ADL tests for threshold cointegration

Jing Li and Junsoo Lee[1]

April 15, 2009

Abstract

In this paper, we propose new tests for threshold cointegration in the autoregressive distributed lag (ADL) model. The indicators in the threshold model are based on either a nonstationary or stationary threshold variable. The cointegrating vector in this paper is not pre-specified. We adopt a supremum Wald type test to account for the so-called Davies problem. The asymptotic null distributions of the proposed tests are free of nuisance parameters. As such, a bootstrap procedure is not required and critical values of the proposed tests are tabulated. A Monte Carlo experiment shows a good finite-sample performance of the proposed tests.

JEL classification: C12; C15; C32

Keywords: Econometric Theory; Time Series

[1]Jing Li, Department of Economics, Assistant Professor, South Dakota State University, Box 504, Brookings, SD 57007-0895, USA (phone: 605-688-4848; Email: Jing.Li@sdstate.edu); Junsoo Lee (Corresponding author), Professor, Department of Economics, Finance and Legal Studies, University of Alabama, Box 870224, AL 35487, USA (phone: 205-348-8978; Email: jlee@cba.ua.edu). We are grateful to Walter Enders, Myung Hwan Seo, Mark Strazicich, and Mark Wohar for their helpful suggestions and comments.

1   Introduction

A large volume of papers document nonlinear and asymmetric adjustments
to a long-run equilibrium in macroeconomic time series. For example, Taylor
and Peel (2000), and Wu and Chen (2001) provide empirical evidence that
deviations from the long-run equilibrium level of exchange rates form a non-
linear adjustment process. The asymmetric dynamics in foreign exchange
rates can be caused by factors such as transaction costs and policy interven-
tion. Actually, there are many important studies that provide evidence of
nonlinear asymmetric behavior in various time series models; see Balke and
Wohar (1998), Sephton (2003), and Michael, Nobay and Peel (1997), among
others.

In this paper, we propose new tests for threshold cointegration in the
single equation framework. Our new tests utilize autoregressive distributed
lag (ADL) models. The properties of the ADL tests differ from the existing
tests using the Engle and Granger (1987, EG) type testing regression, or the
error correction model (ECM). In contrast to the EG or ECM based tests, we
do not test the significance of the coefficient of the cointegrating vector and
cointegrating residuals. Instead, we test the significance of the coefficient
of the lagged dependent variable in an unrestricted fashion. Ericsson and
Mackinnon (2002) refer to this type of cointegration test as an ADL test,
which was initially introduced by Banerjee, Dolado and Mestre (1986, BDM)
and Boswijk (1994, BO). The BDM test involves testing only the coefficient of
the lagged regressand, and requires adding leads of differences of conditioning
variables when strict exogeneity fails. In comparison, the BO test examines
the significance of the coefficients of the lagged regressand as well as the
lagged conditioning variables. In essence, our proposed tests are extensions
of these linear tests to a nonlinear threshold framework. To the best of
our knowledge, no threshold test using ADL models has been previously
considered in the literature. Thus, we wish to contribute to the literature by
filling a gap. We will discuss more details below, but the ADL based threshold
cointegration tests have desirable properties not found in the existing tests
based on the EG or ECM procedures.

Threshold cointegration models employ indicator functions reflecting a
regime change. We consider two different cases in a unified framework where
the indicators are based on either stationary or nonstationary threshold vari-
ables. We provide relevant asymptotic theory for both cases. The difficulty
lies in the treatment of the threshold parameter. Seo (2006) considers a fixed

finite value as a threshold parameter. Given that the threshold parameter is a nuisance parameter, Seo's test is based on the presumption that the probability of a nonstationary variable being less than the fixed threshold value approaches zero asymptotically. If so, Seo's test can be free of the threshold nuisance parameter in large samples. However, Seo (2006) shows that the test based on this asymptotic result exhibits serious size distortions. As a result, he relies on the bootstrap test to improve performance. Our treatment of the threshold parameter is somewhat different. In order to fully exploit the information contained in the threshold value even in small samples, we treat a certain percentile from the empirical distribution of the threshold variable as a threshold parameter. In other words, our grid search is based on a range of percentiles rather than a range of fixed values covering the space of whole real numbers. As a result, the asymptotic distribution of our tests depends on the parameter space of percentiles, which is restricted to lie between 0 and 1. In addition, we do not rely on the presumption that the threshold parameter vanishes in large samples. We perform simulations and demonstrate no serious size distortions in our new tests, even though we do not utilize bootstrapping.

Testing for threshold cointegration involves nonstandard inference in the sense that the threshold parameter (percentile in our case) is not identified under the null hypothesis. To resolve this so-called Davies problem (cf. Davies, 1977 and 1987), a Wald statistic is computed over percentiles of the empirical distribution of the threshold variable. The sup Wald type test is obtained by searching over the range of the percentile parameter. The asymptotic distributions of our tests can be expressed as functionals of (transformed) Brownian motions under the null hypothesis. Moreover, the distributions of our ADL threshold cointegration tests are free of nuisance parameters involving long-run endogeneity. This is possible for the BDM type test when correction for long-run endogeneity is made by adopting the lead-lag procedure of Saikkonen (1991). However, the BO type test does not require such corrections. As in the usual cointegration tests, the limit distributions of the tests depend only on the dimension of the regressors, types of threshold variables and the deterministic terms in the test regression.

Testing for threshold cointegration can be undertaken in either a system model or a single-equation conditional model. System-based tests, as pursued by Seo (2006), have the advantage of assuming away weak exogeneity. But the performance of system-based tests can be improved by utilizing single-equation tests if weak exogeneity holds (cf. Bowswijk and Franses (1992)

and Zivot (2000)). We consider only single-equation tests for some reasons. First, weak exogeneity is found in many applications. In such cases, weak exogeneity would better be imposed for simpler modeling strategies and reduced computational expense. Second, most often a specific cointegration relationship, such as the demand for money, is implied by economic theory. As a result, researchers may focus on one aspect of the economic system. Finally, single-equation tests are easier to compute than system-based tests. No matrix-oriented package is needed for single-equation tests.

The remainder of the paper will proceed as follows. In Section 2, we discuss the issues in testing for threshold cointegration. Section 3 discusses the threshold error-correction model and the weak exogeneity condition. In Section 4, we formally propose the BDM and BO tests. Their asymptotic null distributions are given and asymptotic critical values are tabulated. A Monte Carlo experiment is carried out in Section 5 to illustrate the finite sample performance of new tests. In Section 6, we provide an empirical example. Section 7 concludes.

2 Literature and Issues

In dealing with nonlinear long-run equilibrium models, two issues are mixed. One is whether nonlinearity exists or not, and the other is whether a long-run relationship exists or not. When Balke and Fomby (1997) initially suggest a test for threshold cointegration, they consider a two-step approach. In the first step, they examine the null of no cointegration using the Engle-Granger linear cointegration test. In the second step, they suggest to test whether threshold behavior is present. The two-step procedure entails room for improvement. Clearly, there are good reasons to jointly test for non-linearity and cointegration. Suppose that nonlinearity exists, but one adopts a linear cointegration test in the first step. We can expect that the linear test will hardly reject the null of no cointegration since existing nonlinearity will lead to loss of power. In light of Perron (1989), ignoring nonlinearity will make the test biased toward not rejecting the null of nonstationarity. The linear cointegration test will reject the null of no cointegration only when linear cointegration is present. As such, the linear cointegration test cannot distinguish linear cointegration against nonlinear cointegration. Thus, being unable to reject the null in a linear cointegration test does not necessarily imply the absence of a long-run relationship; the possibility of nonlinear

cointegration still remains. It is obvious that allowing for nonlinearity can increase power. The second step procedure testing for nonlinearity also poses a problem. The second step of testing for the presence of a threshold effect is valid only if cointegration holds and the cointegrating vector is known. It is easy to expect that the null of linearity will be rejected when nonstationarity holds. This is because such tests will diverge if cointegration does not hold. Then, rejection of the null does not necessarily mean nonlinearity. Instead, it can imply nonstationarity or no cointegration. As such, Seo (2006) notes that tests for a threshold effect in a vector error correction model, as suggested in Hansen and Seo (2002), are valid only in a cointegrated system. Thus, the presumption of cointegration seems critical.

The testing procedure developed by Enders and Siklos (2001) is encouraging in this regard. They suggest a formal test for the joint hypothesis of the absence of both nonlinearity and cointegration. They use the Engle-Granger (1987, EG) type testing regression. In particular, they propose two different sets of threshold cointegration tests with threshold autoregressive (TAR) and momentum-TAR (MTAR) adjustments, depending on whether a threshold variable is non-stationary or stationary. In the TAR model, the indicator function takes the value of 1 if the past residual from a cointegrating regression exceeds the threshold value and 0 otherwise. The MTAR model uses the first difference of the cointegrating residuals in the indicator. The threshold cointegration tests of Enders and Siklos (2001), however, also leave room for improvement. In particular, their tests entail limitations found in the Engle-Granger (EG, hereafter) type tests. One well known problem of the EG procedure is that the test imposes a strong restriction that may not hold in practice. Although the coefficients in the long-run cointegration regression usually differ from the short-run adjustment coefficients, these two sets of coefficients are assumed to be equivalent in the EG procedure. Kremers, Ericsson and Dolado (1992) refer to this restriction as a common factor restriction (CFR). As a consequence of imposing the CFR, the EG type tests lose power when the signal-noise ratio increases. This same problem carries over to threshold cointegration tests using the EG type testing procedure.

One may adopt an error correction model (ECM) framework. It is well known that an ECM cointegration test has a mixture of nonstandard and standard normal distributions, where the weight induces a nuisance parameter problem. Zivot (2000) suggests to estimate the weight parameter in a nonparameteric fashion. No such test for threshold cointegration has been suggested in the literature, and a similar procedure could be considered;

however, the ECM version threshold cointegration test also depends on the nuisance parameter, which can make things more complicated.

There are clear operating advantages of using the ADL version tests for threshold cointegration. The ADL based threshold cointegration tests are free of the CFR restriction problem, and they are free of the nuisance parameter dependency problem of the ECM based tests. As such, by adopting the ADL version tests, we can improve upon some limitations found in the existing tests.

