Modelos de elección discreta

Gustavo A. García

ggarci24@eafit.edu.co

Econometría II

Programa de Economía

Universidad EAFIT

Link slides formato html

Link slides formato PDF

En este tema

- Motivación
- Economía y los modelos de elección discreta
- Modelo de probabilidad lineal (MPL)
- Limitaciones del MPL
- Modelos logit y probit
- Qué modelo seleccionar entre el MPL, probit o logit? Preguntémosle a Wooldridge
- Ejercicio aplicado en R: Efectos de la educación sobre el crimen

Lecturas

- Wooldridge, J. (2013). *Introducción a la econometría. Un enfoque moderno*. 5a edición. Cenagage Learning Sección 7.5, Cap 17
- Gujarati, D. y Porter, D. (2010). *Econometría*. 5a edición, Mc Graw Hill. Cap. 15

Motivación

- En los modelos vistos hasta ahora, la variable dependiente y ha tenido un significado *cuantitativo* (por ejemplo, y ha sido una cantidad de dólares, la puntuación en un exámen, un porcentaje, o el logaritmo de éstos). La pregunta que surge es: ¿Qué ocurre cuando se desea usar la regresión múltiple para explicar eventos cualitativos?

- En el caso más sencillo, y que en la práctica se encuentra con frecuencia, el evento que se desea explicar tiene un resultado binario

- Respuesta binaria: la variable y sólo toma los valores 0 y 1. Por ejemplo, y puede indicar

  - si un adulto tiene o no educación universitaria
  - si un estudiante universitario consume o no drogas durante un año escolar
  - si una empresa ha sido absorbida por otra durante un determinado año
  - si el individuo está o no desempleado
  - si el individuo es o no un trabajador informal
  - si el individuo toma transporte público o no para desplazarse a su empleo

Economía y los modelos de elección discreta

La interpretación económica de los modelos de elección discreta parte de la utilidad, donde se especifica que la racionalidad de los agentes económicos hace que se comporten de forma que maximicen la utilidad esperada que les proporciona cada una de las opciones posibles sobre lo que deben decidir

Ejemplos:

Participación laboral: los individuos deciden trabajar si la utilidad que le proporciona esa elección supera la utilidad de no hacerlo (una vez considerados los beneficios y costos de la elección)

Lanzamiento de un nuevo producto: un empresario decidirá lanzar un nuevo producto al mercado si la utilidad que le proporciona este hecho supera a la de no realizarlo

Modo de transporte: los individuos deciden ir a trabajar en transporte público si la utilidad que le proporciona este medio es superior a la utilidad que le reporta transportarse por medio privado

Economía y los modelos de elección discreta

- La formalización teórica parte del supuesto de que la utilidad derivada de una elección, U_{i1} o U_{i2}, es función de las variables explicativas de dicha decisión $\Longrightarrow$ características propias de cada una de las alternativas de elección y de las características personales socioeconómicas y culturales propias del individuo

- Igualmente existe una perturbación aleatoria ϵ_{ij} que recoge las desviaciones que los agentes tienen respecto a lo que sería el comportamiento del agente medio

En este caso el problema de decisión binaria ($j = 0, 1$) se puede plantear en los siguientes términos:
U_{i0}: utilidad que le proporciona al agente i la elección 0
U_{i1}: utilidad que le proporciona al agente i la elección 1
X_{i0}: vector de variables explicativas que caracterizan la elección de la alternativa 0
X_{i1}: vector de variables explicativas que caracterizan la elección de la alternativa 1

Suponiendo, además, linealidad en las funciones, implicaría que:

$$U_{i0} = \overline{U}_{i0} + \epsilon_{i0} = \alpha_0 + X_{i0}\beta + \epsilon_{i0}$$

$$U_{i1} = \overline{U}_{i1} + \epsilon_{i0} = \alpha_1 + X_{i1}\beta + \epsilon_{i1}$$

$\overline{U}_{ij}$ representa las utilidades medias, que pueden ser observadas y son función de una combinación lineal de las variables explicativas observadas: $X_{ij}\beta$

ϵ_{ij} representan aquellos factores de la utilidad asociada a cada una de las alternativas que son desconocidas y que pueden variar según los individuos y según la alternativa

Economía y los modelos de elección discreta

El agente i elegirá la opción 1 si la utilidad de esa elección supera la de la opción 0 y viceversa. Es decir:

$$Y_i = \begin{cases} 1 \text{ si } U_{i1} > U_{i0} \\ 0 \text{ si } U_{i0} > U_{i1} \end{cases}$$

