

CORRELACION ENTRE ALTERNATIVAS: EL MODELO LOGIT JERARQUICO EN PROFUNDIDAD

Marcela Munizaga
Departamento de Ingeniería Civil
Universidad de Chile
Casilla 228-3, Santiago, Chile.
Fax 6718788; e-mail: mamuniza@cec.uchile.cl

Juan de Dios Ortúzar
Departamento de Ingeniería de Transporte
Pontificia Universidad Católica de Chile
Casilla 306 Código 105, Santiago 22, Chile.
Fax: 5530281; e-mail: jos@ing.puc.cl

RESUMEN

En los últimos 20 años los modelos Logit Simple y Logit Jerárquico han sido los más usados en la modelación de elecciones discretas en el contexto de predicciones de demanda por transporte. Sin embargo, la posibilidad (real en la actualidad) de utilizar modelos más poderosos y complejos, como el Probit General, ha motivado un resurgimiento de la discusión sobre cuál es el modelo más adecuado en cada caso. En este contexto, además, se han discutido acaloradamente algunos aspectos del modelo Logit Jerárquico que permanecían relativamente oscuros. En este trabajo presentamos una derivación clara de este modelo, basándonos en la hipótesis de correlación entre las utilidades de algunas alternativas. Además demostramos, en términos generales, que el modelo también tiene la propiedad de reproducir las particiones de mercado observadas en una muestra de calibración. Finalmente, discutimos las fuentes de correlación, cómo se especifican en términos de la matriz de covarianza de las utilidades, y cómo se relacionan con los distintos modelos disponibles para resolver este problema. En particular demostramos que cuando existe correlación entre alternativas, modelos con función de utilidad no-lineal, como el Logit Box-Cox, no son apropiados.

1. INTRODUCCION

Los modelos de elección discreta utilizados en modelación de demanda por transporte se basan en la teoría de la utilidad aleatoria (McFadden, 1974). En ésta, la utilidad individual se modela asociada a ciertas alternativas separables suponiendo constante el consumo de otros bienes. La utilidad U_{in} que el individuo n asocia al bien discreto i , se representa por la suma de una componente observable V_{in} y un error estocástico ε_{in} . Este último recoge el efecto de variables no observadas, diferencias individuales, errores de medición y todo lo que el modelador no es capaz de incorporar a la modelación. De los supuestos asumidos sobre la distribución de los errores ε_{in} se derivan los distintos modelos planteados en la literatura (Ortúzar y Willumsen, 1994).

El modelo más usado hasta ahora es el Logit Simple (LS), pese a que el supuesto de errores independientes e idénticamente distribuidos (iid) impone que los términos de error no estén correlacionados y tengan la misma varianza (McFadden, 1974). Esto se traduce en la propiedad de Independencia de Alternativas Irrelevantes (IAI), que ha sido criticada en muchos casos, como por ejemplo la conocida paradoja del Bus Azul - Bus Rojo (Mayberry, 1973).

El modelo Logit Jerárquico (LJ), que es una generalización del anterior, permite modelar elecciones discretas en algunos casos en que hay correlación entre los términos de error de las alternativas; así, no requiere el supuesto de errores iid, pero impone algunas restricciones en las elasticidades cruzadas entre alternativas. Este modelo, inicialmente aplicado sin tener una clara derivación, como se puede ver por ejemplo en el estudio británico SELNEC (Wilson *et al.*, 1969), fue desarrollado en propiedad por Williams (1977), quien proveyó una derivación de su forma funcional consistente con la teoría de la utilidad aleatoria, una interpretación de todos los parámetros del modelo dentro del marco teórico, restricciones a los parámetros estructurales del modelo que deben cumplirse para que éste sea consistente, y la derivación de medidas de beneficio económico consistentes con la teoría de la utilidad aleatoria.

2. EL MODELO LOGIT JERARQUICO

La génesis del modelo LJ es el reconocimiento del hecho que algunas alternativas pueden ser percibidas como más similares entre sí que otras. Las alternativas "similares" son agrupadas en "nidos" dentro de los cuales se cumple el supuesto iid, y no pudiendo una alternativa pertenecer a más de un nido a la vez. 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. En la Figura 1 se ilustra un caso en que las primeras dos alternativas están correlacionadas, por lo que se representan en un nido al interior del árbol, y se entrega la matriz de covarianza asociada.

