13

Endogeneidad

13.1 Descripción general

Hasta ahora hemos supuesto que las variables explicativas que entran en un modelo de elección discreta son independientes de los factores no observados. Sin embargo, en muchas situaciones las variables explicativas son endógenas, es decir, están correlacionadas o en cualquier caso no son independientes de los factores no observados. Algunos ejemplos son los siguientes:

1. *Atributos no observados de un producto pueden afectar a su precio*.

   Al modelar las elecciones de un consumidor entre diferentes productos, podría ser imposible medir todos los atributos relevantes de los diversos productos. En el caso de los automóviles, por ejemplo, el investigador puede obtener información acerca de la eficiencia del combustible, la longitud, la anchura, la potencia, el peso y muchos otros atributos de cada vehículo ofrecido por los fabricantes, pero atributos tales como la comodidad, belleza del diseño, suavidad de marcha, manejo en curvas, valor esperado de reventa y prestigio no se pueden medir directamente. Sin embargo, es de esperar que el precio del producto refleje estos atributos no observados (es decir, no medidos). Hay dos razones por las cuales el precio se ve afectado. En primer lugar, en la medida en que los atributos no observados supongan un costo para el fabricante, es de esperar que el precio del producto refleje estos costos. En segundo lugar, si los atributos no observados afectan a la demanda del producto, un precio determinado por la interacción entre la oferta y la demanda podría reflejar estas diferencias en la demanda. El resultado final es que el precio está correlacionado con atributos no observados, en lugar de ser independiente como se ha supuesto hasta ahora en este libro.

2. *Los esfuerzos en actividades de marketing pueden estar relacionados con los precios*.

   La publicidad y la promoción de las ventas, a través de acciones como cupones y descuentos, están en todas partes. A menudo, las prácticas habituales de comercialización de las empresas crean una correlación entre el precio de los productos, que observa el investigador, y las actividades de promoción no vinculadas al precio, que por lo general el investigador no puede medir directamente. La correlación puede ir en cualquier dirección. Un producto puede ser promovido a través de una campaña publicitaria junto con descuentos. La publicidad en este caso estaría correlacionada negativamente con el precio: mayor publicidad se produce simultáneamente con precios más bajos. Por el contrario, las empresas pueden aumentar el precio de sus productos para pagar la publicidad, lo que crearía una correlación positiva. En cualquier caso, el precio del producto ya no es independiente de los factores no observados que afectan a las elecciones de los consumidores.

3. *Decisiones interrelacionadas de los decisores*.

   En muchas situaciones, los factores observados que afectan a la elección realizada por una persona están determinados por otra elección de esa persona. Medio de transporte y ubicación de vivienda son un ejemplo destacado. En un modelo de elección del medio habitual de transporte para ir al trabajo, las variables explicativas observadas son por lo general el costo y el tiempo de desplazamiento desde el hogar hasta el lugar de trabajo para cada medio (automóvil, autobús y tren). Sin embargo, las personas que tienden a preferir el transporte público (o que les disgusta menos que a la persona promedio) también podrían tender a comprar o alquilar casas que están cerca del transporte público. Por lo tanto, el tiempo de viaje en transporte es menor para estas personas que para las personas que residen más lejos de la zonas de tránsito. Expresado en términos de factores observados y no observados del modelo de elección del medio de transporte, las actitudes observadas hacia el transporte público, que afectan a la elección del medio de transporte pero que no pueden ser medidas por completo por el investigador, están (negativamente) correlacionadas con el tiempo observado de los viajes en transporte público.

En situaciones como éstas, estimar un modelo sin tener en cuenta la correlación entre factores observados y no observados es inconsistente. La dirección del sesgo a menudo se puede determinar por pura lógica. Por ejemplo, si los atributos no observados deseables están correlacionados positivamente con el precio, la estimación sin tener en cuenta esta correlación se traducirá en un coeficiente de precio estimado que estará sesgado a la baja en magnitud. La razón es obvia: dado que los precios más altos se asocian con atributos deseables, los consumidores evitan los productos más caros *menos* de lo que lo harían si los precios más altos se produjesen sin ningún cambio en compensación en los atributos no observados. Esencialmente, el coeficiente de precio estimado recoge tanto el efecto del precio (que es negativo) como el efecto de los atributos no observados deseables (que son positivos), con estos últimos ocultando el efecto del primero. Un sesgo similar, aunque en la dirección opuesta, se produce si las acciones de marketing consisten en publicidad acompañada de descuentos en los precios. El aumento de la demanda de los productos comercializados proviene tanto del precio más bajo y como de la publicidad (ajena a los precios). El coeficiente de precio estimado recogería los dos efectos, que sumados son mayores al impacto de los precios más bajos por sí mismos.

Varios métodos han sido desarrollados para estimar modelos de elección en presencia de variables explicativas endógenas. En este capítulo se describen estos métodos, delimitando las ventajas y las limitaciones de cada aproximación. En primer lugar describimos el enfoque BLP, desarrollado por Berry, Levinsohn y Pakes (de ahí las iniciales) a través de una serie de publicaciones. Berry (1994) señaló que se pueden incluir constantes en el modelo de elección para capturar el efecto promedio de los atributos del producto (tanto observados y como no observados). Es posible entonces hacer una regresión lineal de las constantes estimadas respecto a los atributos observados, donde la endogeneidad se maneja de la manera habitual, a través de estimación mediante variables instrumentales. En esencia, demostró que la endogeneidad podía extraerse del modelo de elección, que es intrínsecamente no lineal, e insertarlo en un modelo de regresión lineal, donde la endogeneidad puede ser manejada a través de estimación de variables instrumentales estándar. Para aplicar este método, a menudo es necesario estimar un número muy grande de constantes dentro del modelo de elección, lo cual puede ser difícil utilizando métodos de maximización estándar basados en el gradiente. Para abordar esta cuestión, BLP (1995) proporcionó un procedimiento, llamado "la contracción" (*the contraction*), que facilitaba la estimación de estas constantes. Estos dos artículos iniciales estaban basados en modelos agregados, es decir, modelos estimados con datos agregados de cuotas de mercado. Sin embargo, los conceptos son aplicables a datos de elección a nivel individual o a una combinación de datos a nivel agregado e individual, tal y como se utilizan en un artículo posterior por parte de BLP (2004). Casos de uso del enfoque BLP incluyen a Nevo (2001), Petrin (2002), Goolsbee y Petrin (2004), Chintagunta, Dubé y Goh (2005), y Train y Winston (2007), por nombrar sólo unos pocos. Una versión bayesiana del procedimiento ha sido desarrollada por Yang, Chen, y Allenby (2003) y Jiang, Manchanda y Rossi (2007).

El segundo procedimiento que describimos es el enfoque llamado de función de control. Los conceptos que motivan este enfoque datan de los trabajos de Heckman (1978) y Hausman (1978), aunque el primer uso del término "función de control" parece haber sido a cargo de Heckman y Robb (1985). La endogeneidad surge cuando las variables observadas están correlacionadas con factores no observados. Esta correlación implica que los factores no observados, condicionados a las variables observadas, no tienen media cero, como se requiere por lo general para una estimación estándar. Una función de control es una variable que captura la media condicionada, esencialmente "controlando" la correlación. Ríos y Voung (1988) adaptaron estas ideas para manejar la endogeneidad a un modelo probit binario con coeficientes fijos, y Petrin y Train (2009) generalizaron este enfoque a los modelos de elección multinomiales con coeficientes aleatorios. El procedimiento se lleva a cabo en dos pasos. En primer lugar, la variable explicativa endógena (como el precio) es objeto de una regresión respecto variables exógenas. La regresión estimada se utiliza para crear una nueva variable (la función de control) que se introduce en el modelo de elección. El modelo de elección es estimado a continuación, con las variables originales más la nueva variable, lo que permite representar adecuadamente la distribución de los factores no observados condicionados tanto a esta nueva variable como a las originales. Algunos ejemplos de este método los proporcionan Ferreira (2004) y Guervara y Ben-Akiva (2006).

El tercer procedimiento es un enfoque completo de máxima verosimilitud, tal y como lo aplican Villas-Boas y Winer (1999) a un modelo logit multinomial con coeficientes fijos, generalizado por Park y Gupta (de próxima publicación) a modelos de elección con coeficientes aleatorios. El procedimiento está estrechamente relacionado con el enfoque de la función de control, en el sentido de que tiene en cuenta la media condicionada distinta de cero de los factores no observados. Sin embargo, en lugar de implementar secuencialmente los dos pasos (es decir, estimar el modelo de regresión para crear la función de control y a continuación estimar el modelo de elección con esta función de control), los dos pasos se combinan en un criterio de estimación conjunta. Se requieren supuestos adicionales para poder hacer la estimación de forma simultánea, sin embargo, el procedimiento es más eficiente cuando se cumplen estos supuestos.

En las secciones siguientes, se trata cada uno de estos procedimientos, y se acompañan de un caso de estudio utilizando el enfoque BLP.

13.2 El Enfoque BLP

Este procedimiento se describe con mayor facilidad para elecciones entre productos en los que el precio es endógeno. Un conjunto de productos se vende en varios mercados y en cada mercado tiene numerosos consumidores. Los atributos de los productos varían entre mercados, pero no entre consumidores de cada mercado (es decir, todos los consumidores dentro de un mercado determinado se enfrentan a los mismos productos con los mismos atributos). La definición de lo que es un mercado depende del caso práctico. El mercado podría ser una zona geográfica, como en el análisis de Goolsbee y Petrin (2004) relativo a la elección hecha por los hogares entre la oferta televisiva. En este caso concreto, el precio de la televisión por cable y las características ofrecidas (como el número de canales) varían entre ciudades, ya que las franquicias de cable se conceden por los gobiernos locales. Alternativamente, el mercado podría ser definido temporalmente, como en el análisis BLP de la demanda de nuevos vehículos (1995, 2004), donde cada año-modelo consiste en un conjunto de marcas y modelos de vehículos nuevos junto a sus precios y otros atributos. Cada año constituye un mercado en este caso práctico. Si el análisis es entre productos cuyos atributos son los mismos para todos los consumidores, entonces sólo hay un mercado y la diferenciación por mercado es innecesaria.

13.2.1 Especificación

Sea M el número de mercados y J_m el número de opciones disponibles para cada uno de los consumidores en el mercado m. J_m es el número de productos disponibles en el mercado m más, tal vez, dependiendo del análisis, la opción de no comprar ninguno de los productos, que a veces se denomina "el bien externo" [ii] (*the outside good*). El precio del producto j en el mercado m se denota como p_{jm}. Algunos de los atributos de los productos distintos del precio son observados por el investigador y otros no. Los atributos distintos del precio del producto j en el mercado m que sí son observados se denotan por el vector x_{jm}. Los atributos no observados se designan colectivamente como ξ_{jm}, cuyo significado preciso se trata en mayor detalle más adelante.

La utilidad que el consumidor n en el mercado m obtiene del producto j depende de los atributos observados y no observados del producto. Suponga que la utilidad toma la forma

$$U_{njm} = V(p_{jm}, x_{jm}, s_n, \beta_n) + \xi_{jm} + \varepsilon_{njm},$$

donde $\mathrm{s_n}$ es un vector de características demográficas de los consumidores, $\mathrm{V}(\cdot)$ es una función de las variables observadas y las preferencias del consumidor, representadas por el β_n, y ε_{njm} es de tipo valor extremo iid. Observe que ξ_{jm} entra en la utilidad de la misma manera para todos los consumidores; en esta configuración, por lo tanto, ξ_{jm} representa el promedio, o la parte común, de la utilidad que los consumidores obtienen de los atributos no observados del producto j en el mercado m.

