



การเปรียบเทียบการคัดเลือกตัวแบบการถดถอยแบบริดจ์ โดยใช้เกณฑ์ซีพี และสถิติทดสอบซีพี

A Comparison of Selecting Ridge Regression Model by Cp Criteria and C_p Test Statistics

ปาวริศ มีบางไทร(Pavaris Meebangsai)^{1*} ดร.วิชิต หล่อจิระชุนห์กุล(Dr.Vichit Lorchirachoonkul)**

ดร.จิราวัลย์ จิตรถเวช (Dr.Jirawan Jitthavech) ***

(Received: May 21, 2018; Revised: July 9, 2018; Accepted: July 15, 2018)

บทคัดย่อ

การศึกษามีวัตถุประสงค์เพื่อเปรียบเทียบการคัดเลือกตัวแปรอิสระในการวิเคราะห์การถดถอยแบบริดจ์เชิงเส้นพหุคูณ ที่ใช้วิธีการประมาณค่าพารามิเตอร์ริดจ์ 5 วิธี คือ วิธีของโฮเอิร์ล เคนนาร์ด วิธีของ โฮเอิร์ล เคนนาร์ดและบาลด์วิน วิธีของลอว์เลสและแวง วิธีของคาลาฟและชูเกอร์ และวิธีกำลังสองน้อยที่สุด โดยใช้วิธีการคัดเลือกแบบไปข้างหน้า และการกำจัดแบบถอยหลังโดยใช้เกณฑ์ซีพี และสถิติทดสอบซีพี ผลการจำลองข้อมูลสรุปได้ว่าเมื่อขนาดตัวอย่าง 15 30 และ 50 ร้อยละของจำนวนครั้งที่คัดเลือกตัวแบบได้ถูกต้องเมื่อใช้สถิติทดสอบซีพีมีค่าสูงกว่าการใช้เกณฑ์ซีพี ในการใช้วิธีการคัดเลือกแบบไปข้างหน้าและการกำจัดแบบถอยหลัง ซึ่งผลการศึกษาไม่พบตัวแบบ Underspecification และ Misspecification มีเพียงตัวแบบ Overspecification ซึ่งเป็นปัญหาที่มีความรุนแรงในการวิเคราะห์น้อยกว่าตัวแบบในสองกรณีแรก นอกจากนี้พบว่าทุกระดับความสัมพันธ์ของตัวแปรอิสระระหว่าง X_1 และ X_2 มีการคัดเลือกตัวแบบได้ถูกต้องไม่แตกต่างกัน และวิธีการประมาณค่าพารามิเตอร์ริดจ์ด้วยวิธีต่างๆ สามารถคัดเลือกได้ถูกต้องสูงกว่าวิธีกำลังสองน้อยที่สุด เมื่อตัวอย่างมีขนาดเล็ก และสามารถคัดเลือกตัวแบบได้ถูกต้องไม่แตกต่างกันเมื่อตัวอย่างมีขนาดใหญ่ขึ้น

ABSTRACT

The purpose of this study is to compare the independent variable selection in analyzing multiple linear Ridge Regression by Hoerl and Kennard, Hoerl Kennard and Baldwin, Lawless and Wang, Khalaf and Shukur and Ordinary Least Square methods (OLS). Model selection is conducted by forward selection and backward elimination methods using the C_p criteria and C_p test statistics. The simulation results can be concluded that in all sample sizes, the percentage of selecting the correct model based on C_p test statistics by the forward selection and backward elimination method was higher than basing on the C_p criteria. The underspecified model and the misspecified model are not found in this study, only found overspecified model which its effects are less serious than those of the underspecified models. In addition, it is found that, all correlations between X_1 and X_2 give similar correct percentage. Various methods in estimate ridge parameter have percentage of select the correct model higher than OLS when small sample size and no difference when the sample size is larger.

คำสำคัญ: วิธีการถดถอยแบบริดจ์ เกณฑ์ซีพี สถิติทดสอบซีพี

Keywords: Ridge regression, C_p criteria, C_p test statistics

¹Correspondent author: pavaris.m@gmail.com

* นักศึกษา หลักสูตรวิทยาศาสตรมหาบัณฑิต สาขาวิชาสถิติ คณะสถิติประยุกต์ สถาบันบัณฑิตพัฒนบริหารศาสตร์

** รองศาสตราจารย์ สาขาวิชาสถิติ คณะสถิติประยุกต์ สถาบันบัณฑิตพัฒนบริหารศาสตร์

*** ศาสตราจารย์ สาขาวิชาสถิติ คณะสถิติประยุกต์ สถาบันบัณฑิตพัฒนบริหารศาสตร์

บทนำ

การวิเคราะห์การถดถอยเชิงเส้นพหุคูณ เป็นการหารูปแบบความสัมพันธ์ระหว่างตัวแปรตามกับตัวแปรอิสระเพื่อนำไปพยากรณ์ตัวแปรตาม เมื่อจำนวนตัวแปรอิสระมีตั้งแต่สองตัวขึ้นไป ซึ่งตัวแบบที่เหมาะสมในการวิเคราะห์เชิงเส้นพหุคูณควรมีจำนวนตัวแปรอิสระน้อยที่สุด แต่ทำให้ได้สมการถดถอยที่สามารถพยากรณ์ได้ถูกต้องและแม่นยำเพียงพอ เพราะหากมีจำนวนตัวแปรอิสระมากเกินไป ความคลาดเคลื่อนของค่าพยากรณ์มีค่าสูงขึ้นและอาจเกิดปัญหาตัวแปรอิสระบางตัวมีพหุสัมพันธ์ (Multicollinearity) ต่อกันได้ ทำให้ค่าสัมประสิทธิ์การถดถอยที่ได้มีค่าไม่เสถียรและมีความคลาดเคลื่อนสูง ซึ่งมักใช้วิธีการแก้ปัญหาโดยการตัดตัวแปรที่มีความสัมพันธ์ระหว่างกันสูงทิ้งไป หรือเก็บรวบรวมข้อมูลเพิ่มขึ้น เพื่อลดระดับความสัมพันธ์ระหว่างตัวแปรอิสระลง แต่หากข้อมูลที่เก็บรวบรวมเพิ่มทำให้ตัวแปรอิสระยังคงมีพหุสัมพันธ์เชิงเส้นระหว่างกันอยู่ หรือเพิ่มมากขึ้นอาจทำให้ปัญหานี้รุนแรงมากขึ้นได้

