บทที่ 3

วิธีดำเนินการวิจัย

การวิจัยครั้งนี้เป็นการศึกษาเปรียบเทียบวิธีการคัดเลือกสมการถดถอยที่ดีที่สุดภายใต้แนวทางของเบส์ ในการวิเคราะห์ความถดถอยเชิงเส้นพหุคูณ โดยใช้เกณฑ์ค่าเฉลี่ยของค่าคลาดเคลื่อนกำลังสองเฉลี่ยเป็นเกณฑ์ในการตัดสินใจ ซึ่งวิธีการคัดเลือกสมการถดถอยที่นำมาใช้ในการสร้างตัวแบบการถดถอยเชิงเส้นพหุคูณมี 5 วิธีดังนี้

1) วิธีการคัดเลือกตัวแบบที่ดีที่สุด โดยใช้เกณฑ์ข้อสนเทศของเบส์ (Bayes Information Criterion Method (BIC))
2) วิธีการคัดเลือกตัวแปรของเบส์ (Bayesian Variable Selection Method (BVS))
3) วิธีการเฉลี่ยตัวแบบของเบส์ (Bayesian Model Averaging Method) โดยใช้การค้นหาปริภูมิตัวแบบด้วยวิธีออกส์แคม วินโดว์ (Occam 's Window) (BMA_{OCC})
4) วิธีการเฉลี่ยตัวแบบของเบส์ (Bayesian Model Averaging Method) โดยการหาองค์ประกอบของตัวแบบด้วยเทคนิคมอนติคาร์โลโดยใช้ลูกโซ่มาร์คอฟ (Markov Chain Monte Carlo Model Composition(MC^3)) (BMA_{MC3})
5) วิธีการถดถอยแบบขั้นบันได (Stepwise Regression Method (SR))

และใช้เทคนิคมอนติคาร์โล (Monte Carlo Technique) ในการหาข้อสรุปของปัญหาที่ศึกษาตลอดจนใช้โปรแกรม Microsoft Visual Basic 6.0 และโปรแกรม S-plus 2000 บนเครื่องคอมพิวเตอร์ PC ในการประมวลผล ซึ่งขั้นตอนในการวิจัยมีรายละเอียดดังนี้

3.1 การจำลองข้อมูลด้วยวิธีมอนติคาร์โล(Simulation by Monte Carlo Method)

วิธีมอนติคาร์โลเป็นเทคนิคในการจำลองแบบทางคณิตศาสตร์ที่นิยมใช้กันอย่างแพร่หลาย โดยมีการจำลองตัวเลขสุ่ม(random number) มาช่วยในการหาคำตอบของปัญหาที่ต้องการศึกษาซึ่งยังไม่แน่ใจในผลลัพธ์ที่จะเกิดขึ้น เพราะเลขสุ่มมีประโยชน์หลายประการ คือ[1]

---

[1]พจนา แววสวัสดิ์, "การเปรียบเทียบเกณฑ์การคัดเลือกตัวแบบความถดถอยพหุนามแบบติดกลุ่ม," (วิทยานิพนธ์ปริญญามหาบัณฑิต ภาควิชาสถิติ บัณฑิตวิทยาลัย จุฬาลงกรณ์มหาวิยาลัย, 2543), หน้า 32.

1) ทำให้การเลือกตัวอย่างไม่มีความเอนเอียงในการสำรวจหรือทดลองในเรื่องนั้น ๆ ทั้งนี้เพราะเลขสุ่มมาจากแนวคิดเกี่ยวกับการคำนวณความน่าจะเป็น

2) เลขสุ่มจะทำให้ได้มาซึ่งรูปแบบต่าง ๆหรือวิธีการที่สลับซับซ้อนโดยการสร้างสถานการณ์การจำลอง (Simulation)

3) การใช้เลขสุ่มอาจทำเพื่อศึกษาคุณสมบัติทางทฤษฎีของกระบวนการทางสถิติที่มีความสำคัญสำหรับการประมาณค่า ตลอดจนนำไปสู่คำอธิบายเกี่ยวกับอำนาจการทดสอบทางสถิติ (power of statistical tests)

4) เพื่อหาคำตอบในปัญหาทางคณิตศาสตร์ โดยพิจารณาจากการแจกแจงความน่าจะเป็นของปัญหานั้น ๆ

ในการศึกษาวิจัยครั้งนี้จะใช้เทคนิคมอนติคาร์โลในการสร้างข้อมูลที่มีลักษณะการแจกแจงตามที่ต้องการศึกษา โดยขั้นตอนที่สำคัญของการจำลองข้อมูลด้วยวิธีมอนติคาร์โลมี 3 ขั้นตอนดังนี้

ขั้นตอนที่ 1 การสร้างเลขสุ่ม (generate random number)

การสร้างเลขสุ่มเป็นขั้นตอนที่สำคัญในวิธีมอนติคาร์โล ทั้งนี้เนื่องจากหลักการของวิธีมอนติคาร์โลนั้นจะใช้ตัวเลขสุ่มมาช่วยในการหาคำตอบของปัญหา ลักษณะของตัวเลขสุ่มที่ดีจะมีการแจกแจงแบบสม่ำเสมอ(Uniform distribution) ในช่วง [0,1] และเป็นอิสระซึ่งกันและกัน จากนั้นนำเลขสุ่มที่ได้ไปสร้างตัวแปรสุ่มตามลักษณะการแจกแจงที่ต้องการศึกษา เพื่อเป็นข้อมูลสำหรับปัญหานั้น ๆ

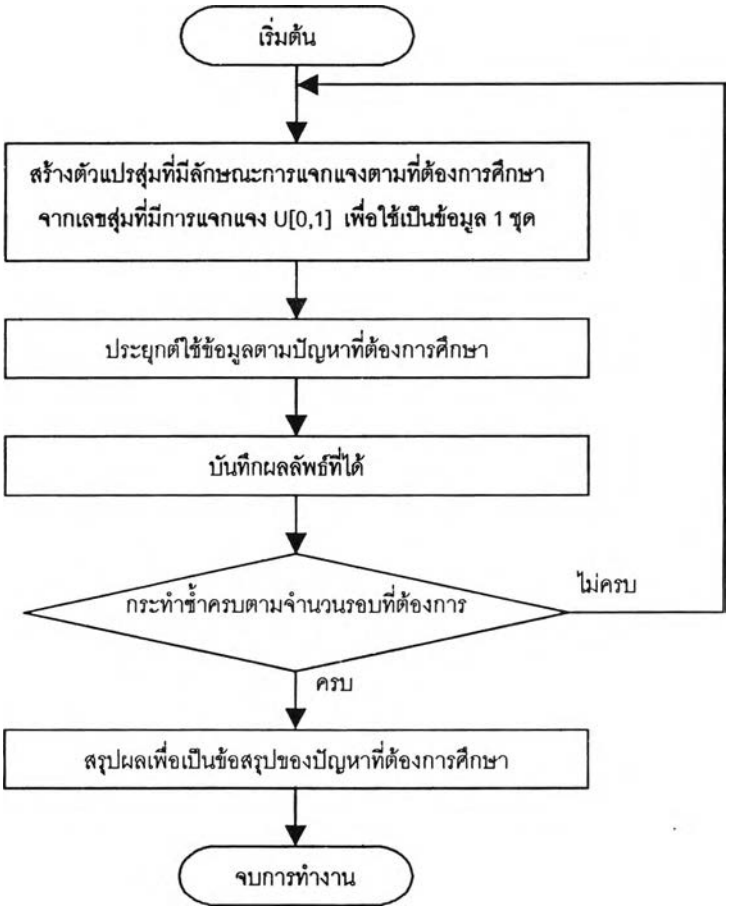
ขั้นตอนที่ 2 การประยุกต์ปัญหาที่ต้องการศึกษาโดยใช้เลขสุ่ม

ขั้นตอนนี้ขึ้นอยู่กับปัญหาที่ต้องการศึกษา ซึ่งเป็นขั้นตอนที่ใช้เลขสุ่มในการหาค่าตามสูตรหรือการคำนวณในปัญหาที่ศึกษา บางปัญหาอาจใช้ตัวเลขสุ่มโดยตรง แต่บางปัญหาอาจใช้ตัวเลขสุ่มเพียงบางขั้นตอนของปัญหาเท่านั้น

