# 4

# GEV

## 4.1 Introducción

El modelo logit estándar exhibe independencia de alternativas irrelevantes (IIA), lo que implica la sustitución proporcional entre alternativas. Como ya comentamos en el capítulo 3, esta propiedad puede ser vista como una restricción impuesta por el modelo o como el resultado natural de un modelo bien especificado, capaz de capturar en la utilidad representativa todas las fuentes de correlación entre alternativas, por lo que sólo queda ruido blanco como inobservado. A menudo, el investigador es incapaz de capturar todas las fuentes de correlación de forma explícita, por lo que las partes no observadas de utilidad están correlacionadas y la IIA no se sostiene. En estos casos se necesita un modelo más general que el logit estándar.

Los modelos generalizados de valor extremo (GEV) constituyen una amplia clase de modelos que exhiben patrones de sustitución variados. El atributo unificador de estos modelos es que las partes no observadas de la utilidad de todas las alternativas se distribuyen conjuntamente de acuerdo a una distribución generalizada de valor extremo. Esta distribución permite correlaciones entre alternativas y, como su propio nombre indica, es una generalización de la distribución de valor extremo univariante que se utiliza en los modelos logit estándar. Cuando todas las correlaciones son cero, la distribución GEV se convierte en el producto de distribuciones de valor extremo independientes y el modelo GEV se convierte en el logit estándar. Por consiguiente, esta clase de modelos incluye el logit simple pero también una variedad de otros modelos. Los test de hipótesis sobre las correlaciones dentro de un modelo GEV se pueden emplear para examinar si las correlaciones son cero, lo que es equivalente a probar si logit estándar proporciona una representación precisa de los patrones de sustitución.

El miembro más ampliamente utilizado de la familia GEV es el logit jerárquico o anidado. Este modelo ha sido aplicado por muchos investigadores en una gran variedad de situaciones, incluyendo la energía, el transporte, la vivienda, las telecomunicaciones y una multitud de otros campos; véase por ejemplo, Ben- Akiva (1973), Train (1986, capítulo 8), Train et al. (1987a), Forinash y Koppelman (1993) y Lee (1999). Su forma funcional es simple en comparación con otros tipos de modelos GEV y proporciona un conjunto flexible de posibles patrones de sustitución. Las secciones 4.2 y 4.3 describen la especificación y estimación de modelos logit jerárquicos. Esta descripción es útil en sí misma, dada la prominencia adquirida por los modelos logit jerárquicos, pero también como base para la comprensión de modelos GEV más complejos. En la Sección 4.4

pasamos a estudiar otros modelos GEV puestos en práctica por algunos investigadores, con especial énfasis en dos de los modelos más prometedores, a saber, el logit combinacional emparejado (*paired combinatorial logit*, PCL) y el logit jerárquico generalizado (*generalized nested logit*, GNL). La sección final del capítulo describe la clase completa de modelos GEV y cómo se generan las nuevas especificaciones de esta clase.

Sólo una pequeña parte de los posibles modelos GEV se han implementado alguna vez. Esto significa que aún no han sido plenamente explotadas todas las capacidades de esta clase de modelos y que nuevas investigaciones en esta área tienen el potencial de encontrar modelos aún más poderosos que los que ya se utilizan actualmente. Un ejemplo de este potencial lo proporciona Karlstrom (2001), que especificó un modelo GEV de una forma nunca empleada anteriormente y que encontró que se ajustaba a sus datos mejor que otros tipos de modelos GEV especificados previamente. Los modelos GEV tienen la ventaja de que las probabilidades de elección suelen tener una forma cerrada, de modo que pueden ser estimadas sin recurrir a la simulación. Sólo por esta razón, los modelos GEV seguirán siendo fuente de nuevas y potentes especificaciones para satisfacer las necesidades de los investigadores.

## 4.2 Logit jerárquico

### 4.2.1 Patrones de sustitución

Un modelo logit jerárquico es apropiado cuando el conjunto de alternativas a las que se enfrenta un decisor puede dividirse en subconjuntos, llamados nidos (*nests*), de tal manera que las siguientes propiedades se cumplen:

1. Para cualesquiera dos alternativas que están en el *mismo* nido, el ratio de probabilidades es independiente de los atributos o de la existencia de cualesquiera otras alternativas. Es decir, se verifica la IIA dentro de cada nido.
2. Para cualesquiera dos alternativas en *diferentes* nidos, el ratio de probabilidades puede depender de los atributos de otras alternativas en los dos nidos. La IIA no se sostiene, en general, entre alternativas en diferentes nidos.
3. 

Un ejemplo puede explicar mejor si un conjunto de alternativas puede ser objeto de una partición de este tipo. Supongamos que el conjunto de alternativas disponibles para un trabajador para ir a su lugar de trabajo consisten en conducir un automóvil en solitario, compartir un automóvil, usar el autobús y usar el tren. Si se elimina cualquiera de estas alternativas, las probabilidades de elección del resto de alternativas se incrementaría (por ejemplo, si el automóvil del trabajador está siendo reparado, por lo que no puede conducir al trabajo por sí mismo, las probabilidades de compartir automóvil, ir en autobús e ir en ferrocarril aumentarían). La pregunta relevante al dividir estas alternativas en grupos es: si quitase una alternativa, ¿en qué proporción aumentaría la probabilidad de cada una de las restantes alternativas? Supongamos que los cambios en las probabilidades se producen como se indica en la tabla 4.1.

*Tabla 4.1. Ejemplo de cumplimiento de la IIA dentro de los nidos de alternativas: cambio en las probabilidades cuando se suprime una alternativa*

| | Probabilidad | | | | |
|---|---|---|---|---|---|
| | | Con alternativa suprimida | | | |
| Alternativa | Original | Automóvil | Automóvil compartido | Autobús | Tren |
| Automóvil | .40 | — | .45 (+12.5%) | .52 (+30%) | .48 (+20%) |
| Automóvil compartido | .10 | .20 (+100%) | — | .13 (+30%) | .12 (+20%) |
| Autobús | .30 | .48 (+60%) | .33 (+10%) | — | .40 (+33%) |
| Tren | .20 | .32 (+60%) | .22 (+10%) | .35 (+70%) | — |

Obsérvese que las probabilidades del autobús y el tren siempre aumentan en la misma proporción cada vez que se elimina una de las otras alternativas. Por lo tanto, la IIA se mantiene entre estas dos alternativas. Pongamos estas dos alternativas en un nido y llamémosle "transporte público". Del mismo modo, la probabilidad del automóvil (en solitario) y el automóvil compartido crecen en la misma proporción cada vez que una de las otras alternativas se elimina. La IIA se mantiene entre estas dos alternativas, por lo que las ponemos en un nido llamado "automóviles". La IIA no se mantiene entre una alternativa del nido “automóviles” y una del nido “transporte público”. Por ejemplo, cuando se elimina la alternativa automóvil, la probabilidad de compartir automóvil se eleva proporcionalmente más que la probabilidad del autobús o del tren. Con nuestros dos nidos, podemos enunciar las pautas de sustitución de forma resumida como: la IIA se mantiene dentro de cada nido, pero no entre nidos. Un modelo logit jerárquico con las dos alternativas de automóvil en un nido y las dos alternativas de transporte público en otro nido, es apropiado para representar esta situación.

Una manera conveniente de representar los patrones de sustitución es con un diagrama en árbol. En el árbol, cada rama representa un subconjunto de alternativas dentro de las cuales la IIA se mantiene y cada hoja de cada rama indica una alternativa. Por ejemplo, el diagrama en árbol de la elección que el trabajador hace del medio de transporte que acabamos de describir se muestra en la figura 4.1. El árbol (de arriba a abajo) se compone de dos ramas, denominadas "automóviles" y "transporte público", para los dos subconjuntos de alternativas, y cada una de las ramas contiene dos sub-ramas para las dos alternativas dentro del subconjunto. Existe sustitución proporcional entre sub-ramas pero no entre ramas.