Para formalizar el modelo, utilizaremos la nomenclatura de Ortúzar y Willumsen (1994), en combinación con la de Daly (1987). En el caso de dos niveles jerárquicos, en que el subíndice i representa una alternativa o nido del nivel superior y j una alternativa del nivel inferior (nido), la función de utilidad se puede representar como:

$$U(i,j) = U_i + U_{j|i} \quad (1)$$

$$U(i,j) = V(i,j) + \varepsilon(i,j) \quad (2)$$

$$V(i,j) = V_i + V_{j/i} \quad (3)$$

$$\varepsilon(i,j) = \varepsilon_i + \varepsilon_{j/i} \quad (4)$$

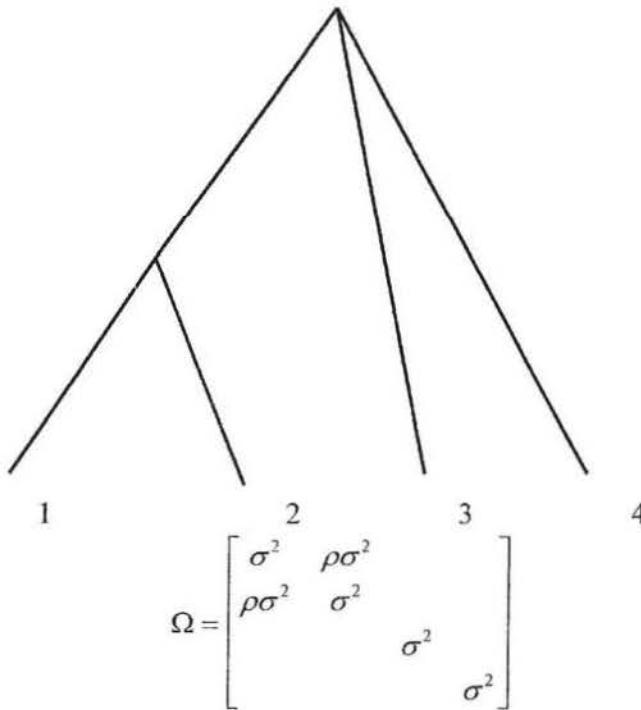


Figura 1: Estructura de nidos y matriz de covarianza del modelo LJ

Williams (1977) demostró que si se asume que los $\varepsilon_{j/i}$ distribuyen Gumbel con varianza σ_{inf}^2 y ε_i distribuye logística con varianza σ_{nido}^2 , entonces $\varepsilon(i,j)$ distribuye Gumbel con varianza igual a $(\sigma_{inf}^2 + \sigma_{nido}^2)$ y la probabilidad de elegir el nido i , y dentro de éste la alternativa j , está dada por:

$$P_{ji} = P_i \cdot P_{j/i} \quad (5)$$

con $P_{j/i} = \frac{\exp(\lambda_i V_{j/i})}{\sum_{k \in C(i)} \exp(\lambda_i V_{k/i})}$ y $P_i = \frac{\exp(\beta V_i)}{\sum_{l \in R} \exp(\beta V_l)}$

en que $V_{j/i}$ es la utilidad representativa de la alternativa j al interior del nido i , en que se considera sólo aquellos atributos que presentan variación al interior del nido. β y λ_i corresponden a los factores de escala del nivel superior y del nido i respectivamente. La utilidad de las alternativas o nidos del nivel superior está dada por la utilidad máxima esperada del nido:

$$\beta V_i = \beta X_i + \frac{\beta}{\lambda_i} \cdot \ln \sum_{k \in C(i)} \exp(\lambda_i V_{k/i}) \quad (6)$$

en que X_i representa la componente de la utilidad asociada a los atributos comunes a todas las opciones del nido. Es fácil demostrar que la separación entre la componente común de la utilidad y el resto no es en realidad necesaria, y la ecuación (6) se puede escribir como:

$$\beta V_i = \frac{\beta}{\lambda_i} \cdot \ln \sum_{k \in C(i)} \exp(\lambda_i V_k) \quad (7)$$

El definir el modelo de esta forma ayuda a evitar confusiones al momento de especificar las constantes modales. En este caso, al igual que en el del LS, se debe incluir $K-1$ constantes específicas, siendo K el número de alternativas elementales. Cada alternativa tendrá asociada una constante específica, con la sola excepción de la alternativa usada como referencia (que asume una constante igual a cero) y los nidos que no deben tener asociadas constantes específicas. Es posible hacer modificaciones a esta especificación siempre y cuando el modelo sea equivalente. Esto aclara la confusión introducida por Bierlaire *et al* (1997) en un artículo cuyo mero título induce a pensar que puede haber problemas de sobre-especificación al introducir constantes al LJ.

El parámetro estructural del nido i , ϕ_i definido como β/λ_i , debe ser mayor que cero y menor que uno para que el modelo sea consistente. En el caso límite en que $\phi_i=1$, el modelo colapsa al LS. El modelo LJ no está restringido en cuanto al número de nidos en serie o en paralelo, pero si hay más de un nido, las condiciones de consistencia de los parámetros estructurales ϕ aplican a cada rama del árbol jerárquico en forma independiente (Ortúzar y Willumsen, 1994).

