

SIGNIFICADO DE LOS ESTIMADORES EN UN MODELO LOGIT DE VARIABLES DEPENDIENTES CUALITATIVAS MULTIPLE

Ursicino Carrascal

RESUMEN.—El significado de los estimadores en modelos LOGIT no es tan inmediato como el que se desprende de la utilización de un modelo lineal. Este artículo pretende aclarar y delimitar tal significado, así como su ilustración práctica.

1. PLANTEAMIENTO

En este artículo vamos a analizar el significado de los estimadores cuando tenemos una variable ficticia como variable a explicar por el modelo. Si planteamos la estimación de un modelo lineal simple, surgirán dos problemas:

— por un lado, la variable endógena toma un número limitado de valores, en el caso más habitual tan sólo dos, 0 ó 1; y admitiendo que los regresores son fijos, la perturbación del modelo tomaría esa misma cantidad de valores, incumpléndose además las hipótesis clásicas¹.

— por otra parte, nada nos asegura que las predicciones de este modelo tomen los contados valores que toma la variable original, por lo que no podrán ser interpretadas de manera coherente.

Ante estas dificultades, la opción es estimar la probabilidad de que cada individuo presente determinada cualidad representada numéricamente por la variable endógena ficticia. Al tratarse de una probabilidad, los valores estimados deberán pertenecer al intervalo $[0,1]$, y no sólo tomar determinados valores, por lo que las predicciones serán fácilmente interpretables.

1 Véase al respecto, Novales (1993, p. 531).

Así pues, en lugar de suponer que nuestra variable endógena depende de los regresores:

$$y_i = \beta_0 + \beta_I x_{Ii} + \dots + \beta_K x_{Ki} + u_i \quad (1)$$

o que la probabilidad p_i de que la variable aleatoria tome un valor dependa de los mismos regresores:

$$p_i = \beta_0 + \beta_I x_{Ii} + \dots + \beta_K x_{Ki} + u_i \quad (2)$$

vamos a suponer que tal probabilidad es el valor de una función de distribución, con lo que aseguramos que la predicción esté en el intervalo $[0,1]$

$$p_i = F(\beta_0 + \beta_I x_{Ii} + \dots + \beta_K x_{Ki}) \quad (3)$$

y estimaremos los parámetros del modelo suponiendo que F es una determinada función de distribución.

2. LOGIT MULTIRRESPUESTA

En un modelo de dos variables la influencia del estimador sobre las probabilidades es evidente: ante un aumento en una unidad en uno de los regresores, si el estimador es positivo aumenta el valor del punto en que se considera la función de distribución y por tanto aumenta la probabilidad de que la variable tome el valor 0, disminuyendo la de 1; si el valor del estimador es negativo, un aumento en el valor del regresor favorece la probabilidad de que la endógena sea 1, en detrimento del otro caso. La variación concreta de la probabilidad depende no sólo del valor de estimador, sino también del punto en que estemos considerando la función de distribución. Así, como muestra Gracia (1988, pp. 31-2) por ejemplo en el caso de un modelo LOGIT:

$$\frac{\partial p_i}{\partial x_{ik}} = \frac{e^{x_i \beta}}{(1 + e^{x_i \beta})^2} \quad (4)$$

Al igual que hemos planteado el modelo anterior para dos valores de la variable endógena, podemos tratar el caso en que la endógena presente más de dos posibilidades. Si los valores que presenta la variable endógena están ordenados (por ejemplo: familias con 1, 2, 3 o más miembros), podemos plantear estimar probabilidades de la siguiente forma:

$$p(Y \leq i / X) = F(\alpha_i + X\beta) \quad (5)$$

siendo i cada una de las clases ordenadas definidas, de modo que estamos estimando modelos que sólo distinguen cada clase por el término indepen-

diente. La probabilidad de pertenecer a una determinada clase i se obtiene como diferencia entre la función de distribución en el punto $\alpha_i + X_i'\beta$, y el valor de la función de distribución en el punto anterior $\alpha_{i-1} + X_i'\beta$.

Para este caso nos planteamos la influencia del valor de los estimadores en las probabilidades estimadas para cada clase en un modelo LOGIT con un solo regresor por simplificar la notación.

