เกณฑ์ใหม่สำหรับการคัดเลือกตัวแบบการถดถอย : กรณีตัวอย่างขนาดเล็ก

วรางคณา เรียนสุทธิ์*
มหาวิทยาลัยทักษิณ ต.บ้านพร้าว อ.ป่าพะยอม จ.พัทลุง 93210

** Corresponding Author: warang27@gmail.com*
รองศาสตราจารย์ สาขาวิชาคณิตศาสตร์และสถิติ คณะวิทยาศาสตร์

ข้อมูลบทความ

ประวัติบทความ :
รับเพื่อพิจารณา : 5 สิงหาคม 2564
แก้ไข : 20 พฤษภาคม 2565
ตอบรับ : 19 กรกฎาคม 2565
DOI : 10.14456/kmuttrd.2022.18

คำสำคัญ : เกณฑ์การคัดเลือกตัวแบบ / ตัวแบบการถดถอย / ความน่าจะเป็นที่จะมีจำนวนตัวแปรมากหรือน้อยเกินไป / ประสิทธิภาพสังเกต L_2

บทคัดย่อ

การศึกษาครั้งนี้มีวัตถุประสงค์เพื่อสร้างเกณฑ์ใหม่สำหรับการคัดเลือกตัวแบบการถดถอย กรณีตัวอย่างขนาดเล็ก โดยเรียกชื่อเกณฑ์ใหม่นี้ว่า New Information Criterion (NIC) ดำเนินการตรวจสอบประสิทธิภาพของเกณฑ์ NIC เปรียบเทียบกับเกณฑ์การคัดเลือกตัวแบบการถดถอยอื่น 3 เกณฑ์ ได้แก่ เกณฑ์ $KICc_C$, $KICc_{SB}$ และ $KICc_{HM}$ ภายใต้สถานการณ์ต่าง ๆ ได้แก่ ความแตกต่างของขนาดตัวอย่าง จำนวนตัวแปรอิสระในสมการ สัมประสิทธิ์การถดถอย ความแปรปรวนของความคลาดเคลื่อน และการแจกแจงของตัวแปรอิสระ จากผลการศึกษา พบว่า เกณฑ์ NIC ที่สร้างขึ้นมีสูตรดังนี้

$$\mathrm{NIC} = \log\left(s^2\right) + \log\left(\frac{n}{n-p}\right) + \frac{\left[(n-p)(2p+3)-2\right]}{(n-p-2)(n-p)}$$

จากผลการเปรียบเทียบประสิทธิภาพ พบว่า ประสิทธิภาพของเกณฑ์ $KICc_{HM}$ ดีที่สุด อย่างไรก็ตาม เกณฑ์นี้สามารถคัดเลือกตัวแบบได้ถูกต้องน้อย ดังนั้น งานวิจัยนี้จึงใช้เกณฑ์ประสิทธิภาพสังเกต L_2 เป็นเกณฑ์การคัดเลือกตัวแบบอีกเกณฑ์หนึ่ง ซึ่งเกณฑ์นี้แนะนำว่าเกณฑ์ NIC คือเกณฑ์ที่ดีที่สุด

New Criteria for Selection of Regression Models: Case of Small-Size Sample

Warangkhana Riansut*
Thaksin University, Banpraw, Pa Phayom, Phatthalung 93210

* *Corresponding Author: warang27@gmail.com*
Associate Professor, Department of Mathematics and Statistics, Faculty of Science.

Article Info

Article History:
Received: August 5, 2021
Revised: May 20, 2022
Accepted: July 19, 2022
DOI : 10.14456/kmuttrd.2022.18

Keywords : Model Selection Criterion / Regression Model / Probability of Overfitting or Underfitting / Observed L_2 Efficiency

Abstract

The objective of this study was to create a new criterion viz. the New Information Criterion (NIC) for the selection of regression models in a small-size sample case. The performance of NIC was compared to those of three other model selection criteria, namely, $KICc_C$, $KICc_{SB}$ and $KICc_{HM}$. The conditions for the simulation were the differences in the sample size, number of parameters in the model, regression coefficient, error variance, and distribution of independent variables. The results showed that NIC exhibits the following formula:

$$NIC = \log\left(s^2\right) + \log\left(\frac{n}{n-p}\right) + \frac{\left[(n-p)(2p+3)-2\right]}{(n-p-2)(n-p)}$$

The performance comparison results revealed that $KICc_{HM}$ performed the best. However, such a criterion could identify the true model less accurately. Therefore, this research used the observed L_2 efficiency as another model selection criterion. This criterion suggested that NIC was the best criterion.

1. บทนำ

การคัดเลือกตัวแบบการถดถอยที่เหมาะสมจากตัวแบบที่เป็นไปได้ทั้งหมดนับเป็นปัญหาหนึ่งที่สำคัญ ในทางปฏิบัติเราไม่สามารถทราบได้เลยว่าตัวแบบที่แท้จริง (True Model) คือตัวแบบใด แต่เราคาดหวังว่าจะสามารถหาตัวแบบที่มาเป็นตัวแทนที่ถูกต้องและเหมาะสมได้ เกณฑ์การคัดเลือกตัวแบบเป็นเครื่องมือหนึ่งที่มีประโยชน์ในการพิจารณาความเหมาะสมของตัวแบบ โดยการประเมินว่าตัวแบบใดที่ให้ความสมดุลที่ดีที่สุด กล่าวคือ สามารถคัดเลือกได้ว่าตัวแบบใดมีจำนวนตัวแปรในสมการที่เพียงพอ ถูกต้อง และเหมาะสมไม่มากและไม่น้อยเกินไป (Goodness of Fit และ Parsimony) เกณฑ์การคัดเลือกตัวแบบเกณฑ์แรกที่เป็นที่รู้จัก คือ เกณฑ์สารสนเทศของอะกะอิเกะ (Akaike's Information Criterion: AIC) [1] เกณฑ์นี้เป็นตัวประมาณไม่เอนเอียงเชิงเส้นกำกับ (Asymptotically Unbiased Estimator) ของความผันแปรของความแตกต่างแบบทางตรงของคูลแบค (Kullback's Directed Divergence) ซึ่งความแตกต่างแบบทางตรงเป็นที่รู้จักกันในนามของสารสนเทศของคูลแบคและลีเบลอ (Kullback-Leibler Information) เพื่อประเมินความแตกต่างระหว่างตัวแบบทางสถิติ 2 ตัวแบบ คือ ตัวแบบที่แท้จริงและตัวแบบโดยประมาณ (Fitted Model) เกณฑ์การคัดเลือกตัวแบบเกณฑ์อื่น ๆ ที่เป็นที่รู้จักต่อมา เช่น เกณฑ์การปรับแก้ของ AIC (AICc) [2] เกณฑ์การปรับแก้ของ AICc (AICu) [3] เกณฑ์สารสนเทศของคูลแบค (Kullback Information Criterion: KIC) [4] เกณฑ์การปรับแก้ของ KIC โดย Cavanaugh ($KICc_C$) [5] เกณฑ์การปรับแก้ของ KIC โดย Seghouane and Bekara ($KICc_{SB}$) [6] และเกณฑ์การปรับแก้ของ KIC โดย Hafidi และ Mkhadri ($KICc_{HM}$) [7] ถึงแม้ว่าเกณฑ์ AIC ยังคงเป็นที่นิยมใช้กันมากที่สุดในการคัดเลือกตัวแบบ แต่ในกรณีตัวอย่างขนาดเล็กหรือสัดส่วนของจำนวนพารามิเตอร์ในตัวแบบเทียบกับขนาดตัวอย่างมีค่าสูง เกณฑ์ AIC มีโอกาสที่จะคัดเลือกตัวแบบที่มีจำนวนตัวแปรมากเกินไป ดังนั้น Hurvich และ Tsai [2] จึงเสนอเกณฑ์ AICc เพื่อแก้ปัญหาดังกล่าว ถึงแม้ว่า Hurvich and Tsai [2] จะเสนอว่าเกณฑ์ AICc เป็นตัวประมาณไม่เอนเอียงของความผันแปรของความแตกต่างแบบทางตรงของคูลแบค แต่ McQuarrie และคณะ [3] ได้สรุปว่าเกณฑ์ AICc มีแนวโน้มที่จะคัดเลือกตัวแบบที่มีจำนวนตัวแปรมากเกินไปเมื่อขนาดตัวอย่างเพิ่มขึ้น ดังนั้น McQuarrie และคณะ [3] จึงเสนอเกณฑ์ AICu ซึ่งเป็นเกณฑ์ที่ดีกว่าเมื่อตัวอย่างมีขนาดปานกลางถึงขนาดใหญ่ ยกเว้นเมื่อตัวแบบที่แท้จริงมีลำดับอนันต์ (Infinite Order) นอกเหนือไปจากนี้ ยังพบการศึกษาของเกณฑ์การคัดเลือกตัวแบบในกรณีตัวอย่างขนาดเล็กอีกกลุ่มหนึ่ง คือ เกณฑ์ KICc ได้แก่ เกณฑ์ $KICc_C$, $KICc_{SB}$ และ $KICc_{HM}$ ซึ่งเกิดจากการปรับแก้ของเกณฑ์ KIC เพื่อแก้ปัญหาการประมาณค่าความแตกต่างแบบสมมาตรของคูลแบค (Kullback's Symmetric Divergence) ต่ำเกินไป อย่างไรก็ตาม เกณฑ์การคัดเลือกตัวแบบการถดถอยที่กล่าวมายังมีข้อดีและข้อเสียในแต่ละประเด็นที่แตกต่างกันไป ดังนั้น การศึกษาครั้งนี้จึงต้องการสร้างเกณฑ์ใหม่สำหรับการคัดเลือกตัวแบบการถดถอย กรณีตัวอย่างขนาดเล็ก และเรียกชื่อเกณฑ์ใหม่นี้ว่า New Information Criterion (NIC) หลังจากนั้น จึงทำการตรวจสอบประสิทธิภาพของเกณฑ์ NIC โดยพิจารณาร้อยละของการคัดเลือกตัวแบบได้ถูกต้อง ร้อยละของการมีจำนวนตัวแปรในสมการมากหรือน้อยเกินไป และประสิทธิภาพสังเกต L_2 (Observed L_2 Efficiency) เปรียบเทียบกับเกณฑ์การคัดเลือกตัวแบบการถดถอยอื่นอีก 3 เกณฑ์ ได้แก่ เกณฑ์ $KICc_C$, $KICc_{SB}$ และ $KICc_{HM}$ เพื่อที่จะสรุปว่าควรใช้เกณฑ์ใดภายใต้สถานการณ์ใด โดยสาเหตุที่ผู้วิจัยเลือกที่จะเปรียบเทียบประสิทธิภาพของเกณฑ์ NIC กับเกณฑ์ $KICc_C$, $KICc_{SB}$ และ $KICc_{HM}$ เนื่องจากเกณฑ์ NIC พัฒนาบนพื้นฐานของเกณฑ์ $KICc_C$ และเกณฑ์ $KICc_C$ เป็นเกณฑ์ในตระกูล KICc เช่นเดียวกับเกณฑ์ $KICc_{SB}$ และ $KICc_{HM}$ การศึกษาครั้งนี้ได้จำลองข้อมูลภายใต้สถานการณ์ต่าง ๆ ได้แก่ ความแตกต่างของขนาดตัวอย่าง จำนวนตัวแปรอิสระในสมการ สัมประสิทธิ์การถดถอย ความแปรปรวนของความคลาดเคลื่อน และการแจกแจงของตัวแปรอิสระ เกณฑ์ที่จะถูกจัดให้เป็นเกณฑ์ที่ดีที่สุดเมื่อมีร้อยละของการคัดเลือกตัวแบบได้ถูกต้องสูงที่สุด ร้อยละของการมีจำนวนตัวแปรในสมการมากหรือน้อยเกินไปต่ำที่สุด และมีประสิทธิภาพสังเกต L_2 สูงที่สุด กล่าวคือ มีค่าเฉลี่ยของประสิทธิภาพสังเกต L_2 สูงที่สุด และมีส่วนเบี่ยงเบนมาตรฐานของประสิทธิภาพสังเกต L_2 ต่ำที่สุด

