

ข้อสอบกลางภาค

วิชา Research Methodology in Finance : MGMG 522

ภาคเรียนที่ 2 ปี 2549

ผู้สอน ดร.ชาญ สรณาคมน์

(คะแนนสอบกลางภาค = 35% ของคะแนนรวม)

(ข้อสอบมีทั้งหมด 5 ข้อ ให้ทำทุกข้อ แต่ละข้อมีคะแนนเท่ากัน)

คำสั่ง

- อนุญาตให้นักศึกษานำเฉพาะสิ่งของต่าง ๆ เหล่านี้เข้าห้องสอบได้เท่านั้น
 - เครื่องคิดเลข (ยกเว้นเครื่องคิดเลขในเครื่องโทรศัพท์เคลื่อนที่ หรือ PDA)
 - ปากกา/ดินสอ/ดินสอกด และ ยางลบ/น้ำยาลบคำผิด
 - กระดาษ A4 หนึ่งแผ่น
- ห้ามยืมสิ่งของใด ๆ ระหว่างกันในขณะที่ทำข้อสอบ (เว้นแต่จะได้รับอนุญาตเสียก่อน)
- ห้ามใช้เครื่องมือสื่อสารทุกชนิดระหว่างการสอบ (ห้ามใช้ PDA และ Notebook ทุกชนิดด้วย)
- อ่านโจทย์ให้ดี แล้วตอบให้ตรงประเด็นคำถาม พยายามอย่าปล่อยคำตอบให้ว่าง
- ตรวจคำตอบให้ดีก่อนส่งกระดาษคำตอบ
- การสอบจะแบ่งออกเป็น 2 ส่วน โดยให้นักศึกษาออกนอกห้องสอบก่อนที่จะทำส่วนที่ 1 เสร็จไม่ว่ากรณีใด ๆ ทั้งสิ้น ยกเว้นแต่จะได้ส่งคืนกระดาษคำตอบและสละสิทธิ์ที่จะทำข้อสอบส่วนที่ 1 ต่อ ภายหลังจากนักศึกษากลับเข้าห้องสอบ ให้ทำข้อสอบในส่วนที่ 2 (นักศึกษาไม่มีสิทธิ์กลับไปดูหรือไปขอตรวจคำตอบ หรือไปแก้ไขกระดาษคำตอบในส่วนที่ 1 แล้วไม่ว่ากรณีใด ๆ ทั้งสิ้น)

ข้อที่ 1. พิจารณาสมการ regression สำหรับการพยากรณ์ราคาบ้านในสหรัฐอเมริกา แล้วตอบคำถาม

$$P_t = 13,196.1 + 16.68Y_t$$

$n = 40$ (ข้อมูลรายปี ระหว่าง ค.ศ. 1964–2003)

โดยที่ P_t = ราคาค่าเช่าเฉลี่ยต่อหลัง ณ ปีที่ t (หน่วย: \$US)

Y_t = ตัวเลข GDP ของสหรัฐอเมริกา ณ ปีที่ t (หน่วย: พันล้าน \$US)

ทั้งตัวแปร P_t และ Y_t อยู่ในรูป nominal term (ยังไม่ได้ปรับผลของเงินเฟ้อออกไป)

คำถาม

- 1.1) จงอธิบายความหมายของตัว intercept (=13,196.1) และ slope coefficient (=16.68) [ให้ระวังเรื่องหน่วยของตัวแปรด้วย]
- 1.2) ปกติเรามักจะเขียนตัวแปร Y อยู่ทางซ้ายมือของสมการ regression เสมอ แต่สมการข้างต้นนั้นตัวแปร Y อยู่ทางขวามือ สมการดังกล่าวเขียนผิดหรือไม่อย่างไร อธิบาย
- 1.3) หากเรา take natural log ทั้งสองข้างของสมการดังกล่าวจะได้ $\ln(P_t) = 4.3 + 0.84\ln(Y_t)$ จงอธิบายความหมายเชิงเศรษฐศาสตร์ของตัว slope coefficient (=0.84)
- 1.4) ถ้าให้ข้อมูลคุณเพิ่มเติมว่า สมการ regression $P_t = 13,196.1 + 16.68Y_t$ มีค่า R-squared = 99.43% คุณแปลกใจหรือไม่? และคุณจะให้คำอธิบายว่าอย่างไรที่ค่า R-squared สูงขนาดนี้?

