บทที่ 1

บทนำ

หอสมุดกลาง สถาบันวิทยบริการ จุฬาลงกรณ์มหาวิทยาลัย

ที่มาและความสำคัญของปัญหา

การพยากรณ์มีความสำคัญต่อการวางแผน และการตัดสินใจเกี่ยวกับการดำเนินงานขององค์กรต่างๆ ไม่ว่าองค์กรนั้นจะเป็นองค์กรขนาดเล็กหรือขนาดใหญ่ องค์กรของรัฐหรือเอกชน หากนักวางแผนหรือผู้ตัดสินใจในองค์กรทราบว่า เหตุการณ์หรือสภาพการณ์ใดจะเกิดขึ้นหรืออาจจะเกิดขึ้นในอนาคตด้วยความเชื่อมั่นสูงระดับหนึ่ง จะทำให้การวางแผนหรือการตัดสินใจในการดำเนินงานเป็นไปได้ดีกว่า อย่างไรก็ตามการเกิดเหตุการณ์หรือสภาพการณ์ โดยทั่วไปเป็นการเกิดภายใต้ความไม่แน่นอนดังนั้นการพยากรณ์ที่ให้ความถูกต้องสูงจึงมีความจำเป็นอย่างยิ่ง

จากที่กล่าวมาแล้วข้างต้น เทคนิคการพยากรณ์จึงมีบทบาทสำคัญเป็นอย่างมาก ซึ่งเทคนิคการพยากรณ์จะประกอบไปด้วย 2 ส่วน ส่วนแรก คือ การสร้างสมการพยากรณ์และการประมาณค่าพารามิเตอร์ ส่วนที่สอง คือการนำสมการพยากรณ์ที่ได้นั้นไปใช้ในการพยากรณ์ ซึ่งค่าพยากรณ์ที่ได้จะมีความคลาดเคลื่อนสูงหรือต่ำเพียงใดจะขึ้นอยู่กับส่วนแรกเป็นสำคัญ ฉะนั้นการประมาณค่าพารามิเตอร์จึงมีความสำคัญต่อความถูกต้องของค่าพยากรณ์ที่จะนำไปใช้ในอนาคต

การวิเคราะห์การถดถอย (Regression Analysis) เป็นเทคนิคหนึ่งของการพยากรณ์ที่นิยมใช้กันอย่างแพร่หลาย เช่น ในวงการธุรกิจ และทางเศรษฐศาสตร์ ฯลฯ ซึ่งได้ใช้วิธีกำลังสองน้อยที่สุดแบบสามัญ (Ordinary Least Squares Method) มาประมาณค่าพารามิเตอร์ในสมการถดถอย ตัวประมาณที่ได้จะเป็นตัวประมาณที่ไม่เอนเอียงเชิงเส้น และมีความแปรปรวนต่ำสุด (Best linear Unbiased Estimator) ซึ่งวิธีการประมาณค่าพารามิเตอร์ที่กล่าวมานี้ จะใช้ได้ดี ภายใต้ข้อสมมติเบื้องต้นเกี่ยวกับความคลาดเคลื่อนสุ่ม (ε_t) คือ

1. ε_t มีเฉลี่ยเป็น 0 ; t=1,2,......,n
2. ε_t ความแปรปรวนของความคลาดเคลื่อนคงที่ $\mathrm{Var}(\varepsilon_t) = \sigma^2$; t=1,2,....,n
3. ε_t และ ε_s ไม่มีความสัมพันธ์กันหรือ $\mathrm{Cov}(\varepsilon_s, \varepsilon_t) = 0$ เมื่อ $s \neq t$

นอกจากข้อสมมติ 3 ข้อข้างต้น ข้อสมมติอีกข้อหนึ่งที่ใช้กันทั่วไปเมื่อทำการอนุมาน คือ สมมติ ε_t (t = 1,2,......,n) มีการแจกแจงแบบปกติ