2. วิธีการวิจัย

การศึกษาครั้งนี้จะศึกษาเฉพาะตัวแบบการถดถอย โดยมีตัวแบบที่แท้จริงหรือตัวแบบที่เกิดจากการจำลองข้อมูลในรูปแบบดังนี้ [8]

$$\mathbf{y}=\mathbf{x}_0\boldsymbol{\beta}_0+\boldsymbol{\varepsilon}_0 \quad (1)$$

และตัวแบบโดยประมาณหรือตัวแบบที่เกิดจากการประมาณค่าพารามิเตอร์ด้วยวิธีการวิเคราะห์การถดถอยในรูปแบบดังนี้

$$\mathbf{y}=\mathbf{x}\boldsymbol{\beta}+\boldsymbol{\varepsilon} \quad (2)$$

โดยที่ $\mathbf{y}$ แทนเวกเตอร์เชิงสุ่มขนาด $n \times \mathbf{x}_0$ และ $\mathbf{x}$ แทนเมตริกซ์ของตัวแปรอิสระขนาด $n \times p_0$ และ $n \times p$ ตามลำดับ $\boldsymbol{\beta}_0$ และ $\boldsymbol{\beta}$ แทนเวกเตอร์ของสัมประสิทธิ์การถดถอยขนาด $p_0 \times 1$ และ $p \times 1$ ตามลำดับ $\boldsymbol{\varepsilon}_0$ และ $\boldsymbol{\varepsilon}$ แทนเวกเตอร์ของความคลาดเคลื่อนที่มีค่าเฉลี่ยเท่ากับ 0 และความแปรปรวนเท่ากับ $\sigma_0^2\mathbf{I}_n$ และ $\sigma^2\mathbf{I}_n$ ตามลำดับ n แทนขนาดตัวอย่าง p_0 แทนจำนวนพารามิเตอร์ในตัวแบบที่แท้จริง p แทนจำนวนตัวประมาณในตัวแบบโดยประมาณ โดยรวมค่าคงที่ด้วย ตัวประมาณภาวะน่าจะเป็นสูงสุด (Maximum Likelihood Estimator) ของ $\boldsymbol{\beta}_0$ และ σ_0^2 แสดงตามลำดับดังนี้

$$\boldsymbol{\beta}=(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y} \text{ และ } \hat{\sigma}^2=\frac{1}{n}(\mathbf{y}-\mathbf{X}\boldsymbol{\beta})'(\mathbf{y}-\mathbf{X}\boldsymbol{\beta})=\frac{SSE}{n} \quad (3)$$

และตัวประมาณไม่เอนเอียง (Unbiased Estimator) ของ σ_0^2 คือ

$$s^2=\frac{1}{n-p}(\mathbf{y}-\mathbf{X}\boldsymbol{\beta})'(\mathbf{y}-\mathbf{X}\boldsymbol{\beta})=\frac{SSE}{n-p}=\frac{n\hat{\sigma}^2}{n-p} \quad (4)$$

โดยที่ SSE แทนผลบวกกำลังสองของความคลาดเคลื่อน (Error Sum of Squares)

จากการศึกษาของ Keerativibool [9] และ Keerativibool and Siripanich [10] ที่พบว่า เกณฑ์ $KICc_C$ ให้ค่าที่ใกล้เคียงกับค่าคาดหมายของตัวประมาณค่าวัดที่สมมาตร (Expected Estimated Symmetric Measure) มากที่สุด และมีค่าของฟังก์ชันการลงโทษ (Penalty Function) มากที่สุด เมื่อเงื่อนไขเป็นจริง โดยเกณฑ์การคัดเลือกตัวแบบที่มีค่าของฟังก์ชันการลงโทษมากจะทำให้ได้ตัวแบบที่มีจำนวนตัวแปรน้อยเกินไปและมีค่าของเกณฑ์การคัดเลือกตัวแบบลู่เข้าสู่ค่าจริงช้า อย่างไรก็ตาม เกณฑ์ประสิทธิภาพสังเกต L_2 แนะนำว่าเกณฑ์ $KICc_C$ เป็นเกณฑ์ที่ดีที่สุด และจากการศึกษาของ McQuarrie และคณะ [3] ได้แทนที่ตัวประมาณภาวะน่าจะเป็นสูงสุดของ σ_0^2 ในสมการที่ (3) ด้วยตัวประมาณไม่เอนเอียงในสมการที่ (4) ลงในเกณฑ์ AICc ทำให้เกณฑ์ AICu มีฟังก์ชันการลงโทษที่มีค่าสูงกว่าจึงแก้ปัญหาการมีจำนวนตัวแปรมากเกินไปเมื่อขนาดตัวอย่างเพิ่มขึ้นได้ ดังนั้นจากองค์ความรู้ทั้ง 2 ผู้วิจัยจึงแทนที่ $\hat{\sigma}^2$ ในสมการที่ (3) ด้วย s^2 ในสมการที่ (4) ลงในสูตรของเกณฑ์ $KICc_C$ ดังนี้

$$KICc_C=\log(\hat{\sigma}^2)+\log\left(\frac{n}{n-p}\right)+\frac{\left[(n-p)(2p+3)-2\right]}{(n-p-2)(n-p)} \quad (5)$$

