ANÁLISIS DE MODELOS MATEMÁTICOS PARA LA PROYECCIÓN Y CORRELACIÓN DE SERIES CRONOLÓGICAS DE ESTUDIOS, PROYECTOS Y TRABAJOS DE INVESTIGACIÓN

M.P. Ing. Luis Milla Lostaunau

RESUMEN

Los estudios, proyectos y trabajos de investigación requieren de modelos matemáticos que optimicen la trayectoria de la curva que se obtenga de una muestra histórica o de un conjunto de datos obtenidos en el terreno o en el laboratorio a fin de que se adecúen sus variables a las mismas.

ABSTRACT

The studies, projects and research works require of mathematical models that optimize the trajectory of the curve that is obtained of a historical sample or of a group of data obtained in the land or in laboratory.

INTRODUCCIÓN

Cuando se quiere realizar estudios de demanda, generalmente en la práctica se encuentra que existen relaciones entre dos o más variables. Se desea frecuentemente expresar esta relación mediante una ecuación matemática que ligue las variables; para ello nos sirve de ayuda la colección de datos que muestran los correspondientes valores de las variables consideradas.

Por ejemplo, supóngase X e Y denotar altura y peso, respectivamente de hombres jóvenes. Entonces una muestra de N individuos daría las alturas X_1 , X_2 , ... X_n y los pesos correspondientes Y_1, Y_2, ... Y_n.

El proceso siguiente es representar los puntos (X_1,Y_1), (X_2, Y_2)....., (X_n,Y_n) en un sistema de coordenadas rectangulares. El conjunto de puntos resultantes se llama diagrama de dispersión. Con el mencionado diagrama es posible representar una curva que se aproxime a los datos y se denomina curva de aproximación.

En la figura (a) se puede ver que los datos se aproximan a una línea recta, por lo tanto se dice que entre las variables existe una relación lineal. En la figura (b) se nota que existe cierta relación entre las variables, pero ésta no es lineal.

"Cuando se quiere realizar estudios, se encuentra relaciones entre dos o más variables, se desea frecuentemente expresar esta relación mediante una ecuación matemática"

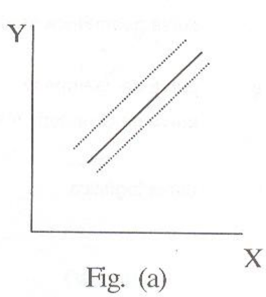

Fig. (a)

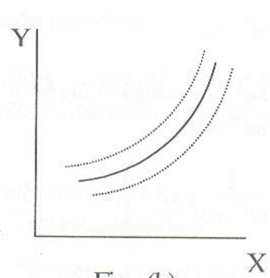

Fig. (b)

ALTERNATIVAS DE MODELOS MATEMÁTICOS DE APROXIMACIÓN TIPO CURVA QUE MÁS SE EMPLEAN PARA DETERMINAR TENDENCIAS HISTÓRICAS

(1)	$y = a + b\,x$	(función lineal)
(2)	$y = a + bx + cx^2$	(función parabólica)
(3)	$y = a(1+i)^x = ab^x$	(función exponencial)
(4)	$y = ax^n$	(función potencial)
(5)	$y = \dfrac{a}{a + bc^{-cx}}$	(función logística)

1. Modelo de ecuaciones de curvas de aproximación

A continuación se anotan varios tipos comunes de curvas de aproximación y sus ecuaciones. Las letras distintas a X e Y representan las constantes. Las variables X e Y se conocen a menudo como las variables independientes y dependientes, respectivamente, aunque pueden intercambiarse.

Estas ecuaciones se llaman polinomiales de primero, segundo, tercero y cuarto y n grados respectivamente, las cuales se llaman funciones lineales, cuadrática, cúbica y cuártica.

(1)	$y = a_0 + a_1\,x$	Línea recta
(2)	$y = a_0 + a_1x + a_2x^2$	Parábola o curva cuadrática
(3)	$y = a_0 + a_1x + a_2x^2 + a_3x^3$	Curva cúbica
(4)	$y = a_0 + a_1x + a_2x^2 + a_3x^3 + a_4x^4$	Curva cuártica
(5)	$y = a_0 + a_1x + a_2x^2 + \dots\dots\dots a_nx^n$	Curva de grado n

2. Otros tipos de modelos prácticos

Para decidir qué curva deberá usarse, es de gran utilidad obtener los diagramas de dispersión de las variables transformadas, para notar cual es la forma de la curva.