El problema fundamental que motiva el enfoque de esta estimación es la endogeneidad de los precios. En particular, el precio de cada producto depende en general de todos sus atributos, tanto los que son observados por el investigador como los que no pueden ser medidos por el investigador, pero sin embargo afectan la demanda y/o los costos del producto. Como resultado, el precio p_{jm}depende de ξ_{jm}.

Supongamos ahora que íbamos a estimar este modelo sin tener en cuenta esta endogeneidad. El modelo de elección incluiría el precio p_{jm} y los atributos observados x_{jm} como variables explicativas. La parte no observada de la utilidad, condicionada a β_n, sería $\varepsilon^*_{njm} = \xi_{jm} + \varepsilon_{njm}$, que incluye la utilidad media de atributos no observados. Sin embargo, dado que p_{jm} depende de ξ_{jm}, este componente no observado, ε^*_{njm}, no es independiente de p_{jm}. Por el contrario, se podría esperar una correlación positiva, estando los atributos no observados más deseables asociados a precios más altos.

La aproximación BLP a este problema consiste en mover ξ_{jm} a la parte observada de la utilidad. Esto se logra mediante la introducción de una constante para cada producto en cada mercado. Sea $\bar{V}$ la porción de $V(\cdot)$ que varía entre productos y mercados, pero que es igual para todos los consumidores. Sea $\tilde{V}$ sea la parte que varía entre consumidores así como entre mercados y productos. Entonces $V(\cdot) =$

[ii] Si el bien externo está incluido el modelo se puede utilizar para predecir la demanda total de productos bajo condiciones alteradas. Si el bien externo no está incluido, el análisis examina la elección de los consumidores entre los productos condicionada a la compra de uno de los productos. El modelo puede ser utilizado para predecir la evolución de las cuotas entre los consumidores que originalmente compraron los productos, pero no se puede usar para predecir los cambios en la demanda total, ya que no incluye los cambios en el número de consumidores que decidieron no comprar cualquiera de los productos. Si el bien externo está incluido, generalmente se considera que su precio es cero.

$\bar{V}(p_{jm}, x_{jm}, \bar{\beta}) + \tilde{V}(p_{jm}, x_{jm}, s_n, \tilde{\beta}_n)$, donde $\bar{\beta}$ son parámetros que son iguales para todos los consumidores y $\tilde{\beta}_n$ son parámetros que varían entre consumidores. Observe que $\bar{V}$ no depende de s_n ya que es constante entre consumidores[iii]. Es más natural pensar en $\bar{V}$ como la representación de la media de V en la población; sin embargo, no tiene por qué ser así. Todo lo que se requiere es que $\bar{V}$ sea constante entre consumidores. La variación en la utilidad de los atributos observados alrededor de esta constante es capturada por $\tilde{V}$, que puede depender de datos demográficos observados y de coeficientes que varían aleatoriamente. La utilidad es entonces

$$U_{njm} = \bar{V}(p_{jm}, x_{jm}, \bar{\beta}) + \tilde{V}(p_{jm}, x_{jm}, s_n, \tilde{\beta}_n) + \xi_{jm} + \varepsilon_{njm}.$$

Reordenando los términos, tenemos

$$U_{njm} = [\bar{V}(p_{jm}, x_{jm}, \bar{\beta}) + \xi_{jm}] + \tilde{V}(p_{jm}, x_{jm}, s_n, \tilde{\beta}_n) + \varepsilon_{njm}.$$

Observe que el término entre paréntesis no varía entre consumidores. Es constante para cada producto en cada mercado. Denotemos esta constante como

$$\delta_{jm} = \bar{V}(p_{jm}, x_{jm}, \bar{\beta}) + \xi_{jm} \tag{13.1}$$

y sustituyámosla en la utilidad

$$U_{njm} = \delta_{jm} + \tilde{V}(p_{jm}, x_{jm}, s_n, \tilde{\beta}_n) + \varepsilon_{njm}. \tag{13.2}$$

Un modelo de elección basado en esta especificación de la utilidad no implica ninguna endogeneidad. Una constante se incluye para cada producto en cada mercado, lo que absorbe ξ_{jm}. La porción restante de utilidad no observada, ε_{njm}, es independiente de las variables explicativas. Las constantes se calculan junto con los otros parámetros del modelo. En esencia, el término que causó el carácter endógeno, es decir, ξ_{jm}, se ha subsumido en la constante de producto-mercado de tal manera que ya no es parte del componente no observado de la utilidad.

El modelo de elección se completa especificando cómo $\tilde{\beta}_n$ varía entre consumidores. Denotemos la densidad de $\tilde{\beta}_n$ como $f(\tilde{\beta}_n|\theta)$, donde θ son parámetros de esta distribución que representan, por ejemplo, la varianza de los coeficientes alrededor de los valores comunes. Teniendo en cuenta que ε_{njm} es de tipo valor extremo iid, la probabilidad de elección es la de un modelo logit mixto:

$$P_{nim} = \int \left[\frac{e^{\delta_{im} + \tilde{V}(p_{im}, x_{im}, s_n, \tilde{\beta}_n)}}{\sum_j e^{\delta_{jm} + \tilde{V}(p_{jm}, x_{jm}, s_n, \tilde{\beta}_n)}} \right] f(\tilde{\beta}_n|\theta) d\tilde{\beta}_n. \tag{13.3}$$

Por lo general, $\tilde{V}$ es lineal con coeficientes $\tilde{\beta}_n$ y variables explicativas que son los atributos observados, p_{jm} y x_{jm}, interactuando quizá con los demográficos, s_n.Pueden especificarse otras distribuciones para ε_{njm}; por ejemplo, Goolsbee y Petrin (2004) asumen que ε_{njm} es una normal conjunta entre productos, de manera que la probabilidad de elección es probit. Además, si disponemos de información sobre el ranking de preferencias del consumidor respecto a las alternativas, total o parcialmente, entonces la probabilidad del ranking se especifica de manera análoga; por ejemplo, Berry, Levinsohn y Pakes (2004) y Train y Winston (2007) disponían de datos sobre el vehículo comprado por cada consumidor así como

[iii] Es posible que $\bar{V}$ incluya datos demográficos agregados, como la renta media del mercado, que varía entre mercados, pero no entre consumidores de cada mercado. Sin embargo, nos abstraemos de esta posibilidad en nuestra notación

sobre su segunda elección del vehículo, y representaron este ranking parcial mediante la inserción de la fórmula logit expandida de la Sección 7.3 dentro de los corchetes de la ecuación (13.3) en lugar de la fórmula logit estándar.

La estimación del modelo de elección descrito en (13.3) proporciona estimaciones de las constantes y de la distribución de preferencias. Sin embargo, no proporciona estimaciones de los parámetros que entran en la parte de la utilidad que es constante entre consumidores; es decir, no proporciona estimaciones de $\bar{\beta}$ en $\bar{V}$. Estos parámetros entran en la definición de las constantes de la ecuación (13.1), lo que constituye un modelo de regresión que puede ser usado para estimar el promedio de las preferencias. Es habitual expresar $\bar{V}$ como una función lineal respecto a los parámetros, de tal manera que (13.1) se convierte en

$$\delta_{jm} = \bar{\beta}'\bar{V}(p_{jm}, x_{jm}) + \xi_{jm} \tag{13.4}$$

donde $\bar{V}(\cdot)$ es una función vectorial de los atributos observados. Es posible estimar una regresión en la que la variable dependiente es la constante para cada producto en cada mercado, y en la que las variables explicativas son el precio y otros atributos observados del producto. El término de error para esta regresión es ξ_{jm}, que se correlaciona con el precio. Sin embargo, los procedimientos para manejar la endogeneidad en los modelos de regresión lineal están bien desarrollados y se describen en cualquier libro de texto de econometría estándar. En particular, la regresión (13.4) se estima a través de variables instrumentales en lugar de mínimos cuadrados ordinarios. Todo lo que se requiere para esta estimación es que el investigador tenga, o pueda calcular, algunas variables exógenas adicionales que se utilizan como instrumentos en lugar del precio endógeno. La selección de instrumentos se trata más tarde como parte del procedimiento de estimación; sin embargo, primero tenemos que afrontar un tema importante que ha estado implícito en nuestra exposición hasta ahora. Nos referimos a la forma de manejar el hecho de que podrían haber (y por lo general hay) un gran número de constantes a estimar, una para cada producto en cada mercado.

13.2.2 La contracción

Como se describió anteriormente, las constantes $\delta_{jm}\ \forall j, m$ se estiman junto con los otros parámetros del modelo de elección. Cuando existen numerosos productos y/o mercados, la estimación de este gran número de constantes puede ser difícil o inviable numéricamente, si uno trata de estimarlas de manera estándar. Por ejemplo, para la elección del vehículo, hay más de 200 marcas y modelos de vehículos nuevos cada año, lo que requiere la estimación de más de 200 constantes para la información correspondiente a cada año. Para 5 años de datos, más de un millar de constantes tendrían que ser estimadas. Si se utilizaran los procedimientos del capítulo 8 para un modelo así, cada iteración implicaría el cálculo de la pendiente respecto a, por ejemplo, más de 1000 parámetros, e invertir una matriz Hessiana de más de 1000 por 1000; también se necesitarían numerosas iteraciones ya que la búsqueda es en un espacio de parámetros de más de 1000 dimensiones.

Por suerte, no es necesario estimar las constantes de la forma estándar. BLP proporciona un algoritmo para estimarlas de forma rápida, dentro del proceso iterativo del resto de parámetros. Este procedimiento se basa en la constatación de que las constantes determinan las cuotas de mercado previstas para cada producto, y que por lo tanto se pueden fijar de tal manera que las cuotas de mercado previstas sean iguales a los porcentajes reales. Para ser precisos, sea S_{jm} sea la cuota o proporción de los consumidores en el mercado de m que eligen el producto j. Para un modelo correctamente especificado, las cuotas *predichas* en cada mercado deben ser iguales a estas cuotas reales (al menos asintóticamente). Podemos encontrar las constantes que hacen cumplir esta igualdad, es decir, que hacen que el modelo prediga las cuotas que realmente se observan. Representemos las constantes a través de un vector $\delta = \langle \delta_{jm}\ \forall j, m \rangle$. Las cuotas previstas son $\hat{S}_{jm}(\delta) = \sum_n P_{njm}/N_m$, donde la suma se extiende sobre los N_m consumidores muestreados en el mercado m. Estas cuotas previstas se expresan como una función de las constantes δ debido a que las constantes afectan a las probabilidades de elección que a su vez afectan a las cuotas previstas.

Recordemos que en la sección 2.8 se describe un procedimiento iterativo para re-calibrar las constantes en un modelo de manera que las cuotas previstas sean iguales a las cuotas reales. Empezando con cualesquiera valores de las constantes, etiquetados $\delta_{jm}^t\ \forall j, m$, las constantes se ajustan iterativamente mediante la fórmula

$$\delta_{jm}^{t+1} = \delta_{jm}^t + ln\left(\frac{S_{jm}}{\hat{S}_{jm}(\delta^t)}\right).$$

Este proceso de ajuste mueve cada constante en la dirección "correcta", en el sentido siguiente. Si, con el valor actual de la constante, la cuota real de un producto excede la cuota predicha, entonces el ratio entre cuota real y predicha (es decir, $S_{jm}/\hat{S}_{jm}(\delta^t)$) es mayor que 1 y $\ln(\cdot)$ es positivo. En este caso, la constante se ajusta al alza, para elevar la proporción predicha. Cuando la cuota real está por debajo de la predicha, el ratio está por debajo de 1 y $\ln(\cdot) < 0$, de manera que la constante se ajusta a la baja. El ajuste se repite iterativamente hasta que las cuotas previstas son iguales a las cuotas reales (dentro de un margen de tolerancia) para todos los productos en todos los mercados.

