*IX Congreso de Ingeniería de Organización*
*Gijón, 8 y 9 de septiembre de 2005*

# Rango de admisibilidad de valores de inclusión en un modelo de elección discreta jerárquico

**J. Nicolás Ibáñez[1], Luis Onieva[1], Jesús Muñuzuri[1], Pablo Cortés[1]**

[1] Grupo de Ingeniería de Organización. Escuela Superior de Ingenieros de Sevilla.
Avda. Descubrimientos, s/n, 41092 Sevilla.
juannicolas@us.es, onieva@us.es, munuzuri@esi.us.es, pca@esi.us.es

**Palabras clave:** Modelos jerárquicos de elección discreta, Valor de inclusión

**Resumen**

*En este trabajo se profundiza en las expresiones de la probabilidad propias de un modelo jerárquico de elección discreta de dos niveles. En la literatura existe una controversia generalizada acerca de los valores admisibles que puede tener el cociente de dos de los parámetros involucrados en el cálculo de estas probabilidades (valor de inclusión), hasta tal extremo de considerar que los modelos con valores de este cociente fuera del rango 0-1 están mal especificados y son contrarios a la teoría más básica de comportamiento de los individuos (teoría RUM). Las probabilidades del modelo se derivan en este trabajo mediante un proceso que denominamos de sucesiva asunción de hipótesis, parecido pero no idéntico al de Ben-Akiva & Lerman (1985), y diferente al de función de valor extremo generalizado propuesto en McFadden (1978), siendo éste último uno de los que conllevan al rango de admisibilidad unitario mencionado. En este trabajo se hace una relajación de este rango unitario, se demuestra que no es contraria a la teoría RUM y se interpreta en qué se traduce en término de los parámetros del modelo.*

## 1. INTRODUCCIÓN A MODELOS NO LINEALES DE PREFERENCIAS JERÁRQUICOS

En un elevado número de trabajos se introducen y aplican modelos no lineales de preferencias de tipo jerárquico (*modelos logit jerárquicos o anidados*) como una herramienta eficiente para relajar algunas de las desventajas de los modelos no lineales más simples, de entre las que se considera como más restrictiva la existencia de un patrón de sustitución proporcional entre alternativas.

Una de las razones principales para la extensión de este tipo de modelos en detrimento de otras generalizaciones del modelo logit simple (probit, logit mixtos, logit heterocedástico, etc.) es el hecho de que mantienen la forma cerrada para las probabilidades asignadas por el modelo a las alternativas entre las que deciden los individuos. Así, el empleo de modelos jerárquicos no hace necesario la simulación numérica de las ecuaciones de probabilidad, redundando en procesos de estimación significativamente más simples.

Sin embargo, pese a esta relativa expansión de los modelos jerárquicos, es común en la literatura que el proceso de obtención de las probabilidades no sea detallado en profundidad. Esta ausencia de detalle se compensa citando a tres trabajos de referencia de finales de los años 70, donde de forma independiente y utilizando demostraciones distintas, se construyen las probabilidades en modelos jerárquicos cumpliendo con el paradigma supuesto al comportamiento de los individuos decisores y denominado de maximización de la utilidad aleatoria (*teoría RUM*). Estos trabajos son los de Williams (1977), Daly & Zachary (1976) y McFadden (1978), si bien el primer manual donde los modelos jerárquicos se presentan explícitamente englobados en un marco más general de teoría de modelos no lineales de preferencias es Ben-Akiva & Lerman (1985).

Por lo general, incluso la cita a estos trabajos de referencia añade más confusión a la discusión sobre modelos jerárquicos, pues al tratarse de demostraciones que adoptan enfoques distintos para probar la adecuación de los modelos jerárquicos a RUM, las conclusiones derivadas acerca de la admisibilidad de los valores de los parámetros en las expresiones de probabilidad son también distintas. En esta ponencia analizamos el trabajo de Ben-Akiva&Lerman, y relajamos a partir de él el rango de admisibilidad para un cociente de parámetros de especial interés en el modelo (*valor de inclusión*).

## 2. *NOTACIÓN Y EJEMPLO DE UN MODELO JERÁRQUICO DE DOS NIVELES*

En este trabajo se abordan estructuras jerárquicas de dos niveles, donde las alternativas elementales que constituyen el nivel inferior son agrupadas sólo una vez en conjuntos o ramas que constituyen el nivel superior (la utilización de estructuras jerárquicas de más de dos niveles es bastante menos frecuente en la literatura y su formulación se obtiene por generalización de este caso más simple). El enfoque utilizado en este trabajo es descomponer un modelo jerárquico de dos niveles en dos modelos logit simples no jerárquicos relacionados.

En la Figura 1 se muestra la notación empleada y la ilustración de una estructura ejemplo para estudiar las preferencias individuales sobre medio de transporte urbano.

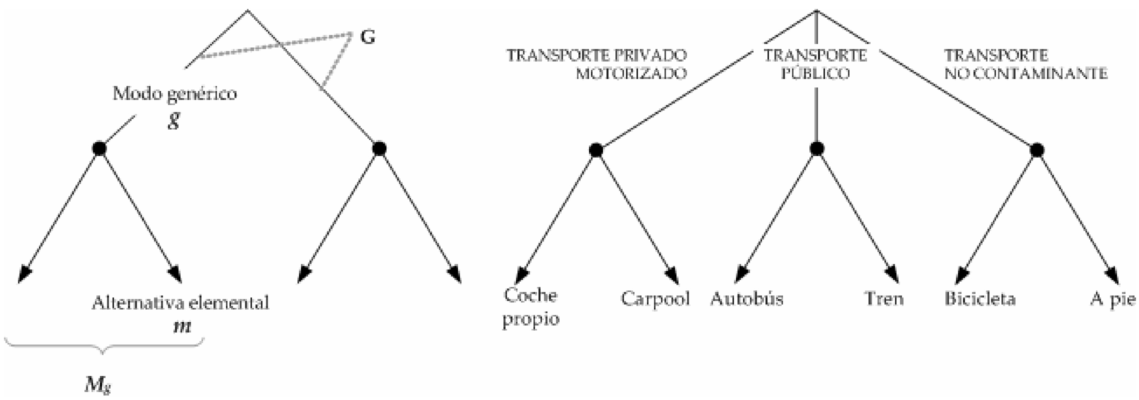


**Figura 1** Notación y ejemplo de un modelo jerárquico de dos niveles

## 3. *PROBABILIDADES DEL MODELO. ASUNCIÓN SUCESIVA DE HIPÓTESIS*

Los modelos jerárquicos son un tipo particular de modelos de elección discreta y por ello su aplicación se hace sobre respuestas de los individuos derivadas de la valoración simultánea de un grupo de alternativas (conjunto de decisión); cada alternativa del conjunto (autobús, coche propio, etc.) está caracterizada por un conjunto de atributos (tiempo de viaje, coste generalizado de desplazamientos, etc). El objetivo de este modelo es cuantificar el peso de los atributos en las respuestas de los individuos.

