หอสมุดกลาง สถาบันวิทยบริการ จุฬาลงกรณ์มหาวิทยาลัย

บทที่ 3

วิธีดำเนินการวิจัย

การวิจัยครั้งนี้มีวัตถุประสงค์เพื่อต้องการเปรียบเทียบวิธีการประมาณค่าพารามิเตอร์เพื่อการพยากรณ์ในสมการถดถอยเชิงเส้นแบบพหุ ใน 3 วิธีคือ วิธีกำลังสองน้อยที่สุด วิธีการแปลงของเพรสและวินส์เทน และวิธีของฮิลเดรธและลู เมื่อความคลาดเคลื่อนมีอัตตสหสัมพันธ์ โดยศึกษาเปรียบเทียบค่ารากที่สองของความคลาดเคลื่อนกำลังสองเฉลี่ยจากการพยากรณ์ ซึ่งได้ทำการพยากรณ์ล่วงหน้าไป 12 คาบเวลา และสำหรับการเปรียบเทียบ จะกระทำภายใต้สถานการณ์ดังนี้คือ ศึกษาค่าสหสัมพันธ์ (ρ) ณ ระดับ 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9, 0.95 และ 0.99 ขนาดตัวอย่างที่ใช้มี 4 ขนาดคือ 15, 30, 45 และ 60

การวิจัยครั้งนี้ใช้เทคนิคการจำลองแบบมอนติคาร์โล (Monte Carlo Simulation Technique) สร้างสถานการณ์ต่างๆ ดังนั้นจะขอกล่าวถึงวิธีการจำลองโดยใช้เทคนิคมอนติคาร์โล แล้วจึงแสดงรายละเอียดของขั้นตอนการวิจัยและโปรแกรมที่ใช้ในการวิจัยตามลำดับ

3.1 วิธีการจำลองโดยใช้เทคนิคมอนติคาร์โล

เทคนิคที่ใช้แก้ปัญหาในการคำนวณทางคณิตศาสตร์นั้นมีอยู่หลายวิธี วิธีการจำลองโดยใช้เทคนิคมอนติคาร์โลเป็นวิธีหนึ่งที่นิยมนำมาใช้แก้ปัญหากันอย่างแพร่หลายในปัจจุบันซึ่งหลักการของการจำลองโดยใช้เทคนิคดังกล่าวจะอาศัยตัวเลขสุ่ม (Random Numbers) มาช่วยในการหาคำตอบของปัญหาที่ต้องการศึกษา

ขั้นตอนของวิธีการจำลองด้วยเทคนิคมอนติคาร์โลที่ใช้กันในปัจจุบัน แบ่งได้เป็น 3 ขั้นตอนดังนี้

3.1.1 การสร้างตัวเลขสุ่ม การใช้ตัวเลขสุ่มเป็นสิ่งที่สำคัญมากในเทคนิคนี้ เพราะว่าหลักการของการจำลองแบบมอนติคาร์โลนั้น ได้ใช้ตัวเลขสุ่มมาช่วยในการค้นหาคำตอบของปัญหา โดยลักษณะของตัวเลขสุ่มที่นำมาใช้ จะมีการแจกแจงแบบยูนิฟอร์มในช่วง (0,1) สำหรับวิธีการสร้างตัวเลขสุ่มมีผู้เสนอไว้หลายวิธี แต่วิธีที่ดีนั้นลักษณะของตัวเลขสุ่มจะมีการแจกแจง

แบบยูนิฟอร์มในช่วง (0,1) และตัวเลขสุ่มแต่ละตัวเป็นอิสระกัน

3.1.2 การนำตัวเลขสุ่มมาประยุกต์ใช้กับปัญหาที่ต้องการศึกษา ขั้นตอนนี้ขึ้นอยู่กับลักษณะของปัญหา บางปัญหาอาจจะไม่ใช้ตัวเลขสุ่มโดยตรง แต่อาจจะมีขั้นตอนอื่นอีกหลายขั้นตอนที่ต้องการใช้ตัวเลขสุ่ม

