).

Because (8) uses the error correction term times indicators as regressors, the ECM test requires that the cointegrating vector, \( \gamma \), be pre-determined. This requirement results in two limitations. First, the ECM test cannot be used when \( \gamma \) is unknown. Second, misspecification of \( \gamma \) will adversely affect the ECM test through both the indicator and the error correction term. In comparison, a poorly-estimated \( \gamma \) affects the ADL model only through the indicator. In this regard TVADLM (1) is more misspecification-robust than TVECM (8).

Model (8) is a system-equation model. The test of Enders and Siklos (2001), nevertheless, is a single-equation test for \( \beta_1^* = \beta_2^* = 0 \) based on the regression

\[ \alpha^*(L) \Delta x_t = \beta_1^* x_{t-1} 1_{1t-1} + \beta_2^* x_{t-1} 1_{2t-1} + e_t^*. \]  

(9)

In order to derive the single-equation model, the ES test assumes \( q \leq 1 \), i.e., only one cointegration relationship possibly exists. Regression (9) essentially replaces the regressand \( \Delta y_t \) in the first equation of (8) with \( \Delta x_t \). To justify this replacement the ES test imposes the restriction that the short-run dynamics of \( \Delta y_t \) and \( \Delta x_t \) involve one common factor. The same restriction is found by Kremers et al. (1992) to be imposed by the Engle-Granger test. The ADL test is based on (1) and therefore does not impose the common factor restriction.

It is instructive to stress that model (1) allows for two types of asymmetry. First, asymmetric adjustment speeds are allowed by letting \( \delta_1 \neq \delta_2 \). Second, asymmetric adjustment roles can be accommodated by the fact that coefficients in one regression exceed those in
other regressions. In comparison the univariate model (9) can characterize only the first type asymmetry, and so is less informative than (1). We will revisit this issue in the application section.

**System-Equation ADL Tests**

This paper proposes a new test for

$$H_0 : \delta_1 = \delta_2 = 0$$

(10)

based on TVADLM (1), called the system-equation ADL test for threshold cointegration. Under Assumption 2, testing (10) is equivalent to testing $\beta_1 = \beta_2 = 0$ in (8). So the ADL test and ECM test are closely related. Li and Lee (2010) show that testing (10) can be undertaken in a single-equation framework if some elements of $\beta_1$ and $\beta_2$ are zeros (weak exogeneity). The system-equation ADL test is more robust than the single-equation ADL test by assuming away weak exogeneity. The alternative hypothesis can be

$$H_1^1: \delta_1 = 0, \delta_2 < 0; \quad H_1^2: \delta_1 < 0, \delta_2 = 0; \quad H_1^3: \delta_1 < 0, \delta_2 < 0. \quad (11)$$

Strictly speaking $H_1^1$ or $H_1^2$ implies that $y_t$ is cointegrated only in one regime. When (10) is rejected, we conclude in this paper that variables are cointegrated, but not necessarily in both regimes. (10) can also be rejected if $y_t$ is linearly cointegrated, i.e., $\delta_1 = \delta_2 \neq 0$. In this sense the threshold cointegration test extends the linear cointegration test, which may lose power whenever nonlinearity is present, c.f., Pippenger and Goering (1993).

Similar to the Dickey-Fuller unit root test, the distribution of the ADL test depends on the deterministic term, $d_t$, in the testing regression (1). We discuss first whether the testing regression includes the intercept term.

(case I): $d_t = 0$ in the testing regression (1);

(case II): $d_t = c_0 \neq 0$ in the testing regression (1).

Following the literature, we make the auxiliary assumption that

**Assumption 3** Under the null hypothesis (10), $y_t$ is generated by (1) with $d_t = 0$.

Assumption 3 states that the true process under the null is vector random walk without
drift. In next section we will discuss the possibility that a drift term is present.

Testing (10) is nonstandard because under (10) the indicators vanish and \( \tau \) cannot be identified. This is the so called Davies’ problem, c.f., Davies (1977, 1987). To resolve this issue we follow Hansen (1996) and calculate the sup-Wald statistic as follows. First, we estimate the cointegrating vector \( \gamma \) by OLS if it is unknown, compute the error correction term \( 4 \), and determine the threshold variable \( 5 \). Second, for given \( \tau \) we define the indicators \( 2 \), fit \( 1 \) by OLS, and obtain the estimated coefficients \( \hat{\delta} = (\hat{\delta}_1(\tau), \hat{\delta}_2(\tau)) \), the residual \( \hat{\epsilon}_t(\tau) \), and the variance matrix \( \hat{\Omega}_\tau = T^{-1} \sum \hat{\epsilon}_t(\tau)\hat{\epsilon}_t(\tau)' \). Next, we stack matrices by letting \( Y_t = (y_{t-11}, y_{t-12}, \ldots, y_{t-1})' \), \( Y_r = [Y_1Y_2 \ldots Y_T]' \), \( e = [e_1e_2 \ldots e_T]' \). Define the annihilation matrix \( Q \equiv I - Z(Z'Z)^{-1}Z' \) where \( Z \) stacks \( (\Delta y_{t-1}, \ldots, \Delta y_{t-p}, dt) \). Let \( F(\tau) \) and \( F^d(\tau) \) denote the Wald statistics for cases I, and II, respectively. They are given by

\[
F(\tau), F^d(\tau) = [\text{vec}(\hat{\delta}_r)]' \left[ \text{var}(\text{vec}(\hat{\delta}_r))^{-1} \right] [\text{vec}(\hat{\delta}_r)] \quad (12)
\]

\[
= [\text{vec}(\hat{\delta}_r)]' \left( Y_r'QY_r \right)^{-1} \otimes \hat{\Omega}_\tau^{-1} \left[ \text{vec}(\hat{\delta}_r) \right] \quad (13)
\]

\[
= [\text{vec}(e'QY_r)]' \left( (Y_r'QY_r)^{-1} \otimes \hat{\Omega}_\tau^{-1} \right) [\text{vec}(e'QY_r)] \quad (14)
\]

\[
= \text{trace} \left[ \hat{\Omega}_\tau^{-1/2}(e'QY_r)(Y_r'QY_r)^{-1}(Y_r'Qe)\hat{\Omega}_\tau^{-1/2} \right], \quad (15)
\]

where \( \hat{\Omega}_\tau^{-1/2} \) is the Cholesky factor of \( \hat{\Omega}_\tau \). Note that the definition of the Wald statistics is (12), but in practice we calculate the Wald statistics using (13). The asymptotic theory is built upon (15). Finally the sup Wald test\(^6\) is

\[
\sup_{\tau \in \Theta} F(\tau), \sup_{\tau \in \Theta} F^d(\tau), \quad (16)
\]

where \( \Theta \subset [0,1] \) is a compact set. Symmetric sets such as \( \Theta = [0.15, 0.85] \) and \( \Theta = [0.05, 0.95] \) are recommended. Following Andrews (1993) and Hansen (1996) certain upper and lower observations of sorted \( z_{t-1} \) are discarded in order to avoid divergent asymptotic distributions.

