Investigaciones Económicas (Segunda época). Vol. XI, n.º 1 (1987) págs. 41-94

PROBLEMAS PRACTICOS EN EL TRATAMIENTO ECONOMETRICO DE DATOS «CROSS-SECTION»

Ignacio MAULEON*

Banco de España

Este trabajo pretende ser un resumen de métodos disponibles para tratar datos en panel, de utilidad para el investigador aplicado. El énfasis se ha puesto en la aplicabilidad de los modelos y se presentan una amplia gama de técnicas (modelos de efectos fijos y aleatorios, modelos con variables dependientes cualitativas y truncadas, aplicaciones de métodos no paramétricos y análisis de varianza).

1. Introducción

Este trabajo está concebido como una guía de apoyo para el investigador aplicado cuyo material son observaciones de corte transversal o en panel, y que no es un especialista en este tipo de econometría. La motivación del trabajo reside en la ausencia de literatura similar sobre este tema, fácilmente accesible, y al interés que recientemente se ha despertado por ella, debido a la abundante disponibilidad de este tipo de datos.

La primera parte del trabajo, presenta una revisión de los tipos de modelos que se pueden aplicar a datos en panel. La presentación se divide en tres partes: *a*) modelos de efectos fijos y aleatorios, *b*) modelos con variables dependientes cualitativas (probit, logit), *c*) modelos con variables dependientes truncadas (tóbit). El énfasis se pone siempre en la aplicabilidad de los modelos, y así la mayor parte o la totalidad de las demostraciones se omiten, a la vez que se incluyen las referencias especializadas sobre el particular. Lo que sí se ha intentado, es explicar el sentido e importancia práctica de los diferentes resultados teóricos (por ejemplo, propiedades de convergencia de algoritmos, etc...). Como la aplicabilidad de los modelos muchas veces no es inmediata, se presentan también métodos sencillos de aplicación (por ejemplo, transformaciones para los modelos de efectos aleatorios).

La segunda parte, se centra en analizar problemas comúnmente encontrados en la formulación y especificación de los modelos. Especial hincapié se pone en la contrastación de modelos, como método para su validación. Una batería am-

* Este trabajo ha sido realizado a petición de R. Sanz, cuya sugerencia agradezco. Agradezco los comentarios de Daniel Peña, Miguel Delgado, y particularmente los de Inmaculada Gallastegui.

plia de contrastes se presenta y explica, a la vez que se exponen expresiones sencillas para calcularlos. Algunos problemas comunes en el análisis de este tipo de datos, también se discuten en esta sección (por ejemplo, ajustes bajos, selección de indicadores, medición de las variables por cocientes, etc...).

En general, ningún investigador aplicado está dispuesto a emplear tiempo, en escribir programas específicos para la aplicación de ciertos métodos. Por eso, para que los métodos se apliquen suele ser necesario disponer de programas estandarizados. La última sección da una breve descripción de algunos de ellos, que disponen de varias de las opciones presentadas en la sección I. También se incluyen las respectivas direcciones comerciales, para facilitar el acceso a dichos programas.

2. Modelos

Las técnicas utilizadas comúnmente en análisis de datos de corte transversal o de panel, son las tradicionales del análisis multivariante: análisis de varianza, correlación canónica, análisis factorial y de componentes principales, y análisis discriminante, principalmente. Sin embargo, en la literatura econométrica se han desarrollado recientemente una serie de métodos, que en gran parte basados en los tradicionales de análisis multivariante, proporcionan un tratamiento más sofisticado de los datos. Más concretamente, estos métodos son: *a*) análisis de datos en panel; *b*) tratamiento de variables cualitativas, y truncadas. El primero es en cierto modo una extensión del análisis de varianza al caso en que haya otras variables explicativas medidas cardinalmente, y que el número de observaciones por casilla sea una. El segundo es un substituto al análisis discriminante, que permite evaluar la proximidad de un individuo a un grupo en términos de probabilidad. Además, el tratamiento de los datos se encuadra dentro del marco del análisis máximo verosímil, lo que permite la aplicación inmediata de todo el aparato estadístico asociado a este enfoque. El tratamiento de las truncaciones, es relativamente nuevo, y requiere, como era de esperar, una técnica más compleja. Sin embargo, casos sencillos pueden ser aplicados con resultados positivos.

Esta sección va a estar dedicada a una presentación de estos dos métodos o técnicas, desde el punto de vista del usuario no especialista. El énfasis se pondrá en cómo aplicar las técnicas, omitiendo en lo posible las demostraciones.

2.1. *Métodos para datos en panel*

La situación típica en este caso, es un grupo de variables observadas para varios individuos a lo largo del tiempo. Se suele suponer que el modelo básico de comportamiento es común a todos los individuos a lo largo del tiempo. Esto puede escribirse como sigue

$$Y_{it} = X_{it}'B + \varepsilon_{it} \qquad \begin{array}{l} i = 1, \ldots, N \\ t = 1, \ldots, T \end{array} \qquad [1]$$

La ventaja de hacer este supuesto, es obviamente, que al disponer de un gran número de observaciones con parámetros comunes B, éstos pueden estimarse más eficientemente. Para aliviar algo la rigidez de este supuesto se suele permitir cierta variabilidad entre las observaciones. Hay dos métodos básicos de introducir esta variabilidad. Uno de ellos supone que la constante de cada ecuación varía para cada individuo y para cada unidad de tiempo de modo que,

$$\varepsilon_{it} = z_{it}'\gamma + e_{it} \tag{2}$$

siendo 'z_{it}' el vector de variables artificiales (dummies) apropiadas. El término 'e_{it}' es un error de media cero, homocedástico e incorrelacionado serialmente y para diferentes individuos. (Modelo de efectos fijos (EF)). El segundo método, supone que las diferencias entre las observaciones individuales son realizaciones de variables aleatorias (modelo de efectos aleatorios (EA)). En este caso tendríamos lo siguiente,

$$\varepsilon_{it} = u_i + v_t + w_{it} \tag{3}$$

donde (u, v, w) son errores aleatorios que cumplen las siguientes características

$$Eu_i = Ev_t = Ew_{it} = 0$$

$$E(u_i u_j) = \begin{cases} \sigma_u^2 & i = j \\ 0 & i \neq j \end{cases}$$

$$E(v_t v_s) = \begin{cases} \sigma_v^2 & t = s \\ 0 & t \neq s \end{cases}$$

$$E(w_{it} w_{js}) = \begin{cases} \sigma_w^2 & i = j, \quad t = s \\ 0 & \text{todos los demás casos} \end{cases}$$

$$E(u_i v_t) = E(u_i w_{jt}) = E(v_t w_{is}) = 0$$

$$\text{(para todo } i, j, t, s) \tag{4}$$

Estos son los dos modelos básicos, de los que pueden obtenerse modelos más sencillos, suponiendo que alguno de los efectos se anula, o mixtos, combinando efectos fijos y aleatorios. En primer lugar, es conveniente presentar una tipología completa de posibles casos, y cómo se pueden aplicar. La discusión sobre la conveniencia de adoptar un enfoque u otro se relega al final de la sección.

En principio, no hay problema teórico especial para aplicar estos modelos. En el caso de efectos fijos basta con definir las matrices de variables artificiales, y en el de efectos aleatorios la matriz de covarianzas de 'ε' para aplicar mínimos cuadrados generalizado (MCG). Sin embargo, pocos programas econométricos en el mercado incorporan como una opción más estos métodos. Además, supongamos que $N = 500$, $T = 2$, un caso muy plausible con datos en panel.

Entonces, la matriz de covarianzas de 'ε' tendría un millón de elementos, y el orden de operaciones requeridas para invertirla sería $(\text{dimensión})^3 = 10^9$, o sea mil millones. Si pensamos en un estimador iterado, el coste empieza a ser considerable, incluso para ordenadores eficientes. Por otra parte, una matriz de un millón de elementos, es problemática para su tratamiento en algunos lenguajes. (Por ejemplo, alguna versión de Fortran sólo admite matrices de menos de 100.000 elementos.)

Afortunadamente, la estructura de las matrices relevantes es muy simple, lo que permite aplicar estos métodos estimando por mínimos cuadrados ordinarios (MCO) el modelo con los datos obtenidos después de una transformación sencilla.

Para ello en primer lugar introducimos la siguiente notación:

$$
\begin{aligned}
Y_{i.} &= \Big(\sum_t Y_{it}\Big)/T \\
Y_{.t} &= \Big(\sum_i Y_{it}\Big)/N \\
Y_{..} &= \Big(\sum_{i,\,t} Y_{it}\Big)/(NT) \\
&= \Big(\sum_i Y_{i.}\Big)/N \\
&= \Big(\sum_t Y_{.t}\Big)/T
\end{aligned} \qquad [5]
$$

Las transformaciones son las siguientes.

a) EFECTOS FIJOS

Si hay únicamente efectos individuales la transformación es,

$$Y_{it}^{+} = Y_{it} - Y_{i.} \qquad [6]$$

y similarmente para X_{it}. A los datos transformados (Y^{+}, X^{+}) se les aplica un procedimiento de regresión mínimo cuadrática estándar.

Si únicamente hay efectos temporales la transformación es,

$$Y_{it}^{+} = Y_{it} - Y_{.t} \qquad [7]$$

Si hay efectos individuales y temporales la transformación es la siguiente:

$$Y_{it}^{+} = Y_{it} - Y_{i.} - Y_{.t} + Y_{..} \qquad [8]$$

Para interpretar mejor esta transformación podemos escribirla del siguiente modo,

$$\begin{aligned} Y_{it} - Y_{i.} - Y_{.t} + Y_{..} &= (Y_{it} - Y_{..}) - (Y_{i.} - Y_{..}) - (Y_{.t} - Y_{..}) \\ &= \bar{Y}_{it} - \bar{Y}_{i.} - \bar{Y}_{.t} \end{aligned} \qquad [9]$$

es decir, que los datos transformados se obtienen restando a la observación individual, la media de grupo individual y la media temporal, estando a su vez todas las observaciones y medias de grupo medidas como desviaciones de sus medias respectivas.

Si se desea estimar los efectos individuales y temporales, es decir el vector 'γ' de [2], se puede llevar a cabo sencillamente, utilizando los resultados de la regresión particionada (véase Apéndice 1).

b) Efectos aleatorios

— *Transformaciones*

Si hay únicamente efectos individuales la transformación es la siguiente,

$$\begin{aligned} Y_{it}^{+} &= Y_{it} - \phi Y_{i.} \\ \phi &= 1 - \sigma_w/(\alpha_w^2 + T\sigma_u^2)^{1/2} \end{aligned} \qquad [10]$$

y similarmente para los regresores. A continuación puede aplicarse MCO a los datos así transformados y todos los estadísticos usuales son válidos.

Si los efectos son sólo temporales, la transformación análoga es,

$$\begin{aligned} Y_{it}^{+} &= Y_{it} - \phi Y_{.t} \\ \phi &= 1 - \sigma_w/(\sigma_w^2 + N\sigma_v^2)^{1/2} \end{aligned} \qquad [11]$$

Si los efectos son temporales e individuales, la transformación es la siguiente,

$$\begin{aligned} Y_{it}^{+} &= Y_{it} - aY_{i.} - bY_{.t} + cY_{..} \\ a &= 1 - \sigma_w/(\sigma_w^2 + T\sigma_u^2)^{1/2} \\ b &= 1 - \sigma_w/(\sigma_w^2 + N\sigma_v^2)^{1/2} \\ c &= \left(\frac{NT\gamma_3}{2}\right) - (b + a) - \frac{1}{2}\left(1 + 2((T\gamma_1 + N\gamma_2) + (b + a) - 3)^{1/2}\right) \\ \gamma_1 &= \sigma_u^2/(\sigma_w^2 + T\sigma_u^2) \\ \gamma_2 &= \sigma_v^2/(\sigma_w^2 + N\sigma_v^2) \\ \gamma_3 &= (\gamma_1\gamma_2)[(2\sigma_w^2 + N\sigma_v^2 + T\sigma_u^2)/(\sigma_w^2 + N\sigma_v^2 + T\sigma_u^2)] \end{aligned} \qquad [12]$$

El método de obtención de esta transformación se indica en el Apéndice 1. (Un caso más sencillo está derivado de Johnston.)

Una situación interesante se produce cuando unos efectos son fijos y otros aleatorios. En general, se suele tomar como efectos aleatorios los individuales para economizar en el número de parámetros estimado. En este caso, puede demostrarse que la transformación adecuada es,

$$Y_{it}^{+} = Y_{it} - Y_{.t} - (KT)Y_{..} \qquad [13]$$

donde 'k' está definido por la siguiente recursión,

$$\phi = 1 - \sigma_w/(\sigma_w^2 + T\sigma_u^2)^{1/2}$$

$$g = \frac{1}{T}\frac{\phi(\phi - 2)}{(\phi - 1)^2}$$

$$h = (\phi/T) + g - \phi g$$

$$k = h + g - hgT$$

La derivación de estas expresiones es algo más delicada que en el caso anterior, pero en todo caso no se requiere ningún argumento especialmente nuevo.

— *Estimación de las varianzas*

Cuando los efectos son aleatorios, las varianzas σ_u^2, σ_w^2, σ_v^2 son por lo general desconocidas. El método a aplicar entonces es el de MCG factibles, es decir, estimar consistentemente las varianzas en un primer paso, para aplicar MCG con estas estimaciones en una segunda etapa. (Podemos notar que en el caso de efectos mixtos, lo adecuado sería aplicar la transformación [7] en primer lugar, antes de realizar la primera regresión). El único problema pendiente es entonces el de definir los estimadores de las varianzas a partir de los errores. Para ello, a continuación se presentan las varianzas derivadas del enfoque de «análisis de varianza» (véase, por ejemplo, Scheffé (1959)). Empezando por el caso más general, consideremos las siguientes medias, definidas análogamente a $Y_{i.}$, $Y_{.t}$ (veáse [5]).

$$\varepsilon_{it} = u_i + v_t + w_{it}$$

$$\varepsilon_{i.} = u_i + v_{.} + w_{i.}$$

$$\varepsilon_{.t} = u_{.} + v_t + w_{.t}$$

$$\varepsilon_{..} = u_{.} + v_{.} + w_{..} \qquad [15]$$

Para obtener una estimación de σ_w^2, observamos que

$$\varepsilon_{it}^{+} = \varepsilon_{it} - \varepsilon_{i.} - \varepsilon_{.t} + \varepsilon_{..} = w_{it} - w_{i.} - w_{.t} + w_{..} = w_{it}^{+} \qquad [16]$$

Utilizando ahora la matriz H del Apéndice 1,

$$\begin{aligned}\sum_{it} E(\varepsilon_{it}^{+})^2 &= E(\varepsilon' H \varepsilon)\\ &= E(w' H w)\\ &= \sigma_w^2 trH\\ &= \sigma_w^2 (T - 1)(N - 1)\end{aligned} \qquad [17]$$

de modo que un estimador insesgado de σ_w^2 será,

$$\hat{\sigma}_w^2 = \frac{\varepsilon^{+\prime}\varepsilon^{+}}{(N - 1)(T - 1)} \qquad [18]$$

Substituyendo ε^+ por su estimación, obtenemos en [18] una estimación consistente. El factor de corrección en este caso es mejor substituirlo por $\big((N - 1)(T - 1) - K\big)$ siendo K la dimensión de 'B' (el número de parámetros estimados).

Similarmente obtenemos

$$\begin{aligned}\varepsilon_{i.} - \varepsilon_{..} &= u_i + w_{i.} - (u_{.} + w_{..}) = a_i\\ E\big(a'a/(N - 1)\big) &= \sigma_u^2 + \sigma_w^2/T\end{aligned} \qquad [19]$$

de modo que a partir de σ_w^2 y utilizando esta expresión (sin esperanzas) obtenemos el estimador de σ_u^2. Es importante señalar que no está garantizado que $\hat{\sigma}_u^2 > 0$. En la práctica cuando $\hat{\sigma}_u^2 < 0$, se suele tomar $\hat{\sigma}_u^2 = 0$. Si sólo hay efectos individuales, esto es equivalente a aplicar MCO y el resultado es probable que sea pobre con lo que se concluye que los efectos son fijos. Sin embargo, es típico en los estudios empíricos que el R^2 sea muy bajo, o en otras palabras que $\hat{\sigma}_w^2$ sea alto. Entonces, es probable que se presente este problema, pero la causa más verosímil es que haya variables omitidas, más específicamente retrasos de todas las variables. En otras palabras, la falta de retrasos entre los regresores en modelos estáticos, es la causa probable de que los ajustes sean bajos $\hat{\sigma}_w^2$ alta en consecuencia y $\hat{\sigma}_u^2 < 0$, de modo que el modelo de efectos aleatorios no sea aplicable.

