MODELO LOGIT MIXTO

El Logit mixto puede aproximar cualquier *modelo de utilidad aleatorio* (RUM), dada su gran flexibilidad. Supera las tres limitaciones tradicionales del Logit: permite la variación aleatoria en los gustos; los patrones de sustitución sin restricciones ni limitaciones; y la correlación en los factores inobservados a lo largo del tiempo. Además, presenta una ventaja respecto al Probit, y es que no depende de la distribución normal. En general es fácil de desarrollar y la simulación de las probabilidades de elección es fácil de calcular (cosa que no siempre pasa con el Probit).

La simulación ha sido la gran limitación del Logit mixto, como del Probit, por los problemas de computación. Su primera aplicación está en Boyd y Mellman (1980) y Cardell y Dubar (1980), aunque más que datos individuales, se trataba de un caso (la demanda de modelos de automóviles), y se emplearon datos a nivel de mercado. Más tarde, Train et al. (1987) y Ben-Akiva *et al*. (1993) emplearon datos individuales de consumidores, pero ello gracias a que la integral sólo tenía dos niveles de integración, lo que permitía el cálculo por cuadratura. A partir de ahí, las mejoras en computación y técnicas de simulación ha permitido aplicar los Logit mixtos con toda su potencia (véanse referencias en Train, 2003, página 138).

Cualquier modelo de comportamiento se puede tratar mediante un Logit mixto. Lo que tienen de común todos los Logits mixtos es la forma funcional de la probabilidad de elección, que es la integral de las probabilidades de un Logit estándar a lo largo de una densidad de parámetros, es decir, con la forma:

$$P_{ni} = \int L_{ni}(\beta)\, f(\beta)\, d\beta$$

Donde $L_{ni}(\beta)$ es la probabilidad de elección Logit evaluada al valor del parámetro β

$$L_{ni}(\beta) = \frac{e^{V_{ni}(\beta)}}{\sum_{j=1}^{J} e^{V_{nj}(\beta)}}$$

y $f(\beta)$ es una función de densidad, no necesariamente normal. $V_{ni}(\beta)$ es la porción observada de la utilidad, que depende del parámetro β. Si la utilidad es lineal en β tendremos que $V_{ni}(\beta) = \beta' x_{ni}$, en cuyo caso las probabilidades del Logit mixto toman la forma

$$P_{ni} = \int \left(\frac{e^{\beta' x_{ni}}}{\sum_j e^{\beta' x_{nj}}} \right) f(\beta) d\beta$$

Por tanto, la probabilidad del Logit mixto es una media ponderada de la formula del Logit para la probabilidad, pero evaluada a distintos valores de β, viniendo dados los pesos por la función de densidad *f*(β). En la literatura estadística, la media ponderada de varias funciones recibe el nombre de *mixed function*, y a la función de densidad que proporciona los pesos se llama *mixing distribution*. El Logit mixto es una mezcla de la función del Logit evaluada a diferentes valores de β con *f*(β) como la *mixing distribution*.

El modelo Logit estándar es un caso particular del Logit mixto. Se da cuando la *mixing distribution* *f*(β) es degenerada para parámetros fijos b: *f*(β) = 1 para β = b y 0 para β ≠ b. En ese caso la probabilidad de elección pasa a ser la fórmula típica del Logit

$$P_{ni} = \frac{e^{b' x_{ni}}}{\sum_j e^{b' x_{nj}}}$$

Pero hay muchas más posibilidades para la *mixing distribution* *f*(β). Puede ser una función discreta, tomando β un conjunto finito de valores concretos. Si β toma M valores diferentes, tales que b_1, b_2,..., b_M, podemos definir una probabilidad para cada uno de ellos, de forma que la probabilidad de que β = b_m sea s_m. Bajo estas condiciones el Logit mixto se convierte en un modelo de clase latente, muy popular en psicología y en marketing. La elección de probabilidades en un caso como ese sería

$$P_{ni} = \sum_{m=1}^{M} s_m \left(\frac{e^{b_m' x_{ni}}}{\sum_j e^{b_m' x_{nj}}} \right)$$

Esto es útil para estimar casos en los que la población se divide en M segmentos, cada uno de ellos con sus propias funciones de comportamiento, o preferencias. La s_m sería la proporción de la población total en el segmento m. Esa probabilidad o proporción puede ser estimada en el modelo junto con los parámetros b de cada segmento.

