

Eleccion Binaria

Walter Sosa-Escudero

Econometria Avanzada

- Y variable aleatoria, toma solo dos valores, uno o cero, asociada a la ocurrencia de un evento (1 ocurre, 0 si no).
Ejemplo: admision a un posgrado, asistir al secundario, etc.
- x vector de K variables explicativas.

Modelo de elección binaria: modelo de la probabilidad de ocurrencia del evento denotado por Y condicional en x :

$$p \equiv Pr(Y = 1|x)$$

Ejemplo: probabilidad de que a una persona la admitan a un posgrado, dadas sus calificaciones, de que un banco quiebre dada su situación financiera, etc.

Importante: dado x , y toma solo los valores (0-1): y condicional en x tiene distribucion de *Bernoulli*.

Entonces:

$$E(y|x) = 1 p + 0 (1 - p) = p$$

y

$$V(y|x) = p (1 - p)$$

El modelo lineal de probabilidad

Un primer modelo podría ser:

$$y = x'\beta + u \quad \text{con} \quad E[u|x] = 0$$

Este es el *modelo lineal de probabilidad (LPM)*. Los parámetros pueden ser estimados consistentemente por MCO, regresando y en x .

Tres razones para descartar LPM

- 1) *Genera predicciones inconsistentes con una probabilidad:*
Notar que:

$$E(y|x) = x'\beta$$

Problema: en nuestro caso $E(y|x) = p$ de modo que $0 \leq E(y|x) = p \leq 1$. Pero en el modelo lineal general $-\infty < x'\hat{\beta} < \infty$: **el LPM puede generar predicciones cuyos valores caen fuera del rango admisible para una probabilidad.**

- 2) u es heterocedastico.

La especificacion lineal implica:

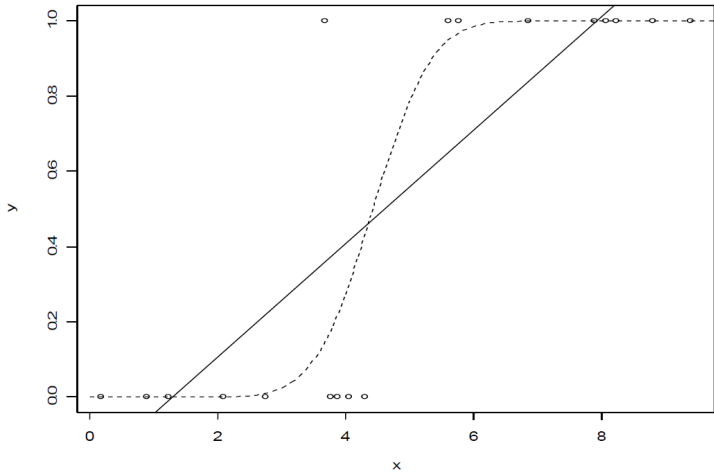
$$V(u_i|x_i) = V(y_i|x_i) = p_i(1 - p_i) = x_i'\beta(1 - x_i'\beta)$$

que, claramente, no es constante. Esta es una razon para abandonar MCO, es facil de arreglar.

- 3) *LPM implica derivadas parciales constantes:*

$$\frac{\partial p}{\partial x_k} = \beta_k$$

que es una constante. Pero, intuitivamente, en la mayoria de los casos relevantes no queremos esto. Efectos 'debiles en las puntas'. Ejemplos: asistencia al secundario, admision a posgrado.



El modelo *no-lineal* que propondremos es:

$$p = F(x'\beta)$$

$F(\cdot)$ tiene las siguientes propiedades:

$$F(-\infty) = 0, F(\infty) = 1, f(z) = dF(z)/dz > 0$$

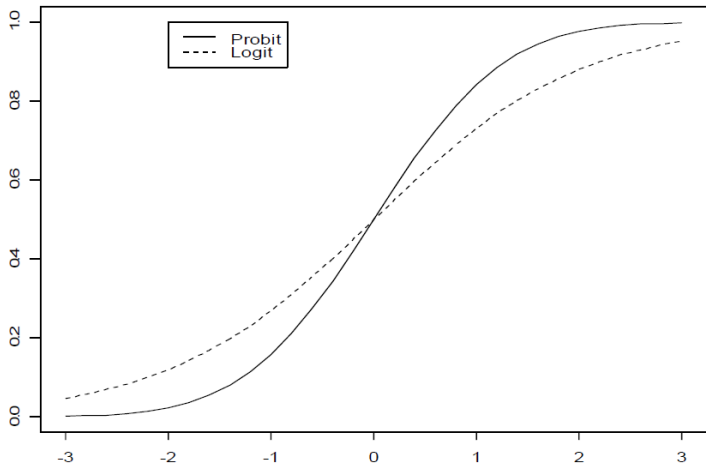
- Probit:

$$F(z) = \int_{-\infty}^z \frac{1}{\sqrt{2\pi}} e^{-\frac{s^2}{2}} ds$$

- Logit:

$$F(z) = \frac{e^z}{1 + e^z}$$

Notar que LMP corresponde a $F(z) = z$.



