การคัดเลือกตัวแปรในตัวแบบการถดถอยเชิงพหุ ด้วยวิธีการค้นหาแบบต้องห้าม

A Tabu Search for Variable Selection in Multiple Linear Regression Models

กานต์ณัฐ ณ บางช้าง จิราวัลย์ จิตรถเวช และ วิชิต หล่อจีระชุณห์กุล
Kannat Na Bangchang Jirawan Jitthavech and Vichit Lorchirachoonkul
คณะสถิติประยุกต์ สถาบันบัณฑิตพัฒนบริหารศาสตร์

บทคัดย่อ

งานวิจัยนี้ได้เสนอวิธีการคัดเลือกตัวแปรในการวิเคราะห์การถดถอยตัวแบบเชิงเส้นพหุ โดยใช้วิธีการค้นหาแบบต้องห้าม ซึ่งมีฟังก์ชันเป้าหมายเป็นค่าความคลาดเคลื่อนกำลังสองเฉลี่ยและค่าเฉลี่ยความคลาดเคลื่อนสัมบูรณ์ แล้วนำผลมาเปรียบเทียบกับเทคนิคการคัดเลือกตัวแบบโดยวิธีการถดถอยแบบขั้นตอน เกณฑ์ที่ใช้ในการเปรียบเทียบ คือ ร้อยละของจำนวนครั้งที่คัดเลือกตัวแบบได้ถูกต้อง การศึกษาได้เปรียบเทียบการคัดเลือกตัวแปรอิสระเข้าสู่ตัวแบบในกรณีที่ไม่มีและมีปัญหาสหสัมพันธ์เชิงเส้นพหุ โดยใช้วิธีการจำลอง กรณีไม่มีปัญหาสหสัมพันธ์เชิงเส้นพหุ ความสามารถในการคัดเลือกตัวแบบของวิธีการถดถอยแบบขั้นตอนและการค้นหาแบบต้องห้ามที่มีฟังก์ชันเป้าหมายเป็นค่าความคลาดเคลื่อนกำลังสองเฉลี่ยมีร้อยละของความถูกต้องใกล้เคียงกัน หากใช้วิธีการค้นหาแบบต้องห้ามที่ใช้ฟังก์ชันเป้าหมายเป็นค่าเฉลี่ยความคลาดเคลื่อนสัมบูรณ์ มีร้อยละความถูกต้องในการคัดเลือกตัวแปรเข้าสู่ตัวแบบต่ำกว่ากรณีที่ใช้ค่าความคลาดเคลื่อนกำลังสองเฉลี่ยเล็กน้อย กรณีที่มีปัญหาสหสัมพันธ์เชิงเส้นพหุ วิธีการค้นหาแบบต้องห้ามที่มีฟังก์ชันเป้าหมายเป็นค่าความคลาดเคลื่อนกำลังสองเฉลี่ยและค่าเฉลี่ยความคลาดเคลื่อนสัมบูรณ์มีร้อยละของความถูกต้องในการคัดเลือกตัวแปรในตัวแบบสูงกว่าวิธีการถดถอยแบบขั้นตอน

คำสำคัญ: การคัดเลือกตัวแปร, การค้นหาแบบต้องห้าม, วิธีการถดถอยแบบขั้นตอน, สหสัมพันธ์เชิงเส้นพหุ

Abstract

This research proposed a variable selection method based on the Tabu Search for multiple linear regression models. In this study two objective functions, mean squared error (MSE) and mean absolute error (MAE), were used in the Tabu Search. The results of Tabu Search were compared with the results obtained by the stepwise regression method based on the hit percentage criterion. The simulations covered both cases, without and with multicollinearity problems. Without the multicollinearity problem, the hit percentages of the stepwise regression method and

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

Tabu Search using the objective function of MSE were almost the same but slightly higher than the Tabu Search using the objective function of the mean absolute error. With the multicollinearity problem, the hit percentages of Tabu Search using both objective functions were higher than the hit percentage of the stepwise regression method.

Keywords: Variable selection, Tabu search, Stepwise regression, Multicollinearity

