

MODELOS DE ELECCION DISCRETA

Revisión y aplicación mediante cuadratura gaussiana

Tesis presentada por

Elio Fabio Sánchez Trujillo
Douglas Gómez Cabrera

para optar al título de
Magíster en Matemáticas Aplicadas

Director

Francisco Iván Zuluaga Díaz
Magíster Matemáticas Aplicadas

Departamento de Ciencias Básicas

Universidad Eafit

Medellín

2008

Nota de aceptación

Jairo Villegas G.
Coordinador de la Maestría

Francisco Zuluaga D.
Director del trabajo

Medellín, Sept. de 2008

Agradecemos a nuestro Director Francisco Iván Zuluaga Díaz, Magíster en Matemáticas Aplicadas, por su dedicación y paciencia, sugerencias y apoyo constante durante este trabajo. Sin su fundamental colaboración este proyecto no hubiese sido posible.

Existen muchas personas que incidieron de una u otra manera para que este proceso de cualificación a nivel de Maestría lo pudiera culminar con éxito. Agradezco de manera especial al amigo y profesor Jairo Villegas, Coordinador de la Maestría, como también a los profesores Gustavo Mejía y Mauro Montealegre.

A la Universidad Eafit, a la Universidad de la Amazonia, que con su iniciativa y apoyo nos permitió cualificarnos a este nivel de conocimiento.

A mi esposa, Graciela, a mis hijas, Karina, Danna y Mellissa, que sin su apoyo, comprensión, Cariño, amor y acompañamiento no hubiera sido posible terminar este proceso, a ellas que son una bendición de Dios, todo mi amor.

Douglas Gómez Cabrera.

Florencia, Caquetá
Septiembre, 2008

Tuve la fortuna de conocer durante las jornadas académicas de esta Maestría a mucha gente brillante y colaboradora como nadie, es allí donde siempre habrá un deuda de gratitud con un gran maestro como es Jairo Villegas. Debo mencionar a docentes como Gustavo Mejía y Mauro Montealegre, pues su apoyo y colaboración me comprometieron a concluir esta tarea.

A la Universidad de la Amazonia, que con su apoyo nos permitió cualificarnos a este nuevo nivel para seguir trabajando por el progreso y desarrollo de nuestra región.

Por supuesto, mantengo una deuda infinita con mi esposa, Carmen Rossy, y mis hijos, Fabio Andrés y Juan José, a quienes les robé muchos días y noches para concluir este sueño. Ellos son **mi Fuerza, mi Luz, mi Amor y mi Razón.**

A Dios y a mi Padre en el seno celestial. A mi Madre, a quien debo Todo, desde el comienzo. (En el más amplio sentido de la palabra.)

Elio Fabio Sánchez Trujillo.

Florencia, Caquetá
Septiembre, 2008

Índice general

Introducción	1
1. Fundamentos	3
1.1. Definiciones Básicas	3
1.2. Consistencia	12
1.3. Normalidad Asintótica	15
2. Modelos de Elección Binaria	17
2.1. Generalidades	17
2.1.1. Introducción	17
2.1.2. El Proceso de Elección	24
2.1.3. Teoría de la Utilidad Aleatoria	25
2.1.4. Maximización de la Utilidad Aleatoria (RUM)	29
2.2. Especificación del modelo	32
2.3. Estimación para modelos de elección binaria	33
2.3.1. Consistencia y normalidad asintótica del Estimador de Máxima Verosimilitud	33
2.3.2. Métodos Iterativos para obtener el Estimador de Máxima Verosimilitud	38
2.3.3. Método del Mínimo Chi-cuadrado de Berkson	39
2.3.4. Comparación entre el Estimador de Máxima Verosimilitud y el Estimador Chi-cuadrado Mínimo	44
2.4. Prueba de Hipótesis	45
2.5. Análisis Discriminante	47
2.6. Predicción Agregada	50
3. Modelos de Respuesta Múltiple	52
3.1. Inferencia Estadística	52
3.2. Análisis Discriminante Multinomial (DA)	60
3.3. Principales Modelos de Respuesta Múltiple	61
3.3.1. Introducción	61

3.3.2. Modelos de Respuesta Múltiple Ordenado	64
3.3.3. Modelos de Respuesta Múltiple No Ordenados	69
3.3.4. Definición	69
3.3.5. Modelo Logit Multinomial	71
3.3.6. Modelo Logit Jerárquico o Anidado (NL)	77
3.3.7. Modelo Probit Multinomial (MNP)	85
4. Aplicación	88
4.1. Introducción	88
4.2. Fuente de datos	88
4.2.1. Caso Estados Unidos	88
4.2.2. Caso Colombia	90
4.3. Metodología	91
4.3.1. Cuadratura Gaussiana	91
4.3.2. Criterio	93
4.4. Resultados	93
4.4.1. Caso Estados Unidos	93
4.4.2. Caso Colombia	94
A. Función de Verosimilitud	96
A.1. Condiciones de regularidad	96
B. Aplicación	97
B.1. Caso EE.UU.	97
B.2. Caso Colombia	101
Bibliografía	105

Resumen

El presente trabajo titulado *MODELOS DE ELECCION DISCRETA Revisión y aplicación mediante cuadratura gaussiana* recoge los fundamentos característicos de estos modelos de gran uso en Microeconometría. Se describen las características de los Modelos de Elección Binaria y de los Modelos de Elección Múltiple o Multinomiales. Por último, se presenta una aplicación práctica de la dinámica laboral de mujeres casadas en Estados Unidos y en Colombia usando Cuadratura Gaussiana. Los modelos estimados son altamente significativos en sus parámetros y el procedimiento de cuadratura gaussiana es estable y confiable.

Palabras Claves: Elección Discreta, Logit, Probit, Estimador de Máxima Verosimilitud, Cuadratura Gaussiana.

This work titled *MODELOS DE ELECCION DISCRETA Revisión y aplicación mediante cuadratura gaussiana* gathers the characteristic foundations of this models which are useful in Microeconometrics. Characteristics of Binary Choice and Multinomial Choice Models are described. Lastly, a practical application about labor dynamic of married women, both, in United States and Colombia using Gaussian Quadrature is presented. Estimated models are highly significant in their parameters and the procedure of Gaussian Quadrature is stable and reliable.

Keywords: Discret Choice, Logit, Probit, Maximum Likelihood Estimator, Gaussian Quadrature.

Introducción

La Economía como ciencia social preocupada por el estudio de la optimización que hacen los agentes de los recursos escasos, está sujeta a los permanentes procesos de decisión que se presentan en la vida cotidiana tanto a nivel microeconómico como macroeconómico.

La Econometría usa herramientas de la Matemática y la Estadística para estudiar planteamientos económicos. En un sentido amplio, usa métodos estadísticos aplicados al análisis económico asumiendo datos agregados y conclusiones generalizadas sobre el comportamiento de las personas y/o firmas. Sin embargo una de las tareas más importantes en el estudio del comportamiento de los agentes económicos es buscar herramientas que permitan cuantificar el problema de una forma objetiva y racional. ([2],[13])

A su vez, la Microeconometría nace por la necesidad de brindar una interpretación económica de las pautas del comportamiento de los agentes económicos o individuos, así como la posibilidad de contrastar estadísticamente las hipótesis efectuadas([15]). Desarrolla todo un conjunto de modelos, métodos y técnicas para el estudio de los recursos escasos dentro de las familias y empresas, y de cómo se emplean estos recursos en el ámbito de la toma de decisiones racionales de los agentes individuales. También facilita la tarea de identificar las características o factores que provocan un comportamiento diferente de los individuos ante los procesos de decisión.

Dado que lo que se intenta representar es el comportamiento del individuo, y éste se plasma en las decisiones que toma, los Modelos de Elección Discreta se han revelado como una de las herramientas más útiles y extendidas. Estos modelos, basados en la teoría de la maximización de la utilidad aleatoria, se han desarrollado entre la Econometría y la Ingeniería del Transporte, que constituye una de sus principales aplicaciones. Los Modelos de Elección Discreta permiten la modelación de variables cualitativas, mediante el uso de técnicas propias de las variables discretas. Una variable es discreta cuando está formada por un número finito de alternativas que miden cualidades. Esta característica exige la codificación como

paso previo a la modelación, proceso por el cual las alternativas de las variables se transforman en códigos o valores cuánticos, susceptibles de ser modelados utilizando técnicas econométricas.

Los Modelos de Elección Discreta se clasifican de acuerdo al número de alternativas incluidas en la variable endógena. Se distinguen los modelos de respuesta dicotómica frente a los denominados modelos de respuesta multinomial o elección múltiple. Según la función utilizada para la estimación de la probabilidad se habla de Modelo Logit o Modelo Probit. Si las alternativas de la variable endógena son excluyentes o incorporan información ordinal se habla de Modelos con datos no ordenados y Modelos con datos ordenados. Dentro de los primeros, según los regresores se refieran a aspectos específicos de la muestra o de las alternativas entre las que se ha de elegir, se distingue entre los modelos multinomiales y los condicionales.

El presente documento titulado ***MODELOS DE ELECCION DISCRETA Revisión y aplicación mediante cuadratura gaussiana***, recoge los fundamentos característicos de estos modelos. En el primer capítulo se fijan los elementos conceptuales esenciales y se presentan definiciones y fundamentos básicos de la teoría. El segundo capítulo se centra en características de los Modelos de Respuesta Binaria. En el tercer capítulo se introducen los conceptos referentes a los Modelos de Elección Múltiple, como extensión de los modelos dicotómicos. El último capítulo se refiere a la implementación concreta de una aplicación que toma como escenario el mercado laboral de mujeres a partir de datos de Estados Unidos y de Colombia. Se plantea un Modelo Discreto de Elección Binaria con el cual se busca explicar la probabilidad de que una mujer casada decida participar en el mercado laboral en estos países mediante variables explicativas comunes.

Capítulo 1

Fundamentos

1.1. Definiciones Básicas

Inicialmente se recuerdan algunas definiciones básicas necesarias para el posterior desarrollo del trabajo. En esta parte referimos a ([2]) y a ([30]) de donde se citan las siguientes definiciones:

Sea Ω un espacio muestral abstracto, se define la colección \mathcal{A} de subconjuntos de Ω que posee ciertas propiedades deseables.

Definición 1. *La colección \mathcal{A} de subconjuntos de Ω se llama un σ -álgebra si satisface las propiedades*

- $\Omega \in \mathcal{A}$
- $E \in \mathcal{A} \Rightarrow \bar{E} \in \mathcal{A}$. (\bar{E} se refiere al complemento de E con respecto a Ω .)
- $E_j \in \mathcal{A}, \quad j = 1, 2, \dots \Rightarrow \bigcup_{j=1}^{\infty} E_j \in \mathcal{A}$

Definición 2. *Una **medida de probabilidad**, denotada por $P(\cdot)$, es un conjunto de funciones de valor real que se define sobre una σ -álgebra \mathcal{A} y satisface las propiedades*

- $E \in \mathcal{A} \Rightarrow P(E) \geq 0$
- $P(\Omega) = 1$
- Si $\{E_j\}$ es una colección contable de conjuntos disjuntos en \mathcal{A} , entonces

$$P\left(\bigcup_j E_j\right) = \sum_j P(E_j)$$

Definición 3. Espacio de probabilidad Dado un espacio muestral Ω , una σ -álgebra \mathcal{A} asociada con Ω y una medida de probabilidad $P(\cdot)$ definida sobre \mathcal{A} , se llama espacio de probabilidad a la tripleta (Ω, \mathcal{A}, P)

Definición 4. Una variable aleatoria en (Ω, \mathcal{A}, P) es una función de valor real definida sobre un espacio muestral Ω , denotada $X(\omega)$ para $\omega \in \Omega$, tal que para cualquier número real x ,

$$\{\omega | X(\omega) < x\} \in \mathcal{A}$$

Definición 5. La Función de Distribución $F(x)$ de una variable aleatoria $X(\omega)$ se define como:

$$F(x) = P\{\omega | X(\omega) < x\} \quad (1.1.1)$$

La función de distribución puede definirse para cualquier variable aleatoria porque una probabilidad se asigna a cada elemento de \mathcal{A} y de

$$\{\omega | X(\omega) < x\}$$

para algún x . Se escribirá

$$P\{\omega | X(\omega) < x\}$$

como

$$P(X < x)$$

Una función de distribución tiene las propiedades

- $F(-\infty) = 0$
- $F(\infty) = 1$
- Es no decreciente y continua hacia la izquierda

Usando una función de distribución, se puede definir el valor esperado de una variable aleatoria si es discreta, continua o mixta. Esto se hace por medio de la integral de Riemann-Stieltjes que es una generalización sencilla de la integral de Riemann. Si X es una variable aleatoria con una función de distribución F y si

$$Y = h(X),$$

donde $h(\cdot)$ es medible en el espacio de Borel. El valor esperado de Y , se denota como $E(Y)$ y se define así: Se divide un intervalo $[a, b]$ en n intervalos con los puntos finales

$$a = x_0 < x_1 < \cdots < x_{n-1} < x_n = b$$

y si x_i^* es un punto arbitrario en $[x_i, x_{i+1}]$. Se define la suma parcial

$$S_n = \sum_{i=0}^{n-1} (x_i^*) [F(x_{i+1}) - F(x_i)] \quad (1.1.2)$$

asociado con esta partición del intervalo $[a, b]$. Si, para algún $\epsilon > 0$, allí existe un número real A y una partición tal que para cada partición más fina y para cualquier elección de x_i^* , $|S_n - A| < \epsilon$, se denominará A a la integral de Riemann-Stieltjes y se denota por:

$$\int_a^b h(x) dF(x)$$

Esta existe si h es una función continua excepto posiblemente para un número contable de discontinuidades, con tal de que, siempre que su discontinuidad coincida con que F , sea continua hacia la derecha. Finalmente, el valor esperado de Y , $E(Y)$ se define como:

$$E(Y) = \int_{-\infty}^{\infty} h(x) dF(x) = \lim \int_a^b h(x) dF(x) \quad (1.1.3)$$

con tal de que el límite (que puede ser $+\infty$ o $-\infty$) exista sin tener en cuenta que $a \rightarrow -\infty$ y $b \rightarrow \infty$.

Si $\frac{dF}{dx}$ existe y es igual a $f(x)$,

$$F(x_{i+1}) - F(x_i) = f(x_i^*)(x_{i+1} - x_i)$$

para algún $x_i^* \in [x_{i+1}, x_i]$ por el teorema del valor medio. Por tanto,

$$E(h(X)) = \int_{-\infty}^{\infty} h(x) f(x) dx \quad (1.1.4)$$

Por otro lado, suponiendo $X = c_i$ con probabilidad p_i , $i = 1, 2, \dots, K$ y tomando $a < c_1$ y $c_k < b$; entonces, para n suficientemente grande, cada intervalo contiene a lo máximo uno de los c_i 's. Entonces, de los n términos en el sumando de 1.1.2 sólo K términos contienen c_i 's no ceros. Por consiguiente

$$\int_a^b h(x) dF(x) = \sum_{i=1}^K h(c_i)p_i \quad (1.1.5)$$

A continuación se definen cuatro tipos de convergencia para una sucesión de variables aleatorias y se muestra la relación entre ellos en varios teoremas.

Definición 6. Convergencia en probabilidad Una sucesión de variables aleatorias $\{X_n\}$ se dice que converge a una variable aleatoria X en probabilidad si

$$\lim_{n \rightarrow \infty} P(|X_n - X| > \epsilon) = 0$$

para cualquier $\epsilon > 0$ También se escribe como:

$$X_n \xrightarrow{p} X$$

o como:

$$\text{plim } X_n = X$$

Definición 7. Convergencia en cuadrados medios Una sucesión $\{X_n\}$ se dice que converge a X en cuadrados medios si

$$\lim_{n \rightarrow \infty} E(X_n - X)^2 = 0.$$

También se escribe como:

$$X_n \xrightarrow{M} X$$

Definición 8. Convergencia en distribución Una sucesión $\{X_n\}$ se dice que converge a X en distribución si la función de distribución F_n de X_n converge a la función de distribución F de X en cada punto de continuidad de F . Se escribe

$$X_n \xrightarrow{d} X$$

y F se llama la **distribución límite** de $\{X_n\}$. Si $\{X_n\}$ y $\{Y_n\}$ tienen la misma distribución límite, se escribe:

$$X_n \xrightarrow{LD} Y_n$$

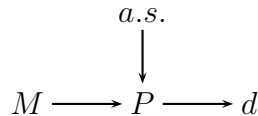
Definición 9. Convergencia casi segura Una sucesión $\{X_n\}$ se dice que converge a X casi seguramente si

$$P \left\{ \omega \mid \lim_{n \rightarrow \infty} X_n(\omega) = X(\omega) \right\} = 1$$

también se escribe como:

$$X_n \xrightarrow{a.s.} X$$

Los próximos cuatro teoremas establecen las relaciones lógicas entre los cuatro modos de convergencia, detallados en la siguiente figura,



Teorema 10. (Chebyshev)

$$E [X_n^2] \rightarrow 0 \Rightarrow X_n \xrightarrow{P} 0$$

Prueba.

$$E [X_n^2] = \int_{-\infty}^{\infty} x^2 dF_n(x) \geq \epsilon^2 \int_S dF_n(x) \quad (1.1.6)$$

donde

$$S = \{x | x^2 \geq \epsilon^2\}$$

pero se tiene que:

$$\begin{aligned} \int_S dF_n(x) &= \int_{-\infty}^{-\epsilon} dF_n(x) + \int_{\epsilon}^{\infty} dF_n(x) \\ &= F_n(-\epsilon) + [1 - F_n(\epsilon)] \\ &= P(X_n < -\epsilon) + P(X_n \geq \epsilon) \\ &\geq P[X_n^2 > \epsilon^2] \end{aligned} \quad (1.1.7)$$

Por consiguiente, de las definiciones (7) y (8) se obtiene que:

$$P[X_n^2 > \epsilon^2] \leq \frac{E(X_n^2)}{\epsilon^2} \quad (1.1.8)$$

que es el teorema que sigue inmediatamente de la definición (9) ■

La desigualdad (1.1.8) se llama la Desigualdad de Chebyshev. Modificando la prueba ligeramente, se puede establecer la siguiente generalización de dicha desigualdad:

$$P[g(X_n) > \epsilon^2] \leq \frac{Eg(X_n)}{\epsilon^2} \quad (1.1.9)$$

donde $g(\cdot)$ es una función continua no negativa.

Obsérvese que la declaración

$$X_n \xrightarrow{M} X \Rightarrow X_n \xrightarrow{P} X,$$

donde X puede ser una constante o una variable aleatoria, se sigue del teorema (10) si se considera $X_n - X$ como el X_n del teorema.

Teorema 11. $X_n \xrightarrow{P} X \Rightarrow X_n \xrightarrow{d} X$

Teorema 12. $X_n \xrightarrow{a.s.} X \Rightarrow X_n \xrightarrow{P} X$

Teorema 13. $X_n \xrightarrow{d} \alpha \Rightarrow X_n \xrightarrow{P} \alpha$

Teorema 14. Si \mathbf{X}_n y \mathbf{X} son K -vectores de variables aleatorias y si $g(\cdot)$ es una función de R^K a R tal que el conjunto E de puntos discontinuos de $g(\cdot)$ es cerrado y $P(\mathbf{X} \in E) = 0$. Si

$$X_n \xrightarrow{d} X,$$

entonces

$$g(X_n) \xrightarrow{d} g(X).$$

La convergencia en distribución de los elementos individuales del vector X_n para los correspondientes elementos del vector X no son suficientes para obtener los resultados anteriores. Sin embargo, si los elementos de X_n son independientes para cada n , la convergencia separada es suficiente.

Teorema 15. Sea X_n un vector de variables aleatorias con un número fijo finito de elementos. Si g es una función de valor real continua en un punto vectorial constante α . Entonces:

$$X_n \xrightarrow{P} \alpha \Rightarrow g(X_n) \xrightarrow{P} g(\alpha)$$

Teorema 16. Si $X_n \xrightarrow{P} X$ y $Y_n \xrightarrow{P} \alpha$, entonces:

1. $X_n + Y_n \xrightarrow{d} X + \alpha$
2. $X_n \cdot Y_n \xrightarrow{d} \alpha X$
3. $(X_n/Y_n) \xrightarrow{d} X/\alpha, \forall \alpha \neq 0$

Ley de los Grandes Números Dada una sucesión de variables aleatorias $\{X_t\}$, se define

$$\bar{X} = n^{-1} \sum_{t=1}^n X_t.$$

Una Ley de Grandes Números (LGN) especifica las condiciones bajo las cuales $\bar{X} - E(\bar{X})$ converge a 0 bien sea en forma casi segura ($\bar{X} - E(\bar{X}) \xrightarrow{a.s.} 0$) o en probabilidad ($\bar{X} - E(\bar{X}) \xrightarrow{P} 0$). La primera se conoce como una *ley fuerte de grandes números* y la última como una *ley débil*. En muchas aplicaciones la forma más simple de mostrar $\bar{X} - E(\bar{X}) \xrightarrow{P} 0$ es mostrar que $\bar{X} - E(\bar{X}) \xrightarrow{M} 0$ y luego aplicar Teorema 10 (Chebyshev). Se enuncian enseguida dos leyes fuertes de grandes números.

Teorema 17. (LGN 1 de Kolmogorov) Sea $\{X_t\}$ una sucesión de variables aleatorias independientes con varianza finita $\text{Var}(X_t) = \sigma_t^2$. Si $\sum_{t=1}^{\infty} \frac{\sigma_t^2}{t^2} < \infty$, entonces

$\bar{X}_n - E(\bar{X}_n)$ converge casi seguramente a cero, esto es, $\bar{X} - E(\bar{X}) \xrightarrow{a.s.} 0$, siendo

$$\bar{X}_n = \frac{1}{n} \sum_{t=1}^n X_t$$

Teorema 18. (LGN 2 de Kolmogorov) Sea $\{X_t\}$ una sucesión de variables aleatorias independientes e idénticamente distribuidas. Entonces una condición suficiente y necesaria para que $\bar{X} \xrightarrow{a.s.} \mu$ es que $E(X_t)$ exista y sea igual a μ .

Teorema 19. (Teorema del límite central) Sea $\{X_t\}$ una sucesión de variables aleatorias independientes e idénticamente distribuidas. Con $E(X_t) = \mu$ y $\text{Var}(X_t) = \sigma_t^2$ entonces

$$\frac{1}{\sqrt{\text{Var}(\bar{X}_n)}}(\bar{X}_n - E(\bar{X}_n)) \xrightarrow{d} N(0,1)$$

Relaciones entre $\lim E$, AE y plim Sea F_n la función de distribución de X_n y $F_n \rightarrow F$ en puntos de continuidad de F . En la Definición 6 se planteó $\text{plim } X_n$. Ahora se define $\lim E$ y AE como sigue

$$\lim E(X_n) = \lim \int_{-\infty}^{\infty} x dF_n(x) \tag{1.1.10}$$

y

$$AE X_n = \int_{-\infty}^{\infty} x dF_n(x) \tag{1.1.11}$$

O dicho de otra forma, AE que se lee *esperanza asintótica* o *media asintótica*, es la media de la distribución límite.

Definición 20. Sea $g_T(\theta)$ una sucesión no negativa de variables aleatorias dependiente de un vector de parámetros θ . Se consideran ahora los tres modos de convergencia uniforme de $g_T(\theta)$ a 0:

1. $P[\lim_{T \rightarrow \infty} \sup_{\theta \in \Theta} g_T(\theta) = 0] = 1$,
2. $\lim_{T \rightarrow \infty} P[\sup_{\theta \in \Theta} g_T(\theta) < \epsilon] = 1$ para algún $\epsilon > 0$,
3. $\lim_{T \rightarrow \infty} \inf_{\theta \in \Theta} P[g_T(\theta) < \epsilon] = 1$ para algún $\epsilon > 0$.

Si se cumple 1.), se dice que $g_T(\theta)$ converge a 0 uniformemente de manera casi segura en $\theta \in \Theta$. Si se cumple 2.), se dice que $g_T(\theta)$ converge a 0 uniformemente en probabilidad en $\theta \in \Theta$. Si se cumple 3.), se tiene que $g_T(\theta)$ converge a 0 semiuniformemente en probabilidad en $\theta \in \Theta$.

Teorema 21. Sean los supuestos:

1. El espacio de parámetro Θ es un subconjunto compacto del K -espacio Euclideo (R^K). (Recordar que θ_0 está en Θ)

2. $Q_T(y, \theta)$ es continua en $\theta \in \Theta$ para todo \mathbf{y} y es una función medible de \mathbf{y} para todo $\theta \in \Theta$,
3. $T^{-1}Q_T(\theta)$ converge a una función no estocástica $Q(\theta)$ uniformemente en probabilidad en $\theta \in \Theta$ cuando T tiende a infinito (∞), y $Q(\theta)$ alcanza un único máximo local a θ_0 . (La continuidad de $Q(\theta)$ se sigue de los supuestos).

$\hat{\theta}_T$ se define como un valor que satisface:

$$Q_T(\hat{\theta}_T) = \max_{\theta \in \Theta} Q_T(\theta)$$

[se entiende que si $\hat{\theta}_T$ no es único, se selecciona un valor apropiado de tal manera que $\hat{\theta}_T(y)$ sea una función medible de y .¹]. Entonces $\hat{\theta}_T$ converge en probabilidad a θ_0

Teorema 22. Sean los supuestos:

1. Sea Θ un subconjunto abierto para el K -espacio Euclideo (así el verdadero valor θ_0 es un punto interior de Θ)
2. $Q_T(y, \theta)$ es una función medible de y para todo $\theta \in \Theta$, y $\frac{\partial Q_T}{\partial \theta}$ existe y es continua en una vecindad abierta $N_1(\theta_0)$ de θ_0 . (Esto implica que Q_T es continua para $\theta \in N_1$)
3. Existe una vecindad abierta $N_2(\theta_0)$ de θ_0 tal que $T^{-1}Q_T(\theta)$ converge a una función no estocástica $Q(\theta)$ uniformemente en probabilidad en $\theta \in N_2(\theta_0)$, y $Q(\theta)$ alcanza un máximo local estricto en θ_0

Sea Θ_T el conjunto de raíces de la ecuación

$$\frac{\partial Q_T}{\partial \theta} = \mathbf{0} \tag{1.1.12}$$

que corresponden al máximo local. Si el conjunto es vacío, el conjunto Θ_T es igual a $\{\mathbf{0}\}$ Entonces, para algún $\epsilon > 0$,

$$\lim_{T \rightarrow \infty} P[\inf_{\theta \in \Theta_T} (\theta - \theta_0)'(\theta - \theta_0) > \epsilon] = 0$$

Teorema 23. Asumiendo supuestos del teorema anterior, y además:

1. $\frac{\partial^2 Q_T}{\partial \theta \partial \theta'}$ existe y es continua en una vecindad abierta convexa de θ_0

¹Esto es posible por un teorema de Jennrich (1969, p.637)

2. $T^{-1} \left(\frac{\partial^2 Q_T}{\partial \theta \partial \theta'} \right)_{\theta_T^*}$ converge en probabilidad a una matriz finita singular:

$$A(\theta_0) = \lim ET^{-1} \left(\frac{\partial^2 Q_T}{\partial \theta \partial \theta'} \right)_{\theta_T^*}$$

para alguna sucesión θ_T^* tal que:

$$\text{plim } \theta_T^* = \theta_0$$

3. $T^{-1/2} \left(\frac{\partial Q_T}{\partial \theta} \right)_{\theta_0} \rightarrow N[\mathbf{0}, \mathbf{B}(\theta_0)]$, donde:

$$\mathbf{B}(\theta_0) = \lim ET^{-1} \left(\frac{\partial Q_T}{\partial \theta} \right)_{\theta_0} \times \left(\frac{\partial Q_T}{\partial \theta'} \right)_{\theta_0}$$

Sea $\{\hat{\theta}_T\}$ una sucesión obtenida para escoger un elemento de Θ_T definido en el teorema anterior tal que:

$$\text{plim } \hat{\theta}_T = \theta_0.$$

($\hat{\theta}_T$ se denomina una raíz consistente). Entonces:

$$\sqrt{T}(\hat{\theta}_T - \theta_0) \rightarrow N[\mathbf{0}, \mathbf{A}(\theta_0)^{-1} \mathbf{B}(\theta_0) \mathbf{A}(\theta_0)^{-1}]$$

Teorema 24. Suponiendo que $g_T(\theta)$ converge en probabilidad a una función no estocástica $g(\theta)$ uniformemente en θ en una vecindad abierta $N(\theta_0)$, entonces $\text{plim } g_T(\hat{\theta}_T) = g(\theta_0)$ si $\text{plim}(\hat{\theta}_T) = \theta_0$ y $g(\theta)$ es continua en θ_0

Definición 25. Estimador de Máxima Verosimilitud Sea $L_T(\theta) = L(y, \theta)$ la densidad conjunta de un vector de T variables aleatorias $\mathbf{y} = (y_1, y_2, \dots, y_T)$ caracterizada por un vector de K parámetros θ . Cuando nos referimos a ella como una función de θ , la llamamos función de verosimilitud. El término Estimador de Máxima Verosimilitud (EMV) se usa frecuentemente para significar dos conceptos diferentes:

1. El valor de θ que maximiza globalmente la función de verosimilitud $L(y, \theta)$ sobre el espacio de parámetros θ ; o
2. Cualquier raíz de la ecuación de verosimilitud

$$\frac{\partial L_T(\theta)}{\partial \theta} = 0 \tag{1.1.13}$$

que corresponda a un máximo local. Aquí lo usaremos en el segundo sentido.

1.2. Consistencia

Las condiciones para la consistencia del *Estimador de Máxima Verosimilitud* (EMV) global o local se pueden obtener del Teorema 21 o 22 haciendo

$$Q_T(\theta) = \log L_T(\theta).$$

Se considera el logaritmo de la función de verosimilitud porque

$$T^{-1} \log L_T(\theta)$$

usualmente converge a una constante finita. Tomando el logaritmo no cambia la localización del máximo local o global.

Ahora, si se asume que y_i ($i = 1, 2, \dots, T$) son independientes e idénticamente distribuidos con función de densidad $f(\cdot, \theta)$, se puede escribir

$$\log L(y, \theta) = \sum_{i=1}^T \log f(y_i, \theta) \quad (1.2.1)$$

En este caso se puede reemplazar el supuesto 3 del Teorema 21 y 22 por los 2 siguientes supuestos:

$$E \left(\sup_{\theta \in \Theta} |\log f(y_t, \theta)| \right) < \infty \quad (1.2.2)$$

y

$$\log f(y_t, \theta) \text{ es una función continua de } \theta \text{ para cada } y_t. \quad (1.2.3)$$

En el Teorema 26 se muestra que los supuestos 1.2.2 y 1.2.3 implican

$$\text{plim } \frac{1}{T} \sum_{i=1}^T \log f(y_i, \theta) = E \log f(y_t, \theta) \text{ uniformemente en } \theta \in \Theta \quad (1.2.4)$$

Además, por desigualdad de Jensen se tiene que

$$E \left(\log \frac{f(y_t, \theta)}{f(y_t, \theta_0)} \right) < \log E \left(\frac{f(y_t, \theta)}{f(y_t, \theta_0)} \right) = 0 \text{ para } \theta \neq \theta_0 \quad (1.2.5)$$

donde la esperanza se toma usando el valor real θ_0 , y por tanto

$$E(\log f(y_t, \theta)) < \log E(f(y_t, \theta_0)) \text{ para } \theta \neq \theta_0 \quad (1.2.6)$$

Como en la ecuación 1.2.6 se tiene que

$$T^{-1} E(\log L_T(\theta)) < T^{-1} E(\log L_T(\theta_0))$$

para $\theta \neq \theta_0$ y para todo T. Sin embargo cuando se toma límite a ambos lados de la desigualdad 1.2.6 cuando T tiende a infinito, se tiene:

$$\lim_{T \rightarrow \infty} T^{-1} E(\log L_T(\theta)) \leq \lim_{T \rightarrow \infty} T^{-1} E(\log L_T(\theta_0))$$

Aquí, generalmente se necesita asumir que: $\lim_{T \rightarrow \infty} T^{-1} E(\log L_T(\theta))$ únicamente se maximice en $\theta = \theta_0$.

