Estimation of Nonlinear Error Correction Models

Myung Hwan Seo[*]
London School of Economics

Discussion paper
No. EM/2007/517
March 2007

The Suntory Centre
Suntory and Toyota International Centres for
Economics and Related Disciplines
London School of Economics and Political Science
Houghton Street
London WC2A 2AE
Tel: 020 7955 6679

---

[*] Department of Economics, London School of Economics, Houghton Street, London WC2A 2AE, United Kingdom. E-mail address: m.seo@lse.ac.uk. This research was supported through a grant from the Economic and Social Science Research Council.

Abstract

Asymptotic inference in nonlinear vector error correction models (VECM) that exhibit regime-specific short-run dynamics is nonstandard and complicated. This paper contributes the literature in several important ways. First, we establish the consistency of the least squares estimator of the cointegrating vector allowing for both smooth and discontinuous transition between regimes. This is a nonregular problem due to the presence of cointegration and nonlinearity. Second, we obtain the convergence rates of the cointegrating vector estimates. They differ depending on whether the transition is smooth or discontinuous. In particular, we find that the rate in the discontinuous threshold VECM is extremely fast, which is n^{3/2}, compared to the standard rate of n: This finding is very useful for inference on short-run parameters. Third, we provide an alternative inference method for the threshold VECM based on the smoothed least squares (SLS). The SLS estimator of the cointegrating vector and threshold parameter converges to a functional of a vector Brownian motion and it is asymptotically independent of that of the slope parameters, which is asymptotically normal.

Keywords: Threshold Cointegration, Smooth Transition Error Correction, Least Squares, Smoothed Least Squares, Consistency, Convergence Rate.

JEL No: C32

1 Introduction

Nonlinear error correction models (ECM) have been studied actively in economics and there are numerous applications. To list only a few, see Michael, Nobay, and Peel (1997) for the application in purchasing power parity, Anderson (1997) in the term structure model of interest rates, Escribano (2004) in Money demand, Psaradakis, Sola, and Spagnolo (2004) in relation between stock prices and dividends, and Sephton (2003) in spatial market arbitrage, and see also a review by Granger (2001). The models include smooth transition ECM of Granger and Teräsvirta (1993), threshold cointegration of Balke and Fomby (1997), Markov switching ECM of Psaradakis et al. (2004), and cubic polynomials of Escribano (2004).

A strand of econometric literature focuses on testing for the presence of nonlinearity and cointegration in an attempt to disentangle the nonstationarity from nonlinearity. A partial list includes Hansen and Seo (2002), Kapetanios, Shin, and Snell (2006) and Seo (2006). Time series properties of various ECMs have been established by Corradi, Swanson, and White (2000) and Saikkonen (2005, 2007) among others. However, the result on estimation is still limited. Most of all, consistency has not been proven except for special cases. It is difficult to establish due to the lack of uniformity in the convergence over the cointegrating vector space as noted by Saikkonen (1995), which derived the consistency of the MLE of a cointegrated system that is nonlinear in parameters but otherwise linear. de Jong (2002) studied consistency of minimization estimators of smooth transition ECMs where the error correction term appears only in a bounded transition function. Another case studied by Kristensen and Rahbek (2008) is that the function is unbounded but becomes linear as the error correction term diverges. Next, estimation of regime-switching and/or discontinuous cases has hardly been studied, which includes important class of models such as smooth transition ECM and threshold cointegration. Hansen and Seo (2002) proposed the MLE under normality but only to make conjecture on the consistency. While it may be argued that the two-step approach by Engle and Granger (1987) can be adopted due to the super-consistency of the cointegrating vector estimate, the estimation error cannot be ignored in nonlinear ECMs as shown by de Jong (2001).[1]

The purpose of this paper is to develop asymptotic theory for a class of nonlinear vector error correction models (VECM). In particular, we consider regime switching VECMs, where each regime exhibits different short-run dynamics and the regime switching depends on the disequilibrium error. Examples include threshold cointegration and smooth transition VECM. First, we establish the square root n consistency for the LS estimator of β. This enables us to employ de Jong (2002) to make asymptotic inference for both short-run and long-run parameters jointly in smooth transition models. Then, we turn to discontinuous models, focusing on the threshold cointegration model, which is particularly popular in practice.

This paper shows that the convergence of the LS estimator of β in the threshold cointegration model is extremely fast at the rate of $n^{3/2}$. This asymptotics is based not on the diminishing threshold asymptotics of Hansen (2000) but on the fixed threshold asymptotics. Two different irregularities contribute to this fast rate. First, the estimating function

[1] It provides an orthogonality condition, under which the two-step approach is valid.

lacks uniformity over the cointegrating vector space as the data becomes stationary at the true value, which is the reason for the super-consistency of the standard cointegrating vector estimates. Second, β takes part in regime switching, which is discontinuous. This model discontinuity also boosts the convergence rate, yielding the super-consistency of the threshold estimate as in Chan (1993). While this fast convergence rate is certainly interesting and has some inferential value, e.g. when we perform sequential test to determine the number of regimes, it makes it very challenging to obtain an asymptotic distribution of the estimator. Even in the stationary threshold autoregression, the asymptotic distribution is very complicated and cannot be tabulated (see Chan 1993). Subsampling is the only way to approximate the distribution in the literature reported by Gonzalo and Wolf (2005), although it would not work when β is estimated due to the involved nonstationarity. Meanwhile, Seo and Linton (2007) proposed the smoothed least squares (SLS) estimation for threshold regression models, which results in the asymptotic normality of the threshold estimate and is applicable to the threshold cointegration model.

We develop the asymptotic distributions of the SLS estimators of β, the threshold parameter γ, and the other short-run parameters. The estimates $\hat{\beta}$ and $\hat{\gamma}$ converge jointly to a functional of Brownian motions, with the rates slightly slower than those of the unsmoothed counterparts. This slow-down in convergence rate has already been observed in Seo and Linton and is the price to pay to achieve standard inference. The remaining regression parameter estimates converge to the Normal as if the true values of β and γ were known. This is not the case if the transition function is smooth. We also show that β can be treated as if known in the SLS estimation of the short-run parameters including γ if we plug in the unsmoothed cointegrating vector estimate due to the fast convergence rate. A set of Monte Carlo experiments demonstrates that this two-step approach is more efficient in finite samples.

This paper is organized as follows. Section 2 introduces the regime switching VECMs and establishes the square root n consistency of the LS estimator of β. Section 3 concentrates on the threshold cointegration model, obtaining the convergence rate of the LS estimator of β and the asymptotic distributions of the SLS estimators of all the model parameters. It also discusses estimation of the asymptotic variances. Finite sample performance of the proposed estimators is examined in Section 4. Section 5 concludes. Proofs of theorems are collected in the appendix.

We make the following conventions throughout the paper. The integral $\int$ is taken over $\mathbb{R}$ unless specified otherwise and the summation $\sum_t$ with respect to t is taken for all available observations for a given sample. The subscript 0 and the hat ^ in any parameter indicate the true value and an estimate of the parameter, respectively, e.g., β_0 and $\hat{\beta}$. For a function g, $\|g\|_2^2 = \int g(x)^2 dx$ and $g^{(i)}$ indicates the i^{th} derivative of g. And, for a random vector x_t and a parameter θ, we write $g_t(\theta) = g(x_t, \theta)$, $g_t = g(x_t, \theta_0)$ and $\hat{g}_t = g\left(x_t, \hat{\theta}\right)$. For example, if $z_t(\beta) = x_t'\beta$, then we write $z_t = x_t'\beta_0$ and $\hat{z}_t = x_t'\hat{\beta}$. The weak convergence of stochastic processes under the uniform metric is signified by $\Rightarrow$.

2 Regime Switching Error Correction Models

Let x_t be a p-dimensional $I(1)$ vector that is cointegrated with single cointegrating vector. It is denoted by $\left(1, \beta'\right)'$, normalizing the first element by 1. Define the error correction term $z_t(\beta) = x_{1t} + x_{2t}'\beta$, where $x_t = (x_{1t}, x_{2t}')'$, and let

$$X_{t-1}(\beta) = \left(1, z_{t-1}(\beta), \Delta_{t-1}'\right)',$$

where Δ_{t-1} denotes the vector of the lagged first difference terms $\left(\Delta x_{t-1}', \cdots, \Delta x_{t-l+1}'\right)'$. Then, consider a two-regime vector error correction model

$$\Delta x_t = A' X_{t-1}(\beta) + D' X_{t-1}(\beta) d_{t-1}(\beta, \gamma) + u_t, \tag{1}$$

where $t = l+1, ..., n$, and $d_t(\beta, \gamma) = d(z_t(\beta), \gamma)$ is a bounded function that controls the transition from one regime to the other regime. It needs not be continuous. Typical examples of the transition function include the indicator function $1\{z_t(\beta) > \gamma\}$ and the logistic function $(1 - \exp(-\gamma_1(z_t(\beta) - \gamma_2)))^{-1}$, where $\gamma = (\gamma_1, \gamma_2)'$.

The threshold cointegration model of Balke and Fomby (1997) and the smooth transition error correction model in Granger and Teräsvirta (1993) can be viewed as special cases. As an alternative, Escribano (2004) used cubic polynomials to capture this type of regime-specific error correction behavior. While the last model is not nested in model (1), all this literature focuses on the nonlinear adjustment based on the magnitude of disequilibrium error. In this regard, Gonzalo and Pitarakis (2006) is different, in which a stationary variable determines the regimes. While we study a two-regime model to simplify our exposition, we expect the models with more than two regimes can be analyzed in a similar way. A symmetric three-regime model can be directly embedded in the two-regime model by replacing the threshold variable z_t with its absolute value $|z_t|$. However, some of the assumptions imposed later on in this section, in particular, the one with the series being $I(1)$ becomes more difficult to verify. See Saikkonen (2007).

We introduce some matrix notation. Define $X(\beta)$, $X_\gamma^*(\beta)$, y, and u as the matrices stacking $X_{t-1}'(\beta)$, $X_{t-1}'(\beta) d_{t-1}(\beta, \gamma)$, Δx_t and u_t, respectively. Let $\lambda = \text{vec}\left((A', D')'\right)$, where vec stacks rows of a matrix. We call by A_z and D_z the columns of A' and D' that are associated with $z_{t-1}(\beta)$ and $z_{t-1}(\beta) d_{t-1}(\beta, \gamma)$, respectively, and by λ_z the collection of A_z and D_z. Then, we may write

$$y = \left[\left(X(\beta), X_\gamma^*(\beta)\right) \otimes I_p\right] \lambda + u.$$

We consider the LS estimation, which minimizes

$$S_n^*(\theta) = \left(y - \left[\left(X(\beta), X_\gamma^*(\beta)\right) \otimes I_p\right] \lambda\right)' \left(y - \left[\left(X(\beta), X_\gamma^*(\beta)\right) \otimes I_p\right] \lambda\right), \tag{2}$$

where $\theta = \left(\beta', \gamma, \lambda'\right)'$. The LS estimator is then defined as

$$\hat{\theta}^* = \arg\min_\theta S_n^*(\theta),$$

where the minimum is taken over a compact parameter space Θ. The concentrated LS is computationally convenient, since it is simple OLS for a fixed (β,γ), *i.e.*

$$\lambda^*(\beta,\gamma)=\left(\begin{bmatrix} X(\beta)'X(\beta) & X(\beta)'X^*_\gamma(\beta) \\ X^*_\gamma(\beta)'X(\beta) & X^*_\gamma(\beta)'X^*_\gamma(\beta)\end{bmatrix}^{-1}\begin{pmatrix} X(\beta)' \\ X^*_\gamma(\beta)'\end{pmatrix}\otimes I_p\right)y,$$

which is then plugged back into (2) for optimization over (β,γ). In practice, the grid search over (β,γ) can be applied. In particular, the grid for β can be set up around a preliminary estimate of β, that can be obtained based on the linear VECM, such as the Johansen's maximum likelihood estimator or the simple OLS estimator as described in Hansen and Seo (2002).

The asymptotic property of the estimator $\hat{\theta}^*$ is nonstandard due to the irregular feature of S_n^*, which does not obey a uniform law of large numbers. Thus, we take a two-step approach. First it is shown that $\hat{\beta}^*=\beta_0+O_p\left(n^{-1/2}\right)$ by evaluating the difference between $\inf S_n^*(\theta)$ and $S_n^*(\theta_0)$, where the infimum is taken over all $\theta\in\Theta$ such that $r_n|\beta-\beta_0|>\delta$ for a sequence r_n such that $r_n\to\infty$ and $r_n/\sqrt{n}\to 0$. Similar approaches were taken by Wu (1981) and Saikkonen (1995) among others. The latter established the consistency of the maximum likelihood estimator of nonlinear transformation of β in the linear model. Second, the consistency of the short-run parameter estimates is established by the standard consistency argument using a uniform law of large numbers.

We assume the following for the consistency of the estimator $\hat{\theta}^*$.

Assumption 1 *(a)* $\{u_t\}$ *is an independent and identically distributed sequence with* $Eu_t=0, Eu_tu_t'=\Sigma$ *that is positive definite.*
(b) $\{\Delta x_t, z_t\}$ *is a sequence of strictly stationary strong mixing random variables with mixing numbers* α_m, $m=1,2,\ldots,$ *that satisfy* $\alpha_m=o\left(m^{-(\alpha_0+1)/(\alpha_0-1)}\right)$ *as* $m\to\infty$ *for some* $\alpha_0\geq 1$, *and for some* $\varepsilon>0$, $E|X_tX_t'|^{\alpha_0+\varepsilon}<\infty$ *and* $E|X_{t-1}u_t|^{\alpha_0+\varepsilon}<\infty$. *Furthermore,* $E\Delta x_t=0$ *and the partial sum process,* $x_{[ns]}/\sqrt{n}, s\in[0,1]$, *converges weakly to a vector Brownian motion* $\mathbf{B}$ *with a covariance matrix* $\mathbf{\Omega}$, *which is the long-run covariance matrix of* Δx_t *and has rank* $p-1$ *such that* $\left(1,\beta_0'\right)\ \mathbf{\Omega}=0$. *In particular, assume that* $x_{2[ns]}/\sqrt{n}$ *converges weakly to a vector Brownian motion* B *with a covariance matrix* Ω, *which is finite and positive definite.*
(c) the parameter space Θ *is compact and bounded away from zero for* λ_z *and there is a function* $\tilde{d}(x)$ *that is monotonic, integrable, and symmetric around zero, and by which* $\sup_{\gamma\in\Theta}|d(x,\gamma)-1\{x>0\}|$ *is bounded.*
(d) Let $u_t(\xi,\lambda,\gamma)$ *be defined as in* (1) *replacing* $z_t(\beta)$ *with* $z_t+\xi$, *where* ξ *belongs to a compact set in* $\mathbb{R}$ *and let*

$$S(\xi,\lambda,\gamma)=E\left(u_t'(\xi,\lambda,\gamma)\,u_t(\xi,\lambda,\gamma)\right).$$

Then, assume that $\frac{1}{n}\sum_t u_t(\xi,\lambda,\gamma)'\,u_t(\xi,\lambda,\gamma)\xrightarrow{p}S(\xi,\lambda,\gamma)$ *uniformly in* (ξ,λ,γ) *on any compact set and* $S(\xi,\lambda,\gamma)$ *is continuous in all its arguments and it is uniquely minimized at* $(\xi,\lambda,\gamma)=(0,\lambda_0,\gamma_0)$.

Condition (a) is common as in Chan (1993). It simplifies our presentation but could be relaxed. While we assume the stationarity and mixing conditions for $\{\Delta x_t, z_t\}$, Saikkonen (2007) provide more primitive conditions on $\{u_t\}$ and the coefficients. They are more stronger than each regime satisfying the standard conditions in the linear VECM. We also focus on the processes without the linear time trend by assuming that $E\Delta x_t = 0$.

