MODELOS MIXED LOGIT: USO Y POTENCIALIDADES

Marcela A. Munizaga y Ricardo Alvarez Daziano
Departamento de Ingeniería Civil, Universidad de Chile. Casilla 228-3, Santiago, Chile.
mamuniza@cec.uchile.cl, r_a_daziano@yahoo.com
http://tamarugo.cec.uchile.cl/~dicidet/

RESUMEN

El uso de modelos de la "familia" Logit ha sido ampliamente difundido en la modelación de demanda por transporte, comprendiendo los más diversos usos, tales como modelos de partición modal, asignación, e incluso distribución y generación de viajes. Sin embargo, los modelos más utilizados, Logit Multinomial y Jerárquico, poseen supuestos simplificatorios que no siempre son sostenibles, y un modelo que permita mayor flexibilidad parece ser deseable. En este último tiempo, y considerando los avances tecnológicos en materia de computación y métodos numéricos, se ha cuestionado el uso de modelos simplificados, dada la cada vez mayor posibilidad de usar modelos más poderosos y complejos que se adapten a un mayor número de situaciones; inquietudes canalizadas hasta ahora hacia el modelo Probit.

En este artículo se estudia teórica y empíricamente el uso y las potencialidades de una nueva alternativa de modelación: el modelo Mixed Logit. Se discute de qué manera los modelos Mixed Logit son capaces de modelar condiciones en las cuales se violan los supuestos de independencia y homoscedasticidad. A través de un análisis de la matriz de covarianza, se distingue aquellas estructuras válidas de error que el modelo soporta. También se discute aspectos asociados a la estimación del modelo y su factibilidad real de aplicación. Se realiza comparaciones entre especificaciones Mixed Logit y otros modelos tales como el Probit y el Logit Jerárquico.

Se concluye que efectivamente se trata de una herramienta poderosa, que podría llegar a constituirse en un modelo de uso común. No obstante, al igual que en el caso de cualquier modelo flexible, es preciso justificar adecuadamente la implementación de cada caso particular en base a consideraciones teóricas.

PALABRAS CLAVE
Modelación, Elecciones Discretas, Mixed Logit

1. INTRODUCCIÓN

Los modelos llamados 'Mixed Logit'' han irrumpido con mucha fuerza en el ambiente teórico de la modelación de demanda por transporte en los últimos años (Ben Akiva y Bolduc, 1996; Brownstone y Train, 1999). Se trata de una alternativa de modelación que podría situarse entre el modelo 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 su formulación en detalle, con una óptica de imparcialidad, verificando la consecuencia de sus hipótesis.

En el contexto de la modelación de elecciones discretas en demanda de transporte, el enfoque más utilizado en la actualidad es el basado en la teoría de la utilidad aleatoria (McFadden, 1974). Según esta teoría, cada individuo *n* tiene una función de utilidad U_{in} asociada a cada una de las alternativas *i*, escogiendo aquella alternativa que maximiza su utilidad. Esta función, propia 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 e_{in} que intenta recoger todos aquellos efectos no incluidos en la componente sistemática de la función de utilidad; por ejemplo, la incapacidad del modelador para observar todas las variables que influyen en la decisión, errores de medición, diferencias entre individuos, percepciones incorrectas entre atributos y la aleatoriedad inherente a la naturaleza humana. De los supuestos que se asumen sobre la distribución del error estocástico se derivan 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 deriva a partir de asumir que los términos de error e_{in} sonn 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 distribuye logística, y que representa correlación en un grupo de alternativas. En síntesis, se trata de estructuras de covarianza (del término de error) muy simples, lo cual es un supuesto simplificatorio 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 error (matriz de covarianza) que los datos permitan estimar, lo cual implica un grado de dificultad de estimación considerable. Este modelo, que aparece como tan deseable desde ese punto de vista, 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).