1. บทนำ

การวิเคราะห์การถดถอย (Regression Analysis) เป็นวิธีการวิเคราะห์ทางสถิติที่เกี่ยวข้องกับการสร้างตัวแบบทางคณิตศาสตร์ เพื่อแสดงความสัมพันธ์ระหว่างตัวแปรตามกับตัวแปรอิสระจำนวน 1 ตัวหรือมากกว่า 1 ตัวก็ได้ [1] ซึ่งรูปแบบที่นิยมใช้กันอย่างแพร่หลาย คือ การวิเคราะห์การถดถอยเชิงเส้นพหุ (Multiple Linear Regression Analysis) โดยเป็นตัวแบบที่ความสัมพันธ์ของตัวแปรตามขึ้นอยู่กับตัวแปรอิสระมากกว่า 1 ตัวแปร ซึ่งทั้งตัวแปรอิสระและตัวแปรตามมีความสัมพันธ์ในรูปเชิงเส้น [2] โดยมีวัตถุประสงค์ เพื่อการพยากรณ์ค่าของตัวแปรตามจากตัวแปรอิสระ ตลอดจนการอนุมานต่างๆเกี่ยวกับตัวแปรตาม การใช้ตัวแบบการถดถอยเชิงเส้นในการพยากรณ์ให้มีประสิทธิภาพขึ้นอยู่กับการเลือกตัวแบบที่มีความเหมาะสม ซึ่งมาจากการคัดเลือกตัวแปรอิสระที่มีอิทธิพลต่อตัวแปรตาม จำนวนที่ไม่มากหรือน้อยเกินไปให้อยู่ในสมการพยากรณ์ เนื่องจากตัวแบบการถดถอยที่มีตัวแปรอิสระมากเกินความจำเป็น ทำให้ค่าพยากรณ์ที่ได้มีความคลาดเคลื่อนสูงและอาจเกิดปัญหาสหสัมพันธ์เชิงเส้นพหุ (Multicollinearity) ในทางตรงกันข้ามถ้าตัวแบบที่สร้างขึ้นขาดตัวแปรอิสระที่สำคัญไป ทำให้ค่าพยากรณ์มีความคลาดเคลื่อนสูงได้เช่นกัน ในปัจจุบันมีเทคนิคในการคัดเลือกตัวแปรอยู่หลายเทคนิคด้วยกัน ที่นิยมใช้กันอย่างแพร่หลาย คือ วิธีพิจารณาทุกตัวแบบที่เป็นไปได้ (All Possible Regression) วิธีเลือกตัวแปรอิสระแบบไปข้างหน้า (Forward Selection) วิธีตัดตัวแปรอิสระออกแบบถอยหลัง (Backward Elimination) วิธีการถดถอยแบบขั้นตอน (Stepwise Regression) [3] นอกจากเทคนิคการคัดเลือกตัวแปรที่กล่าวมาข้างต้นยังมีเทคนิคด้านการหาค่าอุตตมะโดยการจัดกลุ่ม (Combinatorial Optimization) ที่สามารถใช้คัดเลือกตัวแปรอิสระที่เหมาะสมได้เช่นกัน เช่น วิธีการค้นหาแบบต้องห้าม (Tabu Search) เป็นกระบวนการสำหรับแก้ปัญหาเพื่อหาค่าเหมาะสมที่สุด โดยใช้ข้อจำกัดและเวลาในการประมวลผลน้อยที่สุด [4] ซึ่งเป็นเทคนิคที่ใช้การกำหนดเกณฑ์ในการหยุดกระบวนการทำงาน (Algorithm) คือ จำนวนรอบสูงสุดและการลดลงของค่าฟังก์ชันเป้าหมาย (Objective Function) ซึ่งเป็นเทคนิคที่ใช้ระยะเวลาน้อยในการคัดเลือกตัวแปรเข้าสู่ตัวแบบ และมีความคลาดเคลื่อนในการพยากรณ์ต่ำ หรือกล่าวได้ว่าเป็นเทคนิคที่ช่วยในการป้องกันการเกิดปัญหาของตัวแบบที่มีตัวแปรอิสระมากเกินไป (Over Specification) นั่นคือ ได้ตัวแบบที่ประกอบด้วยตัวแปรอิสระมากกว่าตัวแบบจริง

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ในกรณีที่ตัวแปรอิสระไม่มีสหสัมพันธ์เชิงเส้นพหุ การประมาณค่าโดยใช้วิธีกำลังสองน้อยที่สุด (Ordinary Least Squares: OLS) สามารถประมาณค่าสัมประสิทธิ์การถดถอยของตัวแปรอิสระได้เป็นอย่างดี แต่ในกรณีที่ตัวแปรอิสระมีสหสัมพันธ์เชิงเส้นพหุวิธีการประมาณค่าพารามิเตอร์โดยวิธีกำลังสองน้อยที่สุด จะให้ค่าประมาณสัมประสิทธิ์การถดถอย ของตัวแปรอิสระที่มีความคลาดเคลื่อนสูง ซึ่งวิธีการแก้ไขจะใช้ วิธีการถดถอยแบบริดจ์ (Ridge Regression) ซึ่งต้องใช้ดุลยพินิจในการพิจารณาเส้นทางของริดจ์ (Ridge Trace) ดังนั้นผู้วิจัยจึงนำแนวคิดของวิธีการค้นหาแบบต้องห้ามมาใช้ในการคัดเลือกตัวแปรอิสระในกรณีที่ตัวแปรอิสระมีสหสัมพันธ์เชิงเส้นพหุแก่กัน เนื่องจาก Drezner and George [5] ได้เสนอวิธีการคัดเลือกตัวแปรอิสระในการวิเคราะห์การถดถอยเชิงเส้นพหุ โดยวิธีการค้นหาแบบต้องห้าม เปรียบเทียบกับวิธีการถดถอยแบบขั้นตอนและวิธี Maximum R^2 Improvement และ บุษยา ปภาพจน์ [6] ได้ทำการศึกษาเปรียบเทียบวิธีการคัดเลือกตัวแบบที่เหมาะสมที่สุดในการวิเคราะห์การถดถอยเชิงเส้นพหุ โดยเปรียบเทียบการคัดเลือกตัวแบบ 4 วิธี คือ วิธีการค้นหาแบบต้องห้าม วิธีการถดถอยแบบขั้นตอน วิธีเลือกตัวแปรอิสระแบบไปข้างหน้า และวิธี Maximum R^2 Improvement โดยผู้วิจัยทั้งสองทำการศึกษาเฉพาะกรณีที่ไม่มีสหสัมพันธ์เชิงเส้นพหุระหว่างตัวแปรอิสระ