(1)	$Y = \dfrac{1}{a_0 + a_1}$ $\quad o \quad$ $\dfrac{1}{Y} = a_o + a_1 x$	Hipérbola
(2)	$Y = ab^x$ $\quad o \quad$ $\log Y = \log a + (\log b)x = a_0 + a_1 x$	(curva exponencial)
(3)	$Y = ax^b$ $\quad o \quad$ $\log Y = \log a + b \log x$	curva geométrica
(4)	$Y = abx + g$	curva exponencial modificada
(5)	$Y = ax^b + g$	curva geométrica modificada
(6)	$Y = pq^{b^x}$ $\quad o \quad$ $\log Y = \log b + b^x \log q = ab^x + g$	curva de Gompertz
(7)	$Y = pq^{bx} + n$	curva de Gompertz modificada
(8)	$Y = \dfrac{1}{ab^x + g}$ $\quad o \quad$ $\dfrac{1}{Y} = ab^x + g$	curva logística
(9)	$Y = a_0 + a_1(\log x) + a_2(\log x)^2$	

ANÁLISIS DE LAS ECUACIONES

1. Línea Recta.

Es la más sencilla de las curvas de aproximación, la ecuación puede escribirse:

$$Y = a_0 + a_1 x$$

Dados dos puntos cualquiera (X_1,Y_1) y (X_2,Y_2) de la línea, las constantes a_0 y a_1 pueden ser determinadas. La ecuación de la línea resultante puede escribirse:

$$Y - Y_1 = \left(\frac{Y_2 - Y_1}{X_2 - X_1}\right)(X - X_1) \quad o \quad Y - Y_1 = m(X - X_1)$$

Donde:

$m = \dfrac{Y_2 - Y_1}{X_2 - X_1}$ pendiente de la línea y representa el cambio de Y dividido por el correspondiente cambio de X.

2. Método de mínimos cuadrados

Para no emitir una afirmación individual en la construcción de rectas, parábolas u otras curvas de aproximación, en el ajuste a colecciones de datos es conveniente obtener una definición de la "mejor recta de ajuste, "ó" la mejor parábola de ajuste", etc.

Por ejemplo para llegar a una definición consideramos en la figura (c), los puntos representativos de los datos dados (X_1,Y_1), (X_2,Y_2),...(X_n,Y_n). Para un valor dado de X, por ejemplo X_1 habrá una diferencia entre el valor Y_1 y el valor de la curva C, como se ve en la figura (c), se denota esta diferencia por D, que se denomina, desviación y puede ser positivo negativo ó cero. Igual para los valores X_2.....X_n se obtienen desviaciones:

$$D_1^2 + D_2^2 + ... + D_n^2$$

Si esto es pequeño el ajuste es bueno, si es grande el ajuste es malo.

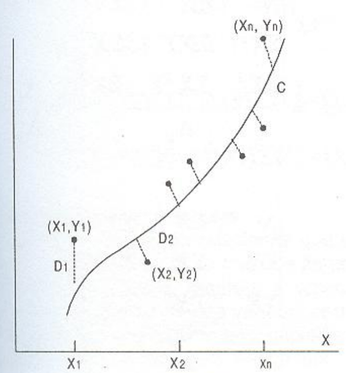

Figura (c)

Se concluye, de todas las curvas de aproximación a una serie de datos puntuales, la curva que tiene la propiedad de que:

$$D_1^2 + D_2^2 + ... + D_n^2$$

es mínimo, se conoce como la mejor curva de ajuste.

3. Recta por mínimos cuadrados

La recta de aproximación por mínimos cuadrados del conjunto de puntos (X_1,Y_1), (X_2,Y_2), ... (X_n, Y_n), tiene la ecuación:

$$Y = a_0 + a_1 x$$

donde las constantes a_0 y a_1 se determina mediante el sistema de ecuaciones.

$$\Sigma Y = a_0 N + a_1 \Sigma x \qquad (1)$$

$$\Sigma XY = a_0 \Sigma X + a_1 \Sigma X^2 \qquad (2)$$

Resolviendo las ecuaciones mostradas:

