Modelos de elección binaria

Prof.: Begoña Álvarez García

Econometría II 2007-2008

Estamos interesados en la ocurrencia o no-ocurrencia de un cierto evento (ej: participación en el mercado laboral; inversión de una empresa en I+D; matriculación en un máster, etc.) y en cómo ese evento depende de un vector de k variables explicativas $x_i = (1, x_{2i}, ..., x_{ki})'$; disponemos para ello de n observaciones. Para formalizar el evento que es objeto de estudio, definimos una variable binaria y_i tal que

$$\begin{cases} y_i = 1 \text{ si el evento} \\ y_i = 0 \text{ si el evento no ocurre} \end{cases}$$

Supondremos que la variable dependiente es aleatoria y que los regresores son no aleatorios.

En principio, podría utilizarse un modelo de regresión lineal para modelar estas decisiones. Sin embargo, como veremos en el apartado siguiente, el uso del modelo de regresión lineal cuando la variable dependiente es discreta presenta diversos problemas y, en la mayoría de los casos, no es el más adecuado para estudiar estos datos. Como alternativa se han formulado otro tipo de modelos conocidos con el nombre de "modelos de elección binaria".

1. MODELO DE PROBABILIDAD LINEAL

Podríamos empezar planteando un modelo de regresión lineal para estudiar esta relación

$$y_i = x_i'\beta + \varepsilon_i \qquad i = 1, ..., n \tag{1}$$

siendo β un vector $k-$dimensional de parámetros desconocidos y u_i un término de error con esperanza nula. Ahora bien, como la variable dependiente toma sólo valor 1 o 0, formular una relación de este tipo tiene diversas consecuencias, la mayoría no deseables, que examinamos a continuación:

1. Como la variable dependiente y_i sólo puede valer 1 o 0, entonces

$$E(y_i) = \Pr(y_i = 1) \times 1 + \Pr(y_i = 0) \times 0 = \Pr(y_i = 1)$$

   Al formular una relación lineal como (1) donde $E(\varepsilon_i) = 0$, $\forall i$, tendremos que

$$E(y_i) = x_i'\beta \quad \Longrightarrow \quad \Pr(y_i = 1) = x_i'\beta$$

   Es decir, decimos que la probabilidad de que la variable dependiente tome el valor 1 depende linealmente de los regresores. Por esta razón, cuando la variable dependiente es binaria, al modelo de regresión se le denomina "modelo de probabilidad lineal". La primera consecuencia de este resultado es que, como las probabilidades han de estar siempre en el intervalo $[0, 1]$, entonces $0 \leq x_i'\beta \leq 1$, $\forall i$; luego, no todos los vectores de β en R^k son válidos, sino sólo aquéllos para los cuales se satisfacen estas n desigualdades.

2. Las predicciones que se obtienen con el modelo de probabilidad lineal pueden salirse fuera del rango observado de valores $[0, 1]$. Supongamos que, para un determinado vector de variables explicativas x_{n+1}' y un vector de estimaciones $\hat{\beta}$ del vector de parámetros β, queremos predecir el valor de la variable dependiente. Parece lógico calcular esta predicción como

$$\hat{y}_{n+1} = x_{n+1}'\hat{\beta}$$

   Es evidente que, en general, el valor predicho no será 0 ni 1. Pero, en realidad, la predicción debe interpretarse como una estimación de $\Pr(y_{n+1} = 1)$ (¿por qué?). La cuestión es que este método no nos garantiza que esas predicciones vayan a estar acotadas en el intervalo $[0, 1]$ y, por tanto, podríamos obtener predicciones absurdas.

3. Al ser binaria la variable dependiente, el término de error u_i del modelo (1) es hete-

rocedástico. Veámoslo:

$$\begin{cases} u_i = 1 - x_i'\beta & \text{cuando } y_i = 1 \\ u_i = -x_i'\beta & \text{cuando } y_i = 0 \end{cases}$$

por tanto

$$\begin{aligned} Var(u_i) &= E(u_i^2) = (1 - x_i'\beta)^2 \times \Pr(y_i = 1) + (-x_i'\beta)^2 \times \Pr(y_i = 0) = \\ &= (1 - x_i'\beta)^2 (x_i'\beta) + (-x_i'\beta)^2 (1 - x_i'\beta) = \\ &= (1 - x_i'\beta)(x_i'\beta) \left[(1 - x_i'\beta) + (-x_i'\beta) \right] = \\ &= (1 - x_i'\beta)(x_i'\beta) \end{aligned}$$

que, en general, es diferente para cada observación. Por lo tanto, la estimación por MCO del modelo (1) es ineficiente y, lo que es más grave, no permite hacer inferencias válidas (¿por qué?). Este último problema podría resolverse utilizando Mínimos Cuadrados Generalizados Factibles. Pero también existen problemas.

