บทที่ 3

วิธีดำเนินการวิจัย

ในการวิจัยครั้งนี้ต้องการเปรียบเทียบวิธีการประมาณค่าพารามิเตอร์ของตัวแบบอนุกรมเวลาเมื่อข้อมูลมีค่าผิดปกติ โดยศึกษาเปรียบเทียบวิธีการประมาณค่าพารามิเตอร์ 3 วิธีคือ วิธีกำลังสองน้อยที่สุดเมื่อยังไม่ปรับปรุงข้อมูล วิธีตัวประมาณภาวะน่าจะเป็นสูงสุดเมื่อปรับปรุงข้อมูลแล้ว และวิธีกำลังสองน้อยที่สุดเมื่อปรับปรุงข้อมูลแล้ว โดยจะศึกษาเปรียบเทียบค่ารากที่สองของค่าเฉลี่ยความคลาดเคลื่อนกำลังสองของค่าพยากรณ์ล่วงหน้า 12 คาบเวลา โดยข้อมูลผิดปกติจะศึกษาในกรณีที่ความคลาดเคลื่อนมีการแจกแจงแบบปลอมปน โดยการแจกแจงที่มาปลอมปนนี้เป็นการแจกแจงที่มีหางยาวกว่าการแจกแจงปกติ

เทคนิคที่ใช้ในการจำลองข้อมูลข้อมูลครั้งนี้อาศัยเทคนิคการจำลองแบบมอนติคาร์โลสร้างสถานการณ์ต่างๆ ดังนั้นจึงขอกล่าวถึงวิธีการจำลองแบบมอนติคาร์โลก่อนที่จะแสดงถึงขั้นตอนการวิจัยและโปรแกรมสำหรับการวิจัยในลำดับต่อไป

วิธีการจำลองโดยใช้เทคนิคมอนติคาร์โล

เทคนิคที่ใช้แก้ปัญหาในการคำนวณทางสถิตินั้นมีอยู่หลายวิธี วิธีการจำลองโดยใช้เทคนิคมอนติคาร์โลเป็นวิธีหนึ่ง ที่นิยมนำมาใช้แก้ปัญหากันอย่างแพร่หลายในปัจจุบัน ซึ่งหลักการของการจำลองโดยใช้เทคนิคดังกล่าว จะใช้เลขสุ่ม(Random Numbers) มาช่วยในการหาคำตอบของปัญหาที่ต้องการศึกษา

ขั้นตอนของวิธีการจำลองด้วยเทคนิคมอนติคาร์โลที่ใช้กันในปัจจุบันแบ่งได้เป็น 3 ขั้นตอน ดังนี้คือ

1. การสร้างตัวเลขสุ่ม การใช้ตัวเลขสุ่มเป็นสิ่งสำคัญมากในเทคนิคนี้ ทั้งนี้เพราะว่าหลักการของการจำลองแบบมอนติคาร์โลนั้น จะใช้ตัวเลขสุ่มมาช่วยในการหาคำตอบของปัญหา โดยลักษณะของตัวเลขสุ่มที่นำมาใช้ จะมีการแจกแจงแบบสม่ำเสมอ (Uniform Distribution) ในช่วง(0,1) สำหรับวิธีการสร้างเลขสุ่มมีผู้เสนอแนะไว้หลายวิธี แต่วิธีที่ดีนั้นลักษณะของเลขสุ่มที่ถูกสร้างขึ้น จะต้องมีการแจกแจงแบบสม่ำเสมอในช่วง(0,1) ตัวเลขสุ่มแต่ละตัวเป็นอิสระต่อกัน และมีช่วงยาวก่อนจะเกิดเลขสุ่มซ้ำ(มีวัฏจักรยาว)

2. การนำเลขสุ่มมาประยุกต์ใช้กับปัญหาที่ต้องการศึกษา ซึ่งขั้นตอนนี้ขึ้นอยู่กับลักษณะของปัญหาที่ศึกษา บางปัญหาอาจใช้เลขสุ่มได้โดยตรง ในขณะที่บางปัญหาอาจต้องใช้ขั้นตอนอื่นอีกหลายขั้นตอน โดยมีการใช้ตัวเลขสุ่มในบางขั้นตอนเท่านั้น

