Estimación Probit

Microeconomía Cuantitativa

R. Mora

Departmento de Economía
Universidad Carlos III de Madrid

Esquema

El modelo de utilidad aleatoria

- $U_m = \beta_m x_m + \varepsilon_m$
- $U_h = \beta_h x_h + \varepsilon_h$

$$\Updownarrow$$

$$\beta_m x_m + \varepsilon_m > \beta_h x_h + \varepsilon_h \Leftrightarrow work = 1$$

$$\Updownarrow$$

$$\beta x + \varepsilon > 0 \Leftrightarrow work = 1$$

donde $\varepsilon = \varepsilon_m - \varepsilon_h$ (la utilidad neta de participación inobservada) y $\beta x = \beta_m x_m - \beta_h x_h$ (la función indice)

El supuesto Probit

- El econometra solo observa *work*, x_m, y x_h

Supuesto Probit : $\varepsilon_h, \varepsilon_m \sim N(0, \Sigma)$

- $\varepsilon \equiv \varepsilon_m - \varepsilon_h | x \sim N\left(0, \sigma^2\right)$
- $Pr\left(work = 1\right) = Pr\left(\varepsilon > -\beta x\right) = Pr\left(\varepsilon \leq \beta x\right)$
- $Pr\left(work = 1\right) = Pr\left(\frac{\varepsilon}{\sigma} \leq \frac{\beta x}{\sigma}\right)$
- $Pr\left(work = 1\right) = \Phi\left(\frac{\beta}{\sigma} x\right)$
  donde Φ es la función de distribución acumulada de la normal estándar

Observabilidad de σ

β y σ son observacionalmente equivalentes a $\beta^* = k\beta$ y $\sigma^* = k\sigma$

$$\Phi\left(\frac{\beta^*}{\sigma^*}x\right) = \Phi\left(\frac{k\beta}{k\sigma}x\right) = \Phi\left(\frac{\beta}{\sigma}x\right), k \neq 0$$

- un número infinito de pares (β^*, σ^*) dan la misma verosimilitud
- las condiciones de identificación asintótica de MV no se cumplen

supuesto identificador: $\sigma = 1$ (por tanto $\varepsilon \backsim N(0,1)$)

- $Pr(work = 1) = \Phi(\beta x)$

Interpretación de las pendientes y los efectos marginales

cuando el control x_j aparece en ambas utilidades U_m y U_h...

- solo el efecto neto en la función índice, $\beta_{mj} - \beta_{hj}$, está identificado

el supuesto de normalidad (no-linealidad)

- la "pendiente neta" $\beta_{mj} - \beta_{hj}$ captura el efecto marginal sobre la función índice βx de un aumento unitario de x_j
- el efecto marginal sobre la probabilidad de participación es más complejo
- si x_j es continua, $\frac{\partial Pr(work=1)}{\partial x_j} = \phi(\beta x)\beta_j$
- si x_j es discreta, $\Delta Pr(work = 1) = \Phi(\beta x_1) - \Phi(\beta x_0)$ donde x_1 es el vector final de controles y x_0 es el vector inicial de controles

Un ejemplo simple

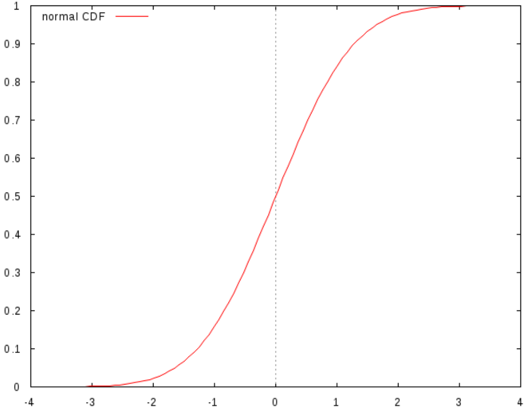
- $U_m = \beta_m^0 + \beta_m^e educ + \beta_m^k kids + \varepsilon_m$ con $\varepsilon_m \sim N\left(0, \sigma_m^2\right)$
- $U_h = \beta_h^0 + \beta_h^e educ + \beta_e^k kids + \varepsilon_h$ con $\varepsilon_h \sim N\left(0, \sigma_h^2\right)$
- $cov\left(\varepsilon_m, \varepsilon_h\right) = \sigma_{m,h}$

Supuesto Probit: $\varepsilon_m - \varepsilon_h | x \sim N(0,1)$

- $Pr\left(work = 1\right) = \Phi(\beta_0 + \beta_e educ + \beta_k kids)$
- $\beta_0 = \beta_m^0 - \beta_h^0$
- $\beta_e = \beta_m^e - \beta_h^e$
- $\beta_k = \beta_m^k - \beta_h^k$
- $var\left(\varepsilon_m - \varepsilon_h\right) = \sigma_m^2 + \sigma_h^2 - 2\sigma_{m,h} = 1$

Una interpretación gráfica

La probabilidad de participar es una función no-lineal de la función índice $\beta_0 + \beta_e educ + \beta_k kids$



