6.

Variables dependientes discretas y limitadas

Existen muchos fenómenos en la actividad económica que responden a elecciones discretas como la decisión de trabajar, la decisión de comprar una bien, la decisión de votar por un candidato, etc.

Para Amemiya (1981), existen dos factores que explican el interés en los modelos de respuesta cualitativa: Primero, los economistas trabajan con modelos que involucran más de una variable discreta, más de dos respuestas, y por supuesto, más de una variable independiente. Segundo, el creciente número de encuestas que se realizan y la posibilidad de trabajar los datos que éstas producen.

A continuación, se desarrollarán algunos modelos estadísticos cuyo objetivo consiste en facilitar la contrastación empírica de la teoría del consumidor. Estos modelos son el de probabilidad lineal, el Logit, el Probit y el Tobit en sus diferentes versiones. Luego se presentará una versión del modelo de autoselección de Heckman y, finalmente, el modelo de variables latentes.

6.1. Especificación del modelo

Suponga que usted desea considerar la ocurrencia de un evento como "comprar un carro"; para describir este evento, definiremos la variable aleatoria dicotómica Y, la cual tomará el valor de 1 si el evento ocurre y 0 si no ocurre. De igual forma, deberemos asumir que la probabilidad del evento depende sobre un vector de variables independientes x* y un vector de parámetros desconocidos θ. El subíndice i denota el i- ésimo individuo. De esta forma, un modelo general dicotómico univariado, se puede expresar como:

(6.1) $p_i = p(Y_i = 1) = G(x^*_i, \theta)$; $i = 1,2,...., n$.

Los Y_i son distribuidos independientemente. Por otro lado, dado que Y_i es la probabilidad de comprar un auto, x^*_i estaría representando aquellas variables que explican Y como el ingreso, el sexo, la edad, el estatus, la educación del individuo i, así como los precios del auto. Ya que (6.1) es muy general, el investigador deberá escoger alguna función $H(x^*_i, \theta)$ sobre un vector de parámetros θ:

(6.2) $p(Y_i = 1) = F[H(x^*_i, \theta)]$

6.2. Formas comunes de las funciones de probabilidad

Considere el siguiente modelo de consumo de automóviles: el consumidor responderá Y=1 si compra el automóvil y Y=0 si no lo compra. Dado que nosotros consideraremos que los factores x^*_i, explican la decisión que toma el consumidor:

(6.3) $$\text{Prob}(Y = 1) = F(\beta' x)$$
$$\text{Prob}(Y = 0) = 1 - F(\beta' x)$$

El conjunto de parámetros β refleja el impacto de los cambios en x sobre la probabilidad. Una primera forma de representar este evento es considerar la siguiente regresión lineal:

(6.4) $F(x, \beta) = \beta' x_i$

Dado que $E[Y] = F(x, \beta)$, el modelo de regresión será:

(6.4.1) $$Y_i = E[Y_i] + (Y_i - E[Y_i])$$
$$= \beta' x_i + \varepsilon_i, \qquad \text{con } \varepsilon_i = Y_i - E[Y_i]$$

que se conoce como el modelo de Probabilidad Lineal. El modelo de Probabilidad Lineal tiene como "debilidad" que $\beta' x_i$ no está restringido a pertenecer entre 0 y 1 como debería ser en términos de la probabilidad. Esto significa que no se satisface en el modelo de probabilidad lineal la condición $0 < \beta' x_i < 1$.
Amemiya (1981) nos indica que este defecto podrá corregirse si definimos F=1 si $F(\beta' x_i) > 1$ y F=0 si $F(\beta' x_i) < 1$, lo cual produce puntos de truncamiento no reales. Este procedimiento fue utilizado en los primeros años de investigación por su simplicidad computacional.

Igualmente, el modelo de probabilidad lineal presenta problemas de heterocedasticidad, ya que ε deberá ser igual a uno o cero. Entonces ε deberá ser igual tanto a $-\beta' x$ como a $1 - \beta' x$ con probabilidades 1-F y F por lo cual $V[\varepsilon|x] = V(Y_i) = \beta' x_i (1 - \beta' x_i)$.

Como menciona Green (1999), restringir $B' x_i$ al intervalo (0, 1) produciría probabilidades y varianzas negativas. Dadas las desventajas del modelo de Probabilidad Lineal, su interés ha ido decayendo, lo cual ha originado que modelos como el Logit o Probit se usen más frecuentemente. Veamos en qué consisten estos modelos.

Suponga que para un vector de regresores se cumple:

(6.5) $\lim_{\beta' x \to +\infty} \text{Prob } (Y_i = 1) = 1$

$\lim_{\beta' x \to +\infty} \text{Prob } (Y_i = 0) = 0$

Asumiendo una distribución normal estándar Φ, obtenemos el modelo Probit:

(6.6) $\text{Prob } (Y = 1) = \int_{-\infty}^{\beta' x} \phi(t)\,dt$

$$= \Phi\,(\beta' x) = \int_{-\infty}^{\beta' x} \frac{1}{\sqrt{2\pi}} e^{-t^2/2} dt$$

De igual forma se obtiene la distribución logística:

(6.7) $\text{Prob } (Y = 1) = \dfrac{e^{\beta' x}}{1 + e^{\beta' x}} = \Lambda(\beta' x) = L(\beta' x)$

Donde $\Lambda(\beta' x)$ es la función de distribución acumulativa logística. Las funciones de distribución Logit y Probit están acotadas entre 0 y 1. La distribución logística tiene varianza igual a $\pi^2/3$. Considere ahora:

(6.8) $L_\lambda(\beta' x) = \dfrac{e^{\lambda\beta' x}}{1 + e^{\lambda\beta' x}}$

Para un valor apropiado de λ se puede aproximar (6.8) a una normal estándar. Esto significa que cuando $\lambda = \pi/\sqrt{3}$ se tendría una distribución con media cero y varianza unitaria. Amemiya (1981) muestra los diferentes valores para los cuales (6.6) y (6.7) difieren a partir de (0,5). Debido a la similitud entre (6.6) y (6.7), es difícil distinguir estadísticamente entre estas funciones, a menos que uno tenga una gran cantidad de datos [Chambers y Cox (1967)].

De esta forma, en modelos univariados dicotómicos no es posible distinguir cuándo usar Logit o Probit, a menos que exista una concentración en la cola dadas las características del problema estudiado. Debido a las anotaciones anteriores, la distribución logística tiende a dar mayores probabilidades a Y = 0 que la distribución normal cuando $\beta' x$ es muy pequeño y probabilidades menores que la distribución normal a Y = 0 cuando $\beta' x$ es muy grande. De esta forma, el tamaño de la muestra podría influenciar los resultados aunque asintóticamente ambas distribuciones no difieren significativamente[12].

Amemiya (1981) muestra las equivalencias en los parámetros entre ambos modelos. Si $\hat{\beta}$ es el parámetro estimado en ambos modelos:

$$(6.9)\quad \begin{cases} \hat{\beta}_{(6.7)} \approx 0.4\hat{\beta}_{(6.6)} & \text{excepto para el término constante} \\ \hat{\beta}_{(6.7)} \approx 0.4\hat{\beta}_{(6.6)} + 0.5 & \text{para el término constante} \end{cases}$$

$$\begin{cases} \hat{\beta}_{(6.7)} \approx 0.25\hat{\beta}_{(6.4)} & \text{excepto para el término constante} \\ \hat{\beta}_{(6.7)} \approx 0.25\hat{\beta}_{(6.4)} + 0.5 & \text{para el término constante} \end{cases}$$

Cuando se desea comparar modelos con diferentes funciones de probabilidad, es mejor comparar directamente las probabilidades que los coeficientes estimados. De esta forma, observaremos las derivadas de las probabilidades con respecto a una variable independiente en particular. Sea x_{ik} el k elemento del vector x_i y sea β_{ik} el k elemento del vector de parámetros β_i, entonces, las derivadas para (6.4), (6.6) y (6.7) serán:

$$(6.10)\quad \begin{cases} \dfrac{\partial}{\partial x_{ik}} \beta' x_i = \beta_k \\ \dfrac{\partial}{\partial x_{ik}} \Phi(\beta' x_i) = \phi(\beta' x_i)\beta_k \\ \dfrac{\partial}{\partial x_{ik}} L(\beta' x_i) = \dfrac{e^{\beta' x_i}}{(1+e^{\beta' x_i})^2}\beta_k \end{cases}$$

Donde ϕ es la función de densidad normal estándar. En el limite, como muestran Amemiya (1981) y Green (1999) la diferencia no es sustancial.

12. Green(1999) menciona que los modelos generan predicciones diferentes si una muestra contiene pocas respuestas afirmativas (Y=1) o pocas respuestas negativas (Y=0) y sí existe una gran variación en la variable independiente de importancia.

6.3. Estimación

A excepción del modelo de Probabilidad Lineal, los modelos Probit y Logit se estiman por máxima verosimilitud donde cada observación es extraída de una distribución de Bernoulli. El modelo con una probabilidad de suceso f(β'x) y observaciones independientes lleva a una probabilidad conjunta o a una función de verosimilitud de la forma:

(6.11) $$Prob(Y_1 = y_1, Y_2 = y_2, Y_3 = y_3, \ldots, Y_n = y_n) = \prod_{y_i=0}[1 - F(\beta' x_i)] \prod_{y_i=1} F(\beta' x_i)$$

Reacomodando términos, (6.11) puede escribirse como:

(6.12) $$L = \prod_i [F(\beta' x_i)]^{y_i} [1 - F(\beta' x_i)]^{1-y_i}$$

La cual es la probabilidad para una muestra de n observaciones. Tomando logaritmos:

(6.13) $$LnL = \sum_i y_i \quad LnF(\beta' x_i) + (1 - y_i) Ln[1 - F(\beta' x_i)]$$

Al derivar con respecto al vector de parámetros se obtiene:

(6.14) $$\frac{\partial LnL}{\partial \beta} = \sum_i \left[\frac{y_i F(\beta' x_i)}{F(\beta' x_i)} + (1 - y_i) - \frac{F(\beta' x_i)}{1 - F(\beta' x_i)} \right] x_i = 0$$

La elección de una F(•) en particular lleva a un modelo empírico. Entre las formas disponibles para calcular (6.13) se encuentra el método de algoritmos de Newton, Newton-Rampson, Máxima verosimilitud. Hoy día, calcular un Logit o un Probit es bastante sencillo, pues estos métodos se encuentran en paquetes estadísticos como el RATS, SAS, SPSS, GAUSS, LIMDEP, E-Views, Easy Reg (de Libre Uso) y el STATA debiendo solamente especificarse qué algoritmo se desea.

6.4. Algunos modelos aplicados

En economía, la tradición de usar modelos Logit y Probit es extensa, aquí menciono tan sólo algunos modelos, quedando en deuda con el resto.

6.4.1. Domencich y McFadden

Considérese a un individuo que toma la decisión entre conducir o usar un método alternativo para ir al trabajo (autobús, metro, etc.). La utilidad que se asocia a cada forma de transporte, está en función de las características Z (principalmente el tiempo y el costo en que se incurre en cada elección) y las características individuales socioeconómicas w, más un término aleatorio de error ε. Definiendo μ_{i1} y μ_{i0} como la utilidad indirecta[13] asociada a la i-ésima persona cuando conduce o toma otro transporte y cuando esta función es lineal:

$$(6.15)\quad \begin{aligned} u_{i0} &= \alpha_0 + \beta' Z_{i0} + \gamma'_0 w_i + \varepsilon_{i0} \\ u_{i1} &= \alpha_0 + \beta' Z_{i1} + \gamma'_0 w_i + \varepsilon_{i1} \end{aligned}$$

De esta forma, la i-ésima persona conduce el automóvil si $u_{i1} > u_{i0}$ y viaja en otro tipo de transporte si $u_{i1} < u_{i0}$. El individuo estaría indeciso si $u_{i1} = u_{i0}$ pero esto sucederá con una probabilidad de cero si ε_{i1} y ε_{i0} son variables continuas aleatorias. Definiendo $y_i = 1$ si la i-ésima persona conduce un automóvil y cero de otra forma, se tendrán las probabilidades:

$$\text{Prob}\left(y_i = 1\right) = \text{Prob}\left(u_{i1} > u_{i0}\right) = \text{Prob}\left[\varepsilon_{i0} - \varepsilon_{i1} < \alpha_1 - \alpha_0 + \beta'\left(Z_{i1} - Z_{i0}\right) + \left(\gamma_1 - \gamma_0\right) w_i\right]$$

$$(6.16)\qquad = F\left[\left(\alpha_1 - \alpha_0\right) + \beta'\left(Z_{i1} - Z_{i0}\right) + \left(\gamma_1 - \gamma_0\right) w_i\right]$$

Donde F es la función de $\varepsilon i_0 - \varepsilon_{i1}$. Si se asume una distribución normal para $\varepsilon_{i1} - \varepsilon_{i0}$ se obtendrá un Probit .

Domencich y McFadden sustentan que el término aleatorio de error está determinado por el tipo de transporte, que a su vez vendrá determinado por una serie de características socioeconómicas que no son observadas por el investigador. Convenientemente debe asumirse que las utilidades de individuos diferentes son distribuidas independientemente, es decir, que la correlación entre ε_{i1} y ε_{i0} es cero.

Por otro lado, la idea de que las constantes α_o y α_1 en (6.15) sean diferentes no es muy consistente con la idea "abstracta de una forma de transporte" ya que esto significa diferencias en el efecto de transportar en una forma especifica más que diferencias en los atributos que son los que explican uno u otro. Por esta razón, los autores deberán asumir que $\alpha_o = \alpha_1$ y excluir el término constante. Sin embargo, las constantes se incluyeron con el fin de ajustar mejor los datos.

El modelo utilizado fue un Logit para una encuesta con 115 individuos en Pisttburg (U.S.A.) en 1967; los resultados fueron:

$$(6.17)\quad \hat{y} = -\underset{(0.51)}{3.82} + \underset{(0.05)}{0.158}Tw - \underset{(0.25)}{0.382}(Aiv - Tss) - \underset{(0.58)}{2.56}(Ac - F) + \underset{(1.07)}{4.94}A\ w - \underset{(1.37)}{2.91}R - \underset{(1.17)}{2.36}Z$$

Desviaciones estándar entre paréntesis, y Tw es el tiempo de tránsito a pie (en minutos), Aiv es el tiempo en automóvil, Tss es el tiempo de transporte entre estacio-

13. En el capítulo siguiente, se discutirá con mayor detalle los modelos de utilidad aleatoria.

nes, Ac es el costo de parqueadero más costos de operar el vehículo (en dólares), F es el precio del pasaje en autobús, A/w es el número de autos por trabajador en el hogar, R es la raza (0 si es blanco, 1 si no es blanco), Z es la ocupación (0 si no es trabajador profesional, 1 si es trabajador profesional). El ingreso y otras variables socioeconómicas se excluyen dado que son comunes en ambas funciones de utilidad.

En el modelo de elección de transporte se puede también definir un índice que nos muestre la propensión a preferir un carro con relación a transitar en algún otro transporte para la i-ésima persona. Sea $y^*_i = \beta' x_i - \xi_i$ donde x_i es el vector de variables que aparece a la derecha de (6.17) y ξ_i corresponde a ε_{i0}-ε_{i1}. Entonces, el modelo de transporte vendrá determinado por la distribución específica de ξ_i. De esta forma si definimos $y_i = 1$ sí y solo sí $y^*_i > 0$ encontraremos un índice continuo.

6.4.2. Lee, L.F.

Lee define la propensión del i-ésimo trabajador de unirse a un sindicato como:

$$y_i^* = \beta_1 + \beta_2 \left[\frac{w_{i1} - w_{i0}}{w_{i0}} \right] + \beta'_3 x_i - \xi_i \tag{6.18}$$

Donde w_{i1} y w_{i0} será el salario cuando el trabajador pertenece a un sindicato y cuando no respectivamente, x_i es un vector de características del i-ésimo trabajador así como los atributos en la industria en la cual está empleado. El trabajador se une a un sindicato ($y_i = 1$) sí y solo sí $y^*_i > 0$. Lee asume una distribución normal para ξ_i y estima los parámetros por máxima verosimilitud para un Probit.

6.4.3. Pencavel

Pencavel estudia cómo inciden en las decisiones de trabajar de la esposa y el esposo la ayuda económica brindada por el gobierno de los Estados Unidos en Seattle y Denver. De esta forma, estima la probabilidad de trabajar de la esposa usando 1657 familias durante 2 años. Las variables que el autor usa son: F igual a uno si la familia pertenece al experimento y cero lo contrario; L igual a uno si el esposo trabaja durante el año anterior al experimento y cero lo contrario; Y igual a uno si la observación es extraída del segundo año de experimento y cero si es extraída del primer año; U igual a uno si el esposo estuvo desempleado durante el año. Se estimaron dos modelos: uno que reporta estimaciones bajo un modelo de probabilidad lineal $\hat{y}_{PL}$ y otro que estima un modelo Logit $\hat{y}_L$, como se observa a continuación:

$$\hat{y}_{PL} = -0.069F + 0.497L + 0.0055Y + 0.043U + 0.036F.L - 0.08F.Y - 0.041F.U \tag{6.19}$$

(6.20) $\hat{y}_L = -\underset{(0.15)}{0.327F} + \underset{(0.126)}{2.305L} + \underset{(0.121)}{0.309Y} + \underset{(0.131)}{0.25U} + \underset{(0.169)}{0.097FL} - \underset{(0.167)}{0.425FY} - \underset{(0.175)}{0.23FU}$

Entre paréntesis, los errores estándar. Por otro lado, los términos constantes fueron incluidos en el modelo pero no fueron reportados. Como observa Pencavel, las probabilidades estimadas por los dos modelos son parecidas.

6.5. Modelo de efectos fijos y aleatorios en datos de panel

Considérese el modelo estructural Probit para datos de panel:[14]

(6.21) $y^*_{it} = \beta' x_{it} + \varepsilon_{it}$; $\varepsilon_{it} \sim N(0,1)$, i = 1,....,n ; t = 1, ...,T_i

$y^*_{it} = 1 si\, y_{it} > 0$ y cero de otra forma

Si las ε_{it} son variables estándar independientes, la naturaleza de los datos de panel es irrelevante. Como el modelo Probit en sí mismo no tiene efectos fijos, esto es, $\varepsilon_{it} = \alpha_i$, remover la heterogeneidad presente en los datos, sobre todo en corte transversal, es bastante complicado, por esto idealmente nosotros debemos especificar que ε_{it} y ε_s sean libremente correlacionados (en el grupo y no entre grupos), lo cual involucrará calcular la probabilidad conjunta de una distribución normal bivariada de dimensión T, lo cual también es problemático.

