บทที่ 3
วิธีดำเนินการวิจัย

การวิจัยในครั้งนี้มีลักษณะเป็นการวิจัยเชิงทดลอง เพื่อต้องการศึกษาและเปรียบเทียบตัวประมาณค่าพารามิเตอร์หรือสัมประสิทธิ์การถดถอยในสมการถดถอยเชิงเส้นพหุคูณ เมื่อตัวแปรตามมีค่าผิดปกติและ/หรือตัวแปรอิสระมีพหุสัมพันธ์กันในระดับต่าง ๆ ซึ่งข้อมูลที่ใช้ในการวิจัยครั้งนี้ ได้จากการจำลองด้วยเทคนิคการจำลองมอนติคาร์โล (Monte Carlo Simulation Technique) และทำการเขียนโปรแกรมด้วยโปรแกรมภาษาฟอร์แทรน (Fortran Version 4.0 for Windows 98) ในการประมวลผลและวิเคราะห์ข้อมูล โดยตัวประมาณสัมประสิทธิ์การถดถอยที่นำมาศึกษาในครั้งนี้ มีดังต่อไปนี้

1. ตัวประมาณกำลังสองน้อยสุด (LS)
2. ตัวประมาณค่าสัมบูรณ์น้อยสุด (LAV)
3. ตัวประมาณริดจ์(RID)
4. ตัวประมาณริดจ์ที่มีค่าสัมบูรณ์น้อยสุด(RLAV)
5. ตัวประมาณริดจ์ที่มีการถ่วงน้ำหนัก(WRID)

ผู้วิจัยจะทำการเปรียบเทียบประสิทธิภาพในการประมาณค่าสัมประสิทธิ์การถดถอยของตัวประมาณการถดถอยทั้ง 5 ตัวดังกล่าว โดยใช้การเปรียบเทียบจากค่ารากที่สองของค่าเฉลี่ยของความคลาดเคลื่อนกำลังสองของการประมาณค่าสัมประสิทธิ์การถดถอยจากทุกตัวประมาณ ซึ่งแต่ละตัวประมาณจะประมาณค่าสัมประสิทธิ์การถดถอยภายใต้สถานการณ์ต่าง ๆ ที่กำหนด ซึ่งมีระดับขนาดตัวอย่าง 6 ระดับ คือ 20 , 30 , 35 , 40 , 50 และ 60 และเนื่องจาก ในการวิจัยครั้งนี้ ผู้วิจัยได้อาศัยเทคนิคการจำลองมอนติคาร์โลมาสร้างข้อมูลในสถานการณ์ต่าง ๆ ดังนั้น ในตอนแรกผู้วิจัยจะกล่าวถึงวิธีการจำลองโดยใช้เทคนิคมอนติคาร์โล แล้วจึงแสดงรายละเอียดของขั้นตอนการวิจัยในลำดับถัดไป ส่วนรายละเอียดของโปรแกรมที่ใช้ในการวิจัยจะแสดงไว้ในภาคผนวก ข. และ ค.

3.1 วิธีการจำลองโดยใช้เทคนิคมอนติคาร์โล

เทคนิคที่ใช้แก้ปัญหาในการคำนวณทางคณิตศาสตร์นั้นมีอยู่หลายวิธี วิธีการจำลองโดยใช้เทคนิคมอนติคาร์โล เป็นวิธีหนึ่งที่นิยมนำมาใช้แก้ปัญหากันอย่างแพร่หลายในปัจจุบัน ซึ่งหลักการของการจำลองโดยใช้เทคนิคดังกล่าวจะใช้เลขสุ่ม (Random Number) มาช่วยในการหาคำตอบของปัญหาที่ต้องการศึกษา ขั้นตอนของวิธีการจำลองด้วยเทคนิคมอนติคาร์โล แบ่งออกได้เป็น 3 ขั้นตอนใหญ่ ๆ ดังต่อไปนี้

