ADL tests for threshold cointegration

Jing Li and Junsoo Lee

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

---

1Jing 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 nonlinear adjustment process. The asymmetric dynamics in foreign exchange rates can be caused by factors such as transaction costs and policy intervention. 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 variables. 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)}
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 nonlinearity 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
co-integration still remains. It is obvious that allowing for non-linearity can increase power. The second step procedure testing for non-linearity also poses a problem. The second step of testing for the presence of a threshold effect is valid only if co-integration holds and the co-integrating vector is known. It is easy to expect that the null of linearity will be rejected when non-stationarity holds. This is because such tests will diverge if co-integration does not hold. Then, rejection of the null does not necessarily mean non-linearity. Instead, it can imply non-stationarity or no co-integration. 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 co-integration 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 non-linearity and co-integration. They use the Engle-Granger (1987, EG) type testing regression. In particular, they propose two different sets of threshold co-integration 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 co-integrating regression exceeds the threshold value and 0 otherwise. The MTAR model uses the first difference of the co-integrating residuals in the indicator. The threshold co-integration 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 co-integration regression usually differ from the short-run adjustment coefficients, these two sets of coefficients are assumed to be equivalent in the EG procedure. Kremer, 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 co-integration tests using the EG type testing procedure.

One may adopt an error correction model (ECM) framework. It is well known that an ECM co-integration test has a mixture of non-standard and standard normal distributions, where the weight induces a nuisance parameter problem. Zivot (2000) suggests to estimate the weight parameter in a non-parametric fashion. No such test for threshold co-integration 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

\[
\begin{align*}
z_t &= d_t + w_t \\
\Delta w_t &= \Pi_1 w_{t-1} I_1 + \Pi_2 w_{t-1} I_2 + \Phi(L) \Delta w_{t-1} + \epsilon_t,
\end{align*}
\]

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 \(I_1\) and \(I_2\), 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 \(\kappa_1 = (\kappa_{11}, \kappa_{12}')'\) 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), $\Pi_1$ and $\Pi_2$ 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 $\Pi_1$ and $\Pi_2$ 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)

$$
\left( \frac{\Delta y_t}{\Delta x_t} \right) = \begin{pmatrix} \kappa_{11} \\ \kappa_{21} \end{pmatrix} \theta_1' z_{t-1} I_{1t} + \begin{pmatrix} \kappa_{12} \\ \kappa_{22} \end{pmatrix} \theta_2' z_{t-1} I_{2t} + \begin{pmatrix} \Phi_1(L) \\ \Phi_2(L) \end{pmatrix} \Delta z_{t-1} + \begin{pmatrix} \epsilon_{1t} \\ \epsilon_{2t} \end{pmatrix}.
$$

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

$$
\begin{align*}
\Delta y_t &= \kappa_{11} \theta_1' z_{t-1} I_{1t} + \kappa_{12} \theta_2' z_{t-1} I_{2t} + \alpha' \Delta x_t + \Phi_1(L) \Delta z_{t-1} + u_{1t} \\
\Delta x_t &= \kappa_{21} \theta_1' z_{t-1} I_{1t} + \kappa_{22} \theta_2' z_{t-1} I_{2t} + \Phi_2(L) \Delta z_{t-1} + \epsilon_{2t},
\end{align*}
$$

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 $\theta$, or equivalently $\kappa_{21} = \kappa_{22} = 0$, then the triangular system becomes

$$
\begin{align*}
\Delta y_t &= \kappa_{11} \theta_1' z_{t-1} I_{1t} + \kappa_{12} \theta_2' z_{t-1} I_{2t} + \alpha' \Delta x_t + \Phi_1^*(L) \Delta z_{t-1} + u_{1t} \\
\Delta x_t &= u_{2t},
\end{align*}
$$

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 $\Delta x_t$. 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 $T \to \infty$,

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

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

$$\Omega = \begin{pmatrix} \Omega_1 & \Omega_{12} \\ \Omega_{21} & \Omega_2 \end{pmatrix},$$

where $\Omega$ 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 $\Omega$. First, because $u_{1t}$ is an innovation process relative to $\{x_t, z_{t-j}, j = 1, 2, \ldots\}$, $\Omega_1 = \sigma^2 = \text{Eu}_1^2$. 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,

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

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

$$C = \begin{pmatrix} c_{11} & c_{12} \\ 0 & c_{22} \end{pmatrix} = \begin{pmatrix} (\Omega_1 - \Omega_{12}\Omega_2^{-1}\Omega_{21})^{1/2} & \Omega_{12}\Omega_2^{-1/2} \\ 0 & \Omega_2^{1/2} \end{pmatrix}$$

8
and we let $\rho^2 = \Omega_1^{-1}\Omega_{12}\Omega_{21}\Omega_2^{-1}$. By definition, $\rho^2$ 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 $\rho$ 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

$$y_t = \hat{\gamma}'x_t + e_t.$$  

(6)

By definition, $e_t = y_t - \hat{\gamma}'x_t$ 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 $\hat{\theta} = (1, -\hat{\gamma}')$ 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”)

