บทที่ 3

วิธีดำเนินการวิจัย

สิ่งที่ต้องการศึกษาในการวิจัยครั้งนี้คือ การประมาณสัมประสิทธิ์การถดถอยพหุเมื่อความผิดพลาดมีการแจกแจงแบบหางยาวกว่าการแจกแจงแบบปกติและมีการแจกแจงแบบเบ้ ในกรณีที่การแจกแจงของความผิดพลาดเป็นแบบหางยาวกว่าการแจกแจงแบบปกติจะศึกษาวิธีการประมาณ 2 วิธีคือ วิธีกำลังสองน้อยที่สุดและวิธี M-estimator ซึ่งใช้เกณฑ์ความแกร่งของ Ramsay โดยจะใช้การแจกแจงแบบปกติปลอมปนและแบบทีในกรณีนี้ ส่วนกรณีที่การแจกแจงความผิดพลาดเป็นแบบเบ้ต้องอาศัยการแปลงที่ใช้การยกกำลังของ Box และ Cox เพื่อศึกษาการประมาณสัมประสิทธิ์การถดถอยพหุทั้งสองวิธี โดยจะใช้การแจกแจงแบบลอกนอร์มอล แกมมา และไวบูลล์ ซึ่งจะพิจารณาประสิทธิภาพการประมาณค่าสัมประสิทธิ์การถดถอยของทั้งสองวิธี โดยใช้เกณฑ์การเปรียบเทียบค่าเฉลี่ยของค่าสัมพัทธ์ของค่าเฉลี่ยความผิดพลาดกำลังสอง (ARMSE) และใช้เกณฑ์การเปรียบเทียบค่าเฉลี่ยของค่าสัมบูรณ์ของค่าแตกต่างของอัตราส่วนค่าเฉลี่ยความผิดพลาดกำลังสอง(AADRM)ด้วย

3.1 วิธีมอนติคาร์โล (Monte Carlo method)

เทคนิคในการจำลองตัวแบบทางคณิตศาสตร์มีอยู่หลายวิธี วิธีมอนติคาร์โลเป็นวิธีหนึ่งที่นิยมใช้กันอย่างแพร่หลายในปัจจุบัน ซึ่งหลักการของวิธีมอนติคาร์โลนั้นเป็นการจำลองตัวเลขสุ่ม (random number) มาช่วยในการหาคำตอบของปัญหาที่ต้องการศึกษา

ในการวิจัยครั้งนี้จะใช้วิธีการมอนติคาร์โลดังกล่าวในการสร้างข้อมูลที่มีสภาพการแจก-แจงตามที่ต้องการ ซึ่งขั้นตอนของวิธีมอนติคาร์โลที่ใช้กันอยู่ในปัจจุบันแบ่งเป็น 3 ขั้นตอนดังนี้

3.1.1 การสร้างตัวเลขสุ่ม การใช้ตัวเลขสุ่มเป็นสิ่งสำคัญมากในวิธีมอนติคาร์โล เพราะว่าหลักการของวิธีมอนติคาร์โลนั้นจะใช้ตัวเลขมาช่วยในการหาคำตอบของปัญหา ซึ่งลักษณะของตัวเลขสุ่มจะมีการแจกแจงแบบสม่ำเสมอ (Uniform distribution)ในช่วง (0, 1) สำหรับวิธีการสร้างตัวเลขสุ่มนั้นมีผู้เสนอไว้หลายวิธี แต่วิธีหนึ่งที่ดีก็คือ วิธีที่ไวท์และชมิดท์ (White and Schemidt 1975 : 421) เสนอไว้กล่าวคือ ลักษณะของตัวเลขสุ่มที่เกิดขึ้นมีการแจกแจงแบบสม่ำเสมอในช่วง (0, 1) และเป็นอิสระซึ่งกันและกัน ซึ่งรายละเอียดแสดงไว้ในภาคผนวก ก

3.1.2 การประยุกต์ปัญหาที่ต้องการศึกษาเพื่อใช้กับตัวเลขสุ่ม ซึ่งขั้นตอนนี้ขึ้นอยู่กับลักษณะของปัญหาที่ต้องการศึกษา บางปัญหาอาจจะไม่ใช้ตัวเลขสุ่มโดยตรง แต่อาจมีขั้นตอนอื่นอีกหลาย ๆ ขั้นตอน ซึ่งบางขั้นตอนต้องใช้ตัวเลขสุ่ม

3.1.3 การทดลองกระทำซ้ำ(replication) เมื่อประยุกต์ปัญหาให้ใช้กับตัวเลขสุ่มได้แล้ว ขั้นตอนต่อไปก็คือ การทดลองโดยใช้กระบวนการของการสุ่ม (random process) มากระทำในลักษณะที่ซ้ำ ๆ กัน เพื่อหาคำตอบของปัญหาที่ต้องการศึกษา

3.2 แผนการทดลอง

การวิจัยในครั้งนี้กำหนดสถานการณ์ต่าง ๆ ที่ต้องการศึกษาโดยสุ่มตัวอย่างจากประชากรที่มีการแจกแจงแบบเดียวกัน ซึ่งลักษณะการแจกแจงแบบหางยาวกว่าการแจกแจงแบบปกติที่ต้องการศึกษาคือ แบบปกติปลอมปน และแบบที สำหรับลักษณะการแจกแจงแบบเบ้ที่ต้องการศึกษา คือ ลอกนอร์มอล แกมมา และไวบูลล์

