MODELOS GENERALIZADOS DE VALORES EXTREMOS (GEV)

Los ***modelos generalizados de valores extremos*** (***generalized extreme value models***, **GEV**) son una familia de modelos con diversos patrones de sustitución. Todos tienen en común que las partes no observadas que explican la utilidad de todas las alternativas están distribuidas conjuntamente según una distribución de valores extremos, lo que permite correlaciones entre alternativas. Si todas las correlaciones fueran cero, la distribución GEV se podría expresar como el producto de distribuciones de valores extremos independientes, y el GEV se reduce a un Logit estándar.

En realidad, los nidos no tienen por qué ser conjuntos disjuntos. Una alternativa puede pertenecer a más de un nido. Se han especificado distintos modelos GEV para nidos con intersecciones. Un planteamiento, para alternativas ordenadas (como el número de coches que posee una familia, que puede ser 0, 1, 2…), es el modelo de valores extremos generalizado ordenado (OGEV, *ordered generalized extreme value*) en el que la correlación entre las utilidades no observadas entre dos alternativas depende de su proximidad en el orden. Cada nido comprende dos alternativas, y se añade un patrón de comportamiento de las correlaciones (más altas en pares cercanos). Un modelo llamado *Logit combinatorial por pares* (*paired combinatorial Logit*, PCL) también crea nidos por pares de alternativas, y cada uno tiene su propia correlación. El OGEV es un PCL con una función de proximidad entre pares de alternativas. También se ha desarrollado un *Logit anidado generalizado* (*generalized nested Logit*, GNL) que incluye el PCL y otros modelos de nidos cruzados como casos especiales.

Todos estos modelos (Logit, Logit anidado, PCL, GNL…) son miembros de la familia de modelos GEV. McFadden (1978) desarrolló un *procedimiento para generar modelos dentro de la familia GEV*, tanto los mencionados como otros que puedan adaptarse mejor a las necesidades de cada caso particular, y que veremos al final.

Tabla 5. *Propiedades comunes de los modelos de elección discreta*

<table>
<tr><th>Limitaciones del Logit</th><th colspan="2">⇒Respuesta</th><th>Limitaciones del Probit</th><th>⇒Respuesta</th></tr>
<tr><td>La propiedad IIA impone restricciones en los patrones de sustitución</td><td>⇒ Modelos de la familia GEV</td><td rowspan="3">Modelos Probit</td><td rowspan="3">Basado en la distribución normal</td><td rowspan="3">⇒Logit Mixto</td></tr>
<tr><td>No puede representar variaciones aleatorias en los gustos</td><td rowspan="2">⇒</td></tr>
<tr><td>No se puede usar con paneles cuando los factores no observados están correlacionados en el tiempo</td></tr>
</table>

Modelo Logit

El modelo Logit tiene tres importantes limitaciones (Train, 2003, p. 46, p. 101): 1) puede representar cambios en los gustos relacionados con características observadas en el individuo que toma decisiones, pero no diferencias en gustos relacionadas con características no observadas; 2) el modelo Logit implica sustituciones proporcionales entre alternativas, por lo que otras formas de sustitución entre alternativas no son posibles; 3) no puede emplearse cuando hay variables inobservadas correlacionadas temporalmente.

Veamos estas características del Logit por partes.

1.

Cuando se estiman las elecciones de los consumidores, los gustos y preferencias de estos por las distintas alternativas pueden ser modeladas por un Logit, siempre que varíen en función de características observadas de los individuos y recogidas en variables del modelo, pero no si esas preferencias varían aleatoriamente o responden a variables inobservadas o no presentes en el modelo.

Un ejemplo. La elección de coche por parte de las economías domésticas depende de dos atributos del coche que las familias observan: el tamaño de su cabina (TC_j) y el precio de venta (PV_j), donde *j* es el modelo de coche. Cada hogar valora ambos atributos de forma distinta, y hay *n* hogares. Tenemos que la función de utilidad de cada uno de ellos será:

$$U_{nj} = \beta^1{}_n TC_j + \beta^2{}_n PV_j + u_{nj}$$

Donde $\beta^1{}_n$ y $\beta^2{}_n$ son parámetros específicos para cada hogar. Los parámetros difieren porque reflejan distintas preferencias, en función de variables de las que tenemos datos, y más en concreto, de variable socioeconómicas de los mismos. Por ejemplo, supongamos que la valoración del tamaño de la cabina depende directamente del número de miembros del hogar (N_n), por lo que podemos hacer:

$$\beta^1{}_n = \sigma N_n$$

donde σ es una constante de proporcionalidad. Si la importancia del precio para un hogar determinado depende inversamente de la renta (R_n), de forma que los hogares con menos renta den más valor al factor precio, tendremos

$$\beta^2_n = \eta/R_n$$

Sustituyendo ahora los parámetros originales por sus equivalentes

$$U_{nj} = (\sigma N_n)TC_j + (\eta/R_n)PV_j + u_{nj}$$

Y de ahí

$$U_{nj} = \sigma(N_nTC_j) + \eta(PV_j/R_n) + u_{nj}$$

Modelo que puede estimarse mediante un Logit. Los parámetros que requieren estimación son σ y η.

La relación entre las variaciones en gustos y las variables observables pueden ser más complejas. Por ejemplo, la valoración del espacio de la cabina de los coches puede depender del tamaño familiar, pero a una tasa decreciente, de forma que $\beta^1_n = \sigma N_n + \varphi N_n^2$, siendo φ negativo. En ese caso tendríamos tres parámetros a estimar

$$U_{nj} = \sigma(N_nTC_j) + \varphi(N_n^2\, TC_j) + \eta(PV_j/R_n) + u_{nj}$$

El modelo Logit empieza a encontrarse con dificultades cuando las preferencias dependen de variables no observadas, o aleatorias (quizás por falta de información). Por ejemplo, la valoración del tamaño de la cabina del coche puede depender también de la altura y peso de los miembros de la familia, de la frecuencia con la que viajan todos juntos, de la existencia de otros vehículos, del tamaño del garaje, etcétera. Supongamos que no tenemos información sobre esas otras variables. En ese caso hay que añadir un término de error μ:

$$\beta^1_n = \sigma N_n + \mu_n$$

También el factor precio puede tener componentes aleatorios o no observados:

$$\beta^2_n = \eta/R_n + \varepsilon_n$$

Sustituyendo en el modelo original tenemos

$$U_{nj} = \sigma(N_nTC_j) + \mu_nTC_j + \eta(PV_j/R_n) + \varepsilon_nPV_j + u_{nj}$$

Donde los componentes μ_nTC_j y ε_nPV_j pasan a formar parte del error, por ser μ_n y ε_n no observables. Por tanto, pasamos a tener

$$U_{nj} = \sigma(N_nTC_j) + \eta(PV_j/R_n) + \theta_{nj}$$

Donde $\theta_{nj} = \mu_nTC_j + \varepsilon_nPV_j + u_{nj}$.

El problema es que este término de error θ_{nj} no puede ser independiente e idénticamente distribuido. La explicación está en que μ_n y ε_n entran en cada alternativa j, por lo que estas estarán correlacionadas entre sí: $Cov(\theta_{nj}, \theta_{nk}) = Var(\mu_n)TC_jTC_k + Var(\varepsilon_n)PV_jPV_k \neq 0$ para cualquier par de coches *j* y *k*.

