ผนวก ข

การวิเคราะห์โดยใช้แบบจำลอง Logit

วิทยานิพนธ์นี้ได้ใช้แบบจำลอง Logit (ธาตรี จันทรโคลิกา, 2542, น. 16-24) เป็นเครื่องมือหนึ่งในการวิเคราะห์ถึงปัจจัยที่มีผลต่อความน่าจะเป็นในการมีหนี้ ซึ่งรายละเอียดต่างๆ รวมทั้งวิธีการประมาณค่าของแบบจำลอง Logit มีดังนี้

ลักษณะของแบบจำลอง Logit

แบบจำลองต่างๆที่นิยมนำมาใช้ในการนำมาใช้วิเคราะห์ข้อมูลที่มีลักษณะของตัวแปรตามเป็นเชิงคุณภาพคือแบบจำลอง Linear Probability แบบจำลอง Probit และแบบจำลอง Logit แต่แบบจำลอง Linear Probability ยังมีข้อบกพร่องในการนำมาวิเคราะห์คือค่าความน่าจะเป็นที่ประมาณได้อาจจะมีค่าอยู่นอกช่วง 0 ถึง 1 ได้ซึ่งขัดแย้งกับหลักความน่าจะเป็น ขณะที่แบบจำลอง Logit และแบบจำลอง Probit มีลักษณะที่คล้ายคลึงกันมากแตกต่างกันเพียงลักษณะของตัวแปรเชิงสุ่มที่ในกรณีแบบจำลอง Protbit มีการกระจายแบบปกติ แต่กรณีของแบบจำลอง Logit มีการกระจายแบบ Logistic ซึ่งเป็นการแจกแจงทั้งสองมีความใกล้เคียงกันมาก จะแตกต่างกันก็เพียงค่าช่วงหางของการแจกแจงเท่านั้น ผลการประมาณค่าแบบจำลอง Probit และ Logit จะให้ผลที่ไม่แตกต่างกันเท่าใดนัก โดยที่แบบจำลอง Logit มีรูปแบบเริ่มต้นดังนี้

กำหนดให้ y* เป็นตัวแปรพฤติกรรมตอบรับ (Response Variable) ซึ่งมีความสัมพันธ์ดังนี้

$$y_i^* = \beta' x_i + \mu_i \qquad \text{.............................(ข.1)}$$

ในทางปฏิบัติตัวแปร y* จะไม่สามารถเก็บข้อมูลได้ (เช่น โอกาสในการมีหนี้) แต่เราจะสามารถเก็บข้อมูลได้เฉพาะค่าที่เกิดขึ้นจริง (เช่น การมีหนี้ที่เกิดขึ้นจริง) ซึ่งต้องนำมาปรับให้เป็นตัวแปรหุ่น (Dummy Variable) โดยที่

y = 1 ถ้า $y^* > 0$

y = 0 ถ้า $y^* \leq 0$ ....(ข.2)

ในกรณีนี้ $\beta' x_i$ จะไม่ใช่ $E(y_i / x_i)$ เหมือนกรณีของแบบจำลองความน่าจะเป็นเชิงเส้นตรง (Linear probability model) แต่เป็นแบบ $E(y_i^* / x_i)$ จากสมการ (ข.1) และ (ข.2) จะได้

$$\begin{aligned} \mathrm{Prob}(y=1) &= \mathrm{Prob}(y^* > 0) \\ &= \mathrm{Prob}(\beta' x_i + \mu_i > 0) \\ &= \mathrm{Prob}(\mu_i > -\beta' x_i) \\ &= 1 - F(-\beta' x_i) \end{aligned} \qquad \text{.............................(ข.3)}$$

โดยที่ F คือ ฟังก์ชั่นการแจกจงสะสมของ u (Cumulative distribution function)

<u>การประมาณค่าแบบจำลอง Logit</u>

ในกรณีค่า y ที่เก็บข้อมูลมาได้จะมีการแจกแจงแบบทวิลักษณ์ (Binomial) ซึ่งค่าความน่าจะเป็นจะถูกกำหนดจากสมการ (ข.3) โดยแปรผันตามค่าของตัวแปรอธิบาย (x_i) ที่จะมีรูปแบบของสมการ Likelihood function (Green, 2000, p. 811-895) ซึ่งวิธีที่ใช้ในการประมาณค่าแบบจำลอง Logit ที่เหมาะสมที่สุดคือวิธี Maximum likelihood ซึ่งแนวคิดในการประมาณค่าของวิธีนี้คือการประมาณโดยหาค่าพารามิเตอร์ที่ทำให้ค่าความน่าจะเป็นร่วมของข้อมูลสูงที่สุด รูปแบบของสมการ Likelihood function คือ

$$L = \prod_{y_i=0} F(-\beta' x_i) \prod_{y_i=1} \left[1 - F(-\beta' x_i)\right] \qquad \text{.......................(ข.4)}$$

จากสมการ (ข.4) เมื่อ u_i มีการแจกจงแบบ Logistic ในแบบจำลอง Logit รูปแบบของสมการความน่าจะเป็นจะมีลักษณะดังนี้

$$F(-\beta' x_i) \quad = \frac{\exp(-\beta' x_i)}{1+\exp(-\beta' x_i)} = \frac{1}{1+\exp(\beta' x_i)} \qquad \text{.......................(ข.5)}$$

ดังนั้น

$$-F(-\beta' x_i) \quad = \frac{\exp(\beta' x_i)}{1+\exp(\beta' x_i)} \qquad \text{.......................(ข.6)}$$

จาก

$$F(-\beta' x_i) \quad = \frac{1}{1+\exp(\beta' x_i)} \qquad \text{.......................(ก.7)}$$

