ข้อสอบปลายภาค

วิชา Research Methodology in Finance : MGMG 522

ภาคเรียนที่ 2 ปี 2547

ผู้สอน ดร.ชาญ สรณาคมน์

(คะแนนสอบปลายภาค = 30% ของคะแนนรวม)

(<u>ข้อสอบมีทั้งหมด 4 ข้อ ให้ทำทุกข้อ แต่ละข้อมีคะแนนเท่ากัน</u>)

<u>คำสั่ง</u>

1. อนุญาตให้นิสิตนำตำรา เอกสารอื่น ๆ และเครื่องคิดเลขเข้าห้องสอบได้ แต่ไม่อนุญาตให้หยิบยืมสิ่งต่าง ๆ เหล่านั้นระหว่างกันในระหว่างการสอบ
2. ไม่อนุญาตให้ใช้เครื่องมือสื่อสารทุกชนิดระหว่างการสอบ ซึ่งรวมถึง PDA และ Notebook ทุกชนิดด้วย
3. อ่านโจทย์ให้ดี แล้วตอบให้ตรงประเด็นคำถาม
4. พยายามอย่าปล่อยคำตอบให้ว่าง
5. ตรวจคำตอบให้ดีก่อนส่งกระดาษคำตอบ
6. ห้ามออกนอกห้องสอบระหว่างการสอบไม่ว่ากรณีใด ๆ ทั้งสิ้น ยกเว้นแต่จะได้ส่งคืนกระดาษคำตอบและสละสิทธิ์ที่จะทำข้อสอบต่อ

ข้อที่ 1. ตารางข้างล่างแสดงค่าเฉลี่ยของคะแนนสอบวัดผลความรู้ภาษาอังกฤษของนักเรียนชั้นมัธยมปีที่ 6 ในแต่ละโรงเรียนทั่วกรุงเทพฯ เรียงตามปี ตั้งแต่ปี พ.ศ. 2540-2546 โดยที่โรงเรียนถูกจำแนกเป็น 3 ประเภทเท่านั้นคือ โรงเรียนรัฐบาล โรงเรียนเอกชน และโรงเรียนนานาชาติ

ปี	โรงเรียนที่ 1	โรงเรียนที่ 2	โรงเรียนที่ 3	โรงเรียนที่ 4	โรงเรียนที่ 5	…	โรงเรียนที่ N
2540	60.1	24.3	45.6	48.0	30.1	…	…
2541	63.4	…	…	…	…	…	…
2542	62.9	…	…	…	…	…	…
2543	58.8	…	…	…	…	…	…
2544	61.0	…	…	…	…	…	…
2545	62.7	…	…	…	…	…	…
2546	65.3	…	…	…	…	…	…

ผลจากการรันสมการ Regression ได้สมการความสัมพันธ์ดังนี้ (ค่าในวงเล็บคือค่า Standard Error)

$$\hat{Y} = 12.63 + 5.55D1 + 48.69D2 + 1.03T1 + 1.01T2 + 0.62T3 + 0.12T4 + 1.07T5 + 0.02T6$$

(1.25) (12.04) (1.53) (1.74) (0.99) (1.45) (0.87) (1.25)

$Adj\text{-}R^2 = 0.51$

โดยที่

Y_i = ค่าเฉลี่ยคะแนนภาษาอังกฤษของนักเรียนชั้นมัธยมปีที่ 6 ในโรงเรียนที่ i (มีค่าระหว่าง 0 – 100 คะแนน)

