# EC435
# บทที่ 6 โคอินทิเกรชัน (Cointegration)

เฉลิมพงษ์ คงเจริญ ©2556

คณะเศรษฐศาสตร์ มหาวิทยาลัยธรรมศาสตร์

November 18, 2019

# spurious regression

กรณีที่ตัวแปรทั้งหมดเป็น $I(1)$ และไม่ได้เป็น cointegrated ตัวทดสอบที่ได้จากแบบจำลองจะไม่เหมาะสม เราจะพบความสัมพันธ์ระหว่างตัวแปรทั้งๆที่จริงๆแล้วตัวแปรทั้งหลายไม่มีความสัมพันธ์กัน (spurious regression)

# spurious regression

กรณีที่ตัวแปรทั้งหมดเป็น $I(1)$ และไม่ได้เป็น cointegrated ตัวทดสอบที่ได้จากแบบจำลองจะไม่เหมาะสม เราจะพบความสัมพันธ์ระหว่างตัวแปรทั้งๆที่จริงๆแล้วตัวแปรทั้งหลายไม่มีความสัมพันธ์กัน (spurious regression)

หากเราพิจารณาอนุกรมเวลาสองกระบวนการที่เป็น $I(1)$ และเป็นอิสระต่อกัน เราสามารถแสดงความสัมพันธ์ของแต่ละตัวแปรได้ดังนี้

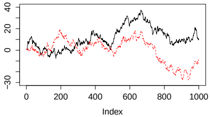
$$y_{1t} = y_{1,t-1} + \varepsilon_{1t} \qquad \varepsilon_{1t} \sim WN(0,1)$$

$$y_{2t} = y_{2,t-1} + \varepsilon_{2t} \qquad \varepsilon_{2t} \sim WN(0,1)$$

โดยที่ $Cov(\varepsilon_{1t}, \varepsilon_{2t}) = 0$ ซึ่งแสดงว่าทั้งสองอนุกรมเวลาเป็นอิสระต่อกัน

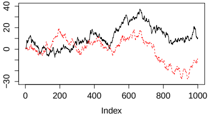
# Spurious regression

Figure: การจำลองกระบวนการ $I(1)$ สองอนุกรมเวลา



# Spurious regression

Figure: การจำลองกระบวนการ $I(1)$ สองอนุกรมเวลา




พิจารณาความสัมพันธ์เชิงถดถอย $y_{1t} = \alpha + \beta y_{2t} + u_t$

```
> summary(lm(y1~y2))
Call:
lm(formula = y1 ~ y2)

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept) 11.43603    0.31329  36.503  < 2e-16 ***
y2           0.16237    0.02885   5.628 2.36e-08 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘’. 0.1 ‘’  1

Residual standard error: 9.905 on 998 degrees of freedom
Multiple R-squared: 0.03077,^^IAdjusted R-squared: 0.02979
F-statistic: 31.68 on 1 and 998 DF,  p-value: 2.363e-08
```

# Spurious regression

หากเราพิจารณาสมการ $\Delta y_{1t} = \beta \Delta y_{2t} + v_t$

```
> summary(lm(diff(y1)~diff(y2)-1))

Call:
lm(formula = diff(y1) ~ diff(y2) - 1)

Coefficients:
         Estimate Std. Error t value Pr(>|t|)
diff(y2) 0.003455   0.030827   0.112    0.911

Residual standard error: 0.9921 on 998 degrees of freedom
Multiple R-squared: 1.258e-05,^^IAdjusted R-squared: -0.0009894
F-statistic: 0.01256 on 1 and 998 DF,  p-value: 0.9108
```

# Spurious regression

หากเราพิจารณาสมการ $\Delta y_{1t} = \beta \Delta y_{2t} + v_t$

```
> summary(lm(diff(y1)~diff(y2)-1))

Call:
lm(formula = diff(y1) ~ diff(y2) - 1)

Coefficients:
         Estimate Std. Error t value Pr(>|t|)
diff(y2) 0.003455   0.030827   0.112    0.911

Residual standard error: 0.9921 on 998 degrees of freedom
Multiple R-squared: 1.258e-05,^^IAdjusted R-squared: -0.0009894
F-statistic: 0.01256 on 1 and 998 DF,  p-value: 0.9108
```

จากตัวอย่างข้างต้นเราจำเป็นต้องระมัดระวังในการพิจารณาความสัมพันธ์ระหว่างอนุกรมเวลาในกรณีที่ตัวแปรที่เราสนใจมีลักษณะเป็น $I(1)$ ซึ่งในกรณีทั่วไปเราจำเป็นต้องปรับให้ตัวแปรดังกล่าวเป็น $I(0)$

# Spurious regression

หากเราพิจารณาสมการ $\Delta y_{1t} = \beta \Delta y_{2t} + v_t$

```
> summary(lm(diff(y1)~diff(y2)-1))

Call:
lm(formula = diff(y1) ~ diff(y2) - 1)

Coefficients:
          Estimate Std. Error t value Pr(>|t|)
diff(y2) 0.003455   0.030827   0.112    0.911

Residual standard error: 0.9921 on 998 degrees of freedom
Multiple R-squared: 1.258e-05,^^IAdjusted R-squared: -0.0009894
F-statistic: 0.01256 on 1 and 998 DF,  p-value: 0.9108
```

จากตัวอย่างข้างต้นเราจำเป็นต้องระมัดระวังในการพิจารณาความสัมพันธ์ระหว่างอนุกรมเวลาในกรณีที่ตัวแปรที่เราสนใจมีลักษณะเป็น $I(1)$ ซึ่งในกรณีทั่วไปเราจำเป็นต้องปรับให้ตัวแปรดังกล่าวเป็น $I(0)$

