บทที่ 5

สรุปผลการวิจัย อภิปรายผลและข้อเสนอแนะ

การวิจัยครั้งนี้เป็นการวิจัยเชิงทดลอง เพื่อศึกษาเปรียบเทียบตัวประมาณริดจ์ 3 วิธี คือวิธี HKB TZE-SAN-LEE และ McD&G ตามลำดับ ในสถานการณ์ต่างๆ ดังนี้

1) การแจกแจงแบบปกติ แบบปกติปลอมปน แบบลอกนอร์มอล และแบบไวบูลล์

2) ขนาดตัวอย่างคือ 10,30,50 และ 100

3) ระดับความสัมพันธ์ .99,.90,.70 เมื่อใช้ตัวแปรอิสระ = 3 และระดับความสัมพันธ์ (.99,.99), (.99,.90) และ (.70,.30) เมื่อใช้ตัวแปรอิสระ = 5

การวิจัยครั้งนี้จำลองข้อมูลโดยเครื่อง IBM 4361 เพื่อสร้างข้อมูลและสถานการณ์ต่างๆ ที่กำหนดโดยการทำซ้ำ 200 ครั้งในแต่ละสถานการณ์

5.1 สรุปผลการวิจัย

1) ความคลาดเคลื่อนมีการแจกแจงแบบปกติ

กรณีที่ตัวแปรอิสระ = 3 6^2 = 0.03 และ 1.0

สำหรับ 6^2 = 0.03 โดยทั่วๆไป วิธี McD&G และ HKB ให้ผลดี แต่เมื่อระดับความสัมพันธ์เพิ่มขึ้น วิธี McD&G ให้ผลดี และวิธี McD&G จะให้ผลดีลดลงเมื่อขนาดตัวอย่างเพิ่มขึ้น สำหรับกรณีที่ 6^2 เพิ่มขึ้นเป็น 1.0 โดยทั่วไป วิธี TZE-SAN-LEE ให้ผลดี และเมื่อระดับความสัมพันธ์เพิ่มขึ้นยังคงได้ผลสรุปเหมือนเดิม ยกเว้นที่ระดับความสัมพันธ์ .99 เมื่อขนาดตัวอย่าง = 10 วิธี HKB ให้ผลดี และทั้ง 3 วิธีจะให้ค่า AMSE ใกล้เคียงกัน

กรณีที่ตัวแปรอิสระ = 5 σ^2 = 0.03 และ 1.0

สำหรับความแปรปรวน = 0.03 โดยทั่วๆไปวิธี McD&G ให้ผลดี เมื่อระดับความสัมพันธ์เพิ่มขึ้นยังคงได้ผลสรุปเหมือนเดิม แต่เมื่อขนาดตัวอย่างเพิ่มขึ้นเป็น 50 และ 100 วิธี TZE-SAN-LEE จะให้ผลดี ในกรณีที่ความแปรปรวนเพิ่มขึ้นเป็น 1.0 โดยทั่วๆไปวิธี HKB ให้ผลดี เมื่อระดับความสัมพันธ์เพิ่มขึ้นยังคงได้ผลสรุปเหมือนเดิม แต่เมื่อขนาดตัวอย่างเพิ่มขึ้น วิธี HKB มีแนวโน้มให้ผลดีลดลง

การเพิ่มจำนวนตัวแปรอิสระจะทำให้ผลสรุปเปลี่ยนไป กล่าวคือ กรณีที่ตัวแปรอิสระ = 3 โดยทั่วๆไปเมื่อระดับความสัมพันธ์เพิ่มขึ้นวิธี McD&G และ TZE-SAN-LEE ให้ผลดี เมื่อ σ^2 = 0.03 และ 1.0 ตามลำดับ แต่เมื่อเพิ่มตัวแปรอิสระเป็น 5 โดยทั่วๆไปเมื่อระดับความสัมพันธ์เพิ่มขึ้น วิธี TZE-SAN-LEE และ HKB ให้ผลดี เมื่อ σ^2 = 0.03 และ 1.0 ตามลำดับ ดังนั้นจำนวนตัวแปรอิสระ ระดับความสัมพันธ์ ความแปรปรวนและขนาดตัวอย่างจึงมีผลต่อทั้ง 3 วิธี โดยที่จำนวนตัวแปรอิสระมีผลมากที่สุด*

