นิพนธ์ต้นฉบับ

การเปรียบเทียบวิธีการประมาณค่าพารามิเตอร์ในการวิเคราะห์การถดถอยพหุคูณ เมื่อเกิดพหุสัมพันธ์ระหว่างตัวแปรอิสระ

อังคณา ฮึกหาญสู้ศัตรู[1]

Abstract

Hukharnsusatrue, A.

A comparison on parameter-estimation methods in multiple regression analysis with existence of multicollinearity among independent variables

Songklanakarin J. Sci. Technol., 2005, 27(6) : 1311-1325

The objective of this research is to compare multiple regression coefficients estimating methods with existence of multicollinearity among independent variables. The estimation methods are Ordinary Least Squares method (OLS), Restricted Least Squares method (RLS), Restricted Ridge Regression method (RRR) and Restricted Liu method (RL) when restrictions are true and restrictions are not true. The study used the Monte Carlo Simulation method. The experiment was repeated 1,000 times under each situation. The analyzed results of the data are demonstrated as follows.

CASE 1: The restrictions are true.

In all cases, RRR and RL methods have a smaller Average Mean Square Error (AMSE) than OLS and RLS method, respectively. RRR method provides the smallest AMSE when the level of correlations is high and also provides the smallest AMSE for all level of correlations and all sample sizes when standard deviation is equal to 5. However, RL method provides the smallest AMSE when the level of correlations is low and middle, except in the case of standard deviation equal to 3, small sample sizes, RRR method provides the smallest AMSE.

Bank of Ayudhya, 1222 Rama III Road, Bang Phongphang, Yan Nawa, Bangkok 10120 Thailand.

---

[1]สต.ม.(สถิติ) เจ้าหน้าที่วิเคราะห์สินเชื่อ 3 ฝ่ายวิเคราะห์สินเชื่อ ธนาคารกรุงศรีอยุธยา จำกัด (มหาชน) 1222 ถนนพระรามที่ 3 แขวงบางโพงพาง เขตยานนาวา กรุงเทพฯ 10120

E-mail: h_aungkana@hotmail.com

รับต้นฉบับ 9 สิงหาคม 2547 รับลงพิมพ์ 21 เมษายน 2548

The AMSE varies with, most to least, respectively, level of correlations, standard deviation and number of independent variables but inversely with to sample size.

CASE 2: The restrictions are not true.

In all cases, RRR method provides the smallest AMSE, except in the case of standard deviation equal to 1 and error of restrictions equal to 5%, OLS method provides the smallest AMSE when the level of correlations is low or median and there is a large sample size, but the small sample sizes, RL method provides the smallest AMSE. In addition, when error of restrictions is increased, OLS method provides the smallest AMSE for all level, of correlations and all sample sizes, except when the level of correlations is high and sample sizes small. Moreover, the case OLS method provides the smallest AMSE, the most RLS method has a smaller AMSE than RRR and RL methods when the level of correlations is low or median and sample sizes are large.

The AMSE varies with, most to least, respectively, error of restrictions, level of correlations, standard deviation and number of independent variables but inversely with to sample sizes, except that error of restrictions does not affect AMSE of OLS method.

Key words : multicollinearity, Ordinary Least Squares, Restricted Least Squares, Restricted Ridge Regression, Restricted Liu

บทคัดย่อ

อังคณา ฮึกหาญสู้ศัตรู

การเปรียบเทียบวิธีการประมาณค่าพารามิเตอร์ในการวิเคราะห์การถดถอยพหุคูณ เมื่อเกิดพหุสัมพันธ์ระหว่างตัวแปรอิสระ

ว. สงขลานครินทร์ วทท. 2548 27(6) : 1311-1325

การวิจัยครั้งนี้มีวัตถุประสงค์เพื่อเปรียบเทียบวิธีการประมาณค่าสัมประสิทธิ์การถดถอยพหุคูณเมื่อเกิดพหุสัมพันธ์ระหว่างตัวแปรอิสระ โดยการเปรียบเทียบวิธีกำลังสองน้อยสุด (Ordinary Least Squares method (OLS)) วิธีกำลังสองน้อยสุดที่ถูกจำกัด (Restricted Least Squares method (RLS)) วิธีริดจ์รีเกรสชันที่ถูกจำกัด (Restricted Ridge Regression method (RRR)) และวิธีลิวที่ถูกจำกัด (Restricted Liu method (RL)) เมื่อข้อจำกัดเป็นจริงและข้อจำกัดไม่เป็นจริง ในการวิจัยครั้งนี้ได้ทำการจำลองแบบด้วยเทคนิคมอนติคาร์โลซึ่งกระทำซ้ำ 1,000 ครั้งในแต่ละสถานการณ์ ผลของการวิจัยสามารถสรุปได้ดังนี้

กรณีที่ 1 ข้อจำกัดเป็นจริง

ในทุกกรณี วิธี RRR และ RL จะให้ค่า AMSE น้อยกว่าวิธี OLS และ RLS ตามลำดับ โดยวิธี RRR ให้ค่า AMSE น้อยที่สุดเมื่อระดับความสัมพันธ์สูง และจะให้ค่า AMSE น้อยที่สุดในทุกระดับความสัมพันธ์ ทุกขนาดตัวอย่าง เมื่อส่วนเบี่ยงเบนมาตรฐานเท่ากับ 5 ส่วนวิธี RL จะให้ค่า AMSE น้อยที่สุดเมื่อระดับความสัมพันธ์ต่ำและปานกลาง ยกเว้นในกรณีที่ส่วนเบี่ยงเบนมาตรฐานเท่ากับ 3 ขนาดตัวอย่างน้อย วิธี RRR ให้ค่า AMSE น้อยที่สุด

ค่า AMSE แปรผันตามปัจจัยต่อไปนี้จากมากไปน้อยคือ ระดับความสัมพันธ์ ส่วนเบี่ยงเบนมาตรฐาน จำนวนตัวแปรอิสระ แต่แปรผกผันกับขนาดตัวอย่าง

กรณีที่ 2 ข้อจำกัดไม่เป็นจริง

ในทุกกรณี วิธี RRR ให้ค่า AMSE น้อยที่สุด ยกเว้นในกรณีที่ส่วนเบี่ยงเบนมาตรฐานเท่ากับ 1 ความคลาดเคลื่อนของข้อจำกัด 5% วิธี OLS จะให้ค่า AMSE น้อยที่สุดเมื่อระดับความสัมพันธ์ต่ำและปานกลาง ขนาดตัวอย่างมาก แต่ถ้าขนาดตัวอย่างน้อย วิธี RL จะให้ค่า AMSE น้อยที่สุด และเมื่อความคลาดเคลื่อนของข้อจำกัดเพิ่มขึ้น วิธี OLS จะให้ค่า AMSE น้อยที่สุดในทุกระดับความสัมพันธ์ ทุกขนาดตัวอย่าง ยกเว้นที่ระดับความสัมพันธ์สูง ขนาดตัวอย่างน้อย วิธี OLS ให้ค่า AMSE น้อยที่สุดและวิธี RLS จะให้ค่า AMSE รองลงมาเป็นส่วนใหญ่เมื่อระดับความสัมพันธ์ต่ำและปานกลาง ขนาดตัวอย่างมาก

ค่า AMSE แปรผันตามปัจจัยต่อไปนี้จากมากไปน้อยคือ ความคลาดเคลื่อนของข้อจำกัด ระดับความสัมพันธ์ ส่วนเบี่ยงเบนมาตรฐาน และจำนวนตัวแปรอิสระ แต่แปรผกผันกับขนาดตัวอย่าง ยกเว้นความคลาดเคลื่อนของข้อจำกัดไม่มีผลต่อค่า AMSE วิธี OLS

ในการวิเคราะห์การถดถอยพหุคูณมีหลักเกณฑ์ว่า การใช้ตัวแปรอิสระที่เหมาะสมมากกว่าหนึ่งตัวขึ้นไปมาช่วยในการอธิบายตัวแปรตามจะทำให้ผลการประมาณค่าตัวแปรตามมีความถูกต้องมากกว่าการใช้ตัวแปรอิสระเพียงตัวเดียวถ้าตัวแปรอิสระนั้นไม่มีความสัมพันธ์กันและมีอิทธิพลต่อตัวแปรตามมากพอสมควร ตัวแบบแสดงความสัมพันธ์ระหว่างตัวแปรอิสระและตัวแปรตามอยู่ในรูปของ

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

เมื่อ $\underset{\sim}{y}$ เป็นเวกเตอร์ของตัวแปรตามขนาด $n \times 1$, X เป็นเมทริกซ์ของตัวแปรอิสระขนาด $n \times (p+1)$, $\underset{\sim}{\beta}$ เป็นเวกเตอร์ของสัมประสิทธิ์การถดถอยพหุคูณขนาด $(p+1) \times 1$ และ $\underset{\sim}{\varepsilon}$ เป็นเวกเตอร์ของค่าความคลาดเคลื่อนที่เกิดขึ้นขนาด $n \times 1$ โดยที่ $E(\underset{\sim}{\varepsilon}) = \underset{\sim}{0}$ และ $\mathrm{cov}(\underset{\sim}{\varepsilon}) = \sigma^2 I_n$

ในการประมาณค่าสัมประสิทธิ์การถดถอยพหุคูณจากตัวแบบดังกล่าว วิธีที่นิยมใช้กันมากคือวิธีกำลังสองน้อยสุด (ordinary least squares method) ซึ่งมีคุณสมบัติเป็นตัวประมาณไม่เอนเอียงและให้ความแปรปรวนต่ำสุดในบรรดาตัวประมาณไม่เอนเอียงเชิงเส้น โดยวิธีกำลังสองน้อยสุดมีข้อสมมติที่ว่าตัวแปรอิสระแต่ละตัวต้องไม่มีความสัมพันธ์เชิงเส้นกับตัวแปรอิสระตัวอื่น แต่ในทางปฏิบัติเป็นไปได้ยากเพราะตัวแปรอิสระต่างๆ ที่นำมาศึกษาส่วนใหญ่มักมีความสัมพันธ์กัน กล่าวคือ ตัวแปรอิสระบางตัวมีพหุสัมพันธ์ (multicollinearity) ทำให้ $|X'X|$ มีค่าเข้าใกล้ศูนย์ส่งผลทำให้ความแปรปรวนของตัวประมาณสัมประสิทธิ์การถดถอยพหุคูณวิธีกำลังสองน้อยสุดมีค่ามาก การประมาณค่าที่ได้ไม่เหมาะสมและขาดความเที่ยงตรง จึงได้มีวิธีพื้นฐานที่สร้างขึ้นเพื่อที่จะปรับปรุงความเที่ยงตรงของตัวประมาณสัมประสิทธิ์วิธีกำลังสองน้อยสุดคือ วิธีกำลังสองน้อยสุดที่ถูกจำกัด (restricted least squares method) โดยหลักการของวิธีนี้ยังคงใช้หลักการของวิธีกำลังสองน้อยสุดเหมือนเดิมเพียงแต่กระทำภายใต้เงื่อนไขที่ทราบข้อจำกัด (restriction) เกี่ยวกับการรวมเชิงเส้น (linear combination) ของสัมประสิทธิ์การถดถอยพหุคูณที่ไม่ทราบค่า ตัวประมาณที่ได้จะเป็นตัวประมาณไม่เอนเอียง (unbiased estimator) ในกรณีที่ข้อจำกัดเป็นจริง แต่เป็นตัวประมาณเอนเอียง (biased estimator) ในกรณีที่ข้อจำกัดไม่เป็นจริงและให้ค่าเฉลี่ยความคลาดเคลื่อนกำลังสองต่ำกว่าวิธีกำลังสองน้อยสุดในกรณีที่ข้อจำกัดเป็นจริงและไม่เป็นจริง

