ข้อสอบกลางภาค

วิชา Research Methodology in Finance : MGMG 522

ภาคเรียนที่ 2 ปี 2549

ผู้สอน ดร.ชาญ สรณาคมน์

(คะแนนสอบกลางภาค = 35% ของคะแนนรวม)

(<u>ข้อสอบมีทั้งหมด 5 ข้อ ให้ทำทุกข้อ แต่ละข้อมีคะแนนเท่ากัน</u>)

<u>คำสั่ง</u>

1. อนุญาตให้นักศึกษานำเฉพาะสิ่งของต่าง ๆ เหล่านี้เข้าห้องสอบได้เท่านั้น
    a. เครื่องคิดเลข (ยกเว้นเครื่องคิดเลขในเครื่องโทรศัพท์เคลื่อนที่ หรือ PDA)
    b. ปากกา/ดินสอ/ดินสอกด และ ยางลบ/น้ำยาลบคำผิด
    c. กระดาษ A4 หนึ่งแผ่น
2. ห้ามยืมสิ่งของใด ๆ ระหว่างกันในขณะทำข้อสอบ (เว้นแต่จะได้รับอนุญาตเสียก่อน)
3. ห้ามใช้เครื่องมือสื่อสารทุกชนิดระหว่างการสอบ (ห้ามใช้ PDA และ Notebook ทุกชนิดด้วย)
4. อ่านโจทย์ให้ดี แล้วตอบให้ตรงประเด็นคำถาม พยายามอย่าปล่อยคำตอบให้ว่าง
5. ตรวจคำตอบให้ดีก่อนส่งกระดาษคำตอบ
6. การสอบจะแบ่งออกเป็น 2 ส่วน โดยห้ามนักศึกษาออกนอกห้องสอบก่อนที่จะทำส่วนที่ 1 เสร็จไม่ว่ากรณีใด ๆ ทั้งสิ้น ยกเว้นแต่จะได้ส่งคืนกระดาษคำตอบและสละสิทธิ์ที่จะทำข้อสอบส่วนที่ 1 ต่อ ภายหลังจากนักศึกษากลับเข้าห้องสอบ ให้ทำข้อสอบในส่วนที่ 2 (นักศึกษาไม่มีสิทธิ์กลับไปดูหรือไปขอตรวจคำตอบหรือไปแก้ไขกระดาษคำตอบในส่วนที่ 1 แล้วไม่ว่ากรณีใด ๆ ทั้งสิ้น)

ข้อที่ 1. Assumption ที่ 1 กำหนดว่า สมการ OLS regression ที่เราจะรันต้อง “correctly specified” จง <u>อธิบายโดยละเอียด</u>ว่า มีกรณีใดบ้างที่เราจะ violate assumption ข้อนี้ และ<u>ผลกระทบที่ตามมา</u>ในแต่ละกรณีมีอะไรบ้าง?

<u>แนวทางการตอบ</u> มี 3 กรณีที่เราจะ violate assumption ข้อนี้

1.1) กรณีขาดตัวแปรที่สำคัญ ผลที่ตามมาคือเราจะได้ biased coefficient estimates, standard errors ของ coefficient estimates ลดลง, และค่า t-value เพิ่มสูงขึ้น

1.2) กรณีมีตัวแปรเกิน ผลที่ตามมาคือเราจะได้ unbiased coefficient estimates, standard errors ของ coefficient estimates เพิ่มขึ้น, และค่า t-value ลดลง

1.3) กรณี functional form ผิดพลาด (เรายังไม่ได้เรียนผลกระทบที่ตามมา)

ข้อที่ 2. พิจารณาสมการพยากรณ์จำนวนผู้มาวิ่งออกกำลังกายที่สนามกีฬาของมหาวิทยาลัยข้างล่างนี้

$Y = 125.0 - 1.5X_1$ ----- (Model #1)

ค่า standard error (.45)

n = 50, Adj-R^2 = .75

$Y = 167.0 - 1.4X_1 - 0.05X_2$ ----- (Model #2)

ค่า standard error (.71) (.03)

n = 50, Adj-R^2 = .73

โดยที่ Y = จำนวนผู้มาวิ่งออกกำลังกายในวันหนึ่ง ๆ (หน่วย: คน)

X_1 = อุณหภูมิสูงสุดในวันนั้น ๆ (หน่วย: องศาเซลเซียส)

