Estimation of Nonlinear Error Correction Models

Myung Hwan Seo*
London School of Economics

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

Discussion paper
No. EM/2007/517
March 2007

* 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

© The author. All rights reserved. Short sections of text, not to exceed two paragraphs, may be quoted without explicit permission provided that full credit, including © notice, is given to the source.

1 Introduction

Nonlinear error correction models (ECM) have been studied actively in economics and there are numerous examples of applications, which include smooth transition ECM of Granger and Teräsvirta (1993), threshold cointegration of Balke and Fomby (1997), Markov switching ECM of Spagnolo, Sola, and Psaradakis (2004) and reviews by Granger (2001). 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, there is an important unresolved issue in this literature, which is the asymptotic properties of estimators of such models. First of all, consistency of least squares (LS) or maximum likelihood estimators (MLE) has not been proven except for special cases. Saikkonen (1995) argued that a general theory for consistency is difficult to establish due to the lack of uniformity in the convergence over the cointegrating vector space and demonstrated that the approach of Wu (1981) and Pötscher and Prucha (1991) can be useful by showing the consistency of an MLE of a cointegrated system that is nonlinear in parameters but otherwise linear. de Jong (2002) studied consistency of a minimization estimator of a smooth transition ECM where the error correction term appears in a bounded transition function only. They also studied asymptotic distributions of smooth models assuming that the cointegrating vector estimators converge at a certain rate. Second, these results are confined to smooth models while the threshold cointegration has acclaimed a large literature of applications as reviewed by Lo and Zivot (2001) and Bec and Rahbek (2004). 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).

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. We first establish the square root n consistency for the cointegrating vector estimates. 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 that is particularly relevant in practice.

This paper shows that the convergence rate of the LS estimator of the cointegrating vector in the threshold cointegration model is extremely fast, which is $n^{3/2}$. 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 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, the cointegrating vector 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. 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 the cointegrating vector is estimated due to the involved nonstationarity. Meanwhile, Seo and Linton (2006) 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 the cointegrating vector, the threshold parameter, and the other short-run parameters. The cointegrating vector estimate and threshold estimate 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 the cointegrating vector and threshold parameter were known. It is worth noting that the estimation of cointegrating vector affects the estimation of short-run parameters in smooth models. We also show that the cointegrating vector can be treated as known in the SLS estimation of the short-run parameters including the threshold parameter if we plug in the unsmoothed cointegrating vector estimate due to the fast convergence rate.

This paper is organized as follows. Section 2 introduces the regime switching VECMs and establish the square root n consistency of the least squares estimator of the cointegrating vector. Section 3 concentrates on the threshold cointegration model, obtaining the convergence rates of the LS estimators of the cointegrating vector and the asymptotic distributions of the SLS estimators of all the model parameters. It also discusses the estimation of the asymptotic variances. Section 4 concludes. Proofs of theorems are collected in the appendix.

We make the following conventions throughout the paper. The integral $\int$ is taken over $\mathcal{R}$ unless specified otherwise. For a function g, $\|g\|_2^2 = \int g(x)^2 \, dx$. The subscript 0 in any parameter and the hat indicate the true value and an estimate of the parameter, respectively, e.g., β_0 and $\hat{\beta}$. And, for a function $g(x_t, \beta)$, we let $g_t = g(x_t, \beta_0)$ and $\hat{g}_t = g\left(x_t, \hat{\beta}\right)$, for example, $z_t = x_t'\beta_0$ and $\hat{z}_t = x_t'\hat{\beta}$ for $z(x_t, \beta) = x_t'\beta$.

2 Regime Switching Error Correction Models

Let x_t be a p-dimensional $I(1)$ vector that is cointegrated with a cointegrating vector β, and $z_t(\beta) = x_t'\beta$. The first element of β is normalized to 1. The vector of the lagged first difference terms $\left(\Delta x_{t-1}', \cdots, \Delta x_{t-l+1}'\right)'$ is denoted as Δ_{t-1}. And let

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

which is a $(pl+2)$-dimensional vector. We 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. Typical examples of the transition function include the indicator function $1\{z_t(\beta) > \gamma\}$, the logistic function $(1 - \exp(-\gamma_1(z_t(\beta) - \gamma_2)))^{-1}$ and $1 - \exp\left(-\gamma^2 z_{t-1}(\beta)^2\right)$. While we may easily generalize the results in this paper to models with more regimes than 2, the distributional features are well exposed in the two-regime model.

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 Least Squares (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 $\theta = \left(\beta', \lambda', \gamma\right)'$. 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).

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. This approach is taken by Wu (1981) and Saikkonen (1995) among others.

The latter established the consistency of the maximum likelihood estimator of nonlinear transformation of the cointegrating vector in a linear model. Second, the consistency of the other short-run parameter estimates are 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* $Eu_t = 0, Eu_t u_t' = \Sigma$ *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* $\varepsilon > 0$, $E\left|X_t X_t^\top\right|^{\alpha_0+\varepsilon} < \infty$ *and* $E\left|X_{t-1}u_t\right|^{\alpha_0+\varepsilon} < \infty$. *Furthermore,* $E\Delta x_t = 0$. *and* $x_{[ns]}/\sqrt{n}$ *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* $p-1$ *s.t.* $\beta_0'\mathbf{\Omega} = 0$. *Let* x_{2t} *be the subvector of* x_t *excluding the first element of* x_t. *Then,* $x_{2[ns]}/\sqrt{n}$ *converges weakly to a vector Brownian motion* B *with a covariance matrix* Ω, *which is positive definite.*
(c) Θ *is compact and* $S_n^*(\theta)/n \xrightarrow{p} S^*(\theta)$ *uniformly in* $\theta \in \Theta$, *which is uniquely minimized at* θ_0, *when* β *is fixed at* β_0. *Moreover,* λ_z *is bounded away from zero, and infimums over* β *and* γ *of* $n^{-2}\sum_{t=1}^n x_{2t-1}x_{2t-1}d_{t-1}(\beta,\gamma)$ *and* $n^{-2}\sum_{t=1}^n x_{2t-1}x_{2t-1}(1-d_{t-1}(\beta,\gamma))$ *are bounded below by a random variable that is positive with probability one.*

Condition (a) is common as in Chan (1993). It simplifies our presentation but could be relaxed. It is referred to Bec and Rahbek (2004) and Saikkonen (2005) for the implication of the primitive conditions on $\{u_t\}$ to the stationarity and the mixing conditions of $\{\Delta x_t, z_t\}$ in the general nonlinear error correction models. They show the existence of such a process as (b). The first element of β is normalized to one and thus (b) introduced B, a subvector of $\mathbf{B}$ that excludes the first element of $\mathbf{B}$. Condition (c) is a set of identifying assumptions. When the cointegrating vector is known the model is a standard nonlinear model that satisfies standard consistency conditions. When $d_{t-1}(\beta,\gamma)$ is the indicator function, Seo and Linton (2006) showed that condition (c) is satisfied, thus establishing the consistency of short-run parameter estimates. The logistic function is continuous and satisfies a uniform law of large numbers. 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 $\lambda_z = 0$ has been studied by de Jong (2002) in the context of a smooth transition error correction model. The last condition in (c) implies that each regime has reasonable proportion of data thus identifying the parameters in each regime. The distributional limit of such quantities are well defined as in e.g. Seo (2005).