ดังนั้นในการวิจัยนี้ผู้วิจัยต้องการศึกษาเปรียบเทียบการคัดเลือกตัวแปรอิสระของตัวแบบการถดถอยเชิงเส้นพหุ ในกรณีที่มีและไม่มีสหสัมพันธ์เชิงเส้นพหุระหว่างตัวแปรอิสระโดยใช้เทคนิคการค้นหาแบบต้องห้าม ที่ใช้ค่าความคลาดเคลื่อนกำลังสองเฉลี่ย (Mean Squares Error: MSE) และค่าเฉลี่ยความคลาดเคลื่อนสัมบูรณ์ (Mean Absolute Error: MAE) เป็นฟังก์ชันเป้าหมายกับวิธีการถดถอยแบบขั้นตอน โดยพิจารณาจากค่าร้อยละของจำนวนครั้งที่คัดเลือกตัวแบบได้ถูกต้องเป็นเกณฑ์ในการเปรียบเทียบ

2. ขอบเขตการศึกษา

ขอบเขตของการศึกษาครั้งนี้ คือ

2.1 ตัวแบบการถดถอยเชิงเส้นพหุที่อยู่ในรูปเชิงเส้นของพารามิเตอร์ ภายใต้ฐานคติ ความคลาดเคลื่อนสุ่มที่มีการแจกแจงแบบปกติที่มีค่าเฉลี่ยเท่ากับศูนย์และค่าความแปรปรวนเท่ากับ σ^2 โดยมีตัวแบบในรูป

$$\mathbf{y} = X\boldsymbol{\beta} + \boldsymbol{\varepsilon} \qquad (1)$$

โดยที่ $\mathbf{y}$ คือ เวกเตอร์ของตัวแปรตาม $\mathbf{y}$ ขนาด $n\times 1$

X คือ เมตริกซ์ของตัวแปรอิสระขนาด $n\times(k+1)$

$\boldsymbol{\beta}$ คือ เวกเตอร์ของพารามิเตอร์ของตัวแบบขนาด $(k+1)\times 1$

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

$\boldsymbol{\varepsilon}$ คือ เวกเตอร์ของความคลาดเคลื่อนสุ่มที่เป็นอิสระต่อกันขนาด $n\times1$ ภายใต้ข้อกำหนด $\boldsymbol{\varepsilon} \sim N_n(\mathbf{0},\sigma^2 I_n)$ เมื่อ I_n คือ เมตริกซ์เอกลักษณ์ขนาด $n\times n$ n คือ ขนาดตัวอย่าง และ k คือ จำนวนตัวแปรอิสระในตัวแบบ

2.2 การศึกษาใช้วิธีการจำลองข้อมูล เพื่อสร้างตัวแบบแท้จริงที่มีตัวแปรอิสระ 4 ตัวที่เกี่ยวข้องกับตัวแปรตาม และค่าความคลาดเคลื่อนสุ่ม ($\sigma^2 = 2500$)

2.3 ตัวแปรอิสระที่จำลองมาจากการแจกแจงแบบเอกรูป ที่ตัวแปรอิสระมีและไม่มีความสัมพันธ์ต่อกัน

2.4 ขนาดตัวอย่างที่ใช้ในการศึกษา คือ $n = 25$

2.5 วิธีที่ใช้ในการศึกษาเพื่อคัดเลือกตัวแปรมี 2 วิธี คือวิธีการค้นหาแบบต้องห้าม และวิธีการถดถอยแบบขั้นตอน โดยกำหนดระดับนัยสำคัญในการเลือกตัวแปรอิสระเข้าในตัวแบบและการคงตัวแปรอิสระไว้ในตัวแบบเป็น 0.05

2.6 ฟังก์ชันเป้าหมายที่ใช้ในการศึกษา โดยวิธีการค้นหาแบบต้องห้าม มี 2 ฟังก์ชัน คือ

$$\text{MSE} = \frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{n-k-1} \qquad (2)$$

$$\text{MAE} = \frac{\sum_{i=1}^{n}|y_i - \hat{y}_i|}{n} \qquad (3)$$

2.7 สร้างตัวแบบเต็มรูปโดยการเพิ่มตัวแปรอิสระเข้ามาในตัวแบบอีก 2 ตัว เพื่อใช้ในการพิจารณาการคัดเลือกตัวแปร