สำหรับการแจกแจงแบบปกติปลอมปนจะกำหนดเปอร์เซนต์การปลอมปนเป็น 1%, 5%, 10% และ 25% ส่วนสเกลแฟคเตอร์มี 2 ระดับ คือ 3 และ 10 ถ้าสเกลแฟคเตอร์สูงจะทำให้เกิดค่าผิดปกติสูง ซึ่งได้แสดงรายละเอียดในตารางที่ 3.1

ตารางที่ 3.1 แสดงค่าสเกลแฟคเตอร์และเปอร์เซนต์การปลอมปนทั้งหมดที่ใช้ในการวิจัยครั้งนี้

c \ p	1	5	10	25
3	(3,1)	(3,5)	(3,10)	(3,25)
10	(10,1)	(10,5)	(10,10)	(10,25)

สำหรับการแจกแจงแบบทีจะสนใจศึกษาเมื่อระดับความเป็นอิสระ = 4, 8 โดยเมื่อระดับความเป็นอิสระน้อยจะทำให้เกิดหางที่หนาและยาวกว่าแบบปกติมากขึ้น

สำหรับการแจกแจงแบบลอกนอร์มอล สนใจศึกษาเมื่อค่า $\mu = 0$, $\sigma^2 = 1$ ซึ่งเป็นค่าเฉลี่ยและค่าความแปรปรวนของการแจกแจงแบบปกติ

สำหรับการแจกแจงแบบแกมมาจะสนใจศึกษาเมื่อ $\beta = 1$, $\alpha = 1, 2, 3$ และ $\beta = 150$, $\alpha = 10$ โดยมีสัมประสิทธิ์ความแปรปรวนเป็น 100%, 70%, 58% และ 32% ตามลำดับ

สำหรับการแจกแจงแบบไวบูลล์จะสนใจศึกษาเมื่อ $\beta = 1$, $\alpha = 1, 2, 3$ และ $\beta = 150$, $\alpha = 10$ โดยมีสัมประสิทธิ์ความแปรปรวนเป็น 100%, 52%, 35% และ 17% ตามลำดับ

ทุก ๆ การแจกแจงดังกล่าว เมื่อใช้จำนวนตัวแปรอิสระที่สนใจ = 3 จะใช้ขนาดตัวอย่าง = 20 สำหรับกรณีจำนวนตัวแปรอิสระ = 5, 10 จะใช้ขนาดตัวอย่าง 50, 100 และ 150 ตามลำดับ

ตารางที่ 3.2 แสดงขนาดตัวอย่างและจำนวนตัวแปรอิสระทั้งหมดที่ใช้ในการวิจัยครั้งนี้

ขนาดตัวอย่าง \ จำนวนตัวแปรอิสระ	3	5	10
20	(20,3)	-	-
50	-	(50,5)	(50,10)
100	-	(100,5)	(100,10)
150	-	(150,5)	(150,10)

ศูนย์วิทยทรัพยากร
จุฬาลงกรณ์มหาวิทยาลัย

3.3 ขั้นตอนในการวิจัย

ขั้นตอนในการวิจัยมีดังนี้คือ

1. การสร้างโปรแกรมย่อย (subroutine) สำหรับการแจกแจงของค่าผิดพลาดตามที่ต้องการศึกษา

2. การสร้างข้อมูล คือ ตัวแปรตาม (y) และตัวแปรอิสระ (X) เมื่อการแจกแจงของค่าผิดพลาดเป็นแบบหางยาวกว่าการแจกแจงแบบปกติ ตัวแปรตาม (y) สร้างจากตัวแปรอิสระ (X) ซึ่งเป็นค่าคงที่ และค่าผิดพลาดที่มีลักษณะการแจกแจงที่ต้องการศึกษา โดยให้ตัวแปรตามมีความสัมพันธ์เชิงเส้นตรงกับตัวแปรอิสระ กรณีที่การแจกแจงของค่าผิดพลาดเป็นแบบเบ้ ตัวแปรตาม (y) สร้างให้มีลักษณะการแจกแจงตามที่ต้องการศึกษา โดยไม่เป็นฟังก์ชันเชิงเส้นตรงกับตัวแปรอิสระ(X) และค่าผิดพลาด

3. การทดสอบความเหมาะสมของตัวแปรอิสระ (X) โดยทดสอบความเป็นอิสระซึ่งกันและกันหรือไม่เกิด multicollinearity

4. การประมาณค่าสัมประสิทธิ์การถดถอยพหุเมื่อการแจกแจงของค่าผิดพลาดเป็นแบบหางยาวกว่าการแจกแจงแบบปกติ โดยวิธีกำลังสองน้อยที่สุด และวิธี M-estimator ซึ่งใช้เกณฑ์ความแกร่งของ Ramsay

5. การประมาณค่าสัมประสิทธิ์การถดถอยพหุเมื่อการแจกแจงของค่าผิดพลาดเป็นแบบเบ้จะอาศัยการแปลงที่ใช้การยกกำลังของ Box และ Cox แล้วประมาณค่าสัมประสิทธิ์การถดถอยพหุโดยวิธีกำลังสองน้อยที่สุด และวิธี M-estimator ซึ่งใช้เกณฑ์ความแกร่งของ Ramsay

6. การหาค่าเฉลี่ยของค่าสัมพัทธ์ของค่าเฉลี่ยความผิดพลาดกำลังสอง และค่าเฉลี่ยของค่าสัมบูรณ์ของค่าแตกต่างของอัตราส่วนความผิดพลาดกำลังสอง ระหว่างวิธีกำลังสองน้อยที่สุดกับวิธี M-estimator ซึ่งใช้เกณฑ์ความแกร่งของ Ramsay

