



MONASH University

Australia

Department of Econometrics and Business Statistics

<http://www.buseco.monash.edu.au/depts/ebs/pubs/wpapers/>

**A New Class of Bivariate Threshold
Cointegration Models**

Biqing Cai, Jiti Gao and Dag Tjøstheim

January 2015

Working Paper 01/15

A New Class of Bivariate Threshold Cointegration Models*

BIQING CAI[†], JITI GAO^{*} AND DAG TJØSTHEIM[†]

[†]University of Bergen, Norway and ^{*}Monash University, Australia

January 28, 2015

Abstract

In this paper, we introduce a new class of bivariate threshold VAR cointegration models. In the models, outside a compact region, the processes are cointegrated, while in the compact region, we allow different kinds of possibilities. We show that the bivariate processes form a $1/2$ -null recurrent system. We also find that the convergence rate for the estimators for the coefficients in the outside regime is \sqrt{T} , while the convergence rate for the estimators for the coefficients in the middle regime is $T^{1/4}$. Also, we show that the convergence rate of the cointegrating coefficient is T , which is same as linear cointegration model. The Monte Carlo simulation results suggest that the estimators perform reasonably well in finite samples. Applying the proposed model to study the dynamic relationship between Federal funds rate and 3-month Treasury bill rate, we find that cointegrating coefficients are the same for the two regimes while the short run loading coefficients are different.

JEL classification: C13; C22; E43

Keywords: β -null recurrent; Cointegration; Markov chain; Threshold VAR models

*The authors would like to thank Heather Anderson for her constructive comments and suggestions on an earlier version of this paper. The work of the first and the third authors was supported financially by the University of Bergen and the Norwegian Research Council. The second author acknowledges financial support by the Australian Research Council Discovery Grants Program under Grant Numbers: DP1096374 and DP130104229.

1 Introduction

Linear cointegration models as introduced by Granger (1981) and Engle and Granger (1987) have been extremely influential in econometrics. But they are surprisingly difficult to extend to the nonlinear case. In fact, no fully satisfying extension seems to exist.

A class of nonlinear models that has enjoyed a similar kind of influence in the time series modeling is the class of threshold models introduced by Tong (1978, 1983). But their analysis has largely been confined to the stationary case. They are difficult to extend to the nonstationary case even in a univariate situation.

To understand some of the reasons why these difficulties exist, it may be useful to go back to the (admittedly quite crude) classifications in stationary-nonstationary and linear-nonlinear models, which results in a 4-fold classification of combinations. The simplest category is the class of linear stationary models, which encompass ARMA models and traditional VAR models. Mixing and martingale methods are standard tools in the statistical analysis of these models. The same methods can essentially be applied to nonlinear stationary models. Verification of ergodicity/mixing plays a more important role but can often be handled using Markov chain techniques. The ordinary threshold model belongs to this category.

Moving to linear nonstationary models one encounters unit root and cointegration models which have been extensively used in econometrics. Since Phillips (1987) and numerous papers after that, the functional central limit theorem, a continuous mapping theorem and weak convergence to stochastic integral have been key techniques to study these models.

When moving to the (possibly largest) class of nonstationary and nonlinear models needed to discuss nonlinear cointegration, new problems arise. To illustrate this, let us take as our starting point the simple case of a bivariate first order linear cointegration model

$$x_t = Ax_{t-1} + e_t, \tag{1.1}$$

where $x_t = (x_{1,t}, x_{2,t})^\tau$, and $\{e_t\}$ is an i.i.d. process. The component processes $\{x_{1,t}\}$ and $\{x_{2,t}\}$ are unit root processes and the matrix A has one eigenvalue $\lambda_1 = 1$ and the other eigenvalue λ_2 with $|\lambda_2| < 1$. The eigenvector of λ_1 is orthogonal to the cointegrating vector β such that $\{\beta^\tau x_t\}$ is stationary. The component processes $\{x_{1,t}\}$ and $\{x_{2,t}\}$ are treated on the same basis in the analysis.

Formally, the nonlinear analogue of (1.1) is

$$x_t = f(x_{t-1}) + e_t, \tag{1.2}$$

where f may be unknown or known up to a parameter vector θ resulting in $f(x, \theta)$. But it is not clear how something like the cointegration concept should be introduced and at the same time retain the statistical techniques used for (1.1). One attempt that has been made is the so-called cointegration regression model where

$$y_t = g(z_t) + u_t, \tag{1.3}$$

where $\{z_t\}$ is univariate nonstationary process and $\{u_t\}$ is univariate stationary process, see, e.g. Park and Phillips (2001). There are basically two approaches to the analysis of (1.3). One, as in e.g., Park and Phillips (2001) or Wang and Phillips (2009a, 2009b), is to let $\{z_t\}$ to be a linear unit root process, and use the so-called local time approximation to the normalized partial sum of nonlinear transformations of the unit root process. The other one is a Markov chain approach as in Karlsen and Tjøstheim (2001) and Karlsen et al. (2007), where $\{z_t\}$ is assumed to be a null recurrent Markov chain, which in some respects is more general than $\{z_t\}$ being unit root but in other respects is more restrictive. The main conceptual difficulty with (1.3) is that from a cointegration point of view, the component processes $\{x_{1,t}\} = \{y_t\}$ and $\{x_{2,t}\} = \{z_t\}$ now have asymmetric roles to play, with $\{y_t\}$ and $\{z_t\}$ being respectively output and input processes in a regression model. The restricted model (1.3) is partly one of mathematical convenience. If, for example, one would like to use the theory developed for (1.1) and (1.3) in (1.2) one basically has to assume that $\{x_t\}$ in (1.3) is a linear cointegrated process, but this leads to a contradiction since linear cointegration is not invariant to nonlinear transformations and $\{x_t\}$ as an output process cannot be linearly cointegrated. This problem does not arise in (1.3) since $\{y_t\}$ does not have to be identified with $\{z_t\}$. Actually the nonlinear problem in (1.2) seems easier to handle with the Markov chain approach because under certain restrictions, null recurrence is invariant to nonlinear transformations.

A less ambitious goal is to restrict nonlinear cointegration to nonlinearity of a threshold type. Then it is straightforward, at least formally, to generalize (1.1) to

$$x_t = Ax_{t-1}1(x_{t-1} \in \mathbb{R}_1) + Bx_{t-1}1(x_{t-1} \in \mathbb{R}_2) + e_t, \quad (1.4)$$

where the regions \mathbb{R}_1 and \mathbb{R}_2 are disjoint subsets of \mathbb{R}^2 such that $\mathbb{R}_1 \cup \mathbb{R}_2 = \mathbb{R}^2$. Then linear cointegration concepts can conceivably be discussed separately for \mathbb{R}_1 and \mathbb{R}_2 and one can have stationary-like behaviour in one region and nonstationary cointegration-like in another, or even different cointegration-like behaviour in each region. A major mathematical difficulty is that in general the $\{x_t\}$ process will be nonstationary and at the same time x_t is the threshold variable deciding the regions \mathbb{R}_1 and \mathbb{R}_2 . Earlier attempts have sought to avoid this problem by introducing an indirect *stationary* threshold variable.

The above problem appears also in the univariate case if one tries to generalize the ordinary unit root processes to threshold unit root processes. We refer to Gao et al. (2013) for a discussion and references. In that paper, the authors were able to construct a theory for the univariate special case of (1.4), and it is our intention to try to extend that theory to a cointegration threshold model such as (1.4) in the present paper.

As mentioned above, existing threshold cointegration models are largely based on stationary threshold variables (cf. Balke and Fomby 1997). Quite frequently these threshold models start from the so-called error correction representation of (1.1), written as

$$\Delta x_t = \alpha\beta^\tau x_{t-1} + e_t,$$

where β is the cointegrating vector and α is the short run loading coefficient (adjustment coefficient). Note that in the bivariate case, α and β are both 2×1 vectors. Simple matrix

manipulation shows that β is orthogonal to the eigenvector of eigenvalue 1 while α is parallel to the other eigenvector. A large part of any nonlinear generalization of error correction models is to provide conditions under which an analogue of the Granger representation theorem holds, implying stationarity of the linear combination $\beta^\tau x_{t-1}$ and the difference Δx_t , see, e.g. Bec and Rahbek (2004), Saikkonen (2005, 2008). In particular, a three-regime threshold error correction model has been proposed by Saikkonen (2008),

$$\Delta x_t = \sum_{s=1}^3 \alpha_s \beta^\tau x_{t-1} 1(\beta^\tau x_{t-1} \in \mathbb{R}_s) + e_t, \quad (1.5)$$

where the threshold variable $\beta^\tau x_{t-1}$ is supposed to be stationary. Compared to the ordinary threshold model, where typically the lag variable x_{t-1} is the threshold variable, this is a somewhat indirect way of formulating a nonstationary threshold VAR model.

In this paper, we start directly from the VAR representation (1.1) and then move to the threshold VAR representation (1.4) and make a first attempt of establishing a theory for such models. In doing so, we will work under some restrictions, most of which we think can be relaxed in subsequent work. First, we only look at the bivariate first order case. Second, the region \mathbb{R}_1 is supposed to be compact. In this region, quite general behavior of $\{x_t\}$ is permitted, i.e. the matrix A in (1.4) can be arbitrary, and including stationary and explosive type behaviours (cf. Gao et al. 2013 in the univariate case). In the region \mathbb{R}_2 , $\{x_t\}$ will be assumed to be a cointegrated process, and we will define precisely what this means. Obviously, there are many generalizations in choosing regions, i.e. there could be more than two regions, and in particular there could be more than one cointegrating region and with different cointegrating vectors. Third, we will limit ourselves to estimation for a given model. Testing problems will be considered in forthcoming papers.

Markov recurrence theory plays an important role in all of this, not the least in determining the different convergence rates for the estimators in different regions. This will be treated in Section 2 and the Appendices. The asymptotic estimation theory is developed in Section 3, whereas simulation experiments are conducted in Section 4. And finally in Section 5, we give an empirical example concerning the dynamic relationship of federal funds rate and 3-month T-bill rate. One reason for using this example is that there is a threshold and in the particular example it will be seen that the general model can be reduced to an error correction threshold model of type (1.5). Finally, Section 6 summarizes this paper. Some basic theory of null recurrent Markov chain is provided in Appendix A. All mathematical proofs are given in the Appendices B and C.

Throughout this paper, all limits are taken “as $T \rightarrow \infty$ ” with T being the number of observations, \rightarrow_d denotes weak convergence, \rightarrow_p denotes convergence in probability, $\rightarrow_{a.s.}$ denotes almost sure convergence. $O_P(\cdot)$ means stochastic order same as, $o_P(\cdot)$ means stochastic order less than. \otimes denotes Kronecker product, vec denotes vectorization, A^τ denotes transpose of a matrix A .

2 Model and Its Probability Properties

2.1 The Model

The bivariate model we are going to examine is

$$x_t = Ax_{t-1}1(x_{t-1} \in D) + Bx_{t-1}1(x_{t-1} \in D^c) + e_t, \quad (2.1)$$

where $t = 1, \dots, T$, $x_t = (x_{1,t}, x_{2,t})^\top$ and $x_0 = (0, 0)^\top$.

Compared with the threshold error correction models as in Saikkonen (2005, 2008), where nonstationarity is disentangled from nonlinearity in the sense that linear cointegration relationship is imposed and the threshold variable is driven by stationary variables, our model is self-excited in the traditional way such that a lagged variable constitutes the threshold variable. Our model is also different from the nonlinear cointegration models in Park and Phillips (2001) or Karlsen et al. (2007) where a right-hand-side variable y is regressed on left-hand-side variable x . In our model, we need not impose x_1 or x_2 to be a regressor and (2.1) is what we believe to be a natural extension of linear VAR models. The appropriateness of VARs is well documented. Stock and Watson (2001) write “This simple (linear VAR) framework provides a systematic way to capture rich dynamics in multiple time series, and the statistical toolkit that came with VARs was easy to use and to interpret. As Sims (1980) and others argued in a series of influential early papers, VARs held out the promise of providing a coherent and credible approach to data description, forecasting, structural inference and policy analysis.”

A crucial tool in our analysis of (2.1) is the theory of null recurrent Markov processes as developed in Karlsen and Tjøstheim (2001). The connection between linear VARs and null recurrent process is pointed out in Myklebust et al. (2012). A first attempt to use this in a threshold context is given in Gao et al. (2013) in the univariate case. In that paper, it is proved that a univariate threshold model is β -null recurrent with $\beta = 1/2$. That proof relies on a result in Meyn and Tweedie (2009, Prop 11.5.4) giving conditions for a univariate threshold process to be recurrent. This result makes it possible to use the Markov splitting technique to obtain $1/2$ -null recurrence. The proof of recurrence in Meyn and Tweedie (2009) is based on a drift criterion using a Liapunov test function. We will essentially use the same technique in our proof of Theorem 1 below, but a nontrivial change of test function is required. Once recurrence is established for (2.1), $1/2$ -null recurrence can be proved using a straightforward adaptation of the proof in Gao et al. (2013).

2.2 Recurrence of the Process

We make the following assumptions, several of which may be relaxed:

Assumption 1. The sequence $\{e_t\}$ is an i.i.d. zero mean random variable on \mathbb{R}^2 with its distribution function being absolutely continuous with respect to the Lebesgue measure. Let $p_e(\cdot, \cdot)$ being the density function, we have $\inf_{e \in C} p_e(e_1, e_2) > 0$ for any compact set C . $\{e_t\}$ has covariance matrix Σ and $E(|e_{i,t}|^4) < \infty$ for $i = 1, 2$.

Assumption 2. D is a compact set in \mathbb{R}^2 .

Assumption 3. $B_{2 \times 2}$ is a matrix with eigenvalue 1 and c with $|c| < 1$, and $A_{2 \times 2}$ can be any matrix. This is a special case of threshold cointegration model.

Remark: Assumption 1 is typical in studying time series properties using Markov chain techniques. Assumption 2 requires that the behaviors of the processes are different inside and outside a compact region, which is similar to Gao et al. (2013). Assumption 3 requires that outside the compact support, the bivariate processes are cointegrated, while in the compact support, we impose no restrictions with both stationary and nonstationary behaviors allowed. As a special case, they may be cointegrated with the same cointegrating vector in the compact set D as in D^c whereas the short run adjustment vectors are different in the two sets.

Based on the above assumptions, we have the following theorem.

Theorem 1. Under Assumptions 1–3, $\{x_t\}$ is a 1/2-null recurrent process.

Remark: More discussion of 1/2-null recurrent process is given in Appendix A. Since the process is 1/2-null recurrent, we can use the splitting chain technique in Appendix A (cf. Karlsen and Tjøstheim 2001) to derive the limit theory for the coefficient estimation for the process.

3 Coefficient Estimation

Convergence rates for coefficient estimates is an important issue in linear VAR cointegration models. In this section, we will discuss the asymptotics for coefficient estimations in the threshold cointegration models. Similar to Lütkepohl (2005), we will first discuss the asymptotics for the transformed model obtained by diagonalizing the B matrix and then discuss the asymptotics for the original model.

