Nonlinear Error Correction Model and Multiple-Threshold Cointegration

Man Wang
Dong Hua University, China

Joint work with N.H.Chan

May 23, 2014

Linear Cointegration:
If $x_{1,t}$ and $x_{2,t}$ are both $I(1)$, and $x_{1,t} + \beta x_{2,t}$ is $I(0)$, then we say that $x_{1,t}$ and $x_{2,t}$ are cointegrated.

Granger representation theorem: $(x_{1,t}, x_{2,t})$ has a vector error correction model (VECM) representation as

$$\begin{bmatrix} \Delta x_{1,t} \\ \Delta x_{2,t} \end{bmatrix} = \begin{bmatrix} c_1 \\ c_2 \end{bmatrix} + \begin{bmatrix} a_1 \\ a_2 \end{bmatrix} z_{t-1} + \begin{bmatrix} b_{11} & b_{12} \\ b_{21} & b_{22} \end{bmatrix} \begin{bmatrix} \Delta x_{1,t-1} \\ \Delta x_{2,t-1} \end{bmatrix} + \begin{bmatrix} u_{1,t} \\ u_{2,t} \end{bmatrix},$$

where $z_{t-1} = x_{1,t-1} + \beta x_{2,t-1}$ is the error correction term, called the equilibrium error.

Threshold Cointegration and TVECM

The linear cointegration implies consistent adjustment towards long-run equilibrium.

Balke and Fomby (1997) proposed threshold cointegration, assuming that the system adjustment follows a threshold model.
The threshold vector ECM (TVECM):

$$\begin{bmatrix} \Delta x_{1,t} \\ \Delta x_{2,t} \end{bmatrix} = \sum_{j=1}^{3} \left(\begin{bmatrix} c_1 \\ c_2 \end{bmatrix}^{(j)} + \begin{bmatrix} a_1 \\ a_2 \end{bmatrix}^{(j)} z_{t-1} + \begin{bmatrix} b_{11} & b_{12} \\ b_{21} & b_{22} \end{bmatrix}^{(j)} \begin{bmatrix} \Delta x_{1,t-1} \\ \Delta x_{2,t-1} \end{bmatrix} \right) 1\{z_{t-1} \text{in } j - \text{th regime}\} + \begin{bmatrix} u_{1,t} \\ u_{2,t} \end{bmatrix}^{(j)}$$

Variables:

- x_t : $p-$ dimensional $I(1)$ vector, cointegrated with single cointegrating vector $(1, \beta^{'})$.
- $z_{t-1}(\beta) = x_{1,t-1} + x^{'}_{2,t-1}\beta$: the error correction term, where $x_t = (x_{1,t}, x^{'}_{2,t})^{'}$.
- $X_{t-1}(\beta) = (1, z_{t-1}(\beta), \Delta x^{'}_{t-1}, \ldots, \Delta x^{'}_{t-l})$, where $\Delta x_{t-i}, i = 1, \ldots, l$ are the lagged difference terms.

An m-regime TVECM:

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

$-\infty = \gamma_0 < \gamma_1 < \ldots < \gamma_m = \infty$ are thresholds and A_j is coefficient in the j-th regime.

Multiple-threshold cointegration is widely used, but estimation theories are limited to the one threshold cointegration (two-regime TVECM).

- Hansen and Seo (2002) constructed the MLE whose consistency is still unknown.
- Seo (2011) gave the asymptotics of the LSE and SLSE.

Goal: estimation of multiple threshold cointegration (m-regime TVECM with $m > 2$).

An $m-$regime TVECM:

$$\Delta x_t = \sum_{j=1}^{m} A_j^{'} X_{t-1}(\beta) 1\{\gamma_{j-1} \leq z_{t-1}(\beta) < \gamma_j\} + u_t, \qquad t = l+1, \ldots, n.$$

We denote:

$$\tilde{X}_j(\beta,\gamma) = \begin{pmatrix} X_l^{'}(\beta)1\{\gamma_{j-1} \leq z_l(\beta) < \gamma_j\} \\ \vdots \\ X_{n-1}^{'}(\beta)1\{\gamma_{j-1} \leq z_{n-1}(\beta) < \gamma_j\} \end{pmatrix},$$

