

การศึกษสมรรถนะการคัดเลือกตัวแปรในการวิเคราะห์การถดถอยเชิงอันดับ  
เมื่อตัวแปรมีความสัมพันธ์เชิงเส้นพหุโดยใช้วิธีการก้าวกระโดดย้อนกลับได้

**A Study of The Performance for Variable Selection of Ordinal Regression  
Models under Multicollinearity Using Reversible Jump Algorithm**

ชลนรรจ์ สุขท้วญาติ<sup>1\*</sup> และ แสงหล้า ชัยมงคล<sup>2</sup>

<sup>1</sup>วิทยาศาสตร์มหาบัณฑิต/ภาควิชาคณิตศาสตร์และสถิติ คณะวิทยาศาสตร์และเทคโนโลยี มหาวิทยาลัยธรรมศาสตร์

คณะวิทยาศาสตร์และเทคโนโลยี มหาวิทยาลัยธรรมศาสตร์

เลขที่ 99 ถนน พหลโยธิน ตำบล คลองหนึ่ง อำเภอ คลองหลวง จังหวัด ปทุมธานี 12121

\* Corresponding author. E-mail : saenglplus@gmail.com

**บทคัดย่อ**

การวิจัยนี้มีวัตถุประสงค์เพื่อศึกษสมรรถนะการคัดเลือกตัวแปรในการวิเคราะห์การถดถอยเชิงอันดับเมื่อตัวแปรมีความสัมพันธ์เชิงพหุโดยใช้วิธีการก้าวกระโดดย้อนกลับได้ สำหรับตัวแบบออดส์เชิงสัดส่วน (PO), ตัวแบบออดส์ที่ไม่เป็นตามสัดส่วน (NPO) และตัวแบบออดส์เชิงสัดส่วนบางส่วน (PPO) เกณฑ์ที่ใช้ในการพิจารณา คือ ร้อยละของการระบุตัวแบบผิดโดยเฉลี่ย กำหนดตัวแปรตอบสนอง 3 ระดับ และตัวแปรอธิบาย 5 ตัวแปร  $X \sim N_5(0, \Sigma_{5 \times 5})$  ที่มีเมทริกซ์สหสัมพันธ์เป็นอิสระกัน, มีความสัมพันธ์แบบคงที่, Toeplitz และ Hub Toeplitz ผลการศึกษาพบว่า สมรรถนะการคัดเลือกตัวแปรด้วยวิธี RJ-MCMC ไม่ขึ้นอยู่กับรูปแบบและระดับความสัมพันธ์ระหว่างตัวแปรอธิบาย และการคัดเลือกตัวแปรของตัวแบบ PO มีสมรรถนะสูงกว่าตัวแบบ NPO

**คำสำคัญ:** ตัวแบบออดส์เชิงสัดส่วน, ตัวแบบออดส์ที่ไม่เป็นตามสัดส่วน, ตัวแบบออดส์เชิงสัดส่วนบางส่วน, มอนติคาร์โลโซมาร์คอฟก้าวกระโดดย้อนกลับได้

**Abstract**

The purpose of this study was to investigate the performance of variable selection in the ordinal regression models when explanatory variables were correlated using a reversible jump method. Three models were considered: Proportional odds model, Non-Proportional odds model and Partial proportional odds model. The criteria of this study is percentage of average misclassification. Each model includes a response variable with three levels and five explanatory variables  $X \sim N_5(0, \Sigma_{5 \times 5})$  with four difference types of correlation matrices: independent, constant correlation structure, Toeplitz, and Hub Toeplitz. The results show that variable selection performance using RJ- MCMC does not depend on the relationship structure and the level of relationship between the explanatory variables. The performance of variable selection of PO models was higher than that of NPO models.

**Keywords:** Proportional odds model (PO), Non-Proportional odds model (NPO), Partial proportional odds model (PPO), Reversible Jump Markov Chain Monte Carlo (RJ-MCMC)

<sup>1</sup>วิทยาศาสตร์มหาบัณฑิต/ภาควิชาคณิตศาสตร์และสถิติ <sup>2</sup>ภาควิชาคณิตศาสตร์และสถิติคณะวิทยาศาสตร์และเทคโนโลยี มหาวิทยาลัยธรรมศาสตร์

## 1. บทนำ

ตัวแบบที่ใช้ศึกษาอิทธิพลของตัวแปรอธิบาย ( $X$ ) หรือปัจจัยที่ส่งผลกระทบต่อตัวแปรตอบสนอง ( $Y$ ) แบบจำแนกประเภทคือ ตัวแบบลอจิสต์ (Logit Model) ที่ประกอบด้วยตัวแปรตอบสนองแบบจำแนกประเภทตั้งแต่ 2 ระดับหรือมากกว่าทั้งแบบไม่มีอันดับและแบบมีอันดับ ส่วนตัวแปรอธิบายโดยทั่วไปเป็นแบบจำแนกประเภททั้งหมดหรือแบบต่อเนื่องทั้งหมดหรือเป็นแบบผสม รูปแบบของตัวแบบลอจิสต์นิยมแปลงจากรูปแบบไม่เชิงเส้นเป็นรูปแบบเชิงเส้นโดยใช้ฟังก์ชันเชื่อมโยงแบบลอจิสต์ โดยตัวแบบอาจมีชื่อเรียกตามเงื่อนไขเพิ่มเติม เช่น ตัวแบบออดส์เชิงสัดส่วน (Proportional odds model: PO) ตัวแบบออดส์เชิงไม่สัดส่วน (Non-Proportional odds model: NPO) ตัวแบบออดส์เชิงสัดส่วนบางส่วน (Partial proportional odds model; PPO)

ตัวแบบออดส์เชิงสัดส่วน (PO) เป็นตัวแบบลอจิสต์ที่นิยมใช้ในกรณีที่ตัวแปรตอบสนองเป็นแบบอันดับที่มีระดับตั้งแต่ 2 ระดับขึ้นไป ตัวแบบนี้อยู่ภายใต้ข้อตกลงเบื้องต้นที่ว่าเวกเตอร์ความชัน ( $\beta$ ) ในตัวแบบที่ส่งผลต่อแต่ละระดับหรือกลุ่มของตัวแปรตอบสนอง ( $Y_j$ );  $j = 1, \dots, J-1$  เท่ากันเมื่อ  $J$  คือ ระดับหรือกลุ่มของตัวแปรตอบสนองทั้งหมด ความสัมพันธ์ของตัวแปรอธิบาย ( $X$ ) และตัวแปรตอบสนอง ( $Y$ ) จะไม่เปลี่ยนแปลงเมื่อระดับหรือกลุ่มของ  $Y$  เปลี่ยนทำให้  $\beta$  ไม่ขึ้นกับ  $j$  เรียกข้อสมมติของความเท่ากันของ log odds ในทุก ๆ จุดตัดว่า proportional odds assumption ภายใต้ข้อจำกัดการเรียงลำดับสุมของตัวแปรตอบสนอง

ตัวแบบออดส์ที่ไม่เป็นตามสัดส่วน (NPO) ตัวแบบจะมีความสัมพันธ์ระหว่างตัวแปรอธิบายและตัวแปรตอบสนองเปลี่ยนไปตามกลุ่มของตัวแปรตอบสนองคือสัมประสิทธิ์การถดถอย ( $\beta$ ) จะแตกต่างกันในแต่ละกลุ่มของตัวแปรตอบสนอง (Erkan Ari และ Zeki Yildiz, 2014) ซึ่งตัวแบบ NPO จะมีความยืดหยุ่นกว่าตัวแบบ PO แต่การใช้ตัวแบบ NPO ยังไม่แพร่หลายมากนัก เนื่องจากเงื่อนไขลำดับสุมอาจเป็นจริงเฉพาะช่วงใดช่วงหนึ่งของตัวแปรอธิบาย (Agresti, 2010) เป็นผลทำให้วิธีการประมาณค่าพารามิเตอร์ของตัวแบบที่ใช้กันโดยทั่วไปไม่สามารถทำได้และในกรณีที่ข้อสมมติของตัวแบบ PO ไม่เป็นจริง การประมาณค่าพารามิเตอร์ของตัวแบบ NPO อาจทำได้ยาก ดังนั้นทางเลือกของการสร้างตัวแบบทำได้โดยการสร้างตัวแบบการถดถอยลอจิสติกทวิภาคจำนวน  $J-1$  สมการแยกกันแต่การทำแบบนี้อาจทำให้ได้ค่าประมาณพารามิเตอร์ของตัวแบบที่แตกต่างไปจากการใช้ตัวแบบ NPO