Sin embargo, en la mayor parte de las aplicaciones del Logit mixto, la función *mixing distribution* $f(\beta)$ es continua, cualquier función continua que determinemos. Por ejemplo, $f(\beta)$ puede ser la función de densidad de β con media b y covarianza W. La probabilidad de elección bajo estas condiciones sería

$$P_{ni} = \int \left(\frac{e^{\beta' x_{ni}}}{\sum_j e^{\beta' x_{nj}}} \right) \phi(\beta \mid b, W) d\beta$$

donde $\phi(\beta \mid b,W)$ es la función de densidad normal con media b y covarianza W. El modelo permite estimar b y W.

Se puede utilizar cualquier otra distribución (lognormal, uniforme, etc…). *Cualquier* comportamiento *maximizador* de la utilidad puede ser representado con solo especificar adecuadamente las variables explicativas y la función de densidad. Se puede ir más allá y representar también muchas formas de comportamiento no maximizador de la utilidad. Las funciones de densidad pueden ser objeto de un test sobre su idoneidad (McFadden y Train 2000; y Chesher y Santos-Silva 2002).

Hay dos tipos de parámetros en los modelos Logit mixtos: 1), los parámetros β que están presentes en la fórmula Logit, que tienen función de densidad $f(\beta)$; y 2) los parámetros que describen esa función de densidad (como b y W, si β sigue una densidad normal con media b y covarianza W). Normalmente, aunque no siempre, el investigador está interesado en estimar estos parámetros de la función *f*.

Si los parámetros que describen o dan cuerpo a la densidad de β se representan con el símbolo θ, podremos escribir la función de densidad como $f(\beta \mid \theta)$. Las probabilidades de elección en el Logit mixto *no* dependen de los valores de β. Las probabilidades se definen como $P_{ni} = \int L_{ni}(\beta) f(\beta|\theta) d\beta$, y son funciones de θ, no de β. Los parámetros β desaparecen al integrar, como ocurre con los ε_{nj}.

En algunos Logit mixtos los β tienen un significado interpretable, dado que representan los gustos de los decisores individuales. En esos casos el investigador puede estar interesado en obtener información sobre los β para cada individuo considerado en la muestra, junto con los parámetros θ que describen la distribución de los β entre los decisores. Existen procedimientos para ello (véase Train, 2003, capítulos 11 y 12).

El siguiente problema que se nos plantea es el de derivar, dentro del Logit mixto, las probabilidades de elección a partir de un comportamiento maximizador de la utilidad. Hay diversas formas, cada una con su interpretación. La más sencilla está basada en los coeficientes aleatorios.

El decisor se enfrenta a *J* alternativas. La utilidad obtenida por el individuo *n* a partir de la alternativa j se expresa como

$$U_{nj} = \beta'_n x_{nj} + \varepsilon_{nj}$$

Donde las x_{nj} son variables observadas relativas a la alternativa *j* y al individuo *n*; β_n es un vector de coeficientes de esas variables para la persona *n*, y representa sus gustos; y ε_{nj} es un término de error *iid* de valores extremos. Los coeficientes, siendo aleatorios, siguen una distribución entre individuos de la población según una función de densidad $f(\beta)$. Esa función de densidad es una función de un conjunto de parámetros θ que representan cosas como la media y la covarianza de los βs en la población. La diferencia esencial con el Logit estándar está en que ahora los β varían entre individuos, en vez de ser fijos y comunes para todos.