$y = \begin{pmatrix} \Delta x_{l+1}^{'}, & \ldots, & \Delta x_n^{'} \end{pmatrix}^{'}$, $u = \begin{pmatrix} u_{l+1}, & \ldots, & u_n \end{pmatrix}^{'}$,
$\lambda = \text{vec}\left(A_1^{'}, \ldots, A_m^{'}\right)$,

Then,

$$y = \left(\left(\tilde{X}_1(\beta,\gamma), \ldots, \tilde{X}_m(\beta,\gamma)\right) \otimes I_p\right) \lambda + u.$$

Let $S_n(\theta) = u'u$,

1. LS estimator:

$$\hat{\theta} = \underset{\theta \in \Theta}{\text{argmin}}\ S_n(\theta). \quad \theta = (\beta, \gamma = (\gamma_1, \gamma_2, \ldots, \gamma_{m-1}), \lambda')' \in \Theta.$$

2. 2-step procedure:

   Fix (β, γ), $y = \left[\left(\tilde{X}_1(\beta,\gamma), \tilde{X}_2(\beta,\gamma), \ldots, \tilde{X}_m(\beta,\gamma)\right) \otimes I_p\right] \lambda + u.$

$$\hat{\lambda}(\beta,\gamma) = \left(\begin{bmatrix} \tilde{X}_1{}'\tilde{X}_1 & \tilde{X}_1{}'\tilde{X}_2 & \ldots & \tilde{X}_1{}'\tilde{X}_m \\ \tilde{X}_1{}'\tilde{X}_2 & \tilde{X}_2{}'\tilde{X}_2 & \ldots & \tilde{X}_2{}'\tilde{X}_m \\ & & \ldots & \\ \tilde{X}_1{}'\tilde{X}_m & \tilde{X}_m{}'X_2 & \ldots & \tilde{X}_m{}'\tilde{X}_m \end{bmatrix}^{-1} \begin{pmatrix} \tilde{X}_1{}' \\ \tilde{X}_2{}' \\ \vdots \\ \tilde{X}_m{}' \end{pmatrix} \otimes I_p \right) y. \quad (2)$$

   Step1: $(\hat{\beta}, \hat{\gamma}) = \underset{(\beta,\gamma) \in \Theta_{\beta,\gamma}}{\text{argmin}}\ S_n(\beta, \gamma, \hat{\lambda}(\beta,\gamma))$,

   Step2: $\hat{\lambda} = \hat{\lambda}(\hat{\beta}, \hat{\gamma})$.

Assumptions 1.

1.1 The parameter space Θ is compact with max $\{|\lambda_{z_1}|, |\lambda_{z_m}|\}$ and $\min_{1\leq i<j\leq m-1}\{|\gamma_i-\gamma_j|\}$ bounded away from zero.

1.2 $\{u_t\}$ is an i.i.d. sequence of random variables with $\mathsf{E}u_t = 0$, $\mathsf{E}u_t u_t^{'} = \Sigma$.

1.3 $\{\triangle x_t, z_t\}$ is a sequence of strictly stationary strong mixing random vectors with mixing numbers $\alpha_m, m = 1, 2, \ldots,$ that satisfy $\alpha_m = o(m^{-(\alpha_0+1)/(\alpha_0-1)})$ as $m \to \infty$ for some $\alpha_0 > 1$, and for some $\varepsilon > 0$. $\mathsf{E}|X_t X_t^{'}|^{\alpha_0-1} < \infty$ and $\mathsf{E}|X_{t-1}u_t|^{\alpha_0+\varepsilon}| < \infty$. Furthermore, $\mathsf{E}\triangle x_t = 0$, and the partial sum process, $x_{[ns]}/\sqrt{n}, s \in [0,1]$, converges weakly to a vector Brownian motion **B** with a covariance matrix $\mathbf{\Omega}$, which is the long-run covariance matrix of $\triangle x_t$ and has rank $p-1$ such that $(1, \beta_0^{'})\mathbf{\Omega} = 0$. In particular, assume that $x_{2[nx]}/\sqrt{n}$ converges weakly to a vector of Brownian motion B with a covariance matrix Ω, which is finite and positive definite.

1.4 Let $u_t(\xi, \gamma, \lambda)$ be defined as the error u_t by replacing $z_t(\beta)$ by $z_t + \xi$, where ξ belongs to a compact set in R and let

$$S(\xi, \gamma, \lambda) = \mathsf{E}(u_t^{'}(\xi, \gamma, \lambda)u_t(\xi, \gamma, \lambda)).$$

Assumptions 2.

