บทที่ 1

บทนำ

1.1 ความเป็นมาและความสำคัญของปัญหา

ในปัจจุบันการดำเนินธุรกิจนั้นหลีกเลี่ยงการแข่งขันกันไม่พ้น ธุรกิจใดมีการวางแผนและการตัดสินใจได้อย่างถูกต้อง และทันต่อเหตุการณ์ที่อาจผันแปรในอนาคตจึงจะสามารถเจริญเติบโตและพัฒนาได้อย่างก้าวไกล การคาดคะเน หรือการพยากรณ์เหตุการณ์ที่จะเกิดขึ้นในอนาคตจึงได้มีบทบาทสำคัญในขบวนการตัดสินใจ เพื่อให้ธุรกิจสามารถดำเนินต่อไปได้อย่างมีประสิทธิภาพ เช่น ผู้บริหารงานที่รับผิดชอบเกี่ยวกับการผลิตและการขายจะต้องพยากรณ์ยอดขายที่จะเกิดขึ้นในอนาคตเพื่อกำหนดแผนการผลิต ฝ่ายการตลาดจะต้องทำการคาดคะเนล่วงหน้าถึงผลของการรณรงค์โฆษณาที่มีต่อยอดขายของบริษัทในระยะยาวเป็นต้น จุดสำคัญของการพยากรณ์ก็เพื่อเป็นประโยชน์ในการกำหนดนโยบายและหาลู่ทางในการแก้ปัญหาต่างๆ ซึ่งในการกำหนดนโยบายนั้นอาจเป็นการกำหนดในระยะสั้น หรืออาจเป็นการวางแผนระยะยาวสำหรับอนาคตได้

ในเทคนิคการพยากรณ์จะมีโครงสร้างอยู่ 2 ส่วน ส่วนแรกคือ การสร้างสมการพยากรณ์ และส่วนที่สองคือ การนำสมการพยากรณ์ที่ได้จากส่วนแรกไปใช้ในการพยากรณ์ซึ่งวิธีการพยากรณ์ที่ใช้มีอยู่หลายวิธี แต่ละวิธีย่อมมีขีดความสามารถที่จำกัดอาจไม่เหมาะสมที่จะนำไปใช้ในสถานการณ์บางอย่าง ทั้งนี้ขึ้นอยู่กับลักษณะของข้อมูล และข้อตกลงเบื้องต้นของแต่ละวิธีการวิเคราะห์การถดถอยเชิงเส้น เป็นวิธีหนึ่งที่สามารถนำไปใช้ในการพยากรณ์หรือคาดคะเนเกี่ยวกับตัวแปรที่ต้องการศึกษา โดยอาศัยความรู้เกี่ยวกับค่าของตัวแปรอื่นๆที่เกี่ยวข้อง ซึ่งตัวแปรที่เกี่ยวข้องกับตัวแปรที่เราต้องการศึกษาอาจมีเพียงตัวเดียวเท่านั้น การวิเคราะห์ในลักษณะนี้เรียกว่า การวิเคราะห์การถดถอยเชิงเส้นอย่างง่าย (Simple Linear Regression) แต่ในบางกรณีตัวแปรที่เกี่ยวข้องกับตัวแปรที่เราต้องการศึกษาอาจมีตั้งแต่สองตัวขึ้นไป การวิเคราะห์ในลักษณะนี้เรียกว่า การวิเคราะห์การถดถอยเชิงเส้นแบบพหุ (Multiple Linear Regression)

ในการวิจัยครั้งนี้สนใจนำเทคนิคการวิเคราะห์สมการถดถอยเชิงเส้นแบบพหุมาใช้ในการพยากรณ์ ซึ่งมีรูปแบบดังนี้

$$Y_t = \beta_0 + \beta_1 X_{1t} + \beta_2 X_{2t} + \varepsilon_t , \quad t = 1,2,\ldots,n$$

โดยที่ Y_t เป็นตัวแปรตาม X_{1t} และ X_{2t} เป็นตัวแปรอิสระที่มีค่าคงที่ β_0, β_1 และ β_2 เป็นพารามิเตอร์ที่ไม่ทราบค่า ε_t เป็นค่าความคลาดเคลื่อนสุ่ม และ n เป็นขนาดตัวอย่าง