3   The model

3.1   Triangular representation of a threshold vector error correction model

Consider an $(n+1)$-dimensional observed series $z_t = (y_t, x_t')', t = 1, \ldots T$, where the regressand y_t is a scalar and T is the number of observations. The two-regime threshold vector error correction model (TV-ECM) augmented by a deterministic term is given by

$$
\begin{aligned}
z_t &= d_t + w_t \\
\Delta w_t &= \Pi_1 w_{t-1} I_{1t} + \Pi_2 w_{t-1} I_{2t} + \Phi(L)\Delta w_{t-1} + \epsilon_t,
\end{aligned}
\tag{1}
$$

where d_t denotes the deterministic term, $\Phi(L)$ involves the p-th order polynomial matrices, and the innovation process is $\epsilon_t \sim iid(0, E\epsilon_t\epsilon_t')$. For expositional ease, we temporarily assume $d_t = 0$ so that $z_t = w_t$. The two regimes are determined by the indicators I_{1t} and I_{2t}, which are specified below. Most economic theories imply a long-run equilibrium in the relationship among integrated variables. Hence the cointegrating vector $\theta = (1, -\gamma')'$ is assumed to be regime-invariant. By definition, the error correction term is $\theta' z_t = y_t - \gamma' x_t$, which measures the deviation from the long-run relationship. Let $\kappa_1 = (\kappa_{11}, \kappa_{21}')'$ and $\kappa_2 = (\kappa_{12}, \kappa_{22}')'$ be the adjustment speed toward the long-run equilibrium in the two regimes, then the loading matrices in (1) are $\Pi_1 = \kappa_1\theta$ and $\Pi_2 = \kappa_2\theta$. The point of the TV-ECM model is that the adjustment speed is allowed to switch across two regimes. The different speed can be theoretically attributed to transaction costs and other factors. We assume that the polynomial term $\Phi(L)$ is regime-invariant so as to facilitate deriving asymptotic results under the null hypothesis. Note that the same TV-ECM model is considered by Seo (2006), but the indicator functions are

defined differently in our tests. The following assumption is made regarding model (1).

Assumption 1. *In model (1), Π_1 and Π_2 have rank 1 or rank 0. If cointegration exists, the cointegration relationship involves y_t (and therefore x_t itself is not cointegrated).*

This assumption is common for single-equation cointegration tests, see Boswijk (1994) and Banerjee *et al.* (1998). Note that if $\kappa_1 = \kappa_2 = 0$, then Π_1 and Π_2 have rank 0. In that case, z_t is not error-correcting and there is no threshold cointegration between y_t and x_t. We wish to illustrate the condition under which the single-equation model of y_t, given x_t, leads to an efficient test, without losing the information contained in the system model. Model (1) can be explicitly written as (assuming $d_t = 0$ for simplicity)

$$
\begin{pmatrix} \Delta y_t \\ \Delta x_t \end{pmatrix} = \begin{pmatrix} \kappa_{11} \\ \kappa_{21} \end{pmatrix} \theta' z_{t-1} I_{1t} + \begin{pmatrix} \kappa_{12} \\ \kappa_{22} \end{pmatrix} \theta' z_{t-1} I_{2t} + \begin{pmatrix} \Phi_1(L) \\ \Phi_2(L) \end{pmatrix} \Delta z_{t-1} + \begin{pmatrix} \epsilon_{1t} \\ \epsilon_{2t} \end{pmatrix}.
$$

We assume the joint normality of $(\epsilon_{1t}, \epsilon'_{2t})' \sim N(0, \Sigma)$ for convenience but this assumption is not essential. We partition Σ conformably into blocks of Σ_{ij}. Then the conditional model of ϵ_{1t} upon ϵ_{2t} is $\epsilon_{1t} = \alpha' \epsilon_{2t} + u_{1t}$, where $\alpha = \Sigma_{22}^{-1} \Sigma_{21}$ and $var(u_{1t}) = \Sigma_{11} - \Sigma_{12} \Sigma_{22}^{-1} \Sigma_{21}$. Note that u_{1t} is the projection error and is uncorrelated with ϵ_{2t}. The above equation has the following triangular representation of a marginal model for Δx_t, and a conditional model of Δy_t on Δx_t

$$
\begin{aligned}
\Delta y_t &= \kappa_{11}^* \theta' z_{t-1} I_{1t} + \kappa_{12}^* \theta' z_{t-1} I_{2t} + \alpha' \Delta x_t + \Phi_1^*(L) \Delta z_{t-1} + u_{1t} \\
\Delta x_t &= \kappa_{21} \theta' z_{t-1} I_{1t} + \kappa_{22} \theta' z_{t-1} I_{2t} + \Phi_2(L) \Delta z_{t-1} + \epsilon_{2t},
\end{aligned}
$$

where $z_t = (y_t, x'_t)'$, $\kappa_{11}^* = \kappa_{11} - \alpha' \kappa_{21}, \kappa_{12}^* = \kappa_{12} - \alpha' \kappa_{22}$ and $\Phi_1^*(L) = \Phi_1(L) - \alpha' \Phi_2(L)$. If the conditioning variable x_t is assumed to be weakly exogenous to θ, or equivalently $\kappa_{21} = \kappa_{22} = 0$, then the triangular system becomes

$$
\begin{aligned}
\Delta y_t &= \kappa_{11} \theta' z_{t-1} I_{1t} + \kappa_{12} \theta' z_{t-1} I_{2t} + \alpha' \Delta x_t + \Phi_1^*(L) \Delta z_{t-1} + u_{1t} \quad (2) \\
\Delta x_t &= u_{2t},
\end{aligned}
$$

where $u_{2t} \equiv \Phi_2(L) \Delta z_{t-1} + \epsilon_{2t}$ by definition. A similar data generating process (DGP) is considered by Boswijk (1994) in a linear framework. Notice that

weak exogeneity requires that the error correction term be absent in the marginal process of Δx_t. As noted previously, weak exogeneity is found in many empirical works such as Hendry and Ericsson (1991); see also Zivot (2000) for various examples. It is clear that testing $\kappa_1 = \kappa_2 = 0$ in (1) amounts to testing $\kappa_{11} = \kappa_{12} = 0$ in the conditional model in (2) under weak exogeneity. Asymptotic properties of subsequent tests are determined by the joint error process of $u_t = (u_{1t}, u'_{2t})'$, for which the following assumption is made.

Assumption 2. *In model (2), the process $\{u_t\}$ is a weakly stationary process that satisfies the multivariate invariance principle. As $T \to \infty$,*

$$T^{-1/2}(u_1 + \ldots + u_{[Tr]}) \Rightarrow B(r), \quad 0 < r \leq 1, \qquad (3)$$

where $[x]$ denotes the nearest integer close to x, and $\Rightarrow$ denotes the weak convergence of the associated probability measure. $B(r)$ is an $(n+1)$ dimensional vector Brownian motion with a long-run covariance matrix

$$\Omega = \left(\begin{array}{cc} \Omega_1 & \Omega_{12} \\ \Omega_{21} & \Omega_2 \end{array} \right),$$

where Ω is positive definite.

Assumption 2 holds under conditions given in Phillips and Durlauf (1986) and Chan and Wei (1988). Assumption 2 allows for a high degree of temporal dependence in $\{u_t\}$. It is instructive to analyze each block of Ω. First, because u_{1t} is an innovation process relative to $\{x_t,\ z_{t-j},\ j = 1, 2, \ldots\}$, $\Omega_1 = \sigma^2 \equiv E u_{1t}^2$. Second, $\Omega_2 = E u_{2t} u'_{2t} + \sum_{j=1}^{\infty}(E u_{2t} u'_{2t-j} + E u_{2t-j} u'_{2t})$, which is a non-diagonal matrix in general unless $\Phi_2(L) = 0$. Finally and most importantly,

$$\Omega_{12} = E u_{1t} u'_{2t} + \sum_{j=1}^{\infty} E u_{1t} u'_{2t-j} + \sum_{j=1}^{\infty} E u_{2t} u'_{1t-j}, \qquad (4)$$

where $E u_{1t} u'_{2t} = \sum_{j=1}^{\infty} E u_{1t} u'_{2t-j} = 0$, since ϵ_t is serially uncorrelated and u_{1t} is orthogonal to ϵ_{2t}. However, $\sum_{j=1}^{\infty} E u_{2t} u'_{1t-j}$ is not necessarily equal to zero unless Δx_t is strictly exogenous to u_{1t}. Let C denote a lower triangular matrix which is the Cholesky decomposition of Ω. We can write $B(r) = CW(r)$, because $CC' = \Omega$. The explicit expression of C is given by

$$C = \left[\begin{array}{cc} c_{11} & c_{12} \\ 0 & c_{22} \end{array} \right] = \left[\begin{array}{cc} \left(\Omega_1 - \Omega_{12}\Omega_2^{-1}\Omega_{21}\right)^{1/2} & \Omega_{12}\Omega_2^{-1/2} \\ 0 & \Omega_2^{1/2} \end{array} \right] \qquad (5)$$

8

and we let $\rho^2 = \Omega_1^{-1}\Omega_{12}\Omega_{21}\Omega_2^{-1}$. By definition, ρ^2 measures the long-run correlation between u_{1t} and u_{2t}. A necessary and sufficient condition for $\rho^2 = 0$ is that $\Omega_{12} = 0$. As shown by Boswijk (1994), the distribution of the linear BDM test depends on the nuisance parameter ρ unless strict exogeneity holds. Note that strict exogeneity is not assumed in this paper. As such, in particular, a special correction is needed for the BDM test. We will talk more about this issue later.

3.2 Indicators

In this paper, we consider two different indicator functions that utilize the change or deviation from the long-run equilibrium as the threshold variable. The deviation from the long-run equilibrium is obtained from the OLS residual e_t in the first stage regression

$$y_t = \widehat{\gamma}' x_t + e_t. \tag{6}$$

By definition, $e_t = y_t - \widehat{\gamma}' x_t$ is the estimated error-correction term. Without (6) we have to grid search the cointegrating vector as well as the threshold value, which becomes infeasible when the cointegrating vector is high-dimensional. An efficient algorithm for a low-dimensional grid search is provided by Hansen and Seo (2002). Notice that $\widehat{\theta} = (1, -\widehat{\gamma}')'$ will be a super-consistent estimate of the cointegrating vector if y_t and x_t are cointegrated. Otherwise, the above regression (6) is spurious, implying that $e_t = O_p(T^{1/2})$ is nonstationary.

