

## NONLINEAR ERROR CORRECTION MODEL AND MULTIPLE-THRESHOLD COINTEGRATION

Man Wang<sup>1</sup>, Ngai Hang Chan<sup>2,3</sup> and Chun Yip Yau<sup>3</sup>

<sup>1</sup>*Dong Hua University*, <sup>2</sup>*Southwestern University of Finance and Economics*  
and <sup>3</sup>*The Chinese University of Hong Kong*

*Abstract:* As an extension of linear cointegration, threshold cointegration has been a vibrant research topic in finance and statistics. Existing estimation procedures of threshold cointegration are usually based on the threshold vector error correction model (TVECM); however, only one threshold cointegration is considered. In this paper, we investigate estimation of the multiple-threshold cointegration that is more widely used in application. Two proposed methods, the LSE and the Smoothed LSE are studied, via the multiple-regime TVECM. The convergence rate of the LSE is obtained and the limiting distribution of the smoothed LSE is developed. To assess the performance of these estimators, a simulation study was conducted, in which the results support the asymptotic theories. As an example, we study the term structure of interest rates by a two-threshold cointegration model.

*Key words and phrases:* Convergence rate, error correction model, multiple-threshold cointegration, smoothed least squares estimator, super consistency.

### 1. Introduction

The idea of co-integration stems from the notion of transforming multiple time series into stationary components, as first discussed in Box and Tiao (1977). This idea was subsequently codified in the notion of cointegration in Granger (1983) and Engle and Granger (1987) in econometrics front. The basic idea of cointegration is that given two (or more) nonstationary time series, a linear (or non-linear) combination of them may be stationary. The classic cointegration implies constant adjustment towards the long-run equilibrium, which was found to be too strict in certain situations. Balke and Fomby (1997) argued that the existence of transaction cost may lead to nonlinear adjustment to long-run equilibrium and proposed the idea of threshold cointegration. Their model suggests that there exist disjoint regimes determined by the size or the sign of the equilibrium error and system adjustments are different from regime to regime. In the past two decades, estimation methods of threshold cointegration have been extensively pursued, including Balke and Fomby (1997), Hansen and Seo (2002) and Seo (2011). Some testing procedures have also been proposed. For instance,

Hansen and Seo (2002) consider testing for the presence of threshold effect assuming the presence of cointegration, and Seo (2006) studies testing for the presence of cointegration within the threshold cointegration model. In applications, this model has been widely used in econometrics and finance, including the study of purchasing power parity (PPP), law of one price (LOP), and the relation between long-term and short-term interest rates, see Walter and Kamol (2004), Enders and Siklos (2001), Enders and Granger (1998), Michael, Nobay, and Peel (1997) and Lo and Zivot (2001), among others.

Although several estimation methods have been proposed for cointegration models, theoretical results are rare. Even estimation consistency may not be obtained by standard methods. As alluded by Saikkonen (1995), the log-likelihood function does not converge uniformly in the whole parameter space, and the directions in which it converges fastest determine the convergence rate of each component of the maximum likelihood estimators. Without uniform convergence, it is difficult to obtain the limiting distribution with Taylor expansions. Existing methods can be divided into two categories: two-step estimation and joint-estimation. The two-step estimation was proposed by Balke and Fomby (1997), based on Engle and Granger's two-step estimation of linear cointegration. They suggested estimating the long-run parameter (cointegrating vector) first and then plugging in the estimated parameter as if it were the true value and estimating the short-run parameters. Although this two-step procedure is based on the super-consistency of the estimate of the cointegrating vector, its global optimization is not yet clear. In fact, for a smooth transition cointegration model, de Jong (2001) found that the estimation error in the first step is non-negligible, unless some regularity conditions are satisfied.

The joint estimation methods estimate the long-run and short-run parameters jointly based on the corresponding threshold vector error correction model (TVECM) representations. For a two-regime TVECM, Hansen and Seo (2002) derived the maximum likelihood estimator (MLE) under the normality assumption. They also proposed a SupLM test for the presence of one threshold cointegration. However, consistency of the MLE is still an open question. For the same two-regime TVECM, Seo (2011) considered the LSE and the smoothed-LSE (SLSE). Under suitable conditions, Seo (2011) showed that the least squares estimates of the cointegrating vector and the threshold converge at rate  $n^{3/2}$  and  $n$ , respectively. In addition, the SLSE of the cointegrating vector and thresholds are found to converge jointly to a functional of Brownian motions, at rates slightly slower than that of the corresponding LSEs in Seo (2011).

Asymptotic theories are mainly focused on one threshold cointegration. However, there is an increasing interest in application of multiple-threshold cointegration. For example, Lo and Zivot (2001) applied a two-threshold VECM to

analyze the PPP and the LOP of tradable goods in the US. The large sample theories of the estimators of multiple-regime TVECM are still unknown, however. For models with multiple threshold parameters, Seo (2011) suggested the sequential estimation discussed in Bai and Perron (1998), but did not develop consistency of the estimator. For the multiple-threshold AR model without cointegration, Li and Ling (2012) first gave the limiting distribution of the LSE of the multiple thresholds. They found that each threshold estimate converges to the minimizer of a compound Poisson process.

The goal of this paper is to develop asymptotic theories of estimation of multiple-threshold cointegration, based on the corresponding multiple-regime TVECM representations. We first study the LSE of a general multiple-regime TVECM and obtain its convergence rate under suitable conditions. The LSEs of the thresholds and cointegrating vector are found to be  $n$ - and  $n^{3/2}$ -consistent, respectively, and the estimate of the slope parameter is asymptotically normal and  $\sqrt{n}$ -consistent. Moreover, the SLSE is explored and its limiting distribution is established. The SLSEs of the cointegrating vector and the multiple thresholds are found to be super-consistent and asymptotically mixed normal; this enables one to conduct statistical inference. The SLSE of the slope parameter has the same limiting distribution as the LSE. Estimation of asymptotic variance of the limiting distribution is given. A simulation study is conducted to assess the two proposed estimators and additionally, a three-regime TVECM is applied to the 120-month and twelve-month interest rates of the US in the period 1952-1991.

The paper is organized as follows. Section 2 investigates the convergence rate of LSE of multiple-regime TVECM. Section 3 establishes the limiting distribution of the SLSE. In Section 4, results of a simulation study and empirical study are reported. Section 5 concludes.

## 2. Least Squares Estimation

### 2.1. Model specification

Let  $\{x_t\}$  be a  $p$ -dimensional vector of  $I(1)$  time series that is cointegrated with a single cointegrating vector, denoted by  $(1, \beta)'$ . Define the error correction term  $z_t(\beta) = x_{1,t} + x'_{2,t}\beta$ , where  $x_t = (x_{1,t}, x'_{2,t})'$ . Then a general multiple-regime TVECM can be written as:

$$\Delta x_t = \sum_{j=1}^m A'_j X_{t-1}(\beta) I(\gamma_{j-1}, \gamma_j, z_{t-1}(\beta)) + u_t, \quad t = l+1, \dots, n, \quad (2.1)$$

where  $m$  is the number of regimes,  $I(\gamma_{j-1}, \gamma_j, z_{t-1}(\beta)) = 1\{\gamma_{j-1} \leq z_{t-1}(\beta) < \gamma_j\}$ ,  $X_{t-1}(\beta) = (1, z_{t-1}(\beta), \Delta x'_{t-1}, \dots, \Delta x'_{t-l+1})'$  with  $\Delta x_{t-i}, i = 1, \dots, l$  being the

lagged first order difference terms,  $\gamma = (\gamma_1, \dots, \gamma_{m-1})'$  is the vector of  $m - 1$  thresholds with  $-\infty = \gamma_0 < \gamma_1 < \dots < \gamma_m = \infty$ , and  $A_j$  is the coefficient in the  $j$ th regime. Throughout this paper,  $(\beta^0, \gamma^0, \lambda^0)$  indicates the true values of parameters,  $z_t = z_t(\beta^0)$  and  $X_t = X_t(\beta^0)$ .

To simplify the notation, (2.1) can be rewritten as:

$$y = \left[ \left( \tilde{X}_1(\beta, \gamma), \dots, \tilde{X}_m(\beta, \gamma) \right) \otimes I_p \right] \lambda + u, \tag{2.2}$$

where  $\tilde{X}_j(\beta, \gamma)$  is the matrix stacking  $X'_{t-1}(\beta)I(\gamma_{j-1}, \gamma_j, z_{t-1}(\beta))$ ,  $j = 1, \dots, m$ ,  $y$  and  $u$  are vectors that stack  $\Delta x_t$  and  $u_t$ , respectively,  $\lambda = \text{vec}((A'_1, A'_2, \dots, A'_m)')$ , and  $\otimes$  denotes the Kronecker product of two matrices. The column in  $A'_j$  associated with  $z_{t-1}(\beta)$  is denoted by  $\lambda^z_j$ .