3.1.3 การทดลองกระทำ เมื่อนำตัวเลขสุ่มมาประยุกต์ให้เข้ากับปัญหาที่ต้องการศึกษาได้แล้ว ขั้นต่อไปก็คือ การทดลองโดยใช้กระบวนการของการสุ่ม (Random Process) มากระทำในลักษณะที่ซ้ำๆกันหลายๆครั้ง เพื่อหาคำตอบที่ต้องการ

3.2 แผนการทดลอง

การวิจัยครั้งนี้ได้กำหนดสถานการณ์ต่างๆ เพื่อศึกษาเปรียบเทียบวิธีประมาณค่าพารามิเตอร์เพื่อการพยากรณ์ โดยใช้สมการถดถอยเชิงเส้นแบบพหุ เมื่อความคลาดเคลื่อนมีอัตตสหสัมพันธ์ โดยที่ลักษณะการแจกแจงของค่าความคลาดเคลื่อนที่ต้องการศึกษามีการแจกแจงแบบปกติ

เริ่มจากสร้างค่าความคลาดเคลื่อนที่มีสหสัมพันธ์กันในรูปแบบ AR(1) และตัวแปรอิสระตามรูปแบบที่กำหนด 3 รูปแบบ ซึ่งมีระดับสหสัมพันธ์ (ρ) 10 ระดับ และขนาดตัวอย่าง 4 ขนาด รวมทั้งสิ้น 120 สถานการณ์ ในแต่ละสถานการณ์จะนำค่าความคลาดเคลื่อนและตัวแปรอิสระที่ได้ไปสร้างตัวแปรตาม แล้วจึงทำการประมาณค่าพารามิเตอร์ด้วยวิธีการทั้ง 3วิธีดังกล่าว

3.3 ขั้นตอนในการวิจัย

ขั้นตอนในการวิจัยมีดังนี้คือ

1. การสร้างโปรแกรมย่อย (Subroutines) สำหรับการแจกแจงของความคลาดเคลื่อน

2. การสร้างข้อมูลคือ ตัวแปรตาม (Y) และตัวแปรอิสระ (X) โดยให้ตัวแปรตามมีความสัมพันธ์เชิงเส้นตรงกับตัวแปรอิสระ

3. การประมาณค่าพารามิเตอร์จากแต่ละวิธีการ

4. การสร้างสมการพยากรณ์ นำค่าพารามิเตอร์ที่ประมาณได้จากขั้นตอนที่ 3 มา

สร้างสมการพยากรณ์ และทำการพยากรณ์ล่วงหน้าไป 12 คาบเวลา

5. การหาค่ารากที่สองของความคลาดเคลื่อนกำลังสองเฉลี่ยจากการพยากรณ์ของแต่ละวิธี

สำหรับรายละเอียดของแต่ละขั้นตอนเป็นดังนี้

3.3.1 <u>การสร้างโปรแกรมย่อย สำหรับการแจกแจงของความคลาดเคลื่อน</u>

การสร้างลักษณะการแจกแจงของความคลาดเคลื่อนตามที่กำหนดในข้อตกลงเบื้องต้น จะต้องใช้ตัวเลขสุ่ม (Random Numbers) ซึ่งมีการแจกแจงแบบยูนิฟอร์มในช่วง (0,1) เป็นพื้นฐานในการสร้าง ซึ่งรายละเอียดของการผลิตและโปรแกรมที่ใช้ชื่อว่า RANDOM แสดงไว้ในภาคผนวก ก สำหรับรายละเอียดในการสร้างการแจกแจงแบบปกติเป็นดังนี้

3.3.1.1 การแจกแจงแบบปกติ โปรแกรมย่อยสำหรับสร้างการแจกแจงแบบปกติจะใช้คำสั่ง NORMAL(MEAN,SD,EX) ซึ่งรายละเอียดของโปรแกรมแสดงในภาคผนวก ก

