บทที่ 5

ผลสรุปการวิเคราะห์และการอภิปรายผล

การวิจัยนี้ได้เสนอวิธีการประมาณค่าสัมประสิทธิ์การถดถอยพหุเมื่อค่าผิดพลาดมีการแจกแจงแบบหางยาวกว่าการแจกแจงแบบปกติและการแจกแจงแบบเบ้ โดยทำการศึกษาเปรียบเทียบวิธีกำลังสองน้อยที่สุดกับวิธี M-estimator ซึ่งใช้เกณฑ์ความแกร่งของ Ramsay ด้วยการพิจารณาค่า MSE ที่คำนวณได้ในแต่ละวิธีการ และใช้เกณฑ์ค่าเฉลี่ยของค่าสัมพัทธ์ของค่าเฉลี่ยความผิดพลาดกำลังสอง (ARMSE) โดยใช้สูตรที่ 1 และ 2 ในแต่ละกรณี และค่าเฉลี่ยของค่าสัมบูรณ์ของค่าแตกต่างของอัตราส่วนค่าเฉลี่ยความผิดพลาดกำลังสอง (AADRM) โดยใช้สูตรที่ 3 และ 4 ในแต่ละกรณี

สูตรที่ 1 กรณีที่ MSE_{OLS} น้อยกว่า MSE_M

$$ARMSE = \frac{MSE_M}{MSE_{OLS}} \times 100$$

สูตรที่ 2 กรณีที่ MSE_M น้อยกว่า MSE_{OLS}

$$ARMSE = \frac{MSE_{OLS}}{MSE_M} \times 100$$

สูตรที่ 3 กรณีที่ MSE_{OLS} น้อยกว่า MSE_M

$$AADRM = \frac{|MSE_{OLS} - MSE_M|}{MSE_M}$$

สูตรที่ 4 กรณีที่ MSE_M น้อยกว่า MSE_{OLS}

$$AADRM = \frac{|MSE_M - MSE_{OLS}|}{MSE_{OLS}}$$

ซึ่งสามารถสรุปผลการวิเคราะห์ได้ดังนี้

5.1 ผลสรุปการเปรียบเทียบวิธีกำลังสองน้อยที่สุดกับวิธี M-estimator

จากการคำนวณหาค่าสัมประสิทธิ์การถดถอยพหุ ด้วยวิธีกำลังสองน้อยที่สุด และวิธี M-estimator ซึ่งใช้เกณฑ์ความแกร่งของ Ramsay และนำค่า MSE ของทั้งสองวิธีมาเปรียบเทียบกันโดยใช้เกณฑ์ของค่าเฉลี่ยของค่าสัมพัทธ์ของค่าเฉลี่ยความผิดพลาดกำลังสอง (ARMSE) และค่าเฉลี่ยของค่าสัมบูรณ์ของค่าแตกต่างของอัตราส่วนของค่าเฉลี่ยความผิดพลาดกำลังสอง (AADRM)

5.1.1 เมื่อค่าผิดพลาดมีการแจกแจงแบบหางยาวกว่าการแจกแจงแบบปกติ ได้ใช้การแจกแจงแบบปกติปลอมปนและแบบทีเป็นกรณีศึกษาการแจกแจงของค่าผิดพลาดที่มีค่าผิดปกติเกิดขึ้นในรูปแบบต่าง ๆ สำหรับการแจกแจงแบบปกติปลอมปนมีทั้งแบบหางยาวเล็กน้อยจนถึงแบบหางยาวมาก ๆ ซึ่งสามารถกำหนดจากสเกลแฟคเตอร์และเปอร์เซนต์การปลอมปน สำหรับการแจกแจงแบบทีมีทั้งแบบที่หางยาวเล็กน้อยและหางยาวมาก ซึ่งกำหนดจากระดับความเป็นอิสระ จะได้ว่ารูปแบบของการแจกแจงของค่าผิดพลาดมีอิทธิพลอย่างมากต่อความสามารถในการประมาณสัมประสิทธิ์ถดถอยพหุ เมื่อค่าผิดปกติเพิ่มมากขึ้น วิธี M-estimator ซึ่งใช้เกณฑ์ความแกร่งของ Ramsay จะประมาณสัมประสิทธิ์การถดถอยพหุได้ดีกว่าวิธีกำลังสองน้อยที่สุด โดยใช้เกณฑ์การเปรียบเทียบค่าเฉลี่ยของค่าสัมพัทธ์ของค่าเฉลี่ยความผิดพลาดกำลังสอง (ARMSE) และค่าเฉลี่ยของค่าสัมบูรณ์ของค่าแตกต่างของอัตราส่วนค่าเฉลี่ยความผิดพลาดกำลังสอง (AADRM) ซึ่งสรุปผลได้ดังนี้