Las expresiones de la utilidad para cada una de las alternativas disponibles para el individuo en cada conjunto de decisión se derivan del siguiente análisis. El tipo de decisiones que realizan los individuos se asume basado en la teoría de comportamiento RUM o de maximización de la utilidad aleatoria. Los primeros trabajos a este respecto son los de Block & Marschak (1960), donde se establece la necesidad de considerar un vector de utilidades *aleatorio* para cada conjunto de decisión $\mathbf{U} = (U_1,...,U_j,\ldots,U_r)'$, de tal forma que la probabilidad asignada por el modelo de que la alternativa *j* sea elegida es la siguiente,

$$P_j = \Pr\left(U_j \geq U_k\right) \text{ para } k = 1,...,r,\ k \neq j$$

En el modelo jerárquico de dos niveles la utilidad (aleatoria) de una alternativa *gm* es.

$$U_{gm} = U_g + U_{m|g} = V_g + \left(V_{m|g} + V_m\right) + \varepsilon_g + \left(\varepsilon_{m|g} + \varepsilon_m\right) \ ,\forall g \in G, \forall m \in M_g$$

donde $g|m$ es la utilidad (aleatoria) de la alternativa elemental *m* condicionada en su pertenencia a la rama de la estructura *g*. Los términos épsilon son aleatorios para conferir a la utilidad el carácter establecido por RUM, y los términos *V* son deterministas y contienen la aportación de los atributos ( $x_{ki}$ ) a las alternativas ($V_i = \sum_k \beta_k x_{ki}$ ).

Haciendo uso de una **hipótesis básica** para evitar términos con el mismo comportamiento se asume que,

$$\varepsilon_{m|g} + \varepsilon_m \rightarrow \varepsilon_{m|g} \quad , \quad V_{m|g} + V_m = V_{m|g}$$

con lo que finalmente la expresión de la utilidad es la siguiente,

$$U_{gm} = U_g + U_{m|g} = V_g + V_{m|g} + \varepsilon_g + \varepsilon_{m|g} \ ,\forall g \in G, \forall m \in M_g$$

El sistema de probabilidades construido en base a la teoría RUM ha de cumplir con las siguientes tres propiedades según Block & Marschak, para así mantener la consistencia más básica del modelo con las decisiones observadas de los individuos (*postulados de la Teoría de la Maximización de la Utilidad Aleatoria* o *postulados RUM*):

$$a) \ 0 \le P_j \le 1, \ \forall j, \forall m \ ; \ \sum_{i=1}^{r} P_i = 1$$

$$b) \ P_j\left(U_1 + \alpha, \ldots, U_r + \alpha\right) = P_j\left(U_1, \ldots, U_r\right), \forall j, \forall \alpha \in \mathbb{R}$$

$$c) \ P_j\left(\lambda U_1, \ldots, \lambda U_r\right) = P_j\left(U_1, \ldots, U_r\right), \forall j, \forall \lambda \in \mathbb{R}^+$$

Aplicando esta teoría RUM con la notación del caso de un modelo jerárquico de dos niveles se tiene que la probabilidad de que un individuo elija una alternativa es la siguiente,

$$P_{gm} = \Pr\left(U_{gm} = \max_{\forall q \in G, \forall h \in M_q} \left\{U_{qh}\right\}\right) = \Pr\left(U_{gm} \ge U_{qh}, \forall q \in G, \forall h \in M_q\right)$$

Se realizan ahora las siguientes dos hipótesis acerca del comportamiento de los términos aleatorios de las utilidades de las alternativas para proceder al cálculo de las expresiones de las probabilidades que éstas sean elegidas,

**HIPÓTESIS 1:** $\varepsilon_{m|g}$ i.i.d. $Gumbel\left(0, \lambda_g\right), \forall m \in M_g$

**HIPÓTESIS 2:** $\left(\varepsilon_g + \varepsilon_{g*}\right)$ i.i.d. $Gumbel\left(0, \mu\right), g \in G$

En este cálculo también se emplea la siguiente propiedad de las distribuciones de valor extremo,

**PROPIEDAD DISTRIBUCIÓN GUMBEL 1:** $\varepsilon_{g*}, \varepsilon_{m|g} \rightarrow Gumbel\left(0, \lambda_g\right)$

La relación entre la Hip1 y la Prop1 es la siguiente:

$$\max_{n \in M_g} U_{gn} = V_g + \varepsilon_g + \max_{n \in M_g}\left(V_{n|g} + \varepsilon_{n|g}\right) \overset{\text{Hip1 \& Prop1}}{=} V_g + \varepsilon_g + Gumbel\left(\frac{1}{\lambda_g} \ln \sum_{n \in M_g} e^{\lambda_g V_{n|g}}, \lambda_g\right)$$

$$= \left(V_g + V_{g*}\right) + \left(\varepsilon_g + \varepsilon_{g*}\right)$$

$$V_{g*} = \frac{1}{\lambda_g} \ln \sum_{n \in M_g} e^{\lambda_g V_{n|g}} \quad , \quad \varepsilon_{g*} \rightarrow Gumbel\left(0, \lambda_g\right)$$

Obsérvese que las distribuciones $\varepsilon_{g*}, \varepsilon_{m|g}$ no so independientes aunque se distribuyan idénticamente, es decir, $\operatorname{cov}(\varepsilon_{g*}, \varepsilon_{m|g}) \neq 0$; este es precisamente uno de los resultados fundamentales del que se derivan las conclusiones de este trabajo.

