

บทที่ 3

วิธีดำเนินการวิจัย

การวิจัยครั้งนี้ต้องการเปรียบเทียบความคลาดเคลื่อนกำลังสองเฉลี่ยของค่าพยากรณ์สำหรับการวิเคราะห์ความถดถอยพหุนามแบบลำดับขั้นของตัวแปรอิสระที่มีกำลังสูงสุดไม่เกิน 2 การสร้างตัวแบบมี 3 วิธีดังนี้

1. การสร้างตัวแบบด้วยวิธีกำจัดตัวแปรอิสระย้อนหลัง
2. การสร้างตัวแบบด้วยวิธีการถดถอยขั้นบันได
3. การสร้างตัวแบบด้วยวิธีของเบส์เซียน

การหาข้อสรุปโดยใช้การจำลองข้อมูลด้วยวิธีมอนติคาร์โล (Simulation by Monte Carlo Method)¹

วิธีมอนติคาร์โลเป็นเทคนิคในการจำลองตัวแบบทางคณิตศาสตร์ ซึ่งหลักการเบื้องต้นนั้นต้องจำลองตัวเลขสุ่ม (random number) เพื่อช่วยในการหาคำตอบของปัญหาที่ต้องการศึกษา โดยในการศึกษาวิจัยนี้จะใช้เทคนิคมอนติคาร์โลในการสร้างข้อมูลที่มีสภาพการแจกแจงตามที่ต้องการศึกษา ขั้นตอนที่สำคัญของการจำลองข้อมูลด้วยวิธีมอนติคาร์โลมี 3 ขั้นตอน คือ

ขั้นตอนที่ 1 การสร้างตัวเลขสุ่ม

การสร้างตัวเลขสุ่มเป็นสิ่งสำคัญมากในวิธีมอนติคาร์โล ทั้งนี้ก็เพราะว่าหลักการของมอนติคาร์โลนั้นจะใช้ตัวเลขสุ่มมาช่วยในการหาคำตอบของปัญหา ลักษณะของตัวเลขสุ่มที่ดีจะมีการแจกแจงสม่ำเสมอ (Uniform Distribution) ในช่วง $[0,1]$ และเป็นอิสระซึ่งกันและกัน

ขั้นตอนที่ 2 การประยุกต์ปัญหาที่ต้องการศึกษามาใช้กับตัวเลขสุ่ม

ขั้นตอนนี้นับอยู่กับลักษณะของปัญหาที่ต้องการศึกษา บางปัญหาอาจใช้ตัวเลขสุ่มโดยตรงแต่บางปัญหาอาจใช้ตัวเลขสุ่มเพียงบางขั้นตอนของปัญหาเท่านั้น

ขั้นตอนที่ 3 การทดลองกระทำ

เมื่อประยุกต์ปัญหาเพื่อใช้กับตัวเลขสุ่มได้แล้วขั้นตอนต่อไปก็คือ การทดลองโดยใช้กระบวนการสุ่ม (random process) มากกระทำในลักษณะที่ซ้ำ ๆ กัน (replication) เพื่อหาคำตอบของปัญหาที่ต้องการศึกษา

¹จะเด็จ สวรรค์ตรานนท์. “การเปรียบเทียบวิธีที่ใช้สำหรับการเลือกสมการถดถอยที่ดีที่สุด.”

(วิทยานิพนธ์ปริญญาโท สาขาวิชาสถิติ บัณฑิตวิทยาลัย จุฬาลงกรณ์มหาวิทยาลัย, 2531.) หน้า 36.

แผนการทดลอง

ผู้วิจัยกำหนดสถานการณ์ในการหาความคลาดเคลื่อนกำลังสองเฉลี่ยของตัวแบบทั้ง 3 วิธีดังนี้

