

การประมาณค่าพารามิเตอร์ในตัวแทนความถดถอยโลจิสติกแบบเกาส์เซียนคอปพูลา

The Parameter Estimation of Logistic Regression Model with Gaussian Copula

ศุภกัญญา บุญมา¹ และ เสกสรรค์ เกียรติสุไพบูรณ์²

บทคัดย่อ

งานวิจัยนี้มีวัตถุประสงค์เพื่อประมาณค่าพารามิเตอร์ในตัวแทนความถดถอยโลจิสติกแบบเกาส์เซียนคอปพูลา สำหรับในกรณีที่ทราบค่าปัจจัยของคอปพูลา Z พบว่า ตัวแบบคอปพูลาโลจิสติกนี้ คือ ตัวแบบโพรบิต ที่ต้องมีการปรับค่าตัวประมาณด้วย $\sqrt{1-\rho}$ เมื่อ ρ คือ ค่าสหสัมพันธ์ในตัวแทนเกาส์เซียนคอปพูลา และเมื่อได้ทำการจำลองข้อมูลเพื่อทดสอบความถูกต้องในการประมาณค่าพารามิเตอร์ในกรณีที่ทราบค่าปัจจัยของคอปพูลา Z และทำการทดสอบเพิ่มเติมในกรณีที่ไมทราบค่าปัจจัยของคอปพูลา Z พบว่า ในกรณีที่ทราบค่าปัจจัยของคอปพูลา Z ได้ให้ผลการประมาณที่ใกล้เคียงกับค่าพารามิเตอร์ ในทางตรงข้ามกรณีที่ไมทราบค่าปัจจัยของคอปพูลา Z ให้ผลการประมาณที่ค่อนข้างสูงกว่าค่าพารามิเตอร์ โดยเฉพาะในระดับความสัมพันธ์ที่สูงขึ้น

คำสำคัญ : ตัวแบบโลจิส, ตัวแบบโพรบิต, เกาส์เซียนคอปพูลา, ตัวแบบโลจิสติกแบบเกาส์เซียนคอปพูลา

Abstract

The objective of this research is to estimate the parameters of a logistic regression model with one-factor Gaussian copula. When the copula factor Z is known, this copula logistic regression model becomes a probit model. In this case, the parameter estimation must be adjusted by the factor of $\sqrt{1-\rho}$ where ρ is the correlation parameter of the Gaussian copula. A simulated data is generated to test the corrected estimation technique when the copula factor value is known, and to test the traditional estimation technique when the copula factor is unknown. When the value of the copula factor is known, the corrected estimation technique yields correct parameter values. On the other hand, when the copula factor value is unknown, the traditional regression estimation technique yields an overestimated parameter values.

Keywords: Logit model, Probit model, Gaussian copula, Logistic with gaussian copula model

¹ นิสิตปริญญาโท, ภาควิชาสถิติ, คณะพาณิชยศาสตร์และการบัญชี, จุฬาลงกรณ์มหาวิทยาลัย

² ผู้ช่วยศาสตราจารย์, ภาควิชาสถิติ, คณะพาณิชยศาสตร์และการบัญชี, จุฬาลงกรณ์มหาวิทยาลัย

1. บทนำ

ในปัจจุบันการวิเคราะห์ข้อมูลด้วยความถดถอยโลจิสติกได้มีการนำมาใช้อย่างแพร่หลาย ในทางการแพทย์ ระบาดวิทยา ทางธุรกิจ การเงิน เป็นต้น โดยในการวิเคราะห์ความถดถอยโลจิสติกมีวัตถุประสงค์เช่นเดียวกันกับการวิเคราะห์ความถดถอยเชิงเส้น เพียงแต่ตัวแปรตามในการวิเคราะห์ความถดถอยโลจิสติกนั้นเป็นตัวแปรจำแนกพวกหรือตัวแปรเชิงคุณภาพ ซึ่งตัวแปรตามในแต่ละค่าสังเกต (Observation) จะมีการแจกแจงแบบเบอร์นูลลีที่เป็นอิสระกัน ดังนั้น ฟังก์ชันความน่าจะเป็นร่วมจะเท่ากับผลคูณของฟังก์ชันความน่าจะเป็นของตัวแปรแต่ละค่าสังเกตนั่นเอง สำหรับตัวแปรอิสระอาจเป็นข้อมูลเชิงปริมาณ หรือข้อมูลเชิงคุณภาพ อย่างใดอย่างหนึ่งหรือทั้งสองอย่าง

