3

Logit

3.1 Probabilidades de elección

Logit es con diferencia el modelo de elección discreta más simple y de uso más extendido. Su popularidad se debe al hecho de que la fórmula para las probabilidades de elección tiene una expresión cerrada y es fácilmente interpretable. Originalmente, la fórmula logit fue obtenida por Luce (1959) a partir de ciertas asunciones sobre las características de las probabilidades de elección, la principal de las cuales era la independencia de alternativas irrelevantes (*independence from irrelevant alternatives*, IIA), tratada en la sección 3.3.2. Marschak (1960) mostró que estos axiomas implicaban que el modelo era consistente con un comportamiento del decisor orientado a la maximización de la utilidad. La relación de la fórmula logit con la distribución de la utilidad no observada (como opuesta a las características de las probabilidades de elección) fue desarrollada por Marley, tal y como citan Luce y Suppes (1965), quienes mostraron que la distribución de valor extremo conduce a la fórmula logit. McFadden (1974) completó el análisis mostrando la relación inversa, es decir, que la fórmula logit para las probabilidades de elección necesariamente implica que la utilidad no observada se distribuye de acuerdo a una distribución de valor extremo. En la ceremonia de entrega de su premio Nobel, McFadden (2001) explicó una historia fascinante sobre el desarrollo de este modelo pionero.

Para obtener el modelo logit, usamos la notación general del Capítulo 2 y añadimos una distribución específica para la utilidad no observada. Un decisor etiquetado como n se enfrenta a J alternativas. La utilidad que el decisor obtiene de la alternativa j se descompone en (1) una parte denominada V_{nj} que es conocida por el investigador a través de algunos parámetros, y (2) una parte ε_{nj} desconocida que es tratada por el investigador como una variable aleatoria: $U_{nj} = V_{nj} + \varepsilon_{nj}\ \forall j$. El modelo logit se obtiene suponiendo que cada ε_{nj} se distribuye independientemente y de forma idénticamente distribuida de acuerdo a una densidad de probabilidad de tipo valor extremo. Esta distribución también se denomina Gumbel y tipo I valor extremo (y en ocasiones, de forma errónea, Weibull). La densidad para cada componente no observado de utilidad es

$$f(\varepsilon_{nj}) = e^{-\varepsilon_{nj}} e^{-e^{-\varepsilon_{nj}}}, \tag{3.1}$$

y la distribución acumulativa es

$$F(\varepsilon_{nj}) = e^{-e^{-\varepsilon_{nj}}}. \tag{3.2}$$

La varianza de esta distribución es $\pi^2/6$. Asumiendo que la varianza es $\pi^2/6$, implícitamente estamos normalizando la escala de la utilidad, como se trata en la sección 2.5. Volveremos sobre este tema y su relevancia para la interpretación del modelo en la siguiente sección. La media de la distribución de valor extremo no es cero, sin embargo, la media es irrelevante ya que sólo las diferencias entre utilidades importan (véase el capítulo 2) y la diferencia entre dos términos aleatorios que tienen la misma media tiene una media igual a cero.

La diferencia entre dos variables de tipo valor extremo se distribuye de forma logística. Es decir, si ε_{nj} y ε_{ni} son de tipo valor extremo iid, entonces $\varepsilon^*_{nji} = \varepsilon_{nj} - \varepsilon_{ni}$ sigue una distribución logística

$$\textit{(3.3)} \qquad \boldsymbol{F}(\varepsilon^*_{nji}) = \frac{\boldsymbol{e}^{\varepsilon^*_{nji}}}{\boldsymbol{1}+\boldsymbol{e}^{\varepsilon^*_{nji}}}.$$

Esta fórmula se utiliza en ocasiones para describir modelos logit binarios, es decir, modelos con dos alternativas. Usar la distribución de valor extremo para los errores (y por lo tanto la distribución logística para las diferencias entre errores) es casi lo mismo que asumir que los errores se distribuyen normalmente y de forma independiente. La distribución de valor extremo tiene colas ligeramente más gruesas que una distribución normal, lo que implica que permite un comportamiento ligeramente más aberrante que la normal. Por lo general, sin embargo, la diferencia entre errores distribuidos según el valor extremo y según distribuciones normales independientes es indistinguible empíricamente.

El supuesto clave del modelo no es tanto la forma de la distribución como que los errores sean independientes entre sí. Esta independencia significa que la parte no observada de la utilidad de una alternativa no está relacionada con la parte no observada de la utilidad de otra alternativa. Es un supuesto bastante restrictivo, por lo que el desarrollo de otros modelos como los descritos en los capítulos 4-6 ha surgido en gran medida con el fin de evitar este supuesto y permitir la existencia de errores correlacionados.

No obstante, es importante observar que la hipótesis de independencia no es tan restrictiva como podría parecer a primera vista y que, de hecho, puede ser interpretada como un resultado natural de un modelo bien especificado. Recuerde del capítulo 2 que ε_{nj} se define como la diferencia entre la utilidad que el decisor obtiene realmente, U_{nj}, y la representación de la utilidad que el investigador ha desarrollado utilizando las variables observadas, V_{nj}. Como tal, ε_{nj} y su distribución dependen de la especificación que el investigador haga de la utilidad representativa; no está definida por la situación de elección *per se*. En este sentido, la hipótesis de independencia permite una interpretación diferente. Bajo la hipótesis de independencia, el error de una alternativa no proporciona al investigador ninguna información sobre el error de otra alternativa diferente. Dicho de forma equivalente, el investigador ha especificado V_{nj} lo suficiente para que el resto de la utilidad (no observada) sea esencialmente "ruido blanco". En un sentido profundo, el objetivo último del investigador es representar la utilidad tan bien que los únicos aspectos que queden sin representar constituyan simplemente ruido blanco; es decir, el objetivo es especificar la utilidad suficientemente bien como para que un modelo logit sea apropiado. Visto de esta forma, logit es el modelo ideal en lugar de una restricción.

Si el investigador considera que la parte no observada de la utilidad está correlacionada entre alternativas dada su especificación de la utilidad representativa, tiene tres opciones: (1) utilizar un modelo diferente que permita errores correlacionados, tales como los descritos en los capítulos 4-6, (2) especificar de nuevo la utilidad representativa de forma que la fuente de correlación quede capturada explícitamente y por lo tanto los errores restantes sean independientes o (3) utilizar el modelo logit bajo la especificación actual de la utilidad representativa y considerar el modelo como una aproximación. La viabilidad de la última opción depende, por supuesto, de los objetivos de la investigación. Las violaciones de los supuestos del modelo logit parecen tener menos efecto en la estimación de las preferencias medias que en el pronóstico de patrones de sustitución (cambio en las preferencias futuras al alterar los atributos de las alternativas). Estas cuestiones se tratan en las siguientes secciones.

Derivamos a continuación las probabilidades de elección logit, siguiendo la aproximación de McFadden (1974). La probabilidad de que el decisor n elija la alternativa i es

$$P_{ni} = Prob(V_{ni} + \varepsilon_{ni} > V_{nj} + \varepsilon_{nj} \;\; \forall j \neq i)$$

(3.4) $$= Prob\big(\varepsilon_{nj} < \varepsilon_{ni} + V_{ni} - V_{nj} \;\; \forall j \neq i\big).$$

Si consideramos ε_{ni} como dado, esta expresión es la distribución acumulativa para cada ε_{nj} evaluada en $\varepsilon_{ni} + V_{ni} - V_{nj}$, que de acuerdo con (3.2) es $exp(-exp(-(\varepsilon_{ni} + V_{ni} - V_{nj})))$. Dado que los εs son independientes, esta distribución acumulativa sobre todo $j \neq i$ es el producto de las distribuciones acumulativas individuales:

$$P_{ni}|\varepsilon_{ni} = \prod_{j \neq i} e^{-e^{-\left(\varepsilon_{ni}+V_{ni}-V_{nj}\right)}}.$$

Por supuesto, ε_{ni} no está realmente dado, por lo que la probabilidad de elección es la integral de $\mathrm{P}_{ni}|\varepsilon_{ni}$ sobre todos los valores de ε_{ni} ponderados por su densidad de probabilidad (3.1):

(3.5) $$P_{ni} = \int \left(\prod_{j \neq i} e^{-e^{-\left(\varepsilon_{ni}+V_{ni}-V_{nj}\right)}}\right) e^{-\varepsilon_{ni}}\, e^{-e^{-\varepsilon_{ni}}}\, d\varepsilon_{ni}.$$

Algunas manipulaciones algebraicas de esta integral resultan en una expresión cerrada y compacta:

(3.6) $$P_{ni} = \frac{e^{V_{ni}}}{\sum_j e^{V_{nj}}},$$

que es la probabilidad de elección logit. El álgebra que obtiene (3.6) de (3.5) se detalla en la última sección de este capítulo.

La utilidad representativa suele especificarse de forma que sea lineal en relación a los parámetros: $V_{nj} = \beta' x_{nj}$, donde x_{nj} es un vector de variables observadas en la alternativa j. Con esta especificación, las probabilidades logit se convierten en

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

En condiciones bastante generales, cualquier función puede aproximarse de forma arbitrariamente precisa por una función lineal en parámetros. La asunción es bastante buena por lo tanto. Es importante destacar que McFadden (1974) demostró que la función logaritmo de la verosimilitud (log-verosimilitud o *log-likelihood*) con estas probabilidades de elección, es globalmente cóncava respecto a los parámetros β, lo cual ayuda en los procedimientos de maximización numérica (como se trata en el Capítulo 8). Numerosos paquetes de software contienen rutinas para la estimación de modelos logit con utilidad representativa lineal en parámetros.

Las probabilidades logit exhiben varias propiedades que son deseables. En primer lugar, P_{ni} está necesariamente entre cero y uno, un requisito para que pueda ser una probabilidad. Cuando V_{ni} crece, lo que refleja una mejora en los atributos observados de la alternativa, manteniendo $\mathrm{V}_{nj} \forall j \neq i$ constante, P_{ni} se acerca a uno. Y P_{ni} se acerca a cero cuando V_{ni} disminuye, ya que la exponencial en el numerador de (3.6) se aproxima a cero a medida que V_{ni} se acerca a $-\infty$. La probabilidad logit para una alternativa nunca es exactamente cero. Si el investigador cree que una alternativa no tiene en realidad ninguna posibilidad de ser elegida por un decisor, puede excluir la alternativa del conjunto de elección. Una probabilidad exactamente igual a 1 se obtiene sólo si el conjunto de elección consiste en una única alternativa.

En segundo lugar, las probabilidades de elección de todas las alternativas suman uno: $\sum_{i=1}^{J} P_{ni} = \sum_i \exp(\mathrm{V}_{ni}) \,/\, \sum_j \exp(\mathrm{V}_{nj}) = 1$. El decisor necesariamente elige una de las alternativas. El denominador de (3.6) es simplemente la suma del numerador sobre todas las alternativas, lo que produce esta propiedad de suma de forma automática. Con logit, así como con algunos modelos un poco más complejos como el logit jerárquico descrito en el Capítulo 4, la interpretación de las probabilidades de elección se ve facilitada al observarse que el denominador sirve para asegurar que las probabilidades sumen uno. En otros modelos, como el logit mixto y el probit, no hay un denominador *per se* que pueda ser interpretado de esta manera.

La relación entre la probabilidad logit y la utilidad representativa es una sigmoidea o función-S, tal y como se muestra en la figura 3.1.

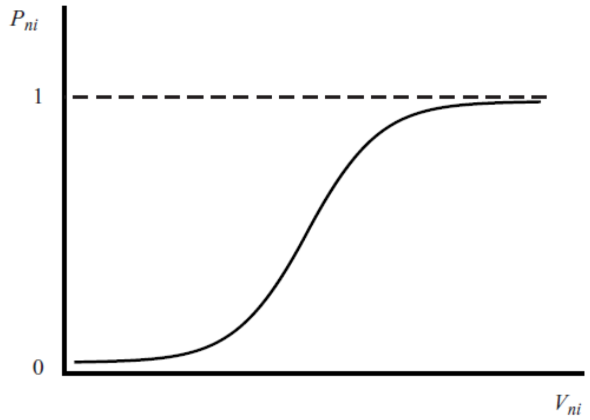

Figura 3.1. Gráfica de la curva logit.

Esta forma tiene implicaciones en el impacto que producen los cambios de las variables explicativas. Si la utilidad representativa de una alternativa es muy baja en comparación con otras alternativas, un pequeño aumento en la utilidad de la alternativa tiene poco efecto sobre la probabilidad de que sea elegida: las otras alternativas son todavía suficientemente mejores por lo que esta pequeña mejora no ayuda mucho. Del mismo modo, si una alternativa es muy superior a las demás en los atributos observados, un aumento adicional en su utilidad representativa tiene poco efecto sobre la probabilidad de elección. El punto en el que el aumento en la utilidad representativa tiene mayor efecto sobre la probabilidad de ser elegida es cuando la probabilidad es próxima a 0.5, lo que significa una probabilidad del 50% de que la alternativa sea elegida. En este caso, una pequeña mejora es decisiva en las elecciones de las personas, induciendo un gran cambio en la probabilidad. La forma sigmoidea de las probabilidades logit se observa en la mayoría de los modelos de elección discreta y tiene implicaciones importantes para los reguladores y responsables de legislar. Por ejemplo, mejorar el servicio de autobuses en zonas en las que el servicio es tan pobre que pocos viajeros viajan en autobús sería menos eficaz, en términos de uso del transporte público, que hacer la misma mejora en áreas donde el servicio de autobús es suficientemente bueno como para que una parte moderada de viajeros ya lo esté usando (pero no tan bueno como para que casi todo el mundo lo esté usando).

La fórmula de probabilidad logit es fácilmente interpretable en el contexto de un ejemplo. Considere en primer lugar una situación de elección binaria: la elección de un hogar entre un sistema de calefacción eléctrica o de gas. Supongamos que la utilidad que obtiene el hogar de cada tipo de sistema depende sólo del precio de compra, el costo anual de operación, el punto de vista del hogar respecto a la conveniencia y calidad de calentarse con cada tipo de sistema, así como la estética de los sistemas dentro de la casa. Los dos primeros factores pueden ser observados por el investigador, pero no el resto de factores. Si el investigador considera que la parte observada de la utilidad es una función lineal de los factores observados, la utilidad de cada sistema de calefacción se puede expresar como: $\mathrm{U}_g = \beta_1 PP_g + \beta_2 OC_g + \varepsilon_g$ y $\mathrm{U}_e = \beta_1 PP_e + \beta_2 OC_e + \varepsilon_e$, donde el subíndice g y e se refieren a gas y electricidad, PP y OC son el precio de compra (*purchase price*) y los costos de operación (*operating cost*), β_1 y β_2 son parámetros escalares y el subíndice n para el hogar se suprime para simplificar la notación. Dado que costos más altos implican menos dinero para gastar en otros bienes, esperamos que la utilidad disminuya a medida que el precio de compra o el costo de operación suban (con todo lo demás constante): $\beta_1 < 0$ y $\beta_2 < 0$.

