บทที่ 1

บทนำ

1.1 ความเป็นมาและความสำคัญของปัญหา

ในการประมาณค่าพารามิเตอร์ของความถดถอยเชิงเส้นแบบง่าย (Simple Linear Regression) นั้น การเลือกใช้วิธีที่เหมาะสมกับลักษณะของข้อมูลที่ต้องการวิเคราะห์นั้นจะต้องคำนึงถึงข้อตกลงเบื้องต้นของวิธีแต่ละวิธีที่ใช้ด้วย ในการวิจัยโดยทั่ว ๆ ไป ปัญหาที่ผู้วิจัยพบก็คือ ลักษณะของค่าความคลาดเคลื่อน (error) ไม่เป็นไปตามข้อตกลงเบื้องต้นกล่าวคือ ค่าความคลาดเคลื่อนมีสหสัมพันธ์กัน และมีความแปรปรวนไม่คงที่ ซึ่งสำหรับการศึกษาความถดถอยเชิงเส้นแบบง่ายที่มีรูปแบบทั่วไปเป็นดังนี้คือ $Y_t = \beta_o + \beta_1 X_t + \varepsilon_t$, $t = 1,2,\ldots, n$ โดยที่ X_t เป็นตัวแปรอิสระที่มีค่าคงที่ Y_t เป็นตัวแปรตาม β_o และ β_1 เป็นพารามิเตอร์ที่ไม่ทราบค่าแทนจุดตัดบนแกน Y และความชันของเส้นถดถอยตามลำดับ ε_t เป็นความคลาดเคลื่อน และ n เป็นขนาดตัวอย่าง โดยปกติการประมาณค่าพารามิเตอร์นั้น ผู้วิจัยมักจะเลือกใช้วิธีกำลังสองต่ำสุด (Ordinary Least Square หรือ OLS) ซึ่งเป็นวิธีที่ให้ตัวประมาณที่มีคุณสมบัติเป็นตัวประมาณเชิงเส้นที่ไม่เอนเอียงและดีที่สุด (Best Linear Unbiased Estimator) ทั้งนี้จะต้องมีข้อตกลงเบื้องต้นเกี่ยวกับความคลาดเคลื่อนดังนี้คือ ความคลาดเคลื่อนจะต้องมีการแจกแจงแบบปกติที่มีค่าเฉลี่ยเป็น 0 ค่าความแปรปรวนเป็น σ^2 และ ε_t , ε_s ไม่มีสหสัมพันธ์ต่อกัน เมื่อ $t \neq s$ อย่างไรก็ตามในทางปฏิบัติ บ่อยครั้งที่เราพบว่า ข้อมูลที่ถูกนำมาใช้วิเคราะห์นั้นมีอยู่ไม่น้อยที่ไม่เป็นไปตามข้อตกลงเบื้องต้นดังกล่าว โดยความคลาดเคลื่อนจะมีสหสัมพันธ์ และความแปรปรวนไม่คงที่ ซึ่งลักษณะเช่นนี้จะทำให้เกิดค่าคลาดเคลื่อนบางกลุ่มมีค่าสูงมาก ๆ หรือต่ำมาก ๆ ซึ่งจะไปมีผลต่อค่าสังเกต (Y) ลักษณะข้อมูลเช่นนี้จะพบมากทางด้านเศรษฐศาสตร์ และอุตสาหกรรม ดังนั้นในกรณีเช่นนี้ ผู้วิจัยอาจเลือกใช้วิธีกำลังสองน้อยที่สุดแบบทั่วไป เพื่อกำจัดลักษณะของความคลาดเคลื่อนอย่างใดอย่างหนึ่ง หรือทั้งสองลักษณะที่ไม่เป็นไปตามข้อตกลงเบื้องต้น

สำหรับในเรื่องการประมาณค่าพารามิเตอร์นั้น ผู้วิจัยต่างมุ่งหวังที่จะได้ตัวประมาณที่ "ดี" เสมอ ส่วนการที่จะตัดสินใจว่าตัวประมาณเช่นไรจึงจะเป็นตัวประมาณที่ "ดี" นั้นขึ้นกับจุดประสงค์ของการวิจัย ซึ่งโดยส่วนใหญ่จะอาศัยคุณสมบัติของตัวประมาณหลายอย่าง เป็นเกณฑ์ คุณสมบัติเหล่านั้นได้แก่ ความไม่เอนเอียง (Unbiasedness) ความคงเส้นคงวา (Consistency) ความพอเพียง (Sufficiency) ประสิทธิภาพ (Efficiency) ความแปรปรวนต่ำสุด (Minimum Variance) ความคลาดเคลื่อนกำลังสองเฉลี่ยต่ำสุด (Minimum Mean Square Error) เป็นต้น สำหรับการพิจารณาความเหมาะสมของตัวประมาณในการวิจัยครั้งนี้ จะพิจารณาจากคุณสมบัติ ความแปรปรวนต่ำสุด เนื่องจากในการเปรียบเทียบเชิงทฤษฎี เราทราบค่าพารามิเตอร์ ดังนั้นจึงสามารถใช้ความแปรปรวนต่ำสุดเป็นเกณฑ์ในการเปรียบเทียบได้ ซึ่งค่าความแปรปรวนต่ำสุดของตัวประมาณจะเป็นค่าที่แสดงถึงความแม่นยำ (Precision) ของตัวประมาณ โดยตัวประมาณที่มีความแม่นยำที่สุดก็คือตัวประมาณที่มีความแปรปรวนต่ำสุด

