บทที่ 1

บทนำ

1.1 ความเป็นมาและความสำคัญของปัญหา

ปัจจุบันเทคนิคการพยากรณ์มีบทบาทต่อวงการต่างๆ เป็นอย่างมาก ซึ่งเทคนิคการพยากรณ์จะมีโครงสร้าง 2 ส่วน ส่วนแรกคือ การสร้างสมการพยากรณ์ ส่วนที่สองคือ การนำสมการพยากรณ์ที่ได้จากส่วนแรกไปใช้ในการพยากรณ์ สำหรับส่วนของการสร้างสมการพยากรณ์นั้น ผู้วิจัยควรจะเลือกวิธีการประมาณค่าพารามิเตอร์ที่ให้ตัวประมาณที่ให้ค่าพยากรณ์มีความคลาดเคลื่อนต่ำ

เป็นที่ทราบกันดีว่า การใช้วิธีกำลังสองน้อยที่สุด (Ordinary Least Squares Method) ประมาณค่าพารามิเตอร์ในสมการถดถอยเชิงเส้นอย่างง่าย จะทำให้ตัวประมาณที่ได้เป็นตัวประมาณที่ไม่เอนเอียงเชิงเส้นมีความแปรปรวนต่ำสุด (Best Linear Unbiased Estimator (BLUE)) และโดยทั่วไปจะสมมติว่า ค่าคลาดเคลื่อนสุ่ม (random errors) ในสมการถดถอยมีการแจกแจงแบบปกติ (Normal Distribution) แต่ถ้าค่าคลาดเคลื่อนมิได้มีการแจกแจงแบบปกติ โดยเฉพาะในกรณีที่ค่าคลาดเคลื่อนมีการแจกแจงแบบหางยาว (Fat-tailed Distribution) ซึ่งมักจะแสดงถึงการเกิดค่าผิดปกติ (Outliers) ขึ้นในตัวแปรตาม (Y) จะได้ว่าวิธีกำลังสองน้อยที่สุดจะให้ตัวประมาณที่ให้ค่าพยากรณ์ที่มีความคลาดเคลื่อนสูง ดังนั้นจึงควรพิจารณาวิธีประมาณค่าพารามิเตอร์วิธีอื่นที่สามารถประมาณค่าพารามิเตอร์ได้ดีกว่าวิธีกำลังสองน้อยที่สุดในกรณีที่มีค่าผิดปกติเกิดขึ้นที่ตัวแปรตาม(Y)

ในการศึกษาครั้งนี้ สนใจที่จะใช้วิธีการประมาณค่าพารามิเตอร์ด้วยวิธีกำลังสองน้อยที่สุด (Ordinary Least Squares Method (OLS)) เปรียบเทียบกับวิธีค่าสัมบูรณ์ต่ำสุด (Least Absolute Value Method (LAV)) ภายใต้การแจกแจงของความคลาดเคลื่อนแบบต่างๆ

1.2 วัตถุประสงค์ของการวิจัย

ศึกษาเปรียบเทียบความคลาดเคลื่อนของการพยากรณ์ จากสมการถดถอยเชิงเส้นอย่างง่ายที่มีการประมาณค่าพารามิเตอร์ด้วยวิธีกำลังสองน้อยที่สุด และวิธีค่าสัมบูรณ์ต่ำสุด

1.3 ข้อตกลงเบื้องต้น

1. สมการถดถอยที่ใช้ในการศึกษาครั้งนี้จะใช้สมการถดถอยเชิงเส้นอย่างง่าย (Simple Linear Regression Equation) โดยมีรูปแบบสมการดังนี้

$$y_t = \beta_1+\beta_2 x_t+\varepsilon_t , \quad t = 1,2,\ldots,n$$

y_t เป็นตัวแปรตาม

x_t เป็นตัวแปรอิสระ

β_i เป็นพารามิเตอร์ที่ไม่ทราบค่า $i = 1,2$

ε_t เป็นค่าความคลาดเคลื่อนสุ่ม

2. ค่าคลาดเคลื่อนสุ่ม (ε_t) เป็นตัวแปรสุ่มที่มีการแจกแจงเหมือนกันและเป็นอิสระซึ่งกันและกัน