El componente no observado de la utilidad de cada alternativa, ε_g y ε_e, varía en los hogares en función de cómo ve cada hogar la calidad, comodidad y estética de cada tipo de sistema. Si estos componentes no observados se distribuyen con una densidad valor extremo iid, entonces la probabilidad de que el hogar elija la calefacción de gas es

(3.7) $$P_g = \frac{e^{\beta_1 PP_g + \beta_2 OC_g}}{e^{\beta_1 PP_g + \beta_2 OC_g} + e^{\beta_1 PP_e + \beta_2 OC_e}}$$

y la probabilidad de que elija la calefacción eléctrica es la misma pero con $\exp(\beta_1 PP_e + \beta_2 OC_e)$ como numerador. La probabilidad de elegir un sistema de gas disminuye cuando su precio de adquisición o costo de operación sube, permaneciendo igual el del sistema eléctrico (suponiendo que β_1 y β_2 son negativos, como es de esperar).

Como en la mayoría de los modelos de elección discreta, el ratio entre los coeficientes de este ejemplo tiene sentido económico. En particular, el ratio β_2/β_1 representa la disposición de la familia a pagar para obtener una reducción de costos de operación. Si β_1 se ha estimado con valor -0.20 y β_2 con valor -1.14, estas estimaciones implicarían que los hogares están dispuestos a pagar hasta (-1,14)/(-0,20)=5,70 dólares más por un sistema cuyo costo de operación anual sea un dólar menos. Esta relación se obtiene de la siguiente manera. Por definición, la disposición de un hogar a pagar por la reducción del costo de operación es el incremento en el precio de compra que mantiene constante la utilidad del hogar dada una reducción en los costos operativos. Tomamos la derivada total de la utilidad respecto al precio de compra y al costo de operación, y fijamos esta derivada igual a cero, por lo que la utilidad no cambia: $\mathrm{dU} = \beta_1 dPP + \beta_2 dOC = 0$. A continuación, resolvemos la ecuación para el cambio en el precio de compra que mantenga la utilidad constante (es decir, que satisfaga esta ecuación) frente a un cambio en los costos de operación: $\partial \mathrm{PP}/\,\partial \mathrm{OC} = -\beta_2/\beta_1$. El signo negativo indica que los dos cambios están en la dirección opuesta: para mantener constante la utilidad, el precio de compra se eleva cuando el costo de operación disminuye.

En esta situación de elección binaria, las probabilidades de elección se pueden expresar de forma más sucinta. Dividiendo el numerador y el denominador de (3.7) por el numerador, y reconociendo que $\exp(\mathrm{a})/\exp(\mathrm{b}) = \exp(\mathrm{a}-\mathrm{b})$, tenemos

$$P_g = \frac{1}{1 + e^{(\beta_1 PP_e + \beta_2 OC_e) - (\beta_1 PP_g + \beta_2 OC_g)}}.$$

En general, las probabilidades logit binarias con utilidades representativas V_{n1} y V_{n2} pueden ser rescritas como $\mathrm{P}_{n1} = 1/(1 + \exp(\mathrm{V}_{n2} - \mathrm{V}_{n1}))$ y $\mathrm{P}_{n2} = 1/(1 + \exp(\mathrm{V}_{n1} - \mathrm{V}_{n2}))$. Si sólo las características sociodemográficas del decisor, s_n, entran en el modelo y los coeficientes de estas variables sociodemográficas se normalizan a cero para la primera alternativa (como se describe en el Capítulo 2), la probabilidad de elección de la primera alternativa es $\mathrm{P}_{n1} = 1/(1 + e^{\alpha' s_n})$, que es la forma que se utiliza en la mayoría de los libros de texto y manuales de informática para el logit binario.

La elección multinomial es una extensión simple. Supongamos que hay un tercer tipo de sistema de calefacción, por ejemplo con petróleo (*oil,* indicado con subíndice "o") como combustible. La utilidad del sistema de petróleo se especifica de la misma forma que para los sistemas de electricidad y gas: $\mathrm{U}_o = \beta_1 PP_o + \beta_2 OC_o + \varepsilon_o$. Con esta opción adicional disponible, la probabilidad de que el hogar elija un sistema de gas es

$$P_g = \frac{e^{\beta_1 PP_g + \beta_2 OC_g}}{e^{\beta_1 PP_g + \beta_2 OC_g} + e^{\beta_1 PP_e + \beta_2 OC_e} + e^{\beta_1 PP_o + \beta_2 OC_o}},$$

que es igual a (3.7), excepto un término adicional incluido en el denominador para representar el calentador de petróleo. Dado que el denominador pasa a ser mayor mientras que el numerador permanece igual, la probabilidad de elegir un sistema de gas es más pequeña cuando está disponible un sistema de petróleo que cuando no lo está, como sería de esperar en el mundo real.

3.2 El parámetro de escala

En el apartado anterior hemos obtenido la fórmula logit bajo el supuesto de que los factores no observados se distribuyen con densidad valor extremo y varianza $\pi^2/6$. Fijar la varianza a $\pi^2/6$ es equivalente a normalizar el modelo respecto a la escala de la utilidad, como se trata en la Sección 2.5. Para hacer estos conceptos más explícitos es útil mostrar el papel que la varianza de los factores no observados juega en los modelos logit.

En general, la utilidad se puede expresar como $\mathrm{U}^*_{nj} = V_{nj} + \varepsilon^*_{nj}$, donde la parte no observada de la utilidad tiene varianza $\sigma^2 \times (\pi^2/6)$. Es decir, la varianza es cualquier número re-expresado como un múltiplo de $\pi^2/6$. Como la escala de la utilidad es irrelevante para el comportamiento, la utilidad se puede dividir por σ sin cambiar el comportamiento. La utilidad pasa a ser $\mathrm{U}_{nj} = V_{nj}/\sigma + \varepsilon_{nj}$ donde $\varepsilon_{nj} = \varepsilon^*_{nj}/\sigma$. Ahora, la parte no observada de la utilidad tiene varianza $\pi^2/6$: $\mathrm{Var}(\varepsilon_{nj}) = \mathrm{Var}(\varepsilon^*_{nj}/\sigma) = (1/\sigma^2)\mathrm{Var}(\varepsilon^*_{nj})$=$(1/\sigma^2) \times \sigma^2 \times (\pi^2/6) = \pi^2/6$. La probabilidad elección es

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

que es la misma fórmula de la ecuación (3.6) pero con la utilidad representativa dividida por σ. Si V_{nj} es lineal en los parámetros con coeficientes β^*, las probabilidades de elección pasan a ser

$$P_{ni} = \frac{e^{(\beta^*/\sigma)' x_{ni}}}{\sum_j e^{(\beta^*/\sigma)' x_{nj}}},$$

Cada uno de los coeficientes está escalado por un factor $1/\sigma$. El parámetro σ se denomina el *parámetro de escala*, ya que escala los coeficientes para reflejar la varianza de la parte no observada de la utilidad.

Sólo es posible estimar la relación β^*/σ; β^* y σ no pueden identificarse por separado. Por lo general, el modelo se expresa en su forma reducida, con $\beta = \beta^*/\sigma$, lo que genera la expresión logit estándar

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

Los parámetros β son los que se estiman en el modelo, pero para la interpretación es útil reconocer que estos parámetros estimados son en realidad las estimaciones de los coeficientes "originales" β^* divididos por el parámetro de escala σ. Los coeficientes estimados indican el efecto de cada variable observada en relación a la varianza de los factores no observados. Una variación mayor en los factores no observados conduce a estimar coeficientes más pequeños, incluso si los factores observados tienen el mismo efecto en la utilidad (es decir, mayor σ significa β inferiores incluso si β^* es el mismo).

El parámetro de escala no afecta al ratio entre dos coeficientes cualesquiera, dado que queda suprimido en la división; por ejemplo, $\beta_1/\beta_2 = (\beta_1^*/\sigma)/(\beta_2^*/\sigma) = \beta_1^*/\beta_2^*$, donde los subíndices se refieren al primer y segundo coeficientes. La disposición a pagar, valor del tiempo y otras medidas de tasas marginales de sustitución no se ven afectadas por el parámetro de escala. Sólo se ve afectada la interpretación de las magnitudes de todos los coeficientes.

Hasta ahora hemos supuesto que la varianza de los factores no observados es la misma para todos los decisores, ya que la misma σ se utiliza para todo n. Supongamos ahora que los factores no observados tienen mayor varianza para unos decisores que para otros. En la Sección 2.5 se trata una situación en la que la varianza de factores no observados es diferente en Boston y en Chicago. Denotemos la varianza para todos los decisores de Boston como $(\sigma^B)^2(\pi^2/6)$ y para los decisores de Chicago como$(\sigma^C)^2(\pi^2/6)$. El ratio de la varianza en Chicago respecto a la de Boston es $k = (\sigma^C/\sigma^B)^2$. Las probabilidades de elección para las personas de Boston pasan a ser

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

y para la gente de Chicago

$$P_{ni} = \frac{e^{(\beta/\sqrt{k})' x_{ni}}}{\sum_j e^{(\beta/\sqrt{k})' x_{nj}}},$$

donde $\beta = \beta^*/\sigma^B$. El ratio de las varianzas k se puede estimar junto con los coeficientes β. Los valores de β estimados se interpretan en relación con la varianza de los factores no observados en Boston y la k estimada proporciona información sobre la varianza en Chicago relativa a la de Boston. Es posible obtener relaciones más complejas permitiendo que la varianza para una observación dependa de más factores. Además, a menudo es de esperar que los datos de diferentes conjuntos de datos puedan tener diferente varianza para factores no observados, dando un parámetro de escala diferente para cada conjunto de datos. Ben-Akiva y Morikawa (1990) y Swait y Louviere (1993) tratan estos temas y proporcionan más ejemplos.

3.3 Potencia y limitaciones de logit

Tres cuestiones permiten dilucidar la potencia que los modelos logit tienen para representar el comportamiento de elección, así como delimitar los límites de esta potencia. Estas cuestiones son: variación de preferencias (*taste variation*), patrones de sustitución y elecciones reiteradas a lo largo del tiempo.

La aplicabilidad de los modelos logit se puede resumir de la siguiente manera:

1. Logit puede representar la variación sistemática de la preferencia (es decir, la variación de preferencia que se relaciona con las características observadas del decisor) pero no la variación de preferencia aleatoria (diferencias en las preferencias que no pueden vincularse a las características observadas).

2. El modelo logit implica sustitución proporcional entre alternativas, dada la especificación de la utilidad representativa hecha por el investigador. Para capturar formas más flexibles de sustitución, se necesitan otros modelos.

3. Si en situaciones de elecciones repetidas, los factores no observados son independientes a lo largo del tiempo, los modelos logit pueden capturar la dinámica de la elección repetida, incluyendo la dependencia del estado. Sin embargo, logit no puede manejar situaciones en las que los factores no observados se correlacionan a lo largo del tiempo.

Elaboramos cada una de estas afirmaciones en las próximas tres subsecciones.

3.3.1 Variación de preferencias

El valor o importancia que los decisores dan a cada atributo de las alternativas varía, en general, para los diferentes decisores. Por ejemplo, el tamaño de un automóvil es probablemente más importante para los hogares con muchos miembros que para los hogares más pequeños. Los hogares con ingresos bajos probablemente están más preocupados por el precio de compra de un bien, en relación al resto de sus características, que los hogares con mayores ingresos. En la elección de en qué barrio vivir, los hogares con niños pequeños estarán más preocupados por la calidad de las escuelas que aquellos hogares sin hijos y así sucesivamente. Las preferencias de los decisores también varían por razones que no están vinculadas a características sociodemográficas observadas, simplemente porque personas diferentes son diferentes. Dos personas que tienen el mismo nivel de ingresos, educación, etc., harán diferentes elecciones, lo que refleja sus preferencias individuales y preocupaciones.

Los modelos logit pueden capturar variaciones de preferencia, pero sólo dentro de ciertos límites. En particular, se pueden incorporar en los modelos logit preferencias que varían sistemáticamente respecto a variables observadas, mientras que preferencias que varían con variables no observadas o puramente al azar no pueden ser manejadas. El siguiente ejemplo ilustra la distinción.

Considere la elección que realizan los hogares en el momento de comprar un automóvil entre marcas y modelos disponibles. Supongamos, por simplicidad, que los dos únicos atributos de los automóviles que el investigador observa son el precio de compra, PP_j para la marca/modelo j, y los centímetros de espacio para los hombros, SR_j, que es una medida del tamaño del interior de un automóvil. El valor que las familias dan a estos dos atributos varía entre los diferentes hogares, por lo que la utilidad se escribe como

$$U_{nj} = \alpha_n SR_j + \beta_n PP_j + \varepsilon_{nj} \tag{3.8}$$

donde α_n y β_n son parámetros específicos para el hogar n.

Los parámetros varían entre los hogares reflejando diferencias en preferencias. Supongamos por ejemplo que el valor otorgado al espacio para los hombros varía con el número de miembros de los hogares, $\mathrm{M_n}$, pero nada más:

$$\alpha_n = \rho M_n,$$

de manera que a medida que aumenta $\mathrm{M_n}$ el valor otorgado al espacio para los hombros, α_n, también aumenta. De forma similar, supongamos que la importancia del precio de compra se relaciona inversamente con el nivel de ingresos, I_n, por lo que los hogares de bajos ingresos dan más importancia al precio de compra:

$$\beta_n = \theta / I_n.$$

Sustituyendo estas relaciones en (3.8) resulta

$$U_{nj} = \rho(M_n SR_j) + \theta(PP_j/I_n) + \varepsilon_{nj}.$$

Bajo el supuesto de que cada ε_{nj} se distribuye valor extremo iid, obtenemos un modelo logit estándar con dos variables describiendo la utilidad representativa, siendo ambas una interacción entre un atributo del vehículo y una característica del hogar.

Podríamos emplear otras especificaciones para la variación en las preferencias. Por ejemplo, podríamos asumir que el valor del espacio para los hombros aumenta con el tamaño del hogar, pero con una tasa decreciente, de modo que $\alpha_n = \rho \mathrm{M}_n + \phi M_n^2$, donde se espera que ρ sea positivo y ϕ negativo. Entonces $U_{nj} = \rho\left(M_n SR_j\right) + \phi(M_n^2 SR_j) + \theta(PP_j/I_n) + \varepsilon_{nj}$, lo que resulta en un modelo logit con tres variables entrando en la utilidad representativa.