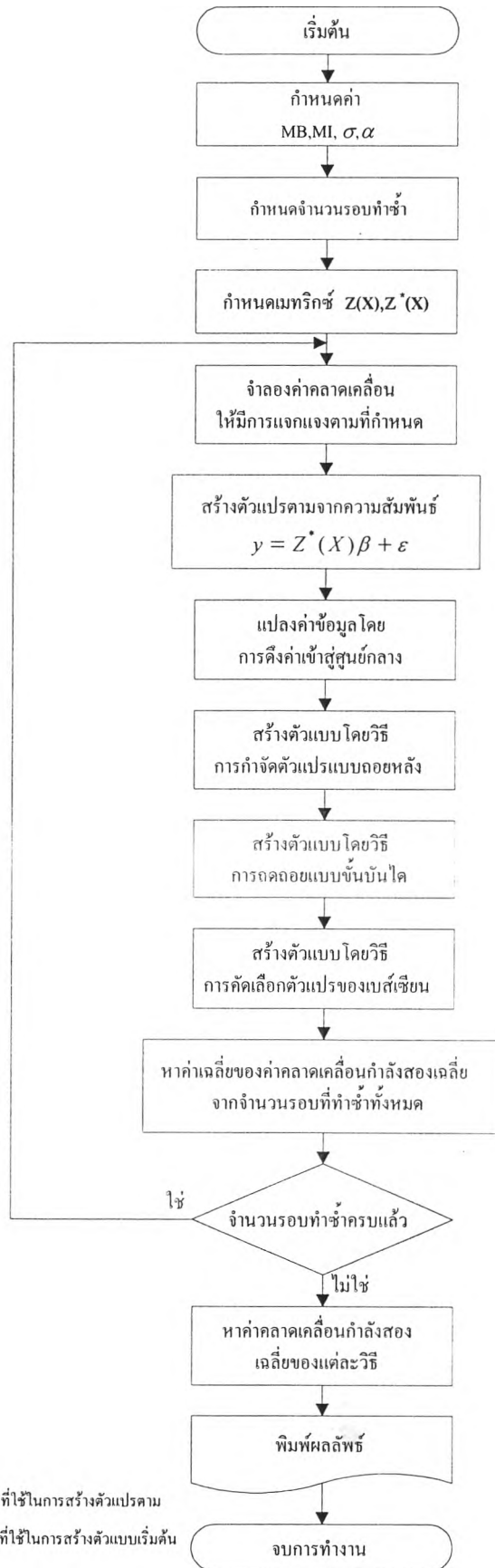
1. เลือกตัวอย่างสุ่มจากประชากรที่มีการแจกแจงเดียวกัน การแจกแจงที่สนใจศึกษา คือ การแจกแจงแบบปกติที่มีค่าเฉลี่ย 0 และส่วนเบี่ยงเบนมาตรฐาน 5,10,20 และ 25
2. กำหนดขนาดตัวอย่างที่ศึกษา คือ 25,50,75 และ 100
3. ตัวแปรอิสระเริ่มต้นสร้างจากการแจกแจงปกติค่าเฉลี่ย 0 ความแปรปรวน 1
4. ระดับนัยสำคัญ คือ 0.05 และ 0.01
5. จำนวนตัวแปรอิสระสูงสุดที่ใช้สำหรับการสร้างตัวแปรตามที่เหมาะสม (MB) คือ 1,2,3,4,5 และ 6
6. จำนวนตัวแปรอิสระสูงสุดที่ใช้สำหรับการสร้างตัวแปรตามในตัวแบบเริ่มต้น (MI) คือ 1,2,3,4,5 และ 6

ขั้นตอนในการศึกษาวิจัย

ขั้นตอนในการศึกษาวิจัยมี ดังนี้ คือ

1. กำหนดลักษณะการแจกแจงของค่าคลาดเคลื่อน ขนาดตัวอย่าง จำนวนตัวแปรอิสระสูงสุดที่ใช้สำหรับการสร้างตัวแปรตามที่เหมาะสม จำนวนตัวแปรอิสระกำลังสูงสุดที่ใช้สำหรับการสร้างตัวแปรตามในตัวแบบเริ่มต้น และระดับนัยสำคัญ
2. สร้างข้อมูลตัวแปรอิสระตามลักษณะการแจกแจงที่ต้องการศึกษา, ข้อมูลตัวพยากรณ์ที่เป็นพหุนามแบบลำดับขั้นของตัวแปรอิสระ และข้อมูลตัวแปรตาม (y) ที่สร้างจากตัวพยากรณ์ที่เป็นพหุนามแบบลำดับขั้นของตัวแปรอิสระ และค่าคลาดเคลื่อนที่มีลักษณะการแจกแจงตามที่ต้องการศึกษา (ε) โดยให้ตัวแปรตามมีความสัมพันธ์เชิงเส้นในพารามิเตอร์กับตัวพยากรณ์
3. ทำการแปลงข้อมูลโดยการดึงค่าเข้าสู่ศูนย์กลาง (Centering)
4. กำหนดตัวแบบเริ่มต้นที่ใช้ในการค้นหาตัวแบบที่เหมาะสม
5. การสร้างตัวแบบโดยวิธีทั้ง 3 วิธี
 - ก) ตัวแปรอิสระย้อนหลัง (Backward Elimination Method)
 - ข) วิธีการถดถอยขั้นบันได (Stepwise Regression Method)
 - ค) วิธีคัดเลือกตัวแปรของเบย์เซียน (Bayesian Variable Selection Method)
6. การหาค่าคลาดเคลื่อนกำลังสองเฉลี่ยของแต่ละตัวแบบ
7. สรุปผลในรูปของตาราง

