การประมาณค่าสัมประสิทธิ์ความถดถอยเชิงพหุเมื่อเกิดพหุสัมพันธ์ด้วยวิธีความถดถอยบูตสแตรปแบบริดจ์

Estimation of Multiple Regression Coefficients with Multicollinearity by Ridge Regression Bootstrapping Method

วริดา พลาศรี (Warida Palasri)* ดร.สุพล ดุรงค์วัฒนา (Dr.Supol Durongwatana)**

บทคัดย่อ

การวิจัยครั้งนี้มีวัตถุประสงค์เพื่อเปรียบเทียบวิธีการประมาณค่าสัมประสิทธิ์ความถดถอยเชิงพหุเมื่อเกิดพหุสัมพันธ์ระหว่างตัวแปรอิสระ โดยเปรียบเทียบวิธีความถดถอยแบบริดจ์และวิธีความถดถอยบูตสแตรปแบบริดจ์ เกณฑ์การเปรียบเทียบที่ใช้ สำหรับการประมาณค่าแบบจุดคือค่าเฉลี่ยความคลาดเคลื่อนกำลังสอง และสำหรับการประมาณค่าแบบช่วงคือค่าความยาวเฉลี่ยของช่วงความเชื่อมั่น ผลการวิจัยสามารถสรุปได้ดังนี้ กรณีการประมาณค่าแบบจุด พบว่ามากกว่า 97% ของสถานการณ์จำลองทั้งหมด วิธีความถดถอยบูตสแตรปแบบริดจ์ให้ค่าเฉลี่ยความคลาดเคลื่อนกำลังสองต่ำสุด กรณีการประมาณค่าแบบช่วง พบว่า 66% ของสถานการณ์จำลองทั้งหมด วิธีความถดถอยบูตสแตรปแบบริดจ์ให้ค่าความยาวเฉลี่ยของช่วงความเชื่อมั่นต่ำกว่าวิธีความถดถอยแบบริดจ์

ABSTRACT

The objective of this study is to compare two methods of estimation for the regression coefficients of multiple regression model with multicollinearity. These two methods are the ridge regression method and the ridge-regression bootstrapping method. The mean squares error is used as the criterion for comparing point estimators and the average length of confidence interval is used as the criterion for comparing interval estimators. The results of the study can be summarized as follows: in case of the point estimation, more than 97% of all simulation situations, the ridge-regression bootstrapping method provides the smallest mean squares error. In case of the interval estimation, 66% of all simulation situations, the average length of confidence interval for the ridge-regression bootstrapping method is shorter than the length from the ridge regression method.

คำสำคัญ : พหุสัมพันธ์ ความถดถอยแบบริดจ์ การบูตสแตรป

Key Words : Multicollinearity, Ridge Regression, Bootstrapping

---

**มหาบัณฑิต หลักสูตรสถิติศาสตรมหาบัณฑิต สาขาวิชาสถิติ คณะพาณิชยศาสตร์และการบัญชี จุฬาลงกรณ์มหาวิทยาลัย*

***รองศาสตราจารย์ ภาควิชาสถิติ คณะพาณิชยศาสตร์และการบัญชี จุฬาลงกรณ์มหาวิทยาลัย*

บทนำ

การวิเคราะห์ความถดถอยเชิงพหุ (Multiple Regression Analysis) เป็นการศึกษาความสัมพันธ์ระหว่างตัวแปรตามหนึ่งตัวกับตัวแปรอิสระตั้งแต่สองตัวขึ้นไป ตัวแบบความถดถอยเชิงเส้นทั่วไป (General Linear Regression Model) ของความสัมพันธ์ระหว่างตัวแปรอิสระและตัวแปรตามมีรูปแบบดังนี้

$$Y_i = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \ldots + \beta_p X_{pi} + \varepsilon_i \; ; i = 1,2,3,\ldots,n$$

เขียนอยู่ในรูปเมตริกซ์ได้ดังนี้

$$\underset{\sim}{Y} = X \underset{\sim}{\beta} + \underset{\sim}{\varepsilon} \qquad (1)$$

เมื่อ $\underset{\sim}{Y}$ เป็นเวกเตอร์ของตัวแปรตามขนาด $n \times 1$, X เป็นเมตริกซ์ของตัวแปรอิสระขนาด $n \times (p+1)$, $\underset{\sim}{\beta}$ เป็นเวกเตอร์ของสัมประสิทธิ์ความถดถอยขนาด $(p+1) \times 1$, $\underset{\sim}{\varepsilon}$ เป็นเวกเตอร์ของความคลาดเคลื่อนขนาด $n \times 1$, n เป็นขนาดตัวอย่าง และ p เป็นจำนวนตัวแปรอิสระ โดยที่ $\underset{\sim}{\varepsilon} \sim N\left(\underset{\sim}{0}, \sigma^2 \mathrm{I}_n\right)$ ซึ่งมี $E\left(\underset{\sim}{\varepsilon}\right) = \underset{\sim}{0}$ และ $Cov\left(\underset{\sim}{\varepsilon}\right) = \sigma^2 \mathrm{I}_n$

ในการประมาณค่าสัมประสิทธิ์ความถดถอยเชิงพหุจากตัวแบบดังกล่าววิธีการประมาณค่าที่นิยมใช้กันมากที่สุดคือวิธีกำลังสองน้อยสุดสามัญ (Ordinary Least Square Method : OLS) เป็นวิธีที่ให้ตัวประมาณค่าที่มีคุณสมบัติเป็นตัวประมาณเชิงเส้นที่ไม่เอนเอียงที่ดีที่สุด (Best Linear Unbiased Estimator : BLUE) โดยที่ตัวประมาณค่าที่ได้อยู่ในรูปของ $\underset{\sim}{\hat{\beta}} = (X'X)^{-1} X' \underset{\sim}{y}$ ในการวิเคราะห์ความถดถอยเชิงพหุข้อมูลที่ใช้อาจก่อให้เกิดปัญหาตัวแปรอิสระมีพหุสัมพันธ์กัน (Multicollinearity) เนื่องจากในการวิเคราะห์ใช้ตัวแปรอิสระมากกว่าหนึ่งตัว ปัญหาที่เกิดขึ้นทำให้เมตริกซ์ $X'X$ เกิดเงื่อนไขที่ไม่ดี (ill-condition) คือ ทำให้ $|X'X|$ มีค่าเล็กลงเข้าใกล้ศูนย์ ซึ่งส่งผลให้สมาชิกของเมตริกซ์ $(X'X)^{-1}$ โดยเฉพาะสมาชิกบนเส้นทแยงมุมมีค่าสูงเกือบเข้าใกล้ค่าอนันต์ เมื่อใช้วิธีกำลังสองน้อยสุดสามัญในการประมาณค่า มีผลทำให้ค่าความแปรปรวนของตัวประมาณค่าสัมประสิทธิ์ความถดถอยมีค่าสูงขึ้น พิจารณาจาก $\mathrm{cov}\left(\underset{\sim}{\hat{\beta}}\right) = \sigma^2 (X'X)^{-1}$ จากค่าของความแปรปรวนที่ได้ ส่งผลต่อตัวประมาณค่าสัมประสิทธิ์ความถดถอยเชิงพหุ ทำให้ขาดความแม่นยำและไม่เที่ยงตรง และเมื่อตัวแปรอิสระมีความสัมพันธ์เชิงเส้นกันอย่างสมบูรณ์ (Perfect Multicollinearity) ผลที่ตามมาจากปัญหาที่เกิดขึ้นคือตัวประมาณค่าสัมประสิทธิ์ความถดถอยไม่สามารถหาค่าได้ เนื่องจาก $|X'X| = 0$ เพราะว่าไม่สามารถหาค่าของ $(X'X)^{-1}$ ได้

