บทที่ 3

วิธีดำเนินการวิจัย

การวิจัยครั้งนี้ต้องการศึกษาเกี่ยวกับการประมาณค่าสัมประสิทธิ์การถดถอยพหุ เมื่อความผิดพลาดมีการแจกแจงแบบเบ้และมีการแจกแจงแบบหางยาวกว่าการแจกแจงแบบปกติ ในกรณีที่การแจกแจงของความผิดพลาดเป็นแบบหางยาวกว่าการแจกแจงแบบปกติจะศึกษาวิธีการประมาณ 4 วิธี คือวิธีกำลังสองน้อยที่สุด วิธี M-estimator โดยใช้ค่า standard deviation ของ ε_i เป็นตัวประมาณสเกล วิธี M-estimator โดยใช้ค่า median absolute deviation เป็นตัวประมาณสเกล และวิธี M-estimator โดยใช้ค่า the modified biweight A-estimator เป็นตัวประมาณสเกล ซึ่งวิธีการประมาณค่าแบบ M-estimator นี้ใช้เกณฑ์ความแกร่งของ Huber เป็นหลัก และพิจารณาการแจกแจงของความผิดพลาดที่มีการแจกแจงแบบปกติปลอมปนและแบบที ส่วนกรณีที่มีการแจกแจงความผิดพลาดเป็นแบบเบ้ต้องอาศัยการแปลงค่าที่อยู่ในรูปยกกำลังของ Box และ Cox เพื่อศึกษาวิธีการประมาณค่าสัมประสิทธิ์การถดถอยเชิงเส้นทั้ง 4 วิธี และพิจารณาการแจกแจงของความผิดพลาดที่มีการแจกแจงแบบลอกนอร์มอล แกมมา และไวบูลล์ โดยจะพิจารณาประสิทธิภาพการประมาณค่าสัมประสิทธิ์การถดถอยของทั้ง 4 วิธีด้วยอัตราส่วนผลต่างของค่าเฉลี่ยความผิดพลาดกำลังสอง* และอัตราส่วนผลต่างของความผิดพลาดกำลังสอง** ซึ่งวิธีดำเนินการวิจัยดังกล่าวมีวิธีการและแผนดำเนินงานดังต่อไปนี้

ศูนย์วิทยทรัพยากร
จุฬาลงกรณ์มหาวิทยาลัย

---

* **อัตราส่วนผลต่างค่าเฉลี่ยของความผิดพลาดกำลังสอง** (ratio of different average mean square error (RDAMSE)) คือผลต่างของวิธีที่ให้ค่าเฉลี่ยความผิดพลาดกำลังสองต่ำที่สุดกับวิธีที่เหลืออีก 4 วิธี หารด้วยค่าเฉลี่ยความผิดพลาดกำลังสองที่มีค่าต่ำสุด

** **อัตราส่วนผลต่างของความผิดพลาดกำลังสอง** (ratio of different mean square error (RDMSE)) คือผลต่างของวิธีที่ให้ค่าเฉลี่ยความผิดพลาดกำลังสองต่ำที่สุดกับวิธีที่เหลืออีก 4 วิธี หารด้วยค่าเฉลี่ยความผิดพลาดกำลังสองที่มีค่าต่ำสุด หลังจากนั้นหาผลรวมของการกระทำซ้ำ แล้วหารด้วยจำนวนครั้งของแต่ละวิธีที่ให้ค่าเฉลี่ยความผิดพลาดกำลังสองต่ำที่สุด

3.1 วิธีมอนติคาร์โล (Monte Carlo method)

เทคนิคการจำลองตัวแบบทางคณิตศาสตร์ที่นิยมใช้วิธีหนึ่ง คือ วิธีมอนติคาร์โล ซึ่งเป็นการจำลองตัวเลขสุ่ม (random number) มาช่วยในการหาคำตอบของปัญหาที่ต้องการศึกษา โดยการสร้างข้อมูลให้มีสภาพการแจกแจงตามที่ต้องการ ซึ่งขั้นตอนของวิธีมอนติคาร์โลที่ใช้กันอยู่ในปัจจุบันแบ่งเป็น 3 ขั้นตอนดังนี้

3.1.1 <u>การสร้างตัวเลขสุ่ม</u> การใช้ตัวเลขสุ่มเป็นสิ่งสำคัญมากในวิธีมอนติคาร์โล เพราะว่าหลักการของวิธีมอนติคาร์โลนั้นจะใช้ตัวเลขสุ่มมาช่วยในการหาคำตอบของปัญหาซึ่งลักษณะของตัวเลขสุ่มจะมีการแจกแจงแบบสม่ำเสมอ (uniform distribution) ในช่วง (0,1) สำหรับวิธีสร้างตัวเลขสุ่มนั้นมีผู้เสนอไว้หลายวิธี แต่วิธีหนึ่งที่ดีก็คือ วิธีของไวท์และชมิดท์ (White and Schemidt 1795:421) ที่ได้เสนอไว้กล่าวคือ ลักษณะของตัวเลขสุ่มที่ได้มีการแจกแจงแบบสม่ำเสมอในช่วง (0,1) และเป็นอิสระซึ่งกันและกัน ซึ่งรายละเอียดเสนอไว้ในภาคผนวก ก.

3.1.2 <u>การประยุกต์ปัญหาที่ต้องการศึกษาเพื่อใช้กับตัวเลขสุ่ม</u> ซึ่งขั้นตอนนี้ขึ้นอยู่กับลักษณะของปัญหาที่ต้องการศึกษา บางปัญหาอาจไม่ใช้ตัวเลขสุ่มโดยตรง แต่จะมีขั้นตอนอื่นอีกหลายๆ ขั้นตอนซึ่งบางครั้งต้องใช้ตัวเลขสุ่ม

3.1.3 <u>การทดลองกระทำซ้ำ</u> (replication) เมื่อประยุกต์ปัญหาให้ใช้กับตัวเลขสุ่มได้แล้ว ขั้นต่อไปก็คือการทดลองโดยกระบวนการของเลขสุ่ม (random process) มากระทำในลักษณะที่กระทำซ้ำๆ กันเพื่อหาคำตอบของปัญหาที่ต้องการศึกษา