Con el empleo de Hip.1, Hip.2 y Prop.1 se obtiene la expresión para las probabilidades para un modelo de elección discreta de dos niveles,

$$P_{gm} = P_g \cdot P_{m|g} = \frac{\exp\left(\mu V_g + \frac{\mu}{\lambda_g} \ln \sum_{n \in M_g} \exp\left(\lambda_g V_{n|g}\right)\right)}{\sum_{q \in G} \exp\left(\mu V_q + \frac{\mu}{\lambda_q} \ln \sum_{n \in M_q} \exp\left(\lambda_q V_{n|q}\right)\right)} \cdot \frac{\exp\left(\lambda_g V_{m|g}\right)}{\sum_{n \in M_g} \exp\left(\lambda_g V_{n|g}\right)} \ , \forall g \in G \ , \forall m \in M_g$$

Esta misma expresión se alcanza si se parte del enfoque de función de extremo generalizado de McFadden (1978); en tal enfoque sin embargo, las hipótesis necesarias sobre el comportamiento de las partes aleatorias de la utilidad establece que ha de cumplirse que,

$$0 \leq \mu / \lambda_g \leq 1 \ , \forall g \in G$$

restricción conocida en la literatura como la de *valores admisibles para el valor de inclusión*. En este trabajo se desarrolla una expresión más general para este rango mediante empleo de la derivación de probabilidades con asunción sucesiva de hipótesis. Este proceso de asunción sucesiva aparece también en Ben-Akiva & Lerman (1985), pero llegando al mismo rango de admisibilidad unitario de McFadden (1978), haciendo que lo que aquí se expone sea una relajación de lo establecido por ambos trabajos.

### *3.1. Hipótesis implícita en Ben-Akiva & Lerman (Hip.3A)*

Ben-Akiva & Lerman (BA&L) realizan una hipótesis que no especifican explícitamente, de cuyo análisis y relajación se extraen rangos de admisibilidad más amplios para el valor de inclusión. La equivalencia de la notación de BA&L con la utilizada más arriba en este trabajo es la siguiente $\mu^m = \mu$ ; $\mu^d = \lambda_g$ ; $\tilde{\varepsilon}_{dm} = \varepsilon_{m|g}$ ; $\tilde{\varepsilon}_m = \varepsilon_g$ ; $\varepsilon'_m = \varepsilon_{g*}$.

En la página 287 de BA&L se establece que,

$$\left(\mu^m / \mu^d\right)^2 = \frac{\operatorname{var}\left(\tilde{\varepsilon}_{dm}\right)}{\operatorname{var}\left(\tilde{\varepsilon}_m + \varepsilon'_m\right)} = \frac{\operatorname{var}\left(\tilde{\varepsilon}_{dm}\right)}{\operatorname{var}\left(\tilde{\varepsilon}_m\right) + \operatorname{var}\left(\tilde{\varepsilon}_{dm}\right)}$$

y que $\tilde{\varepsilon}_m$ y $\tilde{\varepsilon}_{dm}$ son independientes, por lo que $\operatorname{var}\left(\tilde{\varepsilon}_m\right) + \operatorname{var}\left(\tilde{\varepsilon}_{dm}\right) = \operatorname{var}\left(\tilde{\varepsilon}_m + \tilde{\varepsilon}_{dm}\right)$. En términos del valor de inclusión esta independencia se traduce en que,

$$\left(\mu^m / \mu^d\right)^2 = \frac{\operatorname{var}\left(\tilde{\varepsilon}_{dm}\right)}{\operatorname{var}\left(\tilde{\varepsilon}_m + \varepsilon'_m\right)} = \frac{\operatorname{var}\left(\tilde{\varepsilon}_{dm}\right)}{\operatorname{var}\left(\tilde{\varepsilon}_m + \tilde{\varepsilon}_{dm}\right)} \Rightarrow \operatorname{var}\left(\tilde{\varepsilon}_m + \varepsilon'_m\right) = \operatorname{var}\left(\tilde{\varepsilon}_m + \tilde{\varepsilon}_{dm}\right)$$

Lo que los autores están diciendo por tanto es que los valores de las varianzas de las distribuciones $\tilde{\varepsilon}_m + \varepsilon'_m$ y $\tilde{\varepsilon}_m + \tilde{\varepsilon}_{dm}$ son idénticos, probablemente guiados por la relación existente entre $\tilde{\varepsilon}_{dm}$ y $\varepsilon'_m$,

$$\left.\begin{array}{l} \varepsilon'_m = \max\limits_{d \in \mathbf{D}_{nm}} \left(\tilde{V}_d + \tilde{V}_{dm} + \tilde{\varepsilon}_{dm}\right) - V_m \\ \tilde{\varepsilon}_{dm} \rightarrow IID\ Gumbel\ d \in \mathbf{D}_{nm} \end{array}\right\} \Rightarrow \operatorname{var}\left(\varepsilon'_m\right) = \operatorname{var}\left(\tilde{\varepsilon}_{dm}\right)$$

El hecho es que con las hipótesis que se explicitan en BA&L (HIP1 y HIP2 de este trabajo) no es posible establecer la igualdad de $\tilde{\varepsilon}_m + \varepsilon'_m$ y $\tilde{\varepsilon}_m + \tilde{\varepsilon}_{dm}$ si no se realiza la que denominamos Hipótesis 3A (asumida **implícitamente** en BA&L) sobre independencia de los términos aleatorios $\tilde{\varepsilon}_m$ y $\varepsilon'_m$,

$$\operatorname{cov}\left(\tilde{\varepsilon}_m, \varepsilon'_m\right) = 0$$

En la expresión de las probabilidades $P_{gm}$ calculada anteriormente se ha utilizado el resultado conocido de que la probabilidad marginal de elegir una rama *g* de la estructura jerárquica, multiplicada por la probabilidad de elegir una alternativa *m* condicionada en su pertenencia a esta rama, es igual a la probabilidad de elegir la alternativa en cuestión ($P_{gm} = P_g \cdot P_{m|g}$). Para aseverar esta igualdad de probabilidades es necesario asumir que $\operatorname{cov}(\varepsilon_g, \varepsilon_{m|g}) = 0$, la cual no se deriva directamente de las Hipótesis 1 o 2, por lo que le asignamos la etiqueta de Hipótesis 3B, y cuyo cumplimiento no es opcional como el de 3A, sino necesario para obtener la expresión de $P_{gm}$ mencionada,