El decisor conoce el valor de sus propios β_n y sus ε_{nj} para todas las alternativas *j* a las que se enfrenta. Elegirá la alternativa *i* si y sólo si $U_{ni} > U_{nj}\ \forall j \neq i$. En cambio, el investigador observa los x_{nj}, pero no los β_n o los ε_{nj}. Si observara los β_n entonces la probabilidad de elección sería la de una Logit estándar, porque los ε_{nj} son *iid* y siguen una distribución de valores extremos. Dicho de otra forma, la probabilidad condicional al valor de los β_n sería

$$L_{ni}(\beta_n) = \frac{e^{\beta'_n x_{ni}}}{\sum_j e^{\beta'_n x_{nj}}}$$

Dado que el investigador no conoce los β, no puede hacer cálculos condicionando sobre el valor de esos parámetros. La probabilidad de elección incondicionada tiene que ser la integral de $L_{ni}(\beta_n)$ entre todos los posibles valores de β_n:

$$P_{ni} = \int \left(\frac{e^{\beta' x_{ni}}}{\sum_j e^{\beta' x_{nj}}} \right) f(\beta) d\beta$$

El investigador se limita a especificar una distribución para esos coeficientes aleatorios β, y se limita a estimar los parámetros de esa distribución. En la mayor parte de los trabajos empíricos *f*(β) es una normal o log-normal (β ~ N(b,W) o ln β ~ N(b,W)). La distribución log-normal es útil cuando el coeficiente tiene el mismo signo para todos los individuos (por ejemplo, un coeficiente de precios, etc.). Pero también se han tratado otras distribuciones, como la triangular o la uniforme, cuyas funciones de densidad están limitadas en los extremos, a diferencia de la normal; o la distribución de Rayleigh, que se parece a la log-normal pero resulta más fácil de estimar (véase para las referencias y detalles a Train, 2003, página 142). La flexibilidad es muy amplia.

La variación en los gustos y preferencias entre individuos relacionadas con atributos observables del decisor se incorporan al modelo mediante la especificación oportuna de las variables explicativas o mediante la oportuna *mixing distribution*. Modelando ambas cosas se puede especificar correctamente cualquier problema. Por ejemplo, se puede dividir el coste de las alternativas por la renta, para tener en cuenta el coste relativo de las cosas (no todos los individuos tienen la misma renta). Los coeficientes son aleatorios, por lo que representarían la variación entre individuos de la misma renta del valor relativo del coste de las alternativas. Atributos observables de los individuos pueden entrar también en la función *f*(β), si se hace depender uno o varios parámetros de esa función de dichos atributos (como hace Bhat 1998, 2000).

Otra forma de interpretar o plantear un Logit mixto, sin recurrir a los coeficientes aleatorios, es como errores especiales que crean correlaciones entre utilidades de distintas alternativas:

$$U_{nj} = \alpha' x_{nj} + \mu'_n z_{nj} + \varepsilon_{nj}$$

Donde x_{nj} y z_{nj} son vectores de variables observables relativas a la alternativa j, α es un vector de coeficientes fijos, μ es un vector de términos *aleatorios* con media cero y ε_{nj} es *iid* de valores extremos. Los z_{nj} son componentes de error que junto a los ε_{nj} definen la parte estocástica de la utilidad. Por tanto, la parte inobservada de la utilidad sería $\eta_{nj} = \mu'_n z_{nj} + \varepsilon_{nj}$, pudiendo estar correlacionada entre alternativas, dependiendo de la especificación concreta de z_{nj}. Para el Logit estándar z_{nj} es cero, lo que implica que la correlación entre las utilidades de las alternativas no existe. Precisamente esta falta de correlación da lugar a propiedad IIA y a los patrones de sustitución limitados propios del Logit. En cambio, si esos componentes de error son distintos de cero, la utilidad está correlacionada entre alternativas, y tenemos que

$$Cov(\eta_{ni}, \eta_{nj}) = E(\mu_k + \varepsilon_{ni})\ (\mu_k + \varepsilon_{nj}) = \sigma_k$$

La varianza para cada una de las alternativas en el nido k será

$$Var(\eta_{ni}) = E(\mu_k + \varepsilon_{ni})^2 = \sigma_k + \pi^2/6$$

Debido a que la varianza de un término que sigue una distribución de valores extremos, como ε_{ni}, es $\pi^2/6$. Por consiguiente, la correlación entre dos alternativas cualesquiera dentro del nido k será

$$\sigma_k / (\sigma_k + \pi^2/6)$$

Si introducimos la restricción de que la varianza de cada componente de error en cada nido sea la misma para todos los nidos ($\sigma_k = \sigma$, para k=1, ...,K) es lo mismo que imponer que el coeficiente log-sum λ_k (del término I_{nk}) sea el mismo para todos los nidos en un Logit anidado. Desde luego, esta restricción garantiza que el Logit mixto queda normalizado para escala y nivel.