Este algoritmo se puede utilizar para estimar las constantes en lugar de estimarlas a través de los métodos habituales basados en el gradiente. Los otros parámetros del modelo sí se estiman a través de métodos basados en el gradiente, y en cada valor de prueba de estos otros parámetros (es decir, cada iteración en la búsqueda de la optimización de los valores de los otros parámetros), las constantes se ajustan de tal manera que las cuotas previstas son iguales a las cuotas reales para este valor de prueba. Básicamente, el procedimiento que había sido utilizado durante muchos años para la recalibración post-estimación de constantes se utiliza *durante* la estimación, en cada iteración para los otros parámetros.

Berry (1994) demostró que para cualesquiera valores de los otros parámetros en el modelo de elección (es decir, de θ), existe un conjunto único de constantes para las que cuotas previstas son iguales a las cuotas reales. Posteriormente BLP (1995) mostró que el proceso de ajuste iterativo es una contracción, de tal manera que la convergencia a un conjunto único de constantes está garantizada. Cuando se utiliza en el contexto de la estimación en lugar del contexto la calibración post-estimación, el algoritmo ha llegado a ser conocido como "la contracción".

Algunas notas adicionales son útiles en relación a la contracción. En primer lugar, anteriormente hemos definido las cuotas S_{jm} como las cuotas "reales". En la práctica, se pueden utilizar tanto las cuotas de mercado agregadas como las cuotas de la muestra. En algunas situaciones, los datos sobre las cuotas agregadas no están disponibles o no son fiables. Las cuotas de la muestra son consistentes con las cuotas de mercado, siempre que el muestreo sea exógeno. En segundo lugar, el procedimiento impone una restricción o condición sobre la estimación: que las cuotas predichas sean iguales a las cuotas reales. Como vimos en la sección 3.7.1, la estimación de máxima verosimilitud de un modelo logit estándar con constantes específicas de alternativa para cada producto en cada mercado, da necesariamente cuotas previstas iguales a las cuotas de la muestra. Por tanto, la condición de que las cuotas predichas sean iguales a las cuotas de la es consistente con (o más precisamente, es una característica de) la máxima verosimilitud en un modelo logit estándar. Sin embargo, para otros modelos, incluyendo probit y logit mixto, el estimador de máxima verosimilitud no equipara las cuotas previstas con las cuotas de la muestra, incluso cuando se incluyen un conjunto completo de constantes. Las constantes estimadas que se obtienen a través de la contracción, no son por tanto las estimaciones de máxima verosimilitud. Sin embargo, ya que la condición se cumple asintóticamente para un modelo correctamente especificado, imponerlo parece razonable.

13.2.3 Estimación por máxima verosimilitud simulada y variables instrumentales

Hay varias maneras de que los otros parámetros del modelo (es decir, θ y $\bar{\beta}$) puedan ser estimados. El procedimiento más fácil de conceptualizar es el utilizado por Goolsbee y Petrin (2004) y Train y Winston (2007). En estos estudios, el modelo de elección de la ecuación (13.3) se calcula en primer lugar, utilizando máxima verosimilitud simulada (MSL) con la contracción. Este paso proporciona estimaciones de los parámetros que entran en la ecuación (13.2), a saber, las constantes δ_{jm} $\forall j, m$ y los parámetros θ de la distribución de preferencias alrededor de estas constantes. La contracción se utiliza para las constantes, de manera que la maximización de la función log -verosimilitud es sólo sobre θ.

Para ser más precisos, ya que las probabilidades de elección dependen tanto de δ como de θ, esta dependencia se puede denotar funcionalmente como $P_{njm}(\delta, \theta)$. Sin embargo, para cualquier valor dado de θ, las constantes δ están completamente determinadas: son los valores que iguala las cuotas predichas y las cuotas reales cuando este valor de θ se utiliza en el modelo. Por tanto, las constantes calibradas se pueden considerar una función de θ, denotada $\delta(\theta)$. Sustituyendo en la probabilidad de elección, la probabilidad se convierte en una función únicamente de θ: $P_{njm}(\theta) = P_{njm}(\delta(\theta), \theta)$. La función log-verosimilitud se define también como una función de θ: con i_n denotando la alternativa elegida por n, la función log-verosimilitud es $LL(\theta) = \sum_n \ln P_{ni_nm}(\theta)$, donde δ se re-calcula adecuadamente para cualquier θ. Como tal, el estimador es MSL sujeto a la restricción de que las cuotas predichas sean iguales a las cuotas reales (ya sean cuotas de mercado o de la muestra, la que sea que se esté utilizando)[iv].

Una vez que el modelo de elección es estimado, las constantes estimadas se utilizan en la regresión lineal (13.4), que repetimos aquí por conveniencia:

$$\delta_{jm} = \bar{\beta}' \bar{v}(p_{jm}, x_{jm}) + \xi_{jm}$$

Las constantes estimadas del modelo de elección son las variables dependientes en esta regresión, y el precio y otros atributos observados de los productos son las variables explicativas. Puesto que el precio es endógeno en esta regresión, se estima a través de variables instrumentales en lugar de mínimos cuadrados ordinarios. Los instrumentos incluyen los atributos diferentes al precio observados de los productos, x_{jm}, además de al menos un instrumento adicional en lugar del precio. Si nos referimos a todos los instrumentos a través del vector z_{jm}, el estimador de variables instrumentales es el valor de $\bar{\beta}$ que satisface

$$\sum_j \sum_m [\hat{\delta}_{jm} - \bar{\beta}' \bar{v}(p_{jm}, x_{jm})] z_{jm} = 0,$$

donde $\hat{\delta}_{jm}$ es la constante estimada a partir del modelo de elección. Podemos reorganizar esta expresión para expresar el estimador de forma cerrada, como suele mostrarse por lo general en los libros de texto sobre regresión, y como se indica en la Sección 10.2.2:

[iv] Desde un punto de vista de programación, la maximización implica iteraciones dentro de iteraciones. El procedimiento de optimización itera sobre valores de θ en la búsqueda del máximo de la función log-verosimilitud. En cada valor de prueba de θ, la contracción itera sobre los valores de las constantes, ajustándolas hasta que las cuotas previstas son iguales a las cuotas reales en ese valor de prueba de θ.

$$\hat{\hat{\beta}} = \left(\sum_j \sum_m z_{jm} \bar{v}\left(p_{jm}, x_{jm}\right)' \right)^{-1} \left(\sum_j \sum_m z_{jm} \hat{\delta}_{jm} \right).$$

Si el investigador lo desea, la eficiencia puede ser mejorada teniendo en cuenta la covarianza entre las constantes estimadas, a través de mínimos cuadrados generalizados (GLS); véase, por ejemplo, Greene (2000), en relación a la estimación GLS de modelos de regresión lineal.

El problema surge necesariamente en relación a que variables utilizar como instrumentos. Es frecuente, como ya se mencionó, utilizar como instrumentos los atributos observados diferentes del precio bajo el supuesto de que son exógenos[v]. BLP (1994) sugirieron el uso de instrumentos basados en los conceptos de precio. En particular, cada fabricante fijará el precio de cada uno de sus productos de una manera que tenga en consideración la sustitución con sus otros productos, así como la sustitución con productos de otras empresas. Por ejemplo, cuando una empresa está considerando la posibilidad de un aumento de precio para uno de sus productos, los consumidores que abandonarán el consumo de este producto para consumir otro de los productos de la misma empresa no representan la parte más importante de la pérdida (y de hecho hasta podrían representar una ganancia, en función de los márgenes de beneficio) como sí lo son los consumidores que pasarán a consumir productos de otras empresas. Sobre la base de estas ideas, BLP propuso dos instrumentos: los atributos medios distintos del precio de otros productos del mismo fabricante y los atributos medios distintos del precio de los productos de otras empresas. Por ejemplo, en el contexto de la elección de un vehículo en la que, por ejemplo, el peso del vehículo es un atributo observado, los dos instrumentos para el Toyota Camry para un año determinado son (1) el peso medio de todas los modelos de vehículos Toyota en ese año y (2) el peso promedio de todos los vehículos que no son Toyota en ese año[vi]. Train y Winston (2007) utilizaron una extensión de estos instrumentos que refleja hasta qué punto cada producto se diferencia de otros productos del mismo fabricante y de otros fabricantes. En particular, emplearon la suma de las diferencias al cuadrado entre el producto y cada uno de los otros productos del mismo fabricante y de otros fabricantes. Estos dos instrumentos para el producto j en el mercado m son $\mathrm{z}_{jm}^{1} = \sum_{k \in K_{jm}} (x_{jm} - x_{km})^2$, donde K_{jm} es el conjunto de productos ofrecidos en el mercado m por la empresa que produjo el producto j, y $\mathrm{z}_{jm}^{2} = \sum_{k \in S_{jm}} (x_{jm} - x_{km})^2$, donde S_{jm} es el conjunto de productos ofrecidos en el mercado m por todas las empresas, excepto la empresa que produjo el producto j.

En otros contextos, otros instrumentos son apropiados. Goolsebee y Petrin (2004) examinaron las elecciones de los hogares entre las opciones disponibles de canales de televisión, siendo cada ciudad un mercado distinto con diferentes precios y diferentes características distintas del precio, tanto para TV por cable y como por difusión aérea. Siguiendo la práctica sugerida por Hausman (1997), utilizaron los precios ofrecidos en otras ciudades por la misma compañía como instrumentos para cada ciudad. Este instrumento para la ciudad m es $\mathrm{z}_{jm} = \sum_{m' \in K_{jm}} p_{jm'}$, donde K_m es el conjunto de otras ciudades que son atendidas por el operador franquiciado en la ciudad m. El concepto que origina este planteamiento es que los atributos no observados de la televisión por cable en una ciudad determinada (como la

[v] Este supuesto es en gran parte algo conveniente, ya que en general se podría esperar que los atributos no observados de un producto estuviesen relacionados no sólo con el precio sino también con los atributos observados distintos del precio. Sin embargo, un modelo en el que todos los atributos observados son tratados como endógenos dejan poco margen para utilizar instrumentos

[vi] En vez de los promedios, también pueden usarse las sumas, con el número de productos de la misma marca y de otras marcas (es decir, los denominadores de los promedios) entrando también como instrumentos.

calidad de la programación para la franquicia en esa ciudad) están correlacionados con el precio de la televisión por cable en esa ciudad, pero no están correlacionados con el precio de la televisión por cable en otras ciudades. Ellos también incluyeron la tasa de la franquicia de cable impuesta por la ciudad (es decir, impuestos) y la densidad de población de la ciudad.

