Investigaciones Económicas (Segunda época). Vol. XIV, n.º 1 (1990), págs. 3-45

LA ECONOMETRIA DE DATOS DE PANEL*

Manuel ARELLANO
London School of Economics

Olympia BOVER
Centre for Labour Economics,
London School of Economics

Este trabajo presenta una síntesis de los métodos disponibles para el análisis econométrico de datos de panel en un marco unificado. En particular se analiza en que forma las propiedades de diversos estimadores dependen de los supuestos acerca de las variables explicativas, efectos permanentes inobservables y términos de perturbación. Primeramente se estudian modelos lineales estáticos y dinámicos con efectos individuales. A continuación, el análisis se extiende a modelos con variables limitadas con efectos individuales y respuestas dinámicas.

1. Introducción

A lo largo de los 80 ha ido desarrollándose un creciente interés en el uso de los datos de panel en los estudios econométricos. Este interés en parte refleja la disponibilidad de nuevos conjuntos de datos de este tipo, pero sobre todo el creciente escepticismo en la profesión sobre las posibilidades de estimar modelos de comportamiento individual con datos agregados de series temporales. Desde otro ángulo, también refleja la preocupación por las distorsiones que las diferencias inobservables entre individuos introducen en las estimaciones obtenidas a partir de encuestas de corte transversal. Paralelamente a este interés, se han producido considerables avances en las técnicas econométricas en este campo y cambios metodológicos significativos. En este trabajo nos proponemos presentar una síntesis de los métodos disponibles poniendo énfasis en la motivación subyacente en cada caso y su dominio de aplicación. No hemos intentado un compendio exhaustivo y el trabajo, como es natural, refleja nuestros propios intereses. No obstante también tratamos de poner en contexto y ofrecer referencias sobre las áreas que no son estudiadas en mucho detalle.

Se dice que un conjunto de datos es de panel cuando se tienen observaciones de series temporales sobre una muestra de unidades individuales. Es decir,

* Agradecemos los comentarios de Juan José Dolado, José Luis Raymond y de un evaluador anónimo.

un conjunto de individuos son observados en distintos momentos en el tiempo. Digamos que para una variable y_{it} se tienen $i = 1, ..., N$ observaciones de corte transversal y $t = 1, ..., T$ observaciones de series temporales. Algunas veces, i denota países, regiones o sectores industriales en cuyo caso N es relativamente pequeño, mientras que T es grande. Sin embargo, en un típico micropanel de familias o empresas la situación es la contraria. T puede ser tan pequeño como 3, 4 ó 5 mientras que N se refiere a cientos o miles de individuos.

Para el estudio de las propiedades estadísticas de los métodos de inferencia en econometría normalmente se utilizan aproximaciones asintóticas. Si N es pequeño y T es grande se puede pensar que tenemos N variables que explicar $y_{1t}, ..., y_{Nt}$ sobre las cuales tenemos observaciones de series temporales. Es posible que aparezcan restricciones de igualdad relacionando los parámetros de unas ecuaciones con otras, pero esencialmente tenemos un modelo de series temporales multivariante. En este caso las aproximaciones asintóticas relevantes son para N fijo y $T_{\rightarrow\infty}$. Sin embargo, con micropaneles, la utilización de resultados asíntoticos para $T_{\rightarrow\infty}$ no se espera que en general proporcione buenas aproximaciones a las distribuciones de los estimadores y otros estadísticos de interés y por tanto consideramos $N_{\rightarrow\infty}$ y T fijo. De este modo podemos pensar que disponemos de T variables a explicar $y_{i1} ... y_{iT}$ sobre las que tenemos N observaciones de corte transversal. Este es el caso en el que nos vamos a concentrar en este trabajo.

Es importante distinguir los datos de panel de las series temporales de cortes transversales independientes. A menudo, organismos públicos producen encuestas a intervalos regulares de tiempo con criterios similares, pero una nueva muestra aleatoria es obtenida en cada ocasión. Estos datos no permiten la comparación de las observaciones de un individuo particular con su propio pasado y, por tanto, no constituyen un panel[1]. Sin embargo, frecuentemente disponemos de paneles incompletos en el sentido de que se tienen más observaciones temporales para unos individuos que para otros o, siendo el mismo número, los períodos históricos a los que las observaciones corresponden varían. En este caso, no hay ninguna diferencia fundamental con el panel completo, especialmente si las adiciones y eliminaciones de individuos en la muestra se producen de forma aleatoria.

Una de las ventajas principales de los micropaneles es que permiten relajar y contrastar supuestos que están implícitos en los análisis de corte transversal. Hay aquí dos aspectos fundamentales. En primer lugar, la posibilidad de controlar heterogeneidad inobservable. Consideremos una regresión lineal de corte transversal

$$y_i = x_i\beta + u_i \qquad (i = 1, ..., N)$$

[1] Sin embargo, véase Deaton (1985) en donde se explica un método para explotar este tipo de datos formando un pseudo panel de grupos homogéneos. Una muestra de estas características es utilizada en el influyente trabajo de Browning, Deaton y Irish (1985).

Una limitación importante de este tipo de análisis es que es difícil saber si los coeficientes estimados reflejan realmente el impacto de x_i o por el contrario se deben a diferencias inobservables entre los individuos que están correlacionadas con x_i. Si esta heterogeneidad permanece relativamente constante a lo largo del tiempo el panel puede solucionar el problema. Supongamos que tenemos dos observaciones sobre cada unidad i en los períodos 1 y 2:

$$y_{i1} = x_{i1}\beta + \eta_i + v_{i1}$$

$$y_{i2} = x_{i2}\beta + \eta_i + v_{i2}$$

en donde η_i representa diferencias inobservables correlacionadas con x_{i1} y x_{i2}. Nótese que la regresión de $[y_{i2} - y_{i1}]$ sobre $[x_{i2} - x_{i1}]$ identifica β. Por ejemplo y_{it} puede ser el logaritmo de las horas trabajadas por el individuo i en el período t y x_{it} el logaritmo del salario marginal en términos reales. Si el objetivo es estimar la elasticidad de sustitución intertemporal de los salarios, sería muy problemático con una muestra de corte transversal en la medida en que las horas trabajadas dependen no sólo de las variaciones en los salarios marginales presentes sino de las percepciones acerca del flujo de salarios futuros que se pueden considerar como relativamente constantes y son típicamente inobservables. En concreto, en un modelo de este tipo η_i representaría la utilidad marginal de la riqueza (estos modelos han sido estudiados por MaCurdy, 1981 y Heckman y MaCurdy, 1980). Naturalmente si $x_{i1} = x_{i2}$, β desaparece de la ecuación al tomar diferencias y por lo tanto no puede ser estimado. Por desgracia, hay ejemplos importantes de esta situación; un caso que ha sido muy estudiado es la relación entre el logaritmo de los ingresos $[y_{it}]$, nivel de educación (x_i, constante) y «habilidad» (η_i, inobservable) (véase por ejemplo, Griliches, 1977). En realidad, los datos de panel no son tan útiles en este caso como cuando se puede explotar la variación temporal para separar la variación transversal permanente. Sin embargo, existe una abundante literatura sobre modelos que contienen variables explicativas que no varían con el tiempo cuyos coeficientes son identificados gracias a supuestos de no correlación de todas o algunas de las variables explicativas con los efectos individuales (*cf.* Hausman y Taylor, 1981 y artículos relacionados). En ciertos casos la modelización de los efectos permanentes inobservables o de variación lenta en el tiempo es una parte importante del problema económico de interés. Aquí también la disponibilidad de datos de panel puede resolver el problema. Véase por ejemplo Bover (1989) en donde se explota el panel de familias de la Universidad de Michigan (PSID) para modelizar la utilidad marginal de la riqueza en un modelo intertemporal de oferta de trabajo.

El segundo aspecto en el que el panel resulta decisivo respecto a una muestra de corte transversal es la posibilidad de modelizar respuestas dinámicas con microdatos. Ecuaciones con retardos de variables endógenas y exógenas pueden ser especificadas permitiendo la posibilidad de explicar procesos de ajuste. Costes de ajuste, habituación, consideración del futuro, etc. generan autocorrelación en las decisiones de familias y empresas que los modelos empíricos han de ser capaces de reproducir. Por otra parte, en ciertos modelos con expectativas racionales los términos de perturbación tienen la interpretación de «sorpresas» que por tanto están incorrelacionadas con todas las variables retardadas. De este modo, los retardos se convierten de forma natural en las variables instrumentales que permiten identificar los parámetros de interés (*cf.* Hansen y Singleton, 1982). Dada la disponibilidad de retardos, estos modelos son identificables con datos de panel (véase por ejemplo Zeldes, 1989 para la estimación de un modelo de consumo; Hotz, Kydland y Sedlacek, 1988 sobre oferta de trabajo, ambos utilizando el panel PSID; y Hayashi e Inoue, 1989 sobre inversión utilizando un panel de empresas japonesas).

Los modelos autorregresivos con efectos individuales han atraído considerable atención y actualmente se dispone de un amplio repertorio de métodos de estimación. Los efectos individuales generan autocorrelación positiva en los datos y es importante separar esta fuentes de persistencia de la dinámica genuina. No obstante, en paneles cortos (con sólo 3 ó 4 observaciones por individuo) la cantidad de información sobre los coeficientes autorregresivos puede llegar a ser muy escasa.

Cuando se está interesado en contrastar proposiciones de comportamiento o modelizar las reglas de decisión de individuos o empresas, los modelos agregados de series temporales parten de la consideración de un «agente representativo» al que se asimilan las observaciones agregadas. Es una creencia corriente el que la agregación a menudo oculta características interesantes del problema que se analiza. En particular, el marco del agente representativo o supuesto *per cápita* al no considerar efectos distribucionales implica, excepto en casos especiales, una serie de discrepancias entre los coeficientes identificados en el modelo agregado y los parámetros de interés económico objetivo del análisis (*cf.* Stoker, 1986). Por otra parte, muchas veces se puede esperar que la rica variabilidad de corte transversal presente en los paneles de datos ayude a aumentar la precisión de las estimaciones de series temporales incluso si la agregación no fuera un problema. La observación de Arthur Goldberger acerca de la situación actual de la modelización macroeconómica en una reciente entrevista en *Econometric Theory* (Kiefer, 1989) es un buen reflejo del estado de ánimo que está induciendo a un creciente número de investigadores a construir y utilizar datos de panel en su trabajo aplicado:

> *I used to think that one of the reasons for doing modeling was data reduction, that is, you have a massive batch of data and you want to boil it down into more manageable form. But there's a recent article using macroeconomic time series in a major economics journal. I counted the number of numbers that are recorded in that article —the coefficients and standard errors. There are many more numbers reported, not just more digits, more numbers, more figures reported than there were data observations— 20 percent more.*

Los modelos estimados a partir de datos de paneles tienen sus propios problemas y limitaciones. Algunas veces estos problemas pueden llegar a ser tan serios como para plantearse si realmente vale la pena construir paneles en contraposición a encuestas de corte transversal (véase Ashenfelter, Deaton y Solon, 1986, para una discusión pormenorizada en el contexto de las necesidades estadísticas de países en desarrollo). Uno de los problemas fundamentales es el de los errores de medida, especialmente en variables financieras, que son endémicos en las encuestas (en tanto que los agentes tienen un incentivo a no responder o a dar respuestas falsas de forma sistemática; sobre este punto véase por ejemplo Bowden, 1989) y se ven acentuados en los paneles como consecuencia de las entrevistas repetidas a los mismos individuos y sobre todo por considerar ecuaciones en diferencias o en desviaciones con vistas a eliminar componentes fijos inobservables. Griliches y Hausman (1986) presentan procedimientos que explotan la propia estructura del panel para identificar parámetros de interés en presencia de errores en las variables. Otro serio problema es el de la falta de representatividad de la muestra tras una serie de períodos debido a la imposibilidad de encontrar los mismos individuos o a las negativas a responder. En la medida en que este proceso de selección está relacionado con el fenómeno que se quiere modelizar, introducirá sesgos en los estimadores convencionales. Hausman y Wise (1979) presentan un simple modelo en el que se tiene en cuenta este proceso de selección y Hall (1987) también controla el efecto de selección en su análisis de las propiedades dinámicas de la tasa de crecimiento del empleo en un panel de empresas, pero en este campo se ha hecho todavía relativamente poco. Un ejemplo importante es la desaparición de empresas de un panel debido a adquisiciones, fusiones, o bancarrotas cuando se estudian las decisiones de inversión (Pakes y Ericson, 1987, llevan a cabo un análisis no paramétrico de modelos de supervivencia de empresas). Finalmente, el éxito de los modelos de datos de paneles en controlar diferencias permanentes inobservables hasta el momento se limita sobre todo a modelos lineales. Sin embargo, a menudo nos encontramos con modelos teóricos que sugieren componentes de heterogeneidad que entrarían de forma no lineal en las ecuaciones de comportamiento individual. Una importante categoría de modelos no lineales en econometría son los modelos con variables dependientes limitadas (incluyendo los modelos de elección discreta). Por regla general, incluso si la especificación subyacente es lineal, los efectos individuales no se pueden eliminar con una simple transformación y la solución ha de pasar por la introducción de supuestos más restrictivos en la especificación.

Uno de los trabajos más influyentes en el campo de la econometría de datos de panel es el capítulo de Chamberlain en el *Handbook of Econometrics* (Chamberlain, 1984) que engloba y extiende el trabajo anterior más relevante de este autor en al área de datos de panel (Chamberlain, 1980 y 1982). Un análisis pormenorizado de métodos y modelos en la literatura se puede encontrar en la monografía de Hsiao (1986) y también en Hsiao (1985). Otra referencia muy útil es el trabajo de Mauleón (1987), que contiene una introducción a los métodos para datos de panel y a los modelos con variables dependientes limitadas poniendo énfasis en los problemas de aplicación práctica.

El resto de este artículo está organizado de la siguiente forma. En la Sección 2, se discuten los modelos lineales básicos con los métodos de inferencia asociados. Se pone especial énfasis en clarificar la dicotomía entre efectos fijos y efectos aleatorios. La Sección 3 estudia modelos dinámicos y modelos con errores en las variables. La Sección 4 está dedicada a modelos con información en niveles, esto es, modelos en los que la variación transversal no está totalmente contaminada por heterogeneidad inobservable y, por tanto, puede ser posible la estimación de los efectos de variables que no varían con el tiempo. La Sección 5 considera modelos con variables dependientes limitadas y modelos de elección discreta con efectos individuales. Finalmente, la Sección 6 contiene las conclusiones.

Aunque por lo general escribiremos los modelos como sistemas de T ecuaciones, todos los modelos que consideramos en este trabajo son uniecuacionales en el sentido de que el objeto del análisis es una sola ecuación de comportamiento. Si el modelo de interés es un sistema de ecuaciones simultáneas, los métodos que describimos se pueden utilizar para estimar consistentemente ecuación por ecuación. Si hay restricciones entre los parámetros de distintas ecuaciones, éstas se pueden imponer y contrastar en una segunda fase utilizando el método de distancia mínima o de mínimos cuadrados asintóticos (*cf.* Chamberlain, 1982; Gourieroux, Monfort y Trognon, 1985).

De cara a simplificar la presentación, tampoco desarrollamos las complicaciones que surgen cuando el panel es incompleto. Estas complicaciones son de orden técnico y no introducen cambios conceptuales. Una exposición sobre el tema se puede encontrar en Hsiao (1986, Ch. 8). Por otra parte, el programa DPD de Arellano y Bond (1988b) se puede utilizar para la estimación de modelos dinámicos a partir de datos de panel incompleto.

El tipo específico de modelo en el que nos concentramos en este trabajo es un modelo de regresión lineal (o lineal generalizado en el caso de los modelos con variables dependientes limitadas, utilizando la terminología de McCullagh y Nelder, 1983) en el que el objetivo del análisis es la estimación consistente de los parámetros de regresión dada la existencia de una variable explicativa latente, constante en el tiempo, que puede estar correlacionada con las variables explicativas que se observan. Un tipo distinto de modelos con variables latentes (que no consideramos) son aquellos en los que se especifica un sistema de ecuaciones que dependen de varios factores inobservables. El interés en ese caso se centra en la estimación de la estructura de varianzas y covarianzas implicada por la especificación. La fuente de identificación de estos modelos es la presencia de las mismas variables latentes en distintas ecuaciones. Por ello es de esperar que la disponibilidad de datos de panel permita la identificación de modelos más ricos. Los trabajos de Hall y Mishkin (1982) y de Abowd y Card (1989) son ejemplos de aplicaciones importantes en este sentido.