ขั้นตอนที่ 3 การทดลองกระทำ

เมื่อประยุกต์ปัญหาที่ต้องการศึกษาโดยใช้เลขสุ่มแล้ว ขั้นตอนต่อไปคือการทดลองโดยใช้กระบวนการสุ่ม(random process) มาทดลองกระทำซ้ำ ๆ กัน (replication) จำนวนหลายครั้ง เพื่อหาคำตอบของปัญหาที่ต้องการศึกษา ซึ่งการทดลองกระทำซ้ำ ๆ กันนั้นจะเป็นการช่วยลดความไม่แน่นอนของคำตอบได้

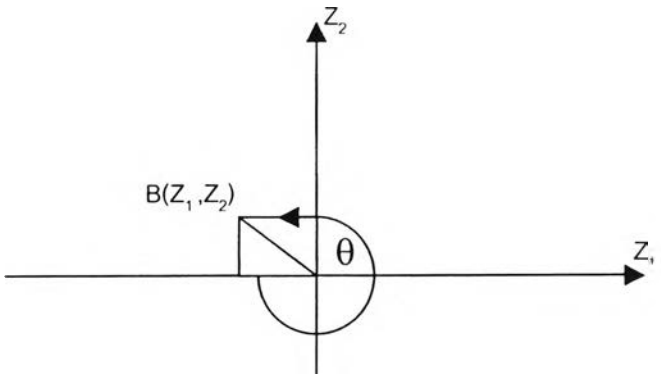
จากหลักการของวิธีมอนติคาร์โลจะเห็นได้ว่า การใช้เลขสุ่มเพื่อเป็นพื้นฐานในการหาคำตอบของปัญหา เป็นวิธีการที่จะนำไปสู่แนวคิดทางทฤษฎีที่เกี่ยวข้องกับการคำนวณ โดยเฉพาะทฤษฎีความน่าจะเป็น ที่จะนำไปสู่การอ้างอิงผลสรุปในสถานการณ์ของข้อมูลจริง เพราะไม่มีผลกระทบจากปัจจัยอื่น ๆ เข้ามาเกี่ยวข้อง ในการทดลองเมื่อกระทำซ้ำ ๆ กันเป็นจำนวนมากแล้ว ความคลาดเคลื่อนอย่างสุ่มที่เกิดขึ้นในการวิเคราะห์หาค่าต่าง ๆ ในแต่ละครั้งจะหมดไป(counter balance) จากขั้นตอนของวิธีมอนติคาร์โล สามารถเขียนผังงานได้ดังนี้



รูปที่ 3.1 แสดงผังงานสำหรับขั้นตอนของวิธีมอนติคาร์โล

3.2 การสร้างเลขสุ่มที่มีการแจกแจงแบบปกติ

การสร้างเลขสุ่มที่มีการแจกแจงแบบปกติในการวิจัยครั้งนี้ใช้วิธีการของบ็อกซ์ และมูลเลอร์ (Box and Muller ,1958) โดยสร้างเลขสุ่มที่มีการแจกแจงแบบปกติมาตรฐานที่มีค่าเฉลี่ยเท่ากับ 0 และมีค่าความแปรปรวนเท่ากับ 1 พร้อม ๆ กัน 2 ค่า และแต่ละค่าเป็นอิสระซึ่งกันและกัน โดยใช้ตัวก่อกำเนิด (generator) Z_1 และ Z_2 ซึ่งพิจารณาดังรูปต่อไปนี้



พิจารณาจากรูปจะได้

(3.1) $Z_1 = B\cos(\theta)$

(3.2) $Z_2 = B\sin(\theta)$

เนื่องจาก $B^2 = Z_1^2 + Z_2^2$ มีการแจกแจงไคกำลังสอง (Chi-square distribution) ด้วยระดับขั้นความเสรีเท่ากับ 2 ซึ่งเทียบเท่ากับการแจกแจงแบบเลขชี้กำลัง (Exponential distribution) ที่ค่าเฉลี่ยเท่ากับ 2 โดยใช้วิธีการแปลงผกผัน (Inverse Tranform Method) จะสามารถสร้างเลขสุ่มที่มีการแจกแจงแบบเลขชี้กำลังได้ดังนี้

(3.3) $$B = (-2\ln r)^{\frac{1}{2}}$$

เมื่อ r เป็นเลขสุ่มที่มีการแจกแจงแบบสม่ำเสมอ (Uniform distribution) ในช่วง [0,1]

โดยการสมมาตรของการแจกแจงแบบปกติ จะได้ว่ามุม θ จะมีการแจกแจงแบบสม่ำเสมอระหว่าง 0 ถึง 2π เรเดียน และรัศมี B ทำมุมกับ θ ซึ่ง B และ θ เป็นอิสระซึ่งกันและกัน

จากสมการ (3.1) (3.2) และ (3.3) เราสามารถสร้างเลขสุ่มที่การแจกแจงแบบปกติมาตรฐานจากเลขสุ่ม 2 ชุด r_1 และ r_2 กล่าวคือ

$$Z_1 = \left(-2\ln r_1\right)^{\frac{1}{2}}\cos\left(2\pi r_2\right)$$

$$Z_2 = \left(-2\ln r_1\right)^{\frac{1}{2}}\sin\left(2\pi r_2\right)$$

ซึ่ง r_1 และ r_2 เป็นเลขสุ่มที่มีการแจกแจงแบบสม่ำเสมอ โดยรายละเอียดและโปรแกรมสำหรับการสร้างเลขสุ่มที่มีการแจกแจงแบบสม่ำเสมอสามารถดูได้จากภาคผนวก

เมื่อได้เลขสุ่มที่มีการแจกแจงแบบปกติมาตรฐานแล้ว สามารถทำการแปลงเลขสุ่มดังกล่าวโดยใช้ฟังก์ชัน

$$m_1 = \mu + \sigma Z_1$$
$$m_2 = \mu + \sigma Z_2$$

จะได้ว่า m_1 และ m_2 มีการแจกแจงแบบปกติด้วยค่าเฉลี่ยเท่ากับ μ และความแปรปรวนเท่ากับ σ^2

3.3 แผนการทดลอง

ผู้วิจัยได้กำหนดสถานการณ์ต่าง ๆ สำหรับการวิจัยครั้งนี้ไว้ดังนี้

1) เลือกตัวอย่างสุ่มเพื่อใช้เป็นค่าคลาดเคลื่อนสุ่มจากประชากรที่มีการแจกแจงเดียวกัน โดยการวิจัยครั้งนี้สนใจศึกษาเฉพาะตัวอย่างสุ่มที่มาจากการแจกแจงแบบปกติที่มีค่าเฉลี่ยเท่ากับ 0 และส่วนเบี่ยงเบนมาตรฐานเท่ากับ 5 10 20 และ 25

2) กำหนดขนาดตัวอย่างที่ศึกษา คือ 25 50 75 และ 100

3) กำหนดจำนวนตัวแปรอิสระที่ศึกษา คือ 3 5 10 12 และ 15 ตัวแปร

4) ระดับนัยสำคัญ คือ 0.01 และ 0.05

5) ค่าคงที่สำหรับวิธี BVS และวิธี BMA_{MC3} มี 4 ระดับ คือ (1,5) (1,10) (10,100) และ (10,500)

6) ค่าคงที่ที่ใช้เป็นเกณฑ์ในการค้นหาปริภูมิตัวแบบสำหรับวิธี BMA_{OCC} เป็น 20 ในกรณีที่ระดับนัยสำคัญเท่ากับ 0.05 และเป็น 100 ในกรณีที่ระดับนัยสำคัญเท่ากับ 0.01

3.4 ขั้นตอนในการดำเนินการวิจัย

ขั้นตอนในการดำเนินการวิจัยมีดังนี้

1) กำหนดลักษณะการแจกแจงของค่าคลาดเคลื่อน ขนาดตัวอย่าง จำนวนตัวแปรอิสระ ระดับนัยสำคัญ ค่าคงที่สำหรับวิธี BVS และวิธี BMA_{MC3} และค่าคงที่ที่ใช้เป็นเกณฑ์ในการค้นหาปริภูมิตัวแบบสำหรับวิธี BMA_{OCC}