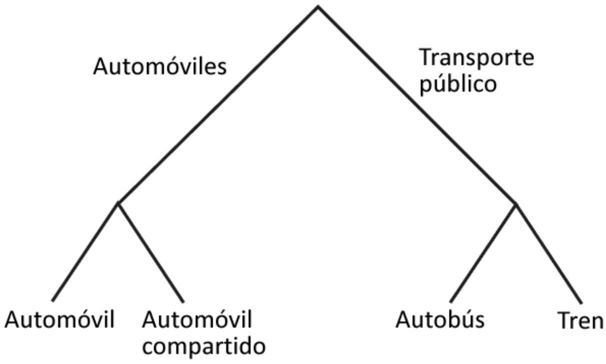


*Figura 4.1. Diagrama en árbol para la elección de medio de transporte.*

### 4.2.2 Probabilidades de elección

Daly y Zachary (1978), McFadden (1978) y Williams (1977) mostraron, de forma independiente y utilizando diferentes pruebas, que el modelo logit jerárquico es consistente con la maximización de la utilidad. Supongamos que dividimos el conjunto de alternativas $j$ en $K$ subconjuntos no solapados $B_1, B_2, \dots, B_K$ y los llamamos nidos. La utilidad que una persona $n$ obtiene de la alternativa $j$ en el nido

$B_K$ se denota, como de costumbre, como $U_{nj} = V_{nj} + \varepsilon_{nj}$, donde $V_{nj}$ es observada por el investigador y $\varepsilon_{nj}$ es una variable aleatoria cuyo valor no es observado por el investigador. El modelo logit jerárquico se obtiene asumiendo que el vector de utilidad no observada $\varepsilon_n = \langle \varepsilon_{n1}, \dots, \varepsilon_{nJ} \rangle$, tiene distribución acumulativa

$$(4.1) \qquad exp\left(-\sum_{k=1}^{K}\left(\sum_{j\in B_k} e^{-\varepsilon_{nj}/\lambda_k}\right)^{\lambda_k}\right).$$

Esta distribución es un tipo de distribución GEV. Es una generalización de la distribución que da lugar al modelo logit. Para logit, cada $\varepsilon_{nj}$ es independiente con una distribución univariante valor extremo. Para esta GEV, la distribución marginal de cada $\varepsilon_{nj}$ es de tipo valor extremo univariante. Sin embargo, los $\varepsilon_{nj}$s se correlacionan dentro de los nidos. Para cualesquiera dos alternativas $j$ y $m$ en el nido $B_K$, $\varepsilon_{nj}$ se correlaciona con $\varepsilon_{nm}$. Para cualesquiera dos alternativas en diferentes nidos, la parte no observada de la utilidad sigue estando no correlacionada: $Cov(\varepsilon_{nj}, \varepsilon_{nm}) = 0$ para cualquier $j \in B_K$ y $m \in B_l$ con $l \neq k$.

El parámetro $\lambda_k$ es una medida del grado de independencia de la utilidad no observada entre las alternativas dentro del nido $k$. Un valor más alto de $\lambda_k$ significa mayor independencia y menor correlación. El estadístico $1 - \lambda_k$ es una medida de correlación, en el sentido de que a medida que $\lambda_k$ aumenta, indicando una menor correlación, este estadístico cae. Como señala McFadden (1978), la correlación en realidad es más compleja que $1 - \lambda_k$, pero $1 - \lambda_k$ se puede utilizar como una indicación de la correlación. Un valor de $\lambda_k = 1$ indica independencia completa dentro del nido $k$, es decir, que no hay correlación. Cuándo $\lambda_k = 1$ para todo $k$, representando independencia entre todas las alternativas en todos los nidos, la distribución GEV se convierte en el producto de términos tipo valor extremo independientes, cuya distribución está expresada en (3.2). En este caso, el modelo logit jerárquico se reduce al modelo logit estándar.

Como han mostrado los autores citados anteriormente, la distribución de los componentes no observados de la utilidad da lugar a la siguiente probabilidad de elección para la alternativa $i \in B_k$:

$$(4.2) \qquad P_{ni} = \frac{e^{V_{ni}/\lambda_k}\left(\sum_{j\in B_k} e^{V_{nj}/\lambda_k}\right)^{\lambda_k - 1}}{\sum_{l=1}^{K}\left(\sum_{j\in B_l} e^{V_{nj}/\lambda_l}\right)^{\lambda_l}}.$$

Podemos utilizar esta fórmula para demostrar que la IIA se mantiene dentro de cada subconjunto de alternativas, pero no entre subconjuntos. Considere las alternativas $i \in B_k$ y $m \in B_l$. Dado que el denominador de (4.2) es el mismo para todas las alternativas, el ratio de probabilidades es el ratio de numeradores:

$$\frac{P_{ni}}{P_{nm}} = \frac{e^{V_{ni}/\lambda_k}\left(\sum_{j\in B_k} e^{V_{nj}/\lambda_k}\right)^{\lambda_k - 1}}{e^{V_{nm}/\lambda_l}\left(\sum_{j\in B_l} e^{V_{nj}/\lambda_l}\right)^{\lambda_l - 1}}.$$

Si $k = l$ (es decir, $i$ y $m$ están en el mismo nido) entonces los factores que aparecen entre paréntesis se anulan y obtenemos

$$\frac{P_{ni}}{P_{nm}} = \frac{e^{V_{ni}/\lambda_k}}{e^{V_{nm}/\lambda_l}}.$$

Este ratio es independiente de todas las otras alternativas. Para $k \neq l$ (es decir, $i$ y $m$ están en diferentes nidos), los factores que aparecen entre paréntesis no se anulan. La relación de probabilidades depende de los atributos de todas las alternativas presentes en los nidos que contienen $i$ y $m$. Obsérvese, sin embargo, que la relación no depende de los atributos de las alternativas de los nidos que no contienen $i$ y $m$. Por lo tanto, un tipo de IIA se mantiene incluso para alternativas en diferentes nidos. Este tipo de IIA puede describirse en términos generales como "independencia de los nidos irrelevantes" o IIN. Con un modelo logit jerárquico, se cumple la IIA sobre alternativas dentro de cada nido y la IIN sobre alternativas de diferentes nidos. Esta propiedad de los modelos logit jerárquicos se refuerza en la siguiente sección en la que descomponemos la probabilidad de elección del logit jerárquico en dos probabilidades logit estándar.

Cuando $\lambda_k = 1$ para todo $k$ (y por tanto $1 - \lambda_k = 0$), indicando que no hay correlación entre los componentes no observados de la utilidad de alternativas dentro de un mismo nido, las probabilidades de elección se convierten simplemente en logit. El modelo logit jerárquico es una generalización del logit que permite un patrón particular de correlación en la utilidad no observada.

El parámetro $\lambda_k$ puede variar para diferentes nidos, lo que refleja distintas correlaciones entre los factores no observados dentro de cada nido. El investigador puede restringir los $\lambda_k$s para que sean el mismo para todos (o algunos) nidos, indicando que la correlación es la misma en cada uno de estos nidos. Es posible usar un test de hipótesis para determinar si aplicar restricciones sobre los valores de las $\lambda_k$s es razonable. Así, por ejemplo, hacer un test sobre la restricción $\lambda_k = 1\ \forall k$ es equivalente a probar si el modelo logit estándar es una especificación razonable para el problema de elección frente al modelo logit jerárquico más general. Estos test se llevan a cabo más fácilmente con el estadístico de ratio de verosimilitud descrito en la Sección 3.8.2.

El valor de $\lambda_k$ debe estar dentro de un rango determinado para que el modelo sea consistente con el comportamiento de maximización de la utilidad. Si $\lambda_k\ \forall k$ está entre cero y uno, el modelo es consistente con la maximización de la utilidad para todos los posibles valores de las variables explicativas. Para $\lambda_k$ mayor que uno, el modelo es coherente con el comportamiento maximizador de la utilidad de algún rango de las variables explicativas, pero no para todos los posibles valores. Kling y Herriges (1995) y Herriges y Kling (1996) proporcionan pruebas de la consistencia del logit jerárquico con la maximización de la utilidad cuando $\lambda_k > 1$; y Train et al. (1987a) y Lee (1999) ofrecen ejemplos de modelos para los cuales $\lambda_k > 1$. Un valor negativo de $\lambda_k$ es inconsistente con la maximización de la utilidad e implica que la mejora de los atributos de una alternativa (como la reducción de su precio) podría disminuir la probabilidad de que la alternativa fuese elegida. Con $\lambda_k$ positivo, el modelo logit jerárquico se aproxima al modelo de "eliminación por aspectos" (*elimination by aspects*) de Tversky (1972) a medida que $\lambda_k \to 0$.