จะได้เกณฑ์ใหม่สำหรับการคัดเลือกตัวแบบการถดถอย กรณีตัวอย่างขนาดเล็ก (New Information Criterion: NIC) ดังนี้

$$NIC=\log(s^2)+\log\left(\frac{n}{n-p}\right)+\frac{\left[(n-p)(2p+3)-2\right]}{(n-p-2)(n-p)} \quad (6)$$

ขั้นตอนการดำเนินการวิจัย แสดงดังนี้

1. สร้างเกณฑ์ใหม่สำหรับการคัดเลือกตัวแบบการถดถอยและเรียกชื่อเกณฑ์ใหม่นี้ว่า New Information Criterion (NIC)

2. จำลองข้อมูลภายใต้สถานการณ์ต่าง ๆ จำนวน 1,000 ชุด ได้แก่ ข้อมูลที่มีความแตกต่างของขนาดตัวอย่าง จำนวนพารามิเตอร์ในตัวแบบ สัมประสิทธิ์การถดถอย ความแปรปรวนของความคลาดเคลื่อน และการแจกแจงของตัวแปรอิสระ ดังนี้

- ขนาดตัวอย่าง 2 ระดับ ได้แก่ n = 15 และ 30
- จำนวนพารามิเตอร์ในตัวแบบ 2 ระดับ ได้แก่ p_0= 3 และ p_0 = 5
- สัมประสิทธิ์การถดถอย 4 ตัวแบบ ได้แก่

  ตัวแบบที่ 1: β_0 = 1, β_1 = 0.5, β_2 = 0.4, β_3 = 0.3, β_4 = 0.2

  ตัวแบบที่ 2: β_0 = 1, β_1 = 0.5, β_2 = 0.4

ตัวแบบที่ 3: $\beta_0 = 1, \beta_1 = \beta_2 = 2$
ตัวแบบที่ 4: $\beta_0 = 1, \beta_1 = \beta_2 = \beta_3 = \beta_4 = 2$

- ความคลาดเคลื่อนมีการแจกแจงปรกติ มีค่าเฉลี่ยเท่ากับ 0 และความแปรปรวน 3 ระดับ ได้แก่ ความแปรปรวนต่ำ ($\sigma_0^2 = 0.25$) ความแปรปรวนปานกลาง ($\sigma_0^2 = 1$) และความแปรปรวนสูง ($\sigma_0^2 = 9$)
- การแจกแจงของตัวแปรอิสระทั้งหมด 6 ตัว จำนวน 2 การแจกแจง ได้แก่ การแจกแจงปรกติมาตรฐาน (Standard Normal Distribution) นั่นคือ $X_0 \sim N(0, I)$ และการแจกแจงเอกรูป (Uniform Distribution) ในช่วง (a, b) ดังนี้

$X_0 \sim U(5, 10)$, $X_2 \sim U(10, 20)$, $X_3 \sim U(7, 9)$, $X_4 \sim U(6, 11)$, $X_5 \sim U(9, 19)$, $X_6 \sim U(4, 8)$

ดังนั้นตัวแบบการถดถอยหรือตัวแบบที่แท้จริงที่พิจารณาในการศึกษาครั้งนี้ คือ

ตัวแบบที่ 1 หรือตัวแบบระบุยากมาก (Very Weakly Identifiable) มีตัวแปรอิสระจำนวน 4 ตัว และมีค่าสัมประสิทธิ์การถดถอยต่ำ

$$y = 1 + 0.5X_1 + 0.4X_2 + 0.3X_3 + 0.2X_4 + \varepsilon_0$$

ตัวแบบที่ 2 หรือตัวแบบระบุยาก (Weakly Identifiable) มีตัวแปรอิสระจำนวน 2 ตัว และมีค่าสัมประสิทธิ์การถดถอยต่ำ

$$y = 1 + 0.5X_1 + 0.4X_2 + \varepsilon_0$$

ตัวแบบที่ 3 หรือตัวแบบระบุง่ายมาก (Very Strongly Identifiable) มีตัวแปรอิสระจำนวน 2 ตัว และมีค่าสัมประสิทธิ์การถดถอยสูง

$$y = 1 + 2X_1 + 2X_2 + \varepsilon_0$$

ตัวแบบที่ 4 หรือตัวแบบระบุง่าย (Strongly Identifiable) มีตัวแปรอิสระจำนวน 4 ตัว และมีค่าสัมประสิทธิ์การถดถอยสูง

$$y = 1 + 2X_1 + 2X_2 + 2X_3 + 2X_4 + \varepsilon_0$$

3. จากการสร้างตัวแปรอิสระทั้งหมด 6 ตัว จะพิจารณาตัวแปรอิสระเข้าตัวแบบครั้งละ 1 ตัว ตามลำดับ ทั้งหมด 6 ครั้ง (Sequentially Nested) โดยไม่ใช้การนำเข้าของตัวแปรอิสระทุกตัวที่เป็นไปได้ (All Possible Subset) เนื่องจากการทบทวนวรรณกรรมพบว่า วิธีการนำตัวแปรอิสระเข้าตัวแบบครั้งละ 1 ตัว ตามลำดับ หรือการนำเข้าทุกตัวที่เป็นไปได้มีผลการทดลองที่ไม่แตกต่างกัน [10] สำหรับการพิจารณาตัวแปรอิสระเข้าตัวแบบ ดำเนินการดังนี้ ครั้งที่ 1 มีค่าคงที่ และ X_1 (p = 2) ครั้งที่ 2 มีค่าคงที่, X_1 และ X_2 (p = 3) ครั้งที่ 3 มีค่าคงที่, X_1, X_2 และ X_3 (p = 4) ครั้งที่ 4 มีค่าคงที่, X_1, X_2, X_3 และ X_4 (p = 5) ครั้งที่ 5 มีค่าคงที่, X_1, X_2, X_3, X_4 และ X_5 (p = 6) และครั้งที่ 6 มีค่าคงที่, X_1, X_2, X_3, X_4, X_5 และ X_6 (p = 7)

4. แต่ละครั้งที่นำตัวแปรอิสระเข้าตัวแบบ จะคำนวณค่าของเกณฑ์ NIC ในสมการที่ (6) เปรียบเทียบกับเกณฑ์การคัดเลือกตัวแบบการถดถอยทั้ง 3 เกณฑ์ ได้แก่ เกณฑ์ $KICc_C$ ในสมการที่ (5), เกณฑ์ $KICc_{SB}$ ในสมการที่ (7) และเกณฑ์ $KICc_{HM}$ ในสมการที่ (8) ดังนี้

$$KICc_{SB} = \log\left(\hat{\sigma}^2\right) + \frac{(p+1)(3n-p-2)}{n(n-p-2)} + \frac{p}{n(n-p)} \quad (7)$$

$$KICc_{HM} = \log\left(\hat{\sigma}^2\right) + \frac{(p+1)(3n-p-2)}{n(n-p-2)} \quad (8)$$

5. สรุปผลว่าทั้งหมด 6 ครั้งที่นำตัวแปรอิสระเข้าตัวแบบครั้งใดมีค่าของเกณฑ์การคัดเลือกตัวแบบการถดถอยต่ำที่สุด จะได้ครั้งนั้นเป็นตัวแบบที่คัดเลือกไว้หรือตัวแบบโดยประมาณ เช่น เมื่อพิจารณาเกณฑ์ NIC ทั้งหมด 6 ครั้ง พบว่าครั้งที่ 5 ให้ค่า NIC ต่ำที่สุด จะได้ตัวแบบที่คัดเลือกประกอบด้วยค่าคงที่, X_1, X_2, X_3, X_4 และ X_5 (p = 6)

6. คำนวณค่าร้อยละของการคัดเลือกตัวแบบได้ถูกต้อง ร้อยละของการมีจำนวนตัวแปรในสมการมากหรือน้อยเกินไป และระยะห่างความคลาดเคลื่อนกำลังสอง (Observed Distance L_2 หรือ Squared Error Distance) ระหว่างตัวแบบที่แท้จริงกับตัวแบบโดยประมาณหาด้วยขนาดตัวอย่าง [3] โดยการคัดเลือกตัวแบบได้ถูกต้องและการมี