El hecho que los modelos 1.2.2 y 1.2.3 impliquen 1.2.4 se sigue del siguiente teorema cuando se hace

$$g_t(\theta) = \log f(y_t, \theta) - E(\log f(y_t, \theta))$$

Teorema 26.

Sea $g_t(y, \theta)$ una función medible de \mathbf{y} en un espacio Euclideano para cada $\theta \in \Theta$, un subconjunto compacto de R^K (K -espacio Euclideano) y una función continua de $\theta \in \Theta$ para cada \mathbf{y} . Asuma $E(g(y, \theta)) = 0$. Sea $\{\mathbf{y}_i\}$ una sucesión de vectores aleatorios independientes e idénticamente distribuidos tales que

$$E\left(\sup_{\theta \in \Theta} |g(y_t, \theta)|\right) < \infty.$$

Entonces $T^{-1} \sum_{t=1}^T g_t(y, \theta)$ converge uniformemente a 0 en probabilidad en $\theta \in \Theta$

Prueba. Particionando Θ en n regiones excluyentes $\Theta_1^n, \Theta_2^n, \dots, \Theta_n^n$ de tal modo que la distancia entre dos puntos cualquiera dentro de cada Θ_i^n tienda a 0 cuando n tiende a ∞ . Sea $\theta_1, \theta_2, \dots, \theta_n$, una sucesión arbitraria de K -vectores tales que $\theta_i \in \Theta_i^n, i = 1, 2, \dots, n$. Entonces escribiendo $g_t(\theta)$ para $g(\mathbf{y}_t, \theta)$, tenemos para cualquier $\epsilon > 0$

$$\begin{aligned} & P \left[\sup_{\theta \in \Theta} \left| T^{-1} \sum_{t=1}^T g_t(\theta) \right| > \epsilon \right] \tag{1.2.7} \\ & \leq P \left[\bigcup_{i=1}^n \left\{ \sup_{\theta \in \Theta_i^n} \left| T^{-1} \sum_{t=1}^T g_t(\theta) \right| > \epsilon \right\} \right] \\ & \leq \sum_{i=1}^n P \left[\sup_{\theta \in \Theta_i^n} \left| T^{-1} \sum_{t=1}^T g_t(\theta) \right| > \epsilon \right] \\ & \leq \sum_{i=1}^n P \left[\left| T^{-1} \sum_{t=1}^T g_t(\theta_i) \right| > \frac{\epsilon}{2} \right] + \sum_{i=1}^n P \left[T^{-1} \sum_{t=1}^T \sup_{\theta \in \Theta_i^n} |g_t(\theta) - g_t(\theta_i)| > \frac{\epsilon}{2} \right] \end{aligned}$$

Puesto que $g_t(\theta)$ es uniformemente continua en $\theta \in \Theta$, tenemos para cada i

$$\lim_{n \rightarrow \infty} \sup_{\theta \in \Theta_i^n} |g_t(\theta) - g_t(\theta_i)| = 0 \quad (1.2.8)$$

pero, puesto que

$$\sup_{\theta \in \Theta_i^n} |g_t(\theta) - g_t(\theta_i)| \leq 2 \sup_{\theta \in \Theta} |g_t(\theta)| \quad (1.2.9)$$

y el lado derecho de la desigualdad (1.2.9) es integrable por nuestros supuestos, (1.2.8) implica por el teorema de convergencia de Lebesgue.²

$$\lim_{n \rightarrow \infty} E \left(\sup_{\theta \in \Theta_i^n} |g_t(\theta) - g_t(\theta_i)| \right) = 0 \quad (1.2.10)$$

uniformemente para i . Se toma n lo suficientemente grande para que el valor esperado en (1.2.10) sea menor que $\epsilon/2$. Finalmente, la conclusión del teorema se sigue del Teorema 18 (LGN 2 de Kolmogorov) cuando T tiende a infinito.

Este teorema puede generalizarse hasta decir que

$$T^{-1} \sum_{t=1}^T g_t(\theta_i)$$

y

$$T^{-1} \sum_{t=1}^T \sup_{\theta \in \Theta_i^n} |g_t(\theta) - g_t(\theta_i)|$$

pueden ser sujetos a una Ley de Grandes Números. ■

Teorema 27. *Sea $g_t(\mathbf{y}, \theta)$ una función medible de \mathbf{y} en un espacio Euclideo para cada t y para cada $\theta \in \Theta$, un subconjunto compacto de R^K (K -espacio Euclideo) y una función continua de θ uniformemente en \mathbf{y} y t . Asumiendo $Eg_t(\mathbf{y}, \theta) = 0$. Sea $\{\mathbf{y}_i\}$ una sucesión de vectores aleatorios independientes y no necesariamente idénticamente distribuidos tales que*

$$E \left(\sup_{\theta \in \Theta} |g_t(\mathbf{y}_t, \theta)|^{1+\delta} \right) \leq M < \infty$$

para algún $\delta > 0$. Entonces

$$T^{-1} \sum_{t=1}^T g_t(\mathbf{y}_t, \theta)$$

converge a 0 uniformemente en probabilidad en $\theta \in \Theta$

²(Royden, 1968, p. 88)

Teorema 28. Sean $\mathbf{y}_1, \mathbf{y}_2, \dots, \mathbf{y}_T$ vectores de constantes. Se define la función de distribución empírica de $(\mathbf{y}_1, \mathbf{y}_2, \dots, \mathbf{y}_T)$ por

$$F_T(\alpha) = T^{-1} \sum_{t=1}^T \chi(\mathbf{y}_t < \alpha)$$

donde χ toma el valor 1 o 0 dependiendo si el evento en su argumento ocurre o no. Se debe notar que $\mathbf{y}_t < \alpha$ significa que cada elemento del vector \mathbf{y}_t es menor que el correspondiente elemento de α . Asumiendo que $g(\mathbf{y}, \theta)$ es una función continua y acotada de \mathbf{y} en el espacio Euclideo y θ en un conjunto compacto Θ . También asumiendo que F_T converge a una función de distribución F . Entonces

$$\lim_{T \rightarrow \infty} T^{-1} \sum_{t=1}^T g(\mathbf{y}_t, \theta) = \int g(\mathbf{y}, \theta) dF(\mathbf{y})$$

uniformemente en θ .

1.3. Normalidad Asintótica

La normalidad asintótica del Estimador de Máxima Verosimilitud o más precisamente una raíz consistente de la ecuación de verosimilitud 1.1.13 se puede analizar haciendo $Q_T = \log L_T$ en el teorema (23). Si $\{y_t\}$ son independientes se puede escribir

$$\log L_T = \sum_{t=1}^T \log f_t(y_t, \theta) \tag{1.3.1}$$

donde f_t es la densidad marginal de y_t . Así bajo consideraciones generales sobre f_t se puede aplicar una ley de grandes números a

$$\frac{\partial^2 \log L_T}{\partial \theta \partial \theta'}$$

y el teorema de límite central a

$$\frac{\partial \log L_T}{\partial \theta}.$$

Aún si los $\{y_i\}$ no son independientes se puede aplicar la ley de los grandes números y el teorema de límite central mientras que el grado de dependencia está limitado en cierta manera. Así se espera se mantengan los supuestos 2 y 3 del teorema (23) en el caso del estimador de máxima verosimilitud.

Sin embargo cuando se usa las características de L_t como una función de densidad conjunta se pueden alcanzar resultados más específicos que el teorema (23), y de

acuerdo con las condiciones de regularidad en la función de verosimilitud³ se deduce que:

$$\mathbf{A}(\theta_0) = -\mathbf{B}(\theta_0) \quad (1.3.2)$$

Por tanto, haciendo (1.3.2) como un supuesto adicional y enunciándolo como un teorema se tiene:

Teorema 29. *Bajo los supuestos del teorema (23) y el supuesto anterior, el estimador de máxima verosimilitud $\hat{\theta}_T$ satisface*

$$\sqrt{T} \left(\hat{\theta}_T - \theta_0 \right) \longrightarrow N \left\{ 0, - \left[\lim E \frac{1}{T} \frac{\partial^2 \log L_T}{\partial \theta \partial \theta'} \Big|_{\theta_0} \right]^{-1} \right\} \quad (1.3.3)$$

Si $\{y_t\}$ son independientes e idénticamente distribuidas con la función de densidad común $f(\cdot, \theta)$ se pueden reemplazar los supuestos 2 y 3 del teorema (23) además del supuesto (1.3.2) con las siguientes condiciones sobre $f(\cdot, \theta)$:

$$\int \frac{\partial f}{\partial \theta} dy = 0 \quad (1.3.4)$$

$$\int \frac{\partial^2 f}{\partial \theta \partial \theta'} dy = 0 \quad (1.3.5)$$

$$\text{plim} \frac{1}{T} \sum_{t=1}^T \frac{\partial^2 \log f}{\partial \theta \partial \theta'} = E \left(\frac{\partial^2 \log f}{\partial \theta \partial \theta'} \right) \quad (1.3.6)$$

Se puede encontrar un conjunto suficiente de condiciones para 1.3.6 haciendo

$$g_t(\theta) = \frac{\partial^2 \log f_t}{\partial \theta_i \partial \theta_j}$$

en el Teorema 26

Teorema 30. (Liapounov CLT) *Sea $\{X_t\}$ independiente con $E(X_t) = \mu_t, V(X_t) = \sigma_t^2$ y*

$$E[|X_t - \mu_t|^3] = m_{3t}.$$

Si

$$\lim_{n \rightarrow \infty} \left[\sum_{t=1}^n \sigma_t^2 \right]^{-1/2} \left[\sum_{t=1}^n m_{3t} \right]^{1/3} = 0,$$

entonces, $Z_n \rightarrow N(0, 1)$

³Ver Apéndice A.

Capítulo 2

Modelos de Elección Binaria

2.1. Generalidades

2.1.1. Introducción

Un individuo en su vida cotidiana se enfrenta a muchas situaciones en las que debe tomar una decisión. A veces, esta elección la hace de manera automática o por mera intuición, sin ser consciente de que, tras cada opción, ha realizado un razonamiento lógico. Sin embargo, en otras ocasiones el individuo debe tomar una decisión que requiere una reflexión más profunda que la mera intuición.

En estos casos, el proceso de elección dependerá de determinados factores, entre ellos sus propias preferencias y condicionantes y acarreará unas consecuencias concretas.

Las repercusiones que se derivan de determinada elección pueden ser de distinta trascendencia. No es lo mismo decidir sobre si viajar en taxi en bus urbano, que decidir si conceder o negar un crédito a una empresa.

De manera análoga, la economía, como ciencia ligada al estudio de la forma en que los agentes emplean los escasos recursos, está sujeta a los continuos procesos de decisión entre las diferentes alternativas que se presentan en la vida cotidiana, tanto a nivel macroeconómico como microeconómico. Ello indica que una de las tareas más importantes en la actitud de los agentes económicos es la búsqueda de herramientas que permitan objetivizar el problema de una forma racional. ([2],[13])

La teoría económica facilita modelos sobre la conducta de los agentes en circunstancias similares, y la Econometría desarrolla métodos para explicar el comportamiento de los individuos ante los procesos de decisión, los cuales se pueden realizar mediante la mera intuición económica o bien a través de un proceso objetivo mediante uno de los modelos específicos distintos a los modelos agregados, ya que estos explican el comportamiento medio del grupo en lugar de interpretar el comportamiento individual. De esta forma, la Econometría pretende dar respuesta

a estas necesidades de la economía, planteando modelos que permitan medir y cuantificar los hechos económicos, con el fin de proporcionar un marco teórico-empírico que facilite los procesos objetivos de toma de decisión.

Por ejemplo, se supone un agente económico, una mujer casada, que se enfrenta a la decisión de participar o no en el mercado laboral. La decisión adoptada lleva asociado unos factores o características propias de la mujer (edad, número de hijos, edad de los hijos, renta familiar, nivel de estudio, etc.). El hecho de que la mujer opte por trabajar lleva asociados unos factores positivos (beneficios) y otros negativos (costos). Ante esta disyuntiva, la mujer puede valorar los beneficios y los costos que le supondría la elección posible. Tras esta valoración de todos los factores (variables) a favor y en contra, se marca su criterio de preferencia en función del cual tomará una decisión.

Es evidente imaginar que la decisión adoptada presenta cierta aleatoriedad ya que la respuesta puede ser distinta para cada una de las mujeres estudiadas, aún en el caso de que los valores de las variables que las condicionasen fuesen idénticos. Es decir, que dos mujeres con las mismas características o condicionantes podrán optar por decisiones diferentes. Sin embargo, a pesar de esas diferencias, las ordenaciones de las preferencias de ambas mujeres tendrán en común algunos rasgos importantes a partir de los que se podría establecer un patrón de comportamiento medio de la población de mujeres.

De esta forma, el estudio de los comportamientos individuales de las mujeres puede servir para modelar el mercado laboral de la mujer, proporcionando un instrumento para la planificación económica y social de las familias.

En general, esta posibilidad implica que a partir de la observación del comportamiento de una muestra de individuos, se pueden obtener conclusiones generales sobre el comportamiento medio de la población a la que pertenece dicha muestra de individuos analizada. Esta posibilidad de establecer resultados generales a partir del análisis de una muestra es un factor clave dentro de los estudios económicos, puesto que permiten establecer medidas de política económica.

Así, cabe poner de manifiesto que para analizar los procesos de decisión se necesita algún instrumento que permita describir las preferencias, así como establecer una ordenación posible de ellas. Por lo tanto, dentro del ámbito de la Econometría, es importante abordar los problemas de la decisión mediante un tipo de modelos que permita analizar las conductas individuales y que, en consecuencia, pueda explicar las preferencias o utilidades de los agentes económicos. La rama de la Econometría que se dedica a estudiar y modelar los procesos de decisión, así como dar solución a los procesos de elección entre diversas alternativas se denomina Microeconometría.

Tradicionalmente, la Econometría se enfrentaba a los problemas de decisión utilizando los modelos agregados, los cuales explican el comportamiento medio del grupo, de tal manera que haciendo uso del modelo general se definía la conducta de los agentes. Sin embargo, cuando se necesitaba analizar los procesos de decisión

individual, y las relaciones causales inherentes al proceso de decisión, es de mayor interés estudiar los modelos desde un punto de vista desagregado (datos individuales) o analizar de cada individuo su comportamiento más probable. Así, estos estudios necesitan un marco de información con datos individuales o desagregados para cada uno de los agentes económicos, lo cual supone un gran esfuerzo en el tratamiento de la información.

La Microeconometría nace por la necesidad de brindar una interpretación económica de las pautas del comportamiento de los agentes económicos o individuos, así como la posibilidad de contrastar estadísticamente las hipótesis efectuadas ([15]). Ha desarrollado un marco teórico-práctico de referencia donde fundamentar el estudio de los procesos de decisión individual, gracias a los avances informáticos que permiten el tratamiento de una gran cantidad de datos individuales. El gran interés de la Microeconometría se centra en que en estos tipos de técnicas econométricas permiten resolver los problemas del tratamiento desagregado (individual) de la información. Estos modelos facilitan, también, la tarea de identificar las características o factores que provocan un comportamiento diferente de los individuos ante los procesos de decisión.

A pesar de la gran utilidad de estos modelos, las técnicas microeconómicas presentan algunas limitaciones. Las estimaciones obtenidas mediante dichas técnicas están determinadas de forma muy importante por los supuestos básicos establecidos, es decir, los supuestos fundamentales (como por ejemplo, la normalidad de las preferencias, errores de medida, simultaneidad de las relaciones, etc.). Estos son aspectos no demasiado importantes desde el punto de vista económico, pero pueden crear serios problemas (inconsistencias) en las estimaciones de los parámetros.

Es importante destacar que las técnicas tradicionales de estimación, como por ejemplo Mínimos Cuadrados Ordinarios (MCO), resultan inapropiados dado el carácter de discontinuidad que tiene el regresando y la no linealidad de la ecuación. Por lo tanto, en estos modelos se deben recurrir a procesos de estimación no lineales, donde la técnica generalmente usada es a través de Máxima Verosimilitud (MV). Además, en Microeconometría se da gran importancia a los diagnósticos que permiten validar correctamente los resultados del modelo, así como en los supuestos establecidos en su formulación teórica. La parte de la Econometría que estudia los problemas teóricos inherentes a este tipo de modelos se denomina Métodos Microeconómicos.

El proceso de elección entre diferentes alternativas está sometido a ciertas limitaciones, que radican en la asignación de un cierto grado de objetividad a cada una de las posibles opciones a las que se enfrenta el individuo económico. La Microeconometría da respuesta a este tipo de problemas asignando la probabilidad de que un individuo elija una determinada alternativa entre un conjunto finito, exhaustivo y mutuamente excluyente de éstas. Dicha probabilidad de elección depende del conjunto de características de cada una de las alternativas, así como

de los condicionantes del individuo decisor.

Así, en el ejemplo anteriormente expuesto sobre la participación laboral de la mujer casada, se puede establecer que el hecho de que la mujer trabaje o no está condicionado por dos tipos de características:

- Características propias de la elección.
- Características propias de los agentes.

Cada uno de estos regresores o características influyen en los procesos de decisión, siendo el papel de la Microeconometría determinar un modelo que permita, dadas esas características, explicar ese proceso de decisión. A tal fin, se establece un valor numérico que permita ordenar las preferencias y que está asociado, en nuestro ejemplo, con el hecho probabilístico de que la mujer decida trabajar.

Por lo tanto, en estos modelos lo que se explican no es el valor que toma el regresando o variable endógena, sino la probabilidad de que el agente económico i elija una determinada alternativa j , que dependerá de los factores que condicionan el proceso de decisión, y de acuerdo con la función de distribución de la probabilidad que se haya supuesto para el caso de elegir una alternativa frente a sus complementarias.

Es frecuente encontrar problemas económicos en donde la decisión a tomar involucra dos o más alternativas, y la variable dependiente que se va a utilizar no tiene un valor concreto o dicho de otra manera, no es cuantitativa y por tanto es necesario cuantificarla a través de categorías mutuamente excluyentes, para mostrar cuál es la alternativa elegida por el individuo. A los modelos que tienen una variable dependiente con estas características se les llama Modelos de Elección Discreta. Cuando el individuo debe decidir entre dos alternativas, la variable dependiente sólo mostrará dos categorías y por tanto este modelo se denomina Modelo de Respuesta Dicotómica o Modelo de Elección Binaria; de la misma manera, cuando el individuo debe decidir entre más de dos alternativas, el modelo se denomina Modelo de Elección Múltiple o Modelo Multinomial. (Ver Tabla 2.1). Estos modelos están determinados por:

- El tipo de función de distribución que se use.
- El proceso a seguir al efectuar la elección (ordenado o secuencial).
- Por las características propias de la elección y de los agentes.

Como el objetivo primordial de estos modelos es hallar la probabilidad que el individuo elija una alternativa determinada, según el valor de las variables exógenas o explicativas, estos modelos requieren un tratamiento econométrico particular. Por tanto, en los modelos microeconómicos adquieren la mayor importancia, las técnicas de recolección de la información o técnicas de muestreo, como quiera que la validez de estos resultados, se verá condicionada por la representatividad de la muestra.

Cuadro 2.1: Tipos de Modelos de Elección Discreta.

Nro de alternativas	Tipo de alternativas	Tipo de función	El regresor se refiere a:	
			Características (de los individuos)	Atributos (de las alternativas)
Modelos de respuesta dicotómica (2 alternativas)	Complementarias	Lineal	Modelo de Probabilidad Lineal Truncado	
		Logística	Modelo Logit	
		Normal Tipificada	Modelo Probit	
Modelos de Respuesta Múltiple (> 2 alternativas)	No Ordenadas	Logística	Logit Multinomial - Logit Anidado - Logit Mixto	Logit Condicional - Logit Anidado - Logit Mixto
		Normal Tipificada	Probit Multinomial Probit Multivariante	Probit Condicional Probit Multivariante
	Ordenadas	Logística	Logit Ordenado	
		Normal Tipificada	Probit Ordenado	

En los Modelos de Elección Discreta, su interpretación económica parte del concepto de utilidad, el cual asume que la racionalidad de los agentes económicos los hace comportarse de tal manera que maximicen la utilidad esperada que les brinda cada una de las alternativas posibles sobre las que deben decidir. El planteamiento teórico de las preferencias de los individuos permite ordenar las diferentes alternativas en función del interés para el individuo decisor y proporciona un marco teórico en el cual se fundamentan los modelos de elección discreta. Este marco teórico permite establecer que la probabilidad que el individuo *i-ésimo* seleccione una de las dos alternativas, depende directamente de que la utilidad brindada por dicha decisión sea mayor que la brindada por la opción alterna. Se asume aquí que la utilidad derivada de una elección U_{i1} o U_{i2} es función de las variables que representan las características propias de cada una de las alternativas y de las características socio-económicas del individuo. También existe una perturbación aleatoria ϵ_{ij} que resume las desviaciones que los agentes tienen respecto al comportamiento del agente medio.

En el caso de decisión binaria ($j = 0, 1$), se puede plantear en los siguientes términos:

- U_{i0} : Nivel de utilidad que le proporciona la elección 0 al agente i .
- U_{i1} : Nivel de utilidad que le proporciona la elección 1 al agente i .
- X_{i0} : Vector de variables explicativas que caracterizan la elección de la alternativa 0 por el agente i .
- X_{i1} : Vector de variables explicativas que caracterizan la elección de la alternativa 1 por el agente i .

Además, suponiendo linealidad en las funciones, implicaría que:

$$\begin{aligned} U_{i0} &= \bar{U}_{i0} + \epsilon_{i0} = \alpha_0 + X_{i0}\beta_0 + \epsilon_{i0} \\ U_{i1} &= \bar{U}_{i1} + \epsilon_{i1} = \alpha_1 + X_{i1}\beta_1 + \epsilon_{i1} \end{aligned} \tag{2.1.1}$$

donde los valores \bar{U}_{i0} y \bar{U}_{i1} representan las utilidades medias asociadas a cada elección, las cuales pueden ser observadas y conocidas a través del proceso de investigación. Las utilidades medias son función de una combinación lineal de las variables explicativas observadas: $X_{i0}\beta_0$ y $X_{i1}\beta_1$. Además, ϵ_{i0} y ϵ_{i1} representan aquellos factores de la utilidad asociada a cada una de las alternativas que son desconocidas y que pueden variar según los individuos y según la alternativa de la que se trate. Se considera que estas son variables aleatorias ruido blanco, es decir, independientemente distribuidas, con esperanza constante igual a cero y varianza constante e igual σ_ϵ^2 ; además, generalmente se supone que siguen una función de distribución normal.

El agente i elegirá la opción uno si la utilidad de esa elección supera la de la opción cero y viceversa. Es decir,

$$Y_i = \begin{cases} 1 & \text{si } U_{i1} > U_{i0} \\ 0 & \text{si } U_{i0} > U_{i1} \end{cases} \tag{2.1.2}$$

Como consecuencia, se puede comprobar que la probabilidad que un individuo elija la opción uno será:

$$\begin{aligned}
 P(Y_i = 1) &= P(U_{i1} > U_{i0}) & (2.1.3) \\
 &= P(\bar{U}_{i1} + \epsilon_{i1} > \bar{U}_{i0} + \epsilon_{i0}) \\
 &= P(\bar{U}_{i1} - \bar{U}_{i0} > \epsilon_{i0} - \epsilon_{i1}) \\
 &= P(\epsilon_{i0} - \epsilon_{i1} < \bar{U}_{i1} - \bar{U}_{i0}) \\
 &= P(\epsilon_{i0} - \epsilon_{i1} < (\alpha_1 + X_{i1}\beta) - (\alpha_0 + X_{i0}\beta)) \\
 &= P(\epsilon_{i0} - \epsilon_{i1} < (\alpha_1 - \alpha_0) + (X_{i1} - X_{i0})\beta) \\
 &= F(X_i\beta) = \int_{-\infty}^{X_i\beta} f(\epsilon_{i0} - \epsilon_{i1})d(\epsilon_{i0} - \epsilon_{i1})
 \end{aligned}$$

con \mathbf{X}_i siendo el vector de atributos observados (medidos) de las alternativas y del propio individuo y β es un vector de parámetros desconocidos, constantes para todos los individuos pero que pueden variar entre una alternativa y otra.

A través de la ecuación 2.1.3 se determina la probabilidad que un individuo elija la opción uno, que depende de la distancia entre utilidades ($\bar{U}_{i1} - \bar{U}_{i0}$) observadas más que de la utilidad real. Así mismo, dicha probabilidad viene dada por el valor de la función de distribución (F en general) en el punto $X_{i1}\beta$, es decir, $F(X_i\beta)$.

Los Modelos de Elección Discreta se utilizan para la modelación de la elección de modo de transporte, (es decir, para determinar la probabilidad de escoger un modo de transporte, vehículo privado o algún tipo de transporte público) a partir de determinadas variables y de las características propias de las alternativas de transporte (costo, distancia, comodidad, duración del viaje, entre otros). También describen el comportamiento del individuo, frente a las diferentes alternativas posibles, usando sistemas de elección probabilísticos. Esto es, en una situación de elección entre un conjunto de alternativas finito, exhaustivo y con alternativas mutuamente excluyentes, los modelos de elección cualitativa determinan la expresión de la probabilidad, P_{ij} , de que el individuo (el elector) i identificado por un vector de características, elija la alternativa j caracterizada por una serie de atributos.

Los modelos de elección discreta más utilizados en la práctica, son: el Logit y el Probit. Ambos son modelos dicotómicos que analizan los problemas asociados a la toma de decisiones cuando los agentes económicos se enfrentan a un proceso de elección, en este caso binario. El criterio de selección entre opciones depende de la probabilidad asociada a cada una de las alternativas que puede tener un individuo.

Para la especificación del modelo de elección binaria, se admite como válida la teoría de la maximización de la utilidad aleatoria ([17]). Esta teoría postula que cada

individuo elige aquella opción que le reporta mayor utilidad (beneficio) y que ésta es aleatoria, lo que obliga a plantear la elección en términos probabilísticos.

2.1.2. El Proceso de Elección

Una elección puede analizarse como una secuencia de decisiones que incluya los siguientes pasos ([4]):

1. Definición del problema de elección
2. Generación de alternativas
3. Evaluación de los atributos de las alternativas
4. Elección
5. Ejecución de la alternativa elegida

Es evidente que los individuos no realizan este proceso siempre que deben efectuar una elección, sino que pueden omitirlo y tomar una decisión, por ejemplo, por hábito, convencionalismo social, intuición, etc. En cualquier caso, estos comportamientos podrían describirse mediante un proceso de elección en el que el decisor generase una única alternativa.

Para poder analizar este proceso será necesario precisar quién se considera decisor (individuo, familia, grupo, etc.), cómo se genera el conjunto de alternativas considerado (aquellas que el decisor juzga disponibles teniendo en cuenta sus características y restricciones personales), qué atributos van a caracterizar a las alternativas y qué reglas se van a seguir para la elección.

Las reglas de decisión describen el mecanismo interno que utiliza el decisor para procesar la información disponible y alcanzar una elección única. Existen numerosas reglas de decisión propuestas ([10]), que se pueden clasificar en las siguientes categorías ([4]):

- **Dominancia:** Una alternativa es dominante respecto a otra si al menos un atributo es mejor y en los demás no es peor. Esta regla puede no conducir a una elección única. Puede añadirse complejidad definiendo un rango de indiferencia (umbral mínimo de diferencia de valores para considerar un atributo mejor que otro)
- **Satisfacción:** Para cada atributo se considera un umbral mínimo de satisfacción para el individuo, una alternativa puede ser eliminada si no alcanza este mínimo para algún atributo.

- Reglas lexicográficas: Se supone que los atributos están ordenados por importancia, de modo que el individuo elige la alternativa que tiene el mayor valor en el atributo más importante, en caso de igualdad se pasa al siguiente atributo (del mismo modo que se realiza el orden alfabético de las palabras en los diccionarios). También se puede utilizar esta regla para ir eliminando las alternativas peores en cada atributo por orden de importancia.
- Eliminación por aspectos: Es la combinación de las reglas lexicográficas y de satisfacción. El individuo comienza por el atributo más importante y elimina las alternativas que no alcanzan su umbral de satisfacción, si todavía hay varias alternativas disponibles pasa al siguiente atributo en importancia (En [27] y [3] se plantea la equivalencia de este tipo de elección con la maximización de la utilidad aleatoria).
- Utilidad: Se considera que los atributos son conmensurables y que el atractivo de una alternativa, expresado por un vector de valores de los atributos, se puede expresar mediante un escalar, denominado en general utilidad. El individuo buscará maximizar esta utilidad (o minimizar sus costos). Esta regla implica un comportamiento de tipo compensatorio, en el que una disminución en un atributo puede ser compensada por una mejora en otro (al contrario de lo que sucedía con las reglas anteriores).

Este comportamiento hipotético de maximización de la utilidad es el que ha fundamentado casi todo el desarrollo de los modelos de elección discreta y para la detección de elecciones que no siguen los principios de máxima utilidad, especialmente con datos de preferencias declaradas, puede consultarse [26] y [25].

2.1.3. Teoría de la Utilidad Aleatoria

Al abordar el problema de modelación de la demanda en un contexto de alternativas discretas, es preciso tener en cuenta los aspectos relevantes que rodean el proceso de toma de decisiones de los individuos. Entre los más importantes cabe señalar los siguientes:

- Definición de la unidad que toma decisiones, determinación del conjunto de elección.
- Selección de los atributos que explican la utilidad de cada alternativa j .
- Determinación de la regla de decisión que sigue la unidad tomadora de decisiones.

El individuo es la unidad típica de toma de decisiones, pero alternativamente también puede considerarse a un grupo, como la familia o el hogar, a una empresa o institución pública. Cada uno de ellos puede representar una única unidad de decisión. Es importante señalar que aunque estudiemos la predicción de la demanda agregada, hay que tener en cuenta las diferencias entre los individuos en el proceso de decisión, debido principalmente a la existencia de distintos patrones de gustos.

Dado un conjunto universal de alternativas (C), el conjunto de elección de un individuo n , (C_n), está determinado por aquellas que éste tiene disponibles; a su vez, este es un subconjunto del conjunto universal ($C_n \subset C$) y representa la región factible en el problema de optimización del consumidor, definida a partir de las distintas restricciones del individuo.

En un contexto de elección modal, el conjunto universal está determinado, por ejemplo, por todos los modos de transporte existentes para realizar un viaje entre el par origen-destino considerado. Es decir, si los modos existentes son automóvil, bus y metro, un individuo particular podría no tener disponible alguna de las alternativas (por ejemplo, si no tiene permiso de conducir no tendría disponible la alternativa automóvil como conductor; si la parada de metro más cercana está a una distancia superior a un kilómetro probablemente no sería factible para él ir en metro, etc) y esto debe ser tomado en cuenta a la hora de determinar su conjunto de elección.

Identificadas las distintas alternativas, cada una de ellas es evaluada en términos de un vector de atributos que representan sus características. Así, por ejemplo, los atributos para el automóvil pueden ser tiempo, costo y confort; para el bus, tiempo, costo, frecuencia y confort, y para el metro, tiempo, costo, frecuencia y confort.