Este efecto es fácil de valorar en la primera clase:

$$\text{como } p(Y = 1 / X) = F(\alpha_1 + \beta X) = \frac{e^{\alpha_1 + \beta X}}{1 + e^{\alpha_1 + \beta X}} \quad (6)$$

$$\text{entonces } \frac{\partial p(Y = 1 / X)}{\partial X} = F(\alpha_1 + \beta X) \frac{\beta}{1 + e^{\alpha_1 + \beta X}} \quad (7)$$

que es lo mismo que aparece en la ecuación 4, de modo que si el signo del estimador es positivo aumenta la probabilidad de pertenecer a esta primera clase ante aumentos en el regresor, y lo opuesto en caso de ser el estimador negativo.

Igualmente es fácil de valorar este efecto en la última clase, n :

$$\text{en este caso } p(Y = n / X) = 1 - F(\alpha_{n-1} + \beta X) = 1 - \frac{e^{\alpha_{n-1} + \beta X}}{1 + e^{\alpha_{n-1} + \beta X}} \quad (8)$$

$$\text{y por lo tanto } \frac{\partial p(Y = n / X)}{\partial X} = -F(\alpha_{n-1} + \beta X) \frac{\beta}{1 + e^{\alpha_{n-1} + \beta X}} \quad (9)$$

de manera que si el signo del estimador es negativo aumenta la probabilidad de pertenecer a esta última clase ante aumentos en el regresor, y lo opuesto en caso de ser el estimador positivo.

Sin embargo esta valoración puede resultar más complicada en el caso del resto de clases. La probabilidad de la clase i será

$$p(Y = i / X) = F(\alpha_i + \beta X) - F(\alpha_{i-1} + \beta X) = \frac{e^{\alpha_i + \beta X}}{1 + e^{\alpha_i + \beta X}} - \frac{e^{\alpha_{i-1} + \beta X}}{1 + e^{\alpha_{i-1} + \beta X}} \quad (10)$$

$$\begin{aligned} \text{y por lo tanto } \frac{\partial p(Y = i / X)}{\partial X} &= F(\alpha_i + \beta X) \frac{\beta}{1 + e^{\alpha_i + \beta X}} - F(\alpha_{i-1} + \beta X) \frac{\beta}{1 + e^{\alpha_{i-1} + \beta X}} = \\ &= \beta e^{\alpha_{i-1} + \beta X} \frac{(e^{\alpha_i - \alpha_{i-1}} - 1)}{(1 + e^{\alpha_i + \beta X})^2} \frac{(1 - e^{\alpha_i + \beta X} e^{\alpha_{i-1} + \beta X})}{(1 + e^{\alpha_{i-1} + \beta X})^2} \end{aligned} \quad (11)$$

Para valorar esta derivada tendremos que tener en cuenta el signo del estimador y el del factor $(1 - e^{\alpha_i + \beta X} e^{\alpha_{i-1} + \beta X})$. El resto de factores son posi-

tivos, incluido $(e^{\alpha_i - \alpha_{i-1}} - 1)$, ya que $\alpha_i > \alpha_{i-1}$, y por lo tanto siempre se cumple que $e^{\alpha_i - \alpha_{i-1}} > 1$, y este factor será positivo.

El signo positivo o negativo del factor $(1 - e^{\alpha_i + \beta X} e^{\alpha_{i-1} + \beta X})$ depende de que $\alpha_i + \beta X + \alpha_{i-1} + \beta X$ sea menor o mayor que cero (respectivamente) o, despejando, que el estimador sea menor o mayor que $c = -(\alpha_i + \alpha_{i-1})/(2X)$, admitiendo que el regresor (X) tome siempre valores positivos.

Así, si $c < 0$, dependiendo de los valores que tome β nos encontraremos en los siguientes casos:

- si $\beta < c$, β es negativo pero $(1 - e^{\alpha_i + \beta X} e^{\alpha_{i-1} + \beta X})$ es positivo, con lo que la derivada es negativa.
- si $c < \beta < 0$, tanto β como $(1 - e^{\alpha_i + \beta X} e^{\alpha_{i-1} + \beta X})$ serían negativos, de modo que la derivada es positiva. Nótese además que según aumenta X , disminuye c en valor absoluto, aproximándose más a cero por el lado de los negativos, estrechándose este intervalo.
- si $\beta > 0$, β es positivo pero $(1 - e^{\alpha_i + \beta X} e^{\alpha_{i-1} + \beta X})$ es negativo, resultando que la derivada es negativa.