2) สร้างข้อมูลตัวแปรอิสระ และค่าคลาดเคลื่อนที่มีลักษณะการแจกแจงตามที่ต้องการศึกษา จากนั้นสร้างข้อมูลตัวแปรตามจากตัวแปรอิสระและค่าคลาดเคลื่อน โดยให้ตัวแปรตามมีความสัมพันธ์เชิงเส้นในพารามิเตอร์กับตัวแปรอิสระ

3) กำหนดตัวแบบเริ่มต้นในการคัดเลือกตัวแบบที่เหมาะสม

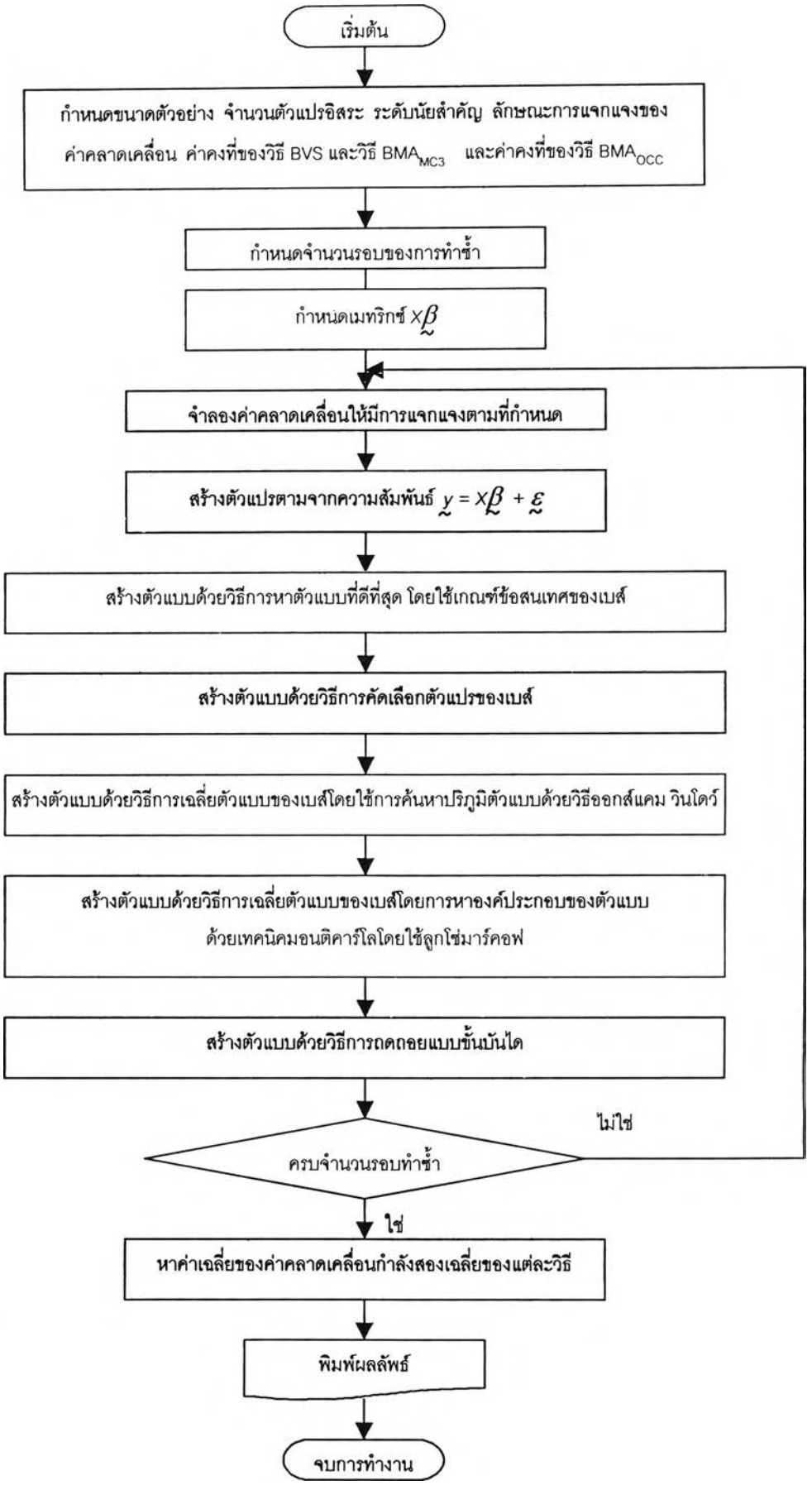
4) ทำการสร้างตัวแบบโดยวิธีการทั้ง 5 วิธี คือ

- วิธีการคัดเลือกตัวแบบที่ดีที่สุด โดยใช้เกณฑ์ข้อสนเทศของเบส์ (Bayes Information Criterion Method (BIC))
- วิธีการคัดเลือกตัวแปรของเบส์ (Bayesian Variable Selection Method (BVS))
- วิธีการเฉลี่ยตัวแบบของเบส์ (Bayesian Model Averaging Method) โดยใช้การค้นหาปริภูมิตัวแบบด้วยวิธีออกแคมส์ วินโดว์ (Occam 's Window) (BMA_{OCC})
- วิธีการเฉลี่ยตัวแบบของเบส์ (Bayesian Model Averaging Method) โดยการหาองค์ประกอบของตัวแบบด้วยเทคนิคมอนติคาร์โลโดยใช้ลูกโซ่มาร์คอฟ (Markov Chain Monte Carlo Model Composition (MC^3)) (BMA_{MC3})
- วิธีการถดถอยแบบขั้นบันได (Stepwise Regression Method (SR))

5) คำนวณค่าคลาดเคลื่อนกำลังสองเฉลี่ยจากการทำซ้ำจำนวน 500 รอบ ของตัวแบบที่ได้จากวิธีการคัดเลือกตัวแบบการถดถอยทั้ง 5 วิธี

6) ทำการเปรียบเทียบค่าคลาดเคลื่อนกำลังสองเฉลี่ยของตัวแบบที่ได้จากวิธีการคัดเลือกตัวแบบทั้ง 5 วิธี โดยสรุปผลในรูปของตารางและรูปภาพ

ผังงานแสดงขั้นตอนในการดำเนินการวิจัย ดังแสดงในรูปที่ 3.2



รูปที่ 3.2 แสดงผังงานสำหรับขั้นตอนในการดำเนินการวิจัย

สำหรับรายละเอียดของแต่ละขั้นตอนมีดังนี้

ขั้นตอนที่ 1

กำหนดลักษณะการแจกแจงของค่าคลาดเคลื่อน ขนาดตัวอย่าง จำนวนตัวแปรอิสระ ระดับนัยสำคัญ ค่าคงที่สำหรับวิธี BVS และวิธี BMA_{MC3} และค่าคงที่ที่ใช้เป็นเกณฑ์ในการค้นหาปริภูมิตัวแบบสำหรับ BMA_{OCC} โดยจะกำหนดตามแผนการทดลองที่กล่าวมาแล้วข้างต้น

ขั้นตอนที่ 2

การสร้างข้อมูลในการวิจัยครั้งนี้ประกอบด้วย ข้อมูลตัวแปรอิสระ ข้อมูลค่าคลาดเคลื่อนที่มีลักษณะการแจกแจงตามที่ต้องการศึกษา และข้อมูลตัวแปรตาม โดยมีรายละเอียดดังนี้

