Serie Apuntes de Clase ΩΒΓ

N° 07
Octubre de 2014

MODELOS DE VARIABLE DEPENDIENTE DISCRETA: EL MODELO LOGIT Y PROBIT

Rafael Bustamante Romaní

UNIVERSIDAD NACIONAL MAYOR DE SAN MARCOS
Universidad del Perú, DECANA DE AMÉRICA

FACULTAD DE CIENCIAS ECONÓMICAS

La **Serie Apuntes de Clase Omega Beta Gamma** tiene por objetivo difundir los materiales de enseñanza generados por los docentes que tienen a su cargo el desarrollo de las asignaturas que forman parte de los Planes de Estudios de las Escuelas Académico-Profesionales de la Facultad de Ciencias Económicas de la Universidad Nacional Mayor de San Marcos. Estos documentos buscan proporcionar a los estudiantes la explicación de algunos temas específicos que son abordados en su formación universitaria.

Encargados de la serie:

Bustamante Romaní, Rafael.
rbustamanter@unmsm.edu.pe

Cisneros García, Juan Manuel.
jcisnerosg@unmsm.edu.pe

Facultad de Ciencias Económicas.
Universidad Nacional Mayor de San Marcos.
Calle Germán Amézaga N° 375.
Ciudad Universitaria, Lima 1. Perú.

La **Serie Apuntes de Clase ΩΒΓ** es promovida y desarrollada por un colectivo de docentes del Departamento de Economía de la Universidad Nacional Mayor de San Marcos.
El contenido de cada publicación es íntegramente responsabilidad de cada autor, no representa necesariamente los puntos de vista de los integrantes del colectivo, ni de la Universidad.

Modelos de variable dependiente limitada

Rafael Bustamante [◊]

Resumen

Este trabajo pretende señalar que las especificaciones del comportamiento económico de los agentes económicos en su proceso de elección individual, en el que la decisión está orientada por la maximización de la utilidad. Para ello se especifican los modelos de probabilidad lineal, los modelos de probabilidad no lineal expresados en los denominados: Logit y Probit en su versión básica. Asimismo se plantea los procedimientos básicos de cómo se deben manejar las variables y los principales indicadores, inferencia estadística, análisis del efecto marginal y las medidas de bondad de ajuste.

Palabras Claves: Modelos de elección discreta, especificaciones, Logit, Probit

Clasificación JEL: C2, C25

[◊] Doctorado en Economía con mención en los Recursos Naturales (c), Universidad Nacional Autónoma de México. MBA Gerencial (c), CENTRUM Pontificia Universidad Católica del Perú. Maestría en Economía con mención en Finanzas, Universidad Nacional Mayor de San Marcos. B. Sc. Economía, UNMSM. Profesor Auxiliar del Departamento de Economía de la UNMSM. Investigador asociado al Instituto de Investigaciones FCE - UNMSM. Contacto: rbustamanter@unmsm.edu.pe.

El autor agradece la colaboración en la elaboración del presente documento a Jimmy Quijano Sicos, alumno de la Facultad de Ciencias Económicas de la UNMSM. Contacto: dxlx_l@hotmail.com.

I. INTRODUCCIÓN

En ocasiones los ciudadanos, en general y los economistas, en particular se encuentran ante situaciones en que deben elegir o decidir entre posibles alternativas. En el caso de que estas alternativas fuesen las de la modelización adquiere un carácter especial denominándose modelos de respuesta dicotómica (o binaria). Algunos de los ejemplos que se pueden plantear de este tipo de modelización son los siguientes:

Una familia puede o no tener vivienda en propiedad (atendiendo a un conjunto de características económicas: nivel de renta o ingresos mensuales, nivel cultural del cabeza de familia, edad del cabeza de familia, etc.).

- Una persona activa puede estar en situación de paro o trabajando.
- Un trabajador se puede plantear afiliarse o no a un sindicato.
- Un trabajador puede optar entre seguir o no una huelga.
- Un ciudadano se puede plantear elegir el medio de transponer, público o privado, para su desplazamiento al lugar de trabajo.
- Un ciudadano decide comprar (cambiar) un coche.
- Un individuo decide suscribir o no una póliza de seguro.
- Una familia se puede plantear el tipo de escuela (pública / privada) a la que desea mandar a sus hijos.
- Un banco o una entidad financiera se encuentra ante la situación de conceder o no un crédito a un agente económico (ciudadano, familia. empresa, corporación).
- Una entidad financiera puede estudiar la probabilidad de si un crédito concedido a un cliente será devuelto o no en la fecha de vencimiento.

El planteamiento de estos modelos se fundamenta en la ecuación:

$$Prob(y_i = 1) = F(\beta' X_i) = Prob(U_{i1} > U_{i0}) \tag{1}$$

Donde se especifica que el individuo se enfrenta a un proceso de decisión entre dos alternativas denominadas 0 y 1, por ejemplo, comprar o no comprar un bien, afiliarse a un sindicato o no, etc., decidiendo realizar aquella que le proporciona mayor utilidad. Así, en el caso de que se opte por la opción 1 se tiene que:

$$Prob(y_i = 1) = F(\beta' X_i) = Prob(U_{i1} > U_{i0}) \tag{2}$$

Además, esta decisión está condicionada al valor de la función de distribución en el punto $\beta' X_i$ y, por tanto, según se establezcan las hipótesis de cómo es esta función de distribución, se especifican diferentes modelos de elección dicotómica. En este tema, se supone que F es una función de distribución uniforme (1) y, por tanto, F= 1. La modelización de la decisión se establecería a través de la ecuación:

$$Prob(y_i = 1) = \beta' X_i \tag{3}$$

Por tanto el Modelo Lineal de Probabilidad quedaría especificado de la siguiente forma:

$$Y_I = \beta' X_i + u_i \tag{4}$$

Una característica específica de este tipo de modelos es la distribución de la muestra, cuya representación gráfica para una sola variable explicativa configura una nube de puntos: que se obtiene a través de su representación sobre un diagrama definido por el regresor en el eje de abscisas y por el regresando, que toma los valores uno o cero, en el eje de ordenadas.

En la nube de puntos de la figura 1 se aprecia que la muestra está dividida en dos subgrupos. Uno de ellos está dispuesto alrededor de la recta Y = 0 que representa al conjunto de individuos que no optaron por la realización de la opción y

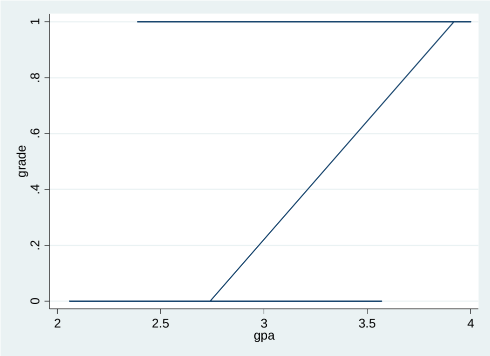
[1]La distribución uniforme continua es una familia de distribuciones de probabilidad para variables aleatorias continuas, tales que cada miembro de la familia, todos los intervalos de igual longitud en la distribución en su rango son igualmente probables. El dominio está definido por dos parámetros, *a* y *b*, que son sus valores mínimo y máximo. La distribución es a menudo escrita en forma abreviada como *U*(*a*,*b*).

el otro subgrupo está dispuesto en torno a la recta Y = 1, que representa a los individuos que optaron por realizarla opción.