$X_2 = X_1^2$

<u>คำถาม</u>

2.1) Model #2 มีปัญหา perfect multicollinearity ระหว่างตัวแปรต้น X_1 กับ X_2 หรือไม่? เพราะว่า X_2 สามารถเขียนเป็นฟังก์ชั่นของ X_1 ได้อย่างสมบูรณ์ดังนี้ คือ $X_2 = X_1^2$ <u>อธิบาย</u>

2.2) จงแปลความหมายค่า coefficient หน้าตัวแปร X_1 และ coefficient หน้าตัวแปร X_2 ว่ามีความหมายว่าอะไร? (<u>ระวังเรื่องหน่วยของการวัดด้วย</u>)

2.3) จะมีผู้มาวิ่งออกกำลังกายที่สนามกีฬาของมหาวิทยาลัยกี่คน หากวันนั้นอุณหภูมิสูงสุดอยู่ที่ 30 องศาเซลเซียส โดยให้ใช้ Model #2 ในการทำนาย

2.4) จงทดสอบสมมุติฐานสัมประสิทธิ์หน้าตัวแปรต้น X_1 และ X_2 (ให้เลือกว่าจะทดสอบด้านเดียวหรือสองด้าน พร้อมระบุเหตุผลประกอบการเลือกด้วย) ณ ระดับนัยสำคัญทางสถิติ 0.05 พร้อมทั้งสรุปผลการทดสอบสมมุติฐาน (ว่า significant หรือ not significant แปลว่าอะไร?)

<u>แนวทางการตอบ</u>

2.1) Model #2 ไม่มีปัญหา perfect multicollinearity ระหว่างตัวแปรต้น X_1 กับ X_2 เพราะถึงแม้ว่า X_2 จะสามารถเขียนเป็นฟังก์ชั่นของ X_1 ได้อย่างสมบูรณ์ดังนี้ คือ $X_2 = X_1^2$ แต่ฟังก์ชั่นดังกล่าวไม่ใช่ฟังก์ชั่นเชิงเส้น จึงไม่ข่าย perfect multicollinearity

2.2) - จาก Model #1 อธิบายได้ว่า ทุก ๆ หนึ่งองศาเซลเซียสที่อุณหภูมิสูงสุดในวันหนึ่ง ๆ เพิ่มขึ้นจะทำให้มีผู้มาวิ่งออกกำลังกายในวันนั้น ๆ ลดลงไปโดยเฉลี่ย 1.5 คน

- จาก Model #2 อธิบายได้ว่า ทุก ๆ หนึ่งองศาเซลเซียสที่อุณหภูมิสูงสุดในวันหนึ่ง ๆ เพิ่มขึ้นจะทำให้มีผู้มาวิ่งออกกำลังกายในวันนั้น ๆ ลดลงไปโดยเฉลี่ย 1.4 คน (ถ้า X_2 คงที่) และทุก ๆ หนึ่งองศาเซลเซียสกำลังสองที่อุณหภูมิสูงสุดในวันหนึ่ง ๆ เพิ่มขึ้นจะทำให้มีผู้มาวิ่งออกกำลังกายในวันนั้น ๆ ลดลงไปโดยเฉลี่ย 0.05 คน (ถ้า X_1 คงที่)

2.3) 80 คน

2.4) จาก Model #2 จะทดสอบสมมุติฐานสัมประสิทธิ์หน้าตัวแปรต้น X_1 และ X_2 ด้านลบด้านเดียว (H_A: $\beta_{X1} < 0$, และ $\beta_{X2} < 0$) เพราะว่า ถ้าอุณหภูมิสูงสุดในวันหนึ่ง ๆ มีค่ามาก แปลว่าอากาศร้อนจัดน่าจะเป็นเหตุให้คนมาวิ่งออกกำลังกายในวันนั้น ๆ ลดลง

- ค่า critical t (ระดับนัยสำคัญ 0.05 และ DF = 50-2-1 = 47) มีค่าประมาณ -1.684
- ค่า t ของ X_1 = -1.4/.71 = -1.97 > critical t (ไม่พิจารณาเครื่องหมายลบ) สรุปว่า reject H_0 แปลว่า อุณหภูมิสูงสุดในวันหนึ่ง ๆ เป็นปัจจัยที่ทำให้คนมาวิ่งออกกำลังกายลดลง
- ค่า t ของ X_2 = -.05/.03 = -1.67 < critical t (ไม่พิจารณาเครื่องหมายลบ) สรุปว่า fail to reject H_0 แปลว่า อุณหภูมิสูงสุดกำลังสองในวันหนึ่ง ๆ ไม่ได้เป็นปัจจัยที่ทำให้คนมาวิ่งออกกำลังกายลดลง