---

*เนื่องจากทั้ง 3 วิธี คือ วิธี HKB, TZE-SAN-LEE และ McD&G มีสูตรการคำนวณหาค่า k เพื่อนำมาบวกกับสมาชิกบนเส้นทแยงมุมของเมตริกซ์ $X'X$ แตกต่างกัน กล่าวคือ ค่า k โดยวิธี HKB มาจากสัดส่วนโดยตรงของจำนวนตัวแปรอิสระ คือ $k = p\hat{\sigma}^2/\hat{\beta}'\hat{\beta}$ ค่า k โดยวิธี TZE-SAN-LEE จะพิจารณาจาก ค่า eigen value ที่น้อยที่สุด และค่า k โดยวิธี McD&G คำนวณจากสมการ $\underline{g}(k)'\underline{g}(k) = \hat{\beta}'\hat{\beta} - \hat{\sigma}^2 \text{ trace}(X'X)^{-1}$ ดังนั้นจำนวนตัวแปรอิสระมีผลต่อทั้ง 3 วิธี เรียงตามลำดับจากมากไปน้อย คือ วิธี HKB McD&G และ TZE-SAN-LEE

2) ความคลาดเคลื่อนมีการแจกแจงแบบปกติปลอมปน

กรณีที่ตัวแปรอิสระ = 3 เปอร์เซนต์การปลอมปน(p) = 5,10 สเกลเฟกเตอร์(c) = 3,10

โดยทั่วไปวิธี TZE-SAN-LEE ให้ผลดี เมื่อระดับความสัมพันธ์เพิ่มขึ้นวิธี McD&G มีแนวโน้มดีขึ้นสำหรับขนาดตัวอย่าง = 10 สำหรับขนาดตัวอย่าง = 30 วิธี HKB มีแนวโน้มดีขึ้น เมื่อขนาดตัวอย่าง = 50 และ 100 วิธี TZE-SAN-LEE ให้ผลดี สำหรับการเพิ่มเปอร์เซนต์การปลอมปนและสเกลเฟกเตอร์ โดยทั่วๆไปยังคงให้ผลสรุปเหมือนเดิม แต่จะทำให้ค่า AMSE มีแนวโน้มเพิ่มขึ้น

กรณีที่ตัวแปรอิสระ = 5 เปอร์เซนต์การปลอมปน(p) = 5,10 สเกลเฟกเตอร์(c) = 3,10

โดยกรณีทั่วไปวิธี HKB ให้ผลดี เมื่อระดับความสัมพันธ์เพิ่มขึ้นวิธี HKB มีแนวโน้มดีขึ้น และเมื่อขนาดตัวอย่างเพิ่มขึ้นยังคงได้ผลสรุปเหมือนเดิม ยกเว้นที่ระดับความสัมพันธ์ (.70,.30) วิธี TZE-SAN-LEE มีแนวโน้มดีขึ้น สำหรับการเพิ่มเปอร์เซนต์การปลอมปน และสเกลแฟกเตอร์ โดยทั่วๆไปยังคงได้ผลสรุปเหมือนเดิม แต่จะทำให้ค่า AMSE เพิ่มมากขึ้น และวิธี HKB มีแนวโน้มดีขึ้น

การเพิ่มตัวแปรอิสระจะทำให้ผลสรุปเปลี่ยนไป กล่าวคือ กรณีที่ตัวแปรอิสระ = 3 เมื่อระดับความสัมพันธ์เพิ่มขึ้น วิธี McD&G ให้ผลดีสำหรับขนาดตัวอย่าง = 10 และจะให้ค่า AMSE ใกล้เคียงกับวิธี TZE-SAN-LEE เมื่อขนาดตัวอย่าง = 30 วิธี HKB ดีและจะให้ค่า AMSE ใกล้เคียงกับวิธี TZE-SAN-LEE เมื่อขนาดตัวอย่าง = 50 และ 100 วิธี TZE-SAN-LEE ให้ผลดี การเพิ่มขนาดตัวอย่างจะทำให้วิธี TZE-SAN-LEE มีแนวโน้มดีขึ้น สำหรับกรณีที่เพิ่มตัวแปรอิสระเป็น 5 เมื่อระดับความสัมพันธ์เพิ่มขึ้น วิธี HKB ให้ผลดี และเมื่อเปอร์เซนต์การปลอมปนและสเกลแฟกเตอร์เพิ่มขึ้น วิธี HKB มีแนวโน้มดีขึ้น ดังนั้นจำนวนตัวแปรอิสระ ระดับความสัมพันธ์ ขนาดตัวอย่าง เปอร์เซนต์การปลอมปน และสเกลแฟกเตอร์จะมีผลต่อทั้ง 3 วิธี โดยที่จำนวนตัวแปรอิสระมีผลมากที่สุด