2.8 เกณฑ์การตัดสินใจ คือ ร้อยละของจำนวนครั้งที่คัดเลือกตัวแบบได้ถูกต้อง ซึ่งคือ

$$\frac{\text{จำนวนครั้งที่คัดเลือกตัวแบบได้ตัวแปรอิสระที่เกี่ยวข้องครบทุกตัว}}{\text{จำนวนครั้งที่สร้างตัวแบบเต็มรูป}} \times 100 \qquad (4)$$

2.9 ข้อมูลในการวิจัยใช้การจำลองแบบมอนติคาร์โล (Monte Carlo Simulation) โดยทำซ้ำจำนวน 500 ครั้ง โปรแกรมที่ใช้ในการจำลองและวิเคราะห์ข้อมูล คือ SAS Version 9.2 และ Microsoft Excel 2007

2.10 โปรแกรมที่ใช้ในการคัดเลือกตัวแบบโดยวิธีการค้นหาแบบต้องห้าม ดัดแปลงมาจาก "TS Directed by direct search methods for nonlinear global optimization" [7] ซึ่งเขียนโปรแกรมด้วย Matlab Version 7 R2010b

3. วิธีการศึกษา

นิยามคำศัพท์ที่เกี่ยวข้อง ดังนี้

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

การคัดเลือกตัวแบบที่ถูกต้อง หมายถึง การคัดเลือกแล้วได้ตัวแบบที่มีตัวแปรอิสระที่เกี่ยวข้องกับตัวแบบทั้ง 4 ตัว

ตัวแบบเต็มรูป (Full Model) หมายถึง ตัวแบบการถดถอยเชิงเส้นพหุที่ประกอบด้วยตัวแปรอิสระจำนวนหนึ่งที่ได้กำหนดขึ้นโดยอาศัยทฤษฎีที่เกี่ยวข้องและหรือการศึกษาในอดีตที่ผ่านมา

ตัวแบบแท้จริง (True Model) หมายถึง ตัวแบบการถดถอยเชิงเส้นพหุที่มีตัวแปรอิสระที่เกี่ยวข้องอยู่ในตัวแบบครบทุกตัวและความคลาดเคลื่อนสุ่มที่กำหนดขึ้นที่ใช้สร้างตัวแปรตาม

งานวิจัยนี้ศึกษาโดยการจำลองข้อมูลให้ตัวแปรอิสระมีการแจกแจงเอกรูป (Uniform distribution) และความคลาดเคลื่อนมีการแจกแจงปกติ (Normal distribution) จำนวนตัวแปรอิสระในตัวแบบเต็มรูป (Full Model) เท่ากับ 6 ตัวแปรและจำนวนตัวแปรอิสระในตัวแบบแท้จริง (True Model) เท่ากับ 4 ตัวแปร โดยมีขั้นตอนการดำเนินงาน ดังนี้

3.1 กรณีที่ข้อมูลไม่มีสหสัมพันธ์เชิงเส้นพหุ

3.1.1 สร้างประชากรที่มีขนาด N=100,000

3.1.2 สร้างตัวแปรอิสระ สัมประสิทธิ์การถดถอย ความคลาดเคลื่อน ดังนี้

3.1.3 สร้างตัวแปรอิสระ 6 ตัวที่เป็นอิสระต่อกัน ให้มีการแจกแจงเอกรูป โดย

$X_1 \sim U(15,80)$ $X_2 \sim U(20,150)$ $X_3 \sim U(10,100)$ $X_4 \sim U(-50,50)$

$X_5 \sim U(-100,100)$ และ $X_6 \sim U(-80,400)$

3.1.4 สร้างความคลาดเคลื่อนที่เป็นอิสระต่อกันให้มีการแจกแจงปกติ $\boldsymbol{\varepsilon} \sim N(0,2500)$

3.1.5 กำหนดค่าสัมประสิทธิ์การถดถอย คือ $\boldsymbol{\beta}' = (100, 24, 15, -8, 5, 0, 0)$ ในการสร้างตัวแปรตาม $\mathbf{y}$ หรือ อีกนัยหนึ่งตัวแปรตาม $\mathbf{y}$ จะขึ้นอยู่กับตัวแปรอิสระ 4 ตัว คือ X_1, X_2, X_3 และ X_4 เท่านั้น

3.1.6 สร้างตัวแปรตาม $\mathbf{y}$ จากตัวแบบการถดถอยต่อไปนี้

$$\mathbf{y} = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \beta_4 X_4 + \boldsymbol{\varepsilon}$$

3.1.7 เมื่อได้ประชากรตามที่ต้องการ ดำเนินการสุ่มตัวอย่างขนาด 25

3.1.8 ทำการคัดเลือกตัวแบบที่เหมาะสมจากวิธีการคัดเลือกตัวแบบทั้ง 2 วิธี คือ วิธีการค้นหาแบบต้องห้าม และวิธีการถดถอยแบบขั้นตอน และคำนวณค่าของฟังก์ชันเป้าหมาย โดยทำซ้ำจำนวน 500 ครั้งจากตัวแบบที่ได้จากการคัดเลือกตัวแบบทั้ง 2 วิธี

3.1.9 คำนวณร้อยละของจำนวนครั้งที่คัดเลือกตัวแบบได้ถูกต้องจากการทำซ้ำจำนวน 500 ครั้ง สำหรับแต่ละวิธี

