

2. La modelización de la demanda de transporte

2.1 Introducción

El objetivo de un sistema de transporte es satisfacer la demanda mediante la provisión de una oferta adecuada que responda a sus exigencias. Los servicios de transporte surgen como consecuencia de la necesidad que tienen los individuos de realizar actividades que implican desplazamientos; por ejemplo, ir al trabajo, llevar los niños al colegio, ir al teatro, al médico, etc. Por esta razón se dice que la demanda de transporte es derivada. En general, no se demanda viajar *per se* sino con el objeto de realizar alguna actividad localizada en el espacio y en el tiempo. Estamos, por tanto, ante una demanda cualitativa y diferenciada en el sentido de que existen viajes por múltiples motivos, a distintas horas del día y que pueden realizarse en distintos modos de transporte. En lo que respecta a la oferta, nos encontramos con un servicio que no puede ser almacenado para ser ofertado, por ejemplo, en períodos donde existe una mayor demanda. Se produce, por tanto, un desequilibrio entre la cantidad ofertada y la demandada debido a la existencia de períodos con mayor nivel de demanda (hora punta) y períodos con menores niveles de demanda (hora valle), siendo imposible el “trasvase” de la capacidad ociosa de un período a otro. Además, la demanda se encuentra

localizada en un espacio, que se va ampliando con cierta celeridad, lo que suele producir problemas de coordinación que afectan al equilibrio del sistema.

Para satisfacer de forma eficiente la demanda de transporte es necesaria la planificación en el corto y largo plazo con un conocimiento bueno y preciso de los distintos atributos que la determinan. Existen varios argumentos que justifican esta planificación. La provisión de infraestructura de transporte es costosa dado su alto grado de indivisibilidad, y requiere, en muchos casos, de largos periodos desde que se planifica su construcción hasta que ésta se ve totalmente finalizada. En segundo lugar, existe una segregación entre el proveedor, generalmente público, de la infraestructura y el operador de la misma, lo que implica la necesidad de una regulación que defina los estándares del servicio de transporte que se desea ofrecer. Por último, la operación del servicio de transporte está asociada a determinadas externalidades (accidentes, contaminación, etc) que distorsionan las decisiones a tomar, ya que los usuarios del sistema no internalizan el coste que su elección de viaje genera.

Los argumentos expuestos justifican la importancia de realizar estudios de demanda que permitan evaluar los proyectos de transporte de la manera más completa posible, en el sentido de que sea factible analizar de forma conjunta los aspectos de la demanda y de la oferta con el fin de conocer el efecto de futuras políticas de transporte que puedan introducirse (por ejemplo, variación de algunos atributos de alternativas ya existentes o la introducción de nuevos modos de transporte).

La modelización de la demanda de transporte se ha basado tradicionalmente en el empleo de dos tipos de modelos: los modelos agregados o de *primera generación* y los modelos desagregados o de *segunda generación*. Los primeros fueron utilizados mayoritariamente en los estudios de transporte hasta finales de los años 70. Estos

modelos emplean datos agregados, por ejemplo, promedio de las variables a nivel zonal para las distintas zonas en las que se divide el área de estudio. La agregación de los datos presenta algunos inconvenientes. En primer lugar, es necesario disponer de una cantidad de datos importante ya que cada observación resulta de la obtención de un promedio de un cierto número de datos individuales. En segundo lugar, la existencia de sesgo entre las unidades agregadas es más probable que en los modelos desagregados. La conducta individual puede ser ocultada por características no identificadas asociadas a la zona, lo que se conoce como *falacia ecológica*³.

Por su parte, los modelos desagregados comienzan a ser populares en los años 80 por el hecho de que presentan ciertas ventajas sobre los modelos utilizados hasta el momento. Este tipo de modelos, se basa en el uso de los datos a nivel individual, lo que permite una mejor comprensión de los comportamientos de viaje, puesto que se basan en las teorías de la elección individual⁴. Las ventajas que presenta este tipo de modelos respecto a los de *primera generación* son varias. En primer lugar, el hecho de utilizar cada dato individual como una observación hace que sea más eficiente en el uso de la información en que los modelos agregados. En segundo lugar, ocupan toda la variabilidad de la que se dispone con los datos individuales. En tercer lugar, los resultados de los modelos a nivel individual pueden ser utilizados a cualquier nivel de agregación⁵. En cuarto lugar, la estimación de los parámetros para cada una de las variables explicativas consideradas en el modelo es explícita, lo que deriva en cierta

³ Un ejemplo para explicar la *falacia ecológica* es el caso de un modelo de frecuencia de viajes por zonas a través del cual se obtiene que el número de viajes por hogar decrece con el ingreso. Al observar los datos individuales se demuestra lo contrario que es lo que cabría esperar. Para un mayor detalle ver Ortúzar y Willumsen, 2001, pp. 221.

⁴ Concretamente, vamos a estudiar la Teoría de la Utilidad Aleatoria en el epígrafe 2.2.3 y a partir de la misma se derivan distintos modelos desagregados o de elección discreta.

⁵ Para la predicción con este tipo de modelos es necesaria la agregación que no es trivial en muchos casos y que será estudiada en el epígrafe 2.4.

flexibilidad para representar las variables de política; además, los parámetros estimados tienen una interpretación directa en términos de la importancia relativa de cada variable explicativa considerada en la elección. Y finalmente, se trata de modelos más estables en el tiempo y en el espacio.

Este capítulo se centra en el análisis de los aspectos más relevantes relativos a la modelización de la demanda de transporte basada en el empleo de modelos desagregados. En la segunda sección se estudian los principales modelos analizando tanto los fundamentos microeconómicos como los econométricos. En la tercera sección se estudian los procedimientos de estimación que se usan habitualmente para estimar este tipo de modelos. En la sección cuarta se aborda el problema de la predicción agregada a partir de estimaciones procedentes de modelos desagregados. Por último, en la sección quinta se analizan las diferentes fuentes de datos empleadas por estos modelos, haciendo especial referencia a la modelización con preferencias declaradas.

2.2 Los modelos de demanda desagregados

2.2.1 Fundamentos microeconómicos

La teoría económica clásica plantea un modelo de comportamiento del consumidor donde las preferencias que definen dicho comportamiento se expresan a través de una función de utilidad que depende de la cantidad consumida de distintos bienes de naturaleza divisible. Este modelo está caracterizado por la maximización de su nivel de satisfacción sujeto a una restricción presupuestaria, que indica que el gasto total en bienes no debe exceder a la renta disponible por el individuo. Postulados posteriores (Lancaster, 1966) que modifican esta teoría, establecen que una fuente de utilidad proviene de las características de los bienes (sabor, color, etc) y no de los bienes *per se*, de modo que unas características determinadas se obtendrían a partir de la compra de un

conjunto de bienes de mercado. La incorporación del consumo de bienes de naturaleza discreta (ver la formalización en McFadden, 1981) permite reformular el problema clásico de comportamiento del consumidor. Siguiendo a Jara-Díaz (1998), la función de utilidad a maximizar va a depender de las cantidades consumidas de bienes de naturaleza continua, representadas en el vector (X) y por la elección entre un conjunto de alternativas discretas representadas por un vector Q_j que recoge sus principales características o atributos. El problema de decisión del consumidor será:

$$\begin{aligned}
 & \underset{x,j}{\text{Max}} U(X, Q_j) \\
 & \text{s.a.} \\
 & \sum P_i X_i + c_j \leq I \\
 & X_i \geq 0 \quad j \in M
 \end{aligned} \tag{2.1}$$

siendo:

P_i , el precio del bien i

X_i , cantidad consumida del bien i

c_j , el coste de la alternativa discreta j

I , la renta del individuo

M , el conjunto de alternativas disponibles

La resolución de las condiciones de primer orden del problema (2.1) para cada j permite obtener las demandas condicionadas $X_j(P, I - c_j, Q_j)$ a la elección discreta j . La sustitución de estas funciones en la utilidad da lugar a la función de utilidad indirecta

condicional (FUIC) $V_j = V_j(P, I - c_j, Q_j)$ ⁶. Esta función representa la máxima utilidad que puede obtener el individuo una vez elegida la alternativa j . Al maximizar en j , el individuo elegirá aquella alternativa que le proporcione máxima utilidad. Si se define $V^* = \text{Max}_{j \in J} V_j(P, I - c_j, Q_j)$, la aplicación de la identidad de Roy permite obtener la demanda de las alternativas discretas:

$$\frac{\frac{\partial V^*}{\partial c_i}}{\frac{\partial V^*}{\partial I}} = \delta_i = \begin{cases} 1 & \text{si } V_i \geq V_j \quad \forall j \neq i \\ 0 & \text{en otro caso} \end{cases} \quad (2.2)$$

que en este caso es una variable discreta.

A partir de V_j podemos obtener la utilidad marginal de la renta y el valor subjetivo de la característica k :

$$\text{Utilidad marginal de la renta: } UMR = \frac{\partial V_j}{\partial I} = - \frac{\partial V_j}{\partial c_j} \quad (2.3)$$

$$\text{Valor subjetivo de la característica } k: SV_k = - \frac{\partial V_j / \partial q_{kj}}{\partial V_j / \partial I} \quad (2.4)$$

Para obtener estas magnitudes es preciso especificar la forma funcional de V_j y estimarla empíricamente. En la modelización basta considerar la parte relevante de V_j en la toma de decisiones del individuo. Ésta se denomina función de utilidad indirecta condicional truncada⁷. En general, cualquier función de utilidad indirecta que cumpla las

⁶ Esta función es continua en $(I - c_j, p, Q_j)$, dos veces diferenciable y homogénea de grado cero en $(I - c_j, p)$, estrictamente cuasiconvexa en p , y además se cumple que $\frac{\partial V_j}{\partial (I - c_j)} > 0$, McFadden (1981).

⁷ Se obtiene al comparar V_i con V_j .

propiedades matemáticas adecuadas⁸ puede ser aproximada a través de su desarrollo de Taylor por una especificación lineal en los parámetros (McFadden, 1981). Muchos de los modelos aplicados en la literatura consideran la aproximación local de primer orden que da lugar a una especificación lineal en los parámetros y en las variables. Bajo esta especificación, la UMR coincide con menos el parámetro del coste de la alternativa y el valor subjetivo de la característica k se obtiene a partir del cociente entre el parámetro de dicha característica y el del coste. Un inconveniente de definir funciones de utilidad lineales es que la renta no juega ningún papel en las decisiones de los individuos, es decir, las elecciones son independientes de la renta. Jara-Díaz y Videla (1989) demuestran que basta considerar una expansión de Taylor de orden igual o superior a dos (términos de orden dos para el coste) para poder considerar la renta en la elección de la alternativa.

2.2.2 La teoría microeconómica del valor del tiempo

Entre los modelos alternativos al modelo clásico de comportamiento del consumidor están aquellos que reconocen que el tiempo influye en las decisiones y en las restricciones a las que se enfrentan los consumidores. La valoración económica de los ahorros de tiempo dedicado al consumo de ciertos bienes o servicios constituye una pieza clave en la evaluación de políticas que tienen como objetivo la reducción de esos tiempos de consumo. En el caso del transporte la obtención de una cuantificación económica de estos ahorros de tiempo cobra una especial relevancia.

⁸ U es continua en (Y, Q_j) para cada j , dos veces diferenciable con continuidad en Y con $\frac{\partial U}{\partial Y} > 0$ y estrictamente cuasiconcava para cada Q_j y cada j .

En este epígrafe se presenta una revisión de los principales modelos que incorporan el tiempo al modelo de comportamiento del consumidor⁹. Podemos distinguir fundamentalmente dos tipos de modelos. En primer lugar, los modelos clásicos que se sitúan en un contexto general donde se asume que los bienes de consumo son de tipo continuo o divisibles (Becker, 1965; DeSerpa, 1971; 1973). En segundo lugar, los modelos desarrollados en el ámbito del transporte que se sitúan en el contexto microeconómico de las elecciones discretas (elección modal, generalmente) y por tanto consideran la posibilidad de consumir bienes de tipo discreto: (Train y McFadden, 1978; Jara-Díaz y Farah, 1987; Bates y Roberts, 1986).

La primera consideración del tiempo en el modelo de comportamiento del consumidor fue desarrollada por Gary Becker en 1965. La idea fundamental es que el consumo de los bienes requiere tiempo (tiempo para comprar, para prepararlos, para consumirlos, etc) y bajo esta hipótesis es necesario incluir una restricción de tiempo, además de la ya conocida restricción de renta.

El tiempo es considerado como un recurso económico no almacenable del cual todos los individuos están dotados en la misma cantidad. Cada individuo dispone del mismo número de unidades de tiempo por período (24 horas al día, 168 horas a la semana, etc). Los consumidores distribuyen su tiempo en trabajar y realizar otras actividades¹⁰, entendiéndose que la realización de una actividad determinada lleva implícito el consumo de varios bienes de mercado y de tiempo. Tenemos entonces, que el tiempo dedicado a trabajar (T_w) más el tiempo que dedicamos a realizar otras actividades (t_i), debe ser igual

⁹ Una revisión más detallada puede verse en González (1997).

