บทที่ 5
สรุปผลการวิจัยและข้อเสนอแนะ

การวิจัยครั้งนี้มีวัตถุประสงค์เพื่อศึกษาและเปรียบเทียบวิธีการประมาณค่าพารามิเตอร์หรือสัมประสิทธิ์ความถดถอยในสมการถดถอยเชิงเส้นพหุคูณ เมื่อตัวอย่างมีขนาดเล็ก โดยจะศึกษาวิธีการประมาณค่า 4 วิธีคือ

1. วิธีกำลังสองน้อยสุด (Ordinary Least Squares method (OLS))
2. วิธีบูตสแตรป (Bootstrap Method (BS))
3. วิธีตัวประมาณ M โดยใช้เกณฑ์ความแกร่งของแรมเซย์ (M-Estimator method by Ramsay (M/R))
4. วิธีตัวประมาณ M โดยใช้เกณฑ์ความแกร่งของตูกี (M-Estimator method by Tukey (M/T))

วิธีกำลังสองน้อยสุดเป็นวิธีพื้นฐานที่นิยมใช้กันอย่างแพร่หลาย แต่เป็นวิธีที่มีประสิทธิภาพต่ำเมื่อขนาดตัวอย่างน้อย วิธีบูตสแตรปเป็นวิธีการประมาณค่าที่ไม่ใช้พารามิเตอร์ (non-parametric estimation) จึงน่าจะเป็นวิธีที่เหมาะสมกับขนาดตัวอย่างน้อย และวิธีตัวประมาณ M เป็นวิธีที่มีความแกร่ง (robustness) จึงเป็นที่น่าสนใจว่าวิธีประมาณค่าวิธีใดที่มีประสิทธิภาพมากที่สุดในสถานการณ์ใด ซึ่งในงานวิจัยนี้ได้กำหนดสถานการณ์ต่างๆไว้ดังนี้

1. จำนวนตัวแปรอิสระที่ศึกษาเท่ากับ 3 และ 6
2. ขนาดตัวอย่างที่ศึกษาเท่ากับ 10, 15, 20, 25 และ 30
3. ลักษณะการแจกแจงของค่าความคลาดเคลื่อนมีดังนี้
    3.1 <u>การแจกแจงปกติ</u> ศึกษาในกรณีที่ค่าเฉลี่ยเท่ากับ 0 กับ 1 และค่าความแปรปรวนเท่ากับ 1, 4 และ 9 ตามลำดับ
    3.2 <u>การแจกแจงปกติปลอมปน</u> ศึกษาในกรณีที่ค่าเฉลี่ยเท่ากับ 0 ค่าความแปรปรวนเท่ากับ 1 ค่าสเกลแฟคเตอร์เท่ากับ 10 และเปอร์เซ็นต์การปลอมปนเท่ากับ 25%, 50% และ 75% ตามลำดับ
    3.3 <u>การแจกแจงแกมมา</u> ศึกษาในกรณีที่พารามิเตอร์เบต้าเท่ากับ 1 และพารามิเตอร์แอลฟาเท่ากับ 0.5, 1, 2 และ 4

เกณฑ์ที่ใช้ในการตัดสินใจว่าวิธีการประมาณค่าสัมประสิทธิ์การถดถอยเชิงเส้นพหุคูณเมื่อขนาดตัวอย่างน้อยวิธีใดที่เหมาะสมที่สุดจะพิจารณาจากค่าเฉลี่ยความคลาดเคลื่อนกำลังสองเฉลี่ย (average of mean square error (AMSE)) นอกจากนี้ยังใช้ค่าอัตราส่วนผลต่างของค่าเฉลี่ยของค่าความคลาดเคลื่อนกำลังสองเฉลี่ย (ratio of different average mean square error (RDAMSE)) เป็นค่าที่ใช้ประกอบการตัดสินใจ ซึ่งวิธีใดให้จากค่าเฉลี่ยความคลาดเคลื่อนกำลังสองเฉลี่ยและค่าอัตราส่วนผลต่างของค่าเฉลี่ยของค่าความคลาดเคลื่อนกำลังสองเฉลี่ยต่ำสุดจะเป็นวิธีที่ดีที่สุด