ในปี ค.ศ.1970 Hoerl และ Kennard ได้เสนอวิธีความถดถอยแบบริดจ์ (Ridge Regression Method) ในการประมาณค่าสัมประสิทธิ์ความถดถอยเชิงพหุเมื่อเกิดปัญหาตัวแปรอิสระมีความสัมพันธ์กันตัวประมาณค่าที่ได้ให้ค่าเฉลี่ยความคลาดเคลื่อนกำลังสองต่ำกว่าวิธีกำลังสองน้อยสุดสามัญ แต่ตัวประมาณค่าที่ได้เป็นตัวประมาณค่าที่เอนเอียง (Biased Estimator) โดยที่ตัวประมาณค่าสัมประสิทธิ์ความถดถอยเชิงพหุด้วยวิธีความถดถอยแบบริดจ์หาได้จากสมการดังต่อไปนี้ $\underset{\sim}{\hat{\beta}}_R = (X'X + kI)^{-1} X' \underset{\sim}{y}$; $k > 0$ เมื่อ k เป็นค่าคงที่ที่มากกว่าศูนย์ เรียกว่าพารามิเตอร์ริดจ์ และ Ozkale (2008) ได้ทำการศึกษาการประมาณค่าสัมประสิทธิ์ความถดถอยเชิงพหุ กรณีที่เกิดปัญหาความไม่คงที่ของความแปรปรวน (Heteroscedasticity) และ/หรือเกิดปัญหาพหุสัมพันธ์ (Multicollinearity) โดยการนำวิธีแจกไนฟ์ (Jackknife Method) มาใช้ในการประมาณค่าด้วยวิธีความถดถอยแบบริดจ์ ผลการศึกษาพบว่าค่าเฉลี่ยความคลาดเคลื่อนกำลังสองและค่าความเอนเอียงของตัวประมาณค่าสัมประสิทธิ์ความถดถอยเชิงพหุมีค่าลดลง เมื่อเทียบกับวิธีการประมาณค่าด้วยวิธีความถดถอยแบบริดจ์ วิธีแจกไนฟ์เสนอโดย Quenouill (1956) เป็นวิธีการทางนอนพาราเมตริกอีกวิธีการหนึ่ง มีหลักการโดยการนำวิธีการสุ่มตัวอย่างซ้ำมาใช้ในการประมาณค่าพารามิเตอร์ สำหรับการวิจัยครั้งนี้ผู้วิจัยได้สนใจศึกษาโดยการนำวิธีบูตสแตรป (Bootstrapping Method) มาใช้ร่วมกับการประมาณ

ค่าด้วยวิธีความถดถอยแบบริดจ์เนื่องจากวิธีบูตสแตรปมีหลักการและแนวคิดที่ได้พัฒนามาจากวิธีแจกไนฟของ Quenouill วิธีบูตสแตรปเสนอโดย Efron (1970) เป็นวิธีการสุ่มตัวอย่างซ้ำจากตัวอย่างที่มีเพียงชุดเดียว ซึ่งหลักของการสุ่มตัวอย่างซ้ำคือสุ่มตัวอย่างแบบใส่คืนขนาดเท่ากับขนาดตัวอย่างที่มีอยู่เพื่อสร้างข้อมูลชุดใหม่

วัตถุประสงค์ของการวิจัย

1. เพื่อศึกษาและเปรียบเทียบค่าเฉลี่ยความคลาดเคลื่อนกำลังสองของตัวประมาณค่าสัมประสิทธิ์ความถดถอยเชิงพหุระหว่างวิธีความถดถอยแบบริดจ์และวิธีความถดถอยบูตสแตรปแบบริดจ์ เมื่อเกิดปัญหาตัวแปรอิสระมีพหุสัมพันธ์กัน สำหรับกรณีการประมาณค่าแบบจุด

2. เพื่อศึกษาและเปรียบเทียบการประมาณค่าความเชื่อมั่นของสัมประสิทธิ์ความถดถอยเชิงพหุระหว่างวิธีความถดถอยแบบริดจ์ และวิธีความถดถอยบูตแสตปแบบริดจ์ เมื่อเกิดปัญหาตัวแปรอิสระมีพหุสัมพันธ์กัน สำหรับกรณีการประมาณค่าแบบช่วง

3. เพื่อหาข้อสรุปของวิธีการประมาณค่าสัมประสิทธิ์ความถดถอยเชิงพหุที่เหมาะสม เมื่อเกิดปัญหาตัวแปรอิสระมีพหุสัมพันธ์กัน สำหรับกรณีการประมาณค่าแบบจุดและการประมาณค่าแบบช่วง

อุปกรณ์และวิธีการวิจัย

วิธีความถดถอยแบบริดจ์ (Ridge Regression Method)

วิธีความถดถอยแบบริดจ์ ใช้หลักการนำค่าคงที่ที่เหมาะสมค่าหนึ่งมาบวกกับสมาชิกในแนวเส้นทแยงมุมของเมตริกซ์ $X'X$ ในสมการ $\hat{\beta}=(X'X)^{-1}X'y$ จะได้ตัวประมาณค่าสัมประสิทธิ์ความถดถอยเชิงพหุโดยวิธีความถดถอยแบบริดจ์ดังต่อไปนี้

การประมาณค่าแบบจุด ตัวประมาณค่าสัมประสิทธิ์ความถดถอยเชิงพหุ คือ

$$\hat{\beta}_R=(X'X+kI)^{-1}X'y \quad ; \quad k>0 \qquad (2)$$

เมื่อ k เป็นค่าพารามิเตอร์ริดจ์ ซึ่งมีค่ามากว่าศูนย์

$$Cov\left(\hat{\beta}_R\right)=\sigma^2\left(X'X+kI\right)^{-1}X'X\left(X'X+kI\right)^{-1} \qquad (3)$$

การประมาณค่าแบบช่วง ระดับความเชื่อมั่นของช่วงความเชื่อมั่นสำหรับ $\hat{\beta}_{Ri}$; $i=0,1,2,\ldots,p$ คือ

$$\hat{\beta}_{Ri}\pm t_{(n-p-1),\frac{\alpha}{2}}S_{ii} \quad ; \quad i=0,1,2,\ldots p \qquad (4)$$

โดยที่ $\hat{\beta}_{Ri}$ คือตัวประมาณค่าสัมประสิทธิ์ความถดถอยเชิงพหุด้วยวิธีความถดถอยแบบริดจ์ของ β_i , $S_{ii}=\sqrt{Var\left(\hat{\beta}_{Ri}\right)}$ คือ ความคลาดเคลื่อนมาตรฐานของ $\hat{\beta}_{Ri}$, $Var\left(\hat{\beta}_{Ri}\right)=S^2R_{ii}$, $S^2=\hat{\sigma}^2=\frac{1}{n-p-1}\left(y-X\hat{\beta}_R\right)'\left(y-X\hat{\beta}_R\right)$, p คือจำนวนตัวแปรอิสระ และ $t_{\frac{\alpha}{2}}$ คือเปอร์เซ็นต์ไทล์ของการแจกแจงที (t Distribution)

วิธีความถดถอยบูตสแตรปแบบริดจ์ (Ridge Regression Bootstrapping Method)

การประมาณค่าด้วยวิธีบูตสแตรปเป็นการประมาณค่าพารามิเตอร์โดยการสุ่มตัวอย่างซ้ำจากตัวอย่างที่มีอยู่ชุดเดียว เพื่อให้ได้ตัวอย่างสุ่มหลายๆ ชุด หลักการของการสุ่มตัวอย่างซ้ำคือสุ่มตัวอย่างแบบใส่คืน (Random Sampling With Replacement) ขนาดเท่ากับตัวอย่างเพื่อสร้างข้อมูลชุดใหม่ แล้วนำมาใช้ในการประมาณค่าพารามิเตอร์ ขั้นตอนของการประมาณค่าสัมประสิทธิ์ความถดถอยเชิงพหุด้วยวิธีบูตสแตรป แบบริดจ์มีขั้นตอนดังต่อไปนี้

1. จากข้อมูล X และ y ที่มี คำนวณหาตัวประมาณค่าสัมประสิทธิ์ความถดถอยเชิงพหุ โดยวิธีความถดถอยแบบริดจ์จากสมการที่ (2)