ผังงานแสดงขั้นตอนในงานวิจัยแสดงดังรูปที่ 3



รูปที่ 3 แสดงผังงานแสดงขั้นตอนในงานวิจัย

สำหรับรายละเอียดของแต่ละขั้นตอนเป็นดังนี้

ขั้นตอนที่ 1

กำหนดลักษณะการแจกแจงของค่าคลาดเคลื่อน ขนาดตัวอย่าง จำนวนตัวแปรอิสระสูงสุดที่ใช้สำหรับการสร้างตัวแปรตามที่เหมาะสม จำนวนตัวแปรอิสระกำลังสูงสุดที่ใช้สำหรับการสร้างตัวแปรตามในตัวแบบเริ่มต้น ระดับนัยสำคัญ จะกำหนดตามแผนการทดลองที่ได้กล่าวมาแล้วข้างต้น

ขั้นตอนที่ 2

การสร้างข้อมูลที่มีลักษณะการแจกแจงที่ต้องการศึกษา จะใช้โปรแกรมภาษาฟอร์แทรน 77 (Fortran 77) บนเครื่อง Sun Sparcwork 2000E ซึ่งการสร้างข้อมูลที่มีลักษณะการแจกแจงแบบต่าง ๆ นั้น จะต้องใช้ตัวเลขสุ่มเป็นพื้นฐานในการสร้างฟังก์ชันที่ใช้ในการสร้างตัวเลขสุ่ม คือ RD(X) เมื่อสร้างตัวเลขสุ่มแล้วจะนำตัวเลขสุ่มที่ได้มา สร้างข้อมูลที่มีการแจกแจงแบบปกติมีค่าเฉลี่ย 0 และความแปรปรวน σ^2 โดยสร้างได้จากโปรแกรมย่อย NORMAL (DMEAN,SIGMA,X)¹ ข้อมูลที่สร้างได้แก่

- 1) ตัวแปรอิสระเริ่มต้น คือ สร้างจากการแจกแจงแบบปกติที่มีค่าเฉลี่ย 0 ความแปรปรวน 1
- 2) ตัวพยากรณ์ที่เป็นทอมพหุนามแบบลำดับชั้นของตัวแปรอิสระเป็นเมทริกซ์ที่สร้างได้จากตัวแปรอิสระเริ่มต้นที่สร้างมาแล้วในตอนต้น และจำนวนตัวแปรอิสระสูงสุดที่ใช้ในการสร้างตัวแปรตาม (MB) ที่กำหนดไว้ในขั้นตอนที่ 1
- 3) ตัวแปรตามให้มีความสัมพันธ์ในเชิงเส้นตรงในพารามิเตอร์กับตัวพยากรณ์ที่ใช้ในการสร้างตัวแปรตาม โดยมีลักษณะการแจกแจงค่าคลาดเคลื่อนตามที่กำหนดในตอนต้น ซึ่งมีรูปแบบดังนี้

$$y = Z(X)\beta + \varepsilon$$

เมื่อ y แทนตัวแปรตาม

$Z(X)$ แทนเมทริกซ์ของตัวพยากรณ์ซึ่งเป็นทอมพหุนามแบบลำดับชั้นของตัวแปรอิสระ

β แทนค่าพารามิเตอร์ที่กำหนด

และ ε แทนค่าคลาดเคลื่อนที่กำหนด ซึ่ง $\varepsilon \sim N(0, \sigma^2 I_n)$

ขั้นตอนที่ 3

ทำการแปลงข้อมูลโดยดึงค่าเข้าสู่ศูนย์กลาง (centering) ซึ่งทำได้โดยการลบค่าของตัวพยากรณ์แต่ละตัวด้วยค่าเฉลี่ยของตัวพยากรณ์นั้น ๆ

¹ ปราณี รัตนัง. “การประมาณสัมประสิทธิ์การถดถอยพหุเมื่อความผิดพลาดมีการแจกแจงแบบเบ้และมีการแจกแจงแบบหางยาวกว่าการแจกแจงปกติ.” (วิทยานิพนธ์ปริญญาโทมหาบัณฑิต ภาควิชาสถิติ จุฬาลงกรณ์มหาวิทยาลัย, 2531) หน้า 117

ขั้นตอนที่ 4

กำหนดตัวแปรเริ่มต้นที่จะทำการค้นหาตัวแปรที่เหมาะสมซึ่งสร้างจากตัวพยากรณ์ที่ได้จากขั้นตอนที่ 2 และจำนวนตัวแปรอิสระสูงสุดที่ใช้ในการสร้างตัวแปรเริ่มต้น (MI) ที่กำหนดในขั้นตอนที่ 1

ขั้นตอนที่ 5

หลังจากที่มีข้อมูลพร้อมแล้วจะทำการสร้างตัวแปรที่เหมาะสมจากทั้ง 3 วิธีด้วยโปรแกรมย่อยดังต่อไปนี้

- ก) โปรแกรมย่อยของวิธีกำจัดตัวแปรแบบถอยหลัง
BACK(N,K,Z,Y,PI,IALP,SSE1,AMSE1,AMSEP1)
- ข) โปรแกรมย่อยของวิธีการถดถอยแบบขั้นบันได
STEP(N,K,Z,Y,PI,IALP,SSE2,AMSE2,AMSEP2)
- ค) โปรแกรมย่อยของวิธีคัดเลือกตัวแปรของเบส์เซียน
BAYE(N,K,PI,Z,Y,IALP,AMSE3,AMSEP3,IP,C)