En la notación que hemos estado utilizando, cada $\lambda_k$ es un parámetro fijo, lo que implica que todos los decisores tienen las mismas correlaciones entre los factores no observados. En realidad, las correlaciones podrían diferir entre decisores en base a sus características observadas. Para dar cabida a esta posibilidad, cada $\lambda_k$ puede ser especificada como una función paramétrica de datos demográficos observados u otras variables, siempre y cuando la función mantenga un valor positivo. Por ejemplo, Bhat (1997a) especifica $\lambda = \exp(\alpha z_n)$, donde $z_n$ es un vector de características del decisor $\mathrm{n}$ y $\alpha$ es un vector de parámetros a estimar, junto con los parámetros que entran en la utilidad representativa. La transformación exponencial asegura que $\lambda$ es positivo.

### 4.2.3 La descomposición en dos logits

La expresión (4.2) no es muy esclarecedora como fórmula. Sin embargo, las probabilidades de elección pueden ser expresadas de una forma alternativa bastante simple y fácilmente interpretable. Sin pérdida de generalidad, el componente observado de utilidad se puede descomponer en dos partes: (1) una parte etiquetada como $W$ constante para todas las alternativas dentro de un nido y (2) una parte etiquetada como $Y$ que varía entre alternativas dentro de un nido. La utilidad se rescribe como

$$(4.3) \qquad U_{nj} = W_{nk} + Y_{nj} + \varepsilon_{nj},$$

para $j \in B_k$, donde:

- $W_{nk}$ sólo depende de variables que describen el nido k. Estas variables difieren entre los nidos pero no entre las alternativas de cada nido.
- $Y_{nj}$ depende de variables que describen la alternativa j. Estas variables varían entre alternativas dentro del nido k.

Observe que esta descomposición es totalmente general, dado que para cualquier $W_{nk}$, $Y_{nj}$ se define como $V_{nj} - W_{nk}$.

Con esta descomposición de la utilidad, la probabilidad logit jerárquica puede escribirse como el producto de dos probabilidades logit estándar. Expresemos la probabilidad de elegir la alternativa $\mathrm{i} \in \mathrm{B}_k$ como el producto de dos probabilidades: la probabilidad de elegir una alternativa dentro del nido $\mathrm{B}_k$ y la probabilidad de elegir concretamente la alternativa i, condicionada a que una alternativa de $\mathrm{B}_k$ ha sido escogida:

$$P_{ni} = P_{ni|B_k} P_{nB_k},$$

donde $P_{ni|B_k}$es la probabilidad condicionada de elegir la alternativa i dado que se ha elegido una alternativa del nido $\mathrm{B}_k$ y $P_{nB_k}$es la probabilidad marginal de elegir una alternativa en el nido $\mathrm{B}_k$ (con la marginalidad aplicada sobre todas las alternativas en $\mathrm{B}_k$). Esta igualdad es exacta, ya que cualquier probabilidad se puede escribir como el producto de una probabilidad marginal y una probabilidad condicionada.

La razón para descomponer $P_{ni}$ en una probabilidad marginal y una condicionada es que, con la fórmula logit jerárquica descrita para $P_{ni}$, las probabilidades marginales y condicionadas toman la forma de logits. En particular, estas probabilidades se pueden expresar como

$$(4.4) \qquad P_{nB_k} = \frac{e^{W_{nk}+\lambda_k I_{nk}}}{\sum_{l=1}^{K} e^{W_{nl}+\lambda_l I_{nl}}},$$

$$(4.5) \qquad P_{ni|B_k} = \frac{e^{Y_{ni}/\lambda_k}}{\sum_{j\in B_k} e^{Y_{nj}/\lambda_k}},$$

donde

$$I_{nk} = ln \sum_{j\in B_k} e^{Y_{nj}/\lambda_k}.$$

La obtención de estas expresiones a partir de la probabilidad de elección (4.2) es una simple cuestión de reordenamiento algebraico. Para los lectores interesados, se proporciona en la sección 4.2.5.

Expresado en palabras, la probabilidad de elegir una alternativa en $\mathrm{B}_k$ toma la forma de la fórmula logit, como si fuese el resultado de un modelo para una elección entre nidos. Esta probabilidad incluye variables $W_{nk}$ que varían entre nidos pero no entre alternativas dentro de cada nido. También incluye una cantidad llamada $I_{nk}$, cuyo significado dilucidaremos posteriormente. La probabilidad de elegir $i$ condicionada al hecho de que una alternativa de $\mathrm{B}_k$ ha sido seleccionada, también está dada por una fórmula logit, como si fuese el resultado de un modelo de elección entre las alternativas dentro del nido. Esta probabilidad condicionada incluye las variables $Y_{nj}$ que varían entre alternativas dentro del nido.

Tenga en cuenta que estas variables se dividen por $\lambda_k$, de modo que cuando $Y_{nj}$ es lineal en los parámetros, los coeficientes que entran en esta probabilidad condicionada son los coeficientes originales divididos por $\lambda_k$. Es habitual referirse a la probabilidad marginal (elección del nido) como al *modelo superior* (*upper model*) y a la probabilidad condicionada (elección de alternativa dentro del nido) como el modelo inferior (*lower model*), lo que refleja sus posiciones relativas en la figura 4.1.

La cantidad $I_{nk}$ enlaza los modelos superior e inferior al traer información desde el modelo inferior hacia el modelo superior. Ben-Akiva (1973) identificó por primera vez la fórmula correcta para este enlace. En particular, $I_{nk}$ es el logaritmo del denominador del modelo inferior. Esta fórmula tiene un significado importante. Recuerde de la explicación sobre el excedente del consumidor en un modelo logit (sección 3.5) que el logaritmo del denominador del modelo logit es la utilidad esperada que el decisor obtiene de la situación de elección, como muestran Williams (1977) y Small y Rosen (1981). La misma interpretación aplica aquí: $\lambda_k I_{nk}$ es la utilidad esperada que recibe el decisor $\mathrm{n}$ de la elección entre las alternativas del nido $\mathrm{B}_k$. La fórmula para la utilidad esperada es igual a la de un modelo logit porque, condicionada a un nido, la elección de las alternativas dentro del nido es de hecho un logit, tal y como se muestra en la ecuación (4.5). $I_{nk}$ es a menudo llamado *valor inclusivo* o *utilidad inclusiva* del nido $\mathrm{B}_k$. También se le llama “término log-suma”, porque es el logaritmo de una suma (de utilidades representativas exponenciadas). El término "precio inclusivo" también se utiliza en ocasiones; sin embargo, en todo caso sería el negativo de $I_{nk}$ lo que se asemejaría más a un precio.

El coeficiente $\lambda_k$ de $I_{nk}$ en el modelo superior es a menudo llamado el coeficiente log-suma. Como se ha mencionado, $\lambda_k$ refleja el grado de independencia entre las partes no observadas de la utilidad de las alternativas del nido $\mathrm{B}_k$, con una $\lambda_k$ menor indicando menor independencia (mayor correlación).

Es apropiado que el valor inclusivo entre como variable explicativa en el modelo superior. Dicho de forma general, la probabilidad de elegir el nido $\mathrm{B}_k$ depende de la utilidad esperada que la persona recibe de ese nido. Esta utilidad esperada incluye la utilidad que recibe sin importar qué alternativa elige dentro del nido, que es $W_{nk}$, además de la utilidad adicional esperada que recibe por ser capaz de elegir la mejor alternativa dentro del nido, que es $\lambda_k I_{nk}$.

