

## MODELOS MIXED LOGIT: ANTECEDENTES TEÓRICOS Y APLICACIONES

Ricardo Alvarez Daziano y Marcela A. Munizaga

Departamento de Ingeniería Civil, Universidad de Chile. Casilla 228-3, Santiago, Chile.

[mamuniza@cec.uchile.cl](mailto:mamuniza@cec.uchile.cl), [ralvarez@cec.uchile.cl](mailto:ralvarez@cec.uchile.cl)

<http://tamarugo.cec.uchile.cl/~dicidet/>

### RESUMEN

El desarrollo de modelos de demanda de transporte se puede describir como una búsqueda de modelos flexibles que se adapten a un mayor número de situaciones prácticas. Sin embargo, en esta búsqueda la flexibilidad ha comprometido la estimabilidad de los modelos. Por un lado aparecen los modelos tradicionales de la "familia" Logit que ofrecen probabilidades de elección cerradas, pero con supuestos simplificadorios que no siempre son sostenibles. Por otro, está el modelo Probit que permite trabajar con una estructura de error general, pero su estimación resulta bastante compleja. En este contexto, caracterizado además por avances tecnológicos en materia de computación y métodos numéricos, se ha cuestionado el uso de modelos simplificados y ha aparecido con fuerza una nueva alternativa de modelación: el modelo Mixed Logit.

En este artículo se revisa detalladamente los antecedentes teóricos que sustentan la formulación del modelo Mixed Logit, a través de un análisis de la matriz de covarianza se discute de qué manera estos modelos son capaces de modelar condiciones en las cuales se violan los supuestos de independencia y homoscedasticidad. Este análisis se complementa con dos aplicaciones numéricas que permiten comprobar la real posibilidad de utilizar este modelo y su capacidad de adaptarse a situaciones prácticas. En los experimentos de simulación se construye bases de datos que permiten controlar objetivamente la bondad de ajuste del modelo, la reproducción de la muestra de calibración y el nivel de respuesta a cambios en los atributos de las alternativas. La aplicación con datos reales pretende validar el estudio empírico y verificar la factibilidad de aplicar herramientas econométricas sofisticadas. A pesar de que su estimación requiere de simulación, se observa que en general el modelo entrega una adecuada reproducción de parámetros y un buen ajuste a los cambios de política.

Se concluye que el Mixed Logit en efecto constituye una alternativa interesante y poderosa para la modelación de elecciones discretas, aunque se debe ser riguroso en la construcción e implementación de una especificación particular, justificando adecuadamente los supuestos utilizados y teniendo claras sus consecuencias previo a la estimación de los parámetros.

## 1. INTRODUCCIÓN

El enfoque más utilizado en la actualidad para modelar elecciones discretas se basa en la teoría de la utilidad aleatoria (McFadden, 1974). Según esta teoría, para cada una de las alternativas  $i$  el individuo  $n$  tiene una función utilidad  $U_{in}$  asociada, escogiendo aquella alternativa que maximiza su utilidad. Esta función inherente del individuo, puede dividirse en una componente sistemática  $V_{in}$ , que recoge el efecto de las variables explicativas (atributos medibles u observables por parte del modelador), y una componente aleatoria  $\varepsilon_{in}$  que intenta recoger todos aquellos efectos no incluidos en la componente sistemática de la función de utilidad. De los supuestos que se asumen sobre la distribución del error estocástico se deriva los distintos modelos planteados en la literatura (Ortúzar y Willumsen, 1994).

Los modelos más utilizados en la actualidad son el Logit Multinomial (McFadden, 1974), que se obtiene a partir de asumir que los términos de error  $\varepsilon_{in}$  son iid Gumbel, y el Logit Jerárquico (Williams, 1977), que se deriva como una extensión del anterior, en que se considera que existe una componente de error adicional, que representa correlación en un grupo de alternativas. En síntesis, se trata de estructuras de covarianza (del término de error) muy básicas, lo cual es un supuesto simplificador que no siempre es sostenible, pero que permite tener modelos simples de entender y usar.

El modelo Probit (Daganzo, 1979), en cambio, se deriva a partir de suponer errores aleatorios distribuidos Normal multivariada, aceptando en teoría cualquier estructura de covarianza que los datos permitan estimar, lo cual implica un grado de dificultad de estimación considerable. A pesar de sus bondades en cuanto a flexibilidad, este modelo ha sido incorporado tímidamente a la práctica, pese a existir desde hace algún tiempo herramientas poderosas que permiten su estimación mediante simulación (ver Munizaga y Ortúzar, 1997).

En este contexto, caracterizado por un compromiso entre estimabilidad y flexibilidad, es que han irrumpido los modelos Mixed Logit -también conocidos como modelos de Error Compuesto o Probit con Kernel Logit- (Ben Akiva y Bolduc, 1996; Brownstone y Train, 1999). Se trata de una alternativa intermedia que se sitúa en algún punto entre el Logit y el Probit. Sus promotores claman que tiene la flexibilidad del Probit, manteniendo parte de la simpleza del Logit. En este trabajo se analiza los antecedentes teóricos que permiten formular este modelo, verificando la consecuencia de sus hipótesis y contrastándolas con las de otros modelos de elección.

## 2. EL MODELO MIXED LOGIT

### 2.1. Formulación

Los orígenes del modelo se remontan a principios de la década de los '80 cuando en los trabajos de Cardell y Dunbar (1980) y Boyd y Melman (1980) se describe un modelo equivalente a los actuales Mixed Logit, con el nombre de modelo Hedónico. Su reciente reaparición con otro nombre y renovados bríos puede deberse a los avances tecnológicos que permiten su estimación en menor tiempo. Recientemente este tipo de modelos ha sido utilizado para modelar diversas situaciones en distintos contextos (Train, 1999; Brownstone y Train, 1999; Algiers *et al.*, 1998).