จากงานวิจัยของสันติ ชีรภัทรและเสกสรร [1] ได้ศึกษาการพยากรณ์มูลค่าความเสี่ยง (Credit Value at Risk: Credit VaR) ซึ่งเป็นการวัดความเสี่ยงที่มีความสำคัญในการจัดการด้านความเสี่ยงของสถาบันทางการเงินด้วยความถดถอยโลจิสติก พบว่าการประมาณค่ามูลค่าความเสี่ยงด้วยความถดถอยโลจิสติกมีค่าประมาณต่ำกว่าค่าพารามิเตอร์ที่ควรจะเป็น (Underestimate) ซึ่งอาจเป็นไปได้ว่ามีการระบุตัวแปรอิสระไม่ครบถ้วนหรือไม่เพียงพอ หรืออาจเกิดเงื่อนไขความผิดปกติบางอย่างที่มีผลต่อความคลาดเคลื่อนของข้อมูลและอาจส่งผลกระทบต่อความเป็นอิสระของตัวแปรตามและทำให้ตัวแปรตามในแต่ละค่าสังเกตมีความสัมพันธ์กันได้ ดังนั้นจึงได้เกิดแนวคิดในการนำความสัมพันธ์เข้ามาประกอบในตัวแปรตามหรือสร้างความสัมพันธ์ระหว่างตัวแปรตามในแต่ละค่าสังเกตขึ้น โดยความสัมพันธ์ที่ทำการศึกษาในครั้งนี้ ได้แก่ เกาส์เซียนคอปูลา (Gaussian copula) ซึ่งเป็นความสัมพันธ์ที่ผ่านทางการแจกแจงร่วมของควอนไทล์ (Quantile) ของตัวแปรสุ่มร่วมที่มีการแจกแจงแบบปกติหลายตัวแปร (Multivariate normal variables) ซึ่งเรียกตัวแบบที่ได้นี้ว่า ตัวแบบความถดถอยโลจิสติกแบบเกาส์เซียนคอปูลา

ในงานวิจัยนี้มีวัตถุประสงค์เพื่อศึกษาและประมาณค่าตัวแบบความถดถอยโลจิสติกแบบเกาส์เซียนคอปูลา พร้อมทั้งเปรียบเทียบความถูกต้องในการประมาณค่าพารามิเตอร์จากตัว

แบบที่ระดับความสัมพันธ์ต่างๆ เพื่อสามารถนำไปใช้ในการวิเคราะห์ข้อมูลต่อไปได้อย่างเหมาะสม

2. ทฤษฎี

ตัวแบบความถดถอยโลจิสติกแบบเกาส์เซียนคอปูลา เป็นตัวแบบโลจิสติกที่ตัวแปรตามในแต่ละค่าสังเกตมีความสัมพันธ์กันด้วยเกาส์เซียนคอปูลา โดยทฤษฎีที่ใช้ในตัวแบบนี้ ได้แก่ เกาส์เซียนคอปูลา ตัวแบบโลจิสติก ตัวแบบพหุคูณ และตัวแบบความถดถอยโลจิสติกแบบเกาส์เซียนคอปูลา โดยมีรายละเอียดดังต่อไปนี้

2.1 เกาส์เซียนคอปูลา

เกาส์เซียนคอปูลา (Gaussian copula) $U = (U_1, U_2, \dots, U_n)$ [2-3] เป็นความสัมพันธ์ที่ผ่านทางการแจกแจงร่วมของควอนไทล์ ของตัวแปรสุ่มร่วมที่มีการแจกแจงแบบปกติ หรืออาจกล่าวได้ว่าเป็นการแจกแจงร่วมของตัวแปรสุ่ม n ตัว ที่มีการแจกแจงแบบปกติ (n -variate normal distribution) ด้วยเมทริกซ์สหสัมพันธ์เชิงเส้น Σ

$$\text{โดย } \Sigma = \begin{bmatrix} 1 & \rho & \cdots & \rho \\ \rho & 1 & \cdots & \rho \\ \vdots & \ddots & \ddots & \vdots \\ \rho & \rho & \cdots & 1 \end{bmatrix}$$

ฟังก์ชันการแจกแจงร่วมของเกาส์เซียนคอปูลามีรูปแบบ ดังนี้

$$C_{\Sigma}^{Ga}(u) = \Phi_{\Sigma}^n(\Phi^{-1}(u_1), \Phi^{-1}(u_2), \dots, \Phi^{-1}(u_n)) \quad (1)$$

โดย $C_{\Sigma}^{Ga}(u)$ คือ ฟังก์ชันการแจกแจงร่วมของเกาส์เซียนคอปูลา $U = (U_1, U_2, \dots, U_n)$ ที่มีเมทริกซ์สหสัมพันธ์เชิงเส้น Σ

เมื่อค่า $U = (U_1, U_2, \dots, U_n)$ เท่ากับ $u = (u_1, u_2, \dots, u_n)$

Φ_{Σ}^n คือ ฟังก์ชันการแจกแจงสะสมของตัวแปรสุ่มที่มีการแจกแจงแบบปกติมาตรฐาน n ตัวแปร ด้วยเมทริกซ์สหสัมพันธ์เชิงเส้น Σ

และ Φ^{-1} คือ ตัวผกผันของฟังก์ชันการแจกแจงสะสมของตัวแปรสุ่มที่มีการแจกแจงแบบปกติ 1 ตัวแปร (Univariate standard normal distribution)