¹⁰ Becker se refiere a estas actividades con el término mercancías básicas (*basic commodities*).

al tiempo total disponible (T). Por tanto, la restricción de tiempo según el modelo de Becker es:

$$\sum_{i=1}^{i=n} t_i q_i + T_w = T \quad (2.5)$$

siendo q_i la cantidad consumida de la mercancía i .

Por su parte, la restricción de renta está compuesta por un lado, por la renta salarial y la renta no salarial (H). La renta no salarial (pensiones, dividendos, pagos de la Seguridad Social, etc) no requiere del consumo de tiempo para su obtención. La renta salarial dependerá del número de horas trabajadas. En este modelo el tiempo de trabajo es endógeno, esto es, el individuo decide libremente el número de horas que dedica a trabajar, lo que significa que puede influir sobre su renta salarial. Por otro lado, tenemos que el individuo dedica su renta a la obtención de aquellos bienes de mercado que le permitan realizar aquellas actividades que desee. Por tanto, la restricción presupuestaria tiene la expresión:

$$\sum_{i=1}^{i=n} P_i q_i = w T_w + H \quad (2.6)$$

siendo

P_i , el precio de la mercancía i ¹¹

w , el salario por hora de trabajo

¹¹ En realidad no existe un mercado donde realizar transacciones con estas mercancías, pero si suponemos que para adquirir la mercancía q_i necesitamos $b_i q_i$ unidades del bien de mercado x_i al precio p_i (es decir $x_i = b_i q_i$) entonces el precio de la mercancía P_i es equivalente a $b_i p_i$.

Al considerar el tiempo que se dedica al consumo, el precio total de la mercancía i no va a ser el precio de mercado, sino éste más el coste de oportunidad del tiempo dedicado a dicho consumo, esto es, el ingreso al que se renuncia debido a que este tiempo se podría dedicar a trabajar. Este precio total del bien se obtiene de despejar las horas de trabajo de la ecuación (2.6) y reemplazar en la ecuación (2.5), de lo que resulta la siguiente restricción que incluye tanto la restricción de tiempo como de renta:

$$\sum_{i=1}^{i=n} (P_i + wt_i) q_i = wT + H \quad (2.7)$$

siendo el precio total de la mercancía i : $P_i + wt_i = P_i^T$ (2.8)

Existe, por tanto, un *compromiso* entre precio y tiempo; precios más altos requieren un tiempo de consumo menor y viceversa. Se pueden describir las elecciones de precio-tiempo a través de una función decreciente que expresa el tiempo de consumo en función del precio de mercado, de manera que el precio total va depender, en definitiva, del precio de mercado. El consumidor minimiza el precio total, decide de forma equilibrada entre el precio de mercado y el coste de oportunidad del tiempo que dedica al consumo de cada bien. La elección de los consumidores está sujeta a su renta salarial; esto es, consumidores de mayores rentas (mayor tasa salarial) estarán dispuestos a pagar mayores precios de mercado para ahorrar en el tiempo que dedican al consumo de los bienes que los consumidores con tasas salariales más bajas. De la condición de primer orden del problema de minimización del precio total se obtiene que el valor del tiempo en el modelo de Becker coincide con la tasa salarial.

El modelo de DeSerpa (1971; 1973) incluye el tiempo dedicado a la realización de actividades y el tiempo de trabajo en la función de utilidad. Además de las restricciones

de renta y tiempo, ya definidas en otros modelos, se incluyen las *restricciones tecnológicas* que expresan la necesidad de un tiempo mínimo para realizar cada actividad, distinguiendo la actividad laboral del resto de actividades. Se define un número mínimo de horas de trabajo con el fin de relajar la hipótesis poco realista acerca de la absoluta flexibilidad que dispone el consumidor para decidir el número de horas que trabaja. De este modo, la formulación del modelo es la siguiente:

$$\begin{aligned}
 & \text{Max} U(X, t, t_w) \\
 & \text{s.a.} \\
 & pX \leq wt_w + H \quad (\lambda) \\
 & \sum_{j=1}^{j=n} t_j + t_w \leq T \quad (\mu) \\
 & t_w \geq t_w^* \quad (\phi) \\
 & t_j \geq t_j^* \quad (\psi_j)
 \end{aligned} \tag{2.9}$$

siendo:

w , salario por hora

t_w , horas de trabajo

H , renta no salarial

T , tiempo total disponible

t_j , el tiempo que dedica a la actividad j

t_w^* , tiempo mínimo de trabajo

t_j^* , tiempo mínimo para la actividad j

A cada restricción se le asocia un multiplicador de *Lagrange*, de manera que λ es la utilidad marginal de la renta, μ la utilidad marginal del tiempo como recurso, ϕ la utilidad marginal por disminuir el número mínimo de horas de trabajo y Ψ_j la utilidad marginal por disminuir el número mínimo de horas dedicadas a la actividad j .

Lo relevante en este modelo es la distinción entre actividades de ocio (*pure leisure activities*) y actividades intermedias (*intermediate activities*). Las actividades de ocio son aquellas a las que el individuo dedica más tiempo del mínimo exigido (la restricción no se cumple con igualdad, por tanto $\Psi_i=0$). Las actividades intermedias son aquellas en las cuales se dedica el tiempo mínimo necesario. En estas actividades nuestra utilidad se ve incrementada cuando se reduce el mínimo tiempo que hay que dedicarles, es decir, $\Psi_j>0$; por tanto la condición de holgura complementaria asegura que la restricción se cumple con igualdad $t_j = t_j^*$.

La valoración marginal del tiempo dedicado a la actividad j es el cociente entre la utilidad marginal del tiempo dedicado a la actividad j y la utilidad marginal de la renta, que es igual a la suma de cuatro elementos: la tasa salarial, la desutilidad/utilidad del trabajo, la disposición a pagar por disminuir las horas de trabajo y la disposición a pagar por ahorrar tiempo en la actividad j que puede ser transferido a actividades de ocio, esto es¹²:

$$\frac{1}{\lambda} \frac{\partial U}{\partial t_j} = w + \frac{1}{\lambda} \frac{\partial U}{\partial t_w} + \frac{\phi}{\lambda} - \frac{\Psi_j}{\lambda} \quad (2.10)$$

¹² Esta expresión se puede obtener a partir de las condiciones de primer orden del problema (2.9)

En el modelo de Becker, los tres últimos términos de la expresión (2.10) son cero porque no hay desutilidad asociada a dedicar tiempo al trabajo, no hay un mínimo de horas de trabajo y no hay restricciones de tiempo mínimo para las distintas actividades que puede realizar el individuo. En este caso, el valor marginal del tiempo dedicado a la actividad j coincide con la tasa salarial, como hemos comentado anteriormente.

De las condiciones de primer orden, también se puede obtener que la valoración marginal del tiempo en la actividad j es igual al valor del tiempo como recurso (μ/λ) menos el valor de ahorrar tiempo en la actividad j (ψ_j/λ), es decir:

$$\frac{1}{\lambda} \frac{\partial U}{\partial t_j} = \frac{\mu}{\lambda} - \frac{\psi_j}{\lambda} \quad (2.11)$$

En las actividades de ocio ($\psi_j = 0$), la valoración marginal del tiempo dedicado a la actividad j coincide con el valor del tiempo como recurso, mientras que para las actividades que son intermedias ($\psi_j \neq 0$), ésta va a ser menor que el valor del tiempo como recurso en la cuantía $\frac{\psi_j}{\lambda}$. Esta magnitud representa el valor del ahorro de una unidad de tiempo en la actividad j , es decir:

$$\frac{\psi_j}{\lambda} = \frac{-\partial U^* / \partial t_j^*}{\partial U^* / \partial I} = \frac{dI}{dt_j^*} \quad (2.12)$$

donde I representa la renta del individuo. En otras palabras, (2.12) representa lo que el individuo está dispuesto a sacrificar de su renta por ahorrar una unidad de tiempo en la actividad intermedia j y poder disfrutar de una unidad adicional de tiempo de ocio. Éste

es el concepto al que convencionalmente se hace referencia como *valor subjetivo del tiempo* en la terminología de evaluación de proyectos de inversión en transporte.

Train y McFadden (1978) plantean un modelo de intercambio entre bienes y ocio donde tratan de analizar cuál es el papel del salario en las decisiones de los consumidores cuando se enfrentan a un problema de elección de modo de transporte y cómo éste varía al considerar distintas formas funcionales en la función utilidad que representa las preferencias individuales.

La función objetivo a la que se enfrenta el consumidor depende del consumo de bienes (G) y del tiempo de ocio disponible (L) sujeto a dos restricciones; la primera establece que el gasto total en bienes consumidos (incluido el gasto en transporte) debe ser igual al total de renta disponible y la segunda, que el tiempo de ocio debe ser igual al total de tiempo disponible una vez descontado el tiempo de dedicado al viaje y al trabajo. El problema de optimización del consumidor es el siguiente:

$$\begin{aligned}
 & \text{Max}U(G_j, L_j) \\
 & \text{s.a.} \\
 & G_j = H + wW - c_j \\
 & L_j = T - W - t_j
 \end{aligned} \tag{2.13}$$

donde t_j y c_j representan el tiempo y coste de viaje, respectivamente, del modo de transporte j . Al sustituir G_j y L_j se obtiene que la función de utilidad depende de las horas de trabajo (W) y el problema se transforma en:

$$\text{Max}_j \left\{ \text{Max}_W U [G_j(W, c_j), L_j(W, t_j)] \right\} \tag{2.14}$$

Al maximizar en W y sustituir en la función objetivo se obtiene la FUIC:

$$U_j = U [G_j(W^*(c_j, t_j), c_j), L_j(W^*(c_j, t_j), t_j)] \quad (2.15)$$

En este modelo, la renta del individuo se determina de forma endógena ya que él elige el número de horas que dedica a trabajar a un salario dado. La forma en que se introduce la variable renta en la función de utilidad va a estar definida por el mapa de curvas de indiferencia entre bienes y ocio del consumidor. Considerando una función de utilidad tipo Cobb-Douglas se obtiene que la tasa salarial puede especificarse ya sea multiplicando al tiempo o bien dividiendo al coste. En el primer caso, se supone implícitamente que las personas con mayor nivel de renta (tasas salariales más altas) perciben más las pérdidas de tiempo en su viaje al trabajo que las personas con menor nivel de renta (tasas salariales más bajas); mientras que en el segundo caso, implícitamente se supone que el trabajador con alto nivel de renta es menos consciente del coste del viaje al trabajo que las personas con bajo nivel de renta. Una consecuencia que se deriva de esta formulación es que tanto la utilidad marginal de la renta como el valor subjetivo del tiempo de viaje son valores constantes, coincidiendo este último con la tasa salarial.

Jara-Díaz y Farah (1987) reformulan el modelo planteado por Train y McFadden (1978) argumentando que, por un lado, la renta no debe determinarse de forma endógena sino exógena debido a que muchos individuos no pueden decidir fácilmente el número de horas que trabajan ni obtener un salario adicional por trabajar más horas de las establecidas en su jornada laboral; y por otro lado, que la proporción gastada en transporte, especialmente para viajes recurrentes, no es despreciable.

Partiendo del modelo de Train y McFadden (1978) obtienen un valor del tiempo que depende directamente de lo que denominan *tasa de gasto* (g), que no es más que la renta

por unidad de tiempo disponible. Si el número de horas de trabajo es fijo, lo importante es el tiempo disponible para gastar nuestra renta y en este caso, el efecto renta es relevante.

Bates y Roberts (1986) proponen una adaptación del modelo de DeSerpa al caso de la elección de alternativa de transporte, diferenciando explícitamente el tiempo dedicado al transporte del dedicado a realizar otras actividades en su formulación del problema del consumidor.

Así definen una función de utilidad que depende de la cantidad consumida de un bien generalizado (X), del tiempo dedicado a realizar una actividad generalizada (q) y del tiempo empleado en cada una de las alternativas de transporte disponibles ($t_j \forall j=1,2,\dots,n$), siendo las elecciones entre alternativas mutuamente excluyentes, como por ejemplo la elección de ruta o de modo de transporte para un viaje específico. Para este modelo, el problema de optimización será:

$$\begin{aligned}
 & \text{Max} U(X, q, t_1, \dots, t_n) \\
 & \text{s.a.} \\
 & pX + \sum \delta_j c_j \leq y \quad (\lambda) \\
 & q + \sum \delta_j t_j \leq T \quad (\mu) \\
 & t_j \geq t_j^* \quad (\psi_j)
 \end{aligned} \tag{2.16}$$

Las restricciones son las ya conocidas, de renta y tiempo, añadiendo respectivamente a cada una de ellas el coste (c_j) y el tiempo (t_j) de las distintas alternativas. Se define una variable ficticia (δ_j) que toma valor 1 si la alternativa j es elegida y 0, en otro caso. Y finalmente, una tercera restricción que define un tiempo mínimo para cada alternativa.

