Available online at www.sciencedirect.com
SCIENCE DIRECT®
Journal of Econometrics 134 (2006) 129–150
JOURNAL OF Econometrics
www.elsevier.com/locate/jeconom

Bootstrap testing for the null of no cointegration in a threshold vector error correction model

Myunghwan Seo*

Department of Economics, London School of Economics, London WC2A 2AE, UK

Available online 15 August 2005

Abstract

We develop a test for the linear no cointegration null hypothesis in a threshold vector error correction model. We adopt a sup-Wald type test and derive its null asymptotic distribution. A residual-based bootstrap is proposed, and the first-order consistency of the bootstrap is established. A set of Monte Carlo simulations shows that the bootstrap corrects size distortion of asymptotic distribution in finite samples, and that its power against the threshold cointegration alternative is significantly greater than that of conventional cointegration tests. Our method is illustrated with used car price indexes.

JEL classification: C12; C15; C32

Keywords: Threshold vector error correction model; Cointegration; Threshold cointegration; Bootstrap

1. Introduction

Threshold cointegration, introduced by Balke and Fomby (1997), generalizes standard linear cointegration to allow adjustment toward long-run equilibrium to be nonlinear and/or discontinuous. Such nonlinear short-run dynamics are predicted by many economic phenomena, such as policy interventions and the presence of transaction costs or any other transaction barriers. Lo and Zivot (2001) provide an

*Tel.: +44 20 7955 7512; fax: +44 20 7955 6592.
E-mail address: m.seo@lse.ac.uk.

0304-4076/$ - see front matter
doi:10.1016/j.jeconom.2005.06.018

extensive review of the growing literature regarding the application of threshold cointegration. Despite significant applied interest, econometric theory has not been developed satisfactorily.

Two testing issues arise: one is testing for the presence of cointegration, that is, the presence of long-run equilibrium, and the other is testing for the linearity of short-run dynamics. Commonly adopted in the literature is the two-step approach proposed by Balke and Fomby (1997), in which the linear no cointegration null hypothesis is first examined against the linear cointegration alternative, and then the linear cointegration null hypothesis is tested against the threshold cointegration alternative. For example, investigating the linearity of the term structure of interest rates, Hansen and Seo (2002) apply the ADF test to the interest rate spread and then apply the SupLM test they developed for a two-regime threshold vector error correction model (TVECM). However, this approach can be quite misleading because the standard cointegration tests can suffer from substantial power loss when the alternative is threshold cointegration, as demonstrated by Pippenger and Goering (2000), Taylor (2001), and the simulation study in this paper. Therefore, a new test is required to examine the linear no cointegration null hypothesis in a TVECM or in a threshold autoregression (TAR), either of which allows both linear and threshold cointegration alternative.[1] Although Enders and Granger (1998) and Enders and Siklos (2001) propose such tests in TAR's, they do not provide a formal distribution theory. No such test is developed in a TVECM.

This paper develops a cointegration test in a TVECM with a prespecified cointegrating vector, in which the linear no cointegration null hypothesis is examined. Economic models often imply simple and known cointegrating vectors, and the use of a prespecified cointegrating vector is common in the literature as in Lo and Zivot (2001) and Hansen and Seo (2002). In addition, the power of the cointegration test improves significantly by using prespecified cointegrating vectors. (See Horvath and Watson (1995) for examples and more discussion.) Unlike the cointegrating vector, however, few economic theories predict the threshold parameter.

The testing is nonstandard since the threshold parameter is not identified under the null hypothesis. The inference problem when a nuisance parameter is not identified under the null hypothesis has been studied by Davies (1987), Andrews and Ploberger (1994), and Hansen (1996), among others. This paper employs the sup-Wald type statistic (hereafter supW) following Davies (1987) and derives its asymptotic distribution based on a newly developed asymptotic theory. This development contributes to the literature by extending the nonlinear nonstationary

[1]Four hypotheses are possible in threshold cointegration models: linear no cointegration, threshold no cointegration, linear cointegration, and threshold cointegration. However, the above two-step approach excludes the threshold no cointegration hypothesis, so does all the existing literature. Throughout this paper, we follow this convention, and the rejection of the linear no cointegration null hypothesis is understood as either linear or threshold cointegration. However, it is also worthwhile to note that testing the threshold no cointegration null hypothesis requires completely different distribution theory than does testing the linear no cointegration null, and therefore, there is no single test examining both the nulls simultaneously. This is another reason the present paper focuses on the linear no cointegration null. The author thanks a referee for directing attention to this point.

asymptotics of Park and Phillips (2001) to a uniform convergence over a class of functions.

This paper proposes a residual-based bootstrap to approximate the distribution of the statistic supW, and establish its consistency. In other words, the distribution of the bootstrapped supW is shown to converge to the same asymptotic distribution of supW by establishing an invariance principle for a bootstrap partial sum process. The bootstrap with nonstationary data has become popular recently, and this paper is the first that shows the consistency of such a bootstrap in a nonlinear model. For more information concerning the bootstraps and their consistency in standard unit root testing, see, for example, Park (2002) and Paparoditis and Politis (2003).

The finite sample properties are investigated and compared to those of the conventional cointegration tests, such as the ADF test and the Wald test of Horvath and Watson (1995). We find that the bootstrap of supW approximates the finite sample distribution of supW reasonably well. Furthermore, it exhibits much higher empirical power against threshold cointegration alternatives than the conventional tests. The discrepancy in power becomes larger as the sample size increases. Therefore, the use of supW is advisable when threshold cointegration is under consideration.

This paper is organized as follows. Section 2 introduces the model and our supW test statistic, and develops an asymptotic distribution of the statistic. We describe the bootstrap procedure and establish its asymptotic validity in Section 3. Finite sample performances of the bootstrap are examined in Section 4. Section 5 illustrates the usefulness of the proposed testing strategy by an application of the law of one price hypothesis. Proofs of all the theorems are presented in the appendix.

2. Testing the linear no cointegration null in a TVECM

We consider a Band-TVECM

$$\Phi(L)\Delta x_t = \alpha_1 z_{t-1} 1\{z_{t-1} \leqslant \gamma_1\} + \alpha_2 z_{t-1} 1\{z_{t-1} > \gamma_2\} + \mu + \varepsilon_t, \tag{1}$$

where $t = 1, \ldots, n$, and $\Phi(L)$ is a qth-order polynomial in the lag operator defined as $\Phi(L) = I - \Phi_1 L^1 - \cdots - \Phi_q L^q$. The error correction term is defined as $z_t = x_t'\beta$ for a known cointegrating vector β. The threshold parameter $\gamma = (\gamma_1, \gamma_2)$ satisfying $\gamma_1 \leqslant \gamma_2$ takes values on a compact set Γ. The model (1) allows for the no-adjustment region in the middle ($\gamma_1 < z_{t-1} \leqslant \gamma_2$), which arises due to the presence of transaction barriers or policy interventions. We employ this model because it has been used in most empirical studies often with restrictions, such as $\alpha_1 = \alpha_2$ and/or $\gamma_1 = -\gamma_2$ imposed. See Lo and Zivot (2001) for a review. The two-regime TVECM ($\gamma_1 = \gamma_2$) is also often used, especially when the sample size is relatively small. The testing strategy developed in this paper can be applied to restricted models with little modification.

There are four possible hypotheses, as mentioned above. Therefore, multiple tests are necessary to distinguish one hypothesis from another. Hansen and Seo (2002) develop a test for the linear cointegration null hypothesis in a two-regime TVECM. In other words, they test the hypothesis $\alpha_1 = \alpha_2$ under the restriction that both are nonzero, and that $\gamma_1 = \gamma_2$. This paper develops a test for the linear no cointegration null hypothesis:

$$H_0 : \alpha_1 = \alpha_2 = 0 \tag{2}$$

in the model (1). It is left to future research to develop a test for the threshold no cointegration null hypothesis.[2] Following the convention in the literature, it is assumed throughout this paper that there is no such case as threshold no cointegration.

The testing problem is not conventional, in the sense that the threshold parameter γ is not identified under the null (2). This is called the Davies problem in the literature following Davies (1977, 1987). A typical approach to the Davies problem is to apply a continuous functional for the statistics defined as a function of the unidentified parameter. (Refer to Hansen, 1996.) Therefore, we do not have to estimate the unidentified nuisance parameter, which cannot be estimated consistently. Davies (1987) proposes taking supremum over the parameter space, and Andrews and Ploberger (1994) argue that some sort of averaging functional is asymptotically optimal. However, the optimality property does not hold in our case, since it is based on the stationarity. Note that the threshold variable z_{t-1} is an integrated process under the null (2).

2.1. Test statistic

We employ the supremum of the Wald (denoted as supW) statistic. The Wald test is a typical method in the linear VECM to test the null hypothesis of no cointegration when there is only one cointegrating relation that is known under the alternative. We adopt the supremum because it is simple and we can avoid the inherent arbitrariness resulting from the choice of weighting function for the averaging. Furthermore, the supW is asymptotically equivalent to the LR statistic under normality. See Watson (1994, pp. 2880–2881) for a discussion.