**HIPÓTESIS 3:** **3A.** $\operatorname{cov}(\varepsilon_g, \varepsilon_{g*}) = 0$ **3B.** $\operatorname{cov}(\varepsilon_g, \varepsilon_{m|g}) = 0$

Antes de abundar en la sección posterior sobre el cálculo del valor de inclusión, se remarca que sólo las Hip.1-2-3B son necesarias para derivar la expresión de las probabilidades y que la asunción de 3B no conlleva necesariamente el cumplimiento de 3A, pues

$$\begin{aligned}\operatorname{var}(\varepsilon_g + \varepsilon_{m|g}) &= \operatorname{var}(\varepsilon_g) + \operatorname{var}(\varepsilon_{m|g}) + 2\operatorname{cov}(\varepsilon_g, \varepsilon_{m|g}) \overset{\substack{\text{Prop1}\\\text{Hip3B}}}{=} \operatorname{var}(\varepsilon_g) + \operatorname{var}(\varepsilon_{g*}) \\ &= \operatorname{var}(\varepsilon_g + \varepsilon_{g*}) - 2\operatorname{cov}(\varepsilon_g, \varepsilon_{g*}) \\ \operatorname{cov}(\varepsilon_g, \varepsilon_{g*}) &= \frac{\pi^2}{12\mu^2} - \frac{\pi^2}{12\lambda_g^2} - \frac{\operatorname{var}(\varepsilon_g)}{2} \neq 0\end{aligned}$$

## 4. *CÁLCULO DEL VALOR DE INCLUSIÓN DE UN MODELO JERÁRQUICO DE DOS NIVELES*

El cociente de los parámetros de forma de las dos distribuciones tipo gumbel (también llamadas de valor extremo tipo I) involucradas en el cálculo de las probabilidades es precisamente el llamado *valor de inclusión* en la literatura, y su relación con las varianzas de las distribuciones supuestas para las partes aleatorias de las utilidades de las alternativas es como sigue,

$$\frac{\operatorname{var}\left(\varepsilon_{m|g}\right)}{\operatorname{var}\left(\varepsilon_g + \varepsilon_{g*}\right)} \overset{\text{Hip1;Hip2}}{=} \frac{\operatorname{var}\left(Gumbel(0,\lambda_g)\right)}{\operatorname{var}\left(Gumbel(0,\mu)\right)} = \frac{\pi^2/(6\lambda_g^2)}{\pi^2/(6\mu^2)} = \left(\mu/\lambda_g\right)^2$$

$$\Downarrow$$

$$\begin{aligned}\left(\mu/\lambda_g\right)^2 &= \frac{\operatorname{var}\left(\varepsilon_{m|g}\right)}{\operatorname{var}\left(\varepsilon_g\right) + \operatorname{var}\left(\varepsilon_{g*}\right) + 2\operatorname{cov}\left(\varepsilon_g, \varepsilon_{g*}\right)} \overset{\text{P1}}{=} \frac{\operatorname{var}\left(\varepsilon_{m|g}\right)}{\operatorname{var}\left(\varepsilon_g\right) + \operatorname{var}\left(\varepsilon_{m|g}\right) + 2\operatorname{cov}\left(\varepsilon_g, \varepsilon_{g*}\right)} \\ &= 1 - \frac{\operatorname{var}\left(\varepsilon_g\right) + 2\operatorname{cov}\left(\varepsilon_g, \varepsilon_{g*}\right)}{\operatorname{var}\left(\varepsilon_g\right) + \operatorname{var}\left(\varepsilon_{m|g}\right) + 2\operatorname{cov}\left(\varepsilon_g, \varepsilon_{g*}\right)} \overset{\text{Hip3B}}{=} 1 - \frac{\operatorname{var}\left(\varepsilon_g\right) + 2\operatorname{cov}\left(\varepsilon_g, \varepsilon_{g*}\right)}{\operatorname{var}\left(\varepsilon_g + \varepsilon_{m|g}\right) + 2\operatorname{cov}\left(\varepsilon_g, \varepsilon_{g*}\right)}\end{aligned}$$

Las varianzas de los términos aleatorios se pueden relacionar de la siguiente forma con las de las utilidades de las alternativas,

$$\text{var}\left(U_{g,m_1}\right)=\text{var}\left(\varepsilon_g+\varepsilon_{m_1|g}\right)\overset{\text{P1;Hip3B}}{=}\text{var}\left(\varepsilon_g+\varepsilon_{m_2|g}\right)=\text{var}\left(U_{g,m_2}\right)$$

$$\text{cov}\left(U_{g,m_1},U_{g,m_2}\right)=\text{cov}\left(\varepsilon_g+\varepsilon_{m_1|g},\varepsilon_g+\varepsilon_{m_2|g}\right)\overset{\text{Hip1;Hip3B}}{=}\text{var}\left(\varepsilon_g\right)$$

Y sustituyendo en la expresión anterior se tiene finalmente que,

$$1-\left(\mu/\lambda_g\right)^2\overset{\substack{\text{P1}\\\text{Hip1}\\\text{Hip3B}}}{=}=\frac{\text{cov}\left(U_{g,m_1},U_{g,m_2}\right)+2\,\text{cov}\left(\varepsilon_g,\varepsilon_{g*}\right)}{\left(\text{var}\left(U_{g,m_1}\right)\text{var}\left(U_{g,m_2}\right)\right)^{1/2}+2\,\text{cov}\left(\varepsilon_g,\varepsilon_{g*}\right)}=\frac{A\,corr\left(U_{g,m_1},U_{g,m_2}\right)+2\,\text{cov}\left(\varepsilon_g,\varepsilon_{g*}\right)}{A+2\,\text{cov}\left(\varepsilon_g,\varepsilon_{g*}\right)}$$