En contraposición al Probit, el Logit puede incorporar tratamientos de efectos fijos, por lo cual:

(6.22) $$Prob(y_{it} = 1) = \frac{e^{\alpha_i + \beta' x_{it}}}{1 + e^{\alpha_i + \beta' x_{it}}}$$

El cual es un modelo no lineal. Chamberlain (1980) sugiere que cuando el conjunto de datos tiene un gran n y el T_i es pequeño, la probabilidad no condicionada para nT observaciones vendrá dada por:

(6.23) $$L = \prod_i \prod_t (F_{it})^{y_{it}} (1 - F_{it})^{1-y_{it}}$$

Chamberlain sugiere maximizar la función condicional de verosimilitud:

(6.24) $$L^c = \prod_i prob\left(Y_{i1} = y_{i1}, Y_{i2} = y_{i2}, \ldots Y_{iT} = y_{it} \mid \sum_{t=1}^{T} y_{it} \right)$$

14. Son datos obtenidos a través de un seguimiento a lo largo del tiempo de secciones cruzadas; por ejemplo, un seguimiento a las familias en diferentes periodos de tiempo (Green, pag 256).

De esta forma, para T observaciones en (6.24) se condiciona la verosimilitud al número de unos que hay en este conjunto de T observaciones. Por ejemplo, suponga que la muestra consiste en un gran número de unidades de corte transversal y que cada observación se da solamente en los períodos 1 y 2. La verosimilitud no condicionada será:

(6.25) $$L = \prod_i prob(Y_{i1} = y_{i1}) prob(Y_{i2} = y_{i2})$$

Si las observaciones son independientes, la función de verosimilitud es el producto de las probabilidades, y para cada par de observaciones se tienen las siguientes posibilidades:

- $y_{i1} = 0$ e $y_{i2} = 0 \Rightarrow$ prob (0,0|suma=0) = 1
- $y_{i1} = 0$ e $y_{i2} = 1 \Rightarrow$ prob (1,1|suma=2) = 1

Como podrá observarse el i-ésimo término en la función de verosimilitud condicionada será uno por lo que su contribución a la función de verosimilitud condicionada es nula. En el caso de que $y_{i1}=0$ y que $y_{i2}=1$, tendremos:

- prob((0,1) | suma = 1) = $\dfrac{prob(0,1)}{prob(0,1) + prob(1,0)}$

En términos generales, la probabilidad condicionada para un par de observaciones vendrá dada por:

(6.26) $$\frac{e^{\beta' x_{i2}}}{e^{\beta' x_{i1}} + e^{\beta' x_{i2}}}$$

Al condicionar las probabilidades a la suma en las dos observaciones se remueve la heterogeneidad existente. Por lo tanto, la función de verosimilitud condicional será el producto de los conjuntos de observaciones para los que la suma no es cero ni T.

6.6. El modelo Logit condicionado

Esta es una versión reciente para incluir los atributos presentes en los bienes. Suponga que exista un modelo de elección no ordenada que provenga de una utilidad aleatoria para el i- ésimo consumo en j elecciones. De esta forma, la utilidad de la elección j es:

(8.27) $$U_{ij} = \beta' Z_{ij} + \varepsilon_{ij}$$

Si el consumidor realiza una elección j en particular, y asumiendo que U_{ij} es el máximo entre j utilidades, el modelo estadístico que depende de la elección j será:

(8.27.1) Prob ($U_{ij} > U_{ik}$) $\forall$ $k \neq j$

Sea y_i una variable aleatoria indicando la elección realizada, McFadden muestra que sí y solo sí las j perturbaciones son independientes e idénticamente distribuidas siguiendo una distribución Weibull:

(6.28) $$F(\varepsilon_{ij}) = exp\left(e^{-\varepsilon_{ij}}\right)$$

(6.28.1) $$prob(y_i = j) = \frac{e^{\beta' Z_{ij}}}{\sum_j e^{\beta' Z_{ij}}}$$

El cual se denomina Logit Condicionado. Usualmente la elección depende de los atributos de Z_{ij}. Por lo tanto, x_{ij}, varía entre los individuos además de las w_i características individuales. De esta manera, el modelo puede plantearse de la forma:

(6.28.2) $$prob(y_i = j) = \frac{e^{\beta' Z_{ij} + \alpha' w_i}}{\sum_j e^{\beta' Z_{ij} + \alpha' w_i}}$$

Schmidt y Strauss (1975), estiman un modelo de ocupación basado en una muestra de 1000 observaciones cuya variable dependiente es la Ocupación, que es igual a 1 si es empleado doméstico, 2 si es obrero no especializado, 3 si es artesano (trabajador manual), 4 si es oficinista y 5 si es trabajador profesional. En el conjunto de variables independientes se incluyeron la constante, la educación, la experiencia, la raza y el sexo. El modelo, incluyendo los estratos sociales, será:

(6.28.2) $$prob(y_i = 1) = \frac{e^{\beta'_j x_i}}{\sum_{k=0}^{4} e^{\beta'_k x_i}}$$

Debe observarse que las probabilidades estimadas dependen de los estratos 1, 2, 3, 4 y 5. De esta forma, el modelo condicional Logit computa las probabilidades relativas a cada estrato, el estrato podrá contener pocos casos o muchos casos. Por lo tanto, la probabilidad de que la i-ésima observación pertenezca a un estrato social s_i, viene dada por:

(6.28.3) $$P(Y_i = 1) = \frac{e^{\sum_{j=0}^{p} \beta'_j x_{ij}}}{\sum_{m \in S_i} e^{\sum_{j=0}^{p} \beta'_j x_{mj}}}$$

6.7. Modelos multinomiales

Un modelo multinomial de respuesta cualitativa se define de la siguiente forma. Asuma que la variable dependiente y_i toma m_i + 1 valores {0, 1, 2, . . ., mi}, entonces el modelo multinomial vendrá dado:

(6.29) $P(y_i = j) = F_{ij}(x^*, \theta)$; $i = 1, 2, \ldots, n$ y $j = 1, 2, \ldots, m_i$.

Donde x* y θ son vectores de variables independientes y parámetros respectivamente. De esta forma, m_i depende de un i en particular cuando los individuos tienen diferentes conjuntos de elección. Para definir el estimador de θ en el modelo (6.29) usualmente se definen $\sum^n_{i=1}(m_i + 1)$ variables binarias, de la forma:

(6.29.1) $y_{ij} = 1$ si $y_i = j$

$= 0$ si $y_i \neq j$; $i = 1, 2, \ldots, n$ y $j = 1, 2, \ldots, m_i$.

La función de verosimilitud viene definida como:

(6.29.2) $$Ln \quad L = \sum_{i=1}^{n} \sum_{j=0}^{m_i} y_{ij} Ln F_{ij}$$

Donde el estimador insesgado $\hat{\theta}$ de θ se define como una solución a la ecuación $\frac{\partial LnL}{\partial \theta} = 0$.

6.7.1. Modelos ordenados

Los modelos multinomiales de respuestas cualitativas se pueden clasificar en modelos ordenados y no ordenados. Un modelo ordenado se define como:
(6.30) $P(Y = j \mid x, \theta) = p(S_j)$

Para alguna medida de probabilidad p, sobre x y θ, y una secuencia finita de intervalos sucesivos { S_j } que depende sobre x y θ tal que $\bigcup_j S_j = \Re$.

En los modelos ordenados, los valores que y toma, corresponden a una partición sobre la línea real. A diferencia de un modelo no ordenado, donde la partición correspondería a particiones no sucesivas sobre la línea real o a particiones de dimensiones mayores sobre el espacio euclideano. En la mayoría de las aplicaciones, el modelo ordenado toma la forma:

(6.31) $P(Y = j \mid x, \alpha, \beta) = F(\alpha_{j+1} - x'\beta) - F(\alpha_j - x'\beta)$; $j = 0, 1, \ldots, m; \alpha_0 = -\infty; \alpha_j \leq \alpha_{j+1}; \alpha_{m+1} = \infty$

Para alguna distribución F, se puede definir un modelo Probit ordenado o un modelo Logit ordenado.

6.7.2. Modelo Logit multinomial

El modelo Logit multinomial se define como:

(6.32) $P(Y_i = J) = \left[\sum_{k=0}^{m_i} exp(x'_{ij}\beta)\right]^{-1} exp(x'_{ij}\beta)$; i =1,2,...,n y j= 0,1,....,m_i.

McFadden (1974)[15] considera el siguiente modelo multinomial derivado del problema del consumidor. Considere a un individuo (i) cuyas utilidades están asociadas con tres alternativas, de la forma siguiente:

(6.33) $U_{ij} = \mu_{ij} + \varepsilon_{ij}$, con j = 0, 1, 2

Donde U_{ij} no es una función estocástica sino determinística. En Mora (2001) se presenta un resumen de u_{ij} estocástica y no estocástica. Por otro lado, y ε_{ij}, es el usual término aleatorio de error. De esta forma, el individuo elige aquella alternativa en la que obtiene la mayor utilidad. El multinomial Logit se puede derivar del problema de maximizar la utilidad sí y solo sí los ε_{ij} son independientes y la función de distribución de ε_{ij} viene dada por $exp[-exp(\varepsilon_{ij})]$. De esta manera, la probabilidad de que el i individuo elija una alternativa j, será:

(6.34) $P(y_i=2) = P(U_{i2} > U_{i1}, U_{i2} > U_{i0})$

$$= P(\varepsilon_2 + \mu_2 - \mu_1 > \varepsilon_1, \varepsilon_2 + \mu_2 - \mu_0 > \varepsilon_0)$$

$$= \int_{-\infty}^{\infty} f(\varepsilon_2)\left[\int_{-\infty}^{\varepsilon_2+\mu_2-\mu_1} f(\varepsilon_1)\partial\varepsilon_1 \int_{-\infty}^{\varepsilon_2+\mu_2-\mu_0} f(\varepsilon_0)\partial\varepsilon_0\right]\partial\varepsilon_2$$

$$= \int_{-\infty}^{\infty} exp(-\varepsilon_2)exp[-exp(-\varepsilon_2)] \times exp[-exp(-\varepsilon_2 - \mu_2 + \mu_1)] \times exp[-exp(-\varepsilon_2 - \mu_2 + \mu_0)]\partial\varepsilon_2$$

$$= \frac{exp(\mu_{i2})}{exp(\mu_{i0}) + exp(\mu_{i1}) + exp(\mu_{i2})}$$

Y tomará una forma parecida a (6.32) sí hacemos $\mu_{i2} - \mu_{i0} = x'_{i2}\beta$ y $\mu_{i1} - \mu_{i0} = x'_{i1}\beta$.

6.8. Variables dependientes limitadas

Existe un gran número de datos cuya observación nos muestra que están limitados o acotados de alguna forma. Este fenómeno lleva a dos tipos de efectos: el truncamiento y la censura.

15. En el capítulo siguiente se profundizará en el modelo de McFadden.

El efecto de truncamiento ocurre cuando la muestra de datos es extraída aleatoriamente de una población de interés, por ejemplo, cuando se estudia el ingreso y la pobreza se establece un valor sobre el cual el ingreso se encuentra por encima o por debajo del mismo. De esta forma, algunos individuos podrán no ser tenidos en cuenta.

Por otro lado, censurar es un procedimiento en el cual los rangos de una variable son limitados a priori por el investigador; este procedimiento produce una distorsión estadística similar al proceso de truncamiento.

6.8.1. Truncamiento

Una distribución truncada es la parte de una distribución no-truncada antes o después de un valor específico; imagínese por ejemplo que nosotros deseamos conocer la distribución de los ingresos anteriores a 100.000 o el número de viajes a una zona mayores de 2, ésta será tan sólo una parte de la distribución total.

6.8.1.2. Densidad de una variable aleatoria truncada

Si una variable continua aleatoria x, tiene una función de densidad de probabilidades, pdf(x), y a es una constante, entonces:

(6.35) $$f(x \mid x > a) = \frac{f(x)}{prob(x > a)}$$

Si x tiene una distribución normal con media μ y desviación estándar σ

(6.36) $$prob(x > a) = 1 - \Phi\left(\frac{a - \mu}{\sigma}\right)$$
$$= 1 - \Phi(\alpha)$$

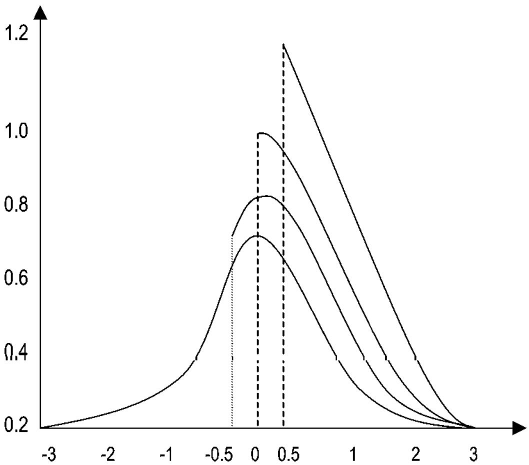
Donde $\alpha = \frac{a - \mu}{\sigma}$ y $\Phi(\alpha)$ es función de densidad acumulativa, entonces la distribución normal truncada será:

(6.37) $$f(x \mid x > a) = \frac{f(x)}{1 - \Phi(\alpha)}$$

$$= \frac{\left(2\pi\sigma^2\right)^{-1/2} e^{-(-x-\mu)^2 / \left(2\sigma^2\right)}}{1 - \Phi(\alpha)}$$

$$= \frac{\left(1/\sigma\right) \phi\left((x - \mu)/\sigma\right)}{1 - \Phi(\alpha)}$$

Donde ϕ será la pdf normal estándar. La distribución normal estándar truncada con $\mu = 0$ y $\sigma = 1$ para a igual a -0.5, 0 y 0.5, será:



GRÁFICA 6.1. Distribuciones normales truncadas.

6.8.1.3. Momentos de una Distribución Truncada

Los momentos de mayor interés son la media y la varianza. Estos momentos son convenientes para hallar la razón inversa de Mills. Si $x \sim N[\mu,\sigma^2]$ con μ constante, entonces la media vendrá dada por E[x|truncamiento] = $\mu + \sigma\lambda(\alpha)$, y la varianza por var[x|truncamiento] = $\sigma^2(1-\delta(\alpha))$ donde $\alpha = (a-\mu)/\sigma$. Por otro lado, nosotros observamos que:

(6.38) $$\lambda(\alpha) = \frac{\phi(\alpha)}{1-\Phi(\alpha)}$$ Si el truncamiento ocurre en x > a

$$\lambda(\alpha) = \frac{-\phi(\alpha)}{\Phi(\alpha)}$$ Si el truncamiento ocurre en x < a

Como $\delta(\alpha) = \lambda(\alpha)(\lambda(\alpha) - \alpha)$ *donde* $0 < \delta(\alpha) < 1 \,\forall\alpha$, y $\dfrac{\partial\phi(\alpha)}{\partial(\alpha)} = -\alpha(\alpha)$.

Donde λ(a) se conoce también como la razón inversa de Mills.

6.8.1.3.4. Estimación por máxima verosimilitud

Tomando el logaritmo de (6.37), y al realizar la suma de los logaritmos de estas densidades, nosotros obtenemos:

(6.39) $$LnL = \frac{-n}{2}\left(Ln(2\pi) + Ln\sigma^2\right) - \frac{1}{2\sigma^2}\sum_i \left(y_i - \beta' x_i\right)^2 - \sum_{i=1}^{n} Ln\left[1 - \Phi\left(\frac{a - \beta' x_i}{\sigma}\right)\right]$$

Las condiciones necesarias para maximizar (6.39) serán:

$$\frac{\partial LnL}{\partial \beta} = \sum_{i=1}^{n} \left[\frac{y_i - \beta' x_i}{\sigma^2} - \frac{\lambda_i}{\sigma} \right] x_i = 0$$

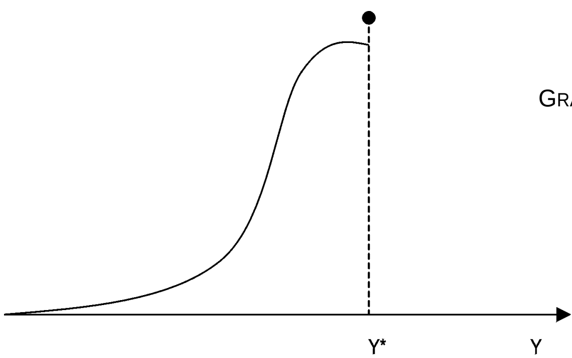
(6.40) $$\frac{\partial LnL}{\partial \sigma^2} = \sum_{i=1}^{n} \left[\frac{-1}{2\sigma^2} + \frac{(y_i - \beta' x_i)^2}{2\sigma^4} - \frac{\alpha_i x_i}{2\sigma^2} \right] = 0$$

$$donde\ \alpha_i = \frac{a - \beta' x_i}{\sigma} \ y \ \lambda_i = \frac{\phi(\alpha_i)}{1 - \Phi(\alpha_i)}$$

6.8.2. Censuramiento

Un procedimiento normal con datos microeconómicos, consiste en censurar la variable dependiente. Cuando la variable dependiente es censurada, los valores en un determinado rango son todos transformados a un valor singular. De esta forma, si definimos una variable aleatoria y transformada de la variable original como:

(6.41) $y = 0$ Si $y^* \leq 0$ $\quad$ $y = y^*$ Si $y^* > 0$



GRÁFICA 6.2. Distribución sensurada.

La distribución correspondiente a $y^* \sim N[\mu,\sigma^2]$ será:

prob $(y=0)$ = prob$(y^* \leq 0)$

$$= \Phi\left(\frac{-\mu}{\sigma}\right)$$

$$= 1 - \Phi\left(\frac{\mu}{\sigma}\right)$$

Si $y^* > 0$ y tiene la densidad de y*, entonces la distribución tiene partes discretas y continuas, donde la probabilidad total será de 1 como se requiere. Para lograr esto, se asigna la probabilidad total en la región censurada al punto de censuramiento.