3.1 Coefficient Estimation for the Transformed System

Using the matrix Q^{-1} composed of eigenvectors of B , we can transform the process into

$$y_t = A_1 y_{t-1} 1(y_{t-1} \in C) + B_1 y_{t-1} 1(y_{t-1} \in C^c) + \varepsilon_t, \quad (3.1)$$

where $B_1 = QBQ^{-1}$, $A_1 = QAQ^{-1}$, $\varepsilon_t = Qe_t$ and C is the compact set obtained by transforming D (see also the proof of Theorem 1). Here, $\{\varepsilon_t\}$ is an i.i.d. sequence with covariance matrix Σ_1 . It is easy to see that $\Sigma_1 = Q\Sigma Q^T$.

We will estimate the coefficient matrices A_1 and B_1 using OLS, with the estimators denoted by \hat{A}_1 and \hat{B}_1 , i.e.,

$$\hat{A}_1 = \left(\sum_{t=1}^T y_t y_{t-1}^T 1(y_{t-1} \in C) \right) \left(\sum_{t=1}^T y_{t-1} y_{t-1}^T 1(y_{t-1} \in C) \right)^{-1}$$

and

$$\widehat{B}_1 = \left(\sum_{t=1}^T y_t y_{t-1}^\tau \mathbf{1}(y_{t-1} \in C^c) \right) \left(\sum_{t=1}^T y_{t-1} y_{t-1}^\tau \mathbf{1}(y_{t-1} \in C^c) \right)^{-1}.$$

Thus

$$\widehat{A}_1 - A_1 = \left(\sum_{t=1}^T \varepsilon_t y_{t-1}^\tau \mathbf{1}(y_{t-1} \in C) \right) \left(\sum_{t=1}^T y_{t-1} y_{t-1}^\tau \mathbf{1}(y_{t-1} \in C) \right)^{-1},$$

and

$$\widehat{B}_1 - B_1 = \left(\sum_{t=1}^T \varepsilon_t y_{t-1}^\tau \mathbf{1}(y_{t-1} \in C^c) \right) \left(\sum_{t=1}^T y_{t-1} y_{t-1}^\tau \mathbf{1}(y_{t-1} \in C^c) \right)^{-1},$$

where $B_1 = \begin{bmatrix} 1 & 0 \\ 0 & c \end{bmatrix}$ and $A_1 = \begin{bmatrix} a_{1,11} & a_{1,12} \\ a_{1,21} & a_{1,22} \end{bmatrix}$.

According to (3.1), we have

$$\Delta y_{1,t} \equiv y_{1,t} - y_{1,t-1} = [(a_{1,11} - 1)y_{1,t-1} + a_{1,12}y_{2,t-1}] \mathbf{1}(y_{t-1} \in C) + \varepsilon_{1,t}$$

and

$$y_{2,t} = cy_{2,t-1} + [a_{1,21}y_{1,t-1} + (a_{1,22} - c)y_{2,t-1}] \mathbf{1}(y_{t-1} \in C) + \varepsilon_{2,t}.$$

Let

$$u_t = \begin{pmatrix} u_{1,t} \\ u_{2,t} \end{pmatrix} \equiv \begin{pmatrix} [(a_{1,11} - 1)y_{1,t-1} + a_{1,12}y_{2,t-1}] \mathbf{1}(y_{t-1} \in C) \\ [a_{1,21}y_{1,t-1} + (a_{1,22} - c)y_{2,t-1}] \mathbf{1}(y_{t-1} \in C) \end{pmatrix}, \quad (3.2)$$

so that

$$\Delta y_{1,t} = u_{1,t} + \varepsilon_{1,t}.$$

According to the Donsker's Lemma, we have

$$\frac{1}{\sqrt{T}} \sum_{t=1}^{[Tr]} \varepsilon_t \rightarrow_d B(r),$$

where $B(r) = (B_1(r), B_2(r))^\tau$ is a two-dimensional Brownian motion with covariance variance matrix Σ_1 . And by Karlsen and Tjøstheim (2001), we have (cf. Proof of (2.8) in Gao et al. 2013):

$$\frac{1}{n(T)} \sum_{t=1}^T u_{1,t} \rightarrow_p \int [(a_{11} - 1)y_1 + a_{1,2}y_2] \mathbf{1}(y \in C) p_s(y) dy \equiv \mu_1,$$

where $n(T)$, being the number of regenerations of the 1/2 null recurrent process $\{y_t\}$, is asymptotically proportional to the number of visits to C and $p_s(y)$ is the invariant density of the bivariate process $\{y_t\}$. Because we have shown that this process is 1/2-null recurrent, we know that $n(T) = O_P(T^{1/2})$. So that, we have:

$$\frac{1}{\sqrt{T}} y_{1,[Tr]} \rightarrow_d B_1(r) + M_{1/2}(r) \mu_1 \equiv P(r) \quad (3.3)$$

with $M_{1/2}(r)$ being a Mittag-Leffler process with parameter 1/2. See Appendix A for more discussion.

3.1.1 Coefficient Estimation in the Middle Regime

An asymptotic distribution for the estimators of the coefficients in the middle regime is given below.

Theorem 2. Under Assumptions 1–3, we have

$$\sqrt{n(T)}\text{vec}\left(\widehat{A}_1 - A_1\right) \rightarrow_d N\left(0, \left(M^{-1} \otimes \Sigma_1\right)\right),$$

$$\text{where } M = \begin{pmatrix} \int y_1^2 1(y \in C) p_s(y) dy & \int y_1 y_2 1(y \in C) p_s(y) dy \\ \int y_1 y_2 1(y \in C) p_s(y) dy & \int y_2^2 1(y \in C) p_s(y) dy \end{pmatrix}.$$

Remark: The proof is given in Appendix B and is an easy extension of Lemma 2.1 in Gao et al. (2013). The convergence rate is $\sqrt{n(T)}$ with normality limiting distribution. Because we know that $\frac{n([Tr])}{\sqrt{T}} \rightarrow_d M_{1/2}(r)$, thus

$$T^{1/4}\text{vec}\left(\widehat{A}_1 - A_1\right) \rightarrow_d MN\left(0, \left(M^{-1}M_{1/2}^{-1}(1) \otimes \Sigma_1\right)\right).$$

That is, if we use deterministic normalization, the limiting distribution is mixed normal.

3.1.2 Coefficient Estimation in the Outside Regime

To prove the asymptotic for the estimators in the outside regime, i.e. for the matrix B_1 , we need the following two Lemmas.

Lemma 1. Under Assumptions 1–3, we have the following results:

(1)

$$\frac{1}{T^2} \sum_{t=1}^T y_{1,t-1}^2 \rightarrow_d \int_0^1 P^2(r) dr,$$

(2)

$$\frac{1}{T} \sum_{t=1}^T y_{2,t-1}^2 \rightarrow_p \frac{1}{1-c^2} \sigma_{\varepsilon_2}^2,$$

(3)

$$\frac{1}{T^{3/2}} \sum_{t=1}^T y_{1,t-1} y_{2,t-1} = o_P(1).$$

Lemma 2. Under Assumptions 1–3, we have the following results:

(1)

$$\frac{1}{T} \sum_{t=1}^T y_{1,t-1} \varepsilon_t \rightarrow_d \int_0^1 P(r) dB(r),$$

(2)

$$\frac{1}{\sqrt{T}} \sum_{t=1}^T y_{2,t-1} \varepsilon_t \rightarrow_d N(0, \sigma_{y_2}^2 \Sigma_1),$$

where $\sigma_{y_2}^2 = \frac{1}{1-c^2} \sigma_{\varepsilon_2}^2$ and Σ_1 is the covariance matrix of ε_t .

Based on Lemmas 1 and 2, we have following theorem.

Theorem 3. Let $H = \begin{pmatrix} T & 0 \\ 0 & T^{1/2} \end{pmatrix}$. Under Assumptions 1–3, then

$$vec[(\widehat{B}_1 - B_1) \times H] \rightarrow_d \begin{pmatrix} \frac{\int_0^1 P(r)dB(r)}{\int_0^1 P(r)^2 dr} \\ N(0, (\sigma_{y_2}^2)^{-1}\Sigma_1) \end{pmatrix}.$$

Remark: We note that the first column of the estimators converges with rate T while the second converges with rate \sqrt{T} . This is similar to the linear cointegration case, although the limiting distribution is different.

3.2 Estimation of the Original Model

3.2.1 Estimation for Coefficients of the Original Model

As in the ordinary linear cointegration model, convergence rates are changed when we transform back to the original model. For the OLS estimator, denoted by \widehat{A} and \widehat{B} , we have

$$\widehat{A} - A = \left(\sum_{t=1}^T e_t x_{t-1}^\tau 1(x_{t-1} \in D) \right) \left(\sum_{t=1}^T x_{t-1} x_{t-1}^\tau 1(x_{t-1} \in D) \right)^{-1},$$

and

$$\widehat{B} - B = \left(\sum_{t=1}^T e_t x_{t-1}^\tau 1(x_{t-1} \in D^c) \right) \left(\sum_{t=1}^T x_{t-1} x_{t-1}^\tau 1(x_{t-1} \in D^c) \right)^{-1}.$$

For the coefficients in the middle regime, the convergence rate is $n(T) \sim T^{1/4}$ as in the univariate case of Gao et al. (2013).

Theorem 4. Under Assumptions 1–3, we have

$$\sqrt{n(T)}vec\left(\widehat{A} - A\right) \rightarrow_d N\left(0, (M_1^{-1} \otimes \Sigma)\right),$$

where $M_1 = \begin{pmatrix} \int x_1^2 1(x \in D) p_s(x) dx & \int x_1 x_2 1(x \in D) p_s(x) dx \\ \int x_1 x_2 1(x \in D) p_s(x) dx & \int x_2^2 1(x \in D) p_s(x) dx \end{pmatrix}$, with $p_s(x)$ being the invariant density of $\{x_t\}$.

For the coefficients in the outside regime, we obtain the same convergence rate as in the linear cointegration case.

Theorem 5. Under Assumptions 1–3, we have

$$\sqrt{T}vec\left(\widehat{B} - B\right) \rightarrow_d N(0, \Sigma_2),$$

where $\Sigma_2 = (Q^\tau \otimes Q^{-1}) \left(\left(\begin{pmatrix} 0 & 0 \\ 0 & (\sigma_{y_2}^2)^{-1} \end{pmatrix} \otimes \Sigma_1 \right) (Q \otimes (Q^{-1})^\tau) \right)$.

3.2.2 Cointegrating Parameter Estimation

As in the ordinary linear cointegration, we can express the cointegrating vector in terms of the inverse transformation matrix $Q = (q_{i,j})$ in the same way as has been done in the representation (3.1) for $\{y_t\}$. The first component of $\{y_t\}$ is a random walk in C^c , whereas the second component displays stationary behaviour in C^c . Hence, since $y_t = Qx_t$, the cointegrating vector in this region is given by $[q_{21} \ q_{22}]$ such that $q_{21}x_{1,t} + q_{22}x_{2,t}$ behaves as a stationary process in C^c .

Often the cointegrating vector is normalized so that it has one component equal to 1. Without loss of generality, we can assume that $q_{22} \neq 0$ and the the cointegrating vector can be normalized to $[\frac{q_{21}}{q_{22}} \ 1]$ so that

$$x_{2,t} = -\frac{q_{21}}{q_{22}}x_{1,t} + \omega_t,$$

with ω_t being a stationary error in D^c . In linear cointegration model, the cointegrating coefficient $\beta = -\frac{q_{21}}{q_{22}}$ is usually estimated by OLS (cf. Engle and Granger 1987)

$$\hat{\beta} = \frac{\sum_{t=1}^T x_{1,t}x_{2,t}}{\sum_{t=1}^T x_{1,t}^2}. \quad (3.4)$$

From the above analysis, we can see that the cointegrating coefficient β is related to unknown transformation matrix Q . In the literature, there are two methods commonly used for cointegrating vector estimation in VAR cointegration models:

- Solving eigenvalue problems in estimated VAR coefficients, such as Johansen (1988);
- Estimation in a triangular system as in Ahn and Reinsel (1990) or Phillips (1991).

We have found the latter to be more convenient in our situation because we can express $x_{1,t}$ as a random walk plus an additional term and $x_{2,t}$ as a linear transformation of $x_{1,t}$ plus an additional term. This enables us to find an explicit expression for the relationship between $x_{2,t}$ and $x_{1,t}$, which makes the mathematical derivation possible. In the following lemma, we find a triangular representation for our threshold VAR cointegration model.

Lemma 3. Under Assumptions 1–3, we have

$$\left\{ \begin{array}{l} x_{1,t} = k_1 \sum_{j=0}^t (u_{1,t-j} + \varepsilon_{1,t-j}) + k_2 \sum_{j=0}^t c^j (u_{2,t-j} + \varepsilon_{2,t-j}) \\ x_{2,t} = \beta x_{1,t} + k_3 \sum_{j=0}^t c^j (u_{2,t-j} + \varepsilon_{2,t-j}), \end{array} \right\}, \quad (3.5)$$

where $k_1 = \frac{q_{22}}{q_{11}q_{22} - q_{12}q_{21}}$, $k_2 = -\frac{q_{12}}{q_{11}q_{22} - q_{12}q_{21}}$, $\beta = \frac{-q_{21}}{q_{22}}$ and $k_3 = \frac{1}{q_{22}}$ with q_{ij} being elements of the matrix Q and $u_t = (A_1Q - B_1Q)x_{t-1}1(x_{t-1} \in D) = (u_{1,t}, u_{2,t})^\tau$ as defined in (3.2).

Based on Lemma 3 and considering the similarity and difference between our model and ordinary VAR cointegration model, we can construct two estimators for β with the first one, denoted by $\hat{\beta}_1$, is defined as

$$\hat{\beta}_1 = \frac{\sum_{t=1}^T x_{1,t}x_{2,t}1(x_{t-1} \in D^c)}{\sum_{t=1}^T x_{1,t}^2 1(x_{t-1} \in D^c)} \quad (3.6)$$

and the second one, denoted by $\widehat{\beta}_2$, is defined as

$$\widehat{\beta}_2 = \frac{\sum_{t=1}^T x_{1,t}x_{2,t}}{\sum_{t=1}^T x_{1,t}^2}. \quad (3.7)$$

Remark: We can see that $\widehat{\beta}_1$ is one estimator using the observations in the outside regime, while $\widehat{\beta}_2$ is another estimator using the observations of the whole sample, which is same as the estimator in the ordinary VAR cointegration model as defined in (3.4).

The following Theorem provides a limiting distribution for $\widehat{\beta}_1$ and $\widehat{\beta}_2$.

Theorem 6. Under Assumptions 1–3, we have

$$T(\widehat{\beta}_1 - \beta) \rightarrow_d \frac{k_1 k_3 \frac{1}{1-c} \int_0^1 P(r) dB_2(r) + \lambda}{k_1^2 \int_0^1 P^2(r) dr},$$

where $k_2 k_3 \frac{1}{1-c^2} \sigma_{\varepsilon_2}^2$ with $P(r)$ and $B_2(r)$ having been defined in Section 3.1, k_1 , k_2 , and k_3 are as defined in Lemma 3 and $\sigma_{\varepsilon_2}^2$ is the variance of $\varepsilon_{2,t}$. Furthermore, the asymptotic distribution of $\widehat{\beta}_2$ is the same as that of $\widehat{\beta}_1$.

