Economía Aplicada

Modelos con variables dependiente binarias

Departamento de Economía
Universidad Carlos III de Madrid

Ver Stock y Watson (capítulo 11)

Modelos con variables dependiente binarias: ¿Cuál es la diferencia?

- Hasta ahora, la variable dependiente (Y) era continua:
  - calificación promedio en una prueba
  - tasa de mortalidad
  - salarios
- ¿Qué pasa si ahora Y es binaria?
  - Y ser o no aceptado en la universidad; X promedio en secundaria, selectividad, otros controles demográficos
  - Y si una persona fuma o no; X impuestos al tabaco, renta, otros controles demográficos
  - Y si se acepta o no una solicitud para una hipoteca; X raza, renta, características de la vivienda, estado civil

Ejemplo: La denegación de hipotecas y raza del individuo. The Boston Fed HMDA Dataset

- Las solicitudes individuales de hipotecas unifamiliares realizadas en 1990 en el área metropolitana de Boston
- 2380 observaciones, recogidas bajo la Ley de Divulgación de Hipotecas (HMDA)
- Variables
  - Variable dependiente: se deniega o concede la hipoteca
  - Variables independientes: el ingreso, la riqueza, la situación laboral, otro préstamo, características de la propiedad, y raza del solicitante.

Modelo de Probabilidad Lineal (MPL)

Un punto de partida natural es el modelo de regresión lineal con un único regresor:

$Y_i = \beta_0 + \beta_1 X_i + u_i$

Pero:

- ¿Qué significa β_1 cuando Y es binaria?
- ¿Qué significa la línea $\beta_0 + \beta_1 X$ cuando Y es binaria?
- ¿Qué significa el valor predicho $\widehat{Y}$ cuando Y es binaria? Por ejemplo, ¿qué quiere decir $\widehat{Y} = 0.26$?

Modelo de Probabilidad Lineal (MPL)

Cuando Y es binaria decimos que es una variable aleatoria Bernouilli:

$$E(Y|X) = 1 * Pr(Y = 1|X) + 0 * Pr(Y = 0|X) = Pr(Y = 1|X)$$

Y bajo el supuesto, $E(u_i|X_i) = 0$:

$$E(Y_i|X_i) = E(\beta_0 + \beta_1 X_i + u_i|X_i) = \beta_0 + \beta_1 X_i,$$

Entonces:

$$E(Y|X) = Pr(Y = 1|X) = \beta_0 + \beta_1 X_i$$

En el MPL, el valor predicho de Y se interpreta como la probabilidad predicha de que $Y = 1$ y β_1, es el cambio en la probabilidad producto de un cambio unitario en X.

Modelo de Probabilidad Lineal (MPL)

- Cuando Y es binaria, el modelo de regresión lineal

$$Y_i = \beta_0 + \beta_1 X_i + \mu_i$$

corresponde a la probabilidad (condicional en X) de que $Y = 1$, la cual es lineal en los parámetros β_0 y β_1

$$Pr(Y = 1|X) = \beta_0 + \beta_1 X_i$$

- El valor predicho es una **probabilidad**:

-$E(Y|X = x) = Pr(Y = 1|X = x) =$ prob. de que $Y = 1$ dado $X = x$

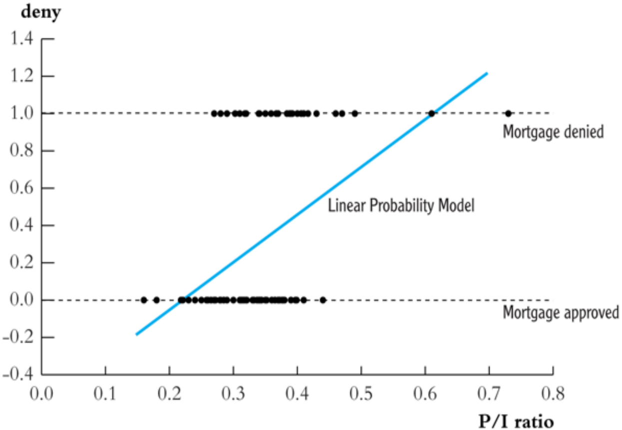
-$\widehat{Y} =$ **la probabilidad predicha** de que $Y = 1$ dado X