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

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 contrary, the threshold model where $d_{t-1}(\beta,\gamma)=1\{z_{t-1}(\beta)>\gamma\}$ has not been studied while the model has been adopted frequently for testing and empirical research. We turn to the so-called threshold cointegration model and develop an asymptotics for that in the next section.

3 Threshold Cointegration Model

Balke and Fomby (1997) introduced the threshold cointegration model to allow for nonlinear and/or asymmetric adjustment process to the equilibrium. The motivation of the model was that the magnitude and/or the sign of the disequilibrium 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 variables. 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.

To resolve this irregularity, Seo and Linton (2006) 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 $\mathcal{K}_t(\beta,\gamma)$ and define $X_\gamma(\beta)$ that stacks $X_{t-1}(\beta)\mathcal{K}_{t-1}(\beta,\gamma)$. Then, we have

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

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

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

Similarly as the concentrated LS estimator, we can define

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

It is worth mentioning that the true model is a threshold model and we employ the smoothing only for the estimation purpose. Also note that $\mathcal{K}\left(\frac{z_t(\beta)-\gamma}{h}\right)\to 1\{z_t(\beta)>\gamma\}$ as $h\to 0$.

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

Assumption 2 *(a) $\mathcal{K}$ is twice differentiable everywhere, $\mathcal{K}^{(1)}$ is symmetric around zero, $\left|\mathcal{K}^{(1)}\right|$ and $\left|\mathcal{K}^{(2)}\right|$ are bounded, and each of the following integrals is finite: $\int \left|\mathcal{K}^{(1)}\right|^4$, $\int \left|\mathcal{K}^{(2)}\right|^2$, and $\int \left|s^2\mathcal{K}^{(2)}(s)\right| ds$.*
(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

$$\int s^{i-1}\mathrm{sgn}(s)\,\mathcal{K}^{(1)}(s)\,ds = 0, \text{ and} \int s^{\vartheta}\mathrm{sgn}(s)\,\mathcal{K}^{(1)}(s)\,ds \neq 0,$$

and $\mathcal{K}(x) - \mathcal{K}(0) \gtrless 0$ if $x \gtrless 0$.
(c) For each integer $i\,(0 \leq i \leq \vartheta)$, and $\eta > 0$, and any sequence $\{h\}$ converging to 0,

$$\lim_{n\to\infty} h^{1-\vartheta}\int_{|hs|>\eta}\left|s^i\mathcal{K}^{(1)}(s)\right| ds = 0, \text{ and } \lim_{n\to\infty} h^{-1}\int_{|hs|>\eta}\left|\mathcal{K}^{(2)}(s)\right| ds = 0.$$

(d) $\limsup_{n\to\infty} nh^{2\vartheta} < \infty$ and

$$\lim_{n\to\infty} h^{-2\vartheta}\int_{|hs|>\eta}\left|\mathcal{K}^{(1)}(s)\right| ds = 0.$$

(e) For some $\mu \in (0,1]$, a positive constant C, and all $x, y \in \mathbb{R}$,

$$\left|\mathcal{K}^{(2)}(x) - \mathcal{K}^{(2)}(y)\right| \leq C\,|x-y|^{\mu}.$$

(f) $(\log n)/\left(nh^2\right) \to 0$ as $n \to \infty$.

These conditions are imposed in Seo and Linton (2006) 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 $\mathcal{K}(0) = 1/2$. Conditions (d) and (h) 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.

The consistency of the smoothed estimator $\hat{\theta}$ is a direct consequence of Theorem 1 under Assumption 2. If the cointegrating vector β were prespecified, then Seo and Linton's consistency applies. For the consistency of $\hat{\beta}$, it is sufficient to check if the condition (c) of Assumption 1 is satisfied, in particular, the last requirement in (c). It follows from the invariance principle in Assumption 1 (b) and the fact that $\mathcal{K}$ and the indicator function are bounded that

$$E\left|\mathcal{K}_{t-1}(\beta,\gamma) - 1\{z_{t-1}(\beta) > \gamma\}\right|^2 = h\int |\mathcal{K}(s) - 1\{s > 0\}|^2\,\phi_{t-1}(hs+\gamma)\,ds + o(1),$$

where ϕ_t is a normal density which is uniformly bounded in t. As $\int |\mathcal{K}(s) - 1\{s>0\}|^2 ds < \infty$ due to Assumption 2, we conclude that $n^{-2}\sum_{t=1}^{n}|x_t x_t' \mathcal{K}_{t-1}(\beta,\gamma) - 1\{z_{t-1}(\beta) > \gamma\}| = o_p(1)$ applying the Cauchy-Schwarz inequality.

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 the continuity of the model. 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) $E\left[X_{t-1}' D_0 D_0' X_{t-1} | z_{t-1} = \gamma_0\right] > 0.$

Then, we obtain the following rate result for the unsmoothed estimator of β and γ.

Theorem 2 *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. Intuitively, $\acute{\gamma} = \hat{\gamma} - x_{t-1}'\left(\hat{\beta} - \beta_0\right)$ behaves like a threshold estimate in a stationary threshold model as $1\left\{z_{t-1}(\beta) > \gamma\right\} = 1\left\{z_{t-1} > \acute{\gamma}\right\}$. Observing that $\sup_t x_{t-1} = O_p\left(n^{1/2}\right)$, we conclude 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 for 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 $\Delta_t = \Delta$. For each positive integer i, define

$$f^{(i)}(z|\Delta) = \partial^i f(z|\Delta) / \partial z^i$$

whenever the derivative exists. In the following, the i^{th} order differentiation is signified by the superscript $^{(i)}$, e.g. $\mathcal{K}^{(1)}(s) = (\partial/\partial s)\mathcal{K}(s)$. Also define

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

and

$$
\begin{aligned}
\sigma_v^2 &= E\left[\left\|\mathcal{K}^{(1)}\right\|_2^2\left(X_{t-1}'D_0u_t\right)^2+\left\|\tilde{\mathcal{K}}^{(1)}\right\|_2^2\left(X_{t-1}'D_0D_0'X_{t-1}\right)^2|z_{t-1}=\gamma_0\right]f\left(\gamma_0\right) \quad (5)\\
\sigma_q^2 &= \mathcal{K}^{(1)}\left(0\right)E\left(X_{t-1}'D_0D_0'X_{t-1}|z_{t-1}=\gamma_0\right)f\left(\gamma_0\right). \quad (6)
\end{aligned}
$$

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

Assumption 4 *(a) $E[|X_t'u_t|^r]<\infty$, $E[|X_t'X_t|^r]<\infty$, 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 $m\rightarrow\infty$.
(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 $M<\infty$, $f^{(i)}\left(z|\Delta\right)$ exists and is a continuous function of z satisfying $\left|f^{(i)}\left(z|\Delta\right)\right|<M$. In addition, $f\left(z|\Delta\right)<M$ for all z and almost every Δ.
(d) The conditional joint density $f\left(z_t,z_{t-m}|\Delta_t,\Delta_{t-m}\right)<M$, for all $\left(z_t,z_{t-m}\right)$ and almost all $\left(\Delta_t,\Delta_{t-m}\right)$.
(e) θ_0 is an interior point of Θ.