The proofs of the main results given above will be given in the Appendices at the end of this paper.

Remark: From this theorem, we can see that the convergence rate of the OLS cointegrating coefficient estimator is T with a nonstandard limiting distribution. The variance of the limiting variable is dependent on the Q -matrix as well as on $P(r)$ and $B_2(r)$. Similar to Phillips and Hansen (1990), we find a second order bias term λ appearing in the limiting distribution, which is quite common in regression with I(1) processes with endogeneity. It is also interesting to see that the two estimators defined in (3.6) and (3.7) share the same asymptotics because the effect of using the middle regime observations is negligible asymptotically.

4 Monte Carlo Simulation

We study two cases: one is that the threshold variables are known, the other is that the threshold variables are unknown. In the theoretical part of this paper, we have examined the asymptotics of the coefficient estimators with known threshold. There exists a natural least squares estimator for the threshold but its asymptotics is difficult to evaluate for our model. A discussion of this is given in Gao et al. (2013) for the univariate nonstationary case based on results by Chan (1993) in the stationary case. Much of that discussion carries over to the present case. It is a challenging research topic to establish a full asymptotic theory for the threshold estimate. In the present paper, we restrict ourselves to examine the threshold estimator by simulation, also investigating its effect on estimation of the autoregressive coefficients and the cointegrating vector.

There is an issue in simulation experiments involving nonstationary I(1) type threshold type processes which is largely absent in the stationary case. For a sample size T , as $T \rightarrow \infty$, the

number of observations in a compact set D will be of order \sqrt{T} (also true for an ordinary non-threshold bivariate cointegration process), whereas the number of observations outside D will be of order $T - \sqrt{T}$. This means that for moderately large values of T and for a moderately small set D , one may encounter simulated realizations that have very few observations in D . Clearly, if this happens in practice, we will not be able to identify D and we will just fit an ordinary linear cointegration model to the data. For this reason, we have discarded simulated realizations where the number of observations in D is less than $0.5\sqrt{T}$. On the other hand, if D is comparably large, for a finite T , it is possible that most of the observations are in D and only a few in D^c . We therefore discard the simulated realizations where the number of observations in D^c is less than $0.1T$. In our empirical example to be presented in Section 5, in fact about two thirds of the data are in D .

4.1 Estimation with Known Thresholds

We are going to estimate the coefficients in Model (2.1) with $B = (b_{ij})$ being a matrix with eigenvalue 1 and c with $|c| < 1$, $A = (a_{ij})$ being any matrix. We simulate $x_t = (x_{1,t}, x_{2,t})^\tau$ with initial value $x_0 = (x_{1,0}, x_{2,0})^\tau = [0, 0]^\tau$, $e_t \sim i.i.d.N(0, \Sigma)$, and we let $\Sigma = cov(e_t) = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$.

The choices of A and B are different for 3 different examples. We do truncation as described above: we use the 1000 replications where the number of observation in the middle regime is larger than $0.5\sqrt{T}$ and the number of observations in the outside regime is larger than $0.1T$. The coefficients are estimated using OLS as defined in Section 3. We also provide the cointegrating parameter estimators: $\hat{\beta}_1, \hat{\beta}_2$, which are defined respectively in (3.6) and (3.7). Theorem 6 shows that the asymptotics for these two estimators are the same.

The means and standard deviations for the estimators for the case where $A = \begin{pmatrix} 1/2 & 0 \\ 0 & 1/2 \end{pmatrix}$, $B = \begin{pmatrix} 1 & 0 \\ 0 & 1/2 \end{pmatrix}$ and $D = [-3, 3] \times [-2, 2]$ are summarized in Table 1 below.

Table 1: Mean and STD of Estimators

T	200	400	800	1600	3200
a11	0.5028 (0.0868)	0.5074 (0.0721)	0.5034 (0.0513)	0.5005 (0.0406)	0.5010 (0.0314)
a12	-0.0058 (0.1143)	-0.0022 (0.0853)	-0.0011 (0.0657)	-0.0027 (0.0488)	-4.7868×10^{-4} (0.0369)
a21	-0.0036 (0.0918)	0.0019 (0.0625)	5.2745×10^{-4} (0.0508)	7.8487×10^{-4} (0.0413)	1.8928×10^{-4} (0.0307)
a22	0.5059 (0.1053)	0.5013 (0.0785)	0.5018 (0.0653)	0.5014 (0.0494)	0.4989 (0.0360)
b11	0.9461 (0.0777)	0.9675 (0.0432)	0.9834 (0.0251)	0.9917 (0.0151)	0.9968 (0.0062)
b12	2.2762×10^{-4} (0.0938)	1.4267×10^{-4} (0.0637)	-7.2827×10^{-4} (0.0435)	-5.1753×10^{-4} (0.0277)	-5.0922×10^{-4} (0.0197)
b21	-0.0042 (0.0792)	-0.0027 (0.0407)	5.1740×10^{-4} (0.0204)	3.9537×10^{-4} (0.0093)	4.6373×10^{-5} (0.0032)
b22	0.4863 (0.0884)	0.4919 (0.0594)	0.4958 (0.0397)	0.4976 (0.0278)	0.4994 (0.0197)
β_1	-0.0035 (0.2032)	-0.0031 (0.0873)	-5.9512×10^{-4} (0.0451)	6.1899×10^{-4} (0.0178)	7.9332×10^{-5} (0.0063)
β_2	-0.0010 (0.0752)	-0.0018 (0.0436)	$-9.5044e \times 10^{-4}$ (0.0250)	$5.9427e \times 10^{-4}$ (0.0123)	$6.3912e \times 10^{-4}$ (0.0055)

In this model, the two processes are independent with the first process being threshold unit root and the second process being a stationary variable. From Theorem 2 and Theorem 3, we know that for this case the coefficients in the middle regime converge with rate $T^{1/4}$,

while the first column of the autoregressive matrix in the outside regime converges with rate T , and the second column of the autoregressive matrix in the outside regime converges with rate \sqrt{T} . The Monte Carlo simulation results are consistent with the theory with a possible small perturbation due to the truncation rule. The rates can be checked by simply comparing the standard errors for various number of observations. We also notice that the cointegrating coefficients $\beta = 0^1$ converge with rate T , which is also consistent with Theorem 6. It is interesting to note that the standard deviation of the estimators with the full sample is less than the standard deviation of the estimators with the sample in the outside (true) regime. This is due to the fact that in this case, both in the middle regime and the outside regime, the two processes are independent. Thus, use of the whole sample provides more information for the (zero) ‘‘cointegrating’’ coefficient.

The results for the case where $A = \begin{pmatrix} 1/2 & 0 \\ 0 & 1/2 \end{pmatrix}$, $B = \begin{pmatrix} 3/4 & 1/4 \\ 1/4 & 3/4 \end{pmatrix}$, and $D = [-3, 3] \times [-2, 2]$ are summarized in Table 2 below.

Table 2: Mean and STD of Estimators

T	200	400	800	1600	3200
a11	0.4943 (0.1040)	0.4942 (0.0821)	0.4990 (0.0582)	0.4986 (0.0469)	0.4990 (0.0405)
a12	-0.0013 (0.1206)	0.0044 (0.0954)	0.0037 (0.0725)	0.0018 (0.0523)	0.0017 (0.0523)
a21	0.0051 (0.1147)	0.0031 (0.0862)	0.0024 (0.0702)	0.0004 (0.0499)	0.0017 (0.0382)
a22	0.5057 (0.1175)	0.5006 (0.0939)	0.4946 (0.0812)	0.4993 (0.0566)	0.5013 (0.0460)
b11	0.7051 (0.0906)	0.7284 (0.0514)	0.7423 (0.0303)	0.7473 (0.0196)	0.7485 (0.0129)
b12	0.2537 (0.0725)	0.2501 (0.0450)	0.2493 (0.0290)	0.2491 (0.0192)	0.2503 (0.0129)
b21	0.2330 (0.0887)	0.2395 (0.0520)	0.2457 (0.0307)	0.2500 (0.0196)	0.2488 (0.0127)
b22	0.7272 (0.0697)	0.7396 (0.0465)	0.7458 (0.0290)	0.7459 (0.0198)	0.7499 (0.0128)
β_1	0.8233 (0.2631)	0.9065 (0.1450)	0.9583 (0.0647)	0.9785 (0.0329)	0.9919 (0.0131)
β_2	0.6499 (0.3102)	0.7823 (0.2484)	0.8953 (0.1529)	0.9497 (0.0831)	0.9843 (0.0292)

In this model, in the outside regime, the two processes are cointegrated with cointegrating vector $[1 \ -1]^{\tau^2}$ while in the middle regime, they are stationary and independent. From Theorem 4 and Theorem 5, we know that for this case the coefficients in the middle regime converge with rate $T^{1/4}$, while the convergence rate for the coefficients in the outside regime

¹In the terminology of Lütkepohl (2005) (P. 250), this model is also a ‘‘cointegration’’ model. However, it specifies a relationship between a stationary variable and a nonstationary variable, so that it is actually not a ‘‘cointegration’’ model according to the definition of Engle and Granger (1987).

²This is due to the fact that the cointegrating vector is orthogonal to the eigenvector of eigenvalue equal to 1 as discussed in the Introduction.

is \sqrt{T} . The simulation results in Table 2 are consistent with the theoretical results. It also appears that the convergence rate for the cointegrating coefficient is T which is consistent with Theorem 6. The cointegrating coefficient estimators using the outside regime sample have smaller standard deviation and less bias because the estimators using the whole sample use the $O_P(\sqrt{T})$ observations in the middle regime as well where these two variables have different relationship from the outside regime.

The result for the case where $A = \begin{pmatrix} 3/4 & 1/8 \\ 1 & 1/2 \end{pmatrix}$, $B = \begin{pmatrix} 3/4 & 1/4 \\ 1/4 & 3/4 \end{pmatrix}$, i.e., both two regimes are cointegration regimes, and $D = [-3, 3] \times [-2, 2]$ is summarized in Table 3 below.

Table 3: Mean and STD of Estimators

T	200	400	800	1600	3200
a11	0.7521 (0.2123)	0.7373 (0.1947)	0.7452 (0.1555)	0.7375 (0.1213)	0.7497 (0.0989)
a12	0.1221 (0.2975)	0.1364 (0.2360)	0.1314 (0.1862)	0.1381 (0.1574)	0.1225 (0.1215)
a21	1.0200 (0.2212)	1.0092 (0.1772)	1.0064 (0.1470)	1.0051 (0.1233)	0.9998 (0.1004)
a22	0.4905 (0.2707)	0.4980 (0.2222)	0.5022 (0.1795)	0.4975 (0.1580)	0.5016 (0.1242)
b11	0.7382 (0.0488)	0.7437 (0.0322)	0.7486 (0.0220)	0.7483 (0.0154)	0.7492 (0.0106)
b12	0.2545 (0.0483)	0.2520 (0.0319)	0.2493 (0.0219)	0.2505 (0.0152)	0.2502 (0.0106)
b21	0.2514 (0.0475)	0.2518 (0.0319)	0.2493 (0.0231)	0.2514 (0.0153)	0.2500 (0.0109)
b22	0.7411 (0.0462)	0.7445 (0.0320)	0.7484 (0.0230)	0.7475 (0.0152)	0.7493 (0.0109)
β_1	0.9742 (0.0413)	0.9826 (0.0226)	0.9877 (0.0139)	0.9933 (0.0079)	0.9959 (0.0044)
β_2	0.9684 (0.0427)	0.9801 (0.0237)	0.9867 (0.0146)	0.9929 (0.0082)	0.9957 (0.0046)

In this model, in the outside regime, the two processes are cointegrated with a cointegrating vector $[1 \ -1]^\tau$ while in the middle regime, they are also cointegrated but with a cointegrating vector $[1 \ -1/2]^\tau$. From Theorem 4 and Theorem 5, we know that for this case the coefficients in the middle regime converge with rate $T^{1/4}$, while the convergence rate for the coefficients in the outside regime is \sqrt{T} . The simulation results in Table 3 are consistent with the theoretical results. We also find that the rate of convergence for the cointegrating coefficient appears to be T which is consistent with Theorem 6. The cointegrating coefficient estimators using the outside regime sample have smaller standard deviation and slightly less bias due to same reasoning as explained for Table 2, but the difference is smaller and the errors themselves much smaller, possibly due to a more similar structure in outside and middle regimes.

Table 4: Mean and STD of Estimators

T	200	400	800	1600	3200
a11	0.4829 (0.1154)	0.4864 (0.0846)	0.4916 (0.0484)	0.4955 (0.0445)	0.4971 (0.0308)
a12	0.0024 (0.1644)	0.0013 (0.1458)	-0.0052 (0.1058)	-0.0013 (0.0658)	-3.5740×10^{-4} (0.0469)
a21	-0.0015 (0.1118)	-0.0015 (0.0735)	0.0048 (0.0537)	-1.7981×10^{-5} (0.0463)	9.5832×10^{-4} (0.0348)
a22	0.5165 (0.1621)	0.4997 (0.1115)	0.5056 (0.0813)	0.4981 (0.0660)	0.5020 (0.0463)
b11	0.9365 (0.1212)	0.9730 (0.0477)	0.9832 (0.0250)	0.9918 (0.0150)	0.9967 (0.0058)
b12	-0.0081 (0.1369)	0.0034 (0.0764)	0.0018 (0.0476)	3.5650×10^{-4} (0.0301)	1.2764×10^{-4} (0.0189)
b21	-0.0030 (0.1148)	6.2183×10^{-5} (0.0470)	4.0212×10^{-4} (0.0201)	4.5382×10^{-4} (0.0095)	1.1201×10^{-4} (0.0032)
b22	0.4689 (0.1262)	0.4879 (0.0767)	0.4948 (0.0445)	0.4971 (0.0295)	0.4993 (0.0187)
τ_1	-3.1458 (0.5528)	-3.1520 (0.4364)	-3.0948 (0.3118)	-3.0504 (0.2348)	-3.0302 (0.1772)
τ_2	2.7868 (0.4679)	2.8974 (0.3243)	2.9752 (0.2487)	2.9808 (0.1931)	2.9880 (0.1352)
τ_3	-1.9672 (0.5357)	-2.0274 (0.4211)	-2.0262 (0.3488)	-2.0308 (0.2631)	-2.0228 (0.1850)
τ_4	1.9840 (0.4961)	2.0136 (0.4245)	2.0386 (0.3488)	2.0334 (0.2622)	2.0228 (0.1850)
β_1	-0.0048 (0.1479)	0.0016 (0.0748)	3.8148×10^{-4} (0.0351)	4.4954×10^{-4} (0.0169)	2.2408×10^{-4} (0.0063)
β_2	-0.0037 (0.0711)	3.9443×10^{-4} (0.0428)	6.1118×10^{-4} (0.0245)	2.5751×10^{-4} (0.0124)	1.9799×10^{-4} (0.0056)

4.2 Estimation with Unknown Thresholds

We use the same data generating processes as in the last subsection. Since we don't know the true thresholds, we estimate them jointly with the autoregressive coefficients. We use a grid search procedure to select the estimates for the true thresholds. We always use the threshold regime $D = [\tau_1 \ \tau_2] \times [\tau_3 \ \tau_4] = [-3 \ 3] \times [-2 \ 2]$, we search τ_1 within -4 and -2 with 10 grid points, search τ_2 within 2 and 4 with 10 grid points, search τ_3 within -3 and -1 with 10 grid points,

search τ_4 within 1 and 3 with 10 grid points³. The thresholds and autoregressive coefficients are chosen such that the sum of square Euclidean norm of the residuals is minimized. I.e.,

$$(\widehat{A}, \widehat{B}, \widehat{\tau}) = \arg \min_{A, B, \tau} \sum_{t=1}^T [\|x_t - Ax_{t-1}\|^2 1(x_{t-1} \in D_\tau) + \|x_t - Bx_{t-1}\|^2 1(x_{t-1} \in D_\tau^c)].$$

The results for the case where $A = \begin{pmatrix} 1/2 & 0 \\ 0 & 1/2 \end{pmatrix}$, $B = \begin{pmatrix} 1 & 0 \\ 0 & 1/2 \end{pmatrix}$ are summarized in Table 4 above.

