Modelos de respuesta binaria

Gustavo A. García

ggarci24@eafit.edu.co

Econometría avanzada II

PhD/Maestría en Economía

Universidad EAFIT

Link slides en formato html

Link slides en formato PDF

En este tema

- Motivación

- Un ejemplo de modelos binarios: elección del modo de pesca

- Efectos marginales

- Ejercicio aplicado en R

Lecturas

- Cameron, A.C., Fotheringham, A., Trivedi, P.K. (2005). *Microeconometrics: Methods and Applications*. Cambridge University Press. Cap. 15

Motivación

- Los modelos de elección discreta o respuesta cualitativa son modelos que tienen como variable dependiente un indicador entre m alternativas mutuamente excluyentes y por lo regular no existe un orden natural de las categorias de la variable dependiente

- El caso más simple de respuesta cualitativa, es el de variable dependiente binaria, es decir, que existe que existe sólo dos resultados $(0 \text{ y } 1)$. Algunos ejemplos:
    - si un individuo está o no empleado
    - si un comprador hace una compra o no
    - si se viaje en transporte público o privado

- Resultados binarios son simples de modelar y su estimación usualmente es por máxima verosimilitud, ya que la distribución de los datos es necesariamente definida por el modelo Bernoulli $\Longrightarrow$ si la probabilidad de un resultado es igual a p, entonces la probabilidad del otro resultado debe ser $(1-p)$

- Para aplicaciones en regresiones, la probabilidad p variará entre individuos como una función de los regresores

- Los dos modelos estándar de variable dependiente binaria son los modelos logit y probit, los cuales especifican diferentes formas funcionales para las probabilidades como una función de los regresores

Ejemplo de un resultado binario: elección del modo de pesca

Supongamos que queremos modelar la elección de pescar desde un bote alquilado o desde un muelle. La variable dependiente es binaria con la siguiente estructura:

$$y_i = \begin{cases} 1 \text{ si se pesca desde un bote privado} \\ 0 \text{ si se pesca desde un muelle} \end{cases}$$

Asumamos que la única variable explicativa es $ln(relp_i)$, donde $relp$ es el precio de pescar en bote relativo al precio de pescar desde un muelle:

$$ln(relp_i) = ln\left(\frac{precio_{bote,i}}{precio_{muelle,i}}\right)$$

El precio de pescar desde el bote o el muelle varía entre individuos debido a diferentes factores, como por ejemplo, a diferencias en acceso

Se espera que la probabilidad de pescar desde un bote disminuya cuando el precio relativo aumente

Ejemplo de un resultado binario: elección del modo de pesca

Para estimar esta relación, se tienen las siguientes opciones:

- **OLS**: una regresión por OLS de y_i sobre $ln(relp_i)$. Sin embargo, esta relación lineal ignora:
  - la forma discreta de la variable dependiente
  - no limita las probabilidades predichas entre 0 y 1
  - el efecto de $ln(relp_i)$ sobre la probabilidad es constante sobre toda las distribución de dicha variable

- **Logit**:

$$p_i = Pr[y_i = 1|x_i] = \frac{exp(\beta_1 + \beta_2 x_i)}{1 + exp(\beta_1 + \beta_2 x_i)}$$

- **Probit**:

$$p_i = Pr[y_i = 1|x_i] = \Phi(\beta_1 + \beta_2 x_i)$$

  donde Φ es la FDA de la normal estándar

Estas funciones no linealies garantizan que:
- $0 < p_i < 1$
- el efecto marginal sea diferente sobre la distribución de la variable x_i:

Logit: $\frac{dp_i}{dx_i} = \frac{exp(\beta_1+\beta_2 x_i)}{(1+exp(\beta_1+\beta_2 x_i))^2}\beta_2$