การสรุปผลการวิจัยจะแบ่งออกเป็น 2 ส่วน คือ

1. ผลสรุปการวิจัยในเชิงพรรณนา

ผลสรุปส่วนนี้จะสรุปผลโดยการพิจารณาจากค่า AMSE และ RDAMSE ทั้ง 3 ขั้นตอนทั้งหมดกล่าวคือจากตารางที่ 4.1 – 4.6 และรูปที่ 4.1 – 4.14 เพื่อให้เห็นชัดเจนว่าการประมาณด้วยวิธีใดเหมาะสมกับสถานการณ์นั้นมากที่สุด และพิจารณาว่าปัจจัยใดบ้างที่มีผลกระทบต่อประสิทธิภาพของตัวประมาณแต่ละวิธีมากที่สุด

2. ผลสรุปการวิจัยในเชิงอนุมาน

ผลสรุปนี้เป็นผลสรุปเพิ่มเติมที่ช่วยสนับสนุนผลในส่วนแรก โดยจะทำการทดสอบสมมติฐานทางสถิติเปรียบเทียบความแตกต่างระหว่างค่า AMSE ของวิธี BS, M/R และ M/T ของทุกสถานการณ์ ซึ่งผลสรุปนี้ไม่ใช่ส่วนสำคัญของงานวิจัยจึงได้แสดงตารางสรุปผลนี้ไว้ในภาคผนวก ก (อยู่ในรูปตารางที่ได้จากผลลัพธ์ของโปรแกรม SPSS12.0)

5.1 ผลสรุปทางสถิติในเชิงพรรณนา

5.1.1 ผลสรุปค่า AMSE ในกรณีที่ค่าความคลาดเคลื่อนมีการแจกแจงปกติ

จากการเปรียบเทียบค่า AMSE พบว่าเมื่อจำนวนตัวแปรอิสระเท่ากับ 3 ทุกระดับค่าเฉลี่ยและความแปรปรวน ตัวประมาณ BS จะเป็นตัวประมาณที่ให้ค่า AMSE ต่ำสุดเมื่อขนาดตัวอย่างเท่ากับ 10 และ 15 แต่เมื่อขนาดตัวอย่างเพิ่มเป็น 20, 25 และ 30 วิธี M/T จะให้ค่า AMSE ต่ำสุด รองลงมาคือวิธี M/R และ OLS ตามลำดับ

ในกรณีที่จำนวนตัวแปรอิสระเท่ากับ 6 ทุกระดับความแปรปรวน เมื่อค่าเฉลี่ยเท่ากับ 0 ตัวประมาณ BS จะเป็นตัวประมาณที่ให้ค่า AMSE ต่ำสุดเมื่อขนาดตัวอย่างเท่ากับ 10, 15 และ 20 แต่เมื่อขนาดตัวอย่างเพิ่มเป็น 25 และ 30 วิธี M/T จะให้ค่า AMSE ต่ำสุด รองลงมาคือวิธี M/R และ OLS และเมื่อค่าเฉลี่ยเท่ากับ 1 ตัวประมาณ BS จะเป็นตัวประมาณที่ให้ค่า AMSE ต่ำสุดเมื่อขนาดตัวอย่าง

เท่ากับ 10 และ 15 แต่เมื่อขนาดตัวอย่างเพิ่มเป็น 20, 25 และ 30 วิธี M/T จะให้ค่า AMSE ต่ำสุด รองลงมาคือวิธี M/R และ OLS ตามลำดับ

