หอสมุดกลาง สถาบันวิทยบริการ จุฬาลงกรณ์มหาวิทยาลัย

บทที่ 1

บทนำ

1.1 ความสำคัญและความเป็นมาของปัญหา

ในงานวิจัยต่าง ๆ ที่ต้องอาศัยเทคนิคการวิเคราะห์ความถดถอยเชิงเส้น Linear Regression Analysis) นั้นเป็นที่ทราบกันดีอยู่แล้วว่าเป้าหมายที่สำคัญของการวิเคราะห์ความถดถอยนั้นคือการพยายามหาหนทางประมาณค่าพารามิเตอร์ของสมการถดถอยโดยวิธีใดวิธีหนึ่งที่เป็นวิธีเหมาะสมที่สุด ซึ่งในการเลือกใช้วิธีที่เหมาะสมกับลักษณะของข้อมูลที่ต้องการวิเคราะห์นั้นจำเป็นจะต้องคำนึงถึงข้อตกลงเบื้องต้นของวิธีแต่ละวิธีที่ใช้ด้วย

สำหรับการศึกษาการวิเคราะห์ความถดถอยเชิงเส้นอย่างง่าย ที่มีรูปแบบทั่วไปเป็นดังนี้คือ $Y_t = \beta_o + \beta_1 X_t + U_t$, $t = 1,2,...,n$ โดยที่ X_t เป็นตัวแปรอิสระที่มีค่าคงที่ Y_t เป็นตัวแปรตาม β_o และ β_1 เป็นพารามิเตอร์ที่ไม่ทราบค่าแทนจุดตัดบนแกน Y และความชันของเส้นถดถอยตามลำดับ U_t เป็นความคลาดเคลื่อน และ n เป็นขนาดตัวอย่าง โดยปกติการประมาณค่าและการทดสอบสมมติฐานเกี่ยวกับพารามิเตอร์นั้น ผู้วิจัยมักเลือกใช้วิธีกำลังสองน้อยที่สุด (Ordinary Least Square : OLS) ซึ่งเป็นวิธีที่ให้ตัวประมาณที่มีคุณสมบัติเป็น BLUE* (Best Linear Unbiased Estimator) และในการทดสอบสมมติฐานก็ให้สถิติทดสอบ (test statistic) ที่มีอำนาจของการทดสอบ (Power of the test) สูง ทั้งนี้จะต้องมีข้อตกลงเบื้องต้นเกี่ยวกับความคลาดเคลื่อนดังนี้คือ ความคลาดเคลื่อนจะต้องมีค่าเฉลี่ยเป็น 0 ค่าความแปรปรวนเป็น σ^2 และ U_i , U_j ไม่มีสหสัมพันธ์ต่อกันเมื่อ $i \neq j$ อย่างไรก็ตามในทางปฏิบัติบ่อยครั้งที่เราพบว่าข้อมูลที่ถูกนำมาใช้วิเคราะห์นั้นมีอยู่ไม่น้อยที่ไม่เป็นไปตามข้อตกลงดังกล่าว และในการวิจัยโดยทั่ว ๆ ไปปัญหาที่ผู้วิจัยพบมากก็คือเหตุการณ์ที่

---

*ทฤษฎี Gauss-Markow ในกลุ่มของ Linear Unbiased Estimator ของ Regression parameters นั้น OLS-Estimator จะเป็นตัวประมาณที่ดีที่สุด (Chow 1983:40-41, Wonnacott and Wonnacott 1970:27-28)

ศูนย์วิทยทรัพยากร จุฬาลงกรณ์มหาวิทยาลัย

ความคลาดเคลื่อนมีความสัมพันธ์ต่อกันกล่าวคือ $E(U_i U_j) \neq 0$ เมื่อ $i \neq j$ ซึ่งสถานการณ์เช่นนี้เราเรียกว่าอัตตสหสัมพันธ์ (Autocorrelation) ดังนั้นถ้าหากข้อมูลที่นำมาศึกษาเกิดปัญหาดังกล่าวทางเลือกที่เป็นไปได้สำหรับวิธีประมาณค่าพารามิเตอร์คือ

1. ใช้วิธีกำลังสองน้อยที่สุดโดยที่ข้อมูลยังคงมีคุณสมบัติที่ไม่เป็นไปตามข้อตกลงเบื้องต้น

2. ใช้วิธีการอื่นที่เหมาะสมกับข้อมูลที่มีคุณสมบัติไม่เป็นไปตามข้อตกลงเบื้องต้น

สำหรับทางเลือกที่หนึ่งนั้นเป็นทางเลือกที่ไม่สู้จะดีนัก เนื่องจากข้อมูลที่เกิดปัญหาอัตตสหสัมพันธ์ขึ้นนั้นมีผลกระทบโดยตรงต่อวิธีนี้ ถ้าเรายังใช้วิธีการนี้อยู่ด้วยเหตุใดเหตุหนึ่งก็ตาม ย่อมจะก่อให้เกิดผลเสียหายโดยเฉพาะในแง่คุณภาพของตัวประมาณพารามิเตอร์ ($\hat{\beta}$) ดังนี้