ตัวแบบออดส์เชิงสัดส่วนบางส่วน (PPO) เป็นตัวแบบผสมของ PO และ NPO (Beredis Peterson et al., 1990) ภายใต้การทดสอบ proportional odds assumption ของตัวแบบ PPO ตัวแปรอธิบายบางตัวมีความสัมพันธ์กับตัวแปรตอบสนองแบบ PO ในขณะที่ตัวแปรที่เหลือจะมีความสัมพันธ์แบบ NPO

การคัดเลือกตัวแบบที่เหมาะสมตามวิธีการทางสถิติแบบดั้งเดิม (Classical statistical framework) จะใช้เกณฑ์ในการเปรียบเทียบตัวแบบ เช่น คัดเลือกตัวแบบด้วย AIC (Akaike's Information Criterion) หรือใช้การทดสอบอัตราส่วนภาวะน่าจะเป็นซึ่งการอนุมานจะทำตามเงื่อนไขในตัวแบบที่เลือกและไม่คำนึงถึงตัวแบบอื่น ๆ อาจทำให้ตัวแปรอธิบายที่มีความสัมพันธ์กับตัวแปรตอบสนองในหลาย ๆ ตัวแปรไม่ถูกเลือกเข้ามาในตัวแบบสุดท้าย ดังนั้น McKinley, T. J. และคณะ (2015) ได้เสนอแนวทางของเบสในการคัดเลือกตัวแบบโดยใช้ความน่าจะเป็นภายหลังของความเกี่ยวพัน (Posterior Probability of Association: PPA) ซึ่งวิธีนี้เป็นการหาระดับความสัมพันธ์ของตัวแปรอธิบายที่มีต่อตัวแปรตอบสนองของตัวแบบหนึ่ง ๆ โดยเฉลี่ยจากตัวแบบทั้งหมดที่เป็นไปได้แต่ตัวแบบ PO และ NPO มีจำนวนพารามิเตอร์ที่ต่างกันคือ ตัวแบบ PO มีพารามิเตอร์ความชันแค่ 1 ตัวในขณะที่ NPO มีจำนวน  $j-1$  ตัวเพื่อแก้ปัญหา McKinley, T. J. และคณะได้ใช้วิธีมอนติคาร์โลโซมาร์คอฟก้าวกระโดดย้อนกลับได้ (Reversible-Jump Markov Chain Monte Carlo: RJ-MCMC) ในการสร้างตัวแบบลอจิสต์โดยที่ไม่จำเป็นต้องรู้โครงสร้างความสัมพันธ์ระหว่างตัวแปรอธิบายและตัวแปรตอบสนองล่วงหน้าว่ามีลักษณะแบบ PO หรือ NPO และวิธีนี้ยังสามารถใช้ในการคัดเลือกตัวแปรของตัวแบบการถดถอยแบบอันดับได้ในการคัดตัวแปรเข้าหรือตัดตัวแปรที่อยู่ในตัวแบบการถดถอย

ออก อาจขึ้นอยู่กับโครงสร้างความสัมพันธ์ระหว่างตัวแปรอธิบายดังนั้นในงานวิจัยนี้ผู้วิจัยจึงสนใจศึกษาประสิทธิภาพของวิธีการ RJ-MCMC ในการคัดเลือกตัวแปรและตัวแบบการถดถอยแบบอันดับกรณีที่ตัวแปรอธิบายมีความสัมพันธ์กัน

## 2. วิธีการ

วิธีการในงานวิจัยนี้เน้นการคัดเลือกตัวแปรของตัวแบบที่เป็นเชิงอันดับโดยใช้วิธีมอนติคาร์โลโซมาร์คอฟกัวกระโดดย้อนกลับได้ ซึ่งนำเสนอโดย McKinley et al. (2015). โดยในงานนี้ได้ศึกษาในกรณีที่ตัวแปรอธิบายมีความสัมพันธ์กันได้พิจารณาแบบความสัมพันธ์ของตัวแปรอธิบายดังนี้ รูปแบบความสัมพันธ์แบบคงที่, รูปแบบความสัมพันธ์แบบ Toeplitz, รูปแบบความสัมพันธ์แบบ Hub Toeplitz และไม่มีความสัมพันธ์กัน

### 2.1 ตัวแบบลอจิสติกสำหรับตัวแปรตอบสนองแบบอันดับ

การกำหนดตัวแบบลอจิสติกสำหรับตัวแปรตอบสนองแบบอันดับโดย  $Y = 1, 2, 3$  และตัวแปรอธิบาย 5 ตัว ในงานวิจัยนี้ได้ศึกษาตัวแบบดังนี้

#### 2.1.1 ตัวแบบลอจิสติกอันดับเชิงสัดส่วน

ตัวแบบลอจิสติกอันดับเชิงสัดส่วน มีรูปแบบดังนี้

$$h(\gamma_{ij}) = \log\left(\frac{\gamma_{ij}}{1 - \gamma_{ij}}\right) = \theta_j - (\beta_1 X_{i1} + \beta_2 X_{i2} + \beta_3 X_{i3} + \beta_4 X_{i4} + \beta_5 X_{i5}) \quad (1)$$

ตัวแบบลอจิสติกอันดับเชิงสัดส่วนมีเงื่อนไขของเวกเตอร์ความชันที่ส่งผลต่อระดับของ  $Y_j$  นั่นคือ โดย  $\beta_j = \beta$  คงที่ในแต่ละระดับของ  $Y_j$  เมื่อ  $j = 1, \dots, J-1$  และ  $J$  คือ ระดับของตัวแปรตอบสนองทั้งหมด ดังนั้นรูปแบบของแต่ละลอจิสติกในรูปแบบเชิงเส้นจะขนานกันทั้ง  $J-1$  เส้น เมื่อ  $\theta_j$  แทน พารามิเตอร์จุดตัด (cut-point) และ  $\beta$  แทน เวกเตอร์ความชัน (Slope) โดยมีเงื่อนไขคือ  $\theta_j < \theta_{j+1}$

#### 2.1.2 ตัวแบบลอจิสติกที่ไม่เป็นตามสัดส่วน

ตัวแบบลอจิสติกที่ไม่เป็นตามสัดส่วนมีรูปแบบดังนี้

$$\log\left(\frac{\gamma_{ij}}{1 - \gamma_{ij}}\right) = \theta_j - (\beta_{1j} X_{i1} + \beta_{2j} X_{i2} + \beta_{3j} X_{i3} + \beta_{4j} X_{i4} + \beta_{5j} X_{i5}) \quad (2)$$

ตัวแบบนี้พารามิเตอร์ถดถอยจะแตกต่างกันตามระดับของกลุ่ม  $j$  และเงื่อนไขของลำดับสุมั้นคือ

$$-\infty < \theta_1 - \beta_1 X < \theta_2 - \beta_2 X < \dots < \theta_{J-1} - \beta_{J-1} X < \infty \quad (3)$$

สำหรับทุก  $X$  และถ้ามีตัวแปรอธิบาย 5 ตัว เมื่อ  $j = 1, \dots, J-2$  จะต้องเป็นจริงสำหรับทุก ๆ ค่าของ  $X_k$  ดังนั้น ถ้าทราบขอบเขตล่างและขอบเขตบนที่เป็นไปได้ของ  $X_k$  แล้วค่า

$$\theta_j - \theta_{j+1} < \sum_{k=1}^5 \min(\beta_{kj} - \beta_{k(j+1)}) X_k^m, [\beta_{kj} - \beta_{k(j+1)}] X_k^M \quad (4)$$