โดยปกติการประมาณค่าพารามิเตอร์ในสมการถดถอยเชิงเส้นนั้น ผู้วิจัยมักเลือกใช้วิธีกำลังสองน้อยที่สุด (Ordinary Least Squares : OLS) ซึ่งเป็นวิธีที่ให้ตัวประมาณที่มีคุณสมบัติเป็นตัวประมาณไม่เอนเอียงเชิงเส้นที่ดีที่สุด (Best Linear Unbiased Estimator : BLUE) ตามทฤษฎีเกาส์-มาร์คอฟ (Gauss-Markov Theorem) ทั้งนี้จะต้องมีข้อตกลงเบื้องต้นเกี่ยวกับความคลาดเคลื่อน ดังนี้คือ ความคลาดเคลื่อนจะต้องมีค่าเฉลี่ยเป็น 0 ค่าความแปรปรวนคงที่ และ $\varepsilon_i, \varepsilon_j$ ไม่มีสหสัมพันธ์ต่อกัน เมื่อ $i \neq j$ แต่ในทางปฏิบัติ เราจะพบว่า ข้อมูลที่นำมาใช้วิเคราะห์นั้นมีอยู่ไม่น้อยที่ไม่เป็นไปตามข้อตกลงเบื้องต้นดังกล่าว เช่น ความคลาดเคลื่อนจะมีสหสัมพันธ์กัน กล่าวคือ $E(\varepsilon_i \varepsilon_j) \neq 0$ เมื่อ $i \neq j$ ซึ่งในสถานการณ์เช่นนี้เราเรียกว่า อัตตสหสัมพันธ์ (Autocorrelation)

โดยทั่วไปข้อมูลอนุกรมเวลาที่นำมาวิเคราะห์ เช่นข้อมูลทางด้านเศรษฐศาสตร์หรือด้านธุรกิจ จะพบว่าความคลาดเคลื่อนมีอัตตสหสัมพันธ์ และรูปแบบที่พบโดยทั่วไปคือ อัตตถดถอยอันดับที่หนึ่ง(First Order Autoregressive) เช่น ราคาข้าวของปีนี้ขึ้นอยู่กับราคาข้าวของปีที่ผ่านมา การเกิดอัตตสหสัมพันธ์ในความคลาดเคลื่อนนั้นอาจเกิดขึ้นได้จากหลายสาเหตุ อาทิเช่น

1. การละเลยตัวแปรอิสระบางตัวที่สำคัญ ปกติแล้วผู้วิจัยมักจะหาตัวแปรอิสระเพื่อเป็นปัจจัยในการอธิบายตัวแปรตามให้ได้ครบถ้วน แต่ในบางครั้งอาจหาตัวแปรอิสระที่เกี่ยวข้องได้ไม่ครบถ้วน หรืออาจตัดตัวแปรอิสระบางตัวทิ้งไป เพราะไม่อาจวัดค่าหรือสังเกตค่าได้ ตัวแปรอิสระที่ถูกละเลยหรือตัดทิ้งเหล่านี้จึงส่งอิทธิพลต่อตัวแปรตามอยู่ภายนอกสมการ โดยส่งอิทธิพลผ่านทางตัวแปรสุ่ม ε ทำให้ตัวแปรสุ่ม ε ขาดความเป็นอิสระได้

2. การกำหนดรูปแบบความสัมพันธ์ทางคณิตศาสตร์ผิด กล่าวคือ ในการศึกษาความสัมพันธ์ระหว่างตัวแปรตามกับตัวแปรอิสระ ซึ่งความจริงตัวแปรทั้งสองมีความสัมพันธ์ไม่เป็นแบบเชิงเส้น เช่น แบบโพลิโนเมียลองศา 3 เป็นต้น แต่ผู้วิจัยกำหนดให้เป็นสมการเชิงเส้น ความคลาดเคลื่อนจึงเกิดขึ้น และความคลาดเคลื่อนเหล่านี้จะไปรวมอยู่ที่ตัวแปรสุ่ม ε ทำให้ตัวแปรสุ่ม ε ขาดความเป็นอิสระ และเกิดปัญหาอัตตสหสัมพันธ์ขึ้น

3. ความคลาดเคลื่อนที่เกิดจากวัดค่าของตัวแปร โดยปกติข้อมูลที่นำมาใช้ในการ

วิเคราะห์ ส่วนใหญ่ได้มาจากงานสำรวจตัวอย่าง ซึ่งในขั้นตอนการสุ่มตัวอย่างนั้น แม้ผู้รวบรวมข้อมูลจะควบคุมงานสำรวจได้ดีเพียงใด ความคลาดเคลื่อนก็ย่อมเกิดขึ้นได้เสมอในหลายขั้นตอน เริ่มตั้งแต่ขั้นวางแผนเตรียมการ จนถึงขั้นประมวลผลและเสนอผล ด้วยเหตุนี้ตัวแปรที่ต้องการศึกษาจึงผนวกเอาความคลาดเคลื่อนที่เกิดขึ้นไว้ และจะส่งผลกระทบต่อค่าความคลาดเคลื่อน(ε) ทำให้ค่าความคลาดเคลื่อนมีความสัมพันธ์กันได้