The results for the case where $A = \begin{pmatrix} 1/2 & 0 \\ 0 & 1/2 \end{pmatrix}$, $B = \begin{pmatrix} 3/4 & 1/4 \\ 1/4 & 3/4 \end{pmatrix}$ are summarized in Table 5 below.

Table 5: Mean and STD of Estimators

T	200	400	800	1600	3200
a11	0.4851 (0.1208)	0.4833 (0.0987)	0.4905 (0.0771)	0.4951 (0.0569)	0.4965 (0.0424)
a12	-0.0132 (0.1706)	-0.0076 (0.1512)	-0.0009 (0.1036)	-0.0019 (0.0831)	-0.0016 (0.0520)
a21	-0.0145 (0.1400)	-0.0098 (0.1174)	-0.0069 (0.0757)	-0.0013 (0.0589)	-0.0004 (0.0452)
a22	0.4776 (0.1702)	0.4930 (0.1317)	0.4908 (0.0955)	0.4942 (0.0855)	0.4965 (0.0560)
b11	0.6935 (0.1181)	0.7288 (0.0623)	0.7418 (0.0402)	0.7470 (0.0311)	0.7475 (0.0169)
b12	0.2604 (0.0902)	0.2501 (0.0534)	0.2481 (0.0317)	0.2485 (0.0215)	0.2501 (0.0149)
b21	0.2321 (0.1250)	0.2407 (0.0584)	0.2491 (0.0343)	0.2482 (0.0219)	0.2496 (0.0151)
b22	0.7284 (0.0833)	0.7368 (0.0588)	0.7409 (0.0391)	0.7469 (0.0312)	0.7480 (0.0180)
τ_1	-3.1554 (0.6576)	-3.1350 (0.5338)	-3.0922 (0.4476)	-3.0576 (0.3680)	-3.0422 (0.3102)
τ_2	2.7104 (0.4213)	2.8227 (0.3842)	2.9386 (0.3619)	2.9664 (0.3029)	2.9908 (0.2598)
τ_3	-1.9982 (0.4661)	-2.0250 (0.3334)	-2.0250 (0.2672)	-2.0052 (0.2032)	-2.0126 (0.1572)
τ_4	1.9912 (0.3953)	1.9994 (0.3269)	2.0118 (0.2678)	2.0048 (0.2008)	2.0052 (0.1320)
β_1	0.8744 (0.1910)	0.9164 (0.1186)	0.9611 (0.0650)	0.9799 (0.0422)	0.9910 (0.0333)
β_2	0.6509 (0.3037)	0.7703 (0.2463)	0.8859 (0.1639)	0.9499 (0.0878)	0.9820 (0.0438)

The results for the case where $A = \begin{pmatrix} 3/4 & 1/8 \\ 1 & 1/2 \end{pmatrix}$, $B = \begin{pmatrix} 3/4 & 1/4 \\ 1/4 & 3/4 \end{pmatrix}$ are summarized in Table 6 below.

³Finer grid points may help to improve the performance of estimations. However, the computational burden is larger.

Table 6: Mean and STD of Estimators

T	200	400	800	1600	3200
a11	0.7477 (0.3931)	0.7370 (0.2325)	0.7469 (0.1816)	0.7502 (0.1461)	0.7432 (0.1119)
a12	0.1312 (0.4943)	0.1159 (0.3425)	0.1386 (0.2730)	0.1248 (0.1982)	0.1220 (0.1484)
a21	1.0482 (0.4027)	1.0368 (0.2507)	1.0354 (0.1801)	1.0214 (0.1414)	1.0151 (0.1178)
a22	0.5249 (0.4988)	0.4961 (0.3573)	0.4884 (0.2454)	0.5060 (0.1902)	0.4993 (0.1502)
b11	0.7375 (0.0495)	0.7445 (0.0329)	0.7476 (0.0228)	0.7492 (0.0155)	0.7493 (0.0110)
b12	0.2556 (0.0484)	0.2518 (0.0326)	0.2501 (0.0227)	0.2497 (0.0154)	0.2501 (0.0110)
b21	0.2529 (0.0479)	0.2498 (0.0325)	0.2502 (0.0226)	0.2499 (0.0160)	0.2505 (0.0107)
b22	0.7397 (0.0465)	0.7462 (0.0320)	0.7476 (0.0225)	0.7490 (0.0161)	0.7489 (0.0107)
τ_1	-3.3992 (0.4957)	-3.3082 (0.5046)	-3.2398 (0.4606)	-3.1708 (0.4219)	-3.1196 (0.3502)
τ_2	2.6464 (0.4679)	2.7156 (0.4172)	2.7766 (0.3926)	2.8564 (0.3407)	2.9024 (0.2979)
τ_3	-2.2136 (0.5325)	-2.1450 (0.4770)	-2.0912 (0.4172)	-2.0916 (0.3787)	-2.0300 (0.2917)
τ_4	1.9288 (0.5402)	1.9714 (0.4730)	1.9900 (0.4360)	2.0162 (0.3610)	2.0346 (0.3076)
β_1	0.9685 (0.0400)	0.9790 (0.0234)	0.9874 (0.0140)	0.9929 (0.0082)	0.9958 (0.0046)
β_2	0.9680 (0.0393)	0.9788 (0.0234)	0.9873 (0.0141)	0.9929 (0.0082)	0.9958 (0.0046)

Comparing these tables with tables in the last subsection, we can see that the estimation of the unknown thresholds plays little role for estimating the autoregressive coefficients in the outside regime and the cointegrating coefficients. The effect of estimated thresholds is larger when the estimating the autoregressive coefficients in the middle regime, but the increases in standard deviations are not very dramatic. This increase is to be expected since typically the middle regime contains fewer observations and hence is more sensitive to misplacement of observations. We can also see that the convergence rates for the threshold estimators are very slow which is quite different from stationary threshold models. Chan (1993) and Chan and Tsay (1998) show that in the stationary case, the estimators for the threshold variables converge with rate T or \sqrt{T} depending on whether the model is “continuous” or not when the data are stationary. However, we find that when nonstationarity is involved, the convergence rate is much slower. This is reasonable because of the slow convergence in the middle regime. A similar phenomenon is also found in Gao et al. (2013) in the threshold unit root model.

5 Relationship between Federal Funds Rate and the T-bill Rate

In this section, we will apply the proposed model to study the dynamic relationship between two key US market interest rates: the Federal funds rate (FF) and the 3-month Treasury bill

rate (TB). The importance of these two interest rates is widely recognized. The Federal Reserve implements its monetary policy by adjusting the FF (see, e.g. Rudebusch 1995) while the 3-month T-bill rate is regarded as default-free rate in the money market and is usually regarded as a proxy of instantaneous interest rate in research (see, e.g. Anderson and Lund 1997).

5.1 Background Introduction

It is well known that interest rates tend to move together. According to the famous expectation hypothesis, the following condition holds:

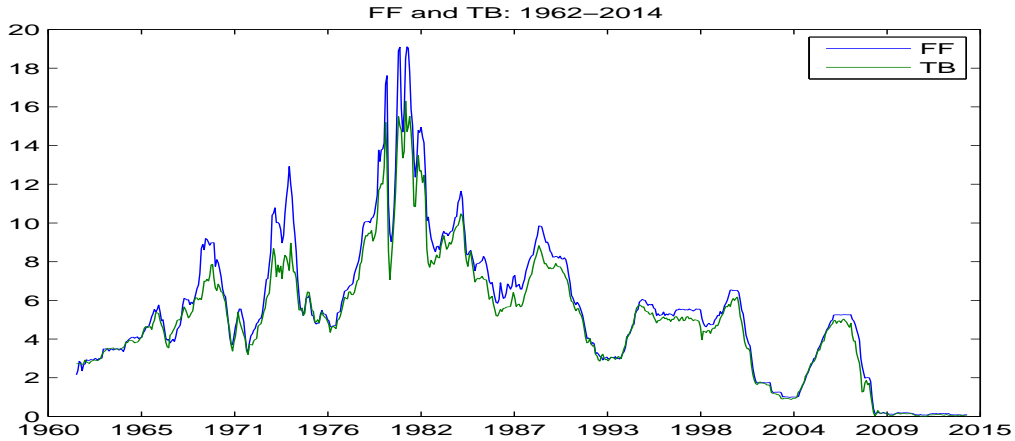
$$R(k, t) = \frac{1}{k} \left[\sum_{j=1}^k E_t(R(1, t + j - 1)) \right] + L(k, t), \quad (5.1)$$

where $R(k, t)$ is the time- t yield with maturity k , E_t is the conditional expectation operator and $L(k, t)$ is the term premia accounting for risk premia and liquidity premia.

Thus, under the assumption of costless and instantaneous portfolio adjustments, equation (5.1) implies that FF and TB move together. At least dating back to Engle and Granger (1987) and Stock and Watson (1988), cointegration models or the corresponding error correction models have been used to study the relationships of interest rates. A detailed analysis of cointegration relationships between T-bill rates with different maturities can be found in Hall et al. (1992). Generally speaking, the empirical evidence shows that these rates co-move in the long run.

However, as pointed out by Anderson (1997) and Bec and Rahbek (2004), existence of transaction costs makes the continuous adjustments impossible which results in nonlinear discontinuous adjustments. Such arguments assert that rational investors will buy and sell bills in an attempt to profit on any yield spread which are not justified by risk, liquidity or other considerations such that only when the yield spread is large enough to cover the transaction costs, yield adjustments occur. Thus, in these two papers, threshold nonlinear error correction models are used such that the model is dependent on a switch variable which is defined as yield spread, which is stationary because they further impose that the cointegrating vector is $[1 - 1]^\tau$ for two regimes depending on whether $|R(k, t) - R(1, t)| > \tau$ or not.

Meanwhile, transaction costs are not the only source of market friction and other frictions such as liquidity constraints may also affect the dynamic behaviours of FF and TB. Furthermore, it is well known that FF is controlled directly by the Federal Reserve as a monetary tool such that it should also contribute to the nonlinearity of the interest rate dynamics because the Federal Reserve may prefer FF to stay in a medium regime rather than being too high or too low. Using a regime switching error correction model, Clarida et al. (2006) find that the interest rate dynamics separate into high interest rate and low interest rate regimes. In summary, we can conclude that both the yield spread and interest rates themselves (especially FF) contribute to the nonlinearity of interest rate dynamics. Using mathematical language, it is possible that the nonlinearity is introduced in the following way: $\tau_1 \leq |FF| \leq \tau_2$ because of the control of the Federal Reserve and $\lambda_1 \leq |FF - \beta TB| \leq \lambda_2$, where β is the cointegrating coefficient



such that $FF - \beta TB$ is the equilibrium error because of the existence of transaction costs. Or equivalently, $\tau_1 \leq |FF| \leq \tau_2$ and $\tau_3 \leq |TB| \leq \tau_4$, because TB and FF series are cointegrated. Thus the model proposed in this paper is useful in studying the relationship between FF and TB. It is also more general and can be used to “test” threshold error correction models, which come out as a special case.

5.2 Empirical Analysis

We use monthly data of TB and FF from the website of Federal Reserve Bank of St. Louis. The sample period is from January, 1962 to May, 2014 and the sample size is 629. The ADF test suggests that both series contain unit roots with p-values for TB and FF being respectively 0.2807 and 0.2648. The following graph shows the time series plot of the two series.

The time series contain several important periods in the U.S.: the stagflation period in the 1970s with both high unemployment and high inflation; the extraordinary high interest rate period in the early 1980s when the Federal Reserve chairman Paul Volcker tried to wring inflation out of the economy by raising interest rates; and the recent years (after 2008) when the Federal Reserve used QE (Quantitative Easing) policy after the financial crisis. From the graph, it seems that the two series move with each other.

We estimate the autoregression matrices by OLS conditional on the threshold variables and the threshold variables by minimizing the sum of Euclidean norm of the residual over some grid points⁴.

The estimation results for the autoregressive coefficients (denote c_{11} , c_{12} , c_{21} , c_{22} as the cor-

⁴We search τ_1 within $[1 \ 5]$, τ_2 within $[\tau_1 + 2 \ \tau_1 + 12]$, τ_3 within $[1 \ 5]$, τ_4 within $[\tau_3 + 2 \ \tau_3 + 12]$, the length of consecutive grid points is 0.2.

responding elements) and eigenvalues and eigenvectors of the estimated autoregressive matrices are listed in following table.

Table 7: Estimation Results

	c_{11}	c_{12}	c_{21}	c_{22}	eig1	eig2
Full Sample	0.7290	0.014	-0.0513	1.0543	0.9965	0.7868
					([-0.7480 - 0.6637] $^{\tau}$)	([-0.9821 - 0.1883] $^{\tau}$)
Middle Regime	0.8667	0.1571	0.0334	0.9679	1.0057	0.8290
					([-0.7490 - 0.6626] $^{\tau}$)	([-0.9723 0.2335] $^{\tau}$)
Outside Regime	0.3818	0.6447	-0.2468	1.2317	0.9532	0.6602
					([-0.7483 - 0.6633] $^{\tau}$)	([-0.9180 - 0.2335] $^{\tau}$)

The estimated thresholds are respectively: $\hat{\tau}_1 = 4$, $\hat{\tau}_2 = 16$, $\hat{\tau}_3 = 1.8$, $\hat{\tau}_4 = 13.8^5$. We also find that for the ordinary VAR model, the sum of Euclidean norm of the residual is 295.7426, while for the threshold model, it is 265.9568, which decrease about 10 percent. Anderson (1997) finds that using a threshold error correction model, the decrease in the coefficient of determination (R^2) compared with linear error correction model is about 8 percent⁶. Even though no formal testing procedure was proposed, we regard this as evidence in favour of a threshold model for interest rates.