Para que el modelo sea estimable, es preciso fijar uno de los factores de escala. Si se fija el valor de β igual a uno, entonces es preciso multiplicar las utilidades al interior del nido por los valores de λ_i correspondientes, que pasar a ser iguales a $1/\phi_i$. Bates (1998) nota que si el escalamiento se hace en el nido, es decir suponiendo λ_i igual a uno, se obtienen estimadores de los parámetros que son equivalentes, pero los estadígrafos t resultan diferentes. Por esto, proponemos normalizar de la única forma en que el modelo es directamente comparable con un LS; es decir, escalando en el nivel superior haciendo β igual a uno.

A continuación demostramos que el modelo LJ cumple con la conocida propiedad del LS de reproducir las particiones de mercado observadas cuando está correctamente especificado en términos de constantes modales. Sea q un individuo cualquiera de la muestra; entonces, la probabilidad de que ella elija la alternativa j en el nivel superior, está dada por:

$$P_{jq} = \frac{\exp(\beta V_{jq})}{\sum_{l \in R} \exp(\beta V_{lj})} \quad (8)$$

Si j es una alternativa elemental (que cuelga de la raíz), entonces se cumple que:

$$V_{jq} = \sum_k \theta_k X_{kj} \quad (9)$$

En caso contrario, si j es un nido, se tiene:

$$V_{jq} = \frac{1}{\lambda_j} \cdot \ln \sum_{i \in j} \exp(\lambda_j V_{iq}) = \frac{1}{\lambda_j} \cdot \ln \sum_{i \in j} \exp \sum_k \lambda_j \theta_k X_{kij} \quad (10)$$

Introduciendo el supuesto que $\beta=1$, la probabilidad de elegir la alternativa i dentro del nido j se puede escribir como:

$$P_{ilj} = \frac{\exp(\sum_k \frac{1}{\phi_j} \theta_k X_{kij})}{\sum_{l \in j} \exp(\sum_k \frac{1}{\phi_j} \theta_k X_{klj})} \quad (11)$$

$$y \quad V_{jq} = \phi_j \ln \sum_{i \in j} \exp(\frac{1}{\phi_j} \sum_k \theta_k X_{kij}) \quad (12)$$

En ese caso, la log-verosimilitud estará dada por:

$$l(\theta) = \sum_{q=1}^Q \sum_{i \in C_q} g_{iq} \ln P_{iq} \quad (13)$$

$$P_{iq} = P_{ilj}^q P_j^q = \frac{\exp(\sum_k \frac{1}{\phi_j} \theta_k X_{kij})}{\sum_{l \in j} \exp(\sum_k \frac{1}{\phi_j} \theta_k X_{klj})} \cdot \frac{\exp \phi_j \ln \sum_{l \in j} \exp(\sum_k \frac{1}{\phi_j} \theta_k X_{klj})}{\sum_{s \in S} \exp \phi_s \ln \sum_{l \in s} \exp(\sum_k \frac{1}{\phi_s} \theta_k X_{klj})} \quad (14)$$

La ecuación (13) se puede re-escribir separando ambas componentes de la probabilidad, quedando:

$$l(\theta) = \sum_{q=1}^Q \sum_{i \in C_q} g_{iq} (\ln P_j^q + \ln P_{ilj}^q) \quad (15)$$

Las condiciones de primer orden que se derivan a partir de derivar con respecto a los parámetros de los atributos son:

$$\frac{\partial l(\theta)}{\partial \theta_h} = \sum_{q=1}^Q \sum_{i \in C_q} g_{iq} \left(\frac{1}{P_j^q} \cdot \frac{\partial P_j^q}{\partial \theta_h} + \frac{1}{P_{ilj}^q} \cdot \frac{\partial P_{ilj}^q}{\partial \theta_h} \right) = 0 \quad (16)$$

Se puede demostrar que la derivada de P_j^q con respecto al atributo h es:

$$\frac{\partial P_j^q}{\partial \theta_h} = P_j^q \left(\sum_{i \in j} P_{ilj}^q X_{hiq} - \sum_{s \in S} P_s^q \sum_{l \in s} P_{llj}^q X_{hlq} \right) \quad (17)$$

La derivada de P_{ilj}^q con respecto al atributo h es:

$$\frac{\partial P_{ilj}^q}{\partial \theta_h} = \frac{P_{ilj}^q}{\phi_j} \left(X_{hiq} - \sum_{l \in j} P_{llj}^q X_{hlq} \right) \quad (18)$$

Para el caso en que θ_h es una constante específica, asociada a la alternativa elemental no anidada e , X_{hiq} toma el valor uno cuando i es igual a e y cero en caso contrario. Además, las alternativas no anidadas pueden definirse como aquellas en que el parámetro ϕ correspondiente a ese nido es igual a uno, y la probabilidad condicional al interior del nido también es igual a uno.