Recordemos que los coeficientes que entran en el modelo inferior se dividen por $\lambda_k$, como se indica en la ecuación (4.5). Estos mismos modelos han sido especificados y estimados sin dividir por $\lambda_k$ en el modelo inferior. Daly (1987) y Greene (2000) describen un modelo de este tipo y el paquete de software STATA lo incluye en su modelo de logit jerárquico en el comando *nlogit*. El paquete NLOGIT permite cualquier especificación. Si los coeficientes en el modelo inferior no se dividen por $\lambda_k$, las probabilidades de elección no son iguales a las expresadas en la ecuación (4.2). Como se muestra en la obtención de estas fórmulas en la sección 4.2.5, se necesita la división por $\lambda_k$ para que el producto de las probabilidades condicionadas y marginales sean iguales a las probabilidades logit jerárquicas dadas por la ecuación (4.2). Sin embargo, el hecho de que el modelo no dé las probabilidades de la ecuación (4.2) no necesariamente significa que sea inadecuado. Koppelman y Wen (1998) y Hensher y Greene (2002) comparan los dos enfoques (dividir por $\lambda_k$ frente a no dividir) y muestran que este último modelo no es consistente con la maximización de la utilidad cuando algún coeficiente es común entre nidos (como un coeficiente de costo que sea el mismo para los medios de transporte autobús y automóvil). Heiss (2002) señala la relación inversa: si no hay coeficientes comunes entre nidos, el segundo modelo es consistente con la maximización de la utilidad, ya que la división necesaria por $\lambda_k$ en cada nido se lleva a cabo de manera implícita (en lugar de explícitamente) al permitir coeficientes separados en cada uno de los nidos, de tal manera que la escala de los coeficientes difiera entre nidos. Por el contrario, cuando los coeficientes son comunes entre los nidos, Heiss encontró que no dividir por $\lambda_k$ conlleva implicaciones contrarias a la intuición.

### *4.2.4 Estimación*

Los parámetros de un modelo jerárquico se pueden estimar mediante técnicas estándar de máxima verosimilitud. Sustituyendo las probabilidades de elección de la expresión (4.2) en la función de log - verosimilitud da una función explícita de los parámetros de este modelo. Los valores de los parámetros que maximizan esta función son, bajo condiciones bastante generales, consistentes y eficientes (Brownstone y Small, 1989).

Existen rutinas en paquetes de software comercial para la estimación de modelos jerárquicos por máxima verosimilitud. Hensher y Greene (2002) ofrecen una guía para estimar logits jerárquicos utilizando software disponible. La maximización numérica a veces es difícil, ya que la función log-verosimilitud no es globalmente cóncava e incluso en áreas cóncavas dista mucho de ser cuadrática. El investigador puede necesitar ayudar a las rutinas probando diferentes algoritmos y/o distintos valores de inicio, como se trata en el Capítulo 8.

En lugar de utilizar máxima verosimilitud, los modelos logit jerárquicos pueden ser estimados consistentemente (pero no de manera eficiente) de una manera secuencial, explotando el hecho de que las probabilidades de elección se pueden descomponer en probabilidades marginales y condicionadas que son logit. Esta estimación secuencial se realiza "de abajo hacia arriba". En primer lugar se estiman los modelos inferiores (para la elección de la alternativa dentro de un nido). Usando los coeficientes estimados, el valor inclusivo se calcula para cada modelo inferior. A continuación, se estima el modelo superior (para la elección del nido) con el valor inclusivo entrando como variable explicativa.

La estimación secuencial presenta dos dificultades que desaconsejan su uso. En primer lugar, los errores estándar de los parámetros de los modelos superiores están sesgados a la baja, como Amemiya (1978) señaló por primera vez. Este sesgo surge debido a que la varianza de la estimación del valor inclusivo que entra en el modelo superior no se incorpora en el cálculo de los errores estándar. Con errores estándar sesgados a la baja, los intervalos de confianza se estiman inferiores a la realidad y los estadísticos-t mayores, por lo que el modelo superior parecerá ser mejor de lo que realmente es. Ben-Akiva y Lerman (1985, p. 298) proporcionan un procedimiento para ajustar los errores estándar para eliminar el sesgo.

En segundo lugar, por lo general es habitual que algunos parámetros aparezcan en varios sub-modelos. La estimación de los diversos modelos superiores e inferiores por separado proporciona estimaciones independientes de cualquier parámetro común que aparezca en el modelo. La estimación simultánea por máxima verosimilitud asegura que los parámetros comunes son forzados a ser los mismos dondequiera que aparezcan en el modelo.

Estas dos complicaciones son síntomas de una circunstancia más general, a saber, que la estimación secuencial de modelos logit jerárquicos, aunque coherente, no es tan eficiente como la estimación simultánea por máxima verosimilitud. Con estimación simultánea, toda la información se utiliza en la estimación de cada parámetro y los parámetros que son comunes entre componentes necesariamente son obligados a ser iguales. Dado que hay software comercial disponible para la estimación simultánea, existen pocas razones para estimar secuencialmente un logit jerárquico. Si surgen problemas en la estimación simultánea, el investigador podría considerar útil estimar el modelo secuencialmente y luego usar las estimaciones secuenciales como valores iniciales en la estimación simultánea. El principal valor de la descomposición del logit jerárquico en sus componentes superior e inferior no proviene de su uso como una herramienta de estimación sino más bien como un dispositivo heurístico: la descomposición ayuda en gran medida a la comprensión del significado y la estructura del modelo logit jerárquico.

### *4.2.5 Equivalencia de las fórmulas del logit jerárquico*

Hemos afirmado en la sección 4.2.3 que el producto de las probabilidades marginales y condicionadas en (4.4) y (4.5) es igual a la probabilidad conjunta de (4.2). Vamos a verificar esta afirmación:

$$P_{ni} = \frac{e^{V_{ni}/\lambda_k}\left(\sum_{j\in B_k} e^{V_{nj}/\lambda_k}\right)^{\lambda_k - 1}}{\sum_{l=1}^{K}\left(\sum_{j\in B_l} e^{V_{nj}/\lambda_l}\right)^{\lambda_l}} \; por\ (4.2)$$

$$= \frac{e^{V_{ni}/\lambda_k}}{\sum_{j\in B_k} e^{V_{nj}/\lambda_k}} \frac{\left(\sum_{j\in B_k} e^{V_{nj}/\lambda_k}\right)^{\lambda_k}}{\sum_{l=1}^{K}\left(\sum_{j\in B_l} e^{V_{nj}/\lambda_l}\right)^{\lambda_l}}$$

$$= \frac{e^{(W_{nk}+Y_{ni})/\lambda_k}}{\sum_{j\in B_k} e^{(W_{nk}+Y_{nj})/\lambda_k}} \frac{\left(\sum_{j\in B_k} e^{(W_{nk}+Y_{nj})/\lambda_k}\right)^{\lambda_k}}{\sum_{l=1}^{K}\left(\sum_{j\in B_l} e^{(W_{nl}+Y_{nj})/\lambda_l}\right)^{\lambda_l}}$$

$$= \frac{e^{W_{nk}/\lambda_k} e^{Y_{ni}/\lambda_k}}{e^{W_{nk}/\lambda_k}\sum_{j\in B_k} e^{Y_{nj}/\lambda_k}} \frac{e^{W_{nk}}\left(\sum_{j\in B_k} e^{Y_{nj}/\lambda_k}\right)^{\lambda_k}}{\sum_{l=1}^{K} e^{W_{nl}}\left(\sum_{j\in B_l} e^{Y_{nj}/\lambda_l}\right)^{\lambda_l}}$$

$$= \frac{e^{Y_{ni}/\lambda_k}}{\sum_{j\in B_k} e^{Y_{nj}/\lambda_k}} \frac{e^{W_{nk}+\lambda_k I_{nk}}}{\sum_{l=1}^{K} e^{W_{nl}+\lambda_l I_{nl}}}$$

$$= P_{ni|B_k} P_{nB_k},$$

donde la penúltima igualdad se debe a que $I_{nk} = \ln \sum_{j\in B_k} e^{Y_{nj}/\lambda_k}$, reconociendo que $\mathrm{e}^{x}\mathrm{b}^{c} = \mathrm{e}^{x + c \ln b}$.