3. ความคลาดเคลื่อนมีการแจกแจงแบบลอกนอร์มอล

กรณีตัวแปรอิสระ = 3 σ^2 = .10,.30,.70

โดยทั่วไปวิธี TZE-SAN-LEE ให้ผลดี เมื่อระดับความสัมพันธ์เพิ่มขึ้นยังได้ผลสรุปเหมือนเดิม ยกเว้นที่ระดับความสัมพันธ์ .99 และขนาดตัวอย่าง = 10 วิธี McD&G ให้ผลดี เมื่อขนาดตัวอย่างและ σ^2 เพิ่มขึ้นยังคงให้ผลสรุปเหมือนเดิม ทั้ง 3 วิธี จะให้ค่า AMSE ใกล้เคียงกันเมื่อ σ^2 = .10 ค่า AMSE มีแนวโน้มเพิ่มขึ้นเมื่อระดับความสัมพันธ์ และ σ^2 เพิ่มขึ้น และจะมีแนวโน้มลดลงเมื่อขนาดตัวอย่างเพิ่มขึ้น

กรณีตัวแปรอิสระ = 5 σ^2 = .10,.30,.70

โดยทั่วไปวิธี HKB ให้ผลดี เมื่อระดับความสัมพันธ์เพิ่มขึ้นยังคงให้ผลสรุปเหมือนเดิม ยกเว้นที่ระดับความสัมพันธ์ (.99,.90) และขนาดตัวอย่าง = 10 วิธี McD&G ให้ผลดี และจะให้ค่า AMSE ใกล้เคียงกับวิธี HKB เมื่อขนาดตัวอย่างและ σ^2 เพิ่มขึ้นยังคงให้ผลสรุปเหมือนเดิม ค่า AMSE มีแนวโน้มเพิ่มขึ้นเมื่อระดับความสัมพันธ์ และ σ^2 เพิ่มขึ้น และจะมีแนวโน้มลดลงเมื่อขนาดตัวอย่างเพิ่มขึ้น ทั้ง 3 วิธีจะให้ค่า AMSE ใกล้เคียงกัน เมื่อขนาดตัวอย่างเพิ่มขึ้นและ σ^2 = .10

การเพิ่มตัวแปรอิสระจะทำให้ผลสรุปเปลี่ยนไป กล่าวคือ กรณีที่ตัวแปรอิสระ = 3 เมื่อระดับความสัมพันธ์เพิ่มขึ้น วิธี TZE-SAN-LEE ให้ผลดี ส่วนวิธี McD&G ให้ผลดีเมื่อขนาดตัวอย่าง = 10 และในกรณีที่เพิ่มตัวแปรอิสระเป็น 5 และระดับความสัมพันธ์เพิ่มขึ้นวิธี HKB ให้ผลดีและมีแนวโน้มดีขึ้นเมื่อ σ^2 เพิ่มขึ้น ดังนั้นจำนวนตัวแปรอิสระ ระดับความสัมพันธ์ ขนาดตัวอย่าง จะมีผลต่อทั้ง 3 วิธี โดยที่จำนวนตัวแปรอิสระมีผลมากที่สุด

