

EXPERIMENTOS CON UN MODELO PROBIT DE PARTICION MODAL PARA SANTIAGO

Juan de Dios Ortázar S.
Departamento de Ingeniería de Transporte
Pontificia Universidad Católica de Chile
Casilla 6177, Santiago 22

Francisco Javier Achondo P.
CIS Ingenieros Consultores
Pedro de Valdivia 1370, 3er piso, Santiago

RESUMEN

El modelo Probit es el más general, pero más difícil de tratar, de los modelos desagregados de elección discreta. Entre sus ventajas está el permitir estructuras de correlación arbitrarias entre las alternativas, diferencias de gustos entre los individuos y la estimación de funciones de demanda para paneles de datos.

Usando información para el corredor Las Condes-Centro de Santiago, que incluye un pequeño panel de 45 individuos cuyas elecciones fueron observadas en 1981 y 1983, se calibran modelos Probit usando la aproximación de Clark y se comparan con modelos Logit más limitados. Para realizar la experiencia se debió agregar la información original, que comprendía datos para nueve alternativas y seis variables explicativas, a sólo tres opciones y tres atributos.

1. INTRODUCCION

Un problema sumamente generalizado en estudios de demanda por transporte, ha sido el contar básicamente con información recolectada para un sólo instante de tiempo. Esto ha dificultado enormemente la tarea de validación de los modelos estimados y no ha permitido inferir el potencial impacto del hábito y los procesos de aprendizaje en el comportamiento respecto a viajes (ver Williams y Ortúzar, 1982).

En los últimos años la literatura se ha preocupado de este problema y se han propuesto variados enfoques para el análisis de elecciones discretas en el tiempo (ver por ejemplo, Heckman, 1981). Una proposición particularmente interesante es la de Daganzo y Sheffi (1982) que permite utilizar, a costa de una transformación matemática, modelos previamente desarrollados para información correspondiente a un instante de tiempo (sección transversal) a datos para un panel (una serie en el tiempo de datos transversales para los mismos individuos).

Johnson y Hensher (1982) han aplicado esta técnica a un panel de dos períodos a fin de examinar el efecto de la experiencia acumulada en el primer período, en la elección del segundo período; también comparan las elasticidades derivadas de modelos estimados con los datos ignorando y considerando en forma explícita su naturaleza temporal, encontrando importantes diferencias en los resultados.

El modelo utilizado es una versión especial del Probit múltiple (Daganzo, 1979); sus características principales y método de estimación se discuten en la sección 2. Los datos se revisan someramente en la sección 3 y los resultados experimentales se presentan en la sección 4. Finalmente en la sección 5 se discuten nuestras principales conclusiones.

2. MODELO PARA UN PANEL DE DOS PERIODOS

Considérese un conjunto de individuos q , pertenecientes a una población o segmento de mercado Q , que tienen la posibilidad de elegir entre un máximo de N alternativas $A = \{A_n\}$ asociadas con una utilidad neta U_n , $n = 1, \dots, N$. En la mayoría de las aplicaciones se asume que las funciones de utilidad son lineales en los parámetros, vale decir:

$$U_n = \sum_k \theta_k Z_{nk} \quad (1)$$

donde θ es un vector de parámetros aleatorios con una cierta distribución en la población y Z_n es un vector k -dimensional de variables explicativas observables (características socio económicas de los individuos y atributos de las alternativas). Esta especificación es la más general posible ya que acepta la existencia de "variaciones en los gustos" de la población.

Ahora bien, si el vector de parámetros se distribuye Normal, el modelo de elección resultante toma la forma Probit múltiple (ver Daganzo, 1979); para estimar sus parámetros se dispone del programa experimental CHOMP desarrollado por Daganzo y Schoenfeld (1978).

Supongamos ahora que las N alternativas disponibles a cada individuo pueden ser elegidas en períodos $j = 1, 2$. La utilidad de una alternativa, que se escribe ahora U_{nj} , puede ser expresada como una función lineal del vector de variables explicativas Z_{nj} del individuo y la alternativa, en el periodo (Daganzo y Sheffi, 1982):

$$U_{nj} = \theta_j^T Z_{nj} \quad (2)$$

A fin de estimar los momentos de θ_j necesitamos información sobre elecciones observadas. Cabe destacar que si hay sólo un periodo j (esto es, el caso transversal) y θ se distribuye Normal, se debe estimar un modelo Probit corriente.