Unlike for nonlinear models with stationary variables, the consistency proof for the nonlinear error correction models is difficult to be established in a general level. It depends crucially on the shape of the nonlinear transformation of the error correction term when the variable takes large values, see Park and Phillips (2001). Condition (c) identifies the shape, which is piece-wise linear in large values of the error correction terms. It is clear that the indicator functions and logistic functions satisfy (c). It distinguishes the current work from previous ones. de Jong (2002) considered the nonlinearity only through a bounded function and Kristensen and Rahbek (2008) through an unbounded function, which becomes linear for the large values of the error correction z_{t-1}. Thus, they do not capture the regime-specific behavior, which makes the consistency proof much different from the previous ones. We do not consider more general functional forms discussed in Saikkonen (2005) and Escribano and Mira (2002). The condition for λ_z is not necessary but convenient for our proof and does not appear to be much restrictive. We note that the case with $\lambda_z = 0$ is similar to the model studied by de Jong (2002) as the error correction term appears only in a bounded function in this case. We also comment on the case where the threshold variable is $|z_{t-1}(\beta)|$. The proof of the consistency goes through almost the same but an additional assumption on λ_z such that $A_z + D_z \neq 0$ will facilitate the direct application of the proof since $1\{|\zeta_t| > \gamma\} \sim 1\{|\zeta_t| > 0\}$ for an integrated process ζ_t.

The conditions in (d) are a standard set of conditions that are imposed to ensure the consistency of nonlinear least squares estimators. More specific set of sufficient conditions to ensure the uniform law of large numbers can be found in Andrews (1987) or Pöscher and Prucha (1991), for instance. It can be easily checked that the commonly used smooth transition functions and the indicator functions for threshold models satisfy such conditions. It implicitly impose the condition that $D_0 \neq 0$ as in the standard threshold model. The identification condition for ξ here is to identify the cointegrating vector at the square root n neighborhood and it is also imposed in de Jong (2002). The conditions in Assumption 1 do not guarantee the existence of a measurable least squares estimator since we do not impose the continuity of the function d. In this case, we can still establish the consistency based on the convergences in outer measure, see e.g. Newey and McFadden (1994). To ease the exposition, we implicitly assume the measurability in the theorem below.

Theorem 1 *Under Assumption 1,* $\hat{\theta}^* - \theta_0$ *is* $o_p(1)$ *and furthermore,* $\sqrt{n}\left(\hat{\beta}^* - \beta_0\right) = o_p(1)$.

When the transition function $d_{t-1}(\beta, \gamma)$ satisfies certain smoothness condition, the asymptotic distribution of $\hat{\theta}^*$ can be derived following the standard approach using the Taylor series expansion. de Jong (2002) explored minimization estimators with nonlinear objective function that involves the error correction term. It derived the asymptotic distributions of such estimators under the assumption that $\sqrt{n}\left(\hat{\beta}^* - \beta_0\right) = O_p(1)$. Thus, we refer to de Jong (2002) for the case with a smooth $d_{t-1}(\beta, \gamma)$. It is worth noting that the asymptotic distribution of the short-run parameter estimates is in general dependent on the estimation error of the cointegrating vector despite its super-consistency due to the nonlinearity of the model. On the other hand, the threshold model has not been studied due to the irregular feature of the indicator function, although the model has been adopted more widely in empirical research. We turn to the so-called threshold cointegration model and develop an asymptotics for the model in the next section.

3 Threshold Cointegration Model

Balke and Fomby (1997) introduced the threshold cointegration model, which corresponds to model (1) with $d(z_t(\beta),\gamma) = 1\{z_t(\beta) > \gamma\}$, to allow for nonlinear and/or asymmetric adjustment process to the equilibrium. That is,

$$\Delta x_t = \begin{cases} A_0 + A_z\ z_{t-1}(\beta) + A_1\ \left(\Delta x'_{t-1}, \cdots, \Delta x'_{t-l+1}\right)' + u_t, & \text{if } z_{t-1}(\beta) \leq \gamma \\ B_1 + B_z\ z_{t-1}(\beta) + B_1\ \left(\Delta x'_{t-1}, \cdots, \Delta x'_{t-l+1}\right)' + u_t, & \text{if } z_{t-1}(\beta) > \gamma \end{cases},$$

where $B = A + D$. The motivation of the model was that the magnitude and/or the sign of the disequilibrium z_{t-1} plays a central role in determining the short-run dynamics (see e.g. Taylor 2001). Thus, they employed the error correction term as the threshold variable. This threshold variable makes the estimation problem highly irregular as the cointegrating vector subjects to two different sorts of nonlinearity. Even when the cointegrating vector is prespecified, the estimation is nonstandard. We introduce a smoothed estimator and study the asymptotic properties of both smoothed and unsmoothed estimators in the following subsections.

To resolve the irregularity of the indicator function, Seo and Linton (2007) introduced a smoothed least squares estimator. To describe the estimator, define a bounded function $\mathcal{K}(\cdot)$ satisfying that

$$\lim_{s\to-\infty} \mathcal{K}(s) = 0, \lim_{s\to+\infty} \mathcal{K}(s) = 1.$$

A distribution function is often used for $\mathcal{K}$. Let $\mathcal{K}_t(\beta,\gamma) = \mathcal{K}\left(\frac{z_t(\beta)-\gamma}{h}\right)$, where $h \to 0$ as $n \to \infty$. To define the smoothed objective function, we replace $d_{t-1}(\beta,\gamma)$ in (1) with $\mathcal{K}_{t-1}(\beta,\gamma)$ and define the matrix $X_\gamma(\beta)$ that stacks $X_{t-1}(\beta)\mathcal{K}_{t-1}(\beta,\gamma)$. Then, we have the smoothed objective function

$$S_n(\theta) = (y - [(X(\beta), X_\gamma(\beta)) \otimes I_p]\lambda)'(y - [(X(\beta), X_\gamma(\beta)) \otimes I_p]\lambda). \tag{3}$$

And, the Smoothed Least Squares (SLS) estimator is defined as

$$\hat{\theta} = \underset{\theta\in\Theta}{\arg\min} S_n(\theta).$$

Similarly as the concentrated LS estimator, we can define

$$\lambda(\beta,\gamma) = \left(\begin{bmatrix} X(\beta)' X(\beta) & X(\beta)' X_\gamma(\beta) \\ X_\gamma(\beta)' X(\beta) & X_\gamma(\beta)' X_\gamma(\beta) \end{bmatrix}^{-1} \begin{pmatrix} X(\beta)' \\ X_\gamma(\beta)' \end{pmatrix} \otimes I_p\right) y, \tag{4}$$

and by profiling we can minimize $S_n(\theta)$ with respect to (β, γ).

It is worth mentioning that the true model is a threshold model and we employ the smoothing only for the estimation purpose. Since

$$\mathcal{K}\left(\frac{z_t(\beta)-\gamma}{h}\right) \to 1\{z_t(\beta) > \gamma\}$$

as $h \to 0$, $S_n(\theta)$ converges in probability to the probability limit of $S_n^*(\theta)$ as $n \to \infty$.

We make the following assumptions regarding the smoothing function $\mathcal{K}$ and the bandwidth parameter h.

Assumption 2 *(a) $\mathcal{K}$ is twice differentiable everywhere, $\mathcal{K}^{(1)}$ is symmetric around zero, $\mathcal{K}^{(1)}$ and $\mathcal{K}^{(2)}$ are uniformly bounded and uniformly continuous. Furthermore, $\int \left|\mathcal{K}^{(1)}(s)\right|^4 ds$, $\int \left|\mathcal{K}^{(2)}(s)\right|^2 ds$, and $\int \left|s^2\mathcal{K}^{(2)}(s)\right| ds$ are finite.*
(b) For some integer $\vartheta \geq 1$ and each integer i $(1 \leq i \leq \vartheta)$, $\int \left|s^i\mathcal{K}^{(1)}(s)\right| ds < \infty$, and

$$\begin{aligned} \int s^{i-1}\operatorname{sgn}(s)\,\mathcal{K}^{(1)}(s)\,ds &= 0, \\ \int s^{\vartheta}\operatorname{sgn}(s)\,\mathcal{K}^{(1)}(s)\,ds &\neq 0, \end{aligned}$$

and $\mathcal{K}(x) - \mathcal{K}(0) \gtreqless 0$ if $x \gtreqless 0$.
(c) Furthermore, for some $\varepsilon > 0$,

$$\begin{aligned} \lim_{n\to\infty} h^{i-\vartheta} \int_{|hs|>\varepsilon} \left|s^i\mathcal{K}^{(1)}(s)\right| ds &= 0, \\ \lim_{n\to\infty} h^{-1} \int_{|hs|>\varepsilon} \left|\mathcal{K}^{(2)}(s)\right| ds &= 0. \end{aligned}$$

(d) The sequence $\{h\}$ satisfies that for some sequence $m \geq 1$,

$$\begin{aligned} nh^3 &\to 0, \\ \log(nm)\left(n^{1-6/r}h^2m^{-2}\right)^{-1} &\to 0, \end{aligned}$$

and

$$h^{-9k/2-3}n^{(9k/2+2)/r+\varepsilon}\alpha_m \to 0$$

where k is the dimension of θ and $r > 4$ is specified in Assumption 4.

These conditions are imposed in Seo and Linton (2007) and common in smoothed estimation as in Horowitz (1992) for example. Condition (b) is an analogous condition to that defining the so-called ϑ^{th} order kernel, and requires a kernel $\mathcal{K}^{(1)}$ that permits negative values when $\vartheta > 1$ and $\mathcal{K}(0) = 1/2$. We impose the condition that $\mathcal{K}(x) - \mathcal{K}(0) \gtreqless 0$ if $x \gtreqless 0$ as we need negative kernels for $\vartheta > 1$. Condition (c) is standard. The standard normal cumulative distribution function clearly satisfies these conditions and see Seo and Linton (2007) for an example with $\vartheta > 1$. Condition (d) serves to determine the rate for h. While this range of rates is admissible, we do not have a sharp bound and thus no optimal rate. It the data were independent and identically distributed, the conditions simplify to $\left(nh^2\right)^{-1}\log n \to 0$ and $nh^3 \to 0$. As will be shown in the following section, the smaller h implies the faster convergence, whereas it may destroy the asymptotic normality if it is too small. This condition is not relevant for the consistency of $\hat{\theta}$. On the other hand, too large h introduces correlation between the threshold estimate $\hat{\gamma}$ and the slope estimate $\hat{\lambda}$.

The following corollary establishes the consistency of the smoothed least squares estimator.

Corollary 2 *Under Assumption 1 and 2, $\hat{\theta} - \theta_0$ is $o_p(1)$ and furthermore, $\sqrt{n}\left(\hat{\beta} - \beta_0\right)$ is $o_p(1)$.*

3.1 Convergence Rates and Asymptotic Distributions

The unsmoothed LS estimator of the threshold parameter is super-consistent in the standard stationary threshold regression and has complicated asymptotic distribution, which depends not only on certain moments but on the whole distribution of data. On the contrary, the smoothed LS estimator of the same parameter exhibits asymptotic normality, while the smoothing slows down the convergence rate. The nonstandard nature of the estimation of threshold models becomes more complicated in threshold cointegration since the thresholding relies on the error correction term, which is estimated simultaneously with the threshold parameter γ. We begin with developing the convergence rates of the unsmoothed estimators of the cointegrating vector β and the threshold parameter γ and then explore the asymptotic distribution of the smoothed estimators.

The asymptotic behavior of the threshold estimator heavily relies on whether the model is continuous or not. We focus on the discontinuous model. The following is assumed.

Assumption 3 (a) *For almost every Δ_t, the probability distribution of z_t conditional on Δ_t has everywhere positive density with respect to Lebesque measure.*
(b) $E\left[X_{t-1}'D_0D_0'X_{t-1}|z_{t-1} = \gamma_0\right] > 0.$

The condition (a) ensures that the threshold parameter γ is uniquely identified and is common in threshold autoregressions. While it is more complicated to verity the condition in general, it is easy to see that the threshold VECM without any lagged term satisfies it since it entails threshold autoregression in z_t. This remark is also relevant to Assumption 4 (c). The discontinuity of the regression function is assumed in the condition (b). At the threshold point γ_0, the change in the regression function is nonzero, which makes the regression function discontinuous. This enables a super-efficient estimation of the threshold parameter γ. If the regression function is continuous, Gonzalo and Wolf (2005) showed that the least squares estimator of the threshold parameter is root n consistent and asymptotically normal in the context of stationary threshold autoregression, which may be used to test for the condition (b).

We obtain the following rate result for the unsmoothed estimator of β and γ.

Theorem 3 *Under Assumption 1 and 3, $\hat{\beta}^* = \beta_0 + O_p\left(n^{-3/2}\right)$ and $\hat{\gamma}^* = \gamma_0 + O_p\left(n^{-1}\right)$.*

It is surprising that the cointegrating vector estimate converges faster than the standard n-rate. Heuristically, $\acute{\gamma} = \hat{\gamma} - x_{2t-1}'\left(\hat{\beta} - \beta_0\right)$ behaves like a threshold estimate in a stationary threshold model as

$$1\left\{z_{t-1}(\beta) > \gamma\right\} = 1\left\{z_{t-1} > \acute{\gamma}\right\}.$$

Since $\sup_{2\leq t\leq n}|x_{t-1}| = O_p\left(n^{1/2}\right)$ and the threshold estimate is super-consistent, it is expected that $\hat{\beta} - \beta_0 = O_p\left(n^{-3/2}\right)$. This fast rate of convergence has an important inferential implication for the short-run parameters as will be discussed later.

We turn to the smoothed estimator for the inference of the cointegrating vector. While subsampling is shown to be valid to approximate the asymptotic distribution of the unsmoothed LS estimator of the threshold parameter in the stationary threshold autoregression (see Gonzalo and Wolf 2005), the extension to the threshold cointegration is not trivial due to the involved nonstationarity. The smoothing of the objective function enables us to develop the asymptotic distribution based on the standard Taylor series expansion. Let $f(\cdot)$ denote the density of z_t and $f(\cdot|\Delta)$ the conditional density given $\Delta_t = \Delta$. Also define

$$\tilde{\mathcal{K}}_1(s) = \mathcal{K}^{(1)}(s)\left(1\{s>0\} - \mathcal{K}(s)\right)$$

and

$$\sigma_v^2 = E\left[\left\|\mathcal{K}^{(1)}\right\|_2^2\left(X_{t-1}'D_0u_t\right)^2 + \left\|\tilde{\mathcal{K}}_1\right\|_2^2\left(X_{t-1}'D_0D_0'X_{t-1}\right)^2|z_{t-1} = \gamma_0\right]f(\gamma_0) \quad (5)$$

$$\sigma_q^2 = \mathcal{K}^{(1)}(0)\,E\left(X_{t-1}'D_0D_0'X_{t-1}|z_{t-1} = \gamma_0\right)f(\gamma_0). \quad (6)$$

First, we set out assumptions that we need to derive the asymptotic distribution.

Assumption 4 *(a)* $E[|X_tu_t'|^r] < \infty$, $E[|X_tX_t'|^r] < \infty$, *for some* $r > 4$,
(b) $\{\Delta x_t, z_t\}$ *is a sequence of strictly stationary strong mixing random variables with mixing numbers* α_m, $m = 1, 2, \ldots,$ *that satisfy* $\alpha_m \leq Cm^{-(2r-2)/(r-2)-\eta}$ *for positive* C *and* η,*as* $m \to \infty$.
(c) For some integer $\vartheta \geq 2$ *and each integer* i *such that* $1 \leq i \leq \vartheta - 1$, *all* z *in a neighborhood of* γ, *almost every* Δ, *and some* $M < \infty$, $f^{(i)}(z|\Delta)$ *exists and is a continuous function of* z *satisfying* $\left|f^{(i)}(z|\Delta)\right| < M$. *In addition,* $f(z|\Delta) < M$ *for all* z *and almost every* Δ.
(d) The conditional joint density $f(z_t, z_{t-m}|\Delta_t, \Delta_{t-m}) < M$, *for all* (z_t, z_{t-m}) *and almost all* (Δ_t, Δ_{t-m}).
(e) θ_0 *is an interior point of* Θ.