2. จากตัวประมาณค่า $\hat{\beta}_R$ โดยวิธีความถดถอยแบบริดจ์ที่ได้ นำมาหาค่า $\hat{\varepsilon}$ โดยที่ $\hat{\varepsilon}=y-\hat{y}=y-X\hat{\beta}_R$

3. จาก $\hat{\varepsilon}_1, \hat{\varepsilon}_2, \ldots, \hat{\varepsilon}_n$ ทำการสุ่มตัวอย่างขนาด n แบบใส่คืนได้ $\hat{\varepsilon}_1^*, \hat{\varepsilon}_2^*, \ldots, \hat{\varepsilon}_n^*$ เมื่อ $\hat{\varepsilon}_i^*$ คือตัวอย่างที่สุ่มได้ตัวที่ i จาก $\hat{\varepsilon}_1, \hat{\varepsilon}_2, \ldots, \hat{\varepsilon}_n$

4. นำค่า $\hat{\varepsilon}_i^*$ มาพิจารณาในสมการจะได้ค่าของเวกเตอร์ของตัวแปรตามใหม่ดังนี้ $\underset{\sim}{y}^* = X\underset{\sim}{\hat{\beta}}_R + \underset{\sim}{\hat{\varepsilon}}^*$

5. นำค่า $\underset{\sim}{y}^*$ และ X คำนวณหาค่า $\underset{\sim}{\hat{\beta}}_R^*$ โดยวิธีความถดถอยแบบริดจ์ ซึ่งตัวประมาณสัมประสิทธิ์ความถดถอยเชิงพหุของข้อมูลชุดใหม่ได้ดังต่อไปนี้ $\underset{\sim}{\hat{\beta}}_R^* = (X'X + kI)^{-1} X' \underset{\sim}{y}^*$

6. ทำตามขั้นตอนในข้อ 2 – 5 ซ้ำกันเท่ากับจำนวนที่บูตสแตรป กำหนดให้จำนวนที่บูตสแตรปเป็น B ครั้ง จะได้ $\underset{\sim}{\hat{\beta}}_R^{*1}, \underset{\sim}{\hat{\beta}}_R^{*2}, \ldots, \underset{\sim}{\hat{\beta}}_R^{*B}$

7. ตัวประมาณค่าสัมประสิทธิ์ความถดถอยเชิงพหุด้วยวิธีความถดถอยบูตสแตรปแบบริดจ์

การประมาณค่าแบบจุด ตัวประมาณค่าสัมประสิทธิ์ความถดถอยเชิงพหุ ใช้ค่าเฉลี่ยที่ได้จากขั้นตอนที่ 6 คือ

$$\overline{\underset{\sim}{\hat{\beta}}}_R^* = \frac{\sum_{j=1}^{B} \underset{\sim}{\hat{\beta}}_R^{*j}}{B} \qquad (5)$$

การประมาณค่าแบบช่วง สำหรับการประมาณค่าแบบช่วงของตัวประมาณค่าสัมประสิทธิ์ความถดถอยเชิงพหุด้วยวิธีความถดถอยบูตส แตรปแบบริดจ์ พิจารณาจากค่าประมาณ $\underset{\sim}{\hat{\beta}}_R^{*1}, \underset{\sim}{\hat{\beta}}_R^{*2}, \ldots, \underset{\sim}{\hat{\beta}}_R^{*B}$ ที่ได้ในขั้นตอนที่ 6 นำมาหาเปอร์เซ็นต์ไทล์ที่ $100\left(\frac{\alpha}{2}\right)$ และเปอร์เซ็นต์ไทล์ที่ $100\left(1-\frac{\alpha}{2}\right)$ จากค่าของเปอร์เซ็นต์ไทล์ที่ได้ มีค่าเท่ากับขีดจำกัดความเชื่อมั่นล่างและขีดจำกัดความเชื่อมั่นบน ตามลำดับ

วิธีดำเนินการวิจัย

ในการวิจัยครั้งนี้ได้กำหนดสถานการณ์ต่างๆ สำหรับการศึกษาการประมาณค่าสัมประสิทธิ์ความถดถอยเชิงพหุ เมื่อเกิดพหุสัมพันธ์ของตัวแบบ $\underset{\sim}{Y} = X\underset{\sim}{\beta} + \underset{\sim}{\varepsilon}$ โดยสร้างสถานการณ์ที่ต้องการศึกษาภายใต้เงื่อนไขดังนี้คือ จำนวนตัวแปรอิสระ (p) ที่ใช้ในการศึกษามี 2,3,4 และ 5 ตัว ขนาดตัวอย่าง (n) ที่ใช้ในศึกษาคือ 15,30,50 และ 100 เวกเตอร์ $\underset{\sim}{\varepsilon}$ มีการแจกแจงแบบปกติ (Normal Distribution) เมื่อค่าเฉลี่ยของความคลาดเคลื่อน ε_i มีค่าเป็นศูนย์นั่นคือ $E\left(\underset{\sim}{\varepsilon}\right) = \underset{\sim}{0}$ และ $Cov\left(\underset{\sim}{\varepsilon}\right) = \sigma^2 \mathrm{I}_n$ โดยที่ส่วนเบี่ยงเบนมาตรฐาน (σ) มีค่าเท่ากับ 1,5 และ 10 ค่าสัมประสิทธิ์ความเชื่อมั่นที่ใช้ในการประมาณค่าแบบช่วงมี 3 ระดับคือ 0.90,0.95 และ 0.99 ตามลำดับ กำหนดค่าสัมประสิทธิ์ความถดถอยให้เป็นค่าคงที่ใดๆ เพื่อสร้างเวกเตอร์ตัวแปรตามจากตัวแบบ $\underset{\sim}{Y} = X\underset{\sim}{\beta} + \underset{\sim}{\varepsilon}$ โดยกำหนด $\underset{\sim}{\beta} = \underset{\sim}{1}$ สร้างตัวแปรอิสระ (X) ให้มีลักษณะความสัมพันธ์ตามที่ต้องการศึกษา โดยในการศึกษาครั้งนี้ใช้วิธีจำลองของ Wichern และ Chrchill (1978) เพื่อสร้างตัวแปรอิสระให้มีความสัมพันธ์กันในระดับต่างๆ ดังต่อไปนี้

$x_{ij} = (1-\alpha^2)^{1/2} Z_{ij} + \alpha Z_{i\,p+1}$; $i = 1,2,\ldots,p$; $j = 1,2,\ldots,n$

โดยที่ Z_{ij} เป็นค่าที่สร้างขึ้นจากการแจกแจงปกติที่มีค่าเฉลี่ยเท่ากับศูนย์และค่าความแปรปรวนเท่ากับหนึ่ง, α^2 เป็นระดับความสัมพันธ์ระหว่างตัวแปร X_i กับ X_j เมื่อ $i = 1,2..,p$; $j = 1,2..,p$; $i \neq j$ และ p เป็นจำนวนตัวแปรอิสระ ระดับความสัมพันธ์ระหว่างตัวแปรอิสระ (ρ) ที่ศึกษามี 3 ระดับคือระดับต่ำ $\rho = 0.30$, ระดับปานกลาง $\rho = 0.60$ และระดับสูง $\rho = 0.90$ โดยที่ ρ คือ ค่าของระดับความสัมพันธ์ระหว่างตัวแปรอิสระ X_i กับ X_j เมื่อ $i = 1,2,..,p; j = 1,2,..,p; i \neq j$ สร้างข้อมูลให้มีลักษณะสัมพันธ์กันเป็นไปตามโครงสร้างที่กำหนด โดยที่เมตริกซ์สหสัมพันธ์ (Correlation Matrix) ของตัวแปรอิสระมีรูปแบบตามตัวอย่างในกรณีของจำนวนตัวแปรอิสระ 3 ตัว ดังนี้

$$Corr\,(x) = \begin{bmatrix} r_{x_1x_1} & r_{x_1x_2} & r_{x_1x_3} \\ r_{x_2x_1} & r_{x_2x_2} & r_{x_2x_3} \\ r_{x_3x_1} & r_{x_3x_2} & r_{x_3x_3} \end{bmatrix} = \begin{bmatrix} 1 & \rho & \rho \\ \rho & 1 & \rho \\ \rho & \rho & 1 \end{bmatrix}$$