ศูนย์วิทยทรัพยากร
จุฬาลงกรณ์มหาวิทยาลัย

1.4 สมมติฐานทางการวิจัย

วิธีการค่าสัมบูรณ์ต่ำสุดจะให้ค่าพยากรณ์ที่มีความคลาดเคลื่อนน้อยกว่าวิธีกำลังสองน้อยที่สุดในกรณีที่ข้อมูลมีค่าผิดปกติ (Outliers) เกิดขึ้นในตัวแปรตาม (Y)

1.5 ขอบเขตของการวิจัย

1. ศึกษาเปรียบเทียบความคลาดเคลื่อนของการพยากรณ์จากสมการถดถอยโดยใช้วิธีการประมาณค่าพารามิเตอร์ 2 วิธี คือ

1.1 วิธีกำลังสองน้อยที่สุด (Ordinary Least Squares Method)

1.2 วิธีค่าสัมบูรณ์ต่ำสุด (Least Absolute Value Method)

2. ลักษณะของตัวแปรอิสระที่นำมาศึกษามีรูปแบบดังต่อไปนี้

2.1 รูปแบบอัตตสหสัมพันธ์อันดับที่ 1 (AR(1))

$X_t = 0.8X_{t-1} + U_t \quad , t = 1,\ldots,n$

$U_t \sim N(0,1)$

$X_o \sim N(0, 1/(1-(0.8)^2))$

2.2 รูปแบบเส้นตรงตามเวลา (Simple Time Trend)

$X_t = t \quad , \quad t = 1,2,\ldots,n$

2.3 รูปแบบแนวโน้มไม่คงที่ (Stochastic Trend)

$X_t = t + U_t \quad , t = 1,2,\ldots,n$

$U_t \sim N(0,1)$

ศูนย์วิทยทรัพยากร
จุฬาลงกรณ์มหาวิทยาลัย

2.4 รูปแบบแนวโน้มตามคาบเวลา (Periodic Trend)

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

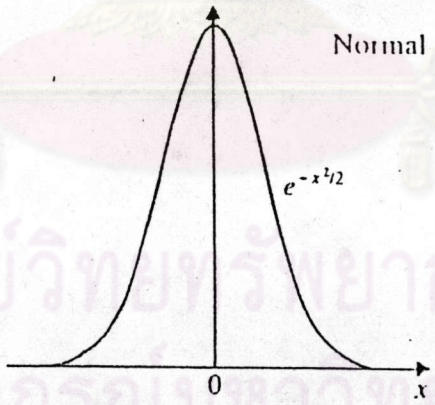
3. ค่า β_1 และ β_2 มีค่าเท่ากับ 1

4. ลักษณะการแจกแจงของความคลาดเคลื่อนสุ่ม (ε_t)

4.1 การแจกแจงแบบปกติ N(0,1)

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

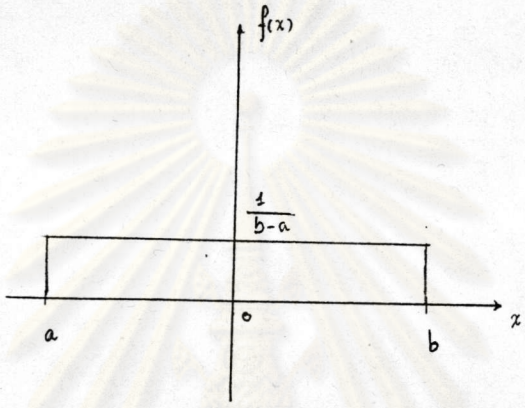
$$\sigma^2 = 1 \ , \mu = 0$$



ศูนย์วิทยทรัพยากร
จุฬาลงกรณ์มหาวิทยาลัย

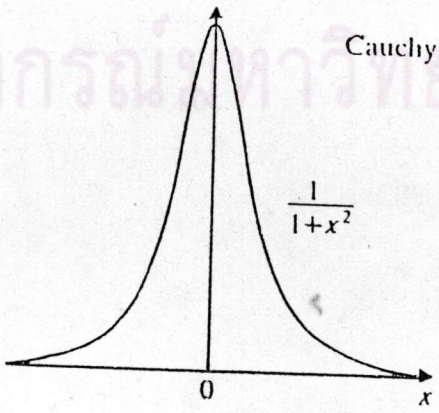
4.2 การแจกแจงแบบสม่ำเสมอ $U(-\sqrt{3}, \sqrt{3})$

$$f(x) = \begin{cases} \frac{1}{2\sqrt{3}} & -\sqrt{3} \leq x \leq \sqrt{3} \\ 0 & \text{อื่นๆ} \end{cases}$$



