หอสมุดกลาง สถาบันวิทยบริการ จุฬาลงกรณ์มหาวิทยาลัย

บทที่ 1

บทนำ

1.1 ความสำคัญและความเป็นมาของปัญหา

การวิเคราะห์การถดถอยเชิงเส้น (Linear Regression Analysis) เป็นการศึกษาถึงความสัมพันธ์ของตัวแปร 2 กลุ่ม กลุ่มหนึ่งเรียกว่าตัวแปรตาม (Dependent variable) ถูกกำหนดขึ้นโดยตัวแปรอีกกลุ่มหนึ่งที่เรียกว่า ตัวแปรอิสระ(Independent variable) ซึ่งอาจมีตัวเดียวหรือหลายตัว และความสัมพันธ์ระหว่างตัวแปรทั้งสองกลุ่มนี้ มีลักษณะความสัมพันธ์เชิงเส้น ซึ่งมีรูปแบบทั่วไปเป็นดังนี้

$$y_t = \beta_0 + \beta_1 x_{1t} + \beta_2 x_{2t} + \ldots + \beta_p x_p + u_t, t=1,2,\ldots,n$$

โดยที่

y_t คือ ตัวแปรตาม

x_{it} คือ ตัวแปรอิสระ ,i=1,2,...,p

β_i คือ พารามิเตอร์ (Parameter) มีชื่อเรียกว่า สัมประสิทธิ์การถดถอย (Regression Coefficients),i=0,1,...,p

u_t คือ ความคลาดเคลื่อนสุ่ม หรือค่าผิดพลาดสุ่ม (Random errors)

n คือ ขนาดตัวอย่าง

สำหรับงานวิจัยโดยทั่วไปที่ต้องอาศัยเทคนิคการวิเคราะห์การถดถอยนั้น เป็นที่ทราบกันดีอยู่แล้วว่าเมื่อเลือกตัวแบบสมการถดถอยตัวแบบใดเป็นตัวแบบทดลองได้แล้ว ขั้นตอนที่สำคัญต่อไปในการวิเคราะห์การถดถอย คือการประมาณค่าพารามิเตอร์ หรือสัมประสิทธิ์

การถดถอย β_i ด้วย $\hat{\beta}_i$ ($i=0,1,\ldots,p$) โดยวิธีใดวิธีหนึ่งที่เหมาะสม ซึ่งวิธีประมาณค่าพารามิเตอร์ที่นิยมใช้กันอย่างแพร่หลายวิธีหนึ่งคือ วิธีกำลังสองน้อยที่สุดแบบสามัญ (Ordinary Least Square Method หรือเรียกย่อว่า วิธี OLS) ซึ่งวิธีนี้ให้ตัวประมาณที่มีคุณสมบัติเป็นตัวประมาณที่ไม่เอนเอียงเชิงเส้นที่ดีที่สุด (Best Linear Unbiased Estimator : BLUE) ตามทฤษฎีเกาส์-มาร์โคฟ (Gauss- Markov Theorem) ทั้งนี้จะต้องมีข้อตกลงเบื้องต้น (Assumptions) ของความคลาดเคลื่อนสุ่ม (u_t) ดังนี้

1. u_t มีค่าคาดหวัง (Expected Value) เป็น 0 ;

   $E(u_t) = 0$

2. ความแปรปรวนของความคลาดเคลื่อนคงที่ ;

   $\sigma^2_{u_t} = \sigma^2$

3. u_i และ u_j ไม่มีสหสัมพันธ์หรือความแปรปรวนร่วม (Covariance);

   $E(u_i u_j) = 0, \quad i \neq j$

ข้อตกลงเบื้องต้นดังกล่าวนี้มีความจำเป็นอย่างยิ่งในการหาตัวประมาณ $\hat{\beta}_i$ ที่เป็น BLUE และในการวิจัยครั้งนี้ ผู้วิจัยสนใจข้อตกลงเบื้องต้นข้อที่ 3 ซึ่งเป็นข้อตกลงเบื้องต้นข้อหนึ่งที่สามารถส่งผลให้เกิดความผิดพลาดร้ายแรงได้ในการอนุมาน ถ้าหากยังคงใช้วิธี OLS ในการประมาณค่าพารามิเตอร์ โดยที่ตัวแบบของการถดถอยไม่เป็นไปตามคุณสมบัติข้อนี้

การที่ความคลาดเคลื่อนมีสหสัมพันธ์กัน หรือเรียกอีกอย่างหนึ่งว่า ความคลาดเคลื่อนมีอัตสหสัมพันธ์กัน (Autocorrelation) นั้น อาจเกิดด้วยสาเหตุหลายประการ เช่น การละเลยตัวแปรอิสระที่สำคัญๆ ซึ่งตัวแปรนั้นมีความสัมพันธ์กับเวลาเป็นต้น