La elaboración del modelo econométrico requiere el ajuste de esa nube de puntos a una función capaz de explicar el comportamiento de la muestra. En el caso del Modelo Lineal de Probabilidad, el ajuste realizadores una recta de regresión (véase figura 1).

Figura 1:

Modelo Lineal de Probabilidad (MPL)



II. MODELO SIMPLE

Desde el punto de vista general se pueden relacionar, a través de una ecuación de comportamiento, una variable endógena que indica un «hecho» o «suceso» (la variable endógena real tan sólo puede tomar los valores uno o cero) en función de una variable explicativa o característica, por ejemplo, el «hecho» de poseer ordenador en función del nivel de renta familiar. Dicho modelo se especificó en el apartado anterior de la forma siguiente:

$$Y_i = \beta' X_i + u_i = [\beta_1 \beta_1]' [1 X_{2i}] \quad (5)$$

Las herramientas metodológicas que se presentan a continuación son aplicables a la información obtenidas en un momento en el tiempo para un grupo determinado de “individuos”, sean estos personas , empresas, bancos, etc. por lo mismo, el componente temporal pierde (momentáneamente) importancia, centrándose ahora el interés en las similitudes o disparidades de ese grupo en determinado instante de tiempo; es así que nuestras observaciones pasarán a tener el subíndice i (y ya no t). Donde i hace referencia al individuo i de la muestra.

Pese a estas características de la información, el uso de MCO no se invalida siempre que la dependiente sea una variable continua sin ninguna limitación, siendo solo necesario ser cuidadoso con la posible heteroscedasticidad del modelo estimado, la misma que debe ser convenientemente corregida. No obstante, cuando la dependiente no satisface estas condiciones, el estimador MCO deja de ser el más apropiado surgiendo otros estimadores de mejores propiedades finitas y asintóticas. Son éstos estimadores el centro del análisis de las siguientes páginas.

Debido a que el problema se centra en la dependiente, dividiremos el análisis sobre la base de las características que esta muestre, distinguiendo entre una dependiente discreta de aquella que siendo continua siempre tiene rangos limitados de trabajo.

II.1 LAS VARIABLES ALEATORIAS BINOMIALES

Son aquellas que solo toman dos valores, tradicionalmente 0 y 1, es decir

$$Y_i = \begin{cases} 1; si\ se\ cumple\ cierta\ condicion \\ 0; de\ cualquier\ otra\ forma \end{cases} \quad (6)$$

$$Y = \beta_{1+}\beta_2 X_2 + \beta_3 X_3 + U \quad (7)$$

*Las variables que nos vienen a la mente: La edad, la formación el estado civil, el número de hijos y ciertas características económicas.

Donde Y es generado mediante la siguiente regla:

Cuando se trata de decidir si o no. Por lo que no resultan adecuados los métodos de regresión.

$$Y_i = \begin{cases} 1; si\ es\ una\ mujer \\ 0; en\ otro\ caso \end{cases} \quad (8)$$

II.2 MODELO DE PROBABILIDAD LINEAL (MPL)

Supongamos que decidimos modelar la variable de (1) usando un modelo lineal de la forma:

$$Y_i = \beta' X_i + U_i \tag{9}$$

Si Y toma valores entre cero y uno. Un modelo simple de regresión lineal de Y sobre X no es apropiado, debido a que pone restricciones inadecuadas sobre los residuos del modelo.
Además del valor pronosticado de Y es $\hat{Y}$

$$E(u_i) = 0 \tag{10}$$

Podemos decir que:

$$E\left(\frac{Y_i}{X_i}\right) = 1^* Prob(Y_i = 1) + 0^* Prob(Y_1 = 0) = Prob(Y_i = 1) \tag{11}$$

$$\left(\frac{Y_i}{X_i}\right) = b' X_i \tag{12}$$

Por lo que se puede concluir que:

$$Prob(Y_i = 1) = b' X_i = E\left(\frac{Y_i}{X_i}\right) \tag{13}$$

Es decir la probabilidad de que la persona trabaje es $b' X_i$, la que por lógica tiene que estar entre cero y uno. No obstante, en el modelo no hay nada que restrinja a Y_i a estarlo. Además se tiene problemas con el error, pues este toma solo dos valores a saber:

Cuadro Nº 1

Si	U_i	Probabilidad:
$y_i = 1$	1-$\beta' X_i$	$-\beta' X_i$ $\quad Prob(y_i = 1)$
$y_i = 0$	-$\beta' X_i$	1-$\beta' X_i$ $\quad Prob(y_i = 0)$
Total	1	

Es decir el error el error es Binomial y no Normal, siendo su varianza igual a:

$$Var\,(U_i) = (1 - \beta' X_i)2(\beta' X_i) + (-\beta' X_i)2(1 - \beta' X_i) = \beta' X_i(1 - \beta' X_i) \qquad (14)$$

De forma tal que, como depende de las observaciones, termina siendo heterocedástico. De esta forma podemos concluir que existen tres grandes limitaciones para el uso del estimador MCO en estos modelos:

- El error es heterocedástico.
- El error no es normal
- Nada restringe a $(Y_i = 1) = b' X_i = E\left(\frac{Y_i}{X_i}\right)$ a estar entre cero y uno.

Los dos primeros problemas pueden ser resueltos con relativa facilidad, utilizando MCG y ampliando la muestra, respectivamente. No obstante no existe forma de resolver el último problema, razón por la cual nos vemos en la necesidad de trabajar con un método que garantice que la probabilidad resultante se mueva entre esos límites; para ello se recurrirá a la función de distribución acumulada del error, la cual será utilizada para obtener el estimador MCO en estos modelos.

II.3 MODELO LINEAL DE PROBABILIDAD PONDERADO 0 MLP ESTIMADO MEDIANTE MÍNIMOS CUADRADOS GENERALIZADOS: SUS LIMITACIONES (Bernardí Cabrer Borrás & Amparo Sancho Pérez, Guada, 2001)

En el apartado anterior se han expuesto los problemas que lleva asociada la estimación por MCO del Modelo Lineal de Probabilidad. Ante estos problemas es necesario buscar una alternativa a la estimación del modelo. Dado que uno de los problemas más importantes que presenta el proceso de la estimación por MICO es la presencia de un problema de heteroscedasticidad, se plantea una solución posible para estimar estos modelos dicotómicos desde la óptica de la linealidad. Para ello se utiliza la estimación por Mínimos Cuadrados Generalizados (MCG) del Modelo Lineal de Probabilidad. A este tipo de modelos se les denomina MLP ponderados.

