ข้อสอบกลางภาค

วิชา Research Methodology in Finance : MGMG 522

ภาคเรียนที่ 2 ปี 2547

ผู้สอน ดร.ชาญ สรณาคมน์

(คะแนนสอบกลางภาค = 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. ปัญหา Serial Correlation และ Heteroskedasticity มีความเหมือนและความแตกต่างกันด้านใดบ้าง?

แนวทางการตอบ

ความเหมือน

ปัญหา Serial Correlation และ Heteroskedasticity จะมีผลลัพธ์ที่ตามมาคล้าย ๆ กัน คือ Coefficient Estimates ที่ได้จาก OLS ถึงแม้จะไม่ bias แต่ว่า Standard Errors ของ Coefficient Estimates จะมีค่าเพิ่มขึ้นแต่ OLS จะรายงานตัวเลขค่า Standard Errors ของ Coefficient Estimates ต่ำกว่าความเป็นจริง ส่งผลไปถึงค่า t ที่ได้จะสูงกว่าความเป็นจริง

ความแตกต่าง

1. Serial Correlation มักเกิดในข้อมูล Time-series ในขณะที่ Heteroskedasticity มักเกิดกับข้อมูลที่เป็น Cross-sectional Data
2. ปัญหา incorrect functional form อาจทำให้เกิด impure serial correlation ได้มากกว่าทำให้เกิดปัญหา impure heteroskedasticity

ข้อที่ 2. จงเปรียบเทียบ overall fit ของสมการ regression โมเดล A กับโมเดล B ว่าสมการใดดีกว่ากัน? ถ้าบอกไม่ได้ว่าโมเดลไหนดีกว่ากัน ให้บอกเหตุผลด้วยว่าทำไมถึงบอกไม่ได้? (ข้อนี้ คุณไม่จำเป็นต้องทราบคำอธิบายว่าตัวแปร Y, X_1, X_2, และ X_3 หมายถึงตัวแปรอะไร? ตัวเลขในวงเล็บแสดงค่าสถิติ t)

Model A: $Y = 0.75 + 1.25X_1 - 35.0X_2 + 0.05X_3$

(1.98) (-2.05) (1.45)

$Adj\text{-}R^2 = 0.792$, $n = 100$, $F = 38.1$

Model B: $\ln Y = 0.03 + 0.12X_1 - 3.15X_2 + 0.01X_3$

(2.49) (-2.14) (1.89)

$Adj\text{-}R^2 = 0.801$, $n = 100$, $F = 38.8$

<u>แนวทางการตอบ</u>

เนื่องจากตัวแปรตามเป็นคนละตัวกันใน 2 โมเดล (ตัวแปรตามในโมเดล A คือ Y ในขณะที่ตัวแปรตามในโมเดล B คือ lnY) การที่จะนำเอาค่า R^2 (หรือ Adj-R^2) มาเปรียบเทียบกันไม่ควรทำ เพราะว่า RSS (Residual Sum of Squares) เป็นคนละตัวกัน กล่าวคือ

ในโมเดล A

$$RSS = \sum \left(Y_i - \hat{Y}_i\right)^2$$

ในขณะที่โมเดล B

$$RSS = \sum \left(\ln Y_i - \ln \hat{Y}_i\right)^2$$

ดังนั้น ในการเปรียบเทียบต้องใช้ค่า quasi-R^2 โดยต้องมีการแปลงค่าบางค่าในโมเดล B เพื่อคำนวณหาค่า quasi-R^2 ซึ่ง

$$\text{quasi} - R^2 = 1 - \frac{\sum [Y_i - \text{antilog}(\ln \hat{Y}_i)]^2}{\sum [Y_i - \overline{Y}]^2}$$

แล้วนำค่า quasi-R^2 ที่คำนวนได้ของโมเดล B ไปเปรียบเทียบกับค่า R^2 ของโมเดล A

ข้อที่ 3. จงอธิบายว่าทำไมการเลือกว่าจะเอาตัวแปรต้นตัวใดไว้ในสมการ regression หรือตัดตัวแปรตัวนั้นออกโดยพิจารณาจากค่า t-value ของตัวแปรนั้น ๆ เพียงอย่างเดียวว่า significant หรือไม่จึง<u>ไม่ควรทำ</u>? <u>อธิบายโดยละเอียด</u>

<u>แนวทางการตอบ</u>

การที่ค่า t มีค่ามาก ๆ มักไม่ค่อยมีปัญหา แต่ถ้าค่า t มีค่าใกล้ ๆ ค่า critical t-value การตัดสินใจจะตัดตัวแปรนั้นทิ้งหรือคงตัวแปรนั้นไว้จะทำได้ลำบาก ยกตัวอย่างเช่น ถ้าเราได้ค่า t = 1.7 เราอาจจะตัดตัวแปรนี้ทิ้งเพราะว่ามันไม่มีนัยสำคัญที่ระดับ 1 เปอร์เซ็นต์ แต่ถ้าสมการเรามีปัญหาตัวแปรขาดอยู่ เป็นไปได้ที่ค่า t = 1.7 นั้นอาจจะน้อยกว่าความเป็นจริงเพราะผลของ bias ใน coefficient หรือเป็นผลมาจาก multicollinearity หรือถ้าค่า t = 1.7 นั้นมีนัยสำคัญที่ 5 เปอร์เซ็นต์ก็อาจจะเป็นไปได้ที่ว่าสมการเรายังมีปัญหา serial correlation หรือ heteroskedasticity เลยทำให้ค่า t มีค่าเกินค่า t ที่แท้จริง