1) เมื่อการแจกแจงเป็นแบบปกติปลอมปน วิธี M-estimator ซึ่งใช้เกณฑ์ความแกร่งของ Ramsay จะประมาณสัมประสิทธิ์การถดถอยพหุได้ดีกว่าวิธีกำลังสองน้อยที่สุดมาก โดยเรียงลำดับของอิทธิพลของค่าผิดปกติที่เกิดขึ้นจากมากไปน้อย ดังนี้

1.1 สเกลแฟคเตอร์คือ ณ สเกลแฟคเตอร์ 10 และ 3

1.2 เปอร์เซนต์การปลอมปน คือ ณ เปอร์เซนต์การปลอมปน 25 10 5 และ 1

1.3 จำนวนตัวแปรอิสระและขนาดตัวอย่าง คือ จำนวนตัวแปรอิสระ 3 และขนาดตัวอย่าง 20 จำนวนตัวแปรอิสระ 5 และขนาดตัวอย่าง 150 100 และ 50 จำนวนตัวแปรอิสระ 10 และขนาดตัวอย่าง 150 100 และ 50

2) เมื่อการแจกแจงเป็นแบบที วิธี M-estimator ซึ่งใช้เกณฑ์ความแกร่งของ Ramsay จะประมาณสัมประสิทธิ์การถดถอยพหุได้ดีกว่าวิธีกำลังสองน้อยที่สุดมาก โดยเรียงลำดับของอิทธิพลของค่าผิดปกติที่เกิดขึ้นจากมากไปน้อย ดังนี้

1.1 ระดับความเป็นอิสระ คือ ระดับความเป็นอิสระ 4 และ 8

1.2 จำนวนตัวแปรอิสระและขนาดตัวอย่าง คือ จำนวนตัวแปรอิสระ 3 และขนาดตัวอย่าง 20 จำนวนตัวแปรอิสระ 5 และขนาดตัวอย่าง 50 100 และ 150 จำนวนตัวแปรอิสระ 10 และขนาดตัวอย่าง 50 100 และ 150

สำหรับจำนวนตัวแปรอิสระและขนาดตัวอย่างมีอิทธิพลไม่มากนักในการเลือกใช้วิธีการประมาณสัมประสิทธิ์การถดถอยพหุ เพราะว่าวิธีกำลังสองน้อยที่สุดและวิธี M-estimator จะพิจารณาถึงค่าผิดพลาดที่เกิดขึ้นในตัวแบบเชิงเส้น และเราได้ทดสอบความเป็นอิสระของตัวแปรอิสระเหล่านั้นแล้ว

5.1.2 เมื่อค่าผิดพลาดมีการแจกแจงแบบเบ้ ได้ใช้การแจกแจงแบบลอกนอร์มอล แกมมา และไวบูลล์ เป็นกรณีศึกษา