Sustituyendo las condiciones de primer orden del problema (2.16) en la aproximación local de primer orden de la función de utilidad, obtenemos una especificación lineal de la FUIC de la alternativa j , de manera que si solo tenemos en cuenta las variables que dependen de la alternativa, la FUIC¹³ toma la siguiente expresión:

$$V_j = -\lambda c_j - \psi_j t_j \quad (2.17)$$

La disposición a pagar por ahorrar tiempo en la alternativa j vendrá dada por el cociente de las utilidades marginales del tiempo y el coste $\left(\frac{\psi_j}{\lambda}\right)$ ¹⁴, que son los parámetros del tiempo y coste respectivamente en la función de utilidad V_j . Este es el caso concreto de una función lineal en los parámetros y en las variables; no obstante, si se consideran aproximaciones locales de la función de utilidad de orden superior se pueden obtener especificaciones no lineales que incluyan interacciones entre variables como las señaladas en Train y McFadden (1978) y Jara-Díaz y Farah (1987). En estos casos, la obtención de la disposición a pagar no corresponde con el cociente de los parámetros del tiempo y coste, sino que debe obtenerse a partir de la expresión (2.4).

2.2.3 La Teoría de la utilidad aleatoria

Cuando abordamos el problema de modelización de la demanda en un contexto de alternativas discretas, es preciso tener en cuenta los aspectos relevantes que rodean el proceso de toma de decisiones de los individuos. Entre los más importantes cabe señalar los siguientes: definición de la unidad que toma decisiones, determinación del conjunto

¹³ Es la parte representativa de la función de utilidad, que comentaremos más adelante.

¹⁴ Obsérvese que este valor coincide con el valor del ahorro de tiempo en la actividad j en el modelo de DeSerpa.

de elección, selección de los atributos que explican la utilidad de cada alternativa j y determinación de la regla de decisión que sigue la unidad tomadora de decisiones.

El individuo es la unidad típica de toma de decisiones, pero alternativamente también puede considerarse a un grupo, como la familia o el hogar, a una empresa o institución pública. Cada uno de ellos puede representar una única unidad de decisión. Es importante señalar que aunque estudiemos la predicción de la demanda agregada, hay que tener en cuenta las diferencias entre los individuos en el proceso de decisión, debido principalmente a la existencia de distintos patrones de gustos.

Dado un conjunto universal de alternativas (C), el conjunto de elección de un individuo n (C_n) está determinado por aquellas que éste tiene disponibles; a su vez, este es un subconjunto del conjunto universal ($C_n \subset C$) y representa la región factible en el problema de optimización del consumidor, definida a partir de las distintas restricciones del individuo.

En un contexto de elección modal, el conjunto universal está determinado por todos los modos de transporte existentes para realizar un viaje entre el par origen-destino considerado. Por ejemplo, si los modos existentes son coche, bus y metro, un individuo particular podría no tener disponible alguna de las alternativas (por ejemplo, si no tiene permiso de conducir no tendría disponible la alternativa coche como conductor; si la parada de metro más cercana está a una distancia superior a un kilómetro probablemente no sería factible para él ir en metro, etc) y esto debe ser tomado en cuenta a la hora de determinar su conjunto de elección.

Identificadas las distintas alternativas, cada una de ellas es evaluada en términos de un vector de atributos que representan sus características. Así, por ejemplo, los atributos

para el coche pueden ser tiempo, coste y confort; para el bus, tiempo, coste, frecuencia y confort, y para el metro, tiempo, coste, frecuencia y confort. Para cada alternativa, los distintos atributos toman diferentes valores. El individuo elige, dados los valores de los atributos, aquella alternativa que considera mejor, estableciendo una regla de decisión que describe el mecanismo interno utilizado para procesar la información disponible y realizar la elección. Los procesos de decisión pueden ser compensatorios, esto es, cambios en uno o más atributos pueden ser compensados por cambios en otros; o no compensatorios, donde los buenos atributos de una alternativa no permiten compensar los malos. Los procesos de naturaleza compensatoria definen una función objetivo (normalmente, una función de utilidad) que exprese el atractivo de la alternativa en términos de sus características. Está basado en el *compromiso* entre atributos, pues al elegir la alternativa con la utilidad más alta, el individuo selecciona aquella que ofrece una combinación óptima de los mismos. Este tipo de modelos no son apropiados para los procesos de decisión caracterizados por discontinuidades, las cuales son mucho mejor expresadas a través de los procesos no compensatorios. Por su parte, estos procesos definen normas o umbrales que restringen el campo de alternativas factibles, como por ejemplo que el coste del viaje no exceda de una determinada cantidad, así se define el coste del viaje máximo. Estas restricciones son consideradas por los individuos en sus procesos de decisión. Se distinguen la elección por eliminación y la satisfacción.

En la elección por eliminación el individuo jerarquiza los atributos (por ejemplo, el tiempo de viaje es más importante que el coste, el coste que el tiempo de espera, etc.) y define un valor mínimo aceptable o umbral para cada uno de ellos. Posteriormente, se jerarquizan las alternativas de mejor a peor en relación al atributo, por ejemplo que el tiempo de viaje no sea mayor de veinte minutos, y todas aquellas alternativas que no satisfagan esta restricción son eliminadas y así sucesivamente hasta quedarnos con una

sola alternativa. Es posible que repetido el proceso tengamos varias alternativas que cumplan todas las restricciones establecidas, en este caso se elige una de manera compensatoria (ver Tverski, 1972).

Por su parte, la satisfacción presenta varias maneras de búsqueda de la estrategia de elección. Por ejemplo, podría tratarse de un proceso cíclico donde los umbrales se modifican secuencialmente hasta encontrar una alternativa. Asimismo, se podría predefinir una regla de manera que la búsqueda se alcance en el punto en el que se satisface la regla. En este caso, no todos los atributos o alternativas tienen que ser considerados.

De los dos procesos de decisión comentados, el que ha sido más utilizado para explicar el comportamiento de elección del individuo es el compensatorio. No obstante, en algunos casos el individuo podrá no considerar este proceso como criterio de elección, y en estas situaciones se analizará si las preferencias del individuo descritas por sus elecciones cumplen con los axiomas que establece la teoría del consumidor. Este tema será tratado más adelante.

La teoría en que se sustenta el comportamiento del consumidor parte del supuesto que los individuos son racionales (*homo economicus*) y su toma de decisiones se basa en alcanzar unos objetivos dadas unas restricciones. En el contexto de las elecciones discretas, estos objetivos se alcanzan maximizando la FUIIC sobre el conjunto de elección del individuo.

La teoría de la utilidad aleatoria¹⁵ constituye el cuerpo teórico que permite tratar empíricamente el problema de las elecciones discretas. A la hora de modelizar las decisiones de los individuos y obtener empíricamente la FUIC, el investigador no dispone de información perfecta; es decir, no puede observar todos los factores que participan en el proceso de elección. Puede darse la situación en que dos individuos con el mismo conjunto de alternativas disponibles, sujetos a las mismas condiciones para realizar un viaje concreto e idénticas características socioeconómicas, elijan distintas alternativas. Para resolver este problema, se acude a la teoría probabilística con la idea de tener en cuenta el efecto de factores de naturaleza aleatoria y de este modo corregir las percepciones incorrectas del analista. Así pues, se define una función de utilidad que presenta dos componentes: una parte observable, determinada por el vector de características o atributos medibles de la alternativa y el vector de características socioeconómicas del individuo; y una componente no observable o aleatoria, definida para incorporar las distintas fuentes de aleatoriedad que han sido identificadas; en primer lugar, cualquier atributo no observado. El vector de atributos de las alternativas es normalmente incompleto, no es posible medir el conjunto total de elementos que el individuo considera en sus decisiones. Otra fuente de aparente aleatoriedad es la variación en los gustos no observada; la función de utilidad puede tener elementos no observados que varían entre individuos. En tercer lugar, los errores en la medición de los distintos atributos y en cuarto lugar, elementos de la función de utilidad no observables que se reemplazan por otros elementos que incluyen variables *proxy*.

¹⁵ Formalizada por McFadden (1974), Domencich y McFadden (1975) y Manski (1977). Desarrollos más recientes se pueden encontrar en Ortúzar y Willumsen (2001).

De este modo, la utilidad de una alternativa está representada por una variable aleatoria que se descompone en la suma de dos componentes: la componente observable (V_{jn}) y la no observable (ε_{jn}), según se establece en la siguiente expresión:

$$U_{jn} = V(Q_{jn}, S_n) + \varepsilon(Q_{jn}, S_n) = V_{jn} + \varepsilon_{jn} \quad (2.18)$$

Tanto la componente observable como la no observable dependen de los atributos de las alternativas y de las características socioeconómicas del individuo.

Al ser la utilidad una variable aleatoria, el individuo n elegirá la alternativa j en su conjunto de opciones disponibles, C_n , con probabilidad igual a:

$$P(j|C_n) = \Pr[V_{jn} + \varepsilon_{jn} \geq V_{in} + \varepsilon_{in}, \forall i \in C_n] \quad (2.19)$$

$$P(j|C_n) = \Pr[\varepsilon_{in} - \varepsilon_{jn} \leq V_{jn} - V_{in}, \forall i \in C_n] \quad (2.20)$$

Para obtener un modelo de utilidad aleatoria específico, es necesario suponer una distribución de probabilidad conjunta para el término de perturbación aleatoria. En función de las distintas hipótesis formuladas acerca de la distribución de los errores se podrán derivar distintos modelos de elección discreta.

En la especificación de la parte observable¹⁶ se distinguen dos tipos de variables: en primer lugar, los atributos que caracterizan a la alternativa j para el individuo n , esto es, tiempo de viaje, coste del viaje, confort, seguridad, etc. Y en segundo lugar, el vector de variables socioeconómicas que caracterizan al individuo n como la renta, la posesión o no de vehículo, el tamaño del hogar, edad, sexo, ocupación, nivel de estudios, etc. Dos

¹⁶ También se suele denominar utilidad sistemática, determinística o medible.

cuestiones se derivan una vez determinadas las variables a considerar. La primera, que constituye el principal problema en la especificación de la parte determinística de la función de utilidad, está relacionada con la determinación de una combinación de atributos y variables socioeconómicas que reflejen una hipótesis razonable sobre el efecto de tales variables en la utilidad. Mientras que la segunda consiste en determinar la forma funcional. En este sentido, resulta interesante trabajar con especificaciones de funciones que, por un lado, definan una relación entre las variables consideradas y que a su vez estén basadas en supuestos teóricos sobre el comportamiento de los individuos, establecidos por el investigador con base en el conocimiento del mercado objetivo que se analiza; por otro lado, que la especificación presente buenas propiedades en el proceso de estimación. Normalmente, se utilizan funciones que sean lineales en los parámetros, donde las variables pueden entrar en forma polinómica, o empleando cualquier otra transformación que permita relacionar los atributos con la función de utilidad. Por otro lado, las variables pueden ser genéricas o específicas; son genéricas cuando el efecto marginal es el mismo para todas las alternativas y específicas cuando la misma variable presenta efectos diferentes en función de la alternativa que se trate. El problema radica en estudiar cuál debe ser la mejor especificación de la parte observable de la función de utilidad dadas las variables y las características del mercado objetivo.

2.3 La teoría de la elección discreta

2.3.1 Principales modelos de elección discreta

Partiendo de la teoría de la utilidad aleatoria se pueden obtener distintos modelos de elección discreta suponiendo una distribución de probabilidad conjunta para el término de la perturbación aleatoria.

Si consideramos la hipótesis de que los errores distribuyen idéntica e independientemente Gumbel con media cero y varianza σ^2 , obtenemos el modelo Logit Multinomial o Logit Simple (MNL). De acuerdo a este modelo, la expresión de la probabilidad de elección de la alternativa j para el individuo n (McFadden, 1974) es:

$$P_{jn} = \frac{e^{\beta V_{jn}}}{\sum_{A_i \in C_n} e^{\beta V_{in}}} \quad (2.21)$$

siendo:

$$\beta = \frac{\pi}{\sigma\sqrt{6}} \quad (2.22)$$

El parámetro β no se puede estimar por separado del resto de los parámetros definidos en la FUIC (V_{in}) y para muchos propósitos se puede suponer, sin pérdida de generalidad, que β toma valor uno.

El modelo MNL descansa sobre el supuesto de independencia de alternativas irrelevantes. Este se expresa prácticamente en que el cociente entre las probabilidades de elección de dos alternativas i y j es constante y sólo depende de las utilidades de ambas alternativas; así, es independiente de las utilidades del resto de las alternativas.

La expresión de esta propiedad es fácil de obtener a partir de (2.21):

$$\frac{P_{in}}{P_{jn}} = e^{\beta(V_i - V_j)} \quad (2.23)$$

En un principio, esta propiedad tuvo su importancia ya que permitía analizar la demanda de nuevas alternativas sin necesidad de reestimar el modelo, bastaba simplemente con conocer sus atributos. El problema se presenta cuando existen alternativas

correlacionadas (por ejemplo, coche como conductor y coche como acompañante, o distintos modos de transporte público como metro y tren), porque el modelo conduce a predicciones sesgadas. Una ilustración interesante de este problema puede verse en Mayberry (1973).

Este modelo es homocedástico, esto es, presenta la misma varianza para todas las alternativas y no permite medir correlación entre ellas, de manera que la matriz de varianza y covarianza presenta forma diagonal:

$$\Sigma = \sigma^2 \begin{pmatrix} 1 & \dots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \dots & 1 \end{pmatrix} \quad (2.24)$$

Otros modelos de elección discreta permiten considerar correlación entre las variables e incluso la existencia de heterocedasticidad¹⁷. Conforme más general es la matriz de covarianza, más compleja es la implementación del modelo y su estimación. El modelo Logit Jerárquico (HL) o Anidado (*hierarchical, nested o tree logit*) resuelve en parte el problema de la independencia de alternativas irrelevantes, de manera que es posible considerar cierto patrón de correlación entre alternativas agrupándolas en jerarquías o nidos (Williams, 1977; Daly y Zachary, 1978). Cada nido o jerarquía es representado por una alternativa compuesta frente al resto de las alternativas disponibles. Un ejemplo de una estructura jerárquica o anidada puede ser la de la Figura 2.1.