Para estimar σ_v^2, utilizamos la expresión análoga a [19],

$$\begin{aligned}\varepsilon_{.t} - \varepsilon_{..} &= v_t + w_{.t} - (v_{.} + w_{.}) = b_t\\ E(b'b)/(J - 1) &= \sigma_v^2 + \sigma_w^2/N\end{aligned} \qquad [20]$$

Si únicamente uno de los efectos está presente, por ejemplo el individual, las expresiones equivalentes a [15] son ahora las siguientes:

$$\begin{aligned}\varepsilon_{it} &= u_i + w_{it}\\ \varepsilon_{i.} &= u_i + w_{i.}\\ \varepsilon_{..} &= u_{.} + w_{..}\end{aligned} \qquad [21]$$

Notamos ahora que,

$$\begin{gathered} \varepsilon_{it} - \varepsilon_{i.} = w_{it} - w_{i.} = a_{it} \\ E\big(a'a/N(T-1)\big) = \sigma_w^2 \end{gathered} \qquad [22]$$

expresión esta última que sirve de base para formular un estimador consistente para σ_w^2 en este caso, de forma análoga a la expresión [18]. Para obtener un estimador de σ_u^2, procedemos del siguiente modo:

$$\begin{gathered} \varepsilon_{i.} - \varepsilon_{..} = u_i + w_{i.} - (u_. + w_{..}) = b_{i.} \\ E\big(b'b/(N-1)\big) = \sigma_u^2 + \sigma_w^2/T \end{gathered} \qquad [23]$$

y a partir de esta última expresión obtenemos un estimador de σ_u^2, dado el estimador de σ_w^2. Como antes, puede ser más correcto corregir las expresiones por el número de parámetros estimado en B (véase párrafo posterior a [18]).

c) Elección de modelo

Una decisión importante en la investigación empírica es cuál de los dos modelos elegir. La respuesta no es simple, y como casi siempre en econometría aplicada, es una cuestión fundamentalmente práctica, que debe ser dejada al juicio del investigador. Sin embargo, hay algunos criterios básicos por los que guiarse.

En primer lugar, podemos adoptar el enfoque clásico del análisis de varianza. Según este enfoque, el modelo de efectos aleatorios es adecuado cuando las observaciones son una muestra aleatoria de una posible población de individuos. Por ejemplo, supongamos que tratamos de medir la diferente sensibilidad hacia la exportación de diferentes sectores industriales. Si analizamos solamente una muestra de los sectores, sería lógico aplicar el modelo de efectos aleatorios, ya que por medio de la muestra tratamos de medir el efecto general, mientras que si analizamos todos, el modelo apropiado sería el de efectos fijos. Sin embargo, podemos considerar siempre que el análisis es condicional a la muestra realizada en el caso de que la muestra seleccionada sea más restringida que la población. Entonces el enfoque de efectos fijos sería apropiado. Si se desea extender las conclusiones a toda la población, entonces el enfoque correcto es el de efectos aleatorios (Mundlak).

Otro problema se plantea en relación a la pertenencia o no a los regresores de retrasos de la variable dependiente. Si los regresores no son estocásticos, puede demostrarse que el modelo de efectos fijos da siempre estimadores insesgados de 'B', ya sea el modelo subyacente de efectos fijos o aleatorios (Wallace y Hussain (1969)). Además, también puede demostrarse que cuando $(T, N) \rightarrow \infty$, los estimadores de variables artificiales y de MCG coinciden. Por otra parte, el estimador de efectos aleatorios (MCG) da estimadores sesgados si el verdadero es de efectos fijos. Si uno o más retrasos de la variable dependiente pertenece a los regresores el análisis es algo más complejo. Puede demostrarse que el estimador de variables artificiales da sesgos (Balestra, Nerlove (1968)). Este sesgo es de orden $(1/T)$ (véase Apéndice 2) de modo que para T pequeño, que es lo usual con datos en panel, puede ser considerable. Una solución si los efectos son aleatorios, se puede obtener con un estimador de dos etapas. En la primera, se estima por variables instrumentales el modelo

$$\begin{aligned} Y_{it} &= \alpha Y_{it-1} + X'_t B + \varepsilon_{it} \\ \varepsilon_{it} &= u_i + w_{it} \end{aligned} \qquad [24]$$

utilizando como instrumentos para $Y_{i,\ t-1}$, una selección de retrasos de X_t ya que podemos escribir por substitución recursiva y siempre que el modelo sea estable $(|\alpha| < 1)$,

$$E(Y_{it}) = X'_t B + \alpha X_{t-1} B + \alpha^2 X_{t-2} B + \cdots \qquad [25]$$

Con las estimaciones de (α y B) se estima ε, y su matriz de covarianzas, para aplicar en un segundo paso, el estimador de MCG.

Sin embargo, hay que resaltar el hecho de que este estimador también da sesgos, que son además de orden $(1/\sqrt{T})$ (véase Apéndice 2), es decir, de mayor orden que los proporcionados por el estimador de efectos fijos. No obstante, la matriz de covarianzas de este estimador para T pequeño es probable que sea mucho menor, de modo que en conjunto, el error cuadrático medio del estimador descrito en ([24], [25]), sea menor que el del estimador de efectos fijos. Para resumir, podemos decir que en el caso de que los retrasos de la variable dependiente no pertenezcan a los regresores, es más robusto el estimador de efectos fijos, y probablemente es preferible de acuerdo a este criterio. Si el modelo es dinámico de tipo autorregresivo, y los efectos individuales son aleatorios, es probable que el estimador de dos etapas que combina variables instrumentales y MCG sea mejor, en la medida en que es probable que su error cuadrático medio sea menor.

Otro criterio, trivial pero no menos importante, es simplemente el de seleccionar el modelo que conduzca a resultados más plausibles o interpretables. Esto es especialmente así en este caso, dada la diversidad de criterios para elegir un modelo u otro. De todas formas ya se ha advertido anteriormente (véase párrafo inmediatamente posterior a [19]), que es típico el fallo del modelo de efectos aleatorios debido a la omisión de variables, lo que causa un R^2 muy bajo.

Un criterio importante en la elección, es también el de la preferencia siempre, manteniendo el resto de condiciones iguales, del modelo de menos parámetros. Un modelo de menos parámetros es más eficiente si las restricciones impuestas por comparación a otro más general son ciertas. Y si son falsas, en gran número de ocasiones, proporciona errores cuadráticos medios menores. Dado que en modelos de datos en panel, lo típico es que T sea pequeño y N alto, utilizar el modelo de efectos aleatorios elimina la necesidad de estimar $(N - 1)$ constantes individuales. Además, estas constantes se estiman con bastante imprecisión dado el tamaño de T. Por ejemplo, si $T = 2$, $N = 500$, introducir 500 constantes individuales, reduce el número de observaciones efectivas para estimar el vector B, de 1.000, a 500, es decir, la mitad. Además, si la selección de variables es apropiada, el ajuste no tiene porqué ser bajo, y el modelo de efectos aleatorios, además de dar resultados aceptables ($\hat{\sigma}_u^2 > 0$), proporciona una estimación mucho más eficiente del vector 'B' y en general de todo el modelo.

d) Un ejemplo ilustrativo

Puede ser ilustrativo finalizar esta discusión con el examen de una situación típica. En general, podemos suponer que T será pequeño (< 10), y N bastante alto. Lo natural es que haya retrasos de la variable dependiente entre los regresores, pues las relaciones económicas son dinámicas. Para economizar en la estimación de parámetros, se puede suponer que los efectos individuales son aleatorios, y los temporales fijos (T «dummies»). Los pasos a seguir para estimar este modelo son los dos siguientes:

1. Estimar el modelo original por variables instrumentales, utilizando como instrumentos para los retrasos de la variable dependiente, retrasos de las variables exógenas. Con la estimación consistente de los errores, estimar su matriz de covarianzas.

2. Aplicar el estimador de mínimos cuadrados generalizados, utilizando la transformación apropiada de los datos dada en [13] y [14]. Las varianzas σ_u^2, σ_w^2 se substituyen por las estimaciones obtenidas en el paso anterior.

e) Otros problemas de interés

Para finalizar la discusión de los modelos para datos en panel, se comentan a continuación algunos tópicos de interés no discutidos anteriormente, y que son *a*) aplicaciones del análisis de covarianzas; *b*) diferenciación del modelo; *c*) estimación por máxima verosimilitud.

En muchas ocasiones puede ser interesante considerar efectos de grupo, más que individuales, pero clasificando a las observaciones de acuerdo a varios criterios (véase, por ejemplo, Martínez Mongay, 1986). Por ejemplo, podemos considerar varios criterios dicotómicos para clasificar empresas como son exporta-no exporta, invierte en *I* y *D* o no, sólo una planta o más, etc. Entonces, puede ser interesante realizar un análisis de efectos al estilo del análisis de varianza (en este caso, cuando hay otros regresores aparte de las constantes, se suele denominar a este análisis de «covarianza», pero esta denominación es arbitraria).

Para simplificar vamos a suponer que sólo hay dos criterios de clasificación, cada uno de ellos con varias categorías. Esto da una clasificación cruzada de (IJ) casillas, siendo I y J el número de categorías respectivamente en cada criterio. Vamos a denominar m_{ij} a la media de cada casilla después de eliminar el efecto de otras variables exógenas. Es decir, el modelo sería

$$Y_{ijp} = X'_{ijp}B + m_{ij} + \varepsilon_{ijp} \tag{26}$$

El análisis de varianza, permite ahora descomponer el efecto total 'm_{ij}' de los dos factores, en contribuciones individuales y un efecto de interacción. Más concretamente podemos escribir,

$$\bar{m}_{ij} = \bar{m}_{i.} + \bar{m}_{.j} + (\bar{m}_{ij} - \bar{m}_{i.} - \bar{m}_{.j})$$

donde

$$\begin{aligned} \bar{m}_{ij} &= m_{ij} - m \\ \bar{m}_{i.} &= m_{i.} - m \\ m_{i.} &= \left(\sum_{j=1}^{J} m_{ij}\right)/J \\ m &= \left(\sum_{i=1}^{I} m_{i.}\right)/I \end{aligned} \tag{27}$$

Es decir, en [27] el efecto total lo descomponemos en los efectos individuales ($m_{i.}$, $m_{.j}$) y un tercer efecto de carácter multiplicativo en el sentido de que no se reduce a la suma de los individuales, o en otras palabras, es un efecto interactivo. En [27] todos los efectos están medidos en desviaciones sobre las medias. Si escribimos la expresión en función de expresiones no centradas tenemos,

$$\begin{aligned} m_{ij} &= m + m_{i.} + m_{.j} + (m_{ij} - m_{i.} - m_{.j} + m) \\ &= \theta + \theta_{i.} + \theta_{.j} + \theta_{ij} \end{aligned} \tag{28}$$

y así, hemos descompuesto el efecto total 'm_{ij}', en un efecto general θ, dos efectos de grupo ($\theta_{i.}$, $\theta_{.j}$) y un efecto debido a la interacción θ_{ij}. Si se introducen más variables se pueden analizar efectos interactivos más complicados, pero el análisis pierde algo de su interés por la dificultad de interpretación. No obstante, aun cuando haya más criterios, el análisis de efectos interactivos de primer orden puede permitir interpretaciones interesantes. En el análisis tradicional de varianza, la descomposición de [28] conduce a estimaciones y contrastes muy sencillos que pueden llevarse a cabo manualmente (de ahí deriva probablemente su popularidad en la época anterior a los ordenadores). En el modelo [26], los estimadores mínimo cuadráticos de m_{ij} estarán dados por

$$\hat{m}_{ij} = \sum_{p} (Y_{ijp} - X_{ijp}'\hat{B})/n_{ij} \tag{29}$$

siendo n_{ij} el número de observaciones en la categoría correspondiente, y B el estimador mínimo cuadrático de B. Los estimadores de (θ, $\theta_{i.}$, $\theta_{.j}$, θ_{ij}) pueden obtenerse ahora substituyendo estos estimadores en las expresiones [27] y [28].

Para hacer contrastes de hipótesis, observamos que las θ's son combinaciones lineales de los m_{ij}, así que basta con aplicar resultados elementales de regresión. Por ejemplo, la varianza de $m_{i.}$ estará dada por,

$$V(\hat{m}_{i.}) = \left(e'\Sigma(m_{ij})e\right)/J^2 \qquad [30]$$

donde 'e' es un vector de unos y $\Sigma(m_{ij})$ la matriz de covarianzas del vector $(m_{i1}, m_{i2}, ..., m_{iJ})$.

Es indudable que este tipo de análisis permite una descomposición muy informativa de las constantes de cada ecuación. Este tipo de análisis puede ser especialmente útil cuando $T = 1$, de modo que no sea posible introducir efectos individuales diferentes para cada una de las observaciones.

Cuando hay efectos individuales constantes, y $T > 1$, una técnica que se ha sugerido en ocasiones para estimar el modelo es diferenciarlo. Es decir, suponiendo que el modelo inicial es

$$Y_{it} = \gamma_i + X_{it}'B + \varepsilon_{it} \qquad [31]$$

diferenciando obtenemos,

$$\Delta Y_{it} = \Delta X_{it}'B + \Delta\varepsilon_{it} \qquad [32]$$

de modo que los efectos individuales desaparecen. La matriz de covarianzas de los errores no es diagonal, pero los estimadores MCO son consistentes al menos.

Si hay retrasos de la variable dependiente entre los regresores, el estimador MCO no es consistente, pero siempre se puede aplicar MCG. De todas formas este método requiere invertir la matriz de covarianzas de 'ε', y dadas las dimensiones de este vector, esto puede ser problemático (véase párrafo posterior a [4]).

El último punto que queda por comentar, es el de la estimación del modelo por métodos de máxima verosimilitud. En el caso del modelo de efectos aleatorios Maddala (1971) demuestra que hay varios óptimos locales. En el caso de modelos dinámicos Anderson y Hsiao (1982) obtienen entre otros resultados, que el estimador de máxima verosimilitud no es consistente en casos en que los estimadores convencionales de efectos fijos y aleatorios lo son. Por otra parte, los programas econométricos existentes en el mercado no incorporan esta posibilidad como una opción rutinaria. Por todo ello, desde el punto de vista aplicado, parece aconsejable descartar el estimador máximo verosímil (sin perjuicio de que cuando $(N, T) \to \infty$, sea equivalente a los estimadores convencionales discutidos en esta sección).

2.2. *Modelos para variables dependientes cualitativas y truncadas*

El tratamiento tradicional de los modelos de variables cualitativas es el análisis discriminante. Este tipo de análisis, básicamente está dirigido a encontrar la combinación lineal de variables, que mejor discrimine entre los individuos de dos o más grupos. Dada esta combinación, un individuo cualquiera cuya procedencia se ignora, puede adscribirse a un grupo o a otro según la proximidad del valor de la mencionada combinación a los valores del grupo. Es decir, si ocurre que

$$|\lambda'(x_i - \bar{x}_1)| < |\lambda'(x_i - \bar{x}_2)| \qquad [33]$$

se concluye que el individuo '*i*' proviene del grupo '1' (λ es la combinación discriminante, '*x*' un vector de '*k*' variables ($\bar{x}_1$, $\bar{x}_2$) las medias de estos vectores en los grupos 1 y 2, y x_i el vector de observaciones para el individuo '*i*').

Este tipo de análisis es discontinuo, y relativamente «ad-hoc». Un tratamiento más interesante de este problema lo proporciona la estimación econométrica de la probabilidad de pertenencia a un grupo o a otro.

Respecto al análisis de modelos truncados, es relativamente nuevo, y no hay modelos equiparables fuera de la literatura econométrica.

La aplicabilidad de este tipo de modelos es enorme, especialmente en datos transversales y de panel. Por ejemplo, podemos pensar en los determinantes de la probabilidad de que una empresa exporte o no (Maravall *et. al* (1986, Alonso (1986)) o en los determinantes de la probabilidad de que una empresa decida aumentar, disminuir o dejar igual su producción (Bergés (1986b)), e incluso el análisis puede ser aplicado a la estimación de las probabilidades de transición en una cadena de Markov (Pérez Gorostegui (1986)). Podemos también pensar en analizar los determinantes de que una empresa decida exportar o no, y si decide exportar cuánto (análisis de variables truncadas). Para modelos discretos, y en lugar de un análisis truncado, puede aplicarse un análisis similar basado en la distribución de Poisson. Por ejemplo, podemos estudiar los determinantes del número de patentes registradas por una empresa determinada.

Esta sección va a dedicarse a la presentación, en primer lugar, de los modelos para variables dependientes cualitativas. En segundo lugar, se discutirán los modelos para variables truncadas.

a) Modelos de respuesta cualitativa

El problema básico que tratan de resolver los modelos de respuesta cualitativa es determinar una función de ciertas variables que permitan atribuir la pertenencia de un individuo a un grupo. El tratamiento del análisis discriminante (*AD*) ya ha sido mencionado en la introducción a esta sección. El enfoque econométrico, consiste en tratar de especificar los determinantes de la probabilidad de pertenencia a un grupo u otro. Además, el problema se plantea dentro del enfoque máximo verosímil lo que se traduce en las propiedades de optimalidad usuales, y en que el contraste de hipótesis y estimación de modelos puede llevarse a cabo con facilidad aplicando resultados generales de máxima verosimilitud.

— *Modelos dicotómicos*

Para motivar el análisis posterior, es útil comenzar por la solución de regresión. Para ello, supongamos que una variable 'Y_i' puede tomar dos valores (1, 0) que representan la pertenencia a dos grupos (por ejemplo, exporta o no exporta), con probabilidades respectivas p_i $(1 - p_i)$. Si suponemos que $p_i = x_i'B$, podemos escribir,

$$Y_i = x_i'B + u_i \tag{34}$$

donde, $Eu_i = 0$, $V(u_i) = p_i(1 - p_i)$. Los coeficientes de B, puede estimarse ahora por MCO, o por MCG en dos etapas, y esta es una solución rápida y sencilla del problema (muy popular en los años sesenta). Sin embargo, una dificultad importante de esta solución reside en la interpretación de $(x_i'B)$ como una probabilidad ya que en la práctica, no hay nada que asegure que pertenezca al intervalo (0, 1).