ดังนั้นถ้าหากข้อมูลที่นำมาศึกษาเกิดปัญหาดังกล่าว และผู้วิจัยยังคงใช้วิธีกำลังสองน้อยที่สุดในการประมาณค่าพารามิเตอร์ ย่อมจะก่อให้เกิดผลเสีย โดยเฉพาะในแง่คุณภาพของตัวประมาณพารามิเตอร์ β_1 จะได้ตัวประมาณที่ไม่เป็นตัวประมาณไม่เอนเอียงเชิงเส้นที่ดีที่สุด เนื่องจากมีความแปรปรวนไม่ต่ำสุด ถึงแม้จะยังคงเป็นตัวประมาณที่ไม่เอนเอียงก็ตาม จากผลนี้สามารถทำให้การอนุมานผิดพลาดร้ายแรงได้ เช่น ถ้าเกิดปัญหาอัตตสหสัมพันธ์ขึ้นโดยเฉพาะอย่างยิ่งอัตตสหสัมพันธ์ทางบวก (Positive Autocorrelation) ค่าความแปรปรวนของตัวประมาณจะมีค่าต่ำกว่าค่าความแปรปรวนที่แท้จริง (underestimated) ดังนั้นการทดสอบสมมติฐานของตัวประมาณอาจปรากฏว่าต่างจากศูนย์ ทั้งๆที่ความจริงแล้วไม่ต่างจากศูนย์ ซึ่งผลดังกล่าว เป็นเหตุให้การพยากรณ์ไม่ถูกต้องและแม่นยำได้

ดังนั้นในกรณีเช่นนี้ ผู้วิจัยอาจเลือกใช้วิธีที่เหมาะสมกว่าการใช้วิธีกำลังสองน้อยที่สุดในการประมาณค่าพารามิเตอร์ เพื่อขจัดลักษณะของความคลาดเคลื่อนที่มีอัตตสหสัมพันธ์กัน ซึ่งจะทำให้การประมาณและการพยากรณ์มีความถูกต้องยิ่งขึ้น ผู้วิจัยได้ศึกษาวิธีการในการประมาณค่าพารามิเตอร์ เมื่อความคลาดเคลื่อนสุ่มมีอัตตสหสัมพันธ์ โดยสนใจที่จะศึกษาวิธีประมาณสามวิธีต่อไปนี้ โดยการเปรียบเทียบค่ารากที่สองของความคลาดเคลื่อนกำลังสองเฉลี่ย (Root Mean Square Error : RMSE) จากการพยากรณ์ เมื่อความคลาดเคลื่อนมีอัตตสหสัมพันธ์อันดับที่หนึ่ง

1. วิธีกำลังสองน้อยที่สุด (Ordinary Least Squares Method)
2. วิธีการแปลงของเพรสและวินส์เทน
   (Prais-Winsten Transformation Method)
3. วิธีของฮิลเดรธและลู (Hildreth-Lu Method)

ในการวิจัยครั้งนี้ ผู้วิจัยไม่ได้ทำการเปรียบเทียบค่าความคลาดเคลื่อนของตัวประมาณพารามิเตอร์ $\hat{\beta}_1$;i=0,1,2 เนื่องจากมีการประมาณค่าพารามิเตอร์หลายค่า ซึ่งจะทำให้สรุปผลได้ยาก ว่าวิธีประมาณค่าพารามิเตอร์วิธีใดเป็นวิธีประมาณค่าพารามิเตอร์ที่ดีที่สุด เช่น วิธีกำลังสองน้อยที่สุด สามารถประมาณ $\hat{\beta}_1$ ได้ดีกว่าวิธีของฮิลเดรธและลู แต่อาจประมาณค่า

$\hat{\beta}_j$ (i≠j) ได้ไม่ดีกว่าวิธีของฮิลเดรธและลูก็ได้ และเหตุผลอีกประการ ก็เพื่อต้องการนำตัวแบบที่ได้ ไปใช้ในการพยากรณ์ค่า Y เพื่อให้เกิดประโยชน์ในด้านการพยากรณ์ต่อไป