When γ is given, the least-squares estimators for the coefficients are the OLS estimators. Thus, write

$$\begin{aligned}\Delta x_t = {} & \hat{\alpha}_1(\gamma) z_{t-1} 1\{z_{t-1} \leqslant \gamma_1\} + \hat{\alpha}_2(\gamma) z_{t-1} 1\{z_{t-1} > \gamma_2\} \\ & + \hat{\mu}(\gamma) + \hat{\Phi}_1(\gamma)\Delta x_{t-1} + \cdots + \hat{\Phi}_q(\gamma)\Delta x_{t-q} + \hat{\varepsilon}_t(\gamma)\end{aligned} \tag{3}$$

and let

$$\hat{\Sigma}(\gamma) = \frac{1}{n}\sum_{t=1}^{n} \hat{\varepsilon}_t(\gamma)\hat{\varepsilon}_t(\gamma)'. \tag{4}$$

We introduce some matrix notations. Let $A = (\alpha_1, \alpha_2)'$, and Z_γ and ε be the matrices stacking $(z_{t-1}1\{z_{t-1} \leqslant \gamma_1\}, z_{t-1}1\{z_{t-1} > \gamma_2\})$ and ε_t', respectively. And define M_{-1} as projection onto the orthogonal space of the constant and the lagged terms

[2] The parameter space for the threshold no cointegration hypothesis is quite complicated (a subset of $\{\alpha_1 = 0 \text{ and } \alpha_2 \neq 0\}$ or $\{\alpha_1 \neq 0 \text{ and } \alpha_2 = 0\}$), since the intercept μ also plays a role in determining the stationarity of the error correction term z_t. Examining α_1 and α_2 is not sufficient for the testing purpose. See Chan et al. (1985) for more details.

$\Delta x_{t-1}, \ldots, \Delta x_{t-q}$. Then the Wald statistic testing the null (2) with a fixed γ is

$$\begin{aligned} W_n(\gamma) &= vec(\hat{A}(\gamma))' \, var(vec(\hat{A}(\gamma)))^{-1} \, vec(\hat{A}(\gamma)) \\ &= vec((Z'_\gamma M_{-1} Z_\gamma)^{-1}(Z'_\gamma M_{-1}\varepsilon))'[(Z'_\gamma M_{-1} Z_\gamma)^{-1} \otimes \hat{\Sigma}(\gamma)]^{-1} \\ &\quad \times vec((Z'_\gamma M_{-1} Z_\gamma)^{-1}(Z'_\gamma M_{-1}\varepsilon)) \\ &= tr\left\{(Z'_\gamma M_{-1}\varepsilon\hat{\Sigma}(\gamma)^{-1/2})'(Z'_\gamma M_{-1} Z_\gamma)^{-1}(Z'_\gamma M_{-1}\varepsilon\hat{\Sigma}(\gamma)^{-1/2})\right\} \end{aligned} \tag{5}$$

and the supremum statistic is defined as

$$\text{supW} = \sup_{\gamma \in \Gamma} W_n(\gamma).$$

2.2. Asymptotic distribution

For the subsequent development of our theory, we make the following assumptions.

Assumption 1. (a) $\{\varepsilon_t\}$ is independent and identically distributed with mean zero, variance Σ and $E|\varepsilon_t|^r < \infty$ for some $r > 4$ and
(b) $\Phi(z) \neq 0$, for all $|z| \leqslant 1$.

Conditions (a) and (b) are assumptions usually imposed to analyse linear models with integrated time series. The threshold variable z_{t-1} is typically assumed to have no deterministic time trend. Therefore, we assume that the time-series x_t has no deterministic time trend. That is, we need an auxiliary assumption that, under the null (2),

$$\mu = 0. \tag{6}$$

A word on notation before we proceed. The symbol $\Rightarrow$ denotes weak convergence with respect to the uniform metric on the parameter space, and $[x]$ signifies the integer part of x. Let $B(r)$ be a standard p-dimensional vector Brownian motion on $[0, 1]$, and write stochastic processes, such as $B(r)$ on $[0, 1]$ as B to achieve notational economy. Similarly, integrals, such as $\int_0^1 B(r)\,\mathrm{d}r$ and $\int_0^1 B(r)\,\mathrm{d}B(r)'$, are written more simply as $\int_0^1 B$ and $\int_0^1 B\,\mathrm{d}B'$, respectively. Let U be a vector Brownian motion with covariance matrix Σ and $W = \beta'\Phi(1)^{-1}U$. All limits in this paper are as the sample size $n \to \infty$. The following development is necessary to derive the limit distribution of supW. Let Θ be a compact set in the real line.

Theorem 1. *If* (2), (6), *and Assumption* 1 *hold*, *then*

(a) $$\frac{1}{n^2}\sum_{t=1}^{n} z_{t-1}^2 1\{z_{t-1} \leqslant \theta\} \Rightarrow \int_0^1 W^2 1\{W \leqslant 0\},$$

(b) $$\frac{1}{n}\sum_{t=1}^{n} z_{t-1} 1\{z_{t-1} \leqslant \theta\}\varepsilon_t \Rightarrow \int_0^1 W 1\{W \leqslant 0\}\,\mathrm{d}U.$$

on Θ.

Note that these convergences are with respect to the uniform metric on the parameter space Θ. This uniformity is important because the supremum type statistic is employed. Although a recent study by Park and Phillips (2001) develops asymptotics involving a broad range of nonlinear transformations of nonstationary variables, the results do not apply here. The limit distributions in the above theorem are functionals of Brownian motions, and the parameter θ disappears. Heuristically, for large n,

$$1\{z_{t-1} \leqslant \theta\} = 1\left\{\frac{z_{t-1}}{\sqrt{n}} \leqslant \frac{\theta}{\sqrt{n}}\right\} \approx 1\{W \leqslant 0\}.$$

By the same reasoning, if the indicator function is replaced by $1\{z_{t-1} > \theta\}$, then we have $1\{W > 0\}$ in the limit.

Even though the supW statistic has the same asymptotic null distribution as the Wald statistic that is constructed by fixing γ at a certain value such as zero, we expect the supW will have better power property unless the chosen value of γ is correct. On the other hand, it may be related to the poor approximation of the asymptotic distribution to the sampling distribution in the finite sample as demonstrated in the Monte Carlo simulation.

Based on this development, we present the limit distribution of our supW statistic.

Theorem 2. *If* (2), (6), *and Assumption* 1 *hold*, *then*

$$sup\,W \Rightarrow tr\left\{\left(\int_0^1 \tilde{B}_1\,\mathrm{d}B'\right)'\left(\int_0^1 \tilde{B}_1\tilde{B}_1'\right)^{-1}\left(\int_0^1 \tilde{B}_1\,\mathrm{d}B'\right)\right\},$$

where B_1 *is the first element of* B, *and*

$$\tilde{B}_1 = \begin{pmatrix} B_1 1\{B_1 \leqslant 0\} - \int_0^1 B_1 1\{B_1 \leqslant 0\} \\ B_1 1\{B_1 > 0\} - \int_0^1 B_1 1\{B_1 > 0\} \end{pmatrix}.$$

Due to the recursion property of the Brownian motion, the limit distribution is well defined. If we do not include the constant term in the estimation, then the demeaned process $B_1 1\{B_1 \leqslant 0\} - \int_0^1 B_1 1\{B_1 \leqslant 0\}$ is replaced by the original process $B_1 1\{B_1 \leqslant 0\}$. Like the conventional cointegrating test statistic (e.g., Horvath and Watson, 1995), the limit distribution does not rely on any nuisance parameter. We

Table 1
Critical values for the supW test

	When β is known					
	w/o constant			w/ constant		
	1%	5%	10%	1%	5%	10%
$p = 2$	14.853	10.968	9.198	17.275	13.078	11.085
$p = 3$	18.417	14.041	12.033	20.527	16.008	13.809
$p = 4$	21.681	16.914	14.750	23.679	18.766	16.370
$p = 5$	24.742	19.679	17.329	26.678	21.419	18.912

tabulate the critical values for both cases with and without the constant. These critical values are obtained by Monte Carlo simulation of 100,000 replications with the sample size of 100,000. Table 1 reports these values.

3. Residual-based bootstrap

In this section, motivated by the asymptotic pivotalness of the supW statistic, we propose a residual-based bootstrap approximation to the distribution of supW and show that the bootstrap is first-order consistent. Refinement of finite sample performance is investigated by Monte Carlo simulations in Section 4.

Since the time-series x_t is nonstationary, we do not resample the data directly. Instead, we make use of the assumption that ε_t is independent and identically distributed. Since it is unobservable, we resample LS residuals of the model (1) independently with replacement. For this reason, the name "residual-based bootstrap" is given.[3]

The bootstrap proceeds as follows. Let $\hat{\gamma} = \arg\min|\hat{\Sigma}(\gamma)|$ and

$$\hat{\Phi}_i = \hat{\Phi}_i(\hat{\gamma}),\ i = 1, \ldots, q \quad \text{and} \quad \hat{\varepsilon}_t = \hat{\varepsilon}_t(\hat{\gamma}). \tag{7}$$

Then, the bootstrap distribution is completely determined by $\hat{\Phi}_i$s and the empirical distribution of $\hat{\varepsilon}_t, \hat{F}$. That is, the bootstrap sample is generated by

$$\Delta x_t^* = \hat{\Phi}_1 \Delta x_{t-1}^* + \cdots + \hat{\Phi}_{q-1} \Delta x_{t-q+1}^* + \hat{\varepsilon}_t^*, \tag{8}$$

where $\hat{\varepsilon}_t^*$ is a random draw from $\hat{F}$. For the initial values of this series we can use the sample values.