En este trabajo supondremos que los vectores de parámetros θ se distribuyen en forma conjunta Normal multivariada, pero son independientes de periodo a periodo (1). Aceptaremos sin embargo la existencia de dependencia estadual, ya que supondremos que la utilidad de la alternativa A_n en el periodo j depende, entre otras cosas, de la elección efectuada en el periodo previo (C_{j-1})

$$U_{nj} = \theta_j^T Z_{nj} + \phi \delta_{n(j-1)} \quad (3)$$

donde ϕ es un parámetro a ser estimado y $\delta_{n(j-1)}$ puede tomar los siguientes valores:

i) 1 si $j = 2$ y $C_1 = A_n$

ii) 0 en otros casos (si $j = 1$ o si $j = 2$ y $C_1 \neq A_n$)

Así, si ϕ es muy grande la elección del periodo anterior probablemente se repetirá; por otro lado, como θ_1 y θ_2 son independientes, si $\phi = 0$ ambas decisiones serán tomadas como si se tratara de distintas personas. Por lo tanto, ϕ tiende a capturar la inercia del sistema. Daganzo y Sheffi (1982) han propuesto reemplazar $\delta_{n(j-1)}$ por $U_{n(j-1)}$ en (3), a fin de tomar en cuenta el hecho de que aquellos individuos que exhiban un alto diferencial de utilidad en el periodo uno van a ser más reacios a cambiar, ceteris paribus, que personas que hayan escogido la misma alternativa pero con un diferencial de utilidad pequeño.

(1) Si no son independientes tenemos el problema de correlación serial o dependencia estadual espuria (ver Heckman, 1981)

Para el caso binario es posible ver fácilmente (Ortízar y Achondo, 1985) que si se define la utilidad de la alternativa elegida en ambos períodos como cero, el modelo anterior puede ser escrito como un modelo probit trinomial equivalente con las siguientes utilidades:

$$w_0 = 0$$

$$w_1 = \beta_1^T (Z_{NC1} - Z_{C1}) \quad (4)$$

$$w_2 = \beta_2^T (Z_{NC2} - Z_{C2}) + w_1$$

donde NC_i indica la alternativa no elegida y C_i denota a la elegida en el período i . Así, es posible estimar el modelo a través de evaluar la probabilidad de que cada individuo en la muestra escoga la alternativa de referencia. La extensión al caso trinomial es trivial, resultando un problema equivalente de tipo (4) pero con cinco opciones.

3. DATOS DISPONIBLES

La información utilizada en este estudio proviene de un banco de datos recolectado en 1981 y 1983 sobre elección de medio de transporte para el viaje al trabajo, en el corredor Las Condes Centro de Santiago, que es analizado en detalle por Ortízar y Achondo (1985). Dicho análisis dio, como resultado secundario, que había 45 individuos cuyo comportamiento había sido observado en ambos años, vale decir, un panel.

El escaso número de observaciones obligó a agrupar los datos; para ello se definieron tres alternativas agregadas:

- Auto particular (auto chofer y auto acompañante)
- Transporte público de superficie (bus y taxi colectivo)
- Metro (Metro y combinaciones)

Al hacer esta agrupación, tres individuos quedaron como cautivos a alguna alternativa (por ejemplo, tenían en realidad disponibles sólo bus y taxi colectivo originalmente).

La Tabla 1 presenta la frecuencia de elección de cada alternativa original en ambos años; entre paréntesis se entrega la información equivalente para la muestra completa. Como se puede ver, no todos los individuos escogen la misma alternativa en ambos períodos; la variación más importante ocurre en el caso del medio Bus-Metro, para el cual en la muestra general hay un incremento importante en el porcentaje que lo elige, y en nuestro panel baja de 14 % a 0.

Para interpretar en mejor forma la información de esta tabla es necesario recordar, en primer lugar, la gran variación en la situación económica del país en ambos años, y en segundo término la restricción aplicada en 1983 a los estacionamientos en el centro de Santiago.