2.1 The probability distribution of z_t has a density with respect to the Lebesgue measure that is continuous, bounded, and everywhere positive, and that the density function $f(z_t|x_{2,t})$ is bounded by $M > 0$ for almost every $x_{2,t},\ t = 1, 2, \ldots, n$.

2.2 $\mathsf{E}[X'_{t-1}(A_j^0 - A_{j+1}^0)(A_j^0 - A_{j+1}^0)' X_{t-1}|z_{t-1} = \gamma_j^0] > 0 \ \ \forall\, j = 1, 2, \ldots, m-1$.

Convergence rate of LSE

Theorem 1

Under Assumptions 1 and 2, $\hat{\theta}$ *is consistent, further* $n^{3/2}(\hat{\beta} - \beta^0)$ *is* $O_p(1)$ *and* $n(\hat{\gamma} - \gamma^0) = O_p(1)$, $n(\hat{\lambda} - \lambda^0)$ *is asymptotically normal.*

Smoothed Least Squares Estimator

The objective function of the LSE of a threshold model is discontinuous. Seo and Linton (2007) proposed to smooth it by replacing $1\{s > 0\}$ by a smooth function $\mathcal{K}(s/h)$, where h is bandwidth.
The minimizer of the smoothed objective function is called the SLSE.
Herein $\mathcal{K}(s)$ is smooth, bounded, with

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

The smoothing is reasonable because

$$\mathcal{K}(s/h) \to 1\{s > 0\}, \text{as } h \to 0.$$

SLSE for an $m-$regime TVECM:

$$\Delta x_t = \sum_{j=1}^{m} A_j^{'} X_{t-1}(\beta) 1\{\gamma_{j-1} \leq z_{t-1}(\beta) < \gamma_j\} + u_t, \quad t = l+1, \ldots, n. \quad (3)$$

- Let

$$\mathcal{K}_{t-1}(\beta, \gamma_{j-1}, \gamma_j) = \mathcal{K}(\frac{z_{t-1}(\beta) - \gamma_{j-1}}{h}) + \mathcal{K}(\frac{\gamma_j - z_{t-1}(\beta)}{h}) - 1,$$

  and replace $1\{\gamma_{j-1} \leq z_{t-1}(\beta) < \gamma_j\}$ with $\mathcal{K}_{t-1}(\beta, \gamma_{j-1}, \gamma_j)$.
- Denote $X_j^*(\beta, \gamma) = X_{t-1}(\beta)\mathcal{K}_{t-1}(\beta, \gamma_{j-1}, \gamma_j)$, then the smoothed objective function become

$$S_n^* = {u^*}^{'} u^*, \text{where} \quad u^* = y - \left[(X_1^*(\beta, \gamma), \ldots, X_m^*(\beta, \gamma)) \otimes I_p \right] \lambda.$$

- $\hat{\theta}^* = \underset{\theta \in \Theta}{\text{argmin}} S_n^*$ is defined as SLSE of $m-$regime TVECM.

Assumptions 3.

3.1 $\mathcal{K}$ is twice differentiable everywhere, $\mathcal{K}^{(1)}$ is symmetric around zero, and $\mathcal{K}^{(1)}$ and $\mathcal{K}^{(2)}$ are uniformly bounded and uniformly continuous. Furthermore, $\int |\mathcal{K}^{(1)}(s)|^4 ds$, $\int |s^2\mathcal{K}^{(1)}(s)|ds$,$\int |\mathcal{K}^{(2)}(s)|^2 ds$ and $\int |s^2\mathcal{K}^{(2)}(s)|ds$ are finite.

3.2 $\mathcal{K}(x) - \mathcal{K}(0)$ has the same sign as x and for some integer $\nu \geq 2$ and each integer i $(1 \leq i \leq \nu)$, $\int |s^i\mathcal{K}^{(1)}(s)|ds < \infty$,

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

3.3 Furthermore, for some $\epsilon > 0$,

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

3.4 $\{h\}$ is a function of n and satisfies that for some sequence $q \geq 1$,

$$nq^3 \to 0,$$
$$\log(nq)(n^{1-6/r}h^2q^{-2})^{-1} \to 0,$$

and $h^{-9k/2-3}n^{(9k/2+2)/r+\epsilon}\alpha_q \to 0$, where k is the dimension of θ and $r > 4$ is specified in Assumptions 4.

