Estimación del Probit Ordinal y del Logit Multinomial

Microeconomía Cuantitativa

R. Mora

Departmento de Economía
Universidad Carlos III de Madrid

Esquema

Introducción

El modelo probit ordinal con tres posibles resultados

$$U = x'\beta + \varepsilon, \varepsilon \sim N(0,1)$$

$$y = 0 \text{ if } U \leq \alpha_1$$
$$y = 1 \text{ if } \alpha_1 < U \leq \alpha_2$$
$$y = 2 \text{ if } \alpha_2 < U$$

- No observamos U, pero sí la elección de cada individuo entre los tres posibles resultados
- Esta elección se representa por y, que es una variable ordinal (es decir, no tiene sentido cardinal)
- El objetivo es estimar β, α_1, y α_2

Logit Multinomial con tres alternativas

$$\Pr(y = 0|x) = 1 - \Pr(y = 1|x) - \Pr(y = 2|x)$$
$$\Pr(y = 1|x) = \frac{\exp(x'\beta_1)}{1 + \exp(x'\beta_1) + \exp(x'\beta_2)}$$
$$\Pr(y = 2|x) = \frac{\exp(x'\beta_2)}{1 + \exp(x'\beta_1) + \exp(x'\beta_2)}$$

- La elección se representa mediante la variable y, que es una variable cualitativa (es decir, no tiene ni sentido cardinal ni ordinal)
- El objetivo es obtener estimaciones para β_1 y β_2

Estimación del modelo de respuesta ordinal

Ejemplo: No-trabaja, tiempo-parcial y tiempo-completo

h^*: horas deseadas de trabajo semanal (en decenas)

$$h^* = -0.5 + 0.07 * educ - 1.0 * kids + \varepsilon, \varepsilon \sim N(0,1)$$

- Si hay solo dos tipos de contratos: a tiempo parcial y a tiempo completo
  - los contratos a tiempo parcial tienen jornadas de como mucho 20 horas a la semana
  - los contratos a tiempo completo tienen jornadas de más de 20 horas a la semana
- En nuestros datos, no observamos h^*sino:
  - $y = 0$ para todos los individuos que no trabajan. ($h^* \leq 0$).
  - $y = 1$ para aquellos que trabajan a tiempo parcial. ($0 < h^* \leq 2$).
  - $y = 2$ para aquellos que trabajan a tiempo completo. ($2 < h^*$).

Normalización

- Sea $y^* = 0.07 * educ - 1.0 * kids + \varepsilon$ de forma que $h^* = -0.5 + y^*$
- No trabajar implica que

$$-0.5 + y^* \leq 0 \Rightarrow y^* \leq 0.5$$
$$(\alpha_1 = 0.5)$$

- Trabajar a tiempo completo implica que

$$-0.5 + y^* > 2 \Rightarrow y^* > 2.5$$

El modelo sin constante

El modelo puede ser normalizado

$$y^* = 0.07 * educ - 1.5 * kids + \varepsilon,\ \varepsilon \sim N(0,1)$$

$$y = 0 \text{ si } y^* \leq 0.5$$
$$y = 1 \text{ si } 0.5 < y^* \leq 2.5$$
$$y = 2 \text{ si } 2.5 < y^*$$

Probabilidades de los extremos

$$
\begin{aligned}
\Pr\left(y=0|x\right) &= \Pr\left(x'\beta+\varepsilon\leq\alpha_1|x\right)\\
&= \Phi\left(-\left(x'\beta-\alpha_1\right)\right)\\
&= 1-\Phi\left(\left(x'\beta-\alpha_1\right)\right)
\end{aligned}
$$

$$
\begin{aligned}
\Pr\left(y=2|x\right) &= \Pr\left(\alpha_2<x'\beta+\varepsilon|x\right)\\
&= \Pr\left(\varepsilon>-\left(x'\beta-\alpha_2\right)|x\right)\\
&= \Phi\left(\left(x'\beta-\alpha_2\right)\right)
\end{aligned}
$$

(Téngase en cuenta que $\Phi\left(a\right)=1-\Phi\left(-a\right)$, porque la distribución normal es simétrica)

Probabilidades intermedias

- In el ejemplo con 3 alternativas solo hay una probabilidad intermedia:

$$\begin{aligned}\Pr(y=1|x) &= \Pr\left(\alpha_1 < x'\beta + \varepsilon \leq \alpha_2|x\right)\\ &= \Pr\left(\varepsilon > -\left(x'\beta-\alpha_1\right), \varepsilon \leq -\left(x'\beta-\alpha_2\right)|x\right)\end{aligned}$$

