El modelo Ordinal y el modelo Multinomial

Microeconomía Cuantitativa

R. Mora

Departmento de Economía
Universidad Carlos III de Madrid

Esquema

Motivación

- Consideramos las siguientes extensiones de los modelos de variable dependiente binaria:
    - Modelos de respuesta ordinal: La variable dependiente toma un número de valores **finitos** y **discretos** que contienen **información ordinal**.
    - Modelos de respuesta multinomial: La variable dependiente toma una serie de valores **finitos** y **discretos** que **NO** contienen **información ordinal**.

- Como en los casos probit y logit, la variable dependiente no es estrictamente continua. La estimación se lleva a cabo utilizando el estimador MV.

Ejemplos de modelos ordinales

- Calificación de crédito, utilizando siete categorías, desde "definitivamente no digno de crédito" a "digno de crédito".
- Decisión de permanecer inactivo, trabajar a tiempo parcial, o trabajar a tiempo completo
- En una regresión de ingresos, el nivel de ingresos se codifica en intervalos: [0,1000), [1000,1500) [1500,2000), [2000, ∞)
- En valoraciones sobre afirmaciones, varias respuestas con contenido ordinal: "completamente en desacuerdo", "en desacuerdo", "algo de acuerdo", "completamente de acuerdo"

Ejemplos de modelos multinomiales

- Elección del modo de transporte: tren, autobús, coche.
- Situación económica: inactivos, desempleados, trabajadores por cuenta propia, empleado.
- Elección campo Educación: ciencia dura, ciencias de la salud, ciencias sociales, humanidades.

Modelos de respuesta ordinal

- Los dos modelos estándar son el **probit ordenado** y el **logit ordenado**.
- El enfoque es equivalente: simplemente usamos para el probit ordenado la CDF normal Φ y para el logit ordenado la CDF logística Λ .
- OLS no funciona porque la variable dependiente no tiene sentido cardinal:
  - Solvencia: 0,1,2,3,4,5 : El cambio de 0 a 1 no tiene que ser "equivalente" al cambio de 4 a 5 .
  - Actividad: inactivo = 0, a tiempo parcial = 1, a tiempo completo = 2: Mientras inactiva es cero horas de trabajo, el código 1 refleja entre 1 y (generalmente) 30 horas de trabajo, y el código 2 refleja más 30 horas de trabajo. Esto implica que no hay proporcionalidad en ir de 0 a 1 y de 1 a 2.

Simplificación

- Los modelos de elección binaria (MLP, probit, logit) podrían ser utilizados agrupando todas las categorías en dos grandes grupos.
- Esta puede ser una solución razonable cuando la muestra es pequeña y las categorías ordinales pueden lógicamente agruparse en dos categorías principales.
- En algunos casos, esta es probablemente una muy mala idea (por ejemplo, con intervalos de ingresos)

- Consideremos tres resultados observados: $y = 0, 1, 2$.
- Consideremos el modelo de variable latente sin constante:

$$y^* = \beta_1 x_1 + ... + \beta_k x_k + \varepsilon$$

donde $\varepsilon \sim \mathcal{N}(0, 1)$.

- Definimos dos puntos de corte: $\alpha_1 < \alpha_2$
- No observamos y^*, pero observamos valores discretos de acuerdo con la siguiente regla

$$y = 0 \text{ si } y^* \leq \alpha_1$$
$$y = 1 \text{ si } \alpha_1 < y^* \leq \alpha_2$$
$$y = 2 \text{ si } \alpha_2 < y^*$$

Ejemplo: actividad

- $y = 0$ si inactivo, $y = 1$ si empleado a tiempo parcial, $y = 2$ si empleado a tiempo completo
- $y^* = \beta_e \times educ + \beta_k \times kids + \varepsilon$, donde $\varepsilon \sim \mathcal{N}(0,1)$
- Entonces

$$\begin{aligned} y &= 0 \text{ si } \beta_e \times educ + \beta_k \times kids + \varepsilon \leq \alpha_1 \\ y &= 1 \text{ si } \alpha_1 < \beta_e \times educ + \beta_k \times kids + \varepsilon \leq \alpha_2 \\ y &= 2 \text{ si } \alpha_2 < \beta_e \times educ + \beta_k \times kids + \varepsilon \end{aligned}$$

- Nótese que podríamos alternativamente incluir una constante β_0 y suponer que $\alpha_1 = 0$.