We first consider the following indicator function (which we call "*Indicator A*")

$$I_{1t} = I(e_{t-d} < e_{t-d}^*(\tau)), \text{ and } I_{2t} = 1 - I_{1t}, \tag{7}$$

where $I_{1t} = 1$ if $e_{t-d} < e_{t-d}^*(\tau)$ and $I_{1t} = 0$ otherwise. In this indicator, the threshold variable e_{t-d} is nonstationary under the null hypothesis of no cointegration; $d \geq 1$ denotes the delay parameter. To focus on our main issues, d is assumed to be known *a priori*. We let $d = 1$ in this paper, but extension to cases with an unknown delay is straightforward. The threshold value, denoted by $e_{t-d}^*(\tau)$, is the τ-th percentile element of $\{e_{t-d}^*\}$ which is the ranked process of $\{e_{t-d}\}$ in an ascending order. In other words, the threshold value is given by the τ-th percentile of the empirical distribution of e_{t-d}. This specification is consistent with the case where the threshold variable is stationary. Thus, we always seek threshold values in the ranked

series of the threshold variable (it is meaningless to grid search the threshold outside the empirical domain of threshold variables). See Hansen (1997) for more discussion on the grid search. It can be shown that the following equivalence holds

$$
\begin{aligned}
I(e_{t-d} \quad < \quad & e_{t-d}^*(\tau)) = I(\sigma_e^{-1} T^{-1/2} e_{t-d} < \sigma_e^{-1} T^{-1/2} e_{t-d}^*(\tau)) \qquad (8) \\
& \Rightarrow \quad I(W(r) < W^*(\tau)),
\end{aligned}
$$

where σ_e^2 is the long-run variance of e_{t-d}, and W^* is the Brownian motion using a sorted time series. We note that the above expression involves the threshold percentile parameter τ, which is a nuisance parameter. An indicator function using a nonstationary threshold variable was previously considered in Seo (2006). However, our treatment of the threshold parameter and the relevant asymptotic distribution will differ from that in Seo (2006), who utilized the following result

$$
I(e_{t-d} < c) = I(\sigma_e^{-1} T^{-1/2} e_{t-d} < \sigma_e^{-1} T^{-1/2} c) \Rightarrow I(W(r) < 0).
$$

Seo (2006) assumed that the threshold parameter c is a fixed value. Under this assumption, he suggests that the rescaled parameter $c^* (= \sigma_e^{-1} T^{-1/2} c)$ will approach zero asymptotically as $T \to \infty$. Thus, the threshold parameter becomes irrelevant in the limiting distribution of his test. In Seo's suggested testing procedure, a grid search is again adopted over the threshold value to obtain the sup Wald statistic. It is also suggested that his supW statistic has the same asymptotic null distribution as the Wald statistic that is constructed by fixing the threshold parameter at a certain value. However, as noted in Seo (2006), his proposed test using the asymptotic distribution does not perform well in finite samples so he relies on a bootstrap procedure. It is useful to note that our tests have different asymptotic properties. First, our tests do not rely on the assumption that the threshold parameter is fixed or irrelevant in the asymptotic distribution. Thus, we allow for the possibility that $c = O_p(T^{1/2})$ when *Indicator A* is used. Second, the distribution of a supW type statistic is different from that of a Wald statistic using fixed threshold parameters. This is so since the supW statistic is an order statistic. Third, using the asymptotic distribution our tests do not exhibit any serious size distortions. Lastly, in our testing procedure, all nuisance parameters are identified and controlled for. As such, our tests do not require bootstrapping.

In our test, the parameter τ can be viewed as a threshold percentile (or more appropriately, a threshold index) rather than a threshold value. The

point is, when a threshold variable is nonstationary, it will be desirable not to restrict the threshold parameter to a fixed value. Treating τ as a percentile parameter has two theoretical advantages. First, it yields an asymptotic distribution that depends on the range of τ. The threshold parameter is estimated via a grid search to obtain a supW type statistic. That means that the information contained in the grid search is preserved and utilized in the asymptotic theory. The presence of a threshold percentile parameter does not pose a problem, since this nuisance parameter appears only under the alternative and will be controlled using the sup Wald type tests. Second, τ always lies in a compact set $\tau \in \Theta = [\underline{\tau}, \overline{\tau}]$, where $\underline{\tau}$ and $\overline{\tau}$ denote the lower and upper bound of Θ, respectively. Thus, the asymptotic distributions of our new tests can be derived by applying convergences with respect to the uniform metric on the compact parameter space Θ. Following Andrews (1993), we set $\Theta = [0.15, 0.85]$ to avoid a divergent asymptotic distribution.

We next consider the "momentum" type indicator function, which is based on a stationary threshold variable. A simple way to achieve stationarity is by taking the first-order difference of the long-run equilibrium error. Then, we use the first differenced series as the threshold variable. We denote this as "*Indicator B*".[2] The MTAR model utilizes the first difference of e_{t-d}. Hence, *Indicator B* is defined as: $I_{1t} = I(\Delta e_{t-d} < \Delta e_{t-d}^*(\tau))$, and $\widetilde{I}_{2t} = 1 - \widetilde{I}_{1t}$ where $\Delta e_{t-d}^*(\tau)$ denotes the τ-th percentile element of the empirical distribution of Δe_{t-d}. This specification indicates that the regime switching depends on whether the difference of the error correction term is less than some threshold value. The required asymptotic result is simpler. Following Caner and Hansen (2001), we can have

$$I(\Delta e_{t-d} < \Delta e_{t-d}^*(\tau)) \Rightarrow I(U(r) < \tau), \qquad (9)$$

where $U(r)$ is a univariate uniform process on $[0,1]$. Since the true DGP is seldom known, both *Indicators A* and *B* are possible. These two indicators were previously considered in Enders and Siklos (2001) for their EG type tests. But they did not consider a percentile threshold parameter and required asymptotic theory differs from ours.

---

[2]The word "momentum" stems from the terminology of Enders and Granger (1998). Enders and Granger (1998) considered this case for unit root tests, and Caner and Hansen (2001) provide the asymptotic theory for unit root tests with this indicator function. We follow Enders and Siklos (2001), who considered the momentum threshold autoregressive (MTAR) model for threshold cointegration tests.

Note that we focus on a two-regime model in this paper. However, the basic model can be easily extended to multiple-regimes, including especially the so-called band threshold models of Balke and Fomby (1997), by redefining the indicators. So far, selecting the number of regimes can be guided by a priori information and economic sense. There is no formal test for the null of k regimes against $k+1$ regimes in a threshold model. Nevertheless, readers should be aware that the grid search will increase exponentially as the number of regimes increases.

4 Single-equation tests for threshold cointegration

We can consider possibly two versions of single-equation tests; one approach uses a pre-specified cointegrating vector (e.g. Kremers *et al.* (1992) and Zivot (2000)), and the other is based on the estimated cointegrating vector. The former approach is proper if the cointegrating vector is known *a priori* and may yield better power than the latter approach, since the cointegrating vector does not need to be estimated. The extension of single-equation tests with a pre-specified cointegrating vector to threshold models is also considered by Li (2006). However, in many empirical applications the cointegrating vector is unknown. In this case, the tests proposed in this paper are more appealing. We extend two linear single-equation tests to test for threshold cointegration. The first—the BO type test, is due to Boswijk (1994), who suggests testing the coefficients of both y_{t-1} and x_{t-1} in the test regression. In contrast, the second-the BDM type test of Banerjee *et al.* (1998), is concerned only with the coefficient of y_{t-1}. Banerjee *et al.* (1998) suggest adding leads of Δx_t to their regression so that the asymptotic results are valid in the absence of strict exogeneity.

4.1 Threshold BO test

The threshold BO test is based on the regression in (2) augmented by the deterministic term d_t, which is usually defined as $d_t = 1$ or $d_t = (1, t)'$

$$\Delta y_t = \pi' d_t + B_1' z_{t-1} I_{1t} + B_2' z_{t-1} I_{2t} + c' q_t + u_{1t}, \qquad (10)$$

where $z_{t-1} = (y_{t-1}, x_{t-1}')'$, and $q_t = (\Delta x_t', \Delta z_{t-1}', \ldots, \Delta z_{t-p}')'$. Note that q_t is stationary by design. Assuming weak exogeneity, we have $B_1 = \kappa_{11}\theta$,

$B_2 = \kappa_{12}\theta$, and $\theta = (1, -\gamma')'$, which is unknown. The indicators in (10) can be either $I_{1t} = I(e_{t-d} < e^*_{t-d}(\tau))$ (*Indicator A*), or $I_{1t} = I(\Delta e_{t-d} < \Delta e^*_{t-d}(\tau))$ (*Indicator B*), and $I_{2t} = 1 - I_{1t}$, where e_t is the residual of (6). As usual, u_{1t} is uncorrelated with q_t. We want to test the null hypothesis of no threshold cointegration

$$H_0 : B_1 = B_2 = 0,$$

against the alternative hypothesis H_1: H_0 is not true. The test of the above hypothesis is nonstandard in the sense that the parameter τ is not identified under H_0. Hypothesis testing involving the nuisance parameters identified only under the alternative was first investigated by Davies (1977 and 1987). Further discussion of the Davies problem can be found in Andrews and Ploberger (1994) and Hansen (1996). The Davies problem is resolved by first constructing a test statistic for each τ, then the final test is a continuous functional of the sequence of statistics. More specifically, let $\widehat{B}(\tau) = (\widehat{B}_1(\tau)', \widehat{B}_2(\tau)')'$ be the OLS estimate of $B = (B_1', B_2')'$ for given τ in (10), and $\widehat{V}(\widehat{B}(\tau))$ be the OLS variance estimate. We denote the test statistics based on *Indicator A* as BO, BO_μ and BO_t, which are associated with $d_t = 0, d_t = 1$ and $d_t = (1, t)'$ respectively. We denote the corresponding test statistics based on *Indicator B* as $\widetilde{BO}, \widetilde{BO}_\mu$ and $\widetilde{BO}_t$, respectively. They are specified as follows

$$BO, BO_\mu, BO_t, \widetilde{BO}, \widetilde{BO}_\mu, \widetilde{BO}_t \equiv \sup_{\tau \in \Theta} \widehat{B}(\tau)'\widehat{V}(\widehat{B}(\tau))^{-1}\widehat{B}(\tau), \qquad (11)$$

where Θ denotes the parameter space of τ. Because I_{1t} and I_{2t} are orthogonal, it follows that the test statistics in (11) can be rewritten as

$$\sup_{\tau \in \Theta} \widehat{B}_1(\tau)'\widehat{V}(\widehat{B}_1(\tau))^{-1}\widehat{B}_1(\tau) + \widehat{B}_2(\tau)'\widehat{V}(\widehat{B}_2(\tau))^{-1}\widehat{B}_2(\tau).$$

That is, the test for a given threshold can be calculated as the sum of two separate Wald statistics for $B_j = 0$, $j = 1, 2$. The functional supremum is used by (11) for its simplicity. Tests involving functionals, such as the exponential test or average test, are proposed by Andrews and Ploberger (1994). In large samples, these tests may be superior due to their greater power. Since the sample size of a macroeconomic research topic can be relatively small, we stick with the simple supremum test. As is common in the literature, we assume that the deterministic term d_t is absent under the null hypothesis, but the asymptotic distribution of the tests is invariant to the coefficients of d_t in the DGP.