Hoerl และ Kennard (1970) ได้ศึกษาวิธีการประมาณค่าสัมประสิทธิ์การถดถอยพหุคูณที่ให้ค่าเฉลี่ยความคลาดเคลื่อนกำลังสองต่ำกว่าวิธีกำลังสองน้อยสุดคือ วิธีริดจ์รีเกรสชัน (Ridge Regression method) โดยตัวประมาณที่ได้จะเป็นตัวประมาณเอนเอียงและให้ค่าเฉลี่ยความคลาดเคลื่อนกำลังสองต่ำกว่าวิธีกำลังสองน้อยสุด Sarkar (1992) ได้นำเอาข้อดีของวิธีริดจ์รีเกรสชันและวิธีกำลังสองน้อยสุดที่ถูกจำกัดมาผสมผสานกันทำให้ตัวประมาณที่ได้มีความเที่ยงตรงมากขึ้นเรียกวิธีการนี้ว่าวิธีริดจ์รีเกรสชันที่ถูกจำกัด (restricted ridge regression method) ซึ่งมีคุณสมบัติเป็นตัวประมาณเอนเอียงและค่าเฉลี่ยความคลาดเคลื่อนกำลังสองต่ำกว่าวิธีกำลังสองน้อยสุดที่ถูกจำกัดและวิธีริดจ์รีเกรสชันในกรณีที่ข้อจำกัดเป็นจริง นอกจากนี้ในกรณีที่ข้อจำกัดไม่เป็นจริงได้มีการกำหนดเงื่อนไขของค่า k ที่เป็นไปได้ที่ทำให้ค่าเฉลี่ยความคลาดเคลื่อนกำลังสองต่ำกว่าวิธีกำลังสองน้อยสุดที่ถูกจำกัดและวิธีริดจ์รีเกรสชัน ต่อมาในปี ค.ศ.1993 ลิวคีเจียน (Liu Kejian) ได้นำเอาข้อดีของวิธีริดจ์รีเกรสชันและวิธีของสไตน์ (Stein) มาผสมผสานกันเรียกวิธีนี้ว่าวิธีลิว (Liu method) โดยตัวประมาณที่ได้จะเป็นตัวประมาณเอนเอียงและให้ค่าเฉลี่ยความคลาดเคลื่อนกำลังสองต่ำกว่าวิธีกำลังสองน้อยสุด จนกระทั่ง Kaciranlar และคณะ (1999) ได้นำเอาข้อดีของวิธีลิวและวิธีกำลังสองน้อยสุดที่ถูกจำกัดมาผสมผสานกันเรียกวิธีนี้ว่าวิธีลิวที่ถูกจำกัด (restricted Liu method) โดยตัวประมาณที่ได้จะเป็นตัวประมาณเอนเอียงและให้ค่าเฉลี่ยความคลาดเคลื่อนกำลังสองต่ำกว่าวิธีกำลังสองน้อยสุดที่ถูกจำกัดและวิธีลิวในกรณีที่ข้อจำกัดเป็นจริง นอกจากนี้ได้กำหนดเงื่อนไขของค่า d ที่ทำให้วิธีลิวที่ถูก

จำกัดมีค่าเฉลี่ยความคลาดเคลื่อนกำลังสองต่ำกว่าวิธีกำลังสองน้อยสุดที่ถูกจำกัดและวิธีลิวในกรณีที่ข้อจำกัดไม่เป็นจริง

จากที่กล่าวมาทั้งหมดข้างต้นผู้วิจัยจึงสนใจเปรียบเทียบวิธีประมาณค่าสัมประสิทธิ์การถดถอยพหุคูณ 4 วิธี คือ วิธีกำลังสองน้อยสุด วิธีกำลังสองน้อยสุดที่ถูกจำกัด วิธีริดจ์รีเกรสชันที่ถูกจำกัด และวิธีลิวที่ถูกจำกัด เพื่อศึกษาว่าการประมาณค่าสัมประสิทธิ์การถดถอยพหุคูณด้วยวิธีใดจะให้ค่าเฉลี่ยความคลาดเคลื่อนกำลังสองน้อยที่สุดเมื่อเกิดพหุสัมพันธ์ระหว่างตัวแปรอิสระ

1. ทฤษฎีและตัวสถิติที่เกี่ยวข้อง

ในที่นี้จะกล่าวถึงการประมาณค่าสัมประสิทธิ์การถดถอยพหุคูณทั้ง 4 วิธี ได้แก่ วิธีกำลังสองน้อยสุด (Ordinary Least Squares (OLS) method) วิธีกำลังน้อยสุดที่ถูกจำกัด (restricted least squares (RLS) method) วิธีริดจ์รีเกรสชันที่ถูกจำกัด (restricted ridge regression (RRR) method) วิธีลิวที่ถูกจำกัด (restricted Liu (RL) method) ซึ่งรายละเอียดต่างๆ มีดังนี้

1.1 วิธีกำลังสองน้อยสุด (OLS)

หลักการของวิธีกำลังสองน้อยสุดคือการเลือกค่า $\underset{\sim}{\beta}$ ซึ่งทำให้ $(\underset{\sim}{y}-X\underset{\sim}{\beta})'(\underset{\sim}{y}-X\underset{\sim}{\beta})$ มีค่าน้อยที่สุด ซึ่งค่าของ $\underset{\sim}{\beta}$ ดังกล่าวจะได้จากการหาอนุพันธ์รวม (total differential) แล้วกำหนดให้เท่ากับศูนย์ ตัวประมาณที่ได้จะมีรูปแบบดังนี้

$$\hat{\underset{\sim}{\beta}} = (X'X)^{-1}X'\underset{\sim}{y} \tag{1.1.1}$$

ตัวประมาณ $\hat{\underset{\sim}{\beta}}$ ในสมการ (1.1.1) ที่ได้จากการประมาณค่าด้วยวิธีกำลังสองน้อยสุดเป็นตัวประมาณไม่เอนเอียงเนื่องจาก $E(\hat{\underset{\sim}{\beta}}) = \underset{\sim}{\beta}$ และมีความแปรปรวนต่ำสุดในบรรดาตัวประมาณไม่เอนเอียงเชิงเส้น นอกจากนี้เมื่อตัวแปรอิสระมีความสัมพันธ์กันคือมีสภาพไม่เหมาะสม (ill-conditioned) เราสามารถพิจารณาผลของตัวแปรอิสระที่มีพหุสัมพันธ์กัน (multicollinearity) กันโดยพิจารณาจากคุณสมบัติ 2 ประการคือ เมทริกซ์ความแปรปรวนร่วมของตัวประมาณ $\hat{\underset{\sim}{\beta}}$ และค่าเฉลี่ยความคลาดเคลื่อนกำลังสองของความแตกต่างระหว่าง $\hat{\underset{\sim}{\beta}}$ และ $\underset{\sim}{\beta}$ ดังต่อไปนี้

$$\text{cov}(\hat{\underset{\sim}{\beta}}) = \sigma^2(X'X)^{-1} \tag{1.1.2}$$

ให้ L_1 คือความแตกต่างระหว่าง $\hat{\underset{\sim}{\beta}}$ และ $\underset{\sim}{\beta}$ ดังนั้น

$$L_1^2 = (\hat{\underset{\sim}{\beta}}-\underset{\sim}{\beta})'(\hat{\underset{\sim}{\beta}}-\underset{\sim}{\beta}) \tag{1.1.3}$$

$$E(L_1^2) = \sigma^2 tr(X'X)^{-1} \tag{1.1.4}$$

ถ้า $\underset{\sim}{\varepsilon}$ มีการแจกแจงปกติจะได้ว่า

$$Var(L_1^2) = 2\sigma^4 tr(X'X)^{-2} \tag{1.1.5}$$

จากสมการ (1.1.4) และ (1.1.5) เราจะพบว่า $E(L_1^2)$ และ $Var(L_1^2)$ ต่างอยู่ในรูปฟังก์ชันผลบวกของสมาชิกในแนวทแยงมุมของเมทริกซ์ $X'X$ ดังนั้นเพื่อความสะดวกในการเปรียบเทียบค่า $\hat{\underset{\sim}{\beta}}$ จึงควรแปลงให้อยู่ในรูปฟังก์ชันของค่าลักษณะเฉพาะ (eigenvalue) ของเมทริกซ์ $X'X$ จะได้ว่า

$$E(L_1^2) = \sigma^2 \sum_{i=1}^{p+1} \frac{1}{\lambda_i}$$

และ $$Var(L_1^2) = 2\sigma^4 \sum_{i=1}^{p+1} \left(\frac{1}{\lambda_i}\right)^2$$

เนื่องจากในกรณีที่ตัวแปรอิสระมีสภาพไม่เหมาะสม กล่าวคือ ค่าลักษณะเฉพาะบางค่าของเมทริกซ์ $X'X$ มีค่าน้อยมากจะทำให้การประมาณค่า $\underset{\sim}{\beta}$ ด้วยวิธีกำลังสองน้อยสุดให้ค่าเฉลี่ยความคลาดเคลื่อนกำลังสองที่มีค่าสูงขึ้น ดังนั้นตัวประมาณ $\hat{\underset{\sim}{\beta}}$ จึงเป็นตัวประมาณไม่เหมาะสม

1.2 วิธีกำลังสองน้อยสุดที่ถูกจำกัด (RLS)

วิธีกำลังสองน้อยสุดที่ถูกจำกัดมีหลักการเหมือนวิธีกำลังสองน้อยสุดเพียงแต่เพิ่มข้อมูลข้อจำกัดเกี่ยวกับการรวมเชิงเส้นของสัมประสิทธิ์การถดถอยพหุคูณที่ไม่ทราบค่าดังนี้

$$R\underset{\sim}{\beta} = \underset{\sim}{r} \tag{1.2.1}$$

เมื่อ R เป็นเมทริกซ์แสดงรูปแบบความสัมพันธ์ของพารามิเตอร์ $\underset{\sim}{\beta}$ ขนาด $q\times(p+1)$ โดยที่ $q < (p+1)$ $\underset{\sim}{\beta}$ เป็นเวกเตอร์ของสัมประสิทธิ์การถดถอยพหุคูณขนาด $(p+1)\times 1$ และ $\underset{\sim}{r}$ เป็นเวกเตอร์ที่ได้จากการรวมเชิงเส้นของสัมประสิทธิ์

การถดถอยพหุคูณขนาด $q \times 1$ ซึ่งตัวประมาณวิธีกำลังสองน้อยสุดที่ถูกจำกัดจะมีรูปแบบดังนี้

$$\underset{\sim}{\beta}^* = \hat{\underset{\sim}{\beta}} + S^{-1}R'(RS^{-1}R')^{-1}(\underset{\sim}{r} - R\hat{\underset{\sim}{\beta}}) \qquad (1.2.2)$$

โดยที่ $R = X'X$ พิจารณาค่าคาดหวังของตัวประมาณจะได้ว่า $E(\underset{\sim}{\beta}^*) = \underset{\sim}{\beta} + S^{-1}R'[RS^{-1}R']^{-1}\underset{\sim}{\delta}$ เมื่อ $r - R\underset{\sim}{\beta} = \underset{\sim}{\delta} = \underset{\sim}{0}$ เป็นกรณีที่ข้อจำกัดนั้นเป็นจริง ตัวประมาณ $\underset{\sim}{\beta}^*$ จะเป็นตัวประมาณไม่เอนเอียง (unbiased estimator) ส่วนในกรณีที่ข้อจำกัดนั้นไม่เป็นจริงคือ $r - R\underset{\sim}{\beta} = \underset{\sim}{\delta} \neq \underset{\sim}{0}$ ตัวประมาณ $\underset{\sim}{\beta}^*$ จะเป็นตัวประมาณเอนเอียง (biased estimator) นอกจากนี้เมื่อพิจารณาความแปรปรวนร่วมของ $\underset{\sim}{\beta}^*$ จะได้ว่า

$$\mathrm{cov}(\underset{\sim}{\beta}^*) = \sigma^2(S^{-1} - S^{-1}R'(RS^{-1}R')^{-1}RS^{-1})$$