การศึกษาวิธีการประมาณค่าของพารามิเตอร์เมื่อลักษณะของความคลาดเคลื่อนมีสหสัมพันธ์ และความแปรปรวนไม่คงที่ ได้มีผู้ศึกษาไว้หลายท่าน โดยที่แต่ละท่านได้ศึกษาลักษณะใดลักษณะหนึ่งเท่านั้น เช่น การศึกษาของโกลเฟลด์และควันท์ เป็นการศึกษาลักษณะของความคลาดเคลื่อนที่มีความแปรปรวนไม่คงที่ การศึกษาของเดอบิน-วัตสัน เป็นการศึกษาลักษณะของความคลาดเคลื่อนที่มีสหสัมพันธ์กัน เป็นต้น จากแนวคิดในการศึกษาแต่ละลักษณะของความคลาดเคลื่อนที่ไม่เป็นไปตามข้อตกลงเบื้องต้นนี้ จึงเป็นสิ่งน่าสนใจที่จะนำเอาแนวความคิดทั้งสองมาศึกษาร่วมกัน เมื่อลักษณะของความคลาดเคลื่อนมีสหสัมพันธ์และความแปรปรวนไม่คงที่ ซึ่งในการวิจัยครั้งนี้จะเปรียบเทียบในรูปของค่าประสิทธิภาพสัมพัทธ์ (Relative Efficiency หรือ RE) ของวิธีกำลังสองต่ำสุดเทียบกับวิธีกำลังสองน้อยที่สุดแบบทั่วไป โดยใช้ค่าความแปรปรวนของตัวประมาณในแต่ละวิธีเป็นค่าเปรียบเทียบ เพื่อจะได้มีข้อสรุปที่แน่นอนในการเลือกใช้วิธีการประมาณค่าพารามิเตอร์ เมื่อความคลาดเคลื่อนมีสหสัมพันธ์ และความแปรปรวนไม่คงที่ซึ่งไม่เป็นไปตามข้อตกลงเบื้องต้นดังกล่าว

ในการศึกษาครั้งนี้ ผู้วิจัยได้ทำการศึกษาวิธีการประมาณค่าพารามิเตอร์ และเนื่องจากการคำนวณค่าความแปรปรวนต่ำสุดของตัวประมาณตามวิธีกำลังสองต่ำสุด หรือตามวิธีกำลังสองน้อยที่สุดแบบทั่วไป เมื่อสภาพความคลาดเคลื่อนไม่เป็นไปตามข้อตกลงเบื้องต้นดังกล่าว

กระทำได้ยากมาก ดังนั้นการวิจัยครั้งนี้จึงได้ทำการศึกษาโดยใช้เทคนิคการจำลองแบบที่เรียกว่า เทคนิคมอนติคาร์โลซิมูเลชั่น (Monte Carlo Simulation Technique) ซึ่งเป็นเทคนิคที่จะทำให้ได้ผลสรุปจากสภาพการณ์ที่เป็นการทดลอง สามารถกำหนดขนาดของกลุ่มตัวอย่าง ลักษณะการแจกแจงของความคลาดเคลื่อน ค่าเฉลี่ย และความแปรปรวนได้

1.2 วัตถุประสงค์ของการวิจัย

1.2.1 เพื่อศึกษาวิธีการประมาณค่าพารามิเตอร์ในการวิเคราะห์ความถดถอยเชิง-เส้น เมื่อลักษณะของความคลาดเคลื่อนมีสหสัมพันธ์ และความแปรปรวนไม่คงที่

1.2.2 เพื่อเปรียบเทียบประสิทธิภาพของตัวประมาณ 5 วิธี ซึ่งจำแนกเป็น

1.2.2.1 ตัวประมาณโดยวิธีกำลังสองต่ำสุด (Ordinary least Square)

