

UNA REVISION Y ANALISIS DE LA ESPECIFICACION DEL MODELO LOGIT JERARQUICO

Juan Antonio Carrasco¹, Juan de Dios Ortúzar S.

Departamento de Ingeniería de Transporte, Pontificia Universidad Católica de Chile
Casilla 306, Cod. 105, Santiago 22, Chile. Fax: (56-2) 553 0281
e-mail: jos@ing.puc.cl

Marcela A. Munizaga M.

Departamento de Ingeniería Civil, Universidad de Chile
Casilla 228-3, Santiago, Chile. Fax: (56-2) 671 8788
e-mail: mamuniza@cec.uchile.cl

RESUMEN

Sorprende que, tras 25 años de haber sido formulado, aún no se comprendan cabalmente algunas propiedades del modelo logit jerárquico y se discuta su especificación. De hecho, este popular modelo ha sido objeto de una serie de acaloradas discusiones recientemente en la literatura. El objetivo de este trabajo es revisar las principales características del modelo en profundidad y examinarlas críticamente, a fin de aclarar todas las controversias. En primer lugar, se presentan los fundamentos teóricos del modelo, de forma concisa y fácil de entender, para luego examinar cuatro controversias en torno a su especificación: la correcta resolución del problema de no identificabilidad (común a todos los modelos de elección discreta), la aplicabilidad de la popular forma funcional NNNL, utilizada en el paquete estadístico ALOGIT, la reproducción de las particiones de mercado observadas y la pertinencia de una nueva especificación alternativa que provoca confusiones en relación a ciertas propiedades del modelo. Aparte del análisis teórico, se utiliza también simulación de Montecarlo para dilucidar algunos de estos aspectos. En este contexto, la principal conclusión del trabajo es que todas las hipótesis del modelo son esenciales para su correcta especificación e interpretación, y no deben ser ignoradas para un adecuado uso del modelo en la práctica.

¹ Actualmente en SDG Chile, Hernando de Aguirre 201, Of. 1301, Santiago, Chile

1. INTRODUCCION

Los modelos de elección discreta utilizados en modelación de demanda por transporte se basan en la teoría de la utilidad aleatoria (Domencich y McFadden, 1975). En ésta, la utilidad U_{iq} que el individuo q asocia a la alternativa discreta i , se representa como la suma de una componente observable V_{iq} y un error estocástico ε_{iq} . 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_{iq} se derivan los distintos modelos de elección discreta (ver Ortúzar, 1982).

En el contexto teórico anterior, los modelos clave de los últimos 25 años han sido el Logit Multinomial (MNL) y el Logit Jerárquico (Nested Logit, NL). Sin embargo, recientemente el NL ha sido objeto de una serie de discusiones y cuestionamientos que involucran no sólo su historia y génesis, sino que también algunas de sus propiedades fundamentales; esto puede tener importantes consecuencias en el uso práctico del modelo, principalmente en su especificación, estimación y en la interpretación de resultados. Por esta razón sentimos la necesidad de una revisión profunda del modelo, examinando críticamente sus propiedades y características fundamentales, a fin de posibilitar su correcta utilización en la práctica.

La Figura 1 muestra una estructura NL completamente general para los propósitos de nuestro análisis. Pese a que los “árboles” pueden tener más de dos niveles, como veremos más adelante, esta estructura es lo suficientemente general desde un punto de vista teórico: todas nuestras conclusiones pueden ser extendidas a tres o más niveles. En este trabajo no nos referiremos a la estimación estadística del modelo, ya que existen excelentes referencias para consultar este aspecto (Ben Akiva y Lerman, 1985; Daly, 1987; Hensher, 1986; Ortúzar y Willumsen, 1994; Hensher y Greene, 1999; Koppelman y Wen, 1998a).

El resto del trabajo está organizado como sigue. La sección 2 revisa y discute los fundamentos del modelo NL; éstas son la base de las especificaciones y propiedades del NL discutidas en la sección 3, donde también se discute la especificación del modelo y las implicancias teóricas de diversas alternativas. Finalmente, la sección 4 sintetiza nuestras principales conclusiones.