ข้อที่ 3. จากการรัน OLS regression โดยใช้ $PCON_i$ เป็นตัวแปรต้น และ UHM_i, TAX_i, และ REG_i เป็นตัวแปรตามได้ผลลัพธ์ดังนี้ (ใช้ข้อมูล 50 มลรัฐ หรือ n = 50)

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	2.697250	0.0097
UHM	60.75874	10.25771	5.923225	0.0000
TAX	-36.46995	13.15459	-2.772413	0.0080
REG	-0.061170	0.042694	-1.432761	0.1587

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 ≈ 3.785 ลิตร)

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

<u>คำถาม</u>

3.1) ใช้ Output จากการรัน Ramsey's Regression Specification Error Test (RESET) ข้างล่างนี้ จงเขียน null hypothesis และ alternative hypothesis, ทดสอบสมมุติฐาน, และเขียนสรุปผลการทดสอบสมมุติฐานว่าเราได้ข้อสรุปว่าอย่างไร?

Ramsey RESET Test:

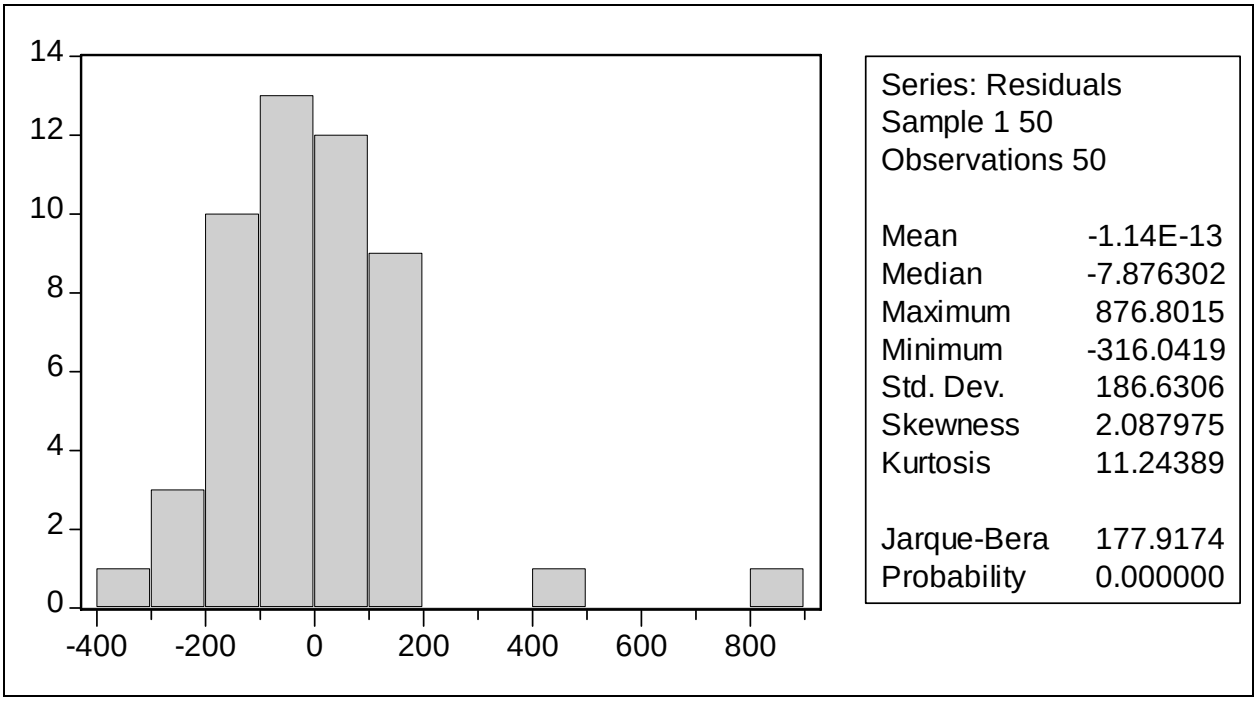
F-statistic	4.885499	Probability	0.005201
Log likelihood ratio	14.66514	Probability	0.002126

Test Equation:
Dependent Variable: PCON
Method: Least Squares
Date: 11/02/06 Time: 21:25
Sample: 1 50
Included observations: 50

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	258.4291	160.7458	1.607688	0.1152
UHM	2.261773	31.63806	0.071489	0.9433
TAX	-14.16436	17.37396	-0.815264	0.4194
REG	0.057035	0.066487	0.857827	0.3957
FITTED^2	0.000938	0.000727	1.289847	0.2040
FITTED^3	-5.79E-07	4.50E-07	-1.287521	0.2048
FITTED^4	1.15E-10	8.40E-11	1.369075	0.1781

