

# Econometría II

## 3.1 Modelos Binarios

Profesora: Graciela Sanroman  
Guía esquemática para las clases

Año: 2016

## **Modelos de elección discreta (de variable cualitativa):**

Los modelos de variable dependiente cualitativa o limitada se corresponden con los modelos de:

- elección discreta
- regresión censurada y truncada.

## Modelos de elección discreta (de variable cualitativa):

- Los agentes eligen alguna alternativa dentro de un conjunto finito de opciones.
- Los agentes pueden ser individuos, hogares, empresas, o cualquier unidad de decisión económica.
- Para pertenecer a este tipo de modelos el conjunto de opciones denominado conjunto de elección debe cumplir con tres características.
  - el conjunto de alternativas debe ser mutuamente excluyente desde la perspectiva del agente
  - las alternativas deben conformar un conjunto exhaustivo, en el que están incluidos todos los eventos posibles
  - la cantidad de alternativas debe ser un número finito.
- Así, en estos modelos la variable a ser explicada,  $y$ , es una variable aleatoria que sólo puede adoptar un número finito y generalmente reducido de valores.

## Modelos de elección discreta

Los modelos de elección discreta se clasifican en:

- modelos cross-section y de datos panel (según el tipo de datos)
- binarios o multinomiales (según la cantidad de alternativas)
- paramétricos o semiparámetricos (según los supuestos respecto a la estructura del modelo)
- dinámicos o estáticos
- un sólo agente o juegos.

## Modelos binarios

En el caso binario es usual suponer que los valores que puede adoptar la variable  $y$  son 0 o 1, y se denomina “éxito” al caso en el cual  $y = 1$ .

En estos modelos estaremos interesados en estimar la probabilidad de que  $y$  adopte un valor determinado.

$$\Pr(y) = \begin{cases} y = 0 & 1 - p \\ y = 1 & p \end{cases} \quad (1)$$

Es fácil corroborar que:

$$E(y) = \Pr(y = 1) = p$$

$$V(y) = E(y - E(y))^2 = p(1 - p)$$

## Modelos binarios

Pero lo que interesará es el análisis condicional, o sea el análisis de  $y \mid x$ , con  $x = (x_1, x_2, \dots, x_k)'$ :

$$E(y \mid x) = \Pr(y = 1 \mid x) = p(x) \quad (2)$$

$$V(y \mid x) = p(x) [1 - p(x)] = \sigma^2(x) \quad (3)$$

Entonces, podemos también ver este problema como uno en el que queremos un modelo para la media condicional  $E(y \mid x) = p(x)$  sabiendo que se trata de un problema intrínsecamente heterocedástico.

## Modelos binarios

En los modelos de elección binaria, de forma análoga que en el caso del modelo de regresión lineal, el interés de la estimación puede estar en:

- a) Estudiar qué variables del vector  $x$  son relevantes para determinar la probabilidad de que  $y$  sea igual a uno (cero).
- b) Estimar los efectos parciales (analizar cómo varía esa probabilidad ante cambios de una variable, por ejemplo,  $x_j$ ).
- c) Predecir  $y$  para distintos valores de  $x$ .

Pero para ello es necesario estimar  $p(x)$ .

## Modelos binarios

Supongamos que

$\{y_i, x_{1i}, x_{2i}\}_{i=1, \dots, N}$  es una muestra aleatoria, donde:

$y_i$  adopta el valor 1 si el individuo  $i$  trabaja y 0 si está desocupado

$x_{1i}$  es la edad del individuo  $i$

$x_{2i}$  adopta el valor 1 si el individuo es mujer y 0 si es hombre,

La probabilidad condicional  $\Pr(y_i \mid \text{Edad} = 35, \text{Sexo} = \text{Mujer})$  estará dada de forma no paramétrica por:

$$\Pr(y_i \mid \text{Edad} = 35, \text{Sexo} = \text{Mujer}) = \frac{\sum_{i=1}^N y_i I(\text{Edad} = 35, \text{Sexo} = \text{Mujer})}{\sum_{i=1}^N I(\text{Edad} = 35, \text{Sexo} = \text{Mujer})}$$

Sin embargo, hay varias razones por las que estaremos interesados en realizar una estimación paramétrica (eficiencia, parsimonia, realizar predicciones) y para realizar una estimación paramétrica es necesario elegir una forma funcional para  $p(x_i)$ .