การประมาณค่า k ซึ่งเป็นค่าพารามิเตอร์ริดจ์ มี 3 วิธี

1. วิธี KS (Khalaf and Shukur, 2005)

$$k = \frac{\lambda_{\max}\hat{\sigma}^2}{(n-p)\hat{\sigma}^2 + \lambda_{\max}\hat{\alpha}_{\max}^2}$$

2. วิธี New HKB (Kibria, 2003)

$$k = \frac{p\hat{\sigma}^2}{\hat{\alpha}'\hat{\alpha}} + \frac{1}{\lambda_{\max}}$$

3. วิธี New LW (Kibria, 2003)

$$k = \frac{p\hat{\sigma}^2}{\sum_{i=1}^{p}\lambda_i\hat{\alpha}^2} + \frac{1}{\lambda_{\max}}$$

โดยที่ p เป็นจำนวนตัวแปรอิสระ $\lambda_{\max}$ เป็นค่าเฉพาะที่มีค่ามากสุดของเมตริกซ์ $X'X$ เมื่อกำหนดให้ P เป็นเมตริกซ์เชิงตั้งฉากขนาด $p \times p$ ซึ่งแต่ละแนวตั้งฉากของเมตริกซ์คือเวกเตอร์เฉพาะ (Eigen Vector) ที่มีค่าสอดคล้องกับค่าเฉพาะ (Eigen Value) $\lambda_1, \lambda_2, \ldots, \lambda_p$ ของเมตริกซ์ $X'X$ และ $X^* = XP$ จะได้ $X^{*'}X^* = P'X'XP = \Lambda = diag\ (\lambda_i)$

$$\hat{\alpha} = \left(X^{*'}X\right)^{-1}X^{*'}y = \Lambda^{-1}X^{*'}y$$

และ $\hat{\sigma}^2 = \frac{y'y - \hat{\alpha}X'y}{n-p}$

การวิจัยครั้งนี้จำลองตามสถานการณ์ต่างๆ โดยใช้เทคนิคมอนติคาร์โลซีมูเลชั่น (Monte Carlo Simulation) โดยใช้โปรแกรม R 2.8.1 ซึ่งมีการทำซ้ำของการจำลอง 500 ครั้งและมีการทำซ้ำของการบูตสแตรป 400 ครั้ง ในทุกสถานการณ์ของการจำลอง ทำการจำลองสถานการณ์ที่ต้องการศึกษาภายใต้เงื่อนไขที่ได้กำหนด แล้วจากนั้นทำการประมาณค่าสัมประสิทธิ์ความถดถอยเชิงพหุทั้ง 2 วิธี สำหรับการประมาณค่าแบบจุด คำนวณค่าเฉลี่ยความคลาดเคลื่อนกำลังสอง และค่าความเอนเอียงกำลังสองเพื่อเปรียบเทียบว่าวิธีการประมาณค่าใดให้ค่าต่ำที่สุด โดยที่สามารถหาได้ดังนี้

ค่าเฉลี่ยความคลาดเคลื่อนกำลังสอง

$$MSE\left(\hat{\beta}\right) = \frac{1}{500}\sum_{r=1}^{500}\left(\hat{\beta}_{(r)} - \beta\right)'\left(\hat{\beta}_{(r)} - \beta\right)$$

ค่าความเอนเอียงกำลังสอง

$$BIAS^2\left(\hat{\beta}\right) = \left(\bar{\hat{\beta}} - \beta\right)'\left(\bar{\hat{\beta}} - \beta\right)$$

เมื่อ $\bar{\hat{\beta}} = \frac{1}{500}\sum_{r=1}^{500}\hat{\beta}_{(r)}$

ซึ่ง $MSE\left(\hat{\beta}\right) = Cov\left(\hat{\beta}\right) + BIAS^2\left(\hat{\beta}\right)$ (6)

โดยที่ r เป็นการทำซ้ำในการจำลอง $r = 1,2,3,\ldots,500$, $\hat{\beta}_{(r)}$ เป็นเวกเตอร์ของตัวประมาณค่าสัมประสิทธิ์ความถดถอยเชิงพหุที่ได้ในแต่ละวิธีของการทำซ้ำครั้งที่ r ในการจำลองของแต่ละสถานการณ์ และ β เป็นเวกเตอร์ของค่าสัมประสิทธิ์ความถดถอยเชิงพหุที่กำหนดไว้เป็นค่าคงที่ใดๆ ซึ่งการวิจัยครั้งนี้กำหนดเท่ากับ 1

สำหรับการประมาณค่าแบบช่วง คำนวณค่าระดับความเชื่อมั่นของช่วงความเชื่อมั่นที่ได้จากแต่ละวิธี หลังจากนั้นคัดเลือกวิธีการประมาณค่าที่ให้ค่าระดับความเชื่อมั่นของการจำลองไม่ต่ำกว่าค่าสัมประสิทธิ์ความเชื่อมั่นที่กำหนด โดยคัดเลือกจากการทดสอบทวินาม แล้วทำการเปรียบเทียบค่าความยาวเฉลี่ยของช่วงความเชื่อมั่นของวิธีที่ให้ค่าระดับความเชื่อมั่นของการจำลองไม่ต่ำกว่าค่าสัมประสิทธิ์ความเชื่อมั่นที่กำหนด โดยพิจารณาว่าวิธีใดให้ค่าความยาวเฉลี่ยของช่วงความเชื่อมั่นต่ำที่สุดกว่ากัน ซึ่งค่าความยาวเฉลี่ยของช่วงความเชื่อมั่น $= \frac{1}{500}\sum_{r=1}^{500}(U_r - L_r)$ โดยที่ U_r , L_r คือ ขอบเขตบนและขอบเขตล่างของช่วงความเชื่อมั่น ในการทำซ้ำครั้งที่ r ซึ่ง $r = 1,2,3,\ldots,500$ จากค่าที่คำนวณได้นำมาสรุปผลการวิจัยในแต่ละสถานการณ์

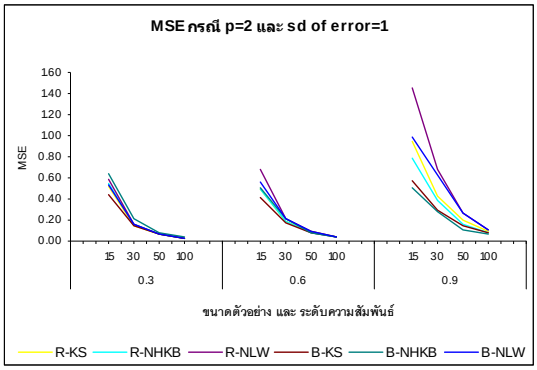
ผลการวิจัยและการอภิปรายผล

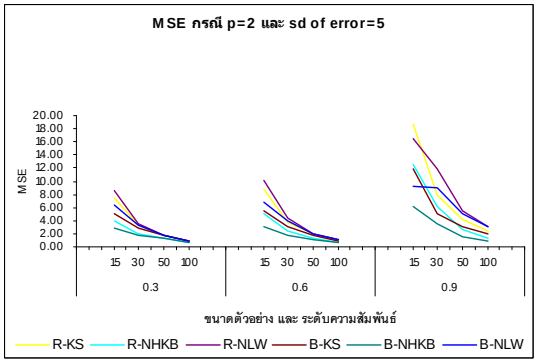
การประมาณค่าแบบจุด

ในกรณีที่จำนวนตัวแปรอิสระเท่ากับ 2 พบว่าเมื่อขนาดตัวอย่างเพิ่มขึ้นค่า MSE ที่ได้จะมีค่าลดลง เมื่อระดับความสัมพันธ์เพิ่มขึ้นค่า MSE ที่ได้จะมีค่าเพิ่มขึ้น ในขณะเดียวกันเมื่อขนาดตัวอย่างเท่ากันและระดับความสัมพันธ์เดียวกันพบว่าเมื่อส่วนเบี่ยงเบนมาตรฐานของความคลาดเคลื่อนเพิ่มขึ้นค่า