Entonces, la derivada de la log-verosimilitud con respecto a θ_h será:

$$\frac{\partial l(\theta)}{\partial \theta_h} = \sum_{q=1}^Q \sum_{i \in C_q} g_{iq} \left\{ \frac{1}{P_j^q} P_j^q \left(\sum_{l \in J} P_{l/j}^q X_{hlq} - \sum_{s \in S} P_s^q \sum_{l \in s} P_{l/s}^q X_{hlq} \right) + \frac{1}{P_{il/j}^q} \frac{P_{il/j}^q}{\phi_j} \left(X_{hlq} - \sum_{l \in J} P_{l/j}^q X_{hlq} \right) \right\} \quad (19)$$

$$= \sum_{q=1}^Q \left\{ \sum_{i \in C_q} g_{iq} \sum_{l \in J} P_{l/j}^q X_{hlq} - \sum_{i \in C_q} g_{iq} \sum_{s \in S} P_s^q \sum_{l \in s} P_{l/s}^q X_{hlq} + \sum_{i \in C_q} g_{iq} X_{hlq} - \sum_{i \in C_q} g_{iq} \sum_{l \in J} P_{l/j}^q X_{hlq} \right\} = 0 \quad (20)$$

Se puede ver fácilmente que evaluando cada uno de estos términos queda:

$$\Leftrightarrow \sum_{q=1}^Q 1 - P_e + g_{eq} - 1 = 0 \quad (21)$$

$$\Leftrightarrow \sum_{q=1}^Q P_e = \sum_{q=1}^Q g_{eq} \quad (22)$$

Queda entonces demostrado que el modelo Logit Jerárquico reproducirá las particiones de mercado observadas en la muestra de calibración, para las alternativas no anidadas.

Planteando las condiciones de primer orden para el caso de la constante específica de una alternativa anidada (a), perteneciente al nido n , definida como $X_{hiq}=1$ si $i=a$ y cero en otro caso, implica:

$$\sum_{q=1}^Q \left\{ \sum_{i \in C_q} g_{iq} \left(\sum_{l \in J} P_{l/j}^q X_{hlq} - \sum_{s \in S} P_s^q \sum_{l \in s} P_{l/s}^q X_{hlq} \right) + \sum_{i \in C_q} g_{iq} \frac{1}{\phi_j} \left(X_{hlq} - \sum_{l \in J} P_{l/j}^q X_{hlq} \right) \right\} = 0 \quad (23)$$

$$\Leftrightarrow \sum_{q=1}^Q \left\{ \sum_{i \in n} g_{iq} P_{a/n}^q - \sum_{i \in C_q} g_{iq} P_n^q P_{a/n}^q + \frac{1}{\phi_n} \left(g_{aq} X_{haq} - \sum_{i \in n} g_{iq} P_{a/n}^q \right) \right\} = 0 \quad (24)$$

$$\Leftrightarrow \sum_{q=1}^Q \left\{ \sum_{i \in n} g_{iq} P_{a/n}^q - P_n^q P_{a/n}^q + \frac{1}{\phi_n} \left(g_{aq} X_{haq} - \sum_{i \in n} g_{iq} P_{a/n}^q \right) \right\} = 0 \quad (25)$$

$$\Leftrightarrow \sum_{q=1}^Q P_n^q P_{a/n}^q = \sum_{q=1}^Q \left\{ \sum_{i \in n} g_{iq} P_{a/n}^q + \frac{1}{\phi_n} \left(g_{aq} X_{haq} - \sum_{i \in n} g_{iq} P_{a/n}^q \right) \right\} \quad (26)$$

$$\Leftrightarrow \sum_{q=1}^Q \left(P_n^q P_{a/n}^q - g_{aq} \right) = \sum_{q=1}^Q \left(\sum_{i \in n} g_{iq} P_{a/n}^q - g_{aq} + \frac{1}{\phi_n} \left(g_{aq} - \sum_{i \in n} g_{iq} P_{a/n}^q \right) \right) \quad (27)$$

$$\Leftrightarrow \sum_{q=1}^Q \left(P_n^q P_{a/n}^q - g_{aq} \right) = \sum_{q=1}^Q \left(\frac{1}{\phi_n} - 1 \right) \left(g_{aq} - \sum_{i \in n} g_{iq} P_{a/n}^q \right) \quad (28)$$

$$\Leftrightarrow \sum_{q=1}^Q (P_n^q P_{a/n}^q - g_{aq}) = \frac{\phi_n - 1}{\phi_n} \sum_{q=1}^Q \left(\sum_{i \in n} g_{iq} P_{a/n}^q - g_{aq} \right) \quad (29)$$