เมื่อ  $X_k^m$  และ  $X_k^M$  คือ ค่าต่ำสุดและค่าสูงสุดของ  $X_k$  ตามลำดับ

#### 2.1.3 ตัวแบบลอจิสติกบางส่วน

ตัวแบบลอจิตออกัสต์บางส่วน นำเสนอโดย Peterson และ Harrell (1990) เป็นตัวแบบผสมระหว่างตัวแบบลอจิตออกัสต์เชิงสัดส่วน (PO) และตัวแบบลอจิตออกัสต์ที่ไม่เป็นตามสัดส่วน (NPO) จะอยู่ในรูปแบบสมการดังนี้

$$\log\left(\frac{\gamma_{ij}}{1-\gamma_{ij}}\right) = \theta_j - \beta^T X_i - \eta_j^T U_i \quad (5)$$

โดย  $i=1, \dots, n$  และ  $j=1, \dots, J-1$  ซึ่ง  $\beta$  เป็นพารามิเตอร์ถดถอยที่สอดคล้องกับตัวแปรอธิบาย  $X_i$  ที่มีโครงสร้างของตัวแบบออกัสต์เชิงสัดส่วนและ  $\eta_j$  เป็นพารามิเตอร์ถดถอยที่สอดคล้องกับตัวแปรอธิบาย  $U_i$  ที่มีโครงสร้างของตัวแบบออกัสต์ที่ไม่เป็นตามสัดส่วน

ความน่าจะเป็นของทั้ง 3 ตัวแบบข้างต้นสามารถกำหนดได้ดังนี้ เมื่อกำหนดให้  $C_i$  คือ กลุ่มที่หน่วยตัวอย่างที่  $i$  อยู่  $C_i$  มีค่าเป็น  $1, 2, \dots, J$  จะได้ความน่าจะเป็นสะสม  $P(C_i \leq j) = \gamma_{ij}$  สามารถสร้าง  $\gamma_{ij}$  โดยเลือกฟังก์ชันเชื่อมโยงที่มีทิศทางเดียวและสามารถหาอนุพันธ์ได้ดังนี้

$$h(\gamma_{ij}) = \theta_j - \mu_i \quad (6)$$

เมื่อ  $\mu_i = \beta_0 + \beta_1 x_{i1} + \dots + \beta_k x_{ik}$ ,  $\beta = (\beta_0, \dots, \beta_k)$  และ  $\theta_j$  คือ จุดตัด (cut point) ที่เป็นตัวแปรต่อเนื่องที่ไม่ทราบค่า โดยที่  $-\infty < \theta_1 < \dots < \theta_{J-1} < \infty$  เพื่อให้ตัวแบบสามารถประมาณค่าได้จะกำหนดให้  $\beta_0 = 0$  ความน่าจะเป็นของแต่ละกลุ่มเป็นดังนี้

$$p_{ij} = \begin{cases} \gamma_{ij} & \text{for } j = 1, \\ \gamma_{ij} - \gamma_{i(j-1)} & \text{for } j = 2, \dots, J-1, \\ 1 - \gamma_{i(J-1)} & \text{for } j = J \end{cases} \quad (7)$$

ดังนั้น ความน่าจะเป็นสะสมของตัวแบบจะเป็นเรียงลำดับ  $0 < \gamma_{i1} < \gamma_{i2} < \dots < \gamma_{i(J-1)} < 1$

## 2.2 รูปแบบความสัมพันธ์ของตัวแปรอธิบาย

### 2.2.1 รูปแบบความสัมพันธ์แบบคงที่

รูปแบบความสัมพันธ์แบบคงที่ เป็นกรณีที่ความสัมพันธ์ของตัวแปรในแต่ละกลุ่มและระหว่างกลุ่มมีความสัมพันธ์คงที่ สมมติตัวอย่างรูปแบบความสัมพันธ์คงที่เมื่อตัวแปรอธิบาย 5 ตัว และ  $\rho_k = 0.5$  มีรูปแบบดังนี้

$$\sum_{k \times k} = \begin{bmatrix} 1 & \rho_k & \dots & \rho_k \\ \rho_k & 1 & \dots & \rho_k \\ \vdots & \vdots & \ddots & \vdots \\ \rho_k & \rho_k & \dots & 1 \end{bmatrix} = \begin{bmatrix} 1 & 0.50 & 0.50 & 0.50 & 0.50 \\ 0.50 & 1 & 0.50 & 0.50 & 0.50 \\ 0.50 & 0.50 & 1 & 0.50 & 0.50 \\ 0.50 & 0.50 & 0.50 & 1 & 0.50 \\ 0.50 & 0.50 & 0.50 & 0.50 & 1 \end{bmatrix}$$

### 2.2.2 รูปแบบความสัมพันธ์แบบ Toeplitz

รูปแบบความสัมพันธ์แบบ Toeplitz คือ ตัวแบบจะมีความสัมพันธ์สูงสำหรับกลุ่มตัวแปรอธิบายที่อยู่ใกล้กันในเมทริกซ์ความสัมพันธ์และตัวแบบจะมีค่าความสัมพันธ์ลดลงสำหรับกลุ่มตัวแปรอธิบายที่ห่างออกไป โดยความสัมพันธ์ระหว่างค่าสังเกตตัวที่  $l$  และ  $m$  คือ  $|l-m|$  จะลดลงอย่างเอกซ์โปเนนเชียล (Exponential) นั่นคือ  $\rho_k^{|l-m|}$  (Hardin et al, 2013) ดังนั้นโครงสร้างของรูปแบบความสัมพันธ์ของประชากรแบบ Toeplitz เป็นดังนี้

$$\Sigma_{k \times k} = \begin{bmatrix} 1 & \rho_k^{|l-m|} & \rho_k^{|l-m|} & \cdots & \rho_k^{|l-m|} \\ \rho_k^{|l-m|} & 1 & \rho_k^{|l-m|} & \cdots & \rho_k^{|l-m|} \\ \rho_k^{|l-m|} & \rho_k^{|l-m|} & 1 & \cdots & \rho_k^{|l-m|} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \rho_k^{|l-m|} & \rho_k^{|l-m|} & \rho_k^{|l-m|} & \cdots & 1 \end{bmatrix}$$

ตัวอย่างมีจำนวนตัวแปรอธิบายทั้งหมด 5 ตัวแปร และกำหนดให้  $\rho_k = 0.5$

$$\Sigma_{k \times k} = \begin{bmatrix} 1 & 0.5^{|1-2|} & 0.5^{|1-3|} & 0.5^{|1-4|} & 0.5^{|1-5|} \\ 0.5^{|2-1|} & 1 & 0.5^{|2-3|} & 0.5^{|2-4|} & 0.5^{|2-5|} \\ 0.5^{|3-1|} & 0.5^{|3-2|} & 1 & 0.5^{|3-4|} & 0.5^{|3-5|} \\ 0.5^{|4-1|} & 0.5^{|4-2|} & 0.5^{|4-3|} & 1 & 0.5^{|4-5|} \\ 0.5^{|5-1|} & 0.5^{|5-2|} & 0.5^{|5-3|} & 0.5^{|5-4|} & 1 \end{bmatrix} = \begin{bmatrix} 1 & 0.5 & 0.025 & 0.125 & 0.0625 \\ 0.5 & 1 & 0.5 & 0.025 & 0.125 \\ 0.025 & 0.5 & 1 & 0.5 & 0.025 \\ 0.125 & 0.025 & 0.5 & 1 & 0.5 \\ 0.0625 & 0.125 & 0.025 & 0.5 & 1 \end{bmatrix}$$

### 2.2.3 รูปแบบความสัมพันธ์แบบ Hub Toeplitz

รูปแบบความสัมพันธ์แบบ Hub Toeplitz จะสมมติว่าเราทราบความสัมพันธ์ระหว่างจุดศูนย์รวมค่าสังเกต (Hub Observation) ซึ่งโดยปกติจะเป็นค่าสังเกตแรกและแต่ละค่าสังเกตอื่น ๆ ซึ่งความสัมพันธ์ระหว่างค่าสังเกตที่ 1 กับค่าสังเกตที่  $m$  จะลดลงตามลำดับที่  $m$  ซึ่ง ( $m = 2, 3, \dots, g$ ) รูปแบบความสัมพันธ์แบบ Hub Toeplitz เป็น