Interpretación

- Al igual que en otros modelos no lineales, los efectos marginales se deben calcular para entender los efectos parciales de un pequeño cambio en la variable explicativa x_j.

- Para los modelos ordinales podemos calcular los efectos marginales sobre las probabilidades predichas usando los mismos principios que con los probit y logit.

Efectos marginales sobre las probabilidades

- Para los modelos de elección binaria nos hemos centrado en los efectos sobre la probabilidad de que y es igual a uno.
- En los modelos ordinales, las cosas no son tan simples porque ahora tenemos más de dos resultados:

$$\frac{\partial \Pr(y=0|x)}{\partial x_j}, \frac{\partial \Pr(y=1|x)}{\partial x_j}, \frac{\partial \Pr(y=2|x)}{\partial x_j}$$

- Si x_j es discreto se calcula como en el caso binario el cambio discreto en las probabilidades predichas asociados con el cambio x_j .

Efectos Marginales

- El efecto parcial de x_j en la probabilidad predicha de
  - el más alto resultado tiene el mismo signo que β_j.
  - el resultado más bajo tiene el signo opuesto a β_j.
  - los resultados intermedios no se pueden deducir del signo de β_j.
- El último resultado se debe a dos efectos compensatorios. Supongamos que $\beta_j > 0$ y aumentamos x_j. La categoría intermedia
  - Puede ser más probable, ya que la probabilidad de la categoría más baja cae.
  - También puede ser menos probable debido a que la probabilidad de la categoría más alta aumenta.
- Por lo general, los efectos parciales de probabilidades intermedias son cuantitativamente pequeños y, a menudo estadísticamente insignificante.

Discusión

¿Cuál es la mejor forma de interpretar los resultados de los modelos ordenados?

- Una opción es estudiar las estimaciones de los parámetros haciendo hincapié en la ecuación de la variable latente subyacente con la que empezamos.
- Otra opción es mirar el efecto sobre el valor esperado de la variable de respuesta ordinal, por ejemplo,

$$\frac{\partial \mathrm{E}(y|x)}{\partial x_j} = \frac{\partial \Pr(y=0|x)}{\partial x_j}\times 0 + \frac{\partial \Pr(y=1|x)}{\partial x_j}\times 1 + \frac{\partial \Pr(y=2|x)}{\partial x_j}\times 2$$

  Esto puede tener mucho sentido si y es una variable numérica, como en la variable de ingreso.
- Alternativamente, simplemente podemos reportar el efecto sobre la probabilidad de observar las categorías ordenadas.

Respuesta Multinomial

- La variable dependiente es tal que
  - Más de dos resultados son posibles.
  - Los resultados no tienen una ordenación natural.

- Una vez más, podríamos agrupar dos o más categorías y así construir una variable de resultado binario de los datos brutos; pero, al hacerlo, eliminamos información potencialmente interesante.

- OLS tampoco es un buen modelo en este contexto.

- Sin embargo, el modelo logit para la opción binaria se puede extender a modelar más de dos resultados.

Modelo de Utilidad Aleatoria

- Supongamos 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 utilidad no observados por el econometra.

Si $x_b'\beta_b + \varepsilon_b \geq \max \ \{x_c'\beta_c + \varepsilon_c, x_t'\beta_t + \varepsilon_t\}$ entonces $y = 0$

Si $x_c'\beta_c + \varepsilon_c > \max \ \{x_b'\beta_b + \varepsilon_b, x_t'\beta_t + \varepsilon_t\}$ entonces $y = 1$

Si $x_t'\beta_t + \varepsilon_t > \max \ \{x_c'\beta_c + \varepsilon_c, x_t'\beta_t + \varepsilon_t\}$ entonces $y = 2$

Notación

- Tenemos dos efectos inobservados netos independientes

$$\varepsilon_{01} = \varepsilon_b - \varepsilon_c$$
$$\varepsilon_{02} = \varepsilon_b - \varepsilon_t$$

  - nótese que $\varepsilon_{12} = \varepsilon_c - \varepsilon_t = \varepsilon_{02} - \varepsilon_{01}$

- Defina

$$x_b'\beta_b - x_c'\beta_c = x'\beta_{01}$$
$$x_b'\beta_b - x_t'\beta_t = x'\beta_{02}$$

Supuesto

$$\{\varepsilon_{01}, \varepsilon_{02}\} \sim F$$

donde F es simétrica.

