Stock, J.H. (1987) "Asymptotic Properties of Least Squares Estimators of Cointegrating Vectors," *Econometrica*, Vol.55, pp.1035 – 1056.

Proposition:

Let $y_{1,t}$ be a scalar, $y_{2,t}$ be a $k \times 1$ vector, and $(y_{1,t}, y'_{2,t})'$ be a $g \times 1$ vector, where $g = k + 1$.

Consider the following model:

$$y_{1,t} = \alpha + \gamma' y_{2,t} + z_t^*$$

$$\Delta y_{2,t} = u_{2,t}$$

$$\begin{pmatrix} z_t^* \\ u_{2,t} \end{pmatrix} = \Psi^*(L)\epsilon_t$$

ϵ_t is a $g \times 1$ i.i.d. vector with $\mathrm{E}(\epsilon_t) = 0$ and $\mathrm{E}(\epsilon_t \epsilon_t') = PP'$.

OLSE is given by:

$$\begin{pmatrix} \hat{\alpha} \\ \hat{\gamma} \end{pmatrix} = \begin{pmatrix} T & \sum y_{2,t}' \\ \sum y_{2,t} & \sum y_{2,t} y_{2,t}' \end{pmatrix}^{-1} \begin{pmatrix} \sum y_{1,t} \\ \sum y_{1,t} y_{2,t} \end{pmatrix}.$$

Define λ_1^*, which is a $g \times 1$ vector, and Λ_2^*, which is a $k \times g$ matrix, as follows:

$$\Psi^*(1)\, P = \begin{pmatrix} \lambda_1^{*\prime} \\ \Lambda_2^* \end{pmatrix}.$$

Then, we have the following results:

$$\begin{pmatrix} T^{1/2}(\hat{\alpha}-\alpha) \\ T(\hat{\gamma}-\gamma) \end{pmatrix} \longrightarrow \begin{pmatrix} 1 & \left(\Lambda_2^* \int W(r)\mathrm{d}r\right)' \\ \Lambda_2^* \int W(r)\mathrm{d}r & \Lambda_2^* \left(\int (W(r))\,(W(r))'\,\mathrm{d}r\right)\Lambda_2^{*\prime} \end{pmatrix}^{-1} \begin{pmatrix} h_1 \\ h_2 \end{pmatrix},$$

where

$$\begin{pmatrix} h_1 \\ h_2 \end{pmatrix} = \begin{pmatrix} \lambda_1^{*\prime} W(1) \\ \Lambda_2^* \left(\int W(r)\,(\mathrm{d}W(r))'\right)\lambda_1^* + \sum_{\tau=0}^{\infty} \mathrm{E}(u_{2,t} z_{t+\tau}^*) \end{pmatrix}.$$

$W(r)$ denotes a g-dimensional standard Brownian motion.

1) OLSE of the cointegrating vector is consistent even though u_t is serially correlated.

2) The consistency of OLSE implies that $T^{-1} \sum \hat{u}_t^2 \longrightarrow \sigma^2$.

3) Because $T^{-1} \sum (y_{1,t} - \bar{y}_1)^2$ goes to infinity, a coefficient of determination, R^2, goes to one.

3.4 Testing Cointegration

3.4.1 Engle-Granger Test

$y_t \sim I(1)$

$y_{1,t} = \alpha + \gamma' y_{2,t} + u_t$

- $u_t \sim I(0) \implies$ Cointegration
- $u_t \sim I(1) \implies$ Spurious Regression

Estimate $y_{1,t} = \alpha + \gamma' y_{2,t} + u_t$ by OLS, and obtain $\hat{u}_t$.

Estimate $\hat{u}_t = \rho \hat{u}_{t-1} + \delta_1 \Delta \hat{u}_{t-1} + \delta_2 \Delta \hat{u}_{t-2} + \cdots + \delta_{p-1} \Delta \hat{u}_{t-p+1} + e_t$ by OLS.