Los modelos Mixed Logit nacen de suponer una función de utilidad  $U_{in}$  conformada por una componente determinística  $V_{in}$ , una componente aleatoria  $\varepsilon_{in}$  independiente e idénticamente distribuida, y uno o más términos aleatorios adicionales. Estos términos de error adicionales pueden ser agrupados en un término aditivo  $\eta_{in}$ , que puede ser función de datos observados de la alternativa, y que permite recoger la presencia de correlación y heteroscedasticidad. Así, la función de utilidad queda definida como

$$U_{in} = V_{in} + \eta_{in} + \varepsilon_{in} \quad (1)$$

donde  $\varepsilon_{in} \sim \text{Gumbel}(0, \lambda)$  y  $\eta_{in} \sim f(\eta/\theta^*)$ , donde  $f$  es una función densidad general y  $\theta^*$  son parámetros fijos que la describen (e.g. media y varianza)<sup>1</sup>. Como  $\varepsilon$  es iid Gumbel, entonces la probabilidad condicional en  $\eta$  de que el individuo  $n$  escoja la alternativa  $i$  corresponde exactamente a un modelo Logit Multinomial (MNL):

$$P_n(i/\eta) = \frac{\exp(V_{in} + \eta_{in})}{\sum_{j=1}^{J_n} \exp(V_{jn} + \eta_{jn})} \quad (2)$$

Para obtenerse la probabilidad de elegir la alternativa, debe evaluarse la expresión anterior sobre todos los valores posibles  $\eta$ , lo que corresponde a la integral de la probabilidad condicional por la función densidad del término aleatorio  $\eta$ , esto es:

$$P_{in} = \int P_n(i/\eta) f(\eta/\theta^*) d\eta \quad (3)$$

Como caso particular, puede suponerse una función de utilidad con la siguiente especificación<sup>2</sup>:

$$U_{in} = \underbrace{\beta' x_{in}}_{V_{in}} + \underbrace{\mu_{in}' z_{in}}_{\eta_{in}} + \varepsilon_{in} \quad (4)$$

En esta expresión se asume que la componente determinística de la utilidad es lineal en los parámetros  $\beta$  que ponderan a los atributos  $x_{in}$ . Por otro lado, se asume que  $\eta$  depende de ciertos parámetros ( $\mu_{in}$ ) y datos observados relacionados con la alternativa  $i$  ( $z_{in}$ ), relación que también se supone lineal en parámetros. Un supuesto adicional, que ha sido utilizado en la mayor parte de los estudios previos (Ben Akiva y Bolduc, 1996; Brownstone y Train, 1999), es que el término  $\mu$  es propio del individuo, sin variar entre alternativas. Es decir:  $\eta_{in} = \mu_n' z_{in}$ .

<sup>1</sup> La distribución de los términos aleatorios generalmente se asume Normal, existiendo diversos argumentos detrás de este supuesto. Otra distribución que ha sido utilizada es la Lognormal, especialmente en aquellos casos en que se quiere restringir el signo de un determinado parámetro.

<sup>2</sup>  $\beta$  es un vector de parámetros de dimensión  $L$  (hay  $L$  variables explicativas en la componente determinística de la función de utilidad);  $x_{in}$  es un vector de atributos de dimensión  $L$ ;  $\mu_{in}$  es un vector aleatorio de dimensión  $K$  cuyas componentes tienen media cero y con matriz de covarianza  $\Omega$ ;  $z_{in}$  es un vector de atributos asociados con la alternativa  $i$  y el individuo  $n$ , y tiene dimensión  $K$ ; finalmente,  $\varepsilon_{in}$  es una variable aleatoria que representa el error estocástico.

## 2.2. Matriz de Covarianza

Dada una función de utilidad como en (4) y considerando además el supuesto usual  $\eta_{in} = \mu_n' z_{in}$ , sea  $z_n$  la matriz de dimensión  $K \times J$  que contiene a los vectores  $z_{in}$  para cada alternativa perteneciente al conjunto de elección del individuo ( $i \in C_n$ ) y  $\varepsilon_n$  un vector aleatorio iid Gumbel con matriz de covarianza  $\Sigma_\varepsilon$  y que contiene a los elementos  $\varepsilon_{in}$ . Si se asume que cada término de  $\mu_n$  tiene una función densidad con media cero y varianza  $\sigma_k^2$  y que el vector en su conjunto tiene una matriz de covarianza  $\Omega$ , entonces la matriz de covarianza del modelo ( $\Sigma$ ), puede escribirse como:

$$\Sigma = z_n' \cdot \Omega \cdot z_n + \Sigma_\varepsilon = z_n' \cdot \Omega \cdot z_n + \sigma_\varepsilon^2 I \quad (5)$$

Es fácil demostrar que la matriz de covarianza obtenida es simétrica y definida positiva, y que su dimensión está bien definida<sup>3</sup>. De esta expresión general se concluye que el modelo es capaz de modelar correlación y heteroscedasticidad entre alternativas. En particular, dependiendo de la especificación que se le dé al término de error adicional, es posible modelar situaciones como alternativas similares y variaciones en los gustos. Si se obtiene la covarianza entre dos alternativas se observa que para  $i, j \in C_n$  con  $i \neq j$ :

$$\text{cov}(U_{in}, U_{jn}) = \sum_{k=1}^K z_{kin} z_{kjn} \sigma_k^2 \quad (6)$$

que en general será distinto de cero si es que para al menos algún  $k$ ,  $\sigma_k^2 > 0$  y  $z_{kin}, z_{kjn} \neq 0$ . En tal caso, se asegura la presencia de correlación entre las alternativas  $i$  y  $j$ . Por otro lado,

$$\text{var}(U_{in}) = \sum_{k=1}^K z_{kin}^2 \sigma_k^2 + \frac{\pi}{6\lambda^2} \quad (7)$$

luego si  $\text{var}(U_{in}) \neq \text{var}(U_{jn})$  se asegura heteroscedasticidad entre dichas alternativas.