2. Efectos fijos y efectos aleatorios

El modelo que vamos a considerar en esta sección es una regresión lineal con efectos individuales de la forma

$$y_{it} = x_{it}'\beta + \eta_i + v_{it} \qquad (i = 1, ..., N; t = 1, ..., T) \qquad [1]$$

en donde x_{it} es un vector $k \times 1$ de variables explicativas, β es el vector de parámetros a estimar, η_i es un efecto individual y v_{it} un término de perturbación. Si el modelo incluye efectos temporales, éstos se suponen incluidos en β en cuyo caso x_{it} contiene las correspondientes variables ficticias de tiempo. Como en todo momento supondremos que T es fijo en el análisis asintótico, es razonable tratar los efectos temporales como coeficientes que se pueden estimar consistentemente. Incluyendo efectos temporales en la ecuación se controla la influencia de todas las posibles variables macroeconómicas sobre el comportamiento individual.

En el modelo de efectos fijos los η_i son tratados como un conjunto de N coeficientes adicionales que se pueden estimar junto con β. Por el contrario, en el modelo de efectos aleatorios tradicional se supone que η_i es una variable aleatoria inobservable independiente de x_{it} que por tanto pasa a formar parte de un término de perturbación compuesto

$$u_{it} = \eta_i + v_{it} \qquad [2]$$

Por esta razón a estos modelos se les llama también modelos con errores compuestos *(error components)*. Una costumbre muy extendida en el trabajo aplicado consiste en estimar ambos modelos para a continuación contrastar «si los efectos son fijos o aleatorios». Este es quizá el malentendido más importante en este campo, del que son responsables en buena medida los primeros trabajos econométricos sobre datos de paneles. El que los efectos se supongan fijos o aleatorios no representa una cualidad intrínseca de la especificación. En realidad, los efectos individuales se pueden considerar siempre aleatorios sin pérdida de generalidad. La distinción crucial es si los efectos están correlacionados o no con las variables observables x_{it}. Si η_i está correlacionado con x_{it} puede ser conveniente hacer inferencia condicional sobre las realizaciones de los η_i en la muestra («efectos fijos»), mientras que si los η_i no están correlacionados con x_{it} es natural hacer inferencia incondicional como ocurre en el modelo de errores compuestos. Los primeros autores en insistir sobre este punto fueron Mundlak (1978) y Chamberlain (1980). A continuación presentamos los métodos de estimación citados los que nos permitirá precisar los comentarios anteriores.

El modelo [1] se puede escribir de forma compacta

$$y = X\beta + C\eta + v \qquad [3]$$

en donde $X = [x_{11} \dots x_{1T} \dots x_{N1} \dots x_{NT}]'$ es $NT \times k$, $y = [y_{11} \dots y_{1T} \dots y_{N1} \dots y_{NT}]'$ es $NT \times 1$ con una notación similar para v, $C = I_N \otimes \iota$, ι es un vector $T \times 1$ de unos y η es el vector $N \times 1$ de efectos individuales. Si estimamos β y η conjuntamente por mínimos cuadrados ordinarios (MCO), utilizando los resultados de la regresión particionada, el estimador de β se puede escribir de la forma siguiente

$$\hat{\beta} = (X'\bar{Q}X)^{-1}X'\bar{Q}y = (\tilde{X}'\tilde{X})^{-1}\tilde{X}'\tilde{y} \qquad [4]$$

en donde $\bar{Q} = I_{NT} - C(C'C)^{-1}C' = I_N \otimes Q$ con $Q = I_T - \iota\iota'/T$, $\tilde{X} = \bar{Q}X$ y $\tilde{y} = \bar{Q}y$. Nótese que $\bar{Q}$ es una matriz simétrica e idempotente lo que justifica la segunda igualdad. Este resultado es muy útil ya que es fácil comprobar que los elementos de $\tilde{X}$ e $\tilde{y}$ son desviaciones con respecto a las medias temporales de las variables originales:

$$\tilde{x}_{it} = x_{it} - \frac{1}{T}\sum_{s=1}^{T} x_{is} \qquad [5]$$

Por tanto se puede obtener $\hat{\beta}$ sin tener que calcular simultáneamente $\hat{\eta}$, sin más que calcular la regresión MCO de $\tilde{y}$ sobre $\tilde{x}$. En la práctica amenudo ésta es la única forma de calcular $\hat{\beta}$ si el número de individuos en la muestra (y por tanto la dimensión de η) es grande. A $\hat{\beta}$ se le llama estimador de covarianza o «intra-grupos».

La interpretación de $\hat{\beta}$ como MCO en el modelo en desviaciones es también muy útil desde el punto de vista teórico ya que ayuda a clarificar sus propiedades cuando T es fijo. En primer lugar, la consistencia de $\hat{\beta}$ no depende de la especificación de η_i porque los efectos son siempre eliminados por la transformación. En segundo lugar, $\hat{\beta}$ sólo es consistente (para T fijo y $N_{\to\infty}$) si $E[\tilde{x}_{it}\,\tilde{v}_{it}] = 0$ lo que requiere la *exogeneidad estricta* de x_{it} con respecto a v_{it} en el sentido de que $E[x_{it}v_{is}] = 0$ para todo t, $s = 1, ..., T$ (en la Sección 3 se discuten modelos con variables predeterminadas).

En el modelo con errores compuestos tradicional se supone que [1] es un modelo de regresión en el que el término de perturbación u_{it} tiene la forma particular dada en [2] con $E[\eta_i] = E[v_{it}] = 0$ y $E[x_{it}\eta_i] = E[x_{it}v_{is}] = 0$. En estas circunstancias, la regresión MCO de y_{it} sobre x_{it} («en niveles») proporciona un estimador consistente de β (ya que $E[x_{it}u_{it}] = 0$). Es importante resaltar que si estos supuestos son ciertos no necesitamos un panel para identificar β: un único corte transversal ($T = 1$) basta para obtener estimaciones consistentes. De otro lado, la especificación [2] genera autocorrelación en u_{it} incluso si no la hay en v_{it}. Obsérvese que con $E[v_{it}v_{is}] = 0$ tenemos $E[u_{it}u_{is}] = \mathrm{Var}[\eta_i]$ para todo $s \neq t$. Esta circunstancia sugiere que se puede conseguir un estimador más eficiente que MCO en niveles utilizando mínimos cuadrados generalizados (MCG). Hausman y Taylor (1981) muestran que cuando $\eta_i \sim iid[0,\sigma^2_\eta]$ y $v_{it} \sim iid[0,\sigma^2]$ una transformación MCG de [1] viene dada por

$$y_{it} - (1-\theta)\bar{y}_i = \beta[x_{it} - (1-\theta)\bar{x}_i] + \theta\eta_i + [v_{it} - (1-\theta)\bar{v}_i] \qquad [6]$$

en donde $\theta = [\sigma^2/[\sigma^2 + T\sigma^2_\eta]]^{1/2}$ y $\bar{y}_i$, $\bar{x}_i$ son las medias temporales de las varia bles. Para que el estimador MCG sea operativo se necesita un estimador previo de θ que a su vez requiere estimadores de σ^2 y σ^2_η. Estos últimos se pueden obtener de la siguiente forma (véase Hausman y Taylor, 1981, pág. 1384):

$$\hat{\sigma}^2 = \frac{1}{N(T-1)}\sum_{i=1}^{N}\sum_{t=1}^{T}[\tilde{y}_{it} - \hat{\beta}'\tilde{x}_{it}]^2$$

$$\hat{\sigma}^2_\eta = \frac{1}{N}\sum_{i=1}^{N}[\bar{y}_i - \hat{\beta}'\bar{x}_i]^2 - \frac{1}{T}\hat{\sigma}^2 \qquad [7]$$

En ambos casos, $\hat{\beta}$ representa el estimador intra-grupos. Este método de MCG se conoce como el estimador de Balestra y Nerlove (*cf.* Balestra y Nerlove, 1966).

En la práctica las estimaciones en niveles (MCO o MCG) y las estimaciones en desviaciones suelen proporcionar resultados muy distintos. Esto es en general una indicación de que hay diferencias inobservables entre individuos que sesgan las estimaciones en niveles. Por otra parte, en paneles en los que la variación transversal es grande y la variación temporal es pequeña (normalmente paneles de familias) las estimaciones intragrupos son mucho más imprecisas que las estimaciones en niveles. En algunos casos hasta el punto de que las estimaciones intragrupos son inservibles. Si este es el caso, todo lo que se puede decir es que la muestra no contiene suficiente información para controlar características permanentes inobservables.

Es conveniente resaltar que incluso en los casos en los que las estimaciones en niveles son consistentes, el método de Balestra y Nerlove sólo es eficiente si los errores son homocedásticos y no hay autocorrelación en los v_{it}. Estos supuestos son particularmente restrictivos con microdatos y no tiene porque esperarse que se cumplan. Una alternativa es utilizar MCO (tanto en niveles como en desviaciones) pero calcular los errores estándar y contrastes de hipótesis de forma que sigan siendo válidos cuando hay heterocedasticidad y autocorrelación utilizando fórmulas del tipo de White (1980). Por ejemplo, la varianza de los estimadores MCO en niveles se puede estimar usando.

$$\text{Var}\,[\hat{\beta}_{\text{MCO}}] = (X'X)^{-1}\,[\textstyle\sum_{i=1}^{N} \sum_{t=1}^{T} \sum_{s=1}^{T} \hat{u}_{it}\,\hat{u}_{is}\,x_{it}\,x'_{is}]\,(X'X)^{-1} \qquad [8]$$

en donde $\hat{u}_{it}$ son residuos MCO en niveles. Una fórmula similar se puede utilizar para el estimador intra-grupos reemplazando niveles por desviaciones con respecto a las medias en [8].

Otro aspecto interesante es que si se supone que los errores son homocedásticos en el corte transversal, al ser T fijo se pueden obtener estimadores MCG tanto en niveles como en desviaciones que son asintóticamente eficientes para cualquier forma de autocorrelación en los errores. En niveles utilizaríamos

$$\hat{\beta}_{\text{MCG}} = (X'[\text{I}_N \otimes \hat{\Omega}^{-1}]X]^{-1}\,X'[\text{I}_N \otimes \hat{\Omega}^{-1}]y \qquad [9]$$

estimando los elementos de la matriz de autocovarianzas Ω mediante:

$$\hat{\omega}_{ts} = \frac{1}{N}\sum_{i=1}^{N} \hat{u}_{it}\,\hat{u}_{is} \qquad [10]$$

$\hat{\beta}_{\text{MCG}}$ es como mínimo tan eficiente en muestras grandes como el estimador de Balestra y Nerlove, y es más eficiente cuando los errores transitorios v_{it} están autocorrelacionados. Un estimador del tipo [9] no se puede utilizar directamente en desviaciones porque en este caso $\hat{\Omega}$ tiene rango $(T-1)$ y no se puede invertir. Una solución es utilizar una inversa generalizada (*cf.* Kiefer, 1980) pero es más simple sustituir las T desviaciones para cada individuo por las

$(T-1)$ primeras diferencias y proceder como en [9] lo que proporciona un estimador con las mismas propiedades que el de Kiefer.

Aunque el modelo de efectos aleatorios tradicional supone que η_i es independiente de x_{it}, es posible especificar un modelo en el que los efectos son aleatorios y están correlacionados con las variables explicativas. Por ejemplo, con $T = 2$ supongamos que la distribución condicional de η_i dados x_{i1} y x_{i2} satisface

$$E[\eta_i \mid x_{i1}, x_{i2}] = \lambda_1' x_{i1} + \lambda_2' x_{i2} \qquad [11]$$

$$\mathrm{Var}[\eta_i \mid x_{i1}, x_{i2}] = \sigma_\eta^2 \qquad [12]$$

Por tanto, $\eta_i = \lambda_1' x_{i1} + \lambda_2' x_{i2} + \varepsilon_i$ en donde ε_i está incorrelacionado con x_{i1} y x_{i2} (en el modelo tradicional de efectos aleatorios $\lambda_1 = \lambda_2 = 0$). Sustituyendo en [1] se obtiene

$$\begin{aligned} y_{i1} &= [\beta + \lambda_1]' x_{i1} + \lambda_2' x_{i2} + \varepsilon_i + v_{i1} \\ y_{i2} &= \lambda_1' x_{i1} + [\beta + \lambda_2]' x_{i2} + \varepsilon_i + v_{i2} \end{aligned} \qquad [13]$$

Una representación equivalente de este sistema es

$$\begin{aligned} y_{i1} &= \lambda_1^{*\prime} x_{i1} + \lambda_2' x_{i2} + [\varepsilon_i + v_{i1}] \\ y_{i2} - y_{i1} &= \beta' [x_{i2} - x_{i1}] + [v_{i2} - v_{i1}] \end{aligned} \qquad [14]$$

en donde $\lambda_1^* = \beta + \lambda_1$. Nótese que la primera ecuación no presenta restricciones. Esto es, toda la información en el modelo acerca de β se encuentra en la segunda ecuación. El estimador óptimo en este caso (el estimador de distancia mínima o el estimador normal máximo verosímil) es el coeficiente de regresión MCO de $(y_{i2} - y_{i1})$ sobre $(x_{i2} - x_{i1})$, que también es el estimador intra-grupos de este modelo. El modelo [13] es un ejemplo de los considerados por Chamberlain (1982 y 1984).

Obsérvese que hemos obtenido el estimador intra-grupos de β por tres caminos distintos. Primero, estimando conjuntamente β y los N efectos individuales (inferencia conjunta). Segundo, transformando la ecuación [1] en desviaciones para eliminar los efectos (inferencia condicional). Tercero, especificando una distribución de probabilidad para los efectos que tiene en cuenta la dependencia entre η y x (inferencia incondicional). El modelo lineal con variables exógenas en sentido estricto es un caso especial en el que los tres principios dan lugar al mismo resultado. Sin embargo, esto no es así en modelos más complejos, como veremos más adelante.

Finalmente pasamos a considerar el problema de contrastar si los efectos están o no correlacionados con x_{it} (esto es, si los estimadores en niveles y en desviaciones estiman o no los mismos parámetros subyacentes). El contraste tradicional propuesto por Hausman (1978) se basa en la comparación directa entre el estimador intra-grupos $\hat{\beta}$ y el estimador de Balestra y Nerlove $\tilde{\beta}_{BN}$.

$$h = [\tilde{\beta}_{BN} - \hat{\beta}]' [\mathrm{Var}(\hat{\beta}) - \mathrm{Var}[\tilde{\beta}_{BN}]]^{-1} [\tilde{\beta}_{BN} - \hat{\beta}] \qquad [15]$$

Bajo la hipótesis nula de efectos incorrelacionados $\text{plim}_{N\to\infty}[\tilde{\beta}_{BN} - \hat{\beta}] = 0$ y h se distribuye asintóticamente como una *ji* – cuadrado. Sin embargo, este procedimiento sólo es válido en las condiciones en que $\tilde{\beta}_{BN}$ es eficiente relativo a $\hat{\beta}$ bajo la hipótesis nula. En la práctica con errores posiblemente heterocedásticos y autocorrelacionados estas condiciones no se cumplen. Un procedimiento alternativo que no sufre de estos problemas y es muy fácil de aplicar consiste en formar un sistema de ecuaciones ampliado combinando las ecuaciones en niveles y las ecuaciones en diferencias para a continuación contrastar dentro del sistema la igualdad de los coeficientes en las variables en niveles y en diferencias. Por ejemplo, de nuevo con $T = 2$ tenemos:

$$\begin{bmatrix} y_{i1} \\ y_{i2} \\ \Delta y_{i2} \end{bmatrix} = \begin{bmatrix} x_{i1} \\ x_{i2} \\ \Delta x_{i2} \end{bmatrix} \beta + \begin{bmatrix} 0 \\ 0 \\ \Delta x_{i2} \end{bmatrix} \alpha + \begin{bmatrix} \nu_{i1} \\ \nu_{i2} \\ \nu_{i3} \end{bmatrix} \quad (i = 1, \dots, N) \qquad [16]$$

β y α se pueden estimar por MCO y calcular estimaciones robustas de sus varianzas. Un test de Wald convencional de la hipótesis $\alpha = 0$ (o de la F) es un contraste de la hipótesis de no correlación de los efectos.