Además de lo anterior TC_j y PV_j varían entre cada alternativa de coches, por lo que $Var(\theta_{nj})$ también lo hace, lo que viola el principio de distribución idéntica de los errores: $Var(\theta_{nj}) = Var(\mu_n)TC_j^2 + Var(\varepsilon_n)PV_j^2 + Var(u_{nj})$.

El modelo Logit proporciona una especificación inadecuada en los casos en los que la variación en los gustos se da en parte de forma aleatoria. Los modelos Logit son bastante robustos ante errores de especificación, por lo que incluso en el caso de variaciones aleatorias en los gustos el Logit podría proporcionar resultados satisfactorios sobre los gustos medios, aunque no hay garantía de ello. Pero incluso si así fuera, el Logit no da información fiable sobre cómo se distribuyen los gustos en torno a esos gustos medios, lo que puede ser importante en muchos casos en los que la media no es la información determinante[1]. Si ocurre esto hay que emplear un Probit.

---

[1] Por ejemplo, cuando se lanza un nuevo producto (un coche) dirigido a un público que no tiene los gustos medios.

2.

Veamos el segundo problema de los modelos Logit. Imaginemos que una característica de una de las alternativas, como el precio de un coche, cae. Podemos interpretarlo como una mejora de sus atributos. En ese caso la alternativa se hace más deseable y la probabilidad de ser elegida por los consumidores o las familias crece. Dado que las probabilidades de elegir cada alternativa debe sumar uno, la probabilidad de las demás debe reducirse. Este impacto en la estructura de probabilidades es muy importante en algunos casos.

Un ejemplo. Imaginemos que un fabricante lanza una cámara fotográfica digital con nuevas prestaciones: es esencial saber cómo afectará, por un lado, a las probabilidades de compra de sus otras cámaras y, por otro lado, a las probabilidades de compra de las cámaras de sus competidores. Quitar clientes a estos es más rentable que las sustituciones de producto dentro del grupo de consumidores de la firma, y tiene además un impacto en la cuota de mercado de la empresa. Estos son los "patrones de sustitución" (*substitution patterns*).

El caso es que el modelo Logit implica un determinado patrón de sustitución de alternativas. En el momento en que queramos investigar lo adecuado de distintos patrones de sustitución alternativos o cuando necesitemos alguno distinto del implícito en el Logit, tendremos que usar otros modelos con otros patrones de sustitución o simplemente, más flexibles.

El tema puede analizarse de dos formas: A), como una restricción en las razones de probabilidades; B), como una restricción sobre las elasticidades cruzadas de las probabilidades.

A.

Restricción en las razones de probabilidades: La *propiedad de independencia respecto de las alternativas irrelevantes*.

Vamos a replantear el modelo Logit pensando en la modelización del comportamiento del consumidor. Supongamos que el consumidor n-ésimo se enfrenta a j alternativas. La utilidad que obtiene el consumidor de la alternativa j tiene dos componentes, uno V_{nj} que depende de una serie de variables conocidas, más una parte desconocida ε_{nj} que se trata como una variable aleatoria. Por tanto

$$U_{nj} = V_{nj} + \varepsilon_{nj}$$

Se supone que cada ε_{nj} está independiente e idénticamente distribuida, según una distribución Gumbel o distribución de valores extremos de tipo I (erróneamente se dice también que según una Weibull, que es la de tipo III). La clave está en que *la diferencia* entre dos variables que siguen una distribución de valores extremos sigue una distribución logística. Por tanto, ***si*** ε_{nj} ***y*** ε_{ni} ***son variables aleatorias de valores extremos de tipo I, tenemos que*** $\varepsilon_{nji} = \varepsilon_{nj} - \varepsilon_{ni}$ ***se regirá por una distribución logística***. Si $\varepsilon_{nji} = z$ tendremos que, como ya sabemos,

$$G(z) = \exp(z) / [1 + \exp(z)]$$

El supuesto clave es la distribución independiente de los errores, lo que significa que la parte no observada de la utilidad en una alternativa no está relacionada con la parte no observada de la utilidad de las otras alternativas. Por tanto, *la utilidad no observada de una alternativa no nos da información sobre la de las demás*. Estas porciones de la utilidad no explicadas por el modelo deberían ser, idealmente, "ruido blanco". *El modelo Probit no requiere este supuesto, que es fuertemente restrictivo*. Otra posibilidad es reespecificar el modelo de la utilidad representativa (los V_{nj}), de forma que se cumpla el supuesto. Por último, puede estimarse el modelo Logit a sabiendas de que no se cumplen los requisitos. Las consecuencias son menos serias para la estimación de las preferencias medias y más serias para los patrones de sustitución.

La probabilidad de que un consumidor *n* elija la opción *i* es

$$P_{ni} = \mathrm{Prob}[U_{ni} > U_{nj} \ \forall j \neq i]$$

$$= \mathrm{Prob}[V_{ni} + \varepsilon_{ni} > V_{nj} + \varepsilon_{nj} \ \forall j \neq i]$$

$$= \mathrm{Prob}[\varepsilon_{nj} < V_{ni} + \varepsilon_{ni} - V_{nj} \ \forall j \neq i]$$

Que es la cdf de ε_{nj} para el punto $V_{ni} + \varepsilon_{ni} - V_{nj}$ siguiendo una distribución de Gumbel. Tomando dicha cdf para esa distribución se llega al resultado (Train, 2003, p. 40):

$$P_{ni} = \frac{e^{V_{ni}}}{\sum_j e^{V_{nj}}}$$

que es la probabilidad siguiendo la distribución logística, sólo que para *j* alternativas en vez de dos.

La utilidad representativa suele modelarse mediante una relación lineal en los parámetros, de forma que $V_{nj} = z = \beta x_{nj}$. Tenemos que P_{ni} está situada entre 0 y 1, que cuando V_{ni} crece P_{ni} se aproxima a 1, si V_{ni} decrece hasta $-\infty$ P_{ni} se aproxima a 0, la probabilidad de una sola alternativa no es nunca exactamente cero (si una alternativa no tiene posibilidades de ser elegida no es tal alternativa), y una probabilidad de 1 exacto sólo se da si el conjunto de alternativas solo cuenta con una posibilidad. Además, la suma de las probabilidades de todas las alternativas ($\sum P_{ni}$ para i = 1,…,j) es igual a 1.

La forma de S que tiene la relación entre la probabilidad Logit y la función de utilidad representativa tiene un significado en lo referente al impacto de un cambio en las variables explicativas sobre la utilidad. Si la utilidad de una alternativa es relativamente muy baja, un pequeño incremento de la misma tiene poco efecto sobre la probabilidad de ser elegida (las otras alternativas siguen siendo mucho mejores); si la utilidad es relativamente muy grande, un pequeño incremento adicional apenas afecta a las probabilidades de elección de cada una de ellas. Cuando la probabilidad está en 0,5 el impacto de un cambio en la utilidad es máximo.