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

(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 $\tau$-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 $\tau$-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

\[ I(e_{t-d} < e_{t-d}^*(\tau)) = I(\sigma_e^{-1}T^{-1/2}e_{t-d} < \sigma_e^{-1}T^{-1/2}e_{t-d}^*(\tau)) \]

\[ \Rightarrow I(W(r) < W^*(\tau)), \]

where \( \sigma_e^2 \) 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 \( \tau \), 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 \( \tau \) 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 $\tau$ as a percentile parameter has two theoretical advantages. First, it yields an asymptotic distribution that depends on the range of $\tau$. 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, $\tau$ always lies in a compact set $\tau \in \Theta = [\tau_l, \tau_r]$, where $\tau_l$ and $\tau_r$ denote the lower and upper bound of $\Theta$, 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 $\Theta$. 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". 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 $I_{2t} = 1 - I_{1t}$ where $\Delta e^*_{t-d}(\tau)$ denotes the $\tau$-th percentile element of the empirical distribution of $\Delta e_{t-d}$. 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

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

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.

---

2The 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 \textit{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 \( \Delta x_t \) 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 \( d_t = (1, t)' \)

\[
\Delta y_t = \pi'd_t + B'_1 z_{t-1} I_{1t} + B'_2 z_{t-1} I_{2t} + c'q_t + u_{1t},
\]

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 \( B_1 = \kappa_{11}\theta \),
\( B_2 = \kappa_1 \theta, \) 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 \( \tau \) 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 \( \tau, \) then the final test is a continuous functional of the sequence of statistics. More specifically, let \( \hat{B}(\tau) = (\hat{B}_1(\tau)', \hat{B}_2(\tau)')' \) be the OLS estimate of \( B = (B_1', B_2')' \) for given \( \tau \) in (10), and \( \hat{V}(\hat{B}(\tau)) \) be the OLS variance estimate. We denote the test statistics based on Indicator A as \( BO, BO_{\mu} \) 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 \( \tilde{BO}, \tilde{BO}_{\mu} \) and \( \tilde{BO}_t, \) respectively. They are specified as follows

\[ BO, BO_{\mu}, BO_t, \tilde{BO}, \tilde{BO}_{\mu}, \tilde{BO}_t \equiv \sup_{\tau \in \Theta} \hat{B}(\tau)' \hat{V}(\hat{B}(\tau))^{-1} \hat{B}(\tau), \]

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

\[ \sup_{\tau \in \Theta} \hat{B}_1(\tau)' \hat{V}(\hat{B}_1(\tau))^{-1} \hat{B}_1(\tau) + \hat{B}_2(\tau)' \hat{V}(\hat{B}_2(\tau))^{-1} \hat{B}_2(\tau). \]

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 $P_G(r)$ is defined as the residual of the continuous time regression $P(r) = \hat{\phi}' G(r) + R_{P,G}(r)$, where $\hat{\phi}$ 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 $W_3(r) = R_{W_1, W_2}(r)$, 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

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

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

where $I_1 = I(W_3 < W_3^*(\tau)), I_2 = I(W_3 \geq W_3^*(\tau))$ and $\tilde{I}_1 = I(U < \tau), \tilde{I}_2 = I(U \geq \tau)$ for given $\tau$. The parameter space $\Theta$ 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 $T \to \infty$

$$BO \Rightarrow A(W), \quad BO_\mu \Rightarrow A(W_\mu), \quad BO_t \Rightarrow A(W_t)$$

$$\tilde{B}O \Rightarrow \tilde{A}(W), \quad \tilde{B}O_\mu \Rightarrow \tilde{A}(W_\mu), \quad \tilde{B}O_t \Rightarrow \tilde{A}(W_t).$$

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 \( z_{t-1} \).

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

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 \( t^2(\hat{\kappa}_{11}(\tau)) \) and \( t^2(\hat{\kappa}_{12}(\tau)) \) in regression (10), where \( \hat{\kappa}_{11}(\tau) \) and \( \hat{\kappa}_{12}(\tau) \) are estimated coefficients of \( y_{t-1} \) in each of two regimes, respectively, for given \( \tau \). Then, under \( H_0 \) and Assumptions 1-2, as \( T \to \infty \)

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

15
where \( Q_j(r) = R_{V^j_1V^j_2}(r) \), \( V^j_1 \) and \( V^j_2 \) 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 \( t^2(\hat{\kappa}_{11}) \) and \( t^2(\hat{\kappa}_{12}) \) involve the nuisance parameter \( \rho^2 \), 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 \( t^2(\hat{\kappa}_{11}) \) and \( t^2(\hat{\kappa}_{12}) \) for each \( \rho \), as in Hansen (1995). This approach requires a kernel-based semiparametric estimation of \( \rho \) 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 \( \Delta x_t \) 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 \( \Delta x_t \) as follows

\[
\Delta y_t = \pi' d_t + \kappa_{11} y_{t-1} I_{1t} + \kappa_{12} y_{t-1} I_{2t} + b_{11} x_{t-1} I_{1t} + b_{12} x_{t-1} I_{2t} + c^* q_t^* + v_t, \tag{12}
\]