6.8.2.1. Momentos

La media de una variable censurada vendrá dada por:

(6.42) $$E(y) = \Phi a + (1-\Phi)(\mu + \sigma\lambda)$$

Y la varianza:

(6.43) $$Var\ (y) = \sigma^2 (1-\Phi)\left[(1-\delta) + (\alpha-\lambda)^2\ \Phi\right]$$

Donde $\Phi\left[\dfrac{a-\mu}{\sigma}\right] = \Phi(\alpha) = prob\left(y^* \le a\right) = \Phi\ ;\ \lambda = \dfrac{\phi}{1-\Phi}\ ;\ \delta = \lambda^2 - \lambda\alpha$

6.8.3. Modelos Tobit

Los modelos Tobit se refiere a modelos censurados o truncados donde el rango de la variable dependiente se restringe de alguna forma. El problema puede verse como:

(6.44) $y = y^*$ Si $y^* > y_0$ ó $y = 0$ Si $y^* \le y_0$

Asumiendo que $y^* = \beta_1 + \beta_2 x + \mu$, con μ el término de perturbación aleatoria, y que y_0 varía entre los hogares pero que dicha variación es conocida, el modelo que genera los datos para una función de verosimilitud con n observaciones independientes será:

(6.45) $$L = \prod_0 F_i(y_{0i}) + \prod_1 f_i(y_i)$$

Donde F_i y f_i son las funciones de distribución y de densidad de y_i*. Por otro lado, Π_0 significa el producto sobre aquellos i para los cuales $y_i^* \le y_{0i}$ y Π_1 significa el producto sobre aquellos i para los cuales $y_i^* > y_{0i}$. Por esta razón, el modelo estándar puede definirse como:

(6.46) $$\begin{gathered} y_i^* = \beta' x_i + \varepsilon_i,\ i = 1,2,3,\ldots,n \\ y_i = y_i^*,\ Si\ y_i^* > 0 \\ y_i = 0,\ Si\ y_i^* \le 0 \end{gathered}$$

Asumiendo que $\varepsilon_i \sim N[0, \sigma^2]$, la función de verosimilitud para el modelo Tobit estándar será:

(6.47) $$L = \prod_0 \left[1 - \Phi\left(\frac{\beta' x_i}{\sigma} \right) \right] \prod_1 \sigma^{-1} \phi \left[\frac{y_i - B' x_i}{\sigma} \right]$$

Donde Φ y ϕ son las funciones de distribución y densidad normal estándar. La función de verosimilitud de la versión truncada del Tobit puede escribirse como:

(6.48) $$L = \prod_0 \left[1 - \Phi\left(\frac{\beta' x_i}{\sigma} \right) \right] \prod_1 \sigma^{-1} \phi \left[\frac{y_i - B' x_i}{\sigma} \right]$$

De igual forma se puede plantear la función:

(6.48.1) $$f(y_{ij}) = \begin{cases} \dfrac{\frac{1}{\sigma}\phi\left[y_{ij} - B'_j\, x_{ij}\right]/\sigma}{1 - \Phi\left[- B' x_{ij}\right]/\sigma}, & Si\ \ y_{ij} > 0 \\ 0 & , de\ otra\ forma \end{cases}$$

Dado el creciente uso de los modelo tipo Tobit, Amemiya realizó la laboriosa tarea de clasificar, los modelos Tobit de acuerdo con similitudes en la función de verosimilitud. La caracterización de los tipos de modelos Tobit es la siguiente:

Tipo	Variable Dependiente		
	y_1	y_2	y_3
1	Censurado	-	-
2	Binario	Censurado	-
3	Censurado	Censurado	-
4	Censurado	Censurado	Censurado
5	Binario	Censurado	Censurado

Las funciones de verosimilitud serán:

Tipo	Función de Verosimilitud
1	$p(y_1<0) \,.\, p(y_1)$
2	$p(y_1<0) \,.\, p(y_1>0, y_2)$
3	$p(y_1<0) \,.\, p(y_1, y_2)$
4	$p(y_1<0, y_3) \,.\, p(y_1, y_2)$
5	$p(y_1<0, y_3) \,.\, p(y_1>0, y_2)$

Cada y_j con j =1,2,3 se asume que tiene una distribución del tipo N($\beta'_j x_j$, σ^2_j) y p denota una probabilidad, o densidad o una combinación. Amemiya toma el producto de cada p sobre la observación que pertenece a una categoría particular determinada por el signo de y_1. De esta forma, el tipo 1, modelo Tobit estándar, la función de verosimilitud [p(y_1<0).p(y_1)], será una notación abreviada para $\prod_0 p\left(y^*_{1i} < 0\right)\prod_1 f_{1i}\left(y_{1i}\right)$ donde f_{1i} es la densidad de N($\beta'_1 x_{1i}$, σ^2_1). De igual forma, en cada tipo de modelo el signo de y_i determina cada una de las dos posibles categorías para las observaciones y la variable censurada se observa en una categoría y no en la otra.

6.8.3.1. Modelo Tobit tipo 2: {p(y_1<0) . p(y_1>0, y_2)}

Este modelo se define como:

$$\begin{aligned} y^*_{1i} &= \beta'_1 x_{1i} + \varepsilon_{1i} \\ y^*_{2i} &= \beta'_2 x_{2i} + \varepsilon_{2i} \\ y_{2i} &= \begin{cases} y^*_{2i} \ \ Si \ \ y^*_{1i} > 0 \\ 0 \ \ Si \ \ y^*_{1i} \leq 0 \end{cases} \end{aligned} \quad ; \text{ i = 1, 2, 3,........., n} \tag{6.49}$$

Donde {ε_{1i}, ε_{2y}} son i.i.d extraídos de una distribución normal bivariada con media cero, varianzas σ^2_1 y σ^2_2 y covarianza σ_{12}. Se asume que solamente el signo de y^*_{1i} se observa y que y^*_{2i} se observa cuando $y^*_{1i}>0$. De igual forma, x_{1i} se observa para todos los i, sin embargo x_{2i} no necesariamente se observa para aquellos i tales que $y^*_{1i} \leq 0$. La función de verosimilitud para este modelo será:

$$L = \prod_0 p\left(y^*_{1i} \leq 0\right)\prod_1 f\left(y_{2i} \mid y^*_{1i} > 0\right)p\left(y^*_{1i} > 0\right) \tag{6.50}$$

Donde Π_0 y Π_1 es el producto para aquel i en el cual y_{21} = 0 y $y_{21} \neq 0$ respectivamente y f(• |y^*_{1i}>0) es la densidad condicional de y^*_{2i} dado y^*_{1i} > 0. Este tipo de modelos fueron originalmente usados por Nelson (1977), Dudley y Montmarquette (1976), Weslin y Gillen (1978). Veamos el modelo de Gronau (1976).

Gronau (1976) asume que el salario ofrecido W^0 a cada ama de casa es independiente de las horas trabajadas H. En este sentido no existe un menú de salarios W^0(H). Dado W^0, una ama de casa solucionará el siguiente problema:

(6.50.1) Maximizar u(C,X)

Sujeto a $W^0H + V = Y$ [Restricción de ingresos]

$C + H = T$ [Restricción de tiempo]

Donde C es el tiempo gastado en el hogar para cuidar los niños, X es un vector de dos bienes, T es tiempo total disponible y V otros ingresos. De esta forma una ama de casa no trabajará si:

(6.50.2) $$\left[\frac{\partial u}{\partial C}\left(\frac{\partial u}{\partial X}\right)^{-1}\right]_{H=0} > W^{0}$$

Y trabajará si la desigualdad contraria se mantiene. Si ella trabaja H y la tasa actual de salario es W^0, entonces:

(6.50.3) $$\frac{\partial u}{\partial C}\left(\frac{\partial u}{\partial X}\right)^{-1} = W$$

La parte derecha de (6.50.2) se conoce como el salario de reserva y, en adelante, se denotará como W^r. Asumiendo que W^0 y W^r son combinaciones lineales de variables independientes más el término de perturbación aleatoria ε_i, el modelo estadístico puede plantearse como:

(6.50.4)
$$W_i^0 = \beta'_2 x_{21} + \varepsilon_{2i}$$
$$W_i^r = \alpha' Z_i + \varepsilon_i$$
$$W_i = \begin{cases} W_i^0 \;\; Si \; W_i^0 > W_i^r \\ 0 \;\; Si \; W_i^0 \leq W_i^r \end{cases} ; \forall i = 1,2,....,n$$

Donde los $\{\varepsilon_{2i}, \varepsilon_i\}$ son i.i.d extraídas de una distribución normal bivariada con media cero, varianzas $\sigma^2_{\varepsilon 2}$ y σ^2_{ε} y covarianza $\sigma_{\varepsilon_2 \varepsilon}$. Note que si se hace $W_i^0 - W_i^r = y_{2i}^* \;\; y \; W_i^0 = y_{2i}^*$ tendremos el modelo (6.49).

6.8.3.2. Modelo Tobit tipo 3: $\{p(y_1<0) \,.\, p(y_1, y_2)\}$

El tipo 3 se define como

(6.51)
$$y_{1i}^* = \beta'_1 x_{1i} + \varepsilon_{1i}$$
$$y_{2i}^* = \beta'_2 x_{2i} + \varepsilon_{2i}$$
$$y_{1i} = \begin{cases} y_{1i}^* \;\; Si \;\; y_{1i}^* > 0 \\ 0 \;\; Si \;\; y_{1i}^* \leq 0 \end{cases} ; \forall i = 1,2,3,\ldots,n$$
$$y_{2i} = \begin{cases} y_{2i}^* \;\; Si \;\; y_{1i}^* > 0 \\ 0 \;\; Si \;\; y_{1i}^* \leq 0 \end{cases}$$

Donde los $\{\varepsilon_{1i}, \varepsilon_{2i}\}$ son extraídos de una distribución normal bivariada con media cero, varianza σ^2_1 y σ^2_2 y covarianza σ_{12}. Este modelo difiere del anterior en que y^*_{1i} también se observa cuando este es positivo. La función de verosimilitud vendrá dada por:

(6.52) $$L = \prod_0 p\left(y^*_{1i} \leq 0\right) \prod_1 f\left(y_{1i}, y_{2i}\right)$$

Donde f(.,.) es la densidad conjunta de y^*_{1i} y y^*_2i. Dado que y^*_{1i} se observa cuando es positivo, todos los parámetros son identificables, incluyendo σ^2_1.

El modelo tipo (3) también se conoce como de autoselección tipo Heckman; a continuación se expondrá en qué consiste.

Heckman (1974) propone un modelo diferente al de Gronau en el sentido de que Heckman incluye la determinación del número de horas trabajadas [H] en el modelo. Al igual que Gronau, Heckman asume que el salario ofrecido W^0 es independiente de las horas trabajadas, además la ecuación de W^0 es la misma de la ecuación de Gronau:

(6.52.1) $W^0_i = \beta'_2 X_{2i} + \varepsilon_{2i}$

Y define:

(6.52.2) $$\frac{\partial u}{\partial C}\left(\frac{\partial u}{\partial X}\right)^{-1} = W^r$$

Por lo tanto:

(6.52.3) $W^r_i = \gamma H_i + \alpha' Z_i + \nu_i$

De esta forma, un individuo trabajará si:

(6.52.4) $W^r_i (H_i = 0) = \alpha' Z_i + \nu_i < W^0_i$

El salario W_i y las horas trabajadas H_i se determinan cuando se soluciona simultáneamente (6.52.1) y (6.52.3), entonces $W^0_i = W^r_i = W_i$. El modelo de Heckman se puede definir como:

(6.52.5) $W_i = \beta'_2 X_{2i} + \varepsilon_{2i}$

(6.52.6) $W_i = \gamma H_i + \alpha' Z_i + \nu_i$

Y para el i que cumple la restricción de las horas trabajadas:

(6.52.7) $H^*_i = \beta'_1 X_{1i} + \varepsilon_{1i} > 0$

Donde $\beta'_1 X_{1i} = \gamma^{-1} (\beta'_2 X_{2i} - \alpha' Z_i)$ y $\varepsilon_{1i} = \gamma^{-1} (\varepsilon_{2i} - \nu_i)$. Se puede observar que (6.52.4) y (6.52.7) son equivalentes dado que $\gamma > 0$.

6.8.3.3. Modelo Tobit tipo 4: { $p(y_1<0, y_3)$. $p(y_1, y_2)$}

El modelo Tobit tipo 4 se define como:

$$
\begin{aligned}
&\boldsymbol{y}^*_{1i} = \beta'_1 \boldsymbol{x}_{1i} + \varepsilon_{1i}\\
&\boldsymbol{y}^*_{2i} = \beta'_2 \boldsymbol{x}_{2i} + \varepsilon_{2i}\\
&\boldsymbol{y}^*_{3i} = \beta'_3 \boldsymbol{x}_{3i} + \varepsilon_{3i}\\
&\boldsymbol{y}_{1i} = \begin{cases} \boldsymbol{y}^*_{1i} \ \ \boldsymbol{Si} \ \ \boldsymbol{y}^*_{1i} > \mathbf{0} \\ \mathbf{0} \ \ \boldsymbol{Si} \ \ \boldsymbol{y}^*_{1i} \leq \mathbf{0} \end{cases} \quad ; \forall \boldsymbol{i} = \mathbf{1,2,3,\ldots,} \boldsymbol{n}\\
&\boldsymbol{y}_{2i} = \begin{cases} \boldsymbol{y}^*_{2i} \ \ \boldsymbol{Si} \ \ \boldsymbol{y}^*_{1i} > \mathbf{0} \\ \mathbf{0} \ \ \boldsymbol{Si} \ \ \boldsymbol{y}^*_{1i} \leq \mathbf{0} \end{cases}\\
&\boldsymbol{y}_{3i} = \begin{cases} \boldsymbol{y}^*_{3i} \ \ \boldsymbol{Si} \ \ \boldsymbol{y}^*_{1i} \leq \mathbf{0} \\ \mathbf{0} \ \ \boldsymbol{Si} \ \ \boldsymbol{y}^*_{1i} > \mathbf{0} \end{cases}
\end{aligned} \tag{6.53}
$$

Donde los {$\varepsilon_{1i}, \varepsilon_{2i}, \varepsilon_{3i}$} son i.i.d extraídos de una distribución normal trivariada. Este modelo difiere del tipo 3 en la adición de y^*_{3i}, el cual se observa solamente si $y^*_{1i} \leq 0$. La función de verosimilitud vendrá dada por:

$$
L = \prod_0 \int_{-\infty}^{0} f_3 \left(y^*_{1i}, y_{3i}\right) dy^*_{1i} \prod_1 f_2 \left(y_{1i}, y_{2i}\right) \tag{6.54}
$$

Donde $f_3(. ,.)$ será la densidad conjunta de y^*_{1i} e y^*_{3i}, y $f_2 (. , .)$ la densidad conjunta de y^*_{1i} e y^*_2i.

Nelson y Olson (1978) proveen el siguiente ejemplo: Sea y^*_1 el tiempo que se usa en entrenamiento escolar vocacional, que es completamente observado si $y^*_1 > 0$. Si esta condición no se cumple entonces pertenecerá al intervalo $(-\infty, 0]$. El tiempo que se usa en educación y^*_2 se observa que pertenece a uno de los intervalos $(-\infty, 0]$, $(0, 1]$ y $(1, \infty)$. El salario y^*_3 se observa completamente. De igual forma, y^*_4 son las horas trabajadas y son observables completamente.

Dado que cada variable pertenece a un conjunto de ecuaciones simultáneas se puede estimar solamente la forma reducida, en un modelo de 2 ecuaciones simultáneas:

$$
\boldsymbol{y}^*_{1i} = \zeta_1 \boldsymbol{y}_{2i} + \alpha_1' \boldsymbol{x}_{1i} + \varepsilon_{1i} \tag{6.54.1}
$$

$$
\boldsymbol{y}_{2i} = \zeta_2 \boldsymbol{y}^*_{1i} + \alpha_2' \boldsymbol{x}_{2i} + \varepsilon_{2i} \tag{6.54.2}
$$

Donde y_{2i} se observa siempre. Por otro lado, y^*_{1i} se observa como y_{1i} si $y^*_{1i} > 0$. Los parámetros estructurales de este modelo se estiman de la siguiente forma:

- Paso 1: Estime los parámetros de la ecuación en la forma reducida para y^*_1 con un Tobit de máxima verosimilitud y para la ecuación reducida y_2 estime los parámetros por mínimos cuadrados ordinarios.
- Paso 2: Reemplace y_{2i} al lado derecho de (6.54.1) por la estimación de los mínimos cuadrados ordinarios obtenido en el paso anterior y estime los parámetros de (6.54.1) a través de un Tobit de máxima verosimilitud.
- Paso 3: Reemplace y^*_{1i} en el lado derecho de (6.54.2) por su estimación obtenida en el paso 1 y estime los parámetros de (6.54.2) por mínimos cuadrados ordinarios.