Se puede ver que se trata de una forma distinta de plantear y justificar un determinado modelo. La forma usual es hacer supuestos directamente sobre la covarianza de  $\varepsilon_{in}$ , como por ejemplo en el caso del Probit. En cambio, en un modelo Mixed Logit lo que se hace es agregar términos que sean fuente de correlación y/o heteroscedasticidad obteniéndose como resultado una determinada estructura de covarianza.

<sup>3</sup> La matriz de covarianza es de dimensión  $J \times J$ . En efecto, como  $\Omega$  es de dimensión  $K \times K$  (con  $K$  el número de componentes aleatorias), y  $z_n$  tiene dimensión  $K \times J$ , entonces  $z_n' \Omega z_n$  es una matriz de dimensión  $J \times J$ ; luego al sumarse esta última con la matriz  $\Sigma_\varepsilon$ , cuya dimensión corresponde a  $J \times J$ , se demuestra finalmente que  $\dim \Sigma = \dim(z_n' \cdot \Omega \cdot z_n + \Sigma_\varepsilon) = J \times J$ .



### 2.3. Algunas Propiedades del Mixed Logit

Probablemente la propiedad más interesante de este modelo es que bajo ciertas condiciones de regularidad cualquier modelo de utilidad aleatoria tiene probabilidades de elección que pueden ser aproximadas tan cerca como se desee por un Mixed Logit (McFadden y Train, 2001). Es así como, por ejemplo, un modelo Mixed Logit con parámetros aleatorios distribuidos Normal puede aproximar a un modelo Probit.

Por otro lado, al permitir modelar alternativas correlacionadas, el Mixed Logit es capaz de levantar el supuesto de independencia de alternativas irrelevantes propio del MNL. De esta manera, los patrones de sustitución entre alternativas son flexibles y son un resultado de los datos y no una imposición particular, como ocurre en el MNL.

La relación entre un modelo Mixed Logit y los modelos de parámetros –formulados con el objetivo de modelar variaciones en los gustos– es directa. En ese sentido, se puede plantear que los parámetros de gusto  $\beta_n$  poseen una media poblacional  $\beta$  más una componente aleatoria  $\mu_n$  que representa la variación en los gustos del individuo  $n$  con respecto a la media. Esto se puede interpretar como una estructura análoga a (4), en la cual  $\eta_{in} = \mu_n' z_{in}$  y  $z = x$  (Brownstone y Train, 1999).

### 2.4. Estimación

La probabilidad de elección de un modelo Mixed Logit como la expresada en la ecuación (3), no posee una expresión matemática cerrada a diferencia del Logit Multinomial o del Jerárquico. Es más, como la integral no puede resolverse analíticamente es preciso usar algún método de integración. En la literatura es posible observar que se prefiere utilizar simulación de Monte Carlo por sobre otros métodos como por ejemplo integración numérica por cuadratura (Bhat, 2000). Además, se puede aprovechar el hecho de que la probabilidad condicional (2) tiene una expresión de forma Logit Multinomial. Así, si se considera  $R$  valores de  $\eta$  obtenidos de su función densidad  $f(\eta/\theta^*)$ , para cada una de las repeticiones es posible calcular

$$P_n(i/\eta^r) = \frac{\exp(V_{in} + \eta_{in}^r)}{\sum_{j=1}^J \exp(V_{jn} + \eta_{jn}^r)}, \quad (8)$$

con  $r=1, \dots, R$ . Luego es posible obtener la probabilidad promedio

$$\tilde{P}(i) = \frac{1}{R} \sum_{r=1}^R P_n(i/\eta^r), \quad (9)$$

y con ella construir la función de verosimilitud simulada

$$SL = \sum_{n \in Q} \sum_{i=1}^J y_{in} \ln \tilde{P}(i) \quad (10)$$

Bajo condiciones de regularidad, el estimador máximo verosímil simulado es consistente y asintóticamente normal. A pesar de que (10) es un estimador insesgado de la probabilidad, su logaritmo natural resulta ser sesgado (Brownstone y Train, 1999); sin embargo, cuando el número de repeticiones crece más rápido que la raíz cuadrada del número de observaciones, el estimador resulta asintóticamente equivalente al estimador máximo verosímil (Hajivassilou y Ruud, 1994).

### 3. EXPERIMENTOS DE SIMULACIÓN

Siguiendo la metodología de Williams y Ortúzar (1982), se realizó un experimento de simulación con el fin de comprobar la factibilidad real de aplicación del modelo. Para ello se creó distintas bases de datos sintéticas considerando medias y desviaciones de los atributos a partir de una base real construida para Santiago. Se consideró una situación de elección modal con cuatro alternativas (auto, bus, metro y taxi) y cuatro variables explicativas (costo, tiempo de viaje, tiempo de acceso, dummy de ingreso).