โดยเมื่อพิจารณาความแตกต่างของความแปรปรวนร่วมระหว่าง $\underset{\sim}{\beta}^*$ กับ $\hat{\underset{\sim}{\beta}}$ จะพบว่า $\mathrm{cov}(\hat{\underset{\sim}{\beta}}) - \mathrm{cov}(\underset{\sim}{\beta}^*) = \sigma^2 [S^{-1}R'(RS^{-1}R')^{-1}RS^{-1}]$ เป็นเมทริกซ์กึ่งบวกแน่นอน (positive semidefinite matrix) ทำให้ตัวประมาณ $\underset{\sim}{\beta}^*$ ที่ได้จากวิธีกำลังสองน้อยสุดที่ถูกจำกัดมีความแปรปรวนน้อยกว่าตัวประมาณ $\underset{\sim}{\beta}$ ที่ได้จากวิธีกำลังสองน้อยสุด ค่าเฉลี่ยความคลาดเคลื่อนกำลังสองของ $\underset{\sim}{\beta}^*$ จะอยู่ในรูปของ

$$MSE(\underset{\sim}{\beta}^*) = \sigma^2 tr(A) + \underset{\sim}{\delta}^{*\prime} S^{-2} \underset{\sim}{\delta}^* \qquad (1.2.3)$$

เมื่อ $A = S^{-1} - S^{-1}R'(RS^{-1}R')^{-1}RS^{-1}, \underset{\sim}{\delta} = r - R\underset{\sim}{\beta}$ และ $\underset{\sim}{\delta}^* = R'(RS^{-1}R')^{-1}\underset{\sim}{\delta}$

เมื่อพิจารณา $MSE(\underset{\sim}{\beta}^*)$ จะพบว่าในกรณีที่ข้อจำกัดนั้นเป็นจริง $\underset{\sim}{\delta} = \underset{\sim}{0}$ จะได้ว่า

$$MSE(\underset{\sim}{\beta}^*) = \sigma^2 tr(A) \qquad (1.2.4)$$

1.3 วิธีริดจ์รีเกรสชันที่ถูกจำกัด (RRR)

หลักการของวิธีนี้คือ ต้องการลดค่าเฉลี่ยความคลาดเคลื่อนกำลังสองของการประมาณ $\underset{\sim}{\beta}^*$ ให้ต่ำลงโดยใช้หลักการของวิธีริดจ์รีเกรสชัน เมื่อพิจารณาค่าเฉลี่ยความคลาดเคลื่อนกำลังสองของการประมาณ $\underset{\sim}{\beta}^*$ การที่จะลดค่าเฉลี่ยความคลาดเคลื่อนกำลังสองให้ต่ำลงจะต้องลดค่า $(X'X)^{-1}$ ให้ต่ำลง ซึ่งก็คือทำให้ $|X'X|$ มีค่าเพิ่มมากขึ้น โดยการบวกค่าคงที่ที่มากกว่าศูนย์กับสมาชิกทุกตัวบนเส้นทแยงมุมของเมทริกซ์ $X'X$ ตัวประมาณริดจ์รีเกรสชันที่ถูกจำกัด $(\underset{\sim}{\beta}_R^*(k))$ จะอยู่ในรูปของ

$$\underset{\sim}{\beta}_R^*(k) = (X'X + kI)^{-1}(X'\underset{\sim}{y} + Sg), \quad k > 0$$

เราสามารถเขียน $\underset{\sim}{\beta}_R^*(k)$ ให้อยู่ในรูปฟังก์ชัน $\underset{\sim}{\beta}^*$ โดยแทนค่า $X'\underset{\sim}{y} + Sg = X'X\underset{\sim}{\beta}^*$ ดังนั้น

$$\underset{\sim}{\beta}_R^*(k) = [I + kS^{-1}]^{-1}\underset{\sim}{\beta}^* = W\underset{\sim}{\beta}^* \qquad (1.3.1)$$

เมื่อ $W = [I + kS^{-1}]^{-1}$ และตัวประมาณ $\underset{\sim}{\beta}_R^*(k) = \underset{\sim}{\beta}^*$ ถ้า $k = 0$

เมื่อพิจารณาค่าคาดหวังของตัวประมาณ $\underset{\sim}{\beta}_R^*(k)$ จะได้ว่า

$$E(\underset{\sim}{\beta}_R^*(k)) = W\underset{\sim}{\beta} + WS^{-1}R'(RS^{-1}R')^{-1}\underset{\sim}{\delta}$$

จากค่าคาดหวังของตัวประมาณ $\underset{\sim}{\beta}_R^*(k)$ จะพบว่าตัวประมาณ $\underset{\sim}{\beta}_R^*(k)$ เป็นตัวประมาณเอนเอียงสำหรับ $\underset{\sim}{\beta}$ ในกรณีที่ $k = 0$ และ $\underset{\sim}{\delta} = \underset{\sim}{0}$ ตัวประมาณ $\underset{\sim}{\beta}_R^*(k)$ จะเป็นตัวประมาณไม่เอนเอียง นอกจากนี้ค่าเฉลี่ยความคลาดเคลื่อนกำลังสองของ $\underset{\sim}{\beta}_R^*(k)$ จะอยู่ในรูปของ

$$MSE(\underset{\sim}{\beta}_R^*(k)) = \sigma^2 tr(WAW') + [WS^{-1}\underset{\sim}{\delta}^* - kS(k)^{-1}\underset{\sim}{\beta}]'[WS^{-1}\underset{\sim}{\delta}^* - kS(k)^{-1}\underset{\sim}{\beta}] \qquad (1.3.2)$$

เมื่อ $\underset{\sim}{\delta}^* = R'(RS^{-1}R')^{-1}\underset{\sim}{\delta}$ และ $S(k)^{-1}$ เป็นเมทริกซ์ผกผันของ $S(k) = (S + kI)$ ในกรณีที่ $\underset{\sim}{\delta} = \underset{\sim}{0}$ คือข้อจำกัดเป็นจริง $MSE(\underset{\sim}{\beta}_R^*(k))$ จะมีรูปแบบ

$$MSE(\underset{\sim}{\beta}_R^*(k)) = \sigma^2 tr(WAW') + k^2\underset{\sim}{\beta}'S(k)^{-2}\underset{\sim}{\beta} \qquad (1.3.3)$$

คุณสมบัติของตัวประมาณวิธีริดจ์รีเกรสชันที่ถูกจำกัดโดยใช้ค่าเฉลี่ยความคลาดเคลื่อนกำลังสองเป็นเกณฑ์

1. ในกรณีที่ข้อจำกัดเป็นจริง $\underset{\sim}{\delta} = \underset{\sim}{0}$

เนื่องจาก S เป็นเมทริกซ์บวกแน่นอน (positive definite matrix) ทำให้สามารถหาเมทริกซ์เชิงตั้งฉาก (orthogonal matrix) ที่ทำให้ $Q'SQ = \Lambda$ เมื่อ $\Lambda = \mathrm{diag}(\lambda_1, \lambda_2, ..., \lambda_{p+1})$ โดยที่ $QQ' = Q'Q = I$ และ $B = Q'AQ$

เมื่อ $A = S^{-1} - S^{-1}R'(RS^{-1}R')^{-1}RS^{-1}$ เป็นเมทริกซ์กึ่งบวกแน่นอน (positive semidefinite matrix) ทำให้สมาชิกในแนวทแยงมุม b_{ii} ของสมาชิก B มีค่ามากกว่าหรือเท่ากับศูนย์ทั้งหมด ใช้ประโยชน์จากคุณสมบัติของฟังก์ชันรอยเมทริกซ์ (trace function) ทำให้ค่าเฉลี่ยความคลาดเคลื่อนกำลังสองของ $\underset{\sim}{\beta}_R^*(k)$ ในสมการ (1.3.3) มีรูปแบบดังนี้

$$MSE(\underset{\sim}{\beta}_R^*(k)) = \sigma^2 \sum_{i=1}^{p+1} \lambda_i^2 b_{ii} / (\lambda_i + k)^2 + k^2 \sum_{i=1}^{p+1} \alpha_i^2 / (\lambda_i + k)^2 \quad (1.3.4)$$

เมื่อ $\underset{\sim}{\alpha} = Q'\underset{\sim}{\beta}$ ช่วงของค่า k ที่ทำให้ $MSE(\underset{\sim}{\beta}_R^*(k)) < MSE(\underset{\sim}{\beta}^*)$ สามารถหาได้จากการนำ $MSE(\underset{\sim}{\beta}_R^*(k))$ ในสมการ (1.3.4) หาอนุพันธ์อันดับที่ 1 เทียบกับ k ซึ่งผลดังกล่าวอยู่ในรูปแบบของ

$$\frac{\partial}{\partial k} MSE(\underset{\sim}{\beta}_R^*(k)) = 2\sum_{i=1}^{p+1} \frac{\lambda_i}{(\lambda_i + k)^3}(k\alpha_i^2 - \sigma^2 \lambda_i b_{ii}) \quad (1.3.5)$$

จากสมการ (1.3.5) จะเห็นได้ว่าเมื่อ $0 < k < \frac{\sigma^2}{\tilde{\alpha}_{\max}^2}$ เมื่อ $\tilde{\alpha}_{\max}^2$ เป็นค่าที่มากที่สุดของ $\tilde{\alpha}_i^2$ ซึ่ง $\tilde{\alpha}_i^2 = \frac{\alpha_i^2}{(\lambda_i b_{ii})}$ จะทำให้ตัวประมาณริดจ์รีเกรสชันที่ถูกจำกัดมีค่าเฉลี่ยความคลาดเคลื่อนกำลังสองต่ำกว่าตัวประมาณกำลังสองน้อยสุดที่ถูกจำกัดในกรณีที่ข้อจำกัดนั้นเป็นจริง

2. ในกรณีที่ข้อจำกัดไม่เป็นจริง

เนื่องจาก $WS^{-1} = S(k)^{-1}$ ทำให้ $MSE(\underset{\sim}{\beta}_R^*(k))$ ในสมการ (1.3.2) มีรูปแบบใหม่ดังนี้

$$\begin{aligned} MSE(\underset{\sim}{\beta}_R^*(k)) &= \sigma^2 tr(WAW') + [S(k)^{-1}\underset{\sim}{\delta}^* - kS(k)^{-1}\underset{\sim}{\beta}]' [S(k)^{-1}\underset{\sim}{\delta}^* - kS(k)^{-1}\underset{\sim}{\beta}] \\ &= \sum_{i=1}^{p+1} \frac{1}{(\lambda_i + k)^2}[\sigma^2\lambda_i^2 b_{ii} + k^2\alpha_i^2 + \delta_i^{*2} - 2k\alpha_i\delta_i^*) \quad (1.3.6) \end{aligned}$$

จากสมการ (1.3.7) หาอนุพันธ์อันดับที่ 1 เทียบกับ k จะได้ว่า

$$\frac{\partial}{\partial k} MSE(\underset{\sim}{\beta}_R^*(k)) = 2\sum_{i=1}^{p+1} \frac{1}{(\lambda_i + k)^3}[k(\lambda_i\alpha_i^2 + \alpha_i\delta_i^*) - (\sigma^2\lambda_i^2 b_{ii} + \delta_i^{*2} + \alpha_i\lambda_i\delta_i^*)] \quad (1.3.7)$$

จากสมการ (1.3.7) จะเห็นได้ว่าเมื่อ $0 < k < k^* < \infty$ โดยที่ $k^* = \frac{\min_i(\sigma^2\lambda_i^2 b_{ii} + \delta_i^* + \alpha_i\lambda_i\delta_i^*)}{\max_i(\lambda_i\alpha_i^2 + \alpha_i\delta_i^*)}$ จะทำให้ตัวประมาณริดจ์รีเกรสชันที่ถูกจำกัดมีค่าเฉลี่ยความคลาดเคลื่อนกำลังสองต่ำกว่าตัวประมาณกำลังสองน้อยสุดที่ถูกจำกัดในกรณีที่ข้อจำกัดนั้นไม่เป็นจริง