เทคนิคเกาส์เซียนคอปพูลา [4] เป็นเทคนิคที่นิยมเป็นอย่างมาก สำหรับการจำลองตัวแปรสุ่มร่วมเมื่อทราบการแจกแจงส่วนริมและสหสัมพันธ์ เนื่องจากเป็นเทคนิคที่มีขั้นตอนในการทำที่ง่าย จึงได้มีการนำเทคนิคนี้มาใช้งานกันอย่างแพร่หลายกว่าเทคนิคอื่นๆ

2.2 ตัวแบบโลจิตและตัวแบบโพรบิต

ตัวแบบโลจิตและตัวแบบโพรบิต [5] เป็นตัวแบบที่ตัวแปรตามมีลักษณะเป็นตัวแปรเชิงคุณภาพ และมีค่าได้เพียง 2 ค่า โดยมีรูปแบบพื้นฐานมาจากแบบจำลองลาเทนท์ (Latent model) [6] โดยกำหนดให้ Y^* เป็นตัวแปรลาเทนท์ มีรูปแบบดังนี้

$$Y_i^* = X_i' \beta_i + e_i \quad (2)$$

ในทางปฏิบัติตัวแปร Y^* ไม่สามารถเก็บข้อมูลหรือสังเกตค่าได้ แต่สามารถเก็บรวบรวมข้อมูลได้เฉพาะค่าที่เกิดขึ้น ซึ่งต้องนำมาปรับให้เป็นตัวแปรหุ่น (Dummy Variable) แทนด้วยค่า Y โดยที่ $Y=1$ ถ้า $Y^* > 0$ และ $Y=0$ ถ้า $Y^* \leq 0$ และจากสมการ (1) และ (2) ข้างต้นจะได้ว่า

$$p_i = P(Y=1) = P(e_i > -X_i' \beta_i) = 1 - F(-X_i' \beta_i) = F(X_i' \beta_i) \quad (3)$$

โดยที่ F คือ ฟังก์ชันการแจกแจงสะสม (Cumulative distribution function) ของค่าความคลาดเคลื่อน (e_i) ส่วนค่า Y ที่เก็บข้อมูลมาไว้จะมีการแจกแจงแบบเบอร์นูลลี (Bernoulli distribution) ที่มีรูปแบบสมการฟังก์ชันภาวะน่าจะเป็น (Likelihood function) ดังนี้

$$L = \prod_i F(X_i' \beta_i)^{y_i} [1 - F(X_i' \beta_i)]^{1-y_i} \quad (4)$$

รูปแบบของสมการ (4) จะขึ้นอยู่กับสมมติฐานที่กำหนดไว้สำหรับการแจกแจงของค่าความคลาดเคลื่อน กล่าวคือ ถ้าการแจกแจงเป็นแบบโลจิสติกแบบจำลองจะเป็นตัวแบบโลจิต ถ้าการแจกแจงของค่าความคลาดเคลื่อนเป็นแบบปกติ (Normal distribution) แบบจำลองจะเป็นตัวแบบโพรบิต

สำหรับในกรณีของตัวแบบ โลจิตมีรูปแบบสมการความน่าจะเป็น ดังนี้

$$\logit(p_i) = \ln\left(\frac{p_i}{1-p_i}\right) = X_i' \beta_i \quad (5)$$

โดยที่

$$p_i = F(X_i' \beta_i) = \frac{1}{1 + \exp(-X_i' \beta_i)} \quad (6)$$

สำหรับในกรณีของตัวแบบโพรบิตมีรูปแบบสมการความน่าจะเป็น ดังนี้

$$\text{probit}(p_i) = \Phi^{-1}(p_i) = \eta_i = X_i' \beta_i \quad (7)$$

โดยที่

$$p_i = \Phi(X_i' \beta_i) = \int_{-\infty}^{\eta_i} \left[\frac{1}{(2\pi)^{1/2}} \right] \exp\left[-\frac{t^2}{2}\right] dt \quad (8)$$

การแจกแจงของตัวแบบโลจิตและตัวแบบโพรบิตมีความใกล้เคียงกันมาก จึงทำให้ผลในการวิเคราะห์ข้อมูลที่ได้จากตัวแบบทั้งสองไม่แตกต่างกันเท่าใดนัก นอกจากขนาดตัวอย่างที่มากพอจึงจะทำให้เห็นความแตกต่างที่ช่วงปลายของการแจกแจง โดยการแจกแจงจากตัวแบบโลจิตจะแบนราบกว่าตัวแบบโพรบิต

2.3 ตัวแบบความถดถอยโลจิสติกแบบเกาส์เซียนคอปพูลา

ตัวแบบความถดถอยโลจิสติกแบบเกาส์เซียนคอปพูลา เป็นตัวแบบ โลจิสติกที่มีความสัมพันธ์แบบคอปพูลาเข้ามาเกี่ยวข้องในตัวแบบ กล่าวคือเป็นตัวแบบที่ตัวแปรตามในแต่ละค่าสังเกตมีความสัมพันธ์กันด้วยเกาส์เซียนคอปพูลา ซึ่งเป็นการสัมพันธ์ผ่านการแจกแจงร่วมของควอนไทล์ของของตัวแปรสุ่มร่วมที่มีการแจกแจงแบบปกติหลายตัวแปร โดยที่มีวัตถุประสงค์เช่นเดียวกันกับตัวแบบความถดถอยโลจิสติกแบบ