Como consecuencia de ello, se puede comprobar que la probabilidad de que un individuo elija la ópción 1 será:

$$\begin{aligned} \text{Prob}(Y_i = 1) &= \text{Prob}(U_{i1} > U_{i0}) \\ &= \text{Prob}(\overline{U}_{i1} + \epsilon_{i1} > \overline{U}_{i0} + \epsilon_{i0}) \\ &= \text{Prob}(\epsilon_{i0} - \epsilon_{i1} < \overline{U}_{i1} - \overline{U}_{i0}) \\ &= \text{Prob}(\epsilon_{i0} - \epsilon_{i1} < (\alpha_1 - \alpha_0) + \beta(X_{i1} - X_{i0})) \\ &= F(X_i\beta) \end{aligned}$$

A través de la anterior ecuación se determina la probabilidad de que un individuo elija la opción 1, que depende de la distancia entre utilidades ($\overline{U}_{i1} - \overline{U}_{i0}$)

Dicha probabilidad viene dada por el valor de la función de distribución F en el punto $X_i\beta$, es decir, $F(X_i\beta)$. Dependiendo al supuesto sobre la función de distribución se llega a diferentes modelos:

- Modelo de probabilidad lineal (MPL): función de distribución uniforme
- Modelo Probit: función de distribución normal tipificada ($N(0,1)$)
- Modelo Logit: función de distribución logística

Modelo de probabilidad lineal (MPL)

¿Qué significa escribir un modelo de regresión múltiple como el siguiente

$$y = \beta_0 + \beta_1 x_1 + \ldots + \beta_k x_k + u$$

cuando y es una variable bivariada?

Como y sólo puede tomar dos valores, los β_j no pueden interpretarse como el cambio en y dado una variación de x_j

El punto clave es que cuando y es una variable binaria que toma los valores 0 y 1, entonces se tiene que:

$$\begin{aligned} \mathrm{E}(y|\mathbf{x}) &= 0 \cdot \mathrm{P}(y = 0|\mathbf{x}) + 1 \cdot \mathrm{P}(y = 1|\mathbf{x}) \\ &= \mathrm{P}(y = 1|\mathbf{x}) \Longrightarrow \text{la probabilidad de éxito} \end{aligned}$$

Con lo cual se llega a la ecuación:

$$E(y|\mathbf{x}) = P(y = 1|\mathbf{x}) = \beta_0 + \beta_1 x_1 + \ldots + \beta_k x_k$$

que indica que la probabilidad de éxito, es decir, $p(\mathbf{x}) = P(y = 1|\mathbf{x})$, es función lineal de las variables x_j

La anterior ecuación es un ejemplo de un modelo de respuesta bivariada y a $P(y = 1|\mathbf{x})$ también se le conoce como la probabilidad de respuesta

Limitaciones del MPL

A un modelo de regresión líneal múltiple en el que la variable dependiente es una variable binaria, se le conoce como modelo de probabilidad lineal (MPL), ya que la probabilidad de respuesta es lineal en los parámetros β_j

β_j mide la variación de la probabilidad de éxito al variar x_j, permaneciendo los demás valores constantes:

$$\Delta \mathrm{P}(y = 1|\mathbf{x}) = \beta_j \Delta x_j$$

Del análisis anterior parece que los MCO pueden extenderse sin dificultad a modelos de regresión con variable dependiente binaria. Sin embargo, no es tan inmediato, ya que el MPL plantea diversos problemas:

- El R^2 no es adecuado como medida de bondad de ajuste
- No normalidad de las perturbaciones u_i
- Varianzas heteroscedásticas de las perturbaciones
- No cumplimiento de $0 \leq \mathrm{E}(y|\mathbf{x}) \leq 1$
- La probabilidad está relacionada en forma lineal con las variables independientes para todos los valores posibles

Limitaciones del MPL

El R^2 no es adecuado como medida de bondad de ajuste

El R^2 de manera convencional tiende a estar subestimado ya que la SCR es más grande de lo habitual

No normalidad de las perturbaciones u_i

El supuesto de normalidad en los u_i es necesario para fines de inferencia estadística

Sin embargo, este supuesto no se mantienen en los MPL ya que, al igual que y, u_i sólo toma dos valores, es decir, que sigue una distribución de Bernoulli

Pero el no cumplimiento del supuesto de normalidad quizá no sea tan crítico como parece, ya que las estimaciones puntuales de MCO aún permanecen insesgadas

Además, puede demostrarse que, conforme el tamaño de la muestra aumenta indefinidamente, los estimadores MCO tienden a tener una distribución normal, así que en muestras grandes, la inferencia estadística del MPL será valida

Limitaciones del MPL

Varianzas heteroscedásticas de las perturbaciones

Aunque no hay correlación serial ($E(u_i) = 0$ y $Cov(u_i, u_j) = 0$ para $i \neq j$), no es posible sostener la afirmación de que las perturbaciones en el MPL son homoscedásticas