3.3.2 <u>การสร้างข้อมูล X และ Y ที่มีความสัมพันธ์เชิงเส้นตรง</u>

มีขั้นตอนต่างๆดังนี้

3.3.2.1 สร้างค่าความคลาดเคลื่อนตามรูปแบบที่กำหนด ดังนี้

$$\varepsilon_t = \rho\varepsilon_{t-1} + v_t \quad , t = 1.2.....n$$

โดย v_t เป็นค่าความคลาดเคลื่อนสุ่มที่มีการแจกแจงแบบปกติ มีค่าเฉลี่ยเป็น 0 และความแปรปรวนเป็น 5 กำหนดระดับสหสัมพันธ์ (ρ) 10 ระดับ คือ 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9, 0.95 และ 0.99 กำหนดขนาดตัวอย่าง (n) ที่ใช้มี 4 ขนาด คือ 15, 30, 45 และ 60 จากนั้นจึงสร้าง ε_t ให้มีการแจกแจงแบบปกติ มีค่าเฉลี่ยและความแปรปรวนตามที่กล่าวมาแล้วในบทที่ 2

3.3.2.2 สร้างค่า X ซึ่งเป็นตัวแปรที่กำหนดรูปแบบที่เป็นลักษณะของข้อมูล

อนุกรมเวลา ซึ่งในการวิจัยครั้งนี้ จะศึกษารูปแบบของตัวแปรอิสระ 3 รูปแบบดังนี้

3.3.2.2.1 $X_{1t} = t \quad , t = 1,2,\ldots,n$

$X_{2t} = t + u_t$

3.3.2.2.2 $X_{1t} = t \quad , t = 1,2,\ldots,n$

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

3.3.2.2.3 $X_{1t} = t + u_t \quad , t = 1,2,\ldots,n$

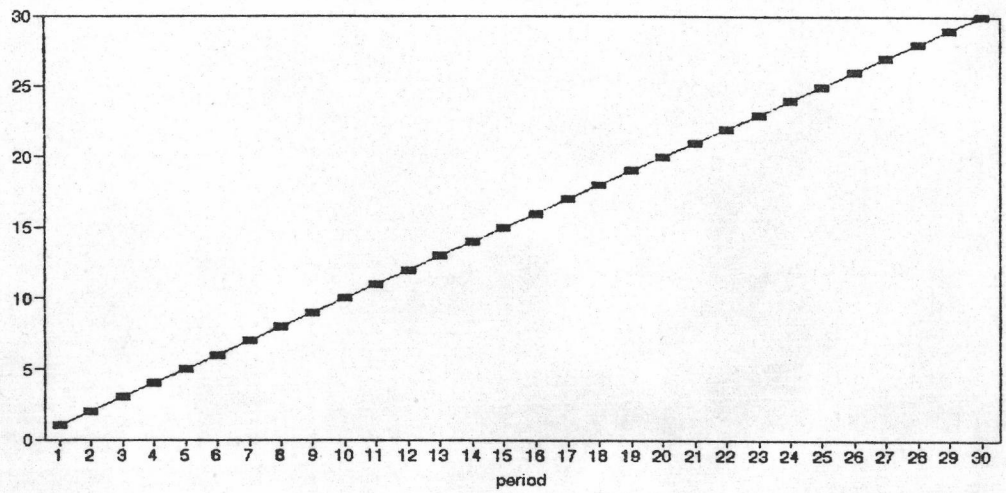
$X_{2t} = t + \cos(2\pi t/12)$

เมื่อ $u_t \sim N(0,10)$

สำหรับรูปแบบของตัวแปรอิสระที่นำมาศึกษาแสดงให้เห็นเป็นตัวอย่างดังรูปที่ 3.1 ถึง 3.3