ดังนี้ คำนวณได้จาก  $\alpha_{k,1} = 1$  และ  $\alpha_{k,m} = \rho_{\max} - \tau_k (m-2)$  เมื่อ  $\tau_k = \frac{(\rho_{\max} - \rho_{\min})}{g_k - 2}$  และ  $2 \leq m \leq g_k$

สมมติตัวอย่างมีจำนวนตัวแปรอธิบายทั้งหมด 5 ตัวแปร และกำหนดให้  $\rho_{\max} = 0.8$  และ  $\rho_{\min} = 0.3$

$$\Sigma_{k \times k} = \begin{bmatrix} 1 & \alpha_{k,2} & \alpha_{k,3} & \cdots & \alpha_{k,g_k} \\ \alpha_{k,2} & 1 & \alpha_{k,2} & \cdots & \alpha_{k,g_k-1} \\ \alpha_{k,3} & \alpha_{k,2} & 1 & \cdots & \alpha_{k,g_k-2} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \alpha_{k,g_k} & \alpha_{k,g_k-1} & \alpha_{k,g_k-2} & \cdots & 1 \end{bmatrix} = \begin{bmatrix} 1 & 0.95 & 0.80 & 0.65 & 0.50 \\ 0.95 & 1 & 0.95 & 0.80 & 0.65 \\ 0.80 & 0.95 & 1 & 0.95 & 0.80 \\ 0.65 & 0.80 & 0.95 & 1 & 0.95 \\ 0.50 & 0.65 & 0.80 & 0.95 & 1 \end{bmatrix}$$

โดยที่  $\Sigma_{k \times k}$  คือ เมทริกซ์สหสัมพันธ์ขนาด ( $k \times k$ )

$k$  คือ จำนวนกลุ่มการจำแนกหรือจำนวนตัวแปรอธิบาย ( $k > 0$ )

$g_k$  คือ ขนาดของจำนวนกลุ่มการจำแนกหรือขนาดตัวอย่าง ( $g_k > 0$ )

$\rho_k$  คือ สหสัมพันธ์ของตัวแปรอธิบายมีค่าอยู่ระหว่าง  $0 \leq \rho_k \leq 1$

## 2.3 วิธีมอนติคาร์โลโซมาร์คอฟกัวกระโดดย้อนกลับได้ในการคัดเลือกตัวแบบและคัดเลือกตัวแปร

### 2.3.1 การคัดเลือกตัวแบบแบบเบสด้วยวิธีมอนติคาร์โลโซมาร์คอฟกัวกระโดดย้อนกลับได้

สมมติว่ามีตัวแบบแข่งขันที่ต้องการเลือกจำนวน  $V$  ตัวแบบ แนวทางการคัดเลือกตัวแบบแบบเบสจะทำได้โดยการประมาณความน่าจะเป็นภายหลังที่ถือว่าตัวแบบ ( $M_v$ ) เป็นตัวแบบที่ถูกคัดเลือกจากตัวแบบที่เป็นไปได้ทั้งหมด ( $M_1, \dots, M_v$ ) โดยความน่าจะเป็นภายหลังกรณีนี้จะเขียนได้ดังนี้

$$P(M_v|D) = \frac{f(D|M_v)P(M_v)}{\sum_{u=1}^V f(D|M_u)P(M_u)} \quad (8)$$

โดย  $P(M_v)$  คือ ความน่าจะเป็นก่อนของตัวแบบ  $M_v$  และ

$$f(D|M_v) = \int_{\Omega_v} f(D|M_v) f(\omega_v|M_u) d\omega_v \quad (9)$$

คือ อินทิเกรตของฟังก์ชันภาวะน่าจะเป็น โดยที่  $D$  คือ ข้อมูล,  $\Omega_v$  คือ ปริภูมิพารามิเตอร์ (หลายมิติ) และ  $\omega_v$  คือ เวกเตอร์พารามิเตอร์ของตัวแบบ  $M_u$

การคัดเลือกตัวแบบโดยใช้มอนติคาร์โลโซมาร์คอฟกัวกระโดดย้อนกลับได้ (Reversible-Jump Markov Chain Monte Carlo: RJ-MCMC) จะกำหนด  $v=i$  เป็นตัวบ่งชี้ของตัวแบบสำหรับการวนซ้ำในรอบนั้น ๆ ให้  $P(M_i \rightarrow M_j)$  คือ ความน่าจะเป็นที่จะกระโดดจากตัวแบบ  $M_i$  ไปตัวแบบ  $M_j$  ซึ่งตัวแบบทั้ง 2 อาจมีจำนวนตัวแปรอธิบายในตัวแบบแตกต่างกันเพื่อให้การข้ามระหว่างตัวแบบของมิติที่แตกต่างกันได้จึงจับคู่พารามิเตอร์ในตัวแบบ  $i$  หรือ  $\omega_i$  กับพารามิเตอร์ของตัวแบบ  $j$  หรือ  $\omega_j$  โดยสร้างชุดพารามิเตอร์ดัมมี่  $u_i$  และ  $u_j$  จะได้  $\dim(\omega_i, u_i) = \dim(\omega_j, u_j)$  ความน่าจะเป็นในการยอมรับของการย้ายจะกำหนดได้ดังนี้

$$\alpha = \min \left[ 1, \frac{f(D|\omega_j, M_j)}{f(D|\omega_i, M_i)} \times \frac{f(\omega_j|M_j)}{f(\omega_i|M_i)} \times \frac{P(M_j)}{P(M_i)} \times \frac{P(M_j \rightarrow M_i)q_u(u_j)}{P(M_i \rightarrow M_j)q_u(u_i)} \times \left| \frac{\partial(\omega_j, u_j)}{\partial(\omega_i, u_i)} \right| \right] \quad (10)$$

โดย  $q_u(u_i)$  และ  $q_u(u_j)$  คือฟังก์ชันความหนาแน่นน่าจะเป็นที่เสนอสำหรับพารามิเตอร์ดัมมี่  $u_i$

และ  $u_j$  ตามลำดับ และ  $\left| \frac{\partial(\omega_j, u_j)}{\partial(\omega_i, u_i)} \right|$  คือค่าดีเทอร์มิแนนต์ของเมทริกซ์จาโคเบียน

### 2.3.2 การคัดเลือกตัวแปรในตัวแบบการถดถอยลอจิสติกแบบอันดับโดยใช้วิธีมอนติคาร์โลโซมาร์คอฟกัวกระโดดย้อนกลับได้

เมื่อมีตัวแปรอธิบายจำนวน  $K$  ตัว แต่ละตัวจะมีรูปแบบความสัมพันธ์กับตัวแปรตอบสนอง  $Y$  ในลักษณะตัวแบบ PO, NPO หรือไม่มีความสัมพันธ์กับตัว  $Y$  อย่างใดอย่างหนึ่งใน 3 รูปแบบนี้ จำนวนตัวแบบที่เป็นไปได้ทั้งหมดเท่ากับ  $3^K$  ตัวแบบ การสร้างตัวแบบจะทำได้โดยการสร้างตัวแปรบ่งชี้สำหรับตัวแปรแต่ละตัว คือ  $S_k$  สำหรับ  $k=1, \dots, K$  โดยกำหนดให้

$$S_k = \begin{cases} 0 & \text{if } X_k \text{ has a PO structure} \\ 1 & \text{if } X_k \text{ has an NPO structure} \\ 2 & \text{if } X_k \text{ is excluded} \end{cases} \quad (11)$$