Si se permite a las varianzas variar para distintos nidos estaremos permitiendo que el coeficiente λ_k del Logit anidado sea diferente entre nidos. Los nidos cruzados o superpuestos se consiguen con *dummies* que identifican los conjuntos de alternativas que se superponen (Bhat, 1998). Para conseguir un Logit heterocedástico a partir de un Logit mixto sólo hay que incluir un componente de error para cada alternativa (Ben-Akiva et al. 2001). El Logit mixto engloba pues a todos los Logits que hemos visto.

Las especificaciones del componente de error y el del coeficiente aleatorio son formalmente equivalentes. Con el enfoque del coeficiente aleatorio la utilidad se especifica mediante la expresión

$$U_{nj} = \beta'_n x_{nj} + \varepsilon_{nj}$$

Siendo β_n aleatorio. Ese coeficiente β_n puede descomponerse en dos partes: su media α y desviaciones respecto a esa media μ_n, de forma que la utilidad queda como

$$U_{nj} = \alpha' x_{nj} + \mu'_n x_{nj} + \varepsilon_{nj}$$

Que tiene componentes de error definidos como $z_{nj} = x_{nj}$.

Ahora, con un enfoque de componente de error, la utilidad sería

$$U_{nj} = \alpha' x_{nj} + \mu'_n z_{nj} + \varepsilon_{nj}$$

Que es equivalente al modelo de parámetro aleatorio con coeficientes fijos para las variables x_{nj} y coeficientes aleatorios con media cero para las variables z_{nj}. Si x_{nj} y z_{nj} se superponen (en el sentido de que algunas variables comunes entran tanto en x_{nj} como en z_{nj}), los coeficientes de esas variables se pueden considerar aleatorios con media α y la misma distribución de μ_n en torno a sus medias.

Aunque los enfoques de coeficientes aleatorios y de componentes de error son equivalentes, elegir un camino u otro afecta a la especificación del Logit mixto. Por ejemplo, si se piensa en términos de parámetros aleatorios lo más lógico es permitir que los coeficientes de cada variable varíen, e incluso que haya correlaciones entre ellos (por ejemplo, Revelt y Train, 1998). Cuando se pretende representar un patrón de sustitución mediante el uso de componentes de error, lo mejor es especificar variables que induzcan correlaciones entre alternativas (Brownstone y Train, 1999).

El Logit mixto no tiene la propiedad de *independencia de las alternativas irrelevantes* (IIA), ni los patrones de sustitución restrictivos del Logit. La razón de probabilidades en el Logit mixto P_{ni}/P_{nj} depende de todos los datos, incluidos los atributos de otras alternativas distintas de i y de j. Los denominadores de la fórmula Logit están dentro de una integral, por lo que no se van. El porcentaje de cambio en la probabilidad de una alternativa debido al cambio en el m-ésimo atributo de otra alternativa es

$$\begin{aligned} E_{nix_{nj}^m} &= -\frac{1}{P_{ni}} \int \beta^m L_{ni}(\beta) L_{nj}(\beta) f(\beta) d\beta \\ &= -\int \beta^m L_{ni}(\beta) \left[\frac{L_{ni}(\beta)}{P_{ni}} \right] f(\beta) d\beta \end{aligned}$$

donde β^m es el m-ésimo elemento de β. Esta elasticidad es distinta para cada alternativa i. En el modelo Logit un diez por ciento de reducción en una alternativa implicaba un diez por ciento de reducción en cada una de las demás. En el Logit mixto el patrón de sustitución depende de la especificación de las variables y de la *mixing distribution*, que se pueden determinar empíricamente.

El porcentaje de cambio en la probabilidad depende de la correlación entre $L_{ni}(\beta)$ y $L_{nj}(\beta)$ a lo largo de diferentes valores de β, que se determinan por la especificación de las variables y de la *mixing distribution* que fije el investigador. Por ejemplo, se puede diseñar una situación en la que una mejora en la alternativa j drene proporcionalmente más de la alternativa i que de la k, y para ello se puede especificar un elemento de x que está positivamente correlacionado con i y j, pero incorrelacionado o negativamente correlacionado entre k y j, con una *mixing distribution* que permita al coeficiente de esa variable variar.