Un ejemplo. Una familia (elemento *n*) tiene que elegir entre dos alternativas: calefacción por gas o eléctrica. La utilidad depende del coste de la instalación CI, del coste de mantenimiento CM, y de una serie de variables inobservables como la valoración estética del sistema una vez instalado o las opiniones sobre el carácter más o menos contaminante de uno y otro tipo de energía. Si la parte observada de la utilidad guarda una relación lineal con las variables observadas tendremos que

$$U_g = \beta_1 CI_g + \beta_2 CM_g + \varepsilon_g$$

$$U_e = \beta_1 CI_e + \beta_2 CM_e + \varepsilon_e$$

Donde es de esperar que $\beta_1<0$ y $\beta_2<0$. Si los componentes inobservados, considerados aleatorios, ε_g y ε_e están idéntica e independiente distribuidos según una distribución de valores extremos, la probabilidad de elegir gas será

$$P_g = \frac{e^{\beta_1 CI_g + \beta_2 CM_g}}{e^{\beta_1 CI_g + \beta_2 CM_g} + e^{\beta_1 CI_e + \beta_2 CM_e}} = \frac{1}{1 + e^{(\beta_1 CI_e + \beta_2 CM_e) - (\beta_1 CI_g + \beta_2 CM_g)}}$$

donde hemos aplicado la propiedad según la cual exp(a)/exp(b) = exp(a-b). Como es obvio, la elección multinomial es una extensión muy sencilla, y basta con añadir más términos a la suma del denominador, lo que, como es lógico, reduce la probabilidad de la opción que aparece en el numerador.

Es interesante señalar que la razón entre coeficientes tiene un significado especial. Así, β_2/β_1 representa la disposición a pagar un incremento en los costes de instalación a cambio de una reducción en los costes de mantenimiento de una unidad monetaria. La idea es que la utilidad permanece sin cambios, por lo que $dU = \beta_1\, dCI + \beta_2\, dCM = 0$, y de ahí obtenemos que $dCI/dCM = -\beta_2/\beta_1$

Para dos alternativas cualesquiera *i* y *k* la razón de probabilidades Logit es

$$\frac{P_{ni}}{P_{nk}} = \frac{\dfrac{e^{V_{ni}}}{\sum_j e^{V_{nj}}}}{\dfrac{e^{V_{nk}}}{\sum_j e^{V_{nj}}}} = \frac{e^{V_{ni}}}{e^{V_{nk}}} = e^{V_{ni}-V_{nk}}$$

esa razón sólo depende de las alternativas i y k, y no de cualesquiera otras que pueda haber. Es por tanto una razón independiente de las alternativas irrelevantes, por lo que el modelo Logit tiene esta propiedad de independencia de las alternativas irrelevantes (*independence from irrelevante alternatives*, IIA). La propiedad IIA puede ser realista en muchas situaciones en las que se plantean elecciones, pero puede ser inapropiada para muchas otras.

Un ejemplo. Una persona se plantea ir al trabajo en coche (c) o en un autobús color azul (aa). Imaginemos que $P_c = P_{aa} = 1/2$, de modo que $P_c/P_{aa} = 1$. Ahora imaginemos que se añade la posibilidad de un autobús rojo (ar), y que la persona lo considera una alternativa exactamente equivalente a la del autobús azul (le da igual), de forma que $P_{ar}/P_{aa} = 1$. En el modelo Logit, debido a la propiedad IIA, el ratio P_c/P_{aa} no cambia por la presencia de una nueva alternativa, por lo que siguen sin cambio los valores $P_c/P_{aa} = 1$ y $P_{ar}/P_{aa} = 1$. Esos dos valores exigen que, necesariamente, $P_c = P_{aa} = P_{ar} = 1/3$. Esto no se corresponde con lo que esperaríamos en la realidad. La aparición del autobús rojo no debería afectar la probabilidad de elegir el coche. Sería de esperar que la probabilidad de tomar un autobús se dividiera entre los dos, el azul y el rojo, si es verdad que a la persona le da igual el color del autobús. Por tanto, lo lógico sería $P_c = ½$ y $P_{aa} = P_{ar} = ¼$. Necesariamente el ratio P_c/P_{aa} debería cambiar con la aparición del autobús rojo, cosa que el modelo Logit no contempla.

Imaginemos que cambia ahora un atributo de las V_{ni}. El impacto de ese cambio en las probabilidades de todas las demás alternativas contenidas en j es una elasticidad cruzada. Podemos preguntarnos cómo aumenta la probabilidad de comprar un coche Honda una mejora en el consumo de gasolina. También por cómo se ve afectada, esta vez negativamente, la probabilidad de compra de un Toyota, que no ha experimentado esa mejora.

Suponemos que z_{ni} es un factor observable que entra en la función de utilidad del agente representativo. El impacto de un cambio de ese factor en la probabilidad de que el agente n elija la opción i es

$$
\begin{aligned}
\frac{\partial P_{ni}}{\partial z_{ni}} &= \frac{\partial\left(e^{V_{ni}} \Big/ \sum_j e^{V_{nj}}\right)}{\partial z_{ni}} \\
&= \frac{e^{V_{ni}}}{\sum_j e^{V_{nj}}} \frac{\partial V_{ni}}{\partial z_{ni}} - \frac{e^{V_{ni}}}{\left(\sum_j e^{V_{nj}}\right)^2} e^{V_{ni}} \frac{\partial V_{ni}}{\partial z_{ni}} \\
&= \frac{\partial V_{ni}}{\partial z_{ni}}\left(P_{ni} - P_{ni}^2\right) \\
&= \frac{\partial V_{ni}}{\partial z_{ni}} P_{ni}\left(1 - P_{ni}\right)
\end{aligned}
$$

Si la utilidad tiene una relación lineal con z_{ni} con un coeficiente β_z, la derivada queda reducida a

$$\partial P_{ni}/z_{ni} = \beta_z P_{ni}(1-P_{ni})$$

La derivada alcanza su valor máximo cuando $P_{ni} = (1-P_{ni})$, lo que se produce cuando $P_{ni} = ½$; la derivada se reduce conforme P_{ni} se aproxima a cero o a 1, ya que uno de los términos siempre se aproximará a cero. Esto concuerda con el hecho de que el impacto de un cambio en cualquier variable es mayor cuando tenemos dudas sobre qué elegir (probabilidades al 50%), y menor cuando nuestras preferencias se decantan fuertemente por o en contra de una las opciones.

El efecto cruzado de z_{nj} sobre la alternativa i se calcula de forma similar:

$$\frac{\partial P_{ni}}{\partial z_{nj}} = \frac{\partial\left(e^{V_{ni}} \Big/ \sum_k e^{V_{nk}} \right)}{\partial z_{nj}}$$

$$= -\frac{e^{V_{ni}}}{\left(\sum_k e^{V_{nk}}\right)^2} e^{V_{nj}} \frac{\partial V_{nj}}{\partial z_{nj}}$$

$$= \frac{\partial V_{nj}}{\partial z_{nj}} P_{ni} P_{nj}$$

Si la el componente explicable de la utilidad V_{nj} tiene una relación lineal con z_{ni}, con un coeficiente βz, la derivada queda reducida a

$$\partial P_{ni}/z_{ni} = -\beta_z P_{ni} P_{nj}$$

Si $\beta_z>0$ la propiedad z_{nj} será deseable, y un incremento de la misma reduce la probabilidad de elegir cualquier otra alternativa distinta de j. De ahí el signo negativo. Además, la reducción en la probabilidad de elección de la opción i es proporcional al valor de la probabilidad antes de que z_{nj} cambiara.