Once we generate a bootstrap sample x_t^* by integrating up Δx_t^*, we follow the same steps calculating supW, that is, (3)–(5) to get

$$\text{supW}^* = \sup_{\gamma \in \Gamma} vec(\hat{A}^*(\gamma))'[(Z_\gamma^{*\prime} M_{-1}^* Z_\gamma^*)^{-1} \otimes \hat{\Sigma}^*(\gamma)]^{-1} vec(\hat{A}^*(\gamma)), \tag{9}$$

where the superscript * indicates the bootstrap counterpart.

Note that we restrict α_1, α_2, and μ at zero-imposing (2) and (6) in the bootstrap data generating process (8). It appears crucial for supW* to have the correct asymptotic distribution. As in the conventional bootstrap unit root test of Basawa et al. (1991), the consistency of $\hat{\alpha}_i$ does not guarantee the consistency of the bootstrap, because of the discontinuity of distribution of a time series at the unit root. Instead of attempting to show the inconsistency, we show the first-order consistency of our bootstrap.

Since the distribution of the bootstrap statistic is defined conditional on each realization of the sample, we need to introduce a notation "$X_n^* \overset{*}{\Rightarrow} X$ in $\mathbf{P}$", meaning that the distance of the laws of X_n^* and X tends to zero in probability. The same notation is also found in Paparoditis and Politis (2003) and Chang and Park (2003).

[3]In contrast, we may also use restricted LS residuals by estimating the model (1) with (2) and (6) imposed. This bootstrap is called difference-based bootstrap. See Park (2002), for example.

The consistency of the bootstrap should be established case-by-case, and that of our bootstrap is established below.

Theorem 3. *If* (2), (6), *and Assumption* 1 *hold*, *then*

$$sup W^{*} \Rightarrow tr\left\{\left(\int_0^1 \tilde{B}_1\,\mathrm{d}B'\right)'\left(\int_0^1 \tilde{B}_1\tilde{B}_1'\right)^{-1}\left(\int_0^1 \tilde{B}_1\,\mathrm{d}B'\right)\right\} \quad in\ \boldsymbol{P}.$$

4. Monte Carlo experiment

The finite sample performance of the supW statistic is examined and compared to the performances of the ADF test and the Wald test by Horvath and Watson (1995) (hereafter HW). The number of simulations is 1000, and that of bootstrap replications is 200. We examine samples of sizes 100 and 250. The statistic supW is computed based on the two-regime TVECM when the sample size is 100, and from the band TVECM when it is 250.

If we adopt the latter with a small sample size, the number of observations in each regime can be too small to estimate the model properly.

In practice, we need a data-dependent way of constructing the parameter space Γ. In case of a stationary threshold variable, an interval between two quantiles of the threshold variable is commonly used in the literature. The interval, however, is not bounded asymptotically if the threshold variable is an integrated process, invalidating the quantile-based approach. Here, we propose the following: First, set $\Theta = [-\bar{\theta}, \bar{\theta}]$ for $\bar{\theta}$ being a quantile of $|z_{t-1}|$, and then compute supW over

$$\left\{\gamma_1, \gamma_2 \in \Theta \text{ s.t. } \gamma_1 \leqslant \gamma_2, \sum_{t=2}^{n} 1\{z_{t-1} \leqslant \gamma_1\} \geqslant m\right.$$

and

$$\left.\sum_{t=2}^{n} 1\{z_{t-1} > \gamma_2\} \geqslant m, \text{ for some } m > 0\right\} \tag{10}$$

and supW* in the same way. The variable $z_t = x_{1t} - x_{2t}$ throughout this section. It is worthwhile to observe that the bootstrap distribution replicates the dependence of the statistic supW on a particular choice of the quantile, unlike the case of the asymptotic distribution.

A few things to be noted: first, we use the quantile of $|z_{t-1}|$ instead of z_{t-1} because the bootstrap sample z_{t-1}^* can evolve symmetrically around zero. Second, we need to use lower and lower quantiles as the sample size increases, to maintain the boundedness of the parameter space. We do not pursue the optimal choice of the quantile here. In our experiment, we use the maximum of $|z_{t-1}|$, which seems to work well with the considered sample sizes. Third, the constraint (10) is necessary for estimating the regime-specific parameters $\alpha_i,\ i = 1, 2$ reasonably, but it is not binding asymptotically due to the recursion property of the Brownian motion. We set $m = 10$ in this experiment.

To examine the size properties of the test statistics, we generate data based on

$$\Delta x_t = \Phi \Delta x_{t-1} + \varepsilon_t, \tag{11}$$

where ε_t is independent and identically distributed with standard bivariate normal distribution. Φ varies among

$$\Phi_0 = \begin{pmatrix} 0 & 0 \\ 0 & 0 \end{pmatrix}, \quad \Phi_1 = \begin{pmatrix} -0.2 & 0 \\ -0.1 & -0.2 \end{pmatrix}, \quad \Phi_2 = \begin{pmatrix} -0.2 & -0.1 \\ -0.1 & -0.2 \end{pmatrix}.$$

To begin with, we assess the accuracy of the asymptotic approximations. For this experiment, we set $\Phi = \Phi_0$, fix Γ as $[-10, 10]$ and try different sample sizes of $100, 500, 1000$ and 3000. Table 2 reports the rejection frequencies of the supW test at 10% nominal sizes. The over-rejection tendency is clear even for the large sample of 3000. Next, we turn to the residual-based bootstrap of supW. In contrast to the asymptotics, the bootstrap performs surprisingly well, even in the small sample of 100, as reported in Table 3. It also reports the rejection frequencies of the other two tests that are based on the asymptotic critical values. All the tests exhibit reasonable size properties.

We turn to the power properties of the tests against the threshold cointegration alternative. We generate data from a simple band TVECM:

$$\Delta x_t = \alpha_1 z_{t-1} 1\{z_{t-1} \leqslant -\theta\} + \alpha_2 z_{t-1} 1\{z_{t-1} > \theta\} + \varepsilon_t,$$

Table 2
Empirical size of supW based on the asymptotic distribution at 10% nominal size

Sample size	100	500	1000	3000
Rejection frequency	.470	.360	.285	.227

Table 3
Size of cointegration tests: the rejection frequencies of supW are based on the bootstrap

	Φ	10% nominal size			5% nominal size		
		Φ_0	Φ_1	Φ_2	Φ_0	Φ_1	Φ_2
$n:100$	supW	.128	.124	.114	.066	.056	.054
	HW	.134	.116	.106	.076	.062	.064
	ADF	.114	.108	.110	.048	.054	.060
$n:250$	supW	.142	.116	.142	.058	.050	.064
	HW	.112	.100	.118	.052	.042	.064
	ADF	.114	.104	.104	.064	.050	.052

Two-regime TVECM is employed when sample size is 100 and band TVECM when 250.

Table 4
Power of cointegration tests: the rejection frequencies of supW are based on the bootstrap

	Threshold (γ)	10% nominal size			5% nominal size		
		γ^0	γ^1	γ^2	γ^0	γ^1	γ^2
Case 1 ($\alpha_2 = (0, .1)$)							
$n: 100$	supW	.464	.436	.362	.324	.276	.236
	HW	.428	.296	.262	.274	.210	.184
	ADF	.372	.242	.218	.222	.122	.124
$n: 250$	supW	.988	.984	.824	.946	.918	.720
	HW	.940	.742	.412	.812	.530	.282
	ADF	.980	.798	.358	.884	.582	.224
Case 2 ($\alpha_2 = (-.3, 0)$)							
$n: 100$	supW	.628	.424	.348	.500	.292	.236
	HW	.516	.308	.268	.366	.218	.158
	ADF	.302	.206	.200	.178	.144	.106
$n: 250$	supW	1.00	.956	.836	.998	.914	.750
	HW	.998	.732	.424	.972	.532	.276
	ADF	.946	.420	.258	.808	.236	.146

Two-regime TVECM is employed when sample size is 100 and band TVECM when it is 250.

where $\theta > 0$. We examine $\theta = 5$, 8, and 10 and denote the corresponding $\gamma = (-\theta, \theta)'$ as γ^0, γ^1, and γ^2, respectively. As θ increases, the no-adjustment region also expands, which may affect the powers of the tests differently. The power of supW is based on the bootstrap p-value due to the severe over-rejection tendency of the asymptotic approximation. We report the case where $\alpha_1 = (-a, 0)'$ and $\alpha_2 = (0, a)'$. To save space, we report the results for $a = 0.1$ only. The variation of a does not alter the results of the comparison much. We also report the case with $\alpha_2 = (-0.3, 0)$ for different types of threshold effect.

Table 4 reports the result: first, the power of supW dominates the powers of the conventional test statistics. Second and more importantly, the power differences increase as the threshold value θ and the sample size n increase. The difference is as large as 49.6% (see the case with $n = 250$, $\gamma = \gamma^2$). Third, we do not observe much difference between HW and ADF. If threshold cointegration is under consideration, the practitioner is advised to employ supW with the residual-based bootstrap to test the linear no cointegration null.[4]

[4]Although not reported here, the power of ADF or HW is greater than that of supW, if the alternative is linear cointegration. This is natural, because the formers are specially designed for the linear alternative. In practice, it may be prudent to apply both conventional tests and the new test, and to interpret the results with caution. In other words, given the nice size properties, rejection by any of the tests may be interpreted as evidence of cointegration, either linear or threshold.