จำนวนตัวแปรในสมการมากหรือน้อยเกินไป อธิบายได้ดังนี้ สมมติสร้างตัวแบบที่แท้จริงตัวแบบที่ 1 มีตัวแปรอิสระจำนวน 4 ตัว การคัดเลือกตัวแบบได้ถูกต้อง หมายถึง ได้ตัวแบบที่มีตัวแปรอิสระจำนวน 4 ตัว ขณะที่การมีจำนวนตัวแปรในสมการมากหรือน้อยเกินไป หมายถึง ได้ตัวแบบที่มีตัวแปรอิสระมากกว่าหรือน้อยกว่า 4 ตัว สำหรับระยะห่างความคลาดเคลื่อนกำลังสองระหว่างตัวแบบที่แท้จริงกับตัวแบบโดยประมาณหารด้วยขนาดตัวอย่าง มีสูตรดังนี้

$$L_2(p) = \frac{1}{n}(\beta_0 - \beta)' \mathbf{X}'\mathbf{X}(\beta_0 - \beta) \quad (9)$$

จากนั้นคำนวณค่าเกณฑ์ประสิทธิภาพสังเกต L_2 ดังนี้

$$\text{ประสิทธิภาพสังเกต } L_2 = \frac{\min_{1\le p\le P} L_2(p)}{L_2(p_s)} \quad (10)$$

โดยที่ P แทนชั้น (Class) ของตัวแบบโดยประมาณที่เป็นไปได้ทั้งหมด p แทนจำนวนตัวประมาณในตัวแบบโดยประมาณ และ p_s แทนจำนวนตัวประมาณในตัวแบบที่คัดเลือกไว้ โดยถ้าตัวแบบที่คัดเลือกไว้ใกล้เคียงกับตัวแบบที่แท้จริง ประสิทธิภาพสังเกต L_2 จะมีค่าสูงขึ้น ดังนั้นเกณฑ์การเลือกตัวแบบที่ดีที่สุดจะคัดเลือกตัวแบบที่ให้ประสิทธิภาพสังเกต L_2 สูงแม้ในกรณีตัวอย่างขนาดเล็กหรือตัวแบบระบุได้ยาก

7. ดำเนินการข้อ 3 – 6 ซ้ำในแต่ละชุดข้อมูลที่จำลองขึ้น จำนวน 1,000 ชุด เพื่อสรุปผลว่าเกณฑ์ใดที่มีโอกาสคัดเลือกตัวแบบได้ถูกต้องมากที่สุด นั่นคือ มีโอกาสได้ตัวแบบที่คัดเลือกไว้ตรงกับตัวแบบที่สร้างขึ้นมากที่สุดและมีประสิทธิภาพสังเกต L_2 สูงที่สุด

3. ผลการวิจัย

จากผลการจำลองข้อมูล จำนวน 1,000 ชุด ได้ผลการวิจัยแสดงดังตารางที่ 1 – 4 ซึ่งพบว่า เมื่อตัวแบบระบุได้ยากถึงยากมากดังตัวแบบที่ 1 และ 2 และความแปรปรวนของความคลาดเคลื่อนมีค่าเพิ่มขึ้น เกณฑ์การคัดเลือกตัวแบบทุกเกณฑ์จะสามารถคัดเลือกตัวแบบได้ถูกต้องน้อย ดังนั้นผู้วิจัยจึงใช้เกณฑ์ประสิทธิภาพสังเกต L_2 ในสมการที่ (10) เป็นเกณฑ์การคัดเลือกตัวแบบอีกเกณฑ์หนึ่ง ได้ผลแสดงดังตารางที่ 5 - 8 และได้ทำการสรุปเกณฑ์การคัดเลือกตัวแบบที่เหมาะสมภายใต้สถานการณ์ต่าง ๆ ไว้ดังตารางที่ 9 โดยพบว่า ถ้าเปรียบเทียบความสามารถในการคัดเลือกตัวแบบได้ถูกต้อง ส่วนใหญ่เกณฑ์ $KICc_{HM}$ มีประสิทธิภาพสูงที่สุด อย่างไรก็ตาม เกณฑ์นี้สามารถคัดเลือกตัวแบบได้ถูกต้องน้อย ดังนั้นจึงพิจารณาเกณฑ์ประสิทธิภาพสังเกต L_2 ประกอบด้วย โดยเกณฑ์นี้แนะนำว่าเกณฑ์ NIC คือเกณฑ์ที่ดีที่สุด เนื่องจากมีค่าเฉลี่ยประสิทธิภาพสังเกต L_2 สูงที่สุดและมีส่วนเบี่ยงเบนมาตรฐานค่าประสิทธิภาพสังเกต L_2 ต่ำที่สุด นอกจากนี้ยังพบว่า เมื่อขนาดตัวอย่างเพิ่มขึ้นหรือตัวแบบสามารถระบุได้ง่ายขึ้น ทุกเกณฑ์สามารถคัดเลือกตัวแบบได้ถูกต้องเพิ่มขึ้น รวมถึงค่าประสิทธิภาพสังเกต L_2 จะเพิ่มขึ้นด้วย และความแปรปรวนของความคลาดเคลื่อนที่เพิ่มขึ้น ส่งผลให้ความถูกต้องของการคัดเลือกตัวแบบลดลง โดยเฉพาะเมื่อตัวแบบระบุยากถึงยากมากดังตัวแบบที่ 1 และ 2 ความแปรปรวนของความคลาดเคลื่อนที่เพิ่มขึ้น จะทำให้ร้อยละของการคัดเลือกตัวแบบได้ถูกต้องลดลงอย่างมาก และมีโอกาสมีจำนวนตัวแปรน้อยเกินไปมากกว่าการมีจำนวนตัวแปรมากเกินไป สำหรับกรณีการแจกแจงของตัวแปรอิสระเปลี่ยนจากการแจกแจงปรกติมาตรฐานเป็นการแจกแจงเอกรูป ร้อยละของการคัดเลือกตัวแบบได้ถูกต้องจะลดลง โดยจะมีโอกาสมีจำนวนตัวแปรน้อยเกินไปมากกว่าการมีจำนวนตัวแปรมากเกินไป

ตารางที่ 1 ร้อยละของการคัดเลือกตัวแบบ เมื่อ n = 15 และตัวแปรอิสระมีการแจกแจงปรกติมาตรฐาน

σ_0^2	ตัวแบบที่	จำนวนตัวแปร	$KICc_C$	$KICc_{SB}$	$KICc_{HM}$	NIC
0.25	1	น้อยเกินไป	90.2	88.5	87.6	95.0
		ถูกต้อง	9.3	11.0	**11.9**	4.7
		มากเกินไป	0.5	0.5	0.5	0.3
0.25	2	น้อยเกินไป	27.8	26.8	26.1	37.0
		ถูกต้อง	68.0	68.3	**68.5**	60.7
		มากเกินไป	4.2	4.9	5.4	2.3
0.25	3	น้อยเกินไป	0.0	0.0	0.0	0.0
		ถูกต้อง	94.4	93.8	93.4	**97.0**
		มากเกินไป	5.6	6.2	6.6	3.0
0.25	4	น้อยเกินไป	0.0	0.0	0.0	0.0
		ถูกต้อง	97.8	96.7	96.5	**98.5**
		มากเกินไป	2.2	3.3	3.5	1.5
1	1	น้อยเกินไป	98.9	98.7	98.6	99.5
		ถูกต้อง	1.0	1.1	1.1	0.5
		มากเกินไป	0.1	0.2	0.3	0.0
1	2	น้อยเกินไป	74.0	72.9	72.2	81.9
		ถูกต้อง	23.8	24.7	**25.2**	17.4
		มากเกินไป	2.2	2.4	2.6	0.7
1	3	น้อยเกินไป	0.1	0.1	0.1	0.3
		ถูกต้อง	95.8	94.8	94.4	**96.8**
		มากเกินไป	4.1	5.1	5.5	2.9
1	4	น้อยเกินไป	2.1	1.9	1.8	3.5
		ถูกต้อง	**96.4**	95.9	95.8	95.9
		มากเกินไป	1.5	2.2	2.4	0.6
9	1	น้อยเกินไป	99.7	99.6	99.3	100.0
		ถูกต้อง	0.2	0.3	**0.5**	0.0
		มากเกินไป	0.1	0.1	0.2	0.0
9	2	น้อยเกินไป	91.8	91.1	90.5	96.0
		ถูกต้อง	6.8	7.1	**7.5**	3.3
		มากเกินไป	1.4	1.8	2.0	0.7
9	3	น้อยเกินไป	42.7	41.3	40.8	52.6
		ถูกต้อง	54.8	55.7	**56.0**	46.1
		มากเกินไป	2.5	3.0	3.2	1.3
9	4	น้อยเกินไป	76.1	73.0	71.8	87.2
		ถูกต้อง	23.0	25.9	**27.1**	12.4
		มากเกินไป	0.9	1.1	1.1	0.4