ปกติ และในงานวิจัยนี้ได้สนใจในกรณีที่ตัวแปรตาม (Y) มีค่าได้เพียง 2 ค่า คือ 0 กับ 1 ดังนั้น $Y_i \sim \text{Ber}(p)$ ด้วยความน่าจะเป็น

$$P(Y_i = y_i) = p_i^{y_i} (1 - p_i)^{1-y_i}; \quad y_i = 0, 1 \quad (9)$$

โดยที่ p_i มาจากตัวแบบโพรบิต

$$p_i = P(Y = 1) = \Phi(X_i' \beta_i) \quad (10)$$

และเกาส์เซียนคอปพูลา $U = (U_1, U_2, \dots, U_n)$ ดังในสมการ (1) ดังนี้

$$C_\Sigma^{Ga}(u) = \Phi_\Sigma^G(\Phi^{-1}(u_1), \Phi^{-1}(u_2), \dots, \Phi^{-1}(u_n))$$

ดังนั้น รูปแบบของสมการฟังก์ชันภาวะน่าจะเป็นของตัวแบบความถดถอยโลจิสติกแบบเกาส์เซียนคอปพูลา เป็นดังนี้

$$L = \int \int_{\min(y_i, 1-p_i)}^{\max(1-p_i, y_i)} \dots \int C_\rho^{Ga}(u_1, u_2, \dots, u_n) du_1 du_2 \dots du_n \quad (11)$$

จะเห็นว่าสมการ (11) มีรูปแบบที่ค่อนข้างซับซ้อน ซึ่งการวิเคราะห์ภาวะความน่าจะเป็นสูงสุด (Maximum likelihood) ทำได้ยาก จึงเพิ่มสมมติฐานบางประการเพื่อให้สามารถวิเคราะห์ภาวะความน่าจะเป็นของตัวแบบได้

จากความสัมพันธ์แบบเกาส์เซียนคอปพูลา ดังแสดงไว้ในสมการ (1) นั้น เราสามารถเขียนตัวแบบเกาส์เซียนคอปพูลา โดยวิธีของตัวแบบปัจจัยเชิงเดี่ยว (one factor model) ดังนี้

$$e_i = \sqrt{\rho} Z + \sqrt{1-\rho} \varepsilon_i \quad (12)$$

และตัวแปรเกาส์เซียนคอปพูลา

$$U = \Phi(e_i) \quad (13)$$

เมื่อ e_i แทน เกาส์เซียนคอปพูลาด้วยสหสัมพันธ์ ρ

Z แทน ค่าปัจจัยของคอปพูลา โดยที่ $Z \sim N(0, 1)$

และ ε_i แทน ความคลาดเคลื่อนของค่าสังเกตที่ i

โดยที่ $\varepsilon_i \sim N(0, 1)$

และจากแบบจำลองโพรบิต ในสมการ (10)

$$p_i = \Phi(X_i' \beta_i)$$

ดังนั้น

$$\eta_i = \Phi^{-1}(p_i) = X_i' \beta \quad (14)$$

หากกำหนดเงื่อนไขให้ค่าปัจจัยของคอปพูลา Z เป็นตัวแปรที่ทราบค่า จะพบว่าโอกาสที่จะเกิดเหตุการณ์เมื่อทราบค่า Z เป็นดังนี้

$$\begin{aligned} P(Y_i = 1 | z) &= P(e_i \leq \Phi^{-1}(p_i)) \\ &= P(\sqrt{\rho} z + \sqrt{1-\rho} \varepsilon_i \leq \Phi^{-1}(p_i)) \\ &= P\left(\varepsilon_i \leq \frac{\Phi^{-1}(p_i) - \sqrt{\rho} z}{\sqrt{1-\rho}}\right) \\ &= P\left(\varepsilon_i \leq \frac{X_i' \beta_i - \sqrt{\rho} z}{\sqrt{1-\rho}}\right) \\ &= \Phi\left(\frac{X_i' \beta_i - \sqrt{\rho} z}{\sqrt{1-\rho}}\right) \end{aligned} \quad (15)$$