ADF Test:

- $H_0:\ \rho = 1$ (Sprious Regression)
- $H_1:\ \rho < 1$ (Cointegration)

$\Longrightarrow$ **Engle-Granger Test**

For example, see Engle and Granger (1987), Phillips and Ouliaris (1990) and Hansen (1992).

Asymmptotic Distribution of Residual-Based ADF Test for Cointegration

# of Refressors,	(a) Regressors have no drift				(b) Some regressors have drift			
excluding constant	1%	2.5%	5%	10%	1%	2.5%	5%	10%
1	−3.96	−3.64	−3.37	−3.07	−3.96	−3.67	−3.41	−3.13
2	−4.31	−4.02	−3.77	−3.45	−4.36	−4.07	−3.80	−3.52
3	−4.73	−4.37	−4.11	−3.83	−4.65	−4.39	−4.16	−3.84
4	−5.07	−4.71	−4.45	−4.16	−5.04	−4.77	−4.49	−4.20
5	−5.28	−4.98	−4.71	−4.43	−5.36	−5.02	−4.74	−4.46

J.D. Hamilton (1994), *Time Series Analysis*, p.766.

3.4.2 Error Correction Representation

VAR(p) model:

$$y_t = \alpha + \phi_1 y_{t-1} + \phi_2 y_{t-2} + \cdots + \phi_p y_{t-p} + \epsilon_t,$$

where y_t, α and ϵ_t indicate $g \times 1$ vectors for $t = 1, 2, \cdots, T$, and ϕ_s is a $g \times g$ matrix for $s = 1, 2, \cdots, p$.

Rewrite:

$$y_t = \alpha + \rho y_{t-1} + \delta_1 \Delta y_{t-1} + \delta_2 \Delta y_{t-2} + \cdots + +\delta_{p-1} \Delta y_{t-p+1} + \epsilon_t,$$

where

$$\rho = \phi_1 + \phi_2 + \cdots + \phi_p,$$

$$\delta_s = -(\phi_{s+1} + \delta_{s+2} + \cdots + \phi_p), \qquad \text{for } s = 1, 2, \cdots, p-1.$$

Again, rewrite:

$$\Delta y_t = \alpha + \delta_0 y_{t-1} + \delta_1 \Delta y_{t-1} + \delta_2 \Delta y_{t-2} + \cdots + +\delta_{p-1} \Delta y_{t-p+1} + \epsilon_t,$$

where

$$\delta_0 = \rho - I_g = -\phi(1),$$

for $\phi(L) = I_g - \delta_1 L - \delta_2 L^2 - \cdots - \delta_p L^p$.

If y_t has h cointegrating relations, we have the following error correction representation:

$$\Delta y_t = \alpha - BA' y_{t-1} + \delta_1 \Delta y_{t-1} + \delta_2 \Delta y_{t-2} + \cdots + +\delta_{p-1} \Delta y_{t-p+1} + \epsilon_t,$$

where $A' y_{t-1}$ is a stationary $h \times 1$ vector (i.e., h I(0) processes), and B and A are $g \times h$ matrices.

Note that $\quad \phi(1) = BA' \quad$ for $\quad \phi(L) = I_g - \delta_1 L - \delta_2 L^2 - \cdots - \delta_p L^p$.

Each row of A' denotes the cointegrating vector, i.e., A' consists of h cointegrating vectors.

Suppose that $\epsilon_t \sim N(0, \Sigma)$. The log-likelihood function is:

$$\begin{aligned}
&\log l(\alpha, \delta_1, \cdots, \delta_{p-1}, B|A) \\
&\quad = -\frac{Tg}{2}\log(2\pi) - \frac{T}{2}\log|\Sigma| \\
&\quad\quad -\frac{1}{2}\sum_{t=1}^{T}(\Delta y_t - \alpha + BA' y_{t-1} - \delta_1 \Delta y_{t-1} - \cdots - \delta_{p-1}\Delta y_{t-p+1})'\Sigma^{-1} \\
&\quad\quad\quad \times(\Delta y_t - \alpha + BA' y_{t-1} - \delta_1 \Delta y_{t-1} - \cdots - \delta_{p-1}\Delta y_{t-p+1})
\end{aligned}$$

Given A and h, maximize $\log l$ with respect to $\alpha, \delta_1, \cdots, \delta_{p-1}, B$.