These assumptions are analogous to those imposed in Seo and Linton (2007) that study the SLS estimator of the threshold regression model. The condition (a) ensures the convergence of the variance covariance estimators but can be weakened. We need stronger mixing condition as set out in (b) than that required for consistency. The conditions (c) - (e) are common in the smoothed estimation as in Horowitz (1992), only (d) being an analogue of a random sample to a dependent sample. In particular, we require more stringent smoothness condition (condition (c)) for the conditional density $f(z|\Delta)$ for the smoothed estimation than for the unsmoothed estimation.

We present the asymptotic distribution below.

Theorem 4 *Suppose Assumption 1 - 4 hold. Let W denote a standard Brownian motion that is independent of B. Then,*

$$\begin{pmatrix} nh^{-1/2}\left(\hat{\beta}-\beta_0\right) \\ \sqrt{nh^{-1}}\left(\hat{\gamma}-\gamma_0\right) \end{pmatrix} \Rightarrow \frac{\sigma_v}{\sigma_q^2}\begin{pmatrix} \int_0^1 BB' & \int_0^1 B \\ \int_0^1 B' & 1 \end{pmatrix}^{-1}\begin{pmatrix} \int BdW \\ W(1) \end{pmatrix}$$

$$\sqrt{n}\left(\hat{\lambda}-\lambda_0\right) \Rightarrow \mathcal{N}\left(0,\left[E\begin{pmatrix} 1 & d_{t-1} \\ d_{t-1} & d_{t-1} \end{pmatrix}\otimes X_{t-1}X_{t-1}'\right]^{-1}\otimes\Sigma\right),$$

and these two random vectors are asymptotically independent. The unsmoothed estimator $\hat{\lambda}^*$ has the same asymptotic distribution as $\hat{\lambda}$.

We make some remarks on the similarities to and differences from the linear cointegration model and the stationary threshold model. First, the asymptotic distribution of $\hat{\beta}$ and $\hat{\gamma}$ is mixed normal, the same asymptotic distribution as that of the standard OLS estimator of the cointegrating vector and the constant in the exogenous case up, to the scaling factor σ_v/σ_q^2. A reading of the proof of this theorem reveals that the linear part does not contribute to the asymptotic variance of $\hat{\beta}$ although β appears in both inside the indicator and the linear part of the model. The factor σ_v^2/σ_q^4 contains the conditional expectation and density at the discontinuity point. It is the asymptotic variance of threshold estimate if the true cointegrating vector were known. Other than the estimation of this factor, the inference can be made in the same way as in the standard OLS case. Second, the cointegrating vector converges faster than the usual n rate but slower than the $n^{3/2}$, which is obtained for the unsmoothed estimator. This is also the case for the estimators $\hat{\gamma}^*$ and $\hat{\gamma}$ for the threshold point γ. Third, as in the stationary threshold model, the slope parameter estimate $\hat{\lambda}$ is asymptotically independent of the estimation of β and γ.

The convergence rates of the estimators $\hat{\beta}$ and $\hat{\gamma}$ depend on the smoothing parameter h in a way that the smaller h accelerates the convergence. This is in contrast to the smoothed maximum score estimation. In the extremum case where $h = 0$, we obtain the fastest convergence, which corresponds to the unsmoothed estimator. The smaller h boosts the convergence rates by reducing the bias but too small a h destroys the asymptotic normality. We do not know the exact order of h where the asymptotic normality breaks down, which requires further research.

The asymptotic independence between the estimator $\hat{\beta}$ and the estimator $\hat{\lambda}$ of the slope parameter λ and the asymptotic normality of $\hat{\lambda}$ contrast the result in smooth transition cointegration models, where the asymptotic distribution of $\hat{\lambda}$ not only draws on the estimation of β but is non-Normal without certain orthogonality condition (see e.g. de Jong 2001, 2002).[2] This is due to the slower convergence of the estimators of β in the smooth transition models. Therefore, it should also be noted that the Engle-Granger type two-step approach, where

[2] In case of $Ez_t = 0$, we can still retain the asymptotic Normality of the slope estimate by estimating (1) after replacing the $z_{t-1}(\beta)$ with $\bar{z}_{t-1}\left(\tilde{\beta}\right) = \left(1,\tilde{\beta}'\right)\left(x_{t-1}-\frac{1}{n}\sum_s x_{s-1}\right)$, for any n-consistent $\tilde{\beta}$, as in de Jong (2001). It is worth noting, however, that this demeaning increases the asymptotic variance.

the cointegrating vector is estimated by the linear regression of x_{1t} on x_{2t} and the estimate is plugged in the error correction model, does not work in our case in the sense that the estimation error affects the asymptotic distribution of $\hat{\lambda}$. Therefore, the above independence result is useful for the construction of confidence interval for the slope parameter λ.

Furthermore, we may propose a two-step approach for the inference of the short-run parameters making use of the fact that the unsmoothed estimator $\hat{\beta}^*$ converges faster than the smoothed estimator $\hat{\beta}$. In principle, we can treat $\hat{\beta}^*$ as if it is the true value β_0. The following corollary states this.

Corollary 5 *Suppose Assumption 1 - 4 hold. Let $\hat{\gamma}(\beta)$ be the smoothed estimator of γ when β is given. Then, $\hat{\gamma}\left(\hat{\beta}^*\right)$ has the same asymptotic distribution as that of $\hat{\gamma}(\beta_0)$, which is $\mathcal{N}\left(0, \frac{\sigma_v^2}{\sigma_q^4}\right)$.*

3.2 Asymptotic Variance Estimation

The construction of confidence interval for the slope parameter λ is straightforward as $\hat{\lambda}$ and $\hat{\lambda}^*$ are just OLS estimators given (β, γ). We may treat the estimates $\hat{\beta}$ and $\hat{\gamma}$ (or $\hat{\beta}^*$ and $\hat{\gamma}^*$) as if they are β_0 and γ_0 due to Theorem 4. We may use either $1\{\hat{z}_{t-1} > \hat{\gamma}\}$ or $\mathcal{K}_{t-1}\left(\hat{\beta}, \hat{\gamma}\right)$ for d_{t-1}. The inference for (β, γ) requires to estimate Ω, σ_v^2, and σ_q^2.[3] The estimation of Ω can be done by applying a standard method of HAC estimation to Δx_t, see e.g., Andrews (1991). Although σ_v^2 and σ_q^2 involve nonparametric objects like conditional expectation and density, we do not have to do a nonparametric estimation as those are limits of the first and second derivatives of the objective function with respect to the threshold parameter γ.

Thus, let

$$\hat{\tau}_t = \frac{1}{2\sqrt{h}} X_{t-1}\left(\hat{\beta}\right)' D\mathcal{K}_{t-1}^{(1)}\left(\hat{\beta}, \hat{\gamma}\right)\hat{u}_t, \tag{7}$$

where $\hat{u}_t$ is the residual from the regression (1), and let

$$\hat{\sigma}_v^2 = \frac{1}{n}\sum_t \hat{\tau}_t^2, \text{ and } \hat{\sigma}_q^2 = \frac{h}{2n} Q_{n22}\left(\hat{\theta}\right),$$

where Q_{n22} is the diagonal element corresponding to γ of the Hessian matrix Q_n, see Appendix for the explicit formulas. Consistency of $\hat{\sigma}_q^2$ is straightforward from the proof of Theorem 4 and that of $\hat{\sigma}_v^2$ can be obtained after a slight modification of Theorem 4 of Seo and Linton (2007).

We can construct confidence interval for γ based on Corollary 5. The estimation of σ_v^2 and σ_q^2 can be done as above with $\beta = \hat{\beta}^*$. Due to the asymptotic normality and independence, the construction of confidence interval is much simpler this way without the need to estimate Ω.

Even though $\hat{\lambda}$ and $\hat{\lambda}^*$ are asymptotically independent of $\left(\hat{\beta}, \hat{\gamma}\right)$ and $\left(\hat{\beta}^*, \hat{\gamma}^*\right)$, they are dependent in finite samples. So, we may not benefit from the imposition of the block diagonal

[3] In the MLE of linear VECM, the likelihood ratio statistic for a hyhothesis on β converges to the Chi-square distribution as the log likelihood function under normality is approximately quadratic. It will be interesting to examine if the same holds true for the threshold cointegration model.

feature of the asymptotic variance matrix. Corollary 5 enables the standard way of constructing confidence interval based on the inversion of t-statistic with jointly estimated covariance matrix. In this case, we may define τ_t in (7) using the score of $u_t(\theta)$ with respect to (γ, λ) for a given $\hat{\beta}^*$. See Seo and Linton (2007) for a more discussion.

4 Monte Carlo Experiments

This section investigates the finite sample performance of the estimators explored in this paper. Of particular interest are the various estimators of the cointegrating vector β and the threshold parameter γ. We compare the unsmoothed LS estimator $\hat{\beta}^*$ of β with the SLS estimator $\hat{\beta}$ and the Johansen's maximum likelihood estimator $\tilde{\beta}$, which is based on the linear VECM. For comparison purpose, we also compute the restricted estimators $\hat{\beta}_0^*$ and $\hat{\beta}_0$, which are the unsmoothed LS and SLS estimators of β when γ is fixed at the true value γ_0. Similarly, $\hat{\gamma}_0^*$ and $\hat{\gamma}_0$ denote the restricted unsmoothed LS and SLS estimators of γ when β is prespecified at the true value β_0. To distinguish the SLS estimator $\hat{\gamma}$ from the two-step SLS estimator, let $\hat{\gamma}_2$ denote the two-step estimator.

The simulation samples are generated from the following process

$$\begin{aligned}
\Delta x_t &= \begin{pmatrix} -1 \\ 0 \end{pmatrix} (x_{1t-1} - \beta_0 x_{2t-1}) + \begin{pmatrix} -2 \\ 0 \end{pmatrix} 1\{x_{1t-1} - \beta_0 x_{2t-1} \leq \gamma_0\} \\
&\quad + \begin{pmatrix} -0.5 \\ 0 \end{pmatrix} (x_{1t-1} - \beta_0 x_{2t-1}) 1\{x_{1t-1} - \beta_0 x_{2t-1} > \gamma_0\} + u_t,
\end{aligned}$$

where $u_t \sim iid\mathcal{N}(0, I_2)$, $t = 0, ..., n$, and $\Delta x_0 = u_0$. This process was considered in Hansen and Seo (2002), who provide us with the finite sample performance of the maximum likelihood estimator of β and γ. We fix $\beta_0 = 1$, and $\gamma_0 = 0$. While the data generating process does not contain any lagged first difference term, the model is estimated with two lagged terms in addition to the error correction term. The estimation is based on the grid search with grid sizes for β and γ being 100 and 500, respectively. The grid for β is set around the Johansen's maximum likelihood estimator $\tilde{\beta}$ as in Hansen and Seo (2002). For the smoothed estimators, we use the standard normal distribution function for $\mathcal{K}$ and set $h = \hat{\sigma} n^{-1/2} \log n$, where $\hat{\sigma}^2$ is the sample variance of the error correction term.

Table 1 summarizes the results of our experiments with various sample sizes $n = 100, 250,$ and 500. We examined the finite sample distributions of the various estimators in terms of mean, root mean squared error (RMSE), mean absolute error (MAE), and selected percentiles from 1000 simulation replications. The RMSE and MAE are reported in log.

The results appear as we expected. The Johansen's maximum likelihood estimator $\tilde{\beta}$ does not perform as well as all the other estimators, which are obtained from estimating the correctly specified threshold cointegration model. The unsmoothed estimators and the restricted estimators outperform the smoothed and the unrestricted counterparts, respectively, in terms of RMSE and MAE. However, careful examination of percentiles reveals that the smoothed estimator $\hat{\beta}$ exhibits the smaller length of interval between five percentile and ninety-five per-

	Mean	RMSE in log	MAE in log	Percentile(%) 5	25	50	75	95
$n = 100$								
$\widetilde{\beta} - \beta_0$	0.001	-2.841	-3.191	-0.083	-0.028	0.000	0.028	0.098
$\widehat{\beta}^* - \beta_0$	0.001	-2.880	-3.293	-0.082	-0.023	-0.001	0.023	0.091
$\widehat{\beta}_0^* - \beta_0$	-0.001	-3.009	-3.617	-0.063	-0.015	-0.001	0.011	0.062
$\widehat{\beta} - \beta_0$	-0.001	-2.854	-3.268	-0.085	-0.025	0.001	0.025	0.084
$\widehat{\beta}_0 - \beta_0$	-0.002	-2.947	-3.470	-0.068	-0.020	-0.001	0.019	0.068
$\widehat{\gamma}^* - \gamma_0$	-0.672	0.138	-0.284	-2.649	-1.153	-0.293	-0.032	0.258
$\widehat{\gamma}_0^* - \gamma_0$	-0.631	0.080	-0.408	-2.617	-1.023	-0.209	-0.038	0.123
$\widehat{\gamma} - \gamma_0$	-0.461	0.149	-0.265	-2.702	-0.945	-0.058	0.241	0.700
$\widehat{\gamma}_0 - \gamma_0$	-0.444	0.090	-0.380	-2.719	-0.862	-0.021	0.192	0.504
$\widehat{\gamma}_2 - \gamma_0$	-0.497	0.146	-0.272	-2.802	-1.014	-0.086	0.219	0.582
$n = 250$								
$\widetilde{\beta} - \beta_0$	0.000	-3.960	-4.241	-0.030	-0.011	0.000	0.011	0.031
$\widehat{\beta}^* - \beta_0$	0.001	-4.302	-4.637	-0.022	-0.007	0.001	0.007	0.023
$\widehat{\beta}_0^* - \beta_0$	0.000	-4.639	-5.133	-0.014	-0.003	0.000	0.003	0.014
$\widehat{\beta} - \beta_0$	0.000	-4.278	-4.626	-0.022	-0.007	0.000	0.007	0.020
$\widehat{\beta}_0 - \beta_0$	0.000	-4.540	-4.866	-0.019	-0.006	0.000	0.006	0.016
$\widehat{\gamma}^* - \gamma_0$	-0.116	-0.928	-1.778	-0.734	-0.108	-0.031	0.015	0.156
$\widehat{\gamma}_0^* - \gamma_0$	-0.102	-1.007	-2.085	-0.526	-0.069	-0.021	-0.002	0.069
$\widehat{\gamma} - \gamma_0$	-0.057	-0.888	-1.677	-0.746	-0.073	0.019	0.113	0.236
$\widehat{\gamma}_0 - \gamma_0$	-0.049	-0.901	-1.826	-0.622	-0.043	0.030	0.100	0.184
$\widehat{\gamma}_2 - \gamma_0$	-0.054	-0.917	-1.704	-0.769	-0.071	0.025	0.105	0.236
$n = 500$								
$\widetilde{\beta} - \beta_0$	0.000	-2.006	-2.147	-0.015	-0.005	0.000	0.006	0.015
$\widehat{\beta}^* - \beta_0$	-0.000	-2.279	-2.425	-0.009	-0.003	-0.000	0.002	0.008
$\widehat{\beta}_0^* - \beta_0$	-0.000	-2.491	-2.698	-0.005	-0.001	-0.000	0.001	0.005
$\widehat{\beta} - \beta_0$	-0.000	-2.253	-2.383	-0.009	-0.003	-0.000	0.003	0.008
$\widehat{\beta}_0 - \beta_0$	-0.000	-2.335	-2.476	-0.007	-0.003	-0.000	0.002	0.008
$\widehat{\gamma}^* - \gamma_0$	-0.006	-1.165	-1.346	-0.115	-0.035	-0.004	0.020	0.096
$\widehat{\gamma}_0^* - \gamma_0$	-0.010	-1.438	-1.636	-0.068	-0.022	-0.007	0.002	0.045
$\widehat{\gamma} - \gamma_0$	0.032	-1.080	-1.199	-0.092	-0.013	0.030	0.076	0.155
$\widehat{\gamma}_0 - \gamma_0$	0.031	-1.213	-1.319	-0.051	-0.007	0.031	0.061	0.122
$\widehat{\gamma}_2 - \gamma_0$	0.028	-1.102	-1.229	-0.086	-0.015	0.025	0.068	0.146

Notation. Johansen's maximum likelihood estimator, $\tilde{\beta}$, the unsmoothed estimators, $\hat{\cdot}^*$, the restricted unsmoothed estimators, $\hat{\cdot}_0^*$, the smoothed estimators, $\hat{\cdot}$, the restricted smoothed estimators, $\hat{\cdot}_0$, and finally the two-step estimator $\hat{\gamma}_2$.