3. Modelos dinámicos

3.1. Inconsistencia del estimador intra-grupos

La especificación más simple con la que podemos analizar los problemas de estimación que surgen con los modelos dinámicos de datos de panel es una autorregresión de primer orden con efectos individuales:

$$y_{it} = \alpha y_{i(t-1)} + \eta_i + v_{it} \, . \qquad [17]$$

El estimador intra-grupos $\hat{\alpha}_{WG}$ es MCO en la ecuación transformada

$$y_{it} - \bar{y}_i = \alpha[y_{i(t-1)} - \bar{y}_{i(-1)}] + [v_{it} - \bar{v}_i] \qquad [18]$$

en donde $\bar{y}_i = T^{-1}\sum_{t=2}^{T} y_{it}$, $\bar{y}_{i(-1)} = T^{-1}\sum_{t=1}^{(T-1)} y_{it}$ y $\bar{v}_i = T^{-1}\sum_{t=2}^{T} v_{it}$. Sin embargo, $[y_{i(t-1)} - \bar{y}_{i(-1)}]$ y $[v_{it} - \bar{v}_i]$ están correlacionados y como consecuencia $\hat{\alpha}_{WG}$ estará sesgado incluso para valores grandes de N cuando T es pequeño. El sesgo asintótico de $\hat{\alpha}_{WG}$ obtenido por Nickell (1981) es

$$\text{plim}_{N\to\infty}[\hat{\alpha}_{WG} - \alpha] = -\frac{(1+\alpha)\, h(\alpha,T)}{(T-1)} \left[1 - \frac{2\alpha h(\alpha,T)}{(T-1)(1-\alpha)} \right]^{-1} \qquad [19]$$

en donde $h(\alpha,T) = 1 - T^{-1}(1-\alpha^T)/(1-\alpha)$. Este resultado supone que los η_i son constantes, el modelo es estacionario y v_{it} es ruido blanco. El Cuadro 1 presenta el valor del sesgo para varios valores de α y T. Como se puede ver, el sesgo disminuye a medida que T crece (es de orden $1/T$). Sin embargo, incluso para $T = 15$, que es un valor grande para un micropanel, el sesgo es de un 22 por 100 con $\alpha = 0.5$. Nótese que para valores de α positivos el sesgo es siempre negativo y que el sesgo no tiende a cero a medida que $\alpha \to 0$.

CUADRO 1
Sesgos asintóticos del estimador intra-grupos en un modelo autorregresivo

$T \backslash \alpha$	0,05	0,50	0,95
2	−0,52	−0,75	−0,97
3	−0,35	−0,54	−0,73
10	−0,11	−0,16	−0,26
15	−0,07	−0,11	−0,17

Un procedimiento alternativo para eliminar los efectos inobservables η_i es tomar primeras diferencias:

$$y_{it} - y_{i(t-1)} = \alpha[y_{i(t-1)} - y_{i(t-2)}] + [v_{it} - v_{i(t-1)}] \quad [20]$$

El estimador MCO en esta ecuación $\hat{\alpha}_d$ también está sesgado, pero en este caso el sesgo no tiende a cero a medida que $T_{\to\infty}$:

$$\text{plim}_{N\to\infty}\,[\hat{\alpha}_d - \alpha] = -\frac{(1+\alpha)}{2} \quad [21]$$

Este sesgo coincide con el del estimador intra-grupos cuando $T = 2$. Finalmente, el sesgo asintótico de MCO en el modelo en niveles $\hat{\alpha}_\ell$ viene dado por

$$\text{plim}_{N\to\infty}\,[\hat{\alpha}_\ell - \alpha] = \frac{\lambda}{\lambda(1-\alpha)^{-1} + (1+\alpha)^{-1}} \quad [22]$$

en donde $\lambda = \sigma^2_\eta/\sigma^2$ y $\sigma^2_\eta = \lim_{N\to\infty} N^{-1}\sum_{i=1}^N \eta_i^2$. Nótese que los sesgos de $\hat{\alpha}_{WG}$ y $\hat{\alpha}_d$ no dependen de los η_i (en realidad no se ven afectados si $\eta_i \equiv 0$) y son negativos. Por el contrario el sesgo de $\hat{\alpha}_\ell$ depende de σ^2_η —la dispersión de los η_i's en la población— y es positivo para $\alpha > 0$.

3.2. *Estimación consistente del modelo autorregresivo*

Anderson y Hsiao (1981) estudiaron la estimación del coeficiente autorregresivo α en [17] por el método de máxima verosimilitud normal (MV). Su conclusión principal es que las propiedades de los estimadores resultantes son muy sensibles a las condiciones iniciales (esto es, la especificación de un modelo para y_{i1}). Normalmente, el inicio del período muestral no coincide con el inicio del proceso dinámico y en cualquier caso no se suele disponer de información *a priori* sobre las condiciones iniciales. Por ello, es deseable utilizar estimadores cuya consistencia no depende de una formulación particular de las condiciones iniciales. Se puede obtener un estimador MV de estas características considerando la función de densidad condicional de las observaciones dado y_{i1}. En concreto, si $T = 3$ la densidad conjunta de y_{i1}, y_{i2}, y_{i3} se puede escribir

$$f[y_{i1}, y_{i2}, y_{i3}] = f_c[y_{i2}, y_{i3} | y_{i1}] f_m[y_{i1}] \quad [23]$$

Si no se imponen restricciones en f_m, el estimador MV de α se puede basar en f_c. Suponiendo que $E[\eta_i|y_{i1}] = \lambda y_{i1}$, $\text{Var}[\eta_i|y_{i1}] = \sigma_\eta^2$, $E[v_{it}^2] = \sigma_t^2$ y $E[v_{i2}v_{i3}] = 0$ tenemos

$$\begin{bmatrix} y_{i2} \\ y_{i3} \end{bmatrix} \Big| y_{i1} \sim N \left[\begin{bmatrix} \pi_2 \\ \pi_3 \end{bmatrix} y_{i1}, \begin{bmatrix} \omega_{22} & \omega_{23} \\ \omega_{32} & \omega_{33} \end{bmatrix} \right] \quad [24]$$

con $\pi_2 = \alpha + \lambda$, $\pi_3 = \alpha(\alpha+\lambda)+\lambda$, $\omega_{22} = \sigma_\eta^2 + \sigma_2^2$, $\omega_{33} = (1+\alpha)^2\sigma_\eta^2 + \sigma_3^2 + \alpha^2 \sigma_2^2$ y $\omega_{23} = (1+\alpha)\sigma_\eta^2 + \alpha\,\sigma_2^2$. Obsérvese que el modelo está exactamente identificado y, por tanto, hay una única solución para α en términos de π_2 y π_3:

$$\alpha = \frac{\pi_3 - \pi_2}{\pi_2 - 1} \quad [25]$$

Los estimadores MV de π_2 y π_3 son los coeficientes de regresión simple de y_{i2} e y_{i3} sobre y_{i1}, respectivamente, $\hat{\pi}_t = [\sum_{i=1}^N y_{i1}\, y_{it}]/[\sum_{i=1}^N y_{i1}^2]$. Por tanto el estimador MV de α es

$$\hat{\alpha} = \frac{\hat{\pi}_3 - \hat{\pi}_2}{\hat{\pi}_2 - 1} = \frac{\sum_{i=1}^N y_{i1}[y_{i3} - y_{i2}]}{\sum_{i=1}^N y_{i1}[y_{i2} - y_{i1}]}$$

$\hat{\alpha}$ tiene una interpretación alternativa muy interesante: es el estimador de variables instrumentales (VI) en la ecuación en primeras diferencias

$$y_{i3} - y_{i2} = \alpha[y_{i2} - y_{i1}] + [v_{i3} - v_{i2}]$$

que utiliza y_{i1} como instrumento para $[y_{i2} - y_{i1}]$. El procedimiento funciona porque y_{i1} está correlacionado con $[y_{i2} - y_{i1}]$ pero no con $[v_{i3} - v_{i2}]$ cuando v_{it} no está autocorrelacionado. La misma técnica se puede utilizar con $T > 3$ sin más que disponer en vectores columna las diferencias e instrumentos disponibles obteniendo

$$\hat{\alpha}_{AH} = \frac{\sum_{i=1}^N \sum_{t=3}^T y_{i(t-2)}[y_{it} - y_{i(t-1)}]}{\sum_{i=1}^N \sum_{t=3}^T y_{i(t-2)}[y_{i(t-1)} - y_{i(t-2)}]} \quad [26]$$

A $\hat{\alpha}_{AH}$ se le denomina estimador de Anderson y Hsiao (AH). Estos autores también propusieron una variante de $\hat{\alpha}_{AH}$ que utiliza $\Delta y_{i(t-2)}$ en lugar de $y_{i(t-2)}$ como instrumento en cuyo caso las sumas temporales en [26] van desde 4 hasta T. $\hat{\alpha}_{AH}$ es consistente para T fijo y $N_{\to\infty}$, para $T_{\to\infty}$ y N fijo, y también cuando ambos T y N tienden a infinito. Sin embargo la equivalencia entre $\hat{\alpha}_{AH}$ y el estimador condicional MV sólo se da para $T = 3$. Para valores $T > 3$, la distribución condicional f_c no sólo presenta restricciones en los coeficientes π_t de la media sino también en las varianzas y covarianzas ω_{ts} por lo que resulta un procedimiento poco atractivo.

Por el contrario, una generalización muy útil del procedimiento de Anderson y Hsiao cuando T es pequeño relativo a N es el siguiente método de variables instrumentales. Como se puede ver en el Cuadro 2, y_{i1} es el único instru-

CUADRO 2
Posibles instrumentos en un modelo autorregresivo en primeras diferencias

Ecuación	Instrumentos válidos
$\Delta y_{i3} = \alpha \Delta y_{i2} + \Delta v_{i3}$	y_{i1}
$\Delta y_{i4} = \alpha \Delta y_{i3} + \Delta v_{i4}$	y_{i1}, y_{i2}
$\vdots$	$\vdots$
$\Delta y_{iT} = \alpha \Delta y_{i(T-1)} + \Delta v_{iT}$	$y_{i1}, y_{i2}, \ldots, y_{i(T-2)}$

mento válido en la ecuación en primeras diferencias para el período 3; para la segunda ecuación tenemos dos instrumentos válidos, y así sucesivamente hasta la última ecuación para la que tenemos $(T-2)$ instrumentos. La estructura de este problema es pues la de un sistema de ecuaciones con instrumentos diferentes para distintas ecuaciones. Si definimos el vector $(T-2)\times 1$ $\bar{v}_i = [\Delta v_{i3}, \ldots, \Delta v_{iT}]'$ y la matriz $(T-2)\times m$, con $m=(T-1)(T-2)/2$:

$$Z_i = \begin{bmatrix} y_{i1} & & & \\ & y_{i1}, y_{i2} & & O \\ & O & \ddots & \\ & & & y_{i1} y_{i2} \ldots y_{i(T-2)} \end{bmatrix}, \qquad [27]$$

el contenido del Cuadro 2 se puede expresar mediante las restricciones de momentos $E[Z_i'\bar{v}_i] = 0$. Nótese que si los errores v_{it} se distribuyen independientemente con varianza constante σ^2, $E[\bar{v}_i\bar{v}_i'] = \sigma^2 H$ en donde H es una matriz simétrica $(T-2)\times(T-2)$ cuyos elementos son iguales a 2 en la diagonal principal, a -1 en las primeras subdiagonales, y a cero en el resto. En este caso se puede comprobar que la matriz de varianzas y covarianzas de $Z_i'\bar{v}_i$ es igual a

$$E[Z_i'\bar{v}_i\bar{v}_i'Z_i] = \sigma^2 E[Z_i'HZ_i] \qquad [28]$$

y que el estimador óptimo de α basado en el Cuadro 2 es

$$\hat{\alpha}_1 = \frac{\sum_i \bar{y}'_{i(-1)} Z_i [\sum_i Z_i'HZ_i]^{-1} \sum_i Z_i'\bar{y}_i}{\sum_i \bar{y}'_{i(-1)} Z_i [\sum_i Z_i'HZ_i]^{-1} \sum_i Z_i'\bar{y}_{i(-1)}} \qquad [29]$$

en donde $\bar{y}_i = [\Delta y_{i3} \ldots \Delta y_{iT}]'$ y $\bar{y}_{i(-1)} = [\Delta y_{i2} \ldots \Delta y_{i(T-1)}]'$. Estimadores de este tipo han sido estudiados por Arellano y Bond (1988a) y por Holtz-Eakin, Newey y Rosen (1988). La forma más útil de interpretarlos es como estimadores del método generalizado de momentos (MGM). Esto es, estimadores que minimizan la discrepancia entre las restricciones de momentos muestrales $\sum_i Z_i'\bar{v}_i/N$ y sus valores de cero en la población (*cf.* Hansen, 1982).

Antes de abandonar esta discusión vale la pena considerar otros dos puntos. En primer lugar, los errores v_{it} no tienen porque ser homocedásticos en el corte transversal o en la dimensión temporal, en cuyo caso la igualdad [28] no se cumple; $\hat{\alpha}_1$ sigue siendo consistente pero se puede obtener un estimador más eficiente si reemplazamos $\Sigma_i Z_i' H Z_i$ en [29] por una estimación más general de la varianza de $Z_i'\bar{v}_i$. En concreto por $\Sigma_i Z_i'\bar{v}_i\bar{v}_i'Z_i$ en donde $\bar{v}_i$ son residuos calculados a partir de $\hat{\alpha}_1$. Al estimador resultante, $\hat{\alpha}_2$, le denominamos «MGM en dos etapas» mientras que $\hat{\alpha}_1$ es «MGM en una etapa».

El segundo punto es preguntarse si las restricciones de momentos $E(Z_i'\bar{v}_i) = 0$ obtenidas a partir del modelo en primeras diferencias son una forma óptima de expresar la información acerca de α, o si por el contrario puede encontrarse una transformación alternativa para eliminar η_i que dé lugar a estimadores MGM más eficientes. La pregunta surge porque las primeras diferencias introducen autocorrelación en los errores y puede sospecharse que ello induzca a ineficiencias. Si una vez eliminados los efectos individuales por medio de primeras diferencias, derivamos la transformación MCG de la ecuación para eliminar la autocorrelación, se obtiene la transformación de desviaciones ortogonales propuesta en Arellano (1988):

$$c_t\left[y_{it} - \frac{1}{(T-t-1)}\,[y_{i(t+1)} + \ldots + y_{iT}]\right]$$
$$= \alpha\left[y_{i(t-1)} - \frac{1}{(T-t-1)}\,[y_{it} + \ldots + y_{i(T-1)}]\right]c_t + v_{it}^* \qquad [30]$$

en donde $c_t^2 = (T-t-1)/(T-t)$ y v_{it}^* es el término de perturbación transformado en la misma forma que la variable dependiente. Se puede comprobar que v_{it}^* es homocedástico y no autocorrelacionado si estas son las propiedades de v_{it}. También se puede comprobar que MCO en [30] es el estimador intragrupos. La ventaja de las desviaciones ortogonales hacia adelante es que los instrumentos en Z_i siguen siendo válidos a diferencia de lo que ocurre con las desviaciones del tipo $[v_{it} - \bar{v}_i]$. Sin embargo, se puede demostrar que MGM en una y dos etapas basados en las restricciones $E[Z_i'v_i^*] = 0$ son numéricamente equivalentes a $\hat{\alpha}_1$ y $\hat{\alpha}_2$. En realidad, cualquier transformación que elimine η_i y no destruya la validez de la matriz de instrumentos Z_i da lugar a los mismos estimadores. Sin embargo, en la práctica puede ocurrir que el número total de instrumentos posibles se considere excesivo para valores particulares de T y N, eliminándose de Z_i las columnas menos relevantes. En este caso, los estimadores en diferencias y en desviaciones ortogonales difieren.

3.3. Modelos con variables exógenas

A continuación pasamos a considerar modelos de regresión dinámicos de la forma

$$y_{it} = \alpha y_{i(t-1)} + \beta x_{it} + \eta_i + v_{it} \qquad [31]$$

en donde x_{it} es escalar para simplificar la presentación. Como se verá, el marco anterior consistente en abordar la estimación en términos de un sistema de ecuaciones con instrumentos diferentes para distintas ecuaciones, sigue siendo útil en este caso.