Then, given h, how do we estimate A? $\Longrightarrow$ Johansen (1988, 1991)

(*) Canonical Correlatoion (正準相関)

$x' = (x_1, x_2, \cdots, x_n)$ and $y' = (y_1, y_2, \cdots, y_m)$, where $n \leq m$.

$$u = a'x = a_1x_1 + a_2x_2 + \cdots + a_nx_n,$$

$$v = b'y = b_1y_1 + b_2y_2 + \cdots + b_my_m,$$

where $\mathrm{V}(u) = \mathrm{V}(v) = 1$ and $\mathrm{E}(x) = \mathrm{E}(y) = 0$ for simplicity.

Define:

$$\mathrm{V}(x) = \Sigma_{xx}, \qquad \mathrm{E}(xy') = \Sigma_{xy}, \qquad \mathrm{V}(y) = \Sigma_{yy}, \qquad \mathrm{E}(yx') = \Sigma_{yx} = \Sigma'_{xy}.$$

The correlation coefficient between u and v, denoted by ρ, is:

$$\rho = \frac{\mathrm{Cov}(u, v)}{\sqrt{\mathrm{V}(u)}\,\sqrt{\mathrm{V}(v)}} = a'\Sigma_{xy}b,$$

where $\mathrm{V}(u) = a'\Sigma_{xx}a = 1$ and $\mathrm{V}(v) = b'\Sigma_{yy}b = 1$.

Maximize $\rho = a'\Sigma_{xy}b$ subject to $a'\Sigma_{xx}a = 1$ and $b'\Sigma_{yy}b = 1$.

The Lagrangian is:

$$L = a'\Sigma_{xy}b - \frac{1}{2}\lambda(a'\Sigma_{xx}a - 1) - \frac{1}{2}\mu(b'\Sigma_{yy}b - 1).$$

Take a derivative with respect to a and b.

$$\frac{\partial L}{\partial a} = \Sigma_{xy}b - \lambda\Sigma_{xx}a = 0,$$
$$\frac{\partial L}{\partial b} = \Sigma'_{xy}a - \mu\Sigma_{yy}b = 0.$$

Using $a'\Sigma_{xx}a = 1$ and $b'\Sigma_{yy}b = 1$, we obtain:

$$\lambda = \mu = a'\Sigma_{xy}b.$$

From the first equation, we obtain:

$$a = \frac{1}{\lambda}\Sigma_{xx}^{-1}\Sigma_{xy}b,$$

which is substituted into the second equation as follows:

$$\frac{1}{\lambda}\Sigma'_{xy}\Sigma_{xx}^{-1}\Sigma_{xy}b - \lambda\Sigma_{yy}b = 0,$$

i.e.,

$$(\Sigma_{yy}^{-1}\Sigma'_{xy}\Sigma_{xx}^{-1}\Sigma_{xy} - \lambda^2 I_m)b = 0,$$

i.e.,

$$|\Sigma_{yy}^{-1}\Sigma'_{xy}\Sigma_{xx}^{-1}\Sigma_{xy} - \lambda^2 I_m| = 0.$$

The solution of λ^2 is given by the maximum eigen value of $\Sigma_{yy}^{-1}\Sigma'_{xy}\Sigma_{xx}^{-1}\Sigma_{xy}$, and b is the corresponding eigen vector.