ศูนย์วิทยทรัพยากร

3.2 <u>แผนดำเนินการวิจัย</u>

จุฬาลงกรณ์มหาวิทยาลัย

ในการวิจัยครั้งนี้ได้กำหนดสถานการณ์ต่างๆ โดยการสุ่มตัวอย่างจากประชากรที่มีลักษณะการแจกแจงแบบเดียวกัน เช่น ลักษณะการแจกแจงแบบหางยาวกว่าการแจกแจงแบบปกติที่ต้องการศึกษาได้แก่ การแจกแจงแบบปกติปลอมปนและแบบที และลักษณะการแจกแจงแบบเบ้ที่ต้องการศึกษาได้แก่ การแจกแจงแบบลอกนอร์มอล แกมมา และไวบูลล์

สำหรับการแจกแจงแบบปกติปลอมปนจะกำหนดเปอร์เซนต์การปลอมปนเป็น 5 10

20 และ 30 ส่วนสเกลแฟกเตอร์มี 3 ระดับคือ 5 10 และ 15 ซึ่งถ้ามีสเกลแฟกเตอร์สูงจะทำให้เกิดค่าผิดปกติสูง สามารถแสดงรายละเอียดได้ในตารางที่ 3.1

ตารางที่ 3.1 แสดงค่าสเกลแฟกเตอร์และเปอร์เซนต์การปลอมปนทั้งหมดที่ใช้ในการวิจัย

C \ P	5	10	20	30
5	(5,5)	(5,10)	(5,20)	(5,30)
10	(10,5)	(10,10)	(10,20)	(10,30)
15	(15,5)	(15,10)	(15,20)	(15,30)

สำหรับการแจกแจงแบบทีจะสนใจศึกษาเมื่อระดับความเป็นอิสระ(DF) = 4 และ 8 โดยเมื่อระดับความเป็นอิสระน้อยจะทำให้เกิดหางหนาและยาวกว่าแบบปกติมากขึ้น

สำหรับการแจกแจงแบบลอกนอร์มอลสนใจศึกษาเมื่อค่า $\mu = 0, \sigma^2 = 1$ ซึ่งเป็นค่าเฉลี่ยและค่าความแปรปรวนของการแจกแจงแบบปกติ

สำหรับการแจกแจงแบบแกมมาจะพิจารณาเมื่อ $\beta = 5, 10$ และ 150 , $\alpha = 1$ และ 2 โดยที่มีประสิทธิภาพความแปรปรวนเป็น 100% และ 70% ตามลำดับ

สำหรับการแจกแจงแบบไวบูลล์จะพิจารณาเมื่อ $\beta = 5, 10$ และ 150 , $\alpha = 1$ และ 2 โดยที่มีประสิทธิภาพความแปรปรวนเป็น 100% และ 52% ตามลำดับ

ขนาดตัวอย่างที่ใช้สำหรับการแจกแจงของความผิดพลาดแบบปกติปลอมปน ลอกนอร์มอล แกมมา และไวบูลล์ เมื่อมีจำนวนตัวแปรอิสระ m = 3 จะใช้ขนาดตัวอย่าง n = 20 และเมื่อมีจำนวนตัวแปรอิสระ m = 5, 10 จะใช้ขนาดตัวอย่าง n = 50, 100 และ 150 ตามลำดับส่วนการแจกแจงของความคลาดเคลื่อนแบบที จะใช้ขนาดตัวอย่าง n = 20 และมีจำนวนตัวแปรอิสระ m = 3, 5 และ 10 ตามลำดับเท่านั้น

3.3 <u>ขั้นตอนของการดำเนินงานวิจัย</u>

ศูนย์วิทยทรัพยากร
จุฬาลงกรณ์มหาวิทยาลัย

การวิจัยนี้สนใจศึกษาการประมาณค่าสัมประสิทธิ์การถดถอยพหุ เมื่อค่าประมาณสเกลเปลี่ยนไป ในที่นี้พิจารณา เมื่อความผิดพลาดของข้อมูลมีการแจกแจงแบบหางยาวกว่าการแจกแจงแบบปกติ และแบบเบ้ โดยการสร้างโปรแกรมหลักจำนวน 2 โปรแกรม สำหรับการแจกแจงที่มีหางยาวกว่าการแจกแจงแบบปกติ และการแจกแจงแบบเบ้ ด้วยภาษาฟอร์แทน IV กับเครื่อง IBM 370/3031 ซึ่งแต่ละโปรแกรมมีขั้นตอนในการดำเนินงานดังนี้

3.3.1 กำหนดลักษณะการแจกแจงของความผิดพลาด (II) ขนาดตัวอย่างและจำนวนตัวแปรอิสระ เพื่อนำไปสร้างตัวแปรตาม (y)

3.3.2 การสร้างข้อมูล คือตัวแปรตาม (y) และตัวแปรอิสระ (X) เมื่อการแจกแจงของความผิดพลาดเป็นแบบหางยาวกว่าการแจกแจงปกติ ตัวแปรตาม (y) สร้างจากตัวแปรอิสระ (X) ซึ่งเป็นค่าคงที่ และค่าความผิดพลาดมีลักษณะการแจกแจงตามที่ต้องการศึกษา โดยให้ตัวแปรตามมีความสัมพันธ์เชิงเส้นตรงกับตัวแปรอิสระ ส่วนกรณีที่การแจกแจงของความผิดพลาดเป็นแบบเบ้ ตัวแปรตาม (y) สร้างให้มีลักษณะการแจกแจงตามที่ต้องการศึกษาโดยไม่เป็นฟังก์ชันเชิงเส้นตรงกับตัวแปรอิสระ (X) และค่าความคลาดเคลื่อน

3.3.3 ทดสอบความเหมาะสมของตัวแปรอิสระ (y) โดยทดสอบความเป็นอิสระซึ่งกันและกันหรือไม่ เรียกว่าการทดสอบ Multicollinearity

3.3.4 การประมาณค่าสัมประสิทธิ์การถดถอยพหุเชิงเส้น เมื่อความผิดพลาดมีการแจกแจงแบบหางยาวกว่าการแจกแจงแบบปกติ โดยใช้วิธีกำลังสองน้อยที่สุด วิธี M-estimator เมื่อใช้ standard deviation ของ ϵ_i เป็นตัวประมาณสเกล (scale estimator) วิธี M-estimator เมื่อใช้ median absolute deviation (MAD) เป็น ตัวประมาณสเกล และวิธี M-estimator เมื่อใช้ the modified biweight A-estimator ที่ C = 9 และ 10 เป็นตัวประมาณสเกล ในที่นี้ใช้เกณฑ์ความแกร่งของ Huber