โดยทั่วไปจะพบว่าข้อมูลทางด้านธุรกิจ ทางเศรษฐศาสตร์ จะถูกเก็บรวบรวมตามเวลาต่อเนื่องกัน เรียกว่า ข้อมูลอนุกรมเวลา (Time Series) ถ้าข้อมูลประเภทนี้มีความสัมพันธ์กับข้อมูลที่ผ่านมา เรียกว่า เกิดอัตสหสัมพันธ์ (Autocorrelation) และในบางครั้งอาจจะพบว่าข้อมูลไม่มีการแจกแจงแบบปกติ โดยอาจจะมีการแจกแจงแบบหางยาวกว่าการแจกแจงแบบปกติ ซึ่งการแจกแจงนี้จะแสดงให้เห็นว่าข้อมูลมีค่าผิดปกติ (Outlier) เกิดขึ้น ดังนั้นถ้าเราใช้วิธีกำลังสองน้อยที่สุดมาใช้ในการประมาณค่าพารามิเตอร์ย่อมไม่เหมาะสม เพราะตัวประมาณที่ได้จะให้ค่าพยากรณ์ที่มีความคลาดเคลื่อนสูง จึงควรจะพิจารณาวิธีอื่นๆ ที่มีความเหมาะสมกว่าวิธีกำลังสองน้อยที่สุดเพื่อจะได้นำไปใช้ในการประมาณค่าพารามิเตอร์ เมื่อความคลาดเคลื่อนเกิดอัตสหสัมพันธ์และมีค่าผิดปกติเกิดขึ้น

ในปี ค.ศ. 1989 Dielman ได้ทำการศึกษาในกรณีที่ความคลาดเคลื่อนมีการแจกแจงไม่เป็นแบบปกติ ได้แก่ การแจกแจงแบบหางยาว จากผลการวิจัยแสดงให้เห็นว่าวิธีค่าสัมบูรณ์ต่ำสุดจะให้ค่าพยากรณ์ที่มีความคลาดเคลื่อนต่ำกว่าวิธีกำลังสองน้อยที่สุด จากผลวิจัยดังกล่าว ในปี ค.ศ. 1994 Dielman และ Rose ได้ทำการศึกษาต่อ โดยใช้วิธีค่าสัมบูรณ์ต่ำสุดเพื่อแก้ปัญหาค่าผิดปกติ ร่วมกับวิธีการแปลงข้อมูลของเพรสและวินส์เทนเพื่อใช้ในการแก้ปัญหาอัตสหสัมพันธ์ ดังนั้นปัญหาทั้ง 2 ลักษณะก็ได้ถูกแก้ไข อีกแนวทางหนึ่งที่สามารถแก้ไขปัญหาดังกล่าว คือ การหาค่าพยากรณ์ร่วมที่เกิดจากการรวมวิธีการพยากรณ์เข้าด้วยกันโดยการให้น้ำหนัก ซึ่งวิธีการให้น้ำหนักมีหลายวิธีด้วยกัน แต่ในการวิจัยครั้งนี้สนใจศึกษาการให้น้ำหนักที่เท่ากันเพื่อความง่ายต่อการคำนวณ จึงเป็นสิ่งที่น่าสนใจว่าวิธีการประมาณวิธีใดที่จะให้ค่าพยากรณ์ที่มีความคลาดเคลื่อนต่ำกว่ากัน

ในการวิจัยครั้งนี้ ผู้วิจัยมีความสนใจที่จะศึกษาเปรียบเทียบความคลาดเคลื่อนจากการพยากรณ์ของ 5 วิธีคือ

1. วิธีกำลังสองน้อยที่สุด (Ordinary Least Squares Method : OLS)
2. วิธีค่าสัมบูรณ์ต่ำสุด (Least Absolute Value Method : LAV)
3. วิธีกำลังสองน้อยที่สุดแบบใช้การแปลงของเพรสและวินส์เทน(OLS using Prais-Winten Transformation Method)
4. วิธีการหาค่าพยากรณ์ร่วม (Combine Forecast Method)
5. วิธีค่าสัมบูรณ์ต่ำสุดแบบใช้การแปลงของเพรสและวินส์เทน (LAV using Prais-Winten Transformation Method)

โดยทำการเปรียบเทียบในสมการถดถอยเชิงเส้นอย่างง่าย เมื่อตัวแปรอิสระมีรูปแบบต่างๆ และความคลาดเคลื่อนมีความสัมพันธ์เป็นอัตสหสัมพันธ์อันดับที่หนึ่งและมีการแจกแจงแบบหางยาวกว่าการแจกแจงแบบปกติ

วัตถุประสงค์ของการวิจัย

เพื่อเปรียบเทียบความคลาดเคลื่อนของค่าพยากรณ์จากวิธีการประมาณค่าพารามิเตอร์ทั้ง 5 วิธีต่อไปนี้ สำหรับสมการถดถอยเชิงเส้นอย่างง่าย เมื่อความคลาดเคลื่อนเกิดอัตสหสัมพันธ์และมีค่าผิดปกติ