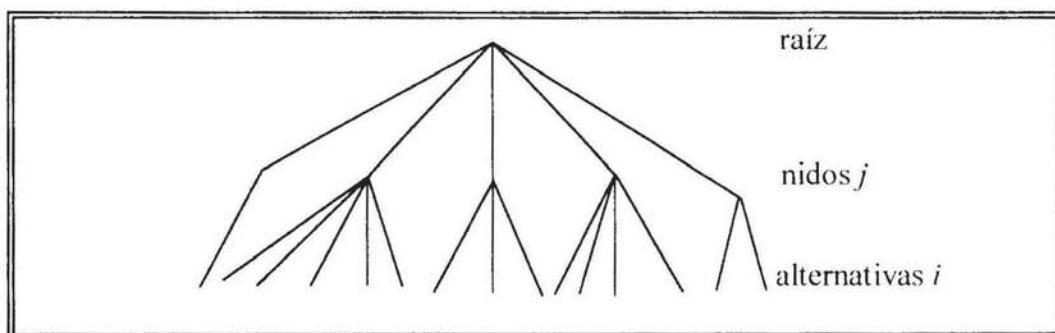


Figura 1: Estructura general de un árbol de dos niveles

2. FUNDAMENTOS DEL MODELO LOGIT JERARQUICO

En su revisión histórica del modelo NL, Ortúzar (2001) presenta a una serie de autores cuyas investigaciones sirvieron de antecedente para el modelo: Wilson (1974), Manheim (1973) y Ben-Akiva (1974), utilizaron versiones intuitivas que – pese a estar basadas en conceptos como probabilidades marginales o maximización de la utilidad – no poseían una construcción rigurosa de su forma funcional ni una interpretación inequívoca de *todos* los parámetros. Williams (1977), es el primer autor en realizar un análisis exhaustivo de las propiedades del NL, principalmente de los costos compuestos (o valores inclusivos), constatando que las versiones anteriores adolecían de fuertes inconsistencias con conceptos microeconómicos. Williams reformuló el modelo, enunciando las condiciones asociadas a sus parámetros estructurales necesarias para la compatibilidad del NL con la teoría de maximización de la utilidad. De esta forma, formuló el NL como un modelo *descriptivo* del comportamiento, completamente coherente con los supuestos básicos microeconómicos. Otros autores, cuyo trabajo completó el desarrollo de los fundamentos teóricos del modelo, son Daly y Zachary (1978), quienes trabajaron en forma simultánea y completamente independiente de Williams, y McFadden (1978; 1981), que generalizó más tarde el trabajo de Williams, Daly y Zachary.

Williams (1977) inicialmente trabajó con un modelo a dos niveles, en el contexto de situaciones bidimensionales como la elección de destino y modo de transporte, definiendo la siguiente función de utilidad:

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

en que i denota alternativas en niveles inferiores y j corresponde al índice del nivel superior al que éstas pertenecen. En términos de una utilidad representativa y términos estocásticos, la expresión (1) da lugar a:

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

$$\text{donde } V(i, j) = V_j + V_{i,j} \quad \text{y} \quad \varepsilon(i, j) = \varepsilon_j + \varepsilon_{i,j} \quad (3)$$

Su definición de los errores estocásticos se puede sintetizar en los siguientes supuestos:

- i) Los errores ε_j y $\varepsilon_{i,j}$ son independientes para todo (i, j)
- ii) Los términos $\varepsilon_{i,j}$ distribuyen Gumbel independiente e idéntico (iid), con parámetro de escala λ
- iii) ε_j posee una distribución tal que la suma entre U_j y el máximo de las utilidades $U_{i,j}$ distribuye Gumbel con parámetro de escala β .

El primer supuesto tiene como consecuencia la siguiente relación entre las varianzas de los errores:

$$\text{Var}(\varepsilon(i, j)) = \text{Var}(\varepsilon_j) + \text{Var}(\varepsilon_{i,j}) \quad (4)$$

es decir, dados los otros dos supuestos.

$$\frac{\pi^2}{6\beta^2} = \sigma_j^2 + \frac{\pi^2}{6\lambda_j^2} \quad (5)$$

o lo que es igual

$$\frac{\beta}{\lambda_j} = \left(1 + \frac{6\sigma_j^2\lambda_j^2}{\pi^2}\right)^{-1/2} \quad (6)$$

resultado que implica la *condición estructural* que el modelo *debe* cumplir (Williams, 1977):

$$\beta \leq \lambda_j \quad (7)$$

Ahora, definiendo el *parámetro estructural* $\phi = \frac{\beta}{\lambda_j}$, la condición (7) pasa a ser $\phi \leq 1$ y cuando $\beta = \lambda_j$ ($\phi = 1$), el NL colapsa al MNL; si $\beta > \lambda_j$ ($\phi > 1$), la estructura jerárquica postulada *en este contexto teórico* es incorrecta. Esta construcción puede generalizarse en dos direcciones. Permitiendo un parámetro de escala diferente λ_j , asociado a cada nido j , tal como lo sugieren Daly y Zachary (1978), y aumentando el número de niveles en serie y paralelo (Williams, 1977; Daly y Zachary, 1978; Sobel, 1980).

Una especificación muy usada del NL es la de dos niveles con diferentes parámetros de escala λ_j en cada nido (Figura 1), cuya forma funcional está dada por:

$$P_{ij} = \frac{\exp(\lambda_j V_{ij})}{\sum_{i \in j} \exp(\lambda_j V_{ij})} \cdot \frac{\exp \beta \left\{ \frac{1}{\lambda_j} \log \left(\sum_{i \in j} \exp(\lambda_j V_{ij}) \right) \right\}}{\sum_{j=1}^m \exp \beta \left\{ \frac{1}{\lambda_j} \log \left(\sum_{i \in j} \exp(\lambda_j V_{ij}) \right) \right\}} \quad (8)$$

En este caso, la condición estructural del modelo pasa a ser:

$$\beta \leq \lambda_j \quad \forall j \Leftrightarrow \phi_j = \frac{\beta}{\lambda_j} \leq 1 \quad \forall j \quad (9)$$

La condición (9) cobra especial importancia al interpretar los resultados de Börsch-Supan (1990), quien muestra que bajo circunstancias muy especiales el parámetro estructural ϕ_j podría ser mayor a uno. Sin embargo, su demostración se basa en el marco teórico de McFadden (1981), quien deriva la condición (9) sin utilizar la relación de las varianzas (4), sino estudiando qué condiciones debe cumplir la función densidad de probabilidad conjunta para ser apropiada matemáticamente. Esta diferencia plantea importantes restricciones en la utilización de la extensión de Börsch-Supan, puesto que implica la imposibilidad de utilizar el valor del parámetro estructural como una forma de conocer la relación jerárquica entre los distintos niveles (ver la discusión en Carrasco y Ortúzar, 2002). Además, *no es posible* utilizar esta extensión en modelos que impliquen separabilidad en

los niveles de elección (ej.: destino-modo), ya que ese supuesto se basa directamente en la condición de las varianzas (4)¹.

