OTROS MODELOS

Los datos de *preferencia revelada* provienen de las elecciones de los individuos observadas en el mundo real. Al elegir, los individuos *revelan* sus preferencias y gustos. Los datos de *preferencias declaradas* (por traducir de alguna forma *stated-preferences*) provienen de encuestas en las que el individuo se enfrenta *hipotéticamente* a un conjunto de alternativas de entre las que tienen que elegir. Los entrevistados afirman cuáles serían sus elecciones en ese caso hipotético. En el primer tipo de preferencias, se observa, por ejemplo, qué coche ha comprado un individuo; en el segundo se pregunta a un individuo cuál compraría si tuviera que elegir, de entre la muestra de alternativas que se le presentan.

Hay ventajas e inconvenientes de uno y otro tipo de datos. La preferencia *revelada* proporciona datos que proceden de elecciones reales, pero las alternativas están limitadas a las que existen o existieron. Pero este tipo de datos no están disponibles, lógicamente, cuando queremos investigar la aceptación de un nuevo producto que aún no está en el mercado. Otro problema es que determinados factores muy importantes para la elección pueden no presentar suficiente variación. Esto suele darse cuando hay precios de mercado como factores relevantes en la elección de un producto o servicio. El precio es muy importante, pero precisamente por eso los oferentes competidores reducen las diferencias entre precios al mínimo en el equilibrio. Precisamente, cuanto más sensible es el demandante al precio, menores serán las diferencias de precios en el equilibrio. Ante un caso así es imposible calcular la respuesta de la demanda ante diferencias en los precios, porque éstas no se observan en los mercados en equilibrio. Si se usaran estos datos, el coeficiente del precio sería poco o nada relevante en la explicación de la demanda, y podríamos concluir (erróneamente) que el precio no es importante. En general, las fuerzas que equilibran los mercados reducen la variación en los datos de aquellas variables más relevantes en los problemas de elección, lo que afecta a los datos de preferencia *revelada*. La única forma de superar esto es acudiendo a datos procedentes de preferencias *declaradas*.

Las preferencias *declaradas* complementan la información que procede de las preferencias *reveladas*. Un cuestionario presenta al individuo distintas alternativas declaradas, descritas y acompañadas de una serie de atributos que presentan la debida variación. El individuo tiene que elegir, como si la situación se le presentara en el mundo real. La ventaja es que puede introducirse la variación o variabilidad que el investigador considere pertinente en los atributos, lo que permite medir mejor la importancia de estos en la determinación de la opción elegida. Si la alternativa no existe en el mundo real (por ser un producto aún no introducido en el mercado, por ejemplo), este método permite considerarla en el proceso de elección igualmente. Por otro lado, lo que un individuo dice que haría ante una situación hipotética no tiene por qué corresponder con lo que de verdad hará en una situación distinta, real.

La combinación de ambos tipos de datos (preferencias *declaradas* y preferencias *reveladas*) permiten superar las limitaciones respectivas. Sin embargo, para ello deben cumplirse ciertos requisitos: 1) en la estimación de las razones de coeficientes, que miden la importancia relativa de los distintos atributos, deben emplearse preferentemente datos de preferencias *declaradas* (dado que se requiere suficiente variación en dichos atributos); 2) las constantes específicas de las alternativas y la escala general de los parámetros deben estimarse a partir de datos de preferencia *revelada*, ya que las constantes y la escala determinan las contribuciones medias de cada atributo.

Con los años se han ido desarrollando procedimientos para estimar modelos de elección discreta sobre una base de datos de preferencias declaradas (*stated*) y reveladas (*revealed*), y a partir de distintos modelos: con modelos Logit, Ben-Akiva and Morikawa (1990), Hensher y Bradley (1993) y Hensher et al. (1999); y con modelos Logit mixtos, Bhat y Castelar (2002) y Brownstone et al. (2000). Estos procedimientos implican algunos cambios en esos modelos, que ya hemos visto. El principal cambio al que obliga la combinación de los dos tipos de datos es que los factores inobservados son distintos para los dos tipos de datos, lo que plantea la necesidad de un tratamiento específico, adaptado al caso. Veamos cómo se hace.

Como siempre, la utilidad que un individuo obtiene de optar por la alternativa *j* en la situación o el momento t sería

$$U_{njt} = \beta' x_{njt} + e_{njt}$$

Donde x_{njt} *no* incluye constantes específicas para cada alternativa y e_{njt} representa el efecto de factores que no son observados por el investigador. Estos factores tienen una media para cada alternativa (que representa el efecto medio de esos factores en la utilidad de cada *j*) y una distribución en torno a esa media. Esta media queda representada por una constante específica para cada alternativa, c_j, y para un Logit estándar la distribución en torno a esa media es de valores extremos (tipo I) con varianza $\lambda^2\pi^2/6$. Se fija la escala de la utilidad normalizando la varianza de la porción inobservada de la utilidad. Mediante ese procedimiento, la utilidad quedaría expresada de la siguiente forma:

$$U_{njt} = (\beta/\lambda)' x_{njt} + c_j/\lambda + \varepsilon_{njt}$$

Donde el error ha quedado normalizado, de forma que $\varepsilon_{njt} = (e_{njt} - c_j)/\lambda$ siendo ahora además una variable con distribución *iid* de valores extremos con varianza $\pi^2/6$. La probabilidad de elección de la alternativa j viene dada por la fórmula Logit para $(\beta/\lambda)' x_{njt} + c_j/\lambda$, y los parámetros estimados son los originales, pero divididos por el factor de escala λ.

