Estimation of Threshold Cointegration

Myung Hwan Seo

London School of Economics

December 2006
Outline

1. Model
   - Estimation Methods
   - Literature

2. Asymptotics
   - Consistency
   - Convergence Rates
   - Asymptotic Distributions

3. Inference

4. Conclusion
Outline

1 Model
   - Estimation Methods
   - Literature

2 Asymptotics
   - Consistency
   - Convergence Rates
   - Asymptotic Distributions

3 Inference

4 Conclusion
Model I

- **Variables:**
  - \( x_t \): \( p \)-dimensional \( I (1) \) vector that is cointegrated with the cointegrating vector \( \beta \). The first element of \( \beta \) is normalized to 1.
  - \( z_t (\beta) = x_t' \beta \): equilibrium error, or error-correction term.
  - \( X_{t-1} (\beta) = (1, z_{t-1} (\beta), \Delta x_{t-1}', \cdots, \Delta x_{t-l+1}') \): the regressor which is a \((pl + 2)\)-dimensional vector.

- **Two-regime threshold vector error correction model (Balke and Fomby 1997)**

\[
\Delta x_t = \begin{cases} 
  A'X_{t-1} (\beta) + u_t, & \text{if } z_{t-1} (\beta) \leq \gamma \\
  (A + D)' X_{t-1} (\beta) + u_t, & \text{if } z_{t-1} (\beta) > \gamma 
\end{cases}
\]

\( t = l + 1, \ldots, n \). That is,

\[
\Delta x_t = A'X_{t-1} (\beta) + D'X_{t-1} (\beta) \cdot 1 \{ z_t (\beta) > \gamma \} + u_t,
\]

where \( 1 \{ \cdot \} \) is the indicator function.
Matrix notation: Let $\alpha = \text{vec}(A)$ and $\delta = \text{vec}(D)$, where \text{vec} stacks rows of a matrix and

\[
y = \begin{pmatrix}
\Delta x_{l+1} \\
\vdots \\
\Delta x_n
\end{pmatrix},
\begin{pmatrix}
u_{l+1} \\
\vdots \\
u_n
\end{pmatrix} = \begin{pmatrix}
u_{l+1} \\
\vdots \\
u_n
\end{pmatrix}, \quad X(\beta) = \begin{pmatrix}
X'_l(\beta) \\
\vdots \\
X'_{n-1}(\beta)
\end{pmatrix},
\begin{pmatrix}
X^*_l(\beta) \cdot 1 \{z_l(\beta) > \gamma\} \\
\vdots \\
X^*_{n-1}(\beta) \cdot 1 \{z_{n-1}(\beta) > \gamma\}
\end{pmatrix}.
\]

Then,

\[
u^*(\theta, \beta) = y - (X(\beta) \otimes I_p) \alpha + (X^*_\gamma(\beta) \otimes I_p) \delta.
\]
Classification of Parameters:

- Long-run parameter: the cointegrating vector $\beta$
- Short-run parameters: collectively we denote them as $\theta$.
  - Slope parameters: $A$ and $D$ or $\alpha$ and $\delta$
  - Threshold parameters: $\gamma$ ($\beta$?)

Estimation method:

1. Least squares: denoted with the superscript $^*$
2. Smoothed Least squares (Seo & Linton 2005)
Least Squares Estimation

Objective Function:

\[ S_n^* (\theta, \beta) = u^* (\theta, \beta)' u^* (\theta, \beta), \]

where \( \theta \) indicates all the short-run parameters \((\alpha', \delta', \gamma)'\).

LS estimator:

\[
\left( \hat{\theta}^*, \hat{\beta}^* \right) = \arg \min_{\theta, \beta} S_n^* (\theta, \beta),
\]

where the minimum is taken over a compact parameter space.