Back to the Cointegration:

Estimate the following two regressions:

$$\Delta y_t = b_{1,0} + b_{1,1}\Delta y_{t-1} + b_{1,2}\Delta y_{t-2} + \cdots + b_{1,p-1}\Delta y_{t-p+1} + u_{1,t}$$

$$y_{t-1} = b_{2,0} + b_{2,1}\Delta y_{t-1} + b_{2,2}\Delta y_{t-2} + \cdots + b_{2,p-1}\Delta y_{t-p+1} + u_{2,t}$$

Obtain $\hat{u}_{i,t}$ for $i = 1, 2$ and $t = 1, 2, \cdots, T$, and compute as follow:

$$\hat{\Sigma}_{11} = \frac{1}{T}\sum_{t=1}^{T} \hat{u}_{1,t}\hat{u}'_{1,t}, \qquad \hat{\Sigma}_{22} = \frac{1}{T}\sum_{t=1}^{T} \hat{u}_{2,t}\hat{u}'_{2,t},$$

$$\hat{\Sigma}_{12} = \frac{1}{T}\sum_{t=1}^{T} \hat{u}_{1,t}\hat{u}'_{2,t}, \qquad \hat{\Sigma}_{21} = \hat{\Sigma}'_{12}.$$

From $\hat{\Sigma}_{22}^{-1}\hat{\Sigma}_{21}\hat{\Sigma}_{11}^{-1}\hat{\Sigma}_{12}$, compute h biggest eigenvalues, denoted by $\hat{\lambda}_1$, $\hat{\lambda}_2$, $\cdots$, $\hat{\lambda}_h$, and the corresponding eigen vectors, denoted by $\hat{a}_1$, $\hat{a}_2$, $\cdots$, $\hat{a}_h$, where $\hat{\lambda}_1 > \hat{\lambda}_2 > \cdots > \hat{\lambda}_h$,

The estimate of A, $\hat{A}$, is given by $\hat{A} = (\hat{a}_1, \hat{a}_2, \cdots, \hat{a}_h)$.

How do we obtain h?

3.5 Testing the Number of Cointegrating Vectors

Trace Test (トレース検定):

$H_0:\ \lambda_{h+1} = 0$ and $H_1:\ \lambda_h > 0.$

$$2(\log l_1 - \log l_0) = -T \sum_{i=h+1}^{g} \log(1 - \hat{\lambda}_i) \longrightarrow \mathrm{tr}(Q),$$

where

$$Q = \left(\int_0^1 W(r)\mathrm{d}W(r)'\right)' \left(\int_0^1 W(r)W(r)'\mathrm{d}r\right)^{-1} \left(\int_0^1 W(r)\mathrm{d}W(r)'\right).$$

Trace Test for # of Cointegrating Relations

# of Random	(a) Regressors have no drift				(b) Some regressors have drift			
Walks ($g - h$)	1%	2.5%	5%	10%	1%	2.5%	5%	10%
1	11.576	9.658	8.083	6.691	6.936	5.332	3.962	2.816
2	21.962	19.611	17.844	15.583	19.310	17.299	15.197	13.338
3	37.291	34.062	31.256	28.436	35.397	32.313	29.509	26.791
4	55.551	51.801	48.419	45.248	53.792	50.424	47.181	43.964
5	77.911	73.031	69.977	65.956	76.955	72.140	68.905	65.063

J.D. Hamilton (1994), *Time Series Analysis*, p.767.

Largest Eigenvalue Test (最大固有値検定):

$H_0: \lambda_{h+1} = 0$ and $H_1: \lambda_h > 0.$

$$2(\log l_1 - \log l_0) = -T \log(1 - \hat{\lambda}_{h+1}) \longrightarrow \text{maxmum eigen value of } Q,$$

Maximum Eigenvalue Test for # of Cointegrating Relations

# of Random	(a) Regressors have no drift				(b) Some regressors have drift			
Walks ($g - h$)	1%	2.5%	5%	10%	1%	2.5%	5%	10%
1	11.576	9.658	8.083	6.691	6.936	5.332	3.962	2.816
2	18.782	16.403	14.595	12.783	17.936	15.810	14.036	12.099
3	26.154	23.362	21.279	18.959	25.521	23.002	20.778	18.697
4	32.616	29.599	27.341	24.917	31.943	29.335	27.169	24.712
5	38.858	35.700	33.262	30.818	38.341	35.546	33.178	30.774

J.D. Hamilton (1994), *Time Series Analysis*, p.768.