4. ความคลาดเคลื่อนมีการแจกแจงแบบไวบูลล์

กรณีที่ตัวแปรอิสระ = 3 α = 1,2 และ 10

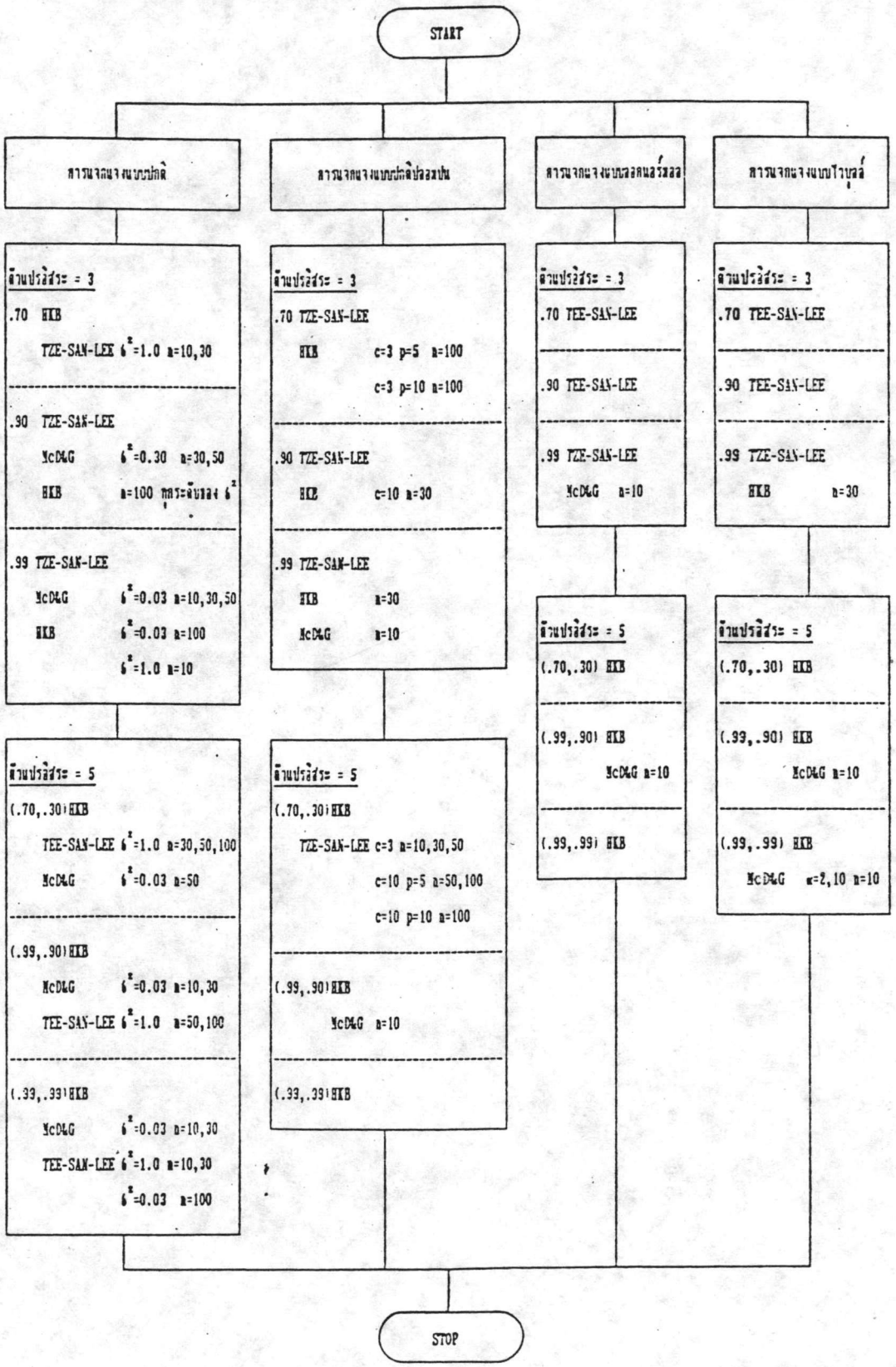
โดยทั่วไปวิธี TZE-SAN-LEE ให้ผลดี เมื่อระดับความสัมพันธ์เพิ่มขึ้นยังคงให้ผลสรุปเหมือนเดิม ยกเว้นที่ระดับความสัมพันธ์ .99 และขนาดตัวอย่าง = 30 วิธี HKB ให้ผลดี และจะให้ค่า AMSE ใกล้เคียงกับวิธี TZE-SAN-LEE สำหรับกรณีที่ขนาดตัวอย่าง และ α เพิ่มขึ้น ยังคงได้ผลสรุปเหมือนเดิม กล่าวคือวิธี TZE-SAN-LEE ให้ผลดี ค่า AMSE มีแนวโน้มลดลงเมื่อขนาดตัวอย่าง และ α เพิ่มขึ้นและค่า AMSE ของทั้ง 3 วิธีจะมีค่าใกล้เคียงกัน เมื่อ α = 10

