

ข้อสอบกลางภาค วิชา Research Methodology in Finance : MGMG 522

ภาคเรียนที่ 2 ปี 2548

ผู้สอน ดร.ชาญ สรณาคมน์

(คะแนนสอบกลางภาค = 30% ของคะแนนรวม)

(ข้อสอบมีทั้งหมด 5 ข้อ ให้ทำทุกข้อ แต่ละข้อมีคะแนนเท่ากัน)

คำสั่ง

1. อนุญาตให้นำตำรา เอกสารอื่น ๆ และเครื่องคิดเลขเข้าห้องสอบได้ แต่ไม่อนุญาตให้หยิบยืมสิ่งต่าง ๆ เหล่านั้นระหว่างกันในช่วงการสอบ
2. ห้ามใช้ฟังก์ชันเครื่องคิดเลขในเครื่องโทรศัพท์เคลื่อนที่
3. ไม่อนุญาตให้ใช้เครื่องมือสื่อสารทุกชนิดระหว่างการสอบ (ห้ามใช้ PDA และ Notebook ทุกชนิดด้วย)
4. อ่านโจทย์ให้ดี แล้วตอบให้ตรงประเด็นคำถาม
5. พยายามอย่าปล่อยคำตอบให้ว่าง และตรวจคำตอบให้ดีก่อนส่งกระดาษคำตอบ
6. หากต้องการออกนอกห้องสอบระหว่างการสอบ ให้ส่งคืนกระดาษคำตอบเสียก่อน

ข้อที่ 1. ค่าต่าง ๆ ที่ได้จากวิธีการ OLS ใน output ของ EViews มีความสัมพันธ์กัน จงเติมตัวเลขในช่องว่างต่อไปนี้พร้อมทั้งแสดงวิธีการคำนวณเพื่อหาตัวเลขต่าง ๆ เหล่านี้ด้วย

Dependent Variable: Y				
Method: Least Squares				
Date: 11/15/05 Time: 15:54				
Sample: 1 20				
Included observations: 20				
Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	103.3971	9.342100	11.06786	0.0000
X	-6.377093	0.0000
R-squared	0.743121	Mean dependent var	169.4000	
Adjusted R-squared	S.D. dependent var	16.32692	
S.E. of regression	8.501763	Akaike info criterion	7.213064	
Sum squared resid	1301.039	Schwarz criterion	7.312637	
Log likelihood	-70.13064	F-statistic	
Durbin-Watson stat	1.451314	Prob(F-statistic)	0.000001	

แนวทางการตอบ จากค่า R^2 เราจะสามารถหาค่า Adjusted R^2 และค่าสถิติ F ได้

$$\text{โดยที่ } \text{Adjusted } R^2 = 1 - (1 - R^2) \frac{(n-1)}{(n-K-1)}$$

$$\text{และ } F = \frac{R^2 / k}{(1 - R^2) / (n - k - 1)}$$

จากนั้นเราจะนำค่า F ไปถอด Square Root ก็จะได้ค่าสถิติ t สำหรับตัวแปรต้น X (กรณีนี้ค่า t จะต้องเป็นค่าที่ติดลบเพราะสัมประสิทธิ์ของตัวแปร X เป็นค่าติดลบ)

จากค่า t เราสามารถคำนวณหาค่า Standard Error ของสัมประสิทธิ์หน้าตัวแปร X ได้จากสมการ

$$\text{Standard Error } (b_x) = \frac{b_x}{t}$$

ผลลัพธ์จะเป็นดังนี้ (ตัวเลขที่คุณคำนวณจะไม่ตรงกับตัวเลขในตารางข้างล่างนี้เนื่องจากการปัดเศษ)

Dependent Variable: Y Method: Least Squares Date: 11/15/05 Time: 15:54 Sample: 1 20 Included observations: 20				
Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	103.3971	9.342100	11.06786	0.0000
X	-6.377093	0.883732	-7.216091	0.0000
R-squared	0.743121	Mean dependent var	169.4000	
Adjusted R-squared	0.728850	S.D. dependent var	16.32692	
S.E. of regression	8.501763	Akaike info criterion	7.213064	
Sum squared resid	1301.039	Schwarz criterion	7.312637	
Log likelihood	-70.13064	F-statistic	52.07197	
Durbin-Watson stat	1.451314	Prob(F-statistic)	0.000001	