ในบางกรณีที่เราอาจจะสามารถศึกษาความสัมพันธ์ระหว่างตัวแปรที่เป็น $I(1)$ ก็ได้ซึ่งในกรณีดังกล่าวคือกรณีที่ตัวแปรเป็น cointegration

# บทนำ

หากเราพิจารณาสมการถดถอยของตัวแปรทั้งสอง อาจจะมีเหตุการณ์เกิดขึ้นได้ 4 เหตุการณ์

*กรณีที่ 1: ทั้ง y และ x เป็น stationary* เราสามารถใช้ OLS ปกติในการวิเคราะห์

# บทนำ

หากเราพิจารณาสมการถดถอยของตัวแปรทั้งสอง อาจจะมีเหตุการณ์เกิดขึ้นได้ 4 เหตุการณ์

*กรณีที่ 1: ทั้ง y และ x เป็น stationary* เราสามารถใช้ OLS ปกติในการวิเคราะห์

*กรณีที่ 2: y และ x เป็น non-stationary แต่ integrated ที่อันดับต่างกัน* เราไม่สามารถใช้ OLS ศึกษาความสัมพันธ์ของทั้งสอง

# บทนำ

หากเราพิจารณาสมการถดถอยของตัวแปรทั้งสอง อาจจะมีเหตุการณ์เกิดขึ้นได้ 4 เหตุการณ์

*กรณีที่ 1: ทั้ง y และ x เป็น stationary* เราสามารถใช้ OLS ปกติในการวิเคราะห์

*กรณีที่ 2: y และ x เป็น non-stationary แต่ integrated ที่อันดับต่างกัน* เราไม่สามารถใช้ OLS ศึกษาความสัมพันธ์ของทั้งสอง

*กรณีที่ 3: y และ x เป็น non-stationary และ integrated ที่อันดับเดียวกัน แต่ $\epsilon$ เป็น random walk* เราไม่สามารถใช้ OLS วิเคราะห์ แต่เราอาจจะพิจารณาสมการที่เรา difference ทั้งสองข้าง

# บทนำ

หากเราพิจารณาสมการถดถอยของตัวแปรทั้งสอง อาจจะมีเหตุการณ์เกิดขึ้นได้ 4 เหตุการณ์

*กรณีที่ 1: ทั้ง $y$ และ $x$ เป็น stationary* เราสามารถใช้ OLS ปกติในการวิเคราะห์

*กรณีที่ 2: $y$ และ $x$ เป็น non-stationary แต่ integrated ที่อันดับต่างกัน* เราไม่สามารถใช้ OLS ศึกษาความสัมพันธ์ของทั้งสอง

*กรณีที่ 3: $y$ และ $x$ เป็น non-stationary และ integrated ที่อันดับเดียวกัน แต่ $\epsilon$ เป็น random walk* เราไม่สามารถใช้ OLS วิเคราะห์ แต่เราอาจจะพิจารณาสมการที่เรา difference ทั้งสองข้าง

*กรณีที่ 4: $y$ และ $x$ เป็น non-stationary โดยมีอันดับ integrated ที่เหมือนกัน และ error term เป็น stationary.* เราเรียกว่า $y$ และ $x$ เป็น **cointegrated.**

## Cointegration

กำหนดให้ $\boldsymbol{Y}_t = (y_{1t}, y_{2t}, ..., y_{nt})'$ แทนเวกเตอร์ $(n \times 1)$ ของอนุกรมเวลาที่เป็น $I(1)$ แล้ว $\boldsymbol{Y}_t$ จะเป็น cointegrated ถ้ามีเวกเตอร์ $\boldsymbol{\beta} = (\beta_1, ..., \beta_n)'$ ที่ทำให้

$$\boldsymbol{\beta}' \boldsymbol{Y}_t = \beta_1 y_{1t} + ... + \beta_n y_{nt} \sim I(0) \qquad (6.1)$$

ซึ่งหมายความว่าผลรวมเชิงเส้น (linear combination) ของตัวแปรที่เป็น $I(1)$ มีลักษณะเป็น $I(0)$

ผลรวมเชิงเส้นมักจะได้มาจากทฤษฎีทางเศรษฐศาสตร์ และเรียกสมการดังกล่าวว่าเป็นความสัมพันธ์ใน**ดุลยภาพระยะยาว (long-run equilibrium)**

# Cointegration

เวกเตอร์ $\beta$ ไม่ได้มีลักษณะเป็นรูปแบบอย่างเดียว (unique) เนื่องจากเวกเตอร์ที่เป็นจำนวนเท่าของ $c\beta'$

# Cointegration

เวกเตอร์ $\boldsymbol{\beta}$ ไม่ได้มีลักษณะเป็นรูปแบบอย่างเดียว (unique) เนื่องจากเวกเตอร์ที่เป็นจำนวนเท่าของ $c\boldsymbol{\beta}'$

เรามักจะสนใจในรูปแบบที่ normalized ซึ่งเขียนเป็นความสัมพันธ์ cointegration ได้เป็น

$$\boldsymbol{\beta}'\boldsymbol{Y}_t = y_{1t} - \beta_2 y_{2t} - \ldots - \beta_n y_{nt} \sim I(0)$$

หรือเป็นสมการ

$$y_{1t} = \beta_2 y_{2t} + \ldots + \beta_n y_{nt} + u_t \qquad (6.2)$$

โดยที่ $u_t \sim I(0)$ และเราเรียกพจน์ $u_t$ ว่า disequilibrium error หรือ cointegrating residuals