5.1.2 ผลสรุปค่า AMSE ในกรณีที่ค่าความคลาดเคลื่อนมีการแจกแจงปกติปลอมปน

จากการเปรียบเทียบค่า AMSE พบว่าเมื่อจำนวนตัวแปรอิสระเท่ากับ 3 ทุกระดับเปอร์เซ็นต์การปลอมปน ตัวประมาณ BS จะเป็นตัวประมาณที่ให้ค่า AMSE ต่ำสุดเมื่อขนาดตัวอย่างเท่ากับ 10, 15 และ 20 แต่เมื่อขนาดตัวอย่างเพิ่มเป็น 25 และ 30 วิธี M/T จะให้ค่า AMSE ต่ำสุด รองลงมาคือวิธี M/R และ OLS ตามลำดับ

ในกรณีที่จำนวนตัวแปรอิสระเท่ากับ 6 ทุกระดับเปอร์เซ็นต์การปลอมปน และทุกระดับขนาดตัวอย่าง ตัวประมาณ M/T จะเป็นตัวประมาณที่ให้ค่า AMSE ต่ำสุด รองลงมาคือวิธี BS, M/R และ OLS ตามลำดับ

5.1.3 ผลสรุปค่า AMSE ในกรณีที่ค่าความคลาดเคลื่อนมีการแจกแจงแกมมา

จากการเปรียบเทียบค่า AMSE พบว่าเมื่อจำนวนตัวแปรอิสระเท่ากับ 3 และ 6 พารามิเตอร์เบต้าเท่ากับ 1 ทุกระดับพารามิเตอร์แอลฟาและทุกระดับขนาดตัวอย่าง ตัวประมาณ M/R จะเป็นตัวประมาณที่ให้ค่า AMSE ต่ำสุด รองลงมาคือวิธี M/T, BS และ OLS ตามลำดับ

5.1.4 ปัจจัยที่มีผลต่อค่า AMSE

ปัจจัยที่มีผลต่อค่า AMSE แบ่งเป็น 2 ประเภทคือ

1. ปัจจัยที่ค่า AMSE แปรผันตาม คือ

1.1 ค่าความแปรปรวนของค่าความคลาดเคลื่อน

เมื่อค่าความแปรปรวนสูงขึ้นแสดงว่าข้อมูลมีการกระจายมากขึ้นหรือกล่าวได้ว่าข้อมูลมีความแตกต่างกันมากขึ้น ดังนั้นโอกาสที่ค่าประมาณจะคลาดเคลื่อนไปจากค่าจริงมากก็จะยิ่งสูงขึ้นด้วย จึงเป็นสาเหตุทำให้ค่า AMSE เพิ่มสูงขึ้น

ค่าความแปรปรวนของค่าความคลาดเคลื่อนน่าจะเป็นปัจจัยที่มีอิทธิพลต่อการเพิ่มขึ้นของค่า AMSE มากที่สุดเพราะเป็นค่าที่กำหนดลักษณะการกระจายของข้อมูลซึ่งจะเห็นได้จากผลการวิเคราะห์เมื่อค่าความคลาดเคลื่อนมีการแจกแจงหรือการกระจายเปลี่ยนไป วิธีการประมาณค่าที่มีประสิทธิภาพที่สุดก็จะเปลี่ยนไปด้วย

1.2 จำนวนตัวแปรอิสระ

เมื่อจำนวนตัวแปรอิสระมากขึ้นแสดงว่ามีพารามิเตอร์ที่จะต้องประมาณค่าเพิ่มขึ้นด้วย ค่าความคลาดเคลื่อนที่เกิดจากการประมาณค่าย่อมเพิ่มขึ้นด้วย ดังนั้นค่า AMSE จึงเพิ่มขึ้น

1.3 เปอร์เซ็นต์การปลอมปนของค่าผิดปกติในค่าความคลาดเคลื่อน

เมื่อมีค่าผิดปกติมากขึ้นทำให้โอกาสที่ค่าประมาณจะผิดพลาดไปจากค่าจริงย่อมมีสูงขึ้นด้วย ค่าความคลาดเคลื่อนจึงมีมากขึ้น