1) <u>ข้อมูลตัวแปรอิสระ</u>

เนื่องจากการวิจัยครั้งนี้ศึกษาเฉพาะการวิเคราะห์ความถดถอยเชิงเส้นพหุคูณในกรณีที่ตัวแปรอิสระเป็นไปตามข้อตกลงเบื้องต้น กล่าวคือ ตัวแปรอิสระเป็นค่าคงที่ ไม่มีพหุสัมพันธ์กัน กำลังสูงสุดของตัวแปรอิสระเป็น 1 และไม่มีอันตรกิริยาระหว่างตัวแปรอิสระ นั่นคือ การวิจัยครั้งนี้ศึกษาเฉพาะการวิเคราะห์ความถดถอยเชิงเส้นพหุคูณในรูปทั่วไป ไม่รวมถึงตัวแบบความถดถอยพหุนาม ดังนั้นในการสร้างข้อมูลตัวแปรอิสระจะสร้างจากการแจกแจงพหุแบบปกติ (multivariate normal distribution) ที่มีเวกเตอร์ค่าเฉลี่ยเท่ากับ 0 และเมทริกซ์ความแปรปรวนร่วมเป็นเมทริกซ์เอกลักษณ์ (identity matrix) ซึ่งจะทำให้ได้ตัวแปรอิสระแต่ละตัวแปรที่มีค่าเฉลี่ยเป็น 0 ความแปรปรวนเป็น 1 และตัวแปรอิสระทุกตัวที่ได้จะไม่มีพหุสัมพันธ์กัน โดยสร้างจากฟังก์ชัน *rmvnorm(n , mean , cov , sd , rho , d)* ในโปรแกรม S-plus 2000 ซึ่งรายละเอียดสามารถดูได้จากภาคผนวก จากนั้นนำข้อมูลตัวแปรอิสระที่ได้เก็บไว้ในแฟ้มข้อมูลสำหรับกรณีจำนวนตัวแปรอิสระในระดับต่าง ๆ เพื่อเรียกใช้ในการประมวลผลต่อไป

2) ข้อมูลค่าคลาดเคลื่อน

การวิจัยครั้งนี้สนใจศึกษาเฉพาะกรณีที่ค่าคลาดเคลื่อนสุ่มมีการแจกแจงแบบปกติที่มีค่าเฉลี่ยเท่ากับ 0 และความแปรปรวนเท่ากับ σ^2 เหมือนกันและเป็นอิสระซึ่งกันและกัน โดยเริ่มจากสร้างเลขสุ่มที่มีการแจกแจงแบบสม่ำเสมอจากโปรแกรมย่อย *Uniform(RandomNumber() , n)* จากนั้นทำการแปลงเลขสุ่มที่ได้โดยใช้วิธีที่กล่าวไว้ในหัวข้อ 3.2 ให้มีการแจกแจงแบบปกติที่มีค่าเฉลี่ย และความแปรปรวนตามที่ต้องการจากโปรแกรมย่อย *Normal(mean ,sigma , X , RandomNumber() ,index , KK , Z2)* ซึ่งโปรแกรมย่อยทั้งสองใช้โปรแกรมภาษาวิชวลเบสิก 6.0 (Microsoft Visual Basic 6.0) บน PC สำหรับรายละเอียดของโปรแกรมสามารถดูได้จากภาคผนวก

3) ข้อมูลตัวแปรตาม

เมื่อมีข้อมูลตัวแปรอิสระ และข้อมูลค่าคลาดเคลื่อนที่มีลักษณะการแจกแจงตามที่ต้องการศึกษาแล้ว จากนั้นจะทำการสร้างข้อมูลตัวแปรตามให้มีความสัมพันธ์เชิงเส้นในพารามิเตอร์กับตัวแปรอิสระ ซึ่งมีรูปแบบความสัมพันธ์ดังนี้

$$\underset{\sim}{y} = X\underset{\sim}{\beta} + \underset{\sim}{\varepsilon}$$

เมื่อ $\underset{\sim}{y}$ เป็นเวกเตอร์ของตัวแปรตาม

X เป็นเมทริกซ์ของตัวแปรอิสระ

$\underset{\sim}{\beta}$ เป็นเวกเตอร์ของพารามิเตอร์ที่กำหนดขึ้น ซึ่งในการวิจัยครั้งนี้กำหนดให้ $\underset{\sim}{\beta}' = (1,1,...,1)_{1\times p}$ นั่นคือกำหนดให้ตัวแปรอิสระทุกตัวมีอิทธิพลเท่ากันในตัวแบบเริ่มต้น

และ $\underset{\sim}{\varepsilon}$ เป็นเวกเตอร์ของค่าคลาดเคลื่อน ซึ่ง $\underset{\sim}{\varepsilon} \sim N(\underset{\sim}{0}, \sigma^2 I_n)$

ขั้นตอนที่ 3

กำหนดตัวแบบเริ่มต้นเป็นตัวแบบเต็มรูป (full model) ในการคัดเลือกตัวแบบที่เหมาะสม

ขั้นตอนที่ 4

หลังจากที่มีข้อมูลครบแล้วจะทำการสร้างตัวแบบที่เหมาะสมจากทั้ง 5 วิธี ดังต่อไปนี้

1) วิธีการคัดเลือกตัวแบบที่ดีที่สุดโดยใช้เกณฑ์ข้อสนเทศของเบส์ (BIC) ใช้โปรแกรมย่อย *BIC(Y() , X() , n , NumX , MSE)*

2) วิธีการคัดเลือกตัวแปรของเบส์ (BVS) ใช้โปรแกรมย่อย *BVS(Y() , X() , n , NumX , IP , CC , MSE)*

3) วิธีการเฉลี่ยตัวแบบของเบส์โดยใช้การค้นหาปริภูมิตัวแบบด้วยวิธีออกส์แคม วินโดว์ (BMA_{OCC}) ใช้โปรแกรมย่อย *BMA.OCC(X , Y , wt , strict , OR , MSE)*

4) วิธีการเฉลี่ยตัวแบบของเบส์โดยการหาองค์ประกอบของตัวแบบด้วยเทคนิคมอนติคาร์โลโดยใช้ลูกโซ่มาร์คอฟ (BMA_{MC3}) ใช้โปรแกรมย่อย *BMA.MC3(Y , X , num.its , M0.var , M0.out, outs.list , PI , K , IP , CC , MSE)*

5) วิธีการถดถอยแบบขั้นบันได (SR) ใช้โปรแกรมย่อย *Stepwise(Y() , X() , n , NumX , Alpha , MSE)*

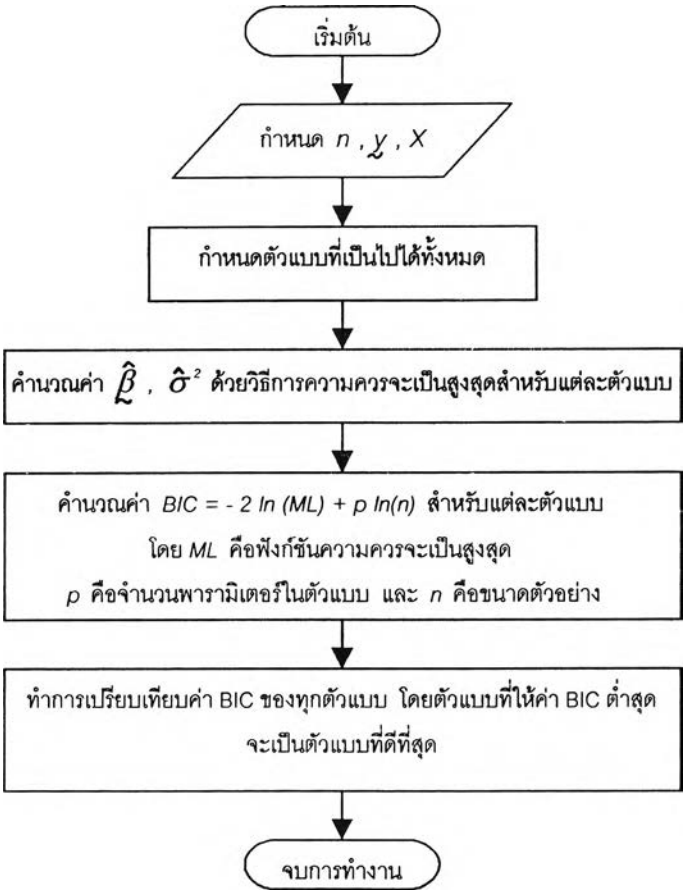
ซึ่งโปรแกรมย่อย BIC(...) BVS(...) แล Stepwise(...) เป็นโปรแกรมย่อยที่สร้างขึ้นโดยใช้ภาษาวิชวลเบสิก (Microsoft Visual Basic 6.0) ส่วนโปรแกรมย่อย BMA.OCC(...) และ BMA.MC3(...) เป็นโปรแกรมย่อยที่สร้างโดยใช้โปรแกรม S-plus 2000 สำหรับรายละเอียดของโปรแกรมสามารถดูได้จากภาคผนวก