จากสมการ (15) จะเห็นว่า เมื่อกำหนดให้ค่าปัจจัยของคอปพูลา Z เป็นตัวแปรที่ทราบค่า จะส่งผลให้เกิดความเป็นอิสระกันของข้อมูล และพบว่าตัวแบบคอปพูลานี้ ก็คือตัวแบบโพรบิตนั่นเองสำหรับเหตุผลในการนำตัวแบบโพรบิตมาใช้เนื่องจากมีรูปแบบที่สอดคล้องกันกับเกาส์เซียนคอปพูลามากกว่าตัวแบบโลจิส ซึ่งทำให้ง่ายต่อการแก้สมการหาตัวประมาณค่าพารามิเตอร์ และจากที่ได้กล่าวมาแล้ว ตัวแบบทั้งสองนี้ให้ผลการศึกษาที่ใกล้เคียงกันจึงสามารถใช้แทนกันได้ นั่นคือ เราสามารถประมาณค่าพารามิเตอร์จากตัวแบบความถดถอยโลจิสติกแบบเกาส์เซียนคอปพูลาได้ด้วยการวิเคราะห์ความถดถอยโพรบิต เมื่อเราทราบค่าปัจจัยของคอปพูลา Z ในการศึกษา นอกจากนี้ค่าประมาณจากการวิเคราะห์ความถดถอยแบบโพรบิตที่ได้ ควรมีการนำมาปรับค่าด้วย $\sqrt{1-\rho}$ จึงจะทำให้ได้ตัวประมาณค่าพารามิเตอร์ที่เหมาะสมสำหรับตัวแบบนี้

3. วิธีการดำเนินงานวิจัย

สำหรับในการดำเนินงานวิจัยครั้งนี้ได้ทำการจำลองข้อมูลเพื่อใช้ในการเปรียบเทียบการประมาณค่าพารามิเตอร์ของตัวแบบความถดถอยโลจิสติกแบบเกาส์เซียนคอปพูลาจากการวิเคราะห์ความถดถอยแบบโพรบิตที่ได้แสดงไว้ในหัวข้อ 2.3

โดยได้ทำการเปรียบเทียบความถูกต้องในการประมาณค่าพารามิเตอร์ระหว่างกรณีไม่ทราบค่าปัจจัยของคอปพูลา Z และกรณีที่ทราบค่าปัจจัยของคอปพูลา Z ทั้งมีการปรับค่าตัวประมาณด้วย $\sqrt{1-\rho}$ และไม่ปรับค่าตัวประมาณ ซึ่งได้กำหนดสถานการณ์การจำลองในกรณีของตัวแปรที่มีตัวแปรอิสระจำนวน 1 ตัวแปร ที่มีการแจกแจงแบบเบอร์นูลลี ด้วยพารามิเตอร์เท่ากับ 0.5 ตัวแปรตามมีการแจกแจงแบบคอปพูลาเบอร์นูลลี ที่ระดับความสัมพันธ์ $\rho = 0, \rho = 0.2, \rho = 0.5$ และ $\rho = 0.8$ โดยกำหนดค่าพารามิเตอร์ของข้อมูลจำลอง คือ $\beta_1 = 1$ และกำหนด β_0 เพื่อให้ความน่าจะเป็นในการเกิดเหตุการณ์เป็น 0.5 ซึ่งสามารถคำนวณค่า $\beta_0 = -0.47$ นอกจากนี้ได้กำหนดจำนวนกลุ่มตัวอย่างในการทดลอง (m) เป็น 1, 5 และ 10 กลุ่ม จำนวนขนาดตัวอย่าง (n) เป็น 100, 500 และ 1,000 และเกณฑ์ในการเปรียบเทียบจะพิจารณาจากค่าเฉลี่ยความคลาดเคลื่อนกำลังสอง (Mean Square Error: MSE) ประกอบกับค่าเฉลี่ยและความคลาดเคลื่อนมาตรฐาน ในการทดลองซ้ำจำนวน 100 รอบ

4. ผลการวิจัย

จากการประมาณค่าพารามิเตอร์ในตัวแบบด้วยความถดถอยพหุคูณ ได้ผลการทดลองโดยแยกเป็นกรณีศึกษา 4 กรณี ได้แก่ กรณีไม่ทราบค่าของปัจจัยคอปพูลา Z กรณีไม่ทราบค่าของปัจจัยคอปพูลา Z และทำการปรับค่าด้วย $\sqrt{1-\rho}$ กรณีทราบค่าของปัจจัยคอปพูลา Z และกรณีทราบค่าของปัจจัยคอปพูลา Z และทำการปรับค่าด้วย $\sqrt{1-\rho}$ ได้ผลการทดลอง ดังนี้

4.1 กรณีไม่ทราบค่าปัจจัยของคอปพูลา Z

สำหรับผลการวิจัยในกรณีนี้ พบว่า ในระดับความสัมพันธ์ที่สูงขึ้นส่งผลให้ค่าประมาณพารามิเตอร์สูงกว่าค่าพารามิเตอร์ที่ควรจะเป็น (Overestimate) อีกทั้งการกระจายของค่าประมาณพารามิเตอร์ในการทดลอง 100 รอบ ก่อนข้างมาก เมื่อพิจารณา ค่าเฉลี่ยความคลาดเคลื่อนกำลังสอง (MSE) ดังแสดงไว้ในตารางที่ 1 จะเห็นว่าที่ระดับ $\rho = 0$ ค่าเฉลี่ยความคลาดเคลื่อนกำลังสองมีค่าต่ำสุดและเพิ่มขึ้นตามระดับความสัมพันธ์ที่สูงขึ้นในทุกกลุ่มขนาดตัวอย่าง โดยเฉพาะในกลุ่มข้อมูลตัวอย่าง 1