Finalmente se muestra la matriz de covarianzas del modelo anterior para el caso particular de dos niveles y cuatro alternativas anidadas de a dos:

$$\frac{\pi^2}{6\beta^2} \begin{bmatrix} 1 & 1-\phi_1^2 & 0 & 0 \\ 1-\phi_1^2 & 1 & 0 & 0 \\ 0 & 0 & 1 & 1-\phi_2^2 \\ 0 & 0 & 1-\phi_2^2 & 1 \end{bmatrix} \quad (10)$$

Se destaca que el modelo NL supone *homocedasticidad* entre alternativas; esto que es fácil de derivar a partir de los fundamentos anteriores, curiosamente no siempre es tomado en cuenta en la literatura² (ver Munizaga *et al*, 2000).

3. CONTROVERSIAS RECENTES EN TORNO A LA ESPECIFICACION DEL NL

Esta sección discute algunas propiedades del NL que han sido objeto de reciente controversia en la literatura, a fin de iluminar ciertos aspectos confusos y permitir un correcto uso del modelo en la práctica. Algunos de estos temas fueron tratados en Munizaga y Ortúzar (1999); lo que aquí se presenta constituye un avance respecto a esa publicación.

3.1. Especificaciones que Resuelven el Problema de No Identificabilidad

Como es sabido, todos los modelos de elección discreta (incluso el MNL) presentan el problema de no-identificabilidad. Por ejemplo, el parámetro β asociado al error Gumbel iid no puede ser estimado en el MNL. El modelo tiene que ser *normalizado*; es decir, los parámetros $\hat{\theta}_k$ son una estimación de los parámetros de la población deflactados por β (ver Ortúzar y Willumsen, 1994). El modelo NL, tal como se presenta en (8), no es estimable pues posee un grado de libertad; para hacerlo, es necesario “fijar” uno de los factores de escala. Considérese, sin pérdida de generalidad, el caso de dos niveles presentado en (8), donde el parámetro superior es β , y λ_j ($j = 1, \dots, J$) son los J parámetros para los correspondientes nidos. En este caso, se pueden definir J parámetros estructurales de la forma:

$$\phi_j = \frac{\beta}{\lambda_j} \quad j = 1, \dots, J \quad (11)$$

¹ Además, se destaca que las condiciones de Börsch-Supan son *locales*, es decir, deben ser probadas para cada valor de la muestra. Ejemplos prácticos de la literatura (ver Herriges y Kling, 1996) plantean serias dudas en torno a la aplicabilidad de éstas (para una discusión más detallada de estos aspectos, ver Carrasco, 2001)

² Es el caso del método discutido en Hensher y Louviere (1998) y Hensher (1999), quienes utilizan el modelo Heterocedástico de Valor Extremo (Bhat, 1995) para encontrar estructuras de árbol prometedoras, suponiendo indirectamente heterocedasticidad e ignorando el hecho que la formación de nidos se basa en las diferencias de correlación entre alternativas y no en la disimilitud de varianzas entre ellas.

La no-identificabilidad implica, en este caso, que si bien todos los parámetros ϕ_j son estimables no se puede conocer el valor de *uno* de los $J + 1$ factores de escala (esto es, los J parámetros λ_j más β) relacionados con la varianza.

3.1.1. Normalizaciones superior e inferior

De acuerdo a las definiciones anteriores, es posible distinguir dos tipos de normalización: la *superior*, en que el parámetro β es el elegido como no identificable y la *inferior*, en que uno de los parámetros λ_j (por ejemplo, $j = r$, con $1 \leq r \leq J$) es seleccionado. Considerérese, una función de utilidad lineal de la forma:

$$\hat{V}_{i,j} = \sum_{k=1}^K \hat{\theta}_k X_{(i,j)}^k \quad (12)$$

en que $X_{(i,j)}^k$ son los K atributos y $\hat{\theta}_k$ sus correspondientes parámetros estimados; éstos están relacionados con los correspondientes a la población (si existen) a través de:

$$\hat{\theta}_k = \beta \theta_k \quad \forall k \quad (13)$$

en el caso de la *normalización superior*; para la *normalización inferior* la relación es:

$$\hat{\theta}_k = \lambda_r \theta_k \quad \forall k \quad (14)$$

Las ecuaciones (13) y (14) permiten ver que la normalización no implica, en términos estrictos, "definir" el parámetro en cuestión como uno (indirectamente asumiendo el valor de la varianza), sino que el parámetro "normalizado" multiplica los coeficientes de la función de utilidad, "mezclándose" con ellos, más que implicando una definición *a priori* de su valor. Así, la especificación del modelo utilizando normalización superior es:

$$P(i,j) = \frac{\exp\left(\frac{1}{\phi_j} V_{i,j}\right)}{\sum_{i' \in J} \exp\left(\frac{1}{\phi_{i'}} V_{i',j}\right)} \cdot \frac{\exp \phi_j \left(\log \left(\sum_{i' \in J} \exp\left(\frac{1}{\phi_{i'}} V_{i',j}\right) \right) \right)}{\sum_{j'=1}^J \exp \phi_{j'} \left(\log \left(\sum_{i' \in J} \exp\left(\frac{1}{\phi_{i'}} V_{i',j'}\right) \right) \right)} \quad (15)$$

y en el caso de la normalización inferior es:

$$P_y = \frac{\exp\left(\frac{\phi_r}{\phi_j} V_{i,j}\right)}{\sum_{i' \in J} \exp\left(\frac{\phi_r}{\phi_{i'}} V_{i',j}\right)} \cdot \frac{\exp \phi_j \left(\log \left(\sum_{i' \in J} \exp\left(\frac{\phi_r}{\phi_{i'}} V_{i',j}\right) \right) \right)}{\sum_{j'=1}^m \exp \phi_{j'} \left(\log \left(\sum_{i' \in J} \exp\left(\frac{\phi_r}{\phi_{i'}} V_{i',j'}\right) \right) \right)} \quad (16)$$