Se pretende modelar el caso en que las alternativas bus y metro son consideradas como similares entre sí. Para construir la parte estocástica de la función utilidad se trabajó con el enfoque planteado por Brownstone y Train (1999), considerado por los autores como un modelo Mixed Logit (ML) “análogo” al Logit Jerárquico (LJ). Este modelo particular se construye agrupando las alternativas en nidos, agregando un parámetro aleatorio asociado a una variable muda propia de cada nido (que indica si la alternativa pertenece o no a éste). La presencia del parámetro aleatorio común para aquellas alternativas que comparten nido, permite obtener una covarianza con presencia de elementos fuera de la diagonal, obteniéndose un patrón de correlación similar al de un LJ, pero no necesariamente una estructura de covarianza análoga. De hecho es posible comprobar cierto compromiso entre correlación y heteroscedasticidad en esta especificación (Munizaga y Alvarez, 2000). Así, se consideró un término de error iid Gumbel(0,  $\lambda$ ) y, adicionalmente, un término de error  $\mu_n$  distribuido Normal(0,  $\sigma_\mu^2$ ) con el fin de modelar correlación. En primer lugar se consideró este último sólo en las alternativas bus y metro, lo que sigue el enfoque planteado por Brownstone y Train, cuya matriz de covarianza resulta heteroscedástica<sup>4</sup>. Luego se consideró errores iid Normal para las alternativas no anidadas con el fin de obtener una matriz homoscedástica correspondiente a una estructura teóricamente modelable por un LJ.

Al asumir  $\mu_n \sim N(0, \sigma_\mu^2)$ , entonces es posible escribir  $\mu_n = s_n \sigma_\mu$  con  $s_n$  distribuido Normal estándar. Como no se conoce  $\sigma_\mu$ , entonces interesa estimar su valor, con el cual queda determinada la distribución del término de error adicional. Nótese que al poseer el modelo un término de error Gumbel, el parámetro estimado se verá escalado y se cumplirá que  $\hat{\sigma}_\mu = \lambda \sigma_\mu$ . Para la estimación de los modelos MNL, LJ y Probit se utilizó un código propio programado en Gauss (Aptech Systems, 1994) basado en la rutina de máxima verosimilitud. Para la estimación del modelo Mixed Logit se utilizó un código flexible en Gauss preparado por Kenneth Train, disponible en su página web<sup>5</sup>.

<sup>4</sup> Debido a que la varianza total del error resulta mayor para las alternativas anidadas (Munizaga y Alvarez, 2000). En ese sentido la especificación original no posee una estructura de covarianza análoga a la de un LJ.

<sup>5</sup> <http://elsa.berkeley.edu/~train/software.html>



Se trabajó con 8000 observaciones, y se consideró un coeficiente de correlación igual a 0,5 entre las alternativas bus y metro<sup>6</sup>, lo que implica que  $\sigma_u = \sigma_e$  (Munizaga y Alvarez, 2000). Primero se generó una base de datos a partir de una matriz de covarianza heteroscedástica, mientras que la segunda base posee una matriz homoscedástica. La varianza total del error se eligió de manera tal que el factor de escala de un modelo MNL fuese igual a uno, asegurándose de que el experimento no fuese completamente determinístico ni aleatorio.

La estimación del Probit se efectuó utilizando el simulador de probabilidades GHK con 10 repeticiones. A su vez, para el Mixed Logit se usó con 200 repeticiones usando números aleatorios basados en series de Halton (Train, 1999). Los valores reportados corresponden a corridas realizadas en un computador personal con un procesador Pentium II de 450 MHz y 64 MB de memoria RAM. Los resultados de las estimaciones del MNL, LJ, Probit y ML para cada situación de elección se presentan en la Tabla 1, en que también se entrega cada valor de referencia. La tabla presenta las estimaciones de los parámetros para cada modelo, el estadígrafo t de significancia estadística y el test t sobre el valor de referencia del parámetro para el ML. Para el LJ, el valor de referencia de es calculado a partir de la correlación simulada.

**Tabla 1: Resultados Simulación**

	<i>Base Heteroscedástica</i>					<i>Base Homoscedástica</i>				
	Ref.	MNL	LJ	Probit	ML	Ref.	MNL	LJ	Probit	ML
<b>Cte Auto</b>	-0,40	-0,3402 (-4,791)	-0,3872 (-5,469)	-0,2434 (-4,743)	-0,2946 (-5,168) [1,848]	-0,40	-0,1489 (-2,143)	-0,2843 (-4,269)	-0,1921 (-3,2748)	-0,4619 (-7,982) [1,070]
<b>Cte Metro</b>	0,20	0,3636 (5,848)	0,3171 (5,194)	0,2267 (5,322)	0,2512 (5,263) [1,072]	0,20	0,4884 (7,982)	0,3605 (6,530)	0,2982 (6,768)	0,1731 (3,666) [-0,569]
<b>Cte Taxi</b>	-0,45	-0,7007 (-12,557)	-0,7698 (-13,065)	-0,4831 (-12,957)	-0,5218 (-12,116) [1,815]	-0,45	-0,1166 (-2,274)	-0,3101 (-5,905)	-0,2043 (-4,159)	-0,4610 (-10,967) [-0,261]
<b>Costo</b>	-0,005	-0,0070 (-11,359)	-0,0070 (-11,279)	-0,0049 (-10,912)	-0,0055 (-11,011) [-0,995]	-0,005	-0,0053 (-8,901)	-0,0052 (-8,453)	-0,0041 (-8,067)	-0,0049 (-9,721) [0,232]
<b>Tiempo de Viaje</b>	-0,08	-0,1044 (-36,560)	-0,1005 (-32,035)	-0,0702 (-30,768)	-0,0804 (-31,172) [-0,148]	-0,08	-0,0835 (-31,464)	-0,0760 (-27,932)	-0,0614 (-21,918)	-0,0791 (-30,845) [0,344]
<b>Tiempo de Acceso</b>	-0,16	-0,2012 (-47,029)	-0,1954 (-41,264)	-0,1379 (-35,499)	-0,1563 (-36,078) [0,860]	-0,16	-0,1765 (-44,919)	-0,1643 (-38,757)	-0,1323 (-23,712)	-0,1596 (-36,698) [0,103]
<b>Ingreso Alto</b>	1,2	1,4928 (24,094)	1,4755 (23,953)	1,0686 (21,306)	1,1776 (21,464) [0,409]	1,2	1,2454 (21,117)	1,2174 (20,828)	0,9998 (15,243)	1,1866 (21,842) [0,247]
$\phi$	0,7071		0,8945 (23,953)			0,7071		0,7458 (22,558)		
$\sigma_u$	0,9069			0,5100 (4,597)	0,7601 (8,352) [1,613]	0,9069			0,5441 (5,880)	0,8472 (9,350) [-0,658]
<b>Iterac.</b>		5	5	6	3		5	5	7	2
<b>l(0)</b>		-0,93469	-0,93426	-0,93688	-0,93291		-1,03180	-1,02919	-1,03138	-1,02867
<b>Tiempo [min]</b>		0,6	0,8	35,5	42,5		0,7	0,8	35,2	152,5