R-squared	0.943461	Mean dependent var	603.7000
Adjusted R-squared	0.935572	S.D. dependent var	677.8267
S.E. of regression	172.0508	Akaike info criterion	13.26263
Sum squared resid	1272864	Schwarz criterion	13.53032
Log likelihood	-324.5659	F-statistic	119.5893
Durbin-Watson stat	2.046010	Prob(F-statistic)	0.000000

3.2) ถ้า Residuals หรือ e_i = Actual $PCON_i$ – Predicted $PCON_i$ ผลจากการทดสอบสมมุติฐานด้วยสถิติ Jarque-Bera ว่า e_i มีการแจงแจงแบบปกติหรือไม่? ได้ผลลัพธ์ตามตารางข้างล่างนี้ จงเขียน null hypothesis และ alternative hypothesis, ทดสอบสมมุติฐาน, และเขียนสรุปผลการทดสอบสมมุติฐานว่าเราได้ข้อสรุปว่าอย่างไร?



<u>แนวทางการตอบ</u>

3.1) H_0 = โมเดล PCON C UHM TAX REG ไม่มีปัญหา specification error

H_A = โมเดล PCON C UHM TAX REG มีปัญหา specification error

ค่า critical F ณ ระดับนัยสำคัญ 0.05, DF = (3, 50-6-1=43) มีค่าประมาณ 2.83

ค่า F ที่ได้จาก RESET Test = 4.885499 > 2.83 สรุปได้ว่า reject H_0 แปลว่า โมเดล PCON C UHM TAX REG มีปัญหา specification error

3.2) H_0 = Residuals มีการแจกแจงแบบปกติ

H_A = Residuals ไม่มีการแจกแจงแบบปกติ

เราไม่จำเป็นต้องรู้ว่าการแจกแจงของตัวสถิติ Jarque-Bera เป็นแบบใด (ถ้าอยากรู้: Jarque-Bera มีการแจกแจงแบบ Chi-squared, DF = 2) แต่ถ้าเมื่อใดก็ตามเห็น p-value ของตัวสถิตินั้น ๆ มีค่าน้อยกว่าระดับนัยสำคัญที่เราจะทดสอบ (เช่น 0.05) เราสามารถสรุปได้ทันทีว่า reject H_0 เช่น ในกรณีนี้ เราสามารถสรุปได้ว่า reject H_0 ซึ่งแปลว่า Residuals ไม่มีการแจกแจงแบบปกติ

ข้อที่ 4. “ในการศึกษาวิจัยเกี่ยวกับเรื่องใดเรื่องหนึ่ง เช่น อัตราเงินเฟ้อ เราก็จะกำหนดให้อัตราเงินเฟ้อเป็นตัวแปรตาม แล้วหน้าที่ของนักวิจัยก็คือ การค้นหาปัจจัยหรือตัวแปรต้นที่มีผลกระทบต่ออัตราเงินเฟ้อ ซึ่งถ้านักวิจัยพยายามทำการศึกษาวิจัยอย่างดีเยี่ยม ท้ายที่สุดแล้ว นักวิจัยก็จะค้นพบแบบจำลองความสัมพันธ์ระหว่างอัตราเงินเฟ้อและปัจจัยที่อธิบายอัตราเงินเฟ้อที่แท้จริงและถูกต้อง” คำกล่าวดังกล่าวถูกหรือผิดอย่างไร? อธิบายโดยละเอียด

แนวทางการตอบ

ผิด เพราะว่า ไม่ว่าเราจะพยายามมากเพียงใดก็ตาม สิ่งที่เราทำได้ก็คือ เราจะได้โมเดลที่มี fit ที่ดี, มีตัวแปรที่มีนัยสำคัญอยู่ในโมเดล, และสามารถใช้พยากรณ์ตัวแปรตามได้ดีในระดับที่เราพอใจ แต่โมเดลที่ถูกต้องแท้จริงไม่มีใครรู้ว่าหน้าตาเป็นอย่างไรและไม่มีใครหาเจอ (ถึงเจอ เราก็ไม่รู้ว่าเราเจอมันแล้ว!)