วิธีการสร้างตัวแปรในแต่ละวิธีมีรายละเอียดดังนี้

ก) การสร้างตัวแปรด้วยวิธีกำจัดตัวแปรแบบถอยหลัง ตัวแปรเริ่มต้นจะมีตัวพยากรณ์ทั้งหมดอยู่ในสมการ และจะตัดตัวพยากรณ์ที่มีค่าสถิติเอฟบางส่วน (Partial F) น้อยที่สุดและไม่มีนัยสำคัญออก ซึ่งค่าสถิติเอฟบางส่วนหาได้ดังนี้

$$F(z_j | z_1, \dots, z_{j-1}, z_{j+1}, \dots, z_k) = \frac{SSR(z_j | z_1, \dots, z_{j-1}, z_{j+1}, \dots, z_k)}{SSE(z_1, \dots, z_k)}$$

$$= \frac{SSR(z_1, \dots, z_k) - SSR(z_1)}{SSE(z_1, \dots, z_k)}$$

เมื่อ $SSR(z_j | z_1, \dots, z_{j-1}, z_{j+1}, \dots, z_k)$ แทนผลบวกกำลังสองของการถดถอยบน z_j โดยที่ $z_1, \dots, z_{j-1}, z_{j+1}, \dots, z_k$ อยู่ในตัวแปรแล้ว

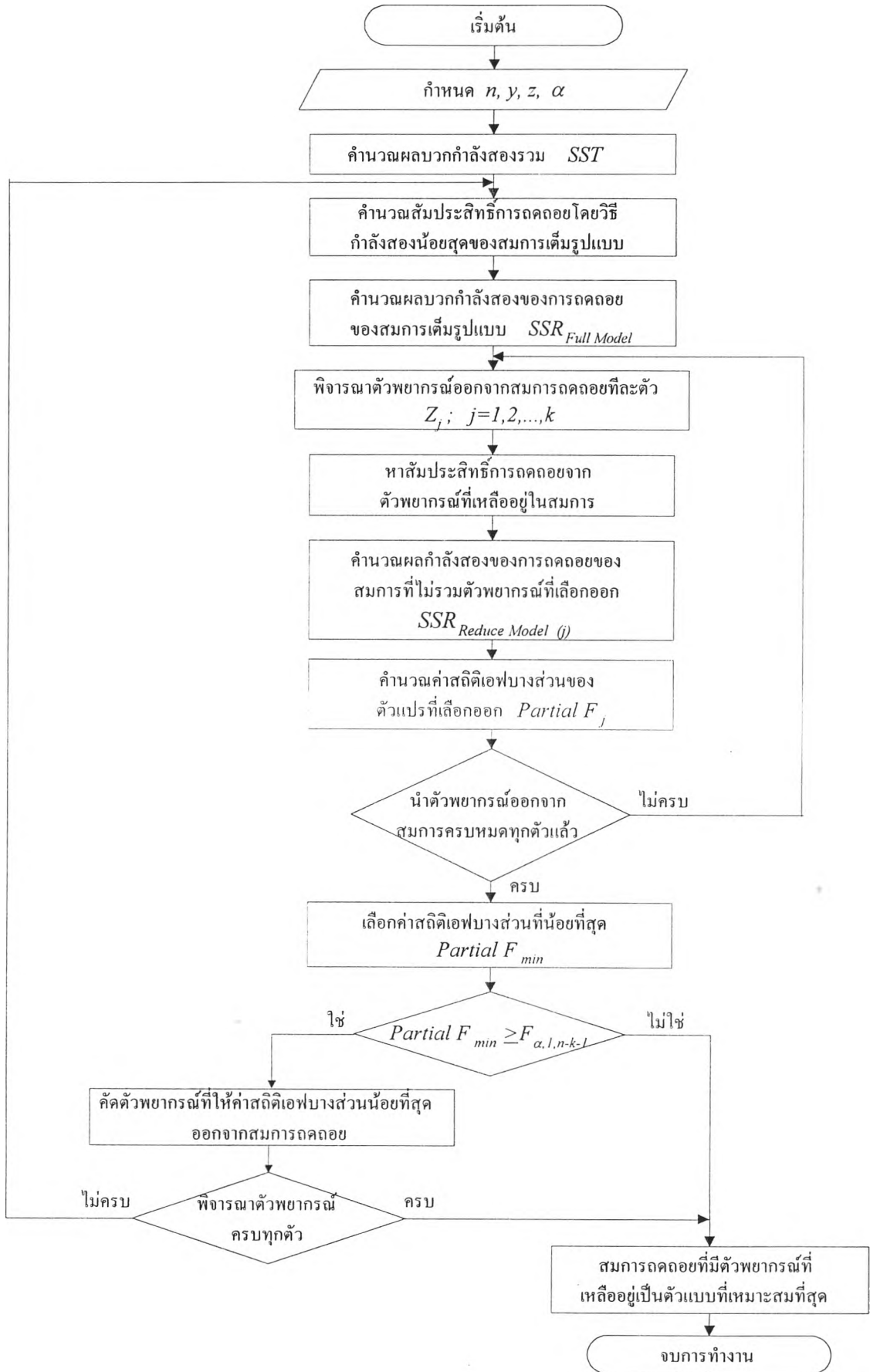
$SSR(z_1, \dots, z_k)$ แทนผลบวกกำลังสองของการถดถอยบน z_1, \dots, z_k