-$\beta_1 =$ el cambio en la probabilidad de que $Y = 1$ para un cambio en una unidad de x:

$$\beta_1 = \frac{Pr(Y=1|X=x+\Delta x) - Pr(Y=1|X=x)}{\Delta x}$$

Ejemplo: MPL, HMDA data

- Denegación de hipoteca frente a proporción de los pagos asociados a la deuda respecto a los ingresos (ratio o relación P/I), para una sub-muestra de los datos de HMDA ($n = 127$)



Ejemplo:MPL, HMDA data

$$\widehat{deny_i} = \widehat{\beta}_0 + \widehat{\beta}_1 PI_i + \widehat{\beta}_2 black_i$$

Model 1: OLS, using observations 1–2380
Dependent variable: deny
Heteroskedasticity-robust standard errors, variant HC1

	Coefficient	Std. Error	t-ratio	p-value
const	−0.0905136	0.0285996	−3.1649	0.0016
pi_rat	0.559195	0.0886663	6.3067	0.0000
black	0.177428	0.0249463	7.1124	0.0000

Mean dependent var	0.119748	S.D. dependent var	0.324735
Sum squared resid	231.8047	S.E. of regression	0.312282
R^2	0.076003	Adjusted R^2	0.075226
$F(2, 2377)$	49.38650	P-value(F)	9.67e–22
Log-likelihood	−605.6108	Akaike criterion	1217.222
Schwarz criterion	1234.546	Hannan–Quinn	1223.527

El MPL: Resumen

-Ventajas:

- sencillo de estimar e interpretar
- la inferencia es la misma que la utilizada en el modelo de regresión múltiple

-Nótese que es necesario utilizar errores estándar robustos porque el MPL es heterocedástico:

$$V(Y|X) = E(Y^2|X) - [E(Y|X)]^2$$

donde $E(Y^2|X) = 1^2 * P(Y=1|X) + 0^2 * P(Y=0|X) = P(Y=1|X)$

de manera que

$$V(Y|X) = P(Y=1|X) - (P(Y=1|X))^2$$

$$V(Y|X) = P(Y=1|X) * (1 - P(Y=1|X))$$

El MPL: Resumen

- Desventajas:

  - El MPL nos da un cambio en la probabilidad predicha para un determinado valor de X que es igual para todos los valores de X, pero esto no tiene sentido cuando la probabilidad de un evento está acotada. Piensa en el ejemplo anterior (HMDA)

  - Del mismo modo, en el MPL las probabilidades predichas pueden ser < 0 o > 1!

- Estas desventajas pueden ser resueltas por medio de modelos de probabilidad no lineales: probit y logit

Modelo probit y logit

- El problema con el MPL es que modela la probabilidad de que $Y = 1$ por medio de una función lineal:

$$Pr(Y = 1|X) = \beta_0 + \beta_1 X$$

- sin embargo lo que queremos es que:

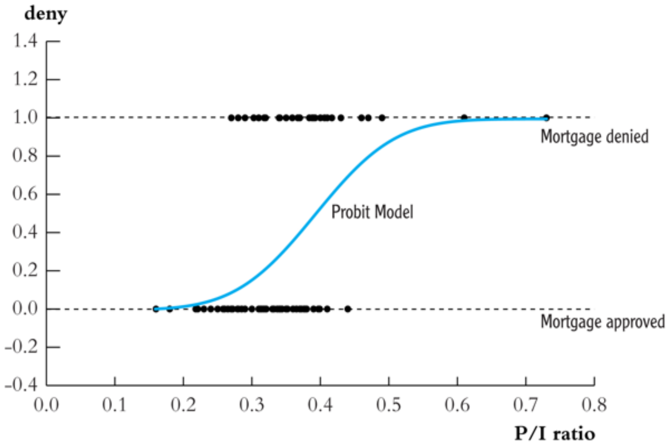
    i. $Pr(Y = 1|X)$ sea creciente en X para $\beta_1 > 0$, y