กรณีที่ตัวแปรอิสระ = 5 α = 1,2 และ 10

โดยทั่วไปวิธี HKB ให้ผลดี เมื่อระดับความสัมพันธ์เพิ่มขึ้นยังคงให้ผลสรุปเหมือนเดิม ยกเว้นเมื่อขนาดตัวอย่าง = 10 ที่ระดับความสัมพันธ์ (.99,.90) α = 1 และที่ระดับความสัมพันธ์ (.99,.99) α = 2,10 วิธี McD&G ให้ผลดี เมื่อขนาดตัวอย่างและ α เพิ่มขึ้น ยังคงให้ผลสรุปเหมือนเดิม กล่าวคือ วิธี HKB ให้ผลดี ค่า AMSE มีแนวโน้มลดลงเมื่อขนาดตัวอย่าง และ α เพิ่มขึ้น และค่า AMSE ของทั้ง 3 วิธีใกล้เคียงกันเมื่อขนาดตัวอย่าง = 10

โดยทั่วไปสำหรับตัวแปรอิสระ = 3 วิธี TZE-SAN-LEE ให้ผลดี เมื่อเพิ่มตัวแปรอิสระเป็น 5 จะมีผลทำให้วิธี HKB ให้ผลดี การเพิ่มตัวแปรอิสระ จะทำให้ผลสรุปเปลี่ยนไป กล่าวคือ กรณีที่ตัวแปรอิสระ = 3 เมื่อระดับความสัมพันธ์เพิ่มขึ้น วิธี TZE-SAN-LEE ให้ผลดี และวิธี HKB จะดีเมื่อขนาดตัวอย่าง = 30 ในกรณีที่เพิ่มตัวแปรอิสระเป็น 5 วิธี HKB ให้ผลดี เมื่อขนาดตัวอย่าง = 10 α = 2 และ 10 วิธี McD&G มีแนวโน้มดีขึ้นเมื่อระดับความสัมพันธ์เพิ่มขึ้น ดังนั้นจำนวนตัวแปรอิสระ ระดับความสัมพันธ์และขนาดตัวอย่าง มีผลต่อทั้ง 3 วิธีโดยที่จำนวนตัวแปรอิสระมีผลมากที่สุด

สรุปผลการเลือกตัวประมาณริดจ์ที่ให้ค่า AMSE ต่ำสุดในสถานการณ์ต่างๆ เป็นแผนผังดังนี้



5.2 การอภิปรายผล

จากการวิเคราะห์เมื่อพิจารณาถึงค่า AMSE, RDAMSE และจำนวน TIMES ของสัมประสิทธิ์การถดถอยพหุด้วยวิธีริดจ์รีเกรสชันพบว่า จำนวนตัวแปรอิสระ ระดับความสัมพันธ์ ลักษณะการแจกแจงของค่าความคลาดเคลื่อนจะมีผลต่อ ค่า AMSE, RDAMSE และจำนวน TIMES ของตัวประมาณริดจ์ ทั้ง 3 วิธี ซึ่งค่า AMSE จะมีแนวโน้มเพิ่มขึ้นเมื่อระดับความสัมพันธ์และการกระจายของค่าความคลาดเคลื่อนมีมาก และจะมีแนวโน้มลดลงเมื่อขนาดตัวอย่างเพิ่มขึ้น

เมื่อความคลาดเคลื่อนมีการแจกแจงแบบปกติ โดยส่วนใหญ่วิธี HKB ให้ผลดี ยกเว้นกรณีที่ตัวแปรอิสระ = 3 ระดับความสัมพันธ์ .90 และ .99 วิธี TZE-SAN-LEE ให้ผลดี

เมื่อความคลาดเคลื่อนมีการแจกแจงแบบปกติปลอมปน โดยส่วนใหญ่วิธี HKB ให้ผลดีสำหรับตัวแปรอิสระ = 5 ยกเว้นที่ระดับความสัมพันธ์(.70,.30) โดยมีสเกลเฟกเตอร์ = 3 เปอร์เซนต์การปลอมปน = 5 และ 10 ขนาดตัวอย่าง = 10,30 และ 100 ,สเกลเฟกเตอร์ = 10 เปอร์เซนต์การปลอมปน = 10 ขนาดตัวอย่าง = 50 และ 100 และที่สเกลเฟกเตอร์ = 10 เปอร์เซนต์การปลอมปน = 10 ขนาดตัวอย่าง = 100 วิธี TZE-SAN-LEE ให้ผลดี สำหรับกรณีที่ตัวแปรอิสระ = 3 โดยส่วนใหญ่วิธี TZE-SAN-LEE ให้ผลดี