La definición de los parámetros α y β en [31] se ha de completar con supuestos acerca de x_{it} y v_{it}. En esta sección hacemos el supuesto de que x_{it} está correlacionado con η_i y que no se dispone de instrumentos externos. Si v_{it} no está autocorrelacionado y x_{it} es una variable *predeterminada,* en el sentido de que $E[x_{it}v_{is}] = 0$ sólo si $s \geq t$, la estimación MGM de α y β es similar a la del caso autorregresivo: los instrumentos válidos para la t-ésima ecuación en primeras diferencias o desviaciones ortogonales vienen dados por el vector $z_{it} = [y_{i1} \ldots y_{i(t-2)} x_{i1} \ldots x_{i(t-1)}]'$. El estimador MGM en dos etapas se puede escribir

$$\hat{\delta} = [W^{*\prime} Z A_N Z' W^*]^{-1} W^{*\prime} Z A_N Z' y^* \qquad [32]$$

en donde $\delta = (\alpha\beta)'$, $w_{it} = [y_{i(t-1)} x_{it}]'$, $Z_i = \text{diag}[z'_{it}]$, $y = [y_{12} \ldots y_{1T} \ldots y_{N2} \ldots y_{NT}]'$, $W = [w_{i2} \ldots w_{iT} \ldots w_{N2} \ldots w_{NT}]'$, $Z = [Z'_1 \ldots Z'_N]'$, W^* e y^* representan primeras diferencias o desviaciones ortogonales de las variables originales, y $A_N = [\Sigma_{i=1}^{N} Z'_i \hat{v}^*_i \hat{v}^{*\prime}_i Z_i]^{-1}$; $\hat{v}^*_i$ es el vector $(T-2)\times 1$ de residuos MGM en una etapa correspondiente al i-ésimo individuo.

Obsérvese que el modelo puede acomodar errores autorregresivos y de media móvil. En primer lugar, si v_{it} sigue un proceso autorregresivo

$$v_{it} = \rho v_{i(t-1)} + \varepsilon_{it} \qquad [33]$$

la ecuación [31] se puede combinar con [33] obteniendo

$$y_{it} = \alpha_1 y_{i(t-1)} + \alpha_2 y_{i(t-2)} + \beta x_{it} + \beta_1 x_{i(t-1)} + \eta^*_i + \varepsilon_{it} \qquad [34]$$

con $\alpha_1 = \alpha + \rho$, $\alpha_2 = -\alpha\rho$, $\beta_1 = -\beta\rho$ y $\eta^*_i = (1-\rho)\eta_i$. Los coeficientes de [34] se pueden estimar por MGM en dos etapas. A continuación se pueden contrastar las restricciones de «factores comunes» utilizando un test de Wald (véase Sargan, 1980) y en caso de ser aceptadas se pueden obtener estimaciones de los coeficientes restringidos por mínimos cuadrados asintóticos (*cf.* Gourieroux, Monfort y Trognon, 1985).

En segundo lugar, si v_{it} es MA(1) en el sentido de que $E[v_{it}v_{i(t-1)}] \neq 0$ aunque $E[v_{it}v_{i(t-k)}] = 0$ para $k > 1$, se puede obtener un estimador consistente de δ utilizando en [32] las columnas de Z que continúan siendo instrumentos válidos con errores MA(1). Esto es, para la t-ésima ecuación transformada tendríamos $z_{it} = [y_{i1} \ldots y_{i(t-3)} x_{i1} \ldots x_{i(t-2)}]'$.

Sin embargo, si v_{it} es MA(q) con $T < q + 3$ (en cuyo caso la matriz de autocovarianzas no tiene restricciones), el modelo con variables predeterminadas no está identificado. En esta situación se puede conseguir la identificación de α y β suponiendo que x_{it} es *exógena en sentido estricto,* esto es, $E(x_{it}v_{is}) = 0$ para todo t y s. Con este supuesto, las x_{it} de todos los períodos son instrumentos válidos en todas las ecuaciones de tal manera que $z_{it} = [x_{it} \ldots x_{iT}]'$ para todo t. Con esta elección de z_{it} el sistema de $(T-2)$ ecuaciones transformadas se puede ver como un sistema incompleto de ecuaciones simultáneas corriente en el que $[x_{i1} \ldots x_{iT}]$ son las variables exógenas. Por otra parta [32] se convierte en el estimador generalizado de mínimos cuadrados en tres etapas (MC3E) propuesto por Chamberlain (1982)[2]. Bhargava y Sargan (1983) consideran una versión en niveles de este mismo modelo al asumir ausencia de correlación entre x_{it} y η_i, y utilizan el método de máxima verosimilitud en lugar de MC3E.

Si x_{it} es exógena en sentido estricto y hay restricciones en la matriz de autocovarianzas (esto es, v_{it} no está autocorrelacionado o si lo está sigue un proceso definido por un pequeño número de parámetros), esta información adicional se puede explotar para obtener estimadores más eficientes que MC3E. Obviamente una posibilidad es continuar utilizando estimadores de la forma [32] ampliando el conjunto de instrumentos en cada ecuación con los retardos válidos de y_{it} lo cual es probable que capture la información más relevante contenida en las restricciones de autocovarianzas. Por ejemplo, si v_{it} es ruido blanco utilizaríamos $z_{it} = [y_{i1} \ldots y_{i(t-2)} x_{i1} \ldots x_{iT}]'$. Otra posibilidad es contrastar y explotar exhaustivamente las restricciones en las covarianzas con un procedimiento MV (como en el caso de Bhargava y Sargan, 1983) o utilizando contrastes de la *ji*-cuadrado y estimadores MCG para sistemas triangulares (*cf.* Arellano, 1989a y 1990). La desventaja del método MV para imponer las restricciones de autocovarianzas es que puede resultar en una pérdida de eficiencia si los errores no son normales (sobre este punto véase Arellano, 1989b).

Los estimadores de Anderson y Hsiao también se pueden utilizar para estimar la ecuación [31] (*cf.* Anderson y Hsiao, 1982). La versión que utiliza $\Delta y_{i(t-2)}$ como instrumento y supone la exogeneidad estricta de x_{it} viene dada por

$$\hat{\delta}_{AHd} = [\textstyle\sum_{i=1}^{N} \sum_{t=4}^{T} \Delta z_{it} \, \Delta w_{it}']^{-1} \sum_{i=1}^{N} \sum_{t=4}^{T} \Delta z_{it} \, \Delta y_{it} \qquad [35]$$

con $z_{it} = [y_{i(t-2)} x_{it}]'$. Por el contrario, el estimador alternativo $\hat{\delta}_{AH\ell}$ utiliza $z_{it}^{*} = [y_{i(t-2)} \, \Delta x_{it}]'$ en lugar de Δz_{it} en [35] y la suma de productos cruzados va desde $t = 3$ a T. En Arellano (1989c) se prueba que hay combinaciones de valores entre 0 y 1 de α y del coeficiente de autocorrelación de primer orden de x_{it} para los cuales no hay correlación entre $\Delta y_{i(t-1)}$ y $\Delta y_{i(t-2)}$ en cuyo caso $\hat{\delta}_{AHd}$ no está identificado. Por otra parte, las comparaciones entre las varianzas asintóticas de $\hat{\delta}_{AHd}$ y $\hat{\delta}_{AH\ell}$ que se llevan a cabo en este trabajo muestran que la varianza de $\hat{\delta}_{AH\ell}$ suele ser mucho más pequeña. Estos resultados son consistentes con las simulaciones efectuadas en Arellano y Bond (1988a). Sin embargo, el resultado más importante de estas simulaciones es que las varianzas de los estimadores MGM son sistemáticamente más pequeñas que las de los estimadores AH. La magnitud de estas diferencias es tal que se puede decir que hay ganancias de eficiencia importantes en la práctica al usar MGM en lugar de AH, a parte de evitar posibles singularidades como en el caso de AHd.

[2] El estimador MC3E convencional sustituiría $A_N = (\sum_{i=1}^{N} Z_i' \hat{v}_i^* \hat{v}_i^{*\prime} Z_i)^{-1}$ en [32] por $A_N = (\sum_{i=1}^{N} Z_i' \hat{\Omega} Z_i)^{-1}$ con $\hat{\Omega} = N^{-1} \sum_{i=1}^{N} \hat{v}_i^* v_i^{*\prime}$. Ambos estimadores son asintóticamente equivalentes en ausencia de heterocedasticidad en el corte transversal, pero en general MC3E generalizado es más eficiente.

3.4. Contrastes de especificación

Los estimadores MGM que utilizan retardos como instrumentos bajo el supuesto de perturbaciones ruido blanco serían inconsistentes si de hecho los errores estuvieran autocorrelacionados, por ello es importante contrastar este supuesto. Si los errores son ruido blanco en niveles esperaremos autocorrelación de primer orden en los residuos en primeras diferencias pero no de segundo orden. Por tanto, podemos utilizar un contraste directo sobre la autocorrelación de segundo orden en los residuos de tipo de multiplicador de Lagrange. Se puede probar que bajo la hipótesis nula de no autocorrelación en v_{it}.

$$m_2 = \frac{1}{\hat{\omega}} \Sigma_{i=1}^{N} \Sigma_{t=4}^{T} \hat{v}_{it}^{*} \hat{v}_{i(t-2)}^{*} \quad \tilde{a} \quad N(0,1) \qquad [36]$$

en donde $\hat{v}_{it}^{*}$ representan en este caso residuos en primeras diferencias y, $\hat{\omega}^2$ es un estimador consistente de la varianza asintótica de $N^{-1/2} \Sigma_{i=1}^{N} \Sigma_{t=4}^{T} \hat{v}_{it}^{*} \hat{v}_{i(t-2)}^{*}$, (*cf.* Arellano y Bond, 1988a).

En la medida en que un estimador MGM del tipo [32] comporta restricciones sobre-identificadoras, éstas se pueden contrastar utilizando un estadístico de Sargan (véase Sargan, 1958 y 1988a; y Hansen, 1982): si la elección de A_N es óptima tenemos que bajo la hipótesis nula de validez de todos los instrumentos en Z:

$$s = [\Sigma_{i=1}^{N} \hat{v}_i^{*\prime} Z_i] A_N [\Sigma_{i=1}^{N} Z_i' \hat{v}_{it}^{*}] \quad \tilde{a} \quad \chi_r^2 \qquad [37]$$

en donde r es la diferencia entre el número de columnas en Z_i y el número de parámetros estimados. El estadístico s puede servir para detectar autocorrelación residual así como otras formas de errores de especificación.

Por último, discutimos el problema de contrastar la existencia de efectos individuales. El punto que queremos resaltar es que no se deben llevar a cabo estos contrastes utilizando residuos de modelos estimados en niveles ya que pueden tener poco o ningún poder para rechazar la hipótesis nula cuando ésta es falsa. El modelo autorregresivo [17] es un buen ejemplo para ver este argumento. Definamos $\alpha_\ell = \text{plim}_{N\to\infty} \hat{\alpha}_\ell$ En [22] tenemos una expresión para el sesgo asintótico $b = (\alpha_\ell - \alpha)$ que muestra que $\alpha_\ell = \alpha$ sólo si $\sigma_\eta^2 = 0$. Supongamos que H_0: $\sigma_\eta^2 = 0$ es la hipótesis nula. Bajo $H_0\, \hat{\alpha}_\ell$ es consistente y podríamos basar un contraste de H_0 en la ausencia de autocorrelación positiva en los residuos MCO. Bajo la hipótesis alternativa, los residuos MCO son analogías muestrales de

$$u_{i\ell t} = y_{it} - \alpha_\ell y_{i(t-1)} = (1-b)\eta_i + [v_{it} - bv_{i(t-1)}] - \alpha b y_{i(t-2)}$$

Por ejemplo, en el caso extremo en que $\alpha = 0$, utilizando $b = \lambda/(1+\lambda)$ se puede ver que la autocorrelación de primer orden en $u_{i\ell t}$ es

$$r_\ell = \frac{E[u_{i\ell t} u_{i\ell(t-1)}]}{E[u_{i\ell t}^2]} = \frac{-\lambda^2}{(1+\lambda)(1+2\lambda)} \qquad [38]$$

Nótese que como $\lambda \geq 0$ tenemos que $r_\ell \leq 0$. Obviamente, en este caso el test no tendría ningún poder. Sin embargo, es posible llevar a cabo un contraste útil de autocorrelación utilizando residuos en niveles construidos a partir de un estimador de α que sea consistente bajo la hipótesis alternativa.

3.5. *MGM en otros contextos: errores en las variables y simultaneidad*

Hasta ahora hemos mostrado la utilidad de abordar los problemas de estimación dinámica con micropaneles en términos de un sistema de ecuaciones transformadas con distintos instrumentos válidos para cada uno de los períodos. Este marco es igualmente útil para otros modelos con datos de paneles. Un caso importante son los modelos con errores en las variables estudiados por Griliches y Hausman (1986).

Como vimos al principio de esta sección, las estimaciones intra-grupos están negativamente sesgadas en modelos dinámicos. A menudo, también aparecen sesgos negativos en las estimaciones intra-grupos de modelos no autorregresivos. Estos sesgos se tienden a explicar por la presencia de errores de medida en las variables independientes que resultan aumentados por las desviaciones. En concreto, si el modelo verdadero es de la forma:

$$y_{it} = \beta x^*_{it} + \eta_i + \varepsilon_{it} \tag{39}$$

pero en lugar de x^*_{it} observamos x_{it} sujeto a un error de medida ξ_{it}:

$$x_{it} = x^*_{it} + \xi_{it} \tag{40}$$

en el modelo en términos de x_{it}

$$y_{it} = \beta x_{it} + \eta_i + v_{it} \tag{41}$$

el término de perturbación $v_{it} = \varepsilon_{it} - \beta\xi_{it}$ está negativamente correlacionado con x_{it}. Por tanto, el sesgo del estimador MCO de β en [41] tendrá un componente positivo debido a la correlación positiva entre x_{it} y η_i, y un componente negativo debido a la correlación negativa entre x_{it} y v_{it} (para ciertos valores de los parámetros podría darse el caso de que el sesgo se anulara). Aunque η_i se puede eliminar transformando [41] en primeras diferencias, desgraciadamente esta transformación dobla la varianza del error de medida (en el caso en que $E[\xi_{it}\xi_{i(t-1)}] = 0$). Sin embargo, como Griliches y Hausman ponen de manifiesto bajo el supuesto de que los errores de medida son ruido blanco o MA(q) con $q < T$, con datos de panel disponemos de instrumentos internos que permiten la identificación de β. Por ejemplo, con $T = 3$ y ξ_{it} ruido blanco, x_{i1} es un instrumento válido en la ecuación en primeras diferencias correspondiente a $t = 3$. De esta forma, dada una especificación particular del modelo, podemos contar el número de restricciones de momentos válidas y utilizar un estimador MGM del tipo presentado en [32] (véase Griliches y Hausman, 1986, para los detalles). El caso en que x_{it} es una variable $0 - 1$ presenta problemas específicos y ha sido estudiado por Chowdhury y Nickell (1985).

A continuación, consideremos el caso en que [41] representa una ecuación estructural en la que x_{it} es una variable endógena. En ciertas ocasiones puede ser necesario utilizar instrumentos externos para identificar β, pero si hay razones teóricas para interpretar v_{it} como una «sorpresa», los retardos de x_{it} e y_{it} serían instrumentos válidos en el modelo en diferencias. Un ejemplo interesante es la relación inversión $-$ q (entre la tasa de inversión (I/K) y el cociente entre el valor de mercado del capital y el coste de reemplazar el stock de capital (q)) estudiada por Hayashi e Inoue (1989). q es endógena en el corte transversal y además es difícil encontrar instrumentos externos cuya correlación con q pero no correlación con v_{it} se pueda justificar teóricamente. Sin embargo, la teoría predice que si v_{it} es ruido blanco los valores pasados y futuros de q estarán incorrelacionados con la perturbación contemporánea. En la notación de [41], quiere decir que en la ecuación en primeras diferencias para el período t el vector de instrumentos válidos es $z_{it} = [x_{i1} \ldots x_{i(t-2)}, x_{i(t+1)} \ldots x_{iT}]'$ que de nuevo se puede explotar óptimamente con un estimador del tipo [32]. En todos estos casos suele haber numerosas restricciones sobre-identificadoras y, por tanto, el test de Sargan proporciona un contraste general de la especificación.

En los modelos que hemos considerado hasta ahora se supone que los términos de perturbación son independientes en el corte transversal. Esto es, que las «sorpresas» que recibe un individuo son idiosincráticas o específicas a dicho individuo. Si hay razones para pensar que las sorpresas macroeconómicas son importantes, se supone que se pueden captar por las variables ficticias de tiempo que formarían parte del vector x. Sin embargo, este tratamiento sólo es válido si las variables macroeconómicas afectan a todos los individuos por igual. En el caso en que las respuestas individuales a las sorpresas macroeconómicas sean heterogéneas, una extensión interesante de la especificación de efectos individuales es la siguiente

$$y_{it} = \beta x_{it} + \psi_t \eta_i + v_{it} \qquad [42]$$

En este caso, η_i se puede eliminar transformando el modelo en pseudodiferencias de la forma $y_{it} - r_t y_{i(t-1)}$ con $r_t = \psi_t / \psi_{t-1}$:

$$y_{it} = r_t y_{i(t-1)} + \beta x_{it} - \beta r_t x_{i(t-1)} + [v_{it} - r_t v_{i(t-1)}] \qquad [43]$$

La transformación da lugar a un modelo que contiene retardos de las variables de la derecha e izquierda de la ecuación original con coeficientes que varían en el tiempo pero satisfacen restricciones de proporcionalidad. Obsérvese que los parámetros ψ_t no están identificados, sólo lo están los cocientes r_t. El vector β y los r_t se pueden estimar en [43] por MGM. Esta especificación ha sido estudiada en este contexto por Chamberlain (1984) y por Holtz-Eakin, Newey y Rosen (1988). En términos más generales, coincide con la formulación de la perturbación en el modelo MIMIC de Joreskog y Goldberger (1975).