หมายเหตุ : ตัวหนา หมายถึง เกณฑ์ที่มีร้อยละของการคัดเลือกตัวแบบได้ถูกต้องสูงที่สุด

ตารางที่ 2 ร้อยละของการคัดเลือกตัวแบบ เมื่อ n = 15 และตัวแปรอิสระมีการแจกแจงเอกรูป

σ_0^2	ตัวแบบที่	จำนวนตัวแปร	$KICc_C$	$KICc_{SB}$	$KICc_{HM}$	NIC
0.25	1	น้อยเกินไป	99.5	99.2	99.2	99.9
		ถูกต้อง	0.4	**0.6**	**0.6**	0.1
		มากเกินไป	0.1	0.2	0.2	0.0
0.25	2	น้อยเกินไป	86.1	85.3	85.2	92.0
		ถูกต้อง	12.3	13.1	**13.2**	7.2
		มากเกินไป	1.6	1.6	1.6	0.8
0.25	3	น้อยเกินไป	4.9	4.4	4.2	8.5
		ถูกต้อง	**89.6**	89.4	89.3	88.0
		มากเกินไป	5.5	6.2	6.5	3.5
0.25	4	น้อยเกินไป	22.3	18.4	17.6	37.2
		ถูกต้อง	75.7	79.4	**79.9**	61.6
		มากเกินไป	2.0	2.2	2.5	1.2
1	1	น้อยเกินไป	99.9	99.8	99.7	99.9
		ถูกต้อง	0.1	0.2	**0.3**	0.1
		มากเกินไป	0.0	0.0	0.0	0.0
1	2	น้อยเกินไป	90.4	90.1	89.5	93.8
		ถูกต้อง	7.2	7.5	**8.0**	5.3
		มากเกินไป	2.4	2.4	2.5	0.9
1	3	น้อยเกินไป	51.4	49.2	48.3	61.4
		ถูกต้อง	45.0	46.5	**47.0**	36.8
		มากเกินไป	3.6	4.3	4.7	1.8
1	4	น้อยเกินไป	85.0	82.7	82.0	92.4
		ถูกต้อง	14.6	16.6	**17.3**	7.5
		มากเกินไป	0.4	0.7	0.7	0.1
9	1	น้อยเกินไป	99.6	99.4	99.4	99.9
		ถูกต้อง	0.4	**0.5**	**0.5**	0.1
		มากเกินไป	0.0	0.1	0.1	0.0
9	2	น้อยเกินไป	93.3	92.6	92.3	96.0
		ถูกต้อง	5.3	5.7	**5.8**	3.6
		มากเกินไป	1.4	1.7	1.9	0.4
9	3	น้อยเกินไป	88.6	88.2	87.4	92.6
		ถูกต้อง	9.2	9.4	**10.1**	6.4
		มากเกินไป	2.2	2.4	2.5	1.0
9	4	น้อยเกินไป	99.1	98.9	98.8	99.6
		ถูกต้อง	0.7	**0.8**	**0.8**	0.4
		มากเกินไป	0.2	0.3	0.4	0.0

หมายเหตุ : ตัวหนา หมายถึง เกณฑ์ที่มีร้อยละของการคัดเลือกตัวแบบได้ถูกต้องสูงที่สุด

ตารางที่ 3 ร้อยละของการคัดเลือกตัวแบบ เมื่อ n = 30 และตัวแปรอิสระมีการแจกแจงปรกติมาตรฐาน

σ_0^2	ตัวแบบที่	จำนวนตัวแปร	$KICc_C$	$KICc_{SB}$	$KICc_{HM}$	NIC
0.25	1	น้อยเกินไป	49.0	46.9	46.6	62.0
		ถูกต้อง	46.6	47.9	**48.2**	35.4
		มากเกินไป	4.4	5.2	5.2	2.6
0.25	2	น้อยเกินไป	1.9	1.9	1.8	3.5
		ถูกต้อง	89.4	88.8	88.4	**91.9**
		มากเกินไป	8.7	9.3	9.8	4.6
0.25	3	น้อยเกินไป	0.0	0.0	0.0	0.0
		ถูกต้อง	91.4	91.0	90.9	**95.9**
		มากเกินไป	8.6	9.0	9.1	4.1
0.25	4	น้อยเกินไป	0.0	0.0	0.0	0.0
		ถูกต้อง	94.3	93.5	93.5	**96.7**
		มากเกินไป	5.7	6.5	6.5	3.3
1	1	น้อยเกินไป	89.4	88.1	87.6	94.5
		ถูกต้อง	8.9	9.9	**10.3**	5.2
		มากเกินไป	1.7	2.0	2.1	0.3
1	2	น้อยเกินไป	41.4	39.7	39.0	51.9
		ถูกต้อง	52.0	52.3	**52.9**	44.7
		มากเกินไป	6.6	8.0	8.1	3.4
1	3	น้อยเกินไป	0.0	0.0	0.0	0.0
		ถูกต้อง	90.0	89.3	88.9	**95.1**
		มากเกินไป	10.0	10.7	11.1	4.9
1	4	น้อยเกินไป	0.0	0.0	0.0	0.0
		ถูกต้อง	93.4	92.7	92.4	**96.0**
		มากเกินไป	6.6	7.3	7.6	4.0
9	1	น้อยเกินไป	98.2	97.8	97.4	99.3
		ถูกต้อง	1.3	1.5	**1.8**	0.4
		มากเกินไป	0.5	0.7	0.8	0.3
9	2	น้อยเกินไป	84.1	83.1	82.8	90.7
		ถูกต้อง	12.0	12.4	**12.5**	7.9
		มากเกินไป	3.9	4.5	4.7	1.4
9	3	น้อยเกินไป	5.8	5.3	5.2	8.9
		ถูกต้อง	87.3	87.4	87.3	**88.3**
		มากเกินไป	6.9	7.3	7.5	2.8
9	4	น้อยเกินไป	14.8	13.6	13.1	22.9
		ถูกต้อง	77.9	78.8	**79.3**	73.7
		มากเกินไป	7.3	7.6	7.6	3.4

หมายเหตุ : ตัวหนา หมายถึง เกณฑ์ที่มีร้อยละของการคัดเลือกตัวแบบได้ถูกต้องสูงที่สุด

ตารางที่ 4 ร้อยละของการคัดเลือกตัวแบบ เมื่อ n = 30 และตัวแปรอิสระมีการแจกแจงเอกรูป

σ_0^2	ตัวแบบที่	จำนวนตัวแปร	$KICc_C$	$KICc_{SB}$	$KICc_{HM}$	NIC
0.25	1	น้อยเกินไป	96.5	96.3	96.2	98.8
		ถูกต้อง	2.8	3.0	3.1	1.1
		มากเกินไป	0.7	0.7	0.7	0.1
0.25	2	น้อยเกินไป	70.2	69.4	68.7	79.6
		ถูกต้อง	25.4	25.9	**26.0**	18.9
		มากเกินไป	4.4	4.7	5.3	1.5
0.25	3	น้อยเกินไป	0.0	0.0	0.0	0.0
		ถูกต้อง	93.2	92.6	92.3	**96.5**
		มากเกินไป	6.8	7.4	7.7	3.5
0.25	4	น้อยเกินไป	0.1	0.1	0.1	0.3
		ถูกต้อง	93.2	92.4	92.1	**95.7**
		มากเกินไป	6.7	7.5	7.8	4.0
1	1	น้อยเกินไป	98.6	98.3	98.3	99.8
		ถูกต้อง	0.9	**1.0**	**1.0**	0.2
		มากเกินไป	0.5	0.7	0.7	0.0
1	2	น้อยเกินไป	86.8	86.1	85.7	92.3
		ถูกต้อง	9.3	9.6	**9.9**	6.1
		มากเกินไป	3.9	4.3	4.4	1.6
1	3	น้อยเกินไป	12.7	11.7	11.4	20.3
		ถูกต้อง	79.9	**80.4**	**80.4**	76.2
		มากเกินไป	7.4	7.9	8.2	3.5
1	4	น้อยเกินไป	20.4	18.9	18.5	31.7
		ถูกต้อง	72.4	73.1	**73.3**	**64.4**
		มากเกินไป	7.2	8.0	8.2	3.9
9	1	น้อยเกินไป	99.5	99.0	98.8	99.9
		ถูกต้อง	0.2	0.4	**0.6**	0.0
		มากเกินไป	0.3	0.6	0.6	0.1
9	2	น้อยเกินไป	89.9	88.8	88.4	94.3
		ถูกต้อง	7.8	8.3	**8.5**	5.0
		มากเกินไป	2.3	2.9	3.1	0.7
9	3	น้อยเกินไป	78.3	77.0	76.6	86.9
		ถูกต้อง	18.2	18.7	**19.0**	12.3
		มากเกินไป	3.5	4.3	4.4	0.8
9	4	น้อยเกินไป	94.9	93.9	93.5	98.1
		ถูกต้อง	4.5	5.3	**5.7**	1.6
		มากเกินไป	0.6	0.8	0.8	0.3