1.2 วัตถุประสงค์ของการวิจัย

เพื่อศึกษาเปรียบเทียบวิธีการประมาณค่าพารามิเตอร์เพื่อการพยากรณ์ในสมการถดถอยเชิงเส้นแบบพหุ เมื่อความคลาดเคลื่อนมีอัตตสหสัมพันธ์อันดับที่หนึ่ง ใน 3 วิธีต่อไปนี้

1. วิธีกำลังสองน้อยที่สุด (Ordinary Least Squares Method)
2. วิธีการแปลงของเพรสและวินส์เทน
   (Prais-Winsten Transformation Method)
3. วิธีของฮิลเดรธและลู (Hildreth-Lu Method)

1.3 สมมติฐานของการวิจัย

ในกรณีที่ค่าอัตตสหสัมพันธ์มีระดับสูง วิธีของฮิลเดรธและลู จะให้ค่าพยากรณ์ที่มีค่ารากที่สองของความคลาดเคลื่อนกำลังสองเฉลี่ยน้อยกว่าวิธีกำลังสองน้อยที่สุด และวิธีการแปลงของเพรสและวินส์เทน

1.4 ข้อตกลงเบื้องต้น

1. ในการวิจัยครั้งนี้ ได้ศึกษารูปแบบสมการถดถอยเชิงเส้นแบบพหุ (Multiple Linear Regression Equation) มีรูปแบบดังนี้

$$Y_t = \beta_o + \beta_1 X_{1t} + \beta_2 X_{2t} + \varepsilon_t \quad , t = 1,2,\ldots,n$$

โดยที่

Y_t เป็นตัวแปรตาม (Dependent Variable)

X_{it} เป็นตัวแปรอิสระ (Independent Variable) ที่ไม่มีความสัมพันธ์ต่อกัน เมื่อ i = 1,2

β_i เป็นพารามิเตอร์ที่ไม่ทราบค่า (Unknown Parameter) ,i=0,1,2

ε_t เป็นค่าความคลาดเคลื่อนสุ่ม (Random Error)

2. ค่าความคลาดเคลื่อน (ε_t) มีสหสัมพันธ์กัน โดยกำหนดรูปแบบความสัมพันธ์เป็นอัตตถดถอยอันดับที่หนึ่ง (First Order Autoregressive) หรือ AR(1) ดังนี้

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

เมื่อ ρ คือ สัมประสิทธิ์สหสัมพันธ์ระหว่าง ε_t กับ ε_{t-1}, $|\rho| < 1$, และข้อตกลงเบื้องต้นของ v_t คือ

$$E(v_t) = 0$$

$$var(v_t) = \sigma_v^2$$

$$E(v_i v_j) = 0 \quad , i \neq j$$

ดังนั้นได้ว่า

$$E(\varepsilon_t) = 0$$

$$V(\varepsilon_t) = \sigma_\varepsilon^2 = \frac{\sigma_v^2}{1-\rho^2} \quad , t = 1,2,\ldots,n$$

1.5 ขอบเขตการวิจัย

1. ศึกษาภายใต้ลักษณะการแจกแจงของความคลาดเคลื่อน (v_t) ที่มีการแจกแจงแบบปกติ (Normal Distribution)

1.1 v_t มีค่าเฉลี่ยเป็น 0 และความแปรปรวนคงที่เป็น $\sigma_v^2 = 5^*$

1.2 ε_t มีค่าเฉลี่ยเป็น 0 และความแปรปรวนคงที่เป็น $\dfrac{\sigma_v^2}{1-\rho^2}$

2. ลักษณะของตัวแปรอิสระที่นำมาศึกษามีรูปแบบดังนี้

2.1 $X_{1t} = t \quad , t = 1,2,\ldots,n$

$X_{2t} = t + u_t$

2.2 $X_{1t} = t \quad , t = 1,2,\ldots,n$

$X_{2t} = t + \cos(2\pi t/12)$

2.3 $X_{1t} = t + u_t \quad , t = 1,2,\ldots,n$

$X_{2t} = t + \cos(2\pi t/12)$

$u_t \sim N(0,10)^{*}$

ซึ่ง $X_t = t$ เรียกว่า รูปแบบเส้นตรงตามเวลา (Simple Time Trend)

$X_t = t + u_t$ เรียกว่า รูปแบบแนวโน้มไม่คงที่ (Stochastic Trend)

$X_t = t + \cos(2\pi t/12)$ เรียกว่า รูปแบบแนวโน้มตามคาบเวลา (Periodic Trend)