Como demuestra la teoría estadística, para una distribución de Bernoulli, la media y la varianza teóricas son $P(y = 1)$ y $P(y = 1)(1 - P(y = 1))$ respectivamente, lo cual revela que la varianza es una función de la media, por lo que es heteroscedástica

Se tiene entonces que:

$$Var(u_i) = P(y = 1)(1 - P(y = 1))$$

Como $P(y = 1) = \beta_0 + \beta_1 x_1 + \ldots + \beta_k x_k$, la varianza de u_i depende de los valores de x y por tanto no es homoscedástica

Recordemos que una forma de resolver el problema de heteroscedásticidad es transformar el modelo ponderandolo por un factor de corrección. En función de esta corrección se podría proceder como sigue:

Paso 1: Estimar el modelo y obtener $\widehat{y_i}$ para calcular el ponderado $\widehat{w}_i = \widehat{y_i}(1 - \widehat{y_i})$
Paso 2: Con $\widehat{w}_i$ se ponderan los datos y se estima el modelo transformado por MCO

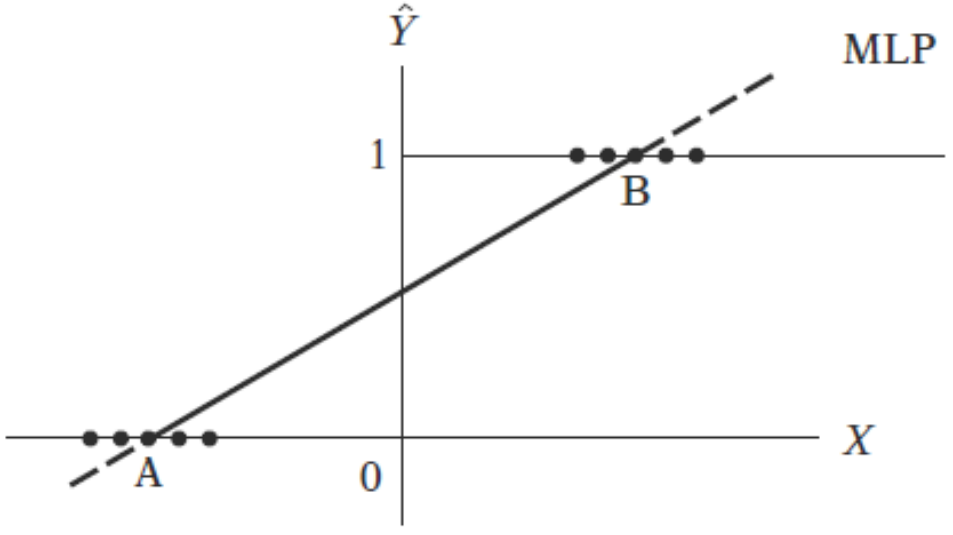
También es posible usar errores estándar corregidos por heteroscedásticidad de White para resolver la heteroscedásticidad, siempre que la muestra sea grande

Limitaciones del MPL

No cumplimiento de $0 \leq \mathrm{E}(y|\mathbf{x}) \leq 1$

Como $E(y|\mathbf{x})$ en los MPL mide la probabilidad condicional de que ocurra el suceso y dado $\mathbf{x}$, ésta debe encontrarse necesariamente entre 0 y 1

Aunque a priori esto es verdadero, no hay garantia de que $\hat{y}$, los estimadores de $\mathrm{E}(y|\mathbf{x})$, cumplan necesariamente esta restricción, y éste es el verdadero problema con la estimación del MPL por MCO



Limitaciones del MPL

$P(y = 1|\mathbf{x}) = E(y = 1|\mathbf{x})$ aumenta linealmente con $\mathbf{x}$

El problema fundamental del MPL es que $P(y = 1|\mathbf{x}) = E(y = 1|\mathbf{x})$ aumenta linealmente con $\mathbf{x}$, es decir, el efecto marginal o incremental de $\mathbf{x}$ permanece constante para cualquier valor de la variable

Por ejemplo, en el análisis sobre la probabilidad ser propietario o no de una vivienda ($y = 1$ si es propietario y $y = 0$ si no es propietario), un determinante importante es el nivel de ingreso

En el MPL cuando el ingreso aumenta marginalmente, la probabilidad de ser propietario de una casa aumenta en la misma cantidad constante