## 4.3 Logit jerárquico de tres niveles

El modelo logit jerárquico que hemos visto hasta este punto se denomina modelo logit jerárquico de dos niveles, ya que hay dos niveles de modelado: las probabilidades marginales (modelo superior) y las probabilidades condicionadas (modelos inferiores). En el caso de la elección del medio de transporte, los dos niveles son: el modelo marginal de automóviles frente a transporte público y los modelos condicionados de tipo de automóvil o tipo de transporte público (automóvil en solitario o compartido condicionado a la elección de automóvil, y autobús o tren condicionado a la elección de transporte público).

En algunas situaciones es apropiado emplear modelos logit jerárquicos de tres o más niveles. Los modelos de tres niveles se obtienen al dividir el conjunto de alternativas en nidos y luego dividir cada nido en sub-nidos. La fórmula de la probabilidad de elección en este caso es una generalización de (4.2) con las sumas adicionales para los sub-nidos dentro de las sumas de los nidos. Ver McFadden (1978) o Ben-Akiva y Lerman (1985) para la fórmula.

Al igual que con un logit jerárquico de dos niveles, las probabilidades de elección para un modelo de tres niveles se pueden expresar como una serie de logits. El modelo superior describe la elección del nido; las modelos intermedios describen la elección de sub-nidos dentro de cada nido; y los modelos inferiores describen la elección de alternativas dentro de cada sub-nido. El modelo superior incluye un valor inclusivo para cada nido. Este valor representa la utilidad esperada que el decisor puede obtener de los sub-nidos dentro del nido. Se calcula como el logaritmo del denominador del modelo intermedio para ese nido. Del mismo modo, los modelos intermedios incluyen un valor inclusivo para cada sub-nido, que representa la utilidad esperada que el decisor puede obtener de las alternativas dentro del sub-nido. Se calcula como el logaritmo del denominador del modelo inferior para cada sub-nido.

Como ejemplo, considere la elección que una familia (hogar) hace de su vivienda dentro de un área metropolitana. La familia elige una entre todas las viviendas disponibles en la ciudad. Las viviendas están disponibles en diferentes barrios de la ciudad y con diferente número de habitaciones. Es razonable asumir que hay factores no observados que son comunes a todas las viviendas en el mismo barrio, tales como la proximidad a tiendas y a lugares de entretenimiento. Por lo tanto, se espera que la parte no observada de la utilidad esté correlacionada entre todas las viviendas dentro de un barrio determinado. También hay factores no observados que son comunes a todas las viviendas con el mismo número de habitaciones, como por ejemplo la comodidad para trabajar en casa. Por lo tanto, esperamos que la utilidad no observada esté aún más correlacionada entre viviendas del mismo tamaño en el mismo barrio que entre viviendas de diferente tamaño en el mismo barrio. Este patrón de correlación puede ser representado por la anidación de las viviendas por barrio y posteriormente sub-anidando por número de dormitorios. Un diagrama en árbol que representa esta situación se puede ver en la figura 4.2 para San Francisco. Hay tres niveles de sub-modelos: la probabilidad de elección de barrio, la probabilidad de elección de número de dormitorios, dado el barrio, y la elección de la vivienda dada la elección de vecindad y número de dormitorios.

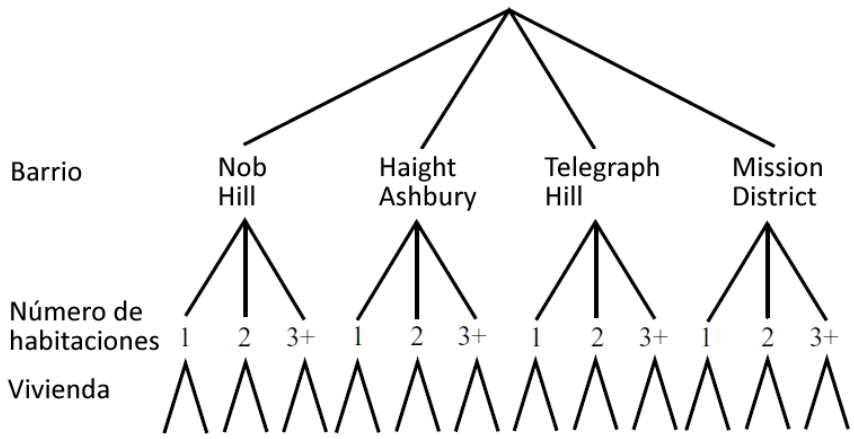


*Figura 4.2. Logit jerárquico de tres niveles.*

Un modelo logit jerárquico con esta estructura de anidación soporta la IIA de las siguientes formas:

1. El ratio de probabilidades de dos viviendas en el mismo barrio y con el mismo número de dormitorios es independiente de las características de todas las demás viviendas. Por ejemplo, la reducción del precio de un apartamento de dos dormitorios en Haight Ashbury extrae probabilidad de elección proporcionalmente de todas las viviendas de un dormitorio en Telegraph Hill.
2. El ratio de probabilidades de dos viviendas en el mismo barrio, pero con diferente número de habitaciones es independiente de las características de las viviendas en otros barrios, pero depende de las características de las viviendas en el mismo barrio que tienen el mismo número de dormitorios que cualquiera de estas viviendas. Bajar el precio de un apartamento de dos dormitorios en Haight Ashbury extrae probabilidad de elección proporcionalmente de viviendas de uno y dos dormitorios en Telegraph Hill, pero extrae de manera desproporcional de viviendas de dos dormitorios en Haight Ashbury en comparación a viviendas de un dormitorio en Haight Ashbury.

3. El ratio de probabilidades de dos viviendas en distintos barrios depende de las características de todas las demás viviendas en esos barrios, pero no de las características de viviendas en otros barrios. Bajar el precio de un apartamento de dos dormitorios en Haight Ashbury extrae probabilidad de elección proporcionalmente de todas las viviendas fuera de Haight Ashbury pero extrae de manera desproporcional de viviendas en Haight Ashbury en relación a las viviendas fuera de Haight Ashbury.

Cada nivel de agrupamiento en un logit jerárquico introduce parámetros que representan el grado de correlación entre alternativas dentro de los nidos. Con el conjunto completo de alternativas dividido en nidos, el parámetro $\lambda_k$ se introduce para el nido k, como se ha descrito para los modelos de dos niveles. Si los nidos se dividen además en sub-nidos, entonces un parámetro $\sigma_{mk}$ se introduce para el sub-nido $\mathrm{m}$ del nido $\mathrm{k}$. Al descomponer la probabilidad en una serie de modelos logit, $\sigma_{mk}$ es el coeficiente del valor inclusivo en el modelo intermedio y $\lambda_k\sigma_{mk}$ es el codeficiente del valor inclusivo en el modelo superior. Del mismo modo que para un logit jerárquico de dos niveles, los valores de estos parámetros deben encontrarse dentro de ciertos rangos para ser consistentes con la maximización de la utilidad. Si $0 < \lambda_k < 1$ y $0 < \sigma_{mk} < 1$, el modelo es consistente con la maximización de la utilidad para todos los niveles de las variables explicativas. Un valor negativo para los parámetros es incompatible con la maximización de la utilidad. Y valores superiores a uno son consistentes sólo para un cierto rango de valores de las variables explicativas.

## 4.4 Solapamiento de nidos

Para los modelos logit jerárquicos que hemos considerado, cada alternativa forma parte de un único nido (y para los modelos de tres niveles, un único sub-nido). Este aspecto de los modelos logit jerárquicos es una restricción en ocasiones inadecuada. Por ejemplo, en nuestro caso de elección de medio de transporte, ponemos los medios de transporte de automóvil en solitario y compartido en un mismo nido ya que tienen algunos atributos no observados similares. Sin embargo, compartir automóvil también tiene algunos atributos no observados que son similares a los del autobús y el tren, como la falta de flexibilidad en la planificación (el trabajador no puede ir a trabajar en cualquier momento del día, sino que tiene que ir en el momento en que el viaje compartido ha sido pactado, de manera similar a tomar una línea de autobús o tren con salidas fijadas). Sería útil tener un modelo en el que la utilidad no observada para la alternativa automóvil compartido pudiese estar correlacionada con la de viajar solo en automóvil y al mismo tiempo estar correlacionada, aunque en un grado diferente, con las alternativas autobús y tren. Dicho de manera equivalente, sería útil poder incluir la alternativa vehículo compartido en dos nidos: en un nido junto a la opción de viajar en automóvil en solitario y en otro nido junto a las opciones de viajar en autobús y tren.