D1 = 1 ถ้าเป็นโรงเรียนมัธยมเอกชน

= 0 ถ้าไม่ได้เป็นโรงเรียนมัธยมเอกชน

D2 = 1 ถ้าเป็นโรงเรียนมัธยมนานาชาติ

= 0 ถ้าไม่ได้เป็นโรงเรียนมัธยมนานาชาติ

T1 = 1 ถ้าข้อมูลเป็นของปี 2541

= 0 ถ้าข้อมูลไม่ใช่ของปี 2541

T2 = 1 ถ้าข้อมูลเป็นของปี 2542

= 0 ถ้าข้อมูลไม่ใช่ของปี 2542

T3 = 1 ถ้าข้อมูลเป็นของปี 2543

= 0 ถ้าข้อมูลไม่ใช่ของปี 2543

T4 = 1 ถ้าข้อมูลเป็นของปี 2544

= 0 ถ้าข้อมูลไม่ใช่ของปี 2544

T5 = 1 ถ้าข้อมูลเป็นของปี 2545

= 0 ถ้าข้อมูลไม่ใช่ของปี 2545

T6 = 1 ถ้าข้อมูลเป็นของปี 2546

= 0 ถ้าข้อมูลไม่ใช่ของปี 2546

ถามว่า

1. โดยเฉลี่ย ณ ปีใดปีหนึ่ง สมการ Regression พยากรณ์ว่านักเรียนมัธยมปีที่ 6 ที่เรียนโรงเรียนนานาชาติทำคะแนนสอบได้<u>ดีกว่าหรือแย่กว่า</u>นักเรียนมัธยมปีที่ 6 ที่เรียนโรงเรียนเอกชน<u>อยู่กี่คะแนน?</u>
2. จากสมการ Regression คุณคิดว่าประเภทโรงเรียน<u>มีความสัมพันธ์หรือไม่อย่างไร</u>กับคะแนนสอบภาษาอังกฤษ? <u>ความสัมพันธ์นั้นมีนัยสำคัญหรือไม่?</u>
3. ถ้าเราติดตามค่าเฉลี่ยของคะแนนสอบภาษาอังกฤษของโรงเรียนใดโรงเรียนหนึ่ง เราพอจะ<u>สรุปอะไรได้บ้าง</u>เกี่ยวกับคะแนนสอบภาษาอังกฤษของนักเรียนมัธยมชั้นปีที่ 6 ในกรุงเทพฯ ช่วงปี พ.ศ. 2540-2546
4. โมเดลนี้ยังขาดตัวแปรที่สำคัญอีกตัวหนึ่ง<u>หรือไม่?</u> กล่าวคือยังขาดตัวแปร D3 โดยที่

   D3 = 1 ถ้าเป็นโรงเรียนมัธยมรัฐบาล

   = 0 ถ้าไม่เป็นโรงเรียนมัธยมรัฐบาล

<u>แนวทางการตอบ</u>

1. นักเรียนมัธยมปีที่ 6 โรงเรียนนานาชาติทำคะแนนได้ดีกว่านักเรียนมัธยมปีที่ 6 โรงเรียนเอกชนอยู่ 43.14 คะแนนโดยเฉลี่ย
2. ประเภทโรงเรียนมีความสัมพันธ์กับคะแนนสอบภาษาอังกฤษอย่างมีนัยสำคัญ (สังเกตุได้จากค่า t ที่คำนวนได้นั้นมากกว่า 2 ทุกตัว) โดยที่โดยเฉลี่ยแล้ว นักเรียนมัธยมปีที่ 6 ในโรงเรียนนานาชาติจะทำคะแนนสอบภาษาอังกฤษได้ดีกว่านักเรียนมัธยมปีที่ 6 ในโรงเรียนเอกชน และนักเรียนมัธยมปีที่ 6 ในโรงเรียนเอกชนจะทำคะแนนสอบภาษาอังกฤษได้ดีกว่านักเรียนมัธยมปีที่ 6 ในโรงเรียนรัฐบาล
3. ค่าเฉลี่ยของคะแนนสอบภาษาอังกฤษของทุกโรงเรียนในช่วงปี พ.ศ. 2540-2546 แทบไม่มีการเปลี่ยนแปลงเลยตลอดช่วงเวลาดังกล่าว
4. ไม่ต้องมี D3 เพราะถ้ามี D3 ในโมเดลจะทำให้เกิดปัญหา Perfect Multicollinearity เกิดขึ้นแล้วจะทำให้ไม่สามารถประมาณค่าสัมประสิทธิ์ต่าง ๆ ของตัวแปรด้วยวิธีการ OLS

ข้อที่ 2. ในสมการ Regression ที่มีตัวแปรตามเป็นแบบ Binary/Dummy Variable (หรือที่เรียกว่า Binary Regression) โมเดลที่เราใช้ไม่ว่าจะเป็น Linear Probability Model, Logit Model, หรือ Probit Model ต่างก็พยากรณ์สิ่งเดียวกัน (บางครั้งต้องมีการแปลงค่าตัวเลขก่อน) <u>สิ่งที่ทั้งสามโมเดลพยากรณ์คืออะไร?</u> และ<u>โมเดลทั้งสามแตกต่างกันในวิธีการอย่างไร?</u>

<u>แนวทางการตอบ</u>

ทั้ง Linear Probability Model, Logit Model, และ Probit Model ต่างก็พยากรณ์สิ่งเดียวกัน คือ เปอร์เซ็นต์ความน่าจะเป็นที่ตัวแปรตามจะมีค่าเท่ากับหนึ่ง ยกตัวอย่างเช่น ถ้าเราต้องการพยากรณ์ว่าผู้สมัครบัตรเครดิตจะได้รับอนุมัติบัตรหรือไม่โดยให้ ตัวแปรตาม Y มี 2 ค่า คือ 1 หรือ 0 โดยที่

Y = 1 หากได้รับอนุมัติบัตร

= 0 หากถูกปฏิเสธ

ทั้ง 3 วิธีใช้ข้อมูลเบื้องต้นเหมือนกัน เช่น ดังตารางข้างล่างนี้

Y	X1	X2	…
0	15	0.5	...
1	17	0.7	...
1	25	0.4	...
0	26	1.1	...
1	30	0.9	...
0	31	0.8	...
...	...	...	...