Cualquier *modelo de utilidad aleatorio* (RUM) puede aproximarse, con cualquier grado de precisión, por Logit mixto, mediante la apropiada elección de variables y de *mixing distribution*, como muestran McFadden y Train (2000) o Dagsvik (1994).

Se podría dar una explicación intuitiva. Si el modelo real responde a la expresión $U_{nj} = \alpha'_n z_{nj}$, donde z_{nj} son variables relativas a la alternativa j, y α sigue cualquier distribución $f(\alpha)$. Cualquier RUM puede ser expresado de esta forma. Por ejemplo, la expresión tradicional $U_{nj} = \beta'_n x_{nj} + \varepsilon_{nj}$ se obtiene haciendo $z'_{nj} = (x'_{nj}, d_j)$, $\alpha' = (\beta'_n, \varepsilon_{nj})$ y $f(\alpha)$ ser la función de densidad conjunta de β_n y ε_{nj} $\forall j$. Una vez conocido α, la elección de la persona está totalmente determinada, ya que U_{nj} queda determinada para cada alternativa j. La probabilidad condicional es en ese caso:

$$q_{ni}(\alpha) = I(\alpha'_n z_{ni} > \alpha'_n z_{nj} \ \forall j \neq i)$$

donde I(.) es un indicador 0-1 que señala si un evento, descrito dentro del paréntesis) ocurre o no. Esa probabilidad condicional es determinista en el sentido de que es cero o uno: condicional a todos los términos aleatorios desconocidos, la elección del decisor queda completamente determinada. La probabilidad de elegir incondicional es la integral de $q_{ni}(\alpha)$ sobre los valores α:

$$Q_{ni} = \int I(\alpha'_n z_{ni} > \alpha'_n z_{nj} \ \forall j \neq i) \, f(\alpha) \, d\alpha$$

Podemos aproximar esa probabilidad mediante un Logit mixto. Escalamos la utilidad por un factor λ, de forma que $U^*_{nj} = (\alpha/\lambda) z_{nj}$. Este escalado no afecta al modelo, ya que el comportamiento no resulta afectado por la escala de la utilidad.

Ahora añadimos un término *iid* de valores extremos ε_{nj}. Al añadir ese término aleatorio de valores extremos el modelo sí cambia, ya que la utilidad de cada alternativa se ve alterada. Es necesario añadir ese término porque sólo así conseguimos tener un Logit mixto. De todas formas, añadir esta variable aleatoria de valores extremos no tiene consecuencias negativas. La probabilidad que nos da el Logit mixto para esa utilidad que hemos especificado es

$$P_{ni} = \int \left(\frac{e^{(\alpha/\lambda)' z_{ni}}}{\sum_j e^{(\alpha/\lambda)' z_{nj}}} \right) f(\alpha) d\alpha$$

Conforme λ se aproxima a cero los coeficientes α/λ en la fórmula Logit crecen y P_{ni} se aproxima a Q_{ni}, que es la probabilidad real. Escalando los coeficientes lo suficiente, el modelo Logit mixto se acerca tanto como se desee al modelo real.

Hemos añadido un término *iid* de valores extremos a la utilidad de cada alternativa, lo que modifica el modelo. Ahora la alternativa con la máxima utilidad antes de añadir el término no tiene por qué serlo después de añadirlo. Cada utilidad recibe una cantidad distinta. No obstante, elevando la escala de utilidad lo suficiente podemos asegurarnos de que la adición de ese término aleatorio de valores extremos no tiene efectos.