Next we derive the asymptotic null distributions of the tests. Let \( W(r, \tau) \) denote the \( n \)-dimensional two-parameter Brownian motion on \( (r, \tau) \in [0,1]^2 \), c.f., Caner and Hansen (2001). When \( \tau = 1 \), \( W(r, 1) \) is the standard one-parameter Brownian motion on \( r \). For notational economy we write \( \int_0^1 \) as \( \int \). The integration is over the first argument of \( W(r, \tau) \) (hold-

\(^6\)This paper focuses on the sup test due to its simplicity. See Andrews and Ploberger (1994) for alternatives.
ing the second argument, \( \tau \), constant) for the stochastic integrations such as \( \int W(r, 1)dW(r, \tau)' \) and \( \int W(r, 1)dW(r, 1)' \). The limit null distributions of the system-equation ADL tests (16) are

**Theorem 1 (case I)** Under (10) and Assumptions 1, 2, 3, as \( T \to \infty \),

\[
\sup_{\tau \in \Theta} F(\tau) \Rightarrow \sup_{\tau \in \Theta} \text{trace} \left[ J_0'J_1^{-1}J_0 \right],
\]

where

\[
J_0 = \begin{pmatrix}
\int W(r, 1)dW(r, \tau)'

\int W(r, 1)dW(r, 1)' - \int W(r, 1)dW(r, \tau)'
\end{pmatrix}, \quad
J_1 = \begin{pmatrix}
\tau \int W(r, 1)W(r, 1)'dr & 0 \\
0 & (1 - \tau) \int W(r, 1)W(r, 1)'dr
\end{pmatrix}.
\]

**Theorem 2 (case II)** Under (10) and Assumptions 1, 2, 3, as \( T \to \infty \),

\[
\sup_{\tau \in \Theta} F^d(\tau) \Rightarrow \sup_{\tau \in \Theta} \text{trace} \left[ J_0^d(J_1^d)^{-1}J_0^d \right],
\]

where

\[
J_0^d = J_0 - \begin{pmatrix}
\tau \left[ \int W(r, 1)dr \right] \left[ \int dW(r, 1)' \right] \\
(1 - \tau) \left[ \int W(r, 1)dr \right] \left[ \int dW(r, 1)' \right]
\end{pmatrix},
\]

\[
J_1^d = J_1 - \begin{pmatrix}
\tau \int W(r, 1)dr \\
(1 - \tau) \int W(r, 1)dr
\end{pmatrix} \begin{pmatrix}
\tau \int W(r, 1)dr \\
(1 - \tau) \int W(r, 1)dr
\end{pmatrix}'.
\]

Notice that the distributions in both theorems depend on \( n \), the dimensional parameter, and \( \Theta \), the searching range. The distributions are free of nuisance parameters such as the Cholesky factor \( P \) and serial correlation characterized by \( \alpha(L) \). Then we can tabulate in Table 1 the critical values of the limit distributions for various \( n \) and \( \Theta \). The critical values are computed as the empirical quantiles from 5,000 independent draws from the simulations for the asymptotic formula in Theorems 1 and 2.

Table 1 shows that the critical value increases as the searching range \( \Theta \) gets wider. This is indicative of the divergent behavior of the sup Wald test. The critical value also increases as the dimensional parameter \( n \) rises. Finally, the critical values of \( \sup_{\tau \in \Theta} F(\tau) \) (reported under column w/o constant in Table 1) are less than those of \( \sup_{\tau \in \Theta} F^d(\tau) \) (reported under
column w/ constant). So as expected, the intercept term shifts the limit distribution.

The limit distributions in (17) and (18) are in terms of TPBM, which involves the parameter $\tau$. By contrast Theorem 2 of Seo (2006) shows that $\tau$ is absent in the limit distribution of the ECM test. This difference is due to the selection of the threshold variable. This paper uses $\Delta x_{t-1}$ as the threshold variable, and grid search is performed over the 15th and 85th percentile of $\Delta x_{t-1}$. Seo uses $x_{t-1}$ as the threshold variable, and grid search is over a set of fixed values. Because $x_{t-1}$ is nonstationary under the null, the fixed value is asymptotically equivalent to zero. This fact implies that the ECM test fails to efficiently use the information contained in the grid search.

Modifications for the Nonzero Drift Component

When the series contains a nonzero drift component under the null hypothesis, the testing regression is modified as

$$\alpha(L)\Delta y_t = (\delta_1 y_{t-1} + c_{01} + c_{11} t) 1_{1t-1} + (\delta_2 y_{t-1} + c_{02} + c_{12} t) 1_{2t-1} + e_t.$$  \hspace{1cm} (19)

Correspondingly, we make the auxiliary assumption that

**Assumption 4** Under the null hypothesis $\delta_1 = \delta_2 = 0$, $y_t$ is generated by (19) with $c_{01} = c_{02} = c_0 \neq 0$ and $c_{11} = c_{12} = 0$.

Notice that the intercept term and trend term in (19) are both multiplied by the indicators. This formulation is intended to facilitate the derivation of the canonical form introduced by Sims et al. (1990). To see this, note that under Assumption 4 $y_t$ consists of a deterministic trend and a stochastic trend. Therefore, for the purpose of developing asymptotic theory, it is necessary to avoid collinearity by “removing” the deterministic trend from $y_t$, and consider the canonical form given by

$$\alpha(L)\Delta y_t = (\delta_1^* \xi_{t-1} + c_{01}^* + c_{11}^* t) 1_{1t-1} + (\delta_2^* \xi_{t-1} + c_{02}^* + c_{12}^* t) 1_{2t-1} + e_t.$$  \hspace{1cm} (20)

where $\xi_{t-1} = y_{t-1} - c_0(t - 1)$, $\delta_1^* = \delta_1$, $\delta_2^* = \delta_2$, $c_{01}^* = c_{01} - c_0$, $c_{02}^* = c_{02} - c_0$, $c_{11}^* = c_{11} + c_0$, and $c_{12}^* = c_{12} + c_0$. Now the key regressor $\xi_{t-1}$ in (20) contains only the stochastic trend. The hypothetical model (20) is just the algebra rearrangement of (19), so testing $\delta_1 = \delta_2 = 0$ in (19) is equivalent to testing $\delta_1^* = \delta_2^* = 0$ in (20). The computation of the ADL test, denoted by $\sup_{\tau \in \Theta} F^d(\tau)$ in this case, is based on (19). The limit distribution is given in the following theorem.
Theorem 3 (case III) Under (10) and Assumptions 1, 2, 4, as $T \to \infty$,