From the table, we see that: (1) For the estimated matrices in two regimes, the larger eigenvalues are close to 1, while the others are obviously less than 1. This suggests that in both regimes, the two series are cointegrated. (2) The eigenvectors associated with the eigenvalues close to 1 are almost the same, which suggests that in both regimes, the cointegrating vectors are almost the same. In this sense, nonlinear error correction models, which assume a common cointegrating vector for all the regimes is suitable to model the relationship of FF and TB. The difference between our model and the nonlinear error correction models when the cointegrating vectors are the same for different regimes is that we use different threshold variables. In the nonlinear error correction models, the threshold variables are assumed to be stationary, while in our model, we do not impose this restriction. As we discussed in Section 5.1, we think that our model can account for more types of market frictions. (3) As we discussed in the Introduction, the eigenvector associated with eigenvalue less than 1 is *parallel to* the short run adjustment vector. We can see that in each regime, the response of FF is larger than the response of TB to deviation of equilibrium⁷ because the absolute size is larger. This suggests that the Federal Reserve tends to adjust FF as a response to the market interest rates. And we also find that the signs of the short run adjustment coefficient of TB to the equilibrium error are different in the two regimes which indicates different behaviors of FF in each regime.

⁵The number of observations in the middle regime is 414.

⁶Of course, this argument is not very precise because we use different data set.

⁷Similarly, Sarno and Thornton (2003) find that most of the adjustment toward long run equilibrium occurs through FF rather than TB.

6 Conclusion

In this paper, we have shown that the processes generated by the class of threshold VAR cointegrated models as specified in (2.1) are 1/2-null recurrent. We then establish an asymptotic theory for autoregressive coefficient estimators in both regimes and the cointegrating coefficient estimators. Applying this model to examine the relationship of Federal funds rate and 3-month T-bill rate, we find that the cointegrating vector is same for the two regimes while the short run loading coefficients are different for these two series.

This paper can be extended in several ways: we can extend the bivariate model to the multivariate models which are still 1/2-null recurrent; as usually discussed in VAR cointegrated models, we can also extend the model to include drift terms and higher order lagged variables; also, the specification of the thresholds can be different. For example, we can allow one cointegrating regime in one direction, the other cointegrating regime in the other direction; finally, as we have discussed in the introduction, specification testing problems need to be considered. These issues are currently investigated by the authors.

Appendix A

To make this paper self-contained, in this Appendix, we introduce necessary notions of Markov chain properties. They are fundamental for deriving the asymptotic theory in this paper. We refer to Karlsen and Tjøstheim (2001), Karlsen et al. (2007) and Nummelin (1984) for more comprehensive treatments.

Let $\{X_t, t \geq 0\}$ be a ϕ -irreducible Markov chain on the state space (E, \mathcal{E}) with transition probability P . This means that for any set $A \in \mathcal{E}$ with $\phi(A) > 0$, we have $\sum_{t=1}^{\infty} P^t(x, A) > 0$ for all $x \in E$. In this paper, we take $E \subseteq \mathbb{R}^2$. We further assume that the ϕ -irreducible Markov chain $\{X_t\}$ is Harris recurrent.

Definition 1. A Markov chain $\{X_t\}$ is Harris recurrent if, given a neighborhood N_x of x with $\phi(N_x) > 0$, $\{X_t\}$ returns to N_x with probability one, for any $x \in E$.

The Harris recurrence allows one to construct a split chain, which decomposes the partial sum of functions of $\{X_t\}$ into blocks of independent and identically distributed (i.i.d.) parts and two asymptotically negligible remaining parts. Let τ_k be the regeneration times, T the number of observations and $n(T)$ the number of regenerations as in Karlsen and Tjøstheim (2001)⁸.

For the process $\{G(X_t) : t \geq 0\}$, defining

$$U_k = \begin{cases} \sum_{t=0}^{\tau_0} G(X_t), & k = 0 \\ \sum_{t=\tau_{k-1}+1}^{\tau_k} G(X_t), & 1 \leq k \leq n(T), \\ \sum_{t=\tau_{N(n)}+1}^T G(X_t), & k = n(T) + 1 \end{cases}$$

where $G(\cdot)$ is a real function defined on \mathbb{R}^d , then we have

$$S_n(G) = \sum_{t=0}^T G(X_t) = U_0 + \sum_{k=1}^{n(T)} U_k + U_{n(T)+1}. \quad (\text{A.1})$$

⁸They use the notation $T(n)$ instead of $n(T)$.

From Nummelin (1984), we know that $\{U_k, k \geq 1\}$ is a sequence of i.i.d. random variables, and U_0 and $U_{n(T)+1}$ converge to zero almost surely (a.s.) when they are divided by the number of regenerations $n(T)$ (using Lemma 3.2 in Karlsen and Tjøstheim 2001).

The distribution and size of the number of regenerations $n(T)$ have no a priori known structure but fully depend on the underlying process $\{X_t\}$. To obtain a specific rate of $n(T)$ in our asymptotic theory for the null recurrent process, we next impose some restrictions on the tail behavior of the distribution of the recurrence times of the Markov chain.

Definition 2. A Markov chain $\{X_t\}$ is β -null recurrent if there exist a small nonnegative function f , an initial measure λ , a constant $\beta \in (0, 1)$, and a slowly varying function $L_f(\cdot)$ such that

$$E_\lambda\left(\sum_{i=1}^T f(X_i)\right) \sim \frac{1}{\Gamma(1+\beta)} T^\beta L_f(T), \quad (\text{A.2})$$

as $T \rightarrow \infty$, where E_λ stands for the expectation with initial distribution λ and $\Gamma(1+\beta)$ is the Gamma function with parameter $1+\beta$.

The definition of a small function f in the above definition can be found in some existing literature (c.f. P. 15 in Nummelin 1984). Assuming β -null recurrence restricts the tail behavior of the recurrence time of the process to be a regularly varying function. In fact, for all small functions f , by Lemma 3.1 in Karlsen and Tjøstheim (2001), we can find an $L_s(\cdot)$ such that (A.2) holds for the β -null recurrent Markov chain with $L_f(\cdot) = \pi_s(f)L_s(\cdot)$, where π_s is an invariant measure of the Markov chain $\{X_t\}$, $\pi_s(f) = \int f(x)\pi_s(dx) = \int f(x)p_s(x)dx$, where s is the small function in the minorization inequality (3.4) of Karlsen and Tjøstheim (2001) and $p_s(x)$ is the invariant density. Letting $L_s(T) = L_f(T)/(\pi_s(f))$ and following the argument in Karlsen and Tjøstheim (2001), we may show that the regeneration number $n(T)$ of the β -null recurrent Markov chain $\{X_t\}$ has the following asymptotic distribution

$$\frac{n(T)}{T^\beta L_s(T)} \rightarrow_d M_\beta(1), \quad (\text{A.3})$$

where $M_\beta(1)$ is the Mittag-Leffler distribution with parameter β . Some examples of univariate β -null recurrent process are given in Karlsen and Tjøstheim (2001) while some examples of multivariate β -null recurrent process are given in Cai and Tjøstheim (2014). In this paper, we show that the bivariate cointegrated process is 1/2-null recurrent process and $L_s(T)$ is a constant. Thus, the mathematical techniques in this appendix are applicable.

Appendix B

Proof of Lemma 1. For (1), using (3.3) and continuous mapping theorem,

$$\frac{1}{T^2} \sum_{t=1}^T y_{1,t-1}^2 \rightarrow_d \int_0^1 P^2(r) dr.$$

For (2), we have

$$\frac{1}{T} \sum_{t=1}^T y_{2,t}^2 = \frac{1}{T} \sum_{t=1}^T \left(\sum_{i=1}^t c^{t-i} u_{2,i} + \sum_{i=1}^t c^{t-i} \varepsilon_{2,i} + c^t y_{2,0} \right)^2$$

and because $y_{2,0} = 0$,

$$\begin{aligned} &= \frac{1}{T} \sum_{t=1}^T \left(\sum_{i=1}^t c^{t-i} u_{2,i} \right)^2 + \frac{1}{T} \sum_{t=1}^T \left(\sum_{i=1}^t c^{t-i} \varepsilon_{2,i} \right)^2 \\ &\quad + \frac{2}{T} \sum_{t=1}^T \left(\sum_{i=1}^t c^{t-i} u_{2,i} \right) \left(\sum_{i=1}^t c^{t-i} \varepsilon_{2,i} \right). \end{aligned}$$

We will first discuss the term $\frac{1}{T} \sum_{t=1}^T \left(\sum_{i=1}^t c^{t-i} \varepsilon_{2,i} \right)^2$. According to Pham and Tran (1985), we know that $\left\{ \left(\sum_{i=1}^t c^{t-i} \varepsilon_{2,i} \right)^2 \right\}$ is a geometrically α -mixing sequence. Thus, according to Theorem 3.1 of Mcleish (1975), we have

$$\frac{1}{T} \sum_{t=1}^T \left(\sum_{i=1}^t c^{2(t-i)} \varepsilon_{2,i}^2 \right) \rightarrow_p \lim_{T \rightarrow \infty} \frac{1}{T} \left(\sum_{t=1}^T \sum_{i=0}^{t-1} c^{2i} \right) \sigma_{\varepsilon_2}^2,$$

where $\sigma_{\varepsilon_2}^2 = E(\varepsilon_{2,1}^2)$. And by the Toeplitz's Lemma, we have:

$$\lim_{T \rightarrow \infty} \frac{1}{T} \left(\sum_{t=1}^T \sum_{i=0}^{t-1} c^{2i} \right) \sigma_{\varepsilon_2}^2 = \left(\sum_{i=0}^{\infty} c^{2i} \right) \sigma_{\varepsilon_2}^2.$$

Thus

$$\frac{1}{T} \sum_{t=1}^T \left(\sum_{i=1}^t c^{2(t-i)} \varepsilon_{2,i}^2 \right) \rightarrow_p \left(\sum_{i=0}^{\infty} c^{2i} \right) \sigma_{\varepsilon_2}^2. \quad (\text{B.1})$$

Next we discuss the term:

$$\begin{aligned} \frac{1}{T} \sum_{t=1}^T \left(\sum_{i=1}^t c^{t-i} u_{2,i} \right)^2 &= \frac{1}{T} \sum_{t=1}^T \left(\sum_{j=1}^t \sum_{i=1}^t c^{t-i} u_{2,i} c^{t-j} u_{2,j} \right) = 1/T \sum_{t=1}^T \sum_{i=1}^t \sum_{j=1}^t (c^i c^j u_{2,t-i} u_{2,t-j}) \\ &= \frac{1}{T} \sum_{t=1}^T \sum_{j=1}^{\infty} \sum_{i=1}^{\infty} (c^i c^j u_{2,t-i} u_{2,t-j} 1(1 \leq i \leq t) 1(1 \leq j \leq t)) \\ &= \frac{1}{T} \sum_{t=1}^T \sum_{j=1}^N \sum_{i=1}^N (c^i c^j u_{2,t-i} u_{2,t-j} 1(1 \leq i \leq t) 1(1 \leq j \leq t)) \\ &\quad + \frac{1}{T} \sum_{t=1}^T \sum_{j=N+1}^{\infty} \sum_{i=1}^N (c^i c^j u_{2,t-i} u_{2,t-j} 1(1 \leq i \leq t) 1(1 \leq j \leq t)) \\ &\quad + \frac{1}{T} \sum_{t=1}^T \sum_{i=1}^N \sum_{j=N+1}^{\infty} (c^i c^j u_{2,t-i} u_{2,t-j} 1(1 \leq i \leq t) 1(1 \leq j \leq t)) \\ &\quad + \frac{1}{T} \sum_{t=1}^T \sum_{i=N+1}^{\infty} \sum_{j=N+1}^{\infty} (c^i c^j u_{2,t-i} u_{2,t-j} 1(1 \leq i \leq t) 1(1 \leq j \leq t)) \\ &= B_1 + B_2 + B_3 + B_4. \end{aligned}$$

For any fixed N , we know that

$$\begin{aligned} &\frac{1}{T} \sum_{j=1}^N \sum_{i=1}^N c^i c^j \sum_{t=1}^T (u_{2,t-i} u_{2,t-j} 1(1 \leq i \leq t) 1(1 \leq j \leq t)) \\ &= \frac{1}{T} \times n(T) \sum_{j=1}^N \sum_{i=1}^N c^i c^j \times 1/n(T) \sum_{t=1}^T (u_{2,t-i} u_{2,t-j} 1(1 \leq i \leq t) 1(1 \leq j \leq t)). \end{aligned}$$

$$\leq C_1 \times 1/T \times n(T) \sum_{j=1}^N \sum_{i=1}^N c^i c^j \times 1/n(T) \sum_{t=1}^T (u_{2,t-i} 1(1 \leq i \leq t))$$

for some constant C_1 , where we have used the compactness of the set C and the definition of $u_{2,t}$.

By Karlsen and Tjøstheim (2001), we have $\frac{1}{n(T)} \sum_{t=1}^T (u_{2,t-i} 1(1 \leq i \leq t)) = O_P(1)$. Thus, $B_1 = O_P(n(T)/T) = o_P(1)$. Moreover,

$$\frac{1}{T} \sum_{t=1}^T \sum_{j=N+1}^{\infty} \sum_{i=1}^N (c^i c^j u_{2,t-i} u_{2,t-j} 1(1 \leq i \leq t) 1(1 \leq j \leq t)) \leq C_2 \sum_{j=N+1}^{\infty} \sum_{i=1}^N c^i c^j = o(1),$$

for some constant C_2 as $N \rightarrow \infty$ and hence $B_2 = o_P(1)$.

The same argument can be used for B_3 and B_4 and thus

$$B_1 + B_2 + B_3 + B_4 = o_P(1). \quad (\text{B.2})$$

By Cauchy-Schwarz inequality and using results (B.1) and (B.2), we have

$$\frac{2}{T} \sum_{t=1}^T \left(\sum_{i=1}^t c^{t-i} u_{2,i} \right) \left(\sum_{i=1}^t c^{t-i} \varepsilon_{2,i} \right) = o_P(1). \quad (\text{B.3})$$

Combining results of (B.1) (B.2) and (B.3), we have

$$\frac{1}{T} \sum_{t=1}^T y_{2,t}^2 \rightarrow_p \left(\sum_{i=0}^{\infty} c^{2i} \right) \sigma_{\varepsilon_2}^2 = \frac{1}{1-c^2} \sigma_{\varepsilon_2}^2.$$

(3) Finally, turning to the off-diagonal terms, we have

$$\sum_{t=1}^T y_{1,t} y_{2,t} = \sum_{t=1}^T \left(\sum_{j=1}^t u_{1,j} + \varepsilon_{1,j} \right) \left(\sum_{i=1}^t c^{t-i} u_{2,i} + c^{t-i} \varepsilon_{2,i} \right) \equiv I_1 + I_2 + I_3 + I_4,$$

where $I_1 = \sum_{t=1}^T \left(\sum_{j=1}^t u_{1,j} \right) \left(\sum_{i=1}^t c^{t-i} u_{2,i} \right)$, $I_2 = \sum_{t=1}^T \left(\sum_{j=1}^t u_{1,j} \right) \left(\sum_{i=1}^t c^{t-i} \varepsilon_{2,i} \right)$, $I_3 = \sum_{t=1}^T \left(\sum_{j=1}^t \varepsilon_{1,j} \right) \left(\sum_{i=1}^t c^{t-i} u_{2,i} \right)$ and $I_4 = \sum_{t=1}^T \left(\sum_{j=1}^t \varepsilon_{1,j} \right) \left(\sum_{i=1}^t c^{t-i} \varepsilon_{2,i} \right)$.

