เฉลยคำถามท้ายบทที่ 6

การวิเคราะห์สมการถดถอยพหุคูณ

(Multiple Regression Analysis)

1. **จงอธิบายความหมายของแบบจำลองสมการพหุคูณในกรณีที่เกิดปัญหาต่อไปนี้**

   ก. ความไม่คงที่ในความแปรปรวนของตัวคลาดเคลื่อน : heteroscedasticity

   ข. สหสัมพันธ์ระหว่างความแปรปรวนของตัวคลาดเคลื่อน : autocorrelation

   ค. สหสัมพันธ์ระหว่างตัวแปรอิสระ : multicollinearity

ตอบ :

ก. ปัญหา heteroscedasticity คือ การที่ความแปรปรวนของตัวคลาดเคลื่อนในแบบจำลองมีค่าไม่คงคงที่ นั่นคือ Var $(u_t) \neq \sigma^2$ ซึ่งขัดกับข้อกำหนดที่ว่า ความแปรปรวนของตัวคลาดเคลื่อน (u_i) จะต้องมีค่าคงที่ทุกค่าของ $i = 1, 2, \ldots, n$ โดยปัญหาดังกล่าวนี้จะเกิดกับข้อมูลประเภท cross – section data ซึ่งจะมีผลทำให้การประมาณค่าด้วยวิธี OLS ขาดลักษณะที่พึงประสงค์ของตัวประมาณค่าที่ดี

ข. ปัญหา autocorrelation คือ เหตุการณ์ที่ค่าความคลาดเคลื่อน u ในแบบจำลองมีความสัมพันธ์ต่อกัน หรือมี covariation ระหว่าง u_i กับ u_j หรือ $E(u_i, u_i)$ ไม่เท่ากับ 0 เมื่อ $i \neq j$ ซึ่งขัดกับสมมุติฐานที่กำหนดว่า $E(u_i, u_i)$ ต้องเท่ากับ 0 เมื่อ $i \neq j$ โดยปัญหา autocorrelation จะทำให้ระดับนัยสำคัญทางสถิติมีค่าสูงกว่าที่ควรจะเป็น ค่า R^2 และ F ที่ได้จากแบบจำลองไม่น่าเชื่อถือ

ค. ปัญหา multicollinearity คือ ปัญหาที่ตัวแปรอิสระในแบบจำลองมีความสัมพันธ์กันเองเกินกว่าระดับที่ยอมรับได้ ซึ่งละเมิดข้อสมมุติฐานที่ว่า ตัวแปรอิสระที่อยู่ในแบบจำลองต้องไม่มีความสัมพันธ์เชิงเส้น การเกิดปัญหานี้จะทำให้ค่าพารามิเตอร์ของตัวแปรอิสระเกิดความผิดพลาด และไม่มีนัยสำคัญทางสถิติ ส่งผลให้ข้อสรุปเกี่ยวกับ marginal effect ของตัวแปรตามที่เกิดจากตัวแปรอิสระนั้นเกิดความผิดพลาดได้

2. **ถ้าเกิดปัญหาดังกล่าวในข้อ 1 จะมีวิธีการทดสอบปัญหาอย่างไร**

ตอบ :

2.1 การทดสอบปัญหา heteroscedasticity

1) the Spearman Rank correlation test มีขั้นตอนในการทดสอบดังนี้

- หาสมการถดถอย : $Y_i = \alpha + \beta X_i + u_i$ จากนั้นคำนวณหาค่า e (ค่า e คือ ค่าคำนวณของ u โดยที่ e = $Y_i - \hat{Y}$
- เรียงลำดับค่า X_i จากต่ำไปสูง หรือจากสูงไปต่ำ ซึ่ง (ในที่นี้เราจะเรียงจากต่ำไปสูง) และเรียงค่า e_i ไปในทำนองเดียวกัน จากนั้นจับคู่โดยยึดเอาลำดับของ X_i เป็นหลัก แล้วคำนวณค่าสัมประสิทธิ์สหสัมพันธ์ของค่า Rank โดยมีสูตร ดังนี้