La limitación del modelo logit surge cuando intentamos permitir variación de preferencias respecto a variables no observadas o puramente al azar. Supongamos, por ejemplo, que el valor del espacio para los hombros varía con el tamaño del hogar y con algunos otros factores (por ejemplo, el tamaño de las propias personas o la frecuencia con la que la familia viaja junta) que pasan desapercibidos para el investigador y, por tanto, son considerados factores aleatorios:

$$\alpha_n = \rho M_n + \mu_n,$$

donde μ_n es una variable aleatoria. Del mismo modo, la importancia del precio de compra consta de sus componentes observados y no observados:

$$\beta_n = \theta/I_n + \eta_n.$$

Sustituyendo en (3.8) resulta en

$$U_{nj} = \rho\left(M_n SR_j\right) + \mu_n SR_j + \theta\left(PP_j/I_n\right) + \eta_n PP_j + \varepsilon_{nj}.$$

Dado que μ_n y η_n no se observan, los términos $\mu_n SR_j$ y $\eta_n PP_j$ pasan a formar parte del componente no observado de la utilidad,

$$U_{nj} = \rho\left(M_n SR_j\right) + \theta\left(PP_j/I_n\right) + \tilde{\varepsilon}_{nj},$$

donde $\tilde{\varepsilon}_{nj} = \mu_n SR_j + \eta_n PP_j + \varepsilon_{nj}$. Los nuevos términos de error $\tilde{\varepsilon}_{nj}$ posiblemente no pueden estar distribuidos idénticamente y de forma independiente, como se requiere para la formulación logit. Desde el momento en que μ_n y η_n entran a formar parte de cada alternativa, $\tilde{\varepsilon}_{nj}$ está necesariamente correlacionada entre las alternativas: $Cov\left(\tilde{\varepsilon}_{nj}, \tilde{\varepsilon}_{nk}\right) = Var(\mu_n) SR_j SR_k + Var(\eta_n) PP_j PP_k \neq 0$ para cualquier pareja de modelos de automóvil j y k. Además, puesto que SR_j y PP_j varían entre alternativas, la varianza de $\tilde{\varepsilon}_{nj}$ varía entre alternativas, violando la asunción de errores idénticamente distribuidos: $Var\left(\tilde{\varepsilon}_{nj}\right) = Var(\mu_n) SR_j{}^2 + Var(\eta_n) PP_j{}^2 + Var(\varepsilon_{nj})$, que es diferente para los diferentes j.

Este ejemplo ilustra la idea general de que cuando las preferencias varían de forma sistemática respecto a las variables observadas, la variación se puede incorporar en los modelos logit. Al contrario, si la variación de preferencia es al menos en parte debida al azar, logit es una mala especificación. Como aproximación, logit podría ser capaz de captar las preferencias promedio bastante bien, incluso cuando las preferencias son al azar, ya que la fórmula logit parece ser bastante robusta frente a malas especificaciones. El investigador podría, por tanto, optar por utilizar logit incluso cuando se sabe que las preferencias tienen un componente aleatorio, en aras de la simplicidad. Sin embargo, no hay garantías de que un modelo logit se vaya a aproximar a las preferencias promedio. E incluso si lo hace, logit no proporcionará información sobre la distribución de las preferencias alrededor de la media. Esta distribución puede ser importante en muchas situaciones, tales como en la predicción de la penetración de mercado que logrará un nuevo producto que atrae a una minoría de personas en lugar de dirigirse a las preferencias promedio. Para incorporar la variación aleatoria de preferencias de forma apropiada y completa, pueden utilizarse en su lugar un modelo probit o un logit mixto.

3.3.2 Patrones de sustitución

Cuando los atributos de una alternativa mejoran (por ejemplo, su precio baja), la probabilidad de ser elegida aumenta. Algunas de las personas que habrían elegido otras alternativas con sus atributos originales ahora eligen esta alternativa en su lugar. Dado que las probabilidades de todas las alternativas suman uno, un aumento en la probabilidad de una alternativa necesariamente implica una disminución de la probabilidad de que otras alternativas sean escogidas. El patrón de sustitución entre alternativas tiene implicaciones importantes en numerosas situaciones.

Por ejemplo, cuando un fabricante de teléfonos celulares lanza un nuevo producto con prestaciones adicionales, está muy interesada en conocer en qué medida el nuevo producto atraerá a los clientes de sus otros modelos de teléfonos celulares en lugar de los teléfonos de la competencia, ya que la empresa logra más beneficios de lo segundo que de lo primero. Como veremos, el patrón de sustitución también afecta a la demanda de un producto y al cambio en la demanda cuando sus atributos cambian. Por lo tanto, los patrones de sustitución son importantes incluso cuando el investigador está interesado sólo en la cuota de mercado (*market share*) sin preocuparse de donde proviene la cuota.

El modelo logit implica un cierto patrón de sustitución entre alternativas. Si la sustitución se produce realmente de esta manera, dada la especificación que el investigador hace de la utilidad representativa, entonces el modelo logit es apropiado.

Sin embargo, para permitir patrones de sustitución más generales y para investigar qué patrón es más preciso, necesitamos modelos más flexibles. La cuestión se puede ver de dos maneras: como una restricción de los ratios de probabilidades de alternativas y/o como una restricción de las elasticidades cruzadas de las probabilidades. Presentamos cada una de estas maneras de caracterizar el problema a continuación.

Propiedad de independencia de alternativas irrelevantes

Para cualesquiera dos alternativas i y k, el ratio de las probabilidades logit es

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

$$= \frac{e^{V_{ni}}}{e^{V_{nk}}} = e^{V_{ni}-V_{nk}}$$

Este ratio no depende de las alternativas que no sean i y k. Es decir, las probabilidades relativas de elegir i sobre k son las mismas sin importar qué otras alternativas estén disponibles o cuáles sean los atributos de las otras alternativas. Dado que el ratio es independiente de alternativas distintas de i y k, se dice que es independiente de alternativas irrelevantes. El modelo logit exhibe esta independencia de alternativas irrelevantes (*independence from irrelevant alternatives*, IIA).

En muchas circunstancias, las probabilidades de elección que exhiben IIA proporcionan una representación precisa de la realidad. De hecho, Luce (1959) consideraba que la IIA es una propiedad de las probabilidades de elección que han sido especificadas apropiadamente. Él derivó el modelo logit directamente de la suposición de que las probabilidades de elección debían cumplir con la IIA, en lugar de (como hemos hecho nosotros) obtener la fórmula logit a partir de un supuesto acerca de la distribución de la utilidad no observada, para luego observar que la IIA es una propiedad resultante del modelo.

Mientras que la propiedad IIA es realista en algunas situaciones de elección, es claramente inadecuada en otras, como señaló por primera vez Chipman (1960) y Debreu (1960). Consideremos el famoso problema del autobús rojo – autobús azul. Un viajero tiene la opción de ir al trabajo en automóvil o tomar un autobús azul. Por simplicidad suponemos que la utilidad representativa de los dos medios de transporte es la misma, de tal manera que las probabilidades de elección son iguales: $P_c = P_{bb} = 1/2$, donde c representa al automóvil (*car*) y bb al autobús azul (*blue bus*). En este caso, el ratio de probabilidades es uno: $P_c/P_{bb} = 1$.

Ahora suponga que se introduce una nueva opción de transporte, un autobús rojo (*red bus*), y que el viajero considera que el autobús rojo es exactamente igual que el autobús azul. La probabilidad de que el viajero elija el autobús rojo es por lo tanto la misma que para el autobús azul, de manera que el ratio de sus probabilidades es uno: $P_{rb}/P_{bb} = 1$. Sin embargo, en el modelo logit el ratio P_c/P_{bb} es el mismo haya o no una nueva alternativa, en este caso, la alternativa del autobús rojo. Por tanto, este ratio se mantiene en uno. Las únicas probabilidades de elección que mantienen ambos ratios constantes, $P_c/P_{bb} = 1$ y $P_{rb}/P_{bb} = 1$, son $P_c = P_{bb} = P_{rb} = 1/3$, que son justamente las probabilidades que el modelo logit predice.

En el mundo real, sin embargo, esperaríamos que la probabilidad de que el viajero eligiese el automóvil se mantuviese igual cuando un nuevo autobús exactamente igual al ya existente estuviese disponible. También esperaríamos que la probabilidad inicial de elegir el autobús se dividiese entre los dos autobuses disponibles, una vez introducido el nuevo autobús. Es decir, podríamos esperar $P_c = 1/2$ y $P_{bb} = P_{rb} = 1/4$. En este caso, el modelo logit, debido a su propiedad IIA, sobreestima la probabilidad de elegir cualquiera de los autobuses y subestima la probabilidad de elegir el automóvil. El ratio de las probabilidades de elección del automóvil y el autobús azul, P_c/P_{bb}, realmente cambia con la introducción del autobús rojo, en lugar de permanecer constante como requiere el modelo logit.

Este ejemplo es bastante rígido y es poco probable que se produzca en la vida real. Sin embargo, el mismo tipo de predicción errónea surge con modelos logit siempre que el ratio de probabilidades para dos alternativas cambie con la introducción o cambio de otra alternativa. Por ejemplo, supongamos que se añade un nuevo medio de transporte que es similar, pero no exactamente igual, a los medios existentes, como un autobús expreso a lo largo de una línea que ya cuenta con servicio de autobús estándar. Podría esperarse que este nuevo medio de transporte redujese la probabilidad del autobús regular en una proporción mayor de lo que se reduciría la probabilidad del automóvil, por lo que el ratio de probabilidades para el automóvil y el autobús regular no permanecería constante. En esta situación, el modelo logit sobrestimaría la predicción de la demanda para los dos tipos de autobuses. Otros ejemplos han sido proporcionados por Ortuzar (1983) y Brownstone y Train (1999).

Sustitución proporcional

El mismo problema puede expresarse en términos de elasticidades cruzadas de las probabilidades logit. Vamos a considerar el cambio de un atributo de la alternativa j. Queremos conocer qué efecto tiene este cambio en las probabilidades de elección de todas las alternativas restante. En la sección 3.6 se obtiene la fórmula de la elasticidad de P_{ni} respecto a una variable que entra en la utilidad representativa de la alternativa j:

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

donde z_{nj} es el atributo de la alternativa j percibido por la persona n y β_z es su coeficiente (o si la variable entra en la utilidad representativa de forma no lineal, β_z sería la derivada de V_{nj} con respecto a z_{nj}) .

Esta elasticidad cruzada es la misma para todas las i: i no entra en la fórmula. Una mejora en los atributos de una alternativa reduce las probabilidades de elección de todas las otras alternativas en el mismo porcentaje. Si la probabilidad de elección de una alternativa se reduce un 10%, entonces todas las otras probabilidades de elección de las restantes alternativas también caen un 10% (excepto, claro está, la alternativa cuyo atributo ha cambiado; su probabilidad se eleva debido a la mejora). Una forma concisa de expresar este fenómeno es indicar que una mejora en una alternativa se extrae proporcionalmente de todas las otras alternativas restantes. Del mismo modo, para una disminución de la utilidad representativa de una alternativa, las probabilidades de todas las restante alternativas se incrementan en el mismo porcentaje.

Este patrón de sustitución, que puede ser llamado *desplazamiento proporcional (proportionate shifting)*, es una manifestación de la propiedad IIA. El ratio de probabilidades de elección de las alternativas i y k se mantiene constante cuando un atributo de la alternativa j cambia sólo si las dos probabilidades varían en la misma proporción. Si denotamos con el superíndice 0 las probabilidades antes del cambio y con 1 después del cambio, la propiedad IIA exige que

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

cuando un atributo de la alternativa j cambia. Esta igualdad sólo puede mantenerse si cada probabilidad cambia en la misma proporción: $P_{ni}^1 = \lambda P_{ni}^0$ y $P_{nk}^1 = \lambda P_{nk}^0$, donde en ambos casos λ es la misma.

La sustitución proporcional puede ser realista para algunas situaciones, en cuyo caso el modelo logit es apropiado. Sin embargo, en muchos casos podemos esperar encontrar otros patrones de sustitución por lo que imponer la sustitución proporcional a través de un modelo logit puede conducir a previsiones poco realistas. Considere una situación que es importante para la *California Energy Comission* (CEC), entidad que tiene la responsabilidad de investigar las políticas reguladoras para promover vehículos de energía eficiente en California, así como reducir la dependencia del estado respecto a la gasolina para los automóviles. Supongamos en aras de la simplicidad que hay tres clases de vehículos: automóviles grandes de gasolina, automóviles pequeños de gasolina y automóviles pequeños eléctricos. Supongamos también que en las condiciones actuales las probabilidades de que un hogar elija cada uno de estos vehículos son 0.66, 0.33 y 0.01 respectivamente. La CEC está interesada en conocer el impacto de subvencionar los automóviles eléctricos. Supongamos que el subsidio es suficiente para elevar la probabilidad de que el automóvil eléctrico sea elegido de 0.01 a 0.10. A través del modelo logit, la probabilidad de elección de cada uno de los automóviles de gasolina se prevé que caiga en el mismo porcentaje. La probabilidad de elección de un automóvil grande de gasolina se reduciría en un diez por ciento, de 0.66 a 0.60, y lo mismo sucedería para el automóvil pequeño de gasolina, reducción del diez por ciento, de 0.33 hasta 0.30. En términos de números absolutos, la probabilidad incrementada del automóvil pequeño eléctrico (0.09) se prevé, según el modelo logit, que provenga dos veces más de los automóviles grandes de gasolina (0.06) que de los automóviles pequeños de gasolina (0.03).

Este patrón de sustitución es claramente poco realista. Dado que el automóvil eléctrico es pequeño, es de esperar que subsidiarlo atraiga más a usuarios de automóviles pequeños de gasolina que a usuarios de automóviles grandes de gasolina. En términos de elasticidades cruzadas, esperaríamos que la elasticidad cruzada de los automóviles pequeños de gasolina respecto a una mejora en los automóviles pequeños eléctricos vaya a ser más alta que la de los automóviles grandes de gasolina. Esta diferencia es importante en el análisis de políticas de la CEC. El modelo logit sobrestimará la predicción de ahorro de gasolina que resultará de la subvención, dado que sobrestimará la sustitución de los automóviles grandes de gasolina (los que más gastan) y subestimará la sustitución de los automóviles pequeños de gasolina. Desde una perspectiva reguladora, esta predicción errónea puede ser crítica, causando que un programa de subsidios parezca más beneficioso de lo que realmente es. Por esta razón la CEC utiliza modelos que son más generales que logit para representar la sustitución entre vehículos. Los modelos logit jerárquico, probit y logit mixto explicados en los capítulos 4-6 ofrecen opciones viables para el investigador.