Concentrated LS estimator: for a fixed \((\beta, \gamma)\),

\[
\begin{bmatrix}
\hat{\alpha}^* (\beta, \gamma) \\
\hat{\delta}^* (\beta, \gamma)
\end{bmatrix}
= \left( \begin{bmatrix}
X (\beta)' X (\beta) & X (\beta)' X^*_\gamma (\beta) \\
X^*_\gamma (\beta)' X (\beta) & X^*_\gamma (\beta)' X^*_\gamma (\beta)
\end{bmatrix}^{-1} \begin{bmatrix}
X (\beta)' \\
X^*_\gamma (\beta)'
\end{bmatrix} \otimes I_p \right) y,
\]

which is then plugged back into \( S_n^* \) for optimization over \((\beta, \gamma)\).
Smoothed Least Squares Estimation (Seo & Linton (2005))

1. Define a bounded function $K(\cdot)$ satisfying that
   \[
   \lim_{s \to -\infty} K(s) = 0, \quad \lim_{s \to +\infty} K(s) = 1.
   \]

2. Let $K_t(\beta, \gamma) = K\left(\frac{z_t(\beta) - \gamma}{\sigma_n}\right)$ for $\sigma_n \to 0$ and replace $1\{z_t(\beta) > \gamma\}$ with $K_t(\beta, \gamma)$:

   \[
   X_{\gamma}(\beta) = \left(\begin{array}{c}
   X'_l(\beta) K_l(\beta, \gamma) \\
   \vdots \\
   X'_{n-1}(\beta) K_{n-1}(\beta, \gamma)
   \end{array}\right)
   \]

   \[
   u(\theta, \beta) = y - (X(\beta) \otimes I_p) \alpha + (X_{\gamma}(\beta) \otimes I_p) \delta.
   \]

3. Then, the Smoothed Least Squares (SLS) estimator is

   \[
   \left(\hat{\theta}, \hat{\beta}\right) = \arg\min_{\theta, \beta} S_n(\theta, \beta),
   \]

   where

   \[
   S_n(\theta, \beta) = u(\theta, \beta)' u(\theta, \beta).
   \]
Smoothed LS II

- Concentration

\[
\begin{bmatrix}
\hat{\alpha}(\beta, \gamma) \\
\hat{\delta}(\beta, \gamma)
\end{bmatrix}
= \left( \begin{bmatrix}
X(\beta)'X(\beta) & X(\beta)'X_\gamma(\beta) \\
X_\gamma(\beta)'X(\beta) & X_\gamma(\beta)'X_\gamma(\beta)
\end{bmatrix} \right)^{-1} \left( \begin{bmatrix}
X(\beta)' \\
X_\gamma(\beta)'
\end{bmatrix} \otimes I_p \right) y.
\]
Applications of Threshold Cointegration:

Literature II

- Testing for threshold cointegration:
  - Threshold effect (assuming the presence of cointegration): Hansen & B. Seo (2002)
- Estimation - No Distribution theory yet.
  - stationary threshold models:
Outline

1. Model
   - Estimation Methods
   - Literature

2. Asymptotics
   - Consistency
   - Convergence Rates
   - Asymptotic Distributions

3. Inference

4. Conclusion
# Review of Asymptotics for stationary threshold models I

<table>
<thead>
<tr>
<th>Model</th>
<th>Least Squares</th>
<th></th>
<th>Smoothed LS</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Slope</td>
<td>Threshold</td>
<td>Threshold</td>
</tr>
<tr>
<td>Discontinuous</td>
<td>$\sqrt{n}$ rate</td>
<td>$n$ rate</td>
<td>$\sqrt{n\sigma_n^{-1}}$ rate</td>
</tr>
<tr>
<td>(Chan)</td>
<td>Normal</td>
<td>complex Poisson</td>
<td>Normal</td>
</tr>
<tr>
<td>Continuous</td>
<td>$\sqrt{n}$ rate</td>
<td>$\sqrt{n}$ rate</td>
<td>$\sqrt{n}$ rate</td>
</tr>
<tr>
<td>(Chan &amp; Tsay)</td>
<td>Normal</td>
<td>Normal</td>
<td>Normal</td>
</tr>
<tr>
<td>Diminishing</td>
<td>$\sqrt{n}$ rate</td>
<td>$n^{1-2\alpha}$ rate</td>
<td>$\sqrt{n^{1-2\alpha}\sigma_n^{-1}}$ rate</td>
</tr>
<tr>
<td>(Hansen)</td>
<td>Normal</td>
<td>Function of BM</td>
<td>Normal</td>
</tr>
</tbody>
</table>