$$\sup_{\tau \in \Theta} F^t(\tau) \Rightarrow \sup_{\tau \in \Theta} \text{trace} \left[ J_0^t (J_1^t)^{-1} J_t^d \right],$$  \hspace{1cm} (21)

where

$$J_0^d = J_0 - C_{21}^* C_{22}^* - C_{31}^* C_{32}^* C_{33}^*$$

$$J_1^d = J_1 - C_{21}^* C_{22}^* C_{21}^* - C_{31}^* C_{32}^* C_{31}^*.$$

$$C_{21} = \left( \begin{array}{cc}
\int W(r, 1) dr & \int rW(r, 1) dr \\
0 & 0
\end{array} \right), \quad C_{22} = \left( \begin{array}{cc}
\tau & \tau \int r dr \\
\tau \int r dr & \tau \int r^2 dr
\end{array} \right), \quad C_{23} = \left( \begin{array}{cc}
W(1, \tau) & \int rW(r, \tau) dr' \\
\int rW(r, \tau) dr & \int r^2W(r, \tau) dr'
\end{array} \right),$$

$$C_{31} = \left( \begin{array}{cc}
0 & 0 \\
(1 - \tau) \int W(r, 1) dr & (1 - \tau) \int rW(r, 1) dr
\end{array} \right), \quad C_{32} = \left( \begin{array}{cc}
1 - \tau & (1 - \tau) \int r dr \\
(1 - \tau) \int r dr & (1 - \tau) \int r^2 dr
\end{array} \right),$$

$$C_{33} = \left( \begin{array}{c}
W(1, 1) - W(1, \tau) \\
[\int rW(r, 1) - \int rW(r, \tau)']
\end{array} \right).$$

Note that the distribution does not depend on $c_0$, the drift term; in particular, whether or not the true value of $c_0$ is zero does not matter. The critical values of the distribution are reported under the column w/trend in Table 1.

Monte Carlo Experiments

This section compares the performances of the ADL test, ECM test and ES test in finite samples. The number of simulations is 5000, and the sample size is 100. We first compare rejection frequencies under the null hypothesis (size) at 5% nominal level. The data generating process (DGP) is given by

$$\Delta y_{1t} = c_1 \Delta y_{1t-1} + e_{1t}$$  \hspace{1cm} (22)

$$\Delta y_{2t} = c_2 \Delta y_{1t-1} + e_{2t}$$  \hspace{1cm} (23)

$$\begin{pmatrix} e_{1t} \\ e_{2t} \end{pmatrix} \sim \text{iidn}(0, \Omega), \quad \Omega = \begin{pmatrix} 1 & d_1 \\ d_1 & 1 \end{pmatrix},$$  \hspace{1cm} (24)

where the short-run dynamics is governed by $c_1$ and $c_2$, and the correlation between the error terms is determined by $d_1$. Their default values are $c_1 = 0.1$, $c_2 = 0.5$, and $d_1 = 0.3$. Panels A, B, C of Figure 1 plot the sizes against varying $c_1$, $c_2$ and $d_1$, respectively, with testing regressions including no intercept term (case I). We can summarize the findings as
follows. First, the size of the ADL test (marked by circle) is close to the nominal level 0.05 in all cases, and is invariant to $c_1$, $c_2$ and $d_1$. This confirms that the limit null distribution in Theorem 1 is indeed free of these parameters. It also implies that the bootstrap is not needed to implement the ADL test. Second, there is evidence that the ES test (marked by triangle) is slightly undersized. Finally, above-nominal size as great as 0.38 is found for the ECM test (marked by square). The severe size distortion found here is comparable to that reported in Table 2 of Seo (2006). These three findings remain largely unchanged in Figure 2 with testing regressions including the intercept term (case II).

Next we investigate the rejection frequencies of the tests under the alternative hypothesis (power). Given the severe size distortion of the ECM test, we compare the size-adjusted power by using the 5% bootstrap critical value obtained under the null hypothesis. The DGP is given by

\begin{align*}
x_{t-1} &= y_{1t-1} - k_2 y_{2t-1} \\
\Delta y_{1t} &= -0.1 x_{t-1} 1(x_{t-1} < 0) + k_1 x_{t-1} 1(x_{t-1} \geq 0) + c_1 \Delta y_{1t-1} + e_{1t} \\
\Delta y_{2t} &= c_2 \Delta y_{1t-1} + e_{2t},
\end{align*}

and (24). The cointegrating vector is $(1,-k_2)$, and the error correcting speed for $y_{1t}$ is determined by $k_1$. Figures 3 and 4 plot the powers against various parameter values for cases I and II, respectively. The powers are shown to rise as $k_1$ deviates from -0.1. When $k_1 = -0.1$, the error correcting speeds are regime-invariant, and the powers of the tests are minimized. The powers seem to depend positively on $c_1$, while remain largely invariant to $d_1$. We see intertwined lines in all panels, meaning no test can dominate others in power. Overall, the closeness of the power lines indicates that the power of the ADL test is comparable to the ES and ECM tests.

**An Application**

This section investigates the bivariate relationship of U.S. 10-year treasury constant yield (10ybond, $y_{1t}$) and the effective federal funds rate (Fedfunds, $y_{2t}$). The monthly data from July 1954 to June 2010 are downloaded from Economic Data-FRED. Panel A of Figure 5 plots the two series. We see both series appear nonstationary (with long “swing”), but tend to move together over time. Therefore we want to look into the possible long run relationship given as

\[ y_{1t} = \gamma_0 + \gamma_1 y_{2t} + x_t, \]
where \( x_t \) is the error correction term. The OLS estimation of (28) is \( x_t = y_{1t} - 2.63 - 0.69y_{2t} \), so the long term rate of 10ybond exceeds the short term rate of Fedfunds by 2.63 when the latter equals zero. Panel B of Figure 5 plots the estimated error correction term. The fact that the error correction term shows less persistence than the two series in Panel A is suggestive of cointegration.