La solución del *AD* ya se ha comentado anteriormente. Una ventaja de este enfoque es que permite manejar con facilidad un número arbitrario de elementos de cada grupo. Si se desea analizar en detalle los determinantes de alternativas poco frecuentes, suele ser más eficiente aumentar la proporción de ese tipo de observaciones por encima de su proporción poblacional. Este tipo de problema puede analizarse también dentro del enfoque máximo verosímil, aunque su tratamiento es mucho más complejo. La raíz del problema estriba en este caso en que la selección de los individuos que pertenecen a la muestra ya no es aleatoria simple, y por tanto, la función de verosimilitud es bastante más complicada (Cosslett (1981)).

El *AD* también permite determinar la probabilidad de pertenencia a un grupo pero con ciertas limitaciones. Por ejemplo, supongamos que un individuo 'i' puede provenir de dos poblaciones (1, 2) con probabilidades conocidas p $(1 - p)$. Supongamos que cada individuo se caracteriza por el valor de un vector de variables aleatorias 'x' y que la función de densidad de x en las poblaciones (1, 2) es $g_1(x)$, $g_2(x)$ respectivamente. Entonces, por el teorema de Bayes obtenemos inmediatamente la expresión

$$\begin{aligned} p(Y_i = 1 \mid x_i) &= \frac{p(Y_i = 1)p(x_i \mid Y_i = 1)}{p(x_i)} \\ &= \frac{pg_1(x_i)}{pg_1(x_i) + (1 - p)g_2(x_i)} \end{aligned} \tag{35}$$

que nos da la probabilidad de pertenencia al grupo 1 del individuo 'i'. Si las $g(\cdot)$ son normales, la expresión [35] se reduce a un modelo de tipo logit (ver más adelante). Además en este caso, el contraste de que una variable es relevante o no en la determinación de 'p_i' puede llevarse a cabo simplemente contrastando la hipótesis $Ex_{1\alpha} = Ex_{2\alpha}$, es decir, que la media de dicha variable sea igual en ambas poblaciones. Para que este contraste sea válido es preciso que las matrices de covarianzas de 'x' en ambas poblaciones sean idénticas, y este hecho unido al supuesto de normalidad limita considerablemente la utilidad de la solución [35] (puede ser útil recordar que las variables cualitativas, son necesariamente no normales).

El enfoque econométrico actual consiste en incorporar al modelo de [34] la restricción de que $0 < p_i < 1$, implícitamente. Para ello, basta con definir

$$p_i = p(x_i'B) \qquad [36]$$

y asegurar que la función escogida $p(\cdot)$ pertenece al intervalo (0, 1). Una clase de funciones que cumple esta condición es precisamente las funciones de distribución de variables aleatorias. Las dos soluciones más populares son la función de probabilidad de una normal y el modelo logístico. Es decir,

$$p(x_i'B) = \int_{-\infty}^{x_i'B} \frac{1}{\sqrt{2\pi}} \exp\left(-(u^2/2)\right) du$$

$$p(x_i'B) = \frac{e^{x_i'B}}{1 + e^{x_i'B}} \qquad [37]$$

En la práctica, ambos modelos suelen dar resultados muy similares, lo que no es de sorprender, dada la semejanza de las funciones de densidad respectivas. El modelo logístico tiene ciertas ventajas desde el punto de vista del cálculo, ya que no hay disponible una solución analítica para la integral normal (aunque se puede aproximar con facilidad como se verá más adelante). Los modelos que utilizan la función Normal se suele denominar «Probit» y los que utilizan la logística, «Logit».

Una vez escogida la función $p(\cdot)$, el paso siguiente es escribir la función de verosimilitud. Como las probabilidades son independientes, podemos escribirla como sigue,

$$\phi = \prod_1 p(x_i'B) \prod_0 \left(1 - p(x_i'B)\right) \qquad [38]$$

donde la opción '1' es elegida con probabilidad p_i y la '0' con $1 - p_i$. Definiendo la variable auxiliar $Y_i = 1$, si el individuo 'i' escoge la alternativa '1', se puede escribir más compactamente [38] en la forma,

$$\phi = \prod_{i=1}^{N} p(x_i'B)^{Y_i}(1 - p(x_i'B))^{(1-Y_i)} \qquad [39]$$

Esta función puede maximizarse ahora con respecto a 'B', introduciéndola en una rutina de optimización. Como de costumbre, se cumplirá que

$$(B - B)\ \sqrt{N_{\tilde{A}}} N(0, \Sigma) \qquad [40]$$

cuando $N \to \infty$, y donde,

$$\Sigma = -\text{plim}\left(1/N\partial^2 \log \phi/(\partial B)(\partial B')\right)^{-1} \quad [41]$$

Puede demostrarse (véase, por ejemplo, Amemiya (1983)) que la matriz de segundas derivadas es cóncava respecto al vector B, y para todo posible valor de este vector. Esto asegura buenas propiedades de convergencia de cualquier algoritmo que resuelva la maximización de [39], y en particular, que no hay máximos locales. De todas formas es interesante analizar explícitamente la iteración de Gauss-Newton. Para ello, observamos en primer lugar que el modelo que da base a la función de verosimilitud [39] puede escribirse similarmente a [34] del modo siguiente,

$$\begin{aligned} Y_i &= p(x_i'B) + u_i \\ Eu_i &= 0, \quad V(u_i) = p_i(1 - p_i) \end{aligned} \quad [42]$$

Este es un modelo no lineal, heterocedástico, y no normal. Sin embargo, cuando $N \to \infty$, la no normalidad es irrelevante, así que parece lógico pensar en corregir el modelo para hacerlo homocedástico y aplicar mínimos cuadrados no lineales. Este último método, puede demostrarse que es equivalente a aplicar MCO en una expansión de Taylor del modelo no lineal, e iterarla. En el modelo que nos ocupa tendríamos

$$\begin{aligned} Y_i &\simeq p(x_i'B_0) + f(x_i'B_0)x_i'(B - B_0) + u_i \\ &= (p_{i0} - f_{i0}x_i'B_0) + f_{i0}x_i'B + u_i \\ &= \delta_i + f_ix_i'B + u_i \end{aligned} \quad [43]$$

siendo 'f' la función de densidad correspondiente a p, y corrigiendo la heterocedasticidad el estimador de B vendría dado por la expresión,

$$\begin{aligned} B = &\left(\sum_{i=1}^{N} \frac{f_i^2}{p_i(1 - p_i)} x_ix_i'\right)^{-1} \\ &* \left(\sum_{i=1}^{N} \frac{f_i}{p_i(1 - p_i)} x_i(Y_i - \delta_i)\right) \end{aligned} \quad [44]$$

En esta expresión (f_i, p_i, δ_i) son funciones de B_0. Entonces, tenemos ya la base para una iteración. De hecho, puede demostrarse que el estimador máximo-verosímil de [39] conduce precisamente a esta solución, que como hemos visto, es de fácil interpretación. Por otra parte, si no se dispone de un programa que incorpore explícitamente la opción de estimar [39], siempre puede programarse por medio de [44] que es una expresión sencilla (o simplemente utilizando [43]) si se dispone de un procedimiento de estimación para mínimos cuadrados no lineales).

Valores para las razones 't' de los estimadores B, pueden ser obtenidos del modo usual, utilizando la estimación de la matriz de covarianzas, dada en [41].

El contraste de hipótesis compuestas, puede instrumentarse por medio del contraste de Wald, e incluso más fácilmente, utilizando el contraste de máxima verosimilitud. Con esto, quedan resueltos los dos problemas más importantes, estimación y contraste del modelo.

Un punto pendiente de interés, que presenta ciertas particularidades en este caso es el de la medida de la bondad del ajuste. En el caso de los mínimos cuadrados ordinarios para un modelo lineal con errores normales, es sabido que se cumple,

$$\begin{aligned} R^2 &= 1 - (\tilde{L}/\hat{L})(2/N) \\ &= 1 - (\hat{\sigma}/\tilde{\sigma})^2 \end{aligned} \qquad [45]$$

siendo L la función de máxima verosimilitud (no su logaritmo), σ la varianza de los errores del modelo, y donde los símbolos ($\hat{}$), ($\tilde{}$) representan estimadores no restringidos y restringidos respectivamente (véase Apéndice 3). Por analogía, podríamos definir una medida de bondad de ajuste basándonos en la función de verosimilitud [39] de la forma siguiente:

$$R^2 = 1 - (\tilde{\phi}/\hat{\phi})(2/N) \qquad [46]$$

La dificultad con esta medida es que nunca puede alcanzar el valor 1 dado que la forma de [39] implica que $(0 < \tilde{\phi} < \hat{\phi} < 1)$. Más concretamente,

$$R^2 < 1 - (\tilde{\phi})(2/N) \qquad [47]$$

cumpliéndose la igualdad sólo cuando ϕ alcanza su valor máximo, es decir, $\phi = 1$. La sugerencia es entonces la medida corregida,

$$R_c^2 = \frac{1 - (\tilde{\phi}/\hat{\phi})^{2/n}}{1 - \tilde{\phi}^{(2/N)}} \qquad [48]$$

y que es igual a uno, cuando ϕ alcanza su valor máximo, e igual a cero cuando $\tilde{\phi} = \hat{\phi}$, como debería ser. (Notar también que $R_c^2 > R^2$.)

Una aplicación muy interesante de este tipo de análisis es la predicción agregada de la proporción poblacional de individuos que optarán por una alternativa ante el cambio de una de las variables del vector x_i. (Por ejemplo, como cambiará la proporción de empresas exportadoras ante un cambio en la desgravación fiscal a la exportación, si es que esta variable es relevante).

La estimación obvia está dada por

$$q = \frac{1}{N} \sum_i Y_i \qquad [49]$$

donde

$$Y_i = p(\bar{x}_i' B) + u_i$$

$$Eu_i = 0, \quad V(u_i) = p_i(1 - p_i)$$

siendo $\bar{x}_i$ el vector de variables determinantes que puede incluir variables controlables (la barra sobre la 'x' quiere indicar que algunas variables se han prefijado). Entonces obtenemos fácilmente,

$$Eq = \frac{1}{N} \sum_i EY_i$$

$$= \frac{1}{N} \sum_i p_i$$

$$V(q) = \frac{1}{N^2} \sum_{i=1}^{N} p_i(1 - p_i) \qquad [50]$$

y por aplicación de un teorema central del límite para sumas de variables con varianzas diferentes pero acotadas con ciertas condiciones (véase, por ejemplo, Theil (1971)), obtenemos normalidad asintótica. Así, para muestras finitas podemos utilizar la siguiente aproximación para establecer intervalos de confianza del estimador 'q',

$$q_{\tilde{A}} N \left(\frac{1}{N} \sum_i p_i, \frac{1}{N^2} \sum_i p_i(1 - p_i) \right) \qquad [51]$$

Esta expresión es, por supuesto, condicional en las distribuciones de 'x' y '$\hat{B}$'. La expresión incondicional para la media puede obtenerse tomando esperanzas respecto a 'x' y '$\hat{B}$' (la varianza es algo más complicada). En todo caso, la solución analítica parece ser extremadamente complicada si es que es posible (véase, por ejemplo, Amemiya (1985)). De todas formas para disponer de una aproximación a grosso modo, puede ser suficiente con la expresión [51] si la muestra inicial es suficientemente representativa de la distribución de 'x', de modo que también ocurra que $\hat{B} \simeq B$.

— *Modelos de respuesta múltiple*

Estos resultados pueden extenderse con facilidad al caso de que la elección, o clasificación del individuo, sea en más de dos categorías (modelos multinominales). En este caso definimos una variable para el individuo 'i', z_i que puede tomar los valores (0, 1, 2, ..., m), y definimos

$$p(z_i = j) = p(x_{ij}'B) = p_{ij} \qquad [52]$$

Se puede comenzar desde un modelo más general en varios sentidos (Amemiya (1985)). Por ejemplo, se puede suponer que el conjunto de posibilidades para cada individuo es diferente de modo que en lugar de 'm' tendríamos 'm_i'. También, puede escribirse la dependencia de la probabilidad de modo más general como p_{ij} $(x'B)$. Para modelos sencillos que son los que vamos a ver aquí, la especificación de [52], junto con el supuesto $m_i = m$ es suficiente.

Para definir la función de verosimilitud, definimos una serie de variables auxiliares del siguiente modo,

$$Y_{ij} = \begin{cases} 1 & \text{si } z_i = j \\ 0 & \text{si } z_i \neq j \end{cases} \qquad [53]$$

y entonces, similarmente a [39] podemos escribir:

$$\phi = \prod_{i=1}^{N} \prod_{j=0}^{m} p(x_{ij}'B)^{Y_{ij}} \qquad [54]$$

Para obtener el estimador máximo verosímil de 'B' maximizamos esta expresión, introduciéndola en cualquier rutina de optimización. Puede demostrarse que se cumplen, al igual que en el caso binomial, los resultados [40] y [41]. También se cumple la concavidad global de ϕ respecto a B, de modo que la convergencia de cualquier rutina de optimización al máximo, no debe plantear problemas especiales. La interpretación del modelo y su estimación, como un problema de estimación de mínimos cuadrados no lineales con errores heterocedásticos, del caso binomial, puede extenderse también al caso multinominal. Para ello a partir de las variables Y_{ij} definidas en [53] definimos los vectores siguientes,

$$\begin{aligned} Y_i' &= (Y_{i1}, Y_{i2}, \ldots, Y_{im}) \\ p_i' &= \left(p(x_{i1}'B);\ \ p(x_{i2}'B), \ldots, p(x_{im}'B)\right) \end{aligned} \qquad [55]$$

y análogamente a [42] escribimos,

$$\begin{aligned} y_i &= p_i + u_i, \quad i = 1, \ldots, N \\ Eu_i &= 0, \quad Eu_iu_i' = \Lambda_i, \quad Eu_iu_j = 0 \quad i \neq j \end{aligned} \qquad [56]$$

donde Λ_i es una matriz de dimensión 'm' cuyos elementos están definidos como sigue

$$\begin{aligned} &p_{ia}(1 - p_{ia}) && : \text{ elemento diagonal} \\ &-p_{ia}p_{ib} && : \text{ elemento fuera de la diagonal} \end{aligned}$$

Estos resultados se derivan del hecho de que el vector 'Y_i' es un vector multivariante, de variables aleatorias discretas, cuya distribución conjunta es una multinomial, donde el número de observaciones es una. Las distribuciones marginales de las variables Y_{ij} son binomiales de parámetros $(1, p_{ij})$ (véase, por ejemplo, Mood *et. al.* (1974) para la distribución multinomial). De forma análoga a [43] podemos expandir [56] por Taylor y aplicar un estimador de mínimos cuadrados generalizados, que iterado, será una solución equivalente a la máxima verosímil. Para ilustrar este resultado podemos efectuar la expansión para una variable cualquiera 'Y_{ij}' de [56] y obtenemos lo siguiente:

$$\begin{aligned} Y_{ij} &\simeq p(x_{ij}'B_0) + f(x_{ij}'B_0)X_{ij}'(B - B_0) + u_{ij} \\ &= (p_{ij0} - f_{ij0}B_0) + f_{ij0}B + u_{ij} \end{aligned} \quad [57]$$

donde $f(\cdot)$ es la función de densidad correspondiente a $p(\cdot)$. Corrigiendo ahora la heterocedasticidad de 'u', y partiendo de un estimador inicial B_0 (por ejemplo, el MCO lineal) podemos utilizar [57] para obtener sucesivos estimadores iterados de 'B'. Como en el caso binomial, esta iteración, de hecho coincide con el estimador máximo verosímil cuando converge.

Una cuestión pendiente en [54] es la forma de las funciones $p(\cdot)$. En el caso binomial, el problema es más sencillo, y como hemos visto cualquier función de distribución es válida. En el caso multinomial, cada individuo 'i', tiene varias posibilidades con probabilidad p_{ij}. Similarmente al modelo dicotómico en el que $p_i + (1 - p_i) = 1$, debe cumplirse que la suma de todas las probabilidades sea la unidad. En este caso esto implica entonces que $\sum_j^m p_{ij} = 1$, para todo 'i'. Es decir, que la forma de la función que elijamos debe cumplir esta propiedad. Una función que la cumple, además de la usual $0 < p_{ij} < 1$ es la logística multinomial que es la siguiente:

$$p(x_{ij}'B) = \exp(x_{ij}'B) / \left(\sum_{h=0}^{m} \exp(x_{ih}'B) \right) \quad [58]$$

y que como fácilmente se observa, es una simple generalización de la logística binomial [37]. Esta función tiene además la ventaja de su facilidad de cálculo. La definición de p_i, a partir de funciones de densidad normales es muy compleja y de cálculo costoso de forma que sólo en casos muy sencillos se pueden llevar a cabo los cálculos. La escasa evidencia disponible, no demuestra en todo caso que los modelos probit multinomiales sean claramente superiores a los logísticos. (Hausman y Wise (1978.)

— *Independencia de alternativas irrelevantes*

La fórmula logística multinomial tiene una propiedad que en ciertos casos puede no ser deseable. Vamos a analizarla por medio del famoso ejemplo de Mc Fadden (1974). Supongamos que un individuo se enfrenta a la decisión de elegir un medio de transporte, coche o autobús. Si el individuo es indiferente frente a estas dos alternativas, esperamos que se cumpla,

$$p(Y_i = 1)/p(Y_i = 0) = (1/2)/(1/2) = 1 \quad [59]$$

siendo '1' la elección del autobús y '0' la del coche.