# Cointegration

เวกเตอร์ $\boldsymbol{\beta}$ ไม่ได้มีลักษณะเป็นรูปแบบอย่างเดียว (unique) เนื่องจากเวกเตอร์ที่เป็นจำนวนเท่าของ $c\boldsymbol{\beta}'$

เรามักจะสนใจในรูปแบบที่ normalized ซึ่งเขียนเป็นความสัมพันธ์ cointegration ได้เป็น

$$\boldsymbol{\beta}'\boldsymbol{Y}_t = y_{1t} - \beta_2 y_{2t} - \ldots - \beta_n y_{nt} \sim I(0)$$

หรือเป็นสมการ

$$y_{1t} = \beta_2 y_{2t} + \ldots + \beta_n y_{nt} + u_t \qquad (6.2)$$

โดยที่ $u_t \sim I(0)$ และเราเรียกพจน์ $u_t$ ว่า disequilibrium error หรือ cointegrating residuals

ในดุลยภาพระยะยาวค่า disequilibrium error จะเท่ากับศูนย์ และความสัมพันธ์ระยะยาวจะเท่ากับ $y_{1t} = \beta_2 y_{2t} + \ldots + \beta_n y_{nt}$

# Error Correction Models

หากเกิดภาวะที่ในระยะสั้นตัวแปรขยับออกจากดุลยภาพในระยะยาว จะมีกลไกในการปรับตัวเข้าสู่ดุลยภาพระยะยาวอย่างไร ซึ่งสามารถอธิบายผ่านแบบจำลอง Error Correction Model (ECM)

# Error Correction Models

หากเกิดภาวะที่ในระยะสั้นตัวแปรขยับออกจากดุลยภาพในระยะยาว จะมีกลไกในการปรับตัวเข้าสู่ดุลยภาพระยะยาวอย่างไร ซึ่งสามารถอธิบายผ่านแบบจำลอง Error Correction Model (ECM)

หากเราพิจารณา cointegration ที่มีตัวแปรสองตัวแปรที่มีลักษณะ $I(1)$ คือ $y_{1t}$ และ $y_{2t}$ โดยมีความสัมพันธ์ cointegration เขียนได้เป็น $y_{1t} - \beta_2 y_{2t} \sim I(0)$ Engle and Granger (1987) แสดงให้เห็นว่ามีแบบจำลอง error correction model (ECM) ที่สามารถแสดงได้ในรูป

$$\Delta y_{1t} = c_1 + \alpha_1(y_{1,t-1} - \beta_2 y_{2,t-1}) + \sum_{j=1}^{p} \psi_{11}^{j} \Delta y_{1,t-j} + \sum_{j=1}^{p} \psi_{12}^{j} \Delta y_{2,t-j} + \varepsilon_{1t}$$

$$\Delta y_{2t} = c_2 + \alpha_2(y_{1,t-1} - \beta_2 y_{2,t-1}) + \sum_{j=1}^{p} \psi_{21}^{j} \Delta y_{1,t-j} + \sum_{j=1}^{p} \psi_{22}^{j} \Delta y_{2,t-j} + \varepsilon_{2t}$$

# Error Correction Models

หากเกิดภาวะที่ในระยะสั้นตัวแปรขยับออกจากดุลยภาพในระยะยาว จะมีกลไกในการปรับตัวเข้าสู่ดุลยภาพระยะยาวอย่างไร ซึ่งสามารถอธิบายผ่านแบบจำลอง Error Correction Model (ECM)

หากเราพิจารณา cointegration ที่มีตัวแปรสองตัวแปรที่มีลักษณะ $I(1)$ คือ $y_{1t}$ และ $y_{2t}$ โดยมีความสัมพันธ์ cointegration เขียนได้เป็น $y_{1t} - \beta_2 y_{2t} \sim I(0)$ Engle and Granger (1987) แสดงให้เห็นว่ามีแบบจำลอง error correction model (ECM) ที่สามารถแสดงได้ในรูป

$$\Delta y_{1t} = c_1 + \alpha_1 (y_{1,t-1} - \beta_2 y_{2,t-1}) + \sum_{j=1}^{p} \psi_{11}^{j} \Delta y_{1,t-j} + \sum_{j=1}^{p} \psi_{12}^{j} \Delta y_{2,t-j} + \varepsilon_{1t}$$

$$\Delta y_{2t} = c_2 + \alpha_2 (y_{1,t-1} - \beta_2 y_{2,t-1}) + \sum_{j=1}^{p} \psi_{21}^{j} \Delta y_{1,t-j} + \sum_{j=1}^{p} \psi_{22}^{j} \Delta y_{2,t-j} + \varepsilon_{2t}$$

สมการข้างต้นแสดงพฤติกรรมเชิงพลวัตรของ $y_{1t}$ และ $y_{2t}$ โดยแสดงการปรับตัวระยะสั้นเมื่อตัวแปรตัวใดตัวหนึ่งเคลื่อนตัวออกไปจากดุลยภาพ ซึ่งแบบจำลอง ECM ส่งผลให้เกิดประโยชน์ในการวิเคราะห์ข้อมูลการเงิน