3.1.10 เปรียบเทียบค่าความคลาดเคลื่อนกำลังสองเฉลี่ย (MSE) ของตัวแบบและร้อยละของจำนวนครั้งที่คัดเลือกตัวแบบได้ถูกต้องจากวิธีการคัดเลือกตัวแปรทั้ง 2 วิธี

3.2 กรณีที่ข้อมูลมีสหสัมพันธ์เชิงเส้นพหุ

3.2.1 สร้างประชากรที่มีขนาด N=100,000

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

3.2.2 กำหนดค่าสัมประสิทธิ์สหสัมพันธ์ (ρ) ระหว่างตัวแปรอิสระ X_1 และ X_4 เท่ากับ 0.999 โดยใช้วิธีรากที่สอง (square root method) [8] มีขั้นตอนดังนี้

1) จำลองตัวแปรสุ่ม U_1 และ U_2 จากเลขสุ่มที่มีการแจกแจงเอกรูป โดยอิสระกัน

2) คำนวณ $X_1=\mu_1+\sigma_1 U_1$ โดย μ_1 และ σ_1 คือ ค่าเฉลี่ยและส่วนเบี่ยงเบนมาตรฐานของตัวแปร X_1 ตามลำดับ

3) คำนวณ $X_4=\mu_4+\sigma_4(\rho U_1+\sqrt{1-\rho^2}U_2)$ โดย μ_4 และ σ_4 คือ ค่าเฉลี่ยและค่าเบี่ยงเบนมาตรฐานของตัวแปร X_4 ตามลำดับ

3.2.3 สร้างความคลาดเคลื่อนที่เป็นอิสระต่อกันให้มีการแจกแจงปกติ $\varepsilon \sim N(0,2500)$

3.2.4 กำหนดค่าสัมประสิทธิ์การถดถอย คือ $\boldsymbol{\beta}' = (100, 24, 15, -8, 5, 0, 0)$ สำหรับสร้างตัวแปรตาม $\mathbf{y}$

3.2.5 สร้างตัวแปรตาม $\mathbf{y}$ จากตัวแบบการถดถอยต่อไปนี้

$$\mathbf{y} = \beta_0+\beta_1X_1+\beta_2X_2+\beta_3X_3+\beta_4X_4+\boldsymbol{\varepsilon}$$

3.2.6 เมื่อได้ประชากรตามที่ต้องการ ดำเนินการสุ่มตัวอย่างขนาด 25 แล้วคำนวณค่าสัมประสิทธิ์สหสัมพันธ์ระหว่าง X_1 และ X_4 ในแต่ละตัวอย่างที่สุ่มมาได้

3.2.7 เนื่องจากปัญหาสหสัมพันธ์เชิงเส้นพหุ เป็นปัญหาที่พบบ่อยในตัวอย่าง [9] ดังนั้นผู้วิจัยจึงพิจารณาแบ่งกลุ่มของตัวอย่างที่ได้จากการสุ่มตามค่าสัมประสิทธิ์สหสัมพันธ์ระหว่าง X_1 และ X_4 ดังนี้

$0.495 \le r \le 0.504$ จัดให้อยู่ในกลุ่มค่าสัมประสิทธิ์สหสัมพันธ์เท่ากับ 0.50

$0.945 \le r \le 0.954$ จัดให้อยู่ในกลุ่มค่าสัมประสิทธิ์สหสัมพันธ์เท่ากับ 0.95

$0.955 \le r \le 0.964$ จัดให้อยู่ในกลุ่มค่าสัมประสิทธิ์สหสัมพันธ์เท่ากับ 0.96

$0.965 \le r \le 0.974$ จัดให้อยู่ในกลุ่มค่าสัมประสิทธิ์สหสัมพันธ์เท่ากับ 0.97

$0.975 \le r \le 0.984$ จัดให้อยู่ในกลุ่มค่าสัมประสิทธิ์สหสัมพันธ์เท่ากับ 0.98

$0.985 \le r \le 0.994$ จัดให้อยู่ในกลุ่มค่าสัมประสิทธิ์สหสัมพันธ์เท่ากับ 0.99

ทำซ้ำเช่นนี้จนครบ 500 ครั้ง ในแต่ละกลุ่มค่าสัมประสิทธิ์สหสัมพันธ์และขนาดตัวอย่าง

3.2.8 ทำการคัดเลือกตัวแบบที่เหมาะสมจากวิธีการคัดเลือกตัวแบบทั้ง 2 วิธี คือ วิธีการค้นหาแบบต้องห้าม และวิธีการถดถอยแบบขั้นตอน และคำนวณค่าความคลาดเคลื่อนกำลังสองเฉลี่ย (MSE) โดยทำซ้ำจำนวน 500 ครั้งจากตัวแบบที่ได้จากการคัดเลือกตัวแบบทั้ง 2 วิธี

3.2.9 คำนวณร้อยละของจำนวนครั้งที่คัดเลือกตัวแบบได้ถูกต้องจากการทำซ้ำจำนวน 500 ครั้ง ในแต่ละวิธี