Assumptions 4.

4.1 $\mathsf{E}[|X_t u_t^{'}|^r] < \infty$, $\mathsf{E}[|X_t X_t^{'}|^r] < \infty$, for some $r > 4$.

4.2 $\{\triangle x_t, z_t\}$ is a sequence of strictly stationary strongly mixing random vectors with mixing coefficients $\alpha_m, m = 1, 2, \ldots,$ that satisfy $\alpha_m \leq Cm^{-(2r-2)/(r-2)-\eta}$ for positive C and η, as $m \to \infty$.

4.3 For some integer $\nu \geq 2$ and each integer i such that $1 \leq i \leq \nu - 1$, for all z in a neighborhood of threshold $\gamma_j, j = 1, \ldots, m-1$, for almost every $x_{2,t}$ and some $M < \infty, f^{(i)}(z|x_{2,t})$ exists and is a continuous function of z satisfying $|f^{(i)}(z|x_{2,t})| < M$. In addition, $f(z|x_{2,t}) < M$ for all z and almost every $x_{2,t}$.

4.4 The conditional joint density $f(z_t, z_{t-m}|x_{2,t}, x_{2,t-m}) < M$, for all (z_t, z_{t-m}) and almost every $(x_{2,t}, x_{2,t-m})$.

4.5 θ^0 is an interior point of Θ.

Limiting distribution of $\hat{\theta}^*$.

Theorem 2

Under Assumptions 1-4,

$$\begin{pmatrix} nh^{-1/2}(\hat{\beta}^* - \beta_0) \\ \sqrt{nh^{-1}}(\hat{\gamma}^* - \gamma^0) \end{pmatrix}$$

$$\stackrel{d}{\Longrightarrow} \begin{pmatrix} \sigma_q^2 \int_0^1 BB' & \sigma_{q_1}^2 \int_0^1 B' & \sigma_{q_2}^2 \int_0^1 B' & \cdots & \sigma_{q_{m-1}}^2 \int_0^1 B' \\ \sigma_{q_1}^2 \int_0^1 B & \sigma_{q_1}^2 & 0 & \cdots & 0 \\ & & & \cdots & \\ \sigma_{q_{m-1}}^2 \int_0^1 B & 0 & 0 & \cdots & \sigma_{q_{m-1}}^2 \end{pmatrix}^{-1} \begin{pmatrix} \int Bd\sum_{j=1}^{m-1} \sigma_{v_j} W_j \\ \sigma_{v_1} W_1(1) \\ \sigma_{v_2} W_2(1) \\ \vdots \\ \sigma_{v_{m-1}} W_{m-1}(1) \end{pmatrix},$$

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

and $\begin{pmatrix} nh^{-1/2}(\hat{\beta}^* - \beta_0) \\ \sqrt{nh^{-1}}(\hat{\gamma}^* - \gamma^0) \end{pmatrix}$ *is asymptotically independent of* $\sqrt{n}(\hat{\lambda}^* - \lambda^0)$. *Here* $\stackrel{d}{\Longrightarrow}$ *stands for convergence in distribution.*

Notations in Theorem 2 are as follows:

- $B, W_1, \ldots, W_{m-1}$ are mutually independent Brownian motions.
- $I_j = 1\{\gamma^0_{j-1} \le z_{t-1} < \gamma^0_j\}$, here γ^0 is the true value of threshold parameters.
- $\tilde{\mathcal{K}}_1(s) = \mathcal{K}^{(1)}(1(s>0) - K(s))$. For $j = 1, \ldots, m-1$,
  $\sigma^2_{v_j} = \mathsf{E}\left[F_j | z_{t-1} = \gamma^0_j\right] f_Z(\gamma^0_j)$, where
  $F_j = ||\mathcal{K}^{(1)}||^2_2 (X'_{t-1}(A^0_j - A^0_{j+1})u_t)^2 + ||\tilde{\mathcal{K}_1}||^2_2 (X'_{t-1}(A^0_j - A^0_{j+1})(A^0_j - A^0_{j+1})' X_{t-1})$.
- For $j = 1, \ldots, m-1$,
  $\sigma^2_{q_j} = \mathcal{K}^{(1)}(0)\mathsf{E}[X'_{t-1}(A^0_j - A^0_{j+1})(A^0_j - A^0_{j+1})' X_{t-1})|z_{t-1} = \gamma^0_j] f_Z(\gamma^0_j)$
  and $\sigma^2_q = \sum_{j=1}^{m-1} \sigma^2_{q_j}$.