เมื่อความคลาดเคลื่อนมีการแจกแจงแบบลอกนอร์มอลและไวบูลล์ โดยส่วนใหญ่วิธี HKB ให้ผลดี ยกเว้นกรณีที่ตัวแปรอิสระ = 3 วิธี TZE-SAN-LEE ให้ผลดี

กรณีที่ค่าความคลาดเคลื่อนมีการแจกแจงแบบลอกนอร์มอล ตัวประมาณริดจ์ทั้ง 3 วิธี จะให้ค่า AMSE และ RDAMSE ใกล้เคียงกัน เมื่อ σ^2 = .10 และกรณีที่ความคลาดเคลื่อนมีการแจกแจงแบบไวบูลล์ ตัวประมาณริดจ์ทั้ง 3 วิธีจะให้ค่า AMSE และ RDAMSE ใกล้เคียงกัน เมื่อ α = 10

ทุกการแจกแจงของค่าความคลาดเคลื่อน จำนวนตัวแปรอิสระจะมีผลต่อตัวประมาณริดจ์ทั้ง 3 วิธีมากที่สุด และไม่มีตัวประมาณริดจ์วิธีใดที่จะให้ค่า AMSE, RDAMSE น้อยที่สุด และจำนวน TIMES มากที่สุดสำหรับทุกกรณี

เมื่อความคลาดเคลื่อนมีความแจกแจงแบบสมมาตรและเบ้ ในบางกรณีจะให้ผลสรุปต่างกัน เนื่องจากการประมาณค่า k โดยวิธี HKB และ McD&G ขึ้นอยู่กับค่าประมาณความคลาดเคลื่อนซึ่งได้มาจากวิธีกำลังสองน้อยที่สุด ดังนั้นเมื่อความคลาดเคลื่อนมีการแจกแจกแตกต่างจากการแจกแจงแบบปกติจึงมีผลกระทบต่อทั้ง 2 วิธีโดยตรง อย่างไรก็ตามจำนวน

ตัวแปรอิสระ ระดับความสัมพันธ์และขนาดตัวอย่างก็มีอิทธิพลต่อค่า AMSE ของทั้ง 3 วิธี เนื่องจากค่าความคลาดเคลื่อนที่เกิดขึ้นไม่ได้เกิดจากระบบเชิงเส้นเพียงอย่างเดียว แต่จะเกิดจากการมี multicollinearity ระหว่างตัวแปรอิสระด้วย

5.3 ข้อเสนอแนะ

1. การเลือกใช้ตัวประมาณริดจ์ จะวิเคราะห์สัมประสิทธิ์การถดถอยพหุ โดยวิธีริดจ์รีเกรสชัน เมื่อตัวแปรอิสระมีความสัมพันธ์กันสูง และจะเลือกตัวประมาณริดจ์จากวิธีที่ให้ค่า AMSE และ RDAMSE น้อยที่สุด ซึ่งขึ้นอยู่กับจำนวนตัวแปรอิสระ ระดับความสัมพันธ์ ลักษณะการแจกแจงของค่าความคลาดเคลื่อน และขนาดตัวอย่าง

2. ควรศึกษาวิจัยวิธีการสถิติวิธีอื่น เช่นวิธี principal component ในการประมาณสัมประสิทธิ์การถดถอยพหุ เมื่อตัวแปรอิสระมีความสัมพันธ์กันสูง และนำมาเปรียบเทียบกับวิธีริดจ์รีเกรสชัน จากวิธีที่ให้ค่า AMSE และ RDAMSE น้อยที่สุด

3. ควรศึกษาเพิ่มเติมในกรณีที่มีข้อมูลของตัวแปรอิสระสูญหาย และมีความสัมพันธ์กันสูง

4. งานวิจัยนี้ศึกษาเปรียบเทียบตัวประมาณริดจ์ สำหรับตัวแปรอิสระเท่ากับ 3 ที่ระดับความสัมพันธ์ .70,.90 และ .99 สำหรับตัวแปรอิสระเท่ากับ 5 ที่ระดับความสัมพันธ์ (.70,.30),(.99,.90)และ(.99,.99) เท่านั้น ผู้สนใจควรศึกษาเพิ่มเติมที่ระดับความสัมพันธ์อื่นนอกเหนือจากงานวิจัยนี้

5. ควรศึกษาวิธีการหาค่า k (ridge estimator) ที่เหมาะสมโดยใช้หลักการของวิธี binary search