We introduce some additional notation to succinctly express our asymptotic results. Let $P(r)$ and $G(r)$ be two stochastic processes. First, the continuous residual process $R_{P,G}(r)$ is defined as the residual of the continuous time regression $P(r) = \widehat{\phi}' G(r) + R_{P,G}(r)$, where $\widehat{\phi}$ solves $\min_\phi \int ||P(r) - \phi' G(r)||^2 dr$. Explicitly,

$$R_{P,G}(r) = P(r) - \left(\int G(r)P(r)dr \right)' \left(\int G(r)G(r)'dr \right)^{-1} G(r).$$

We let $W_3(r) = R_{W1,W2}(r)$, where $W = (W_1, W_2')'$ is a standard $(n+1)$ dimensional vector Brownian motion. Let W_3^* be the sorted W_3 and $U(r)$ be a uniform process on $[0,1]$. Then, we define

$$A(H) = \sup_{\tau \in \Theta} \sum_{j=1}^{2} \int H \cdot I_j dW_1 \left(\int H \cdot I_j H' \right)^{-1} \int H \cdot I_j dW_1$$

$$\widetilde{A}(H) = \sup_{\tau \in \Theta} \sum_{j=1}^{2} \int H \cdot \widetilde{I}_j dW_1 \left(\int H \cdot \widetilde{I}_j H' \right)^{-1} \int H \cdot \widetilde{I}_j dW_1$$

where $I_1 = I(W_3 < W_3^*(\tau)), I_2 = I(W_3 \geq W_3^*(\tau))$ and $\widetilde{I}_1 = I(U < \tau), \widetilde{I}_2 = I(U \geq \tau)$ for given τ. The parameter space Θ is set as $\Theta = [0.15, 0.85]$. The limit distribution of the threshold BO test (11) under H_0 is given in the following theorem.

Theorem 1 *Let $W(r) = (W_1(r), W_2(r)')'$ be an $(n+1)$-dimensional standard vector Brownian motion. Also, we let $W_\mu(r) = R_{W,1}(r)$, and $W_t(r) = R_{W,(1,r)}(r)$. Under the null hypothesis H_0 and Assumptions 1-2, as $T \to \infty$*

$$BO \Rightarrow A(W), \quad BO_\mu \Rightarrow A(W_\mu), \quad BO_t \Rightarrow A(W_t)$$
$$\widetilde{BO} \Rightarrow \widetilde{A}(W), \quad \widetilde{BO}_\mu \Rightarrow \widetilde{A}(W_\mu), \quad \widetilde{BO}_t \Rightarrow \widetilde{A}(W_t).$$

Some remarks here are helpful. Most importantly, there is no nuisance parameter in the null distributions of the BO threshold cointegration tests. In other words, all distributions are similar. Thus, critical values can be tabulated for different cases when the asymptotic null distributions depend on (i) the dimension of $W(r)$, which is the same as the dimension of z_t,

(ii) different types of indicators, and (iii) the deterministic term d_t. Note that the distributions are free of any other nuisance parameters. Actually, our asymptotic result is consistent with the result in Theorem 1 of Boswijk (1994) showing that all other nuisance parameters are controlled for, while we additionally take into account of the presence of threshold parameters and their indicators.

4.2 Threshold BDM test

The threshold BDM test is an alternative way to test the same null hypothesis of no threshold cointegration. Recall that in (10), $B_1 = \kappa_{11}\theta$, $B_2 = \kappa_{12}\theta$ and $\theta = (1, -\gamma')'$ is the cointegrating vector. Therefore, testing $B_1 = B_2 = 0$ is equivalent to testing $\kappa_{11} = \kappa_{12} = 0$, which is the BDM test. Therefore, one difference between the two tests is that the BDM test is concerned with the coefficients of the subvector of z_{t-1}. In order to discuss the asymptotic distribution of the threshold BDM test, the following notations are introduced.

Definition 1. *Let ξ_t be an $(n+1)$ dimensional random walk process such that $\Delta\xi_t \sim iid(0, I)$. Define an $(n+1)$ dimensional $V^1(r, \tau)$, $V^2(r, \tau), \widetilde{V}^1(r, \tau)$ and $\widetilde{V}^2(r, \tau)$ as stochastic processes on $[0,1]^2$, such that $T^{-1/2}\xi_{t-1}I_{1t} \Rightarrow V^1(r, \tau)$, $T^{-1/2}\xi_{t-1}I_{2t} \Rightarrow V^2(r, \tau)$, $T^{-1/2}\xi_{t-1}\widetilde{I}_{1t} \Rightarrow \widetilde{V}^1(r, \tau)$ and $T^{-1/2}\xi_{t-1}\widetilde{I}_{2t} \Rightarrow \widetilde{V}^2(r, \tau)$.*

Here, V processes can be treated as stochastic limits of the product of a random walk process and an indicator. They are not necessarily a vector version of the two-parameter Brownian motion introduced by Caner and Hansen (2001), who assume a stationary threshold variable. Thus, our asymptotic distribution differs from that of Caner and Hansen (2001). The following Corollary is to explain why a BDM test based on regression (10) is limited in practice.

Corollary 1. *Use Indicator A and consider the squared OLS t-ratios of $t^2(\widehat{\kappa}_{11}(\tau))$ and $t^2(\widehat{\kappa}_{12}(\tau))$ in regression (10), where $\widehat{\kappa}_{11}(\tau)$ and $\widehat{\kappa}_{12}(\tau)$ are estimated coefficients of y_{t-1} in each of two regimes, respectively, for given τ. Then, under H_0 and Assumptions 1-2, as $T \to \infty$*

$$t^2(\widehat{\kappa}_{1j}(\tau)) \Rightarrow \int Q_j dW^\rho \left(\int Q_j Q_j'\right)^{-1} \int Q_j dW^\rho, \quad j = 1, 2,$$

where $Q_j(r) = R_{V_1^j, V_2^j}(r), V_1^j$ and V_2^j are the first and remaining columns of $V^j(r, \tau)$ $j = 1, 2$, respectively, and $W^\rho = (1 - \rho^2)^{1/2} W_1 + \rho W_2$.

Corollary 1 shows that distributions of $t^2(\widehat{\kappa}_{11})$ and $t^2(\widehat{\kappa}_{12})$ involve the nuisance parameter ρ^2, which is the long-run cross correlation between u_{1t} and u_{2t}. Hence in general these distributions are not similar. There are two approaches to conduct inference based on squared t-ratios. The first approach is to tabulate critical values of $t^2(\widehat{\kappa}_{11})$ and $t^2(\widehat{\kappa}_{12})$ for each ρ, as in Hansen (1995). This approach requires a kernel-based semiparametric estimation of ρ for the real data, and then comparing the test statistic to the corresponding critical value. The second approach is to follow Saikkonen (1991), and include leads of Δx_t to diagonalize u_{1t} and u_{2t}. Our threshold BDM test will adopt the second approach. But, each has limitations. The first approach requires choosing the bandwidth and proper kernels, while the second entails selection of the lead length.

Therefore, we base our threshold BDM test on (10) augmented with leads of Δx_t as follows

$$\Delta y_t = \pi' d_t + \kappa_{11} y_{t-1} I_{1t} + \kappa_{12} y_{t-1} I_{2t} + b_{11} x_{t-1} I_{1t} + b_{12} x_{t-1} I_{2t} + c^{*'} q_t^* + v_t, \quad (12)$$

where $q_t^* = (\Delta x_t', \Delta x_{t+1}', \dots, \Delta x_{t+m}', \Delta z_{t-1}', \dots, \Delta z_{t-p}')'$. Notice that compared to (10), q_t^* in (12) adds leads of Δx_t. Following Saikkonen (1991), the maximal value m for the lead terms are chosen such that the new error process v_t in (12) satisfies $E v_t u_{2t+j} = 0$ for all j and the long-run correlation between v_t and u_{2t} becomes zero. If we can define the long-run covariance matrix Ω^* for $(v_t, u_{2t})'$ similarly as in (5), then Ω^* is diagonal.

Intuitively, v_t can be viewed as the projection error of u_{1t} upon leads of u_{2t}. Thus, by construction, no effect of leads of u_{2t} remains in v_t. Only a finite number of leads are needed, because the joint stationarity of u_{1t} and u_{2t} implies that the effect of u_{2t+j} on u_{1t} dies out quickly as j increases. Hence, it is reasonable to assume that $E u_{1t} u_{2t+j} \simeq 0$ for $j > q$ and q is big enough. In practice, choosing q can be guided by the extent to which $\widehat{u}_{1t-j}$ is correlated with $u_{2t} = \Delta x_t$, where $\widehat{u}_t$ is the residual of (10). If the correlation is weak, then $q = 1$ or 2 is recommended. Our objective is to test the null hypothesis of no cointegration

$$H_0^* : \kappa_{11} = \kappa_{12} = 0,$$

from (12) against the alternative hypothesis that H_1^*: H_0^* is not true. Let $\widehat{\kappa}(\tau) = (\widehat{\kappa}_{11}(\tau), \widehat{\kappa}_{12}(\tau))'$ be the OLS estimate of $\kappa = (\kappa_{11}, \kappa_{12})'$ for given

τ in (12) and $\widehat{V}(\widehat{\kappa}(\tau))$ be the OLS variance estimate. The test statistics BDM, BDM_μ and BDM_t are based on *Indicator A;* they are associated with $d_t = 0, d_t = 1$ and $d_t = (1, t)'$, respectively. The test statistics $\widetilde{BDM}, \widetilde{BDM}_\mu$ and $\widetilde{BDM}_t$ are based on *Indicator B* in a similar manner. These statistics are specified as follows

$$BDM, BDM_\mu, BDM_t, \widetilde{BDM}, \widetilde{BDM}_\mu, \widetilde{BDM}_t \equiv \sup_{\tau \in \Theta} \widehat{\kappa}(\tau)' \widehat{V}(\widehat{\kappa}(\tau))^{-1} \widehat{\kappa}(\tau). \tag{13}$$

Because I_{1t} is orthogonal to I_{2t}, the test statistics in (13) can be equivalently calculated as

$$\sup_{\tau \in \Theta} t^2(\widehat{\kappa}_{11}(\tau)) + t^2(\widehat{\kappa}_{12}(\tau)),$$

where $t^2(\widehat{\kappa}_{11}(\tau))$ and $t^2(\widehat{\kappa}_{12}(\tau))$ are defined in Corollary 1.