แนวทางการตอบ

- 1.1) ความหมายของตัว intercept คือ ถ้าตัวเลข GDP เป็นศูนย์ ราคาบ้านจะเท่ากับ \$US 13,196.1 และความหมายของ slope coefficient คือทุก ๆ หนึ่งพันล้านเหรียญสหรัฐฯ ที่ GDP เพิ่มขึ้นในแต่ละปีจะทำให้ราคาบ้านโดยเฉลี่ยสูงขึ้น 16.68 เหรียญสหรัฐฯ
- 1.2) สมการดังกล่าวไม่ได้เขียนผิดแต่อย่างใด ปกติเราจะใช้สัญลักษณ์ Y แทนตัวแปรตาม แต่ก็ไม่ได้ผิดแต่อย่างใดหากเราจะใช้สัญลักษณ์ Y แทนตัวแปรต้น
- 1.3) slope coefficient ของสมการ $\ln(P_t) = 4.3 + 0.84\ln(Y_t)$ บอกเราว่าหาก GDP เพิ่มขึ้น 1% จะทำให้ราคาบ้านเพิ่มขึ้น 0.84%
- 1.4) ไม่แปลกใจ เหตุที่ค่า R-squared สูงขนาดนี้เป็นเพราะว่าทั้ง P_t และ Y_t ต่างก็ประมาณการค่าเงินเฟ้อทั้งคู่ คล้าย ๆ กับเรากำลังรัน regression โดยใช้สมการ
อัตราเงินเฟ้อ (วัดจากราคาบ้านที่เปลี่ยนไป) = $a + b \cdot$ อัตราเงินเฟ้อ (วัดจากตัวเลข GDP ที่เปลี่ยนไป)
จึงไม่น่าแปลกที่เราจะได้ค่า R-squared สูงขนาดนี้ อีกประการหนึ่ง ค่า R-squared สำหรับ time-series regression ตามปกติมักมีค่าสูงอยู่แล้ว

ข้อที่ 2. พิจารณาสมการพยากรณ์จำนวนผู้มาวิ่งออกกำลังกายที่สนามกีฬาของมหาวิทยาลัยข้างล่างนี้

$$Y = 125.0 - 15.0X_1 - 1.0X_2 + 1.5X_3 \text{ ----- (Model \#1)}$$

$$\text{ค่า } t \quad \quad (-3.5) \quad (-1.3) \quad (4.1)$$

$$\text{Adj-}R^2 = .75$$

$$Y = 123.0 - 14.0X_1 + 5.5X_2 - 3.7X_4 \text{ ---- (Model \#2)}$$

$$\text{ค่า } t \quad (-2.1) \quad (3.5) \quad (-1.7)$$

$$\text{Adj-}R^2 = .73$$

โดยที่ Y = จำนวนผู้มาวิ่งออกกำลังกายในวันหนึ่ง ๆ (หน่วย: คน)

X_1 = ปริมาณฝนในวันนั้น ๆ (หน่วย: มิลลิเมตร)

X_2 = ตัวแปรหุ่น โดยมีค่าเท่ากับ 1 หากอากาศวันนั้นปลอดโปร่ง (= 0 ในกรณีอื่น ๆ)

X_3 = อุณหภูมิสูงสุดในวันนั้น ๆ (หน่วย: องศาเซลเซียส)

X_4 = จำนวนวิชาเรียนที่มี term paper หรือการบ้านที่นักศึกษาต้องส่งงานในวันรุ่งขึ้น (หน่วย: วิชา)

คำถาม

- 2.1) ให้พิจารณาเลือกโมเดลที่คุณคิดว่าดีกว่า พร้อมทั้งอธิบายเหตุผลในการเลือกโมเดลนั้นตามเกณฑ์ “the four criteria”
- 2.2) ทำไม coefficient estimate หน้าตัวแปร X_2 จึงมีเครื่องหมายต่างกันได้ทั้งสอง model อธิบาย

แนวทางการตอบ

- 2.1) เลือก model #2 เพราะถึงแม้ค่า Adj- R^2 ของ model #2 จะต่ำกว่า model #1 แต่ก็อยู่ในระดับค่อนข้างสูง สัมประสิทธิ์หน้าตัวแปรทุกตัวมีเครื่องหมายที่สอดคล้องกับทฤษฎีหรือ common sense ทั่ว ๆ ไป ดูค่า t โดยรวมก็อยู่ในเกณฑ์ใช้ได้
- 2.2) สาเหตุที่ coefficient estimate หน้าตัวแปร X_2 มีเครื่องหมายต่างกันได้ อาจเป็นเพราะใน model #1 อาจมีปัญหาตัวแปรขาด (omitted variable) เลยทำให้สัมประสิทธิ์หน้าตัวแปร X_2 เกิดการ bias downward