5. Empirical illustration

We illustrate the use of our test strategy using the price indexes of used car markets from 29 different locations in the US, and investigate the Law of One Price (LOP) hypothesis. The hypothesis has been frequently examined in the literature, and Lo and Zivot (2001), among others, study the hypothesis and the presence of threshold effect in the adjustment toward the long-run equilibrium for many categories of goods. We use the same data set as they do, namely the US Bureau of Labor Statistics Monthly Consumer Price Indexes for the period of December 1986–June 1996 (115 observations), as plotted in Fig. 1.

Lo and Zivot (2001) proceed as follows: first, they employ the prespecified cointegrating vector $(1, -1)'$, which is implied by the LOP for log prices. Second, having 29 locations in hand, they pick a benchmark city, New Orleans, and construct 28 bivariate systems of log prices, each of which consists of the benchmark city and one of the remaining 28 locations. Third, for each system, they apply the two-step approach; that is, first, they test the linear no cointegration null hypothesis and then the linear cointegration null hypothesis. In the second step, they employ a two-regime TVECM with no lagged term to maintain a reasonable degree of freedom in the small sample size.

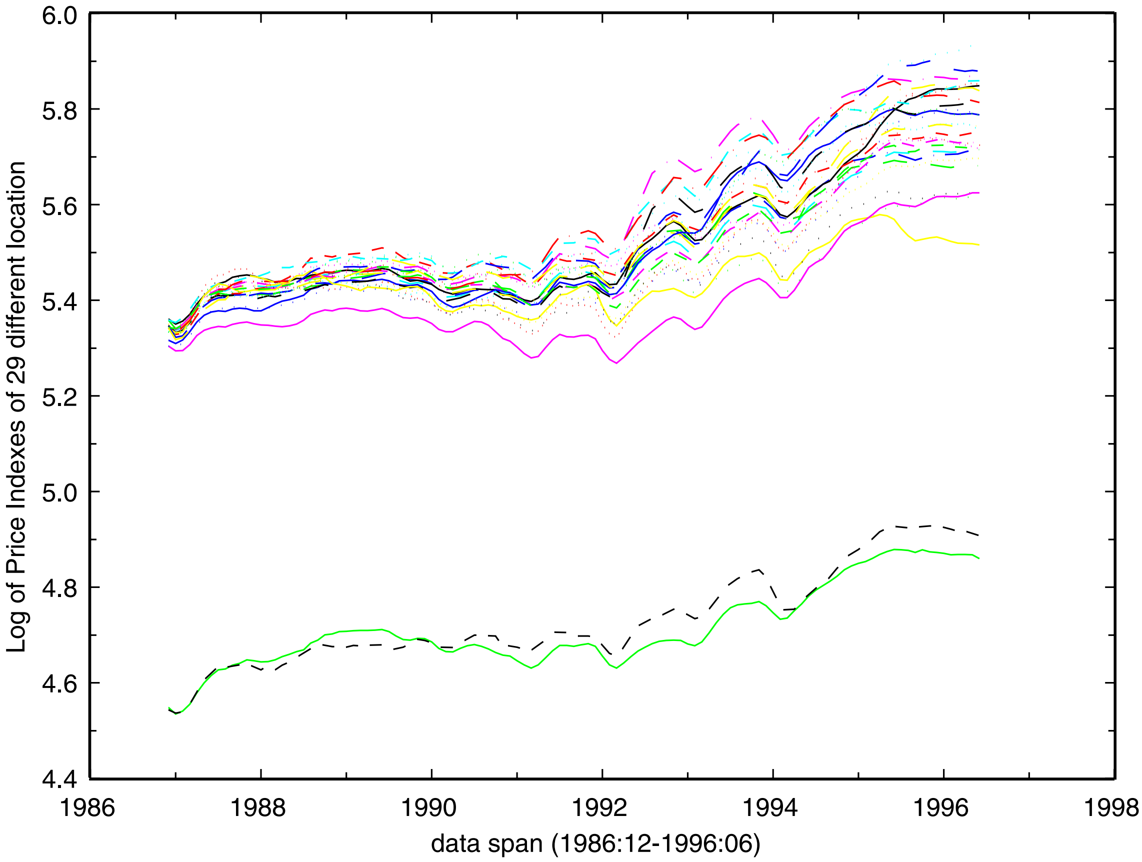

Fig. 1. Plot of data.

Similarly, we consider the 28 bivariate systems and test the linear no cointegration null hypothesis using the test developed in this paper, and compare it with conventional cointegration tests, such as the HW test and the ADF test. And we employ the two-regime model ((1) with $\gamma_1 = \gamma_2$). For the lag length selection, both no-lag and Schwarz Information Criterion (BIC) are used for robustness. Based on the simulation results, we apply the residual-based bootstrap with 1000 replications for supW and the asymptotic distribution for the others.

We focus on the used car indexes, among others, because the uncertainty about the quality of a used car, in addition to the transaction cost, may generate a clearer threshold effect in the adjustment than the others, and therefore, the discrepancy in the rejection frequencies between the new test and the conventional tests can be clear as in the simulation study. For example, the relationship between the changes of New York price indexes and the lagged spread of New York and New Orleans price indexes is given in Fig. 2. The dashed line represents the estimated threshold point from the two-regime TVECM. It looks quite clear from Fig. 2 that the adjustment of the price is not linearly related to the price spread.

Table 5 reports the results. Conventional cointegration tests, such as HW and ADF, does not reject the linear no cointegration null hypothesis in most cases. On the contrary, supW without no lagged terms rejects the null in 21 systems out of 28 at 10% nominal size. If we include some lagged terms based on BIC, then the number

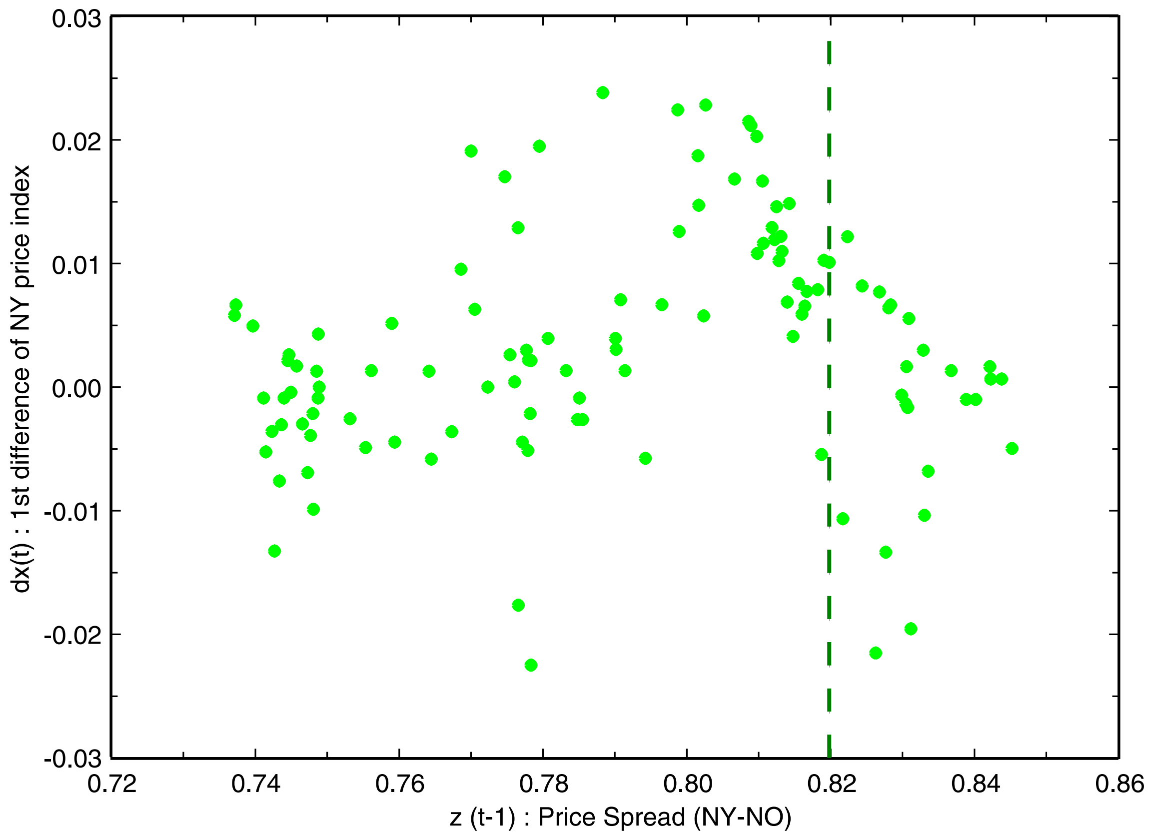

Fig. 2. Threshold property of data.