2. ปัจจัยที่ค่า AMSE แปรผกผัน คือ

2.1 ขนาดตัวอย่าง

การเพิ่มขนาดตัวอย่างทำให้ตัวประมาณมีค่าใกล้เคียงค่าจริงมากยิ่งขึ้น (ตามกฎของเลขจำนวนมาก) ซึ่งจะช่วยลดผลกระทบหรือลดอิทธิพลจากค่าความคลาดเคลื่อนลงได้ ดังนั้นค่า AMSE ซึ่งเป็นค่าที่ใช้วัดความคลาดเคลื่อนที่เกิดขึ้นจึงมีแนวโน้มลดลงด้วย

5.2 ผลสรุปทางสถิติในเชิงอนุมาน

ผลสรุปส่วนนี้จะพิจารณาความแตกต่างของวิธีการประมาณค่า 3 วิธีคือวิธี BS, M/R และ M/T ด้วยการทดสอบสมมติฐานทางสถิติโดยจะใช้วิธีการวิเคราะห์ความแปรปรวนจำแนกทางเดียว (one-way Analysis of Variance(one-way ANOVA)) ซึ่งมีค่าสถิติเอฟ (F test) เป็นค่าที่ใช้ในการทดสอบเมื่อค่า AMSE ของแต่ละวิธีมีความแปรปรวนเท่ากัน ในกรณีที่ความแปรปรวนไม่เท่ากันจะใช้ค่าสถิติของบาวน์ (Brown's test) แทนเพราะการใช้ค่าสถิติดังกล่าวเป็นการทดสอบที่มีความแกร่งมากกว่าการใช้ค่าเอฟในกรณีที่การทดสอบไม่เป็นไปตามเงื่อนไขของการวิเคราะห์ความแปรปรวน การทดสอบในส่วนนี้จะใช้โปรแกรม SPSS12.0 ช่วยในการประมวลผลโดยแสดงไว้ในภาคผนวก ก

ผลการวิเคราะห์สรุปได้ว่าสถานการณ์ที่วิธีการประมาณค่าทั้ง 3 วิธีมีความแตกต่างกันที่ระดับนัยสำคัญ 0.05 คือในกรณีที่ค่าความคลาดเคลื่อนมีการแจกแจงแกมมาเมื่อ

1. จำนวนตัวแปรอิสระเท่ากับ 3 ค่าพารามิเตอร์แอลฟาเท่ากับ 0.5 และพารามิเตอร์เบต้าเท่ากับ 1 เมื่อขนาดตัวอย่างเท่ากับ 20, 25 และ 30

2. จำนวนตัวแปรอิสระเท่ากับ 3 ค่าพารามิเตอร์แอลฟาเท่ากับ 1 และพารามิเตอร์เบต้าเท่ากับ 1 เมื่อขนาดตัวอย่างเท่ากับ 20, 25 และ 30
3. จำนวนตัวแปรอิสระเท่ากับ 3 ค่าพารามิเตอร์แอลฟาเท่ากับ 2 และพารามิเตอร์เบต้าเท่ากับ 1 เมื่อขนาดตัวอย่างเท่ากับ 30
4. จำนวนตัวแปรอิสระเท่ากับ 6 ค่าพารามิเตอร์แอลฟาเท่ากับ 0.5 และพารามิเตอร์เบต้าเท่ากับ 1 เมื่อขนาดตัวอย่างเท่ากับ 20, 25 และ 30