# ตัวอย่างที่ 6.2

ตัวอย่าง ECM สำหรับราคาหุ้นและเงินปันผล

กำหนดให้ล็อกของราคาหุ้น ($s_t$) และล็อกของเงินปันผล ($d_t$) เป็น $I(1)$ ถ้าล็อกของสัดส่วนเงินปันผลต่อราคา ($\log(D_t/S_t)$) เป็น $I(0)$ ดังนั้น $d_t - s_t \sim I(0)$ หรือเขียนเป็นสมการความสัมพันธ์ระยะยาวได้เป็น

$$d_t = s_t + \mu + u_t \qquad (6.4)$$

โดยที่ $u_t \sim I(0)$ และสามารถเขียน ECM ได้ดังนี้

$$\Delta s_t = c_s + \alpha_s(d_{t-1} - s_{t-1} - \mu) + \varepsilon_{st} \qquad (6.5)$$

$$\Delta d_t = c_d + \alpha_d(d_{t-1} - s_{t-1} - \mu) + \varepsilon_{dt} \qquad (6.6)$$

# ตัวอย่างที่ 6.2

สมมุติให้ $\alpha_s = 0.5$ และ $\alpha_d = 0$ จะได้

$$\Delta s_t = c_s + 0.5(d_{t-1} - s_{t-1} - \mu) + \varepsilon_{st}$$
$$\Delta d_t = c_d + \varepsilon_{dt}$$

ดังนั้นมีเพียงราคาหุ้นที่ตอบสนองต่อ disequilibrium error โดยที่ค่าเฉลี่ยของการเปลี่ยนแปลงของตัวแปรทั้งสองจะขึ้นอยู่กับข้อมูลในอดีต

$$E(\Delta s_t | F_{t-1}) = c_s + 0.5(d_{t-1} - s_{t-1} - \mu)$$
$$E(\Delta d_t | F_{t-1}) = c_d$$

# multiple cointegrating relationship

เวกเตอร์ของอนุกรมเวลา $\boldsymbol{Y}_t = (y_{1t}, ..., y_{nt})'$ ซึ่งเป็นอนุกรมเวลาที่ cointegrated กันอาจจะมีความสัมพันธ์ระยะยาวได้ $r$ รูปแบบ ซึ่ง $o < r < n$ เช่นกรณีที่ $\boldsymbol{Y}_t = (y_{1t}, y_{2t}, y_{3t})'$ อาจจะมีความสัมพันธ์ในรูป $\beta_{11}y_{1t} + \beta_{12}y_{2t} + \beta_{13}y_{3t} \sim I(0)$ และ $\beta_{21}y_{1t} + \beta_{22}y_{2t} + \beta_{23}y_{3t} \sim I(0)$ และเราเรียกเวกเตอร์ $(\beta_{11} \quad \beta_{12} \quad \beta_{13})'$ และ $(\beta_{21} \quad \beta_{22} \quad \beta_{23})'$ ว่า cointegrating vector

# การทดสอบ cointegration

หากเรากำหนดให้ $\boldsymbol{Y}_t$ เป็นเวกเตอร์ $n \times 1$ ของอนุกรมเวลา $I(1)$ ซึ่งมี $r$ $(0 < r < n)$ cointegrating vectors ซึ่งแสดงเป็นสัญลักษณ์ได้ดังนี้

$$\begin{pmatrix} \beta_1' \boldsymbol{Y}_t \\ \vdots \\ \beta_r' \boldsymbol{Y}_t \end{pmatrix} = \begin{pmatrix} u_{1t} \\ \vdots \\ u_{rt} \end{pmatrix} \sim I(0) \qquad (6.7)$$

การทดสอบ cointegration อาจจะแบ่งได้เป็นสองกรณีคือ

1. การทดสอบว่ามีความสัมพันธ์ cointegrating อย่างน้อยหนึ่งรูปแบบหรือไม่ (Engle and Granger(1986))
2. การทดสอบว่ามี cointegrating vector $0 \leq r < n$ หรือไม่ (Johansen (1988))

# การทดสอบ cointegration

การทดสอบในประเด็นแรกได้ถูกนำเสนอโดย Engle and Granger (1986) ซึ่งเสนอการทดสอบสองขั้นตอน (two-step cointegration test)ซึ่งประกอบด้วย

1. สร้าง cointegrating residuals โดยที่ $u_t = y_{1t} - \beta_2 y_{2t} - \ldots - \beta_n y_{nt}$
2. ทดสอบ unit root ของ $u_t$ เพื่อแสดงว่า $u_t$ เป็น $I(0)$ หรือไม่ โดยสมมุติฐานหลักของการทดสอบนี้คือไม่มี cointegration ซึ่งสมมุติฐานดังกล่าวสอดคล้องกับ $u_t$ เป็น $I(1)$

การทดสอบ cointegration ตามวิธีการของ Engle and Granger สามารถแบ่งออกได้เป็นสองกรณีคือ

*กรณีที่หนึ่ง* เราทราบความสัมพันธ์ของอนุกรมเวลาที่เราสนใจ

*กรณีที่สอง* เราจำเป็นต้องประมาณค่าสำหรับ cointegrating vector และสร้าง cointegrating vector จาก $\hat{u}_t = \hat{\boldsymbol{\beta}}' \boldsymbol{Y}_t$

# การทดสอบกรณีทราบ cointegration vector

กำหนดให้ $\boldsymbol{Y}_t$ เป็นเวกเตอร์ $n \times 1$ ของ $I(1)$ และ cointegrated กันโดยที่มีความสัมพันธ์สามารถแสดงได้ด้วย cointegrating vector $\boldsymbol{\beta}$ ซึ่ง $u_t = \boldsymbol{\beta}' \boldsymbol{Y}_t$ เราต้องการทดสอบสมมุติฐานหลัก $H_0 : u_t = \boldsymbol{\beta}' \boldsymbol{Y}_t \sim I(1)$ หรือไม่มี cointegration กับสมมุติฐานรอง
$H_1 : u_t = \boldsymbol{\beta}' \boldsymbol{Y}_t \sim I(0)$ หรือมี cointegration