6.8.3.4. Modelo Tobit tipo 5: $\{p(y_1<0, y_3) . p(y_1>0, y_2)\}$

El modelo Tobit tipo 5 se deriva del tipo 4, y se omite la ecuación para y_{1i}. Dado que solamente observamos el signo de y^*_{1i}, tendremos:

$$y^*_{1i} = \beta'_1 x_{1i} + \varepsilon_{1i}$$
$$y^*_{2i} = \beta'_2 x_{2i} + \varepsilon_{2i}$$
$$y^*_{3i} = \beta'_3 x_{3i} + \varepsilon_{3i}$$

(6.55) $$y_{2i} = \begin{cases} y^*_{2i} & Si\ y^*_{1i} > 0 \\ 0 & Si\ y^*_{1i} \leq 0 \end{cases} \quad \forall i = 1,2,3,\ldots,n$$

$$y_{3i} = \begin{cases} y^*_{3i} & Si\ y^*_{1i} \leq 0 \\ 0 & Si\ y^*_{1i} > 0 \end{cases}$$

Donde los $\{\varepsilon_{1i}, \varepsilon_{2i}, \varepsilon_{3i}\}$ son i.i.d extraídas de una distribución normal trivariada. La función de verosimilitud para este modelo será:

(6.56) $$L = \prod_0 \int_{-\infty}^{0} f_3\left(y^*_{1i}, y_{3i}\right) dy^*_{1i} \prod_0 \int_0^{\infty} f_2\left(y^*_{1i}, y_{2i}\right) dy^*_{1i}$$

Donde $f_3(.,.)$ y $f_2(.,.)$ se definen como la densidad conjunta de y^*_{1i} y y_{3i} . Por otro lado, $f_2(.,.)$ es la densidad conjunta de y^*_{1i} y y^*_{2i}. Lee (1978) desarrolla el siguiente modelo: Sea y^*_{2i} el logaritmo de la tasa de salario del i-ésimo trabajador en el caso de que él o ella se unan a un sindicato. Cuando el trabajador decide unirse o no al sindicato, este evento estará determinado por el signo de la variable (y^*_{1i}):

(6.57) $$y^*_{1i} = y^*_{2i} - y_{3i} + \alpha' Z_i + \varepsilon_i$$

Dado que nosotros observamos y^*_{2i} solamente sí el trabajador se une al sindicato y sí no se une al sindicato (y^*_{3i}), el logaritmo del salario observado, denotado por y_i, se define como:

(6.57.1)
$$\begin{aligned} y_i &= y^*_{2i} \ \ si \ \ y^*_{1i} > 0 \\ y_i &= y^*_{3i} \ \ si \ \ y^*_{1i} \leq 0 \end{aligned}$$

Lee asume que x_2 y x_3, las variables independientes en (y_{2i}) y (y^*_{3i}), incluyen características individuales de las firmas y trabajadores tales como la localización regional, el tamaño de la ciudad, la educación, la experiencia, la raza, el sexo y la riqueza. Por otro lado, Z incluye características individuales y variables que representan el costo monetario y no-monetario de ser miembro del sindicato. Dado que y^*_1 no es observado a excepción del signo, la varianza de y^*_1 puede asumirse como unitaria. Lee estima el modelo en dos etapas tipo Heckman aplicado separadamente a (y^*_2) y (y^*_3). Amemiya (1994) define también el Tobit tipo 5 como:

(6.57.2)
$$\begin{aligned} y^*_{ji} &= B'_j x_{ji} + \varepsilon_{ji} \\ Z^*_{ji} &= \gamma'_j S_{ji} + V_{ji} \\ \gamma_i = y^*_{ki} \ \ si \ \ Z^*_{ki} &= \max_j Z^*_{ji} \ ; \ j = 1,2,3,... i = 1,2,3... \end{aligned}$$

Donde cada y_i, x_{ji}, S_{ji} son observados. Se asume que $\{\varepsilon_{ji}, V_{ji}\}$ son i.i.d a través de i, pero correlacionado a través de j, y para cada i y j las dos variables pueden estar correlacionadas. La forma de verosimilitud vendrá dada por:

(6.57.3)
$$L = \prod_1 f\left(y^*_{1i} \mid Z_{1i} \ \ es \ el \ max\right) p_{1i} \times \prod_2 f\left(y^*_{2i} \mid Z_{2i} \ \ es \ el \ max\right) p_{2i} \times \prod_j f\left(y^*_{ji} \mid Z_{ji} \ \ es \ el \ max\right) p_{ji}$$

Donde Π_j es el producto sobre aquel i para el cual Z^*_{ji} es el máximo (max) y $p_{ji} = p(\ Z^*_{ji}$ máximo).

6.9. Contrastes de especificación

De la mano con el desarrollo de las formas de estimación de los modelos, la literatura ha venido ofreciéndonos una serie de contrastes para conocer la "bondad" de los modelos estimados. El origen de estos contrastes se remonta a los trabajos de Rao (1947) en lo que se conoce como "contraste Score" o "contraste de puntuación". Posteriormente Silvey (1959) propone el contraste de multiplicadores de Lagrange que no es otra cosa que el mismo contraste de Rao.

El contraste de multiplicadores de Lagrange no es el único que se pueda usar, pues están el de Hausman (1978) y el contraste de momentos condicionales [Newey

(1985) y Tauchen (1985)]. Para Pagan y Vella (1989) el uso del contraste de especificación en variables dependientes limitadas no es muy común debido a la dificultad computacional de los mismos.

Los contrastes de especificación que se desarrollarán serán: El contraste de Rao ó contraste de puntuación; el contraste de especificación de Hausman, el cual parte de los trabajos de Durbin (1954) y por lo tanto se conoce también como Durbin-Hausman o Durbin-Wu-Hausman debido a los trabajos de Wu (1973); el contraste de la matriz de información de White (1982) y el contraste de momentos condicionales sugerido por Newey (1985) y Tauchen (1985).

6.9.1. Contraste de Rao ó contraste de puntuación

Suponga que existen n observaciones independientes $y_1, y_2, y_3,......, y_n$ con funciones de densidad idénticas $f(y, \theta)$ donde θ es un vector $p \times 1$ de p parámetros. Entonces la función de verosimilitud $L(\theta)$, el vector de puntuación (Score vector) $d(\theta)$, y la matriz de información $I(\theta)$ vienen definidas como:

$$L(\theta) = \sum_{i=1}^{n} \ln f(y_i, \theta) = \sum_{i=1}^{n} L_i(\theta)$$

$$d(\theta) = \frac{\partial L(\theta)}{\partial \theta}\text{, para algún q, } E[d(\theta)] = 0$$

$$I(\theta) = \operatorname{cov}[d(q)] = E[d(\theta)\, d(\theta)'] = E\left[\frac{\partial^2 L}{\partial\theta\partial\theta'}\right]$$

El estimador máximo verosímil $\hat{\Theta}$ viene dado por la ecuación $d(\hat{\Theta}) = 0$. La hipótesis a probar será $H_0 = h(\theta) = 0$, donde $h(\theta)$ es una vector de dimensión r de θ ($r \leq p$) con una constante dada c. Rao (1948), propone el siguiente estadístico:

(6.58) $dI(\bar{\Theta})[I(\bar{\Theta})]^{-1} d(\bar{\Theta})$

Donde ($\bar{\Theta}$) es el estimador máximo verosímil restringido de (θ) bajo Ho. Si Ho es cierto, entonces $d(\bar{\Theta})$ se espera que tienda a cero. Rao muestra que el estadístico Score tiene una distribución chi-cuadrada (χ^2) con r grados de libertad bajo Ho.

Breusch y Pagan (1980) sugieren usar este estadístico como un contraste de especificación. La ventaja del contraste de puntuación consiste en que depende solamente de los estimadores máximos verosímiles del modelo restringido, ya que tanto el vector de puntuación como la matriz de información se basan en el modelo total. Una extensión del contraste de puntuación consiste en un estimador general $\sqrt{n}$ más que en la restricción máximo verosímil, a esta extensión se le denomina el contraste Neyman-Rao [Hall y Mathiason (1990)].

Para modelos de elección binarios, Davidson y Mackinon (1989) muestran que el contraste de puntuación con base en la matriz de información puede ser computado fácilmente y Orme (1992) muestra que las propiedades se conservan en los modelos Tobit[16].

6.9.2. El contraste Durbin-Hausman

El contraste de especificación sugerido por Hausman se basa en la comparación de dos conjuntos de parámetros estimados: Sea $\bar{\Theta}$ un estimador de θ el cual es eficiente bajo Ho, pero inconsistente bajo H_1, y $\overset{+}{\Theta}$ un estimador de θ el cual es consistente bajo Ho y H_1 pero ineficiente bajo Ho. Sea $d = \bar{\Theta} - \overset{+}{\Theta}$ y $\mathrm{var}(\sqrt{n}\,\bar{\Theta}) = V_1$ y $\mathrm{var}(\sqrt{n}\,\overset{+}{\Theta}) = V_0$. El contraste de Hausman se basa en el resultado de $\mathrm{var}(\sqrt{n}\,d) = V_1 - V_0$ bajo Ho (Rao 1973,pp.317). El test estadístico consiste en $(\sqrt{n}d)'(V_1 - V_0)^{-1}(\sqrt{n}\,d)$ que tiene una distribución chi-cuadrada con p grados de libertad, que es la dimensión de θ. En el caso de que $(V_1 - V_0)^{-1}$ no exista, se puede usar la inversa generalizada [Rao y Mitra (1971)], pero la chi-cuadrada tendrá menores grados de libertad. Existe una variación del tipo Hausman, en el que ambos estimadores son inconsistentes bajo H_1, aunque $\mathrm{pLim}\bar{\Theta} \neq \mathrm{pLim}\overset{+}{\Theta}$ bajo H_1[ver Ruud (1984)].

6.9.3. El contraste de la matriz de información de White

La matriz de información de White(1982) se basa en el hecho de que en un modelo especificado correctamente, tendremos:

$$(6.59)\quad E[d(\theta)\, d(\theta)'] = E\left[-\frac{\partial^2 L}{\partial\theta\partial\theta'}\right]$$

Si nosotros consideramos el estadístico $d(\theta)\, d(\theta)' + \left[\frac{\partial^2 L}{\partial\theta\partial\theta'}\right]$ éste deberá tener media cero para un modelo especificado correctamente. Como problema de esta constante se encuentra el hecho de que al tratar de obtener la varianza de este contraste se requieren derivadas de orden superior de al de L. Lancaster (1984) ha mostrado que este estadístico se puede obtener a partir de la regresión:

$$(6.60)\quad r = GC_1 + ZC_2 + \varepsilon_i$$

16. Como existen diferentes formas del contraste Score, que dependen de los estimadores de la matriz de información, en datos pequeños, las propiedades del estimador son bastante malas con respecto a la matriz de información cuando se usa el método del gradiente del producto externo (outer product gradient) [Davidson y Mackinon (1989), Orme (1990)]. Una alternativa consiste en tomar el hesiano como lo observa Taylor (1991).

Donde r es un vector de unos, G es una matriz $n \times p$ cuyos i,j elementos son:

$G_{i,j}(\theta) = \dfrac{\partial L_i(\theta)}{\partial \theta_j}$, i = 1,2,3,...,n; j = 1,2,3,...,p; Z es una matriz cuyos elementos típicos son:

$$\frac{\partial^2 L_i(\theta)}{\partial\theta_j \partial\theta_k} + \frac{\partial L_i(\theta)}{\partial\theta_j}\frac{\partial L_i(\theta)}{\partial\theta_k} \quad \forall\ j = 1,2,...,p;\ k = 1,2,...,j.$$

Z y G son evaluados en el estimador restringido máximo verosímil. El número de columnas en Z será $\frac{1}{2}(k^2 + k)$. Cox (1983) y Chesher (1984) demuestran que el contraste de la matriz de información puede ser interpretado como un contraste de puntuación para rechazar heterogeneidad o una variación de los parámetros en θ.

6.9.4. El contraste de momentos condicionados (CM)

El contraste de momentos condicionados fue sugerido por Newey (1985) y Tauchen (1985), y se basa en la premisa de que bajo una especificación correcta, no solamente se tiene $E[d(\theta)] = 0$ sino también condiciones de sobreidentificación como $E[m(y, \theta)] = 0$. El contraste de momentos condicionados puede reducirse a la regresión:

$$(6.61)\quad i = \hat{G}C + b\hat{m} + \varepsilon_i$$

Donde G es la matriz { $G_{ij} = \dfrac{\partial L_i(\theta)}{\partial \theta_j}$ } e i es un vector de unos. $\hat{G}$ es ortogonal a i dado que $i'\hat{G} = 0$ son las ecuaciones de solución máximo verosímil $\hat{\theta}$ y $\hat{m}$ será ortogonal a i si la condición de momentos se satisface. De esta forma, un contraste para la condición de momentos consiste en la hipótesis b = 0 en la ecuación (6.61). En el caso de que existan r condiciones de momentos, se define la matriz $\hat{m} = m(\hat{\theta})$ de $n \times r$, y se contrasta la hipótesis b = 0 en la regresión (6.61). Aun cuando $i'\hat{G} = 0$ es importante incluir $\hat{G}$ en la regresión [Mackinon (1992, pp,132)]. La regresión artificial se basa en la igualdad informacional (6.59) la cual es válida solamente cuando la densidad completa de las observaciones es especificada correctamente.

Un último contraste de especificación proviene de Pregibon (1980); este contraste consiste en que si la regresión está bien especificada, no se deberían encontrar variables independientes significativas. De esta forma, una clase de error consistirá en que la variable dependiente necesitase algún tipo de transformación en su relación con las variables dependientes, en lo que se conoce como una función de unión (LINK FUNCTION). El contraste consiste en la regresión:

(6.61.1) $y = \beta' x_i + \varepsilon_i$

Donde β´ serán los paramentos estimados, entonces $\hat{\beta}' x_i$ deberá ser significativo y $(\hat{\beta}' x_i)^2$ no. El modelo también se usa para especificación en transformaciones de las variables independientes, de esta forma, si en un modelo inicial Logit o Probit $(\hat{\beta}' x_i)^2$ da significativo entonces alguna variable independiente podría necesitar algún tipo de transformación, una vez realizada esta $(\hat{\beta}' x_i)^2$ no deberá ser significativo.

6.9.5. Contrastes de heterocedasticidad

Entre los primeros trabajos sobre Heterocedasticidad realizados por Maddala y Nelson (1975) se argumentan que una regresión con Heterocedasticidad en los errores, los estimadores son consistentes pero ineficientes. En el caso del Tobit, el estimador máximo verosímil (ML) es inconsistente en la presencia de Heterocedasticidad [Brannas y Laitila (1989)].

En un modelo de regresión, la comprobación de Heterocedasticidad se realiza con base en los residuos del modelo de mínimos cuadrados. Pagan y Park (1993) sugieren que los contrastes existentes para probar heterocedasticidad pueden ser considerados como un contraste de momentos condicionados(CM). La condición de momentos para un contraste CM, será:

(6.61) $$\frac{1}{n}\sum E\left(Z_i(\varepsilon^2{}_i - \sigma^2)\right) = 0$$

Siendo ε_i el error con varianza σ^2 bajo Homocedasticidad, y Z_i es indicador mal especificado. Por ejemplo, sí asumimos que $\mathrm{Var}(\mu_i) = \sigma^2(1 + Z_i\gamma)^2$ se deberá mostrar que esta condición se sigue del contraste de puntuación en el caso del Tobit. En el caso de un Probit o Logit, sea $y^*_i = \beta' x_i + \varepsilon_i$; $\varepsilon_i \sim$ i.i.d$(0, \sigma^2)$, el cual es observado solamente cuando $y_i = \begin{cases} 1 \text{ Si } y^*_i > 0 \\ 0 \text{ De otra forma} \end{cases}$. Dado que y^*_i es observado como una variable dicotómica, solamente β/σ es estimable. Asumiendo que $\sigma = 1$, entonces:

(6.63) Probabilidad $[y_i = 1] = F(\beta' x_i)$

De esta forma, el logaritmo máximo verosímil L será:

(6.63.1) L = $\sum_{i=1}^{n}\left[y_i F(\beta' x_i)+(1-y_i)(1-F(\beta' x_i))\right]=\sum L_i$

Sea $G_{ij}(\beta) = \dfrac{\partial L_i}{\partial \beta_j}=\left[\dfrac{y_i}{F(\beta' x_i)}-\dfrac{1-y_i}{1-F(\beta' x_i)}\right] f_i \quad x_{ij}$

$$= \left[\frac{y_i - F(\beta' x_i)}{F(\beta' x_i)(1-F(\beta' x_i))}\right] f_i \quad x_{ij}$$

Donde f_i es la primera derivada de Fi, es decir, $f(\beta' x_i)$ es la función de densidad. El j-ésimo elemento del vector de puntuación d(β) es igual a $\sum_{i=1}^{n} G_{ij}(\beta)$. Si nosotros deseamos probar la hipótesis de que $\beta=0$ entonces se deberá obtener un contraste de puntuación con nR^2 en una regresión artificial de $\boldsymbol{i}$ sobre G(β) donde $\boldsymbol{i}$ es un vector de unos. Davidson y Mackinnon (1984) denotan este contraste como LM_1.