Dado que la suma de todas las probabilidades debe ser igual a uno, cuando cambia una variable observada los cambios provocados sobre las probabilidades tienen que sumar cero, es decir, la suma de las probabilidades resultantes tienen que volver a ser uno. En los modelos Logit esto se cumple, y es fácil de demostrar:

$$\sum_{i=1}^{j} \frac{\partial P_{ni}}{\partial z_{nj}} = \frac{\partial V_{nj}}{\partial z_{nj}} P_{nj}(1-P_{nj}) + \sum_{i\neq j}\left(-\frac{\partial V_{nj}}{\partial z_{nj}}\right) P_{nj} P_{ni}$$

$$= \frac{\partial V_{nj}}{\partial z_{nj}} P_{nj}\left[\left(1-P_{nj}\right) - \sum_{i\neq j} P_{ni}\right]$$

$$= \frac{\partial V_{nj}}{\partial z_{nj}} P_{nj}\left[\left(1-P_{nj}\right) - \left(1-P_{nj}\right)\right]$$

$$= 0$$

El concepto de elasticidad es más común entre economistas, entre otras cosas porque elimina las unidades de medida de las variables. La elasticidad de P_{ni} con respecto a z_{ni}, que es una de las variables que explican la utilidad de la alternativa i, será:

$$\begin{aligned} E_{iz_{ni}} &= \frac{\partial P_{ni}}{\partial z_{ni}} \frac{z_{ni}}{P_{ni}} \\ &= \frac{\partial V_{ni}}{\partial z_{ni}} P_{ni}(1-P_{ni}) \frac{z_{ni}}{P_{ni}} \\ &= \frac{\partial V_{ni}}{\partial z_{ni}} z_{ni}(1-P_{ni}) \end{aligned}$$

Si el componente explicable de la utilidad V_{ni} tiene una relación lineal con z_{ni}, con un coeficiente βz, tendremos

$$E_{iz_{ni}} = \beta_z z_{ni}(1-P_{ni})$$

Por otro lado, la elasticidad cruzada de P_{ni} con respecto a una variable z_{nj} que explica la utilidad de una alternativa j (V_{nj}):

$$\begin{aligned} E_{iz_{nj}} &= \frac{\partial P_{ni}}{\partial z_{nj}} \frac{z_{nj}}{P_{ni}} \\ &= -\frac{\partial V_{nj}}{\partial z_{nj}} z_{nj} P_{nj} \end{aligned}$$

Si el componente explicable de la utilidad V_{nj} tiene una relación lineal con z_{nj}, con un coeficiente βz, tendremos

$$E_{iz_{nj}} = -\beta_z z_{nj} P_{nj}$$

Esta elasticidad cruzada es igual para todas las alternativas i (no depende de i), por lo que un cambio en una de las variables que explican la utilidad de j altera las probabilidades de todas las demás alternativas en un mismo porcentaje. Esta es la *propiedad de IIA*. Una mejora en una alternativa "se alimenta" proporcionalmente de todas las demás.

Este patrón de sustitución se conoce como *patrón de sustitución proporcional*, y es un reflejo directo de la propiedad IIA: el ratio de probabilidades entre las alternativas *i* y *k* se mantiene constante cuando un atributo de la alternativa *j* cambia, y ello es así porque ambas probabilidades cambian en la misma proporción.

La propiedad IIA implica que, denotando 0 la posición antes del cambio y 1 la posición después de un cambio en un atributo z_{nj} de la alternativa j para el agente n,

$$\frac{P_{ni}^{1}}{P_{nk}^{1}} = \frac{P_{ni}^{0}}{P_{nk}^{0}}$$

lo que sólo puede darse si cada probabilidad cambia exactamente en la misma proporción.

El Logit y la sustitución proporcional van de la mano. Cuando este patrón de sustitución es razonable, el uso del Logit tiene sentido, pero no así en otros casos, en los que el Logit puede llevar a resultados poco realistas.

Un ejemplo. Si queremos evaluar el impacto de medidas de fomento de vehículos poco contaminantes, y tenemos tres alternativas (coches grandes de gas, coches pequeños de gas y coches pequeños eléctricos, con probabilidades de ser elegidos 0,66, 0,33 y 0,01 respectivamente), el impacto de un subsidio a los coches eléctricos que aumenta su probabilidad de 0,01 a 0,1, según el Logit, reduciría las probabilidades de las otras dos opciones en un 10%, al 0,6 y 0,3. Los coches grandes de gas reducen su probabilidad en 0,06 puntos, y los pequeños de gas en 0,03, por lo que la pérdida en los primeros es el doble que en los segundos, en términos absolutos. Esto no es muy realista ya que, en cierto sentido, los coches pequeños compiten entre sí más intensamente que con los coches grandes. Sería de esperar que la elasticidad cruzada para los coches pequeños de gas sea mayor que para los grandes, y no igual. *El modelo Logit haría una predicción errónea (exagerada) del gas ahorrado por el subsidio, pues pronosticaría una reducción de coches grandes demasiado alta*.

No obstante, la propiedad *IIA* del Logit tiene sus ventajas. Cuando es realista, es fácil estimar parámetros del modelo operando sobre una submuestra de posibles alternativas. La consistencia del estimador no resulta afectada por esta exclusión, ya que las probabilidades *relativas* en la submuestra no dependen de las demás alternativas y sus atributos. Otras veces podemos estar interesados sólo en un conjunto de alternativas. Por ejemplo, en el medio de transporte seleccionado por ejecutivos para ir de Madrid a Barcelona, considerando el AVE y el puente aéreo, y excluyendo del análisis otras muchas posibilidades (autobús, otros trenes, coche…).

Existen distintos tipos de test para comprobar empíricamente si la propiedad IIA se da en un determinado caso (McFadden, 1978). Una posibilidad es estimar el modelo con todas las alternativas y sólo con algunas, y comprobar si los ratios de probabilidades cambian de forma significativa (Hausman y McFadden 1984 proporcionan un estadístico para ello). Otra posibilidad es meter las variables (atributos) de la utilidad de una alternativa en otra alternativa. Si el supuesto IIA no se cumpliera dichas variables deberían ser significativas para la utilidad de esas otras alternativas (hay dos procedimientos, uno de McFadden 1987 y otro de Train, Ben-Akiva y Atherton, 1989). Con el tiempo se han desarrollado versiones más generales del Logit (los *generalized extreme value models*, GEV y el *mixed Logit*, que veremos más adelante), de los que el Logit es un caso particular que se obtiene mediante unas restricciones en los parámetros. La significatividad de éstas se puede comprobar fácilmente.

3.

En ocasiones podemos observar repetidas elecciones de un mismo individuo a lo largo del tiempo. Estos datos que muestran distintas elecciones para un mismo individuo son los *datos de panel*. También pueden presentarse ante el individuo una serie de alternativas hipotéticas, cuyos atributos pueden ir variando para recoger el impacto de dichos cambios en la elección del individuo ("stated preferences" o preferencias declaradas).

El modelo Logit puede emplearse con paneles de datos siempre que se cumpla el requisito de que los factores inobservados que afectan a las elecciones sean independientes entre las distintas elecciones efectuadas a lo largo del tiempo. Que los factores observados entre elecciones puedan estar relacionados –influyendo las elecciones pasadas en las presentes (*state dependence*) o con retardos en los efectos de los cambios de atributos en las elecciones– no presenta problemas, pero los inobservados no pueden estarlo.