Economic intuition says that the adjustment speed of the interest rate pair may depend on \( x_t \), which measures the gap between them. Big gap supposedly leads to fast adjustment. At this point an interesting question is, which level of \( x_t \) will trigger the fast adjustment. In order to answer that, we grid search the threshold value \( \tau \) by minimizing the determinant\(^7\) of \( \hat{\Omega} = T^{-1} \sum \hat{e}_t \hat{e}_t' \), where \( \hat{e}_t \) denotes the OLS residual of the two-regime TVADLM\(^8\)

\[
\alpha_1(L) \Delta y_{1t} = \delta_{11} y_{t-1} 1_{1t-1} + \delta_{12} y_{t-1} 1_{2t-1} + d_1 + e_{1t} \\
\alpha_2(L) \Delta y_{2t} = \delta_{21} y_{t-1} 1_{1t-1} + \delta_{22} y_{t-1} 1_{2t-1} + d_2 + e_{2t}.
\]

(29) (30)

The model above excludes trend term since Figure 5 shows both series are not trending. We follow the parsimony principle and include one lag (let \( p = 1 \)) in \( \alpha_1(L) \) and \( \alpha_2(L) \). The threshold value is estimated as -1.2, and a dash line corresponding to that value is drawn in Panel B of Figure 5. It seems that below the dash line the error series is less persistent (or moves more quickly) than above the line. This observation echoes the economic intuition, and signifies regime-switching.

Table 2 reports the estimation results for (29) and (30). There are two main findings about the estimated coefficients. First, asymmetric adjustment speeds are confirmed by the fact that the absolute values of the coefficients in lower regime one (with \( 1_{1t-1} = 1(z_t < -1.2) \)) are greater than those in upper regime two (with \( 1_{2t-1} = 1(z_t \geq -1.2) \)). For instance, in equation (30), the estimated coefficient of \( y_{1t-1} \) is 0.27 and significant in regime one but -0.02 and insignificant in regime two. To see what this means, notice that in regime one the short term rate Fedfunds exceeds the long term rate 10ybond by an amount greater than 1.2. In other words, the yield curve becomes highly inverted in regime one. This situation is quite unusual, so market is anticipated to act quickly. The estimation result indicates that fast adjustment indeed occurs in regime one.

The second main finding is, the absolute values of two estimated coefficients in (30) are significantly greater than those in (29). Take the coefficient for the regressor \( y_{2t-1} 1_{1t-1} \). It is

---

\(^7\)Minimizing the trace produces similar results.

\(^8\)The two-regime model is used to achieve simple exposition. The model presented in this section is not necessarily the optimal one.
-0.26 and significant in (30), but 0.03 and insignificant in (29). Recall that the regressands are differenced 10ybond in (30) and differenced Fedfunds in (29). Hence the second finding indicates that the two interest rates play asymmetric roles in the adjustment process. More explicitly, Fedfunds has two big coefficients (0.27 and -0.26) and therefore plays bigger role than 10ybond. This result is once again intuitively appealing. Fedfunds plays bigger role just because it is a policy tool frequently manipulated by the Federal Reserve.

Note that both findings above are unavailable from the linear Engle-Granger type model which pre-excludes nonlinearity. The single-equation threshold model used by Enders and Siklos (2001), however, can only detect the asymmetric adjustment speeds, but not the asymmetric roles. This section serves as one example showing that the system-equation threshold model is inherently more informative than the single-equation model, and the nonlinear model is more general than the linear model. The estimated model seems adequate as the Ljung-Box Q test with one lag finds no serial correlation for the residual.

Finally we test the null hypothesis of no threshold cointegration ($H_0 : \delta_{ij} = 0, (i, j = 1, 2)$) using the proposed system-equation ADL test. Neither the ECM test of Seo (2006) nor the single-equation ADL test of Li and Lee (2010) can be used here, because $\gamma_0$ and $\gamma_1$ in (28) are unknown, and because both variables are error-correcting (weak exogeneity fails). We let $\Theta = [0.15, 0.85]$, and the ADL test of case II equals 183.0, greater than the corresponding 5% critical value 37.7. The Enders-Siklos test equals 20.8, so rejects the null hypothesis as well. We conclude that 10ybond and Fedfunds are cointegrated. There is fast adjustment associated with highly inverted yield curve, and Fedfunds plays bigger adjustment role than 10ybond.
Conclusion

In this article we develop a new system-equation ADL test for threshold cointegration. The new test relaxes restrictions such as known cointegrating vectors and weak exogeneity imposed by previous tests. Unlike the single-equation test, the new test allows for multiple cointegration relationships. The proposed test employs an improved grid search for the threshold value. As a result, information contained in the grid search is used efficiently regardless of sample sizes. The limit null distribution of the proposed test, expressed as functionals of two-parameter Brownian motion, can be tabulated since no nuisance parameters are involved. Monte Carlo simulations demonstrate that the size distortion of our test is negligible, so bootstrap is unwanted. The power of our test is comparable to existing tests. We provide an empirical application of the new test, in which equilibrium between long term and short term interest rates is found and twofold asymmetries are highlighted. In particular, statistical evidence is reported for asymmetric adjustment roles; such evidence is eluded in the single-equation model.
Mathematical Proof

The asymptotic theory is based on the following lemma. Suppose \( e_t \) is an i.i.d \((n \times 1)\) sequence with mean zero, finite fourth moments, and \( Ee_t e_t' = \Omega = PP' \) where \( P \) is the Cholesky factor. Let \( u_t = D(L)e_t \) denote a stationary linear vector sequence with \( \sum_{s=0}^{\infty} |D_{ij}^s| < \infty \) for each \( i, j = 1, \ldots, n \). Define \( \xi_t = \sum_{j=1}^{t} u_j, \Lambda = (D_0 + D_1 + \ldots)P = D(1)P, \) and \( \Lambda_1 = \sum_{i=1}^{\infty} E(1(z_{t-1} < \tau)1(z_{t-1-i} < \tau)u_t u_{t-i}') \). \( W(r, \tau) \) denotes the two-parameter Brownian motion introduced by Caner and Hansen (2001).

Lemma 1

\[
\begin{align*}
a & : T^{-1/2} \sum_{t=1}^T 1(z_{t-1} < \tau)e_t \Rightarrow PW(r, \tau) \\
b & : T^{-1} \sum_{t=1}^{T} \xi_{t-1}1(z_{t-1} < \tau)e_t' \Rightarrow \Lambda \left[ \int W(r, 1)dW(r, \tau) \right] P' \\
c & : T^{-2} \sum_{t=1}^{T} \xi_{t-1}\xi_{t-1}'1(z_{t-1} < \tau) \Rightarrow \tau \Lambda \left[ \int W(r, 1)dW(r, 1)'dr \right] \Lambda' \\
d & : T^{-1} \sum_{t=1}^{T} \xi_{t-1}1(z_{t-1} < \tau)u_t' \Rightarrow \Lambda \left[ \int W(r, 1)dW(r, \tau) \right] \Lambda' + \Lambda_1
\end{align*}
\]