แต่วิธีการแตกต่างกันดังนี้

1. Linear Probability Model ใช้วิธีประมาณค่าสัมประสิทธิ์แบบตรง ๆ ด้วย OLS โดยเรารันสมการ OLS Regression โดยใช้ตัวแปรต้นและตัวแปรตามดังตารางข้างบน สมการ Regression ที่ได้ คือ

   $\hat{D} = b_0 + b_1 X_1 + b_0 X_2 + ...$

   โดยที่ เมื่อแทนค่า X_1, X_2, … ในสมการดังกล่าว เราจะได้ค่า D คือ เปอร์เซ็นต์ความน่าจะเป็นที่ผู้สมัครจะได้รับอนุมัติบัตรเครดิต ซึ่งค่าเปอร์เซ็นต์ที่ได้อาจจะมากกว่า 100 เปอร์เซ็นต์หรือเป็นค่าติดลบก็เป็นได้

2. วิธี Logit Model เราก็ยังจะใช้ข้อมูลตัวแปรต้นและตัวแปรตามดังตารางข้างบน แต่ในโปรแกรม Eviews ให้ระบุว่าจะใช้วิธี Logit ซึ่งเครื่องจะ estimate สมการออกมาในรูป $\ln\left[\frac{D}{1-D}\right] = b_0 + b_1X_1 + b_0X_2 + ...$ เวลาเราจะหาเปอร์เซ็นต์ความน่าจะเป็นว่าผู้สมัครบัตรเครดิตจะได้รับอนุมัติบัตรเครดิตหรือไม่เราต้องแปลงจาก $\ln\left[\frac{D}{1-D}\right]$ ให้เป็นค่า D เสียก่อน ซึ่งค่า D ที่แปลงเสร็จแล้วจะมีค่าอยู่ระหว่าง 0 ถึง 100 เปอร์เซ็นต์อย่างแน่นอน
3. วิธี Probit Model เราก็ยังใช้ข้อมูลตัวแปรต้นและตัวแปรตามดังตารางข้างบน แต่ในโปรแกรม Eviews ให้ระบุว่าจะใช้วิธี Probit ซึ่งเครื่องจะ estimate สมการออกมาในรูป $\hat{Z} = b_0 + b_1X_1 + b_0X_2 + ...$ เวลาเราจะหาเปอร์เซ็นต์ความน่าจะเป็นว่าผู้สมัครบัตรเครดิตจะได้รับอนุมัติบัตรเครดิตหรือไม่เราต้องแปลงจาก ค่า Z-Score ให้เป็นค่า D เสียก่อนโดยเปิดตาราง Standard Normal ซึ่งค่า D ที่แปลงเสร็จแล้วจะมีค่าอยู่ระหว่าง 0 ถึง 100 เปอร์เซ็นต์อย่างแน่นอน

ข้อที่ 3. จงอธิบายว่า ระบบสมการโครงสร้าง (Structural Equations) <u>แตกต่าง</u>จากสมการเดียว (Single Equation) <u>อย่างไร?</u> และ<u>เพราะอะไร</u>การประมาณค่าสัมประสิทธิ์ของตัวแปรใน Structural Equations <u>จึงไม่ควรทำด้วยวิธีการ OLS แบบธรรมดา?</u>

<u>แนวทางการตอบ</u>

Single Equation อธิบายความสัมพันธ์ระหว่างตัวแปร X กับตัวแปร Y ในความหมายเชิงนัย ๆ ว่า การเปลี่ยนแปลงในตัวแปรต้น X ก่อให้เกิดการเปลี่ยนแปลงในตัวแปรตาม Y โดยที่ความสัมพันธ์ระหว่างตัวแปร X กับตัวแปร Y นั้นสามารถอธิบายได้ด้วยสมการเดียวและไม่มีสมการอื่นใดที่อธิบายความสัมพันธ์ระหว่างตัวแปร X กับตัวแปร Y อีกแล้ว ในขณะที่ Structural Equations อธิบายว่ามีอิทธิพลของตัวแปรต้นบางตัวทำให้เกิดการเปลี่ยนแปลงในตัวแปรตาม (ความสัมพันธ์อันนี้ระบุไว้ในสมการหนึ่งในระบบสมการโครงสร้าง) และในขณะเดียวกันผลจากการเปลี่ยนแปลงในตัวแปรตามกลับมาส่งผลทำให้ตัวแปรต้นตัวนั้นเปลี่ยนแปลง (ความสัมพันธ์ส่วนหลังนี้ระบุไว้ในสมการอีกอันหนึ่ง ซึ่งเป็นคนละสมการในระบบสมการโครงสร้าง)