1.2.2.2 ตัวประมาณโดยวิธีกำลังสองน้อยที่สุดแบบทั่วไป (Generalized Least Square) ซึ่งจำแนกเป็น

1. วิธีกำลังสองน้อยที่สุดแบบทั่วไปในการแก้ปัญหาเฉพาะลักษณะของความคลาดเคลื่อนที่มีความแปรปรวนไม่คงที่

2. วิธีกำลังสองน้อยที่สุดแบบทั่วไปในการแก้ปัญหาเฉพาะลักษณะของความคลาดเคลื่อนที่มีสหสัมพันธ์กัน

3. วิธีกำลังสองน้อยที่สุดแบบทั่วไปในการแก้ปัญหาทั้งลักษณะของความคลาดเคลื่อนที่มีสหสัมพันธ์กัน และความแปรปรวนไม่คงที่

4. วิธีกำลังสองน้อยที่สุดแบบทั่วไปในการแก้ปัญหาทั้งลักษณะของความคลาดเคลื่อนที่มีสหสัมพันธ์กัน และความแปรปรวนไม่คงที่ โดยอาศัยวิธีการแปลงข้อมูล

1.3 สมมติฐานของการวิจัย

ลักษณะการกระจายของความคลาดเคลื่อนที่มีสหสัมพันธ์กัน และความแปรปรวนไม่คงที่ ที่มีรูปแบบต่าง ๆ กัน จะมีผลให้ค่าความแปรปรวนของตัวประมาณแตกต่างกัน

1.4 ข้อตกลงเบื้องต้น

1.4.1 ความคลาดเคลื่อนเป็นตัวแปรสุ่มต่อเนื่องที่มีสหสัมพันธ์และความแปรปรวนไม่คงที่ ดังนี้

1.4.1.1 ค่าความแปรปรวนของความคลาดเคลื่อนไม่คงที่นั่นคือ

$$V(\varepsilon_t) = \sigma_t^2 \quad , t = 1,2,\ldots, n$$

โดยสามารถจำแนกค่าความแปรปรวนออกเป็น

1. ค่าความแปรปรวนของความคลาดเคลื่อนผันแปรตามตัวแปรอิสระ

$$V(\varepsilon_t) = \sigma^2 X_t^2$$

2. ค่าความแปรปรวนของความคลาดเคลื่อนผันแปรตามตัวแปรตาม

$$V(\varepsilon_t) = \sigma^2 \left[E(Y_t) \right]^2$$

$$= \sigma^2 (\beta_o + \beta_1 X_t)^2$$

3. ค่าความแปรปรวนของความคลาดเคลื่อนผันแปรแบบสุ่ม

$$V(\varepsilon_t) = \sigma^2 Z_t$$

เมื่อ Z_t มีการแจกแจงแบบยูนิฟอร์ม

1.4.1.2 ค่าความคลาดเคลื่อนมีสหสัมพันธ์กัน โดยกำหนดรูปแบบความสัมพันธ์ เป็นอัตตะถดถอยเรียงลำดับค่าแรก (first order autoregressive process)

$$\varepsilon_t = \rho \varepsilon_{t-1} + V_t$$

โดยที่ $|\rho| < 1$ และมีข้อตกลงเบื้องต้นสำหรับ V_t คือ

$$E(V_t) = 0$$

$$V(V_t) = \sigma^2 V_t$$

$$E(V_t V_{t-s}) = 0 \quad , s \neq 0$$

1.4.2 การวิจัยครั้งนี้ถือว่า ความแปรปรวนเป็นดัชนีสำคัญที่จะใช้เป็นเกณฑ์ในการเลือกตัวประมาณ ซึ่งตัวประมาณทุกตัวเป็นตัวประมาณที่ไม่เอนเอียง

1.5 ขอบเขตของการวิจัย

1.5.1 ศึกษาความแปรปรวนของตัวประมาณ ของวิธีกำลังสองต่ำสุด และวิธีกำลังสองน้อยที่สุดแบบทั่วไปทั้ง 4 รูปแบบ เมื่อความคลาดเคลื่อนมีการแจกแจงแบบปกติที่มีสหสัมพันธ์กัน และความแปรปรวนไม่คงที่ ซึ่งค่าความแปรปรวนผันแปรตามตัวแปรอิสระตัวแปรตาม และแบบสุ่ม

1.5.2 ในกรณีความคลาดเคลื่อนมีสหสัมพันธ์ จะศึกษาเมื่อค่าสหสัมพันธ์เป็น 0.3 0.5 0.7 และ 0.9 ในวิจัยนี้สหสัมพันธ์หมายถึง autocorrelation ซึ่งอาจใช้คำว่าอัตตะสหสัมพันธ์ได้