Por otro lado, si $c > 0$, según los valores que tome β nos encontraremos en los siguientes casos:

- si $\beta < 0$, β es negativo pero $(1 - e^{\alpha_i + \beta X} e^{\alpha_{i-1} + \beta X})$ es positivo, de modo que la derivada es negativa.
- si $0 < \beta < c$, β y $(1 - e^{\alpha_i + \beta X} e^{\alpha_{i-1} + \beta X})$ son positivos, con lo que la derivada es positiva. Nótese de nuevo que según aumenta X , disminuye c , aproximándose más a cero y estrechándose este intervalo.
- si $\beta > c$, β es positivo pero $(1 - e^{\alpha_i + \beta X} e^{\alpha_{i-1} + \beta X})$ es negativo, lo que supone que la derivada es negativa.

En el siguiente ejemplo² vamos a ilustrar gráficamente lo que estamos comentando. Hemos estimado la probabilidad de que una familia tenga 1, 2, 3 ó 4 o más miembros en función de su renta familiar (en millones de ptas.), con datos de la Encuesta de Presupuestos Familiares 1990-91 española. En esta estimación consideramos a las familias ordenadas según su tamaño, de modo que estimamos la probabilidad de tener hasta cierto número de miembros.

En la tabla 1 aparecen los estimadores del término constante, de la RENTA, y los correspondientes a familias de dos y de tres miembros, de modo que para un nivel de renta dado, las probabilidades de que la familia tenga los distintos tamaños son:

$$p(\text{TAM} = 1) = F(-0.7788761 - 0.65949 * \text{renta}) = 1 / (1 + e^{-(-0.7788761 - 0.65949 * \text{renta})})$$

$$p(\text{TAM} = 2) = F(1.72460196 - 0.65949 * \text{renta}) - F(-0.7788761 - 0.65949 * \text{renta})$$

$$p(\text{TAM} = 3) = F(2.99150065 - 0.65949 * \text{renta}) - F(1.72460196 - 0.65949 * \text{renta})$$

$$p(\text{TAM} = 4) = 1 - F(2.99150065 - 0.65949 * \text{renta})$$

TABLA 1

SAS		11:19 Saturday, August 3, 1996 1	
Probit Procedure			
Class Level Information			
Class	Levels	Values	
TAM	4	1 2 3 4	
Number of observations used = 16386			

SAS			11:19 Saturday, August 3, 1996 2		
Probit Procedure					
Data Set	= WORK.A				
Dependent Variable	= TAM				

Weighted Frequency Counts for the Ordered Response Categories

Level	Count
1	2174
2	4735
3	4425
4	5052

Log Likelihood for LOGISTIC -20742.63509

Probit Procedure

Variable DF Estimate Std Err ChiSquare Pr>Chi Label/Value

INTERCPT	1	-0.778876	1 0.032496	574.4824	0.0001	Intercept
RENTA	1	-0.65949	0.01469	2015.749	0.0	
INTER.2	1	1.72460196	0.022953			2
INTER.3	1	2.99150065	0.027499			3

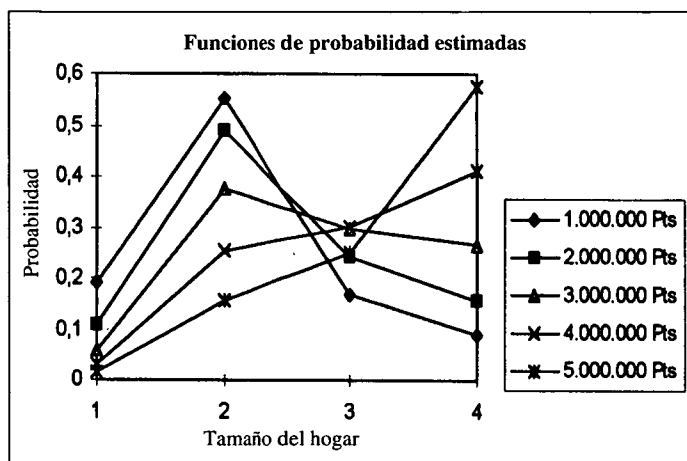
Estimated Covariance Matrix

	INTERCPT	RENTA	INTER.2	INTER.3
INTERCPT	0.001056	-0.00330978	-0.000295	-0.000279
RENTA	-0.00330978	0.02157657	-0.000631569	-0.00118315
INTER.2	-0.000295	-0.000631569	0.000527	0.000492
INTER.3	-0.000279	-0.00118315	0.000492	0.000756