3.3.5 การประมาณค่าสัมประสิทธิ์การถดถอยพหุเชิงเส้น เมื่อความผิดพลาดมีการแจกแจงแบบเบ้จะอาศัยการแปลงที่อยู่ในรูปยกกำลังของ Box และ Cox แล้วประมาณค่าสัมประสิทธิ์การถดถอยพหุเชิงเส้นโดยใช้วิธีกำลังสองน้อยที่สุด วิธี M-estimator เมื่อใช้ standard deviation ของ ϵ_i เป็นตัวประมาณสเกล (scale estimator) วิธี M-estimator เมื่อใช้ median absolute deviation (MAD) เป็นตัวประมาณสเกล และวิธี M-estimator

ศูนย์วิทยทรัพยากร
จุฬาลงกรณ์มหาวิทยาลัย

เมื่อใช้ the modified biweight A-estimator ที่ C = 9 และ 10 เป็นตัวประมาณสเกล ในที่นี้ใช้เกณฑ์ความแกร่งของ Huber

3.3.6 คำนวณหาค่าความผิดพลาดกำลังสองของการประมาณค่าสัมประสิทธิ์การถดถอย (mean square error) อัตราส่วนผลต่างของความผิดพลาดกำลังสองของการประมาณค่าสัมประสิทธิ์การถดถอย โดยเทียบกับวิธีที่ให้ค่าความผิดพลาดกำลังสองต่ำที่สุด, ค่าความแปรปรวนของผลต่างความผิดพลาดกำลังสอง ค่าเฉลี่ยของความผิดพลาดกำลังสอง (AMSE) ค่าความแปรปรวนของค่าเฉลี่ยความผิดพลาดกำลังสอง (STD) และอัตราส่วนผลต่างค่าเฉลี่ยของความผิดพลาดกำลังสอง โดยเทียบกับวิธีที่ให้ค่าเฉลี่ยความผิดพลาดกำลังสองต่ำที่สุด (DIFF) ของตัวประมาณทั้ง 4 วิธี

3.3.7 สรุปผลในรูปของตาราง

3.4 ผังงานวิธีดำเนินการวิจัย

ขั้นตอนในการดำเนินงานสามารถแสดงเป็นผังงานได้ ดังรูปที่ 3.1 ส่วนรายละเอียดของโปรแกรมแสดงไว้ในภาคผนวก ค. และความหมายของสัญลักษณ์ที่ใช้ในผังงานมีดังนี้

II รูปแบบของการแจกแจงของความผิดพลาด

N จำนวนค่าสังเกต

M จำนวนพารามิเตอร์

IC สเกลแฟคเตอร์

P เปอร์เซนต์การปลอมปน

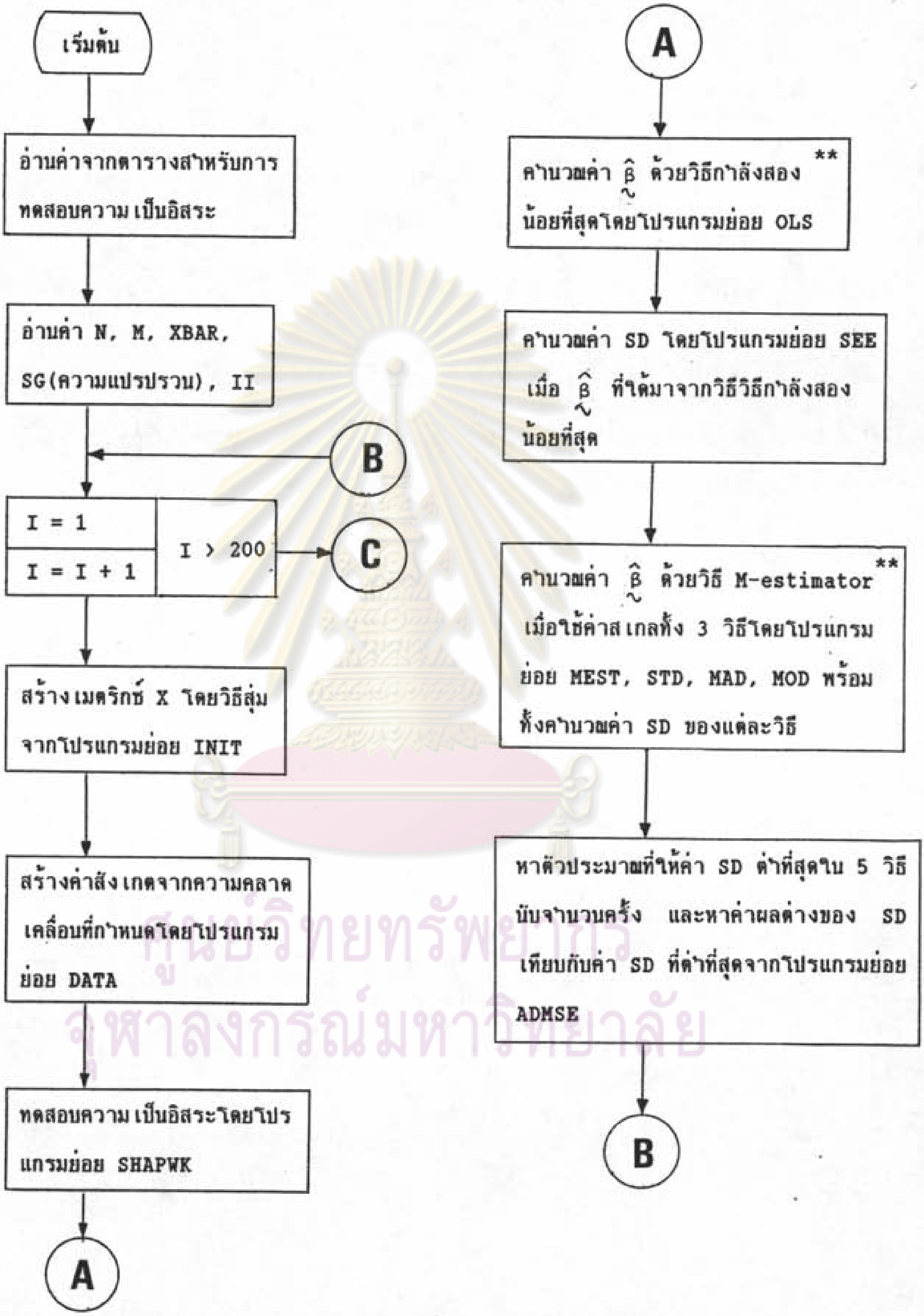
AMSE ค่าเฉลี่ยของความผิดพลาดกำลังสอง

STD ค่าเฉลี่ยของความแปรปรวนของความผิดพลาดกำลังสอง

ARDIF ผลต่างของค่าเฉลี่ยของความผิดพลาดกำลังสองเมื่อเทียบกับค่าเฉลี่ยของความผิดพลาดกำลังสอง