Supongamos ahora que el conjunto de posibilidades de elección se amplía de forma que la elección a la que se enfrenta ahora un individuo es coche (0) autobús rojo (1), o autobús azul (2). Parece lógico suponer que la diferencia entre autobús rojo y azul no puede ser mayor que la diferencia entre la elección de coche o autobús. Si el individuo es indiferente ante los dos tipos de autobuses, lo esperable es que se cumpla lo siguiente

$$\begin{aligned} p(Y_i = 0) &= 1/2, \quad p(Y_i = (1 \text{ ó } 2) = 1/2 \\ p(Y_i = 1) &= 1/4, \quad p(Y_i = 2) = 1/4 \end{aligned} \qquad [60]$$

de forma que se cumplirá

$$p(Y_i = 1)/p(Y_i = 0) = 1/2 \qquad [61]$$

Si las tres alternativas fuesen totalmente diferentes y el individuo fuese indiferente ante ellas, esperaríamos que se cumpliese lo siguiente

$$\begin{gathered} p(Y_i = 0) = p(Y_i = 1) = p(Y_i = 2) = 1/3 \\ p(Y_i = 1)/p(Y_i = 0) = 1 \end{gathered} \qquad [62]$$

Es evidente que en este ejemplo, lo lógico es suponer que las dos alternativas autobús rojo o azul son muy parecidas, de modo que en el caso de [62], estaríamos sobrevalorando las probabilidades de elección de las dos alternativas (1/3 frente a 1/4). Vemos también en [62] que el cociente entre la probabilidad de las dos elecciones, coche o autobús rojo, es el mismo que en el caso de [59], en el que sólo hay un tipo de autobús. Es decir, la fórmula de [62] no se ve afectada por la existencia o no de otras alternativas próximas. Esta propiedad se suele denominar «independencia de alternativas irrelevantes», y en el caso de este ejemplo, es claramente no deseable. Consideremos ahora qué ocurre con la fórmula multinomial. Entonces tenemos lo siguiente,

$$p(Y_i = 1)/p(Y_i = 0) = \exp(x_{i1}'B)/\exp(x_{i0}'B) \qquad [63]$$

Para la derivación de esta fórmula, es irrelevante que el conjunto de alternativas sea (0, 1) ó (0, 1, 2), y que sean similares o no. Ello es así, porque estas alternativas adicionales sólo afectan al denominador de [58], que desaparece en la formulación de [63]. Vemos entonces que el modelo logístico multinomial cumple esta propiedad que puede ser no deseable en muchos casos. Para evitar esta dificultad McFadden ha desarrollado el modelo logístico condicional, o anidado. Una ventaja de este modelo es que se reduce al logístico simple (que cumple la propiedad de independencia de alternativas irrelevantes), cuando cierto parámetro es cero. Entonces, la propiedad discutida puede contrastarse formalmente estableciendo como hipótesis nula, que dicho parámetro es cero. Este modelo es, sin embargo, algo complicado, y desde el punto de vista práctico, probablemente puede resolverse la dificultad mencionada de un modo mucho más sencillo. El ejemplo anterior, puede ser enfocado, alternativamente al modelo de McFadden, como un problema de elección secuencial: en primer lugar se elige entre autobús o coche, y si se elige el autobús, en una segunda fase, se elige entre autobús rojo o azul. Vamos a denominar la elección de coche por '0', y la elección de autobús por '1'. La elección de autobús rojo se representará por $(1, 0)$ y la de azul por $(1, 1)$. Gráficamente, el problema de decisión puede representarse por el siguiente árbol,

$$\left\langle \begin{array}{l} 0 \\ 1 \quad \left\langle \begin{array}{l} 1 \\ 0 \end{array}\right. \end{array}\right. \qquad [64]$$

El modelo que explica las probabilidades puede ser escrito ahora como sigue,

$$\begin{aligned} p(Y_i = 0) &= p(x_{1i}'B_1) \\ p(Y_i = 1) &= 1 - p(x_{1i}'B_1) \\ p(Y_i = (1, 1)) &= (1 - p(x_{1i}'B_1))p(x_{2i}'B_2) \\ p(Y_i = (1, 0)) &= (1 - p(x_{1i}'B_1))(1 - p(x_{2i}'B_2)) \end{aligned} \qquad [65]$$

La función de máxima verosimilitud es entonces,

$$\begin{aligned} \phi &= \prod_{i=1}^{N} p(x_{1i}'B_1)^{Y_i}(1 - p(x_{1i}'B_1))^{1-Y_i} \\ &* \prod_{J=1}^{N_1} p(x_{2i}'B_2)^{w_j}(1 - p(X_{2i}'B_2)^{1-w_j} \end{aligned} \qquad [66]$$

donde $w_j = 1$ si se elige $(1, 1)$ y cero si se elige $(1, 0)$ y N_1 es el número de individuos que eligen la alternativa 1 en la primera fase. La enorme ventaja de este enfoque es que la función de máxima verosimilitud se reduce al producto de dos funciones binomiales independientes, de forma que su optimización es muy sencilla. Es evidente también que este enfoque ya no cumple la propiedad de «independencia de alternativas irrelevantes».

El esquema de decisión de [64] puede generalizarse con facilidad a modelos más complicados. Por ejemplo, si el árbol de decisiones es el siguiente:

$$\left\langle \begin{array}{l} 0 \quad \left\langle \begin{array}{l} 0 \\ 1 \end{array}\right. \\ 1 \quad \left\langle \begin{array}{l} 0 \\ 1 \end{array}\right. \end{array}\right. \qquad [67]$$

la definición de las probabilidades estará completa si añadimos a [65] la siguiente especificación

$$\begin{aligned} p(Y_i = (0, 0)) &= p(x_{1i}'B)p(x_{3i}'B_3) \\ p(Y_i = (0, 1)) &= p(x_{1i}'B)(1 - p(x_{3i}'B_3)) \end{aligned} \qquad [68]$$

La función de verosimilitud sería ahora,

$$\begin{aligned}\phi = &\prod_{i=1}^{N} p_{1i}^{Y_i}(1 - p_{1i})^{1-Y_i} \\ &* \sum_{j=1}^{N_0} p_{3j}^{w_j}(1 - p_{3j})^{1-w_j} \\ &* \prod_{k=1}^{N_1} p_{2k}^{z_k}(1 - p_{2k})^{1-z_k}\end{aligned} \qquad [69]$$

donde el significado de los símbolos se deduce fácilmente por analogía al caso [66]. Con igual facilidad pueden tratarse los casos multinomiales, que ya no se especifican aquí, por tratarse de un mero ejercicio práctico.

— *Predicción*

Una de las aplicaciones más útiles del modelo logístico multinomial, es la predicción del grado de aceptación de nuevas alternativas. Por ejemplo, supongamos que en [58] podemos escribir

$$x_{ij}'B = x_i'B_1 + x_j'B_2 \qquad [70]$$

donde 'x_1' son características individuales y 'x_j' características asociadas a la elección 'j' (p_{ij}, puede ser, por ejemplo, la probabilidad de que una empresa exporte más a una zona económica determinada, si desaparecen determinadas barreras arancelarias, medidas por x_j). Si podemos determinar el valor 'x_j' para una nueva alternativa $j = m + 1$, la probabilidad de que el individuo 'i' la elija viene dada por la expresión

$$p(z_i = m + 1) = \exp\,(x_{i,\,m+1}'B)/\left(\sum_{h=0}^{m+1} \exp\,(x_{ih}'B)\right) \qquad [71]$$

y las probabilidades de elección de las diferentes alternativas se modifican similarmente (cambia solamente el denominador). Si se desean establecer intervalos de confianza para la proporción total de la población que escogerá esta nueva alternativa, puede hacerse de modo similar al caso binomial [51]. En el caso multinomial definimos las '$m + 1$' variables siguientes:

$$z_j = \sum_{i=1}^{N} Y_{ij}, \quad j = 0, \ldots, m + 1 \qquad [72]$$

cuya esperanza y varianzas vienen dadas por,

$$\begin{aligned}E(z_j) &= \sum_{i=1}^{N} p_{ij} \\ V(z_j) &= \sum_{i=1}^{N} p_{ij}(1 - p_{ij})\end{aligned} \qquad [73]$$

ya que para extracciones independientes (diferentes i's) las variables son independientes. Definiendo

$$r_{m+1} = (z_{m+1}/N) \qquad [74]$$

donde N es el número total de individuos en la muestra, tendremos asintóticamente la siguiente distribución,

$$r_{m+1} \tilde{A} N\left(\left(\sum_{i=1}^{N} p_{ij}\right)/N \ ; \ \left(\sum_{i=1}^{N} p_{ij}(1 - p_{ij})\right)/N^2\right) \qquad [75]$$

a partir de la cual pueden establecerse predicciones con intervalos, para la elección de una nueva modalidad. La aplicabilidad de esta técnica en análisis microeconómico de medidas de política económica, es obviamente, muy amplia.

— *Problemas específicos con datos en panel*

Todos los modelos vistos en este apartado pueden aplicarse a observaciones en panel (individuos a través del tiempo). Del mismo modo que en los modelos anteriores, será preciso instrumentar un modelo que pueda dar cuenta de la heterogeneidad individual y temporal. Los modelos propuestos en la literatura, desembocan en funciones de verosimilitud muy complejas o incluso intratables (Heckman y Willis (1977), Heckman (1981)). Sin embargo, es posible introducir al menos cierta heterogeneidad individual y temporal, y tener en cuenta aspectos dinámicos, dentro del esquema visto hasta ahora. Esto además, tiene la ventaja de que el modelo propuesto aquí, puede estimarse con programas existentes en el mercado. El modelo básico propuesto aquí es el siguiente

$$\begin{aligned} p_{ijt} &= p(d_i + \gamma_j + \delta_t + x_{ijt}'B_1 + x_{ijt-1}'B_2) \\ i &= 1, ..., N \\ j &= 0, 1, ..., m \\ t &= 1, ..., T \end{aligned} \qquad [76]$$

El modelo incorpora $(N - 1)$ constantes para los efectos individuales, m para las $(m + 1)$ posibles elecciones, y $(T - 1)$ para los efectos temporales. Además, los aspectos dinámicos se tienen en cuenta, por medio de los retrasos de los regresores. La función de verosimilitud es una simple generalización de [54] y está dada por la expresión siguiente,

$$\phi = \prod_{i=1}^{N} \prod_{j=0}^{m} \prod_{t=1}^{T} p_{ijt}^{Y_{ijt}} \qquad [77]$$

donde $Y_{ijt} = 1$ si en 't' el individuo 'i' elige la opción 'j' y cero en los demás casos. Todo el análisis de este modelo puede conducirse ahora como en el caso

$T = 1$ (en realidad es la misma situación con (NT) observaciones en lugar de N).

b) Modelos de respuesta truncada

Un modelo de respuesta truncada o de «variable dependiente limitada», es un modelo en el que la variable dependiente está restringida de alguna manera. Este tipo de modelo y análisis se suele denominar «Tobit», por analogía a los modelos «probit» y «logit», ya que fue Tobin (1958) quien lo inició. El modelo se denomina «truncado» si no se observa ninguna variable fuera de ciertos límites, y «censurado» si al menos las variables independientes se observan fuera de los límites especificados. Por ejemplo, es razonable suponer que una empresa decidirá exportar solamente si alcanza cierto umbral de tamaño, que le permita soportar ciertos costes fijos asociados a la exportación. Alcanzando ese umbral, el volumen de exportación puede ser también dependiente del tamaño.

Alternativamente, podemos suponer que el umbral es función de una serie de variables que determinan que haya beneficio, o no (entre los determinantes de ese umbral, podemos suponer que están la proporción de ventas interiores sobre las totales, el grado de participación de capital extranjero, y la relación capital trabajo (Maravall, Torres, 1986). Si lo hay, se decide exportar, y tanto más cuanto mayor sea el beneficio esperado. Un tipo de razonamiento similar puede aplicarse igualmente al análisis de la inversión en Investigación y Desarrollo.

El caso más sencillo de este tipo de modelos ocurre cuando la variable dependiente es explicada por un conjunto de regresores, y sólo se observa si es positiva. Por ejemplo, el volumen de exportación de una empresa estará determinado por una serie de variables exógenas. Pero la exportación es nula, cuando el valor determinado por esos regresores es negativo (y no observado por lo tanto). Este tipo de modelos puede especificarse estadísticamente del siguiente modo,

$$\begin{aligned} Y_i^+ &= x_i'B + u_i \\ Y_i &= \begin{cases} Y_i^+ & Y_i^+ > 0 \\ 0 & Y_i^+ \leqslant 0 \end{cases} \end{aligned} \qquad [78]$$

es decir, sólo se observa la variable 'Y_i^+' si es positiva. Si es negativa, en general sólo observamos los regresores 'x_i'.

Los puntos fundamentales que hay que discutir en relación a este enfoque desde el punto de vista práctico son dos: 1) métodos de estimación; 2) formulación de modelos. A partir de [78] e introduciendo hipótesis más sofisticadas pueden tratarse situaciones muy complejas en el marco de los modelos truncados. El problema práctico fundamental será cómo estimarlos una vez formulados. Vamos a empezar la discusión por los métodos de estimación del modelo [78] para presentar a continuación generalizaciones del modelo.

— *Métodos de estimación*

El primer método de estimación que se aplicó a estos modelos fue el de mínimos cuadrados. Supongamos que aplicamos MCO en [78] para las observaciones positivas. El modelo es ahora

$$Y_i = x_i'B + u_i, \quad Y_i > 0 \tag{79}$$

Pero entonces tenemos lo siguiente,

$$E(u_i|Y_i > 0) = E(u_i|u_i > -x_i'B) \neq 0 \tag{80}$$

que es distinto de cero, si la distribución de 'u' es simétrica alrededor del origen (por ejemplo, $u - \mathcal{N}(0, \sigma)$). En otras palabras, la media de la distribución condicional es distinta de cero, aunque la media incondicional sea cero.

Supongamos ahora que añadimos las observaciones, para las que $Y_i^+ \leqslant 0$ poniendo Y_i en lugar de Y_i^+ como variable dependiente. El modelo es entonces,

$$\begin{aligned} Y_i &= x_iB + u_i & Y_i^+ &> 0 \\ 0 &= x_iB + u_i - Y_i^+, & Y_i^+ &\leqslant 0 \end{aligned} \tag{81}$$

que también puede escribirse compactamente como sigue,

$$Y = XB + u - \begin{bmatrix} 0 \\ Y^+ \end{bmatrix} \tag{82}$$

Similarmente a [80] tenemos ahora,

$$E(Y_i^+ \mid Y_i^+ < 0) \neq 0 \tag{83}$$

es decir, que esta situación, sería un caso típico de errores en la variable dependiente cuya media no es cero.

La conclusión es entonces, que tanto si aplicamos MCO a los valores observados, como si lo hacemos con todos, los estimadores serán sesgados ya que no se cumple un supuesto básico, necesario para la insesgadez de los MCO, y que es el supuesto de que la media de los errores sea cero (o al menos constante).

Una alternativa inmediata es obtener el sesgo de [80] para incorporarlo a la ecuación. Puede demostrarse por integración directa lo siguiente,

$$\begin{gathered} E(u_i | y_i > 0) = \sigma\lambda(x_i'B/\sigma) \\ u_i - \mathcal{N}(0, \sigma) \\ \lambda(\cdot) = \phi(\cdot)/\Phi(\cdot) \end{gathered} \tag{84}$$

siendo $\phi(\cdot)$ y $\Phi(\cdot)$ las funciones de densidad y distribución de una normal (0, 1), respectivamente.

Es claro entonces que podemos escribir

$$
\begin{aligned}
Y_i &= x_i'B + \sigma\lambda(x_i'B/\sigma) + \varepsilon_i \\
E(\varepsilon_i \mid x_i) &= E\varepsilon_i = 0
\end{aligned}
\qquad [85]
$$

y por lo tanto, podemos estimar los parámetros del modelo en esta expresión por algún procedimiento de mínimos cuadrados. En primer lugar, observamos en [78] que el modelo puede entenderse dentro de los modelos de respuesta cualitativa, es decir, $Y_i > 0$ o bien $Y_i < 0$. Aunque es obvio intuitivamente, que en este caso se pierde mucha información si sólo tenemos en cuenta el signo de Y_i (véase más adelante), también es claro que podemos estimar el vector $\alpha = B/\sigma$, por medio de un probit en una primera etapa. Con este estimador, podemos calcular $\hat{\lambda}_i = \lambda(x'_i\hat{\alpha})$, y estimar B por MCO en [85]. Este es el estimador propuesto por Heckman (1976) y puede demostrarse su consistencia y normalidad asintótica.

Los errores del modelo son heterocedásticos y, puede aplicarse la corrección de White para obtener estimadores consistentes de la matriz de covarianzas. Este estimador es muy útil como punto de partida para métodos más eficientes, que en general, son no lineales (vgr. el máximo verosímil).