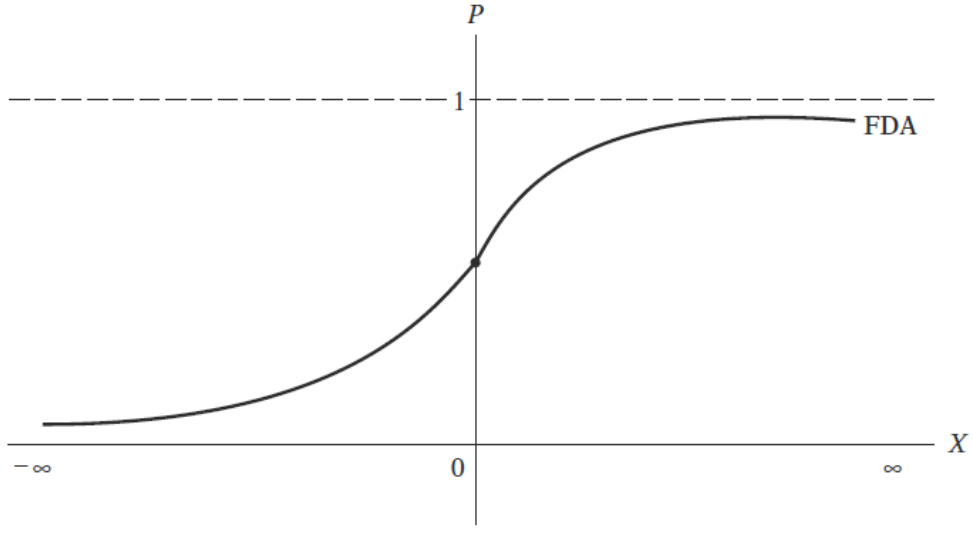
Sin embargo, esto no ocurre en la realidad, ya que se espera que la probabilidad de ser propietario se encuentre relacionado en forma no lineal con el ingreso: familias con ingresos bajos tienen menor probabilidad de ser propietarios de una vivienda comparadas a familias con mayores ingresos, que son más propensas a tener viviendas propias

Limitaciones del MPL

Por consiguiente, lo que necesitamos es un modelo (probabilístico) que tenga estas dos características:

- a medida que una variable X aumente, $P(y = 1|\mathbf{x}) = \mathrm{E}(y = 1|\mathbf{x})$ también aumenta pero no se sale del intervalo $0 - 1$
- la relación entre $P(y = 1|\mathbf{x})$ y alguna X es no lineal, es decir, nos acercamos a probabilidades de cero a tasas más lentas cuando disminuye alguna X y nos acercamos a probabilidades de uno cada vez más lentas a medida que alguna X se hace más grande

El modelo que deseamos tendría la siguiente forma:



Las funciones de distribución acumuladas (FDA) más comúnmente utilizadas son:

- la logística $\Longrightarrow$ modelo logit
- la normal $\Longrightarrow$ modelo probit

Modelos logit y probit

Especificación

Para evitar las limitaciones del MPL, considere una clase de modelos de respuesta binaria de la forma:

$$P(y = 1|\mathbf{x}) = G(\beta_0 + \beta_1 x_1 + \ldots + \beta_k x_k) = G(\beta_0 + \mathbf{x}\boldsymbol{\beta})$$

donde G es una función que asume $0 < G(z) < 1$, para todos los número reales z, lo cual asegura que las probabilidades de respuesta estimada estén estrictamente entre cero y uno

Las dos funciones no lineales más comúnmente estudiadas son:

El modelo logit:

$$G(z) = \Lambda(z) = \frac{exp(z)}{1 + exp(z)}$$

El modelo probit:

$$G(z) = \Phi(z) = \int_{-\infty}^{z} \phi(v)dv$$

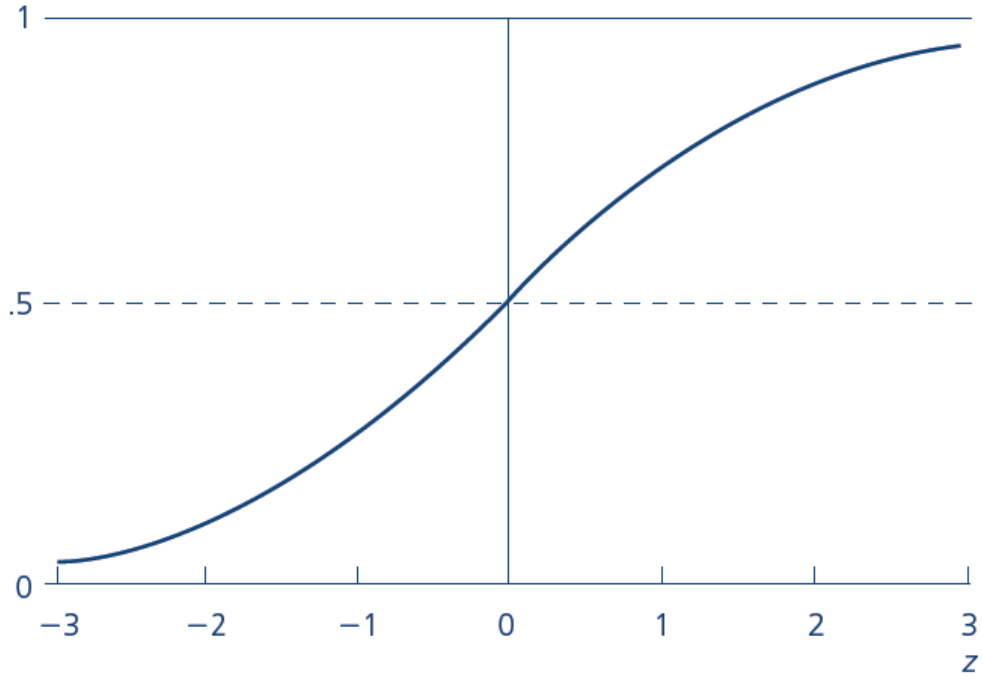
donde $\phi(z)$ es la densidad normal estándar:

$$\phi(z) = (2\pi)^{-1/2} exp(-z^2/2)$$

Modelos logit y probit

Especificación

Gráficamente estas funciones de distribución acumuladas tiene la siguiente forma:



Se observa que son funciones crecientes, aumenta con más rapidez en $z = 0$, $G(z) \to 0$ a medida que $z \to -\infty$, y $G(z) \to 1$ a medida que $z \to \infty$

Modelos logit y probit

Especificación

Los modelos logit y probit pueden derivarse a partir de un modelo de variable latente subyacente. Sea y^* una variable inobservable, o latente, determinada por:

$$y^* = \beta_0 + \mathbf{x}\boldsymbol{\beta} + e, \;\; y = 1[y^* > 0]$$

La función $1[\cdot]$ recibe el nombre de función indicadora, que asume el valor de 1 si el evento dentro de corchetes es verdadero y de 0 si no lo es. Por tanto,

$$y = \begin{cases} 1 \text{ si } y^* > 0 \\ 0 \text{ si } y^* \leq 0 \end{cases}$$

Ejemplo: la decisión de participar en el mercado laboral
Una persona decide participar en el mercado laboral si el salario potencial de mercado (w) es mayor al salario de reserva (w^*), siendo este último el salario mínimo que estaría dispuesto a recibir si decidiera trabajar. Si la valoración del mercado de su tiempo excede el valor implícito del tiempo, el individuo optará por la actividad laboral. Por otra parte, si el individuo otorga un valor mayor a su tiempo que lo que hace el mercado, no optará por la actividad laboral. Note, entonces, que hay dos variables latentes, inobservables, w y w^*, y la variable que sí se observa es la de participación laboral, esto es:

$$\text{Participación laboral} = \begin{cases} 1 \text{ si } w > w^* \\ 0 \text{ si } w \leq w^* \end{cases}$$

Modelos logit y probit

Especificación

Supuestos sobre e

- e es independiente de $\mathbf{x}$
- e tiene la distribución logística estándar o la distribución norma estándar
- lo anterior supone que e se distribuye simétrica en torno a cero, lo cual significa que $1 - G(-z) = G(z)$

A partir de la ecuación de $y^* = \beta_0 + \mathbf{x}\boldsymbol{\beta} + e$ y los supuestos establecidos, se puede calcular la probabilidad de respuesta de y:

$$\begin{aligned} \mathrm{P}(y = 1|\mathbf{x}) &= \mathrm{P}(y^* > 0|\mathbf{x}) \\ &= \mathrm{P}[e > -(\beta_0 + \mathbf{x}\boldsymbol{\beta})|\mathbf{x}] \\ &= 1 - G[-(\beta_0 + \mathbf{x}\boldsymbol{\beta})] \\ &= G(\beta_0 + \mathbf{x}\boldsymbol{\beta}) \end{aligned}$$

Modelos logit y probit

Especificación

En la mayoría de las aplicaciones de los modelos de respuesta binaria, la meta principal es explicar los efectos de las x_j sobre la probabilidad de respuesta $P(y = 1|\mathbf{x})$

En este punto, debe tenerse en cuenta que la magnitud de β_j no son, por sí mismas, útiles (en contraste con el MPL), esto debido a la naturaleza no lineal de $G(\cdot)$. Es decir, que las estimaciones de los β_j en los modelos logit y probit no muestran los efectos de las x_j sobre $P(y = 1|\mathbf{x})$ dada la no linealidad de los modelos

Para calcular el efecto parcial se procede como sigue, dependiendo el tipo de variable