Despejando a_0 de la ecuación (1) y reemplazando a en (2)

$$a_0 = \frac{\Sigma Y - a_1 \Sigma x}{N}$$

Reemplazando a_0 en (2)

$$\Sigma XY = \left(\frac{\Sigma Y - a_1 \Sigma X}{N}\right)\Sigma x + a_1 \Sigma x^2 \qquad (3)$$

efectuando operaciones:

$$\Sigma XY = \frac{\Sigma X \Sigma Y}{N} - \frac{a_1 (\Sigma X)(\Sigma X)}{N} + a_1 \Sigma X^2 \qquad (4)$$

Factorizando

$$\frac{N\Sigma XY - (\Sigma X)(\Sigma Y)}{N} = a_1\left[\frac{N\Sigma X^2 - (\Sigma X)(\Sigma X)}{N}\right] \qquad (5)$$

Luego despejando a_1:

$$a_1 = \frac{N\Sigma XY - (\Sigma X)(\Sigma Y)}{N\Sigma X^2 - (\Sigma X)(\Sigma X)} = \frac{N\Sigma XY - (\Sigma X)(\Sigma Y)}{N\Sigma X^2 - (\Sigma X)^2} \qquad (6)$$

$$a_1 = \frac{N\Sigma XY - (\Sigma X)(\Sigma Y)}{N\Sigma X^2 - (\Sigma X)^2} \qquad (7)$$

Reemplazando a_1 en la ecuación (1)

$$\Sigma Y = a_0 N + \Sigma X \left[\frac{N\Sigma XY - (\Sigma X)(\Sigma Y)}{N\Sigma X^2 - (\Sigma X)^2} \right] \quad (8)$$

Multiplicando los paréntesis y despejando a

$$\frac{N(\Sigma X^2)(\Sigma Y) - (\Sigma Y)(\Sigma X)^2 - N(\Sigma X)(\Sigma XY) + (\Sigma X)(\Sigma X)(\Sigma Y)}{N\left[N\Sigma X^2 - (\Sigma X)^2\right]} = a_0$$

$$\frac{N(\Sigma X^2)(\Sigma Y) - (\Sigma Y)(\Sigma X)^2 - N(\Sigma X)(\Sigma XY) + (\Sigma X)^2(\Sigma Y)}{N\left[N\Sigma X^2 - (\Sigma X)^2\right]} = a_0$$

$$\frac{\left[(\Sigma X^2)(\Sigma Y) - (\Sigma X)(\Sigma XY)\right]}{\left[N\Sigma X^2 - (\Sigma X)^2\right]} = a_0$$

Por lo tanto

$$a_0 = \frac{(\Sigma X^2)(\Sigma Y) - (\Sigma X)(\Sigma XY)}{N\Sigma X^2 - (\Sigma X)^2}$$

RELACIONES NO LINEALES

Las relaciones no lineales pueden reducirse a veces a relaciones lineales mediante una adecuada transferencia de las variables.

1. Modelo de Parábolas de mínimos cuadrados.

La parábola de aproximación de mínimos cuadrados a la serie de puntos (X_1,Y_1) (X_2,Y_2)... (X_n, Y_n) tiene la ecuación:

$$Y = a_0 + a_1 X + a_2 X^2$$

donde las constantes se determinan resolviendo las ecuaciones:

$$\Sigma Y = a_0 N + a_1 \Sigma X + a_2 \Sigma X^2 \quad (1)$$

$$\Sigma XY = a_0 \Sigma X + a_1 \Sigma X^2 + a_2 \Sigma X^3 \quad (2)$$

$$\Sigma X^2 Y = a_0 \Sigma X^2 + a_1 \Sigma X^3 + a_2 \Sigma X^4 \quad (3)$$

Expresando las ecuaciones forma matricial

$$\begin{bmatrix} N & \Sigma X & \Sigma X^2 \\ \Sigma X & \Sigma X^2 & \Sigma X^3 \\ \Sigma X^2 & \Sigma X^3 & \Sigma X^4 \end{bmatrix} \begin{bmatrix} a_0 \\ a_1 \\ a_2 \end{bmatrix} = \begin{bmatrix} \Sigma Y \\ \Sigma XY \\ \Sigma X^2 Y \end{bmatrix}$$

