บทที่ 3

วิธีดำเนินการวิจัย

ในการวิจัยครั้งนี้ต้องการเปรียบเทียบวิธีการประมาณค่าสูญหายในข้อมูลอนุกรมเวลา โดยศึกษาเปรียบเทียบวิธีการประมาณค่าสูญหาย 2 วิธี คือ วิธี Between-Forecast Estimation และ วิธี Fixed-Point Smoothing โดยทำการศึกษาภายใต้สถานการณ์ต่างๆ ดังนี้ (1) ค่าพารามิเตอร์ที่แตกต่างกัน ซึ่งถูกกำหนดโดยระดับความรุนแรงของอัตตสหสัมพันธ์ (Autocorrelation Function; ACF) และอัตตสหสัมพันธ์บางส่วน (Partial Autocorrelation Function ; PACF) และคุณสมบัติของรูปแบบที่ศึกษาแต่ละรูปแบบ (2) ขนาดตัวอย่างหรือขนาดอนุกรมเวลา 3 ขนาด คือ 50,75 และ100 (3) ตำแหน่งที่ข้อมูลสูญหาย ซึ่งถูกแบ่งออกเป็น 3 ช่วงคือ ต้น กลาง และปลาย (4) จำนวนข้อมูลสูญหาย คือ 1 ค่า และ 2 ค่า

เทคนิคที่ใช้ในการจำลองข้อมูลครั้งนี้อาศัยเทคนิคการจำลองแบบมอนติคาร์โล(Monte Carlo Simulation Method) สร้างสถานการณ์ต่าง ๆ ดังนั้นจึงขอกล่าวถึงวิธีการจำลองแบบมอนติคาร์โลก่อนที่จะแสดงถึงขั้นตอนการวิจัยและโปรแกรมสำหรับการวิจัยในลำดับต่อไป

3.1 วิธีการจำลองแบบมอนติคาร์โล (Monte Carlo Simulation Technique)

วิธีมอนติคาร์โลเป็นวิธีหนึ่งที่นิยมใช้กันอย่างแพร่หลายในปัจจุบันโดยอาศัยตัวเลขสุ่ม (Random Numbers) มาช่วยในการหาคำตอบของปัญหาที่ต้องการศึกษาขั้นตอนของมอนติคาร์โลที่ใช้กันในปัจจุบัน แบ่งได้เป็น 3 ขั้นตอนดังนี้

3.1.1 การสร้างตัวเลขสุ่ม (Random Number Generation)การสร้างตัวเลขสุ่มเป็นสิ่งที่สำคัญมากเนื่องจากหลักการของมอนติคาร์โลนั้นจะใช้ตัวเลขสุ่มมาช่วยในการหาคำตอบของปัญหาโดยลักษณะของตัวเลขสุ่มจะมีการแจกแจงแบบยูนิฟอร์มในช่วง (0,1)และเป็นอิสระกัน

3.1.2 การประยุกต์ปัญหาที่ต้องการศึกษามาใช้กับตัวเลขสุ่ม ซึ่งขั้นตอนนี้ขึ้นอยู่กับลักษณะของปัญหาที่ต้องการศึกษา บางปัญหาอาจจะไม่ใช้ตัวเลขสุ่มโดยตรง

3.1.3 การทดลองกระทำโดยใช้กระบวนการของการสุ่ม (Random process) มากระทำในลักษณะที่ซ้ำ ๆ กัน เพื่อหาคำตอบของปัญหาที่ต้องการศึกษา

3.2 แผนการทดลอง

เนื่องจากการวิจัยครั้งนี้มุ่งที่จะศึกษาความสามารถในการประมาณค่าสูญหายในข้อมูลอนุกรมเวลาของวิธีการทั้ง 2 วิธี ณ สถานการณ์ที่แตกต่างกันของรูปแบบอนุกรมเวลา 5 รูปแบบ ในระดับความรุนแรงของอัตตสหสัมพันธ์และอัตตสหสัมพันธ์บางส่วนที่แตกต่างกันในแต่ละรูปแบบ รวมทั้งสิ้น 19 สถานการณ์ ขนาดตัวอย่าง 3 ขนาด จำนวนข้อมูลสูญหายและตำแหน่งที่สูญหายแตกต่างกันรวมทั้งหมด 6 สถานการณ์ ซึ่งรวมทั้งสิ้น 342 สถานการณ์ โดยทำการเปรียบเทียบวิธีการประมาณค่าทั้ง 2 วิธีในทุกสถานการณ์ เพื่อหาวิธีการประมาณค่าสูญหายที่ดีที่สุดในแต่ละสถานการณ์