Table 5
Least-squares estimates for 28 bivariate systems with the benchmark city, New Orleans

City	$\hat{\alpha}_1$	$\hat{\alpha}_2$	City	$\hat{\alpha}_1$	$\hat{\alpha}_2$
NY	.166(.029)	.185(.031)	Po	.056(.016)	.067(.018)
	.15(.026)	.168(.028)		.027(.014)	.031(.017)
PH	.089(.028)	.099(.03)	BU	−.003(.022)	−.014(.025)
	.064(.025)	.07(.027)		−.037(.013)	−.051(.015)
CH	.172(.031)	.191(.034)	DA	.126(.023)	.158(.028)
	.158(.031)	.178(.033)		.023(.02)	.026(.025)
LA	−.144(.049)	−.166(.053)	AT	.101(.035)	.11(.038)
	−.156(.043)	.177(.047)		.106(.026)	.115(.028)
SF	−.163(.047)	−.187(.051)	AN	−.061(.019)	−.082(.022)
	−.157(.046)	−.18(.051)		−.048(.016)	−.062(.019)
Bo	.096(.019)	.111(.021)	DN	.119(.054)	.131(.058)
	.079(.016)	.097(.019)		.132(.037)	.141(.04)
Cl	.086(.024)	.106(.026)	DT	.154(.046)	.167(.049)
	.087(.022)	.104(.024)		.147(.039)	.16(.043)
Ci	−.071(.02)	−.094(.024)	MI	.101(.019)	.12(.022)
	−.067(.019)	−.085(.023)		.069(.017)	.081(.019)
DC	−.362(.067)	−.399(.072)	KC	.064(.015)	.079(.018)
	−.262(.061)	−.289(.067)		.026(.014)	.03(.017)
Ba	.127(.029)	.142(.032)	HS	−.105(.046)	−.12(.05)
	.111(.032)	.123(.035)		−.092(.041)	−.107(.044)
SL	−.045(.019)	−.062(.022)	HO	−.168(.037)	−.191(.041)
	−.056(.016)	−.071(.018)		−.142(.032)	−.162(.036)
MS	−.098(.022)	−.126(.026)	PI	−.019(.019)	−.032(.021)
	−.056(.022)	−.072(.027)		−.046(.014)	−.06(.016)
Ma	.079(.014)	.1(.016)	TA	−.1(.044)	.221(.1)
	.06(.011)	.076(.012)		−.076(.036)	.216(.082)
SD	−.105(.03)	−.113(.032)	SE	.101(.017)	.122(.02)
	−.089(.026)	−.097(.028)		.054(.019)	.063(.023)

The lag lengths are selected by BIC. The standard errors are in the parentheses. Cities Abbv.: 1. Anchorage AN, 2. Atlanta AT, 3. Baltimore BT, 4. Boston BO, 5. Buffalo BU, 6. Chicago CH, 7. Cleveland CL, 8. Cincinnati CI, 9. Dallas DA, 10. Denver DN, 11. Detroit DT, 12. Honolulu HO, 13. Houston HS, 14. Kansas City KC, 15. Los Angeles LA, 16. Miami MA, 17. Milwaukee MI, 18. Minneapolis MS, 19. New York NY, 20. Philadelphia PH, 21. Pittsburgh PI, 22. Portland PO, 23. San Diego SD, 24. San Francisco SF, 25. Seattle SE, 26. St. Louis SL, 27. Tampa Bay TA, 28. Washington DC, Benchmark: New Orleans NO.

of rejected systems drops to 9, while HW can reject only one. In particular, ADF cannot reject the null in any of 28 systems. The estimates for $\alpha_i's$ are reported in Table 5, while we need be cautious about the interpretation. Note that positive values of $\alpha_i's$ do not directly imply that the series is explosive in the TVECM (Table 6).

6. Concluding remarks

We have developed the supW test examining the linear no cointegration null hypothesis in the Band TVECM, a residual-based bootstrap for the test, and

Table 6
Cointegration tests for LOP using the price indexes of used car markets from 29 different locations (28 bivariate systems with the benchmark city, New Orleans)

Cities	sup $W^{\text{no lag}}$ (p-value)	$HW^{\text{no lag}}$ (10% : 8.3)	sup W^{BIC} (p-value)	HW^{BIC} (10% : 8.3)	ADF (10% : −2.6)
NY	0	3.652	0.009	1.214	.685
PH	.034	3.2	0.661	1.109	−.303
CH	0	4.509	0.22	0.285	1.602
LA	0	4.261	0.372	1.339	.863
SF	.235	6.133	0.7	1.549	.326
BO	0	14.162	0.293	0.7	3.589
CL	.195	4.036	0.432	5.371	−1.862
CI	0	15.905	0.71	2.017	3.843
DC	.446	0.439	0.839	0.822	.1
BA	.009	1.609	0.061	3.621	−.322
SL	0	8.83	0.087	0.03	2.969
MS	0	3.98	0.567	0.062	1.151
MA	0	3.484	0.052	0.541	.889
SD	.15	5.997	0.299	3.789	−.946
PO	.035	2.406	0.515	1.572	.895
BU	0	7.623	0.038	0.707	2.736
DA	0	2.153	0.034	0.264	.456
AN	.002	4.019	0.091	1.204	.359
AT	.02	8.795	0.064	7.714	−.625
DN	.029	12.761	0.044	12.938	−2.051
DT	.313	2.654	0.536	0.342	.56
ML	.043	3.833	0.677	2.86	.307
KC	0	2.161	0.299	0.507	−.032
HS	.019	1.029	0.506	0.511	−1.223
HO	.011	0.433	0.155	1.061	.647
PL	0	10.482	0.888	0.575	3.198
TA	.435	1.031	0.556	2.714	−.987
SE	.154	2.882	0.854	0.416	.378
Number of rejections					
10%	21	6	9	1	0
5%	21	4	4	1	0
1%	14	2	1	0	0

required asymptotic results to justify the test and the bootstrap. Furthermore, we have demonstrated the advantage of this test over the conventional cointegration tests within the context of the commonly used two-step approach.

Two issues have not been studied enough in the literature nor in this paper. First is how to test the linear no cointegration null hypothesis when we need to estimate the cointegrating vector because there is no relevant economic theory prespecifying it. The earlier version of this paper proposed a recursive estimation method, but the power improvement over the convention tests is not as good as expected. Second, the null of threshold no cointegration has never been studied, although the possibility of such a hypothesis can be raised conceptually. The chief difficulties are that the economic meaning of the hypothesis should be clarified and, second, the distribution theory under the hypothesis appears to be completely different from the one developed in this paper.

Acknowledgements

I would like to thank Bruce E. Hansen for his excellent advice, encouragement and insightful comments. I gratefully acknowledge helpful comments from an anonymous associate editor and two referees.

Appendix A. Proofs of Theorems

In the following proofs, let $\bar{\theta} = \max_{\theta \in \Gamma} |\theta|$ and $\sup_{\theta}$ means $\sup_{\theta \in \Theta}$ to simplify notation. The asymptotic distribution of supW does not depend on whether it is computed from two-regime or Band TVECM.

A.1. Proof of Theorem 1.

(a) We first show that

$$\sup_{\theta} \left| \frac{1}{n^2} \sum_{t=1}^{n} (z_{t-1}^2 1\{z_{t-1} \leqslant \theta\}) - \frac{1}{n^2} \sum_{t=1}^{n} (z_{t-1}^2 1\{z_{t-1} \leqslant 0\}) \right| \to 0. \tag{12}$$

To see this, note that, for arbitrary $\theta \in \Theta$,

$$\begin{aligned} &\left| \frac{1}{n^2} \sum_{t=1}^{n} (z_{t-1}^2 1\{z_{t-1} \leqslant \theta\}) - \frac{1}{n^2} \sum_{t=1}^{n} (z_{t-1}^2 1\{z_{t-1} \leqslant 0\}) \right| \\ &\quad \leqslant \frac{1}{n^2} \sum_{t=1}^{n} (z_{t-1}^2 |1\{z_{t-1} \leqslant \theta\} - 1\{z_{t-1} \leqslant 0\}|) \\ &\quad \leqslant \frac{1}{n^2} \sum_{t=1}^{n} (z_{t-1}^2 1\{|z_{t-1}| \leqslant \bar{\theta}\}) \\ &\quad \leqslant \frac{1}{n^2} \sum_{t=1}^{n} \bar{\theta}^2 \to 0. \end{aligned} \tag{13}$$

Next, if (2), (6), and Assumption 1 hold, then it follows from multivariate invariance principle (see, for example, Proposition 18.1 in Hamilton, 1994) and the continuous mapping theorem that

$$\frac{z_{[nr]}}{\sqrt{n}} = \frac{x'_{[nr]}\beta}{\sqrt{n}} \Rightarrow \beta' \Phi(1)^{-1} U(r) = W(r), \quad 0 \leqslant r \leqslant 1.$$

Furthermore, the transformation $s^2 1\{s \leqslant 0\}$ is continuous in s. Therefore, we have from the continuous mapping theorem that

$$\frac{1}{n^2}\sum_{t=1}^{n}(z_{t-1}^2 1\{z_{t-1}\leqslant 0\}) = \int_0^1 \left(\frac{z_{[nr]}}{\sqrt{n}}\right)^2 1\left\{\frac{z_{[nr]}}{\sqrt{n}}\leqslant 0\right\} \Rightarrow \int_0^1 W^2 1\{W\leqslant 0\}.$$

Remark 1. Since $x'_{[nr]}\beta/\sqrt{n}$ is a sequential empirical process (van der Vaart and Wellner (1996) p. 225: by Beverage–Nelson decomposition we can apply the result with a little adjustment), $x'_{[nr]}\beta/\sqrt{n}$ converges weakly to a tight limit in a metric space equipped with supremum norm whose marginal distribution is a Brownian motion W. Since (12) does not depend on $\beta \in \mathscr{R}^p$, the above result readily extends to the weak convergence on $\Theta \otimes \mathscr{R}^p$.