Es en este contexto que en los últimos años aparecen los modelos Mixed Logit (también conocidos como modelos de Error Compuesto o Probit con Kernel Logit), como una alternativa intermedia que se sitúa en algún punto entre el Logit y el Probit. La idea central de este tipo de modelos es considerar más de una componente aleatoria; de esta forma, además de una componente Gumbel iid, con lo cual el modelo básico es Logit, se agrega otras componentes que permiten modelar correlación y/o heteroscedasticidad. Esto permite ganar generalidad, pero la estimación deja de ser simple como en el caso del modelo Logit, y al igual que en el caso del Probit, se requiere simulación.

Ya se ha dicho que la distribución del término de error estocástico juega un rol fundamental en los modelos de elección discreta, y que los más utilizados suponen una distribución Gumbel homoscedástica e independiente. Si el punto es incorporar modelos que permitan estructuras de error más generales, es importante analizar qué estructuras sería deseable poder estimar y por qué. Estamos hablando de la posible existencia de correlación y heteroscedasticidad (distinta varianza) en los términos de error. En ambos casos, éstas se pueden dar entre alternativas y entre observaciones. El caso de correlación entre alternativas (presente por ejemplo cuando el usuario percibe algunas alternativas como más similares entre sí que otras) es asimilado bajo ciertas restricciones por el modelo Logit Jerárquico, que permite una matriz de covarianza diagonal por bloques y homoscedástica (ver Munizaga y Ortúzar, 1999 a; b). Sin embargo, muchos casos de correlación y heteroscedasticidad, fácilmente asociables a situaciones prácticas, no se pueden tratar adecuadamente con los modelos tradicionales (Munizaga et al, 1997). Por ello, resulta interesante encontrar un modelo más general que se adapte a situaciones más sofisticadas.

2. EL MODELO MIXED LOGIT

2.1 Formulación

La idea de los modelos Mixed Logit como tal, no es una idea nueva, modelos de estas características han sido propuestos con anterioridad. Por ejemplo se puede citar los trabajos de Cardell y Dunbar (1980) y Boyd y Melman (1980), en que un modelo equivalente a los actuales Mixed Logit es descrito con el nombre de modelo Hedónico. Su reciente reaparición con otro nombre y renovados bríos puede deberse a que los avances tecnológicos en computación y métodos numéricos permiten ahora su estimación en menor tiempo. Recientemente este tipo de modelos ha sido utilizado para modelar diversas situaciones (Train, 1999; Brownstone y Train, 1999) y una aplicación directa al transporte se reporta en Algers *et al.* (1998), donde se estima valores subjetivos del tiempo a partir de un modelo de este tipo.

En términos más específicos que lo ya adelantado, los modelos Mixed Logit nacen de suponer una función de utilidad U_{in} conformada por una componente determinística V_{in}, una componente aleatoria e_{in} independiente e idénticamente distribuida, y uno o más términos aleatorios adicionales. Estos términos de error adicionales pueden ser agrupados en un término aditivo h_{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:

$$U_{in} = V_{in} + h_{in} + e_{in} \tag{1}$$

Se puede ver que se trata de una forma distinta de justificar un determinado modelo. La forma usual es hacer supuestos directamente sobre la distribución del término de error e_{in}, como por ejemplo en el caso del Probit. En cambio, en un modelo Mixed Logit lo que se hace es construir una estructura de error diferente agregando términos que sean fuente de correlación y/o heteroscedasticidad.

Se asume que e es iid Gumbel, mientras que h sigue una función de distribución cualquiera definida por una densidad $f(h/q^*)$, donde q^* son parámetros fijos que la describen (e.g. media y varianza)[1]. Como e es iid Gumbel, la probabilidad condicional en h de que el individuo n escoja la alternativa i corresponde al modelo Logit Multinomial (o Logit Simple):

$$P_n(i/h) = L_{in}(h) = \frac{e^{V_{in}+h_{in}}}{\sum_j e^{V_{jn}+h_{jn}}} \tag{2}$$

Por lo tanto, la probabilidad de elegir la alternativa corresponde a la integral de la probabilidad condicional sobre todos los posibles valores de h, lo que depende de los parámetros de su distribución, esto es:

$$P_{in} = \int L_{in}(h) f(h/q^*) dh \tag{3}$$

Como caso particular, puede suponerse una función de utilidad con la siguiente especificación[2]:

$$U_{in} = \underbrace{b^t x_{in}}_{V_{in}} + \underbrace{m_{in}^t z_{in}}_{h_{in}} + e_{in} \tag{4}$$