$SSR(z_j)$ แทนผลบวกกำลังสองของการถดถอยบน z_j

และ $SSE(z_1, \dots, z_k)$ แทนผลบวกกำลังสองของค่าคลาดเคลื่อนเมื่อ z_1, \dots, z_k อยู่ในตัวแปร

เมื่อตัดตัวพยากรณ์ไปแล้วจะคำนวณค่าสัมประสิทธิ์สหสัมพันธ์ของตัวพยากรณ์ที่ยังเหลืออยู่ จากนั้นคำนวณค่าสถิติเอฟบางส่วนของตัวพยากรณ์ที่อยู่ในสมการใหม่แล้วทดสอบนัยสำคัญ ทำซ้ำเช่นนี้จนกระทั่งค่าสถิติเอฟบางส่วนที่น้อยที่สุดมีนัยสำคัญ หรือไม่มีตัวพยากรณ์เหลือในตัวแปร ระดับนัยสำคัญที่ใช้ คือ 0.05 และ 0.01

ผังงานแสดงวิธีการสร้างตัวแปรด้วยวิธีกำจัดตัวแปรแบบถอยหลังแสดงดังรูปที่ 4

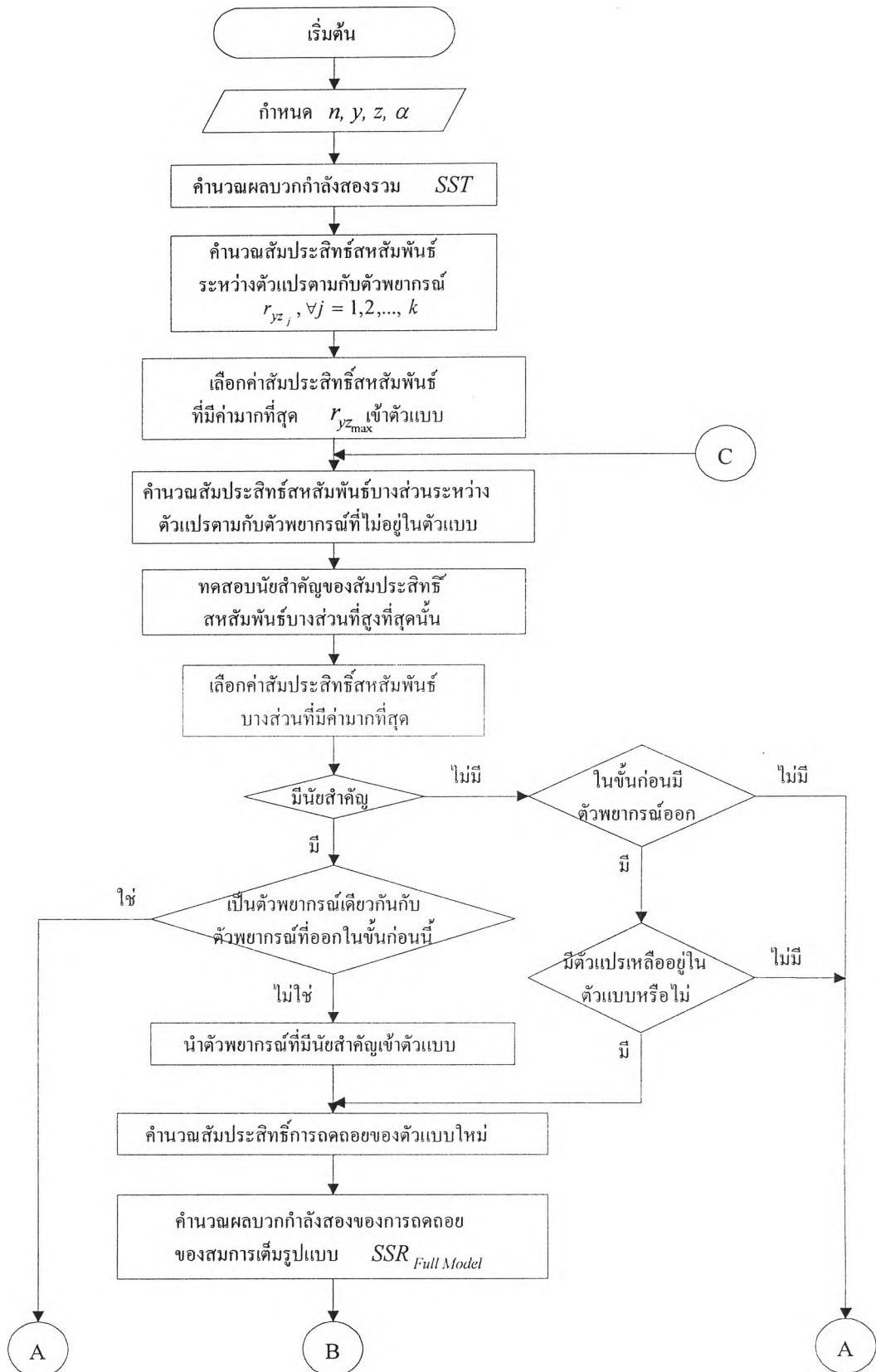


รูปที่ 4 แสดงผังการสร้างตัวแบบด้วยวิธีกำจัดตัวแปรแบบถดถอยหลัง

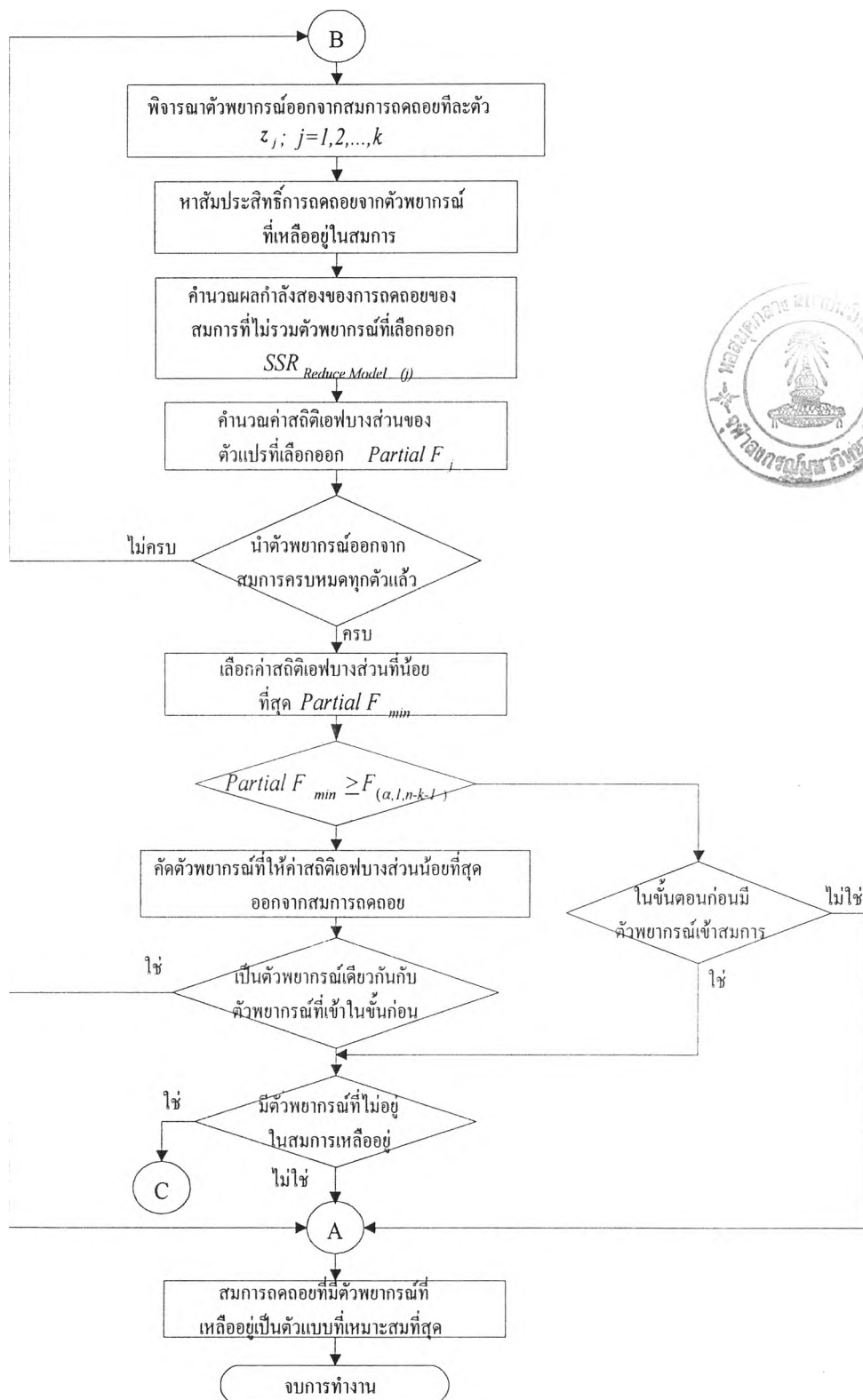