For I_1 , we have

$$I_1^2 = \left(\sum_{t=1}^T \left(\sum_{j=1}^t u_{1,j} \right) \left(\sum_{i=1}^t c^{t-i} u_{2,i} \right) \right)^2 \leq \sum_{t=1}^T \left(\sum_{j=1}^t u_{1,j} \right)^2 \sum_{t=1}^T \left(\sum_{i=1}^t c^{t-i} u_{2,i} \right)^2.$$

In view of $\sum_{t=1}^T \left(\sum_{j=1}^t u_{1,j} \right)^2 = O_P(T^2)$ and $\sum_{t=1}^T \left(\sum_{i=1}^t c^{t-i} u_{2,i} \right)^2 = o_P(T)$ by (B.2), we have

$$I_1 = o_P(T^{3/2}). \quad (\text{B.4})$$

For I_4 , by Hansen (1992) we obtain

$$I_4 = O_P(T). \quad (\text{B.5})$$

For I_3 , using results (B.4) (B.5) and by Cauchy-Schwarz inequality,

$$I_3^2 \leq \sum_{t=1}^T \left(\sum_{j=1}^t \varepsilon_{1,j} \right)^2 \sum_{t=1}^T \left(\sum_{i=1}^t c^{t-i} u_{2,i} \right)^2 \leq O_P(T^2) o_P(T) = o_P(T^3).$$

For I_2 , we have

$$I_2 = \sum_{t=1}^T \left(\sum_{j=1}^t u_{1,j} \right) \left(\sum_{i=1}^t c^{t-i} \varepsilon_{2,i} \right) \equiv \sqrt{T} \sum_{t=1}^T U_{1,Tt} \kappa_t,$$

where $U_{1,Tt} = \frac{1}{\sqrt{T}} \sum_{i=1}^t u_{1,i}$ and $\kappa_t = (\sum_{i=1}^t c^{t-i} \varepsilon_{2,i})$.

Observe that $U_{1,Tt} \rightarrow_d \mu_1 M_{1/2}([t/T])$, where μ_1 is defined in Section 3.1 and $M_{1/2}(\cdot)$ is a Mittag-Leffler process, which is almost surely continuous. Meanwhile, we have

$$\sup_t E|E(\kappa_t | \mathcal{F}_{t-m})| \rightarrow 0,$$

as $m \rightarrow \infty$, where \mathcal{F}_{t-m} is the σ -algebra generated by $\{\varepsilon_{t-m}, \varepsilon_{t-m-1}, \dots, \varepsilon_1\}$. Thus applying Theorem 3.3 of Hansen (1992), we have

$$I_2 = o_P(T^{3/2}). \quad (\text{B.7})$$

Combining the results (B.4) (B.5) (6) and (B.7), we have proved (3) in Lemma 1. \square

Proof of Lemma 2. (1) In order to show this, we first notice that $\{\varepsilon_t\}$ is a martingale difference sequence (mds) with respect to filtration \mathcal{F}_t being a σ -algebra generated by $\{\varepsilon_t, \varepsilon_{t-1}, \dots, \varepsilon_1\}$. Thus, according to Theorem 2.1 of Hansen (1992), we have

$$\frac{1}{T} \sum y_{1,t-1} \varepsilon_t \rightarrow_d \int_0^1 P(r) dB(r),$$

where $B(r) = [B_1(r), B_2(r)]^\tau$ is a two-dimensional Brownian motion with covariance matrix Σ_1 .

(2) Note also that $y_{2,t-1} \varepsilon_t$ is two-dimensional mds and define

$$\xi_{1,t} \equiv \frac{1}{\sqrt{T}} y_{2,t-1} \varepsilon_{1,t}$$

and the partial sum process $S_{1,t} \equiv \sum_{i=1}^t \xi_{1,i} = \sum_{i=1}^t \frac{1}{\sqrt{T}} y_{2,i-1} \varepsilon_{1,i}$.

We then have

$$\sum_{t=1}^T E[\xi_{1,t}^2 | \mathcal{F}_{t-1}] = \frac{1}{T} \sum_{t=1}^T y_{2,t-1}^2 E[\varepsilon_{1,t}^2 | \mathcal{F}_{t-1}] = \frac{1}{T} \sum_{t=1}^T y_{2,t-1}^2 \sigma_{\varepsilon_1}^2 \rightarrow_p \left(\sum_{i=0}^{\infty} c^{2i} \right) \sigma_{\varepsilon_2}^2 \sigma_{\varepsilon_1}^2$$

by (2) of Lemma 1.

To prove a martingale central limit theorem, we need to verify the ‘‘conditional Liapounov condition’’, i.e., $\sum_{t=1}^T E[\xi_{1,t}^{2+2\delta} | \mathcal{F}_{t-1}] = o_P(1)$ for some $\delta > 0$. This requires us to use the fact that $\varepsilon_{1,t}$ is independent of \mathcal{F}_{t-1} and then to verify

$$\sum_{t=1}^T E[\xi_{1,t}^4 | \mathcal{F}_{t-1}] \leq C_1 \frac{1}{T^2} \sum_{t=1}^T y_{2,t-1}^4 = O_P(T^{-1}),$$

for some constant C_1 .

According to the definition of $y_{2,t-1}$ and Minkowski inequality

$$\sum_{t=1}^T y_{2,t-1}^4 \leq \left\{ \left[\sum_{t=1}^T \left[\sum_{i=1}^{t-1} (c^{t-i} u_{2,i})^4 \right]^{1/4} + \left[\sum_{t=1}^T \left[\sum_{i=1}^{t-1} (c^{t-i} \varepsilon_{2,i})^4 \right]^{1/4} \right]^4 \right\}.$$

As we have shown in the proof of (2) of Lemma 1 $\{\sum_{i=1}^{t-1} c^{t-i} \varepsilon_{2,i}\}$ is geometrically α -mixing with finite fourth moment, so that $\sum_{t=1}^T E(\sum_{i=1}^{t-1} c^{t-i} \varepsilon_{2,i})^4 \leq C_2 T$ for some constant C_2 .

By the boundedness of $u_{2,t}$ uniformly in t and then using the same procedure as in the proof of (2) of Lemma 1,

$$\sum_{t=1}^T \left[\sum_{i=1}^{t-1} c^{t-i} u_{2,i} \right]^4 \leq \max_{t=1, \dots, T} \left(\sum_{i=1}^{t-1} c^{t-i} u_{2,i} \right)^2 \sum_{t=1}^T \left[\left(\sum_{i=1}^{t-1} c^{t-i} u_{2,i} \right)^2 \right]$$

$$\leq C_3 \sum_{t=1}^T \left[\left(\sum_{i=1}^{t-1} c^{t-i} u_{2,i} \right)^2 \right] = O_P(T).$$

Thus, we have shown $\sum_{t=1}^T E \left[\xi_{1,t}^4 | \mathcal{F}_{t-1} \right] = O_P\left(\frac{1}{T}\right)$. Applying the classical central limit theorem for martingales (cf. Gaenssler and Joos 1992), we have

$$S_{1,T} \rightarrow_d N\left(0, \left(\sum_{i=0}^{\infty} c^{2i} \right) \sigma_{\varepsilon_2}^2 \sigma_{\varepsilon_1}^2\right).$$

Similarly for $S_{2,t} = \sum_{i=1}^t \frac{1}{\sqrt{T}} y_{2,i-1} \varepsilon_{2,i}$, we have

$$S_{2,T} \rightarrow_d N\left(0, \left(\sum_{i=0}^{\infty} c^{2i} \right) \sigma_{\varepsilon_2}^2 \sigma_{\varepsilon_2}^2\right).$$

Notice that $\sigma_{\varepsilon_1}^2$ and $\sigma_{\varepsilon_2}^2$ are diagonal terms of Σ_1 . Applying the Cramer-Wold device, we verify (2) of Lemma 2. □

Proof of Lemma 3. By (3.1)

$$Qx_t = A_1 Qx_{t-1} 1(x_{t-1} \in D) + B_1 Qx_{t-1} 1(x_{t-1} \in D^c) + \varepsilon_t.$$

Writing this in terms of $u_t = (A_1 Q - B_1 Q)x_{t-1} 1(x_{t-1} \in D)$,

$$\begin{cases} q_{11}x_{1,t} + q_{12}x_{2,t} = q_{11}x_{1,t-1} + q_{12}x_{2,t-1} + u_{1,t} + \varepsilon_{1,t}, \\ q_{21}x_{1,t} + q_{22}x_{2,t} = c(q_{21}x_{1,t-1} + q_{22}x_{2,t-1}) + u_{2,t} + \varepsilon_{2,t}. \end{cases}$$

From these equations

$$q_{21}x_{1,t} + q_{22}x_{2,t} = \sum_{j=0}^t c^j (u_{2,t-j} + \varepsilon_{2,t-j}), \quad q_{11}x_{1,t} + q_{12}x_{2,t} = \sum_{j=0}^t (u_{1,t-j} + \varepsilon_{1,t-j}).$$

Thus, we have

$$q_{11}x_{1,t} + q_{12}x_{2,t} = q_{11}x_{1,t} + q_{12} \frac{1}{q_{22}} \left(\sum_{j=0}^t c^j (u_{2,t-j} + \varepsilon_{2,t-j}) - q_{21}x_{1,t} \right) = \sum_{j=0}^t (u_{1,t-j} + \varepsilon_{1,t-j}),$$

and

$$\left(q_{11} - \frac{q_{12}q_{21}}{q_{22}} \right) x_{1,t} = \sum_{j=0}^t (u_{1,t-j} + \varepsilon_{1,t-j}) - \frac{q_{12}}{q_{22}} \sum_{j=0}^t c^j (u_{2,t-j} + \varepsilon_{2,t-j}).$$

Hence,

$$\begin{cases} x_{1,t} = \frac{q_{22}}{q_{11}q_{22} - q_{12}q_{21}} \sum_{j=0}^t (u_{1,t-j} + \varepsilon_{1,t-j}) - \frac{q_{12}}{q_{11}q_{22} - q_{12}q_{21}} \sum_{j=0}^t c^j (u_{2,t-j} + \varepsilon_{2,t-j}), \\ x_{2,t} = \frac{-q_{21}}{q_{22}} x_{1,t} + \frac{1}{q_{22}} \left(\sum_{j=0}^t c^j (u_{2,t-j} + \varepsilon_{2,t-j}) \right). \end{cases}$$

□

Appendix C

Proof of Theorem 1. There exists a matrix Q such that $QBQ^{-1} = B_1 = \begin{pmatrix} 1 & 0 \\ 0 & c \end{pmatrix}$, i.e., $B = Q^{-1}B_1Q$. Then by (2.1),

$$Qx_t = QA x_{t-1}1(x_{t-1} \in D) + QQ^{-1}B_1Qx_{t-1}1(x_{t-1} \in D^c) + Qe_t.$$

Moreover with $y_t = Qx_t$,

$$y_t = A_1y_{t-1}1(Q^{-1}y_{t-1} \in D) + B_1y_{t-1}1(Q^{-1}y_{t-1} \in D^c) + Qe_t,$$

where $A_1 = QAQ^{-1}$ and $B_1 = QBQ^{-1}$, i.e.,

$$y_t = A_1y_{t-1}1(y_{t-1} \in C) + B_1y_{t-1}1(y_{t-1} \in C^c) + \varepsilon_t,$$

with C being also a compact set.

We will use the drift criterion (V1) of Meyn and Tweedie (2009, P. 189) to discuss the null recurrence property of the above model. Inspired by Meyn and Tweedie (2009, P. 280), we define the following test function:

$$\begin{aligned} V(y) &= \ln(u - ay_1)1(y_1 < -R)1(-S \leq y_2 \leq S) + \ln(v + by_1)1(y_1 > R)1(-S \leq y_2 \leq S) \\ &\quad + |y_2|[1(y_2 < -S) + 1(y_2 > S)], \end{aligned}$$

where in our case, $a = b = u = v = 1$ (cf. Meyn and Tweedie 2009, P. 281), and where S and R are positive constants.

Moreover, we take $V(y) = 0$ within $[-R, R] \times [-S, S]$ and we also assume that $C \subset [-R, R] \times [-S, S]$. We need to show that $E(V(y_t)|y) - V(y) \leq 0$, where $E(V(y_t)|y) \equiv E(V(y_t)|y_{t-1} = y)$ with y being in the complement of $[-R, R] \times [-S, S]$. We decompose the proof into the following 4 steps.

(1) when $-S \leq y_2 \leq S$ and $y_1 < -R$, then $V(y) = \ln(1 - y_1)$ and

$$\begin{aligned} E(V(y_t)|y) &= E(\ln(1 - y_1 - \varepsilon_{1,t})1(y_1 + \varepsilon_{1,t} < -R)1(-S \leq cy_2 + \varepsilon_{2,t} \leq S)) \\ &\quad + E(\ln(1 + y_1 + \varepsilon_{1,t})1(y_1 + \varepsilon_{1,t} > R)1(-S \leq cy_2 + \varepsilon_{2,t} \leq S)) \\ &\quad + E((cy_2 + \varepsilon_{2,t})1[cy_2 + \varepsilon_{2,t} > S]) + E(-(cy_2 + \varepsilon_{2,t})[1[cy_2 + \varepsilon_{2,t} < -S]]) \\ &\leq E(\ln(1 - y_1 - \varepsilon_{1,t})1(y_1 + \varepsilon_{1,t} < -R)) + E(\ln(1 + y_1 + \varepsilon_{1,t})1(y_1 + \varepsilon_{1,t} > R)) \\ &\quad + E((cy_2 + \varepsilon_{2,t})1[cy_2 + \varepsilon_{2,t} > S]) + E(-(cy_2 + \varepsilon_{2,t})[1[cy_2 + \varepsilon_{2,t} < -S]]) \\ &\equiv I_1 + I_2 + I_3 + I_4. \end{aligned}$$

According to the proof of Proposition 11.5.4 of Meyn and Tweedie (2009), by choosing $R \rightarrow \infty$, we have $I_1 + I_2 \leq \ln(1 - y_1)$ when $y_1 < -R$. Meanwhile, the fact that the second moment of $\varepsilon_{2,t}$ exists gives

$$\begin{aligned} I_3 &= E((cy_2 + \varepsilon_{2,t})1[cy_2 + \varepsilon_{2,t} > S]) = E((cy_2 + \varepsilon_{2,t})1[\varepsilon_{2,t} > S - cy_2]) \\ E((cy_2 + \varepsilon_{2,t})1[\varepsilon_{2,t} > S - cy_2]) &= cy_2Pr[\varepsilon_{2,t} > S - cy_2] + E(\varepsilon_{2,t}[\varepsilon_{2,t} > S - cy_2]). \end{aligned}$$

The moment assumption on $\{\varepsilon_t\}$ then implies $cy_2Pr[\varepsilon_{2,t} > S - cy_2] \leq cSPr[\varepsilon_{2,t} > S - cy_2] = o(1)$ and $E(\varepsilon_{2,t}1[\varepsilon_{2,t} > S - cy_2]) = o(1/S)$. The same argument is valid for I_4 , so that when $R \rightarrow \infty$ and $S \rightarrow \infty$, we have $E(V(y_t)|y) \leq V(y)$.