El siguiente método de estimación a considerar, es el de máxima verosimilitud. Si las funciones de densidad y distribución de 'u' son 'f' y 'p' respectivamente, la función de máxima verosimilitud del modelo «censurado» con observaciones $i = 1, ..., N$ está dada por

$$\Psi = \prod_{-} p\,(Y_i^+ \leq 0) \prod_{+} f(Y_i) \qquad [86]$$

donde ($-$) indica que el producto se toma sobre las observaciones para las que se cumple $Y_i^+ \leq 0$, y ($+$) para $Y_i^+ > 0$. Para simplificar vamos a hacer el supuesto usual de que 'u' sigue una distribución normal. Entonces, en el caso censurado [86] puede ser escrito como

$$\Psi = \prod_{-} (1 - \Phi(x_i'B/\sigma)) \prod_{+} \sigma^{-1}\phi((Y_i - x_i'B)/\sigma) \qquad [87]$$

De forma análoga, si los datos están truncados la función de verosimilitud es simplemente,

$$\Psi = \prod_{+} \phi((Y_i - x_i'B)/\sigma)/(\sigma\Phi(x_i'B/\sigma)) \qquad [88]$$

Vamos a concentrarnos en el análisis de [87] que es el caso más frecuente. En primer lugar observamos que por medio de una operación trivial podemos escribir [87] en una forma que sugiere una interpretación interesante. Es decir,

$$
\begin{aligned}
\Psi = &\left(\prod_{-} (1 - \Phi(x_i'B/\sigma)) \prod_{+} \Phi(x_i'B/\sigma)\right) \times \\
&\times \left(\prod_{+} \phi((Y_i - X_i'B)/\sigma)/\sigma\Phi(x_i'B/\sigma)\right)
\end{aligned}
\qquad [89]
$$

Esta expresión permite descomponer la función de verosimilitud del caso censurado en dos productos. El primero es simplemente un probit binomial y el segundo, la función de máxima verosimilitud truncada. En otras palabras, se analizan primero los determinantes de la decisión cualitativa y a continuación los determinantes de la dimensión de 'Y' supuesto que se ha elegido la alternativa '+'. Esto sugiere que un método de estimación de (B/σ) es simplemente aplicar un método tipo probit, maximizando la primera parte de [87]. El estimador será consistente y asintóticamente normal, pero obviamente, no será eficiente al no tener en cuenta toda la información contenida en la segunda parte de la función máximo verosímil [87].

El estimador que maximiza conjuntamente toda la expresión [89], que es el de máxima verosimilitud, será lógicamente el más eficiente. Puede demostrarse su consistencia y normalidad asintótica (Amemiya (1973)). Tampoco caben esperar muchos problemas en la iteración pues la función (Log Ψ) es globalmente cóncava (Olsen (1978)). El único problema pendiente es entonces la selección de un punto inicial para la estimación.

Podemos tomar como valores iniciales los estimadores obtenidos por alguno de los métodos anteriores. Lógicamente, cuanto mejor sea el estimador inicial más rápida será la convergencia, y por esto, en general será preferible empezar la iteración con estimadores consistentes.

Una observación importante respecto al estimador máximo verosímil es su falta de robustez frente a cambios en determinados supuestos. En el caso lineal clásico, es bien sabido que la distribución asintótica de este estimador no se ve afectada por la no normalidad. También, los estimadores siguen siendo consistentes en presencia de heterocedasticidad. No es así, sin embargo, en el caso de los Tobits. Aunque los contrastes de estas hipótesis en este caso son complicados, es importante en el trabajo empírico, llevar a cabo siquiera una comprobación de algún tipo (por ejemplo, un gráfico de $\hat{u}_1^2$, puede ayudar a detectar heterocedasticidad). Si se detectan estos problemas, lo más sencillo es tratar de modificar la función de densidad del error en el modelo original, de forma que se suavicen.

— *Formulación del modelo*

El modelo de [78] se puede generalizar en varias direcciones, de modo que permita dar cabida a situaciones bastante más complicadas. En todos los casos, generalmente el problema se reduce a escribir la función de verosimilitud, y obtener un estimador inicial para la iteración. Las propiedades usuales de los estimadores máximo verosímiles se cumplen en general, y tampoco suelen presentarse problemas graves de convergencia. Respecto a los estimadores iniciales, es preferible que sean consistentes, y el enfoque de Heckman suele ser muy útil para obtenerlos. De todas formas, estimadores inconsistentes pueden ser válidos también aunque la convergencia sea más lenta. Vamos a analizar una posible generalización que ha gozado de cierta popularidad en trabajos empíricos (Gronau (1973), por ejemplo). Otras posibles vías de generalizar el modelo se indicarán, simplemente.

Consideremos entonces el modelo siguiente,

$$
\begin{aligned}
Y_{1i}^{+} &= x_{1i}'B_1 + u_{1i} \\
Y_{2i}^{+} &= x_{2i}'B_2 + u_{2i} \\
Y_{2i} &= \begin{cases} Y_{2i}^{+} & Y_{1i} > 0 \\ 0 & Y_{1i}^{+} \leq 0 \end{cases} \\
u_i' &= (u_{1i}, u_{2i}), \quad u - i \cdot i \cdot d \cdot \mathcal{N}(0, \Sigma)
\end{aligned} \qquad [90]
$$

donde Y_1^+ no es observada pero sí, su signo. Por ejemplo, supongamos que la exportación de una empresa es 'Y_2^+', y 'Y_1^+' es el beneficio esperado de la actividad exportadora, o cualquier otro umbral que determine la conveniencia de iniciar o no la actividad exportadora. Para dar otra interpretación clarificadora del modelo, supongamos que existe un umbral mínimo de exportaciones z_i^+ por debajo del cual la empresa 'i' no encuentra beneficioso exportar, y que además ese umbral no se observa. El modelo sería ahora

$$
\begin{aligned}
z_i^{+} &= x_{1i}'B_1 + u_{1i} \\
Y_{2i}^{+} &= x_{2i}'B_2 + u_{2i} \\
Y_{2i} &= \begin{cases} Y_{2i}^{+} & Y_{2i}^{+} > z_i^{+} \\ 0 & Y_{2i}^{+} \leq z_i^{+} \end{cases}
\end{aligned} \qquad [91]
$$

Definiendo $Y_{1i}^+ = Y_{2i}^+ - z_i^+$, el modelo puede expresarse en la forma [90].

La función de verosimilitud puede ser escrita del siguiente modo

$$
\Psi = \prod_{-} p(Y_{1i}^{+} \leq 0) \prod_{+} p(Y_{1i} > 0) f(Y_{2i} | Y_{1i}^{+} > 0) \qquad [92]
$$

Esta expresión no es operativa, pero puede ser reescrita de forma más manejable. Para ello notamos lo siguiente

$$
\begin{aligned}
f(Y_{2i} | Y_{1i} > 0) p(Y_{1i} > 0) &= \int_0^\infty f(Y_{1i} | Y_{2i}) \, dY_{1i} \\
&= \int_0^\infty f(Y_{1i} | Y_{2i}) f(Y_{2i}) \, dY_{1i}
\end{aligned} \qquad [93]
$$

Dado el supuesto de normalidad de los errores, la distribución condicional de Y_{1i}, será,

$$
Y_{1i} | Y_{2i} - \mathcal{N}(X_{1i}B_1 + (Y_{2i} - x_{2i}B_2)\sigma_{12}/\sigma_2^2; \ (\sigma_1^2 - \sigma_{12}^2 / \sigma_2^2)) \qquad [94]
$$

o más compactamente $\mathcal{N}(\mu_i, w_i)$. Podemos escribir entonces la función de verosimilitud como,

$$
\Psi = \prod_{-} (1 - \Phi(x_{1i}'B_1/\sigma_1)) \prod_{+} \Phi(\mu_i/w_i) \phi(x_{2i}'B_2/\sigma_2) \qquad [95]
$$

expresión esta última, que al estar completamente especificada puede introducirse ya en cualquier rutina de optimización. Una última observación importante relacionada con [95] es que sólo depende de σ_1 a través de (B_1/σ_1) y (σ_{12}/σ_1) de modo que sólo estos cocientes son identificables. El medio más sencillo de solucionar el problema es hacer $\sigma_1 = 1$.

Respecto a estimadores iniciales, puede derivarse un estimador consistente del tipo de Heckman. Sin embargo, una solución algo inferior, aunque sencilla y rápida, es la siguiente: 1) estimar B_1 por medio de un probit. (El estimador será consistente aunque no eficiente), 2) estimar B_2 por MCO utilizando sólo los valores observados de 'Y_2^+'. Con este estimador inicial puede intentarse ahora la optimización de [95].

Una posible ampliación del modelo ([90], [91]) consiste en añadir una tercera ecuación en la siguiente forma,

$$\begin{aligned} Y_{3i}^+ &= x_{3i}B_3 + u_{3i} \\ Y_{3i} &= \begin{cases} Y_{3i}^+ & Y_{1i}^+ > 0 \\ 0 & Y_{1i}^+ \leq 0 \end{cases} \end{aligned} \qquad [96]$$

y suponer $u' = (u_1, u_2, u_3)$, $u - \mathcal{N}(0, \Sigma)$. En el caso de las exportaciones, Y_3^+ podría representar, por ejemplo, el número de países a los que se exporta. Alternativamente, podríamos añadir al modelo base, la especificación,

$$Y_{1i} = \begin{cases} Y_{1i}^+ & Y_{1i}^+ > 0 \\ 0 & Y_{1i}^+ \leq 0 \end{cases} \qquad [97]$$

si Y_{1i} es observado para valores positivos. Volviendo al caso de las exportaciones, 'Y_1^+' podría representar ahora el volumen de exportación e 'Y_2^+' el número de países a los que se exporta. Las generalizaciones de [96] y [97] pueden también combinarse en un modelo conjunto. En principio, puede pensarse en muchas otras ampliaciones del modelo básico, pero hay que recordar siempre que los modelos excesivamente complicados no suelen dar buen resultado empírico. Desde el punto de vista aplicado es, en general, más útil la originalidad que la complejidad. De todas formas, las ampliaciones indicadas anteriormente cubren la mayor parte de las aplicaciones empíricas hasta la actualidad.

Un problema práctico importante para la aplicación de todos estos modelos es la disponibilidad de programas informáticos. Excepto para el caso más simple, no existen apenas programas disponibles lo que limita bastante su aplicabilidad. No obstante, algunos programas como el TSP (versión 4.1) incluyen una rutina que optimiza una función de verosimilitud arbitraria suministrada por el usuario. Un problema en este caso es que no se dispone de una expresión analítica exacta para la función de distribución de una normal. Una aproximación numérica aceptable para valores de $|\theta| < 3$ es la siguiente,

$$\Phi(\theta) \simeq \frac{1}{2} + \frac{1}{\sqrt{2\pi}}\left(\theta - \frac{\theta^3}{6} + \frac{\theta^5}{40} - \frac{\theta^7}{336} + \frac{\theta^9}{3.456} - \frac{\theta^{11}}{42.240}\right) \qquad [98]$$

que se obtiene expandiendo Φ (·) por Taylor alrededor del origen. (El método más rápido de calcular la expansión es utilizando resultados relacionados con los polinomios Hermitianos). Para acelerar el cálculo de la aproximación, las sucesivas potencias de 'θ' pueden calcularse recursivamente, es decir, $\theta = x'B$, $\theta^2 = \theta\theta$, $\theta^3 = \theta^2\theta$, $\theta^5 = \theta^3\theta^2$, etc. Una precaución necesaria es que no haya demasiados valores atípicos ($|\theta| > 3$), ya que entonces la calidad de la aproximación empieza a deteriorarse. Una pregunta natural que surge en relación a este problema es si es posible encontrar distribuciones, cuyas funciones de densidad y distribución posean expresiones analíticas sencillas, lo que simplificaría mucho la iteración. Aunque las hay (por ejemplo, la logística) el problema es que necesitamos obtener distribuciones condicionales y marginales para escribir la función de verosimilitud, y la distribución normal es muy sencilla de manejar para obtener estos resultados.

— *Modelos de Poisson*

Dentro de esta sección y antes de finalizarla, conviene discutir brevemente los modelos de Poisson. Supongamos que la variable discreta 'Y' puede tomar sólo valores positivos $Y = (0, 1, 2, ...)$. La distribución de Poisson cumple estas características para valores de λ positivos, y su función de densidad o cuantía está dada por

$$p(Y = r) = e^{-\lambda}\lambda^r/r!, \quad r = 0, 1, 2, ... \quad ; \quad \lambda > 0 \qquad [99]$$

Esta distribución puede proporcionar un tratamiento alternativo a los modelos truncados, cuando la variable dependiente es discreta, ya que puede tomar el valor cero, pero no valores negativos. Por otra parte, cuando la variable dependiente es discreta, en principio no es adecuado utilizar un modelo truncado del tipo [78] con un error distribuido de acuerdo a una función de densidad continua (por ejemplo, Fair (1978)). Una aplicación típica de estos modelos ha sido a la investigación de la relación entre el número de patentes e inversión en I y D (Hausman *et. al.* (1981)).

Para establecer el modelo, supongamos que disponemos de una serie de observaciones independientes 'Y_i', $i = 1 ... N$, que siguen la distribución de Poisson. Al estudiar los determinantes de 'Y_i', y como se debe asegurar que $\lambda_i > 0$ suponemos una relación exponencial, es decir,

$$\lambda_i = \exp (x_i'B) \qquad [100]$$

La función de verosimilitud será como de costumbre,

$$\Psi = \prod_{i=1}^{N} p_i \qquad [101]$$

estando p_i definido por [99] y [100]. Un estimador inicial para el algoritmo es el sugerido por Lancaster (1974), cuya derivación puede explicarse más o menos intuitivamente como sigue. En primer lugar y por las características de una Poisson podemos escribir,

$$Y_i = \lambda_i + u_i \quad ; \quad Eu_i = 0 \quad ; \quad V(u_i) = \lambda_i \qquad [102]$$

Entonces, tomando logaritmos y expandiendo por Taylor,

$$LY_i = L\lambda_i + \frac{1}{\lambda_i + \delta u_i} u_i \quad ; \quad |\delta| < 1 \qquad [103]$$

Definamos ahora,

$$w_i = u_i/(\lambda_i + \delta u_i) = \varepsilon_i/(\lambda_i^{1/2} + \delta\varepsilon_i) \qquad [104]$$

donde

$$\varepsilon_i = u_i/\lambda_i, \quad V(\varepsilon_i) = 1 \qquad [105]$$

Si la aproximación es buena de modo que $\delta \simeq 0$ tenemos que $Ew_i = 0$, $V(w_i) = (1/\lambda_i)$, de modo que podemos escribir [103] como,

$$LY_i \simeq x_i'B + w_i, \quad Ew_i = 0, \quad V(w_i) = 1/\lambda_i \qquad [106]$$

y obtener un estimador de B por un procedimiento de mínimos cuadrados generalizados. Como $EY_i = \lambda_i$, un estimador lógico en un primer paso de λ_i es Y_i, así que el estimador MCG de B en [106] es

$$B = (X'YX)^{-1}XYq \qquad [107]$$

donde $Y = \text{diag}\ (Y_i)$, y, $q_i = LY_i$ (si $Y_i = 0$ se puede poner $Y_i = 0{,}5$ (Maddala (1983)).

Finalmente, hay que notar que una desventaja de la distribución de Poisson es que la esperanza es igual a la varianza, lo que en principio parece algo restrictivo.

3. Formulación y especificación

Esta sección va a estar dedicada al análisis de algunos problemas prácticos que se presentan al utilizar los modelos presentados en la sección anterior, los puntos discutidos se van a agrupar y presentar por este orden, en tres grupos: 1) formulación de modelos; 2) contrastes de validación; 3) problemas específicos. La clasificación no pretende ser exhaustiva, y los problemas comunes con modelos de series temporales se omiten, ya que están suficientemente analizados en otro tipo de literatura.

3.1. *Formulación de modelos*

a) Variables medidas en cocientes

Respecto a la formulación, un problema que se suele plantear con frecuencia es el análisis de modelos con la variable dependiente medida como un cociente. Este tipo de transformación, es útil para evitar los problemas de heterocedasticidad y correlaciones espurias. Por ejemplo, consideremos el modelo,

$$Y_i = x_i B + u_i, \quad V(u_i) = kx_i^2 \qquad [108]$$

Para estimar este modelo, el procedimiento óptimo es corregir la heterocedasticidad, lo que en este modelo puede hacerse reescribiendo [108] en la forma siguiente,

$$\left(\frac{Y_i}{x_i}\right) = B + \varepsilon_i, \quad \varepsilon_i = (u_i/x_i) \qquad V(\varepsilon_i) = k \qquad [109]$$

Con datos de tipo transversal («cross-section») también se plantea el problema de correlaciones espurias similar al de la correlación entre variables tendenciales, que suele aparecer en series temporales. Medir las variables en cocientes es un modo de suavizar el problema. También, la heterocedasticidad puede ser simplemente reflejo del hecho de que la varianza de una variable estocástica tendencial tiene que crecer, pues de otro modo la variable se haría determinística con el paso del tiempo (es decir, si cuando $T \to \infty$, $m \to \infty$, entonces $\sigma/m \to 0$, lo que es generalmente considerado como una propiedad no deseable). En resumen, medir las variables dependientes como cocientes es una solución práctica muy extendida a los problemas interrelacionados de correlaciones espurias y heterocedasticidad.