Las ecuaciones (15) y (16) muestran que desde un punto de vista práctico las especificaciones resultantes son equivalentes a definir el parámetro correspondiente como uno en la forma funcional general (8) del NL. Sin embargo, tal como se discutirá más abajo la opción de normalizar en el nivel inferior implica decidir cuál de ellos usar, con potenciales confusiones cuando los parámetros de escala son diferentes, tal como se puede ver de las ecuaciones (13) y (14). Como un importante aspecto de la modelación es comunicar los resultados a los tomadores de decisiones, esto puede constituir un problema no menor.

3.1.2. Consideraciones teóricas

Ambas normalizaciones son posibles de homologar mediante las siguientes relaciones:

$$\hat{\phi}_j^{\text{sup}} = \hat{\phi}_j^{\text{inf}} \quad \forall j \quad (17)$$

$$\hat{\theta}^{\text{sup}} = \hat{\phi}_r \hat{\theta}^{\text{inf}} \quad (18)$$

en que los superíndices *inf* y *sup* indican normalización superior e inferior, respectivamente. Las ecuaciones (17) y (18) muestran que *ambas* especificaciones son equivalentes y, por ende, compatibles con los principios de maximización de la utilidad. Sin embargo, es interesante notar que dependiendo de la normalización elegida existirá una diferencia en el valor de los parámetros estimados. Esta disimilitud no es importante para casos en que interese la razón de coeficientes, como la estimación puntual del valor subjetivo del tiempo (Gaudry *et al.*, 1989), donde los factores de escala se cancelan, llegando al mismo resultado independientemente de la normalización. También el valor de las elasticidades es el mismo (Daly, 2001a).

Sin embargo, esta disimilitud puede ser importante si se desea comparar parámetros de un NL con, por ejemplo, los de un MNL. En este caso, *sólo es posible cotejar sus coeficientes con los obtenidos en la normalización superior*, pues en el MNL, los coeficientes estimados $\hat{\theta}$ corresponden al producto de los coeficientes poblacionales θ y el parámetro de escala. Para el caso del NL, tal como se aprecia en (13) y (14), sólo los coeficientes $\hat{\theta}^{\text{sup}}$ comprenden al parámetro β asociado a la varianza *total* de los errores de cada alternativa; los coeficientes $\hat{\theta}^{\text{inf}}$ consideran al parámetro λ_r que sólo está ligado a la varianza del nido donde se normaliza.

La discusión anterior permite concluir que, si bien ambas especificaciones son consistentes con la teoría, existen una serie de interesantes razones para preferir la normalización superior:

- La posibilidad de comparar directamente coeficientes estimados del NL con los del MNL
- La mayor sencillez que implica situar como parámetro de referencia al *único* que está relacionado directamente con la varianza total de las alternativas
- La mayor simplicidad en la forma funcional de la probabilidad en este caso.

3.1.3. Experimentos de simulación

Para analizar las características de las normalizaciones anteriores, se realizaron diversos experimentos de simulación que se presentan en detalle en Carrasco (2001). Estos siguieron el

marco teórico general utilizado por Williams y Ortúzar (1982), y fueron implementados como un caso realista donde sólo las variables explicativas y la opción elegida se encontraban disponibles en el proceso de estimación. Los atributos (costo, tiempo de viaje y tiempo de espera) fueron generados asumiendo una distribución Normal, con media y desviación estándar basadas en los datos de Ortúzar y Donoso (1985), y los coeficientes de las funciones de utilidad también tomaron valores realistas basados en los resultados de Ortúzar e Ivelic (1987) para la misma base de datos. Los errores estocásticos fueron construidos de acuerdo al marco teórico de Williams, presentado en las ecuaciones (1)-(4) y los modelos fueron estimados con el paquete estadístico GAUSS (Aptech Systems, 1994), usando el método de máxima verosimilitud.

La Tabla 1 presenta estimaciones utilizando las normalizaciones superior e inferior para una muestra de 64.000 individuos simulados³. La normalización inferior posee como referencia el parámetro λ_r . Como se puede verificar, los resultados confirman las relaciones (17) y (18) entre los parámetros de ambas normalizaciones. Por ejemplo, para el parámetro *Costo*, resulta la siguiente expresión: $\hat{\theta}^{sup} = \hat{\phi}_r \hat{\theta}^{inf} \Rightarrow -0,0056 \cong 0,5335 \cdot (-0,0104)$