En esta expresión se asume que la componente determinística de la utilidad es lineal en los parámetros b que ponderan a los atributos x_{in}. Por otro lado, se asume que h depende de ciertos parámetros (m_{in}) y datos observados relacionados con la alternativa i (z_{in}), relación que también se supone lineal en parámetros. Un supuesto adicional es que el término m es propio del individuo, sin variar entre alternativas. Es decir:

$$h_{in} = m_n^t z_{in} \tag{5}$$

Esta especificación es la que ha sido utilizada en la mayor parte de los estudios previos (Ben Akiva y Bolduc, 1996; Brownstone y Train, 1999).

---

1 En términos prácticos, 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 log-normal, especialmente en aquellos casos en que se quiere restringir el signo de un determinado parámetro.

2 b 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; m_{in} es un vector aleatorio de dimensión K cuyas componentes tienen media cero y con matriz de covarianza Ω; z_{in} es un vector de atributos asociados con la alternativa i y el individuo n, y tiene dimensión K; finalmente, e_{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 (5), 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 e_n un vector aleatorio iid Gumbel con matriz de covarianza Σ_e y que contiene a los elementos e_{in}. Si se asume que cada término de m_n tiene una función densidad con media cero y varianza s_k^2 y que el vector en su conjunto tiene una matriz de covarianza Ω, entonces la matriz de covarianza del modelo (Σ), puede escribirse como:

$$\Sigma = z_n^t \cdot \Omega \cdot z_n + \Sigma_e = z_n^t \cdot \Omega \cdot z_n + s_e^2 I \tag{6}$$

Es claro que la dimensión de la matriz de covarianza obtenida está bien definida[3] y de esta expresión general se concluye que el modelo es capaz de modelar correlación y heteroscedasticidad entre alternativas. En efecto, si obtenemos la covarianza entre dos alternativas se observa que para $i, j \in C_n$ con $i \neq j$:

$$\mathrm{cov}(U_{in}, U_{jn}) = \sum_{k=1}^{K} z_{kin} z_{kjn} s_k^2 \tag{7}$$

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

En cuanto a la varianza,

$$\mathrm{var}(U_{in}) = \sum_{k=1}^{K} z_{kin}^2 s_k^2 + \frac{p}{6 l^2} \tag{8}$$

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

2.3 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, 1997). Es así como, un modelo Mixed Logit con parámetros aleatorios distribuidos normal, puede aproximar a un

[3] La matriz de covarianza es de dimensión $J \times J$. En efecto, como Ω es de dimensión $K \times K$ (con K el número de componentes aleatorias), y z_n tiene dimensión $K \times J$, entonces $z_n^t \cdot \Omega \cdot z_n$ es una matriz de dimensión $J \times J$; luego al sumarse esta última con la matriz Σ_e, que es de dimensión $J \times J$, se obtiene finalmente que $\dim \Sigma = \dim\left(z_n^t \cdot \Omega \cdot z_n + \Sigma_e\right) = J \times J$.

modelo Probit.

Además el modelo Mixed Logit, al permitir modelar la presencia de correlación entre alternativas, es capaz de levantar el supuesto de independencia de alternativas irrelevantes propio del modelo Logit Multinomial. En otras palabras, los patrones de sustitución entre alternativas son flexibles. En efecto, dada una probabilidad tipo Mixed Logit (9), se puede demostrar que la razón entre probabilidades de dos alternativas depende de todo el conjunto de alternativas disponibles.

$$P_{in} = \int_{-\infty}^{\infty}\cdots\int_{-\infty}^{\infty}\left\{\frac{\exp(b^{t}x_{in} + m_n^{t}z_{in})}{\sum_{l=1}^{J}\exp(b^{t}x_{1\mathrm{n}} + m_n^{t}z_{1\mathrm{n}})}\right\} f(m_{1n})\cdot\ldots\cdot f(m_{Kn})\,dm_{1n}\ldots dm_{Kn} \tag{9}$$

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 utilizar simulación. Sin embargo, se puede aprovechar el hecho de que la probabilidad condicional (2) tiene una expresión de forma logit multinomial.