13.2.4 Estimación por GMM

Como se dijo anteriormente, es posible usar varios métodos para la estimación, no sólo máxima verosimilitud. BLP (1995,2004), Nevo (2001) y Petrin (2002) utilizaron un estimador basado en un método generalizado de momentos (GMM). Este procedimiento es una versión generalizada del método de momentos simulados (MSM) que se describe en el capítulo 10, aumentado por los momentos de la ecuación de regresión. Las condiciones de momentos se crean a partir de las probabilidades de elección como

$$\sum_n \sum_j \left(d_{njm} - P_{njm}\right) z_{njm} = 0, \tag{13.5}$$

donde d_{njm} es la variable dependiente, que es 1 si el consumidor n en el mercado m escoge la alternativa j y 0 en caso contrario, y z_{njm} es un vector de instrumentos que varía entre productos y mercados (como los atributos observados distintos del precio y las funciones de los mismos), así como entre consumidores en cada mercado (tales como las variables demográficas interactuando con atributos distintos del precio). En la estimación, las probabilidades exactas se sustituyen por su versión simulada. Tenga en cuenta que esta es la misma fórmula que la mostrada en la sección 10.2.2, en relación a la estimación MSM de modelos de elección. Estas condiciones de momentos, cuando se satisfacen, implican que la media observada de los instrumentos para las alternativas elegidas, $\sum_{\mathrm{n}} \sum_{\mathrm{j}} d_{njm} z_{njm}/N_m$, es igual a la media predicha por el modelo, $\sum_{\mathrm{n}} \sum_{\mathrm{j}} P_{njm} z_{njm}/N_m$. Observe que, tal como se describe en la sección 3.7, para un modelo logit estándar, estas condiciones de momentos son la condición de primer orden para la estimación de máxima verosimilitud, donde los instrumentos son las variables explicativas del modelo. En otros modelos, esta condición de momentos no es igual a la condición de primer orden de la máxima verosimilitud, de tal manera que se produce una pérdida de eficiencia. Sin embargo, tal como se describe en el capítulo 10 en la comparación entre MSL y MSM, la simulación de estos momentos es no sesgada dado un simulador no sesgado de la probabilidad de elección, de manera que MSM es consistente para un número fijo de valores extraídos al azar en la simulación. En contraste, MSL es consistente sólo cuando se considera que el número de valores al azar aumenta con el tamaño de la muestra

La condición de momentos para la ecuación de regresión se crean como

$$\sum_j \sum_m \xi_{jm} z_{jm} = 0,$$

donde z_{jm} son instrumentos que varían entre productos y mercados, pero no entre las personas en cada mercado (como los atributos observados distintos del precio y las funciones de los mismos). Estos momentos pueden ser re-escritos para incluir los parámetros del modelo de forma explícita, suponiendo una especificación lineal para $\bar{V}$:

$$\sum_j \sum_m \left[\delta_{jm} - \bar{\beta}' \bar{v}\left(p_{jm}, x_{jm}\right)\right] z_{jm} = 0. \tag{13.6}$$

Como vimos en la sección previa, estos momentos definen el estimador de variables instrumentales estándar de los coeficientes de regresión.

Los parámetros del sistema son $\bar{\beta}$, que capturan elementos de preferencias que son iguales para todos los consumidores, y θ, que representa la variación en las preferencias entre consumidores. Las constantes δ_{jm} también se pueden considerar parámetros, ya que se estiman junto con los otros parámetros. Alternativamente, se pueden considerar funciones de los otros parámetros, como vimos en la subsección anterior, calculadas para igualar las cuotas predichas y reales para cualesquiera valores dados de los otros parámetros. Para las distinciones en los párrafos siguientes, consideramos que los parámetros son $\bar{\beta}$ y θ, sin las constantes. Bajo esta terminología, la estimación se realiza de la siguiente manera.

Si el número de condiciones de momentos en (13.5) y (13.6) combinados es igual al número de parámetros (y tiene una solución), entonces el estimador se define como el valor de los parámetros que satisface todas las condiciones de momentos. El estimador es el MSM descrito en el capítulo 10 aumentado con momentos adicionales para la regresión. Si el número de condiciones de momentos supera el número de parámetros, entonces no hay ningún conjunto de valores de los parámetros que pueda satisfacer todas las condiciones. En este caso, el número de condiciones independientes se reduce mediante el uso de un método generalizado de momentos (GMM)[vii]. El estimador GMM se describe más fácilmente definiendo los momentos específicos de observación:

$$g^1_{njm} = (d_{njm} - P_{njm})z_{njm},$$

que son los términos en (13.5), y

$$g^2_{njm} = [\delta_{jm} - \bar{\beta}'\bar{v}(p_{jm}, x_{jm})]z_{jm},$$

que son los términos en (13.6). Agrupe estos dos vectores en uno solo, $g_{njm} = \langle g^1_{njm}, g^2_{njm}\rangle$, señalando que el segundo conjunto de momentos se repite para cada consumidor en el mercado m. Las condiciones de momentos se pueden escribir de manera sucinta como $\mathrm{g} = \sum_n \sum_j g_{njm} = 0$. El estimador GMM es el valor del parámetro que minimiza la forma cuadrática $g'\Theta^{-1}g$, donde Θ es una matriz de ponderación definida positiva. La covarianza asintótica de g es la matriz de ponderación óptima, calculada como $\sum_n \sum_j g_{njm}g'_{njm}$. Ruud (2000, capítulo 21) proporciona un estudio útil sobre estimadores GMM.

13.3 Lado de la oferta

El lado de la oferta de un mercado puede ser importante por varias razones. En primer lugar, en la medida en que los precios del mercado están determinados por la interacción entre la oferta y la demanda, la predicción de los resultados del mercado debido a un cambio en las condiciones del mismo requiere una comprensión tanto de la oferta como de la demanda. Un ejemplo importante se plantea en el contexto del análisis antimonopolio de las fusiones de empresas. Antes de la fusión, cada empresa fija los precios de sus propios productos con el fin de maximizar sus propios beneficios. Cuando dos empresas se fusionan, establecen los precios de sus productos para maximizar su beneficio conjunto, lo que puede producir, y por lo general así sucede, precios diferentes a los existentes cuando las empresas competían entre sí. Una tarea central del Departamento de Justicia (*Department of Justice*) y la Comisión Federal de Comercio (*Federal Trade Commission*) en el momento de decidir si aprueba o no

[vii] Cuando se utilizan probabilidades simuladas en las condiciones, el procedimiento puede ser denominado quizá de forma más exacta como estimador mediante el método generalizado de momentos simulados (GMSM). Sin embargo, nunca he visto usar este término; en su lugar, GMM se utiliza para denotar tanto los momentos simulados y como no simulado

una fusión es pronosticar el impacto de la misma sobre los precios. Esta tarea conlleva generalmente modelar el lado de la oferta, es decir, los costos marginales y la política de precios de las empresas, así como la demanda de cada producto como una función del precio.

En segundo lugar, es de esperar en muchos contextos que las empresas fijen sus precios, no sólo en base al costo marginal o a algún margen aplicado sobre el costo marginal, sino también sobre la base de la demanda de su producto y el impacto de los cambios de precios en su demanda. En estos contextos, los precios observados contienen información acerca de la demanda de los productos y la elasticidad respecto al precio. Esta información, si se extrae correctamente, puede ser utilizada en la estimación de los parámetros de la demanda. El investigador podría, por lo tanto, optar por examinar la oferta como un aspecto de la estimación de la demanda, aun cuando entre los objetivos del investigador no esté la predicción de la oferta *per se*.

En los párrafos siguientes, se describen varios tipos de comportamiento de los precios que pueden observarse en los mercados y se muestra cómo se puede combinar la especificación de este comportamiento con la estimación de la demanda vista anteriormente. En cada caso, se requieren algunas suposiciones sobre el comportamiento de las empresas. El investigador debe decidir, por lo tanto, si incorpora la oferta en el análisis bajo estas suposiciones o si por el contrario estima la demanda sin considerar la oferta, utilizando los métodos descritos en la sección anterior. El investigador se enfrenta a una disyuntiva inherente. Incorporar la oferta en el análisis tiene el potencial de mejorar las estimaciones de la demanda y ampliar el uso del modelo. Sin embargo, implica hacer suposiciones sobre el comportamiento de los precios que podrían no ser reales, de forma que estimar la demanda sin considerar la oferta podría ser más seguro.

13.3.1 Costo Marginal

Un principio básico de la teoría económica es que los precios dependen del costo marginal (*marginal cost*, MC). El costo marginal de un producto depende de sus atributos, incluyendo tanto los atributos que son observados por parte del investigador, x_{jm}, como los que no son observador, ξ_{jm}. El costo marginal también depende de los precios de los medios de producción, como alquileres y salarios, y otras "palancas de cambio de los costos" (*cost shifters*). El investigador observa algunas de estas palancas de cambio de los costos, etiquetadas como c_{jm} y no observa otras. A menudo se asume que el costo marginal es separable en términos no observadas, de tal manera que toma la forma de una regresión:

$$MC_{jm} = W(x_{jm}, c_{jm}, \gamma) + \mu_{jm}, \tag{13.7}$$

donde W(·) es una función con parámetros γ. El término de error en esta ecuación, μ_{jm}, depende en general de los atributos no observados ξ_{jm} y de otras palancas de cambio de los costos no observadas.

Los costos marginales se relacionan con los precios de manera diferente, dependiendo del mecanismo de fijación de precios que esté operando en el mercado. Examinamos a continuación las posibilidades más habituales.

13.3.2 Precios MC

En los mercados con competencia perfecta, el precio de cada producto es presionado a la baja, en equilibrio, hacia su costo marginal. Por lo tanto, los precios en un mercado competitivo con costos marginales separables son

$$p_{jm} = W(x_{jm}, c_{jm}, \gamma) + \mu_{jm}, \tag{13.8}$$

Esta ecuación de precios puede ser utilizada conjuntamente con cualquiera de las dos formas descritas anteriormente para la estimación de la demanda, cada uno de los cuales se abordan a continuación.

MSL e IV, con precios MC

Supongamos que se utiliza el método descrito en la sección 13.2.3; es decir, el modelo de elección se estima por máxima verosimilitud y luego se hace una regresión de las constantes estimadas respecto a los atributos del producto, usando variables instrumentales para tener en cuenta la endogeneidad de los precios. Con esta estimación, las variables que entran en la ecuación de precios MC son los instrumentos adecuados para usar en las variables instrumentales. Cualquieras palancas de cambio de costos observadas c_{jm} sirven como instrumentos, junto con los atributos observados distintos del precio x_{jm}. En esta configuración, la ecuación de precios simplemente proporciona información acerca de los instrumentos adecuados. Como la ecuación de precios no depende de parámetros de la demanda, no proporciona ninguna información, más allá de los instrumentos, que pueda mejorar la estimación de los parámetros de la demanda.

El investigador podría querer estimar la ecuación de precios a pesar de que no incluya parámetros de la demanda. Por ejemplo, podría querer predecir los precios y la demanda en condiciones de cambio en las palancas de costos. En este caso, la ecuación (13.8) se estima mediante mínimos cuadrados ordinarios. La estimación de la demanda y la oferta se realiza tres pasos: (1) se estima el modelo de elección mediante MSL, (2) se estima la regresión de las constantes de atributos observados del producto mediante variables instrumentales y (3) se estima la ecuación de precios mediante mínimos cuadrados ordinarios. De hecho, si las variables que entran en la ecuación de fijación de precios son los únicos instrumentos, entonces la ecuación de precios se calcula implícitamente como parte del paso 2, incluso si el investigador no estima la ecuación de precios de manera explícita en el paso 3. Recuerde de los textos estándar relativos a regresión que la estimación de variables instrumentales es equivalente la de mínimos cuadrados en dos etapas. En nuestro contexto, las variables instrumentales en el paso 2 se pueden obtener en dos etapas: (i) estimando la ecuación de precios por mínimos cuadrados ordinarios y utilizando los coeficientes estimados para predecir los precios, y luego (ii) haciendo una regresión de las constantes de los atributos del producto por mínimos cuadrados ordinarios utilizando el precio pronosticado como variable explicativa en lugar del precio observado. Por lo tanto, el proceso conjunto de estimación es el siguiente : (1) estimamos el modelo de elección, (2) estimamos la ecuación de precios por mínimos cuadrados ordinarios y (3) estimamos la ecuación para las constantes por mínimos cuadrados ordinarios utilizando los precios predichos en lugar de los precios reales.