Cuando combinamos preferencias declaradas y preferencias reveladas se presenta el siguiente problema: las constantes específicas de cada alternativa y el factor de escala no tienen por qué ser iguales para ambos tipos de datos. Esos parámetros reflejan la influencia de los factores inobservados en la utilidad, y éstos no serán los mismos en uno y otro caso, ni en media ni en varianza. En el caso del experimento que conduce a las preferencias declaradas, los factores inobservados están más controlados, o al menos igualados entre alternativas, lo que reduce la importancia de estos. Aunque se pida al encuestado que no tenga en cuenta más factores que los explicitados en el experimento.

Esas diferencias deben ser tenidas en cuenta, y para ello hay que separar las constantes y factores de escala para ambos tipos de dato. Nombramos, por ejemplo, c^s_j y c^r_j a las constantes o efectos medios de los factores inobservados para la alternativa j en los casos de preferencia revelada y preferencia declarada. Hacemos lo mismo para los factores de escala con λ^s y λ^r, que son proporcionales a las desviaciones típicas de las distribuciones en torno a las medias representadas por las constantes. La normalización requiere ahora igualar uno dos los dos parámetros de escala a uno, convirtiendo así al otro parámetro en una razón de parámetros. Si hacemos $\lambda^r = 1$ tendremos una expresión para la utilidad en la situación t bajo preferencias declaradas, y otra para las preferencias reveladas:

$$U_{njt} = (\beta/\lambda^s)' x_{njt} + c^s_j/\lambda^s + \varepsilon_{njt}$$

$$U_{njt} = \beta' x_{njt} + c^r_j + \varepsilon_{njt}$$

Muchos programas informáticos no pueden estimar el factor de escala que afecta separadamente a los coeficientes en el modelo de preferencias declaradas. A veces hay que introducir ese parámetro extra programando uno sus propias rutinas (Hensher y Bradley, 1993, muestran cómo hacerlo para un Logit anidado).

Al emplearse simultáneamente ambas fuentes de información, las estimaciones del vector de parámetros β dependerá de la variación en los atributos (el vector *x*), de forma que si toda la variación se encuentra en los datos de procedencia "declarada", serán estos datos los que determinen ese valor estimado. En tanto que los datos de preferencia revelada contengan variación, ésta aportará algo a la estimación del parámetro.

Por otro lado, las constantes específicas de cada alternativa son distintas para cada fuente de datos, y por tanto se estiman separadamente, en el sentido de que cada tipo de datos determina sus constantes. Esta separación es interesante por cuanto podemos aislar la fuente de problemas que acompaña a cada una de las fuentes en algunos casos. Por ejemplo, las preferencias declaradas pueden mostrar una diferencia respecto a las elecciones que se hacen en el mundo real. La constante muestra la probabilidad media de comprar un producto (una alternativa). Si hay un sesgo en las preferencias declaradas podemos usar la constante calculada con datos de preferencias reveladas, con lo que nuestras predicciones ganarían en realismo. De la misma forma, podemos usar el factor de escala de las preferencias reveladas en vez del procedente de preferencias declaradas, para poder hacer predicciones más realistas.

***

A veces nos encontramos con alternativas clasificadas u ordenadas en un ranking (*ranked data*). En experimentos de preferencias declaradas es posible pedir al entrevistado no que identifique la alternativa que prefiere, olvidándose de las demás, sino que las ordene todas por orden de prelación. Puede pedirse que elija una, y después que elija otra de las restantes, y así sucesivamente; o bien se le puede pedir que las ordene todas de una vez. Se supone que al ordenar las alternativas el entrevistado está ordenando sus preferencias.

Estos datos ordenados (*ranked data*) pueden tratarse mediante un Logit estándar o un Logit mixto, sin especiales modificaciones. Sí es necesario preparar los datos de una forma determinada, que veremos ahora. Un Probit, en cambio, necesita alguna adaptación para trabajar con datos ordenados, aunque dicho ajuste no es sustancial.

En el caso del Logit estándar la probabilidad de cualquier ordenación de alternativas es el producto de una serie de fórmulas Logit. Por ejemplo, si se le presentan a un individuo 4 alternativas (A, B, C, D), y éste las ordena de forma que la preferida es C, y después B, D y A, y si suponemos que la utilidad de cada alternativa tiene un factor aleatorio que sigue una distribución de valores extremos, entonces la probabilidad de esa ordenación concreta puede expresarse como la probabilidad (Logit) de elegir C de entre las 4 opciones, por la probabilidad (Logit) de elegir B de las 3 opciones restantes, por la probabilidad de elegir D de entre las 2 restantes. Expresado matemáticamente:

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

Siendo j = A, B, C, D y con ε_{nj} *iid* de valores extremos. Entonces

$$P_{C,B,D,A} = \frac{e^{\beta' x_{nC}}}{\sum_{j=A,B,C,D} e^{\beta' x_{nj}}} \frac{e^{\beta' x_{nB}}}{\sum_{j=A,B,D} e^{\beta' x_{nj}}} \frac{e^{\beta' x_{nD}}}{\sum_{j=A,D} e^{\beta' x_{nj}}}$$

La distribución de valores extremos hace ese resultado posible, por lo que la fórmula no se aplica a modelos Probit, basados en la distribución normal.

Como puede observarse, la fórmula implica que la ordenación de las 4 alternativas puede descomponerse en 3 procesos de elección independientes. Estas 3 elecciones se conocen como *pseudo-observaciones*, porque cada ranking completo, que es una sola observación para cada individuo, se descompone en varios *rankings* que parecen distintas observaciones (casos separados). En general, un ranking de J alternativas proporciona J-1 pseudo-observaciones en un modelo Logit estándar. En la primera pseudo-observación todas las alternativas están disponibles y la variable dependiente identifica la primera opción o alternativa. En la segunda pseudo-observación la primera alternativa elegida queda descartada, las restantes pasan a constituir el conjunto de opciones (*choice set*) y la variable dependiente identifica la segunda opción preferida. Cuando se preparan los datos para la estimación Logit, las variables explicativas se repiten J-1 veces.