The asymptotic distributions of the threshold BDM tests can be derived using a partitioned regression. We decompose the expressions in Definition 1 into $V^j(r, \tau) = (V_1^j, V_2^{j\prime})$ and $\widetilde{V}^j(r, \tau) = (\widetilde{V}_1^j, \widetilde{V}_2^{j\prime})$ $j = 1, 2$. We also let $V^{j\mu}(r) = (V_1^{j\mu}, V_2^{j\mu\prime})'$, $V^{jt}(r) = (V_1^{jt}, V_2^{jt\prime})'$, $\widetilde{V}^{j\mu}(r) = (\widetilde{V}_1^{j\mu}, \widetilde{V}_2^{j\mu\prime})'$, and $\widetilde{V}^{jt}(r) = (\widetilde{V}_1^{jt}, \widetilde{V}_2^{jt\prime})'$. Then, we express these terms as $V^{j\mu}(r) = R_{V^j,1}(r)$, $V^{jt}(r) = R_{V^j,(1,r)}(r)$, $\widetilde{V}^{j\mu}(r) = R_{\widetilde{V}^j,1}(r)$, and $\widetilde{V}^{jt}(r) = R_{\widetilde{V}^j,(1,r)}(r)$, respectively. We utilize the following functional

$$A_1(H_1, H_2) = \sup_{\tau \in \Theta} \sum_{j=1}^2 \int H_j dW_1 \left(\int H_j H_j' \right)^{-1} \int H_j dW_1.$$

Theorem 2 *Under the null hypothesis H_0^* and Assumptions 1-2, we have as $T \to \infty$*

$$BDM \ \Rightarrow \ A_1(Q_1, Q_2), \quad BDM_\mu \Rightarrow A_1(Q_1^\mu, Q_2^\mu), \quad BDM_t \Rightarrow A_1(Q_1^t, Q_2^t)$$
$$\widetilde{BDM} \ \Rightarrow \ A_1(\widetilde{Q}_1, \widetilde{Q}_2), \quad \widetilde{BDM}_\mu \Rightarrow A_1(\widetilde{Q}_1^\mu, \widetilde{Q}_2^\mu), \quad \widetilde{BDM}_t \Rightarrow A_1(\widetilde{Q}_1^t, \widetilde{Q}_2^t),$$

where $Q_j(r) = R_{V_1^j, V_2^j}(r)$, $Q_j^\mu(r) = R_{V_1^{j\mu}, V_2^{j\mu}}(r)$, $Q_j^t(r) = R_{V_1^{jt}, V_2^{jt}}(r)$, $\widetilde{Q}_j(r) = R_{\widetilde{V}_1^j, \widetilde{V}_2^j}(r)$, $\widetilde{Q}_j^\mu(r) = R_{\widetilde{V}_1^{j\mu}, \widetilde{V}_2^{j\mu}}(r)$ and $\widetilde{Q}_j^t(r) = R_{\widetilde{V}_1^{jt}, \widetilde{V}_2^{jt}}(r)$ for $j = 1, 2$.

Theorem 2 shows that after including leads of Δx_t, the distributions of the threshold BDM tests are asymptotically similar. In the next section, we will investigate finite-sample performance of the BO and BDM threshold cointegration tests.

5   Monte Carlo Experiment

Asymptotic critical values of the threshold BO and BDM threshold cointegration tests are reported in Table 1. These critical values are computed by simulating the asymptotic distributions of the relevant statistics given in Theorem 1 and Theorem 2. We use the asymptotic sample size $T = 1,000$ with $10,000$ replications. The parameter space is set as $\Theta = [0.15, 0.85]$. n denotes the number of conditioning variables in Table 1. In general, the critical values with $d_t \neq 0$ are greater than those with $d_t = 0$.

[Insert Table 1]

In Table 2, we report the sizes of the BO and BDM threshold cointegration tests for finite samples. The DGP is given by (1) with $d_t = 0, n = 1$, $\Pi_1 = \Pi_2 = 0$ (so the null hypothesis holds), and $\epsilon_t \sim iid(0, E\epsilon_t\epsilon_t')$. Explicitly, we consider

$$
\begin{aligned}
\Delta y_t &= \phi_1 \Delta y_{t-1} + \epsilon_{1t} \\
\Delta x_t &= \phi_2 \Delta y_{t-1} + \phi_3 \Delta x_{t-1} + \epsilon_{2t}
\end{aligned}
\tag{14}
$$

where $E\epsilon_{1t}\epsilon_{2t} = \rho_1$ and $E\epsilon_{1t}^2 = E\epsilon_{2t}^2 = 1$. For DGP (14), we can show that

$$
\begin{aligned}
u_{1t} &\equiv \epsilon_{1t} - \rho_1 \epsilon_{2t} \\
u_{2t} &\equiv \Delta x_t = (1 - \phi_3 L)^{-1} \left[\phi_2 (1 - \phi_1 L)^{-1} \epsilon_{1t-1} + \epsilon_{2t} \right].
\end{aligned}
\tag{15}
$$

It is clear from (15) that ϕ_1 and ϕ_2 are the key parameters yielding a nonzero value of ρ. Hence, we let $\phi_3 = 0$ in (14) without a loss of generality. From (15), it follows that $Eu_{1t}\epsilon_{1t} = 1 - \rho_1^2$ and $Eu_{1t}\epsilon_{2t} = 0$. Some algebra demonstrates that $Eu_{1t}u_{2t-j} = 0$ for all $j \geq 0$, $Eu_{2t}u_{1t-1} = \phi_2(1 - \rho_1^2)$ and $Eu_{2t}u_{1t-j} = \phi_2\phi_1^{j-1}(1-\rho_1^2)$ for all $j \geq 2$. Then it follows from (4) that $\Omega_{12} = \phi_2(1-\rho_1^2)/\phi_1$. We next let $\Theta^{dgp} = (\rho_1, \phi_2, \phi_1)$. Similarly, Ω_2 can be shown to be a more intricate function of Θ^{dgp}. However, for our purposes, it suffices to know that ρ is a function of Θ^{dgp}. If we can show that test sizes are invariant to the parameters in Θ^{dgp}, which is the determinant of ρ, then it is safe to say that the sizes are invariant to ρ. Hansen (1995) examines this issue from a different perspective and estimates ρ, using kernel-based methods, for each of his simulated data sets. From this procedure, he can show whether the sizes are invariant to the estimated value of ρ, say, $\widehat{\rho}$. While it seems difficult

to directly control ρ with any DGP, we can still control the value of ρ with our DGP indirectly, but in an effective and time efficient manner.

The BO and BDM tests are based on the testing regression (10) and (12) respectively. Thus, we use different indicators for the $\widetilde{BO}$ and $\widetilde{BDM}$ tests but the testing regression remains the same. For completeness, we also consider the $BDMN$ and $\widetilde{BDMN}$ tests, which are supposedly the BDM and $\widetilde{BDM}$ tests but *without* using the lead of Δx_t in (12). We expect that the $BDMN$ and $\widetilde{BDMN}$ tests can suffer size distortions due to a lack of correction for violating strict exogeneity and we examine how serious the problem will be. We let $p = 1$ in (10) and $p = q = 1$ in (12). Moreover, we set $d_t = 0$ in (10) and (12), since a similar pattern is found for the models with nonzero deterministic terms. We consider different combinations of ρ_1, ϕ_2 and ϕ_1. Note that $\Omega_{12} = 0$ (so $\rho = 0$) when $\phi_2 = 0, \phi_1 = 0.4$. In the other two cases using $(\phi_2 = 0.5, \phi_1 = 0.1)$ and $(\phi_2 = 0.5, \phi_1 = 0.4)$, we have $\rho \neq 0$. In all simulations, we use the 5% nominal size with 2,000 replications.

[Insert Table 2]

We can summarize our main findings as follows. First, no obvious size distortion is found in our simulations using different combinations of ρ_1, ϕ_2 and ϕ_1. This result is consistent with Theorems 1 and 2, which state that our test distributions are free of ρ. Second, the size distortion in finite samples is alleviated (at a slow speed) as the sample size increases. Overall, our new tests tend to over-reject the null only mildly. Third, contrary to our initial expectation, excluding the leads of Δx_t in (12) does not introduce any serious size distortions in the BDM tests. This fact indicates that the BDM type tests omitting lead terms do not suffer much in the threshold models we consider with our DGPs. Fourth, in terms of size distortions, there is no apparent difference between the BO and BDM tests or the tests using different indicators. Finally, the size distortions of our proposed tests are much smaller than in Seo's test (see Table 2 in Seo, 2006). Note that this is so even though we do not adopt a bootstrap method.

The power of our proposed tests is investigated using the following DGP

$$\begin{align}
\Delta y_t &= -0.1\varepsilon_{t-1}I_{1t} + k\varepsilon_{t-1}I_{2t} + \phi_1\Delta y_{t-1} + \epsilon_{1t} \qquad (16) \\
\Delta x_t &= k_2\varepsilon_{t-1}I_{1t} + k_3\varepsilon_{t-1}I_{2t} + \phi_2\Delta y_{t-1} + \epsilon_{2t},
\end{align}$$

where $\varepsilon_{t-1} = y_{t-1} - \gamma'x_{t-1}$, $I_{1t} = I(\varepsilon_{t-1} < c)$ and $I_{12} = 1 - I_{1t}$. Some remarks here about the DGP (16) will be helpful. First, by assuming that Δx_t is not

error-correcting (so $k_2 = k_3 = 0$) in (16), we impose weak exogeneity of x_t in the cointegrating vector. Second, we consider only the case where regime switching is based on the level (rather than the difference) of ε_t. Accordingly, only the BO and BDM type tests are considered in power simulations. In order to focus on the key parameters of τ, γ and k, we let $E\epsilon_{1t}\epsilon_{2t} = \rho_1 = 0.7$ and $\phi_2 = 0.5$ in (16). Note that γ and k measure the long-run relationship among variables and the short-run adjustment speed, respectively.

We have compared the power of our tests to the tests of Enders and Siklos (2001), and denote their tests as the ES test in Table 3. To evaluate the ES test in our context, we first run the regression (6) and obtain the residual e_t. Next, we run the regression

$$\Delta e_t = \varphi_1 e_{t-1} I_{1t} + \varphi_2 e_{t-1} I_{2t} + \varphi_3 \Delta e_{t-1} + v_t, \tag{17}$$

where $I_{1t} = I(e_{t-1} < c)$ and $I_{2t} = 1 - I_{1t}$. Then, the ES test is given as $\sup F(\tau)$, where $F(\tau)$ is the F statistic for $\varphi_1 = \varphi_2 = 0$ in (17) for a given τ. We define τ as the percentile parameter of the rescaled threshold value of c. Since the ES test is the extended Engle-Granger residual-based cointegration test in the threshold framework, in principle, the ES test can lose power when it implicitly imposes the common factor restriction as discussed in Section 2.