Table 1: Distribution of Estimators

centile than the unsmoothed estimator $\hat{\beta}^*$. And the percentiles of all the estimators of β seem very symmetric around the medians, which are mostly zeros.

Similar observation is made for the comparison among the estimators of the threshold parameter γ. The unsmoothed estimators have smaller RMSE and MAE than the smoothed ones. The knowledge on the true value of β helps reduce RMSE and MAE of the estimators of γ. The two-step estimator $\hat{\gamma}_2$, which employs more efficient estimator $\hat{\beta}^*$ than $\hat{\beta}$, indeed improves upon $\hat{\gamma}$ and even has the smaller RMSE than the restricted estimator $\hat{\gamma}_0$ when $n = 250$. The distributions of all the estimators of γ appear quite asymmetric and have large negative biases when $n = 100$. However, for $n = 500$ the distributions become more or less symmetric around the medians and the biases are much smaller than those for $n = 100$. We also note that the biases of the unsmoothed estimators are much bigger than those of smoothed estimators when $n = 100$ and $n = 250$.

5 Conclusion

We have established the consistency of the LS estimators of the cointegrating vector in general regime switching VECMs, validating the application of some of existing results on the joint estimation of long-run and short-run parameters for models with smooth transition. We also provided asymptotic inference methods for threshold cointegration, establishing the convergence rates and asymptotic distributions of the LS and SLS estimators of the model parameters. In particular, the theory and Monte Carlo experiments indicate that the inference on the threshold parameter can be improved upon by using the two-step estimator.

While we only considered two-regime models, our results might be extended to multiple-regime models, provided that the assumptions on stationarity and invariance principle in Assumption 1 hold true. In that case, we may consider the sequential estimation strategy discussed in Bai and Perron (1998) and Hansen (1999) . A sequence of estimations and tests can determine the number of regimes and the threshold parameter. The LM test by Hansen and Seo (2002) can be employed to test for the presence of the second break without modification due to the fast rate of convergence of the cointegrating vector estimators.

It is also possible to think of the case with more than one cointegrating relation if p is greater than 2. In this case, the threshold variable can be understood as a linear combination of those cointegrating vectors. However, the models commonly used in empirical applications are bivariate and the estimation of such a model is more demanding and thus we leave it for a future research.

Proof of Theorems

A word on notation. Throughout this section, C stands for a generic constant that is finite.

The proof of Theorem 1 makes use of the following lemma, which might be of independent interest.

Lemma 6 *Let $\{Z_n\}$ be a sequence of random variables and $\{r_n\}$ be a sequence of positive numbers such that $r_n \to \infty$. If $a_n Z_n = o_p(1)$ for any sequence $\{a_n\}$ such that $a_n/r_n \to 0$,* *then*

$$r_n Z_n = O_p(1).$$

Proof of Lemma 6 Let $X_n = r_n Z_n$ and suppose $X_n \neq O_p(1)$. Then, there exists $\varepsilon > 0$ such that for any K,

$$\limsup_{n\to\infty} \Pr(|X_n| > K) > \varepsilon.$$

Thus, there exists n_1 such that $\Pr(|X_{n_1}| > 1) > \varepsilon$. Similarly one can find $n_2 > n_1$ such that $\Pr(|X_{n_2}| > 2) > \varepsilon$, and $n_3 < n_4 < \cdots$, accordingly. Let $b_n = 1$ for $n \leq n_1$, $b_n = 2$ for $n_1 < n \leq n_2$, etc. Then, it is clear from the construction that $b_n \to \infty$, and that $\Pr(|X_n| > b_n) > \varepsilon$, infinitely often (*i.o.*), which implies that $X_n/b_n \neq o_p(1)$. However, given the condition of the lemma, $X_n/b_n = (r_n/b_n) Z_n = o_p(1)$ since $(r_n/b_n)/r_n \to 0$. This yields contradiction. ■

Proof of Theorem 1

Let $\Theta_{r_n,\delta} = \{\theta \in \Theta : r_n |\beta - \beta_0| > \delta\}$. Supremums and infimums in this proof are taken on the set $\Theta_{r_n,\delta}$, unless specified otherwise. To show that $\hat{\beta}^* - \beta_0 = o_p(r_n^{-1})$ we need to show that for every $\delta > 0$,

$$\Pr\left\{\inf_\theta S_n^*(\theta)/n - S_n^*(\theta_0)/n > 0\right\} \to 1. \tag{8}$$

Let $X_{\beta,\gamma}^* = \left(X(\beta), X_\gamma^*(\beta)\right) \otimes I_p$ and rewrite (2) as

$$S_n^*(\theta) = y'y + \lambda' X_{\beta,\gamma}^{*\prime} X_{\beta,\gamma}^* \lambda - 2y' X_{\beta,\gamma}^* \lambda.$$

Let $\eta = \sqrt{n}(\beta - \beta_0)$ and r_n be a sequence of real numbers such that $\sqrt{n} \geq r_n \to \infty$ and $r_n/\sqrt{n} \to 0$ as $n \to \infty$. Then,

$$|\eta| = r_n |\beta - \beta_0| \left(\sqrt{n}/r_n\right) \geq \delta\sqrt{n}/r_n \to \infty \tag{9}$$

for any $\beta \in \Theta_{r_n,\delta}$.

Note that

$$\frac{1}{n}S_n^*(\theta) - \frac{1}{n}S_n^*(\theta_0) = \frac{1}{n}\lambda' X_{\beta,\gamma}^{*\prime} X_{\beta,\gamma}^* \lambda - \frac{2}{n}y' X_{\beta,\gamma}^* \lambda + \frac{1}{n}y'y - \frac{1}{n}S_n^*(\theta_0)$$

and that $y'y/n - S_n^*(\theta_0)/n = y'y/n - u'u/n$ converges in probability to a positive constant as in the standard linear regression due to Assumption 1 (c) and this term is free of the parameter θ. Therefore, it is sufficient to show that

$$\begin{aligned}
& \Pr\left\{\inf_\theta \left(\frac{1}{n}\lambda' X_{\beta,\gamma}^{*\prime} X_{\beta,\gamma}^* \lambda - 2\frac{1}{n}y' X_{\beta,\gamma}^* \lambda\right) \geq 0\right\} \\
= & \Pr\left\{\inf_\theta \left(|\eta| \frac{1}{n}\lambda' \frac{X_{\beta,\gamma}^{*\prime} X_{\beta,\gamma}^*}{|\eta|^2}\lambda - 2\frac{1}{n}\frac{y' X_{\beta,\gamma}^*}{|\eta|}\lambda\right) \geq 0\right\} \\
\to & \ 1,
\end{aligned} \tag{10}$$

as $n \to \infty$. This follows from (9) if we show that (i)

$$\sup_{\theta} \left| \frac{1}{n} \frac{y' X^*_{\beta,\gamma}}{|\eta|} \lambda \right| = O_p(1), \tag{11}$$

and (ii) $\inf_\theta \frac{1}{n} \lambda' \frac{X^{*\prime}_{\beta,\gamma} X^*_{\beta,\gamma}}{|\eta|^2} \lambda$ converges weakly to a random variable that is positive with probability one.

Show (i) first. Note that $y' X^*_{\beta,\gamma}/n$ consists of the sample means of the product of Δx_t and $\left(1, z_{t-1}(\beta), \Delta'_{t-1}\right)$ and that of Δx_t and $\left(1, z_{t-1}(\beta), \Delta'_{t-1}\right) d_{t-1}(\beta, \gamma)$. However, as λ and $d_{t-1}(\beta,\gamma)$ are bounded, it is sufficient to observe that $\frac{1}{n}\sum_t |\Delta x'_t|$, and $\frac{1}{n}\sum_t \left|\Delta x_t \Delta'_{t-1}\right|$ are $O_p(1)$, and that

$$\frac{1}{n|\eta|} \sum_t |z_{t-1}(\beta) \Delta x'_t| \le \frac{1}{n\delta} \sum_t |z_{t-1} \Delta x'_t| + \frac{1}{n} \sum_t \left| \Delta x_t \frac{x'_{2t-1}}{\sqrt{n}} \right| = O_p(1),$$

by the law of large numbers for $|z_{t-1}\Delta x'_t|$, the weak convergence of $x_{2t}/\sqrt{n}$ and the Cauchy-Schwarz inequality.

We consider (ii) now. Let $\dot\eta = \eta/|\eta|$. Since $\frac{1}{n} X^{*\prime}_{\beta,\gamma} X^*_{\beta,\gamma}$ is the matrix of the sample means of the outer product of

$$\left(\left(1, z_{t-1}(\beta), \Delta'_{t-1}\right), \left(1, z_{t-1}(\beta), \Delta'_{t-1}\right) d_{t-1}(\beta,\gamma)\right)',$$

we may write

$$\frac{1}{n|\eta|^2} \lambda' X^{*'}_{\beta,\gamma} X^*_{\beta,\gamma} \lambda = \lambda'_z \left(\Xi^*_n(\beta,\gamma) \otimes I_p\right) \lambda_z + R_n(\theta),$$

where

$$\Xi^*_n(\beta,\gamma) = \begin{pmatrix} \frac{1}{n^2}\sum_t \left(x'_{2t-1}\dot\eta\right)^2 & \frac{1}{n^2}\sum_t \left(x'_{2t-1}\dot\eta\right)^2 d_{t-1}(\beta,\gamma) \\ \frac{1}{n^2}\sum_t \left(x'_{2t-1}\dot\eta\right)^2 d_{t-1}(\beta,\gamma) & \frac{1}{n^2}\sum_t \left(x'_{2t-1}\dot\eta\right)^2 d^2_{t-1}(\beta,\gamma) \end{pmatrix},$$

and

$$\sup_\theta |R_n(\theta)| = O_p\left(\left|\frac{r_n}{\sqrt{n}}\right|^2\right),$$

by the same reasoning to show (i). Since λ_z is bounded away from zero by Assumption 1 (c), it is sufficient to show that

$$\Xi^*_n(\beta,\gamma) \Rightarrow \left(\dot\eta'\Omega\dot\eta\right) \begin{pmatrix} \int_0^1 W^2 & \int_0^1 W^2 1\{W>0\} \\ \int_0^1 W^2 1\{W>0\} & \int_0^1 W^2 1\{W>0\} \end{pmatrix}, \tag{12}$$

where W is the standard Brownian motion. Note that the matrix in (12) is positive definite and free of parameters up to a constant multiple $\left(\dot\eta'\Omega\dot\eta\right)$, which is positive and bounded away from zero.

Now we show (12). It follows from Assumption 1 (b) and the continuous mapping theorem that

$$\frac{1}{n^2}\sum_t \left(x'_{2t-1}\dot\eta\right)^2 \Rightarrow \int_0^1 \left(B'\dot\eta\right)^2 =^d \int_0^1 W^2\left(\dot\eta'\Omega\dot\eta\right), \tag{13}$$

and that

$$\frac{1}{n^2}\sum_t \left(x_{2t-1}'\dot{\eta}\right)^2 1\left\{x_{2t-1}'\dot{\eta}>0\right\} \Rightarrow \int_0^1 \left(B'\dot{\eta}\right)^2 1\left\{B'\dot{\eta}>0\right\} =^d \left(\dot{\eta}'\Omega\dot{\eta}\right)\int_0^1 W^2 1\left\{W>0\right\}.$$

Then, it remains to show that

$$\sup_\theta \left|\frac{1}{n^2}\sum_t \left(x_{2t-1}'\dot{\eta}\right)^2 \left(1\left\{x_{2t-1}'\dot{\eta}>0\right\} - d\left(z_{t-1}\left(\beta\right),\gamma\right)\right)\right| = o_p\left(1\right) \tag{14}$$

and that

$$\sup_\theta \left|\frac{1}{n^2}\sum_t \left(x_{2t-1}'\dot{\eta}\right)^2 \left(1\left\{x_{2t-1}'\dot{\eta}>0\right\} - d^2\left(z_{t-1}\left(\beta\right),\gamma\right)\right)\right| = o_p\left(1\right). \tag{15}$$

Since $\left(1\left\{x>0\right\}-g^2\left(x\right)\right) \leq \left(\sup_x\left|g\left(x\right)\right|+1\right)\left|1\left\{x>0\right\}-g\left(x\right)\right|$ for any bounded function g and $\sup_{\dot{\eta},t}\left(x_{2t-1}'\dot{\eta}/\sqrt{n}\right)^2 = O_p\left(1\right)$, it is sufficient to show that

$$\sup_\theta \frac{1}{n}\sum_t \left|d\left(z_{t-1}\left(\beta\right),\gamma\right) - 1\left\{x_{2t-1}'\dot{\eta}>0\right\}\right| \leq R_{1n}+R_{2n} = o_p\left(1\right),$$

where

$$\begin{aligned} R_{1n} &= \sup_\theta \frac{1}{n}\sum_t \left|1\left\{z_{t-1}\left(\beta\right)>0\right\} - 1\left\{x_{2t-1}'\dot{\eta}>0\right\}\right| \\ R_{2n} &= \sup_\theta \frac{1}{n}\sum_t \left|d\left(z_{t-1}\left(\beta\right),\gamma\right) - 1\left\{z_{t-1}\left(\beta\right)>0\right\}\right|. \end{aligned}$$

Due to (9) and the fact that $z_{t-1}\left(\beta\right) = z_{t-1} + \frac{x_{2t-1}'\dot{\eta}}{\sqrt{n}}\left|\eta\right|$,

$$\frac{1}{n}\sum_t \left|1\left\{z_{t-1}\left(\beta\right)>0\right\} - 1\left\{x_{2t-1}'\dot{\eta}>0\right\}\right| \leq \frac{1}{n}\sum_t 1\left\{\left|\frac{x_{2t-1}'\dot{\eta}}{\sqrt{n}}\right| \leq \left|\frac{\delta r_n z_{t-1}}{\sqrt{n}}\right|\right\}. \tag{16}$$

For any $m_n>0$,

$$\begin{aligned} \sup_{|\dot{\eta}|=1} 1\left\{\left|\frac{x_{2[nr]}'\dot{\eta}}{\sqrt{n}}\right| \leq \left|\frac{\delta r_n z_{t-1}}{\sqrt{n}}\right|\right\} &\leq 1\left\{\inf_{|\dot{\eta}|=1}\left|\frac{x_{2[nr]}'\dot{\eta}}{\sqrt{n}}\right| \leq \left|\frac{\delta r_n z_{t-1}}{\sqrt{n}}\right|\right\} \\ &\leq 1\left\{\inf_{|\dot{\eta}|=1}\left|\frac{x_{2[nr]}'\dot{\eta}}{\sqrt{n}}\right| \leq m_n\right\} + 1\left\{\left|\frac{\delta r_n z_{t-1}}{\sqrt{n}}\right| > m_n\right\} \end{aligned} \tag{17}$$

Consider the first term in (17). Since $\frac{x_{2[nr]}'\dot{\eta}}{\sqrt{n}}$ converges weakly on $\left\{0\leq r\leq 1\right\}\times\left\{\dot{\eta}:\left|\dot{\eta}\right|=1\right\}$,

$$E1\left\{\inf_{|\dot{\eta}|=1}\left|\frac{x_{2[nr]}'\dot{\eta}}{\sqrt{n}}\right| \leq m_n\right\} \to \Pr\left\{\inf_{|\dot{\eta}|=1}\left|B_r'\dot{\eta}\right| \leq 0\right\} = 0,$$

for any $r\in[0.1]$ and for any decreasing sequence $m_n\to 0$. And for the second term in (17), we observe that $E1\left\{\left|\frac{\delta r_n z_{t-1}}{\sqrt{n}}\right| > m_n\right\}$ is the same for all t and that it goes to zero as

$m_n\sqrt{n}/r_n \to \infty$. Consider the right hand side term of (16). Letting $m_n \to 0$ and $m_n\sqrt{n}/r_n \to \infty$, we observe that

$$E\frac{1}{n}\sum_t 1\left\{\left|\frac{x'_{2t-1}\dot{\eta}}{\sqrt{n}}\right| \leq \left|\frac{\delta r_n z_{t-1}}{\sqrt{n}}\right|\right\} = \int_0^1 E1\left\{\left|\frac{x'_{2[nr]}\dot{\eta}}{\sqrt{n}}\right| \leq \left|\frac{\delta r_n z_{[nr]}}{\sqrt{n}}\right|\right\} dr \to 0,$$

by the dominated convergence theorem. This shows that $R_{1n} = o_p(1)$.