เราสามารถทดสอบสมมุติฐานข้างต้นได้ด้วยการทดสอบยูนิทรูทเช่น ADF หรือ PP กับตัวแปร $u_t$ ซึ่งเราจะสรุปว่า $\boldsymbol{Y}_t$ เป็น cointegrated ถ้าเราปฏิเสธสมมุติฐานหลัก

# ตัวอย่างที่ 6.3

ในตัวอย่างนี้เราจะพิจารณาความสัมพันธ์ระหว่างราคาปัจจุบัน (spot price: $S_t$) และราคาฟิวเจอร์ (future price: $F_t$) ของอัตราแลกเปลี่ยนโดยที่ตัวแปรทั้งสองมีความสัมพันธ์ตามแบบจำลอง cost of carry model ซึ่งล็อกของราคาปัจจุบัน ($s_t$) จะเท่ากับล็อกของราคาฟิวเจอร์ ($f_t$) บวกกับค่าคงที่(ซึ่งสะท้อนต้นทุนในการถือสินทรัพย์)

$$f_t = s_t + c$$

ดังนั้นในกรณีนี้เราสามารถเขียน $u_t = f_t - s_t$ หรือ $\beta = (1 \quad -1)'$

```
> set50_m <- read.csv("C:/teaching/ec435/r/set50_m.csv")
> lfutures<-log(set50_m$futures)
> lspot<-log(set50_m$spot)
> u<-lfutures-lspot
> adfTest(u, lags=2, type=c("nc"))
Title:
 Augmented Dickey-Fuller Test
Test Results:
  PARAMETER:
    Lag Order: 2
  STATISTIC:
    Dickey-Fuller: -7.2144
  P VALUE:
    0.01
```

# การทดสอบกรณีที่เราไม่ทราบค่า cointegrating vector

ในกรณีนี้เราจำเป็นต้องประมาณค่า cointegrating vector ขึ้นมา

ขั้นตอนแรกของการทดสอบเราจำเป็นต้องประมาณค่า $\beta$ โดยที่เราอาจจะเขียนในรูปบรรทัดฐาน (normalized) ซึ่งเขียนเป็นสมการได้

$$y_{1t} = c + \beta_2 y_{2t} + \ldots + \beta_n y_{nt} + u_t \qquad (6.8)$$

การทดสอบ cointegration จะเป็นการทดสอบ unit root ของ $\hat{u}_t = y_{1t} - \hat{c} - \hat{\beta}_2 y_{2t} - \ldots - \hat{\beta}_n y_{nt}$ โดยที่ $\hat{c}, \hat{\beta}_2, \ldots, \hat{\beta}_n$ เป็นตัวประมาณค่า OLS ของสัมประสิทธิ์

Phillips and Ouliaris (1990) ได้แสดงให้เห็นว่าตัวทดสอบ ADF และ PP ของ cointegrating residual ในกรณีที่เราต้องประมาณค่าสัมประสิทธิ์ของ cointegrating vector ไม่ได้มีการแจกแจงแบบ ADF distribution แบบปกติ

# ตัวอย่างที่ 6.4

ต่อเนื่องจากตัวอย่างที่ผ่านมา ในกรณีนี้สมมุติว่าเราไม่ทราบค่าของ cointegrating vector ดังนั้นสมมุติให้ $\boldsymbol{\beta} = (1 \quad -\beta_2)'$ ดังนั้นเราสามารถพิจารณาสมการถดถอย $s_t = c + \beta_2 f_t + u_t$ ซึ่งหลังจากที่เราประมาณค่าด้วย OLS เราสามารถที่จะหาค่า $\hat{u}_t = s_t - \hat{c} - \hat{\beta}_2 f_t$ ตามขั้นตอนในคำสั่งข้างล่าง

```
> m1<-lm(lfutures~lspot)
> summary(m1)
Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept) -0.044097   0.023113  -1.908   0.0587 .
lspot        1.006206   0.003535 284.610   <2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘’. 0.1 ‘’  1

Residual standard error: 0.01349 on 124 degrees of freedom
Multiple R-squared:  0.9985,^^IAdjusted R-squared:  0.9985
F-statistic: 8.1e+04 on 1 and 124 DF,  p-value: < 2.2e-16

> uhat<-m1$residuals
```

# ตัวอย่างที่ 6.4

```
> adfTest(uhat,lags=2, type=c("nc"))

Title:
 Augmented Dickey-Fuller Test

Test Results:
  PARAMETER:
    Lag Order: 2
  STATISTIC:
    Dickey-Fuller: -8.4232
  P VALUE:
    0.01
```

# การประมาณค่า ECM ด้วย OLS

สมมุติว่าเราสนใจตัวแปร $(y_{1t} \quad y_{2t})'$ ซึ่งมีความสัมพันธ์ cointegration แสดงได้ด้วยสมการ $y_{1t} = \beta_2 y_{2t} + u_t$ (หรือ $y_{1t} - \beta_2 y_{2t} = u_t \sim I(0)$) และเรามีตัวประมาณค่า $\hat{\beta}_2$ ที่มีคุณลักษณะเป็น consistent และเราต้องประมาณค่า ECM ในรูปแบบ

$$\Delta y_{1t} = c_1 + \alpha_1 (y_{1,t-1} - \beta_2 y_{2,t-1}) + \sum_{i=1}^{p} \psi_{11}^{j} \Delta y_{1,t-j} + \sum_{i=1}^{p} \psi_{12}^{j} \Delta y_{2,t-j} + \varepsilon_{1t} \quad (6.9)$$

$$\Delta y_{2t} = c_2 + \alpha_2 (y_{1,t-1} - \beta_2 y_{2,t-1}) + \sum_{i=1}^{p} \psi_{21}^{j} \Delta y_{1,t-j} + \sum_{i=1}^{p} \psi_{22}^{j} \Delta y_{2,t-j} + \varepsilon_{2t} \quad (6.10)$$