(b) Denote $S_n(\theta) = 1/n\sum_{t=1}^{n} z_{t-1}1\{z_{t-1}\leqslant\theta\}\varepsilon_t$. Similarly in (a), we first show that

$$\sup_{\theta}|S_n(\theta)-S_n(0)| \to_{\mathrm{p}} 0. \tag{14}$$

Note that for any $\theta \in \Theta$,

$$|S_n(\theta)-S_n(0)| \leqslant \frac{1}{n}\sum_{t=1}^{n}|z_{t-1}1\{|z_{t-1}|\leqslant\bar{\theta}\}\varepsilon_t|. \tag{15}$$

Since the choice of θ is arbitrary, the supremum in (14) is bounded above by the right-hand side term in (15), which vanishes asymptotically because

$$\begin{aligned} E\left\{\frac{1}{n}\sum_{t=1}^{n}|z_{t-1}1\{|z_{t-1}|\leqslant\bar{\theta}\}\varepsilon_t|\right\} &\leqslant \frac{1}{n}\sum_{t=1}^{n}\sqrt{E[|\varepsilon_t|^2]}\sqrt{E[z_{t-1}^2 1\{|z_{t-1}|\leqslant\bar{\theta}\}]} \\ &\leqslant \sqrt{E[|\varepsilon_t|^2]\bar{\theta}^2}\,\frac{1}{n}\sum_{t=1}^{n}\sqrt{E[1\{|z_{t-1}|\leqslant\bar{\theta}\}]} \\ &\to 0 \end{aligned} \tag{16}$$

as $n\to\infty$. For the convergence (16), note that, for any $\varepsilon>0$,

$$\begin{aligned} \lim_{n\to\infty}\frac{1}{n}\sum_{t=1}^{n}E[1\{|z_{t-1}|\leqslant\bar{\theta}\}] &= \lim_{n\to\infty}\int_0^1 E\left[1\left\{\left|\frac{z_{[nr]}}{\sqrt{n}}\right|\leqslant\frac{\bar{\theta}}{\sqrt{n}}\right\}\right] \\ &\leqslant \lim_{n\to\infty}\int_0^1 E\left[1\left\{\left|\frac{z_{[nr]}}{\sqrt{n}}\right|\leqslant\varepsilon\right\}\right] \\ &= \int_0^1 \lim_{n\to\infty} E\left[1\left\{\left|\frac{z_{[nr]}}{\sqrt{n}}\right|\leqslant\varepsilon\right\}\right] \\ &= \int_0^1 E[1\{|W(r)|\leqslant\varepsilon\}] \\ &= E\left[\int_0^1 1\{|W(r)|\leqslant\varepsilon\}\right]. \end{aligned}$$

Since $\bar{\theta}/\sqrt{n}<\varepsilon$ for a large enough n, we have the inequality. Then, the remaining equalities follow from the dominated convergence theorem, the definition of convergence in distribution, and the Tonelli theorem. Since this choice of ε is arbitrary and $\int_0^1 1\{|W(r)|\leqslant\varepsilon\}\to_{\text{a.s.}} 0$ as $\varepsilon\to 0$ (refer to the definition of local time), we see that the last term vanishes by applying the dominated convergence theorem again.

Next, the convergence

$$\frac{1}{n}\sum_{t=1}^{n} z_{t-1}1\{z_{t-1}\leqslant 0\}\varepsilon_t \Rightarrow \int_0^1 W1\{W\leqslant 0\}\,\mathrm{d}U \tag{17}$$

follows from Kurtz and Protter (1991) since $s1\{s\leqslant 0\}$ is continuous. □

A.2. Proof of Theorem 2.

We consider nonsingular transformations on x_t. Specifically, let T concatenate β and $e_2,\ldots,e_p$, where $e_i=(0,\ldots 0,1,0,\ldots,0)'$, the ith element being 1, so that the first element of $T'x_t$ corresponds to the cointegration term z_t and the cointegrating vector becomes $(1,0,\ldots,0)'$ for the transformed time series. This can be put more clearly by postmultiplying T on both sides of (1) so that (1) can be written as

$$\begin{aligned}\Delta\tilde{x}_t' &= \Delta x_t'T\\ &= z_{t-1}1\{z_{t-1}\leqslant\gamma_1\}\alpha_1'T + z_{t-1}1\{z_{t-1}>\gamma_2\}\alpha_2'T\\ &\quad + \mu'T + \Delta x_{t-1}'\Phi_1'T + \cdots + \Delta x_{t-q}'\Phi_q'T + \varepsilon_t'T\\ &= z_{t-1}1\{z_{t-1}\leqslant\gamma_1\}\tilde{\alpha}_1 + z_{t-1}1\{z_{t-1}>\gamma_2\}\tilde{\alpha}_2\\ &\quad + \tilde{\mu} + \Delta x_{t-1}'\tilde{\Phi}_1 + \cdots + \Delta x_{t-q}'\tilde{\Phi}_q + e_t',\end{aligned}$$

where $e_t'=\varepsilon_t'T$, $\tilde{\alpha}_1=\alpha_1'T$, and $\tilde{\mu}$ and $\tilde{\Phi}_i's$ are defined accordingly. First, note that $\tilde{\alpha}_1=\tilde{\alpha}_2=0$ is equivalent to $\alpha_1=\alpha_2=0$, since T is positive definite. Therefore, we can test $\tilde{\alpha}_1=\tilde{\alpha}_2=0$ instead of $\alpha_1=\alpha_2=0$. Second, the OLS residual, $\hat{e}_t(\gamma)$, from this model is identical to $\hat{\varepsilon}_t'(\gamma)T$ for the OLS residual, $\hat{\varepsilon}_t(\gamma)$, from the original model, and $\hat{\Sigma}_e=T'\hat{\Sigma}T$. Thus, the Wald test statistic to test $\tilde{\alpha}_1=\tilde{\alpha}_2=0$ is equivalent to our original test statistic. To see this, write the test statistic in this case

$$\begin{aligned}&\operatorname{tr}\left\{(Z_\gamma'M_{-1}e\hat{\Sigma}_e^{-1/2})'(Z_\gamma'M_{-1}Z_\gamma)^{-1}(Z_\gamma'M_{-1}e\hat{\Sigma}_e^{-1/2})\right\}\\ &\quad=\operatorname{tr}\left\{(Z_\gamma'M_{-1})'(Z_\gamma'M_{-1}Z_\gamma)^{-1}(Z_\gamma'M_{-1}\varepsilon T(T'\hat{\Sigma}T)^{-1}T'\varepsilon')\right\}\end{aligned}$$

and observe that the second line is equivalent to (5). Hence, the argument above is just a rewriting of (5).

Let $\imath$ be a vector of ones, $\bar{\Delta}X_{-1}$ be the matrix of stacking the demeaned lagged terms $\bar{\Delta}x_{t-j}'$s, and $\bar{Z}_\gamma$ be the demeaned Z_γ. And note that for $j=1,\ldots,q$,

$$\sup_\theta\left|\frac{1}{n\sqrt{n}}\sum_{t=1}^{n} z_{t-1}1\{0<z_{t-1}\leqslant\theta\}\Delta x_{t-j}'\right|\leqslant\frac{1}{n\sqrt{n}}\sum_{t=1}^{n}|z_{t-1}1\{|z_{t-1}|\leqslant\bar{\theta}\}\Delta x_{t-j}'| \to_{\mathrm{p}} 0 \tag{18}$$

by the same reasoning as in the Proof of Theorem 1(b), and that

$$\frac{1}{n\sqrt{n}}\sum_{t=1}^{n} z_{t-1}1\{z_{t-1}\leqslant 0\}\Delta x'_{t-j}\to_{\mathrm{p}}0 \tag{19}$$

by Theorem 3.3 of Hansen (1992), in which the result is shown for $j=1$, and the same proof applies for the other j's. Furthermore, $1/n\sqrt{n}\sum_{t=1}^{n} z_{t-1}1\{z_{t-1}\leqslant 0\}\,1/n\sum_{t=1}^{n}\Delta x'_{t-j}\to_{\mathrm{p}}0$, since $E\Delta x'_{t-j}=0$. Therefore, we have

$$\frac{1}{n\sqrt{n}}\,Z'_{\gamma}\bar{\Delta}X_{-1}=\mathrm{o_p}(1).$$

Furthermore, we can write the projection onto the space of the constant and the lagged terms as the sum of two orthogonal projections, that is, the projections onto $\imath$ and onto the $\bar{\Delta}X_{-1}$. Therefore,

$$\begin{aligned}
&\frac{1}{n}\,Z'_{\gamma}M_{-1}e\hat{\Sigma}_e^{-1/2}\\
&\quad=\frac{1}{n}\,Z'_{\gamma}(I_n-\imath(\imath'\imath)^{-1}\imath'-\bar{\Delta}X_{-1}(\bar{\Delta}X'_{-1}\bar{\Delta}X_{-1})^{-1}\bar{\Delta}X'_{-1})e\hat{\Sigma}_e^{-1/2}\\
&\quad=\frac{1}{n}\,\bar{Z}'_{\gamma}e\hat{\Sigma}_e^{-1/2}-\frac{1}{n\sqrt{n}}\,Z'_{\gamma}\bar{\Delta}X_{-1}\left(\frac{1}{n}\,\bar{\Delta}X'_{-1}\bar{\Delta}X_{-1}\right)^{-1}\frac{1}{\sqrt{n}}\bar{\Delta}X'_{-1}e\hat{\Sigma}_e^{-1/2}\\
&\quad=\frac{1}{n}\,\bar{Z}'_{\gamma}\,e\hat{\Sigma}_e^{-1/2}+\mathrm{o_p}(1)
\end{aligned} \tag{20}$$

and similarly

$$\begin{aligned}
\frac{1}{n^2}Z'_{\gamma}M_{-1}Z_{\gamma}&=\frac{1}{n^2}\,\bar{Z}'_{\gamma}\bar{Z}_{\gamma}-\frac{1}{n\sqrt{n}}\,Z'_{\gamma}\bar{\Delta}X_{-1}\left(\frac{1}{n}\bar{\Delta}X'_{-1}\bar{\Delta}X_{-1}\right)^{-1}\left(\frac{1}{n\sqrt{n}}Z'_{\gamma}\bar{\Delta}X_{-1}\right)'\\
&=\frac{1}{n^2}\bar{Z}'_{\gamma}\bar{Z}_{\gamma}+\mathrm{o_p}(1).
\end{aligned} \tag{21}$$