Tabla 1: Estimación de las normalizaciones superior e inferior del NL

Coeficiente	Ref. ^a	Normalización Superior	Normalización Inferior
Costo	-0,005	-0,0056 (-30,9)	-0,0104 (-29,5)
Tiempo de Viaje	-0,080	-0,0952 (-76,6)	-0,1784 (-72,7)
Tiempo de Espera	-0,160	-0,1880 (-94,0)	-0,3524 (-81,1)
Constante Alt. 2	0,200	0,1997 (11,4)	0,3743 (11,6)
Constante Alt. 3	-0,200	-0,2684 (-14,0)	-0,5031 (-13,5)
Constante Alt. 4	-0,400	-0,4632 (-25,3)	-0,8681 (-23,1)
ϕ_1	0,550	0,5335 (62,7)	0,5335 (62,7)
ϕ_2	0,710	0,6705 (60,8)	0,6705 (60,8)
Log-verosimilitud		-0,9434	-0,9434

^a Los valores estimados teóricos son: $\hat{\theta}_{SUP} = \beta \theta_{REF}$ y $\hat{\theta}_{INF} = \lambda_r \theta_{REF}$ con $\beta = 1,2$ y $\lambda_r = 2,18$; *SUP* e *INF* corresponden a las normalizaciones superior e inferior y *REF* es el valor de referencia.

Otro aspecto destacable en la Tabla 1 es la disimilitud entre los test *t* en ambos casos. Esto ha sido examinado por Bates (1999) y Koppelman y Wen (1998b), quienes muestran que esta disimilitud está fuertemente relacionada con la diferencia en el valor de las varianzas y covarianzas de los parámetros estructurales y los coeficientes correspondientes. Si bien nuestros resultados confirman la diferencia de tests *t* para una estructura más compleja, vale destacar que las disimilitudes no resultaron mayores a un 15% y en la mayoría de los casos fueron menores a un 10% (en todas las muestras utilizadas, ver Carrasco, 2001). Luego, aunque estos resultados no son completamente generales, sirven para mostrar que la diferencia entre los test *t* de ambas normalizaciones no es crítica (para el típico valor 1,96 una reducción de un 15% es 1,67 y una de un 10% es 1,76). Este resultado va en la misma dirección que Daly (2000), quien arguye que esta disimilitud es otro recordatorio de que los test *t* son aproximados y no deben ser utilizados como

³ Es interesante mencionar que no se encontró influencia del tamaño muestral, grado de correlación o varianza de las variables independientes, en el valor de los parámetros ni en los tests *t* (aparte de su magnitud).

un criterio estricto de selección del mejor modelo. Un test que sí es estricto es el de razón de verosimilitud (ver Ortúzar, 1982), y además es invariante a la normalización.

3.2. Especificaciones UMNL y NNNL

A continuación se discute la controversia originada por Koppelman y Wen (1998a, 1998b) sobre dos posibles especificaciones del NL que se pueden encontrar en la literatura: la UMNL (*utility maximising nested logit*) y NNNL (*non normalised nested logit*), y en que han participado Daly (2001b), Hensher y Greene (1999), Hunt (2000) y Koppelman *et al* (2001).

3.2.1. Forma funcional de las especificaciones

La especificación UMNL es la definida originalmente por McFadden (1981). Luego, en el caso simple de dos niveles, la función de probabilidad es:

$$P(i, j) = \frac{\exp\left(\frac{\hat{V}_{i/J}}{\phi_j}\right)}{\sum_{i' \in J} \exp\left(\frac{\hat{V}_{i'/J}}{\phi_j}\right)} \cdot \frac{\exp\phi_j \left(\log\left(\sum_{i' \in J} \exp\left(\frac{\hat{V}_{i'/J}}{\phi_j}\right)\right)\right)}{\sum_{j'=1}^J \exp\phi_{j'} \left(\log\left(\sum_{i' \in J} \exp\left(\frac{\hat{V}_{i'/J}}{\phi_{j'}}\right)\right)\right)} \quad (19)$$

La especificación NNNL surgió debido a las dificultades que planteaba la estimación simultánea del modelo en sus inicios y es la base del método pionero propuesto por Daly (1987). En este caso, la función de probabilidad del modelo es idéntica a la anterior, excepto por la omisión de la inversa del parámetro estructural en las alternativas elementales; esto es:

$$P(i, j) = \frac{\exp\left(\hat{V}_{i/J}\right)}{\sum_{i' \in J} \exp\left(\hat{V}_{i'/J}\right)} \cdot \frac{\exp\phi_j \left(\log\left(\sum_{i' \in J} \exp\left(\hat{V}_{i'/J}\right)\right)\right)}{\sum_{j'=1}^J \exp\phi_{j'} \left(\log\left(\sum_{i' \in J} \exp\left(\hat{V}_{i'/J}\right)\right)\right)} \quad (20)$$

Esta especificación no es compatible con los principios fundamentales del NL si el modelo tiene atributos genéricos y/o si la correlación no es la misma en todos los nidos. En este último caso, es fácil ver que si se agrega una constante a la utilidad de todas las opciones, las probabilidades de elección cambian, violando la propiedad de *invarianza traslacional*. Ahora bien, contrariamente a lo que señalan Hensher y Greene (1999) y Hunt (2000), es importante notar que esta controversia *no* está relacionada con el problema de no-identificabilidad; lo que es fácil de comprobar si se compara la expresión (20) con las relaciones (15) y (16).