Esta solución no obstante, presenta algunos problemas. Para analizarlos, consideremos la siguiente especificación,

$$C_i = \bar{C}_i + u_i, \quad \bar{C}_i = EC_i, \quad Eu_i = 0$$
$$0 < C_i < 1$$
$$\bar{C}_i = X_i' B \qquad [110]$$

La restricción en la variable dependiente, no tiene porqué presentarse, pero muy frecuentemente se dará (por ejemplo, proporción de exportaciones sobre ventas totales). En principio, podemos aplicar ahora MCO a este modelo, y serán insesgados. Aunque los estimadores no serán eficientes, ya que 'u' no puede tener una distribución normal al estar acotado, serán asintóticamente normales. Sin embargo, como en el caso del modelo binomial, no existe, en general, garantía de que el valor estimado de c_i esté entre cero y uno, es decir, que $0 < x_i'\hat{B} < 1$. Esta última propiedad es deseable, y una forma de garantizarla similarmente al caso de los probits y logits sería hacer

$$\bar{C}_i = F(x_i'B) \tag{111}$$

y escoger una función $F(\cdot)$ que garantizase la restricción en $\bar{C}_i$ (por ejemplo, una función de distribución). Sin embargo, esta especificación es no lineal, lo que siempre complica el análisis. Si $F(\cdot)$ es la función logística, es fácil comprobar que

$$\log\left(\bar{C}_i/(1 - \bar{C}_i)\right) = x_i'B \tag{112}$$

así que una alternativa obvia es utilizar el modelo,

$$\log\left(C_i/(1 - C_i)\right) = x_i'B + \varepsilon_i, \quad \varepsilon_i - \mathcal{N}(0, \sigma) \tag{113}$$

pues ahora la variable dependiente puede tomar valores en el intervalo $(-\infty, +\infty)$. Si el intervalo de la variable es $0 < C_i < k$, basta con considerar $0 < (C_i/k) < 1$ y si es $(0, +\infty)$, la transformación es simplemente $\log(C_i)$.

En cualquier caso, la solución de medir las variables en cocientes no tiene por qué ser la óptima. Una alternativa es simplemente estimar el modelo suponiendo errores heterocedásticos. Por ejemplo, si 'x_i' es ahora un vector en el modelo [108] y $V(u_i) = k(x_i'B)^2$ se puede estimar el modelo en dos etapas: 1) se estima por MCO, para obtener estimadores consistentes de B; 2) se corrige la heterocedasticidad y se estima por MCG (véase, por ejemplo, Theil (1971)).

b) Contrastes de asociación no paramétricos

Un tipo de análisis muy útil en este contexto, es el de contrastes no paramétricos de asociación. El término «no paramétrico» hace referencia simplemente al hecho de que se contrasta la asociación entre las variables sin suponer ningún tipo de modelo para los datos. La utilidad de este análisis en este contexto es debida, en parte, a que en muchas ecuaciones los datos en panel son variables cualitativas, es decir clasificadas por categorías. El análisis no paramétrico, en sí mismo, y como complemento de otros análisis está entonces especialmente indicado. La ventaja de estos contrastes es su robustez. Aparentemente, el precio es una pérdida de potencia como en los modelos generales frente a los restringidos, pero probablemente no es así, ya que la pregunta a la que responden estas técnicas es más imprecisa. En otras palabras, la robustez o validez del método, la obtenemos al contrastar simplemente la existencia de relación entre 'x' e 'Y' en lugar de su elasticidad, por ejemplo. El modelo es más general pero la pregunta es más imprecisa.

A continuación se presentan una serie de problemas que este tipo de método puede analizar. Supongamos que se disponen de observaciones clasificadas de acuerdo a dos criterios en las categorías de forma que 'n_{ij}' es el número de elementos en la casilla (i, j), $i = 1, ..., I; j = 1, ..., J$. Utilizaremos la notación siguiente:

$$n_{i.} = \sum_{j} n_{ij}, \quad n_{..} = \sum_{i,\, j} n_{ij} = n$$

y similarmente para casos con más subíndices y otros símbolos. El contraste básico de independencia entre las dos variables, puede llevarse a cabo por medio de la expresión,

$$n\left(\sum_{i,\, j=1}^{I,\, J} (\hat{p}_{ij} - \tilde{p}_{ij})/\tilde{p}_{ij}\right) \tilde{A} X^2((I - 1)(J - 1)) \qquad [114]$$

donde

$$\begin{aligned} \hat{p}_{ij} &= n_{ij}/n \\ \tilde{p}_{ij} &= \tilde{p}_{i.}\tilde{p}_{.j} \\ \tilde{p}_{i.} &= n_{i.}/n, \quad \tilde{p}_{.j} = n_{.j}/n \end{aligned} \qquad [115]$$

(la derivación puede encontrarse en cualquier libro de estadítica. Por ejemplo, Mood *et. al.* (1974). El símbolo $\tilde{A}$ significa asintóticamente, cuando $n \to \infty$).

Otros contrastes menos comúnmente utilizados aunque también muy útiles son los de asociación parcial y total en grupos de variables. Por ejemplo, supongamos que se desea contrastar la asociación entre una variable 'i' y otras dos 'j', 'k'. (Esto sería un caso equivalente al contraste de R^2 en la regresión lineal). La hipótesis puede contrastarse por medio de la expresión,

$$n\left(\sum_{i,\, j,\, k=1}^{I,\, J,\, K} (\hat{p}_{ijn} - \tilde{p}_{ijn})/\tilde{p}_{ijn}\right) \tilde{A} X^2(g \cdot l) \qquad [116]$$

donde

$$\begin{aligned} g.l &= (IJK - 1) - ((I - 1) + (JK - 1)) \\ \hat{p}_{ijk} &= n_{ijk}/n \\ \tilde{p}_{ijk} &= \tilde{p}_{i..}\tilde{p}_{.jk} \\ \tilde{p}_{i..} &= n_{i..}/n \\ \tilde{p}_{.jk} &= n_{.jk}/n \end{aligned} \qquad [117]$$

En el mismo ejemplo de las tres variables, supongamos que se desea contrastar la hipótesis de asociación entre las variables 'i' y 'j' condicionadas a 'K'. El contraste en este caso puede llevarse a cabo con la misma expresión de [116] donde ahora los símbolos se definen de la siguiente forma,

$$
\begin{aligned}
g.l &= (IJK - 1) - \big((IK - 1) + (JK - 1)\big) \\
\hat{p}_{ijn} &= n_{ijk}/n \\
\hat{p}_{ijk} &= \hat{p}_{i.k}\hat{p}_{.jk}/\hat{p}_{..k} \\
\hat{p}_{i.k} &= n_{i.k}/n \\
\hat{p}_{.jn} &= n_{.jk}/n \\
\hat{p}_{..k} &= \sum_i \hat{p}_{i.k} = \sum_j \hat{p}_{.jk}
\end{aligned}
\qquad [118]
$$

Los contrastes [117], [118] son de máxima verosimilitud, y su derivación es relativamente directa. Pueden elaborarse, también, medidas de asociación total y parcial, aunque su interpretación es algo discutible. De todas formas, llegados a este punto es más útil especificar modelos y estimar la asociación concreta de las variables en ellos.

Un tercer punto importante en este contexto es el de la reducción del número de variables a analizar. En general, suele haber una cantidad considerable de variables disponibles potencialmente explicativas. El método de probar su capacidad explicativa una a una puede ser costoso, y además implica problemas estadísticos. Los métodos de reducción de datos pueden ser útiles entonces, en la fase exploratoria del trabajo (véase, por ejemplo, Fernández Alvarez (1986)). Estos métodos son fundamentalmente, ciertas técnicas de análisis factorial y el análisis de componentes principales, con y sin rotaciones. En alguna medida, el análisis de correlaciones canónicas también puede servir para este propósito (veáse, por ejemplo, Anderson (1958)).

3.2. *Contrastes de validación*

El segundo punto mencionado al comienzo de esta sección hace referencia a los contrastes de validación. En general, puede decirse que es preciso someter los modelos estimados a una batería amplia de contrastes de especificación. Esta es la única forma de suavizar los aspectos subjetivos, inevitables en todo proceso de especificación práctica de modelos econométricos. Una batería bastante amplia de contrastes la constituye el siguiente conjunto:

a) estabilidad,
b) no omisión de variables significativas,
c) heterocedasticidad,
d) normalidad,
e) datos atípicos,
f) simultaneidad,
g) linealidad.

En Pagan y Hall (1983) puede encontrarse una presentación unificada de la mayoría de estos contrastes, como análisis de residuos. A continuación, se presenta brevemente la motivación de los contrastes y cómo llevarlos a cabo.

El contraste de estabilidad puede llevarse a cabo con facilidad por medio de la estimación máximo verosímil del modelo en dos submuestras. Dado que en general, el tamaño de la muestra no suele ser un problema con datos en panel, es conveniente reservar una parte para contrastar el modelo, y especificarlo con el resto de los datos. El contraste será entonces,

$$-2(L - (L_1 + L_2)) \underset{A}{\sim} X^2(k) \qquad [119]$$

donde 'k' son los parámetros del modelo, L, L_1 y L_2 las funciones de verosimilitud evaluadas con el estimador obtenido en la muestra conjunta y en las dos submuestras, y el símbolo $\underset{A}{\sim}$ significa asintóticamente (en este caso debe entenderse como $(N_1, N_2) \to \infty$). Aparte del objetivo obvio de contrastar la homogeneidad del modelo en las dos submuestras, el contraste de estabilidad es una salvaguardia frente a los problemas del agotamiento de los datos («data mining») y el «pretesting». Para ejemplificar el problema del agotamiento de los datos, consideremos los 'K' coeficientes 't' obtenidos al regresar una variable 'Y' en un vector 'x', y supongamos que todas las variables son independientes (por ejemplo, porque se ha diseñado así un experimento). Para simplificar, vamos a suponer que la distribución de las 't's' es independiente (que es aproximadamente cierto para T alto). Como de costumbre, la región crítica será $|t_i| > 2$, de modo que si un 't_i' pertenece a ella decidiremos que la variable correspondiente es significativa. Para T alto sabemos que

$$p[|t_i| > 2] \simeq 0{,}05 \qquad [120]$$

Definiendo la región crítica por 'C' consideremos ahora la siguiente probabilidad

$$\begin{aligned} p[\text{al menos un } t_i \in C] &= 1 - p[\text{ningún } t_i \in C] \\ &\simeq 1 - \prod_i^k p[t_i \in C] \\ &= 1 - (0{,}95)^K \qquad [121] \end{aligned}$$

Si el número de regresores es 10, esto quiere decir que la probabilidad de que encontremos al menos un regresor significativo es aproximadamente del 40 por 100 y del 80 por 100 si $K = 30$. En otras palabras, este análisis responde a la intuición de que las correlaciones obtenidas después de numerosas pruebas, pueden deberse a la casualidad.

Para analizar el problema del «pretesting», consideremos el proceso de decisión usual, según el cual, el estimador que adoptamos para el parámetro 'B' en una regresión está dado por la siguiente regla,

$$\tilde{B} = \begin{cases} \hat{B} & \text{si } |t| > 2 \\ 0 & \text{si } |t| < 2 \end{cases} \qquad [122]$$

donde $\hat{B}$ es el estimador MCO y 't' su coeficiente asociado de student. Es un hecho conocido que $E\hat{B} = B$. Consideremos ahora la esperanza de $\tilde{B}$. Tenemos entonces lo siguiente,

$$E\tilde{B} = E(\hat{B}||t| > 2)p(|t| > 2) \qquad [123]$$

que en general será distinto de B. Es decir, que este estimador es sesgado. Tampoco la varianza de $\tilde{B}$ es la de $\hat{B}$, ni $\tilde{B}$ está distribuido normalmente. Si tenemos en cuenta que en un proceso de especificación, decisiones de este tipo se superponen, la distribución final de los estadísticos 't' puede ser muy diferente de una $\mathcal{N}(0, 1)$.

Un método de paliar estos dos problemas, consiste en reestimar el modelo con datos nuevos, y comprobar su homogeneidad. Por ello, son especialmente importantes los contrastes de estabilidad. Conviene recordar también que un método de aliviar el carácter espúrio de una correlación es explicarla, o interpretarla, y en este punto, la teoría económica puede jugar, en principio, un papel estadístico importante.

Una vez que se ha llegado a una especificación tentativa del modelo, suele ser conveniente compararlo con un modelo ampliado en el que entran todas las variables potencialmente significativas. Este procedimiento es útil para detectar variables significativas omitidas en determinados pasos de la especificación debido a sesgos, etc.

La heterocedasticidad es un problema especialmente importante con datos en panel, como ya se ha comentado. Una hipótesis muy razonable es suponer que la media es proporcional a la varianza, es decir,

$$Y_i = x_i B + \varepsilon_i, \quad V(\varepsilon_i) = k(EY_i)^2 \qquad [124]$$

Una forma sencilla de contrastar esta hipótesis consiste en los dos pasos siguientes. En el primer paso se estima [124] por MCO y se obtiene $\hat{\varepsilon}_i$. El contraste se lleva ahora a cabo en un segundo paso por medio de la expresión

$$\mathcal{N}R^2(\hat{\varepsilon}_1^2 | \alpha_o + x_i\alpha_1) - \chi^2(1) \qquad [125]$$

es decir, comparando el R^2 obtenido al regresar los residuos al cuadrado de la primera regresión, en las X's, con una χ^2 de un grado de libertad.

La normalidad de los residuos puede contrastarse con la expresión

$$\mathcal{N}\left(\frac{\gamma_1^2}{6\sigma^6} + \frac{\gamma_2^2}{24\sigma^8}\right) A\chi^2(2)$$

donde

$$\gamma_1 = \Sigma\, \hat{\varepsilon}_i^3 / \mathcal{N}$$

$$\gamma_2 = (\Sigma\, \hat{\varepsilon}_i^4 / \mathcal{N}) - 3(\Sigma\, \hat{\varepsilon}_i^2 / \mathcal{N}) \qquad [126]$$

y $\hat{\varepsilon}_i$ son los residuos de la regresión.

Para interpretar este contraste adecuadamente, es preciso asegurar que no existen datos atípicos, ya que es especialmente sensible a ellos, al depender de potencias elevadas de los residuos.

El método usual para detectar datos atípicos suele consistir en establecer una banda ($\pm 2\sigma$) o bien ($\pm 3\sigma$), fuera de la cual, un residuo se considera atípico. Sobre esta práctica conviene hacer una puntualización. Por medio de un análisis similar al de [121] obtenemos lo siguiente (suponiendo normalidad)

$$p[\text{al menos un } |\hat{\varepsilon}_i| > 3\sigma] \simeq 1 - 0{,}99^N \qquad [127]$$

y si N es 100, valor bajo para datos en panel, esta probabilidad es aproximadamente el 64 por 100. Es decir, si se observa un dato fuertemente alejado de la muestra, y que además se sabe que coincide con un acontecimiento anormal, es conveniente intervenirlo de algún modo si esto afecta a los resultados. En otro caso, lo más razonable es no manipular los datos ya que corremos el peligro de acabar trabajando con una muestra de observaciones «inventada».

La simultaneidad en datos transversales, no requiere ningún análisis especial. Es suficiente con sustituir el índice 'i' por el 't', en los modelos simultáneos usuales, y todos los resultados son válidos. Para contrastar la simultaneidad de una variable se puede aplicar el contraste de Hausman, que es un contraste de máxima verosimilitud. Supongamos que tenemos un modelo de dos variables dado por,

$$\begin{aligned} Y_{1i} &= aY_{2i} + z_{1i}'b + \varepsilon_{1i} \\ Y_{2i} &= \alpha Y_{1i} + z_{2i}'B + \varepsilon_{2i} \end{aligned} \qquad [128]$$

y en el que hay restricciones de modo que el modelo está identificado. La forma reducida para Y_{2i}, que se obtiene al resolver [128] será,

$$Y_{2i} = z_i'p_2 + V_{2i} \qquad [129]$$

donde 'z_i' incluye todas las variables no repetidas de z_1 y z_2. El contraste de Hausman se lleva a cabo ahora en dos pasos. Primero se estima [129] por MCO y se obtiene un estimador de los errrores $\hat{V}_{2i}$. En segundo lugar, se contrasta la significatividad de 'θ' en la siguiente ecuación, estimada por MCO,

$$Y_{1i} = aY_{2i} + z'_{1i}b + \theta\hat{V}_{2i} + w_i \qquad [130]$$

Si se acepta la hipótesis nula $\theta = 0$, se concluye que no hay simultaneidad y se puede estimar la ecuación por MCO (si 'a' y 'α' son distintos de cero, siempre habrá simultaneidad aunque sea poco relevante).

Finalmente, con datos en panel, puede ser importante contrastar la no linealidad. Esto puede hacerse en principio con facilidad, añadiendo regresores del tipo x^2, x^3, ... La no linealidad no suele darse tanto en series temporales, pues al superponerse con los aspectos dinámicos, éstos dan cuenta de ella. En un modelo con datos transversales, precisamente un DW lejano a 2, puede ser un signo de no linealidad (por ejemplo, un $DW \simeq 0$, sería indicio de correlación positiva. Si los residuos son una función cuadrática del índice 'i', esto es lo que ocurre). La utilidad del estadístico DW en este contexto se reduce simplemente a que casi todos los programas econométricos lo suelen incluir rutinariamente en todas las salidas.

Los contrastes de heterocedasticidad, normalidad, y simultaneidad en modelos de variables cualitativas y truncadas, son mucho más complicados, como ya se ha comentado en la sección anterior. En cualquier caso, puede ser siempre útil utilizar por lo menos técnicas descriptivas (por ejemplo, hacer un gráfico conjunto de las variables $\hat{\varepsilon}_i^2$, $\hat{E}Y_i$).