## Modelos binarios

Dos alternativas básicas:

- El modelo de probabilidad lineal (MPL) que supone linealidad es decir que  $p(x_i)$  se puede representar como  $x_i'\beta$ .
- Los modelos índice (o indicador) de los cuales veremos 2:
  - Probit,  $p(x_i) = \Phi(x_i'\beta)$  siendo  $\Phi(\cdot)$  la función de distribución de una v.a.  $N(0, 1)$ .
  - Logit,  $p(x_i) = \Lambda(x_i'\beta)$  siendo  $\Lambda(\cdot)$  la función de distribución de una v.a. logística estándar.

Notar que  $x$  es un vector de dimensión  $k$  que puede incluir variables transformadas (tales como el logaritmo, el cuadrado o el cubo) o interacciones entre variables.

## El modelo de probabilidad lineal

El modelo de probabilidad lineal para variables dependientes binarias puede especificarse como,

$$\Pr(y_i = 1 \mid \mathbf{x}_i) = \mathbf{x}_i' \boldsymbol{\beta} = \beta_0 + x_{1i}\beta_1 + x_{2i}\beta_2 + \dots + x_{ki}\beta_k$$

En este modelo el efecto parcial  $\partial \Pr(y_i = 1 \mid \mathbf{x}_i) / \partial x_j = \beta_j$

El modelo es intrínsecamente heterocedástico:

- MCO será consistente pero ineficiente
- podemos estimar por MCO y calcular los errores estándar a través de la fórmula de White.

## El modelo de probabilidad lineal

Dado que en este modelo la forma de la varianza del estimador es conocida, los estimadores MCG serán asintóticamente más eficientes. Para estimar utilizando MCG dos etapas:

**ETAPA 1:** obtener estimaciones consistentes  $\hat{\beta}_{MCO}$  y predecir  $\hat{p}_i = \mathbf{x}_i' \hat{\beta}_{MCO}$

**ETAPA 2:** Se obtienen los estimadores  $\hat{\beta}_{MCG}$  aplicando la regresión MCO a

$$\frac{y_i}{\hat{\sigma}_i}, \frac{1}{\hat{\sigma}_i}, \frac{x_{1i}}{\hat{\sigma}_i}, \dots, \frac{x_{Ki}}{\hat{\sigma}_i} \quad i = 1..N.$$

$$\text{donde } \hat{\sigma}_i^2 = \hat{p}_i(1 - \hat{p}_i)$$

# El modelo de probabilidad lineal: limitaciones y ventajas

## Limitaciones:

- Predicciones  $\hat{p}_i$  no necesariamente pertenecen al intervalo  $[0, 1]$
- Efectos parciales constantes  $\partial P(y_i = 1 \mid \mathbf{x}_i) / \partial x_k = \beta_k \quad \forall \mathbf{x}_i$ .

## Ventajas:

- Es lineal, por lo cual las propiedades de los estimadores son ampliamente conocidas
- Ofrece buenas estimaciones de los efectos parciales para valores de  $\mathbf{x}_i$  cercanos a su media (en ausencias de otros problemas de especificación)
- Problemas tales como errores de medida en las variables explicativas, presencia de heterogeneidad inobservable u otras fuentes de endogeneidad son más fácilmente tratadas ya que podemos utilizar VI y MGM y en particular todos los estimadores vistos en modelos lineales para datos de panel.

## Modelos “Índice”: Probit y Logit

En términos más generales puedo escribir el modelo de probabilidad condicional:

$$\begin{aligned}P(y_i = 1 \mid \mathbf{x}_i) &= p(\mathbf{x}_i) = p_i \\&= G(\mathbf{x}_i' \frac{\beta^*}{\sigma}) = G(\mathbf{x}_i' \beta) = G(r)\end{aligned}$$

donde

$\mathbf{x}_i = (1, x_{2i}, \dots, x_{Ki})$  es un vector  $K \times 1$  de características del individuo  $i$

$\beta = (\beta_1, \beta_2, \dots, \beta_K)$  es un vector  $K \times 1$  de parámetros desconocidos

denominamos a  $r = \mathbf{x}_i' \beta = \beta_1 + \beta_2 x_{2i} \dots + \beta_K x_{Ki}$  función “Índice”

Notar que no hay identificación separada de  $\beta^*$  y  $\sigma$ , sólo identifico  $\beta = \frac{\beta^*}{\sigma}$  (up to scale) por lo que para estimar el modelo se supone  $\sigma = 1$ .