จะได้

$$P_i = F(+\beta' x_i) \quad = \frac{1}{1+\exp(-\beta' x_i)} \qquad \text{.......................(ก.8)}$$

$$(1+\exp(-\beta' x_i))P_i = 1 \qquad \text{.......................(ข.9)}$$

$$\exp(-\beta' x_i) = \frac{1-P_i}{P_i} \qquad \text{.....................(ข.10)}$$

ถอด Natural logarithms ของสมการ (ข.10) จะได้

$$\beta' x_i = \log(\frac{P_i}{1-P_i}) \qquad \text{.....................(ข.11)}$$

$$\log(\frac{P_i}{1-P_i}) = \beta' x_i \qquad \text{.....................(ข.12)}$$

การประมาณค่าแบบจำลอง Logit จะเริ่มจากสมการ Likelihood function สมการ (ข.4)

$$L = \prod_{y_i=0} F(-\beta' x_i) \prod_{y_i=1} \left[1 - F(-\beta' x_i)\right] \qquad \text{.......................(ข.4)}$$

ซึ่งสามารถเขียนให้อยู่ในรูปของการแจกแจงแบบ Logistic ได้ดังนี้

$$L = \prod_{i=1}^{n} \left(\frac{1}{1+\exp(\beta' x_i)}\right)^{1-y_i} \left(\frac{\exp(\beta' x_i)}{1+\exp(\beta' x_i)}\right)^{y_i} \qquad \text{.....................(ข.13)}$$

$$= \frac{\exp(\beta') \sum_{i=1}^{n} x_i y_i}{\prod_{i=1}^{n} \left[1+\exp(\beta' x_i)\right]} \qquad \text{.....................(ข.14)}$$

กำหนดให้ $t^* = \sum_{i=1}^{n} x_i y_i$ จากนั้นทำการประมาณค่าเพื่อหาค่า Maximum-likelihood ของ β จาก

$$\log L = \beta' t^* - \sum_{i=1}^{n} \log\left[1+\exp(\beta' x_i)\right] \qquad \text{.....................(ข.15)}$$

โดยการกำหนดให้ค่า $\frac{\partial \log L}{\partial \beta} = 0$ จะได้

$$\frac{\partial \log L}{\partial \beta} = \sum_{i=1}^{n} \frac{\exp(\beta' x_i)}{1+\exp(\beta' x_i)} x_i + t^* = 0 \qquad \text{.....................(ข.16)}$$

จะสามารถหาค่า β โดยการหาค่า first order condition ของ Likelihood function

<u>การทดสอบสมมติฐานและพิจารณาความเหมาะสมของสมการ</u>

การทดสอบสมมติฐานแบ่งออกเป็นสองส่วนคือการทดสอบความมีนัยสำคัญของสัมประสิทธิ์

$$H_0 : \beta_i = 0$$
$$H_1 : \beta_i \neq 0$$

การทดสอบสมมติฐานข้างต้นนี้พิจารณาจากค่า Z-stat ที่ได้เปรียบเทียบกับค่า critical Z ณ ระดับความเชื่อมั่นต่างๆ หากค่า Z-stat มากกว่า critical Z เราจะปฏิเสธสมมติฐานที่ว่าค่าสัมประสิทธิ์เท่ากับศูนย์

การทดสอบสมมติฐานอีกส่วนเป็นการพิจารณาความน่าเชื่อถือของสมการโดยมีสมมติฐานว่า

$$H_0 : \beta_0 = \beta_1 = ... = \beta_n = 0$$
$$H_1 : \beta_0 \neq \beta_1 \neq ... \neq \beta_n \neq 0$$

การทดสอบสมมติฐานนี้จะพิจารณาจากค่า log-likelihood ratio (LR)

$$LR = 2[\ln L - \ln L_r]$$

เมื่อ

Restricted log-likelihood ($\ln L_r$) คือค่า log-likelihood กรณีที่สัมประสิทธิ์ทุกตัวเท่ากับศูนย์

Unrestricted log-likelihood ($\ln L$) คือค่า log-likelihood กรณีสมการปกติ

ค่า LR ที่ได้จะมีการแจกแจงแบบไคสแควร์และมี degree of freedom เท่ากับจำนวนตัวแปรอิสระของสมการ การทดสอบสมมติฐานจึงต้องทำการเปรียบเทียบค่า LR กับ ค่า critical Chi-square ณ ระดับความเชื่อมั่นต่างๆ หากค่า LR มากกว่า critical Chi-square เราจะปฏิเสธสมมติฐานที่ว่าค่าสัมประสิทธิ์ทุกตัวมีค่าเท่ากับศูนย์

ในส่วนของการพิจารณาความเหมาะสมของสมการนั้น ยังสามารถแสดงได้ในรูปของค่า R^2 แต่จากแบบจำลองที่ใช้มีลักษณะที่ไม่ได้เป็นเส้นตรง (nonlinear) ซึ่งประมาณค่าด้วยวิธีแบบ Maximum likelihood ดังนั้นจึงต้องทำการพิจารณาจากค่า Pseudo-R^2 แทน ซึ่งในหลายวิธีของการคำนวณ Pseudo-R^2 คือ McFadden's-R^2 สามารถคำนวณได้ดังนี้

$$Pseudo - R^2 = 1 - \frac{\ln L}{\ln L_o}$$

โดยที่

lnL คือ ค่า log-likelihood ของแบบจำลองปกติที่มีค่าคงที่ และสัมประสิทธิ์อย่างน้อยหนึ่งตัวในสมการ

$\ln L_o$ คือ ค่า log-likelihood ของแบบจำลองที่มีเพียงค่าคงที่เท่านั้น