Ventajas de la IIA

Como acabamos de mencionar, la propiedad de IIA del modelo logit puede ser poco realista en muchos entornos. Sin embargo, cuando la IIA refleja la realidad (o una aproximación adecuada de la realidad), se obtienen considerables ventajas en su uso. En primer lugar, debido a la IIA, es posible estimar los parámetros del modelo de forma consistente en un subconjunto de alternativas para cada decisor estudiado. Por ejemplo, en una situación con 100 alternativas, el investigador podría, con el fin de reducir el tiempo de cálculo, estimar sobre un subconjunto de 10 alternativas para cada persona de la muestra, con la elección de la persona incluida así como 9 alternativas adicionales seleccionadas aleatoriamente entre las restantes 99 . Dado que las probabilidades relativas dentro de un subconjunto de alternativas no se ven afectadas por los atributos o por la existencia de alternativas fuera del subconjunto, la exclusión de alternativas en la estimación no afecta a la consistencia del estimador. Los detalles de este tipo de estimación se ofrecen en la sección 3.7.1. Este hecho tiene gran importancia práctica. En el análisis de situaciones de elección para las que el número de alternativas es grande, la estimación en un subconjunto de alternativas puede ahorrar cantidades sustanciales de tiempo de computación. En casos extremos, el número de alternativas podría ser tan grande como para impedir por completo la estimación si no fuese posible utilizar un subconjunto de las alternativas.

Otro uso práctico de la propiedad IIA surge cuando el investigador sólo está interesado en examinar elecciones entre un subconjunto de alternativas y no entre todas las alternativas. Por ejemplo, considere un investigador que está interesado en la comprensión de los factores que afectan a la elección que realizan los trabajadores entre el automóvil y el autobús como medios de transporte para ir al trabajo. El conjunto completo de alternativas incluye medios como caminar, andar en bicicleta, motocicletas, patines, etc. Si el investigador pensó que la propiedad IIA se cumplía adecuadamente en este caso, podría estimar un modelo de elección con sólo automóvil y autobús como alternativas y excluir de la muestra analizada a los trabajadores que utilizaron otros medios de transporte. Esta estrategia le ahorraría al investigador un tiempo y costo considerables en el desarrollo de los datos de los otros medios, sin limitar su capacidad de examinar los factores relacionados con el automóvil y el autobús.

Pruebas de IIA

Si la IIA se cumple en un entorno particular es una cuestión empírica, susceptible de investigación estadística. Las primeras pruebas para verificar la IIA fueron desarrolladas por McFadden et al . (1978). Se sugieren dos tipos de pruebas. En primer lugar, el modelo se puede re-estimar en un subconjunto de alternativas. Según la IIA, el ratio de probabilidades para cualesquiera dos alternativas debería ser el mismo estén o no estén presentes otras alternativas. Como resultado, si la IIA se observa realmente, las estimaciones obtenidas de los parámetros en el subconjunto de alternativas no deberían ser significativamente diferentes a las obtenidas usando el total de las alternativas. Un test de la hipótesis de que los parámetros en el subgrupo son los mismos que los parámetros para el conjunto completo de alternativas constituye una prueba de la IIA. Hausman y McFadden (1984) proporcionan un estadístico adecuado para este tipo de prueba. En segundo lugar, el modelo se puede re-estimar con nuevas variables intercambiadas entre alternativas, es decir, con las variables de una alternativa formando parte de la utilidad de otra alternativa. Si el ratio de probabilidades de las alternativas i y k en realidad depende de los atributos y de la existencia de una tercera alternativa j (en violación de la IIA), los atributos de la alternativa j entrarán significativamente en la utilidad de las alternativas i o k dentro de una especificación logit. Por lo tanto, una prueba de si las variables cruzadas entre alternativas entran en el modelo constituye una prueba de la IIA. McFadden (1987) desarrolló un procedimiento para realizar este tipo de prueba con regresiones: para ello usaba los residuos del modelo logit original como variable dependiente y las variables cruzadas entre alternativas, apropiadamente especificadas, como variables explicativas. Train et al. (1989) muestran cómo este procedimiento se puede realizar convenientemente dentro del modelo logit mismo.

La llegada de modelos que no presentan la IIA y especialmente el desarrollo de software para la estimación de estos modelos, hace que las pruebas de IIA sean más fáciles que antes. Para especificaciones más flexibles, tales como GEV y logit mixto, el modelo logit simple con IIA es un caso especial que surge bajo ciertas restricciones sobre los parámetros del modelo más flexible. En estos casos, la IIA puede ser probada mediante pruebas de estas limitaciones. Por ejemplo, un modelo logit mixto se convierte en un logit simple si la distribución de mezcla tiene varianza cero. La IIA se puede probar mediante la estimación de un logit mixto, probando posteriormente si la varianza de la distribución mixta es cero en la práctica.

Una prueba de la IIA como una restricción a un modelo más general necesariamente opera bajo la suposición de que el modelo más general es en sí mismo una especificación apropiada. Las pruebas sobre subconjuntos de alternativas (Hausman y McFadden, 1984) y sobre variables cruzadas entre alternativas (McFadden, 1987; Train et al, 1989), aunque son más difíciles de realizar, operan bajo hipótesis menos restrictivas. El contrapunto a esta ventaja, por supuesto, es que cuando la IIA falla, estas pruebas no proporcionan tanta orientación sobre la especificación correcta a usar en lugar del modelo logit.

3.3.3 Datos de panel

En muchas ocasiones el investigador puede observar numerosas elecciones realizadas por cada decisor. Por ejemplo, en los estudios de ocupación, se observa si las personas en la muestra trabajan o no trabajan en cada mes durante varios años. Un investigador interesado en las dinámicas de elección de compra de un automóvil podría obtener datos sobre las compras actuales y pasadas de vehículos en una muestra de hogares. En las encuestas usadas en investigación de mercados, a los encuestados a menudo se les pide responder una serie de preguntas de elección hipotética, llamadas experimentos de "preferencia declarada". Para cada uno de estos experimentos, se describe al encuestado un conjunto de productos alternativos con diferentes atributos y se le pide que indique cuál es el producto que elegiría. Al encuestado se le administra una batería de este tipo de preguntas, variando cada vez los atributos de los productos con el fin de determinar cómo la elección del entrevistado cambia cuando cambian los atributos. Por lo tanto, el investigador puede observar la secuencia de opciones elegidas por cada encuestado. Los datos que representan repeticiones de elecciones como éstas se llaman datos de panel.

Si los factores no observados que afectan a los decisores son independientes entre las elecciones repetidas, un modelo logit puede utilizarse para examinar los datos de panel de la misma forma como se usaría para examinar datos obtenidos de una sola vez. Cualesquiera dinámicas relacionadas con factores observados que entren en el proceso de decisión, como la dependencia del estado del decisor (por la cual elecciones pasadas de la persona influyen en sus elecciones presentes) o la respuesta diferida a cambios en los atributos, se puede acomodar al modelo. Sin embargo, dinámicas asociadas a factores no observados no se pueden manejar, dado que se supone que los factores no observados no guardan relación entre elecciones.

La utilidad que el decisor n obtiene de la alternativa j en el período o situación de elección t es

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

Si ε_{njt} se distribuye con densidad valor extremo, independiente respecto a n, j y sobre todo, t, entonces, usando la misma prueba descrita en (3.6), las probabilidades de elección son

$$P_{nit} = \frac{e^{V_{nit}}}{\sum_j e^{V_{njt}}}. \tag{3.9}$$

Cada situación de elección de cada decisor se convierte en una observación independiente. Si se especifica que la utilidad representativa de cada período dependa sólo de las variables correspondientes a ese período, por ejemplo, $V_{njt} = \beta' x_{njt}$, donde x_{njt} es un vector de variables que describen la alternativa j tal y como se presenta al decisor n en el período t, entonces no hay ninguna diferencia esencial entre el modelo logit con datos de panel y con datos puntuales.

Podemos capturar aspectos dinámicos del comportamiento especificando que la utilidad representativa en cada período dependa de variables observadas en otros períodos. Por ejemplo, una respuesta diferida al precio puede representarse mediante la introducción del precio en el período $t-1$ como variable explicativa en la utilidad del período t. La introducción de los precios en períodos futuros puede realizarse, tal y como hace Adamowicz (1994), para capturar la anticipación de los consumidores a futuros cambios de precios. Bajo los supuestos del modelo logit, la variable dependiente en períodos anteriores también se puede introducir como variable explicativa. Supongamos por ejemplo que existe una inercia o formación de hábitos en las elecciones de las personas de tal manera que tienden a quedarse con la alternativa que hayan elegido previamente a menos que otra alternativa proporcione una utilidad suficientemente mayor para justificar un cambio. Este comportamiento se puede capturar como $V_{njt} = \alpha y_{nj(t-1)} + \beta x_{njt}$, donde $y_{nj(t-1)} = 1$ si n eligió j en el período t y 0 en caso contrario. Con $\alpha > 0$, la utilidad de la alternativa j en el período actual es mayor si ya se consumía la alternativa j en el período anterior. La misma especificación también puede capturar una especie de búsqueda de la variedad. Si α es negativo, el consumidor obtiene mayor utilidad de no elegir la misma alternativa que eligió en el último período. Son posibles numerosas variaciones sobre estos conceptos. Adamowicz (1994) introduce el número de veces que la alternativa ha sido elegida previamente en lugar de usar un simple indicador para la elección inmediatamente anterior. Erdem (1996) introduce los atributos de las alternativas elegidas previamente, con la utilidad de cada alternativa en el período actual dependiendo de la similitud de sus atributos con los atributos previamente experimentados.

La inclusión de la variable dependiente diferida no induce inconsistencia en la estimación, ya que para un modelo logit se supone que los errores deben ser independientes en el tiempo. La variable dependiente diferida $y_{nj(t-1)}$ no está correlacionada con el error actual ε_{njt} debido a esta independencia. La situación es análoga a los modelos de regresión lineal, en los que una variable dependiente diferida puede añadirse sin inducir sesgo, siempre y cuando los errores sean independientes a lo largo del tiempo.

Por supuesto, la asunción de errores independientes en el tiempo es severa. Por lo general, es de esperar que existan algunos factores que no están siendo observados por el investigador que afecten a cada una de las elecciones de los decisores. En particular, si hay dinámicas en los factores observados, el investigador también podría esperar que haya dinámicas en los factores no observados. En estas situaciones, el investigador puede utilizar un modelo como el probit o el logit mixto que permiten que los factores no observados estén correlacionados en el tiempo, o re-especificar la utilidad representativa para incorporar de forma explícita las fuentes de las dinámicas no observadas en el modelo, de tal manera que los errores restantes sean independientes en el tiempo.

3.4 Utilidad representativa no lineal

En algunos contextos, el investigador encontrará útil permitir que los parámetros entren en la utilidad representativa de forma no lineal. La estimación pasa a ser más difícil, ya que la función log-verosimilitud (*log-likelihood*) puede que no sea globalmente cóncava y los procedimientos informáticos capaces de tratar esta situación no están tan ampliamente disponibles como para modelos logit con utilidad lineal respecto a los parámetros. Sin embargo, los aspectos del comportamiento que el investigador está estudiando pueden incluir parámetros que sólo son interpretables cuando entran en la utilidad de forma no lineal. En estos casos, el esfuerzo necesario para desarrollar un código propio puede estar justificado. Dos ejemplos ilustran este punto.

Ejemplo 1: El balance entre bienes y ocio

Considere la posibilidad de elección de los trabajadores respecto al medio de transporte a su lugar de trabajo (automóvil o autobús). Supongamos que los trabajadores también eligen el número de horas que quieren trabajar basándose en el equilibrio estándar entre bienes y ocio. Train y McFadden (1978) desarrollaron un procedimiento para examinar estas decisiones interrelacionadas. Como veremos a continuación, los parámetros de la función de utilidad de los trabajadores respecto a los bienes y al ocio entran de forma no lineal en la utilidad de los medios de transporte.

Supongamos que las preferencias de los trabajadores con respecto a los bienes G (*goods*) y al ocio L (*leisure*) están representados por una función de utilidad Cobb-Douglas de la forma

$$U = (1-\beta)lnG + \beta lnL.$$

El parámetro β refleja la preferencia relativa de los trabajadores entre obtener bienes y disponer tiempo de ocio, donde una β mayor implica una mayor preferencia por el ocio en relación con los bienes. Cada trabajador tiene una cantidad fija de tiempo (24 horas al día) y se enfrenta a un salario fijo por hora trabajada, w. En el modelo estándar de bienes-ocio, el trabajador elige el número de horas a trabajar que maximiza U, sujeto a las restricciones de que (1) el número de horas trabajadas más el número de horas de ocio es igual al número de horas disponibles y (2) el valor de los bienes consumidos es igual al salario por hora w por el número de horas trabajadas.

Cuando se añade el medio de transporte al modelo, las restricciones sobre el tiempo y el dinero cambian. Cada medio de transporte resta una cierta cantidad de tiempo y de dinero. Condicionado a la elección del automóvil como medio de transporte, el trabajador maximiza su utilidad U sujeta a la restricción de que (1) el número de horas trabajadas más el número de horas de ocio es igual al número de horas disponibles después de *restar el tiempo pasado en el automóvil para ir al trabajo* y (2) el valor de los bienes consumidos es igual al salario por hora w por el número de horas trabajadas *menos el costo de conducir hasta el trabajo*. La utilidad asociada con la elección de viajar en automóvil es el mayor valor posible de U que se puede alcanzar bajo estas restricciones. Del mismo modo, la utilidad de viajar en autobús al trabajo es el valor máximo de U que se puede obtener teniendo en cuenta el tiempo y el dinero que quedan después de restar el tiempo y el dinero consumidos en el viaje en autobús. Train y McFadden obtuvieron los valores de maximización de U condicionados a cada medio de transporte. Para la U dada anteriormente, estos valores son

$$U_j = -\alpha\left(c_j/w^{\beta} + w^{1-\beta}t_j\right)\, para\ j = automóvil\ y\ autobús.$$

El costo del viaje se divide por w^{β} y el tiempo de viaje se multiplica por $w^{1-\beta}$. El parámetro β, que denota la preferencia relativa de los trabajadores por los bienes o por el ocio, entra en la expresión de la utilidad de la elección de cada medio de transporte de forma no lineal. Dado que este parámetro tiene sentido, el investigador desearía estimarlo dentro de esta utilidad no lineal en lugar de utilizar una aproximación mediante un modelo lineal en parámetros.