where \( q_t^* = (\Delta x'_{t}, \Delta x'_{t+1}, \ldots, \Delta x'_{t+m}, \Delta z'_{t-1}, \ldots, \Delta z'_{t-p})' \). Notice that compared to (10), \( q_t^* \) in (12) adds leads of \( \Delta x_t \). Following Saikkonen (1991), the maximal value \( m \) for the lead terms are chosen such that the new error process \( v_t \) in (12) satisfies \( Ev_t u_{2t+j} = 0 \) 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 \( \Omega^* \) for \( (v_t, u_{2t})' \) similarly as in (5), then \( \Omega^* \) 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 \( Eu_{1t} u_{2t+j} \approx 0 \) for \( j > q \) and \( q \) is big enough. In practice, choosing \( q \) can be guided by the extent to which \( \hat{u}_{1t-j} \) is correlated with \( u_{2t} = \Delta x_t \), where \( \hat{u}_{t} \) 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 \( \hat{\kappa}(\tau) = (\hat{\kappa}_{11}(\tau), \hat{\kappa}_{12}(\tau))' \) be the OLS estimate of \( \kappa = (\kappa_{11}, \kappa_{12})' \) for given
\( \tau \) in (12) and \( \hat{V}(\hat{\tau}(\cdot)) \) be the OLS variance estimate. The test statistics \( BDM, BDM_\mu \) and \( BDM_t \) are based on Indicator A; they are associated with \( d_t = 0, d_2 = 1 \) and \( d_t = (1, t)' \), respectively. The test statistics \( \widehat{BDM}, \widehat{BDM}_\mu \) and \( \widehat{BDM}_t \) are based on Indicator B in a similar manner. These statistics are specified as follows

\[
BDM, BDM_\mu, BDM_t, \widehat{BDM}, \widehat{BDM}_\mu, \widehat{BDM}_t \equiv \sup_{\tau \in \Theta} \hat{\tau}(\cdot)\hat{V}(\hat{\tau}(\cdot))^{-1}\hat{\tau}(\cdot).
\]

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

\[
\sup_{\tau \in \Theta} t^2(\hat{\tau}_{11}(\cdot)) + t^2(\hat{\tau}_{12}(\cdot)),
\]

where \( t^2(\hat{\tau}_{11}(\cdot)) \) and \( t^2(\hat{\tau}_{12}(\cdot)) \) 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 \( V_j(r, \tau) = (V_j^1, V_j^2) \) and \( \tilde{V}_j(r, \tau) = (\tilde{V}_j^1, \tilde{V}_j^2) \) \( j = 1, 2 \). We also let \( V^{j\mu}(r) = (V_j^{j\mu}, V_j^{j\mu})' \), \( V^{jt}(r) = (V_j^{jt}, V_j^{jt})' \), \( \tilde{V}^{j\mu}(r) = (\tilde{V}_j^{j\mu}, \tilde{V}_j^{j\mu})' \), and \( \tilde{V}^{jt}(r) = (\tilde{V}_j^{jt}, \tilde{V}_j^{jt})' \). Then, we express these terms as \( V^{j\mu}(r) = R_{V_j^{1\mu}}(r) \), \( V^{jt}(r) = R_{V_j^{1t}}(r) \), \( \tilde{V}^{j\mu}(r) = R_{\tilde{V}_j^{1\mu}}(r) \), and \( \tilde{V}^{jt}(r) = R_{\tilde{V}_j^{1t}}(r) \), 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 \( T \to \infty \)

\[
BDM \Rightarrow A_1(Q_1, Q_2), \quad BDM_\mu \Rightarrow A_1(Q_1^\mu, Q_2^\mu), \quad BDM_t \Rightarrow A_1(Q_1^t, Q_2^t)
\]

\[
\widehat{BDM} \Rightarrow A_1(\tilde{Q}_1, \tilde{Q}_2), \quad \widehat{BDM}_\mu \Rightarrow A_1(\tilde{Q}_1^\mu, \tilde{Q}_2^\mu), \quad \widehat{BDM}_t \Rightarrow A_1(\tilde{Q}_1^t, \tilde{Q}_2^t),
\]

where \( Q_j(r) = R_{V_j^1, V_j^2}(r), Q_j^\mu(r) = R_{V_j^{1\mu}, V_j^{2\mu}}(r), Q_j^t(r) = R_{V_j^{1t}, V_j^{2t}}(r), \tilde{Q}_j(r) = R_{\tilde{V}_j^1, \tilde{V}_j^2}(r), \tilde{Q}_j^\mu(r) = R_{\tilde{V}_j^{1\mu}, \tilde{V}_j^{2\mu}}(r) \) and \( \tilde{Q}_j^t(r) = R_{\tilde{V}_j^{1t}, \tilde{V}_j^{2t}}(r) \) for \( j = 1, 2 \).

Theorem 2 shows that after including leads of \( \Delta x_t \), 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.
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 $d_t = 0$.