TABLA N° 1
Frecuencia de elecciones en ambos períodos

Alternativa	% que la elige	
	1981	1983
1. Auto chofer	9 (112)	10 (125)
2. Auto acompañante	3 (33)	3 (41)
3. Taxi colectivo	3 (29)	5 (47)
4. Metro	9 (109)	9 (125)
5. Bus	2 (52)	3 (108)
6. Auto chofer Metro	8 (94)	7 (112)
7. Auto accomp. Metro	- (22)	2 (39)
8. Taxi colec. Metro	2 (27)	3 (25)
9. Bus Metro	6 (39)	- (77)
	42 (517)	42 (699)

La Tabla 2 presenta las elecciones observadas en 1981 y 1983 con las alternativas agregadas, para esta sub muestra de 42 individuos; como se puede ver, aún de esta forma se detectan importantes variaciones entre ambos años; esto es, la existencia de hábito no parece ser decisiva.

Tabla N° 2
Elecciones observadas al agrupar alternativas

Alternativa Elegida	Año	
	1981	1983
1	12	13
2	5	8
3	25	21

La Tabla 3 presenta finalmente el movimiento desde y hacia cada alternativa entre 1981 y 1983, de acuerdo a información recolectada en la muestra del segundo periodo. Como se puede ver el efecto "hábito" aparece más fuerte en los casos de Metro y transporte público de superficie (al menos 50 % en la diagonal); otro aspecto interesante de esta tabla, es la importante proporción de individuos que cambia de auto a otros medios, probablemente debido a la reducción en los sitios de estacionamiento en el centro antes mencionada.

TABLA N° 3

Cambio modal declarado entre 1981 y 1983

Alternativa escogida en 1981	Alternativa escogida en 1983		
	1	2	3
1. Auto	5	9	4
2. T. Público de Sup.	0	4	1
3. Metro	8	1	16

4. RESULTADOS DE LA MODELACION

En esta sección presentamos los resultados de estimar modelos probit para dos períodos y tres alternativas que, empleando la transformación descrita por Ortázar y Achondo (1985), conducen a modelos equivalentes para cinco alternativas con las siguientes funciones de utilidad:

$$\begin{aligned}
 V_0 &= 0 \\
 V_1 &= \theta_1 X_1 + \theta_2 X_2 \\
 V_2 &= \theta_1 X_3 + \theta_2 X_4 \\
 V_3 &= \theta_1 X_5 + \theta_2 X_6 + \theta_3 V_1 \\
 V_4 &= \theta_1 X_7 + \theta_2 X_8 + \theta_3 V_2
 \end{aligned} \tag{5}$$

en que X_1 y X_2 corresponden a la diferencia entre los valores de la variable Tiempo de Viaje de las dos alternativas no elegidas y la elegida, en el periodo 1; X_5 y X_7 es lo mismo para el periodo 2; X_3 y X_4 es lo equivalente para la variable Tiempo de Acceso en el periodo 1 y X_6 y X_8 , lo mismo para el periodo 2. Inicialmente se tenía considerado estimar modelos incluyendo la variable Costo de Viaje, pero esta resultó con signo contrario a la intuición en todas nuestras especificaciones.

Los valores de las variables anteriores se construyeron, a partir de los datos originales utilizando el siguiente procedimiento (Ortázar, 1983):

- para la opción escogida, el valor que tomaban en la alternativa elegida originalmente

- para las alternativas disponibles, el valor de la alternativa original con mayor utilidad de acuerdo a un modelo logit simple estimado para la muestra total de cada período (ver Ortázar, 1986).

Para estimar los modelos es también necesario especificar al programa CHOMP una estructura y parámetros de la matriz de varianza-covarianza asociada al modelo. La estructura más general para ésta, en nuestro caso, es la siguiente:

$$\Sigma = \begin{matrix} 0 & 0 & 0 & 0 & 0 \\ 0 & \sigma_{11} & 0 & \sigma_{13} & \sigma_{14} \\ 0 & 0 & \sigma_{22} & \sigma_{23} & \sigma_{24} \\ 0 & \sigma_{13} & \sigma_{23} & \sigma_{33} & 0 \\ 0 & \sigma_{14} & \sigma_{24} & 0 & \sigma_{44} \end{matrix} \quad (6)$$

que puede ser restringida a diversas versiones con menor cantidad de parámetros si no se cuenta con suficientes datos (2). Por esta misma razón, en la modelación utilizamos sólo las dos variables explicativas modales mencionadas anteriormente (tiempo de viaje y tiempo de acceso) y ninguna variable socio económica.