ถ้าเกิดปัญหาของความคลาดเคลื่อนมีอัตตสหสัมพันธ์กัน และยังคงใช้วิธี OLS ในการประมาณค่าพารามิเตอร์ β_1 จะได้ตัวประมาณที่ไม่เป็น BLUE (มีความแปรปรวนไม่ต่ำสุด) ถึงแม้จะเป็นตัวประมาณที่ไม่เอนเอียงก็ตาม จากผลนี้สามารถทำให้การอนุมานผิดพลาดได้ ตัวอย่างเช่น ถ้าค่าความคลาดเคลื่อนมีอัตตสหสัมพันธ์ทางบวก (Positive Autocorrelation) จะมีผลทำให้ค่าเฉลี่ยของความคลาดเคลื่อนกำลังสอง (Mean Square Error) มีค่าต่ำกว่าค่าจริงมาก เพราะฉนั้นการทดสอบนัยสำคัญของตัวประมาณอาจปรากฏว่าต่างจากศูนย์ ทั้งที่ความจริงแล้วไม่ต่างจากศูนย์ ดังนั้นจึงควรตรวจสอบค่าความคลาดเคลื่อนว่าเป็นไปตามข้อตกลงเบื้องต้นหรือไม่ เมื่อข้อมูลเป็นข้อมูลอนุกรมเวลา ก่อนที่จะเลือกวิธีประมาณค่าพารามิเตอร์

รูปแบบอัตตสหสัมพันธ์ของความคลาดเคลื่อนที่ผู้วิจัยสนใจในการวิจัยครั้งนี้ คือ อัตตสหสัมพันธ์ตำแหน่งที่ 1 AR(1) (First-Order Autoregressive Model) ในสมการถดถอยเชิงเส้นอย่างง่าย (Simple Linear Regression Equation) ที่มีรูปแบบทั่วไปดังนี้

$$y_t = \beta_0 + \beta_1 x_t + u_t \quad ,t = 1,2,\ldots,n$$

โดยที่

u_t คือ ความคลาดเคลื่อนสุ่ม ซึ่งกำหนดรูปแบบอัตตสหสัมพันธ์เป็น AR(1) ดังนี้

$$u_t = \rho u_{t-1} + v_t \quad ,t = 1,2,\ldots,n$$

เมื่อ ρ คือ สัมประสิทธิ์อัตตสหสัมพันธ์ตำแหน่งที่ 1, $-1 \leq \rho \leq 1$

v_t คือ ความคลาดเคลื่อนสุ่ม

y_t คือ ตัวแปรตาม

x_t คือ ตัวแปรอิสระ

β_0, β_1 คือ พารามิเตอร์

สำหรับการทดสอบอัตตสหสัมพันธ์ตำแหน่งที่ 1 AR(1) ของความคลาดเคลื่อน นอกจากตัวสถิติทดสอบเดอร์บิน-วัตสัน (Durbin-Watson Statistic) แล้ว ยังมีตัวสถิติ

อื่นๆ ที่ได้มีผู้คิดค้นขึ้นสำหรับการทดสอบอัตตสหสัมพันธ์ตำแหน่งที่ 1 ด้วยเช่นกัน และผู้วิจัยมีความคิดว่า ในบรรดาตัวสถิติเหล่านั้นน่าจะมีตัวสถิติบางตัวที่มีอำนาจการทดสอบสูงกว่าตัวสถิติทดสอบเดอร์บิน-วัตสันในบางเงื่อนไข หรือในบางสถานการณ์ ด้วยเหตุนี้ผู้วิจัยจึงสนใจศึกษาการทดสอบอัตตสหสัมพันธ์ในความคลาดเคลื่อน (u_t) ที่มีรูปแบบเป็น AR(1) โดยที่จะศึกษาตัวสถิติอื่นนอกเหนือจากตัวสถิติทดสอบเดอร์บิน-วัตสัน และเมื่อลักษณะการแจกแจงของความคลาดเคลื่อน (v_t) เป็นแบบต่างๆกันคือแบบสมมาตร แบบสม่ำเสมอ แบบหางยาว และแบบเบ้ และตัวแปรอิสระ (x_t) มี 4 รูปแบบ

การวิจัยครั้งนี้ผู้วิจัยจะศึกษาเปรียบเทียบตัวสถิติทดสอบ 3วิธี คือ ตัวสถิติทดสอบของเดอร์บิน-วัตสัน (Durbin-Watson Test: DW) ตัวสถิติทดสอบอัตราส่วนวอนนิวแมน (Von Nuemann Ratio Test : VN) และตัวสถิติทดสอบเกียรี่ (Geary Test : G)