Ejemplo 2: Agregación geográfica

Se han desarrollado modelos ampliamente utilizados para la elección que los viajeros hacen de sus destinos para diferentes tipos de viajes, tales como viajes dentro de un área metropolitana para ir de compras. Habitualmente, el área metropolitana se divide en zonas y los modelos dan la probabilidad de que una persona elija viajar a una zona en particular. La utilidad representativa para cada zona depende del tiempo y el costo de los viajes a la zona, además de otro tipo de variables, tales como cantidad de población residente y número de personas empleadas en comercios, que reflejan motivos por los que la gente puede desear visitar la zona. Estas últimas variables se llaman variables de *atracción*; las etiquetaremos con el vector a_j para la zona j. Puesto que son estas variables de atracción las que dan lugar a los parámetros que entran en la utilidad de forma no lineal, asumimos por simplicidad que la utilidad representativa depende únicamente de estas variables.

La dificultad en la especificación de la utilidad representativa proviene del hecho de que la decisión del investigador en relación a qué tamaño asignar a cada zona es bastante arbitraria. Sería útil tener un modelo que no fuese sensible al nivel de agregación usado en la definición de las zonas. Si dos zonas se combinan, sería bueno para el modelo dar una probabilidad de viajar a la zona combinada que sea la misma que la suma de las probabilidades de viajar a las dos zonas originales. Esta consideración impone restricciones a la forma de la utilidad representativa.

Considere las zonas j y k que, cuando se combinan, se etiquetan como zona c .La población y el empleo en la zona combinada son necesariamente las sumas de la población y el empleo de las dos zonas originales: $a_j + a_k = a_c$. Con el fin de que los modelos den la misma probabilidad de elección de estas zonas antes y después de su fusión, el modelo debe satisfacer

$$P_{nj} + P_{nk} = P_{nc},$$

que para los modelos logit toma la forma

$$\frac{e^{V_{nj}} + e^{V_{nk}}}{e^{V_{nj}} + e^{V_{nk}} + \sum_{l \neq j,k} e^{V_{nl}}} = \frac{e^{V_{nc}}}{e^{V_{nc}} + \sum_{l \neq j,k} e^{V_{nl}}}$$

Esta igualdad se cumple sólo cuando $exp(V_{nj}) + exp(V_{nk}) = exp(V_{nc})$. Si la utilidad representativa se especifica como $V_{nl} = \ln(\beta' a_l)$ para todas las zonas l, entonces la igualdad se cumple: $exp(\ln(\beta' a_j)) + exp(\ln(\beta' a_k)) = \beta' a_j + \beta' a_k = \beta' a_c = exp(\ln(\beta' a_c))$. Por lo tanto, para especificar un modelo de elección de destino que no sea sensible al nivel de agregación zonal, la utilidad representativa tiene que ser especificada con los parámetros dentro de una operación logarítmica.

3.5 Excedente del consumidor

Para el análisis de políticas reguladoras, el investigador a menudo está interesado en medir el cambio en el excedente que obtiene el consumidor (*consumer surplus*) por efecto de una política en particular. Por ejemplo, si se está considerando una nueva alternativa como la construcción de un sistema de tren ligero en la ciudad, es importante medir los beneficios del proyecto para ver si se justifican los costos. Del mismo modo, un cambio en los atributos de una alternativa puede tener un impacto en el excedente que recibe el consumidor cuya evaluación es importante. La degradación de la calidad del agua de los ríos daña a los pescadores que ya no pueden pescar de forma efectiva en los sitios dañados. La medición de este daño en términos monetarios es un elemento central de la acción legal contra el contaminador. A menudo, es importante evaluar los efectos distributivos de una política, por ejemplo, cómo la carga de un impuesto es soportado por diferentes grupos de la población.

Bajo los supuestos logit, el excedente percibido por el consumidor asociado a un conjunto de alternativas toma una forma cerrada que es fácil de calcular. Por definición, el excedente del consumidor de una persona concreta es la utilidad, expresada en dólares, que la persona recibe en la situación de elección. El decisor escoge la alternativa que ofrece la mayor utilidad. El excedente del consumidor es por lo tanto $CS_n = (1/\alpha_n) max_j(U_{nj})$, donde α_n es la utilidad marginal de los ingresos: $dU_n/dY_n = \alpha_n$, siendo Y_n los ingresos de la persona n. La división por α_n traduce utilidad a dólares, ya que $1/\alpha_n = dY_n/dU_n$. El investigador no observa U_{nj} y por lo tanto no se puede utilizar esta expresión para calcular el excedente del consumidor del decisor. En lugar de ello, el investigador observa V_{nj} y conoce la distribución de la porción restante de la utilidad. Con esta información, el investigador es capaz de calcular el excedente del consumidor esperado:

$$E(CS_n) = \frac{1}{\alpha_n} E\left[max_j\left(V_{nj} + \varepsilon_{nj}\right)\right],$$

donde la esperanza se calcula sobre todos los valores posibles de ε_{nj}. Williams (1977) y Small y Rosen (1981) muestran que si cada ε_{nj} se disitribuye como valor extremo iid y la utilidad es lineal respecto al ingreso (de modo que α_n es constante respecto a ingresos), entonces esta expectativa se convierte en

(3.10)
$$E(CS_n) = \frac{1}{\alpha_n} ln\left(\sum_{j=1}^{J} e^{V_{nj}}\right) + C,$$

donde C es una constante desconocida que representa el hecho de que el nivel absoluto de utilidad no puede ser medido. Como veremos a continuación, esta constante es irrelevante para definir una política y puede ser ignorada.

Observe que el argumento entre paréntesis de esta expresión es el denominador de la probabilidad de elección logit (3.6). Excepto por la división y la adición de constantes, el excedente del consumidor esperado en un modelo logit es simplemente el logaritmo del denominador de la probabilidad de elección. A menudo esta expresión es llamada el término log-suma (*log-sum*). Esta semejanza entre las dos fórmulas no tiene significado económico, en el sentido de que no hay nada en el denominador de una probabilidad de elección que se relacione necesariamente con el excedente del consumidor. Es simplemente el resultado de la fórmula matemática de la distribución de valor extremo. Sin embargo, la relación hace el cálculo del excedente del consumidor esperado extremadamente fácil, lo que representa otra de las muchas comodidades de usar el modelo logit.

Bajo la interpretación estándar de la distribución de los errores, tal y como se describe en el último párrafo del punto 2.3, $E(CS_n)$ es el promedio del excedente del consumidor en la sub-población de personas que tienen las mismas utilidades representativas que la persona n. El excedente total del consumidor en la población se calcula como la suma ponderada de $E(CS_n)$ sobre una muestra de decisores, con las ponderaciones reflejando el número de personas en la población que afrontan la misma utilidad representativa que la persona de la muestra.

El cambio en el excedente del consumidor que resulta de un cambio en las alternativas y/o en el conjunto de elección se calcula a partir de (3.10). En particular, $E(CS_n)$ se calcula dos veces: primero en las condiciones previas al cambio y de nuevo en las condiciones posteriores al cambio. La diferencia entre los dos resultados es el cambio en el excedente del consumidor:

$$\Delta E(CS_n) = \frac{1}{\alpha_n}\left[ln\left(\sum_{j=1}^{J^1} e^{V_{nj}^1}\right) - ln\left(\sum_{j=1}^{J^0} e^{V_{nj}^0}\right)\right],$$

donde los superíndices 0 y 1 se refieren a antes y después del cambio. El número de alternativas puede cambiar (por ejemplo, se puede añadir una nueva alternativa) así como los atributos de las alternativas. Dado que la constante desconocida C entra en el excedente del consumidor esperado tanto antes como después del cambio, desaparece de la diferencia y por lo tanto puede ser ignorada cuando se calculan los cambios en el excedente del consumidor.

Para calcular el cambio en el excedente del consumidor, el investigador debe conocer o haber estimado la utilidad marginal del ingreso, α_n. Por lo general, un precio o un costo variable forma parte de la utilidad representativa, en cuyo caso el negativo de su coeficiente es α_n por definición. (Un coeficiente asociado a un precio o a un costo es negativo; el negativo de un coeficiente negativo da una α_n positiva). Por ejemplo, en la elección entre el automóvil y el autobús, la utilidad es $U_{nj} = \beta_1 t_{nj} + \beta_2 c_{nj}$, donde t es el tiempo, c es el costo y tanto β_1 como β_2 son negativos, lo que indica que la utilidad disminuye a medida que el tiempo o el costo de un viaje aumenta. El negativo del coeficiente de costo $-\beta_2$ es la cantidad en que la utilidad se incrementa debido a una disminución de un dólar en los costos. Una reducción de un dólar en costos es equivalente a un incremento de un dólar en ingresos, ya que la persona puede gastar el dólar que ahorra en costos de viaje de la misma manera que si hubiese obtenido un dólar adicional de ingresos. Por consiguiente, la cantidad $-\beta_2$ es el aumento en la utilidad generada por un incremento de un dólar en los ingresos: la utilidad marginal del ingreso. En este caso, la cantidad es la misma para todo n. Si c_{nj} ha entrado en la utilidad representativa interactuando con características de la persona que no sean el ingreso, como en el producto $c_{nj}H_n$ donde H_n es el tamaño del hogar, entonces la utilidad marginal del ingreso sería $-\beta_2 H_n$, cantidad que varía para diferentes n.

A lo largo de esta explicación hemos asumido que α_n está fijada para cada persona con independencia de sus ingresos. La fórmula (3.10) para la esperanza del excedente del consumidor depende de manera crítica del supuesto de que la utilidad marginal de los ingresos es independiente de los ingresos. Si la utilidad marginal de los ingresos cambia con los ingresos, necesitamos una fórmula más complicada, dado que α_n mismo se convierte en una función de los cambios en los atributos. McFadden (1999) y Karlstrom (2000) proporcionan procedimientos para el cálculo de los cambios en el excedente del consumidor en estas condiciones.

Las condiciones para el uso de la expresión (3.10) en realidad son menos estrictas de lo que hemos indicado. Dado que sólo los cambios en el excedente del consumidor son relevantes para el análisis de políticas reguladoras, la fórmula (3.10) se puede utilizar si la utilidad marginal del ingreso es constante en el rango de cambios en el ingreso que implícitamente se están considerando en la política reguladora. Por lo tanto, para cambios en políticas reguladoras que cambian el excedente del consumidor en pequeñas cantidades por persona en relación con los ingresos, la fórmula se puede utilizar a pesar de que la utilidad marginal del ingreso en realidad varíe con el ingreso.

La suposición de que α_n no depende de los ingresos tiene implicaciones en la especificación de la utilidad representativa. Como ya se ha mencionado, α_n por lo general se toma como el valor absoluto del coeficiente de precio o de costo. Por lo tanto, si el investigador tiene previsto utilizar su modelo para estimar los cambios en el excedente del consumidor y quiere aplicar la fórmula (3.10), no se puede especificar que este coeficiente dependa de los ingresos. En el ejemplo de la elección del medio de transporte, el costo puede aparecer multiplicado por el tamaño del hogar, de manera que el coeficiente de costo y, por lo tanto, la utilidad marginal del ingreso, varíe entre hogares de diferente tamaño. Sin embargo, si el costo aparece dividido por el ingreso del hogar, el coeficiente de costo pasa a depender de los ingresos, algo que viola el supuesto necesario para la expresión (3.10). Esta violación puede no ser importante para los pequeños cambios en el excedente del consumidor, pero sin duda pasa a ser importante para los grandes cambios.

3.6 Derivadas y elasticidades

Dado que las probabilidades de elección son una función de las variables observadas, a menudo es útil conocer en qué medida cambian dichas probabilidades en respuesta a un cambio en algún factor observado. Por ejemplo, en el caso de un hogar eligiendo marca y modelo de automóvil a comprar, una pregunta natural es: ¿en qué medida aumenta la probabilidad de escoger un automóvil dado si se mejora su eficiencia en el uso de combustible? Desde el punto de vista de los demás fabricantes de automóviles, una cuestión relacionada es: ¿hasta qué punto descenderá la probabilidad de que los hogares elijan, por ejemplo, un Toyota, si mejora la eficiencia de uso de combustible de un Honda?

Para abordar estas cuestiones, calculamos las derivadas de las probabilidades de elección. El cambio en la probabilidad de que el decisor n elija la alternativa i dado un cambio en un factor observado z_{ni} que forma parte de la utilidad representativa de esa alternativa (manteniendo constante la utilidad representativa de las otras alternativas) es

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

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

$$= \frac{\partial V_{ni}}{\partial z_{ni}}\left(P_{ni} - P_{ni}{}^2\right)$$

$$= \frac{\partial V_{ni}}{\partial z_{ni}} P_{ni}(1 - P_{ni})$$

Si la utilidad representativa es lineal en z_{ni} con coeficiente β_z, la derivada se convierte en $\beta_z P_{ni}(1 - P_{ni})$. Esta derivada es máxima cuando $P_{ni} = 1 - P_{ni}$, algo que sucede cuando $P_{ni} = 0.5$. Inversamente, disminuye a medida P_{ni} se aproxima a cero o a uno. La curva de probabilidad sigmoidea de la figura 3.1 es consistente con estos hechos. Dicho de forma intuitiva, el efecto de un cambio en una variable observada es mayor cuando las probabilidades de elección indican un alto grado de incertidumbre en cuanto a la elección. A medida que la elección se hace más cierta (es decir, las probabilidades se acercan a cero o a uno), el efecto de un cambio en una variable observada disminuye.

También se puede determinar en qué medida cambia la probabilidad de elegir una alternativa particular cuando cambia una variable observada relacionada con *otra* alternativa. Supongamos que z_{nj} se refiere a un atributo de la alternativa j ¿Cómo cambia la probabilidad de elegir la alternativa i cuando z_{nj} aumenta? tenemos

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

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

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

Cuando V_{nj} es lineal en z_{nj} con coeficiente β_z, entonces esta derivada cruzada se convierte en $-\beta_z P_{ni} P_{nj}$. Si z_{nj} es un atributo deseable, de manera que β_z es positivo, incrementar z_{nj} disminuye la probabilidad de elegir cada alternativa que no sea j. Además, la disminución de la probabilidad es proporcional al valor de la probabilidad antes de que z_{nj} cambiase.