สำหรับรายละเอียดของแต่ละขั้นตอนเป็นดังนี้

3.3.1 การสร้างโปรแกรมย่อยสำหรับสร้างการแจกแจงของค่าผิดพลาดตามที่ต้องการศึกษา

การสร้างลักษณะการแจกแจงของค่าผิดพลาดทุกรูปแบบที่ต้องการศึกษานั้น ใช้โปรแกรมภาษาฟอร์แทรนโฟร์ (FORTRAN IV) โดยใช้กับเครื่อง IBM 370/3031 ซึ่งการสร้างลักษณะการแจกแจงแบบต่าง ๆ นั้น จะต้องใช้ตัวเลขสุ่มเป็นพื้นฐานในการสร้าง สำหรับรายละเอียดในการสร้างการแจกแจงแบบต่าง ๆ เป็นดังนี้

I16486160

3.3.1.1 การแจกแจงแบบปกติ ฟังก์ชันที่ใช้ในการแจกแจงแบบปกติจะใช้วิธีของ Box และ Muller (1958) ซึ่งเป็นวิธีที่สร้างการแจกแจงแบบปกติที่มีค่าเฉลี่ยเป็น μ และความแปรปรวนเป็น σ^2 โดย DMEAN และ $(SIGMA)^2$ คือ ค่าเฉลี่ยและความแปรปรวนที่กำหนด สำหรับรายละเอียดแสดงไว้ในภาคผนวก ก การใช้ฟังก์ชันนี้ใช้คำสั่ง NORMAL (DMEAN, SIGMA) ผลลัพธ์คือ NORMAL ที่มีค่าเฉลี่ยและความแปรปรวนที่กำหนด

3.3.1.2 การแจกแจงแบบปกติปลอมปน สร้างโดยการใช้วิธีการแปลงข้อมูลมาจากการแจกแจงแบบปกติ ผลลัพธ์คือ NORMAL โดยได้มาจาก NORMAL (DMEAN, SIGMA) ด้วยความน่าจะเป็น (1-P) เมื่อ DMEAN SIGMA และ P เป็นค่าเฉลี่ย ความแปรปรวน และสัดส่วนการปลอมปน และได้มาจาก NORMAL (DMEAN, SG2) ด้วยความน่าจะเป็น P เมื่อ SG2 เป็นผลคูณของ IC กับ SIGMA IC เป็นสเกลแฟคเตอร์สำหรับความแปรปรวน สำหรับรายละเอียดแสดงไว้ในภาคผนวก ก

3.3.1.3 การแจกแจงแบบที ฟังก์ชันที่ใช้คือ TDIST (NDF, DMEAN,SIGMA) โดยที่ผลลัพธ์คือค่าของ TDIST มีระดับความเป็นอิสระ เป็น NDF ซึ่งสร้างจากสัดส่วนของการแจกแจงแบบปกติ และการแจกแจงแบบไคสแควร์ มีระดับความเป็นอิสระเป็น NDF ค่าเฉลี่ยและความแปรปรวนของการแจกแจงแบบปกติเป็น DMEAN และ SIGMA สำหรับรายละเอียดแสดงไว้ในภาคผนวก ก.

3.3.1.4 การแจกแจงแบบลอกนอร์มอล หาได้จากการหา Exponential ของฟังก์ชัน NORMAL (DMEAN, SIGMA) เมื่อ DMEAN และ SIGMA เป็นค่าเฉลี่ยและความแปรปรวนของการแจกแจงแบบปกติ สำหรับรายละเอียดแสดงไว้ในภาคผนวก ก.

3.3.1.5 การแจกแจงแบบแกมมา ฟังก์ชันที่ใช้คือ GAMMA1 (ALPHA1, BETA1) ฟังก์ชันนี้สร้างโดยวิธี Reproductive property ซึ่งรายละเอียดแสดงไว้ในภาคผนวก ก. ฟังก์ชัน GAMMA1 จะกำหนดค่า ALPHA1 และ BETA1 เพื่อให้ได้ค่าเฉลี่ยและความแปรปรวนตามที่ต้องการ ผลลัพธ์คือค่า GAMMA1 จะมีการแจกแจงแบบแกมมามีค่าเฉลี่ยเป็น BETA1 * ALPHA1 และความแปรปรวนเป็น $(BETA)^2$*ALPHA1

3.3.1.6 การแจกแจงแบบไวบูลล์ ฟังก์ชันที่ใช้คือ WEIBUL (ALPHA1, BETA1) ฟังก์ชันนี้สร้างโดยวิธี Inverse transformation ซึ่งรายละเอียดแสดงไว้ในภาคผนวก ก. ฟังก์ชัน WEIBul จะกำหนดค่า ALPHA1, BETA1 เพื่อให้ได้ค่าเฉลี่ย และความแปรปรวนตามที่ต้องการ ผลลัพธ์คือ ค่า WEIBUL จะมีการแจกแจงแบบไวบูลล์มีค่าเฉลี่ยเป็น BETA1*(1+1/ALPHA1) และความแปรปรวนเป็น $(BETA1)^2*[\Gamma(1+2/ALPHA1) - \Gamma^2(1+1/ALPHA1)$