Veamos cuáles serían los pasos para estimar el modelo por Mínimos Cuadrados Generalizados:

- Se estima el modelo (1) por MCO, sin tener en cuenta la heterocedasticidad. Se obtienen las predicciones $\hat{y}_i$ para toda la muestra (recordar que estas predicciones representan estimaciones de la probabilidad condicional de que $y_i = 1$)
- Las estimaciones $\hat{y}_i$ se utilizan para estimar la varianza de las perturbaciones aleatorias para cada individuo mediante la expresión $\hat{w}_i = \hat{y}_i(1 - \hat{y}_i)$.
- Si los valores estimados $\hat{y}_i$ son mayores que la unidad, se deben sustituir por la unidad (si son menores que 0, se sustituyen por 0), para que sean coherentes. Pero esto provoca otro problema: los valores correspondientes de $\hat{w}_i$ en esos casos serían 0, y como estos valores se utilizan como ponderadores en el método de estimación por MCG, eso supone que tendríamos que dividir los valores de las variables del modelo por 0. Para evitar estos problemas se puede proceder de dos formas:

  - Se eliminan estas observaciones de la muestra y, por tanto, se pierde información.
  - Se sustituyen los valores mayores o iguales a la unidad por 0.999 y los menores o iguales a cero por 0.001.

- Se pondera el modelo (1) dividiendo ambos miembros de la ecuación por $\sqrt{\hat{w}_i} = \sqrt{\hat{y}_i(1-\hat{y}_i)}$, con el fin de transformar el modelo en homocedástico.

Finalmente se estima el modelo transformado por MCO, y ésas serán las estimaciones por MCG. Las estimaciones son eficientes. Sin embargo queda sin resolver un problema importante: las probabilidades predichas a partir de las estimaciones MCG no tienen por qué estar acotadas en [0,1].

Todos estos inconvenientes vienen originados porque, en realidad, la formulación misma del modelo de probabilidad no es totalmente lógica: las funciones lineales no están acotadas ni inferior ni superiormente, mientras que las probabilidades sí lo están.

A continuación, vamos a revisar otros modelos de elección binaria que no presentan estos problemas.

2 MODELOS DE ELECCIÓN BINARIA: PROBIT Y LOGIT

Dado que la mayoría de problemas que nos encontramos con el modelo de probabilidad lineal se derivan de que intentamos expresar una probabilidad, que ha de estar comprendida entre 0 y 1, a través de una forma lineal, que en principio no está acotada, una solución es acotar esa forma funcional. Es decir, en lugar de proponer una relación $y_i = x_i'\beta + u_i$, formulamos la siguiente relación

$$y_i = F(x_i'\beta) + u_i \tag{2}$$

donde $F(.)$ es una función de R en R, β un vector de parámetros desconocidos $k-$dimensional y u_i un término de error con $E(u_i) = 0$. Esta formulación implica que

$$\begin{aligned} E(y_i) &= F(x_i'\beta) && (3) \\ \Longrightarrow\ & \Pr(y_i = 1) = F(x_i'\beta) && (4) \end{aligned}$$

Las igualdades (2) y (3) son totalmente equivalentes por ser y_i una variable binaria. En la práctica, es más habitual formular los modelos para variables binarias utilizando la forma de la ecuación (3). Obsérvese que esta formulación implica suponer que la variable dependiente sigue una distribución binomial tal que

$$\begin{cases} y_i = 1 & \Pr(y_i = 1) = F(x_i'\beta) \\ y_i = 0 & \Pr(y_i = 0) = 1 - F(x_i'\beta) \end{cases}$$