1.5.3 กำหนดระดับนัยสำคัญ 0.01 และ 0.05 ในการทดสอบความคลาดเคลื่อนมีสหสัมพันธ์ และความแปรปรวนไม่คงที่

1.5.4 ศึกษาขนาดตัวอย่าง 15 30 45 และ 60

1.5.5 ในการวิจัยครั้งนี้ จำลองการทดลองขึ้นโดยใช้เทคนิคมอนติคาร์โลขึ้นจากเครื่องคอมพิวเตอร์ IBM 370/3031 ซึ่งจะกระทำซ้ำกัน 300 ครั้งในแต่ละสถานการณ์*

---

*จากการทดลองจำลองสถานการณ์หนึ่งซ้ำกัน 300 ครั้ง และ 500 ครั้ง พบว่าผลลัพธ์ที่ได้ในการทดลองทั้งสองให้ผลไม่แตกต่างกันมากนัก เพื่อเป็นการไม่สิ้นเปลืองชั่วโมงการทำงานของคอมพิวเตอร์จึงเลือกการกระทำซ้ำกันเพียง 300 ครั้ง และสามารถดูผลเปรียบเทียบการทดลองจำลองสถานการณ์ทั้งสองได้จากภาคผนวก ข

หอสมุดกลาง สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

1.6 คำจำกัดความ

1.6.1 ความคลาดเคลื่อนกำลังสองเฉลี่ย (Mean Square Error) หรือ MSE ของตัวประมาณ คือค่าที่แสดงว่าค่าต่าง ๆ ของตัวประมาณแตกต่างจากค่าจริงเพียงไร โดยวัดในรูปค่าเฉลี่ยกำลังสองของค่าแตกต่างนั้น กล่าวคือ ถ้า $\hat{\theta}$ เป็นตัวประมาณของพารามิเตอร์ θ แล้ว ความคลาดเคลื่อนกำลังสองเฉลี่ยของ $\hat{\theta}$ คือ $E(\hat{\theta} - \theta)^2$

1.6.2 ความแปรปรวน (Variance) ของตัวประมาณ คือค่าที่แสดงว่าค่าต่าง ๆ ของตัวประมาณแตกต่างจากค่าเฉลี่ยของค่าประมาณเพียงไร โดยวัดในรูปค่าเฉลี่ยกำลังสองของค่าแตกต่างนั้น กล่าวคือ ถ้า $\hat{\theta}$ เป็นตัวประมาณของพารามิเตอร์ θ แล้ว ความแปรปรวนของ $\hat{\theta}$ คือ $E(\hat{\theta} - E(\hat{\theta}))^2$

1.6.3 ตัวประมาณเชิงเส้นที่ดีที่สุดและไม่เอนเอียง (Best Linear Unbiased Estimator) หรือ BLUE เป็นคุณสมบัติหนึ่งของตัวประมาณ โดยตัวประมาณ $\hat{\theta}$ จะมีคุณสมบัติเป็น BLUE ของพารามิเตอร์ θ ถ้า $\hat{\theta}$ มีคุณสมบัติครบ 3 ข้อต่อไปนี้คือ

1.6.3.1 เป็นฟังก์ชันเชิงเส้นของตัวอย่างสุ่ม

1.6.3.2 เป็นตัวประมาณที่ไม่เอนเอียง

1.6.3.3 เป็นตัวประมาณที่มีความแปรปรวนต่ำสุด

1.6.4 ประสิทธิภาพสัมพัทธ์ (Relative Efficiency) หรือ RE ของตัวประมาณ $\hat{\theta}_1$ เมื่อเทียบกับ $\hat{\theta}_2$ เป็นการเปรียบเทียบค่าความแปรปรวนของตัวประมาณ 2 ตัว ที่ประมาณค่าพารามิเตอร์เดียวกัน ในรูปอัตราส่วนของค่าความแปรปรวนของตัวประมาณหนึ่งต่อค่าความแปรปรวนของอีกตัวหนึ่ง หรืออาจจะพิจารณาจากสูตร

$$RE(\hat{\theta}_1 , \hat{\theta}_2) = \frac{V(\hat{\theta}_2)}{V(\hat{\theta}_1)}$$

1.7 ประโยชน์ของการวิจัย

เพื่อช่วยให้ข้อสรุปที่เป็นประโยชน์ต่อผู้ใช้ในการเลือกใช้ตัวประมาณพารามิเตอร์ได้อย่างมีประสิทธิภาพและเหมาะสม เมื่อความคลาดเคลื่อนไม่เป็นไปตามข้อตกลงเบื้องต้น