3.3.2 การสร้างข้อมูลให้มีความสัมพันธ์เชิงเส้นตรง

3.3.2.1 เมื่อการแจกแจงของค่าผิดพลาดเป็นแบบหางยาวกว่าการแจกแจงแบบปกติ

ในการวิจัยครั้งนี้จะสร้างตัวแปรอิสระ X ซึ่งเป็นค่าคงที่ก่อน แล้วจึงสร้างตัวแปร $\underset{\sim}{y}$ ที่มีความสัมพันธ์เชิงเส้นกับ X ให้มีลักษณะการแจกแจงตามความผิดพลาดที่กำหนด ตามรูปแบบของความสัมพันธ์คือ $\underset{\sim}{y} = X\underset{\sim}{\beta} + \underset{\sim}{\varepsilon}$ เมื่อ $\underset{\sim}{\beta}$ เป็นค่าพารามิเตอร์ที่ถูกกำหนดขึ้นมา และ $\underset{\sim}{\varepsilon}$ เป็นความผิดพลาดที่มีการแจกแจงเป็นแบบต่าง ๆ ดังที่กล่าวแล้วใน 3.3.1.2 ถึง 3.3.1.3 สำหรับการสร้าง $\underset{\sim}{y}$ นั้น จะเริ่มจากการกำหนดจำนวนตัวแปรอิสระและขนาดตัวอย่างที่ต้องการศึกษา พารามิเตอร์ $\underset{\sim}{\beta}$, ค่าเฉลี่ย ค่าความแปรปรวน และลักษณะการแจกแจงของค่าผิดพลาด รวมทั้ง ค่าคงที่ X แล้วจึงสร้าง $\underset{\sim}{y}$ ตามรูปแบบดังกล่าว

3.3.2.2 เมื่อการแจกแจงของค่าผิดพลาดเป็นแบบเบ้

ในการวิจัยครั้งนี้จะสร้างตัวแปรอิสระ X ซึ่งมีค่าคงที่ก่อน แล้วจึงสร้างตัวแปร $\underset{\sim}{y}$ ให้มีการแจกแจงเป็นแบบเบ้ ตาม 3.3.1.4 ถึง 3.3.1.6 แล้วอาศัยการแปลงที่ใช้การยกกำลังของ Box และ Cox เพื่อสร้างตัวแปร $\underset{\sim}{y}^{(\lambda)}$ ที่มีความสัมพันธ์เชิงเส้นกับ X ตามรูปแบบของความสัมพันธ์คือ $\underset{\sim}{y}^{(\lambda)} = X\underset{\sim}{\beta} + \underset{\sim}{\varepsilon}$ เมื่อค่า $\underset{\sim}{\beta}$ เป็นค่าพารามิเตอร์ที่ถูกกำหนดขึ้นมา λ เป็นค่าพารามิเตอร์ของการแปลงที่ใช้การยกกำลังของ Box และ Cox $\underset{\sim}{\varepsilon}$ เป็นค่าผิดพลาดที่มีการแจกแจงเป็นแบบสมมาตร, เชิงเส้น และเป็นแบบปกติโดยประมาณ

3.3.3 การทดสอบความเหมาะสมของตัวแปรอิสระ (X) โดยการทดสอบความเป็นอิสระซึ่งกันและกัน หรือไม่เกิด multicollinearity

เมื่อสร้างตัวแปรอิสระ (X) ให้มีค่าคงที่แล้วจะทำการทดสอบความเป็นอิสระซึ่งกันและกัน เนื่องจากถ้าตัวแปรอิสระเกิด multicollinearity จะทำให้เกิดการประมาณสัมประสิทธิ์การถดถอยพหุด้วยวิธีกำลังสองน้อยที่สุด และวิธี M-estimator ถูกลดประสิทธิภาพ (Montgomery (ค.ศ.1982 : 364-381), Mason และ Gunst (ค.ศ.1985 : 401-407)) ผลของการที่ตัวแปรอิสระมีความสัมพันธ์กัน หรือเกิด multicollinearity จะทำให้เมตริกซ์ X เกือบจะเป็น singular matrix นั่นคือ determinant มีค่าน้อย และ inverse มีค่าสูง ซึ่งจะทำให้ความแปรปรวนของสมการการถดถอยสูง และมีผลทำให้ยอมรับ H0 บ่อยครั้งขึ้น ทำให้การสรุปความสัมพันธ์ของ $\underset{\sim}{y}$ กับ X ผิดไป

การทดสอบว่าเกิด multicollinearity หรือไม่ โดยการทดสอบสมมติฐานดังนี้

<u>สมมติฐานของการทดสอบ</u> H_0 : ไม่เกิด multicollinearity

เทียบกับ H_1 : เกิด multicollinearity

<u>ณ. ระดับนัยสำคัญ</u> α = 0.05 และ 0.01

<u>ค่าสถิติทดสอบ</u> $$F^{(m-2)}_{(n-m+1)} = \frac{R_i^2/(m-2)}{(1-R_i^2)/(n-m+1)}$$ (ไม่รวม intercept)

โดย R_i^2 เป็นของ Regression X_i on $X(j\neq i)$ $i,j = 1,\ldots,m$

<u>ขอบเขตวิกฤต</u> F จากตาราง ที่ระดับนัยสำคัญ 0.05 และ 0.01

<u>การตัดสินใจ</u> จะปฏิเสธ H_0 เมื่อ F (คำนวณ) >F (ตาราง)

3.3.4 การประมาณค่าสัมประสิทธิ์การถดถอยพหุเมื่อการแจกแจงของค่าผิดพลาดเป็นแบบหางยาวกว่าการแจกแจงแบบปกติ

