ESPECIFICACION DEL MODELO

Cuatro especificaciones alternativas son presentadas para la estimación del modelo de uso de anticonceptivos anteriormente descrito. Estas son: Mínimos Cuadrados Ordinarios (MCO), Mínimos Cuadrados Ponderados (MCP), Logit y de Probabilidad Conicionada (COPRO).

La aplicación de un modelo de regresión lineal cuando la variable dependiente es dicotómica es muy compleja. Este tipo de modelo supone que los individuos enfrentan la elección entre dos alternativas, y que la elección que haga depende de las características y atributos del individuo. Una finalidad de los modelos de elección cualitativa es determinar la probabilidad de que un individuo con un conjunto de atributos hará una elección por otra. Más precisamente, deseamos encontrar una relación entre un conjunto de atributos que distinguen a un individuo y la probabilidad de que éste hará una elección determinada (por ejemplo, elegir entre usar o no anticonceptivos).

A.- MINIMOS CUADRADOS ORDINARIOS.

Nos interesamos en los factores que determinan el que una mujer use o no anticonceptivos. Asumiendo que esta decisión depende del nivel de educación de la mujer (medido en años) y efectos aleatorios, la forma de este modelo es:

$$(1) \qquad Y_i = a + bZ_i + e_i \qquad i = 1, 2, \ldots, N$$

donde Y_i es una variable aleatoria que toma el valor de 1 si la iésima mujer usa anticonceptivos, y el valor de 0 si no lo hace. Z_i es el nivel de educación de ésta iésima mujer y e_i es una variable independiente distribuida aleatoriamente con media igual a 0. Suponemos que Z_i es fija, o que su aleatoriedad está distribuida independientemente de E_i y N es el número de observaciones. Los parámetros a estimar son a y b.

La interpretación de la ecuación (1) como un modelo de probabilidad lineal se obtiene cuando se toma el valor esperado de la variable dependiente Y_i.

(2) $E\,(Y_i) = a + bZ_i$

Aun cuando Y_i puede tomar sólo los valores de 0 y 1, podemos describir la distribución probabilística de Y al establecer:

$$P_i = \text{Prob}\,(Y_1 = 1) \qquad y \qquad 1 - P_i = \text{Prob}\,(Y_i = 0)$$

por lo tanto

(3) $E\,(Y_i) = 1\,(P_i) + (1 - P_i) = P_i$

De este modo, la ecuación 1 puede ser intepretada como describiendo la probabilidad de que una mujer utilice anticonceptivos, dado la información sobre su nivel de educación. Notamos que la pendiente de la regresión (b) mide el efecto en la probabilidad de utilizar anticonceptivos de cambios de un año de educación en el individuo.

Puesto que la probabilidad debe estar entre 0 y 1, se tiene tiene la siguiente restricción

(4) $0 \leq E\,(Y_i) = P_i \leq 1$

No obstante que esto es cierto a priori, no existe garantía de que $\hat{Y}_i$, la estimación de P_i, satisfaga necesariamente esta restricción, lo cual es uno de los problemas asociados con los modelos de probabilidad lineal (MPL). Cuando intentamos utilizar MPL para predicciones, se presenta una debilidad seria en el modelo. Aún si P_i se restringe al intervalo unitario, predicciones fuera de este intervalo pueden ser formuladas para valores de las variables explicativas fuera del rango en la muestra, aún si los coeficientes (las estimaciones de a y b) son derivados minimizando la suma de residuales cuadrados sujeta a la condición de que estas predicciones caigan dentro del intervalo unitario. La determinación de la estimación de los parámetros de mínimos cuadrados sujeta a la restricción en forma de desigualdad.

$$0 \leq \hat{Y}_i \leq 1$$

es un problema de estimación no-lineal, la cual en este caso puede resolverse utilizando una técnica de programación matemática. La introducción de la restricción en forma de desigualdad conduce a estimar parámetros con menor varianza, pero no existe garantía de que la esti mación esté insesgada. El problema de predicciones fuera del intervalo unitario con el modelo ordinario de mímimos cuadrados sugiere un problema adicional asociado con la especificación del modelo. El problema surge porque demasiadas observaciones en una muestra determinada pueden ser sorteadas por atributos cuyos valores están asociados con valores extremos de probabilidad de elección (0 ó 1).