If the cointegrating vector is known, these results will apply.
Assumption (1)

(a) \( \{u_t\} \) is an independent and identically distributed sequence with \( \mathbb{E}u_t = 0, \mathbb{E}u_t u_t' = \Sigma \) that is positive definite. Furthermore, it has a density \( \text{wrt Lebesque measure that is everywhere positive.} \)

(b) \( \{\Delta x_t, z_t\} \) is a sequence of strictly stationary strong mixing random variables with mixing numbers \( \alpha_m, m = 1, 2, \ldots \), that satisfy \( \alpha_m = o\left(m^{-\left(\alpha_0 + 1\right)/\left(\alpha_0 - 1\right)}\right) \) as \( m \to \infty \) for some \( \alpha_0 \geq 1 \), and for some \( \varepsilon > 0 \), \( \mathbb{E}\left|X_t X_t^\top\right|^{\alpha_0 + \varepsilon} < \infty \) and \( \mathbb{E}\left|X_{t-1} u_t\right|^{\alpha_0 + \varepsilon} < \infty \). Furthermore, \( \mathbb{E}\Delta x_t = 0 \) and \( x_{[ns]} / \sqrt{n} \) converges weakly to a vector Brownian motion \( \mathbf{B} \) with the covariance matrix \( \Sigma \), which is the long-run covariance matrix of \( \Delta x_t \) and has rank \( p - 1 \) s.t. \( \beta_0' \Sigma = 0. \)

(c) \( \mathbb{E}\left[X_{t-1}' D_0 D_0' X_{t-1} | z_{t-1}\right] > 0 \) a.s.

- Condition (c) implies the discontinuity of the model
Theorem (1)

Under Assumption 1, $\sqrt{n} \left( \hat{\beta}^* - \beta_0 \right)$ and $\left( \hat{\theta}^* - \theta_0 \right)$ are $o_p(1)$. In addition, assume that $s^2 \{s > 0\} - \mathcal{K}(s) < M < \infty$, for any $s \in \mathcal{R}$. Then, $\sqrt{n} \left( \hat{\beta} - \beta_0 \right)$ and $\left( \hat{\theta} - \theta_0 \right)$ are $o_p(1)$.

- Sketch of Proof
  1. Show $\sqrt{n} \left( \hat{\beta}^* - \beta_0 \right) = O_p(1)$ by contradiction.
  2. Show $\sqrt{n} \left( \hat{\beta}^* - \beta_0 \right)$ and $\left( \hat{\theta}^* - \theta_0 \right)$ are $o_p(1)$ based on ULLN.
  3. Show the difference between $S_n$ and $S_n^*$ are asymptotically uniformly negligible.
Consistency

**Theorem (1)**

*Under Assumption 1, \( \sqrt{n} \left( \hat{\beta}^* - \beta_0 \right) \) and \( \left( \hat{\theta}^* - \theta_0 \right) \) are \( o_p \left( 1 \right) \). In addition, assume that \( s^2 |1 \{s > 0\} - K(s)| < M < \infty \), for any \( s \in \mathcal{R} \). Then, \( \sqrt{n} \left( \hat{\beta} - \beta_0 \right) \) and \( \left( \hat{\theta} - \theta_0 \right) \) are \( o_p \left( 1 \right) \).*

**Sketch of Proof**

1. Show \( \sqrt{n} \left( \hat{\beta}^* - \beta_0 \right) = O_p \left( 1 \right) \) by contradiction.
2. Show \( \sqrt{n} \left( \hat{\beta}^* - \beta_0 \right) \) and \( \left( \hat{\theta}^* - \theta_0 \right) \) are \( o_p \left( 1 \right) \) based on ULLN.
3. Show the difference between \( S_n \) and \( S_n^* \) are asymptotically uniformly negligible.
Convergence Rate