ในการศึกษาเปรียบเทียบนั้น จะศึกษาความสามารถในการควบคุมความผิดพลาดประเภทที่ 1 (Type I Error) ถ้าตัวสถิติทดสอบใดสามารถควบคุมความผิดพลาดประเภทที่ 1 ได้ จึงจะศึกษาเปรียบเทียบอำนาจการทดสอบ (Power of the test) ของตัวสถิติทดสอบเหล่านั้นต่อไปภายใต้สถานการณ์ต่าง ๆ เพื่อจะได้เสนอแนะตัวสถิติทดสอบที่เหมาะสมสำหรับการทดสอบอัตตสหสัมพันธ์ตำแหน่งที่ 1 ในสมการถดถอยเชิงเส้นอย่างง่ายในสถานการณ์เหล่านั้น

1.2 วัตถุประสงค์ของการวิจัย

ในการวิจัยนี้ ต้องการศึกษาเปรียบเทียบตัวสถิติทดสอบ3วิธีต่อไปนี้ ในการทดสอบอัตตสหสัมพันธ์ตำแหน่งที่ 1 ในการวิเคราะห์สมการถดถอยเชิงเส้นอย่างง่าย เพื่อหาข้อสรุปเกี่ยวกับความเหมาะสมของการทดสอบดังกล่าว โดยพิจารณาจากความน่าจะเป็นของความผิดพลาดประเภทที่ 1 และอำนาจการทดสอบ

1. ตัวสถิติทดสอบของเดอร์บิน-วัตสัน (Durbin-Watson Test: DW)

ศูนย์วิทยทรัพยากร
จุฬาลงกรณ์มหาวิทยาลัย

2. ตัวสถิติทดสอบอัตราส่วนวอนนิวแมน (Von Nuemann Ratio Test : VN)

3. ตัวสถิติทดสอบเกียรี่ (Geary Test : G)

1.3 สมมุติฐานของการวิจัย

ในกรณีที่ค่าอัตตสหสัมพันธ์มีค่าน้อยและขนาดตัวอย่างเล็ก สถิติทดสอบอัตราส่วนวอนนิวแมน จะให้ค่าอำนาจการทดสอบสูงกว่าสถิติทดสอบเดอร์บิน-วัตสัน และสถิติทดสอบเกียรี่

1.4 ข้อตกลงเบื้องต้น

สมการถดถอยเชิงเส้นอย่างง่ายที่ใช้ในการวิจัย มีรูปแบบสมการดังนี้

$y_t = \beta_0 + \beta_1 x_t + u_t$, $t = 1,2,\ldots,n$

โดยที่

y_t คือ ตัวแปรตาม

x_t คือ ตัวแปรอิสระ

β_0, β_1 คือ พารามิเตอร์

u_t คือ ความคลาดเคลื่อนสุ่ม

โดยกำหนดรูปแบบอัตตสหสัมพันธ์ของ u_t เป็นแบบ AR(1)

$u_t = \rho u_{t-1} + v_t$

เมื่อ ρ คือ สัมประสิทธิ์อัตตสหสัมพันธ์ตำแหน่งที่ 1, $-1 \leq \rho \leq 1$

v_t คือ ความคลาดเคลื่อนสุ่ม

1.5 ขอบเขตการวิจัย

1.5.1 ศึกษาภายใต้ลักษณะการแจกแจงของความคลาดเคลื่อนสุ่ม (v_t) 4 แบบ คือ

ศูนย์วิทยทรัพยากร
จุฬาลงกรณ์มหาวิทยาลัย

หอสมุดกลาง สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

1.5.1.1 การแจกแจงแบบปกติ ; N(0,1)

$$f(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2} \qquad -\infty < x < \infty$$

$$\sigma^2 = 1 \; ; \; \mu = 0$$

1.5.1.2 การแจกแจงแบบสม่ำเสมอ ; $U(-\sqrt{3},\sqrt{3})$

$$f(x) = \frac{1}{2\sqrt{3}} \qquad -\sqrt{3} \leq x \leq \sqrt{3}$$

$$= 0 \qquad \text{อื่น ๆ}$$

1.5.1.3 การแจกแจงแบบเอกซ์โพเนนเชียล ; EX(1)

$$f(x) = e^{-x} \qquad x \geq 0$$

$$= 0 \qquad \text{อื่น ๆ}$$

1.5.1.4 การแจกแจงแบบโคชี ; C(0,1)

$$f(x) = \frac{1}{\pi(1+x^2)} \qquad -\infty < x < \infty$$

ศูนย์วิทยทรัพยากร
จุฬาลงกรณ์มหาวิทยาลัย

1.5.2 ลักษณะของตัวแปรอิสระที่นำมาศึกษามี 4 แบบ คือ

1.5.2.1 รูปแบบเส้นตรงตามเวลา (Simple Time Trend)