3. การทดลองกระทำ เมื่อประยุกต์ปัญหาที่สนใจให้ใช้กับตัวเลขสุ่มได้แล้ว ขั้นต่อไปคือ การทดลองโดยใช้กระบวนการสุ่ม (Random Process) มากระทำในลักษณะซ้ำๆกัน (Replication) เพื่อหาคำตอบของปัญหาที่ต้องการศึกษา

การวางแผนการทดลอง

การวิจัยครั้งนี้ต้องการเปรียบเทียบความสามารถในการพยากรณ์ เมื่อมีการประมาณค่าพารามิเตอร์ของอนุกรมเวลาโดยวิธีการประมาณค่า 3 วิธี โดยมีแผนการทดลองดังนี้

1. ศึกษาในกรณีที่ข้อมูลอนุกรมเวลามี 2 รูปแบบคือ AR(1) และ MA(1)
2. ในแต่ละรูปแบบของอนุกรม กำหนดค่าพารามิเตอร์ 5 ระดับคือ 0.3 , 0.5 , 0.6 , 0.7 และ 0.8
3. ขนาดตัวอย่างมี 4 ขนาดคือ 40 , 60 , 80 และ 120
4. สัดส่วนของการปลอมปนมี 4 ระดับคือ 0.03 , 0.05 , 0.08 และ 0.10
5. การแจกแจงของความคลาดเคลื่อน(a_t) มีรูปแบบฟังก์ชันคือ

การแจกแจงปกติปลอมปน

$$f(X) = (1-p)N(0,100) + pH$$

เมื่อ H คือการแจกแจงที่มาปลอมปน ในการวิจัยนี้สนใจศึกษา 2 การแจกแจงคือ

$$H \sim N(0, C^2\sigma^2) \quad \text{การแจกแจงปกติ}$$

ในที่นี้กำหนด C = 5 และ 10 และ $\sigma^2 = 100$

$$\sim L(0, \beta) \quad \text{การแจกแจงแบบลาปลาซ}$$

ในที่นี้กำหนด β = 1 และ 10

การเปรียบเทียบจะพิจารณาจากค่ารากที่สองของค่าเฉลี่ยความคลาดเคลื่อนกำลังสองของค่าพยากรณ์ล่วงหน้า 12 คาบเวลา เพื่อหาวิธีที่ดีที่สุดในแต่ละสถานการณ์ โดยสถานการณ์ที่สนใจในการวิจัยครั้งนี้มีรวมทั้งสิ้น 640 สถานการณ์

ขั้นตอนการวิจัย

1. จำลองความคลาดเคลื่อน (a_t) ของข้อมูลอนุกรมเวลาจากที่กำหนดให้ในแผนการทดลอง

2. จำลองข้อมูลอนุกรมเวลา (z_t) ตามรูปแบบที่กำหนด ซึ่งมี 2 รูปแบบคือ AR(1) และ MA(1)

3. ประมาณค่าพารามิเตอร์ของอนุกรมเวลาทั้ง 3 วิธี

4. ปรับค่าข้อมูลผิดปกติ

5. ประมาณค่าพารามิเตอร์ของอนุกรมเวลาที่ปรับปรุงข้อมูลแล้ว

6. คำนวณค่าพยากรณ์ทั้ง 3 วิธี

7. คำนวณค่าความคลาดเคลื่อนจากการพยากรณ์ แล้วทำการเปรียบเทียบ

ซึ่งรายละเอียดของแต่ละขั้นตอนมีดังนี้

1. การจำลองค่าความคลาดเคลื่อน

การจำลองความคลาดเคลื่อนจะใช้การสร้างโปรแกรมย่อยสำหรับการสร้างการแจกแจงของความคลาดเคลื่อนที่ต้องการศึกษา ซึ่งมีการแจกแจงแบบต่างๆดังนี้

1.1 การแจกแจงแบบปกติ โปรแกรมย่อยสำหรับสร้างการแจกแจงแบบปกติจะใช้ฟังก์ชันย่อย NORMA(RMEAN,VAR,EX1) รายละเอียดแสดงไว้ในภาคผนวก ค.