ข) วิธีการสร้างตัวแบบด้วยวิธีการถดถอยขั้นบันได จะเริ่มจากการคัดเลือกตัวพยากรณ์ที่มีความสัมพันธ์กับตัวแปรตาม (r_{zy}) มากที่สุดเข้าสู่สมการถดถอย โดยที่ $r_{zy}^2 = 1 - \frac{SSE}{SST}$ จากนั้นจะหาสัมประสิทธิ์สหสัมพันธ์บางส่วนระหว่างตัวแปรตามกับตัวพยากรณ์แต่ละตัวที่ยังไม่อยู่ในสมการ โดยถือว่าได้รวมตัวพยากรณ์ z_j ไว้ในตัวแบบแล้ว และเลือกตัวพยากรณ์ที่มีค่าสัมประสิทธิ์สหสัมพันธ์บางส่วนที่สูงที่สุดและมีนัยสำคัญเข้าสู่ตัวแบบ ถัดมาเป็นขั้นตอนการคัดเลือกตัวพยากรณ์ออก โดยจะตัดตัวพยากรณ์ที่มีค่าสถิติเอฟบางส่วนน้อยที่สุดและไม่มีนัยสำคัญออกจากตัวแบบ จะทำซ้ำการคัดเลือกตัวพยากรณ์เข้าและออก โดยมีเงื่อนไขการหยุดเมื่อข้อใดข้อหนึ่งต่อไปนี้เป็นจริง