4. Modelos con información en niveles

En el modelo básico de «efectos fijos» y en los modelos considerados en la Sección 3 se supone que todas las variables explicativas están potencialmente correlacionadas con los efectos individuales y por tanto sólo estimadores basados en desviaciones temporales de las observaciones originales pueden ser consistentes. Sin embargo si algunas de las variables explicativas no están correlacionadas con los efectos, los niveles de las variables contienen información acerca de los parámetros de interés cuya utilización daría lugar a estimadores más eficientes, que especialmente en los casos en que la variación temporal sea escasa puede resultar crucial en la práctica. Por otra parte, esta información en niveles puede ser suficiente para identificar coeficientes de variables explicativas que no varían a lo largo del tiempo y están correlacionadas con los efectos.

Empecemos considerando un modelo similar a [1] pero que contiene un vector $g \times 1$ de variables explicativas constantes en el tiempo z_i:

$$y_{it} = x'_{it}\beta + z'_i\gamma + \eta_i + v_{it} \qquad [44]$$

Si transformamos [44] en desviaciones no sólo eliminamos de la ecuación η_i sino también $z'_i\gamma$ por lo que únicamente β se puede estimar directamente. No obstante, el vector γ está identificado si z_i está incorrelacionado con η_i y v_{it} sean cuales sean las propiedades de x_{it} con relación a η_i. Por ejemplo, si $\hat{\beta}$ es el estimador intragrupos de β y $\mathrm{Cov}[x_{it}v_{is}] = 0$ para todo t, s, podemos estimar consistentemente γ por MCO en la ecuación

$$\bar{y}_i - \hat{\beta}'\bar{x}_i = z'_i\gamma + [\eta_i + \bar{v}_i - \bar{x}'_i(\hat{\beta} - \beta)] \qquad [45]$$

Por el contrario si z_i está correlacionado con η_i, en general γ no está identificado a menos de que se disponga de variables instrumentales externas incorrelacionadas con η_i pero correlacionadas con z_i. La situación cambia si suponemos que hay al menos tantas variables x_{it} incorrelacionadas con η_i como variables z_i correlacionadas con η_i. En este caso las medias temporales de las primeras se pueden utilizar como variables instrumentales para las segundas en [45]. Esta es la idea básica en los métodos de estimación desarrollados por Hausman y Taylor (1981). Bhargava y Sargan (1983) también consideran un modelo con correlación entre los regresores y los efectos en su estudio sobre modelos dinámicos con errores compuestos. Posteriormente, Amemiya y MaCurdy (1986) y Breusch, Mizon y Schmidt (1989) han propuesto conjuntos de instrumentos más amplios del modelo de Hausman y Taylor que dan lugar en general a aumentos de eficiencia asintótica.

En esta sección seguimos utilizando nuestra estrategia de estimación MGM consistente en presentar los modelos de panel como sistemas de ecuaciones transformadas con potencialmente distintos instrumentos válidos para cada una de las ecuaciones. Los estimadores de la Sección 3 estaban basados en la información «intra-grupos» representada por el sistema de $(T-1)$ ecuaciones en desviaciones ortogonales o en primeras diferencias. Puesto que ahora vamos a hacer supuestos que también nos permiten utilizar la información en niveles («entre-grupos»), es natural completar ese sistema con una ecuación en las medias de las variables. Con referencia al modelo [44] tenemos

$$\begin{aligned} y^*_{i1} &= x^{*\prime}_{i1}\beta + v^*_{i1} \\ &\vdots \\ y^*_{i(T-1)} &= x^{*\prime}_{i(T-1)}\beta + v^*_{i(t-1)} \\ \bar{y}_i &= \bar{x}'_i\beta + z'_i\gamma + [\eta_i + \bar{v}_i] \end{aligned} \qquad [46]$$

en donde y^*_{it}, x^*_{it} y v^*_{it} representan desviaciones ortogonales de las variables originales de la forma definida en [30]. Obsérvese que [46] es una transformación lineal no-singular del sistema de T ecuaciones [44], por tanto, ambos contienen la misma información. La utilidad de esta transformación está en separar la variación intra-grupos de la variación entre-grupos poniendo de manifiesto como se combinan ambas en la estimación y cuales son las restricciones de momentos necesarias en cada caso. Amemiya y MaCurdy (1986) consideran una transformación similar utilizando las $(T-1)$ primeras desviaciones con respecto a las medias totales en lugar de desviaciones ortogonales. La diferencia está en que en [46] los errores transformados siguen estando incorrelacionados si los errores v_{it} están incorrelacionados. A continuación presentamos los estimadores de Hausman y Taylor (HT), Amemiya y MaCurdy (AM) y Breusch, Mizon y Schmidt (BMS) en el contexto de [46] siguiendo la formulación de Arellano y Bover (1989).

4.1. Hausman-Taylor y estimadores relacionados

Definamos las particiones $x_{it} = [x'_{1it}, x'_{2it}]'$ de dimensiones $[k_1 \times 1, k_2 \times 1]$ y $z_i = [z'_{1i}, z'_{2i}]'$ de dimensiones $[g_1 \times 1, g_2 \times 1]$. x_{1it} y z_{1i} están incorrelacionadas con η_i pero x_{2it} y z_{2i} están correlacionadas. Todas las variables son estrictamente exógenas con respecto a los errores aleatorios v_{it}. Como se puede esperar, la clasificación de variables entre correlacionadas e incorrelacionadas con los efectos es decisiva y debe contrastarse. Por ejemplo, Chowdhury y Nickell (1985) estudian con el panel de Michigan el efecto sobre el logaritmo de los ingresos individuales del número de años de educación $[S_i]$, baja por enfermedad en períodos anteriores $[e_{it}]$, pertenencia a un sindicato $[c_{it}]$ y el desempleo en períodos anteriores $[u_{it}]$. El modelo también incluye la edad al cuadrado $[a^2_{it}]$, la raza $[r_i]$ y una serie de variables ficticias de tiempo. S_i siempre se supone correlacionada con η_i mientras que a^2_{it} y r_i son siempre tratadas como incorrelacionadas. Utilizando métodos del tipo HT, Chowdhury y Nickell contrastan y rechazan la ausencia de correlación de e_{it} y c_{it} con los efectos pero no de u_{it} que por tanto podría servir como instrumento (aunque pobre) para S_i.

Supongamos que η_i y v_{it} son independiente e idénticamente distribuidos (*iid*) $\eta_i \sim iid[0, \sigma^2_\eta]$, $v_{it} \sim iid[0, \sigma^2]$ con $E[\eta_i v_{it}] = 0$. Bajo estos supuestos es fácil comprobar que la matriz de covarianzas del sistema transformado [46] es una matriz diagonal de la forma

$$\mathrm{Var}[v^*_{i1} \ldots v^*_{i(T-1)} \bar{u}_i] = \sigma^2 \begin{bmatrix} I_{T-1} & 0 \\ 0 & \frac{1}{\theta^2 T} \end{bmatrix}$$

en donde $\theta^2 = \sigma^2/[\sigma^2 + T\sigma^2_\eta]$. Por tanto, para conseguir la transformación MCG con matriz de covarianzas múltiplo de la matriz unidad basta con ponderar la última ecuación de [46] obteniendo

$$\begin{aligned} y^*_{i1} &= x^*_{i1}\beta + v^*_{i1} \\ &\vdots \\ y^*_{i(T-1)} &= x^*_{i(T-1)}\beta + v^*_{i(T-1)} \\ \theta T^{\frac{1}{2}}\bar{y}_i &= \theta T^{\frac{1}{2}}\bar{x}_i'\beta + \theta T^{\frac{1}{2}} z_i'\gamma + \theta T^{\frac{1}{2}}[\eta_i + \bar{v}_i] \end{aligned} \qquad [47]$$

Es conveniente escribir este sistema en forma compacta

$$y^+_i = W^+_i\delta + u^+_i \qquad [48]$$

en donde $y^+_i = [y^*_{i1} \ldots y^*_{i(T-1)}\, \theta T^{\frac{1}{2}}\bar{y}_i]'$, $u^+_i = [v^*_{i1} \ldots v^*_{i(T-1)}\, \theta T^{\frac{1}{2}}\bar{u}_i]'$, $\delta = (\beta', \gamma')'$ y

$$W^+_i = \begin{bmatrix} x^*_{i1} & 0 \\ \vdots & \vdots \\ x^{*\prime}_{i(T-1)} & 0 \\ \theta T^{\frac{1}{2}}\bar{x}_i' & \theta T^{\frac{1}{2}} z_i' \end{bmatrix} = \begin{bmatrix} X^*_i & 0 \\ \hdashline \multicolumn{2}{c}{\theta T^{\frac{1}{2}}\bar{w}_i'} \end{bmatrix}$$

con $X^*_i = [x^*_{i1} \ldots x^*_{i(T-1)}]'$ y $\bar{w}_i = [\bar{x}_i' z_i']'$. MCO en este sistema coincide con el estimador MCG (Balestra-Nerlove) introducido en la Sección 2 (con la diferencia de que ahora el modelo contiene variables explicativas que no varían con el tiempo):

$$\begin{aligned} \hat{\delta}_{MCG} &= [\sum_i W^{+\prime}_i W^+_i]^{-1} \sum_i W^{+\prime}_i y^+_i \\ &= [\sum_i W_i'Qy_i + \theta^2 T \sum_i \bar{w}_i\bar{w}_i']^{-1} [\sum_i W_i'Qy_i + \theta^2 T \sum_i \bar{w}_i\bar{y}_i'] \end{aligned} \qquad [48']$$

en donde $y_i = [y_{i1} \ldots y_{iT}]'$, $W_i = \begin{bmatrix} x_{i1} \ldots x_{iT} \\ z_i \ \ldots z_i \end{bmatrix}'$ y Q es el operador intra-grupos. La equivalencia entre las dos expresiones de $\hat{\delta}_{MCG}$ en [48'] se puede ver teniendo en cuenta que $\Sigma_i X^{*\prime}_i X^*_i = \Sigma_i \tilde{X}_i'\tilde{X}_i$ en donde $\tilde{X}_i$ es la matriz T×k de desviaciones con respecto a las medias de las variables x_{it}. $\hat{\delta}_{MCG}$ es inconsistente en este caso puesto que parte de las variables en $\bar{x}_i$ y z_i incluidas en la última ecuación de [47] están correlacionadas con la perturbación $[\eta_i + \bar{v}_i]$. Como β está identificado a partir de la información intra-grupos (β se puede estimar consistentemente en las primeras $(T-1)$ ecuaciones por MCO), la cuestión que se plantea es si podemos encontrar instrumentos válidos para la última ecuación que permitan la identificación de γ.

HT proponen utilizar $\bar{x}_{1i}$ como instrumento para z_{2i}. Por tanto, es necesario que $k_1 \geq g_2$ para que el estimador HT esté identificado. El método HT es un estimador por variables instrumentales de [47] de la forma[3]

$$\tilde{\delta} = \left[\sum_i W_i^{+\prime} Z_i \left[\sum_i Z_i^{\prime} Z_i\right]^{-1} \sum_i Z_i^{\prime} W_i^{+}\right]^{-1} \sum_i W_i^{+\prime} Z_i \left[\sum_i Z_i^{\prime} Z_i\right]^{-1} \sum_i Z_i^{\prime} y_i^{+} \qquad [49]$$

que utiliza la siguiente matriz de instrumentos:

$$Z_i = \begin{bmatrix} x_{i1}^{*} & 0 & 0 \\ \vdots & \vdots & \vdots \\ x_{i(T-1)}^{*} & 0 & 0 \\ 0 & z_{1i}^{\prime} & \bar{x}_{1i}^{\prime} \end{bmatrix} = \begin{bmatrix} X_i^{*} & 0 \\ 0 & m_{1i}^{\prime} \end{bmatrix}$$

con $m_{1i}^{\prime} = [z_{1i}^{\prime} \; \bar{x}_{1i}^{\prime}]$. Para calcular [49] necesitamos una estimación consistente previa de θ que a su vez depende de σ^2 y σ_η^2. $\hat{\sigma}^2$ se puede obtener usando los residuos intra-grupos en la forma expresada en [7]. Sin embargo, para obtener $\hat{\sigma}_\eta^2$ necesitamos un estimador consistente $\hat{\delta} = (\hat{\beta}^{\prime} \hat{\gamma}^{\prime})$ del vector completo de parámetros. En la propuesta de HT, $\hat{\beta}$ es el estimador intra-grupos y $\hat{\gamma}$ es el estimador IV en [45] utilizando m_{1i} como instrumentos para z_i. A continuación σ_η^2 se obtiene calculando

$$\hat{\sigma}_\eta^2 = \frac{1}{N} \Sigma_{i=1}^{N} [\bar{y}_i - \hat{\delta}^{\prime} \bar{w}_i]^2 - \frac{1}{T} \hat{\sigma}^2$$

AM observan que la variación temporal de x_{1it} se puede explotar utilizando $x_{1i1}, x_{1i2}, \ldots, x_{1iT}$ como T instrumentos distintos para z_{2i} en lugar de $\bar{x}_{1i}$. El estimador AM es también de la forma [49] pero con $m_{2i}^{\prime} = [z_{1i}^{\prime} \, x_{1i1}^{\prime} \ldots x_{1iT}^{\prime}]$ en lugar de m_{1i} en la definición de la matriz de instrumentos Z_i. Una consecuencia inmediata de utilizar más instrumentos es que la identificación del estimador AM tan sólo requiere que $Tk_1 \geq g_2$. En general, AM es más eficiente que HT aunque el aumento de eficiencia puede ser modesto si la variación temporal de x_{1it} es escasa. Se podría argumentar que HT descansa en una condición de exogeneidad con respecto a η_i más débil que AM, ya que la consistencia de HT sólo requiere que la media temporal de x_{1it} y η_i estén incorrelacionados mientras que AM necesita que período a período los x_{1it} tengan cero correlación con η_i. Sin embargo, es difícil justificar la primera condición en ausencia de la segunda. En particular, querría decir que HT podría ser inconsistente si se calculara utilizando un subconjunto de los períodos de tiempo $t = 1, \ldots, T$ (*cf.* Amemiya y MaCurdy, 1986, pág. 877).

Finalmente, como sugieren Bhargava y Sargan (1983, pág. 1648), puede ser razonable suponer que las desviaciones $\tilde{x}_{2it}$ están incorrelacionadas con los efectos aunque x_{2it} esté correlacionada. Esta situación se da si las covarianzas entre los x_{2it} y η_i son constantes en el tiempo. Es decir, cuando x_{2it} se puede escribir en la forma

$$x_{2it} = \kappa \eta_i + \zeta_{it} \qquad [50]$$

[3] La forma en que presentamos los estimadores HT, AM y BMS es distinta a la forma en que aparecen en los artículos originales y está basada en Arellano y Bover (1989).

en donde $E[\zeta_{it}\eta_i] = 0$ y por tanto $\kappa = E[\eta_i x_{2it}]/E[\eta_i^2]$ que no depende de t. En este caso las variables $\tilde{x}_{2i1} \ldots \tilde{x}_{2i(T-1)}$ ($\tilde{x}_{2iT}$ es colineal con los anteriores) se pueden utilizar como instrumentos adicionales para z_{2i}. Este es precisamente el estimador propuesto por BMS que es de nuevo de la forma [49] con $m'_{3i} = [z'_{1i}\, x'_{1i1} \ldots x'_{1iT}\, \tilde{x}'_{2i1} \ldots \tilde{x}'_{2i(T-1)}]$. La condición de identificación es $Tk_1 + (T-1)\, k_2 \geq g_2$. Si los instrumentos adicionales son válidos (lo que se puede contrastar) BMS es más eficiente que AM y HT. Obsérvese, que en el caso del estimador BMS, a diferencia de HT y AM, los instrumentos de cada ecuación son también instrumentos válidos en todas las otras ecuaciones del sistema, por tanto el estimador del sistema [47] por mínimos cuadrados en dos etapas con m_{3i} como vector de instrumentos (esto es, [49] con $Z_i = I_T \otimes m'_{3i}$) es numéricamente equivalente al estimador BMS.