GMM con precios MC

Si se utiliza la estimación GMM, las variables explicativas en la ecuación de precios se convierten en instrumentos $z_{jm} = \langle x_{jm}, c_{jm} \rangle$, que entran en los momentos indicados en las ecuaciones (13.5) y (13.6). La ecuación de precios proporciona condiciones de momentos adicionales basados en estos instrumentos:

$$\sum_j \sum_m \left[p_{jm} - W\left(x_{jm}, c_{jm}, \gamma\right)\right] z_{jm} = 0.$$

Los parámetros de la demanda pueden ser estimados sin estos momentos adicionales. O estos momentos adicionales pueden ser incluidos en la estimación GMM, junto con los definidos en (13.5) y (13.6). El estimador es el valor de los parámetros del modelo de demanda y de la ecuación de precios que minimiza $g'\Theta^{-1}g$, donde g incluye los momentos de la ecuación de precios y Θ incluye su covarianza.

13.3.3 Margen fijo sobre el costo marginal

Las empresas podrían fijar sus precios como un margen fijo sobre el costo marginal, siendo este margen independiente de la demanda. Esta forma de fijación de precios es considerada por la gente de negocios como una práctica omnipresente. Por ejemplo, Shim y Sudit (1995) informan de que esta forma de fijación de precios es utilizada por más del 80 por ciento de los directivos de las empresas fabricantes. Por supuesto, es difícil saber cómo interpretar las declaraciones de los directivos acerca de su política de precios, ya que cada gerente puede pensar que ellos fijan sus precios con un margen fijo por encima del costo marginal y sin embargo, el tamaño de este margen " fijo" puede ser que varíe entre directivos en relación con la demanda de los productos de los distintos fabricantes.

Esta forma de fijación de precios tiene las mismas implicaciones para la estimación de la demanda que los precios MC. El precio es $p_{jm} = \mathrm{k}MC_{jm}$ por alguna constante k. La ecuación de fijación de precios es la misma de antes, la ecuación (13.8), pero con los coeficientes incorporando ahora el factor k. Todos los otros aspectos de la estimación, ya sea usando MSL y IV o GMM, son los mismas que en el caso de precios MC.

13.3.4 Precios de monopolio y equilibrio de Nash para empresas con un solo producto

Considere una situación en la que cada producto es ofrecido por una empresa independiente. Si sólo hay un producto en el mercado, entonces la empresa es monopolista. Se supone que el monopolista fija el precio para maximizar sus beneficios, manteniendo fijos todos los demás precios de la economía. Si hay múltiples productos, entonces las empresas son oligopolistas. Asumimos que cada oligopolista maximiza sus beneficios, dados los precios del resto de empresas. Este supuesto implica que cada oligopolista utiliza la misma condición de maximización de beneficios de un monopolista, manteniendo fijos todos los demás precios (incluidos los precios de sus rivales). El equilibrio de Nash se produce cuando ninguna empresa es inducida a cambiar su precio, dados los precios del resto de empresas.

Sean p_j y q_j el precio y la cantidad del producto j, donde el subíndice que se refiere a los mercados se omite por conveniencia, ya que estamos examinando el comportamiento de las empresas en un mercado determinado. Las ganancias de la empresa son $\pi_j = p_j q_j - \mathrm{TC}(q_j)$, donde TC es el costo total. Los beneficios son máximos cuando

$$\frac{d\pi_j}{dp_j} = 0$$

$$\frac{d(p_j q_j)}{dp_j} - MC_j \frac{dq_j}{dp_j} = 0$$

$$q_j - p_j \frac{dq_j}{dp_j} = MC_j \frac{dq_j}{dp_j}$$

$$p_j + q_j \left(\frac{dq_j}{dp_j}\right)^{-1} = MC_j \qquad (13.9)$$

$$p_j + p_j \frac{q_j}{p_j} \left(\frac{dq_j}{dp_j}\right)^{-1} = MC_j$$

$$p_j + (p_j/e_j) = MC_j,$$

donde MC_jes el costo marginal del producto j y e_j es la elasticidad de la demanda para el producto j respecto a su precio. Esta elasticidad depende de todos los precios, debido a que la elasticidad para un producto es diferente para diferentes niveles de precio (y por lo tanto, diferentes cantidades) para ese producto así como los otros productos. Tenga en cuenta que la elasticidad es negativa, lo que implica que $p_j + (p_j/e_j)$ es menor que p_j de tal manera que el precio está por encima del costo marginal, como es de esperar[viii]. Sustituyendo la especificación en (13.7) para MC y añadiendo subíndices para los diferentes mercados, tenemos

$$p_{jm} + (p_{jm}/e_{jm}) = W(x_{jm}, c_{jm}, \gamma) + \mu_{jm}, \tag{13.10}$$

que es igual a la ecuación de precios bajo precios MC, (13.8), excepto que el lado izquierdo añade (p_{jm}/e_{jm}) al precio.

Los parámetros de costos marginales pueden estimarse después de los parámetros de demanda. En el contexto de MSL y IV, el proceso consiste en los mismos tres pasos que usamos con precios MC: (1) estimar el modelo de elección mediante MSL, (2) estimar la regresión de las constantes respecto al precio y a otros atributos mediante variables instrumentales, y (3) estimar la ecuación de precios mediante mínimos cuadrados ordinarios. El único cambio es que ahora la variable dependiente en la ecuación de precios no es el precio en sí, sino $p_{jm} + (p_{jm}/e_{jm})$. Este término incluye la elasticidad de la demanda, que no es observado directamente. Sin embargo, dados unos parámetros estimados de la demanda en los pasos 1 y 2, la elasticidad de la demanda puede ser calculada y utilizada en el paso 3.

El hecho de que la ecuación de fijación de precios incluya la elasticidad de la demanda implica que los precios observados contienen información acerca de los parámetros de demanda. El procedimiento de estimación secuencial que acabamos de describir no utiliza esta información. En el contexto de GMM, utilizar esta información adicional es sencillo, al menos conceptualmente. Se definen condiciones de momentos adicionales a partir de la ecuación de precios como

$$\sum_j \sum_m \left(p_{jm} + (p_{jm}/e_{jm}) - W(x_{jm}, c_{jm}, \gamma)\right) z_{jm} = 0.$$

La única diferencia con el procedimiento utilizado con precios MC es que ahora las condiciones de momentos incluyen el término adicional (p_{jm}/e_{jm}). En cada valor de prueba de los parámetros en la estimación GMM, la elasticidad se calcula y se inserta en este momento.

13.3.5 Precios de monopolio y equilibrio de Nash para empresas multiproducto

Ahora generalizaremos el análisis de la sección anterior para permitir que cada empresa pueda vender más de un producto en un mercado. Si sólo hay una empresa que ofrece todos los productos en el mercado, esta empresa es un monopolista multiproducto. De lo contrario, el mercado es un oligopolio con empresas multiproducto. El mercado de los vehículos nuevos es un ejemplo representativo, donde cada fabricante, como Toyota, ofrece numerosas marcas y modelos de vehículos. La regla de fijación de precios se diferencia de la situación de un único producto por empresa en que ahora cada empresa

[viii] La condición puede ser reorganizada para que tome la forma que utilizada a menudo para monopolistas y oligopolistas de un único producto: $(p_j - MC_j)/p_j = -1/e_j$, donde el margen expresado como un porcentaje del precio está inversamente relacionado con la magnitud de la elasticidad.

debe tener en cuenta el impacto de su precio de un producto en la demanda de sus otros productos. Emplearemos el supuesto estándar, como antes, de que cada empresa fija precios con el fin de maximizar los beneficios, dados los precios de los productos de las otras empresas.

Considere una empresa que ofrece un conjunto de k productos. Los beneficios de la empresa son $\pi = \sum_{\mathrm{j} \in \mathrm{K}} p_j q_j - \mathrm{TC}(q_j\ \forall \mathrm{j} \in \mathrm{K})$. La empresa elige el precio del producto $\mathrm{j} \in \mathrm{K}$ que maximiza sus beneficios:

$$d\pi/dp_j = 0$$

$$q_j + \sum_{k \in K} (p_k - MC_k)\left(dq_k/dp_j\right) = 0.$$

Condiciones análogas aplican simultáneamente a todos los productos de la empresa. Las condiciones para todos los productos de la empresa se pueden expresar de forma sencilla mediante el uso de notación matricial. Agrupemos los precios de los K productos en el vector p, las cantidades en el vector q y los costos marginales en el vector MC. Definimos una matriz $\mathrm{K} \times \mathrm{K}$ con las derivadas de la demanda respecto al precio, D, donde el elemento (i, j)-ésimo es $\left(dq_j/dp_i\right)$. Siendo así, los precios que maximizan los beneficios de los productos de la empresa satisfacen

$$q + D(p - MC) = 0$$

$$D^{-1}q + p - MC = 0$$

$$p + D^{-1}q = MC.$$

Tenga en cuenta que esta última ecuación simplemente es una versión generalizada de la regla del monopolista mono-producto dada en (13.9). Se maneja de la misma manera en la estimación. Con estimación MSL y IV, se estima después del modelo de la demanda, usando los parámetros de la demanda para calcular D^{-1}. Con GMM, la ecuación de precios se incluye como una condición de momentos adicional, calculando D^{-1} para cada valor de prueba de los parámetros.

13.4 Funciones de control

El enfoque BLP no siempre es aplicable. Si las cuotas de mercado observadas para algunos productos en algunos mercados son cero, el enfoque BLP no puede ser implementado, ya que las constantes para estos mercados de productos no están identificadas. (Cualquier constante finita genera una cuota de mercado prevista estrictamente positiva, que superaría la cuota real de cero). Un ejemplo es el estudio a cargo de Martin (2008) sobre la elección que hacen los consumidores entre bombillas incandescentes y bombillas fluorescentes compactas (*compact fluorescent lightbulbs*, CFLs), en el que la publicidad y las promociones se produjeron con una frecuencia semanal y con variación entre diferentes tiendas y, sin embargo, era común que una tienda no vendiese ningún CFL en una semana determinada. La endogeneidad también puede surgir entre los propios decisores y no sólo entre mercados (es decir, grupos de decisiones), de tal manera que la endogeneidad puede no quedar absorbida por las constantes de producto-mercado. Por ejemplo, supongamos que las personas a las que les gusta el transporte público optan por vivir cerca de zonas de tráfico de tal manera que el tiempo de tránsito en su elección del medio de transporte es endógeno. No es posible estimar constantes para cada decisor, ya que las constantes serían infinitas (para la alternativa elegida) e negativamente infinitas (para las alternativas no elegidas), prediciendo perfectamente las elecciones, sin dejar información para la estimación de parámetros. Aún cuando la aproximación BLP pudiese implementarse, el investigador podría querer evitar la complicación de aplicar la contracción que habitualmente se requiere en el enfoque BLP.