วิธีการสร้างตัวแบบในแต่ละวิธี มีรายละเอียดดังนี้

1) **วิธีการคัดเลือกตัวแบบที่ดีที่สุดโดยใช้เกณฑ์ข้อสนเทศของเบส์** (Bayes Information Criterion Method (BIC))

วิธีการคัดเลือกตัวแบบที่ดีที่สุดโดยใช้เกณฑ์ข้อสนเทศของเบส์จะเริ่มจากการสร้างตัวแบบที่เป็นไปได้ทั้งหมดจากตัวแบบเริ่มต้นที่กำหนดไว้ เช่นในตัวแบบเริ่มต้นมีตัวแปรอิสระจำนวน k ตัวแปร จะได้ว่าตัวแบบที่เป็นไปได้ทั้งหมดมีจำนวน 2^k ตัวแบบ จากนั้นคำนวณค่า BIC สำหรับแต่ละตัวแบบ โดยตัวแบบที่ให้ค่า BIC ต่ำสุดจะเป็นตัวแบบที่ให้ค่าพยากรณ์ถูกต้องและแม่นยำมากที่สุด

ผังงานแสดงขั้นตอนของวิธีการคัดเลือกตัวแบบที่ดีที่สุดโดยใช้เกณฑ์ข้อสนเทศของเบส์ ดังแสดงในรูปที่ 3.3



รูปที่ 3.3 แสดงผังงานสำหรับขั้นตอนของวิธีการคัดเลือกตัวแบบที่ดีที่สุดโดยใช้เกณฑ์ข้อสนเทศของเบส์

2) **วิธีการคัดเลือกตัวแปรของเบส์** (Bayesian Variable Selection Method (BVS))

ขั้นตอนของวิธีการคัดเลือกตัวแปรของเบส์มีดังนี้

1) กำหนดความน่าจะเป็นก่อน และค่าคงที่ของวิธีการคัดเลือกตัวแปรของเบส์

2) กำหนดตัวแบบเริ่มต้นเป็นตัวแบบเต็มรูป คือ มีตัวแปรอิสระครบทุกตัวแปรในสมการถดถอย

3) คำนวณค่าประมาณของสัมประสิทธิ์การถดถอยและค่าคลาดเคลื่อนกำลังสองเฉลี่ยของตัวแบบเต็มรูปโดยวิธีกำลังน้อยสุด เพื่อเป็นค่าเริ่มต้นที่ใช้ในการสุ่มพารามิเตอร์ $\underset{\sim}{\beta}$ และ σ^2

4) กำหนดจำนวนรอบในการค้นหาพารามิเตอร์ ซึ่งในการวิจัยครั้งนี้กำหนดจำนวนรอบเท่ากับ 5,000 รอบ ในกรณีที่จำนวนตัวแปรอิสระเป็น 3 5 และ 10 ตัวแปร และจำนวนรอบเท่ากับ 10,000 รอบ ในกรณีที่จำนวนตัวแปรอิสระเป็น 12 และ 15 ตัวแปร

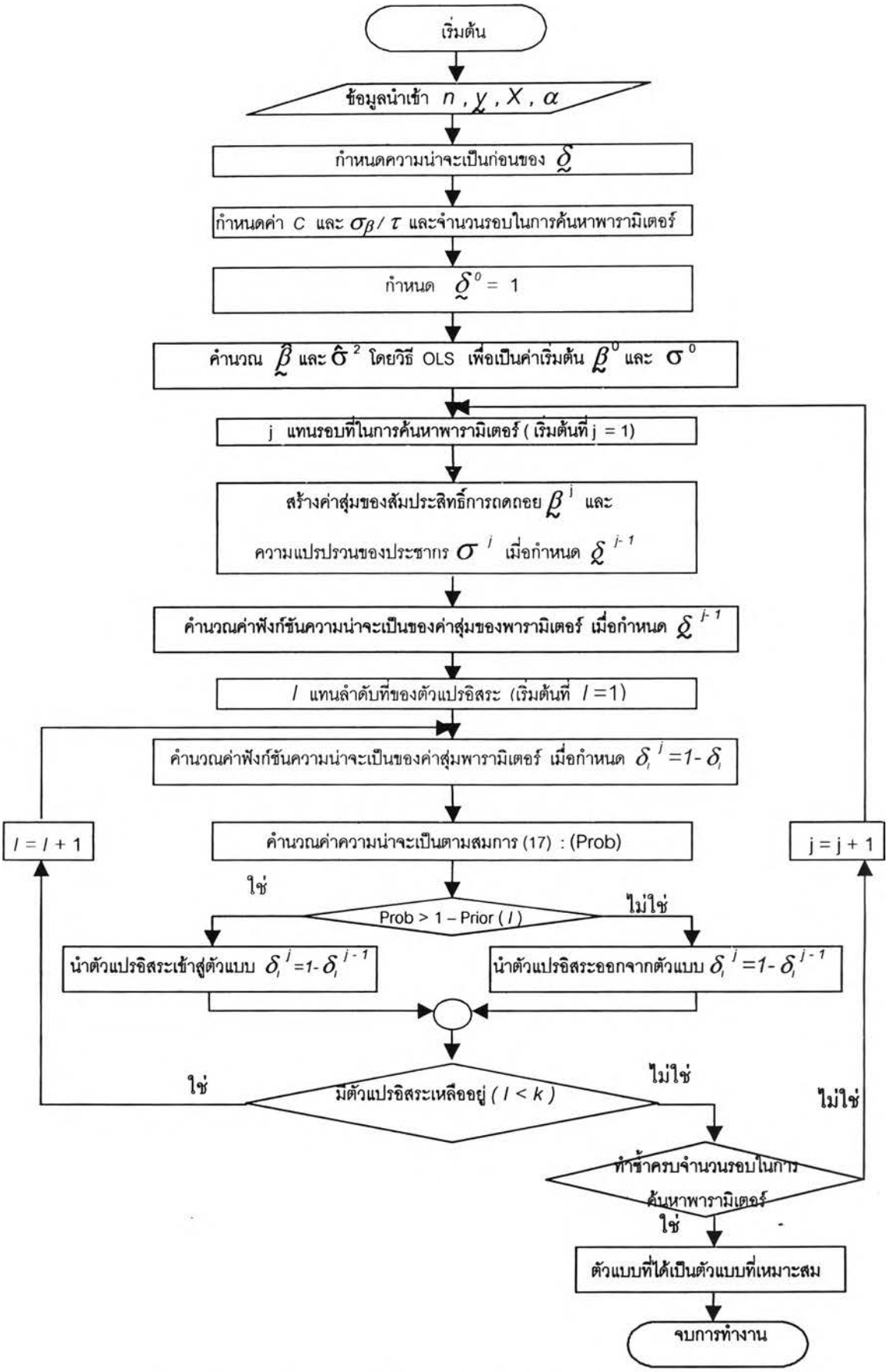
5) ทำการสุ่มค่าพารามิเตอร์ $\underset{\sim}{\beta}$ และ σ^2 จากการแจกแจงพหุแบบปกติ (multivariate normal distribution) และการแจกแจงแบบแกมมาผกผัน (inverse gamma distribution) ตามลำดับ

6) คำนวณหาฟังก์ชันความน่าจะเป็นของค่าที่สุ่มได้ เพื่อนำไปหาค่าความน่าจะเป็นแบบมีเงื่อนไขภายหลังจากทราบข้อมูลแล้ว

7) พิจารณาตัวแปรที่อยู่ในสมการถดถอย โดยตัวแปรที่จะอยู่ในสมการถดถอยจะต้องมีค่าความน่าจะเป็นแบบมีเงื่อนไขภายหลังจากทราบข้อมูลแล้ว มากกว่าค่าความน่าจะเป็นก่อน

8) เมื่อทำซ้ำครบจำนวนรอบในการค้นหาพารามิเตอร์แล้ว จะเลือกตัวแบบที่มีความถี่ในการถูกเลือกสูงสุดเป็นตัวแบบที่ดีที่สุดในการพยากรณ์