1.2 การแจกแจงแบบลาปลาซ โปรแกรมย่อยสำหรับสร้างการแจกแจงแบบลาปลาซจะได้จากฟังก์ชันย่อย DOUB(RMEAN,VAR,EX) รายละเอียดแสดงไว้ในภาคผนวก ค.

2. การจำลองข้อมูลอนุกรมเวลา

สำหรับการสร้างข้อมูลอนุกรมเวลาที่มีข้อมูลผิดปกติจะใช้โปรแกรมย่อย GEN(RMEAN,VAR) ในการสร้างข้อมูลทั้งในรูปแบบ AR(1) และ MA(1) โดยเรียกใช้ฟังก์ชันย่อย AR1(NALL,RMEAN,VAR) สำหรับรูปแบบ AR(1) และเรียกใช้ฟังก์ชันย่อย MA1(NALL,RMEAN,VAR)สำหรับรูปแบบ MA(1) ด้วยความน่าจะเป็น 1-p และใช้ฟังก์ชันย่อย EAR1(NALL,RMEAN,VAR,EEA) สำหรับรูปแบบ AR(1) และใช้ฟังก์ชันย่อย EMA1

(NALL,RMEAN,VAR,EEC) สำหรับรูปแบบ MA(1) ด้วยความน่าจะเป็น p เมื่อความคลาดเคลื่อนถูกปลอมปนด้วยการแจกแจงแบบปกติ และใช้ฟังก์ชันย่อย LAR1(NALL,RMEAN, VAR,EEA) สำหรับรูปแบบ AR(1) และใช้ฟังก์ชันย่อย LMA1(NALL,RMEAN,VAR,EEC) สำหรับรูปแบบ MA(1) ด้วยความน่าจะเป็น p เมื่อความคลาดเคลื่อนถูกปลอมปนด้วยการแจกแจงลาปลาซ สำหรับรายละเอียดแสดงไว้ในภาคผนวก ค.

2.1. การสร้างตัวแปร z_t ตามรูปแบบ AR(1) มีขั้นตอนในการสร้างดังนี้

2.1.1 ข้อมูลที่มีรูปแบบปกติ

สร้าง z_0 ให้มีการแจกแจงปกติ ค่าเฉลี่ย 0 ความแปรปรวน $\frac{\sigma^2}{1-\phi_1^2}$ และสร้าง a_t ; $t=1,...,n$ ให้มีการแจกแจงแบบปกติ ค่าเฉลี่ย 0 ความแปรปรวน σ_a^2

จากนั้นสร้าง z_t ; $t=1,...,n$ ให้มีรูปแบบความสัมพันธ์คือ

$$z_t = \phi_1 z_{t-1} + a_t$$

2.1.2 ข้อมูลที่มีค่าผิดปกติรูปแบบ $N(0,C^2\sigma^2)$

สร้าง z_0 ให้มีการแจกแจงปกติ ค่าเฉลี่ย 0 ความแปรปรวน $\frac{\sigma^2}{1-\phi_1^2}$ และสร้าง a_t ; $t=1,...,n$ ให้มีการแจกแจงแบบปกติ ค่าเฉลี่ย 0 ความแปรปรวน $C^2\sigma_a^2$

จากนั้นสร้าง z_t ; $t=1,...,n$ ให้มีรูปแบบความสัมพันธ์คือ

$$z_t = \phi_1 z_{t-1} + a_t$$

2.1.3 ข้อมูลที่มีค่าผิดปกติรูปแบบ $L(0,\beta)$

สร้าง z_0 ให้มีการแจกแจงปกติ ค่าเฉลี่ย 0 ความแปรปรวน $\frac{\sigma^2}{1-\phi_1^2}$ และสร้าง a_t ; $t=1,...,n$ ให้มีการแจกแจงแบบลาปลาซ $L(0,\beta)$

2.2. การสร้างค่าตัวแปร z_t ตามรูปแบบ MA(1) มีขั้นตอนในการสร้างดังนี้

2.2.1 ข้อมูลที่มีรูปแบบปกติ

สร้าง a_t ; $t = 0,...,n$ ให้มีการแจกแจงปกติค่าเฉลี่ย 0 ความแปรปรวน $\sigma_a^2 = 100$