ก) การประมาณค่าสัมประสิทธิ์การถดถอยพหุโดยวิธีกำลังสองน้อยที่สุด

ก.1) โปรแกรมย่อยที่ใช้ในการประมาณค่าสัมประสิทธิ์การถดถอยพหุจากข้อมูลจะใช้โปรแกรมย่อย OLS(B) โดย B เป็นเวกเตอร์ของสัมประสิทธิ์การถดถอยพหุ ซึ่งเมตริกซ์ของ X จะใช้ร่วมกับโปรแกรมหลัก เมตริกซ์ X จะประกอบด้วย ตัวแปรอิสระ และตัวแปรตาม ขนาด M2 X N, M2 คือจำนวนตัวแปรอิสระ M ตัวและตัวแปรตาม 1 ตัว N คือจำนวนข้อมูล โปรแกรมย่อย OLS(B) มีขั้นตอนการทำงานดังนี้

ขั้นตอน	ความหมาย
1	คำนวณค่า $(X'X)$ และ $(X'y)$ ในเมตริกซ์ของ X ประกอบด้วยตัวแปรอิสระ และตัวแปรตาม โดยเพิ่มเวกเตอร์ y ในแนวตั้งที่ M2 (M2 = M+1)
2	คำนวณค่า $(X'X)^{-1}$ เรียกใช้โปรแกรมย่อย INVS (M,A)
3	คำนวณ $\hat{\beta}$ โดยใช้ $B = (X'X)^{-1}(X'y)$

ก.2) โปรแกรมย่อย สำหรับหาเมตริกซ์ผกผัน (inverse matrix) คือโปรแกรมย่อย INVS (M,A) เมื่อ M เป็นขนาดของเมตริกซ์ผกผันที่ต้องการหา A เป็นเมตริกซ์ผกผันที่ต้องการ

ก.3) โปรแกรมย่อยสำหรับประมาณค่าผิดพลาด $\hat{\varepsilon}_i$ เท่ากับ $y_i - \hat{y}_i$ คือโปรแกรมย่อย YRESID (X,N,M,YHAT,YRES,B) โดยที่ X เป็นเมตริกซ์ของตัวแปรอิสระและตัวแปรตาม N คือจำนวนข้อมูล M คือจำนวนตัวแปรอิสระ YHAT คือค่าประมาณ $\hat{y}_i$ YRES คือค่าผิดพลาด ε_i และ B คือสัมประสิทธิ์การถดถอยพหุ

ข) การประมาณค่าสัมประสิทธิ์การถดถอยพหุ โดยวิธี M-estimator ซึ่งใช้เกณฑ์ความแกร่งของ Ramsay

การประมาณค่าสัมประสิทธิ์การถดถอยพหุ โดยวิธี M-estimator เกณฑ์ความแกร่งของ Ramsay มีค่า a = 0.3 เรียกใช้โปรแกรมย่อย MEST (B1, B) เมื่อ B1 เป็นค่าเริ่มต้น ซึ่งมีขั้นตอนการทำงานดังนี้

ขั้นตอน	ความหมาย
1	เริ่มต้นเมื่อ $k = 0, \hat{\beta}_k$ คือ B1 เป็นค่าประมาณสัมประสิทธิ์การถดถอยพหุซึ่งคำนวณจากวิธีกำลังสองน้อยที่สุด
2	คำนวณ $\hat{\varepsilon}_i(\beta_k) = y_i - X_i'\beta_k$, $i = 1,2,\ldots,n$ โดย $\hat{\varepsilon}_i(\beta_k)$ คือค่า YRES โดยเรียกใช้โปรแกรมย่อย YRESID (X,N,M,YHAT,YRES,B) ซึ่งแทน B ด้วย $\hat{\beta}_k$ และใช้ E แทน YRES
3	คำนวณ ค่า s จาก median $\lvert \varepsilon_i$ - median $(\varepsilon_i)\rvert/0.6745$ โดย SIGMA คือ s ซึ่งจะเรียกใช้โปรแกรมย่อย RANK (N,E) สำหรับเรียงค่าจากน้อยไปมาก เพื่อจะหาค่า MEDIAN และ S_2 ซึ่งเป็น absolute ของ E กับ MEDIAN
4	คำนวณค่า $(Wk)_i = \exp(-a\lvert\varepsilon_i/s\rvert)$ โดยใช้ W แทน $(Wk)_i$
5	หาค่า $\hat{\beta}_{k+1} = (X'WkX)^{-1}(X'Wky)$ เมื่อ Wk เป็นเมตริกซ์ทแยงมุมของค่าถ่วงน้ำหนักจากข้อ 4 ใช้ B แทน β_{k+1} และเรียกใช้โปรแกรมย่อย OLS1(B) ซึ่งจะคำนวณค่า B โดยมีการถ่วงน้ำหนัก W
6	ถ้า $\lvert\hat{\beta}_{k+1} - \hat{\beta}_k\rvert/\lvert\hat{\beta}_k\rvert \leq 0.001$ ทุกค่าพารามิเตอร์ M ตัวแล้ว $\hat{\beta}_{k+1}$ จะเป็นตัวประมาณที่ต้องการ โดยใช้ B แทน β_{k+1} และ B1 แทน β_k ถือว่าจบกระบวนการ ถ้า $\lvert\hat{\beta}_{k+1} - \hat{\beta}_k\rvert/\lvert\hat{\beta}_k\rvert > 0.001$ ในแต่ละค่าของพารามิเตอร์ M ตัวแล้ว
7	ตรวจสอบว่า จำนวนครั้ง k ถ้า $k \geq 20$ แล้วจะหยุดการทำงาน และ $\hat{\beta}_{k+1}$ เป็นตัวประมาณที่ต้องการ ถ้า $k < 20$ แล้วจะหารอบ(iteration) ถัดไปโดย $k = k+1$ แล้วกลับไปขั้นที่ 4