Varios tipos de modelos GEV se han especificado con nidos solapados, de manera que una alternativa pueda ser miembro de más de un nido. Vovsha (1997), Bierlaire (1998) y Ben-Akiva y Bierlaire (1999) han propuesto diversos modelos llamados logits jerárquicos cruzados (*cross-nested logits*, CNLs) que contienen múltiples nidos solapados. Small (1987) consideró una situación en la que las alternativas tienen un orden natural, como por ejemplo el número de automóviles que posee un hogar (0, 1, 2, 3...) o la destinación para ir de compras, con las zonas de tiendas ordenadas por la distancia desde la residencia del decisor. Small especificó un modelo, llamado modelo ordenado generalizado de valor extremo (*ordered generalized extreme value*, OGEV ) en el que la correlación en utilidad no observada entre dos alternativas depende de su proximidad en la ordenación. Este modelo tiene nidos solapados como en el CNL, pero cada nido consta de dos alternativas y se impone un patrón en las correlaciones (mayor correlación para los pares más cercanos). Small (1994) y Bhat (1998b) describieron una versión anidada del OGEV, que es similar a un logit jerárquico excepto en que los modelos inferiores (modelos de elección de alternativas dada la elección de un nido) son OGEV en lugar de logit estándar. Chu (1981,

1989) propuso un modelo llamado logit combinacional emparejado (*paired combinatorial logit* , PCL) en el que cada par de alternativas constituye un nido con su propia correlación. Con $J$ alternativas, cada alternativa es miembro de $J-1$ nidos, y la correlación de su utilidad no observada con cada una de las otras alternativas puede ser estimada. Wen y Koppelman (2001) han desarrollado un modelo logit jerárquico generalizado (*generalized nested logit*, GNL) que incluye cruzados como casos especiales el PCL y otros modelos jerárquicos. En los apartados siguientes describo los modelos PCL y GNL, el primero debido a su simplicidad y el segundo debido a su generalidad.

### 4.4.1 Logit combinacional emparejado (PCL)

Cada par de alternativas se considera que es un nido. Dado que cada alternativa está emparejada con cada una de las otras alternativas, cada alternativa es miembro de $J-1$ nidos. Un parámetro etiquetado como $\lambda_{ij}$ indica el grado de independencia entre las alternativas $i$ y $j$. Dicho de forma equivalente: $1-\lambda_{ij}$ es una medida de la correlación existente entre la utilidad no observada de la alternativa $i$ y de la alternativa $j$. Este parámetro es análogo a $\lambda_k$ en un modelo logit jerárquico, donde $\lambda_k$ indica el grado de independencia entre alternativas dentro del nido y $1-\lambda_k$ es una medida de correlación dentro del nido. Y como sucede con el logit jerárquico, el modelo PCL se convierte en un logit estándar cuando $\lambda_{ij}=1$ para todos los pares de alternativas.

Las probabilidades de elección para el modelo PCL son

$$(4.6) \qquad P_{ni} = \frac{\sum_{j\neq i} e^{V_{ni}/\lambda_{ij}} (e^{V_{ni}/\lambda_{ij}} + e^{V_{nj}/\lambda_{ij}})^{\lambda_{ij}-1}}{\sum_{k=1}^{J-1} \sum_{l=k+1}^{J} (e^{V_{nk}/\lambda_{kl}} + e^{V_{nl}/\lambda_{kl}})^{\lambda_{kl}}}.$$

La suma en el numerador es sobre el total de $J-1$ nidos en los que la alternativa $i$ se encuentra. Para cada uno de estos nidos, el término que se añade a la suma es el mismo que el numerador de la probabilidad logit jerárquica (4.2). Así, el modelo PCL es como el logit jerárquico salvo que permite que $i$ esté en más de un nido. El denominador en la probabilidad PCL también tiene la misma forma de un logit jerárquico: es la suma sobre todos los nidos de la suma de los $exp(V/\lambda)$s dentro del nido, elevado a la potencia adecuada $\lambda$. Si $\lambda_{ij}$ está entre cero y uno para todos los pares $ij$, el modelo es consistente con la maximización de la utilidad para todos los niveles de los datos. Es fácil verificar que $P_{ni}$ se convierte en la fórmula logit estándar cuando $\lambda_{ij}=1\ \forall i,j$. En su aplicación práctica, Koppelman y Wen (2000) comprobaron que el modelo PCL obtenía mejores resultados que el logit jerárquico o el logit estándar.

El investigador puede hacer un test de hipótesis sobre si $\lambda_{ij}=1$ para algunos o todos los pares, mediante el test de ratio de verosimilitudes de la Sección 3.8.2. La aceptación de la hipótesis para un par de alternativas implica que no hay una correlación significativa en la utilidad no observada para ese par. El investigador también puede añadir cierta estructura al patrón de correlación. Por ejemplo, es posible asumir que las correlaciones son iguales entre un grupo de alternativas; esta asunción se impone fijando $\lambda_{ij}=\lambda_{kl}$ para todo $i,j,k$ y $l$ dentro del grupo. El modelo OGEV de Small es un modelo PCL en el que se especifica que $\lambda_{ij}$ sea una función de la proximidad entre $i$ y $j$. Con un número grande de alternativas, el investigador tendrá probablemente necesidad de imponer alguna forma de estructura en las $\lambda_{ij}$s, simplemente para evitar la proliferación de parámetros que afloran con un valor de $J$ grande. Esta proliferación de parámetros, uno para cada par de alternativas, es lo que hace al modelo PCL tan flexible. El objetivo del investigador es aplicar esta flexibilidad con sentido en su situación particular.

Como se ha visto casi al final de la sección 2.5, ya que la escala y el nivel de utilidad son irrelevantes, como máximo pueden estimarse $J(J-1)/2-1$ parámetros de covarianza en un modelo de elección discreta. Un modelo PCL contiene $J(J-1)/2\ \lambda$s: una $\lambda$ por cada alternativa emparejada con cada una del resto de alternativas, teniendo en cuenta que $i$ emparejado con $j$ es lo mismo que $j$ emparejado con

$i$. El número de $\lambda$s excede el número máximo de parámetros de covarianza identificables exactamente en uno. El investigador, por lo tanto, debe fijar al menos una restricción sobre las $\lambda$s. Esto puede lograrse mediante la normalización de una de las $\lambda$s a 1. Si se impone cierta estructura sobre el patrón de correlación, como se describe en el párrafo anterior, en general esta estructura impondrá la normalización de forma automática.

### 4.4.2 Logit jerárquico generalizado (GNL)

Los nidos de alternativas son etiquetados $B_1, B_2, \ldots, B_K$. Cada alternativa puede ser miembro de más de un nido. Es importante destacar que una alternativa puede estar en un nido en diversos grados. Dicho de otra manera, una alternativa se distribuye entre los nidos, estando más presente en unos nidos que en otros. Un parámetro de *asignación* $\alpha_{jk}$indica en qué medida la alternativa $j$ es un miembro del nido $k$. Este parámetro debe ser no negativo: $\alpha_{jk} \geq 0\ \forall j, k$. Un valor nulo significa que la alternativa no está presente en el nido en absoluto. La interpretación se facilita al imponer que los parámetros de asignación sumen uno entre los nidos para cualquier alternativa: $\sum_k \alpha_{jk} = 1\ \forall j$. Bajo esta condición, $\alpha_{jk}$ refleja la porción de la alternativa que se asigna a cada nido.