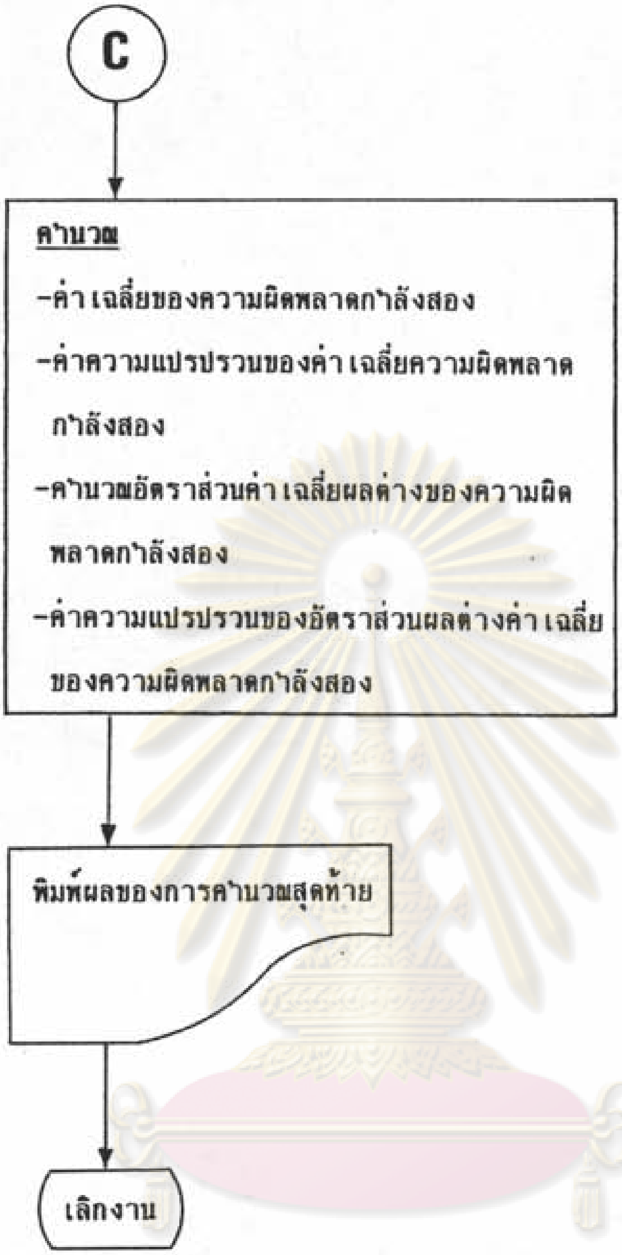
ศูนย์วิทยทรัพยากร
จุฬาลงกรณ์มหาวิทยาลัย

รูปที่ 3.1 แสดงผังงานสำหรับการประมาณค่าสัมประสิทธิ์การถดถอยพหุ เมื่อค่าประมาณสเกลเปลี่ยนไป



ศูนย์วิทยทรัพยากร
จุฬาลงกรณ์มหาวิทยาลัย

** กรณีที่ค่าความผิดพลาดมีการแจกแจงแบบเบ้จะต้องทำการแปลงข้อมูลที่อยู่ในรูปยกกำลังของ Box และ Cox ก่อนการประมาณค่าทุกวิธี



ศูนย์วิทยทรัพยากร
จุฬาลงกรณ์มหาวิทยาลัย

จากรูปที่ 3.1 สามารถแยกอธิบายแต่ละขั้นตอนได้ดังนี้

3.4.1 เริ่มต้นระบบโดยการอ่านค่าจากตารางสำหรับการทดสอบความเป็นอิสระได้แก่ ตารางค่าสัมประสิทธิ์การทดสอบของ Shapiro และ Wilk และตาราง percentage point ของการทดสอบ Shapiro และ Wilk ทั้งนี้พิจารณาตามขนาดของตัวอย่างค่าสังเกต n = 20 50 100 และ 150 ซึ่งรายละเอียดแสดงไว้ในภาคผนวก ข.

3.4.2 อ่านค่าขนาดตัวอย่าง (N) จำนวนตัวแปรอิสระ (M) ค่าเฉลี่ย (XBAR) ค่าความแปรปรวน (SG) ของการแจกแจงแบบปกติ และลักษณะการแจกแจงของประชากร (II)

3.4.3 สร้างเมตริกซ์ X ที่มีการแจกแจงแบบปกติมีขนาด n x m โดยวิธีสุ่มพร้อมทั้งสร้างเมตริกซ์ $(X'X)^{-1}$ จากโปรแกรมย่อย INIT

3.4.4 สร้างค่าสังเกตจากความคลาดเคลื่อนที่กำหนดตาม II จากโปรแกรมย่อย DATA ได้แก่

3.4.4.1 <u>การแจกแจงปกติ</u> ฟังก์ชันที่ใช้ในการแจกแจงแบบปกติจะใช้วิธีของ Box และ Muller (1958) ซึ่งเป็นวิธีที่สร้างการแจกแจงแบบปกติที่มีค่าเฉลี่ยเป็น μ และความแปรปรวนเป็น σ^2 โดย DMEAN และ $(SIGMA)^2$ คือค่าเฉลี่ยและความแปรปรวนที่กำหนดสำหรับรายละเอียดแสดงไว้ในภาคผนวก ก. การใช้ฟังก์ชันนี้ใช้คำสั่ง NORMAL(DMEAN,SIGMA) ผลลัพธ์คือ NORMAL ที่มีค่าเฉลี่ยและความแปรปรวนที่กำหนด

3.4.4.2 <u>การแจกแจงแบบปกติปลอมปน</u> สร้างโดยการใช้วิธีการแปลงข้อมูลมาจากการแจกแจงแบบปกติที่ได้มาจาก NORMAL(DMEAN,SIGMA) ด้วยความน่าจะเป็น (1-P) เมื่อ DMEAN SIGMA และ P เป็นค่าเฉลี่ย ความแปรปรวนและเปอร์เซนต์การปลอมปน และได้มาจาก NORMAL(DMEAN,SG2) ด้วยความน่าจะเป็น P เมื่อ SG2 เป็นผลคูณของ IC กับ SIGMA IC เป็นสเกลแฟกเตอร์สำหรับความแปรปรวนโดยที่รายละเอียดแสดงไว้ในภาคผนวก ก.