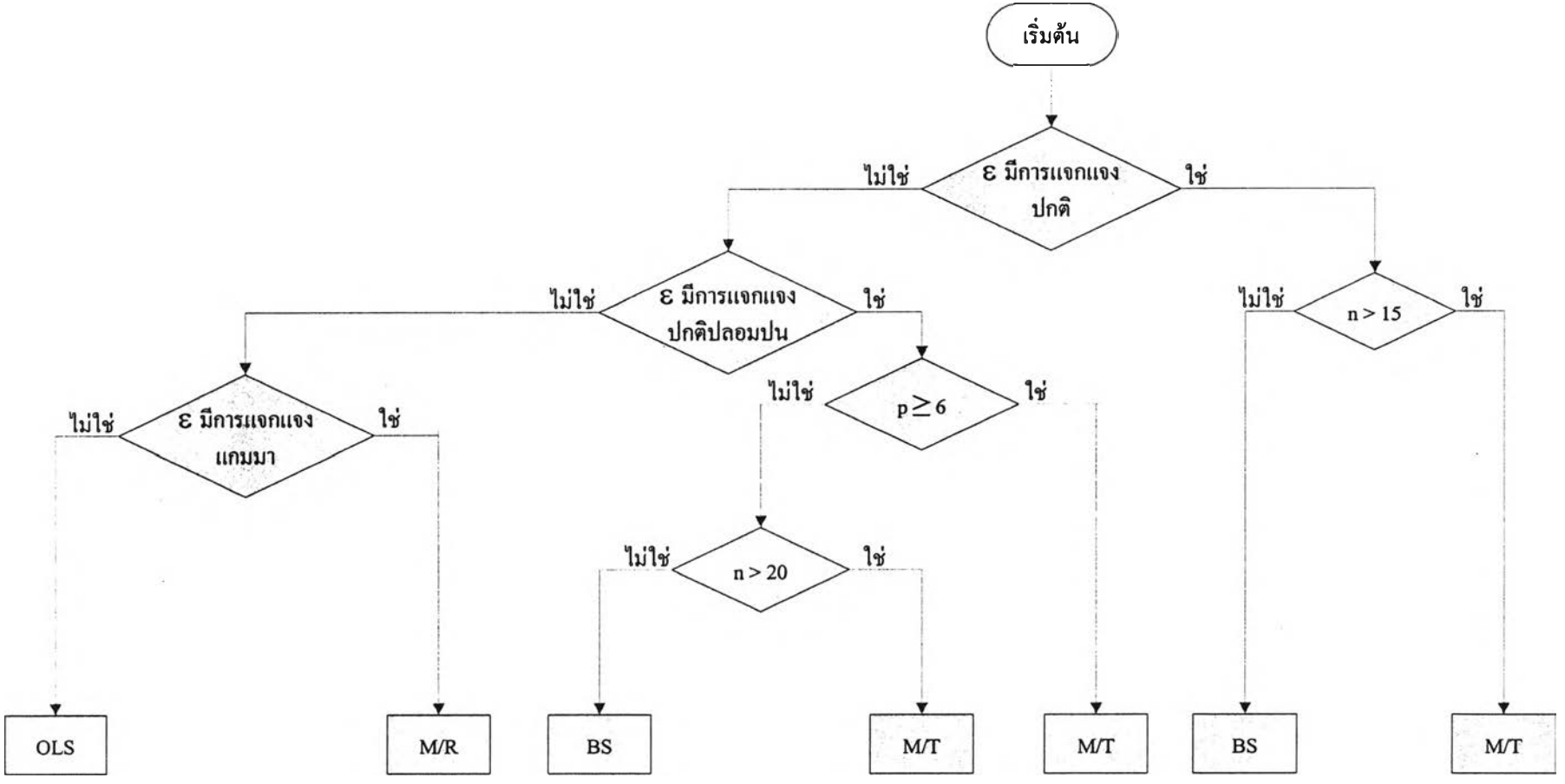
โดยที่ในสถานการณ์ดังกล่าวทั้งหมดเมื่อพิจารณาจากแผนภาพค่าเฉลี่ยพบว่าวิธี M/R จะให้ค่า AMSE ต่ำสุดรองลงมาคือวิธี M/T และ BS ตามลำดับ

ผลสรุปดังกล่าวแสดงให้เห็นว่าทั้ง 3 วิธีจะมีค่า AMSE ความแตกต่างกันมากที่สุดในกรณีที่ค่าความคลาดเคลื่อนมีการแจกแจงแกมมา จำนวนตัวแปรอิสระเท่ากับ 3 ค่าพารามิเตอร์แอลฟาเท่ากับ 0.5 และพารามิเตอร์เบต้าเท่ากับ 1 โดยที่เมื่อขนาดตัวอย่างเพิ่มขึ้นทั้ง 3 วิธีจะยิ่งมีความแตกต่างกันมากขึ้น