La obtención del modelo HL se basa en los siguientes supuestos:

¹⁷ Distinta varianza entre alternativas.

Se agrupa en jerarquías o nidos a todos los subconjuntos de opciones correlacionadas entre sí, es decir, aquellas que presentan ciertas semejanzas.

Cada jerarquía o nido es representado por una alternativa compuesta frente a las demás que están disponibles a la población.

La estimación sigue un proceso secuencial donde se estima primero un MNL para aquellas opciones más similares que se han agrupado en un nido, por ejemplo para las alternativas A_3 y A_4 del nido inferior de la figura 2.1, omitiendo todas aquellas variables¹⁸ que tengan el mismo valor para este subconjunto de opciones, $A^I(q)$. La necesidad de omitir estas variables se debe al hecho de que el MNL trabaja en diferencias. Éstas son introducidas posteriormente en el nido superior, puesto que afectan a la elección entre la jerarquía inferior y el resto de opciones en $A^S(q)$.

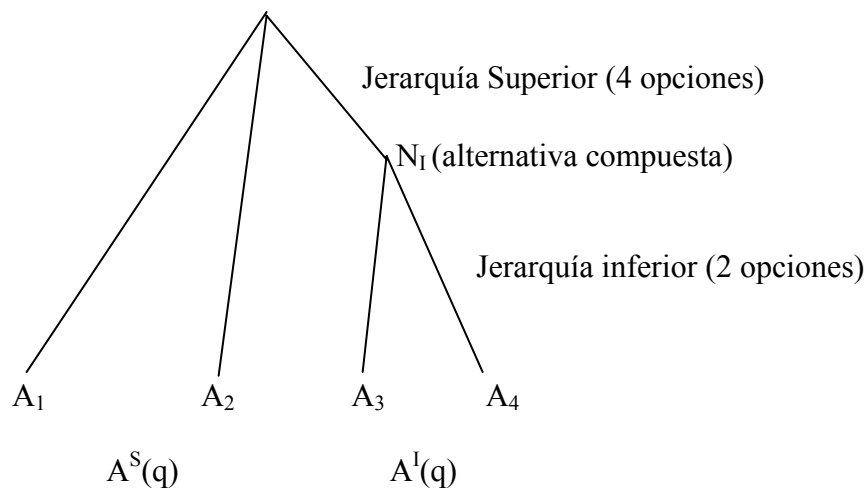


Figura 2.1: Estructura del Logit Jerárquico

La forma en que se introduce el nido inferior en la jerarquía superior es a través de la alternativa compuesta N_1 , a la cual se le asocia una utilidad representativa de todo el

¹⁸ Denominadas como vector W .

nido, que va a tener dos componentes (ver Sobel, 1980; Ortúzar, 1983); la primera considera como variable la utilidad máxima esperada de las alternativas del nido (EMU, *expected maximum utility*):

$$EMU = \ln \sum_{A_j \in A^I(q)} \exp(V_j) \quad (2.25)$$

siendo V_j la utilidad de la alternativa j del nido, en la cual no se tiene en cuenta el vector de variables W , es decir, aquellas que toman el mismo valor para las alternativas del subconjunto de alternativas $A^I(q)$.

Y la segunda que considera el vector de los atributos comunes a todos los componentes del nido, W . De este modo, la utilidad compuesta del nido viene dada por:

$$\tilde{U}_j = \phi EMU + \alpha W \quad (2.26)$$

donde ϕ y α son parámetros a estimar.

Realizado todo el proceso de estimación para el nido inferior, se estima un MNL para la jerarquía superior donde se incluye la alternativa compuesta y el resto de alternativas contenidas en el nido superior.

De acuerdo con este modelo, la probabilidad de que un individuo n elija la alternativa del nido $A_j \in A^I(q)$ es igual a la probabilidad de que elija la alternativa compuesta (en el nido superior) por la probabilidad condicional de que opte por A_j , en la jerarquía inferior, dado que escogió la alternativa compuesta.

En el caso del ejemplo mostrado en la Figura 2.1 donde hay dos niveles jerárquicos, esta probabilidad sería:

$$P_{jn} = \frac{\exp(\phi EMU + \alpha W + \theta^S Z^S)}{\sum_{A_i \in A^S(q)} \exp(\phi EMU + \alpha W + \theta^S Z^S)} \cdot \frac{\exp(\theta^I Z_j^I)}{\sum_{A_i \in A^I(q)} \exp(\theta^I Z_i^I)} \quad (2.27)$$

denotando los superíndices I y S (como en la Figura 2.1) los niveles jerárquicos inferior y superior, respectivamente; Z^I y Z^S los vectores de los atributos de las alternativas disponibles en ambos niveles y θ , α y ϕ , los parámetros a estimar.

La consistencia interna requerida por el modelo establece, para el caso de dos niveles jerárquicos¹⁹, que:

$$0 < \phi \leq 1 \quad (2.28)$$

Si ϕ es igual a 1, el modelo HL es matemáticamente equivalente al MNL. Si ϕ es menor que 0 significa que el aumento de la utilidad de una alternativa del nido disminuirá la probabilidad de elegir dicha opción. Si ϕ es igual a 0 significa que el aumento de la utilidad de una alternativa del nido no alteraría la probabilidad de elección de la alternativa. Si ϕ es mayor que 1, el aumento de la utilidad de una alternativa del nido aumentaría tanto su probabilidad de elección como la del resto de las alternativas del nido.

Si el número de jerarquías o nidos definidos es superior a 2, la consistencia interna del modelo vendría definida de la siguiente manera:

$$0 < \phi_1 \leq \phi_2 \leq \dots \leq \phi_s \leq 1 \quad (2.29)$$

siendo ϕ_I el parámetro correspondiente al nido del nivel inferior y ϕ_S al nido de nivel superior de cada rama del árbol (Ortúzar y Willumsen, 2001).

¹⁹ El considerado en la Figura 2.1 y en la definición de la probabilidad de elección de la alternativa j para el individuo n .

La matriz de covarianza para este modelo para el caso que hemos supuesto en nuestro planteamiento será:

$$\Sigma = \begin{pmatrix} \sigma^2 & \dots & \mathbf{0} \\ \vdots & \ddots & \sigma_{34}^2 \\ \mathbf{0} & \sigma_{43}^2 & \sigma^2 \end{pmatrix} \quad (2.30)$$

Al igual que en el modelo MNL, se trata de un modelo homocedástico y presenta correlación entre las alternativas 3 y 4 que están agrupadas en un mismo nido (la covarianza entre la alternativa 3 y 4 es distinta de cero y $\sigma_{34} = \sigma_{43}$). Para una explicación más clara de este modelo ver Carrasco y Ortúzar (2002).

A pesar de que el modelo HL permite resolver, parcialmente, el problema de la correlación entre alternativas y se presenta como un modelo más completo que el MNL, éste posee aún algunas limitaciones (Ortúzar y Willumsem, 2001):

No permite considerar variaciones en los gustos, al igual que el MNL, ni funciones de utilidad heterocedásticas.

No permite tratar los problemas de correlación entre alternativas pertenecientes a distintos nidos; sólo considera las interdependencias entre opciones de los nidos que se haya especificado en la estructura jerárquica a modelizar.

Las posibles estructuras jerárquicas aumentan conforme aumenta el número de alternativas en proporción no lineal. Sobel (1980) encuentra 26 estructuras posibles para un conjunto de cuatro alternativas. No obstante, el problema combinatorio se reduce al estudiar aquellas estructuras que teórica o intuitivamente son más razonables.

Uno de los problemas potenciales de la estimación del modelo por el método secuencial es la transmisión de errores y distinto valor de los parámetros en los distintos nidos (ver Ortúzar *et al.*, 1987). Sin embargo, éstos se han podido solucionar gracias a la aparición de distintos programas²⁰ que permiten la estimación simultánea del modelo.

El modelo Probit Multinomial (MNP) se obtiene de suponer una distribución de probabilidad conjunta para el término de perturbación aleatoria Normal Multivariada con media cero y matriz de covarianza arbitraria (Daganzo, 1979). En este sentido, el MNP permite considerar heterocedasticidad y correlación entre los distintos términos del error. La posibilidad de poder definir cualquier matriz de covarianzas otorga gran flexibilidad y le confiere al modelo MNP la capacidad para estudiar una gama más amplia de comportamientos.

Una cuestión importante de este modelo es la dificultad a la hora de establecer la función de verosimilitud debido a que no existe una expresión cerrada para la probabilidad de elección de las alternativas. Para más de tres opciones, no es posible, analíticamente, estimar el modelo y hay que solventar este problema empleando procedimientos de simulación.

Los problemas observados en la estimación de modelos MNP derivaron en su no utilización por muchos años, y también en la búsqueda de modelos flexibles que fueran más fácilmente tratables desde el punto de vista computacional. Así surgió el modelo Logit Mixto (ML) que presenta una flexibilidad igual a la del MNP y una implementación computacional marginalmente menos compleja. No obstante, avances en las técnicas de simulación y algoritmos de búsqueda, así como en poder de los

²⁰ Software: ALOGIT (Daly, 1987; 1992); LIMDEP (Economic Software, Inc. 1995); HieLoW (Bierlaire y Vandevyvere 1995); GAUSS (Aptech Systems 1994).

ordenadores, hacen que ambos modelos sean factibles de implementar en la práctica; de hecho, actualmente el mayor problema es la mayor dificultad asociada a la interpretación de los resultados.

McFadden y Train (2000) demuestran que el modelo ML puede aproximar cualquier modelo de utilidad aleatoria. Entre sus principales ventajas, tenemos que, con respecto al MNL, permite estudiar la variación en los gustos, patrones de sustitución no restrictivos y correlación entre elementos no observados en el tiempo. Y con respecto al MNP, que no está restringido a la distribución Normal; de hecho es posible definir otras distribuciones²¹ de probabilidad y es algo más sencilla la simulación de sus probabilidades de elección.

La definición del ML está basada en la forma funcional de sus probabilidades de elección. Las probabilidades de elección son las integrales de una probabilidad logit estándar, donde los parámetros no tienen por qué ser fijos y se comportan según una determinada distribución de probabilidad caracterizada por su función de densidad. De este modo:

$$P_{jn} = \int S_{jn}(\beta) f(\beta) d\beta \quad (2.31)$$

siendo

$$S_{jn}(\beta) = \frac{e^{V_{jn}(\beta)}}{\sum_{i=1}^I e^{V_{in}(\beta)}} \quad (2.32)$$

²¹ Revelt y Train (1998) y Ben-Akiva y Bolduc (1996) han especificado distribuciones Normal o Lognormal, mientras que Revelt y Train (2000), Hensher y Greene (2001), Sillano y Ortúzar (2002) y Train (2001) han utilizado además distribuciones Uniformes y Triangulares.

la probabilidad logit evaluada para el parámetro β , $f(\beta)$ la función de densidad y V_{jn} la parte determinística de la función de utilidad.

Partiendo de un modelo de comportamiento maximizador de utilidad podemos derivar un modelo ML que permita estudiar variación en los gustos, definiendo un ML de parámetros aleatorios. Para ello se especifica que los parámetros de las variables explicativas de la función de utilidad para cada una de las alternativas varíen según una distribución de probabilidad. Si queremos definir un modelo que considere la existencia de correlación entre las funciones de utilidad de las distintas alternativas, formularemos un modelo de componentes de error. En este modelo, alguno de los parámetros de las variables explicativas de la función de utilidad son fijos y otros aleatorios. Esto da lugar a que el error aleatorio tenga dos componentes: un término que distribuye iid Gumbel y uno adicional que distribuye con media cero.

La utilidad de la alternativa j para el individuo n en el modelo de parámetros aleatorios está dada por:

$$U_{jn} = \beta_n x_{jn} + \varepsilon_{jn} \quad (2.33)$$

siendo

ε_{jn} , el término de perturbación aleatoria que distribuye iid Gumbel.

x_{jn} , el conjunto de las variables explicativas de la alternativa j .

β_n , un vector de coeficientes, el cual sigue una determinada distribución, $\beta \sim f(b, \Omega)$ que depende de los parámetros b y Ω (generalmente, vector de medias y matriz de varianza y covarianza).

Por su parte, la función de utilidad de la alternativa j para el modelo de componentes de error tiene la expresión:

$$U_{jn} = \alpha x_{jn} + \mu_n z_{jn} + \varepsilon_{jn} \quad (2.34)$$

en que

α , es el vector de parámetros de las variables explicativas de la alternativa j , que es un vector de coeficientes fijos.

x_{jn} y z_{jn} , son vectores de variables observadas.

μ_n , vector de errores aleatorios con media cero.

ε_{jn} , vector de errores que distribuye iid Gumbel.