Considere ahora el estadístico basado en la matriz de información. Diferenciando L_i dos veces con respecto a β, obtenemos:

(6.64) $\dfrac{\partial^2 L_i}{\partial\beta\partial\beta'}=\dfrac{\partial}{\partial\beta'}\left[\dfrac{y_i f(\beta' x_i)}{F(\beta' x_i)}-\dfrac{1-y_i f(\beta' x_i)}{1-F(\beta' x_i)}\right] \quad x_i$

$$\frac{\partial^2 L_i}{\partial\beta\partial\beta'}=\left[\frac{y_i F(\beta' x_i) f'(\beta' x_i)-f^2(\beta' x_i)}{F^2(\beta' x_i)}-\frac{(1-y_i)\,1-F(\beta' x_i) f'(\beta' x_i)+f^2(\beta' x_i)}{(1-F(\beta' x_i))^2}\right] \quad x_i x'_i$$

Tomando $E(y_i) = F(\beta' x_i)$ entonces:

(6.64.1) I(b) = $E\left[-\sum_{i=1}^{n}\dfrac{\partial^2 L}{\partial\beta\partial\beta'}\right]=\sum_{i=1}^{n}\dfrac{f^2(\beta' x_i)}{F(\beta' x_i)(1-F(\beta' x_i))} x_i x'_i$

Davidson y Mackinnon sugieren el siguiente contraste de puntuación: Defina la matriz R(β) de $n \times p$ cuyos elementos típicos son:

(6.64.2) $R_{ij}(\beta) = \left[F(\beta' x_i)(1-F(\beta' x_i))\right]^{-\frac{1}{2}} f(\beta' x_i) x_{ij}$

El n-vector, r(β) tendrá como elemento típico:

(6.64.3) $r_i(\beta) = y_i\left[\dfrac{1-F(\beta' x_i)}{F(\beta' x_i)}\right]^{\frac{1}{2}}-(1-y_i)\left[\dfrac{F(\beta' x_i)}{1-F(\beta' x_i)}\right]^{\frac{1}{2}}$

De esta forma, la matriz de información será $I(\beta) = R'(\beta)R(\beta)$ y el vector de puntuación $d(\beta) = R'(\beta)r(\beta)$. De aquí, el contraste estadístico será:

(6.64.4) $LM_2 = d'Id = r'R(R'R)^{-1}R'r$

El cual no es más que nR^2 de la regresión artificial $r = Rc + \varepsilon_i$. Davidson y Mackinnon arguyen que LM_1 tiene propiedades débiles con relación a LM_2. Para comprobar Heterocedasticidad, se especifica $Var(\mu_i) = (1 + \gamma' Z_i)^2$ y se contrasta Ho: $\gamma = 0$. Por otro lado, Davidson y Mackinnon especifican $Var(\mu_i) = \exp(2\gamma' Z_i)$ y Harvey (1976) $Var(\mu_i) = [\exp(\gamma' Z_i)]^2$. Reemplazando $\beta' x_i$ en el modelo de regresión por $\beta' x_i / (1 + \gamma' Z_i)$ y haciendo $\hat{\beta}$ el estimador máximo verosímil de β bajo Ho: $\gamma = 0$ entonces se reemplaza β por $\hat{\beta}$ en LM_1 y LM_2. La matriz G será ahora de $n \times (p + m)$ donde m es la dimensión de γ, de esta forma:

(6.65) $$\left.\frac{\partial L_i}{\partial \beta}\right|_{\beta=\hat{\beta},\gamma=0} = \left[\frac{y_i f\left(\frac{\hat{\beta}' x_i}{1+\gamma' Z_i}\right)}{F\left(\frac{\hat{\beta}' x_i}{1+\gamma' Z_i}\right)} - \frac{1 - y_i f\left(\frac{\hat{\beta}' x_i}{1+\gamma' Z_i}\right)}{1 - F\left(\frac{\hat{\beta}' x_i}{1+\gamma' Z_i}\right)}\right] x_i$$

(6.65.1) $$\left.\frac{\partial L_i}{\partial \gamma}\right|_{\beta=\hat{\beta},\gamma=0} = \left[\frac{y_i f\left(\frac{\hat{\beta}' x_i}{1+\gamma' Z_i}\right)}{F\left(\frac{\hat{\beta}' x_i}{1+\gamma' Z_i}\right)} - \frac{1 - y_i f\left(\frac{\hat{\beta}' x_i}{1+\gamma' Z_i}\right)}{1 - F\left(\frac{\hat{\beta}' x_i}{1+\gamma' Z_i}\right)}\right] \hat{\beta}' x_i(-Z_i)$$

La matriz R es ahora $n \times (p + m)$ y en la i-ésima fila tendremos:

$$[R_i(\beta), R_i(\gamma)] = \left[F\left(\frac{\beta' X_i}{1+\gamma' Z_i}\right)\left(1 - F\left(\frac{\beta' X_i}{1+\gamma' Z_i}\right)\right)\right]^{-1/2} f\left(\frac{\beta' X_i}{1+\gamma' Z_i}\right)\left[X_i, \beta' \chi_i(-Z_i)\right]$$

El Score vector $d(\hat{\beta}, 0)$ es $R'r$ el cual es un vector $(p + m)$:

$$\left[f\left(\frac{\beta' x_i}{1+\gamma' Z_i}\right)\left(y_i - F\left(\frac{\beta' x_i}{1+\gamma' Z_i}\right)\right)\right]\left[F\left(\frac{\beta' x_i}{1+\gamma' Z_i}\right)\left(1 - F\left(\frac{\beta' x_i}{1+\gamma' Z_i}\right)\right)\right]^{-1} x_i = 0$$

$$\left[f\left(\frac{\beta' x_i}{1+\gamma' Z_i}\right)(y_i - F\left(\frac{\beta' x_i}{1+\gamma' Z_i}\right))\right]\left[F\left(\frac{\beta' x_i}{1+\gamma' Z_i}\right)(1 - F\left(\frac{\beta' x_i}{1+\gamma' Z_i}\right))\right]^{-1}(\beta' x_i)Z_i = 0$$

Dado que $\left[f\left(\frac{\beta' x_i}{1+\gamma' Z_i}\right)(y_i - F\left(\frac{\beta' x_i}{1+\gamma' Z_i}\right))\right]\left[F\left(\frac{\beta' x_i}{1+\gamma' Z_i}\right)(1 - F\left(\frac{\beta' x_i}{1+\gamma' Z_i}\right))\right]^{-1}$ es el residuo generalizado [Gauriéroux, Monfort y Trognon (1987)] y denotándolo por φ, entonces (6.65) puede escribirse como:

(6.66) $\frac{1}{n}\sum \hat{\varphi}_i\, x_i = 0a$ y $\frac{1}{n}\sum \hat{\varphi}(\beta' x_i)Z_i = 0$

De esta forma, el contraste de heterocedasticidad involucra solamente $\hat{\varphi}_i$ y no $\hat{\varphi}_i^{\,2}$. Veamos qué sucede en el modelo Tobit.

Lee y Maddala (1985) usan la forma funcional $Var(\mu_i) = \sigma^2 i = G(\alpha + \delta' Z_i)$ y sugieren probar $\delta = 0$. Asumiendo que $\sigma_i = \sigma + \delta Z_i$ y definiendo un indicador I_i como: $I_i = \begin{cases} 1\ Si\, y^*_i > 0 \\ 0\ De\ otra\ forma \end{cases}$ el logaritmo de la función de verosimilitud L, viene dado por:

(6.67) $$L = \sum_{i=1}^{n} I_i\left[-\frac{1}{2}Ln\sigma^2{}_i - \frac{1}{2\sigma^2{}_i}(y_i - \beta' x_i)^2\right] + \sum_{i=1}^{n}(1-I_i)Ln\left[1-\Phi\left(\frac{\beta' x_i}{\sigma_i}\right)\right]$$

Bajo Ho: $\delta = 0$, entonces:

(6.68) $$\frac{\partial L}{\partial \delta} \approx \sum\left[I_i\left(\frac{(y_i - \beta' x_i)^2}{\sigma^2} - 1\right) + (1-I_i)\left[\lambda_i\left(\frac{\beta' x_i}{\sigma_i}\right)\right]Z_i\right]$$

Donde λ_i es la razón de Mills, esto es:

(6.68.1) $$\lambda_i = \frac{\phi\left(\frac{\beta' x_i}{\sigma}\right)}{1-\Phi\left(\frac{\beta' x_i}{\sigma}\right)}$$

El vector de puntuación, será proporcional a:

(6.68.2) $$\sum\left(\hat{\varepsilon}^2{}_i - \sigma^2\right)Z_i$$

La condición de momentos para probar Heterocedasticidad en el contraste CM, es:

(6.68.3) $$\frac{1}{n}\sum\left(\hat{\varepsilon}^2{}_i - \sigma^2\right)Z_i = 0$$

De esta forma, el test para comprobar Heterocedasticidad se reduce al CM. Otro contraste comúnmente usado consiste en el contraste de las razones de verosimilitud o LR-test, el cual consiste en:

(6.69) $LR = -2[\text{Ln}\,\hat{L}_r - \text{Ln}\,\hat{L}\,]$

Donde $\hat{L}_r$ y $\hat{L}$ son los logaritmos de las funciones de verosimilitud del modelo con heterocedasticidad y el modelo sin heterocedasticidad. De esta forma, nosotros podemos probar Heterocedasticidad sobre una varianza del tipo $Var(\varepsilon_i) = (1 + \gamma' Z_i)^2$ ó $Var(\varepsilon_i) = \exp(2\gamma' Z_i)$ ó $Var(\varepsilon_i) = [\exp(\gamma' Z_i)]^2$ bajo Ho: $\gamma=0$. Por lo tanto, si no existe evidencia estadística para rechazar Ho existirá homocedasticidad y de forma contraria existirá. Además, $LR \sim \chi^2_{d-dr}$ donde d y dr son los grados de libertad asociados al modelo general y al modelo con heterocedasticidad.

6.9.6. Contrastes de normalidad

Cuando no existe normalidad, existen sesgos en los estimadores[Goldberger (1983)]. Pagan y Vella (1989) sugieren un contraste CM para normalidad con base en el tercer y cuarto momento de los residuos. En términos generales, ya se trate de un Logit, Probit o Tobit, el problema consiste en evaluar el momento $E[\mu^p_i \mid y_i = 0]$. Lee y Maddala (1985) sugieren usar el método de recursividad para los momentos de

$\mu_i = \left(\frac{y_i - \beta' x_i}{\sigma}\right)$, el cual consiste en:

(6.70) $$E\left[\left(\mu_i^{p-1} \middle| y_i = 0\right)\right] = p\sigma^2\left[\mu_i^{p-1} \middle| y_i = 0\right] - \sigma\lambda_i(-\beta' x_i)^p$$

Para $p \geq 1$ y $E(\mu_i \mid y_i = 0) = -\sigma\lambda_i$ donde λ_i es la razón de Mills. Definiendo el indicador:

$I_i = \begin{cases} 1\ Si\, y_i > 0 \\ 0\ De\ otra\ forma \end{cases}$, entonces:

(6.71) $$\theta_{1i} = E\left(\mu^3{}_i \middle| y_i\right) = I_i\mu^3{}_i - (1 - I_i)\sigma\lambda_i(2\sigma^2 + (\beta' x_i)^2)$$

(6.72) $\theta_{2i} = E\left(\mu^4{}_i - 3\sigma^4 \middle| y_i\right) = I_i(\mu^4{}_i - 3\sigma^4) + (1 - I_i)\sigma\lambda_i(\beta' x_i)(3\sigma^2 + (\beta' x_i)^2)$

De esta forma, el contraste de normalidad se basa en θ_{1i} y θ_{2i} [Pagan y Vella (1989) y Skeels y Vella (1993)]. Para implementar este contraste se necesitan los residuos, pero no se pueden obtener de ($y_i - \hat{\beta}' x_i$) pues éstos no tienen media cero. Una forma de corregir este problema consiste en obtener los residuos generalizados:

(6.73) $\hat{\eta}_i = -\hat{\sigma}(1 - I_i)\hat{\lambda}_i + I_i \hat{\mu}_i$

Para el Probit, el valor esperado del error será:

(6.74) $E(\mu_i | y_i) = \phi_i(y_i - \Phi_i)\Phi_i^{-1}(1 - \Phi_i)^{-1}$

El residuo generalizado $\hat{\eta}$ para un modelo de elección discreto, viene dado por:

(6.75) $\hat{\eta}_i = \hat{f}_i (y_i - \hat{F}_i)\hat{F}_i^{-1}(1 - \hat{F}_i)^{-1}$

Powell (1986) desarrolla un estimador de cuadrados ordinarios censurado simétricamente (SCLS) para un Tobit de la forma siguiente: Se eliminan las observaciones para las cuales $\hat{\beta}'xi < 0$. A continuación, se hace $y_i = 2\hat{\beta}' x_i$. De esta forma, los y_i son distribuidos simétricamente sobre ($0, 2\hat{\beta}' x_i$) y los errores ($y_i - \hat{\beta}' x_i$) son distribuidos simétricamente sobre $-\hat{\beta}' x_i$ y $\hat{\beta}' x_i$, por lo cual tendrán media cero.

Newey (1985) parte de una distribución F_i, bajo la hipótesis alternativa de que $F_i = \Phi(\mu_i + \gamma_1 \mu_i^2 + \gamma_2 \mu_i^3)$ donde $\mu_i = \left(\frac{\beta' x_i}{\sigma}\right)$. Las condiciones de momentos resultantes serán:

(6.76) $\frac{1}{n}\sum \hat{\mu}_i^2 \hat{\mu}_i = 0$ y $\frac{1}{n}\sum \hat{\mu}_i^3 \hat{\mu}_i = 0$

Bera, Jarque y Lee (1984) construyen el siguiente contraste: Suponga una función de densidad g(m) que satisface la ecuación diferencial:

(6.77) $\frac{\partial \mathbf{Lng}(\mu)}{\partial \mu} = \frac{\mathbf{c}_1 + \mu}{\mathbf{c}_0 + \mathbf{c}_1\mu + \mathbf{c}_2\mu^2}$

Si μ tiene media cero, esto implica que $a = -b_1$, después de reparametrizar:

(6.78) $$\frac{\partial Lng(\mu)}{\partial\mu} = \frac{c_1\mu}{c_0\, c_1\mu + c_2\mu^2}$$

De esta forma, si se hace $c_1 = 0, c_2 = 0$ y $c_0 = \sigma^2$ obtendremos una densidad de $N(0, \sigma^2)$, por lo tanto, un contraste de normalidad consiste en contrastar $c_1 = 0, c_2 = 0$. Smith (1989) sugiere usar polinomios ortogonales y construir un contraste de puntuación para los supuestos distribucionales. Bajo condiciones generales, se puede escribir una densidad $h(Z, \theta, \psi)$ de Z como el producto de otra densidad $f(Z, \theta)$ con momentos finitos de todos los órdenes y una serie de polinomios ortogonales $p_k(Z, \theta)$; de esta forma tendríamos:

(6.79) $$h(Z, \theta, \psi) = f(Z,\theta)\sum_{k=0}^{\infty} a_k(\theta,\psi)\, p_k(Z,\theta)$$

Donde $a_0(\theta, \psi) = 1$ y $p_0(Z, \theta) = 1$. La hipótesis de Smith para normalidad consiste en hacer Ho: $a_k(\theta, \psi) = 0$ para $k > 1$.

Newey (1987) considera un contraste de normalidad tipo Hausman de la forma:

(6.80) $$y^*_i = Z_t\,\delta + \varepsilon_i \text{ con } y_i = \begin{cases} 1 \text{ Si } y^*_i > 0 \\ 0 \text{ De otra forma} \end{cases}$$

Donde Z_t incluye variables endógenas. El contraste se basa en la diferencia entre un estimador Tobit máximo verosímil (MLE), $\hat{\delta}$, y un estimador SCLS, $\hat{\delta}_s$ mínimo cuadrado censurado simétricamente. El contraste de normalidad de Hausman, será:

(6.81) $$Ho : n(\hat{\delta}_s - \hat{\delta})'\left[\nu(\hat{\delta}_s - \hat{\delta})\right]^{-1}(\hat{\delta}_s - \hat{\delta})$$

Los supuestos distribucionales del contraste podrían ser afectados por especificaciones erróneas en la ecuación [Blundell y Meghir (1986)]. Por otro lado, cuando existen distorsiones en el tamaño el contraste podría sobrerrechazar la hipótesis nula [Newey (1987)]. Algunos autores como Chesher, Lancaster e Irish (1985) proponen usar métodos gráficos para detectar fallas en los supuestos distribucionales. Aunque el procedimiento es menos formal, podría informarnos preliminarmente sobre problemas de distribución, y consiste en tomar los residuos $\hat{\mu}_i$ y computar la función de distribución usando el método de Kaplan-Meier (KPM). La comparación visual de esta función con respecto a la distribución F se hace a través de graficar F-1[KPM($\hat{\mu}_i$)] contra

$\hat{\mu}_i$ en el caso de una distribución normal F = Φ. Si el modelo es correcto, la gráfica que resulta será continua [Horowitz y Neuman (1989)].

6.9.7. Contraste de correlación contemporánea

Considere el siguiente Probit bivariado:

$$(6.82)\quad \begin{aligned} y_1^* &= \beta'_1 x_{1i} + \varepsilon_{1i} \\ y_2^* &= \beta'_2 x_{2i} + \varepsilon_{2i} \end{aligned}$$

Donde y^*_1 y y^*_2 son variables dependientes dicotómicas y {ε_1 y ε_2} son normales bivariadas con media cero varianza unitaria y correlación ρ. El vector de puntuación para este modelo viene dado por:

$$(6.83)\quad \frac{\partial L}{\partial \beta_1} = \sum x_{1i}\hat{\varepsilon}_{1i}\,,\ \frac{L}{\partial \beta_2} = \sum x_{2i}\hat{\varepsilon}_{2i}\,,\ \frac{L}{\partial \rho} = \sum \hat{\varepsilon}_{1i}\hat{\varepsilon}_{2i}$$

Donde $\hat{\varepsilon}_{1i}$ y $\hat{\varepsilon}_{2i}$ son los residuos generalizados para las dos ecuaciones Probit. El contraste estadístico viene dado por nR^2 de una regresión artificial de un vector de unos sobre ($x_1\hat{\varepsilon}_1$, $x_2\hat{\varepsilon}_2$, $\hat{\varepsilon}_1\hat{\varepsilon}_2$). Una alternativa simple consiste en computar los residuos generalizados, su correlación al cuadrado r^2, y usar nr^2 como una chi-cuadrado con un grado.

Kiefer (1982) desarrolla un contraste Score para un Probit multivariado: él comienza con el supuesto general de que la matriz de correlación de los errores es R y crea un contraste de puntuación para la hipótesis R = I. Kiefer también desarrolla un contraste para la hipótesis $\rho=0$ cuando R = $(1-\rho)I + \rho ee'$, donde e es un vector de unos; este contraste es bastante conveniente en modelos de efectos aleatorios con datos de panel.