These assumptions are analogous to those imposed in Seo and Linton (2006) that study the SLS estimator of the threshold regression model. The condition (a) ensures the convergence of the variance covariance estimators. We need stronger mixing condition as set out in (b). The conditions (c) - (e) are common in the smoothed estimation as in Horowitz (1992), only (d) being an analogue of an *iid* sample to a dependent sample.

We present the asymptotic distribution below.

Theorem 3 *Suppose Assumption 1 - 2 hold. Let W denote a standard Brownian motion that is independent of B.Then,*

$$
\begin{aligned}
\begin{pmatrix} nh^{-1/2}\left(\hat{\beta}-\beta_0\right)\\ \sqrt{nh^{-1}}\left(\hat{\gamma}-\gamma_0\right)\end{pmatrix} &\Rightarrow \frac{\sigma_v}{\sigma_q^2}\begin{pmatrix}\int_0^1 BB' & \int_0^1 B\\ \int_0^1 B' & 1\end{pmatrix}^{-1}\begin{pmatrix}\int BdW\\ W\left(1\right)\end{pmatrix}\\
\sqrt{n}\left(\hat{\lambda}-\lambda_0\right) &\Rightarrow \mathcal{N}\left(0,\left[E\begin{pmatrix}1 & d_{t-1}\\ d_{t-1} & d_{t-1}\end{pmatrix}\otimes X_{t-1}X_{t-1}'\right]^{-1}\otimes\Sigma\right),
\end{aligned}
$$

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

The asymptotic distribution for $\hat{\beta}$ should be understood as that excluding its first element since it is normalized to one. We make some remarks on the similarities to and differences from the linear cointegration model and the stationary threshold model. First, both $\hat{\beta}$ and $\hat{\gamma}$, which appear inside the indicator function, are correlated in a similar form as that of the constant and cointegrating vector estimates in linear models, as γ is a constant within the indicator function. In fact, a reading of the proof of this theorem reveals that the linear part does not contribute to the asymptotic variance although β appears in the linear part of the model. Thus, the variance is different as it contains conditional expectations and the density at the discontinuity point. Note that σ_v^2/σ_q^4 is the variance of threshold estimate if the true cointegrating vector were known. 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 threshold estimates $\hat{\gamma}^*$ and $\hat{\gamma}$. 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 $\hat{\beta}$ and $\hat{\gamma}$ depend on the smoothing parameter h in a way the smaller h accelerate 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 $\hat{\beta}$ and the slope estimate $\hat{\lambda}$ 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 (see e.g. de Jong 2001, 2002).[1] This is due to the slower convergence of the cointegrating vector estimators. Therefore, it should also be noted that the Engle-Granger type two-step approach, where 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 a nonlinear error correction model. 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 4 *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 $\hat{\gamma}^*$)

[1] In case of $\mathbf{E}z_t = 0$, we can still retain the asymptotic Normality of the slope estimate by estimating (1) after replacing the $z_{t-1}(\beta)$ with

$$\bar{z}_{t-1}\left(\tilde{\beta}\right) = \bar{x}_{t-1}'\tilde{\beta} = \left(x_{t-1} - \frac{1}{n}\sum_s x_{s-1}\right)'\tilde{\beta},$$

for any n-consistent $\tilde{\beta}$, as in de Jong (2001). It is worth noting, however, that this demeaning increases the asymptotic variance.

as if they are β_0 and γ_0 due to Theorem 3. 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. The estimation of Ω can be done by applying a standard method for 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 derivative 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 of (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 3 and that of $\hat{\sigma}_v^2$ can be obtained after a slight modification of Theorem 4 of Seo and Linton (2006).

We can construct confidence interval for γ based on Corollary 4. 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 feature of the asymptotic variance matrix. Corollary 4 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 (2006) for a more discussion.

4 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 in such models with smooth transition. We also provided an asymptotic inference method for threshold cointegration, establishing the convergence rates and asymptotic distributions.

While we only considered two-regime models, it should be straightforward to extend our results to multiple-regime models. 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 without modification due to the super fast convergence rate of the cointegrating vector.

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

Proof of Theorems

Proof of Theorem 1

Let $\Theta_{r_n,\delta} = \{\theta \in \Theta : r_n |\beta - \beta_0| > \delta\}$. To show that $\hat{\beta}^* - \beta_0 = o_p(r_n^{-1})$ we need to show that for every $\delta > 0$,

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

Let $X^*_{\beta,\gamma} = \left[X(\beta) : X^*_\gamma(\beta)\right]$ and rewrite (2) as

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

Let $\eta = \sqrt{n}(\beta - \beta_0)$ and r_n be a sequence such that $r_n \to \infty$ and $r_n/\sqrt{n} \to 0$ as $n \to \infty$. Then, $\eta \to \infty$ for any $\beta \in \Theta_{r_n,\delta}$. We note that $S_n^*(\theta_0)/n$ and $y'y/n$ are $O_p(1)$ by the law of large numbers and show that $y' X^*_{\beta,\gamma}\lambda/n = O_p(|\eta|)$ while $\lambda' X^{*\prime}_{\beta,\gamma} X^*_{\beta,\gamma}\lambda/n = O_p\left(|\eta|^2\right)$. In particular, we show that $\inf_{\theta\in\Theta_{r_n,\delta}} \lambda' X^{*\prime}_{\beta,\gamma} X^*_{\beta,\gamma}\lambda/\left(n|\eta|^2\right)$ converges in distribution to a random variable which is positive with probability one, thus proving (8).