## Modelos “Índice”: Probit y Logit

$G(.)$  es una función que proyecta ese índice en la probabilidad condicional. Queremos  $0 < G(.) < 1$ , entonces una fdp es una opción razonable. Las más usadas son la Normal estándar y la Logit estándar (media 0 y varianza  $\frac{\pi^2}{3}$ ). Así

$$\begin{aligned} G(r) &= \Phi(r) \\ &= \int_{-\infty}^r \frac{1}{\sqrt{2\pi}} \exp\left\{-\frac{1}{2}x^2\right\} dx \end{aligned}$$

$$\begin{aligned} G(r) &= \Lambda(r) \\ &= \frac{\exp(r)}{1 + \exp(r)} \end{aligned}$$

Las predicciones de estos modelos serán predicciones de la probabilidad condicional y asumirán la forma  $\hat{p}_i = G(r_i) = G(\mathbf{x}'_i \hat{\beta})$ .

## Estimación de los modelos Probit y Logit

Cuando utilizamos este tipo de modelos nos enfrentamos a una estimación en la cual la forma de la función de verosimilitud es conocida ya que cada observación seguirá una distribución Bernoulli. Así, la contribución de cada individuo a la verosimilitud será

$$p_i^{y_i} (1 - p_i)^{1-y_i}$$

entonces, la función de verosimilitud de una muestra aleatoria  $\{y_i, \mathbf{x}_i\}$ ,  $i = 1, \dots, N$  será

$$\mathcal{L}(\beta) = \prod_{i=1}^N p_i^{y_i} (1 - p_i)^{1-y_i} \quad (4)$$

de donde la log-verosimilitud

$$L(\beta) = \sum_{i=1}^N l_i(\beta) \quad (5)$$

donde  $l_i(\beta) = y_i \log(p_i) + (1 - y_i) \log(1 - p_i)$  y  $p_i = G(\mathbf{x}_i' \beta)$ .

## Estimación de los modelos Probit y Logit

Entonces, los estimadores máximo verosímiles (EMV) de  $\beta$  serán

$$\hat{\beta}_{MV} = \hat{\beta} = \arg \max_b \sum_{i=1}^N l_i(b). \quad (6)$$

Las CPO del problema estarán dadas por:

$$\sum_{i=1}^N \frac{\partial l_i(\hat{\beta})}{\partial b} = 0. \quad (7)$$

Sea  $g(r) = \frac{\partial G(r)}{\partial r}$ , entonces  $\frac{\partial l_i(b)}{\partial b} = \frac{g(\mathbf{x}'_i b) \mathbf{x}_i [y_i - G(\mathbf{x}'_i b)]}{G(\mathbf{x}'_i b) [1 - G(\mathbf{x}'_i b)]} = \frac{g(\mathbf{x}'_i b) \mathbf{x}_i [y_i - p_i]}{p_i (1 - p_i)}$ . Es importante notar que esta última expresión evaluada en el verdadero valor de  $\beta$  corresponde con el tanteo (o score) de la log-verosimilitud condicional de la observación  $i$ -ésima.



## Estimación de los modelos Probit y Logit

El valor esperado del hessiano condicional en  $\mathbf{x}_i$  resulta:

$$E [H_i (\beta) \mid \mathbf{x}_i] = - \frac{[g(\mathbf{x}'_i \beta)]^2}{G(\mathbf{x}'_i \beta) [1 - G(\mathbf{x}'_i \beta)]} \mathbf{x}_i \mathbf{x}'_i \quad (8)$$

y por lo tanto

$$\text{Avar} \left( \hat{\beta} \right) = \left\{ \sum_{i=1}^N \frac{[\hat{g}_i]^2}{\hat{p}_i (1 - \hat{p}_i)} \mathbf{x}_i \mathbf{x}'_i \right\}^{-1} = \hat{V} \quad (9)$$