El término de la mano derecha debe anularse para que se cumpla que el modelo se comporta como un MNL al interior del nido. Esto es claro si se considera la sumatoria sólo sobre aquellas observaciones en que la elección recae sobre alguna de las alternativas del nido (el resto de los elementos son evidentemente nulos). En consecuencia, debe cumplirse además que:

$$\sum_{q=1}^Q P_n^q P_{a/n}^q = \sum_{q=1}^Q g_{aq} \quad (30)$$

lo cual garantiza que el modelo LJ también reproducirá las particiones de mercado observadas en la muestra de calibración para las alternativas anidadas.

3. CORRELACION

El caso mas sencillo y conocido de violación al supuesto de independencia de los errores, es cuando existen opciones percibidas como más similares entre sí que otras, por el usuario. Este problema es claramente ilustrado por la paradoja del bus azul - bus rojo (Mayberry, 1973), en que de hecho las alternativas son idénticas para el usuario. Supongamos una situación inicial de elección entre dos alternativas (auto y bus), en que ambas tienen idéntico nivel de utilidad media (V); en este caso, la probabilidad de elección de cada una de acuerdo al LS sería:

$$P = \frac{e^V}{e^V + e^V} = \frac{1}{2} \quad (31)$$

Si el operador de buses decide pintar la mitad de sus vehículos de color rojo y la otra mitad de color azul, se tendría una situación de elección con tres alternativas (auto, bus azul y bus rojo) en que el nivel de utilidad media seguiría siendo el mismo (V). Si se omite el efecto de la existencia de correlación entre las alternativas bus azul y bus rojo, el modelo LS predeciría que en este caso las probabilidades de elección serían:

$$P = \frac{e^V}{e^V + e^V + e^V} = \frac{1}{3} \quad (32)$$

es decir, en promedio ahora un tercio de los usuarios elegiría auto, un tercio elegiría bus azul e igual proporción bus rojo. Este resultado, que podría indicar a los operadores que con sólo pintar los buses de distinto color pueden aumentar considerablemente su participación de mercado, es claramente erróneo y no se sostiene al modelar correctamente, tomando en consideración la existencia de correlación entre las opciones similares (iguales en este caso extremo).

Otro caso ilustrativo, que de hecho sirvió a Williams (1977) para plantear el LJ, es el de elección conjunta de destino y modo de transporte. Si se supone que la decisión de elección de destino es realizada en el nido superior, que existe un conjunto de destinos D y un conjunto de modos M , se puede postular una función de utilidad del tipo $U(d,m) = Ud + Udm$. Acá Ud representa la utilidad del destino d

y tiene asociada un término de error con varianza σ_D^2 ; U_{dm} representa la utilidad de viajar en el modo m al destino d , y tiene asociado un término de error con varianza σ_{DM}^2 .

En términos del parámetro de escala de la distribución Gumbel, en el nodo inferior se debe cumplir que $\sigma_{DM}^2 = \pi^2/6\lambda^2$, en que λ es el factor de escala a ese nivel. En el nivel superior, la relación entre los parámetros de escala y la varianza es: $\pi^2/6\beta^2 = \sigma_D^2 + \pi^2/6\lambda^2$. Daganzo y Kusnic (1993) estudian la relación general entre varianza y parámetros de escala en el modelo LJ, demostrando que la covarianza entre las utilidades de dos opciones i y j , está dada por:

$$\text{cov}(U_i, U_j) = \frac{\pi^2}{6} [1 - \phi_{Y(i,j)}^2] \quad (33)$$

en que U_i , U_j son las utilidades de las alternativas i y j respectivamente. $Y(i,j)$ corresponde al ancestro común más cercano entre ambas alternativas, utilizando notación de grafos; es decir, si $i=j$ entonces $Y(i,j)=i$, si i y j no pertenecen a un mismo nodo, entonces el ancestro común más cercano será el nodo seminal 0. ϕ toma por convención los valores uno para el nodo seminal y cero para los nodos elementales (alternativas); el resto de los valores debe estar entre cero y uno, y debe cumplirse además que el ϕ de un ancestro sea siempre mayor que el de sus descendientes.