Note that $y' X^*_{\beta,\gamma}/n$ consists of sample means of the product of Δx_t and $\left(1, z_{t-1}(\beta), \Delta'_{t-1}\right)$ and of Δx_t and $\left(1, z_{t-1}(\beta), \Delta'_{t-1}\right) d_{t-1}(\beta,\gamma)$. However, as d is 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

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

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

We can proceed similarly with the matrix $\lambda' X^{*\prime}_{\beta,\gamma} X^*_{\beta,\gamma}\lambda/n$. We can easily see that the leading terms in the matrix are $\frac{1}{n}\sum_t z_{t-1}(\beta)^2$ and $\frac{1}{n}\sum_t d_{t-1}(\beta,\gamma)^2 z_{t-1}(\beta)^2$. Note that

$$\frac{1}{n|\eta|^2}\sum_t d_{t-1}(\beta,\gamma)^2 z_{t-1}(\beta)^2 = \frac{1}{n}\sum_t d_{t-1}(\beta,\gamma)^2 \left(\frac{x'_{t-1}\upsilon}{\sqrt{n}}\right)^2 + o_p(1) = O_p(1),$$

where $|\upsilon| = 1$. Thus,

$$\frac{1}{n|\eta|^2}\lambda' X^{*\prime}_{\beta,\gamma} X^*_{\beta,\gamma}\lambda = \lambda'_z\left(\Xi^*_1/|\eta|^2 \otimes I_p\right)\lambda_z + o_p(1),$$

uniformly in $\Theta_{r_n,\delta}$, where

$$\Xi^*_1 = \begin{pmatrix} \frac{1}{n}\sum_t z_{t-1}(\beta)^2 & \frac{1}{n}\sum_t d_{t-1}(\beta,\gamma)^2 z_{t-1}(\beta)^2 \\ \frac{1}{n}\sum_t d_{t-1}(\beta,\gamma)^2 z_{t-1}(\beta)^2 & \frac{1}{n}\sum_t d_{t-1}(\beta,\gamma)^2 z_{t-1}(\beta)^2 \end{pmatrix}.$$

Note that the weak limit of $\Xi_1^*/|\eta|^2$ on $\Theta_{r_n,\delta}$ is a positive definite matrix with probability one since $\frac{1}{n|\eta|^2}\sum_t z_{t-1}(\beta)^2 \Longrightarrow \int (B'v)^2$ and

$$\lim_{n\to\infty}\frac{1}{n|\eta|^2}\sum_t z_{t-1}(\beta)^2 > \lim_{n\to\infty}\frac{1}{n|\eta|^2}\sum_t d_{t-1}(\beta,\gamma)^2 z_{t-1}(\beta)^2 > 0$$

almost surely by Assumption 1. Since $|\lambda_z|$ is bounded away from zero, the infimum of $\lambda_z'\left(\Xi_1^*/|\eta|^2\otimes I_p\right)\lambda_z$ on $\Theta_{r_n,\delta}$ converges in distribution to a random variable that is positive with probability one by the continuous mapping theorem.

Since $\hat{\beta}^*-\beta_0 = o_p\left(r_n^{-1}\right)$ for any $r_n/\sqrt{n}\to 0$, we should have $\hat{\beta}^*-\beta_0 = O_p\left(n^{-1/2}\right)$. Otherwise, there exist $\varepsilon>0$ and an increasing sequence $b_n\to\infty$ such that $\Pr\{\sqrt{n}\left|\hat{\beta}^*-\beta_0\right| > b_n$ *infinitely often*$\} > \varepsilon$ and $b_n/\sqrt{n}\to 0$. Now, let $c_n=\frac{\sqrt{n}}{b_n}$, then $c_n\to\infty$ and $c_n/\sqrt{n}\to 0$ and thus $c_n\left(\hat{\beta}^*-\beta_0\right) = o_p(1)$, which yields contradiction since

$$\Pr\{c_n\left|\hat{\beta}^*-\beta_0\right| = \sqrt{n}\left|\hat{\beta}^*-\beta_0\right|/b_n > 1 \textit{ infinitely often}\} > 0.$$

Now, let $\bar{\Theta}_\delta = \{\theta : \delta < \sqrt{n}(\beta-\beta_0) < C$ for a finite constant $C\}$. Let $\acute{\gamma} = \gamma - x_{t-1}'(\beta-\beta_0)$ and assume $\acute{\gamma}$ lies in a compact set without loss of generality as $\sup_t x_{t-1} = O_p(\sqrt{n})$ and $\hat{\beta}^*-\beta_0 = O_p\left(n^{-1/2}\right)$ (see e.g. de Jong 2002 p.256). Then, $y'X_{\beta,\gamma}^*/n = y'X_{\beta_0,\acute{\gamma}}^*/n + o_p(1)$, uniformly in $\bar{\Theta}_\delta$. Furthermore,

$$\lambda' X_{\beta,\gamma}^* X_{\beta,\gamma}^*\lambda = \lambda' X_{\beta_0,\acute{\gamma}}^{*\prime} X_{\beta_0,\acute{\gamma}}^*\lambda + \lambda_z'\left(\Xi_2^*\otimes I_p\right)\lambda_z + o_p(1),$$

uniformly in $\bar{\Theta}_\delta$, where Ξ_2 is the matrix obtained by replacing $z_{t-1}(\beta)$ in Ξ_1 with $x_{t-1}(\beta-\beta_0)$. Thus,

$$S_n^*(\theta)/n = u^*(\beta_0,\acute{\gamma},\lambda)' u^*(\beta_0,\acute{\gamma},\lambda)/n + \lambda_z'\left(\Xi_2^*/n\otimes I_p\right)\lambda_z + o_p(1),$$

uniformly in $\bar{\Theta}_\delta$, where $u^*(\beta_0,\gamma,\lambda) = y - X_{\beta_0,\gamma}^*\lambda$. Note that $u^*(\beta_0,\gamma,\lambda)' u^*(\beta_0,\gamma,\lambda)/n$ converges uniformly in probability to a limit that is greater than or equal to $p\lim_{n\to\infty} S_n^*(\theta_0)/n$ for all (γ,λ) by Assumption 1. Furthermore, as for Ξ_1^*, Ξ_2^*/n converges to a positive definite matrix as $C > \sqrt{n}(\beta-\beta_0) > \delta > 0$. Thus, $\inf_{\theta\in\bar{\Theta}_\delta}\lambda_z'\left(\Xi_2^*\otimes I_p\right)\lambda_z$ converges in distribution a positive random variable. Therefore, (8) is proven for $r=1/2$.

Next, we turn to the consistency of the short-run parameters. Since $\sqrt{n}\left(\hat{\beta}^*-\beta_0\right) = o_p(1)$, we have $\Xi_2^*/n = o_p(1)$ and

$$S_n^*\left(\acute{\theta}\right)/n = u^*(\beta_0,\acute{\gamma},\lambda)' u^*(\beta_0,\acute{\gamma},\lambda)/n + o_p(1).$$

Then, the standard consistency proof of using the uniform convergence of the first term on the right hand side applies to the estimates of $\acute{\gamma}$ and λ (see e.g. Theorem 1 in Seo and Linton 2006). As $\acute{\gamma} = \gamma + o_p(1)$, the consistency of $\hat{\gamma}^*$ also follows. Again the same argument applies for the smoothed estimator of the short-run parameters as above. Therefore, the proof is complete. ■