    ii. $0 \leq Pr(Y = 1|X) \leq 1$ para todos los valores de X

- Esto requiere una función de distribución acumulada, que garantiza que para cualquier valor de los parámetros y de X define probabilidades, con valores en el intervalo [0,1].

- Una posibilidad es utilizar una con forma de "S."

Una forma funcional "S"



Modelo Probit

- El modelo Probit modela la probabilidad de $Y = 1$ usando la función de distribución acumulada de una distribución normal estándar: $\Phi(z)$, evaluada en $z = \beta_0 + \beta_1 X$.

El modelo Probit puede ser expresado como,

$Pr(Y = 1|X) = \Phi(\beta_0 + \beta_1 X)$

- donde $\Phi(.)$ es la función de densidad normal acumulada y $z = \beta_0 + \beta_1 X$ es el $z-$valor o $z - index$ de un modelo probit.

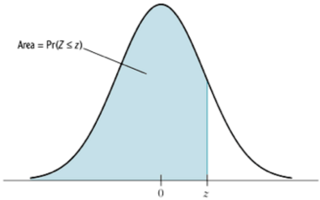
- Ejemplo: Supongamos $\beta_0 = -2$, $\beta_1 = 3$, $X = 0,4$, entonces

$Pr(Y = 1|X = .4) = \Phi(-2 + 3 * 0,4) = \Phi(-0,8)$

$Pr(Y = 1|X = .4) =$ área bajo la función de densidad a la izquierda de $z = -.8$, que es ...

Modelo Probit

TABLE 1 The Cumulative Standard Normal Distribution Function, $\Phi(z) = \Pr(Z \text{ " } z)$



z	Second Decimal Value of z									
	0	1	2	3	4	5	6	7	8	9
−2.9	0.0019	0.0018	0.0018	0.0017	0.0016	0.0016	0.0015	0.0015	0.0014	0.0014
−2.8	0.0026	0.0025	0.0024	0.0023	0.0023	0.0022	0.0021	0.0021	0.0020	0.0019
−0.8	0.2119	0.2090	0.2061	0.2033	0.2005	0.1977	0.1949	0.1922	0.1894	0.1867
−0.7	0.2420	0.2389	0.2358	0.2327	0.2296	0.2266	0.2236	0.2206	0.2177	0.2148
−0.6	0.2743	0.2709	0.2676	0.2643	0.2611	0.2578	0.2546	0.2514	0.2483	0.2451
−0.5	0.3085	0.3050	0.3015	0.2981	0.2946	0.2912	0.2877	0.2843	0.2810	0.2776
−0.4	0.3446	0.3409	0.3372	0.3336	0.3300	0.3264	0.3228	0.3192	0.3156	0.3121

$$\Pr(z \leq -0.8) = .2119$$

Modelo Logit

El modelo Logit modela la probabilidad de $Y = 1$ dado X, como la función de distribución acumulada para una función de distribución logistica, evaluada en $z = \beta_0 + \beta_1 X$:

$$Pr(Y = 1|X) = F(\beta_0 + \beta_1 X)$$

donde F es la función de distribución acumulada para una logística:

$$F(\beta_0 + \beta_1 X) = \frac{1}{1+e^{-(\beta_0+\beta_1 X)}}$$

Modelo Logit

- Ejemplo: Supongamos $\beta_0 = -2$, $\beta_1 = 3$, $X = 0,4$, entonces
$Pr(Y = 1|X = .4) = \frac{1}{1+e^{-(-3+2*0.4)}} = 0.0998$

¿Por qué preocuparnos del modelo logit si ya tenemos el probit?