Note that (18), (19), and Theorem 1 is sufficient for the consistency of $\hat{\Phi}'_j s$, $\hat{\mu}$, and $\hat{\Sigma}$.[5] It follows from (20), (21), and Theorem 1 that

$$\frac{1}{n}\,Z'_{\gamma}M_{-1}e\hat{\Sigma}_e^{-1/2}\Rightarrow\int_0^1\left(\tilde{W}-\int_0^1\tilde{W}\right)\mathrm{d}B'=\omega\int_0^1\tilde{B}_1\,\mathrm{d}B',$$

$$\frac{1}{n^2}\,Z'_{\gamma}M_{-1}Z_{\gamma}\Rightarrow\int_0^1\left(\tilde{W}-\int_0^1\tilde{W}\right)\left(\tilde{W}-\int_0^1\tilde{W}\right)'=\omega^2\int_0^1\tilde{B}_1\tilde{B}'_1,$$

where $\tilde{W}=(W1\{W\leqslant 0\},W1\{W>0\})'$. B, B_1, and $\tilde{B}_1$ are as defined in the statement of this theorem. Note that W can be written ωB_1 for the long-run variance of $x'_t\beta,\omega^2$, since $x'_t\beta$ is the first element of the transformed series, and that

[5]The same logic that shows the consistency of these estimators in the standard linear VECM applies here. In other words, our TVECM replaces z_{t-1} with $z_{t-1}1\{z_{t-1}\leqslant\gamma_1\}$ and $z_{t-1}1\{z_{t-1}>\gamma_2\}$. Therefore, we only have to show that $1/n\sqrt{n}\sum_{t=1}^{n} z_{t-1}1\{z_{t-1}\leqslant\gamma_1\}\Delta x_{t-j}=\mathrm{o_p}(1)$, instead of $1/n\sqrt{n}\sum_{t=1}^{n} z_{t-1}\Delta x_{t-j}=\mathrm{o_p}(1)$, and that $1/n\sqrt{n}\sum_{t=1}^{n} z_{t-1}1\{z_{t-1}\leqslant\gamma_1\}=\mathrm{O_p}(1)$, instead of $1/n\sqrt{n}\sum_{t=1}^{n} z_{t-1}=\mathrm{O_p}(1)$.

$1\{W \leqslant 0\} = 1\{\omega B_1 \leqslant 0\} = 1\{B_1 \leqslant 0\}$. Therefore, the remaining steps for the convergence of supW follow from the continuous mapping theorem. $\square$

A.3. Proof of Theorem 3.

The superscript * is used to signify the bootstrap sample, statistics and so on. For example, $\mathbf{P}^*(\mathbf{E}^*)$ is the bootstrap probability(expectation) conditional on the realization of the original sample and $\overset{*}{\Rightarrow}$, $\mathrm{O}_\mathrm{p}^*(1)$ and $\mathrm{o}_\mathrm{p}^*(1)$ are defined with respect to $\mathbf{P}^*$. For a detailed description of these notations, see Chang and Park (2003). Let $u_t = \Delta x_t$.

We first establish a bootstrap invariance principle:

$$\frac{1}{\sqrt{n}} \sum_{t=1}^{[nr]} \varepsilon_t^* \overset{*}{\Rightarrow} U(r) \quad \text{in } \mathbf{P}. \tag{22}$$

As in Theorem 2.2 of Park (2002) who establishes an invariance principle for the sieve bootstrap, we prove (22) based on the strong approximation (see Lemma 2.1 of Park (2002) which is a rewriting of the result in Sahanenko (1982) and Theorem 1 of Einmahl (1987) for a multidimensional version). In particular, Einmahl (1987) requires the existence of the second moment. Since any projection matrix is positive semi-definite, however, we have $1/n \sum_{t=1}^n \hat{\varepsilon}_t \hat{\varepsilon}_t' \leqslant 1/n \sum_{t=1}^n \varepsilon_t \varepsilon_t'$ and thus

$$\mathbf{E}^* \varepsilon_t^* \varepsilon_t^{*\prime} = \frac{1}{n} \sum_{t=1}^n \hat{\varepsilon}_t \hat{\varepsilon}_t' \leqslant \frac{1}{n} \sum_{t=1}^n \varepsilon_t \varepsilon_t' \to \Sigma \quad \text{a.s.} \tag{23}$$

by the strong law of large numbers. Since $\mathbf{E}^* \varepsilon_t^* \varepsilon_t^{*\prime}$ is finite for almost every realization, we can apply the strong approximation conditional on each realization of data. Therefore, the invariance principle (22) follows if we show that $\mathbf{E}^* |\varepsilon_t^*|^r = \mathrm{O}_\mathrm{p}(1)$ for $r > 2$. Since $(\sum_{i=1}^c a_i)^r \leqslant c^{r-1} \sum_{i=1}^c |a_i|^r$, without loss of generality, we have[6]

$$\begin{aligned}
\frac{1}{n} \sum_{t=1}^n |\hat{\varepsilon}_t - \varepsilon_t|^r \leqslant & \sum_{j=1}^q |\hat{\Phi}_j - \Phi_j|^r \frac{1}{n} \sum_{t=1}^n |u_{t-j}|^r + |\hat{\mu}|^r \\
& + |n\hat{\alpha}_1|^r \frac{1}{n^{1+r}} \sum_{t=1}^n |z_{t-1} 1\{z_{t-1} \leqslant \gamma_1\}|^r \\
& + |n\hat{\alpha}_2|^r \frac{1}{n^{1+r}} \sum_{t=1}^n |z_{t-1} 1\{z_{t-1} > \gamma_2\}|^r \\
= & \ \mathrm{o}_\mathrm{p}(1)
\end{aligned}$$

and therefore,

$$\mathbf{E}^* |\varepsilon_t^*|^r = \frac{1}{n} \sum_{t=1}^n |\hat{\varepsilon}_t|^r \leqslant \frac{1}{n} \sum_{t=1}^n |\hat{\varepsilon}_t - \varepsilon_t|^r + \frac{1}{n} \sum_{t=1}^n |\varepsilon_t|^r \to_\mathrm{p} \mathbf{E} |\varepsilon_t|^r. \tag{24}$$

Next we develop the bootstrap invariance principle with $u_t^* = \hat{\Phi}(L)^{-1} \varepsilon_t^*$.

[6] The consistency of the estimators is established in the middle of the proof of Theorem 2.

Let $\tilde{u}_t^* = \hat{\Phi}(1)^{-1}\sum_{j=1}^{q}(\sum_{i=j}^{q}\hat{\Phi}_i)u_{t-j+1}^*$. Then we have

$$\frac{1}{\sqrt{n}}\sum_{t=1}^{[nr]} u_t^* = \hat{\Phi}(1)^{-1}\frac{1}{\sqrt{n}}\sum_{t=1}^{[nr]}\varepsilon_t^* + \frac{1}{\sqrt{n}}(\tilde{u}_0^* - \tilde{u}_{[nr]}^*)$$
$$\overset{*}{\Rightarrow} \Phi(1)^{-1}U(r) \quad \text{in } \mathbf{P}.$$

The equality follows from (13) of Park (2002), and the convergence holds by the consistency of $\hat{\Phi}_j$, $j = 1, \ldots, q$, and by Theorem 3.3 of Park (2002) which shows $\max_{0\leqslant r\leqslant 1}|\tilde{u}_{[nr]}^*/\sqrt{n}| = o_p(1)$. Therefore, by the continuous mapping theorem, we have

$$\frac{1}{\sqrt{n}}z_{[nr]}^* = \frac{1}{\sqrt{n}}\sum_{t=1}^{[nr]}\beta' u_t^* \overset{*}{\Rightarrow} \beta'\Phi(1)^{-1}U(r) \quad \text{in } \mathbf{P}. \tag{25}$$

Equipped with the invariance principle (25), the proof of Theorem 1 is also valid with the bootstrap sample. Since (12) is still valid with z_t replaced by z_t^*, the continuous mapping theorem and the invariance principle (25) yield

$$\frac{1}{n^2}\sum_{t=1}^{n} z_{t-1}^{*2}1\{z_{t-1}^*\leqslant\theta\} \overset{*}{\Rightarrow} \int_0^1 W^2 1\{W\leqslant 0\} \quad \text{in } \mathbf{P}.$$