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

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

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

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

It is clear from (15) that $\phi_1$ and $\phi_2$ are the key parameters yielding a nonzero value of $\rho$. Hence, we let $\phi_3 = 0$ in (14) without a loss of generality. From (15), it follows that $E\epsilon_{1t}\epsilon_{2t} = 1 - \rho_1^2$ and $E\epsilon_{1t}\epsilon_{2t} = 0$. Some algebra demonstrates that $E\epsilon_{1t}\epsilon_{2t-j} = 0$ for all $j \geq 0$, $E\epsilon_{2t}\epsilon_{1t-1} = \phi_2 (1 - \rho_1^2)$ and $E\epsilon_{2t}\epsilon_{1t-j} = \phi_2 \phi_1^{j-1} (1 - \rho_1^2)$ for all $j \geq 2$. Then it follows from (4) that $\Omega_{12} = \phi_2 (1 - \rho_1^2) / \phi_1$. We next let $\Theta^{dp} = (\rho_1, \phi_2, \phi_1)$. Similarly, $\Omega_2$ can be shown to be a more intricate function of $\Theta^{dp}$. However, for our purposes, it suffices to know that $\rho$ is a function of $\Theta^{dp}$. If we can show that test sizes are invariant to the parameters in $\Theta^{dp}$, which is the determinant of $\rho$, then it is safe to say that the sizes are invariant to $\rho$. Hansen (1995) examines this issue from a different perspective and estimates $\rho$, 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 $\rho$, say, $\hat{\rho}$. While it seems difficult
to directly control $\rho$ with any DGP, we can still control the value of $\rho$ 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 $BO$ and $BDM$ tests but the testing regression remains the same. For completeness, we also consider the $BDMN$ and $\tilde{BDM}$ tests, which are supposedly the $BDM$ and $\tilde{BDM}$ tests but without using the lead of $\Delta x_t$ in (12). We expect that the $BDMN$ and $\tilde{BDMN}$ 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 $\rho_1$, $\phi_2$ and $\phi_1$. 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.

We can summarize our main findings as follows. First, no obvious size distortion is found in our simulations using different combinations of $\rho_1$, $\phi_2$ and $\phi_1$. This result is consistent with Theorems 1 and 2, which state that our test distributions are free of $\rho$. 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 $\Delta x_t$ 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

$$
\Delta y_t = -0.1\varepsilon_{t-1}I_{1t} + k\varepsilon_{t-1}I_{2t} + \phi_1\Delta y_{t-1} + \epsilon_{1t} \\
\Delta x_t = k_2\varepsilon_{t-1}I_{1t} + k_3\varepsilon_{t-1}I_{2t} + \phi_2\Delta y_{t-1} + \epsilon_{2t},
$$

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 $\Delta x_t$ 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 \( \varepsilon_t \). Accordingly, only the BO and BDM type tests are considered in power simulations. In order to focus on the key parameters of \( \tau, \gamma \) and \( k \), we let \( E\varepsilon_1\varepsilon_{2t} = \rho_1 = 0.7 \) and \( \phi_2 = 0.5 \) in (16). Note that \( \gamma \) 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

\[
\Delta e_t = \varphi_1 e_{t-1}I_{1t} + \varphi_2 e_{t-1}I_{2t} + \varphi_3 \Delta e_{t-1} + v_t,
\]

(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 \( \tau \).

We define \( \tau \) 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.

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 \( \Delta x_t \). 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 \( \Delta x_t \) 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 \( \tau \) 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 \( \phi_1, \gamma \) and \( k \). That is, even though a power function is not explicitly derived, we can
see that the power is an increasing function of $\phi_1$ and a decreasing function of $\gamma$ 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 $t$:

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

where $k$ is the marginal propensity to consume, $\alpha$ is a parameter and $\epsilon_t$ 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 $\epsilon_t$ 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

\[ \ln y_{2t} - 1.0239 \ln y_{1t} = -0.6213 + \epsilon_t. \]

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, \( \text{Indicator A} \) with \( I_t = I(\epsilon_{t-1} < e_{t-1}^*(\tau)) \) and \( \text{Indicator B} \) with \( \tilde{I}_t = I(\Delta e_{t-1} < \Delta e_{t-1}^*(\tau)) \), are considered. Specifically, the following threshold ADL regression is fitted:

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

where \( I_t \) can be replaced with \( \tilde{I}_t \) if \( \text{Indicator B} \) is adopted. Most important, the adjustment speeds toward the long-run equilibrium, as measured by \( \beta_i \) (\( i = 1, 2, 3, 4 \)), 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 \( \Delta y_{1t} \) and \( \Delta y_{2t} \) is included in the regression following the parsimony principle. Here, the lag-selection is guided by the partial autocorrelation function (PACF) (not presented in the paper) of \( \Delta y_{2t} \). In our case, the PACF indicates little serial correlation after one lag. Readers should keep in mind, however, that the final results depend on the selection of RHS variables. We wish to point out that \( \Delta y_{1t} \) 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 \( \epsilon_{t-1} \). In a more general model, the threshold variable could be \( \epsilon_{t-d} \), where the delay lag \( d \) is estimated by a grid search. The threshold variable \( \epsilon_{t-1} \) 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 \( \epsilon_{t-1} \). 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), re-
spectively. To check the robustness of our results, we also provide estimation
results for the model without lagged values of $\Delta y_{1t}$ and $\Delta y_{2t}$. 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

$$ H'_0: \beta_1 = \beta_2 = 0; \quad \text{BDM Test}, $$

$$ H''_0: \beta_1 = \beta_2 = \beta_3 = \beta_4 = 0; \quad \text{BO Test}. $$

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 coin-
tegration 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 pa-
rameter, 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