จากนั้นสร้าง z_t ; $t = 1,...,n$ ให้มีรูปแบบความสัมพันธ์คือ

$$z_t = a_t - \theta_1 a_{t-1}$$

2.2.2 ข้อมูลที่มีค่าผิดปกติรูปแบบ $N(0,C^2\sigma^2)$

สร้างa_t ; $t=1,...,n$ ให้มีการแจกแจงแบบปกติ ค่าเฉลี่ย 0 ความแปรปรวน $C^2\sigma_a^2$

จากนั้นสร้าง z_t ; $t = 1,...,n$ ให้มีรูปแบบความสัมพันธ์คือ

$$z_t = a_t - \theta_1 a_{t-1}$$

2.2.3 ข้อมูลที่มีค่าผิดปกติรูปแบบ $L(0,\beta)$

สร้าง a_t ; $t=1,...,n$ ให้มีการแจกแจงแบบลาปลาซ $L(0,\beta)$

จากนั้นสร้าง z_t ; $t = 1,...,n$ ให้มีรูปแบบความสัมพันธ์คือ

$$z_t = a_t - \theta_1 a_{t-1}$$

หลังจากสร้างข้อมูลดังข้างรูปแบบต้นแล้วจะใช้โปรแกรมย่อย GEN โดยกำหนดสัดส่วนค่าผิดปกติเท่ากับ 0.03 , 0.05 , 0.08 และ 0.10

3. การประมาณค่าพารามิเตอร์ของข้อมูลอนุกรมเวลาจากทั้ง 3 วิธี

3.1 วิธีกำลังสองน้อยที่สุดเมื่อยังไม่ปรับปรุงข้อมูล

รูปแบบ AR(1) ประมาณค่า ϕ_1 โดยใช้สูตร

$$\hat{\phi}_1 = \frac{\sum_{t=2}^{n} z_t z_{t-1}}{\sum_{t=3}^{n} z_{t-1}^2}$$

รูปแบบ MA(1) ประมาณค่า θ_1 โดยใช้สูตร

หอสมุดกลาง สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

$$\hat{\theta}_1 = -\frac{\sum_{t=2}^{n} z_t a_{t-1}}{\sum_{t=3}^{n} a_{t-1}^2}$$

3.2 วิธีตัวประมาณภาวะน่าจะเป็นสูงสุดเมื่อปรับปรุงข้อมูลแล้ว

รูปแบบ AR(1) ประมาณค่า ϕ_1 จากสมการ $f(\phi_1) = 0$

$$f(\phi_1) = \phi_1^3\left[z_1^2 - \sum_{t=2}^{n} z_{t-1}^2 - nz_1^2 + n\sum_{t=2}^{n} z_{t-1}^2\right] + \phi_1^2(2-n)\sum_{t=2}^{n} z_t z_{t-1}$$
$$+ \phi_1\left[-z_1^2 - \sum_{t=2}^{n} z_t^2 + nz_1^2 - n\sum_{t=2}^{n} z_{t-1}^2\right] + n\sum_{t=2}^{n} z_t z_{t-1}$$

การแก้สมการหาค่า ϕ_1 ใช้เทคนิคการแบ่งครึ่งช่วง[1] (Bisection Algorithm) มาช่วยในการคำนวณ

รูปแบบ MA(1) ประมาณค่า θ_1 จากสมการ $f(\theta_1) = 0$

$$f(\theta_1) = \frac{-\theta_1}{1-\theta_1^2} + \frac{(n+1)\theta_1^{2n+1}}{1-\theta_1^{2n+2}} - \frac{n}{2\sum_{t=0}^{n} a_t^2}\left[\sum_{t=2}^{n} Z_t a_{t-1} + \theta_1 \sum_{t=3}^{n} a_{t-1}^2\right]$$

การแก้สมการหาค่า θ_1 ใช้เทคนิคการแบ่งครึ่งช่วง มาช่วยในการคำนวณ

3.3 วิธีกำลังสองน้อยที่สุดเมื่อปรับปรุงข้อมูลแล้ว

ประมาณค่าพารามิเตอร์โดยใช้สูตรเดียวกันกับ วิธีกำลังสองน้อยที่สุดเมื่อยังไม่ปรับปรุงข้อมูล