$$r = 1 - 6\frac{\sum D_i^2}{n(n^2-1)}$$

โดยที่ : D คือ ผลต่างระหว่างลำดับของ X_i กับ e_i

n คือ จำนวนค่าสังเกต

- พิจารณาค่า Rank คือถ้ามีค่าสูงหรือมีค่าเข้าใกล้ 1 แสดงว่า X_i กับ e_i มีความสัมพันธ์กัน นั่นคือ เกิดปัญหา heteroscedasticity แต่ถ้า Rank มีค่าต่ำ หรือเข้าใกล้ 0 แสดงว่า X_i กับ e_i ไม่มีความสัมพันธ์กัน นั่นคือ ไม่เกิดปัญหา heteroscedasticity
- กรณีมีตัวแปรอิสระหลายตัวก็ให้หาสัมประสิทธิ์สหสัมพันธ์ของ e_t กับตัวแปรอิสระตัวอื่น ๆ ในลักษณะเดียวกัน แล้วพิจารณาค่า Rank ที่คำนวณได้ว่ามีค่าต่ำหรือสูง

2) the Goldfeld – Quandt test มีขั้นตอนดังนี้

- กำหนดสมมุติฐานที่ต้องการทดสอบ ดังนี้

สมมุติฐานหลัก H_o: u_t เป็น homoscedasticity

สมมุติฐานรอง H_a: u_t เป็น heteroscedasticity

- เรียงลำดับ X_i จากต่ำไปสูง หรือสูงไปต่ำ
- ตัดจำนวนตัวอย่างตรงกลางออกจากจำนวนตัวอย่างทั้งหมด เพื่อแบ่งตัวอย่างออกเป็น 2 ชุด คือ ชุดที่มีค่าต่ำ และชุดที่มีค่าสูง
- ประมาณค่าสมการถดถอยทั้ง 2 กลุ่มแล้วคำนวณหาผลบวกของ sum of square residuals: $\sum e_i^2$

$\sum e_1^2$ คือ ตัวรบกวนจากกลุ่ม X_i ที่มีค่าต่ำ ซึ่งมี df = [(n – c)/2] – K

$\sum e_2^2$ คือ ตัวรบกวนจากกลุ่ม X_i ที่มีค่าสูง ซึ่งมี df = [(n – c)/2] – K

โดยที่ : c คือ จำนวนตัวอย่างที่ถูกตัดออก

K คือ จำนวนพารามิเตอร์ที่คำนวณ

คำนวณค่า F – statistic

$$F^* = \sum e_2^2 / \sum e_1^2$$

- หาค่า F จากตาราง ณ ระดับนัยสำคัญทางสถิติ 5% และมี degree of freedom คือ

$$v_1 = v_2 = (n - c - 2K)/2$$

เปรียบเทียบค่า F^* กับ $F_{ตาราง}$ ถ้า $F^* > F_{(n-c-2k)/2,\ \alpha}$ แสดงว่าปฏิเสธ H_o หมายความว่าเกิดปัญหา heteroscedasticity แต่ถ้า $F^* < F_{(n-c-2k)/2,\ \alpha}$ แสดงว่ายอมรับ H_o หมายความว่าไม่เกิดปัญหา heteroscedasticity

2.2 การทดสอบปัญหา autocorrelation

1) Durbin–Watson test โดยมีขั้นตอนในการทดสอบดังนี้

- ตั้งสมมุติฐาน $H_o : \rho = 0$

  $H_a : \rho \neq 0$

- คำนวณค่า Durbin–Watson: D.W. หรือ d หาได้จากสูตร

$$d = \frac{\sum_{t=2}^{n}(u_t - u_{t-1})^2}{\sum_{t=1}^{n} u_t^2}$$

- นำค่า D.W. หรือค่า d ที่คำนวณได้มาเปรียบเทียบกับค่า d ที่ได้จากตาราง ณ ระดับนัยสำคัญที่กำหนด ภายใต้ขนาดตัวอย่าง n และจำนวนตัวแปรอิสระ k