Notamos que a pesar de que en MCO no requiere que las perturbaciones (e's) estén normalmente distribuidas, suponemos que sí lo están para los propósitos de inferencia estadística. El supuesto de normalidad para las perturbaciones ya no puede mantenerse en los modelos de probabilidad lineal a causa de que, como Y_i, e_i es binomial, es decir e_i toma dos valores solamente. Para demostrarlo, escribiremos la educación (1) como:

$$e_i = Y_i - a - bZ_i$$

$$\text{cuando} \quad Y_i = 1 \qquad e_i = 1 - a - bZ_i$$

$$Y_i = 0 \qquad e_i = - a - bZ_i$$

y obviamente e_i no puede suponerse estar normalmente distribuido.

Dada la distribución de e, el supuesto usual de que la varianza $\text{Var}(e_i) = \sigma^2$ ya no es aceptable. El carácter Bernoulli de Y_i implica una varianza para e_i de

$$(5) \quad \text{Var}\,(e_i) = (1 - a - b_i)^2\, P_i + (-a - bZ_i)^2\,(1-P_i) = P_i\,(1-P_i)$$

Puesto que la varianza de e_i depende de P_i, los e's son heteroscedásticos y el uso de MCO producirá estimaciones ineficientes y predicciones imprecisas. Las observaciones para los cuales P_i es cercano a 0 ó 1 tendrán relativamente baja varianza, mientras que las observaciones cercanas a .5 tendrán mayores varianzas. La presencia de heteroscedasticidad resulta en una pérdida de eficiencia, pero en sí misma no produce sesgos o inconsistencia en la estimación de los parámetros.

B. MODELO DE MINIMOS CUADRADOS PONDERADOS.

El punto débil de MCP es que el supuesto clásico de homoscedasticidad es insostenible. Sabemos que en la presencia de heteroscedasticidad los estimadores MCO, aunque insesgados, no son eficientes. Es decir, que no tienen una varianza mínima. En el modelo de MCO, la varianza de los errores está dado por:

$$\text{Var}(e_i) = P_i\,(1-P_i)$$

y es el modelo de mínimos cuadrados generalizados el que parece ser el más apropiado. Aunque la varianza de e_i depende del valor esperado de Y_i, una forma de resolver la heteroscedasticidad es transformar la información dividiendo ambos lados de la equación 1 por h_i, donde

$$h_i = \left[P_i\,(1 - P_i) \right]^{1/2}$$

y (6) $Y_i/h_i = (a/h_i) + (bZ_i/h_i) + e_i/h_i$

el término de perturbación (e/h) será ahora homoscedástico. La dificultad obvia es que P_i, al igual que h_i, es desconocida. Para estimar h_i, utilizamos el siguiente procedimiento de dos etapas:

a) Del modelo de MCO obtenemos los valores estimados de la probabilidad $\hat{P}_i = \hat{Y}_i$

y después usamos

$$\left[\hat{P}_i\,(1 - \hat{P}_i)\right]^{1/2}$$

para obtener $\hat{h}_i$, la estimación de h_i

b) Después utilizamos $\hat{h}_i$ para transformar la información como en la ecuación 6, y correr la regresión de MCO con ella.

La dificultad con MCP es que no existe seguridad en que $\hat{Y}_i$ predicha caiga en el intervalo unitario. Si alguna $\hat{Y}_i$ cae fuera, las observaciones deben ser igualadas ya sea a 0.01 ó 0.99 o descartadas del modelo. Debido a estos problemas, enfoques alternativos al modelo de respuesta cualitativamente dicotómica son desarrollados en los cuales P_i es restringido al intervalo unitario. Uno de tales enfoques es el modelo Logit.

C. EL MODELO LOGIT

El problema relamente serio en la estimación del modelo de de probabilidad lineal es que la estimación de las probabilidades puede no caer dentro del intervalo unitario. El modelo Logit es una técnica de estimación especial que transforma el modelo original de tal modo que la estimación de las probabilidades caiga dentro del intervalo unitario. Dado que nuestra preocupación principal en la estimación de modelos de elección binaria, en la de interpretar la variable dependiente como la probabilidad de elegir, dada la información sobre los atributos del individuo; parece razonable utilizar algunas nociones de probabilidad como base para la transformación.

Suponga que el evento "E" es una acción tomada por un individuo para maximizar su utilidad. Asimismo que existe un índice teórico I_i, determinado por una variable explicativa Z_i como en el MPL. Se supone que el índice I_i es una variable continua, aleatoria y normalmente distribuida. Así.

(7) $I_i = a + bZ_i$

y entre mayor sea I_i, mayor la probabilidad de ocurrencia del evento "E". Puesto que la probabilidad debe caer entre 0 y 1, la relación monotónica entre I_i y la probabilidad condicional de ocurrencia del evento E, dado I_i, $[P_r(E/I_i)]$ debe suponerse tomar la forma general de una función de densidad acumulativa.