Proof of Lemma 1-a: Let \( v_t \) be an i.i.d \((n \times 1)\) vector with mean zero and unity variance \( Ev_t v_t' = I \). Notice that \( \{1(z_{t-1} < \tau)v_t\}, i = (1, \ldots, n) \), is a strictly stationary and ergodic martingale difference process with variance \( E(1(z_{t-1} < \tau)v_t)^2 = \tau \). For any \((r, \tau)\), the martingale difference central limit theorem implies that

\[
T^{-1/2} \sum_{t=1}^{[Tr]} 1(z_{t-1} < \tau)v_t \rightarrow^d N(0, r\tau),
\]

and the asymptotic covariance kernel is

\[
E \left( T^{-1/2} \sum_{t=1}^{[Tr_1]} 1(z_{t-1} < \tau_1)v_t \right) \left( T^{-1/2} \sum_{t=1}^{[Tr_2]} 1(z_{t-1} < \tau_2)v_t \right) = (r_1 \wedge r_2)(\tau_1 \wedge \tau_2).
\]

The proof of the stochastic equicontinuity of \( T^{-1/2} \sum_{t=1}^{[Tr]} 1(z_{t-1} < \tau)v_t \) is lengthy and can be found in Caner and Hansen (2001). The final result is deduced after applying the Cramer-Wold device and noticing that \( e_t = P v_t \). Notice that (3) generalizes the traditional multi-
variate invariance principle since when $\tau = 1$ it follows that $1(z_{t-1} < \tau) = 1$ with probability one, and (3) becomes $T^{-1/2}\sum_{i=1}^{\lfloor T \rfloor} e_i \Rightarrow PW(r, 1)$ where $W(r, 1)$ is the standard multivariate one-parameter Brownian motion.

**Proof of Lemma 1-b**: We apply the multivariate Beveridge-Nelson decomposition to $u_t$ and obtain $u_t = D(1)e_t + \tilde{e}_{t-1} - \tilde{e}_t$ where $\tilde{e}_t = \tilde{D}(L)e_t$, $\tilde{D}(L) = \sum_{j=0}^{\infty} \tilde{d}_j L^j$ and $\tilde{d}_j = \sum_{s=j+1}^{\infty} d_s$. The stated result is obtained by using Theorem 2.2 in Kurtz and Protter (1991) and Lemma 1-a above.

**Proof of Lemma 1-c**: This is a trivial extension of result 3 of Theorem 3 of Caner and Hansen (2001) to the multivariate case.

**Proof of Lemma 1-d**: This is a trivial extension of result (e) of Lemma 3.1 of Phillips and Durlauf (1986) to the two-parameter Brownian Motion case.

**Proof of Theorem 1**: We prove the result when the testing regression is a TVADL(1):

$$\Delta y_t = \delta_1 y_{t-1}1_{it-1} + \delta_2 y_{t-1}1_{2t-1} + \alpha_1 \Delta y_{t-1} + e_t.$$  

The proof can be easily extended to a general TVADL(p) model. Under the null hypothesis $\delta_1 = \delta_2 = 0$, it follows that $\alpha(L)\Delta y_t = e_t$, $\alpha(L) = I_n - \alpha_1 L$. Define $u_t = D(L)e_t$, $D(L) = \alpha(L)^{-1}$, then $y_{t-1} = \xi_{t-1}$ in Lemma 1. Let $\tilde{y}_{t-1} = (y'_{t-1}1_{it-1}, y'_{t-1}1_{2t-1})'$, $\delta = (\delta_1, \delta_2)$, $x_{t-1} = (\tilde{y}_{t-1}', \Delta y_{t-1}')'$, and $\theta = (\delta, \alpha_1)$. Consider the scaling matrix

$$\Upsilon = \text{diag}(TI_n, T^{1/2}I_n)$$

We have

$$\Upsilon (\hat{\theta} - \theta) = \left[ \Upsilon^{-1} \left( \sum_{t=1}^{T} x_{t-1}x'_{t-1} \right) \Upsilon^{-1} \right]^{-1} \left[ \Upsilon^{-1} \sum_{t=1}^{T} x_{t-1}e_t \right],$$

17
In order to obtain $B_{11}$ note $1_{1t-1} = 1(j_{t-1} < \tau)$, $1_{2t-1} = 1 - 1(j_{t-1} < \tau)$, and $1_{1t-1}1_{2t-1} = 0.$ The off-diagonal elements of $B_{11}$ are due to the orthogonality of $1_{1t-1}$ and $1_{2t-1}$. The diagonal terms are due to Lemma 1-c, and the facts that $T^{-2} \sum_{t=1}^{T} y_{t-1} y'_{t-1} \Rightarrow \Lambda \left[ \int W(r, 1)W(r, 1)'dr \right] \Lambda'$, cf. Lemma 3.1.(b) of Phillips and Durlauf (1986), and

$$T^{-2} \sum_{t=1}^{T} y_{t-1} y'_{t-1} 1_{2t-1} = T^{-2} \sum_{t=1}^{T} y_{t-1} y'_{t-1}(1 - 1(j_{t-1} < \tau)) \Rightarrow (1 - \tau)\Lambda \left[ \int W(r, 1)W(r, 1)'dr \right] \Lambda'.$$

$B_{12} = o_p(1)$ since Lemma 1-d implies $T^{-3/2} \sum_{t=1}^{T} \bar{y}_{t-1} \Delta y'_{t-1} = T^{-1/2}T^{-1} \sum_{t=1}^{T} \bar{y}_{t-1} \Delta y'_{t-1} = T^{-1/2}O_p(1) = o_p(1)$. $B_{22}$ is due to the Weak Law of Large Number. We obtain $C_1$ by Lemma 1-b, and because $T^{-1} \sum_{t=1}^{T} e'_t \Rightarrow \Lambda \left[ \int W(r, 1)dW(r, 1)' \right] P'$, cf. Proposition 18.1.(f) of Hamilton (1994), and

$$T^{-1} \sum_{t=1}^{T} y_{t-1} 1_{2t-1} e'_t =$$

$$T^{-2} \sum_{t=1}^{T} y_{t-1}(1 - 1(j_{t-1} < \tau)) e'_t \Rightarrow \Lambda \left[ \int W(r, 1)dW(r, 1)' - \int W(r, 1)dW(r, \tau)' \right] P'.$$

$C_2$ is due to the Central Limit Theorem.