4.3 การแจกแจงแบบโคชี $C(0,1)$

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

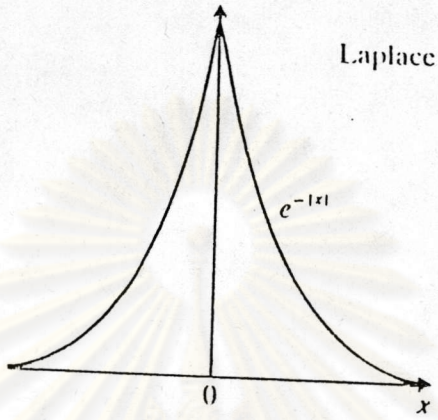


ศูนย์วิทยทรัพยากร
จุฬาลงกรณ์มหาวิทยาลัย

หอสมุดกลาง สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

4.4 การแจกแจงแบบลาปลาซ L(0,1)

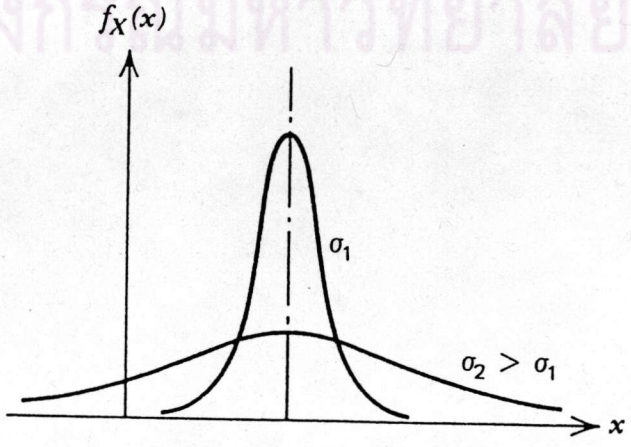
$$f(x) = \frac{1}{2} e^{-|x|} \qquad -\infty < x < \infty$$



4.5 การแจกแจงแบบปกติปลอมปน 0.85N(0,1)+ 0.15N(0,25)

$$f(x) = 0.85 \frac{1}{\sqrt{2\pi}} e^{-0.5x^2} + 0.15 \frac{1}{5\sqrt{2\pi}} e^{-0.5x^2/25}$$

$$-\infty < x < \infty$$



ศูนย์วิทยทรัพยากร
จุฬาลงกรณ์มหาวิทยาลัย

5. ขนาดตัวอย่างที่นำมาศึกษามี 3 ระดับ คือ 15 , 30 , 50

6. สร้างแบบจำลองข้อมูลให้มีสถานการณ์ตามที่ต้องการศึกษา ด้วยเทคนิคมอนติคาร์โลซิมูเลชั่น (Monte Carlo Simulation Technique) โดยเครื่องคอมพิวเตอร์ AMDAHL 5860 และเขียนโปรแกรมด้วยภาษาฟอร์แทรน

1.6 ประโยชน์ที่คาดว่าจะได้รับ

1.6.1 ผลการศึกษาเปรียบเทียบสามารถบอกได้ว่า ถ้ามีข้อมูลอยู่ชุดหนึ่งซึ่งมีค่าผิดปกติเกิดขึ้นควรจะใช้วิธีกำลังสองน้อยที่สุดหรือวิธีค่าสัมบูรณ์ต่ำสุดในการประมาณค่าพารามิเตอร์เพื่อนำไปใช้สร้างสมการพยากรณ์ ที่จะให้ค่าพยากรณ์ที่มีค่าคลาดเคลื่อนต่ำ