Proof of Theorem 2

As in the proof of Theorem 1, let $\eta = \sqrt{n}\left(\beta - \beta_0\right)$. Since $\hat{\eta}^* = o_p(1)$, it follows that $\sup_t \frac{x_t'}{\sqrt{n}} \hat{\eta}^* = o_p(1)$ due to the invariance principle for $x_t/\sqrt{n}$. Then, $\frac{x_t'}{\sqrt{n}}\eta + \gamma$ can be confined to a compact set without loss of generality. Thus, let $\zeta = \frac{x_t'}{\sqrt{n}}\eta + \gamma$ and consider a new parametrization of the model:

$$\dot{u}_t\left(\dot{\theta}\right) = \Delta x_t - A'X_{t-1} - D'X_{t-1}1\left\{z_{t-1} > \zeta\right\} - \left(A_z + D_z 1\left\{z_{t-1} > \zeta\right\}\right)\frac{x_{t-1}'}{\sqrt{n}}\eta,$$

where $\dot{\theta} = \left(\eta', \lambda', \zeta\right)'$. Note that for any θ there is $\dot{\theta}$ such that $u_t(\theta) = \dot{u}_t\left(\dot{\theta}\right)$ for each realization of data and vice versa. Also note that $\theta_0 = \dot{\theta}_0$ and the least squares estimates of θ and $\dot{\theta}$ must be the same. Here, we show that $\hat{\zeta} = \gamma_0 + O_p\left(n^{-1}\right)$, which implies that $\hat{\zeta} = \frac{x_t'}{\sqrt{n}}\hat{\eta}^* + \hat{\gamma}^* = O_p\left(n^{-1}\right)$ and thus $\sqrt{n}\left(\hat{\beta}^* - \beta_0\right)$ and $\hat{\gamma}^*$ are $O_p\left(n^{-1}\right)$. Partition $\dot{\theta} = \left(\dot{\theta}_1', \zeta\right)'$.

Let $\Theta_c = \left\{\dot{\theta} : \left|\dot{\theta} - \theta_0\right| < c\right\}$. 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 later. It is sufficient to show the following claim for $\hat{\zeta}$ to be n-consistent.

Claim *For any $\varepsilon > 0$, there is $c > 0$ and $K > 0$ such that with probability greater than $1 - \varepsilon$, if $\dot{\theta} \in \Theta_c$,* and $|\zeta| > K/n$, *then* $\left(\dot{S}_n\left(\dot{\theta}\right) - \dot{S}_n\left(\dot{\theta}_1, \zeta_0\right)\right) > 0$, *where* $\dot{S}_n\left(\dot{\theta}\right) = \sum_t \dot{u}_t\left(\dot{\theta}\right)'\dot{u}_t\left(\dot{\theta}\right)$.

Proof of Claim First assume $\zeta > 0$ and $\zeta_0 = 0$ for simplicity. Let $u_t = u_t(\theta_0) = \dot{u}_t\left(\dot{\theta}_0\right)$ and $\dot{u}_{1t}\left(\dot{\theta}\right) = \Delta x_t - A'X_{t-1} - D'X_{t-1}1\left\{z_{t-1} > \zeta\right\}$ and $\dot{u}_{2t}\left(\dot{\theta}\right) = \dot{u}_t(\theta) - \dot{u}_{1t}\left(\dot{\theta}\right)$. Then,

$$\begin{aligned}
\left(\dot{S}_n\left(\dot{\theta}\right) - \dot{S}_n\left(\dot{\theta}_1, 0\right)\right)/n &= \frac{1}{n}\sum_t\left[\dot{u}_{1t}\left(\dot{\theta}\right)'\dot{u}_{1t}\left(\dot{\theta}\right) - \dot{u}_{1t}\left(\dot{\theta}_1, 0\right)'\dot{u}_{1t}\left(\dot{\theta}_1, 0\right)\right] \\
&\quad + \frac{1}{n}\sum_t\left[\dot{u}_{2t}\left(\dot{\theta}\right)'\dot{u}_{2t}\left(\dot{\theta}\right) - \dot{u}_{2t}\left(\dot{\theta}_1, 0\right)'\dot{u}_{2t}\left(\dot{\theta}_1, 0\right)\right] \\
&\quad - \frac{2}{n}\sum_t\left[\dot{u}_{1t}\left(\dot{\theta}\right)'\dot{u}_{2t}\left(\dot{\theta}\right) - \dot{u}_{1t}\left(\dot{\theta}_1, 0\right)'\dot{u}_{2t}\left(\dot{\theta}_1, 0\right)\right],
\end{aligned}$$

which are denoted as $D_{1n}\left(\dot{\theta}\right), D_{2n}\left(\dot{\theta}\right)$, and $D_{3n}\left(\dot{\theta}\right)$, respectively. After some tedious algebra, we can show that

$$\begin{aligned}
\left|D_{2n}\left(\dot{\theta}\right)\right| &= \left|\frac{1}{n}\sum_t\left(\frac{x_{t-1}'}{\sqrt{n}}\eta\right)^2\left(D_z'D_z - 2A_z'D_z\right)1\left\{0 < z_{t-1} \le \zeta\right\}\right| \\
&\le \frac{1}{n}\sum_t 1\left\{0 < z_{t-1} \le \zeta\right\}\sup_t\left(\frac{x_{t-1}'}{\sqrt{n}}\eta\right)^2\max_{\dot{\theta}\in\Theta_c}\left|D_z'D_z - 2A_z'D_z\right|,
\end{aligned}$$

which is $O_p(\zeta c)$ because $\frac{1}{n}\sum_t 1\left\{0 < z_{t-1} \le \zeta\right\} = O_p(\zeta)$ due to Chan (1993, Claim 2), $\sup_t\left(\frac{x_{t-1}'}{\sqrt{n}}\right) = O_p(1), |\eta| < c$, and the parameter space bounded. Similarly, we can show that $D_{3n}\left(\dot{\theta}\right) = O_p\left(\zeta c\right)$ as

$$D_{3n}\left(\dot{\theta}\right) = \frac{2}{n}\sum_t \left(X_{t-1}'D\left(A_z + D_z\right) - \left(\Delta x_t - A'X_{t-1}\right)' D_z\right)\left(\frac{x_{t-1}'}{\sqrt{n}}\eta\right) 1\left\{0 < z_{t-1} \leq \zeta\right\}.$$

Thus, for the given $\varepsilon > 0$, we can choose a constant M such that $\Pr\left\{\left|D_{2n}\left(\dot{\theta}\right)\right| > Mc\right\} < \varepsilon$ and $\Pr\left\{\left|D_{3n}\left(\dot{\theta}\right)\right| > Mc\right\} < \varepsilon$. On the other hand, Chan (1993, Claim 1) also showed that $D_{1n}\left(\dot{\theta}\right) > \bar{c}\zeta$ for some $\bar{c} > 0$ for all large n and for all $K/n < \zeta < c$ by choosing c small. Then, letting $c < \bar{c}/M$, we have $\dot{S}_n\left(\dot{\theta}\right) - \dot{S}_n\left(\dot{\theta}_1, 0\right) > 0$ for all large n with probability greater than $1-\varepsilon$ as desired. The case where $\zeta < 0$ can be done similarly. ■

Proof of Theorem 3

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

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

where $D_n = diag\left(\left(h/n\right)^{1/2} I_{p-1}, \sqrt{h}, I_{2(pl+2)}\right)$ and $\tilde{\theta}$ lies between $\hat{\theta}$ and θ_0.