- 1) ไม่มีตัวพยากรณ์ที่สามารถเข้าหรือออกจากตัวแบบในขั้นถัดกัน
- 2) ตัวพยากรณ์ที่เข้าและออกจากตัวแบบในขั้นถัดกันเป็นตัวเดียวกัน
- 3) ในขั้นตอนการออกไม่มีตัวพยากรณ์เหลืออยู่ในตัวแบบ
- 4) ในขั้นตอนการเข้าไม่เหลือตัวพยากรณ์ที่ไม่อยู่ในสมการ

ผังงานแสดงวิธีการสร้างตัวแบบด้วยวิธีการถดถอยขั้นบันได แสดงดังรูปที่ 5



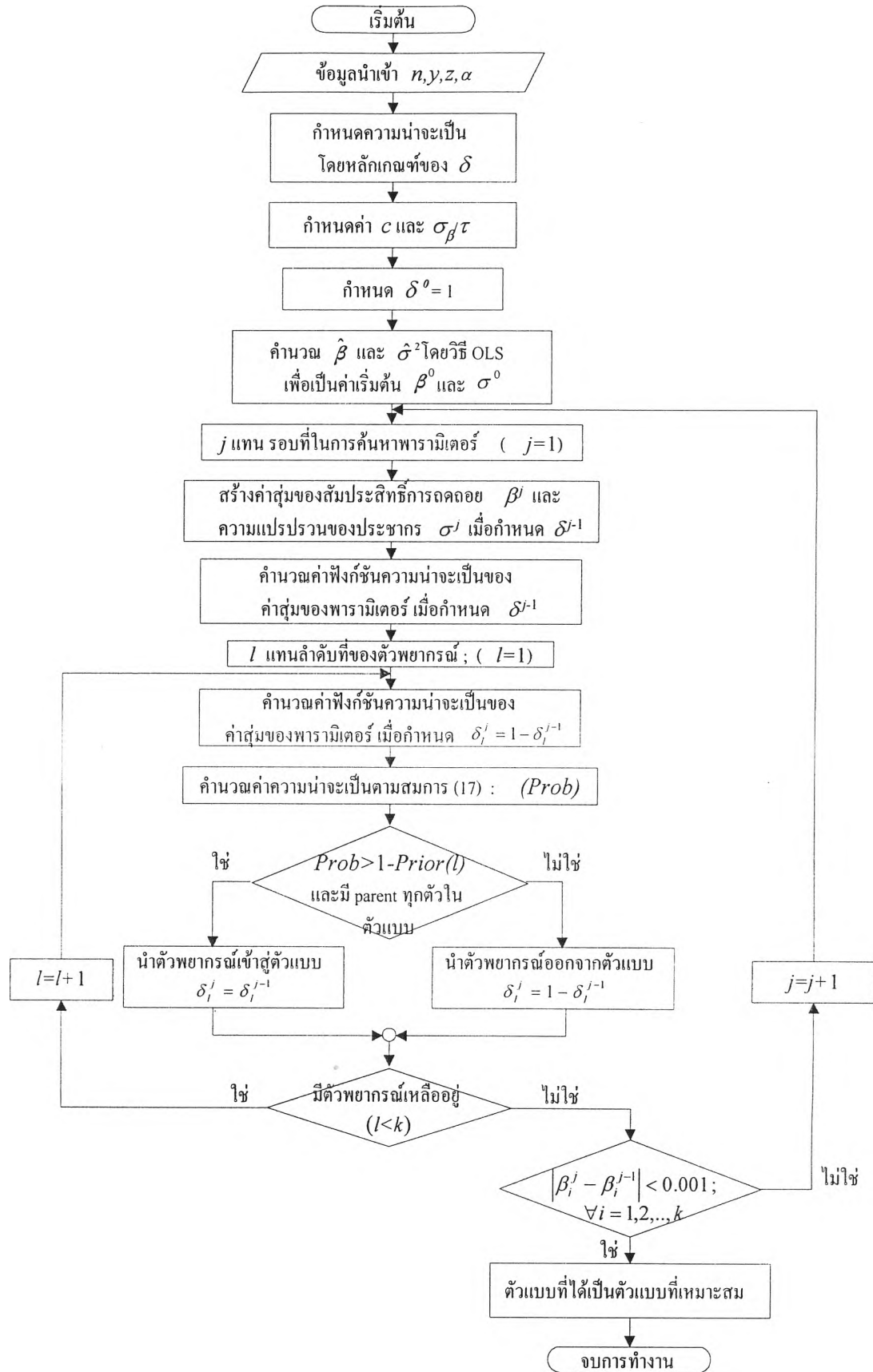
รูปที่ 5 แสดงผังงานแสดงการสร้างตัวแบบด้วยวิธีถดถอยขั้นบันได