3.3.5 การประมาณสัมประสิทธิ์การถดถอยพหุ เมื่อการแจกแจงของค่าผิดพลาดเป็นแบบเบ้

ก) การประมาณสัมประสิทธิ์การถดถอยพหุโดยวิธีกำลังสองน้อยที่สุดจากการแปลงที่ใช้การยกกำลังของ Box และ Cox โปรแกรมย่อยที่ใช้คือ BOXCOX (B, FL1) ซึ่ง B เป็นค่าสัมประสิทธิ์ของการถดถอยพหุ FL1 คือ $\hat{\lambda}$ ซึ่งเป็นค่าประมาณ λ ของกำลังการแปลงของ Box และ Cox โดยมีขั้นตอนการทำงานดังนี้

ขั้นตอน	ความหมาย
1	คำนวณหาค่าเฉลี่ยเรขาคณิต (geometric mean) : G ของเวกเตอร์ $\underset{\sim}{y}$ โดย y_i จะต้องไม่เท่ากับ 0 หรือค่าเลขลบ แล้วคำนวณค่า $\underset{\sim}{y}1 = \underset{\sim}{y} / G$
2	กำหนดค่าเริ่มต้น ST ค่าสุดท้าย FIN เป็นขอบเขตเริ่มต้นในการหาค่า $\hat{\lambda}$ หรือ FL_1 โดยใช้เทคนิคของการหา $\hat{\lambda}$ ที่มีค่าผลบวกกำลังสองของค่าผิดพลาด หรือที่มีค่า $\hat{\sigma}^2_{\varepsilon\|\hat{\lambda}}$ ต่ำสุด รายละเอียดแสดงในภาคผนวก ข. ซึ่ง SSE2 แทน $\hat{\sigma}^2_{\varepsilon\|\hat{\lambda}}$ ที่มีค่าต่ำสุด จะได้ค่าประมาณ FL1 และ B

ข) การประมาณสัมประสิทธิ์การถดถอยพหุ โดยวิธี M-estimator จากเทคนิคการแปลงที่ใช้การยกกำลังของ Box และ Cox โปรแกรมย่อยที่เรียกใช้คือ BOXC (B,FL1) ซึ่ง B เป็นค่าสัมประสิทธิ์การถดถอยพหุ FL1 คือ $\hat{\lambda}$ ซึ่งเป็นค่าประมาณ λ ของการแปลงที่ใช้การยกกำลังของ Box และ Cox สำหรับโปรแกรมย่อย BOXC(B,FL1) นี้จะใช้คำตอบของโปรแกรม BOXCOX(B,FL1) ในการประมาณสัมประสิทธิ์การถดถอยพหุ และ λ ซึ่งเป็นค่าประมาณของการแปลง

ค) โปรแกรมย่อยสำหรับคำนวณค่าผลบวกกำลังสองของค่าผิดพลาด

โปรแกรมย่อยที่ใช้ SUMSQ(FLX,SSE,K,L) โดย FLX เป็นค่า λ ที่ต้องการใช้ในการแปลงข้อมูล ให้อยู่ในรูปตัวแบบเชิงเส้น $\underset{\sim}{y}^{(\lambda)} = X\underset{\sim}{\beta} + \underset{\sim}{\varepsilon}$ SSE คือ $\sigma^2_{\underset{\sim}{\varepsilon}\|\hat{\lambda}}$ ที่คำนวณได้จากตัวแบบเชิงเส้นนี้ K เป็นค่าที่ใช้สำหรับบอกตำแหน่งของ SSE ว่าเป็นของค่าจุดเริ่มต้น ค่ากลาง หรือค่าสุดท้าย ในโปรแกรมที่เรียกใช้คือ BOXCOX และ BOXC L เป็นค่าสำหรับเลือกวิธีการประมาณสัมประสิทธิ์การถดถอยพหุ โดย L = 1 ใช้วิธี OLS(B), L = 2 ใช้วิธี M-estimator MEST(B1,B) โดยไม่คำนวณค่า B1 และ L = 3 ใช้วิธี M-estimator MEST(B1,B) โดยคำนวณค่า B1 ด้วย

ง) โปรแกรมย่อยสำหรับการแปลงข้อมูล โดยเทคนิคการแปลงที่ใช้การยกกำลังของ Box และ Cox โปรแกรมย่อยที่ใช้คือ BCOX(FL1) ซึ่ง FL1 คือค่า λ สำหรับการแปลงข้อมูลหรือ y1 ให้อยู่ในรูปของตัวแบบเชิงเส้น $\underset{\sim}{y}^{(\lambda)} = X\underset{\sim}{\beta} + \underset{\sim}{\varepsilon}$ โดย y1 ใช้ข้อมูลร่วมกับโปรแกรมที่เรียกใช้ คือ SUMSQ และผลลัพธ์ที่จะถูกแทนในเมตริกซ์ X ในแนวตั้ง ที่ M2 เพื่อใช้ในการประมาณสัมประสิทธิ์การถดถอยพหุต่อไป