La consistencia de los estimadores anteriores depende de que se cumpla la condición $E[Z'_i u_i^+] = 0$, esto es, $E[x_{it}^* v_{it}^*] = 0, t = 1, \ldots, T-1$, y $E[m_{pi}\, [\eta_i + \bar{v}_i]] = 0$ (con $p = 1, 2$ ó 3). Estas restricciones de momentos requieren que todas las variables explicativas sean estrictamente exógenas con relación a v_{it} y, por tanto, no se ven afectadas si v_{it} es heterocedástico o está autocorrelacionado. Sin embargo, si v_{it} no es ruido blanco con varianza constante, [47] no es la transformación MCG por lo que su uso no tiene ninguna justificación adicional comparada con [46]. La ventaja de estimar [46] es que no requiere una estimación previa de θ. Habiendo obtenido estimadores del tipo [49] en el sistema [46] se puede mejorar la eficiencia calculando MGM en dos etapas. Esto es, reemplazando $[\Sigma_i Z'_i Z_i]$ en la fórmula del estimador por $[\Sigma_i Z'_i \hat{u}_i \hat{u}_i Z_i]$ en donde $\hat{u}_i$ son residuos de [46] en la primera etapa. En principio, ninguno de estos estimadores es completamente eficiente ya que no utilizan la transformación MCG adecuada. Con autocorrelación y heterocedasticidad arbitrarias en el corte transversal, esta transformación se podría llevar a cabo utilizando una extensión multivariante de los métodos semiparamétricos del tipo propuesto por Robinson (1987). Sin embargo, estos métodos son difíciles de calcular y todavía se tiene poca experiencia acerca de su comportamiento en las aplicaciones prácticas.

La situación es mucho más simple en el caso en que $E[u_i u'_i] = \Omega$ en donde $u_i = [u_{i1} \ldots u_{iT}]'$. Esto es, los errores son homocedásticos en el corte transversal pero se permite autocorrelación y heterocedasticidad temporal arbitrarias en v_{it}. En la Sección 2 (ecuación [9]) vimos como obtener estimadores MCG eficientes en este caso. En principio, los estimadores de esta sección se pueden generalizar de forma similar: primero obtenemos la transformación MCG del sistema utilizando la descomposición de $\Omega^{-1/2}$ (o un estimador consistente de esta descomposición) y a continuación estimamos por IV el sistema transformado. El problema con HT y AM es que al tener distintos instrumentos válidos para distintas ecuaciones, al alterar la transformación los instrumentos también se han de modificar ya que de otra forma los estimadores resultantes podrían ser inconsistentes (este caso corresponde al Modelo 4 de AM). Este problema no se plantea con el estimador BMS: si m_{3i} es un vector de instrumentos válido en todas las ecuaciones lo seguirá siendo para cualquier transformación de los errores del sistema. Por tanto, una generalización del estimador BMS con Ω arbitraria es mínimos cuadrados en tres etapas (MC3E) aplicado al sistema [44] de ecuaciones originales utilizando m_{3i} como vector de instrumentos:

$$\tilde{\delta}_{MC3E} = \left[\sum_i W_i' Z_i \left[\sum_i Z_i' \hat{\Omega} Z_i\right]^{-1} \sum_i Z_i' W_i\right]^{-1} \sum_i W_i' Z_i \left[\sum_i Z_i' \hat{\Omega} Z_i\right]^{-1} \sum_i Z_i' y_i \qquad [51]$$

en donde $Z_i = I_T \otimes m_{3i}'$, $\hat{\Omega} = N^{-1} \sum_{i=1}^{N} \hat{u}_i \hat{u}_i'$ y $\hat{u}_i$ es el vector de residuos por mínimos cuadrados en dos etapas. $\hat{\delta}_{MC3E}$ es asintóticamente equivalente a BMS cuando $v_{it} \sim iid\,[0, \sigma^2]$ y más eficiente que BMS cuando los v_{it} están autocorrelacionados. En realidad, no es difícil comprobar que si utilizamos un estimador de Ω en [51] de la forma $\tilde{\Omega} = \hat{\sigma}^2 I_T + \hat{\sigma}_\eta^2 u'$ obtenemos el estimador BMS (véase Arellano y Bover, 1989).

4.2. *Modelos con variables predeterminadas*

Bhargava y Sargan (1983) consideran un modelo del tipo Hausman y Taylor con la adición de la variable endógena retardada:

$$y_{it} = \alpha y_{i(t-1)} + x_{it}'\beta + z_i'\gamma + \eta_i + v_{it} \qquad [52]$$

Puesto que por definición las variables endógenas retardadas están correlacionadas con los errores en niveles cuando estos contienen un efecto permanente, el vector de instrumentos relativo a [52] sigue siendo del mismo tipo que en los casos estáticos considerados anteriormente. En el modelo de Bhargava y Sargan este vector es

$$d_i = [z_{1i}' \, x_{1i0}' \dots x_{1iT}' \, \tilde{x}_{2i0}' \dots \tilde{x}_{2i(T-1)}']'$$

que es igual al utilizado posteriormente por BMS con la única diferencia de que ahora suponemos que se observa el corte transversal para $t = 0$ por conveniencia de notación. Bhargava y Sargan completan el sistema de T ecuaciones [52] con las siguientes ecuaciones de forma reducida

$$y_{i0} = d_i'\pi_0 + \varepsilon_{i0} \qquad [53]$$

$$z_{2i} = \pi_1 d_i + \varepsilon_{i1} \qquad [54]$$

$$\bar{x}_{2i} = \pi_2 d_i + \varepsilon_{i2} \qquad [55]$$

y las identidades

$$x_{2it} = \tilde{x}_{2it} + \bar{x}_{2i} \qquad (t = 0, \dots, T-1) \qquad [56]$$

Bhargava y Sargan proponen estimar inicialmente el sub-sistema de T ecuaciones [52] por el método de máxima verosimilitud con información limitada (MVIL) sin imponer restricciones en la forma de la matriz $T \times T$ Ω de covarianzas $E[u_{it} u_{is}] = \omega_{ts}$. También desarrollan una versión alternativa del estimador MVIL que impone las restricciones del modelo de errores compuestos en Ω. Por otra parte, la estimación por MVIL de un sub-conjunto de ecuaciones simultáneas es asintóticamente equivalente a la estimación por MC3E de esas mismas ecuaciones cuando no se utilizan restricciones en las covarianzas (véase por ejemplo, Sargan, 1988b). Por tanto el estimador MC3E en [51] con $\delta = (\alpha\beta'\gamma')'$,

$$W_i = \begin{bmatrix} y_{i0} \cdots y_{i(T-1)} \\ x_{i1} \ldots x_{iT} \\ z_i \ \ldots z_i \end{bmatrix}'$$

y $Z_i = I_T \otimes d_i'$ tiene la misma varianza asintótica que el estimador MVIL de Bhargava y Sargan con Ω sin restricciones.

En el caso en que v_{it} no está autocorrelacionado, las perturbaciones en diferencias o en desviaciones ortogonales hacia adelante están incorrelacionadas con los retardos de la variable endógena, que se pueden utilizar como instrumentos adicionales. Para ello necesitamos utilizar el sistema transformado

$$\begin{aligned} y_{i1}^{*} &= \alpha y_{i0}^{*} + x_{i1}^{*\prime}\beta + v_{i1}^{*} \\ &\vdots \\ y_{i(T-1)}^{*} &= \alpha y_{i(T-2)}^{*} + x_{i(T-1)}^{*\prime}\beta + v_{i(T-1)}^{*} \\ \bar{y}_i &= \alpha\bar{y}_{i(-1)} + \bar{x}_i'\beta + z_i'\gamma + [\eta_i + \bar{v}_i] \end{aligned} \qquad [57]$$

usando como matriz de instrumentos:

$$Z_i = \begin{bmatrix} d_i' y_{i0} & & 0 \\ & d_i' y_{i0} y_{i1} \ddots & \\ 0 & & d_i' y_{i0} \ldots y_{i(T-2)} \; d_i' \end{bmatrix}$$

Distintas variantes de estimadores MGM de este tipo se pueden construir dependiendo de las propiedades que se supongan acerca de las variables explicativas que varían con el tiempo, incluyendo casos en los que un subconjunto de las x_{it} son variables predeterminadas en lugar de estrictamente exógenas, pudiendo éstas a su vez estar o no correlacionadas con los efectos.

En la medida en que haya restricciones sobre-identificadoras, las especificaciones más restrictivas de las variables x_{it} son contrastables en el marco MGM. Por ejemplo, Holtz-Eakin (1988) combina restricciones de momentos en niveles y en primeras diferencias para construir contrastes de la existencia de efectos permanentes en modelos autorregresivos.

5. Modelos con variables dependientes limitadas

Con frecuencia las decisiones individuales que queremos modelizar son de carácter cualitativo como la participación de un individuo en el mercado de trabajo o el pago de dividendos en una empresa. En estos casos la variable dependiente y_{it} es binaria y usualmente toma el valor $y_{it} = 1$ si la acción se produce e $y_{it} = 0$ en caso contrario. Por tanto, lo que observamos en la muestra son distintas frecuencias de ocurrencia de un suceso determinado para grupos de individuos que comparten ciertas características observables. Es natural pues modelizar la probabilidad condicional de ocurrencia del suceso dado un conjunto de variables exógenas. Esta estrategia equivale a aplicar el análisis de regresión a este caso:

$$E\,[y_{it}\,|\,x_i] = Pr[y_{it} = 1\,|\,x_i] \qquad [58]$$

en donde $x_i = [x'_{i1} \dots x'_{iT}]'$. Si suponemos que la función de regresión es lineal, obtenemos el modelo lineal de probabilidad $E[y_{it}\,|x_i] = x'_{it}\beta$ que no parece muy adecuado sabiendo que esta función debe ser una probabilidad que varía entre cero y uno. Una generalización del modelo lineal apropiada en este contexto consiste en suponer que las variables exógenas sólo intervienen mediante la combinación lineal $x'_{it}\beta$ que a su vez afecta a y_{it} a través de una función de distribución de probabilidad acumulada univariante: $E[y_{it}\,|\,x_i] = F[x'_{it}\beta]$. Los modelos *logit* y *probit* son los casos especiales que más se utilizan en la práctica. En el primero, F es la distribución logística mientras que en el segundo es la normal $N(0,1)$. En ambos casos obtenemos un modelo de regresión no-lineal que se puede estimar por el método de máxima verosimilitud.

Con datos de panel es natural preguntarse si en un modelo binario el panel puede ser útil para controlar diferencias inobservables entre individuos como ocurre en el modelo lineal. Añadiendo efectos individuales a la especificación anterior tenemos

$$E[y_{it}\,|\,x_i,\ \eta_i] = F[x'_{it}\beta + \eta_i] \qquad [59]$$

A partir de esta ecuación en general no es posible derivar una condición en la que no aparezcan los η_i y que se pueda utilizar como base para estimar β. Por ello en modelos no lineales las inferencias acerca de β condicionando en η_i sólo son posibles para determinadas especificaciones de $F(.)$. Sigue siendo posible utilizar el método de efectos fijos consistente en estimar conjuntamente β y los η_i por máxima verosimilitud, y el método de efectos aleatorios especificando una distribución de probabilidad para η_i, pero como veremos más adelante el primer método es en general inconsistente cuando T es fijo y los supuestos requeridos para el segundo son a menudo considerados excesivamente restrictivos.

Un marco muy útil para los modelos de elección cualitativa consiste en introducir una variable aleatoria latente continua y^*_{it} que se relaciona con las evaluaciones de las funciones objetivo de los agentes. El suceso ocurre si y sólo si $y^*_{it} \geq 0$. En estos términos el modelo [59] se puede escribir como un modelo de regresión lineal para y^*_{it} junto con un supuesto acerca de la relación entre la variable endógena latente y^*_{it} y la variable endógena observable y_{it}:

$$y^*_{it} = \beta' x_{it} + \eta_i + v_{it} \qquad (t = 1, \dots, T: \quad i = 1, \dots, N) \qquad [60]$$

$$y_{it} = \mathbf{1}\,[y^*_{it} \geq 0] \qquad [61]$$

en donde $1(A)$ es la función indicador del suceso A que toma el valor uno si A es cierto y cero en caso contrario. El término v_{it} es una perturbación aleatoria con distribución F. Una de las ventajas de este marco es que es una forma natural de utilizar modelos teóricos de elección en la especificación econométrica. Por ejemplo, en un análisis de participación laboral femenina basado en un modelo de búsqueda, y_{it}^* se puede interpretar como la diferencia entre salarios de reserva y salarios de mercado (*cf.* Heckman, 1974). Otra ventaja importante es que es un marco estadístico muy amplio que permite acomodar y relacionar la clase de modelos conocidos por la denominación genérica de «variables dependientes limitadas» (VDL). Por ejemplo, reemplazando [61] por

$$y_{it} = \max\,[y_{it}^*, 0] \qquad [62]$$

obtenemos el modelo de regresión censurado o modelo Tobit (*cf.* Tobin, 1958; Amemiya, 1973). En este caso la variable dependiente es continua para valores positivos pero tiene un punto de acumulación en cero. En el modelo de regresión truncado la especificación es similar con la diferencia de que el punto de acumulación no se observa. En el modelo Tobit una sola ecuación gobierna dos decisiones: primeramente, la de llevar a cabo la acción, y en segundo lugar la elección de un valor para y_{it} en el caso en que la acción se produzca. En el modelo generalizado de selectividad las dos decisiones se separan introduciendo una segunda ecuación que gobierna la primera decisión (*cf.* Heckman, 1976 y 1979):

$$I_{it} = \gamma' z_{it} + \varepsilon_{it} \qquad [63]$$

$$y_{it} = \begin{cases} y_{it}^* & \text{si } I_{it} \geq 0 \\ 0 & \text{en caso contrario} \end{cases} \qquad [64]$$

Esta especificación y sus variantes son también útiles para corregir la falta de aleatoriedad en muestras en las que el proceso de selección de las observaciones es endógeno.

Aunque en esta sección nos vamos a concentrar en el estudio de los problemas asociados con la introducción de efectos individuales y respuestas dinámicas en los modelos VDL, hay situaciones de interés microeconométrico en que estos modelos no son la mejor forma de representar una variable no continua. Este es el caso de variables en las que las observaciones son números enteros positivos de bajo orden en las que el cero puede aparecer con una probabilidad no despreciable (por ejemplo, el número de intervalos de baja por enfermedad en un año, el número de coches en una familia, etc.). Para este tipo de datos los modelos basados en la distribución Poisson ofrecen un punto de partida razonable (véase Hausman, Hall y Griliches (1984), y Gourieroux, Monfort y Trognon, 1984).

5.1. Efectos individuales en modelos VDL

Empecemos considerando el caso en que la ecuación [60] representa una extensión VDL del modelo tradicional con errores compuestos. Esto es, las variables explicativas x_{it} se suponen exógenas con respecto a la perturbación compuesta $u_{it} = \eta_i + v_{it}$, y por tanto independientes del efecto individual η_i. El modelo se completa con un supuesto de observabilidad acerca de la relación entre y_{it} e y^*_{it}, y con un supuesto acerca de la distribución de $u_i = [u_{i1} \ldots u_{iT}]'$. Asumiendo que u_i es normal multivariante $N(0, \Omega)$ y la correspondencia [61] obtenemos un modelo *probit* multivariante que en principio puede ser estimado por máxima verosimilitud. La dificultad es que la evaluación de la función de verosimilitud requiere el cálculo de integrales múltiples de orden T que resultan computacionalmente intratables cuando T es mayor que 3 ó 4. No obstante, en el caso en que $\Omega = \sigma^2 I_T + \sigma^2_\eta \iota\iota'$ la integral múltiple se puede escribir como una integral univariante de un producto de funciones de distribución acumuladas normales que es relativamente más fácil de evaluar (sobre este punto véase Heckman, 1981, y Butler y Moffitt, 1982). Estos comentarios también se aplican al modelo Tobit y a los modelos generalizados de selectividad. Recientemente, Hajivassiliou y McFadden (1989) han aplicado el método de momentos simulados de McFadden (1989) a modelos VDL con errores compuestos en los que se supone que v_{it} sigue un proceso autorregresivo de primer orden. Todos estos métodos son complicados y sensibles a la especificación de Ω incluso manteniendo la normalidad, sin embargo la estimación consistente de β no ofrece especiales dificultades. Una posibilidad es obtener T estimaciones ML o por algún método robusto de β ecuación por ecuación (por tanto, ignorando la correlación entre los u_{it}) para a continuación combinarlas en un sólo estimador utilizando una media ponderada (un método de distancia mínima). Otra posibilidad en el caso *probit* es utilizar el estimador MGM propuesto por Avery, Hansen y Hotz (1983).