Estos modelos son siempre heterocedásticos, es decir, las observaciones no tienen todas la misma varianza ya que

$$Var(y_i) = (1 - F(x_i'\beta))F(x_i'\beta)$$

Para que un modelo de elección binaria esté totalmente especificado, debe decirse qué función $F(.)$ se utiliza. Las especificaciones habituales son:

a) Función identidad $\quad F(x_i'\beta) = x_i'\beta$

(Modelo de probabilidad lineal)

b) Distribución normal estándar $\quad F(x_i'\beta) = \Phi(x_i'\beta) = \int_{-\infty}^{X_i'\beta} \frac{1}{2\pi} \exp(-t^2/2)dt$

(Modelo probit)

c) Distribución logística $\quad F(x_i'\beta) = \Lambda(x_i'\beta) = \frac{\exp(x_i'\beta)}{1+\exp(x_i'\beta)} = \frac{1}{1+\exp(-x_i'\beta)}$

(Modelo logit)

Aunque en la literatura se han sugerido otras distribuciones diferentes a la normal y la logística, en las aplicaciones econométricas éstas han sido prácticamente las únicas que se han utilizado.

Como puede observarse, bajo las especificaciones Logit o Probit la función de regresión es no lineal en los parámetros β, por tanto el modelo no puede estimarse por *mínimos cuadrados ordinarios*, sino que debemos recurrir a la estimación por *máxima verosimilitud.* Esto se analiza en más detalle en la Sección 2.4.

3 LOS MODELOS DE ELECCIÓN BINARIA COMO MODELOS CON VARIABLE LATENTE

Hasta aquí, los modelos de elección binaria se han introducido como generalización de los modelos lineales, introduciendo una función $F(.)$ que tome valores en $[0,1]$ para asegurar que las predicciones obtenidas con esta modelización se puedan interpretar como probabilidades. Pero existe una interpretación de estos modelos mucho más atractiva para los economistas.

En muchos casos, la variable dependiente binaria que queremos analizar es el resultado de una decisión: el decisor (persona, empresa, país) tiene que optar entre realizar una determinada acción o no; para ello compara la utilidad (beneficios, bienestar, etc.) que, dadas sus características, le reporta esa acción. Si esa utilidad (beneficios, bienestar, etc.) es positiva, realiza la acción y en caso contrario no. Para ilustrar esta idea, utilizaremos el siguiente ejemplo:

> **EJEMPLO** : Supongamos que y_i representa la decisión de participación en el mercado laboral de una mujer ($y_i = 1$ si la mujer trabaja, 0 en otro caso) y consideremos que estos dos posibles resultados, trabajar o no trabajar, van asociados a dos niveles de utilidad diferentes
>
> $$\begin{aligned} U^*_{y=1} &= x_i'\beta_1 + u_{i1} \\ U^*_{y=0} &= x_i'\beta_0 + u_{i0} \end{aligned}$$
>
> donde x_i representa el vector de variables de control que determinan las preferencias por participar en el mercado laboral (edad, estado civil, hijos, renta familiar, tasa de desempleo local, etc.), β_1 y β_0 son vectores de parámetros desconocidos y u_{i1} y u_{i0} representan una serie de componentes de inobservables de las preferencias (gustos) específicos de cada una de las opciones. Bajo esta caracterización, una mujer decidirá participar en el mercado laboral si la utilidad de trabajar es mayor que la de no hacerlo. Es decir, la elección de participar se realiza cuando $U^*_{y=1} > U^*_{y=0} \Longrightarrow U^*_{y=1} - U^*_{y=0} > 0$. Denotemos por y_i^* a la

diferencia $U^*_{y=1} - U^*_{y=0}$. La variable y_i^* no es observable, por eso recibe el nombre de "variable latente". Ahora podemos describir la decisión

$$\begin{aligned} y_i &= 1(y_i^* > 0) \\ &= 1(x_i'\beta_1 + u_{i1} - x_i'\beta_0 + u_{i0} > 0) \\ &= 1\left[x_i'(\beta_1 - \beta_0) + u_{i1} - u_{i0} > 0\right] \end{aligned}$$

donde 1(.) es la función indicador que toma valor 1 si el evento entre paréntesis se cumple y 0 en otro caso. Aquí no podemos identificar ambos conjutos de parámetros β_1 y β_0, pero sí su diferencia $(\beta_1 - \beta_0)$. De modo que podemos reparametrizar la variable latente de la siguiente forma:

$$y_i^* = x_i'(\beta_1 - \beta_0) + u_{i1} - u_{i0} = x_i'\beta + u_i^*$$

siendo u_i^* una variable aleatoria con media 0. Distinguiremos varios casos según qué hipótesis hagamos sobre este término de error.