Un aspecto lógicamente necesario de las derivadas de las probabilidades de elección es que, cuando una variable observada cambia, los cambios en las probabilidades de elección sumen cero. Esto es una consecuencia del hecho de que las probabilidades deben sumar a uno antes y después del cambio; la demostración de este hecho para modelos logit es la siguiente:

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

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

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

$$= 0.$$

En términos prácticos, si una alternativa se mejora de manera que su probabilidad de ser elegida aumenta, la probabilidad adicional necesariamente debe extraerse de otras alternativas. Para aumentar la probabilidad de una alternativa es necesario disminuir la probabilidad de otra alternativa. Aunque es obvio, este hecho es a menudo olvidado por planificadores que quieren mejorar la demanda de una alternativa sin reducir la demanda de otras alternativas.

Los economistas suelen medir la respuesta a los cambios mediante elasticidades en lugar de derivadas, dado que las elasticidades están normalizadas por las unidades de las variables. Una elasticidad es el cambio porcentual en una variable asociado a un cambio del uno por ciento en otra variable. La elasticidad de P_{ni} respecto a z_{ni}, una variable que entra en la utilidad de la alternativa i, es

$$E_{iz_{ni}} = \frac{\partial P_{ni}}{\partial z_{ni}} \frac{z_{ni}}{P_{ni}}$$

$$= \frac{\partial V_{ni}}{\partial z_{ni}} P_{ni}(1 - P_{ni}) \frac{z_{ni}}{P_{ni}}$$

$$= \frac{\partial V_{ni}}{\partial z_{ni}} z_{ni}(1 - P_{ni}).$$

Si la utilidad representativa es lineal en z_{ni} con coeficiente β_z, entonces $E_{iz_{ni}} = \beta_z z_{ni}(1 - P_{ni})$.

La elasticidad cruzada de P_{ni} respecto a una variable que entra en la especificación de la alternativa j es

$$E_{iz_{nj}} = \frac{\partial P_{ni}}{\partial z_{nj}} \frac{z_{nj}}{P_{ni}}$$

$$= -\frac{\partial V_{nj}}{\partial z_{nj}} z_{nj} P_{nj}.$$

que en el caso de que la utilidad sea lineal se reduce a $E_{iz_{nj}} = -\beta_z z_{nj} P_{nj}$. Tal y como se ha analizado en la sección 3.3.2, esta elasticidad cruzada es igual para todas las alternativas i: un cambio en un atributo de la alternativa j cambia las probabilidades de todas las otras alternativas en el mismo porcentaje. Esta propiedad de las elasticidades cruzadas de logit es una manifestación, o re-expresión, de la propiedad de IIA de las probabilidades de elección logit.

3.7 Estimación

Manski y McFadden (1981) y Cosslett (1981) describen métodos de estimación para varios procedimientos de muestreo. Nosotros trataremos en esta sección la estimación para el esquema de muestreo más habitual. En primer lugar, describimos la estimación cuando la muestra es exógena y todas las alternativas se utilizan en la estimación. Luego tratamos la estimación en un subconjunto de alternativas y con ciertos tipos de muestras basadas en la propia elección (es decir, no exógenas).

3.7.1 Muestra exógena

Consideremos en primer lugar la situación en que la muestra se ha seleccionado exógenamente, es decir, la muestra se ha seleccionado aleatoriamente o aleatoriamente por estratos, con los estratos definidos sobre factores exógenos a la elección que se desea analizar. Si el procedimiento de muestreo se relaciona con la elección que se desea analizar (por ejemplo, si se está examinando la elección del medio de transporte y la muestra se obtiene mediante la selección de gente en los autobuses y se une a una selección de personas reclutadas en peajes) entonces necesitaremos procedimientos de estimación más complejos en general, como se trata en la próxima sección. También asumimos que las variables explicativas son exógenas a la situación de elección. Es decir, las variables que entran en la utilidad representativa son independientes del componente no observado de utilidad.

Con el propósito de hacer la estimación, obtenemos una muestra de N decisores. Dado que las probabilidades de elección logit tienen una expresión cerrada, podemos aplicar los procedimientos habituales de máxima verosimilitud (*maximum-likelihood*). La probabilidad de que la persona n elija la alternativa que realmente hemos observado que eligió se puede expresar como

$$\prod_i (P_{ni})^{y_{ni}}$$

donde $y_{ni} = 1$ si la persona n eligió i y cero en caso contrario. Observe que dado que $y_{ni} = 0$ para todas las alternativas no elegidas y P_{ni} elevado a la potencia cero es 1, este término es simplemente la probabilidad de la alternativa elegida.

Asumiendo que la elección de cada decisor es independiente de las elecciones del resto de decisores, la probabilidad de que cada persona de la muestra haya elegido la alternativa que realmente hemos observado que eligió es

$$L(\beta) = \prod_{n=1}^{N} \prod_i (P_{ni})^{y_{ni}},$$

donde β es un vector que contiene los parámetros del modelo. La función logaritmo de verosimilitud (*log-likelihood*) es por lo tanto

$$LL(\beta) = \sum_{n=1}^{N} \sum_i y_{ni} \ln P_{ni} \quad (3.11)$$

y el estimador es el valor de β que maximiza esta función. McFadden (1974) muestra que $LL(\beta)$ es globalmente cóncava para una especificación de la utilidad lineal en parámetros y numerosos paquetes de software estadístico disponibles en el mercado permiten estimar estos modelos. Cuando los parámetros entran en la utilidad representativa de forma no lineal, el investigador puede tener que escribir su propio código para hacer la estimación utilizando los procedimientos descritos en el Capítulo 8.

En este caso, la estimación de la máxima verosimilitud puede ser rescrita y reinterpretada de una manera que ayuda a comprender la naturaleza de las estimaciones. En el máximo de la función de verosimilitud, su derivada con respecto a cada uno de los parámetros es cero:

$$\text{(3.12)} \qquad \frac{LL(\beta)}{d\beta} = 0$$

La estimación de máxima verosimilitud son por lo tanto los valores de β que satisfacen esta condición de primer orden. Por conveniencia, hagamos que la utilidad representativa sea lineal en parámetros: $V_{nj} = \beta' x_{nj}$. Esta restricción no es necesaria, pero hace que la notación y el análisis sean más concisos. Usando (3.11) y la fórmula para las probabilidades logit, se muestra al final de este apartado que la condición de primer orden (3.12) se convierte en

$$\text{(3.13)} \qquad \sum_n \sum_i (y_{ni} - P_{ni}) x_{ni} = 0$$

Reorganizando y dividiendo ambos lados por N, obtenemos

$$\text{(3.14)} \qquad \frac{1}{N} \sum_n \sum_i y_{ni} x_{ni} = \frac{1}{N} \sum_n \sum_i P_{ni} x_{ni}$$

Esta expresión es fácilmente interpretable. Sea $\bar{x}$ el promedio de x entre las alternativas elegidas por los individuos muestreados: $\bar{x} = (1/\mathrm{N}) \sum_n \sum_i y_{ni} x_{ni}$. Sea $\hat{x}$ la media de x sobre las elecciones previstas de los decisores en la muestra: $\hat{x} = (1/\mathrm{N}) \sum_n \sum_i P_{ni} x_{ni}$. El promedio observado de x en la muestra es $\bar{x}$, mientras que $\hat{x}$ es la media predicha. Según (3.14), estas dos medias son iguales cuando el estimador de verosimilitud es máximo. Es decir, las estimaciones de máxima verosimilitud de β son aquellas que hacen que el promedio pronosticado de cada variable explicativa sea igual al promedio observado en la muestra. En este sentido, las estimaciones inducen al modelo a reproducir los promedios observados en la muestra.

Esta propiedad del estimador de máxima verosimilitud para los modelos logit adquiere un significado especial para constantes específicas de alternativa. Una constante específica de alternativa es el coeficiente de una variable indicador (*dummy variable*) que identifica a una alternativa. Una variable indicador para la alternativa j es una variable cuyo valor en la utilidad representativa de la alternativa i es $d_i^j = 1$ para $i = j$ y cero en caso contrario. Según (3.14) la constante estimada es la que cumple

$$\frac{1}{N} \sum_n \sum_i y_{ni} d_i^j = \frac{1}{N} \sum_n \sum_i P_{ni} d_i^j$$

$$S_j = \widehat{S_J},$$

donde S_j es la proporción o cuota de personas de la muestra que eligieron la alternativa j y $\widehat{S_J}$ es la proporción prevista para la alternativa j. Con constantes específicas de alternativa, las proporciones previstas para la muestra son iguales a las proporciones observadas. Por lo tanto, el modelo estimado es correcto en promedio dentro de la muestra. Esta característica es similar a la función que cumple una constante en un modelo de regresión lineal, donde la constante asegura que la media del valor predicho de la variable dependiente sea igual a su promedio observado en la muestra.

La condición de primer orden (3.13) proporciona otra interpretación importante. La diferencia entre la elección real de una persona, y_{ni}, y la probabilidad de esta elección, P_{ni} , es un error del modelo o residuo. El lado izquierdo de (3.13) es la covarianza muestral de los residuos con las variables explicativas. Las estimaciones de máxima verosimilitud son, por lo tanto, los valores de las βs que hacen que esta covarianza sea cero, es decir, que hacen que los residuos no estén correlacionados con las variables explicativas. Esta condición para las estimaciones logit es la misma que se aplica en modelos de regresión lineal. Para un modelo de regresión $y_n = \beta' x_n + \varepsilon_n$, la estimación ordinaria de mínimos cuadrados son los valores de β que hacen que $\sum_n (y_n - \beta' x_n) x_n = 0$. Este hecho se comprueba resolviendo para β: $\beta = (\sum_n x_n x_n')^{-1} (\sum_n x_n y_n')$, que es la fórmula para el estimador ordinario de mínimos cuadrados. Dado que $y_n - \beta' x_n$ es el residuo en el modelo de regresión, los estimadores hacen que los residuos no estén correlacionados con las variables explicativas.

Bajo esta interpretación, los estimadores pueden verse como una forma de proporcionar una muestra análoga a las características de la población. Hemos supuesto que las variables explicativas son exógenas, lo que significa que no están correlacionadas dentro de la población con los errores del modelo. Dado que las variables y los errores no están correlacionados en la población, tiene sentido elegir estimadores que hagan las variables y los residuos no correlacionados en la muestra. Los estimadores hacen justamente eso: proporcionan un modelo que reproduce en la muestra la covarianza nula que se produce en la población.

Los estimadores que resuelven las ecuaciones de la forma (3.13) se dice que emplean el método de los momentos, ya que para definir el estimador utilizan condiciones sobre los momentos (correlaciones en este caso) entre los residuos y las variables. Volveremos a estos estimadores cuando hablemos de la estimación asistida por simulación en el capítulo 10.

Hemos afirmado sin pruebas que (3.13) es la condición de primer orden para el estimador de máxima verosimilitud del modelo logit. Vamos a mostrar esta prueba ahora. La función de logaritmo de la verosimilitud (3.11) puede ser re-expresada como

$$LL(\beta) = \sum_n \sum_i y_{ni} ln P_{ni}$$

$$= \sum_n \sum_i y_{ni} ln \left(\frac{e^{\beta' x_{ni}}}{\sum_j e^{\beta' x_{nj}}} \right)$$

$$= \sum_n \sum_i y_{ni} (\beta' x_{ni}) - \sum_n \sum_i y_{ni} ln \left(\sum_j e^{\beta' x_{nj}} \right)$$

La derivada de la función log-verosimilitud se convierte en

$$\frac{dLL(\beta)}{d\beta} = \frac{\sum_n \sum_i y_{ni} (\beta' x_{ni})}{d\beta} - \frac{\sum_n \sum_i y_{ni} ln \left(\sum_j e^{\beta' x_{nj}} \right)}{d\beta}$$

$$= \sum_n \sum_i y_{ni} x_{ni} - \sum_n \sum_i y_{ni} \sum_j P_{nj} x_{nj}$$

$$= \sum_n \sum_i y_{ni} x_{ni} - \sum_n \left(\sum_j P_{nj} x_{nj} \right) \sum_i y_{ni}$$

$$= \sum_n \sum_i y_{ni} x_{ni} - \sum_n \left(\sum_j P_{nj} x_{nj} \right)$$

$$= \sum_n \sum_i (y_{ni} - P_{nj}) x_{ni}$$

Estableciendo este derivada igual a cero obtenemos la condición de primer orden (3.13).

Estimación en un subconjunto de alternativas

En algunas situaciones, el número de alternativas que enfrenta el decisor es tan grande que la estimación de los parámetros del modelo es muy costosa o incluso imposible. Con un modelo logit, la estimación puede realizarse en un subconjunto de las alternativas sin producir inconsistencia. Por ejemplo, un investigador que examina una situación de elección que involucra 100 alternativas puede estimar sobre un subconjunto de 10 alternativas para cada decisor de la muestra, incluyendo siempre la alternativa elegida por cada persona así como 9 alternativas más, seleccionadas al azar entre las 99 restantes. Si todas las alternativas tienen las mismas oportunidades de ser seleccionadas dentro del subconjunto, entonces la estimación puede realizarse en el subconjunto de alternativas como si fuera el conjunto completo. Si las alternativas tienen desigual probabilidad de ser seleccionadas, se requieren procedimientos de estimación más complejos. El procedimiento se describe como sigue.

Supongamos que el investigador ha utilizado algún método específico para la selección aleatoria de alternativas que forman el subconjunto que se utiliza en la estimación para cada decisor de la muestra. Denotamos el conjunto completo de alternativas como F y un subconjunto de alternativas como K. Sea $q(K|i)$ la probabilidad, bajo el método empleado por el investigador para seleccionar la muestra, de que el subconjunto K sea seleccionado teniendo en cuenta que el decisor eligió la alternativa i. Asumiendo que el subconjunto incluye necesariamente la alternativa elegida, tenemos que $q(K|i) = 0$ para cualquier K que no incluya i. La probabilidad de que la persona n elija la alternativa i del conjunto completo es P_{ni}. Nuestro objetivo es obtener una fórmula para la probabilidad de que la persona elija la alternativa i condicionada a que el investigador haya seleccionado el subconjunto K para él. Esta probabilidad condicionada se denota $P_n(i|K)$.