Para cada alternativa, los distintos atributos toman diferentes valores. El individuo elige, dados los valores de los atributos, aquella alternativa que considera mejor, estableciendo una regla de decisión que describe el mecanismo interno utilizado para procesar la información disponible y realizar la elección. Los procesos de decisión pueden ser compensatorios, esto es, cambios en uno o más atributos pueden ser compensados por cambios en otros; o no compensatorios, donde los buenos atributos de una alternativa no permiten compensar los malos. Los procesos de naturaleza compensatoria definen una función objetivo (normalmente, una función de utilidad) que exprese el atractivo de la alternativa en términos de sus características. Está basado en el compromiso entre atributos, pues al elegir la alternativa con la utilidad más alta, el individuo selecciona aquella que ofrece una combinación óptima de los mismos. Este tipo de modelos no son apropiados para los procesos de decisión caracterizados por discontinuidades, las cuales son mucho mejor expresadas a través de los procesos no compensatorios. Por su parte, estos procesos definen

normas o umbrales que restringen el campo de alternativas factibles, como por ejemplo que el costo del viaje no exceda de una determinada cantidad, así se define el costo del viaje máximo. Estas restricciones son consideradas por los individuos en sus procesos de decisión. Se distinguen la elección por eliminación y la satisfacción.

- En la **elección por eliminación** el individuo jerarquiza los atributos (por ejemplo, el tiempo de viaje es más importante que el costo, el costo que el tiempo de espera, etc.) y define un valor mínimo aceptable o umbral para cada uno de ellos. Posteriormente, se jerarquizan las alternativas de mejor a peor en relación al atributo, por ejemplo que el tiempo de viaje no sea mayor de veinte minutos, y todas aquellas alternativas que no satisfagan esta restricción son eliminadas y así sucesivamente hasta quedarnos con una sola alternativa. Es posible que repetido el proceso tengamos varias alternativas que cumplan todas las restricciones establecidas, en este caso se elige una de manera compensatoria (Ver [27]).
- Por su parte, **la satisfacción** presenta varias maneras de búsqueda de la estrategia de elección. Por ejemplo, podría tratarse de un proceso cíclico donde los umbrales se modifican secuencialmente hasta encontrar una alternativa. Asimismo, se podría predefinir una regla de manera que la búsqueda se alcance en el punto en el que se satisface la regla. En este caso, no todos los atributos o alternativas tienen que ser considerados.

De los dos procesos de decisión comentados, el que ha sido más utilizado para explicar el comportamiento de elección del individuo es el de Satisfacción. No obstante, en algunos casos el individuo podrá no considerar este proceso como criterio de elección, y en estas situaciones se analizará si las preferencias del individuo descritas por sus elecciones cumplen con los axiomas que establece la teoría del consumidor.

La teoría en que se sustenta el comportamiento del consumidor parte del supuesto que los individuos son racionales (*homo economicus*) y su toma de decisiones se basa en alcanzar unos objetivos dadas unas restricciones. En el contexto de las elecciones discretas, estos objetivos se alcanzan maximizando la Función de Utilidad Indirecta Condicional (FUIC) sobre el conjunto de elección del individuo.

La teoría de la utilidad aleatoria¹ constituye el cuerpo teórico que permite tratar empíricamente el problema de las elecciones discretas. A la hora de modelar las decisiones de los individuos y obtener empíricamente la Función de Utilidad Indirecta Condicional (FUIC), el investigador no dispone de información perfecta; es decir, no

¹Formalizado por [19], [9] y [17]. Desarrollos más recientes se pueden encontrar en [24]

puede observar todos los factores que participan en el proceso de elección. Puede darse la situación en que dos individuos con el mismo conjunto de alternativas disponibles, sujetos a las mismas condiciones para realizar un viaje concreto e idénticas características socioeconómicas, elijan distintas alternativas. Para resolver este problema, se acude a la teoría probabilística con la idea de tener en cuenta el efecto de factores de naturaleza aleatoria y de este modo corregir las percepciones incorrectas del analista. Así pues, se define una función de utilidad que presenta dos componentes: **Una parte observable**, determinada por el vector de características o atributos medibles de la alternativa y el vector de características socioeconómicas del individuo; y **Una componente no observable o aleatoria**, definida para incorporar las distintas fuentes de aleatoriedad que han sido identificadas; En primer lugar cualquier atributo no observado. El vector de atributos de las alternativas es usualmente incompleto, no es posible medir el conjunto total de elementos que el individuo considera en sus decisiones. Otra fuente de aparente aleatoriedad es la variación no observada en los gustos: la función de utilidad puede tener elementos no observados que varían entre individuos; en tercer lugar los errores en la medición de los distintos atributos así como los elementos de la función de utilidad no observables que se reemplazan por otros elementos (variables **proxy**).

De este modo, la utilidad U_{jn} que obtiene el individuo n al elegir la alternativa j es una variable aleatoria que se descompone en la suma de dos componentes: La componente observable (V_{jn}) y la no observable (ϵ_{jn}) según se establece en la siguiente expresión:

$$U_{jn} = V_{jn} + \epsilon_{jn} \quad j = 1, \dots, J, \quad n = 1, \dots, N, \quad j \in \mathbf{C}_n \quad (2.1.4)$$

Tanto la componente observable como la no observable dependen de los atributos de las alternativas y de las características socioeconómicas del individuo. La probabilidad que el individuo n elija la alternativa j dado un conjunto de opciones disponibles \mathbf{C}_n es:

$$\begin{aligned} P(j|\mathbf{C}_n) &= P[V_{jn} + \epsilon_{jn} \geq V_{in} + \epsilon_{in}] \quad \forall i \in \mathbf{C}_n \\ &= P[\epsilon_{in} - \epsilon_{jn} \leq V_{jn} + V_{in}] \end{aligned} \quad (2.1.5)$$

Para especificar la parte observable (V_{jn}) se usan dos tipos de variables:

- Los atributos que caracterizan a la alternativa j para el individuo n . En el ejemplo estos atributos son tiempo de viaje, costo del viaje, confort, seguridad, etc.
- El vector de variables socioeconómicas que caracterizan al individuo n . Por ejemplo la renta, la posesión o no de vehículo, el tamaño del hogar, edad, sexo, ocupación, nivel de estudios, etc.

Dos cuestiones se derivan una vez determinadas las variables a considerar: La primera, que constituye el principal problema en la especificación de la parte

determinística de la función de utilidad, está relacionada con la determinación de una combinación de atributos y variables socioeconómicas que reflejen una hipótesis razonable sobre el efecto de tales variables en la utilidad. Mientras que la segunda consiste en determinar la forma funcional. En este sentido, resulta interesante trabajar con especificaciones de funciones que definan una relación entre las variables consideradas y que a su vez estén basadas en supuestos teóricos sobre el comportamiento de los individuos, establecidos por el investigador con base en el conocimiento del mercado objetivo que se analiza. Además que la especificación presente buenas propiedades en el proceso de estimación.

Usualmente, se utilizan funciones que sean: Lineales en los parámetros, donde las variables pueden entrar en forma polinómica, o empleando cualquier otra transformación que permita relacionar los atributos con la función de utilidad. Por otra parte, las variables pueden ser genéricas o específicas; son genéricas cuando el efecto marginal es el mismo para todas las alternativas y específicas cuando la misma variable presenta efectos diferentes en función de la alternativa que se trate. El problema radica en estudiar cuál debe ser la mejor especificación de la parte observable de la función de utilidad dadas las variables y las características del mercado objetivo.

2.1.4. Maximización de la Utilidad Aleatoria (RUM)

En los **Modelos de Elección Discreta** se representa el comportamiento de un individuo que se enfrenta a la elección de una única alternativa entre un conjunto finito de alternativas disponibles. La teoría de la utilidad aleatoria parte de un individuo perfectamente racional que siempre opta por la alternativa que le supone una mayor utilidad. La utilidad que le reporta al individuo n la alternativa j es U_{nj} , completamente conocida por el que realiza la elección. Desde el punto de vista económico, U_{nj} es una función de utilidad indirecta condicional ². De este modo el individuo n escogerá la alternativa i si y sólo si

$$U_{ni} > U_{nj}, \forall j \neq i \quad (2.1.6)$$

Note que en el modelo (2.1.4), la variable aleatoria determinística V_{nj} será una función de una serie de atributos observados (medidos) de las alternativas y del propio individuo, que se denominarán \mathbf{x}_{nj} . Esta función requerirá para su calibración el ajuste de unos parámetros β .

$$V_{nj} = f(\mathbf{x}_{nj}|\beta) \quad (2.1.7)$$

En la mayor parte de los modelos de elección discreta que se emplean en la actualidad, se emplean funciones lineales en los parámetros para representar esta

²Para ver el paso de la utilidad directa a la utilidad indirecta condicional puede verse ([5])

utilidad representativa y se considera que los parámetros β son constantes para todos los individuos pero pueden variar entre unas alternativas y otras, de este modo:

$$V_{nj} = \sum_{k=1}^K \beta_{kj} \mathbf{x}_{knj} \quad (2.1.8)$$

En general se define ϵ_{nj} como la diferencia entre el valor real de la utilidad y el valor que observa el analista. De este modo, las características de ϵ_{nj} van a depender de la forma en la que el analista haya representado la elección. El error no está definido para una situación de elección en sí, sino para la especificación que el analista haya hecho de esa situación.³

Si el analista fuese capaz de determinar con absoluta precisión el valor de todas las variables que explican la conducta del elector, así como la forma precisa en la que influyen, el valor del error sería cero. Si únicamente fuese debido a los errores de medición de las variables podría ser un simple ruido blanco. En la medida en que este error se deba a la ausencia en el modelo de variables explicativas o a diferencias en la especificación de la influencia de éstas en el comportamiento, la distribución de estos errores y las correlaciones entre los errores de las distintas alternativas, los distintos individuos o las elecciones de un mismo individuo podrán ser diferentes.

Si bien la elección del individuo, bajo las hipótesis mencionadas, es plenamente determinista; el analista no va a conocer el valor de las utilidades reales, por lo que lo único que va a poder conocer es la probabilidad de que una alternativa sea la de mayor utilidad.

Como se ha comentado, el término ϵ_{nj} se trata como un error aleatorio de media cero. Si se denomina $f(\epsilon_n)$ a la función de densidad conjunta del vector aleatorio $\epsilon_n = (\epsilon_{n1}, \epsilon_{n2}, \dots, \epsilon_{nJ})$ se puede estimar la probabilidad que el individuo escoja cada una de las alternativas, dado que conoce la utilidad representativa V_{jn} . La probabilidad de que el individuo n escoja la alternativa i será:

$$\begin{aligned} P_{ni} &= P(U_{ni} > U_{nj}), \quad \forall j \neq i \\ &= P(V_{ni} + \epsilon_{ni} \geq V_{nj} + \epsilon_{nj}), \quad \forall j \neq i \\ &= P(\epsilon_{nj} - \epsilon_{ni} \leq V_{ni} - V_{nj}), \quad \forall j \neq i \end{aligned}$$

Esta es la probabilidad de que cada término aleatorio $\epsilon_{nj} - \epsilon_{ni}$ sea inferior a una cantidad observada $V_{ni} - V_{nj}$, por lo tanto se trata de una función de distribución acumulada. A partir de la función de densidad conjunta $f(\epsilon_n)$ se puede calcular esta

³Ver [28].

probabilidad como:

$$\begin{aligned} P_{ni} &= P(\epsilon_{nj} - \epsilon_{ni} \leq V_{ni} - V_{nj}), \quad \forall j \neq i \\ &= \int_i I(\epsilon_{nj} - \epsilon_{ni} \leq V_{ni} - V_{nj}) f(\epsilon_n) d\epsilon_n \quad \forall j \neq i \end{aligned} \quad (2.1.9)$$

Siendo $\mathbf{I}(\cdot)$ una función indicadora que valga uno si el individuo ha escogido la alternativa i y cero en otro caso. Dependiendo de la distribución que cada modelo asuma para el ϵ_{nj} se tendrá un valor cerrado de esta integral (en modelos como el Logit o el Logit jerárquico) o tendrá que ser evaluada numéricamente por simulación (como en el probit o el Logit mixto).

El proceso de calibrado del modelo consistirá en la estimación de los valores de los coeficientes β_{kj} . ($k = 1, \dots, K$ $j = 1, \dots, J$). El procedimiento más empleado es tomar como estimadores aquellos valores que maximicen la verosimilitud de la muestra utilizada. La función de verosimilitud de una muestra determinada será en general:

$$L(\beta) = \prod_{n=1}^N \prod_i (P_{ni})^{y_{ni}} \quad (2.1.10)$$

con

$$y_{ni} = \begin{cases} 1, & \text{si el individuo } n \text{ ha escogido la alternativa } i; \\ 0, & \text{e.o.c.} \end{cases}$$

Para facilitar la maximización numérica se trabaja con el logaritmo neperiano de la verosimilitud, esto es,

$$\mathcal{L} = \log L(\beta) = \sum_{n=1}^N \sum_i y_{ni} \ln(P_{ni}) \quad (2.1.11)$$

El estimador es el valor de β que maximiza la función de log-verosimilitud. Para hallar el estimador de β existen diversos procedimientos numéricos, en algunos casos se utilizan métodos de estimación, principalmente con modelos que requieren simulación para calcular la probabilidad, como es el caso del **Método de Máxima Verosimilitud Simulada** (MSL) el cual requiere condiciones estrictas para que el estimador sea consistente, eficiente y asintóticamente normal.

La elección observada únicamente proporciona información acerca de cuál de las utilidades es mayor. Por lo tanto, no se puede determinar la escala ni el nivel de utilidades, ya que un cambio en éstos, igual para todas las utilidades, no va a modificar el hecho de que la utilidad de una alternativa sea mayor que la de otra. En este caso se podrá normalizar el modelo en cuanto a escala y nivel, en un proceso

conocido como identificación.

Por este motivo, no pueden estimarse todos los parámetros de los modelos, ni todos los elementos de la matriz de covarianza⁴ que caracteriza la distribución conjunta de los términos de error.

2.2. Especificación del modelo

Un modelo de elección binaria se define como

$$P(Y_i = 1) = F(\mathbf{X}_i\boldsymbol{\beta}), \quad i = 1, 2, \dots, n \quad (2.2.1)$$

donde

- $\{Y_i\}$ es una sucesión de variables aleatorias binarias independientes que toman el valor uno o cero
- \mathbf{X}_i es un vector de constantes conocidas de tamaño K
- $\boldsymbol{\beta}$ es un vector de parámetros desconocidos de tamaño K
- F es una función conocida, que puede ser función lineal o no lineal de las variables originales

Entre las formas funcionales de F más frecuentemente usadas están:

$$F(x\boldsymbol{\beta}) = \begin{cases} x\boldsymbol{\beta} \text{ en cuyo caso 2.2.1 se denomina } \textit{Modelo de Probabilidad Lineal}, \\ \Phi(x\boldsymbol{\beta}) = \int_{-\infty}^x \frac{1}{\sqrt{2\pi}} \exp[-(t^2/2)] dt \text{ llamado } \textit{Modelo Probit}, \\ \Lambda(x\boldsymbol{\beta}) = \frac{e^{x\boldsymbol{\beta}}}{1 + e^{x\boldsymbol{\beta}}} \text{ y en este caso 2.2.1 se conoce como } \textit{Modelo Logit}. \end{cases}$$

La estimación e interpretación de los modelos probabilísticos lineales plantea una serie de problemas que han llevado a la búsqueda de otros modelos alternativos que permitan estimaciones más fiables para las variables respuesta. Para evitar que la variable endógena estimada pueda encontrarse fuera del rango $(0, 1)$, las alternativas disponibles llevan a utilizar modelos de probabilidad no lineales, donde la función de especificación utilizada garantice un resultado en la estimación comprendido

⁴Se denominará simplifcadamente matriz de covarianza a la matriz de varianzas-covarianzas

en el rango 0-1. Como las funciones de distribución cumplen este requisito, ya que son funciones continuas que toman valores comprendidos entre 0 y 1, se tienen varias alternativas para selección de $F(\cdot)$. Sin embargo, las más habituales son la función de distribución logística, que ha dado lugar al modelo Logit, y la función de distribución de la normal tipificada, que ha dado lugar al modelo Probit. Tanto los modelos Logit como los Probit relacionan, por tanto, la variable endógena Y_i con las variables explicativas X_{ki} a través de una función de distribución.

Para el modelo Logit, se utiliza la función logística por lo que su especificación se convierte en:

$$\begin{aligned}
 P(Y_i = 1) &= \frac{1}{1 + e^{-(\beta_1 + \beta_2 X_{2i} + \dots + \beta_k X_{ki})}} + u_i = \frac{e^{\beta_1 + \beta_2 X_{2i} + \dots + \beta_k X_{ki}}}{1 + e^{\beta_1 + \beta_2 X_{2i} + \dots + \beta_k X_{ki}}} + u_i \\
 &= \frac{e^{X_i \beta}}{1 + e^{X_i \beta}} + u_i \quad (2.2.2)
 \end{aligned}$$

Ahora, en el caso del modelo Probit la función de distribución utilizada es la normal tipificada, con lo que el modelo se especifica a través de la siguiente expresión

$$P(Y_i = 1) = \int_{-\infty}^{X_i \beta} \frac{1}{\sqrt{2\pi}} \exp \left[-\left(\frac{s^2}{2} \right) \right] ds + u_i \quad (2.2.3)$$

Dada la similitud existente entre las curvas de la normal tipificada y de la logística, los resultados estimados por ambos modelos no difieren mucho entre sí⁵. Sin embargo existen diferencias operativas, entre ellas el cálculo de la función de distribución normal es más complejo que el de la logística, ya que la primera sólo puede calcularse en forma de integral. Esto hace que el modelo Logit sea usado con mayor frecuencia en los estudios empíricos.

2.3. Estimación para modelos de elección binaria

2.3.1. Consistencia y normalidad asintótica del Estimador de Máxima Verosimilitud

Para derivar las propiedades asintóticas del Estimador de Máxima Verosimilitud (EMV) es necesario asumir los siguientes supuestos para el modelo 2.2.1:

Supuesto 2.3.1. *La forma funcional F tiene derivada f y $0 < F(x) < 1$, además con $f(x) > 0$ para toda x , y derivada de segundo orden f' .*

⁵Discrepan en la rapidez con que las curvas se aproximan a los valores extremos, y así la función logística es más achatada que la normal al alcanzar, esta última, más rápidamente los valores extremos (0 y 1)

Supuesto 2.3.2. *El espacio de parámetros B es un subconjunto abierto acotado de un espacio Euclideo k -dimensional*

Supuesto 2.3.3. *$\{x_i\}$ son uniformemente acotados en i y $\lim_{n \rightarrow \infty} n^{-1} \sum_{i=1}^n x_i x_i'$ es una matriz no singular finita. Además la distribución empírica de $\{x_i\}$ converge a una función de distribución.*

Ambos modelos, Probit y Logit, satisfacen el supuesto 2.3.1. El acotamiento de $\{x_i\}$ es algo restrictivo y hasta podría eliminarse. Sin embargo, se asume para hacer más sencilla las pruebas de consistencia y normalidad asintótica.

La función de verosimilitud del modelo 2.2.1 está dada por

$$L = \prod_{i=1}^n [F(x_i \beta)]^{y_i} [1 - F(x_i \beta)]^{1-y_i}$$

De modo que:

$$\log L = \sum_{i=1}^n y_i \log F(x_i \beta) + \sum_{i=1}^n (1 - y_i) \log[1 - F(x_i \beta)] \quad (2.3.1)$$

Ahora, para hallar el Estimador de Máxima Verosimilitud, se debe encontrar la derivada parcial de la ecuación anterior con respecto a β , es decir $\frac{\partial \log L}{\partial \beta}$. Así:

$$\begin{aligned} \frac{\partial \log L}{\partial \beta} &= \frac{\partial [\sum_{i=1}^n y_i \log F(x_i' \beta) + \sum_{i=1}^n (1 - y_i) \log[1 - F(x_i' \beta)]]}{\partial \beta} \\ &= \sum_{i=1}^n \frac{y_i f(x_i' \beta) x_i - F(x_i' \beta) f(x_i' \beta) x_i}{F(x_i' \beta)(1 - F(x_i' \beta))} \end{aligned}$$

Luego:

$$\frac{\partial \log L}{\partial \beta} = \sum_{i=1}^n \frac{y_i - F(x_i' \beta)}{F(x_i' \beta)(1 - F(x_i' \beta))} f(x_i' \beta) x_i$$

Es decir que el Estimador de Máxima Verosimilitud $\hat{\beta}$ es una solución (**si existe**) de

$$\frac{\partial \log L}{\partial \beta} = \sum_{i=1}^n \frac{y_i - F(x_i' \beta)}{F(x_i' \beta)(1 - F(x_i' \beta))} f(x_i' \beta) x_i = 0 \quad (2.3.2)$$

Para probar la consistencia de $\widehat{\beta}$ se deben verificar los supuestos del Teorema 22. Los supuestos 1 y 2 se satisfacen claramente. Para verificar el supuesto 3 se debe usar el teorema 27. Si se define

$$g_i(y, \beta) = [y - F(\mathbf{x}_i' \beta)] \log F(\mathbf{x}_i' \beta), g_i(y, \beta)$$

en una vecindad compacta de β_0 que satisface todas las condiciones para $g_t(y, \theta)$ en el teorema, debido a los supuestos 2.3.1 y 2.3.3. Además, $\lim_{n \rightarrow \infty} n^{-1} \sum_{i=1}^n F_{i0} \log F_i$ existe por el teorema 28. Por tanto, $n^{-1} \sum_{i=1}^n y_i \log F_i$ converge a $\lim_{n \rightarrow \infty} n^{-1} \sum_{i=1}^n F_{i0} \log F_i$; $F_{i0} = F(\mathbf{x}_i' \beta_0)$, uniformemente en probabilidad en $\beta \in N(\beta_0)$, una vecindad abierta de β_0 .

Un resultado similar se puede obtener para el segundo término del lado derecho de 2.3.1. Por tanto

$$\begin{aligned} Q(\beta) &= \text{plim } n^{-1} \log L \\ &= \lim_{n \rightarrow \infty} n^{-1} \sum_{i=1}^n F_{i0} \log F_i \\ &\quad + \lim_{n \rightarrow \infty} n^{-1} \sum_{i=1}^n (1 - F_{i0}) \log (1 - F_i) \end{aligned} \tag{2.3.3}$$

donde la convergencia es uniforme en $\beta \in N(\beta_0)$. Debido a los supuestos asumidos se puede derivar dentro de la operación límite en 2.3.3, para obtener

$$\frac{\partial Q}{\partial \beta} = \lim_{n \rightarrow \infty} n^{-1} \sum_{i=1}^n \frac{F_{i0}}{F_i} f_i \mathbf{x}_i - \lim_{n \rightarrow \infty} n^{-1} \sum_{i=1}^n \frac{1 - F_{i0}}{1 - F_i} f_i \mathbf{x}_i, \tag{2.3.4}$$

El cual se desaparece cuando $\beta = \beta_0$. Puesto que

$$\left. \frac{\partial^2 Q}{\partial \beta \partial \beta'} \right|_{\widehat{\beta}_0} = - \lim_{n \rightarrow \infty} n^{-1} \sum_{i=1}^n \frac{f_{i0}^2}{F_{i0} (1 - F_{i0})} \mathbf{x}_i \mathbf{x}_i', \tag{2.3.5}$$

Q alcanza un estricto máximo local en $\beta = \beta_0$. Así el supuesto 3 del teorema 22 se mantiene, con lo cual se prueba la consistencia de $\widehat{\beta}$.

Una solución de (2.3.2) puede no existir. Sin embargo, la eventual irresolubilidad no afecta la consistencia de la prueba por que la probabilidad de que esto ocurra se aproxima a cero cuando n tiende a infinito, como se muestra en el teorema 22.

Para demostrar la probabilidad asintótica se usa el teorema 29, lo cual significa que se deben verificar los supuestos 1, 2 y 3 del teorema 23 y la ecuación (1.3.2).

Derivando (2.3.2) con respecto a β se tiene que:

$$\begin{aligned} \frac{\partial^2 \log L}{\partial \beta \partial \beta'} &= - \sum_{i=1}^n \left[\frac{y_i - F_i}{F_i (1 - F_i)} \right]^2 f_i^2 \mathbf{x}_i \mathbf{x}_i' \\ &+ \sum_{i=1}^n \left[\frac{y_i - F_i}{F_i (1 - F_i)} \right] f_i' \mathbf{x}_i \mathbf{x}_i' \end{aligned} \quad (2.3.6)$$

Por lo anterior se satisface el supuesto 1 del teorema 23. También se puede verificar que los supuestos 2.3.1 y 2.3.3 implica el supuesto 2 del teorema 23 usando los teoremas 24 y 27. De (2.3.2) se tiene que:

$$\frac{1}{\sqrt{n}} \frac{\partial \log L}{\partial \beta} \Big|_{\beta_0} = \frac{1}{\sqrt{n}} \sum_{i=1}^n \frac{y_i - F_{i0}}{F_{i0} (1 - F_{i0})} f_{i0} \mathbf{x}_i \quad (2.3.7)$$

Del Teorema 19 se obtiene la normalidad asintótica de (2.3.7), porque cada término en la sumatoria es independiente con el segundo momento acotado y el tercer momento absoluto debido a los Supuestos 2.3.1, 2.3.2 y 2.3.3. Así, se satisface el supuesto 3 del 23, y se obtiene

$$\frac{1}{\sqrt{n}} \frac{\partial \log L}{\partial \beta} \Big|_{\beta_0} \rightarrow N(\mathbf{0}, \mathbf{A}) \quad (2.3.8)$$

donde

$$\mathbf{A} = \lim_{n \rightarrow \infty} \frac{1}{n} \sum_{i=1}^n \frac{f_{i0}^2}{F_{i0} (1 - F_{i0})} \mathbf{x}_i \mathbf{x}_i' \quad (2.3.9)$$

Finalmente, de la ecuación 2.3.6 se obtiene

$$\lim_{n \rightarrow \infty} \frac{1}{n} E \frac{\partial^2 \log L}{\partial \beta \partial \beta'} \Big|_{\beta_0} = -\mathbf{A} \quad (2.3.10)$$

verificando así la ecuación (1.3.2). Por tanto, se tiene que:

$$\sqrt{n} (\hat{\beta} - \beta_0) \rightarrow N(\mathbf{0}, \mathbf{A}^{-1}) \quad (2.3.11)$$

Concavidad global de la función Verosimilitud en los modelos Logit y Probit

Concavidad global significa que

$$\frac{\partial^2 \log L}{\partial \beta \partial \beta'}$$

es una matriz definida negativa para $\beta \in B$. Por expansión de Taylor se obtiene

$$\begin{aligned} \log L(\beta) &= \log L(\widehat{\beta}) + \left. \frac{\partial \log L}{\partial \beta^i} \right|_{\widehat{\beta}} (\beta - \widehat{\beta}) \\ &+ \frac{1}{2} (\beta - \widehat{\beta})' \left. \frac{\partial^2 \log L}{\partial \beta \partial \beta^i} \right|_{\widehat{\beta}^*} (\beta - \widehat{\beta}) \end{aligned} \quad (2.3.12)$$

donde β^* se encuentra entre β y $\widehat{\beta}$, concavidad global implica que

$$\log L(\widehat{\beta}) > \log L(\beta)$$

para $\beta \neq \widehat{\beta}$ si $\widehat{\beta}$ es una solución de (2.3.2). Se debe probar la concavidad global para los modelos Probit y Logit.

1. Para el modelo Logit se tiene

$$\Lambda(x) = \Lambda(x)(1 - \Lambda(x)) \quad \text{y} \quad \frac{\partial^2 \Lambda}{\partial x^2} = \frac{\partial \Lambda}{\partial x} (1 - 2\Lambda) \quad (2.3.13)$$

Ahora, insertando esta ecuación en 2.3.6 con $F(x) = \Lambda$ se obtiene

$$\frac{\partial^2 \log L}{\partial \beta \partial \beta^i} = - \sum_{i=1}^n \Lambda_i (1 - \Lambda_i) \mathbf{x}_i \mathbf{x}_i' \quad (2.3.14)$$

Donde

$$\Lambda_i = \Lambda(\mathbf{x}_i' \beta)$$

Entonces del supuesto 2.3.3 se obtiene la concavidad global.

2. Para el modelo Probit se tiene que haciendo:

$$F_i = \Phi_i$$

$$f_i = \phi_i$$

y

$$f_i' = -\mathbf{x}_i' \beta \phi_i$$

donde ϕ_i es la función de densidad de $N(0, 1)$, y reemplazando en (2.3.6) se obtiene

$$\begin{aligned} \frac{\partial^2 \log L}{\partial \beta \partial \beta^i} &= - \sum_{i=1}^n \phi_i \Phi_i^{-2} (1 - \Phi_i)^{-2} [(y_i - 2y_i \Phi_i + \Phi_i^2) \phi_i \\ &+ (y_i - \Phi_i) \Phi_i (1 - \Phi_i) \mathbf{x}_i' \beta] \mathbf{x}_i \mathbf{x}_i' \end{aligned} \quad (2.3.15)$$

Por tanto se debe mostrar la positividad de

$$g_y(x) \equiv (y - 2y\Phi + \Phi^2)\phi + (y - \Phi)\Phi(1 - \Phi)x$$

para $y = 1$ y para $y = 0$.

- Primero, sea $y = 1$. Como

$$g_1(x) = (1 - \Phi)^2(\phi + \Phi x)$$

es suficiente demostrar que $\phi + \Phi x > 0$. La desigualdad se satisface si $x \geq 0$, así se asuma $x < 0$. Pero esto es equivalente a mostrar

$$\phi > (1 - \Phi)x \text{ para } x > 0 \quad (2.3.16)$$

que se obtiene de la identidad

$$x^{-1} \exp(-x^2/2) = \int_x^\infty (1 + y^{-2}) \exp(-y^2/2) dy \quad (2.3.17)$$

- Segundo, sea $y = 0$

$$g_0(x) = \Phi^2[\phi - (1 - \Phi)x]$$

que es absolutamente positivo si $x \leq 0$ y es positivo si $x > 0$ debido a 2.3.16. Demostrando así la concavidad global para el caso Probit.

2.3.2. Métodos Iterativos para obtener el Estimador de Máxima Verosimilitud

Los métodos iterativos se pueden usar para calcular una raíz de la ecuación 2.3.2. Para los modelos Logit y Probit, la iteración es simple por la concavidad global probada en la sección anterior.