Let  $\theta = (\beta', \gamma', \lambda)'$  and  $\Theta$  be the compact parameter space. If  $u(\theta) = y - \left[ \left( \tilde{X}_1(\beta, \gamma), \dots, \tilde{X}_m(\beta, \gamma) \right) \otimes I_p \right] \lambda$  and  $S_n(\theta) = u'(\theta)u(\theta)$ , then the LSE is defined as  $\hat{\theta} = \text{argmin}_{\theta \in \Theta} S_n(\theta)$ .

Here  $S_n(\theta)$  is not continuous with respect to  $\gamma$ , and  $\hat{\theta}$  does not have a closed form. Therefore, a grid search is necessary. Observing that once  $(\beta, \gamma)$  is fixed,  $\tilde{X}_j(\beta, \gamma)$  is observed and we denote it by  $\tilde{X}_j$ ,  $j = 1, \dots, m$ , so that the  $\hat{\lambda}$  that minimizes  $S_n(\beta, \gamma, \lambda)$  is simply the ordinary LSE:

$$\hat{\lambda}(\beta, \gamma) = \left( \left( \begin{matrix} \tilde{X}'_1 \tilde{X}_1 & \tilde{X}'_1 \tilde{X}_2 & \dots & \tilde{X}'_1 \tilde{X}_m \\ \tilde{X}'_2 \tilde{X}_1 & \tilde{X}'_2 \tilde{X}_2 & \dots & \tilde{X}'_2 \tilde{X}_m \\ \vdots & \vdots & \ddots & \vdots \\ \tilde{X}'_m \tilde{X}_1 & \tilde{X}'_m \tilde{X}_2 & \dots & \tilde{X}'_m \tilde{X}_m \end{matrix} \right)^{-1} \begin{pmatrix} \tilde{X}'_1 y \\ \tilde{X}'_2 y \\ \vdots \\ \tilde{X}'_m y \end{pmatrix} \otimes I_p \right) y. \tag{2.3}$$

Hence, a grid search over  $(\beta, \gamma)$  is suggested, as follows, and  $\hat{\lambda}$  is then calculated by (2.3).

1. For fixed  $(\beta, \gamma)$ , calculate  $\hat{\lambda}(\beta, \gamma)$  based on (2.3). Then plug  $\beta, \gamma$  and  $\hat{\lambda}(\beta, \gamma)$  into  $S_n(\theta)$  and denote  $S_n(\beta, \gamma, \hat{\lambda}(\beta, \gamma))$  by  $S_n^c(\beta, \gamma)$ .
2. For a fixed  $\beta$ ,  $S_n^c(\beta, \gamma)$  takes at most  $(n - l)! / [(m - 1)!(n - m - l + 1)!]$  possible values with different  $\gamma$ , so we can find its minimum by enumeration and denote it by  $S_n^{cc}(\beta)$ . Once  $\beta$  is fixed,  $z_t(\beta)$ ,  $t = l, \dots, n - 1$ , are observed and can be arranged in an increasing order  $z_{(1)}(\beta), \dots, z_{(n-l)}(\beta)$ . As  $S_n^c(\beta, \gamma)$  is a constant over the  $(m - 1)$ -dimensional cube  $[z_{(j_1)}(\beta), z_{(j_1+1)}(\beta)] \times \dots \times [z_{(j_{m-1})}(\beta), z_{(j_{m-1}+1)}(\beta)]$  of  $\gamma = (\gamma_1, \gamma_2, \dots, \gamma_{m-1})'$ , it is sufficient to choose  $\gamma_j$  from  $z_{(i)}(\beta)$ ,  $i = 1, \dots, n - l$  and  $j = 1, \dots, m - 1$  in order to get all possible values of  $S_n^c(\beta, \gamma)$ . If we further require at least one data point for each regime, then  $\gamma_j < \gamma_{j+1}$ , for  $j = 1, \dots, m - 1$ . In this way, the number of different values of  $S_n^c(\beta, \gamma)$  is  $(n - l)! / [(m - 1)!(n - m - l + 1)!]$ .

3. The minimizer of  $S_n^{cc}(\beta)$  is obtained within the grids set over  $R^{p-1}$ . This could be time consuming when  $p$  increases; to enhance the efficiency,  $\beta$  can be searched around a preliminary estimate.

The preliminary estimate of  $\beta$  is always chosen to be Johansen's MLE based on a misspecified linear cointegration model. This estimate performs well in such a grid search, as reported in Hansen and Seo (2002), Seo (2011), and the simulation studies reported in Section 4.

**Remark 1.** When the sample size or the number of thresholds is large, the grid search algorithm for estimating multiple threshold parameters in the cointegrating systems carries a high computational burden. Two procedures that potentially yield more computationally efficient solutions are the sequential testing approach and the genetic algorithm. The sequential testing procedure applies the test to a single threshold case in Hansen and Seo (2002), repeatedly so that multiple thresholds can be detected one by one. Specifically, a single threshold test is conducted for a given time series. If a threshold is found, then one proceeds by testing for the presence of an additional threshold in each regime, until no threshold is detected in all the regimes. Since repeated testings are conducted, bias can be introduced and further investigation is required to establish the consistency of this procedure. Interestingly, the idea of the supremum  $F$ -type test of Bai and Perron (1998) for sequential testing of structural-change could also be considered to derive a modified test statistic, which accounts for the sequential nature.

On the other hand, the genetic algorithm is a stochastic search heuristic that mimics the process of natural selection. It seeks an approximate solution to minimizing  $S_n(\theta)$  with respect to the  $\theta$ . This procedure has shown some promises in problems such as multiple change-point detection of piece-wise stationary time series (Davis, Lee, and Rodriguez-Yam (2006)) and the multiple thresholds estimation of threshold autoregressive models (Yau, Tang, and Lee (2015)). As such, we expect the genetic algorithm to be useful in the current multiple thresholds cointegration setting. This computational issue will be dealt with in a future project.

## 2.2. Convergence rate

To establish consistency of  $\hat{\theta}$ , we need the following assumptions.

### Assumption 1.

- 1.1.  $\{u_t\}$  is an i.i.d. sequence of random vectors with  $Eu_t = 0$ , and  $Eu_t u_t' = \Sigma$  is positive definite.

- 1.2.  $\{\Delta x_t, z_t\}$  is a sequence of strictly stationary and strongly mixing random vectors with mixing numbers  $\alpha_H, H = 1, 2, \dots$ , that satisfy for some  $\alpha_0 > 1$ ,  $\alpha_H = o(H^{-(\alpha_0+1)/(\alpha_0-1)})$  as  $H \rightarrow \infty$ , and for some  $\varepsilon > 0$ ,  $E\|X_t X_t'\|^{\alpha_0+\varepsilon} < \infty$  and  $E\|X_{t-1} u_t'\|^{\alpha_0+\varepsilon} < \infty$ , where  $\|A\| = \text{tr}(A'A)^{1/2}$ . Furthermore,  $E\Delta x_t = 0$ , and the partial sum process,  $x_{[ns]}/\sqrt{n}, s \in [0, 1]$ , converges weakly to a vector of Brownian motions  $\mathbf{B}$  with a covariance matrix  $\mathbf{\Omega}$ , that is the long-run covariance matrix of  $\Delta x_t$  and has rank  $p - 1$  such that  $(1, \beta^{0'})\mathbf{\Omega} = 0$ . In particular, assume that  $x_{2,[ns]}/\sqrt{n}$  converges weakly to a vector of Brownian motions  $B$  with a covariance matrix  $\Omega$ , which is finite and positive definite.
- 1.3. The parameter space  $\Theta$  is compact with  $\min_{1 \leq i < j \leq m-1} \{|\gamma_i - \gamma_j|\}$  and  $\max\{|\lambda_1^z|, |\lambda_m^z|\}$  bounded away from zero, where  $|\lambda_j^z|$  is the Euclidean norm of the vector  $\lambda_j^z$ .
- 1.4. Let  $u_t(\xi, \gamma, \lambda)$  be the error  $u_t$  when  $z_{t-1}(\beta)$  is replaced by  $z_{t-1} + \xi$ , where  $\xi$  belongs to a compact set in  $R$ , and let  $S(\xi, \gamma, \lambda) = E(u_t'(\xi, \gamma, \lambda) u_t(\xi, \gamma, \lambda))$ . Assume that  $1/n \sum_t u_t'(\xi, \gamma, \lambda) u_t(\xi, \gamma, \lambda) \xrightarrow{P} S(\xi, \gamma, \lambda)$  uniformly in  $(\xi, \gamma, \lambda)$  on any compact set and  $S(\xi, \gamma, \lambda)$  is continuous in all its arguments and is uniquely minimized at  $(\xi, \gamma, \lambda) = (0, \gamma^0, \lambda^0)$ .

Assumptions 1.1, 1.2 and 1.4 are the same as those in Seo (2011). They are standard assumptions in time series and explanations are given in Seo (2011).