กลุ่มจะเห็นว่าค่าเฉลี่ยความคลาดเคลื่อนกำลังสองมีค่าสูงขึ้นค่อนข้างมากในแต่ละระดับความสัมพันธ์ที่สูงขึ้น และเมื่อจำนวนกลุ่มตัวอย่าง (m) และขนาดตัวอย่าง (n) ในการทดลองเพิ่มขึ้นส่งผลให้ค่าเฉลี่ยความคลาดเคลื่อนกำลังสองมีค่าลดลงในทุกระดับความสัมพันธ์ จากผลการทดลองในกรณีนี้แสดงให้เห็นว่าการประมาณค่าพารามิเตอร์โดยอาศัยแบบจำลองพหุคูณยังไม่เหมาะสม ซึ่งทำให้ได้ค่าประมาณสูงกว่าค่าพารามิเตอร์ที่ควรจะเป็น

สำหรับค่าเฉลี่ยความคลาดเคลื่อนกำลังสอง (MSE) ในกรณีนี้ ได้แสดงผลดังตารางต่อไปนี้

ตารางที่ 1 ค่าเฉลี่ยความคลาดเคลื่อนกำลังสองในกรณีไม่ทราบค่าปัจจัยของคอปพูลา Z

n	m	ค่าเฉลี่ยความคลาดเคลื่อนกำลังสอง (MSE)			
		$\rho = 0$	$\rho = 0.2$	$\rho = 0.5$	$\rho = 0.8$
100	1	0.0697	0.6131	3.2550	11.4260
	5	0.0103	0.0273	0.0719	0.1768
	10	0.0050	0.0147	0.0380	0.0865
500	1	0.0479	0.3942	1.3490	8.0767
	5	0.0024	0.0310	0.0881	0.2302
	10	0.0011	0.0140	0.0400	0.0879
1000	1	0.0098	0.1540	1.1070	7.0681
	5	0.0010	0.0198	0.0634	0.1736
	10	0.0004	0.0105	0.0327	0.0796

4.2 กรณีไม่ทราบค่าปัจจัยของคอปพูลา Z และทำการปรับค่าด้วย $\sqrt{1-\rho}$

สำหรับในกรณีนี้ พบว่า ในระดับความสัมพันธ์ที่สูงขึ้นกรณีจำนวนกลุ่มตัวอย่าง 1 กลุ่ม มีค่าประมาณพารามิเตอร์สูงกว่าค่าพารามิเตอร์ที่ควรจะเป็น (Overestimate) โดยที่จำนวนกลุ่มตัวอย่าง 5 และ 10 กลุ่ม มีค่าประมาณพารามิเตอร์ต่ำกว่าค่าพารามิเตอร์ที่ควรจะเป็น (Underestimate) อีกทั้งการกระจายของค่าประมาณพารามิเตอร์ในการทดลอง 100 รอบ

ค่อนข้างมาก เมื่อพิจารณาค่าเฉลี่ยความคลาดเคลื่อนกำลังสอง (MSE) ได้ผลดังตารางต่อไปนี้

ตารางที่ 2 ค่าเฉลี่ยความคลาดเคลื่อนกำลังสองในกรณีไม่ทราบค่าปัจจัยของคอปพูลา Z และปรับค่าตัวประมาณด้วย $\sqrt{1-\rho}$

n	m	ค่าเฉลี่ยความคลาดเคลื่อนกำลังสอง			
		$\rho = 0$	$\rho = 0.2$	$\rho = 0.5$	$\rho = 0.8$
100	1	0.0700	0.4724	1.4509	1.7109
	5	0.0103	0.0258	0.0777	0.1995
	10	0.0050	0.0162	0.0614	0.1848
500	1	0.0479	0.2991	0.5742	1.1685
	5	0.0024	0.0291	0.0818	0.2001
	10	0.0012	0.0167	0.0656	0.1914
1000	1	0.0098	0.1160	0.4718	0.9965
	5	0.0010	0.0205	0.0731	0.1960
	10	0.0005	0.0148	0.0655	0.1948

จากตารางข้างต้นจะเห็นว่าที่ระดับ $\rho = 0$ ค่าเฉลี่ยความคลาดเคลื่อนกำลังสองมีค่าต่ำสุดและเพิ่มขึ้นตามระดับความสัมพันธ์สูงขึ้นไปในทุกกลุ่มขนาดตัวอย่าง เมื่อจำนวนกลุ่มตัวอย่าง และขนาดตัวอย่าง ในการทดลองเพิ่มขึ้นส่งผลให้ค่าเฉลี่ยความคลาดเคลื่อนกำลังสองมีค่าลดลงในทุกระดับความสัมพันธ์ และค่าเฉลี่ยความคลาดเคลื่อนกำลังสองในกรณี 4.2 นี้ จะเห็นว่ามีค่าต่ำกว่าในกรณี 4.1 โดยเฉพาะในระดับที่มีความสัมพันธ์สูง ซึ่งกรณีนี้เป็นการปรับค่าตัวประมาณจากกรณี 4.1 จึงทำให้ความคลาดเคลื่อนลดลงเนื่องจากค่าประมาณในกรณี 4.1 มีค่าประมาณพารามิเตอร์สูงกว่าค่าพารามิเตอร์ที่ควรจะเป็น อย่างไรก็ตามผลการประมาณค่าพารามิเตอร์ในกรณีนี้ยังไม่เหมาะสม ซึ่งยังได้ค่าประมาณต่ำกว่าค่าพารามิเตอร์ที่ควรจะเป็น