$$a_0 = \frac{\begin{vmatrix} \Sigma Y & \Sigma X & \Sigma X^2 \\ \Sigma XY & \Sigma X^2 & \Sigma X^3 \\ \Sigma X^2 Y & \Sigma X^3 & \Sigma X^4 \end{vmatrix}}{\begin{vmatrix} N & \Sigma X & \Sigma X^2 \\ \Sigma X & \Sigma X^2 & \Sigma X^3 \\ \Sigma X^2 & \Sigma X^3 & \Sigma X^4 \end{vmatrix}} = \frac{\Delta_0}{\Delta_p}$$

$$a_1 = \frac{\begin{vmatrix} N & \Sigma X & \Sigma X^2 \\ \Sigma X & \Sigma XY & \Sigma X^3 \\ \Sigma X^2 & \Sigma X^2 Y & \Sigma X^4 \end{vmatrix}}{\Delta_p} = \frac{\Delta_1}{\Delta_p}$$

$$a_2 = \frac{\begin{vmatrix} N & \Sigma X & \Sigma Y \\ \Sigma X & \Sigma X^2 & \Sigma XY \\ \Sigma X^2 & \Sigma X^3 & \Sigma X^2 Y \end{vmatrix}}{\Delta_p} = \frac{\Delta_2}{\Delta_p}$$

Resolviendo las matrices separadamente, por el método de los cofactores.

$$\Delta_p = \begin{vmatrix} N & \Sigma X & \Sigma X^2 \\ \Sigma X & \Sigma X^2 & \Sigma X^3 \\ \Sigma X^2 & \Sigma X^3 & \Sigma X^4 \end{vmatrix} = N\begin{vmatrix} \Sigma X^2 & \Sigma X^3 \\ \Sigma X^3 & \Sigma X^4 \end{vmatrix} - \Sigma X\begin{vmatrix} \Sigma X & \Sigma X^3 \\ \Sigma X^2 & \Sigma X^4 \end{vmatrix} + \Sigma X^2\begin{vmatrix} \Sigma X & \Sigma X^2 \\ \Sigma X^2 & \Sigma X^3 \end{vmatrix} =$$

$$= N\left[(\Sigma X^2)(\Sigma X^4) - (\Sigma X^3)^2\right] - \Sigma X\left[(\Sigma X)(\Sigma X^4) - (\Sigma X^3)(\Sigma X^2)\right] + \Sigma X^2\left[(\Sigma X)(\Sigma X^3) - (\Sigma X^2)^2\right] =$$

$$= N(\Sigma X^2)(\Sigma X^4) - N(\Sigma X^3)^2 - (\Sigma X)^2(\Sigma X^4) + (\Sigma X)(\Sigma X^2)(\Sigma X^3) + (\Sigma X)(\Sigma X^2)(\Sigma X^3) - (\Sigma X^2)^3 =$$

$$= N(\Sigma X^2)(\Sigma X^4) - N(\Sigma X^3)^2 - (\Sigma X)^2(\Sigma X^4) - (\Sigma X^2)^3.$$

Calculando D_1:

$$\Delta_1 = \begin{vmatrix} \Sigma Y & \Sigma X & \Sigma X^2 \\ \Sigma XY & \Sigma X^2 & \Sigma X^3 \\ \Sigma X^2 Y & \Sigma X^3 & \Sigma X^4 \end{vmatrix} = \Sigma Y\begin{vmatrix} \Sigma X^2 & \Sigma X^3 \\ \Sigma X^3 & \Sigma X^4 \end{vmatrix} - \Sigma X\begin{vmatrix} \Sigma XY & \Sigma X^3 \\ \Sigma X^2 Y & \Sigma X^4 \end{vmatrix} + \Sigma X^2\begin{vmatrix} \Sigma XY & \Sigma X^2 \\ \Sigma X^2 Y & \Sigma X^3 \end{vmatrix}$$

$$= \Sigma Y\left[(\Sigma X^2)(\Sigma X^4) - (\Sigma X3)^2\right] - \Sigma X\left[(\Sigma XY)(\Sigma X^4) - (\Sigma X^3)(\Sigma X^2 Y)\right] + \Sigma X^2\left[(\Sigma XY)(\Sigma X^3) - (\Sigma X^2)(\Sigma X^2 Y)\right] =$$

$$= (\Sigma Y)(\Sigma X^2)(\Sigma X^4) - (\Sigma Y)(\Sigma X^3)^2 - (\Sigma X)(\Sigma XY)(\Sigma X^4) + (\Sigma X^3)(\Sigma X^2 Y)(\Sigma X) + (\Sigma X^2)(\Sigma XY)(\Sigma X^3) - (\Sigma X^2)^2(\Sigma X^2 Y)$$