ในการประมาณค่าสัมประสิทธิ์การถดถอยพหุคูณด้วยวิธีริดจ์รีเกรสชันที่ถูกจำกัดจะต้องเลือกค่า k ให้มีค่าเหมาะสม เพื่อที่จะทำให้ค่าความแปรปรวนร่วมของตัวประมาณสัมประสิทธิ์การถดถอยพหุคูณลดลงและค่าเฉลี่ยความคลาดเคลื่อนกำลังสองของตัวประมาณริดจ์รีเกรสชันที่ถูกจำกัดมีค่าต่ำสุด

1.4 วิธีลิวที่ถูกจำกัด (RL)

หลักการของวิธีนี้จะใช้หลักการของตัวประมาณลิวที่จะลดค่าเฉลี่ยความคลาดเคลื่อนกำลังสองของการประมาณ $\underset{\sim}{\beta}^*$ ให้ต่ำลง กล่าวคือ การบวกค่าคงที่ค่าหนึ่งให้กับสมาชิกในแนวทแยงมุมของเมทริกซ์ $X'X$ เนื่องจากตัวประมาณริดจ์มีปัญหาในการประมาณค่าพารามิเตอร์ k ดังนั้นตัวประมาณที่ใช้หลักการของลิวจึงแก้ปัญหาดังกล่าวโดยการบวกค่าคงที่ k ที่มีค่าเท่ากับหนึ่งให้กับสมาชิกทุกตัวในแนวทแยงมุมของเมทริกซ์ $X'X$ และเทอมของ $X'\underset{\sim}{y} + Sg = X'X\underset{\sim}{\beta}^*$ จึงนำ $d\underset{\sim}{\beta}^*$ มาบวกเพราะพารามิเตอร์ d ที่ใช้หลักการของตัวประมาณสไตน์หาได้ง่าย ดังนั้นการประมาณสัมประสิทธิ์การถดถอยพหุคูณด้วยวิธีลิวที่ถูกจำกัดอยู่ในรูปของ

$$\underset{\sim}{\beta}_d^*(d) = (X'X + I)^{-1}(X'\underset{\sim}{y} + Sg + d\underset{\sim}{\beta}^*)$$

เราสามารถเขียน $\underset{\sim}{\beta}_d^*(d)$ ให้อยู่ในรูปของฟังก์ชัน $\underset{\sim}{\beta}^*$ โดยแทนค่า $X'\underset{\sim}{y} + Sg = X'X\underset{\sim}{\beta}^*$ จะได้ว่าตัวประมาณดังกล่าวอยู่ในรูปของ

$$\begin{aligned} \underset{\sim}{\beta}_d^*(d) &= (S + I)^{-1}(S + dI)\underset{\sim}{\beta}^* \\ &= F_d\underset{\sim}{\beta}^* \quad (1.4.1) \end{aligned}$$

เมื่อ $F_d = (S+I)^{-1}(S+dI)$ ถ้า $d = 1$ จะทำให้ $\underset{\sim}{\beta}_d^*(d) = \underset{\sim}{\beta}^*$ และพิจารณาค่าคาดหวังของ $\underset{\sim}{\beta}_d^*(d)$ จะได้ว่า

$$E(\underset{\sim}{\beta}_d^*(d)) = F_d \underset{\sim}{\beta} + F_d S^{-1} R'(RS^{-1}R')^{-1} \underset{\sim}{\delta}$$

จากค่าคาดหวังของตัวประมาณ $\underset{\sim}{\beta}_d^*(d)$ จะทำให้ทราบว่าตัวประมาณ $\underset{\sim}{\beta}_d^*(d)$ เป็นตัวประมาณเอนเอียง เนื่องจาก $E(\underset{\sim}{\beta}_d^*(d)) \neq \underset{\sim}{\beta}$ ในกรณีที่ $\underset{\sim}{\delta} = 0$ คือข้อจำกัดเป็นจริง และ $d = 1$ ตัวประมาณ $\underset{\sim}{\beta}_d^*(d)$ จะเป็นตัวประมาณไม่เอนเอียง นอกจากนี้ค่าเฉลี่ยความคลาดเคลื่อนกำลังสองของ $\underset{\sim}{\beta}_d^*(d)$ จะอยู่ในรูปของ

$$MSE(\underset{\sim}{\beta}_d^*(d)) = \sigma^2 tr(F_d A F_d') + \left\| F_d \underset{\sim}{\beta} + F_d S^{-1} \underset{\sim}{\delta}^* - \underset{\sim}{\beta} \right\|^2 \quad (1.4.2)$$

เมื่อ $\delta_i^* = R'(RS^{-1}R')^{-1} \underset{\sim}{\delta}$ และ $A = S^{-1} - S^{-1}R'(RS^{-1}R')^{-1} RS^{-1}$

จากสมการ (1.4.2) ในกรณีที่ข้อจำกัดนั้นเป็นจริง คือ $\underset{\sim}{\delta} = \underset{\sim}{0}$ ค่าเฉลี่ยความคลาดเคลื่อนกำลังสองของ $\underset{\sim}{\beta}_d^*(d)$ จะอยู่ในรูปของ

$$MSE(\underset{\sim}{\beta}_d^*(d)) = \sigma^2 tr(F_d A F_d') + (d-1)^2 \underset{\sim}{\beta}'(S+I)^{-2} \underset{\sim}{\beta} \quad (1.4.3)$$

คุณสมบัติของตัวประมาณลิวที่ถูกจำกัดโดยใช้ค่าเฉลี่ยความคลาดเคลื่อนกำลังสองเป็นเกณฑ์

1. ในกรณีที่ข้อจำกัดเป็นจริง $\underset{\sim}{\delta} = \underset{\sim}{0}$

ใช้ประโยชน์จากคุณสมบัติของฟังก์ชันรอยเมทริกซ์ (trace function) ทำให้ค่าเฉลี่ยความคลาดเคลื่อนกำลังสองของ $\underset{\sim}{\beta}_d^*(d)$ ในสมการ (1.4.3) มีรูปแบบดังนี้

$$MSE(\underset{\sim}{\beta}_d^*(d)) = \sigma^2 \sum_{i=1}^{p+1} \frac{(\lambda_i + d)^2}{(\lambda_1 + 1)^2} b_{ii} + (d-1)^2 \sum_{i=1}^{p+1} \frac{\alpha_i^2}{(\lambda_i + 1)^2} = f(d) \quad (1.4.4)$$

เนื่องจาก $f(d) = MSE(\underset{\sim}{\beta}_d^*(d))$ เป็นฟังก์ชันนูน เราสามารถแก้ปัญหาเพื่อหาค่าพารามิเตอร์ d ที่ทำให้ $f(d)$ มีค่าต่ำสุดได้ โดยการหาอนุพันธ์อันดับที่ 1 ของ $f(d)$ เทียบกับ d แล้วให้เท่ากับศูนย์ ซึ่งค่า d_{opt} ที่ได้จะมีเพียงคำตอบเดียว คือ

$$d_{opt} = 1 - \sigma^2 \sum_{i=1}^{p+1} \frac{b_{ii}}{\lambda_1 + 1} \Bigg/ \sum_{i=1}^{p+1} \frac{b_{ii}\sigma^2 + \alpha_i^2}{(\lambda_1 + 1)^2} \quad (1.4.5)$$

จากสมการ (1.4.5) เมื่อกำหนด $\underset{\sim}{\beta}$ และ σ^2 การใช้ $d = d_{opt}$ จะทำให้ $MSE(\underset{\sim}{\beta}_d^*(d))$ มีค่าต่ำสุด แต่ว่า d_{opt} ขึ้นอยู่กับค่าพารามิเตอร์ที่ไม่ทราบค่าดังกล่าว เนื่องจากตอนนี้เราพิจารณาที่ข้อจำกัดเป็นจริง ตัวประมาณ d_{opt} ที่ดีที่สุดควรได้มาจากการแทนที่ α_i และ σ^2 ใน d_{opt} ด้วย $\tilde{\alpha}_i$ และ $\tilde{\sigma}^2$ ตามลำดับ ที่มาจากการประมาณด้วยวิธี RLS เราเรียกการประมาณ d_{opt} ที่ได้นี้ว่า ตัวประมาณที่ให้ค่าเฉลี่ยความคลาดเคลื่อนกำลังสองต่ำสุดที่มาจาก RLS (RLS-based minimum MSE estimator) ตัวประมาณที่ได้มีรูปแบบดังนี้

$$\hat{d}_{RLS} = 1 - \tilde{\sigma}^2 \sum_{i=1}^{p+1} \frac{b_{ii}}{\lambda_1 + 1} \Bigg/ \sum_{i=1}^{p+1} \frac{b_{ii}\tilde{\sigma}^2 + \tilde{\alpha}_i^2}{(\lambda_1 + 1)^2}$$

2. ในกรณีที่ข้อจำกัดไม่เป็นจริง $\underset{\sim}{\delta} \neq \underset{\sim}{0}$

ในกรณีที่ข้อจำกัดไม่เป็นจริงตัวประมาณ $\underset{\sim}{\beta}_d^*(d)$ และ $\underset{\sim}{\beta}^*$ จะเป็นตัวประมาณค่าเอนเอียงทั้งคู่ เราสามารถแทน $F_d - I = (d-1)(S+I)^{-1}$ ในสมการ (1.4.2) ซึ่งผลดังกล่าวจะมีรูปแบบดังนี

$$MSE(\underset{\sim}{\beta}_d^*(d)) = \sigma^2 tr(F_d A F_d') + \left\| (d-1)(S+I)^{-1} \underset{\sim}{\beta} + F_d S^{-1} \underset{\sim}{\delta}^* \right\|^2$$

$$= \sum_{i=1}^{p+1} \frac{1}{(\lambda_i + 1)^2} \left[\left(\sigma^2 b_{ii} + \frac{\tilde{\delta}_i^2}{\lambda_i^2} \right) (\lambda_i + d)^2 + (d-1)^2 \alpha_i^2 + \frac{2(d-1)\alpha_i \tilde{\delta}_i (\lambda_i + d)}{\lambda_i} \right] \quad (1.4.6)$$

โดยที่ $\tilde{\delta} = Q'\underset{\sim}{\delta}^*$ จากสมการ (2.4.6) เราจะให้ $MSE(\underset{\sim}{\beta}_d^*(d)) = h(\underset{\sim}{d})$ ซึ่ง $h(d)$ เป็นฟังก์ชันนูนโดยแท้ใน d ค่า d ที่ทำให้ $MSE(\underset{\sim}{\beta}_d^*(d)) < MSE(\underset{\sim}{\beta}^*)$ นั้นพารามิเตอร์ d จะต้องมีค่าอยู่ในช่วงเปิดที่มีจุดปิดดังนี้ $d_1 = 1$ และ $d_2 = 1$ -

$$\frac{2tr\left\{(S+I)^{-1}(\sigma^2 A+S^{-1}\underset{\sim}{\delta}^*\underset{\sim}{\beta}'+S^{-2}\underset{\sim}{\delta}^*\underset{\sim}{\delta}^{*'})\right\}}{tr\{(S+I)^{-2}(\sigma^2 A+\underset{\sim}{\beta}\underset{\sim}{\beta}'+2S^{-1}\underset{\sim}{\delta}^*\underset{\sim}{\beta}'+S^{-2}\underset{\sim}{\delta}^*\underset{\sim}{\delta}^{*'})\}}$$ เมื่อ