La forma de relacionar esto con el supuesto de que similitud entre opciones implica correlación, proviene de suponer que si dos opciones son percibidas como similares esto implica que existe una componente del error que el usuario asigna a ambas alternativas. Esta componente común, desconocida, se traduce en la presencia de elementos fuera de la diagonal en la matriz de covarianza. En el caso de destino y modo de transporte, en que la elección de destino se realiza en un nivel superior, esta componente común es precisamente la componente asociada al destino, que es obviamente percibida como igual. En este caso, si se supone arbitrariamente que la varianza σ_D^2 del término de error distribuido Logística, asociado al destino, es de la misma magnitud que la varianza σ_{DM}^2 del término de error distribuido Gumbel, asociado al modo, entonces el coeficiente de correlación es igual a 0,5; ya que la componente común entre dos opciones (σ_D^2) corresponde al 50% de la varianza del nivel superior ($\sigma_D^2 + \sigma_{DM}^2$). Por otro lado, en el caso de la paradoja del bus azul-bus rojo, ambas alternativas son idénticas, y por lo tanto el coeficiente de correlación es igual a 1,0.

El modelo LJ permite tratar correlación siempre y cuando ésta sea modelable a través de una estructura de árbol como la descrita anteriormente, en que una alternativa no puede pertenecer a dos ramas simultáneamente; es decir, la matriz de covarianza debe ser diagonal por bloques. Si analizamos esto en términos del parámetro ϕ de la estructura jerárquica, se tiene:

$$\phi = \frac{\beta}{\lambda} = \sqrt{\frac{\sigma_{DM}^2}{\sigma_D^2 + \sigma_{DM}^2}} \quad (34)$$

- $\rho = 0 \Rightarrow \sigma_D = 0 \Rightarrow \beta/\lambda = 1,0$
- $\rho = 0,5 \Rightarrow \sigma_D = \sigma_{DM} \Rightarrow \beta/\lambda = 1/\sqrt{2} = 0,707$
- $\rho = 1,0 \Rightarrow \sigma_{DM} = 0 \Rightarrow \beta/\lambda = 0$

Aplicando la formulación desarrollada por Daganzo y Kusnic (1993) se puede comprobar que se llega a exactamente los mismos resultados. Por ejemplo, suponiendo que el parámetro ϕ del nido es igual a 0,707, al aplicar la ecuación (20) se llega a:

- $i=j$, i alternativa elemental $\rightarrow \text{cov}(U_i, U_i) = \sigma^2 = \pi^2/6(1-\phi) = \pi^2/6$
- i, j distintos pero en un mismo nido $\rightarrow \text{cov}(U_i, U_j) = \pi^2/6(1-\phi, 0,707^2) = 0,5\sigma^2$

de lo cual se desprende que el coeficiente de correlación ρ es igual a 0,5 y, en consecuencia, ambos resultados son compatibles. Es trivial verificar que en los otros dos casos esto también se cumple; el primero ($\rho=0$), corresponde al modelo LS en que la componente común es nula y por lo tanto no hay correlación. El segundo caso corresponde a uno con coeficiente de correlación igual a 0,5, en que la componente común tiene tanto peso como las componentes individuales. Finalmente, se puede ver que el tercer caso ($\rho=1$) sólo tiene sentido cuando se trata efectivamente de dos alternativas idénticas, como en la paradoja del bus-azul y bus-rojo. Curiosamente en este caso, emblemático para introducir el modelo LJ, como $\phi=0$ el modelo se hace indeterminado y no se puede estimar.

En casos intermedios, se puede ver que a mayor correlación menor ϕ y *vice-versa*. Sin embargo, como ya fuera dicho, la aplicabilidad del modelo LJ se remite a aquellas estructuras de error que puedan ser descritas por alguna estructura de árbol. Esto presenta problemas, por ejemplo, cuando se modela la elección de modo en el caso de modos combinados, en que el viaje tiene una parte en un modo y otra en otro, ya que una alternativa no puede pertenecer simultáneamente a dos nidos (ver Ortúzar, 1983). Lo mismo sucede con el caso de elección de ruta, en que varias rutas pueden tener componentes en común. Además, se debe imponer que todas las alternativas dentro de un nido estén idénticamente correlacionadas.

En síntesis, las fuentes de correlación¹ son similitud entre alternativas (caso extremo: bus azul - bus rojo), y presencia de alternativas con una componente común (por ejemplo cuando se modela como alternativas disponibles para elección modal: auto, bus, metro, auto-metro y bus-metro). Para cada par de alternativas similares, o con una componente en común, debe aparecer un elemento fuera de la diagonal en la matriz de covarianza. Para los casos mencionados se debe cumplir además que todos los elementos fuera de la diagonal sean positivos, pero podría encontrarse ejemplos de fuentes de correlación negativa. Cabe señalar que el modelo LJ no permite modelar la presencia de correlación negativa.