3.2.2. Comparación entre ambas especificaciones

En relación con la consistencia de ambas especificaciones con el principio de maximización de la utilidad, es claro que el UMNL cumple con esta propiedad debido a que es un caso particular de la familia GEV (General Extreme Value, ver McFadden, 1981). En cambio, aunque el NNNL puede ser consistente con una idea general de maximización de la utilidad (Daly, 2000), es claro que *no* es compatible con una de las propiedades fundamentales del modelo NL si la función de utilidad tiene coeficientes genéricos. Por lo tanto, coeficientes genéricos, estimados con una especificación NNNL, generan un modelo *inconsistente con el principio de la maximización de la utilidad*.

Sin embargo, es interesante notar que algunas condiciones particulares permiten que el NNNL sea consistente con el UMNL. Considérese, una vez más el NL de dos niveles (8); además definase a $\bar{\theta}_y$ y a θ_y como los vectores de parámetros de los atributos de las alternativas (i,j) , correspondientes al modelo de Williams y al NNNL, respectivamente. Es fácil ver que la relación entre ellos resulta ser:

$$\theta_y = \lambda_j \bar{\theta}_y \quad \forall i, j \quad (21)$$

De (21) se pueden derivar dos condiciones de equivalencia entre ambas especificaciones:

- i) Cuando los árboles tienen iguales parámetros estructurales ϕ en cada nivel; en el marco teórico de Williams esto implica que todos los parámetros λ , asociados a error estocástico al interior del nido (y su respectiva correlación), tienen el mismo valor.
- ii) Cuando el modelo no posee parámetros genéricos en alternativas en distintos nidos, en una especificación lineal de la función de utilidad. Esta segunda condición es menos restrictiva y probablemente válida en gran cantidad de casos reales.

Además, es posible modificar la especificación NNNL de forma que sea compatible con la teoría en los casos en que no es equivalente con la UMNL. Esto se logra utilizando nidos artificiales y ciertas restricciones (Bates, 1999; Hensher y Greene, 1999; Koppelman y Wen 1998b), tal como se recomienda en ALOGIT (Daly, 1992). En el Anexo se presentan tres ejemplos, con la estructura que se desea estimar a la izquierda y a la derecha la que se debe codificar en ALOGIT para obtener un resultado consistente con la teoría. Sin embargo, cabe destacar que estas modificaciones pueden implicar, en ciertos casos, la construcción de árboles contra-intuitivos y la necesidad de re-escalar parámetros para que sean equivalentes a los del modelo UMNL (ver Carrasco, 2001). Además, mientras más parámetros estructurales tenga un modelo, mayor será la complejidad que adquiera el árbol modificado del NNNL.

En síntesis, la especificación UMNL es claramente preferible. Una razón que ha sido esgrimida en favor de la especificación NNNL es la de eficiencia computacional, ya que posee una función de verosimilitud más sencilla (Daly, 1987); no obstante, sin duda la principal es que esta especificación es utilizada por ALOGIT, que es uno de los paquetes de estimación más populares en la práctica.

3.3. Sobre la Recuperación de Particiones de Mercado Observadas en el NL

Se ha discutido en forma tangencial a la literatura, el cumplimiento o no de la propiedad de reproducir las particiones de mercado observadas en el caso del modelo NL. Es claro que esta propiedad se cumple en el modelo Logit simple. En Munizaga y Ortúzar (1999) se realizó una demostración matemática cuyo resultado se interpretó diciendo que la propiedad también se cumplía para el modelo NL. No obstante, luego de analizar en profundidad la demostración, y validar la interpretación de resultados con simulación, pudimos concluir que la propiedad se cumplía en forma exacta sólo a nivel de nidos, pero se observaban pequeñas diferencias entre las particiones de mercado observadas y las predichas por el modelo para las alternativas que forman parte de un nido. Estas diferencias se minimizan cuando el parámetro estructural ϕ se acerca a uno (ver Carrasco, 2001).

3.4. Otra Especificación del NL y la Modelación de Heterocedasticidad

Hensher y Greene (1999) y Hunt (2000) proponen una definición alternativa de los coeficientes del NL que puede acarrear algunas confusiones acerca de algunas propiedades del modelo. Ellos definen un parámetro de escala para cada nido (γ_j) y otros para cada alternativa (μ_i) asociada a cada nido del que penden, tal como se muestra en la Figura 2.

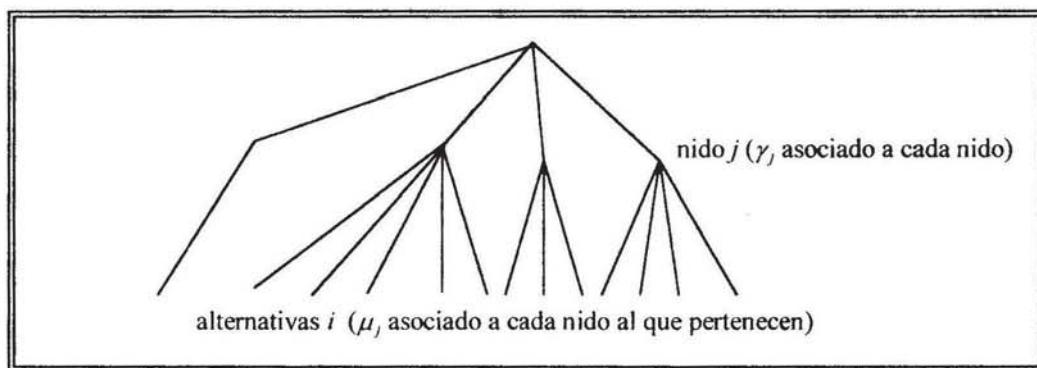


Figura 2: Definición alternativa de parámetros en el NL

La diferencia más importante entre esta especificación y la de Williams, es la incorporación de parámetros al nivel de las alternativas elementales en vez de un único parámetro asociado a la raíz. De esta forma, la nueva especificación posee $2J$ parámetros de escala en vez de los tradicionales $J + 1$ (en que J es el número de nidos). En esta visión alternativa, la probabilidad de elección está dada por:

$$P_j = \frac{\exp(\mu_j V_{i/J})}{\sum_{i \in J} \exp(\mu_j V_{i/J})} \cdot \frac{\exp \gamma_j \left\{ \frac{1}{\mu_j} \log \left(\sum_{i \in J} \exp(\mu_j V_{i/J}) \right) \right\}}{\sum_{j'=1}^m \exp \gamma_{j'} \left\{ \frac{1}{\mu_{j'}} \log \left(\sum_{i \in J'} \exp(\mu_{j'} V_{i/J'}) \right) \right\}} \quad (22)$$

Si se redefine las variables μ como λ y γ como β en (22) -sin considerar su significado teórico- es posible encontrar una especificación muy similar a la de Williams (8), excepto por el hecho de que se postulan diferentes parámetros β_j en cada nido (ver 23):

$$P_y = \frac{\exp(\lambda_j V_{i,j})}{\sum_{i \in j} \exp(\lambda_j V_{i,j})} \cdot \frac{\exp \beta_j \left\{ \frac{1}{\lambda_j} \log \left(\sum_{i \in j} \exp(\lambda_j V_{i,j}) \right) \right\}}{\sum_{j'=1}^m \exp \beta_{j'} \left\{ \frac{1}{\lambda_{j'}} \log \left(\sum_{i \in j'} \exp(\lambda_{j'} V_{i,j'}) \right) \right\}} \quad (23)$$

Esta especificación permitiría, indirectamente, la existencia de heterocedasticidad entre alternativas generalizando el NL de Williams y McFadden. Sin embargo, en realidad es altamente inconveniente en la práctica, pues introduce confusión al resolver el problema de no-identificabilidad del modelo. En efecto, en este caso es necesario normalizar J parámetros; es decir, se definen como no estimables o *todos* los parámetros β_j (normalización superior) o *todos* los parámetros λ_j (normalización inferior). Esto es diferente a las normalizaciones discutidas anteriormente, donde para resolver el problema de no-identificabilidad era necesario definir *sólo un* parámetro de escala como no-identificable.

Como consecuencia de lo anterior, la normalización superior resulta ser la misma de Williams (es decir, está correcta); sin embargo, la normalización inferior conduce a la especificación NNNL con todos los potenciales problemas de compatibilidad con la teoría, explicados anteriormente. Además, con esta especificación alternativa se producen serios problemas en estructuras *parcialmente degeneradas* (árboles con algunos nidos que contienen una sola opción), que no ocurren si se utiliza la especificación tradicional. Así, cuando se requiera modelar correlación y heteroscedasticidad, parece más natural usar el modelo Mixed Logit (ver la discusión en Alvarez y Munizaga, 2001) en que, por construcción, sucede que al introducir correlación también se introduce heteroscedasticidad en la matriz de covarianzas.

4. CONCLUSIONES

El análisis de las propiedades fundamentales del modelo NL ha servido para examinar algunas controversias relacionadas con su especificación. Primero se analizó las normalizaciones que permiten resolver el problema de no-identificabilidad, argumentando la conveniencia de la normalización superior. Luego se estudió la disimilitud de test t entre ambas especificaciones, argumentando que este popular estadígrafo sólo debe ser utilizado como referencia y no como una restricción absoluta. También se analizó los casos en que la popular especificación NNNL del modelo no es compatible con la teoría. Finalmente se estudió una nueva especificación del NL, propuesta por Hensher y Greene (1999) y Hunt (2000), mostrando que provoca importantes confusiones en vez de facilitar el uso e interpretación del modelo, por lo que no se recomienda su utilización. Esperamos que nuestra discusión permita un mejor uso del modelo NL en la práctica.

AGRADECIMIENTOS

Agradecemos a John Bates, Andrew Daly, David Hensher, Frank Koppelman y Huw Williams por sus comentarios y sugerencias a versiones iniciales de este trabajo. No obstante, todos los errores

y afirmaciones incorrectas que permanezcan son, por supuesto, de nuestra completa responsabilidad. Finalmente queremos agradecer el apoyo financiero del Fondo Nacional de Desarrollo Científico y Tecnológico (FONDECYT).