- Puesto que $-(x'\beta-\alpha_1) < -(x'\beta-\alpha_2)$:

$$\begin{aligned}\Pr\left(\varepsilon \leq -\left(x'\beta-\alpha_2\right)|x\right) - \Pr\left(\varepsilon < -\left(x'\beta-\alpha_1\right)|x\right)\\ = \Phi\left(-\left(x'\beta-\alpha_2\right)\right) - \Phi\left(-\left(x'\beta-\alpha_1\right)\right)\end{aligned}$$

$$\Pr(y=1|x) = \Phi(x'\beta-\alpha_1) - \Phi(x'\beta-\alpha_2)$$

La esperanza condicional y MCO

$$\Pr(y=0|x) = 1-\Phi\left(\left(x'\beta-\alpha_1\right)\right)$$
$$+\Pr(y=1|x) = \Phi\left(x'\beta-\alpha_1\right)-\Phi\left(x'\beta-\alpha_2\right)$$
$$+\Pr(y=2|x) = \Phi\left(\left(x'\beta-\alpha_2\right)\right)$$

....

$$\sum \Pr(y=j|x) = 1 \text{ (Las probabilidades suman uno)}$$

- En general, MCO no funciona porque la esperanza condicional de la variable dependiente no es lineal:

$$\begin{aligned} E(y|x) &= 0\times\Pr(y=0|x)+1\times\Pr(y=1|x)+2\times\Pr(y=2|x) \\ &= \Pr(y=1|x)+2\times\Pr(y=2|x) \\ &= \Phi\left(x'\beta-\alpha_1\right)+\Phi\left(x'\beta-\alpha_2\right) \end{aligned}$$

Estimación MV

La probabilidad de una observación es

$$\begin{aligned}\Pr(y|x) = &\left(1-\Phi\left(x'\beta-\alpha_1\right)\right)^{1(y=0)}\times\\ &\left(\Phi\left(x'\beta-\alpha_1\right)-\Phi\left(x'\beta-\alpha_2\right)\right)^{1(y=1)}\times\\ &\left(\Phi\left(x'\beta-\alpha_2\right)\right)^{1(y=2)}\end{aligned}$$

Por ello, para N observaciones la verosimilitud es:

$$\begin{aligned}\mathrm{L}(b) = \prod_{i=1}^{N}\Big\{&\left(1-\Phi\left(x_i'b-\alpha_1\right)\right)^{1(y_i=0)}\times\\ &\left(\Phi\left(x_i'b-\alpha_1\right)-\Phi\left(x_i'b-\alpha_2\right)\right)^{1(y_i=1)}\times\\ &\left(\Phi\left(x_i'b-\alpha_2\right)\right)^{1(y_i=2)}\Big\}\end{aligned}$$

Efectos marginales

Opciones para presentar los resultados:

- Cuando la ecuación de la variable latente tiene una interpretación sencilla (como en el ejemplo), presentarla es una buena manera de informar sobre el modelo.
- Por otra parte, los efectos marginales de las probabilidades de cada una de las categorías se pueden calcular.
  - Cuando la variable independiente es discreta, efectos marginales se pueden calcular como en el caso binario.
- Por último, se puede estimar el efecto sobre el valor esperado de la variable observada.

Efectos marginales cuando el regressor es continuo

Efectos marginales cuando el regresor es continuo

$$\frac{\partial \mathtt{Pr}\left(y=0|x\right)}{\partial x_j} = -\phi\left(x'\beta-\alpha_1\right)\beta_j$$

$$\frac{\partial \mathtt{Pr}\left(y=1|x\right)}{\partial x_j} = \left(\phi\left(x'\beta-\alpha_1\right)-\phi\left(x'\beta-\alpha_2\right)\right)\beta_j$$

$$\frac{\partial \mathtt{Pr}\left(y=2|x\right)}{\partial x_j} = \phi\left(x'\beta-\alpha_2\right)\beta_j$$

El modelo Logit multinomial

Modelo de Utilidad Aleatoria

- Suponiendo que hay tres alternativas de transporte: bus, coche, tren:

$$U_b = x_b'\beta_b + \varepsilon_b$$
$$U_c = x_c'\beta_c + \varepsilon_c$$
$$U_t = x_t'\beta_t + \varepsilon_t$$

  donde $\{\varepsilon_b, \varepsilon_c, \varepsilon_t\}$ son los efectos en la utiliad inobservados por el econometra
- Sea $y = 0$ si bus, $y = 1$ si coche, and $y = 2$ si tren.
  - y no tiene sentido ni cardinal ni ordinal!