Al modelo Logit con alternativas clasificadas (*ranked alternatives*) se le suele llamar un *Logit expandido* (*exploded Logit*), pues cada observación se multiplica en varias pseudo-observaciones para la estimación. Para algunas aplicaciones puede verse Beggs et al. (1981), Chapman y Staelin (1982) y Hausman y Ruud (1987).

Un modelo Logit mixto puede estimarse con datos ordenados con el mismo mecanismo de expansión. Si asumimos ahora que β es aleatorio con una función de densidad g(β | θ), donde θ son parámetros de esa distribución, condicionando por β, la probabilidad de una determinada clasificación seleccionada por la persona es un producto de Logits, como el de la ecuación anterior. La probabilidad incondicional es la integral de ese producto sobre la densidad de β, de forma que:

$$P_{C,B,D,A} = \int \left(\frac{e^{\beta' x_{nC}}}{\sum_{j=A,B,C,D} e^{\beta' x_{nj}}} \frac{e^{\beta' x_{nB}}}{\sum_{j=A,B,D} e^{\beta' x_{nj}}} \frac{e^{\beta' x_{nD}}}{\sum_{j=A,D} e^{\beta' x_{nj}}} \right) g(\beta|\theta) d\beta$$

Este Logit mixto con alternativas ordenadas se estima por los procedimientos usuales aplicados a los datos de panel, siendo el panel de datos una repetición J-1 veces del conjunto de observaciones original de las variables explicativas. En el Logit mixto cada individuo tiene sus propios coeficientes y, lo que es más importante, sus coeficientes afectan a toda su clasificación u ordenación de alternativas. Quiere esto decir que las pseudo-observaciones están correlacionadas.

Un modelo Probit también admite estos datos ordenados. Imaginemos que las utilidades de las alternativas son idénticas a las que acabamos de ver, sólo que los términos de error siguen una distribución conjuntamente normal, de forma que

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

Para j=A,B,C,D donde ε_n = (ε_{nA},…, ε_{nD})' sigue una distribución N(0,Ω). Como ya sabemos, la probabilidad de una determinada ordenación es

$$P_{C,B,D,A} = \text{Prob}(U_{nC} > U_{nB} > U_{nD} > U_{nA})$$

Descomponer esa probabilidad conjunta en condicionales y una marginal no funciona con un Probit como con un Logit, dado que las probabilidades condicionales no se transforman en probabilidades incondicionales como ocurre cuando los errores son independientes. Hay que buscar otra vía. Con los modelos Probit resultaba mucho más cómodo trabajar con diferencias entre las utilidades que con las utilidades en sí mismas. Por tanto, hacemos

$$\begin{aligned}
\tilde{U}_{njk} &= U_{nj} - U_{nk} \\
\tilde{x}_{njk} &= x_{nj} - x_{nk} \\
\tilde{\varepsilon}_{njk} &= \varepsilon_{nj} - \varepsilon_{nk}
\end{aligned}$$

Ahora podemos expresar la probabilidad de una ordenación particular como

$$\begin{aligned}
P_{C,B,D,A} &= \Pr ob(U_{nC} > U_{nB} > U_{nD} > U_{nA}) \\
&= \Pr ob(\tilde{U}_{nBC} < 0,\ \tilde{U}_{nDB} < 0,\ \tilde{U}_{nDA} < 0)
\end{aligned}$$

Para expresar esta probabilidad definimos una matriz de transformación M que calcula las diferencias, como la matriz que transformaba, en el Probit, la matriz de covarianzas de los errores en la matriz de covarianzas de las diferencias entre errores (recordamos: $\tilde{\Omega} = M\Omega M'$). El procedimiento con datos ordenados es el mismo, pero la matriz de transformación es distinta ahora.

Amontonamos las alternativas A a D, de forma que la utilidad se exprese en forma de vector como $U_n = V_n + \varepsilon_n$, donde $\varepsilon_n \sim N(0,\Omega)$. Definimos una matriz 3x4

$$M = \begin{pmatrix} 0 & 1 & -1 & 0 \\ 0 & -1 & 0 & 1 \\ 1 & 0 & 0 & -1 \end{pmatrix}$$

La matriz tiene una fila por cada desigualdad en el argumento de la probabilidad- Cada fila contiene un 1 y un -1, junto con dos ceros, donde el 1 y el -1 identifica las alternativas están siendo restadas en la desigualdad. Con esa matriz la probabilidad de las alternativas ordenadas sería

$$
\begin{aligned}
P_{C,B,D,A} &= \Pr ob(U_{nC} > U_{nB} > U_{nD} > U_{nA}) \\
&= \Pr ob(\tilde{U}_{nBC} < 0, \tilde{U}_{nDB} < 0, \tilde{U}_{nDA} < 0) \\
&= \Pr ob(MU_n < 0) \\
&= \Pr ob(MV_n + M\varepsilon_n < 0) \\
&= \Pr ob(M\varepsilon_n < -MV_n)
\end{aligned}
$$

Las diferencias entre errores definidas por $M\varepsilon_n$ están distribuidas según una distribución conjunta normal, con media cero y covarianzas $M\Omega M'$. La probabilidad de que estas diferencias entre errores correlacionadas sean menores que $-MV_n$ se calcula mediante simulación con un simulador GHK. Pueden verse aplicaciones en Hajivassiliou y Ruud (1994) y Schechter (2001).

***

Un caso distinto al anterior es cuando se pide a los encuestados una ordenación cualitativa de entre varias opciones (*ordered responses*).