Donde se han de cumplir las siguientes restricciones para los términos involucrados,

$$\left.\begin{aligned}A=\text{var}\left(U_{g,m}\right)&\overset{\text{Hip3B}}{=}=\text{var}\left(\varepsilon_g\right)+\frac{\pi^2}{6\lambda_g^2}\geq 0\\ \frac{\pi^2}{6\mu^2}&=\text{var}\left(\varepsilon_g+\varepsilon_{g*}\right)\\ &=\text{var}\left(\varepsilon_g\right)+\text{var}\left(\varepsilon_{g*}\right)+2\,\text{cov}\left(\varepsilon_g,\varepsilon_{g*}\right)\overset{\text{P1}}{=}\\ &=A+2\,\text{cov}\left(\varepsilon_g,\varepsilon_{g*}\right)\geq 0\end{aligned}\right\}\Rightarrow A=\frac{\pi^2}{6\mu^2}-2\,\text{cov}\left(\varepsilon_g,\varepsilon_{g*}\right)\geq 0 \;\Downarrow\; \frac{\text{cov}\left(\varepsilon_g,\varepsilon_{g*}\right)}{\pi^2/(12\mu^2)}\leq 1$$

Nótese cómo no se permite cualquier covarianza positiva entre los componentes aleatorios $\varepsilon_g,\varepsilon_{g*}$, pero sí negativa. La no existencia de limitación conceptual alguna para suponer valores negativos de esta covarianza es la que provoca la relajación del rango de admisibilidad para el valor de inclusión.

Existe otra varianza de la que no hemos especificado su valor pero que debe ser por consistencia no negativa,

$$\text{var}\left(\varepsilon_g\right)=\frac{\pi^2}{6\mu^2}-\frac{\pi^2}{6\lambda_g^2}-2\,\text{cov}\left(\varepsilon_g,\varepsilon_{g*}\right)\geq 0$$

Así por tanto, la condición necesaria (y suficiente) para que todas las varianzas de los términos aleatorios sean positivas es hacer que,

$$\text{cov}\left(\varepsilon_g,\varepsilon_{g*}\right)\leq\frac{\pi^2}{12\mu^2}-\frac{\pi^2}{12\lambda_g^2}\;\Rightarrow\;\boxed{\left(\mu/\lambda_g\right)^2\leq 1-\frac{\text{cov}\left(\varepsilon_g,\varepsilon_{g*}\right)}{\pi^2/\left(12\mu^2\right)}}\quad\text{(R-1)}$$

Ahora puede verse como la asunción 3A hecha implícitamente en BA&L estableciendo la independencia entre $\varepsilon_g$ y $\varepsilon_{g*}$, no necesaria para la derivación de las probabilidades, conduce inexorablemente a que $0\leq\mu/\lambda_g\leq 1$.

Reordenando algo más la expresión anterior para el valor de inclusión se obtiene la expresión para la correlación entre las utilidades de dos alternativas que pertenecen a la misma rama *g* de la estructura jerárquica,

$$\left.\begin{aligned} corr\left(U_{g,m_1},U_{g,m_2}\right) &= \frac{\left[1-\left(\mu/\lambda_g\right)^2\right]\pi^2/(6\mu^2)-2\operatorname{cov}\left(\varepsilon_g,\varepsilon_{g^*}\right)}{\pi^2/(6\mu^2)-2\operatorname{cov}\left(\varepsilon_g,\varepsilon_{g^*}\right)} \\ &= 1-\frac{\left(\mu/\lambda_g\right)^2\pi^2/(6\mu^2)}{\pi^2/(6\mu^2)-2\operatorname{cov}\left(\varepsilon_g,\varepsilon_{g^*}\right)} \end{aligned}\right\} \Rightarrow 0 \le \frac{\left(\mu/\lambda_g\right)^2\pi^2/(6\mu^2)}{\pi^2/(6\mu^2)-2\operatorname{cov}\left(\varepsilon_g,\varepsilon_{g^*}\right)} \le 2$$

Debido a que el denominador de la igualdad es siempre positivo debido a cálculos anteriores, se tiene que la condición que ha de cumplir el valor de inclusión para hacer que la correlación esté en el intervalo $(-1,1)$ es precisamente que

$$\left(\mu/\lambda_g\right)^2 \le 2\left(1-\frac{\operatorname{cov}\left(\varepsilon_g,\varepsilon_{g^*}\right)}{\pi^2/\left(12\mu^2\right)}\right) \text{ (R-2),}$$

la cual siempre se cumple mientras que R-1 se mantenga; por tanto, nuestra relajación del valor de inclusión sigue ofreciendo resultados plausibles para la correlación entre alternativas.

Por otro lado, la correlación entre alternativas que pertenecen a ramas distintas es nula, si bien la independencia de la HIP1 ha de ser extendida a partes aleatorias condicionadas que pertenecen a distintas ramas y asumir independencia entre las partes aleatorias de la utilidad a nivel de rama,

(**HIPÓTESIS 3C**: $\operatorname{cov}(\varepsilon_g,\varepsilon_h)=0\ ,\forall g,h\in G, g\neq h$)

$$\operatorname{cov}\left(U_{g_1,m},U_{g_2,n}\right)=\operatorname{cov}\left(\varepsilon_{g_1}+\varepsilon_{m|g_1},\varepsilon_{g_2}+\varepsilon_{n|g_2}\right)\overset{\text{Hip3B}}{=}\operatorname{cov}(\varepsilon_{g_1},\varepsilon_{g_2})+\operatorname{cov}\left(\varepsilon_{m|g_1},\varepsilon_{n|g_2}\right)\overset{\text{Hip1;Hip3C}}{=}0$$

Para ilustrar los resultados a los que se ha llegado en esta sección se realiza una representación gráfica del valor permitido para el valor de inclusión en un modelo jerárquico de dos niveles en función de la covarianza de los términos $\varepsilon_g$ y $\varepsilon_{g*}$.