Los pasos a realizar son los siguientes:

Se estima el modelo $Y_i = \beta' X_i + U_i$mediante el método de los MCO sin tener en cuenta el problema de la heteroscedasticidad, obteniendo el valor estimado Y_i; que se puede considerar como la estimación de la esperanza condicional o bien de la probabilidad condicional. Las estimaciones $\widehat{w}_i$ se utilizan para estimar la varianza de la perturbación aleatoria, $\widehat{w}_i$ luego se obtiene la siguiente estimación de la varianza:

$$\widehat{w}_i = \widehat{Y}_i(1-\widehat{Y}_i) \qquad (15)$$

Si los valores estimados de Y_i son mayores que la unidad se deben sustituir por la unidad. En este caso el valor resultante de Y_i; será cero. Este hecho provocaría serios problemas al utilizar la variable$\widehat{w}_i$ como ponderador (se tendría que dividir un número por cero). Es por ello por lo que, en definitiva, se opta entre las dos opciones alternativas siguientes:

- Se eliminan estas observaciones, con lo que se pierde información (se reduce el tamaño de la muestra), por lo que los estimadores que se obtienen ya no son robustos.
- Se sustituyen los valores mayores o iguales a la unidad por 0,999.

Si los valores estimados de Y_i son negativos (menores que cero) se deben sustituir por cero. En este caso el valor resultante de $\widehat{w}_i$ también será cero. Esto provocaría serios problemas al utilizar la variable sí; como ponderador (se tendría que dividir un número por cero). Es por ello por lo que, en definitiva, se opta entre las dos alternativas siguientes:

- Se eliminan estas observaciones con lo que se pierde información (se reduce el tamaño de la muestra), por lo que los estimadores que se obtienen ya no son robustos.
- Se sustituyen los valores menores o iguales a cero por 0.001.

Se pondera el modelo dividiendo ambos miembros de la ecuación por la desviación típica estimada, $\sqrt{\widehat{w}_i} = \sqrt{P_i(1-P_i)}$, con el fin de transformar el modelo en homoscedástico, esto es:

$$Y = \beta_1 + \beta_2 X_{2i} + \beta_3 X_{3i} + \cdots + \beta_k X_{ki} + u_i \quad (16)$$

$$\frac{Y_i}{\sqrt{\widehat{w}_i}} = \frac{\beta_1}{\sqrt{\widehat{w}_i}} + \beta_1 \frac{X_{2i}}{\sqrt{\widehat{w}_i}} + \beta_1 \frac{X_{3i}}{\sqrt{\widehat{w}_i}} + \cdots + \beta_1 \frac{X_{ki}}{\sqrt{\widehat{w}_i}} + \frac{u_i}{\sqrt{\widehat{w}_i}} \quad (17)$$

La estimación del modelo transformado, mediante el método MCO es equivalente a aplicar MCG en el modelo y en ambos casos se obtienen estimaciones eficientes de los coeficientes de regresión.

Los problemas asociados a la estimación del MLP mediante son análogos a los que presenta la estimación del modelo por MCO ya que:

Aunque se puede demostrar que las estimaciones llevadas a cabo mediante MCG son eficientes, en la práctica no se garantiza que los resultados obtenidos para las estimaciones de la variable Y_i, que son estimaciones de la probabilidad Pi, no puedan ser negativos o bien mayores que uno. Es decir, que las estimaciones de la variable Y_i i, o predicciones de P_i , pueden tomar valores fuera del intervalo (0,1).

- ✓ Dado que se omiten aquellas observaciones que no resultan coherentes con una interpretación probabilística, los estimadores obtenidos por MCG no son robustos.

- ✓ El coeficiente de determinación continúa siendo excesivamente bajo (subestimación del coeficiente de determinación).

- ✓ Debido a la pérdida del término independiente en el modelo, ya que se han ponderado todas las variables de la ecuación por $\hat{w}_i$, la suma de todas las probabilidades no será necesariamente igual a la unidad.

- ✓ La omisión del término independiente puede provocar que la suma de los residuos sea distinta de cero. Este error de especificación del modelo puede tener consecuencias sobre el coeficiente de determinación (puede ser negativo), la función de verosimilitud estimada a partir de los residuos (la suma de los residuos no es necesariamente igual a cero) y los estadísticos que se obtienen a partir de ella.

- ✓ La no normalidad de las perturbaciones aleatorias se sigue manteniendo a pesar de la transformación realizada. Por tanto, los tests de significación tradicionales quedan invalidados (los tests paramétricos: t-Student. F de Snedecor, etc.). No obstante, el tamaño de la muestra tiene un papel importante en estos modelos ya que si ésta es suficientemente grande, los contrastes sí que sirven de forma asintótica.

- ✓ Todos estos problemas nos llevan a la búsqueda de modelos y métodos de estimación alternativos (generalmente no lineales) que solucionen los efectos producidos por los estimadores anteriormente expuestos

II.4 LOS MODELOS PROBABILISTICOS LOGIT Y PROBIT

Los modelos dicotómicos modelaran los problemas asociados a la toma de decisiones cuando los agentes económicos se enfrentan a un proceso de elección binaria. El criterio de selección entre opciones depende de la probabilidad asociada a cada una de las alternativas posibles que puede tener un individuo.

El Modelo Lineal de Probabilidad (MPL) no es capaz de dar una respuesta adecuada a los problemas que presentan los procesos de decisión dicotómica. Por esta razón este tema se dedica a un planteamiento no lineal de los modelos de elección dicotómica que, sin duda, solucionan algunos de los problemas asociados al MLP.

El proceso de elección de un individuo en un modelo dicotómico depende de que la utilidad que obtiene el individuo en una opción supere la utilidad que le proporciona la opción complementaria. Es decir el individuo opta por la alternativa uno $(y_i = 1)$ frente a la alternativa cero $(y_i = 0)$ si la utilidad que le proporciona esta opción. U_{i1} supera la de la opción cero, U_{i0}.

Ahora bien esta utilidad depende de los valores que toman las características del agente económico y de la opción a elegir que serán las variables del problema representadas mediante la combinación lineal $\beta' X_i$.

Desde el punto de vista formal, se tiene:

$$Prob(Y_i = 1) = F(\beta' X_i) = Prob\ (U_{i1} > U_{i0}) \qquad (18)$$

Dependiendo de la función de distribución que se asocia al proceso de decisión, $\beta' X_i$[2], el modelo especificado es diferente.