Model specification:

$$\Delta x_t = \begin{pmatrix} -1 \\ 0.8 \end{pmatrix} z_{t-1} 1\{z_{t-1} < \gamma_{10}\} + \begin{pmatrix} -1 \\ 0 \end{pmatrix} z_{t-1} 1\{\gamma_{10} \leq z_{t-1} < \gamma_{20}\}$$
$$+ \begin{pmatrix} -1 \\ 0.3 \end{pmatrix} z_{t-1} 1\{z_{t-1} \geq \gamma_{20}\} + u_t,$$

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

Case 1: $\gamma^0 = (-1, 1)$, Case 2: $\gamma^0 = (-3, 3)$.

Sample sizes n=100 and 250 for case 1, and n=250 for case 2.
800 replications for each experiment.

Notation of estimators.

- $\tilde{\beta}$: the Johansen's MLE.
- $\hat{\beta}, \hat{\gamma}$: the LSE.
- $\hat{\beta}^*, \hat{\gamma}^*$ the SLSE.
- $\hat{\gamma}_r(\hat{\gamma}_r^*)$ the restricted LSE (SLSE) when β^0 is given.
- $\hat{\beta}_s$ and $\hat{\gamma}_s$ ($\hat{\beta}_s^*$ and $\hat{\gamma}_s^*$) are the sequential LSE (SLSE). That is, the two thresholds are selected sequentially.

Comparison of estimation for different n.

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

Conclusion: the performance of the LSE and SLSE is much improved when $n \uparrow$.

Comparison of estimation for different γs.

	mean	sd	MAE in log
case 1, n=250,	$\gamma^0 = c(1, -1)$		
$\hat{\beta} - \beta^0$	7.8611e-06	0.00054	-7.7729
$\hat{\gamma}_1 - \gamma_1^0$	0.1821	0.9852	-0.2869
$\hat{\gamma}_2 - \gamma_2^0$	-0.0679	1.6558	0.3177
$\hat{\beta}^* - \beta^0$	1.7008e-05	0.00058	-7.7381
$\hat{\gamma}_1^* - \gamma_1^0$	0.1448	1.0029	-0.2587
$\hat{\gamma}_2^* - \gamma_2^0$	-0.0525	1.6951	0.3334
case 2, n=250,	$\gamma^0 = c(3, -3)$		
$\hat{\beta} - \beta^0$	-3.0012e-02	0.2433	-3.4921
$\hat{\gamma}_1 - \gamma_1^0$	0.1623	0.5483	-1.3868
$\hat{\gamma}_2 - \gamma_2^0$	-0.6184	1.1678	-0.2426
$\hat{\beta}^* - \beta^0$	-3.0029e-02	0.2433	-3.4920
$\hat{\gamma}_1^* - \gamma_1^0$	0.1566	0.5421	-1.3881
$\hat{\gamma}_2^* - \gamma_2^0$	-0.6757	1.2554	-0.1726

Conclusion: the performance of the LSE and SLSE declines when magnitude of $\gamma \uparrow$.

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

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

For estimators of β,

- The Johansen's MLE does not perform as well as the other estimators.
- The LSE outperforms the SLSE.

For estimators of γ,

- The LSE $\hat{\gamma}$ performs slightly better than the SLSE $\hat{\gamma}^*$.
- $\hat{\gamma}$ and $\hat{\gamma}^*$ shows superiority over the sequential estimators $\hat{\gamma}_s$ and $\hat{\gamma}_s^*$ respectively.
- $\hat{\gamma}_r$ and $\hat{\gamma}_r^*$ outperform the unrestricted estimators $\hat{\gamma}$ and $\hat{\gamma}^*$. Improvement from $\hat{\gamma}$ to $\hat{\gamma}_r$ is limited.

Conclusions.
Simulation results agree with the the theories developed.
With our choice of $\mathcal{K}$ and h, the SLSE performs almost as we as the LSE and is therefore recommended.

Campbell and Shiller (1987): the term structure of interest rates implies that long-term and short-term interest rates are cointegrated with $\alpha = (1, -1)$.