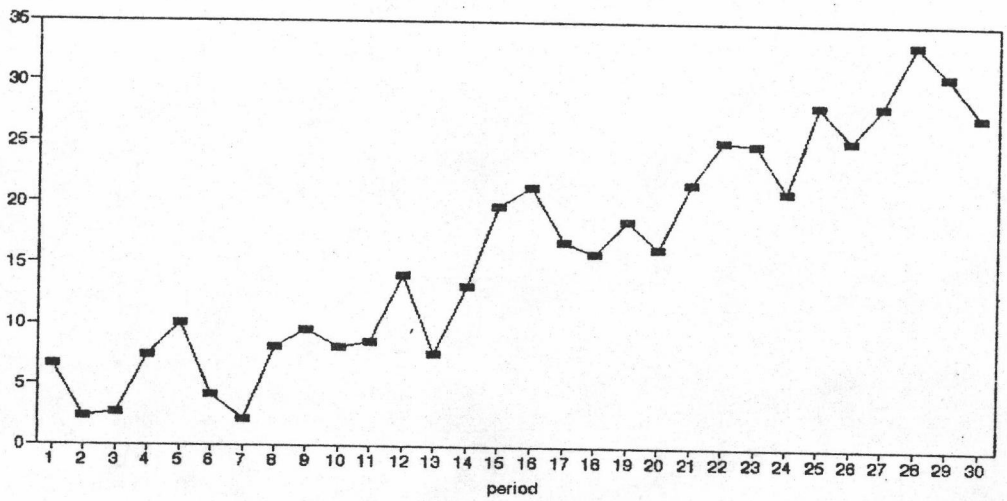
รูปที่ 3.1 รูปแบบเส้นตรงตามเวลา (Simple Time Trend)

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



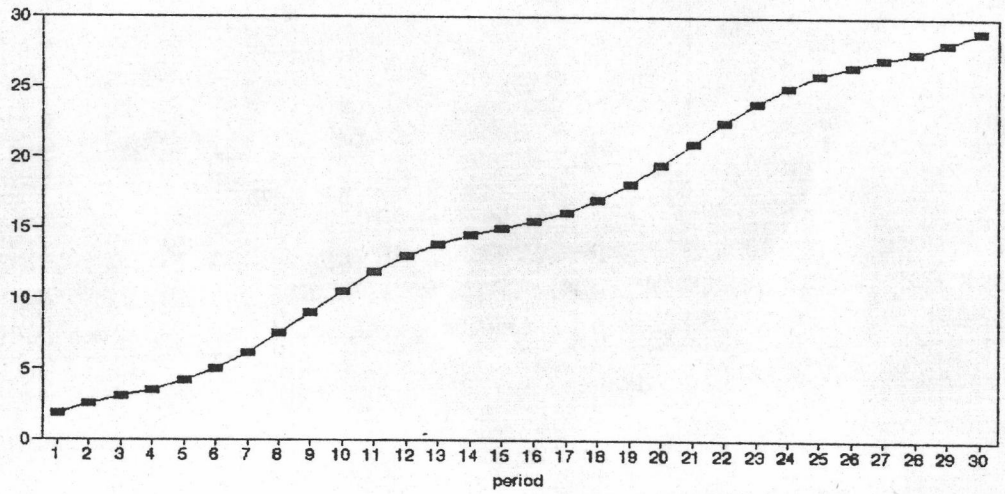
รูปที่ 3.2 รูปแบบแนวโน้มไม่คงที่ (Stochastic Trend)

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



รูปที่ 3.3 รูปแบบแนวโน้มตามคาบเวลา (Periodic Trend)

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



3.3.2.3 สร้างค่าตัวแปรตาม Y ซึ่งมีรูปแบบสมการดังนี้

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

โดยที่ β_o และ β_1 เป็นค่าพารามิเตอร์ที่ถูกกำหนดขึ้นมา

3.3.3 <u>การประมาณค่าพารามิเตอร์จากแต่ละวิธีการ</u>

เมื่อสร้างข้อมูล (X,Y) ที่มีความสัมพันธ์เชิงเส้นต่อกัน และรูปแบบความคลาดเคลื่อนตามที่กำหนดแล้ว จากนั้นจึงนำข้อมูล (X,Y) ที่ได้ไปประมาณค่าพารามิเตอร์ โดยในการวิจัยครั้งนี้ได้เสนอวิธีการประมาณค่าพารามิเตอร์ไว้ 3 วิธี ซึ่งมีรายละเอียดดังนี้