1. วิธีกำลังสองน้อยที่สุด (Ordinary Least Squares Method : OLS)
2. วิธีค่าสัมบูรณ์ต่ำสุด (Least Absolute Value Method : LAV)
3. วิธีกำลังสองน้อยที่สุดแบบใช้การแปลงของเพรสและวินส์เทน (OLS using Prais-Winten Transformation Method)
4. วิธีการหาค่าพยากรณ์ร่วม (Combine Forecast Method)
5. วิธีค่าสัมบูรณ์ต่ำสุดแบบใช้การแปลงของเพรสและวินส์เทน (LAV using Prais-Winten Transformation Method)

สมมติฐานของการวิจัย

ภายใต้สถานการณ์เมื่อความคลาดเคลื่อนสุ่ม ε_t เกิดอัตสหสัมพันธ์อันดับที่ 1 และมีค่าผิดปกติเกิดขึ้น วิธีการประมาณค่าพารามิเตอร์ด้วยวิธีค่าสัมบูรณ์ต่ำสุดแบบใช้การแปลงของเพรสและวินส์เทนจะให้ค่าพยากรณ์ที่ให้ค่ารากที่สองของค่าเฉลี่ยความคลาดเคลื่อนกำลังสองจากการพยากรณ์ต่ำกว่าวิธีอื่นๆ

ข้อตกลงเบื้องต้น

1. ในการศึกษาครั้งนี้ จะใช้รูปแบบของสมการถดถอยเชิงเส้นอย่างง่าย โดยมีรูปแบบดังนี้

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

โดยที่ y_t เป็นตัวแปรตาม

x_t เป็นตัวแปรอิสระ

β_i เป็นพารามิเตอร์ที่ไม่ทราบค่า ; $i = 0,1$

ε_t เป็นความคลาดเคลื่อนสุ่ม

n เป็นขนาดตัวอย่าง

2. ความคลาดเคลื่อน ε_t มีการแจกแจงเหมือนกันซึ่งมีสหสัมพันธ์กัน และมีความแปรปรวนคงที่

$$\mathrm{Var}(\varepsilon_t) = \sigma^2$$

3. ความคลาดเคลื่อน ε_t มีสหสัมพันธ์กัน ซึ่งกำหนดรูปแบบความสัมพันธ์เป็น Frist order Autoregressive [AR(1)]

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

โดยที่ $|\rho| < 1$ และมีข้อตกลงเบื้องต้นของ v_t คือ

$$\mathrm{E}(v_t) = 0$$
$$\mathrm{Var}(v_t) = \sigma_v^2$$
$$\mathrm{E}(v_s, v_t) = 0 \qquad \text{เมื่อ } s \neq t$$

4. ในการหาค่าพยากรณ์ร่วม จะใช้วิธีค่าสัมบูรณ์ต่ำสุดร่วมกับวิธีกำลังสองน้อยที่สุดแบบใช้การแปลงของเพรสและวินส์เทน

ขอบเขตของการวิจัย

1. รูปแบบตัวแปรอิสระเป็นดังนี้

1.1 รูปแบบปกติ (Normal)

$$x_t = u_t \qquad ; t = 1,2,......,n$$
$$u_t \sim \mathrm{N}(10,5)\,*$$

1.2 รูปแบบอัตสหสัมพันธ์อันดับที่ 1 (Frist Order Autoregressive ; AR(1))

$$x_t = 0.6^{**} x_{t-1} + u_t \qquad ; t = 1,2,......,n$$
$$u_t \sim \mathrm{N}(0,5)\,*$$
$$x_0 \sim \mathrm{N}(0, \frac{5}{1-0.6^2})$$

* การวิจัยครั้งนี้ได้ทดลองที่ค่าเฉลี่ยและความแปรปรวนอื่นๆ ได้ผลสรุปไม่แตกต่างกัน

** การวิจัยครั้งนี้ได้ทดลองที่สัมประสิทธิ์ค่าอื่นๆ นอกเหนือจาก 0.6 ได้ผลสรุปไม่แตกต่าง

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

หอสมุดกลาง สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

2. การแจกแจงความคลาดเคลื่อนสุ่ม v_t ที่สนใจศึกษา คือ

2.1 การแจกแจงแบบปกติ (Normal Distribution) ฟังก์ชันความหนาแน่นอยู่ในรูปของ

$$f(x)=\frac{1}{\sqrt{2\pi\sigma^2}}e^{-\frac{1}{2\sigma^2}(x-\mu)^2} \qquad -\infty<\mu<\infty,-\infty<x,\infty,\sigma>0$$