Re-parametrización del modelo

$$\varepsilon_{01} \equiv \varepsilon_b - \varepsilon_c, \qquad \varepsilon_{02} \equiv \varepsilon_b - \varepsilon_t$$
$$x'\beta_{01} \equiv x_b'\beta_b - x_c'\beta_c, \qquad x'\beta_{02} \equiv x_b'\beta_b - x_t'\beta_t$$

- Supuesto: $\{\varepsilon_{01}, \varepsilon_{02}\} \sim F$, donde F es simétrica.

$$\Pr(y=0|x) = F\left(x'\beta_{01}, x'\beta_{02}\right)$$
$$\Pr(y=1|x) = F\left(-x'\beta_{01}, x'\left(\beta_{02}-\beta_{01}\right)\right)$$
$$\Pr(y=2|x) = F\left(-x'\beta_{02}, -x'\left(\beta_{02}-\beta_{01}\right)\right)$$

El Logit Multinomial

Cuando el vector $\{\varepsilon_b, \varepsilon_c, \varepsilon_t\}$ tiene una distribución de valor extremo, entonces tenemos el **Logit Multinomial**:

$$\Pr(y = 0|x) = 1 - \Pr(y = 1|x) - \Pr(y = 2|x)$$
$$\Pr(y = 1|x) = \frac{\exp(x'\beta_1)}{1 + \exp(x'\beta_1) + \exp(x'\beta_2)}$$
$$\Pr(y = 2|x) = \frac{\exp(x'\beta_2)}{1 + \exp(x'\beta_1) + \exp(x'\beta_2)}$$

- OLS no funciona porque la esperanza condicionada no es lineal
- La estimación MV proporciona estimadores consistentes y asintóticamente normales
  - En `gretl`, el comando `logit` estima el Logit multinomial cuando la variable dependiente es discreta y no binaria si se usa la opción `--multinomial`

Ejemplo: Coche, bicicleta, y tren

$$\Pr(car|x) = 1 - \Pr(bicycle|x) - \Pr(train|x)$$
$$\Pr(bicycle|x) = \frac{\exp(x'\beta_1)}{1+\exp(x'\beta_1)+\exp(x'\beta_2)}$$
$$\Pr(train|x) = \frac{\exp(x'\beta_2)}{1+\exp(x'\beta_1)+\exp(x'\beta_2)}$$

donde

$$x'\beta_1 = 1 - 0.1 * age - .1 * income - 3 * kids + 5 * center$$
$$x'\beta_2 = 1 - .1 * income + 1.5 * kids + 2 * center$$

Observa que $\beta_1 = \beta_{bicycle} - \beta_{car}$ y que $\beta_2 = \beta_{train} - \beta_{car}$.
¿Cuál es la valoración relativa entre *train* y *bicycle*?

$$x'\beta_1 = 1 - 0.1 * age - .1 * income - 3 * kids + 5 * center$$
$$x'\beta_2 = 1 - .1 * income + 1.5 * kids + 2 * center$$

- $\beta_{1,age} = -.1 < 0$: A medida que las personas envejecen, tiene más valor usar el coche que usar la bicicleta.

- $\beta_{1,income} = \beta_{2,income} = -.1 < 0$: Cuanto mayor sea el ingreso, más probable es el coche en vez del tren o la bicicleta.

- $\beta_{1,kids} = -3, \beta_{2,kids} = 1.5$: entonces $\beta_{car,kids} - \beta_{train,kids} = -1.5 < 0$ y $\beta_{bicycle,kids} - \beta_{train,kids} = -3 - 1.5 = -4.5 < 0$: Cuantos más niños, más probable es el tren frente al coche y la bicicleta.

- Si el viaje pasa por el centro de la ciudad, el tren y la bicicleta son más probables que el coche (¿y cómo se valora el tren

logit transport const age income kids center --multinomial

Model 3: Multinomial Logit, using observations 1–5000
Dependent variable: transport, Standard errors based on Hessian

	Coefficient	Std. Error	z	p-value
transport = 2				
const	0.522376	0.501274	1.0421	0.2974
age	−0.0969857	0.00890363	−10.8928	0.0000
income	−0.0957303	0.00589953	−16.2268	0.0000
kids	−2.78035	0.222090	−12.5190	0.0000
center	5.29508	0.422875	12.5216	0.0000
transport = 3				
const	0.867895	0.197401	4.3966	0.0000
age	0.00590025	0.00495462	1.1909	0.2337
income	−0.100911	0.00351216	−28.7320	0.0000
kids	1.42980	0.0869553	16.4429	0.0000
center	1.89417	0.0899837	21.0502	0.0000

Mean dependent var	2.253800	S.D. dependent var	0.915616
Log-likelihood	−2687.761	Akaike criterion	5395.521
Schwarz criterion	5460.693	Hannan–Quinn	5418.363

Number of cases ‘correctly predicted’ = 3787 (75.7 percent)
Likelihood ratio test: $\chi^2(8) = 3712.464$ [0.0000]

Resumen

- No podemos estimar por MCO ni el probit ordinal ni el logit multinomial.
- La estimación MV proporciona estimadores que son consistentes y asintóticamente normales.
- Los efectos marginales se pueden calcular como en el caso binario.