บทที่ 1

บทนำ

ที่มาและความสำคัญของปัญหา

ปัจจุบันได้มีการนำเอาเทคนิคการพยากรณ์เชิงปริมาณไปใช้อย่างแพร่หลาย โดยเฉพาะอย่างยิ่งในวงการธุรกิจ ทั้งนี้อาจเป็นเพราะภาวะการแข่งขัน และความสลับซับซ้อนในวงการธุรกิจที่มีมากขึ้น เนื่องจากการพยากรณ์จะช่วยให้สามารถคาดการณ์ต่าง ๆ ได้ล่วงหน้า ซึ่งเป็นประโยชน์อย่างยิ่งต่อธุรกิจนั้น ๆ ที่จะได้อาศัยข้อมูลจากการพยากรณ์มาช่วยในการวางแผนงาน และการตัดสินใจสำหรับการดำเนินงานต่าง ๆ

การพยากรณ์เชิงปริมาณเป็นการพยากรณ์ที่อาศัยหลักวิชาการ โดยใช้สูตรสมการวิธีการทางสถิติ หรือแบบจำลองต่าง ๆ มาช่วยในการพยากรณ์ ซึ่งแบ่งออกเป็น 2 ประเภทใหญ่ ๆ คือ ประเภทที่ 1 การพยากรณ์เชิงอธิบาย (Explanatory Forecasting) เป็นการพยากรณ์โดยมีแนวความคิดว่า พฤติกรรมของสิ่งที่จะพยากรณ์ถูกกำหนดหรือถูกอธิบายโดยสิ่งอื่น ๆ ซึ่งมีความสัมพันธ์บางลักษณะกับสิ่งที่จะพยากรณ์ ได้แก่ เทคนิคการวิเคราะห์การถดถอย (Regression Analysis) และการจำลองเศรษฐมิติ (Econometric Model) และประเภทที่ 2 การพยากรณ์แบบอนุกรมเวลา (Time Series Forecasting) เป็นการพยากรณ์โดยมีแนวความคิดว่า พฤติกรรมของสิ่งที่จะพยากรณ์ควรจะเพียงพอที่จะพยากรณ์พฤติกรรมในอนาคตได้ ได้แก่ เทคนิคการทำให้เรียบ (Smoothing Technique) , การพยากรณ์แบบการกรองแบบปรับได้ (Adaptive Filtering) และวิธีอนุกรมเวลา Box-Jenkins เป็นต้น

การวิเคราะห์การถดถอยพหุ (Multiple Regression Analysis) เป็นเทคนิคหนึ่งที่นำมาใช้ในการพยากรณ์กันมาก โดยการใช้ค่าของตัวแปรอิสระ (Independent Variables) ตั้งแต่ 2 ตัวขึ้นไปมาพยากรณ์ค่าของตัวแปรตาม (Dependent Variable) ซึ่งตัวแปรทั้งสองประเภทนี้มีความสัมพันธ์กันในลักษณะใดลักษณะหนึ่ง ตัวอย่างเช่น ราคาสินค้าขึ้นอยู่กับต้นทุนสินค้า รายได้ประชาชาติ และส่วนแบ่งการตลาด เป็นต้น

โดยปกติการวิเคราะห์การถดถอย จะใช้ข้อมูลหรือค่าสังเกตที่ครบสมบูรณ์ของแต่ละตัวแปรที่พิจารณา แต่ในทางปฏิบัติผู้วิเคราะห์มักจะประสบกับปัญหาเกี่ยวกับการรวบรวมข้อมูลที่ต้องการใช้ศึกษา กล่าวคือค่าสังเกตของตัวแปรบางค่าที่ต้องการศึกษาสูญหายไป ซึ่งการสูญหายไปนี้อาจเกิดขึ้นโดยไม่ได้ตั้งใจ หรือเนื่องจากไม่ได้เก็บค่าโดยจงใจ หรืออาจเนื่องมาจากค่าใช้จ่ายในการเก็บค่าสังเกตมีจำนวนจำกัด หรืออาจเกิดจากเวลา หรือสภาวะแวดล้อม ที่ต้องทำให้ค่าสังเกตบางค่านั้นหายไป