Algunas de las alternativas que se han implementado para resolver este problema incluyen el enfoque basado en la función de control, que se aborda en esta sección, y una versión específicamente diseñada de la aproximación mediante función de control que utiliza máxima verosimilitud con información completa, que se trata en la siguiente sección.

La configuración para el enfoque basado en la función de control sigue muy de cerca la especificación de modelos de regresión de ecuaciones simultáneas. Sin embargo, dado que los modelos de elección son no lineales, es necesario tratar una capa adicional de complejidad, y esta complejidad adicional puede restringir la aplicabilidad del método más de lo que podríamos esperar de entrada dada la analogía con los modelos lineales.

Denotemos la variable explicativa endógena para el decisor n y la alternativa j como y_{nj}. No diferenciamos los mercados, como en el enfoque BLP, porque permitimos la posibilidad de que las variables endógenas varíen entre decisores y no sólo entre mercados. La variable endógena podría ser el precio, el tiempo de tránsito o lo que sea relevante en un caso concreto dado. En las siguientes secciones, se abordan los problemas que pueden surgir cuando la variable endógena es el precio.

La utilidad que el consumidor n obtiene del producto j se expresa como

$$U_{nj} = V\left(y_{nj}, x_{nj}, \beta_n\right) + \varepsilon_{nj}, \tag{13.11}$$

donde x_{nj} son variables exógenas observadas relativas a la persona n y el producto j (incluyendo datos demográficos observados). El término no observado ε_{nj} no es independiente de y_{nj} como se requiere para la estimación estándar. Dejemos que la variable explicativa endógena se exprese como una función de los instrumentos observados y de los factores no observados:

$$y_{nj} = W\left(z_{nj}, \gamma\right) + \mu_{nj}, \tag{13.12}$$

donde ε_{nj} y μ_{nj} son independientes de z_{nj}, pero μ_{nj} y ε_{nj} están correlacionados. La correlación entre μ_{nj} y ε_{nj} implica que y_{nj} y ε_{nj} están correlacionados, que es lo que nos concierne. Asumimos para nuestra explicación inicial que μ_{nj} y ε_{nj} son independientes para todo los $k \neq$ j.

Consideremos ahora la distribución de ε_{nj} condicionada a μ_{nj}. Si esta distribución condicionada toma una forma conveniente, entonces podemos utilizar un enfoque de función de control para estimar el modelo. Descompongamos ε_{nj} en su media condicionada a μ_{nj} y en desviaciones alrededor de esta media: $\varepsilon_{nj} = E\left(\varepsilon_{nj}\middle|\mu_{nj}\right) + \tilde{\varepsilon}_{nj}$. Por construcción, las desviaciones no se correlacionan con μ_{nj} y por lo tanto no se correlacionan con y_{nj}. La esperanza condicionada es una función de μ_{nj} (y quizás de otras variables); se denomina la función de control y se denotada como $\mathrm{CF}(\mu_{nj}, \lambda)$, donde λ son los parámetros de esta función. El caso más simple es aquel en el que $E\left(\varepsilon_{nj}\middle|\mu_{nj}\right) = \lambda\mu_{nj}$, de manera que la función de control es simplemente μ_{nj} por un coeficiente a estimar. Expondremos las diferentes motivaciones para usar diversas funciones de control más adelante. Sustituyendo la media condicionada y las desviaciones en la ecuación de utilidad, tenemos

$$U_{nj} = V\left(y_{nj}, x_{nj}, \beta_n\right) + CF\left(\mu_{nj}, \lambda\right) + \tilde{\varepsilon}_{nj}. \tag{13.13}$$

Las probabilidades de elección se obtienen de la distribución condicionada de las desviaciones $\tilde{\varepsilon}_{nj}$. Definamos $\tilde{\varepsilon}_n = \langle\tilde{\varepsilon}_{nj}\forall j\rangle$ y $\mu_n = \langle\mu_{nj}\forall j\rangle$. La distribución condicionada de $\tilde{\varepsilon}_n$ se denota como $g(\tilde{\varepsilon}_n|\mu_n)$ y la distribución de β_n es $f(\beta_n|\theta)$. La probabilidad de elección es entonces

$$P_{nj} = Prob(U_{nj} > U_{nk}\ \forall k \neq j)$$

(13.14) $$= \int\int I(V_{nj} + CF_{nj} + \tilde{\varepsilon}_{nj} > V_{nk} + CF_{nk} + \tilde{\varepsilon}_{nk}\ \forall k \neq j) g(\tilde{\varepsilon}_n|\mu_n) f(\beta_n|\theta) d\tilde{\varepsilon} d\beta_n,$$

donde se utilizan las siguientes abreviaturas:

$$V_{nj} = V(p_{nj}, x_{nj}, \beta_n)$$

$$CF_{nj} = CF(\mu_{nj}, \lambda).$$

Este es un modelo de elección como cualquier otro, con la función de control formando parte como variable explicativa adicional. Tenga en cuenta que la integral es sobre la distribución condicionada de $\tilde{\varepsilon}$ en lugar del ε original. Por construcción, $\tilde{\varepsilon}$ no está correlacionado con la variable endógena, mientras que el ε original sí lo estaba. Básicamente, la parte de ε que está correlacionada con y_{nj} se introduce explícitamente como variable explicativa adicional, concretamente, la función de control, de tal manera que la parte restante no está correlacionada.

El modelo se estima en dos etapas. En primer lugar, se estima la ecuación (13.12). Es una regresión con la variable endógena como variable dependiente y con instrumentos exógenos como variables explicativas. Los residuos de esta regresión proporcionan estimaciones de las μ_{nj}s. Estos residuos se calculan como $\hat{\mu}_{nj} = y_{nj} - W(z_{nj}, \hat{\gamma})$ usando los parámetros estimados $\hat{\gamma}$. En segundo lugar, el modelo de elección se estima con $\hat{\mu}_{nj}$ entrando en la función de control. Es decir, las probabilidades de elección en (13.14) se estiman por máxima verosimilitud, con $\hat{\mu}_{nj}$ y/o una función paramétrica de $\hat{\mu}_{nj}$ entrando como variables explicativas adicionales.

La cuestión central con el enfoque de la función de control es la especificación de la propia función de control y de la distribución condicionada de $\tilde{\varepsilon}_n$. En algunas situaciones, hay formas naturales de especificar estos elementos del modelo. En otras situaciones, es difícil o incluso imposible, especificarlos de una manera que represente la realidad de forma significativa. La aplicabilidad del enfoque depende de que el investigador sea capaz de especificar estos términos de manera significativa.

Algunos ejemplos de especificaciones de la función de control son las siguientes:

1. Sea

   $U_{nj} = V(y_{nj}, x_{nj}, \beta_n) + \varepsilon_{nj}$

   $y_{nj} = W(z_{nj}, \gamma) + \mu_{nj}$

   y asumamos que ε_{nj} y μ_{nj} siguen una distribución normal conjunta con media cero y matriz de covarianza constante para todo j. Debido a las propiedades de las distribuciones normales, la esperanza de ε_{nj} condicionada a μ_{nj} es $\lambda\mu_{nj}$, donde λ refleja la covarianza, y las desviaciones alrededor de la media, $\tilde{\varepsilon}_{nj}$, son normales con varianza constante. En este caso, la función de control es $\mathrm{CF}(\mu_{nj}, \lambda) = \lambda\mu_{nj}$. La utilidad es

   $U_{nj} = V(y_{nj}, x_{nj}, \beta_n) + \lambda\mu_{nj} + \varepsilon_{nj}$

   El modelo de elección es probit, con el residuo de la ecuación de precios como una variable adicional. Tenga en cuenta sin embargo que la varianza de $\tilde{\varepsilon}_{nj}$ difiere de la varianza de ε_{nj}, de manera que la escala en el modelo probit estimado es diferente de la escala original. Si $\beta_n$es aleatorio, entonces el modelo es un probit mixto.

2. Supongamos que ε_{nj} consta de una parte distribuida normalmente que se correlaciona con y_{nj} y una parte que se distribuye valor extremo iid. En concreto,

   $U_{nj} = V(y_{nj}, x_{nj}, \beta_n) + \varepsilon_{nj}^1 + \varepsilon_{nj}^2$

   $y_{nj} = W(z_{nj}, \gamma) + \mu_{nj}$

   donde ε_{nj}^1 y μ_{nj} son conjuntamente normales y ε_{nj}^2 es valor extremo iid. La distribución condicionada de ε_{nj}^1 es, como en el ejemplo anterior, normal con media $\lambda\mu_{nj}$ y varianza constante. Sin embargo, la distribución condicionada de ε_{nj}^2 es igual a su distribución no condicionada dado que ε_{nj}^2 y μ_{nj} son independientes. La utilidad se convierte en

   $U_{nj} = V(y_{nj}, x_{nj}, \beta_n) + \lambda\mu_{nj} + \tilde{\varepsilon}_{nj}^1 + \varepsilon_{nj}^2$

   donde $\tilde{\varepsilon}_{nj}^1$ es normal con media cero y varianza constante. Este componente de error puede expresarse como $\tilde{\varepsilon}_{nj}^1 = \sigma\eta_{nj}$, donde η_{nj}, es una normal estándar. La utilidad se convierte así en

   $U_{nj} = V(y_{nj}, x_{nj}, \beta_n) + \lambda\mu_{nj} + \sigma\eta_{nj} + \varepsilon_{nj}^2.$

   La probabilidad de elección es la de un modelo logit mixto, mezclado respecto a los componentes de error normales $\sigma\eta_{nj}\ \forall j$ así como a los elementos aleatorios de β_n. La desviación estándar de los errores condicionados, σ, es estimada, a diferencia del ejemplo anterior.

3. La notación se puede generalizar para permitir correlación entre ε_{nj} y μ_{nk} para $\mathrm{k} \neq \mathrm{j}$. Definamos $\varepsilon_{nj} = \varepsilon_{nj}^1 + \varepsilon_{nj}^2$ como en el ejemplo anterior, excepto que ahora supondremos que los vectores agrupados ε_n^1 y μ_n son conjuntamente normales. En este caso, ε_n^1 condicionada a μ_n es normal con media $\mathrm{M}\mu_n$y varianza Ω, donde M y Ω son matrices de parámetros. Agrupando utilidades y funciones explicativas, tenemos

   $U_n = V(y_n, x_n, \beta_n) + \mathrm{M}\mu_n + L\eta_n + \varepsilon_n^2,$

   donde L es el factor Choleski triangular inferior de Ω y η_n es un vector variables normales estándar iid. Dado que los elementos de $\tilde{\varepsilon}_{nj}^2$ son iid valor extremo, el modelo es logit mixto, mezclado en relación a los componentes de error η_n y a los elementos aleatorios de β_n. Los residuos de todos los productos forman parte de la utilidad de cada producto.

13.4.1 Relación con el comportamiento de los precios

Como se estableció con anterioridad, la principal limitación del enfoque de función de control es la necesidad de especificar dicha función de control así como la distribución condicionada del nuevo término no observado $\tilde{\varepsilon}$. En algunas situaciones, la verdadera distribución condicionada es tan compleja que no se puede obtener, de manera que cualquier intento de especificación que use distribuciones estándar, como la normal, es necesariamente incorrecta. Estas cuestiones se explican más fácilmente usando el ejemplo de la política de precios de las empresas, donde el precio es la variable endógena, pero pueden surgir bajo cualquier tipo de variable endógena, dependiendo de la forma en que se determine la misma.