La generalización es sencilla. La utilidad que un individuo *n* obtiene de la alternativa *j* en el momento *t* es

$$U_{njt} = V_{njt} + \varepsilon_{njt} \qquad \forall\ j,t$$

Donde ε_{njt} sigue una distribución de valores extremos, y debe ser independiente entre los distintos *n*, *j* y también *t*.

La probabilidad de elección será ahora

$$P_{nit} = \frac{e^{V_{nit}}}{\sum_j e^{V_{njt}}}$$

Si la utilidad representativa viene explicada para cada individuo sólo por variables referidas a un mismo período t, entonces $V_{njt} = \beta\mathbf{x}_{njt}$ donde $\mathbf{x}_{njt}$ es un vector de variables que afectan al individuo n que se enfrenta a la alternativa j en el momento t, y se puede usar un Logit igual con datos de sección cruzada y con datos de panel.

Los supuestos del modelo Logit no impiden que una de las variables que componen $\mathbf{x}_{njt}$ provenga de un período anterior, o posterior. Incluso la variable *dependiente* en un período anterior puede incorporarse, para representar *hábitos* o inercias en el consumo, que requieren de un estímulo suficientemente fuerte para romperse: $V_{njt} = \alpha y_{nj(t-1)} + \beta\mathbf{x}_{njt}$, siendo $y_{nj(t-1)} = 1$ si la opción j se eligió en el período t-1. Hay muchas posibles variantes de lo mismo: Adamowicz (1994) el número de veces que esa alternativa se ha seleccionado en el pasado; Erdem (1996) los atributos de las alternativas elegidas en el pasado, para considerar la similaridad entre alternativas. Si los errores son independientes en el tiempo, no habrá inconsistencia en la estimación. *Gracias a ese supuesto la variable $y_{nj(t-1)}$ y ε_{njt} están incorrelacionadas*. Es una restricción fuerte.

Si los factores inobservados están correlacionados en el tiempo hay que descartar el Logit y optar por un Logit mixto o por un Probit, a no ser que podamos explicitar esas dependencias de forma que el error que quede cumpla con la propiedad de independencia en el tiempo.

Modelo Logit Anidado

Cuando la propiedad IIA no se cumple, porque quedan variables explicativas por explicitar, lo que genera correlaciones que violan el supuesto iid, hay que emplear modelos más generales. Los ***modelos generalizados de valores extremos*** **(*generalized extreme value models*, GEV)** son una familia de modelos con diversos patrones de sustitución. Todos tienen en común que las *partes no observadas* que explican la utilidad de todas las alternativas están distribuidas *conjuntamente* según una *distribución de valores extremos*, lo que permite correlaciones entre alternativas. Si todas las correlaciones fueran cero, la distribución GEV se podría expresar como el producto de distribuciones de valores extremos independientes, y el GEV se reduce a un Logit estándar. Se pueden hacer test sobre las correlaciones en un modelo GEV, y comprobar así si estas son cero o no. Otra ventaja de los GEV es que las probabilidades *tienen una solución analítica*, por lo que no es necesario recurrir a la simulación para su estimación.

Dentro de la familia de modelos GEV, además del propio Logit, el más extendido es el **Logit anidado (*nested Logit*).** Su principal ventaja es que su forma funcional es relativamente simple, comparada con otros modelos de la familia GEV[2], y permite un conjunto amplio y variado de patrones de sustitución.

Un *Logit anidado* es apropiado para el caso en que el conjunto de alternativas al que se enfrenta un decisor se puede dividir en subconjuntos, llamados *nidos*, de forma que:

1. Para dos alternativas que están en el mismo nido la razón entre probabilidades es independiente de los atributos o existencia de otras alternativas, o lo que es lo mismo, dentro de cada nido se cumple la propiedad IIA.

2. Para dos alternativas situadas en nidos distintos la razón de probabilidades puede depender de los atributos de las otras alternativas en cualquiera de los dos nidos, esto es, IIA no se cumple en general para alternativas de nidos distintos.

Un ejemplo. Un trabajador tiene cuatro opciones para acudir al trabajo: coche, andando, autobús o tren. Si una alternativa desaparece las restantes alternativas ven aumentada su probabilidad. La pregunta clave a la hora de dividir en nidos es cómo cambian esas probabilidades. La siguiente tabla muestra un caso hipotético (Train, 2003, p. 82):

[2] Como puede ser el caso de los Logit combinatoriales por pares (*paired combinatorial Logit*, PCL), o el Logit anidado generalizado (*generalized nested Logit*, GNL). Aún no se han explorado todos los posibles modelos de la familia GEV.

	Original	Quitando una alternativa			
		Coche	Andando	Autobús	Tren
Coche	0,40	-	0,45 (+12,5%)	0,52 (+30%)	0,48 (+20%)
Andando	0,10	0,20 (+100%)	-	0,13 (+30%)	0,12 (+20%)
Autobús	0,30	0,48 (+60%)	0,33 (+10%)	-	0,40 (+33%)
Tren	0,20	0,32 (+60%)	0,22 (+10%)	0,35 (+70%)	-

Las alternativas de autobús y tren cambian en una misma proporción cuando se elimina la posibilidad de ir en coche o andando (60% y 10% respectivamente). Por tanto, para esas dos alternativas se cumple la propiedad IIA y se puede crear un nido llamado, por ejemplo, "público". El coche y el ir andando están asociados de la misma forma frente a la eliminación de la posibilidad del autobús o el tren (30% y 20% respectivamente), por lo que se puede formar un nido de nombre "privado" con ellas. Cuando desaparece la posibilidad de ir en coche la probabilidad de ir andando crece más que la de tomar un autobús o un tren (100% frente a 60%). Por ello ir andando *no* pertenece al mismo nido.

Se ha demostrado que este modelo Logit anidado es compatible con la maximización de utilidad (por ejemplo, McFadden, 1978). Si dividimos el conjunto de alternativas posibles *j* en *k* subconjuntos disjuntos llamados B_1, B_2, …, B_k, éstos serán los nidos. La utilidad que una persona n obtiene de la alternativa *j* alojada en el nido B_k será

$$U_{nj} = V_{nj} + \varepsilon_{nj}$$

Donde V_{nj} es observable y ε_{nj} no lo es. El modelo Logit anidado se obtiene asumiendo que el vector de utilidades inobservado, $\varepsilon_n = (\varepsilon_{n1}, \varepsilon_{n2}, \ldots, \varepsilon_{nj})$ tiene una *función de distribución de probabilidades* (cdf):

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

que es un tipo de distribución de valores extremos GEV, una generalización de la que genera el Logit. En éste cada ε_{nj} es independiente y sigue una distribución de valores extremos. En el modelo GEV, en cambio, los ε_{nj} están correlacionados dentro de cada nido. Para dos alternativas cualesquiera *j* y *m* dentro del nido B_k, ε_{nj}

estará correlacionado con ε_{nm}. En cambio, para dos alternativas situadas en nidos distintos los ε_{nj} no estarán correlacionados, es decir, $cov(\varepsilon_{nj}, \varepsilon_{nm})=0$ para cualquier *j* que pertenezca a B_k y cualquier m que pertenezca a $B\ell$, siendo $k\neq\ell$.