#### 2.3.2.1 การคัดตัวแปรเข้าหรือออกจากตัวแบบ

การคัดตัวแปรที่ยังไม่อยู่ในตัวแบบเข้ามาเพิ่มในตัวแบบหรือตัดตัวแปรที่อยู่ในตัวแบบออกจะทำให้ความน่าจะเป็น  $P_{jump}$  การคัดตัวแปรเพิ่มเข้าในตัวแบบจะกำหนดให้ความน่าจะเป็นในการคัดตัวแปรเข้าในตัวแบบด้วยโครงสร้าง PO และ NPO คือ  $p_{PO}$  และ  $p_{NPO} = 1 - p_{PO}$  ตามลำดับ กำหนดฟังก์ชันการเพิ่มตัวแปรด้วยโครงสร้าง PO คือ

$$g_{0 \rightarrow PO}(u_1) = u_1 = \beta_k \quad (12)$$

เมื่อ  $u_1$  ถูกสุ่มจากการแจกแจงที่มีฟังก์ชันความหนาแน่นน่าจะเป็น  $q_u(\cdot)$  และการเพิ่มตัวแปรใน  
ตัวแบบโดยโครงสร้าง NPO จะกำหนดฟังก์ชันดังนี้

$$q_{0 \rightarrow NPO}(u_1, \dots, u_{J-1}) = (u_1, \dots, u_{J-1}) = (\beta_{k1}, \dots, \beta_{k(J-1)}) \quad (13)$$

โดย  $u_1, \dots, u_{J-1}$  เป็นอิสระกันและมีการแจกแจงเหมือนกันถูกสุ่มมาจากการแจกแจงที่มีฟังก์ชัน  
ความหนาแน่นน่าจะเป็น  $q_u(\cdot)$  ความน่าจะเป็นการยอมรับการเพิ่มตัวแปรด้วยตัวแบบ PO หรือ  $0 \rightarrow PO$  คือ

$$\alpha = \min \left[ 1, \frac{f(Y|\beta', \theta^{(i)})}{f(Y|\beta^{(i)}, \theta^{(i)})} \times \frac{f(\beta'_k)}{1} \times \frac{1}{q_u(u_1)} \times \frac{1}{p_{PO}} \right] \quad (14)$$

สำหรับการย้ายจากตัวแบบ  $0 \rightarrow NPO$  ความน่าจะเป็นการยอมรับคือ

$$\alpha = \min \left[ 1, \frac{f(Y|\beta', \theta^{(i)})}{f(Y|\beta^{(i)}, \theta^{(i)})} \times \frac{\left[ \prod_{j=1}^{J-1} f(\beta'_{kj}) \right] \left[ \prod_{j=2}^{J-1} f(\theta_j^{(i)} | \theta_{j-1}^{(i)}, \beta'_j, \beta'_{j-1}) \right]}{\left[ \prod_{j=2}^{J-1} f(\theta_j^{(i)} | \theta_{j-1}^{(i)}, \beta_j^{(i)}, \beta_{j-1}^{(i)}) \right]} \times \frac{1}{q_u(u_1, \dots, u_{J-1})} \times \frac{1}{1 - p_{PO}} \right] \quad (15)$$

กำหนด  $u_j$  เป็นตัวแปรสุ่มที่อิสระกันและมีการแจกแจงเหมือนกันที่มีการแจกแจงปกติค่าเฉลี่ย  
เป็น 0 และ  $\sigma_{P\beta}^2$  คือ ความแปรปรวนที่นำเสนอหรือเขียนได้เป็น  $u_j \sim N(0, \sigma_{P\beta}^2)$  ส่วนการคัดตัวแปรออกจากตัว  
แบบทำได้โดยย้อนกลับกระบวนการการคัดตัวแปรเข้าโดยเปลี่ยนความน่าจะเป็นที่ยอมรับได้ให้เป็นความน่าจะเป็นของ  
การคัดตัวแปรออก

### 2.3.2.2 การคัดตัวแปรเข้าหรือออกจากตัวแบบ

กรณีที่มีตัวแปรตัวใหม่เพิ่มเข้ามาในตัวแบบ RJ-MCMC จะปรับปรุงตัวแบบ 2 ประการคือ ประการ  
แรกจะทำการปรับปรุงค่าพารามิเตอร์ของตัวแปรที่ละตัว (สมมติตัวแปร  $k$ ) โดยลำดับการปรับปรุงจะทำแบบสุ่มซึ่ง  
ค่าพารามิเตอร์ค่าใหม่จะถูกแทนค่าเดิมด้วยความน่าจะเป็น  $P_{move}$  และประการที่สองจะทำการสลับโครงสร้างของ  
ความสัมพันธ์จากแบบ  $PO \rightarrow NPO$  (ถ้าตัวแปร  $K$  มีโครงสร้าง PO) หรือ  $NPO \rightarrow PO$  (ถ้าตัวแปร  $K$  มี  
โครงสร้าง NPO) ด้วยความน่าจะเป็น  $1 - P_{move}$

ขั้นตอนการสลับโครงสร้างความสัมพันธ์จะอาศัยขั้นตอนก๊าวกระโดดย้อนกลับได้โดยการสลับ  
 $NPO \rightarrow PO$  หรือเชื่อมโยง  $(\beta_{k1}, \dots, \beta_{k(J-1)}) \rightarrow \beta_k$  ทำได้โดยการกำหนดฟังก์ชันดังนี้

$$g_{NPO \rightarrow PO}(\beta_{k1}, \dots, \beta_{k(J-1)}) = (\bar{\beta}_k, \bar{\beta}_k - 2\beta_{k2}, \dots, \bar{\beta}_k - 2\beta_{k(J-1)}) \\ = (\beta_k, u_1, \dots, u_{J-2}) \quad (16)$$

โดย  $\bar{\beta}_k = (J-1)^{-1} \sum_{j=1}^{J-1} \beta_{kj}$  และการสลับกลับจะใช้ฟังก์ชันผกผันนั้นคือ

$$g_{PO \rightarrow NPO}(\beta_k, u_1, \dots, u_{J-2}) = \frac{J-1}{2} \left( \beta_k(4-J) + \sum_{j=1}^{J-2} u_j, (\beta_k - u_1), \dots, (\beta_k - u_{J-2}) \right) \\ = (\beta_{k1}, \dots, \beta_{k(J-1)}) \quad (17)$$

ความน่าจะเป็นในการยอมรับสำหรับการสลับ  $PO \rightarrow NPO$  คือ

$$\alpha = \min \left[ 1, \frac{f(Y|\beta', \theta^{(i)})}{f(Y|\beta^{(i)}, \theta^{(i)})} \times \frac{\prod_{j=1}^{J-1} f(\beta'_{kj})}{f(\beta_k^{(i)})} \times \frac{\prod_{j=2}^{J-1} f(\theta_j^{(i)}|\theta_{j-1}^{(i)}, \beta'_j, \beta'_{j-1})}{\prod_{j=2}^{J-1} f(\theta_j^{(i)}|\theta_{j-1}^{(i)}, \beta_j^{(i)}, \beta_{j-1}^{(i)})} \right. \\ \left. \times \frac{1}{q_u(u_1, \dots, u_{J-2})} \times (J-1) \left( \frac{J-1}{2} \right)^{J-2} \right]$$

(18)

โดยในทอมสุดท้ายคือค่าสมบูร์นของดีเทอร์มีแนนของเมทริกซ์จาโคเบียน เช่นเดียวกับความน่าจะเป็นในการยอมรับในการสลับ  $NPO \rightarrow PO$  คือ

$$\alpha = \min \left[ 1, \frac{f(Y|\beta', \theta^{(i)})}{f(Y|\beta^{(i)}, \theta^{(i)})} \times \frac{f(\beta'_k)}{\prod_{j=1}^{J-1} f(\beta'_{kj})} \times \frac{\prod_{j=2}^{J-1} f(\theta_j^{(i)}|\theta_{j-1}^{(i)}, \beta'_j, \beta'_{j-1})}{\prod_{j=2}^{J-1} f(\theta_j^{(i)}|\theta_{j-1}^{(i)}, \beta_j^{(i)}, \beta_{j-1}^{(i)})} \right. \\ \left. \times \frac{q_u(u_1, \dots, u_{J-2})}{1} \times \left( \frac{1}{J-1} \right) \left( \frac{2}{J-1} \right)^{J-2} \right] \quad (19)$$