Convergence of T_n

Let

$$\begin{aligned} e_t\left(\theta\right) &= \Delta x_t - A'X_{t-1} - D'X_{t-1}\mathcal{K}_{t-1}\left(\beta,\gamma\right) - \left(A_z + D_z\mathcal{K}_{t-1}\left(\beta,\gamma\right)\right)x_{t-1}'\left(\beta - \beta_0\right) \\ &= u_t - \left(A - A_0\right)' X_{t-1} - \left(D - D_0\right)' X_{t-1}d_{t-1} \\ &\quad - D'X_{t-1}\left(\mathcal{K}_{t-1}\left(\beta,\gamma\right) - d_{t-1}\right) - \left(A_z + D_z\mathcal{K}_{t-1}\left(\beta,\gamma\right)\right)x_{t-1}'\left(\beta - \beta_0\right), \end{aligned} \tag{9}$$

recalling that $X_{t-1}'\left(\beta\right)D = X_{t-1}'D + x_{t-1}'\left(\beta - \beta_0\right)D_z'$, and note that

$$\begin{aligned} \frac{\partial e_t\left(\theta\right)'}{\partial\beta} &= -x_{2t-1}\left[A_z' + D_z'\mathcal{K}_{t-1}\left(\beta,\gamma\right) + \left(D_z'x_{t-1}'\left(\beta - \beta_0\right) + X_{t-1}'D\right)\mathcal{K}_{t-1}^{(1)}\left(\beta,\gamma\right)/h\right] \\ \frac{\partial e_t\left(\theta\right)'}{\partial\gamma} &= -\left(X_{t-1}'D + D_z'x_{t-1}'\left(\beta - \beta_0\right)\right)\mathcal{K}_{t-1}^{(1)}\left(\beta,\gamma\right)/h \\ \frac{\partial e_t\left(\theta\right)'}{\partial\lambda} &= \begin{pmatrix} -\left(X_{t-1}'D + D_z'x_{t-1}'\left(\beta - \beta_0\right)\right)\otimes I_p \\ -\mathcal{K}_{t-1}\left(\beta,\gamma\right)\left(X_{t-1}'D + D_z'x_{t-1}'\left(\beta - \beta_0\right)\right)\otimes I_p \end{pmatrix}. \end{aligned} \tag{10}$$

Note that x_{2t-1} appears in (10) due to the normalization restriction. The asymptotic distribution of $\sqrt{n}D_nT_n\left(\theta_0\right)/2 = \frac{1}{\sqrt{n}}D_n\sum_t \frac{\partial e_t(\theta_0)'}{\partial\theta}e_t\left(\theta_0\right)$ has been developed in Seo and Linton (2006) except for the first part of β. Thus, we focus on

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

where

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

Since v_{1t} is a martingale difference sequence, $\frac{1}{n}\sum_t x_{2t-1} v_{1t}\sqrt{h} = O_p\left(\sqrt{h}\right)$ due to the convergence of stochastic integrals (see e.g. Kurtz and Protter 1991). We derive the convergence of $\frac{1}{n}\sum_t x_{2t-1} v_{3t}/\sqrt{h}$. Then, similar argument yields that $\frac{1}{n}\sum_t x_{2t-1} v_{3t}\sqrt{h} = o_p(1)$ as $h \to 0$. Let $\bar{v}_{3t} = \left(v_{3t} - Ev_{3t}\right)/\sqrt{h}$, then $\bar{v}_{3t}$ is a zero mean strong mixing array Seo and Linton (2006, Lemma 2) has shown that $\sqrt{n/h}Ev_{3t} \to 0$ and $\operatorname{var}\left[(hn)^{-1/2}\sum_t v_{3t}\right] = \operatorname{var}\left[v_{3t}/\sqrt{h}\right] + o_p(1) \to \sigma_v^2$, 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 $n^{-1/2}\sum_{t=2}^{[nr]}\left(\Delta x_{2t-1}', \bar{v}_{3t}/\sigma_v\right) \Longrightarrow (B', W)$ due to Assumption 4 and the invariance principle of Wooldridge and White (1988, Theorem 2.11). For the independence between B and W, see Lemma 2 of Seo and Linton (2006) to get $\sum_{s,t=1}^{n} E\Delta x_s \bar{v}_{3t} = o(1)$. Here, similar argument leads to the asymptotic independence between the score for β and λ. For the relation between the asymptotic distributions of $\hat{\gamma}$ and $\hat{\beta}$, note that $v_{3t} = h\frac{\partial e_t(\theta_0)'}{\partial \gamma} e_t(\theta_0)$.

For the convergence of $\frac{1}{n}\sum_t x_{2t-1}\bar{v}_{3t}$, we resort to Hansen (1992, Theorem 3.1). While Hansen imposes moment condition higher than 2, we show here that the second moment condition is sufficient by studying Hansen's proof directly. We find that the moment condition higher-than-2 is used to show that $\sup_{t \le n} n^{-1/2}\sum_{k=1}^{\infty} E\left[\bar{v}_{3t+k}|\mathcal{F}_t\right] = o_p(1)$, where $\mathcal{F}_t$ is the natural filtration at time t. Using the Markov inequality and mixing inequality, he obtains

$$
\Pr\left\{\left|\sup_{t\le n} n^{-1/2}\sum_{k=1}^{\infty} E\left[\bar{v}_{3t+k}|\mathcal{F}_t\right]\right| \ge \varepsilon\right\} \le \frac{CE\left|\bar{v}_{3t}\right|^p}{\varepsilon^p n^{p/2-1}}, \tag{11}
$$

which converges to zero provided $p > 2$ and $E\left|\bar{v}_{3t}\right|^p < \infty$. Now we show that while $E\left|\bar{v}_{3t}\right|^p$ is not bounded for $p > 2$ but diverges slower than $n^{p/2-1}$. As $\sqrt{n/h}Ev_{3t} \to 0$ and the part with u_t can be done in the same manner, we focus on

$$
\begin{aligned}
& E\left|\mathcal{K}_{t-1}^{(1)} X_{t-1}' D_0 D_0' X_{t-1}\left(\mathcal{K}_{t-1} - d_{t-1}\right)\right|^p h^{-p/2} \\
= \;& h^{-p/2}\int \left|X' D_0 D_0' X\right|^p \left|\mathcal{K}^{(1)}\left(\frac{z-\gamma_0}{h}\right)\left(\mathcal{K}\left(\frac{z-\gamma_0}{h}\right) - 1\left(z > \gamma_0\right)\right)\right|^p f(z|X)\, dz dP_X \\
= \;& h^{1-p/2}\int \left|X' D_0 D_0' X\right|^p \left|\mathcal{K}^{(1)}(s)\left(\mathcal{K}(s) - 1(s)\right)\right|^p f\left(hs + \gamma_0|X\right) ds dP_X,
\end{aligned}
$$