ถ้า	$0 < d < d_L$	ผลลัพธ์ positive autocorrelation
	$d_L < d < d_U$	ผลลัพธ์ ไม่สามารถสรุปได้
	$d_U < d < 4\text{-}d_U$	ผลลัพธ์ ไม่เกิดปัญหา autocorrelation
	$4\text{-}d_U < d < 4\text{-}d_L$	ผลลัพธ์ ไม่สามารถสรุปได้
	$4\text{-}d_L < d < 4$	ผลลัพธ์ negative autocorrelation

2) Run Test มีขั้นตอนในการทดสอบดังนี้

- ตั้งสมมุติฐาน

  $H_o: \rho = 0$

  $H_a: \rho \neq 0$

- หาค่า Runs test จากตาราง Residual

  โดยที่ : N = total number of observations

  N_1= number of + symbols

  N_2= number of – symbols

  n = number of runs

- เปรียบเทียบค่า Runs test จากข้อ 2 กับค่า Runs test จากตาราง

  เราจะยอมรับ H_o ถ้า $n \geq run_{table}$; ไม่เกิดปัญหา autocorrelation

  และจะปฏิเสธ H_o ถ้า $n \leq run_{table}$; เกิดปัญหา autocorrelation

2.3 การทดสอบปัญหา multicolinearity จะพิจารณาได้จากค่า partial correlation coefficient ระหว่างตัวแปรอิสระด้วยกัน ถ้าค่า pearson correlation มีค่ามากกว่า 0.7 ก็อาจจะถือว่าตัวแปรอิสระคู่นั้น ๆ มีความสัมพันธ์กันมาก ควรได้รับการแก้ไข

3. **เมื่อทดสอบพบปัญหาดังกล่าวในข้อ 1 จะมีวิธีการแก้ปัญหาอย่างไร**

ตอบ :

สมการทั่วไป

1. วิธีแก้ปัญหา heteroscedasticity ต้องเปลี่ยนรูปแบบสมการดั่งเดิม เพื่อทำให้ค่ารบกวนของตัวแบบใหม่คงที่จากนั้นดำเนินการหาค่าพารามิเตอร์โดยวิธี OLS สำหรับการแปลงสมการจากรูปแบบเดิมเป็นรูปแบบใหม่มีดังนี้

- หาความสัมพันธ์ของรูปแบบที่แน่นอนระหว่างตัวรบกวนและค่าตัวแปรอิสระ (X_i) ที่ทำให้เกิดปัญหา heteroscedasticity

$$\sigma^2_{ui} = f(X_i)$$

กำหนดความแปรปรวนของตัวรบกวนในรูปแบบที่แน่นอน ซึ่งสมมุติว่าเราทราบค่ารูปแบบดังกล่าว ดังนี้

$$E(u_i^2) = \sigma^2_{ui} = k^2X_i^2$$

นำค่า Square root ของเทอมตัวแปรอิสระ ($\sqrt{X_i}$) ไปหารตัวแปรทุกตัว

$$\frac{Y_i}{\sqrt{X_i}} = b_0\frac{1}{\sqrt{X_i}} + b_1\frac{X_i}{\sqrt{X_i}} + \frac{u_i}{\sqrt{X_i}}$$

สมการที่เกิดจากการเปลี่ยนแปลงนี้ความแปรปรวนของตัวแปรรบกวนใหม่ $\frac{u_i}{\sqrt{X_i}}$ จะมีค่าคงที่พิสูจน์ได้ดังนี้

$$\text{Var}\left(\frac{u_i}{\sqrt{X_i}}\right) = E\left(\frac{u_i}{\sqrt{X_i}}\right)^2 = \frac{1}{X_i^2}E(u_i^2) = \frac{1}{X_i^2}k^2X_i^2 = k^2 = \text{ค่าคงที่}$$

- คำนวณค่าพารามิเตอร์ของสมการที่แปลงแล้ว ด้วยวิธี OLS หรือที่นิยมใช้คือ Weighted Least Squares
- นำค่าพารามิเตอร์ของสมการใหม่กลับไปแทนยังสมการเดิม