Next, it follows from Assumption 1 (c) that

$$\sup_\theta \left| d\left(z_{t-1}(\beta), \gamma\right) - 1\left\{z_{t-1}(\beta) > 0\right\}\right| \leq \tilde{d}\left(\inf_\eta \left| z_{[nr]} + \frac{x'_{2[nr]}\eta}{\sqrt{n}}\right|\right), \tag{18}$$

where $[nr] = t-1$. Due to (9), for each r and for any sequence $\{\varepsilon_n\}$ such that $\varepsilon_n \to 0$ and $\varepsilon_n\sqrt{n}/r_n \to \infty$ as $n \to \infty$,

$$\begin{aligned}
& E\left[\tilde{d}\left(\inf_\eta \left| z_{[nr]} + \frac{x'_{2[nr]}\eta}{\sqrt{n}}\right|\right)\right] \\
\leq \quad & E\left[\tilde{d}\left(\delta\frac{\sqrt{n}}{r_n}\left(\inf_{|\dot{\eta}|=1}\left|\frac{x'_{2[nr]}\dot{\eta}}{\sqrt{n}}\right| - \varepsilon_n\right)\right) 1\left\{\sup_\eta \left|\frac{z_{[nr]}}{|\eta|}\right| \leq \varepsilon_n\right\} 1\left\{\inf_{|\dot{\eta}|=1}\left|\frac{x'_{2[nr]}\dot{\eta}}{\sqrt{n}}\right| > 2\varepsilon_n\right\}\right] \\
& + CE\left(1\left\{\left|\frac{z_{[nr]}}{\delta\sqrt{n}} r_n\right| > \varepsilon_n\right\} + 1\left\{\inf_{|\dot{\eta}|=1}\left|\frac{x'_{2[nr]}\dot{\eta}}{\sqrt{n}}\right| \leq 2\varepsilon_n\right\}\right).
\end{aligned}$$

As $\inf_{|\dot{\eta}|=1}\left|\frac{x'_{2[nr]}\dot{\eta}}{\sqrt{n}}\right| = O_p(1)$, $E1\left\{\inf_{|\dot{\eta}|=1}\left|\frac{x'_{2[nr]}\dot{\eta}}{\sqrt{n}}\right| < 2\varepsilon_n\right\} \to 0$. Due to the fact that $\varepsilon_n\sqrt{n}/r_n \to \infty$, we have $E1\left\{\left|\frac{z_{[nr]}}{\delta\sqrt{n}} r_n\right| > \varepsilon_n\right\} \to 0$, and

$$\begin{aligned}
& E\tilde{d}\left(\delta\frac{\sqrt{n}}{r_n}\left(\inf_{|\dot{\eta}|=1}\left|\frac{x'_{2[nr]}\dot{\eta}}{\sqrt{n}}\right| - \varepsilon_n\right)\right) 1\left\{\sup_\eta \left|\frac{z_{[nr]}}{|\eta|}\right| \leq \varepsilon_n\right\} 1\left\{\inf_{|\dot{\eta}|=1}\left|\frac{x'_{2[nr]}\dot{\eta}}{\sqrt{n}}\right| > 2\varepsilon_n\right\} \\
\leq \quad & \tilde{d}\left(d\varepsilon_n\sqrt{n}/r_n\right) \to 0.
\end{aligned}$$

Thus, it follows from the dominated convergence theorem that the integral of (18) over r on the interval $[0,1]$ goes to zero. This shows $R_{2n} = o_p(1)$, thus, completing the proof of (8). Then, it follows from Lemma 6 that $\sqrt{n}\left(\hat{\beta}^* - \beta_0\right) = O_p(1)$.

We turn to show that $\hat{\theta} - \theta = o_p(1)$ and $\sqrt{n}\left(\hat{\beta}^* - \beta_0\right) = o_p(1)$. When $d(x,\gamma)$ is continuous in all its arguments, we can resort to Theorem 1 of de Jong (2002). And it can be readily generalized to the case where the limit objective function is continuous. Recalling Assumption 1 (d) and reading the proof of Theorem 1 of de Jong (2002), we see that we only have to show his equation $(A.9)$ with $a_n = 0$, which corresponds to

$$\frac{1}{n}\sum_t \left[u_t\left(\frac{x'_{2t-1}\eta}{\sqrt{n}}, \lambda, \gamma\right)' u_t\left(\frac{x'_{2t-1}\eta}{\sqrt{n}}, \lambda, \gamma\right) - S\left(\frac{x'_{2t-1}\eta}{\sqrt{n}}, \lambda, \gamma\right)\right] \xrightarrow{p} 0, \tag{19}$$

uniformly in (η, γ) on any compact set. However, for any η and γ and $C > 0$, the term in

(19) is bounded by

$$\sup_{\lambda,\gamma,|\xi|\leq C}\left|\frac{1}{n}\sum_t\left[u_t(\xi,\lambda,\gamma)'u_t(\xi,\lambda,\gamma)-S(\xi,\lambda,\gamma)\right]\right|+\frac{1}{n}\sum_t 1\left\{\sup_\eta\left|\frac{x_{2t-1}'\eta}{\sqrt{n}}\right|>C\right\},$$

whose first term converges to zero in probability by assumption and the second term can be made arbitrarily small by choosing C large enough due to the weak convergence of $\frac{x_{2t-1}'\eta}{\sqrt{n}}$. Therefore, the proof is complete. ■

Proof of Corollary 2

As in the proof of Theorem 1, we need to show that

$$\Pr\left\{\inf_\theta S_n(\theta)/n-S_n(\theta_0)/n>0\right\}\to 1,$$

where we follow the notational convention of the proof the theorem that infimums and supremums are assumed to be taken over $\Theta_{r_n,\delta}$ unless specified otherwise. As $S_n(\theta_0)/n$ does not contain any $I(1)$ variable, the result in Seo and Linton (2007) applies so that it is sufficient to show (10) with $X_{\beta,\gamma}^*$ replaced by $X_{\beta,\gamma}=(X(\beta),X_\gamma(\beta))\otimes I_p$. However, (11) is obvious as $\mathcal{K}$ is bounded. Thus, we need to show (12), that is,

$$\begin{aligned}\Xi_n(\beta,\gamma) &= \begin{pmatrix}\frac{1}{n^2}\sum_t\left(x_{2t-1}'\dot{\eta}\right)^2 & \frac{1}{n^2}\sum_t\left(x_{2t-1}'\dot{\eta}\right)^2\mathcal{K}_{t-1}(\beta,\gamma)\\ \frac{1}{n^2}\sum_t\left(x_{2t-1}'\dot{\eta}\right)^2\mathcal{K}_{t-1}(\beta,\gamma) & \frac{1}{n^2}\sum_t\left(x_{2t-1}'\dot{\eta}\right)^2\mathcal{K}_{t-1}^2(\beta,\gamma)\end{pmatrix}\\ &\Rightarrow \left(\dot{\eta}'\Omega\dot{\eta}\right)\begin{pmatrix}\int_0^1 W^2 & \int_0^1 W^2 1\{W>0\}\\ \int_0^1 W^2 1\{W>0\} & \int_0^1 W^2 1\{W>0\}\end{pmatrix},\end{aligned}$$

which follows if we show (14) and (15) with d replaced by $\mathcal{K}$. The proof hinges on (18), which is the part where d matters. However, it still holds replacing d with $\mathcal{K}$ and taking supremum over both θ and h on any interval including zero. This establishes that $\hat{\beta}$ is $\sqrt{n}$-consistent. Remaining part of the proof follows if the conditions on $\mathcal{K}$ satisfy Assumption 1 (d). In other words, defining $\tilde{u}_t(\xi,\lambda,\gamma)$ like $u_t(\xi,\lambda,\gamma)$ replacing the indicator with $\mathcal{K}$, we need to show the uniform convergence of its sample mean to $S(\xi,\lambda,\gamma)$, which follows from Seo and Linton (2007) as $\tilde{u}_t(\xi,\lambda,\gamma)$ consists of stationary variables. This completes the proof. ■

A word on notation. Having proved Theorem 1 and Corollary 2, we write hereafter $\theta=\left(\eta',\gamma,\lambda'\right)'$ and $\theta_1=(\eta',\gamma)'$ with slight abuse of notation. To further simplify notation assume $\gamma_0=0$ and thus θ_{10} is a vector of zeros.

Proof of Theorem 3

Let $\Theta_c=\{\theta:|\theta-\theta_0|<c\}$. Due to the consistency shown in Theorem 1, we may restrict the parameter space to Θ_c for some $c>0$, which will be specified below. It is sufficient to show the following claim for θ_1 to be n-consistent as in Chan (1993).

Claim I *For any $\varepsilon > 0$, there exist $c > 0$ and $K > 0$ such that*

$$\liminf_{n\to\infty} \Pr\left\{S_n(\theta) - S_n(0,\lambda) > 0 \text{ for all } \theta \in \Theta_{c,K}\right\} > 1 - \varepsilon,$$

where $\Theta_{c,K} = \Theta_c \cap \{\theta : |\theta_1| > K/n\}$.

Proof of Claim I: Let $\gamma_t = \gamma - \frac{x'_{2t}}{\sqrt{n}}\eta$ and $u_t(\theta) = u_{1t}(\theta) + u_{2t}(\theta)$, where

$$\begin{aligned} u_{1t}(\theta) &= u_t - (A - A_0)' X_{t-1} - (D - D_0)' X_{t-1} 1\left\{z_{t-1} > \gamma_{t-1}\right\} \\ &\quad - D_0' X_{t-1}\left(1\left\{z_{t-1} > \gamma_{t-1}\right\} - 1\{z_{t-1} > 0\}\right) \\ u_{2t}(\theta) &= -\left(A_z + D_z 1\left\{z_{t-1} > \gamma_{t-1}\right\}\right)\frac{x'_{2t-1}}{\sqrt{n}}\eta. \end{aligned}$$

Since $u_{2t}(0,\lambda) = 0$, $(S_n(\theta) - S_n(0,\lambda))/n = D_{1n}(\theta) + D_{2n}(\theta)$, where

$$\begin{aligned} D_{1n}(\theta) &= \frac{1}{n}\sum_t \left[u_{1t}(\theta)' u_{1t}(\theta) - u'_{1t}(0,\lambda) u_{1t}(0,\lambda)\right] \\ D_{2n}(\theta) &= \frac{1}{n}\sum_t u_{2t}(\theta)' u_{2t}(\theta) - \frac{2}{n}\sum_t u_{1t}(\theta)' u_{2t}(\theta). \end{aligned}$$

Note that $\vec{x}_{2n} = \sup_{t<n}\left|\frac{x_{2t-1}}{\sqrt{n}}\right| = O_p(1)$ and thus $\vec{\gamma}_n = \sup_{t<n}\left|\gamma_{t-1}\right| = O_p(|\theta_1|) = O_p(c)$ for $\theta \in \Theta_c$. Then,

$$\left|\frac{1}{n}\sum_t u_{2t}(\theta)' u_{2t}(\theta)\right| \leq O\left(\left[\vec{x}_{2n}\right]^2 |\eta|^2\right) = O_p(c|\eta|). \tag{20}$$

Since $\{u_t\}$ is a martingale difference sequence, so is the sequence $\left\{u_t 1\left\{z_{t-1} > \gamma_{t-1}\right\}\right\}$, implying

$$\frac{1}{n}\sum_t u_t 1\left\{z_{t-1} > \gamma_{t-1}\right\}\frac{x'_{2t-1}}{\sqrt{n}} = o_p(1),$$

see e.g. Hansen (1992). This implies that

$$\left|\frac{1}{n}\sum_t u_{1t}(\theta)' u_{2t}(\theta)\right| \leq o_p(|\eta|) + O_p(c|\eta|) + o_p(|\eta|), \tag{21}$$

because $\left|\frac{1}{n}\sum_t X_{t-1} 1\left\{z_{t-1} > \gamma_{t-1}\right\}\frac{x'_{2t-1}}{\sqrt{n}}\right| \leq \frac{1}{n}\sum_t\left|X_{t-1}\frac{x'_{2t-1}}{\sqrt{n}}\right| = O_p(1)$ and

$$\begin{aligned} &\frac{1}{n}\sum_t\left|X_{t-1}\left(1\left\{z_{t-1} > \gamma_{t-1}\right\} - 1\{z_{t-1} > 0\}\right) 1\left\{z_{t-1} > \gamma_{t-1}\right\}\frac{x'_{2t-1}}{\sqrt{n}}\right| \\ \leq\ & \vec{x}_{2n}\frac{1}{n}\sum_t |X_{t-1}|\ 1\left\{|z_{t-1}| \leq \left|\gamma_{t-1}\right|\right\} \\ =\ & O_p(1)\, O_p(c). \end{aligned}$$

The meaning of $o_p(|\eta|)$ is that the term can be bounded for all large n by the product of $|\eta|$ and an arbitrary small constant with probability $1-\varepsilon$ for any $\varepsilon > 0$. The same is true for

$O_p\left(c\left|\eta\right|\right)$ since c can be chosen arbitrary small. Thus, we conclude from (20) and (21) that for any $m, \varepsilon > 0$, there is $c > 0$ such that

$$\liminf_{n\to\infty} \Pr\left\{\left|D_{2n}\left(\theta\right)\right| \leq m\left|\theta_1\right| \text{ for all } \theta \in \Theta_c\right\} > 1-\varepsilon. \tag{22}$$

On the other hand, as we show below, for any $\varepsilon > 0$ and all sufficiently large n, there exists some constant $m' > 0$ such that $D_{1n}\left(\theta\right) > m'\left|\theta_1\right|$ with probability $1-\varepsilon$, which will complete the proof of Claim I as m is arbitrary. By direct calculation as above, we may write

$$\begin{aligned} D_{1n}\left(\theta\right) &= \left(D_0' + O_p\left(c\right)\right)\frac{1}{n}\sum_t X_{t-1}X_{t-1}' 1\left\{\left|z_{t-1}\right| \leq \left|\gamma_{t-1}\right|\right\} D_0 \\ &\quad + \left(D_0' + O\left(c\right)\right)\frac{1}{n}\sum_t X_{t-1}u_t 1\left\{\left|z_{t-1}\right| \leq \left|\gamma_{t-1}\right|\right\} + O_p\left(c^2\right). \end{aligned}$$

We first argue that the same as (22) can be said for the last two terms. It is obvious for the last term $O_p\left(c^2\right)$ and it is left to show that for any $\varepsilon, \delta > 0$ and sufficiently large n

$$\Pr\left\{\sup_{\theta\in\Theta_{c,K}}\left|\frac{1}{n}\sum_t u_t 1\left\{\left|z_{t-1}\right| \leq \left|\gamma_{t-1}\right|\right\}\right| / \left|\theta_1\right| > \delta\right\} < \varepsilon. \tag{23}$$

The other terms in X_{t-1} can be analyzed similarly. To show (23), we consider a grid

$$\Theta_b = \left\{\theta_1 = \left(b^{i_1}, ..., b^{i_k}\right)' K/n : \left|\theta_1\right| < c, \text{ and } i_1, ..., i_k = 1, 2, ...\right\},$$

for $b > 1$ and first show (23) when the supremum is taken over Θ_b. By the Markov inequality

$$\begin{aligned} &\Pr\left\{\sup_{\Theta_b}\left|\frac{1}{n}\sum_t u_t 1\left\{\left|z_{t-1}\right| \leq \left|\gamma_{t-1}\right|\right\}\right| / \left|K/n\left(b^{i_1}, ..., b^{i_k}\right)'\right| > \delta\right\} \\ \leq &\sum_{i_1,...,i_k} \frac{E\left|\frac{1}{n}\sum_t u_t 1\left\{\left|z_{t-1}\right| \leq \left|\gamma_{t-1}\right|\right\}\right|^2}{\left|K/n\left(b^{i_1}, ..., b^{i_k}\right)'\right|^2 \delta^2} \\ = &\sum_{i_1,...,i_k} \frac{1}{\delta^2 K^2 \left|\left(b^{i_1}, ..., b^{i_k}\right)\right|^2} \sum_t E u_t^2 1\left\{\left|z_{t-1}\right| \leq \left|\gamma_{t-1}\right|\right\}. \end{aligned} \tag{24}$$

However, since the conditional distribution, say, F_{t-1} of z_{t-1} given x_{2t-1} has a density, which is bounded by M, an expansion of it yields that

$$\sum_t EF_{t-1}\left(\left|\gamma_{t-1}\right|\right) \leq M\frac{1}{n}\sum_t E\left|\frac{x_{2t}}{\sqrt{n}}\right|^2 K\left|\left(b^{i_1}, ..., b^{i_k}\right)\right| = O\left(K\left|\left(b^{i_1}, ..., b^{i_k}\right)\right|\right). \tag{25}$$

Furthermore, $b > 1$ implies that $\sum_{i_1,...,i_k}\left|\left(b^{i_1}, ..., b^{i_k}\right)\right|^{-1} < \infty$. Thus, letting K large makes the term in (24) smaller than any given $\varepsilon > 0$.