(-) Test t con respecto a cero [-] Test t con respecto al parámetro de referencia

<sup>6</sup> Se trabajó con coeficientes de correlación mayores, pero se optó por reportar los resultados para este nivel, debido a que generalmente los grados de correlación prácticos no resultan sustancialmente altos.

El modelo ML permite recuperar adecuadamente los valores de todos los parámetros de gusto con que fue generada la base de datos, lo cual es mostrado por el estadígrafo  $t$ , que es menor que 1,96 en todos los casos. El MNL presenta parámetros que son distintos de los valores verdaderos, especialmente en el caso de correlación y heteroscedasticidad. Cabe hacer notar que para la base heteroscedástica se observa ciertos efectos de escala sobre los parámetros estimados por el modelo LJ, lo que reafirma ciertos estudios empíricos previos (Munizaga y Álvarez, 2000). Por otro lado, a pesar del alto número de observaciones, el Probit presenta ciertas dificultades para recoger los parámetros asociados a la correlación presente.

En la Tabla 2 se muestra los cambios de política considerados para efectuar un análisis de respuesta de los modelos (Williams y Ortúzar, 1982); es posible ver que las políticas definidas corresponden a cambios fuertes en los valores de los atributos, aumentando al doble o disminuyendo a la mitad algunos valores en cada caso. Como medida de error para cada una de las políticas definidas, se utiliza el índice Chi cuadrado de Gunn y Bates (1982), reportado en la Tabla 3; éste se calcula como  $\chi^2 = \sum_i \frac{(\hat{N}_i - N_i)^2}{N_i}$ , donde  $\hat{N}_i$  es el número de individuos que elige la alternativa  $i$  según la predicción hecha por el modelo, y  $N_i$  es el número de individuos que elige la alternativa  $i$  de acuerdo al modelo de simulación.

Tabla 2: Cambios de Política

	Costo de Viaje				Tiempo de Viaje				Tiempo de Acceso			
	Auto	Bus	Metro	Taxi	Auto	Bus	Metro	Taxi	Auto	Bus	Metro	Taxi
P1	2,0											
P2			2,0						1,5			
P3		2,0				0,5						
P4		0,5	1,5			2,0	0,3					
P5									1,5			2,0
P6	2,0		0,5		0,5	1,5				2,0		

Tabla 3: Índice  $\chi^2$ 

	Base Heteroscedástica				Base Homoscedástica			
	MNL	LJ	Probit	ML	MNL	LJ	Probit	ML
Base	0,00	0,03	0,40	0,00	0,00	0,00	0,53	0,01
P1	10,21	10,62	9,00	6,01	1,78	1,79	0,67	2,52
P2	9,65	9,81	7,50	5,65	2,52	0,91	0,81	1,42
P3	4,32	3,41	1,16	1,41	34,34	11,79	11,88	11,37
P4	10,35	6,64	3,41	8,03	11,63	5,44	13,46	5,41
P5	8,99	8,24	3,96	5,16	9,23	8,25	1,54	6,01
P6	1,87	3,40	3,33	2,64	4,15	5,01	4,91	6,64

$$\chi^2_{95\%,3} = 7,815$$

Las mayores diferencias entre las predicciones de los modelos y los valores simulados (realidad virtual) se encuentran para el MNL para ambos casos (homo/heteroscedástico). Las predicciones del LJ y ML son bastante similares (Ver Tabla 4), prácticamente indistinguibles, para el caso homoscedástico, siendo ambas a su vez muy parecidas a lo que se ha llamado "realidad virtual". Sin embargo, si la base es heteroscedástica se observa algunas pequeñas diferencias entre las predicciones de ambos modelos.