Supongamos que se tiene el siguiente modelo:

$$Y^* = X_I \beta + u_i \qquad (19)$$

Dónde:

$E(u_i) = 0$, podemos decir que: $E\left(\frac{Y_i}{X_i}\right) = \beta' X_i$

En el que Y^*_i no es una variable observable e igual, por ejemplo, al "número de horas deseadas de trabajo" la variable que se observa es Y_i, la misma que toma el valor de uno si $Y^*_i > 0$ y de 0 si $Y^*_i < 0$. Note que ahora $\beta' X_i$ La función de verosimilitud pertinente, para los n individuos de una muestra, estaría dado por:

Debido a que el modelo de probabilidad lineal tiene intrínsecamente defectos se consideran modelos alternativos.

Para ello planteamos el siguiente caso:

Si se tiene una muestra de "N" observaciones con las "n" primera observaciones iguales a cero, y las "N-n" observaciones restantes iguales a uno, entonces la distribución binomial de la variable aleatoria Y_i se puede escribir como una Bernoulli:

---

[2]Un enfoque alternativo para analizar el problema de la elección de un agente es el enfoque de la variable latente.

$$Prob(Y_i = y_i) = \alpha^{y_i}(1-\alpha)^{y_i} \qquad (20)$$

Observaciones:

Cada evento tiene dos resultados posibles $Y_i = 1(Y_i{}^* > 0)$ ó $Y_i = 0(Y_i{}^* \leq 0)$.

Cada evento es independiente. Es decir, el resultado (éxito o fracaso) de cualquier ensayo (o evento) es independiente de cualquier otro evento.

La probabilidad de permanencia permanece constante de evento a evento.

Además si definimos a la variable latente:

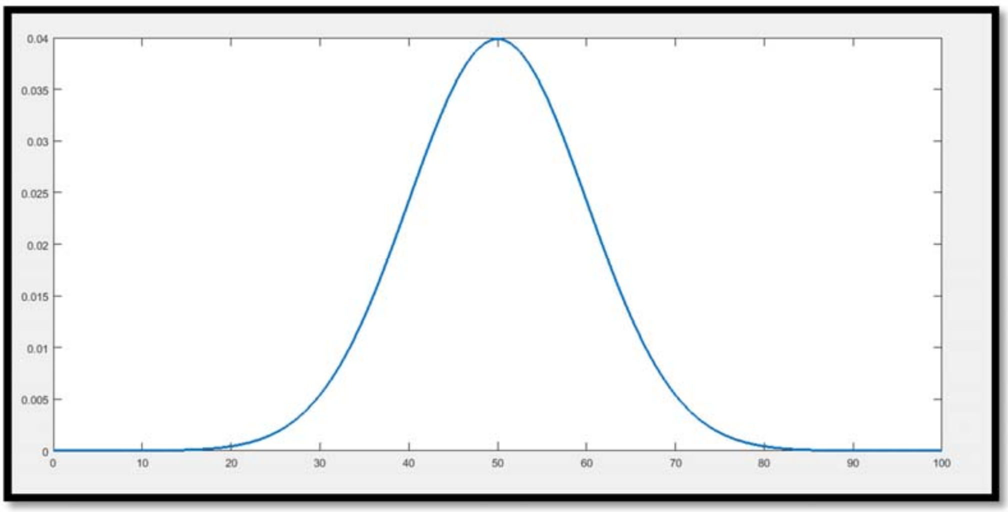
$$\boldsymbol{Y_I^* = \beta' X_i + u_i} \qquad (21)$$

$$\boldsymbol{y_i = g(y_i^*) = \begin{cases} 1, si\ y_i^* > 0\ ocurre\ el\ evento \\ 0 \quad si\ no\ ocurre\ el\ evento \end{cases}} \qquad (22)$$

$$Prob\ (y_i^* > 0) Prob(y_i = 1) = Prob(\beta' X_i + u_i > 0) = Prob\ (-\beta' X_i < u_i)$$

Si asumimos que u_i tiene una función de densidad continua y simétrica tenemos como se muestra en la figura N 2

Figura N 2



Por propiedad de simetría tenemos que:

$$Prob\ (u_i > -\beta' X_i) = 1 - Prob(u_i > -\beta' X_i) = 1 - F(-\beta' X_i) = F(\beta' X_i) = Prob(u_i < \beta' X_i)$$

Entonces:

Volviendo a lo planteado

Planteamos la FV:

$$L = F(Y_1, Y_2, \dots, Y_N) = F(Y_1)F(Y_2)F(Y_3)\dots F(Y_N)$$

$$L = F(Y_1, Y_2, \dots, Y_N) = Prob(Y_i = y_1)Prob(Y_i = y_2)\dots Prob(Y_i = y_n)$$

$$xProb(Y_i = y_{n+1})Prob(Y_i = y_{n+2})Prob(Y_i = y_{n+3})\dots Prob(Y_i = y_N)$$

$$L = \alpha^{y_1}(1-\alpha)^{1-y_1}\alpha^{y_2}(1-\alpha)^{1-y_2}\alpha^{y_3}(1-\alpha)^{1-y_3}\dots\ \alpha^{y_n}(1-\alpha)^{1-y_n}*$$

$$\alpha^{y_{n+1}}(1-\alpha)^{1-y_{n+1}}\alpha^{y_{n+2}}(1-\alpha)^{1-y_{n+2}}\alpha^{y_{n+3}}(1-\alpha)^{1-y_{n+3}}\dots\ \alpha^{y_N}(1-\alpha)^{1-y_N}$$

$$L = \left(F(\beta' X_i)\right)^{y_1}(1 - F(\beta' X_i))^{1-y_1}\left(F(\beta' X_i)\right)^{y_2}(1 - F(\beta' X_i))^{1-y_2}\left(F(\beta' X_i)\right)^{y_3}(1 - F(\beta' X_i))^{1-y_{32}}$$

$$\dots\ \alpha^{y_n}\left(1 - F(\beta' X_i)\right)^{1-y_n} * F(\beta' X_i)^{y_{n+1}}(1 - F(\beta' X_i))^{1-y_{n+1}}F(\beta' X_i)^{y_{n+2}}(1 - F(\beta' X_i))^{1-y_{n+2}}$$

$$F(\beta' X_i)^{y_{n+3}}(1 - F(\beta' X_i))^{1-y_{n+3}}\dots F(\beta' X_i)^{y_N}(1 - F(\beta' X_i))^{1-y_N} \qquad (23)$$

Dónde:

$$Prob\ (y_i = 1) = F(\beta' X_i) \qquad (24)$$

Entonces tenemos la función de Verosimilitud que viene expresada por:

$$L = \prod_1^n(1 - F(\beta' X_i))^{1-y_1}\prod_{i=n+1}^N(F(\beta' X_i))^{y_1} = \prod_1^n(1 - P_I)^{1-y_1}\prod_{i=n+1}^N(P_I)^{y_1} \qquad (24)$$

$$para\ i = 1,2,3, \dots n \Longrightarrow y_i = 0$$

$$para\ i = n+1, n+2, \dots, N \Longrightarrow y_i = 1$$

Si F (u) es normal estándar estaríamos hablando del modelo Probit, mientras que si fuera logística nos referiríamos al modelo logit. Cabe mencionar que como ambas funciones son simétricas podemos concluir que:

$$Prob(Y_i = 1) = 1 - F(\beta' X_i) = F(\beta' X_i)$$

Comparemos un poca más de estas funciones. La principal diferencia entre ellas es la **amplitud** de sus colas, ya que la logística tiene cola más ancha. Por los mismos que los resultados que se obtienen con cada una de ellas no son comparables. Dado que en el modelo Probit el uso de una normal estándar arroja β's estandarizados (Siendo σ=1), la comparación con los β's logit requiere estandarizar estos últimos también para lo cual hay que dividir los estimados entre la desviación estándar, que es igual a $\frac{\pi}{\sqrt{3}}$.