6.9.8. Contraste de sesgos de selección

El contraste para sesgos de selección fue el primer contraste de especificación en modelos con variables dependientes limitadas. Este contraste fue desarrollado por Gronau (1974) y Heckman (1979). En términos generales se le conoce como el contraste de Heckman. El problema planteado parte del modelo de autoselección tipo Heckman, de la forma:

$$(6.84)\quad y^*_{1i} = \beta'_1 x_{1i} + \varepsilon_{1i}$$

Y la ecuación de selección es:

(6.85) $y^*_{2i} = \beta'_2 x_{2i} + \varepsilon_{2i} > 0$

De esta forma, y^*_{1i} es observado e igual a y_{1i} sí y solo sí $y^*_{2i} > 0$. Por lo tanto, y^*_{1i} es censurado por la ecuación de selección. Es también de esperar que ε_{1i} y ε_{2i} tengan un grado de correlación ρ. Dado que y^*_{2i} es observado como una variable dicotómica, normalizando (6.85) a través de Var(ε_{2i}) = 1, deberemos asumir:

(6.86) $\begin{pmatrix} \varepsilon_{1i} \\ \varepsilon_{2i} \end{pmatrix} \sim N\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \sigma^2 & \rho\sigma \\ \rho\sigma & 1 \end{pmatrix} \right)$

Una estimación por mínimos cuadrados ordinarios de (6.84) da como resultado estimadores sesgados de β_1 dado que E [$\varepsilon_{1i} \mid y^*_{2i} > 0$] es diferente de cero; esta expresión viene dada por $\rho\sigma\lambda_i$ donde λ_i es la razón de Mills. De esta forma (6.84) en términos de la variable observada y_{1i} puede escribirse como:

(6.87) $y_{1i} = \beta'_1 x_{1i} + \rho\sigma\lambda_i + \varepsilon_{1i}$ donde E[ε_{1i}] = 0

Una inspección de (6.87) nos muestra la naturaleza del error, esto es, la omisión de λ_i. Dado que λ_i no es observado, Heckman sugiere obtener el estimador preliminar de $\hat{\lambda}_i$ con base en $\hat{\beta}_2$, el Probit estimado de (6.85) y luego estimar (6.87) por mínimos cuadrados ordinarios. Adicionalmente si ρ es igual a cero no existirán sesgos. Melino (1982) muestra que el contraste de significancia de λ_i en (6.87) es un contraste de puntuación sobre $\rho = 0$. El modelo puede ser estimado en dos etapas, sin embargo, existen restricciones: Si x_{2i} contiene solamente una constante entonces λ_i es una constante y el coeficiente $\rho\sigma$ no es estimable; si λ_i es una función lineal de los componentes de x_{1i} entonces se producirá multicolinealidad, esto ocurre cuando x_{2i} contiene solamente variables dummy (falsas) y x_{1i} incluye las mismas variables dummy y sus combinaciones. Usando ML el problema puede resolverse, sin embargo la estimación bajo ML tiene como problema que se podría entrar en un LOOP al buscar convergencia, y en muchas ocasiones podría encontrarse máximo locales y no globales. Recientemente se ha popularizado el estimador ML provisto por el programa LIMDEP para este tipo de Modelos; Nawata (1993a, 1993b) ha demostrado cómo los estimadores obtenidos por LIMDEP no son confiables. Olsen (1982) muestra que la función de verosimilitud para el modelo de selección tiene un único máximo condicionado sobre ρ. Olsen, sugiere que uno puede obtener estimadores ML condicionados sobre ρ y examinarlos en ρ. El procedimiento ha sido usado por Nawata, mostrando que este procedimiento da resultados más confiables que usar el programa LIMDEP.

6.9.9. Contraste de estabilidad

No es muy común contrastar estabilidad en modelos de variables dependientes limitadas, sin embargo, Anderson (1987) abre el camino en este tipo de contrastes. Anderson propone comparar el logaritmo de la verosimilitud cuando el modelo es regresado sobre un período, con respecto a un período posterior. El trabajo se inspira en el contraste de estabilidad de Chow, extendiéndose el uso de las variables dummy a los modelos Tobit y Probit. Hoffman y Pagan (1989) sugieren, siguiendo a Anderson, definir primero un período de 1 hasta s y un período de s+1 hasta s+S, y elaborar el estadístico:

(6.88) $$\hat{\tau} = \frac{1}{n}\sum_{t=s+1}^{s+S} \hat{d}_t$$

Donde $\hat{d}_t$ son los estimadores de puntuación. La media de los períodos del estimador posterior deberá ser cercana o igual a cero en el caso de estabilidad en los parámetros. La varianza asintótica de $\hat{\tau}$ será $S(1+k)I_{\theta\theta}$ [Hoffman y Pagan (1989)] donde K = s/S. De esta forma, el contraste estadístico será:

(6.89) $$\frac{1}{S(1+k)}\hat{\tau}' \hat{I}_{\theta\theta}^{-1} \hat{\tau}$$

el cual sigue una distribución chi-cuadrada con s grados de libertad. Este contraste puede ser aplicado a cualquier modelo de variables dependientes limitadas y se estima por máxima verosimilitud.

6.9.10. Contraste de exogeneidad

En modelos de ecuaciones simultáneas que involucran variables dependientes limitadas, Groger (1990) considera un contraste de exogeneidad tipo Hausman a través de una estimación de mínimos cuadrados ordinarios no-linelaes. Smith y Blundell (1986) consideran el siguiente modelo:

(6.90) $$\begin{matrix} y^*_{1i} = y'_{2i}\gamma_1 + x'_{1i}\beta_1 + \varepsilon_{1i} \\ y'_{2i} = x'_i\pi_2 + \gamma'_{2i} + \varepsilon_{2i} \end{matrix} \; ; \; y_{1i} = \begin{cases} y^*_{1i} & Si \;\; y^*_{1i} > 0 \\ 0 & De\ otra\ forma \end{cases}$$ y

$$\begin{pmatrix} \varepsilon_{1i} \\ \varepsilon_{2i} \end{pmatrix} \sim IN\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \sigma_{11} & \sigma_{12} \\ \sigma_{21} & \Sigma_{22} \end{pmatrix}\right)$$

Donde y^*_{1i} no es observado. La hipótesis para probar exogeneidad consiste en Ho: $\sigma_{12} = 0$. Smith y Blundell sugieren:

- Suponga que $\varepsilon_{1i} = \nu'_{2i}\lambda + e_{1i}$ y sustitúyalo en (6.90)
- Denote $Var(e_{1i}) = \sigma_{12}$
- Estime $y^*_{1i} = y'_{2i}\,\gamma_1 + x'_{1i}\beta_1 + \nu'_{2i}\,\alpha + E_{1i}$ como un Probit y pruebe la hipótesis Ho: $\alpha = 0$. Smith y Blundell a este estimador, el estimador condicional ML.

Rivers y Voung (1988) consideran el mismo modelo que Blundell y Smith pero en un contexto Probit y Vella y Verbeek (1993) lo consideran para el caso de un modelo de panel.

6.10. Variables latentes

Las variables latentes representan conceptos unidimensionales en su más pura forma, puede decirse que se trata de variables abstractas como inteligencia, paisaje, etc. Como todas las variables latentes corresponden a conceptos, ellas son variables hipotéticas que varían en su grado de abstracción: inteligencia, clase social, poder y expectativas son variables latentes abstractas creadas en la teoría. Variables menos abstractas son la educación y el tamaño de la población.

Un ejemplo es la hipótesis de Emile Durkheim sobre la relación inversa entre la cohesión social y el suicidio: la cohesión social se refiere a la solidaridad de grupo, la cual es una variable abstracta; el suicidio es directamente observable, pero la relación directa-indirecta es muy débil de acuerdo con la misma clasificación de los suicidios.

Un modelo latente se acompaña de un conjunto de ecuaciones estructurales que resumen las relaciones entre las variables latentes. Bollen (1989) usa las relaciones entre la democracia política y la industrialización en países desarrollados, para introducir la noción de modelos de variables latentes. Dado que algunas sociedades han alternado entre dictaduras y regímenes electorales, es difícil discernir si la asociación realmente existe. La democracia política se refiere a la extensión de los derechos políticos (imparcialidad de las elecciones) y libertades políticas (libertad de prensa) en un país. La industrialización es el grado en el cual la economía de una sociedad se caracteriza por el proceso de manufactura mecanizado, esto implica riqueza social, población educada, avances en el estándar de vida, y éstas son las oportunidades de una democracia.

Suponga que se tienen tres variables latentes aleatorias: democracia política en 1965 y 1960 e industrialización en 1960. Uno podría asumir que la democracia política en 1965 es una función de la democracia política e industrialización de 1960. No existe nada que nos diga que el nivel de industrialización es una variable latente exógena (independiente) y se simboliza como ξ_1, esta es exógena, en tanto sus causas están por fuera del modelo. La variable democracia política es una variable latente endógena, ella está determinada por variables en el modelo, cada variable latente es representada por η_i. De esta forma, la democracia política en 1960 es representada por η_1 y la democracia política en 1965 por η_2, las variables latentes endógenas son parcialmente explicadas en el modelo y el componente no explicado γ_i es un término aleatorio; de esta forma, el modelo de variables latentes para el ejemplo será:

(6.91)
$$\eta_1 = \gamma_{11}\xi_1 + \gamma_1$$
$$\eta_2 = \boldsymbol{B}_{21}\eta_1 + \gamma_{21}\xi_1 + \gamma_2$$

Escribiéndolo en notación matricial:

(6.91.1)
$$\begin{bmatrix} \eta_1 \\ \eta_2 \end{bmatrix} = \begin{bmatrix} \boldsymbol{0} & \boldsymbol{0} \\ \boldsymbol{B}_{21} & \boldsymbol{0} \end{bmatrix} \begin{bmatrix} \eta_1 \\ \eta_2 \end{bmatrix} + \begin{bmatrix} \gamma_{11} \\ \gamma_{21} \end{bmatrix} [\xi_1] + \begin{bmatrix} \gamma_1 \\ \gamma_2 \end{bmatrix}$$

Compactamente, tenemos:

(6.92) $\eta = \boldsymbol{B}\eta + \gamma\xi + \gamma$

Además, E(η) = 0, E(ξ) =0, E(γ) = 0 y γ no esta correlacionado con ξ. Por otro lado, (I-B) es no singular. De esta forma (6.92) es una matriz general que representa las ecuaciones estructurales para un modelo de variable latentes.

6.10.1. Ecuaciones estructurales con variables observadas

La ecuación (6.92) es una representación general de ecuaciones estructurales con variables observadas de la forma:

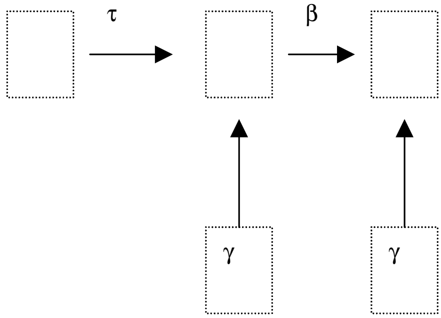
(6.93) $\boldsymbol{y} = \boldsymbol{B}\boldsymbol{y} + \Gamma\boldsymbol{x} + \gamma$

Donde y es un vector de variables endógenas de m$\times$1, B es una matriz de coeficientes de m$\times$m, Γ es una matriz de coeficientes de m$\times$n, x es un vector de variables exógenas n$\times$1 y γ es un vector de errores en las ecuaciones de m$\times$1. Los γ representan los errores aleatorios en las relaciones entre los y_is y los x_is. El supuesto estándar consiste en que los errores γ no están no-correlacionados con x. Por otro lado, E(γ^2_{ik}) = Var γ_i $\forall$ k y Cov(γ_{ik} , γ_{il}) = 0 $\forall$ k $\neq$ l.

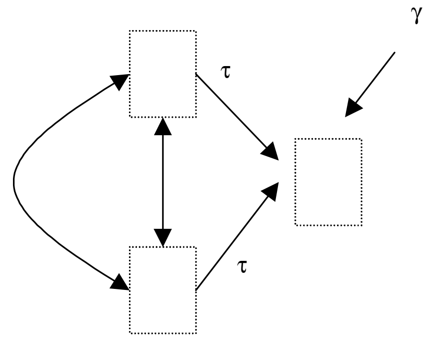
El modelo implícito para las ecuaciones estructurales con variables observadas, será:

(6.93.1)
$$\boldsymbol{y} = \eta \ \textbf{\textit{donde}} \ \boldsymbol{y} = \boldsymbol{m} \times \boldsymbol{1} \ \textbf{\textit{vector de var iables observadas}}$$
$$\boldsymbol{x} = \xi \ \textbf{\textit{donde}} \ \boldsymbol{x} = \boldsymbol{n} \times \boldsymbol{1} \ \textbf{\textit{vector de var iables observadas}}$$

Los modelos recursivos son sistemas de ecuaciones que tienen casualidad no recíproca como en 6.3.a o retroalimentación como en 6.3.b. Cuando esto es cierto es posible denotar a B como una matriz triangular inferior, adicionalmente la matriz de covarianzas de los errores en las ecuaciones (Ψ) será diagonal, esto significa que los errores para una ecuación no están correlacionados con los errores de las otras ecuaciones. Por ejemplo, si y_1 causa y_2, y_2 no tiene efecto sobre y_1 ni directamente, ni a través de alguna relación con otras variables, veamos:

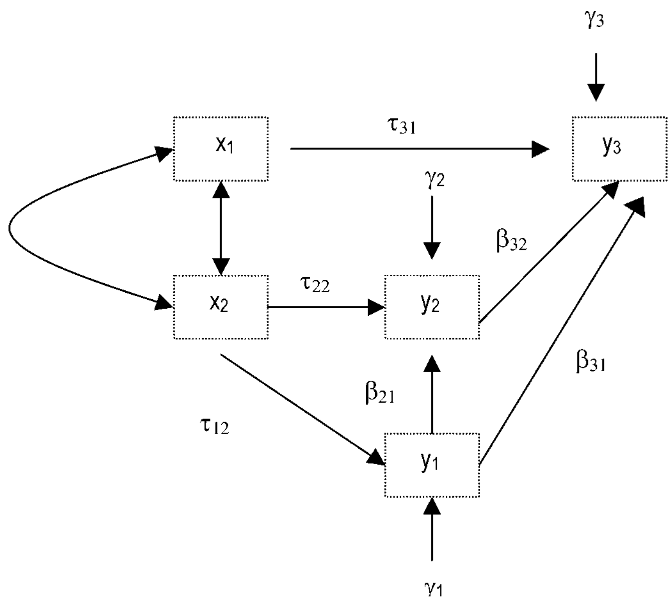


GRÁFICA 6.3.a. Causalidad unidireccional.



GRÁFICA 6.3.b. Retroalimentación.

McDonald y Clelland (1984) proponen el siguiente modelo, el cual representa el deseo de unión de los trabajadores textiles en el suroeste de los Estados Unidos. Donde x_1 representa los años en la fábrica de textiles, x_2 representa la edad, y_1 la sumisión al empresario, y_2 el apoyo a las actividades de los trabajadores y, y_3 la propensión hacia las uniones.



GRÁFICA 6.3.c. Retroalimentación en un modelo de sindicatos.

El ordenamiento causal de McDonald y Clelland consiste en que la sumisión influencia la actitud hacia el activismo y_2 y las uniones y_3, y el activismo afecta el sentimiento de unión. Por otro lado, los γ's errores no están correlacionados a través de las ecuaciones. La ecuación en términos matriciales será:

$$\begin{bmatrix} y_1 \\ y_2 \\ y_3 \end{bmatrix} = \begin{bmatrix} 0 & 0 & 0 \\ \beta_{21} & 0 & 0 \\ \beta_{31} & \beta_{32} & 0 \end{bmatrix} \begin{bmatrix} y_1 \\ y_2 \\ y_3 \end{bmatrix} + \begin{bmatrix} 0 & \tau_{12} \\ 0 & \tau_{22} \\ \tau_{31} & 0 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} + \begin{bmatrix} \gamma_1 \\ \gamma_2 \\ \gamma_3 \end{bmatrix} \tag{6.94}$$

6.10.2. La Matriz de covarianzas

La hipótesis del modelo de ecuación estructural general, consiste en

$$\sum = \sum(\theta) \tag{6.95}$$

Donde Σ es la matriz de covarianzas de la población de y e x, y $\Sigma(\theta)$ es la matriz de covarianzas como función de un modelo de parámetros libres en θ. La ecuación (6.92) implica que cada elemento de la matriz de covarianzas es una función de uno o más parámetros del modelo. La relación entre Σ y $\Sigma(\theta)$ es fundamental para entender la identificación y estimación del modelo ajustado.

La matriz de covarianzas $\Sigma(\theta)$ reúne los siguientes elementos: Primero, la matriz de covarianzas de y. Segundo, la matriz de covarianzas de x con y. Tercero, la matriz de covarianzas de x. Considérese primero $\Sigma_{yy}(\theta)$, esto es, la matriz de covarianzas de y:

$$\begin{aligned} \Sigma_{yy}\ \theta &= E(yy') \\ &= E\left[(I-B)^{-1}(\Gamma x+\gamma)\left((I-B)^{-1}(\Gamma x+\gamma)'\right)\right] \\ &= E\left[(I-B)^{-1}(\Gamma x+\gamma)(x'\Gamma'+\gamma')(I-B)^{-1'}\right] \\ &= (I-B)^{-1}\left[E(\Gamma xx'\Gamma')+E(\Gamma x\gamma')+E(\gamma x'\Gamma')+E(\gamma\gamma')\right](I-B)^{-1'} \\ &= (I-B)^{-1}(\Gamma\Phi\Gamma'+\Psi)(I-B)^{-1'} \end{aligned} \tag{6.95.1}$$

Donde Φ es la matriz de covarianzas de x, Ψ es la matriz de covarianzas de y. La matriz de covarianza de x, $\Sigma_{xx}(\theta)$, es igual a Φ, esto es:

$$\sum\nolimits_{xx}(\theta) = \mathbf{E(xx')} = \mathbf{\Phi} \tag{6.95.2}$$

La parte final de la matriz de covarianzas es $\Sigma_{xy}(\theta)$, esto es, la covarianza de x con y:

$$\begin{aligned}
&\textstyle\sum_{xy}(\theta) = E(xy') \\
(6.95.3) \qquad &= E\left[x\left((I-B)^{-1}(\Gamma x+\gamma)\right)'\right] \\
&= \Phi\Gamma'(I-B)^{-1'}
\end{aligned}$$

De esta forma, encontramos que $\Sigma(\theta)$ será:

$$(6.96) \quad \sum(\theta) = \begin{bmatrix} (I-B)^{-1}(\Gamma\Phi\Gamma'+\Psi)(I-B)^{-1'} & (I-B)^{-1}\Gamma\Phi \\ \Phi\Gamma'(I-B)^{-1'} & \Phi \end{bmatrix}$$

6.10.3. Identificación

La identificación del modelo (6.96) con una o más ecuaciones, requiere una investigación de cuáles parámetros son conocidos y desconocidos. Por parámetros conocidos entiéndase aquellos que pueden ser identificados, estos parámetros generalmente son características de la población y de la distribución de las variables observadas como las varianzas y covarianzas para los cuales los estimadores de la muestra son consistentes. Los parámetros desconocidos son aquellos parámetros cuyo estatus de identificación no es conocido, estableciendo entonces el investigador cuándo existen valores únicos para estos.

El modelo se encuentra identificado si se muestra que los parámetros desconocidos son funciones solamente de los parámetros identificados, y que estas funciones llevan a soluciones únicas. Suponga que la Var(y) es el parámetro identificado, θ_1 y θ_2 son parámetros desconocidos, y la ecuación que los relaciona es Var(y) = $\theta_1 + \theta_2$. La identificación deberá establecer cuando se alcanzan valores únicos de θ_1 y θ_2 en esta ecuación. Claramente con dos parámetros desconocidos θ_1, θ_2 en una sola ecuación, la identificación no es posible. Para algún valor dado de Var(y) un conjunto infinito de valores de θ_1 y θ_2 satisfacen dicha ecuación. Sin embargo, adicionando una segunda ecuación $\theta_1 = \theta_2$ se puede asegurar la identificación por la cual cada parámetro será igual a $\frac{Var(y)}{2}$.