3.4.4.3 <u>การแจกแจงแบบที</u> ฟังก์ชันที่ใช้คือ TDIST(NDF,DMEAN,SIGMA) โดยที่ผลลัพธ์คือค่าของ TDIST มีระดับความเป็นอิสระเป็น NDF ซึ่งสร้างจากสัดส่วนของการแจกแจงแบบปกติ และการแจกแจงแบบไคสแควร์มีระดับความเป็นอิสระเป็น NDF ค่าเฉลี่ยและความแปรปรวนของการแจกแจงแบบปกติเป็น DMEAN และ SIGMA สำหรับรายละเอียดแสดงไว้ในภาค

ศูนย์วิทยทรัพยากร
จุฬาลงกรณ์มหาวิทยาลัย

ผนวก ก.

3.4.4.4 <u>การแจกแจงแบบลอกนอร์มอล</u> หาได้จากการหา exponential ของฟังก์ชัน NORMAL(DMEAN,SIGMA) เมื่อ DMEAN และ SIGMA เป็นค่าเฉลี่ยและความแปรปรวนของการแจกแจงแบบปกติสำหรับรายละเอียดแสดงไว้ในภาคผนวก ก.

3.4.4.5 <u>การแจกแจงแบบแกมมา</u> ฟังก์ชันที่ใช้คือ GAMMA1(ALPHA1,BETA1) ฟังก์ชันนี้สร้างโดยวิธี reproductive property ซึ่งรายละเอียดแสดงไว้ในภาคผนวก ก. ฟังก์ชัน GAMMA1 จะกำหนดค่า ALPHA1 และ BETA1 เพื่อให้ได้ค่าเฉลี่ยและความแปรปรวนตามที่ต้องการ ผลลัพธ์คือค่า GAMMA1 จะมีการแจกแจงแบบแกมมามีค่าเฉลี่ยเป็น BETA1 x ALPHA1 และความแปรปรวนเป็น $BETA1^2$ x ALPHA1

3.4.4.6 <u>การแจกแจงแบบไวบูลล์</u> ฟังก์ชันที่ใช้คือ WEIBUL(ALPHA1,BETA1) ฟังก์ชันนี้สร้างโดยวิธี inverse transformation ซึ่งรายละเอียดแสดงไว้ในภาคผนวก ก. ฟังก์ชัน WEIBUL จะกำหนดค่า ALPHA1 และ BETA1 เพื่อให้ได้ค่าเฉลี่ยและความแปรปรวนตามที่ต้องการ ผลลัพธ์คือ ค่า WEIBUL จะมีการแจกแจงแบบไวบูลล์มีค่าเฉลี่ยเป็น BETA1 x (1-1/ALPHA1) และความแปรปรวนเป็น $BETA1^2 \times [\Gamma(1+2/ALPHA1)-\Gamma^2(1+1/ALPHA1)]$

หลังจากนั้นกำหนดค่าสังเกต $\underset{\sim}{y}$ ให้มีความสัมพันธ์เชิงเส้นโดยพิจารณาการแจกแจงของค่าผิดพลาด ดังนี้

3.4.4.7 <u>เมื่อการแจกแจงของค่าผิดพลาดเป็นแบบหางยาวกว่าการแจกแจงแบบปกติ</u>

ในการวิจัยครั้งนี้จะสร้างตัวแปรอิสระ X ซึ่งมีค่าคงที่ก่อน แล้วจึงสร้างตัวแปร $\underset{\sim}{y}$ ที่มีความสัมพันธ์เชิงเส้นกับ X ให้มีลักษณะการแจกแจงตามความผิดพลาดที่กำหนด II และตามรูปแบบของความสัมพันธ์คือ $\underset{\sim}{y} = X\underset{\sim}{\beta} + \underset{\sim}{\varepsilon}$ เมื่อ $\underset{\sim}{\beta}$ เป็นค่าพารามิเตอร์ที่ถูกกำหนดขึ้นมา และ $\underset{\sim}{\varepsilon}$ เป็นความผิดพลาดที่มีการแจกแจงเป็นแบบต่างๆดังที่กล่าวแล้วใน 3.4.4.1 ถึง 3.4.4.3 สำหรับการสร้าง $\underset{\sim}{y}$ นั้นจะเริ่มจากการกำหนดจำนวนตัวแปรอิสระ และขนาดตัวอย่างที่ต้องการศึกษา พารามิเตอร์ $\underset{\sim}{\beta}$ ค่าเฉลี่ย ค่าความแปรปรวน และลักษณะการแจกแจงของค่าผิดพลาด

ศูนย์วิทยทรัพยากร
จุฬาลงกรณ์มหาวิทยาลัย

รวมทั้งค่าคงที่ X แล้วจึงสร้าง y ตามรูปแบบดังกล่าว

3.4.4.8 เมื่อการแจกแจงของค่าผิดพลาดเป็นแบบเบ้

ในการวิจัยครั้งนี้จะสร้างตัวแปรอิสระ X ซึ่งมีค่าคงที่ก่อน แล้วจึงสร้างตัวแปร $\underset{\sim}{y}$ ให้มีการแจกแจงแบบเบ้ตาม 3.4.4.4 ถึง 3.4.4.6 แล้วอาศัยการแปลงที่อยู่ในรูปยกกำลังของ Box และ Cox เพื่อสร้างตัวแปร $\underset{\sim}{y}^{(\lambda)}$ ที่มีความสัมพันธ์เชิงเส้นกับ X ตามรูปแบบของความสัมพันธ์คือ $\underset{\sim}{y}^{(\lambda)} = X\underset{\sim}{\beta} + \underset{\sim}{\varepsilon}$ เมื่อ $\underset{\sim}{\beta}$ เป็นค่าพารามิเตอร์ที่ถูกกำหนดขึ้นมา λ เป็นค่าพารามิเตอร์ของการแปลงที่อยู่ในรูปยกกำลังของ Box และ Cox และ $\underset{\sim}{\varepsilon}$ เป็นความผิดพลาดที่มีการแจกแจงเป็นแบบสมมาตรเชิงเส้นและเป็นแบบปกติโดยประมาณ