3.3.3.1 การประมาณค่าพารามิเตอร์ด้วยวิธีกำลังสองน้อยที่สุด

สูตรการประมาณค่าพารามิเตอร์ $\underset{\sim}{\beta}$ คือ

$$\hat{\underset{\sim}{\beta}} = (X'X)^{-1}(X'\underset{\sim}{Y})$$

3.3.3.2 การประมาณค่าพารามิเตอร์ด้วยวิธีการแปลงของเพรสและวินส์เทน

ในขั้นตอนแรกทำการประมาณค่า ρ ก่อน โดยใช้สูตร

$$\hat{\rho} = \frac{\sum_{t=2}^{n} e_t e_{t-1}}{\sum_{t=3}^{n} e_{t-1}^2}$$

เมื่อ e_t และ e_{t-1} คือ ค่าความคลาดเคลื่อนกำลังสองน้อยที่สุด (Ordinary Least Squares Residuals)

จากนั้นจึงนำค่าประมาณ ρ ไปแทนในสมการ

$$Y_t\sqrt{1-\rho^2} = \beta_o\sqrt{1-\rho^2} + \beta_1 X_{1t}\sqrt{1-\rho^2} + \beta_2 X_{2t}\sqrt{1-\rho^2} + v_t$$

เมื่อ $t = 1$

$$Y_t-\rho Y_{t-1} = \beta_0(1-\rho) + \beta_1(X_{1t}-\rho X_{1t-1}) + \beta_2(X_{2t}-\rho X_{2t-1}) + v_t$$

เมื่อ $t = 2,3,\ldots,n$

คำนวณหาค่าประมาณพารามิเตอร์ $\underset{\sim}{\beta}$ โดยวิธี OLS จะได้

$$\underset{\sim}{\hat{\beta}} = (X^{*\prime}X^{*})^{-1}(X^{*\prime}\underset{\sim}{Y}^{*})$$

โดยที่

$$\underset{\sim}{Y}^{*} = \begin{bmatrix} Y_1\sqrt{1-\rho^2} \\ Y_2-\rho Y_1 \\ Y_3-\rho Y_2 \\ \cdot \\ \cdot \\ \cdot \\ Y_t-\rho Y_{t-1} \end{bmatrix} \qquad X^{*} = \begin{bmatrix} \sqrt{1-\rho^2} & X_{11}\sqrt{1-\rho^2} & X_{21}\sqrt{1-\rho^2} \\ 1-\rho & X_{12}-\rho X_{11} & X_{22}-\rho X_{21} \\ 1-\rho & X_{13}-\rho X_{12} & X_{23}-\rho X_{22} \\ \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot \\ 1-\rho & X_{1t}-\rho X_{1t-1} & X_{2t}-\rho X_{2t-1} \end{bmatrix}$$

3.3.3.3 การประมาณค่าพารามิเตอร์ด้วยวิธีของฮิลเดรธและลู

ในขั้นตอนแรกทำการประมาณค่า ρ ก่อน โดยหาค่าประมาณ ρ จากเทคนิคการค้นหาแบบกริด (Grid Search Technique) จากนั้นจึงนำค่าประมาณ ρ ที่ได้แทนลงในสมการ

$$Y_t-\rho Y_{t-1} = \beta_0(1-\rho) + \beta_1(X_{1t}-\rho X_{1t-1}) + \beta_2(X_{2t}-\rho X_{2t-1}) + v_t$$

เมื่อ $t = 2,3,\ldots,n$

คำนวณหาค่าประมาณพารามิเตอร์ $\underset{\sim}{\beta}$ โดยวิธี OLS จะได้

$$\underset{\sim}{\hat{\beta}} = (X^{*\prime}X^{*})^{-1}(X^{*\prime}\underset{\sim}{Y}^{*})$$

โดยที่

$$\underset{\sim}{Y}^{*} = \begin{bmatrix} Y_2-\rho Y_1 \\ Y_3-\rho Y_2 \\ \cdot \\ \cdot \\ \cdot \\ Y_t-\rho Y_{t-1} \end{bmatrix} \qquad X^{*} = \begin{bmatrix} 1-\rho & X_{12}-\rho X_{11} & X_{22}-\rho X_{21} \\ 1-\rho & X_{13}-\rho X_{12} & X_{23}-\rho X_{22} \\ \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot \\ 1-\rho & X_{1t}-\rho X_{1t-1} & X_{2t}-\rho X_{2t-1} \end{bmatrix}$$