Este principio general deberá mantenerse para ecuaciones estructurales más complicadas. Los parámetros cuya identificación es desconocida están en θ y θ contiene t libres parámetros restringidos(no-redundantes) de B, Γ, Φ y Ψ. La ecuación que relaciona Σ a θ es la hipótesis de la estructura de la covarianza $\Sigma = \Sigma(\theta)$. Este principio general lo enuncia Bollen de la siguiente forma "Si un parámetro desconocido en θ se escribe como una función de uno o más elementos de Σ, este parámetro es identificado, si todos los parámetros desconocidos en θ son identificados, entonces el modelo es identificado".

Un conjunto más apropiado de identificación es conocido como la regla t, la regla del B nulo, la regla recursiva y las condiciones de rango y orden.

6.10.3.1. Regla t

Esta es la condición más sencilla, pero no es una condición suficiente. La regla t, parte de que el número de elementos no-redundantes en la matriz de covarianzas de las variables observadas deberá ser mayor o igual al número de parámetros desconocidos en θ, esto es:

$$t \leq \left(\tfrac{1}{2}\right)(p+q)(p+q+1) \tag{6.97}$$

Donde, $p+q$ es el número de variables observadas y t es el número de parámetros libres en θ.

6.10.3.2. Regla del B nulo

En un modelo multiecuacional donde las variables que no son endógenas afectan a alguna variable endógena, la matriz B es cero. Veamos:

$$\begin{gathered} y_1 = \tau_{11} x_1 + \tau_{12} x_2 + \gamma_1 \\ y_2 = \tau_{21} x_1 + \tau_{22} x_3 + \gamma_2 \\ cov(x_i, \gamma_j) = 0 \, \forall i = 1,2,3; j = 1,2 \end{gathered} \tag{6.98}$$

La matriz B es cero dado que y_1 no afecta a y_2, ni y_2 afecta y_1. De esta forma se establece que la identificación de algún modelo donde B es cero, los parámetros desconocidos en Γ, Φ y Ψ son funciones de los parámetros identificados de Σ. Sustituyendo $B=0$ en (6.96) y particionando Σ, obtenemos:

$$\begin{bmatrix} \Sigma_{yy} & \Sigma_{yx} \\ \Sigma_{xy} & \Sigma_{xx} \end{bmatrix} = \begin{bmatrix} \Gamma\Phi\Gamma'\Psi & \Gamma\Phi \\ \Phi\Gamma' & \Phi \end{bmatrix} \tag{6.99}$$

Como puede observarse $\Phi = \Sigma_{xx}$, por lo cual Φ es identificado. Por otro lado:

$$\Phi\Gamma' = \Sigma_{xy} = \Sigma_{xx}\Gamma' = \Gamma' = \Sigma_{xx}^{-1}\Sigma_{xy} \tag{6.100}$$

De esta forma, Γ es una función conocida que se identifica a partir de las matrices de covarianzas y que en sí misma es identificada. También podemos observar que:

$$\begin{aligned} \Psi &= \Sigma_{yy} - \Gamma\Phi\Gamma' \\ &= \Sigma_{yy} - \Sigma_{yx}\Sigma_{xx}^{-1}\Sigma_{xx}\Sigma_{xx}^{-1}\Sigma_{xy} \\ &= \Sigma_{yy} - \Sigma_{yx}\Sigma_{xx}^{-1}\Sigma_{xy} \end{aligned} \tag{6.101}$$

Entonces, cuando B = 0, Γ, Φ y Ψ, pueden escribirse como funciones de las matrices de covarianzas identificadas de las variables observadas. Si los errores de una ecuación no están correlacionados con aquellos de las otras ecuaciones, en un sistema (Ψ es diagonal), entonces esas ecuaciones pueden tratarse como separadas o no relacionadas. Si Ψ no es diagonal y los errores de las últimas dos ecuaciones están correlacionadas, entonces tal modelo será llamado "Seemingly unrelated regresions". La regla B nula es una condición suficiente para identificar un modelo.

6.10.3.3. Regla recursiva

A diferencia de la regla anterior, la regla recursiva no requiere que B = 0; para aplicar la regla recursiva B deberá ser una matriz triangular, y Ψ diagonal. Una condición más exacta para B, consiste en que ésta sea una matriz triangular inferior. Si ambas condiciones se mantienen, el modelo está identificado. Supongamos el modelo de sindicatos de la gráfica 6.3.c:

$$\begin{aligned} \boldsymbol{y}_1 &= \tau_{12}\boldsymbol{x}_2 + \gamma_1 \\ \boldsymbol{y}_2 &= \beta_{21}\boldsymbol{y}_1 + \tau_{22}\boldsymbol{x}_2 + \gamma_2 \\ \boldsymbol{y}_3 &= \beta_{31}\boldsymbol{y}_1 + \beta_{32}\boldsymbol{y}_2 + \tau_{31}\boldsymbol{x}_1 + \gamma_3 \end{aligned} \tag{6.102}$$

Una propiedad para todos los modelos recursivos consiste en que para una ecuación dada, el término de error γ no esté correlacionado con las variables explicatorias. De esta forma, cov $(x_2, \gamma_2) = 0$, cov $(x_3, \gamma_1) = 0$ y cov $(x_2, \gamma_3) = 0$, y de igual forma para cov $(\gamma_2, y_1) =$ cov $(\gamma_2, \gamma_2 x_2 + \gamma_1) = 0$. Así, γ_2 no está correlacionado con y_1 y x_2, las dos variables explicatorias de la segunda ecuación, y de igual forma cov $(\gamma_3, y_1) = 0$ y cov $(\gamma_3, y_2) = 0$. En general, para la i-ésima ecuación en algún modelo recursivo, γ_i no está correlacionado con las variables endógenas, las cuales son variables explicatorias en esa ecuación; esto se debe a que las variables endógenas están en función de las variables exógenas y de los errores de las otras ecuaciones, los cuales no están correlacionados con γ_i.

6.10.3.4. Condiciones de rango y orden

Si una condición de restricción en una ecuación se determina a partir de las variables excluidas, entonces "una condición necesaria para que una ecuación sea identificada consiste en que el número de variables excluidas de la ecuación sea al menos p-1 ". Considere el modelo:

(6.103) $$y_i = \left[\boldsymbol{B}'_i \mid \gamma'_i\right]\boldsymbol{Z}_i + \gamma_i$$

Multiplicando ambos lados por N y tomando el valor esperado:

(6.104) $$\sigma'_{y_i x} = \left[\boldsymbol{B}'_i \mid \gamma'_i\right]\Sigma_{Z_i x}$$

Si B'_i y γ'_i son funciones solamente de los elementos de las covarianzas de σ'_{yix} y Σ_{Zix}, ellos son identificados. Una condición necesaria consiste en que el número de ecuaciones en (6.104) sea al menos igual al número de parámetros libres desconocidos en $\left[\boldsymbol{B}'_i \mid \gamma'_i\right]$. El número de ecuaciones es el número de elementos en σ'_{yix}, esto es, q. De esto se sigue que q covarianzas con y_i resultan de q variables en x. El número de parámetros desconocidos en B'_i es (p-1) y en γ'_i es q. De esto se deduce que, con q ecuaciones en (p-1) + q desconocidos, B'_i y γ'_i no pueden ser identificados, entonces los (p-1) + q desconocidos deberán ser reducidas a q.

La condición de orden se define propiamente como "si las variables excluidas son solamente el tipo de restricciones, entonces las (p-1) variables deberán ser excluidas de la i-ésima ecuación para que sea posible la identificación" Bollen. Suponga el siguiente modelo:

(6.105) $$\begin{bmatrix} y_1 \\ y_2 \\ y_3 \end{bmatrix} = \begin{bmatrix} 0 & \beta_{12} & 0 \\ \beta_{21} & 0 & 0 \\ \beta_{31} & \beta_{32} & 0 \end{bmatrix}\begin{bmatrix} y_1 \\ y_2 \\ y_3 \end{bmatrix} + \begin{bmatrix} \tau_{11} & 0 \\ 0 & \tau_{22} \\ 0 & 0 \end{bmatrix}\begin{bmatrix} x_1 \\ x_2 \end{bmatrix} + \begin{bmatrix} \gamma_1 \\ \gamma_2 \\ \gamma_3 \end{bmatrix}$$

Y suponga que en la primera ecuación β_{13} y γ_{12} no sean restringidas a cero, entonces:

(6.105.1) $$y_1 = \boldsymbol{B}_{12}y_2 + \boldsymbol{B}_{13}y_3 + \tau_{11}x_1 + \tau_{12}x_2 + \gamma_1$$

El cual tiene la misma forma que (6.103). Multiplicando ambos lados por las variables exógenas y tomando valores esperados:

(6.105.2) $$\begin{aligned} \operatorname{cov}(y_1, x_1) &= \beta_{12}\operatorname{cov}(y_2, x_1) + \beta_{13}\operatorname{cov}(y_3, x_1) + \gamma_{11}\operatorname{var}(x_1) + \gamma_{12}\operatorname{cov}(x_2, x_1) \\ \operatorname{cov}(y_1, x_1) &= \beta_{12}\operatorname{cov}(y_2, x_2) + \beta_{13}\operatorname{cov}(y_3, x_1) + \gamma_{12}\operatorname{var}(x_2) + \gamma_{11}\operatorname{cov}(x_1, x_2) \end{aligned}$$

El resultado son dos ecuaciones para cuatro desconocidas (B_{12}, B_{13}, γ_{11}, γ_{12}) y es claro entonces que una única solución no es posible, la condición de orden requiere (p-1) o 2 exclusiones de la primera ecuación. La especificación de $\beta_{13}=0$ y $\gamma_{12}=0$ satisface este requerimiento, así que a partir de la primera ecuación, se encuentra una condición suficiente para la identificación.

Una forma para revisar la condición de orden para todas las ecuaciones en el modelo consiste en formar una matriz C, la cual es [(1-B)| -Γ]. Así, para cada fila se cuenta el número de ceros, si una fila tiene (p-1) o más ceros, esta es una condición de orden. Para el ejemplo anterior, tendremos:

$$(6.105.3) \quad C = \begin{bmatrix} 1 & -\beta_{12} & 0 & -\tau_{11} & 0 \\ -\beta_{21} & 1 & 0 & 0 & -\tau_{22} \\ -\beta_{31} & -\beta_{32} & 1 & 0 & 0 \end{bmatrix}$$

Cada fila en (6.105.3) tiene (p-1) o 2 exclusiones, así que se satisface la condición de orden. De igual forma la condición de orden nos brinda una regla para detectar sobreidentificación para modelos no recursivos. Con Ψ libres de sobreidentificación, ésta podrá ocurrir si es posible producir una nueva ecuación con la misma forma pero con parámetros diferentes de una ecuación anterior, a través de usar una combinación lineal de las otras ecuaciones en un modelo, esto ocurre cuando dos o más ecuaciones tienen restricciones idénticas, veamos:

$$(6.106) \quad \begin{aligned} y_1 &= \beta_{12} y_2 + \tau_{11} x_1 + \tau_{12} x_2 + \gamma_1 \\ y_2 &= \beta_{21} y_1 + \tau_{22} x_2 + \gamma_2 \end{aligned}$$

Suponga que se excluye x_2 de ambas ecuaciones ($\tau_{12} = \tau_{22} = 0$), entonces la matriz C será:

$$(6.106.1) \quad \begin{bmatrix} \mathbf{1} & -\beta_{12} & -\tau_{11} & \mathbf{0} \\ -\beta_{21} & \mathbf{1} & \mathbf{0} & \mathbf{0} \end{bmatrix}$$

La condición de orden para ambas ecuaciones consiste en hallar si cada ecuación tiene una exclusión (p–1). Multiplicando la segunda fila de (6.106.1) por una constante a y al sumar este resultado a la primera fila, se obtiene:

$$(6.106.2) \quad \begin{bmatrix} 1-\beta_{21}a & -\beta_{12}+a & -\tau_{11} & 0 \\ -\beta_{21} & 1 & 0 & 0 \end{bmatrix}$$

Dividiendo cada elemento de la primera fila por (1-aβ_{21}) se obtiene:

$$(6.106.3) \quad C^* = \begin{bmatrix} 1 & -\beta_{12}^* & -\tau_{11}^* & 0 \\ -\beta_{21} & 1 & 0 & 0 \end{bmatrix}$$

Donde $\beta^{*}_{21} = \left(\frac{-\beta_{12} + a}{1 - a\beta_{21}}\right) y - \tau^{*}_{11} = \frac{-\tau_{11}}{(1 - a\beta_{21})}$. La primera ecuación representada por la primera fila de C* en (6.106.3) tiene la misma forma y las mismas variables excluidas que C en (6.105.3). De esta forma, existirá un infinito conjunto de valores para β^{*}_{12} y de τ^{*}_{11} pero que no son iguales al verdadero B_{12} y τ_{11}, es por esta razón que B_{12} y τ_{11} no son identificados incluso cuando se satisface la condición de orden. Este procedimiento es sencillo con dos ecuaciones, pero en sistemas complejos multiecuacionales no es tan viable. Para determinar la regla del rango con C { [(I-B)| -Γ] } revise la identificación para la i-ésima ecuación, borre todas las columnas de C que no tengan ceros en la i-ésima fila de C y use las columnas que quedan para construir una nueva matriz C. Y de esta forma, una condición suficiente y necesaria para la identificación de la i-ésima ecuación consistirá en que el rango de C_i sea igual a p-1. Esta es la condición de rango de Bollen.

A manera de ilustración considere (6.106) y la C matriz es (6.106.1) y examinemos la identificación de la primera ecuación: solamente existen ceros en la primera fila en la cuarta columna, de esta forma se borran las 3 columnas y C_i queda como:

(6.106.3.1) $$C_1 = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$$

Dado que el rango de una matriz o vector es el número de filas independientes y columnas, con ambos elementos de C_1 cero, su rango será cero; con un rango menor que uno, la primera ecuación no es identificable como se demostró anteriormente. Para la segunda ecuación de C en (6.106.1) C_2 será:

(6.106.3.2) $$C_2 = \begin{bmatrix} -\gamma_{11} & 0 \\ 0 & 0 \end{bmatrix}$$

Excepto cuando γ_{11} es cero, el rango de C_2 será uno, lo cual satisface la condición de rango y entonces la segunda ecuación es identificada.

Haavelmo (1953) desarrolla un modelo de propensión marginal a consumir, esto es, la parte del ingreso disponible para comprar bienes de consumo. De esta forma, si este ingreso es igual a los gastos en inversión (x_1), a un gasto en consumo y_2 y los gastos en consumo están en función del ingreso disponible y_1 más un término de error γ, el sistema de dos ecuaciones será:

(6.107) $$\begin{aligned} y_1 &= y_2 + x_1 \\ y_2 &= \beta_{21} y_1 + \gamma_2 \end{aligned}$$

En términos matriciales:

$$\text{(6.107.1)} \quad \begin{bmatrix} y_1 \\ y_2 \end{bmatrix} = \begin{bmatrix} 0 & 1 \\ \beta_{21} & 0 \end{bmatrix} \begin{bmatrix} y_1 \\ y_2 \end{bmatrix} + \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} x_1 & x_2 \end{bmatrix} + \begin{bmatrix} 0 \\ 1 \end{bmatrix} \begin{bmatrix} 0 & \gamma_2 \end{bmatrix}$$

La matriz C será:

$$\text{(6.107.2)} \quad C = \begin{bmatrix} 1 & -\beta_{12} & -\tau_{11} & 0 \\ -\beta_{21} & 1 & 0 & 0 \end{bmatrix}$$

La matriz, C_2 será:

$$\text{(6.107.3)} \quad C_2 = \begin{bmatrix} -1 \\ 0 \end{bmatrix}$$

El rango es uno, lo cual satisface la condición de rango, por lo tanto es identificada.