3.3 ขั้นตอนในการวิจัย

ขั้นตอนในการวิจัยแบ่งได้เป็น 5 ขั้นตอนคือ

1. สร้างค่าตัวแปร z_t ตามรูปแบบที่กำหนด ซึ่งมีด้วยกัน 5 รูปแบบคือ AR(1), AR(2), MA(1), MA(2) และ ARMA(1,1)
2. กำหนดตำแหน่งสูญหายของข้อมูล
3. ประมาณค่าพารามิเตอร์ของอนุกรมเวลาที่มีข้อมูลสูญหาย
4. ประมาณค่าสูญหายทั้ง 2 วิธี
5. หาค่าความคลาดเคลื่อนกำลังสองเฉลี่ยและทำการเปรียบเทียบ

ซึ่งแต่ละขั้นตอนมีรายละเอียดดังนี้

ศูนย์วิทยทรัพยากร
จุฬาลงกรณ์มหาวิทยาลัย

3.3.1 สร้างค่าตัวแปร z_t ตามรูปแบบที่กำหนด

3.3.1.1 สร้างค่าตัวแปร z_t ตามรูปแบบ AR(1) มีขั้นตอนในการสร้างดังนี้

1. สร้าง z_0 ให้มีการแจกแจงปกติ ค่าเฉลี่ย 0 ความแปรปรวน $\frac{\sigma^2}{1-\phi_1^2}$ และสร้าง a_t ; t=1,___,n+s ให้มีการแจกแจงปกติ ค่าเฉลี่ย 0 ความแปรปรวน $\sigma_a^2 = 100$ ซึ่งวิธีการแจกแจงปกติกระทำโดยวิธี Box-Muller (รายละเอียดในการแสดงไว้ในภาคผนวก ก)

2. จากนั้นสร้าง z_t ; t=1,___,n+s ให้มีรูปแบบความสัมพันธ์ คือ $z_t = \phi_1 z_{t-1} + a_t$ จะได้

$$z_1 = \phi_1 z_0 + a_1$$

$$z_2 = \phi_1 z_1 + a_2$$

$$\vdots$$

$$z_{n+s} = \phi_1 z_{n+s-1} + a_{n+s}$$

3.3.1.2 สร้างค่าตัวแปร z_t ตามรูปแบบ AR(2) มีขั้นตอนในการสร้างดังนี้

1. สร้าง z_0 และ z_1 ให้มีการแจกแจงปกติ ค่าเฉลี่ย 0 ความแปรปรวน $\frac{1-\phi_2}{1+\phi_2}\frac{\sigma^2}{(1-\phi_2)^2-\phi_1^2}$ และสร้าง a_t ; t=2,___,n+s ให้มีการแจกแจงปกติ ค่าเฉลี่ย 0 ความแปรปรวน $\sigma_a^2 = 100$

2. จากนั้นสร้าง z_t ; t=2,___,n+s ให้มีรูปแบบความสัมพันธ์ คือ $z_t = \phi_1 z_{t-1} + \phi_2 z_{t-2} + a_t$ จะได้

$$z_2 = \phi_1 z_1 + \phi_2 z_0 + a_2$$

$$z_3 = \phi_1 z_2 + \phi_2 z_1 + a_3$$

$$\vdots$$

$$z_{n+s} = \phi_1 z_{n+s-1} + \phi_2 z_{n+s-2} + a_{n+s}$$

3.3.1.3 สร้างค่าตัวแปร z_t ตามรูปแบบ MA(1) มีขั้นตอนในการสร้างดังนี้

ศูนย์วิทยทรัพยากร
จุฬาลงกรณ์มหาวิทยาลัย

1. สร้าง a_t ; t=0,___,n+s ให้มีการแจกแจงปกติ ค่าเฉลี่ย 0 ความแปรปรวน $\sigma_a^2 = 100$