Caso 1: Suponemos que el término de error tiene distribución normal estándar, es decir $u_i^* \sim N(0,1)$. Entonces, la variable observada y_i satisface que

$$\begin{aligned} \Pr(y_i &= 1) = \Pr(y_i^* > 0) = \Pr(x_i'\beta + u_i^* > 0) \\ &= \Pr(u_i^* > -x_i'\beta) = 1 - \Phi(-x_i'\beta) = \Phi(x_i'\beta) \end{aligned}$$

Obsérvese que, de nuevo, hemos llegado a un *modelo Probit.*

Caso 2: Suponemos que el término de error tiene distribución normal $u_i^* \sim N(0, \sigma^2)$, con σ^2 desconocida. Los parámetros del modelo son ahora β y σ^2. Razonando como antes, obtenemos:

$$\begin{aligned} \Pr(y_i &= 1) = \Pr(y_i^* > 0) = \Pr(x_i'\beta + u_i^* > 0) \\ &= \Pr(u_i^* > -x_i'\beta) = \Pr\left(\frac{u_i^*}{\sigma} > \frac{-x_i'\beta}{\sigma}\right) = \\ &= 1 - \Phi\left[-x_i'\left(\frac{\beta}{\sigma}\right)\right] = \Phi\left[x_i'\left(\frac{\beta}{\sigma}\right)\right] \end{aligned}$$

El problema que tenemos en este caso es que no podemos identificar los parámetros β y σ de forma separada. ¿Qué significa esto? Significa que existen infinitas combinaciones de valores de β y σ que dan lugar a un mismo valor del ratio β/σ. Por ello, generalmente suele suponerse que $\sigma = 1$ (con lo cual estaríamos en el Caso 1). Si no quiere hacerse explícita esta hipótesis, podemos denotar por $\gamma = \beta/\sigma$ y suponer que la variable observada sigue un modelo Probit con coeficiente γ. En la práctica, este supuesto no plantea ningún inconveniente pues, por ejemplo, la significatividad de la j-ésima variable explicativa se estudia contrastando $\beta_j = 0$, lo que equivale a contrastar $\gamma_j = 0$.

Caso 3: Suponemos que el término de error tiene distribución logística. Siguiendo el mismo razonamiento que en los casos anteriores

$$\begin{aligned}\Pr(y_i &= 1) = \Pr(y_i^* > 0) = \Pr(x_i'\beta + u_i^* > 0) \\ &= \Pr(u_i^* > -x_i'\beta) = 1 - \Lambda\left(-x_i'\beta\right) = \Lambda\left(x_i'\beta\right)\end{aligned}$$

En este caso, tenemos un modelo Logit.

4. ESTIMACIÓN POR MÁXIMO VEROSIMILITUD DE LOS MODELOS DE ELECCIÓN BINARIA

Los economistas aplicados están interesados en la utilización de los modelos de elección binaria para explicar comportamientos observados. Para ello, necesitan estimar los parámetros de estos modelos utilizando una muestra de datos. Sin embargo, como hemos visto anteriormente, excepto el modelo de probabilidad lineal, los modelos de elección binaria son fundamentalmente no-lineales en parámetros, lo cual inmediatamente excluye el método de mínimos cuadrados ordinarios como estrategia de estimación. Este problema se resuelve utilizando máxima verosimilitud como método de estimación.

La técnica de máxima verosimilitud se centra en la probabilidad de observar ciertas realizaciones de las decisiones objeto de estudio y las características de los individuos que las toman. Consideremos una muestra de n observaciones $\{y_i, x_i\}$ extraída de la población, donde y_i es una variable binaria. Suponiendo

$$y_i = 1(y_i^* > 0) = 1(x_i'\beta + u_i^* > 0)$$

y suponiendo que las u_i son idependientes e idénticamente distribuidas, el procedimiento de máxima verosimilitud permite encontrar el valor de los parámetros β que con mayor probabilidad han generado los datos $\{y_i, x_i\}$.