การวิเคราะห์การถดถอยแบบริดจ์ (Ridge Regression) เป็นการวิเคราะห์เพื่อประมาณพารามิเตอร์ของตัวแบบการถดถอยในกรณีที่ตัวแปรอิสระมีพหุสัมพันธ์กันสูง ซึ่งวิธีนี้ไม่จำเป็นต้องตัดตัวแปรอิสระที่มีความสัมพันธ์กันสูงออกจากตัวแบบ หลักการของการถดถอยแบบริดจ์คือการทำให้ค่าประมาณสัมประสิทธิ์การถดถอยเชิงเส้นพหุคูณมีค่าเพิ่มขึ้นเล็กน้อย ซึ่งจะทำให้ได้ตัวประมาณที่เอนเอียงแต่ความแปรปรวนของตัวประมาณค่าต่ำลงและส่งผลให้ค่าความคลาดเคลื่อนกำลังสองเฉลี่ยของตัวประมาณค่าต่ำลง เนื่องจากค่าประมาณความคลาดเคลื่อนของสัมประสิทธิ์การถดถอยเป็นฟังก์ชันของ $(X'X)^{-1}$ ดังนั้นการลดค่าความคลาดเคลื่อนดังกล่าวให้ต่ำลงจึงต้องพยายามลดค่า $(X'X)^{-1}$ ให้ต่ำลงโดยการบวกค่าคงตัวที่มีค่าน้อย ๆ ให้กับเมทริกซ์ $(X'X)^{-1}$ ที่มีการแปลงรูปด้วยสเกลหน่วยความยาวของตัวแปรตาม และตัวแปรอิสระ การแปลงรูปนั้นต้องลบค่าสังเกตทุกตัวด้วยค่าเฉลี่ยและหารด้วยรากที่สองของค่าผลบวกกำลังสองของตัวแปรนั้นๆ

จากตัวแบบการถดถอยเชิงเส้นพหุคูณในรูปของเมทริกซ์

$$y = X\beta + \varepsilon \quad (1)$$

โดยที่ y เป็นเวกเตอร์ค่าสังเกตของตัวแปรตาม y ขนาด $n \times 1$
 X เป็นเมทริกซ์ของตัวแปรอิสระขนาด $n \times (p+1)$
 β เป็นเวกเตอร์ของพารามิเตอร์ของตัวแบบขนาด $(p+1) \times 1$
 ε เป็นเวกเตอร์ของความคลาดเคลื่อนสุ่ม $n \times 1$

ฐานคติของการถดถอยเชิงเส้นพหุ ความคลาดเคลื่อนสุ่ม ε มีการแจกแจงแบบปกติ มีค่าเฉลี่ยเท่ากับ 0 และความแปรปรวนเท่ากับ $\sigma^2 I$ นั่นคือ $\varepsilon \sim N_n(0, \sigma^2 I)$

เมื่อ I เป็นเมทริกซ์เอกลักษณ์ขนาด $n \times n$
 n เป็นขนาดตัวอย่าง
 p เป็นจำนวนตัวแปรอิสระในตัวแบบ

การแปลงรูปด้วยสเกลหน่วยความยาวทำดังนี้ $w_{ij} = \frac{x_{ij} - \bar{x}_j}{\sqrt{SS_j}}$; $i = 1, 2, \dots, n$; $j = 1, 2, \dots, p$ และ $y_i^0 = \frac{y_i - \bar{y}}{\sqrt{SST}}$; $i = 1, 2, \dots, n$ เมื่อ $SS_j = \sum_{i=1}^n (x_{ij} - \bar{x}_j)^2$ เป็นผลบวกกำลังสองของตัวแปรอิสระ x_j และ $SST = \sum_{i=1}^n (y_i - \bar{y})^2$ เป็นผลบวกกำลังสองของตัวแปรตาม y



ตัวแบบการถดถอยที่มีการแปลงรูปอยู่ในรูปแบบดังนี้

$$\mathbf{y}^0 = \mathbf{W}\boldsymbol{\delta} + \boldsymbol{\varepsilon} \quad (2)$$

ตัวแบบการถดถอยในสมการ (1) เมทริกซ์ $\mathbf{X}$ มีขนาดเท่ากับ $n \times (p + 1)$ สมการ (2) เมทริกซ์ $\mathbf{W}$ มีขนาดเท่ากับ $n \times p$ และ $\mathbf{y}^0$ มีขนาดเท่ากับ $n \times 1$ เมื่อ p เป็นจำนวนตัวแปรอิสระในตัวแบบ

จากตัวแบบการถดถอยภายใต้ข้อสมมติพื้นฐานเกี่ยวกับ $\boldsymbol{\varepsilon}_i$ ในการวิเคราะห์การถดถอย ตัวประมาณพารามิเตอร์ของตัวแบบด้วยวิธีกำลังน้อยที่สุด ภายใต้ตัวแบบการถดถอยหลังการแปลงรูปตาม (2) จะได้สัมประสิทธิ์การถดถอยมาตรฐานดังนี้

$$\hat{\boldsymbol{\delta}} = (\mathbf{W}'\mathbf{W})^{-1} \mathbf{W}'\mathbf{y}^0 \quad (3)$$

และได้เมทริกซ์ความแปรปรวนของ $\hat{\boldsymbol{\delta}}$ เท่ากับ $\text{Var}(\hat{\boldsymbol{\delta}}) = \sigma^2 (\mathbf{W}'\mathbf{W})^{-1}$

ในปี 1970 Hoerl and Kennard ได้นำเสนอวิธีการวิเคราะห์ถดถอยริดจ์ ในกรณีที่ตัวแปรอิสระมีพหุสัมพันธ์เชิงเส้นต่อกัน โดยทำการบวกพารามิเตอร์ริดจ์ (Ridge Parameter) ซึ่งเป็นค่าคงตัวน้อยๆ ที่มากกว่าศูนย์ กับสมาชิกทุกตัวบนเส้นทแยงมุมของเมทริกซ์ $\mathbf{W}'\mathbf{W}$ จะได้ตัวประมาณค่าสัมประสิทธิ์การถดถอยแบบริดจ์เท่ากับ

$$\hat{\boldsymbol{\delta}} = (\mathbf{W}'\mathbf{W} + k\mathbf{I})^{-1} \mathbf{W}'\mathbf{y}^0 \quad (4)$$

ตัวประมาณ $\hat{\boldsymbol{\delta}}$ ที่ได้มีสมบัติเป็นตัวประมาณค่าที่มีความเอนเอียง แต่มีความแปรปรวนของตัวประมาณค่าต่ำกว่าตัวประมาณค่าด้วยวิธีกำลังสองน้อยที่สุด แต่ในการประมาณค่าสัมประสิทธิ์การถดถอยแบบริดจ์จะต้องมีการประมาณค่าพารามิเตอร์ริดจ์ (k) ซึ่งการศึกษาครั้งนี้ได้ใช้วิธีการประมาณค่าพารามิเตอร์ริดจ์ 5 วิธี ได้แก่