- Long-run vs. Short-run parameters
- Threshold vs. Slope parameters
- Smoothed vs. Unsmoothed estimator

**Theorem (2)**

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

- In consequence, \( \hat{\alpha}^* \) and \( \hat{\delta}^* \) converge at the rate of \( \sqrt{n} \).
Convergence Rate

- Long-run vs. Short-run parameters
- Threshold vs. Slope parameters
- Smoothed vs. Unsmoothed estimator

Theorem (2)

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

- In consequence, \( \hat{\alpha}^* \) and \( \hat{\delta}^* \) converge at the rate of \( \sqrt{n} \).
Assumption (2)

(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 \( \alpha_m, m = 1, 2, \ldots, \) that satisfy
\[
\alpha_m \leq Cm^{-(2r-2)/(r-2)-\eta} \quad \text{for positive } C \text{ and } \eta, \text{as } m \to \infty.
\]

(c) For some integer \( h \geq 1 \) and each integer \( i \) such that \( 1 \leq i \leq h, \) all \( z \) in a neighborhood of \( \gamma, \) almost every \( \Delta, \) and some \( M < \infty, f^{(i)} (z|\Delta) \) exists and is a continuous function of \( z \) satisfying \( |f^{(i)} (z|\Delta)| < M. \) In addition, \( f (z|\Delta) < M \) for all \( z \) and almost every \( \Delta. \)

(d) and 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 \( (\Delta_t, \Delta_{t-m}). \)

(e) \( \theta_0 \) is an interior point of \( \Theta. \)
Assumption (3)

(a) $\mathcal{K}$ is twice differentiable everywhere, $|\mathcal{K}(1) (\cdot)|$ and $|\mathcal{K}(2) (\cdot)|$ are uniformly bounded, and each of the following integrals is finite:

\[
\int |\mathcal{K}(1)|^4, \int |\mathcal{K}(2)|^2, \int |v^2 \mathcal{K}(2) (v)| dv.
\]

(b) For some integer $h \geq 1$ and each integer $i$ ($1 \leq i \leq h$),

\[
\int |v^i \mathcal{K}(1) (v)| dv < \infty, \text{ and}
\]

\[
\int s^{i-1} \text{sgn} (s) \mathcal{K}(1) (s) ds = 0,
\]

and

\[
\int s^h \text{sgn} (s) \mathcal{K}(1) (s) ds \neq 0,
\]

and $\mathcal{K} (x) - \mathcal{K} (0) \geq 0$ if $x \geq 0$. 

Seo
Threshold Cointegration
Assumption (3 cont’d)

(c) For each integer \( i (0 \leq i \leq h) \), and \( \eta > 0 \), and any sequence \( \{\sigma_n\} \) converging to 0,

\[
\lim_{n \to \infty} \sigma_n^{i-h} \int_{|\sigma_n s| > \eta} \left| s^i \mathcal{K}^{(1)} (s) \right| ds = 0,
\]

and

\[
\lim_{n \to \infty} \sigma_n^{-1} \int_{|\sigma_n s| > \eta} \left| \mathcal{K}^{(2)} (s) \right| ds = 0.
\]

(d) \( \limsup_{n \to \infty} n \sigma_n^{2h} < \infty \) and

\[
\lim_{n \to \infty} \sigma_n^{-2h} \int_{|\sigma_n s| > \eta} \left| \mathcal{K}^{(1)} (s) \right| ds = 0.
\]
Assumption (3 cont’d)