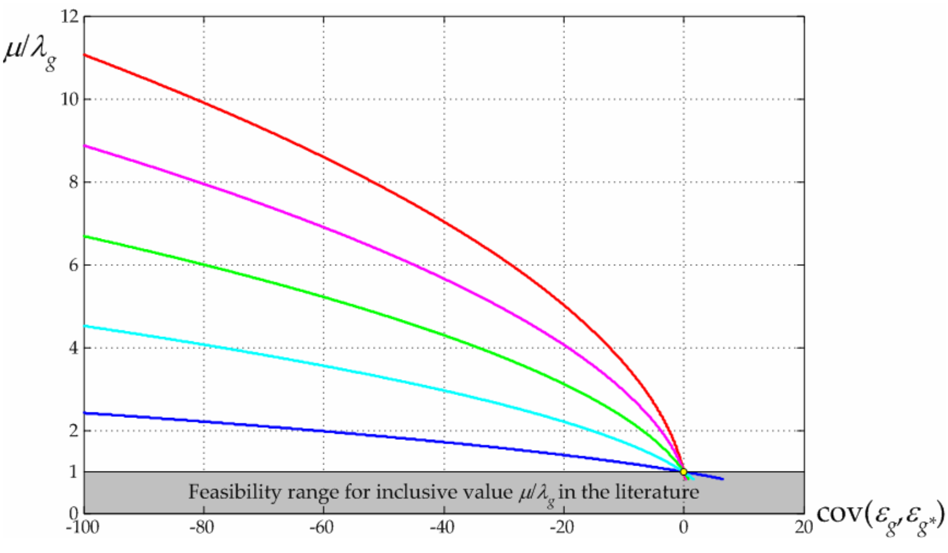


**Figura 2** Rango admisible el valor de inclusión en función de la relajación de la hipótesis 3A $\operatorname{cov}\left(\varepsilon_g,\varepsilon_{g^*}\right)\neq 0$

El límite superior para la covarianza es $\pi^2/(12\mu^2)$, pues existe esta limitación básica para hacer que todas las varianzas involucradas en el modelo sean no negativas; obsérvese simplemente que valores mayores de la covarianza ocasionan valores de $(\mu/\lambda_g)^2$ negativos. Una vez especificada una covarianza y el valor de $\mu$, el rango de admisibilidad del cociente $\mu/\lambda_g$ es el que va desde 0 hasta la curva correspondiente.

### *4.1. Interpretación de covarianzas negativas en la Hipótesis 3A*

En este apartado se aborda la explicación del significado de una covarianza negativa entre los términos $\varepsilon_g$ y $\varepsilon_{g*}$. La distribución de $\varepsilon_g$ es la más neutra al razonamiento que se ha seguido, en tanto que no se ha necesitado de la especificación del valor concreto de su varianza.

En el proceso de decisión entre ramas de la estructura que da lugar al cálculo de las probabilidades $P_1,\ldots,P_g,\ldots,P_G$ ($\sum_{g=1}^{G} P_g = 1$) participan como parte aleatoria de las utilidades $(\varepsilon_1+\varepsilon_{1*}),\ldots,(\varepsilon_g+\varepsilon_{g*}),\ldots,(\varepsilon_G+\varepsilon_{G*})$, conteniendo $\varepsilon_g$ la parte aleatoria asociada exclusivamente a la rama *g*, y $\varepsilon_{g*}$ la asociada a las alternativas elementales que componen o cuelgan de esta rama y entre las que eligen los individuos. La misma definición de $\varepsilon_{g*}$ establece su correlación con las i.i.d variables aleatorias $\varepsilon_{1|g},\ldots,\varepsilon_{m|g},\ldots,\varepsilon_{M_g|g}$.

Una covarianza negativa entre $\varepsilon_g$ y $\varepsilon_{g*}$ implica que mientras menos cuenten los atributos en explicar la utilidad asociada en exclusiva a nivel de rama (mayor importancia de $\varepsilon_g$), menor es la aleatoriedad que mide $\varepsilon_{g*}$, la cual esta directa y positivamente relacionada con las i.i.d variables aleatorias $\varepsilon_{1|g},\ldots,\varepsilon_{m|g},\ldots,\varepsilon_{M_g|g}$; por tanto, mientras mayor importancia tenga $\varepsilon_g$ menor habrá de tener al menos alguna de las $M_g$ variables aleatorias elementales.

En términos del ejemplo sobre transporte interurbano de la Figura 1 esta covarianza negativa se puede explicar mediante los dos resultados siguientes: primero, se puede considerar que los términos aleatorios de la rama *transporte público* se deben por ejemplo (y principalmente) a no incluir explícitamente como un atributo del modelo la mala imagen que este tipo de transporte tiene en la ciudad, no apareciendo por tanto en la parte determinista de la utilidad y sí en la aleatoria, y segundo, se puede considerar que las partes aleatorias específicas de las alternativas *autobús* y *tren* se deben principalmente a no incluir la puntualidad de cada medio de transporte en la parte determinista. Mientras peor es la imagen menor se espera que sea $\varepsilon_{TPublico}$, y mientras mayor es la puntualidad mayor se espera que sea $\varepsilon_{autobus|TPublico}$ y $\varepsilon_{tren|TPublico}$, y con ellos mayor se espera que sea $\varepsilon_{TPublico*}$.

La existencia de $\mathrm{cov}(\varepsilon_{TPublico},\varepsilon_{TPublico*})$ negativa implica que una peor imagen del sistema de transporte público está relacionada con una percepción de más puntualidad de los autobuses y/o trenes. Pueden existir situaciones en las que esta correlación negativa tenga sentido (si bien no serán las más comunes en general), por ejemplo, inmediatamente después de que se anuncie un plan de mejora del servicio de trenes, la imagen del transporte público sigue siendo mala, y los autobuses siguen afectados por una mala puntualidad, pero no así el uso del tren, cuya $\varepsilon_{tren|TPublico}$ (y también por tanto $\varepsilon_{TPublico*}$) puede aumentar porque cuando se realiza la encuesta ya se notan los primeros cambios en el sistema de tren; cuanto peor es la imagen del sistema de transporte más se espera que impacte en las decisiones de los individuos la puntualidad del tren.

Si no se considera lógica en ninguna situación posible una explicación como la anterior, habrá de suponerse la Hip3A ($\mathrm{cov}(\varepsilon_g,\varepsilon_{g*})=0$) y con ello queda el rango de admisibilidad unitario consensuado en la literatura.