ผังงานแสดงขั้นตอนของวิธีการคัดเลือกตัวแปรของเบส์ ดังแสดงในรูปที่ 3.4

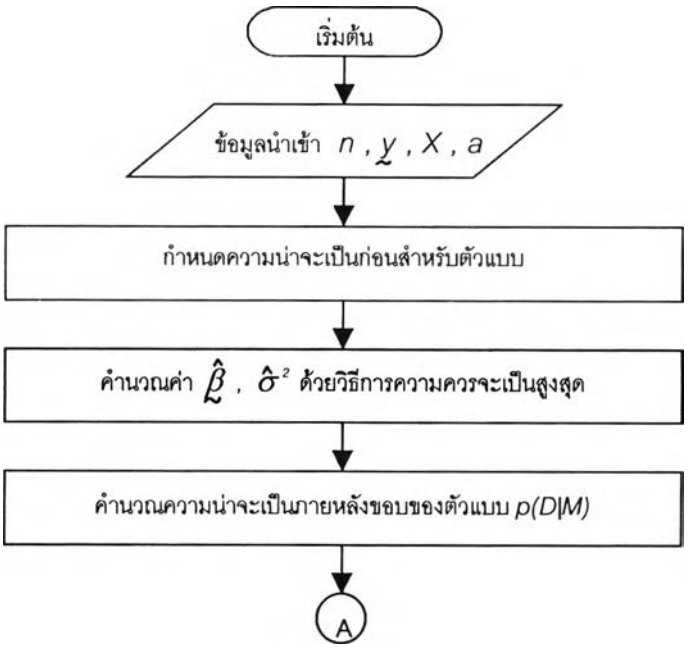


รูปที่ 3.4 แสดงผังงานสำหรับขั้นตอนของวิธีการคัดเลือกตัวแปรของเบส์

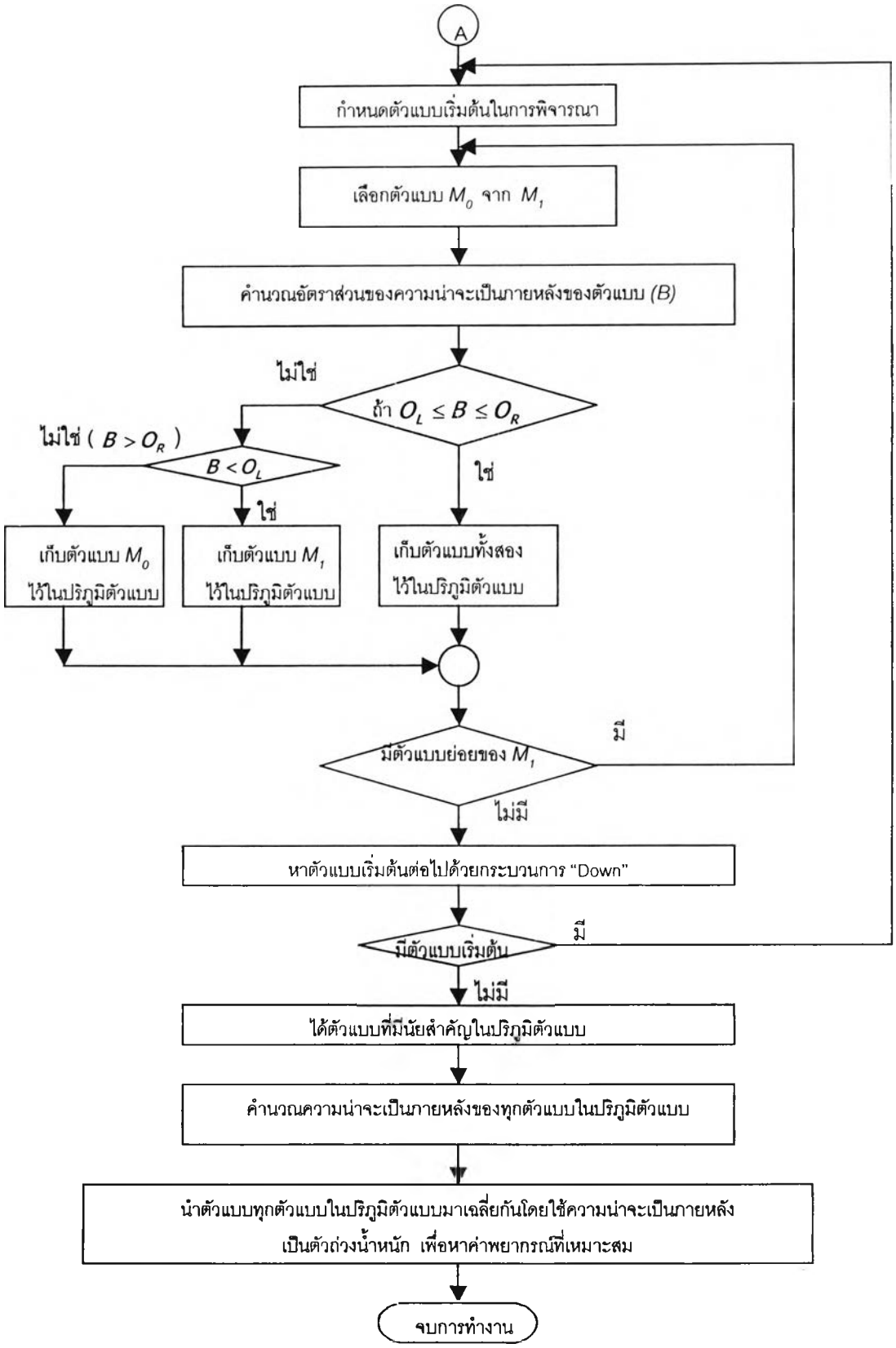
3) **วิธีการเฉลี่ยตัวแบบของเบส**(Bayesian Model Averaging Method) **โดยใช้การค้นหาปริภูมิตัวแบบด้วยวิธีออกแคมส์ วินโดว์** (Occam 's Window) (BMA_{occ})

วิธีการนี้เริ่มจากกำหนดความน่าจะเป็นก่อนสำหรับตัวแบบ คำนวณค่าประมาณพารามิเตอร์ของตัวแบบด้วยวิธีการความควรจะเป็นสูงสุด คำนวณความน่าจะเป็นภายหลังขอบของตัวแบบ และกำหนดตัวแบบเริ่มต้นในการพิจารณาเป็นตัวแบบเต็มรูป จากนั้นทำการค้นหาปริภูมิตัวแบบด้วยวิธีออกส์แคม วินโดว์ (Occam 's Window) ด้วยกระบวนการ "Up" และกระบวนการ "Down" ดังที่กล่าวไว้ในบทที่ 2 เมื่อได้ตัวแบบที่มีนัยสำคัญในปริภูมิตัวแบบแล้ว ทำการคำนวณความน่าจะเป็นภายหลังของทุกตัวแบบในปริภูมิตัวแบบ นำทุกตัวแบบในปริภูมิตัวแบบมาเฉลี่ยกันโดยใช้ความน่าจะเป็นภายหลังของแต่ละตัวแบบเป็นตัวถ่วงน้ำหนัก เพื่อหาค่าพยากรณ์ที่เหมาะสม

ผังงานแสดงขั้นตอนของวิธีการเฉลี่ยตัวแบบของเบส์โดยใช้การค้นหาปริภูมิตัวแบบด้วยวิธีออกแคมส์ วินโดว์ ดังแสดงในรูปที่ 3.5



รูปที่ 3.5 แสดงผังงานสำหรับขั้นตอนของวิธีการเฉลี่ยตัวแบบของเบส์โดยใช้การค้นหาปริภูมิตัวแบบด้วยวิธีออกแคมส์ วินโดว์