2. การแก้ปัญหา autocorrelation

ต้องประมาณค่า ρ : autoregressive coefficient จากนั้นนำค่า ρ ไปคูณกับข้อมูลเดิมของตัวแปรทุกตัว แล้วนำผลคูณของตัวแปรล่าช้าหนึ่งช่วงเวลา ไปลบกับข้อมูลเดิมหลังจากแปลงสมการแล้วค่าคำนวณที่ได้จากวิธี OLS จะมีคุณสมบัติพึ่งปรารถนา แสดงได้โดยสมการต่อไปนี้

สมการเดิม (ซึ่งเกิดปัญหา autocorrelation)

$$Y_t = a + bX_t + u_t \qquad \text{... (1)}$$

สมการในรูปล่าช้า 1 ช่วงเวลา

$$Y_{t-1} = a + bX_{t-1} + u_{t-1} \qquad \text{... (2)}$$

ค่า ρ ไปคูณทั้งสมการ

$$\rho Y_{t-1} = \rho a + \rho bX_{t-1} + \rho u_{t-1} \qquad \text{... (3)}$$

(2) – (3) จะได้

$$(Y_{t-1} - \rho Y_{t-1}) = (a - \rho a) + (bX_{t-1} - \rho bX_{t-1}) + (u_{t-1} - \rho u_{t-1}) \qquad \text{... (4)}$$

เขียนให้อยู่ในรูป $Y_t^* = a^* + bX_t^* + v_t$

โดยที่ : $Y_t^* = (Y_{t-1} - \rho Y_{t-1})$, $a^* = (a - \rho a)$

$bX_t^* = (bX_{t-1} - \rho bX_{t-1})$, $v_t = (u_{t-1} - \rho u_{t-1})$

3. การแก้ปัญหา multicollinearity

- เพิ่มขนาดตัวอย่างหรือชุดข้อมูลให้มากขึ้น
- ตัดตัวแปรใดตัวแปรหนึ่งทิ้งโดยพิจารณาจากความสำคัญของตัวแปรด้วยทฤษฎี ว่าตามทฤษฎีแล้วตัวแปรใดมีความสัมพันธ์มากกว่าก็เลือกตัวแปรนั้นมาใช้แล้วตัดตัวแปรที่มีความสำคัญน้อยกว่าออก
- ถ้าไม่อยากตัดตัวแปรออกเลยก็อาจจะหาความสัมพันธ์ของตัวแปรทั้ง 2 แล้วนำส่วนที่ไม่มีความสัมพันธ์กันมาใช้

4. **จงอธิบายวิธีการทดสอบอำนาจการพยากรณ์**

ตอบ :

การทดสอบว่าสมการประมาณค่า $Y = a + b_1X_1 + b_2X_2 + ... + b_kX_k$ มีอำนาจการพยากรณ์สูงต่ำเพียงใดนั้น ทดสอบได้ง่าย ๆ คือ

- ทำการแทนค่าตัวแปรอิสระทุกตัวจะได้ค่า ณ เวลาต่างๆ $t = 1, 2, 3, ... , n$
- นำค่า $\hat{Y}$ ที่ได้เป็นตัวแปรอิสระ ดูความสัมพันธ์ของตัวแปรตาม Y_t โดย

$$Y_t = a + b\hat{Y}$$

ทำการวิเคราะห์โดยวิธี OLS ได้ค่า a และ b ถ้าทดสอบได้ว่า a = 0 และ b = 1 หรือค่า R^2 เข้าใกล้ 1 แสดงว่า $\hat{Y}$ เป็นตัวพยากรณ์ที่มีประสิทธิภาพ หรือสมการ $\hat{Y} = \hat{\alpha} + \hat{\beta}_1X_1 + \hat{\beta}_2X_2 + ... + \hat{\beta}_kX_k$ เป็นสมการที่สามารถนำไปใช้ในการทำนายผลของตัวแปรตาม Y ได้เป็นอย่างดี