El modelo de parámetros aleatorios es equivalente al modelo de componentes de error cuando el vector de parámetros β se descompone en su media α más las desviaciones μ_n ($\beta_n = \alpha + \mu_n$). De forma recíproca, si $z_{jn} = x_{jn}$ el modelo de componentes de error es equivalente a un modelo de parámetros aleatorios; y si $z_{jn} \neq x_{jn}$ se obtendría un modelo de parámetros aleatorios con coeficientes fijos para x_{jn} y coeficientes aleatorios con media cero para z_{jn} .

2.3.2 Procedimientos de estimación

Existen diversos procedimientos para estimar modelos de elección discreta. Los que se emplean con más frecuencia son el método de máxima verosimilitud y el método de

máxima verosimilitud simulada. Los modelos MNL y HL se estiman por el primer método, mientras que los modelos MNP²² y ML se estiman por el segundo. Además de estos métodos, es posible aplicar técnicas Bayesianas para obtener parámetros individuales en modelos de tipo ML.

El método de estimación de máxima verosimilitud está basado en que una determinada muestra puede ser obtenida de varias poblaciones distintas; existe una probabilidad mayor de que sea generada por una cierta población que por otras. La idea del método de estimación se puede ilustrar considerando una muestra de n observaciones de una variable $Z=\{Z_1, Z_2, \dots, Z_n\}$ representativa de una población caracterizada por un vector de parámetros θ , por ejemplo, su media y varianza. La variable Z es una variable aleatoria que tiene asociada una función de densidad, $f(Z,\theta)$, que depende de los valores de los parámetros θ . Siendo independientes en la muestra los valores de la variable Z , podemos escribir la función de densidad conjunta:

$$f(Z_1, Z_2, \dots, Z_n | \theta) = f(Z_1 | \theta) \cdot f(Z_2 | \theta) \cdot \dots \cdot f(Z_n | \theta) \quad (2.35)$$

donde estadísticamente se interpreta que f es una función con Z variables y un vector de parámetros θ fijo. Considerando f como función de θ en (2.35) se obtiene la función de verosimilitud $L(\theta)$. Al maximizar con respecto a θ , se obtienen los estimadores por máxima verosimilitud. Estos representan los parámetros que reproducen con mayor probabilidad la muestra observada (Ortúzar, 2001).

²² Existen otros métodos de estimación como la integración numérica o la aproximación de Clark, véase Daganzo (1979).

En el caso de una muestra de N elecciones discretas independientes, la función de verosimilitud está representada por el producto de las probabilidades de que cada individuo elija la alternativa efectivamente seleccionada:

$$L(\theta) = \prod_{n=1}^N \prod_{A_j \in A(n)} (P_{jn})^{g_{jn}} \quad (2.36)$$

siendo

$$g_{jn} = \begin{cases} 1, & \text{si } A_j \text{ es elegida por } n \\ 0, & \text{en otro caso} \end{cases} \quad (2.37)$$

Es habitual maximizar el logaritmo natural de la función de verosimilitud $L(\theta)$, $l(\theta)$, porque se obtiene el mismo resultado y es más manejable. La nueva función es:

$$l(\theta) = \log L(\theta) = \sum_{n=1}^N \sum_{A_j \in A(n)} g_{jn} \log P_{jn} \quad (2.38)$$

El conjunto de parámetros obtenidos al maximizar la log-verosimilitud sigue una distribución asintóticamente Normal, $N(\theta, \Sigma^2)$, siendo la expresión de la varianza:

$$\Sigma^2 = \frac{-1}{E\left(\frac{\partial^2 l(\theta)}{\partial \theta^2}\right)} \quad (2.39)$$

La obtención de los parámetros implica la aplicación de procedimientos numéricos iterativos. Cuando la función de utilidad es lineal en los parámetros, el método converge rápidamente y siempre existe un máximo único. Esto es una ventaja para la estimación del modelo pues es fácil disponer de software apropiado.

Una cuestión a destacar es que no se puede utilizar como medida de bondad del ajuste el R^2 del método de mínimos cuadrados ordinarios, que está basado en los residuos estimados, porque si el MNL especifica constantes específicas para cada alternativa, se puede asegurar que siempre reproduciría las cuotas de mercado de cada una de ellas. La constante específica recoge el efecto de variables que no han sido consideradas en la especificación del modelo a estimar. Así, no es apropiado comparar la suma de probabilidades de elegir una alternativa con el número total de observaciones de las alternativas que fueron elegidas como medida de bondad del ajuste porque esta condición siempre se cumple si el MNL especifica constantes para todas las alternativas.

El método de estimación por máxima verosimilitud simulada fue propuesto por Lerman y Manski (1981) para la estimación del modelo MNP. Se basa en evaluar la probabilidad de elección generando un número de simulaciones de una distribución Nomal Multivariada. Se considera un éxito cuando U_i es el valor más alto. La proporción de éxitos se aproxima a la probabilidad de elección para un número de simulaciones suficientemente grande. Este método presentaba algunos problemas en la práctica que se han podido resolver en los últimos años gracias a los avances en los procesos de simulación en la modelización de las elecciones discretas. El enfoque propuesto por McFadden (1989) no precisa evaluar la integral múltiple sustituyendo la probabilidad de elección en la ecuación de momentos por un simulador insesgado. El método de momentos simulados puede considerarse el precursor del modelo ML.

Los dos problemas de este método en la práctica eran:

Si el número de éxitos fuera igual a cero, la log verosimilitud tendería a infinito y el método colapsa. Para resolver este problema Lerman y Manski (1981) recomiendan

sustituir el ratio del número de éxitos (N_i) sobre el número total de simulaciones (N) por el cociente $(N_i+I)/(N+J)$ siendo J el número de alternativas; sin embargo, esta modificación introduce sesgo.

El error relativo asociado con el método de simulación es inversamente proporcional a la raíz cuadrada del número de éxitos. Esto implica hacer más simulaciones lo que se traduce en un método más caro.

Para el modelo ML, tenemos que la probabilidad de elección de la alternativa j para el individuo n tiene la expresión general (2.31). Si consideramos que existe variación en los gustos, los parámetros β_n distribuyen con cierta media y covarianza de manera que la probabilidad de elección (2.31) depende de dos tipos de parámetros; por un lado, del vector de coeficientes que representan los gustos en la población, β_n y por otro, de los parámetros (τ^*) asociados a la distribución de estos gustos en la población ($f(\beta_n | \tau^*)$).

La probabilidad de la secuencia de elecciones de un individuo n es:

$$P_n(\tau^*) = \int S_n(\beta_n) f(\beta_n | \tau^*) d\beta_n \quad (2.40)$$

donde S_n representa la probabilidad conjunta de elección del individuo n en distintas situaciones de elección t dado un conjunto de parámetros β_n fijo.

$$S_n(\beta_n) = \prod_{it} [P_{jnt}]^{g_{jnt}} = \prod_{it} \left[\frac{e^{V_{int}(\beta_n)}}{\sum_{it} e^{V_{jnt}(\beta_n)}} \right]^{g_{int}} \quad (2.41)$$

siendo $g_{int} = \begin{cases} 1, & \text{si en la secuencia } t \text{ elige } i \\ 0, & \text{en otro caso} \end{cases}$

y $f(\beta_n | \tau^*)$ la función de densidad de β_n .

y la función de verosimilitud tiene la expresión:

$$L(\tau^*) = \prod_n P_n(\tau^*) \quad (2.42)$$

La función de log-verosimilitud:

$$l(\tau^*) = \sum_n \ln P_n(\tau^*) \quad (2.43)$$

La integral (2.40) no puede resolverse analíticamente y, por tanto, no es posible maximizar la función de verosimilitud. Esta probabilidad se aproxima por simulación y se maximiza la función de log-verosimilitud simulada. El procedimiento de simulación es el siguiente:

Dar valores iniciales a la media y a la covarianza, τ^* ; dados estos valores se simulan los valores de β_n a partir de su distribución, $f(\beta_n/\tau^*)$.

Conocidos los valores de β_n , que han sido obtenidos en la primera etapa, se calcula la probabilidad de la secuencia de elección para el individuo n , $S_n(\beta_n)$.

Este proceso se repite para varias simulaciones de β_n y el promedio se toma como aproximación de la probabilidad de elección:

$$PS_n(\tau^*) = \frac{1}{R} \sum_{r=1}^R S_n(\beta_n^r) \quad (2.44)$$

siendo R el número de repeticiones y β_n^r la simulación r -ésima de la distribución f .

El estimador que se obtiene por este procedimiento de simulación es consistente y asintóticamente Normal bajo condiciones de regularidad.

También se puede utilizar el método de estimación Bayesiana (HB); éste requiere dos distribuciones: una *a priori* y otra *a posteriori*. La distribución *a priori* añade información sobre los valores de los parámetros (β) del modelo y está caracterizada por su función de densidad $P(\beta)$; y representa el conocimiento previo que se tiene sobre los parámetros antes de la obtención de los datos, esto es, las elecciones observadas. Además se obtiene información adicional, concretamente una muestra de elecciones independientes para un conjunto de individuos. Así tendremos un conjunto de T elecciones para cada individuo n , $y_n = \{y_{n1}, \dots, y_{nT}\}$ y el conjunto de elecciones de la muestra total será $Y = \{y_1, \dots, y_N\}$. Con este conjunto Y de elecciones se actualiza la información *a priori* obteniéndose, una nueva función de densidad $P(\beta|Y)$ que caracteriza a la distribución *a posteriori*. Esta es una distribución condicional de los parámetros del modelo dada la información *a priori* y el conjunto de elecciones observadas. El teorema de la probabilidad condicionada de Bayes establece la relación definida entre la distribución *a priori* y la distribución *a posteriori*. Tenemos que:

$$\begin{aligned} P(\beta, Y) &= P(\beta|Y) \cdot P(Y) \\ P(\beta, Y) &= P(Y|\beta) \cdot P(\beta) \end{aligned} \tag{2.45}$$

y por tanto,

$$P(\beta|Y) = \frac{P(Y|\beta) \cdot P(\beta)}{P(Y)} \tag{2.46}$$

donde $P(Y)$ es la probabilidad marginal de Y que es constante (no depende de los valores de los parámetros del modelo). Esta probabilidad es la constante de normalización, que asegura que la suma de las probabilidades sea igual a uno (propiedad requerida para las funciones de densidad); por otro lado, $P(Y|\beta)$ es la probabilidad de observar la elección dados los valores de los parámetros del modelo. Este valor es conocido como la verosimilitud de los datos.

La distribución *a posteriori* se puede expresar como una proporción de la verosimilitud de los datos y de la distribución *a priori*, tal que:

$$P(\beta|Y) \propto P(Y|\beta) \cdot P(\beta) \quad (2.47)$$

Es posible dar una interpretación clásica para los estimadores bayesianos lo que permite su obtención de éstos para los mismos objetivos de análisis que los estimadores máximo verosímiles. Así el Teorema de Bernstein-von Mises²³ (Bernstein, 1917; von Mises, 1931) establece que:

Dado un modelo con parámetros β , siendo el verdadero parámetro β^* , el estimador máximo verosímil $\hat{\beta}$ y el estimador bayesiano $\bar{\beta}$ se tiene que:

La distribución *a posteriori* de β converge a una distribución Normal con matriz de covarianza $-H^{-1}/N$ al aumentar el tamaño de la muestra²⁴. Es decir,

²³ El planteamiento inicial del teorema es de Laplace (1820) y los trabajos posteriores de Bernstein (1917) y de von Mises (1931) dieron lugar a lo que hoy se conoce como el Teorema de Bernstein –von Mises. En realidad, la explicación y notación corresponden a la utilizada por Train (2002).

²⁴ $-H$ es la matriz de información y se define como $-H = -E \left[\frac{\partial^2 \log P(y_n | \beta^*)}{\partial \beta \partial \beta'} \right]$ para que sea definida positiva, Train (2002).

$\sqrt{N}(\beta - \bar{\beta}) \xrightarrow{d} N(0, -H^{-1})$. Esto implica que la desviación estándar de la distribución *a posteriori* corresponde a los errores estándar del enfoque clásico.

La media de la distribución *a posteriori* converge al máximo de la función de verosimilitud. Esto es, $\sqrt{N}(\bar{\beta} - \hat{\beta}) \xrightarrow{p} 0$. Este resultado es una implicación del primer punto. La distribución *a posteriori* es asintóticamente Normal y su media y su máximo coinciden. Asimismo, la distribución *a posteriori* es asintóticamente proporcional a la función de verosimilitud y desaparece la diferencia entre los estimadores máximo verosímil y bayesiano. Esto se debe al hecho de que la distribución *a posteriori* es proporcional a la distribución *a priori* y para muestras grandes desaparece el efecto de la segunda sobre la primera.

La distribución *a posteriori* se distribuye asintóticamente igual que el máximo de la función de verosimilitud. Es decir, $\sqrt{N}(\bar{\beta} - \beta^*) \xrightarrow{d} N(0, -H^{-1})$. Esto implica que el estimador bayesiano es equivalente al estimador máximo verosímil desde el punto de vista clásico.