Esta probabilidad condicionada se obtiene de la siguiente manera. La probabilidad conjunta de que el investigador seleccione el subconjunto K y el decisor elija la alternativa i es $Prob(K, i) = q(K|i)Pni$. La probabilidad conjunta también se puede expresar con la condicionada opuesta como $Prob(K, i) = Pni(i|K)Q(K)$ donde $Q(K) = \sum_{j \in F} P_{nj} q(K|j)$ es la probabilidad marginal de que el investigador seleccione el subconjunto K sobre todas las alternativas que la persona podría elegir. Igualando estas dos expresiones y despejando $P_n(i|K)$, tenemos

$$P_n(i|K) = \frac{P_{ni}\, q(K|i)}{\sum_{j \in F} P_{nj} q(K|j)}$$

$$= \frac{e^{V_{ni}}\, q(K|i)}{\sum_{j \in F} e^{V_{nj}} q(K|j)}$$

$$(3.15) \qquad \frac{e^{V_{ni}}\, q(K|i)}{\sum_{j\in K} e^{V_{nj}} q(K|j)}$$

donde en la segunda línea hemos cancelado los denominadores de P_{ni} y P_{nj} $\forall j$, y en la tercera, la igualdad utiliza el hecho de que $q(K|j) = 0$ para cualquier j que no esté en K.

Supongamos que el investigador ha diseñado el proceso de selección de manera que $q(K|j)$ es la misma para todos los $j \in K$. Esta propiedad se produce, por ejemplo, si el investigador asigna la misma probabilidad de selección a todas las alternativas no escogidas, de modo que la probabilidad de escoger j en el subconjunto cuando i es la opción escogida por el decisor es la misma probabilidad de escoger i en el subconjunto cuando j es la opción elegida. McFadden (1978) llama a esto la "propiedad de condicionamiento uniforme", ya que el subconjunto de alternativas tiene una probabilidad uniforme (igual) de ser seleccionado, condicionada a que cualquiera de sus miembros haya sido escogido por el decisor. Cuando esta propiedad se cumple, $q(K|j)$ desaparece de la expresión anterior y la probabilidad se convierte en

$$P_n(i|K) = \frac{e^{V_{ni}}}{\sum_{j\in K} e^{V_{nj}}},$$

que es simplemente la fórmula logit para una persona que se enfrenta a las alternativas disponibles en el subconjunto K.

La función log-verosimilitud condicionada en virtud de la propiedad de condicionamiento uniforme es

$$CLL(\beta) = \sum_n \sum_{i\in K_n} y_{ni} ln \frac{e^{V_{ni}}}{\sum_{j\in K_n} e^{V_{nj}}},$$

donde K_n es el subconjunto seleccionado para la persona n . Esta función es la misma que la función log-verosimilitud dada en (3.11) excepto que el subconjunto de alternativas K_n sustituye, para cada persona de la muestra, al conjunto completo. La maximización de CLL proporciona un estimador consistente de β. Sin embargo, dado que hay información excluida de CLL que sí está incorporada en LL (es decir, información sobre alternativas que no están en cada subconjunto) el estimador basado en CLL no es eficiente.

Supongamos que el investigador diseña un proceso de selección que no presenta la propiedad de condicionamiento uniforme. En este caso, la probabilidad $q(K|i)$ se puede incorporar en el modelo como variable separada. La expresión en (3.15) se puede reescribir como

$$P_n(i|K) = \frac{e^{V_{ni}+ln\, q(K|i)}}{\sum_{j\in K} e^{V_{nj}+ln\, q(K|j)}}$$

Una variable z_{nj} calculada como $\ln q(K_n|j)$ se añade a la utilidad representativa de cada alternativa. El coeficiente de esta variable está limitado a 1 en la estimación.

Nos podemos plantear la siguiente cuestión: ¿por qué un investigador va a querer diseñar un procedimiento de selección que no satisfaga la propiedad de condicionamiento uniforme, cuando satisfacer esta propiedad hace que la estimación sea tan sencilla? Un ejemplo de los beneficios potenciales que tiene el condicionamiento no uniforme la proporcionan Train et al. (1987a) en su estudio de la demanda de telecomunicaciones. La situación de elección en este caso incluye una enorme cantidad de alternativas que representan las diferentes posibilidades de llamadas por hora del día, distancia y duración de las mismas. La gran mayoría de alternativas casi nunca fueron elegidas por los decisores dentro de la población. Si se hubiesen seleccionado las alternativas con igual probabilidad para cada alternativa, habría sido muy probable que los subconjuntos resultantes hubiesen consistido en su totalidad en alternativas que casi nunca fueron elegidas, junto con la alternativa realmente elegida por la persona. Comparar la alternativa elegida de una persona con un grupo de alternativas altamente indeseables por el individuo proporciona poca información sobre las razones que han motivado la elección de la persona. Para evitar este problema, se seleccionaron las alternativas en proporción a sus cuotas de mercado en la población (o para ser más precisos, estimaciones de las cuotas de mercado de la población). Este procedimiento incrementó la probabilidad de que las alternativas relativamente deseables se incluyesen en cada subconjunto de alternativas usado en la estimación.

3.7.2 Muestras basadas en la elección

En algunas situaciones, una muestra tomada sobre la base de factores exógenos incluiría pocas personas que han optado por alternativas concretas. Por ejemplo, en una elección de calentadores de agua, una muestra aleatoria de hogares en la mayoría de las áreas incluiría sólo un pequeño número de hogares que hayan elegido sistemas de calefacción de agua solares. Si el investigador está particularmente interesado en los factores que afectan a la penetración en el mercado de los dispositivos solares, una muestra al azar tendría que ser muy grande para asegurar un número razonable de hogares con calor solar.

En situaciones como éstas, el investigador podría optar por seleccionar la muestra, o parte de la muestra, sobre la base de la elección que se analiza. Por ejemplo, el investigador que desea estudiar los calentadores de agua podría complementar una muestra aleatoria de hogares con hogares que se sabe (quizá a través de registros de ventas en las tiendas si el investigador tiene acceso a los mismos) que han instalado recientemente calentadores de agua solares.

Las muestras seleccionadas sobre la base de las elecciones de los decisores pueden ser puramente basadas en la elección o un híbrido entre selección basada en la elección y en factores exógenos. En una muestra totalmente basada en la elección, la población se divide en grupos por cada alternativa elegida y los decisores se extraen al azar dentro de cada grupo, aunque en proporciones diferentes. Por ejemplo, un investigador que esté estudiando cómo eligen las personas la ubicación de su lugar de residencia y esté interesado en identificar los factores que contribuyen a que la gente elija una ubicación en particular, podría extraer al azar dentro de esa ubicación concreta uno de cada L hogares y extraer aleatoriamente del resto de ubicaciones uno de cada M hogares, donde M es mayor que L. Este procedimiento asegura que el investigador dispone de un número adecuado de personas de la zona de interés en la muestra. Una muestra híbrida es como la elaborada por el investigador interesado en la calefacción solar de agua del ejemplo anterior, en la que una muestra exógena se complementa con una muestra extraída sobre la base de las decisiones de los hogares.

La estimación de los parámetros del modelo con muestras extraídas al menos parcialmente sobre la base de la elección del decisor es bastante compleja en general y varía con el método exacto del procedimiento de muestreo. Para los lectores interesados, Ben-Akiva y Lerman (1985, pp 234-244) proporcionan un estudio útil. Uno de los resultados es particularmente relevante, ya que permite a los investigadores estimar modelos logit sobre muestras basadas en la elección sin necesidad de emplear procedimientos de estimación complejos. Este resultado, debido a Manski y Lerman (1977), se puede describir de la siguiente manera. Si el investigador está utilizando una muestra basada totalmente en la elección de los decisores e incluye una constante específica de alternativa en la utilidad representativa para cada alternativa, estimar un modelo logit como si la muestra fueses exógena produce estimaciones consistentes para todos los parámetros del modelo a excepción de las constantes específicas de alternativa. Además, estas constantes resultan sesgadas por un factor conocido y por lo tanto se pueden ajustar de manera que las constantes ajustadas sean consistentes. En particular, la esperanza de la constante estimada para la alternativa j, denominada $\widehat{\alpha}_j$, está relacionada con la verdadera constante α_j^* por

$$E(\widehat{\alpha_j}) = {\alpha_j}^* - ln(A_j/S_j),$$

donde $\mathrm{A_j}$ es la proporción de decisores en la población que eligió la alternativa j y $\mathrm{S_j}$ es la proporción en la muestra basada en la elección que eligió la alternativa j . En consecuencia, si $\mathrm{A_j}$ es conocido (es decir, si las cuotas de mercado de la población son conocidas para cada alternativa) entonces $\widehat{\alpha_j}$ es un estimador consistente de la constante específica de alternativa, el cual se estima en la muestra basada en la elección y al que se le suma el logaritmo del ratio entre la proporción de esa alternativa en la población y en la muestra.

3.8 Bondad de ajuste y pruebas de hipótesis

Trataremos a continuación la bondad de ajuste y los test de hipótesis en el contexto de los modelos logit, en los cuales la función logaritmo de verosimilitud (log-verosimilitud) se puede calcular con exactitud. Los conceptos aplican a otros modelos, con los debidos ajustes por la varianza de la simulación, cuando la función log-verosimilitud se simula en lugar de calcularse con exactitud.

3.8.1 Bondad de ajuste

Un estadístico denominado *índice de ratio de verosimilitud* (*likelihood ratio index*) se utiliza a menudo con modelos de elección discreta para medir lo bien que se ajustan a los datos. Dicho de forma más precisa, el estadístico mide lo bien que el modelo, con sus parámetros estimados, se comporta en comparación con un modelo en el que todos los parámetros son iguales a cero (que generalmente es equivalente a no tener ningún modelo en absoluto). Esta comparación se realiza sobre la base de la función log-verosimilitud, evaluada tanto para los parámetros estimados como para todos los parámetros iguales a cero.

El índice de ratio de verosimilitud se define como

$$\rho = 1 - \frac{LL(\hat{\beta})}{LL(0)},$$

donde $LL(\hat{\beta})$ es el valor de la función log-verosimilitud en los parámetros estimados y $LL(0)$ es su valor cuando todos los parámetros se igualan a cero. Si los parámetros estimados no lo hacen mejor, en términos de la función de verosimilitud, que los parámetros nulos (es decir, si el modelo estimado no es mejor que no tener modelo) entonces $LL(\hat{\beta}) = LL(0)$ y por lo tanto $\rho = 0$. Este es el valor más bajo que ρ puede tomar (dado que si $LL(\hat{\beta})$ fuese inferior a $LL(0)$, entonces $\hat{\beta}$ no sería la estimación de máxima verosimilitud).

En el extremo opuesto, supongamos que el modelo estimado ha resultado ser tan bueno que podríamos predecir perfectamente la elección de cada decisor de la muestra. En este caso, la función de verosimilitud en los parámetros estimados sería igual a uno, ya que la probabilidad de observar las elecciones que realmente se hicieron es uno. Y dado que el logaritmo de uno es cero, la función log-verosimilitud sería cero en los parámetros estimados. Si $LL(\hat{\beta}) = 0$ entonces $\rho = 1$. Este es el valor más alto que ρ puede tomar. En resumen, el índice de ratio de verosimilitud va desde cero, cuando los parámetros estimados no son mejores que los parámetros nulos, hasta uno, cuando los parámetros estimados predicen perfectamente las decisiones de los decisores incluidos en la muestra.

Es importante destacar que el índice de ratio de verosimilitud no es en absoluto similar en su interpretación a la R^2 utilizada en regresiones, a pesar de que los dos estadísticos tienen el mismo rango. R^2 indica el porcentaje de la variación de la variable dependiente que se "explica" por el modelo estimado. El coeficiente de verosimilitud no tiene un sentido intuitivamente interpretable para los valores que se encuentran entre los extremos (cero y uno). Es el porcentaje de incremento en la función log-verosimilitud por encima del valor resultante cuando fijamos a cero los parámetros (ya que $\rho = 1 - LL(\hat{\beta})/LL(0) = (LL(0) - LL(\hat{\beta}))/LL(0))$. Sin embargo, el significado de un aumento en dicho porcentaje no está claro. En la comparación de dos modelos estimados con los mismos datos y con el mismo conjunto de alternativas (tal que $LL(0)$ es igual para ambos modelos), por lo general es válido decir que el modelo con el ρ más alto se ajusta mejor a los datos. Pero eso no es más que decir que es preferible incrementar el valor de la función log-verosimilitud. Dos modelos estimados en muestras que no sean idénticas o con un conjunto diferente de alternativas para cualquier decisor en la muestra no se pueden comparar a través de sus valores de índice de ratio de verosimilitud.

Otro estadístico de bondad de ajuste que se utiliza ocasionalmente, pero que realmente debería ser evitado, es el "porcentaje correctamente predicho". Este estadístico se calcula identificando para cada decisor en la muestra la alternativa con la probabilidad más alta de ser elegida, basándose en el modelo estimado, y determinando si esta alternativa fue o no la que realmente el decisor escogió. El porcentaje de decisores en la muestra para los cuales la alternativa más probable y la alternativa elegida es la misma se denomina "porcentaje correctamente predicho".

Este estadístico incorpora una noción opuesta al significado de las probabilidades y al propósito de especificar probabilidades de elección. El estadístico está basado en la idea de que el investigador predice que el decisor elegirá la alternativa para la que el modelo ha estimado una probabilidad de elección más alta. Sin embargo tal y como vimos en la formulación de las probabilidades de elección en el capítulo 2, el investigador no tiene suficiente información para predecir la elección del decisor. El investigador sólo tiene información suficiente como para indicar la probabilidad de que el decisor elija cada alternativa. Cuando calcula probabilidades de elección, el investigador está afirmando que si la situación de elección se repitiese varias veces (o fuese afrontada por muchas personas con los mismos atributos) cada alternativa sería elegida una determinada proporción de las veces. Esto es bastante diferente a decir que la alternativa con la probabilidad más alta será la elegida cada vez.

Un ejemplo puede ser útil para comprender la diferencia. Supongamos que un modelo estimado predice probabilidades de elección de 0.75 y 0.25 en una situación de elección con dos alternativas. Estas probabilidades significan que si 100 personas se enfrentan a las utilidades representativas que dieron lugar a estas probabilidades (o una persona enfrenta estas utilidades representativas 100 veces) la mejor predicción que el investigador puede hacer sobre cuántas personas podrían elegir cada alternativa son 75 y 25. Sin embargo, el estadístico "porcentaje correctamente predicho" está basado en la idea de que la mejor predicción para cada persona es la alternativa con mayor probabilidad. Esta noción podría predecir que una alternativa sería elegida por las 100 personas, mientras que la otra alternativa nunca sería elegida. El procedimiento olvida el sentido de las probabilidades, obviamente da cuotas de mercado inexactas y parece implicar que el investigador tiene información perfecta.