Supongamos que la elección es entre productos y que la variable endógena y_{nj} es el precio p_{nj}, por lo que podemos hablar de distintos comportamientos de los precios y de si estos comportamientos pueden tener cabida dentro del enfoque basado en la función de control. Consideremos en primer lugar una situación en la que la función de control se puede aplicar fácilmente: precios MC. La utilidad que el consumidor n obtiene del producto j es

$$U_{nj} = V(p_{nj}, x_{nj}, \beta_n) + \varepsilon_{nj}^1 + \varepsilon_{nj}^2,$$

donde ε_{nj}^1 está correlacionada con el precio y ε_{nj}^2 se distribuye valor extremo iid. Por ejemplo, ε_{nj}^1 podría representar atributos no observados del producto. Los precios varían entre personas porque diferentes personas están en diferentes mercados o los precios se establecen por separado para personas o grupos de personas (por ejemplo, para tener en cuenta los costos de transporte de cada cliente). Supongamos, además, que las empresas fijan los precios iguales al costo marginal, que se especifica como $MC_{nj} = W(z_{nj}, \gamma) + \mu_{nj}$, donde z_{nj} son variables exógenas que afectan el costo marginal (incluyendo los atributos observados del producto) y μ_{nj} captura el efecto de palancas de cambio de los costos no observadas (incluyendo los atributos no observados del producto). La ecuación de precios es entonces

$$p_{nj} = W(z_{nj}, \gamma) + \mu_{nj}.$$

Supongamos que ε_{nj}^1 y μ_{nj} son conjuntamente normales, con la misma matriz de covarianza para todo j. Podría surgir correlación, por ejemplo, si los atributos no observados afectan a la utilidad, así como a los costos, entrando así tanto en ε_{nj}^1 como μ_{nj}. Al igual que en el segundo ejemplo dado anteriormente, la utilidad se convierte en

$$U_{nj} = V(p_{nj}, x_{nj}, \beta_n) + \lambda\mu_{nj} + \sigma\eta_{nj} + \varepsilon_{nj}^2,$$

donde η_{nj} es una normal estándar iid. El modelo se estima en dos pasos: en primer lugar, se estima la ecuación de precios y se retienen sus residuos, $\hat{\mu}_{nj}$. En segundo lugar, se estima el modelo de elección usando estos residuos como variables explicativas. El modelo es un logit mixto, mezclado respecto a los nuevos componentes de error η_{nj}. La misma especificación es aplicable si las empresas fijan precios con un margen constante sobre el costo marginal. Con esta política de precios, la ecuación de precios es lo suficientemente simple como para que suposiciones razonables sobre los términos observados en el modelo den una distribución condicionada para los términos no observados de la utilidad que pueda ser obtenida y que sea conveniente.

Consideremos ahora los precios en situación de monopolio o en equilibrio de Nash, donde el precio depende de la elasticidad de la demanda así como del costo marginal. Como se mostró anteriormente en relación con el enfoque BLP, la ecuación de precios para un monopolista mono-producto o un oligopolista de Nash es

$$p_{nj} + (p_{nj}/e_{nj}) = MC$$

$$p_{nj} = -(p_{nj}/e_{nj}) + W(z_{nj}, \gamma) + \mu_{nj}.$$

Supongamos ahora que ε_{nj}^1 y μ_{nj} presentan una distribución normal conjunta. La distribución de ε_{nj}^1 condicionada a μ_{nj} sigue siendo normal. Sin embargo, μ_{nj} ya no es el único error en la ecuación de precios y por lo tanto no podemos obtener una estimación de μ_{nj} sobre la que condicionar. A diferencia de los precios MC, el componente no observado de la demanda, ε_{nj}^1, entra en la ecuación de precios a través de la elasticidad. La ecuación de precios incluye dos términos no observados: μ_{nj} y una transformación altamente no lineal de ε_{nj}^1 entrando a través de e_{nj}.

Si reescribimos la ecuación de precios de forma que pueda ser estimada, con una parte observada y un error separable, tendríamos

$$p_{nj} = Z_j(z_{nj}, \gamma) + \mu_{nj}^*,$$

donde z_n es un vector con todas las variables exógenas observadas que afectan el costo marginal y a la elasticidad, $Z_j(\cdot)$ es una función paramétrica de estas variables y μ_{nj}^* son las desviaciones no observadas del precio en torno a esta función. Podemos estimar esta ecuación y conservar su residuo, que es una estimación de μ_{nj}^*. Sin embargo, μ_{nj}^* no es μ_{nj}; por el contrario, μ_{nj}^*incorpora tanto μ_{nj} como los componentes no observados del margen basado en la elasticidad p_{nj}/e_{nj}. La distribución de ε_{nj}^1 condicionada a este μ_{nj}^* no se ha obtenido, y podría no ser obtenible. Dada la forma en que ε_{nj}^1 entra en la ecuación de precios a través de la elasticidad, su distribución condicionada desde luego no es normal aunque su distribución no condicionada sea normal. De hecho, su distribución condicionada ni siquiera es independiente de las variables exógenas.

Villas- Boas (2007) ha propuesto una dirección alternativa para gestionar esta situación. Señala que la dificultad que nos hemos encontrado anteriormente en la especificación de una función de control apropiada surge del supuesto de que los costos marginales son separables en términos no observados, más que del supuesto de que los precios están relacionados con las elasticidades. Supongamos por el contrario que el costo marginal es una función general de los términos observados y no observados: $\mathrm{MC} = W^*(z_{nj}, \gamma, \mu_{nj})$. En muchos sentidos, ese supuesto de términos no observados y no separables es más realista, dado que es de esperar que las palancas de cambio de costo no observadas, en general, interactúen con palancas de costo observadas. En virtud de esta función general de costo, Villas-Boas muestra que para cualquier especificación de la función de control y de la distribución de ε_{nj}^1 condicionada a esta función de control, existe una función de costo marginal $W^*(\cdot)$ y una distribución de los términos no observados μ_{nj}y ε_{nj}^1 que es consistente con ellas. Este resultado implica que el investigador puede aplicar el enfoque basado en la función de control, incluso cuando los precios dependen de la elasticidad, a sabiendas de que existe cierta función de costo marginal y unas distribuciones de los términos no observables que hacen que el enfoque sea consistente. Por supuesto, esta prueba de la existencia no proporciona ninguna orientación sobre qué función de control y qué distribución condicionada son más razonables, lo que debe seguir siendo una cuestión importante para el investigador.

13.5 Enfoque de máxima verosimilitud

El enfoque de máxima verosimilitud es similar al de la función de control, exceptuando el hecho de que los parámetros del modelo se estiman de forma simultánea en lugar de secuencialmente. Al igual que con el enfoque de función de control, la utilidad viene dada por la ecuación (13.13) y la variable explicativa endógena se especifica en (13.12). Para permitir que la notación sea más compacta, agrupamos cada uno de los términos entre alternativas, de tal manera que las dos ecuaciones resultan

$$U_n = V(y_n, x_n, \beta_n) + \varepsilon_n \tag{13.15}$$

$$y_n = W(z_n, \gamma) + \mu_n. \tag{13.16}$$

En lugar de especificar la distribución condicionada de ε_n dado μ_n, el investigador especifica su distribución conjunta, la cual denotamos como $\mathrm{g}(\varepsilon_n, \mu_n)$. Usando la ecuación (13.16), μ_n se puede expresar como una función de los datos y de los parámetros: $\mu_n = y_n - W(z_n, \gamma)$. De esta forma, la distribución conjunta de ε_ny y_n es $\mathrm{g}(\varepsilon_n, y_n - W(z_n, \gamma))$. Denominemos a la alternativa elegida como i. La probabilidad de los datos observados para la persona n es la probabilidad de que la variable explicativa endógena tome el valor y_n y de que la alternativa escogida sea la i. Condicionada respecto a β_n, esta probabilidad es

$$P_n(\beta_n) = \int I(U_{ni} > U_{nj}\ \forall j \neq i) g(\varepsilon_n, y_n - W(z_n, \gamma)) d\varepsilon_n.$$

Si β_n es aleatorio, entonces $P_n(\beta_n)$ se mezcla respecto a su distribución. La probabilidad resultante P_n se inserta en la función log-verosimilitud: $LL = \sum_n \ln(P_n)$. Esta LL se maximiza respecto a los parámetros del modelo. En lugar de estimar (13,16) primero y usar luego los residuos en la probabilidad de elección, los parámetros de (13.16) y el modelo de elección se estiman de forma simultánea.

El tercer ejemplo del enfoque de función de control (que es el ejemplo más general) se puede adaptar a este procedimiento de máxima verosimilitud. La utilidad agrupada es

$$U_n = V(y_n, x_n, \beta_n) + \varepsilon_n^1 + \varepsilon_n^2$$

Siendo la variable endógena agrupada:

$$y_n = W(z_n, \gamma) + \mu_n,$$

donde $W(\cdot)$ tiene ahora valores vectoriales. Cada elemento de ε_n^2 se asume distribuido tipo valor extremo iid. Asumamos que ε_n^1 y μ_n presentan una distribución conjuntamente normal con media cero y covarianza Ω. Su densidad se denota como $\phi(\varepsilon_n^1, \mu_n|\Omega)$. La probabilidad de que forma parte de la función log-verosimilitud para la persona n que eligió la alternativa i es

$$P_n = \int\int \frac{e^{V(y_{ni}, x_{ni}, \beta_n) + \varepsilon_{ni}^1}}{\sum e^{V(y_{nj}, x_{nj}, \beta_n) + \varepsilon_{nj}^1}} \phi(\varepsilon_n, (y_n - W(z_n, \gamma))|\Omega) f(\beta_n|\theta) d\varepsilon_n d\beta_n.$$

Esta probabilidad se inserta en la función log-verosimilitud y se maximiza respecto a γ (los parámetros que relacionan la variable explicativa endógena con los instrumentos), θ (que describe la distribución de las preferencias que afectan a la utilidad) y Ω (la covarianza de los términos no observados correlacionados ε_n^1 y μ_n). Por supuesto, en cualquier caso real concreto, es posible aplicar restricciones a Ω para reducir el número de parámetros. Por ejemplo, Park y Gupta (en un trabajo que será publicado próximamente) asumen que ε_{nj}^1 y μ_{nk} no están correlacionados para $k \neq j$.

Los enfoques basados en función de control y en máxima verosimilitud proporcionan una solución de compromiso que es habitual en econometría: generalidad frente a eficiencia. El enfoque de máxima verosimilitud requiere una especificación de la distribución conjunta de ε_n y μ_n, mientras que la función de control requiere una especificación de la distribución condicionada de ε_n dado un μ_n. Cualquier distribución conjunta implica una distribución condicionada particular, pero cualquier distribución condicionada dada no implica necesariamente una distribución conjunta particular. Puede haber numerosas distribuciones conjuntas que generen una misma distribución condicionada. Por tanto, el enfoque de función de control es más general que el enfoque de máxima verosimilitud: es aplicable para cualquier distribución conjunta que sea coherente con la distribución condicionada especificada. Sin embargo, si la distribución conjunta puede especificarse correctamente, el enfoque de máxima verosimilitud es más eficiente que el de función de control, simplemente por el hecho de que es la solución de máxima verosimilitud para todos los parámetros.