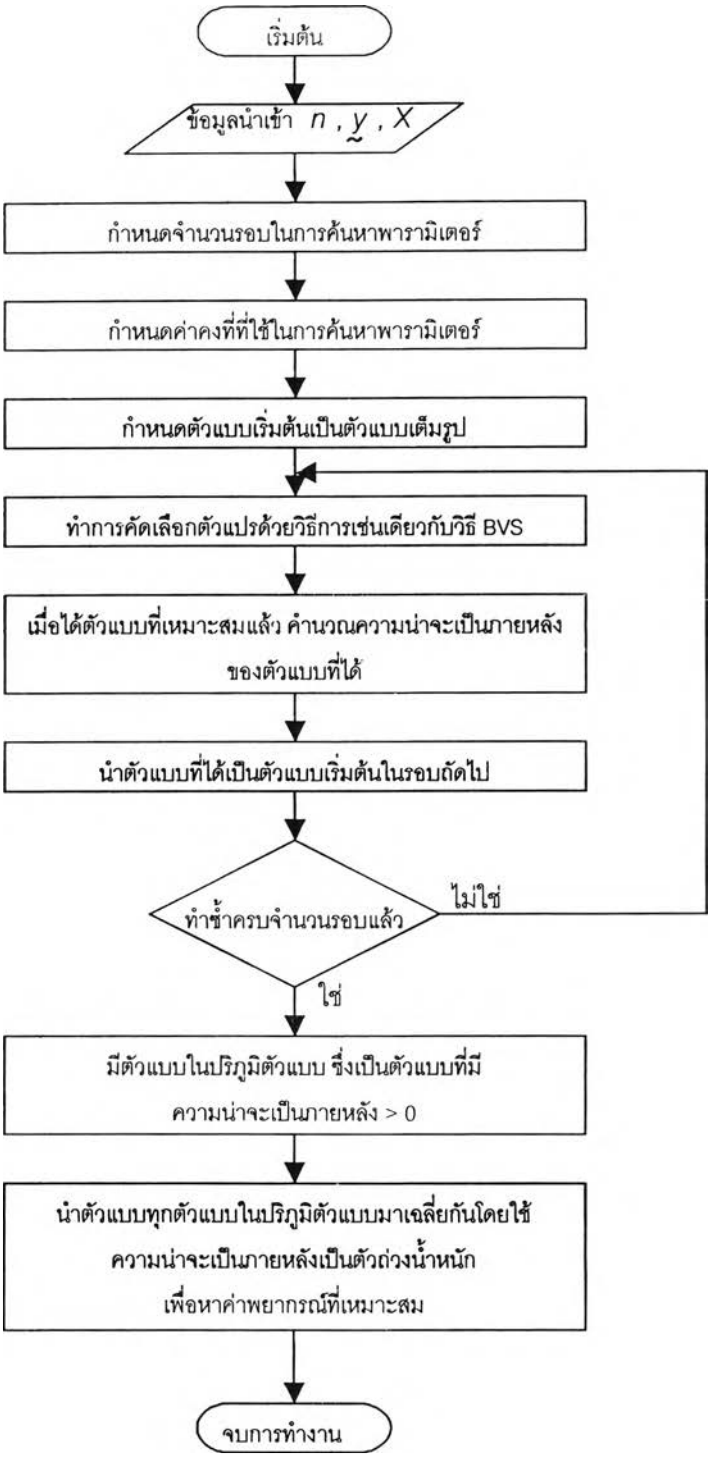


รูปที่ 3.5 (ต่อ) แสดงผังงานสำหรับขั้นตอนของวิธีการเฉลี่ยตัวแบบของเบส์โดยใช้การค้นหาปริภูมิตัวแบบด้วยวิธีออกแคมส์วินโดว์

4) **วิธีการเฉลี่ยตัวแบบของเบส** (Bayesian Model Averaging Method) **โดยการหาองค์ประกอบของตัวแบบด้วยเทคนิคมอนติคาร์โลโดยใช้ลูกโซ่มาร์คอฟ** (Markov Chain Monte Carlo Model Composition (MC^3)) (BMA_{MC3})

วิธีการนี้เริ่มต้นจากกำหนดตัวแบบเริ่มต้นเป็นตัวแบบเต็มรูป กำหนดจำนวนรอบและค่าคงที่ในการค้นหาพารามิเตอร์ จากนั้นทำการวนซ้ำเพื่อคำนวณความน่าจะเป็นภายหลังของแต่ละตัวแบบ และทำการคัดเลือกตัวแปรด้วยวิธีการเช่นเดียวกับวิธีการคัดเลือกตัวแปรของเบส์ เมื่อทำซ้ำครบจำนวนรอบที่กำหนดไว้แล้วจะได้ตัวแบบที่มีค่าความน่าจะเป็นภายหลังมากว่าศูนย์อยู่ในปริภูมิตัวแบบ นำตัวแบบทุกตัวแบบในปริภูมิตัวแบบมาเฉลี่ยกันโดยใช้ความน่าจะเป็นภายหลังของแต่ละตัวแบบเป็นตัวถ่วงน้ำหนัก เพื่อหาค่าพยากรณ์ที่เหมาะสม

ผังงานแสดงขั้นตอนของวิธีการเฉลี่ยตัวแบบของเบส์โดยการหาองค์ประกอบของตัวแบบด้วยเทคนิคมอนติคาร์โลโดยใช้ลูกโซ่มาร์คอฟ ดังแสดงในรูปที่ 3.6



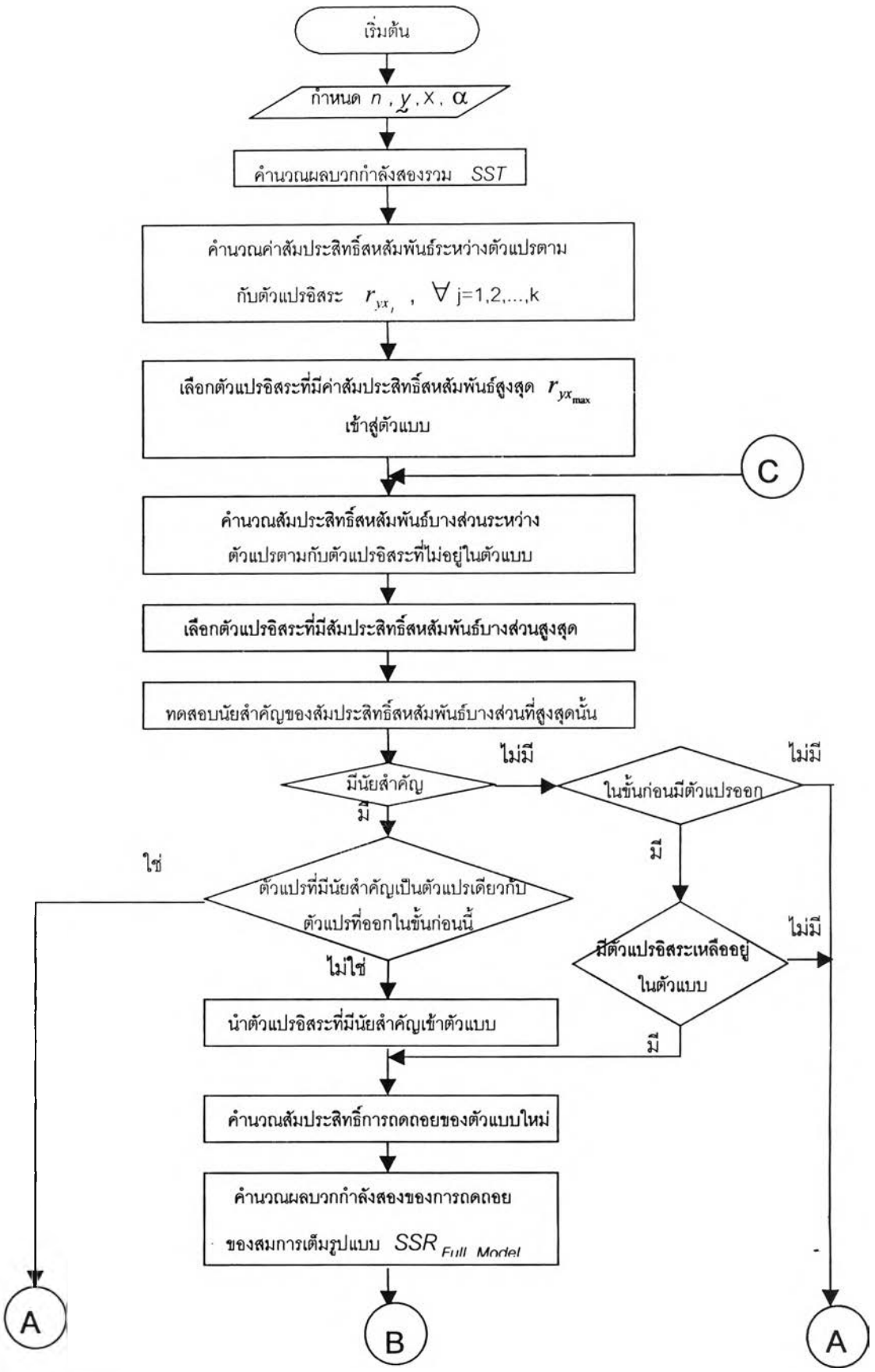
รูปที่ 3.6 แสดงผังงานสำหรับขั้นตอนของวิธีการเฉลี่ยตัวแบบของเบส์โดยการหาองค์ประกอบของตัวแบบด้วยเทคนิคมอนติคาร์โลโดยใช้ลูกโซ่มาร์คอฟ