Imaginemos un caso con dos alternativas. Si en el modelo original, con su escala original, la alternativa 1 tiene una utilidad 0,5 unidades superior a la alternativa 2, la 1 será elegida. Ahora añadimos un término de error de valores extremos. Debido a su varianza, podría ocurrir perfectamente que el valor de la utilidad de la alternativa 2 supere al de la alternativa 1 (por al menos 0,5 unidades). Por tanto, el modelo original ha sido alterado de forma significativa. Ahora escalamos el modelo original por un factor de 10 unidades (λ=10). La alternativa 1 supera a la 2 en 5 unidades en este modelo original reescalado. Al añadir ahora una variable aleatoria de valores extremos es sumamente improbable que una diferencia de esa magnitud se contrarreste. Estamos añadiendo ε_{n1} a la alternativa 1, y ε_{n2} a la 2, y ε_{n2} tendría que superar en más de 5 unidades a ε_{n1} para que se produjera la reversión, lo que es sumamente improbable, por no decir imposible. En cualquier caso, puede hacerse el factor de escala tan grande como queramos (λ=100; λ=1000). Es precisamente el añadido de esa variable aleatoria de valores extremos lo que convierte al modelo original en un Logit mixto, y esto no lo altera de forma significativa si la utilidad se escala al alza lo suficientemente.

El investigador no *tiene que* escalar la utilidad para poder trabajar con un Logit mixto. Simplemente, es un método efectivo para conseguir la aproximación deseada entre el modelo original (*RUM*) y el Logit mixto que se pretende estimar. A veces no es necesario. El modelo original puede, por ejemplo, incluir una variable *iid* distinta para cada alternativa. En un caso así, si asumir una distribución de valores extremos no es problemático, la escala de la utilidad queda determinada por la varianza del término *iid*. El investigador sólo tiene que preocuparse, en un caso así, por encontrar un conjunto de variables y una *mixing distribution* capaces de captar bien el resto de la utilidad, normalmente aquellas partes que presentan correlaciones entre alternativas o heterocedasticidad.

El Logit mixto, como el Probit, tiene que estimarse por medio de métodos de simulación. La utilidad es $U_{nj} = \beta'_n x_{nj} + \varepsilon_{nj}$, y los coeficientes β_n siguen una distribución $f(\beta \mid \theta)$, donde θ se refiere al conjunto de parámetros de esa distribución (como la media y la covarianza de β). El investigador especifica la forma funcional de $f(.)$ y desea estimar los parámetros θ. La probabilidad de elegir la alternativa i es

$$P_{ni} = \int L_{ni}(\beta) f(\beta \mid \theta)\, d\beta$$

Siendo

$$L_{ni}(\beta) = \frac{e^{\beta' x_{ni}}}{\sum_{j=1}^{J} e^{\beta' x_{nj}}}$$

Las probabilidades se aproximan por simulación para cualquier valor de θ: 1) se toma una muestra aleatoria de β a partir de la distribución $f(\beta \mid \theta)$ y se marca como β^r siendo r = 1 la primera muestra; 2) calcular la fórmula Logit $L_{ni}(\beta^r)$ con esta muestra; 3) repetir los pasos 1) y 2) muchas veces, y promediar los resultados. Esa media es la probabilidad (simulada)

$$\breve{P}_{ni} = \frac{1}{R} \sum_{r=1}^{R} L_{ni}(\beta^r)$$

siendo R el número de muestras. $\breve{P}_{ni}$ es un estimador insesgado de P_{ni}, por construcción; su varianza se reduce conforme R crece; y es estrictamente positivo, por lo que ln $\breve{P}_{ni}$ está definido, siendo esto muy útil para aproximar el logaritmo de la función de verosimilitud (*log-likelihood function*). Además, $\breve{P}_{ni}$ es dos veces diferenciable en los parámetros θ y en las variables *x*, lo que facilita la búsqueda numérica de la función de máxima verosimilitud, y el cálculo de elasticidades. Por último, $\breve{P}_{ni}$ suma uno entre todas las alternativas, que es muy importante siempre que se hacen predicciones sobre la base del modelo.

Las probabilidades obtenidas por simulación se insertan en el logaritmo de la función de verosimilitud (*log-likelihood function*) para tener un logaritmo de verosimilitud simulado:

$$SSL = \sum_{n=1}^{N}\sum_{j=1}^{J} d_{nj} \ln \breve{P}_{nj}$$

donde d_{nj} = 1 si el individuo *n* elije *j*, e igual a cero en otro caso. El estimador de máxima verosimilitud simulado (*maximum simulated likelihood estimator*, MSLE) es el valor de θ que maximiza SLL (véase el capítulo 10 de Train, 2003, para sus propiedades).