4. ปรับค่าข้อมูลผิดปกติ

ตัวสถิติสำหรับทดสอบค่าผิดปกติคือ

$$\hat{\tau}(t) = \frac{\hat{\omega}_{(t)}}{\left[\mathrm{Var}(\hat{\omega})\right]^{1/2}}$$

รูปแบบ AR(1)

---

[1] Richard L. Burden, J. Douglas Faires. Numerical Analysis: (Boston : Prindle Weber & Schmidt,1985),pp.29

$$\hat{\omega}_t = z_t - \frac{\phi_1}{1+\phi_1^2}(z_{t-1}+z_{t+1}) \quad [2]$$

$$\mathrm{Var}(\hat{\omega}) = (1+\phi^2)^{-1}\sigma_a^2$$

รูปแบบ MA(1)

$$\hat{\omega}_t = \frac{1}{1+\theta_1^2}\left[a_t + \theta_1 a_{t+1} + \theta_1^2 a_{t+2} + \ldots\right]$$

$$\mathrm{Var}(\hat{\omega}) = \sigma_a^2 (1-\theta^2) \quad [3]$$

ปรับค่าผิดปกติในอนุกรม

$$\tilde{Z}_t = Z_t - \hat{\omega}_t$$

5. ประมาณค่าพารามิเตอร์ของอนุกรม

หลังจากปรับค่าผิดปกติในอนุกรมแล้วทำการประมาณค่าพารามิเตอร์ของอนุกรมอีกครั้งโดยใช้สูตรในการคำนวณเหมือนในขั้นตอนที่ 3

6. คำนวณค่าพยากรณ์จากทั้ง 3 วิธี

รูปแบบ AR(1) สมการที่ใช้ในการพยากรณ์คือ

$$\hat{Z}_t = \hat{\phi}_1 z(t-1) \qquad ; t = n+1$$
$$= \hat{\phi}_1 \hat{z}(t-1) \qquad ; t = n+2,\ldots,n+12$$

รูปแบบ MA(1) สมการที่ใช้ในการพยากรณ์คือ

$$\hat{z}(t) = -\hat{\theta}_1 a_{t-1} \qquad ; t = n+1,\ldots,n+12$$

---

[2] A. J. Fox, "Outlier In Time Series", Journal of the Royal Statistical Society,Ser. B,34(1972),pp.352

[3] Cheng I, Tiao G. C. and Chen C. "Estimation of Time Series Parameters in the Presence of Outliers" Technometries. ,30(1988),pp.194

7. คำนวนค่าความคลาดเคลื่อนจากการพยากรณ์แล้วทำการเปรียบเทียบ

การทดลองในสถานการณ์หนึ่งๆเมื่อได้ค่าพยากรณ์ล่วงหน้า 12 คาบเวลาครบทั้ง 3 วิธีแล้วจะนำค่าพยากรณ์มาเปรียบเทียบกับค่าจริง เพื่อคำนวนหาค่าความคลาดเคลื่อนกำลังสองของแต่ละคาบเวลาและทำซ้ำเช่นเดิมจนครบ 500 ครั้งแล้วจึงคำนวหาค่ารากที่สองของค่าเฉลี่ยความคลาดเคลื่อนกำลังสองเฉลี่ย 12 คาบเวลาพยากรณ์ ตามสูตร

$$\text{RMSE} = \frac{1}{12}\sum_{i=1}^{12}\sqrt{\sum_{i=1}^{500}\frac{(z_{ti} - \hat{z}_{ti})^2}{500}}$$

ซึ่งขั้นตอนของการทดลองดังกล่าวนี้สรุปเป็นผังงานได้ดังรูปที่ 3.1

โปรแกรมที่ใช้ในการวิจัย

โปรแกรมที่ใช้ในการวิจัยทั้งหมดเขียนด้วยภาษาฟอร์แทรน โดยใช้เครื่อง IBM AMDAHL 5860 ซึ่งในแต่ละสถานการณ์ของการทดลองลักษณะของการทำงานของโปรแกรมจะเหมือนกันสำหรับรายละเอียดของโปรแกรมจะแสดงไว้ในภาคผนวก ค.

รูปที่ 3.1 แสดงผังงานสำหรับหาค่าความคลาดเคลื่อนจากการพยากรณ์