1. $\hat{\beta}$ ยังเป็นตัวประมาณค่าที่ไม่เอนเอียง (Unbiased Estimator) แต่โดยทั่วไปจะมีประสิทธิภาพ (Efficiency) ต่ำกว่าวิธีอื่น ๆ

2. ถ้าเกิดปัญหาอัตตสหสัมพันธ์ขึ้นโดยเฉพาะอย่างยิ่ง อัตตสหสัมพันธ์ทางบวก[1] (Positive Autocorrelation) ค่าประมาณ σ^2 จะมีค่าต่ำกว่าความเป็นจริง ยังผลให้ $\hat{v}(\hat{\beta})$ ต่ำกว่าความเป็นจริงไปด้วยเนื่องจากในการประมาณค่า β โดยอาศัยวิธีกำลังสองน้อยที่สุดนั้น

$$\hat{v}(\hat{\beta}) = \sigma^2 (X'X)^{-1}$$

จากค่า $\hat{V}(\hat{\beta})$ ที่ต่ำ ๆ ในกรณีนี้จึงเป็นเลขลวงตา หากนักวิจัยไม่ระมัดระวังและมีความเข้าใจในเรื่องเหล่านี้ อาจเข้าใจผิดคิดว่างานของตนมีคุณภาพสูงทั้ง ๆ ที่ความจริงมิได้เป็นเช่นนั้น จากผลกระทบเหล่านี้แสดงให้เห็นว่าผลสรุปที่ได้จากการประมาณค่าพารามิเตอร์ในทางเลือกที่หนึ่งจะบิดเบือนหรือผิดเพี้ยนไปจากความเป็นจริง

สำหรับทางเลือกที่สองจะเป็นทางเลือกที่น่าสนใจในการนำมาใช้แก้ปัญหาข้างต้นกล่าวคือ ในกรณีที่ข้อมูลที่มีอยู่นั้นมีปัญหาอัตตสหสัมพันธ์ การใช้วิธีการประมาณค่าพารามิเตอร์ที่

---

[1]มนตรี พิริยะกุล, เทคนิคการวิเคราะห์สมการถดถอย เล่ม 2 (โรงพิมพ์มหาวิทยาลัยรามคำแหง, กรุงเทพมหานคร, 2526), หน้า 55.

เหมาะสมมาทำการประมาณค่าแทนวิธีกำลังสองน้อยที่สุดนั้นจะทำให้การวิเคราะห์และการแปลผลมีความถูกต้องมากขึ้น ซึ่งวิธีการที่เหมาะสมก็มีหลายวิธีอาทิเช่นวิธี GLS, WLS เป็นต้น

จากผลกระทบที่เกิดจากปัญหาอัตตสหสัมพันธ์นั้น นับว่าจำเป็นอย่างยิ่งที่จะต้องทำการตรวจสอบข้อมูลที่มีอยู่ว่าเป็นไปตามข้อตกลงเบื้องต้นดังกล่าวหรือไม่ เพราะถ้าตรวจพบว่าเกิดความขัดแย้งกับข้อตกลงก็จะได้แก้ไขให้ถูกต้อง โดยการหาวิธีที่เหมาะสมกว่าแทนการใช้วิธีกำลังสองน้อยที่สุดเพื่อทำให้งานวิจัยมีคุณภาพสูง สถิติทดสอบอัตตสหสัมพันธ์นั้นมีหลายตัวด้วยกันเช่นการทดสอบเดอร์บิน-วัตสัน (Durbin-Watson test:DW) การทดสอบเบเรนบรูตและเวบบ์ (Berenblut-Webb test : G) การทดสอบเกียรี่ (Geary test) เป็นต้น ดังนั้นจึงเป็นสิ่งที่น่าจะศึกษาว่าสถิติทดสอบใดที่เหมาะสมสำหรับทดสอบอัตตสหสัมพันธ์ทางบวก ตลอดจนแนวทางในการเลือกใช้สถิติทดสอบที่เหมาะสมกับลักษณะการแจกแจงของความคลาดเคลื่อนแบบปกติและแบบสมมาตรหางยาว ซึ่งคุณสมบัติที่จะชี้ให้เห็นว่าสถิติทดสอบนั้นมีความเหมาะสมหรือไม่ก็คือความแกร่ง (Robussness) และอำนาจของการทดสอบ (Power of the test) นั่นเอง โดยที่สถิติทดสอบที่มีความแกร่งนั้นจะต้องไม่ทำให้ความน่าจะเป็นของความคลาดเคลื่อนประเภทที่ 1 (type I error) หรือความคลาดเคลื่อนประเภทที่ 2 (type II error) มากขึ้น ถ้าข้อตกลงเบื้องต้นเกิดผิดพลาดไป (Sawat Pratoomraj 1970:1) และด้วยเหตุว่าการคำนวณค่าความน่าจะเป็นของความคลาดเคลื่อนประเภทที่ 1 และอำนาจของการทดสอบเมื่อลักษณะข้อมูลและการแจกแจงของความคลาดเคลื่อนเป็นแบบต่าง ๆ นั้นกระทำได้ยาก ดังนั้นผู้วิจัยจึงเลือกศึกษาโดยวิธีมอนติคาร์โล (Monte Carlo method) ซึ่งเป็นวิธีที่ศึกษาในรูปของการจำลอง (Simulation) โดยอาศัยตัวเลขสุ่มเทียม (Psudo random number) และสามารถระบุขนาดตัวอย่าง ความแปรปรวน ค่าเฉลี่ย และลักษณะการแจกแจงของความคลาดเคลื่อนได้ตามที่ผู้วิจัยต้องการ