El parámetro λ_k es una medida del grado de independencia de la utilidad inobservada entre alternativas dentro de la cesta B_k. Cuanto mayor es el valor de λ_k, hasta 1, mayor será la independencia y menor la correlación dentro de B_k, por lo que ese parámetro, o $1-\lambda_k$, puede considerarse un índice de la correlación. Cuando $\lambda_k = 1$ para todos los k, de forma que la independencia se da entre todas las alternativas de todos los nidos, la distribución GEV se convierte en el producto de términos de valores extremos independientes, con una distribución de probabilidad de tipo Gumbel: $F(\varepsilon_{nj}) = e^{-e^{-\varepsilon_{nj}}}$. Entonces, el Logit anidado se convierte en un Logit normal.

La distribución arriba mencionada, para un Logit anidado, lleva a una probabilidad de elección de la alternativa i del nido B_k:

$$P_{ni} = \frac{e^{V_{ni}/\lambda_k}\left(\sum_{j\in B_k} e^{V_{ni}/\lambda_k}\right)^{\lambda_k - 1}}{\sum_{\ell=1}^{K}\left(\sum_{j\in B_k} e^{V_{ni}/\lambda_k}\right)^{\lambda_\ell}}$$

Consideremos dos alternativas, $i \in B_k$ y $m \in B_\ell$. Dado que el denominador es igual para todas las alternativas, la razón de probabilidades es la razón de los numeradores:

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

Si $k=\ell$, es decir, si *m* e *i* están en el mismo nido, los términos entre paréntesis se igualan y nos queda la ya conocida expresión, independiente de cualquier otra alternativa distinta de ese par

$$\frac{P_{ni}}{P_{nm}} = \frac{e^{V_{ni}/\lambda_k}}{e^{V_{nm}/\lambda_\ell}}$$

Si $k \neq \ell$, es decir, si *m* e *i* no están en el mismo nido, los términos entre paréntesis no se igualan y la razón de probabilidades dependerá del resto de alternativas en las cestas de *m* e *i* y sus atributos, aunque no de los atributos de las alternativas que están en otras cestas.

Existe por tanto una forma de propiedad IIA para alternativas de nidos distintos, algo así como una "independencia de nidos irrelevantes" (IIN). *Con un modelo Logit anidado la propiedad IIA se mantiene para las alternativas dentro de cada nido, y la propiedad IIN para las alternativas situadas en distintos nidos.* Cuando $\lambda_k = 1$ para todos los k (y por tanto $1 - \lambda_k = 0$ para todos los k), no hay correlación entre las utilidades no observadas dentro de cada nido, y las probabilidades de selección son las del Logit normal, que no es más que un caso particular, siendo el anidado más general en el sentido de que permite un determinado patrón de correlaciones entre las utilidades no observadas.

El parámetro λ_k puede diferir entre nidos, lo que indica que las correlaciones entre utilidades inobservadas son distintas dentro de cada nido. Sobre cualquier restricción de igualdad entre estos parámetros se pueden hacer contrastes de hipótesis. Si imponemos la restricción de que $\lambda_k = 1$ para todos los k nidos, y hacemos un contraste de hipótesis, estaremos contrastando si el modelo Logit normal es una especificación razonable frente al Logit anidado. Por otro lado, el valor de λ_k debe moverse entre cero y uno para que el modelo sea coherente con la maximización de utilidad para *todos* los valores posibles de las variables explicativas. Cuando $\lambda_k < 0$ una mejora en uno de los atributos de una alternativa reduce la probabilidad de elegir otras, lo que contradice el comportamiento maximizador de la utilidad. No se han diferenciado los λ_k entre individuos, cuando en verdad podrían diferir, según las características de los mismos. Por tanto, el parámetro λ_k se puede poner en función de distintas variables, siempre que λ_k sea positivo.

La probabilidad de elección puede expresarse de otra forma. Descomponemos la parte observable de la utilidad: una constante para todas las alternativas dentro de un nido (W) y otra parte que varía entre alternativas dentro del nido. La utilidad sería ahora

$$U_{nj} = W_{nk} + Y_{nj} + \varepsilon_{nj}$$

Para $j \in B_k$. Tenemos que W_{nk} depende sólo de las variables que caracterizan el nido B_k (estas variables varían entre nidos, pero no entre alternativas dentro de un mismo nido). Por otro lado, Y_{nk} depende de variables que caracterizan la alternativa *j* (estas variables varían entre alternativas dentro de B_k). Se verifica que, con carácter general, $Y_{nj} \equiv V_{nj} - W_{nk}$, con lo que siempre podremos hacer esta reformulación a partir del planteamiento básico según el cual $U_{nj} = V_{nj} + \varepsilon_{nj}$.

Gracias a esta separación de V_{nj} en dos términos podemos escribir la probabilidad del Logit anidado como el producto de dos probabilidades Logit estándar. Así, la probabilidad de elegir una alternativa *i* que forma parte del nido B_k se podrá expresar como el producto de la probabilidad de elegir cualquiera de las alternativas del nido B_k por la probabilidad de que sea elegida precisamente la alternativa *i* dado que se ha seleccionado previamente el nido B_k. Por tanto:

$$P_{ni} = P_{ni/B_k} P_{nB_k}$$

Toda probabilidad puede escribirse como el producto de una probabilidad *marginal* (P_{nBk}) y una probabilidad *condicional* ($P_{ni/Bk}$). A la probabilidad marginal, para elección de nido, se la conoce también como el modelo superior (*upper model*), y a la probabilidad condicional (elección de alternativas dentro del nido elegido) como el modelo inferior (*lower model*). En el caso del Logit anidado, estas probabilidades adoptan la forma de probabilidades de elección Logit estándar. En efecto,

$$P_{nB_k} = \frac{e^{W_{nk} + \lambda_k I_{nk}}}{\sum_{\ell=1}^{K} e^{W_{n\ell} + \lambda_\ell I_{n\ell}}}$$

$$P_{ni\backslash B_k} = \frac{e^{Y_{ni}/\lambda_k}}{\sum_{j \in B_k} e^{Y_{nj}/\lambda_k}}$$

siendo

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

Para una demostración de la igualdad $P_{ni} = P_{ni/B_k} P_{nB_k}$ véase Train (2003, p. 90).

Intuitivamente el resultado es simple: primero se aplica un Logit para elegir entre alternativas formadas por nidos, y la probabilidad asociada depende de W_{nk}, que son variables que varían entre nidos, pero no entre alternativas alojadas en cada nido; después se aplica otro Logit para elegir entre alternativas que forman parte de un mismo nido, que ha sido previamente seleccionado, lo que depende de Y_{nj}, variables que varían entre alternativas que pertenecen a ese nido. Por su parte, el elemento I_{nk} une el modelo inferior con el modelo superior, llevando información del inferior al superior. Se trata del logaritmo del denominador del modelo inferior, y el logaritmo del denominador de la probabilidad de elección de un Logit es la utilidad que el decisor espera obtener de la situación de elección. Se le conoce como *valor inclusivo*, *utilidad inclusiva* o *precio inclusivo* del nido B_k. Por otro lado, λ_k refleja el grado de independencia entre las utilidades inobservadas de las alternativas del nido B_k, indicando un λ_k más bajo menos independencia (más correlación). De la misma forma, $\lambda_k I_{nk}$ es la utilidad esperada por el decisor *n* de la elección de alternativas en el nido B_k. La idea es que la probabilidad de elegir el nido B_k depende de la utilidad que el decisor espera obtener al elegir entre las alternativas de B_k. Esta utilidad esperada incluye: la utilidad que recibe sin importar la alternativa del nido que finalmente elige (W_{nk}), más la utilidad esperada por estar en disposición de elegir la mejor alternativa del nido ($\lambda_k I_{nk}$).