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

<table>
<thead>
<tr>
<th></th>
<th>BO</th>
<th>BO$_\mu$</th>
<th>BO$_t$</th>
<th></th>
<th>BO</th>
<th>BO$_\mu$</th>
<th>BO$_t$</th>
<th></th>
<th>BO</th>
<th>BO$_\mu$</th>
<th>BO$_t$</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>n</td>
<td>1%</td>
<td>5%</td>
<td>10%</td>
<td>1%</td>
<td>5%</td>
<td>10%</td>
<td>1%</td>
<td>5%</td>
<td>10%</td>
<td>1%</td>
<td>5%</td>
<td>10%</td>
</tr>
<tr>
<td>2</td>
<td>27.53</td>
<td>22.56</td>
<td>19.95</td>
<td>24.67</td>
<td>22.11</td>
<td>23.27</td>
<td>27.47</td>
<td>24.77</td>
<td>23.47</td>
<td>24.67</td>
<td>22.11</td>
<td>23.27</td>
</tr>
<tr>
<td>3</td>
<td>32.09</td>
<td>27.06</td>
<td>24.48</td>
<td>29.87</td>
<td>29.35</td>
<td>26.50</td>
<td>29.87</td>
<td>26.50</td>
<td>29.87</td>
<td>29.35</td>
<td>26.50</td>
<td>29.87</td>
</tr>
<tr>
<td>4</td>
<td>37.36</td>
<td>31.58</td>
<td>28.97</td>
<td>31.38</td>
<td>30.89</td>
<td>31.47</td>
<td>36.30</td>
<td>34.37</td>
<td>33.79</td>
<td>31.38</td>
<td>30.89</td>
<td>31.47</td>
</tr>
<tr>
<td>5</td>
<td>42.33</td>
<td>36.33</td>
<td>33.41</td>
<td>38.10</td>
<td>35.24</td>
<td>41.19</td>
<td>37.71</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>BDM</td>
<td>BDM$_\mu$</td>
<td>BDM$_t$</td>
<td></td>
<td>BDM</td>
<td>BDM$_\mu$</td>
<td>BDM$_t$</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>n</td>
<td>1%</td>
<td>5%</td>
<td>10%</td>
<td>1%</td>
<td>5%</td>
<td>10%</td>
<td>1%</td>
<td>5%</td>
<td>10%</td>
<td>1%</td>
<td>5%</td>
<td>10%</td>
</tr>
<tr>
<td>3</td>
<td>22.44</td>
<td>17.90</td>
<td>15.42</td>
<td>26.34</td>
<td>20.91</td>
<td>18.32</td>
<td>29.61</td>
<td>23.98</td>
<td>21.31</td>
<td>26.34</td>
<td>20.91</td>
<td>18.32</td>
</tr>
<tr>
<td>4</td>
<td>25.28</td>
<td>20.08</td>
<td>17.43</td>
<td>29.01</td>
<td>23.34</td>
<td>20.43</td>
<td>31.94</td>
<td>26.02</td>
<td>22.90</td>
<td>29.01</td>
<td>23.34</td>
<td>20.43</td>
</tr>
<tr>
<td>5</td>
<td>27.47</td>
<td>21.91</td>
<td>19.33</td>
<td>31.01</td>
<td>25.00</td>
<td>22.19</td>
<td>34.77</td>
<td>27.93</td>
<td>24.88</td>
<td>31.01</td>
<td>25.00</td>
<td>22.19</td>
</tr>
</tbody>
</table>

Table 1. Critical values of proposed ADL tests for threshold cointegration
Table 2. Sizes of ADL tests for threshold cointegration at the 5% level