La densidad

Supuesto: muestra iid

- sea β_0 el verdadero valor
- entonces, en el modelo Probit

$$Pr\left(work\,|x\right)=\begin{cases}\Phi\left(\beta_0 x\right) \text{ si } work=1\\ 1-\Phi\left(\beta_0 x\right) \text{ si } work=0\end{cases}$$

La verosimilitud de una observación

- la verosimilitud reemplaza en la densidad el verdadero vector β_0 por un valor cualquiera β
- entonces la verosimilitud para el individuo i es

$$L_i(\beta) = \begin{cases} \Phi(\beta x_i) \text{ si } work_i = 1 \\ 1 - \Phi(\beta x_i) \text{ si } work_i = 0 \end{cases}$$

- o, de forma más compacta,

$$L_i(\beta) = [\Phi(\beta x_i)]^{work_i} [1 - \Phi(\beta x_i)]^{1-work_i}$$

La log-verosimilitud

- primero, tomamos logaritmos

$$l_i(\beta) = work_i \log(\Phi(\beta x_i)) + (1 - work_i) \log(1 - \Phi(\beta x_i))$$

- luego juntamos la muestra entera

$$l(\beta) = \sum_{i=1}^{n} l_i(\beta)$$

- por tanto

$$l(\beta) = \sum_i \{work_i \log(\Phi(\beta x_i)) + (1 - work_i)) \log(1 - \Phi(\beta x_i))\}$$

Estimación MV

definición

- El estimador MV es el vector $\hat{\beta}^{ML}$ tal que

$$\hat{\beta}^{ML} = \arg\max_{\beta} l(\beta)$$

- debido a la naturaleza no-lineal de la maximización de la verosimilitud, no hay fórmulas explícitas para los estimadores MV del modelo Probit
- en vez de expresiones algebraicas, se utilizan algoritmos númericos de maximización, necesitándose normalmente solo unas pocas iteraciones
- en `gretl`, se utiliza un algoritmo cuasi-Newton (el BFGS)

Un Control que Clasifica Perfectamente

- supón que la dummy D_i predice perfectamente $work_i$ en la muestra en el sentido que $work_i = 1 \Leftrightarrow D_i = 1$
- si $\beta x = \beta_0 + \beta_D D$, entonces $\beta x = \left\{ \begin{array}{ll} \beta_0 + \beta_D & \text{si } work = 1 \\ \beta_0 & \text{si } work = 0 \end{array} \right|$
- y la log-versosimilitud será creciente en β_D:

$$l(\beta) = \sum_i \{work_i \log(\Phi(\beta_0 + \beta_D)) + (1 - work_i)) \log(1 - \Phi(\beta_0))\}$$

- por tanto, no puede haber un estimador MV

El problema de predición perfecta

- en general, supón que $\widetilde{\beta}$ predice perfectamente $work_i$ en la muestra en el sentido de que para un escalar k, $\widetilde{\beta}x_i > k$ si y solo si $work_i = 1$
- entonces lo mismo ocurre para cualquier múltiplo de $\widetilde{\beta}$ y la log-verosimilitud no tendrá un máximo
- ésto se puede deber a varias razones
  - un control puede ser un clasificador perfecto: (hay que quitarlo)
  - el modelo puede estar trivialmente mal especificado (como predecir participación entre trabajadores)
  - la muestra puede simplemente no ser suficientemente grande

Propiedades asintóticas y contrastes

bajo condiciones generales, el estimador MV es consistente, asintóticamente normal, y asintóticamente eficiente

- podemos construir contrastes asintóticos tipo t e intervalos de confianza (como hacemos en MCO, MC2E, y VI)
- restriciones de exclusión (por ejemplo, $H_0 : \beta_j = 0 \text{ and } \beta_k = 0$)
  - el multiplicador de Lagrange solo requiere la estimación del modelo bajo la nula
  - el test de Wald solo requiere la estimación del modelo sin restringir
  - la razón de verosimilitudes (LR) requiere la estimación de ambos modelos

El contraste de razón de verosimilitudes

Hipótesis anidadas

- se basa en la diferencia de las log-verosimilitudes bajo la hipótesis nula y la alternativa
- restringir un modelo no puede generar mayores log-verosimilitudes

$$LR = 2\left(l_{ur}\left(\hat{\beta}_{ur}^{ML}\right) - l_{r}\left(\hat{\beta}_{r}^{ML}\right)\right) \xrightarrow{a} \chi_q$$

donde q es el número de restricciones

Resumen

- no todos los parámetros del modelo de utilidad aleatoria pueden ser estimados
- el modelo Probit identifica cómo cada control afecta la probabilidad de participación
- la estimación MV del modelo Probit requiere el uso de algoritmos numéricos de maximización
- bajo condiciones generales los estimadores MV son consistentes y asintóticamente normales y eficientes
- constrastes de significativad y contrastes de restricciones generales son fáciles de implementar en el modelo Probit