- La principal razón es histórica: el modelo logit es computacionalmente menos intensivo pero hoy en día estas ventajas son de menor importancia

- En la practica, los modelos logit y probit son bastante similares y los resultados no dependen de la elección entre uno de ellos.

Interpretación de los coeficientes y pendientes

En contraste con el modelo lineal, en los modelos probit y logit los parámetros no corresponden al efecto marginal sobre la variable dependiente de un cambio en una de las variables de control.

En estos modelos el efecto será:

- en el caso de que x_j sea continua, $\frac{\partial Pr(y=1)}{\partial x_j} = f(\beta x)\beta_j$
- en caso de que x_j sea discreta, $\Delta Pr(y=1) = F(\beta x_1) - F(\beta x_0)$

- donde $f(.)$ y $F(.)$ son las funciones de densidad y de distribución acumulada, respectivamente.

Interpretación de los coeficientes y pendientes

- Específicamente con $z = \beta_0 + \beta_1 x_1 + \ldots + \beta_k x_k$

  Logit:

  - $f(z) = \frac{e^{-z}}{(1+e^{-z})^2}$
  - $F(z) = \frac{1}{1+e^{-z}}$

  Probit:

  - $f(z) = \phi(z)$
  - $F(z) = \Phi(z)$

Estimación e Inferencia en los modelos Logit y Probit

Nos centraremos en el modelo Probit:

$$Pr(Y=1|X) = \Phi(\beta_0 + \beta_1 X)$$

Podríamos utilizar mínimos cuadrados no lineales. Sin embargo, un estimador más eficiente (menor varianza) es **el estimador de Máxima Verosimilitud**

El estimador de Máxima Verosimilitud de los coeficientes en el modelo probit

- La función de máxima verosimilitud es la densidad condicional de $Y_1, \ldots, Y_n$ dado $X_1, \ldots, X_n$, como función de los parametros desconocidos (β's)
- El estimador de máxima verosimilitud es (EMV) es el valor de β's que maximiza la función de máxima verosimilitud.
- El EMV es el valor de β's que mejor describe la distribución de los datos.
- En muestras grandes, el EMV es:
    - consistente
    - normalmente distribuido
    - eficiente (tiene la menor varianza entre todos los estimadores)

El estimador de Máxima Verosimilitud de los coeficientes en el modelo probit

Datos: $Y_1, \ldots, Y_n$, i.i.d.

La derivación de la verosimilitud parte por definir la densidad de Y_1:

$Pr(Y_1 = 1|X) = \Phi(\beta_0 + \beta_1 X_1)$ y $Pr(Y_1 = 0) = (1 - \Phi(\beta_0 + \beta_1 X_1))$,
entonces

$$Pr(Y_1 = y_1|X_1) = \Phi(\beta_0 + \beta_1 X_1)^{y_1} * (1 - \Phi(\beta_0 + \beta_1 X_1))^{(1-y_1)} \qquad y_1 = 1, 0$$

$$Pr(Y_1 = y_1|X_1) = \Phi(z_1)^{y_1} * (1 - \Phi(z_1))^{(1-y_1)}$$

con $z_1 = \beta_0 + \beta_1 X_1$

El estimador de Máxima Verosimilitud. Probit

Función de verosimilitud para un modelo probit, la densidad conjunta de $Y_1, \ldots, Y_n$ dado $X_1, \ldots, X_n$, como función de los β:

$$f(\beta; Y_1, \ldots, Y_n | X_1, \ldots, X_n) = \{\Phi(z_1)^{y_1} * (1 - \Phi(z_1))^{(1-y_1)}\}\{\Phi(z_2)^{y_2} * (1 - \Phi(z_2))^{(1-y_2)}\} \\ \ldots \{\Phi(z_n)^{y_n} * (1 - \Phi(z_n))^{(1-y_n)}\}$$

- $\widehat{\beta}^{EMV}$ maximiza la función de verosimilitud.