5.3 ผลสรุปการเลือกวิธีการประมาณค่าสัมประสิทธิ์การถดถอยเชิงเส้นพหุคูณเมื่อตัวอย่างมีขนาดเล็ก

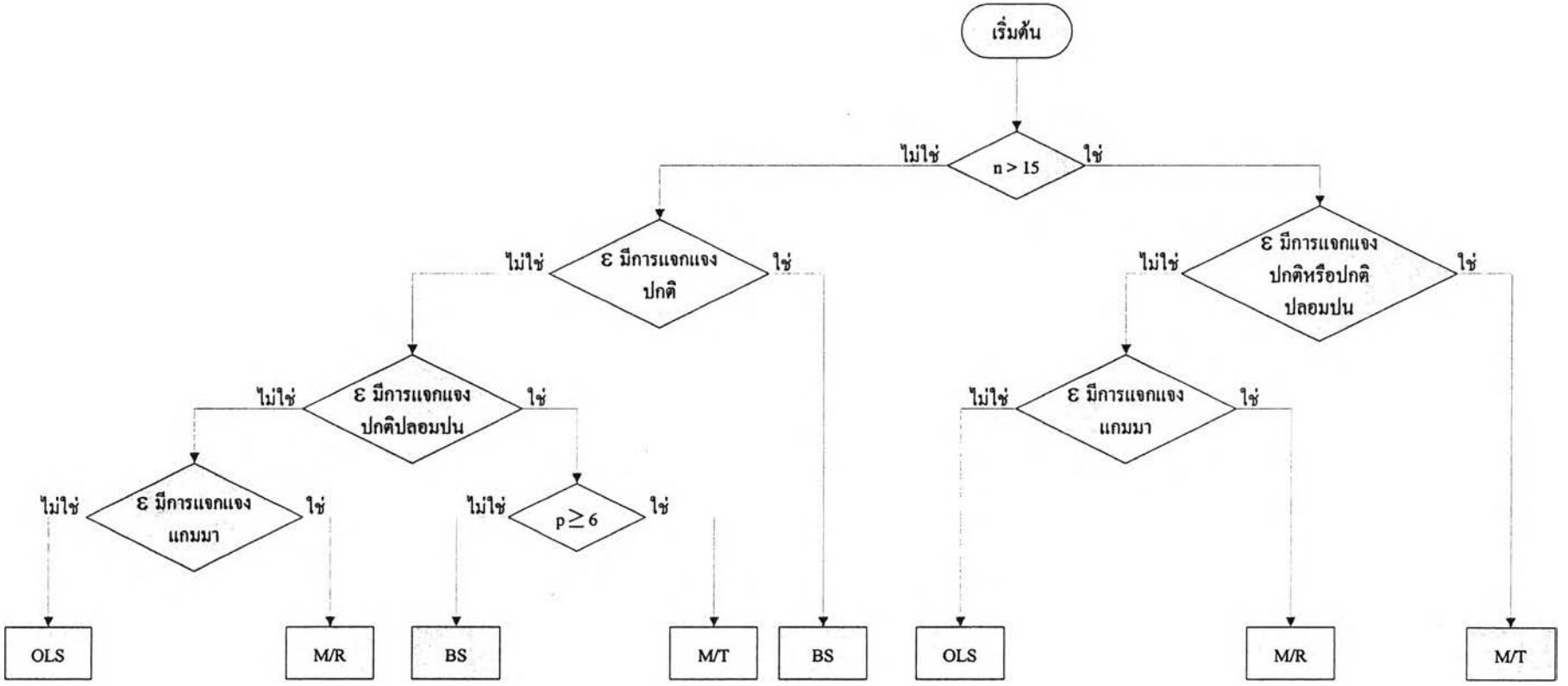
ในงานวิจัยนี้ได้ผลสรุปการเลือกวิธีการประมาณค่าสัมประสิทธิ์การถดถอยเชิงเส้นพหุคูณเมื่อตัวอย่างมีขนาดเล็กว่าควรพิจารณาจากลักษณะการแจกแจงของค่าความคลาดเคลื่อนก่อนเพราะจากผลการวิจัยจะได้ว่าวิธี BS จะเป็นวิธีที่ดีที่สุดเมื่อค่าความเคลื่อนมีการแจกแจงปกติหรือมีการแจกแจงปกติปลอมปน (การแจกแจงสมมาตรหางยาว) เมื่อตัวอย่างมีขนาดเล็ก (อยู่ในช่วงไม่เกิน 20) วิธี M/T จะเป็นวิธีที่ดีที่สุดเมื่อค่าความเคลื่อนมีการแจกแจงปกติหรือมีการแจกแจงปกติปลอมปน (การแจกแจงสมมาตร) โดยเฉพาะเมื่อมีค่าเฉลี่ยไม่เท่ากับ 0 มีค่าผิดปกติมากและมีขนาดตัวอย่างมาก (อยู่ในช่วงมากกว่า 15 ขึ้นไปหรือมากกว่า 20 ในบางกรณี) และวิธี M/R จะเป็นวิธีที่ดีที่สุดเมื่อค่าความเคลื่อนมีการแจกแจงแกมมา (การแจกแจงเบ้ขวา) ทุกระดับค่าพารามิเตอร์แอลฟาและทุกระดับขนาดตัวอย่าง ผลสรุปดังกล่าวอาจเขียนในรูปของผังงานได้ดังนี้