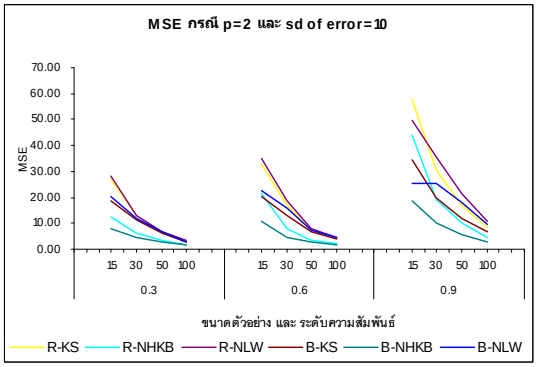
MSE ในแต่ละวิธีจะสูงขึ้น โดยที่วิธีความถดถอยบูตสแตรปแบบริดจ์จะให้ค่า MSE ต่ำกว่าวิธีความถดถอยแบบริดจ์ ซึ่งส่วนใหญ่วิธีการประมาณค่า k ด้วยวิธี New HKB ให้ค่า MSE ต่ำที่สุด รองลงมาคือวิธี KS และวิธี New LW ตามลำดับ แสดงดังภาพที่ 1

สำหรับกรณีที่จำนวนตัวแปรอิสระเท่ากับ 3,4 และ 5 ค่า MSE มีแนวโน้มในลักษณะเดียวกันกับในกรณีจำนวนตัวแปรอิสระเท่ากับ 2 แต่ค่า MSE ที่ได้จะมีค่าเพิ่มสูงขึ้นตามจำนวนตัวแปรอิสระ พิจารณาได้จากภาพที่ 2-4

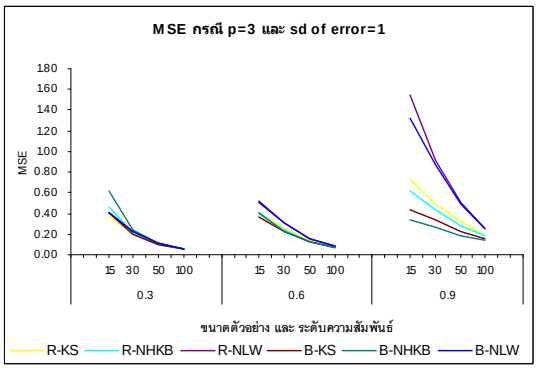


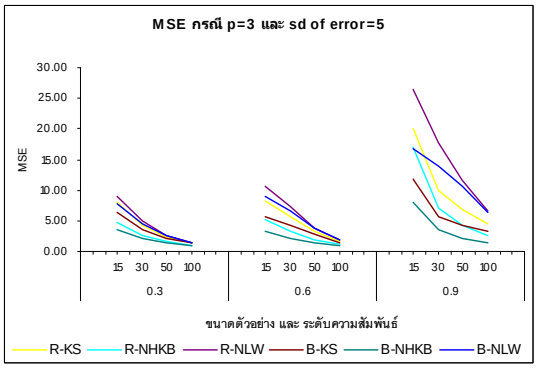


ภาพที่ 1 กราฟแสดงการเปรียบเทียบค่า MSE กรณีที่จำนวนตัวแปรอิสระเท่ากับ 2 และค่าส่วนเบี่ยงเบนมาตรฐานของความคลาดเคลื่อนเท่ากับ 1,5 และ 10 ตามลำดับ

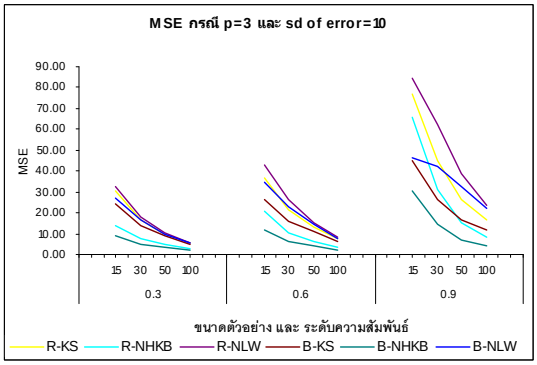


ภาพที่ 1 กราฟแสดงการเปรียบเทียบค่า MSE กรณีที่จำนวนตัวแปรอิสระเท่ากับ 2 และค่าส่วนเบี่ยงเบนมาตรฐานของความคลาดเคลื่อนเท่ากับ 1,5 และ 10 ตามลำดับ (ต่อ)

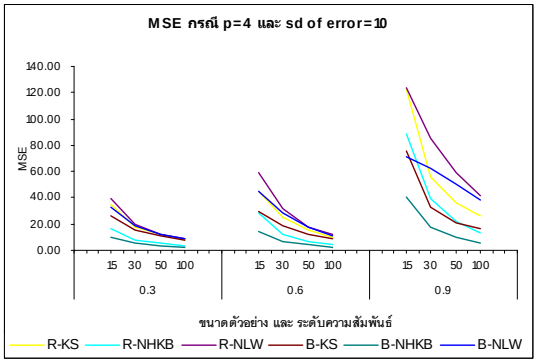




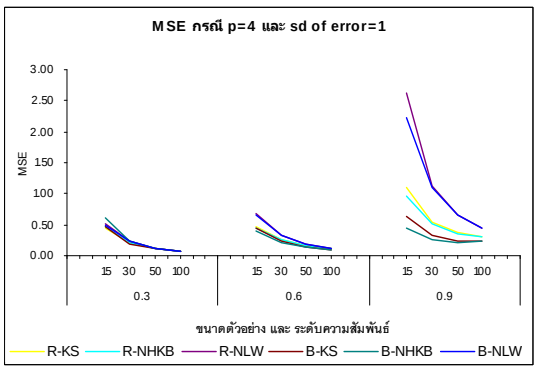
ภาพที่ 2 กราฟแสดงการเปรียบเทียบค่า MSE กรณีที่จำนวนตัวแปรอิสระเท่ากับ 3 และค่าส่วนเบี่ยงเบนมาตรฐานของความคลาดเคลื่อนเท่ากับ 1,5 และ 10 ตามลำดับ

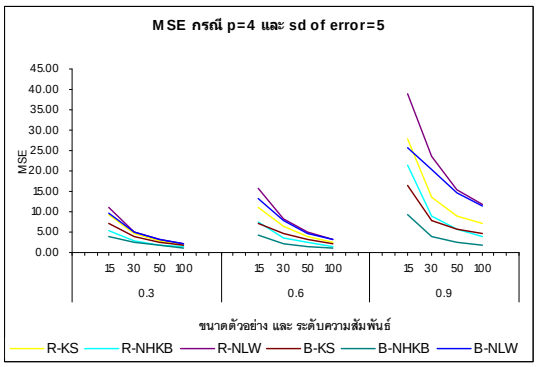


ภาพที่ 2 กราฟแสดงการเปรียบเทียบค่า MSE กรณีที่จำนวนตัวแปรอิสระเท่ากับ 3 และค่าส่วนเบี่ยงเบนมาตรฐานของความคลาดเคลื่อนเท่ากับ 1,5 และ 10 ตามลำดับ (ต่อ)