Con

$$\hat{p}_i = G(\mathbf{x}'_i \hat{\beta})$$

y

$$g(\mathbf{x}'_i \beta) = \phi(\mathbf{x}'_i \beta) = \frac{1}{\sqrt{2\pi}} \exp \left\{ -\frac{1}{2} [\mathbf{x}'_i \beta]^2 \right\} \quad (\text{Probit})$$

$$g(\mathbf{x}'_i \beta) = \Lambda(\mathbf{x}'_i \beta) [1 - \Lambda(\mathbf{x}'_i \beta)] = \frac{\exp(\mathbf{x}'_i \beta)}{[1 + \exp(\mathbf{x}'_i \beta)]^2} \quad (\text{Logit})$$

## Contrastes de hipótesis en los modelos Probit y Logit

El vector  $\hat{\beta}$  tendrá una distribución asintótica normal con media  $\beta$ . El error estándar (asintótico) de  $\hat{\beta}_j$  estará dado por la raíz cuadrada del  $j$ -ésimo elemento de la diagonal de  $\hat{V}$  en (9). Por lo tanto es posible calcular los estadísticos  $t$  y construir intervalos de confianza a la manera usual para realizar contrastes respecto a los parámetros  $\beta_j$ .

Notar que los contrastes sólo tienen validez asintótica, debido a que no conocemos su distribución en muestras pequeñas. Es decir que en el contraste

$$H_0 \beta_j = 0 \text{ frente a } H_1 \beta_j \neq 0$$

el estadístico de la prueba, bajo  $H_0$ , se distribuye normal estándar si la muestra es muy grande

$$\frac{\hat{\beta}_j}{\text{ee}(\hat{\beta}_j)} \stackrel{a}{\sim} N(0, 1)$$

## Contrastes de hipótesis en los modelos Probit y Logit

Para realizar contrastes múltiples, considere el modelo

$$P(y_i = 1 \mid \mathbf{x}_i, \mathbf{z}_i) = G(\mathbf{x}_i' \boldsymbol{\beta} + \mathbf{z}_i' \boldsymbol{\gamma})$$

donde  $\mathbf{x}_i, \mathbf{z}_i$  son vectores de dimensión  $K_1$  y  $K_2$  respectivamente.

Queremos realizar el contraste de exclusión  $\mathbf{H}_0 : \boldsymbol{\gamma} = \mathbf{0}$ .

Utilizar tanto los estadísticos de Wald, RV o LM, los tres son asintóticamente equivalentes, y bajo la  $H_0$ , convergen en distribución a una  $\chi^2$

$$W \stackrel{a}{\sim} \chi_{K_2}^2$$

$$RV \stackrel{a}{\sim} \chi_{K_2}^2$$

$$LM \stackrel{a}{\sim} \chi_{K_2}^2$$

## Medidas de bondad de ajuste

Las medidas de bondad de ajuste más usuales son:

1. Proporción de predicciones correctas.

Sea  $vc$  un “valor de corte”, se define

$$\hat{y}_i = \begin{cases} 0 & \text{si } \hat{p}_i < vc \\ 1 & \text{si } \hat{p}_i \geq vc \end{cases}$$

Habitualmente  $vc$  es establecido igual a 0.50, o también se establece en la frecuencia muestral de unos.

## Medidas de bondad de ajuste

Una medida del ajuste del modelo podría ser la proporción total de aciertos, no obstante, ésta será en general una mala medida.

Deberemos tener en cuenta separadamente la proporción de unos (ceros) predichos correctamente, o sea,

$\hat{y}_i$	0	1
$y_i$		
0	$\widehat{pr0} = \frac{\sum_i (1-y_i)(1-\hat{y}_i)}{\sum_i (1-\hat{y}_i)}$	
1		$\widehat{pr1} = \frac{\sum_i y_i \hat{y}_i}{\sum_i \hat{y}_i}$

A partir de estos estadísticos, se pueden obtener medidas sintéticas, tales como combinaciones lineales convexas de  $\widehat{pr1}$  y  $\widehat{pr0}$ , o el  $\min(\widehat{pr1}, \widehat{pr0})$ .