สรุป การเลือกตัวแปรโดยพิจารณาจากค่า t ของ coefficient เพียงอย่างเดียวไม่ควรทำ

ข้อที่ 4. นี่เป็นผลจากการศึกษาจำนวนผู้เข้าชมทีมบาสเก็ตบอล Los Angeles Lakers โดยตัวแปรตาม Y_t คือ จำนวนผู้เข้าชม (มีหน่วยเป็นคน) ณ ครั้งที่ t และตัวแปรต้นมีดังนี้

L_t = เปอร์เซ็นต์การชนะของทีม Los Angeles Lakers (จำนวนครั้งที่ทีมชนะทั้งหมดก่อนครั้งที่ t หารด้วยจำนวนครั้งที่เล่นทั้งหมด*100%)

P_t = เปอร์เซ็นต์การชนะของทีมคู่แข่งที่ Los Angeles Lakers จะต้องเล่นด้วยในครั้งที่ t (จำนวนครั้งที่ทีมคู่แข่งชนะก่อนครั้งที่ t หารด้วยจำนวนครั้งที่เล่นทั้งหมด*100%)

W_t = 1 ถ้าวันที่แข่งขันกันเป็นวันศุกร์ เสาร์ หรืออาทิตย์ (W_t = 0 สำหรับวันอื่น ๆ ที่ไม่ใช่วันหยุดสุดสัปดาห์)

ผลการรันสมการ regression ออกมาเป็นดังนี้ (ตัวเลขในวงเล็บแสดงค่า Standard Error)

$$Y_t = 14{,}123 + 20L_t + 2{,}600P_t + 900W_t$$
$$\qquad\qquad (500) \quad (1{,}000) \quad (300)$$

DW = 0.85, n = 44, Adj-R^2 = 0.46

A.) ให้ทดสอบสมมุติฐานด้านเดียวของ coefficient หน้าตัวแปรแต่ละตัว ณ ระดับนัยสำคัญ 1 เปอร์เซ็นต์

B.) ให้ทดสอบหา Serial Correlation ด้านเดียว ณ ระดับนัยสำคัญ 5 เปอร์เซ็นต์

C.) พิจารณาตัวแปร L_t กับ P_t คุณแปลกใจกับผลการทดสอบสมมุติฐานที่ทำในข้อ A. หรือไม่ พูดอีกนัยหนึ่ง ตัวแปร L_t เป็นตัวแปรเกิน (irrelevant variable) <u>หรือไม่?</u> <u>อธิบายเหตุผลประกอบ</u>

<u>แนวทางการตอบ</u>

A.) coefficient หน้าตัวแปรทุกตัวน่าจะมีเครื่องหมายเป็นบวก ดังนั้นสมมุติฐานจะเป็น

$H_0: \beta_i \leq 0$

$H_A: \beta_i > 0$

Critical t (one-sided, 1%, DF = 44-3-1 = 40) = 2.423

ค่า t หน้าตัวแปร L_t = 20/500 = 0.04 → ไม่ significant

ค่า t หน้าตัวแปร P_t = 2,600/1,000 = 2.60 → significant

ค่า t หน้าตัวแปร W_t = 900/300 = 3.00 → significant

B.) สำหรับ test for serial correlation ใช้ค่าสถิติ DW โดยที่สมมุติฐานสำหรับการทดสอบจะเป็น

H_0: ไม่มี serial correlation ด้านบวก

H_A: มี serial correlation ด้านบวก

ค่า Critical DW (one-sided, 5%, k'=4, n ≈ 45) → D_L = 1.34, D_U = 1.72

ค่า DW ที่ได้คือ 0.85 < D_L → significant (แปลว่ามี serial correlation ด้านบวก)

C.) ตัวแปร L_t มีค่า t ต่ำ แต่ทางทฤษฎีแล้วตัวแปร L_t ไม่น่าจะเป็นตัวแปรเกิน เพราะว่าการที่ทีม Los Angeles Lakers ชนะบ่อย ๆ น่าจะมีผลทำให้คนเข้าชมทีม Los Angeles Lakers แข่งขันมากขึ้น สามารถที่เราเจอว่าค่า t หน้าตัวแปร L_t มีค่าต่ำอาจเป็นเพราะว่า (1) มีตัวแปรขาดที่ยังไม่ได้ใส่ในโมเดล หรือ (2) ตัวแปร L_t กับ P_t มีปัญหา Multicollinearity กันเลยทำให้ค่า t ของตัวแปรบางตัวลดลง