Enseguida se tratará la iteración del método de score y se le dará una adecuada interpretación. El método de score se define como

$$\hat{\beta}_2 = \hat{\beta}_1 - \left\{ \left[E \frac{\partial^2 \log L}{\partial \beta \partial \beta'} \right]_{\hat{\beta}_1} \right\}^{-1} \frac{\partial \log L}{\partial \beta} \Big|_{\hat{\beta}_1} \quad (2.3.18)$$

donde $\hat{\beta}_1$ es un estimador inicial de β_0 y $\hat{\beta}_2$ es el estimador de segunda pasada. La iteración se repite hasta que una sucesión de estimadores obtenidos sea convergente. Al usar 2.3.2 y 2.3.6 la ecuación anterior se puede escribir como

$$\hat{\beta}_2 = \left[\sum_{i=1}^n \frac{\hat{f}_i^2}{\hat{F}_i(1 - \hat{F}_i)} \mathbf{x}_i \mathbf{x}_i' \right]^{-1} \sum_{i=1}^n \frac{\hat{f}_i}{\hat{F}_i(1 - \hat{F}_i)} \mathbf{x}_i [y_i - \hat{F}_i + \hat{f}_i \mathbf{x}_i' \hat{\beta}_1], \quad (2.3.19)$$

donde

$$\widehat{F}_i = F(\mathbf{x}_i' \widehat{\beta}_1)$$

y

$$\widehat{f}_i = f(\mathbf{x}_i' \widehat{\beta}_1)$$

De la iteración 2.3.19 es posible hacer una interesante interpretación. De 2.2.1 se obtiene

$$y_i = F(\mathbf{x}_i' \beta_0) + u_i$$

donde

$$E u_i = 0$$

y

$$V u_i = F(\mathbf{x}_i' \beta_0) [1 - F(\mathbf{x}_i' \beta_0)]$$

el cual es un modelo de regresión no lineal heterocedástico. Expandiendo $F(\mathbf{x}_i' \beta_0)$ en una serie de Taylor alrededor de $\beta_0 = \widehat{\beta}_1$ y organizando términos se tiene

$$y_i - \widehat{F}_i + \widehat{f}_i \mathbf{x}_i' \widehat{\beta}_1 \cong \widehat{f}_i \mathbf{x}_i' \beta_0 + u_i \quad (2.3.20)$$

Así $\widehat{\beta}_2$ definido en 2.3.19 se puede interpretar como el estimador de Mínimos Cuadrados Ponderados (MCP) de β_0 aplicado a 2.3.20 con $V u_i$ estimado por $\widehat{F}_i (1 - \widehat{F}_i)$. Por esta razón el método la iteración del método de score en el Modelo de Elección Discreta, algunas veces se le conoce como Iteración de Mínimos Cuadrados Ponderados No Lineal (MCPNL). (Walker y Duncan, 1967)

2.3.3. Método del Mínimo Chi-cuadrado de Berkson

Existen muchas variaciones del método Chi-cuadrado ($MIN\chi^2$), uno de los cuales es el método de Berkson. Por ejemplo, el Estimador de Mínimos Cuadrados Factibles Generalizados (MCFG) definido en Amemiya⁶ es un estimador ($MIN\chi^2$). Otro ejemplo es el estimador Barankin-Gurland mencionado en Amemiya⁷. Una característica común de estos estimadores es que el minimando evaluado en el estimador es asintóticamente distribuido como chi-cuadrado, de donde se deriva su nombre.

El método $MIN\chi^2$ en el contexto del Modelo Discreto, fué propuesto primero por Berkson⁸, para el modelo Logit pero puede usarse para cualquier modelo de Elección Discreta. Es útil solamente cuando existen muchas observaciones de la variable dependiente y que tienen el mismo valor del vector de variables independientes \mathbf{x}

⁶Amemiya, Sección 6.2

⁷Amemiya, Sección 4.2.4

⁸1944

(algunas veces referido como "muchas observaciones por celda"). Así que $F(\mathbf{x}'\beta_0)$ para el valor específico de \mathbf{x} puede ser estimado con precisión por la frecuencia relativa de y cuando ésta es igual a 1.

Para explicar el método en detalle se deben definir nuevos símbolos. Suponga que el vector \mathbf{x}_i toma T valores vectoriales distintos $x_{(1)}, x_{(2)}, \dots, x_{(T)}$ y clasifica los enteros $(1, 2, \dots, n)$ en T conjuntos disjuntos I_1, I_2, \dots, I_T por la regla: $i \in I_t$ si $x_{(i)} = x_{(t)}$. Sea $P_{(t)} = P(y_i = 1)$ si $i \in I_t$. Además, sea n_t el número de enteros contenidos en I_t ,

$$r_t = \sum_{i \in I_t} y_i$$

y

$$\widehat{P}_{(t)} = r_t/n_t$$

Nótese que $\{\widehat{P}_{(t)}\}$ constituye los estadísticos suficientes del modelo. También se escribirá $\mathbf{x}_{(t)}$, $P_{(t)}$ y $\widehat{P}_{(t)}$ como \mathbf{x}_t , P_t y \widehat{P}_t si no hay ambigüedad.

De 2.2.1 se tiene que

$$P_t = F(\mathbf{x}_t' \beta_0), t = 1, 2, \dots, T. \quad (2.3.21)$$

si F es uno a uno, lo cual es consecuencia del supuesto 2.3.1. Se puede invertir la relación anterior para obtener

$$F^{-1}(P_t) = \mathbf{x}_t' \beta_0 \quad (2.3.22)$$

donde F^{-1} denota la función inversa de F . Al expandir $F^{-1}(\widehat{P}_t)$ en una serie de Taylor alrededor de P_t (lo cual es posible por el supuesto 2.3.1) se obtiene

$$\begin{aligned} F^{-1}(\widehat{P}_t) &= \mathbf{x}_t' \beta_0 + \left. \frac{\partial F^{-1}}{\partial P_t} \right|_{P_t^*} (\widehat{P}_t - P_t) \\ &= \mathbf{x}_t' \beta_0 + \frac{1}{f[F^{-1}(P_t^*)]} (\widehat{P}_t - P_t) \\ &\equiv \mathbf{x}_t' \beta_0 + v_t + w_t, \end{aligned} \quad (2.3.23)$$

donde P_t^* se encuentra entre \widehat{P}_t y P_t ,

$$v_t = \frac{1}{f[F^{-1}(P_t)]} (\widehat{P}_t - P_t)$$

y

$$w_t = \left(\frac{1}{f[F^{-1}(P_t^*)]} - \frac{1}{f[F^{-1}(P_t)]} \right) (\widehat{P}_t - P_t)$$

El hecho que v_t y w_t dependan de n ha sido eliminado.

Puesto que

$$V v_t \equiv \sigma_t^2 = \frac{P_t(1 - P_t)}{n_t f^2 [F^{-1}(P_t)]} \quad (2.3.24)$$

y debido a que w_t es $O(n_t^{-1})$ y por tanto puede ser ignorado para n_t grandes (como mostraremos más adelante), 2.3.23 define aproximadamente un modelo de regresión lineal heterocedástico. El estimador $MIN\chi^2$ denotado $\tilde{\beta}$ se define como el estimador Mínimo Cuadrado Ponderado (MCP) aplicado a 2.3.23 ignorando w_t . Se puede estimar σ_t^2 por $\hat{\sigma}_t^2$ el cual se obtuvo sustituyendo \hat{P}_t por P_t en 2.3.24. Así

$$\tilde{\beta} = \left(\sum_{t=1}^T \hat{\sigma}_t^{-2} \mathbf{x}_t \mathbf{x}_t' \right)^{-1} \sum_{t=1}^T \hat{\sigma}_t^{-2} \mathbf{x}_t F^{-1}(\hat{P}_t) \quad (2.3.25)$$

se probará la consistencia y normalidad asintótica de $\tilde{\beta}$ (cuando n tiende a infinito con T fijo) bajo supuestos 2.3.1, 2.3.3 y el siguiente supuesto adicional:

Supuesto 2.3.4. $\lim_{n \rightarrow \infty} (n_t/n) = c_t \neq 0$ para $t = 1, 2, \dots, T$ donde T es un entero fijo.

Se asume el supuesto 2.3.4 para simplificar el análisis. Sin embargo, si $c_t = 0$ para algún t se puede proceder como si las observaciones para ese t no existieran e imponer los supuestos 2.3.3 y 2.3.4 en las observaciones restantes.

Insertando la ecuación (2.3.23) en la (2.3.25) y reorganizando términos, se obtiene

$$\sqrt{n} (\tilde{\beta} - \beta_0) = \left(\frac{1}{n} \sum_{t=1}^T \hat{\sigma}_t^{-2} \mathbf{x}_t \mathbf{x}_t' \right)^{-1} \frac{1}{\sqrt{n}} \sum_{t=1}^T \hat{\sigma}_t^{-2} \mathbf{x}_t (v_t + w_t) \quad (2.3.26)$$

Como T es fijo, se tiene por Teorema (15) que

$$\text{plim}_{n \rightarrow \infty} \frac{1}{n} \sum_{t=1}^T \hat{\sigma}_t^{-2} \mathbf{x}_t \mathbf{x}_t' = \sum_{t=1}^T \bar{\sigma}_t^{-2} \mathbf{x}_t \mathbf{x}_t' \quad (2.3.27)$$

donde

$$\bar{\sigma}_t^{-2} = c_t f^2 \frac{[F^{-1}(P_t)]}{[P_t(1 - P_t)]}$$

También, al aplicar el Teorema (16) se obtiene

$$\begin{aligned}
 \frac{1}{\sqrt{n}} \sum_{t=1}^T \widehat{\sigma}_t^{-2} \mathbf{x}_t (v_t + w_t) &= \sum_{t=1}^T \frac{1}{\sqrt{n}} (\widehat{\sigma}_t^{-1}) \widehat{\sigma}_t^{-1} \sigma_t \mathbf{x}_t \\
 &\times (\sigma_t^{-1} v_t + \sigma_t^{-1} w_t) \\
 &\stackrel{LD}{\equiv} \sum_{t=1}^T \bar{\sigma}_t^{-1} \mathbf{x}_t \sigma_t^{-1} v_t
 \end{aligned} \tag{2.3.28}$$

debido a que

$$\begin{aligned}
 \text{plim } n^{-1/2} \sigma_t^{-1} &= \bar{\sigma}_t^{-1} \\
 \text{plim } \widehat{\sigma}_t^{-1} \sigma_t &= 1 \text{ y} \\
 \text{plim } \sigma_t^{-1} w_t &= 0
 \end{aligned} \tag{2.3.29}$$

Pero, debido a que $\{v_t\}$ son independientes, el vector $(\sigma_1^{-1} v_1, \sigma_2^{-1} v_2, \dots, \sigma_T^{-1} v_T)$ converge a $N(\mathbf{0}, \mathbf{I}_T)$. Por tanto

$$\sum_{t=1}^T \bar{\sigma}_t^{-1} \mathbf{x}_t \sigma_t^{-1} v_t \rightarrow N \left(\mathbf{0}, \sum_{t=1}^T \bar{\sigma}_t^{-2} \mathbf{x}_t \mathbf{x}_t' \right) \tag{2.3.30}$$

Finalmente, de (2.3.26) hasta (2.3.30) y Teorema (16) se tiene que

$$\sqrt{n} (\tilde{\beta} - \beta_0) \rightarrow N \left[\mathbf{0}, \left(\sum_{t=1}^T \bar{\sigma}_t^{-2} \mathbf{x}_t \mathbf{x}_t' \right)^{-1} \right] \tag{2.3.31}$$

Puesto que \mathbf{A} definido en 2.3.9 es también

$$\mathbf{A} = \sum_{t=1}^T \bar{\sigma}_t^{-2} \mathbf{x}_t \mathbf{x}_t'$$

bajo el supuesto 2.3.4, y las ecuaciones (2.3.11) y (2.3.31) muestran que el estimador $MIN\chi^2$ tiene la misma distribución asintótica que el Estimador de Máxima Verosimilitud. El estimador $MIN\chi^2$ es más fácil de calcular que el Estimador de Máxima Verosimilitud por que el primero es explícitamente definido en ecuación (2.3.25), mientras que el último requiere iteración. Sin embargo, el método $MIN\chi^2$ requiere un gran número de observaciones en cada celda. Si el modelo contiene varias variables independientes, esto puede ser imposible de obtener. En la siguiente subsección se compararán los dos estimadores en más detalle.

En el modelo Probit,

$$F^{-1}(\hat{P}_t) = \Phi^{-1}(\hat{P}_t).$$

Aunque Φ^{-1} no tiene una forma explícita puede ser fácilmente evaluada numéricamente. La función $\Phi^{-1}(\cdot)$ se conoce como la transformación Probit.

En el modelo Logit, se puede escribir explícitamente

$$\Lambda^{-1}(\hat{P}_t) = \log \left[\frac{\hat{P}_t}{(1 - \hat{P}_t)} \right] \quad (2.3.32)$$

la cual es llamada transformación Logit o Logaritmo de las razones impares.

Cox⁹ mencionó la siguiente modificación de 2.3.32:

$$\Lambda_c^{-1}(\hat{P}_t) = \log \left\{ \frac{[\hat{P}_t + (2n_t)^{-1}]}{[1 - \hat{P}_t + (2n_t)^{-1}]} \right\}. \quad (2.3.33)$$

Esta modificación tiene dos ventajas sobre la 2.3.32:

1. La transformación de la Ecuación (2.3.33) puede siempre estar definida mientras 2.3.32 puede no estar definida si $\hat{P}_t = 0$ o 1 (Sin embargo, no es recomendable usar la modificación de Cox cuando n_t es pequeña)
2. Se puede mostrar que

$$E\Lambda_c^{-1}(\hat{P}_t) - \Lambda^{-1}(P_t)$$

es del orden de n_t^{-2} , mientras que

$$E\Lambda^{-1}(\hat{P}_t) - \Lambda^{-1}(P_t)$$

es del orden de n_t^{-1}

Así se ha probado la consistencia y la normalidad asintótica del estimador $MIN\chi^2$ de Berkson asumiendo que $\mathbf{x}_i = \mathbf{x}_t$ para $i \in I_T$. Sin embargo, en la práctica puede ocurrir que un investigador procede como si $\mathbf{x}_i = \mathbf{x}_t$ para $i \in I_T$ aún si $\mathbf{x}_i \neq \mathbf{x}_t$ por que las observaciones individuales \mathbf{x}_i no están disponibles y sólo esta disponible su media grupal

$$\mathbf{x}_t \equiv n_t^{-1} \sum_{i \in I_t} \mathbf{x}_i.$$

En este caso, el estimador $MIN\chi^2$ es generalmente inconsistente.

⁹Cox, 1970 pag.33

2.3.4. Comparación entre el Estimador de Máxima Verosimilitud y el Estimador Chi-cuadrado Mínimo

En un modelo simple donde el vector \mathbf{x}_t consiste de 1 y de una variable independiente sencilla y donde T es pequeño, la media exacta y la varianza del Estimador de Máxima Verosimilitud y el estimador $MIN\chi^2$ pueden calcularse por medio de un método directo. Berkson 1955, 1957 lo hizo así para el modelo Logit y Probit respectivamente y encontró que el Error Cuadrado Medio exacto para el estimador $MIN\chi^2$ era más pequeño que todos los ejemplos considerados.

Amemiya¹⁰ obtuvo la fórmula para el sesgo del orden de n^{-1} y el Error Cuadrado Medio del orden de n^{-2} de MLE y de $MIN\chi^2$ en el modelo general Logit. El método que empleó en su estudio fué: Usando la Ecuación (2.3.13) y el esquema de muestreo descrito en la sección anterior, la ecuación normal 2.3.2 se reduce a

$$\sum_{t=1}^T n_t \left[\widehat{P}_t - \Lambda(\mathbf{x}_t' \beta) \right] \mathbf{x}_t = 0 \quad (2.3.34)$$

La ecuación anterior se refiere a la definición del Estimador de Máxima Verosimilitud $\widehat{\beta}$ implícitamente como una función de $\widehat{P}_1, \widehat{P}_2, \dots, \widehat{P}_T$, algo como

$$\widehat{\beta} = g\left(\widehat{P}_1, \widehat{P}_2, \dots, \widehat{P}_T\right).$$

Expandiendo g en una serie de Taylor alrededor de P_1, P_2, \dots, P_T y notando que

$$g(P_1, P_2, \dots, P_T) = \beta_0$$

se obtiene

$$\begin{aligned} \widehat{\beta} - \beta_0 &\cong \sum_t g_t u_t + \frac{1}{2} \sum_t \sum_s g_{ts} u_t u_s \\ &+ \frac{1}{6} \sum_t \sum_s \sum_r g_{tsr} u_t u_s u_r, \end{aligned} \quad (2.3.35)$$

donde

$$u_t = \widehat{P}_t - P_t$$

y g_t, g_{ts} y g_{tsr} representan las derivadas parciales de primer, segundo y tercer orden de g evaluadas en (P_1, P_2, \dots, P_T) , respectivamente. El sesgo del Estimador de Máxima Verosimilitud al orden de n^{-1} se obtiene tomando la esperanza de los dos primeros términos del lado derecho de la Ecuación (2.3.35). El Error Cuadrado Medio del Estimador de Máxima Verosimilitud al orden de n^{-2} se obtiene calculando

¹⁰Amemiya, 1980b

el Error Cuadrado Medio del lado derecho de la ecuación anterior, ignorando los términos de orden menor a n^{-2} . No es necesario considerar términos más altos en la expansión de Taylor porque el Eu_t^k para $k \geq 5$ es a lo más del orden de n_t^{-3} . Una expansión de Taylor para el estimador $MIN\chi^2 \tilde{\beta}$ se obtiene expandiendo el lado derecho de la Ecuación 2.3.25 alrededor de P_t .

Usando estas fórmulas, Amemiya calculó los Errores Cuadrados Medios aproximados del Estimador de Máxima Verosimilitud y el estimador $MIN\chi^2$ en algunos ejemplos tanto artificiales como empíricos y encontró que el estimador $MIN\chi^2$ tiene un Error Cuadrado Medio más pequeño que el Estimador de Máxima Verosimilitud en todos los ejemplos considerados. Sin embargo, la diferencia entre las dos matrices del Error Cuadrado Medio pueden no ser ni definida positiva ni definida negativa (Ghosh y Sinha, 1981). De hecho Davis 1984 mostró ejemplos en los cuales el Estimador de Máxima Verosimilitud tiene un Error Cuadrado Medio menor de n^{-2} y ofreció un argumento intuitivo que mostró que para un T grande, es más probable que el Estimador de Máxima Verosimilitud tenga un error cuadrático medio más pequeño.

Smith, Savin y Robertson (1984) realizaron un estudio Monte Carlo del Modelo Logit con una variable independiente y encontraron que aunque en estimación puntual el $MIN\chi^2$ era mejor que el Estimador de Máxima Verosimilitud, como en los estudios de Berkson y Amemiya, la convergencia de la distribución del estimador $MIN\chi^2$ a una distribución normal era algunas veces relativamente lenta, siendo poco satisfactoria en una instancia como $n = 480$.

2.4. Prueba de Hipótesis

Para probar una hipótesis en un único parámetro, se puede realizar una prueba normal estándar usando la normalidad asintótica bien sea del Estimador de Máxima Verosimilitud o $MIN\chi^2$. Una hipótesis lineal puede probarse usando métodos generales. El problema de escoger un modelo entre varias alternativas puede ser resuelto bien sea por el Criterio de Información de Akaike (AIC) o por la prueba de Cox de hipótesis no anidadas.

Aquí se tratará sólo una prueba chi-cuadrado basada en el estimador $MIN\chi^2$ de Berkson debido a que esto no es un caso especial de las pruebas discutidas. La prueba estadística es la *Weighted Sum of Squared Residuals* (WSSR)¹¹ de la ecuación 2.3.23

¹¹Suma Ponderada de los Cuadrados Residuales (SPCR)

definida por

$$WSSR = \sum_{i=1}^T \hat{\sigma}_t^{-2} \left[F^{-1} \left(\hat{P}_t \right) - \mathbf{x}_t' \tilde{\beta} \right]^2 \quad (2.4.1)$$

En el modelo de regresión normal heterocedástico $y \sim N(\mathbf{X}\beta, \mathbf{D})$ con \mathbf{D} conocido,

$$\left(y - \mathbf{X}\hat{\beta}_G \right)' \mathbf{D}^{-1} \left(y - \mathbf{X}\hat{\beta}_G \right)$$

se distribuye como χ_{T-K}^2 . De esto se puede deducir que WSSR como se define anteriormente es asintóticamente distribuido como χ_{T-K}^2 .

Se usa este hecho para escoger entre el modelo no restringido

$$P_t = F \left(\mathbf{x}_t' \beta_0 \right)$$

y el modelo restringido

$$P_t = F \left(\mathbf{x}_{1t}' \beta_{10} \right)$$

donde \mathbf{x}_{1t} y β_{10} son los primeros $K - q$ elementos de \mathbf{x}_t y β_0 respectivamente. Sean $WSSR_u$ y $WSSR_c$ los valores de la Ecuación 2.4.1 derivados de los modelos no restringido y restringido, respectivamente. Se debería escoger el modelo no restringido si y sólo si

$$WSSR_c - WSSR_u > \chi_{q,\alpha}^2 \quad (2.4.2)$$

donde $\chi_{q,\alpha}^2$ denota el valor crítico α % de χ_q^2 .

Para escoger entre modelos no anidados, se puede usar la siguiente variación del Criterio de Información de Akaike:

$$AIC = \frac{1}{2} WSSR + K \quad (2.4.3)$$

Esto puede justificarse sobre la base que en el modelo de regresión normal heterocedástico mencionado antes,

$$\left(y - \mathbf{X}\hat{\beta}_G \right)' \mathbf{D}^{-1} \left(y - \mathbf{X}\hat{\beta}_G \right) = -2 \log L$$

aparte de un término constante.

En lugar de 2.4.1 se puede usar

$$\sum_{t=1}^T n_t \left[\hat{P}_t \left(1 - \hat{P}_t \right) \right]^{-1} \left[\hat{P}_t - F \left(\mathbf{x}_t' \tilde{\beta} \right) \right]^2, \quad (2.4.4)$$

puesto que esta es asintóticamente equivalente a 2.4.1.

2.5. Análisis Discriminante

El propósito del Análisis discriminante es medir las características de un individuo o de un objeto y con base en las medidas, clasificarlo en uno de dos posibles grupos.

Es posible enunciar el problema estadísticamente como sigue: Supóngase que el vector de variables aleatorias \mathbf{x}^* se genera de acuerdo a alguna densidad g_1 o g_0 , se va a clasificar una observación dada en \mathbf{x}^* , denotada \mathbf{x}_i^* , en el grupo caracterizado por g_1 o g_2 . Es útil definir $y_i = 1$ si \mathbf{x}^* se genera por g_1 y $y_i = 0$ si es generado por g_0 . Se debe predecir y_i basados en \mathbf{x}^* . La información esencial necesaria para la predicción es la probabilidad condicional $P(y_i = 1|\mathbf{x}_i^*)$.

Se ignorará el problema de la predicción dada la probabilidad condicional y se encamina la pregunta de cómo especificar y estimar la probabilidad condicional.

Por Teorema de Bayes se tiene

$$P(y_i = 1|\mathbf{x}_i^*) = \frac{g_1(\mathbf{x}_i^*) q_1}{g_1(\mathbf{x}_i^*) q_1 + g_0(\mathbf{x}_i^*) q_0} \quad (2.5.1)$$

donde q_1 y q_0 denotan las probabilidades marginales $P(y_i = 1)$ y $P(y_i = 0)$ respectivamente. Se evaluará la ecuación anterior, asumiendo que g_1 y g_0 son las densidades de $N(\mu_1, \Sigma_1)$ y $N(\mu_0, \Sigma_0)$, respectivamente. Este supuesto se enuncia formalmente como

$$\begin{aligned} \mathbf{x}_i^* | (y_i = 1) &\sim N(\mu_1, \Sigma_1) \\ \mathbf{x}_i^* | (y_i = 0) &\sim N(\mu_0, \Sigma_0) \end{aligned} \quad (2.5.2)$$

Esta es la forma más comúnmente usada de Análisis Discriminante, también llamado Análisis Discriminante Normal.

Bajo la Ecuación (2.5.2), el modelo (2.5.1) se reduce al siguiente modelo Logit Cuadrático:

$$P(y_i = 1|\mathbf{x}_i^*) = \Lambda\left(\beta_{(1)} + \beta_{(2)}' \mathbf{x}_i^* + \mathbf{x}_i^{*'} \mathbf{A} \mathbf{x}_i^*\right), \quad (2.5.3)$$

donde

$$\begin{aligned} \beta_{(1)} &= \frac{1}{2} \mu_0' \Sigma_0^{-1} \mu_0 - \frac{1}{2} \mu_1' \Sigma_1^{-1} \mu_1 + \log q_1 - \log q_0 \\ &\quad - \frac{1}{2} \log |\Sigma_1| + \frac{1}{2} \log |\Sigma_0| \end{aligned} \quad (2.5.4)$$

$$\beta_{(2)} = \Sigma_1^{-1} \mu_1 - \Sigma_0^{-1} \mu_0 \quad (2.5.5)$$

y

$$\mathbf{A} = \frac{1}{2} (\Sigma_0^{-1} - \Sigma_1^{-1}) \quad (2.5.6)$$

En el caso especial $\Sigma'_1 = \Sigma_0$ el cual se asume frecuentemente en aplicaciones econométricas, se tiene $\mathbf{A} = \mathbf{0}$; por tanto, la Ecuación 2.5.3 se reduce al Modelo Logit Lineal:

$$P(y_i = 1 | \mathbf{x}_i^*) = \Lambda(\mathbf{x}_i^* \beta) \quad (2.5.7)$$

donde

$$\beta_{(1)} + \beta_{(2)} \mathbf{x}_i^* = \mathbf{x}_i^* \beta$$

de acuerdo la notación de la sección 2.2

Se considera la estimación Máxima Verosimilitud los parámetros $\mu_1, \mu_0, \Sigma_1, \Sigma_0, q_1$ y q_0 basados en las observaciones $(y_i, \mathbf{x}_i^*), i = 1, 2, \dots, n$. La determinación de q_1 y q_0 varía de acuerdo a los autores. Se ha adoptado aquí el enfoque de Warner (1963) y se trata a q_1 y q_0 como parámetros desconocidos a estimar. La función verosimilitud puede escribirse como

$$\mathbf{L} = \prod_{i=1}^n [g_1(\mathbf{x}_i^*) q_1]^{y_i} [g_0(\mathbf{x}_i^*) q_0]^{1-y_i} \quad (2.5.8)$$

Igualando las derivadas de $\log L$ a 0 produce los siguientes estimadores Máxima Verosimilitud:

1.
$$\hat{q}_1 = \frac{n_1}{n}, \quad (2.5.9)$$

donde $n_1 = \sum_{i=1}^n y_i$

2.
$$\hat{q}_0 = \frac{n_0}{n}, \quad (2.5.10)$$

donde $n_0 = n - n_1$

3.
$$\hat{\mu}_1 = \frac{1}{n_1} \sum_{i=1}^n y_i \mathbf{x}_i^* \quad (2.5.11)$$

4.
$$\hat{\mu}_0 = \frac{1}{n_0} \sum_{i=1}^n (1 - y_i) \mathbf{x}_i^* \quad (2.5.12)$$

5.

$$\widehat{\Sigma}_1 = \frac{1}{n_1} \sum_{i=1}^n y_i (\mathbf{x}_i^* - \widehat{\mu}_1) (\mathbf{x}_i^* - \widehat{\mu}_1)' \quad (2.5.13)$$

y

6.

$$\widehat{\Sigma}_0 = \frac{1}{n_0} \sum_{i=1}^n (1 - y_i) (\mathbf{x}_i^* - \widehat{\mu}_0) (\mathbf{x}_i^* - \widehat{\mu}_0)' \quad (2.5.14)$$

Si $\Sigma_1 = \Sigma_0 (\equiv \Sigma)$ como se asume frecuentemente, las Ecuaciones 2.5.13 y 2.5.14 deberían reemplazarse por

$$\begin{aligned} \widehat{\Sigma} = & \frac{1}{n} \left[\sum_{i=1}^n y_i (\mathbf{x}_i^* - \widehat{\mu}_1) (\mathbf{x}_i^* - \widehat{\mu}_1)' \right. \\ & \left. + \sum_{i=1}^n (1 - y_i) (\mathbf{x}_i^* - \widehat{\mu}_0) (\mathbf{x}_i^* - \widehat{\mu}_0)' \right] \quad (2.5.15) \end{aligned}$$

Los estimadores Máxima Verosimilitud de $\beta_{(1)}$, $\beta_{(2)}$ y \mathbf{A} se obtienen insertando estos estimados al lado derecho de las Ecuaciones 2.5.4, 2.5.5 y 2.5.6.

El análisis Discriminante es frecuentemente usado en análisis de elección de modos de transporte.

Al modelo definido por la Ecuación 2.5.2 con $\Sigma_1 = \Sigma_0$ y por la Ecuación 2.5.7, se le llama Modelo de Análisis Discriminante (MAD) y al estimado de $\beta \equiv (\beta_{(1)}, \beta_{(2)})'$ obtenido insertando las Ecuaciones 2.5.11, 2.5.12 y 2.5.15 en Ecuaciones 2.5.4 y 2.5.5 con $\Sigma_1 = \Sigma_0$ se le conoce como el estimador AD, denotado por β_{DA}' . De otro lado si se asume sólo la Ecuación 2.5.7 y no la Ecuación 2.5.2, se tiene un modelo Logit. Se denota el Estimador de Máxima Verosimilitud Logit de β por $\widehat{\beta}_\Lambda$.

El rendimiento relativo de los dos estimadores dependerá críticamente de la distribución real asumida para \mathbf{x}_i^* . Si se asume Ecuación 2.5.2 con $\Sigma_1 = \Sigma_0$ además de la Ecuación 2.5.7, el estimador de DA es el Estimador de Máxima Verosimilitud genuino y por tanto debería ser asintóticamente más eficiente que el Estimador de Máxima Verosimilitud Logit. Sin embargo, si no se asume la Ecuación 2.5.2, el estimador Análisis Discriminante (DA) pierde su consistencia en general, mientras el Estimador de Máxima Verosimilitud Logit la mantiene. Así se esperaría que el Estimador de Máxima Verosimilitud Logit sea más robusto.

2.6. Predicción Agregada

Se considera ahora el problema de predecir la proporción agregada $r \equiv n^{-1} \sum_{i=1}^n y_i$ en el Modelo de Elección Discreta (2.2.1). Esto es un problema frecuente de importancia práctica para los hacedores de políticas. Por ejemplo en el problema de elección de transporte, sería deseable saber la proporción de gente que, en una comunidad, usen el tránsito cuando se aplica una nueva tarifa y los otros valores de las variables independientes x prevalecen. Se asume que β (sin el subíndice 0) se ha estimado del anterior muestreo. Sin embargo, para simplificar el análisis se asumirá que el valor estimado de β es igual al valor real.

La predicción de r debería hacerse basándose en la distribución condicional de r dado $\{\mathbf{x}_i\}$. Cuando n es grande, se presenta la siguiente distribución asintótica:

$$r \stackrel{A}{\sim} N \left[n^{-1} \sum_{i=1}^n F_i, n^{-2} \sum_{i=1}^n F_i (1 - F_i) \right] \quad (2.6.1)$$

Una vez se calcula la media y la varianza asintótica se tiene toda la información necesaria para predecir r .

Si se observa cada \mathbf{x}_i , $i = 1, 2, \dots, n$, se puede calcular adecuadamente la media y la varianza de la Ecuación (2.6.1). Sin embargo, puesto que es más real asumir que no se puede observar cada \mathbf{x}_i se considerará el problema de cómo estimar la media y la varianza para la distribución (2.6.1) en esa situación. Para ese propósito se asumirá que $\{\mathbf{x}_i\}$ son variables aleatorias independientes e idénticamente distribuidas con la misma función de distribución k-variada G . Entonces la media y la varianza asintótica de r puede estimarse por $EF(\mathbf{x}'\beta)$ y $n^{-1}EF(1 - F)$ respectivamente, donde E es la esperanza tomada con respecto a G .

Westin (1974) estudió la evaluación de EF cuando $F = \Lambda$ (distribución logística) y $\mathbf{x} \sim N(\mu_x, \Sigma_x)$. El notó que la densidad de

$$p \equiv \Lambda(\mathbf{x}'\beta)$$

está dada por

$$f(p) = \frac{1}{\sqrt{2\pi}\sigma} \frac{1}{p(1-p)} \exp \left\{ -\frac{1}{2\sigma^2} \left[\log \left(\frac{p}{1-p} \right) - \mu \right]^2 \right\}, \quad (2.6.2)$$

donde

$$\mu = \mu_x' \beta$$

y

$$\sigma^2 = \beta' \Sigma_x \beta$$

Puesto que la media de esta densidad no tiene una expresión de forma cerrada, se debe evaluar

$$\int pf(p) dp$$

numéricamente para valores de μ y σ^2 .

McFadden y Reid 1975 mostraron que si $F = \Phi$ (distribución normal estándar) y $\mathbf{x} \sim N(\mu_x, \Sigma_x)$ como antes, se tiene

$$E\Phi(\mathbf{x}'\beta) = \Phi\left[(1 + \sigma^2)^{-1/2} \mu\right] \quad (2.6.3)$$

Así la evaluación de EF es mucho más sencilla que en el caso Logit.