2. จากนั้นสร้าง z_t; t=1,___,n+s ให้มีรูปแบบความสัมพันธ์ คือ $z_t = a_t - \theta_1 a_{t-1}$ จะได้

$$z_1 = a_1 - \theta_1 a_0$$

$$z_2 = a_2 - \theta_1 a_1$$

$$\vdots$$

$$z_{n+s} = a_{n+s} - \theta_1 a_{n+s-1}$$

3.3.1.4 <u>สร้างค่าตัวแปร z_t ตามรูปแบบ MA(2)</u> มีขั้นตอนในการสร้างดังนี้

1. สร้าง a_t ; t=0,___,n+s ให้มีการแจกแจงปกติ ค่าเฉลี่ย 0 ความแปรปรวน $\sigma_a^2 = 100$

2. จากนั้นสร้าง z_t; t=2,___,n+s ให้มีรูปแบบความสัมพันธ์ คือ $z_t = a_t - \theta_1 a_{t-1} - \theta_2 a_{t-2}$ จะได้

$$z_2 = a_2 - \theta_1 a_1 - \theta_2 a_0$$

$$z_3 = a_3 - \theta_1 a_2 - \theta_2 a_1$$

$$\vdots$$

$$z_{n+s} = a_{n+s} - \theta_1 a_{n+s-1} - \theta_2 a_{n+s-2}$$

3.3.1.5 <u>สร้างค่าตัวแปร z_t ตามรูปแบบ ARMA(1,1)</u> มีขั้นตอนในการสร้างดังนี้

1. สร้าง z_0 ให้มีการแจกแจงปกติ ค่าเฉลี่ย 0 ความแปรปรวน $\frac{(1+\theta_1^2-2\phi_1\theta_1)\sigma^2}{1-\phi_1^2}$ และสร้าง a_t ; t=0,___,n+s ให้มีการแจกแจงปกติค่าเฉลี่ย 0 ความแปรปรวน $\sigma_a^2 = 100$

2. จากนั้นสร้าง z_t; t=1,___,n+s ให้มีรูปแบบความสัมพันธ์ คือ $z_t = \phi_1 z_{t-1} + a_t - \theta_1 a_{t-1}$ จะได้

$$z_1 = \phi_1 z_0 + a_1 - \theta_1 a_0$$

$$z_2 = \phi_1 z_1 + a_2 - \theta_1 a_1$$

$$\vdots$$

ศูนย์วิทยทรัพยากร
จุฬาลงกรณ์มหาวิทยาลัย

$$z_{n+s} = \phi_1 z_{n+s-1} + a_{n+s} - \theta_1 a_{n+s-1}$$

หลังจากสร้างตัวแปร z_t ตามรูปแบบที่กำหนดใน 3.3.1.1-3.3.1.5 ก็จะทำการตัดค่า z_t จำนวน s ตัวแรกทิ้ง (ในการศึกษาครั้งนี้ s = 200) เพื่อไม่ให้ข้อมูลที่ได้ขึ้นกับค่าเริ่มต้นที่กำหนดไว้ ดังนั้นจะได้ z_t จำนวน n ตัวตามต้องการ

3.3.2 กำหนดตำแหน่งสูญหายของข้อมูล

1. แบ่งช่วงอนุกรมเวลาออกเป็น 3 ช่วงเท่า ๆกัน ได้ดังนี้

ขนาดตัวอย่าง	ต้น	กลาง	ท้าย
50	1-17	18-34	35-50
75	1-25	26-50	51-75
100	1-34	35-67	68-100

2. กำหนดตำแหน่งสูญหายโดยอาศัยตัวเลขสุ่มที่มีการแจกแจงยูนิฟอร์มในช่วง (a,b) โดย a : คือตำแหน่งแรกของช่วง

b : คือตำแหน่งสุดท้ายของช่วง

เช่น ต้องการหาตำแหน่งสูญหายของช่วงต้น และ n=50

ถ้า m=1 ตำแหน่งสูญหายคือตัวเลขสุ่มที่ได้จากการแจกแจงยูนิฟอร์มในช่วง (1,17)

m=2 ตำแหน่งสูญหายตำแหน่งแรกคือตัวเลขสุ่มที่ได้จากการแจกแจงยูนิฟอร์มในช่วง (1,17) และตำแหน่งสูญหายตำแหน่งที่ 2 คือตำแหน่งถัดไป