ข้อที่ 3. ให้ใช้ EViews Outputs ข้างล่างนี้ตอบคำถาม

Dependent Variable: PCON Method: Least Squares Date: 10/30/06 Time: 18:41 Sample: 1 50 Included observations: 50				
Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	389.5715	144.4328		
UHM	60.75874	10.25771		
TAX	-36.46995	13.15459		
REG	-0.061170	0.042694		
R-squared	0.924190	Mean dependent var		603.7000
Adjusted R-squared	0.919246	S.D. dependent var		677.8267
S.E. of regression	192.6203	Akaike info criterion		13.43594
Sum squared resid	1706718.	Schwarz criterion		13.58890
Log likelihood	-331.8984	F-statistic		186.9258
Durbin-Watson stat	2.208489	Prob(F-statistic)		0.000000

โดยที่ $PCON_i$ = ปริมาณการบริโภคน้ำมันในมลรัฐที่ i^{th} (หน่วย: ล้านล้าน BTU)

UHM_i = ระยะทางรวมของถนนในมลรัฐที่ i^{th} (หน่วย: ไมล์)

TAX_i = อัตราภาษีน้ำมันในมลรัฐที่ i^{th} (หน่วย: cents ต่อแกลลอน; 1 แกลลอน $\approx$ 3.785 ลิตร)

REG_i = จำนวนรถยนต์ที่จดทะเบียนในมลรัฐที่ i^{th} (หน่วย: พันคัน)

คำถาม

- 3.1) จงทดสอบสมมติฐานสัมประสิทธิ์หน้าตัวแปรต้น UHM, TAX, และ REG (ให้เลือกว่าจะทดสอบด้านเดียวหรือสองด้าน พร้อมระบุเหตุผลประกอบการเลือกด้วย) ณ ระดับนัยสำคัญทางสถิติ 0.05 พร้อมทั้งสรุปผลการทดสอบสมมติฐาน (ว่า significant หรือ not significant แปลว่าอะไร?)
- 3.2) จงทดสอบสมมติฐานด้วย F-test ณ ระดับนัยสำคัญทางสถิติ 0.05 พร้อมทั้งสรุปผลการทดสอบสมมติฐาน (ว่า significant หรือ not significant แปลว่าอะไร?)
- 3.3) พิจารณาผลจากการรัน Chow Test โดยแบ่งข้อมูลออกเป็นกลุ่มมลรัฐทางตอนบนของสหรัฐรวม Alaska (observation ที่ 1-28) กับกลุ่มมลรัฐทางตอนใต้ของสหรัฐรวม Hawaii (observation ที่ 29-50) ได้ผลลัพธ์ตามข้างล่างนี้

Chow Breakpoint Test: 29		
F-statistic	3.079540	Probability 0.026001
Log likelihood ratio	12.85945	Probability 0.011983

จงเขียน null hypothesis และ alternative hypothesis, ทดสอบสมมติฐาน, และเขียนสรุปผลการทดสอบสมมติฐานว่าเราได้ข้อสรุปว่าอย่างไร

แนวทางการตอบ

- 3.1) จะทดสอบสมมุติฐานสัมประสิทธิ์หน้าตัวแปรต้น UHM และ REG ด้านบวก ($H_A: \beta_{UHM} > 0$, และ $\beta_{REG} > 0$) ด้านเดียวเพราะระยะทางรวมของถนนและจำนวนรถยนต์ที่จดทะเบียนที่มากขึ้นจะทำให้การบริโภคน้ำมันสูงตามไปด้วย และจะทดสอบด้านลบ ($H_A: \beta_{TAX} < 0$) ด้านเดียวสำหรับตัวแปร TAX เพราะภาษีที่เพิ่มขึ้นจะทำให้การบริโภคน้ำมันลดลง ค่า Critical t ณ ระดับนัยสำคัญ 0.05 และ $DF = 50 - 3 - 1 = 46$ มีค่าประมาณ 1.684

ผลลัพธ์ UHM กับ TAX sig แต่ REG ไม่ sig แปลว่าตัวแปร UHM มีความสัมพันธ์ทางบวกกับปริมาณการบริโภคน้ำมัน และ TAX มีความสัมพันธ์ทางลบกับปริมาณการบริโภคน้ำมัน แต่ REG ไม่มีผลต่อปริมาณการบริโภคน้ำมัน (ดูตารางข้างล่างนี้ประกอบ)