3.4.5 การทดสอบความเหมาะสมของตัวแปรอิสระ X โดยทดสอบความเป็นอิสระซึ่งกันและกัน เนื่องจากถ้าตัวแปรอิสระเกิด multicollinearity จะทำให้การประมาณค่าสัมประสิทธิ์การถดถอยพหุด้วยวิธีกำลังสองน้อยที่สุด และวิธี M-estimator ลดประสิทธิภาพลง (Mongomery(ค.ศ.1982 หน้า 364-381) Mason และ Gunst (ค.ศ.1985 หน้า 401-407)) และผลของการที่ตัวแปรอิสระมีความสัมพันธ์กันหรือเกิด multicollinearity จะทำให้เมตริกซ์ X เกือบจะเป็น sigular matrix นั่นคือ determinant มีค่าน้อยและ inverse มีค่าสูงซึ่งจะทำให้ความแปรปรวนของสมการถดถอยมีค่าสูง และมีผลทำให้ยอมรับ H_o บ่อยครั้งดังนั้นการสรุปความสัมพันธ์ของ $\underset{\sim}{y}$ และ X อาจเกิดความผิดพลาดได้

การทดสอบว่าเกิด multicollinearity หรือไม่อาจทำการทดสอบสมมติฐานดังนี้

สมมติฐานของการทดสอบ H_o : ไม่เกิด multicollinearity

เทียบกับ H_a : เกิด multicollinearity

ณ.ระดับสำคัญ α = 0.05 และ 0.01

ตัวสถิติทดสอบ $$F_{(m-2),(n-m+1)} = \frac{R_i^2/(m-2)}{(s-R_i^2)/(n-m+1)}$$ (ไม่รวม intercept)

โดยที่ R_i^2 เป็นสหสัมพันธ์ (correlation) ของ Regression ที่ X_i บน X $(j \neq i)$;

ศูนย์วิทยทรัพยากร
จุฬาลงกรณ์มหาวิทยาลัย

$i,j = 1,\ldots, m$

ขอบเขตวิกฤต F จะหาได้จากตาราง ฌ.ระดับนัยสำคัญ 0.05 และ 0.01

การตัดสินใจ จะปฏิเสธ H_o เมื่อ F (คำนวณ) > F (ตาราง)

3.4.6 การประมาณค่าสัมประสิทธิ์การถดถอยพหุด้วยวิธีกำลังสองน้อยที่สุด

3.4.6.1 เมื่อความผิดพลาดมีการแจกแจงเป็นแบบหางยาวกว่าการแจกแจงแบบปกติ โปรแกรมย่อยที่ใช้ในการประมาณค่าสัมประสิทธิ์การถดถอยคือ OLS(B) โดยที่ B เป็นเวกเตอร์ของสัมประสิทธิ์การถดถอยพหุ ซึ่งเมตริกซ์ของ X จะใช้ร่วมกับโปรแกรมหลัก เมตริกซ์ X ประกอบด้วยตัวแปรอิสระ M ตัวและตัวแปรตาม 1 ตัว N คือจำนวนข้อมูล และโปรแกรมย่อย OLS(B) จะประกอบด้วยโปรแกรมย่อยอีกดังนี้

ก) โปรแกรมย่อยสำหรับหาเมตริกซ์ผกผัน (inverse matrix) คือโปรแกรม INVS(M,A) เมื่อ M เป็นขนาดของเมตริกซ์ผกผันที่ต้องการหาและ A เป็นเมตริกซ์ผกผันที่ต้องการ

ข) โปรแกรมย่อยสำหรับประมาณค่าผิดพลาด คือโปรแกรม YRESID (X,N,M,YHAT,YRES,B) โดยที่ X เป็นเมตริกซ์ของตัวแปรอิสระและตัวแปรตาม N คือจำนวนข้อมูล M คือจำนวนตัวแปรอิสระ YHAT คือค่าประมาณ ε_i YRES คือค่าผิดพลาด และ B คือสัมประสิทธิ์การถดถอยพหุ

3.4.6.2 เมื่อการแจกแจงของค่าผิดพลาดมีการแจกแจงแบบเบ้ จะอาศัยการแปลงที่อยู่ในรูป Box และ Cox โดยโปรแกรย่อย BOXCOX(B,FL1) ซึ่ง B คือค่าสัมประสิทธิ์การถดถอยพหุ FL1 คือ $\hat{\lambda}$ ซึ่งเป็นค่าประมาณ λ ที่อยู่ในรูปยกกำลังของ Box และ Cox โดยมีขั้นตอนการทำงานดังนี้

ศูนย์วิทยทรัพยากร
จุฬาลงกรณ์มหาวิทยาลัย

ขั้นตอน	ความหมาย
1	คำนวณค่าเฉลี่ยเรขาคณิต (geometric mean(G)) ของเวกเตอร์ y โดย y_i มีค่ามากกว่า 0 แล้วคำนวณค่า y1 = y/G
2	กำหนดค่าเริ่มต้น ST ค่าสุดท้าย FIN เป็นขอบเขตในการหาค่า $\hat{\lambda}$ หรือ FL1 โดยใช้เทคนิคของการหา $\hat{\lambda}$ ที่มีค่าผลบวกกำลังสองของค่าผิดพลาด หรือที่มีค่า $\hat{\sigma}^2_{\varepsilon\|\hat{\lambda}}$ ต่ำสุด รายละเอียดแสดงในภาคผนวก ข. ซึ่ง SSE2 แทน $\hat{\sigma}^2_{\varepsilon\|\hat{\lambda}}$ ที่มีค่าต่ำสุดซึ่งใช้ในการประมาณค่า FL1 และ B