4.3 กรณีทราบค่าปัจจัยของคอปพูลา Z

สำหรับกรณีนี้ พบว่าในระดับความสัมพันธ์ที่สูงขึ้นส่งผลให้ค่าประมาณพารามิเตอร์สูงกว่าค่าพารามิเตอร์ที่ควรจะเป็น

ค่อนข้างมากส่วนค่าความคลาดเคลื่อนมาตรฐานพบว่ามีความสูงมากนักเมื่อเทียบในกรณี 4.1 และกรณี 4.2 สำหรับค่าเฉลี่ยความคลาดเคลื่อนกำลังสอง ได้ผลดังตารางต่อไปนี้

ตารางที่ 3 ค่าเฉลี่ยความคลาดเคลื่อนกำลังสองในกรณีทราบค่าปัจจัยของคอปพูลา Z

n	m	ค่าเฉลี่ยความคลาดเคลื่อนกำลังสอง (MSE)			
		$\rho = 0$	$\rho = 0.2$	$\rho = 0.5$	$\rho = 0.8$
100	1	0.0697	0.6131	3.2551	11.4260
	5	0.0109	0.0245	0.1428	1.1036
	10	0.0050	0.0149	0.1223	1.0606
500	1	0.0479	0.3942	1.3491	8.0767
	5	0.0025	0.0121	0.1118	0.9385
	10	0.0012	0.0102	0.1077	0.9373
1000	1	0.0098	0.1540	1.1070	7.0681
	5	0.0011	0.0101	0.1088	0.9736
	10	0.0005	0.0089	0.1058	0.9339

จากตารางค่าเฉลี่ยความคลาดเคลื่อนกำลังสองข้างต้น พบว่าค่าเฉลี่ยความคลาดเคลื่อนกำลังสองมีค่าเพิ่มขึ้นตามระดับความสัมพันธ์สูงขึ้นไปในทุกกลุ่มขนาดตัวอย่าง โดยเฉพาะในกลุ่มข้อมูลตัวอย่าง 1 กลุ่มจะเห็นว่าค่าเฉลี่ยความคลาดเคลื่อนกำลังสองมีค่าสูงขึ้นค่อนข้างมากในแต่ละระดับความสัมพันธ์ที่สูงขึ้น และเมื่อจำนวนกลุ่มตัวอย่าง (m) และขนาดตัวอย่าง (n) ในการทดลองเพิ่มขึ้นส่งผลให้ค่าเฉลี่ยความคลาดเคลื่อนกำลังสองมีค่าลดลงในทุกระดับความสัมพันธ์

4.4 กรณีไม่ทราบค่าปัจจัยของคอปพูลา Z และการปรับค่าด้วย $\sqrt{1-\rho}$

สำหรับกรณีนี้ พบว่าเมื่อจำนวนกลุ่มตัวอย่างเพิ่มขึ้นส่งผลให้ในทุกระดับความสัมพันธ์มีค่าเฉลี่ยของค่าประมาณ และค่าความคลาดเคลื่อนมาตรฐานค่อนข้างใกล้เคียงกันมาก อีกทั้งค่าเฉลี่ยของค่าประมาณพารามิเตอร์ยังใกล้เคียงกับค่าพารามิเตอร์มากในทุกระดับความสัมพันธ์ และมีการกระจาย

ของข้อมูลค่าประมาณพารามิเตอร์ค่อนข้างน้อย ยกเว้นกรณีจำนวนกลุ่มข้อมูล 1 กลุ่ม ที่ให้ค่าประมาณพารามิเตอร์ที่สูงกว่าค่าพารามิเตอร์ โดยเฉพาะในระดับความสัมพันธ์ที่สูงขึ้น

สำหรับค่าเฉลี่ยความคลาดเคลื่อนกำลังสองในกรณีนี้ ได้ผลดังตารางต่อไปนี้

ตารางที่ 4 ค่าเฉลี่ยความคลาดเคลื่อนกำลังสองในกรณีทราบค่าปัจจัยของคอปพูลา Z และปรับค่าตัวประมาณด้วย $\sqrt{1-\rho}$

n	m	ค่าเฉลี่ยความคลาดเคลื่อนกำลังสอง			
		$\rho = 0$	$\rho = 0.2$	$\rho = 0.5$	$\rho = 0.8$
100	1	0.0697	0.4724	1.4507	1.7109
	5	0.0109	0.0102	0.0111	0.0106
	10	0.0050	0.0043	0.0051	0.0057
500	1	0.0479	0.2991	0.5742	1.1685
	5	0.0025	0.0023	0.0017	0.0011
	10	0.0012	0.0010	0.0008	0.0008
1000	1	0.0098	0.1160	0.4718	0.9965
	5	0.0011	0.0010	0.0007	0.0024
	10	0.0005	0.0005	0.0004	0.0004