- pero este valor no tiene expresión explícita! Entonces el EMV debe ser resuelto por métodos numéricos

- En muestras grandes:

- $\widehat{\beta} s^{EMV}$, son consistentes

- $\widehat{\beta} s^{EMV}$, están normalmente distribuidos

- $\widehat{\beta} s^{EMV}$, son asintóticamente eficientes entre todos los estimadores (bajo el supuesto de que el modelo probit es el correcto)

El estimador de Máxima Verosimilitud. Probit

- Los errores estándar def $\widehat{\beta} s^{EMV}$ son calculados automáticamente
- Para contrastes e intervalos de confianza se procede de la forma usual
- Todo se hace extensivo para múltiples X's

El estimador de Máxima Verosimilitud. Logit

- La única diferencia entre los modelos probit y logit es la forma funcional para la probabilidad: Φ es remplazada por la función acumulada para una función logística. Salvo por esto, la verosimilitud es igual
- Al igual que en el modelo probit,
    - $\widehat{\beta} s^{EMV}$ son consistentes
    - Sus errores estándar pueden ser calculados
    - Contrastes e intervalos de confianza se procede de la forma usual

Medidas de bondad del ajuste para los modelos Logit y Probit

- El R^2 y el $\bar{R}^2$ no tienen sentido. Otras dos medidas son generalmente utilizadas:

- La **porcentaje (o proporción) de predicciones correctas** $=$ fracción de Y's para los cuales la probabilidad predicha es $> 50\%$ cuando $Y_i = 1$, o es $< 50\%$ cuando $Y_i = 0$. Donde la elección del umbral de 50% es arbitraria.

- El **pseudo-**R^2: mide el grado de mejora en el ajuste del modelo del log de la verosimilitud respecto al modelo sin Xs.

Comandos básicos en `gretl` para la estimación de modelos probit y logit

- `probit`: estimación por máxima verosimilitud de un modelo probit
- `omit/add`: contraste de significación conjunta
- `$yhat`: probabilidad predicha
- `$lnl`: log de la verosimilitud del modelo estimado
- `pdf(N,z)`: entrega la densidad para una distribución normal estándar
- `cdf(N,z)`: entrega la distribución acumulada para una distribución normal estándar
- `logit`: estimación por máxima verosimilitud de un modelo logit

probit *depvar* *indvars* --robust --verbose --p-values

- *depvar* debe ser una variable binaria $\{0, 1\}$ (en caso contrario, otro modelo debe ser estimado o recibiremos un mensaje de error)
- las pendientes son evaluadas en las medias de las variables
- por defecto, los errores son calculados usando el inverso del Hessiano
- la salida reporta el estadístico χ^2_q para el contraste conjunto de todas las pendientes iguales a cero
- options:
  1. `--robust`: matriz de covarianzas robustas
  2. `--p-values`: reporta el p-valor en vez de las estimaciones de las pendientes.
  3. `--verbose`: muestra información acerca de las iteraciones numéricas

Impacto de la fertilidad sobre la participación laboral femenina

Usando los datos contenidos en el archivo `fertility.gdt` :

- estime un MPL que explique si una mujer ha trabajado o no durante el último año como función de las variables `morekids`, `agem1`, `black`, `hispan`, y `othrace`. Interprete los parámetros.
- Utilizando el modelo anterior, ¿Cuál es el impacto sobre la probabilidad de trabajar asociado a que una mujer tenga más de dos hijos?
- Utilizando el modelo anterior y asumiendo que la edad de la madre es una variable continua, ¿Cuál es el impacto sobre la probabilidad de trabajar asociado a un cambio marginal en la educación de la madre?
- Responda las preguntas anteriores utilizando un modelo probit y logit.