Assumption 1.3 is about the parameter space, especially on  $\lambda_j^z$  and  $\gamma$ . This assumption excludes the possibility of a reduced model with less than  $m$  regimes by requiring  $\gamma_j < \gamma_{j+1}, j = 1, 2, \dots, m-2$ . Additionally, it requires the coefficient of  $z_{t-1}(\beta)$  be bounded away from zero in the outer regimes, which means that  $z_{t-1}(\beta)$  should be included in the error correction model in at least one of the two outer regimes. This assumption simplifies the analysis of the consistency of the LSE. For the LSE of a two-regime TVECM, Seo (2011) made a similar assumption requiring  $(\lambda_1^z, \lambda_2^z)$  to be bounded away from zero, which means, at least one of the two regimes has  $z_{t-1}(\beta)$  included in the error correction model.

With Assumptions 1, we can establish the consistency of  $\hat{\theta}$ . To obtain the super-consistency of  $\hat{\beta}$  and  $\hat{\gamma}$ , we need another assumption.

### Assumption 2.

- 2.1. The probability distribution of  $\{z_t\}$  has a density with respect to Lebesgue measure that is continuous, bounded, and everywhere positive, and the density function  $f(z_t|x_{2,t})$  is bounded by  $K > 0$  for almost every  $x_{2,t}, t = 1, 2, \dots, n$ .
- 2.2. There exist nonrandom vectors  $W_j^* = (1, w_{j,1}, w_{j,2}, \dots, w_{j,l})'$  with  $w_{j,1} = \gamma_j^0$ , such that  $(A_j^0 - A_{j+1}^0)'W_j^* \neq \mathbf{0}_p, \forall j = 1, 2, \dots, m-1$ .

Assumption 2 is common in estimation of threshold model, as used in Chan (1993), Gonzalo and Wolf (2005), and Li and Ling (2012). Assumption 2.1 is for identification of thresholds, as used in Seo (2011). Assumption 2.2 implies that the regression model is discontinuous at each threshold point. These assumptions imply that with positive probability,  $X'_{t-1}(A_j^0 - A_{j+1}^0)(A_j^0 - A_{j+1}^0)'X_{t-1}$  is bigger than a positive constant. It is found from the proof of convergence rates of  $\hat{\gamma}$  and  $\hat{\beta}$  that the discontinuity of the regression model at threshold  $\gamma_j^0$  is crucial for the super consistency of  $\hat{\gamma}_j$ ,  $j = 1, \dots, m - 1$ .

**Theorem 1.** *Under Assumptions 1 and 2,  $n^{3/2}(\hat{\beta} - \beta^0)$  is  $O_p(1)$ ,  $n(\hat{\gamma} - \gamma^0)$  is  $O_p(1)$  and*

$$\sqrt{n}(\hat{\lambda}^* - \lambda^0) \xrightarrow{d} \mathcal{N} \left( \mathbf{0}, \left[ E \left( \begin{pmatrix} I_1 & 0 & \cdots & 0 \\ 0 & I_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & I_m \end{pmatrix} \otimes X_{t-1} X'_{t-1} \right) \right]^{-1} \otimes \Sigma \right),$$

where  $I_j = 1\{\gamma_{j-1}^0 \leq z_{t-1}(\beta^0) < \gamma_j^0\}$  and  $\xrightarrow{d}$  stands for convergence in distribution.

For a multiple-regime TAR model that does not involve cointegration, Li and Ling (2012) found that the LSE of threshold converges at rate  $n$  and the LSE of the slope parameter is asymptotically normal. Gonzalo and Pitarakis (2002) considered estimating the multiple thresholds of a multiple-regime TAR model in a sequential way and established the  $n$ -convergence rate of the sequential estimates. Theorem 1 shows that for the multiple-regime TVECM, the convergence rate of the LSE of thresholds is still  $n$ , although the model is much more complicated under the presence of cointegration. As pointed out by Seo (2011), the  $n^{3/2}$  convergence rate is surprising, but reasonable; since the estimated threshold variable  $z_{t-1}(\hat{\beta}) = z_{t-1} + (x'_{2,t-1}/\sqrt{n})\sqrt{n}(\hat{\beta} - \beta^0)$  and  $\sup_{1 \leq t \leq n} |x_{t-1}| = O_p(n^{1/2})$ , it is expected that  $\hat{\gamma}$  and  $\sqrt{n}\hat{\beta}$  should converge at the same rate.

The limiting distribution of  $(\hat{\beta}', \hat{\gamma}')$  is difficult to establish because the objective function of LSE is highly irregular: the threshold variable  $z_t(\beta)$  needs to be estimated and its estimate is nonstationary if  $\hat{\beta}$  does not equal  $\beta^0$ . Besides, the objective function is discontinuous with respect to the cointegrating vector and the thresholds. The nonstationarity is an intrinsic difficulty, but the discontinuity can be removed with some smoothing techniques, in which case we can apply a Taylor expansion to derive the limiting distribution of the estimator.

### 3. Limiting Distribution of Smoothed LSE

#### 3.1. Limiting distribution

The smoothed LSE (SLSE) was first proposed by Seo and Linton (2007), and subsequently used for the one threshold cointegration (Seo (2011)). The key point of SLSE is to approximate the indicator function  $1\{z_{t-1}(\beta) > \gamma_1\}$  by a smooth function  $\mathcal{K}_{t-1,h}(\beta, \gamma_1) = \mathcal{K}((z_{t-1}(\beta) - \gamma_1)/h)$ , where  $h = h(n)$  is the bandwidth that approaches 0 when  $n \rightarrow \infty$  and  $\lim_{n \rightarrow \infty} \mathcal{K}_{t-1,h}(\beta, \gamma_1) = 1\{z_{t-1}(\beta) > \gamma_1\}$ . Herein  $\mathcal{K}$  is a smooth and bounded function satisfying  $\lim_{s \rightarrow -\infty} \mathcal{K}(s) = 0$  and  $\lim_{s \rightarrow +\infty} \mathcal{K}(s) = 1$ .

To extend the SLSE to multiple-threshold cointegration, we propose to approximate the indicator function  $1\{\gamma_{j-1} \leq z_{t-1}(\beta) < \gamma_j\}$  by  $\mathcal{K}_{t-1,h}(\beta, \gamma_{j-1}, \gamma_j) = \mathcal{K}((z_{t-1}(\beta) - \gamma_{j-1})/h) + \mathcal{K}((\gamma_j - z_{t-1}(\beta))/h) - 1$ . It is immediate that  $\lim_{n \rightarrow \infty} \mathcal{K}_{t-1,h}(\beta, \gamma_{j-1}, \gamma_j) = 1\{\gamma_{j-1} \leq z_{t-1}(\beta) < \gamma_j\}$ . In this way, Model (2.2) is smoothed to

$$y = [(X_1^*(\beta, \gamma), X_2^*(\beta, \gamma), \dots, X_m^*(\beta, \gamma)) \otimes I_p] \lambda + u^*, \quad (3.1)$$

where  $X_j^*(\beta, \gamma)$  is the matrix stacking  $X'_{t-1}(\beta) \mathcal{K}_{t-1,h}(\beta, \gamma_{j-1}, \gamma_j)$  and the rest are defined in the same way as at (2.2).

Let  $u^*(\theta) = y - [(X_1^*(\beta, \gamma), X_2^*(\beta, \gamma), \dots, X_m^*(\beta, \gamma)) \otimes I_p] \lambda$ , and  $S_n^*(\theta) = u^{*\prime}(\theta) u^*(\theta)$ . Since  $S_n^*(\theta)$  converges in probability to the limit of  $S_n(\theta)$  as  $n \rightarrow \infty$ , their minimizers will be close to each other. Hence, we define the SLSE of Model (2.2) as the LSE of Model (3.1):  $\hat{\theta}^* = \operatorname{argmin}_{\theta \in \Theta} S_n^*(\theta)$ . We first study consistency of  $\hat{\theta}^*$ , which requires an assumption.

#### Assumption 3.

- 3.1.  $\mathcal{K}$  is twice differentiable everywhere,  $\mathcal{K}^{(1)}$  is symmetric around zero, and  $\mathcal{K}^{(1)}$  and  $\mathcal{K}^{(2)}$  are uniformly bounded and uniformly continuous, where  $g^{(i)}$  indicates the  $i$ th derivative of  $g$ . Further,  $\int |\mathcal{K}^{(1)}(s)|^4 ds$ ,  $\int |s^2 \mathcal{K}^{(1)}(s)| ds$ ,  $\int |\mathcal{K}^{(2)}(s)|^2 ds$ , and  $\int |s^2 \mathcal{K}^{(2)}(s)| ds$  are finite.
- 3.2.  $\mathcal{K}(s) - \mathcal{K}(0)$  has the same sign as  $s$ , and for some integer  $\nu \geq 2$  and each integer  $i$  ( $1 \leq i \leq \nu$ ),  $\int |s^i \mathcal{K}^{(1)}(s)| ds < \infty$ ,  $\int s^{i-1} \operatorname{sgn}(s) \mathcal{K}^{(1)}(s) ds = 0$ , and  $\int s^{\nu-1} \operatorname{sgn}(s) \mathcal{K}^{(1)}(s) ds \neq 0$ .
- 3.3. For some  $\epsilon > 0$ ,
 
$$\lim_{n \rightarrow \infty} h^{i-\nu} \int_{|hs| > \epsilon} |s^i \mathcal{K}^{(1)}(s)| ds = 0 \text{ and } \lim_{n \rightarrow \infty} h^{-1} \int_{|hs| > \epsilon} |s^i \mathcal{K}^{(2)}(s)| ds = 0.$$
- 3.4.  $h$  is a function of  $n$  and satisfies that for some sequence  $q \geq 1$ ,  $nh^3 \rightarrow 0$ ,  $\log(nq)(n^{1-6/r} h^2 q^{-2})^{-1} \rightarrow 0$ , and  $h^{-9k/2-3} n^{(9k/2+2)/r+\epsilon} \alpha_q \rightarrow 0$ , where  $k$  is the dimension of  $\theta$  and  $r > 4$  is specified in Assumption 4.