4 GMM (Generalized Mothod of Moments, 一般化積率法)

1. Method of Moments (積率法):

   Regression Model: $y_t = x_t\beta + \epsilon_t$

   From the assumption, $\mathrm{E}(x_t'\epsilon_t) = 0$.

   The sample mean is given by:

$$\frac{1}{T}\sum_{t=1}^{T} x_t'\epsilon_t = \frac{1}{T}\sum_{t=1}^{T} x_t'(y_t - x_t\beta) = 0.$$

Therefore,

$$\beta_{MM} = \left(\frac{1}{T}\sum_{t=1}^{T} x_t' x_t\right)^{-1}\left(\frac{1}{T}\sum_{t=1}^{T} x_t' y_t\right),$$

which is equivalent to OLS.

2. Generalized Mothod of Moments (GMM, 一般化積率法):

$$\mathrm{E}\left(h(\theta; w_t)\right) = 0$$

θ is a $k \times 1$ parameter vector to be estimated.

w_t is an observed vector $w_t = (y_t, x_t)$.

$h(\theta; w_t)$ is a $r \times 1$ vector function, where $r \geq k$.

Define $g(\theta; W_T)$ as follows:

$$g(\theta; W_T) = \frac{1}{T}\sum_{t=1}^{T} h(\theta; w_t),$$

where $W_T = \{w_T, w_{T-1}, \cdots, w_1\}$.

Compute:

$$\min_{\theta}\Big(g(\theta; W_T)\Big)' S^{-1}\Big(g(\theta; W_T)\Big)$$

The solution of θ, denoted by $\hat{\theta}_T$, corresponds to the GMM estimator, where

S is defined as follows:

$$S = \lim_{T\to\infty} \frac{1}{T} \sum_{t=1}^{T} \sum_{\tau=-\infty}^{\infty} \mathrm{E}\left((h(\theta; w_t))\,(h(\theta; w_{t-\tau}))'\right).$$

In empirical studies, S is replaced by its estimate, i.e., $\hat{S}_T$.

When $h(\theta; w_t)$, $t = 1, 2, \cdots, T$, are not serially correlated, the following $\hat{S}_T$ is consistent, i.e.,

$$\hat{S}_T = \frac{1}{T} \sum_{t=1}^{T} \Big(h(\hat{\theta}_T; w_t)\Big)\Big(h(\hat{\theta}_T; w_t)\Big)' \longrightarrow S.$$

When $h(\theta; w_t)$, $t = 1, 2, \cdots, T$, are serially correlated,

$$\hat{S}_T = \hat{\Gamma}(0) + \sum_{\tau=1}^{q} k(\frac{\tau}{q+1})(\hat{\Gamma}(\tau) + \hat{\Gamma}(\tau)'),$$

where $\hat{\Gamma}(\tau) = \dfrac{1}{T} \displaystyle\sum_{t=\tau+1}^{T} h(\hat{\theta}_T; w_t) h(\hat{\theta}_T; w_{t-s})'$.

$k(x) = 1 - x \implies$ Bartlett kernel (Newwey-west estimator),

$k(x) \implies$ Parzen kernel, and etc.

Then, we obtain:

$$\sqrt{T}(\hat{\theta}_T - \theta) \longrightarrow N\left(0, (DS^{-1}D')^{-1}\right),$$

where

$$D = \frac{\partial g(\theta; W_T)}{\partial \theta'}.$$

Note that D is a $r \times k$ matrix.

Let $\hat{D}_T$ be an estimate of D.

The variance estimator of $\hat{\theta}_T$ is given by:

$$\hat{D}_T = \frac{\partial g(\hat{\theta}_T; W_T)}{\partial \theta'}.$$