การประมาณค่าสัมประสิทธิ์ของตัวแปรในสมการโครงสร้างด้วยวิธีการ OLS แบบธรรมดาไม่ควรทำเพราะว่าสมการโครงสร้างได้ violate ข้อสมมุติฐานดั่งเดิมที่สำคัญอันหนึ่งที่ OLS ระบุไว้ว่าจะต้องมีเพื่อให้สัมประสิทธิ์ที่ได้จากวิธีการ OLS นั้นไม่มีความเอนเอียง สมมุติฐานดั่งเดิมที่สำคัญที่สมการโครงสร้าง violate ก็คือ สมมุติฐานที่ว่าตัวแปรต้นทุกตัวไม่มีความสัมพันธ์ใด ๆ กับ error term ซึ่งถ้าเราติดตามผลของการเปลี่ยนแปลงในตัวแปร X, Y และ error term ในระบบสมการโครงสร้างเราจะพบว่า ทุกครั้งที่ค่า error term เปลี่ยนแปลงจะทำให้ตัวแปรต้น X บางตัว

เปลี่ยนแปลงไปทุกครั้ง นั่นก็แปลว่า error term มีความสัมพันธ์กับตัวแปรต้น X ตัวนั้น ผลที่ตามมาก็คือ การใช้ OLS หาค่าสัมประสิทธิ์จะได้ค่าสัมประสิทธิ์ที่มีความเอนเอียง (biased)

ข้อที่ 4. จงอธิบายวิธีการประมาณค่าสัมประสิทธิ์ของตัวแปรด้วยวิธีการสองขั้นตอนของสมการโครงสร้างดังต่อไปนี้

$$Y_{1t} = \beta_0 + \beta_1 Y_{2t} + \beta_2 X_t + \beta_3 Z_t + \beta_4 W_t + \beta_5 P_t + \varepsilon_{1t}$$

$$Y_{2t} = \alpha_0 + \alpha_1 Y_{1t} + \alpha_2 Q_t + \alpha_3 L_t + \varepsilon_{2t}$$

โดยที่

ตัวแปร Y_{1t} และ Y_{2t} เป็นตัวแปร endogenous variables

ตัวแปร X_t, Z_t, W_t, P_t, Q_t, และ L_t เป็นตัวแปร pre-determined variables

และตัวแปร ε_{1t} และ ε_{2t} เป็น classical error terms

<u>แนวทางการตอบ</u>

<u>ขั้นตอนแรก</u>

รัน OLS แบบธรรมดา โดยใช้ specification ข้างล่างสองอันนี้

$$\hat{Y}_{2t} = b_0 + b_1 X_t + b_2 Z_t + b_3 W_t + b_4 P_t + b_5 Q_t + b_6 L_t$$

$$\hat{Y}_{1t} = a_0 + a_1 X_t + a_2 Z_t + a_3 W_t + a_4 P_t + a_5 Q_t + a_6 L_t$$

เราก็จะได้สัมประสิทธิ์ b_0-b_6 และ a_0-a_6 จากนั้นเอาค่าต่าง ๆ ของตัวแปร X_t, Z_t, W_t, P_t, Q_t, และ L_t ไปคูณกับสัมประสิทธิ์ที่ได้มาจาก OLS เพื่อที่จะให้ได้ค่า $\hat{Y}_{1t}$ และ $\hat{Y}_{2t}$ เพื่อนำไปใช้เป็นตัวแปรต้นในขั้นตอนที่สอง

<u>ขั้นตอนที่สอง</u>

รัน OLS แบบธรรมดา โดยใช้ specification ข้างล่างสองอันนี้

$$Y_{1t} = \beta_0 + \beta_1 \hat{Y}_{2t} + \beta_2 X_t + \beta_3 Z_t + \beta_4 W_t + \beta_5 P_t + \varepsilon_{1t}$$

↑ มาจากขั้นตอนแรก

$$Y_{2t} = \alpha_0 + \alpha_1 \hat{Y}_{1t} + \alpha_2 Q_t + \alpha_3 L_t + \varepsilon_{2t}$$

↑ มาจากขั้นตอนแรก