1.2 วัตถุประสงค์ของการวิจัย

ในการวิจัยครั้งนี้ต้องการศึกษาเปรียบเทียบการทดสอบที่ใช้ทดสอบอัตตสหสัมพันธ์ทางบวกของความคลาดเคลื่อนในการวิเคราะห์ความถดถอยเชิงเส้นอย่างง่าย 3 วิธีคือ

1. การทดสอบเดอร์บินและวัตสัน (DW)
2. การทดสอบเบเรนบรูตและเวบบ์ (G)
3. การทดสอบเกียรี่ (Geary)

ศูนย์วิทยทรัพยากร
จุฬาลงกรณ์มหาวิทยาลัย

เพื่อหาข้อสรุปเกี่ยวกับความเหมาะสมของการทดสอบดังกล่าวภายใต้ลักษณะการแจกแจงของความคลาดเคลื่อนแบบปกติ (Normal Distribution) และแบบโลจิสติค (Logistic Distribution) แบบดับเบิ้ลเอ็กซ์โปเนนเชียล (Double Exponential Distribution) ซึ่งเป็นการที่มีลักษณะสมมาตรหางยาว (Long tail distribution, Heavy tail distribution) โดยพิจารณาจากอำนาจการทดสอบ (1 - β) และความน่าจะเป็นของความคลาดเคลื่อนประเภทที่ 1 (type I error)

1.3 สมมติฐานของการวิจัย

ภายใต้ลักษณะการแจกแจงของความคลาดเคลื่อนที่ไม่ได้มีการแจกแจงแบบปกตินั้นการทดสอบเบีเรนบรูตและเวบบ์จะให้ค่าอำนาจของการทดสอบสูงสุด

1.4 ข้อตกลงเบื้องต้น

สมการถดถอยที่ใช้ในการศึกษาเป็นแบบสมการถดถอยอย่างง่ายโดยมีรูปแบบเป็น

$$Y_t = \beta_o + \beta_1 X_t + U_t \quad , \quad t = 1,\dots,n$$

X_t คือตัวแปรอิสระที่มีค่าคงที่

Y_t คือตัวแปรตาม

β_o, β_1 คือพารามิเตอร์

U_t คือความคลาดเคลื่อน

กำหนดให้รูปแบบของอัตตสหสัมพันธ์เป็นแบบ AR(1)[1]

$$U_t = \rho U_{t-1} + V_t$$

ρ คือ สหสัมพันธ์ระหว่าง U_t กับ U_{t-1}

---

[1]มนตรี พิริยะกุล, เทคนิคการวิเคราะห์สมการถดถอยเล่ม 2, (โรงพิมพ์มหาวิทยาลัยรามคำแหง, กรุงเทพมหานคร, 2526), หน้า 46-52

การวิจัยครั้งนี้ถือว่าความแกร่งและอำนาจของการทดสอบเป็นดัชนีที่ผู้วิจัยจะใช้เป็นเกณฑ์ในการเลือกการทดสอบที่เหมาะสมในการนำไปใช้ทดสอบอัตตสหสัมพันธ์ในแต่ละการแจกแจงของความคลาดเคลื่อนที่กำหนดในวัตถุประสงค์ของการวิจัย

1.5 <u>ขอบเขตของการวิจัย</u>

1. ศึกษาภายใต้ลักษณะการแจกแจงของความคลาดเคลื่อนที่มีการแจกแจงแบบปกติ (Normal Distribution) โลจิสติค (Logistic Distribution) และดับเบิ้ลเอ็กซ์โปเนนเชียล (Double Exponential Distribution)