## 5. *RANGO DEL VALOR DE INCLUSIÓN Y CONSISTENCIA CON RUM*

En esta sección se abunda en los argumentos expuestos en la literatura para justificar la necesidad de rangos de admisibilidad unitarios para el valor de inclusión. En Louviere et al. (2000, p.146) se afirma que $\mu/\lambda_g \leq 1$, $\forall \lambda_g \in G$ es condición suficiente para que un modelo jerárquico sea consistente con la teoría RUM. Los resultados de este trabajo se han derivado de la aplicación sucesiva de los postulados RUM para calcular $P_g$ y $P_{m|g}$, y por ello, aun cuando $\mathrm{cov}(\varepsilon_g, \varepsilon_{g*}) \neq 0$ y $\mu/\lambda_g > 1$ se ha demostrado que se tienen modelos consistentes con RUM, por lo que en efecto, la condición $\mu/\lambda_g \leq 1$, $\forall \lambda_g \in G$ es suficiente pero NO necesaria.

Por otro lado, en Hensher & Greene (2002, p.13), se advierte de que un sistema de probabilidades que sea consistente con RUM no puede verse afectado por la adición de una constante a las utilidades de todas sus alternativas. Esta propiedad se cumple con el sistema de probabilidades presentado en este trabajo para cualquier valor de $\mu/\lambda_g$, pues la restricción coincide precisamente con el postulado RUM-b utilizado en el cálculo de las expresiones de las probabilidades. Así por tanto, tampoco se puede imponer un rango de admisibilidad unitario para el valor de inclusión bajo este argumento.

El último de los elementos analizado que liga los valores admisibles del valor de inclusión con la consistencia del modelo en estudio con la teoría RUM es las elasticidades de las probabilidades de las alternativas. En la literatura se argumenta que si $\mu/\lambda_g > 1$, entonces la adición de un valor constante positivo a una de las alternativas provoca una disminución de la probabilidad de dicha alternativa, al considerar este hecho contrario al desempeño que se espera del modelo, se restringe el valor de inclusión al intervalo 0,1. Esta situación de menores probabilidades para mayores utilidades de una alternativa es contraria a priori a los postulados RUM que se han empleado en este trabajo, pues

$$\Pr\left(U_{gm} + K = \max_{\forall q \in G, \forall h \in M_q} \left\{U_{qh}\right\}\right) \geq \Pr\left(U_{gm} = \max_{\forall q \in G, \forall h \in M_q} \left\{U_{qh}\right\}\right) \ , \ \forall K \geq 0$$

y además,

$$\frac{\partial P_{gm}}{\partial V_{gm}} = \lambda_g \left[\left(1 - P_{m|g}\right) + \mu/\lambda_g \left(1 - P_g\right) P_{m|g}\right] P_{gm} \geq 0 \ \ \forall \mu/\lambda_g \geq 0$$

Queda demostrado por tanto que la relajación realizada para el rango de admisibilidad del valor de inclusión en función de los valores de $\mathrm{cov}(\varepsilon_g, \varepsilon_{g*})$ es consistente con un comportamiento adecuado de las probabilidades, al menos en lo que respecta a las elasticidades directas.

El resto de elasticidades (las cruzadas o no directas) del modelo jerárquico de dos niveles se comportan de la siguiente forma,

$$\frac{\partial P_{gm}}{\partial V_{gn}} = \lambda_g \left[\left(\mu/\lambda_g - 1\right) P_{n|g} - \mu/\lambda_g \, P_{gn}\right] P_{gm}$$

$$\frac{\partial P_{gm}}{\partial V_{hn}} = -\lambda_h \, \mu/\lambda_h \, P_{hn} P_{gm}$$

El comportamiento esperado es que estas elasticidades cruzadas sean negativas, es decir que si aumenta la utilidad de una alternativa, disminuya la probabilidad de elección de cada una de las restantes; la elasticidad cruzada entre alternativas en diferentes ramas es negativa para cualquier valor de $\mu/\lambda_h$, mientras que la cruzada entre alternativas bajo la misma rama es negativa sólo si $\mu/\lambda_g \leq 1/(1-P_g)$. Sin embargo, imponer que esta elasticidad cruzada en la misma rama sea menor que la elasticidad directa es menos restrictivo, pues se tiene que,

$$\frac{\partial P_{gm}}{\partial V_{gm}} \geq \frac{\partial P_{gm}}{\partial V_{gn}} \; iff \quad 1 + C \cdot P_{m|g} \geq C \cdot P_{n|g} \Rightarrow \mu/\lambda_g \leq \frac{2}{1-P_g}$$

$$C = \mu/\lambda_g\left(1-P_g\right) - 1$$

Si se tienen valores para el valor de inclusión mayores a $2/(1-P_g)$ se tendrá la situación de que un incremento en la utilidad de una alternativa provoca un aumento en la probabilidad de que ésta resulte elegida menor al aumento de que otra alternativa de la misma rama resulte elegida; la explicación reside en la correlación negativa entre $\varepsilon_g$ y $\varepsilon_{g*}$. Es importante recalcar que los postulados RUM son consistentes con esta situación. Si se prefiere descartar la situación en la que ocurre esta mayor elasticidad cruzada que la directa, se ha de restringir el valor de la covarianza de la siguiente forma,

$$\frac{\pi^2}{12\mu^2}\left(1 - \frac{4}{\left(1-P_g\right)^2}\right) \leq \operatorname{cov}\left(\varepsilon_g, \varepsilon_{g*}\right) \leq \frac{\pi^2}{12\mu^2} \quad \Rightarrow \quad \frac{2}{1-P_g} \geq \left(1 - \frac{\operatorname{cov}\left(\varepsilon_g, \varepsilon_{g*}\right)}{\pi^2/(12\mu^2)}\right)^{1/2} \geq 0$$

$$\Downarrow$$

$$\mu/\lambda_g \leq \left(1 - \frac{\operatorname{cov}\left(\varepsilon_g, \varepsilon_{g*}\right)}{\pi^2/\left(12\mu^2\right)}\right)^{1/2} \leq \frac{2}{1-P_g}$$

En la Figura 3 se muestran los límites máximo y mínimo permitidos para la covarianza entre $\varepsilon_g$ y $\varepsilon_{g*}$ en función de la probabilidad marginal $P_g$, y en la Figura 4 las relajaciones que se permiten para el valor de inclusión asociado a esa rama *g*.