ข้อที่ 2. หากเรามีกลุ่มตัวอย่างเป็นพนักงาน 100 คนซึ่งเป็นเพศชายทั้งหมด แล้วทำการประมาณค่าสมการข้างล่างด้วยวิธี OLS เราจะได้สมการความสัมพันธ์ดังนี้

$$\text{Salary} = 1205 + 2251 \cdot \text{YEAR} + 15000 \cdot \text{GPA}$$

โดยที่ Salary คือ เงินเดือนของพนักงาน (บาท)

YEAR คือ ประสบการณ์การทำงาน (ปี)

และ GPA คือ เกรดเฉลี่ยตอนสำเร็จการศึกษาระดับปริญญาตรีของพนักงาน (0.00-4.00)

ถ้าเราเพิ่มตัวแปร Dummy ที่ชี้แทนเพศ โดยกำหนดให้เพศชาย = 1 และเพศหญิง = 0 ถามว่าเมื่อเรารันสมการ OLS อีกครั้งโดยกำหนดสมการ regression ใหม่เป็น

$$\text{Salary} = b_0 + b_1 \cdot \text{YEAR} + b_2 \cdot \text{GPA} + b_3 \cdot \text{Dummy}$$

สัมประสิทธิ์ b_2 ที่ได้จะเป็นบวกหรือเป็นลบหรือเป็นศูนย์หรือผลลัพธ์จะเป็นอย่างไร? เพราะอะไร? อธิบาย

แนวทางการตอบ

เราจะรัน OLS ไม่ได้เพราะ OLS ไม่สามารถคำนวณหาว่าถ้า Dummy เปลี่ยนแปลงไปหนึ่งหน่วยจะทำให้ Y เปลี่ยนแปลงไปเท่าไร เพราะค่าของ Dummy เท่ากันทุก ๆ observation หรือพูดอีกนัยหนึ่งคือเรากำลังประสบกับปัญหา perfect multicollinearity โดยที่ $Dummy = 1 + 0 \cdot (X1 \text{ หรือ } X2 \text{ ก็ได้})$

ข้อที่ 3. การทดสอบสมมติฐานเพื่อหาปัญหา autocorrelation จะได้ข้อสรุปแตกต่างจากข้อสรุปจากการทดสอบสมมติฐานแบบทั่ว ๆ ไปอย่างไร? และจงทดสอบสมมติฐานเพื่อดูว่าโมเดลข้างล่างนี้มีปัญหา autocorrelation หรือไม่?

Dependent Variable: Y Method: Least Squares Date: 11/07/05 Time: 14:03 Sample: 1960 1999 Included observations: 40				
Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	-102.7258	8.622617	-11.91353	0.0000
X1	1.757755	0.675734	2.601253	0.0134
X2	3.802866	0.172320	22.06862	0.0000
X3	-0.125559	0.167131	-0.751265	0.4574
R-squared	0.988828	Mean dependent var	101.7500	
Adjusted R-squared	0.987897	S.D. dependent var	71.41300	
S.E. of regression	7.856401	Akaike info criterion	7.055174	
Sum squared resid	2222.029	Schwarz criterion	7.224062	
Log likelihood	-137.1035	F-statistic	1062.115	
Durbin-Watson stat	1.465020	Prob(F-statistic)	0.000000	

แนวทางการตอบ

การทดสอบสมมติฐานทั่ว ๆ ไปจะมีข้อสรุป 2 อย่างคือ (1) Reject the null hypothesis หรือไม่ก็ (2) Cannot reject the null hypothesis แต่การทดสอบสมมติฐานเพื่อหาปัญหา autocorrelation จะมีข้อสรุป 3 อย่างคือ

1. Reject the null hypothesis (คือ มีปัญหา autocorrelation)
2. Cannot reject the null hypothesis (คือ ไม่มีปัญหา autocorrelation)
3. Inconclusive (คือ สรุปไม่ได้ว่ามีปัญหา autocorrelation หรือไม่)

กรณีโมเดลที่ให้มา $n = 40$, $K = 3$, ไปเปิดตารางจะได้ค่า $D_L = 1.25$ และ $D_U = 1.57$