5.1.2.1 เมื่อค่าผิดพลาดมีการแจกแจงแบบลอกนอร์มอล ซึ่งมักเกิดขึ้นในการวิเคราะห์งานทางด้านเศรษฐศาสตร์ เช่น การวิเคราะห์การกระจายรายได้ เป็นต้น สำหรับการวิจัยครั้งนี้ได้ใช้การแจกแจงแบบลอกนอร์มอลที่มีค่าเฉลี่ย (μ) = 0 และค่าความแปรปรวน (σ^2) = 1 ในการประมาณสัมประสิทธิ์การถดถอยพหุเมื่อค่าผิดพลาดมีการแจกแจงแบบเบ้นั้นจะอาศัยเทคนิคในการแปลงข้อมูลของ Box และ Cox เพื่อแปลงข้อมูลให้เข้าสู่ภาวะปกติเสียก่อน จะได้ว่าเมื่อใช้เกณฑ์การเปรียบเทียบค่าเฉลี่ยของค่าสัมพัทธ์ของค่าเฉลี่ยความผิดพลาดกำลังสอง และค่าเฉลี่ยของค่าสัมบูรณ์ของค่าแตกต่างของอัตราส่วนค่าเฉลี่ยความผิดพลาดกำลังสองแล้ว วิธีกำลังสองน้อยที่สุดจะประมาณสัมประสิทธิ์การถดถอยพหุได้ใกล้เคียงกับวิธี M-estimator เมื่อใช้จำนวนตัวแปรอิสระและขนาดตัวอย่างแบบต่าง ๆ

5.1.2.2 เมื่อค่าผิดพลาดมีการแจกแจงแบบแกมมา ซึ่งมักเกิดขึ้นในการวิเคราะห์งานทางด้านเศรษฐศาสตร์ เช่น การวิเคราะห์การกระจายรายได้ เป็นต้นในการประมาณสัมประสิทธิ์การถดถอยพหุเมื่อค่าผิดพลาดมีการแจกแจงแบบเบ้นั้นจะอาศัยเทคนิคในการแปลงข้อมูล

ของ Box และ Cox เพื่อแปลงข้อมูลให้เข้าสู่ภาวะปกติเสียก่อน ผลปรากฏว่าเมื่อใช้เกณฑ์การเปรียบเทียบค่าเฉลี่ยของค่าสัมพัทธ์ของค่าเฉลี่ยความผิดพลาดกำลังสองและค่าเฉลี่ยของค่าสัมบูรณ์ของค่าแตกต่างของอัตราส่วนค่าเฉลี่ยความผิดพลาดกำลังสองแล้ว วิธีกำลังสองน้อยที่สุดจะประมาณสัมประสิทธิ์การถดถอยพหุได้ใกล้เคียงกับวิธี M-estimator เมื่อใช้สัมประสิทธิ์ของความแปรปรวน CV = 100% (β=1 , α=1), CV = 70% (β = 1 , α=2), CV = 58% (β=1, α=3) และ CV = 17% (β=150,α=10) และเมื่อใช้จำนวนตัวแปรอิสระและขนาดตัวอย่างแบบต่าง ๆ

5.1.2.3 เมื่อค่าผิดพลาดมีการแจกแจงแบบไวบูลล์ ซึ่งมักเกิดขึ้นในงานทางด้านวิศวกรรม เช่น เป็นตัวแบบสำหรับการกระจายของการเกิดข้อบกพร่อง (time to failure) ของเครื่องจักรหรืออุปกรณ์ไฟฟ้า ในการประมาณสัมประสิทธิ์การถดถอยพหุเมื่อค่าผิดพลาดมีการแจกแจงแบบเบ้นั้นจะอาศัยเทคนิคในการแปลงข้อมูลของ Box และ Cox เพื่อแปลงข้อมูลให้เข้าสู่ภาวะปกติเสียก่อน ผลปรากฏว่าเมื่อใช้เกณฑ์การเปรียบเทียบค่าเฉลี่ยของค่าสัมพัทธ์ของค่าเฉลี่ยความผิดพลาดกำลังสอง และค่าเฉลี่ยของค่าสัมบูรณ์ของค่าแตกต่างของอัตราส่วนค่าเฉลี่ยความผิดพลาดกำลังสองแล้ว วิธีกำลังสองน้อยที่สุดจะประมาณสัมประสิทธิ์การถดถอยพหุได้ใกล้เคียงกันเมื่อใช้สัมประสิทธิ์ของความแปรปรวน CV = 100% (β=1 , α=1), CV = 52% (β=1, α=2), CV = 35% (β = 1, α=3) และ CV = 17% (β=150,α=10) และเมื่อใช้จำนวนตัวแปรอิสระและขนาดตัวอย่างแบบต่าง ๆ