ดังนั้นในการวิเคราะห์การถดถอย ถ้าข้อมูลที่เก็บรวบรวมมาเพื่อทำการวิเคราะห์เป็นไปตามข้อตกลงเบื้องต้น และมีข้อมูลครบสมบูรณ์ทุกตัวในขอบเขตของการพิจารณาก็ไม่เกิดปัญหาในการวิเคราะห์ แต่ถ้าหากข้อมูลที่รวบรวมมาได้นั้นมีบางตัวค่าสูญหายไป และไม่สามารถตามไปเก็บเพิ่มเติมได้ ทำให้ข้อมูลของตัวอย่างไม่สมบูรณ์อาจจะทำให้เกิดปัญหาในการวิเคราะห์ ผู้วิจัยอาจจะแก้ปัญหาโดยตัดค่าสังเกตชุดนั้นทิ้งไป ในกรณีนี้อาจจะมีผลทำให้ขนาดตัวอย่างมีจำนวนน้อยลง และส่งผลให้ค่าพยากรณ์มีความคลาดเคลื่อนสูง และที่สำคัญยิ่งก็คือทำให้สูญเสียรายละเอียดบางอย่างไป ซึ่งอาจจะมีผลกระทบต่อผลสรุปของการวิเคราะห์นั้น ๆ ได้ โดยทั่วไปการวิเคราะห์การถดถอยนั้น วิธีที่นิยมใช้ในการประมาณพารามิเตอร์หรือสัมประสิทธิ์การถดถอย (Regression Coefficient) คือวิธีกำลังสองน้อยที่สุด (Least Squares Method) แต่เมื่อค่าสังเกตสูญหายไปจะไม่สามารถประมาณได้ดีด้วยวิธีดังกล่าว นอกจากจะประมาณค่าสังเกตที่สูญหายก่อนที่จะใช้วิธีกำลังสองน้อยที่สุด

ในการวิจัยครั้งนี้ ผู้วิจัยได้ทำการศึกษาวิธีการหลายวิธีในการประมาณค่าที่สูญหายของตัวแปรตามในการวิเคราะห์การถดถอยเชิงเส้นพหุ และใช้วิธีกำลังสองน้อยที่สุดหาสัมประสิทธิ์การถดถอยเพื่อหาสมการถดถอยเชิงเส้นพหุในการพยากรณ์ ซึ่งวิธีการประมาณค่าสูญหายที่สนใจคือ

1. วิธีสูญหาย
2. วิธีค่าเฉลี่ย
3. วิธีสมการถดถอย
4. วิธีอีเอ็ม (EM Algorithm)
5. วิธีการของฮันท์ (Hunt's Method)

วัตถุประสงค์ของการวิจัย

1. เพื่อศึกษาวิธีการประมาณค่าสูญหายของตัวแปรตาม เพื่อการพยากรณ์ด้วยสมการถดถอยเชิงเส้นพหุ

2. เพื่อเปรียบเทียบวิธีการประมาณค่าสูญหายของตัวแปรตามทั้ง 5 วิธีด้วยการเปรียบเทียบความคลาดเคลื่อนของค่าพยากรณ์ที่ได้จากแต่ละวิธี

ข้อตกลงเบื้องต้น

1. สมการถดถอยที่ใช้ในการศึกษาครั้งนี้จะใช้สมการถดถอยเชิงเส้นพหุ (Multiple Linear Regression Equation) โดยมีรูปแบบสมการดังนี้

$$y_t = \beta_0 + \beta_1 x_{1t} + \beta_2 x_{2t} + \varepsilon_t \quad ; t = 1,..,n+m$$

เมื่อ y_t เป็นตัวแปรตาม

x_{1t}, x_{2t} เป็นตัวแปรอิสระตัวที่ 1 และ 2

β_i เป็นพารามิเตอร์ที่ไม่ทราบค่า $i = 0,1,2$

ε_t เป็นค่าความคลาดเคลื่อน

$n+m$ เป็นจำนวนค่าสังเกตทั้งหมด

n เป็นจำนวนค่าสังเกตที่ไม่สูญหาย

m เป็นจำนวนค่าสังเกตที่สูญหาย

2. ค่าความคลาดเคลื่อน (ε_t) เป็นตัวแปรสุ่มที่มีลักษณะดังนี้

2.1 $E(\varepsilon_t) = 0$

2.2 $V(\varepsilon_t) = \sigma^2$

2.3 $E(\varepsilon_t \varepsilon_k) = 0 \quad ; t \neq k$

2.4 $\varepsilon_t \sim N(0,\sigma^2)$

3. การสูญหายจะเกิดที่ตัวแปรตามเท่านั้น และเป็นการสูญหายแบบสุ่ม

สมมุติฐานของการวิจัย

ค่าพยากรณ์ที่ได้จากการประมาณค่าสูญหายของตัวแปรตามในการวิเคราะห์การถดถอยเชิงเส้นพหุด้วยวิธีอีเอ็ม (EM Algorithm) และวิธีการของฮันท์ (Hunt's Method) จะให้ค่าคลาดเคลื่อนโดยเฉลี่ยต่ำกว่าวิธีอื่น ๆ และวิธีอีเอ็ม (EM Algorithm) กับวิธีการของฮันท์ (Hunt's Method) จะให้ผลที่ไม่ต่างกัน