จากตารางค่าเฉลี่ยความคลาดเคลื่อนกำลังสองข้างต้น พบว่าค่าเฉลี่ยความคลาดเคลื่อนกำลังสองมีค่าค่อนข้างใกล้เคียงกันในทุกระดับความสัมพันธ์ในทุกกลุ่มขนาดตัวอย่าง และมีค่าค่อนข้างน้อยเมื่อเทียบกับกรณีอื่นๆ ยกเว้นกลุ่มข้อมูลตัวอย่าง 1 กลุ่ม มีค่าเฉลี่ยความคลาดเคลื่อนกำลังสองค่อนข้างมากเมื่อเทียบกับจำนวนตัวอย่างอื่นๆ และเมื่อจำนวนกลุ่มตัวอย่าง (m) และขนาดตัวอย่าง (n) ในการทดลองเพิ่มขึ้นส่งผลให้ค่าเฉลี่ยความคลาดเคลื่อนกำลังสองมีค่าลดลงในทุกระดับความสัมพันธ์ด้วย จากผลการทดลองนี้แสดงว่า การประมาณค่าโดยอาศัยแบบจำลองโพธิท ในกรณีทราบค่าปัจจัยของคอปพูลา Z และปรับค่าตัวประมาณด้วย $\sqrt{1-\rho}$ มีความเหมาะสมสำหรับใช้ในการประมาณค่าพารามิเตอร์ในตัวแบบนี้ ซึ่งสอดคล้องกับทฤษฎีในหัวข้อ 2.3

5. สรุปผล

จากการศึกษาและประมาณค่าพารามิเตอร์ในตัวแบบความถดถอยโลจิสติกแบบเกาส์เซียนคอปพูลา โดยวิธีการประมาณค่าพารามิเตอร์สำหรับกรณีที่ทราบค่าปัจจัยของคอปพูลา Z ในแต่ละการศึกษานั้น พบว่าตัวแบบคอปพูลาเป็นคือตัวแบบโพธิท ดังนั้นจึงใช้วิธีการวิเคราะห์ความถดถอยโพธิทในการหาค่าตัวประมาณพารามิเตอร์ได้ โดยตัวประมาณที่ได้ควรนำมาปรับค่าด้วย $\sqrt{1-\rho}$ จึงทำให้ได้ตัวประมาณที่เหมาะสมสำหรับตัวแบบนี้ในทุกระดับความสัมพันธ์ ที่จำนวนกลุ่มทดลองมากกว่า 1 กลุ่ม นอกจากนี้ในการประมาณค่าพารามิเตอร์ในตัวแบบความถดถอยโลจิสติกแบบเกาส์เซียนคอปพูลาสำหรับงานวิจัยครั้งนี้ ใช้วิธีการวิเคราะห์ความถดถอยโพธิทในการหาค่าตัวประมาณพารามิเตอร์สามารถทำได้เฉพาะในกรณีที่ทราบค่าปัจจัยของคอปพูลา Z และปรับค่าด้วย $\sqrt{1-\rho}$ เท่านั้น ซึ่งเป็นที่น่าสนใจที่จะทำการประมาณค่าด้วยวิธีอื่น ที่สามารถครอบคลุมกรณีไม่ทราบค่าปัจจัยของคอปพูลา Z ต่อไป

เอกสารอ้างอิง

- 1 S. Tirapat and S. Kiatsupaibul, "Credit value at risk via credit scoring model" Simulation Society Research Workshop, 2007, pp. 1-4.
- 2 B. R. Nelsen, An Introduction to Copulas, Springer, Second Edition, 1999.
- 3 E. Luciano, Umberto Cherubini and Walter Vecchiato, Copula Method in Finance, John Wiley & Sons, Inc., 2004.
- 4 จิตติมา จิรเศรษฐศิริ, การจำลองตัวแปรสุ่มร่วมด้วยเทคนิคคอปพูลาเมื่อทราบการแจกแจงส่วนริมและสหสัมพันธ์. วิทยานิพนธ์ปริญญาโทบริหารธุรกิจ, จุฬาลงกรณ์มหาวิทยาลัย, 2548.
- 5 A. D. Powers, Yu Xie, Statistical Methods for Categorical Data Analysis, Academic Press, USA, 2000.
- 6 ชาติรี จันทรโคติกา, "แบบจำลองตัวแปรไม่ต่อเนื่อง" เอกสารประกอบการบรรยายเศรษฐมิติ 2, 2542, หน้า 16-24.