3.4.7 การหาค่าเฉลี่ยความผิดพลาดกำลังสอง (Mean Square Error (MSE)) โดยโปรแกรมย่อย SEE(B,SE) เมื่อ B ที่ใช้มาจากวิธีกำลังสองน้อยที่สุดโดยการหาผลรวมของ MSE ของ $\hat{\underset{\sim}{\beta}}$ ที่ได้จากการประมาณเทียบกับ $\underset{\sim}{\beta}$ ของประชากรทุกค่าของ β_i ; i = 1,...., M แล้วหารด้วย M

$$\text{MSE ของวิธีกำลังสองน้อยที่สุด} = \frac{\sum_{i}^{M}(\hat{\beta}_i(\text{OLS}) - \hat{\beta}_i)^2}{M}$$

เมื่อ $\hat{\beta}_{i(OLS)}$ เป็น $\hat{\beta}_i$ ของวิธีกำลังสองน้อยที่สุด

3.4.8 การประมาณค่าสัมประสัมประสิทธิ์การถดถอยพหุด้วยวิธี M-estimator โดยใช้ค่าประมาณสเกลทั้ง 3 วิธี โปรแกรมที่เรียกใช้คือ MEST ซึ่งพิจารณาตามการแจกแจงของค่าผิดพลาดดังนี้

3.4.8.1 เมื่อความผิดพลาดเป็นแบบหางยาวกว่าการแจกแจงแบบปกติ จะเรียกโปรแกรมย่อยในการประมาณค่าสัมประสิทธิ์การถดถอยแยกวิธีตามการประมาณค่าสเกลดังนี้

ก) โปรแกรมย่อย BAR(B1,BMBAR,SIGMA) เมื่อ BMBAR คือค่า $\hat{\underset{\sim}{\beta}}$ ที่ได้จากการประมาณค่าโดยวิธี M-estimator เมื่อใช้ the standard deviation of location เป็นค่าประมาณสเกล และใช้เกณฑ์ความแกร่งของ Huber ที่ b = 2

ศูนย์วิทยทรัพยากร
จุฬาลงกรณ์มหาวิทยาลัย

ข) โปรแกรมย่อย MAD(B1,BMMAD,SIGMA) เมื่อ BMMAD คือค่า $\hat{\beta}$ ที่ได้จากการประมาณค่าโดยวิธี M-estimator เมื่อใช้ median absolute deviation เป็นค่าประมาณสเกล และใช้เกณฑ์ความแกร่งของ Huber เมื่อ b = 2 และ SIGMA คือค่าประมาณสเกล

ค) โปรแกรมย่อย MMOD(BMMAD,KCC,SIGMA1,BMMOD,SIGMA) เมื่อ BMMOD คือค่า $\underset{\sim}{\beta}$ ที่ได้จากการประมาณค่าโดยวิธี M-estimator เมื่อใช้ the modified biweight A-estimator เป็นค่าประมาณสเกล และใช้เกณฑ์ความแกร่งของ Huber เมื่อ b = 2 SIGMA คือค่าประมาณสเกลที่ได้จากวิธี MAD BMMAD คือค่าสัมประสิทธิ์การถดถอยที่ได้จากวิธี MAD SIGMA1 คือค่าประมาณสเกลของวิธี MMOD และ KCC คือค่าคงที่ C = 9 หรือ 10

3.4.8.2 เมื่อความผิดพลาดเป็นแบบเบ้ จะอาศัยการแปลงที่อยู่ในรูปยกกำลังของ Box และ Cox และประมาณค่าสัมประสิทธิ์การถดถอยพหุโดยใช้โปรแกรมย่อยตามลำดับดังนี้

ก) การประมาณค่าสัมประสิทธิ์การถดถอยพหุ โดยวิธี M-estimator ที่ใช้ค่า standard deviation of location เป็นตัวประมาณสเกล จากเทคนิคการแปลงที่อยู่ในรูปยกกำลังของ Box และ Cox โปรแกรมย่อยที่เรียกใช้คือ BOXC1(B,FL1,SIGMA) ซึ่ง B เป็นค่าสัมประสิทธิ์ของการถดถอยพหุ FL1 คือ λ ซึ่งเป็นค่าประมาณที่อยู่ในรูปยกกำลังของ Box และ Cox สำหรับโปรแกรมย่อย BOXC1(B,FL1,SIGMA) นี้ใช้คำตอบที่ได้จากโปรแกรม BOXCOX(B,FL1) ในการประมาณสัมประสิทธิ์การถดถอยพหุ และ $\hat{\lambda}$ ซึ่งเป็นค่าประมาณของการแปลง

ข) การประมาณค่าสัมประสิทธิ์การถดถอยพหุ โดยวิธี M-estimator ที่ใช้ค่า median absolute deviation เป็นตัวประมาณสเกล จากเทคนิคการแปลงที่อยู่ในรูปยกกำลังของ Box และ Cox โปรแกรมย่อยที่เรียกใช้คือ BOXC2(B,FL1,SIGMA) สำหรับโปรแกรมย่อย BOXC2(B,FL1,SIGMA) นี้ประมวลผลในทำนองเดียวกับโปรแกรม BOXC1 (B,FL1, SIGMA) และใช้คำตอบที่ได้จากโปรแกรม BOXCOX(B,FL1) ในการประมาณสัมประสิทธิ์การถด-

ศูนย์วิทยทรัพยากร
จุฬาลงกรณ์มหาวิทยาลัย

ถอยพหุ และ $\hat{\lambda}$ ซึ่งเป็นค่าประมาณของการแปลง

ค) การประมาณค่าสัมประสิทธิ์การถดถอยพหุ โดยวิธี M-estimator ที่ใช้ค่า the modified biweight A-estimator เป็นตัวประมาณสเกล จากเทคนิคการแปลงที่อยู่ในรูปยกกำลังของ Box และ Cox โปรแกรมย่อยที่เรียกใช้คือ BOXC3 (B,FL1,SIGMA) สำหรับโปรแกรมย่อย BOXC3(B,FL1,SIGMA) นี้ประมวลผลในทำนองเดียวกับโปรแกรม BOXC1 (B,FL1,SIGMA) และ BOXC2(B,FL1,SIGMA) และใช้คำตอบที่ได้จากโปรแกรม BOXCOX(B, FL1) ในการประมาณสัมประสิทธิ์การถดถอยพหุ และ $\hat{\lambda}$ ซึ่งเป็นค่าประมาณของการแปลง