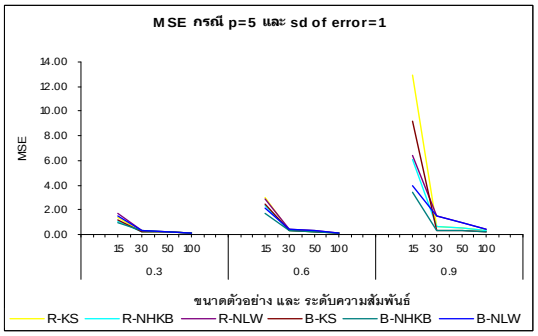


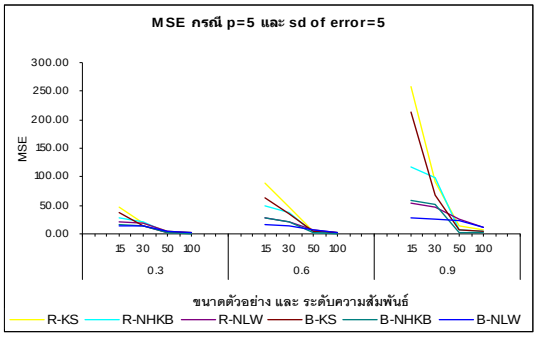
ภาพที่ 3 กราฟแสดงการเปรียบเทียบค่า MSE กรณีที่จำนวนตัวแปรอิสระเท่ากับ 4 และค่าส่วนเบี่ยงเบนมาตรฐานของความคลาดเคลื่อนเท่ากับ 1,5 และ 10 ตามลำดับ (ต่อ)



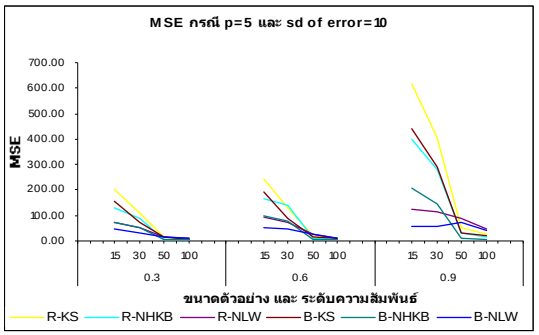


ภาพที่ 3 กราฟแสดงการเปรียบเทียบค่า MSE กรณีที่จำนวนตัวแปรอิสระเท่ากับ 4 และค่าส่วนเบี่ยงเบนมาตรฐานของความคลาดเคลื่อนเท่ากับ 1,5 และ 10 ตามลำดับ





ภาพที่ 4 กราฟแสดงการเปรียบเทียบค่า MSE กรณีที่จำนวนตัวแปรอิสระเท่ากับ 5 และค่าส่วนเบี่ยงเบนมาตรฐานของความคลาดเคลื่อนเท่ากับ 1,5 และ 10 ตามลำดับ



ภาพที่ 4 กราฟแสดงการเปรียบเทียบค่า MSE กรณีที่จำนวนตัวแปรอิสระเท่ากับ 5 และค่าส่วนเบี่ยงเบนมาตรฐานของความคลาดเคลื่อนเท่ากับ 1,5 และ 10 ตามลำดับ (ต่อ)

โดยที่

R-KS หมายถึงวิธีความถดถอยแบบริดจ์ โดยใช้ค่า k ด้วยวิธี KS

R-NHKB หมายถึงวิธีความถดถอยแบบริดจ์ โดยใช้ค่า k ด้วยวิธี New HKB

R-NLW หมายถึงวิธีความถดถอยแบบริดจ์ โดยใช้ค่า k ด้วยวิธี New LW

B-KS หมายถึงวิธีความถดถอยบูตสแตรปแบบริดจ์ โดยใช้ค่า k วิธี KS

B-NHKB หมายถึงวิธีความถดถอยบูตสแตรปแบบริดจ์ โดยใช้ค่า k ด้วยวิธี New HKB

B-NLW หมายถึงวิธีความถดถอยบูตสแตรปแบบริดจ์ โดยใช้ค่า k ด้วยวิธี New LW

MSE หมายถึงค่า $MSE\left(\hat{\beta}_{\sim}\right)$

$BIAS^2$ หมายถึงค่า $BIAS^2\left(\hat{\beta}_{\sim}\right)$

จากภาพที่ 1-4 พบว่าเกือบทุกสถานการณ์ของการจำลองทั้งหมด การประมาณค่าสัมประสิทธิ์ความถดถอยเชิงพหุด้วยวิธีความถดถอยบูตสแตรปแบบริดจ์ให้ค่า MSE ต่ำที่สุด โดยส่วนใหญ่วิธีการประมาณค่า k ด้วยวิธี New HKB จะให้ค่า MSE ต่ำสุด รองลงมาคือวิธี KS และวิธี New LW ตามลำดับ และพบว่าเมื่อขนาดตัวอย่างใหญ่ ระดับความสัมพันธ์อยู่ในระดับต่ำและส่วนเบี่ยงเบนมาตรฐานของความคลาดเคลื่อนมีค่าน้อย วิธีของการประมาณค่าแต่ละวิธีจะให้ค่า MSE ใกล้เคียงกันมาก และมีค่าน้อย ในทางตรงกันข้ามเมื่อขนาดตัวอย่างมีขนาดเล็ก ระดับความสัมพันธ์สูงขึ้นและส่วนเบี่ยงเบนมาตรฐานของความคลาดเคลื่อนมีค่ามากขึ้น จะพบความแตกต่างของค่า MSE ได้อย่างชัดเจนและค่า MSE ที่ได้มีค่าสูงขึ้น นั่นคือค่า MSE มีแนวโน้มลดลงเมื่อขนาดตัวอย่างเพิ่มขึ้น และมีแนวโน้มเพิ่มขึ้นเมื่อระดับความสัมพันธ์สูงขึ้นและส่วนเบี่ยงเบนมาตรฐานของความคลาดเคลื่อนมีค่ามากขึ้น เพราะเมื่อขนาดตัวอย่างเพิ่มขึ้นทำให้ค่าเฉพาะของเมตริกซ์ $X'X$ มีค่าลดลงจึงส่งผลทำให้ค่า MSE มีแนวโน้มเพิ่มขึ้น และเมื่อส่วนเบี่ยงเบนมาตรฐานของความคลาดเคลื่อนเพิ่มมากขึ้นทำให้เกิดความคลาดเคลื่อนมากขึ้นส่งผลให้ค่า MSE มีแนวโน้มเพิ่มขึ้น นอกจากนี้กรณีของการเพิ่มจำนวนตัวแปรอิสระ พบว่าถ้าระดับความสัมพันธ์ของตัวแปรอิสระต่ำ เมื่อจำนวนตัวแปรอิสระเพิ่มขึ้นค่า MSE มีค่าเพิ่มขึ้นเล็กน้อยและมีแนวโน้มที่จะคงที่หรือลดลงเมื่อจำนวนตัวแปรอิสระเพิ่มขึ้น เพราะว่ามีปัจจัยที่มีปัญหาพหุสัมพันธ์กันค่อนข้างน้อยมาอธิบายสมการถดถอย แต่ในกรณีที่ระดับความสัมพันธ์ของตัวแปรอิสระสูงจะทำให้เกิดปัญหาพหุสัมพันธ์มากขึ้น การเพิ่มจำนวนตัวแปรอิสระทำให้ค่า MSE สูงขึ้น เพราะไม่สามารถบอกได้ว่าการที่มีจำนวนตัวแปรอิสระหลายตัวจะสามารถอธิบายได้ดีขึ้น ด้วยสาเหตุที่มีปัญหาพหุสัมพันธ์กันสูงจากปัจจัยที่ส่งผลต่อการเปลี่ยนแปลงของค่า MSE สามารถสรุปได้ดังนี้

1. ขนาดตัวอย่าง ค่า MSE มีแนวโน้มลดลงเมื่อขนาดตัวอย่างเพิ่มขึ้น เพราะเมื่อขนาดตัวอย่างเพิ่มขึ้น หมายความว่าเราได้ข้อมูลมากขึ้นเพื่อนำมาใช้ในการอธิบายสมการความถดถอยเชิงพหุ จึงส่งผลทำให้ค่า MSE ลดลง

2. ระดับความสัมพันธ์ค่า MSE มีแนวโน้มเพิ่มขึ้นเมื่อระดับความสัมพันธ์เพิ่มขึ้น เพราะเมื่อ

ระดับความสัมพันธ์เพิ่มขึ้นทำให้ค่าเฉพาะของเมตริกซ์ $X'X$ มีค่าลดลง จึงส่งผลทำให้ค่า MSE มีค่าเพิ่มขึ้น