1.6.2 เพื่อเป็นแนวทางในการศึกษาและวิจัยต่อไป

1.7 วิธีดำเนินการวิจัย

1.7.1 ศึกษาวิธีการประมาณค่าพารามิเตอร์ในสมการถดถอยเชิงเส้นอย่างง่าย และเขียนโปรแกรมจำลองค่าสังเกตของตัวแปรในตัวแบบที่ต้องการศึกษา และเขียนโปรแกรมสำหรับคำนวณค่าพารามิเตอร์ของสมการถดถอยเชิงเส้นอย่างง่ายของวิธีกำลังสองน้อยที่สุดและวิธีค่าสัมบูรณ์ต่ำสุด

1.7.2 ศึกษาเปรียบเทียบค่ารากที่สองของค่าเฉลี่ยของความคลาดเคลื่อนกำลังสองของการพยากรณ์ที่ได้จากการประมาณค่าพารามิเตอร์ด้วยวิธีกำลังสองน้อยที่สุด และวิธีค่าสัมบูรณ์ต่ำสุด ข้อมูลที่ใช้ในการวิจัยครั้งนี้ได้จากการจำลองขึ้นในเครื่องคอมพิวเตอร์ AMDAHL 5860 โดยใช้เทคนิคมอนติคาร์โลซิมูเลชั่น และทำซ้ำ 200 รอบ ในแต่ละสถานการณ์

1.8 เกณฑ์ที่ใช้ในการเปรียบเทียบ

เกณฑ์ที่ใช้ในการเปรียบเทียบค่ารากที่สองของค่าเฉลี่ยของความคลาดเคลื่อนกำลังสองของการพยากรณ์ ($RMSFE_t$) ซึ่งค่า $RMSFE_t$ นี้คำนวณจาก

$$RMSFE_t = \sqrt{\left(\sum_{i=1}^{200} \frac{(\hat{y}_{t\,i} - y_{t\,i})^2}{200} \right)} \quad ,t= 1,\ldots,12$$

เมื่อ t แทน คาบเวลาที่ทำการพยากรณ์

i คือ รอบที่ทำการพยากรณ์ , i มีค่าตั้งแต่ 1 - 200 รอบ

$\hat{y}_{t\,i}$ คือ ค่าพยากรณ์ ณ คาบเวลาที่ t รอบที่ i

$y_{t\,i}$ คือ ค่าข้อมูลจริง ณ คาบเวลาที่ t รอบที่ i

โดยเกณฑ์ที่ใช้จะพิจารณาจากค่าเปอร์เซนต์ความแตกต่าง ($P.D._t$) ระหว่างค่ารากที่สองของค่าเฉลี่ยของความคลาดเคลื่อนกำลังสองจากการพยากรณ์ของวิธีค่าสัมบูรณ์ต่ำสุด ($RMSFE_t(LAV)$) และวิธีกำลังสองน้อยที่สุด ($RMSFE_t(OLS)$) นั่นคือ

$$P.D._t = \frac{RMSFE_t(OLS) - RMSFE_t(LAV)}{RMSFE_t(OLS)} \times 100$$

1.9 คำจำกัดความ

คำจำกัดความที่ใช้ในการวิจัยมีดังนี้

- ค่าผิดปกติ (Outliers) คือ ค่าสังเกตที่มีค่าสูง หรือต่ำกว่าปกติในกลุ่มค่าสังเกตเดียวกัน

ศูนย์วิทยทรัพยากร
จุฬาลงกรณ์มหาวิทยาลัย

- ค่ารากที่สองของค่าเฉลี่ยของความคลาดเคลื่อนกำลังสองของการพยากรณ์ (The Square Root of the Mean Squared Forecast Error ($RMSFE_t$)) หมายถึง

$$RMSFE_t = \sqrt{\left(\sum_{i=1}^{200} (\hat{y}_{t\,i} - y_{t\,i})^2 \right) \Big/ 200} \quad ,t= 1,\ldots,12$$

เมื่อ t แทน คาบเวลาที่ทำการพยากรณ์

i คือ รอบที่ทำการพยากรณ์ , i มีค่าตั้งแต่ 1 - 200 รอบ

$\hat{y}_{t\,i}$ คือ ค่าพยากรณ์ ณ คาบเวลาที่ t รอบที่ i

$y_{t\,i}$ คือ ค่าข้อมูลจริง ณ คาบเวลาที่ t รอบที่ i

ศูนย์วิทยทรัพยากร
จุฬาลงกรณ์มหาวิทยาลัย