Next, if a_1 and a_2 are any two adjacent points in Θ_b, then $\left|a_1 - a_2\right| \leq \left|a_1\right|\left(b-1\right)K/n \leq c\left(b-1\right)K/n$. Also, let γ_{1t} and γ_{2t} denote $\gamma_t's$ corresponding to any two points (say, θ_{11} and

$\theta_{12},$) lying between a_1 and a_2, then

$$\begin{aligned}
& \left|1\left\{|z_{t-1}| \leq \left|\gamma_{1t-1}\right|\right\} - 1\left\{|z_{t-1}| \leq \left|\gamma_{2t-1}\right|\right\}\right| \\
\leq \quad & 1\left\{\left|\gamma_{1t-1}\right| - \sup_{t,\theta_{11},\theta_{12}} |\gamma_{1t} - \gamma_{2t}| \leq |z_{t-1}| \leq \left|\gamma_{1t-1}\right| + \sup_{t,\theta_{11},\theta_{12}} |\gamma_{1t} - \gamma_{2t}|\right\}
\end{aligned}$$

and

$$\sup_{t,\theta_{11},\theta_{12}} |\gamma_{1t} - \gamma_{2t}| \leq \vec{x}_{2n}\ c\,(b-1)\,\frac{K}{n},$$

where the supremum is taken over $t < n$ and over θ_{11} and θ_{12} between a_1 and a_2. Since a_1 and a_2 were chosen arbitrary, the supremum can be extended to the collection of all θ_{11} and θ_{12} that lie between any two adjacent point in Θ_b and by the same reasoning as (25),

$$E \sup_{\theta_{11},\theta_{12}} \left|\sum_t u_t\left(1\left\{|z_{t-1}| \leq \left|\gamma_{1t-1}\right|\right\} - 1\left\{|z_{t-1}| \leq \left|\gamma_{2t-1}\right|\right\}\right)\right| = O\left(c\,(b-1)\,K\right).$$

Since b can be chosen arbitrarily close to 1, this completes the proof of (23).

Turning to the first term of D_{1n}, let F denote the distribution function of $|z_t|$ then it follows from the standard uniform law of large numbers that

$$\sup_{|x|,|\theta_1|\leq C} \frac{1}{n}\sum_t \left|1\left\{|z_{t-1}| \leq |x'\theta_1|\right\} - F\left(|x'\theta_1|\right)\right| \xrightarrow{p} 0,$$

for any $C < \infty$. Since C is arbitrary and $\vec{x}_{2n} = O_p(1)$,

$$\sup_{|\theta_1|\leq C} \frac{1}{n}\sum_t \left|1\left\{|z_{t-1}| \leq \left|\dot{\gamma}_{t-1}\right|\right\} - F\left(\left|\dot{\gamma}_{t-1}\right|\right)\right| \xrightarrow{p} 0. \tag{26}$$

Since $x_{2[nr]}/\sqrt{n} \Rightarrow B(r)$ and F is continuous, it follows from (26) that

$$\frac{1}{n}\sum_t 1\left\{|z_{t-1}| \leq \left|\dot{\gamma}_{t-1}\right|\right\} \Rightarrow \int_0^1 F\left(\left|\gamma - B(r)'\eta\right|\right)dr \overset{d}{=} \int_0^1 F\left(\left|\gamma + \eta'\Omega\eta W(r)\right|\right)dr,$$

where W is a standard Brownian motion and the equality in distribution follows from the normality of B. Then, for all large n,

$$\begin{aligned}
& \Pr\left\{\frac{1}{n}\sum_t 1\left\{|z_{t-1}| \leq \left|\dot{\gamma}_{t-1}\right|\right\} - m_5|\theta_1| \geq 0 \text{ for all } \theta_1 \in \Theta_c\right\} \\
\geq \quad & \Pr\left\{\int_0^1 F\left(\left|\gamma + \eta'\Omega\eta W(r)\right|\right)dr - m_5|\theta_1| \geq 0 \text{ for all } \theta_1 \in \Theta_c\right\} - \varepsilon/2 \\
\geq \quad & 1 - \varepsilon,
\end{aligned}$$

where the last inequality is shown below in several steps.

First, there is m_1 such that $F(z) \geq m_1 z$ for all $z \in [0,1]$ and $\sup_{0\leq r\leq 1}|W(r)| = O_p(1)$ and $|\gamma|$ and $|\eta|$ are bounded by c. Thus, choosing c small we can argue that for any $\varepsilon > 0$, with probability greater than $1-\varepsilon$,

$$\int_0^1 F\left(\left|\gamma+\eta'\Omega\eta W\left(r\right)\right|\right)dr \geq \int_0^1 m_1\left|\gamma+\eta'\Omega\eta W\left(r\right)\right|dr.$$

Second, since $|\gamma| \leq c$, if $|W(r)| \geq c$, then $|\gamma+\eta'\Omega\eta W(r)| \geq 2\eta'\Omega\eta\,|W(r)| \geq 2m_2\,|\eta|\,|W(r)|$, for some m_2 as Ω is positive definite. Thus, choosing c small, $\Pr\left\{\inf_{0\leq r\leq 1}|W(r)| \geq c\right\} > 1-\varepsilon$ and thus with probability greater $1-\varepsilon$,

$$\int_0^1 m_1\left|\gamma+\eta'\Omega\eta W\left(r\right)\right|dr \geq m_3\left|\eta\right|\int_0^1\left|W\left(r\right)\right|dr.$$

Third, for any $\varepsilon > 0$, there is m_4 such that $\Pr\left\{\int_0^1|W(r)|\,dr > m_4\right\} > 1-\varepsilon$. Thus, we can conclude that for any $\varepsilon > 0$, there exists a constant $m_5 > 0$ such that

$$\Pr\left\{\int_0^1 F\left(\left|\gamma+\eta'\Omega\eta W\left(r\right)\right|\right)dr - m_5\left|\theta_1\right| \geq 0\right\} > 1-\varepsilon/2.$$

■

Proof of Theorem 4

To derive the limit distribution of the SLS estimator $\hat{\theta}$, define $T_n(\theta) = \frac{\partial S_n(\theta)}{n\partial\theta}$ and $Q_n(\theta) = \frac{\partial^2 S_n(\theta)}{n\partial\theta\partial\theta'}$. Then, by the mean value theorem,

$$\sqrt{n}D_n^{-1}\left(\hat{\theta}-\theta_0\right) = \left(D_nQ_n\left(\tilde{\theta}\right)D_n\right)^{-1}\sqrt{n}D_nT_n\left(\theta_0\right),$$

where D_n is a diagonal matrix, whose first $p-1$ elements are $(h/n)^{1/2}$, the p-th element is $\sqrt{h}$, and the others are 1, and $\tilde{\theta}$ lies between $\hat{\theta}$ and θ_0. Recall that $X'_{t-1}(\eta)D = X'_{t-1}D + \frac{x'_{2t-1}\eta}{\sqrt{n}}D'_z$ and write the residuals of the SLS as

$$\begin{aligned} e_t\left(\theta\right) &= u_t - \left(A-A_0\right)'X_{t-1} - \left(D-D_0\right)'X_{t-1}d_{t-1} \\ &\quad - D'X_{t-1}\left(\mathcal{K}_{t-1}\left(\eta,\gamma\right)-d_{t-1}\right) - \left(A_z + D_z\mathcal{K}_{t-1}\left(\eta,\gamma\right)\right)\frac{x'_{2t-1}\eta}{\sqrt{n}}. \end{aligned} \tag{27}$$

Then,

$$\frac{\partial e_t\left(\theta\right)'}{\partial\eta} = -x_{2t-1}\left[A'_z + D'_z\mathcal{K}_{t-1}\left(\eta,\gamma\right) + \left(D'_z\frac{x'_{2t-1}\eta}{\sqrt{n}} + X'_{t-1}D\right)\frac{\mathcal{K}^{(1)}_{t-1}\left(\eta,\gamma\right)}{h}\right] \tag{28}$$

$$\frac{\partial e_t\left(\theta\right)'}{\partial\gamma} = -\left(X'_{t-1}D + D'_z\frac{x'_{2t-1}\eta}{\sqrt{n}}\right)\frac{\mathcal{K}^{(1)}_{t-1}\left(\eta,\gamma\right)}{h} \tag{29}$$

$$\frac{\partial e_t\left(\theta\right)'}{\partial\lambda} = \begin{pmatrix} -\left(X'_{t-1}D + D'_z\frac{x'_{2t-1}\eta}{\sqrt{n}}\right)\otimes I_p \\ -\mathcal{K}_{t-1}\left(\eta,\gamma\right)\left(X'_{t-1}D + D'_z\frac{x'_{2t-1}\eta}{\sqrt{n}}\right)\otimes I_p \end{pmatrix}.$$

Convergence of T_n

The asymptotic distribution of

$$\sqrt{n}D_nT_n(\theta_0)/2=\frac{1}{\sqrt{n}}D_n\sum_t\frac{\partial e_t(\theta_0)'}{\partial\theta}e_t(\theta_0)$$

has been developed in Seo and Linton (2007) except for the part corresponding to η. Thus, we focus on the convergence of

$$\frac{\sqrt{h}}{2n}\sum_t\frac{\partial e_t(\theta_0)'}{\partial\eta}e_t(\theta_0)=-\frac{1}{n}\sum_t x_{2t-1}(\sqrt{h}v_{1t}+\sqrt{h}v_{2t}+v_{3t}/\sqrt{h}),$$

where

$$\begin{aligned}v_{1t}&=A_{z0}'u_t+D_{z0}'\mathcal{K}_{t-1}u_t,\\ v_{2t}&=(A_{z0}'+D_{z0}'\mathcal{K}_{t-1})D_0'X_{t-1}(\mathcal{K}_{t-1}-d_{t-1}),\\ v_{3t}&=\mathcal{K}_{t-1}^{(1)}X_{t-1}'D_0[u_t-D_0'X_{t-1}(\mathcal{K}_{t-1}-d_{t-1})],\end{aligned}$$

and that of covariances of $\sqrt{n}D_nT_n(\theta_0)$.

Since v_{1t} is a martingale difference array, $\frac{1}{n}\sum_t x_{2t-1}v_{1t}=O_p(1)$ due to the convergence of stochastic integrals (see e.g. Kurtz and Protter 1991). The convergence of $n^{-1}h^{-1/2}\sum_t x_{2t-1}v_{2t}$ is similar to that of $\frac{1}{n\sqrt{h}}\sum_t x_{2t-1}v_{3t}$, which we will establish here. Then, it follows that

$$\frac{1}{n}\sum_t x_{2t-1}(v_{1t}+v_{2t})\sqrt{h}=o_p(1),$$

as $h\to0$. Let $\bar{v}_{3t}=(v_{3t}-Ev_{3t})/\sqrt{h}$, then $\bar{v}_{3t}$ is a zero mean strong mixing array. Seo and Linton (2007, Lemma 2) has shown that $\sqrt{n/h}Ev_{3t}\to0$ and var $\left[(hn)^{-1/2}\sum_t v_{3t}\right]=$ var $\left[v_{3t}/\sqrt{h}\right]+o(1)\to\sigma_v^2$, which is defined in (5). This implies that

$$\frac{1}{n}\sum_t x_{2t-1}Ev_{3t}/\sqrt{h}=o_p(1),$$

and that

$$n^{-1/2}\sum_{t=2}^{[nr]}\begin{pmatrix}\Delta x_{2t-1}\\ \bar{v}_{3t}\end{pmatrix}\Rightarrow\begin{pmatrix}B\\ \sigma_v^2W\end{pmatrix},\qquad(30)$$

due to Assumption 4 and the invariance principle of Wooldridge and White (1988, Theorem 2.11), where W is a standard Brownian motion that is independent of B. For the independence between B and W, see Lemma 2 of Seo and Linton (2007), which shows that $\sum_{s,t=1}^n E\Delta x_s\bar{v}_{3t}=o(1)$.

For the convergence of $\frac{1}{n}\sum_t x_{2t-1}\bar{v}_{3t}$, we resort to Hansen (1992, Theorem 3.1). Checking his conditions, we observe that the moment condition for $\bar{v}_{3t}$ is not met, that is, $E|\bar{v}_{3t}|^p\to\infty$ for $p>2$. However, it is not necessary but used to show that $\sup_{1\le t\le n}n^{-1/2}\sum_{k=1}^\infty E[\bar{v}_{3t+k}|\mathcal{F}_t]=o_p(1)$, where $\mathcal{F}_t$ is the natural filtration at time t. Using the Markov inequality, he obtains that for $p > 2$

$$\Pr\left\{\left|\sup_{1\le t\le n} n^{-1/2}\sum_{k=1}^{\infty} E\left[\bar{v}_{3t+k}|\mathcal{F}_t\right]\right| \ge \varepsilon\right\} \le \frac{CE\left|\bar{v}_{3t}\right|^p}{\varepsilon^p n^{p/2-1}} \to 0 \tag{31}$$

if $E\left|\bar{v}_{3t}\right|^p < \infty$. Now we show that while $E\left|\bar{v}_{3t}\right|^p$ is not bounded for $p > 2$ but diverges at a rate slower than $n^{p/2-1}$ so that (31) still holds. As $\sqrt{n/h}Ev_{3t} \to 0$ and the part associated with u_t can be done in the same manner, we focus on

$$\begin{aligned}
& E\left|\mathcal{K}_{t-1}^{(1)}X_{t-1}'D_0D_0'X_{t-1}\left(\mathcal{K}_{t-1}-d_{t-1}\right)\right|^p h^{-p/2} \\
= \; & h^{-p/2}\int \left|X'D_0D_0'X\right|^p \left|\mathcal{K}^{(1)}\left(\frac{z-\gamma_0}{h}\right)\left(\mathcal{K}\left(\frac{z-\gamma_0}{h}\right)-1\left(z>\gamma_0\right)\right)\right|^p f\left(z|X\right)dzdP_X \\
= \; & h^{1-p/2}\int \left|X'D_0D_0'X\right|^p \left|\mathcal{K}^{(1)}\left(s\right)\left(\mathcal{K}\left(s\right)-1\left(s\right)\right)\right|^p f\left(hs+\gamma_0|X\right)dsdP_X,
\end{aligned}$$

where P_X is the distribution of X_{t-1} and the last equality follows by the change-of-variables. Note that f is bounded almost every X, $\mathcal{K}(s) - 1(s)$ is bounded, $\left|\mathcal{K}^{(1)}\right|^p$ is integrable, and $E\left|X'D_0D_0'X\right|^p < \infty$. As $h^{1-p/2}/n^{p/2-1} \to 0$ for $p > 2$, we conclude that (31) converges to zero. Therefore,

$$\frac{\sqrt{h}}{2\sigma_v n}\sum_t \frac{\partial e_t\left(\theta_0\right)'}{\partial\eta}e_t\left(\theta_0\right) \Rightarrow \int_0^1 BdW. \tag{32}$$

Finally, note that a similar argument that showed the asymptotic independence in (30) yields the asymptotic independence between the scores for η and λ. For the covariance between the scores for η and γ, note that $v_{3t} = h\frac{\partial e_t(\theta_0)'}{\partial\gamma}e_t\left(\theta_0\right)$.