Nuestros resultados experimentales se presentan en las Tablas 4 a 7. En la primera se muestran modelos logit simple correspondientes a juntar las observaciones de ambos períodos como si se tratara de distintos individuos ("naïve pooling"). Como se puede ver, la variable Costo/Ingreso resulta con signo contrario a la intuición, por lo que se debe retirar de la función de utilidad; el modelo es sólo discreto en sus índices de bondad de ajuste, lo que no debe extrañar dado el escaso número de observaciones disponibles. La variable Tiempo de Viaje tiene buena significación y la variable Tiempo de Acceso es sólo significativa al 80 % de confianza.

(2) Cabe destacar que en principio todos estos parámetros son estimables, puesto que en matrices de varianza-covarianza para N alternativas, que tienen $N(N+1)/2$ parámetros libres, $N+1$ son redundantes y se deben fijar en valores de normalización arbitrarios (ver Albricht et. al., 1977; Dansie, 1985); en nuestro caso N vale 5, por lo que habría 6 parámetros arbitrarios, pero hemos fijado, por construcción del modelo, a 7 de ellos en 0.

TABLA N° 4

Modelos logit simple para todos los datos

Variable	Coeficiente (test t)	
	MNL-1	MNL-2
TVIAJE	-0,0996 (-2,4)	-0,0997 (-2,4)
TACCESO	-0,0640 (-1,3)	-0,0617 (-1,3)
COSTO/INGRESO	0,0122 (0,9)	-
AUTO	-0,8110 (-2,0)	-0,7076 (-1,8)
T. PUBLICO DE SUP.	-0,6957 (-1,7)	-0,6830 (-1,7)
 1(O)	-86,20	-86,20
1(E)	-71,69	-71,15
Rho cuadrado	0,168	0,175
Tamaño muestral	84	84

La Tabla 5 presenta modelos estimados con la información para 1983; es interesante destacar que ésta permite incluir una variable Hábito, que vale 1 si el individuo escoge la misma alternativa que en 1981 y 0 en otros casos. En una modelación similar, pero utilizando todo el conjunto de los datos, esta variable había resultado decisiva (ver Ortízar y Achondo, 1985); aca, si bien los parámetros de las variables restantes cambian levemente en magnitud y significancia, el efecto es menos marcado.

TABLA N° 5

Modelos logit simple para 1983

Variable	Coeficiente (test t)		
	MNL-1	MNL-2	MNL-H
TVIAJE	-0,1233 (-2,2)	-0,1158 (-2,1)	-0,0945 (-1,6)
TACCESO	-0,0848 (-1,3)	-0,0671 (-1,1)	-0,0603 (-0,9)
COSTO/INGRESO	0,0234 (1,9)	-	-
HABITO	-	-	0,6835 (1,9)
AUTO	-0,5950 (-1,1)	-0,3329 (-0,7)	-0,1459 (-0,3)
T. PUB. DE SUP.	-0,2157 (-0,4)	-0,1984 (-0,4)	-0,0403 (-0,1)
 1(O)	-42,49	-42,49	-42,49
1(E)	-35,83	-36,36	-34,55
Rho cuadrado	0,157	0,144	0,187
Tamaño muestral	42	42	42

La Tabla 6 muestra diversos modelos Probit estimados para el conjunto de datos, sin incorporar la información de que se trata de los mismos individuos en dos períodos diferentes.

TABLA N° 6
Modelos probit independiente para todos los datos

Variable	Coeficiente (test t)		
	MNP-1	MNP-2	MNP-3
TVIAJE	-0,0701 (-2,6)	-0,0341 (-2,7)	-0,0278 (-2,7)
TACCESO	-0,0383 (-2,0)	-0,0182 (-1,3)	-0,0144 (-1,4)
AUTO	-0,5299 (*)	-0,2506 (*)	-0,1925 (-9,9)
T. PUB. DE SUP.	-0,5754 (*)	-0,2729 (*)	-0,2257 (-4,2)
$(\sigma_{11})^{\frac{1}{2}}$	1,0000 (**)	0,4829 (7,3)	0,3701 (5,2)
$(\sigma_{22})^{\frac{1}{2}}$	1,0000 (**)	0,4829 (7,3)	0,4274 (8,1)
$(\sigma_{33})^{\frac{1}{2}}$	1,0000 (**)	0,4829 (7,3)	0,2366 (2,7)
 l(0)	 -92,28	 -92,28	 -92,28
l(0)	-72,82	-72,82	-72,61
Rho cuadrado	0,211	0,211	0,213
Tamaño muestral	84	84	84