เพื่อหาว่าโมเดลนี้มีปัญหา autocorrelation หรือไม่ เราจะใช้การทดสอบสมมติฐานสองด้าน (two-tailed d test) ปรากฏว่า ค่า Durbin-Watson ของเรตกอยู่ช่วงระหว่าง D_L ถึง D_U (inconclusive) ดังนั้นเราจึงไม่สามารถสรุปได้ว่าโมเดลที่ให้มามีปัญหา autocorrelation หรือไม่

ข้อที่ 4. พิจารณาสมการ OLS regression ข้างล่างนี้ (ค่า Std. Error อยู่ในวงเล็บ)

$$\text{GPA} = 1.05 + 0.005 \cdot \text{GMAT} + 0.20 \cdot \text{EXPERIENCE} - 0.10 \cdot \text{AGE} + 0.25 \cdot \text{BIZ}$$

(0.001) (0.20) (0.10) (0.10)

$$\text{Adjusted } R^2 = 0.20 \quad N = 1000$$

โดยที่ GPA = เกรดเฉลี่ยของนักศึกษา MBA ตอนจบการศึกษา

GMAT = คะแนน GMAT ตอนสอบเข้า (คะแนนเต็ม = 800)

EXPERIENCE = ประสบการณ์การทำงานก่อนเข้าเรียน MBA (ปี)

AGE = อายุของนักศึกษา (ปี)

และ BIZ = 1 ถ้านักศึกษาคนนั้นจบปริญญาตรีสาขาบริหารธุรกิจหรือสาขาใกล้เคียง หรือ
= 0 หากจบปริญญาตรีสาขาอื่น ๆ

คำสั่ง

A.) จงทดสอบสมมุติฐานสัมประสิทธิ์ของตัวแปรต้นแต่ละตัว (ให้กำหนดเครื่องหมายเองและให้กำหนดเองว่าจะทดสอบข้างเดียวหรือสองข้างตามความเหมาะสม) ตัวแปรต้นทั้งสี่ตัวมีความสัมพันธ์กับตัวแปรตามหรือไม่อย่างไร? มีนัยสำคัญหรือไม่?

B.) เราสามารถหาค่า F เพื่อทำ F-test ได้หรือไม่อย่างไร? ถ้าทำได้ โมเดลนี้โดยรวมมีนัยสำคัญหรือไม่?

หมายเหตุ: การทดสอบสมมุติฐานใด ๆ ก็ตาม ให้ใช้ระดับนัยสำคัญ .05 ในการทดสอบ

แนวทางการตอบ A.)

- b_{GMAT} น่าจะมีเครื่องหมายเป็นบวก จึงทดสอบสมมุติฐานข้างเดียวด้านบวก ค่า t ที่ได้เท่ากับ 5 ซึ่งหมายถึงมีนัยสำคัญทางสถิติที่ระดับ .05 สรุปได้ว่า คะแนน GMAT ตอนสอบเข้ากับ GPA ตอนจบมีความสัมพันธ์ในทิศทางเดียวกัน
- $b_{\text{EXPERIENCE}}$ น่าจะมีเครื่องหมายเป็นบวก จึงทดสอบสมมุติฐานข้างเดียวด้านบวก ค่า t ที่ได้เท่ากับ 1 ซึ่งหมายถึงไม่มีนัยสำคัญทางสถิติที่ระดับ .05 สรุปได้ว่า ประสบการณ์การทำงานไม่ได้ส่งผลทางบวกต่อ GPA ตอนจบ
- b_{AGE} น่าจะมีค่าไม่เท่ากับศูนย์ จึงทดสอบสมมุติฐานสองข้าง ค่า t ที่ได้เท่ากับ 1 ซึ่งหมายถึงไม่มีนัยสำคัญทางสถิติที่ระดับ .05 สรุปได้ว่า อายุของนักศึกษาไม่มีผลต่อ GPA
- b_{BIZ} น่าจะมีเครื่องหมายเป็นบวก จึงทดสอบสมมุติฐานข้างเดียวด้านบวก ค่า t ที่ได้เท่ากับ 2.5 ซึ่งหมายถึงมีนัยสำคัญทางสถิติที่ระดับ .05 สรุปได้ว่า นักศึกษาที่เรียนปริญญาตรีสาขาบริหารธุรกิจหรือสาขาใกล้เคียงจะมี GPA สูงกว่านักศึกษาที่เรียนปริญญาตรีสาขาอื่น ๆ