1. การสร้างตัวเลขสุ่ม การใช้ตัวเลขสุ่มเป็นสิ่งสำคัญมากในเทคนิคนี้ ทั้งนี้ เป็นเพราะว่า หลักการจำลองด้วยเทคนิคมอนติคาร์โลนั้น จะใช้เลขสุ่มมาช่วยในการหาคำตอบของปัญหา โดยลักษณะของตัวเลขสุ่มที่นำมาใช้จะมีการแจกแจงสม่ำเสมอในช่วง (0,1) สำหรับวิธีการสร้างตัวเลขสุ่ม มีผู้เสนอไว้หลายวิธี แต่วิธีที่ดีนั้นลักษณะของเลขสุ่มที่ถูกสร้างขึ้นจะต้องมีการแจกแจงสม่ำเสมอในช่วง (0,1) ตัวเลขสุ่มแต่ละตัวเป็นอิสระต่อกันและมีช่วงยาวก่อนจะเกิดเลขสุ่มซ้ำ (มีวัฏจักรยาว)

2. การนำตัวเลขสุ่มมาประยุกต์ใช้กับปัญหาที่ต้องศึกษา ขั้นตอนนี้ขึ้นอยู่กับลักษณะของปัญหา บางปัญหาอาจจะไม่ได้ใช้เลขสุ่มโดยตรง แต่ใช้ในการผลิตเลขสุ่มที่มีการแจกแจงแบบอื่นต่อไป

3. การทดลองกระทำซ้ำ เมื่อนำตัวเลขสุ่มมาประยุกต์ให้เข้ากับปัญหาที่ต้องการศึกษาได้แล้ว ขั้นต่อไปคือ การทดลองโดยใช้กระบวนการของการสุ่ม (Random Process) มากระทำในลักษณะซ้ำ ๆ กัน หลาย ๆ ครั้ง เพื่อหาคำตอบของปัญหาที่ต้องการ

3.2 การวางแผนการทดลอง

การวิจัยในครั้งนี้ ต้องการประมาณค่าพารามิเตอร์หรือสัมประสิทธิ์การถดถอยในสมการถดถอยเชิงเส้นพหุคูณ เมื่อตัวแปรอิสระมีพหุสัมพันธ์และ/หรือตัวแปรตามมีค่าผิดปกติ โดยเริ่มต้นจากการสร้างค่าความคลาดเคลื่อนที่มีค่าผิดปกติในระดับต่าง ๆ และสร้างค่าตัวแปรอิสระที่มีสหสัมพันธ์กันในระดับต่าง ๆ ที่กำหนด ภายใต้ขนาดตัวอย่างต่าง ๆ กัน ซึ่งลักษณะของความคลาดเคลื่อนที่ต้องการศึกษา คือ การแจกแจงปกติ และการแจกแจงปกติปลอมปน ซึ่งจะมีการแจกแจงที่หางยาวกว่าการแจกแจงปกติ จากนั้น จะนำค่าความ

คลาดเคลื่อนและค่าตัวแปรอิสระที่ได้ ไปสร้างค่าตัวแปรตาม แล้วจึงทำการประมาณค่าพารามิเตอร์ด้วยตัวประมาณพารามิเตอร์ทั้ง 5 ตัว แล้วจึงหาค่ารากที่สองของค่าเฉลี่ยความคลาดเคลื่อนกำลังสองของตัวประมาณการถดถอยพหุคูณแต่ละตัว และทำการเปรียบเทียบค่ารากที่สองของค่าเฉลี่ยความคลาดเคลื่อนกำลังสองของตัวประมาณการถดถอยที่ได้

3.3 วิธีการทดลอง

ทำการเขียนโปรแกรมคอมพิวเตอร์ภาษาฟอร์แทรน เพื่อสร้างข้อมูลให้เป็นไปตามการทดลองที่กำหนด ซึ่งวิธีการทดลองแบ่งออกเป็น 5 ขั้นตอน ดังต่อไปนี้

1. การสร้างข้อมูลของความคลาดเคลื่อน($\underset{\sim}{\varepsilon}$)

2. การสร้างข้อมูลของตัวแปรอิสระ(X)

3. การสร้างข้อมูลของตัวแปรตาม($\underset{\sim}{y}$) จากความสัมพันธ์ $\underset{\sim}{y} = X\underset{\sim}{\beta} + \underset{\sim}{\varepsilon}$

4. การประมาณค่าพารามิเตอร์ด้วยตัวประมาณการถดถอยแต่ละตัว

5. การหาค่ารากที่สองของค่าเฉลี่ยความคลาดเคลื่อนกำลังสอง ของตัวประมาณการถดถอยแต่ละตัว และเปรียบเทียบค่ารากที่สองของค่าเฉลี่ยความคลาดเคลื่อนกำลังสองของตัวประมาณการถดถอยที่ได้