En el caso en que se esté modelando modos combinados, podría pensarse en una componente del error asociada a todas las alternativas, con varianza σ^2 y otra asociada a cada uno de los modos en particular de varianza σ_i^2 para el modo i . Las componentes comunes son la causa de correlación y la matriz de covarianza tendrá una forma como la descrita en la Figura 2.

¹ No se debe confundir la correlación entre las componentes no observadas de error (ya sea entre alternativas o entre observaciones) con la correlación entre atributos, que si bien de estar presente también introduce problemas, se trata de un problema diferente en su naturaleza conocido en econometría como multicolinealidad.

	auto	bus	metro	auto-metro	bus-metro
auto	$\sigma^2 + \sigma_a^2$			σ_a^2	
bus		$\sigma^2 + \sigma_b^2$			σ_b^2
metro			$\sigma^2 + \sigma_m^2$	σ_m^2	σ_m^2
auto-metro	σ_a^2		σ_m^2	$\sigma^2 + \sigma_a^2 + \sigma_m^2$	σ_m^2
bus-metro		σ_b^2	σ_m^2	σ_m^2	$\sigma^2 + \sigma_b^2 + \sigma_m^2$

Figura 2: Ejemplo de matriz de covarianza con correlación cruzada

Ahora bien, también podría plantearse que la magnitud de estas componentes comunes depende de la proporción del viaje que se realice en cada modo. Es decir que no es lo mismo, en términos de percepción del término de error, si un viaje tiene el 10% en auto y el 90% en metro que si ocurre lo contrario. Esto se puede modelar incorporando un factor que multiplique el término de la varianza común, de forma proporcional al porcentaje del viaje que se realizaría en ese modo.

Otra fuente de correlación cruzada que ha sido reportada en la literatura, es la presencia de segmentos de ruta comunes en modelos de asignación estocástica (Sheffi, 1985). Al definir las distintas rutas potencialmente utilizables en una red, cada arco podrá formar parte de más de una ruta. Si consideramos que la utilidad de una ruta determinada estará dada por la suma de los atributos de sus arcos, es claro que la existencia de arcos comunes en varias rutas inducirá la existencia de elementos fuera de la diagonal en la matriz de covarianza en alguna forma que no necesariamente pueda ser modelada como una estructura de árbol (Yai *et al.*, 1997). Modelos que permiten tratar estructuras más sofisticadas de correlación entre alternativas son el Probit General (Daganzo, 1979) y el modelo de Componentes de Error o Logit Mixto (ver Train, 1998).

En la práctica profesional de nuestro país se ha planteado que modelos LS, usando transformadas Box-Cox en las variables que componen la función de utilidad, pueden obviar la necesidad de utilizar modelos LJ en presencia de correlación entre opciones. Este planteamiento se basa en suponer que al incorporar no-linealidades, de alguna manera se está permitiendo que el modelo incorpore la correlación. A continuación se muestra que el introducir transformadas Box-Cox en las variables, no induce correlación en el modelo LS.

Transformada Box-Cox y correlación

La transformada de Box-Cox se define como (ver Ortúzar y Willumsen, 1994):

$$X^{(\lambda)} = \begin{cases} (X^\lambda - 1) / \lambda & \text{si } \lambda \neq 0 \\ \log X & \lambda = 0 \end{cases} \quad (35)$$

La componente determinística de la función de utilidad del modelo LS, transformada Box-Cox en los atributos, se puede escribir como:

$$V_{jq} = \sum_k \theta_k X_{kjq}^{(\lambda)} \quad (36)$$

Dado que $U_j = V_j + \varepsilon_j$, la covarianza se puede escribir como:

$$\text{cov}(U_j, U_i) = \text{cov}(V_j + \varepsilon_j, V_i + \varepsilon_i)$$

$$= E[(\sum_k \theta_k X_{kj}^{(\lambda_k)} + \varepsilon_j - E(\sum_k \theta_k X_{kj}^{(\lambda_k)} + \varepsilon_j)) \cdot (\sum_k \theta_k X_{ki}^{(\lambda_k)} + \varepsilon_i - E(\sum_k \theta_k X_{ki}^{(\lambda_k)} + \varepsilon_i))] \quad (37)$$

Es claro que las componentes determinísticas en cada elemento del producto se anulan, a pesar de la no-linealidad. Sobreviven entonces sólo los términos aleatorios, quedando:

$$\text{cov}(U_j, U_i) = E[(\varepsilon_j - E(\varepsilon_j)) \cdot (\varepsilon_i - E(\varepsilon_i))] = \text{cov}(\varepsilon_j, \varepsilon_i) \quad (38)$$

Siendo este último término nulo, por definición del modelo. Analizando el caso desde este punto de vista (analítico), es claro que el introducir transformadas Box-Cox a las variables explicativas no permite que el modelo recoja correlación en los términos de error. Mas aún, en el caso inverso, en que la especificación correcta es no lineal (por ejemplo Box-Cox) pero el modelador asume una función de utilidad lineal, se puede demostrar que el modelo será sesgado, pero los términos de error estarán libres de correlación. Por lo tanto es claro que no existe ninguna relación entre el uso de transformadas Box-Cox y la existencia de correlación entre las componentes no observables de la utilidad.