B.) จากค่า Adjusted $R^2 = 0.2$ จะสามารถหาค่า R^2 ได้ = 0.2032 และจะสามารถหาค่า F ได้ = 63.4375 ซึ่งค่า F เยอะขนาดนี้ แปลว่า โมเดลนี้โดยรวมมีนัยสำคัญทางสถิติ นั่นก็หมายความว่า ตัวแปรต้นทั้ง 4 ตัวโดยรวมสามารถร่วมกันอธิบาย GPA ตอนจบของนักศึกษา MBA ได้

ข้อที่ 5. ต่อเนื่องจากข้อที่ 4. คุณแปลกใจหรือไม่ที่ผลการรัน OLS ออกมาเป็นเช่นที่คุณสรุปตามคำตอบของคุณในข้อที่ 4. ถ้าคุณรู้สึกแปลกใจกับผลที่ได้ ให้อธิบายโดยละเอียดว่าปัญหาที่เกิดขึ้นน่าจะมาจากสาเหตุอะไรได้บ้าง? และคุณจะทดสอบด้วยวิธีการทางสถิติเพิ่มเติมอย่างไรเพื่อยืนยันสาเหตุของปัญหาเหล่านั้น? (Hints: ปัญหาหลัก ๆ ก็มี omitted variable, irrelevant variable, multicollinearity, และ incorrect functional form ของตัวแปร AGE)

แนวทางการตอบ แปลกใจที่ตัวแปรบางตัวไม่มีนัยสำคัญ อาจเป็นเพราะ

1. มีปัญหา multicollinearity ระหว่าง AGE กับ EXPERIENCE เพราะคนทำงานมานานย่อมมีอายุมาก การใส่ตัวแปรต้นทั้ง 2 ตัวนี้เลยอาจทำให้ตัวแปรทั้ง 2 นี้ไม่มีนัยสำคัญทางสถิติ (multicollinearity ไปเพิ่มค่า Std. Error และไปลดค่า t)
2. AGE กับ EXPERIENCE อาจเป็นตัวแปรเกินมาจริง ๆ ก็ได้ (เลยฟ้องออกมาด้วยค่า t ที่น้อย)
3. Functional form ของตัวแปร AGE อาจจะผิดก็ได้ เพราะช่วงอายุน้อย ๆ คนเราอาจจะเรียนได้ไม่ค่อยดีนัก พออายุมากขึ้นหน่อยอาจจะเรียนได้ดีขึ้น แต่พออายุมากเกินจุด ๆ หนึ่งไปก็จะเรียนได้ไม่ค่อยดี กล่าวคือน่าจะมีพจน์ AGE^2 เป็นตัวแปรต้นด้วย
4. โมเดลนี้อาจจะขาดตัวแปรที่สำคัญไปก็ได้ เลยทำให้ coefficients บางตัวมัน bias แล้วแสดงผลออกมาเป็นว่า AGE กับ EXPERIENCE ไม่มีนัยสำคัญทางสถิติ

การทดสอบหา multicollinearity ระหว่าง AGE กับ EXPERIENCE อาจทำได้โดยการทดสอบ simple correlation test หรือหาค่า VIFs ของ AGE กับ EXPERIENCE

การทดสอบหา omitted variable กับ irrelevant variable ต้องทำโดยทดลองเพิ่มหรือตัดตัวแปรออก แล้วพิจารณาว่านัยสำคัญทางสถิติ และค่า Adjusted R^2 เพิ่มขึ้นหรือลดลงหรือไม่อย่างไร อีกอย่างคือต้องเช็คดูทฤษฎีอีกครั้งว่าการจะอธิบาย GPA ต้องใช้ตัวแปรต้นอะไรบ้าง

ส่วนการทดสอบ functional form อาจใช้การ plot graph ระหว่าง GPA กับตัวแปรต้นแต่ละตัวเพื่อดูความสัมพันธ์ว่าเป็นเชิงเส้นหรือไม่ ถ้าความสัมพันธ์ไม่เป็นเชิงเส้น เราอาจจะใช้การแปลงรูปตัวแปร (take log หรือยกกำลังสอง) หรือเพิ่มพจน์กำลังสองของตัวแปรเข้าไปดู แล้วพิจารณาว่านัยสำคัญทางสถิติ และค่า Adjusted R^2 เพิ่มขึ้นหรือลดลงหรือไม่อย่างไร