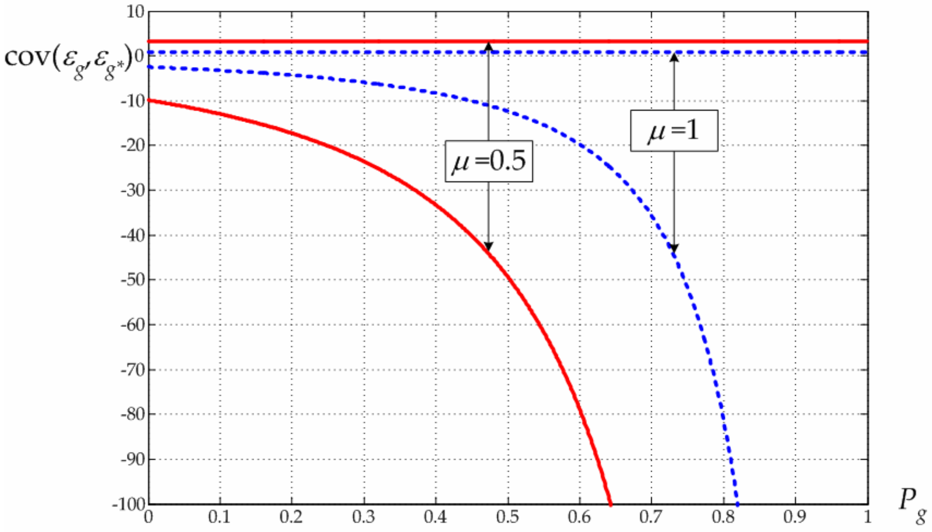


**Figura 3** Rango admisible para las covarianzas

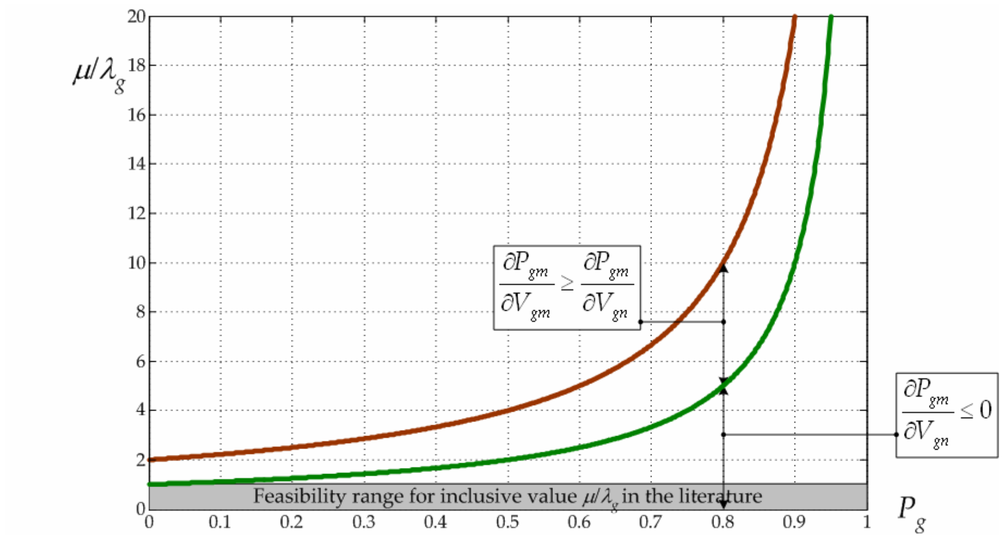


**Figura 4** Rango admisible para el valor de inclusión y efecto en las elasticidades

## 6. *CONCLUSIONES*

En este trabajo se ha relajado una de las hipótesis que se asumen implícitamente en la literatura especializada (Ben-Akiva&Lerman, 1985) cuando se derivan las expresiones de la probabilidad de un modelo de elección discreta jerárquico de dos niveles. Esta relajación no es contraria a la teoría RUM que sirve como base de los modelos de elección discreta y se traduce en permitir valores de inclusión mayores que uno en estos modelos. Mediante este trabajo se ofrece una explicación de las implicaciones que estos valores mayores que uno tienen sobre el modelo y de por qué son consistentes con la teoría RUM.

La asunción de los postulados RUM para derivar los dos modelos logit sucesivos que dan lugar a la expresión de la probabilidad de un modelo logit jerárquico de dos niveles no conlleva la necesidad de asumir la Hip.3A de $\text{cov}(\varepsilon_g, \varepsilon_{g^*}) = 0$ hecha en Ben-Akiva & Lerman (1985), por lo que el rango admisible con las distribuciones aleatorias involucradas en este proceso de asunción sucesiva de hipótesis es,

$$0 \leq \mu/\lambda_g \leq \left(1 - 12\mu^2 \text{cov}\left(\varepsilon_g, \varepsilon_{g^*}\right) \big/ \pi^2\right)^{1/2}$$

Las covarianzas en cada rama $g$ están limitadas superiormente por $\pi^2 / (12\mu^2)$ para que los términos aleatorios de las utilidades estén correctamente definidos, y limitadas inferiormente por $\pi^2 \left(1 - 4(1 - P_g)^2\right) \big/ (12\mu^2)$ para que las elasticidades cruzadas de alternativas en la rama $g$ no sean mayores que las elasticidades directas.

## 7. *REFERENCIAS*


- BEN-AKIVA, LERMAN (1985): *Discrete Choice Analysis.* MIT Press
- BLOCK, MARSCHAK (1960) Random orderings and stochastic theories of responses. In Marschak, J. (1974) *Economic Information, Decision and Prediction: Selected Essays* (Volume 1)
- DALY, ZACHARY (1976) Improved multiple choice models. PTRC, London
- HENSHER, GREENE (2002) Specification and Estimation of Nested Logit Models, *Transp. Res. B*, 36,1-18
- LOUVIERE, HENSHER, SWAIT (2000): *Stated Choice Methods: Analysis and Application*. Cambridge.
- MCFADDEN (1978) Modelling the choice of residential location. In Karlqvist, A. et al. (eds) *Spatial Interaction Theory and Residential Location*. North-Holland, Amsterdam
- WILLIAMS (1977) On the formation of travel demand models and economic evaluation measures of user benefit. *Environment and Planning A*, 9 (3), pp285-344