ข้อที่ 5. มีผลงานวิจัยชิ้นหนึ่งในต่างประเทศอ้างว่า (1) คนที่หน้าตาดีไม่ว่าผู้หญิงหรือผู้ชายจะได้เงินเดือนมากกว่าคนที่หน้าตาธรรมดา และ (2) ผู้ชายที่หน้าตาดีจะได้เงินเดือนมากกว่าคนอื่น ๆ (ถ้าปัจจัยอื่น ๆ เหมือนกัน เช่น ตำแหน่งงานที่ทำ จำนวนปีของประสบการณ์การทำงาน สถานที่ทำงาน)

ให้ Y = เงินเดือน (หน่วย: บาท/เดือน)

X_1 = ตัวแปรบังคับตัวที่ 1

X_2 = ตัวแปรบังคับตัวที่ 2

X_3 = ตัวแปรบังคับตัวที่ 3

D_1 = ตัวแปร dummy มีค่า = 1 ถ้าคนนั้นมีหน้าตาดี, = 0 ถ้าหน้าตาธรรมดา

D_2 = ตัวแปร dummy มีค่า = 1 ถ้าคนนั้นเป็นเพศชาย, = 0 ถ้าเป็นเพศหญิง

คำถาม

5.1) สมมุติว่าคุณต้องการทดสอบข้อสรุปผลงานวิจัยในต่างประเทศข้อที่ 1 ว่า “คนที่หน้าตาดีไม่ว่าผู้หญิงหรือผู้ชายจะได้เงินเดือนมากกว่าคนที่หน้าตาธรรมดา” หรือไม่โดยใช้ข้อมูลของคนไทย ให้เขียนโมเดลสมการ regression ที่คุณจะใช้ทดสอบ และถ้าข้อสรุปสำหรับประเทศไทยเหมือนกับข้อสรุปจากผลงานวิจัยในต่างประเทศ สัมประสิทธิ์หน้าตัวแปร Dummy ที่ใช้ต้องมากกว่าศูนย์หรือน้อยกว่าศูนย์และ/หรือ มีนัยสำคัญทางสถิติหรือไม่อย่างไร (บังคับว่าในโมเดลสมการ regression ของคุณจะต้องมีตัวแปร X_1, X_2, และ X_3 อยู่ด้วย)

5.2) สมมุติว่าคุณต้องการทดสอบข้อสรุปผลงานวิจัยในต่างประเทศข้อที่ 2 ว่า “ผู้ชายที่หน้าตาดีจะได้เงินเดือนมากกว่าคนอื่น ๆ” หรือไม่โดยใช้ข้อมูลของคนไทย ให้เขียนโมเดลสมการ regression ที่คุณจะใช้ทดสอบ และถ้าข้อสรุปสำหรับประเทศไทยเหมือนกับข้อสรุปจากผลงานวิจัยในต่างประเทศ สัมประสิทธิ์หน้าตัวแปร Dummy ที่ใช้ต้องมากกว่าศูนย์หรือน้อยกว่าศูนย์ และ/หรือ มีนัยสำคัญทางสถิติหรือไม่อย่างไร (บังคับว่าในโมเดลสมการ regression ของคุณจะต้องมีตัวแปร X_1, X_2, และ X_3 อยู่ด้วย)

<u>แนวทางการตอบ</u>

5.1) การทดสอบข้อสรุปผลงานวิจัยในต่างประเทศข้อที่ 1 ว่า “คนที่หน้าตาดีไม่ว่าผู้หญิงหรือผู้ชายจะได้เงินเดือนมากกว่าคนที่หน้าตาธรรมดา” เราจะรัน Y C X_1 X_2 X_3 D_1 โดยถ้าข้อสรุปสำหรับประเทศไทยเหมือนกับข้อสรุปจากผลงานวิจัยในต่างประเทศ สัมประสิทธิ์หน้าตัวแปร D_1 ต้องมากกว่าศูนย์และมีนัยสำคัญทางสถิติ

5.2) การทดสอบข้อสรุปผลงานวิจัยในต่างประเทศข้อที่ 2 ว่า “ผู้ชายที่หน้าตาดีจะได้เงินเดือนมากกว่าคนอื่น ๆ” เราจะรัน Y C X_1 X_2 X_3 D_1*D_2 โดยถ้าข้อสรุปสำหรับประเทศไทยเหมือนกับข้อสรุปจากผลงานวิจัยในต่างประเทศ สัมประสิทธิ์หน้าตัวแปร D_1*D_2 ต้องมากกว่าศูนย์และมีนัยสำคัญทางสถิติ (ถ้าตอบว่าจะรัน Y C X_1 X_2 X_3 D_1 D_2 ก็พออนุโลม)