Para cualquier vector β, la probabilidad de observar los valores de y_i, dados los valores de x_i (y suponiendo que las observaciones son independientes) puede escribirse como

$$L(\beta \mid x_i) = \prod_{i=1}^{n} \Pr(y_i \mid x_i; \beta) = \prod_{i=1}^{n} \Pr(y_i = 0 \mid x_i; \beta)^{1-y_i} . \Pr(y_i = 1 \mid x_i; \beta)^{y_i}$$

Es decir, la probabilidad total de observar los valores de y_i que tenemos en la muestra es igual al producto de probabilidades individuales asociada con cada una de las realizaciones. Tomando logaritmos tenemos la función de log-verosimilitud

$$\ln L(\beta \mid x_i) = \sum_{i=1}^{n} [(1 - y_i). \ln \Pr(y_i = 0 \mid x_i; \beta)] + \sum_{i=1}^{n} y_i. \ln \Pr(y_i = 1 \mid x_i; \beta)$$

El estimador máximo verosímil (MV) se define como el valor $\hat{\beta}_{MV}$ que maximiza esta función, suponiendo que este máximo existe y es único.

Modelo Probit:

$$\begin{aligned} \Pr(y_i &= 1 \mid x_i; \beta) = \Phi(x_i'\beta) \\ \Pr(y_i &= 0 \mid x_i; \beta) = 1 - \Phi(x_i'\beta) \end{aligned}$$

y, por tanto

$$L(\beta \mid x_i) = \sum_{i=1}^{n} [(1 - y_i). \ln(1 - \Phi(x_i'\beta))] + \sum_{i=1}^{n} y_i \ln \Phi(x_i'\beta)$$

Modelo Logit:

$$\begin{aligned} \Pr(y_i &= 1 \mid x_i; \beta) = \Lambda(x_i'\beta) = \frac{\exp(x_i'\beta)}{1 + \exp(x_i'\beta)} \\ \Pr(y_i &= 0 \mid x_i; \beta) = 1 - \Lambda(x_i'\beta) = \frac{1}{1 + \exp(x_i'\beta)} \end{aligned}$$

y, por tanto

$$\ln L(\beta \mid x_i) = \sum_{i=1}^{n}[(1-y_i).\ln(1-\Lambda(x_i'\beta))] + \sum_{i=1}^{n} y_i \ln \Lambda(x_i'\beta)$$

Las condiciones de primer orden que permiten obtener el estimador MV se obtienen derivando la función de log-verosimilitud respecto a β e igualando a cero. Tanto en el caso Probit como en el Logit, la función de log-verosimilitud es globalmente cóncava, por tanto estos modelos tienen, a lo sumo, una única solución que define explícitamente al estimador MV del modelo.

5. INTERPRETACIÓN DE LOS COEFICIENTES Y COMPARACIÓN DE LOS MODELOS

5.1 Interpretación de los parámetros

Dada una variable dependiente binaria y_i y un vector $(k \times 1)$ de variables explicativas x_i, hemos visto tres especificaciones alternativas para modelar la probabilidad condicional:

$$\begin{aligned}
\text{Modelo de probabilidad lineal} \quad &: \quad \Pr(y_i = 1 \mid x_i; \beta) = x_i'\beta \\
\text{Probit} \quad &: \quad \Pr(y_i = 1 \mid x_i; \beta) = \Phi(x_i'\beta) \\
\text{Logit} \quad &: \quad \Pr(y_i = 1 \mid x_i; \beta) = \Lambda(x_i'\beta)
\end{aligned}$$

Todas estas funciones son monótonas crecientes en sus argumentos, esto aporta una primera interpretación intuitiva de los parámetros:

> *Si el parámetro β_j asociado con la variable x_j es positivo (negativo), entonces la probabilidad condicional* $\Pr(y_i = 1 \mid x_i; \beta)$ *aumentará (disminuirá) cuando x_j crece.*

Pero, generalmente, los economistas están interesados no sólo en el signo del efecto, sino también en su magnitud. La medida más habitual que se utiliza en los modelos de elección binaria es el *efecto marginal sobre la probabilidad condicional* $\Pr(y_i = 1 \mid x_i; \beta)$ *de un* *aumento en una unidad en la variable* x_j. Cada uno de los modelos de elección binaria que hemos visto conllevan diferentes representaciones de estos efectos marginales:

$$\begin{aligned}
\text{Modelo de probabilidad lineal} &: \frac{\partial \Pr(y_i = 1 \mid x_i; \beta)}{\partial x_{ij}} = \beta_j \\
\text{Probit} &: \frac{\partial \Pr(y_i = 1 \mid x_i; \beta)}{\partial x_{ij}} = \phi(x_i'\beta)\beta_j \\
\text{Logit} &: \frac{\partial \Pr(y_i = 1 \mid x_i; \beta)}{\partial x_{ij}} = \frac{\exp(x_i'\beta)}{[1 + \exp(x_i'\beta)]^2}\beta_j
\end{aligned}$$

Lo que sugieren estos resultados es que las estimaciones puntuales de los parámetros poblacionales β no son directamente comparables. Obsérvese que, mientras el efecto marginal en el modelo de probabilidad lineal es constante, los efectos marginales en el Probit y Logit dependen del valor de las variables explicativas.