El procedimiento para la estimación de la distribución *a posteriori* de los parámetros en un modelo de tipo ML es el denominado Bayes Jerárquico (HB) que fue desarrollado por Allenby (1997), SawtoothSoftware (1999) y Train (2001). La distribución *a posteriori* de los parámetros a estimar es proporcional a la verosimilitud de los datos y a la distribución *a priori*. Se supone, por tanto, que los parámetros distribuyen según una determinada función de densidad conocidos los valores de la media y varianza poblacionales, esto es, $\beta_n \square \phi(b, \Sigma)$. Asimismo, la media poblacional distribuye

Normal mientras que la varianza poblacional distribuye Wishart invertida, quedando definida la distribución *a priori* como $k(b, \Sigma)^{25}$. De esta manera, tenemos:

$$K(b, \Sigma, \beta_n \forall n | Y) \propto \prod_n L(y_n | \beta_n) \phi(\beta_n | b, \Sigma) k(b, \Sigma) \quad (2.48)$$

siendo $L(y_n | \beta_n)$ la probabilidad de que la persona n realice la secuencia de elecciones observada condicionada al vector de parámetros β_n .

La estimación del modelo implica un proceso iterativo que requiere de valores²⁶ iniciales de los parámetros. Este proceso aplica el muestreo de Gibbs para obtener simulaciones de la distribución *a posteriori* en la ecuación (2.48). El procedimiento tiene los siguientes pasos:

Se genera un nuevo valor de b condicionado a los valores de Σ y de $\beta_n \forall n$. El nuevo valor se muestrea de una distribución Normal con media igual al valor medio de los parámetros β_n en la población y matriz de covarianza conocida Σ .

Se simula un nuevo valor para la varianza Σ a partir de una distribución Wishart invertida condicionado a los valores de b y β_n .

Una vez obtenidos los nuevos valores para la media y la varianza (b y Σ , respectivamente), se generan nuevos valores para los parámetros β_n mediante el algoritmo de Metrópolis-Hastings (SawtoothSoftware, 1999). Este algoritmo se basa en

²⁵ Se puede asumir independencia entre b y Σ y entonces $k(b, \Sigma) = \gamma(b)\psi(\Sigma)$.

²⁶ Estos valores iniciales pueden tomarse de estudios previos y/o de suposiciones del investigador. Suelen considerarse distribuciones poco informativas (por ejemplo media cero y varianza grande).

la obtención sucesiva de valores de β_n hasta encontrar aquel valor que consigue el mejor ajuste a los datos, de manera que no sea posible alcanzar un ajuste superior.

Cada uno de estos pasos es repetido en cada iteración. El conjunto de parámetros b , Σ y β_n se re-estiman en cada una de ellas. Este procedimiento de simulación basado en el muestreo de Gibbs converge a la distribución *a posteriori* de todos los parámetros. La convergencia del procedimiento se asegura con un número elevado de simulaciones, entre 100.000 y 200.000. El muestreo de la distribución *a posteriori* requiere de nuevas simulaciones, una vez que el procedimiento converge. Dado que cada iteración se construye a partir de la iteración anterior, existe correlación entre las simulaciones. Para evitarlo se extraen una de cada diez o más. El estimador bayesiano del parámetro es el valor medio de la distribución *a posteriori*.

Como en el caso de método clásico de estimación, es posible realizar distintos test de ajuste general (Ortúzar y Willumsem, 2001). En cada una de las iteraciones se calcula el valor de la función de verosimilitud de los datos con los parámetros estimados hasta ese punto de la simulación. Con este dato se puede calcular la log-verosimilitud como una variable adicional que es reestimada en cada iteración. Alcanzada la convergencia, con los datos de esta variable adicional se obtiene su distribución *a posteriori*. Tenemos así un valor de la log-verosimilitud²⁷ que sirve como indicador del ajuste del modelo y permite realizar los tests de selección de modelos.

Para la estimación de parámetros individuales, además del método de Bayes Jerárquico descrito anteriormente, existe un segundo procedimiento desarrollado por Revelt y Train (2000). Su planteamiento se concreta en obtener la distribución de gustos

²⁷ Este valor presenta las mismas propiedades de la log-verosimilitud estimada por el método clásico.

individuales condicionado a las elecciones observadas de los individuos y la distribución de gustos poblacionales caracterizados por su media y su varianza partiendo de la discusión de Allenby y Rossi (1999).

Dada una secuencia de T_n elecciones y_n y dada la distribución de los gustos en la población, esto es, conocidos los parámetros poblacionales, la densidad condicional de cualquier parámetro individual se expresa a través de la regla de Bayes como:

$$h(\beta_n | y_n, b, \Sigma) = \frac{P(y_n | \beta_n) \cdot g(\beta_n | b, \Sigma)}{P(y_n | b, \Sigma)} \quad (2.49)$$

La esperanza condicional de β_n se calcula integrando sobre todo el dominio de su distribución, cuya expresión formal es:

$$\begin{aligned} E(\beta_n | y_n, b, \Sigma) &= \int \beta_n \cdot h(\beta_n | y_n, b, \Sigma) \cdot d\beta_n = \\ &= \frac{\int \beta_n \cdot P(y_n | \beta_n) \cdot g(\beta_n | b, \Sigma) d\beta_n}{P(y_n | b, \Sigma)} = \\ &= \frac{\int \beta_n \cdot P(y_n | \beta_n) \cdot g(\beta_n | b, \Sigma) \cdot d\beta_n}{\int P(y_n | \beta_n) \cdot g(\beta_n | b, \Sigma) \cdot d\beta_n} \end{aligned} \quad (2.50)$$

La aproximación simulada de esta expresión se obtiene promediando R observaciones de β_n^r . Estas observaciones son extraídas de la función de densidad poblacional y cada una de ellas es ponderada por la probabilidad logit condicionada a cada observación. La expresión es, por tanto:

$$ES(\beta_n | y_n, b, \Sigma) = \frac{\sum_{r=1}^R \beta_n^r \cdot P(y_n | \beta_n^r)}{\sum_{r=1}^R P(y_n | \beta_n^r)} \quad (2.51)$$

Esta expresión permite obtener los estimadores de los parámetros individuales.

Por otra parte, Revelt y Train (2000) proponen un método alternativo donde los coeficientes del modelo están condicionados por las elecciones individuales. Así la relación de proporcionalidad se puede establecer partiendo de la ecuación (2.49), puesto que el denominador es un valor constante que no depende de los parámetros individuales. Tenemos, entonces:

$$h(\beta_n | y_n, b, \Sigma) \propto P_n(y_n | \beta_n) \cdot g(\beta_n | b, \Sigma) \quad (2.52)$$

Las observaciones de la distribución de los parámetros individuales (2.49) se obtienen utilizando el algoritmo de Metrópolis-Hastings (Chib y Greenberg, 1995) que consiste en extraer simulaciones de manera iterativa de la distribución de los parámetros individuales condicionada a los parámetros poblaciones, $g(\beta_n | b, \Sigma)$, que serán aceptadas como valores de los parámetros individuales si incrementan la verosimilitud de los datos. Cada valor muestreado se almacena después de un cierto (elevado²⁸) número de iteraciones. Finalizado el proceso iterativo se construye la distribución de los parámetros calculando las frecuencias observadas para los parámetros individuales durante el proceso de simulación. Este método es bastante sencillo de implementar.

Cabe señalar dos importantes diferencias entre este método y el método HB. La primera está en la forma en que la incertidumbre sobre los parámetros de la población, b y Σ , es definida y calculada. En el HB para calcular la matriz de covarianza se especifica una distribución *a priori* para los parámetros poblaciones que refleja el conocimiento y/o creencia previa sobre éstos antes de examinar los datos, de manera que la matriz de

²⁸ Algunas miles de iteraciones.

covarianza de la distribución *a posteriori* se calcula a partir de la regla de Bayes combinando esta información *a priori* con información exógena (es decir, los datos). Mientras que en este método la matriz de covarianza se obtiene por la maximización de la log-verosimilitud y es calculada en base a una formulación asintótica. La segunda diferencia está en que el HB simula las distribuciones *a posteriori* de los parámetros poblacionales y de cada uno de los parámetros individuales; mientras que el otro requiere calcular la función de verosimilitud para los parámetros de la población.

La información *a priori* es más sencilla de introducir en el análisis bayesiano. Sin embargo, para trabajar con distribuciones distintas a la Normal es más sencillo implementar el método de Revelt y Train (2000).

2.4 La predicción de demanda con modelos desagregados

En la mayoría de las ocasiones, la estimación de modelos de elección discreta que permiten obtener las probabilidades individuales tiene como objetivo final su empleo para la predicción de demanda agregada y su variación ante cambios de política. Pasamos del ámbito individual o desagregado al agregado, fase que requiere de la agregación de las estimaciones obtenidas a nivel individual. Esta agregación puede introducir sesgos y por otra parte, el proceso mismo de agregación no es trivial en la mayoría de los casos²⁹.

Partiendo de un modelo desagregado en el que la probabilidad de que el individuo n elija la alternativa j es:

²⁹ Sólo cuando el modelo desagregado es lineal, entonces el proceso de agregación es trivial; se trata solamente de sustituir los promedios de las variables explicativas en la especificación del modelo.

$$P_{jn} = f_j(\vec{X}_n) \quad (2.53)$$

donde \vec{X}_n , es el vector de variables explicativas, esto es, de atributos de la alternativa j y f_j , es la función de elección que puede ser logit, probit, etc., para la alternativa j . La proporción de mercado estimada de la alternativa j es el valor esperado de las probabilidades individuales sobre el grupo, es decir:

$$P_{jN} = \frac{1}{N} \sum_{j=1}^N f_j(\vec{X}_n) \quad (2.54)$$

siendo N el número de individuos en el grupo. Este procedimiento es conocido como enumeración.

Asimismo, conocida la distribución del vector de atributos para toda la población, el problema de determinar la probabilidad de elección de la alternativa j en el ámbito agregado se reduce al cómputo de la siguiente integral:

$$P_{jN} = \int_{\vec{X}_N} f_j(\vec{X}) g(\vec{X}_N) d\vec{X}_j \quad (2.55)$$

donde $g(\vec{X}_N)$ es la distribución de probabilidad conjunta de las variables explicativas, \vec{X}_j , para el grupo de individuos N .

Ambas expresiones, (2.54) y (2.55), presentan cierta complejidad, y no se utilizan en la práctica ya sea porque requieren una importante cantidad de datos, o bien porque sólo es posible de resolver mediante procedimientos numéricos complejos. Existen diversas técnicas que permiten aproximar estas expresiones. Las más utilizadas en la práctica son el enfoque inocente, el método de clasificación y la enumeración muestral.

El *enfoque inocente* consiste en reemplazar los valores promedio de los atributos en la función individual. Así la probabilidad (2.53) se aproxima por:

$$P_j \approx f_j(\bar{X}_{jn}; \bar{\theta}) \quad (2.56)$$

Este método sólo es apropiado cuando la función de elección, f_j , es lineal. En el caso que sea no lineal, se pueden cometer sesgos importantes en la predicción. Una manera de reducir este sesgo es dividiendo la población en un número finito de clases homogéneas y aplicar en cada una de ellas el enfoque inocente. Este método se denomina *método de clasificación* y en este caso la predicción agregada está dada por la siguiente expresión:

$$P_{jN} = \sum_{c \in N} \frac{n_c}{N} f_j(\bar{X}_c) \quad (2.57)$$

donde \bar{X}_c es el vector que representa el promedio de los atributos de la clase c y n_c/N es la proporción de individuos en dicha clase. El problema en este método surge a la hora de determinar el número de clases y cómo van a estar definidas. McFadden y Reid (1975) definen un criterio para determinar las clases en función del tamaño de la varianza de las diferencias entre los distintos pares de utilidades estimadas y demuestran que con ocho clases definidas de esta forma, el sesgo de agregación es despreciable.

El *método de enumeración muestral* consiste en aplicar el método de enumeración³⁰ a una muestra representativa de la población. Este enfoque implica una serie de etapas; en primer lugar, predecir el comportamiento de cada individuo en la muestra y obtener el promedio de tales predicciones. En segundo lugar, predecir el comportamiento de cada

³⁰ Ecuación (2.54)

individuo para las condiciones de cambio y obtener el promedio de las mismas; y en tercer lugar, calcular las diferencias entre las predicciones de los pasos 1 y 2 con el objetivo de determinar el efecto agregado que tuvo la aplicación de la política.

Los tres pasos descritos no generan ninguna complicación adicional siempre y cuando la muestra sea aleatoria. Si la muestra es estratificada, es necesario ponderar³¹ cada observación conforme a su representatividad en la población. Este método resulta eficiente cuando el conjunto de elección no es muy grande y las predicciones se realizan a corto plazo. En el largo plazo, no se puede garantizar que la distribución de alternativas y atributos en la muestra se mantenga constante a lo largo del tiempo. Para estos casos, Daly y Gunn (1986) han propuesto aplicar este método para una muestra artificial.

El método de clasificación coincide con el enfoque inocente cuando sólo existe una clase. Y si en el método de enumeración muestral cada individuo es una clase, este enfoque es idéntico al método de clasificación.

Existen otros métodos de agregación basados en la obtención de la distribución del vector de atributos a lo largo de la población empleando diferentes procedimientos. En Ben-Akiva y Lerman (1985) se encuentra una revisión detallada de estos métodos.