6.10.3.5. Resumen de las reglas de identificación

Las reglas de identificación para ecuaciones estructurales con variables observadas asumiendo que no existen errores de medición, esto es $y = By + \Gamma_x + \gamma$, será:

Reglas de identificación	Evalúa	Requisitos	Condición necesaria	Condición suficiente
t-regla	Modelo	$t \le (1/2)(p+q)(p+q+1)$	Sí	No
Regla B nulo	Modelo	B = 0	No	Sí
Regla recursiva	Modelo	B triangular, Ψ diagonal	No	Sí
Condición de Orden	Ecuación	Restricción $\ge$ p-1, Ψ libre	Sí	No
Condición de rango	Ecuación	Rango (C_i) = p-1, Ψ libre	Sí	Sí

6.10.4. Estimación

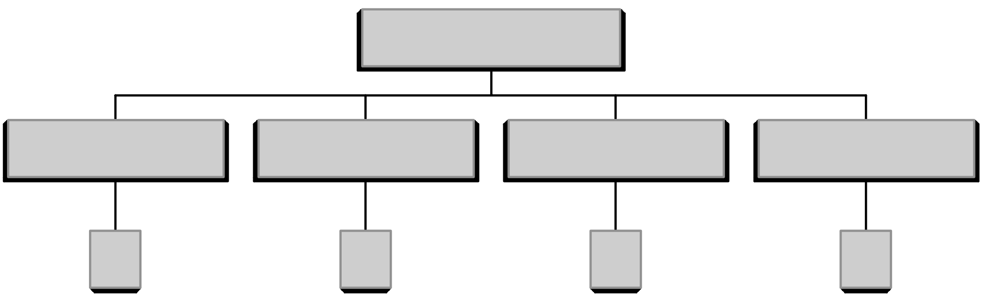
El procedimiento de estimación se deriva de la relación de la matriz de covarianzas de las variables observadas a los parámetros estructurales. De esta forma:

$$\text{(6.108)} \quad \sum(\theta) = \begin{bmatrix} (I-B)^{-1}(\Gamma\Phi\Gamma' + \Psi)(I-B)^{-1'} & (I-B)^{-1}\Gamma\Phi \\ \Phi\Gamma'(I-B)^{-1'} & \Phi \end{bmatrix}$$

Si el modelo de ecuaciones estructurales es correcto y los parámetros de la población son conocidos, entonces Σ será igual a $\Sigma(\theta)$. La forma de estimar el modelo de ecuaciones estructurales se realiza a través de una función de máximo verosimilitud, donde dicha función se ajusta maximizando:

(6.109) $$F_{ML} = Log\,|\,\Sigma(\theta)| + traza\left(S\,\Sigma^{-1}(\theta)\right) - Log\,|\,S\,| - p + q$$

Donde S es la matriz de covarianzas para y_i y x_i. A través de este procedimiento se obtendrán estimaciones maximoverosímiles. Mora (1997), supone el siguiente modelo: Supóngase que el paisaje rural es una variable latente. Debido a que existen diferentes características que determinan un paisaje supondremos que este tiene cuatro indicadores principales, como se observa en la siguiente gráfica:



Con $$\begin{aligned} P_1 &= \lambda_{11}\xi_1 + \delta_1 \\ P_2 &= \lambda_{21}\xi_1 + \delta_2 \\ P_3 &= \lambda_{31}\xi_1 + \delta_3 \\ P_4 &= \lambda_{41}\xi_1 + \delta_4 \end{aligned}$$

(6.110) $$p_i^* = \Lambda_p \xi_p + \delta \;;$$
$$\varepsilon(\delta_i) = 0; cov(\xi_i, \lambda_i) = 0 \forall : i..n; \delta_{iid}(0, \sigma_{jj}^2)$$

Donde el paisaje p^*_i es la variable latente y ξ_i es el verdadero paisaje. El P indicador del paisaje en P^*_i sirve como indicador de la variable latente ξ_i, el verdadero paisaje. Sin pérdida de generalidad, si el indicador es centrado alrededor de cero, de tal forma que P^*_i tiene un valor extremo, con los parámetros consistentes de los elementos de λ, la varianza de P^*_i, Φ, y la varianza del error. Tomando segundos momentos a ambos lados de la ecuación (6.110)

(6.111) $\sum(\theta) = E(PP') = E(\Lambda_p P + \delta)(P'\Lambda_p + \delta_l) = \Lambda_p \Phi \Lambda'_p + \Theta_\delta$

(6.112) Cov(ξ_i) = $\Lambda_p \Phi \Lambda'_p + \Theta_\delta = \sum(\theta)$

Donde $\Theta_\delta = \begin{bmatrix} \delta_1 & 0 & 0 & 0 \\ 0 & \delta_2 & 0 & 0 \\ 0 & 0 & \delta_3 & 0 \\ 0 & 0 & 0 & \delta_4 \end{bmatrix}$; $\bar{\lambda} = \begin{bmatrix} \lambda_{11} \\ \lambda_{21} \\ \lambda_{31} \\ \lambda_{41} \end{bmatrix}$; $\Phi = (\Phi_{11})$

Siguiendo a Bollen (1989), el modelo anterior puede ser identificado siempre que t $\leq$ 1/2(p)(p+1), es decir, t $<$ 10 factores desconocidos. A través de (6.111) se obtiene:

(6.113) $\sum(\theta) = \begin{bmatrix} \lambda_{11}{}^2\Phi_{11} + Var\delta_1 & & & \\ \lambda_{21}\lambda_{11}\Phi_{11} & \lambda_{21}{}^2\Phi_{11} + Var\delta_2 & & \\ \lambda_{31}\lambda_{11}\Phi_{11} & \lambda_{31}\lambda_{21}\Phi_{11} & \lambda_{31}{}^2\Phi_{11} + Var\delta_3 & \\ \lambda_{41}\lambda_{11}\Phi_{11} & \lambda_{41}\lambda_{21}\Phi_{11} & \lambda_{41}\lambda_{31}\Phi_{11} & \lambda_{41}{}^2\Phi_{11} + Var\delta_4 \end{bmatrix}$

En (6.113) existen 10 elementos desconocidos, escribiendo cada elemento en la matriz de covarianzas P1, P2, P3, P4 en términos de sus correspondientes parámetros estructurales y haciendo $\lambda_{11} = 1$ para escalar ξ_1 y poder identificar $\Sigma(\theta)$, se encuentran 9 ecuaciones con 9 elementos desconocidos, lo que lleva a una solución para los parámetros desconocidos. Con este resultado, se calcula (6.113). De Saris (1978) se conoce además:

(6.114) $P_i^* = \Phi \Lambda_p{}' \sum^{-1} P_i$

Se podrá observar además que $P_{ji(j=1..k)}$ es un indicador imperfecto de p^*_i si $\sigma_{jj} \rangle 0$ y si $\sigma_{jj} = 0$ $P_{ji(j=1..k)}$ es un indicador perfecto de p^*_i. Entonces, obtenemos (6.114) como una aproximación del paisaje y su función de verosimilitud vendrá dada por:

(6.115) $likelh = \left[\left|\sum(\theta)\right|\right]^{-\frac{N}{2}} e^{-\frac{N}{2} trace\left[S_{WW} \sum(\theta)^{-1}\right]}$

Donde $S_{ww} = \dfrac{P'P}{N}$. Para estimar (6.115) se requiere estimar (6.113) a través de máxima verosimilitud; los resultados encontrados usando una encuesta sobre paisaje, que se describe en el capítulo 9, fueron:

$$\lambda(i) = \begin{bmatrix} 1 \\ 22,8316 \\ 0,467748 \\ 1,596955 \end{bmatrix} \quad \Phi_{11} = [-0.00026] \quad \Theta_\delta = \begin{bmatrix} 0.01349 & & & \\ 0 & 0.152026 & & \\ 0 & 0 & 0.012205 & \\ 0 & 0 & 0 & 0.019107 \end{bmatrix}$$

$$\sum(\theta) = \begin{bmatrix} 0.013288 & & & \\ -0.00596 & 0.015987 & & \\ -0.00012 & -0.00279 & 0.012148 & \\ -0.00042 & -0.00952 & -0.00019 & 0.018442 \end{bmatrix}$$

Bollen (1989), provee el siguiente contraste de sobreidentificación:

$$\text{(6.116)} \quad FML\left(s\hat{\sum}(\theta)\right) = -Ln\left|\sum(\theta)\right| + Tr\left(s\sum^{-1}(\theta)\right) - Log\,|\,s\,| - q$$

Y $FML \approx \chi^2_{q,gl}$ donde $Ho = \sum = \sum(\theta)$ MAIC = Min AIC. El resultado encontrado fue $FML = 0.389563$; dado este valor, bajo H_0 no existe sobreidentificación, esto es, no existe suficiente evidencia estadística para rechazar H_0.

Bibliografía

AIGNER, D.J AND GOLDBERGER. (1977). Latent variables in socioeconomic models, North Holland.

AMEMIYA, T. (1994). Introduction to statistics and econometrics, Harvard University Press.

———. (1981). "Qualitative response models: a survey", Journal of economics history, vol. XIX, pp. 1483-1536.

———. (1974). " Tobit models : A Survey ", Journal of econometrics, 24, pp- 3-63.

ANDERSON, G.J. (1987). "Prediction test in limited dependent variable models", Journal of econometrics, 34, pp.253-61.

BERA, A.K., JARQUE, C.M. Y L.F, LEE. (1984). "Testing the normality assumption in limited dependent variable models", International economic review, 25, pp.563-78.

BLUNDELL, R. AND C, MEGHIR. (1986). "Selection criteria for microeconomic model of labor supply", Journal of applied econometrics, 1, pp.55-82.

BOLLEN, K. (1989). Structural equations with latent variables, John Wiley Sons.Inc.

BRANNAS, K AND T, LAITILA. (1989). "Heterocedasticity in the tobit model", Statistical papers, 30, pp.185-196.

BREUSH, T.S AND A.R, PAGAN. (1980). "The lagrange multiplier test and its applications to model specification in econometrics", Review of economic studies, 47, pp. 239-53.

BYRON, R.P. (1969) " A simple model for estimation demand systems under separable utility assumptions". Review of economic studies, pp. 261-274.

CESARIO, F.J. AND J. L, KNETSH. (1970). "Time bias in recreation benefit estimates", Water resources research, 6, pp.700-704.

CHAMBERS, E.A AND COX, D.R. (1967)."Discrimation between alternative binary response models". Biometrika. Vol 4, nums (3-4), pp. 573-78.

CHAMBERLAIN, G. (1980). "Analysis of covariance with qualitative data". Review of economic studies, Jan, pp. 225-38.

CHESHER, A.D. (1984). "Testing for neglected hetoregeneity", Econometrica, vol. 52, pp.865-72.

———, LANCASTER, T. AND M. IRISH. (1985). "On detecting the failure of distributions assumptions", Annals de L´INSEE, 59, pp. 7-44.

COX, D.R. (1983). "Some remarks on overdispersion", Biometrika , 70,pp.269-74.

DAVIDSON, R AND J.G, MACKINNON. (1989). "Testing for consistency using artificial regressions" Econometric Theory, 5, pp.363-84.

————.(1984). "Convenient specification test for Logit and Probit models", Journal of econometrics, 25, pp. 241-62.

DOMENCICH, T.A AND D, MCFADDEN. (1975). Urban travel demand, Amsterdam, North holland.

DUDLEY, L. AND MONTMARQUETLE, C. (1976). " A model of the supply of bilateral foreign aid" American economic review, March pp.132-42.

DURBIN, J. (1954). "Errors in variables", Review of the international statistical institute, 22, pp.22-32.

EYE VON, A AND C, CLOGG. (1994). Latent variable analysis: Applications for development research, Sage Publications.

GOLDBERGER, A.S. (1983). "Abnormal selection bias" en Karlin, Amemiya y Goodman (comps.), Studies in econometrics, Time series and multivariate analysis, New York, Academic Press.

GAURIÉROUX, C.A, MONFORT, E.R Y A, TROGNON. (1987). "Generalized residuals", Journal of econometrics, 34, pp.5-32.

GREEN, W.H. (1999). Análisis econométrico, Tercera edición, Prentice Hall Iberia.

GROGGER, C. (1991). "Model for truncated counts", Journal of applied econometrics, vol. 6, pp. 225-38.

GROGGER, J. (1990). "A simple test for exogeneity in probit, logit and poisson regression models", Economics letters, 33, pp. 329-32.

GRONAU R. (1976). "The allocation time of Israel women". Journal of Political economic, Aug, pp.201-20.

HAUSMAN, J. (1978). "Specification test in econometrics", Econometrica, vol.46, pp.1251-71.

HALL, W.J AND D.J, MATHIASON. (1990). "On large sample estimation and testing in parametric models", International statistical review, 58, pp.77-97.

HARVEY, A. (1976). "Estimating regression models with multiplicative heteroskedasticity". Econometric, 44, pp. 461-465.

HECKMAN, J. (1979). "Sample selection bias as a specification error", Econometrica, 47, pp.153-61.

———-.(1998). "Detecting discrimination", The journal of economic perspectives, vól 12, num.2, pp. 101-18, spring.

———. (1974). "Shadow prices, market wages and labor supply", Econometrica, 42, pp.679-94.

HOFFMAN, D AND A.R, PAGAN. (1989). "Post-sample prediction test for generalized method of moments estimators", Oxford bulletin of economics and statistics, 51, pp. 333-44.

HOROWITZ, J.L AND G.R, NEUMAN. (1989). "Specification testing in censored regression models: parametric and semiparametric methods", Journal of applied econometrics, 4, pp.s61-s86.

KEALY, M.J AND R.C, BISHOP. (1986). "Theoretical and empirical specifications issues in travel cost demand studies", American journal of agricultural economics, august, pp. 660-67.

KIEFER, N.M. (1982). "Testing for dependence in multivariate Probit models", Bometrika, 69, pp.161-6.

LANCASTER, T. (1984). "Test of specification in econometrics", Econometrics review, 3, pp.211-42.

LEE, L.F. (1978). "Unionism and wage rates : A simultaneous equations model with qualitative and limited dependent variables. International economic review Jun, pp. 415-34.

MCDONALD, J. A, AND D.A CLELLAND. (1984). "Textile workers and union sentimen". Social Forces 63; pp.502-521.

MCFADDEN, D. (1974). "Conditional logit analysis of qualitative choice behavior" in fronters in econometrics edited by P. Zarembka, New York: Academic Press, pp. 105-42.

———-.(1981). "Econometric models of probabilistic choice", in structural analysis discrete date, edited by C.F Mansky and D. McFadden. Cambridge Mass: Mit Press.

MACKINNON, J.G. (1992). "Model specification test and artificial regressions", Journal of economic literature, vol. 30, pp.102-46.

MACREALY, G AND C.M, DAYTON. (1994). "Latent class model for longitudinal assessment of trait acquisition", en Alexander Von Eye, Clifford C Clogg (comps.), Latent variable analysis: application for development research.

MADDALA, G.S. (1995) "Specification test in limited dependent variable models" en Phillips, P.C.B.,

NELSON, F. "Censored regression models with unobserved stochastic censoring thresholds". Econometric 6, pp.309-327.

NELSON, F Y L.. OLSON. "Specification and estimation of a simultaneous equation model with limited dependent variables". International economic review, Oct , pp.695-710.

PAGAN, A. R AND PARK, Y. (1993). "Testin for heteroskedasticity in G.S maddala". C.R Rao an H.D vinod (eds), Handook of Statistic Vol. 11 Amsterdam : North-Holland, pp.489-518.

PENCAVEL, J.H. (1979). "Market work decisions and unemployment of husbands and wives in the seatle and Denver income maintenance experiments". Mimeographed. April.

RAO, C.R AND MITRA, S.K. (1971). "Generalid inverses of matrices and its applications, New York : Wiley.

SRINVASAN, T.N (comps.), Advances in econometrics and qualitative variables, Basil Blackwell.

——— (1994) Econometrics methods and applications, Volume II, Edward Edgar Publishing.

——— (1983) Dependent and qualitative variables in econometrics, Cambridge, Cambridge University Press.

——— PHILLIPS, P.C.B., SRINVASAN, T.N (1995). Advances in econometrics and qualitative variables, Basil Blackwell.

———, AND F.D, NELSON. (1975). "Specification errors in limited dependent variable models" NBER Working paper series, No 96.

———, AND L.F, LEE. (1985). "The common structure of test for selectivity bias, serial correlation, heterocedasticity and non-normality in the Tobit model", International economic review, 26,pp.1-20

MELINO, A. (1982). "Testing for sample selection bias ", Review of economic studies, 49, pp.151-3.

MORA, J.J. (1997). "Aspectos microeconómicos del paisaje", Boletín Socioeconómico, núm. 30, pp.81-97.

——— (2001). El efecto de las características socioeconómicas sobre la consistencia en la toma de decisiones: Un análisis experimental, Borradores de Economía y Finanzas, 01, mayo, Universidad Icesi.

NAWATA, K. (1993a). "A note on the estimation of models with sample selection biases", Economic letters, 42, pp.15-24.

——— (1993b). "Estimation of sample selection biases models", Manuscript, Dept of economics, University of Western, Australia.

NEWEY, W.K. (1985). "Maximum likelihood specification testing and conditional moment test", Econometrica, vol. 53, pp.1047-73.

——— (1987). "Specification test for distributional assumptions in the Tobit model", Journal of econometrics, vol. 34, pp.125-45.

OLSEN, R. (1982). "Distributional test for selectivity bias and more robust likelihood estimator", International economic review, 23, pp. 233-40.

ORME, R.J. (1990) "The small-sample performance of the information matrix test", Journal of econometrics, 46, pp.309-31.

——— (1992). "Efficient score test for heterocedasticity in microeconometrics", Econometrics review, 11, pp. 235-52.

PAGAN, A.R AND F, VELLA. (1989). "Diagnostic test for models based on unit record data: a survey", Journal of applied econometrics, 4, pp. 175-94.

———, AND Y, PARK. (1993). "Testing for heterocedasticity " en G.S.Maddala., C.R. Rao and H. D. Vinod (comps.) Handbook of statistic, vol.11, Amsterdam: North-Holland, pp.489-518.

POWELL, J.L. (1986). "Symmetrically trimmed least squares estimation for Tobit models", Econometrica, vol. 54, pp.1435-60.

PREGINON, D. (1980). " Goodness of link test for generalized linear models", Applied statistics, 29, pp.15-24.

RAO, C.R. (1947) "Large sample test of statistical hypothesis concerning several parameters with applications to problems of estimation", Proceedings of the Cambridge philosophical society, 44, pp.50-7.

——— (1973). Linear statistical inference and its applications, New York, John Wiley and Sons.

——— AND MITRA, S.K. (1971). Generalized inverses of matrices and its applications, New York, John Wiley and Sons.

RUUD, P.A. (1984). "Test of specification in econometrics", Econometric reviews, 3, pp.211-42.

SCHMID, P. AND STRAUSS, R.P (1975). " The prediction of occupation using multiple logit model". International economic review; June, pp. 47-86.

SILVEY, S.D. (1959). "The lagrangian multiplier test", Annals of mathematical statistics, núm 7.pp.389-407.

SKEELS, C.L AND F, VELLA. (1993). The performance of conditional moment test in Tobit and Probit models. ANU, Camberra.

SMITH, R. (1989). "On the use of distributional misspecification checks in limited dependent variable models", Economic journal, supplement, 99, pp.178-92.

———, AND BLUNDELL, R. (1986) "An exogeneity test for simultaneous equation Tobit model with application to labor supply", Econometrica, vol. 54, pp. 679-85.

WEDDERBURN. (1976). "Quasi-likelihood functions, generalized linear models, and the gauss newton method", Biometrika, vol.61, num. 3, pp. 439.

WESTIN, R.B AND GUILLEN, D.W. (1978)." Parking location and transit demand : A case study of endogenous attributes in disaggregate mode choice functions". Journal of econometric, 8, pp. 75-101.

TAYLOR, L. (1991). "Testing exclusion restricctions for misspecified Tobit model", Economics letters, 37, pp.411-16.

TAUCHEN, G. (1985). "Diagnostic testing and evaluation of maximum likelihood models", Journal of econometrics, 30, pp.415-43.

TUKEY, J.W. (1949). "One degree of freedom for non-additivity", Biometrics, 5, pp. 232-242.

WHITE, H. (1982). "Maximum likelihood estimation of misspecified models", Econometrica, vol 50, pp.1-25.

WU, D.M. (1973). "Alternative test of independence between stochastic regressors and disturbances", Econométrica, vol.41, pp.134-39.