Se define un parámetro $\lambda_k$ para cada nido que realiza la misma función que en los modelos logit jerárquicos, es decir, indica el grado de independencia entre alternativas dentro del nido: un valor mayor de $\lambda_k$ se traduce en una mayor independencia y menor correlación.

La probabilidad de que la persona $n$ elija la alternativa $i$ es

$$(4.7)\qquad P_{ni} = \frac{\sum_k \left(\alpha_{ik} e^{V_{ni}}\right)^{1/\lambda_k} \left(\sum_{j \in B_k} (\alpha_{jk} e^{V_{nj}})^{1/\lambda_k}\right)^{\lambda_k - 1}}{\sum_{l=1}^{K} \left(\sum_{j \in B_l} (\alpha_{jl} e^{V_{nj}})^{1/\lambda_l}\right)^{\lambda_l}}.$$

Esta fórmula es similar a la probabilidad logit jerárquica dada en la ecuación (4.2), excepto que el numerador es una suma sobre todos los nidos que contienen la alternativa $i$, con ponderaciones aplicadas a estos nidos. Si cada alternativa entra en un solo nido, con $\alpha_{jk} = 1$ para $j \in B_k$ y cero en caso contrario, el modelo se convierte en un modelo logit jerárquico. Y si además, $\lambda_k = 1$ para todos los nidos, entonces el modelo se convierte en el logit estándar. Wen y Koppelman (2001) obtienen varios modelos jerárquicos cruzados como casos especiales del modelo GNL.

Para facilitar la interpretación, la probabilidad GNL se puede descomponer como

$$P_{ni} = \sum_k P_{ni|B_k} P_{nk},$$

donde la probabilidad del nido $k$ es

$$P_{nk} = \frac{\left(\sum_{j \in B_k} (\alpha_{jk} e^{V_{nj}})^{1/\lambda_k}\right)^{\lambda_k}}{\sum_{l=1}^{K} \left(\sum_{j \in B_l} (\alpha_{jl} e^{V_{nj}})^{1/\lambda_l}\right)^{\lambda_l}}.$$

y la probabilidad de la alternativa $i$ dado el nido $k$ es

$$P_{ni|B_k} = \frac{(\alpha_{ik} e^{V_{ni}})^{1/\lambda_k}}{\sum_{j \in B_k} (\alpha_{jk} e^{V_{nj}})^{1/\lambda_k}}.$$

## 4.5 Logit heterocedástico

En lugar de capturar las correlaciones entre alternativas, el investigador puede desear simplemente permitir que la varianza de factores no observados difiera entre alternativas. Steckel y Vanhonacker (1988), Bhat (1995) y Recker (1995) describen un tipo de modelo GEV, llamado valor extremo heterocedástico (*heteroskedastic extreme value*, HEV) que es igual al logit excepto por permitir una varianza diferente para cada alternativa. La utilidad se especifica como $U_{nj} = V_{nj} + \varepsilon_{nj}$, donde $\varepsilon_{nj}$ se distribuye independientemente como valor extremo con varianza $(\theta_j \pi)^2/6$. No hay correlación en factores no observados entre las alternativas, sin embargo, la varianza de los factores no observados es diferente para alternativas diferentes. Para ajustar la escala global de la utilidad, la varianza de una alternativa se normaliza a $\pi^2/6$, que es la varianza de la distribución normalizada de valor extremo. Las varianzas de las otras alternativas se estiman posteriormente en relación a la varianza normalizada.

Las probabilidades de elección para este logit heterocedástico son (Bhat, 1995)

$$P_{ni} = \int \left[ \prod_{j \neq i} e^{-e^{-(V_{ni} - V_{nj} + \theta_i w)/\theta_j}} \right] e^{-e^{-w}} e^{-w} dw.$$

donde $w = \varepsilon_{ni}/\theta_i$. La integral no tiene una forma cerrada, sin embargo, se puede aproximar por simulación . Tenga en cuenta que $exp(-exp(-w))exp(-w)$ es la densidad de la distribución valor extremo, facilitada en la Sección 3.1. Por lo tanto, $P_{ni}$ es la integral del factor entre corchetes sobre la densidad de valor extremo. Puede ser simulada de la siguiente manera: (1) Extraiga un valor de la distribución de valor extremo, usando el procedimiento descrito en la Sección 9.2.3. (2) Para esta extracción de $w$, calcule el factor entre paréntesis, es decir $\prod_{j \neq i} \exp(-\exp(-(V_{ni} - V_{nj} + \theta_i w)/\theta_j))$. (3) Repita los pasos 1 y 2 múltiples veces y promedie los resultados. Este promedio es una aproximación de $P_{ni}$. Bhat (1995) muestra que, puesto que la integral es sólo unidimensional, las probabilidades logit heterocedásticas se pueden calcular eficazmente mediante cuadratura numérica en lugar de simulación.

## 4.6 La familia GEV

Pasamos ahora a describir el proceso que McFadden (1978) desarrolló para generar modelos GEV. Utilizando este proceso, el investigador puede desarrollar nuevos modelos GEV que se adapten mejor a las circunstancias específicas de su problema de elección. Como ejemplo, se muestra cómo se utiliza el procedimiento para generar modelos que ya hemos comentado, es decir, logit, logit jerárquico y logit combinacional emparejado. El mismo procedimiento puede ser aplicado por un investigador para generar nuevos modelos con propiedades que cumplan con sus necesidades de investigación.

Para simplificar la notación, vamos a omitir el subíndice $n$ que denota al decisor. También, ya que vamos a utilizar $\exp(V_j)$ varias veces, lo referiremos de forma más compacta como $Y_j$. Es decir, $Y_j = \exp(V_j)$. Tenga presente que $Y_j$ es necesariamente positivo.

Considere una función G que depende de $Y_j$ para todo j. Denotamos esta función $G = G(Y_1, \dots, Y_J)$. Sea $G_i$ la derivada de G con respecto a $Y_i$: $G_i = \partial G / \partial Y_i$. Si esta función reúne ciertas condiciones, entonces es posible basar un modelo de elección discreta en ella. En particular, si G satisface las condiciones que se enumeran en el párrafo siguiente, entonces

$$P_i = \frac{Y_i G_i}{G} \tag{4.8}$$

es la probabilidad de elección de un modelo de elección discreta que es consistente con la maximización de la utilidad. Cualquier modelo que pueda ser formulado de esta manera es un modelo GEV. Por tanto, esta fórmula define la familia de modelos GEV.

Las propiedades que la función debe cumplir son las siguientes:

1. $\mathrm{G} \geq 0$ para todos los valores positivos de $Y_j$ $\forall j$.
2. G es una función homogénea de grado uno. Es decir si cada $Y_j$ se incrementa en una proporción determinada ρ, G se incrementa también en esa misma proporción ρ: $G(\rho Y_1, \dots, \rho Y_J) = \rho G(Y_1, \dots, Y_J)$. En realidad, Ben-Akiva y Francois (1983) mostraron que esta condición podía relajarse para permitir cualquier grado de homogeneidad. Mantenemos el uso de grado uno, ya que al hacerlo la condición es más fácil de interpretar y es consistente con la descripción original de McFadden.
3. $\mathrm{G} \rightarrow \infty$ cuando $Y_j \rightarrow \infty$ para cualquier j.
4. Las derivadas parciales cruzadas de G cambian de signo de una manera particular. Es decir, $\mathrm{G}_i \geq 0$ para todo i, $\mathrm{G}_{ij} = \partial \mathrm{G}_i / \partial \mathrm{Y}_j \leq 0$ para todo $j \neq i$, $\mathrm{G}_{ijk} = \partial \mathrm{G}_{ij} / \partial \mathrm{Y}_k \geq 0$ para cualquier $\mathrm{i}, \mathrm{j}, \mathrm{k}$ distintos, y así sucesivamente para derivadas parciales cruzadas de mayor orden.