3.2.10 เปรียบเทียบค่าความคลาดเคลื่อนกำลังสองเฉลี่ย (MSE) ของตัวแบบและร้อยละของจำนวนครั้งที่คัดเลือกตัวแบบได้ถูกต้องจากเทคนิคการคัดเลือกตัวแปรทั้ง 2 วิธี

3.3 ขั้นตอนการคัดเลือกตัวแบบโดยวิธีการค้นหาแบบต้องห้าม

3.3.1 เริ่มต้นโดยพิจารณากำหนดพิสัยของค่าพารามิเตอร์ จากวิธีกำลังสองน้อยที่สุด ดังนี้

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

b_0 = (50,150) , b_1 = (10,50) , b_2 = (0,30) , b_3 = (-40,0) , b_4 = (0,50) , b_5 = (-5,5) และ b_6 = (-10,10) กำหนดระยะที่อยู่ในรายการต้องห้ามเท่ากับ 10 รอบ

3.3.2 สุ่มค่าเริ่มต้นของพารามิเตอร์จากในพิสัยที่กำหนด

3.3.3 นำค่าเริ่มต้นของพารามิเตอร์มาคำนวณในฟังก์ชันเป้าหมาย MSE หรือ MAE

3.3.4 สร้างเซตคำตอบข้างเคียงของค่าพารามิเตอร์ปัจจุบัน

3.3.5 ตรวจสอบรายการต้องห้าม หากพบว่าไม่เกินระยะที่กำหนด จะไม่ยกเลิกการเป็นรายการต้องห้าม

3.3.6 ดำเนินการค้นหาค่าพารามิเตอร์จากเซตคำตอบข้างเคียงแล้วนำมาคำนวณในฟังก์ชันเป้าหมาย

3.3.7 พิจารณาหาค่าพารามิเตอร์ชุดต่อมา หลังจากนั้นนำมาคำนวณในฟังก์ชันเป้าหมาย

3.3.8 พิจารณาหาค่าพารามิเตอร์ จนครบ 20 รอบ

3.4 ขั้นตอนการคัดเลือกตัวแบบโดยวิธีการถดถอยแบบขั้นตอน

กำหนดจำนวนตัวแปรอิสระในตัวแบบเต็มรูปเท่ากับ k ตัวแปร

3.4.1 เลือกตัวแปรอิสระตัวที่ i เข้าสู่ตัวแบบ พิจารณาจาก $\max_{1\le i\le k}(r_{yi})$ โดยที่ r_{yi} เป็นค่าสัมประสิทธิ์สหสัมพันธ์ระหว่างตัวแปร y กับตัวแปร x_i จากนั้นทดสอบสมมติฐาน $H_0 : \beta_i = 0$ เทียบกับ $H_1 : \beta_i \neq 0$ โดยใช้สถิติทดสอบทีที่องศาเสรีเท่ากับ $n-k-1$ ถ้ายอมรับสมมติฐานหลักจะหยุดการพิจารณาเลือกตัวแปรอิสระเข้าสู่ตัวแบบ แต่ถ้าปฏิเสธสมมติฐานหลักแสดงว่าตัวแปรอิสระตัวที่ i มีส่วนในการอธิบายความผันแปรของตัวแปรตาม จึงสามารถคงตัวแปรอิสระตัวที่ i ไว้ในตัวแบบ

3.4.2 เลือกตัวแปรอิสระตัวที่ j เข้าสู่ตัวแบบเป็นตัวถัดไป พิจารณาจาก $\max_{\substack{1\le j\le k \\ j\neq i}}(r_{yj.i})$ โดยที่ $r_{yj.i}$ เป็นค่าสัมประสิทธิ์สหสัมพันธ์บางส่วน (Partial Correlation) ระหว่างตัวแปรตาม y กับตัวแปรอิสระ x_j เมื่อค่าของตัวแปรอิสระ x_i คงที่ จากนั้นทดสอบสมมติฐาน $H_0 : \beta_j = 0$ เทียบกับ $H_1 : \beta_j \neq 0$ โดยใช้สถิติทดสอบเอฟบางส่วน ถ้ายอมรับสมมติฐานหลักแสดงว่าตัวแปรอิสระตัวที่ j ไม่มีส่วนในการอธิบายความผันแปรของตัวแปรตามเมื่อมีตัวแปรอิสระตัวที่ i อยู่ในตัวแบบ จึงนำตัวแปรอิสระตัวที่ j ออกจากตัวแบบ แต่ถ้าปฏิเสธสมมติฐานหลักแสดงว่าตัวแปรอิสระตัวที่ j มีส่วนในการอธิบายความผันแปรของตัวแปรตามเมื่อมีตัวแปรอิสระตัวที่ i อยู่ในตัวแบบ จึงสามารถคงตัวแปรอิสระตัวที่ j ไว้ในตัวแบบ