1. วิธีกำลังสองน้อยสุด ($k = 0$)
2. วิธี Hoerl and Kennard [1]

$$k_{HK} = \frac{\hat{\sigma}^2}{\hat{\alpha}_{\max}^2} \quad (5)$$

เมื่อ $\hat{\sigma}^2 = \frac{\mathbf{y}^{0'}\mathbf{y}^0 - \hat{\boldsymbol{\alpha}}'\mathbf{T}'\mathbf{W}'\mathbf{y}^0}{n - p}$, $\hat{\boldsymbol{\alpha}} = \mathbf{T}'\hat{\boldsymbol{\delta}}$, โดย $\mathbf{T}$ เป็นเมทริกซ์ของเวกเตอร์เฉพาะ (Eigenvector) ของเมทริกซ์ $\mathbf{W}'\mathbf{W}$, $\hat{\alpha}_i$ เป็นสมาชิกตัวที่ i ของเวกเตอร์ $\hat{\boldsymbol{\alpha}}$, $\hat{\alpha}_{\max} = \max(\hat{\alpha}_1, \hat{\alpha}_2, \dots, \hat{\alpha}_p)$ และ p เป็นจำนวน

พารามิเตอร์ที่ต้องประมาณค่าในตัวแบบที่อยู่ในสมการ (2)

3. วิธี Hoerl Kennard and Baldwin [2]

$$k_{HKB} = \frac{p\hat{\sigma}^2}{\hat{\boldsymbol{\alpha}}'\hat{\boldsymbol{\alpha}}} \quad (6)$$

4. วิธี Lawless and Wang [3]

$$k_{LW} = \frac{p\hat{\sigma}^2}{\sum_{i=1}^p \lambda_i \hat{\alpha}_i^2} \quad (7)$$

เมื่อ λ_i เป็นค่าเฉพาะที่ i ของเมทริกซ์ $\mathbf{W}'\mathbf{W}$

5. วิจิ Khalaf and Shukar [4]

$$k_{KS} = \frac{\lambda_{\max} \hat{\sigma}^2}{\lambda_{\max} \hat{\sigma}_{\max}^2 + (n-p) \hat{\sigma}^2} \quad (8)$$

เมื่อ Λ คือ เมทริกซ์ของเวกเตอร์ค่าเฉพาะ (Eigenvalue) ของเมทริกซ์ $\mathbf{W}'\mathbf{W}$

λ_i เป็นค่าเฉพาะตัวที่ i ของเมทริกซ์ Λ และ $\lambda_{\max} = \max_{1 \leq i \leq p} \lambda_i$

นิสาชล [5] ได้ศึกษาการเปรียบเทียบการคัดเลือกตัวแปรอิสระที่มีปัญหาสหสัมพันธ์เชิงเส้นพหุ ด้วยวิธีการถดถอยแบบบริดจ์และการค้นหาแบบต้องห้าม พบว่า วิธีการค้นหาแบบต้องห้ามที่มีฟังก์ชันเป้าหมายเป็นค่าความคลาดเคลื่อนกำลังสองเฉลี่ยปรับด้วยฟังก์ชันการลงโทษ มีร้อยละของการคัดเลือกได้ตัวแบบจริงสูงและค่อนข้างคงที่ โดยไม่ขึ้นกับขนาดตัวอย่างและค่าสัมประสิทธิ์สหสัมพันธ์

การคัดเลือกตัวแบบการถดถอยมีได้หลายเกณฑ์ ที่นิยมใช้ คือ สัมประสิทธิ์ตัวกำหนดปรับปรุง (Adjusted Coefficient of Determination) [6] หรือ R_{adj}^2 นักวิจัยเข้าใจว่าตัวแบบใดให้ R_{adj}^2 ค่าสูงที่สุด แสดงว่าตัวแบบนั้นเหมาะสมที่สุด แต่ความเป็นจริงตัวแบบที่ให้ค่า R_{adj}^2 สูงที่สุด จะทำให้ตัวแบบที่ได้รับการคัดเลือก มีจำนวนตัวแปรที่มากเกินไป (Overspecification) [7] ด้วยเหตุนี้ มอลโลว์ [8] ได้เสนอเกณฑ์การคัดเลือกตัวแบบที่เกี่ยวข้องกับค่าคลาดเคลื่อนกำลังสองเฉลี่ย (Mean Square Error: MSE) ของค่าประมาณ $\hat{\mathbf{y}}$ หาดด้วย σ^2 คือ เกณฑ์ซีพี ซึ่งเขียนเป็นสูตรได้ดังนี้

$$\Gamma_p = \frac{1}{\sigma^2} \left\{ E \left[(\hat{\mathbf{y}} - E(\hat{\mathbf{y}}))' (\hat{\mathbf{y}} - E(\hat{\mathbf{y}})) \right] + (E(\hat{\mathbf{y}}) - E(\mathbf{y}))' (E(\hat{\mathbf{y}}) - E(\mathbf{y})) \right\}$$

จากสมการข้างต้นพบว่า สถิติซีพีสามารถเขียนได้ในรูปของ $\text{Var}(\hat{\mathbf{y}})$ และความเอนเอียงของค่าทำนาย $\hat{\mathbf{y}}$ สามารถเขียนได้ในรูปของ

$$\Gamma_p = \frac{1}{\sigma^2} \left\{ \text{Var}(\hat{\mathbf{y}}) + (E(\mathbf{y}))' (\mathbf{I} - \mathbf{H}) E(\mathbf{y}) \right\}$$

เนื่องจาก $(\mathbf{I} - \mathbf{H})$ เป็นเมทริกซ์นิพจน์และสมมาตร (Symmetric Idempotent Matrix) สามารถเขียนได้เป็น

$$\begin{aligned} \Gamma_p &= \frac{1}{\sigma^2} \left\{ p\sigma^2 - \text{tr}(\mathbf{I} - \mathbf{H})\sigma^2 + E[\mathbf{y}'(\mathbf{I} - \mathbf{H})\mathbf{y}] \right\} \\ &= \frac{1}{\sigma^2} \left\{ p\sigma^2 - (n-p)\sigma^2 + E(SSE) \right\} \end{aligned}$$

$$\text{ดังนั้น} \quad \Gamma_p = \frac{E(SSE)}{\sigma^2} - (n-2p)$$

เมื่อ $E(\mathbf{y})$ เป็นค่าคาดหวังของตัวแปรตามจากตัวแบบการถดถอย

$E(\hat{\mathbf{y}})$ เป็นค่าคาดหวังของค่าทำนายจากตัวแบบที่มีพารามิเตอร์ p และ $\mathbf{H} = \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'$

ให้ $\hat{\sigma}^2 = \text{MSE}$ เป็นตัวประมาณค่าที่ไม่เอนเอียงของ σ^2 ของตัวแบบเต็มรูป และแทนค่าของ $E(SSE)$ ด้วย SSE_p และ ให้ C_p เป็นตัวประมาณค่าของ Γ_p ดังนั้น

$$C_p = \frac{\text{SSE}_p}{\text{MSE}} - n + 2p \quad (9)$$

ตัวแบบการถดถอยที่เหมาะสมเป็นตัวแบบที่มีค่า C_p และมีค่าใกล้เคียงกับ p ถ้า C_p มีค่าเท่ากับ p ตัวประมาณ C_p จะไม่มีความเอนเอียง ถ้า C_p มีค่ามากกว่า p ตัวประมาณ C_p จะมีความเอนเอียงมาก ถ้า C_p มีค่าน้อยกว่า p ตัวประมาณ C_p จะมีความเอนเอียงน้อย [9]