Westin y McFadden no consideraron la evaluación de la varianza asintótica de r , la cual constituye una importante pieza de información para el propósito de predicción.

Otra deficiencia de estos estudios es que la variabilidad debida a la variación de β se ignora totalmente. Se sugiere un modo parcialmente bayesiano de tratar este problema. Dado un $\hat{\beta}$ estimado de β se trata β como una variable aleatoria con la distribución $N(\hat{\beta}, \Sigma_\beta)$. Un estimado de la matriz de covarianza asintótica del estimador $\hat{\beta}$ puede usarse para Σ_β .

Se hace referencia al modelo (2.6.1) como la distribución asintótica de r condicionalmente en $\{\mathbf{x}_i\}$ y β . La distribución de r condicional en $\{\mathbf{x}_i\}$ pero no en β , por lo tanto, está dado por

$$r \stackrel{A}{\sim} N \left[E_\beta n^{-1} \sum_{i=1}^n F_i, E_\beta n^{-2} \sum_{i=1}^n F_i (1 - F_i) + \mathbf{V}_\beta n^{-1} \sum_{i=1}^n F_i \right] \quad (2.6.4)$$

Finalmente, si todas las observaciones en $\{\mathbf{x}_i\}$ no están disponibles y $\{\mathbf{x}_i\}$ pueden referirse como variables aleatorias independientes e idénticamente distribuidas, se puede aproximar la distribución (2.6.4) por

$$r \stackrel{A}{\sim} N [E_x E_\beta F, n^{-1} E_x E_\beta F (1 - F) + E_x \mathbf{V}_\beta F] \quad (2.6.5)$$

Capítulo 3

Modelos de Respuesta Múltiple

3.1. Inferencia Estadística

Asumiendo que la variable dependiente y_i toma $m_i + 1$ valores $0, 1, \dots, m_i$, un modelo de elección múltiple multinomial general se define como:

$$P(y_i = j) = F_{ij}(\mathbf{x}^*, \boldsymbol{\theta}), \quad \begin{array}{l} i = 1, 2, \dots, n \text{ y} \\ j = 1, 2, \dots, m_i, \end{array} \quad (3.1.1)$$

donde \mathbf{x}^* y $\boldsymbol{\theta}$ son vectores de variables y parámetros independientes, respectivamente. La ecuación anterior se podría escribir eventualmente como $P_{ij} = F_{ij}$. Existe también la posibilidad que no todas las variables y parámetros independientes estén incluidas en el argumento de cada F_{ij} . Nótese que $P(Y_i = 0)$ ($\equiv F_{i0}$) no necesita especificarse porque debe ser igual a uno menos la suma de las m_i probabilidades definidas en 3.1.1.

Es importante hacer que m_i dependa de i porque en muchas aplicaciones los individuos enfrentan diferentes conjuntos de elección. Por ejemplo, en análisis de selección del modo de transporte, viajar por metro no se incluye en el conjunto de elecciones de quienes viven fuera de su área de servicio.

Para definir el **Estimador de Máxima Verosimilitud** de $\boldsymbol{\theta}$ en el modelo 3.1.1 es útil definir $\sum_{i=1}^n (m_i + 1)$ variables binarias.

$$\begin{array}{l} y_{ij} = 1 \text{ si } y_i = j \\ y_{ij} = 0 \text{ si } y_i \neq j, i = 1, 2, \dots, n \text{ y} \\ j = 0, 1, \dots, m_i, \end{array} \quad (3.1.2)$$

Luego se puede escribir la función Log verosimilitud como

$$\log L = \sum_{i=1}^n \sum_{j=0}^{m_i} y_{ij} \log F_{ij}, \quad (3.1.3)$$

lo cual es una generalización natural de

$$\log L = \sum_{i=1}^n y_i \log F(\mathbf{x}'_i \beta) + \sum_{i=1}^n (1 - y_i) \log [1 - F(\mathbf{x}'_i \beta)] \quad (3.1.4)$$

El Estimador de Máxima Verosimilitud $\hat{\theta}$ de θ se define como una solución de la ecuación normal

$$\frac{\partial \log L}{\partial \theta} = \mathbf{0}$$

Muchos de los resultados sobre el Estimador de Máxima Verosimilitud en el caso binario se mantienen aún para el modelo 3.1.1. Un conjunto razonable de condiciones suficientes para su consistencia y normalidad asintótica se pueden encontrar usando los teoremas definidos en la sección inicial de Fundamentos como se hizo para el caso binario. También se puede mostrar la equivalencia del método de scoring y la iteración Mínimos Cuadrados Ponderados No Lineal (NLWLS)¹ (*NLGLS* para ser exacto). Pero aquí se demostrarán éstos conceptos bajo un modelo un poco menos general que 3.1.1. Asumiendo

$$P(y_i = j) = F_j(\mathbf{x}'_{i1}\beta_1, \mathbf{x}'_{i2}\beta_2, \dots, \mathbf{x}'_{iH}\beta_H), \quad (3.1.5)$$

$$i = 1, 2, \dots, n \text{ y } j = 1, 2, \dots, m,$$

donde H es un entero fijo. En la mayoría de los ejemplos específicos del Modelo Multinomial se tiene $H = m$ pero aquí no es necesario. Note que se ha asumido ahora que m no depende de i para simplificar el análisis.

Al replantear los supuestos 2.3.1, 2.3.2 y 2.3.3 con las siguientes modificaciones obvias se obtienen los siguientes supuestos:

Supuesto 3.1.1. F_{ij} tiene derivadas parciales

$$f_{ij}^k \equiv \frac{\partial F_{ij}}{\partial (\mathbf{x}'_{ik}\beta_k)}$$

y derivadas parciales de segundo orden

$$f_{ij}^{kl} \equiv \frac{\partial f_{ij}^k}{\partial (\mathbf{x}'_{il}\beta_l)}$$

para cada i, j, k, l y $0 < F_{ij} < 1$ y $f_{ij}^k > 0$ para cada i, j y k .

¹Nonlinear Weighted Least Squares, Walker y Duncan, 1967

Supuesto 3.1.2. *El espacio de parámetros B es un subconjunto acotado abierto de un espacio Euclideo.*

Supuesto 3.1.3. *$\{\mathbf{x}_{ih}\}$ son uniformemente acotados en i para cada h y*

$$\lim_{n \rightarrow \infty} n^{-1} \sum_{i=1}^n \mathbf{x}_{ih} \mathbf{x}'_{ih}$$

es una matriz finita no singular para cada h . Además la función empírica de distribución de $\{\mathbf{x}_{ih}\}$ converge a una función de distribución.

Bajo estos supuestos el Estimador de Máxima Verosimilitud $\hat{\beta}$ de

$$\beta = (\beta'_1, \beta'_2, \dots, \beta'_H)'$$

puede mostrarse consistente y asintóticamente normal. Para derivar la matriz asintótica de covarianza, se empieza así:

Derivando 3.1.3 con respecto a β_k , se obtiene:

$$\frac{\partial \log L}{\partial \beta_k} = \sum_{i=1}^n \sum_{j=0}^m y_{ij} F_{ij}^{-1} f_{ij}^k \mathbf{x}_{ik} \quad (3.1.6)$$

Derivando 3.1.6 con respecto a β'_l se tiene que :

$$\frac{\partial^2 \log L}{\partial \beta_k \partial \beta'_l} = - \sum_{i=1}^n \sum_{j=0}^m y_{ij} F_{ij}^{-2} f_{ij}^k f_{ij}^l \mathbf{x}_{ik} \mathbf{x}'_{il} + \sum_{i=1}^n \sum_{j=0}^m y_{ij} F_{ij}^{-1} f_{ij}^{kl} \mathbf{x}_{ik} \mathbf{x}'_{il} \quad (3.1.7)$$

Al tomar la esperanza y notando

$$\sum_{j=0}^m f_{ij}^{kl} = 0,$$

se obtiene que

$$- E \frac{\partial^2 \log L}{\partial \beta_k \partial \beta'_l} = \sum_{i=1}^n \sum_{j=0}^m F_{ij}^{-1} f_{ij}^k f_{ij}^l \mathbf{x}_{ik} \mathbf{x}'_{il} \equiv \mathbf{A}_{kl} \quad (3.1.8)$$

Definiendo $\mathbf{A} = \{\mathbf{A}_{kl}\}$. Entonces se puede mostrar bajo los supuestos 3.1.1, 3.1.2 y 3.1.2 que

$$\sqrt{n} (\hat{\beta} - \beta) \rightarrow N \left(0, \lim_{n \rightarrow \infty} n \mathbf{A}^{-1} \right) \quad (3.1.9)$$

Para mostrar las equivalencias entre el método de scoring y el de iteración NLWLS (Minimos Cuadrados Ponderados No Lineal)², es conveniente reescribir (3.1.6) y (3.1.8) usando la siguiente notación vectorial:

$$\begin{aligned} \mathbf{y}_i &= (y_{i1}, y_{i2}, \dots, y_{im})' \\ \mathbf{F}_i &= (F_{i1}, F_{i2}, \dots, F_{im})' \\ \mathbf{f}_i^k &= (f_{i1}^k, f_{i2}^k, \dots, f_{im}^k)' \\ \mathbf{\Lambda}_i &= E (y_i - \mathbf{F}_i)(y_i - \mathbf{F}_i)' = \mathbf{D}(\mathbf{F}_i) - \mathbf{F}_i \mathbf{F}_i' \end{aligned} \quad (3.1.10)$$

donde $\mathbf{D}(\mathbf{F}_i)$ es la matriz diagonal de la cual el j -ésimo elemento diagonal es F_{ij} . Entonces, usando las identidades

$$\sum_{j=0}^m F_{ij} = 1$$

y

$$\sum_{j=0}^m f_{ij}^k = 0$$

y notando:

$$\mathbf{\Lambda}_i^{-1} = \mathbf{D}^{-1}(\mathbf{F}_i) + F_{i0}^{-1} \mathbf{I} \mathbf{I}' \quad (3.1.11)$$

donde \mathbf{I} es un m -vector de unos, se obtiene de (3.1.6) y (3.1.8)

$$\frac{\partial \log L}{\partial \beta_k} = \sum_{i=1}^n \mathbf{x}_{ik} \mathbf{f}_i^{k'} \mathbf{\Lambda}_i^{-1} (\mathbf{y}_i - \mathbf{F}_i) \quad (3.1.12)$$

y

$$- E \frac{\partial^2 \log L}{\partial \beta_k \partial \beta_l'} = \sum_{i=1}^n \mathbf{x}_{ik} \mathbf{f}_i^{k'} \mathbf{\Lambda}_i^{-1} \mathbf{f}_i^l \mathbf{x}_{il}' \quad (3.1.13)$$

Suponga que $\bar{\beta}$ es el estimador inicial de β , sean $\bar{\mathbf{F}}_i$, $\bar{\mathbf{f}}_i^k$ y $\bar{\mathbf{\Lambda}}_i$, respectivamente, \mathbf{F}_i , \mathbf{f}_i^k , y $\mathbf{\Lambda}_i$ evaluados en $\bar{\beta}$ y definiendo

$$\bar{\mathbf{A}}_{kl} = \sum_{i=1}^n \mathbf{x}_{ik} \bar{\mathbf{f}}_i^{k'} \bar{\mathbf{\Lambda}}_i^{-1} \bar{\mathbf{f}}_i^l \mathbf{x}_{il}' \quad (3.1.14)$$

y

²(Ver Amemiya, 1976b)

$$\bar{\mathbf{c}}_k = \sum_{i=1}^n \mathbf{x}_{ik} \bar{\mathbf{f}}_i^{k'} \bar{\Lambda}_i^{-1} \sum_{h=1}^H \bar{\mathbf{f}}_i^h \mathbf{x}'_{ih} \bar{\beta}_h + \sum_{i=1}^n \mathbf{x}_{ik} \bar{\mathbf{f}}_i^{k'} \bar{\Lambda}_i^{-1} (\mathbf{y}_i - \bar{F}_i) \quad (3.1.15)$$

Entonces el estimador de segunda pasada $\bar{\beta}$ obtenido por la iteración de método de scoring iniciado desde $\hat{\beta}$ está dado por

$$\bar{\beta} = \bar{\mathbf{A}}^{-1} \bar{\mathbf{c}}, \quad (3.1.16)$$

donde

$$\bar{\mathbf{A}} = \{ \bar{\mathbf{A}}_{kl} \}$$

y

$$\bar{\mathbf{c}} = \left(\bar{\mathbf{c}}'_1, \bar{\mathbf{c}}'_2, \dots, \bar{\mathbf{c}}'_m \right)' .$$

Es posible interpretar (3.1.16) como una iteración NLGLS. De (3.1.5) se obtiene:

$$\mathbf{y}_i = \mathbf{F}_i + \mathbf{u}_i, \quad i = 1, 2, \dots, n \quad (3.1.17)$$

donde $E\mathbf{u}_i = \mathbf{0}$, $E\mathbf{u}_i \mathbf{u}'_i = \Lambda_i$, y $E\mathbf{u}_i \mathbf{u}'_j = \mathbf{0}$ para $i \neq j$. Expandiendo \mathbf{F}_i en series de Taylor alrededor de $\mathbf{x}'_{ih} \bar{\beta}_h$, $h = 1, 2, \dots, H$, se obtiene

$$\mathbf{y}_i - \bar{\mathbf{F}}_i + \sum_{h=1}^H \bar{\mathbf{f}}_i^h \mathbf{x}'_{ih} \bar{\beta}_h \cong \sum_{h=1}^H \bar{\mathbf{f}}_i^h \mathbf{x}'_{ih} \bar{\beta}_h + \mathbf{u}_i, \quad i = 1, 2, \dots, n \quad (3.1.18)$$

La aproximación 3.1.18 puede referirse como un modelo de regresión lineal de mn observaciones con una matriz de covarianza no escalar. Si se estima Λ_i por $\bar{\Lambda}_i$ y aplicando Estimador Mínimos Cuadrados Factibles Generalizados (FGLS) a 3.1.18, el estimador resultante es precisamente 3.1.16.

Luego, se deriva el estimador MIN χ^2 de β en el modelo 3.1.5, siguiendo a Amemiya (1976b). Ahora, en la sección 2.3.3, se asume que hay muchas observaciones con el mismo valor de variables independientes y asumiendo que el vector de variables independientes toma T valores distintos, la ecuación 3.1.5 se puede escribir como:

$$P_{tj} = F_j \left(\mathbf{x}'_{t1} \beta_1, \mathbf{x}'_{t2} \beta_2, \dots, \mathbf{x}'_{tH} \beta_H \right), \quad (3.1.19)$$

$$t = 1, 2, \dots, T \text{ y } j = 1, 2, \dots, m$$

Para definir el estimador MIN χ^2 de β , se debe poder invertir m ecuaciones en 3.1.19 para cada t y resolverlas para H variables $\mathbf{x}'_{t1} \beta_1, \mathbf{x}'_{t2} \beta_2, \dots, \mathbf{x}'_{tH} \beta_H$. Asumiendo por el momento que $H = m$.

Si además, se asume que el Jacobiano no desaparece, se obtiene

$$\begin{aligned} \mathbf{x}'_{tk}\beta_k &= G_k(P_{t1}, P_{t2}, \dots, P_{tm}), \\ t &= 1, 2, \dots, T \text{ y } k = 1, 2, \dots, m \end{aligned} \quad (3.1.20)$$

Como en la sección 2.3.3, se define

$$r_{tj} = \sum_{i \in I_t} y_{ij},$$

donde I_t es el conjunto de i para el cual

$$\mathbf{x}_{ih} = \mathbf{x}_{th}$$

para todo h , y

$$\hat{P}_{tj} = \frac{r_{tj}}{n_t},$$

donde n_t es el número de enteros en I_t . Expandiendo

$$G_k(\hat{P}_{t1}, \hat{P}_{t2}, \dots, \hat{P}_{tm})$$

en series de Taylor alrededor de

$$(P_{t1}, P_{t2}, \dots, P_{tm})$$

y usando 3.1.20, se obtiene

$$\begin{aligned} G_k(\hat{P}_{t1}, \hat{P}_{t2}, \dots, \hat{P}_{tm}) &\cong \mathbf{x}'_{tk}\beta_k + \sum_{j=1}^m g_{tk}^j (\hat{P}_{tj} - P_{tj}), \\ t &= 1, 2, \dots, T \text{ y } k = 1, 2, \dots, m \end{aligned} \quad (3.1.21)$$

donde

$$g_{tk}^j = \frac{\partial G_k}{\partial P_{tj}}.$$

La ecuación 3.1.21 es una generalización de 2.3.22, pero acá se escribe como una ecuación aproximada, ignorando un término de error que corresponde a w_t en 2.3.22.

La ecuación 3.1.21 es una ecuación de regresión lineal aproximada con una matriz de covarianza no escalar que depende de

$$\mathbf{\Lambda}_t \equiv E(\hat{\mathbf{P}}_t - \mathbf{F}_t)(\hat{\mathbf{P}}_t - \mathbf{F}_t)' = n_t^{-1} [\mathbf{D}(\mathbf{F}_t) - \mathbf{F}_t \mathbf{F}_t'],$$

donde

$$\hat{\mathbf{P}}_t = (\hat{P}_{t1}, \hat{P}_{t2}, \dots, \hat{P}_{tm})'$$

y

$$\mathbf{F}_t = (F_{t1}, F_{t2}, \dots, F_{tm})'$$

El estimador MIN χ^2 $\tilde{\beta}$ de β se define como Estimador Mínimos Cuadrados Factibles Generalizados (FGLS) aplicado a 3.1.21, usando

$$\hat{\Lambda}_t = n_t^{-1} [\mathbf{D}(\hat{P}_t) - \hat{\mathbf{P}}_t \hat{\mathbf{P}}_t']$$

como un estimador de Λ_t .

La consistencia y normalidad asintóticamente de $\tilde{\beta}$ puede probarse usando un método similar al empleado en el 2.3.22. Sea Ω^{-1} la matriz de covarianza asintótica de $\tilde{\beta}$, esto es,

$$\sqrt{n}(\tilde{\beta} - \beta) \rightarrow N(\mathbf{0}, \lim_{n \rightarrow \infty} n\Omega^{-1}).$$

Entonces se puede deducir de 3.1.21 que el k, l -ésimo sub-bloque de Ω está dado por

$$\Omega_{kl} = \sum_t \mathbf{x}_{tk} \left\{ \underline{\mathbf{G}}_t' \Lambda_t \underline{\mathbf{G}}_t \right\}_{kl}^{-1} \mathbf{x}'_{tl}, \quad (3.1.22)$$

donde $\underline{\mathbf{G}}_t'$ es una matriz $m \times m$ cuya k -ésima fila es igual a

$$(g_{tk}^1, g_{tk}^2, \dots, g_{tk}^m)$$

y $\{\}_{kl}^{-1}$ denota el k, l -ésimo elemento de la inversa de la matriz dentro de $\{\}$.

Ahora, se obtiene de la ecuación 3.1.13 que:

$$\mathbf{A}_{kl} = \sum_t \mathbf{x}_{tk} \left\{ \underline{\mathbf{F}}_t' \Lambda_t^{-1} \underline{\mathbf{F}}_t \right\}_{kl} \mathbf{x}'_{tl}, \quad (3.1.23)$$

donde $\underline{\mathbf{F}}_t'$ es una matriz $m \times m$ cuya k -ésima fila es igual a

$$(f_{tk}^1, f_{tk}^2, \dots, f_{tk}^m).$$

Así la equivalencia asintótica del estimador MIN χ^2 y el Estimador de Máxima Verosimilitud se sigue de la identidad

$$\underline{\mathbf{G}}_t^{-1} = \underline{\mathbf{F}}_t'$$

En el desarrollo anterior se asumió $H = m$.

Asumiendo $H < m$, se puede invertir la ecuación 3.1.19 y obtener

$$\begin{aligned} \mathbf{x}'_{tk}\beta_k &= G_k(P_{t1}, P_{t2}, \dots, P_{tm}), \\ t &= 1, 2, \dots, T \text{ y } k = 1, 2, \dots, H, \end{aligned} \tag{3.1.24}$$

Pero la selección de la función G_k no es única. Amemiya (1976b) mostró que el estimador MIN χ^2 es asintóticamente eficiente sólo cuando se escoge la función G_k correcta dentro de las muchas posibles. Este hecho disminuye la utilidad del método. Para este caso Amemiya (1977c) propuso el siguiente método, el cual siempre conduce a un estimador asintóticamente eficiente:

Paso 1: Se usa alguna G_k en (3.1.24) y defina

$$\hat{\mu}_{tk} = G_k\left(\hat{P}_{t1}, \hat{P}_{t2}, \dots, \hat{P}_{tm}\right)$$

Paso 2: En (3.1.18) se reemplaza $\bar{\mathbf{F}}_i$ y $\bar{\mathbf{f}}_i^h$ por \mathbf{F}_i y \mathbf{f}_i^h evaluado en $\hat{\mu}_{tk}$ y se reemplaza $x'_{ih}\bar{\beta}_h$ por $\hat{\mu}_{tk}$

Paso 3: Se aplica FGLS en (3.1.18) usando $\mathbf{\Lambda}_i$ evaluado en $\hat{\mu}_{tk}$.

Se concluye esta subsección generalizando la Suma Ponderada de Residuos Cuadrados³ definido por los modelos 2.4.1 y 2.4.4 al caso multinomial.

No se escribirá el análogo de 2.4.1 explícitamente porque requeriría una complicada notación aunque la idea es sencilla. A cambio, se afirmará que es de la forma

$$\left(y - \mathbf{X}\hat{\beta}_G\right) \hat{\Sigma}^{-1} \left(y - \mathbf{X}\hat{\beta}_G\right)$$

obtenido de la ecuación de regresión 3.1.21.

Esta se distribuye asintóticamente chi-cuadrado con grados de libertad igual a mT menos el número de parámetros de regresión.

Ahora, considerando la generalización de 2.4.4. Se define el vector $\tilde{F}_t = F_t\left(\tilde{\beta}\right)$. Entonces, el análogo de 2.4.4 es

³Weighted Sum of Squared Residuals (WSSR)

$$\begin{aligned}
 WSSR &= \sum_{t=1}^T \left(\hat{\mathbf{P}}_t - \tilde{\mathbf{F}}_t \right)' \hat{\Lambda}_t^{-1} \left(\hat{\mathbf{P}}_t - \tilde{\mathbf{F}}_t \right) \\
 &= \sum_{t=1}^T n_t \left(\hat{\mathbf{P}}_t - \tilde{\mathbf{F}}_t \right)' \left[\mathbf{D}(\hat{\mathbf{P}}_t)^{-1} + \hat{P}_{t0}^{-1} \mathbf{I} \mathbf{I}' \right] \left(\hat{\mathbf{P}}_t - \tilde{\mathbf{F}}_t \right) \\
 &= \sum_{t=1}^T n_t \sum_{j=0}^m \frac{(\hat{P}_{tj} - \tilde{F}_{tj})^2}{\hat{P}_{tj}}
 \end{aligned} \tag{3.1.25}$$

Este es asintóticamente equivalente al análogo de 2.4.1.

3.2. Análisis Discriminante Multinomial (DA)

El modelo del análisis discriminante analizado en la sección 2.5 sobre Modelos de Elección Binaria se puede generalizar para producir análisis discriminante multinomial definido por

$$\mathbf{x}_i^* | (y_i = j) \sim N(\mu_j, \Sigma_j) \tag{3.2.1}$$

y

$$P(y_i = j) = q_j \tag{3.2.2}$$

para $i = 1, 2, \dots, n$ y $j = 0, 1, \dots, m$. Mediante la Regla de Bayes se obtiene

$$P(y_i = j | \mathbf{x}_i^*) = \frac{g_j(\mathbf{x}_i^*) q_j}{\sum_{k=0}^m g_k(\mathbf{x}_i^*) q_k} \tag{3.2.3}$$

donde g_j es la función de densidad de $N(\mu_j, \Sigma_j)$. Tal como se obtuvo el modelo (2.5.3) del modelo (2.5.1) se puede obtener de (3.2.3)

$$\frac{P(y_i = j | \mathbf{x}_i^*)}{P(y_i = 0 | \mathbf{x}_i^*)} = \Lambda(\boldsymbol{\beta}_{j(1)} + \boldsymbol{\beta}'_{j(2)} \mathbf{x}_i^* + \mathbf{x}_i^{*'} \mathbf{A} \mathbf{x}_i^*) \tag{3.2.4}$$

donde $\boldsymbol{\beta}_{j(1)}$, $\boldsymbol{\beta}_{j(2)}$, y \mathbf{A} son similares a (2.5.3), (2.5.4) y (2.5.5) excepto que los subíndices 1 y 0 se deben cambiar a j y 0 respectivamente.

De igual manera, el término $\mathbf{x}_i^{*'} \mathbf{A} \mathbf{x}_i^*$ desaparece si todos los Σ 's son idénticos. Si se escribe

$$\boldsymbol{\beta}_{j(1)} + \boldsymbol{\beta}'_{j(2)} \mathbf{x}_i^* = \boldsymbol{\beta}'_j \mathbf{x}_i,$$

el modelo Análisis Discriminante con varianzas idénticas se puede escribir exactamente en la forma de 3.3.25, excepto por una modificación de los subíndices de β y \mathbf{x} ⁴.

3.3. Principales Modelos de Respuesta Múltiple

3.3.1. Introducción

Los modelos de elección dicotómica o binaria se pueden generalizar para el caso de más de dos opciones, dando origen a los **Modelos de Respuesta Múltiple**, donde se generaliza el proceso de elección de tal forma que el agente económico se enfrenta a varias alternativas posibles.

Estas alternativas se pueden presentar de forma ordenada, no ordenada o secuencial, dando origen a un tratamiento específico de la variable endógena del modelo y, por consiguiente, a diferentes tipos de modelación. Surgen así los modelos denominados

- **Modelos de Respuesta Ordenada** Surgen cuando las decisiones de los individuos pueden ordenarse.

- **Modelos de Respuesta No Ordenada.**

Es el modelo donde el proceso de elección no implica ninguna ordenación. Entre los modelos de Respuesta no Ordenada se destacan:

- **Modelo Logit Multinomial o Modelo Logit Simple(MNL)**
- **Modelo Probit Multinomial (MNP)**
- **Modelo Multinomial Logit Jerárquico o Anidado (NL)**

Los Modelos de Respuesta Múltiple, analizan la elección que un individuo realiza en función de un conjunto de variables explicativas o regresores. Éstas pueden ser de dos tipos:

- Las que se refieren a las **Características Propias del Individuo**, por ejemplo: la edad, nivel de renta, distancia al puesto de trabajo, etc., y
- Las que se refieren a los **Aspectos Específicos de cada Alternativa concreta**, por ejemplo: tiempo empleado en el trayecto según se vaya a pie, en automóvil o en autobús, costo monetario del medio al realizar el desplazamiento, el precio de la leche según la marca elegida, el precio del bien o servicio según la elección realizada.

⁴Los ejemplos de los modelos multinomiales se encuentran en los artículos de Powers y otros(1978) y Uhler(1968), los cuales se resumen por Amemiya(1981)

La variable dependiente o regresando se suele construir asignando el valor 0 a la primera alternativa, 1 a la segunda, y así sucesivamente, hasta el número de alternativas o categorías menos uno. Así, en el caso del problema de decidir cómo ir a trabajar, se asigna el valor 0 a la opción ir a pie, 1 a la opción ir en automóvil y 2 a la alternativa ir en autobús, siendo en total tres las alternativas entre las que el agente debe decidir.

El enfoque teórico de este tipo de modelos se fundamenta en la teoría de la utilidad del agente económico. En este caso, se supone que el agente económico es racional y que elige la alternativa u opción que le va a brindar una mayor utilidad. Además, el modelo se puede concebir como un problema de decisión, en el sentido de que se debe elegir una opción entre un conjunto de M alternativas, que se plantea en los siguientes términos:

Suponiendo que $U_{i0}, U_{i1}, \dots, U_{im}, \dots, U_{i(M-1)}$ representan las utilidades de las M alternativas⁵ para el individuo i -ésimo, las variables $X_{i0}^*, X_{i1}^*, \dots, X_{im}^*, \dots, X_{i(M-1)}^*$ son el conjunto de características propias de la elección tal y como las percibe el individuo (precio del trayecto, duración del desplazamiento, etc) y X_i^{**} es el conjunto de características personales del individuo (edad, nivel de renta, etc). Se supone, además, linealidad en las funciones⁶ de tal forma que la especificación del modelo sería:

$$\begin{aligned}
 U_{i0} &= V_{i0} + \epsilon_{i0} = \beta_0 + X_{i0}^* \delta + X_i^{**} \gamma_0 + \epsilon_{i0} & (3.3.1) \\
 U_{i1} &= V_{i1} + \epsilon_{i1} = \beta_1 + X_{i1}^* \delta + X_i^{**} \gamma_1 + \epsilon_{i1} \\
 &\dots \\
 U_{im} &= V_{im} + \epsilon_{im} = \beta_m + X_{mj}^* \delta + X_i^{**} \gamma_m + \epsilon_{im} \\
 &\dots \\
 U_{i(M-1)} &= V_{i(M-1)} + \epsilon_{i(M-1)} = \beta_{(M-1)} + X_{i(M-1)}^* \delta + X_i^{**} \gamma_{(M-1)} + \epsilon_{i(M-1)}
 \end{aligned}$$

El individuo elige una determinada opción si la utilidad que le proporciona dicha alternativa es mayor que la utilidad que le proporciona el resto de alternativas. Es decir:

⁵Se refiere a diferenciales de utilidad, no a niveles

⁶El supuesto de linealidad es importante, puesto que si se reemplaza U_i por $\log U_i$ los resultados no son los mismos

$$Y_i = \begin{cases} 0 & \text{si } U_{i0} > U_{im} \forall m \neq 0 \\ 1 & \text{si } U_{i1} > U_{im} \forall m \neq 1 \\ 2 & \text{si } U_{i2} > U_{im} \forall m \neq 2 \\ \dots & \\ (M-1) & \text{si } U_{i(M-1)} > U_{im} \forall m \neq (M-1) \end{cases} \quad (3.3.2)$$

Un enfoque alternativo para plantear el problema de la elección entre múltiples alternativas es el enfoque de la variable latente. En él se supone la existencia de una variable inobservable o latente no limitada en su rango de variación, Y_i^* , que depende de las características propias de la decisión, X_i^* , o del individuo X_i^{**} . Sobre esta variable latente se aplica una regla de observabilidad que genera las alternativas que se observan en la realidad. Desde este punto de vista, los valores de la variable real u observada Y_i que mide las distintas categorías se fundamentan de acuerdo con el siguiente patrón o esquema:

$$Y_i = \begin{cases} 0 & \text{si } Y_i^* \leq c_1 \\ 1 & \text{si } c_1 \leq Y_i^* \leq c_2 \\ \dots & \\ (M-1) & \text{si } c_{(M-1)} \geq Y_i^* \end{cases} \quad (3.3.3)$$

donde c_1, c_2, \dots, c_{M-1} son los valores de los umbrales o barreras. Formalmente se puede expresar el modelo de respuesta múltiple a través de la relación siguiente:

$$Y_i^* = F(X_i\beta) + u_i = F(Z_i + u_i)$$

donde:

- Y_i^* : Es una variable latente (no observada)
- $F(\cdot)$: Es una función no lineal de una combinación lineal de las características o índice
- $Z_i : X_i\beta$ es el índice del modelo
- u_i : Es una variable aleatoria

Por tanto, en el proceso de especificación de los modelos de respuesta múltiple influyen los siguientes elementos:

1. El tipo de función $F(\cdot)$ elegida: Generalmente se utilizan la función de distribución Logística, la función de distribución de la Normal y la función Valor Extremo o función de Gompit.