Assumption 3 is imposed in Seo and Linton (2007) for the consistency of SLSE of a two-regime threshold regression model and is common in the smoothed estimation method, see Horowitz (1992). When we extend the SLSE to the multiple-threshold regression model, we need to deal with the interaction of adjacent regimes like  $X'_{t-1}(\beta)\mathcal{K}_{t-1,h}(\beta, \gamma_{j-1}, \gamma_j)X_{t-1}(\beta)\mathcal{K}_{t-1,h}(\beta, \gamma_j, \gamma_{j+1})$ . It turns out that under Assumption 3, those interaction terms are negligible when establishing consistency of SLSE. One example of  $\mathcal{K}$  is constructed based on normal distribution and  $h$  can be chosen to be  $n^{-1/2}\log(n) * C$ , where  $C$  is a constant. Further details can be found in Section 4.

**Theorem 2.** *Under Assumptions 1 and 3,  $\hat{\theta}^* - \theta^0$  is  $o_p(1)$ , and  $\sqrt{n}(\hat{\beta}^* - \beta^0)$  is  $o_p(1)$ .*

With  $\mathcal{K}$  satisfying Assumptions 3, the argument in proof of Lemma 5 in Appendix holds if we replace the indicator function by  $\mathcal{K}$ . Therefore, Theorem 2 can be proved. Since the objective function of the SLSE is smooth, we can apply a Taylor expansion to derive the limiting distribution, which requires an additional assumption.

**Assumption 4.**

- 4.1.  $E\|X_t u'_t\|^r < \infty$ ,  $E\|X_t X'_t\|^r < \infty$ , for some  $r > 4$ .
- 4.2.  $\{(\Delta x'_t, z_t)'\}$  is a sequence of strictly stationary and strongly mixing random vectors with mixing coefficients  $\alpha_H$ ,  $H = 1, 2, \dots$ , satisfying for positive  $C$  and  $\eta$ ,  $\alpha_H \leq CH^{-(2r-2)/(r-2)-\eta}$  as  $H \rightarrow \infty$ .
- 4.3. Let  $\Delta_{t,t-l+1} = (\Delta x_t, \Delta x_{t-1}, \dots, \Delta x_{t-l+1})$  and  $f(z_t|\Delta_{t,t-l+1})$  be the conditional density function. For some integer  $\nu \geq 2$  and each integer  $i$  such that  $1 \leq i \leq \nu - 1$ , for all  $z$  in a neighborhood of threshold  $\gamma_j^0$ ,  $j = 1, \dots, m - 1$ , for almost every  $\Delta_{t,t-l+1} = \Delta$  and some  $K < \infty$ ,  $f^{(i)}(z|\Delta)$  exists and is a continuous function of  $z$  satisfying  $|f^{(i)}(z|\Delta)| < K$ . In addition,  $f(z|\Delta) < K$  for all  $z$  and almost every  $\Delta$ .
- 4.4. The conditional joint density  $f(z_t, z_{t-H}|\Delta_{t,t-l+1}, \Delta_{t-H,t-l-H+1}) < K$ , for all  $(z_t, z_{t-H})$  and almost every  $(\Delta_{t,t-l+1}, \Delta_{t-H,t-l-H+1})$ .
- 4.5.  $\theta^0$  is an interior point of  $\Theta$ .

Assumption 4 is analogous to assumptions imposed in Seo and Linton (2007), Seo (2011), and Horowitz (1992). These assumptions are common in standard smooth estimation methods, and are interpreted in Seo (2011).

**Theorem 3.** *Under Assumptions 1, 2, 3, and 4,*

$$\begin{pmatrix} nh^{-1/2}(\hat{\beta}^* - \beta^0) \\ \sqrt{nh^{-1}}(\hat{\gamma}^* - \gamma^0) \end{pmatrix} \xrightarrow{d} \begin{pmatrix} \sigma_q^2 \int_0^1 BB' & \sigma_{q_1}^2 \int_0^1 B & \sigma_{q_2}^2 \int_0^1 B & \dots & \sigma_{q_{m-1}}^2 \int_0^1 B \\ \sigma_{q_1}^2 \int_0^1 B' & \sigma_{q_1}^2 & 0 & \dots & 0 \\ \sigma_{q_2}^2 \int_0^1 B' & 0 & \sigma_{q_2}^2 & \dots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \sigma_{q_{m-1}}^2 \int_0^1 B' & 0 & 0 & \dots & \sigma_{q_{m-1}}^2 \end{pmatrix}^{-1} \begin{pmatrix} -\int_0^1 B d \sum_{j=1}^{m-1} \sigma_{v_j} W_j \\ \sigma_{v_1} W_1(1) \\ \sigma_{v_2} W_2(1) \\ \vdots \\ \sigma_{v_{m-1}} W_{m-1}(1) \end{pmatrix},$$

$$\sqrt{n}(\hat{\lambda}^* - \lambda^0) \xrightarrow{d} \mathcal{N} \left( \mathbf{0}, \left[ E \left( \begin{pmatrix} I_1 & 0 & \dots & 0 \\ 0 & I_2 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & I_m \end{pmatrix} \otimes X_{t-1} X'_{t-1} \right) \right]^{-1} \otimes \Sigma \right),$$

and  $\begin{pmatrix} nh^{-1/2}(\hat{\beta}^* - \beta^0) \\ \sqrt{nh^{-1}}(\hat{\gamma}^* - \gamma^0) \end{pmatrix}$  and  $\sqrt{n}(\hat{\lambda}^* - \lambda^0)$  are asymptotically independent.

Notations of Theorem 3 are given below. In view of Jensen’s inequality,  $\int_0^1 BB' - \int_0^1 B \int_0^1 B'$  is positive definite with probability 1. It follows that the inverse matrix appearing in the joint limiting distribution of  $\hat{\beta}^*$  and  $\hat{\gamma}^*$  exists with probability 1. The normal distribution to which  $\sqrt{n}(\hat{\lambda}^* - \lambda^0)$  converges is the same as that of  $\sqrt{n}(\hat{\lambda} - \lambda^0)$ . Both are the same as when  $\beta^0$  and  $\gamma^0$  are known.

1.  $B$  is the vector of Brownian motions defined in Assumption 1.1, and  $W_1, \dots, W_{m-1}$  are mutually independent standard Brownian motions that are independent of  $B$ .  $\int_0^1 BB'$  stands for  $\int_0^1 B(t)B'(t)dt$ .
2.  $I_j$  are the same as defined in Theorem 1.
3.  $\tilde{\mathcal{K}}_1(s) = \mathcal{K}^{(1)}(s)(1\{s > 0\} - K(s))$ . For  $j = 1, \dots, m-1$ ,  $\sigma_{v_j}^2 = E \left[ F_j | z_{t-1} = \gamma_j^0 \right] f_Z(\gamma_j^0)$ , where  $F_j = \|\mathcal{K}^{(1)}\|_2^2 (X'_{t-1}(A_j^0 - A_{j+1}^0)u_t)^2 + \|\tilde{\mathcal{K}}_1\|_2^2 (X'_{t-1}(A_j^0 - A_{j+1}^0)(A_j^0 - A_{j+1}^0)'X_{t-1})^2$ , with  $\|g\|_2 = (\int_{-\infty}^{\infty} g(x)^2 dx)^{1/2}$ .
4. For  $j = 1, \dots, m-1$ ,  $\sigma_{q_j}^2 = \mathcal{K}^{(1)}(0)E[X'_{t-1}(A_j^0 - A_{j+1}^0)(A_j^0 - A_{j+1}^0)'X_{t-1}]|z_{t-1} = \gamma_j^0] f_Z(\gamma_j^0)$  and  $\sigma_q^2 = \sum_{j=1}^{m-1} \sigma_{q_j}^2$ .