Similarly, to show

$$\frac{1}{n}\sum_{t=1}^{n} z_{t-1}^* 1\{z_{t-1}^*\leqslant\theta\}\varepsilon_t^* \overset{*}{\Rightarrow} \int_0^1 W 1\{W\leqslant 0\}\,\mathrm{d}U \quad \text{in } \mathbf{P}, \tag{26}$$

we need to establish (14) with z_t^* and ε_t^*. As in the proof of Theorem 1, we note that

$$\begin{aligned} E^*\left\{\frac{1}{n}\sum_{t=1}^{n}|z_{t-1}^* 1\{|z_{t-1}^*|\leqslant\bar{\theta}\}\varepsilon_t^*|\right\} &\leqslant \frac{1}{n}\sum_{t=1}^{n}\sqrt{E^*[|\varepsilon_t^*|^2]}\sqrt{E^*[z_{t-1}^{*2}1\{|z_{t-1}^*|\leqslant\bar{\theta}\}]} \\ &\leqslant \sqrt{E^*[|\varepsilon_t^*|^2]\bar{\theta}^2}\,\frac{1}{n}\sum_{t=1}^{n}\sqrt{E^*[1\{|z_{t-1}^*|\leqslant\bar{\theta}\}]} \\ &= o_p(1), \end{aligned}$$

since the first term in the third line $E^*[|\varepsilon_t^*|^2] = o_p(1)$ by (24) and the second term is also $o_p(1)$ by the same reasoning yielding (16). That is, for any $\varepsilon>0$, by the dominated convergence theorem and the invariance principle (25),

$$\begin{aligned} \lim_{n\to\infty}\frac{1}{n}\sum_{t=1}^{n}E[1\{|z_{t-1}^*|\leqslant\bar{\theta}\}] &= \lim_{n\to\infty}\int_0^1 E\left[1\left\{|z_{[nr]}^*/\sqrt{n}|\leqslant\bar{\theta}/\sqrt{n}\right\}\right] \\ &\leqslant \lim_{n\to\infty}\int_0^1 E\left[1\left\{|z_{[nr]}^*/\sqrt{n}|\leqslant\varepsilon\right\}\right] \\ &= \int_0^1 \lim_{n\to\infty}E\left[1\left\{|z_{[nr]}^*/\sqrt{n}|\leqslant\varepsilon\right\}\right] \\ &= \int_0^1 E[1\{|W(r)|\leqslant\varepsilon\}], \end{aligned}$$

which can be made arbitrarily small choosing ε sufficiently small for the same reason in the Proof of Theorem 1. Since the transformation $x \to x1\{x \leqslant 0\}$ is continuous and the second moment of ε_t^* is finite a.s. by (23), we apply Kurtz and Protter (1991) to get (26).

Then, we need to go over the Proof of Theorem 2 with the bootstrap sample. Many parts of this step resemble Chang and Park (2003), which establishes the consistency of the bootstrap of the ADF test. The main difference is that we have $z_{t-1}^* 1\{z_{t-1}^* \leqslant \gamma_1\}$ and $z_{t-1}^* 1\{z_{t-1}^* > \gamma_2\}$ instead of z_{t-1}^*. Therefore, the convergences (18) and (19) with the bootstrap sample are of main concern.

First, note that $1/n \bar{\Delta} X_{-1}^{*\prime} \bar{\Delta} X_{-1}^*$ and $1/\sqrt{n} \bar{\Delta} X_{-1}^{*\prime} e^*$ are $\mathrm{O}_\mathrm{p}(1)$ by Lemma 3.3 of Chang and Park (2003). Then, (18) holds with the bootstrap sample as in (26). Second, we show (19), i.e.,

$$\frac{1}{n\sqrt{n}} \sum_{t=1}^{n} z_{t-1}^* 1\{z_{t-1}^* \leqslant 0\} u_{t-j}^{*\prime} = \mathrm{o}_\mathrm{p}(1)$$

by following the proof of Theorem 3.3 of Hansen (1992). Since $z_{t-1}^* 1\{z_{t-1}^* \leqslant 0\}/\sqrt{n}$ converges to $W1\{W \leqslant 0\}$ by the continuous mapping theorem, it remains to show that $1/n \sum_{t=1}^{n} |u_t^*| = \mathrm{O}_\mathrm{p}(1)$ and $|1/n \sum_{t=1}^{n} u_t^*| = \mathrm{o}_\mathrm{p}(1)$. The first is straightforward from the inequality (25) in Park (2002) since $\mathbf{E}^* |\varepsilon_t^*|^r = \mathrm{O}_\mathrm{p}(1)$. For the second, note that we can write, by the Beverage–Nelson decomposition and by the Proof of Theorem 3.3 of Park (2002),

$$\mathbf{E}^* \left| \frac{1}{n} \sum_{t=1}^{n} u_t^* \right| = \hat{\Phi}(1)^{-1} \mathbf{E}^* \left| \frac{1}{n} \sum_{t=1}^{n} \varepsilon_t^* \right| + \mathrm{o}_\mathrm{p}(1)$$

and that $\mathbf{E}^* |1/n \sum_{t=1}^{n} \varepsilon_t^*| = \mathrm{o}_\mathrm{p}^*(1)$ by Theorem 1 of Andrews (1988). Therefore, we have shown that $(1/n\sqrt{n}) Z_\gamma^{*\prime} \bar{\Delta} X_{-1}^* = \mathrm{o}_\mathrm{p}(1)$, which in turn implies that (20) and (21) hold with the bootstrap sample. The remaining steps are straightforward, since they are basically the same as in the bootstrap of the ADF test. $\square$

References

Andrews, D.W.K., 1988. Laws of large numbers for dependent non-identically distributed random variables. Econometric Theory 4, 458–467.

Andrews, D., Ploberger, W., 1994. Optimal tests when a nuisance parameter is present only under the alternative. Econometrica 62, 1383–1414.

Balke, N., Fomby, T., 1997. Threshold cointegration. International Economic Review 38, 627–645.

Basawa, I.V., Mallik, A.K., McCormick, W.P., Reeves, J.H., Taylor, R.L., 1991. Bootstrapping unstable first-order autoregressive processes. The Annals of Statistics 19, 1098–1101.

Chan, K.S., Petruccelli, J.D., Tong, H., Woolford, S.W., 1985. A multiple-threshold AR(1) model. Journal of Applied Probability 22, 267–279.

Chang, Y., Park, J.Y., 2003. A sieve bootstrap for the test of a unit root. Journal of Time Series Analysis 24, 379–400.

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

Davies, R., 1987. Hypothesis testing when a nuisance parameter is present only under the alternative. Biometrika 74, 33–43.

Einmahl, U., 1987. A useful estimate in the multidimensional invariance principle. Probability Theory and Related Fields 76, 81–101.
Enders, W., Granger, C., 1998. Unit root tests and asymmetric adjustment with an example using the term structure of interest rates. Journal of Business and Economic Statistics 16, 304–311.
Enders, W., Siklos, P.L., 2001. Cointegration and threshold adjustment. Journal of Business and Economic Statistics 19 (2), 166–176.
Hamilton, J.D., 1994. Time Series Analysis. Princeton University Press, New Jersey.
Hansen, B., 1992. Convergence to stochastic integrals for dependent heterogeneous processes. Econometric Theory 8, 489–500.
Hansen, B., 1996. Inference when a nuisance parameter is not identified under the null hypothesis. Econometrica 64, 413–430.
Hansen, B., Seo, B., 2002. Testing for two-regime threshold cointegration in vector error correction models. Journal of Econometrics 110, 293–318.
Horvath, M., Watson, M., 1995. Testing for cointegration when some of the cointegrating vectors are prespecified. Econometric Theory 11, 984–1014.
Kurtz, T., Protter, P., 1991. Weak limit theorems for stochastic integrals and stochastic differential equations. Annals of Probability 19, 1035–1070.
Lo, M., Zivot, E., 2001. Threshold cointegration and nonlinear adjustment to the law of one price. Macroeconomic Dynamics 5, 533–576.
Paparoditis, E., Politis, D.N., 2003. Residual-based block bootstrap for unit root testing. Econometrica 71, 813–855.
Park, J., Phillips, P., 2001. Nonlinear regressions with integrated time series. Econometrica 69, 117–161.
Park, J.Y., 2002. An invariance principle for sieve bootstrap in time series. Econometric Theory 18, 469–490.
Pippenger, M., Goering, G., 2000. Additional results on the power of unit root and cointegration tests under threshold processes. Applied Economic Letters 7, 641–644.
Sahanenko, A., 1982. On unimprovable estimates of the rate of convergence in invariance principle. In: Gnedenko, B.V., Puri, M.L., Vincze, I. (Eds.), Nonparametric Statistical Inference, vol. 2. North-Holland, Amsterdam, pp. 779–784.
Taylor, A.M., 2001. March, Potential pitfalls for the purchasing power parity puzzle-sampling and specification biases in mean-reversion tests of the law of one price. Econometrica 69, 473–498.
van der Vaart, A.W., Wellner, J.A., 1996. Weak Convergence and Empirical Process. Springer, New York.
Watson, M.W., 1994. Vector autoregressions and cointegration. In: Engle, R.F., McFadden, D.L. (Eds.), Handbook of Econometrics, vol. 4. North-Holland, Amsterdam, pp. 2843–2915.