El procedimiento es simple. Se toman distintas muestras aleatorias para cada observación, lo que preserva la independencia entre decisores de las probabilidades simuladas que entran en SLL. La probabilidad simulada del Logit mixto puede relacionarse con los métodos de simulación de aceptación- rechazo (*accept-reject*, AR) que se emplean en modelos Probit. Para cualquier modelo de utilidad aleatorio (RUM), el simulador AR se construye de la forma siguiente: 1) se toma una muestra aleatoria de los términos aleatorios; 2) se calcula la utilidad de cada alternativa a partir de esa muestra, y se identifica qué alternativa tiene la máxima utilidad de todas; 3) se repiten los pasos 1) y 2) muchas veces; 4) se calcula la probabilidad simulada para una alternativa como la proporción de muestras para las que la alternativa tiene la máxima utilidad. El simulador AR es insesgado por construcción, pero no es estrictamente positivo para cualquier número finito de muestras. Tampoco es suave, sino que presenta un escalón: puede ser constante para un rango de parámetros para los que la alternativa con la máxima utilidad no cambia para ninguna muestra, y con saltos cuando un cambio en los parámetros cambian la alternativa con mayor utilidad. Los métodos numéricos de maximización no se pueden emplear con el simulador AR debido a esas características. Para tratar este problema, el simulador AR se suaviza reemplazando la función-indicador 0-1 *con la fórmula del Logit*. Escalando la probabilidad adecuadamente, el simulador suavizado puede acercarse al simulador AR tanto como queramos (véase Train, 2003, sección 5.6.2).

Este simulador del Logit mixto puede verse como un simulador AR suavizado *con la fórmula del Logit* para cualquier RUM: se toma una muestra, se calculan las utilidades, se insertan esas utilidades en la fórmula Logit y se promedian los resultados. Un Logit mixto puede aproximar cualquier RUM, y este teorema puede entenderse desde esta perspectiva. El simulador AR basado en Logit suavizado puede aproximarse tanto como queramos al simulador AR para cualquier modelo, con sólo escalar suficientemente la utilidad. Dado que el simulador para el Logit mixto equivale a un simulador AR para un Logit suavizado, el Logit mixto simulado puede aproximarse tanto como queramos al simulador AR de cualquier modelo.

Si añadimos la posibilidad de elecciones repetidas para cada decisor incluido en la muestra a partir de la cual estimamos, estaremos ante el caso más general de los datos de panel. Hay distintas especificaciones posibles.

La especificación más simple permite que los coeficientes que entran en la fórmula de la utilidad varíen entre personas pero sean constantes entre períodos para cada persona. La utilidad derivada de la alternativa *j* en el período *t* para la persona *n* es

$$U_{njt} = \beta_n x_{njt} + \varepsilon_{njt}$$

Siendo ε_{njt} *iid* de valores extremos a lo largo del tiempo, personas y alternativas. Consideremos ahora una secuencia de alternativas, una para cada período de tiempo, $i = (i_1, \ldots, i_T)$. Condicionado al valor de β, la probabilidad de que la persona opte por esa secuencia de alternativas es el producto de fórmulas Logit:

$$L_{ni}(\beta) = \prod_{t=1}^{T} \left[\frac{e^{\beta_n' x_{ni_t t}}}{\sum_j e^{\beta_n' x_{ni_t t}}} \right]$$

dado que los ε_{njt} son independientes a lo largo del tiempo. La probabilidad incondicional es la integral de ese producto para todos los valores de β:

$$P_{ni} = \int L_{ni}(\beta) f(\beta) d\beta$$

La única diferencia entre un Logit mixto con elecciones repetidas y uno con sólo un proceso de elección por individuo es que el integrando consiste en un producto de fórmulas Logit, una para cada período de tiempo, en vez de sólo una fórmula Logit. La probabilidad se calcula mediante simulación de forma similar. Se toma una muestra de β a partir de su distribución. La fórmula Logit se calcula para cada período, y acto seguido se hace lo propio con el producto. Este proceso se repite para muchas muestras, y los resultados se promedian.