<table>
<thead>
<tr>
<th>Test</th>
<th>$T$</th>
<th>$\rho_1$</th>
<th>$\phi_2 = 0, \phi_1 = 0.4$</th>
<th>$\phi_2 = 0.5, \phi_1 = 0.1$</th>
<th>$\phi_2 = 0.5, \phi_1 = 0.4$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$BO$</td>
<td>100</td>
<td>0.3</td>
<td>0.066</td>
<td>0.070</td>
<td>0.085</td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.7</td>
<td>0.069</td>
<td>0.075</td>
<td>0.075</td>
</tr>
<tr>
<td></td>
<td>250</td>
<td>0.3</td>
<td>0.072</td>
<td>0.061</td>
<td>0.063</td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.7</td>
<td>0.065</td>
<td>0.059</td>
<td>0.063</td>
</tr>
<tr>
<td>$\tilde{BO}$</td>
<td>100</td>
<td>0.3</td>
<td>0.077</td>
<td>0.082</td>
<td>0.073</td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.7</td>
<td>0.066</td>
<td>0.073</td>
<td>0.064</td>
</tr>
<tr>
<td></td>
<td>250</td>
<td>0.3</td>
<td>0.076</td>
<td>0.065</td>
<td>0.066</td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.7</td>
<td>0.075</td>
<td>0.067</td>
<td>0.059</td>
</tr>
<tr>
<td>$BDM$</td>
<td>100</td>
<td>0.3</td>
<td>0.067</td>
<td>0.061</td>
<td>0.075</td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.7</td>
<td>0.060</td>
<td>0.077</td>
<td>0.065</td>
</tr>
<tr>
<td></td>
<td>250</td>
<td>0.3</td>
<td>0.065</td>
<td>0.063</td>
<td>0.068</td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.7</td>
<td>0.061</td>
<td>0.054</td>
<td>0.072</td>
</tr>
<tr>
<td>$\tilde{BDM}$</td>
<td>100</td>
<td>0.3</td>
<td>0.057</td>
<td>0.068</td>
<td>0.059</td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.7</td>
<td>0.055</td>
<td>0.060</td>
<td>0.054</td>
</tr>
<tr>
<td></td>
<td>250</td>
<td>0.3</td>
<td>0.058</td>
<td>0.060</td>
<td>0.057</td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.7</td>
<td>0.065</td>
<td>0.052</td>
<td>0.049</td>
</tr>
<tr>
<td>$BDMN$</td>
<td>100</td>
<td>0.3</td>
<td>0.069</td>
<td>0.068</td>
<td>0.081</td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.7</td>
<td>0.059</td>
<td>0.077</td>
<td>0.070</td>
</tr>
<tr>
<td></td>
<td>250</td>
<td>0.3</td>
<td>0.068</td>
<td>0.066</td>
<td>0.068</td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.7</td>
<td>0.058</td>
<td>0.053</td>
<td>0.068</td>
</tr>
<tr>
<td>$\tilde{BDMN}$</td>
<td>100</td>
<td>0.3</td>
<td>0.059</td>
<td>0.070</td>
<td>0.058</td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.7</td>
<td>0.053</td>
<td>0.055</td>
<td>0.050</td>
</tr>
<tr>
<td></td>
<td>250</td>
<td>0.3</td>
<td>0.056</td>
<td>0.063</td>
<td>0.052</td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.7</td>
<td>0.062</td>
<td>0.051</td>
<td>0.049</td>
</tr>
</tbody>
</table>

Note: The threshold $BO$ and $\tilde{BO}$ tests, which are defined in (11), are testing $B_1 = B_2 = 0$ in (10). The threshold $BDM$ and $\tilde{BDM}$ tests, defined in (13), are testing $\kappa_{11} = \kappa_{12} = 0$ in (12). The $BDMN$ and $\tilde{BDMN}$ tests are the same as the $BDM$ and $\tilde{BDM}$ tests, except that the lead terms of $\Delta x_t$ are omitted in (12).
### Table 3. Powers of ADL and ES tests for threshold cointegration at the 5% level

<table>
<thead>
<tr>
<th>$c$</th>
<th>Test</th>
<th>$T$</th>
<th>$\phi_1$</th>
<th>$\gamma = 0.2, k = -0.1$</th>
<th>$\gamma = 0.2, k = -0.3$</th>
<th>$\gamma = 0.7, k = -0.3$</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.0</td>
<td>BO</td>
<td>100</td>
<td>0.1</td>
<td>0.283</td>
<td>0.558</td>
<td>0.333</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>0.4</td>
<td>0.579</td>
<td>0.848</td>
<td>0.464</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>250</td>
<td>0.1</td>
<td>0.893</td>
<td>0.995</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0.996</td>
<td>0.100</td>
<td>0.987</td>
</tr>
<tr>
<td></td>
<td>BDM</td>
<td>100</td>
<td>0.1</td>
<td>0.205</td>
<td>0.471</td>
<td>0.342</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>0.4</td>
<td>0.396</td>
<td>0.742</td>
<td>0.451</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>250</td>
<td>0.1</td>
<td>0.730</td>
<td>0.973</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0.951</td>
<td>0.999</td>
<td>0.967</td>
</tr>
<tr>
<td></td>
<td>ES</td>
<td>100</td>
<td>0.1</td>
<td>0.037</td>
<td>0.103</td>
<td>0.122</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>0.4</td>
<td>0.043</td>
<td>0.189</td>
<td>0.120</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>250</td>
<td>0.1</td>
<td>0.216</td>
<td>0.610</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0.383</td>
<td>0.764</td>
<td>0.289</td>
</tr>
<tr>
<td>2.0</td>
<td>BO</td>
<td>100</td>
<td>0.1</td>
<td>0.283</td>
<td>0.473</td>
<td>0.201</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>0.4</td>
<td>0.579</td>
<td>0.822</td>
<td>0.337</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>250</td>
<td>0.1</td>
<td>0.893</td>
<td>0.911</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0.966</td>
<td>0.100</td>
<td>0.963</td>
</tr>
<tr>
<td></td>
<td>BDM</td>
<td>100</td>
<td>0.1</td>
<td>0.205</td>
<td>0.368</td>
<td>0.212</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>0.4</td>
<td>0.396</td>
<td>0.690</td>
<td>0.321</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>250</td>
<td>0.1</td>
<td>0.730</td>
<td>0.953</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0.951</td>
<td>0.997</td>
<td>0.906</td>
</tr>
<tr>
<td></td>
<td>ES</td>
<td>100</td>
<td>0.1</td>
<td>0.037</td>
<td>0.061</td>
<td>0.085</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>0.4</td>
<td>0.043</td>
<td>0.140</td>
<td>0.098</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>250</td>
<td>0.1</td>
<td>0.216</td>
<td>0.528</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0.383</td>
<td>0.752</td>
<td>0.216</td>
</tr>
</tbody>
</table>