รูปที่ 5.1 แสดงขั้นตอนสรุปเลือกวิธีการประมาณค่าสัมประสิทธิ์การถดถอยเชิงเส้นพหุคูณเมื่อตัวอย่างมีขนาดเล็กในทางทฤษฎี



หมายเหตุ 1. จำนวนตัวแปรอิสระต้องน้อยกว่าขนาดตัวอย่าง ($p < n$)

รูปที่ 5.2 แสดงขั้นตอนสรุปเลือกวิธีการประมาณค่าสัมประสิทธิ์การถดถอยเชิงเส้นพหุคูณเมื่อตัวอย่างมีขนาดเล็กในทางปฏิบัติ



หมายเหตุ 1. จำนวนตัวแปรอิสระต้องน้อยกว่าขนาดตัวอย่าง ($p < n$)

2. ในทางปฏิบัติกรณีที่ขนาดตัวอย่างมากกว่า 15 การเลือกใช้ตัวประมาณ M อาจพิจารณาจากค่าผิดปกติได้ กล่าวคือถ้าพบว่ามีค่าผิดปกติที่รุนแรงมากให้ใช้เกณฑ์ความแกร่งของแรมเซย์ แต่ถ้าพบว่ามีค่าผิดปกติเกิดขึ้นจำนวนมากให้ใช้เกณฑ์ความแกร่งของตูกี

5.4 ข้อเสนอแนะ

การประมาณค่าสัมประสิทธิ์การถดถอยเชิงเส้นพหุคูณเมื่อตัวอย่างมีขนาดเล็กในแต่ละสถานการณ์ ควรเลือกใช้วิธีการประมาณค่าดังนี้