[Insert Table 3]

From Table 3, we first observe that the power of our tests increases as the sample size increases, implying that the proposed tests are consistent. Second, it is clear that the BO and BDM tests dominate the ES test. In addition, there is some weak evidence that the BO test outperforms the BDM test. This comes as no surprise, considering that the BDM test requires adding additional regressors in the form of leads of Δx_t. We can say that the BO test is based on a more parsimonious model. As a result, we expect that the power gain of the BO test over the BDM test will increase as the long-run correlation increases (because more leads of Δx_t are needed for the BDM test in this case). Third, when c changes from 0 to 2, the power of all tests worsens. This is because more observations cluster into one regime when τ increases. In other words, it is less likely for any regime switching to occur as c increases. As c increases, less information is left in the other regime for estimation and inference. Fourth, it is clear that the power depends on ϕ_1, γ and k. That is, even though a power function is not explicitly derived, we can

see that the power is an increasing function of ϕ_1 and a decreasing function of γ and k. The reason for a negative relationship between the power and k is intuitive-the deviation from the null hypothesis increases as k decreases, so the power should also increase. Note that there is no threshold effect when $\gamma = 0.2$ and $k = -0.1$. This explains why the power is invariant to changes in the threshold values. In short, the BO test performs better than any of other tests in terms of size and power. Given this, we recommend using the BO threshold cointegration test for empirical research.

6   Application Example

In this section, we implement and examine an application of our new ADL threshold cointegration tests. We are interested in the relationship between U.S. consumption (y_2) and GDP (y_1). A textbook model implies the following long-run relationship at time t :

$$y_{2t} = k y_{1t}^{\alpha} e^{\epsilon_t}, \quad (0 < k < 1),$$

where k is the marginal propensity to consume, α is a parameter and ϵ_t captures other determinants of consumption. Taking logs and rearranging gives

$$\ln y_{2t} - \alpha \ln y_{1t} = \ln k + \epsilon_t.$$

This equation states that if ϵ_t is stationary, and both consumption and GDP are nonstationary, then the cointegrating vector for consumption and GDP is given by $(1, -\alpha)$. We are interested in examining whether the short-run adjustment speed toward the long-run equilibrium is the same regardless of the previous state of the system. The above equation says little about the short-run dynamics. However, the short-run dynamics can be represented by the threshold ADL model. We will estimate the model using quarterly U.S. data obtained from Economic Data–FRED at the web site of the Federal Reserve Bank of St. Louis. The data that we employ covers the time period 1947:Q1–2006:Q1, which consists of 237 observations and is not seasonally adjusted. Figure 1 plots the time series of the natural log of consumption and GDP. Both time series appear nonstationary, but tend to move together; thus, the deviation between the two series is likely to converge to a constant number and suggests cointegration.

In this paper, we want to relax the assumption of a pre-specified cointegrating vector and consider estimating the cointegrating vector. Therefore,

the threshold ADL model is appropriate and threshold cointegration tests are suggested. First the estimated cointegrating vector is given by the following regression

$$\ln y_{2t} - 1.0239 \ln y_{1t} = -0.6213 + e_t.$$

Note that the coefficient of $\ln y_{1t}$ is slightly different from 1. Figure 1 also plots the estimated error correction term given by $e_t = \ln y_{2t} - 1.0239 \ln y_{1t} + 0.6213$. From Figure 1, it is safe to conclude that the error correction term is free of a deterministic trend. Two indicators, *Indicator A* with $I_t = I(e_{t-1} < e_{t-1}^*(\tau))$ and *Indicator B* with $\widetilde{I}_t = I(\Delta e_{t-1} < \Delta e_{t-1}^*(\tau))$, are considered. Specifically, the following threshold ADL regression is fitted:

$$\Delta y_{2t} = \beta_0 + \beta_1 y_{2t-1} I_t + \beta_2 y_{2t-1}(1 - I_t) + \beta_3 y_{1t-1} I_t + \beta_4 y_{1t-1}(1 - I_t) + \\ \beta_5 \Delta y_{1t} + \beta_6 \Delta y_{1t-1} + \beta_7 \Delta y_{2t-1} + e_t,$$

where I_t can be replaced with $\widetilde{I}_t$ if *Indicator B* is adopted. Most important, the adjustment speeds toward the long-run equilibrium, as measured by β_i ($i = 1, 2, 3, 4$), are allowed to vary in the threshold model. Thus, the conventional ADL model is a special case of the threshold ADL model when $\beta_1 = \beta_2$ and $\beta_3 = \beta_4$. Notice that only one lag of Δy_{1t} and Δy_{2t} is included in the regression following the the parsimony principle. Here, the lag-selection is guided by the partial autocorrelation function (PACF) (not presented in the paper) of Δy_{2t}. In our case, the PACF indicates little serial correlation after one lag. Readers should keep in the mind, however, that the final results depend on the selection of RHS variables. We wish to point out that Δy_{1t} must be present as a regressor, since the single-equation model is conditioned on y_{1t}.

To focus on the main issues, we assume here that the regime switching depends on e_{t-1}. In a more general model, the threshold variable could be e_{t-d}, where the delay lag d is estimated by a grid search. The threshold variable e_{t-1} is nonstationary under the null hypothesis of no threshold cointegration. The threshold value reported in Table 4 is obtained by maximizing the Wald statistic (threshold Boswijk) for the null

$$H_0 : \beta_1 = \beta_2 = \beta_3 = \beta_4 = 0$$

between the lower and upper 15% percentiles of the sorted series e_{t-1}. Alternatively, the threshold value can be determined by minimizing the residual sum of squares. The diagram at the bottom of Figure 1 plots the residual

sum of squares of the threshold ADL regression against the various values of the error correction term. It is clear that the diagram is roughly V-shaped, with possible multiple threshold values. However, for simplicity, we focus here on the threshold ADL model with two regimes.

Table 4 reports the estimation results of the conditional threshold ADL regression based on *Indicator A* (Panel A) and *Indicator B* (Panel B), respectively. To check the robustness of our results, we also provide estimation results for the model without lagged values of Δy_{1t} and Δy_{2t}. Note that looking at the case with *Indicator A*, the threshold value determined by maximizing the Wald statistics is the same as the one (displayed in Figure 1) minimizing RSS. Table 4 also reports the ADL threshold cointegration tests for the following two null hypotheses

$$H_0' : \beta_1 = \beta_2 = 0; \quad \texttt{BDM Test},$$

$$H_0'' : \beta_1 = \beta_2 = \beta_3 = \beta_4 = 0; \quad \texttt{BO Test}.$$

It is clear from Table 4 that the null hypothesis of no threshold cointegration is rejected in all tests. Therefore, we confirm that consumption and GDP are threshold-cointegrated with a cointegrating vector of (1, -1.0239).

7 Summary and Concluding Remarks

In this paper, we develop new threshold cointegration tests based on ADL type models. Our tests extend the linear ADL cointegration tests previously suggested by Boswijk (1994) and Banerjee et al. (1986). Our ADL threshold cointegration tests are new in the literature and perform reasonably well. Using our new threshold cointegration tests, we can resolve or improve upon some problems found in the existing tests. In particular, our threshold cointegration ADL tests are not subject to the size distortions or power loss found in some existing tests. Our new tests are free of the nuisance parameters that generate problems in the ECM or EG based tests.

We provide the relevant asymptotic theory for two different cases in a unified framework where the indicator function is represented by either a nonstationary or stationary threshold variable. We treat the threshold parameter, which is not identified under the null, as a percentile parameter. The resulting sup Wald statistics are obtained by searching over the range between 0 and 1, regardless of different types of indicators. Our tests are

asymptotic based tests that do not rely on the bootstrap procedure. Since our tests are single equation based tests, the condition of weak exogeneity is warranted. It seems plausible, however, that our findings might generalize to a system of equations. This topic remains as a subject for future research.

References

[1] Andrews, D. (1993), "Tests for Parameter Instability and Structural Change with Unknown Change Point," *Econometrica*, 61, 821–856.

[2] Andrews, D. and W. Ploberger (1994), "Optimal Tests When a Nuisance Parameter Is Present Only Under the Alternative," *Econometrica*, 62, 1383–1414.

[3] Balke, N., and T. Fomby (1997), "Threshold Cointegration, *International Economic Review*, 38, 627-645.

[4] Balke, N., and M. Wohar (1998), "Nonlinear Dynamics and Covered Interest Rate Parity," *Empirical Economics*, 23, 535 - 559.

[5] Banerjee, A., J.J. Dolado, D. Hendry, and G.W. Smith (1986), "Exploring Equilibrium Relationships in Econometrics Through Static Models: Some Monte Carlo Evidence," *Oxford Bulletin of Economics and Statistics*, 48, 3, 253-277.

[6] Boswijk, H. P. (1994), "Testing for an Unstable Root in Conditional and Structural Error Correction Models," *Journal of Econometrics*, 63, 37-60.

[7] Caner, M., and B. Hansen (2001), "Threshold Autoregression with a Unit Root," *Econometrica*, 69, 1555-1596.

[8] Chan, N. H. and C. Z. Wei (1988), "Limiting Distributions of Least Squares Estimates of Unstable Autoregressive Processes," *Annals of Statistics*,16, 367-401.

[9] Davies, R. (1977), "Hypothesis Testing When a Nuisance Parameter is Present only Under the Alternative," *Biometrika,* 64 247–254.

[10] Davies, R. (1987), "Hypothesis Testing When a Nuisance Parameter is Present only Under the Alternative," *Biometrika,* 74, 33–43.

[11] Enders, W., and C. W. J. Granger (1998), "Unit Root Tests and Asymmetric Adjustment with an Example Using the Term Structure of Interest Rates, *Journal of Business and Economic Statistics*, 16, 304-311.

[12] Enders, W. and P. Siklos (2001), "Cointegration and Threshold Adjustment," *Journal of Business and Economic Statistics* 19, 166 – 76.

[13] Engle, R. F. and C. W. J. Granger (1987), "Cointegration and Error Correction: Representation, Estimation and Testing," *Econometrica*, 55, 251-276.

[14] Ericsson, N. R., and J.G. MacKinnon (2002), "Distributions of Error Correction Tests for Cointegration," *Econometrics Journal*, 5, 285-318.

[15] Granger, C. W. J. (1981), "Some Properties of Time Series Data and Their Use in Econometric Model Specification," *Journal of Econometrics*, 16, 121-130.

[16] Hansen, B. (1995), "Rethinking the univariate approach to unit root tests: How to use covariates to increase power," *Econometric Theory*, 11, 1148-1171.