Dependent Variable: PCON Method: Least Squares Date: 10/30/06 Time: 14:10 Sample: 1 50 Included observations: 50				
Variable	Coefficient	Std. Error	t-Statistic	Critical t
C	389.5715	144.4328		
UHM	60.75874	10.25771	5.923	1.684
TAX	-36.46995	13.15459	-2.772	-1.684
REG	-0.061170	0.042694	-1.433	1.684
R-squared	0.924190	Mean dependent var	603.7000	
Adjusted R-squared	0.919246	S.D. dependent var	677.8267	
S.E. of regression	192.6203	Akaike info criterion	13.43594	
Sum squared resid	1706718	Schwarz criterion	13.58890	
Log likelihood	-331.8984	F-statistic	186.9258	
Durbin-Watson stat	2.208489	Prob(F-statistic)	0.000000	

- 3.2) จะทดสอบสมมุติฐานด้วย F-test ณ ระดับนัยสำคัญทางสถิติ 0.05 พร้อมทั้งสรุปผลการทดสอบสมมุติฐาน ว่า Sig หรือ ไม่ Sig แปลว่าอะไร?

ค่า $F = 186.9258 > \text{Critical } F = 2.84$ ($DF = 3, 46$) สรุปได้ว่า Sig แปลว่าโมเดลโดยรวมใช้ได้ ตัวแปรต้นทั้ง 3 ตัวในสมการสามารถอธิบายตัวแปรตามได้ดี

- 3.3) จาก Chow Test ได้ค่า $F = 3.07954 > \text{Critical } F = 2.61$ ($DF = 4, 42$) สรุปได้ว่า Sig แปลว่าปัจจัย UHM, TAX, และ REG มีผลต่อปริมาณการบริโภคน้ำมันแตกต่างกันระหว่างกลุ่มมลรัฐทางเหนือและกลุ่มมลรัฐทางใต้

ข้อที่ 4. เนื่องจากการทดสอบทางสถิติอย่างเป็นทางการ (formal test for multicollinearity problem) เพื่อหาปัญหา multicollinearity นั้นยังไม่มี เราพอจะมีแนวทางหรือ guideline อะไรบ้างในการทดสอบหาปัญหา multicollinearity อธิบายโดยละเอียด

แนวทางการตอบ

1. ใช้ Simple Correlation Test (ถ้าค่า $|r| > 0.8$ หรือ t-test ของค่า r ปรากฏผลว่า sig แสดงว่าอาจมีปัญหา multicollinearity)
2. ดูค่า Variance Inflation Factor (ถ้า $VIF > 5$ แสดงว่าอาจมีปัญหา multicollinearity)

ข้อที่ 5. ผลการรันสมการ regression ใน EViews โดยระบุรูปแบบสมการ $Y = C + X_1 + X_2 + X_3$ ได้ผลดังนี้

Dependent Variable: Y Method: Least Squares Date: 10/30/06 Time: 16:26 Sample: 1 45 Included observations: 45				
Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	1.293485	2.037718	0.634771	0.5291
X1	0.000152	0.000380	0.399589	0.6915
X2	0.004196	0.000608	6.902785	0.0000
X3	-0.233926	0.176105	-1.328331	0.1914
R-squared	0.957605	Mean dependent var	12.25111	
Adjusted R-squared	0.954503	S.D. dependent var	11.66951	
S.E. of regression	2.489107	Akaike info criterion	4.746412	
Sum squared resid	254.0218	Schwarz criterion	4.907005	
Log likelihood	-102.7943	F-statistic	308.6998	
Durbin-Watson stat	1.570738	Prob(F-statistic)	0.000000	

ให้ทำ General F-Test โดยกำหนดสมมุติฐานที่จะทดสอบดังนี้

$$H_0: \beta_1 = 0$$

$$\beta_3 = 0$$

$$H_A: H_0 \text{ is not true.}$$

คำถาม ผลการทดสอบสมมุติฐานเป็นอย่างไร? และสรุปได้ว่าอย่างไร?