2. กำหนดให้ $\underset{\sim}{\beta}' = (10 , 1.5)$ ในประชากรทุกรูปแบบที่ศึกษา*

3. สร้างค่า X ที่เป็นค่าคงที่ 2 รูปแบบคือ

3.1 กำหนดให้

3.2 เลือกโดยสุ่มจากประชากรที่มีการแจกแจงปกติมีค่าเฉลี่ยเป็น 200 และค่าความแปรปรวนเป็น 100*

4. จำนวนข้อมูลที่ศึกษามีขนาดเป็น 15 30 และ 50

5. ค่า ρ ที่ศึกษาเป็น 0.0 0.1 0.3 0.5 0.7 และ 0.9

6. ระดับนัยสำคัญที่ศึกษาคือ 0.01 และ 0.05

7. ความคลาดเคลื่อนของการแจกแจง มีค่าเฉลี่ยเป็น 0 และค่าความแปรปรวนเป็น 1**

---

* การวิจัยครั้งนี้กำหนดค่าพารามิเตอร์เป็นค่าดังกล่าว เนื่องจากได้ทดลองกระทำ ณ ค่าต่าง ๆ แล้วปรากฏว่าไม่ว่าจะเป็นค่าใดก็ตาม ผลสรุปไม่เปลี่ยนแปลง

** การวิจัยครั้งนี้กำหนดความแปรปรวนเท่ากับ 1 เนื่องจากได้ทดลองกระทำ ณ ค่าความแปรปรวนอื่น ๆ แล้วปรากฏว่าไม่ว่าจะเป็นเท่าใดก็ตามผลสรุปไม่เปลี่ยนแปลง

ศูนย์วิทยทรัพยากร
จุฬาลงกรณ์มหาวิทยาลัย
หอสมุดกลาง สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

8. การวิจัยครั้งนี้ทำการทดลองด้วยเครื่องคอมพิวเตอร์ IBM 370/3010 ด้วยเทคนิคมอนติคาร์โล โดยใช้โปรแกรมที่เขียนด้วยภาษาฟอร์แทรน (FORETRAN) การทดสอบกระทำซ้ำกัน 1,000 ครั้ง ในแต่ละสถานการณ์ของการทดลอง

1.6 คำจำกัดความ

อัตตสหสัมพันธ์ (Autocorrelation) หมายถึงเหตุการณ์ที่ตัวแปรสุ่ม U มีความสัมพันธ์ต่อกันหรือมี Covariation ระหว่าง U_i กับ U_j กล่าวคือ $E(U_iU_j) \neq 0$ เมื่อ $i \neq j$

ความแกร่ง (Robustness) ของการทดสอบ หมายถึงคุณสมบัติของการทดสอบที่ไม่แสดงความไวต่อการเปลี่ยนแปลงของปัจจัยอื่นที่ไม่ใช่ปัจจัยที่ต้องการทดสอบ เช่นการเบี่ยง-เบนไปจากข้อตกลงเบื้องต้นของการทดสอบอันจะมีผลต่อความน่าจะเป็นของความคลาดเคลื่อนประเภทที่ 1 และความน่าจะเป็นของความคลาดเคลื่อนประเภทที่ 2

อำนาจของการทดสอบ (Power of the test) หมายถึงความน่าจะเป็น (Probability) ที่จะปฏิเสธสมมติฐานว่าง (Null Hypothesis) เมื่อสมมติฐานว่างนั้นผิดซึ่งจะมีค่าเท่ากับ $1 - \beta$ เมื่อ β คือความน่าจะเป็นของความคลาดเคลื่อนประเภทที่ 2 (Kirk 1969:555, Minium 1978:364)

ความคลาดเคลื่อนประเภทที่ 1 (Type I error) เป็นความผิดพลาดที่เกิดจากการปฏิบัติสมมติฐานว่าง เมื่อสมมติฐานว่างนั้นถูก

ความคลาดเคลื่อนประเภทที่ 2 (Type II error) เป็นความผิดพลาดที่เกิดจากการไม่ปฏิเสธสมมติฐานว่าง เมื่อสมมติฐานว่างนั้นผิด

ศูนย์วิทยทรัพยากร
จุฬาลงกรณ์มหาวิทยาลัย

1.7 ประโยชน์ของการวิจัย

1. เป็นแนวทางในการตัดสินใจว่าควรเลือกใช้สถิติทดสอบใดจึงจะมีอำนาจของการทดสอบสูงสุดในการทดสอบอัตตสหสัมพันธ์ในการวิเคราะห์ความถดถอยเชิงเส้นอย่างง่ายตลอดจนนำไปใช้ได้เหมาะสมกับลักษณะการแจกแจงของความคลาดเคลื่อนทั้งแบบปกติและแบบสมมาตรหางยาว

2. เป็นแนวทางในการวิจัยเพื่อเปรียบเทียบตัวสถิติอื่น ๆ ต่อไป