[17] Hansen, B. (1997), "Approximate Asymptotic P-values for Structural Change Tests," *Journal of Business and Economic Statistics*, 15, 60-67.

[18] Hansen, B. and B. Seo (2002), "Testing for Two-regime Threshold Cointegration in Vector Error Correction Models," *Journal of Econometrics*, 110, 293-318.

[19] Hendry, D. F. and N. R. Ericsson (1991), "Modeling the Demand for Narrow Money in the United Kingdom and the United States," *European Economic Review*, 35, 833-881.

[20] Kremers, J.J.M., N.R. Ericsson and J.J. Dolado (1992), "The Power of Cointegration Tests," *Oxford Bulletin of Economics and Statistics*, 54, 3, 325-348.

[21] Kurtz, T and P. Protter (1991), "Weak Limit Theorems for Stochastic Integrals and Stochastic Differential Equations," *Annals of Probability* 19, 1035-1070.

[22] Li, J. (2006), "Testing Linearity Against Threshold Autoregressive Models," Working Paper, University of Alabama.

[23] Michael, P., R. A. Nobay, and D. A. Peel (1997), "Transaction Costs and Nonlinear Adjustment in Real Exchange Rates: An Empirical Investigation," *Journal of Political Economy*, 105, 862-879.

[24] Phillips, P. C. B., and S. Durlauf (1986), "Multiple Time Series Regression with Integrated Processes," *Review of Economic Studies*, 53, 473-495.

[25] Phillips. P. C. B. and S. Ouliaris (1990), "Asymptotic Properties of Residual Based Tests for Cointegration," *Econometrica*, 58, 165-193.

[26] Seo, M. (2006), "Bootstrap Testing for the Null of No Cointegration in a Threshold Vector Error Correction Model" *Journal of Econometrics*, forthcoming.

[27] Saikkonen, P. (1991), "Asymptotically Efficient Estimation of Cointegration Regressions," *Econometric Theory*, 7, 1-21.

[28] Sephton, P. S. (2003), "Spatial Market Arbitrage and Threshold Cointegration," *American Journal of Agricultural Economics*, 85, 1041–1046.

[29] Taylor, M. P. and D. A. Peel (2000), "Nonlinear Adjustment, Long-Run Equilibrium and Exchange Rate Fundamentals," *Journal of International Money and Finance,* 19, 33-53.

[30] Tong, H. (1983), Threshold Models in Non-linear Time Series Analysis, Lecture Notes in Statistics, Vol. 21. Springer, Berlin.

[31] Tong, H. (1990), Non-Linear Time Series: A Dynamic System Approach, Oxford University Press, Oxford.

[32] Wu, J. and Chen, S. (2001), "Nominal Exchange-Rate Prediction: Evidence from a Nonlinear Approach," *Journal of International Money and Finance,* 20, 521-532.

[33] Zivot, E. (2000), "The Power of Single Equation Tests for Cointegration When the Cointegrating Vector is pre-specified," *Econometric Theory*, 16, 407-439.

Table 1. Critical values of proposed ADL tests for threshold cointegration

n	BO			BO_μ			BO_t		
	1%	5%	10%	1%	5%	10%	1%	5%	10%
1	20.96	16.60	14.66	24.00	19.04	16.90	26.98	22.07	19.57
2	27.53	22.56	19.95	30.09	24.67	22.11	32.97	27.47	24.74
3	32.09	27.06	24.48	34.75	29.35	26.50	38.03	32.36	29.34
4	37.36	31.58	28.97	39.93	33.88	30.99	42.86	36.53	33.79
5	42.33	36.33	33.41	44.29	38.10	35.24	48.36	41.19	37.71
n	BDM			BDM_μ			BDM_t		
	1%	5%	10%	1%	5%	10%	1%	5%	10%
1	17.15	12.97	11.15	20.71	15.96	13.63	23.79	18.96	16.62
2	20.68	16.01	13.76	23.72	18.66	16.24	26.95	21.57	19.01
3	22.44	17.90	15.42	26.34	20.91	18.32	29.61	23.98	21.31
4	25.28	20.08	17.43	29.01	23.34	20.43	31.94	26.02	22.90
5	27.47	21.91	19.33	31.01	25.00	22.19	34.77	27.93	24.88

n	$\widetilde{BO}$			$\widetilde{BO}_\mu$			$\widetilde{BO}_t$		
	1%	5%	10%	1%	5%	10%	1%	5%	10%
1	19.98	15.65	13.65	23.88	18.66	16.36	26.15	21.44	19.17
2	26.01	21.08	18.68	28.66	23.43	20.90	31.56	26.21	23.60
3	29.90	25.34	22.68	33.16	27.91	25.15	36.70	30.67	27.61
4	34.99	29.58	26.78	38.34	32.26	29.36	40.78	34.98	32.01
5	40.12	33.99	31.03	43.29	36.43	33.40	45.15	39.23	36.03
n	$\widetilde{BDM}$			$\widetilde{BDM}_\mu$			$\widetilde{BDM}_t$		
	1%	5%	10%	1%	5%	10%	1%	5%	10%
1	15.95	11.92	9.94	20.07	15.03	12.90	22.92	18.12	15.87
2	18.57	13.88	11.66	21.46	16.84	14.53	24.89	19.76	17.18
3	20.04	15.40	13.15	23.77	18.49	16.09	27.23	21.41	18.68
4	22.50	17.17	14.65	25.63	20.37	17.54	28.29	23.00	20.08
5	24.34	19.06	16.55	27.81	21.57	18.87	30.71	24.53	21.51

Table 2. Sizes of ADL tests for threshold cointegration at the 5% level

Test	T	ρ_1	$\phi_2 = 0, \phi_1 = 0.4$	$\phi_2 = 0.5, \phi_1 = 0.1$	$\phi_2 = 0.5, \phi_1 = 0.4$
BO	100	0.3	0.066	0.070	0.085
		0.7	0.069	0.075	0.075
	250	0.3	0.072	0.061	0.063
		0.7	0.065	0.059	0.063
$\widetilde{BO}$	100	0.3	0.077	0.082	0.073
		0.7	0.066	0.073	0.064
	250	0.3	0.076	0.065	0.066
		0.7	0.075	0.067	0.059
BDM	100	0.3	0.067	0.061	0.075
		0.7	0.060	0.077	0.065
	250	0.3	0.065	0.063	0.068
		0.7	0.061	0.054	0.072
$\widetilde{BDM}$	100	0.3	0.057	0.068	0.059
		0.7	0.055	0.060	0.054
	250	0.3	0.058	0.060	0.057
		0.7	0.065	0.052	0.049
$BDMN$	100	0.3	0.069	0.068	0.081
		0.7	0.059	0.077	0.070
	250	0.3	0.068	0.066	0.068
		0.7	0.058	0.053	0.068
$\widetilde{BDMN}$	100	0.3	0.059	0.070	0.058
		0.7	0.053	0.055	0.050
	250	0.3	0.056	0.063	0.052
		0.7	0.062	0.051	0.049

Note: The threshold BO and $\widetilde{BO}$ tests, which are defined in (11), are testing $B_1 = B_2 = 0$ in (10). The threshold BDM and $\widetilde{BDM}$ tests, defined in (13), are testing $\kappa_{11} = \kappa_{12} = 0$ in (12). The $BDMN$ and $\widetilde{BDMN}$ tests are the same as the BDM and $\widetilde{BDM}$ tests, except that the lead terms of Δx_t are omitted in (12).

Table 3. Powers of ADL and ES tests for threshold cointegration at the
5% level

c	Test	T	ϕ_1	$\gamma = 0.2, k = -0.1$	$\gamma = 0.2, k = -0.3$	$\gamma = 0.7, k = -0.3$
0.0	BO	100	0.1	0.283	0.558	0.333
			0.4	0.579	0.848	0.464
		250	0.1	0.893	0.995	0.924
			0.4	0.996	0.100	0.987
	BDM	100	0.1	0.205	0.471	0.342
			0.4	0.396	0.742	0.451
		250	0.1	0.730	0.973	0.897
			0.4	0.951	0.999	0.967
	ES	100	0.1	0.037	0.103	0.122
			0.4	0.043	0.189	0.120
		250	0.1	0.216	0.610	0.388
			0.4	0.383	0.764	0.289
2.0	BO	100	0.1	0.283	0.473	0.201
			0.4	0.579	0.822	0.337
		250	0.1	0.893	0.911	0.782
			0.4	0.966	0.100	0.963
	BDM	100	0.1	0.205	0.368	0.212
			0.4	0.396	0.690	0.321
		250	0.1	0.730	0.953	0.750
			0.4	0.951	0.997	0.906
	ES	100	0.1	0.037	0.061	0.085
			0.4	0.043	0.140	0.098
		250	0.1	0.216	0.528	0.281
			0.4	0.383	0.752	0.216

Note: The threshold BO and $\widetilde{BO}$ tests, which are defined in (11), are testing $B_1 = B_2 = 0$ in (10). The threshold BDM and $\widetilde{BDM}$ tests, defined in (13), are testing $\kappa_{11} = \kappa_{12} = 0$ in (12). The $BDMN$ and $\widetilde{BDM}N$ tests are the same as the BDM and $\widetilde{BDM}$ tests, except that the lead terms of Δx_t are omitted in (12).

Table 4: Conditional threshold ADL model of consumption

Panel A

	β_0	β_1	β_2	β_3	β_4	β_5	β_6	β_7	$e^*(\tau)$	τ
Coefficients	-0.02	-0.03	-0.03	0.03	0.03	0.61	-0.42	0.12	-0.00269	0.52
$t-$ratio	-0.94	-1.05	-1.23	1.11	1.25	9.39	-3.32	1.87		
	BO	59.62*		BDM	35.01*					

	β_0	β_1	β_2	β_3	β_4	β_5			$e^*(\tau)$	τ
Coefficients	-0.02	-0.03	-0.03	0.03	0.03	0.53			-0.00336	0.53
$t-$ratio	-0.75	-0.99	-0.93	1.01	0.93	7.08				
	BO	33.56*		BDM	21.49*					

Panel B

	β_0	β_1	β_2	β_3	β_4	β_5	β_6	β_7	$\Delta e^*(\tau)$
Coefficients	-0.05	-0.10	-0.09	0.10	0.09	0.63	-0.34	0.07	-0.00688
$t-$ratio	-3.22	-3.39	-3.80	3.45	3.79	9.40	-2.91	0.83	
	BO	41.72*		BDM	30.17*				

	β_0	β_1	β_2	β_3	β_4	β_5			$\Delta e^*(\tau)$
Coefficients	-0.03	-0.06	-0.05	0.07	0.05	0.54			0.00541
$t-$ratio	-2.05	-2.83	-2.06	2.81	2.10	7.50			
	BO	45.65*		BDM	41.55*				

Note: * significant at 5% level.