3. กำหนด $\beta_o = 10,\ \beta_1 = 5,\ \beta_2 = 1$

เนื่องจากผู้วิจัยได้ทดลองที่ค่า β_1 อื่นๆ พบว่าเมื่อ β_1 เปลี่ยนค่าไปจะไม่มีผลกระทบต่อผลสรุปของแต่ละวิธีในการพยากรณ์ สำหรับในการวิจัยครั้งนี้ ผู้วิจัยจึงขอกำหนดเป็นค่าดังกล่าว

4. ศึกษาเมื่อค่าสหสัมพันธ์ (ρ) เป็น 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9, 0.95 และ 0.99

5. ขนาดตัวอย่าง (n) ที่นำมาศึกษามี 4 ระดับ คือ 15, 30, 45 และ 60

6. ศึกษาค่ารากที่สองของความคลาดเคลื่อนกำลังสองเฉลี่ยจากการพยากรณ์ล่วงหน้า 12 คาบเวลา และเฉลี่ยของ 12 คาบเวลา

7. การวิจัยครั้งนี้ จำลองข้อมูลขึ้นตามสถานการณ์ที่ต้องการศึกษา โดยใช้เทคนิคการจำลองแบบมอนติคาร์โล (Monte Carlo Simulation Technique) จากเครื่องคอมพิวเตอร์ AMDAHL 5860 โดยใช้ภาษาฟอร์แทน (Fortran) ทำการจำลองแบบซ้ำๆกัน 500 ครั้ง ในแต่ละสถานการณ์ของการวิจัย

1.6 เกณฑ์ที่ใช้ในการเปรียบเทียบ

เกณฑ์ที่ใช้ในการเปรียบเทียบใช้ ค่ารากที่สองของความคลาดเคลื่อนกำลังสองเฉลี่ยจากการพยากรณ์ ซึ่งสูตรการคำนวณมีดังนี้

---

* การวิจัยครั้งนี้ได้ทดลองที่ค่าเฉลี่ยและความแปรปรวนค่าอื่นๆ ได้ผลสรุปไม่แตกต่าง จึงเลือกค่าเฉลี่ยและความแปรปรวนเป็นค่าคงที่ใดๆ

$$RMSE_t = \sqrt{\frac{\sum_{i=1}^{500} (Y_{ti} - \hat{Y}_{ti})^2}{500}} \quad , \quad t = 1,2,\ldots,12$$

เมื่อ t คือ คาบเวลาที่ทำการพยากรณ์

Y_{ti} คือ ค่า Y จริง ณ คาบเวลาที่ t รอบที่ i

$\hat{Y}_{ti}$ คือ ค่า Y จากการพยากรณ์ ณ คาบเวลาที่ t รอบที่ i

1.7 คำจำกัดความ

1. อัตตสหสัมพันธ์ (Autocorrelation) คือเหตุการณ์ที่ตัวแปรสุ่ม ε มีความสัมพันธ์ต่อกัน กล่าวคือ $cov(\varepsilon_i, \varepsilon_j) \neq 0$ เมื่อ $i \neq j$

2. ค่ารากที่สองของความคลาดเคลื่อนกำลังสองเฉลี่ย (Root Mean Square Error) หรือ RMSE จากการพยากรณ์ คือ ค่าที่แสดงว่าค่าต่างๆจากการพยากรณ์ แตกต่างจากค่าจริงเพียงไร โดยวัดในรูป ค่ารากที่สองของค่าเฉลี่ยกำลังสองของค่าแตกต่างระหว่างค่าพยากรณ์ $\hat{Y}$ และค่าจริง Y ซึ่งสูตรการคำนวณมีดังนี้

$$RMSE_t = \sqrt{\frac{\sum_{i=1}^{500} (Y_{ti} - \hat{Y}_{ti})^2}{500}} \quad , \quad t = 1,2,\ldots,12$$

1.8 ประโยชน์ที่คาดว่าจะได้รับ

1. เพื่อเป็นแนวทางให้ผู้วิจัยสามารถเลือกใช้วิธีประมาณค่าพารามิเตอร์ที่จะนำไปใช้ในการสร้างสมการพยากรณ์ได้อย่างเหมาะสมและมีประสิทธิภาพ เมื่อความคลาดเคลื่อนมีอัตตสหสัมพันธ์อันดับที่หนึ่ง

2. สามารถเป็นแนวทางในการศึกษาเพื่อเลือกวิธีประมาณค่าพารามิเตอร์ที่นำไปใช้ในการพยากรณ์ เมื่อความคลาดเคลื่อนมีอัตตสหสัมพันธ์ในรูปแบบอื่นๆได้ต่อไป