หมายเหตุ : ตัวหนา หมายถึง เกณฑ์ที่มีร้อยละของการคัดเลือกตัวแบบได้ถูกต้องสูงที่สุด

ตารางที่ 5 ประสิทธิภาพสังเกต L_2 เมื่อ n = 15 และตัวแปรอิสระมีการแจกแจงปรกติมาตรฐาน

σ_0^2	ตัวแบบที่	ค่าสถิติ	$KICc_C$	$KICc_{SB}$	$KICc_{HM}$	NIC
0.25	1	$\bar{X}$	0.5683	0.5845	**0.5904**	0.5020
		S.D.	0.2767	0.2770	0.2762	**0.2758**
0.25	2	$\bar{X}$	0.7948	0.7985	**0.8006**	0.7359
		S.D.	0.3225	0.3189	**0.3171**	0.3537
0.25	3	$\bar{X}$	0.9616	0.9573	0.9547	**0.9793**
		S.D.	0.1622	0.1707	0.1751	**0.1201**
0.25	4	$\bar{X}$	0.9877	0.9823	0.9817	**0.9923**
		S.D.	0.0858	0.1006	0.1017	**0.0647**
1	1	$\bar{X}$	0.6567	0.6571	**0.6578**	0.6423
		S.D.	0.2430	0.2432	0.2427	**0.2423**
1	2	$\bar{X}$	0.6604	0.6625	**0.6636**	0.6382
		S.D.	0.2896	0.2907	0.2914	**0.2842**
1	3	$\bar{X}$	0.9720	0.9663	0.9642	**0.9776**
		S.D.	0.1406	0.1517	0.1557	**0.1296**
1	4	$\bar{X}$	**0.9752**	0.9731	0.9728	0.9683
		S.D.	**0.1334**	0.1364	0.1358	0.1560
9	1	$\bar{X}$	0.8956	0.8941	0.8903	**0.9202**
		S.D.	0.2041	0.2052	0.2110	**0.1723**
9	2	$\bar{X}$	0.8947	0.8901	0.8860	**0.9209**
		S.D.	0.2108	0.2166	0.2219	**0.1744**
9	3	$\bar{X}$	0.7287	0.7370	**0.7394**	0.6688
		S.D.	0.3319	0.3270	**0.3259**	0.3440
9	4	$\bar{X}$	0.5360	0.5602	**0.5683**	0.4416
		S.D.	0.3201	0.3237	0.3256	**0.2919**

หมายเหตุ : ตัวหนา หมายถึง ประสิทธิภาพสังเกต L_2 มีค่าเฉลี่ยสูงที่สุดและมีส่วนเบี่ยงเบนมาตรฐานต่ำที่สุด

ตารางที่ 6 ประสิทธิภาพสังเกต L_2 เมื่อ n = 15 และตัวแปรอิสระมีการแจกแจงเอกรูป

σ_0^2	ตัวแบบที่	ค่าสถิติ	$KICc_C$	$KICc_{SB}$	$KICc_{HM}$	NIC
0.25	1	$\bar{X}$	0.7954	0.7913	0.7919	**0.8105**
		S.D.	0.2200	0.2234	0.2234	**0.2118**
0.25	2	$\bar{X}$	0.7788	0.7764	0.7768	**0.7965**
		S.D.	0.2424	0.2430	0.2431	**0.2372**
0.25	3	$\bar{X}$	**0.9284**	0.9271	0.9267	0.9121
		S.D.	**0.2187**	0.2204	0.2209	0.2447
0.25	4	$\bar{X}$	0.8289	0.8597	**0.8649**	0.7056
		S.D.	0.3140	0.2869	**0.2812**	0.3864
1	1	$\bar{X}$	0.9230	0.9168	0.9137	**0.9350**
		S.D.	0.1753	0.1838	0.1892	**0.1521**
1	2	$\bar{X}$	0.8999	0.8982	0.8947	**0.9213**
		S.D.	0.2179	0.2193	0.2228	**0.1892**
1	3	$\bar{X}$	0.6886	0.6985	**0.7016**	0.6327
		S.D.	0.3301	0.3287	**0.3286**	0.3361
1	4	$\bar{X}$	0.5073	0.5255	**0.5321**	0.4475
		S.D.	0.2830	0.2873	0.2890	**0.2623**
9	1	$\bar{X}$	0.9549	0.9517	0.9480	**0.9739**
		S.D.	0.1779	0.1834	0.1895	**0.1347**
9	2	$\bar{X}$	0.9490	0.9435	0.9406	**0.9678**
		S.D.	0.1883	0.1983	0.2044	**0.1504**
9	3	$\bar{X}$	0.8121	0.8098	0.8075	**0.8350**
		S.D.	0.2449	0.2467	0.2481	**0.2287**
9	4	$\bar{X}$	0.8198	0.8162	0.8144	**0.8366**
		S.D.	0.2132	0.2153	0.2162	**0.2020**

หมายเหตุ : ตัวหนา หมายถึง ประสิทธิภาพสังเกต L_2 มีค่าเฉลี่ยสูงที่สุดและมีส่วนเบี่ยงเบนมาตรฐานต่ำที่สุด

ตารางที่ 7 ประสิทธิภาพสังเกต L_2 เมื่อ n = 30 และตัวแปรอิสระมีการแจกแจงปรกติมาตรฐาน

σ_0^2	ตัวแบบที่	ค่าสถิติ	$KICc_C$	$KICc_{SB}$	$KICc_{HM}$	NIC
0.25	1	$\bar{X}$	0.7485	0.7562	**0.7567**	0.6700
		S.D.	0.2842	0.2821	**0.2820**	0.3032
0.25	2	$\bar{X}$	0.9288	0.9245	0.9222	**0.9430**
		S.D.	0.2161	0.2222	0.2248	**0.1985**
0.25	3	$\bar{X}$	0.9407	0.9382	0.9373	**0.9708**
		S.D.	0.2013	0.2049	0.2068	**0.1465**
0.25	4	$\bar{X}$	0.9701	0.9662	0.9662	**0.9815**
		S.D.	0.1276	0.1348	0.1348	**0.1037**
1	1	$\bar{X}$	0.6215	0.6268	**0.6284**	0.5741
		S.D.	0.2454	0.2457	**0.2446**	0.2456
1	2	$\bar{X}$	0.7248	0.7269	**0.7307**	0.6731
		S.D.	0.3294	0.3285	**0.3277**	0.3419
1	3	$\bar{X}$	0.9324	0.9272	0.9247	**0.9676**
		S.D.	0.2112	0.2194	0.2224	**0.1477**
1	4	$\bar{X}$	0.9643	0.9603	0.9586	**0.9779**
		S.D.	0.1407	0.1481	0.1517	**0.1138**
9	1	$\bar{X}$	0.8183	0.8155	0.8103	**0.8375**
		S.D.	0.2215	0.2238	0.2269	**0.2082**
9	2	$\bar{X}$	0.7997	0.7937	0.7915	**0.8281**
		S.D.	0.2479	0.2526	0.2539	**0.2289**
9	3	$\bar{X}$	0.9180	0.9182	0.9174	**0.9201**
		S.D.	**0.2238**	0.2244	0.2255	0.2254
9	4	$\bar{X}$	0.8726	0.8793	**0.8832**	0.8321
		S.D.	0.2537	0.2469	**0.2428**	0.2964

หมายเหตุ : ตัวหนา หมายถึง ประสิทธิภาพสังเกต L_2 มีค่าเฉลี่ยสูงที่สุดและมีส่วนเบี่ยงเบนมาตรฐานต่ำที่สุด