Convergence of Q_n

To begin with, we claim the following lemma.

Lemma 7 *Under the conditions of this theorem, $h^{-1}\left(\hat{\gamma}-\gamma_0\right)$ and $h^{-1}\hat{\eta}$ are $o_p(1)$.*

Proof of Lemma 7 Recalling the formula (27) and (29), we may write

$$\frac{1}{n}\sum_t \frac{\partial}{\partial\gamma}e_t\left(\theta\right)'e_t\left(\theta\right) = T_{1n}\left(\theta\right) + \cdots + T_{10n}\left(\theta\right), \tag{33}$$

where

$$\begin{aligned}
T_{1n}\left(\theta\right) &= -\frac{1}{nh}\sum_t \mathcal{K}_{t-1}^{(1)}\left(\eta,\gamma\right)X_{t-1}'Du_t, \\
T_{2n}\left(\theta\right) &= tr\left[D'\frac{1}{nh}\sum_t \mathcal{K}_{t-1}^{(1)}\left(\eta,\gamma\right)X_{t-1}X_{t-1}'\left(A-A_0\right)\right] \\
T_{3n}\left(\theta\right) &= tr\left[D'\frac{1}{nh}\sum_t X_{t-1}X_{t-1}'\mathcal{K}_{t-1}^{(1)}\left(\eta,\gamma\right)\left(\mathcal{K}_{t-1}\left(\eta,\gamma\right)-d_{t-1}\right)D\right],
\end{aligned}$$

and till we reach

$$T_{10n}(\theta) = \frac{1}{nh} D_z' \eta \sum_t \frac{x_{2t-1}}{\sqrt{n}} \mathcal{K}_{t-1}^{(1)}(\eta,\gamma)\left(A_z + D_z \mathcal{K}_{t-1}(\eta,\gamma)\right) \frac{x_{2t-1}'\eta}{\sqrt{n}}.$$

As we proved the consistency of the smoothed least squares estimator of θ in Corollary 2, we can find a sequence $r_n \to 0$ and that $\Pr\left\{\left|\hat{\theta} - \theta_0\right| > r_n\right\} \to 0$. Without loss of generality, we restrict the parameter space to $\Theta_{r_n} = \{|\theta - \theta_0| \leq r_n\}$. We first show that $\sup_{\Theta_{r_n}} |T_{in}(\theta)| = o_p(1)$, for all $i \neq 3$, which implies that

$$T_{3n}\left(\hat{\theta}\right) = o_p(1), \tag{34}$$

since (33) equals zero at $\theta = \hat{\theta}$ due to the first order condition of the SLS estimation.

First take $T_{1n}(\theta)$. In particular, consider

$$\left|\frac{1}{nh}\sum_t \mathcal{K}^{(1)}\left(\frac{z_{t-1} + x_{2t-1}'\eta/\sqrt{n} - \gamma}{h}\right) u_t\right|,$$

and the other terms in the vector can be handled similarly. As in Seo and Linton, we may assume u_t is bounded by some constant $C < \infty$ and divide the parameter space for θ_1 of Θ_{r_n} into Γ_n non-overlapping pieces $\Theta_{ni}, i = 1, ..., \Gamma_n$, and the distance between any two points in each piece is smaller than equal to $\mu h^{2+\varepsilon}$ for some $\mu > 0$. Then, $\Gamma_n = O\left(h^{-3(p-1)(1+\varepsilon/2)}\right)$. An application of a Hoeffding type inequality for martingale difference sequence (Azuma 1967) yields that for $\theta_{1i} \in \Theta_{ni}$ and some constant C_1

$$\begin{aligned}
& \Pr\left\{\sup_{i\leq\Gamma_n}\left|\frac{1}{nh}\sum_t u_t \mathcal{K}^{(1)}\left(\frac{z_{t-1} + x_{2t-1}'\eta_i/\sqrt{n} - \gamma_i}{h}\right)\right| > \eta\right\} \\
\leq & \sum_{i\leq\Gamma_n}\Pr\left\{\left|\frac{1}{nh}\sum_t u_t \mathcal{K}^{(1)}\left(\frac{z_{t-1} + x_{2t-1}'\eta_i/\sqrt{n} - \gamma_i}{h}\right)\right| > \eta\right\} \\
\leq & \Gamma_n \exp\left(-nh^2\eta^2/C_1\right) \\
\to & 0,
\end{aligned}$$

as $nh^2 \to \infty$. And for any θ_1 and θ_2 in Θ_{ni},

$$\begin{aligned}
& \frac{1}{nh}\sum_t \left[\mathcal{K}^{(1)}\left(\frac{z_{t-1} + x_{2t-1}'\eta_1/\sqrt{n} - \gamma_1}{h}\right) - \mathcal{K}^{(1)}\left(\frac{z_{t-1} + x_{2t-1}'\eta_2/\sqrt{n} - \gamma_2}{h}\right)\right] u_t \\
\leq & \frac{1}{nh}\sum_t \sup_{\eta,\gamma}\left|\mathcal{K}^{(2)}\left(\frac{z_{t-1} + x_{2t-1}'\eta/\sqrt{n} - \gamma}{h}\right) u_t \frac{x_{2t-1}}{\sqrt{n}}\right| \frac{\mu h^{2+\varepsilon}}{h} \\
\leq & O_p\left(h^{\varepsilon}\right),
\end{aligned}$$

as $\sup_{1\leq t\leq n}\left|\frac{x_{2t-1}}{\sqrt{n}}\right| = O_p(1)$ and $\mathcal{K}^{(2)}$ is bounded.

Next, we study the convergence of $T_{3n}(\theta)$. The same analysis applies to $T_{in}s$ for $i = 2, 4, 5, ..., 10$, to yield that they are all $o_p(1)\,s$. Note for example that T_{2n} is the product of $(A - A_0)$, which is bounded by r_n, and of a sample mean, which is $O_p(1)$ applying the same method as below.

Define

$$T_t(x,\theta_1) = X_{t-1}X_{t-1}'\mathcal{K}^{(1)}\left(\frac{z_{t-1}+x'\eta-\gamma}{h}\right)\left(\mathcal{K}\left(\frac{z_{t-1}+x'\eta-\gamma}{h}\right)-1\{z_{t-1}+x'\eta>\gamma\}\right),$$

and

$$\varsigma(x,\theta_1) = E(T_t(x,\theta_1)).$$

Furthermore, letting $\dot{\gamma}=\gamma-x'\eta$, define

$$\begin{aligned}\dot{T}_t(\dot{\gamma}) &= X_{t-1}X_{t-1}'\mathcal{K}^{(1)}\left(\frac{z_{t-1}-\dot{\gamma}}{h}\right)\left(\mathcal{K}\left(\frac{z_{t-1}-\dot{\gamma}}{h}\right)-1\{z_{t-1}>\dot{\gamma}\}\right),\\ \dot{\varsigma}(\dot{\gamma}) &= E\dot{T}_t(\dot{\gamma}).\end{aligned}$$

Assume that x belongs to a compact set Θ_x and thus $\dot{\gamma}$ lies within an interval Θ_{gn}, shrinking to zero as $|\theta_1|\le r_n\to 0$. Then, it is clear that

$$\sup_{x\in\Theta_x,\theta_1\in\Theta_{r_n}}\left|\frac{1}{nh}\sum_t(T_t(x,\theta_1)-\varsigma(x,\theta_1))\right| \le \sup_{\dot{\gamma}\in\Theta_{gn}}\left|\frac{1}{nh}\sum_t\left(\dot{T}_t(\dot{\gamma})-\dot{\varsigma}(\dot{\gamma})\right)\right| = o_p(1), \quad (35)$$

where the last equality is due to Lemma 4 (17) of Seo and Linton (2007). Since (35) holds for any compact set Θ_x and $\sup_t\left|n^{-1/2}x_{2t-1}\right|=O_p(1)$, (35) holds true when x is replaced with $n^{-1/2}x_{2t-1}$. This implies that

$$\sup_{\Theta_{r_n}}\left|T_{3n}(\theta)-tr\left[D'\frac{1}{nh}\sum_t\varsigma\left(\frac{x_{2t-1}}{\sqrt{n}},\theta_1\right)D\right]\right| = o_p(1), \quad (36)$$

and thus

$$\frac{1}{nh}\sum_t\varsigma\left(\frac{x_{2t-1}}{\sqrt{n}},\hat{\theta}_1\right)=o_p(1), \quad (37)$$

due to (34). Furthermore,

$$\frac{1}{nh}\sum_t\varsigma\left(\frac{x_{2t-1}}{\sqrt{n}},\theta_{10}\right)=\frac{1}{nh}\sum_t\dot{\varsigma}(0)=o(1), \quad (38)$$

where the last equality is due to (24) of Seo and Linton (2007). Expanding the $(1,1)$ element of (37) around θ_{10} yields that for $\tilde{\theta}_1$ between $\hat{\theta}_1$ and θ_{10}

$$\frac{1}{nh}\sum_t\varsigma_{11}\left(\frac{x_{2t-1}}{\sqrt{n}},\hat{\theta}_1\right)=\frac{1}{nh}\sum_t\varsigma_{11}\left(\frac{x_{2t-1}}{\sqrt{n}},\theta_{10}\right)+\frac{\partial}{\partial\theta_1}\frac{1}{nh}\sum_t\varsigma_{11}\left(\frac{x_{2t-1}}{\sqrt{n}},\tilde{\theta}_1\right)\left(\hat{\theta}_1-\theta_{10}\right),$$

where ς_{11} denotes the $(1,1)$ element of ς. Given (37) and (38), this implies that $h^{-1}\left(\hat{\theta}_1-\theta_{10}\right)=o_p(1)$ if

$$\frac{\partial}{\partial\theta_1}\frac{1}{n}\sum_t\varsigma_{11}\left(\frac{x_{2t-1}}{\sqrt{n}},\tilde{\theta}_1\right)\neq o_p(1). \quad (39)$$

However,

$$\begin{aligned}
& \frac{\partial}{\partial\theta_1} E\left[\mathcal{K}^{(1)}\left(\frac{z_{t-1}+x'\eta-\gamma}{h}\right)\left(\mathcal{K}\left(\frac{z_{t-1}+x'\eta-\gamma}{h}\right)-1\left\{z_{t-1}+x'\eta>\gamma\right\}\right)\right] \\
= \;& E\left[\mathcal{K}^{(2)}\left(\frac{z_{t-1}+x'\eta-\gamma}{h}\right)\left(\mathcal{K}\left(\frac{z_{t-1}+x'\eta-\gamma}{h}\right)-1\left\{z_{t-1}+x'\eta>\gamma\right\}\right)\frac{1}{h}\right](x',-1)' \\
& +\left[E\mathcal{K}^{(1)}\left(\frac{z_{t-1}+x'\eta-\gamma}{h}\right)^2\frac{1}{h}-\mathcal{K}^{(1)}(0)\, f(\gamma-x'\eta)\right](x',-1)',
\end{aligned}$$

where f denotes the density of z_t. Using change-of-variables formula,

$$\begin{aligned}
E\mathcal{K}^{(1)}\left(\frac{z_{t-1}-\dot{\gamma}}{h}\right)^2\frac{1}{h} &= \int_{-\infty}^{\infty}\mathcal{K}^{(1)}\left(\frac{z-\dot{\gamma}}{h}\right)^2\frac{1}{h}f(z)\,dz \\
&= \int_{-\infty}^{\infty}\mathcal{K}^{(1)}(z)^2 f(hz+\dot{\gamma})\,dz \\
&\to \int_{-\infty}^{\infty}\mathcal{K}^{(1)}(z)^2\,dz\cdot f(0),
\end{aligned}$$

by the dominated convergence theorem as h and $\dot{\gamma}$ go to zero. Similarly,

$$\begin{aligned}
& E\left[\mathcal{K}^{(2)}\left(\frac{z_{t-1}-\dot{\gamma}}{h}\right)\left(\mathcal{K}\left(\frac{z_{t-1}-\dot{\gamma}}{h}\right)-1\left\{z_{t-1}>\dot{\gamma}\right\}\right)\frac{1}{h}\right] \\
\to\; & \int_{-\infty}^{\infty}\mathcal{K}^{(2)}(z)\left(\mathcal{K}(z)-1\{z>0\}\right)dz\cdot f(0).
\end{aligned}$$

Furthermore, it follows from the integral by parts that

$$\int_{-\infty}^{\infty}\mathcal{K}^{(1)}(s)^2\,ds-\int_{-\infty}^{\infty}\mathcal{K}^{(2)}(s)\left(1\{s>0\}-\mathcal{K}(s)\right)ds=-\int_{-\infty}^{\infty}\mathcal{K}^{(2)}(s)\,1\{s>0\}\,ds=\mathcal{K}'(0).$$

Since $\sup_{1\le t\le n,\theta_1\in\Theta_{r_n}}\left|\frac{x'_{2t-1}}{\sqrt{n}}\eta-\gamma\right|=o_p(1)$, we obtain

$$\begin{aligned}
& \frac{\partial}{\partial\theta_1}\frac{1}{n}\sum_t \varsigma_{11}\left(\frac{x_{2t-1}}{\sqrt{n}},\tilde{\theta}_1\right) \\
= \;& 2\int_{-\infty}^{\infty}\mathcal{K}^{(2)}(z)\left(\mathcal{K}(z)-1\{z>0\}\right)dz\cdot f(0)\;\frac{1}{n}\sum_t\left(\frac{x'_{2t-1}}{\sqrt{n}},-1\right)'+o_p(1).
\end{aligned}$$

Thus, (39) follows since $\int_{-\infty}^{\infty}\mathcal{K}^{(2)}(z)\left(\mathcal{K}(z)-1\{z>0\}\right)dz>0$ and $f(0)>0$. ■

In view of Lemma 7 and the proof, we can restrict the parameter space to

$$\Theta_n=\{\theta\in\Theta:|\lambda-\lambda_0|<r_n, h^{-1}|\gamma-\gamma_0|<r_n, h^{-1}|\eta|<r_n\},$$

for a sequence $r_n\to 0$. Let

$$Q_n^a(\theta)=\frac{1}{n}\sum_t\frac{\partial e_t(\theta)'}{\partial\theta}\frac{\partial e_t(\theta)}{\partial\theta'},\;\; Q_n^b(\theta)=\frac{1}{n}\sum_t\sum_{i=1}^{p}\frac{\partial^2 e_{it}(\theta)}{\partial\theta\partial\theta'}e_{it}(\theta),$$

where the subscript i of a matrix (or a vector) indicates the i^{th} Column (element) of the matrix (the vector). Then, $Q_n(\theta) = 2Q_n^a(\theta) + 2Q_n^b(\theta)$. Start with $Q_n^b(\theta)$, in particular, with

$$\begin{aligned} & -\sum_t \frac{\partial^2 e_{it}(\theta)}{\partial\eta\partial\eta'} e_{it}(\theta) \\ = & \sum_t x_{2t-1}x_{2t-1}' \left(2D_{zi}\frac{\mathcal{K}_{t-1}^{(1)}(\eta,\gamma)}{h} + X_{t-1}(\eta)' D_i \frac{\mathcal{K}_{t-1}^{(2)}(\eta,\gamma)}{h^2}\right) e_{it}(\theta). \qquad (40) \end{aligned}$$