Es decir:

$$\frac{\beta_L \sqrt{3}}{\pi} vs \beta_p \qquad (25)$$

Dado que no hay forma de saber a priori cómo se comportan los errores de los modelos que queremos estimar, y que la diferencia entre estas funciones es relativamente útil, la elección entre logit y probit dependerá **del mejor ajuste** que se logre utilizando una indistintamente del otro.

$$Prob\ (Y_i = 1) = \Phi(X_i\beta) = \int_{-\infty}^{X_i\beta} \frac{1}{\sqrt{2\pi}} exp\left(\frac{-z^2}{2}\right) dz \qquad (26)$$

La función de distribución estándar restringe a la probabilidad entre uno y cero.

$$\Phi(Z) = 1 \Phi(Z) = 0$$

$$Lim \to \infty\ Lim \to \infty$$

II.5 EL MODELO PROBIT

Definimos una variable latente tal que:

$$Y_i^* = X_i\beta + \varepsilon_i \tag{27}$$

Aplicando la regla de observalidad que generan las alternativas que se dan en la realidad. Desde este punto de vista el modelo dicotómico se expresaría como:

$$G(Y_i^*) = Y_i = \begin{cases} 1 \dots \dots \dots \dots si\ Y_i^* > 0 \\ 0 \dots de\ otra\ forma \end{cases} \tag{28}$$

Donde $G(.)$ es una transformación monótona que se le aplica a Y_i^*

$\varepsilon_i \sim N(0, \sigma^2)$.

Y_i^* Tiene media $\beta' X_i$ y varianza σ^2, y la relación establecida entre la probabilidad de que el suceso ocurra y el valor de la variable latente es monótona creciente. Cada individuo realiza la elección comparando su valor$\beta' X_i$ con el valor crítico Y_i^* que refleja sus preferencias.

Suponiendo que es una variable aleatoria, un individuo elegiría la opción 1 si $\beta' X_i \geq Y_i^*\ y\ 0$ en caso contrario.

Entonces tenemos:

$$Prob\ (Y_i = 1) = Prob\ (Y_i{}^* > 0)$$

$$= Prob\ (X_i\beta + \varepsilon_i > 0)$$

$$= Prob(\varepsilon_i > -X_i\beta)$$

Estandarizando esta relación:

$$Prob\ (Y_i = 1) Prob(\frac{\varepsilon_i}{\sigma} > -X_i \frac{\beta}{\sigma}) \tag{29}$$

Donde σ^2 es la varianza de ε_i.

Debido a la propiedad de simetría de las funciones de distribución Logística y Normal tenemos lo siguiente.

$$Prob\ (Y_i = 1) = Prob(\frac{\varepsilon_i}{\sigma} > -X_i\frac{\beta}{\sigma})$$

$$= Prob(\frac{\varepsilon_i}{\sigma} < X_i\frac{\beta}{\sigma})$$

$$Prob\ (Y_i = 1) = \Phi(X_i\frac{\beta}{\sigma})$$

De esto se obtiene lo siguiente.

$$Prob\ (Y_i = 0) = 1 - Prob\ (Y_i = 1) = 1 - Prob\ \left(\frac{\varepsilon_i}{\sigma} < \frac{X_i\beta}{\sigma}\right) = 1 - \Phi(X_i\beta)$$

Si además suponemos que tenemos una muestra independientemente de n observaciones con las m primeras iguales a cero, y las n-m restantes iguales a uno, entonces se tiene lo siguiente:

La variable aleatoria Y_i se comporta como un proceso Bernoulli[3].

Distribuida la función de Verosimilitud para la muestra es:

$$L = F(Y_1, Y_2, \dots, Y_n) = F(Y_1)F(Y_2)F(Y_3)\dots F(Y_n) =$$

$$Prob(Y_1 = 0).Prob(Y_2 = 0)\dots.Prob(Y_m = 0).Prob(Y_{m+1} = 0).Prob(Y_{m+2} = 0)\dots\ Prob(Y_n = 1)$$

$$L = \alpha^{y_1}(1-\alpha)^{1-y_1}\alpha^{y_2}(1-\alpha)^{1-y_2}\alpha^{y_3}(1-\alpha)^{1-y_3}\dots\ \alpha^{y_m}(1-\alpha)^{1-y_m}\dots\ \alpha^{y_N}(1-\alpha)^{1-y_N}$$

$$\alpha = Prob(Y = 1) = F(\beta' X_i)$$

$$1 - \alpha = Prob(Y = 0) = 1 - F(\beta' X_i)$$

$$L = \left(F(\beta' X_i)\right)^{y_1}\left(1 - F(\beta' X_i)\right)^{1-y_1}\left(F(\beta' X_i)\right)^{y_2}\left(1 - F(\beta' X_i)\right)^{1-y_2}\dots$$

$$\dots\left(F(\beta' X_i)\right)^{y_m}\left(1 - F(\beta' X_i)\right)^{1-y_m}\left(F(\beta' X_i)\right)^{y_{m+1}}\left(1 - F(\beta' X_i)\right)^{1-y_{m+1}}\dots$$

$$\dots\left(F(\beta' X_i)\right)^{y_n}\left(1 - F(\beta' X_i)\right)^{1-y_n}$$

[3] $Prob(Y = y_i) = \alpha^{y_1}(1-\alpha)^{1-y_1}$ Dónde: $\alpha = Prob(Y = 1)$, $1 - \alpha = Prob(Y = 0)$

Poniendo la serie en forma estandarizada:

$$\prod_{i=1}^{m}\left[1-\Phi(\frac{\beta'}{\sigma}X_i)\right]\prod_{i=1}^{n}\left[\Phi(\frac{\beta'}{\sigma}X_i)\right] \tag{30}$$

$$=\coprod_{i=1}^{n}\Phi(\frac{\beta}{\sigma}X_i)^{y_1}\left[1-\Phi\left(\frac{\beta}{\sigma}X_i\right)\right]^{1-y_1}=\coprod_{i=1}^{n}(P_i)^{y_1}[1-P_i]^{1-y_1} \tag{31}$$

De la cual obtenemos la función Log-Likelihood:

$$l=\left(\frac{\beta}{\sigma}\right)=\ln(L)$$

$$=\sum_{i=1}^{n}Y_i\ln\left[\Phi X_i\frac{\beta}{\sigma}\right]+(1-Y).\ln\left[1-\Phi X_i\frac{\beta}{\sigma}\right] \tag{32}$$

Note que la función Likelihood está acotada superiormente por cero, debido a que $0\leq\Phi(.)\leq 1$ esto implica que:

$\ln\Phi(.)\leq 0$ y $\ln(1-\Phi(.))\leq 0$

Otro importante aspecto de trabajar con funciones de distribución estandarizadas es que los parámetros β y σ siempre aparecen juntos. Por consiguiente ellos no pueden ser identificados separadamente solo importa el ratio $\frac{\beta}{\sigma}$.