De todas formas el modelo con efectos independientes de las variables explicativas es de interés limitado y, por tanto, debemos considerar los problemas asociados con los modelos VDL con efectos fijos y modelos con efectos aleatorios correlacionados con x_{it}. Neyman y Scott (1948) estudiaron las propiedades del método ML cuando el vector de parámetros a estimar se puede dividir en un subconjunto fijo de parámetros de interés y otro subconjunto de parámetros cuyo número crece con el tamaño de la muestra (parámetros accidentales). En nuestro caso estas nociones corresponden al modelo de efectos fijos con $[\beta, \eta_1, \eta_2, \ldots, \eta_N]$. La conclusión es que por lo general el estimador ML de β resuelto conjuntamente con los de los η_i es inconsistente (con T fijo)[4]. Una excepción importante es el modelo de regresión lineal estudiado en la Sección

[4] Heckman (1981b) lleva a cabo simulaciones de Monte Carlo para evaluar la importancia de estos sesgos en varios modelos *probit* multiperíodo. Los sesgos no son importantes en los modelos estáticos pero sí en las versiones dinámicas. Por otra parte, los resultados de las simulaciones de Hotz y Miller (1988) utilizando un modelo generalizado de selectividad con $T = 10$, indican que los sesgos son suficientemente pequeños para que la estimación conjunta de los efectos y el resto de los parámetros sea útil en su caso.

2. Por otra parte si disponemos de un estadístico s tal que podemos escribir la siguiente factorización de la función de verosimilitud conjunta

$$\mathscr{L}[y; \beta, \eta_1 \ldots \eta_N] = \mathscr{L}(y \mid s; \beta)\mathscr{L}[s; \beta, \eta_1 \ldots \eta_N] \quad [65]$$

la maximización de la función de verosimilitud condicional $\mathscr{L}(y \mid s; \beta)$ que depende de β pero no de los η_i proporcionará en general estimaciones consistentes de β (*cf.* Andersen, 1973 y Chamberlain, 1980). En el modelo de regresión lineal, s es el vector de medias individuales y $\mathscr{L}(y \mid s; \beta)$ corresponde a la función de verosimilitud del modelo en desviaciones con respecto a las medias. Además para este modelo como ya sabemos los estimadores ML de β conjuntos y condicionales coinciden (y son iguales al estimador intra-grupos). Los estimadores de este tipo son muy atractivos porque no requieren ningún supuesto acerca de la distribución de los efectos individuales. Por desgracia no es siempre posible encontrar un estadístico que opere la descomposición [65] en cuyo caso el estimador ML condicional no es una alternativa[5].

El modelo *logit* es un caso en el que se puede aplicar la estrategia de la función de verosimilitud condicional. Consideremos de nuevo el modelo dado por [60] y [61] con $T = 2$ asumiendo que los v_{it} son *iid* con la función logística $F(v) = e^v/[1 + e^v]$ y dejando sin especificar los η_i. Esto es

$$E[y_{it} \mid x_i, \eta_i] = Pr[y_{it} = 1 \mid x_i, \eta_i]$$

$$= Pr[v_{it} \leq \beta' x_{it} + \eta_i \mid x_i, \eta_i] = \frac{\exp[\beta' x_{it} + \eta_i]}{1 + \exp[\beta' x_{it} + \eta_i]} \quad [66]$$

Si condicionamos en las medias individuales como en el modelo lineal (o simplemente $y_{i1} + y_{i2}$ con $T = 2$) obtenemos una probabilidad que no depende de η_i:

$$Pr[y_{i2} = 1 \mid x_i, \eta_i, y_{i1} + y_{i2} = 1] = \frac{\Pr[y_{i2} = 1, y_{i1} + y_{i2} = 1 \mid x_i, \eta_i]}{\Pr[y_{i1} + y_{i2} = 1 \mid x_i, \eta_i]}$$

$$= \frac{\Pr[y_{i1} = 0 \mid x_i, \eta_i] \ \Pr[y_{i2} = 1 \mid x_i, \eta_i]}{\Pr[y_{i1} = 0 \mid x_i, \eta_i] \Pr[y_{i2} = 1 \mid x_i, \eta_i] + \Pr[y_{i1} = 1 \mid x_i, \eta_i] \ \Pr[y_{i2} = 0 \mid x_i, \eta_i]}$$

$$= F(\beta'[x_{i2} - x_{i1}]) \quad [67]$$

Obsérvese que si $y_{i1} + y_{i2}$ es 0 ó 2, Pr $[y_{i2} = 1 \mid x_i, \eta_i, y_{i1} + y_{i2}]$ es 0 ó 1, respectivamente. Por tanto sólo las observaciones con $y_{i1} + y_{i2} = 1$ contienen información sobre β en este modelo. De esta forma obtenemos una función de verosimilitud *logit* para esas observaciones en la que los dos resultados posibles son $[y_{i1} = 0, y_{i2} = 1]$ y $[y_{i1} = 1, y_{i2} = 0]$ con las variables explicativas en primeras diferencias $[x_{i2} - x_{i1}]$. Este procedimiento fue propuesto por Andersen

[5] Véase Cox y Reid (1987), Cox (1988) y Cruddas, Reid y Cox (1989) sobre el uso de funciones de verosimilitud condicionales aproximadas en conexión con el problema de los parámetros accidentales.

(1970); una discusión más completa que la nuestra se puede encontrar en Chamberlain (1980) en donde también se analiza el caso en que $T > 2$.

Este procedimiento se puede ver como el equivalente binario de las primeras diferencias en el modelo lineal. El problema es que en este caso a diferencia del lineal, los estimadores resultantes sólo son consistentes si los errores son en realidad logísticos, homocedásticos y no autocorrelacionados. Manski (1987) ha demostrado que en un modelo como el anterior

$$\text{med}[y_{i2}-y_{i1} \mid x_i, y_{i1}+y_{i2}=1] = \text{sgn}E[y_{i2}-y_{i1} \mid x_i] = \text{sgn}[\beta'[x_{i2}-x_{i1}]] \qquad [68]$$

en donde med(.) es la mediana y sgn(.) es la función signo definida por sgn(u) $= -1$ si $u < 0$, sgn(u) $= 0$ si $u = 0$ y sgn(u) $= 1$ si $u > 0$. El resultado [68] no requiere la distribución logística, sino que es válido para cualquier distribución de v_{it} con tal de que ésta sea estacionaria (esto es, que la distribución de $v_{i1} \mid x_i, \eta_i$ sea de la misma forma que la distribución de $v_{i2} \mid x_i, \eta_i$); por tanto permite correlación entre v_{i1} y v_{i2} y heterocedasticidad en el corte transversal. Con el supuesto adicional de que al menos una de las variables x_{it} sea continua y acotada, Manski prueba que el vector de coeficientes estandarizado $\beta/\|\beta\|$ está identificado. Desafortunadamente [68] no se puede utilizar para definir un estimador MGM porque la función sgn no es invertible. Manski propone utilizar el estimador de *score* máximo (*cf.* Manski, 1985) que maximiza la siguiente función criterio

$$s(\beta) = \Sigma_{i=1}^{N} [y_{i2}-y_{i1}] \,\text{sgn}[\beta' \Delta x_{i2}] \qquad [69]$$

y también demuestra su consistencia. La función $s(\beta)$ tendrá en general varios máximos locales en cuyo caso la elección de valores iniciales de los parámetros puede ser crucial. Otro problema de este estimador es que no es asintóticamente normal a la tasa de convergencia $\sqrt{N}$ por lo que no es posible calcular errores estándar y contrastes de hipótesis asintóticos en la forma habitual.

A continuación pasamos a considerar modelos en los que se especifica la distribución condicional de los efectos dado x_i (modelos con «efectos aleatorios» correlacionados con x_{it}). En particular, si junto con [60] suponemos que

$$\eta_i \mid x_i \sim N[\lambda_1' x_{i1} + \ldots + \lambda_T' x_{iT}, \sigma_\eta^2] \qquad [70]$$

$$v_i \mid x_i \sim N(0, \Omega) \qquad [71]$$

obtenemos un modelo de regresión VDL multivariante del tipo considerado por Chamberlain (1984):

$$y_{it}^* = \pi_{t1}' x_{i1} + \ldots + \pi_{tT}' x_{iT} + \varepsilon_{it} = \pi_t' x_i + \varepsilon_{it} \quad (t = 1, \ldots, T) \qquad [72]$$

en donde $\pi_{tt} = \beta + \lambda_t$, $\pi_{ts} = \lambda_s$ para $t \neq s$ y $[\varepsilon_{i1} \ldots \varepsilon_{iT}]$ tiene una distribución normal conjunta. La estimación eficiente de este sistema está sujeta a los mismos problemas computacionales que hemos discutido en relación con los modelos de errores compuestos. Sin embargo, es relativamente sencillo construir estimadores de β basados en estimaciones consistentes de los coeficientes π_t obtenidas ecuación por ecuación. Definamos $\hat{\pi} = [\hat{\pi}'_1 ... \hat{\pi}'_T]'$, en donde $\hat{\pi}_t$ es el estimador ML de π_t en la ecuación t-ésima [72] (por ejemplo $\hat{\pi}_t$ puede ser un *probit* uniecuacional si el modelo se completa con [61] o un *Tobit* si se trata de [62]). Como β y $\lambda = [\lambda'_1 ... \lambda'_T]'$ son funciones de π, podemos estimarlos conjuntamente minimizando la distancia cuadrática entre $\hat{\pi}$ y $\pi(\beta, \lambda)$:

$$d(\beta,\lambda) = [\hat{\pi} - \pi(\beta,\lambda)]' V^{-1} [\hat{\pi} - \pi(\beta,\lambda)] \qquad [73]$$

La elección óptima para V es un estimador consistente de la varianza asintótica de $\hat{\pi}$. Este es el método de distancia mínima propuesto por Chamberlain (1984).

La minimización directa de [73] tiene varios inconvenientes. Uno de ellos es que nos obliga a estimar conjuntamente β y λ a pesar de que normalmente estaremos únicamente interesados en β y el vector λ puede contener un gran número de coeficientes. Por otra parte la evaluación de [73] requiere fórmulas explícitas de las restricciones en los coeficientes π implicados por el modelo. Finalmente, estas restricciones serán a menudo no-lineales haciendo necesario el uso de métodos iterativos en esta etapa. Un procedimiento alternativo es el siguiente. Tomando desviaciones con respecto a las medias en [60] los efectos η_i son eliminados. Si a continuación tomamos esperanzas condicionales a x_i obtenemos

$$E[\tilde{y}^*_{it} | x_i] = \beta' \tilde{x}_{it} \qquad [74]$$

Obsérvese que de acuerdo con [72] $E[y^*_{it} | x_i] = \pi'_t x_i$, por tanto definiendo $\pi^+_t = \pi_t - [\pi_1 + ... + \pi_T]/T$ tenemos

$$\pi^{+\prime}_t x_i = \beta' \tilde{x}_{it} \quad (t = 1, ..., T) \qquad [75]$$

En particular, esto quiere decir que β resuelve

$$\beta = [\sum_i \sum_t \tilde{x}_{it} \tilde{x}'_{it}]^{-1} \sum_i \sum_t \tilde{x}_{it} \pi^{+\prime}_t x_i \qquad [76]$$

por lo que un simple estimador consistente de β se puede obtener reemplazando π^+_t por $\hat{\pi}^+_t = \hat{\pi}_t - [\hat{\pi}_1 + ... + \hat{\pi}_T]/T$ en la ecuación anterior. En definitiva, se trata de calcular predicciones de las variables latentes y^*_{it} utilizando la forma reducida [72]: $\hat{y}^*_{it} = \hat{\pi}'_t x_i$, para a continuación estimar β mediante la regresión intra-grupos de $\hat{y}^*_{it}$ sobre x_{it}. Se puede demostrar que el estimador obtenido de esta forma es asintóticamente equivalente a un estimador de distancia mínima que utiliza un valor ineficiente de V en [73] (véase Bover y Arellano, 1988). Para obtener un estimador eficiente (relativo a un $\hat{\pi}$ dado) reescribamos [75] en la forma

$$\pi^{+\prime}_t x_i = \beta' Q'_t x_i$$

en donde Q_l es una matriz de selección $k \times Tk$ tal que $\tilde{x}_{it} = Q_l' x_i$. De este modo tenemos que

$$\pi_l^+ = Q_l \beta$$

lo cual sugiere considerar estimadores de β de la forma

$$\hat{\beta} = \left[\sum_t^T \sum_s^T Q_t' A_{ts} Q_s \right]^{-1} \sum_t^T \sum_s^T Q_t' A_{ts} \hat{\pi}_s^+ \qquad [77]$$

en donde las A_{ts} son matrices de ponderaciones de orden $Tk \times Tk$. El estimador intra-grupos con las predicciones descrito anteriormente se puede obtener a partir de esta fórmula con

$$A_{ts} = \begin{cases} \Sigma_{i=1}^N x_i x_i' & \text{para } t = s \\ 0 & \text{para } t \neq s \end{cases}$$

Sin embargo la elección eficiente de A_{ts} viene dada por el (t,s)-ésimo bloque de la inversa de la varianza asintótica de $\hat{\pi}^+$. No obstante, si el incremento de eficiencia es pequeño puede ser aconsejable utilizar el estimador intra-grupos con las predicciones por razones de simplicidad computacional.

Los modelos en los que se especifica la distribución de $\eta_i | x_i$ introducen más estructura que los modelos condicionales. Por esta razón se puede esperar que los primeros sean más eficientes pero menos robustos que los segundos. En particular, los métodos anteriores del tipo Chamberlain son sensibles a la especificación de $E(\eta_i | x_i)$. Sin embargo la ausencia de resultados suficientemente generales en la teoría de la estimación condicional de modelos no lineales, sugiere que un camino fructífero es considerar modelos de efectos aleatorios con especificaciones más generales de la distribución de $\eta_i | x_i$. Por ejemplo, un estimador consistente no paramétrico de $E[y_{it}^* | x_i]$ proporcionaría predicciones mucho más robustas de y_{it}^* sobre las que basar un estimador intra-grupos.

5.2. Modelos VDL dinámicos

En el estudio empírico de sucesos individuales a lo largo del tiempo se suele observar que los individuos que han experimentado un suceso en el pasado (por ejemplo, un período de desempleo o de enfermedad) tienen una probabilidad más alta de volverlo a experimentar que aquellos que no lo experimentaron. Los modelos con efectos permanentes que hemos considerado en esta Sección dan lugar a dependencia temporal en los estados individuales. Sin embargo es una dependencia espúrea en el sentido de que tras controlar por los η_i ésta desaparece. En el caso de un modelo de desempleo, estaríamos suponiendo que los individuos difieren en su propensión a experimentar el desempleo pero que teniendo en cuenta estas diferencias (observables e inobservables), el haber estado desempleado no aumenta las posibilidades de volverlo a estar en el futuro. Por el contrario, en un modelo con dependencia estructural en los estados, incluso teniendo en cuenta posibles diferencias inobservables entre individuos [η_i], la experiencia de un suceso en el pasado modifica la probabilidad de ocurrencia futura. Esto es, en un modelo binario tendríamos que

$$\Pr[y_{it} = 1 \mid y_{i(t-1)} = 1, x_i, \eta_i] \neq \Pr[y_{it} = 1 \mid x_i, \eta_i] \qquad [78]$$

La especificación de modelos con heterogeneidad y de modelos con dependencia estructural en los estados ha sido estudiada con detalle en Heckman (1981a), en donde se puede encontrar una completa tipología de modelos binarios dinámicos.

Las consideraciones anteriores sugieren un modelo VDL en el que las evaluaciones de utilidad de los agentes (la variable latente y^*_{it}) dependen de los estados pasados:

$$y^*_{it} = \lambda d_{i(t-1)} + \beta' x_{it} + \eta_i + v_{it} \qquad [79]$$

en donde $d_{it} = 1[y_{it} = y^*_{it}]$. d_{it} es el indicador de estado asociado con la observabilidad de y^*_{it} y por tanto en el modelo binario $d_{it} = y_{it}$. Si eliminamos las variables explicativas, los efectos individuales y completamos el modelo con [61] obtenemos un proceso de Markov de primer orden.