By virtue of Theorem 3 and the definitions of  $\sigma_{q_j}^2$ ,  $\sigma_q^2$  and  $\sigma_{v_j}^2$ , it is found that the asymptotic variance of  $\hat{\beta}^*$  is determined by  $\mathcal{K}$ ,  $h$ , discontinuities of Model (2.2) and densities of  $z_t$  at the  $m - 1$  thresholds. By simple calculations based

on Theorem 3, the limiting distribution of  $\hat{\beta}^*$  is obtained as

$$nh^{-1/2}(\hat{\beta}^* - \beta^0) \xrightarrow{d} \frac{1}{\sigma_q^2} \left( \int_0^1 BB' - \int_0^1 B \int_0^1 B' \right)^{-1} \left( \sum_{j=1}^{m-1} \sigma_{v_j} [W_j(1) \int B - \int BdW_j] \right).$$

Theorem 3 shows that the SLSEs of the multiple thresholds are not asymptotically independent. The reason is that the convergence rate of  $\hat{\beta}^*$  is not fast enough to make the estimation error negligible. The following Corollary states that when  $\beta^0$  is known, the smoothed least squares estimators of the  $m - 1$  thresholds are asymptotically independent. Therefore, making use of the LSE  $\hat{\beta}$  that converges faster than  $\hat{\beta}^*$ , an alternative estimation method for the thresholds can be developed. First, we obtain  $\hat{\beta}$ , then plug it into the TVECM as if it were true and estimate the thresholds by the SLSE, which is denoted as  $\hat{\gamma}^*(\hat{\beta})$ . The following corollary shows that  $\hat{\gamma}^*(\hat{\beta})$  has the same limiting distribution as if  $\beta^0$  were known.

**Corollary 1.** *Suppose that Assumptions 1, 2, 3 and 4 hold. Let  $\hat{\gamma}^*(\beta)$  be the smoothed LSE of  $\gamma$  when  $\beta$  is given. Then,  $\sqrt{nh^{-1}}(\hat{\gamma}^*(\hat{\beta}) - \gamma^0)$  has the same asymptotic distribution as that of  $\sqrt{nh^{-1}}(\hat{\gamma}^*(\beta^0) - \gamma^0)$ , and  $\sqrt{nh^{-1}}(\hat{\gamma}^*(\beta^0) - \gamma^0) \xrightarrow{d} N(0, V)$  with  $V = \text{diag}(\sigma_{v_1}^2/\sigma_{q_1}^4, \dots, \sigma_{v_{m-1}}^2/\sigma_{q_{m-1}}^4)$ , where  $\sigma_{v_j}^2$  and  $\sigma_{q_j}^2$  are the same as in Theorem 3.*

This corollary is useful for constructing confidence intervals for  $\gamma$ . Estimation of the asymptotic variance of  $\hat{\gamma}^*(\hat{\beta})$  is much easier than the case based on the result given in Theorem 3, since the vector of Brownian motions  $B$  is not involved. Further, estimates of the  $m - 1$  thresholds are asymptotically independent, hence we can construct confidence intervals for each of them individually, as will be discussed in Section 3.3. This corollary can be proved by similar arguments used in Corollary 2 of Seo (2011).

### 3.2. Asymptotic variance

Given the asymptotic normality of the SLSE, we have to estimate the variances of the limiting distributions for inference. Since  $\hat{\lambda}^*$  is asymptotically independent of  $(\hat{\beta}^*, \hat{\gamma}^*)$ , we can estimate its asymptotic variance first; this involves the expectation of  $1\{\gamma_{j-1}^0 \leq z_{t-1}(\beta^0) < \gamma_j^0\}, j = 1, \dots, m - 1$ . Since  $\hat{\beta}^*$  and  $\hat{\gamma}^*$  are super-consistent, we can input them into the indicator function as if they were the true values. Then  $\hat{\lambda}^*$  is the ordinary LSE and estimation of its asymptotic variance is standard.

Inference for  $\beta$  and  $\gamma$  is more complicated. It involves estimating  $\Omega, \sigma_{v_j}^2$ , and  $\sigma_{q_j}^2, j = 1, 2, \dots, m - 1$ ,  $\Omega$  being the variance matrix of the multiple-dimensional

Brownian motions to which  $x_{2,[ns]}/\sqrt{n}$ ,  $s \in [0, 1]$  converges. Since  $\Delta x_t$  is a stationary  $p$ -dimensional time series, estimation of  $\Omega$  is standard, see Andrews (1991). The variances  $\sigma_{v_j}^2$  and  $\sigma_{q_j}^2$  that appear in joint limiting distribution of  $(\hat{\beta}^*, \hat{\gamma}^*)$  are limits of certain elements of the score and Hessian matrix of  $S_n^*(\theta)$ . We estimate them based on a Taylor expansion of  $S_n^*(\theta)$ .

With  $\hat{\theta}^* = \operatorname{argmin}_{\theta \in \Theta} S_n^*(\theta)$  and a Taylor expansion, we have:

$$\sqrt{n}(D_n Q_n(\theta)|_{\tilde{\theta}} D_n) D_n^{-1}(\hat{\theta}^* - \theta^0) = -\sqrt{n} D_n T_n(\theta^0),$$

where  $\tilde{\theta}$  lies between  $\theta^0$  and  $\hat{\theta}^*$ ,  $Q_n(\theta) = \partial^2 S_n^*(\theta)/2n\partial\theta\partial\theta'$ ,  $T_n = \partial S_n^*(\theta)/2n\partial\theta$ , and  $D_n = \operatorname{diag}(\sqrt{h/n}\mathbf{1}'_{p-1}, \sqrt{h}\mathbf{1}'_{m-1}, \mathbf{1}'_{k_\lambda})$ , with  $k_\lambda$  the length of  $\lambda$  and  $\mathbf{1}_k$  the vector of length  $k$  and all elements being 1.

Through the convergence of  $D_n Q_n(\theta) D_n$  in a neighborhood of  $\theta^0$ , we find suitable quantities to estimate  $\sigma_{q_j}^2$ . From the proof of Theorem 3, we know that

$$D_n Q_n(\hat{\theta}^*) D_n \xrightarrow{d} \begin{pmatrix} \sigma_q^2 \int_0^1 B(s)B(s)' ds & \sigma_{q_1}^2 \int_0^1 B(s) ds & \cdots & \sigma_{q_{m-1}}^2 \int_0^1 B(s) ds & 0 \\ \sigma_{q_1}^2 \int_0^1 B(s)' ds & \sigma_{q_1}^2 & \cdots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ \sigma_{q_{m-1}}^2 \int_0^1 B(s)' ds & 0 & \cdots & \sigma_{q_{m-1}}^2 & 0 \\ \mathbf{0} & 0 & \cdots & 0 & N \end{pmatrix},$$

where  $N$  is the asymptotic covariance matrix of  $\sqrt{n}(\hat{\lambda}^* - \lambda^0)$ .

As a result, we propose to estimate  $\sigma_{q_j}^2$  by  $(\hat{\sigma}_{q_j}^*)^2 = (D_n Q_n(\hat{\theta}^*) D_n)_{p-1+j, p-1+j}$ , the  $(p-1+j)$ th diagonal element of  $D_n Q_n(\hat{\theta}^*) D_n$ , for  $j = 1, 2, \dots, m-1$ . Then the proof of Theorem 3 ensures the consistency of  $\sigma_{q_j}^2$ .

From the proof of the convergence of  $\sqrt{n} D_n T_n(\theta^0)$ , we propose an estimation method for  $\sigma_{v_j}^2$ . Since  $\sqrt{n} D_n T_n(\theta^0)$  converges in distribution to

$$\left( -\int B' d \sum_{j=1}^{m-1} \sigma_{v_j} W_j, \sigma_{v_1} W_1(1), \sigma_{v_2} W_2(1), \dots, \sigma_{v_{m-1}} W_{m-1}(1), 0'_{k_\lambda} \right)',$$

$\sigma_{v_j}^2$  is the asymptotic variance of  $\sqrt{nh}[\partial S_n^*(\theta^0)/2n\partial\gamma_j]$ . Let  $\phi_{n,t}^{\gamma_j} = [\partial u_t^{*'}(\theta^0)/\partial\gamma_j] u_t^{*'}(\theta^0)$ , then it is shown in the proof of Theorem 3 that  $\lim_{n \rightarrow \infty} \sqrt{n} E(\sqrt{h}\phi_{n,t}^{\gamma_j}) = 0$  and

$$\lim_{n \rightarrow \infty} \operatorname{Var} \left( \sqrt{nh} \frac{\partial S_n^*(\theta^0)}{2n\partial\gamma_j} \right) = \lim_{n \rightarrow \infty} \operatorname{Var} \left( \sqrt{h}\phi_{n,t}^{\gamma_j} \right).$$

As a result, we estimate  $\sigma_{v_j}^2$  by  $(\hat{\sigma}_{v_j}^*)^2 = (1/n) \sum_{t=1}^n \left( \sqrt{h}\phi_{n,t}^{\gamma_j^*} \right)^2$ , where  $\phi_{n,t}^{\gamma_j^*} = [\partial u_t^{*'}(\hat{\theta}^*)/\partial\gamma_j] u_t^{*'}(\hat{\theta}^*) = X'_{t-1}(\hat{\beta}^*)(\hat{A}_{j+1}^* - \hat{A}_j^*) \{[\mathcal{K}^{(1)}((\hat{\gamma}_j^* - z_{t-1}(\hat{\beta}^*))/h)]/h\} \hat{u}_t^*(\hat{\theta}^*)$ .