เราสามารถแทนค่า $\hat{\beta}_2$ ลงในสมการเหมือนเป็นพารามิเตอร์ที่เราทราบค่า

เราสามารถพิจารณา disequilibrium error $(y_{1,t-1} - \beta_2 y_{2,t-1})$ เหมือนเป็นตัวแปรที่ทราบค่า เนื่องจากตัวแปรทุกตัวในสมการ (6.9) และ (6.10) เป็น $I(0)$ เราสามารถประมาณค่าสมการดังกล่าวด้วย OLS

# ตัวอย่างที่ 6.5

เราต้องการประมาณค่าสมการ ECM สำหรับล็อกของราคาปัจจุบันและล็อกของราคาฟิวเจอร์ซึ่งเราจำเป็นต้องหาอันดับของค่าล่าของผลต่างที่เหมาะสม ซึ่งอาจจะใช้ AIC, BIC ในการเลือก แต่ในที่นี้เราสมมุติให้แบบจำลองของเป็นดังนี้

$$\Delta s_t = c_1 + \alpha_1 \hat{u}_{t-1} + \psi_{11} \Delta s_{t-1} + \psi_{12} \Delta f_{t-1} + \varepsilon_{ft} \qquad (6.11)$$

$$\Delta f_t = c_2 + \alpha_2 \hat{u}_{t-1}) + \psi_{21} \Delta s_{t-1} + \psi_{22} \Delta f_{t-1} + \varepsilon_{st} \qquad (6.12)$$

```
> ecm1<-lm(dlfutures[2:125]~uhat[1:124]+dlfutures[1:124]+dlspot[1:124])
> summary(ecm1)
Coefficients:
                  Estimate Std. Error t value Pr(>|t|)
(Intercept)       0.004115   0.006281   0.655   0.5136
uhat[1:124]      -1.152182   0.598961  -1.924   0.0568 .
dlfutures[1:124]  0.071715   0.478109   0.150   0.8810
dlspot[1:124]     0.137282   0.498649   0.275   0.7836
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘’. 0.1 ‘’  1

Residual standard error: 0.06973 on 120 degrees of freedom
Multiple R-squared:  0.06491,^^IAdjusted R-squared:  0.04154
F-statistic: 2.777 on 3 and 120 DF,  p-value: 0.04424
```

# VAR models และ cointegration

Granger representation theorem เชื่อมโยง cointegration กับ error correction models

Johansen สร้างแบบจำลอง cointegration และ error correction models โดยใช้โครงร่าง vector autoregression

พิจารณาตัวแปรในระดับ level ในรูป $VAR(p)$ โดยมีตัวแปรอยู่ในรูปเวกเตอร์ $(n \times 1)$ เขียนเป็นสัญลักษณ์ $\boldsymbol{Y}_t$

$$\boldsymbol{Y}_t = \boldsymbol{\Pi}_1 \boldsymbol{Y}_{t-1} + ... + \boldsymbol{\Pi}_p \boldsymbol{Y}_{t-p} + \boldsymbol{\epsilon}_t, t = 1, ..., T$$

ถ้า VAR(p) มีลักษณะเป็น unit root แสดงว่าตัวแปรบางตัวหรือทั้งหมดใน $\boldsymbol{Y}_t$ มีลักษณะเป็น $I(1)$ และตัวแปรเหล่านั้นอาจจะ cointegrate กันได้

# Cointegrated VAR

ความสัมพันธ์ลักษณะ cointegrating สามารถพิจารณาได้อย่างชัดเจนขึ้นหากเราแปลง level VAR เป็น **vector error correction model (VECM)**

$$\Delta \boldsymbol{Y}_t = \boldsymbol{\Pi}\boldsymbol{Y}_{t-1} + \boldsymbol{\Gamma}_1\Delta\boldsymbol{Y}_{t-1} + \ldots + \boldsymbol{\Gamma}_{p-1}\Delta\boldsymbol{Y}_{t-p+1} + \boldsymbol{\epsilon}_t$$

โดยที่เมตริกซ์ $\boldsymbol{\Pi}$ ถูกเรียกว่า *long run impact matrix* และ $\boldsymbol{\Gamma}_k$ ถูกเรียกว่า *short run impact matrices*

ในแบบจำลอง VECM เราพบว่า $\Delta\boldsymbol{Y}_t$ และ Lag ของ $\Delta\boldsymbol{Y}_t$ เป็น $I(0)$

จะเห็นได้ว่า เหลือเพียงพจน์ $\boldsymbol{\Pi}\boldsymbol{Y}_{t-1}$ ที่มีโอกาสเป็น $I(1)$ และหากเราพิจารณา $\Delta\boldsymbol{Y}_t$ เราพบว่า $\Delta\boldsymbol{Y}_t$ จะเป็น $I(0)$ ถ้า $\boldsymbol{\Pi}\boldsymbol{Y}_{t-1}$ เป็น $I(0)$