Calculando D_2

$$\Delta_2 = \begin{vmatrix} N & \Sigma Y & \Sigma X^2 \\ \Sigma X & \Sigma XY & \Sigma X^3 \\ \Sigma X^2 & \Sigma X^2 Y & \Sigma X^4 \end{vmatrix} = N\begin{vmatrix} \Sigma XY & \Sigma X^3 \\ \Sigma X^2 Y & \Sigma X^4 \end{vmatrix} - \Sigma Y\begin{vmatrix} \Sigma X & \Sigma X^3 \\ \Sigma X^2 & \Sigma X^4 \end{vmatrix} + \Sigma X^2\begin{vmatrix} \Sigma X & \Sigma XY \\ \Sigma X^2 & \Sigma X^2 Y \end{vmatrix} =$$

$$= N\left[(\Sigma XY)(\Sigma X^4) - (\Sigma X^3)(\Sigma X^2 Y)\right] - \Sigma Y\left[(\Sigma X)(\Sigma X^4) - (\Sigma X^2)(\Sigma X^3)\right] + \Sigma X^2\left[(\Sigma X)(\Sigma X^2 Y) - (\Sigma X^2)(\Sigma XY)\right] =$$

$$N(\Sigma XY)(\Sigma X^4) - N(\Sigma X^3)(\Sigma X^2 Y) - (\Sigma Y)(\Sigma X)(\Sigma X^4) + (\Sigma Y)(\Sigma X^2)(\Sigma X^3) + (\Sigma X^2)(\Sigma X)(\Sigma X^2 Y) - (\Sigma X^2)^2(\Sigma XY)$$

Calculando D_3

$$\Delta_3 = \begin{vmatrix} N & \Sigma X & \Sigma Y \\ \Sigma X & \Sigma X^2 & \Sigma XY \\ \Sigma X^2 & \Sigma X^3 & \Sigma X^2 Y \end{vmatrix} = N\begin{vmatrix} \Sigma X^2 & \Sigma XY \\ \Sigma X^3 & \Sigma X^2 Y \end{vmatrix} - \Sigma X\begin{vmatrix} \Sigma X & \Sigma XY \\ \Sigma X^2 & \Sigma X^2 Y \end{vmatrix} + \Sigma Y\begin{vmatrix} \Sigma X & \Sigma XY \\ \Sigma X^2 & \Sigma X^2 Y \end{vmatrix}$$

$$= N\left[(\Sigma X^2)(\Sigma X^2 Y) - (\Sigma XY)(\Sigma X^3)\right] - \Sigma X\left[(\Sigma X)(\Sigma X^2 Y) - (\Sigma X2)(\Sigma XY)\right] + \Sigma Y\left[(\Sigma X)(\Sigma X^2 Y) - (\Sigma XY)(\Sigma X^2)\right] =$$

$$= N(\Sigma X^2)(\Sigma X^2 Y) - N(\Sigma XY)(\Sigma X^3) - (\Sigma X)^2(\Sigma X^2 Y) + (\Sigma X)(\Sigma X^2)(\Sigma XY) + (\Sigma Y)(\Sigma X)(\Sigma X^2 Y) - (\Sigma Y)(\Sigma XY)(\Sigma X^2)$$

2. Modelos multiecuacionales

La cuantificación de las relaciones que existen entre variables, cuando se trata de estudiar un fenómeno económico, requiere a veces de la introducción de más de una variable endógena. La formulación total del modelo, necesita tantas ecuaciones como variables endógenas. Por lo que el empleo de modelos multiecuacionales es una necesidad inevitable que surge a medida que la teoría económica quiere abarcar un mayor campo de estudio, es decir que se haga más general. La estimación de modelos uniecuacionales completos ya estaba resuelta por la Estadística.

En este caso se suponen G variables endógenas (y) y K predeterminadas (Z). Suponemos, además, que no existen en el modelo relaciones contables (en caso de existir pueden eliminarse previamente) lo que equivale decir que en todas las ecuaciones hay que incluir una perturbación aleatoria ξ por las razones econométricas.