รูปที่ 5 (ต่อ) แสดงผังงานแสดงการสร้างตัวแบบด้วยวิธีลดหย่อนขั้นบันได

ค) วิธีการสร้างตัวแบบโดยวิธีคัดเลือกตัวแปรของเบส์เซียน ตัวแบบเริ่มต้นจะมีลักษณะเป็นตัวแบบเต็มรูป คือมีตัวพยากรณ์ทั้งหมดอยู่ในสมการ จากนั้นคำนวณค่าประมาณของสัมประสิทธิ์การถดถอยและค่าความคลาดเคลื่อนกำลังสองเฉลี่ยของตัวแบบเต็มรูปโดยวิธีกำลังสองน้อยที่สุด เพื่อเป็นค่าเริ่มต้นที่จะใช้ในการสุ่มค่าพารามิเตอร์ β และ σ^2 ทำการสุ่มค่าพารามิเตอร์ β และ σ^2 จากการแจกแจงที่ได้กล่าวมาแล้วข้างต้น หลังจากนั้นจึงคำนวณหาฟังก์ชันความน่าจะเป็นของค่าสุ่มที่ได้ เพื่อนำไปหาค่าความน่าจะเป็นแบบมีเงื่อนไขภายหลังจากที่ทราบข้อมูลแล้ว เกณฑ์ในการคัดเลือกจะพิจารณาจากความน่าจะเป็นโดยหลักเกณฑ์ที่ตัวพยากรณ์แต่ละตัวจะมีโอกาสอยู่ในตัวแบบซึ่งกล่าวไว้แล้วในบทที่ 2 การพิจารณาตัวพยากรณ์ที่จะเข้าสู่สมการมีเกณฑ์ดังนี้ ถ้าความน่าจะเป็นแบบมีเงื่อนไขภายหลังจากที่ทราบข้อมูลแล้วมีค่ามากกว่าความน่าจะเป็นโดยหลักเกณฑ์ และมีตัวพยากรณ์ที่เป็น parent ของตัวพยากรณ์นั้น ๆ ทุกตัวอยู่ในสมการแล้ว แต่ถ้าไม่เป็นไปตามเงื่อนไขแล้ว ตัวพยากรณ์นั้นจะต้องออกจากตัวแบบ

ผังงานแสดงวิธีการสร้างตัวแบบด้วยวิธีคัดเลือกตัวแปรของเบส์เซียน แสดงดังรูปที่ 6



รูปที่ 6 แสดงผังงานแสดงวิธีการสร้างตัวแบบด้วยวิธีคัดเลือกตัวแปรของเบส์เซียน

ขั้นตอนที่ 6

หลังจากได้ตัวแบบที่เหมาะสมในแต่ละวิธีแล้วจะคำนวณค่าคลาดเคลื่อนกำลังสองเฉลี่ยซึ่งได้จากโปรแกรมย่อย MSQE(N,K,Z,Y,B,AMSE)

ขั้นตอนที่ 7

เมื่อได้ค่าคลาดเคลื่อนกำลังสองเฉลี่ยของแต่ละวิธีแล้วเมื่อทำซ้ำตามจำนวนรอบที่กำหนด จะนำผลมาสรูปลงในตารางเพื่อแสดงการเปรียบเทียบหรือศึกษาแนวโน้มของแต่ละวิธี