## Medidas de bondad de ajuste

### 2. Pseudo- $R^2$ de McFadden.

Se basa en comparar la log-verosimilitud resultante en la estimación con las  $K$  variables explicativas  $\left(L\left[\hat{\beta}\right]\right)$  y compararlo con la log-verosimilitud de un modelo que incluye sólo una constante ( $L_0$ ).

$$\text{Pseudo-}R^2 = 1 - \frac{L\left(\hat{\beta}\right)}{L_0}. \quad (10)$$

Notar que éste estadístico está relacionado con el estadístico **RV** pero no es igual, ya que **RV** =  $2\left(L\left(\hat{\beta}\right) - L_0\right)$ .

## Medidas de bondad de ajuste

### 3. Criterio de información de Akaike

$$AIC = -L\left(\hat{\beta}\right) + K$$

donde  $K$  es la dimensión del vector  $\beta$ .

Este estadístico particularmente útil para contrastar modelos no anidados. Como es habitual el criterio es elegir el modelo que minimice el estadístico  $AIC$ .

## **Medidas de bondad de ajuste**

Es importante destacar que ninguno de los estadísticos antes definidos tiene una distribución asintótica conocida. Recordar que, no obstante, es posible realizar contrastes estadísticos sobre la significación conjunta de todos (o varios) de los coeficientes de un modelo mediante los contrastes de Wald, RV y LM, tal como se explicó en la sub-sección anterior.



## **Efectos parciales en los modelos Probit y Logit**

Una diferencia importante de los modelos Probit y Logit con respecto al Modelo de Probabilidad Lineal es que, en los primeros, los efectos parciales dependerán del valor de  $\mathbf{x}_i$ .

## Efectos parciales cuando la variable $x$ es continua

Si  $x_j$  es una variable continua, entonces es posible definir el efecto parcial como,

$$\begin{aligned}\frac{\partial P(y_i = 1 \mid \mathbf{x}_i)}{\partial x_j} &= \frac{\partial P(y_i = 1 \mid 1, x_{1i}, \dots, x_{ji}, x_{Ki})}{\partial x_j} \\ &= \beta_j g(\mathbf{x}_i' \beta)\end{aligned}$$

El efecto parcial será diferente para los diferentes individuos. Es necesario decidir en que valores se evalúan, las dos alternativas más obvias:

- Obtener la media muestral del efecto parcial  $\beta_j \left[ \frac{1}{N} \sum_{i=1}^N g(\mathbf{x}_i' \beta) \right]$ .
- Evaluarlo para  $\bar{\mathbf{x}}$  (el valor de la media muestral de  $\mathbf{x}_i$ ), o sea, el efecto parcial es  $\beta_j [g(\bar{\mathbf{x}}' \beta)] = \beta_j [g(\beta_1 + \beta_2 \bar{x}_2 + \dots \beta_K \bar{x}_K)]$  donde

$$\bar{x}_j = \frac{1}{N} \sum_{i=1}^N x_{ji}.$$

Es importante tener en cuenta que el efecto parcial de  $\beta_j$  es una función del vector  $\beta$ , por lo cual, para calcular los errores estándar y construir intervalos de confianza para los efectos parciales se utilizará el método delta.

## Efectos parciales cuando la variable $x$ es binaria

Suponga que  $x_K$  es una variable binaria, o sea, adopta sólo dos posibles valores, por ejemplo 0 y 1, el efecto parcial de  $x_K$  estará definido por la diferencia,

$$G(\beta_1 + \beta_2 x_{2i} \dots + \beta_{K-1} x_{k-1i} + \beta_K) \\ - G(\beta_1 + \beta_2 x_{2i} \dots + \beta_{K-1} x_{k-1i})$$

En este caso también es necesario definir donde evaluar el efecto parcial y se aplica lo anterior, o sea, dos alternativas son obtener la media muestral de esta expresión o evaluar esta expresión para el promedio de las  $x$  que no son  $x_K$ .