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 $E\left|X' D_0 D_0' X\right|^p < \infty$. As $h^{1-p/2} n^{p/2-1} \to 0$, we conclude that (11) converges to zero. Therefore,

$$
\frac{\sqrt{h}}{2\sigma_v n}\sum_t \frac{\partial e_t(\theta_0)'}{\partial \beta} e_t(\theta_0) \Rightarrow \int_0^1 B dW. \tag{12}
$$

Convergence of Q_n

First, we show that $h^{-1}(\hat{\gamma}-\gamma_0)$ and $h^{-1}\hat{\eta}$ are $o_p(1)$. Define $\zeta, \dot{\theta}$ and Θ_{r_n} for some sequence $r_n \to 0$ as in the proof of Theorem 2. With slight abuse of notation, let $\mathcal{K}_{t-1}^{(1)}(\zeta) = \mathcal{K}^{(1)}\left(\frac{z_{t-1}-\zeta}{h}\right)$. Since $\sup_t \frac{x'_{t-1}}{\sqrt{n}}\hat{\eta} = o_p(1)$ and $\mathcal{K}_{t-1}^{(1)}(\beta,\gamma)$ is bounded, we may write

$$\begin{aligned}
\frac{-1}{2n}\sum_t \frac{\partial e_t\left(\dot{\theta}\right)'}{\partial \gamma} e_t\left(\dot{\theta}\right) &= \frac{1}{n}\sum_t \left(X'_{t-1}D + D'_z x'_{t-1}(\beta-\beta_0)\right)\mathcal{K}_{t-1}^{(1)}(\beta,\gamma)\, e_t\left(\dot{\theta}\right) \\
&= \frac{1}{n}\sum_t X'_{t-1}D\mathcal{K}_{t-1}^{(1)}(\zeta)\, e_t\left(\beta_0,\dot{\theta}_2\right) + o_p(1),
\end{aligned}$$

uniformly in $\dot{\theta} \in \Theta_{r_n}$, where $\dot{\theta} = \left(\beta, \dot{\theta}_2\right)$. However, Lemma 5 of Seo and Linton shows that $h^{-1}\left(\hat{\zeta}-\gamma_0\right) = o_p(1)$ using the convergence of $\frac{1}{n}\sum_t X'_{t-1}D\mathcal{K}_{t-1}^{(1)}(\zeta)\, e_t\left(\beta_0,\dot{\theta}_2\right)$, which corresponds to $T_n^{\psi}(\theta)$ there, and that of its expectation. Thus, $h^{-1}(\hat{\gamma}-\gamma_0)$ and $h^{-1}\hat{\eta}$ are $o_p(1)$. Now let $\Theta_{0n} = \{\theta \in \Theta : |\theta-\theta_0| < r_n, h^{-1}|\gamma-\gamma_0| < r_n, h^{-1}|\eta| < r_n\}$.

Next, we may write

$$\begin{aligned}
Q_n(\theta)/2 &= Q_n^a(\theta) + Q_n^b(\theta) \\
&= \frac{1}{n}\sum_t \frac{\partial e_t(\theta)'}{\partial \theta}\frac{\partial e_t(\theta)}{\partial \theta'} + \sum_{i=1}^{p}\frac{1}{n}\sum_t \frac{\partial^2 e_{it}(\theta)}{\partial\theta\partial\theta'} e_{it}(\theta),
\end{aligned}$$

where $e_{it}(\theta)$ is the i^{th} element of $e_t(\theta)$. Start with $Q_n^b(\theta)$, in particular,

$$\begin{aligned}
& -\sum_t \frac{\partial^2 e_{it}(\theta)}{\partial\beta\partial\beta'} e_{it}(\theta) \\
&= \sum_t x_{2t-1}x'_{2t-1}\left(2D_{zi}\frac{\mathcal{K}_{t-1}^{(1)}(\beta,\gamma)}{h} + \left(X'_{t-1}D_i + D_{z_i}\frac{x'_{t-1}\eta}{\sqrt{n}}\right)\frac{\mathcal{K}_{t-1}^{(2)}(\beta,\gamma)}{h^2}\right) e_{it}(\theta) \qquad (13)
\end{aligned}$$

where M_i indicates the i^{th} Column of a matrix M and the i^{th} element if it is a vector. We show that

$$\begin{aligned}
& \frac{h}{n^2}\sum_t x_{2t-1}x'_{2t-1}X'_{t-1}D_i\frac{\mathcal{K}_{t-1}^{(2)}(\beta,\gamma)}{h^2} e_{it}(\theta) \qquad (14) \\
& \xrightarrow{p} E\left(D'_{0i}X'_{t-1}X_{t-1}D_{0i}|z_{t-1}=\gamma_0\right) f(\gamma_0)\int \tilde{\mathcal{K}}^{(2)}\int_0^1 BB',
\end{aligned}$$

uniformly in Θ_{0n}, where $\int \tilde{\mathcal{K}}^{(2)} = \int \mathcal{K}^{(2)}(s)\left(\mathcal{K}(s) - 1\{s>0\}\right)$, and then we can see that the other terms in (13) are $o_p(1)$ as $h \to 0$ and $\sup_t \frac{x'_{t-1}\eta}{h\sqrt{n}} = o_p(1)$ uniformly in Θ_{0n}. Recall (9) and first observe that

$$\frac{1}{hn^2}\sum_t x_{2t-1}x'_{2t-1}X'_{t-1}D_i\mathcal{K}_{t-1}^{(2)}(\beta,\gamma)\, u_{it} = O_p\left((nh)^{-1/2}\right),$$

uniformly in Θ_{0n} following the same reasoning for (12). While we have $\mathcal{K}_{t-1}^{(2)}(\beta,\gamma)$ with $(\beta,\gamma) \neq (\beta_0,\gamma_0)$, we may write $\mathcal{K}_{t-1}^{(2)}(\beta,\gamma) = \mathcal{K}^{(2)}\left(\frac{z_{t-1}-\gamma_0}{h} + g\right)$ and restrict g into a set

$\{|g| < r_n\}$ for a sequence $r_n \to 0$, as $\sup_t x'_{t-1} |\beta - \beta_0| / h + |\gamma - \gamma_0| / h = o_p(1)$. Thus, it does not affect the derivation. Similarly, we can show that all the other terms are $o_p(1)$ except the following:

$$\frac{1}{hn^2} \sum_t x_{2t-1} x'_{2t-1} X'_{t-1} D_i \mathcal{K}^{(2)}_{t-1}(\beta, \gamma) \left(X'_{t-1} D_i \left(\mathcal{K}_{t-1}(\beta, \gamma) - d_{t-1} \right) \right),$$

which converges in probability to the limit in (14) uniformly in Θ_{0n}. This follows from Hansen (1992, Theorem 3.3), the invariance principle, and the fact that

$$\begin{aligned} & E X'_{t-1} D_i \mathcal{K}^{(2)}_{t-1}(\beta, \gamma) \left(X'_{t-1} D_i \left(\mathcal{K}_{t-1}(\beta, \gamma) - d_{t-1} \right) \right) / h \\ \to \quad & E \left(D'_{0i} X'_{t-1} X_{t-1} D_{0i} | z_{t-1} = \gamma_0 \right) f(\gamma_0) \int \tilde{\mathcal{K}}^{(2)}, \end{aligned}$$

uniformly in Θ_{0n} as in (12) using the change-of-variables.