2. Los diferentes criterios de elección ante los que se presenta el individuo: Es decir, si el problema presentado es un proceso de elección ordenado, una elección ordenada jerarquizada o no ordenada.
3. El tipo de variables a considerar en el estudio o características (regresores): Es necesario determinar aquellas que son propias del individuo y que, por tanto, son comunes a todas las alternativas (edad, renta, nivel de estudios, etc.) y las características específicas de cada una de las alternativas (precio, tiempo, distancia, etc.). En la elección a realizar intervendrán ambos tipos de características. Así, en la elección del medio de transporte influye tanto el nivel de renta del usuario como la velocidad o tiempo de duración del desplazamiento, además del precio del tiquete, en su caso.

3.3.2. Modelos de Respuesta Múltiple Ordenado

Definición

Una definición general de un modelo ordenado es

$$P(y = j | \mathbf{x}, \boldsymbol{\theta}) = p(S_j)$$

para alguna medida de probabilidad p dependiendo de \mathbf{x} y $\boldsymbol{\theta}$ y una secuencia finita de intervalos sucesivos $\{S_j\}$ dependiendo de \mathbf{x} y $\boldsymbol{\theta}$ tal que $\cup_j S_j = R$, donde R es la recta real.

En el modelo ordenado los valores que toma corresponden a una partición de la recta real, mientras en el no ordenado corresponden bien sea a una partición no sucesiva de la recta real o a una partición de un espacio Euclideo de mayor dimensión.

En la mayoría de las aplicaciones el modelo ordenado es de la forma

$$P(y = j | \mathbf{x}, \boldsymbol{\alpha}, \boldsymbol{\theta}) = F(\boldsymbol{\alpha}_{j+1} - \mathbf{x}'\boldsymbol{\theta}) - F(\boldsymbol{\alpha}_j - \mathbf{x}'\boldsymbol{\theta}), j = 0, 1, \dots, m, \quad (3.3.4)$$

$$\alpha_0 = -\infty, \alpha_j \leq \alpha_{j+1}, \alpha_{m+1} = \infty,$$

para alguna función de distribución F . Si $F = \phi$, 3.3.4 define el modelo probit ordenado; y si $F = \Lambda$, define el modelo Logit ordenado.⁷

⁷Pratt (1981) mostró que la función log-verosimilitud del modelo 3.3.4 basada en observaciones $(y_i, \mathbf{x}_i), i = 1, 2, \dots, n$, en (y, \mathbf{x}) es globalmente cóncava si f , derivada de F , es positiva y $\log f$ es cóncava

El modelo 3.3.4 se deduce considerando una variable aleatoria continua no observada y^* que determina el resultado de y por la regla

$$y = j \text{ si y sólo si } \alpha_j < y^* < \alpha_{j+1}, \quad j = 0, 1, \dots, m \quad (3.3.5)$$

Si la función de distribución de $y^* - \mathbf{x}'\beta$ es F , el modelo 3.3.5 implica 3.3.4.

Especificación

El modelo de respuesta múltiple ordenado relaciona la variable Y_i con las variables, X_{2i}, \dots, X_{ki} a través de la siguiente ecuación:

$$Y_i^* = F(X_i\beta) + u_i = F(Z_i + u_i)$$

Donde:

- Y_i^* : Es una variable latente (no observada) que cuantifica las distintas categorías
- $F(\cdot)$: Es una función no lineal del tipo Logística, distribución Normal o bien la distribución Valor Extremo
- $X_i\beta$: Es una combinación lineal de las variables o características que se denomina índice del modelo y se denota por Z_i
- u_i : Es una variable aleatoria

Para los valores de la variable real u observada, Y_i , que mide las distintas categorías, en el caso de cuatro categorías, el esquema de la variable se fundamenta de acuerdo con el siguiente patrón:

$$Y_i = \begin{cases} 0 & \text{si } Y_i^* \leq c_1 \\ 1 & \text{si } c_1 \leq Y_i^* \leq c_2 \\ 2 & \text{si } c_2 \leq Y_i^* \leq c_3 \\ 3 & \text{si } c_3 \geq Y_i^* \end{cases} \quad (3.3.6)$$

Se debe subrayar que existe una correspondencia entre el orden de los valores de la variable real u observada y el orden de la variable latente, es decir, que si $Y_i < Y_j$ implica que $Y_i^* < Y_j^*$.

Desde el punto de vista gráfico, un modelo ordenado en el que se contemplan cuatro categorías (0, 1, 2 y 3) y tan sólo interviene un único regresor o característica se representan en el eje de abscisas los valores que toman las características, mientras que en el eje de ordenadas se representan las tres categorías.

La probabilidad de elegir cada una de las categorías de Y_i viene definida por la siguiente relación:

$$\begin{aligned}
 P(Y_i = 0/X_i, \boldsymbol{\beta}, c) &= F(c_1 - X_i\boldsymbol{\beta}) & (3.3.7) \\
 P(Y_i = 1/X_i, \boldsymbol{\beta}, c) &= F(c_2 - X_i\boldsymbol{\beta}) - F(c_1 - X_i\boldsymbol{\beta}) \\
 P(Y_i = 2/X_i, \boldsymbol{\beta}, c) &= F(c_3 - X_i\boldsymbol{\beta}) - F(c_2 - X_i\boldsymbol{\beta}) \\
 P(Y_i = 3/X_i, \boldsymbol{\beta}, c) &= 1 - F(c_3 - X_i\boldsymbol{\beta})
 \end{aligned}$$

Donde $F(\cdot)$ es la función de distribución o de densidad acumulada de la ecuación elegida en la especificación del modelo.

Los valores de los umbrales o barreras c_m y los valores de $\boldsymbol{\beta}$ se estiman conjuntamente mediante el método de la Máxima Verosimilitud, y se debe cumplir la siguiente restricción:

$$c_1 < c_2 < c_3$$

En el caso particular que la función de distribución elegida sea la de la normal, es decir, que el modelo especificado objeto del análisis sea el Probit Ordenado, la probabilidad de elegir cada una de las categorías de Y_i se puede reescalar y vendrá definida por la siguiente relación:

$$\begin{aligned}
 P(Y_i = 0/X_i, \boldsymbol{\beta}, c) &= \Phi(c_1 - X_i\boldsymbol{\beta}) & (3.3.8) \\
 P(Y_i = 1/X_i, \boldsymbol{\beta}, c) &= \Phi(c_2 - X_i\boldsymbol{\beta}) - \Phi(c_1 - X_i\boldsymbol{\beta}) \\
 P(Y_i = 2/X_i, \boldsymbol{\beta}, c) &= \Phi(c_3 - X_i\boldsymbol{\beta}) - \Phi(c_2 - X_i\boldsymbol{\beta}) \\
 P(Y_i = 3/X_i, \boldsymbol{\beta}, c) &= 1 - \Phi(c_3 - X_i\boldsymbol{\beta})
 \end{aligned}$$

Interpretación

La interpretación de los Modelos Ordenados es análoga a la de los modelos Logit, Probit y Valor Extremo. Así, la interpretación de los parámetros del Modelo Ordenado se puede efectuar a través de las derivadas parciales, en concreto a través de los efectos marginales, mientras que la comparación entre distintas situaciones se puede efectuar a través del cociente *odds*.

Únicamente se interpretará el Modelo Ordenado Probit ya que la explicación del Modelo Ordenado Logit y el Modelo Ordenado Valor Extremo es análogo.

1. La interpretación de los parámetros

La interpretación de los parámetros se puede efectuar a través de las derivadas parciales para los distintos modelos estudiados; así, en el caso del Modelo

Ordenado Probit descrito a través del sistema 3.3.8, el efecto marginal del regresor para cada categoría es:

$$\begin{aligned}
 \frac{\partial P(Y_i = 0)}{\partial X_k} &= -\phi(c_1 - X_i\beta)\beta_k & (3.3.9) \\
 \frac{\partial P(Y_i = 1)}{\partial X_k} &= -\phi(c_2 - X_i\beta)\beta_k - \phi(c_1 - X_i\beta)\beta_k \\
 \frac{\partial P(Y_i = 2)}{\partial X_k} &= -\phi(c_3 - X_i\beta)\beta_k + \phi(c_2 - X_i\beta)\beta_k \\
 \frac{\partial P(Y_i = 3)}{\partial X_k} &= -\phi(c_3 - X_i\beta)\beta_k
 \end{aligned}$$

En general, los coeficientes estimados en los Modelos Ordenados no cuantifican directamente el incremento en la probabilidad dado el aumento unitario en la correspondiente variable independiente. La magnitud de la variación en la probabilidad depende del nivel original de ésta y, por tanto, de los valores iniciales de todos y cada uno de los regresores y de sus coeficientes. Por tanto, mientras el signo de los coeficientes sí que indica perfectamente la dirección del cambio, la magnitud de la variación depende del valor concreto que tome la función de densidad, lo cual depende de la pendiente de dicha función en el punto X_i (X_i es igual al vector $[X_{2i}, \dots, X_{ki}]$). Naturalmente, cuanto más elevada sea dicha pendiente, mayor será el impacto del cambio en el valor del regresor, el cual a su vez va a incidir en el valor de la variable latente Y_i^* .

2. Interpretación de *La ratio Odds o riesgo*

El cociente entre la utilidad de que se elija una opción frente al resto de alternativas, $(M - 1)$ opciones, se mide a través de la probabilidad asignada a esta frente a la utilidad del resto de opciones y se le denomina odds. Que se cuantifica mediante la siguiente proporción:

$$\frac{P_i}{1 - P_i}$$

El *odds* también, se utiliza para comparar situaciones distintas o bien para interpretar el modelo. En efecto, el cociente entre *odds* se emplea para comparar utilidades de distintas situaciones, o bien para interpretar el modelo. Así, en el caso, que se compare la situación del individuo i con la del individuo j (situación de referencia) se tiene:

$$\text{Cociente entre odds} = \frac{\frac{P_i}{(1-P_i)}}{\frac{P_j}{(1-P_j)}} \quad (3.3.10)$$

En el caso de que el valor de la **Ratio Odds** sea:

- Mayor que Uno: La utilidad de la situación i es más elevada que la de la situación j .
- Menor que Uno: La utilidad de la situación j es más elevada que la de la situación i .
- La unidad : Las utilidades de las situaciones i y j son iguales.

Estimación Máximo-Verosimilitud

La función de probabilidad conjunta de un modelo de elección múltiple ordenado parte del supuesto de que se dispone de una muestra de tamaño I ($i = 1, 2, \dots, I$) bajo la hipótesis de independencia entre los distintos individuos y se puede expresar a través de la siguiente relación:

$$Prob(u_1, u_2, \dots, u_i, \dots, u_I) = Prob(u_1)Prob(u_2) \dots Prob(u_i) \dots Prob(u_I) = \prod_{i=1}^I Prob(u_i) \quad (3.3.11)$$

De forma alternativa, utilizando la función de probabilidad en términos de la variable Y_i , se tiene:

$$Prob(Y_1, Y_2, \dots, Y_i, \dots, Y_I) = Prob(Y_1)Prob(Y_2) \dots Prob(Y_i) \dots Prob(Y_I) = \prod_{i=1}^I Prob(Y_i) \quad (3.3.12)$$

teniendo en cuenta que Y_i puede tomar el valor cero, uno, dos, tres, ..., se obtiene:

$$\prod_{i \in Y_i=0} Prob(Y_i = 0) \prod_{i \in Y_i=1} Prob(Y_i = 1) \prod_{i \in Y_i=2} Prob(Y_i = 2) \prod_{i \in Y_i=3} Prob(Y_i = 3) \dots \quad (3.3.13)$$

a partir de la función de la probabilidad conjunta 3.3.13 se puede obtener la función de verosimilitud, que se define como:

$$L = \prod_{i \in Y_i=0} Prob(Y_i = 0) \prod_{i \in Y_i=1} Prob(Y_i = 1) \prod_{i \in Y_i=2} Prob(Y_i = 2) \prod_{i \in Y_i=3} Prob(Y_i = 3) \dots \quad (3.3.14)$$

mientras que el logaritmo de la función de verosimilitud es:

$$\mathcal{L} = \ln L = \sum_{i \in Y_i=0} Prob(Y_i = 0) + \sum_{i \in Y_i=1} Prob(Y_i = 1) + \sum_{i \in Y_i=2} Prob(Y_i = 2) \dots \quad (3.3.15)$$

Al sustituir en 3.3.15 la probabilidad de cada opción, $\text{Prob}(Y_i = \dots)$, por distintas especificaciones se obtiene el logaritmo de la función de verosimilitud del Modelo Ordenado. Las funciones más utilizadas en la especificación son la Logística y la de distribución Normal obteniendo los Modelos Ordenado Logit y Ordenado Probit respectivamente.

El logaritmo de la función de verosimilitud, en particular la ecuación 3.3.15, no es lineal, con lo cual para obtener los estimadores máximo-verosímiles de los parámetros se deben aplicar métodos de estimación no lineales a través de algún algoritmo de optimización.

Los estimadores obtenidos por el proceso de máxima verosimilitud son consistentes y asintóticamente eficientes. Igualmente son asintóticamente normales, con lo que los contrastes de hipótesis son asintóticos. Así, cuando \mathbf{I} (tamaño de la muestra) tiende a infinito, el contraste de significación individual de los parámetros se puede realizar a través de una distribución normal. En este caso, todos los tests propuestos para estudiar la bondad del ajuste en los modelos de variable dicotómica son válidos para los Modelos Ordenados.

3.3.3. Modelos de Respuesta Múltiple No Ordenados

3.3.4. Definición

Los modelos de elección múltiple entre alternativas no ordenadas pueden venir generados por modelos de utilidad aleatoria, modelos que suponen que los individuos son agentes racionales, que disponen de información perfecta, y que se enfrentan a un conjunto de alternativas a las que les asocian una utilidad. Ahora bien, desde el punto de vista del investigador, esta utilidad no es directamente observable, descomponiéndose en dos componentes, una parte observable V_{nj} , que dependerá de un conjunto de atributos medibles para cada individuo y alternativa, y una aleatoria ϵ_{nj} . Una formulación habitual es la del modelo de utilidad aleatoria aditivo:

$$U_{jn} = V_{jn} + \epsilon_{jn}, j = 1, \dots, J \quad (3.3.16)$$

donde U_{jn} es la utilidad que al individuo n -ésimo le reporta la alternativa j , y, J es el número de alternativas disponibles.

Un individuo escogerá siempre la alternativa que le proporcione mayor utilidad, de modo que si el individuo n -ésimo selecciona la alternativa j , es porque la utilidad reportada U_{jn} es la mayor de todas.

$$U_{nj} \geq U_{nk} \Leftrightarrow V_{nj} - V_{nk} \geq \epsilon_{nk} - \epsilon_{nj}, \forall k \neq j, k = 1, \dots, J \quad (3.3.17)$$

La decisión observada revela cuál de las alternativas proporciona mayor utilidad, pero no sus utilidades, que no son observables.

Dado que no se conoce con exactitud el componente aleatorio, y por tanto, no se puede determinar con certeza si (ecuación anterior) se cumple, se debe pasar a un marco probabilístico. Así, la probabilidad de que el individuo n escoja la alternativa j viene dada por

$$P(Y_n = j) = P_{nj} = P(\epsilon_{nk} \leq \epsilon_{nj} + (V_{nj} - V_{nk})), \forall k \neq j, k = 1, \dots, J \quad (3.3.18)$$

Dependiendo de la especificación del componente aleatorio ϵ_{nj} y de la parte medible V_{nj} se pueden generar distintos modelos. Dentro de los modelos de elección discreta los más utilizados son:

- El Logit Multinomial (Mc Fadden (1974)), que se deriva a partir de asumir que los términos de error ϵ_{nj} presentan una distribución Gumbel homoscedástica e independiente, y
- El Logit Jerárquico o Anidado (Williams (1977)), que desarrollado como una extensión del anterior, en el que se considera una partición disjunta de las alternativas en grupos de tal forma que las alternativas incluidas en cada grupo presentan correlación en sus términos de error; siguiendo estos también una distribución Gumbel.

Sin embargo, ambos modelos poseen supuestos simplificadores que no siempre son sostenibles. Su principal limitación es la propiedad conocida con el nombre de **Independencia de Alternativas Irrelevantes** y que consiste en que el cociente de probabilidades entre dos alternativas no depende de las características o atributos del resto de las alternativas.

En el modelo Logit Jerárquico al agrupar las alternativas en subgrupos de características similares, esta propiedad se mantiene entre las alternativas de un mismo subgrupo, pero no entre ellos. Esta propiedad aunque simplifica el proceso de estimación, supone una restricción importante en la modelación del comportamiento de los individuos que no parece muy razonable en numerosas situaciones, puesto que la hipótesis subyacente es que la elección entre dos alternativas cualesquiera no depende de las características o atributos de una tercera.

Los diferentes modelos de comportamiento de **maximización de la utilidad aleatoria** se diferencian principalmente en la especificación que suponen para la densidad de los factores no observados $f(\epsilon_n)$.

Los principales modelos de elección discreta son:

- Logit (MNL): errores Gumbel independientes e idénticos (iid.). Por lo tanto no admite correlación entre alternativas ni entre elecciones y requiere homoscedasticidad (varianzas iguales para todas las alternativas y para todas las elecciones).
- Probit (MNP): se supone que los factores no observados distribuyen conjuntamente normal: $(\epsilon_{n1}, \epsilon_{n2}, \dots, \epsilon_{nJ}) \sim N(0, \Omega)$. Con una matriz de covarianza Ω completa, el modelo puede reflejar cualquier patrón de correlación y admite heteroscedasticidad. Sus dificultades estriban en la necesidad de simulación para el cálculo de las integrales y en el hecho de que sólo admite la distribución normal para los coeficientes, cuando en ocasiones puede ser oportuno establecer otras distribuciones.
- Modelos de Valor Extremo Generalizados (GEV): se admiten patrones de correlación entre alternativas. Uno de los más sencillos y utilizados es el Logit jerárquico o anidado (NL), en el que las alternativas se estructuran en grupos llamados nidos, de modo que los factores no observados tienen la misma correlación para las alternativas en el interior del nido y no existe correlación con las restantes. En general no requieren simulación para su estimación.

3.3.5. Modelo Logit Multinomial

Si consideramos la hipótesis de que los errores distribuyen idéntica e independientemente Gumbel con media cero y varianza σ^2 , obtenemos el **Modelo Logit Multinomial o Logit Simple (MNL)**. Es decir, El modelo Logit (o Logit Multinomial, MNL) se obtiene suponiendo que cada ϵ_{nj} distribuye según una distribución de Gumbel, o valor extremo tipo I ⁸, independiente e idéntica (es decir, con la misma varianza): $\epsilon_{nj} \sim \mathbf{Gumbel iid}$.

La elección de una distribución de Gumbel no se debe a la sencillez de la fórmula Logit obtenida, sino que se obtiene al tomar en consideración que los residuos de cada alternativa no son interesantes en sí mismos, sino que sólo interesa el máximo de las utilidades aleatorias. Bajo condiciones poco restrictivas (residuos con distribución exponencial, normal, Weibull o Gumbel), el máximo entre alternativas se distribuye según la distribución de Gumbel⁹ o valor extremo tipo I con parámetro de localización 0.

Si una variable ϵ se distribuye según una distribución de Gumbel¹⁰ de parámetro

⁸Gumbel, 1958

⁹véase Sørensen, 2003: 19

¹⁰Las principales propiedades de esta distribución se pueden consultar en Ben-Akiva y Lerman (1985: 104-105)

de localización η y parámetro de escala μ , su función de distribución es:

$$F(\epsilon) = \exp \left[-e^{-\mu}(\epsilon - \eta) \right], \mu > 0 \quad (3.3.19)$$

y su función de densidad es:

$$F(\epsilon) = \mu e^{-\mu}(\epsilon - \eta) \exp \left[-e^{-\mu}(\epsilon - \eta) \right] \quad (3.3.20)$$

La moda de la distribución es η , la media es $\eta + \gamma/\mu$, siendo γ la constante de Euler ($\sim 0,577$), la varianza¹¹ es $\pi^2/6\mu^2$.

En el modelo Logit se emplea la distribución Gumbel con parámetro de localización cero. Debe señalarse que, aunque en muchas ocasiones se afirma que los errores tienen media nula, en realidad su media es γ/μ (al ser una distribución sesgada), si bien este valor no es observable al trabajarse siempre con diferencias entre utilidades, y la diferencia entre dos términos aleatorios con la misma media tiene media cero.

La diferencia entre dos distribuciones de Gumbel, con parámetro de escala μ y parámetro de localización nulo, distribuye logística, es decir,

$$\epsilon_{nji}^* = \epsilon_{nj} - \epsilon_{ni}$$

tiene la función de distribución:

$$F(\epsilon_{nji}^*) = \frac{e^{\mu\epsilon_{nji}^*}}{1 + e^{\mu\epsilon_{nji}^*}} \quad (3.3.21)$$

La distribución logística es muy similar a la distribución normal. Su varianza es el doble de la Gumbel, por lo que para la logística estándar la varianza¹² es $\pi^2/3$. De este modo, cuando se habla de errores Gumbel en los modelos de la familia Logit debe entenderse que esa expresión se refiere al error de cada alternativa, si bien lo observable es la diferencia con distribución logística. Puede verse la similitud entre estas distribuciones en la Figura 3.1, en la que se aprecia también el sesgo de la distribución Gumbel.

Para calcular la probabilidad de que el individuo n escoja la alternativa i de acuerdo con la ecuación

$$\begin{aligned} P_{ni} &= P(U_{ni} > U_{nj}, \forall j \neq i) \\ P_{ni} &= P(V_{ni} + \epsilon_{ni} \geq V_{nj} + \epsilon_{nj} \forall j \neq i) \\ P_{ni} &= P(\epsilon_{nj} - \epsilon_{ni} \leq V_{ni} + V_{nj} \forall j \neq i) \end{aligned}$$

¹¹Una deducción analítica de estos valores para $\eta = 0$ se puede encontrar en Louviere, Hensher y Swait (2000: 142-143)

¹²Para una deducción formal de los momentos de esta distribución puede consultarse Gumbel, 1958: 126-128

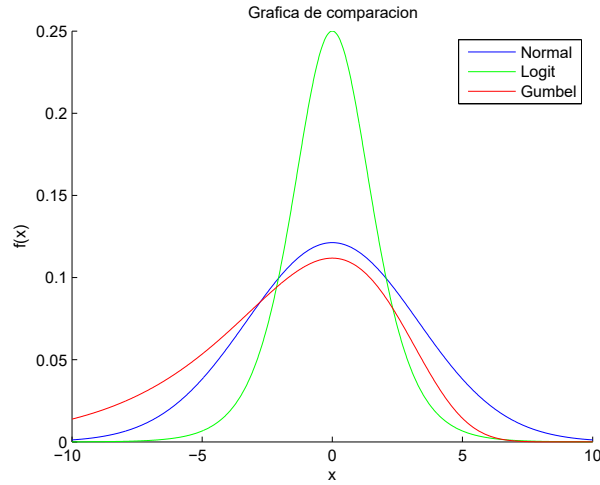


Figura 3.1: Comparación entre distribuciones normal, Gumbel y logística

se obtiene la conocida expresión del modelo Logit:

$$P_{ni} = \frac{e^{\mu V_{ni}}}{\sum e^{\mu V_{nj}}} \quad (3.3.22)$$

El parámetro de escala no es identificable, por lo que habitualmente se fija su valor en 1. De este modo, los coeficientes del modelo (en el caso de una formulación lineal en los parámetros), se están multiplicando por este parámetro, relacionado con la varianza de los errores. Esto tiene importancia al comparar resultados de modelos diferentes (la normalización habitual es diferente para el modelo probit) o en el caso de parámetros de escala diferentes: En lo sucesivo se supondrá que el parámetro de escala es 1 y se omitirá de la formulación, por lo que la probabilidad pasa a ser:

$$P_{ni} = \frac{e^{V_{ni}}}{\sum e^{V_{nj}}} \quad (3.3.23)$$

En el caso de que la parte observada de la utilidad sea lineal en los parámetros, que es la especificación más habitual, $V_{nj} = \mathbf{x}_{nj}\boldsymbol{\beta}$ donde $\boldsymbol{\beta}$ es el vector traspuesto de $\boldsymbol{\beta} = (\boldsymbol{\beta}_1, \dots, \boldsymbol{\beta}_k)$, entonces la probabilidad se puede expresar como:

$$P_{ni} = \frac{e^{\mathbf{x}_{ni}\boldsymbol{\beta}}}{\sum_j e^{\mathbf{x}_{nj}\boldsymbol{\beta}}} \quad (3.3.24)$$

o escrita de otra forma¹³,

¹³(McFadden, 1974)

$$P_{ni} = \frac{\exp(\mathbf{x}'_{nj}\beta)}{\sum_{k=0}^{m_n} \exp(\mathbf{x}'_{nk}\beta)}, \quad n = 1, 2, \dots, t \text{ y } j = 0, 1, \dots, m_n, \quad (3.3.25)$$

donde se puede asumir $\mathbf{x}_{n0} = 0$ sin pérdida de generalidad.

Si bien Train (2003: 41) afirma que bajo condiciones generales, cualquier función se puede aproximar de forma arbitrariamente cercana mediante una función lineal en los parámetros (teniendo en cuenta que las variables entren transformadas¹⁴), existen circunstancias que pueden hacer más adecuadas funciones no lineales¹⁵. Por otro lado, la función de log-verosimilitud de la muestra, en el modelo lineal, es globalmente cóncava en los parámetros β , lo que facilita su maximización numérica¹⁶.

Por otra parte, el **Modelo Logit Multinomial (MNL)** descansa sobre el supuesto de **Independencia de Alternativas Irrelevantes (IAI)**. Éste se expresa prácticamente en que el cociente entre las probabilidades de elección de dos alternativas i y j es constante y sólo depende de las utilidades de ambas alternativas; así, es independiente de las utilidades del resto de las alternativas. La expresión de esta propiedad es fácil de obtener a partir de la ecuación anterior, así:

$$\frac{P_{ni}}{P_{nj}} = e^{\beta(V_i - V_j)} \quad (3.3.26)$$

En un principio, esta propiedad tuvo su importancia ya que permitía analizar la demanda de nuevas alternativas sin necesidad de reestimar el modelo, bastaba simplemente con conocer sus atributos. El problema se presenta cuando existen alternativas correlacionadas (por ejemplo, automóvil como conductor y automóvil como acompañante, o distintos modos de transporte público como metro y bus), porque el modelo conduce a predicciones sesgadas.¹⁷

Como el MNL supone errores Gumbel independientes e idénticamente distribuidos, entonces no acepta correlación ni heterocedasticidad. Luego la matriz de covarianza asociada a este modelo es **diagonal**. Además como la varianza del término Gumbel se asocia con el factor de escala, se tiene

¹⁴véase McFadden, 1981: 220 y Ben-Akiva y Lerman, 1985: 62-64

¹⁵Train (2003: 56-58) presenta ejemplos de parámetros que no entran linealmente en la función de utilidad

¹⁶McFadden, 1973

¹⁷Una ilustración interesante de este problema puede verse en Mayberry (1973)

$$\Sigma = \sigma^2 \begin{pmatrix} 1 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & 1 \end{pmatrix} = \sigma^2 \mathbf{I}_{J_n \times J_n} = \frac{\pi^2}{6\lambda^2} \mathbf{I}_{J_n \times J_n} \quad (3.3.27)$$

Para que el modelo sea identificable, se debe fijar el valor del factor de escala. En la mayoría de los casos implícitamente se hace el supuesto que el factor de escala (λ) es igual a uno.

Otros modelos de elección discreta permiten considerar correlación entre las variables e incluso la existencia de heterocedasticidad. Conforme más general es la matriz de covarianza, más compleja es la implementación del modelo y su estimación.

La estimación del modelo Logit se realiza a partir de la maximización de la función de log-verosimilitud de la muestra, así:

$$\log L = \sum_{i=1}^n \sum_{j=0}^{m_i} y_{ij} \log P_{ij} \quad (3.3.28)$$

Siguiendo a McFadden(1974) se mostrará la concavidad global de 3.3.28.

A partir de derivar la ecuación 3.3.28 con respecto a β , se tiene que:

$$\frac{\partial \log L}{\partial \beta} = \sum_i \sum_j \frac{y_{ij}}{P_{ij}} \frac{\partial P_{ij}}{\partial \beta} \quad (3.3.29)$$

donde \sum_i y \sum_j denotan $\sum_{i=1}^n$ y $\sum_{j=0}^{m_i}$, respectivamente.

Ahora, derivando la ecuación 3.3.29 una vez más, se obtiene:

$$\frac{\partial^2 \log L}{\partial \beta \partial \beta'} = \sum_i \sum_j \frac{y_{ij}}{P_{ij}} \left[\frac{\partial^2 P_{ij}}{\partial \beta \partial \beta'} - \frac{1}{P_{ij}} \frac{\partial P_{ij}}{\partial \beta} \frac{\partial P_{ij}}{\partial \beta'} \right] \quad (3.3.30)$$

Por tanto, derivando la ecuación 3.3.25 y después de cierta manipulación se tiene que:

$$\frac{\partial P_{ij}}{\partial \beta} = P_{ij} (\mathbf{x}_{ij} - \bar{\mathbf{x}}_i) \quad (3.3.31)$$

donde,

$$\bar{\mathbf{x}}_i = \frac{\sum_k \exp(\mathbf{x}'_{ik}\beta)\mathbf{x}_{ik}}{\sum_k \exp(\mathbf{x}'_{ik}\beta)}$$

y

$$\begin{aligned} \frac{\partial^2 P_{ij}}{\partial\beta\partial\beta'} &= P_{ij} (\mathbf{x}_{ij} - \bar{\mathbf{x}}_i) (\mathbf{x}_{ij} - \bar{\mathbf{x}}_i)' \\ &- \frac{P_{ij} \sum_k \exp(\mathbf{x}'_{ik}\beta) \mathbf{x}_{ik}\mathbf{x}'_{ik}}{\sum_k \exp(\mathbf{x}'_{ik}\beta)} + P_{ij}\bar{\mathbf{x}}_i\bar{\mathbf{x}}_i' \end{aligned} \quad (3.3.32)$$

Por último, insertando las ecuaciones 3.3.31 y 3.3.32 en la ecuación 3.3.30 se obtiene:

$$\frac{\partial^2 \log L}{\partial\beta\partial\beta'} = \sum_i \sum_j P_{ij} (\mathbf{x}_{ij} - \bar{\mathbf{x}}_i) (\mathbf{x}_{ij} - \bar{\mathbf{x}}_i)' \quad (3.3.33)$$

el cual, curiosamente, no depende de y_{ij} . Como $P_{ij} > 0$ en este modelo, la matriz 3.3.33 es definida negativa a menos que $(\mathbf{x}_{ij} - \bar{\mathbf{x}}_i)' \alpha = 0$ para cada i y j para algún $\alpha \neq 0$. Puesto que tal evento es extremadamente improbable, se puede concluir para todos los propósitos prácticos que la Función Log Verosimilitud es Globalmente Cóncava en el Modelo Logit multinomial.

Las principales limitaciones del modelo Logit se centran en los siguientes aspectos:

- El modelo, al suponer los parámetros β constantes, no admite variaciones aleatorias en los gustos de los individuos. Se supone que todos ellos tienen la misma respuesta ante una variación en las variables explicativas. Las variaciones sistemáticas sí se pueden reflejar introduciendo, en la función de utilidad, interacciones de las variables de nivel de servicio con las características del individuo
- El modelo exhibe la citada propiedad de **independencia de alternativas irrelevantes** (IIA), dado que, por la formulación de la probabilidad, se obtiene que el cociente de las probabilidades de elección de dos alternativas no depende de las propiedades de las restantes alternativas:

$$\frac{P_{in}}{P_{nk}} = \frac{e^{V_{ni}}}{e^{V_{nk}}} = e^{\beta(V_{ni}-V_{nk})} \quad (3.3.34)$$

Como se ha dicho, otros modelos permiten otros patrones de sustitución admitiendo la existencia de correlación entre alternativas, el más sencillo es el Logit jerárquico.