Por tanto el modelo puede escribirse, en su forma más general de la forma siguiente:

$$\beta_{11}y_{1t}+....+\beta_{1G}y_{Gt}+\gamma_{11}Z_{1t}+....+\gamma_{1k}Z_{kt}=\varepsilon_{1t}$$

$$\beta_{G1}y_{1t}+....+\beta_{GG}y_{Gt}+\gamma_{G1}Z_{1t}+....+\gamma_{GkZkt}=\varepsilon_{Gt}$$

(1)

para t=1,2,.....T

en donde se supone una muestra de T observaciones que pueden ser temporales o atemporales.

En notación matricial detallada se tiene:

$$\begin{bmatrix}\beta_{11}....\beta_{G}\\ \vdots\\ \beta_{G1}....\beta_{GG}\end{bmatrix}\begin{bmatrix}y^{1t}\\ \vdots\\ y_{Gt}\end{bmatrix}+\begin{bmatrix}\gamma_{11}....\gamma_{1k}\\ \vdots\\ \gamma_{G1}....\gamma_{Gk}\end{bmatrix}\begin{bmatrix}Z_{1t}\\ \vdots\\ Z_{kt}\end{bmatrix}=\begin{bmatrix}\varepsilon_{1t}\\ \vdots\\ \varepsilon_{Gt}\end{bmatrix}$$

para t = 1,2, ... T

Haciendo:

$$\beta=\begin{bmatrix}\beta_{11} & \beta_{1G}\\ \\ \beta_{G1} & \beta_{GG}\end{bmatrix};\quad \Gamma=\begin{bmatrix}\gamma_{11}....\gamma_{1k}\\ \vdots \quad\quad \vdots\\ \gamma_{G1}....\gamma_{Gk}\end{bmatrix}$$

$$y=\begin{bmatrix}y_{1t}\\ \vdots\\ y_{Gt}\end{bmatrix};\quad Z_t=\begin{bmatrix}Z_{1t}\\ \vdots\\ Z_{kt}\end{bmatrix};\quad \varepsilon_t=\begin{bmatrix}\varepsilon_{1t}\\ \vdots\\ \varepsilon_{Gt}\end{bmatrix} \quad (3)$$

El sistema (1) se puede escribir en notación matricial condensado:

$$\beta y_t+\Gamma z_t=\varepsilon_t$$

para t = 1,2,... T

Para que el sistema sea completo debe verificarse:

$$|\beta|\neq 0$$

(2)

"Se consideran modelos de regresión en los cuales la variable dependiente o de respuesta pueden ser en sí misma de naturaleza dicótoma, 1 ó 0"

MODELOS DE REGRESIÓN CON LA VARIABLE DEPENDIENTE DICÓTOMA. LOS MODELOS MLP, LOGIT, PROBIT Y TOBIT

En este caso se supone implícitamente que la variable dependiente Y era cuantitativa, mientras que la variables explicativas pueden ser cuantitativas y cualitativas o una mezcla de las dos. Se consideran modelos de regresión en los cuales la variable dependiente o de respuesta pueden ser en sí misma de naturaleza dicótoma, tomando valor de 1 ó 0.

1. Variable dependiente dicótoma

Supongamos que se desea estudiar la participación de la fuerza laboral de hombres adultos en función de la tasa de desempleo, de la tasa de salarios promedio, del ingreso familiar, de la educación, etc. Una persona o bien está en la fuerza laboral o no lo está. Por tanto la variable dependiente que es la participación de la fuerza laboral, solamente pueden tener dos valores: 1 si la persona está en la fuerza laboral y 0 si él o ella no lo está.

La característica en esta situación es que la variable dependiente es del tipo que produce una respuesta de sí o no; es decir es dicótoma por naturaleza.

Existen cuatro enfoques más comunes utilizados para la estimación de estos modelos.