$$\frac{\partial p}{\partial x_k} = \beta_k f(x'_i \beta)$$

,que no es constante. Notar que:

$$\text{sgn}(\partial p / \partial x_k) = \text{sgn} \beta_k$$

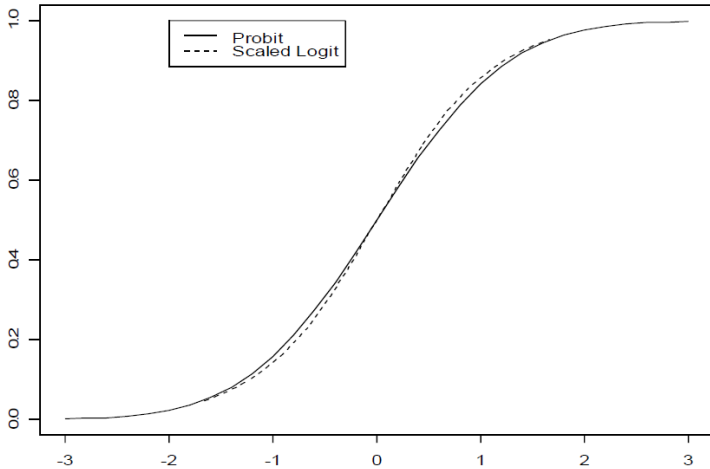
de modo que el signo de la derivada es interpretable, no su valor, la derivada depende de *donde* se la evalúe. Alternativa: derivadas en las medias:

$$\frac{\partial p}{\partial x_k}_{x=\bar{x}} = \beta_k f(\bar{x}' \beta_k),$$

el *efecto marginal en las medias*. Podríamos hacerlo en cualquier punto. O graficar esta derivada moviendo solo una variable y fijando las restantes en algún punto interesante.

Logits o probits?

- Si X es logistica, $V(X) = \pi^2/3$, entonces $V(Z = X\sqrt{3}/\pi) = 1$. Se puede mostrar que la distribución de Z (logistica re-escalada) es *muy* similar a la normal estandar.
- En general, los coeficientes del logit exceden a los de probit en $\pi/\sqrt{3}$.
- Utilizar cualquier especificacion.



$(y_i, x_i), i = 1, \dots, n$, muestra iid. Y_i tiene distribucion de Bernoulli con $p_i = Pr(y_i = 1)$

La funcion de verosimilitud sera:

$$L(\beta) = \prod_{y_i=1} p_i \prod_{y_i=0} (1 - p_i) = \prod_{i=1}^n p_i^{y_i} (1 - p_i)^{1-y_i}$$

y su logaritmo:

$$\begin{aligned} l(\beta) &= \sum_{i=1}^n [y_i \ln p_i + (1 - y_i) \ln(1 - p_i)] \\ &= \sum_{i=1}^n [y_i \ln F(x'_i \beta) + (1 - y_i) \ln(1 - F(x'_i \beta))] \end{aligned}$$

Las condiciones de primer orden son:

$$\sum_{i=1}^n \frac{(y_i - F_i) f_i x_{ki}}{F_i(1 - F_i)} = 0, \quad k = 1, \dots, K$$

con $F_i \equiv F(x_i' \hat{\beta})$, $f_i \equiv f(x_i' \hat{\beta})$.

- Es un sistema de K ecuaciones *no-lineales* con K incógnitas.
- No satisface las condiciones del teorema de la función implícita: es imposible 'despejar' $\hat{\beta}$.
- Existe una solución siempre que los regresores sean linealmente independientes y no haya un *clasificador perfecto*. Ver notas de clase.

Habiendo estimado los parametros por maxima verosimilitud:

$$\sqrt{n} \left(\hat{\beta}_{MV} - \beta_0 \right) \xrightarrow{d} N(0, V_{MV})$$

- *Significatividad individual*: $H_0 : \beta_k = \beta_{0k}$ vs. $H_A : \beta_k \neq \beta_{0k}$

$$t_k = \frac{\hat{\beta}_k - \beta_{k0}}{\sqrt{\hat{V}(\hat{\beta}_k)/n}} \xrightarrow{d} N(0, 1)$$

- *Significatividad de un grupo de coeficientes:* H_0 : r coeficientes del modelo son cero, H_A : alguno de ellos es $\neq 0$

$$LR = -2[\ln \hat{L}_r - \ln \hat{L}_{nr}] \sim \chi^2(r)$$

En donde \hat{L} y \hat{L}_r son, respectivamente, el valor de la verosimilitud en el el modelo restringido y sin restringir. Este estadístico tiene distribución asintótica χ^2 con r grados de libertad, en donde r es el número de restricciones.

- *Pseudo-R*²:

$$LR = 1 - \frac{\ln L}{\ln L_0}$$

L es el valor máximo de la función de verosimilitud bajo el modelo original y L_0 es el valor correspondiente al modelo con solo una constante. Mide el incremento en la capacidad explicativa por considerar un modelo mas completo que solo una constante.

$$\hat{p}_i \equiv x_i' \hat{\beta}$$

$$\hat{y}_i \equiv 1[\hat{p}_i > c] \quad (\text{generalmente, } c = 0.5)$$

$$y_i^c \equiv 1[\hat{y}_i = y_i] \quad (1 \text{ si prediccion correcta})$$

El porcentaje de predicciones correctas es:

$$H = \frac{\sum_{i=1}^n y_i^c}{n}$$

- Sensible a la eleccion de c . Errores simetricos.
- Un modelo trivial tiene $H \geq 0.5$