การคัดเลือกตัวแบบโดยใช้เกณฑ์ซีพีเมื่อตัวอย่างมีขนาดเล็ก จะทำให้ตัวแบบที่ได้รับการคัดเลือกมีจำนวนตัวแปรอิสระในตัวแบบมากเกินไป (Overspecification) ด้วยเหตุนี้ รวบรวมและอ้างถึงใน รจนา [10] ได้เสนอเกณฑ์การคัดเลือกตัวแบบขึ้นมาใหม่ โดยลดโอกาสที่จะทำให้ตัวแบบที่ได้รับการคัดเลือกมีจำนวนตัวแปรอิสระมากกว่าหรือน้อยกว่าความเป็นจริงจึงทำให้เกณฑ์ที่เสนอขึ้นใหม่นี้คัดเลือกรูปแบบได้ถูกต้องมากขึ้นเมื่อตัวอย่างมีขนาดเล็ก

Akaike [11] ได้เสนอเกณฑ์ข้อสนเทศของอาไคเคะ (Akaike's Information Criterion : AIC) หรือเกณฑ์เอไอซี โดยตัวแบบที่เหมาะสมคือตัวแบบที่ให้ค่า AIC น้อยที่สุด แต่เกณฑ์เอไอซีมีความผิดพลาดในการคัดเลือกตัวแบบที่เหมาะสมสูงเมื่อตัวอย่างมีขนาดเล็ก และเมื่อตัวอย่างมีขนาดเล็กเมื่อเทียบกับจำนวนพารามิเตอร์ที่ต้องการประมาณ จะทำให้ตัวแบบที่ได้รับการคัดเลือกมีจำนวนตัวแปรอิสระในตัวแบบมากเกินไป (Overspecification)

จากเกณฑ์การคัดเลือกตัวแบบที่ได้เสนอข้างต้น จะทำการคัดเลือกตัวแบบที่เหมาะสมโดยพิจารณาจากตัวแบบที่ให้ค่าเกณฑ์นั้นๆ ต่ำที่สุด เช่น หากพิจารณาจากเกณฑ์ซีพี จะต้องคำนวณค่า C_p ทุกตัวแบบที่เป็นไปได้ ตัวแบบใดให้ค่า C_p ต่ำที่สุด ถือว่าตัวแบบนั้นเหมาะสมที่สุด นอกจากนี้ผู้วิจัยจะต้องใช้วิจารณ์ฐานในการพิจารณาว่าควรเลือกตัวแบบใด เพราะตัวแบบที่พิจารณาอาจให้ค่าของเกณฑ์ที่ใกล้เคียงกัน แต่ค่าของเกณฑ์ที่ลดลงอาจจะไม่มีนัยสำคัญ จึงควรหยุดกระบวนการคัดเลือก ดังนั้น พลากร [12] จึงได้เสนอสถิติทดสอบเพื่อคัดเลือกตัวแบบการถดถอยเชิงเส้นพหุคูณ คือสถิติทดสอบซีพี และเปรียบเทียบความถูกต้องของการเลือกตัวแบบที่ใช้เกณฑ์ซีพีด้วยเกณฑ์ต่ำสุด สถิติทดสอบที่นำเสนอได้แก่

$$T_{C_p} = \frac{(C_p + n - 2p) - (n - p_{full})}{\sqrt{2(n - p_{full})}} \quad (10)$$

$$= \frac{(C_p - 2p) + p_{full}}{\sqrt{2(n - p_{full})}} \rightarrow N(0,1) \quad (11)$$

สถิติทดสอบ T_{C_p} มีการแจกแจงเข้าสู่การแจกแจงแบบปกติ

เมื่อ C_p เป็นค่าของเกณฑ์ซีพี ที่คำนวณได้จากสมการที่กำลังพิจารณา

p เป็นจำนวนตัวประมาณค่าของสัมประสิทธิ์การถดถอยในสมการที่กำลังพิจารณา

p_{full} เป็นจำนวนตัวประมาณค่าของสัมประสิทธิ์การถดถอยในสมการที่มีตัวแปรอิสระครบทุกตัว

และ n เป็นขนาดของตัวอย่าง

สถิติทดสอบนี้ใช้ในการทดสอบสมมติฐาน $H_0 : \Gamma_p = \Gamma_{full}$ เทียบกับ $H_a : \Gamma_p \neq \Gamma_{full}$ เมื่อ Γ_{full} เป็นค่าคาดหมายของค่าคลาดเคลื่อนกำลังสองของค่าพยากรณ์กับค่าคาดหมายของ y หาดด้วย σ^2 จากตัวแบบการถดถอยเต็มรูป จะปฏิเสธ H_0 เมื่อ $|T_{C_p}|$ มีค่ามากกว่า $Z_{\frac{\alpha}{2}}$ ซึ่งหมายถึง ไม่สามารถนำตัวแปรอิสระเข้าสู่ตัวแบบหรือออกจากตัวแบบได้

ในการวิจัยครั้งนี้ผู้วิจัยได้ใช้วิธีการคัดเลือกตัวแปรเข้าสู่ตัวแบบการถดถอยแบบบริดจ์โดยใช้ด้วยวิธีเลือกตัวแปรอิสระไปข้างหน้า (Forward Selection) และ วิธีตัดตัวแปรอิสระออกแบบถดถอยหลัง (Backward Elimination) โดยใช้เกณฑ์ซีพี (Cp Criterion) และสถิติทดสอบซีพี (Cp-Test) ในการวิเคราะห์การถดถอยแบบบริดจ์ ที่มีการประมาณค่าพารามิเตอร์วิธี 5 วิธี คือ วิธีของ Hoerl and Kennard (HK) วิธีของ Hoerl Kennard and Baldwin (HKB) วิธีของ Lawless and Wang (LW) วิธี Khalaf and Shukar (KS) และวิธีกำลังสองน้อยที่สุด (LS)

วัตถุประสงค์ของงานวิจัย

เพื่อเปรียบเทียบความถูกต้องในการคัดเลือกตัวแปรเข้าสู่ตัวแบบการถดถอยแบบบริดจ์โดยใช้ด้วยวิธีเลือกตัวแปรอิสระไปข้างหน้า และ วิธีตัดตัวแปรอิสระออกแบบถอยหลัง โดยใช้เกณฑ์ซีพี และสถิติทดสอบซีพี เป็นเกณฑ์ในการพิจารณาตัวแบบที่ใช้วิธีการประมาณค่าพารามิเตอร์บริดจ์ 5 วิธี คือ วิธีของ Hoerl and Kennard วิธีของ Hoerl Kennard and Baldwin วิธีของ Lawless and Wang วิธี Khalaf and Shukar และวิธีกำลังสองน้อยที่สุด