1. El modelo lineal de probabilidad (MLP)
2. El modelo logit
3. El modelo probit
4. El modelo tobit (regresión censurada)

2. Modelo lineal de probabilidad (MLP)

Veamos el siguiente modelo simple:

$$Y_i=\beta_1+\beta_2X_i+\mu_i \quad (1)$$

donde X = ingreso familiar
Y = 1 si la familia posee una casa
= 0 si la familia no posee una casa

El modelo (1) expresa la variable dicótoma Yi como una función lineal de las variables explicativas Xi; se denominan modelos lineales de probabilidad (MLP), puesto que E(Xi/Yi), la esperanza condicional de y_i dado x_i, puede ser interpretado como la probabilidad condicional de que el evento sucede dado x_i, es decir,

$$P_r(Y_i=1/X_i)$$

En el caso anterior de la probabilidad de que una familia posea una casa y tenga un ingreso de una cierta cantidad Xi. La justificación del nombre MLP para modelos como (1) puede ser la siguiente:

Suponiendo que:

$$E(\mu_i) = 0$$

como es lo usual (para obtener estimadores insesgados), se obtiene:

$$E(Y_i / X_i) = \beta 1 + \beta_2 X_i \quad (2)$$

P_i = probabilidad de que Y......... probabilidad de que (es decir de que el evento no ocurra), la variable y tiene la siguiente distribución:

Y_i	Probabilidad
0	1 - P_i

$$Total\ \frac{P_i}{1}$$

Por definición de esperanza matemática, se obtiene:

$$E(Y_i) = 0(4 - P_i) + 1(P_i) = P_i \quad (3)$$

Comparando (2) y (3) se puede igualar

$$E(Y_i / X_i) = \beta_1 + \beta_2 X_i = P_i \quad (4)$$

Es decir la esperanza condicional del modelo (1) puede interpretarse como la probabilidad condicional Y_i.

La probabilidad P_i debe encontrarse entre 0 y 1 y se tiene la restricción

$$0 \leq E(Y_i / X_i) \leq 1$$

Es decir la esperanza condicional o probabilidad condicional debe encontrarse entre 0 y 1.

3. Modelo Logit

Representación de la propiedad de vivienda antes mencionado.

$$P_i = E(Y = 1 / Xi)$$

$$P_i = E(Y = 1 / Xi) = \frac{1}{1 + e^{-(\beta_1 + \beta_2 Xi)}} \quad (5)$$

Pi probabilidad de poseer una casa dada por (5) entonces (1-P_i) la probabilidad de no poseer una casa.

$$1 - P_i = \frac{1}{1 + e^{-z_i}}$$

$$1 - P_i = \frac{1}{1 + e^{-z_i}} \quad (6)$$

$$\frac{P_i}{1 - P_i} = \frac{1 + e^{z_i}}{1 + e^{-z_i}} = e^{z_i} \quad (7)$$

Tomando logaritmos

$$Li = \ln\left(\frac{P_i}{1 - P_i}\right) = Z_i = \beta_1 + \beta_2 X_i \quad (8)$$

4. Modelo Probit

Para mostrar el modelo Probit suponemos que el ejemplo de propiedad de vivienda, la decisión iesima familia de poseer una casa o de no poseerla depende de un índice de conveniencia no observable Ii que está determinado por una o varias variables explicativas, por ejemplo el ingreso Xi, de tal manera que entre mayor sea el índice Ii, mayor será la probabilidad de que la familia posea vivienda. El índice Ii se expresa como:

$$Ii = \beta_1 + \beta_2 X_i$$

Donde Xi es el ingreso de la iésima familia.

Dado el supuesto de normalidad, la probabilidad de Ii sea menos o igual que Ii puede ser calculada a partir de:

$$P_i = P_r(Y=1) = P_r(Ii \leq Ii) = F(Ii) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{T} e^{-t^2/2}\, dt$$

$$= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\beta_1 + \beta_2 X_i} e^{-t^2/2}\, dt \quad (10)$$

Donde t es una variable normal estandarizada, es decir

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

4. Modelo Tobit

En términos matemáticos se puede expresar el modelo Tobit como

$$Y_i = \beta_1 + \beta_2 X_{2i} + \mu_{2i} = \text{ si RHS>0} \quad (11)$$

= 0, en los demás casos

Donde RHS = lado derecho

BIBLIOGRAFÍA

- Manual de Econometría de Kleir Lawrence R. Aguilas – Madrid 1958.
- Introducción a la Econometría de Kleir Lawrence R. Aguilas – Madrid 1968.
- Fundamentos y posibilidades de la Econometría de Alfonso García Barbancho – Ed. Ariel. Barcelona 1962.
- Contribución a la Teoría de la Política Económica de George R. Fewel R. Fondo de Cultura Económica 1987.
- Planificación y Modelos Econométricos de Aznar Grasa A. – Ed. Pirámide – Madrid 1978.
- Econometría de Domador Gujarati – Ed. Mc Graw Hill 1998.