(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) For some sequence $m_n \geq 1$, and $\varepsilon > 0$,

$$\log (nm_n) \left( n^{1-6/\rho} \sigma_n^2 m_n^{-2} \right)^{-1} \rightarrow 0$$

$$\sigma_n^{3k-1} n^{3/\rho + \varepsilon} \alpha_{m_n} \rightarrow 0.$$
Condition (f) serves to determine the rate for $\sigma_n$. When the data are i.i.d. and the regressors possess a moment generating function, the conditions can be weakened to

$$\frac{\log(n)}{n\sigma_n^2} \rightarrow 0,$$

since $\alpha m_n = 0$ and we can set $m_n = 1$ in this case. Although condition (e) provides permissible rates for the bandwidth selection, it may not be sharp.
Let $B$ be the limit Brownian motion of the partial sum process of $\Delta x_t$, whose covariance matrix is $\Omega$, and

$$
\sigma_v^2 = \mathbb{E} \left[ \|K^{(1)}\|_2^2 \left( X'_{t-1} D_0 u_t \right)^2 + \|\tilde{K}^{(1)}\|_2^2 \left( X'_{t-1} D_0 D'_0 X_{t-1} \right)^2 | z_{t-1} = \gamma_0 \right] f(\gamma_0)
$$

$$
\sigma_q^2 = K^{(1)}(0) \mathbb{E} \left( X'_{t-1} D_0 D'_0 X_{t-1} | z_{t-1} = \gamma_0 \right) f(\gamma_0),
$$

where $K^{(i)}$ is the $i$-th derivative of $K$, $\tilde{K}^{(1)}(s) = K^{(1)}(s) \left( 1 \{s > 0\} - K(s) \right)$, and $\|g\|_2^2 = (\int g^2)$.
Theorem

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

$$\left( \begin{array}{c}
n\sigma_n^{-1/2} \left( \hat{\beta} - \beta_0 \right) \\
\sqrt{n}\sigma_n^{-1} \left( \hat{\gamma} - \gamma_0 \right)
\end{array} \right) \overset{d}{\to} \frac{\sigma_v}{\sigma_q^2} \left( \begin{array}{cc}
\int_0^1 B B' & \int_0^1 B \\
\int_0^1 B' & 1
\end{array} \right)^{-1} \left( \begin{array}{c}
\int B dW \\
W(1)
\end{array} \right),
$$

$$\sqrt{n} \left( \begin{array}{c}
\hat{\alpha} - \alpha_0 \\
\hat{\delta} - \delta_0
\end{array} \right) \overset{d}{\to} \mathcal{N} \left( 0, \left[ \mathbf{E} \left( \begin{array}{cc}
1 & d_{t-1} \\
d_{t-1} & d_{t-1}
\end{array} \right) \otimes X_{t-1} X'_{t-1} \right]^{-1} \otimes \Sigma \right)$$

and they are asymptotically independent. The unsmoothed estimator $\hat{\alpha}^*$ and $\hat{\delta}^*$ have the same asymptotic distribution as $\hat{\alpha}$ and $\hat{\delta}$. 

Asymptotic Distribution VII
Corollary

Let \( \hat{\gamma} (\beta) \) be the smoothed estimator of \( \gamma \) when \( \beta \) is given. Then, \( \hat{\gamma} (\hat{\beta}^*) \) has the same asymptotic distribution as that of \( \hat{\gamma} (\beta_0) \), which is \( N \left( 0, \frac{\sigma_v^2}{\sigma_q^4} \right) \).

Thus, we do not need to estimate the long-run variance for the inference for \( \gamma \).

Suppose that \( \beta_n = \beta + O_p (n^{-1}) \). For example, \( \beta_n \) can be obtained from the simple OLS of the first element of \( x_t \) to the other elements of \( x_t \), as in Engle-Granger procedure. Note that, by multiplying \( \beta_0 \) both sides of the model

\[
\Delta x_t = \begin{cases} 
A'X_{t-1} (\beta) + u_t, & \text{if } z_{t-1} (\beta) \leq \gamma \\
(A + D)' X_{t-1} (\beta) + u_t, & \text{if } z_{t-1} (\beta) > \gamma 
\end{cases}
\]
we obtain a threshold autoregressive process for $z_t = x_t' \beta_0$

$$
\Delta z_t = \begin{cases} 
  a_0 z_{t-1} + a_1 \Delta z_{t-1} + \cdots + u_t, & \text{if } z_{t-1} \leq \gamma \\
  (a_0 + d_0) z_{t-1} + (a_1 + d_1) \Delta z_{t-1} + \cdots + u_t, & \text{if } z_{t-1} > \gamma
\end{cases}
$$