Por ejemplo, cuando se pregunta sobre la actuación de un determinado político, y se ofrecen las opciones *muy buena*, *buena*, *ni buena ni mala*, *mala* y *muy mala*; o bien, por un determinado producto que se acaba de probar, existiendo una calificación de 1 a 7 (de peor a mejor); o sobre la probabilidad de comprar un coche en los próximos meses, con la posibilidad de que no sea *nada probable*, *algo probable* o *muy probable*.

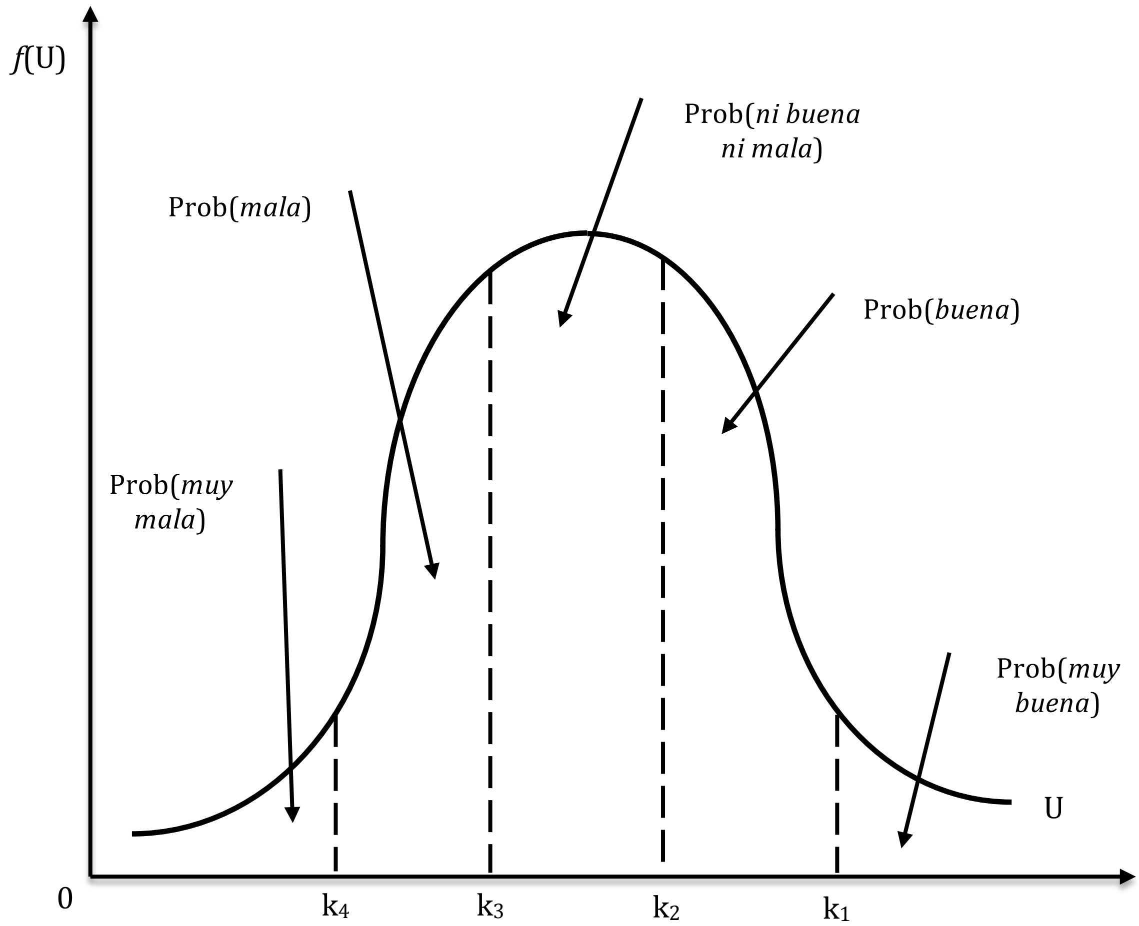
La clave aquí está en que las posibles respuestas ya implican un determinado orden. Podríamos pensar que el encuestado, al elegir, opta entre alternativas, como en cualquiera de los casos de elección que hemos venido viendo. Sin embargo, en el modelo Logit estándar se supone que los errores de cada alternativa son independientes, y eso contradice el hecho de que las alternativas tengan ya un orden. En efecto, esto implica que entre las alternativas hay más similaridad entre las “cercanas” y menos entre las opciones más “lejanas”. Esto puede abordarse mediante un Logit anidado, un Logit mixto o un Probit que tengan en cuenta el patrón de similaridades entre alternativas que acabamos de mencionar. Por ejemplo, podemos emplear un Probit con correlación entre las alternativas, siendo mayor entre alternativas similares y menor entre alternativas más alejadas. Los resultados de ese modelo podrían ser buenos, pero en realidad no se ajustaría bien a la naturaleza de los datos. Todos estos modelos plantean una función de utilidad para cada alternativa. Después se comparan los niveles de utilidad proporcionados por cada una de ellas. Pero en estos casos de opciones ordenadas no parece que cada una de las alternativas deba tener su propia función de utilidad detrás, ya que son valoraciones subjetivas de un mismo fenómeno al que se enfrenta el encuestado.

Es más razonable pensar que el encuestado tiene una función de utilidad (y sólo una) asociada al caso, y elige una opción en función del nivel de utilidad. Imaginemos un caso en el que se pregunta a una persona por la actuación de un político. Una variable *U* mide el nivel de utilidad, o de satisfacción, de la actuación del político, y en función del nivel que alcanza esa variable *U* su valoración será más o menos positiva. Al encuestado se le ofrecen 5 posibles respuestas (*muy buena*, *buena*, *ni buena ni mala*, *mala* o *muy mala*). Cada una de estas respuestas implica un rango en el nivel de satisfacción U. Las fronteras entre esos rangos son unos niveles mínimos de satisfacción k_i, de manera que si *U* queda igual o por encima de k_1 la respuesta será muy buena; si *U* está por debajo de k_1 pero por encima de k_2, responderá buena; etc.