Since $h \to 0$ and $\sup_{1\le t\le n}\left|\frac{x_{2t-1}'\eta}{h\sqrt{n}}\right| = o_p(1)$, $X_{t-1}(\eta)' D_i = X_{t-1}' D_i + D_{z_i}\frac{x_{2t-1}'\eta}{\sqrt{n}} = X_{t-1}' D_i + o_p(1)$ and the leading term in (40) with the normalization is

$$\frac{h}{n^2}\sum_t x_{2t-1}x_{2t-1}' X_{t-1}' D_i \frac{\mathcal{K}_{t-1}^{(2)}(\eta,\gamma)}{h^2} e_{it}(\theta) \Rightarrow \tilde{\sigma}_{q_i}^2 \int \tilde{\mathcal{K}}_2 \int_0^1 BB', \qquad (41)$$

where $\tilde{\mathcal{K}}_2(s) = \mathcal{K}^{(2)}(s)(\mathcal{K}(s) - 1\{s>0\})$ and $\tilde{\sigma}_{q_i}^2 = E\left(D_{0i}' X_{t-1}' X_{t-1} D_{0i} | z_{t-1} = 0\right) f(0)$. The convergence (41) can be achieved in the same way as in Lemma 7. That is, decomposing it as in the lemma, we can easily show that $T_{in}'s$ are negligible except for $i = 3$. For T_{3n}, the same as (35) holds with

$$\varsigma(x,\theta) = \frac{1}{hn^2}\sum_t xx' E\left[X_{t-1}' D_i \mathcal{K}^{(2)}\left(\frac{z_{t-1}}{h} + \dot{\gamma}\right)\left(\mathcal{K}\left(\frac{z_{t-1}}{h} + \dot{\gamma}\right) - 1\left\{\frac{z_{t-1}}{h} + \dot{\gamma} > 0\right\}\right)\right].$$

where $\dot{\gamma} = \frac{x'\eta - \gamma}{h}$. Since $\sup_{t\le n, \theta\in\Theta_n}\left|\frac{x_{2t-1}'\eta - \gamma}{h}\right| = o_p(1)$ and by the change-of-variables technique

$$\frac{1}{h} E\left(X_{t-1}' D_i\right)^2 \mathcal{K}_{t-1}^{(2)}\left(\mathcal{K}_{t-1} - d_{t-1}\right) \to E\left(\left(D_i' X_{t-1}\right)^2 | z_{t-1} = 0\right) f(0) \int \tilde{\mathcal{K}}_2,$$

the convergence in (41) follows.

The remaining terms in $Q_n^b(\theta)$ are $\frac{\partial^2 e_{it}(\theta)}{\partial\lambda\partial\lambda'} = 0$,

$$\begin{aligned} \frac{\partial^2 e_{it}(\theta)}{\partial\gamma\partial\eta'} &= \left(D_{zi}\frac{\mathcal{K}_{t-1}^{(1)}(\eta,\gamma)}{h} + X_{t-1}(\eta)' D_i \frac{\mathcal{K}_{t-1}^{(2)}(\eta,\gamma)}{h^2}\right) x_{2t-1}', \\ \frac{\partial^2 e_{it}(\theta)}{\partial\gamma\partial\gamma'} &= -X_{t-1}(\eta)' D_i \frac{\mathcal{K}_{t-1}^{(2)}(\eta,\gamma)}{h^2}; \ \frac{\partial^2 e_{it}(\theta)}{\partial\lambda\partial\gamma'} = \begin{pmatrix} 0 \\ \frac{\mathcal{K}_{t-1}^{(1)}(\eta,\gamma)}{h} X_{t-1}(\eta) \end{pmatrix} \otimes I_i, \end{aligned}$$

and

$$\frac{\partial^2 e_{it}(\theta)}{\partial\lambda\partial\eta'} = -\begin{pmatrix} \epsilon_2 \\ \mathcal{K}_{t-1}(\eta,\gamma)\epsilon_2 + \frac{\mathcal{K}_{t-1}^{(1)}(\eta,\gamma)}{h} X_{t-1}(\eta) \end{pmatrix} x_{2t-1}' \otimes I_i$$

where $\epsilon_2 = (0, 1, 0, ..., 0)$ whose dimension is $(pl + 2)$. As their convergence can be analyzed similarly as above, we omit the details and conclude that

$$D_n Q_n^b D_n \Rightarrow \tilde{\sigma}_q^2 \int \tilde{\mathcal{K}}_2 \begin{pmatrix} -\int_0^1 BB' & \int_0^1 B & 0 \\ \int_0^1 B' & 1 & 0 \\ 0 & 0 & 0 \end{pmatrix},$$

where $\tilde{\sigma}_q^2 = \sum_{i=1}^p \tilde{\sigma}_{q_i}^2$. Similarly,

$$D_n Q_n^a D_n \Rightarrow \begin{bmatrix} \left\|\mathcal{K}^{(1)}\right\|_2^2 \tilde{\sigma}_q^2 \begin{pmatrix} \int_0^1 BB' & -\int_0^1 B \\ -\int_0^1 B' & 1 \end{pmatrix} & 0 \\ 0 & E\begin{pmatrix} 1 & d_{t-1} \\ d_{t-1} & d_{t-1} \end{pmatrix} \otimes X_{t-1}X_{t-1}' \otimes I_p \end{bmatrix}.$$

Finally, note that $\left\|\mathcal{K}^{(1)}\right\|_2^2 - \int \tilde{\mathcal{K}}_2 = \mathcal{K}^{(1)}(0)$ by an application of the integral by parts, which yields the desired result.

The convergence of $\hat{\lambda}^*$ is a direct consequence of Theorem 3, which obtains the convergence rates of $\hat{\gamma}^*$ and $\hat{\eta}^*$, and Theorem 5 of Seo and Linton (2007). ■

Proof of Corollary 5

Let $\eta_n = \eta + O_p\left(n^{-1}\right)$, $\theta_2 = (\gamma, \lambda)$. The consistency of $\hat{\theta}_2$ is obvious since we established the consistency in Corollary 2 based on the $\sqrt{n}$-consistency of $\hat{\beta}$. For the limit distribution, let $T_{2n}(\theta_2; \eta)$ and $Q_{2n}(\theta_2; \eta)$ denote the score and hessian of the sum of squares function with a fixed η. But, we have already derived the convergence of the Hessian $Q_n(\theta)$ for $\hat{\eta} = \eta_0 + o_p(1)$. Thus, we only have to examine the score T_{2n}. Let

$$\begin{aligned} \ddot{e}_t(\theta_2; \eta_n) &= u_t - (A - A_0)' X_{t-1} - (D - D_0)' X_{t-1} d_{t-1} \\ &\quad - D' X_{t-1}\left(\mathcal{K}_{t-1}(\eta_n, \gamma) - d_{t-1}\right) - \left(A_z + D_z \mathcal{K}_{t-1}(\eta_n, \gamma)\right) x_{2t-1}'(\eta_n - \eta_0), \\ \frac{\partial \ddot{e}_t(\theta_2)'}{\partial \theta_2} &= \begin{pmatrix} -\left(X_{t-1}' D + D_z' x_{2t-1}'(\eta_n - \eta_0)\right) \mathcal{K}_{t-1}^{(1)}(\eta_n, \gamma)/h \\ -\left(X_{t-1}' D + D_z' x_{2t-1}'(\eta_n - \eta_0)\right) \otimes I_p \\ -\mathcal{K}_{t-1}(\eta_n, \gamma)\left(X_{t-1}' D + D_z' x_{2t-1}'(\eta_n - \eta_0)\right) \otimes I_p \end{pmatrix} \end{aligned}$$

Then, we need to show that

$$\frac{\sqrt{h}}{\sqrt{n}} \sum_{t=1}^n \left| \frac{\partial \ddot{e}_t(\theta_{20}; \eta_n)'}{\partial \theta_2} \ddot{e}_t(\theta_{20}; \eta_n) - \frac{\partial \ddot{e}_t(\theta_{20}; \eta_0)'}{\partial \theta_2} \ddot{e}_t(\theta_{20}; \eta_0) \right| = o_p(1).$$

As the arguments are all similar for each term, we show that

$$\begin{aligned}
&\frac{\sqrt{h}}{\sqrt{n}}\sum_{t=1}^{n}\left|X_{t-1}'D_0D_0'X_{t-1}\left(\mathcal{K}_{t-1}\left(\eta_n,\gamma_0\right)-\mathcal{K}_{t-1}\right)\right| \\
\leq\; & \frac{\sqrt{n}}{\sqrt{h}}\left|\left(\eta_n-\eta_0\right)\right|\sup_{1\leq t\leq n}\left|\frac{x_{2t-1}'}{\sqrt{n}}\right|\frac{1}{n}\sum_{t=1}^{n}\left|X_{t-1}'D_0D_0'X_{t-1}\mathcal{K}_{t-1}^{(1)}\left(\tilde{\eta},\gamma_0\right)\right|=o_p\left(1\right),
\end{aligned}$$

where $\tilde{\eta}$ lies between η_0 and η_n, since $\left(\eta_n-\eta_0\right)\sqrt{n}/\sqrt{h}=o_p\left(1\right)$ and $\mathcal{K}^{(1)}$ is bounded. This completes the proof. ■

References

Anderson, H. M. (1997): "Transaction Costs and Non-linear Adjustment towards Equilibrium in the US Treasury Bill Market," *Oxford Bulletin of Economics and Statistics*, 59(4), 465–84.

Andrews, D. W. K. (1987): "Consistency in nonlinear econometric models: a generic uniform law of large numbers," *Econometrica*, 55(6), 1465–1471.

——— (1991): "Heteroskedasticity and autocorrelation consistent covariance matrix estimation," *Econometrica*, 59, 817–858.

Azuma, K. (1967): "Weighted sums of certain dependent random variables," *The Tohoku Mathematical Journal. Second Series*, 19, 357–367.

Bai, J., and P. Perron (1998): "Estimating and testing linear models with multiple structural changes," *Econometrica*, 66, 47–78.

Balke, N., and T. 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," *The Econometrics Journal*, 7(2), 628–651.

Chan, K. S. (1993): "Consistency and Limiting Distribution of the Least Squares Estimator of a Threshold Autoregressive Model," *The Annals of Statistics*, 21, 520–533.

Corradi, V., N. R. Swanson, and H. White (2000): "Testing for stationarity-ergodicity and for comovements between nonlinear discrete time Markov processes," *Journal of Econometrics*, 96(1), 39–73.

de Jong, R. M. (2001): "Nonlinear estimation using estimated cointegrating relations," *Journal of Econometrics*, 101(1), 109–122.

——— (2002): "Nonlinear minimization estimators in the presence of cointegrating relations," *Journal of Econometrics*, 110(2), 241–259.

Engle, R. F., and C. W. J. Granger (1987): "Co-integration and error correction: representation, estimation, and testing," *Econometrica*, 55(2), 251–276.

Escribano, A. (2004): "Nonlinear Error Correction: The Case of Money Demand in the United Kingdom (1878-2000)," *Macroeconomic Dynamics*, 8(1), 76–116.

Escribano, A., and S. Mira (2002): "Nonlinear error correction models," *Journal of Time Series Analysis*, 23(5), 509–522.

Gonzalo, J., and J.-Y. Pitarakis (2006): "Threshold Effects in Cointegrating Relationships," *Oxford Bulletin of Economics and Statistics*, 68(s1), 813–833.

Gonzalo, J., and M. Wolf (2005): "Subsampling Inference in Threshold Autoregressive Models," *Journal of Econometrics*, 127(201-224), 209–233.

Granger, C., and T. Terasvirta (1993): *Modelling nonlinear economic relationships.* Oxford University Press.

Granger, C. W. J. (2001): "Overview of Nonlinear Macroeconometric Empirical Models," *Macroeconomic Dynamics*, 5(4), 466–81.

Hansen, B. (1992): "Convergence to Stochastic Integrals for Dependent Heterogeneous Processes," *Econometric Theory*, 8, 489–500.

Hansen, B., and B. Seo (2002): "Testing for two-regime threshold cointegration in vector error correction models," *Journal of Econometrics*, 110, 293–318.

Hansen, B. E. (1999): "Threshold effects in non-dynamic panels: estimation, testing, and inference," *Journal of Econometrics*, 93(2), 345–368.

——— (2000): "Sample splitting and threshold estimation," *Econometrica*, 68, 575–603.

Horowitz, J. L. (1992): "A smoothed maximum score estimator for the binary response model," *Econometrica*, 60(3), 505–531.

Kapetanios, G., Y. Shin, and A. Snell (2006): "Testing for Cointegration in Nonlinear Smooth Transition Error Correction Models," *Econometric Theory*, 22.

Kristensen, D., and A. Rahbek (2008): "Likelihood-Based Inference in Nonlinear Error-Correction Models," *Journal of Econometrics*, forthcoming.

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

Lo, M., and E. Zivot (2001): "Threshold Cointegration and Nonlinear Adjustment to the Law of One Price," *Macroeconomic Dynamics*, 5(4), 533–576.

Michael, P., A. R. Nobay, and D. A. Peel (1997): "Transactions Costs and Nonlinear Adjustment in Real Exchange Rates: An Empirical Investigation," *Journal of Political Economy*, 105(4), 862–79.

Newey, W. K., and D. McFadden (1994): "Large sample estimation and hypothesis testing," in *Handbook of econometrics, Vol. IV*, vol. 2, pp. 2111–2245. North-Holland, Amsterdam.

Park, J., and P. Phillips (2001): "Nonlinear Regressions with Integrated Time Series," *Econometrica*, 69, 117–161.

Pötscher, B. M., and I. R. Prucha (1991): "Basic structure of the asymptotic theory in dynamic nonlinear econometric models. I. Consistency and approximation concepts," *Econometric Reviews*, 10(2), 125–216.

Psaradakis, Z., M. Sola, and F. Spagnolo (2004): "On Markov error-correction models, with an application to stock prices and dividends," *Journal of Applied Econometrics*, 19(1), 69–88.

Saikkonen, P. (1995): "Problems with the asymptotic theory of maximum likelihood estimation in integrated and cointegrated systems," *Econometric Theory*, 11(5), 888–911.

——— (2005): "Stability results for nonlinear error correction models," *Journal of Econometrics*, 127, 69–81.

——— (2007): "Stability of Regime Switching Error Correction Models under Linear Cointegration," *Econometric Theory*, 24, 294–318.

Seo, M. (2006): "Bootstrap Testing for the Null of No Cointegration in a Threshold Vector Error Correction Model," *Journal of Econometrics*, 134(1), 129–150.

Seo, M., and O. Linton (2007): "A Smoothed Least Squares Estimator For The Threshold Regression," *Journal of Econometrics*, 141, 704.

Sephton, P. S. (2003): "Spatial market arbitrage and threshold cointegration," *American Journal of Agricultural Economics*, 85, 1041.

Taylor, A. M. (2001): "Potential Pitfalls for the Purchasing Power Parity Puzzle? Sampling and Specification Biases in Mean-Reversion Tests of the Law of One Price," *Econometrica*, 69(2), 473–498.

Wooldridge, J. M., and H. White (1988): "Some invariance principles and central limit theorems for dependent heterogeneous processes," *Econometric Theory*, 4(2), 210–230.

Wu, C.-F. (1981): "Asymptotic theory of nonlinear least squares estimation," *The Annals of Statistics*, 9(3), 501–513.