## Efectos parciales cuando la variable $x$ es discreta

Suponga que  $x_K$  es una variable discreta ordinal, que adopta valores en el conjunto de los número enteros, por ejemplo  $x_{Ki}$  mide los años de educación completos de un individuo, el efecto parcial de  $x_K$  estará definido por la diferencia,

$$G [\beta_1 + \beta_2 x_{2i} \dots + \beta_{K-1} x_{k-1i} + \beta_K (x_{Ki} + 1)] \\ - G [\beta_1 + \beta_2 x_{2i} \dots + \beta_{K-1} x_{k-1i} + \beta_K x_{Ki}]$$

Para evaluar este efecto parcial se aplican las alternativas de obtener la media muestral de esta expresión o de evaluar esta expresión para el promedio de las  $x$ .<sup>1</sup>

---

<sup>1</sup>Formas funcionales de las funciones  $G(.)$  y  $g(.)$  para los modelos Probit y Logit son las previamente definidas.

## Probit y Logit como modelos de variable latente

Los modelos probit y logit se pueden también entender como modelos de variable latente o modelos de utilidad estocástica (RUM).

Supongamos que un individuo enfrenta un problema de decisión binaria del tipo (participar o no participar); específicamente supongamos que estoy estudiando la decisión de enviar un niño a una guardería.

Definamos  $w_0$  como la utilidad que obtiene la madre si no lo envía a guardería y  $w_1$  si lo envía.

$$w_0 = h_0 + \varepsilon_0$$

$$w_1 = h_1 + \varepsilon_1$$

donde  $h_j$  es un componente determinístico y  $\varepsilon_j$  es un componente aleatorio.

## Probit y Logit como modelos de variable latente

La decisión dependerá del signo de  $w_1 - w_0$ , al cual puedo expresarlo como,

$$\begin{aligned}w_1 - w_0 &= (h_1 - h_0) + (\varepsilon_1 - \varepsilon_0) \\w^* &= h + \varepsilon\end{aligned}$$

$$\begin{aligned}y_i &= \mathbf{1}(w^* > 0) \\&= \mathbf{1}(h > -\varepsilon)\end{aligned}$$

Si supongo que  $\varepsilon$  se distribuye  $N(0,1)$  entonces el modelo probit además de ser un modelo estadístico tiene una interpretación estructural,

$$\begin{aligned}\Pr(y_i = 1) &= \Pr(w_i^* > 0) \\&= \Pr(-h_i < \varepsilon) \\&= 1 - \Phi(-h_i) \\&= \Phi(h_i)\end{aligned}$$

## Probit y Logit como modelos de variable latente

Similarmente si supongo que  $\varepsilon$  se distribuye logistica, tengo

$$\begin{aligned}\Pr(y_i = 1) &= \Pr(w_i^* > 0) \\ &= \Pr(-h_i < \varepsilon_i) \\ &= 1 - \Lambda(-h_i) \\ &= \Lambda(h_i)\end{aligned}$$

## Probit y Logit como modelos de variable latente

Pero estamos considerando modelos microeconómicos, en los cuales lo que más interesa es analizar el comportamiento condicional, es decir, el objetivo del modelo es tener en cuenta  $h$  distintas para los individuos, En un modelo condicional no siempre puedo observar separadamente  $h_0$  o  $h_1$ , pero supongo que la diferencia entre ambas depende de características observables del individuo ( $x_i$ ) o de los atributos de las distintas alternativas ( $z_j$ ).

$$v_i = x_i' \beta + z_j' \delta$$

$$\begin{aligned} y_i \mid x_i, z_j &= \mathbf{1}(w_i^* > 0) \\ &= \mathbf{1}(x_i' \beta + z_j' \delta > -\varepsilon) \end{aligned}$$



Para estimar este modelo podría suponer  $\varepsilon$  se distribuye como una variable normal,

$$\Pr(y_i | x_i, z_j) = \Phi(x_i' \beta + z_j' \delta)$$

pero también podría suponer que es una v.a. logit

$$\Pr(y_i | x_i) = \Lambda(x_i' \beta + z_j' \delta)$$

## Problemas de especificación en los modelos Probit y Logit

- En el modelo de regresión lineal MCO es consistente si se cumple la condición de ortogonalidad entre el regresor y el término de perturbación. Los problemas de heterocedasticidad, autocorrelación y no-normalidad del término de error no afectaban la consistencia del estimador.
- En los modelos de elección discreta la consistencia del Estimador Máximo Versosímil (EMV) también depende del cumplimiento de la condición de ortogonalidad pero también y crucialmente de otros supuestos acerca del término de error. Si éste es,
  - heterocedástico
  - sigue una distribución de probabilidad diferente a la que se asume,

entonces el EMV no es consistente.