3.3.3 ประมาณค่าพารามิเตอร์ของอนุกรมเวลา

นำข้อมูล z_t ที่มีข้อมูลสูญหายมาประมาณค่าพารามิเตอร์ตามรูปแบบที่กำหนดได้โดยเทคนิคของบ็อกซ์-เจนกินส์ ด้วยโปรแกรมสำเร็จรูป SPSSX

3.3.4 การประมาณค่าสูญหายทั้ง 2 วิธี

หลังจากการประมาณค่าพารามิเตอร์ให้กับอนุกรมเวลาแล้ว ขั้นต่อไปคือการประมาณค่าสูญหายทั้ง 2 วิธีคือ

ศูนย์วิทยทรัพยากร
จุฬาลงกรณ์มหาวิทยาลัย

1. <u>วิธี Between-Forecast Estimation</u> มีขั้นตอนดังนี้

1.1 นำข้อมูล $z_1,\ldots,z_r$ คืออนุกรมเวลาช่วงก่อนข้อมูลสูญหาย มาทำการพยากรณ์ (forecast) ค่าสูญหาย จะได้ $z_r(l)$; $l=1,\ldots,m$ ตามหลักการพยากรณ์ของบ็อกซ์-เจนกินส์

1.2 นำข้อมูล $z_{r+m+1},\ldots,z_T$; T = ขนาดอนุกรมเวลา(n) คืออนุกรมเวลาช่วงหลังข้อมูลสูญหายมาทำการพยากรณ์ย้อนหลัง (backforecast) จะได้ $z_{r+m+1}(m+1-l)$; $l=1,\ldots,m$ ตามหลักการพยากรณ์ย้อนหลังของบ็อกซ์-เจนกินส์

1.3 หาค่าถ่วงน้ำหนัก c_l และ d_l ; $l=1,\ldots,m$ ตามรูปแบบและจำนวนข้อมูลสูญหาย

1.4 ประมาณค่าสูญหาย $z_{r,r+m+1}(l)$ ซึ่ง

$$z_{r,r+m+1}(l) = c_l z_r(l) + d_l z_{r+m+1}(m+1-l) \quad ; \quad l=1,\ldots,m$$

2. <u>วิธี Fixed-Point Smoothing</u> มีขั้นตอนดังนี้

2.1 สร้างเมตริกซ์ F , G , Z(t/t) และ P(t/t) ; t=0 ตามรูปแบบและค่าพารามิเตอร์ที่ได้จากขั้นตอนที่ 3.3.2

2.2 ขั้นตอนนี้เป็นขั้นตอนของการประมาณค่าของอนุกรมเวลาเมื่อเวลาเปลี่ยนไปในแต่ละหน่วยเวลา ซึ่งมีด้วยกัน 2 ขบวนการดังนี้

2.2.1 คำนวณ Z(t+1/t),P(t+1/t), ε(t+1) และ y(t+1) ; $t=1,\ldots,n$ ตามเทคนิคการย้อนกลับของ Kalman (Kalman recursive)

2.2.2 ขบวนการนี้เริ่มเมื่อขั้นตอนที่ 2.2.1 ทำการคำนวณจนกระทั่งพบข้อมูลสูญหาย ขบวนการนี้จะทำการประมาณค่า $y(\tau)$; $\tau = \tau_1,\ldots,\tau_m$ ด้วย $y(\tau/t)$ โดยคำนวณค่า $y(\tau/t)$; $t = r+1,\ldots,T$ ตามหลักการ Fixed-Point Smoothing ดังที่ได้กล่าวแล้วในบทที่ 2