The consistency of  $(\hat{\sigma}_{v_j}^*)^2$  can be proved by an argument similar to that of the proof of Theorem 4 of Seo and Linton (2007). With  $(\hat{\sigma}_{v_j}^*)^2$ ,  $(\hat{\sigma}_{q_j}^*)^2$ ,  $j =$

1, 2, ...,  $m - 1$ , and  $B$  being simulated based on estimates of  $\Omega$ , we can estimate the variance of the joint limiting distribution of  $(\hat{\beta}^*, \hat{\gamma}^*)$  and construct confidence intervals for  $\beta$  and  $\gamma$  jointly.

If we are only interested in inference for  $\gamma$ , its confidence intervals can be constructed by the alternative estimation method based on Corollary 1. We estimate  $\hat{\beta}$  by the LSE and plug it into the TVECM as if it were the true value, then the SLSE of the multiple thresholds converge to independent normal variables. By Corollary 1, this two-step estimator  $\hat{\gamma}^*(\hat{\beta})$  has an asymptotic variance matrix that does not involve  $B$ . Therefore, we only need to estimate  $\sigma_{v_j}^2$  and  $\sigma_{q_j}^2$  with aforementioned methods, without worrying about  $\Omega$ .

#### 4. Simulation and Empirical Studies

##### 4.1. Simulation study

We focused on the LSE and the SLSE of multiple-threshold cointegration, especially estimates of the cointegrating vector and thresholds, and assessed their performance through a simulation study.

The model was

$$\begin{aligned} \Delta x_t = & \begin{pmatrix} -1 \\ 0 \end{pmatrix} z_{t-1} + \begin{pmatrix} 0 \\ 0.3 \end{pmatrix} z_{t-1} 1\{z_{t-1} > \gamma_2^0\} \\ & + \begin{pmatrix} 0 \\ 0.8 \end{pmatrix} z_{t-1} 1\{z_{t-1} < \gamma_1^0\} + u_t, \end{aligned} \quad (4.1)$$

where  $z_{t-1} = x_{1,t-1} + \beta^0 x_{2,t-1}$ ,  $\beta^0 = -2$ ,  $\gamma_1^0 < \gamma_2^0$ ,  $u_t \sim \text{i.i.d } N(0, I_2)$  for  $t = l + 1, \dots, n$ , and  $\Delta x_0 = u_0$ .

This model is a simple extension of the two-regime TVECM of Hansen and Seo (2002) and Seo (2011). We estimated (4.1) based on a three-regime TVECM that contains  $\Delta x_{t-1}$  and  $\Delta x_{t-2}$ , although the true model contains no lagged terms of  $\Delta x_t$ . This should be reasonable since we are only interested in estimators of the cointegrating vector and thresholds.

We considered two cases: experiments with sample sizes  $n = 100$  and  $n = 250$  were performed for case I,  $\gamma = (-1, 1)$  and one experiment with a sample size  $n = 250$  was performed for case II,  $\gamma = (-3, 3)$ . Each experiment had 800 simulation replications. For the SLSE, we followed Seo's (2011) suggestion and chose  $\mathcal{K} = \Phi(s) + s\phi(s)$  and  $h = \hat{\sigma}^* n^{-1/2} \log n$ , where  $\Phi$  and  $\phi$  are, respectively, the cumulative distribution function and the probability density function of  $N(0, 1)$ , and  $\hat{\sigma}^*$  is the sample standard deviation of  $z_t(\hat{\beta}^*)$ .

As discussed in Section 3, the LSE and the SLSE are to be searched over grids in  $R^3$  for (4.1), which is time consuming. To reduce computing time, we constructed grids in two steps.

Table 1. Comparison of estimation for different  $n$ , case I.

	mean	sd	MAE in log
$n = 100$			
$\hat{\beta} - \beta^0$	2.34e-05	0.0026	-6.2415
$\hat{\gamma}_1 - \gamma_1^0$	0.4162	1.409	0.1381
$\hat{\gamma}_2 - \gamma_2^0$	0.2049	1.777	0.3913
$\hat{\beta}^* - \beta^0$	5.19e-05	0.0029	-6.1557
$\hat{\gamma}_1^* - \gamma_1^0$	0.4571	1.474	0.1779
$\hat{\gamma}_2^* - \gamma_2^0$	0.2095	1.838	0.4253
$n = 250$			
$\hat{\beta} - \beta^0$	7.8611e-06	0.00054	-7.7729
$\hat{\gamma}_1 - \gamma_1^0$	0.1821	0.9852	-0.2869
$\hat{\gamma}_2 - \gamma_2^0$	-0.0679	1.6558	0.3177
$\hat{\beta}^* - \beta^0$	1.7008e-05	0.00058	-7.7381
$\hat{\gamma}_1^* - \gamma_1^0$	0.1448	1.0029	-0.2587
$\hat{\gamma}_2^* - \gamma_2^0$	-0.0525	1.6951	0.3334

1. Denote Johansen's MLE by  $\tilde{\beta}$ . We used  $\tilde{\beta}$  as a preliminary estimate and set 100 grids around it.
2. For each grid point of  $\beta$ , we calculated  $z_t(\beta)$ ,  $t = 3, \dots, n-1$  and determined the interval  $(\min\{z_t(\beta)\}, \max\{z_t(\beta)\})$ . For sample sizes  $n = 100$  and  $n = 250$ , we set 20 and 50 grids in the interval respectively.

Exact implementation of the grid search is described at the end of Section 2.1.

We present the estimation results of the LSE, SLSE, Johansen's MLE, and other estimators.  $\hat{\gamma}_r^*$  is the restricted SLSE when  $\beta^0$  is known and  $\hat{\gamma}_r$  is the restricted LSE when  $\beta^0$  is known. The LSE and the SLSE are both joint estimators that estimate all parameters simultaneously. We compare them with the sequential estimators which estimate one threshold at a time. Define  $\hat{\gamma}_s$  and  $\hat{\gamma}_s^*$  as the sequential LSE and sequential SLSE respectively, similarly we have  $\hat{\beta}_s$  and  $\hat{\beta}_s^*$ . The performance of different estimators are reported in Tables 1, 2 and 3, in terms of mean, standard deviation (sd) of the biases and mean absolute error (MAE) of the 800 replications.

Table 1 shows that for case I,  $\gamma = (-1, 1)$ , the performances of the LSE and the SLSE of  $\beta$  and  $\gamma$  are much improved when  $n$  increases, in terms of mean, sd and MAE. In Table 2, results of the two estimators of  $\beta$  and  $\gamma$  are reported for  $n = 250$  and different  $\gamma$ s. When the magnitude of thresholds increases, the overall performances of  $\hat{\gamma}$  and  $\hat{\gamma}^*$  decline. One possible explanation to the decline in performance is that when the magnitude of  $\gamma$  increases, most of the data falls in the middle regime as a result of normal sample setting and the TVECM model behaves more like a linear ECM. Therefore, only a small fraction of the data is

Table 2. Comparison of estimation for different  $\gamma$ s.

	mean	sd	MAE in log
case I, $n = 250$ , $\gamma = c(-1, 1)$			
$\hat{\beta} - \beta^0$	7.8611e-06	0.00054	-7.7729
$\hat{\gamma}_1 - \gamma_1^0$	0.1821	0.9852	-0.2869
$\hat{\gamma}_2 - \gamma_2^0$	-0.0679	1.6558	0.3177
$\hat{\beta}^* - \beta^0$	1.7008e-05	0.00058	-7.7381
$\hat{\gamma}_1^* - \gamma_1^0$	0.1448	1.0029	-0.2587
$\hat{\gamma}_2^* - \gamma_2^0$	-0.0525	1.6951	0.3334
case II, $n = 250$ , $\gamma = c(-3, 3)$			
$\hat{\beta} - \beta^0$	-3.0012e-02	0.2433	-3.4921
$\hat{\gamma}_1 - \gamma_1^0$	0.1623	0.5483	-1.3868
$\hat{\gamma}_2 - \gamma_2^0$	-0.6184	1.1678	-0.2426
$\hat{\beta}^* - \beta^0$	-3.0029e-02	0.2433	-3.4920
$\hat{\gamma}_1^* - \gamma_1^0$	0.1566	0.5421	-1.3881
$\hat{\gamma}_2^* - \gamma_2^0$	-0.6757	1.2554	-0.1726

Table 3. Result of different estimators for case I,  $n=250$ .