- El modelo no admite correlación entre las observaciones, lo que puede hacer poco adecuada su utilización para datos de panel y encuestas en preferencias declaradas con varias respuestas por individuo, en los que existen factores no observados correlacionados a lo largo del tiempo, si bien si pueden acomodarse dinámicas relacionadas con factores observados (puede verse una discusión en Train, 2003: 54-56).
- El modelo no admite diferencias en la variabilidad de la evaluación de la utilidad para distintas alternativas (p.ej. el hecho de que el tiempo de viaje en vehículo privado sea mucho más variable que el tiempo de viaje en metro) o distintos individuos.

Para superar las limitaciones del MNL han ido surgiendo en la literatura un amplio abanico de modelos, de diferentes complejidades. La mayoría de ellos parten de la base del modelo Logit (residuos Gumbel), a excepción, principalmente, del modelo probit (residuos normales).

Una limitación adicional del modelo Logit es que las **colas** de la función de probabilidad son simétricas y delgadas, lo que puede no corresponderse con la realidad en algunas situaciones¹⁸.

3.3.6. Modelo Logit Jerárquico o Anidado (NL)

El Modelo Logit Jerárquico o Anidado (NL) permite agrupar las alternativas que el analista considera similares en conjuntos, conocidos como nidos. Dentro de cada nido se mantiene la **Independencia de Alternativas Irrelevantes (IAI)**, mientras que entre nidos diferentes la razón de probabilidades entre alternativas puede depender de las propiedades de otras alternativas de los nidos, pero no de las de otros nidos diferentes. En este sentido se podría hablar de una independencia de nidos irrelevantes¹⁹.

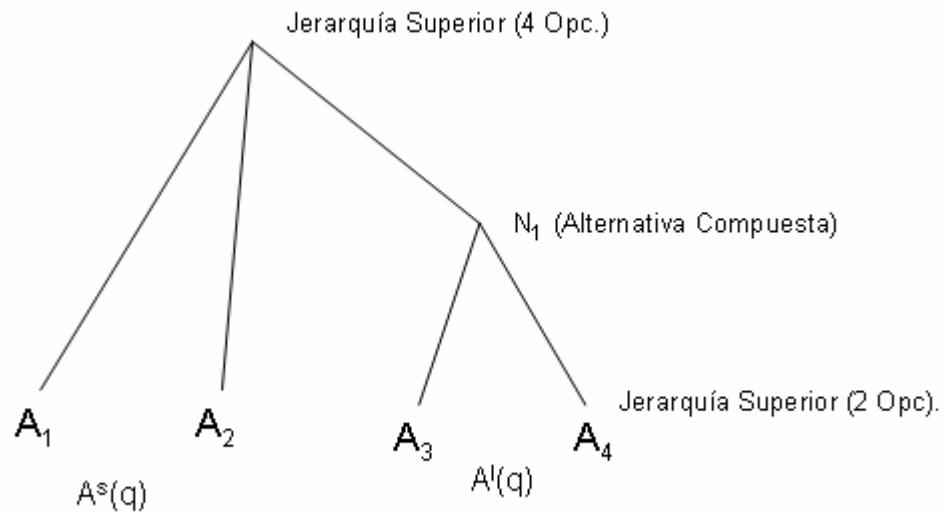
Pueden existir nidos tanto en serie como en paralelo, dando lugar a la estructura de árbol que caracteriza a estos modelos. Aparecen unos parámetros de escala adicionales en cada nido que, en cierto modo, reflejan la estructura de correlación. Uno de estos parámetros debe ser fijado para evitar problemas de identificación, existiendo dos estrategias diferentes para realizarlo, inferior o superior, en función de si la restricción se aplica en uno de los nidos inferiores o en el nido principal.

¹⁸Para abordar esta limitación se ha desarrollado la familia de modelos Dogit (Gaudry y Dagenais, 1978, 1979) e IPT-L (Inverse Power Transformation Logit, Gaudry, 1981)

¹⁹ (Train, 2003: 84)

La estructura de la matriz de covarianza obtenida es diagonal por bloques. Cuando dos o más alternativas están agrupadas en un nido, los elementos no diagonales correspondientes serán distintos de cero. Los elementos de la diagonal siguen siendo idénticos, manteniéndose la homocedasticidad característica del MNL²⁰.

El modelo Logit Jerárquico o Anidado resuelve en parte el problema de la **Independencia de Alternativas Irrelevantes (IAI)**, de manera que es posible considerar cierto patrón de correlación entre alternativas agrupándolas en jerarquías o nidos²¹. Cada nido o jerarquía es representado por una alternativa compuesta frente al resto de las alternativas disponibles. Un ejemplo de una estructura jerárquica o anidada puede ser la de la siguiente figura.



La obtención del modelo NL se basa en los siguientes supuestos:

- Se agrupa en jerarquías o nidos a todos los subconjuntos de opciones correlacionadas entre sí, es decir, aquellas que presentan ciertas semejanzas.
- Cada jerarquía o nido es representado por una alternativa compuesta frente a las demás que están disponibles a la población.

²⁰(véase Munizaga y Álvarez-Daziano, 2001; Espino, 2003: 36)

²¹Williams, 1977; Daly y Zachary, 1978

La estimación sigue un proceso secuencial donde se estima primero un MNL para aquellas opciones más similares que se han agrupado en un nido, por ejemplo para las alternativas A_3 y A_4 del nido inferior de la figura, omitiendo todas aquellas variables que tengan el mismo valor²² para este subconjunto de opciones, $A^I(q)$. La necesidad de omitir estas variables se debe al hecho de que el MNL trabaja en diferencias. Éstas son introducidas posteriormente en el nido superior, puesto que afectan a la elección entre la jerarquía inferior y el resto de opciones en $A^S(Q)$.

La forma en que se introduce el nido inferior en la jerarquía superior es a través de la alternativa compuesta N_I , a la cual se le asocia una utilidad representativa de todo el nido, que va a tener dos componentes²³:

- la primera considera como variable la Utilidad Máxima Esperada²⁴ de las alternativas del nido:

$$EMU = \ln \sum_{A_j \in A^I(q)} \exp(V_j) \quad (3.3.35)$$

siendo V_j la utilidad de la alternativa j del nido, en la cual no se tiene en cuenta el vector de variables \mathbf{W} , es decir, aquellas que toman el mismo valor para las alternativas del subconjunto de alternativas $A^I(q)$.

- Y la segunda que considera el vector de los atributos comunes a todos los componentes del nido, \mathbf{W} . De este modo, la utilidad compuesta del nido viene dada por:

$$\tilde{U}_j = \phi EMU + \alpha W \quad (3.3.36)$$

donde ϕ y α son parámetros a estimar.

Realizado todo el proceso de estimación para el nido inferior, se estima un MNL para la jerarquía superior donde se incluye la alternativa compuesta y el resto de alternativas contenidas en el nido superior.

De acuerdo con este modelo, la probabilidad de que un individuo n elija la alternativa del nido $A_j \in A^I(q)$ es igual a la probabilidad de que elija la alternativa compuesta (en el nido superior) por la probabilidad condicional de que opte por A_j , en la jerarquía inferior, dado que escogió la alternativa compuesta.

En el caso del ejemplo mostrado en la figura donde hay dos niveles jerárquicos, esta probabilidad sería:

²²Denominadas como vector W

²³(ver Sobel, 1980; Ortúzar, 1983)

²⁴EMU, Expected Maximum Utility

$$P_{nj} = \frac{\exp(\phi EMU + \alpha W + \theta^S Z^S)}{\sum_{A_j \in A^S(q)} \exp(\phi EMU + \alpha W + \theta^S Z^S)} * \frac{\exp(\theta^I Z_j^I)}{\sum_{A_j \in A^I(q)} \exp(\theta^I Z_i^I)} \quad (3.3.37)$$

denotando los superíndices **I** y **S** (como en la Figura 1) los niveles jerárquicos inferior y superior, respectivamente; Z^I y Z^S los vectores de los atributos de las alternativas disponibles en ambos niveles y θ , α y ϕ , los parámetros a estimar.

La consistencia interna requerida por el modelo establece, para el caso de dos niveles jerárquicos²⁵, que:

$$0 < \phi \leq 1 \quad (3.3.38)$$

- Si ϕ es igual a 1, el modelo NL es matemáticamente equivalente al MNL.
- Si ϕ es menor que 0 significa que el aumento de la utilidad de una alternativa del nido disminuirá la probabilidad de elegir dicha opción.
- Si ϕ es igual a 0 significa que el aumento de la utilidad de una alternativa del nido no alteraría la probabilidad de elección de la alternativa.
- Si ϕ es mayor que 1, el aumento de la utilidad de una alternativa del nido aumentaría tanto su probabilidad de elección cómo la del resto de las alternativas del nido.
- Si el número de jerarquías o nidos definidos es superior a 2, la consistencia interna del modelo vendría definida de la siguiente manera:

$$0 < \phi_1 \leq \phi_2 \leq \phi_3 \cdots \leq \phi_S \leq 1 \quad (3.3.39)$$

siendo ϕ_1 el parámetro correspondiente al nido del nivel inferior y ϕ_S al nido de nivel superior de cada rama del árbol²⁶.

La matriz de covarianza para este modelo para el caso que hemos supuesto en nuestro planteamiento será:

$$\Sigma = \begin{pmatrix} \sigma^2 & \cdots & 0 \\ \vdots & \ddots & \sigma_{34}^2 \\ 0 & \sigma_{43}^2 & \sigma^2 \end{pmatrix} \quad (3.3.40)$$

²⁵considerado en la figura y en la definición de la probabilidad de elección de la alternativa j para el individuo n

²⁶(Ortúzar y Willumsen, 2001)

Al igual que en el modelo MNL, se trata de un modelo homocedástico y presenta correlación entre las alternativas 3 y 4 que están agrupadas en un mismo nido (la covarianza entre la alternativa 3 y 4 es distinta de cero y $\sigma_{34} = \sigma_{43}$).²⁷

A pesar de que el modelo NL permite resolver, parcialmente, el problema de la correlación entre alternativas y se presenta como un modelo más completo que el MNL, éste posee aún algunas limitaciones²⁸:

- No permite considerar variaciones en los gustos, al igual que el MNL, ni funciones de utilidad heterocedásticas.
- No permite tratar los problemas de correlación entre alternativas pertenecientes a distintos nidos; sólo considera las interdependencias entre opciones de los nidos que se haya especificado en la estructura jerárquica a modelizar.
- Las posibles estructuras jerárquicas aumentan conforme aumenta el número de alternativas en proporción no lineal.²⁹ No obstante, el problema combinatorio se reduce al estudiar aquellas estructuras que teórica o intuitivamente son más razonables.
- Uno de los problemas potenciales de la estimación del modelo por el método secuencial es la transmisión de errores y distinto valor de los parámetros en los distintos nidos³⁰. Sin embargo, éstos se han podido solucionar gracias a la aparición de distintos programas que permiten la estimación simultánea del modelo.

En el modelo Logit multinomial, definido anteriormente, se recaló su debilidad cuando algunas de las alternativas son similares. El modelo Logit anidado (o no independiente) mitiga esa debilidad hasta cierto punto³¹.

Se analizará un modelo tricótonomo en detalle y luego se generalizarán los resultados obtenidos a un caso multinomial general.

Consideremos el modelo bus rojo-bus azul una vez más como ilustración. Sean

$$U_j = \mu_j + \epsilon_j, j = 0, 1y2$$

las utilidades asociadas con automóvil, bus rojo, y bus azul. (Para evitar complicaciones innecesarias en la notación, se ha suprimido el subscrito i). Se recalca

²⁷Para una explicación más clara de este modelo ver Carrasco y Ortúzar (2002)

²⁸(Ortúzar y Willumsem, 2001)

²⁹Sobel (1980) encuentra 26 estructuras posibles para un conjunto de cuatro alternativas

³⁰(ver Ortúzar et al., 1987)

³¹Este modelo se le atribuye a McFadden (1997) y se desarrolla con mayor detalle en un artículo posterior de McFadden (1981)

antes que no es razonable asumir la independencia entre ϵ_1 y ϵ_2 , aunque ϵ_0 se puede asumir independiente de las otras dos. McFadden sugirió la siguiente distribución bivariada como una forma conveniente de tener en cuenta una correlación entre ϵ_1 y ϵ_2 :

$$F(\epsilon_1, \epsilon_2) = \exp \left\{ - \left[\exp(-\rho^{-1}\epsilon_1) + \exp(-\rho^{-1}\epsilon_2) \right]^\rho \right\}, \quad 0 < \rho \leq 1 \quad (3.3.41)$$

Johnson y Kotz (1972, p.256) llamó a esta distribución *distribución de valor extremo bivariada tipo B de Gumbel*. Se puede mostrar que el coeficiente de correlación es $1 - \rho^2$. Si el $\rho = 1$ (el caso de independencia), $F(\epsilon_1, \epsilon_2)$ llega a ser el producto de dos distribuciones de valor extremo Tipo I - en otras palabras, el modelo Logit multinomial. Como para ϵ_0 , se asume $F(\epsilon_0) = \exp[-\exp(-\epsilon_0)]$ como en el modelo Logit multinomial.

Bajo estas presunciones se puede mostrar:

$$P(y = 0) = \frac{\exp(\mu_0)}{\exp(\mu_0) + [\exp(\rho^{-1}\mu_1) + \exp(\rho^{-1}\mu_2)]^\rho} \quad (3.3.42)$$

y

$$P(y = 1 | y \neq 0) = \frac{\exp(\rho^{-1}\mu_1)}{\exp(\rho^{-1}\mu_1) + \exp(\rho^{-1}\mu_2)} \quad (3.3.43)$$

Las otras probabilidades se pueden deducir de 3.3.42 y 3.3.43. Por lo tanto estas dos ecuaciones definen un modelo Logit anidado en el caso tricotómico. Dividiendo el numerador y el denominador de 3.3.42 por el $\exp(\mu_0)$ y los de 3.3.43 por

$$\exp(-\rho^{-1}\mu_1),$$

notamos que las probabilidades dependen de $\mu_2 - \mu_0, \mu_1 - \mu_0$ y ρ . Normalmente se especificaría

$$\mu_j = \mathbf{x}'_j \beta, \quad j = 0, 1, 2.$$

La forma de estas dos probabilidades es intuitivamente atractiva. La ecuación 3.3.43 muestra que la selección entre las dos alternativas similares se hace de acuerdo a un modelo binario Logit, mientras 3.3.42 sugiere que la escogencia entre carro y no carro es también igual a un modelo Logit excepto que se usa un cierto tipo de promedio ponderado del $\exp(\mu_1)$ y $\exp(\mu_2)$.

Para obtener 3.3.42, nótese que

$$\begin{aligned}
 P(y = 0) &= P(U_0 > U_1, U_0 > U_2) & (3.3.44) \\
 &= P(\mu_0 + \epsilon_0 > \mu_1 + \epsilon_1, \mu_0 + \epsilon_0 > \mu_2 + \epsilon_2) \\
 &= \int_{-\infty}^{\infty} \left\{ \int_{-\infty}^{\epsilon_0 + \mu_0 - \mu_1} \left[\int_{-\infty}^{\epsilon_0 + \mu_0 - \mu_2} \exp(-\epsilon_0) \right. \right. \\
 &\times \left. \exp[-\exp(-\epsilon_0)] f(\epsilon_1, \epsilon_2) d\epsilon_2 \right] d\epsilon_1 \Big\} d\epsilon_0 \\
 &= \int_{-\infty}^{\infty} \exp(-\epsilon_0) \exp[-\exp(-\epsilon_0)] \times \exp\left(-\left\{\exp[-\rho^{-1}(\epsilon_0 + \mu_0 - \mu_1)]\right. \right. \\
 &+ \left. \left. \exp[-\rho^{-1}(\epsilon_0 + \mu_0 - \mu_2)]\right\}^{\rho}\right) d\epsilon_0 \\
 &= \int_{-\infty}^{\infty} \exp(-\epsilon_0) \exp[-\alpha \exp(-\epsilon_0)] d\epsilon_0 \\
 &= \alpha^{-1}
 \end{aligned}$$

donde,

$$\alpha = 1 + \exp(-\mu_0) \left[\exp(\rho^{-1}\mu_1) + \exp(\rho^{-1}\mu_2) \right]^{\rho}$$

Para obtener 3.3.43, primero se puede observar que

$$\begin{aligned}
 P(y = 1 | y \neq 0) &= P(U_1 > U_2 | U_1 > U_0 \text{ o } U_2 > U_0) & (3.3.45) \\
 &= P(U_1 > U_2)
 \end{aligned}$$

donde la última igualdad se obtiene del particular supuesto distribucional. Por tanto, se tiene que

$$\begin{aligned}
 P(U_1 > U_2) &= P(\mu_1 + \epsilon_1 > \mu_2 + \epsilon_2) & (3.3.46) \\
 &= \int_{-\infty}^{\infty} \left[\int_{-\infty}^{\epsilon_1 + \mu_1 - \mu_2} f(\epsilon_1, \epsilon_2) d\epsilon_2 \right] d\epsilon_1 \\
 &= \int_{-\infty}^{\infty} \exp(-\epsilon_1) \left\{ 1 + \exp[-\rho^{-1}(\mu_1 - \mu_2)] \right\}^{\rho-1} \\
 &\times \exp\left(-\exp(-\epsilon_1) \times \left\{ 1 + \exp[-\rho^{-1}(\mu_1 - \mu_2)] \right\}^{\rho}\right) d\epsilon_1 \\
 &= \left\{ 1 + \exp[-\rho^{-1}(\mu_1 - \mu_2)] \right\}^{-1}
 \end{aligned}$$

donde en la tercera igualdad se usa

$$\begin{aligned}
 \frac{\partial F(\epsilon_1, \epsilon_2)}{\partial \epsilon_1} &= \left[\exp(-\rho^{-1}\epsilon_1) + \exp(-\rho^{-1}\epsilon_2) \right]^{\rho-1} & (3.3.47) \\
 &\times \exp(-\rho^{-1}\epsilon_1) F(\epsilon_1, \epsilon_2)
 \end{aligned}$$

El modelo definido por 3.3.42 y 3.3.43 se reduce al modelo Logit multinomial si $\rho = 1$. Por consiguiente la hipótesis IAI, que es equivalente a la hipótesis $\rho = 1$, puede probarse contra la hipótesis alternativa del modelo Logit Anidado por cualquiera de las tres pruebas asintóticas que son:

- La prueba de proporción de probabilidad (The Likelihood ratio test (LRT))
- La prueba de Wald's (Wald, 1943)
- La prueba Rao's score (Rao, 1947)

Para generalizar el Modelo Logit Anidado tricotómico definido por 3.3.42 y 3.3.43 al caso general de $m + 1$ respuestas. Supone que los $m + 1$ enteros $0, 1, \dots, m$ pueden particionarse naturalmente en S grupos para que cada grupo consista en alternativas similares. Escribiendo la partición como

$$(0, 1, 2, \dots, m) = B_1 \cup B_2 \cup \dots \cup B_s \quad (3.3.48)$$

Donde \cup denotan la unión. Luego McFadden sugiere la distribución conjunta

$$F(\epsilon_0, \epsilon_1, \dots, \epsilon_m) = \exp\left\{-\sum_{s=1}^S a_s \left[\sum_{j \in B_s} \exp(-\rho_s^{-1} \epsilon_j)\right]^{\rho_s}\right\} \quad (3.3.49)$$

Entonces puede mostrarse que

$$\sum_{j \in B_s} P_j = \frac{a_s [\sum_{j \in B_s} \exp(\rho_s^{-1} \mu_j)]^{\rho_s}}{\sum_{s=1}^S a_s [\sum_{j \in B_s} \exp(-\rho_s^{-1} \epsilon_j)]^{\rho_s}}, s = 1, 2, \dots \quad (3.3.50)$$

y

$$P(y = j | j \in B_s) = \frac{\exp(\rho_s^{-1} \mu_j)}{\sum_{k \in B_s} \exp(\rho_s^{-1} \mu_k)} s = 1, 2, \dots \quad (3.3.51)$$

Note que 3.3.50 y 3.3.51 son generalizaciones de 3.3.42 y 3.3.43 respectivamente. Por tanto, estas probabilidades definen completamente el modelo. Podemos interpretar

$$\left[\sum_{j \in B_s} \exp(\rho_s^{-1} \mu_j)\right]^{\rho_s}$$

como una clase de media ponderada de $\exp(\mu)$ para $j \in B$.

El modelo Logit anidado definido por 3.3.50 y 3.3.51 se puede estimar por MLE, pero también puede estimarse consistentemente mediante un método natural de dos pasos, el cual es computacionalmente más simple. Suponga que $\mu_j = \mathbf{x}'_j \boldsymbol{\beta}$.

- Primero, la parte de la función de verosimilitud que es el producto de las probabilidades condicionales de la forma 3.3.51 se maximiza para producir un estimador de $\rho_s^{-1}\beta$
- Segundo, este estimador se inserta en el lado derecho 3.3.50, y el producto de 3.3.50 sobre s e i (el cual se suprime) se maximiza para obtener los estimadores de $\rho's$ y $a's$ (uno de los $a's$ se puede establecer arbitrariamente)³².

En un modelo Logit anidado a gran escala, el método de los dos pasos es especialmente útil debido a la infactibilidad próxima de MLE. Otro mérito del método de dos pasos es que produce estimadores consistentes que se pueden usar para iniciar una iteración Newton- Raphson para calcular MLE, como lo hacen Hausman y McFadden.

3.3.7. Modelo Probit Multinomial (MNP)

El Modelo Probit Multinomial (MNP) se obtiene de suponer una distribución de probabilidad conjunta para el término de perturbación aleatoria Normal Multivariada con media cero y matriz de covarianza arbitraria³³.

En este sentido, el Modelo Probit Multinomial MNP permite considerar heterocedasticidad y correlación entre los distintos términos del error. La posibilidad de poder definir cualquier matriz de covarianzas otorga gran flexibilidad y le confiere al Modelo Probit Multinomial MNP la capacidad para estudiar una gama más amplia de comportamientos.

Una cuestión importante de este modelo es la dificultad a la hora de establecer la función de verosimilitud debido a que no existe una expresión cerrada para la probabilidad de elección de las alternativas. Para más de tres opciones, no es posible, analíticamente, estimar el modelo y hay que solventar este problema empleando procedimientos de simulación.

Los problemas observados en la estimación de modelos MNP derivaron en su no utilización por muchos años, y también en la búsqueda de modelos flexibles que fueran más fácilmente tratables desde el punto de vista computacional. Así surgió el **Modelo Logit Mixto (ML)**.

³² La matriz de covarianza asintótica de estos elementos es dada por McFadden (1981)

³³(Daganzo, 1979)

El modelo probit supone que los errores siguen conjuntamente una distribución normal. Esto es, (de acuerdo con el Teorema Central del Límite), la distribución resultante de la suma de una gran cantidad de componentes independientes no observadas, por lo que parece un supuesto razonable para el término de error³⁴.

El modelo Probit permite un rango de comportamientos humanos mucho más amplio del admitido por el MNL³⁵, al ser capaz de incluir **correlaciones**, **heterocedasticidades** y **variaciones en los gustos**.

Su única limitación reside en la utilización de la distribución normal que, si bien en la mayoría de las situaciones proporciona una representación adecuada de las componentes aleatorias³⁶, puede ser inadecuada en ciertas situaciones, y conducir a predicciones erróneas. Es el caso de coeficientes que únicamente tengan sentido que posean un determinado signo.

Si se introduce la función de densidad de la distribución normal en la ecuación

$$\begin{aligned} P_{ni} &= P(\epsilon_{nj} - \epsilon_{ni} \leq V_{ni} - V_{nj} \forall j \neq i) \\ &= \int_i I(\epsilon_{nj} - \epsilon_{ni} \leq V_{ni} - V_{nj} \forall j \neq i) f(\epsilon_n) d\epsilon_n \end{aligned}$$

la integral resultante carece de expresión analítica, y debe ser resuelta por simulación. Existen diversos procedimientos para resolver esta integral (integración numérica, aproximación de Clark, etc.), el más empleado es el simulador GWK (Geweke-Hajivassiliou-Keane), desarrollado a partir de los trabajos de estos autores: Geweke (1989 y 1991), Hajivassiliou³⁷ y Keane (1990, 1994). Este estimador permite reducir la dimensión de la integral.

Varias experiencias han verificado su utilidad y relativa exactitud³⁸.

La dificultad relativa de esta integración ha provocado que el modelo probit haya sido poco empleado en la práctica para la elección modal, a pesar de su potencia³⁹. En los últimos años, los reseñados avances en la simulación, junto con el

³⁴(Ben-Akiva y Lerman, 1985: 129, Munizaga, 1997: 7)

³⁵Hausman y Wise, 1978

³⁶Train, 2003: 101

³⁷referenciado en Hajivassiliou y McFadden, 1998

³⁸Hajivassiliou, McFadden y Ruud, 1996; Börsh-Supan y Hajivassiliou, 1993. Puede encontrarse una descripción de estos simuladores en Ortúzar (2000b: 150-157) o en Train (2003: 119-137)

³⁹Según Sørensen (2003: 30) se ha empleado en modelos de asignación cuando existen abundantes superposiciones entre rutas

desarrollo de la informática, han soslayado parcialmente estos inconvenientes. Sin embargo, la aparición del modelo Logit mixto ha relegado de nuevo al modelo probit.

La identificación del modelo probit es mucho más compleja que la del modelo Logit multinomial, en el que las limitaciones impuestas a la matriz de covarianza por la estructura del modelo eliminan ya numerosos parámetros.

En el modelo probit la matriz de covarianza es completa, y no todos sus elementos van a poder ser estimados de forma independiente. Es necesario fijar elementos en la matriz de covarianza de las diferencias entre utilidades⁴⁰.

Sea $U_j, j = 0, 1, 2, \dots, m$ la utilidad estocástica asociada con la **j-ésima** alternativa para un individuo particular. El **Modelo Probit Multinomial**, es el modelo en el cual U_j se distribuyen conjunta y normalmente⁴¹.

Este modelo raramente se ha usado en la práctica hasta hace poco, debido a su dificultad computacional (excepto, por supuesto, cuando $m = 1$ en cuyo caso el modelo se reduce a un modelo del probit binario que se discutió antes). Para ilustrar la complejidad del problema, considere el caso de $m = 2$. Por tanto, para evaluar $P(y = 2)$, por ejemplo, se debe calcular la integral múltiple

$$\begin{aligned} P(y = 2) &= P(U_2 > U_1, U_2 > U_0) && (3.3.52) \\ &= \int_{-\infty}^{\infty} \int_{-\infty}^{U_2} \int_{-\infty}^{U_2} f(U_0, U_1, U_2) dU_0 dU_1 dU_2 \end{aligned}$$

donde f es una densidad normal trivariada. El cálculo directo de tal integral es complicado incluso para el caso de $m = 3$. Además, 3.3.52 debe evaluarse en cada paso de un método iterativo para maximizar la función de verosimilitud.

Hausman y Wise (1978) estimaron un modelo probit tricotonomo para explicar la elección modal entre manejar el automóvil propio, tomar colectivo, y tomar un autobús, para 557 obreros en Washington D.C.

Ellos evaluaron integrales de la forma (3.3.52) usando la expansión de series y notaron que el método es factible para un modelo con más de cinco alternativas.

⁴⁰Walker, 2001: 34-35

⁴¹Dicho modelo se propuso primero por Aitchison y Bennett (1970)

Capítulo 4

Aplicación

4.1. Introducción

La simulación es el método numérico más frecuentemente usado para la estimación de Modelos de Elección Discreta desde [20], así como en recientes estudios y textos de [28]. El objetivo ahora es presentar el desarrollo de la aplicación pero usando otro método de estimación, basado en el enfoque de Cuadratura Gaussiana (Ver [7], [12]).

La fórmula de Cuadratura Gaussiana ha sido popularizada por ([7]) para la estimación del componente de error del modelo panel probit. En ([6]) se resaltan algunas dificultades numéricas y estadísticas al aplicarlos a modelos con efectos de grupo. Con un número mediano o grande de individuos en un grupo, la función de verosimilitud del modelo calculado por la fórmula de Cuadratura Gaussiana puede ser numéricamente inestable, y peor aún, imposible de calcular. La inferencia estadística puede ser también insegura. Pero estas dificultades se pueden resolver con un algoritmo cuidadosamente diseñado y la selección adecuada del número de puntos de cuadratura. Sin embargo, con un número muy grande de individuos en un grupo, la fórmula Cuadratura Gaussiana de la integral puede presentar grandes errores de aproximación numérica.

4.2. Fuente de datos

4.2.1. Caso Estados Unidos

Estos primeros datos se toman del Panel Study Income Dynamics (PSID) para el mercado laboral de los Estados Unidos. Consta de 19 variables sobre mujeres blancas casadas en edades entre 30 y 60 años.

Como se expresa en la descripción detallada que se encuentra en su sitio Web, el Panel Study Income Dynamics (PSID) [29]

“es un examen longitudinal de una muestra representativa de los individuos estadounidenses (hombres, mujeres, y niños) y sus familias. Se desarrolla desde 1968. Los datos se recogen anualmente, y los archivos de datos contienen la total extensión de la información recogida en el curso del estudio. Los datos del PSID se utilizan para análisis de sección-cruzada, longitudinales e intergeneracionales, y para estudiar tanto a individuos como a familias. El estudio acentúa los aspectos dinámicos del comportamiento económico y demográfico, pero contiene una amplia gama de medidas, incluyendo las sociológicas y psicológicas. Entre 1968 y 1988, el PSID recogió la información referente a casi 37.500 individuos y cubren tanto como 21 años de sus vidas”.

El PSID proporciona una amplia variedad de información de la familia y a nivel individual, así como una cierta información sobre las localizaciones en las cuales residen los hogares. El foco central de los datos es económico y demográfico, con detalles sustanciales en orígenes y cantidades de renta, empleo, cambios de la composición de la familia, y localización residencial. El contenido de naturaleza más sociológica o psicológica también se incluye en algunas partes del estudio.

A continuación se procede a presentar un caso de aplicación utilizando información tomada de la Base de Datos PSID. El archivo contiene datos sobre la participación laboral de mujeres casadas. Se trata de información referente a 753 mujeres, de las cuales 428 reportan estar trabajando durante un determinado año. Los datos se agrupan por rango de edades, resultando en 11 grupos diferentes.

Aquí se plantea un modelo discreto de elección binaria con el cual se busca explicar la probabilidad de que una mujer casada decida participar en el mercado laboral. En este caso la variable dependiente **inlf** tomará el valor de 1 si la mujer encuestada está trabajando y 0 en cualquier otro caso. Las variables explicativas de la participación de la mujer en la fuerza laboral son:

educ Número de años de estudios.

exper Número de años de experiencia laboral.

exper2 La variable **exper** elevada al cuadrado.

ingresop Ingreso propio (en miles).

hijos6 Número de hijos menores de 6 años.

4.2.2. Caso Colombia

Los datos usados en esta parte del trabajo se toman de la Encuesta Continua de Hogares realizada por el DANE pertenecientes al segundo trimestre del año 2005 referente al mercado laboral en las principales ciudades colombianas. Consta de 2929 registros de mujeres en edades entre 15 y 90 años.