Así es conveniente normalizar σ a uno, así se podrá analizar β También recordar que σ puede presentar problemas de heteroscedasticidad para tal caso hay que realizar las correcciones pertinentes.

II.6 ESPECIFICACIÓN

El modelo Probit relaciona, a través de una función no lineal, la variable Y_i con un conjunto de variables: $X_{2i}, X_{3i}, \dots, X_{ki}$, que definen la combinación lineal siguiente:

$$[1X_{2i}, X_{3i}, \dots, X_{ki}][\beta_1\beta_2, \dots, \beta_k]' = \begin{bmatrix} \beta_1 \\ \beta_2 \\ \cdot \\ \cdot \\ \cdot \\ \beta_k \end{bmatrix} [1X_{2i}, X_{3i}, \dots, X_{ki}] = \beta{'}X_i \qquad (33)$$

Así pues, la especificación del modelo Probit se efectúa a través de la ecuación de distribución de la normal:

$$Y_i = \int_{-\infty}^{\beta{'}X_i} \frac{1}{(2\pi)} exp^{-\frac{1}{2}S^2} ds + u_i \qquad (34)$$

Donde la variable $\beta{'}X_i$ es el índice que define el modelo Probit y s es una variable «muda» de integración con media cero v varianza una. De forma compacta, el modelo se puede escribir:

$$Y_i = \Phi(\beta{'}X_i) + u_i \qquad (35)$$

Si conocidos (dados) los valores de las características X; se asigna una probabilidad, por ejemplo P, para que la variable Y; valga la unidad, se tiene:

$$Prob(Y_i = 1) = P_i \qquad (36)$$

Para los mismos valores de las variables X_i, la probabilidad de que la variable Y_i valga cero es (1 - P_i), puesto que la suma de ambas probabilidades debe ser igual a la unidad. En este caso se tiene:

$$Prob(Y = 0/X_i) = (1 - P_i) \qquad (37)$$

¿Cómo se puede estimar el valor de P_i, o ¿cómo se puede cuantificar la utilidad de elegir la opción uno?

ALTERNATIVA 1

Si se calcula la esperanza de Y_ii en términos probabilísticos se obtiene:

$E\left(\frac{Y_i}{X_i}\right) = (valorY_i = 0\,)\big(Prob(Y_i = 0)\big) + (valorY_i = 1\,)\big(Prob(Y_i = 1)\big) = 0(1 - P_i) + 1.P_i = P_i$ (38)

Además, el valor esperado de la variable Y; condicionado a un valor concreto de las variables Xi se puede obtener a partir del modelo, a través de la siguiente relación:

$E(Y_i/X_i) = E\ (\Phi(\beta' X_i) + u_i) = \Phi(\beta' X_i)$ (39)

Igualando 1 con 2 se tiene:

$P_i = \Phi(\beta' X_i)$ (40)

Es decir, que la probabilidad de que ocurra el «hecho», para unos valores concretos de las variables explicativas, $Prob(Y_i = 1)$ se puede medir a través del valor asignado mediante el modelo Probit especificado.

ALTERNATIVA 2

Otro planteamiento para estimar un valor concreto del regresando del modelo (4.13) se obtiene a partir de la ecuación:

$Y_i - \ \Phi(\beta' X_i) = u_i$ (41)

INTERPRETACIÓN DEL MODELO PROBIT

Una vez estimado el modelo, un valor concreto del regresando cuantifica, a través de la probabilidad, la utilidad de elegir la opción 1, cuya expresión es:

$\hat{Y}_i = \hat{P}_i = \Phi(\hat{\beta} X_i)$ (42)

La interpretación de los parámetros del modelo Probit se puede efectuar a través de las derivadas parciales. La derivada parcial del modelo Probit respecto a la variable Xki, si es derivable, es igual a:

$$\frac{\partial \hat{P}_i}{\partial X_{ik}} = \frac{\partial \Phi(\hat{\beta} X_i)}{\partial X_{ik}} \beta_k = \phi(\hat{\beta} X_i)\beta_k \qquad (43)$$

Una cuestión importante a tener en cuenta en los modelos Probit es la interpretación de los distintos elementos que intervienen en su especificación.

Así pues, a través de ambas alternativas, se obtiene que el modelo estimado cuantifica la probabilidad de elegir la opción 1.

II.5 MODELO LOGIT

Especificación

El modelo Logit relaciona la variable Y_i; con las variables $X_1, X_2, \dots, X_{k-1}, X_k$; a través de la siguiente ecuación:

$$Y_i = \frac{1}{1+e^{-(\beta_1+\beta_2 X_{2i}+\beta_3 X_{3i}+\cdots+\beta_k X_{ki})}} + u_i \qquad (44)$$

O bien de forma compacta:

$$Y_i = \frac{1}{1+e^{\beta' X_i}} + u_i = \frac{e^{\beta' X_i}}{1+e^{\beta' X_i}} + u_i \qquad (45)$$

La forma funcional, el modelo se puede escribir como:

$$Y_i = \wedge(\beta' X_i) + u_i \qquad (46)$$

Dónde:

$\wedge(.)$: Hace referencia a la función de distribución Logística.

u_i: es una variable aleatoria que se distribuye normal. Las variables o características X_i; son fijas en el muestreo.

La variable dependiente Y puede tomar los valores cero o la unidad.

La interpretación del modelo Logit se puede efectuar a partir del siguiente hecho: conocidos (dados) los valores de las características X_i; se les asigna una probabilidad por ejemplo P_i, de que la variable Y_i valga la unidad. Así se tiene:

II.6 VALIDACIÓN Y CONTRASTES DE HIPÓTESIS

En el campo de los modelos de elección discreta, se pueden construir los contrastes habituales, sobre un coeficiente o un conjunto de coeficientes. A partir de estimaciones consistentes y asintóticamente[4] eficientes de la matriz de varianzas covarianzas del modelo.

Podemos diferenciar dos situaciones distintas: el contraste de una hipótesis sobre un parámetro individual y la significatividad estadística del modelo en su conjunto. a través de los tests fundamentados en la función de verosimilitud y en la bondad del ajuste.