1. กรณีที่ค่าความคลาดเคลื่อนเป็นไปตามข้อสมมติของการวิเคราะห์ความถดถอยกล่าวคือมีการแจกแจงปกติที่เป็นอิสระซึ่งกันและกันด้วยค่าเฉลี่ยเท่ากับศูนย์และมีความแปรปรวนคงที่ ควรเลือกใช้วิธีบูตสแตรป แต่เมื่อขนาดตัวอย่างมากควรเลือกใช้วิธีตัวประมาณ M โดยใช้เกณฑ์ความแกร่งของตูกีจะได้ตัวประมาณที่มีประสิทธิภาพดีขึ้น
2. กรณีที่ค่าความคลาดเคลื่อนมีการแจกแจงหางยาวกว่าการแจกแจงปกติและมีจำนวนตัวแปรอิสระไม่มากนักวิธีบูตสแตรปยังคงเป็นวิธีที่มีประสิทธิภาพแต่เมื่อค่าเฉลี่ยไม่เท่ากับศูนย์ (กล่าวคือไม่เป็นไปตามเงื่อนไขของการวิเคราะห์ความถดถอย) วิธีตัวประมาณ M โดยใช้เกณฑ์ความแกร่งของตูกีจะมีประสิทธิภาพดีกว่า
3. กรณีที่ค่าความคลาดเคลื่อนมีการแจกแจงเบ้ขวาควรเลือกใช้วิธีตัวประมาณ M โดยใช้เกณฑ์ความแกร่งของแรมเซย์ซึ่งกรณีนี้สามารถใช้ได้ในทุกระดับขนาดตัวอย่าง

ในทางปฏิบัตินั้นพบว่าการจะกำหนดว่าขนาดตัวอย่างเท่าใดจึงจะเรียกว่าขนาดตัวอย่างเล็ก (small sample size) หรือขนาดตัวอย่างใหญ่ (large sample size) นั้นไม่สามารถกำหนดได้โดยง่าย ซึ่งมีนักสถิติหลายท่านที่ได้ศึกษาถึงขนาดตัวอย่างที่เหมาะสมกับการวิเคราะห์ความถดถอยเชิงเส้นพหุคูณ เช่นในปี พ.ศ.2534 กรีน (Green) ได้กำหนดสูตรในการคำนวณขนาดตัวอย่างเท่ากับ $n \geq 50 + 8p$ เมื่อ p คือจำนวนตัวแปรอิสระ และในปีพ.ศ.2538 เรคอฟ (Raykov) กล่าวไว้ว่าขนาดตัวอย่างควรมีมากกว่าจำนวนตัวแปรอิสระอย่างน้อย 5 เท่า เป็นต้น ซึ่งผู้สนใจอาจใช้เกณฑ์ดังกล่าวเป็นการกำหนดขนาดตัวอย่างได้

การวิจัยครั้งนี้ผู้วิจัยศึกษาถึงวิธีประมาณค่าสัมประสิทธิ์การถดถอยเชิงเส้นพหุคูณเมื่อตัวอย่างมีขนาดเล็กในกรณีที่ตัวแปรอิสระแต่ละตัวเป็นข้อมูลเชิงปริมาณ (quantitative data) และเป็นอิสระซึ่งกันและกัน ผู้ที่สนใจสามารถศึกษาในกรณีที่ตัวแปรอิสระมีความสัมพันธ์กันหรือมีตัวแปรอิสระบางตัวที่เป็นข้อมูลเชิงคุณภาพ (Qualitative data) ได้ ซึ่งมีวิธีการประมาณค่าบางวิธีเช่น วิธีการประมาณค่าแบบริดจ์ (Ridge estimation) หรือวิธีวิเคราะห์องค์ประกอบหลัก (Principle component analysis) เป็นวิธีที่เหมาะสำหรับกรณีที่ตัวแปรอิสระมีความสัมพันธ์กัน เป็นที่น่าสนใจว่าเมื่อตัวอย่างมีขนาดเล็ก วิธีเหล่านี้จะมีประสิทธิภาพมากหรือน้อยเพียงใดหรือจะมีวิธีการประมาณวิธีอื่นอีกหรือไม่ที่มีประสิทธิภาพมากกว่า ในกรณีที่มีค่าผิดปกติ ผู้ที่สนใจยังสามารถศึกษาในกรณีที่มีค่าผิดปกติในตัวแปรอิสระหรือในตัวแปรตามได้ ซึ่งในงานวิจัยนี้ศึกษาเฉพาะในกรณีที่มีค่าผิดปกติในค่าความคลาดเคลื่อนเท่านั้น ยังมีตัวประมาณที่มีความแกร่งตัวอื่นๆที่อาจมีประสิทธิภาพในกรณีดังกล่าวได้