Cuando x_j es una variable continua

$$\frac{\partial P(y = 1|\mathbf{x})}{\partial x_j} = g(\beta_0 + \mathbf{x}\boldsymbol{\beta})\beta_j$$

donde g es la función de densidad ($dG(z)/dz$). En los modelos logit y probit, $G(\cdot)$ es una FDA estrictamente crecientes y, por tanto $g(z) > 0$ para toda z

Por consiguiente, el efecto parcial de x_j sobre $P(y = 1|\mathbf{x})$ depende de $\mathbf{x}$ a través de la cantidad positiva $g(\beta_0 + \mathbf{x}\boldsymbol{\beta})$, lo cual significa que el efecto parcial siempre tiene el mismo signo de β_j

Modelos logit y probit

Especificación

Cuando x_j es una variable binaria

En este caso se estaría calculando el efecto parcial de cambiar x_j de cero a uno. Por ejemplo, asumiendo que x_1 es binaria se tiene:

$$G(\beta_0 + \beta_1 + \beta_2 x_2 + \ldots + \beta_k x_k) - G(\beta_0 + \beta_2 x_2 + \ldots + \beta_k x_k)$$

Por ejemplo, si y es un indicador de ir a trabajo en bicicleta y x_1 es una variable binaria que indica si hay o no cicloruta en el trayecto al trabajo, entonces la anterior ecuación nos muestra el cambio en la probabilidad de utilizar la bici debido a la existencia de la infraestructura de transporte

Observe que saber el signo de β_1 es suficiente para determinar si la intervención o programa tuvo efecto positivo o negativo, pero para calcular la magnitud del efecto, se debe estimar a partir de la anterior ecuación

También se puede usar dicha ecuación para otros tipos de variables discretas, por ejemplo el número de niños. Si x_k denota esta última variable, el efecto sobre la probabilidad de que x_k cambie de c_k a $c_k + 1$ sería:

$$G(\beta_0 + \beta_1 x_1 + \beta_2 x_2 + \ldots + \beta_k (c_k + 1)) - G(\beta_0 + \beta_2 x_2 + \ldots + \beta_k c_k)$$

Modelos logit y probit

Interpretación los efectos marginales

En comparación al MPL, el costo de usar los modelos probit y logit es que los efectos parciales son más difíciles de calcular debido a que el factor de escala, $g(\beta_0 + \mathbf{x}\boldsymbol{\beta})$, depende de $\mathbf{x}$, es decir, de todas las variables explicativas

Una posibilidad es insertar valores interesantes para las x_j, como por ejemplo, las medias, las medianas, los mínimos, los máximos y los cuartiles superiores e inferiores, y ver cómo cambia $g(\beta_0 + \mathbf{x}\boldsymbol{\beta})$

El anterior procedimiento puede ser muy informativo, pero es demasiado tedioso, ya que se debe pensar para x_j que valor asignarle. Para resolver esto, en la literatura (y en los softwares) se calculan dos tipos de efectos parciales

- El efecto parcial en el promedio
- El efecto parcial promedio

Modelos logit y probit

Interpretación los efectos marginales

Efecto parcial en el promedio

La idea aquí es reemplazar cada variable explicativa con su promedio muestral, esto implica que el el efecto marginal para una variable continua sería:

$$\frac{\partial P(y=1|\mathbf{x})}{\partial x_j} = g(\widehat{\beta}_0 + \overline{\mathbf{x}}\widehat{\boldsymbol{\beta}})\widehat{\beta}_j = g(\widehat{\beta}_0 + \widehat{\beta}_1\overline{x}_1 + \widehat{\beta}_2\overline{x}_2 + \ldots + \widehat{\beta}_k\overline{x}_k)\widehat{\beta}_j$$

Y si x_1 fuera una variable discreta sería:

$$G(\widehat{\beta}_0 + \widehat{\beta}_1 + \widehat{\beta}_2\overline{x}_2 + \ldots + \widehat{\beta}_k\overline{x}_k) - G(\widehat{\beta}_0 + \widehat{\beta}_2\overline{x}_2 + \ldots + \widehat{\beta}_k\overline{x}_k)$$

Se obtiene entonces el efecto parcial de x_j para la persona promedio en la muestra

Limitaciones

- Si alguna de las variables explicativas son discretas, sus promedios no representan a nadie en la muestra. Por ejemplo, si $x_1 = 1$ si es mujer y el 47.5% de la muestra son mujeres, no tiene sentido incorporar $\overline{x} = 0.475$. Una posibilidad es fijar las variables binarias en 1 o 0
- Si una variable explicativa continua aparece como una función no lineal, por ejemplo, como un log o en una forma cuadrática, no es claro si se quiere promediar la función no lineal o insertar el promedio en la función no lineal. Los softwares se quedan con la primera opción