5.2 การอภิปรายผล

จากผลการวิเคราะห์ครั้งนี้ เมื่อพิจารณาถึง MSE ของวิธีกำลังสองน้อยที่สุด และวิธี M-estimator เมื่อค่าผิดพลาดมีการแจกแจงแบบหางยาวกว่าการแจกแจงแบบปกติโดยใช้การแจกแจงแบบปกติปลอมปนและแบบทีเป็นกรณีศึกษา ผลปรากฏว่า การแจกแจงของค่าผิดพลาดที่มีค่าผิดปกติเกิดขึ้นในรูปแบบต่าง ๆ มีทั้งแบบหางยาวเล็กน้อย และแบบหางยาวมาก ๆ ดังกล่าวค่อนข้างมีอิทธิพลต่อการประมาณสัมประสิทธิ์การถดถอยพหุมากกว่าขนาดตัวอย่างและจำนวนตัวแปรอิสระ และเมื่อพิจารณาถึง MSE ของวิธีกำลังสองน้อยที่สุดและวิธี M-estimator เมื่อค่าผิดพลาดมีการแจกแจงแบบเบ้ โดยใช้การแจกแจงแบบลอกนอร์มอล แกมมา และไวบูลล์ ผลปรากฏว่าการแจกแจงที่มีรูปแบบเบ้ดังกล่าว โดยอาศัยเทคนิคการแปลงที่ใช้การยกกำลังของ Box และ Cox

สำหรับการแปลงข้อมูลให้เข้าสู่ภาวะปกติแล้ว วิธีกำลังสองน้อยที่สุดและวิธี M-estimator สามารถใช้ในการประมาณค่าสัมประสิทธิ์การถดถอยพหุได้ใกล้เคียงกัน เมื่อใช้เกณฑ์การเปรียบเทียบค่าเฉลี่ยของค่าสัมพัทธ์ของค่าเฉลี่ยความผิดพลาดกำลังสอง (ARMSE) และค่าเฉลี่ยของค่าสัมบูรณ์ของค่าแตกต่างของอัตราส่วนของค่าเฉลี่ยความผิดพลาดกำลังสอง

5.3 ข้อเสนอแนะ

ในการประมาณค่าสัมประสิทธิ์การถดถอยพหุ เมื่อค่าผิดพลาดมีการแจกแจงแบบหางยาวกว่าการแจกแจงแบบปกติ และการแจกแจงของค่าผิดพลาดที่มีค่าผิดปกติเกิดขึ้นมาก ควรเลือกใช้วิธี M-estimator ซึ่งใช้เกณฑ์ความแกร่งของ Ramsay เนื่องจากวิธีนี้สามารถลดอิทธิพลของค่าผิดปกติลงได้ นอกจากวิธี M-estimator ซึ่งใช้เกณฑ์ความแกร่งของ Ramsay แล้ว ยังสามารถศึกษาวิธี M-estimator ซึ่งใช้เกณฑ์ความแกร่งของ Huber ซึ่งพยายามลดอิทธิพลของค่าผิดปกติเมื่อค่าผิดพลาดนั้นเกินขอบเขตที่กำหนด สำหรับวิธีของ Andrew ควรใช้กับข้อมูลที่มีค่าผิดปกติมาก และวิธีของ Hampel ควรใช้สำหรับข้อมูลที่มีความผิดปกติปานกลางโดยลดอิทธิพลลงเป็นช่วง ๆ และยังสามารถศึกษาค่าผิดพลาดที่มีการแจกแจงเป็นแบบอื่น ได้แก่ แบบดับเบิ้ลเอ็กซ์โปเนนเชียล เป็นต้น

ในการประมาณค่าสัมประสิทธิ์การถดถอยพหุเมื่อค่าผิดพลาดมีการแจกแจงแบบเบ้ สำหรับวิธีนี้ควรใช้เทคนิคการแปลงที่ใช้การยกกำลังของ Box และ Cox ในการแปลงข้อมูลให้เข้าสู่ภาวะปกติเสียก่อน แล้วทำการตรวจสอบว่าข้อมูลหลังจากการแปลงนั้นมีการแจกแจงเป็นแบบปกติหรือไม่ ถ้ามีการแจกแจงเป็นแบบปกติจึงเลือกใช้ วิธีกำลังสองน้อยที่สุดหรือวิธี M-estimator ซึ่งจะให้ผลลัพธ์ใกล้เคียงกัน แต่วิธีกำลังสองน้อยที่สุดจะทำได้สะดวกและรวดเร็วกว่า