รายละเอียดของแต่ละขั้นตอนข้างต้น เป็นดังนี้

1. การสร้างข้อมูลของความคลาดเคลื่อน($\underset{\sim}{\varepsilon}$)

1.1 การแจกแจงปกติ

สร้างค่าความคลาดเคลื่อน ε_i ให้มีการแจกแจงปกติด้วยค่าเฉลี่ย(μ) เท่ากับ 0 และความแปรปรวน (σ^2) เท่ากับ 3 โดยมีสูตรในการสร้างค่า ดังต่อไปนี้

$$NORMAL_1 = \mu + \sigma Z_1$$
$$NORMAL_2 = \mu + \sigma Z_2$$

โดยการเขียนโปรแกรม กระทำโดยการเรียกฟังก์ชัน NORMAL(MEAN,SD,IX) ซึ่งแสดงรายละเอียดในการสร้างค่าไว้ในภาคผนวก ข. (หน้า 308)

1.2 การแจกแจงปกติปลอมปน

จากฟังก์ชันการแจกแจง

$$f(\varepsilon) = (1-P)N(0,3) + PN(0,3C^2)$$

สร้างค่าความคลาดเคลื่อน ε_i ให้มีการแจกแจงปกติด้วยค่าเฉลี่ย(μ) เท่ากับ 0 และความแปรปรวน (σ^2) เท่ากับ 3 ด้วยความน่าจะเป็นเท่ากับ (1-P) โดยการเรียกฟังก์ชัน NORMAL(MEAN,SD,IX) และสร้างค่าความคลาดเคลื่อน ε_i ให้มีการแจกแจงปกติด้วยค่าเฉลี่ย(μ) เท่ากับ 0 และความแปรปรวน (σ^2) เท่ากับ $3C^2$ ด้วยความน่าจะเป็นเท่ากับ P โดยการเรียกฟังก์ชัน NORMAL(MEAN,CSD,IX) ซึ่งแสดงรายละเอียดไว้ในภาคผนวก ข. (หน้า 308)

ในการวิจัยครั้งนี้ กำหนดระดับสัดส่วนการปลอมปน(P) เท่ากับ 0.05 , 0.08 , 0.10 และ 0.20
และกำหนดระดับค่าผิดปกติ 3 ระดับ คือ ระดับเล็กน้อย ปานกลาง และรุนแรง โดยใช้เกณฑ์การกำหนดขนาดค่าผิดปกติด้วย Box Plot จะได้ว่า
เมื่อค่า C = 4 - 6 ตัวแปรตามจะมีค่าผิดปกติระดับเล็กน้อย - ปานกลาง
และ เมื่อค่า C = 12 - 13 ตัวแปรตามจะมีค่าผิดปกติระดับรุนแรง
ดังนั้น ในการวิจัยครั้งนี้ เลือกกำหนดค่า C = 4 เมื่อตัวแปรตามมีค่าผิดปกติระดับเล็กน้อย C = 6 เมื่อตัวแปรตามมีค่าผิดปกติระดับปานกลาง และ C = 13 เมื่อตัวแปรตามมีค่าผิดปกติระดับรุนแรง

2. การสร้างข้อมูลของตัวแปรอิสระ(X)

การวิจัยในครั้งนี้ กำหนดจำนวนตัวแปรอิสระที่ใช้ในการวิจัยเท่ากับ 3 ตัวแปร ได้แก่ ตัวแปรอิสระ x_1 , x_2 และ x_3 โดยกำหนดให้มีพหุสัมพันธ์ระหว่างตัวแปรอิสระ x_1 กับ x_2 และมีระดับสหสัมพันธ์ระหว่างตัวแปรอิสระ x_1 กับ x_2 (ρ) จำนวน 7 ระดับ คือ 0.1 ,

0.3 , 0.5 , 0.7 , 0.9 , 0.95 และ 0.99 และทำการจำลองตัวแปรอิสระ x_1 กับ x_2 โดยใช้วิธีจำลองของ Wilchern และ Churchill(1978) ซึ่งมีสูตรดังนี้

$$x_{ij} = (1-\rho^2)^{1/2}.Z_{ij} + \rho Z_{i3} \qquad ; i = 1,2,\ldots,n \quad , j = 1,2$$