3.8.2 Test de hipótesis

Como sucede con las regresiones, para probar hipótesis sobre los parámetros individuales en los modelos de elección discreta – por ejemplo probar si el parámetro es cero - se usan estadísticos-t estándar (*t-statistics*). Para hipótesis más complejas, casi siempre puede usarse un test de ratio de verosimilitud de la siguiente manera. Considere la hipótesis nula H que puede ser expresada como restricciones sobre los valores de los parámetros. Dos de las hipótesis más comunes son (1) varios parámetros son iguales a cero y (2) dos o más parámetros son iguales entre ellos. La estimación de máxima verosimilitud restringida de los parámetros (etiquetados $\hat{\beta}^H$) es el valor de β que da el mayor valor de LL sin violar las restricciones de la hipótesis nula H. Definamos el ratio de verosimilitudes $R = L(\hat{\beta}^H)/L(\hat{\beta}\,)$, donde $\hat{\beta}^H$ es el valor máximo (restringido) de la función de verosimilitud (sin logaritmo) bajo la hipótesis nula H y $\hat{\beta}$ es el máximo sin restricciones de la función de verosimilitud. Al igual que en los test de ratios de verosimilitud para modelos diferentes a los de elección discreta, el estadístico de test definido como $-2logR$ se distribuye chi-cuadrado con un número de grados de libertad igual al número de restricciones que implica la hipótesis nula. Por lo tanto, el estadístico de test de hipótesis es $-2(LL(\hat{\beta}^H) - LL(\hat{\beta}))$. Dado que el logaritmo de la verosimilitud es siempre negativo, esto es simplemente dos veces la (magnitud de la) diferencia entre los máximos restringidos y no restringidos de la función log-verosimilitud. Si este valor supera al valor crítico de chi-cuadrado con los grados de libertad apropiados, entonces la hipótesis nula es rechazada.

Hipótesis nula I: Los coeficientes de varias variables explicativas son cero

Para probar esta hipótesis, estime el modelo dos veces: una vez con estas variables explicativas incluidas y una segunda vez sin ellas (ya que excluyendo las variables obliga a que sus coeficientes sean cero). Observe el valor máximo de la función log-verosimilitud para cada estimación; dos veces la diferencia entre estos valores máximos es el valor del estadístico de test. Compare el estadístico de test con el valor crítico de chi-cuadrado con un número de grados de libertad igual al número de variables explicativas excluidas de la segunda estimación.

Hipótesis nula II: Los coeficientes de las dos primeras variables don las mismas

Para probar esta hipótesis, estime el modelo dos veces: una vez con cada una de las variables explicativas entrando por separado en el modelo, incluyendo las dos primeras; luego con las dos primeras variables reemplazadas por una única variable que es la suma de las dos variables (dado que sumar las variables obliga a sus coeficientes a ser iguales). Observe el valor máximo de la función log-verosimilitud para cada una de las estimaciones. Multiplique la diferencia de estos valores máximos por dos y compare esta cifra con el valor crítico de chi-cuadrado con un grado de libertad.

3.9 Estudio de un caso: predicción para un nuevo sistema de tráfico

Una de las primeras aplicaciones de los modelos logit que constituye una prueba importante de sus capacidades, surgió a mediados de la década de 1970 en el área de la Bahía de San Francisco. Un nuevo sistema de tren, llamado *Bay Area Rapid Transit (BART)* había sido construido. Daniel McFadden obtuvo una subvención de la *National Science Foundation* para aplicar modelos logit a la elección del medio de transporte de los viajeros en el área de la bahía y usar los modelos para predecir el número de pasajeros del sistema BART. Tuve la suerte de trabajar como su asistente de investigación en este proyecto. Una muestra de los pasajeros fue seleccionada antes de que el BART fuese abierto al público. Sobre esta muestra se estimaron modelos de elección sobre el medio de transporte. Estas estimaciones proporcionaron información importante sobre los factores que entran en juego en las decisiones de los viajeros, incluyendo el valor otorgado al ahorro de tiempo. Posteriormente, se utilizaron los modelos para pronosticar las decisiones que los pasajeros incluidos en la muestra harían una vez que el BART estuviese disponible. Cuando el BART fue abierto al público, los pasajeros fueron contactados de nuevo y se observaron sus elecciones reales de medios de transporte. La proporción prevista de usuarios del BART se comparó con la proporción realmente observada. Los modelos lograron predecir bastante bien las proporciones, de forma mucho más precisa que los procedimientos empleados por los consultores del BART, que no habían utilizado los modelos de elección discreta.

El equipo del proyecto recolectó datos de 771 viajeros antes de la apertura del BART. Se consideraron cuatro medios de transporte como opciones disponibles para viajar al trabajo: (1) la conducción de un automóvil por uno mismo, (2) tomar el autobús y caminar hasta la parada de autobús, (3) tomar el autobús y conducir hasta la parada de autobús y (4) compartir automóvil entre viajeros. El tiempo y el costo de viajar en cada medio de transporte se determinaron para cada viajero, basándose en la ubicación del domicilio y del trabajo de cada persona. El tiempo de viaje se diferenció entre tiempo caminando a pie (para el modo autobús – a pie), tiempo de espera (para los dos modos de autobús) y tiempo en vehículo (para todos los modos). También se registraron las características de los viajeros, incluyendo ingresos, tamaño del hogar, número de vehículos y número de conductores en el hogar, y si el viajero era el cabeza de familia. Con toda esta información se estimó un modelo logit con utilidad lineal en los parámetros.

El modelo estimado se muestra en la tabla 3.1, extraído de Train (1978).

Tabla 3.1. Modelo Logit de elección de medio de transporte para viajar al trabajo

Variable explicativa[a]		Coeficiente	Estadístico-t
Costo dividido por salario después de impuestos, minutos (1-4)		−0.0284	4.31
Tiempo dentro del automóvil, minutos (1, 3, 4)		−0.0644	5.65
Tiempo dentro del transporte público, minutos (2, 3)		−0.0259	2.94
Tiempo a pie, minutos (2, 3)		−0.0689	5.28
Tiempo de espera en transbordos, minutos (2, 3)		−0.0538	2.30
Número de transbordos (2, 3)		−0.1050	0.78
Tiempo de paso del primer autobús, minutos (2, 3)		−0.0318	3.18
Ingresos del hogar con techo de $7500 (1)		0.00000454	0.05
Ingresos del hogar – $7500 con suelo 0, techo $3000 (1)		−0.0000572	0.43
Ingresos del hogar – $10500 con suelo 0, techo $5000 (1)		−0.0000543	0.91
Número de conductores en el hogar (1)		1.02	4.81
Número de conductores en el hogar (3)		0.990	3.29
Número de conductores en el hogar (4)		0.872	4.25
Indicador de si el trabajador es cabeza de familia (1)		0.627	3.37
Densidad de empleo en la ubicación del trabajo (1)		−0.0016	2.27
Ubicación del hogar en o cerca del distrito principal de negocios (1)		−0.502	4.18
Autos por conductor con techo 1 (1)		5.00	9.65
Autos por conductor con techo 1 (3)		2.33	2.74
Autos por conductor con techo 1 (4)		2.38	5.28
Indicador de sólo automóvil (1)		−5.26	5.93
Indicador de autobús con acceso en automóvil (3)		−5.49	5.33
Indicador de compartición de automóvil (4)		−3.84	6.36
Índice de ratio de verosimilitud	0.4426		
Log-verosimilitud en convergencia	−595.8		
Número de observaciones	771		
Valor del tiempo ahorrado como % del salario:			
Tiempo en automóvil	227		3.20
Tiempo en transporte público	91		2.43
Tiempo a pie	243		3.10
Tiempo de espera en transbordo	190		2.01

a La variable entra en los medios de transporte indicados en paréntesis y es cero en el resto de medios.
Medios de transporte: 1. Sólo Automóvil. 2. Autobús con acceso a pies. 3. Autobús con acceso en automóvil. 4. Automóvil compartido.

El costo del viaje se dividió por el salario del viajero para reflejar la expectativa de que los trabajadores con salarios más bajos van a estar más preocupados por el costo que los trabajadores mejor remunerados. El tiempo de viaje pasado en el vehículo entra en el modelo por separado para automóvil y autobús, con el fin de indicar que los viajeros podrían encontrar el tiempo empleado en el autobús más o menos molesto que el tiempo durante el que conducen un automóvil. Los viajes en autobús a menudo implican transbordos y estos transbordos pueden ser pesados para los viajeros. Es por ello que el modelo incluye el número de transbordos y el tiempo estimado de espera en los transbordos. El tiempo transcurrido entre el paso de dos autobuses para la primera línea de autobús que el pasajero tomaría se incluye como una medida de la cantidad máxima de tiempo que la persona tendría que esperar usando este medio.

Los coeficientes estimados de costo y de los diversos componentes de tiempo proporcionan información sobre el valor del tiempo. Por definición, el valor del tiempo es el costo extra en que una persona estaría dispuesta a incurrir para ahorrar tiempo. La utilidad toma la forma $U_{nj} = \alpha c_{nj}/w_n + \beta t_{nj} + \cdots$, donde c es el costo y t es el tiempo. La derivada total con respecto a los cambios en el tiempo y el costo es $dU_{nj} = (\alpha/w_n) dc_{nj} + \beta dt_{nj}$, que igualamos a cero y resolvemos para dc/dt para poder encontrar el cambio en el costo que mantiene inalterada la utilidad de un cambio en el tiempo: $dc/dt = -(\beta/\alpha)w_n$. Por tanto, el valor del tiempo es una proporción β/α del salario de la persona. Los valores estimados del tiempo se presentan en la parte inferior de la tabla 3.1. El tiempo ahorrado al viajar en autobús es valorado como un 91% del salario ((-0.0259/-0.0284) × 100), mientras que el ahorro de tiempo de viaje en automóvil vale más del doble: 227% del salario. Esta diferencia sugiere que los pasajeros consideran la conducción considerablemente más pesada que ir en autobús, cuando se evalúa por minuto invertido en el viaje. Los viajeros parece que eligen el automóvil no porque les guste la conducción en sí, sino porque conducir es generalmente más rápido. Caminar se considera más molesto que esperar un autobús (243% del salario frente a 190%) y esperar un autobús es más molesto que viajar en él.

Los ingresos entran en la utilidad representativa de la alternativa "sólo automóvil". Entran de una forma lineal por tramos para permitir la posibilidad de que un ingreso adicional tenga un impacto diferente dependiendo del nivel general de ingresos. Ninguna de las variables de ingreso contribuye significativamente. Al parecer, dividir el costo del viaje por los salarios suprime cualquier efecto que el ingreso pudiera tener en la elección del medio de transporte de un viajero. Es decir, los salarios más altos inducen al viajero a preocuparse menos por los gastos del viaje, pero no inducen una predilección por la conducción más allá del impacto que tiene a través de los costos. El número de personas y el número de vehículos por conductor en el hogar tienen un efecto significativo en la elección del medio de transporte, tal y como se esperaba. Se han incluido también constantes específicas de alternativa, con la constante para la alternativa "autobús - a pie" normalizada a cero.

El modelo en la tabla 3.1 se utilizó para predecir la elección de medio de transporte de los pasajeros después de la inauguración del BART. Se consideró como conjunto de elección los cuatro medios de transporte enumerados anteriormente más dos modalidades del BART que se diferenciaban en función de si la persona necesitaba usar el autobús o conducir para llegar a la estación del BART. La tabla 3.2 presenta las cuotas de mercado previstas y reales para cada medio de transporte. La demanda del BART se estimó que sería de un 6.3%, en comparación con el 6.2% que realmente obtuvo. La estrecha correspondencia entre predicción y valor real es notable.

Tabla 3.2. Predicciones para post-apertura del BART

	Cuota real	Cuota prevista
Sólo automóvil	59.90	55.84
Autobús con acceso a pie	10.78	12.51
Autobús con acceso en automóvil	1.426	2.411
BART con acceso a pie	0.951	1.053
BART con acceso en automóvil	5.230	5.286
Automóvil compartido	21.71	22.89

Las cifras del cuadro 3.2 tienden a enmascarar varias complicaciones que surgieron en la predicción. Por ejemplo, caminar hasta la estación del BART fue incluido originalmente como un medio de transporte separado. El modelo predijo esta opción muy pobremente, sobre-estimando el número de personas que iban a caminar hasta el BART por un factor de doce. El problema fue investigado y se encontró que se debía principalmente a las diferencia entre la experiencia de caminar a las estaciones del BART y la de caminar hacia el autobús, debido a los barrios en los que se ubicaban las estaciones del BART. Estos problemas se analizan con mayor detenimiento por McFadden et al. (1977)

3.10 Obtención de las probabilidades logit

Se afirmó sin pruebas en la sección 3.1 que si el componente no observado de utilidad se distribuye con una densidad iid valor extremo para cada alternativa, las probabilidades de elección toman la forma de la ecuación (3.6). Nos proponemos a continuación demostrar este resultado. De (3.5) tenemos

$$P_{ni} = \int_{s=-\infty}^{\infty} \left(\prod_{j \neq i} e^{-e^{-(s+V_{ni}-V_{nj})}} \right) e^{-s} e^{-e^{-s}} ds$$

donde s es ε_{ni}. Nuestra tarea consiste en evaluar esta integral. Observando que $V_{ni} - V_{ni} = 0$ y agrupando términos en el exponente de e, tenemos

$$P_{ni} = \int_{s=-\infty}^{\infty} \left(\prod_{j} e^{-e^{-(s+V_{ni}-V_{nj})}} \right) e^{-s} ds$$

$$= \int_{s=-\infty}^{\infty} exp\left(-\sum_{j} e^{-(s+V_{ni}-V_{nj})} \right) e^{-s} ds$$

$$= \int_{s=-\infty}^{\infty} exp\left(-e^{-s} \sum_{j} e^{-(V_{ni}-V_{nj})} \right) e^{-s} ds$$

Definimos $t = \exp(-s)$ tal que $-\exp(-s)\, ds = dt$. Tenga en cuenta que a medida que s tiende a infinito, t se aproxima a cero, y cuando s se acerca a menos infinito, t se convierte infinitamente grande. Usando este nuevo término llegamos a

$$P_{ni} = \int_{\infty}^{0} exp\left(-t \sum_{j} e^{-(V_{ni}-V_{nj})} \right) (-dt)$$

$$= \int_{0}^{\infty} exp\left(-t \sum_{j} e^{-(V_{ni}-V_{nj})} \right) dt$$

$$= \left. \frac{exp\left(-t \sum_{j} e^{-(V_{ni}-V_{nj})} \right)}{-\sum_{j} e^{-(V_{ni}-V_{nj})}} \right|_{0}^{\infty}$$

$$= \frac{1}{\sum_{j} e^{-(V_{ni}-V_{nj})}} = \frac{e^{V_{ni}}}{\sum_{j} e^{V_{nj}}},$$

según se pretendía demostrar.