3.3.5 <u>หาค่าความคลาดเคลื่อนกำลังสองเฉลี่ยและทำการเปรียบเทียบ</u>

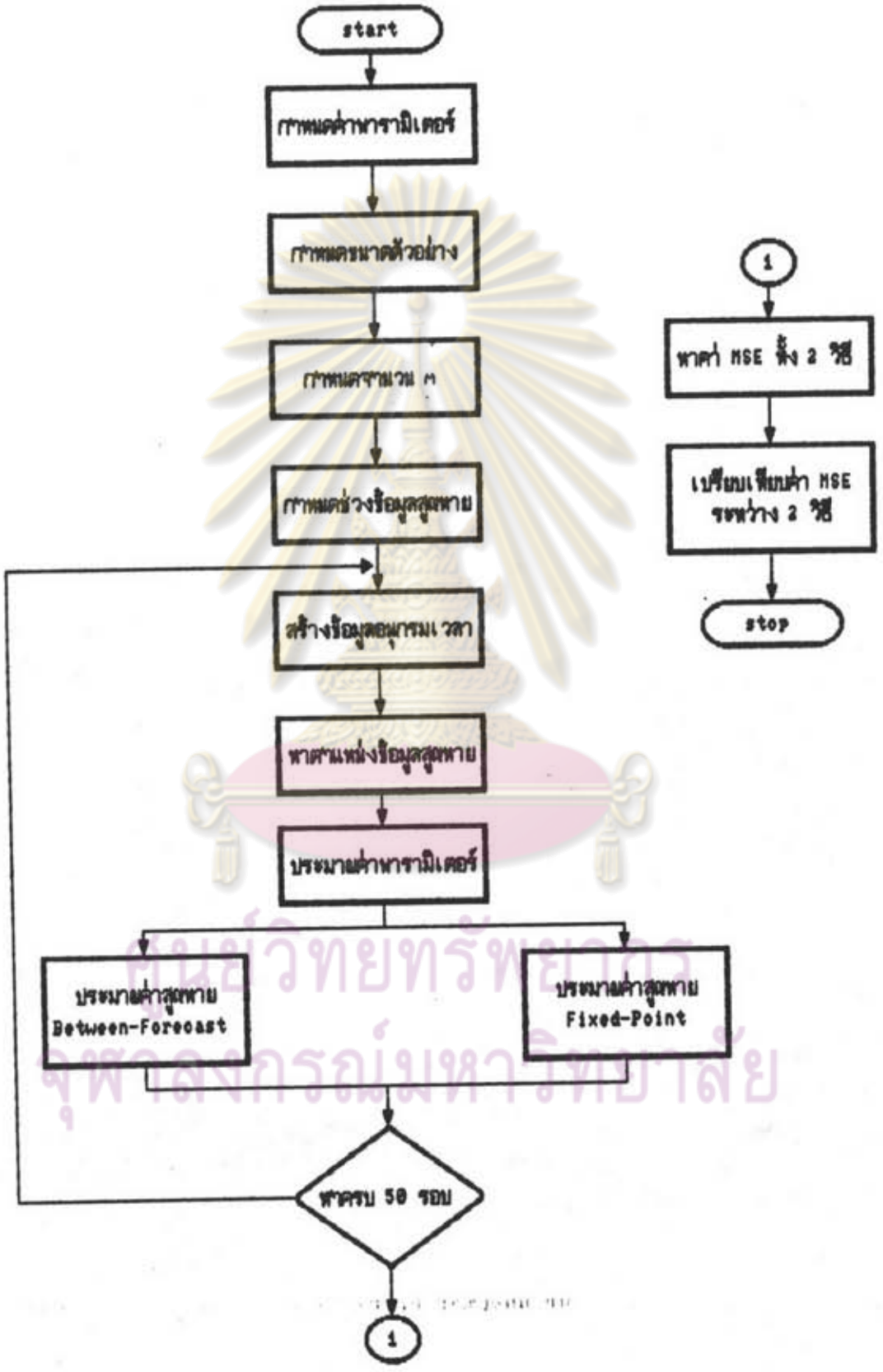
หลังจากการคำนวณค่าประมาณได้จากทั้ง 2 วิธีแล้ว จึงนำค่าประมาณกับค่าจริงมาหาผลต่างเพื่อคำนวณค่าความคลาดเคลื่อนกำลังสองเฉลี่ยของทั้ง 2 วิธีการในสถานการณ์นั้นๆ จนครบทุกสถานการณ์

จากขั้นตอนต่าง ๆ ที่ได้กล่าวมาแล้วข้างต้นสามารถสรุปเป็นผังงานได้ดังรูปที่ 3.1

ศูนย์วิทยทรัพยากร
จุฬาลงกรณ์มหาวิทยาลัย

หอสมุดกลาง สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

รูปที่ 3.1 แสดงผังงานของขั้นตอนการดำเนินงาน



ศูนย์วิทยทรัพยากร
จุฬาลงกรณ์มหาวิทยาลัย