Entonces, si se considera R valores de h obtenidos de su función densidad $f(h/q^*)$, para cada una de las repeticiones es posible calcular

$$P_n(i/h^r) = L_{in}(h^r) = \frac{e^{V_{in}+h^{r}_{in}}}{\sum_j e^{V_{jn}+h^{r}_{jn}}}, \tag{10}$$

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

$$\tilde{P}(i) = \frac{1}{R}\sum_{r=1}^{R} L_{in}(h^r), \tag{11}$$

y con ella construir la función de verosimilitud simulada

$$SL = \sum_q \sum_{i=1}^{J} y_{qi} \ln \tilde{P}(i) \tag{12}$$

Bajo condiciones de regularidad, el estimador máximo verosímil simulado es consistente y asintóticamente normal. A pesar de que (12) 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).

2.5 Mixed Logit y Logit Jerárquico

Un tema que ha sido fuente de confusión es que la especificación Mixed Logit podría hacerse equivalente a un modelo Logit Jerárquico. Este último modelo fue concebido para tratar correlación entre alternativas, agrupando alternativas *similares* en *nidos* dentro de los cuales se cumple el supuesto de errores iid (Williams, 1977). La jerarquización en nidos implica una determinada estructura de la matriz de covarianza, pues si dos o más alternativas están agrupadas en un nido, los elementos no diagonales correspondientes serán distintos de cero.

Brownstone y Train (1999) presentan un modelo Mixed Logit que ellos llaman "análogo" a un Logit Jerárquico. Este modelo particular se construye agrupando las alternativas en nidos; luego, en la función de utilidad se agrega una variable muda para cada nido que indica si la alternativa pertenece o no a éste. A cada una de estas variables mudas se le asocia un parámetro aleatorio común entre las alternativas. De esta manera se logra un modelo con una estructura de correlación tal que en aquellas alternativas que pertenecen a un mismo nido aparece un término fuera de la diagonal en la matriz de covarianza. Los autores concluyen que de esta forma se logra un patrón de correlación igual al de un Logit Jerárquico. Sin embargo, lo correcto es comparar la matriz de covarianza de ambos modelos.

A modo de ejemplo, supongamos un caso en el cual existen tres alternativas disponibles para un determinado individuo. Estas alternativas son auto, bus y metro. Supongamos además que las alternativas bus y metro se encuentran correlacionadas, por ser percibidas como más similares entre sí que el auto. Este caso, que corresponde a un Logit Jerárquico con un nido de transporte público podría modelarse como un Mixed Logit con la siguiente especificación de acuerdo a Brownstone y Train (1999):

$$
\begin{aligned}
U_{auto} &= C_{auto} + b^{t}_{auto} x_{auto} + e_{auto} \\
U_{bus} &= b^{t}_{bus} x_{bus} + m + e_{bus} \\
U_{metro} &= C_{metro} + b^{t}_{metro} x_{metro} + m + e_{metro}
\end{aligned}
\tag{13}
$$

donde m es un término aleatorio con media cero y varianza s^{2}_{m}, y $e_{.}$ es un término iid Gumbel con varianza s^{2}_{e}. Es fácil ver que la matriz de covarianza de este modelo es:

$$
\Sigma = \begin{bmatrix} s^{2}_{e} & 0 & 0 \\ 0 & s^{2}_{m} + s^{2}_{e} & s^{2}_{m} \\ 0 & s^{2}_{m} & s^{2}_{m} + s^{2}_{e} \end{bmatrix}
\tag{14}
$$

Esta matriz posee términos fuera de la diagonal indicando correlación entre las alternativas bus y metro, sin embargo, es heteroscedástica. Por lo tanto este modelo no es en realidad equivalente al Logit Jerárquico en términos de estructura de errores, ya que este último es homoscedástico por definición.