**Tabla 4: Particiones de Mercado**

	Alt.	Base Heteroscedástica							Base Homoscedástica						
		Base	P1	P2	P3	P4	P5	P6	Base	P1	P2	P3	P4	P5	P6
BASE	1	3225	2482	2556	2754	2961	2837	3818	3244	2542	2596	2837	2973	2919	3665
	2	1056	1146	1190	2262	157	1341	157	918	1090	1113	2050	162	1290	158
	3	2625	3017	2862	2103	3858	3327	2897	2498	2843	2729	1975	3626	3171	2799
	4	1094	1355	1392	881	1024	495	1128	1340	1525	1562	1138	1239	620	1378
MNL	1	3227	2445	2495	2675	2941	2713	3806	3244	2512	2548	2820	3014	2836	3703
	2	1055	1232	1266	2281	193	1393	162	918	1065	1093	1869	193	1239	159
	3	2624	3034	2908	2166	3882	3398	2865	2498	2867	2758	2165	3557	3289	2717
	4	1094	1289	1331	878	983	496	1166	1340	1556	1601	1145	1236	636	1421
LJ	1	3217	2438	2500	2681	2927	2717	3743	3242	2510	2577	2847	3005	2859	3588
	2	1058	1234	1269	2317	182	1391	157	919	1066	1097	1929	168	1228	145
	3	2628	3037	2901	2116	3907	3391	2931	2499	2869	2735	2066	3594	3262	2843
	4	1096	1291	1329	887	984	502	1169	1340	1556	1591	1159	1232	650	1425
PROBIT	1	3238	2466	2513	2725	3000	2791	3752	3263	2545	2589	2860	3068	2895	3643
	2	1067	1239	1275	2305	175	1405	150	932	1082	1115	1938	165	1270	139
	3	2610	2984	2859	2103	3830	3324	2928	2478	2826	2709	2080	3528	3227	2784
	4	1090	1318	1361	875	997	486	1170	1332	1553	1594	1129	1240	614	1435
ML	1	3224	2439	2491	2715	2945	2780	3746	3245	2504	2579	2861	3013	2884	3570
	2	1057	1221	1257	2297	180	1369	158	920	1065	1095	1927	167	1227	145
	3	2625	3008	2877	2092	3889	3345	2948	2498	2864	2724	2059	3591	3255	2854
	4	1094	1332	1375	896	986	506	1148	1338	1568	1602	1153	1228	633	1431

#### 4. APLICACIÓN A UNA BASE DE DATOS REAL

Como una forma de validar el estudio empírico se utilizó la base de datos reales del corredor Las Condes – Centro (Ortúzar y Donoso, 1983), básicamente por la estrictez de su construcción como por el hecho de que ha sido ampliamente estudiada, lo que permite contrastar fácilmente los resultados con estudios previos. La muestra consta de 697 observaciones y 9 modos alternativos. Se trabajó con las siguientes variables de servicio: TDV (tiempo de viaje), TCAM (tiempo de caminata), TESP (tiempo de espera), C/w (costo dividido por tasa salarial<sup>7</sup>). Se supuso la existencia de correlación entre alternativas de transporte público, por lo que se consideró la estructura anidada presentada en la Figura 1<sup>8</sup>. Los modelos estimados fueron el MNL, Probit Independiente, LJ, Probit homoscedástico, y dos especificaciones ML. La primera (MLHe), heteroscedástica, que sólo considera un término adicional dentro del nido; la segunda (MLHo), homoscedástica, que considera además términos de error adicionales independientes en las alternativas no anidadas. Los resultados de la estimación se reportan en la Tabla 5. El coeficiente de correlación al interior del nido se puede obtener, para el LJ, a través del parámetro estructural  $\phi$ , mientras que para el Probit corresponde a un parámetro más a estimar. En el caso del ML, lo que se estima es  $\hat{\sigma}_\mu$  y se puede demostrar fácilmente que:

$$\hat{\rho}_{ML} = \frac{6\hat{\sigma}_\mu^2}{6\hat{\sigma}_\mu^2 + \pi^2} \quad (11)$$

<sup>7</sup> Se considera esta variable como una aproximación a un modelo de tasa salarial.

<sup>8</sup> Se probó con otras estructuras de nidos. En este trabajo se reporta la estructura que permite hacer el mayor número de comparaciones.