#### 2.4 การจำลองข้อมูล

ในการวิจัยนี้กำหนดตัวแปรอธิบาย 5 ตัวแปร คือ  $(x = x_1, x_2, x_3, x_4, x_5)$  สุ่มจากการแจกแจงปรกติหลายตัวแปร (Multivariate Normal Distribution) ที่มีเวกเตอร์ค่าเฉลี่ย 0 และมีเมทริกซ์ความแปรปรวนและความแปรปรวนรวมเป็น  $\Sigma_{5 \times 5}$  เมื่อกำหนดรูปแบบโครงสร้างความสัมพันธ์ 4 แบบคือ โครงสร้างความสัมพันธ์เป็นอิสระกัน, ความสัมพันธ์แบบคงที่, ความสัมพันธ์แบบ Toeplitz และความสัมพันธ์แบบ Hub Toeplitz โดยกำหนดสัมประสิทธิ์สหสัมพันธ์เป็น  $r = 0, 0.3, 0.5, 0.8$

สถานการณ์การจำลองในการศึกษามี 3 กรณีดังนี้

1. กำหนดเป็นตัวแบบ PO โดยกำหนดให้ตัวแปร 4 ตัวมีความสัมพันธ์กับ  $Y$  ในรูปแบบของ PO และสุ่มอีกหนึ่งตัวแปรให้ไม่มีความสัมพันธ์ ( $\beta = 0$ )
2. กำหนดเป็นตัวแบบ NPO ทำเช่นเดียวกับการที่ 1 แต่ให้ความสัมพันธ์ของตัวแปร 4 ตัวอยู่ในรูปแบบ NPO
3. กำหนดเป็นตัวแบบ PPO ทำโดยจะสุ่มตัวแปร 1 ตัวให้ไม่มีความสัมพันธ์และอีก 1 ตัวให้มีความสัมพันธ์แบบ NPO ส่วนตัวแปรที่เหลือให้มีความสัมพันธ์แบบ PO

จำลองพารามิเตอร์ถดถอย  $\beta_{kj}$  และพารามิเตอร์จุดตัด  $\theta_j$  โดยสุ่มจากการแจกแจงปรกติที่มีค่าเฉลี่ยเป็น  $\mu$  และความแปรปรวน  $\sigma^2$  เมื่อ  $k=1, 2, \dots, 5$  และ  $j=1, 2$  โดยสร้างตามเงื่อนไขในข้อ 2.1 ดังนี้ กรณีที่เป็นตัวแบบ PO กำหนดให้  $\theta_1 \sim N(-1, 0.1^2)$ ,  $\theta_2 \sim N(1, 0.1^2) + N(0, 0.1^2)$ ,  $\beta_1, \beta_3, \beta_4, \beta_5$  สุ่มจากการแจกแจงปรกติที่มีค่าเฉลี่ยเป็น 0 และความแปรปรวน  $5^2$  และ  $\beta_2 = 0$  ในกรณีที่เป็นตัวแบบ NPO กำหนดให้  $\theta_1 \sim N(4, 0.1^2)$ ,  $\theta_2 \sim N(6, 0.1^2) + N(0, 0.1^2)$ ,  $\beta_{11} = 0$ ,  $\beta_{21} \sim N(3, 0.5^2)$ ,  $\beta_{31} \sim N(1, 2^2)$ ,  $\beta_{41} \sim N(4, 1^2)$ ,  $\beta_{51} \sim N(2, 5^2)$ ,  $\beta_{21} = 0$ ,  $\beta_{22} \sim N(2, 3^2)$ ,  $\beta_{32} \sim N(1, 4^2)$ ,  $\beta_{42} \sim N(6, 2^2)$ ,  $\beta_{52} \sim N(0, 1^2)$  และกรณีที่เป็นตัวแบบ PPO กำหนดให้  $\theta_1 \sim N(-1, 0.1^2)$ ,  $\theta_2 \sim N(1, 0.1^2) + N(0, 0.1^2)$ ,  $\beta_1 \sim N(0, 0.1^2)$ ,  $\beta_2 = 0$ ,  $\beta_3, \beta_{41}, \beta_{42}$  และ  $\beta_5$  สุ่มจากการแจกแจงปรกติที่มีค่าเฉลี่ยเป็น 0 และความแปรปรวน  $\beta_5$

สำหรับตัวแปรตอบสนองกำหนดให้มี 3 ระดับ คือ  $Y=1, 2$  และ 3 โดยสุ่มจากการแจกแจง  $u$  เมื่อ  $u \sim U(0,1)$  เป็นจุดตัด (cut-point) ในการแบ่งกลุ่มตัวแปรอธิบาย โดย  $Y=1$  ด้วยความน่าจะเป็น  $p_{i1}$ ,  $Y=2$  ด้วยความน่าจะเป็น  $p_{i2}$  และ  $Y=3$  ด้วยความน่าจะเป็น  $p_{i3}$  ซึ่งจะเปรียบเทียบได้ดังนี้ กรณีที่  $u < p_{i1}$  จะให้  $Y=1$  กรณีที่  $p_{i1} \leq u \leq p_{i1} + p_{i2}$  จะให้  $Y=2$  และกรณีที่  $u \geq p_{i1} + p_{i2}$  จะได้  $Y=3$  (โดยแต่ละกลุ่มของตัวแปรตอบสนอง ( $Y$ ) ต้องมีอย่างน้อย 5% ถ้าไม่ถึง 5% ต้องทำการสุ่มตัวอย่างใหม่) ให้ขนาดตัวอย่าง  $n=500$  ทำการจำลองข้อมูลซ้ำ 100 ชุด คัดเลือกตัวแปรด้วยวิธี RJ-MCMC ใช้จำนวนลูกโซ่ 2 โซ่ จำนวนรอบซ้ำ 20,000 รอบและตัด 10,000 รอบแรกทิ้ง (burn-in) และใช้ชุดคำสั่ง BayesOrd จาก <https://github.com/tjmckinley/BayesOrd> เกณฑ์ที่ใช้ในการพิจารณา คือ ร้อยละของการระบุตัวแบบผิด (misclassification) โดยเฉลี่ย โดยทำการเปรียบเทียบค่าความน่าจะเป็นภายหลังของความเกี่ยวพัน (Posterior Probability of Association: PPA) ของตัวแปรในแต่ละรูปแบบซึ่งค่า PPA คำนวณได้ดังนี้

$$PPA = \frac{1}{1 + BF_{01} \times \text{prior odds}} = \frac{BF_{10} \times \text{prior odds}(H_1)}{1 + BF_{10} \times \text{prior odds}(H_1)}$$

โดย  $\text{prior odds}(H_1) = \frac{\Pr(H_1)}{\Pr(H_0)}$ ,  $BF_{01} = \frac{\Pr(x|H_0)}{\Pr(x|H_1)}$  และ  $BF_{10} = \frac{\Pr(x|H_1)}{\Pr(x|H_0)}$

เมื่อ  $H_0$ : ตัวแปร  $X$  มีรูปแบบความสัมพันธ์กับ  $Y$  ตามที่กำหนด  
และ  $H_1$ : ตัวแปร  $X$  มีรูปแบบความสัมพันธ์กับ  $Y$  ไม่เป็นไปตามที่กำหนด

### 3. ผลและอภิปราย

ตารางที่ 1 แสดงร้อยละการคัดเลือกตัวแปรและรูปแบบความสัมพันธ์โดยเฉลี่ยของตัวแบบ PO จากการเปรียบเทียบความน่าจะเป็นภายหลังของความเกี่ยวพันของตัวแปรในแต่ละรูปแบบจากการคำนวณค่า PPA โดยกำหนดตัวแบบเป็น PO เมื่อตัวแปรอธิบายมีความสัมพันธ์ในรูปแบบต่าง ๆ โดยถ้าค่า PPA มีมากที่สุด จะกำหนดให้ตัวแปรมีโครงสร้างเป็นตัวแบบนั้น ๆ กรณีที่ศึกษานี้มีจำนวนสถานการณ์ที่เป็นไปได้ เมื่อกำหนดให้ตัวแปรอธิบายทั้ง 4 มีความสัมพันธ์กับตัวแปรตอบสนองในลักษณะของตัวแบบ PO และอีกหนึ่งตัวไม่มีความสัมพันธ์กับตัวแปรตอบสนองพบว่า ตัวแปรทั้ง 4 ตัวถูกคัดเข้ามาในสมการในทุกกรณีของรูปแบบความสัมพันธ์และทุกระดับของสัมประสิทธิ์สหสัมพันธ์ (ร้อยละของการคัดออกโดยเฉลี่ยเท่ากับ 0) และส่วนใหญ่จะถูกคัดเข้าด้วยโครงสร้างความสัมพันธ์แบบ PO มีเพียงไม่กี่ร้อยละ 1 ที่ตัวแปรจะถูกคัดเข้าด้วยโครงสร้าง NPO ในขณะที่ตัวแปรที่ไม่มีความสัมพันธ์กับตัวแปรตอบสนองจะไม่ถูกคัดเลือกเข้าในตัวแบบมีเพียงส่วนน้อย (ไม่เกินร้อยละ 3) ที่ถูกคัดเข้าในตัวแบบด้วยโครงสร้างแบบ PO