La correlación entre las alternativas bus y metro está dada por:

$$corr(U_{bus}, U_{metro}) = r_{bus,metro} = \frac{s_m^2}{s_m^2 + s_e^2} = \frac{1}{1 + \frac{s_e^2}{s_m^2}} \qquad (15)$$

Luego,

- Si $s_m^2 << s_e^2$, entonces $r_{bus,metro} \rightarrow 0$
- Si $s_m^2 = s_e^2$, entonces $r_{bus,metro} = 0{,}5$
- Si $s_m^2 >> s_e^2$, entonces $r_{bus,metro} \rightarrow 1$

De los casos presentados, es claro que mientras mayor sea la desviación de m en relación a la del error iid Gumbel, mayor será la correlación obtenida. Este es el resultado esperado, pues justamente m es el término común que impone la presencia de correlación entre las alternativas bus y metro.

2.6 Mixed Logit y Probit

Tal como se mencionara con anterioridad, el modelo Mixed Logit se construye a partir de asumir términos de error adicionales que pueden implicar una matriz de covarianza heteroscedástica y con correlación. Contrariamente, en el caso del Probit se asume un solo término de error aleatorio con matriz de covarianza general. En efecto, un modelo Probit multinomial se deriva del hecho de suponer que dada una función de utilidad $U_{in} = V_{in} + e_{in}$, el vector $e_n = (e_{1n}, \ldots, e_{Jn})^t$ distribuye Normal multivariada con matriz de covarianza Σ general.

El modelo Probit tampoco posee una expresión cerrada de la probabilidad de elección, por lo que se vuelve necesario utilizar algún tipo de aproximación o simulación. El método de simulación más utilizado es el de máxima verosimilitud simulada a partir del estimador de Geweke-Hajivassilou-Keane (Börsch-Supan y Hajivassilou, 1993), el cual reduce recursivamente la dimensión de la integral hasta llegar a un problema equivalente en el cual se requiere de repeticiones de una normal unidimensional truncada. Las probabilidades simuladas de esta forma resultan insesgadas, continuas y diferenciables. Las simulaciones para los modelos Probit y Mixed Logit tienen distinta dimensión: *J-1* para el caso del Probit[4] y *K* para el caso del Mixed Logit. De esta forma, si *K<J-1* se obtiene una ventaja por sobre el Probit pues se está ganando por el lado de trabajar con una dimensión menor. Esto se da cuando el número de parámetros aleatorios incorporados al modelo Mixed Logit es menor que el número de alternativas.

---

[4] Debido a que se basan en las diferencias $e_j - e_i$, con *i* la alternativa escogida y *j* cada una de las *J-1* alternativas restantes.

3. RESULTADOS EMPÍRICOS PRELIMINARES

Como una forma de validar el análisis teórico previo, se realizó un experimento de simulación siguiendo la metodología de Williams y Ortúzar (1982). Los valores medios y desviaciones de los atributos considerados, fueron obtenidos de una base real construida para la ciudad de Santiago de Chile. Para la estimación de los modelos Logit Multinomial y Logit Jerárquico 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[5].

El caso considerado corresponde a una elección modal con tres alternativas (por ejemplo auto, bus y metro) y dos variables explicativas (costo y tiempo de viaje). El proceso de elección discreta fue simulado a partir de valores dados para las variables independientes y los parámetros de gusto.

Se trabajó con un error Gumbel iid y, adicionalmente, con un término distribuido Normal presente en las alternativas bus y metro. Con este tipo de especificación, análoga a la presentada en (13), se pretende modelar el caso en que las alternativas bus y metro son consideradas como similares entre sí. Es decir, el modelo adecuado para modelar la elección simulada correspondería a un Mixed Logit. Interesa estudiar cómo se adaptan los modelos Logit Multinomial y Jerárquico a esta elección.

Como resultado empírico preliminar se presenta un caso con 4000 observaciones, en el cual se consideró $s_m = s_e$, lo cual implica un coeficiente de correlación de 0,5 entre las alternativas bus y metro, y $s_m = 3 \cdot s_e$, lo que implica un coeficiente de correlación de 0,75.