Figura 1: Estructura Jerárquica

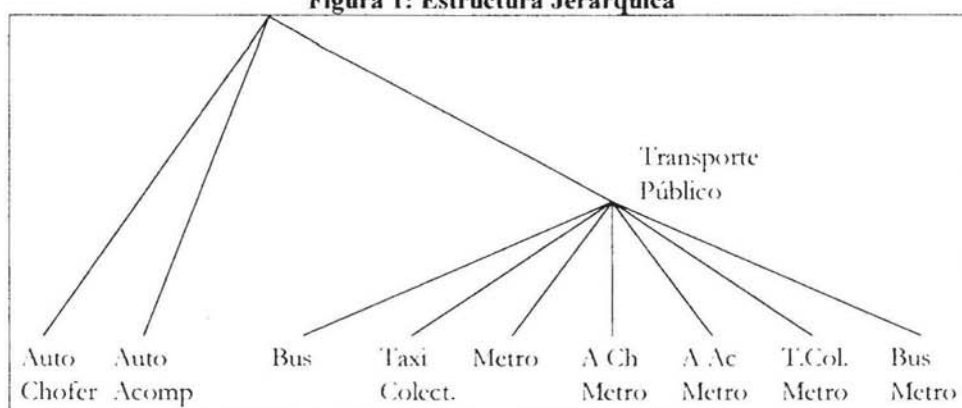


Tabla 5: Resultados Base Las Condes – Centro

	MNL	Probit Ind.	LJ	Probit	MLHe	MLHo
<b>Auto Ac.</b>	0,1482 (0,368)	0,0092 (0,039)	0,0281 (0,066)	-0,0040 (0,005)	0,0833 (0,194)	-0,0087 (0,019)
<b>Taxi Col.</b>	0,7229 (1,826)	0,3688 (1,575)	0,7217 (1,818)	0,4083 (0,493)	1,1953 (2,391)	1,0986 (2,112)
<b>Metro</b>	4,6685 (10,280)	3,2713 (11,087)	4,4970 (8,231)	3,2466 (3,000)	5,5311 (8,283)	5,2850 (7,613)
<b>Bus</b>	2,2387 (5,025)	1,2460 (4,574)	2,1074 (0,335)	1,2705 (1,919)	2,8913 (5,017)	2,7154 (4,587)
<b>Auto Ch – Metro</b>	0,1868 (0,942)	0,1277 (0,905)	0,3350 (1,571)	0,1512 (0,524)	0,4834 (1,661)	0,4005 (1,415)
<b>Auto Ac – Metro</b>	0,9853 (2,202)	0,6251 (2,238)	1,1664 (2,609)	0,6600 (0,7516)	1,5016 (2,747)	1,3699 (2,451)
<b>T Col – Metro</b>	0,4988 (1,164)	0,3425 (1,281)	0,6620 (1,452)	0,3896 (0,394)	1,0353 (1,930)	0,9128 (1,645)
<b>Bus – Metro</b>	1,7888 (4,184)	1,1266 (4,106)	1,8777 (4,375)	1,1164 (1,018)	2,3699 (4,356)	2,2272 (3,959)
<b>TDV</b>	-0,0823 (-4,743)	-0,0554 (-4,031)	-0,0907 (-4,002)	-0,0550 (-1,774)	-0,0951 (-4,936)	-0,0909 (-4,783)
<b>TCAM</b>	-0,1610 (-8,625)	-0,1077 (-8,662)	-0,1531 (-7,019)	-0,1067 (-7,702)	-0,1904 (-7,488)	-0,1807 (-7,007)
<b>TESP</b>	-0,2359 (-2,238)	-0,1475 (-2,028)	-0,2170 (-1,966)	-0,1484 (-0,800)	-0,2741 (-2,498)	-0,2641 (-2,426)
<b>C/w</b>	-0,0244 (-3,647)	-0,0143 (-2,028)	-0,0228 (-2,854)	-0,0142 (-1,979)	-0,0253 (-3,318)	-0,0267 (-3,424)
<b>SEXO</b>	-0,2951 (-1,361)	-0,1531 (-1,169)	-0,2627 (-1,269)	-0,1479 (-1,049)	-0,2830 (-1,273)	-0,2923 (-1,305)
<b>AUTOLIC</b>	2,3606 (5,786)	1,4889 (5,902)	2,2018 (4,842)	1,4736 (1,730)	2,5321 (5,690)	2,5308 (5,496)
$\phi$			0,9181 (6,575)			
$\sigma_u$					1,6061 (3,026)	0,8974 (2,191)
$\rho$			0,1571*	0,0730 (0,299)	0,6106*	0,3287*
<b>Iteraciones</b>	6	9	5	27	6	8
<b>l(0)</b>	-1,36456	-1,37835	-1,36439	-1,37828	-1,35857	-1,35989
<b>Tiempo Conv [min]</b>	5,4	3,8	39,8	226,4	37,4	55,7

(\*) Estimado a partir de  $\phi$  y  $\sigma_u$  respectivamente.



Es posible verificar que los parámetros no poseen variaciones notables entre modelos, salvo algunas diferencias con las constantes modales y ciertos parámetros para los modelos Probit. Sin embargo, sí es posible observar diferencias importantes en el grado de correlación estimado por los distintos modelos. Mientras los modelos LJ y Probit obtienen una correlación muy baja, el MLHe obtiene un valor considerablemente alto. Este valor baja prácticamente a la mitad cuando se impone una estructura homoscedástica. Dos posibles explicaciones se pueden enunciar para este extraño resultado. Por un lado, puede postularse que la base resulta ser heteroscedástica y correlacionada, por ello al imponer homoscedasticidad se subestima la correlación. Por otro lado, puede ser que el número de observaciones que posee la muestra no sea lo suficientemente alto como para poder recoger de manera correcta efectos complejos de la estructura de covarianza.

## 5. CONCLUSIONES

El modelo Mixed Logit ofrece una alternativa interesante y útil de modelación. Permite modelar situaciones prácticas en las que se espera el levantamiento de ciertos supuestos simplificadoros, propios de los modelos clásicos y más utilizados. Su matriz de covarianza depende de la especificación que se le dé a los términos de error adicionales que se consideren, y puede ser tan general como se desee. En ese sentido, ofrece una estructura más flexible que otros modelos. En particular se reconoce su capacidad de tratar con alternativas correlacionadas y variaciones en los gustos expresada a través de parámetros aleatorios. Sin embargo, se debe ser riguroso con la especificación que se esté utilizando, justificando adecuadamente en base a consideraciones teóricas, la estructura de modelo previo a la estimación de los parámetros.

En este trabajo se presentó dos aplicaciones numéricas, una basada en experimentos de simulación y otra con datos reales, ambas en un contexto de alternativas similares. Se muestra empíricamente que el modelo presentado por Brownstone y Train (1999) como un Mixed Logit equivalente a un Logit Jerárquico en realidad no lo es. Sin embargo, para el nivel de correlación reportado, si el ML no se ajusta para obtener una matriz de covarianza homoscedástica, entonces las particiones predichas por ambos no presentan severas diferencias. Además, en términos teóricos, esta especificación siempre compromete la homoscedasticidad al definir correlación, lo que puede ser visto como un problema si se desea comparar con el Logit Jerárquico, o como una ventaja por la ganancia en flexibilidad.

## AGRADECIMIENTOS

Este estudio fue parcialmente financiado por Fondecyt.

## REFERENCIAS

Algers, S., Bergstrom, P., Dahlberg, M. y Dillen, J. (1999) Mixed logit estimation of the value of travel time. Working Paper, Department of Economics, Uppsala University.