3.3.6 การหาค่าเฉลี่ยของค่าสัมพัทธ์ของค่าเฉลี่ยความผิดพลาดกำลังสอง และค่าเฉลี่ยของค่าแตกต่างของอัตราส่วนความผิดพลาดกำลังสองระหว่างวิธีกำลังสองน้อยที่สุดกับ วิธี M-estimatorซึ่งใช้เกณฑ์ความแกร่งของ Ramsay

ก) การหาค่าเฉลี่ยความผิดพลาดกำลังสอง (Mean Square Error(MSE)) สำหรับการประมาณสัมประสิทธิ์การถดถอยพหุ โดยวิธีกำลังสองน้อยที่สุด และวิธี M-estimator ใช้วิธีหาผลรวมของ MSE ของ $\underset{\sim}{\hat{\beta}}$ ที่ได้จากการประมาณกับ $\underset{\sim}{\beta}$ ของประชากร ทุกค่าของ β_i, $i = 1,\ldots,m$ แล้วหารด้วย m

$$\text{MSE ของวิธีกำลังสองน้อยที่สุด} = \frac{\sum_{i=1}^{m} (\hat{\beta}_{i(OLS)} - \beta_i)^2}{m}$$

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

$$\text{MSE ของวิธี M-estimator} = \frac{\sum_{i=1}^{m} (\hat{\beta}_{i(m)} - \beta_i)^2}{m}$$

เมื่อ $\hat{\beta}_{i(m)}$ เป็น $\hat{\beta}_i$ ของวิธี M-estimator

ข) การหาค่าเฉลี่ยของค่าสัมพัทธ์ของค่าเฉลี่ยความผิดพลาดกำลังสอง (Average of Relative Mean Square Error(ARMSE)) ระหว่างวิธีกำลังสองน้อยที่สุดกับ วิธี M-estimator ซึ่งใช้เกณฑ์ความแกร่งของ Ramsay เขียนได้ดังนี้

$$\text{ARMSE} = \begin{cases} \dfrac{MSE_M}{MSE_{OLS}} \times 100, & MSE_{OLS} < MSE_M \\ \dfrac{MSE_{OLS}}{MSE_M} \times 100, & MSE_M < MSE_{OLS} \end{cases}$$

โดยที่ MSE_{OLS} แทน MSE ของวิธีกำลังสองน้อยที่สุด

MSE_M แทน MSE ของวิธี M-estimator

ค) การหาค่าเฉลี่ยของค่าสัมบูรณ์ของค่าแตกต่างของค่าเฉลี่ยความผิดพลาดกำลังสอง (Average of Absolute value of Different Ratio of Mean Square Error(AADRM)) ระหว่างวิธีกำลังสองน้อยที่สุดกับ วิธี M-estimator ซึ่งใช้เกณฑ์ความแกร่งของ Ramsay

$$AADRM = \begin{cases} \dfrac{|MSE_{OLS} - MSE_M|}{MSE_M} & , \quad MSE_{OLS} < MSE_M \\ \dfrac{|MSE_M - MSE_{OLS}|}{MSE_{OLS}} & , \quad MSE_M < MSE_{OLS} \end{cases}$$