ตรงข้ามกับกรณีที่กำหนดให้เป็นตัวแบบ NPO ในตารางที่ 2 จะพบว่ามีตัวแปรบางตัวที่มีความสัมพันธ์กับตัวแบบแต่ไม่ถูกคัดเข้ามาในตัวแบบ (ร้อยละการคัดออกอยู่ในช่วง 0.25-8) และมีร้อยละของการระบุความสัมพันธ์ผิดเป็นแบบ PO อยู่ในช่วง 36-39.5 ในขณะที่ตัวแปรที่กำหนดให้ไม่มีความสัมพันธ์กับตัวแบบมีร้อยละของการคัดเลือกเข้าในตัวแบบในรูปแบบ NPO (อยู่ในช่วงร้อยละ 3-11) สูงกว่า PO (อยู่ในช่วงร้อยละ 0-2)

ตารางที่ 3 ในกรณีที่กำหนดตัวแบบ PPO ตัวแปรอธิบายที่กำหนดให้มีความสัมพันธ์แบบ PO จะไม่ถูกคัดเข้ามาในตัวแบบด้วยร้อยละ 30 (โดยประมาณ) และไม่ถูกระบุว่ามีความสัมพันธ์ผิดเป็นแบบ NPO ยกเว้นในกรณีที่รูปแบบความสัมพันธ์คงที่ เมื่อ  $r = 0.5, 0.8$  กับรูปแบบความสัมพันธ์ Toeplitz ที่  $r = 0.8$  แต่ด้วยร้อยละเพียง 0.33 เท่านั้น ตรงข้ามกับตัวแปรที่กำหนดให้มีความสัมพันธ์แบบ NPO จะถูกคัดเข้ามาในตัวแบบด้วยความสัมพันธ์ที่ถูกต้องทุกกรณี สำหรับตัวแปรที่กำหนดให้ไม่มีความสัมพันธ์กับตัวแบบจะถูกคัดเข้าในตัวแบบเฉพาะในรูปแบบ PO ด้วยร้อยละไม่เกิน 5

ตารางที่ 1 ร้อยละการคัดเลือกตัวแปรและรูปแบบความสัมพันธ์โดยเฉลี่ยของตัวแบบ PO

| รูปแบบความสัมพันธ์ | สหสัมพันธ์ | ตัวแบบ            | ร้อยละการคัดเลือกตัวแปรและรูปแบบความสัมพันธ์โดยเฉลี่ย |      |        | ร้อยละการระบุตัวแบบผิด |
|--------------------|------------|-------------------|-------------------------------------------------------|------|--------|------------------------|
|                    |            |                   | PO                                                    | NPO  | คัดออก |                        |
| อิสระกัน           | $r = 0$    | PO                | 99.75                                                 | 0.25 | 0.00   | 0.25                   |
|                    |            | ไม่มีความสัมพันธ์ | 3.00                                                  | 0.00 | 97.00  | 3.00                   |
| คงที่              | $r = 0.3$  | PO                | 100.00                                                | 0.00 | 0.00   | 0.00                   |
|                    |            | ไม่มีความสัมพันธ์ | 2.00                                                  | 1.00 | 97.00  | 3.00                   |
| คงที่              | $r = 0.5$  | PO                | 99.75                                                 | 0.25 | 0.00   | 0.25                   |
|                    |            | ไม่มีความสัมพันธ์ | 1.00                                                  | 0.00 | 99.00  | 1.00                   |
|                    | $r = 0.8$  | PO                | 99.75                                                 | 0.25 | 0.00   | 0.25                   |
|                    |            | ไม่มีความสัมพันธ์ | 2.00                                                  | 0.00 | 98.00  | 2.00                   |
| Toeplitz           | $r = 0.3$  | PO                | 99.00                                                 | 1.00 | 0.00   | 1.00                   |
|                    |            | ไม่มีความสัมพันธ์ | 2.00                                                  | 0.00 | 98.00  | 2.00                   |
|                    | $r = 0.5$  | PO                | 100.00                                                | 0.00 | 0.00   | 0.00                   |
|                    |            | ไม่มีความสัมพันธ์ | 2.00                                                  | 0.00 | 98.00  | 2.00                   |

ตารางที่ 1 (ต่อ) ร้อยละการคัดเลือกตัวแปรและรูปแบบความสัมพันธ์โดยเฉลี่ยของตัวแบบ PO

|              |                                  |                   |        |      |       |      |
|--------------|----------------------------------|-------------------|--------|------|-------|------|
| Toeplitz     | $r = 0.8$                        | PO                | 100.00 | 0.00 | 0.00  | 0.00 |
|              |                                  | ไม่มีความสัมพันธ์ | 3.00   | 0.00 | 97.00 | 3.00 |
| Hub Toeplitz | $r_{\min} = 0.3, r_{\max} = 0.8$ | PO                | 100.00 | 0.00 | 0.00  | 0.00 |
|              |                                  | ไม่มีความสัมพันธ์ | 1.00   | 0.00 | 99.00 | 1.00 |

ตารางที่ 2 ร้อยละการคัดเลือกตัวแปรและรูปแบบความสัมพันธ์โดยเฉลี่ยของตัวแบบ NPO

| รูปแบบความสัมพันธ์ | สหสัมพันธ์ | ตัวแบบ            | ร้อยละการคัดเลือกตัวแปรและรูปแบบความสัมพันธ์โดยเฉลี่ย |       |        | ร้อยละการระบุตัวแบบผิด |
|--------------------|------------|-------------------|-------------------------------------------------------|-------|--------|------------------------|
|                    |            |                   | PO                                                    | NPO   | คัดออก |                        |
| อิสระกัน           | $r = 0$    | NPO               | 36.00                                                 | 63.75 | 0.25   | 36.25                  |
|                    |            | ไม่มีความสัมพันธ์ | 1.00                                                  | 3.00  | 96.00  | 4.00                   |
| คงที่              | $r = 0.3$  | NPO               | 39.50                                                 | 59.75 | 0.75   | 40.25                  |
|                    |            | ไม่มีความสัมพันธ์ | 2.00                                                  | 4.00  | 94.00  | 6.00                   |
|                    | $r = 0.5$  | NPO               | 36.50                                                 | 62.00 | 1.50   | 38.00                  |
|                    |            | ไม่มีความสัมพันธ์ | 1.00                                                  | 8.00  | 91.00  | 9.00                   |
|                    | $r = 0.8$  | NPO               | 37.50                                                 | 54.50 | 8.00   | 45.50                  |
|                    |            | ไม่มีความสัมพันธ์ | 0.00                                                  | 11.00 | 89.00  | 11.00                  |
| Toeplitz           | $r = 0.3$  | NPO               | 38.50                                                 | 60.25 | 1.25   | 39.75                  |
|                    |            | ไม่มีความสัมพันธ์ | 1.00                                                  | 3.00  | 96.00  | 4.00                   |

|                 |                   |                   |       |       |       |       |
|-----------------|-------------------|-------------------|-------|-------|-------|-------|
|                 | $r = 0.5$         | NPO               | 37.50 | 62.00 | 0.50  | 38.00 |
|                 |                   | ไม่มีความสัมพันธ์ | 2.00  | 5.00  | 93.00 | 7.00  |
|                 | $r = 0.8$         | NPO               | 39.00 | 57.00 | 4.00  | 43.00 |
|                 |                   | ไม่มีความสัมพันธ์ | 1.00  | 5.00  | 94.00 | 6.00  |
| Hub<br>Toeplitz | $r_{\min} = 0.3,$ | NPO               | 39.50 | 56.25 | 4.25  | 43.75 |
|                 | $r_{\max} = 0.8$  | ไม่มีความสัมพันธ์ | 1.00  | 9.00  | 90.00 | 10.00 |