Aptech Systems (1994) **GAUSS User's Manual**. Maple Valley, CA.

Ben-Akiva, M.E. y Bolduc, D. (1996) Multinomial probit with a logit kernel and a general parametric specification of the covariance structure. Working Paper, Department d'Economie, Université Laval.

Bhat, C. (2000) Quasi-Random Maximum Simulated Likelihood Estimation of the Mixed Multinomial Logit Model. **9<sup>th</sup> International Association for Travel Behaviour Research Conference (IATBR)**, Queensland, Australia

Boyd, J. y Melman, R. (1980) The effect of fuel economy standards on the US automotive market: an hedonic demand analysis. **Transportation Research**, Vol 14A, 367-378.

Brownstone, D. y Train, K. (1999) Forecasting new product penetration with flexible substitution patterns. **Journal of Econometrics**, Vol 89, 109-129.

Cardell, N. y Dunbar, F. (1980) Measuring the societal impacts of automobile downsizing. **Transportation Research**, Vol 14A, 423-434.

Daganzo, C.F. (1979) **Multinomial Probit: The Theory and its Applications to Travel Demand Forecasting**. Academic Press, Nueva York.

Gunn, H.F. y Bates, J.J. (1982) Statistical aspects of travel demand modelling. **Transportation Research**, Vol 16A, 371-382.

Hajivassiliou, V y Ruud, P. (1994) Classical estimation methods for LDV models using simulation. En R. Engle y D. McFadden (Eds.), **Handbook of Econometrics** Vol IV. Elsevier, Nueva York.

McFadden, D (1974) Conditional logit analysis of qualitative choice behavior. En P. Zarembka (Ed.), **Frontiers in Econometrics**. Academic Press, Nueva York.

McFadden, D. y Train, K. (2001) Mixed MNL models for discrete response. **Journal of Applied Econometrics** (en imprenta).

Munizaga, M.A. y Alvarez, R. (2000) Modelos mixed logit: uso y potencialidades. **Actas del XI Congreso Panamericano de Ingeniería de Tránsito y Transporte**, Noviembre, Gramado, Brasil, 523-535.

Munizaga, M.A. y Ortúzar, J. de D. (1997) On the applicability of the multinomial probit model. **Proceedings of the 25th European Transport Forum**, Vol. P415. PTRC Education and Research Services Ltd., Londres.

Munizaga, M.A. y Ortúzar, J. de D. (1999a) Correlación entre alternativas: el modelo Logit Jerárquico en profundidad. **Actas del IX Congreso Chileno de Ingeniería de Transporte**, 109-120, Sociedad Chilena de Ingeniería de Transporte, Santiago.



Munizaga, M.A. y Ortúzar, J. de D. (1999b) Nested logit modelling: some hard facts. **27th European Transport Conference**, Vol P324. PTRC Education and Research Services Ltd., Londres. 25-36.

Ortúzar, J de D y Donoso, P. C. F.. (1983) Survey design, implentation, data coding and evaluation for the estimation of disaggregate choice models in Santiago, Chile. **2nd International Conference on Survey Methods in Transport**, Septiembre, Hungerford Hill Village, Australia

Ortúzar, J. de D. y Willumsen, L.G. (1994) **Modelling Transport**. Segunda Edición, John Wiley and Sons, Chichester.

Train, K. (1999) Halton sequences for mixed logit. Working paper, Department of Economics, University of California at Berkeley.

Williams, H.C.W.L. (1977) On the formation of travel demand models and economic evaluation measures of user benefit. **Environment and Planning**, Vol 9A, 285-344.

Williams, H.C.W.L. y Ortúzar, J. de D. (1982) Behavioural theories of dispersion and the mis-specification of travel demand models. **Transportation Research**, Vol 16B, 167-219

## APÉNDICE

Sea  $i$  una alternativa perteneciente al nido  $k$ . La utilidad de esta alternativa, utilizando la especificación ML considerada, corresponde a:  $U_{in} = V_{in} + \mu_{kn} + \varepsilon_{in}$ , con  $\varepsilon_{in} \sim \text{Gumbel}(0, \lambda)$  y  $\mu_{kn} \sim f(0, \sigma_\mu^2)$ . Es fácil ver que  $\text{VAR}(U_{in}) = \sigma_\mu^2 + \sigma_\varepsilon^2$  y  $\text{COV}(U_{in}, U_{jn}) = \sigma_\mu^2$  ssi  $j \in k$ . Este tipo de estructura de la matriz de covarianza, implica el siguiente coeficiente de correlación:

$$\rho = \frac{\sigma_\mu^2}{\sigma_\mu^2 + \sigma_\varepsilon^2} \quad (\text{A.1})$$

Si  $\mu_{kn} \sim N(0, \sigma_\mu^2)$ , entonces  $\mu_{kn} = s_{kn}\sigma_\mu$  con  $s_{kn}$  distribuido Normal estándar y el parámetro estimado será tal que  $\hat{\sigma}_\mu = \lambda\sigma_\mu$ . Considerando (A.1) y la relación entre el factor de escala y la varianza del término Gumbel, resulta directo demostrar que se cumple (11).

$$\hat{\sigma}_\mu = \frac{\pi}{\sqrt{6}\sigma_\varepsilon}\sigma_\mu \Leftrightarrow \frac{\sigma_\varepsilon^2}{\sigma_\mu^2} = \frac{\pi^2}{6\hat{\sigma}_\mu^2} \Leftrightarrow \frac{\sigma_\mu^2}{\sigma_\mu^2 + \sigma_\varepsilon^2} = \frac{6\hat{\sigma}_\mu^2}{6\hat{\sigma}_\mu^2 + \pi^2}$$