3.3.4 <u>การสร้างสมการพยากรณ์</u>

เมื่อได้ค่าประมาณพารามิเตอร์จากแต่ละวิธีแล้ว จะทำการประมาณซ้ำจนครบ 500 ครั้ง จากนั้นจึงนำค่าประมาณพารามิเตอร์มาหาค่าเฉลี่ย เพื่อนำค่าประมาณที่ได้ไปเข้าสมการพยากรณ์ เพื่อหาค่าพยากรณ์ Y โดยจะทำการพยากรณ์ล่วงหน้าไป 12 คาบเวลา ซึ่งแต่ละวิธีที่ศึกษามีรูปแบบของสมการพยากรณ์ค่า Y ดังนี้

3.3.4.1 วิธีกำลังสองน้อยที่สุด

สมการพยากรณ์ค่า Y_t คือ

$$\hat{Y}_t = \hat{\beta}_o + \hat{\beta}_1 X_{1t} + \hat{\beta}_2 X_{2t}$$

3.3.4.2 วิธีการแปลงของเพรสและวินส์เทน

สมการพยากรณ์ค่า Y_t คือ

$$\hat{Y}_t = \hat{\rho} Y_{t-1} + \hat{\beta}_o(1-\hat{\rho}) + \hat{\beta}_1(X_{1t}-\hat{\rho}X_{1t-1}) + \hat{\beta}_2(X_{2t}-\hat{\rho}X_{2t-1})$$

3.3.4.3 วิธีของฮิลเดรธและลู

สมการพยากรณ์ค่า Y_t คือ

$$\hat{Y}_t = \hat{\rho}Y_{t-1} + \hat{\beta}_0(1-\hat{\rho}) + \hat{\beta}_1(X_{1t}-\hat{\rho}X_{1t-1}) + \hat{\beta}_2(X_{2t}-\hat{\rho}X_{2t-1})$$

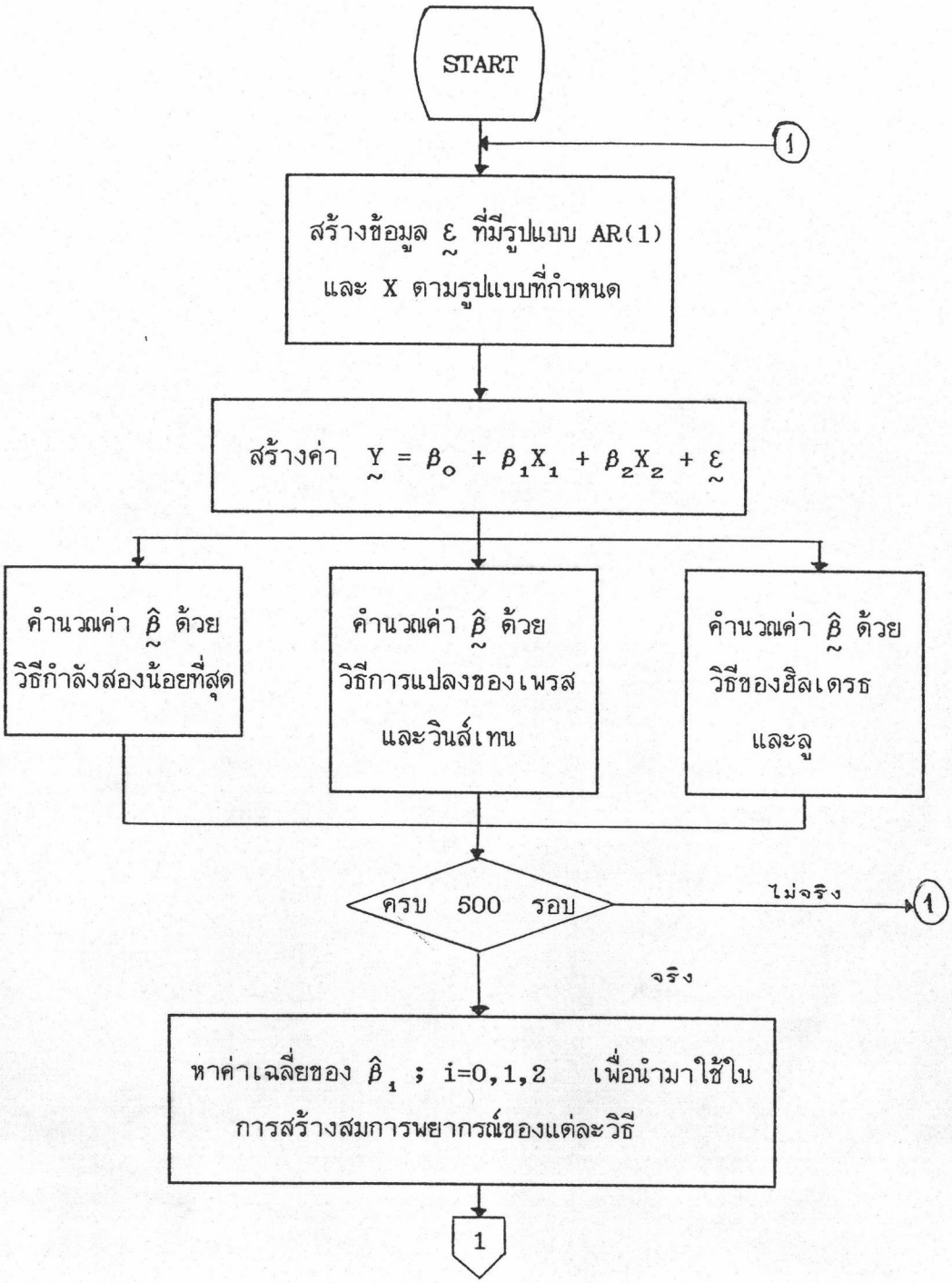
3.3.5 <u>การหาค่ารากที่สองของความคลาดเคลื่อนกำลังสองเฉลี่ยจากการพยากรณ์</u>