where

$$\Upsilon^{-1} \left( \sum_{t=1}^{T} x_{t-1} x'_{t-1} \right) \Upsilon^{-1} \Rightarrow B = \begin{pmatrix} B_{11} & B_{12} \\ B'_{12} & B_{22} \end{pmatrix},$$

$$\Upsilon^{-1} \sum_{t=1}^{T} x_{t-1} e_t \Rightarrow C = \begin{pmatrix} C_1 \\ C_2 \end{pmatrix},$$

$$B_{11} = \begin{pmatrix} \tau \Lambda \left[ \int W(r, 1)W(r, 1)'dr \right] \Lambda' & 0 \\ 0 & (1 - \tau)\Lambda \left[ \int W(r, 1)W(r, 1)'dr \right] \Lambda' \end{pmatrix},$$

$$B_{12} = o_p(1),$$

$$B_{22} = E \Delta y_{t-1} \Delta y'_{t-1},$$

$$C_1 = \begin{pmatrix} \Lambda \left[ \int W(r, 1)dW(r, \tau)' \right] P' \\ \Lambda \left[ \int W(r, 1)dW(r, 1)' - \int W(r, 1)dW(r, \tau)' \right] P' \end{pmatrix},$$

$$C_2 = O_p(1).$$
Because $\Upsilon^{-1}\left(\sum_{t=1}^{T} x_{t-1}'x_{t-1}'\right) \Upsilon^{-1}$ is asymptotically block-diagonal, it follows that

$$\Upsilon \left( \hat{\theta} - \theta \right) \Rightarrow \begin{pmatrix} B_{11}^{-1}C_1 \\ B_{22}^{-1}C_2 \end{pmatrix}.$$ 

In particular, $B_{11}^{-1}C_1$ indicates that

$$TI_n(\hat{\delta}_1 - \delta_1) \Rightarrow \tau^{-1}\Lambda^{-1} \left[ \int W(r, 1)W(r, 1)'dr \right]^{-1} \left[ \int W(r, 1)dW(r, \tau)' \right] P',$$

and

$$TI_n(\hat{\delta}_2 - \delta_2) \Rightarrow (1 - \tau)^{-1}\Lambda^{-1} \left[ \int W(r, 1)W(r, 1)'dr \right]^{-1} \left[ \int W(r, 1)dW(r, 1)' - \int W(r, 1)dW(r, \tau)' \right] P'.$$

This implies the super-consistency of $\hat{\delta}$, and the consistency of $\hat{\Omega}_r$. For given $\tau$ let $Y_r$ and $e$ be the matrices stacking $(y_{t-1}1_{t-1}, y_{t-1}1_{2t-1})$ and $e_t$, respectively. Define $Q$ as projection onto the orthogonal space of the lagged term $\Delta y_{t-1}$. Write the OLS estimates for $(\delta_1, \delta_2)$ as $\hat{\delta}_r = (\hat{\delta}_1, \hat{\delta}_2)$. Then the Wald statistic for $H_0 : \delta_1 = \delta_2 = 0$ is

$$F(\tau) = [\text{vec}(\hat{\delta}_r)]'\left[\text{var}(\text{vec}(\hat{\delta}_r))^{-1}\right][\text{vec}(\hat{\delta}_r)]$$

$$= [\text{vec}(\hat{\delta}_r)]' \left[[Y_r'QY_r]^{-1} \otimes \hat{\Omega}_r^{-1}\right]^{-1}[\text{vec}(\hat{\delta}_r)]$$

$$= [\text{vec}(e'QY_r)]' \left[[Y_r'QY_r]^{-1} \otimes \hat{\Omega}_r^{-1}\right]^{-1}[\text{vec}(e'QY_r)]$$

$$= \text{trace} \left[\hat{\Omega}_r^{-1/2}(e'QY_r)(Y_r'QY_r)^{-1}(Y_r'Qe)\hat{\Omega}_r^{-1/2}\right].$$

We can show that

$$T^{-1}Y_r'Qe = T^{-1}\Sigma\tilde{Y}_{t-1}e_t - T^{-1/2}(T^{-1}\Sigma\tilde{Y}_{t-1}\Delta y_{t-1}') \left(T^{-1}\Sigma\Delta y_{t-1}\Delta y_{t-1}'\right)^{-1} (T^{-1/2}\Sigma\Delta y_{t-1}e_t')$$

$$= T^{-1}\Sigma\tilde{Y}_{t-1}e_t + o_p(1) \Rightarrow C_1,$$

and

$$T^{-2}Y_r'QY_r = T^{-2}\Sigma\tilde{Y}_{t-1}\tilde{Y}_{t-1}' - T^{-1}(T^{-1}\Sigma\tilde{Y}_{t-1}\Delta y_{t-1}') \left(T^{-1}\Sigma\Delta y_{t-1}\Delta y_{t-1}'\right)^{-1} (T^{-1}\Sigma\Delta y_{t-1}\tilde{Y}_{t-1}')$$

$$= T^{-2}\Sigma\tilde{Y}_{t-1}\tilde{Y}_{t-1}' + o_p(1) \Rightarrow B_{11}.$$
Combining with $\hat{\Omega}_{\tau}^{-1/2} = P^{-1} + o_p(1)$ we can show

$$F(\tau) = \text{trace} \left[ \hat{\Omega}_{\tau}^{-1/2}(e'QY_{\tau})(Y_{\tau}'QY_{\tau})^{-1}(Y_{\tau}'Qe)\hat{\Omega}_{\tau}^{-1/2} \right]$$

$$= \text{trace} \left[ \hat{\Omega}_{\tau}^{-1/2}C_1B_{11}^{-1}C_1\hat{\Omega}_{\tau}^{-1/2} \right] + o_p(1)$$

$$\Rightarrow \text{trace} \left[ J_0^TJ_1^{-1}J_0 \right],$$

where

$$J_0 = \begin{pmatrix}
\int W(r,1)dW(r,\tau) \\
\int W(r,1)dW(r,1)' - \int W(r,1)dW(r,\tau)'
\end{pmatrix}$$

$$J_1 = \begin{pmatrix}
\tau \left[ \int W(r,1)W(r,1)'dr \right] & 0 \\
0 & (1 - \tau) \left[ \int W(r,1)W(r,1)'dr \right]
\end{pmatrix}.$$

The remaining steps for the convergence of $\sup F$ follow from the continuous mapping theorem. End of Proof.