3. ส่วนเบี่ยงเบนมาตรฐานของความคลาดเคลื่อน ค่า MSE มีแนวโน้มเพิ่มขึ้นเมื่อค่าของส่วนเบี่ยงเบนมาตรฐานของความคลาดเคลื่อนเพิ่มขึ้น เพราะเมื่อส่วนเบี่ยงเบนมาตรฐานของความคลาดเคลื่อนเพิ่มขึ้น หมายความว่าความคลาดเคลื่อนมีการกระจายมากขึ้น ส่งผลให้การกระจายตัวของตัวแปรตามเพิ่มขึ้นทำให้ค่าประมาณของพารามิเตอร์ที่ได้สูงขึ้น จึงทำให้ค่า MSE มีค่าเพิ่มขึ้น

4. จำนวนตัวแปรอิสระค่า MSE มีแนวโน้มเพิ่มขึ้นเมื่อจำนวนตัวแปรอิสระเพิ่มขึ้น เพราะเมื่อจำนวนตัวแปรอิสระเพิ่มขึ้นทำให้เกิดความสัมพันธ์ระหว่างกลุ่มตัวแปรอิสระเพิ่มขึ้น จึงทำให้ค่า MSE มีค่าเพิ่มขึ้น

จากสถานการณ์จำลองทั้งหมดเมื่อทำการเปรียบเทียบสถานการณ์ของวิธีการประมาณค่าที่ให้ค่า MSE น้อยที่สุด พบว่าวิธีความถดถอยบูตสแตรปแบบริดจ์เป็นวิธีที่มีจำนวนสถานการณ์ที่ให้ค่า MSE ต่ำที่สุดมากสุด โดยมีเปอร์เซ็นต์เท่ากับ 97% ซึ่งค่า MSE ต่ำสุดที่ได้อยู่ในสถานการณ์ที่ระดับความสัมพันธ์เท่ากับ 0.6 และ 0.9 ส่วนอีก 3% จากสถานการณ์ทั้งหมดเป็นของวิธีความถดถอยแบบริดจ์ ซึ่งค่า MSE ต่ำสุดที่ได้อยู่ในสถานการณ์ที่ระดับความสัมพันธ์เท่ากับ 0.3 พิจารณาตามวิธีการประมาณค่า k วิธีที่มีจำนวนสถานการณ์มากที่สุดที่ให้ค่า MSE ต่ำที่สุดคือ วิธี New HKB รองลงมาคือวิธี KS และวิธี New LW ตามลำดับ และพบว่าเมื่อระดับความสัมพันธ์มีค่าเพิ่มขึ้นการประมาณค่า k ด้วยวิธี New HKB จะมีเปอร์เซ็นต์เพิ่มขึ้นและวิธีอื่นๆ มีค่าลดลง แสดงว่าวิธีการประมาณค่านี้มีความเหมาะสมเมื่อตัวแปรอิสระมีความสัมพันธ์กันสูง

ค่า $BIAS^2$ มีค่าลดลงเมื่อขนาดตัวอย่างเพิ่มขึ้นในทุกๆ ระดับความสัมพันธ์ แต่จะมีค่าเพิ่มขึ้นเมื่อส่วนเบี่ยงเบนมาตรฐานของความคลาดเคลื่อนเพิ่มขึ้น จำนวนตัวแปรอิสระเพิ่มขึ้น และจำนวนขนาดตัวอย่างเล็ก ซึ่งการประมาณค่าด้วยวิธีความถดถอยแบบริดจ์จะให้ค่า $BIAS^2$ น้อยกว่าวิธีความถดถอยบูตสแตรปแบบริดจ์ในทุกๆ ค่า k โดยที่วิธีการประมาณค่า k ด้วยวิธี New LW ให้ค่า $BIAS^2$ น้อยที่สุด รองลงมาคือวิธี KS และวิธี New HKB ตามลำดับ จากค่าของ MSE ที่ได้และความสัมพันธ์ของสมการ (6) แสดงว่าวิธีความถดถอยแบบริดจ์ให้ค่าความแปรปรวนของตัวประมาณสัมประสิทธิ์ความถดถอยเชิงพหุสูงกว่าวิธีความถดถอยบูตสแตรปแบบริดจ์ในทุกๆ ค่า k นั่นคือการประมาณค่าด้วยวิธีความถดถอยบูตสแตรปแบบริดจ์จะให้ค่าความแปรปรวนลดลงและให้ค่า $BIAS^2$ เพิ่มขึ้นสูงกว่าวิธีความถดถอยแบบริดจ์ แต่อัตราการเพิ่มของค่า $BIAS^2$ น้อยกว่าอัตราการลดของความแปรปรวน ทำให้วิธีความถดถอยบูตสแตรปแบบริดจ์ในทุกๆ ค่า k มีค่า MSE ต่ำกว่าวิธีความถดถอยแบบริดจ์ในเกือบทุกสถานการณ์

การประมาณค่าแบบช่วง

พบว่าโดยสถานการณ์จำลองทั้งหมด การประมาณค่าด้วยวิธีความถดถอยแบบริดจ์เป็นวิธีที่มีสถานการณ์ที่ให้ค่าสัมประสิทธิ์ความเชื่อมั่นที่ได้จากการจำลองไม่ต่ำกว่าค่าสัมประสิทธิ์ความเชื่อมั่นที่กำหนดมากกว่าวิธีความถดถอยบูตสแตรปแบบริดจ์ เพราะว่าจากการประมาณค่าแบบจุดด้วยวิธีความถดถอยแบบริดจ์ เป็นวิธีการประมาณค่าที่ให้ค่าความแปรปรวนสูงกว่าวิธีความถดถอยบูตสแตรปแบบริดจ์ แต่วิธีความถดถอยแบบริดจ์ให้ค่าความเอนเอียงน้อยกว่าวิธีความถดถอยบูตสแตรปแบบริดจ์ จึงส่งผลให้โอกาสที่ช่วงความเชื่อมั่นที่คำนวณได้จากวิธีความถดถอยแบบริดจ์คลุมค่าพารามิเตอร์เริ่มต้นที่กำหนดมีค่าสูง ผลที่ตามมาคือจำนวนสถานการณ์ที่ค่าสัมประสิทธิ์ความเชื่อมั่นที่ได้จากการจำลองจากวิธีความถดถอยแบบริดจ์ให้ค่าไม่ต่ำกว่าค่าสัมประสิทธิ์ความเชื่อมั่นที่กำหนดมีจำนวนมากกว่าวิธีความถดถอยบูตสแตรปแบบริดจ์ สามารถสรุปแยกตาม

วิธีการประมาณค่า k แต่ละวิธี โดยที่วิธีที่ให้จำนวนสถานการณ์ที่มีค่าสัมประสิทธิ์ความเชื่อมั่นที่ได้จากการจำลองไม่ต่ำกว่าค่าสัมประสิทธิ์ความเชื่อมั่นที่กำหนดจากมากไปน้อยดังนี้ อันดับแรกคือวิธี R-KS รองลงมาคือวิธี R-NLW วิธี B-KS วิธี B-NLW วิธี R-NHKB และวิธี B-NHKB ตามลำดับแต่ในทางตรงกันข้ามโดยส่วนใหญ่ของสถานการณ์ทั้งหมด วิธีความถดถอยบูตสแตรปแบบริดจ์เป็นวิธีที่ให้ค่าความยาวเฉลี่ยของช่วงความเชื่อมั่นต่ำที่สุด เนื่องจากการประมาณค่าด้วยวิธีความถดถอยบูตสแตรปแบบริดจ์ให้ค่าความแปรปรวนต่ำกว่าวิธีความถดถอยแบบริดจ์ จึงส่งผลให้ค่าความยาวเฉลี่ยของช่วงความเชื่อมั่นจากวิธีความถดถอยบูตสแตรปแบบริดจ์มีค่าต่ำกว่า และเมื่อหลังจากคัดเลือกวิธีการประมาณค่าที่ให้ค่าระดับความเชื่อมั่นของการจำลองไม่ต่ำกว่าค่าสัมประสิทธิ์ความเชื่อมั่นที่กำหนดแล้ว พบว่า 66% ของสถานการณ์จำลองทั้งหมด วิธีความถดถอยบูตสแตรปแบบริดจ์มีค่าความยาวเฉลี่ยของช่วงความเชื่อมั่นต่ำกว่าวิธีความถดถอยแบบริดจ์ เมื่อพิจารณาจากวิธีการประมาณค่า k วิธีที่ให้จำนวนของสถานการณ์ที่มีค่าความยาวเฉลี่ยของช่วงความเชื่อมั่นต่ำที่สุดจำนวนมากที่สุดคือวิธี B-NHKB รองลงมาคือวิธี B-KS วิธี R-NHKB วิธี R-KS วิธี R-NLW และวิธี B-NLW ตามลำดับ ปัจจัยที่ส่งผลต่อการเปลี่ยนแปลงของตัวประมาณค่าแบบช่วง สามารถสรุปได้ดังนี้