(*) Valores de t superiores a 15, probablemente erróneos; se debe recordar que CHOMP es un programa experimental

(**) Valores fijados, no estimados

Para comparar estos valores con los de la Tabla 4 y entre sí, es necesario recordar algunas propiedades de las formas funcionales estimadas. En efecto, la función de probabilidad del modelo logit tiene la forma:

$$P_i = \exp(\alpha(\beta Z_i)) / \sum_j \exp(\alpha(\beta Z_j)) \quad (7)$$

en que Z son las variables explicativas y α está relacionado con la desviación estándar σ de las funciones de utilidad (recordar que la matriz de varianza-covarianza es diagonal y que todas las varianzas son iguales por construcción en este modelo), a través de la siguiente ecuación (Cochrane, 1975):

$$\alpha = 1,2825/\sigma \quad (8)$$

Por lo tanto, al no poder estimar α directamente (esto es equivalente a suponerlo igual a 1), en efecto se impone la condición de normalización de que el valor de σ sea igual a 1,2825 (ver Ortázar e Ivelic, 1986).

Ahora bien, en el modelo probit binario, que es el caso más fácil de examinar, se tiene que la forma de la probabilidad de elección es la siguiente (Daganzo, 1979):

$$P_1 = \Phi((\Omega(Z_1 - Z_2)/(\sigma_{11} + \sigma_{22} - 2\sigma_{12})) \quad (9)$$

en que Φ es la función normal acumulada estándar. Es fácil ver que si la matriz de varianza-covarianza es independiente e idéntica, como en el caso del modelo logit, los parámetros Ω estimados corresponden a las utilidades marginales deflactadas por 1,4142 σ .

Por lo tanto, para comparar los valores de Ω estimados para los modelos MNP-1 y MNP-2 de la Tabla 6, basta con dividir los correspondientes al segundo caso por $\sigma = 0,4829$; para comparar a los primeros con los de MNL-2 en la Tabla 4, se les debe multiplicar por 1,2825. En este caso se detectan mayores diferencias ya que las funciones de probabilidad no son las mismas.

La Tabla 7 presenta los resultados de la estimación del modelo probit múltiple correspondiente al panel de datos, de acuerdo a la especificación auxiliar (5) y a la matriz de varianza covarianza (6).

TABLA N° 7

Modelo probit múltiple para el panel de datos

Variable	Coeficiente (test t) MNP-P
TVIAJE	-0,0113 (-2,5)
TACCESO	-0,0076 (-5,7)
θ_3	0,4506 (1,2)
$(\sigma_{11})^{\frac{1}{2}}$	0,1646 (2,6)
$(\sigma_{22})^{\frac{1}{2}}$	0,0503 (1,0)
$(\sigma_{33})^{\frac{1}{2}}$	0,5700 (1,8)
$(\sigma_{44})^{\frac{1}{2}}$	0,0837 (6,7)
$\sigma_{13}/\sigma_{11}\sigma_{33}$	1,0000 (5,4)
$\sigma_{14}/\sigma_{11}\sigma_{44}$	0,9090 (2,9)
$\sigma_{23}/\sigma_{22}\sigma_{33}$	0,6677 (2,1)
$\sigma_{24}/\sigma_{22}\sigma_{44}$	0,4746 (1,4)
1(O)	-62,98
1(E)	-52,18
Rho cuadrado	0,172
Tamaño muestral	84

Los resultados anteriores muestran, como se esperaba, un efecto "hábito" moderado a través del valor 0,45 (que no resulta estadísticamente distinto de cero) del parámetro θ_3 que debe interpretarse de igual forma que ϕ en la ecuación (4). La raíz cuadrada de los términos de la diagonal de la matriz de varianza covarianza (6), indica claramente que las funciones de utilidad de las alternativas del modelo auxiliar (5) no son idénticas. Finalmente, los términos fuera de la diagonal (se presentan los parámetros correspondientes al coeficiente de correlación entre cada par de funciones de utilidad para las que se supuso que ésta no era cero), indican en algunos casos una fuerte dependencia estadual.