พิจารณาจุดปิด d_2 จุดปิด d_2 สามารถมากกว่าหรือเท่ากับหรือน้อยกว่า $d_1 = 1$

2. วิธีดำเนินการวิจัย

ข้อมูลที่ใช้ในการวิจัยครั้งนี้ได้จากการจำลอง (Simulation) ขึ้นโดยใช้โปรแกรมภาษาฟอร์แทรน 77 และกระทำซ้ำ 1,000 รอบในแต่ละสถานการณ์ซึ่งมีแผนการทดลองและขั้นตอนที่ใช้ในการวิจัยดังต่อไปนี้

2.1 แผนการทดลอง

ในการวิจัยครั้งนี้กำหนดสถานการณ์ต่างๆ ที่ต้องการศึกษาดังนี้

1. เลือกกลุ่มตัวอย่างจากประชากรโดยกำหนดให้ประชากรมีการแจกแจงแบบปกติ เมื่อ $\mu = 0, \sigma = 1, 3$ และ 5

2. ขนาดตัวอย่างที่ใช้ในการศึกษาเท่ากับ 30, 50 และ 100

3. จำนวนตัวแปรอิสระที่ใช้ในการศึกษาเท่ากับ 3 และ 5

4. ความคลาดเคลื่อนของข้อจำกัดที่ใช้ในการศึกษาคือ 5%, 10% และ 15%

5. ระดับความสัมพันธ์ของตัวแปรอิสระ

เรากำหนดระดับความสัมพันธ์ของตัวแปรอิสระแบ่งเป็น 3 ระดับคือ

ระดับต่ำ ค่า ρ มีค่าอยู่ในช่วง 0.1 ถึง 0.3
ระดับปานกลาง ค่า ρ มีค่าอยู่ในช่วง 0.4 ถึง 0.6
ระดับสูง ค่า ρ มีค่าอยู่ในช่วง 0.7 ถึง 0.9

กรณีตัวแปรอิสระเท่ากับ 3 ระดับความสัมพันธ์ของตัวแปรอิสระที่ศึกษาคือ

ระดับต่ำ $\rho = (0.1, 0.2, 0.3)$
ระดับปานกลาง $\rho = (0.4, 0.5, 0.6)$
ระดับสูง $\rho = (0.7, 0.8, 0.9)$

โดยค่า ρ ในวงเล็บคือความสัมพันธ์ระหว่าง X_1 กับ X_2, X_1 กับ X_3 และ X_2 กับ X_3 ตามลำดับ

กรณีตัวแปรอิสระเท่ากับ 5 ระดับความสัมพันธ์ของตัวแปรอิสระที่ศึกษาคือ

ระดับต่ำ $\rho = (0.1, 0.2, 0.3, 0.3)$
ระดับปานกลาง $\rho = (0.4, 0.5, 0.6, 0.6)$
ระดับสูง $\rho = (0.7, 0.8, 0.9, 0.9)$

โดยค่า ρ ในวงเล็บคือความสัมพันธ์ระหว่าง X_1 กับ X_2, X_1 กับ X_3, X_2 กับ X_3 และ X_4 กับ X_5 ตามลำดับ

2.2 ขั้นตอนในการวิจัย

ขั้นตอนในการวิจัยมีดังนี้

1. สร้างข้อมูลของความคลาดเคลื่อนให้มีลักษณะตามที่ต้องการศึกษา โดยใช้วิธีของบอกซ์ (Box) และมุลเลอร์ (Muller)

2. สร้างข้อมูลของตัวแปรอิสระ (X) ให้มีระดับความสัมพันธ์ตามที่กำหนดและสร้างข้อมูลของตัวแปรตาม ($\underset{\sim}{y}$) จากรูปแบบความสัมพันธ์ $\underset{\sim}{y} = X\underset{\sim}{\beta} + \underset{\sim}{\varepsilon}$ โดยกำหนดให้ $\underset{\sim}{\beta}$ เป็นค่าคงที่ใดๆ คือ $\beta_0 = 1, \beta_1 = 1, \beta_2 = 1, \beta_3 = 1, \beta_4 = 1$ และ $\beta_5 = 1$ ซึ่งในการทดลองนี้จะใช้ข้อจำกัด $\beta_1 + \beta_2 + \beta_3$ $(R\underset{\sim}{\beta} = \underset{\sim}{r})$

3. ประมาณค่าสัมประสิทธิ์การถดถอยพหุคูณด้วยวิธี OLS, RLS, RRR และ RL ซึ่งวิธีริดจ์รีเกรสชันที่ถูกจำกัดจะต้องทำการประมาณค่าพารามิเตอร์ k โดยใช้วิธีการค้นหาข้อมูลแบบลำดับ (Sequential Search) ในกรณีที่ข้อจำกัดเป็นจริงแล้วจึงคำนวณค่าประมาณจากวิธีริดจ์รีเกรสชันที่ถูกจำกัด ส่วนวิธีลิวที่ถูกจำกัดจะต้องทำการประมาณค่าพารามิเตอร์ d โดยใช้ $\hat{d}_{RLS}$ ในกรณีที่ข้อจำกัดเป็นจริงแล้วจึงคำนวณค่าประมาณจากวิธีลิวที่ถูกจำกัด

4. หาค่าเฉลี่ยความคลาดเคลื่อนกำลังสองของสัมประสิทธิ์การถดถอยพหุคูณโดยใช้วิธี OLS, RLS, RRR และ RL ในกรณีที่ข้อจำกัดเป็นจริง พร้อมทั้งเปรียบเทียบค่าเฉลี่ยความคลาดเคลื่อนกำลังสองของแต่ละวิธีและสรุปผลที่ได้ ซึ่งสูตรในการหาค่าเฉลี่ยความคลาดเคลื่อนกำลังสองหาได้จากสูตรดังนี้

$$MSE_j = \frac{1}{1000}\sum_{i=1}^{1000}\left(\hat{\beta}_{ij} - \beta_j\right)^2$$

$$AMSE = \frac{1}{p+1}\sum_{j=1}^{p+1} MSE_j$$

เมื่อ β_j แทนตัวประมาณสัมประสิทธิ์การถดถอยพหุคูณตัวที่ j
$\hat{\beta}_{ij}$ แทนตัวประมาณสัมประสิทธิ์การถดถอย

พหุคูณตัวที่ j จากการประมาณครั้งที่ i

MSE_j แทนค่าความคลาดเคลื่อนกำลังสองเฉลี่ยของค่าประมาณสำหรับสัมประสิทธิ์การถดถอย $\tilde{\beta}_j$

AMSE แทนค่าเฉลี่ยของความคลาดเคลื่อนกำลังสองเฉลี่ยของสัมประสิทธิ์การถดถอยพหุคูณ

5. กรณีข้อจำกัดไม่เป็นจริงกลับไปทำขั้นตอนที่ 1 และ 2 อีกครั้ง โดยขั้นตอนที่ 2 สร้าง $\tilde{\beta}^{**}$ ให้เป็นข้อมูลชุดใหม่แทน $\tilde{\beta}$ ที่เป็นข้อมูลชุดเดิมเพื่อที่จะทำให้ผลรวมเชิงเส้นของข้อจำกัดที่ใช้ $(\beta_1, \beta_2, \beta_3 = 3)$ มีความคลาดเคลื่อนไปจากผลรวมเชิงเส้นของค่าพารามิเตอร์จริง $(\tilde{\beta}^{**})$ 5%, 10% และ 15% ตามลำดับ จากนั้นทำตามขั้นตอนที่ 3 และ 4 ต่อไป โดยการหาค่าพารามิเตอร์ *k* และ *d* จะใช้วิธีการค้นหาข้อมูลแบบลำดับ

3. ผลการวิจัย

ผลที่ได้จากการจำลองข้อมูลเป็นไปตาม Table 1-8

3.1 กรณีที่ข้อจำกัดเป็นจริง

จาก Table 1 เมื่อจำนวนตัวแปรอิสระเท่ากับ 3 จะพบว่า ในทุกกรณี วิธี RRR และ RL จะให้ค่า AMSE น้อยกว่าวิธี OLS และ RLS ตามลำดับ โดยวิธี RL ให้ค่า AMSE น้อยที่สุดเมื่อระดับความสัมพันธ์ต่ำและปานกลาง ภายใต้ทุกขนาดตัวอย่าง แต่วิธี RRR จะให้ค่า AMSE น้อยที่สุดเมื่อระดับความสัมพันธ์สูง ภายใต้ทุกขนาดตัวอย่าง ส่วนในกรณีที่ส่วนเบี่ยงเบนมาตรฐานเพิ่มขึ้น โดยทั่วไปวิธี RRR จะให้ค่า AMSE น้อยที่สุดในทุกขนาดตัวอย่างและทุกระดับความสัมพันธ์ยกเว้นเมื่อส่วนเบี่ยงเบนมาตรฐานเท่ากับ 3 วิธี RL ให้ค่า AMSE น้อยที่สุดเมื่อระดับความสัมพันธ์ต่ำ ภายใต้ทุกขนาดตัวอย่างและระดับความสัมพันธ์ปานกลาง เมื่อขนาดตัวอย่างเท่ากับ 50 และ 100

จาก Table 2 เมื่อจำนวนตัวแปรอิสระเท่ากับ 5 จะพบว่าในทุกกรณี วิธี RRR และ RL จะให้ค่า AMSE น้อยกว่าวิธี OLS และ RLS ตามลำดับ โดยวิธี RL ให้ค่า AMSE น้อยที่สุดเมื่อระดับความสัมพันธ์ต่ำและปานกลาง ภายใต้ทุกขนาดตัวอย่าง แต่วิธี RRR จะให้ค่า AMSE น้อยที่สุดเมื่อระดับความสัมพันธ์สูง ภายใต้ทุกขนาดตัวอย่าง ส่วนในกรณีที่ส่วนเบี่ยงเบนมาตรฐานเพิ่มขึ้น โดยทั่วไปวิธี RRR จะให้ค่า AMSE น้อยที่สุดในทุกขนาดตัวอย่างและทุกระดับความสัมพันธ์ ยกเว้นเมื่อส่วนเบี่ยงเบนมาตรฐานเท่ากับ 3 วิธี RL ให้ค่า AMSE น้อยที่สุดที่ระดับความสัมพันธ์ต่ำ เมื่อขนาดตัวอย่างเท่ากับ 50 และ 100 และระดับความสัมพันธ์ปานกลาง เมื่อขนาดตัวอย่างเท่ากับ 100

3.2 กรณีข้อจำกัดไม่เป็นจริง

จาก Table 3-5 เมื่อจำนวนตัวแปรอิสระเท่ากับ 3 จะพบว่าในกรณีข้อจำกัดมีความคลาดเคลื่อน 5% ส่วนเบี่ยงเบนมาตรฐานเท่ากับ 1 วิธี RL จะให้ค่า AMSE น้อยที่สุดเมื่อระดับความสัมพันธ์ต่ำและปานกลาง ขนาดตัวอย่างเท่ากับ 30 และ 50 แต่วิธี OLS ให้ค่า AMSE

Table 1. AMSE of multiple regression coefficient estimators, in case the distribution of error is normal distribution, number of independent variables equals 3 and restrictions are true.