La estimación del Mixed Logit se realizó 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. El tiempo total de estimación del ML es de alrededor de 15 minutos; mientras que para los modelos MNL y LJ éste resulta ser menor a un minuto.

Los resultados de las estimaciones del modelo Logit Multinomial (MNL), del Mixed Logit (ML) y del Logit Jerárquico (LJ) se presentan en la Tabla 1, en que también se entregan los valores 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 modelo ML. Para el LJ, el valor de referencia de f es calculado a partir de la correlación simulada.

---

[5] http://elsa.berkeley.edu/~train/software.html

Tabla 1: 4000 Observaciones Correlacionadas.

	$r_{bus,metro} = 0{,}5$				$r_{bus,metro} = 0{,}75$			
	Referencia	MNL	LJ	ML	Referencia	MNL	LJ	ML
Cte Auto	-0,4	-0,1907 (-2,464)	-0,3237 (-3,728)	-0,4460 (-2,553) [-0,2634]	-0,4	0,2012 (2,624)	-0,1732 (-2,203)	-0,4084 (-1,578) [-0,0324]
Cte Metro	-0,2	-0,2059 (-3,195)	-0,2204 (-4,282)	-0,2939 (-3,675) [-0,6407]	-0,2	0,0543 (0,810)	-0,0622 (-1,847)	-0,1454 (-1,732) [0,6507]
Costo	-0,005	-0,0051 (-6,192)	-0,005 (-6,180)	-0,0067 (-4,434) [-0,2634]	-0,005	-0,0014 (-1,806)	-0,0013 (-1,784)	-0,0031 (-1,634) [0,9964]
Tiempo	-0,08	-0,0767 (-15,163)	-0,0628 (-7,347)	-0,0841 (-13,145) [-1,1743]	-0,08	-0,0586 (-11,536)	-0,0328 (-4,151)	-0,0767 (-11,266) [0,4778]
f	0,7071		0,7453 (6,393)	-	0,5		0,4281 (3,889)	-
s_m	1,2825498		-	1,3975 (2,6381) [0,2170]	3,8476495			3,3202 (3,132) [-0,4975]
Iteraciones		4	4	4		6	5	4
L-verosim.		-1,02960	-1,02910	-1,02919		-1,03894	-1,03673	-1,03678

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. Los modelos MNL y LJ arrojan estimadores de estos parámetros que son distintos de los valores verdaderos, especialmente en el caso de mayor correlación. Cabe hacer notar que se observa una relación entre los parámetros estimados por el modelo LJ y los del ML. Al dividir ambos parámetros en cada base de datos, se encuentra un valor relativamente constante, que es mayor en el caso de mayor correlación. Esto se explica porque la presencia de heteroscedasticidad afecta al factor de escala que multiplica los parámetros. En el caso del modelo ML la componente común del error (μ) es fijada en un cierto valor en cada repetición de la simulación, por lo tanto, el factor de escala característico de la distribución Gumbel es el correspondiente al término aleatorio ε solamente $(l = \frac{p}{\sqrt{6}s_e})$. En cambio, en el caso del modelo LJ, aún abstrayéndose de la heteroscedasticidad, es la suma de ambas componentes de error la que se supone distribuida Gumbel, y en ese caso el factor de escala es menor; si todas las alternativas tuvieran la misma varianza del término de error, el factor de escala del LJ correspondería a $l = \frac{p}{\sqrt{6(s_e^2 + s_m^2)}}$.

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.

Tabla 2: Cambios de Política (porcentaje de cambio en tiempo y costo).

	c_{auto}	c_{bus}	c_{metro}	t_{auto}	t_{bus}	t_{metro}
P1		-50%				
P2				+100%		
P3					+100%	
P4					-50%	
P5		-40%				
P6		-50%	+100%		-50%	+100%

En la Tabla 3 se reporta el índice Chi cuadrado de Gunn y Bates (1982) como medida de error para cada una de las políticas definidas; este se calcula como $c^2 = \sum_i \frac{(\hat{N}_i - N_i)^2}{N_i}$, 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 3: Índice Chi-cuadrado