REFERENCIAS

- Aptech Systems (1994) **GAUSS User's Manual**. Mapple Valley.
- Alvarez, R. y Munizaga, M.A. (2001) Modelos mixed logit: antecedentes teóricos y aplicaciones. **X Congreso Chileno de Ingeniería de Transporte**, Universidad de Concepción, 8-12 Octubre 2001.
- Bates, J. (1999) Comunicación privada.
- Ben-Akiva, M.E. (1974) Structure of passenger travel demand models. **Transportation Research Record** **526**, 26-42.
- Ben-Akiva, M.E. y Lerman, S.R. (1985) **Discrete Choice Analysis: Theory and Application to Travel Demand**. MIT Press, Cambridge, Mass.
- Bhat, C.R. (1995) A heteroskedastic extreme value model of intercity travel mode choice. **Transportation Research** **29B**, 471-483.
- Börsch-Supan, A. (1990) Recent developments in flexible discrete choice models: nested logit analysis versus simulated moments probit analysis. En M.M. Fisher, P. Nijkamp e Y.Y. Papageorgiou (eds.), **Behavioural Modelling of Spatial Choices and Processes**. North Holland, Amsterdam.
- Carrasco, J.A. (2001) Elección Discreta de Alternativas Homocedásticas Correlacionadas: El Modelo Logit Jerárquico en Profundidad. Tesis de Magister, Departamento de Ingeniería de Transporte, Pontificia Universidad Católica de Chile.
- Carrasco, J.A. y Ortúzar, J. De D. (2002) A review and assessment of the nested logit model. **Transport Reviews** **22** (en imprenta).
- Daly, A.J. (1987) Estimating “tree” logit models. **Transportation Research** **21B**, 251-267.
- Daly, A.J. (1992) **ALOGIT 3.2 User's Guide**. Hague Consulting Group, The Hague.
- Daly, A.J. (2000, 2001a) Comunicaciones privadas.
- Daly, A.J. (2001b). Alternative tree logit models: comments on a paper by Koppelman and Wen. **Transportation Research** **35B** (en imprenta).
- Daly, A.J. y Zachary, S. (1978) Improved multiple choice models. En D.A. Hensher y M.Q. Dalvi (eds.), **Determinants of Travel Choice**. Saxon House, Sussex.

Domencich, T.A. y McFadden, D. (1975) **Urban Travel Demand: A Behavioural Analysis**, American Elsevier, Nueva York.

Gaudry, M.J.I., Jara-Díaz, S.R. y Ortúzar, J. de D. (1989) Value of time sensitivity to model specification. **Transportation Research** **23B**, 151-158.

Hensher, D.A. (1986) Sequential and full information maximum likelihood estimation of a nested logit model. **The Review of Economics and Statistics** **LXVIII**, 657-720.

Hensher D.A. (1999) HEV choice models as a search engine for the specification of nested logit tree structures. **Marketing Letters** **10**, 339-349

Hensher, D.A. y Greene, W.H. (1999) Specification and estimation of nested logit models. Working Paper, Centre for Transport Studies, University of Sydney.

Hensher D.A. y Louviere, J.J. (1998) A comparison of elasticities derived from multinomial logit, nested logit and heterocedastic extreme value SP-RP discrete choice models. **8th World Conference of Transportation Research**, Amberes.

Herriges, J. y Kling, C. (1995) An empirical investigation of the consistency of nested logit models with utility maximization. **Economic Letters** **50**, 33-39.

Hunt, G.L. (2000) Alternative nested logit model structures and the special case of partial degeneracy. **Journal of Regional Science** **40**, 89-113.

Koppelman, F.S. y Wen, C.H. (1998a) Alternative nested logit models: structure, properties and estimation. **Transportation Research** **32B**, 289-298.

Koppelman, F.S. y Wen, C.H. (1998b) Different nested logit models: which are you using? **Transportation Research Record** **1645**, 1-7.

Koppelman F.S., Sethi, V. y Wen, C.H. (2001) Alternative nested logit models: a response to comments by Andrew Daly on an earlier paper by Frank Koppelman and Chieh-Hua-Wen. Working Paper, Northwestern University.

Manheim, C.F. (1973) Practical implications of some fundamental properties of travel demand models. **Highway Research Record** **244**, 21-38

McFadden, D. (1978) Modelling the choice of residential location. En A. Karlqvist, L. Lundqvist, F. Snickars y J.W. Weibull (eds.), **Spatial Interaction Theory and Planning Models**. North Holland, Amsterdam.

McFadden, D. (1981) Econometric models of probabilistic choice. En C.F. Manski y D. McFadden (eds.), **Structural Analysis of Discrete Data with Econometric Applications**. The MIT Press, Cambridge, Mass.

Munizaga, M.A., Heydecker, B.G. y Ortúzar, J. de D. (2000) Representation of heteroskedasticity in discrete choice models. **Transportation Research** 34B, 219-240

Munizaga, M.A. y Ortúzar, J. de D. (1999) Correlación entre alternativas: el modelo logit jerárquico en profundidad. En J.E. Coeymans y P. Sommariva (eds.), **Actas del IX Congreso Chileno de Ingeniería de Transporte**, Sociedad Chilena de Ingeniería de Transporte, Santiago.

Ortúzar, J. de D. (1982) Fundamentals of discrete multimodal choice modelling. **Transport Reviews** 2, 47-48.

Ortúzar, J. de D. (2001) On the development of the of nested logit model. **Transportation Research** 35B, 213-216.

Ortúzar J. de D. y Donoso, P.C.F. (1985) Modal choice modelling for several alternatives: application of disaggregate demand models in Santiago, Chile. En B.G. Hutchinson, P. Nijkamp y M. Batty (eds.), **Optimization and Discrete Choice in Urban Systems**. Springer-Verlag, Berlin.

Ortúzar J. de D. e Ivelic, A.M. (1987) Effects of using more accurately measured level-of-service variables on the specification and stability of mode choice models. **Proceedings 15th PTRC Summer Annual Meeting**. University of Bath, England.

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

Sobel, K.L. (1980) Travel demand forecasting with the nested multinomial logit model, **59th Annual meeting of the Transportation Research Board**, Washington, D.C.

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.

Wilson A.G. (1974) **Urban and Regional Models in Geography and Planning**. John Wiley and Sons, Chichester.

ANEXO: EJEMPLOS DE TRANSFORMACION DE LA ESPECIFICACION NNNL