ดังนั้น $\boldsymbol{\Pi}\boldsymbol{Y}_{t-1}$ จะเป็นเมตริกซ์ที่ระบุ cointegrating relations หากตัวแปร cointegrate กัน

# Cointegrated VAR

ถ้า VAR(p) process มีลักษณะเป็น unit roots เราพบว่า $\mathbf{\Pi}$ จะเป็น singular matrix และถ้า $\mathbf{\Pi}$ เป็น singular แล้วมันจะมี reduced rank ($rank(\mathbf{\Pi}) = r < n$ ซึ่งค่าดังกล่าวเป็นไปได้สองกรณีคือ

1. $rank(\mathbf{\Pi}) = 0$ แสดงว่า $\mathbf{\Pi}$ และ $\boldsymbol{Y}_t$ เป็น $I(1)$ และตัวแปรมิได้ cointegrate กัน และแบบจำลอง VECM จะลดลงเหลือแค่ VAR(p-1) ของตัวแปรในรูป first differences
2. $0 < rank(\mathbf{\Pi}) = r < n$ กรณีนี้แสดงว่า $\boldsymbol{Y}_t$ เป็น $I(1)$ โดยที่มี $r$ linear independent cointegrating vectors

# Johansen’s Methodology for Modeling Cointegration

ขั้นตอนในการสร้างแบบจำลอง Cointegration ของ Johansen คือ

- ระบุ Order และประมาณค่า VAR(p) สำหรับ $\boldsymbol{Y}_t$
- สร้าง likelihood ratio tests สำหรับการทดสอบ rank of $\boldsymbol{\Pi}$ เพื่อระบุจำนวน cointegrating vectors
- หากจำเป็นเราอาจจะสมมุติ normalization และระบุเงื่อนไขของ cointegrating vectors
- จาก normalized cointegrating vectors ที่ได้ระบุไว้ เราสามารถประมาณค่า cointegrated VECM ด้วย MLE

## Specification of Deterministic Terms

Johansen (1995) ใช้ specification ของ VECM ในรูปแบบต่อไปนี้

$$\boldsymbol{Y}_t = \boldsymbol{\Phi} \boldsymbol{D}_t + \boldsymbol{\Pi}_1 \boldsymbol{Y}_{t-1} + \ldots + \boldsymbol{\Pi}_p \boldsymbol{Y}_{t-p} + \boldsymbol{\epsilon}_t, t = 1, \ldots, T$$

โดยที่ $\boldsymbol{\Phi} \boldsymbol{D}_t = \boldsymbol{\mu}_t = \boldsymbol{\mu}_0 + \boldsymbol{\mu}_1 t$ เป็น deterministic terms โดยที่พฤติกรรมของ deterministic term ของ $\boldsymbol{Y}_t$ สามารถแยกออกได้เป็น 5 กรณี:

1. Model $H_2(r)$ : $\boldsymbol{\mu}_0$ (no constant)
2. Model $H_1^*(r)$ : $\boldsymbol{\mu}_t = \boldsymbol{\mu}_0 = \boldsymbol{\alpha} \boldsymbol{\rho}_0$ (restricted constant)
3. Model $H_1(r)$ : $\boldsymbol{\mu}_t = \boldsymbol{\mu}_0$ (unrestricted constant)
4. Model $H^*(r) = \boldsymbol{\mu}_t = \boldsymbol{\mu}_0 + \boldsymbol{\alpha} \boldsymbol{\rho}_1 t$ (restricted trend)
5. Model $H(r) = \boldsymbol{\mu}_t = \boldsymbol{\mu}_0 + \boldsymbol{\mu}_1 t$ (unrestricted constant and trend)

# Likelihood Ratio Tests for the Number of Cointegrating Vectors

หากเราใช้ $H(r)$ เป็นสัญลักษณ์แทน unrestricted cointegrated VECM แล้ว แบบจำลอง $I(1)$ ของ $H(r)$ สามารถเขียนเป็นเงื่อนไขของ rank ของ $\mathbf{\Pi}$ มีค่าน้อยกว่าหรือเท่ากับ $r$ เราสามารถสร้าง nested set ของแบบจำลอง

$$H(0) \subset \ldots \subset H(r) \subset \ldots \subset H(n)$$

โดยที่ $H(0)$ แสดงถึง non-cointegrated VAR model โดยที่ $\mathbf{\Pi} = \mathbf{0}$ และ $H(n)$ แสดงถึง unrestricted stationary VAR(p)

เนื่องจาก rank ของ $\mathbf{\Pi}$ เป็นตัวระบุจำนวน cointegrating relationships ในตัวแปร $\boldsymbol{Y}_t$ ดังนั้น Johansen ได้เสนอตัวทดสอบ LR statistics สำหรับจำนวน cointegrating relationships ซึ่งเป็น LR statistics สำหรับการระบุ rank ของ $\mathbf{\Pi}$

# Johansen's Trace Statistics

การทดสอบของ Johensen มีพื้นฐานจากการตัวประมาณค่า eigenvalues $\hat{\lambda}_1 > \hat{\lambda}_2 > \ldots > \hat{\lambda}_n$ ของเมตริกซ์ $\mathbf{\Pi}$ ( rank ของเมตริกซ์ มีค่าเท่ากับจำนวนของ eigenvalue ที่มีค่าไม่เท่ากับศูนย์)