ตารางที่ 8 ประสิทธิภาพสังเกต L_2 เมื่อ n = 30 และตัวแปรอิสระมีการแจกแจงเอกรูป

σ_0^2	ตัวแบบที่	ค่าสถิติ	$KICc_C$	$KICc_{SB}$	$KICc_{HM}$	NIC
0.25	1	$\bar{X}$	0.6811	0.6812	**0.6816**	0.6810
		S.D.	0.2298	0.2298	0.2291	**0.2240**
0.25	2	$\bar{X}$	0.6889	**0.6899**	0.6884	0.6778
		S.D.	0.2788	0.2803	0.2813	**0.2671**
0.25	3	$\bar{X}$	0.9532	0.9488	0.9471	**0.9748**
		S.D.	0.1800	0.1882	0.1906	**0.1373**
0.25	4	$\bar{X}$	0.9630	0.9591	0.9570	**0.9750**
		S.D.	0.1452	0.1520	0.1565	**0.1226**
1	1	$\bar{X}$	0.8234	0.8202	0.8191	**0.8591**
		S.D.	0.2396	0.2425	0.2432	**0.2101**
1	2	$\bar{X}$	0.8309	0.8264	0.8241	**0.8647**
		S.D.	0.2481	0.2514	0.2526	**0.2182**
1	3	$\bar{X}$	0.8712	0.8739	**0.8742**	0.8404
		S.D.	0.2699	0.2682	**0.2676**	0.2984
1	4	$\bar{X}$	0.8423	0.8480	**0.8511**	0.7711
		S.D.	0.2746	0.2701	**0.2665**	0.3289
9	1	$\bar{X}$	0.9361	0.9284	0.9273	**0.9561**
		S.D.	0.2021	0.2160	0.2183	**0.1685**
9	2	$\bar{X}$	0.9229	0.9143	0.9118	**0.9539**
		S.D.	0.2248	0.2370	0.2398	**0.1750**
9	3	$\bar{X}$	0.7204	0.7174	0.7173	**0.7319**
		S.D.	0.2550	0.2593	0.2599	**0.2429**
9	4	$\bar{X}$	0.6810	0.6804	0.6803	**0.6874**
		S.D.	**0.2282**	0.2290	0.2290	0.2307

หมายเหตุ : ตัวหนา หมายถึง ประสิทธิภาพสังเกต L_2 มีค่าเฉลี่ยสูงที่สุดและมีส่วนเบี่ยงเบนมาตรฐานต่ำที่สุด

ตารางที่ 9 เกณฑ์การคัดเลือกตัวแบบที่เหมาะสมภายใต้สถานการณ์ต่าง ๆ

ตัวแบบที่	n	σ_0^2	การแจกแจงของตัวแปรอิสระ							
			การแจกแจงปรกติมาตรฐาน				การแจกแจงเอกรูป			
			เกณฑ์	% ถูกต้องสูงสุด	ประสิทธิภาพสังเกต L_2		เกณฑ์	% ถูกต้องสูงสุด	ประสิทธิภาพสังเกต L_2	
					$\bar{X}$ สูงสุด	S.D. ต่ำสุด			$\bar{X}$ สูงสุด	S.D. ต่ำสุด
1	15	0.25	$KICc_{HM}$	11.9	$KICc_{HM}$	**NIC**	$KICc_{SB}$, $KICc_{HM}$	0.6	**NIC**	**NIC**
		1	$KICc_{SB}$, $KICc_{HM}$	1.1	$KICc_{HM}$	**NIC**	$KICc_{HM}$	0.3	**NIC**	**NIC**
		9	$KICc_{HM}$	0.5	**NIC**	**NIC**	$KICc_{SB}$, $KICc_{HM}$	0.5	**NIC**	**NIC**
	30	0.25	$KICc_{HM}$	48.2	$KICc_{HM}$	$KICc_{HM}$	$KICc_{HM}$	3.1	$KICc_{HM}$	**NIC**
		1	$KICc_{HM}$	10.3	$KICc_{HM}$	$KICc_{HM}$	$KICc_{SB}$, $KICc_{HM}$	1.0	**NIC**	**NIC**
		9	$KICc_{HM}$	1.8	**NIC**	**NIC**	$KICc_{HM}$	0.6	**NIC**	**NIC**
2	15	0.25	$KICc_{HM}$	68.5	$KICc_{HM}$	$KICc_{HM}$	$KICc_{HM}$	13.2	**NIC**	**NIC**
		1	$KICc_{HM}$	25.2	$KICc_{HM}$	**NIC**	$KICc_{HM}$	8.0	**NIC**	**NIC**
		9	$KICc_{HM}$	7.5	**NIC**	**NIC**	$KICc_{HM}$	5.8	**NIC**	**NIC**
	30	0.25	**NIC**	91.9	**NIC**	**NIC**	$KICc_{HM}$	26.0	$KICc_{SB}$	**NIC**
		1	$KICc_{HM}$	52.9	$KICc_{HM}$	$KICc_{HM}$	$KICc_{HM}$	9.9	**NIC**	**NIC**
		9	$KICc_{HM}$	12.5	**NIC**	**NIC**	$KICc_{HM}$	8.5	**NIC**	**NIC**
3	15	0.25	**NIC**	97.0	**NIC**	**NIC**	$KICc_{C}$	89.6	$KICc_{C}$	$KICc_{C}$
		1	**NIC**	96.8	**NIC**	**NIC**	$KICc_{HM}$	47.0	$KICc_{HM}$	$KICc_{HM}$
		9	$KICc_{HM}$	56.0	$KICc_{HM}$	$KICc_{HM}$	$KICc_{HM}$	10.1	**NIC**	**NIC**
	30	0.25	**NIC**	95.9	**NIC**	**NIC**	**NIC**	96.5	**NIC**	**NIC**
		1	**NIC**	95.1	**NIC**	**NIC**	$KICc_{SB}$, $KICc_{HM}$	80.4	$KICc_{HM}$	$KICc_{HM}$
		9	**NIC**	88.3	**NIC**	$KICc_{C}$	$KICc_{HM}$	19.0	**NIC**	**NIC**
4	15	0.25	**NIC**	98.5	**NIC**	**NIC**	$KICc_{HM}$	79.9	$KICc_{HM}$	$KICc_{HM}$
		1	$KICc_{C}$	96.4	$KICc_{C}$	$KICc_{C}$	$KICc_{HM}$	17.3	$KICc_{HM}$	**NIC**
		9	$KICc_{HM}$	27.1	$KICc_{HM}$	**NIC**	$KICc_{SB}$, $KICc_{HM}$	0.8	**NIC**	**NIC**
	30	0.25	**NIC**	96.7	**NIC**	**NIC**	**NIC**	95.7	**NIC**	**NIC**
		1	**NIC**	96.0	**NIC**	**NIC**	$KICc_{HM}$	73.3	$KICc_{HM}$	$KICc_{HM}$
		9	$KICc_{HM}$	79.3	$KICc_{HM}$	$KICc_{HM}$	$KICc_{HM}$	5.7	**NIC**	$KICc_{C}$

4. วิจารณ์ผลการวิจัย

จากผลการศึกษาข้างต้นผู้วิจัยขอวิจารณ์ผลดังนี้ เกณฑ์การคัดเลือกตัวแบบที่เหมาะสมภายใต้สถานการณ์ต่าง ๆ มีความหลากหลาย แต่ส่วนใหญ่เกณฑ์ NIC มีความเหมาะสมกับตัวอย่างขนาดเล็กมากกว่าเกณฑ์ $KICc_C$, $KICc_{SB}$ และ $KICc_{HM}$ สอดคล้องกับการศึกษาของ Al-Subaihi [11] ที่พบว่าไม่มีเกณฑ์ใดเลยที่มีประสิทธิภาพภายใต้ทุกสถานการณ์ที่พิจารณา ซึ่งแต่ละเกณฑ์จะมีความเหมาะสมแตกต่างกันไปตามข้อมูลที่กำลังศึกษา และสอดคล้องกับการศึกษาของ Whittaker และ Furlow [12] ที่พบว่า ไม่ควรระบุว่าเกณฑ์การคัดเลือกตัวแบบเกณฑ์ใดเหมาะสมที่สุดในทุกสถานการณ์ที่เกิดขึ้น แต่ประสิทธิภาพจะขึ้นอยู่กับเงื่อนไขของข้อมูลที่ศึกษาด้วย การศึกษาครั้งนี้ยังพบว่า กรณีตัวอย่างมีขนาดเล็กร้อยละของการคัดเลือกตัวแบบได้ถูกต้องจะต่ำ และเมื่อตัวอย่างมีขนาดใหญ่ขึ้นหรือตัวแบบสามารถระบุได้ง่ายขึ้น ทุกเกณฑ์สามารถคัดเลือกตัวแบบได้ถูกต้องเพิ่มขึ้น รวมถึง