Finalmente, se debe señalar que en la etapa predictiva puede aparecer un problema serio cuando se producen cambios en la varianza de los errores de los datos usados en cada una de las distintas fases: de estimación y de predicción. Esto puede ser debido a un cambio en el método de medición de los datos o a un cambio real en la componente

³¹ El factor de ponderación se calcula como el cociente entre la incidencia de la observación en la población y en la muestra.

aleatoria. Ortúzar e Ivelic (1987) muestran que si los errores en la fase de estimación son menores que los errores en la fase predictiva, la demanda será sobrestimada y viceversa.

2.5 Las fuentes de datos

2.5.1 Preferencias Reveladas

Las Preferencias Reveladas (PR) son datos que reflejan el comportamiento actual de los individuos en sus decisiones de viaje. Estos se obtienen a partir de encuestas que permiten recoger información de las variables que explican la utilidad de las distintas alternativas y de las elecciones realizadas. Hasta la mitad de los años 80 éste fue el tipo de datos más utilizado en la modelización de la demanda de transporte. Sin embargo, presentan una serie de limitaciones en términos de comprensión del comportamiento de viajes (ver Ortúzar y Willumsem, 2001):

Las observaciones de las elecciones actuales pueden no presentar suficiente variabilidad para la construcción de buenos modelos que permitan evaluar proyectos y realizar predicciones.

Las variables más interesantes suelen estar correlacionadas, por ejemplo, el tiempo de viaje y el coste. En este caso, es difícil, separar dicho efecto en la modelización y por lo tanto también en la fase predictiva.

No es posible el estudio de variables latentes³². Pueden existir factores que dominen el comportamiento actual, lo que dificulta detectar la importancia relativa de otras

³² Las variables latentes son variables difíciles de medir o que no presentan suficiente variabilidad.

variables igualmente importantes como el confort, la seguridad, la puntualidad del servicio, etc.

No permiten estudiar los efectos de nuevas políticas, como por ejemplo la introducción de un nuevo modo de transporte.

No existe información completa sobre las condiciones del mercado, lo que dificulta determinar el conjunto real de elecciones disponibles.

Están sujetas a potencialmente importantes errores de medida, especialmente de nivel de servicio. Esto se intenta resolver realizando el investigador las mediciones para el conjunto de alternativas disponibles.

2.5.2 Preferencias Declaradas

Las Preferencias Declaradas (PD) son datos que tratan de reflejar lo que los individuos harían ante determinadas situaciones hipotéticas construídas por el investigador. Las PD se desarrollaron inicialmente en el ámbito de la investigación de mercado y comenzaron a ser utilizadas en la modelización de transporte a fines de los años 70. A diferencia de los datos de PR, que entregan información sobre los viajes que realiza un individuo habitualmente, los datos de PD informan sobre los viajes que el individuo realizaría si, por ejemplo, se introdujera un nuevo modo de transporte, se mejorase la calidad del servicio, se ofreciese una ruta alternativa más rápida, etc.

La posibilidad de diseñar experimentos de PD permite, en principio, resolver los problemas que presentan las PR (ver Ortúzar y Willumsen, 2001):

- ✓ Se puede ampliar el rango de variación hasta el nivel en el que existe un *compromiso* entre las distintas alternativas consideradas en el diseño.

- ✓ En la construcción de los escenarios se puede evitar la existencia de correlación entre variables.
- ✓ Permiten incorporar tanto atributos como alternativas no disponibles en el momento del análisis.
- ✓ Se puede aislar el efecto de un determinado atributo así como considerar variables latentes.
- ✓ El conjunto de elección se puede pre-especificar.
- ✓ En el diseño, se pueden evitar los errores de medida.

No obstante, no podemos estar seguros de que el individuo se comporte como dice que haría cuando contesta a una encuesta de PD. Es importante, por tanto, diseñar ejercicios que sean plausibles y realistas para que el entrevistado se implique en el juego correctamente. Los tipos de errores clasificados para este tipo de datos son cuatro (ver por ejemplo Bradley y Kroes, 1990):

Sesgo de afirmación: El entrevistado contesta, consciente o inconscientemente, lo que cree que el entrevistador quiere.

- *Sesgo de racionalización:* El entrevistado intenta ser racional en sus respuestas con el objetivo de justificar su comportamiento en el momento de la entrevista.
- *Sesgo de política:* El entrevistado contesta con el objetivo de influir en las decisiones de política en función de su creencia de cómo pueden afectar los resultados de la encuesta.
- *Sesgo de no restricción:* A la hora de responder no se toma en cuenta todas las restricciones que afectan a su comportamiento, de manera que las respuestas no son reales.

En un ejercicio de PD se pueden distinguir tres elementos principalmente. En primer lugar, tenemos la situación en la que el individuo se encuentra para declarar sus preferencias; ésta puede ser una situación real (un viaje que realice en este momento como ir al trabajo) o hipotética (un viaje que realizaría en el futuro dadas una serie de condiciones), y constituye el contexto de decisión. En segundo lugar, se deben seleccionar las alternativas, normalmente hipotéticas aunque algunas de ellas pueden existir en la actualidad, que se presentan en el ejercicio como función de un conjunto de atributos. En tercer lugar, está la forma en que los individuos pueden declarar sus preferencias (ver Ortúzar y Garrido, 2000); las más frecuentes son: Jerarquización (*Ranking*), Escalamiento o Elección Generalizada (*Rating*) y Elección (*Choice*).

Un punto importante del experimento de PD es la selección de los atributos a considerar en cada alternativa del ejercicio. Una forma de identificar los atributos más relevantes es realizando un grupo de discusión (*focus group*) con una muestra representativa de individuos. Un grupo de discusión no es más que una reunión de individuos dirigida por un moderador de manera que los participantes hablen sobre el tema considerado y durante el desarrollo de la misma se obtengan tanto los elementos clave en el análisis como una mejor comprensión del problema tratado.

El número de atributos a considerar para cada juego es determinado por el investigador; no obstante, la literatura recomienda que no debe ser muy elevado (no más de cuatro) por cada juego para evitar el efecto fatiga (Carson *et al.* 1994) o que contesten de manera lexicográfica (Saelensminde, 1999), tema que trataremos más adelante.

Las unidades de medida en que se expresan los atributos pueden ser triviales en algunos casos, como el tiempo o el coste, pero no lo son en otros como el confort, la seguridad, etc. Estos atributos requieren un estudio más detallado en grupos de discusión, unido a

la realización encuestas piloto que ayuden a encontrar la métrica más adecuada³³. La definición de niveles es también decisión del investigador; sólo se recomienda definir mayores niveles de variación para las variables consideradas más importantes en la elección a estudiar (Wittink *et al*, 1982), pero se debe cuidar que esto no sesgue los resultados del experimento.

El diseño factorial completo determina el número total de opciones dados los atributos y los niveles de variación para cada uno de ellos. Si a es el número de atributos y n el número de niveles de cada atributo, el número total de opciones es n^a . Si tenemos un diseño con dos atributos a dos niveles y tres atributos a tres niveles, el número de opciones resultantes es 108 ($2^2 \times 3^3$). En la práctica no es factible presentar a un individuo 108 opciones para que declare sus preferencias. Este número se puede reducir si consideramos un diseño factorial fraccional, que no es otra cosa que un diseño formado por un subconjunto de opciones del diseño factorial completo. La diferencia entre un diseño factorial completo y un diseño factorial fraccionado está en que mientras el primero permite medir todos los efectos principales e interacciones entre las variables, en el segundo se sacrifica la medición de algunas (o todas) las interacciones con el fin de reducir el número de opciones. Los efectos principales se definen como la respuesta de pasar al siguiente nivel de la variable en cuestión manteniendo constante el resto de variables, mientras que las interacciones consideran los efectos debidos a la variación conjunta de más de una variable; por ejemplo, en la elección modal el efecto de la interacción entre el tiempo y el coste. Normalmente, los cambios en la elección están determinados por los efectos principales y según Louviere (1988):

³³ En la siguiente sección se presenta el proceso de generación del diseño experimental, donde la definición de la variable latente considerada requirió de un trabajo minucioso.

- El 80% o más de la varianza de los datos lo explican los efectos principales.
- Las interacciones de dos términos explican más de un 2% o 3% de la varianza.
- Las interacciones de tres términos explican una proporción muy pequeña de la varianza, del orden del 0.5% al 1% y rara vez sobre el 2% o el 3%.
- Las interacciones de mayor orden explican una proporción minúscula de la varianza de los datos.

Si sólo se desea analizar los efectos principales, el diseño de 108 opciones anterior se reduciría a 16, y a 27 opciones si se considera interacciones de dos términos (Kocur *et al*, 1982), que son números más manejables. No obstante, la literatura (Bradley y Daly, 1994) recomienda no presentar más de 10 (entre 8 y 10) situaciones de elección a cada individuo (ver la discusión de Cherchi y Ortúzar, 2002). En casos en que se requiera estudiar interacciones y el número de opciones sea elevado, se puede utilizar el diseño en bloques; para esto se subdivide la muestra y a cada submuestra se le presenta uno de los bloques (Louviere *et al*, 2000). El tamaño de cada submuestra recomendado (Kocur *et al*, 1982) es de un mínimo de 30 personas por grupo, aunque en la práctica este tamaño parece bajo (ver también la discusión de Ortúzar y Willumsen, 2001).

Las formas de obtener la respuesta en un experimento de PD son, como comentamos anteriormente, Jerarquización, Escalamiento y Elección:

Jerarquización: Se presentan todas las opciones simultáneamente al individuo y se le pide que las ordene en función de sus preferencias, de más a menos preferida. Al ordenar las opciones, el individuo está jerarquizando los valores de utilidad de forma que la opción más preferida le reportará un mayor nivel de utilidad.

Escalamiento: Se le pide al individuo que exprese su grado de preferencia para una opción utilizando una escala arbitraria que puede ser numérica (de 1 a 5 o de 1 a 10) o semántica; por ejemplo: 1 = siempre elijo A, 2 = probablemente elijo A, 3 = ninguna; 4 = probablemente elijo B, 5 = siempre elijo B.

Elección: El individuo selecciona una de las distintas opciones que se le presentan que pueden ser dos (elección binaria) o más de dos (elección múltiple). Se considera que ésta es la forma más sencilla de responder a una encuesta de PD para un individuo porque es la forma habitual en la que toma decisiones. En estos casos, puede incluirse la alternativa “ninguna de ellas” para no forzar al entrevistado a elegir cuando ninguna le parece conveniente.

Un aspecto importante en este tipo de diseños es la necesidad de hacer creíble las distintas opciones presentadas al individuo. Para ello podemos definir el contexto en el que el experimento se realiza y en ese contexto hacer factibles las opciones de elección. Por ejemplo, para que sea considerado viable que exista un bus que realice un recorrido en un tiempo menor que un vehículo privado, puede ser necesario que existan una serie de condiciones tales como: carriles exclusivos para el bus, preferencia en los semáforos, etc. Este tipo de elementos se pueden considerar en el contexto en el que se presenta el experimento haciendo más realistas las opciones.

2.5.3 Datos Mixtos

Con el objetivo de aprovechar las ventajas de cada tipo de datos, PR y PD, e intentar evitar sus desventajas se plantea la *estimación con datos mixtos*³⁴ cuya formalización

³⁴ También conocida con el nombre de *estimación mixta*.

econométrica se debe a Ben-Akiva y Morikawa (1990). Esta estimación se basa en la utilización de ambos conjuntos de datos, PR más PD, en la estimación econométrica de modelos de demanda de transporte de forma que sea posible aprovechar las complementariedades de estas dos fuentes de información.

Desde el punto de vista econométrico, la diferencia fundamental entre los datos de PR y de PD es que presentan distintos tipos de error. Como ya hemos comentado, los datos de PR tienen errores en la medición de las variables independientes³⁵; mientras que los datos de PD presentan errores en la variable dependiente³⁶, debido a que no existe compromiso entre lo que el individuo declara que haría y lo que realmente hará cuando se dé la misma situación.

De acuerdo a su naturaleza, las distintas fuentes de datos se clasifican en: *datos primarios* y *datos secundarios*. La fuente de *datos primarios* proporciona información directa acerca de los parámetros del modelo mientras que la fuente de *datos secundarios* proporciona información indirecta. En el caso de modelos de elección modal la principal fuente de información sería una encuesta de PR, *fuentes de datos primarios*, y la información adicional proporcionada por la encuesta de PD se consideraría como *fuentes de datos secundarios*.

La estructura econométrica de estimación conjunta de ambas fuentes de datos se basa en la idea que la diferencia entre los errores de los datos de PR y PD se puede especificar considerando términos de error con diferente varianza. Si ε es el error estocástico de los datos de PR y η el de los datos de PD, podemos expresar la diferencia entre las varianzas a través de la expresión:

³⁵ En nuestro caso, las variables independientes son los atributos de las distintas alternativas.

³⁶ En nuestro caso, la variable independiente es la alternativa elegida.

$$\sigma_{\varepsilon}^2 = \mu^2 \cdot \sigma_{\eta}^2 \quad (2.58)$$

siendo μ un parámetro desconocido.