การวิจัยครั้งนี้ศึกษากรณีข้อมูลมีลักษณะเป็นเชิงปริมาณ (Quantitative data) และตัวแปรอิสระในตัวแบบเชิงเส้นเป็นแบบคงที่ (fixed effect model) ผู้ที่สนใจอาจจะศึกษาการประมาณค่าสัมประสิทธิ์การถดถอยพหุสามารถศึกษาในกรณีที่ข้อมูลมีลักษณะเป็นเชิงคุณภาพ (Qualitative data) และตัวแปรอิสระในตัวแบบเชิงเส้นแบบสุ่ม (Random effect model)

การศึกษาเปรียบเทียบวิธีการประมาณสัมประสิทธิ์การถดถอยพหุระหว่างวิธีกำลังสองน้อยที่สุดกับวิธี M-estimator อาจจะใช้เกณฑ์การเปรียบเทียบค่าเฉลี่ยของค่าเบี่ยงเบนของค่า

ประมาณต่อค่าจริงของค่าสัมประสิทธิ์การถดถอย โดยเขียนสูตรดังนี้คือ $\frac{1}{m}\sum_{i=1}^{m}\left|\frac{\hat{\beta}_i - \beta_i}{\hat{\beta}}\right|$ ซึ่งผู้วิจัยได้ทดลองศึกษาจากค่ารากที่สองของค่าเฉลี่ยความผิดพลาดกำลังสองต่อค่าเฉลี่ยของค่าประมาณ = $\frac{\sqrt{MSE}}{\bar{\hat{\beta}}}$ สำหรับกรณีที่ค่าผิดพลาดมีการแจกแจงแบบปกติปลอมปน ณ สเกล-แฟคเตอร์ 3 และ 10 ใช้เปอร์เซนต์การปลอมปนเท่ากับ 1 5 10 และ 25 เมื่อจำนวนตัวแปรอิสระเท่ากับ5 และใช้ขนาดตัวอย่างเท่ากับ 50 ผลการทดลองเป็นดังนี้คือ ณ สเกลแฟคเตอร์เท่ากับ 3 ในกรณีที่ค่าเฉลี่ยความผิดพลาดกำลังสองของวิธีกำลังสองน้อยที่สุดน้อยกว่าของวิธี M-estimator ที่เปอร์เซนต์การปลอมปนเท่ากับ 1 5 10 และ 25 ได้ค่า $\frac{\sqrt{MSE_{OLS}}}{\bar{\hat{\beta}}}$ เท่ากับ 0.71 0.72 0.75 และ 0.86 ตามลำดับ ส่วนกรณีที่ค่าเฉลี่ยความผิดพลาดกำลังสองของวิธี M-estimator น้อยกว่าวิธีกำลังสองน้อยที่สุด ได้ค่า $\frac{\sqrt{MSE_M}}{\bar{\hat{\beta}}}$ เท่ากับ 0.70 0.71 0.73 และ 0.79 ตามลำดับ สำหรับกรณีที่ใช้สเกลแฟคเตอร์เท่ากับ 10 ในกรณีที่ค่าเฉลี่ยความผิดพลาดกำลังสองของวิธีกำลังสองน้อยที่สุดน้อยกว่าของวิธี M-estimator ใช้เปอร์เซนต์การปลอมปนเท่ากับ 1 5 10 และ 25 ได้ค่า $\frac{\sqrt{MSE_{OLS}}}{\bar{\hat{\beta}}}$ เท่ากับ 0.74 0.85 0.92 และ 1.22 ตามลำดับ ส่วนกรณีที่ค่าเฉลี่ยความผิดพลาดกำลังสองของวิธี M-estimator น้อยกว่าวิธีกำลังสองน้อยที่สุด ได้ค่า $\frac{\sqrt{MSE}}{\bar{\hat{\beta}}}$ เท่ากับ 0.74 0.78 0.86 และ 1.04 ตามลำดับ

ศูนย์วิทยทรัพยากร
จุฬาลงกรณ์มหาวิทยาลัย