Determinar el efecto marginal de una variable explicativa tiene sentido cuando dicha variable es continua. Si el regresor es una variable discreta, lo natural es determinar la probabilidad de "éxito" que se obtiene para cada posible valor del regresor y comparar después esas probabilidades.

Vamos a ilustrar esto con un ejemplo.

EJEMPLO: Formulamos un modelo logit para analizar cómo afectan los ingresos familiares y el tipo de población (rural o urbana) donde reside la familia en la decisión de vivir en casa propia o alquilada. Las variables son:

$y_i = 1$ si la familia vive en casa propia, 0 si vive de alquiler

x_{2i} = ingresos mensuales (medios) de la familia en el último año (medidos en diez miles de euros)

x_{3i} =1 si la familia vive en una ciudad con más de 10.000 habitantes, 0 en caso contrario.

Estamos formulando el siguiente modelo:

$$\Pr(y_i = 1 \mid x_{2i}, x_{3i}) = \Lambda(\beta_1 + \beta_2 x_{2i} + \beta_3 x_{3i}) = \frac{\exp(\beta_1 + \beta_2 x_{2i} + \beta_3 x_{3i})}{1 + \exp(\beta_1 + \beta_2 x_{2i} + \beta_3 x_{3i})}$$

Supongamos que estimamos el modelo y obtenemos $\hat{\beta}_1 = -4$, $\hat{\beta}_2 = 30$ y $\hat{\beta}_3 = -0.05$. Veamos qué tipo de información podemos extraer con estos datos:

a) *Predicción de probabilidades*: Supongamos una familia A urbana, con unos ingresos de 1000 euros al mes, la probabilidad de vivir en una casa de propiedad es

$$\Lambda(-4 + 30 \times 0.1 + (-0.5) \times 1) = 0.182$$

b) *Efecto de la variación en una variable explicativa continua*: Si la familia A aumenta sus ingresos mensuales en 100 euros, la probabilidad de que sea propietaria de la casa donde reside es

$$\Lambda(-4 + 30 \times 0.11 + (-0.5) \times 1) = 0.231$$

es decir, la probabilidad aumenta 0.049. Pero veamos cómo el efecto de este cambio no es el mismo si los ingresos de partida son diferentes:

Supongamos una familia B, urbana, y con ingresos de 2500 euros al mes. La probabilidad de que resida en casa de propiedad es $\Lambda(3) = 0.952$. Si sus ingresos aumentan en 100 euros, la probabilidad pasa a ser $\Lambda(3.3) = 0.964$, es decir, aumenta 0.012.

En general, para cualquier familia urbana, el efecto marginal de la variable ingresos mensuales se puede calcular de forma más directa aplicando la expresión que se presentó más arriba, esto es

$$\frac{\partial \Pr(y_i = 1 \mid x_{2i}, x_{3i})}{\partial x_{ij}} = \frac{\exp(-4.5 + 30x_{2i})}{[1 + \exp(-4.5 + 30x_{2i})]^2} \times 30$$

c) *Efecto de la variación en una variable artificial*: Se trata de comparar la probabilidad de poseer la casa donde se reside, según la familia sea urbana o rural. En esta comparación también hay que tener en cuenta los ingresos mensuales de la familia. Para ello, o bien se fija un valor de los ingresos, o bien se ofrece un gráfico en el que, para cada uno de los posibles valores de los ingresos, se puede observar cuál es la probabilidad de ser propietario cuando se reside en un núcleo rural o urbano. En este segundo caso, se elabora un gráfico con los ingresos mensuales en el eje de abscisas y las probabilidades en el eje de ordenadas, y se representan dos curvas: "Probabilidad de que una familia urbana con ingresos mensuales x_2 posea la casa en la que reside"$=\Lambda(-4.5+30x_2-0.5)$; y "Probabilidad de que una familia rural con ingresos mensuales x_2 posea la casa en la que reside"$=\Lambda(-4.5+30x_2)$.