1. ในกรณีที่ปัจจัยอื่นๆ คงที่ พบว่าค่าความยาวเฉลี่ยของช่วงความเชื่อมั่นมีแนวโน้มลดลงเมื่อขนาดตัวอย่างเพิ่มขึ้น และค่าสัมประสิทธิ์ความเชื่อมั่นที่ได้จากการจำลองมีแนวโน้มมีค่าเพิ่มขึ้น เมื่อขนาดตัวอย่างเพิ่มขึ้น ส่งผลให้ค่าสัมประสิทธิ์ความเชื่อมั่นที่ได้จากการจำลองมีแนวโน้มให้ค่าไม่ต่ำกว่าค่าสัมประสิทธิ์ความเชื่อมั่นที่กำหนด

2. ในกรณีที่ปัจจัยอื่นๆ คงที่ พบว่าค่าความยาวเฉลี่ยของช่วงความเชื่อมั่นมีแนวโน้มเพิ่มขึ้นเมื่อระดับความสัมพันธ์ระหว่างตัวแปรอิสระมีค่าเพิ่มขึ้น

3. ในกรณีที่ปัจจัยอื่นๆ คงที่ พบว่าค่าความยาวเฉลี่ยของช่วงความเชื่อมั่นมีแนวโน้มเพิ่มขึ้น เมื่อส่วนเบี่ยงเบนมาตรฐานของความคลาดเคลื่อนมีค่าเพิ่มขึ้น

4. ในกรณีที่ปัจจัยอื่นๆ คงที่ พบว่าค่าความยาวเฉลี่ยของช่วงความเชื่อมั่นมีแนวโน้มเพิ่มขึ้น เมื่อค่าสัมประสิทธิ์ความเชื่อมั่นที่กำหนดมีค่าเพิ่มขึ้น

สรุปผลการวิจัย

การประมาณค่าแบบจุด

จากสถานการณ์ของการจำลองทั้งหมด โดยส่วนใหญ่วิธีความถดถอยบูตสแตรปแบบริดจ์เป็นวิธีที่มีค่า MSE ของตัวประมาณค่าสัมประสิทธิ์ความถดถอยเชิงพหุต่ำกว่าวิธีความถดถอยแบบริดจ์ โดยที่ค่าความแปรปรวนของตัวประมาณค่าที่ได้ด้วยวิธีความถดถอยบูตสแตรปแบบริดจ์มีค่าน้อยกว่าวิธีความถดถอยแบบริดจ์ แต่ค่าความเอนเอียงที่ได้จะมากกว่าวิธีความถดถอยแบบริดจ์ ซึ่งโดยส่วนใหญ่ของสถานการณ์ทั้งหมด วิธี B-NHKB เป็นวิธีการประมาณค่าที่ดีสุด โดยที่ค่า MSE จะมีค่าแปรผันตามระดับความสัมพันธ์ระหว่างตัวแปรอิสระ ส่วนเบี่ยงเบนมาตรฐานของความคลาดเคลื่อน และจำนวนตัวแปรอิสระ แต่จะแปรผกผันกับขนาดตัวอย่าง

การประมาณค่าแบบช่วง

จากสถานการณ์ของการจำลองทั้งหมดการประมาณค่าด้วยวิธีความถดถอยแบบริดจ์เป็นวิธีที่มีสถานการณ์ที่ให้ค่าสัมประสิทธิ์ความเชื่อมั่นที่ได้จากการจำลองไม่ต่ำกว่าค่าสัมประสิทธิ์ความเชื่อมั่นที่กำหนดมากกว่าวิธีความถดถอยบูตสแตรปแบบริดจ์ แต่ในทางตรงกันข้ามโดยส่วนใหญ่ของสถานการณ์จำลองทั้งหมด วิธีความถดถอยบูตสแตรปแบบริดจ์เป็นวิธีที่ให้ค่าความยาวเฉลี่ยของช่วงความเชื่อมั่นต่ำสุดหลังจากคัดเลือกวิธีการประมาณค่าที่ให้ค่าระดับความเชื่อมั่นของการจำลองไม่ต่ำกว่าค่าสัมประสิทธิ์ความเชื่อมั่นที่กำหนดแล้ว

ข้อเสนอแนะ

1. ในด้านการนำไปใช้ประโยชน์วิธีความถดถอยบูตสแตรปแบบริดจ์ ควรที่จะนำมาใช้เมื่อเกิดความสัมพันธ์ระหว่างตัวแปรอิสระค่อนข้างสูงและขนาดตัวอย่างมีขนาดใหญ่

2. ในด้านการวิจัยเพื่อเป็นแนวทางให้ผู้ที่สนใจศึกษาเพิ่มเติม ควรขยายผลการวิจัยออกไปให้เกิดประโยชน์ยิ่งขึ้น โดยทำการศึกษาในสถานการณ์อื่นๆ นอกเหนือจากงานวิจัยครั้งนี้

กิตติกรรมประกาศ

วิทยานิพนธ์ฉบับนี้สำเร็จลุล่วงไปได้ด้วยความช่วยเหลือและเอาใจใส่อย่างดียิ่งของอาจารย์ที่ปรึกษาวิทยานิพนธ์ ผู้วิจัยขอกราบขอบพระคุณต่อท่านอาจารย์เป็นอย่างสูง ที่กรุณาให้คำปรึกษาและแนะนำเกี่ยวกับวิทยานิพนธ์ด้วยดีเสมอมา ในที่นี้ผู้วิจัยขอกราบขอบพระคุณอาจารย์ที่เป็นประธานกรรมการสอบวิทยานิพนธ์และกรรมการสอบวิทยานิพนธ์ทุกท่าน ที่กรุณาให้คำแนะนำ ตรวจสอบ และแก้ไขวิทยานิพนธ์ฉบับนี้ให้สมบูรณ์ยิ่งขึ้น สุดท้ายนี้ผู้วิจัยขอกราบขอบพระคุณ บิดา มารดา และพี่ชายทั้งสองคน ที่ช่วยส่งเสริม สนับสนุนและให้กำลังใจเสมอมา รวมทั้งขอบคุณ ญาติๆ พี่ๆ น้องๆ และเพื่อนๆ ทุกคนที่คอยช่วยเหลือและให้กำลังใจมาโดยตลอด

เอกสารอ้างอิง

Efron, B. and Tibshirani, R. 1993. An Introduction to the Bootstrap. Chapman and Hall, New York.

Hoerl, A.E., and Kennard, R.W., 1970. Ridge regression:Biased estimation for nonorthogonal problems. Technometrics. 12: 55-69.

Khalaf, G., and Shukur, G. 2005. Choosing ridge parameters for regression problems. Communications in Statistics—Theory and Methods A. 34:1177–1182.

Kibria, B. M. Golam. 2003. Performance of some new ridge regression estimators. Communications in Statistics—Theory and Methods. 32:419–435.

Ozkale, M.R. 2008. A jackknifed ridge estimator in linear regression model with heteroscedastic or correlated errors. Statistics and Probability Letters. 78:3159-3169.

Quenouille, M.H., 1956. Notes on bias in estimation. Biometrika. 43:353-360.