3.3. *Problemas específicos de economía industrial*

a) Efecto de variables no observadas

El tercer punto mencionado al comienzo de esta sección hace referencia a problemas específicos de estudios empíricos de economía industrial, que quizás pueden ser generalizables a otros estudios basados en datos en panel. Un problema bastante común es la selección y uso de indicadores sintéticos de varias variables (este problema aparece, por ejemplo, en los trabajos de Bergés (1986) y Fariñas y Romero (1986a)). El uso de indicadores puede plantearse econométricamente como un problema de variables no observadas. Vamos a analizar en primer lugar el efecto de las variables no observadas para pasar en segundo lugar a discutir criterios de elección. Consideremos entonces el modelo definido por las siguientes ecuaciones,

$$\begin{aligned} Y^+ &= x^+a + \varepsilon \\ Y &= Y^+ + u \\ x &= x^+ + V \end{aligned} \qquad [131]$$

donde x^+ e Y^+ no son observables, pero sí lo son x e Y. Supongamos también que los errores tienen media cero y son ruido blanco. Además suponemos

$$E\varepsilon u = E\varepsilon v = Euv = EY^+u = Ex^+v = Ex^+\varepsilon = 0 \qquad [132]$$

El parámetro de interés es 'a' y para estimarlo, escribimos la primera ecuación en función de las variables observables. Obtenemos entonces,

$$\begin{aligned} Y &= xa + (\varepsilon + u - aV) \\ &= xa + W, \quad EXw = -a\sigma_V^2 \end{aligned} \qquad [133]$$

Si no disponemos de variables instrumentales, el único método posible en la práctica es aplicar mínimos cuadrados ordinarios. No es difícil demostrar que para un número elevado de observaciones se cumple

$$\begin{aligned} \hat{a} &\simeq a(1 - \sigma_V^2/\sigma_x^2) \\ R^2 &\simeq \hat{a}^2\sigma_x^2/\sigma_y^2 \end{aligned} \qquad [134]$$

Como $\sigma_x^2 = \sigma^2(V) + \sigma^2(x^+)$, se cumple que $|\hat{a}| < a$, es decir, que el uso de indicadores sesga los estimadores hacia cero (aunque no su signo) y también las medidas de ajuste. Todo ello hace que sea más dificultoso detectar la relación buscada. Consideremos ahora el problema de la elección de indicadores y supongamos que tenemos el siguiente modelo

$$\begin{aligned} Y &= ax + \varepsilon \\ Z &= bx + V \\ W &= cx + u \end{aligned} \qquad [135]$$

donde el parámetro de interés es 'a', pero la variable 'x' no es observada y en su lugar disponemos de dos indicadores observados 'z' y 'w'. Suponemos también que se cumplen las condiciones usuales de ortogonalidad entre regresores y errores, y de independencia entre estos últimos.

No es difícil demostrar que al sustituir 'x' por 'Z' en la primera ecuación de [135], el sesgo obtenido será

$$\text{plim } \hat{a} - a = a(1 - \sigma_v^2/\sigma_z^2)/b - a$$

y el obtenido al utilizar 'w' se obtiene análogamente. Como los parámetros de esta expresión no son identificables, no es de mucha utilidad. En la práctica, la alternativa es recurrir a técnicas multivariantes de reducción de dimensionalidad, más concretamente, el análisis factorial, el de componentes principales, y el análisis de correlaciones canónicas.

Otro problema bastante común es el de los ajustes excesivamente bajos. Aunque en principio esto no tiene porqué ser anormal, implica que las relaciones económicas entre las variables son débiles. Esta conclusión parece difícil de creer, y la explicación que surge naturalmente es que hay variables omitidas. Para ilustrar este hecho recordamos que en el modelo

$$Y = x_1B_1 + x_2B_2 + \varepsilon, \quad x_1'x_2 = 0 \qquad [136]$$

el R^2 puede descomponerse en la forma

$$R^2 = R_1^2 + R_2^2 \qquad [137]$$

siendo R_1 y R_2 los coeficientes de correlación total entre (Y, x_1), (Y, x_2) respectivamente. Consideremos ahora el modelo dinámico,

$$Y = aY_{-1} + xB + \varepsilon, \quad Y'_{-1}x = 0 \qquad [138]$$

si $a = 0{,}7$, la contribución de Y_{-1} al R^2 es $a^2 = 0{,}49$. Si $R^2 = 0{,}7$, resulta que la contribución de Y_{-1} al R^2 es aproximadamente el 70 por 100. Este ejemplo ilustra la contribución relativa de los aspectos dinámicos a los ajustes de ecuaciones. Como de hecho, las relaciones económicas son dinámicas, omitir este aspecto puede ser una explicación de los pobres ajustes obtenidos en muchos estudios empíricos (aparte de los sesgos que esta omisión implica).

b) Criterios de elección

Finalmente, un problema algo más sutil es el de la selección de los datos con criterios preespecificados (muestreo no aleatorio). Por ejemplo, se plantea cuando se seleccionan empresas de acuerdo a su tamaño, a que coticen o no en Bolsa, etc. Consideremos el siguiente modelo

$$Z_i = m + \varepsilon_i, \quad E\varepsilon_i = 0, \quad E(\varepsilon_i\varepsilon_j) = 0 \quad i \neq j, \quad E\varepsilon_i^2 = \sigma^2 \qquad [139]$$

y supongamos que seleccionamos las observaciones de forma que Z_i pertenece a la muestra si $Z_i > Z^+$. Es evidente entonces que

$$E(\varepsilon | Z > z^+) \neq 0 \qquad [140]$$

y así, la estimación usual de la media 'm', estará sesgada. Lo mismo ocurre en el modelo de regresión, en el que $m = x_i'B$, y por lo tanto, los estimadores MCO serán sesgados. Si los valores $Z_i < z^+$ no se observan, el modelo cae dentro de la categoría de los modelos «censurados» discutido en la sección 2. De todas formas, es un problema que aunque sea de difícil solución en muchos casos, es importante ser consciente de que existe.

4. Programas

Es un hecho conocido que la evolución del software en general, incluido el econométrico, en los últimos años es muy rápida. Esta evolución rápida propiciada por la evolución técnica hace obsoleto rápidamente cualquier análisis detallado de los programas existentes. No obstante, si pueden darse algunas ideas generales sobre las disponibilidades econométricas para el usuario no especialista.

Existe una variedad considerable de programas que poseen varias, o la mayoría de las modalidades de estimación sugeridas en la sección 1. Casi todos los programas están disponibles en versión iterativa, y para ordenadores personales

(PC). El precio varía considerablemente de unos a otros, pero las versiones para PC son relativamente baratas, ya que en general se venden de una vez por todas (no hay contratos temporales de alquiler debido a la facilidad de sortear los sistemas de seguridad).

Los programas que incluyen alguna o varias de las opciones para datos de «cross-section» son los clásicos de análisis multivariante a los que se ha añadido rutinas econométricas, y los programas econométricos tradicionales. Entre los primeros están los siguientes: BMDP, SAS, SPSS; y entre los segundos el TSP y el SORITEC. Todos estos programas disponen de opciones para tratar los problemas discutidos en este trabajo.

Desde el punto de vista de los modelos de la sección 2, un programa puede ser evaluado atendiendo a ciertas características que se comentan brevemente a continuación. En primer lugar, para estimar los modelos de efectos fijos (EF) y efectos aleatorios (EA) con datos de panel, es suficiente con un programa de regresión ordinaria. También es conveniente que los datos puedan ser manipulados con facilidad de modo que se puedan calcular medias por subgrupos, etc. Si el número de datos a analizar es alto, es conveniente también que el programa disponga de un método de lectura eficiente (en general, con formato fijo, no libre). En lo que se refiere a modelos de respuesta cualitativa, es suficiente con que un programa ofrezca una de las dos posibilidades, probit o logit, ya que en la práctica suelen dar resultados similares. Para los modelos multinomiales, basta con disponer de la opción logística. La versión probit multinomial es muy costosa, y los modelos logísticos condicionales, que son de uso complicado, pueden ser sustituidos por modelos binomiales o multinomiales secuenciales (véase sección 2).

Respecto a los modelos truncados, con el básico, pueden tratarse una gran variedad de casos prácticos. Es raro que los programas ofrezcan versiones más complicadas pues, en general, tienen que ser diseñados a la medida del problema que se quiere tratar. Tampoco suele ofrecer ningún programa modelos para variables dependientes discretas (Poisson u otras distribuciones). Lo que sí es útil en estos casos, es disponer de una rutina que optimice una función arbitraria suministrada por el usuario. Para que esta opción sea útil, el programa debe ser lo suficientemente flexible como para permitir una definición sencilla de la función máximo verosímil. Finalmente, cuando se dispone de un elevadísimo número de observaciones, es muy útil disponer de una opción que permita seleccionar observaciones para la estimación (por ejemplo, en la fase de especificación, esta opción puede hacer menos gravosa económicamente la estimación del modelo, sin perjuicio de que en la fase final se estime con toda la muestra).

Un programa que incluye muchas de estas opciones es el TSP en su versión 4.1. En la versión para PC, no todas las opciones están disponibles en la actualidad, pero lo natural es que en un futuro próximo lo estén. Una ventaja adicional de este programa es su precio: para dar una idea, el coste de la versión 4.1 para una institución que posea la versión anterior es de 1.000 dólares (el programa se vende de una vez por todas, no se alquila). La versión para micros, tiene un precio de 300 dólares y también se vende definitivamente. El SORITEC es un programa bastante parecido en opciones y sólo es ligeramente más caro. De todas formas hay algunas diferencias entre ambos. El resto de los programas mencionados, SAS, BMDP, SPSS, son fundamentalmente programas de análisis multivariante, aunque también incluyen bastantes opciones econométricas, especialmente el SAS. El BMDP está pensado fundamentalmente para análisis de datos médicos y biológicos. El SPSS para análisis sociológicos y el SAS es un programa muy amplio que se vende por módulos que incluyen, un bloque econométrico, otro para gráficos, etc. Este último programa es especialmente versátil para el manejo básico de datos (transformaciones, agrupaciones, etc.), aunque es uno de los más caros. Una de sus ventajas es que dispone de opciones para tratar eficientemente, modelos con un número elevado de datos, como son, selección de observaciones, cálculo de estimadores a partir de estadísticos suficientes, etc.

Aunque suele ser fácil comunicarse con los distribuidores, se ha pensado que puede ser útil finalizar esta sección, recogiendo las direcciones comerciales de los programas mencionados. Las direcciones son las siguientes:

TSP

TSP international
928 Mears Court
Stanford, California 94305
USA

SORITEC

North-holland
(a division of Elsevier Science Publishers)
Attn:E. van Koten
P.O. Box 1991
1000 BZ Amsterdam
The Netherlands

BMDP

BMDP Program Librarian
Health Sciences Computing Facility
AV-111, CHS
University of California
Los Angeles, CA. 90024
ISA

SAS

SAS software limited
The Centre, 68 High Street
Weybridge, Surrey KT13 8BL
United Kingdom

SPSS

SPSS
Avelingen West
P.D. Box 115
4200 AC Gorinchem
The Netherlands

Apéndice 1. Obtención de las transformaciones para la aplicación de los modelos de efectos fijos y aleatorios

Consideremos el modelo siguiente,

$$Y_{it} = z_{it}'\gamma + x_{it}'B + \varepsilon_{it} \quad \begin{array}{l} i = 1, ..., N \\ t = 1, ..., T \end{array} \qquad [A1]$$

donde 'ε_{it}' son errores no observados, 'z_{it}' representa un conjunto de variables artificiales temporales e individuales ($N - 1$ individuales, $T - 1$ temporales, y una constante general), y 'x_{it}' agrupa el resto de las variables que determinan el comportamiento de la variable de interés 'y_{it}'.

Definamos,

$$\begin{aligned} Y_i' &= (Y_{i1}, Y_{i2}, ..., Y_{it}) \\ Y' &= (Y_1', Y_2', ..., Y_N') \end{aligned} \qquad [A2]$$

y similarmente 'z', Y', X'. Utilizando los resultados de la regresión particionada, obtenemos para la estimación mínimo cuadrática del vector β en [A1], la siguiente expresión,

$$\beta = (X'M_zX)^{-1}X'M_zY \qquad [A3]$$

donde

$$\begin{aligned} M_z &= I - Z(Z'Z)^{-1}Z' \\ M_zM_z &= M_z \end{aligned} \qquad [A4]$$

de modo que el resultado de [A3] puede obtenerse también regresando las variables (X, Y) independientemente en Z en un primer paso, y en un segundo paso regresando los residuos de Y en los de X. Este resultado es particularmente útil en el presente contexto, pues permite simplificar considerablemente la estimación de β que es el vector de parámetros de interés. Más concretamente, puede demostrarse lo siguiente,

$$M_z = I_{NT} - A - B + J_{NT} \qquad [A5]$$

donde

$$A = (I_N \otimes \mathcal{J}_T)\frac{1}{T}$$

$$B = (I_N \otimes \mathcal{J}_T)\frac{1}{N}$$

$$C = (I_N \otimes \mathcal{J}_T)\frac{1}{NT} = I_{NT}\left(\frac{1}{NT}\right) \qquad [A6]$$

y donde $\mathcal{J}_N$ es una matriz cuadrada de unos y de dimensión N, y I_{NT} es la matriz unidad de orden NT.

Es sencillo comprobar las siguientes relaciones,

$$A^2 = A$$
$$B^2 = B$$
$$C^2 = C$$
$$AB = C$$
$$AC = C$$
$$BC = C \qquad [A7]$$

de modo que efectivamente obtenemos $M_z^2 = M_z$. Dadas las definiciones de [A6] obtenemos los siguientes datos transformados,

$$Y^+ = M_z Y$$
$$Y_{it}^+ = Y_{it} - Y_{i.} - Y_{.t} + Y_{..} \qquad [A8]$$

donde

$$Y_{i.} = \sum_{t=1}^{T} Y_{it}/T$$

$$Y_{.t} = \sum_{i=1}^{N} Y_{it}/N$$

$$Y_{..} = \sum_{i,\ t} Y_{it}/(NT) \qquad [A9]$$

y similarmente para X y X^+. Es decir, basta con llevar a cabo la sencilla transformación de [A8] en todos los datos, y aplicar mínimos cuadrados ordinarios, a los nuevos datos para obtener el estimador de B. Si se desea obtener estimadores de los efectos individuales y temporales, se puede recurrir a la segunda parte de los resultados de la regresión particionada, es decir,

$$\hat{\gamma} = (Z'Z)^{-1} \quad Z'(Y - x\hat{B}) \qquad [A10]$$

Definiendo $\hat{V} = (Y - x\hat{B})$, es fácil ver que se obtienen los siguientes estimadores,

$$\hat{\gamma}_i = \hat{v}_{i.}$$
$$\hat{\gamma}_t = \hat{v}_{.t} \qquad [A11]$$

donde $\hat{v}_{i.}$, $\hat{v}_{.t}$ están definidos análogamente a $Y_{i.}$, $Y_{.t}$ en [A9]. Estos estimadores cumplen la restricción

$$\frac{1}{N} \sum_i \hat{\gamma}_i = \hat{V}_{..} = \frac{1}{T} \sum_t \hat{\gamma}_t \qquad [A12]$$

como efectivamente deberá ocurrir, ya que el número de efectos identificables en un modelo de estas características es $N + T - 1$.

Finalmente, los estimadores de $\hat{\varepsilon}$ se obtienen por medio de la expresión,

$$\hat{\varepsilon}_{it} = \hat{V}_{it} - \hat{V}_{i.} - \hat{V}_{.t} + \hat{V}_{..} \qquad [A13]$$

y el estimador de σ_ε^2, utilizando la expresión usual que en este caso es,

$$\hat{\sigma}_\varepsilon^2 = \hat{\varepsilon}'\hat{\varepsilon}/((N - 1)(T - 1) - K) \qquad [A14]$$

siendo 'k' la dimensión de B.

Supongamos ahora que en lugar del modelo [A1] de efectos fijos, el comportamiento de 'Y_{it}' se describe mejor por el modelo de efectos aleatorios expuesto a continuación:

$$\begin{aligned} Y_{it} &= X_{it}'B + \varepsilon_{it} \qquad & i = 1, ..., N \\ \varepsilon_{it} &= u_{it} + v_t + w_{it} \qquad & t = 1, ..., T \end{aligned} \qquad [A16]$$

donde (u, v, w) son errores aleatorios incorrelacionados entre sí y para diferentes valores de sus subíndices, de media cero y varianza constante. Puede demostrarse que la inversa de la matriz de covarianzas de 'ε' es proporcional a la siguiente matriz

$$H = I - (\gamma_1 T)\, A - (N\gamma_2)\, B - (\gamma_3 TN) C \qquad [A17]$$

donde

$$\gamma_1 = \sigma_u^2/(\sigma_w^2 + T\sigma_u^2)$$
$$\gamma_2 = \sigma_v^2/(\sigma_w^2 + T\sigma_v^2)$$
$$\gamma_3 = \gamma_1\gamma_2[(2\sigma_w^2 + N\sigma_v^2 + T\sigma_u^2)/(\sigma_w^2 + N\sigma_v^2 + T\sigma_u^2)] \qquad [A18]$$

y las matrices A, B, C han sido definidas en [A6].