Next, note that $\frac{\partial^2 e_{it}(\theta)}{\partial \lambda \partial \lambda'} = 0$ and

$$\begin{aligned} & \begin{pmatrix} \frac{\partial^2 e_{it}(\theta)}{\partial \gamma \partial \beta'} & \frac{\partial^2 e_{it}(\theta)}{\partial \gamma \partial \gamma'} \\ \frac{\partial^2 e_{it}(\theta)}{\partial \lambda \partial \beta'} & \frac{\partial^2 e_{it}(\theta)}{\partial \lambda \partial \gamma'} \end{pmatrix} \\ = & \begin{pmatrix} \left(D_{zi} \mathcal{K}^{(1)}_{t-1}(\beta, \gamma) / h + X_{t-1}(\beta)' D_i \mathcal{K}^{(2)}_{t-1}(\beta, \gamma) / h^2 \right) x'_{2t-1} & -X_{t-1}(\beta)' D_i \mathcal{K}^{(2)}_{t-1}(\beta, \gamma) / h^2 \\ -\begin{pmatrix} \epsilon_2 \\ \mathcal{K}_{t-1}(\beta, \gamma) \epsilon_2 + \frac{\mathcal{K}^{(1)}_{t-1}(\beta, \gamma)}{h} X_{t-1}(\beta) \end{pmatrix} x'_{2t-1} \otimes I_i & \begin{pmatrix} 0 \\ \frac{\mathcal{K}^{(1)}_{t-1}(\beta, \gamma)}{h} X_{t-1}(\beta) \end{pmatrix} \otimes I_i \end{pmatrix}, \end{aligned}$$

where $\epsilon_2 = (0, 1, 0, ..., 0)$ whose dimension is $(pl + 2)$. The same reasoning as above applies to $\frac{\partial^2 e_{it}(\theta)}{\partial \gamma \partial \beta'}$ and $\frac{\partial^2 e_{it}(\theta)}{\partial \lambda \partial \beta'}$ and Seo and Linton (2006, Corollary 3) to the remaining terms, and thus we conclude

$$D_n Q_n^b D_n \Rightarrow \tilde{\sigma}_q^2 \int \tilde{\mathcal{K}}^{(2)}_1 \begin{pmatrix} -\int_0^1 BB' & \int_0^1 B & 0 \\ \int_0^1 B' & 1 & 0 \\ 0 & 0 & 0 \end{pmatrix},$$

where $\tilde{\sigma}_q^2 = \sum_{i=1}^p E \left(D'_{0i} X'_{t-1} X_{t-1} D_{0i} | z_{t-1} = \gamma_0 \right) f(\gamma_0)$. 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 $\left\| \mathcal{K}^{(1)} \right\|_2^2 - \int \tilde{\mathcal{K}}^{(2)}_1 = \mathcal{K}^{(1)}(0)$ by an application of the integral by parts, which yield the desired result. ■

Proof of Corollary 4

We examine the score function T_n and the Hessian Q_n as functions of β. But, we have already derived the convergence of the Hessian Q_n for $\hat{\beta} = \beta_0 + O_p\left(hn^{-1/2}\right)$. Thus, we only have to examine the score T_n. Let $\beta_n = \beta + O_p\left(n^{-3/2}\right)$, $\ddot{\theta} = (\gamma, \lambda)$ and

$$
\begin{aligned}
\ddot{e}_t\left(\ddot{\theta};\beta_n\right) &= u_t - (A - A_0)' X_{t-1} - (D - D_0)' X_{t-1} d_{t-1} \\
&\quad - D' X_{t-1} \left(\mathcal{K}_{t-1}(\beta_n, \gamma) - d_{t-1}\right) - \left(A_z + D_z \mathcal{K}_{t-1}(\beta_n, \gamma)\right) x'_{t-1} (\beta_n - \beta_0), \\
\frac{\partial \ddot{e}_t\left(\ddot{\theta}\right)'}{\partial \ddot{\theta}} &= \begin{pmatrix} -\left(X'_{t-1} D + D'_z x'_{t-1}(\beta_n - \beta_0)\right) \mathcal{K}^{(1)}_{t-1}(\beta_n, \gamma)/h \\ -\left(X'_{t-1} D + D'_z x'_{t-1}(\beta_n - \beta_0)\right) \otimes I_p \\ -\mathcal{K}_{t-1}(\beta_n, \gamma)\left(X'_{t-1} D + D'_z x'_{t-1}(\beta_n - \beta_0)\right) \otimes I_p \end{pmatrix}
\end{aligned}
$$

Then, some straightforward but tedious algebra yields that

$$
\frac{\sqrt{h}}{\sqrt{n}} \sum_{t=1}^{n} \left| \frac{\partial \ddot{e}_t\left(\ddot{\theta}_0;\beta_n\right)'}{\partial \ddot{\theta}} \ddot{e}_t\left(\ddot{\theta}_0;\beta_n\right) - \frac{\partial \ddot{e}_t\left(\ddot{\theta}_0;\beta_0\right)'}{\partial \ddot{\theta}} \ddot{e}_t\left(\ddot{\theta}_0;\beta_0\right) \right| = o_p(1).
$$

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

$$
\begin{aligned}
&\frac{\sqrt{h}}{\sqrt{n}} \sum_{t=1}^{n} \left| X'_{t-1} D_0 D'_0 X_{t-1} \left(\mathcal{K}_{t-1}(\beta_n, \gamma_0) - \mathcal{K}_{t-1}\right) \right| \\
\leq\ & \frac{1}{\sqrt{nh}} \sup_t \left| x'_{t-1} n (\beta_n - \beta_0) \right| \frac{1}{n} \sum_{t=1}^{n} \left| X'_{t-1} D_0 D'_0 X_{t-1} \mathcal{K}^{(1)}_{t-1}\left(\tilde{\beta}, \gamma_0\right) \right| = o_p(1),
\end{aligned}
$$

where $\tilde{\beta}$ lies between β_0 and β_n. The remaining terms can be analyzed similarly. ■

References

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

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.

GONZALO, J., AND M. WOLF (2005): "Subsampling Inference in Threshold Autoregressive Models," *Journal of Econometrics*.

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.

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

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.

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.

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.

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.

SAIKKONEN, P. (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*, forthcoming.

SEO, M. (2005): "Unit Root Test in a Threshold Autoregression: Asymptotic Theory and Residual-based Block Bootstrap," http://sticerd.lse.ac.uk/dps/em/em484.pdf.

——— (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 (2006): "A Smoothed Least Squares Estimator For Threshold Regression Models," *Journal of Econometrics*, forthcoming.

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

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.