II.7 CONTRASTE INDIVIDUAL DE UN COEFICIENTE

Dadas las propiedades estadísticas de los estimadores máximo-verosímiles y su distribución asintótica según una normal, se puede plantear el siguiente contraste de hipótesis sobre un coeficiente de regresión aislado.

En efecto, dado un grado de significación de α se acepta la hipótesis nula sobre β_k con una probabilidad de $(1-\alpha)$ si se cumple la desigualdad probabilística siguiente:

$$Prob\left(-N_{\frac{\alpha}{2}} < \frac{\hat{\beta}_K - \beta_K}{S_{\hat{\beta}_K}} < N_{\frac{\alpha}{2}}\right) = 1-\alpha \qquad (47)$$

4. Es importante recalcar que la función de verosimilitud, bajo ciertas condiciones de regularidad, es uno de los principales candidatos a poseer las propiedades asintóticas.

II.8 PRUEBAS ESTADÍSTICAS BASADAS EN LA FUNCIÓN DE VEROSIMILITUD Y EN EL LOGARITMO DE LA FUNCIÓN DE VEROSIMILITUD (LOG LIKELIHOOD)

A partir de la función de verosimilitud estimada de los residuos o bien de su logaritmo[5]:

$$l(\beta,\sigma^2) = ln[L(\beta,\sigma^2)] = -\frac{-n}{2}ln2\pi - \frac{n}{2}ln\sigma^2 - \frac{1}{2\sigma^2}\sum_{1}^{n}\left(y_i - F(\beta x_i)\right)^2 \qquad (48)$$

Se pueden construir distintos contrastes de hipótesis. El criterio general para la elección entre distintos modelos es el siguiente: se prefiere aquel modelo que presente un valor de la función de verosimilitud mayor.

Además, a partir de la función de verosimilitud se pueden construir distintos contrastes de hipótesis. Entre otros, cabe destacar el de la Razón de Verosimilitud, que sirve para realizar las pruebas hipótesis entre dos modelos que presentan la misma variable endógena. El primer modelo se estima bajo la hipótesis nula (modelo con restricciones), cuya función de verosimilitud se denota por L_R; mientras que el segundo modelo se estima bajo la hipótesis alternativa (modelo sin restricciones). Cuya función de verosimilitud se denota por L_{SR}.A partir de estas dos funciones de verosimilitud *(LcR* Y *LsR)* se construye la **<u>Razón de Verosimilitud</u>**[6] entre ambas funciones, que se define como:

$$LR = -2\ln(\lambda) = -2ln\left(\frac{L_{CR}}{L_{SR}}\right) = 2(lnL_{CR} - lnL_{SR}) = -2(l_{CR} - l_{SR}) \qquad (49)$$

El estadístico $LR = -2\ln(\lambda)$ se distribuye según una λ^2; con un número de grados de libertad igual al número de restricciones.

[5]La expresión de función de verosimilitud que se expone a continuación tan sólo se cumple cuando el tamaño de la muestra tiende a infinito (suficientemente grande). En este tipo de modelos la función de verosimilitud no se puede simplificar ya que el estimador de la varianza del modelo y los estimadores de los coeficientes de regresión no son independientes

[6] En el Eviews es LR statistic (K df) donde K df son los grados de libertad igual al número de variables explicativas en el modelo. Estos se ven cuando estimo el modelo ya sea Logia o Probit.

Akaike (1973)[7] propone una corrección a los estadísticos anteriores por el número de parámetros del modelo (coeficientes de regresión). La expresión del estadístico de Akaike (AIC) es:

$$AIC = \frac{2K}{n} - \frac{2l}{n} \tag{50}$$

y sirve para comparar la bondad del ajuste entre dos modelos. Según este criterio es preferible aquel modelo que presente un valor del AIC menor, es decir siempre un modelo es más parsimonioso que en el sentido de que tiene mayor poder explicativo y esto se ve con el AIC

Una alternativa al criterio propuesto por Akaike es el propuesto por **Schwarz** (1978) que se define como:

$$SC = Schawrz = \frac{K^* \ln n}{n} - \frac{2l}{n} \tag{51}$$

Dicho estadístico, al igual que el AIC de .akaike, sirve para comparar la bondad del ajuste entre dos modelos (no es necesario que presenten la misma variable endógena).

En este caso se tiene en cuenta explícitamente el tamaño de la muestra. Según este criterio es preferible aquel modelo que presente un valor del estadístico de Schwarz menor.

Otra alternativa al criterio propuesto por Akaike es el propuesto por **Hannan--Quinn (1979)** que se define como:

$$H - Q = Hannan - Quinn = \frac{2^* K^* \ln(\ln n)}{n} - \frac{2l}{n} \tag{52}$$

Dicho estadístico, al igual que el AIC de Akaike, sirve para comparar la bondad del ajuste entre dos modelos. En este caso se tiene en cuenta explícitamente el tamaño de la muestra. Según este criterio es preferible aquel modelo que presente un valor del estadístico de Hannan-Quinn menor.

[7] En el Eviews es Akaike info criterion

II.9 MEDIDAS DE BONDAD DEL AJUSTE

Dado que las pruebas tradicionales de bondad del ajuste, tales como el R2,no son válidas en los modelos en los que la variable endógena toma exclusivamente los valores uno o cero, se van a proponer unas medidas alternativas que midan la bondad del ajuste del modelo a los datos.

II.9.1 R2 PROPUESTO POR MCFADDEN (1974) (MC FADDEN R-SQUARED)

Que se define como:

$$R^2 McFadden = \rho^2 = 1 - \left(\frac{\ln L_{SR}}{\ln L_{CR}}\right) = 1 - \left(\frac{L_{SR}}{L_{CR}}\right) \quad (53)$$

Donde, InLcR es el logaritmo de la función de verosimilitud del modelo restringido (con restricciones) que se obtiene bajo la hipótesis nula:

$$H_0: \beta_2 = \beta_3 = \cdots = \beta_K = \cdots = \beta_K = 0 \quad (54)$$

El estadístico propuesto no tiene una interpretación tan directa como el R^2en el modelo de regresión lineal. En concreto, toma el valor uno cuando la predicción es perfecta es decir, la probabilidad estimada de que se produzca el evento es uno cuando éste efectivamente se produzca y cero cuando no se produzca el evento. Por su parte, el estadístico tomará el valor cero cuando ambas funciones de verosimilitud sean iguales. El problema es que, fuera de estos dos valores extremos, el estadístico no tiene un significado tan intuitivo como el coeficiente de determinación. Algunos autores han señalado que, en realidad, lo que este estadístico mide es el porcentaje de «incertidumbre» en los datos explicada por el modelo. Como Regla práctica, es de esperar que un buen modelo tenga un ρ^2Entre 0.2 y 0.4 (Colin Cameron & Trivedi, 2005).