	$r_{bus,metro} = 0{,}5$			$r_{bus,metro} = 0{,}75$		
	MNL	ML	LJ	MNL	ML	LJ
P1	9,151	10,131	9,814	1,250	2,339	2,441
P2	22,885	5,369	3,692	82,383	7,009	6,450
P3	1,287	1,949	1,797	33,536	2,542	2,473
P4	8,562	7,402	7,255	45,094	5,764	5,476
P5	6,403	7,455	7,248	1,008	1,375	1,331
P6	70,712	27,747	28,167	119,496	5,653	5,948

Las mayores diferencias entre las predicciones de los modelos y los valores simulados (realidad virtual) se encuentran para el MNL en el caso de mayor correlación. Las predicciones del LJ y ML son muy parecidas, siendo ambas a su vez muy parecidas a lo que se ha llamado “realidad virtual”. Llama la atención el hecho de que las predicciones del ML y las del LJ sean muy parecidas, a pesar de las diferencias de formulación de los modelos que ya hemos señalado (homo-heteroscedasticidad). De hecho, se puede decir que al menos en este caso un grado una heteroscedasticidad como la descrita aparece como un problema menor para el modelo LJ.

4. CONCLUSIONES

El modelo Mixed Logit aparece como una alternativa de modelación razonable en términos prácticos, al permitir su estimación y uso con un computador personal en un tiempo moderado, permitiendo modelar correlación y heteroscedasticidad. Su estructura está determinada por la forma del término que se agrega a la función de utilidad y puede llegar a ser tan general como se desee. En este contexto, puede constituirse en un "competidor" del modelo Probit, que hasta hace poco era considerado como el único o principal camino de flexibilizar la modelación de elecciones discretas considerando correlación y heteroscedasticidad, y también un competidor para los modelos Logit Multinomial y Logit Jerárquico, que son los más usados hasta ahora.

La diferencia central entre el Probit (que es su principal competidor) y el Mixed Logit está en la forma en que en cada uno se especifica su estructura. En el caso del Probit, se hace directamente en la matriz de covarianza, y en el caso del Mixed Logit se hace directamente en la función de utilidad, y la matriz de covarianza aparece como un resultado. En términos de estimación el modelo Mixed Logit parece requerir un tiempo de procesamiento un poco menor que el Probit.

Se ha mostrado aquí que el modelo presentado por Brownstone y Train (1999) como un caso particular de Mixed Logit que es equivalente al Logit Jerárquico, salvo por la diferencia en las distribuciones en ambos casos, en realidad no lo es en términos teóricos, ya que el modelo propuesto es heteroscedástico. Sin embargo, evidencia empírica reducida y preliminar parece indicar que efectivamente las predicciones son similares al aplicar ambos modelos a una misma base de datos.

Desde nuestro punto de vista, el modelo ML es una alternativa de modelación interesante y útil, y lo importante sigue siendo tener claras las propiedades del modelo que se está utilizando y justificar adecuadamente, en base a consideraciones teóricas, una cierta estructura de modelo previo a la estimación de los parámetros.

Agradecimientos
Este estudio fue parcialmente financiado por Fondecyt.

Referencias bibliográficas

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'Economique, Université Laval.

Börsch-Supan, A. y Hajivassiliou, V.A. (1993) Smooth unbiased multivariate probability simulators for maximum likelihood estimation of limited dependent variable models. *Journal of Econometrics 58*, 347-368.

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

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

Cardell, N. y Dunbar, F. (1980) Measuring the societal impacts of automobile downsizing. *Transportation Research 14 A*, 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 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. (1997) Mixed MNL models for discrete response. Working Paper, Department of Economics, University of California at Berkeley.

Munizaga, M.A., Heydecker, B.G. y Ortúzar, J. de D (1997) On the error structure of discrete choice models. *Traffic Engineering and Control 38*, 593-597.

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 Willumsen, L.G. (1994) *Modelling Transport*. Segunda Edición, John Wiley and Sons, Chichester.

Revelt, D y Train, K.E. (1998) Mixed logit with repeated choices: household's choices of appliance efficiency level. *Review of Economics and Statistics 80*, 647-657.

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 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 16B*, 167-219.