La razón es que los supuestos respecto al término de error afectan no sólo el segundo momento y momentos de orden mayor sino también la media condicional.

## Problemas de especificación en los modelos Probit y Logit: variables omitidas independientes

Consideremos el problema de omisión de variables independientes. Este problema suele denominarse “heterogeneidad inobservable”.

Consideremos el siguiente modelo de variable latente:

$$y_i^* = x_i' \beta + \eta_i + v_i$$

donde

$$y_i = I(y_i^* > 0)$$

$$v_i \mid x_i, \eta_i \sim N(0, 1)$$

## Problemas de especificación en los modelos Probit y Logit: variables omitidas independientes

Supongo que  $\eta_i$  es independiente de  $x$  y  $\eta_i \sim N(0, \sigma_\eta^2)$ . Bajo estos supuestos el término de error compuesto

$$\varepsilon_i = \eta_i + v_i$$

El modelo estructural de interés es

$$\Pr(y_i = 1 \mid x_i, \eta_i) = \Pr[v_i > -(x_i' \beta + \eta_i)] = \Phi(x_i' \beta + \eta_i)$$

donde  $x$  es un vector  $1 \times K$  con  $x_1 = 1$ .

## Problemas de especificación en los modelos Probit y Logit: variables omitidas independientes

Estamos interesados en el efecto parcial de  $x_j$  en la probabilidad de éxito, manteniendo  $\eta_i$  (y los restantes elementos de  $x$ ) constantes.

$\varepsilon_i$  es independiente de  $x$  y tiene una distribución  $N(0, \sigma_\eta^2 + 1)$ . Entonces

$$\Pr(y_i = 1 \mid x_i) = \Pr(\eta_i + v_i > -x_i' \beta \mid x_i) = \Phi(x_i' \frac{\beta}{\sigma})$$

donde  $\sigma = \sqrt{\sigma_\eta^2 + 1}$ . Entonces se deduce inmediatamente que un modelo probit de  $y$  sobre  $x$  ofrece una estimación consistente de  $\frac{\beta}{\sigma}$ .

La presencia de un sesgo de atenuación originado en la presencia de heterogeneidad inobservable en los modelos probit ha concitado afirmaciones del tipo “en los modelos probit la presencia de heterogeneidad inobservable es un problema más serio que en los modelos lineales debido a que, aun cuando la variable omitida fuese independiente de los regresores las estimaciones probit serán inconsistentes”.

## Problemas de especificación en los modelos Probit y Logit: variables omitidas independientes

Aquí se ha demostrado que efectivamente un probit de  $y$  sobre  $x$  estima consistentemente  $\frac{\beta}{\sigma}$  en lugar de  $\beta$ ; no obstante, debemos recordar que en los modelos no lineales usualmente se desea estimar los efectos parciales y no los coeficientes; y para el objetivo de obtener la dirección de los efectos o el efecto relativo de los diversos regresores estimar  $\frac{\beta}{\sigma}$  es tan bueno como estimar  $\beta$ .

Para una variable continua  $x_j$ , podríamos estar interesados en estimar:

$$\frac{\partial \Pr(y_i = 1 \mid x_i, \eta_i)}{\partial x_j} = \beta_j \phi(x_i' \beta + \eta_i)$$

para diversos valores de  $x$  y  $\eta$ .

## Problemas de especificación en los modelos binarios: endogeneidad en los regresores

Consideremos  $\eta_i$  correlacionada con  $x_i$ , o depende en cierta medida de  $x$  (por ejemplo  $Var(\eta_i | x_i) = h(x_i)$ ).

En este caso la omisión de  $\eta_i$  tendrá graves consecuencias: ni los estimadores ni los efectos parciales promedios serán consistentes. Similar que en los modelos de regresión lineal.