Figure 1:

Appendix

Proof of Theorem 1

For given τ, we have $I_{1t} = 1(e_{t-d} < e^*_{t-d}(\tau))$, $I_{2t} = 1 - I_{1t}$. We let $X_1 = (z_1 I_{11}, \ldots, z'_T I_{1T})'$, $X_2 = (z'_1 I_{21}, \ldots, z'_T I_{2T})'$, $\Delta y = (\Delta y_1, \ldots, \Delta y_T)'$, and $q = (q'_1, \ldots, q'_T)'$, and define the projection matrix $M = I - q(q'q)^{-1}q'$. It follows that $\widehat{B}_j = (X'_j M X_j)^{-1}(X'_j M \Delta y)$ for $j = 1, 2$ and $\widehat{V}(\widehat{B}_j) = \widehat{\sigma}^2 (X'_j M X_j)^{-1}$, where $\widehat{\sigma}^2 = T^{-1}(\Delta y - X_1 \widehat{B}_1 - X_2 \widehat{B}_2)' M(\Delta y - X_1 \widehat{B}_1 - X_2 \widehat{B}_2)$ is the estimate of $\text{Var}(u_{1t})$. Following Lemma 1 of Boswijk (1994), we can express the vector z_t under H_0 as

$$z_t = \Gamma S_t + \eta_t,$$

where Γ is a nonsingular matrix, $S_t = \sum_{j=1}^{t} u_t$ with $u_t = (u_{1t}, u'_{2t})'$ and η_t is a stationary process. Basically, this is a modification of the Granger representation theorem. By Lemma 2.2(c) of Phillips and Ouliaris (1990) and the continuous mapping theorem (CMT), we can show that $I_{1t} \Rightarrow I(W_3 < W_3^*(\tau))$. By Assumption 2 and the CMT, it follows from Kurtz and Protter (1991) that

$$\left(T^{-2} \sum_{t=1}^{T} S_{t-1} S'_{t-1} I_{1t}, T^{-1} \sum_{t=1}^{T} S_{t-1} u_{1t} I_{1t} \right) \Rightarrow \left(\int B I_1 B', \int B I_1 dB_1 \right)$$

$$\left(T^{-2} \sum_{t=1}^{T} S_{t-1} S'_{t-1} I_{2t}, T^{-1} \sum_{t=1}^{T} S_{t-1} u_{1t} I_{2t} \right) \Rightarrow \left(\int B I_2 B', \int B I_2 dB_1 \right),$$

where $I_1 = I(W_3 < W_3^*(\tau))$, $I_2 = I(W_3 \geq W_3^*(\tau))$. Also,

$$\left(\begin{array}{c} T^{-1} \sum_{t=1}^{T} S_{t-1} q_t I_{jt}, \ T^{-1} \sum_{t=1}^{T} q_t q'_t, \ T^{-1/2} \sum_{t=1}^{T} w_t u_{1t}, \\ T^{-3/2} \sum_{t=1}^{T} S_{t-1} \eta_t I_{jt}, \ T^{-1/2} \sum_{r=1}^{T} \eta_t q'_t, \ T^{-1/2} \sum_{r=1}^{T} \eta_t u_{1t} \end{array} \right) = O_p(1).$$

for $j = 1, 2$. Let $S'_{-j} = (S'_1 I_{jt}, \ldots, S'_T I_{jT})$ and $\eta'_{-j} = (\eta'_1 I_{jt}, \ldots, \eta'_T I_{jT})$ for $j = 1, 2$ and $v' = (u_{11}, \ldots, u_{1T})$. It follows

$$T^{-2} X'_1 M X_1 = T^{-2}(\Gamma S'_{-1} M S_{-1} \Gamma' + \eta'_1 M \eta_1) = T^{-2} \Gamma S'_{-1} M S_{-1} \Gamma' + o_p(1)$$
$$= T^{-2} \Gamma(S'_{-1} S_{-1} - S'_{-1} q(q'q) q' S_{-1}) \Gamma' = T^{-2} \Gamma S'_{-1} S_{-1} \Gamma' + o_p(1)$$
$$= \Gamma \int B \cdot 1_1 B' dr \Gamma' + o_p(1).$$

Similarly, we can show that $T^{-2} X'_2 M X_2 = \Gamma \int B I_1 B' dr \Gamma' + o_p(1)$, $T^{-1} X'_1 M v = \Gamma \int B I_1 dB_1 + o_p(1)$ and $T^{-1} X'_2 M v = \Gamma \int B I_2 dB_1 + o_p(1)$. Let C denote a

lower triangular matrix which is the Cholesky decomposition of Ω. So we can rewrite $B(r) = CW(r)$. The asymptotic distribution of $\widehat{B}_j$ can be written as

$$T\widehat{B}_1 \;\Rightarrow\; \left(\Gamma C \int WI_1W'drC'\Gamma'\right)^{-1}\Gamma C \int WI_1dW_1\sigma \qquad (18)$$

$$T\widehat{B}_2 \;\Rightarrow\; \left(\Gamma C \int WI_2W'drC'\Gamma'\right)^{-1}\Gamma C \int WI_2dW_1\sigma,$$

and their covariances converge weakly to

$$T^2\widehat{V}(\widehat{B}_1) \;\Rightarrow\; \sigma^2\left(\Gamma C \int WI_1W'drC'\Gamma'\right)^{-1} \qquad (19)$$

$$T^2\widehat{V}(\widehat{B}_2) \;\Rightarrow\; \sigma^2\left(\Gamma C \int WI_2W'drC'\Gamma'\right)^{-1}.$$

Combining these results yields for a given τ

$$BO(\tau) \Rightarrow \sum_{j=1}^{2} \int WI_jdW_1 \left(\int WI_jW'\right)^{-1} \int WI_jdW_1.$$

The final claim follows after applying CMT to $BO(\tau)$ over Θ. The proof for the case where an intercept or a time trend is added is analogous. For the tests based on *Indicator B*, note that $I_{1t} \Rightarrow I(U < \tau)$ by Caner and Hansen (2001). The procedure for the proof of these tests is the same as that for tests based on *Indicator A*, except that the expression for I_{1t} is changed.

Proof of Corollary 2

First note that under the null hypothesis, $T^{-1/2}z_{t-1} \Rightarrow \Gamma CW(r)$. Then by Definition 1, it is clear that

$$T^{-1/2}z_{t-1}I_{1t} \Rightarrow \Gamma CV^1 \equiv (Z_1^1, Z_2^{1\prime}), \;\; T^{-1/2}z_{t-1}I_{2t} \Rightarrow \Gamma CV^2 \equiv Z^2. \qquad (20)$$

Since 1_{1t} is orthogonal to 1_{2t}, simple algebra leads to

$$R_{Z_1^1,(Z_2^1,Z^2)}(r) = R_{Z_1^1,Z_2^1}(r). \qquad (21)$$

Then a partitioned regression based on (18), (19) and (21) implies that

$$t^2(\widehat{\kappa}_{11}) \Rightarrow \widehat{\sigma}^{-2} \int R_{Z_1^1,Z_2^1}(r)dB_1 \left(\int R_{Z_1^1,Z_2^1}(r)R_{Z_1^1,Z_2^1}(r)'\right)^{-1} \int R_{Z_1^1,Z_2^1}(r)dB_1.$$
$$(22)$$

Given (20) and the fact that $\Gamma = (\Gamma_{ij})$ and $C = (c_{ij})$ are both lower triangular (cf. Boswijk 1994 lemma 1), we can show that

$$R_{Z_1^1, Z_2^1}(r) = \Gamma_{11} c_{11} Q_1(r),$$

where $Q_1(r) = R_{V_1^1, V_2^1}(r)$. Substitute $R_{Z_1^1, Z_2^1}(r) = \Gamma_{11} c_{11} Q_1(r)$ into (22), and we have

$$t^2(\widehat{\kappa}_{11}) \Rightarrow \widehat{\sigma}^{-2} \int Q_1 dB_1 \left(\int Q_1 Q_1' \right)^{-1} \int Q_1 dB_1,$$

where the factor $\Gamma_{11} c_{11}$ cancels out in the end. The final result follows by recalling that $B(r) = CW(r)$, where C is given in (5). Thus $B_1(r) = c_{11} W_1(r) + c_{12} W_2(r)$, $B_2(r) = c_{22} W_2(r)$ and

$$
\begin{aligned}
\widehat{\sigma}^{-1} dB_1 &= \widehat{\sigma}^{-1} c_{11} dW_1(r) + \widehat{\sigma}^{-1} c_{12} dW_2(r) \\
&= (1 - \rho^2)^{1/2} dW_1 + \rho dW_2 + o_p(1)
\end{aligned}
$$

since $\rho^2 = \Omega_{12} \Omega_{21} / \Omega_2 \Omega_1$ and $\widehat{\sigma}^2 = \Omega_1 + o_p(1)$. The proof for $t^2(\widehat{\kappa}_{12})$ is similar.

Proof of Theorem 3

Define $u_t^* = (v_t, u_{2t}')'$, where v_t is the error in (12) and $\Delta x_t = u_{2t}$. The multiple invariance principle leads to

$$T^{-1/2}(u_1^* + \ldots + u_{[Tr]}^*) \Rightarrow B^*(r), \quad 0 < r \leq 1,$$

where $B^*(r) = (B_1^*(r), B_2^*(r)')' = C^* W(r)$ and

$$
C^* = \begin{bmatrix} c_{11} & 0 \\ 0 & c_{22} \end{bmatrix} = \begin{bmatrix} \Omega_1^{*1/2} & 0 \\ 0 & \Omega_2^{1/2} \end{bmatrix}.
$$

Notice that in this case the Cholesky decomposition for the long-run variance of $B^*(r)$ is diagonal by construction, because $E v_t u_{2t+j} = 0$ for any j. Using similar arguments adopted in proving Corollary 2, we can show that for $j = 1, 2$,

$$t^2(\widehat{\kappa}_{1j}) \Rightarrow \widehat{\sigma}_v^{-2} \int Q_j dB_1^* \left(\int Q_j Q_j' \right)^{-1} \int Q_j dB_1^*.$$

Because the matrix C^* is diagonal, and $\widehat{\sigma}_v^2 = \Omega_1^* + o_p(1)$, it follows that

$$\widehat{\sigma}_v^{-1} dB_1^* = \widehat{\sigma}_v^{-1} \Omega_1^{*1/2} dW_1(r) = dW_1(r) + o_p(1),$$

as claimed. The proof for other cases where $d_t \neq 0$ and different indicators are used is analogous.