วิธีการวิจัย

1. วิธีดำเนินการ

งานวิจัยนี้ ใช้วิธีการจำลองข้อมูลในการเปรียบเทียบการคัดเลือกตัวแบบโดยในตัวแบบมีจำนวนตัวแปรอิสระ 5 ตัว คือ x_1, x_2, x_3, x_4 และ x_5 โดยตัวแปรอิสระ x_1, x_2 และ x_3 เกี่ยวข้องกับตัวแปรตาม y และตัวแปรอิสระ x_4 และ x_5 มีความสัมพันธ์กันสูงในระดับ 0.95 0.99 0.999 และ 0.9999 ส่วนตัวแปรอิสระ x_4 และ x_5 เป็นตัวแปรอิสระที่ไม่เกี่ยวข้องกับตัวแปรตาม การศึกษาใช้ขนาดตัวอย่างเท่ากับ 15 30 และ 50 ในแต่ละกรณีทำซ้ำจำนวน 100 ครั้ง มีขั้นตอนการดำเนินงานดังนี้

1.1 สร้างประชากรขนาด $N = 1,000,000$ จำนวน 4 ประชากร ตามค่าสัมประสิทธิ์สหสัมพันธ์ของตัวแปรอิสระ x_1 และ x_2 เท่ากับ 0.95, 0.99, 0.999 และ 0.9999 โดยตัวแปรอิสระทั้ง 5 ตัวมีการแจกแจงปกติดังนี้ $x_1 \sim N(10, 2^2)$, $x_2 \sim N(20, 3^2)$, $x_3 \sim N(15, 3^2)$, $x_4 \sim N(30, 2^2)$, $x_5 \sim N(25, 4^2)$ และความคลาดเคลื่อนสุ่ม $\varepsilon \sim N(0, 1)$

1.2 กำหนดค่าพารามิเตอร์ของตัวแบบการถดถอยดังนี้ $\beta' = [1 \ 2 \ 5 \ 0 \ 0]$

1.3 สร้างตัวแบบการถดถอย $y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \beta_4 x_4 + \beta_5 x_5 + \varepsilon$

1.4 เมื่อได้ประชากรตามต้องการ ดำเนินการสุ่มตัวอย่างขนาด 15 30 และ 50

1.5 ทำการคัดเลือกสมการถดถอย เพื่อให้ได้ 1 สมการที่มีตัวแปรอิสระ x_1, x_2 และ x_3 ที่เกี่ยวข้องกับตัวแปรตามครบทุกตัว กล่าวคือได้สมการ $\hat{y} = \hat{\beta}_0 + \hat{\beta}_1 x_1 + \hat{\beta}_2 x_2 + \hat{\beta}_3 x_3$ โดยใช้วิธีการคัดเลือกตัวแปรอิสระไปข้างหน้า และวิธีการตัดตัวแปรอิสระออกแบบถอยหลัง ด้วยวิธีการประมาณค่าพารามิเตอร์บริดจ์ (Ridge Parameter) 5 วิธี คือ (1) วิธี Hoerl and Kennard (2) วิธี Hoerl Kennard and Baldwin (3) วิธี Lawless and Wang (4) วิธี Khalaf and Shukar และ (5) วิธีการประมาณค่าพารามิเตอร์กำลังสองน้อยที่สุด (LS) และได้ใช้วิธีการเลือกตัวแปรอิสระ โดยใช้เกณฑ์ซีพี (C_p Criterion) และ สถิติทดสอบซีพี (C_p -test) เป็นเกณฑ์ในการตัดสินใจ

1.6 คำนวณ ร้อยละของจำนวนครั้งในแต่ละกรณีที่มีวิธีการคัดเลือกตัวแปรอิสระเข้าสู่สมการได้ถูกต้องตามสมการ $\hat{y} = \hat{\beta}_0 + \hat{\beta}_1 x_1 + \hat{\beta}_2 x_2 + \hat{\beta}_3 x_3$ หากสมการที่คัดเลือกได้แตกต่างจากสมการนี้ ถือว่าการคัดเลือกตัวแปรไม่ถูกต้องซึ่งแบ่งได้เป็น 3 ลักษณะดังนี้ ถ้าสมการที่คัดเลือกได้มีตัวแปรอิสระ x_4 หรือ x_5 หรือทั้ง 2 ตัวแปรอยู่ในสมการที่คัดเลือกได้มีลักษณะ Overspecification หากสมการที่คัดเลือกได้มีตัวแปรอิสระ x_1, x_2 และ x_3 หายไปบางตัวเรียกว่าสมการที่ได้มีลักษณะ Underspecification และหากสมการที่คัดเลือกได้มีตัวแปร x_1, x_2 และ x_3 หายไปบางตัว และมีตัวแปร x_4 หรือ x_5 บางตัวอยู่ในสมการถือว่าสมการที่ได้มีลักษณะ Misspecification



2. ขั้นตอนการคัดเลือกตัวแบบ

งานวิจัยนี้ได้ทำการคัดเลือกตัวแปรอิสระโดยใช้วิธีการเลือกแบบไปข้างหน้า (Forward Selection) และวิธีการกำจัดตัวแปรแบบถอยหลัง (Backward Elimination) โดยมีขั้นตอนการคัดเลือกตัวแบบของแต่ละเกณฑ์ดังนี้

2.1 ขั้นตอนการคัดเลือกตัวแปรอิสระโดยเกณฑ์ซีฟี่ด้วยวิธีการเลือกแบบไปข้างหน้า

เนื่องจากตัวแปรอิสระ x_1 และ x_2 มีความสัมพันธ์กันสูงและต้องการคัดเลือกตัวแปรอิสระเข้าสู่สมการโดยใช้เกณฑ์ C_p จึงเริ่มต้นสมการด้วยการมีตัวแปรอิสระ x_1 และ x_2 อยู่ในสมการ มีรายละเอียดในการดำเนินงานดังนี้

- (1) คำนวณค่าซีฟี่ จากทุกสมการที่มีตัวแปรอิสระ 2 ตัว คือ x_1 และ x_2 สมมติให้เป็น

$$C_p(x_1, x_2)$$

- (2) เมื่อนำตัวแปรอิสระที่เหลือ $x_i, i=3,4,5$ เข้าสู่สมการทีละ 1 ตัวแปร คำนวณค่าซีฟี่ของ

ทั้ง 3 สมการ คือ $C_p(x_1, x_2, x_i), i=3,4,5$

- (3) พิจารณาสมการที่สอดคล้องกับค่าซีฟี่ต่ำสุด $C_{p\min} = \min_{3 \leq i \leq 5} C_p(x_1, x_2, x_i)$

- (4) เปรียบเทียบค่า $C_p(x_1, x_2)$ กับ $C_{p\min}$ ถ้า $C_{p\min}$ มีค่าน้อยกว่า $C_p(x_1, x_2)$