(El método para obtener [A17], consiste en obtener la descomposición de la matriz de covarianzas de 'ε' en sus raíces y vectores característicos. La inversa se obtiene inmediatamente sustituyendo en la descomposición la matriz diagonal de raíces características por su inversa. El interés del método estriba, por supuesto, en que las raíces y vectores mencionados son relativamente fáciles de detener. Este método es debido a Hannan.)

El estimador óptimo de 'B' en [A16] se obtiene ahora aplicando mínimos cuadrados generalizados. Esto requiere entonces la formación de los siguientes productos,

$$\begin{aligned} &(Y'HX) \\ &(X'HX) \end{aligned} \qquad [A19]$$

Supongamos que podemos descomponer H en un producto de la forma $H = ww$, donde $w = w'$. Entonces, los productos de [A19] pueden escribirse del modo siguiente

$$\begin{aligned} Y'HX &= Y'w'wX = Y^{+\prime}X^{+} \\ X^{+} &= wX \\ Y^{+} &= wY \end{aligned} \qquad [A20]$$

Es decir, que para obtener los mínimos cuadrados generalizados basta con aplicar mínimos cuadrados ordinarios a las variables transformadas en (Y^+, X^+). Supongamos ahora que la matriz w tiene la siguiente forma,

$$w = (I - aA - bB + cC) \qquad [A21]$$

Utilizando las propiedades de [A7] e igualando coeficientes, obtenemos los siguientes valores para los coeficientes (a, b, c) que garantizan la propiedad $ww = H$:

$$\begin{aligned} a &= 1 - \sigma_w/\sqrt{\sigma_w^2 + T\sigma_u^2} \\ b &= 1 - \sigma_w/\sqrt{\sigma_w^2 + N\sigma_V^2} \\ c &= \left(\frac{NT\gamma_3}{2} - (b + a)\right) - 1/2(1 + \sqrt{4((T\gamma_1 + N\gamma_2) + (b + a) - 3)} \end{aligned}$$

[A22]

de modo que finalmente, las transformaciones de [A20] pueden escribirse como sigue,

$$Y_{it}^{+} = Y_{it} - aY_{i.} - bY_{.t} + cY_{..} \qquad [A23]$$

y similarmente X_{it}^{+}.

El caso en que uno de los efectos es fijo y otro aleatorio, es algo más complejo, pero pueden obtenerse las transformaciones análogamente, utilizando los procedimientos descritos en este apéndice para los casos de efectos fijos y aleatorios conjuntamente. (Esta técnica es la utilizada por Johnston en un caso más sencillo.)

Apéndice 2. Ordenes de magnitud de los sesgos en el modelo dinámico

Para centrar la atención en el problema de interés, vamos a suponer que el modelo dinámico está descrito por la ecuación siguiente

$$\begin{aligned} Y_{it} &= \alpha Y_{it-1} + \varepsilon_{it} \quad i = 1, \ldots, N \\ \varepsilon_{it} &= u_i + w_{it} \quad t = , \ldots, T \end{aligned} \qquad [A24]$$

donde los símbolos poseen el significado usual, descrito en la sección anterior. La matriz de covarianzas de 'e' en este caso está dada por,

$$\begin{aligned} \Sigma &= (I_{NT} + (I_N \otimes J_T)\rho)\sigma_w^2 \\ \rho &= \sigma_u^2/\sigma_w^2 \end{aligned} \qquad [A25]$$

Consideramos en primer lugar, el sesgo en la estimación de 'α' obtenido al aplicar el modelo de efectos fijos. La matriz de variables artificiales entonces es,

$$Z = (I_N \otimes L_T) \qquad [A26]$$

y su matriz asociada M_z,

$$\begin{aligned} M_z &= I - z(z'z)^{-1}z' \\ &= I - (I_N \otimes J_T)1/T \\ &= I - B \end{aligned} \qquad [A27]$$

Para calcular el sesgo, podemos considerar que $N \to \infty$ y T es un valor fijo. Entonces, tomamos límites en probabilidad, en la expresión mínimo cuadrática, que tenderán a las esperanzas de las expresiones relevantes. Es fácil ver que la expresión de la que dependerá el sesgo es,

$$\frac{1}{NT} \sum_{i,t} E(Y_{i,\,t-1}\varepsilon_{it}^{+}) \qquad [A28]$$

donde

$$\begin{aligned} \varepsilon_{it}^{+} &= (I - B)\varepsilon \\ &= \varepsilon_{it} - \varepsilon_{i.} \\ &= w_{it} - w_{i.} \end{aligned} \qquad [A29]$$

Consideremos ahora

$$
\begin{aligned}
E(Y_{i,\,t-1}\varepsilon_{it}^{+}) &= -EY_{i,\,t-1}w_{i.} = -EY_{i,\,t-1}\left(\sum_t w_{it}/T\right) \\
&= -\frac{1}{T}\sum_{s=1}^{t-1}(EY_{i,\,t-1}w_{is}) \\
&= -\frac{\sigma_w^2}{T}(1+\alpha+\alpha^2+\cdots+\alpha^{t-2}) \\
&= -\frac{\sigma_w^2}{T}\left(\sum_{s=0}^{t-2}\alpha^s\right) \\
&< -\frac{\sigma_w^2}{T}\left(\frac{1}{1-\alpha}\right) \quad (|\alpha|<1) \\
&= 0\left(\frac{1}{T}\right)
\end{aligned}
\qquad [A30]
$$

Es evidente ahora que la expresión de [A28] es $0(1/T)$. Es decir, que el sesgo no desaparece a no ser que $T \rightarrow \infty$. Lo que esto significa, es que si T es un valor pequeño, aunque N sea alto, el tamaño del sesgo puede ser bastante considerable, y ésta suele ser precisamente la situación con datos en panel.

Podemos observar también que las expresiones de [A29] son idénticas si en el modelo original, los efectos aleatorios 'u' son sustituidos por efectos fijos: en otras palabras, el sesgo de [A30] es el mismo, sean los efectos verdaderos, fijos o aleatorios.

Consideremos ahora los sesgos obtenidos por aplicación del método de mínimos cuadrados generalizados. La transformación de [A29] se sustituye ahora por la siguiente,

$$\varepsilon_{jt}^{+} = \varepsilon_{it} - \phi\varepsilon_{i.} = w_{it} - \phi w_{i.} + u_i(1-\phi) \qquad [A31]$$

donde $\phi = 1 + 0(1/\sqrt{T})$ (véase [A22] para el caso general).

La expresión a considerar en este caso es

$$\frac{1}{NT}\sum_{i,\,t} E(Y_{i,\,t-1}^{+}\varepsilon_{it}^{+}) \qquad [A32]$$

ya que la matriz de transformación de los datos no es idempotente (véase [A20] y [A21]).

Como ocurre que

$$
\begin{aligned}
E(Y_{it}u_i) &= \sigma_u^2 \\
E(Y_{i.}u_i) &= \frac{1}{T}\sum_t E(Y_{it}u_i) = \sigma_u^2
\end{aligned}
\qquad [A33]
$$

es fácil obtener el siguiente orden de magnitud,

$$E(Y^{+}_{i,\,t-1}\varepsilon^{+}_{it}) = E(Y_{i,\,t-1} - \phi y_{i.})(\varepsilon_{it} - \phi\varepsilon_{i.}) = 0(1/\sqrt{T}) \qquad [A34]$$

ya que $\phi = 1 + 0(1/\sqrt{T})$. Es decir, el sesgo del estimador de mínimos cuadrados generalizados será de mayor orden de magnitud, aun en el supuesto de que los efectos sean aleatorios. Para interpretar este resultado, notamos en primer lugar, que si $\sigma^2_u \rightarrow \infty$, $\phi \rightarrow 1$, y el estimador tiende al de efectos fijos [A22]. Entonces, si σ^2_u, es alto en relación a σ^2_w, será más conveniente utilizar el estimador de efectos fijos. Pero si σ^2_u es bajo en relación a σ^2_w es probable que la varianza del estimador de mínimos cuadrados generalizados sea bastante menor, de modo que compense el sesgo, y dé finalmente un error cuadrático medio menor. Hay que resaltar, que en la práctica, suele darse precisamente que σ^2_u es bastante bajo en relación a σ^2_w.

Apéndice 3. Cálculo de las medidas de ajuste en modelos de variables dependientes cualitativas

En el caso del modelo lineal, la estimación restringida consiste en hacer $B = 0$ en el modelo

$$Y_i = \alpha + x_i{'}B + u_i \qquad [A35]$$

es decir, que en el modelo restringido, solamente se estima la media de Y.

En el caso del modelo logístico binomial [37] haciendo $B = 0$, obtenemos

$$p(Y_i = 1) = \frac{1}{1 + 1} = 1/2 \qquad [A36]$$

si no hay constante. Si la hay, ocurre lo siguiente $X_i{'}B = B_0 + x_i{'}B_1$

$$p(Y_i = 1) = \frac{a}{1 + a}, \quad a = e^{B_0} \qquad [A37]$$

de forma que $p_i = p = (N_1/N)$ siendo N_1 el número de individuos que escogen la alternativa 1. Entonces, la función de verosimilitud restringida es,

$$\phi = (N_1/N)^{N_1}(1 - N_1/N)^{N-N_1} \qquad [A38]$$

Si $N = 2N_1$, es fácil comprobar que $0{,}75\ R^2_c = R^2$, de modo que la corrección de [48] es substancial. El caso probit binomial se analiza similarmente.

Para analizar el caso logístico multinomial, notamos, en primer lugar, que es preciso normalizar la expresión [58]. Para ello multiplicando el numerador y denominador de esta expresión por exp (α) obtenemos lo siguiente:

$$\exp\,(\alpha + x_{ij}'B) \Big/ \sum_{h=0}^{m} \exp\,(\alpha + x_{ij}'B) \qquad [A39]$$

Si hay una constante en x_{ij}, podemos escribir

$$x_{ij}'B = B_0 + x_{1ij}'B_1$$
$$\alpha + x_{ij}'B = (\alpha + B_0) + x_{1ij}'B_1 = x_{ij}'B^+ \qquad [A40]$$

de modo que la constante no está identificada. Una manera de eliminar esta indeterminación es hacer $\alpha = -x_{i0}'B$, de modo que [A39] se convierte en,

$$\exp\,(Z_{ij}'B) \Big/ \left(1 + \sum_{j=1}^{m} \exp\,(Z_{ij}'B)\right) \qquad [A41]$$
$$Z_{ij}'B = (x_{ij} - x_{i0})'B$$

con lo cual la constante queda eliminada. Para evaluar el caso restringido, haciendo $B = 0$ obtenemos, $p_{ij} = p_j = (1/m)$. Un supuesto más razonable es el siguiente:

$$x_{ij}'B = \gamma_j + x_{1ij}'B_1 \qquad [A42]$$

de forma que

$$Z_{ij}'B = (\gamma_j - \gamma_0) + (x_{1ij} - x_{1i0})'B_1 = \pi_j + z_{ij}^{+}{}'B_1 \qquad [A43]$$

y ahora en la estimación restringida haremos $B_1 = 0$ obteniendo,

$$p_{ij} = p_j = (N_j/N)$$
$$\phi = \prod_{j=1}^{m} (N_j/N)^{N_j} \qquad [A44]$$

Otra medida del éxito del modelo frecuentemente utilizada, es la proporción de casos correctamente clasificados sobre casos totales.

Referencias

Alonso, J. A. (1986): «Especialización exportadora y tipología de la empresa: un análisis multivariante», *Investigaciones Económicas.*

Amemiya T. (1973): «Regression analysis when the dependent variable is truncated normal», *Económetrica.*

Amemiya, T. (1985): *Advanced Econometrics*. Basil Blackwell. Oxford.
Anderson, T W. and Hsiao, Ch. (1982): «Formulation and estimation of dynamic models using panel data», *Journal of Econometrics*.
Anderson, T. W. (1958): *An introduction to multivariate statistical analysis*. Wiley and Sons. New York.
Balestra, P. and Nerlove, M. (1966): «Pooling cross section and time series data in the estimation of a dynamic model: the demand for natural gas», *Econométrica*.
Berges, A. (1986a): «La medición de la dimensión empresarial: una comparación internacional», *Investigaciones Económicas*.
Berges, A. (1986b): «Expectativas de demanda y planes de producción en las empresas», *Investigaciones Económicas*.
Calvet, J., Coll, J., Salamero, A. y Sansalvado, M. (1986): «Análisis sistemático de variables en la economía industrial española: rentabilidad, productividad y competitividad», *Investigaciones Económicas*.
Cosslett, S. (1981): «Maximum likelihood estimator for choice-based samples», *Econométrica*.
Cuervo, C. (1986): «Inversión y financiación en la empresa industrial española», *Investigaciones Económicas*.
Chamberlain, G. (1984): «Panel data», *Handbook of Econometrics*.
Duran, J. J. y Lamothe, P. (1986): «Análisis empírico del comportamiento económico-financiero de las grandes empresas españolas participadas por el capital extranjero», *Investigaciones Económicas*.
Espitia, M., Salas, V. y Yagüe, M. J. (1986): «Generación y reparto de beneficios en los mercados industriales españoles: contraste empírico a partir del ratio '*q*'», *Investigaciones Económicas*.
Espitia, M. y Salas, V. (1986): «'*q*' de Tobin y regulación: aplicación al sector eléctrico español», *Investigaciones Económicas*.
Fariñas, J. C. y Rodríguez, L. (1986): «Los resultados de las mayores empresas industriales de la CEE y España (1973-1982): un estudio comparativo», *Investigaciones Económicas*.
Fair, R. (1978): «A theory of extramarital affairs», *Journal of Political Economy*.
Fernández, A. I. (1986): «La gestión de la cartera de créditos en una entidad financiera», *Investigaciones Económicas*.
Gronau, R. (1973): «The effect of children on the housewife's value of time», *Journal of Political Economy*.
Hausman, J. and Wise, D. (1978): «A conditional probit model for qualitative choice: discrete decisions recognizing interdependence and heterogeneous preferences», *Econométrica*.
Hausman, J., Wall, B. and Griliches, Z. (1981): «Econometric models for count data with and application to the patents R. D. relationship», Discussion paper NBER.
Heckman, J. (1976): «Simultaneous equations model with continuous and discrete endogenous variables and structural shifts», en *Studies in non linear estimation*, ed. Goldfeld Quandt. Cambridge: Ballinger.
Heckman, J. (1981): «Heterogeneity and state dependence in dynamic models of labour supply», in *Conference on low income labour markets*, ed. S. Rosen University of Chicago Press.
Heckman, J. y Willis, R. (1977): «A beta logistic model for the analysis of sequential labour force participation of married women», *Journal of Political Economy*.
Jaumandreu, J. y Mato G. (1986): «Concentración industrial en España: 1973-81», Jornadas de Economía Industrial.
Johnston, J. (1972): *Econometrics*. McGraw-Hill. New York.
Lancaster, T. (1974): «Some econometric of counts of events». Discussion paper. University of Hull.
Maddala, G. S. (1983): *Limited-dependent and qualitative variables in econometrics*. Cambridge University Press.
Maddala, G. S. (1971): «The use of variance components in pooling cross section and time series data», *Econométrica*.

Mc Fadden, D. (1974): «The measurement of urban travel demand», *Journal of public economics*.
Mc Fadden, D. L. (1984): «Econometric analysis of cualitative response models». *Handbook of Econometrics*.
Maravall, F. y Torres, A. (1986): «Comportamiento exportador de las empresas y competencia imperfecta», *Investigaciones Económicas*.
Martínez Mongay, C. (1986): «El análisis de las diferencias de eficiencia entre empresas. Aplicación de un modelo econométrico a la PYME española en 1980», *Investigaciones Económicas*.
Mood, A., Graybill, F. and Boes, D. (1974): *Introduction to the theory of statistics*. Mc Graw-Hill.
Mundlak, Y. (1978): «On the pooling of time series and cross section data», *Econométrica*.
Olsen, R. (1978): «Note on the uniqueness of the maximum likelihood estimator for the Tobit model», *Econométrica*.
Pagan, A. y Hall, R. (1983): «Diagnostic tests as residual analysis», *Econometric Reviews*.
Pérez, E. (1986): «Cadenas de Markov, teoría de la información y cuotas de venta», Jornadas de Economía Industrial.
Plackett, R. (1974): *The analysis of categorical data*. London Griffin.
Scheffe, H. (1959): *The analysis of variance*. Wiley and Sons. Toronto.
Salvador, R. y Sole, F. (1986): «La teoría de Marris y el crecimiento de las empresas españolas», Jornadas de Economía Industrial.
Tobin, J. (1958): «Estimation of relationships for limited dependent variables», *Econométrica*.
Theil, H. (1971): *Principles of econometrics*. Wiley and Sons. New York.
Wallace, T. D. and Hussain, A. (1969): «The use of error components models in combining cross section with time series data», *Econométrica*.
White, H. (1980): «A heteroskedasticity-consistent covariance matrix estimator and a direct test for heteroskedasticity», *Econométrica*.

Abstract

This paper is a survey of currently available techniques for the statistical treatment of panel data. The paper is adressed to the applied researcher, so that the more theoretical aspects are avoided. The main topics dealt with are, the fixed and random effects models, qualitative, censored, and truncated models, and some applications of the analysis of variance and non parametric methods.

Recepción del original, octubre de 1986.
Versión final, diciembre de 1986.