เมื่อ $Z_{i1},\ldots,Z_{i3}$ แทน ตัวเลขสุ่มที่มีการแจกแจงปกติมาตรฐานที่เป็นอิสระต่อกัน (Independent Standard Normal Random Number)

ρ แทน สหสัมพันธ์ระหว่างตัวแปรอิสระ x_1 กับ x_2

และสำหรับตัวแปรอิสระ x_3 ในการวิจัยครั้งนี้ กำหนดให้มีการแจกแจงปกติ ที่มีค่าเฉลี่ยเท่ากับ 0* ความแปรปรวนเท่ากับ 5* และไม่มีสหสัมพันธ์กับตัวแปรอิสระ x_1 และ x_2 และทำการจำลองโดยการเรียกโปรแกรมย่อย GENX(X,N,Q,IX,CORR) ซึ่งแสดงรายละเอียดไว้ในภาคผนวก ข. (หน้า 308)

3. การสร้างข้อมูลของตัวแปรตาม($\underset{\sim}{y}$)

สร้างค่าตัวแปรตาม $\underset{\sim}{y}$ จากตัวแบบการถดถอยพหุคูณเชิงเส้น ซึ่งอยู่ในรูปแบบดังนี้

$$y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \beta_3 x_{i3} + \varepsilon_i \quad ; i = 1,2,\ldots,n$$

หรือเขียนในรูปเมทริกซ์ ได้ดังนี้

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

---

* การวิจัยครั้งนี้ได้ทดลองให้ค่าเฉลี่ยและความแปรปรวน มีค่าอื่น ๆ ได้ผลลัพธ์ไม่แตกต่างกัน ซึ่งได้แสดงตัวอย่างของผลลัพธ์จากการใช้ค่าเฉลี่ยและความแปรปรวนค่าอื่น ไว้ในภาคผนวก ง. (หน้า 345)

โดยการวิจัยในครั้งนี้ กำหนดให้ ค่าจริงของพารามิเตอร์ $\underset{\sim}{\beta} = \underset{\sim}{1}$* และทำการจำลองโดยการเรียกโปรแกรมย่อย BUILDY(Y,BETA,ERROR,Y,XB,N,Q)

เมื่อได้ข้อมูลของตัวแปรตาม $\underset{\sim}{y}$ แล้ว จะนำค่าของตัวแปรตาม $\underset{\sim}{y}$ มาตรวจสอบค่าผิดปกติโดยใช้กราฟแบบ Box Plot โดยมีเกณฑ์ในการกำหนดค่าผิดปกติ** ดังต่อไปนี้

ถ้าพบว่า ค่าของตัวแปรตาม $\underset{\sim}{y}$ ทุกค่าอยู่ในช่วง (Q1 – 1.5(IQR) , Q3 + 1.5(IQR)) จะถือว่า ข้อมูลของตัวแปรตาม $\underset{\sim}{y}$ ทุกค่ามีค่าปกติ

เมื่อ Q1 คือ ค่าควอไทล์ที่ 1 (First Quatile) ของตัวแปรตาม $\underset{\sim}{y}$

Q3 คือ ค่าควอไทล์ที่ 3 (Third Quatile) ของตัวแปรตาม $\underset{\sim}{y}$

และ IQR(Interquatile Range) = Q3 – Q1

ถ้าพบว่า ค่าของตัวแปรตาม $\underset{\sim}{y}$ มีค่าอยู่ในช่วง (Q1 – 3(IQR) , Q1 - 1.5(IQR)) หรือ อยู่ในช่วง (Q3 + 1.5(IQR) , Q3 + 3(IQR)) จะถือว่า ค่าของตัวแปรตาม $\underset{\sim}{y}$ ที่อยู่ในช่วงเหล่านี้เป็นค่าผิดปกติระดับเล็กน้อย - ปานกลาง

ถ้าพบว่า ค่าของตัวแปรตาม $\underset{\sim}{y}$ มีค่าน้อยกว่าค่าของ Q1 – 3(IQR) หรือ มีค่ามากกว่าค่าของ Q3 + 3(IQR) จะถือว่า ค่าของตัวแปรตาม $\underset{\sim}{y}$ ที่อยู่ในช่วงเหล่านี้เป็นค่าผิดปกติระดับรุนแรง