ก็สามารถนำตัวแปรอิสระ x_i ที่อยู่ในสมการที่สอดคล้องกับค่าซีฟี่ที่ต่ำสุดเข้าสู่สมการได้ หากค่า $C_{p\min}$ มีค่ามากกว่าหรือเท่ากับ $C_p(x_1, x_2)$ จะไม่สามารถนำตัวแปรอิสระเข้าสู่สมการได้และสิ้นสุดการคัดเลือกตัวแปรอิสระ

- (5) ถ้ามีการนำตัวแปรอิสระเข้าสู่สมการในขั้นตอน (4) ให้ทำซ้ำในกระบวนการขั้นที่ (2)

และ (3) โดยพิจารณาจากตัวแปรที่เหลืออยู่

2.2 ขั้นตอนการคัดเลือกตัวแบบโดยใช้วิธีการเลือกแบบไปข้างหน้าสำหรับสถิติทดสอบซีฟี่

การคัดเลือกตัวแปรอิสระในกรณีนี้ กระทำเหมือนกับขั้นตอน (1) – (5) ทุกขั้นตอนโดยใช้วิธี

การทดสอบค่า C_p แทนการพิจารณา ค่า C_p ต่ำสุด หากค่าของสถิติทดสอบของ C_p ไม่มีนัยสำคัญที่ระดับ α ตัวแปรอิสระที่สอดคล้องกับค่าสถิติทดสอบ C_p สามารถนำเข้าสู่สมการได้กระบวนการคัดเลือกจะดำเนินการเช่นนี้ถึงจนกระทั่งไม่สามารถนำตัวแปรอิสระเข้าสู่สมการได้อีก

2.3 ขั้นตอนการคัดเลือกตัวแบบโดยใช้วิธีการกำจัดตัวแปรแบบถอยหลัง สำหรับเกณฑ์ซีฟี่

- (1) คำนวณค่าซีฟี่ จากตัวแบบที่มีตัวแปรอิสระทุกตัว สมมติให้เป็น $C_p(x_1, x_2, x_3, x_4, x_5)$

- (2) เมื่อนำตัวแปรอิสระ $x_i, i=3,4,5$ ออกจากสมการทีละ 1 ตัวแปร คำนวณค่าซีฟี่ของทั้ง

3 สมการ คือ $C_p(x_1, x_2, x_3, x_4), C_p(x_1, x_2, x_3, x_5), C_p(x_1, x_2, x_4, x_5)$

- (3) พิจารณาสมการที่สอดคล้องกับค่าซีฟี่ต่ำสุด

$$C_{p\min} = \min[C_p(x_1, x_2, x_3, x_4), C_p(x_1, x_2, x_3, x_5), C_p(x_1, x_2, x_4, x_5)]$$

- (4) เปรียบเทียบค่า $C_p(x_1, x_2, x_3, x_4, x_5)$ กับ $C_{p\min}$ ถ้า $C_{p\min}$ มีค่าน้อยกว่า

$C_p(x_1, x_2, x_3, x_4, x_5)$ ก็สามารถนำตัวแปรอิสระ x_i ที่อยู่ในสมการที่สอดคล้องกับค่าซีฟี่ที่ต่ำสุดออกจากสมการได้ หากค่า $C_{p\min}$ มีค่ามากกว่าหรือเท่ากับ $C_p(x_1, x_2, x_3, x_4, x_5)$ จะไม่สามารถนำตัวแปรอิสระออกจากสมการได้และสิ้นสุดการคัดเลือกตัวแปรอิสระ โดยห้ามนำตัวแปร x_1 และ x_2 ออกจากตัวแบบ เนื่องจากตัวแปรอิสระ x_1 และ x_2 มีความสัมพันธ์กันสูง

- (5) ถ้ามีการนำตัวแปรอิสระออกจากสมการในขั้นตอน (4) ให้ทำซ้ำในกระบวนการขั้นที่

(2) และ (3) โดยพิจารณาจากตัวแปรที่เหลืออยู่

2.4 ขั้นตอนการคัดเลือกตัวแบบโดยใช้วิธีการกำจัดตัวแปรแบบถอยหลัง สำหรับสถิติทดสอบซีพี

การคัดเลือกตัวแปรอิสระในกรณีนี้ กระทำเหมือนกับขั้นตอน (1) – (5) ทุกขั้นตอนโดยใช้วิธีการทดสอบค่า C_p แทนการพิจารณา ค่า C_p ต่ำสุด หากค่าของสถิติทดสอบของ C_p ไม่มีนัยสำคัญที่ระดับ α ตัวแปรอิสระที่สอดคล้องกับค่าสถิติทดสอบ C_p สามารถนำออกจากสมการได้ กระบวนการคัดเลือกจะดำเนินการเช่นนี้จนกระทั่งไม่สามารถนำตัวแปรอิสระออกจากสมการได้อีก

ผลการวิจัย

ผลการคัดเลือกตัวแบบด้วยเกณฑ์ซีพีและใช้สถิติทดสอบซีพี เมื่อจำแนกตามขนาดตัวอย่าง ที่ระดับนัยสำคัญ 0.05 ได้ผลดังตารางที่ 1

จากตารางที่ 1	HK เป็นการประมาณค่าพารามิเตอร์รีดจ์โดยใช้วิธีของ Hoerl and Kennard
	HKB เป็นการประมาณค่าพารามิเตอร์รีดจ์โดยใช้วิธีของ Hoerl Kennard and Baldwin
	LW เป็นการประมาณค่าพารามิเตอร์รีดจ์โดยใช้วิธีของ Lawless and Wang
	KS เป็นการประมาณค่าพารามิเตอร์รีดจ์โดยใช้วิธีของ Khalaf and Shukar
	LS เป็นการใส่ค่าพารามิเตอร์รีดจ์ เท่ากับ 0 (วิธีกำลังสองน้อยที่สุด)

เมื่อขนาดตัวอย่างเท่ากับ 15 พบว่า การคัดเลือกตัวแบบโดยใช้วิธีการเลือกแบบไปข้างหน้าและวิธีการกำจัดตัวแปรแบบถอยหลัง ใช้สถิติทดสอบซีพี คัดเลือกตัวแบบได้ถูกต้องมากกว่าเกณฑ์ซีพี และคัดเลือกตัวแบบได้ถูกต้องมากที่สุดเมื่อใช้วิธีการประมาณค่าพารามิเตอร์รีดจ์ โดยใช้วิธีของ Lawless and Wang เมื่อตัวแปรอิสระ x_1 และ x_2 มีสัมประสิทธิ์สหสัมพันธ์เท่ากับ 0.9999