5) **วิธีการถดถอยแบบขั้นบันได** (Stepwise Regression Method (SR))

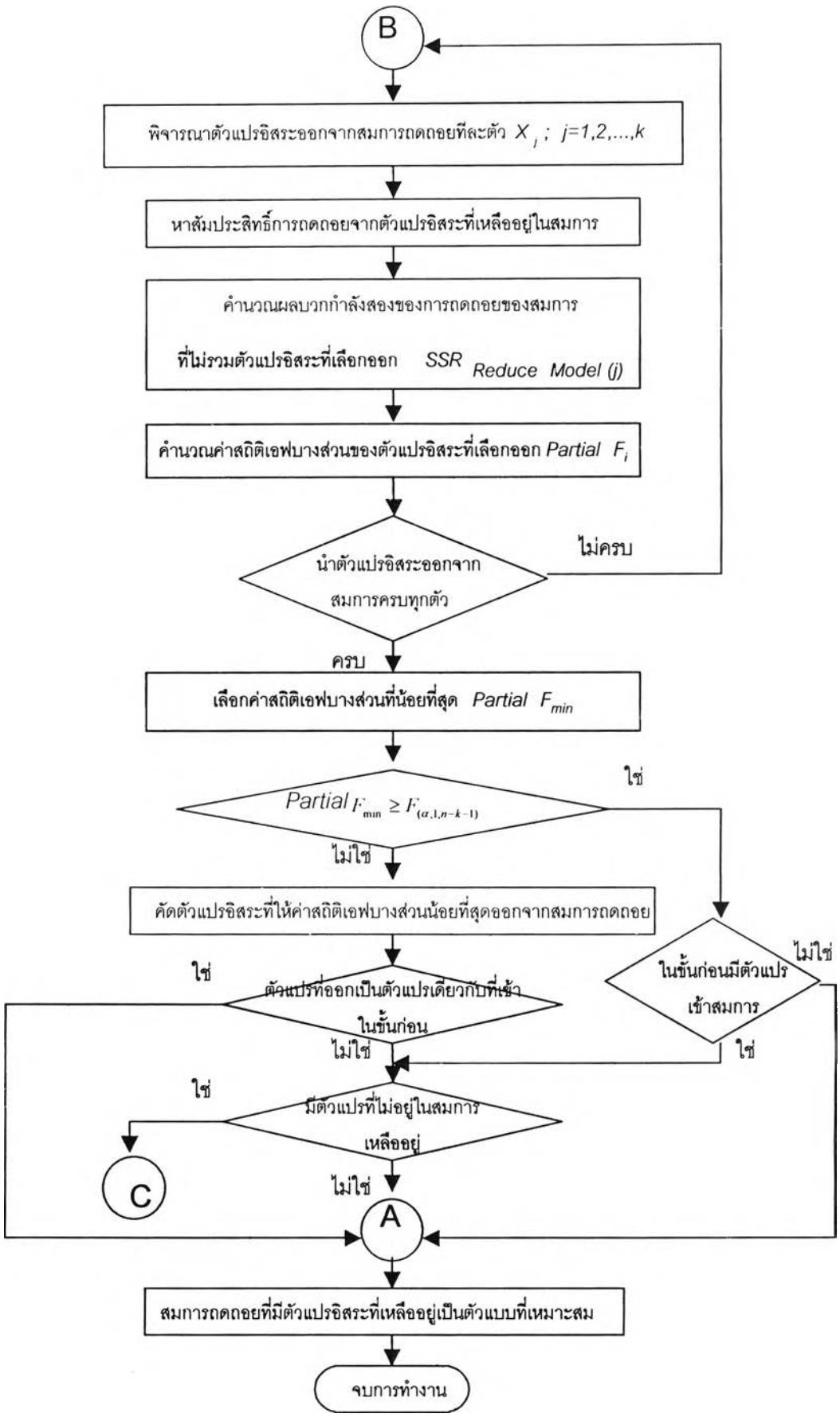
วิธีการนี้จะเริ่มจากการคัดเลือกตัวแปรอิสระที่มีความสัมพันธ์กับตัวแปรตาม (r_{xy}) มากที่สุด (X_j) เข้าสู่สมการถดถอย โดยที่ $r_{xy}^2 = 1 - \frac{SSE}{SST}$ จากนั้นจะหาสัมประสิทธิ์สหสัมพันธ์บางส่วนระหว่างตัวแปรตามกับตัวแปรอิสระแต่ละตัวที่ยังไม่อยู่ในสมการถดถอย โดยถือว่าได้รวมตัวแปรอิสระ X_j ไว้ในตัวแบบแล้ว และเลือกตัวแปรอิสระที่มีค่าสัมประสิทธิ์สหสัมพันธ์บางส่วนสูงที่สุดและมีนัยสำคัญเข้าสู่ตัวแบบ ขั้นตอนต่อไปเป็นการคัดเลือกตัวแปรอิสระออก โดยจะตัดตัวแปรอิสระที่มีค่าสถิติเอฟบางส่วนน้อยที่สุด และไม่มีนัยสำคัญออกจากตัวแบบ จะทำซ้ำการคัดเลือกตัวแปรอิสระเข้าและออกจากตัวแบบจนกระทั่งเงื่อนไขการหยุดข้อใดข้อหนึ่งต่อไปนี้เป็นจริง

1) ไม่มีตัวแปรอิสระที่สามารถเข้าหรือออกจากตัวแบบในขั้นถัดกัน
2) ตัวแปรอิสระที่เข้าและออกจากตัวแบบในขั้นถัดกันเป็นตัวแปรเดียวกัน
3) ในขั้นตอนการออกไม่มีตัวแปรอิสระเหลืออยู่ในตัวแบบ
4) ในขั้นตอนการเข้าไม่เหลือตัวแปรอิสระที่ไม่อยู่ในสมการถดถอย

ผังงานแสดงขั้นตอนของวิธีการถดถอยแบบขั้นบันได ดังแสดงในรูปที่ 3.7



รูปที่ 3.7 แสดงผังงานสำหรับขั้นตอนของวิธีการถดถอยแบบขั้นบันได



รูปที่ 3.7 (ต่อ) แสดงผังงานสำหรับขั้นตอนของวิธีการถดถอยแบบขั้นบันได

ขั้นตอนที่ 5

คำนวณหาค่าเฉลี่ยของค่าคลาดเคลื่อนกำลังสองเฉลี่ย (AMSE) จากการทำซ้ำจำนวน 500 รอบ ของทั้ง 5 วิธีเพื่อใช้เป็นเกณฑ์ในการตัดสินใจ และคำนวณค่าส่วนเบี่ยงเบนมาตรฐานของค่าคลาดเคลื่อนกำลังสองเฉลี่ย รวมถึงค่าอัตราส่วนผลต่างของค่าเฉลี่ยของค่าคลาดเคลื่อนกำลังสองเฉลี่ย (RDAMSE) ของทั้ง 5 วิธีดังกล่าวข้างต้น เพื่อเป็นเกณฑ์ประกอบการตัดสินใจ

ขั้นตอนที่ 6

ทำการเปรียบเทียบค่าเฉลี่ยของค่าคลาดเคลื่อนกำลังสองเฉลี่ย(AMSE) สำหรับตัวแบบที่ได้จากการคัดเลือกตัวแบบการถดถอยทั้ง 5 วิธี โดยสรุปผลในรูปของตารางและรูปภาพเพื่อแสดงการเปรียบเทียบและศึกษาแนวโน้มของแต่ละวิธี

สำหรับรายละเอียดของโปรแกรมทั้งหมดสามารถดูได้จากภาคผนวก