Level of Corr.	n	Standard deviation											
		σ = 1				σ = 3				σ = 5			
		OLS	RLS	RRR	RL	OLS	RLS	RRR	RL	OLS	RLS	RRR	RL
low	30	0.04251	0.03795	0.03707	**0.03702***	0.36061	0.32156	0.30370	**0.30338***	1.01281	0.92879	**0.78402***	0.79834
	50	0.02119	0.01913	0.01898	**0.01897***	0.18870	0.17212	0.16301	**0.16290***	0.51973	0.47812	**0.43020***	0.43379
	100	0.00970	0.00916	0.00915	**0.00913***	0.08928	0.08247	0.08085	**0.08081***	0.24579	0.22908	**0.21904***	0.21944
middle	30	0.06239	0.05755	0.05307	**0.05287***	0.54152	0.48798	**0.40955***	0.41046	1.50979	1.36884	**1.06089***	1.09391
	50	0.03060	0.02862	0.02686	**0.02683***	0.27536	0.24859	0.22005	**0.21994***	0.76490	0.69052	**0.57320***	0.58404
	100	0.01495	0.01413	0.01359	**0.01358***	0.13457	0.12360	0.11651	**0.11639***	0.37382	0.34333	**0.31024***	0.31243
high	30	0.19453	0.18792	**0.11464***	0.11827	1.75379	1.69127	**0.94105***	0.97416	4.88831	4.69798	**2.55004***	2.68051
	50	0.08655	0.08366	**0.06181***	0.06323	0.75196	0.72596	**0.48648***	0.50058	2.08878	2.01655	**1.29194***	1.34060
	100	0.04818	0.04729	**0.03674***	0.03727	0.39959	0.38964	**0.29485***	0.30632	1.10441	1.08234	**0.78423***	0.81619

n = sample size; OLS = Ordinary Least Squares method; RLS = Restricted Least Squares method;
RRR = Restricted Ridge Regression method; RL = Restricted Liu method; corr. = correlations.
*** indicates the method providing the smallest AMSE.**

Table 2. AMSE of multiple regression coefficient estimators, in case distribution of error is normal distribution, number of independent variables equals 5 and restrictions are true.

Level of Corr.	n	σ = 1 OLS	σ = 1 RLS	σ = 1 RRR	σ = 1 RL	σ = 3 OLS	σ = 3 RLS	σ = 3 RRR	σ = 3 RL	σ = 5 OLS	σ = 5 RLS	σ = 5 RRR	σ = 5 RL
low	30	0.05558	0.04746	0.04609	**0.04599***	0.50020	0.42716	**0.36293***	0.36341	1.38943	1.18656	**0.91540***	0.94256
	50	0.02558	0.02272	0.02256	**0.02254***	0.23021	0.20450	0.19348	**0.19334***	0.63948	0.56805	**0.50669***	0.51203
	100	0.01257	0.01124	0.01121	**0.01120***	0.11313	0.10113	0.09884	**0.09879***	0.31425	0.28092	**0.26683***	0.26755
middle	30	0.07710	0.07116	0.06486	**0.06455***	0.69391	0.64041	**0.47606***	0.47875	1.92752	1.77891	**1.20981***	1.26051
	50	0.03767	0.03386	0.03147	**0.03143***	0.31200	0.29034	**0.25625***	0.25640	0.86667	0.80650	**0.66107***	0.67557
	100	0.01800	0.01693	0.01615	**0.01614***	0.15656	0.14693	0.13817	**0.13805***	0.43489	0.40813	**0.36655***	0.36953
high	30	0.22989	0.22249	**0.12881***	0.13334	2.06905	2.00241	**0.99091***	1.03224	5.74736	5.56224	**2.65626***	2.81050
	50	0.11353	0.11025	**0.07564***	0.07730	0.93180	0.90223	**0.57046***	0.58990	2.58832	2.50619	**1.50126***	1.56295
	100	0.05932	0.05784	**0.04414***	0.04470	0.47990	0.46652	**0.34610***	0.35341	1.33305	1.29588	**0.91312***	0.94034

Standard deviation

n = sample size; OLS = Ordinary Least Squares method; RLS = Restricted Least Squares method;
RRR = Restricted Ridge Regression method; RL = Restricted Liu method; corr. = correlations.
*** indicates the method providing the smallest AMSE.**

Table 3. AMSE of multiple regression coefficient estimators, in case distribution of error is normal distribution, number of independent variables equals 3 and error of restrictions equals 5%.

Level of Corr.	n	σ = 1 OLS	σ = 1 RLS	σ = 1 RRR	σ = 1 RL	σ = 3 OLS	σ = 3 RLS	σ = 3 RRR	σ = 3 RL	σ = 5 OLS	σ = 5 RLS	σ = 5 RRR	σ = 5 RL
low	30	0.04251	0.03997	0.03972	**0.03972***	0.36061	0.32366	0.30974	**0.30909***	1.01281	0.93096	**0.79268***	0.80584
	50	0.02119	0.02114	0.02112	**0.02112***	0.18870	0.17422	0.16664	**0.16651***	0.51973	0.48031	**0.43537***	0.43853
	100	**0.00970***	0.00971	0.00973	0.00972	0.08928	0.08443	0.08347	**0.08344***	0.24579	0.23108	**0.22260***	0.22290
middle	30	0.06239	0.05963	0.05609	**0.05596***	0.54152	0.52018	**0.41544***	0.41810	1.50979	1.37117	**1.06851***	1.10034
	50	0.03060	0.02971	0.02925	**0.02924***	0.27536	0.25078	**0.22365***	0.22392	0.76490	0.69281	**0.57819***	0.58848
	100	**0.01495***	0.01569	0.01568	0.01558	0.13457	0.12565	**0.11938***	0.11948	0.37382	0.34547	**0.31414***	0.31616
high	30	0.19453	0.19003	**0.11838***	0.12188	1.75379	1.69287	**0.94529***	0.97608	4.88831	4.69907	**2.55532***	2.68465
	50	0.08655	0.08636	**0.06504***	0.06636	0.75196	0.72875	**0.49095***	0.50462	2.08878	2.01945	**1.29719***	1.34480
	100	0.04818	0.04923	**0.03924***	0.03968	0.39959	0.39119	**0.29798***	0.30919	1.10441	1.08349	**0.78747***	0.81883

Standard deviation

n = sample size; OLS = Ordinary Least Squares method; RLS = Restricted Least Squares method;
RRR = Restricted Ridge Regression method; RL = Restricted Liu method; corr. = correlations.
*** indicates the method providing the smallest AMSE.**

น้อยที่สุดเมื่อขนาดตัวอย่างเท่ากับ 100 และวิธี RRR ให้ค่า AMSE น้อยที่สุดเมื่อระดับความสัมพันธ์สูง ภายใต้ทุกขนาดตัวอย่าง ในกรณีที่ส่วนเบี่ยงเบนมาตรฐานเพิ่มขึ้น วิธี RRR จะให้ค่า AMSE น้อยที่สุดในทุกขนาดตัวอย่างและทุกระดับความสัมพันธ์ ยกเว้นเมื่อส่วนเบี่ยงเบนมาตรฐานเท่ากับ 3 วิธี RL ให้ค่า AMSE น้อยที่สุดเมื่อระดับความสัมพันธ์ต่ำภายใต้ทุกขนาดตัวอย่าง และในกรณีข้อจำกัดมีความคลาดเคลื่อนเพิ่มขึ้น ส่วนเบี่ยงเบนมาตรฐานเท่ากับ 1 วิธี OLS จะให้ค่า AMSE น้อยที่สุดในทุกขนาดตัวอย่างและทุกระดับความสัมพันธ์ ยกเว้นระดับความสัมพันธ์สูง วิธี RRR ให้ค่า AMSE น้อยที่สุดเมื่อขนาดตัวอย่างเท่ากับ 30 นอกจากนี้เมื่อส่วนเบี่ยงเบนมาตรฐานเพิ่มขึ้น วิธี RRR จะให้ค่า AMSE น้อยที่สุดทุกขนาดตัวอย่างและทุกระดับความสัมพันธ์ และสังเกตได้ว่ากรณีที่วิธี OLS ให้ค่า AMSE น้อยที่สุด วิธี RLS จะให้ค่า AMSE รองลงมาเป็นส่วนใหญ่เมื่อระดับความสัมพันธ์ต่ำและปานกลาง ขนาดตัวอย่างมาก

จาก Table 6-8 เมื่อจำนวนตัวแปรอิสระเท่ากับ 5 จะพบว่า ในกรณีข้อจำกัดมีความคลาดเคลื่อน 5% ส่วนเบี่ยงเบนมาตรฐานเท่ากับ 1 วิธี RL จะให้ค่า AMSE

Table 4. AMSE of multiple regression coefficient estimators, in case distribution of error is normal distribution, number of independent variables equals 3 and error of restrictions equals 10%.

Level of Corr.	n	Standard deviation											
		σ = 1				σ = 3				σ = 5			
		OLS	RLS	RRR	RL	OLS	RLS	RRR	RL	OLS	RLS	RRR	RL
low	30	**0.04251***	0.04593	0.04635	0.04640	0.36061	0.32971	**0.31903***	0.31932	1.01281	0.93709	**0.80596***	0.81782
	50	**0.02119***	0.02710	0.02741	0.02743	0.18870	0.17927	**0.17407***	0.17422	0.51973	0.48644	**0.44468***	0.44736
	100	**0.00970***	0.01682	0.01696	0.01697	0.08928	0.08720	**0.08690***	0.08692	0.24579	0.23689	**0.23008***	0.23026
middle	30	**0.06239***	0.06572	0.06317	0.06323	0.54152	0.52640	**0.42636***	0.43097	1.50979	1.37752	**1.08111***	1.11159
	50	**0.03060***	0.03588	0.03569	0.03569	0.27536	0.25705	**0.23160***	0.23182	0.76490	0.69917	**0.58776***	0.59738
	100	**0.01495***	0.02148	0.02161	0.02161	0.13457	0.13153	**0.12633***	0.12641	0.37382	0.35145	**0.32222***	0.32404
high	30	0.19453	0.19685	**0.12682***	0.13011	1.75379	1.69919	**0.95471***	0.98504	4.88831	4.70488	**2.56581***	2.69402
	50	0.08655	0.09133	**0.07347***	0.07466	0.75196	0.73683	**0.50158***	0.51470	2.08878	2.02762	**1.30909***	1.35520
	100	**0.04818***	0.05142	0.04825	0.04849	0.39959	0.39699	**0.30633***	0.31118	1.10441	1.08891	**0.79632***	0.82671

n = sample size; OLS = Ordinary Least Squares method; RLS = Restricted Least Squares method;
RRR = Restricted Ridge Regression method; RL = Restricted Liu method; corr. = correlations.
*** indicates the method providing the smallest AMSE.**

Table 5. AMSE of multiple regression coefficient estimators, in case distribution of error is normal distribution, number of independent variables equals 3 and error of restrictions equals 15%.

Level of Corr.	n	Standard deviation											
		σ = 1				σ = 3				σ = 5			
		OLS	RLS	RRR	RL	OLS	RLS	RRR	RL	OLS	RLS	RRR	RL
low	30	**0.04251***	0.05585	0.05700	0.05711	0.36061	0.35971	**0.33317***	0.33363	1.01281	0.94717	**0.82380***	0.83426
	50	**0.02119***	0.03701	0.03753	0.03758	0.18870	0.19026	**0.18556***	0.18573	0.51973	0.49651	**0.45807***	0.46023
	100	**0.00970***	0.02637	0.02659	0.02661	**0.08928***	0.09979	0.10017	0.10018	0.24579	0.24651	**0.24143***	0.24148
middle	30	**0.06239***	0.07583	0.07452	0.07455	0.54152	0.53665	**0.44120***	0.44611	1.50979	1.38789	**1.09872***	1.12765
	50	**0.03060***	0.04614	0.04617	0.04620	0.27536	0.26740	**0.24374***	0.24402	0.76490	0.70963	**0.60187***	0.61071
	100	**0.01495***	0.03110	0.03138	0.03140	**0.13457***	0.14124	0.13722	0.13728	0.37382	0.36125	**0.33444***	0.33601
high	30	0.19453	0.20839	**0.14020***	0.14313	1.75379	1.71022	**0.96934***	0.99733	4.88831	4.71540	**2.58165***	2.70862
	50	**0.08655***	0.10460	0.08694	0.08799	0.75196	0.75020	**0.51830***	0.53074	2.08878	2.04109	**1.32752***	1.37176
	100	**0.04818***	0.06188	0.05738	0.05759	0.39959	0.40705	**0.31981***	0.32423	1.10441	1.09858	**0.81073***	0.83581

n = sample size; OLS = Ordinary Least Squares method; RLS = Restricted Least Squares method;
RRR = Restricted Ridge Regression method; RL = Restricted Liu method; corr. = correlations.
*** indicates the method providing the smallest AMSE.**