- *Muy buena* si $U \geq k_1$
- *Buena* si $k_1 > U \geq k_2$
- *Ni buena ni mala* si $k_2 > U \geq k_3$
- *Mala* si $k_3 > U > k_4$
- *Muy mala* si $k_4 > U$

El investigador observa distintas variables explicativas relacionadas con la persona que responde a la pregunta, como la afiliación política, la renta, etc. Aquí también hay factores inobservados, que son aleatorios. La distribución de estos factores inobservados determina la probabilidad de cada una de las cinco respuestas.

La variable *U* se distribuye en torno a $\beta'x$ con una distribución determinada por la de ε. Hay puntos de separación determinados por $k_1, \ldots, k_4$. La probabilidad de que la persona conteste muy mala es el área bajo la curva de la distribución y a la izquierda de k_4 (véase el gráfico). Y lo mismo puede decirse de cada una de las restantes porciones.



Una vez se especifica una distribución para ε las probabilidades se calculan con suma facilidad. Si, por ejemplo, ε sigue una distribución logística, la función de distribución sería F(ε) = exp(ε)/(1+exp(ε)). La probabilidad de responder muy mala (opinión) sería

$$\text{Prob}(\textit{muy mala}) = \text{Prob}(U < k_4)$$

$$= \text{Prob}(\beta' x + \varepsilon < k_4)$$

$$= \text{Prob}(\varepsilon < k_4 - \beta' x)$$

$$= \frac{e^{k_4 - \beta' x}}{1 + e^{k_4 - \beta' x}}$$

Por otro lado

$$\text{Prob}(\textit{mala}) = \text{Prob}(k_4 < U < k_3)$$

$$= \text{Prob}(k_4 < \beta' x + \varepsilon < k_3)$$

$$= \text{Prob}(k_4 - \beta' x < \varepsilon < k_3 - \beta' x)$$

$$= \text{Prob}(\varepsilon < k_3 - \beta' x) - \text{Prob}(\varepsilon < k_4 - \beta' x)$$

$$= \frac{e^{k_3 - \beta' x}}{1 + e^{k_3 - \beta' x}} - \frac{e^{k_4 - \beta' x}}{1 + e^{k_4 - \beta' x}}$$

Lo mismo con el resto de probabilidades. La estimación se lleva a cabo por máxima verosimilitud. Los parámetros estimados consisten en las β (que nos dan el impacto de las variables explicativas en la opinión de la gente sobre la actuación del político), así como las k_i.

A este modelo se le conoce como *Logit ordenado* (*ordered Logit*). Las diferencias con el Logit estándar son evidentes: el *Logit binario* estándar tiene dos alternativas y dos funciones de utilidad, cada una asociada a una alternativa. En cambio, el *Logit ordenado* tiene múltiples alternativas y una sola función de utilidad. Sin embargo, la fórmula de las probabilidades es muy similar. Eso se debe a que si dos variables aleatorias son *iid* de valores extremos entonces su diferencia sigue una distribución logística. Por tanto, si asumimos que las dos utilidades de un *Logit binario* son *iid* de valores extremos, la diferencia de utilidades tendrá un término de error logístico, como el *Logit ordenado*.

Si suponemos que los ε siguen una distribución estándar normal en vez de una logística obtenemos un modelo diferente, pero igualmente operativo. La fórmula Logit binaria queda sustituida por la función de distribución normal estándar. El modelo recibe el nombre de *Probit ordenado* (*ordered Probit*). En nuestro ejemplo, las fórmulas de las probabilidades serían

$$\text{Prob}(\textit{muy mala}) = \text{Prob}(\varepsilon < k_4 - \beta' x)$$

$$= \Phi(k_4 - \beta' x)$$

y

$$\text{Prob}(\textit{mala}) = \text{Prob}(\varepsilon < k_3 - \beta' x) - \text{Prob}(\varepsilon < k_4 - \beta' x)$$

$$=\Phi(k_4 - \beta' x) - \Phi(k_4 - \beta' x)$$

donde Φ es la función de distribución normal estándar.

Si queremos permitir que los parámetros varíen aleatoriamente en la población tendremos que especificar una versión *mixta* del modelo (véase, por ejemplo, Bhat, 1999). Si el parámetro β sigue una función de densidad $g(\beta \mid \theta)$, las probabilidades de este *Logit mixto ordenado* serán simplemente las probabilidades del *Logit ordenado* integradas a lo largo de la función de densidad g(.). Por ejemplo,

$$\text{Prob}(\textit{muy mala}) = \int \left(\frac{e^{k_4 - \beta' x}}{1 + e^{k_4 - \beta' x}} \right) g(\beta|\theta)\, d\beta$$

y

$$\text{Prob}(\textit{mala}) = \int \left(\frac{e^{k_3 - \beta' x}}{1 + e^{k_3 - \beta' x}} - \frac{e^{k_4 - \beta' x}}{1 + e^{k_4 - \beta' x}} \right) g(\beta|\theta)\, d\beta$$

Estas probabilidades se calculan mediante simulación, como hacíamos con los Logit mixtos, tomando valores de β a partir de la función g(.), calculando la probabilidad logit ordenada para cada muestra y promediando los resultados. El Probit mixto ordenado se calcula de la misma forma.