**Proof of Theorem 2:** Only a sketch of the proof is provided since it is similar to that of Theorem 1. Let the testing regression be a TVADL(1) with the intercept term:

$$\Delta y_t = \delta_1 y_{t-1}1_{l-1} + \delta_2 y_{t-1}1_{2t-1} + c_0 + \alpha_1 \Delta y_{t-1} + e_t.$$

Under the null hypothesis $\delta_1 = \delta_2 = 0$ and $c_0 = 0$ it follows $y_{t-1} = \xi_{l-1}$ in Lemma 1. For given $\tau$ let $Y_{\tau}, \Delta Y_{-1}$ and $e$ be the matrices stacking $(y_{t-1}1_{l-1}, y_{t-1}1_{2t-1})$, $\Delta y_{t-1}$ and $e_t$, respectively. Let $Q^d$ be the projection onto the orthogonal space of the intercept term and $\Delta y_{t-1}$. Because $E\Delta y_{t-1} = 0$, $Q^d$ can be written as the sum of the projection onto the orthogonal space of intercept term and the projection onto the orthogonal space of $\Delta y_{t-1}$. Next we can show that

$$T^{-1}Y_{\tau}'Q^d e = T^{-1}Y_{\tau}'[I - \iota(\iota')\iota'] - \Delta Y_{-1}(\Delta Y_{-1}'\Delta Y_{-1})^{-1}\Delta Y_{-1}' e$$

$$= T^{-1}Y_{\tau}'[I - \iota(\iota')\iota'] e + o_p(1)$$

$$= T^{-1}\Sigma\tilde{y}_{t-1}e_t' - (T^{-3/2}\Sigma\tilde{y}_{t-1})(T^{-1/2}\Sigma e_t') + o_p(1)$$

$$\Rightarrow C^d_1 \equiv C_1 - C_2,$$
where

$$C_2 = \left( \frac{\tau \Lambda \left[ \int W(r, 1)dr \right] \left[ \int dW(r, 1) \right] P'}{(1 - \tau) \Lambda \left[ \int W(r, 1)dr \right] \left[ \int dW(r, 1) \right] P'} \right) \equiv \Lambda C_2^* P'. $$

We can also show

$$T^{-2}Y'_tQ^dY_t = T^{-2}Y'_t[I - \iota(\iota')\iota'] - \Delta Y_{-1}(\Delta Y'_{-1}(\Delta Y_{-1})^{-1}\Delta Y'_{-1})Y_t$$

$$= T^{-2}Y'_t[I - \iota(\iota')\iota']Y_t + o_p(1)$$

$$= T^{-2}\Sigma\hat{y}_{t-1}\hat{y}'_{t-1} - (T^{-3/2}\Sigma\hat{y}_t)(T^{-3/2}\Sigma\hat{y}'_t) + o_p(1)$$

$$\Rightarrow B_{11}^d = B_{11} - B_2 B'_2,$$

where

$$B_2 = \left( \frac{\tau \Lambda \left[ \int W(r, 1)dr \right]}{(1 - \tau) \Lambda \left[ \int W(r, 1)dr \right]} \right) \equiv \Lambda B_2^*.$$ 

It follows that

$$F^d(\tau) = \text{trace} \left[ \hat{\Omega}_{\tau}^{-1/2}(\iota'Q^dY_t)(Y'_tQ^dY_t)^{-1}(Y'_tQ^d\iota)\hat{\Omega}_{\tau}^{-1/2} \right]$$

$$= \text{trace} \left[ \hat{\Omega}_{\tau}^{-1/2}C_1^d(B_{11}^d)^{-1}C_1^d\hat{\Omega}_{\tau}^{-1/2} \right] + o_p(1)$$

$$\Rightarrow \text{trace} \left[ J_0^d (J_1^d)^{-1}J_0^d \right],$$

where

$$J_0^d = J_0 - C_2^*$$

$$J_1^d = J_1 - B_2^* B'_2.$$ 

**Proof of Theorem 3:** The testing regression is given by

$$\Delta y_t = (\delta_1 y_{t-1} + c_{01} + c_{11} t)1_{1t-1} + (\delta_2 y_{t-1} + c_{02} + c_{12} t)1_{2t-1} + e_t.$$ 

For easy of exposition, the above regression excludes the lagged term $\Delta y_{t-1}$ since $\Delta y_{t-1}$ is asymptotically irrelevant, as shown by the proofs of Theorems 1 and 2. Under the null hypothesis $\delta_1 = \delta_2 = 0$, $c_{11} = c_{12} = 0$, and $c_{01} = c_{02} = c_0$ it follows that $y_t = c_0 t + \xi_t$. Because $y_t$ is collinear with the time trend when $c_0 \neq 0$, we consider the following hypothetical
regression in the canonical form

\[ \Delta y_t = (\delta^*_1 \xi_{t-1} + c_0^* + c_1^* t)1_{1t-1} + (\delta^*_2 \xi_{t-1} + c_0^* + c_12^* t)1_{2t-1} + e_t, \]

where \( \xi_{t-1} = y_{t-1} - c_0(t - 1), \) \( \delta^*_1 = \delta_1, \) \( \delta^*_2 = \delta_2, \) \( c_0^* = c_{01} - c_0, \) \( c_0^* = c_{02} - c_0, \) \( c_11^* = c_{11} + c_0, \) and \( c_{12} = c_{12} + c_0. \) For given \( \tau, \) stacking \( (\xi_{t-1}1_{1t-1}, \xi_{t-1}1_{2t-1}) \) as \( X_\tau, \) \( (1_{1t-1}, t1_{1t-1}) \) as \( M_1, \) and \( (1_{2t-1}, t1_{2t-1}) \) as \( M_2. \) Because \( 1_{1t-1}1_{2t-1} = 0, \forall t, \) the projection onto the orthogonal space of \( (1_{1t-1}, t1_{1t-1}, 1_{2t-1}, t1_{2t-1}), \) denoted by \( Q', \) is

\[ Q' = I - M_1(M_1M_1)^{-1}M_1' - M_2(M_2M_2)^{-1}M_2'. \]

Next we can show that

\[ T^{-1}X'_rQ'e = T^{-1}X'_r[I - M_1(M_1M_1)^{-1}M_1' - M_2(M_2M_2)^{-1}M_2']e, \]

\[ T^{-1}X'_r e \Rightarrow C_1, \]

\[ T^{-1}X'_rM_1(M_1M_1)^{-1}M_1'e = (T^{-1}X'_rM_1M_1^{-1}) (M_1^{-1}M_1M_1^{-1}) (M_1^{-1}M_1') \Rightarrow C_{21}C_{22}C_{23}, \]

\[ T^{-1}X'_rM_2(M_2M_2)^{-1}M_2'e = (T^{-1}X'_rM_2M_2^{-1}) (M_2^{-1}M_2M_2^{-1}) (M_2^{-1}M_2') \Rightarrow C_{31}C_{32}C_{33}, \]

where

\[
C_{21} = \begin{pmatrix} \tau \Lambda \int W(r, 1)dr & \tau \Lambda \int rW(r, 1)dr \\ 0 & 0 \end{pmatrix} \equiv \Lambda C_{21}^*,
\]

\[
C_{22} = \begin{pmatrix} \tau & \tau \int rdr \\ \tau \int rdr & \tau \int r^2dr \end{pmatrix},
\]

\[
C_{23} = \begin{pmatrix} W(1, \tau)P' \\ \int rdrW(r, \tau)'P' \end{pmatrix} \equiv C_{23}^* P'.
\]

\[
C_{31} = \begin{pmatrix} 0 & 0 \\ (1 - \tau)\Lambda \int W(r, 1)dr & (1 - \tau)\Lambda \int rW(r, 1)dr \end{pmatrix} \equiv \Lambda C_{31}^*,
\]

\[
C_{32} = \begin{pmatrix} 1 - \tau & (1 - \tau) \int rdr \\ (1 - \tau) \int rdr & (1 - \tau) \int r^2dr \end{pmatrix},
\]

\[
C_{33} = \begin{pmatrix} [W(1, 1) - W(1, \tau)]P' \\ [\int rdrW(r, 1)' - \int rdrW(r, \tau)']P' \end{pmatrix} \equiv C_{33}^* P'.
\]