Esta consideración determina que las funciones de utilidad³⁷ de la alternativa j para cada una de las fuentes de datos sean:

$$\begin{aligned} U_j^{PR} &= V_j^{PR} + \varepsilon_j = \theta \cdot X_j^{PR} + \alpha \cdot Y_j^{PR} + \varepsilon_j \\ \mu U_j^{PD} &= \mu(V_j^{PD} + \eta_j) = \mu(\theta \cdot X_j^{PD} + \omega \cdot Z_j^{PD} + \eta_j) \end{aligned} \quad (2.59)$$

donde θ , α y ω son los parámetros a estimar. X_j^{PR} y X_j^{PD} son atributos comunes de la alternativa j para los datos de PR y de PD respectivamente; mientras que Y_j^{PR} y Z_j^{PD} son atributos no comunes de la alternativa j para cada conjunto de datos.

Al multiplicar la función de utilidad de los datos de PD por el parámetro desconocido μ , ecuación (2.59), se consigue que el error estocástico de este tipo de datos tenga la misma varianza³⁸ que los datos de PR.

Normalmente, se supone que los datos de PD deberían tener más ruido que los datos de PR, si éste es el caso, el valor de μ , conocido como el *coeficiente de escala* del modelo, debiera estar entre 0 y 1. Si el valor resultara mayor que 1 estaría indicando que los datos con mayor nivel de ruido son los de PR.

Cuando los errores distribuyen Gumbel con media cero y distinta varianza, las probabilidades de elección están dadas por las siguientes expresiones (Ben-Akiva y Morikawa, 1990):

³⁷ Las funciones de utilidad están basadas en el enfoque de la *utilidad aleatoria* expuesto en el epígrafe 2.2.3.

³⁸ La varianza viene definida por la ecuación (2.58).

$$P_j^{PR} = \frac{\exp(\theta X_j^{PR} + \alpha Y_j^{PR})}{\sum_i \exp(\theta X_i^{PR} + \alpha Y_i^{PR})} \quad (2.60)$$

$$P_j^{PD} = \frac{\exp\{\mu(\theta X_j^{PD} + \omega Z_j^{PD})\}}{\sum_i \exp\{\mu(\theta X_i^{PD} + \omega Z_i^{PD})\}}$$

siendo la función de verosimilitud conjunta:

$$L(\theta, \mu, \alpha, \omega) = \left(\prod_{h=1}^{H^{PR}} \prod_{A_j \in A(n)} (P_{jn}^{PR})^{g_{jn}} \right) \cdot \left(\prod_{h=1}^{H^{PD}} \prod_{A_j \in A(n)} (P_{jn}^{PD})^{g_{jn}} \right) \quad (2.61)$$

donde g_{jn} se define como en la ecuación (2.37)

Los parámetros del modelo se obtienen maximizando la función de verosimilitud que es no-lineal debido a que μ multiplica al resto de los parámetros en la función de utilidad PD. Para la estimación de este tipo de modelos se han desarrollado dos procedimientos: estimación simultánea y estimación secuencial³⁹.

La estimación simultánea fue propuesta por Bradley y Daly (1997). Esta se basa en proponer una estructura jerárquica como la de la Figura 2.2 para resolver la no linealidad (en los parámetros) en la ecuación (2.59). Según esta estructura las alternativas de PR caen directamente del nido raíz y las alternativas PD estarían, cada una de ellas, incluidas en un nido de alternativa única, siendo μ el parámetro que acompaña a la utilidad representativa de cada nido (EMU). El ejemplo mostrado en la Figura 2.2 corresponde a un caso de tres alternativas de PR y tres alternativas de PD.

³⁹ Swait *et al* (1994) han propuesto otro método de estimación secuencial similar al desarrollado por Ben-Akiva y Morikawa (1990).

Para cada observación de PR se considera que no están disponibles las alternativas de PD y para cada observación de PD se considera que no están disponibles las alternativas de PR. La modelización de la elección en este caso se realiza mediante un HL. Este procedimiento de estimación simultánea se puede generalizar a más de dos conjuntos de datos, basta con añadir los nidos artificiales para las alternativas de los conjuntos de datos adicionales teniendo en cuenta que el coeficiente de escala sea el mismo para cada conjunto de datos. Dado que se trata de un modelo tipo HL, la estimación se puede realizar con software estándar como por ejemplo ALOGIT (Daly, 1992).

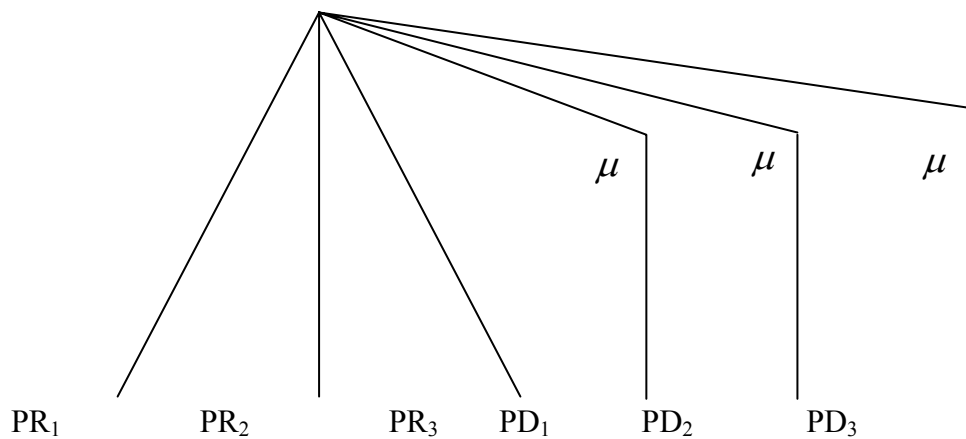


Figura 2.2: Estructura de árbol para estimación con datos mixtos.

El método secuencial fue propuesto por Ben-Akiva y Morikawa (1990). Se basa en un algoritmo con los siguientes tres pasos:

Se estima un modelo con los datos de PD según la expresión (2.59) con el objetivo de estimar los parámetros $\mu\theta$ y $\mu\omega$. Obtenidos estos valores se define una nueva variable, tal que:

$$V_j^{PR} = \mu\theta \cdot X_j^{PR} \quad (2.62)$$

Se estima un modelo con los datos de PR incluyendo la nueva variable definida anteriormente, (2.62), para obtener los parámetros λ y α según la expresión

$$U_j^{RP} = \lambda \bar{V}_j^{PR} + \alpha Y_j^{PR} + \varepsilon_j, \text{ siendo } \lambda = 1/\mu.$$

Finalmente, se ponderan X_j^{PD} y Z_j^{PD} por el *coeficiente de escala* del modelo, μ . Se obtiene así un conjunto de *datos secundarios* que tienen el mismo error aleatorio⁴⁰ que el conjunto de *datos primarios*, lo que posibilita que se estimen los parámetros de forma simultánea trabajando conjuntamente con ambos datos.

Ortúzar y Willumsen (2001) recomiendan utilizar el método simultáneo, que es más fácil de implementar, pues aplicaron tanto éste como el método secuencial al mismo conjunto de datos y los resultados fueron muy parecidos.

Una cuestión a resolver es si es posible combinar dos (o varias) fuentes de datos. Para combinar varias fuentes de datos, se supone que son iguales los parámetros de las variables que son comunes a las distintas fuentes de datos. Para el caso de combinar PR y PD sería θ en las funciones de utilidad de PR y PD, expresión (2.59). Louviere *et al* (2000) proponen estimar modelos MNL para los datos de PR y PD definiendo la misma especificación para las variables comunes a ambos datos. Estimados los parámetros, representar gráficamente los parámetros comunes a PR y PD (θ^{PR} y θ^{PD}), ver Figura 2.3:

⁴⁰ Ambos errores tienen media cero e igual varianza.

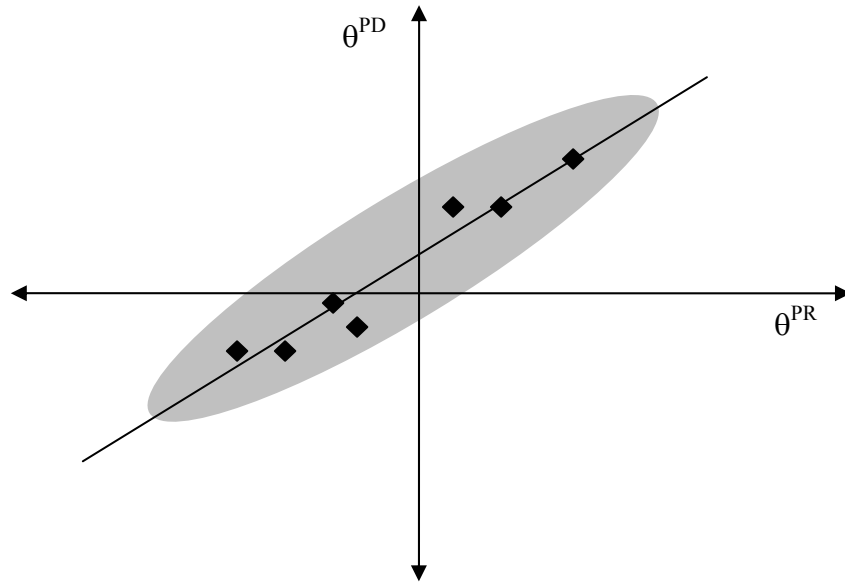


Figura 2.3: Representación de parámetros de PR y PD

Si los parámetros son iguales o parecidos debe resultar una nube de puntos que se aproxima a una línea recta, en este caso se podría estimar el modelo con datos mixtos. La situación contraria, nos estaría indicando que no se debe estimar el modelo combinando las fuentes de datos. Puede darse el caso de que algunos de los parámetros no se concentren en la nube de puntos que da lugar a una línea recta, en este caso se debería realizar una combinación parcial de datos, definiendo iguales parámetros para aquellas variables que resultan iguales entre PR y PD y para los que no, parámetros específicos para cada base de datos. Por ejemplo, según la Figura 2.4, los parámetros específicos serían para las variables tiempo de aparcamiento y frecuencia.

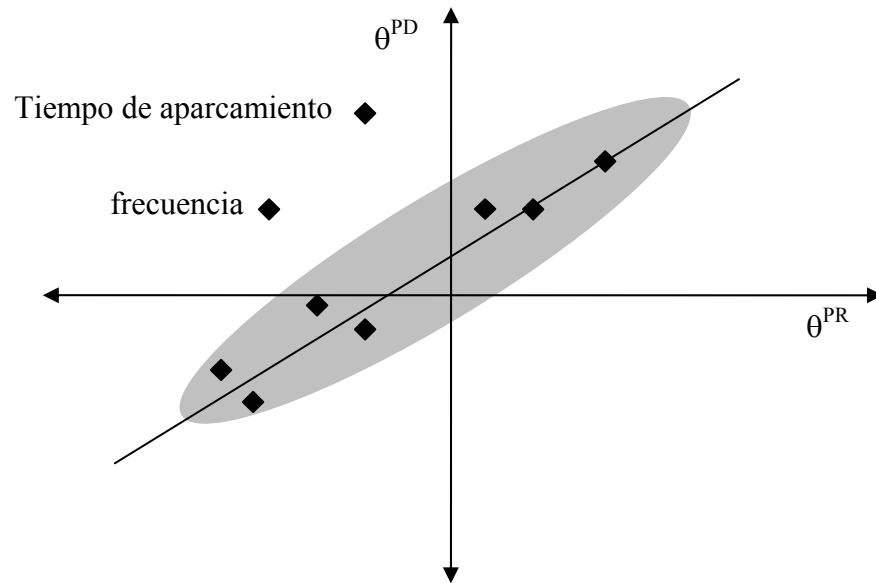


Figura 2.4: Representación de parámetros de PR y PD no iguales

Para la predicción Louviere *et al* (2000) proponen que el modelo contenga las constantes específicas de las alternativas de PR así como todos los parámetros estimados conjuntamente. Si los parámetros no son estimados conjuntamente, surge el problema de cuáles considerar, sugieren que si el parámetro de PR⁴¹ no estimado conjuntamente tiene el signo incorrecto, deberíamos utilizar el correspondiente al parámetro de PD y si existe una diferencia importante entre el parámetro de PR y de PD, no está claro cuál debe ser la solución; se podría estudiar cuán de sensible es la elección al parámetro con los datos de PR o una muestra de validación (holdout data set).

Alternativamente, existe un test para verificar la hipótesis de homogeneidad de los parámetros propuesto por Swait y Louviere (1993) cuyo procedimiento es el siguiente:

⁴¹ Según explican estos autores, es más frecuente que los datos de PR den lugar a estimaciones de parámetros con signo intuitivamente incorrectos (pp. 246).

Estimar un MNL para cada conjunto de datos de manera independiente, según la expresión (2.21).

Estimar el modelo conjuntamente según el método manual propuesto por los autores en dicho artículo.

Calcular el estadístico siguiente:

$$\chi^2 = -2 \left[(L^{PR} + L^{PD}) - L^{PR+PD} \right] \square \chi_{\theta-1}^2 \quad (2.63)$$

Este test es generalizable a cualquier número de bases de datos.

Si se acepta la hipótesis nula, existe homogeneidad de parámetros entre ambas bases de datos, y por tanto, se pueden combinar ambos datos. Si no se aceptara, probaríamos con la representación gráfica, pues el test puede rechazarse porque algunos parámetros no sean iguales, situación de la Figura 2.4.