Note: The threshold $BO$ and $\tilde{BO}$ tests, which are defined in (11), are testing $B_1 = B_2 = 0$ in (10). The threshold $BDM$ and $\tilde{BDM}$ tests, defined in (13), are testing $\kappa_{11} = \kappa_{12} = 0$ in (12). The $BDMN$ and $\tilde{BDMN}$ tests are the same as the $BDM$ and $\tilde{BDM}$ tests, except that the lead terms of $\Delta x_t$ are omitted in (12).
Table 4: Conditional threshold ADL model of consumption

<table>
<thead>
<tr>
<th>Panel A</th>
<th>$\beta_0$</th>
<th>$\beta_1$</th>
<th>$\beta_2$</th>
<th>$\beta_3$</th>
<th>$\beta_4$</th>
<th>$\beta_5$</th>
<th>$\beta_6$</th>
<th>$\beta_7$</th>
<th>$e^*(\tau)$</th>
<th>$\tau$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Coefficients</td>
<td>-0.02</td>
<td>-0.03</td>
<td>-0.03</td>
<td>0.03</td>
<td>0.03</td>
<td>0.61</td>
<td>-0.42</td>
<td>0.12</td>
<td>-0.00269</td>
<td>0.52</td>
</tr>
<tr>
<td>$t-$ratio</td>
<td>-0.94</td>
<td>-1.05</td>
<td>-1.23</td>
<td>1.11</td>
<td>1.25</td>
<td>9.39</td>
<td>-3.32</td>
<td>1.87</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$BO$</td>
<td>59.62*</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$BDM$</td>
<td>35.01*</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Panel B</th>
<th>$\beta_0$</th>
<th>$\beta_1$</th>
<th>$\beta_2$</th>
<th>$\beta_3$</th>
<th>$\beta_4$</th>
<th>$\beta_5$</th>
<th>$\beta_6$</th>
<th>$\beta_7$</th>
<th>$\Delta e^*(\tau)$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Coefficients</td>
<td>-0.02</td>
<td>-0.03</td>
<td>-0.03</td>
<td>0.03</td>
<td>0.03</td>
<td>0.53</td>
<td></td>
<td></td>
<td>-0.00336</td>
</tr>
<tr>
<td>$t-$ratio</td>
<td>-0.75</td>
<td>-0.99</td>
<td>-0.93</td>
<td>1.01</td>
<td>0.93</td>
<td>7.08</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$BO$</td>
<td>33.56*</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$BDM$</td>
<td>21.49*</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Panel B</th>
<th>$\beta_0$</th>
<th>$\beta_1$</th>
<th>$\beta_2$</th>
<th>$\beta_3$</th>
<th>$\beta_4$</th>
<th>$\beta_5$</th>
<th>$\beta_6$</th>
<th>$\beta_7$</th>
<th>$\Delta e^*(\tau)$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Coefficients</td>
<td>-0.03</td>
<td>-0.06</td>
<td>-0.05</td>
<td>0.07</td>
<td>0.05</td>
<td>0.54</td>
<td></td>
<td></td>
<td>0.00541</td>
</tr>
<tr>
<td>$t-$ratio</td>
<td>-2.05</td>
<td>-2.83</td>
<td>-2.06</td>
<td>2.81</td>
<td>2.10</td>
<td>7.50</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$BO$</td>
<td>45.65*</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$BDM$</td>
<td>41.55*</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

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 \( \hat{B}_j = (X'_j M X_j)^{-1} (X'_j M \Delta y) \) for \( j = 1, 2 \) and \( \hat{V}(\hat{B}_j) = \hat{\sigma}^2 (X'_j M X_j)^{-1} \), where \( \hat{\sigma}^2 = T^{-1} (\Delta y - X_1 \hat{B}_1 - X_2 \hat{B}_2)' M (\Delta y - X_1 \hat{B}_1 - X_2 \hat{B}_2) \) 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 \( \Gamma \) is a nonsingular matrix, \( S_t = \sum_{j=1}^t u_t \) with \( u_t = (u_{1t}, u_{2t})' \) and \( \eta_t \) 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), it follows from Kurtz and Protter (1991) that

\[
\begin{align*}
(T^{-2} \sum_{t=1}^T S_{t-1} S_{t-1}' I_{1t}, T^{-1} \sum_{t=1}^T S_{t-1} u_{1t} I_{1t}) & \Rightarrow \left( \int B_I B', \int B_I d\Gamma \right), \\
(T^{-2} \sum_{t=1}^T S_{t-1} S_{t-1}' I_{2t}, T^{-1} \sum_{t=1}^T S_{t-1} u_{1t} I_{2t}) & \Rightarrow \left( \int B_{I2} B', \int B_{I2} d\Gamma \right),
\end{align*}
\]