II.9.2 EL ESTADÍSTICO *LR* (LR-STATISTIC)

Si denotamos con L_{SR} el valor de la función de verosimilitud respecto a todos los parámetros y L_{cR} es la función de verosimilitud que se obtiene bajo la hipótesis nula (o modelo con restricciones):

$$H_0: \beta_2 = \beta_3 = \cdots = \beta_k = 0$$

se puede definir el estadístico:

$$LR = -2\ln(\lambda) = -2\left(\frac{L_{CR}}{L_{SR}}\right) = -2(\ln L_{CR} - \ln L_{SR}) \qquad (55)$$

El estadístico $LR = -2\ln(\lambda)$ se distribuye según una λ^2; con un número de grados de libertad igual al número de restricciones (K-1).

II.9.3 PROPORCIÓN DE PREDICCIONES CORRECTAS (EXPECTACIÓN-PREDICCIÓN).

Se puede representar, también, una medida de la bondad del ajuste al 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 Yi, de forma que:

Tabla N 2: Clasificación de predicciones

		Valor real del regresando Y_i	
		$Y_i = 1$ Frecuencia = $I_{.1}$	$Y_i = 0$ Frecuencia = $I_{.2}$
Predicción de $\hat{Y}_i$	$\hat{Y}_i > C$ Frecuencia = $I_{1.}$	Frecuencia de aciertos = I_{11} (Predicción Correcta)	Frecuencia de errores = I_{12} (Predicción errónea)
	$\hat{Y}_i < C$ Frecuencia = I_2	Frecuencia de errores =I_{12} (Predicción errónea)	Frecuencia de aciertos =I_{22} (Predicción correcta)

Analizando la tabla se comprueba que las frecuencias dispuestas en la diagonal principal corresponden a las predicciones correctas, mientras que las frecuencias de la diagonal secundaria son las que no se adecuan al modelo.

Se define el estadístico pseudo coeficiente de determinación de la predicción.

$$Pseudo\ R^2 de\ predicción = \frac{Predicciones\ correctas}{Frecuencia\ total} = \frac{I_{11}+I_{22}}{I_{11}+I_{22}+I_{33}+I_{44}} \qquad (56)$$

En el caso de los modelos MLP, Logit, Probit y Valor Extremo se asigna generalmente el valor de predicción igual a uno cuando $\hat{Y}_i > 0.5$ e igual a cero cuando $\hat{Y}_i < 0.5$. (Bernardí Cabrer Borrás & Amparo Sancho Pérez, Guada, 2001)

III. PROCEDIMIENTO PARA ESTIMAR UN MODELO (Beltran Barco, 2001)

Para estimar correctamente un modelo discreto se sugiere seguir los pasos que se explican a continuación (Beltran Barco, 2001):

1. Analizar la matriz de correlaciones entre la dependiente y el conjunto de posibles explicativas. A partir de ella se busca rescatar dos cosas:

- Establecer el grado de relación de las explicativas y la dependiente como su signo esperado.
- Establecer la posible correlación entre explicativas potenciales, regla práctica, si dos variables tienen una correlación mayor a 75% debe elegir entre ellas a aquella que ajuste mejor; no incluir a ambas: el modelo.

2. Analizar tablas cruzadas entre la dependiente y las explicativas que mostraron en 1 ser las más relacionadas con la primera. A través de este análisis pretende confirmar la dirección y magnitud de la relación.

3. Estimar la ecuación con todas las explicativas que aparecieron como relevantes en 1 y 2. Una vez corrido el modelo dejar aquellas explicativas que tengan el signo esperado y cuya probabilidad asociada a t no sea mayor a 10% o 15%. Nótese que en el caso de los modelos discretos él t reduce su validez por lo que se relaja la necesidad de ser muy estrictos respecto de las conclusiones que arroja este test. Esta regla práctica se da porque el t estadístico pierde relevancia debido a problemas de multicolinealidad.

Uno de los resultados claves del modelo estimado es la predicción de probabilidad asociada a la variable dependiente, la misma que puede ser determinada para la media muestral o para individuos con características específicas dentro de la muestra.

III.1 EFECTO IMPACTO, MARGINAL O EFECTO ESCALA (Greene, 1997)

Determinar los efectos impactos de las variables explicativas del modelo. En caso de una variable explicativa continua k éste sería igual a:

$$EI = \frac{\partial Proba(Y_i=1)}{\partial X_K} = f(\beta' X_i)\hat{\beta}_{ki} \quad (57)$$

Acá la probabilidad no permanece constante, sino que depende de la función de densidad $f(\hat{\beta}X_i)$ que viene a ser la derivada de la función de distribución acumulada Logit o Probit ya explicados[8]

El mismo que puede ser evaluado en la media muestral o para un conjuntoespecífico de valores de las explicativas.

- En el caso de una variable explicativa discreta tendría que calcularse la diferencia de la probabilidad cuando dicha variable toma un valor u otro. Pe ejemplo, si estamos analizando la decisión de trabajar y la variable explicativa de interés es el sexo de la persona, definido como 1 si es hombre y 0 si e mujer, el efecto impacto de la misma sobre la probabilidad de trabajar sería:

$$EI_{x_2} = F(\beta_1 + \beta_2(1) + \beta_3 X_{3i} + \cdots + \beta_k X_{ki}) - F(\beta_1 + \beta_2(0) + \beta_3 X_{3i} + \cdots + \beta_k X_{ki}) \quad (58)$$

En este caso también podría calcularse el efecto para la media muestral o paracaracterísticas determinadas del individuo.

Note que cualquiera sea el tipo de variable explicativa, el efecto impacto arroja el cambio de la probabilidad, en puntos porcentuales, frente a la variación en una unidad de la explicativa, razón por la cual su utilidades mayor cuando analizamos explicativas discretas. Se determina la elasticidad de la probabilidad respecto de cambios en las variables explicativas. La misma puede definirse como para las variables explicativa k:

$$n_k EI_{X_K} * \left(\frac{\bar{X}_k}{F(\beta' X_i)}\right) \quad (59)$$

[8] Tener en cuenta que en la formulación de estos modelos la especificación es con la funcion de distribución acumulada y no la función de densidad.

La elasticidad indica el cambio porcentual en la probabilidad ante una variación de 1 % en la variable explicativa de interés, razón por la cual resulta más conveniente estimarla para explicativas continúas. No obstante, dado que carece de unidades, la elasticidad puede servir también para *rankear* todas las variables explicativas de acuerdo con su importancia relativa en el modelo.

Bibliografía

Colin Cameron , A., & Trivedi, P. (2005). *Microeconometrics: Methods and Applications.* (C. U. Press, Ed.) New York.

Beltran Barco, A. (2001). *Econometria de Corte Transversal.* Notas de Clase.

Bernardí Cabrer Borrás, & Amparo Sancho Pérez, Guada. (2001). *Microeconometría y Decisión.* Ediciones Pirámide, .

Greene, W. (1997). *Análisis Econometrico* (Tercera ed.). Prentice Hall.