But, the plug-in estimate $\hat{\alpha}(\beta_n)$ and $\hat{\delta}(\beta_n)$ do not have the same asymptotic distribution as $\hat{\alpha}(\beta_0)$ and $\hat{\delta}(\beta_0)$.

To treat the estimate $\beta_n$ as the true value $\beta_0$, we first need

$$
\frac{1}{\sqrt{n}} \sum_{t} \Delta x_{t-k} K_{t-1} (\beta_n, \gamma_n) x'_{t-1} (\beta_n - \beta_0) = o_p (1)
$$
In case of $\mathbb{E}z_t = 0$, we can still retain the Normality by replacing the $z_{t-1} \left( \tilde{\beta} \right)$ 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 the asymptotic variance increases by doing this.
Outline

1 Model
   - Estimation Methods
   - Literature

2 Asymptotics
   - Consistency
   - Convergence Rates
   - Asymptotic Distributions

3 Inference

4 Conclusion
Asymptotic Variance Estimation:

- long-run covariance matrix $\Omega$: see Andrews (1991) for example.
- variance of threshold estimate ($\sigma_v$ and $\sigma_q$): Let

$$\hat{\tau}_t = \frac{1}{\sqrt{n}} X_{t-1} \left( \hat{\beta} \right)' D K_{t-1}^{(1)} \left( \hat{\beta}, \hat{\gamma} \right) \hat{u}_t,$$

where $\hat{u}_t$ is the regression residual, and let

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

where $Q_{n22}$ is the diagonal element corresponding to $\gamma$ of the second derivative of $S_n$. Refer to Seo and Linton (2005).

- variance of slope estimate: the same as linear regression.

We do not have to estimate the density and conditional expectation by a nonparametric method. Hansen (2000) uses a nonparametric method to estimate the asymptotic variance; the introduction of a smoothing parameter is necessary.
Testing for two thresholds: To test the null of one threshold against two threshold, we can extend the SupLM test statistic of Hansen and Seo (2002) that examines the null of no threshold. Due to the fast convergence rate of the least squares estimator of $\beta$ and $\gamma$, we can treat the estimated $\beta$ and $\gamma$ as if known. For the p-value, the fixed regressor bootstrap and residual bootstrap can be employed as described there.
The paper establishes the consistency and convergence rates of the LSE and SLSE of the threshold vector error correction model. In particular,

- The LSE of the cointegrating vector estimate converges extremely fast. This validates a two-step estimation, in which the cointegrating vector is first estimated by LS and the other short-run parameters by SLS.
- The LSE of the slope is asymptotically not affected by the estimation of the other parameters.

The limit distribution of the SLSE is derived and thus providing a way of inference for the cointegrating vector.

- Multiple-Regime Threshold: We can estimate the thresholds simultaneously or sequentially. Hansen (1999) and Bai and Perron (1998).
- More than one cointegrating relation.
Conclusion

- The paper establishes the consistency and convergence rates of the LSE and SLSE of the threshold vector error correction model. In particular,
  - The LSE of the cointegrating vector estimate converges extremely fast. This validates a two-step estimation, in which the cointegrating vector is first estimated by LS and the other short-run parameters by SLS.
  - The LSE of the slope is asymptotically not affected by the estimation of the other parameters.
- The limit distribution of the SLSE is derived and thus providing a way of inference for the cointegrating vector.
- Multiple-Regime Threshold: We can estimate the thresholds simultaneously or sequentially. Hansen (1999) and Bai and Perron (1998).
- More than one cointegrating relation.