22
Thus,
\[ T^{-1}X'_t Q^t e \Rightarrow C_1 - C_{21} C_{22} C_{23} - C_{31} C_{32} C_{33}. \]

In a similar fashion we can show
\[ T^{-2}X'_t Q^t X_t \Rightarrow B_{11} - C_{21} C_{22} C_{21}' - C_{31} C_{32} C_{31}'. \]

Finally,
\[
F^t(\tau) = \text{trace} \left[ \hat{\Omega}_\tau^{-1/2}(e'Q^t X_t)(X'_t Q^t X_t)^{-1}(X'_t Q^t e)\hat{\Omega}_\tau^{-1/2}' \right] \\
\Rightarrow \text{trace} \left[ J_0'^t (J_1^t)^{-1} J_d^t \right],
\]

where
\[
J_0^d = J_0 - C_{21}^* C_{22} C_{23}^* - C_{31}^* C_{32} C_{33}^* \\
J_1^d = J_1 - C_{21} C_{22} C_{21}' - C_{31} C_{32} C_{31}'.
\]
References


Davies, R. B. (1977). Hypothesis testing when a nuisance parameter is present only under the alternative. *Biometrika* 64, 247–254.

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


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


## Table 1: Asymptotic critical values

<table>
<thead>
<tr>
<th>( \Phi = [0.10, 0.90] )</th>
<th>w/o constant</th>
<th>w/ constant</th>
<th>w/ trend</th>
</tr>
</thead>
<tbody>
<tr>
<td>1% 5% 10% 1% 5% 10% 1% 5% 10%</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>( n = 2 )</td>
<td>35.0 30.6 28.3</td>
<td>43.8 38.1 35.1</td>
<td>62.4 54.8 51.5</td>
</tr>
<tr>
<td>( n = 3 )</td>
<td>61.4 55.3 52.1</td>
<td>71.8 64.7 61.4</td>
<td>95.1 87.3 83.5</td>
</tr>
<tr>
<td>( n = 4 )</td>
<td>95.5 88.1 83.9</td>
<td>107.1 99.3 95.1</td>
<td>135.5 126.9 122.5</td>
</tr>
<tr>
<td>( n = 5 )</td>
<td>138.0 128.8 123.8</td>
<td>151.6 141.8 136.8</td>
<td>184.5 175.0 169.7</td>
</tr>
<tr>
<td>( \Phi = [0.15, 0.85] )</td>
<td>w/o constant</td>
<td>w/ constant</td>
<td>w/ trend</td>
</tr>
<tr>
<td>1% 5% 10% 1% 5% 10% 1% 5% 10%</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>( n = 2 )</td>
<td>35.5 30.2 27.9</td>
<td>43.5 37.7 34.7</td>
<td>60.0 54.3 51.1</td>
</tr>
<tr>
<td>( n = 3 )</td>
<td>61.3 55.2 52.1</td>
<td>71.3 64.1 60.6</td>
<td>95.1 86.9 82.8</td>
</tr>
<tr>
<td>( n = 4 )</td>
<td>95.4 87.7 83.6</td>
<td>106.6 99.2 94.6</td>
<td>136.1 126.7 122.2</td>
</tr>
<tr>
<td>( n = 5 )</td>
<td>137.2 127.9 123.4</td>
<td>150.8 140.8 135.6</td>
<td>184.1 174.2 169.0</td>
</tr>
<tr>
<td>( \Phi = [0.20, 0.80] )</td>
<td>w/o constant</td>
<td>w/ constant</td>
<td>w/ trend</td>
</tr>
<tr>
<td>1% 5% 10% 1% 5% 10% 1% 5% 10%</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>( n = 2 )</td>
<td>35.4 29.6 27.3</td>
<td>43.9 37.1 34.0</td>
<td>59.4 53.5 50.4</td>
</tr>
<tr>
<td>( n = 3 )</td>
<td>60.7 54.2 50.9</td>
<td>70.5 63.3 59.9</td>
<td>94.2 86.1 82.4</td>
</tr>
<tr>
<td>( n = 4 )</td>
<td>95.9 86.8 82.7</td>
<td>106.8 98.1 94.0</td>
<td>135.8 126.0 121.4</td>
</tr>
<tr>
<td>( n = 5 )</td>
<td>135.8 125.9 121.6</td>
<td>149.7 140.3 135.0</td>
<td>185.4 173.4 168.2</td>
</tr>
</tbody>
</table>
Table 2: Estimation Results

<table>
<thead>
<tr>
<th>Equation (29)</th>
<th>Regressors</th>
</tr>
</thead>
<tbody>
<tr>
<td>Coefficients</td>
<td>$y_{t-1}$</td>
</tr>
<tr>
<td>T Values</td>
<td>-1.52</td>
</tr>
<tr>
<td>Ljung-Box Q Test</td>
<td>2.48</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Equation (30)</th>
<th>Regressors</th>
</tr>
</thead>
<tbody>
<tr>
<td>Coefficients</td>
<td>$y_{t-1}$</td>
</tr>
<tr>
<td>T Values</td>
<td>4.81**</td>
</tr>
<tr>
<td>Ljung-Box Q Test</td>
<td>1.65</td>
</tr>
</tbody>
</table>

Tests for Threshold Cointegration

<table>
<thead>
<tr>
<th>ADL Test</th>
<th>$\sup_{\tau \in [0.15, 0.85]} F^d(\tau) = 138.0^{**}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Enders-Siklos Test</td>
<td>20.8**</td>
</tr>
</tbody>
</table>

Note: ** significant at 5% level.
Figure 1: Rejection Frequency of Various Tests under the Null Hypothesis at 5% Level with Intercept Term Excluded.
Figure 2: Rejection Frequency of Various Tests under the Null Hypothesis at 5% Level with Intercept Term Included.
Figure 3: Rejection Frequency of Various Tests under the Alternative Hypothesis at 5% Level with Intercept Term Excluded.
Figure 4: Rejection Frequency of Various Tests under the Alternative Hypothesis at 5% Level with Intercept Term Included.
Figure 5: Time Series Plot