---

* การวิจัยครั้งนี้ได้ทดลองให้ค่าของพารามิเตอร์ $\underset{\sim}{\beta}$ มีค่าอื่น ๆ ได้ผลลัพธ์ไม่แตกต่างกัน ซึ่งได้แสดงตัวอย่างของผลลัพธ์จากการใช้ค่าของพารามิเตอร์ $\underset{\sim}{\beta}$ ค่าอื่น ไว้ในภาคผนวก จ. (หน้า 357)

** ข้อมูลได้จากหนังสือ " Modem Data Analysis : A First Course in Applied Statistics " (Hamilton , 1990)

จากการทดสอบ ได้ว่า เมื่อกำหนดระดับสเกลแฟคเตอร์(C) เท่ากับ 4 - 6 จะได้ค่าผิดปกติในระดับเล็กน้อย - ปานกลาง (เนื่องจาก จะได้ค่าที่อยู่ในช่วง (Q1 – 3(IQR) , Q1 - 1.5(IQR)) หรือ อยู่ในช่วง(Q3 + 1.5(IQR) , Q3 + 3(IQR)) และเมื่อกำหนดระดับสเกลแฟคเตอร์(C) เท่ากับ 12 - 13 จะได้ค่าผิดปกติในระดับรุนแรง (เนื่องจาก จะได้ค่าที่มีค่าน้อยกว่าค่าของ Q1 – 3(IQR) หรือ มีค่ามากกว่าค่าของ Q3 + 3(IQR)) โดยแสดงรายละเอียดในการกำหนดขนาดของค่าผิดปกติไว้ในภาคผนวก ก. (หน้า 306) ซึ่งในการวิจัยครั้งนี้ เลือกกำหนดค่า C เท่ากับ 4 เมื่อตัวแปรตามมีค่าผิดปกติระดับเล็กน้อย ค่า C เท่ากับ 6 เมื่อตัวแปรตามมีค่าผิดปกติระดับปานกลาง และ ค่า C เท่ากับ 13 เมื่อตัวแปรตามมีค่าผิดปกติระดับรุนแรง ซึ่งใช้ค่าต่าง ๆ เหล่านี้ เป็นตัวประกอบในการผลิตค่าของตัวแปรตาม $\underset{\sim}{y}$ ที่ใช้ในการทดลองครั้งนี้

4. การประมาณค่าพารามิเตอร์ด้วยตัวประมาณการถดถอยแต่ละตัว

เมื่อสร้างข้อมูลของตัวแปรอิสระ(X) และตัวแปรตาม($\underset{\sim}{y}$) ได้แล้ว ขั้นต่อไป คือ การนำข้อมูลของตัวแปรอิสระ(X) และตัวแปรตาม($\underset{\sim}{y}$) ที่ได้ ไปประมาณค่าพารามิเตอร์ $\underset{\sim}{\beta}$ ซึ่งในการวิจัยครั้งนี้ ได้ทำการประมาณค่าพารามิเตอร์ด้วยตัวประมาณการถดถอยเชิงเส้นพหุคูณ 5 ตัว ดังต่อไปนี้

4.1 ตัวประมาณกำลังสองน้อยสุด

ประมาณค่าพารามิเตอร์ $\underset{\sim}{\beta}$ ด้วยตัวประมาณกำลังสองน้อยสุด $\hat{\underset{\sim}{\beta}}_{LS}$ จากสมการ (2.1) ในบทที่ 2 ซึ่งมีรูปแบบตัวประมาณดังนี้

$$\hat{\underset{\sim}{\beta}}_{LS} = (X'X)^{-1}.X'\underset{\sim}{y}$$

4.2 ตัวประมาณค่าสัมบูรณ์น้อยสุด

ประมาณค่าพารามิเตอร์ $\underset{\sim}{\beta}$ ด้วยวิธีซิมเพล็กซ์(Simplex Method) จะได้ตัวประมาณค่าสัมบูรณ์น้อยสุด ซึ่งมีรูปแบบตัวประมาณดังนี้

$$\hat{\beta}_0 = \hat{\beta}_0^{+} - \hat{\beta}_0^{-}$$

$$\hat{\beta}_1 = \hat{\beta}_1^{+} - \hat{\beta}_1^{-}$$

$$\hat{\beta}_2 = \hat{\beta}_2^{+} - \hat{\beta}_2^{-}$$

และ $$\hat{\beta}_3 = \hat{\beta}_3^{+} - \hat{\beta}_3^{-}$$

โดยที่ $\hat{\beta}_0^{+}, \hat{\beta}_0^{-}, \hat{\beta}_1^{+}, \hat{\beta}_1^{-}, \hat{\beta}_2^{+}, \hat{\beta}_2^{-}, \hat{\beta}_3^{+}, \hat{\beta}_3^{-}$ ได้จากการแก้ปัญหาโปรแกรมเชิงเส้นในหัวข้อ 2.1.2 ด้วยวิธีซิมเพล็กซ์