Entonces

$$\begin{aligned}\Pr(y=0|x) &= \Pr\left(x'\beta_{01}+\varepsilon_{01}\geq 0, x'\beta_{02}+\varepsilon_{02}\geq 0|x\right)\\ &= \Pr\left(\varepsilon_{01}\geq -\left(x'\beta_{01}\right), \varepsilon_{02}\geq -\left(x'\beta_{02}\right)|x\right)\end{aligned}$$

Dada la simetría,

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

Logit Multinomial

- Debemos modelar la probabilidad de que un individuo pertenece a la categoría j condicional a tener características x:

$$\Pr(y = j|x)$$

- Cuando el vector $\{\varepsilon_b, \varepsilon_c, \varepsilon_t\}$ se distribuye conjuntamente con una distribucion de valor extremo tenemos el modelo Logit Multinomial:

$$\Pr(y = 0|x) = 1 - \Pr(y = 1|x) - \Pr(y = 2|x)$$

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

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

- La principal diferencia en comparación con el logit binario es que ahora hay dos vectores de parámetros, β_1 y β_2.
- En el caso general con J posibles respuestas, hay $J-1$ vectores de parámetros.
- Esto hace que la interpretación de los coeficientes sea más difícil que en los modelos de elección binaria.

Interpretación con tres alternativas

El caso más sencillo es cuando β_{1j} y β_{2j} tienen el mismo signo.

- Si β_{1j} y β_{2j} son positivos entonces un aumento en x_j hace menos probable que el individuo pertenece a la categoría 0 ...
- y $\Pr(y_i = 1|x_i) + \Pr(y_i = 2|x_i)$ aumenta
- para saber cómo se asigna este incremento total entre estas dos probabilidades, tenemos que mirar los efectos marginales: la derivada parcial es muy compleja y el efecto marginal $\frac{\partial \Pr(y=1|x)}{\partial x_j}$ de hecho, puede ser negativo incluso si $\beta_{1j} > 0$!

Independence of irrelevant alternatives (IIA)

- Una limitación del modelo logit es que la relación de la probabilidad de cualesquiera dos alternativas l y m depende sólo de los vectores de parámetro β_l y β_m y las variables explicativas x

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

- La inclusión o exclusión de otras categorías es irrelevante para la relación de las dos probabilidades.
- Este comportamiento se conoce como la "independencia de alternativas irrelevantes", y puede llevar a un comportamiento contrario a lo que sugeriría la intuición.

Ejemplo: IIA puede ser contra-intuitivo

- Las personas pueden ir al trabajo por tres medios de transporte: autobús azul, autobús rojo o tren.

- Las personas eligen una de estas alternativas; el econometra modeliza con un logit multinomial esta decisión, y obtiene una estimación de

$$\frac{\Pr(y = red|x)}{\Pr(y = train|x)} = \exp\left(x'\left(\beta_{red} - \beta_{train}\right)\right)$$

- Supongamos que la empresa de autobuses ahora elimina el autobús azul del conjunto de opciones, ¿cree que $\frac{\Pr(y=red|x)}{\Pr(y=train|x)}$ sería igual que antes?

Otros modelos multinomiales

- Hay muchos otros modelos econométricos que se pueden utilizar para modelar modelos de respuesta multinomial:
  - Probit multinomial
  - Logit condicional
  - Logit anidado

- Su estudio supera el objetivo del curso.

Resumen

- Cuando la variable dependiente tiene un número finito de valores discretos, podemos extender los modelos probit y logit
    - Cuando la variable dependiente contiene información ordinal, entonces podemos utilizar probit ordinal y logit ordinal.
    - Cuando la variable dependiente no contiene ninguna información ordinal, podemos utilizar modelos multinomiales. Un ejemplo es el logit multinomial.

- Estos son todos modelos no lineales, y todos ellos pueden ser estimados por MV.