5. CONCLUSIONES

Se han estimado diversos modelos para un conjunto de datos consistente en 42 individuos cuyas elecciones fueron observadas en dos períodos diferentes. Los resultados indican que hay inercia en el proceso de elección y que cuando ésta se toma en cuenta, los valores asociados a los coeficientes de las variables de servicio en la función de utilidad de los modelos cambian.

La Tabla 8 presenta una comparación de la valoración relativa del tiempo de viaje respecto al tiempo de acceso en los principales modelos estimados; como se puede ver esta difiere bastante, sugiriendo también diferentes elasticidades y estimaciones de funciones de valoración como el valor subjetivo del tiempo (*).

TABLA N° 8

Ponderación relativa del tiempo de acceso respecto
del tiempo de viaje en los distintos modelos estimados

Modelo	TACCESO/TVIA
MNL-2 (Tabla 4)	0,619
MNL-H (Tabla 5)	0,638
MNP-3 (Tabla 6)	0,518
MNP-P (Tabla 7)	0,672

AGRADECIMIENTOS

Este trabajo contó con financiamiento parcial de la Dirección de Investigación de la Pontificia Universidad Católica de Chile (DIUC) y del FONDECYT.

(*) Desgraciadamente no contábamos con estimaciones para el coeficiente de la variable costo de viaje.

REFERENCIAS

- ALBRIGHT, R.L., LERMAN, S.R. y MANSKI, C.F. (1977) Report on the development of an estimation program for the multinomial probit model. Preliminary Report preparado para la Federal Highway Administration, Washington, D.C.
- COCHRANE, R.A. (1975) A possible economic basis for the gravity model. Journal of Transport Economics and Policy, Vol. 9, No. 1, 34-49.
- DAGANZO, C.F. (1979) Multinomial Probit: The Theory and its Applications to Demand Forecasting. Academic Press, Nueva York.
- DAGANZO, C.F. y SCHOENFELD, L. (1978) CHOMP user's manual. Research Report UCB-ITS-RR-78-7, Institute of Transportation Studies, University of California at Berkeley.
- DAGANZO, C.F. y SHEFFI, Y. (1982) A note on the treatment of categorical time series data with discrete choice models. 1982 ORSA/TIMS Meeting, San Diego, California.
- DANSIE, B.R. (1985) Parameter estimability in the multinomial probit model. Transportation Research, Vol. 19B, No. 6, 526-528.
- HECKMAN, J.J. (1981) Statistical models for discrete panel data. En C.F. Manski y D. McFadden (eds.), Structural Analysis of Discrete Data With Econometric Applications. MIT Press, Cambridge.
- JOHNSON, L.W. y HEMSHER, D.A. (1982) Application of multinomial probit to a two-period panel data set. Transportation Research, Vol. 16A, No. 5/6, 457-464.
- ORTUZAR, J. de D. (1983) Nested logit models for mixed-mode travel in urban corridors. Transportation Research, Vol. 17A, No. 4, 283-299.
- ORTUZAR, J. de D. (1986) The cultural and temporal transferability of discrete choice disaggregate modal split models. World Conference on Transport Research, Four Seasons Hotel, 11-15 Mayo 1986, Vancouver.
- ORTUZAR, J. de D. y ACHONDO, F.J. (1985) A pseudo-panel discrete mode choice model: preliminary findings. 13th PTRC Summer Annual Meeting, University of Sussex, 15-18 Julio 1985, Brighton.
- ORTUZAR, J. de D. e IVELIC, A.M. (1986) Efectos de la desagregación temporal de variables de servicio en la especificación y estabilidad de funciones de demanda. IV Congreso Panamericano de Ingeniería de Tránsito y Transporte, Hotel Carrera, 1-4 Diciembre 1986, Santiago.
- WILLIAMS, H.C.W.L. y ORTUZAR, J. de D. (1982) Behavioural theories of dispersion and the mis-specification of travel demand models. Transportation Research, Vol. 16B, No. 3, 167-219.