Modelos logit y probit

Interpretación los efectos marginales

Efecto parcial promedio

Bajo este método se evita pensar sobre qué valores insertar para las variables explicativas. Aquí la idea, entonces, es que se promedia los efectos parciales individuales a través de la muestra. Para una variable continua sería:

$$\left[n^{-1} \sum_{i=1}^{n} g(\widehat{\beta}_0 + \mathbf{x}_i \widehat{\boldsymbol{\beta}}) \right] \widehat{\beta}_j$$

Para el caso en que x_1 sea discreta, se tiene:

$$n^{-1} \sum_{i=1}^{n} [G(\widehat{\beta}_0 + \widehat{\beta} + \widehat{\beta}_2 x_2 + \ldots + \widehat{\beta}_k x_k) - G(\widehat{\beta}_0 + \widehat{\beta}_2 x_2 + \ldots + \widehat{\beta}_k x_k)]$$

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

Modelos logit y probit

Estimación por máxima verosimilitud

Debido a la naturaleza no lineal de los modelos logit y probit la estimación se realiza por máxima verosimilitud

Para obtener el estimador de máxima verosimilitud, condicional sobre las variables explicativas, se necesita la densidad de y_i dada $\mathbf{x}_i$. Esto se puede escribir como:

$$f(y|\mathbf{x}_i;\boldsymbol{\beta}) = [G(\mathbf{x}_i\boldsymbol{\beta})]^y[1 - G(\mathbf{x}_i\boldsymbol{\beta})]^{1-y}, \ \text{con} \ \ y = 0, 1$$

Se puede ver con facilidad que cuando $y = 1$, se obtiene $G(\mathbf{x}_i\boldsymbol{\beta})$ y cuando $y = 0$, se obtiene $1 - G(\mathbf{x}_i\boldsymbol{\beta})$. La función de log-verosimilitud para la observación i es una función de los parámetros y los datos ($\mathbf{x}_i, y_i$) y se obtiene al aplicar log a la ecuación de arriba, esto es:

$$\ell_i(\boldsymbol{\beta}) = y_i log[G(\mathbf{x}_i\boldsymbol{\beta})] + (1 - y_i)log[1 - G(\mathbf{x}_i\boldsymbol{\beta})]$$

El log-verosimilitud para un tamaño de muestra n se obtiene al sumar a través de todas las observaciones:

$$\mathcal{L}(\boldsymbol{\beta}) = \sum_{i=1}^{n} \ell_i(\boldsymbol{\beta})$$

La estimación máxima verosímil de $\boldsymbol{\beta}$, denotada como $\widehat{\boldsymbol{\beta}}$, maximiza esta log-verosimilitud

Modelos logit y probit

Medidas de bondad de ajuste

El R^2 de McFadden o pseudo R^2

Dado que el R^2 convencional no es adecuado para medir la bondad de ajuste del modelo, la literatura propone el siguiente estadístico:

$$\text{pseudo}R^2 = 1 - \frac{\ln L_{NR}}{\ln L_R}$$

donde lnL_R es el log de la función de verosimilitud del modelo restringido bajo la hipótesis nula: $H_0 : \beta_2 = \beta_3 = \ldots = \beta_k = 0$, y lnL_{NR} es log de la función de verosimilitud para el modelo no restringido

Modelos logit y probit

Medidas de bondad de ajuste

Proporción de predicciones correctas

Una medida de bondad de ajuste puede ser considerar el porcentaje de predicciones correctas que proporciona la estimación. Para ello, consideramos un valor verdadero de y_i y el obtenido a partir de la estimación o predicción $\hat{y}_i$ de forma que:

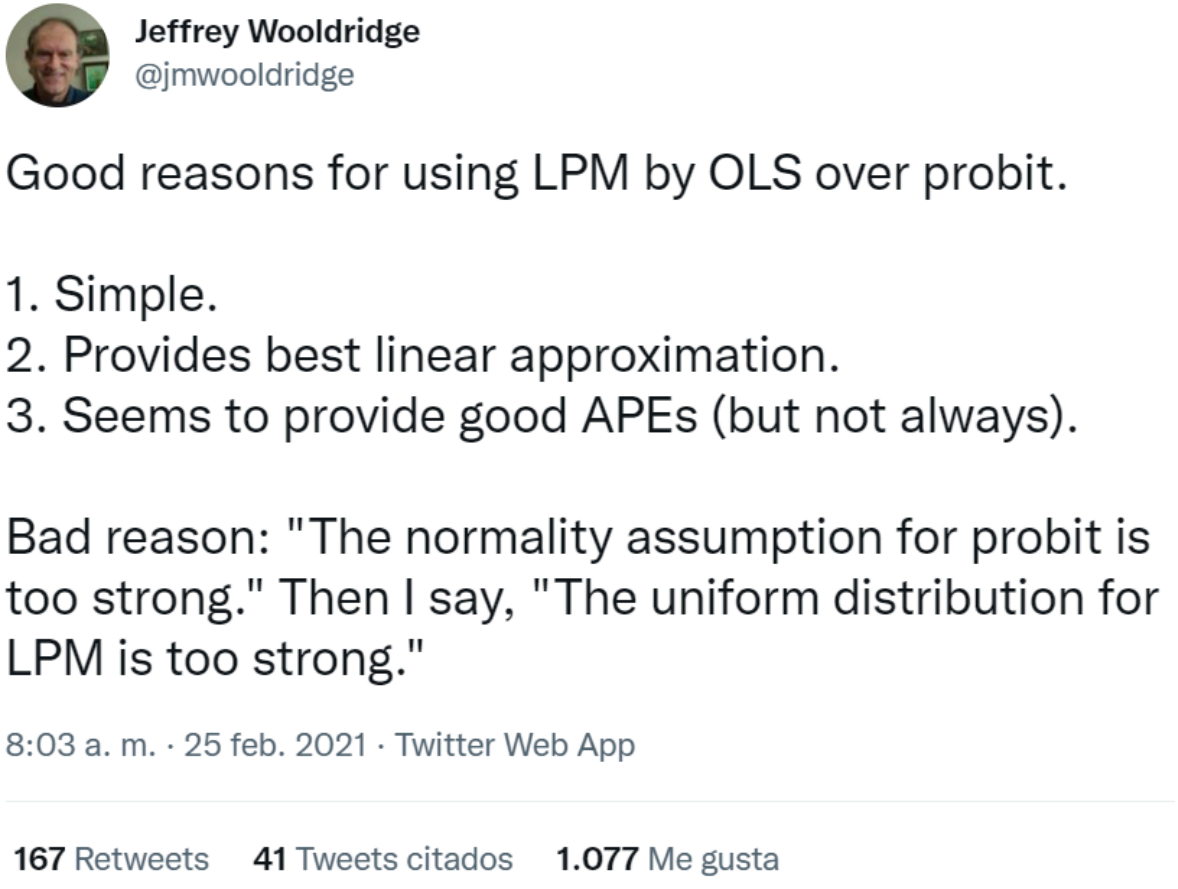
		Valor real de y_i	
		$y_i = 0$	$y_i = 1$
Predicción de $\widehat{y}_i$	$\widehat{y}_i < c$	Predicción correcta (I_1)	Predicción errónea (I_2)
	$\widehat{y}_i > c$	Predicción errónea (I_3)	Predicción correcta (I_4)

La idea es obtener la frecuencia con la que le modelo predice correctamente $y_i = 1$ y $y_i = 0$, así como la proporción de predicciones correctas en general. Así el porcentaje de predicciones correctas será

$$\% \text{ predicciones correctas} = \frac{\text{Predicciones correctas}}{\text{Frecuencia total}} = \frac{I_1 + I_4}{I_1 + I_2 + I_3 + I_4}$$

Por lo general se define el umbral $c = 0.5$ pero tiene criticas, por ejemplo, cuando uno de los resultados es poco probable: Por ejemplo, si $\overline{y} = 0.08$ (sólo 8% de éxitos en la muestra) podría ser que nunca se prediga $y_i = 1$ debido a que la probabilidad estimada de éxito nunca es mayor que 0.5. Una alternativa es usar la fracción de éxitos en la muestra como umbral: 0.08

Qué modelo seleccionar entre el MPL y probit o logit? Preguntémosle a Wooldridge



Jeffrey Wooldridge
@jmwooldridge

Good reasons for using LPM by OLS over probit.

1. Simple.
2. Provides best linear approximation.
3. Seems to provide good APEs (but not always).

Bad reason: "The normality assumption for probit is too strong." Then I say, "The uniform distribution for LPM is too strong."

8:03 a. m. · 25 feb. 2021 · Twitter Web App

167 Retweets **41** Tweets citados **1.077** Me gusta

Link al tweet

Ejercicio aplicado: Efectos de la educación sobre el crimen

En este ejercicio aplicado se va a analizar los efectos de la educación sobre la probabilidad de ir prisión. Esta aplicación se base en el paper: Lochner, L. y Moretti, E. (2004). "The Effect of Education on Crime: Evidence from Prison Inmates, Arrests, and Self-Reports", *American Economic Review*, 94(1):155-189 (http://www.nber.org/papers/w8605)

Las principales variables a analizar son:

- prision: variable binaria igual a 1 si la persona está en prisión, y 0 no
- educ: años de escolaridad
- age: edad
- AfAm: variable binaria igual a 1 para afroamericano, 0 no

En los siguientes links se encuentran los datos y el código utilizado en R:

- Datos
- Código en R