13.6 Caso de estudio: elección de consumidores entre vehículos nuevos

Un caso ilustrativo del enfoque BLP lo proporcionan Train y Winston (2007). Su estudio examinó la cuestión de por qué los tres grandes fabricantes (estadounidenses) de automóviles han estado perdiendo cuota de mercado. Como parte del estudio, estimaron un modelo de elección de los compradores de vehículos nuevos. Dado que muchos de los atributos de los vehículos nuevos no son observados por el investigador, y sin embargo afectan a los precios, es de esperar que el precio sea endógeno.

La estimación se realizó sobre una muestra de consumidores que compraron o alquilaron vehículos nuevos en el año 2000, junto con los datos relativos a las cuotas de mercado para cada marca y modelo de ese año. El análisis no incluyó un bien externo, y como tal, muestra la demanda condicionada a la compra de un vehículo nuevo. Para cada comprador muestreado, la información de la encuesta incluyó la marca y el modelo del vehículo que la persona que compró, más una lista de los vehículos que la persona declaró que tuvo en consideración en el momento de comprar. El vehículo elegido y los vehículos tenidos en cuenta fueron tratados como un ranking, con el vehículo elegido en primera posición y los vehículos considerados clasificados en el orden en que la persona los mencionó. El modelo de elección se especificó como un "logit explotado" (*exploded logit*) para la probabilidad del ranking de la persona, mezclado respecto a una distribución de coeficientes aleatorios. Vea la Sección 7.3 para una extensa explicación relativa a esta especificación para datos de ordenación o ranking. El modelo de elección incluyó constantes para cada marca y modelo de vehículo, así como variables explicativas y coeficientes aleatorios para capturar las variaciones en las preferencias entre consumidores. El análisis distinguía 200 marcas y modelos, y utilizaba la contracción para calcular las 199 constantes (con un normalizada a cero) dentro de la estimación de máxima verosimilitud de los otros parámetros. Las constantes estimadas fueron posteriormente objeto de una regresión respecto a los atributos observados de los vehículos, incluyendo el precio. Puesto que el precio fue considerado endógeno, esta regresión se estimó mediante variables instrumentales en lugar de mínimos cuadrados ordinarios.

Tabla 13.1. Modelo logit explotado para la elección de nuevo vehículo

	Parámetro	Error estándar
Precio dividido por el ingreso de los hogares		
Coeficiente medio	−1.6025	0.4260
Desviación estándar del coeficiente	0.8602	0.4143
Índice de Reparaciones Reportado por el Consumidor (Consumer Report’s repair index), para mujeres de 30 o más años	0.3949	0.0588
Vehículos lujosos o deportivos, para personas que alquilan	0.6778	0.4803
Furgoneta, para hogares con adolescentes	3.2337	0.5018
Todoterrenos o familiares, para hogares con adolescentes	2.0420	0.4765
(1+número de concesionarios en un radio de 50 millas del hogar del comprador)	1.4307	0.2714
Potencia: desviación estándar del coeficiente	0.0045	0.0072
Consumo de combustible (l/mpg): desviación estándar del coeficiente	−102.15	20.181
Camioneta ligera, furgoneta o pickup: desviación estándar del coeficiente	6.8505	2.5572
Número de compras previas consecutivas de un vehículo GM	0.3724	0.1471
Número de compras previas consecutivas de un vehículo GM, si el hogar es rural	0.3304	0.2221
Número de compras previas consecutivas de un vehículo Ford	1.1822	0.1498
Número de compras previas consecutivas de un vehículo Chrysler	0.9652	0.2010
Número de compras previas consecutivas de un vehículo japonés	0.7560	0.2255
Número de compras previas consecutivas de un vehículo europeo	1.7252	0.4657
Lealtad al fabricante, componente de error: desviación estándar	0.3453	0.1712
SLL en convergencia	−1994.93	

La tabla 13.1 muestra las estimaciones de los parámetros que se refieren a la variación de las preferencias entre consumidores. Estos son los parámetros que se estimaron por máxima verosimilitud respecto a la probabilidad del ranking de marcas y modelos de cada comprador, usando la contracción de las constantes. Los coeficientes estimados tienen las siguientes implicaciones:

- El precio dividido por los ingresos entra como variable explicativa, con el fin de captar la idea de que los hogares con ingresos más altos dan menos importancia al precio que los hogares con menores ingresos. Se asigna a la variable un coeficiente aleatorio distribuido normalmente, cuya media y desviaciones estándar se calcularon. La media estimada es negativa, como se esperaba, y la desviación estándar estimada es considerablemente grande y significativa (con un estadístico-*t* en torno a 2), lo que indica una variación considerable en respuesta al precio.
- El Índice de Reparaciones Reportado por el Consumidor (*Consumer Report's repair index*) entra para las mujeres de 30 o más años, y no para los hombres y las mujeres más jóvenes. Esta distinción fue descubierta a través de pruebas de variables demográficas alternativas interactuado con el índice de reparaciones.
- Las personas que alquilan su vehículo tienen mayor preferencia por el lujo y por los vehículos deportivos de la gente que compra su vehículo.
- Se observa que los hogares con adolescentes tienen una mayor preferencia por furgonetas, todoterrenos y coches familiares que otros hogares.
- Se encontró que las ubicaciones de los concesionarios afectan a las elecciones de los hogares. Uno de los propósitos del análisis de Train y Winston fue investigar el impacto de las ubicaciones de los concesionarios en la elección del vehículo, para ver si los cambios en las ubicaciones podían explicar parte de la pérdida de cuota de mercado de los 3 grandes fabricantes. El modelo de la demanda indica que, como era de esperar, la probabilidad de comprar una marca y un modelo determinado se eleva cuando hay más concesionarios que venden esa marca y modelo dentro de un radio de 50 millas del hogar.
- La potencia, el consumo de combustible y una variable indicadora referida a si el vehículo es una camioneta, entran con coeficientes aleatorios. Estas variables no varían entre consumidores, sólo entre marcas y modelos. Como resultado, el producto del coeficiente medio por la variable es absorbido por las constantes de marca/modelo, de forma que sólo la desviación estándar es calculada en el modelo de elección. Para ser precisos, cada una de estas variables entra a formar parte de la utilidad como $\beta_n x_j$, con β_n aleatorio. Descomponemos β_n en su media $\bar{\beta}$ y sus desviaciones $\tilde{\beta}_n$. El impacto medio $\bar{\beta}x_j$ no varía entre consumidores y se convierte en parte de la constante del vehículo j. Las desviaciones $\tilde{\beta}_n x_j$ entran en el modelo de elección por separado de las constantes, y la desviación estándar de $\tilde{\beta}_n$ es estimada. Las estimaciones indican una considerable variación en las preferencias de los consumidores respecto a la potencia, la eficiencia de combustible y al hecho de que los vehículos sean camionetas.
- El último conjunto de variables captura la lealtad de los consumidores hacia los fabricantes. Cada variable se especifica como el número de compras pasadas consecutivas de un vehículo de un fabricante determinado (o grupo de fabricantes). Los coeficientes estimados indican que la lealtad a los fabricantes europeos es mayor, seguida por la lealtad a Ford. Curiosamente, los hogares rurales resultan ser más leales a GM que los hogares urbanos. Estas variables de fidelidad son un tipo de variable dependiente diferida, que introducen aspectos econométricos debido a la posibilidad de que existan factores no observados correlacionados en serie. Winston y Train tratan estos temas y tienen en cuenta la correlación en serie, al menos en parte, en su procedimiento de estimación.

El modelo de elección de la tabla 13.1 incluye constantes para cada marca y modelo de vehículo. Las constantes estimadas fueron objeto de una regresión respecto a los atributos del vehículo para estimar los aspectos de la demanda que son comunes a todos los consumidores. La tabla 13.2 muestra los coeficientes estimados de esta regresión, utilizando variables instrumentales para tener en cuenta la endogeneidad del precio.

Tabla 13.2. Regresión de constantes respecto a atributos de los vehículos

	Parámetro	Error estándar
Precio (precio de venta recomendado por el fabricante, en miles de dólares)	−0.0733	0.0192
Potencia dividida por peso (en toneladas)	0.0328	0.0117
Cambio automático	0.6523	0.2807
Ancho (en pulgadas)	0.0516	0.0127
Largo menos ancho (en pulgadas)	0.0278	0.0069
Consumo de combustible (en galones por milla)	−31.641	23.288
Vehículo de lujo o deportivo	−0.0686	0.2711
Todoterreno o familiar	0.7535	0.4253
Furgoneta	−1.1230	0.3748
Camioneta pickup	0.0747	0.4745
Chrysler	0.0228	0.2794
Ford	0.1941	0.2808
General Motors	0.3169	0.2292
Europeo	2.4643	0.3424
Coreano	0.7340	0.3910
Constante	−7.0318	1.4884
R-cuadrado	0.394	

Como era de esperar, el precio entra en el modelo con un coeficiente negativo, lo que indica que a los consumidores no les gusta el precio más alto, manteniendo todo lo demás igual (es decir, suponiendo que no hay cambios en los atributos de un vehículo para compensar el precio más alto). Curiosamente, cuando la regresión se estimó por mínimos cuadrados ordinarios, haciendo caso omiso de la endogeneidad, el coeficiente de precio estimado fue considerablemente menor: -0.0434 en comparación con la estimación de -0.0733 obtenida utilizando variables instrumentales. La dirección de esta diferencia es la esperada, ya que una correlación positiva entre el precio y los atributos no observados crea un sesgo a la baja en la magnitud del coeficiente de precio. El tamaño de esta diferencia indica la importancia de tener en cuenta la endogeneidad.

Los otros coeficientes estimados tienen los signos esperados. La estimación indica que los consumidores valoran la potencia adicional, como se evidencia por el coeficiente positivo de la relación potencia-peso. Los consumidores también prefieren tener transmisión automática de serie (manteniendo el precio del vehículo constante). El tamaño del vehículo se mide tanto por su distancia entre ejes como por su longitud respecto a la distancia entre ejes. Ambas medidas entran positivamente en el modelo, y la distancia entre ejes obtiene un coeficiente mayor que la longitud respecto a la distancia entre ejes. Este resultado es de esperar, ya que la distancia entre ejes es generalmente un mejor indicador del espacio interior del vehículo que la longitud. El coeficiente negativo de consumo de combustible implica que los consumidores prefieren una mayor eficiencia de combustible, lo que reduce el consumo por milla recorrida.

Tenga en cuenta que el precio entra en ambas partes del modelo: en la regresión de la tabla 13.2, que captura los impactos que son constantes entre consumidores, y en el modelo de elección de la tabla 13.1, que captura impactos que difieren entre consumidores. Tomando ambas partes conjuntamente, el precio forma parte de la utilidad con un coeficiente: $-0{,}0773 - 1.602/Ingresos + 0{,}860 * \eta/Ingresos$, donde η es un término aleatorio normal estándar. El coeficiente de precio tiene un componente constante, una parte que varía con el ingreso del hogar y una parte que varía aleatoriamente entre hogares con el mismo nivel de ingresos.

La endogeneidad de los precios ha sido manejada adecuadamente en este caso de estudio mediante la inclusión de constantes en el modelo de elección para absorber los atributos no observados y usando posteriormente variables instrumentales al hacer la regresión de las constantes respecto al precio y a otros atributos observados, es decir, a través del enfoque BLP. Un caso de estudio sobre el enfoque de función de control lo proporcionan Petrin y Train (2009), y del enfoque de máxima verosimilitud lo proporcionan Park y Gupta (de próxima publicación).