	mean	sd	MAE in log
$\hat{\beta} - \beta^0$	2.1381e-05	0.00045	-7.9221
$\hat{\beta} - \beta^0$	7.8611e-06	0.00054	-7.7729
$\hat{\beta}^* - \beta^0$	1.7008e-05	0.00058	-7.7381
$\hat{\beta}_s - \beta^0$	2.0909e-05	0.00053	-7.7798
$\hat{\beta}_s^* - \beta^0$	3.9289e-06	0.00052	-7.8185
LSE and SLSE			
$\hat{\gamma}_1 - \gamma_1^0$	1.8213e-01	0.9852	-0.2869
$\hat{\gamma}_2 - \gamma_2^0$	-6.7928e-02	1.6557	0.3177
$\hat{\gamma}_1^* - \gamma_1^0$	1.4489e-01	1.0028	-0.2587
$\hat{\gamma}_2^* - \gamma_2^0$	-5.2532e-02	1.6951	0.3334
(sequential)LSE and SLSE			
$\hat{\gamma}_{s,1} - \gamma_1^0$	4.8364e-02	0.971	-0.3241
$\hat{\gamma}_{s,2} - \gamma_2^0$	6.6848e-01	1.6701	0.3741
$\hat{\gamma}_{s,1}^* - \gamma_1^0$	-4.5039e-02	0.9764	-0.3141
$\hat{\gamma}_{s,2}^* - \gamma_2^0$	6.8553e-01	1.763	0.4305
(restricted) LSE and SLSE			
$\hat{\gamma}_{r,1} - \gamma_1^0$	1.7321e-01	0.993	-0.2475
$\hat{\gamma}_{r,2} - \gamma_2^0$	-6.3433e-02	1.6141	0.2866
$\hat{\gamma}_{r,1}^* - \gamma_1^0$	8.0666e-02	0.9798	-0.2910
$\hat{\gamma}_{r,2}^* - \gamma_2^0$	-3.5652e-02	1.6840	0.3346

used to estimate the threshold parameter, leading to slower convergence of the estimate.

Next we compare the performances of the proposed estimators with other

estimators mentioned in this section, see Table 3. Five estimators of  $\beta$  are reported: Johansen's MLE, LSE, SLSE, sequential LSE, and sequential SLSE, all of which estimate  $\beta$  very accurately. In terms of mean, Johansen's MLE does not perform as well as the other estimators, which are obtained from estimating the correctly specified multiple-threshold cointegration model. The LSE outperforms the SLSE, agreeing with the theoretical result that the LSE converges faster.

We also compare the performances of different estimators of thresholds  $\gamma$ . The LSE  $\hat{\gamma}$  has slightly smaller MAE than the SLSE  $\hat{\gamma}^*$ , supporting the theory of faster convergence rate of  $\hat{\gamma}$ . The joint estimators  $\hat{\gamma}$  and  $\hat{\gamma}^*$  show superiority over the sequential estimators  $\hat{\gamma}_s$  and  $\hat{\gamma}_s^*$ , respectively, in terms of mean. Although the sequential estimators can reduce computing time of grid search, their accuracies are not satisfactory; This was also reported by Lo and Zivot (2001). The restricted estimators  $\hat{\gamma}_r$  and  $\hat{\gamma}_r^*$  outperform the unrestricted estimators  $\hat{\gamma}$  and  $\hat{\gamma}^*$ , respectively, which means knowledge of the true value of  $\beta$  helps to improve the performance of estimators. However, the improvement from  $\hat{\gamma}$  to  $\hat{\gamma}_r$  is limited, the reason of which may be the extremely fast convergence rate  $n^{3/2}$  of  $\hat{\beta}$ .

The simulation study supports the theoretic results obtained in Sections 3 and shows the superiority of the LSE and the SLSE over sequential estimators. With our choice of  $\mathcal{K}$  and  $h$ , the SLSE performs almost as well as the LSE. The SLSE is recommended because of its asymptotic normality.

#### 4.2. Term structure of interest rates

In economics and finance, yields of different maturities appear to move together. This observation inspired many theories and empirical studies; the present value model of stock prices, expectation theory of interest rates, market segmentation theory, and preferred habitat theory. These studies are based on the hypothesis that interest rates of a security with different maturities should not deviate too much from each other.

Let  $r_t$  and  $R_t$  be the interest rates of a one-period and multi-period bonds, respectively. Campbell and Shiller (1987) suggested that the term structure of interest rates implies a linear cointegration relationship between  $r_t$  and  $R_t$  with cointegrating vector  $(1, -1)$ .

The development of threshold cointegration led to many empirical studies that considered threshold cointegration through a TVECM. For detailed discussions, see Hansen and Seo (2002), and Enders and Granger (1998). Hansen and Seo (2002) considered using a two-regime TVECM to describe the dynamics of the long-term and short-term interest rates as,

$$\begin{pmatrix} \Delta R_t \\ \Delta r_t \end{pmatrix} = \begin{cases} \mu_1 + \rho_1 z_{t-1}(\beta) + \Gamma_1 \begin{pmatrix} \Delta R_{t-1} \\ \Delta r_{t-1} \end{pmatrix} + u_t, & z_{t-1} \leq \gamma, \\ \mu_2 + \rho_2 z_{t-1}(\beta) + \Gamma_2 \begin{pmatrix} \Delta R_{t-1} \\ \Delta r_{t-1} \end{pmatrix} + u_t, & z_{t-1} > \gamma. \end{cases} \quad (4.2)$$



Herein  $z_t(\beta) = R_t - \beta r_t$ ,  $\mu_j$  is an intercept and  $\rho_j, \Gamma_j$  are slope parameters in the  $j$ -th regime,  $j = 1, 2$ .

The two regimes stand for different situations: the long-term interest rate  $R_t$  is relatively low or high compared to the short-term interest rate  $r_t$ . Although  $R_t \geq r_t$  is typical,  $R_t < r_t$  was occurred in history. For example, in November 2004, the yield curve for UK Government bonds was partially inverted. The yield for the ten-year bond was 4.68%, but was only 4.45% for the thirty-year bond.

We now take one step further to consider a three-regime TVECM. Data are from the monthly interest rate series of the United States studied in McCulloch and Kwon (1993). This data was used widely by many authors including Campbell (1995), who investigated interest rates of different maturities ranging from one month to 120 months, in the period 1952-1991, and found many pairs of them to be cointegrated. Hansen and Seo (2002) examined the same pairs of interest rates in this period and found some pairs to be threshold cointegrated.

We chose the same pair of the twelve-month and 120-month bonds as Hansen and Seo (2002), who modeled them by a two-regime TVECM with one lagged terms of  $\Delta R_t$  and  $\Delta r_t$ . They estimated the model by the MLE as

$$\begin{aligned} \Delta R_t &= \begin{cases} 0.54 + 0.34z_{t-1} + 0.35\Delta R_{t-1} - 0.17\Delta r_{t-1} + u_{1,t}, & \text{if } z_{t-1} \leq -0.63, \\ 0.01 - 0.02z_{t-1} - 0.08\Delta R_{t-1} + 0.09\Delta r_{t-1} + u_{1,t}, & \text{if } z_{t-1} > -0.63, \end{cases} \\ \Delta r_t &= \begin{cases} 1.45 + 1.41z_{t-1} + 0.92\Delta R_{t-1} - 0.04\Delta r_{t-1} + u_{2,t}, & \text{if } z_{t-1} \leq -0.63, \\ -0.04 + 0.04z_{t-1} - 0.07\Delta R_{t-1} + 0.23\Delta r_{t-1} + u_{2,t}, & \text{if } z_{t-1} > -0.63, \end{cases} \end{aligned}$$

where  $z_{t-1} = R_{t-1} - \hat{\beta}r_{t-1}$  with  $\hat{\beta} = 0.984$ . The estimated cointegrating vector  $(1, -0.984)$  is very close to Campbell and Shiller's (1987) theory:  $(1, -1)$ . The two regimes determined by the threshold estimate  $-0.63$  contained 8% and 92% of the data and are defined as extreme regime and typical regime, respectively.