4.3 ตัวประมาณริดจ์

ประมาณค่าพารามิเตอร์ $\underset{\sim}{\beta}$ ด้วยตัวประมาณการถดถอยริดจ์ $\hat{\underset{\sim}{\beta}}_{RID}$ จากสมการ (2.5) ในบทที่ 2 ซึ่งมีรูปแบบตัวประมาณดังนี้

$$\hat{\underset{\sim}{\beta}}_{RID} = \left(X'X + kI\right)^{-1} X' \underset{\sim}{y}$$

และใช้วิธีการประมาณค่า k ที่คิดค้นโดย Hoerl , Kennard และ Bladwin (1975) โดยอิงพื้นฐานจากตัวประ มาณกำลังสองน้อยสุด(LS) ซึ่งแทนด้วย k_{HKB} เขียนได้ในรูปแบบดังนี้

$$k_{HKB} = \frac{qs^2}{\hat{\underset{\sim}{\beta}}_{LS}' \hat{\underset{\sim}{\beta}}_{LS}}$$

เมื่อ $s^2 = \dfrac{\left(\underset{\sim}{y} - X\hat{\underset{\sim}{\beta}}_{LS}\right)'\left(\underset{\sim}{y} - X\hat{\underset{\sim}{\beta}}_{LS}\right)}{n-q}$ เป็น ตัวประมาณพารามิเตอร์ σ^2

4.4 ตัวประมาณริดจ์ที่มีค่าสัมบูรณ์น้อยสุด

ประมาณค่าพารามิเตอร์ $\underset{\sim}{\beta}$ ด้วยตัวประมาณริดจ์ที่มีค่าสัมบูรณ์น้อยสุด $\hat{\underset{\sim}{\beta}}_{RLAV}$ จากสมการ (2.8) ในบทที่ 2 ซึ่งมีรูปแบบตัวประมาณดังนี้

$$\underset{\sim RLAV}{\hat{\beta}} = \left(X'X + k^{*}I\right)^{-1} X'\underset{\sim}{y}$$

โดยใช้ค่า k^{*} ที่ได้จากวิธีการประมาณค่า k ที่คิดค้นโดย Hoerl , Kennard และ Bladwin (1975) แต่อิงพื้นฐานจากตัวประมาณค่าสัมบูรณ์ต่ำสุด(LAV) แทนการอิงพื้นฐานจากตัวประมาณกำลังสองน้อยสุด(LS) ซึ่งมีรูปแบบค่า k^{*} เป็นดังนี้

$$k^{*} = \frac{q s_{LAV}^{2}}{\underset{\sim LAV}{\hat{\beta}}' \underset{\sim LAV}{\hat{\beta}}}$$

เมื่อ $s_{LAV}^{2} = \dfrac{\left(\underset{\sim}{y} - X\underset{\sim LAV}{\hat{\beta}}\right)'\left(\underset{\sim}{y} - X\underset{\sim LAV}{\hat{\beta}}\right)}{n-q}$ เป็น ตัวประมาณพารามิเตอร์ σ^{2}

4.5 ตัวประมาณริดจ์ที่มีการถ่วงน้ำหนัก

ประมาณค่าพารามิเตอร์ $\underset{\sim}{\beta}$ ด้วยตัวประมาณริดจ์ที่มีการถ่วงน้ำหนัก $\underset{\sim WRID}{\hat{\beta}}$ จากสมการ (2.7) ในบทที่ 2 ซึ่งมีรูปแบบตัวประมาณดังนี้

$$\underset{\sim WRID}{\hat{\beta}} = \left(X'WX + kI\right)^{-1} X'W\underset{\sim}{y}$$