R_t and r_t: interest rates of the twelve-month and 120-month bonds of the US during in the period 1952-1991.

Hansen and Seo (2002) found them to be threshold cointegrated via a two-regime TVECM estimated by the MLE:

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

$$\Delta r_t = \begin{cases} 1.45 + 1.41 z_{t-1} + 0.92 \Delta R_{t-1} - .04 \Delta r_{t-1} + u_{2,t}, & \text{if } z_{t-1} \leq -0.63, \\ -0.04 + 0.04 z_{t-1} - 0.07 \Delta R_{t-1} + 0.23 \Delta r_{t-1} + u_{2,t}, & \text{if } z_{t-1} > -0.63, \end{cases}$$

where $z_{t-1} = R_{t-1} - 0.984 r_{t-1}$. The estimated cointegrating vector $(1, -0.984)$ is very close to $(1, -1)$ and (8%,92$\%$) of data fall into the two regimes.

We take one step further and consider a three-regime TVECM. By the LSE, the model is estimated as:

$$\Delta R_t = \begin{cases} 0.53 + 0.37 z_{t-1} + .37\Delta R_{t-1} - .28\Delta r_{t-1} + u_{1,t}, & \text{if } z_{t-1} \le -0.62, \\ 0.44 - 0.94 z_{t-1} - 0.06\Delta R_{t-1} + 0.11\Delta r_{t-1} + u_{1,t}, & \text{if } 1.58 \ge z_{t-1} > -0.62, \\ -0.02 - 0.00 z_{t-1} - 0.02\Delta R_{t-1} + 0.11\Delta r_{t-1} + u_{1,t}, & \text{if } z_{t-1} > 1.58, \end{cases}$$

$$\Delta r_t = \begin{cases} 1.45 + 1.43 z_{t-1} + 0.74\Delta R_{t-1} - .28\Delta r_{t-1} + u_{2,t}, & \text{if } z_{t-1} \le -0.62, \\ -0.42 + 1.0 z_{t-1} - 0.14\Delta R_{t-1} - 0.29\Delta r_{t-1} + u_{2,t}, & \text{if } 1.58 \ge z_{t-1} > -0.62, \\ -0.01 - 0.01 z_{t-1} + 0.17\Delta R_{t-1} + 0.25\Delta r_{t-1} + u_{2,t}, & \text{if } z_{t-1} > 1.58. \end{cases}$$

$(1, \hat{\beta}) = (1, -0.983)$, data in three regimes (8%, 72%, 20%).

For the SLSE, the model is estimated as:

$$\Delta R_t = \begin{cases} 0.56 + 0.41 z_{t-1} + .21\Delta R_{t-1} - .25\Delta r_{t-1} + u_{1,t}, & \text{if } z_{t-1} \le -0.84, \\ 0.73 - 0.33 z_{t-1} - 0.096\Delta R_{t-1} - 0.23\Delta r_{t-1} + u_{1,t}, & \text{if } 1.58 \ge z_{t-1} > -0.84, \\ 0.03 - 0.05 z_{t-1} + 0.04\Delta R_{t-1} + 0.08\Delta r_{t-1} + u_{1,t}, & \text{if } z_{t-1} > 1.58, \end{cases}$$

$$\Delta r_t = \begin{cases} 1.82 + 1.65 z_{t-1} + 0.42\Delta R_{t-1} - .22\Delta r_{t-1} + u_{2,t}, & \text{if } z_{t-1} \le -0.84, \\ 0.65 - 0.25 z_{t-1} - 0.21\Delta R_{t-1} - 0.02\Delta r_{t-1} + u_{2,t}, & \text{if } 1.58 \ge z_{t-1} > -0.84, \\ 0.002 - 0.02 z_{t-1} + 0.08\Delta R_{t-1} + 0.21\Delta r_{t-1} + u_{1,t}, & \text{if } z_{t-1} > 1.58. \end{cases}$$

$(1, \hat{\beta}^*) = (1, -0.981)$, data in three regimes (5.4%, 73.15%, 21.45%).

A three-regime TVECM seems to be more reasonable:

1. Situations of R_t relatively high and R_t relatively low are both abnormal.
2. The three-regime models keep the data in the left regime and divide the rest to two more regimes.
3. Both the two models are different in the middle and right regimes.

They both indicate: market is more sensitive to low R_t than high R_t.

Thank You!