น้อยที่สุด เมื่อระดับความสัมพันธ์ต่ำและปานกลาง ขนาดตัวอย่างเท่ากับ 30 และ 50 แต่วิธี OLS ให้ค่า AMSE น้อยที่สุดเมื่อขนาดตัวอย่างเท่ากับ 100 และวิธี RRR ให้ค่า AMSE น้อยที่สุดเมื่อระดับความสัมพันธ์สูง ทุกขนาดตัวอย่าง กรณีที่ส่วนเบี่ยงเบนมาตรฐานเพิ่มขึ้นวิธี RRR จะให้ค่า AMSE น้อยที่สุดทุกขนาดตัวอย่างและทุกระดับความสัมพันธ์ ยกเว้นเมื่อส่วนเบี่ยงเบนมาตรฐานเท่ากับ 3 วิธี RL ให้ค่า AMSE น้อยที่สุด เมื่อระดับความสัมพันธ์ต่ำทุกขนาดตัวอย่าง กรณีข้อจำกัดมีความคลาดเคลื่อน 10% ส่วนเบี่ยงเบนมาตรฐานเท่ากับ 1 ส่วนใหญ่วิธี OLS ให้ค่า AMSE น้อยที่สุดเมื่อระดับความสัมพันธ์ต่ำและปานกลางภายใต้ขนาดตัวอย่างเท่ากับ 50 และ 100 และวิธี RRR จะให้ค่า AMSE น้อยที่สุดเมื่อขนาดตัวอย่างน้อย ระดับความสัมพันธ์สูง กรณีที่ส่วนเบี่ยงเบนมาตรฐานเพิ่มขึ้น วิธี RRR จะให้ค่า AMSE น้อยที่สุดทุกขนาดตัวอย่างและทุกระดับความสัมพันธ์ และในกรณีข้อจำกัดมีความคลาดเคลื่อน 15% ส่วนเบี่ยงเบนมาตรฐานเท่ากับ 1 วิธี OLS จะให้ค่า AMSE น้อยที่สุดทุกขนาดตัวอย่างและทุกระดับความสัมพันธ์ ยกเว้นระดับความสัมพันธ์สูง วิธี RRR ให้ค่า AMSE น้อยที่สุดเมื่อขนาดตัวอย่าง 30 และ 50 กรณี

Table 6. AMSE of multiple regression coefficient estimators, in case distribution of error is normal distribution, number of independent variables equals 5 and error of restrictions equals 5%.

Level of Corr.	n	Standard deviation											
		σ = 1				σ = 3				σ = 5			
		OLS	RLS	RRR	RL	OLS	RLS	RRR	RL	OLS	RLS	RRR	RL
low	30	0.05558	0.04957	0.04809	**0.04800***	0.50020	0.42926	0.36645	**0.36580***	1.38943	1.18865	**0.91962***	0.94567
	50	0.02558	0.02419	0.02419	**0.02419***	0.23021	0.20600	0.19611	**0.19594***	0.63948	0.56959	**0.51036***	0.51533
	100	**0.01257***	0.01261	0.01266	0.01265	0.11313	0.10251	0.10075	**0.10071***	0.31425	0.28230	**0.26932***	0.26994
middle	30	0.07710	0.07376	0.06705	**0.06673***	0.69391	0.64307	**0.47884***	0.48121	1.92752	1.78163	**1.21336***	1.26311
	50	0.03767	0.03599	0.03349	**0.03347***	0.31200	0.29202	**0.25922***	0.25984	0.86667	0.80813	**0.66462***	0.67880
	100	**0.01800***	0.01895	0.01891	0.01878	0.15656	0.14856	**0.14037***	0.14048	0.43489	0.40976	**0.36927***	0.37214
high	30	0.22989	0.22759	**0.13167***	0.13629	2.06905	2.00851	**0.99504***	1.03873	5.74736	5.56934	**2.66171***	2.81561
	50	0.11353	0.11313	**0.07900***	0.08057	0.93180	0.90487	**0.57395***	0.59054	2.58832	2.50859	**1.50466***	1.56650
	100	0.05932	0.05979	**0.04735***	0.04787	0.47990	0.46969	**0.34986***	0.35720	1.33305	1.29928	**0.91717***	0.94449

n = sample size; OLS = Ordinary Least Squares method; RLS = Restricted Least Squares method; RRR = Restricted Ridge Regression method; RL = Restricted Liu method; corr. = correlations.
*** indicates the method providing the smallest AMSE.**

Table 7. AMSE of multiple regression coefficient estimators, in case distribution of error is normal distribution, number of independent variables equals 5 and error of restrictions equals 10%.

Level of Corr.	n	Standard deviation											
		σ = 1				σ = 3				σ = 5			
		OLS	RLS	RRR	RL	OLS	RLS	RRR	RL	OLS	RLS	RRR	RL
low	30	0.05558	0.05590	**0.05405***	0.05414	0.50020	0.43558	**0.37248***	0.37349	1.38943	1.19496	**0.92779***	0.95263
	50	**0.02558***	0.02854	0.02870	0.02871	0.23021	0.21039	**0.20139***	0.20158	0.63948	0.57402	**0.51696***	0.52153
	100	**0.01257***	0.01673	0.01685	0.01686	0.11313	0.10663	**0.10537***	0.10541	0.31425	0.28643	**0.27459***	0.27512
middle	30	0.07710	0.08151	**0.07351***	0.07386	0.69391	0.65088	**0.48615***	0.48818	1.92752	1.78950	**1.22167***	1.27032
	50	**0.03767***	0.04120	0.03897	0.03907	0.31200	0.29719	**0.26567***	0.26613	0.86667	0.81326	**0.67183***	0.68561
	100	**0.01800***	0.02282	0.02291	0.02292	0.15656	0.15343	**0.14593***	0.14604	0.43489	0.41464	**0.37535***	0.37809
high	30	0.22989	0.24187	**0.13997***	0.14463	2.06905	2.02379	**1.00507***	1.04604	5.74736	5.58563	**2.67317***	2.82669
	50	0.11353	0.11803	**0.08759***	0.08909	0.93180	0.91353	**0.58313***	0.59880	2.58832	2.51701	**1.51401***	1.57575
	100	0.05932	0.06342	**0.05581***	0.05632	0.47990	0.47855	**0.35933***	0.36666	1.33305	1.30836	**0.92724***	0.95441

n = sample size; OLS = Ordinary Least Squares method; RLS = Restricted Least Squares method; RRR = Restricted Ridge Regression method; RL = Restricted Liu method; corr. = correlations.
*** indicates the method providing the smallest AMSE.**

ที่ส่วนเบี่ยงเบนมาตรฐานเพิ่มขึ้น วิธี RRR จะให้ค่า AMSE น้อยที่สุดทุกขนาดตัวอย่างและทุกระดับความสัมพันธ์ และสังเกตได้ว่ากรณีที่วิธี OLS ให้ค่า AMSE น้อยที่สุด วิธี RLS จะให้ค่า AMSE รองลงมาเป็นส่วนใหญ่เมื่อระดับความสัมพันธ์ต่ำและปานกลาง และขนาดตัวอย่างมีค่ามาก

จาก Table 1-8 จะเห็นได้ว่าค่า AMSE มีแนวโน้มลดลงเมื่อขนาดตัวอย่างเพิ่มขึ้นเพราะเมื่อขนาดตัวอย่างเพิ่มขึ้นทำให้ความแปรปรวนลดลงจึงส่งผลทำให้ค่า AMSE ลดลง และมีแนวโน้มเพิ่มขึ้นเมื่อระดับความสัมพันธ์เพิ่มขึ้นเพราะระดับความสัมพันธ์เพิ่มขึ้นทำให้ค่าลักษณะเฉพาะของเมทริกซ์ $X'X$ มีค่าลดลง จึงส่งผลทำให้ค่า AMSE มีแนวโน้มเพิ่มขึ้น นอกจากนี้เมื่อส่วนเบี่ยงเบนมาตรฐานและจำนวนตัวแปรอิสระเพิ่มขึ้นก็มีผลทำให้ค่า AMSE มีแนวโน้มเพิ่มขึ้น เพราะเมื่อส่วนเบี่ยงเบนมาตรฐานเพิ่มขึ้นทำให้เกิดความคลาดเคลื่อนมากขึ้นและเมื่อจำนวนตัวแปรอิสระเพิ่มขึ้นทำให้เกิดความสัมพันธ์ระหว่างกลุ่มตัวแปรอิสระเพิ่มขึ้น จึงส่งผลทำให้ค่า AMSE มีแนวโน้มเพิ่มขึ้น ในขณะเดียวกันเมื่อพิจารณาที่ข้อจำกัดจะเห็นได้ว่าที่ข้อจำกัดมีความคลาดเคลื่อนเกิดขึ้นจะทำให้ค่า AMSE เพิ่มขึ้น ซึ่งมีผลกระทบต่อวิธี RLS RRR และ RL เพราะทั้ง

Table 8. AMSE of multiple regression coefficient estimators, in case distribution of error is normal distribution, number of independent variables equals 5 and error of restrictions equals 15%.

Level of Corr.	n	Standard deviation											
		σ = 1				σ = 3				σ = 5			
		OLS	RLS	RRR	RL	OLS	RLS	RRR	RL	OLS	RLS	RRR	RL
low	30	**0.05558***	0.06646	0.06411	0.06421	0.50020	0.44613	**0.38294***	0.38372	1.38943	1.20549	**0.93989***	0.96346
	50	**0.02558***	0.03578	0.03609	0.03612	0.23021	0.21767	**0.20967***	0.20989	0.63948	0.58134	**0.52649***	0.53062
	100	**0.01257***	0.02360	0.02378	0.02381	0.11313	0.11350	**0.11277***	0.11281	0.31425	0.28643	**0.28265***	0.28306
middle	30	**0.07710***	0.09441	0.08484	0.08524	0.69391	0.66383	**0.49797***	0.49965	1.92752	1.80251	**1.23471***	1.28214
	50	**0.03767***	0.04792	0.04791	0.04793	0.31200	0.30586	**0.27556***	0.27586	0.86667	0.82189	**0.68268***	0.69598
	100	**0.01800***	0.03093	0.03115	0.03116	0.15656	0.16155	**0.15473***	0.15483	0.43489	0.42277	**0.38478***	0.38737
high	30	0.22989	0.26534	**0.15357***	0.15820	2.06905	2.04827	**1.02103***	1.06179	5.74736	5.61111	**2.69074***	2.84375
	50	0.11353	0.13695	**0.10138***	0.10280	0.93180	0.92821	**0.59798***	0.61366	2.58832	2.53144	**1.52926***	1.59069
	100	**0.05932***	0.07772	0.06948	0.06997	0.47990	0.49308	**0.37446***	0.38173	1.33305	1.32312	**0.94334***	0.97009

n = sample size; OLS = Ordinary Least Squares method; RLS = Restricted Least Squares method;
RRR = Restricted Ridge Regression method; RL = Restricted Liu method; corr. = correlations.
*** indicates the method providing the smallest AMSE.**

3 วิธีเป็นวิธีที่ใช้ข้อจำกัดเกี่ยวกับพารามิเตอร์ ($\underset{\sim}{\beta}$) ในการสร้างตัวประมาณสัมประสิทธิ์การถดถอยพหุคูณซึ่งส่งผลทำให้ค่า AMSE เพิ่มขึ้น ขณะที่วิธี OLS เป็นวิธีที่ไม่มีข้อจำกัดเกี่ยวกับพารามิเตอร์เข้ามาเกี่ยวข้องจึงทำให้ค่า AMSE คงที่ และเมื่อข้อจำกัดมีความคลาดเคลื่อนมากขึ้นทำให้ค่า AMSE ของวิธี RLS RRR และ RL มีค่าเพิ่มขึ้น