Probit: $\frac{dp_i}{dx_i} = \phi(\beta_1 + \beta_2 x_i)\beta_2$

Ejemplo de un resultado binario: elección del modo de pesca

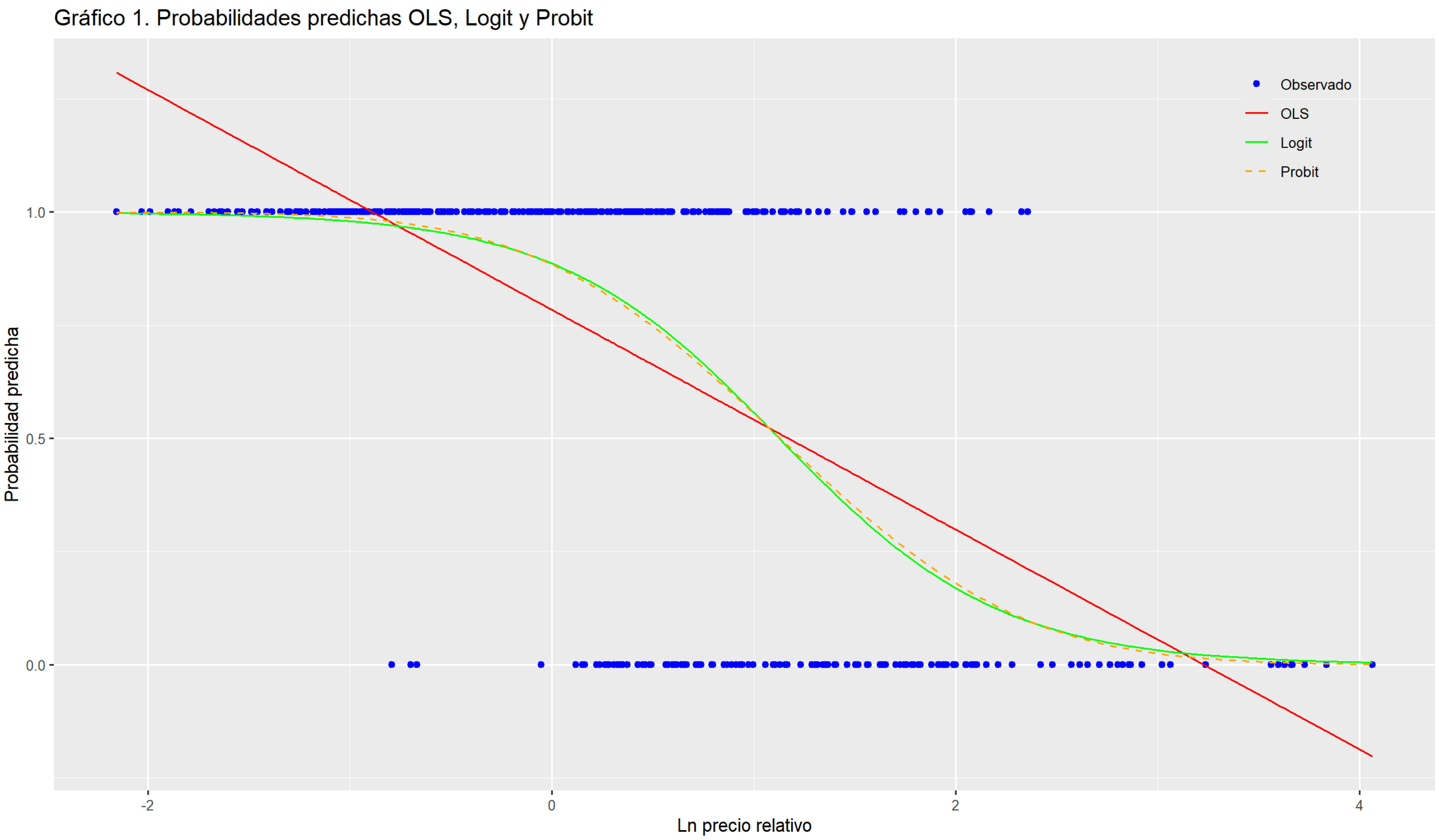
Tabla 1. Elección del modo de pesca
(1=bote; 0=muelle)

	OLS	Logit	Probit
ln relp	-0.243***	-1.823***	-1.056***
	(0.010)	(0.145)	(0.075)
Constante	0.784***	2.053***	1.194***
	(0.013)	(0.169)	(0.088)
Num.Obs.	630	630	630
R2	0.463		
R2 Adj.	0.462		
Log.Lik.	-195.167	-206.827	-204.411

* p < 0.1, ** p < 0.05, *** p < 0.01

Nota: Errores estándar en paréntesis

Ejemplo de un resultado binario: elección del modo de pesca



Efectos marginales

En términos generales el modelo de elección binaria es determinado por:

$$y_i = \begin{cases} 1 \text{ con probabilidad } p \\ 0 \text{ con probabilidad } 1-p \end{cases}$$

$$p_i = Pr[y_i = 1|\mathbf{x}_i] = F(\mathbf{x}_i^{'}\boldsymbol{\beta}),$$

donde $F(\cdot)$ es la función de distribución que puede ser logística (logit) o normal estándar (probit)

El mayor interés en estos modelos es determinar el efecto marginal de cambios en un regresor sobre la porbabilidad condicional que $y = 1$. Asumiendo un regresor continuo, el efecto marginal viene determinado por:

$$\frac{\partial Pr[y_i = 1|\mathbf{x}_i]}{\partial x_{ij}} = F^{'}(\mathbf{x}_i^{'}\boldsymbol{\beta})\beta_j,$$

donde $F^{'}(z) = \partial F(z)/\partial z$

Efectos marginales

Existen dos tipos de efectos marginales:

Efecto parcial en el promedio

$$\frac{\partial Pr[y_i = 1|\mathbf{x}_i]}{\partial x_{ij}} = F^{'}(\overline{\mathbf{x}}_i^{'}\hat{\boldsymbol{\beta}})\hat{\beta}_j$$

Efecto parcial promedio

$$\frac{\partial Pr[y_i = 1|\mathbf{x}_i]}{\partial x_{ij}} = \left[n^{-1}\sum_{i=1}^{n} F^{'}(\mathbf{x}_i\hat{\boldsymbol{\beta}})\right]\hat{\beta}_j$$

Importante: la interpretación de los efectos marginales se debe dar en término de puntos porcentuales y NO en términos de porcentajes

Por ejemplo, en la probabilidad de empleo, se encuentra un efecto marginal de la educación de 0.03, lo cual indica que un año de educación incrementa la probabilidad de estar empleado en 3 puntos porcentuales

Ejercicio aplicado en R

En este ejercicio se va a utilizar la base de datos *fishing* de Cameron y Trivedi (2005), sobre el modo de pesca

Archivos a descargar:

- Descripción de los datos
- Código