(2) The proof for the case $y_1 > R$ and $-S \leq y_2 \leq S$ is the same as for the step (1).

(3) Now we study the case $y_2 > S$ and any fixed $y_1 \in \mathbb{R}$, then $V(y_1, y_2) = y_2$ and

$$\begin{aligned} E(V(y_t)|y) &= E(\ln(1 - y_1 - \varepsilon_{1,t})1(y_1 + \varepsilon_{1,t} < -R)1(-S \leq cy_2 + \varepsilon_{2,t} \leq S)) \\ &+ E(\ln(1 + y_1 + \varepsilon_{1,t})1(y_1 + \varepsilon_{1,t} > R)1(-S \leq cy_2 + \varepsilon_{2,t} \leq S)) \\ &+ E((cy_2 + \varepsilon_{2,t})1[cy_2 + \varepsilon_{2,t} > S]) + E(-(cy_2 + \varepsilon_{2,t})[1(cy_2 + \varepsilon_{2,t} < -S)]) \\ &\equiv L_1 + L_2 + L_3 + L_4. \end{aligned}$$

For any fixed y_1 , we have $L_1 + L_2 < u$, with u being a constant depending on y_1 . Meanwhile, we have

$$\begin{aligned} L_3 &= E((cy_2 + \varepsilon_{2,t})1[cy_2 + \varepsilon_{2,t} > S]) = cy_2 Pr(cy_2 + \varepsilon_{2,t} > S) + E(\varepsilon_{2,t}1[\varepsilon_{2,t} > S - cy_2]) \\ &= cy_2 + O(1). \end{aligned}$$

and $L_4 = E(-(cy_2 + \varepsilon_{2,t})[1(cy_2 + \varepsilon_{2,t} < -S)]) = O(1/S)$.

Hence, we obtain $E(V(y_t)|y) \leq cy_2 + u \leq y_2$, with u being a constant and $S \rightarrow \infty$.

(4) The proof for the case of $y_2 < -S$ is similar to step (3).

Thus we have proved the recurrence of the process $\{y_t\}$ according to the same reasoning as Proposition 11.5.4 of Meyn and Tweedie (2009). By Theorem 2 of Myklebust et al. (2012), we then know that $\{x_t\}$ is also recurrent. It is obvious that it cannot be positive recurrent due to the nonstationary behaviour far out. Now we are ready to prove 1/2-null recurrence of the bivariate process $\{x_t\}$. This has to do with the tail behaviour of the regeneration times and is being done as in Gao et al. (2013). We just sketch the main argument. We let the set D be the set of regenerating the process $\{x_t\}$. It is compact and thus small if the distribution of $\{e_t\}$ is absolutely continuous with respect to the Lebesgue measure.

Let S_r be the recurrence time, i.e., $S_r = \min\{n \geq 1 : x_n \in D\}$. So that $Pr(S_r > n) = Pr(x_1 \in D^c \cap x_2 \in D^c \dots \cap x_n \in D^c)$. Let $x'_t = Bx'_{t-1} + e_t$, and S'_r be the recurrence time of $\{x'_t\}$. Then, using the same reasoning as in Gao et al. (2013), $Pr(S_r > n)$ behaves as $Pr(S'_r > n)$ asymptotically.

According to Myklebust et al. (2012), we know that $\{x'_t\}$ is 1/2-null recurrent process and $Pr(S'_r > n) = O(n^{-1/2})$, and hence, $Pr(S_r > n) = O(n^{-1/2})$. Thus, the process $\{x_t\}$ is 1/2-null recurrent. \square

Proof of Theorem 2. Observe that

$$\hat{A}_1 - A_1 = \left(\sum_{t=1}^T \varepsilon_t y_{t-1}^\tau 1(y_{t-1} \in C) \right) \left(\sum_{t=1}^T y_{t-1} y_{t-1}^\tau 1(y_{t-1} \in C) \right)^{-1}.$$

We then have

$$\frac{1}{n(T)} \left(\sum_{t=1}^T y_{t-1} y_{t-1}^\tau 1(y_{t-1} \in C) \right) \rightarrow_p \begin{pmatrix} \int y_1^2 1(y \in C) p_s(y) dy & \int y_1 y_2 1(y \in C) p_s(y) dy \\ \int y_1 y_2 1(y \in C) p_s(y) dy & \int y_2^2 1(y \in C) p_s(y) dy \end{pmatrix},$$

and similar to Lemma 2.3 of Gao et al. (2013),

$$\frac{1}{\sqrt{n(T)}} \text{vec} \left(\left(\sum_{t=1}^T \varepsilon_t y_{t-1}^\tau 1(y_{t-1} \in C) \right) \right) \rightarrow_d$$

$$N \left(0, \begin{pmatrix} \int y_1^2 1(y \in C) p_s(y) dy & \int y_1 y_2 1(y \in C) p_s(y) dy \\ \int y_1 y_2 1(y \in C) p_s(y) dy & \int y_2^2 1(y \in C) p_s(y) dy \end{pmatrix} \otimes \Sigma_1 \right).$$

Thus, we have derived

$$\sqrt{n(T)} \text{vec} \left(\widehat{A}_1 - A_1 \right) \rightarrow_d N \left(0, (M^{-1} \otimes \Sigma_1) \right).$$

□

Proof of Theorem 3. Observe that

$$\begin{aligned} (\widehat{B}_1 - B_1) \times H &= \left(\sum_{t=1}^T \varepsilon_t y_{t-1}^\tau 1(y_{t-1} \in C^c) \right) \left(\sum_{t=1}^T y_{t-1} y_{t-1}^\tau 1(y_{t-1} \in C^c) \right)^{-1} \times H \\ &= \left(\sum_{t=1}^T \varepsilon_t y_{t-1}^\tau 1(y_{t-1} \in C^c) \times H^{-1} \right) \left(H^{-1} \sum_{t=1}^T y_{t-1} y_{t-1}^\tau 1(y_{t-1} \in C^c) H^{-1} \right)^{-1} \\ &= \left(\sum_{t=1}^T \varepsilon_t y_{t-1}^\tau \times H^{-1} - \sum_{t=1}^T \varepsilon_t y_{t-1}^\tau 1(y_{t-1} \in C) \times H^{-1} \right) \\ &\quad \times \left(H^{-1} \times \sum_{t=1}^T (y_{t-1} y_{t-1}^\tau - y_{t-1} y_{t-1}^\tau 1(y_{t-1} \in C)) \times H^{-1} \right)^{-1} \\ &= \left(\sum_{t=1}^T \varepsilon_t y_{t-1}^\tau \times H^{-1} \right) \left(H^{-1} \times \sum_{t=1}^T y_{t-1} y_{t-1}^\tau \times H^{-1} \right)^{-1} + o_P(1) \\ &= \left\{ \left(T^{-1} \sum_{t=1}^T y_{1,t-1} \varepsilon_t \right) \left(T^{-2} \sum_{t=1}^T y_{1,t-1}^2 \right)^{-1} \left(T^{-1/2} \sum_{t=1}^T y_{2,t-1} \varepsilon_t \right) \left(T^{-1} \sum_{t=1}^T y_{2,t-1}^2 \right)^{-1} \right\} + o_P(1) \end{aligned}$$

by the result of Theorem 2 and then (3) of Lemma 1. Theorem 3 then follows from Lemmas 1 and 2 as well as the continuous mapping theorem.

□

Proof of Theorem 4. This follows similarly from that of Theorem 2.

□

Proof of Theorem 5. Observe that

$$\begin{aligned} (\widehat{B}_1 - B_1) &= \left(\sum_{t=1}^T \varepsilon_t y_{t-1}^\tau 1(y_{t-1} \in C^c) \right) \left(\sum_{t=1}^T y_{t-1} y_{t-1}^\tau 1(y_{t-1} \in C^c) \right)^{-1} \\ &= \left(\sum_{t=1}^T Q e_t x_{t-1}^\tau Q^\tau 1(x_{t-1} \in D^c) \right) \left(\sum_{t=1}^T Q x_{t-1} x_{t-1}^\tau Q^\tau 1(x_{t-1} \in D^c) \right)^{-1} \\ &= Q \left(\sum_{t=1}^T e_t x_{t-1}^\tau 1(x_{t-1} \in D^c) \right) \left(\sum_{t=1}^T x_{t-1} x_{t-1}^\tau 1(x_{t-1} \in D^c) \right)^{-1} Q^{-1} = Q \left(\widehat{B} - B \right) Q^{-1} \end{aligned}$$

by the proof of Theorem 3. Taking the different convergence rates of Theorem 3 into account, it follows by an easy computation that $\sqrt{T} \text{vec} \left(\widehat{B} - B \right) \rightarrow_d N(0, \Sigma_2)$, where

$$\Sigma_2 = (Q^\tau \otimes Q^{-1}) \left(\begin{pmatrix} 0 & 0 \\ 0 & (\sigma_{y_2}^2)^{-1} \end{pmatrix} \otimes \Sigma_1 \right) (Q \otimes (Q^{-1})^\tau).$$

□

Proof of Theorem 6. From (3.7), we have

$$\widehat{\beta}_2 - \beta = \frac{\sum_{t=1}^T x_{1,t} \eta_t}{\sum_{t=1}^T x_{1,t}^2}, \quad (\text{C.1})$$

where $\eta_t = k_3(\sum_{j=0}^t c^j(u_{2,t-j} + \varepsilon_{2,t-j}))$ and $x_{1,t} = k_1 y_{1,t} + z_t$ with $z_t = k_2 \sum_{j=0}^t c^j(u_{2,t-j} + \varepsilon_{2,t-j})$. In Section 3.1, we have shown that $\frac{1}{\sqrt{T}} y_{1,[Tr]} \rightarrow_d P(r)$. Following (2) of Lemma 1 and Cauchy-Schwarz inequality,

$$\frac{1}{T^2} \sum_{t=1}^T x_{1,t}^2 = \frac{1}{T^2} k_1^2 \sum_{t=1}^T y_{1,t}^2 + \frac{1}{T^2} k_2^2 \sum_{t=1}^T z_t^2 + \frac{2}{T^2} k_1 k_2 \sum_{t=1}^T y_{1,t} z_t = \frac{1}{T^2} k_1^2 \sum_{t=1}^T y_{1,t}^2 + o_P(1).$$

Then, by the continuous mapping theorem

$$\frac{1}{T^2} \sum_{t=1}^T x_{1,t}^2 \rightarrow_d k_1^2 \int_0^1 P^2(r) dr. \quad (\text{C.2})$$

Furthermore, we have

$$\frac{1}{T} \sum_{t=1}^T x_{1,t} \eta_t = \frac{1}{T} \sum_{t=1}^T x_{1,t} (k_3 (\sum_{j=0}^t c^j (u_{2,t-j} + \varepsilon_{2,t-j}))). \quad (\text{C.3})$$

Let us now consider

$$\begin{aligned} & \frac{1}{T} \sum_{t=1}^T x_{1,t} (\sum_{j=0}^t c^j (u_{2,t-j})) = \frac{1}{T} \sum_{j=0}^T c^j (\sum_{t=1}^T x_{1,t} u_{2,t-j}) \\ & = \frac{1}{T} \sum_{j=0}^{m_T} c^j (\sum_{t=1}^T x_{1,t} u_{2,t-j}) + \frac{1}{T} \sum_{j=m_T}^T c^j (\sum_{t=1}^T x_{1,t} u_{2,t-j}) \equiv I_1 + I_2. \end{aligned}$$

The definition of $x_{1,t}$ in (3.5) gives

$$\begin{aligned} I_1 &= \frac{1}{T} \sum_{j=0}^{m_T} c^j (\sum_{t=1}^{T-j} x_{1,t} u_{2,t}) = \sum_{j=0}^{m_T} c^j (\sum_{t=1}^{T-j} (x_{1,t-j-1} + k_1 \sum_{i=t-j}^t (u_{1,i} + \varepsilon_{1,i}) + z_t - z_{t-j-1})) u_{2,t-j} \\ &\equiv I_{11} + I_{12} + I_{13} + I_{14}, \end{aligned}$$

where $I_{11} = \frac{1}{T} \sum_{j=0}^{m_T} c^j \sum_{t=1}^{T-j} x_{1,t-j-1} u_{2,t-j}$, $I_{12} = \frac{1}{T} \sum_{j=0}^{m_T} c^j \sum_{t=1}^{T-j} k_1 \sum_{i=t-j}^t (u_{1,i} + \varepsilon_{1,i}) u_{2,t-j}$, $I_{13} = \frac{1}{T} \sum_{j=0}^{m_T} c^j \sum_{t=1}^{T-j} z_t u_{2,t-j}$ and $I_{14} = -\frac{1}{T} \sum_{j=0}^{m_T} c^j \sum_{t=1}^{T-j} z_{t-j} u_{2,t-j}$.

By the definition of u_t , $|x_{1,t-1} u_{2,t}| < C' |u_{2,t}|$ for some constant C' . Following Lemma 3.3 of Karlsen and Tjøstneim (2001),

$$E|I_{11}| \leq C_1 \frac{1}{T} \sum_{j=0}^{m_T} c^j T^{1/2+\epsilon} = O\left(\frac{T^{1/2+\epsilon}}{T}\right)$$

for any $\epsilon > 0$.

Meanwhile, using Cauchy-Schwarz inequality implies

$$\begin{aligned} |I_{12}| &\leq \frac{1}{T} \sum_{j=0}^{m_T} c^j \left| \sum_{t=1}^{T-j} k_1 \sum_{i=t-j}^t (u_{1,i} + \varepsilon_{1,i}) u_{2,t-j} \right| \\ &\leq C_2 \frac{1}{T} \sum_{j=0}^{m_T} c^j j \sqrt{\sum_{t=1}^T (u_{1,t} + \varepsilon_{1,t})^2 \sum_{t=1}^T u_{2,t}^2} = O_P\left(\frac{1}{T}\right) \sqrt{O_P(T) O_P(\sqrt{T})} = O_P(T^{-1/4}). \end{aligned}$$

In addition, we have

$$|I_{13}| \leq \frac{1}{T} \sum_{j=0}^{m_T} c^j \sqrt{\sum_{t=1}^T z_t^2 \sum_{t=1}^T u_{2,t}^2} = O_P(T^{-1/4})$$

by (2) of Lemma 1 and $\sum_{t=1}^T u_{2,t}^2 = O_P(\sqrt{T})$. Similarly, $I_{14} = o_P(1)$.

For I_2 , using the fact that $u_{2,t}$ is bounded uniformly in t ,

$$\frac{1}{T} E \left| \sum_{j=m_T}^T c^j \left(\sum_{t=1}^T x_{1,t} u_{2,t-j} \right) \right| \leq C_3 \sum_{j=m_T}^T c^j \sum_{t=1}^T E |x_{1,t}| \leq C_4 T^2 \sum_{j=m_T}^T c^j = o(1)$$

by letting $m_T = \lceil T^a \rceil$ with $0 < a < 1$.