3.4.3 หลังจากนั้น ทดสอบสมมติฐาน $H_0 : \beta_i = 0$ เทียบกับ $H_1 : \beta_i \neq 0$ โดยใช้สถิติทดสอบเอฟบางส่วน โดยถือว่าตัวแปรอิสระตัวที่เข้ามาใหม่อยู่ในตัวแบบและพิจารณาตัวแปรอิสระ

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ทุกตัวที่เข้ามาก่อนหน้านี้เป็นตัวแปรตัวสุดท้ายที่เข้ามาในตัวแบบ ถ้ายอมรับสมมติฐานหลักแสดงว่าตัวแปรอิสระตัวที่ i ไม่มีส่วนในการอธิบายความผันแปรของตัวแปรตามเมื่อมีตัวแปรอิสระตัวที่ j เข้ามาอยู่ในตัวแบบ จึงนำตัวตัวแปรอิสระตัวที่ i ออกจากตัวแบบ แต่ถ้าปฏิเสธสมมติฐานหลักแสดงว่าตัวแปรอิสระตัวที่ i มีส่วนในการอธิบายความผันแปรของตัวแปรตามเมื่อมีตัวแปรอิสระตัวที่ j เข้ามาอยู่ในตัวแบบ จึงสามารถคงตัวแปรอิสระตัวที่ i ไว้ในตัวแบบ

3.4.4 ทำซ้ำ 3.4.2 และ 3.4.3 จนกว่าจะไม่สามารถนำตัวแปรอิสระเข้ามาในตัวแบบหรือไม่สามารถนำตัวแปรอิสระออกจากตัวแบบได้

4. ผลการศึกษา

ตารางที่ 1 ร้อยละของจำนวนครั้งที่คัดเลือกตัวแบบได้ถูกต้อง จำแนกตามค่าสัมประสิทธิ์สหสัมพันธ์ระหว่าง x_1 กับ x_4 และวิธีที่ใช้ในการคัดเลือกตัวแบบ เมื่อ $\varepsilon_i \sim N(0,2500)$, $n = 25$

$r < 0.1$			$0.965 \leq r \leq 0.974$		
วิธีที่ใช้	ฟังก์ชันเป้าหมาย	ร้อยละ	วิธีที่ใช้	ฟังก์ชันเป้าหมาย	ร้อยละ
Tabu	MSE	88.60	Tabu	MSE	71.00
Tabu	MAE	85.60	Tabu	MAE	70.40
Stepwise	MSE	89.40	Stepwise	MSE	10.20
$0.495 \leq r \leq 0.504$			$0.975 \leq r \leq 0.984$		
วิธีที่ใช้	ฟังก์ชันเป้าหมาย	ร้อยละ	วิธีที่ใช้	ฟังก์ชันเป้าหมาย	ร้อยละ
Tabu	MSE	79.40	Tabu	MSE	71.20
Tabu	MAE	79.20	Tabu	MAE	69.40
Stepwise	MSE	45.20	Stepwise	MSE	9.20
$0.945 \leq r \leq 0.954$			$0.985 \leq r \leq 0.994$		
วิธีที่ใช้	ฟังก์ชันเป้าหมาย	ร้อยละ	วิธีที่ใช้	ฟังก์ชันเป้าหมาย	ร้อยละ
Tabu	MSE	71.40	Tabu	MSE	70.60
Tabu	MAE	70.20	Tabu	MAE	69.60
Stepwise	MSE	15.60	Stepwise	MSE	7.40

$0.955 \leq r \leq 0.964$		
วิธีที่ใช้	ฟังก์ชันเป้าหมาย	ร้อยละ
Tabu	MSE	70.80
Tabu	MAE	70.00
Stepwise	MSE	12.40

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

5. สรุปผลการศึกษา

ในการสรุปผลการศึกษา จะทำการกำหนดระดับของค่าสัมประสิทธิ์สหสัมพันธ์ของตัวแปรอิสระ X_1 และ X_4 ดังนี้

ค่าสัมประสิทธิ์สหสัมพันธ์	ระดับความสัมพันธ์
ต่ำกว่า 0.1	ต่ำ
0.495 – 0.504	กลาง
0.945 – 0.994	สูง

กรณีตัวแปรอิสระ X_1 และ X_4 มีสหสัมพันธ์เชิงเส้นพหุในระดับต่ำ วิธีการคัดเลือกตัวแปรอิสระโดยวิธีการถดถอยแบบขั้นตอน มีร้อยละของตัวแบบที่คัดเลือกได้ถูกต้องสูงที่สุด ในขณะที่วิธีการค้นหาแบบต้องห้ามเมื่อใช้ฟังก์ชันเป้าหมายทั้ง 2 ฟังก์ชัน มีร้อยละของการคัดเลือกตัวแบบต่ำกว่าวิธีการถดถอยแบบขั้นตอนเล็กน้อย

กรณีตัวแปรอิสระ X_1 และ X_4 มีสหสัมพันธ์เชิงเส้นพหุในระดับกลางและระดับสูง วิธีการคัดเลือกตัวแปรอิสระโดยใช้วิธีการค้นหาแบบต้องห้าม เมื่อใช้ฟังก์ชันเป้าหมายทั้ง 2 ฟังก์ชันมีร้อยละของการคัดเลือกตัวแบบได้ถูกต้องมากกว่าวิธีการถดถอยแบบขั้นตอน