Hay poca intuición económica que motive estas propiedades, particularmente la última. Sin embargo, es fácil verificar si una función las cumple. La falta de intuición detrás de las propiedades es al mismo tiempo una bendición y una maldición. La desventaja es que el investigador tiene poca orientación sobre cómo especificar una función G que proporcione un modelo que satisfaga las necesidades de su investigación. La ventaja es que el enfoque puramente matemático permite al investigador generar modelos que él no podría haber desarrollado confiando solamente en su intuición económica. Karlstrom (2001) ofrece un ejemplo: arbitrariamente especificó una G (en el sentido de que no se basaba en conceptos de comportamiento) y se encontró que la fórmula de probabilidad resultante se ajustaba mejor a sus datos que logit, logit jerárquico y PCL.

Para ilustrar el proceso de generación podemos mostrar cómo logit, logit jerárquico y los modelos PCL se obtienen bajo las especificaciones apropiadas de G.

**Logit**

Sea $\mathrm{G} = \sum_{j=1}^{J} Y_j$. Esta G exhibe las cuatro propiedades requeridas: (1) La suma de $Y_j$s positivas es positiva. (2) Si todas las $Y_j$s son incrementadas por un factor $\rho$, G se incrementa en ese mismo factor. (3) Si alguna $Y_j$ se incrementa sin límite, entonces G también lo hace. (4) La primera derivada parcial de G es $\mathrm{G}_i = \partial \mathrm{G} / \partial \mathrm{Y}_j = 1$, que cumple el criterio de que $\mathrm{G}_i > 0$. Y las derivadas de orden superior son todas cero, que cumple claramente el criterio expuesto, ya que son $\geq 0$ o $\leq 0$ como se requiere.

Si insertamos esta función G y su primera derivada $\mathrm{G}_i$ en (4.8), la probabilidad de elección resultante es

$$P_i = \frac{Y_i G_i}{G}$$

$$= \frac{Y_i}{\sum_{j=1}^{J} Y_i}$$

$$= \frac{e^{V_i}}{\sum_{j=1}^{J} e^{V_i}},$$

que es la fórmula logit.

**Logit jerárquico**

Las $J$ alternativas se dividen en $K$ nidos etiquetados $B_1, \dots, B_K$. Sea

$$G = \sum_{l=1}^{K} \left( \sum_{j \in B_l} Y_j^{1/\lambda_l} \right)^{\lambda_l} ,$$

con cada $\lambda_k$ entre cero y uno. Las tres primeras propiedades son fáciles de verificar. Para la cuarta propiedad, calculamos la primera derivada parcial

$$G_i = \lambda_k \left( \sum_{j \in B_k} Y_j^{1/\lambda_k} \right)^{\lambda_k - 1} \frac{1}{\lambda_k} Y_i^{(1/\lambda_k)-1}$$

$$= Y_i^{(1/\lambda_k)-1} \left( \sum_{j \in B_k} Y_j^{1/\lambda_k} \right)^{\lambda_k - 1}$$

para $i \in B_k$. Dado que $Y_j \geq 0\ \forall j$, tenemos que $\mathrm{G}_i \geq 0$, según se requería. La segunda derivada parcial cruzada es

$$G_{im} = \frac{\partial G_i}{\partial Y_m}$$

$$= (\lambda_k - 1) Y_i^{(1/\lambda_k)-1} \left( \sum_{j \in B_k} Y_j^{1/\lambda_k} \right)^{\lambda_k - 2} \frac{1}{\lambda_k} Y_m^{(1/\lambda_k)-1}$$

$$= \frac{\lambda_k - 1}{\lambda_k} (Y_i Y_m)^{(1/\lambda_k)-1} \left( \sum_{j \in B_k} Y_j^{1/\lambda_k} \right)^{\lambda_k - 2}$$

para $m \in B_k$ y $m \neq i$. Con $\lambda_k \leq 1$, $G_{ij} \leq 0$, según se requería. Para $j$ en un nido diferente a $i$, $G_{ij} = 0$, que también cumple el criterio. Derivadas parciales cruzadas de orden superior se calculan de manera similar; exhiben la propiedad requerida si $0 < \lambda_k \leq 1$.

La probabilidad de elección se convierte en

$$P_i = \frac{Y_i G_i}{G}$$

$$= \frac{Y_i Y_i^{(1/\lambda_k)-1} \left( \sum_{j \in B_k} Y_j^{1/\lambda_l} \right)^{\lambda_k - 1}}{\sum_{l=1}^{K} \left( \sum_{j \in B_l} Y_j^{1/\lambda_l} \right)^{\lambda_l}}$$

$$= \frac{Y_i^{1/\lambda_k} \left(\sum_{j\in B_k} Y_j^{1/\lambda_l}\right)^{\lambda_k - 1}}{\sum_{l=1}^{K} \left(\sum_{j\in B_l} Y_j^{1/\lambda_l}\right)^{\lambda_l}}$$

$$= \frac{(e^{V_i})^{1/\lambda_k} \left(\sum_{j\in B_k} (e^{V_j})^{1/\lambda_l}\right)^{\lambda_k - 1}}{\sum_{l=1}^{K} \left(\sum_{j\in B_l} (e^{V_j})^{1/\lambda_l}\right)^{\lambda_l}}$$

$$= \frac{e^{V_i/\lambda_k} \left(\sum_{j\in B_k} e^{V_j/\lambda_l}\right)^{\lambda_k - 1}}{\sum_{l=1}^{K} \left(\sum_{j\in B_l} e^{V_j/\lambda_l}\right)^{\lambda_l}}$$

Que es la fórmula logit jerárquica ( 4.2).

**Logit combinacional emparejado**

Sea

$$G = \sum_{k=1}^{J-1} \sum_{l=k+1}^{J} (Y_k^{1/\lambda_{kl}} + Y_l^{1/\lambda_{kl}})^{\lambda_{kl}}.$$

Las propiedades requeridas se verifican de la misma forma que para el logit jerárquico. Tenemos

$$G_i = \sum_{j\neq i} \lambda_{ji} (Y_i^{1/\lambda_{ij}} + Y_j^{1/\lambda_{ij}})^{\lambda_{ij} - 1} \frac{1}{\lambda_{ij}} Y_i^{(1/\lambda_{ij}) - 1}$$

$$= \sum_{j\neq i} Y_i^{(1/\lambda_{ij}) - 1} (Y_i^{1/\lambda_{ij}} + Y_j^{1/\lambda_{ij}})^{\lambda_{ij} - 1}.$$

De esta forma, la probabilidad de elección es

$$\mathrm{P_i} = \frac{\mathrm{Y_i G_i}}{\mathrm{G}}$$

$$= \frac{Y_i \sum_{j\neq i} Y_i^{(1/\lambda_{ij}) - 1} (Y_i^{1/\lambda_{ij}} + Y_j^{1/\lambda_{ij}})^{\lambda_{ij} - 1}}{\sum_{k=1}^{J-1} \sum_{l=k+1}^{J} (Y_k^{1/\lambda_{kl}} + Y_l^{1/\lambda_{kl}})^{\lambda_{kl}}}$$

$$= \frac{\sum_{j\neq i} Y_i^{(1/\lambda_{ij})} (Y_i^{1/\lambda_{ij}} + Y_j^{1/\lambda_{ij}})^{\lambda_{ij} - 1}}{\sum_{k=1}^{J-1} \sum_{l=k+1}^{J} (Y_k^{1/\lambda_{kl}} + Y_l^{1/\lambda_{kl}})^{\lambda_{kl}}}$$

$$= \frac{\sum_{j\neq i} e^{V_i/\lambda_{ij}} (e^{V_i/\lambda_{ij}} + e^{V_j/\lambda_{ij}})^{\lambda_{ij} - 1}}{\sum_{k=1}^{J-1} \sum_{l=k+1}^{J} (e^{V_k/\lambda_{kl}} + e^{V_l/\lambda_{kl}})^{\lambda_{kl}}}$$

que es la fórmula del modelo PCL (4.6).

**Logit jerárquico generalizado**

El lector puede verificar que las probabilidades GNL en la ecuación (4.7) se obtienen de

$$G = \sum_{k=1}^{K} \left( \sum_{j \in B_k} (\alpha_{jk} Y_j)^{1/\lambda_k} \right)^{\lambda_k} .$$

Usando el mismo proceso, los investigadores pueden generar otros modelos GEV.