Thus, we have

$$\frac{1}{T} \sum_{t=1}^T x_{1,t} \eta_t = \frac{1}{T} \sum_{t=1}^T x_{1,t} (k_3 \left(\sum_{j=0}^t c^j \varepsilon_{2,t-j} \right)) + o_P(1). \quad (\text{C.4})$$

By (3.5), we deduce

$$\begin{aligned} & \frac{1}{T} \sum_{t=1}^T x_{1,t} (k_3 \left(\sum_{j=0}^t c^j \varepsilon_{2,t-j} \right)) \\ &= \frac{1}{T} \sum_{t=1}^T (k_1 \sum_{j=0}^t (u_{1,t-j} + \varepsilon_{1,t-j}) + k_2 \sum_{j=0}^t c^j (u_{2,t-j} + \varepsilon_{2,t-j})) (k_3 \left(\sum_{j=0}^t c^j \varepsilon_{2,t-j} \right)) \equiv L_1 + L_2, \end{aligned}$$

where $L_1 = \frac{1}{T} \sum_{t=1}^T (k_1 \sum_{j=0}^t (u_{1,t-j} + \varepsilon_{1,t-j})) (k_3 \left(\sum_{j=0}^t c^j \varepsilon_{2,t-j} \right))$ and $L_2 = \frac{1}{T} \sum_{t=1}^T (k_2 \sum_{j=0}^t c^j (u_{2,t-j} + \varepsilon_{2,t-j})) (k_3 \left(\sum_{j=0}^t c^j \varepsilon_{2,t-j} \right))$.

By (B.1) and (B.3),

$$L_2 \rightarrow_p k_2 k_3 \left(\sum_{i=0}^{\infty} c^{2i} \right) \sigma_{\varepsilon_2}^2 = k_2 k_3 \frac{1}{1-c^2} \sigma_{\varepsilon_2}^2. \quad (\text{C.5})$$

For L_1 , Theorem 3.1 of Hansen (1992) will be applied. First of all, from the functional central limit theorem for an α -mixing sequence (cf. Davidson 1994, Theorem 29.6),

$$\frac{1}{\sqrt{T}} \sum_{t=1}^{\lceil Tr \rceil} \left(\sum_{j=0}^t c^j \varepsilon_{2,t-j} \right) \rightarrow_d \left(\sum_{i=0}^{\infty} c^i \right) B_2(r) = \frac{1}{1-c} B_2(r).$$

Define that \mathcal{G}_t is the σ algebra generated by $\{\varepsilon_{1,t}, \varepsilon_{1,t-1}, \dots, \varepsilon_{1,1}\}$ and define $\zeta_{t+1} = k_3 \left(\sum_{j=0}^t c^j \varepsilon_{2,t-j} \right)$. Construct the following martingale difference sequence:

$$\delta_t = \sum_{i=0}^{\infty} (E_t \zeta_{t+i} - E_{t-1} \zeta_{t+i}) = k_3 \sum_{i=0}^{\infty} c^i \varepsilon_{2,t-1} = k_3 \frac{1}{1-c} \varepsilon_{2,t-1},$$

where E_t is the conditional expectation with respect to information \mathcal{G}_t .

Then, we have following martingale decomposition (cf. Hansen 1992, P. 492 or Phillips and Solo 1992, Lemma 2.1)

$$\zeta_t = k_3 \sum_{j=0}^{t-1} c^j (\varepsilon_{2,t-1-j}) = \delta_t + \nu_{t-1} - \nu_t,$$

where $\nu_t = k_3 \sum_{j=0}^{t-1} \frac{c^{j+1}}{1-c} \varepsilon_{2,t-1-j}$.

Simple calculation (cf. Hansen 1992, Equation 3) shows

$$L_1 = k_1 k_3 \frac{1}{T} \sum_{t=1}^T \left(\sum_{j=0}^t (u_{1,t-j} + \varepsilon_{1,t-j}) \right) \left(\sum_{j=0}^t c^j \varepsilon_{2,t-j} \right)$$

$$\begin{aligned}
&= k_1 k_3 \frac{1}{T} \sum_{t=1}^T \left(\sum_{j=0}^t (u_{1,t-j} + \varepsilon_{1,t-j}) \right) (\delta_{t+1} + \nu_t - \nu_{t+1}) \\
&= k_1 k_3 \frac{1}{T} \sum_{t=1}^T \left(\sum_{j=0}^t (u_{1,t-j} + \varepsilon_{1,t-j}) \right) \delta_{t+1} + k_1 k_3 \frac{1}{T} \sum_{t=1}^T (u_{1,t} + \varepsilon_{1,t}) \nu_t - k_1 k_3 \frac{1}{T} \left(\sum_{t=1}^T u_{1,t} + \varepsilon_{1,t} \right) \nu_{T+1}.
\end{aligned}$$

By Cauchy-Schwarz inequality,

$$k_1 k_3 \frac{1}{T} \left(\sum_{t=1}^T u_{1,t} + \varepsilon_{1,t} \right) \nu_{T+1} = o_P(1).$$

By the law of large numbers for an α -mixing sequence,

$$k_1 k_3 \frac{1}{T} \sum_{t=1}^T \varepsilon_{1,t} \nu_t \rightarrow_p 0.$$

And by Cauchy-Schwarz inequality again,

$$k_1 k_3 \frac{1}{T} \sum_{t=1}^T u_{1,t} \nu_t \leq k_1 k_3 \frac{1}{T} \sqrt{\sum_{t=1}^T u_{1,t}^2 \sum_{t=1}^T \nu_t^2} = \frac{1}{T} \sqrt{O_P(T^{1/2}) O_P(T)} = O_P(T^{-1/4}).$$

Then Theorem 3.1 of Hansen (1992) implies

$$L_1 \rightarrow_d k_1 k_3 \frac{1}{1-c} \int_0^1 P(r) dB_2(r). \quad (\text{C.6})$$

Finally, by (C.1-C.6), using the continuous mapping theorem, we have

$$T(\widehat{\beta}_2 - \beta) \rightarrow_d \frac{k_1 k_3 \frac{1}{1-c} \int_0^1 P(r) dB_2(r) + \lambda}{k_1^2 \int_0^1 P^2(r) dr}, \quad (\text{C.7})$$

where $\lambda = k_2 k_3 \frac{\sigma_{\varepsilon_2}^2}{1-c^2}$.

To finish the proof, it suffices to show that the two estimators are equivalent. Note that

$$T(\widehat{\beta}_1 - \beta) = \frac{\frac{1}{T} \sum_{t=1}^T x_{1,t} \eta_t 1(x_{t-1} \in D^c)}{\frac{1}{T^2} \sum_{t=1}^T x_{1,t}^2 1(x_{t-1} \in D^c)}. \quad (\text{C.8})$$

To show the equivalence of the two estimators, we need only to show

$$\frac{1}{T} \sum_{t=1}^T x_{1,t} \eta_t 1(x_{t-1} \in D) = o_P(1) \quad (\text{C.9})$$

and

$$\frac{1}{T^2} \sum_{t=1}^T x_{1,t}^2 1(x_{t-1} \in D) = o_P(1). \quad (\text{C.10})$$

We first show (C.9), we have

$$\frac{1}{T} \sum_{t=1}^T x_{1,t} \eta_t 1(x_{t-1} \in D) = \frac{1}{T} \sum_{t=1}^T [x_{1,t-1} + k_1(u_{1,t} + \varepsilon_{1,t}) + z_t - z_{t-1}] \eta_t 1(x_{t-1} \in D).$$

Using Cauchy-Schwarz inequality, it is easy to show that $\frac{1}{T} \sum_{t=1}^T x_{1,t-1} \eta_t 1(x_{t-1} \in D) = o_P(1)$, $\frac{1}{T} \sum_{t=1}^T k_1(u_{1,t} + \varepsilon_{1,t}) \eta_t 1(x_{t-1} \in D) = o_P(1)$, $\frac{1}{T} \sum_{t=1}^T z_t \eta_t 1(x_{t-1} \in D) = o_P(1)$ and $\frac{1}{T} \sum_{t=1}^T z_{t-1} \eta_t 1(x_{t-1} \in D) = o_P(1)$. Thus, (C.9) has been proved.

To show (C.10), we have

$$\begin{aligned} \frac{1}{T^2} \sum_{t=1}^T x_{1,t}^2 1(x_{t-1} \in D) &= \frac{1}{T^2} \sum_{t=1}^T [x_{1,t-1} + k_1(u_{1,t} + \varepsilon_{1,t}) + z_t - z_{t-1}]^2 1(x_{t-1} \in D) \\ &\leq 4 \frac{1}{T^2} \sum_{t=1}^T (x_{1,t-1})^2 1(x_{t-1} \in D) + 4 \frac{1}{T^2} \sum_{t=1}^T (k_1(u_{1,t} + \varepsilon_{1,t}))^2 1(x_{t-1} \in D) \\ &\quad + 4 \frac{1}{T^2} \sum_{t=1}^T z_t^2 1(x_{t-1} \in D) + 4 \frac{1}{T^2} \sum_{t=1}^T z_{t-1}^2 1(x_{t-1} \in D). \end{aligned}$$

It is straightforward to show that the four terms are all $o_P(1)$. Therefore, (C.10) has been proved. \square

References

- Ahn, S.K., and G.C., Reinsel, 1990. Estimation for partially nonstationary multivariate autoregressive models. *Journal of American Statistical Association*, 65, 813-823.
- Anderson, H.M., 1997. Transaction costs and non-linear adjustment towards equilibrium in the US Treasury market. *Oxford Bulletin of Economics and Statistics*, 59, 465-484.
- Anderson, T.G. and J., Lund, 1997. Estimating continuous-time stochastic volatility models of the short-term interest rate. *Journal of Econometrics*, 77, 343-377.
- Balke, N.S. and T.B., Fomby, 1997. Threshold cointegration. *International Economic Review*, 38, 627-645.
- Bec, F., and A., Rahbek, 2004. Vector equilibrium correction models with non-linear discontinuous adjustments. *Econometrics Journal*, 2, 628-651.
- Cai, B. and D., Tjøstheim, 2014. Nonparametric regression estimation for multivariate null recurrent processes. University of Bergen working paper.
- Chan, K.S., 1993. Consistency and limiting distribution of the least squares estimator of a threshold autoregressive model. *Annals of Statistics*, 21, 520-533.
- Chan, K.S. and R.S., Tsay, 1998. Limiting properties of the least squares estimator of a continuous threshold autoregressive model. *Biometrika*, 85, 413-426.
- Clarida, R.H., L., Sarno, M.P., Taylor and G., Valente, 2006. The role of asymmetries and regime shifts in term structure of interest rates. *Journal of Business*, 79, 1193-1224.
- Davidson, J., 1994. *Stochastic Limit Theory: an Introduction for Econometricians*. Oxford University Press.
- Engle, R.F. and C.W.J., Granger, 1987. Cointegration and error correction: representation, estimation, and testing. *Econometrica*, 55, 251-276.
- Gaenssler, P. and K., Joos, 1992. Another view on martingale central limit theorems. *Stochastic Processes and their Applications*, 40, 181-197.
- Gao, J., D., Tjøstheim and J., Yin, 2013. Estimation in threshold autoregressive models with a stationary and a unit root regime. *Journal of Econometrics*, 172, 1-13.
- Granger, C.W.J., 1981. Some properties of time series data and their use in econometric models specification. *Journal of Econometrics*, 16, 121-130.
- Hall, A.D., H.M., Anderson and C.W.J., Granger, 1992. A cointegration analysis of Treasury bill yields. *Review of Economics and Statistics*, 74, 116-126.
- Hall, P. and C.C., Heyde, 1980. *Martingale Limit Theory and Its Applications*. Academic Press, New York.
- Hansen, B.E., 1992. Convergence to stochastic integrals for dependent heterogeneous processes. *Econometric Theory*, 8, 489-500.
- Karlsen, H.A. and D., Tjøstheim, 2001. Nonparametric estimation in null recurrent time series. *Annals of Statistics*, 29, 372-416.

- Karlsen, H.A., T., Myklebust and D., Tjøstheim, 2007. Nonparametric estimation in a nonlinear cointegration type model. *Annals of Statistics*, 35, 252-299.
- Johansen, S., 1988. Statistical analysis of cointegration vectors. *Journal of Economic Dynamics and Control*, 12, 231-254.
- Lütkepohl, H., 2005. *New Introduction to Multiple Time Series Analysis*. Springer, Berlin.
- Meleish, D.L., 1975. A maximal inequality and dependent strong laws. *Annals of Probability*, 3, 829-839.
- Meyn, S. and R.L., Tweedie, 2009. *Markov Chain and Stochastic Stability*. Cambridge University Press, 2nd edition.
- Myklebust, T., Karlsen, H.A. and D., Tjøstheim, 2012. Null recurrent unit root processes. *Econometric Theory*, 28, 1-41.
- Nummelin, E., 1984. *General Irreducible Markov Chains and Non-negative Operators*. Cambridge University Press.
- Park, J.Y. and P.C.B., Phillips, 1999. Asymptotics for nonlinear transformations of integrated time series. *Econometric Theory*, 15, 269-298.
- Park, J.Y. and P.C.B., Phillips, 2001. Nonlinear regressions with integrated time series. *Econometrica*, 69, 117-161.
- Pham, T.D. and L.T., Tran, 1985. Some mixing properties of time series models. *Stochastic Processes and Their Applications*, 19, 297-303.
- Phillips, P.C.B., 1987. Time series regression with a unit root. *Econometrica*, 55, 277-301.
- Phillips, P.C.B., 1991. Optimal inference in cointegrated systems. *Econometrica*, 59, 283-306.
- Phillips, P.C.B. and B.E., Hansen, 1990. Statistical inference in instrument variable regression with I(1) processes. *Review of Economic Studies*, 57, 99-125.
- Phillips, P.C.B. and V., Solo, 1992. Asymptotics for linear processes. *Annals of Statistics*, 20, 971-1001.
- Rudebusch, G.D., 1995. Federal reserve interest rate, targeting rational expectations, and the term structure. *Journal of Monetary Economics*, 35, 245-274.
- Saikkonen, P., 2005. Stability results for nonlinear error correction models. *Journal of Econometrics*, 127, 69-81.
- Saikkonen, P., 2008. Stability of regime switching error correction models under linear cointegration. *Econometric Theory*, 24, 294-318.
- Sims, C.A., 1980. Macroeconomics and reality. *Econometrica*, 48, 1-48.
- Stock, J.H. and M.W., Watson, 1988. Testing for common trends. *Journal of American Statistical Association*, 83, 1097-1107.
- Stock, J.H. and M.W., Watson, 2001. Vector autoregression. *Journal of Economic Perspectives*, 15, 101-115.
- Tong, H., 1978. On a threshold model. In *Pattern Recognition and Signal Processing*, ed. by Chen C.H., Sijthoff and Noordhoff, Amsterdam.
- Tong, H., 1983. *Threshold Models in Nonlinear Time Series Analysis*. Springer-Verlag, New York.
- Wang, Q. and P.C.B., Phillips, 2009a. Asymptotic theory for local time density and nonparametric cointegration regression. *Econometric Theory*, 25, 710-738.
- Wang, Q. and P.C.B., Phillips, 2009b. Structural nonparametric cointegration regression. *Econometrica*, 77, 1901-1948.
- Watson, M.W., 1994. Vector autoregression and cointegration. In *Handbook of Econometrics*, Vol. 4, ed. by Engle, R.F. and D.L., McFadden, North-Holland.