Si el modelo que se especifica es un Probit, el procedimiento para derivar los efectos marginales de las variables y los cambios en la probabilidad inducidos por alguna variable artificial es similar. De hecho, aunque las estimaciones de los parámetros β que se obtienen en un Probit no son directamente comparables con las correspondientes al Logit, los efectos marginales de las variables explicativas que se obtienen a partir de estos dos modelos sí son muy parecidos.

También existe una transformación bastante popular para poder comparar directamente los parámetros estimados a partir de un Logit y un Probit. Esta transformación es: $\hat{\beta}_{\text{Probit}}$ t $0.625\hat{\beta}_{\text{Logit}}$.

5.2 El "odds-ratio"

El "odds" es el cociente entre la probabilidad de elegir una opción y la probabilidad de elegir la otra, es decir

$$\text{Odds}=\frac{\Pr(y_i=1\mid x_i;\beta)}{1-\Pr(y_i=1\mid x_i;\beta)}$$

Veamos cuál es su expresión en las especificaciones Logit y Probit:

Logit:

$$\text{Odds}=\frac{\frac{1}{1+\exp(-x_i'\beta)}}{1-\frac{1}{1+\exp(-x_i'\beta)}}=\frac{1}{\exp(-x_i'\beta)}=\exp(x_i'\beta)$$

Probit:

$$\text{Odds}=\frac{\Phi(x_i'\beta)}{1-\Phi(x_i'\beta)}$$

Si el Odds>1 significa que, para el individuo i, la opción $y_i = 1$ es más probable que la opción $y_i = 0$ o, en otras palabras, que el individuo i obtiene mayor utilidad de la opción 1 que de la opción 0. Si el Odds<1 la interpretación sería la contraria. Y si Odds=1, ambas opciones son igual de probables, es decir, el individuo es indiferente ante ambas opciones.

En muchas ocasiones, puede resultar ilustrativo comparar el Odds entre dos situaciones diferentes. Por ejemplo, supongamos un individuo de referencia j en el cual se fijan todas las variables contenidas en x_i en su valor medio y se compara su Odds con el de otro individuo i que difiere de j en el valor de una o más variables explicativas. Podemos construir el Odds-ratio o Cociente entre Odds de esos individuos como

$$\text{Odds ratio o cociente entre Odds}=\frac{\frac{P_i}{1-P_i}}{\frac{P_j}{1-P_j}}$$

Si el Odds-ratio>1 significa que el individuo i tiene una preferencia mayor por la opción $y_i = 1$ frente a la $y_i = 0$ que el individuo j.

Siguiendo el mismo razonamiento, podemos deducir la interpretación de los valores del Odds-ratio menores o iguales que 1.

6. INFERENCIA ESTADÍSTICA EN MODELOS DE ELECCIÓN BINARIA

Para realizar inferencias sobre los parámetros de los modelos de elección binaria, necesitamos disponer de estimaciones de las varianzas de dichos parámetros en cada modelo. En el modelo de probabilidad lineal son fáciles de obtener puesto que la estimación se puede llevar a cabo por MCO. En el caso de los modelos Probit y Logit, la distribución de los estimadores MV para una muestra de tamaño n es

$$\sqrt{n}(\hat{\beta} - \beta) \overset{a}{\vee} N(0, I(\hat{\beta})^{-1})$$

donde $I(\hat{\beta})^{-1}$ está denotando la inversa de la matriz de información. Existen varias formas de estimar esta matriz. Afortunadamente, los programas econométricos calculan directamente la matriz de varianzas y covarianzas de $\hat{\beta}$, así como los errores estándar. La inferencia estadística y los contrastes de hipótesis se pueden llevar a cabo utilizando las técnicas habituales.