Para salvar este problema se puede estimar el modelo utilizando el MPL y:

- en un corte transversal estimar por VI o MGM si dispongo de un instrumento válido. Este procedimiento es relativamente sencillo y en general aporta una buena medida de los efectos promedio.
- si dispongo de un panel utilizar todas las opciones estudiadas para modelos lineales de datos de panel
- bajo algunos supuestos adicionales será posible estimar de forma consistente. Para describir estos métodos es necesario distinguir el caso en el cual el regresor es continuo de aquel en el que es una variable binaria.

## Problemas de especificación en los modelos binarios: utilización de una fdp errónea

Suponga que el modelo verdadero es tal que  $\varepsilon_i \sim iid$  con fdp  $F$ .

$$y^* = x_i' \beta + \varepsilon_i$$

$$y = I[y^* > 0] = I[\varepsilon_i > -x_i' \beta]$$

Entonces,

Modelo verdadero:  $y_i = F(x_i' \beta) + u_i$  donde  $E(u_i | x_i) = 0$ .

Si suponemos  $\varepsilon_i | x_i \sim iidN(0, 1)$  (Probit),

Modelo estimado:

$$y_i = \Phi(x_i' \beta) + u_i^*$$

$$u_i^* = u_i + F(x_i' \beta) - \Phi(x_i' \beta)$$

Es sencillo observar que si  $F \neq \Phi$  entonces  $E(u_i^* | x_i) \neq 0$  y el EMV que utiliza la fdp Normal es inconsistente.

Es importante notar que el sesgo que se produce en la estimación puede afectar no sólo los coeficientes sino también las probabilidades estimadas.



## **Problemas de especificación en los modelos binarios: otros potenciales problemas**

Horowitz (1993) realizó una serie de simulaciones y arribó a las siguientes conclusiones, respecto a utilizar un modelo Logit cuando la verdadera distribución es otra:

- (a) Los errores son pequeños cuando la verdadera distribución es unimodal y homocedástica.
- (b) Los errores pueden ser muy grandes cuando la verdadera distribución es bimodal o heterocedástica.

Es importante notar, no obstante, que en el caso binario, la diferencia entre las estimaciones probit y logit es muy pequeña, es decir los errores cometidos son pequeños si una distribución logit se modeliza como probit o viceversa, en particular serán más reducidos cuanto más distribuídos se presenten los unos y ceros de la muestra.

## Problemas de especificación en los modelos binarios: Heterocedasticidad

$$y_i^* = x_i' \beta + \varepsilon_i \quad y_i = I[y_i^* > 0]$$

Probit homocedástico:

$$\varepsilon_i \sim N(0, 1)$$

$$\Pr(y_i = 1 | x_i) = \Phi(x_i' \beta)$$

Probit heterocedástico:

$$\varepsilon_i \sim N(0, [\exp(z_i' \gamma)]^2)$$

$$\Pr(y_i = 1 | x_i, z_i) = \Phi\left(\frac{x_i' \beta}{\exp(z_i' \gamma)}\right)$$

Log-verosimilitud modelo probit heterocedástico

$$\ln L = \sum_{i=1}^N \left\{ y_i \ln \Phi\left(\frac{x_i' \beta}{\exp(z_i' \gamma)}\right) + (1 - y_i) \ln \left(1 - \Phi\left(\frac{x_i' \beta}{\exp(z_i' \gamma)}\right)\right) \right\}$$

## Referencias:

- \*\*\* Wooldridge, J. M. (2001) Introducción a la Econometría: un enfoque moderno, Thomson Learning, México. (2a. Edición en español, 2006). Capítulo 7 (7.5 y 7.6) y Capítulo 17.
- \*\* Wooldridge, J. (2002) Econometric Analysis of Cross Section and Panel Data, MIT Press. Capítulo 15 (15.1 a 15.6), Capítulo 16, Capítulo 17 (17.1 a 17.4), Capítulo 19 (19.1 a 19.5 y 19.6.1).
- \*\* Mroz, T. (1987) "The Sensitivity of an Empirical Models of Married Women's of Work to Economic and Statistical Assumptions", Econometrica, 55(4), 765-799.
- \* Cameron A. C. y P.K. Trivedi (2009) Microeconometrics Using Stata, Stata Press. Capítulos 14, 15 y 16.
- \* Greene, W. H. (1999) Análisis Económico. 3a. Edición. Prentice Hall Iberia, Madrid. Capítulos 19 y 20.