La encuesta de Hogares es el sistema de recolección de datos que en nuestro país ha venido utilizando el DANE (Departamento Administrativo Nacional de Estadística) desde 1970 para obtener información de familias y hogares colombianos. Desde 1976 se conoció como Encuesta Nacional de Hogares cubriendo las principales ciudades (Bogotá, Medellín, Cali, Barranquilla, Bucaramanga, Pasto y Manizales) y sus correspondientes áreas metropolitanas. La Encuesta Nacional de Hogares presentaba limitaciones principalmente en la frecuencia trimestral y la rotación de personal operativo que realizaba la captura de la información dificultando el seguimiento de los hogares no informantes.

En el segundo trimestre de 2001 se dió paso a un nuevo sistema conocido como Encuesta Continua de Hogares que se caracteriza por una mayor cobertura, frecuencia en la recolección y presentación de datos (mensual en lugar de trimestral) así como cambios importantes en los conceptos que permiten la estimación de la fuerza de trabajo. Se presenta así una nueva caracterización en la definición de ciertas variables que modifican los análisis macro y microeconómicos de los indicadores del mercado laboral.

La Encuesta Continua de Hogares tiene como objetivo principal proporcionar información básica acerca del tamaño y estructura de la fuerza de trabajo de la población colombiana (empleo, desempleo e inactividad). Además, permite obtener datos de variables adicionales como: sexo, edad, estado civil, educación, entre otras.

De manera similar al caso anterior, se plantea un modelo discreto de elección binaria con el cual se busca explicar la probabilidad de que una mujer casada decida participar en el mercado laboral nacional. En este caso la variable **partic** tomará el valor de 1 si la mujer encuestada está trabajando y 0 en cualquier otro caso. Las variables usadas para explicar la participación de la mujer en la fuerza laboral son:

educat Número de años de estudios.

exper Número de años de experiencia laboral.

exper2 La variable **exper** elevada al cuadrado.

ingresop Ingreso propio (en miles).

hijos6 Número de hijos menores de 6 años.

Para este caso, los datos se agruparon por ciudad, correspondientes con las capitales de departamento objeto de estudio de la Encuesta Continua de Hogares resultando 13 grupos diferentes.

4.3. Metodología

4.3.1. Cuadratura Gaussiana

Considerando el siguiente modelo de elección binaria de componente aleatorio:

$$y_{ij}^* = x_{ij}\beta + \varepsilon_{ij} \quad (4.3.1)$$

para grupos $j = 1, \dots, J$ e individuos $i = 1, \dots, N_j$ en el j -ésimo grupo, donde x es un vector de variables exógenas. La perturbación en 4.3.1 se genera por un componente de error

$$\varepsilon_{ij} = u_j + v_{ij} \quad (4.3.2)$$

donde u_j y v_{ij} son mutuamente independientes, u_j son i.i.d. para todo j y v_{ij} son i.i.d. para todo i y j . El signo de la variable latente y^* determina la variable dicotómica observada I como $I = 1$ si $y^* > 0$ y 0 de otro modo. Sea F la función de distribución condicional de ε condicional en u , y sea f la función de densidad de u . La función de Log-Verosimilitud para el modelo es:

$$L = \sum_{j=1}^J \ln [Prob(I_{1j}, \dots, I_{N_j,j})] \quad (4.3.3)$$

donde,

$$Prob(I_{1j}, \dots, I_{N_j,j}) = \int_{-\infty}^{\infty} f(u_{ij}) \prod_{i=1}^{N_j} [F(b_{ij}|u_j) - F(a_{ij}|u_j)] du_u$$

$a_{ij} = -x_{ij}\beta$ y $b_{ij} = \infty$ si $I_j = 1$, y $a_{ij} = -\infty$ y $b_{ij} = -x_{ij}\beta$ si $I_j = 0$. Para un Modelo Probit, tanto u como v se asumen normalmente distribuidas. Una normalización típica para el modelo Probit especifica una media cero y una varianza uno para la disturbancia total ε . Sea ρ la correlación de individuos dentro de un grupo. Para este modelo normalizado de componente aleatorio ρ es también la varianza de u . Por tanto, u_j es $N(0, \rho)$ y v $N(0, 1 - \rho)$. La función Log-Verosimilitud para este modelo de componente de error Probit, puede ser escrito de manera más exacta, usando la simetría de una densidad normal como:

$$L = \sum_{j=1}^J \ln \left\{ \int_{-\infty}^{\infty} \frac{1}{(2\pi)^{\frac{1}{2}}} e^{-\frac{1}{2}u^2} \prod_{i=1}^{N_j} \Phi \left[D_{ij} \left(\frac{x_{ij}\beta + \rho^{\frac{1}{2}}u}{(1 - \rho)^{\frac{1}{2}}} \right) \right] du \right\} \quad (4.3.4)$$

donde Φ es la función de distribución normal estándar, y D es un indicador de signo tal que $D_{ij} = 1$ si $I_{ij} = 1$ y $D_{ij} = -1$ en cualquier otro caso.

La probabilidad conjunta de respuesta individuales dentro de un grupo en (4.3.4) involucra una integral simple, cuyo integrando es un producto de funciones de probabilidad univariadas. En ([7]) demostraron que estas probabilidades conjuntas podrían ser correctamente calculadas por la Cuadratura Gauss-Hermite. La Cuadratura Gaussiana calcula la integral

$$\int_{-\infty}^{\infty} e^{-z^2} g(z) dz$$

usando la fórmula de integración:

$$\int_{-\infty}^{\infty} e^{-z^2} g(z) dz = \sum_{m=1}^M w_m g(z_m) \quad (4.3.5)$$

donde M es el número definido de puntos; las z s y las w s son las abscisas y los pesos de los M puntos Gauss-Hermite. Si g en (4.3.5) es un polinomio de grado menor que $2M - 1$, el cálculo de la integración de Gauss-Hermite es exacta.

La teoría se basa en Polinomios Ortogonales ¹. Las abscisas y los pesos están disponibles de Stroud y Secrets (1966) y Abramowitz y Stegun (1964). Prest et al. (1992) proporcionan programas de cómputo para generar las abscisas y los pesos con un número específico de puntos. Dados los puntos z y los pesos w Gauss-Hermite, la verosimilitud en 4.3.4 puede ser evaluada por la fórmula Gauss-Hermite como:

$$L = \sum_{j=1}^J \ln \left\{ \sum_{m=1}^M w_m \frac{1}{(\pi)^{\frac{1}{2}}} \prod_{i=1}^{N_j} \Phi \left[D_{ij} \left(\frac{x_{ij} \beta + (2\rho)^{\frac{1}{2}} z_m}{(1-\rho)^{\frac{1}{2}}} \right) \right] \right\} \quad (4.3.6)$$

En ([7]) muestran la utilidad de la Cuadratura Gaussiana con un ejemplo de Datos Panel consistente de 1550 unidades muestrales con un máximo de 11 periodos para cada uno. Basados en la estabilidad de los estimados sugirieron que la cuadratura de 2 o 4 puntos serían suficientes. El enfoque de Cuadratura Gaussiana es computacionalmente eficiente comparado con otras técnicas de Cuadratura, tales como la de integración trapezoidal ². Por tanto, el enfoque de Cuadratura Gaussiana se usa muy frecuentemente en la literatura de la econometría empírica.

¹ver por ejemplo Press et al., 1992, capítulo 4

²Heckman y Willis, 1975

4.3.2. Criterio

Como elemento central de la aplicación se calcula la función de verosimilitud del modelo mediante el enfoque de Cuadratura Gaussiana y se investiga la validez de la aplicabilidad de la técnica numérica. El principio es recalculando la estimación usando dos números diferentes de puntos de cuadratura y comparar la log-verosimilitud y los coeficientes entre el modelo original y los otros dos modelos estimados. Si el cambio es menor que 0.01 % entonces la elección del punto de cuadratura no afecta significativamente los resultados. Para un cambio mayor a 0.1 % se debe dudar del enfoque de cuadratura para el modelo.

4.4. Resultados

4.4.1. Caso Estados Unidos

Los resultados de la ejecución de la estimación se presentan al final del trabajo, más exactamente en el Apéndice B (Aplicación). En la Tabla 1 del mismo se muestran los resultados producto de utilizar un procedimiento de cuadratura Gauss-Hermite adaptativo, con 8 puntos, utilizando como grupos las edades. Observando el test de Wald, las variables elegidas resultan ser conjuntamente significativas en explicar la probabilidad de que una mujer casada participe en el mercado laboral. Si observamos el estadístico de prueba z , todas resultan ser individualmente significativas. Además de tener los signos adecuados.

Por ejemplo una mayor cantidad de niños menores de 6 años afecta negativamente la probabilidad de que una mujer casada decida participar en el mercado laboral. Es de resaltar el signo negativo del coeficiente asociado a **exper2**. Esto significa que a medida que aumente la experiencia, la probabilidad de participar en el mercado laboral aumenta (signo positivo de *exper* pero a una tasa decreciente). Igualmente mayor número de años de estudios afecta positivamente la probabilidad de participar en el mercado laboral, y una conclusión que parece también lógica y consistente con la realidad, se observa que un factor que hace que las mujeres casadas tengan que buscar trabajo es que el ingreso familiar no sea suficiente. Por tanto a mayor ingreso familiar, menor probabilidad de que las mujeres casadas tengan que buscar trabajo. Además:

lnsig2u Representa el logaritmo del efecto individual.

Sigma Muestra la desviación estándar del efecto individual.

rho Es el porcentaje de la varianza total explicado por el componente de varianza al nivel del panel.

Cuando rho es cero, el componente de varianza no es importante, y el estimador panel no es diferente del estimador agrupado (estimador pooled). Para contrastar esto se utiliza un test de razón de verosimilitudes, como se observa se puede rechazar la hipótesis nula, y por consiguiente nuestra estimación es adecuada.

En la Tabla 2 se presentan los efectos marginales y la probabilidad que una mujer con unas características promedio participe en el mercado laboral de Estados Unidos.

Uno de los elementos claves que valida la utilización del procedimiento de cuadratura es la estabilidad de los parámetros estimados mediante la utilización de diferentes puntos de cuadratura. En esta parte del trabajo se estiman dos modelos alternativos uno con 12 puntos y otro con 16 puntos, basados en la siguiente concepto: si los coeficientes no cambian en mas que una diferencia relativa del 0,01 %, la elección de los puntos de cuadratura no afecta significativamente el resultado, y los resultados pueden ser interpretados con confianza; sin embargo si los resultados cambian de manera apreciable con una diferencia relativa mayor o igual al 1 %, el método de cuadratura será inestable y no conveniente. Los resultados de este análisis de sensibilidad se presentan en la tabla 3 y 4 donde se puede observar que las diferencias relativas son inferiores al 0,01 % por lo consiguiente el procedimiento de cuadratura usado es estable y confiable.

4.4.2. Caso Colombia

En la Tabla 5 se presentan los resultados después de utilizar un procedimiento de cuadratura Gauss-Hermite adaptativo, con 8 puntos, utilizando como grupos las ciudades. Observando el test de Wald, las variables elegidas resultan ser conjuntamente significativas en explicar la probabilidad de que una mujer casada participe en el mercado laboral. Si observamos el estadístico de prueba z , todas resultan ser individualmente significativas. Además de tener los signos adecuados.

Por ejemplo, la variable **hijos6**, a mayor cantidad de niños menores de 6 años, la probabilidad de que una mujer casada decida participar en el mercado laboral se afecta negativamente. Es de resaltar el signo negativo del coeficiente asociado a **exper2**. Esto quiere decir que en la medida que aumente la experiencia, la probabilidad de participar en el mercado laboral aumenta (signo positivo de exper pero a una tasa decreciente). Igualmente, la variable **educ** indica que a mayor número de años de estudios la probabilidad de participar en el mercado laboral se ve afectada positivamente y una conclusión que parece también lógica y consistente con la realidad, se observa que una variable que hace que las mujeres casadas tengan que buscar trabajo es que el ingreso familiar no sea suficiente. Por tanto a mayor ingreso familiar, menor probabilidad de que las mujeres casadas tengan que buscar

trabajo. También:

lnsig2u Representa el logaritmo del efecto individual.

Sigma Muestra la desviación estándar del efecto individual.

rho Es el porcentaje de la varianza total explicado por el componente de varianza al nivel del panel.

Cuando $\rho = 0$, el componente de varianza no es importante, y el estimador panel no es diferente del estimador agrupado (estimador pooled). Para contrastar esto se utiliza un test de razón de verosimilitudes, como se puede observar se puede rechazar la hipótesis nula, y por consiguiente nuestra estimación es adecuada.

En la Tabla 6 del Apéndice B se presentan los efectos marginales y la probabilidad de que una mujer con unas características promedio participe en el mercado laboral.

Uno de los elementos claves que valida la utilización del procedimiento de cuadratura es la estabilidad de los parámetros estimados mediante la utilización de diferentes puntos de cuadratura. En esta parte del presente trabajo se estiman dos modelos alternativos uno con 8 puntos y otro con 12 puntos, la idea anterior aún se mantiene. Los resultados de este análisis de sensibilidad se presentan en la tabla 7 y 8, obsérvese que las diferencias relativas son inferiores al 0,01 % por lo consiguiente el procedimiento de cuadratura es también estable y confiable.

A manera de conclusión, se puede enunciar que los resultados de la aplicación son consistentes con estudios teóricos similares, (Ver [14]) además se presenta en la siguiente tabla la comparación de algunos parámetros estimados para la regresión probit tanto para el caso de Estados Unidos como para Colombia:

Comparativo	EE.UU.	Colombia
ingresop	-.0140223	-5.71e-08
exper	.1216683	.0368012
hijos6	-.7917483	-.0262662
educat	.056473	.0136696
Log verosimilit.	-415.6071	-1807.4794
P(partic) (predict)	0.5778022	0.5445877

Como se puede apreciar, los signos de los coeficientes se mantienen en los dos casos y la probabilidad que una mujer casada y con unas características promedio, participe en el mercado laboral son aproximadamente iguales para los dos países: 0.5778 y 0.5446 para Estados Unidos y Colombia respectivamente.

Apéndice A

Función de Verosimilitud

A.1. Condiciones de regularidad

1. $x_i, i = 1, \dots, n, \quad i.i.d. \sim f(x_i; \theta_0)$.
2. Si $\theta \neq \theta_0 \Rightarrow f(x_i; \theta) \neq f(x_i; \theta_0)$.
3. $\theta \in \Theta, \Theta$ es compacto.
4. $\ln f(x_i; \theta)$ continua para cada $\theta \in \Theta$ con probabilidad uno.
5. $E[\sup_{\theta} |\ln f(x; \theta)|] < \infty$.
6. θ_0 es un punto interior de Θ .
7. $f(x; \theta)$ es dos veces diferenciable y estrictamente positiva en un entorno de θ_0 .
8. $\int \sup_{\theta} \nabla_{\theta} f(x; \theta) dx < \infty$ y $\int \sup_{\theta} \nabla_{\theta\theta} f(x; \theta) dx < \infty$
9. $I(\theta_0, x) \equiv E(s(\theta_0, x)s(\theta_0, x)')$ existe y es no-singular.
10. $E[\sup_{\theta} \|H(x; \theta)\|] < \infty$.

Las condiciones 1. a 5. son necesarias para consistencia, el resto también hace falta para normalidad asintótica. Ver [23] para mas detalles.

Apéndice B

Aplicación

B.1. Caso EE.UU.

TABLA 1

```
Regresion probit efectos aleat.          Numero de obs      =      753
Variable grupo (i): grupo                Numero de grupos   =      11

Efectos aleat u_i ~ Gaussiano           Obs por grupo: min =      6
                                           prom =      68.5
                                           max =      105
                                           Chi2 Wald(5)      =      158.66
                                           Prob > chi2       =      0.0000

Log verosimilit. = -415.6071
```

	inlf	Coef.	Err. Std.	z	P> z	[95% Interval Conf.]	
ingresop		-.0140223	.0048943	-2.87	0.004	-.0236149	-.0044298
exper		.1216683	.0188973	6.44	0.000	.0846302	.1587063
exper2		-.0020395	.0006166	-3.31	0.001	-.003248	-.000831
hijos6		-.7917483	.1206739	-6.56	0.000	-1.028265	-.5552319
educ		.1326852	.0252501	5.25	0.000	.083196	.1821744
_cons		-1.99356	.3317149	-6.01	0.000	-2.643709	-1.34341
/lnsig2u		-1.726882	.606595			-2.915786	-.5379772
Sigma		.4217086	.1279032			.2327261	.764152
rho		.1509869	.0777593			.0513787	.3686583

```
Test razon-verosiml. de rho=0: chibar2(01) = 23.30 Prob >= chibar2 = 0.000
```


TABLA 2

Probabilidad que una mujer casada con unas características promedio, participe en el mercado laboral es: 0.57780221

variable	dy/dx	Err. Std.	z	P> z	[95% I.C.]	X
ingresop	-.0067914	.00186	-3.66	0.000	-.010428 -.003155	20.129
exper	.0466689	.00712	6.55	0.000	.032708 .060629	10.6308
exper2	-.0009576	.00023	-4.14	0.000	-.001411 -.000504	178.039
hijos6	-.2094332	.04033	-5.19	0.000	-.288478 -.130388	.237716
educ	.056473	.00958	5.90	0.000	.037702 .075244	12.2869

TABLA 3

Reajuste del modelo intpoints() = 12

Iteracion 0: log likelihood = -415.60604

Iteracion 1: log likelihood = -415.60602

Regresion probit efectos aleat.

Numero de obs = 753

Variable Grupo(i): grupo

Numero de grupos = 11

Efectos Aleat. u_i ~ Gaussiano

Obs por grupo: min = 6

prom = 68.5

max = 105

Chi2 Wald(5) = 158.11

Log verosimilit. = -415.60602

Prob > chi2 = 0.0000

inlf	Coef.	Err. Std.	z	P> z	[95% Interval Conf.]
ingresop	-.0140176	.0048962	-2.86	0.004	-.023614 -.0044212
exper	.1216736	.0188992	6.44	0.000	.0846319 .1587153
exper2	-.0020391	.0006167	-3.31	0.001	-.0032479 -.0008303
hijos6	-.7920814	.1210485	-6.54	0.000	-1.029332 -.5548307
educ	.1326736	.025253	5.25	0.000	.0831786 .1821686
_cons	-1.994287	.3324886	-6.00	0.000	-2.645953 -1.342621
/lnsig2u	-1.724547	.6102672			-2.920649 -.5284452
sigma_u	.4222011	.1288278			.232161 .7678026
rho	.1512864	.0783576			.0511422 .3708796

Reajuste del modelo intpoints() = 16

Iteracion 0: log verosimil. = -415.60604

Iteracion 1: log verosimil. = -415.60602

Regresion probit efectos aleat.

Numero de obs = 753

Variable Grupo(i): grupo

Numero de grupos = 11

Efectos Aleat. $u_i \sim$ Gausiano

Obs por grupo: min = 6

prom = 68.5

max = 105

Chi2 Wald(5) = 158.11

Log verosimilit. = -415.60602

Prob > chi2 = 0.0000

	inlf	Coef.	Err.Std.	z	P> z	[95% Interval Conf.]	
ingresop		-.0140176	.0048962	-2.86	0.004	-.023614	-.0044212
exper		.1216736	.0188992	6.44	0.000	.0846319	.1587153
exper2		-.0020391	.0006167	-3.31	0.001	-.0032479	-.0008303
hijos6		-.7920814	.1210485	-6.54	0.000	-1.029332	-.5548307
educ		.1326736	.025253	5.25	0.000	.0831786	.1821686
_cons		-1.994287	.3324886	-6.00	0.000	-2.645953	-1.342621
/lnsig2u		-1.724547	.6102672			-2.920649	-.5284452
sigma_u		.4222011	.1288278			.232161	.7678026
rho		.1512864	.0783576			.0511422	.3708796

TABLA 4

Chequeo de Cuadratura

	Cuadratura ajustada 8 puntos	Cuadratura comparada 12 puntos	Cuadratura comparada 16 puntos	
Log verosimilit	-415.6071	-415.60602	-415.60602	
		.00108309	.00108309	Diferencia
		-2.606e-06	-2.606e-06	Diferencia relativa
inlf:	-.01402233	-.0140176	-.0140176	
ingresop		4.732e-06	4.732e-06	Diferencia

		-.00033746	-.00033746	Diferencia relativa

inlf:	.12166826	.12167357	.12167357	
exper		5.312e-06	5.312e-06	Diferencia
		.00004366	.00004366	Diferencia relativa

inlf:	-.00203948	-.00203912	-.00203912	
exper2		3.568e-07	3.568e-07	Diferencia
		-.00017494	-.00017494	Diferencia relativa

inlf:	-.79174834	-.79208144	-.79208144	
hijos6		-.0003331	-.0003331	Diferencia
		.00042071	.00042071	Diferencia relativa

inlf:	.13268519	.1326736	.1326736	
educ		-.00001159	-.00001159	Diferencia
		-.00008736	-.00008736	Diferencia relativa

inlf:	-1.9935596	-1.9942868	-1.9942868	
_cons		-.00072723	-.00072723	Diferencia
		.00036479	.00036479	Diferencia relativa

lnsig2u:	-1.7268816	-1.7245469	-1.7245469	
_cons		.00233467	.00233467	Diferencia
		-.00135196	-.00135196	Diferencia relativa

B.2. Caso Colombia

TABLA 5

Regresion probit efectos aleat.	Numero de obs	=	2928
Variable Grupo(i): codig	Numero de grupos	=	13
Efectos aleat $u_i \sim$ Gaussiano	Obs por grupo: min	=	178
	prom	=	225.2
	max	=	286
Log verosimilit. = -1807.4794	Chi2 Wald(5)	=	277.24
	Prob > chi2	=	0.0000

partic	Coef.	Std. Err.	z	P> z	[95% Conf. Interval]	
ingreso	-5.71e-08	1.97e-08	-2.90	0.004	-9.57e-08	-1.85e-08
exper	.0368012	.006254	5.88	0.000	.0245436	.0490588
exper2	-.0008962	.0000984	-9.11	0.000	-.001089	-.0007033
hijos6	-.0427552	.0318552	-1.34	0.180	-.1051903	.0196799
educat	.0375077	.006714	5.59	0.000	.0243484	.050667
_cons	-.2420671	.1455181	-1.66	0.096	-.5272773	.0431431
/lnsig2u	-2.855496	.4450994			-3.727875	-1.983117
sigma_u	.2398484	.0533782			.1550609	.370998
rho	.0543979	.0228954			.0234793	.1209869

Prueba razon-verosimilitud de rho=0: chibar2(01) = 72.59 Prob >= chibar2 = 0.000

TABLA 6

$y = \text{Pr}(\text{partic})$ (predict) = Probabilidad que una mujer casada con unas características promedio, participe en el mercado laboral es: 0.54458778

variable	dy/dx	Std. Err.	z	P> z	[95% C.I.]		X
ingreso	-2.43e-08	.00000	-3.17	0.002	-3.9e-08	-9.3e-09	959835
exper	.0135194	.00245	5.51	0.000	.008711	.018328	26.8801
exper2	-.0003399	.00004	-8.76	0.000	-.000416	-.000264	973.871
hijos6	-.0262662	.01236	-2.13	0.034	-.050491	-.002041	.587773
educat	.0136696	.00261	5.24	0.000	.00856	.018779	8.98429

TABLA 7

Reajuste del modelo intpoints() = 8

Iteracion 0: log verosimilit. = -1807.4794
 Iteracion 1: log verosimilit. = -1807.4794

Regresion probit efectos aleat.	Numero de obs	=	2928
Variable Grupo (i): codig	Numero de grupos	=	13
Efectos aleat. u_i ~ Gaussian	Obs por grupo: min	=	178
	prom	=	225.2
	max	=	286
	Wald chi2(5)	=	277.24
Log versoimilit = -1807.4794	Prob > chi2	=	0.0000

partic	Coef.	Std. Err.	z	P> z	[95% Conf. Interval]	
ingreso	-5.71e-08	1.97e-08	-2.90	0.004	-9.57e-08	-1.85e-08
exper	.0368011	.006254	5.88	0.000	.0245435	.0490588
exper2	-.0008962	.0000984	-9.11	0.000	-.001089	-.0007033
hijos6	-.0427553	.0318552	-1.34	0.180	-.1051904	.0196798
educat	.0375077	.006714	5.59	0.000	.0243484	.050667
_cons	-.2420475	.1455354	-1.66	0.096	-.5272916	.0431966
/lnsig2u	-2.855496	.4450975			-3.727871	-1.983121
sigma_u	.2398485	.053378			.1550612	.3709974
rho	.0543979	.0228953			.0234794	.1209866

Reajuste del modelo intpoints() = 12

Iteracion 0: log verosimilit = -1807.4794
 Iteracion 1: log verosimilit = -1807.4794

Regresion probit efectos aleat.	Numero de obs.	=	2928
Variable grupo(i): codig	Number of groups	=	13
Efectos Aleator u_i ~ Gaussian	Obs por grupo: min	=	178
	prom	=	225.2
	max	=	286

Log versosimilit = -1807.4794 Wald chi2(5) = 277.24
 Prob > chi2 = 0.0000

partic	Coef.	Std. Err.	z	P> z	[95% Conf. Interval]	
ingreso	-5.71e-08	1.97e-08	-2.90	0.004	-9.57e-08	-1.85e-08
exper	.0368011	.006254	5.88	0.000	.0245435	.0490588
exper2	-.0008962	.0000984	-9.11	0.000	-.001089	-.0007033
hijos6	-.0427553	.0318552	-1.34	0.180	-.1051904	.0196798
educat	.0375077	.006714	5.59	0.000	.0243484	.050667
_cons	-.2420475	.1455354	-1.66	0.096	-.5272916	.0431966
/lnsig2u	-2.855496	.4450975			-3.727871	-1.983121
sigma_u	.2398485	.053378			.1550612	.3709974
rho	.0543979	.0228953			.0234794	.1209866

TABLA 8

	Chequeo de Cuadratura			
	Cuadratura ajustada 4 puntos	Cuadratura comparada 8 puntos	Cuadratura comparada 12 puntos	
Log verosimilit.	-1807.4794	-1807.4794	-1807.4794	
		.00001241	.00001241	Diferencia
		-6.865e-09	-6.865e-09	Diferencia relativa
partic: ingreso	-5.712e-08	-5.712e-08	-5.712e-08	
		-1.822e-14	-1.822e-14	Diferencia
		3.190e-07	3.190e-07	Diferencia relativa
partic: exper	.03680117	.03680115	.03680115	
		-2.516e-08	-2.516e-08	Diferencia
		-6.838e-07	-6.838e-07	Diferencia relativa
partic: exper2	-.00089616	-.00089616	-.00089616	
		1.630e-10	1.630e-10	Diferencia
		-1.819e-07	-1.819e-07	Diferencia relativa
partic: hijos6	-.04275517	-.0427553	-.0427553	
		-1.300e-07	-1.300e-07	Diferencia

		3.041e-06	3.041e-06	Diferencia relativa

partic:	.03750771	.03750771	.03750771	
educat		-5.370e-09	-5.370e-09	Diferencia
		-1.432e-07	-1.432e-07	Diferencia relativa

partic:	-.24206709	-.24204749	-.24204749	
_cons		.0000196	.0000196	Diferencia
		-.00008095	-.00008095	Diferencia relativa

lnsig2u:	-2.8554961	-2.8554958	-2.8554958	
_cons		2.661e-07	2.662e-07	Diferencia
		-9.318e-08	-9.321e-08	Diferencia relativa

Bibliografía

- [1] Amemiya, T. (1981) "Qualitative Response Models: A Survey," *Journal of Econometrics Literature*, 19, 4, pp. 481-536.
- [2] Amemiya, T. (1985) *Advanced Econometrics* Harvard University Press : Cambridge, Massachusetts..
- [3] Batley, R.P.; Daly, A.J. (2004) "Departure time choice under uncertainty: a model of random utility maximisation for wagers" *European Transport Conference*, Strasbourg.
- [4] M. Ben Akiva, S. Lerman, (1985) "Discrete Choice Analysis: Theory and application to Travel Demand", MIT Press, Cambridge, Mass .
- [5] Blayac, T. y Causse, A. (2001) "Value of travel time: a theoretical legitimization of some nonlinear representative utility in discrete choice models," *Transportation Research*, Vol.35B.
- [6] Borjas y Sueyoshi. (1994). *Journal of Econometrics*, vol. 64, pp. 164-182
- [7] Butler, J.S., y Moffitt, R. (1982) "A computationally efficient quadrature procedure for the one-factor multinomial factor," *Econometrica*, 50, 761-764.
- [8] Cabrer Borrás, C;. Sancho Pérez, A; Serrano Domingo, G. (2001) *Microeconomía y Decisión*. Ed. Pirámide.
- [9] Domenich, T., y McFadden, D.L. (1975) "Urban Travel Demand: A Behavioral Analysis." North Holland Publishing Co.
- [10] Foerster, J. F. (1979). "Mode choice decision process models: a comparison of compensatory and non-compensatory structures". *Transportation Research*, 13A:17-28.
- [11] Geweke, J; Keune, and Runkle, D,. (1994) "Alternative computational approaches to inference in the multinomial probit model" *Review of Econometrics and Statistics*. 76, 609-632.
- [12] Geweke, K,. (1996) "Monte Carlo simulation and numerical integration." *Handbook of Computational Economics*. 731-800, Elsevier Science, Amsterdam.

- [13] Green W. H., (2002) *Econometric Analysis*. Prentice Hall; 5th edition.
- [14] Lee, L-f. (2000) "A numerically stable quadrature procedure for the one-factor random-component discrete choice model," *Journal of Econometrics*, Volume 95, Number 1, March 2000 , pp. 117-129(13).
- [15] Maddala. G. S., (2001) *Introduction to Econometrics*. Wiley, 3rd edition.
- [16] Maddala. G. S., (1983) "Limited Dependant and Qualitative variables in Econometrics" *Introduction to Econometrics*. Cambridge University Press, New York.
- [17] Manski, C. (1977) *The structure of random utility models*. Theory and Decision 8. 229-254.
- [18] Manski. Ch.; McFadden D., (1981). "Alternative Estimators and Sample Design for Discrete Choice Analysis" *Structural Analysis of Discrete data with Econometrics Application*. MIT Press, Cambridge, 2-30.
- [19] McFadden, D.(1974). "Conditional Logit Analysis of Qualitative Choice Behavior." *Frontiers in Econometrics*. Academic Press. 105-142.
- [20] McFadden, D.(1984). "Econometric Analysis of Qualitative Response Models." *Handbook of Econometrics*. Vol 2 Amsterdam.
- [21] McFadden, D.(1987). "Regression based Especification Test for the Multinomial Logit Model." *Journal of Econometrics*. 34. 63-82.
- [22] McFadden, D.(2000). "Economic Choices." *Prize Lecture*. December 8, 2000. 330-365.
- [23] Newey, Whitney, and Daniel McFadden (1994) *Handbook of Econometrics*, Volume 4, North Holland.
- [24] Ortúzar, J. de D. y L. G. Willumsen (2001) *Modelling Transport Third Edition*, John Wiley and Sons, Ltd., West Sussex, England.
- [25] Rouwendal, J. and Blaeij, A.T. de (2004). "Inconsistent and lexicographic choices in stated preference analysis." *Tinbergen Institute Discussion Paper*, Tinbergen Institute, Amsterdam TI 2004-038/3.
- [26] Sælensminde, K. (2002). *The impact of choice inconsistencies in stated choice studies Environmental and Resource Economics*, vol. 23, pp. 403-420
- [27] Tversky, A. (1972). "Elimination by aspects: a theory of choice", *Psychological review*, 79, pp.281-299.
- [28] Train, K. (2003). *Discrete Choice Methods with Simulation* Cambridge University Press, p. 19.

- [29] PSID Web-Site: <http://psidonline.isr.umich.edu/>
- [30] Zuluaga Díaz F. (2005) *Econometría de Datos de Panel: Revisión y una Aplicación*. Tesis de Maestría, Departamento de Ciencias Básicas, Universidad EAFIT, Medellín, Col. 111 pp.