En la tabla 2 aparecen las probabilidades de que la familia tenga 1, 2, 3 o 4 miembros para distintos niveles de renta:

TABLA 2

	1.000.000 Pts.	2.000.000 Pts.	3.000.000 Pts.	4.000.000 Pts.	5.000.000 Pts.
1 persona	0,191798495	0,109305372	0,059673108	0,031773418	0,016686531
2 personas	0,55186774	0,490732272	0,377198564	0,254539752	0,155124229
3 personas	0,167827441	0,241873956	0,296741161	0,301162743	0,252293003
4 personas	0,088506324	0,1580884	0,266387167	0,412524087	0,575896237



Atendiendo a los criterios antes mencionados, como β es negativo, la probabilidad de que la familia tenga un solo miembro disminuye al aumentar su renta, y aumenta la probabilidad de que tenga 4 miembros o más, como se puede apreciar en la tabla 2 y en el gráfico.

TABLA 3

	1.000.000 Pts.	2.000.000 Pts.	3.000.000 Pts.	4.000.000 Pts.
Factor c				
2 persona	-0,47286293	-0,236431465	-0,157620977	-0,118215733
3 personas	-2,358051305	-1,179025653	-0,786017102	-0,589512826
$\beta-c$				
2 personas	-0,18662707	-0,423058535	-0,501869023	-0,541274268
3 personas	1,698561305	0,519535653	0,126527102	-0,069977174

En la tabla 3 aparecen los valores c calculados para el resto de casos posibles, así como la diferencia entre β y c . En los casos en que esta diferencia sea negativa, como el estimador también lo es, el efecto sobre la

derivada es negativo, luego la probabilidad de pertenecer a ese grupo disminuye cuando aumenta la renta, como se aprecia para el caso de familias de dos miembros. Si esta diferencia es positiva, y como el estimador es negativo, el efecto sobre la derivada es positivo, esto es, aumenta la probabilidad de pertenecer a este tipo de familia al aumentar la renta, lo que ocurre en todos los casos excepto cuando la renta es de cuatro millones para familias con tres miembros, de modo que la probabilidad de pertenecer a este tipo de familia es mayor si tiene una renta de cuatro millones que de cinco, como se observa en la tabla y en el gráfico donde la línea de cinco millones sobrepasa a la de cuatro sólo para el último tipo de familia, mientras que este cambio ya se aprecia para el resto de rentas con hogares de tres miembros.

CONCLUSIONES

En este trabajo se ha pretendido aclarar el significado de los estimadores en los modelos LOGIT multirrespuesta, obteniéndose un criterio de delimitación claro para el análisis de los estimadores, combinando el signo del estimador y el de su diferencia con lo que hemos denominado factor c.

REFERENCIAS

- Amemiya, T. (1988): «Modelos de respuesta cualitativa: un examen». Traducido por Daniel Manzano Romero. *Cuadernos Económicos del I.C.E.*, nº 39, pp. 173-243.
- Carrascal, U. (1997): *Aplicaciones estadísticas y econométricas con SAS*. Editorial RAMA. Madrid.
- Gracia Díez, M. (1988): «Modelos con variable dependiente cualitativa y de variación limitada». *Cuadernos Económicos del I.C.E.*, nº 39, pp. 7-49.
- Johnston, J. (1987): *Métodos de Econometría*. Traducido por Sánchez Fernández, J. y García Barbancho, A. Editorial Vicens-Vives. Barcelona.
- Judge, G. G., Hill, R. C., Griffiths, W. E., Lütkepohl, H. y Lee, T. C. (1988): *Introduction to the theory and practice of Econometrics*. 2ª Edición. John Wiley & Sons. Nueva York.
- Maddala, G. S. (1985): *Limited-dependent and qualitative variables in Econometrics*. Cambridge University Press. Nueva York.
- Maddala, G. S. (1996): *Introducción a la Econometría*. Traducido por Jolly Vallejo, J. C. Editorial Prentice-Hall Hispano Americana S. A. Méjico.
- McKelvey, R. D. y Zavoina, W. (1975): «A Statistical model for the analysis of ordinal level dependent variables» *Journal of Mathematical Sociology*, Vol. 4, pp. 103-20.
- Novalés, A. (1993): *Econometría 2ª Edición*. Editorial McGraw-Hill. Madrid.