2.2 การแจกแจงแบบปกติปลอมปน (Contaminated Normal Distribution) ฟังก์ชันการแจกแจงอยู่ในรูปของ

$$f(x)=(1-p)N(0,5)+pN(0,5c^2)$$

เมื่อ p คือ สัดส่วนของการปลอมปน (percent of contamination) โดยกำหนดให้

p=0.05,0.08 และ 0.10

c คือ สเกลแฟคเตอร์ (scale factor) โดยกำหนดให้ c=5 และ 10

2.3 ฟังก์ชันการแจกแจงอยู่ในรูปของ

$$f(x)=(1-p)N(0,5)+pL(0,\beta)$$

เมื่อ p คือ สัดส่วนของการปลอมปน (percent of contamination) โดยกำหนดให้

p=0.05,0.08 และ 0.10

$L(0,\beta)$ คือการแจกแจงลาปลาซ (Laplace Distribution) ซึ่งมีฟังก์ชันความหนาแน่นอยู่ในรูปของ

$$f(x)=\frac{1}{2\beta}e^{-\left|\frac{x-\alpha}{\beta}\right|} \qquad -\infty<x<\infty,-\infty<\alpha,\infty,\beta>0$$

โดยที่ $E(x)=\alpha$, $Var(x)=2\beta^2$ ในที่นี้กำหนดให้ $\alpha=0$, $\beta=8$ และ 15

3. กำหนดให้ $\beta_0=10^*,\beta_1=1^*$

4. ระดับสหสัมพันธ์ (ρ) มี 5 ระดับ ซึ่งกำหนดให้เท่ากับ 0.1,0.3,0.5,0.7 และ 0.9

5. ขนาดตัวอย่าง (n) มี 5 ระดับ คือ 20,30,40,50 และ 60

* การวิจัยครั้งนี้ได้ทดลองที่ค่าสัมประสิทธิ์การถดถอยที่ค่าอื่นๆ ได้ผลสรุปไม่แตกต่างกัน

เกณฑ์ที่ใช้ในการตัดสินใจ

ในการวิจัยครั้งนี้ จะพิจารณาค่าความคลาดเคลื่อนจากการพยากรณ์ล่วงหน้าไป 12 คาบเวลา การพิจารณาความคลาดเคลื่อนจากการพยากรณ์ อาจจะพิจารณาเป็นรายเดือน หรือรายไตรมาส ซึ่งเกณฑ์ที่ใช้ในการตัดสินใจว่าค่าพยากรณ์ที่ได้จากแต่ละวิธีใดมีความถูกต้องมากที่สุด โดยจะพิจารณาจากการเปรียบเทียบค่ารากที่สองของค่าเฉลี่ยความคลาดเคลื่อนกำลังสองจากการพยากรณ์ (RMSFE) เฉลี่ย 12 คาบเวลา

$$\text{RMSFE} = \frac{1}{12}\sum_{j=1}^{12}\sqrt{\sum_{i=1}^{700}\frac{(y_{ji} - \hat{y}_{ji})}{700}}$$

y_{ji} คือ ค่าของข้อมูลจริง ณ คาบเวลาที่ j รอบที่ i

$\hat{y}_{ji}$ คือ ค่าพยากรณ์ ณ คาบเวลาที่ j รอบที่ i

j คือ คาบเวลาในการพยากรณ์ ; j = 1,2,......,12

i คือ จำนวนรอบของการทำซ้ำ ; i = 1,2,......,700

คำจำกัดความ

1. อัตสหสัมพันธ์ (Autocorrelation) คือ เหตุการณ์ที่ตัวแปรสุ่มมีความสัมพันธ์ในตัวเอง กล่าวคือ $Cov(\varepsilon_s, \varepsilon_t) \neq 0$ เมื่อ $s \neq t$

2. ค่าผิดปกติ (Outlier) คือ ค่าสังเกตที่มีค่าสูง หรือต่ำกว่าปกติในกลุ่มค่าสังเกตเดียวกัน

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

ประโยชน์ที่คาดว่าจะได้รับ

1. เพื่อใช้เป็นแนวทางในการตัดสินใจเลือกวิธีการประมาณค่าพารามิเตอร์ ที่ทำให้ค่าพยากรณ์มีความถูกต้องมากที่สุด เมื่อความคลาดเคลื่อนเกิดอัตสหสัมพันธ์และมีค่าผิดปกติ

2. เพื่อใช้เป็นแนวทางในการศึกษาต่อ เมื่อความคลาดเคลื่อนมีความสัมพันธ์ในรูปแบบอื่นๆ นอกเหนือจากกรณีที่ทำการศึกษาในวิจัยนี้