ขอบเขตของการวิจัย

1. กำหนดการสูญหายของข้อมูลเฉพาะข้อมูลของตัวแปรตาม และสูญหายอย่างสุ่ม

2. ลักษณะของตัวแปรอิสระที่นำมาศึกษามีรูปแบบดังต่อไปนี้

2.1 รูปแบบที่ 1

$x_{1t} = t$

$x_{2t} = t + u_t \quad ; \; u_t \sim N(0,9)$

2.2 รูปแบบที่ 2

$x_{1t} = t$

$x_{2t} = t + \cos(2\pi t/4)$

2.3 รูปแบบที่ 3

x_{1t} , x_{2t} มีการแจกแจงแบบปกติ N(20,60)

เมื่อ t = 1,2,...,n+m

3. กำหนดพารามิเตอร์ $\beta_0 = 10$, $\beta_1 = 1$ และ $\beta_2 = 2$

4. ค่าความคลาดเคลื่อนมีการแจกแจงปกติ $N(0,\sigma^2)$ โดยกำหนดให้ σ = 5,10,15, 20 และ 25

5. ขนาดตัวอย่างที่ใช้ในการศึกษาเท่ากับ 10,20,30,50 และ 70

6. ข้อมูลตัวแปรตามที่สูญหายร้อยละ 10,20,30,40,50,60 และ 70

7. การวิจัยครั้งนี้ได้จำลองข้อมูลให้มีสถานการณ์ตามที่กำหนดข้างต้น โดยใช้เทคนิคการจำลองแบบมอนติคาร์โล (Monte Carlo Simulation Technique) จากเครื่องคอมพิวเตอร์ AMDAHL 5860 เขียนโปรแกรมด้วยภาษาฟอร์แทรน (Fortran) ทำการจำลองข้อมูลซ้ำ 200 รอบในแต่ละสถานการณ์

เกณฑ์การตัดสินใจ

เกณฑ์การตัดสินใจว่าการประมาณค่าสูญหายด้วยวิธีใดใช้ได้ดีกว่า จะพิจารณาโดยการเปรียบเทียบค่าความคลาดเคลื่อนระหว่างค่าพยากรณ์ของตัวแปรตามกับค่าจริง ในรูปของค่ารากที่สองของค่าเฉลี่ยของความคลาดเคลื่อนกำลังสอง (The Squares Root of Mean Squares Error : RMSE) วิธีการใดให้ค่า RMSE ต่ำกว่าจะเป็นวิธีการประมาณที่ดีกว่า โดยคำนวณจากสูตร

$$\text{RMSE} = \frac{\sum_{t=1}^{12}\sqrt{\frac{\sum_{i=1}^{200}(y_{ti}-\hat{y}_{ti})^2}{200}}}{12}$$

โดยที่ y_{ti} คือค่าสังเกตของข้อมูลตัวที่ t หรือคาบเวลาที่ t ในการทำซ้ำรอบที่ i

$\hat{y}_{ti}$ คือค่าพยากรณ์ของข้อมูลตัวที่ t หรือคาบเวลาที่ t ในการทำซ้ำรอบที่ i

i คือจำนวนรอบของการทำซ้ำ ; i = 1,2,..,200

t คือค่าสังเกตของข้อมูลหรือคาบเวลาของการพยากรณ์

; t = (n+m)+1,(n+m)+2,..,(n+m)+12

ประโยชน์ที่คาดว่าจะได้รับ

1. เพื่อเป็นแนวทางในการตัดสินใจเลือก วิธีการประมาณค่าสูญหายของตัวแปรตามในสมการถดถอยเชิงเส้นพหุเพื่อการพยากรณ์

2. เพื่อเป็นแนวทางในการศึกษา และเปรียบเทียบวิธีการประมาณค่าสูญหายในสถานการณ์อื่น ๆ ต่อไป