เมื่อขนาดตัวอย่างเท่ากับ 30 พบว่า การคัดเลือกตัวแบบโดยใช้วิธีการเลือกแบบไปข้างหน้าและวิธีการกำจัดตัวแปรแบบถอยหลัง ใช้สถิติทดสอบซีพี คัดเลือกตัวแบบได้ถูกต้องมากกว่าเกณฑ์ซีพี และคัดเลือกตัวแบบได้ถูกต้องมากที่สุดเมื่อใช้วิธีการประมาณค่าพารามิเตอร์รีดจ์ วิธีของ Lawless and Wang โดยคัดเลือกตัวแบบด้วยวิธีการเลือกแบบไปข้างหน้า โดยใช้สถิติทดสอบซีพี เมื่อตัวแปรอิสระ x_1 และ x_2 มีสัมประสิทธิ์สหสัมพันธ์เท่ากับ 0.9999 หรือคัดเลือกตัวแบบด้วยวิธีการกำจัดตัวแปรแบบถอยหลัง โดยใช้สถิติทดสอบซีพีเมื่อตัวแปรอิสระ x_1 และ x_2 มีสัมประสิทธิ์สหสัมพันธ์เท่ากับ 0.9999 พบว่า สามารถคัดเลือกตัวแบบได้ถูกต้องร้อยละ 100 ทั้ง 5 วิธี

การคัดเลือกตัวแบบโดยใช้วิธีการเลือกแบบไปข้างหน้าและวิธีการกำจัดตัวแปรแบบถอยหลังเมื่อใช้สถิติทดสอบซีพี คัดเลือกตัวแบบได้ถูกต้องมากกว่าเกณฑ์ซีพี เมื่อขนาดตัวอย่างเท่ากับ 50 และทุกขนาดความสัมพันธ์ของตัวแปรอิสระระหว่าง x_1 และ x_2 และการใช้สถิติทดสอบซีพีสามารถคัดเลือกตัวแบบได้ถูกต้อง 100% ทั้ง 5 วิธี

อภิปรายและสรุปผลการวิจัย

จากการจำลองข้อมูล ให้ตัวแบบการถดถอยมีตัวแปรอิสระในตัวแบบ 5 ตัว โดยมีตัวแปรที่เกี่ยวข้องกับตัวแปรตาม 3 ตัว มีตัวแปรอิสระ 1 คู่ คือ x_1 และ x_2 มีความสัมพันธ์กันสูงในระดับ 0.95, 0.99, 0.999 และ 0.9999 และตัวแปรอิสระอีก 2 ตัว คือ x_4 และ x_5 เป็นตัวแปรที่ไม่เกี่ยวข้องกับตัวแปรตาม ได้ผลการคัดเลือกตัวแบบดังนี้

ในทุกขนาดตัวอย่างและทุกระดับความสัมพันธ์ของตัวแปรอิสระ x_1 และ x_2 พบว่าการคัดเลือกตัวแบบด้วยวิธีการเลือกแบบไปข้างหน้าและวิธีการกำจัดตัวแปรแบบถอยหลัง โดยใช้สถิติทดสอบซีพี คัดเลือกตัวแบบได้ถูกต้องมากกว่าเกณฑ์ซีพี ซึ่งสอดคล้องกับงานวิจัยของคุณพลากร เพราะสถิติทดสอบซีพีจะคัดเลือกตัวแปรอิสระออกจากตัว



แบบและเอาเข้าสู่ตัวแบบโดยการทดสอบความมีนัยสำคัญในการคัดเลือก แต่เกณฑ์ซีพี จะคัดเลือกตัวแปรอิสระออกจากตัวแบบและเข้าสู่ตัวแบบโดยพิจารณาจากค่าที่ต่ำที่สุด ซึ่งค่าของเกณฑ์ซีพีอาจจะลดลงอย่างไม่มีนัยสำคัญ จะส่งผลให้มีการคัดเลือกตัวแบบที่ผิดพลาดทำให้มีตัวแปรอิสระเกินขึ้นมา และพบว่าเมื่อขนาดตัวอย่างเพิ่มขึ้น ร้อยละของการคัดเลือกตัวแบบที่คัดเลือกได้ถูกต้องเพิ่มขึ้นซึ่งเป็นไปตามทฤษฎีของเกณฑ์ซีพี ที่คัดเลือกตัวแบบได้ดีเมื่อตัวอย่างมีขนาดใหญ่ และสถิติทดสอบซีพี คัดเลือกตัวแบบได้ดีเมื่อขนาดตัวอย่างใหญ่พอ เนื่องจากสถิติทดสอบซีพีเป็นสถิติทดสอบเชิงกำกับ

เมื่อพิจารณาประสิทธิภาพในการคัดเลือกตัวแบบ ความแตกต่างระหว่างความสัมพันธ์ของตัวแปรอิสระระหว่าง x_1 และ x_2 พบว่ามีการคัดเลือกตัวแบบได้ถูกต้องไม่แตกต่างกัน และเมื่อพิจารณาประสิทธิภาพในการคัดเลือกตัวแบบระหว่างวิธีการประมาณค่าพารามิเตอร์วิธีต่างๆ พบว่า มีการคัดเลือกตัวแบบได้ถูกต้องไม่แตกต่างกัน แต่สามารถคัดเลือกได้ถูกต้องสูงกว่าวิธีกำลังสองน้อยที่สุด เมื่อตัวอย่างมีขนาดเล็ก และสามารถคัดเลือกตัวได้ถูกต้องไม่แตกต่างกัน เมื่อตัวอย่างมีขนาดใหญ่ขึ้น

ข้อเสนอแนะ

ในการวิจัยครั้งต่อไป ควรศึกษาการคัดเลือกตัวแบบโดยใช้เกณฑ์อื่นๆ และสถิติทดสอบอื่นๆ เช่น เกณฑ์เอไอซี หรือ สถิติทดสอบเอไอซี หรือ ใช้ตัวแบบการถดถอยแบบริดจ์ที่ใช้วิธีการประมาณค่าพารามิเตอร์วิธีอื่นๆ นอกจากนี้ อาจจะนำไปประยุกต์ใช้กับตัวแบบการถดถอยพหุตัวแปร (Multivariate Regression Analysis) ในกรณีที่ตัวแปรอิสระมีพหุสัมพันธ์กันสูง เป็นต้น

กิตติกรรมประกาศ

ขอขอบคุณกรุงเทพมหานครที่สนับสนุนทุนในการศึกษาต่อ และขอขอบคุณ ศ.ดร.จิราวัลย์ จิตรถเวช และ รศ.ดร.วิจิต หล่อจิระชุมหะกุล คณะสถิติประยุกต์ สถาบันบัณฑิตพัฒนบริหารศาสตร์ ที่ช่วยให้คำปรึกษาตลอดจนแนวคิดที่เป็นประโยชน์อย่างมากสำหรับการทำวิจัยเรื่องนี้

เอกสารอ้างอิง

1. Hoerl AE, Kennard RW. Ridge Regression: Biased Estimation to Nonorthogonal Problem. *Technometrics*. 1970a; 12(2):55-67
2. Hoerl AE, Kennard RW, Baldwin K. Ridge Regression: Some Simulations. *Communications in Statistics*, 1975; 4(2): 105-123.
3. Lawless J, Wang PA. Simulation study of ridge and other regression Estimators. *Communication Statistics – Theory and Methods*, 1976; 5(4): 307-323.
4. Khalaf G, Shukur G. Choosing ridge parameter for regression Problem. *Communication Statistics in Theory and Methods*, 2005; 34: 1177-1182.
5. Nisachon N. A Comparison of Variable Selection with Multicollinearity by Ridge Regression and Tabu Search [Msc thesis]. Bangkok: National Institute of Development Administration; 2013. Thai.
6. McQuarrie A, Robert S, Tsai CL. The model Selection Criterion AICu. *Statistics and Probability Letters*, 1997; 34: 2585-292.
7. Sheather SJ. A Modern Approach to Regression with R. New York: Springer; 2009.