6.1 Medidas de bondad de ajuste.—

Para juzgar cuándo un modelo de elección binaria ajusta bien los datos observados existen varias medidas que se basan en los principios del R^2 utilizado en los modelos lineales estimados por MCO (esta medida, obviamente, no puede ser utilizada en los modelos de elección binaria). Existen dos medidas bastante populares que se construyen a partir de la verosimilitud del modelo completo L_{SR} (evaluada en los β estimados) y la verosimilitud de un modelo en el que sólo incluimos una constante L_R (evaluada en la estimación de dicha constante). Estas medidas fueron desarrolladas por Cragg y Uhler (1970) y McFadden (1974). La formulación de estas medidas es la siguiente:

$$\begin{aligned} \text{Cragg y Uhler} \quad &: \quad \textit{pseudo-}R^2 = (L_{SR}^{2/n} - L_R^{2/n})/(1 - L_R^{2/n}) \\ \text{McFadden} \quad &: \quad \textit{pseudo-}R^2 = 1 - \ln L_{SR} / \ln L_R \end{aligned}$$

Una forma alternativa de evaluar la bondad de ajuste de un modelo de elección binaria es la que se basa en la comparación de las predicciones del modelo con los valores observados. Para hacer esto, necesitamos establecer previamente qué consideramos una predicción correcta. La idea es la siguiente: denotemos $\hat{P}_i = \Pr(y_i = 1 \mid x_i; \hat{\beta})$, es decir, la predicción de la probabilidad de que y_i tome valor 1 obtenida a partir del modelo de elección binaria estimado (Probit o Logit). Pero esta probabilidad está entre 0 y 1, ¿a partir de qué valor consideramos que el modelo predice una realización de $y_i = 1$ para el individuo? Una posibilidad es suponer

$$\hat{y}_i = 1(\hat{P}_i > 0.5).$$

La proporción de predicciones correctas se define entonces como

$$\mathbf{P} = \frac{1}{n} \sum_{i=1}^{n} 1(y_i = \hat{y}_i)$$

y constituye una medida de bondad de ajuste del modelo, aunque no siempre es fiable.

6.2 Contraste de significación de la regresión.—

Para contrastar la significatividad conjunta de todas las variables explicativas en un modelo estimado por MV, podemos utilizar el contraste de ratio de verosimilitudes. Es decir, estamos contrastando

$$
\begin{aligned}
H_0 &: \quad \beta_2 = \beta_3 = ... = \beta_k \\
H_1 &: \quad \text{al menos un } \beta_j = 0, \quad j = 2, ..., k
\end{aligned}
$$

El modelo restringido sería un modelo en el que sólo hubiera término constante, y el modelo sin restringir, el modelo con todas las variables explicativas. Para implementar este contraste, necesitamos disponer de las funciones de verosimilitud evaluadas en su máximo, para estos dos modelos. El estadístico de contraste es el siguiente:

$$-2\ln(L_R/L_{SR}) = 2(\ln L_{SR} - \ln L_R) \sim \chi^2_r$$

donde r es igual al número de restricciones (en este caso, número de variables explicativas del modelo). Como siempre, la regla de decisión del contraste viene determinada por la comparación del valor del estadístico de contraste con el valor crítico de una χ^2_r para un nivel de significación α determinado (0.05, 0.01 habitualmente).

Este contraste puede ser utilizado también para contrastar la significatividad de un subconjunto de variables explicativas.

> Nótese la similitud de este contraste con el contraste F de restricciones lineales en un modelo estimado por MCO. Recordemos: el estadístico F mide el aumento de la suma de cuadrados de los residuos cuando se eliminan las variables del modelo. El ratio de verosimilitudes se basa en la diferencia de las funciones de log-verosimilitud para los modelos restringido y sin restringir. La idea es que, como el estimador MV maximiza la función de log-verosimilitud, cuando se eliminan variables obtenenmos log-verosimilitudes más pequeñas -o al menos no más grandes-. El estadístico, entonces, trata de revelar si la disminución en la log-verosimilitud es lo bastante grande como para llegar a la conclusión de que las variables eliminadas son importantes.

6.3 Contrastes de especificación en modelos de elección binaria.—

Consideremos dos importantes problemas de especificación con los que podemos encontrarnos: omisión de variables relevantes y heterocedasticidad. En el contexto de los modelos de elección binaria, las consecuencias de estos problemas son todavía más graves:

1. Si omitimos una variable relevante del modelo, incluso en el caso en que esta variable esté incorrelacionada con las otras variables explicativas incluidas en el modelo, los parámetros estimados serán inconsistentes.

2. Si las perturbaciones del modelo de regresión latente son heterocedásticas, los estimadores MV son inconsistentes y la matriz de varianzas y covarianzas incorrecta. Este resultado es especialmente preocupante porque, en la mayoría de aplicaciones de modelos de elección binaria, se trabaja con datos microecómicos que suelen tener naturaleza heterocedástica.