$$x_t = t \qquad , t = 1,2,\ldots,n$$

1.5.2.2 รูปแบบแนวโน้มไม่คงที่ (Stochastic Trend)

$$x_t = t + n_t \quad , t = 1,2,\ldots,n$$

$$n_t \sim N(0,1)$$

1.5.2.3 รูปแบบอัตตสหสัมพันธ์ตำแหน่งที่ 1 (First Order Autoregressive (AR(1))

$$x_t = 0.8x_{t-1} + n_t \quad , t = 1,2,\ldots,n$$

$$n_t \sim N(0,1)$$

$$x_0 \sim N\left(0, \frac{1}{1-0.8^2}\right)$$

1.5.2.4 รูปแบบแนวโน้มตามคาบเวลา (Periodic Trend)

$$x_t = t + \cos(2\pi t/12) \quad , t = 1,2,\ldots,n$$

1.5.3 กำหนด $\beta_0 = 1$

$\beta_1 = 1$

ในประชากรทุกรูปแบบที่ศึกษา*

1.5.4 ขนาดตัวอย่าง แบ่งออกเป็น 3 ขนาด คือ

ขนาดเล็ก n = 10, 15, 20

ขนาดกลาง n = 30

ขนาดใหญ่ n = 50

---

* การวิจัยครั้งนี้ ได้ทดลองค่าของ β_0, β_1 ที่ค่าอื่นๆ บางค่า ปรากฏว่าผลสรุปไม่แตกต่างกัน

ศูนย์วิทยทรัพยากร
จุฬาลงกรณ์มหาวิทยาลัย

1.5.5 ค่า ρ แบ่งออกเป็น 3 ระดับ คือ

ระดับต่ำ $\rho = 0.1, 0.3$

ระดับกลาง $\rho = 0.5$

ระดับสูง $\rho = 0.7, 0.9$

ค่าอัตสหสัมพันธ์ (ρ) มีค่าอยู่ในช่วง $-1 \leq \rho \leq 1$ แต่เนื่องจากในทางปฏิบัติ โดยเฉพาะข้อมูลทางธุรกิจและทางเศรษฐศาตร์ ส่วนมากจะพบว่ามีค่าเป็นบวก และค่าของ ρ มีความสมมาตรกัน ผู้วิจัยจึงศึกษาเฉพาะ $\rho \geq 0$

1.5.6 ระดับนัยสำคัญที่ศึกษาคือ 0.05

1.5.7 การวิจัยครั้งนี้จำลองข้อมูลขึ้นตามสถานการณ์ที่ต้องการศึกษา โดยใช้เทคนิคการจำลองแบบมอนติคาร์โล (Monte Carlo Simulation Technique) จากเครื่องคอมพิวเตอร์ AMDAHL 5860 โดยใช้ภาษาฟอร์แทรน (FORTRAN) ทำการจำลองแบบซ้ำๆกัน 1,000 ครั้ง ในแต่ละสถานการณ์ของการวิจัย

1.6 คำจำกัดความ

1.6.1 ความผิดพลาดประเภทที่ 1 (Type I Error) เป็นความผิดพลาดที่เกิดจากการปฏิเสธสมมุติฐาน H_0 เมื่อ H_0 ถูกต้อง

1.6.2 ความผิดพลาดประเภทที่ 2 (Type II Error) เป็นความผิดพลาดที่เกิดจากการยอมรับสมมุติฐาน H_0 เมื่อ H_0 ไม่ถูกต้อง

1.6.3 อำนาจการทดสอบ (Power of the Test) เป็นความน่าจะเป็นที่ปฏิเสธสมมุติฐาน H_0 เมื่อ H_0 ไม่ถูกต้อง

ศูนย์วิทยทรัพยากร
จุฬาลงกรณ์มหาวิทยาลัย

1.6.4 อัตตสหสัมพันธ์ (Autocorrelation) หมายถึง ตัวแปรสุ่ม u_t , $t = \pm1, \pm2, \ldots, \pm n$ มีความสัมพันธ์ต่อกัน กล่าวคือ $Cov(u_i, u_j) \neq 0$ เมื่อ $i \neq j$

1.7 ประโยชน์ที่คาดว่าจะได้รับ

1.7.1 เพื่อเป็นแนวทางในการตัดสินใจเลือกตัวสถิติที่เหมาะสม ในการตรวจสอบอัตตสหสัมพันธ์ตำแหน่งที่ 1 ในการวิเคราะห์สมการถดถอยเชิงเส้นอย่างง่าย ที่ให้อำนาจการทดสอบสูงสุด

1.7.2 เพื่อเป็นแนวทางในการศึกษาเปรียบเทียบตัวสถิติอื่นๆ ต่อไป

ศูนย์วิทยทรัพยากร
จุฬาลงกรณ์มหาวิทยาลัย