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

\[
\begin{align*}
(T^{-1} \sum_{t=1}^T S_{t-1} q_t I_{jt}, T^{-1} \sum_{t=1}^T S_{t-1} q_t I_{jt}, T^{-1} \sum_{t=1}^T \eta_t I_{jt}, T^{-1} \sum_{t=1}^T \eta_t I_{jt}) & = O_p(1),
\end{align*}
\]

for \( j = 1, 2 \). Let \( S_{-j} = (S_{-j} I_{jt}, \ldots, S_{-j} I_{jt}) \) and \( \eta_{-j} = (\eta_{-j} I_{jt}, \ldots, \eta_{-j} I_{jt}) \) for \( j = 1, 2 \) and \( v' = (u_{11}, \ldots, u_{1T}) \). It follows

\[
\begin{align*}
T^{-2} X'_1 M X_1 & = T^{-2} (\Gamma S'_{-1} M S_{-1} \Gamma' + \eta'_{-1} M \eta_{-1}) = T^{-2} \Gamma S'_{-1} M S_{-1} \Gamma' + o_p(1) \\
& = T^{-2} \Gamma (S'_{-1} S_{-1} - S'_{-1} q(q'q)q'S_{-1}) \Gamma' = T^{-2} \Gamma S'_{-1} S_{-1} \Gamma' + o_p(1) \\
& = \Gamma \int B \cdot 1_1 B' d\Gamma' + o_p(1) \quad \text{for} \quad j = 1, 2.
\end{align*}
\]

Similarly, we can show that \( T^{-2} X'_2 M X_2 = \Gamma \int B_{I2} B' d\Gamma' + o_p(1), T^{-1} X'_1 M v = \Gamma \int B_I d\Gamma + o_p(1) \) and \( T^{-1} X'_2 M v = \Gamma \int B_{I2} d\Gamma + o_p(1) \). Let \( C \) denote a
lower triangular matrix which is the Cholesky decomposition of $\Omega$. So we can rewrite $B(r) = CW(r)$. The asymptotic distribution of $\hat{B}_j$ can be written as

$$T \hat{B}_1 \Rightarrow \left( \Gamma C \int W W' dC'T' \right)^{-1} \Gamma C \int W_1 W_1 \sigma$$

(18)

and their covariances converge weakly to

$$T^2 \hat{V}(\hat{B}_1) \Rightarrow \sigma^2 \left( \Gamma C \int W W' dC'T' \right)^{-1}$$

$$T^2 \hat{V}(\hat{B}_2) \Rightarrow \sigma^2 \left( \Gamma C \int W_1 W W' dC'T' \right)^{-1} \Gamma C \int W_1 W_1 \sigma,$$

(19)

Combining these results yields for a given $\tau$

$$BO(\tau) \Rightarrow \sum_{j=1}^{2} \int W_1 W_1 \left( \int W_1 W W' dC'T' \right)^{-1} \int W_1 W_1 \sigma.$$

The final claim follows after applying CMT to $BO(\tau)$ over $\Theta$. 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

$$T^{-1/2} z_{t-1} I_{1t} \Rightarrow \Gamma CV \equiv (Z_1, Z_2'), T^{-1/2} z_{t-1} I_{2t} \Rightarrow \Gamma CV \equiv Z^2.$$

(20)

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

$$R_{Z_1, Z_2}(r) = R_{Z_1, Z_2}(r).$$

(21)

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

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

(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

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

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

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

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

$$\tilde{\sigma}^{-1}dB_1 = \tilde{\sigma}^{-1}c_{11}dW_1(r) + \tilde{\sigma}^{-1}c_{12}dW_2(r) = (1 - \rho^2)^{1/2}dW_1 + \rho dW_2 + o_p(1)$$

since $\rho^2 = \Omega_{12}\Omega_{21}/\Omega_{22}$ and $\tilde{\sigma}^2 = \Omega_1 + o_p(1)$. The proof for $t^2(\hat{\kappa}_{12})$ 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

$$T^{-1/2}(u^*_1 + \ldots + u^*_{[T]} \Rightarrow B^*(r), \quad 0 < r \leq 1,$$

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 $E\tilde{v}_t u_{2t+j} = 0$ for any $j$. Using similar arguments adopted in proving Corollary 2, we can show that for $j = 1, 2,$

$$t^2(\hat{\kappa}_{ij}) \Rightarrow \tilde{\sigma}^{-2} \int Q_j dB_1^* \left( \int Q_j Q_j' \right)^{-1} \int Q_j dB_1^*.$$  

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

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

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