Note: ไม่ต้องพิจารณาว่าตัวแปรในสมการ regression คืออะไร

$$\text{สูตร } F = \frac{(RSS_C - RSS_U)/M}{RSS_U/(n - K - 1)}$$

Dependent Variable: Y Method: Least Squares Date: 10/30/06 Time: 16:37 Sample: 1 45 Included observations: 45				
Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	-1.297983	0.593725	-2.186169	0.0344
X1	0.000131	0.000383	0.341527	0.7344
X2	0.004295	0.000609	7.053571	0.0000
R-squared	0.955781	Mean dependent var	12.25111	
Adjusted R-squared	0.953675	S.D. dependent var	11.66951	
S.E. of regression	2.511658	Akaike info criterion	4.744103	
Sum squared resid	264.9538	Schwarz criterion	4.864548	
Log likelihood	-103.7423	F-statistic	453.9056	
Durbin-Watson stat	1.520596	Prob(F-statistic)	0.000000	

Dependent Variable: Y Method: Least Squares Date: 10/30/06 Time: 17:21 Sample: 1 45 Included observations: 45				
Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	4.311446	2.891487	1.491083	0.1434
X1	0.002689	0.000142	18.96739	0.0000
X3	-0.381937	0.253945	-1.504014	0.1401
R-squared	0.908336	Mean dependent var	12.25111	
Adjusted R-squared	0.903971	S.D. dependent var	11.66951	
S.E. of regression	3.616217	Akaike info criterion	5.473074	
Sum squared resid	549.2350	Schwarz criterion	5.593518	
Log likelihood	-120.1442	F-statistic	208.0969	
Durbin-Watson stat	2.069910	Prob(F-statistic)	0.000000	

Dependent Variable: Y Method: Least Squares Date: 10/30/06 Time: 17:37 Sample: 1 45 Included observations: 45				
Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	1.223303	2.009723	0.608693	0.5460
X2	0.004431	0.000154	28.69324	0.0000
X3	-0.230999	0.174184	-1.326177	0.1919
R-squared	0.957440	Mean dependent var	12.25111	
Adjusted R-squared	0.955413	S.D. dependent var	11.66951	
S.E. of regression	2.464080	Akaike info criterion	4.705855	
Sum squared resid	255.0111	Schwarz criterion	4.826299	
Log likelihood	-102.8817	F-statistic	472.4220	
Durbin-Watson stat	1.550197	Prob(F-statistic)	0.000000	

Dependent Variable: Y Method: Least Squares Date: 10/30/06 Time: 17:33 Sample: 1 45 Included observations: 45				
Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	0.133812	0.815021	0.164182	0.8704
X1	0.002753	0.000137	20.05316	0.0000
R-squared	0.903399	Mean dependent var	12.25111	
Adjusted R-squared	0.901152	S.D. dependent var	11.66951	
S.E. of regression	3.668901	Akaike info criterion	5.481088	
Sum squared resid	578.8160	Schwarz criterion	5.561384	
Log likelihood	-121.3245	F-statistic	402.1293	
Durbin-Watson stat	1.988638	Prob(F-statistic)	0.000000	

Dependent Variable: Y Method: Least Squares Date: 10/30/06 Time: 17:33 Sample: 1 45 Included observations: 45				
Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	-1.330566	0.579959	-2.294240	0.0267
X2	0.004496	0.000148	30.44229	0.0000
R-squared	0.955658	Mean dependent var	12.25111	
Adjusted R-squared	0.954627	S.D. dependent var	11.66951	
S.E. of regression	2.485725	Akaike info criterion	4.702432	
Sum squared resid	265.6896	Schwarz criterion	4.782728	
Log likelihood	-103.8047	F-statistic	926.7327	
Durbin-Watson stat	1.507588	Prob(F-statistic)	0.000000	

Dependent Variable: Y Method: Least Squares Date: 10/30/06 Time: 17:37 Sample: 1 45 Included observations: 45				
Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	30.82843	7.736607	3.984748	0.0003
X3	-1.820592	0.740797	-2.457611	0.0181
R-squared	0.123162	Mean dependent var	12.25111	
Adjusted R-squared	0.102771	S.D. dependent var	11.66951	
S.E. of regression	11.05362	Akaike info criterion	7.686820	
Sum squared resid	5253.848	Schwarz criterion	7.767116	
Log likelihood	-170.9534	F-statistic	6.039853	
Durbin-Watson stat	1.534652	Prob(F-statistic)	0.018094	

แนวทางการตอบ

Critical F ณ DF = (2, 41) มีค่าประมาณ 3.23

$$F = \frac{(265.6896 - 254.0218)/2}{254.0218/(45 - 3 - 1)} = 0.9416 < \text{Critical F} \text{ สรุปได้ว่า ไม่ significant}$$

แปลว่า สมการ regression ที่กำหนดไว้ตอนแรก (Y C X1 X2 X3) มี statistical fit ไม่แตกต่างจากสมการ regression ที่มีตัวแปรต้น X2 ตัวเดียว เรามักจะสรุปว่าเราจะใช้สมการที่มี X2 ตัวเดียว