ง) โปรแกรมย่อยสำหรับการคำนวณค่าผลบวกกำลังสองของค่าผิดพลาด โปรแกรมที่เรียกใช้คือ SUMBAR(FLX,SSE,K,SIGMA) SUMMAD(FLX,SSE,K,SIGMA) และ SUMMOD(FLX,SSE,K,BMMAD,SIGMA1,KCC,SIGMA) โดยที่ FLX เป็นค่า $\hat{\lambda}$ ที่ต้องการใช้ในการแปลงข้อมูล $\underset{\sim}{y}$ ให้อยู่ในรูปตัวแบบเชิงเส้นดังนี้

(3.4.8.2.1) $$\underset{\sim}{y}^{(\lambda)} = X\underset{\sim}{\beta} + \underset{\sim}{\varepsilon}$$

SSE คือ $\hat{\sigma}^2_{\varepsilon|\lambda}$ ที่คำนวณได้จากตัวแบบเชิงเส้นนี้ K เป็นค่าใช้สำหรับบอกตำแหน่งของ SSE ว่าเป็นค่าของจุดเริ่มต้น ค่ากลาง หรือค่าสุดท้ายในโปรแกรมที่เรียกใช้คือ BOXCOX BOXC1 BOXC2 และ BOXC3 SIGMA เป็นค่าประมาณสเกลของการประมาณสเกลในแต่ละวิธี

ค) โปรแกรมย่อยสำหรับการแปลงข้อมูลโดยเทคนิคการแปลงที่อยู่ในรูปยกกำลังของ Box และ Cox โปรแกรมย่อยที่ใช้คือ BCOX(FL1) ซึ่ง FL1 คือค่า $\hat{\lambda}$ สำหรับการแปลงข้อมูลหรือ y1 ให้อยู่ในรูปตัวแบบเชิงเส้น (3.4.8.2.1) โดยที่ y1 ใช้ข้อมูลร่วมกับโปรแกรมที่เรียกใช้คือ SUMBAR SUMMAD หรือ SUMMOD และผลลัพธ์จะถูกแทนเมตริกซ์ X ในแนวตั้งที่ M2 เพื่อใช้ในการประมาณสัมประสิทธิ์การถดถอยพหุต่อไป

หลังจากการประมาณค่าสัมประสิทธิ์การถดถอยพหุแล้ว อันดับต่อไปคือการหาค่าเฉลี่ยความผิดพลาดกำลังสองของการประมาณค่าสัมประสิทธิ์การถดถอยพหุด้วยวิธี BAR MAD และ MOD ที่ C = 9 และ 10 โดยการหาผลรวมของ MSE ของ $\hat{\beta}$ ที่ได้จากการประมาณเทียบกับ β ของประชากรทุกค่าของ β_i ;i = 1,..., M แล้วหารด้วย M

ศูนย์วิทยทรัพยากร
จุฬาลงกรณ์มหาวิทยาลัย

$$\text{MSE ของวิธี BAR} = \frac{\sum_{i}^{M} (\hat{\beta}_{i(BAR)} - \beta_i)^2}{M}$$

เมื่อ $\hat{\beta}_{i(BAR)}$ เป็น $\hat{\beta}_i$ ของวิธี M-estimator ที่ใช้ standard deviation of location เป็นตัวประมาณสเกล

$$\text{MSE ของวิธี MAD} = \frac{\sum_{i}^{M} (\hat{\beta}_{i(MAD)} - \beta_i)^2}{M}$$

เมื่อ $\hat{\beta}_{i(MAD)}$ เป็น $\hat{\beta}_i$ ของวิธี M-estimator ที่ใช้ median absolute deviation เป็นตัวประมาณสเกล

$$\text{MSE ของวิธี MOD} = \frac{\sum_{i}^{M} (\hat{\beta}_{i(MOD)} - \hat{\beta}_i)^2}{M}$$

เมื่อ $\hat{\beta}_{i(MOD)}$ เป็น $\hat{\beta}_i$ ของวิธี M-estimator ที่ใช้ modified biweight A-estimator ที่ C = 9 และ 10 เป็นตัวประมาณสเกล

3.4.9 หาค่าอัตราส่วนผลต่างความผิดพลาดกำลังสอง (Ratio of Different Mean Square Error (RDMSE)) จากวิธีที่ให้ค่าเฉลี่ยความผิดพลาดกำลังสองน้อยที่สุดมีค่าต่ำสุดกับวิธีที่เหลืออีก 4 วิธี หลังจากนั้นหาผลรวมของการกระทำซ้ำ แล้วหารด้วยจำนวนครั้งของแต่ละวิธีที่ให้ค่า MSE ต่ำที่สุด

ศูนย์วิทยทรัพยากร
จุฬาลงกรณ์มหาวิทยาลัย

$$\text{RDMSE} = \frac{\sum^{TIMES} \left[\frac{MSE_{nom} - MSE_{min}}{MSE_{min}}\right] \times 100}{TIMES}$$

เมื่อ MSE_{min} แทนวิธีที่ให้ค่า MSE ต่ำที่สุด

MSE_{nom} แทนวิธีที่ให้ค่า MSE โดยทั่วไปที่ไม่ใช่ค่าต่ำที่สุด

TIMES คือจำนวนครั้งของแต่ละวิธีที่ให้ค่า MSE ต่ำที่สุด

3.4.10 กลับไปกระทำซ้ำจากข้อ 3.4.3 ถึง 3.4.9 จนกระทั่งครบ 200 ครั้งแล้ว

หาค่าอัตราส่วนผลต่างค่าเฉลี่ยความผิดพลาดกำลังสอง (Ratio of Different Average Mean Square Error (RDAMSE)) จากวิธีที่ให้ค่าเฉลี่ยความผิดพลาดกำลังสองน้อยที่สุดมีค่าต่ำสุดกับวิธีที่เหลืออีก 4 วิธี

$$RDMSE = (AMSE_{nom} - AMSE_{min})/AMSE_{min}$$

เมื่อ $AMSE_{min}$ แทนวิธีที่ให้ค่า AMSE ต่ำที่สุด

$AMSE_{nom}$ แทนวิธีที่ให้ค่า AMSE โดยทั่วไปที่ไม่ใช่ค่าต่ำที่สุด

3.4.11 พิมพ์รายงานผลการคำนวณสุดท้ายและจบการทำงาน

ศูนย์วิทยทรัพยากร
จุฬาลงกรณ์มหาวิทยาลัย