We model the data by a three-regime TVECM with one lagged term of  $\Delta R_t$  and  $\Delta r_t$ , as an analog of (4.2). The model is estimated by the two proposed estimators. By the LSE, the model was estimated as

$$\begin{aligned} \Delta R_t &= \begin{cases} 0.543 + 0.344z_{t-1} + 0.351\Delta R_{t-1} - 0.176\Delta r_{t-1} + u_{1,t}, & \text{if } z_{t-1} \leq -0.62, \\ 0.005 - 0.019z_{t-1} - 0.046\Delta R_{t-1} + 0.090\Delta r_{t-1} + u_{1,t}, & \text{if } 1.58 \geq z_{t-1} > -0.62, \\ 0.891 - 0.439z_{t-1} - 0.097\Delta R_{t-1} + 0.044\Delta r_{t-1} + u_{1,t}, & \text{if } z_{t-1} > 1.58, \end{cases} \\ \Delta r_t &= \begin{cases} 1.467 + 1.391z_{t-1} + 0.750\Delta R_{t-1} - 0.142\Delta r_{t-1} + u_{2,t}, & \text{if } z_{t-1} \leq -0.62, \\ -0.023 + 0.009z_{t-1} - 0.120\Delta R_{t-1} + 0.195\Delta r_{t-1} + u_{2,t}, & \text{if } 1.58 \geq z_{t-1} > -0.62, \\ 0.794 - 0.350z_{t-1} + 0.162\Delta R_{t-1} + 0.032\Delta r_{t-1} + u_{2,t}, & \text{if } z_{t-1} > 1.58. \end{cases} \end{aligned}$$

For the SLSE,  $\mathcal{K}$  and  $h$  were determined as in the simulation study. The

model was estimated as

$$\Delta R_t = \begin{cases} 0.575 + 0.340z_{t-1} + 0.174\Delta R_{t-1} - 0.147\Delta r_{t-1} + u_{1,t}, & \text{if } z_{t-1} \leq -0.84, \\ 0.025 - 0.051z_{t-1} - 0.050\Delta R_{t-1} - 0.091\Delta r_{t-1} + u_{1,t}, & \text{if } 1.57 \geq z_{t-1} > -0.84, \\ 0.840 - 0.418z_{t-1} - 0.085\Delta R_{t-1} + 0.002\Delta r_{t-1} + u_{1,t}, & \text{if } z_{t-1} > 1.57, \end{cases}$$

$$\Delta r_t = \begin{cases} 1.748 + 1.572z_{t-1} + 0.484\Delta R_{t-1} + 0.081\Delta r_{t-1} + u_{2,t}, & \text{if } z_{t-1} \leq -0.84, \\ 0.002 - 0.031z_{t-1} + 0.324\Delta R_{t-1} + 0.210\Delta r_{t-1} + u_{2,t}, & \text{if } 1.57 \geq z_{t-1} > -0.84, \\ 0.757 - 0.325z_{t-1} + 0.041\Delta R_{t-1} - 0.016\Delta r_{t-1} + u_{2,t}, & \text{if } z_{t-1} > 1.57. \end{cases}$$

The two estimated cointegrating vectors are (1,-0.983) and (1,-0.981). Although the estimators provide different estimates of thresholds, they both divide the data into three regimes with similar groups of percentages, 8%, 72%, 20% for the LSE and 5.4%, 73.15% and 21.45% for the SLSE.

We compare the three-regime TVECM models with the two-regime TVEM model of Hansen and Seo (2002). It is obvious that they have almost the same amount of data in the left regime, and the three-regime models further put the remaining data into two regimes. Error-correction appears to occur only in the two regimes that stand for abnormal situations. The short-rate equations have larger error-correction effects when the spread is low and the long-rate equations have larger error-correction effects when the spread is high.

In the two estimated three-regime TVECMs, the coefficients in the left regime have larger absolute value than that in the right regime, indicating that the system is more sensitive to low  $R_t$  than to high  $R_t$ . This observation reflects the market behavior: because the former phenomenon is more unusual and unstable than the latter, the market adjusts more rapidly.

## 5. Conclusion

Existing asymptotic theories of estimation of threshold cointegration are limited to the one threshold case. We investigated the LSE and SLSE of multiple-threshold cointegration, based on the multiple-regime TVECM representation. The convergence rate of LSE was obtained and the limiting distribution of SLSE was established. For the two proposed estimation methods, estimates of thresholds and cointegration vector are super-consistent, while estimates of slope parameters are  $\sqrt{n}$ -consistent and asymptotically normal. The convergence rate of SLSE is slower than that of LSE, as a cost of developing the limiting distribution. Simulation study supports the theoretical result of the the convergence rate of the two estimators.

## Acknowledgement

We would like to thank the Editor and two anonymous referees for helpful comments and useful references, which led to an improved version of this paper.

This research was supported in part by grants from the General Research Fund of HKSAR-RGC-GRF Nos 400313 and 14300514 (Chan), HKSAR-RGC-ECS 405012 and HKSAR-RGC-GRF 405113 and 14300514 (Yau) and Fundamental Research Funds for the Central Universities 14D210809 (Wang).

## References

- Andrews, D. (1991). Heteroskedasticity and autocorrelation consistent covariance matrix estimation. *Econometrica* **59**, 817-858.
- Bai, J. and Perron, P. (1998). Estimating and testing linear models with multiple structural changes. *Econometrica* **66**, 47-78.
- Balke, N. S. and Fomby, T. B. (1997). Threshold cointegration. *Internat. Econom. Rev.* **38**, 627-645.
- Box, G. E. P. and Tiao, G. C. (1977). A canonical analysis of multiple time series. *Biometrika* **64**, 355-365.
- Campbell, J. (1995). Some lessons from the yield curve. *J. Econom. Perspectives* **9**, 129-152.
- Campbell, J. Y. and Shiller, R. J. (1987). Cointegration and tests of present value models. *J. Political Econom.* **95**, 1062-1088.
- Chan, K. (1993). Consistency and limiting distribution of the least squares estimator of a threshold autoregressive model. *Ann. Statist.* **21**, 520-533.
- de Jong, R. (2001). Nonlinear estimation using estimated cointegrating relations. *J. Econom.* **101**, 109-122.
- de Jong, R. (2002). Nonlinear minimization estimators in the presence of cointegrating relations. *J. Econom.* **110**, 241-259.
- Davis, R. A., Lee, T. C. M. and Rodriguez-Yam, G. A. (2006). Structural break estimation for nonstationary time series models. *J. Amer. Statist. Assoc.* **101**, 223-239.
- Enders, W. and Granger, C. (1998). Unit-root tests and asymmetric adjustment with an example using the term structure of interest rates. *J. Business Econom. Statist.* **16**, 304-311.
- Enders, W. and Siklos, P. (2001). Cointegration and threshold adjustment. *J. Business Econom. Statist.* **19**, 166-176.
- Engle, R. F. and Granger, C. W. J. (1987). Co-integration and error correction: Representation, estimation, and testing. *Econometrica* **55**, 251-276.
- Gonzalo, J. and Pitarakis, J.-Y. (2002). Estimation and model selection based inference in single and multiple threshold models. *J. Econom.* **110**, 319-352.
- Gonzalo, J. and Wolf, M. (2005). Subsampling inference in threshold autoregressive models. *J. Econom.* **127**, 201-224.
- Granger, C. W. J. (1983). Cointegrated variables and error correcting model. Discussion paper, **83-13a**. Department of Economics, University of California, San Diego .
- Hansen, B. E. and Seo, B. (2002). Testing for two-regime threshold cointegration in vector error-correction models. *J. Econom.* **110**, 293-318.
- Horowitz, J. L. (1992). A smoothed maximum score estimator for the binary response model. *Econometrica* **60**, 505-531.
- Li, D. and Ling, S. (2012). On the least squares estimation of multiple-regime threshold autoregressive models. *J. Econom.* **167**, 240-253.

- Lo, M. C. and Zivot, E. (2001). Threshold cointegration and nonlinear adjustment to the law of one price. *Discussion Papers in Economics at the University of Washington* 0030, Department of Economics at the University of Washington.
- McCulloch, H. and Kwon, H. (1993). US term-structure data, 1947-1991. Ohio State University Working Paper, **93-6**.
- Michael, P., Nobay, A. R. and Peel, D. A. (1997). Transactions costs and nonlinear adjustment in real exchange rates: An empirical investigation. *J. Polit. Econom.* **105**, 862-879.
- Saikkonen, P. (1995). Problems with the asymptotic theory of maximum likelihood estimation in integrated and cointegrated systems. *Econom. Theory* **11**, 888-911.
- Seo, M. H. (2006). Bootstrap testing for the null of no cointegration in a threshold vector error correction model. *J. Econom.* **134**, 129-150.
- Seo, M. H. (2011). Estimation of nonlinear error correction models. *Econom. Theory* **27**, 201-234.
- Seo, M. H. and Linton, O. (2007). A smoothed least squares estimator for threshold regression models. *J. Econom.* **141**, 704-735.
- Walter, E. and Kamol, C. (2004). Threshold cointegration and purchasing power parity in the pacific nations. *Appl. Ergonom.* **36**, 889-896.
- Yau, C. Y., Tang, C. M. and Lee, T. C. M. (2015). Estimation of multiple-regime threshold autoregressive models with structural breaks. *J. Amer. Statist. Assoc.* **110**, in press.

Department of Finance, Dong Hua University, No.1882, Yan-an Road West, Shanghai, 200051, China.

E-mail: wangman@dhu.edu.cn

Department of Statistics, The Chinese University of Hong Kong, Shatin, N.T. Hong Kong.

School of Statistics, Southwestern University of Finance and Economics, Chengdu, China, PRC 611130.

E-mail: nhchan@sta.cuhk.edu.hk

Department of Statistics, The Chinese University of Hong Kong, Shatin, N.T. Hong Kong.

E-mail: cyyau@sta.cuhk.edu.hk

(Received June 2014; accepted May 2015)