ข้อสังเกตที่น่าสนใจจากการวิจัยนี้ คือ กรณีตัวแปรอิสระ X_1 และ X_4 มีสหสัมพันธ์เชิงเส้นพหุในระดับต่ำ วิธีการคัดเลือกตัวแปรอิสระโดยวิธีการถดถอยแบบขั้นตอน จะได้ค่าสัมประสิทธิ์การถดถอยของตัวแปรอิสระทุกตัวมีค่าใกล้เคียงกับค่าที่กำหนดในการจำลอง แต่วิธีการค้นหาแบบต้องห้ามเมื่อใช้ฟังก์ชันเป้าหมายทั้ง 2 ฟังก์ชัน จะให้ค่าสัมประสิทธิ์การถดถอยของตัวแปรอิสระ $X_1(\beta_1)$, $X_2(\beta_2)$ และ $X_4(\beta_4)$ สูงกว่าค่าที่กำหนดในการจำลอง ในขณะที่ค่าสัมประสิทธิ์การถดถอยของตัวแปรอิสระ $X_3(\beta_3)$ ต่ำกว่าค่าที่กำหนดในการจำลอง

กรณีตัวแปรอิสระ X_1 และ X_4 มีสหสัมพันธ์เชิงเส้นพหุในระดับกลางและระดับสูง วิธีการคัดเลือกตัวแปรอิสระโดยใช้วิธีการค้นหาแบบต้องห้าม เมื่อใช้ฟังก์ชันเป้าหมายทั้ง 2 ฟังก์ชัน จะได้ค่าสัมประสิทธิ์การถดถอยของตัวแปรอิสระทุกตัวมีค่าใกล้เคียงกับค่าที่กำหนดในการจำลองในทุกสถานการณ์ ในขณะที่การคัดเลือกตัวแปรอิสระโดยวิธีการวิธีการถดถอยแบบขั้นตอน จะให้ค่าสัมประสิทธิ์การถดถอยของตัวแปรอิสระ (β_0) มีเครื่องหมายผิดไปจากที่กำหนดไว้ในการจำลอง ซึ่งข้อสังเกตข้างต้นเป็นสิ่งที่น่าสนใจที่ผู้วิจัยจะทำการศึกษาในโอกาสต่อไป

กิตติกรรมประกาศ

ผู้วิจัยขอขอบคุณ รศ.ดร.วิชิต หล่อจีระชุณห์กุล ที่ให้คำแนะนำและข้อเสนอแนะที่ดีทำให้งานวิจัยชิ้นนี้สำเร็จลุล่วงไปด้วยดี

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

เอกสารอ้างอิง

[1] วิชิต หล่อจีระชุณห์กุล และจิราวัลย์ จิตรถเวช, 2548. เทคนิคการพยากรณ์. พิมพ์ครั้งที่ 3, โครงการส่งเสริมและพัฒนาเอกสารวิชาการ สถาบันบัณฑิตพัฒนบริหารศาสตร์, กรุงเทพฯ. [Vichit Lorchirachoonkul and Jirawan Jitthavech, 2005. Forecasting Techniques. 3rd ed. National Institute of Development Administration. Bangkok. (in Thai)]

[2] ทรงศิริ แต้สมบัติ, 2548. การวิเคราะห์การถดถอย. พิมพ์ครั้งที่ 3, มหาวิทยาลัยเกษตรศาสตร์, กรุงเทพฯ. [Songsiri Taesombut, 2005. Regression Analysis. 3rd ed. Kasetsart University. Bangkok. (in Thai)]

[3] Montgomery, D.C., Elizabeth, A.P. and Vining, G.G., 2006. Introduction to Linear Regression Analysis. 4th ed. John Willey & Sons, Inc., USA.

[4] Glover, F., 1990. Tabu Search. *A Tutorial Interface*, 20, 74-94.

[5] Drezner, Z. and George, A., 1999. Tabu Search model selection in multiple regression analysis. *Communication in Statistics - Simulation and Computation*, 28, 349-367.

[6] บุษยา ปภาพจน์, 2548. การเปรียบเทียบวิธีการคัดเลือกตัวแบบที่ดีที่สุดในการวิเคราะห์การถดถอยเชิงเส้นพหุ. วิทยานิพนธ์ปริญญามหาบัณฑิต มหาวิทยาลัยเกษตรศาสตร์. [Busaya Pabhapote, 2005. A Comparison of methods for selecting the best model in multiple linear regression analysis. M.S. Thesis, Kasetsart University. (in Thai)]

[7] Hedar, A. and Fukushima, M., 2003. TS directed by direct search methods for nonlinear global optimization. *Technical report*, 7, 7-18.

[8] Rubinstein, R.Y. and Kroese, D.P., 2008. Simulation and Monte Carlo Method. 2nd ed., John Wiley & Sons, Inc., New Jersey.

[9] Gujarati, D.N., 2006. Essentials of Econometrics. 3rd ed., McGraw-Hill, New York.

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้