***

Cuando hay varias respuestas ordenadas entrelazadas o conectadas de alguna forma el problema cambia un poco. Podemos imaginar una pregunta adicional a la relativa a la actuación del político, como una valoración de la situación económica, también con 5 opciones. En ese caso ambas respuestas estarán relacionadas. A la variable *U* que medía la satisfacción por la actuación del político se une ahora una variable *W* que mide la valoración de la situación económica. Como de costumbre, podemos explicitar unas funciones de utilidad para ambas variables

$$U = \beta' x + \varepsilon$$

$$W = \alpha' z + \mu$$

Las mismas variables pueden formar parte de *x* y de *z*, lo que explicaría en parte la conexión entre *U* y *W*. Pero esa relación puede deberse también a los factores inobservados. Para considerar esta posibilidad asumimos una distribución conjuntamente normal para ε y μ, con una correlación ρ (y varianzas unitarias por normalización). Los puntos de corte serán de nuevo $k_1, \dots, k_4$ para *U*, y $c_1, \dots, c_4$ para *W*. Queremos obtener la probabilidad de cada posible combinación de respuestas a las dos preguntas.

Veamos el caso en que la persona encuestada responde *muy mala* a la opinión sobre el político y *muy mala* a la situación económica:

$$\text{Prob}(\textit{muy mala}/\textit{muy mala})$$

$$= \text{Prob}(U < k_4 \text{ y } U < c_4)$$

$$= \text{Prob}(\varepsilon < k_4 - \beta' x \text{ y } \mu < c_4 - \alpha' z)$$

$$= \text{Prob}(\varepsilon < k_4 - \beta' x)$$

$$\text{x Prob}(\mu < c_4 - \alpha' z \mid \varepsilon < k_4 - \beta' x)$$

Aplicando la descomposición de una probabilidad en una condicional y una marginal. De la misma forma:

$$\text{Prob}(\textit{muy mala}/\textit{buena})$$

$$= \text{Prob}(U < k_4 \text{ y } c_2 < U < c_1)$$

$$= \text{Prob}(\varepsilon < k_4 - \beta' x \text{ y } c_2 - \alpha' z < \mu < c_1 - \alpha' z)$$

$$= \text{Prob}(\varepsilon < k_4 - \beta' x)$$

$$\text{x Prob}(c_2 - \alpha' z < \mu < c_1 - \alpha' z \mid \varepsilon < k_4 - \beta' x)$$

El resto de combinaciones se calcula de forma similar, y lo mismo puede decirse de más de dos preguntas enlazadas. A este modelo se le conoce como *Probit ordenado multivariante* (*multivariate ordered probit*). Las probabilidades se calculan por simulación, empleando un simulador como el GHK (Train, 2003, pp. 126 y siguientes). Sin embargo, a diferencia del Probit estándar, para el que la normal conjunta está truncada sólo por un lado, el Probit multivariante presenta truncamientos por ambos lados. Hajivassiliou y Ruud (1994) tienen un tratamiento explícito del GHK con truncamiento por ambos lados.

***

La *valoración contingente* es otra posible variante más, con sus particularidades (Hausman, 1993). En este tipo de encuestas se registra la reacción de la persona a una cantidad que fija el encuestador, por ejemplo, si acepta o rechaza una determinada cantidad referida a un caso objeto de estudio, como una cantidad de dinero por un servicio, o una subida de impuestos a cambio de alguna actuación pública, etc. Puede haber preguntas encadenadas. Si el entrevistado acepta la cantidad, por ejemplo, se podría repetir la pregunta para una cantidad superior, o inferior, hasta quedar determinado un límite, dentro de un rango. ¿Cuándo se emplean estas técnicas? Cuando la falta de mercado impide observar precios, y la preferencia revelada no tiene cabida. Un ejemplo son los estudios medioambientales.

Cuando sólo se hace una pregunta el método es de acotación simple (*single-bounded*), ya que la respuesta sólo proporciona una referencia acerca de la disposición a pagar: sabemos que la persona acepta la cantidad, pero no si aceptaría una cantidad superior, y de qué magnitud; o que la rechaza, pero no por cuánto margen. Cuando se plantea una segunda pregunta el método es de acotación doble (*double-bounded*). Si la persona rechaza la cantidad, la segunda pregunta puede determinar qué cantidad sí estaría dispuesta a pagar; y si la acepta, qué cantidad determinaría ya el rechazo. No obstante, puede que la respuesta a la segunda pregunta insista en el rechazo o la aceptación de la respuesta a la primera, con lo que no tendríamos necesariamente un rango determinado desde ambos extremos.

La cantidad que se incluye en la primera pregunta puede variar para cada persona encuestada. Con las respuestas se estima la distribución de la propensión a pagar. El procedimiento de estimación es muy similar al empleado para los Logit y Probits ordenados, sólo que los puntos de corte vienen determinados aquí por el diseño del cuestionario en vez de ser estimados como parámetros.

Si W_n representa la propensión a pagar de una persona n, tendremos que W_n variará entre personas con una distribución $f(W \mid \theta)$, donde θ son los parámetros de la distribución (como la media y la varianza). El objetivo del investigador es estimar esos parámetros en la población. Supongamos que el investigador diseña un cuestionario *single-bounded* con un valor de referencia distinto para cada persona de la muestra, siendo éste k_n. La persona responde "sí" siempre que $W_n > k_n$ y "no" en otro caso. El investigador asume que W_n sigue una distribución normal en la población, con media $\bar{W}$ y varianza σ^2.