ตารางที่ 3 ร้อยละการคัดเลือกตัวแปรและรูปแบบความสัมพันธ์โดยเฉลี่ยของตัวแบบ PPO

| รูปแบบ<br>ความสัมพันธ์ | สหสัมพันธ์ | ตัวแบบ            | ร้อยละการคัดเลือกตัวแปรและรูปแบบ<br>ความสัมพันธ์โดยเฉลี่ย |        |        | ร้อยละการระบุ<br>ตัวแบบผิด |
|------------------------|------------|-------------------|-----------------------------------------------------------|--------|--------|----------------------------|
|                        |            |                   | PO                                                        | NPO    | คัดออก |                            |
| อิสระกัน               | $r = 0$    | PO                | 70.00                                                     | 0.00   | 30.00  | 30.00                      |
|                        |            | NPO               | 0.00                                                      | 100.00 | 0.00   | 0.00                       |
|                        |            | ไม่มีความสัมพันธ์ | 1.00                                                      | 0.00   | 99.00  | 1.00                       |
| คงที่                  | $r = 0.3$  | PO                | 70.00                                                     | 0.00   | 30.00  | 30.00                      |
|                        |            | NPO               | 0.00                                                      | 100.00 | 0.00   | 0.00                       |
|                        |            | ไม่มีความสัมพันธ์ | 1.00                                                      | 0.00   | 99.00  | 1.00                       |
|                        | $r = 0.5$  | PO                | 68.33                                                     | 0.34   | 31.33  | 31.67                      |
|                        |            | NPO               | 0.00                                                      | 100.00 | 0.00   | 0.00                       |
|                        |            | ไม่มีความสัมพันธ์ | 0.00                                                      | 0.00   | 100.00 | 0.00                       |

ตารางที่ 3 (ต่อ) ร้อยละการคัดเลือกตัวแปรและรูปแบบความสัมพันธ์โดยเฉลี่ยของตัวแบบ PPO

| รูปแบบ<br>ความสัมพันธ์ | สหสัมพันธ์ | ตัวแบบ            | ร้อยละการคัดเลือกตัวแปรและรูปแบบ<br>ความสัมพันธ์โดยเฉลี่ย |        |        | ร้อยละการระบุ<br>ตัวแบบผิด |
|------------------------|------------|-------------------|-----------------------------------------------------------|--------|--------|----------------------------|
|                        |            |                   | PO                                                        | NPO    | คัดออก |                            |
| คงที่                  | $r = 0.8$  | PO                | 67.00                                                     | 0.33   | 32.67  | 33.00                      |
|                        |            | NPO               | 0.00                                                      | 100.00 | 0.00   | 0.00                       |
|                        |            | ไม่มีความสัมพันธ์ | 2.00                                                      | 1.00   | 97.00  | 3.00                       |
| Toeplitz               | $r = 0.3$  | PO                | 69.33                                                     | 0.00   | 30.67  | 30.67                      |
|                        |            | NPO               | 0.00                                                      | 100.00 | 0.00   | 0.00                       |
|                        |            | ไม่มีความสัมพันธ์ | 2.00                                                      | 0.00   | 98.00  | 2.00                       |
|                        | $r = 0.5$  | PO                | 68.00                                                     | 0.00   | 32.00  | 32.00                      |
|                        |            | NPO               | 0.00                                                      | 100.00 | 0.00   | 0.00                       |
|                        |            | ไม่มีความสัมพันธ์ | 5.00                                                      | 0.00   | 95.00  | 5.00                       |
|                        | $r = 0.8$  | PO                | 67.67                                                     | 0.33   | 32.00  | 32.33                      |
|                        |            | NPO               | 0.00                                                      | 100.00 | 0.00   | 0.00                       |

|                 |                                       |                   |       |        |       |       |
|-----------------|---------------------------------------|-------------------|-------|--------|-------|-------|
|                 |                                       | ไม่มีความสัมพันธ์ | 3.00  | 0.00   | 97.00 | 3.00  |
| Hub<br>Toeplitz | $r_{\min} = 0.3,$<br>$r_{\max} = 0.8$ | PO                | 68.00 | 0.33   | 31.67 | 32.00 |
|                 |                                       | NPO               | 0.00  | 100.00 | 0.00  | 0.00  |
|                 |                                       | ไม่มีความสัมพันธ์ | 2.00  | 0.00   | 98.00 | 2.00  |

#### 4. บทสรุป

ผู้จากการศึกษาและจำลองตัวแบบ PO, NPO และ PPO ภายใต้สถานการณ์ที่ตัวแปรอธิบายมีรูปแบบความสัมพันธ์และระดับของสัมประสิทธิ์สหสัมพันธ์ต่างกัน ผลการศึกษาพบว่าสมรรถนะการคัดเลือกตัวแปรในการวิเคราะห์การถดถอยเชิงอันดับด้วยขั้นตอนวิธีก้าวกระโดดย้อนกลับได้มีแนวโน้มที่ไม่ขึ้นอยู่กับรูปแบบความสัมพันธ์และระดับความสัมพันธ์ระหว่างตัวแปรอธิบายและเป็นวิธีที่ไม่คัดเลือกตัวแปรที่ไม่มีความสัมพันธ์กับตัวแบบเข้ามาในตัวแบบ การคัดเลือกตัวแปรของตัวแบบ PO มีสมรรถนะสูงกว่าตัวแบบ NPO โดยในตัวแบบ NPO นั้นมีตัวแปรบางตัวจะถูกคัดเข้ามาในตัวแบบด้วยโครงสร้างของ PO ในขณะที่ตัวแปรที่มีโครงสร้างแบบ PO จะไม่ถูกคัดเข้ามาในตัวแบบด้วยโครงสร้างของ NPO ไม่ว่าจะเป็นในตัวแบบ PO หรือ ตัวแบบ PPO ทั้งนี้เนื่องจากตัวแบบ NPO มีพารามิเตอร์มากกว่าตัวแบบ PO จึงมีความแตกต่างระหว่างมิติของพารามิเตอร์ทำให้ในการดำเนินงานตัวแบบ PO ซึ่งเป็นตัวแบบที่ง่ายกว่าจึงถูกเลือกมากกว่า

#### 5. เอกสารอ้างอิง

- Agresti, A. (2010). **Categorical Data Analysis**. 3<sup>rd</sup> ed. Canada: John Wiley & Sons.
- Ari, E., and Yildiz, Z. (2014). Parallel Linear Assumption in Ordinal Logistic Regression and Analysis Approaches. **International Interdisciplinary Journal of Scientific Research**, 1(3), 1-16.
- Gelman, A., Carlin, J. B., Stern, H. S., David B. Dunson, Vehtari, A., and Rubin, D. B. (2014). **Bayesian Data Analysis**. 3<sup>rd</sup> ed. United States: CRC Press.
- Hardin, J., Garcia, S. R., and Golan, D. (2013). A Method for Generating Realistic Correlation Matrices. **The Annals of Applied Statistics**, 7(3), 1733-1762. doi: 10.1214/13-AOAS638.
- Hastie, D., and Green, P. J. (2011). Model choice using reversible jump Markov Chain Monte Carl. **Statistica Neerlandica**, 66(3), 309-338. doi: 10.1111/j.1467-9574.2012.00516.x.
- McKinley, T. J., Morters, M., and Wood, J. L. N. (2015). Bayesian Model Choice in Cumulative Link Ordinal Regression Models. **International Society for Bayesian Analysis**, 10,1-30. doi: 10.1214/14-BA884.
- Peterson, B., Harrell, F. E., and Jr. (1990). Partial Proportional Odds Models for Ordinal Response Variables. **Journal of the Royal Statistical Society Series C (Applied Statistics)**, 39(2), 205-217.
- Tutz, G., and Scholz, T. (2003). Ordinal regression modelling between proportional odds and non-proportional odds. **Technical report, Institute of Statistics**, University of Munich.