การทดลองในสถานการณ์หนึ่งๆ เมื่อได้ค่าพยากรณ์ Y ล่วงหน้า 12 คาบเวลา ในแต่ละวิธีทั้ง 3 แล้ว จะนำค่าพยากรณ์มาหาความแตกต่างกับค่าจริง Y เพื่อคำนวณหาค่าความคลาดเคลื่อนกำลังสองของแต่ละคาบเวลา และทำซ้ำเช่นเดิมจนครบ 500 ครั้ง แล้วจึงคำนวณหาค่ารากที่สองของความคลาดเคลื่อนกำลังสองเฉลี่ยในแต่ละคาบเวลา ($RMSE_t$) จากสูตรดังนี้

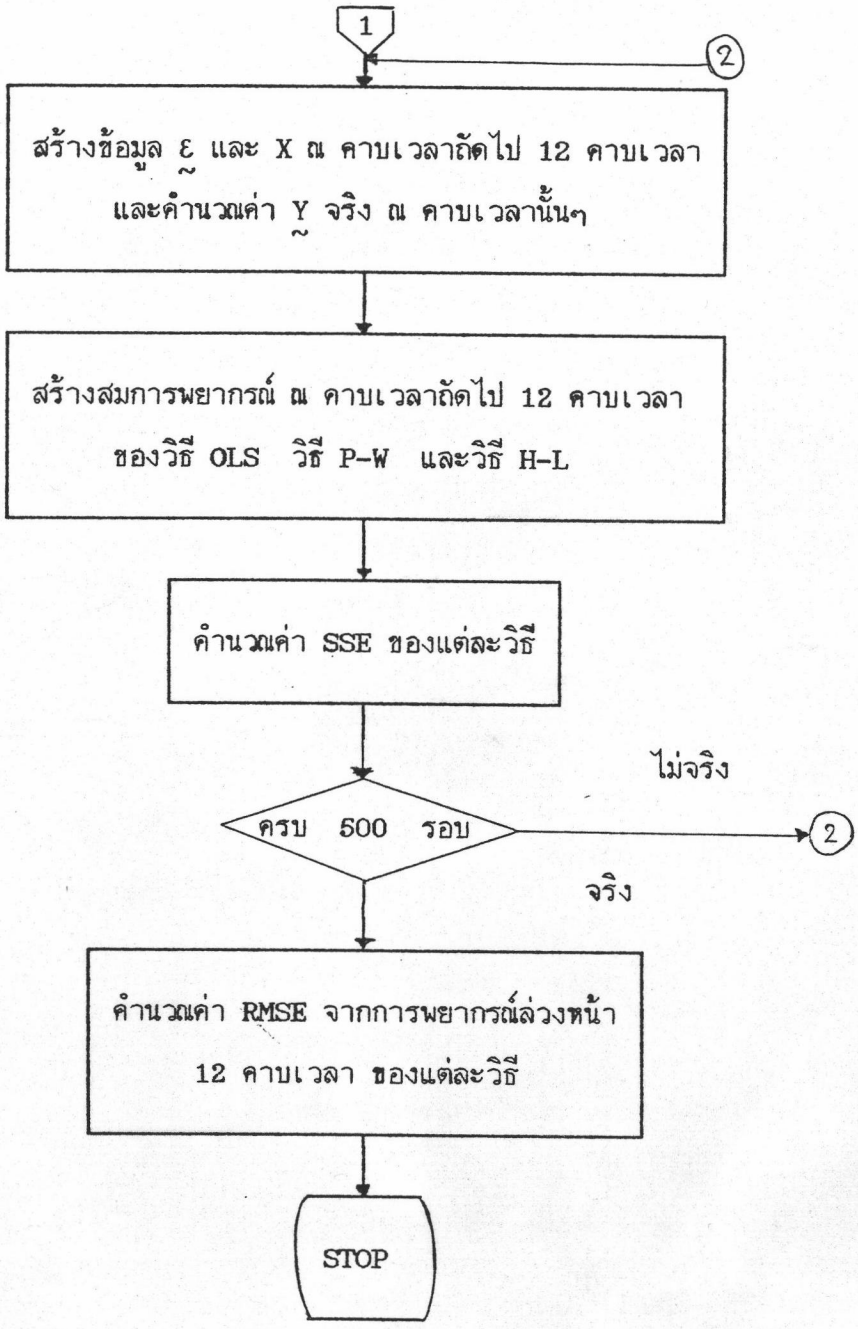
$$RMSE_t = \sqrt{\frac{\sum_{i=1}^{500} (Y_{ti} - \hat{Y}_{ti})^2}{500}} \quad , t = 1,2,\ldots,n$$

จากนั้นจะนำค่า $RMSE_t$ ของวิธีการทั้ง 3 วิธี มาทำการเปรียบเทียบ โดยจะเปรียบเทียบในแต่ละคาบเวลาทั้งหมด 12 คาบเวลา แล้วจึงเปรียบเทียบโดยเฉลี่ยทั้ง 12 คาบเวลาอีกครั้ง และจะทำการทดลองเช่นนี้โดยการเปลี่ยนรูปแบบตัวแปรอิสระ ขนาดตัวอย่าง และระดับสหสัมพันธ์ จนครบทุกรูปแบบที่ต้องการศึกษา

รูปที่ 3.4 แสดงผังงานสำหรับหาค่ารากที่สองของความคลาดเคลื่อนกำลังสองเฉลี่ยจากการพยากรณ์



รูปที่ 3.4(ต่อ)


1
2
สร้างข้อมูล ε และ X ณ คาบเวลาถัดไป 12 คาบเวลา
และคำนวณค่า Y จริง ณ คาบเวลานั้นๆ
สร้างสมการพยากรณ์ ณ คาบเวลาถัดไป 12 คาบเวลา
ของวิธี OLS วิธี P-W และวิธี H-L
คำนวณค่า SSE ของแต่ละวิธี
ไม่จริง
ครบ 500 รอบ
2
จริง
คำนวณค่า RMSE จากการพยากรณ์ล่วงหน้า
12 คาบเวลา ของแต่ละวิธี
STOP