ค่าประสิทธิภาพสังเกต L_2 จะเพิ่มขึ้นด้วย สอดคล้องกับการศึกษาของ Al-Subaihi [11] ที่พบว่า ปัจจัยต่าง ๆ เช่น ขนาดตัวอย่าง จำนวนตัวแปรตาม จำนวนตัวแปรอิสระ และความสัมพันธ์ระหว่างตัวแปรตาม ล้วนมีบทบาทในการตัดสินใจว่าควรใช้เกณฑ์ใด สอดคล้องกับการศึกษาของ Acquah [13] ที่พบว่า ประสิทธิภาพของเกณฑ์การคัดเลือกตัวแบบแต่ละเกณฑ์จะขึ้นอยู่กับขนาดของตัวอย่าง ระดับของความไม่สมมาตร และปริมาณของความคลาดเคลื่อนในตัวแบบ และสอดคล้องกับการศึกษาของ Sangthong [14] ที่พบว่า ประสิทธิภาพของเกณฑ์การคัดเลือกตัวแบบค่อนข้างต่ำเมื่อจำนวนกลุ่มมีขนาดเล็ก และประสิทธิภาพการคัดเลือกตัวแบบจะสูงขึ้นเมื่อจำนวนกลุ่มมากขึ้น สำหรับกรณีความคลาดเคลื่อนมีความแปรปรวนเพิ่มขึ้น ทำให้ทุกเกณฑ์การคัดเลือกตัวแบบมีประสิทธิภาพลดลง โดยเฉพาะเมื่อตัวแบบระบุยากถึงยากมากดังตัวแบบที่ 1 และ 2 ความแปรปรวนของความคลาดเคลื่อนที่เพิ่มขึ้น จะทำให้ร้อยละของการคัดเลือกตัวแบบได้ถูกต้องลดลงอย่างมาก และมีโอกาสที่จะคัดเลือกตัวแบบที่มีจำนวนตัวแปรน้อยเกินไปมากกว่าการมีจำนวนตัวแปรมากเกินไป ดังนั้นควรตรวจสอบความแปรปรวนของความคลาดเคลื่อนหลังจากการสร้างตัวแบบทุกครั้ง เนื่องจากมีผลกระทบต่อความถูกต้องของเกณฑ์การคัดเลือกตัวแบบ ซึ่งกรณีความแปรปรวนของความคลาดเคลื่อนเพิ่มขึ้นแล้วทำให้ร้อยละของการคัดเลือกตัวแบบได้ถูกต้องลดลงนั้นสอดคล้องกับการศึกษาของ Acquah [13] ที่พบว่า ประสิทธิภาพของเกณฑ์การคัดเลือกตัวแบบแต่ละเกณฑ์จะขึ้นอยู่กับหลาย ๆ ปัจจัย เช่น ปริมาณของความคลาดเคลื่อนในตัวแบบ และสอดคล้องกับการศึกษาของ Keerativibool and Siripanich [10] ที่พบว่า ประสิทธิภาพของทุกเกณฑ์การคัดเลือกตัวแบบจะลดลงเมื่อความแปรปรวนของความคลาดเคลื่อนมีค่าเพิ่มขึ้น

5. สรุปผลการวิจัย

การศึกษาครั้งนี้ได้เสนอเกณฑ์ใหม่สำหรับการคัดเลือกตัวแบบการถดถอย กรณีตัวอย่างขนาดเล็ก โดยเรียกชื่อเกณฑ์ใหม่นี้ว่า New Information Criterion (NIC) มีสูตรดังนี้

$$NIC = \log\left(s^2\right) + \log\left(\frac{n}{n-p}\right) + \frac{\left[(n-p)(2p+3)-2\right]}{(n-p-2)(n-p)}$$

ผลการจำลองข้อมูลจำนวน 1,000 ชุด ภายใต้สถานการณ์ต่าง ๆ ได้แก่ ความแตกต่างของขนาดตัวอย่าง จำนวนตัวแปรอิสระในสมการ สัมประสิทธิ์การถดถอย ความแปรปรวนของความคลาดเคลื่อน และการแจกแจงของตัวแปรอิสระ แล้วเปรียบเทียบประสิทธิภาพของเกณฑ์ NIC เทียบกับเกณฑ์ $KICc_C$, $KICc_{SB}$ และ $KICc_{HM}$ พบว่า ส่วนใหญ่เกณฑ์ $KICc_{HM}$ สามารถคัดเลือกตัวแบบได้ถูกต้องดีกว่าเกณฑ์อื่น ๆ อย่างไรก็ตาม เกณฑ์นี้สามารถคัดเลือกตัวแบบได้ถูกต้องน้อย ดังนั้นผู้วิจัยจึงใช้เกณฑ์ประสิทธิภาพสังเกต L_2 ประกอบการพิจารณา โดยพบว่าเกณฑ์ NIC มีประสิทธิภาพสูงที่สุด เนื่องจากมีค่าเฉลี่ยประสิทธิภาพสังเกต L_2 สูงที่สุดและมีส่วนเบี่ยงเบนมาตรฐานค่าประสิทธิภาพสังเกต L_2 ต่ำที่สุด

6. กิตติกรรมประกาศ

การศึกษาครั้งนี้ได้รับงบประมาณสนับสนุนจากสำนักงานสภานโยบายการอุดมศึกษา วิทยาศาสตร์ วิจัยและนวัตกรรมแห่งชาติ โดยมหาวิทยาลัยทักษิณ ประจำปีงบประมาณ พ.ศ. 2564

7. เอกสารอ้างอิง

1. Akaike, H., 1974, "A New Look at the Statistical Model Identification," *IEEE Transactions on Automatic Control,* 19 (6), pp. 716-723.

2. Hurvich, C.M. and Tsai, C.L., 1989, "Regression and Time Series Model Selection in Small Samples," *Biometrika,* 76 (2), pp. 297-307.

3. McQuarrie, A.D.R., Shumway, R.H. and Tsai, C.L., 1997, "The Model Selection Criterion AICu," *Statistics and Probability Letters,* 34 (3), pp. 285-292.

4. Cavanaugh, J.E., 1999, "A Large-Sample Model Selection Criterion Based on Kullback's Symmetric Divergence," *Statistics and Probability Letters,* 42 (4), pp. 333-343.

5. Cavanaugh, J.E., 2004, "Criteria for Linear Model Selection Based on Kullback's Symmetric Divergence," *Australian and New Zealand Journal of Statistics,* 46 (2), pp. 257-274.

6. Seghouane, A.K. and Bekara, M., 2004, "A Small Sample Model Selection Criterion Based on Kullback's Symmetric Divergence," *IEEE Transactions on Signal Processing,* 52 (12), pp. 3314-3323.

7. Hafidi, B. and Mkhadri, A., 2006, "A Corrected Akaike Criterion Based on Kullback's Symmetric Divergence: Applications in Time Series, Multiple and Multivariate Regression," *Computational Statistics and Data Analysis,* 50 (6), pp. 1524-1550.

8. Montgomery, D.C., Peck, E.A. and Vining, G.G., 2006, Introduction to Linear Regression Analysis, 5th ed., Wiley, New York. p. 67.

9. Keerativibool, W., 2014, "Unifying the Derivations of Kullback Information Criterion and Corrected Versions," *Thailand Statistician,* 12 (1), pp. 37-53.

10. Keerativibool, W. and Siripanich, P., 2017, "Comparison of the Model Selection Criteria for Multiple Regression Based on Kullback-Leibler's Information," *Chiang Mai Journal of Science,* 44 (2), pp. 699-714.

11. Al-Subaihi, Ali A., 2007, Variable Selection in Multivariable Regression Using SAS/IML. [Online], Available: https://www.jstatsoft.org/article/view/v007i12/mv.pdf. [28 May 2016]

12. Whittaker, T.A. and Furlow, C.F., 2009, "The Comparison of Model Selection Criteria When Selecting Among Competing Hierarchical Linear Models," *Journal of Modern Applied Statistical Methods,* 8 (1), pp. 172-193.

13. Acquah, H.G., 2010, "Comparison of Akaike Information Criterion (AIC) and Bayesian Information Criterion (BIC) in Selection of an Asymmetric Price Relationship," *Journal of Development and Agricultural Economics,* 2 (1), pp. 1-6.

14. Sangthong, M., 2019, "A Study of the Effectiveness of Model Selection Criteria for Multilevel Analysis," *Burapha Science Journal,* 24 (1), pp. 156-169. (In Thai)