Si los coeficientes en el modelo inferior no se dividen por λ_k la elección de probabilidades cambia, y la igualdad del producto de probabilidades marginales y condicionadas con las probabilidades Logit ya no se da. Esto no quiere decir que el modelo sin λ_k sea inapropiado, aunque se pueden dar casos en que así sea. Por ejemplo, si no dividimos podemos encontrar que el modelo no es consistente con la maximización de utilidad cuando algún coeficiente es común entre nidos (por ejemplo, un coeficiente compartido entre dos medios de transporte), o que algunos resultados con contraintuitivos. Stata incluye el λ_k en su comando nLogit.

Hay modelos Logit anidados de tres niveles, o más, en vez de los dos que acabamos de ver (superior e inferior). En el ejemplo que hemos visto había que elegir primero entre el transporte público o privado, y después entre cada una de sus opciones. Pero puede haber casos en los que sean necesarios más niveles de decisión. Estos modelos se forman dividiendo el conjunto de alternativas en nidos (k), y éstos en subnidos (m). Las fórmulas se complican (pueden verse en McFadden, 1978, y Ben-Akiva y Lerman, 1985). No obstante, las probabilidades de elección también pueden expresarse como el producto de probabilidades Logit. El modelo superior describe la elección de nidos, el medio la elección de subnidos, y el inferior la elección de alternativas dentro de cada subnido. El superior incluye un *valor inclusivo* para cada nido ($\lambda_k \sigma_{mk}$), que representa la utilidad esperada por el decisor de los subnidos que hay en cada nido. Se calcula como el logaritmo del denominador del modelo intermedio para cada nido. Los modelos intermedios incluyen otro valor inclusivo para cada subnido (σ_{mk}), que representa la utilidad esperada de las alternativas dentro de cada subnido. Se calcula como el logaritmo del denominador del modelo inferior para cada subnido. Para garantizar la consistencia con el comportamiento maximizador de la utilidad del consumidor para todos los valores de las variables, es necesario que $0<\lambda_k<1$ y $0<\sigma_{mk}<1$. Los valores negativos no son compatibles con dicho comportamiento maximizador, y los superiores a 1 sólo lo son para un rango determinado de valores de las variables explicativas.

Un ejemplo. Una familia que tiene que elegir qué casa comprar decide sobre diversos parámetros. Uno de ellos puede ser el barrio, con muchas alternativas. Se supone que hay factores inobservables comunes a todas las alternativas de un mismo barrio (seguridad, entretenimiento, accesibilidad), por lo que se esperaría una cierta correlación entre todas las utilidades inobservadas de las mismas. Otro criterio puede ser el número de habitaciones de la casa, y también aquí pueden esperarse correlaciones entre las utilidades inobservadas (más habitaciones permiten trabajar en casa, por ejemplo), por lo que las alternativas en un mismo barrio con las mismas habitaciones deberían tener correlaciones aún más fuertes entre las utilidades inobservadas. Esta circunstancia recomienda separar los barrios en nidos, y el número de habitaciones en subnidos. Habría tres modelos anidados: el primero para elegir barrio, el segundo para elegir número de habitaciones dado el barrio, y el tercero para elegir la casa en concreto, dado el barrio y número de habitaciones seleccionados.

Las probabilidades cumplen la propiedad IIA de la siguiente forma. La razón de probabilidades para dos apartamentos en el mismo barrio y con el mismo número de habitaciones es independiente de las características del resto de apartamentos en el mercado. Si baja el precio de un apartamento de dos habitaciones en Malasaña, baja la probabilidad de seleccionar un apartamento de una habitación en el barrio de Salamanca en la misma proporción. La razón de probabilidades de dos apartamentos en el mismo barrio pero con distinto número de habitaciones es independiente de las características de los apartamentos en otros barrios, pero sí depende de las características de los apartamentos en el mismo barrio que tienen el mismo número de habitaciones que cualquiera de las dos alternativas en cuestión.

Otros Modelos GEV Más Generales

En realidad, los nidos no tienen por qué ser conjuntos disjuntos. Una alternativa puede pertenecer a más de un nido. En el ejemplo del transporte privado y público, el coche puede compartir atributos inobservados con el autobús, relacionados con el estado de las carreteras, el riesgo de retrasos por embotellamientos, etc. Podría definirse un tercer nido de "transporte por carretera" con el coche y el autobús, que ya están en los nidos de "privado" y "público". Se han especificado distintos modelos GEV para nidos con intersecciones o nidos cruzados. Un planteamiento, para alternativas ordenadas (como el número de coches que posee una familia, que puede ser 0, 1, 2…), es el **modelo de valores extremos generalizado ordenado (OGEV, *ordered generalized extreme value*)** en el que la correlación entre las utilidades no observadas entre dos alternativas depende de su proximidad en el orden. Cada nido comprende dos alternativas, y se añade un patrón de comportamiento de las correlaciones (más altas en pares cercanos). Un modelo llamado ***Logit combinatorial por pares* (*paired combinatorial Logit*, PCL)** también crea nidos por pares de alternativas, y cada uno tiene su propia correlación. El OGEV es un PCL en el que λ_{ij} sigue una función de proximidad entre *i* y *j*. También se ha desarrollado un ***Logit anidado generalizado* (*generalized nested Logit*, GNL)** que incluye el PCL y otros modelos de nidos cruzados como casos especiales.

El PCL se basa en la construcción de nidos por pares de alternativas, de la siguiente forma: si hay *j* alternativas, cada alternativa forma pares con todas las demás, estando presente en *j-1* nidos. Un parámetro λ_{ij} indica el grado de independencia entre dos alternativas, *i* y *j*. El significado es el mismo del parámetro λ_k en el caso del Logit anidado: indica el grado de independencia entre alternativas dentro del nido B_k, siendo 1- λ_k un índice de la correlación existente. De la misma forma que el Logit mixto, el PCL se convierte en un Logit estándar cuando λ_{ij} = 1 para todo par de alternativas i y j. La probabilidad de elección para el modelo PCL es

$$P_{ni} = \frac{\sum_{j \neq i} e^{V_{ni}/\lambda_{ij}} \left(e^{V_{ni}/\lambda_{ij}} + e^{V_{nj}/\lambda_{ij}} \right)^{\lambda_{ij}-1}}{\sum_{k=1}^{J-1} \sum_{\ell=k+1}^{J} \left(e^{V_{nk}/\lambda_{k\ell}} + e^{V_{n\ell}/\lambda_{k\ell}} \right)^{\lambda_{k\ell}}}$$

La suma del numerador afecta a todos los j-1 nidos en los que la alternativa i está presente. El PCL es como el Logit anidado excepto en que se permite aquí que i esté en más de un nido. El denominador también adopta la forma de un Logit anidado, es decir, la suma de todos los nidos, cada uno con su suma de exp(V/λ) elevada a la potencia λ. Se puede contrastar la hipótesis de que λ_{ij} = 1 para todo par de alternativas i y j o para algún conjunto de pares. Si se acepta la hipótesis no habrá evidencia de una correlación entre las utilidades inobservadas de ese par. También se pueden contrastar otras hipótesis, como igualdades entre grupos de parámetros

λ. Cuando el número de alternativas J es muy grande se hace necesario imponer una cierta estructura en los λ_{ij} para reducir su número, aunque es esta abundancia la que hace del PCL un modelo tan flexible. El número total de parámetros λ_{ij} es igual a J(J-1)/2, es decir, uno por cada par de alternativas, teniendo en cuenta que $\lambda_{ij} = \lambda_{ji}$. El número de covarianzas estimables es una menos, es decir, J(J-1)/2-1, por lo que hay que imponer al menos una restricción sobre los λ_{ij} (suele hacerse una normalización, igualando uno de los parámetros a 1).