La probabilidad del "sí" es

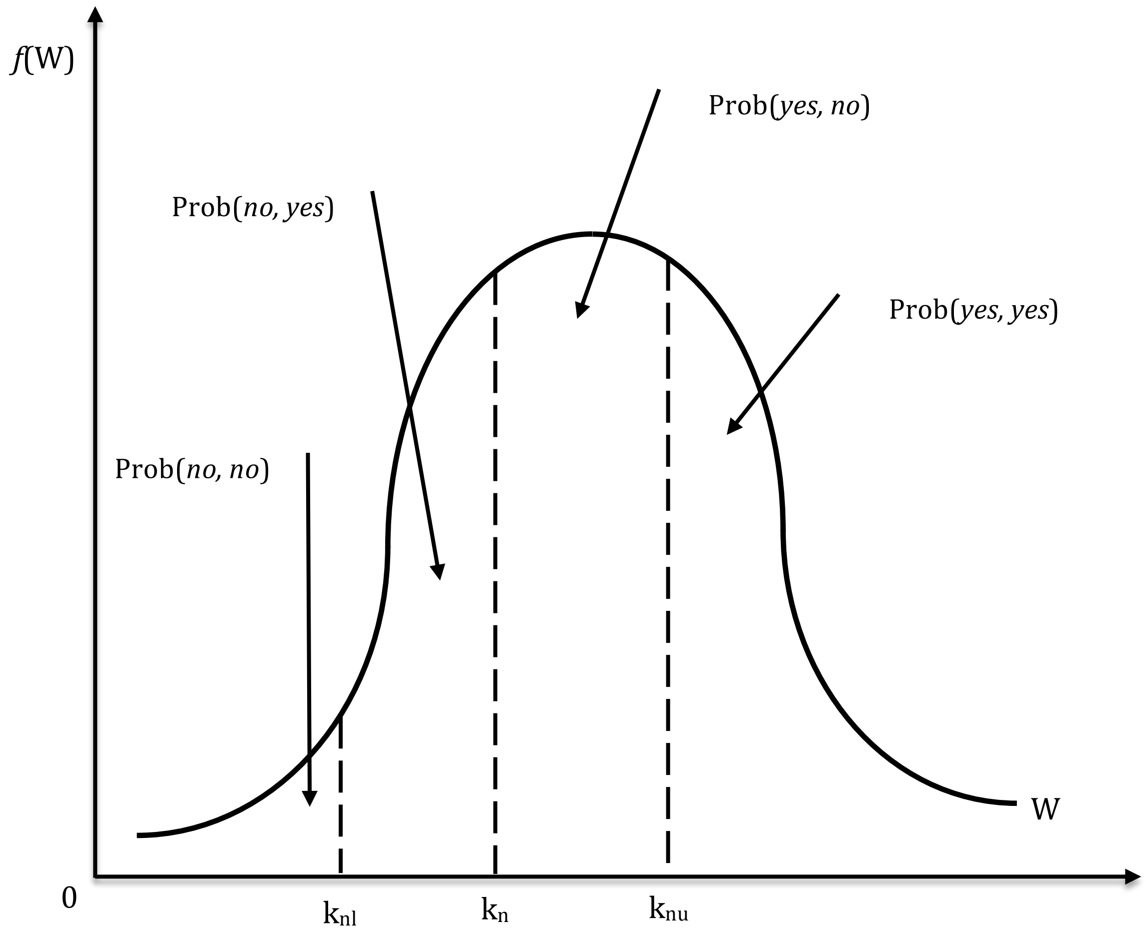
$$\mathrm{Prob}(W_n > k_n) = 1 - \mathrm{Prob}(W_n < k_n) = 1 - \Phi((k_n - \bar{W})/\sigma),$$

Y la probabilidad del "no" es $\Phi((k_n - \bar{W})/\sigma)$ donde $\Phi(.)$ es la distribución de probabilidad estándar normal. La función de verosimilitud es pues

$$\sum_n y_n \ln[1 - \Phi((k_n - \bar{W})/\sigma)] + (1 - y_n) \ln[\Phi((k_n - \bar{W})/\sigma)]$$

donde $y_n=1$ si la persona n dice "sí" y 0 en otro caso. Cuando maximizamos esa función de máxima verosimilitud obtenemos estimaciones de $\bar{W}$ y σ.

Si el cuestionario es *double-bounded* el procedimiento es similar al descrito, si bien el valor de referencia de la segunda pregunta será ahora k_{nu} si la persona responde "sí" a la primera, donde $k_{nu}>k_n$; y k_{nl} si la persona responde "no" a la primera pregunta, siendo $k_{nl}<k_n$. Hay cuatro posibles secuencias de respuestas a las dos preguntas, cuyas probabilidades quedan representadas en el siguiente gráfico:



Las probabilidades serían:

- (no, no): $P = Prob(W_n < k_{nl}) = \Phi((k_{nl} - \bar{W})/\sigma)$
- (no, sí): $P = Prob(k_{nl} < W_n < k_n) = \Phi((k_n - \bar{W})/\sigma) - \Phi((k_{nl} - \bar{W})/\sigma)$
- (sí, no): $P = Prob(k_n < W_n < k_{nu}) = \Phi((k_{nu} - \bar{W})/\sigma) - \Phi((k_n - \bar{W})/\sigma)$
- (sí, sí): $P = Prob(W_n > k_{nu}) = 1 - \Phi((k_{nu} - \bar{W})/\sigma)$

Estas probabilidades entran en la función de verosimilitud, que se maximiza para obtener las estimaciones de $\bar{W}$ y σ. Se pueden emplear otras estimaciones en lugar de la normal. Por ejemplo, la log-normal es atractiva si el investigador asume el supuesto de que todas las personas tienen una propensión a pagar positiva; también se puede diseñar una distribución tal que un buen número de observaciones tengan un valor cero, pues puede haber una proporción elevada de la población que no esté dispuesta a pagar nada, más una log-normal para los demás.

Se pueden aumentar las dimensiones del modelo (el número de preguntas en la encuesta) para reflejar que la demanda de algunos bienes puede estar relacionada, de forma que si estamos dispuestos a pagar por unos lo estemos también para pagar por otros. Como vimos en el Probit ordenado multirespuesta, el simulador GHK es útil cuando se asumen variables conjuntamente distribuidas como normales.