Se pueden incluir variables exógenas a la fórmula de la utilidad en un momento dado, para representar así respuestas retardadas o anticipadas a cambios en las variables, de la misma forma que en el caso del Probit. Sin embargo, en el caso del Logit mixto, esto no implica cambiar el procedimiento de estimación. Condicionando por β_n, los únicos términos aleatorios que quedan en el Logit mixto son los ε_{nj}, que son independientes en el tiempo. Una variable dependiente retardada que entre en los U_{njt} estará incorrelada con esos términos de error restantes para el período t, ya que esos términos son independientes en el tiempo.

Las probabilidades condicionadas (condicionadas a β) son por tanto las mismas que acabamos de ver en la expresión de $L_{ni}(\beta)$, pero con las *x* incluyendo variables dependientes retardadas. La probabilidad incondicional es la integral de esa probabilidad condicional para todos los valores de β, que es exactamente la expresión anterior para P_{ni}. Desde este punto de vista, ***el Logit mixto es más conveniente que el Probit para la representación de los hábitos*** (*state dependence*), ya que las variables dependientes retardadas se pueden añadir al Logit mixto sin ajustar la fórmula de la probabilidad ni el método de simulación. Erdem (1996) y Johannesson y Lundin (2000) estudian la formación de *hábitos* y la búsqueda de variedades mediante un Logit mixto que también trata la variación aleatoria en los gustos.

Si los datos y las elecciones no se observan desde el primer momento del proceso (desde el primer problema de elección que se presenta al individuo) el tema de las *condiciones iniciales* debe ser tratado debidamente, exactamente como ocurría con el Probit. El investigador debe arreglárselas para representar la probabilidad de la primera elección observada, que depende de las previas, no observadas. Heckman y Singer (1996) proporcionan algunas vías para tratar este problema. Si el investigador observa la secuencia de elecciones desde el principio el problema no aparece, en cuyo caso el uso el Logit mixto facilita mucho la inclusión de variables dependientes retardadas para captar inercias u otros tipos de *state dependence*, como los hábitos. Contra lo que pudiera parecer, no es tan raro poder observar la secuencia de acontecimientos desde el principio. Las *stated-preferences* (respuestas a una serie de situaciones en las que se plantean elecciones hipotéticas) son un buen ejemplo.

En la especificación y en casi todas las aplicaciones se asume que los coeficientes βn son constantes en el tiempo (distintas elecciones) para un mismo individuo. Este supuesto es apropiado si los gustos del individuo son estables en el tiempo, al menos durante el período en que se produce la secuencia de elecciones. Sin embargo, ese supuesto se puede alterar, permitiendo que los coeficientes varíen para cada persona en el tiempo. Se puede hacer de diversas formas. Por ejemplo, los gustos de cada persona podrían estar serialmente correlados en el tiempo (entre rondas de elección), de forma que la utilidad se especifique como

$$U_{njt} = \beta_{nt} x_{njt} + \varepsilon_{njt}$$

$$\beta_{nt} = b + \tilde{\beta}_{nt}$$

$$\tilde{\beta}_{nt} = \rho \tilde{\beta}_{n(t-1)} + \mu_{nt}$$

Donde *b* es fijo y μ_{nt} es *iid* entre los distintos *n* y *t*. La simulación de la probabilidad para la secuencia de elecciones tiene los siguientes pasos: 1) se toma una muestra μ^r_{n1} para el período inicial, y se calcula la fórmula Logit para ese período haciendo $\beta^r_{n1} = b + \mu^r_{n0}$; 2) se toma otra muestra μ^r_{n2} para el segundo período, y se calcula $\beta^r_{n2} = b + \rho\mu^r_{n1} + \mu^r_{n2}$, para después calcular la fórmula Logit basada en ese β^r_{n2}; 3) se continúa para todos los períodos (que son T); 4) se calcula el producto de los T

Logits; 5) se repiten los pasos 1-4 para diversas secuencias de muestras; 6) se promedian los resultados. El peso de la simulación es mayor aquí que con coeficientes constantes en el tiempo para cada persona, ya que se requieren T veces más muestras.

Para un caso práctico de aplicación de Logit mixto puede verse Train (1999), que analiza la elección de zonas de pesca, también empleado, pero con coeficientes correlacionados, en Train (1998) y aplicando un modelo desarrollado por Desvousges et al. (1996). Este caso puede verse explicado en Train (2003), página 151 a 154.