ซึ่งการวิจัยในครั้งนี้ จะเลือกใช้ค่าของน้ำหนัก $w_{ii} = \dfrac{1}{|e_i|}$ ซึ่งเสนอโดย Pfaffenberger และ Dielman (1990)

เมื่อ e_i แทน ส่วนเหลือ(Residual)จากการใช้ตัวประมาณค่าสัมบูรณ์ต่ำสุด(LAV) ในการประมาณค่าตัวแปรตาม $\underset{\sim}{y}$

และใช้วิธีการประมาณค่า k ที่คิดค้นโดย Hoerl , Kennard และ Bladwin (1975) โดยอิงพื้นฐานจากตัวประ มาณกำลังสองน้อยสุด(LS) ซึ่งแทนด้วย k_{HKB} เขียนได้ในรูปแบบดังนี้

$$k_{HKB} = \frac{q s^{2}}{\underset{\sim LS}{\hat{\beta}}' \underset{\sim LS}{\hat{\beta}}}$$

เมื่อ $s^2 = \dfrac{\left(\underset{\sim}{y} - X\underset{\sim LS}{\hat{\beta}}\right)'\left(\underset{\sim}{y} - X\underset{\sim LS}{\hat{\beta}}\right)}{n-q}$ เป็น ตัวประมาณพารามิเตอร์ σ^2

5. การหาค่ารากที่สองของค่าเฉลี่ยความคลาดเคลื่อนกำลังสองของตัวประมาณการถดถอยแต่ละตัว

เมื่อได้ค่าประมาณพารามิเตอร์จากตัวประมาณการถดถอยทั้ง 5 ตัวแล้ว นำมาหาค่าผลรวมกำลังสองของความคลาดเคลื่อนจากการประมาณค่าพารามิเตอร์แต่ละตัว โดยนำค่าประมาณที่ได้จากตัวประมาณแต่ละตัว ลบออกด้วยค่าจริงของพารามิเตอร์ ซึ่งกำหนดให้ $\underset{\sim}{\beta} = 1$ แล้วยกกำลังสองสะสมไว้ในแต่ละครั้ง และทำการทดลองเช่นเดิมจนครบ 1,000 ครั้ง แล้วจึงคำนวณหาค่ารากที่สองของค่าเฉลี่ยของความคลาดเคลื่อนกำลังสอง ของตัวประมาณการถดถอยพหุคูณเชิงเส้น(RMSE) ซึ่งมีสูตรในการคำนวณดังนี้

$$RMSE_i = \sqrt{\sum_{j=1}^{1,000} \frac{\left(\beta_{ij} - \hat{\beta}_{ij}\right)^2}{1,000}} \quad , i = 0,1,2,3$$

เมื่อ β_{ij} แทน ค่าจริงของพารามิเตอร์ตัวที่ i ในสมการถดถอย ในการจำลองรอบที่ j

$\hat{\beta}_{ij}$ แทน ค่าประมาณของพารามิเตอร์ตัวที่ i ในสมการถดถอย ในการจำลองรอบที่ j

$RMSE_i$ แทน ค่ารากที่สองของค่าเฉลี่ยความคลาดเคลื่อนกำลังสองของตัวประมาณพารามิเตอร์ตัวที่ i ในสมการถดถอย

จากนั้นจึงทำการเปรียบเทียบค่าRMSE ที่ได้จากตัวประมาณการถดถอยทั้ง 5 ตัว เพื่อพิจารณาว่าตัวประมาณใดมีประสิทธิภาพสูงสุดในแต่ละสถานการณ์ และทำการทดลองเช่นนี้โดยเปลี่ยนระดับพหุสัมพันธ์ของตัวแปรอิสระ ระดับสัดส่วนการปลอมปนของความคลาดเคลื่อน ระดับขนาดของค่าผิดปกติ และระดับขนาดตัวอย่าง จนกระทั่งครบทุกรูปแบบของสถานการณ์ที่ต้องการศึกษา และเพื่อให้ง่ายแก่การเข้าใจ จึงได้แสดงผังงานซึ่งแสดงขั้นตอนการวิจัยทั้งหมดในรูปที่ 3.1 ดังต่อไปนี้

รูปที่ 3.1 แสดงผังงานสำหรับหาค่ารากที่สองของค่าเฉลี่ยความคลาดเคลื่อนกำลังสองของตัวประมาณการถดถอยทั้ง 5 ตัว