***

Hemos visto los modelos Logit mixto y Logit mixto ordenado. Siguiendo la misma lógica pueden desarrollarse modelos mixtos de todas clases. Cualquier modelo cuyas probabilidades pueden escribirse como una función de parámetros puede convertirse en "mixto" permitiendo que los parámetros sean aleatorios e integrando la función sobre la distribución de los parámetros. La probabilidad se calcula por simulación tomando muestras aleatorias de la distribución, calculándose la función para cada muestra obtenida y promediando los resultados. Dos posibles ejemplos serían el *Logit mixto anidado* y el *Probit mixto*.

El modelo Logit mixto, como sabemos, no exhibe la propiedad de *independencia respecto de las alternativas irrelevantes* (IIA) propia del Logit estándar, y puede adaptarse a cualquier patrón de sustitución mediante la apropiada especificación de variables y *mixing distribution*. Este hecho ha llevado a algunos a considerar superados los modelos Logit anidados. Un Logit mixto proporciona patrones de correlación y sustitución análogos a los de un Logit anidado. Por ejemplo, podemos considerar un Logit anidado con dos nidos para las alternativas A y B. Siempre que los coeficientes log-sum (λ_k) se encuentren entre 0 y 1 la sustitución dentro de cada nido será mayor que la sustitución entre nidos. Este patrón de sustitución puede ser representado en un Logit mixto mediante la especificación de una variable *dummy* para cada nido y permitiendo que los coeficientes de esas *dummies* sean aleatorios (restringiendo, para la especificación, las medias a cero siempre que se incluya un conjunto completo de constantes específicas para cada alternativa, e igualando las dos varianzas).

Al especificar un Logit mixto de esa forma la simulación queda afectada. Ésta se emplea para aproximar integrales cuando una forma cerrada de las mismas no existe. *La solución analítica siempre es más exacta, y la simulación debe usarse sólo cuando ésta no es posible*. Cuando se emplea un Logit mixto para representar los patrones de sustitución de un Logit anidado, cosa que es posible, sustituye la forma cerrada de la integral del Logit anidado por una integral sin forma cerrada que hay que simular. Esa sustitución implica una pérdida de precisión. El Logit mixto sólo aportaría dos ventajas: 1) puede hacer más fácil la estimación de numerosas estructuras de nidos, incluyendo nidos entrecruzados, y 2) si otros coeficientes son también aleatorios, un Logit mixto es necesario de todas formas.

Esta segunda razón sugiere la posibilidad de emplear un *Logit mixto anidado*. Supóngase que el investigador cree que alguno de los coeficientes en el modelo son aleatorios y también que los factores inobservados están correlacionados entre alternativas en una forma tal que puede representarse mediante un Logit anidado. Puede especificarse un *Logit mixto anidado* que represente esa situación. Condicionado a los coeficientes que entran en la utilidad, la probabilidad de elección es la propia de un Logit anidado, con una integral de forma cerrada que puede calcularse analíticamente. La probabilidad incondicional es la fórmula del Logit anidado, pero integrada sobre la distribución de los coeficientes aleatorios. La estimación presenta dificultades, y una *estimación bayesiana jerárquica* (Train, 2003, capítulo 12), que evita maximizar la función de verosimilitud, puede ser la mejor opción.

El Probit mixto es un caso distinto. El modelo Probit tiene una característica especial: *todos los términos aleatorios entran en la utilidad linealmente y están distribuidos aleatoriamente de forma que la propia utilidad sigue una distribución normal*. Esta característica, que es también una limitación, puede relajarse con un *Probit mixto*. Supongamos que algunos términos aleatorios entran en la utilidad de forma no lineal, o no están distribuidos aleatoriamente, pero que condicionando a esa situación la utilidad sigue una distribución normal. Por ejemplo, el coeficiente de un precio podría seguir una log-normal para asegurar así que es negativo para todo el mundo, mientras que el resto de coeficientes pueden ser fijos o aleatorios normales, siendo los términos de error finales conjuntamente normales. Un Probit mixto sería una solución apropiada para una especificación como esa. Condicionando a los coeficientes de los precios las probabilidades de elección siguen la fórmula estándar del Probit. Las probabilidades incondicionadas son la integral de esa fórmula Probit sobre la distribución del coeficiente de precios. Por tanto, son necesarias dos capas de simulación para obtener las probabilidades: 1) se extrae una muestra aleatoria de la distribución del coeficiente de precios; y 2) para esa muestra, el simulador GHK u otro simulador Probit se aplica a aproximar la *probabilidad de elección condicional*. Este proceso se repite muchas veces y los resultados se promedian.

Este proceso de simulación requiere mucho tiempo, pues el simulador GHK se aplica para cada muestra aleatoria del coeficiente de precios. Por otro lado, el Probit mixto permite evitar algunas de las dificultades prácticas que pueden surgir con el Logit mixto. Por ejemplo, cuando hay heterocedasticidad pura (varianzas distintas para la utilidad de cada alternativa) o un patrón de correlación enter alternativas fijo (una matriz de covarianzas que no depende de las variables), puede ser más fácil estimar un Probit en lugar de especificar numerosos componentes de error en un Logit mixto. *La especificación de la covarianza y la heterocedasticidad puede ser más compleja en un Logit mixto que en un Probit, porque los términos aleatorios iid de valores extremos se suman necesariamente a cualesquiera otros elementos aleatorios que el investigador especifique*. El Probit ofrece una especificación menos tortuosa en esos casos, salvo que haya términos aleatorios no-normales, en cuyo caso sería necesario un Probit mixto. Éste preserva la facilidad del Probit para representar una covarianza fija para los errores aditivos. El único coste de un Probit aquí es el coste o tiempo de computación.