8. Mallows CL. Some Comments on CP. *Technometrics*, 1973; 15(4): 661-675
9. Songsiri T. *Regression Analysis* .3rd ed. Kasetsart University (BKK); 2005.
10. Rotjana K. A Comparison of the Model Selection Criterion in Multiple Linear Regression Analysis [Msc thesis]. Bangkok: Thammasart University; 2006. Thai.
11. Akaike H. Information Theory and an Extension of the Maximum Likelihood Principle. In 2nd International Symposium on Information Theory. B.N. Petrov and F.Cski, eds. Akademiai Kiado, Budapest; 1973; 15(4): 267-281
12. Palakorn S. Test Statistics for Selecting Multiple Linear Regression Models [Msc thesis]. Bangkok: National Institute of Development Administration; 2010. Thai.



ตารางที่ 1 ร้อยละของตัวแบบที่คัดเลือกได้ จำแนกตามวิธีการคัดเลือก วิธีการประมาณค่าพารามิเตอร์รีดจ์
ค่าสัมประสิทธิ์สหสัมพันธ์ และขนาดตัวอย่าง

วิธีการ คัดเลือก	เกณฑ์การ ตัดสิน	วิธีการประมาณค่าพารามิเตอร์รีดจ์									
		LS		HK		HKB		LW		KS	
		ถูก	ไม่ถูก	ถูก	ไม่ถูก	ถูก	ไม่ถูก	ถูก	ไม่ถูก	ถูก	ไม่ถูก
$n = 15, r = 0.95$											
Forward	C_p	63	37	68	32	79	21	78	22	68	32
	$C_p - test$	90	10	96	4	98	2	98	2	96	4
Backward	C_p	63	37	68	32	78	22	77	23	68	32
	$C_p - test$	90	10	96	4	98	2	98	2	96	4
$n = 15, r = 0.99$											
Forward	C_p	63	37	69	31	80	20	78	22	69	31
	$C_p - test$	90	10	95	5	98	2	98	2	95	5
Backward	C_p	63	37	69	31	79	21	77	23	69	31
	$C_p - test$	90	10	95	5	98	2	98	2	95	5
$n = 15, r = 0.999$											
Forward	C_p	63	37	71	29	78	22	81	19	71	29
	$C_p - test$	90	10	96	4	98	2	98	2	96	4
Backward	C_p	63	37	71	29	76	25	79	21	71	29
	$C_p - test$	90	10	96	4	98	2	98	2	96	4
$n = 15, r = 0.9999$											
Forward	C_p	63	37	68	32	73	27	82	18	68	32
	$C_p - test$	90	10	94	6	98	2	99	1	94	6
Backward	C_p	63	37	68	32	71	29	80	20	68	32
	$C_p - test$	90	10	94	6	98	2	99	1	94	6
$n = 30, r = 0.95$											
Forward	C_p	70	30	73	27	76	24	75	25	73	27
	$C_p - test$	99	1	99	1	99	1	99	1	99	1
Backward	C_p	68	32	73	27	76	24	75	25	73	27
	$C_p - test$	99	1	99	1	99	1	99	1	99	1
$n = 30, r = 0.99$											
Forward	C_p	70	30	73	27	76	24	76	24	73	27
	$C_p - test$	99	1	99	1	99	1	99	1	99	1
Backward	C_p	68	32	73	27	76	24	76	24	73	27
	$C_p - test$	99	1	99	1	99	1	99	1	99	1

ตารางที่ 1 ร้อยละของตัวแบบที่คัดเลือกได้ จำแนกตามวิธีการคัดเลือก วิธีการประมาณค่าพารามิเตอร์รีดจ์
ค่าสัมประสิทธิ์สหสัมพันธ์ และขนาดตัวอย่าง (ต่อ)

วิธีการ คัดเลือก	เกณฑ์การ ตัดสิน	วิธีการประมาณค่าพารามิเตอร์รีดจ์									
		LS		HK		HKB		LW		KS	
		ถูก	ไม่ถูก	ถูก	ไม่ถูก	ถูก	ไม่ถูก	ถูก	ไม่ถูก	ถูก	ไม่ถูก
$n = 30, r = 0.999$											
Forward	C_p	70	30	72	28	75	25	75	25	72	28
	$C_p - test$	99	1	99	1	99	1	99	1	99	1
Backward	C_p	68	32	72	28	75	25	75	25	72	28
	$C_p - test$	99	1	99	1	99	1	99	1	99	1
$n = 30, r = 0.9999$											
Forward	C_p	70	30	72	28	75	25	76	24	72	28
	$C_p - test$	99	1	99	1	99	1	100	0	99	1
Backward	C_p	68	32	71	29	75	25	76	24	71	29
	$C_p - test$	100	0	100	0	100	0	100	0	100	0
$n = 50, r = 0.95$											
Forward	C_p	73	27	75	25	76	24	75	25	75	25
	$C_p - test$	100	0	100	0	100	0	100	0	100	0
Backward	C_p	73	27	75	25	76	24	75	25	75	25
	$C_p - test$	100	0	100	0	100	0	100	0	100	0
$n = 50, r = 0.99$											
Forward	C_p	73	27	75	25	77	23	76	24	75	25
	$C_p - test$	100	0	100	0	100	0	100	0	100	0
Backward	C_p	73	27	75	25	77	23	76	24	75	25
	$C_p - test$	100	0	100	0	100	0	100	0	100	0
$n = 50, r = 0.999$											
Forward	C_p	73	27	74	26	75	25	75	25	74	26
	$C_p - test$	100	0	100	0	100	0	100	0	100	0
Backward	C_p	73	27	74	26	75	25	75	25	74	26
	$C_p - test$	100	0	100	0	100	0	100	0	100	0
$n = 50, r = 0.9999$											
Forward	C_p	73	27	74	26	75	25	74	26	74	26
	$C_p - test$	100	0	100	0	100	0	100	0	100	0
Backward	C_p	73	27	74	26	75	25	74	26	74	26
	$C_p - test$	100	0	100	0	100	0	100	0	100	0