โดยที่ MSE_{OLS} แทน MSE ของวิธีกำลังสองน้อยที่สุด

MSE_M แทน MSE ของวิธี M-estimator

ศูนย์วิทยทรัพยากร
จุฬาลงกรณ์มหาวิทยาลัย

ตารางที่ 3.3 แสดงลักษณะการทำงานของโปรแกรมทั้งหมดที่ใช้ในการวิจัย

อันดับที่	ชื่อโปรแกรม	คุณสมบัติของโปรแกรม	ชื่อโปรแกรมที่เรียกใช้
โปรแกรมหลัก 4	MAIN 1	- คำนวณค่าประมาณของพารามิเตอร์ $\hat{\beta}$ จากวิธีการประมาณ 2 วิธี เมื่อความผิดพลาดมีการแจกแจงแบบหางยาวกว่าการแจกแจงแบบปกติ 1. แบบกำลังสองน้อยที่สุด 2. แบบ M-estimator	INIT, DATA OLS และ MEST
โปรแกรมหลัก 2	MAIN 2	- คำนวณค่าประมาณของพารามิเตอร์ $\hat{\beta}$ จากวิธีการประมาณ 2 วิธี เมื่อความผิดพลาดมีการแจกแจงแบบเบ้ 1. Box และ Cox 2. Box และ Cox แบบแกร่ง	INIT, DATA BOXCOX และ BOXC
SUBROUTINE และ FUNCTION		- คำนวณค่า MSEOLS , MSEM, ARMSE และ AADRM - และคำนวณค่าเฉลี่ยและ SD ของ ARMSE และ AADRM	
1	RAND	สร้างเลขสุ่มที่มีการแจกแจงแบบสม่ำเสมอ	-
2	NORMAL	สร้างเลขสุ่มที่มีการแจกแจงแบบปกติ	RAND
3	TDIST	สร้างเลขสุ่มที่มีการแจกแจงแบบที	RAND,NORMAL
4	GAMMA1	สร้างเลขสุ่มที่มีการแจกแจงแบบแกมมา	RAND
5	WEIBUL	สร้างเลขสุ่มที่มีการแจกแจงแบบไวบูลล์	RAND
6	OLS	คำนวณค่าพารามิเตอร์ $\hat{\beta}$ โดยวิธีกำลังสองน้อยที่สุด	INVS
7	OLS1	คำนวณค่าพารามิเตอร์ $\hat{\beta}$ โดยวิธีกำลังสองน้อยที่สุดแบบถ่วงน้ำหนัก	INVS
8	MEST	คำนวณค่าพารามิเตอร์ $\hat{\beta}$ โดยวิธี M-estimator	INVS,RANK
9	YRESID	คำนวณค่า residual	OLS,OLS1,MEST
10	INVS	คำนวณค่าของ Inverse matrix แบบ swapping	-
11	INIT	กำหนดค่าเริ่มต้นของตัวแปรและอ่านค่าตัวแปรและสร้างเมตริกซ์ของตัวแปรอิสระ X	NORMAL
12	DATA	สร้างเวกเตอร์ของตัวแปรตาม $\underline{Y}$ ให้มีการแจกแจงแบบ 1. ปกติปลอมปน	RAND,NORMAL
		2. ที	TDIST
		3. ลอกนอร์มอล	LOGNOR
		4. แกมมา	GAMMA1
		5. ไวบูลล์	WEIBUL
13	BOXCOX	คำนวณค่าตัวพารามิเตอร์ของ BOX&COX โดยใช้วิธีกำลังสองน้อยที่สุด	SUMSQ
14	BOXC	คำนวณค่าตัวพารามิเตอร์ของ BOX&COX โดยใช้วิธีแกร่งแบบ M	SUMSQ
15	BCOX	แปลงค่า Yi ในแต่ละค่าของ λ	-
16	SUMSQ	คำนวณหาค่า $\hat{\beta}$ และ SSE ซึ่งเลือกหาได้ทั้งวิธี OLS และ M	BCOX,OLS, MEST
17	RANK	เรียงอันดับค่าของข้อมูลจากน้อยไปมาก	-
18	SHAPWK	คำนวณค่าพารามิเตอร์ W ของ Shapiro-Wilk ในการทดสอบว่าข้อมูลมีการแจกแจงเป็นแบบปกติ	-

โปรแกรมทั้งหมดที่ใช้ในการวิจัยแสดงไว้ในภาคผนวก ค

แสดงผังงานสำหรับโปรแกรมหลักที่ 1 MAIN PROGRAM 1

CALL SUBROUTINE		งานที่ทำ
(1) SUBROUTINE INIT, DATA	สร้างค่า Y, X	(1) สร้างตัวแปรสุ่มที่มีการแจกแจงแบบ - ปกติแบบปลอมปน - ที
(2) SUBROUTINE OLS	หาค่า $\hat{\beta}$ OLS โดยวิธี OLS	(2) คำนวณค่าประมาณ $\hat{\beta}$ สัมประสิทธิ์การถดถอย โดยวิธีกำลังสองน้อยที่สุด
(3) SUBROUTINE MEST	หาค่า $\hat{\beta}$ M โดยวิธี M-estimator	(3) คำนวณค่าประมาณ $\hat{\beta}$ สัมประสิทธิ์การถดถอย โดยวิธี M-estimator
	หาค่า MSE_{OLS} MSE_M, ARMSE AADRM	(4) คำนวณค่า MSE_{OLS}, MSE_M, ARMSE และ AADRM ทำซ้ำตั้งแต่ขั้นที่ 1 (สร้างค่า y) ถึงขั้นที่ 4 จนครบ 200 ครั้ง
	คำนวณค่าเฉลี่ย, SD ของ ARMSE, AADRM	(5) คำนวณค่าเฉลี่ย, SD ของ ARMSE และ AADRM ที่ MSE_M น้อยกว่า MSE_{OLS} MSE_{OLS} " MSE_M

ศูนย์วิทยทรัพยากร
จุฬาลงกรณ์มหาวิทยาลัย

แสดงผังงานสำหรับโปรแกรมหลักที่ 2 MAIN PROGRAM 2

CALL SUBROUTINE		งานที่ทำ
(1) SUBROUTINE INIT, DATA	สร้างค่า Y, X	(1) สร้างตัวแปรสุ่มที่มีการแจกแจงแบบ - ลอกนอร์มอล - แกมมา - ไวบูลล์
(2) SUBROUTINE BOXCOX	หาค่า $\hat{\beta}$ B&C โดยวิธี OLS	(2) คำนวณค่าสัมประสิทธิ์การถดถอย $\hat{\beta}$ วิธี OLS โดยอาศัยการแปลงข้อมูลแบบ Box และ Cox
(3) SUBROUTINE BOXC	หาค่า $\hat{\beta}$ B&C โดยวิธี M-estimator	(3) คำนวณค่าสัมประสิทธิ์การถดถอย $\hat{\beta}$ วิธี M-estimator โดยอาศัยการแปลงข้อมูลแบบ Box และ Cox
	หาค่า MSE_{OLS} MSE_M, ARMSE และ AADRM	(4) คำนวณค่า MSE_{OLS}, MSE_M, ARMSE และ AADRM ทำซ้ำตั้งแต่ขั้นที่ 1 (สร้างค่า y) ถึงขั้นที่ 4 จนครบ 200 ครั้ง
	คำนวณค่าเฉลี่ย, SD ของ ARMSE และ AADRM	(5) คำนวณค่าเฉลี่ย, SD ของ ARMSE และ AADRM ที่ MSE_M น้อยกว่า MSE_{OLS} MSE_{OLS} " MSE_M

ศูนย์วิทยทรัพยากร
จุฬาลงกรณ์มหาวิทยาลัย