El GNL se caracteriza porque cada alternativa puede estar en más de un nido, pero esta vez *en distintos grados*. Hay un parámetro de asignación α_{jk} (≥ 0) que indica el grado en el que la alternativa j es miembro del nido k. Si el parámetro tiene valor cero la alternativa no está presente en k. La suma de todos los parámetros para el nido k es igual a 1, es decir, $\sum_k \alpha_{jk} = 1 \ \forall j$, lo que permite interpretar cada parámetro como la porción de cada alternativa que se localiza en cada nido.

De nuevo aparece el parámetro λ_k indicando el grado de independencia entre alternativas que pertenecen a un mismo nido *k*. Un valor mayor para ese parámetro indica más independencia y menos correlación.

La probabilidad de que la persona *n* escoja la alternativa *i* es

$$P_{ni} = \frac{\sum_k \left(\alpha_{ik} e^{V_{ni}}\right)^{1/\lambda_k} \left(\sum_{j \in B_k} \left(\alpha_{jk} e^{V_{nj}}\right)^{1/\lambda_k}\right)^{\lambda_k - 1}}{\sum_{\ell=1}^{K} \left(\sum_{j \in B\ell} \left(\alpha_{j\ell} e^{V_{nj}}\right)^{1/\lambda_\ell}\right)^{\lambda_\ell}}$$

Es una expresión muy similar a la del Logit anidado, sólo que el numerador es la suma de todos los nidos que contienen la alternativa *i*, según sus "pesos" en los mismos. Es obvio que si cada alternativa entra sólo en un nido, de forma que $\alpha_{jk} = 1$ el modelo se convierte en un Logit anidado. Si además de eso $\lambda_k = 1$ para todos los nidos, el modelo se convierte en un Logit estándar.

Para facilitar la interpretación del GNL la probabilidad puede descomponerse de la siguiente forma:

$$P_{ni} = \sum_k P_{ni \backslash B_k} P_{nk}$$

donde tenemos una probabilidad asociada a elegir un nido k particular:

$$P_{ni} = \frac{\left(\sum_{j\in B_k}\left(\alpha_{jk}e^{V_{nj}}\right)^{1/\lambda_k}\right)^{\lambda_k}}{\sum_{\ell=1}^{K}\left(\sum_{j\in B\ell}\left(\alpha_{j\ell}e^{V_{nj}}\right)^{1/\lambda_\ell}\right)^{\lambda_\ell}}$$

y una probabilidad asociada a elegir la alternativa i dentro del nido *k*, que es:

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

Todos estos modelos (Logit, Logit anidado, PCL, GNL…) son miembros de la familia de modelos GEV. McFadden (1978) desarrolló una **procedimiento para generar modelos dentro de la familia GEV**, tanto los mencionados como otros que puedan adaptarse mejor a las necesidades de cada caso particular.

Omitiremos la alusión al decisor *n*, y definimos, para simplificar aún más la notación, $\exp(V_j) \equiv Y_j$, que es necesariamente positivo. Ahora definimos una función G que depende de todos los Y_j, de forma que $G = G(Y_1, Y_2, \ldots, Y_j)$ y $G_i = \partial G/\partial Y_i$. Si G satisface determinadas condiciones, podremos decir que

$$P_i = Y_i G_i / G$$

donde P_i es la probabilidad de elección para un modelo de elección discreta consistente con la maximización de utilidad. Todos los modelos así generados serán de la familia GEV. ¿Cuáles son esas condiciones?

1. $G\geq 0$ para todos los valores positivos de las Y_j $\forall j$.

2. G es homogénea de grado uno. Esto quiere decir que si todas las Y_j se multiplican por ρ tendremos que G aumenta en la misma proporción, es decir, $G = G(\rho Y_1, \rho Y_2, \ldots, \rho Y_j) = G = \rho G(Y_1, Y_2, \ldots, Y_j)$.

3. $G\rightarrow\infty$ cuando $Y_j\rightarrow\infty$ para cualquier j.

4. Las derivadas cruzadas parciales de G cambian de signo de una forma determinada: $G_i\geq 0$ para todo i; G_{ij} $(=\partial G_i/\partial Y_j)\leq 0$ para todo $j\neq i$; G_{ijk} $(=\partial G_{ij}/\partial Y_k)\geq 0$ para todo $j\neq i\neq k$; etc., alternándose los signos de la misma forma para las derivadas cruzadas siguientes.

Estas propiedades no tienen interpretación económica, aunque son fácilmente comprobables. La falta de sentido económico dificulta la especificación de la función G a emplear en cada caso, aunque la abstracción permite desarrollar y probar cualquier tipo de especificación que cumpla con los requisitos.

Para el Logit, la función G es

$$G = \sum_{j=1}^{J} Y_j$$

Siendo $G_i = \partial G / \partial Y_i = 1$, tenemos que

$$\begin{aligned} P_i &= \frac{Y_i G_i}{G} \\ &= \frac{Y_i}{\sum_{j=1}^{J} Y_j} \\ &= \frac{e^{V_i}}{\sum_{j=1}^{J} e^{V_j}} \end{aligned}$$

que es la probabilidad para el Logit.

La función G para el Logit anidado es

$$G = \sum_{\ell=1}^{K} \left(\sum_{j \in B_\ell} Y_j^{1/\lambda_\ell} \right)^{\lambda_\ell}$$

donde las J alternativas están distribuidas en K conjuntos disjuntos ($B_1, B_2, \ldots, B_k$), y cada λ_k se mueve entre cero y 1. La primera derivada con respecto a Y_i para $i \in B_k$ es

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

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

Y haciendo las sustituciones oportunas en $P_i = Y_i G_i / G$, incluyendo $\exp(V_j) \equiv Y_j$ tenemos la expresión ya vista de la probabilidad para el modelo Logit anidado.

De la misma forma el PCL parte de una función G del tipo

$$G = \sum_{k=1}^{J-1} \sum_{\ell = k+1}^{J} \left(Y_k^{1/\lambda_{k\ell}} + Y_\ell^{1/\lambda_{k\ell}} \right)^{\lambda_{k\ell}}$$

Mientras que el GNL parte de una del tipo

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

Véase Train (2003, p. 100) para la obtención de las probabilidades, y la verificación de las propiedades de las funciones G.