4. สรุปผลการวิจัย

4.1 กรณีข้อจำกัดเป็นจริง

จากการเปรียบเทียบกับวิธีต่างๆ จะเห็นได้ว่าในทุกกรณี วิธี RRR และ RL จะให้ค่า AMSE น้อยกว่าวิธี OLS และ RLS ตามลำดับ โดยในสถานการณ์ที่มีผลทำให้วิธี RRR มีค่า AMSE น้อยที่สุดคือ ระดับความสัมพันธ์ที่สูงระหว่างตัวแปรอิสระ ส่วนเบี่ยงเบนมาตรฐานมาก จำนวนตัวแปรอิสระมาก และขนาดตัวอย่างน้อย ส่วนสถานการณ์ที่มีผลทำให้วิธี RL มีค่า AMSE น้อยที่สุดคือระดับความสัมพันธ์ที่ต่ำระหว่างตัวแปรอิสระ ส่วนเบี่ยงเบนมาตรฐานน้อย จำนวนตัวแปรอิสระน้อย และขนาดตัวอย่างมาก

วิธีการต่างๆ ที่ใช้ในการวิจัยจะให้ค่า AMSE ลดลงเมื่อขนาดตัวอย่างเพิ่มขึ้น และให้ค่า AMSE เพิ่มขึ้นเมื่อระดับความสัมพันธ์ ส่วนเบี่ยงเบนมาตรฐาน และจำนวนตัวแปรอิสระเพิ่มขึ้น ตามลำดับ

4.2 กรณีข้อจำกัดไม่เป็นจริง

จากการเปรียบเทียบกับวิธีต่างๆ จะเห็นได้ว่าสถานการณ์ที่มีผลทำให้วิธี RRR มีค่า AMSE น้อยที่สุดคือข้อจำกัดมีความคลาดเคลื่อนน้อย ระดับความสัมพันธ์ระหว่างตัวแปรอิสระสูง ส่วนเบี่ยงเบนมาตรฐานมาก จำนวนตัวแปรอิสระมาก และขนาดตัวอย่างน้อย และสถานการณ์ที่มีผลทำให้วิธี RL มีค่า AMSE น้อยที่สุดคือข้อจำกัดมีความคลาดเคลื่อนน้อย ระดับความสัมพันธ์ระหว่างตัวแปรอิสระต่ำ ส่วนเบี่ยงเบนมาตรฐานน้อย จำนวนตัวแปรอิสระน้อย และขนาดตัวอย่างน้อย ส่วนสถานการณ์ที่มีผลทำให้วิธี OLS มีค่า AMSE น้อยที่สุดคือข้อจำกัดมีความคลาดเคลื่อนมาก ระดับความสัมพันธ์ระหว่างตัวแปรอิสระต่ำ ส่วนเบี่ยงเบนมาตรฐานน้อย จำนวนตัวแปรอิสระน้อย และขนาดตัวอย่างมาก

วิธีการต่างๆ ที่ใช้ในการวิจัยจะให้ค่า AMSE ลดลงเมื่อขนาดตัวอย่างเพิ่มขึ้น และให้ค่า AMSE เพิ่มขึ้นเมื่อความคลาดเคลื่อนของข้อจำกัด ระดับความสัมพันธ์ ส่วนเบี่ยงเบนมาตรฐาน และจำนวนตัวแปรอิสระเพิ่มขึ้นตามลำดับ ยกเว้นกรณีวิธี OLS ซึ่งความคลาดเคลื่อนของข้อจำกัดไม่มีผลต่อค่า AMSE เพราะวิธี OLS เป็นวิธีที่ไม่มีข้อจำกัดเกี่ยวกับพารามิเตอร์เข้ามาเกี่ยวข้องจึงทำให้ค่า AMSE คงที่

5. ข้อเสนอแนะ

ผลการวิจัยในครั้งนี้มีข้อเสนอแนะ 2 ด้าน คือ

5.1 ด้านนำไปใช้ประโยชน์ (เมื่อตัวแปรอิสระมีพหุสัมพันธ์)

การเลือกใช้ตัวประมาณ

เนื่องจากอาจทำให้เกิดความยุ่งยากกรณีที่ไม่ทราบการแจกแจงของความคลาดเคลื่อน และความคลาดเคลื่อนของข้อจำกัด ดังนั้นเพื่อความสะดวกในการนำไปใช้งานผู้วิจัยได้สรุปสถานการณ์ต่างๆ ในกรณีที่สามารถประมาณค่าระดับความสัมพันธ์ของตัวแปรอิสระได้ โดยทราบจำนวนตัวแปรอิสระ และขนาดตัวอย่าง ดังต่อไปนี้

1. กรณีข้อจำกัดเป็นจริง (จำนวนตัวแปรอิสระเท่ากับ 3 และ 5)

- ระดับความสัมพันธ์ต่ำ ควรเลือกใช้วิธี RL เมื่อขนาดตัวอย่าง 30, 50 และ 100
- ระดับความสัมพันธ์ปานกลาง ควรเลือกใช้วิธี RL เมื่อขนาดตัวอย่าง 30, 50 และ 100
- ระดับความสัมพันธ์สูง ควรเลือกใช้วิธี RRR เมื่อขนาดตัวอย่าง 30, 50 และ 100

2. กรณีข้อจำกัดไม่เป็นจริง (จำนวนตัวแปรอิสระเท่ากับ 3 และ 5)

- ระดับความสัมพันธ์ต่ำ ควรเลือกใช้วิธี RRR เมื่อขนาดตัวอย่าง 30 และวิธี OLS ที่ขนาดตัวอย่าง 50 และ 100
- ระดับความสัมพันธ์ปานกลาง ควรเลือกใช้วิธี RRR ที่ขนาดตัวอย่าง 30 และ 50 และวิธี OLS เมื่อขนาดตัวอย่าง 100
- ระดับความสัมพันธ์สูง ควรเลือกใช้วิธี RRR เมื่อขนาดตัวอย่าง 30, 50 และ 100

จากการเลือกใช้ตัวประมาณข้างต้นมีรายละเอียดเพิ่มเติมในการเลือกใช้ได้ดังนี้

กรณีข้อจำกัดเป็นจริง

เนื่องจากวิธี RRR ให้ค่า AMSE น้อยกว่าวิธี RL ไม่มากนัก โดยทั่วไปเราสามารถใช้วิธี RL ในการแก้ปัญหาการเกิดพหุสัมพันธ์ได้ทุกกรณีเพราะวิธี RL ไม่ต้องยุ่งยากในการคำนวณค่า k เหมือนวิธี RRR

กรณีข้อจำกัดไม่เป็นจริง

จากผลสรุปในข้างต้น กรณีที่ทราบความคลาดเคลื่อนของข้อจำกัดและส่วนเบี่ยงเบนมาตรฐานของการแจกแจง พบว่าวิธี OLS จะให้ค่า AMSE น้อยที่สุดเมื่อความคลาดเคลื่อนของข้อจำกัดมาก ส่วนเบี่ยงเบนมาตรฐานน้อย ในขณะที่วิธี RRR ก็จะให้ค่า AMSE น้อยที่สุดเมื่อส่วนเบี่ยงเบนมาตรฐานมาก เราจะเห็นได้ว่าความคลาดเคลื่อนของข้อจำกัดและส่วนเบี่ยงเบนมาตรฐานต่างมีผลต่อค่า AMSE ของวิธี RRR และ OLS และเมื่อพิจารณาในทางปฏิบัติโดยส่วนใหญ่ความคลาดเคลื่อนของข้อจำกัดมากและส่วนเบี่ยงเบนมาตรฐานก็มากด้วย จึงทำให้ผู้วิจัยเลือกใช้เกณฑ์ที่ความคลาดเคลื่อนของข้อจำกัดและส่วนเบี่ยงเบนมาตรฐานมีค่ามากในการเลือกใช้วิธีการประมาณค่าสัมประสิทธิ์การถดถอยพหุคูณดังกล่าวข้างต้น (กรณีข้อจำกัดไม่เป็นจริง) เพื่อจะได้ไม่โน้มเอียงไปในวิธีการใดวิธีการหนึ่ง โดยกรณีที่วิธี RRR ให้ค่า AMSE น้อยกว่าวิธี OLS ไม่มากนักจะเลือกใช้วิธี OLS เพราะวิธี OLS มีการคำนวณที่ง่ายกว่าวิธี RRR มาก

5.2 ด้านการวิจัย

1) ควรศึกษาวิจัยการประมาณค่าพารามิเตอร์ k และ d ที่ง่ายและสะดวกรวดเร็วมาช่วยในการประมาณค่าสัมประสิทธิ์ของวิธีริดจ์รีเกรสชันที่ถูกจำกัดและวิธีลิวที่ถูกจำกัด

2) ควรทำการศึกษาภายใต้ข้อจำกัดเกี่ยวกับพารามิเตอร์ $\underset{\sim}{\beta}$ ที่มากกว่าข้อจำกัดเดียว

3) ควรศึกษาเพิ่มเติมเมื่อจำนวนตัวแปรอิสระมากกว่า 5

4) ควรศึกษาและเปรียบเทียบวิธีการทางสถิติวิธีอื่นที่ใช้ในการประมาณค่าสัมประสิทธิ์การถดถอยพหุคูณเมื่อเกิดพหุสัมพันธ์ระหว่างตัวแปรอิสระ

5) ควรมีการศึกษาที่ขนาดตัวอย่างเล็กกว่า 30 เช่น n = 10 หรือ 20 เป็นต้น ซึ่งน่าจะทำให้วิธี RRR ให้ค่า AMSE น้อยกว่าวิธีอื่นมากขึ้น

6) ถ้าตัวแปรอิสระแต่ละคู่มีระดับความสัมพันธ์หลายระดับปนกัน การที่จะสรุปว่าระดับความสัมพันธ์อยู่ระดับใดนั้นควรมองภาพรวมว่าส่วนใหญ่ตัวแปรอิสระมีความสัมพันธ์อยู่ในระดับใด

7) ข้อจำกัดไม่เป็นจริงคือกรณีที่หาข้อจำกัดไม่ได้ซึ่งในทางปฏิบัติจะเกิดกรณีนี้มากที่สุด

กิตติกรรมประกาศ

การศึกษานี้เป็นส่วนหนึ่งของงานวิจัยที่ได้รับทุนสนับสนุนงานวิจัยระดับบัณฑิตวิทยาลัยจากทบวงมหาวิทยาลัย ประจำปี 2546 ขอขอบคุณอาจารย์ที่ปรึกษาและอาจารย์ทุกท่านที่ได้ให้ความรู้ คำแนะนำและคำปรึกษาตลอดจนการแก้ไขข้อบกพร่องต่างๆ เป็นอย่างดีมาโดยตลอด และบัณฑิตวิทยาลัยที่ให้การสนับสนุนในการทำวิทยานิพนธ์ของนักศึกษาปริญญาโท จนทำให้งานวิจัยสำเร็จลุล่วงไปด้วยดี

เอกสารอ้างอิง

Hoerl, E. and Kennard, Robert W. 1970. Ridge Regression: Biased Estimation for Nonorthogonal Problem. Technometrics 42: 80-86.

Kaciranlar, S., Sakallioglu, S. and Akdeniz, F.A. 1999. New Biased Estimator in Linear Regression and Detialed Analysis of the Widely-Analysed Dataset on Portland Cement: Sankhya, Series B, 61: 443-459.

Kaciranlar, S. and Akdeniz, F. 2001. More on the Biased Estimator in Linear Regression: Sankhya, Series B, 63: 321-325.

Rao, C.R. 1973. Linear Statistical Inference and Its Applications, Wiley, New York.

Sarkar, N. 1992. A New Estimator Combining the Ridge Regression and the Restricted Least Squares Methods of Estimation: Communications in Statistics and Theory Methods, Series A, 21: 1987-2000.