4. CONCLUSIONES

Sobre la base de una derivación clara y fiel a sus orígenes del modelo Logit Jerárquico, hemos aclarado algunos puntos que permanecían oscuros, como por ejemplo la correcta especificación de constantes específicas, incluyendo una demostración de que este modelo también cumple la propiedad de reproducir las particiones de mercado observadas en la muestra de calibración. Hemos aclarado que no existe ninguna relación entre el uso de transformadas Box-Cox en la función de utilidad y la existencia de correlación en la componente no observable de la función de utilidad. Hemos discutido las fuentes de correlación, y qué casos pueden ser modelados dentro del marco del modelo Logit Jerárquico, discutiendo además detalladamente la relación entre la magnitud de la correlación y el parámetro estructural del modelo, que permite interpretar de mejor manera los resultados al momento de analizar modelos estimados.

Queremos agregar que nuestra experiencia con datos simulados y con datos reales, nos indica que estimar estadísticamente la magnitud de la correlación, ya sea a través del parámetro ϕ del modelo Logit Jerárquico, o directamente en el caso del modelo Probit, es extremadamente difícil, y requiere gran cantidad de datos de buena calidad para lograrlo (ver Munizaga *et al*, 1997). Sin embargo, recomendamos se siga usando el modelo Logit, tanto en sus versiones Simple como Jerárquica, ya que ambos tienen un comportamiento notablemente robusto ante cambios considerables de las condiciones en que fueron estimados.

AGRADECIMIENTOS

Este trabajo ha sido parcialmente financiado por FONDECYT, proyecto 1980722. Agradecemos a Juan Antonio Carrasco su minuciosa revisión de nuestras demostraciones y a Rodrigo Garrido sus sugerencias para hacer más fácilmente comprensible el documento.

REFERENCIAS

- Bates, JJ (1998) Comunicación privada.
- Bierlaire, M, Lotan, T y Toint, P (1997) On the overspecification of multinomial and nested logit models due to alternative specific constants. *Transportation Science* **31**, 363-371.
- Daganzo, CF (1979) *Multinomial Probit: The Theory and its Applications to Travel Demand Forecasting*. Academic Press, Nueva York.
- Daganzo, CF y Kusnic, M (1993) Two properties of the nested logit model. *Transportation Science* **27**, 395-400.
- Daly, AJ (1987) Estimating 'tree' logit models. *Transportation Research* **21B**, 251-268.
- Mayberry, JP (1973) Structural requirements for abstract-mode models of passenger transportation. En R.E. Quandt. (ed.), *The Demand for Travel: Theory and Measurement*. Health and Co., Lexington, Mass.
- McFadden, D (1974) Conditional logit analysis of qualitative choice behavior. En P Zarembka (ed.), *Frontiers in Econometrics*. Academic Press, Nueva York.
- Munizaga, MA, Heydecker, BG y Ortúzar, J de D (1997) On the error structure of discrete choice models. *Traffic Engineering and Control* **38**, 593-597.
- Ortúzar, J de D (1983) Nested logit models for mixed-mode travel in urban corridors. *Transportation Research* **17A**, 283-299.
- Ortúzar, J de D y Willumsen, LG (1994) *Modelling Transport*. Segunda Edición, John Wiley and Sons, Chichester.
- Sheffi, Y. (1985) *Urban Transportation Networks*. Prentice-Hall, Englewood Cliffs, N.J.
- Train, K (1998) Recreation demand models with taste differences over people. *Land Economics* **74**, 230-239.
- 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 HCWL y Ortúzar, J de D (1982) Behavioural theories of dispersion and the mis-specification of travel demand models. *Transportation Research* **16B**, 167-219.
- Wilson, AG, Hawkins, AF, Hill, GJ y Wagon, DJ (1969) Calibration and testing of the SELNEC transport model. *Regional Studies* **3**, 337-350.
- Yai, T, Iwakura, S y Morichi, S (1997) Multinomial probit with structured covariance for route choice behaviour. *Transportation Research* **31B**, 195-207.