**Johansen's Trace Statistic**

Johansen's LR statistic ทดสอบสมมุติฐานในลักษณะเป็น nested hypotheses

$$H_0(r_0) : r = r_0 \quad \text{vs} \quad H_1(r_0) : r > r_0$$

โดยที่ LR statistic ถูกเรียกว่า *trace statistic* และคำนวณได้ดังนี้

$$LR_{trace}(r_0) = -T \sum_{i=r_0+1}^{n} \ln(1 - \hat{\lambda}_i)$$

ถ้า $rank(\mathbf{\Pi}) = r_0$ แล้ว $\hat{\lambda}_{r_0+1}, \ldots, \hat{\lambda}_n$ จะมีค่าเข้าใกล้ศูนย์ และ $LR_{trace}(r_0)$ ควรจะมีค่าน้อย

การแจกแจง(เมื่อตัวอย่างใหญ่ asymptotic) ของ $LR_{trace}(r_0)$ เมื่อ $H_0$ เป็นจริงจะเป็น multivariate distribution ของ Dickey-Fuller unit-root distribution

# Sequential Procedure

Johansen เสนอการทดสอบในลักษณะที่เป็นอันดับ (sequential testing procedure) โดยที่กระบวนการดังกล่าวจะระบุจำนวนของ cointegrating vectors ได้อย่าง consistently

- ขั้นตอนแรกเราทดสอบ $H_0(r=0)$ กับ $H_1(r>0)$ ถ้าเราไม่สามารถปฏิเสธสมมุติฐานหลัก แสดงว่าไม่มี cointegrating vectors
  ถ้าเราปฏิเสธ $H_0(r=0)$ เราสามารถสรุปได้ว่า มี cointegrating vector อย่างน้อยหนึ่งค่า และทดสอบในขั้นต่อไป
- ทดสอบ $H_0(r=1)$ กับ $H_1(r>1)$ ถ้าเราไม่สามารถปฏิเสธสมมุติฐานหลัก เราสรุปว่ามี cointegrating vector เท่ากับหนึ่ง
  แต่ถ้า เราปฏิเสธสมมุติฐานหลัก แสดงว่า มี cointegrating vectors อย่างน้อยสองค่า
- กระบวนการที่เป็นอันดับดังกล่าวจะดำเนินไปเรื่อยๆจนกระทั่งสมมุติฐานหลักไม่สามารถปฏิเสธ สรุปว่ามีความสัมพันธ์เท่ากับ $r_0$

# Johansen's Maximum Eigenvalue Statistic

Johansen ยังได้เสนอการทดสอบ LR statistic สำหรับสมมุติฐานต่อไปนี้

$$H_0(r_0) : r = r_0 \quad \text{vs} \quad H_1(r_0) : r = r_0 + 1$$

โดยเราเรียก LR statistic ว่า *maximum eigenvalue statistic* และคำนวณได้ดังนี้

$$LR_{max}(r_0) = -T \ln(1 - \hat{\lambda}_{r_0+1})$$

# ตัวอย่างที่ 6.6

เราจะพิจารณาข้อมูลในตัวอย่างที่ 6.5 ด้วยแบบจำลอง VECM

```
> fsprice<-cbind(lfutures,lspot)
> library(vars)
> var.mod<-VAR(fsprice, lag.max=4, ic=c("SC"))
> var.mod

VAR Estimation Results:
=======================

Estimated coefficients for equation lfutures:
============================================
Call:
lfutures = lfutures.l1 + lspot.l1 + const

lfutures.l1    lspot.l1       const
 0.07057597  0.91654897  0.08517149


Estimated coefficients for equation lspot:
=========================================
Call:
lspot = lfutures.l1 + lspot.l1 + const

lfutures.l1    lspot.l1       const
 -0.1508610   1.1346435   0.1098629
```

# ตัวอย่างที่ 6.6

แบบจำลองที่เหมาะสมคือ $VAR(1)$ package ที่ใช้ในการทดสอบ Johansen และประมาณค่า VECM คือ urca โดยที่คำสั่งที่ใช้ในการทดสอบ Johansen คือ *ca.jo* โดยที่เราต้องระบุข้อมูล (*fsprice*) วิธีในการทดสอบ (trace) รูปแบบของสมการ (*const*) และจำนวน lag ในแบบจำลอง *VAR* อย่างไรก็ตามใน package นี้กำหนด lag ขั้นต่ำเท่ากับ 2 (VECM(1))

```
> library(urca)
> fsprice.rc<-ca.jo(fsprice, type=c("trace"), ecdet=c("const"), K=2)
> summary(fsprice.rc)

######################
# Johansen-Procedure #
######################

Test type: trace statistic , without linear trend and constant in cointegration

Eigenvalues (lambda):
[1]  3.779375e-01  1.865708e-02 -2.281477e-17

Values of teststatistic and critical values of test:

          test 10pct  5pct  1pct
r <= 1 |  2.34  7.52  9.24 12.97
r = 0  | 61.20 17.85 19.96 24.60
```

# ตัวอย่างที่ 6.6

เราสามารถสร้างสมการเวกเตอร์เอเรอคอเรกชัน ได้ด้วยคำสั่ง *cajorls* โดยระบุว่าใช้รูปแบบจาก *fsprice.rc* และจำนวนความสัมพันธ์เท่ากับ 1 ($r = 1$)

```
> fsprice.vecm<-cajorls(fsprice.rc, r=1)
> fsprice.vecm
$rlm

Call:
lm(formula = substitute(form1), data = data.mat)

Coefficients:
              lfutures.d  lspot.d
ect1          -1.1705     -0.2296
lfutures.dl1  -1.0964     -0.3301
lspot.dl1      1.3179      0.5226

$beta
                   ect1
lfutures.l2   1.00000000
lspot.l2     -1.00496480
constant      0.03639024
```