Un modelo dinámico conceptualmente distinto aparece si suponemos que las decisiones del presente dependen de las evaluaciones de utilidad pasadas en lugar de los estados pasados:

$$y^*_{it} = \alpha y^*_{i(t-1)} + \beta' x_{it} + \eta_i + v_{it} \qquad [80]$$

Este tipo de modelo aparece de forma natural en problemas de maximización intertemporal de utilidad con hábitos persistentes. Por ejemplo, si [80] representa el nivel de demanda deseado de un bien, la contribución a la función de utilidad intertemporal en el período t dependería de una pseudo diferencia entre y^*_{it} e $y^*_{i(t-1)}$. En principio, el efecto de $y^*_{i(t-1)}$ se podría separar del de $d_{i(t-1)}$ combinando [79] y [80] en un modelo que englobara a ambos.

Empecemos considerando la estimación de un modelo de efectos aleatorios de la forma [80]. Si especificamos la distribución $y^*_{i1} \mid x_i$, este modelo es una simple extensión del modelo estático considerado anteriormente. Suponiendo que

$$y^*_{i1} \mid x_i \sim N[\mu'_1 x_{i1} + \ldots + \mu'_T x_{iT}, \sigma^2] \qquad [81]$$

junto con [70] y [71] obtenemos para [80] una forma reducida igual a [72] en la que los coeficientes π_{ts} son funciones de α, β, λ y $\mu = [\mu'_1 \ldots \mu'_T]'$. El criterio de distancia mínima [73] se sigue pudiendo utilizar para el nuevo conjunto de restricciones en π así como el estimador intra-grupos. Tomando desviaciones y esperanzas condicionales en [80] tenemos[6].

$$E[\tilde{y}^*_{it} | x_i] = \alpha E[\tilde{y}^*_{i(t-1)} | x_i] + \beta' \tilde{x}_{it} \qquad [82]$$

o

$$\pi^{+\prime}_{0t} x_i = \alpha \pi^{+\prime}_{1(t-1)} x_i + \beta' \tilde{x}_{it} = \delta' \tilde{w}_{it} \qquad [83]$$

en donde $\pi^+_{0t} = \pi_t - [\pi_2 + ... + \pi_T]/(T-1)$, $\pi^+_{1(t-1)} = \pi_{t-1} - [\pi_1 + ... + \pi_{T-1}]/(T-1)$, $\delta = (\alpha\ \beta')'$ y $\tilde{w}_{it} = [\pi^{+\prime}_{1(t-1)} x_i, \tilde{x}'_{it}]'$, por lo que δ resuelve

$$\delta = [\sum_i \sum_t \tilde{w}_{it} \tilde{w}'_{it}]^{-1} \sum_i \sum_t \tilde{w}_{it} [\pi^{+\prime}_{0t} x_i] \qquad [84]$$

Reemplazando $\tilde{w}_{it}$ en [84] por $[\hat{\pi}^{+\prime}_{1(t-1)} x_i, \tilde{x}'_{it}]'$ y π^+_{0t} por $\hat{\pi}^+_{0t}$ obtenemos un estimador consistente y asintóticamente normal de δ que equivale a llevar a cabo la regresión intra-grupos de las predicciones $\hat{y}^*_{it}$ sobre $\hat{y}^*_{i(t-1)}$ y x_{it}. Una vez se dispone de estimaciones de la matriz de covarianzas de $\hat{\pi}$, la obtención de estimaciones consistentes de la varianza de estos estimadores intra-grupos es relativamente simple. Las fórmulas se pueden encontrar en Bover y Arellano (1988). En cuanto a la eficiencia y la robustez de estos estimadores, las mismas consideraciones que hicimos para el modelo estático se aplican en este caso.

Obsérvese que las estimaciones $\hat{y}^*_{it}$ de $E[y^*_{it} | x_i]$ sólo se han de calcular una vez y a partir de ahí se pueden utilizar como si fueran las observaciones de la variable dependiente en la estimación de modelos alternativos con o sin retardos en y^*_{it} utilizando el procedimiento intra-grupos. En general, una característica atractiva de los métodos del tipo de Chamberlain es la separación entre la búsqueda de una especificación adecuada al nivel de la forma reducida y al nivel de la ecuación estructural. La forma funcional y los supuestos de observabilidad y de distribución se pueden contrastar en la forma reducida hasta que se disponga de un $\hat{y}^*_{it}$ estadísticamente satisfactorio, para después concentrarse en la ecuación de interés.

La estimación de [79] es más problemática porque en ese caso la especificación del modelo es difícilmente compatible con una forma reducida lineal. Por ejemplo, si el modelo es binario o Tobit tomando esperanzas condicionales en [79] obtenemos

$$E[y^*_{it} | x_i] = \lambda \Pr[y^*_{i(t-1)} \geq 0 | x_i] + \beta' x_{it} + E[\eta_i | x_i] \qquad [85]$$

Incluso si $E[\eta_i | x_i]$ es lineal, $\Pr[y^*_{i(t-1)} \geq 0 | x_i]$ será en general una función no lineal de x_i. Por otra parte como el término de perturbación $\varepsilon_{it} = y^*_{it} - E[y^*_{it} | x_i]$

[6] Obsérvese que en este modelo $\tilde{y}^*_{it} = y^*_{it} - \frac{1}{(T-1)} [y^*_{i2} + ... + y^*_{iT}]$ mientras que $\tilde{y}^*_{i(t-1)} =$ $= y^*_{i(t-1)} - \frac{1}{(T-1)} [y^*_{i1} + ... + y^*_{i(T-1)}]$ por tanto en nuestra notación $\tilde{y}^*_{i(t-1)}$ no coincide con el valor de $\tilde{y}^*_{it}$ retardado un período. Dado el contexto, creemos innecesario introducir una notación especial para diferenciar ambos conceptos.

contiene un componente binario, el supuesto de normalidad es difícil de justificar. Si dispusiéramos de estimaciones no-paramétricas consistentes de $E[y^*_{it}|x_i]$ y $E[d_{it}|x_i]$ lo suficientemente generales, sería natural extender el procedimiento intra-grupos mencionado anteriormente para estimar λ y β.

Heckman (1981a y 1981b) considera la aplicación de métodos máximo verosímiles a la estimación de modelos binarios con dependencia estructural en los estados. Si el modelo no contiene efectos η_i, los métodos de estimación de cadenas de Markov con variables exógenas son apropiados (*cf.* Amemiya, 1985). Si por el contrario se incluye η_i, la especificación de las condiciones iniciales y^*_{i1} es crucial. Heckman (1981b) sugiere un procedimiento ML aproximado basado en la aproximación de $y^*_{i1}|x_i$ mediante un modelo *probit* lineal.

La identificación de los modelos de efectos aleatorios estudiada hasta ahora se basa en la condición de exogeneidad estricta $E[v_{it}|x_i] = 0$ y, por tanto, los estimadores asociados son consistentes cualquiera que sea la correlación entre los v_{it} para distintos períodos. Sin embargo, es posible la identificación de modelos VDL con tan sólo variables predeterminadas basándonos a cambio en supuestos de ausencia de autocorrelación en los errores v_{it}. El tratamiento es similar al desarrollado en la Sección 3 para modelos lineales. En esta exposición vamos a considerar únicamente el caso más simple que consiste en una autorregresión en y^*_{it} con efectos individuales:

$$y^*_{it} = \alpha y^*_{i(t-1)} + \eta_i + v_{it} \quad (t = 1, ..., T) \qquad [86]$$

en donde los errores v_{it} no están autocorrelacionados. El modelo se completa con un supuesto de observabilidad relacionando y^*_{it} con y_{it}. Supongamos inicialmente que $T = 3$ (observamos y_{i1}, y_{i2}, y_{i3}) y además

$$\eta_i | y^*_{i1} \sim N[\lambda y^*_{i1}, \sigma^2_\eta] \qquad [87]$$

$$\begin{bmatrix} v_{i2} \\ v_{i3} \end{bmatrix} | y^*_{i1} \sim N\left[0, \begin{bmatrix} \sigma^2_2 & 0 \\ 0 & \sigma^2_3 \end{bmatrix}\right] \qquad [88]$$

entonces

$$E[y^*_{it}|y^*_{i1}] = \pi_{t1} y^*_{i1} \quad (t = 2,3) \qquad [89]$$

con $\pi_{21} = \alpha + \lambda$ y $\pi_{31} = \alpha(\alpha+\lambda) + \lambda$. π_{21} y π_{31} se pueden estimar separadamente por máxima verosimilitud utilizando las observaciones con $y^*_{i1} > 0$ (nótese que como estamos condicionando en y^*_{i1} se trata de una selección de muestra exógena relativa a π_{21} y π_{31}) y a continuación obtener

$$\hat{\alpha} = [\hat{\pi}_{31} - \hat{\pi}_{21}]/[\hat{\pi}_{21} - 1] \qquad [90]$$

Con $T = 3$, $[y^*_{i2}\ y^*_{i3}]|y^*_{i1}$ se distribuye normalmente sin restricciones de covarianzas por lo que la única ineficiencia incurrida en [90] proviene de la estimación separada de π_{21} y π_{31}. Sin embargo con $T > 3$ la varianza de $[y^*_{i2} ... y^*_{iT}]|y^*_{i1}$ está sujeta a restricciones. Ignorar estas restricciones basando nuestro estimador de α exclusivamente en las medias condicionales de la forma [89] puede dar lugar a estimaciones excesivamente imprecisas (equivaldría a utilizar únicamente y_{i1} como instrumento para todas las ecuaciones en el modelo autorregresivo de la Sección 3). Por otra parte, la estimación ML conjunta de $(T-1)$ ecuaciones VDL con restricciones de covarianzas es en general computacionalmente intratable. Una alternativa que permite explotar parte de la información en las covarianzas es ampliar la forma reducida [89] considerando un número creciente de variables condicionales. Por ejemplo, con $T = 4$ tras transformar [86] a primeras diferencias, podemos tomar esperanzas condicionales dado y^*_{i1} en la ecuación para $t = 3$ y dados y^*_{i1} e y^*_{i2} en la ecuación para $t = 4$:

$$E[y^*_{i3} - y^*_{i2} | y^*_{i1}] = \alpha E[y^*_{i2} - y^*_{i1} | y^*_{i1}]$$

$$E[y^*_{i4} - y^*_{i3} | y^*_{i1}, y^*_{i2}] = \alpha E[y^*_{i3} - y^*_{i2} | y^*_{i1}, y^*_{i2}] \qquad [91]$$

Suponiendo que estas esperanzas son lineales tenemos:

$$\psi_{11} y^*_{i1} = \alpha\, \psi_{12} y^*_{i1}$$

$$\psi'_{21} \begin{bmatrix} y^*_{i1} \\ y^*_{i2} \end{bmatrix} = \alpha\, \psi'_{22} \begin{bmatrix} y^*_{i1} \\ y^*_{i2} \end{bmatrix} \qquad [92]$$

en donde $\psi'_{11} = \pi_{31} - \pi_{21}$, $\psi_{12} = \pi_{21} - 1$, etc. Las ecuaciones [92] sugieren estimar α utilizando

$$\hat{\alpha} = \frac{\hat{\psi}_{12} A_{11} \hat{\psi}_{11} + \hat{\psi}'_{22} A_{22} \hat{\psi}_{21} + \hat{\psi}_{12} A_{12} \hat{\psi}_{21} + \hat{\psi}'_{22} A_{21} \hat{\psi}_{11}}{\hat{\psi}_{12} A_{11} \hat{\psi}_{12} + \hat{\psi}'_{22} A_{22} \hat{\psi}_{22} + \hat{\psi}_{12} A_{12} \hat{\psi}_{22} + \hat{\psi}'_{22} A_{21} \hat{\psi}_{12}} \qquad [93]$$

en donde los $\hat{\psi}$ son estimaciones consistentes de los ψ y los A_{ts} son ponderaciones. ψ_{11} y ψ_{12} se pueden estimar por ML en el sub-panel con observaciones $y^*_{i1} > 0$, mientras que ψ_{21} y ψ_{22} se pueden estimar en el sub-panel con $y^*_{i1} > 0$ y $y^*_{i2} > 0$. Por tanto, la «forma reducida» para $T = 4$ está formada por las dos ecuaciones en [89] y por

$$E[y^*_{it} | y^*_{i1}, y^*_{i2}] = [y^*_{i1}\, y^*_{i2}]' \pi_{t2} \qquad [94]$$

Finalmente, nótese que en el caso en que $y^*_{it} = y_{it}$ (y por tanto el modelo es lineal), es natural utilizar estimaciones MCO de π_{t1} y π_{t2}. Se puede comprobar que en esta situación el estimador [93] coincide con el estimador MGM de Arellano y Bond (1988a). No obstante, hay una diferencia sustancial entre las condiciones necesarias para que un estimador del tipo [93] sea consistente en los modelos VDL en comparación con las del modelo lineal. En concreto, en el modelo lineal no es necesario que las esperanzas condicionales [89] y [94] sean lineales; estas expresiones se pueden interpretar simplemente como predictores lineales. Por otra parte, las estimaciones MCO de π_{t1} y π_{t2} son consistentes para una clase muy amplia de distribuciones y situaciones de heterogeneidad, mientras que los estimadores ML utilizados en el caso VDL no son en general robustos a errores de especificación en la distribución de las perturbaciones y de los efectos.

6. Conclusiones

Para el análisis econométrico de modelos de comportamiento individual es natural utilizar datos sobre las características y las decisiones de familias y empresas, puesto que estas observaciones suelen ser las que más se aproximan al nivel al que han sido formuladas las teorías en las que se basan los modelos. Por otra parte, con frecuencia estamos interesados en estudiar la variación temporal en el comportamiento de individuos heterogéneos, en cuyo caso las muestras de corte transversal que contienen una sola observación por individuo resultan insuficientes. Por ejemplo, es muy difícil contrastar un modelo de consumo basado en la teoría del ciclo vital con datos agregados puesto que es difícil encontrar razones para que la teoría se cumpla en los datos agregados. Sin embargo, es también difícil llevar a cabo un contraste utilizando datos provenientes de encuestas en ausencia de series temporales de consumo individuales. En consecuencia, evitar problemas de agregación y facilitar el seguimiento del comportamiento individual en el tiempo son dos grandes ventajas de los datos de panel sobre los datos de series temporales y los datos de corte transversal, respectivamente. Naturalmente, la importancia cualitativa de estas ventajas dependerá de las aplicaciones y de los conjuntos de datos específicos que se consideren.

La otra gran ventaja de los datos de panel es que permiten la estimación de modelos que tienen en cuenta diferencias permanentes entre los individuos aunque estas no se observen. La relevancia práctica de este punto ha sido demostrada repetidas veces: son muy numerosas las aplicaciones en que los resultados de regresiones basadas en la variación intra-individual son significativamente distintos de los resultados basados en la variación entre individuos. Estas discrepancias sugieren que en las regresiones de corte transversal el supuesto *ceteris paribus* no se cumple, porque los regresores están correlacionados con las diferencias inobservables incluidas en los términos de perturbación.

La econometría de datos de panel se suele asociar a los modelos con efectos individuales, ya que son los que requieren el desarrollo de métodos específicos de inferencia. La mayor parte de este trabajo se ha dedicado a exponer los problemas que surgen en una serie de modelos econométricos habituales (modelos de regresión estáticos, dinámicos, con errores en las variables y modelos VDL) cuando se incluyen efectos individuales, y también a presentar las soluciones disponibles. La capacidad de los datos de panel para obtener estimaciones consistentes en presencia de efectos individuales varía de unos modelos a otros, y en algunos casos, como en los modelos con errores de medida, los sesgos del segundo tipo pueden resultar exacerbados como consecuencia de los intentos para corregir los primeros.

Referencias

Abowd, J. M. y Card, D. (1989): «On the Covariance Structure of Earnings and Hours Changes», *Econometrica, 57,* págs. 411-445.

Amemiya, T. (1973): «Regression Analysis When the Dependent Variable is Truncated Normal», *Econometrica, 41,* págs. 997-1016.

Amemiya, T. (1985): *Advanced Econometrics,* Blackwell, Oxford.

Amemiya, T. y MaCurdy, T. E. (1986): «Instrumental-Variable Estimation of an Error-Components Model», *Econometrica, 54,* págs. 869-881.

Andersen, E. B. (1970): «Asymptotic Properties of Conditional Maximum Likelihood Estimators», *Journal of the Royal Statistical Society,* Series B, *32,* págs. 283-301.

Andersen, E. B. (1973): *Conditional Inference and Models for Measuring,* Mentalhygiejnisk Forsknings Institut, Copenhagen.

Anderson, T. W. y Hsiao, C. (1981): «Estimation of Dynamic Models with Error Com ponents», *Journal of the American Statistical Association, 76,* págs. 598-606.

Anderson, T. W. y Hsiao, C. (1982): «Formulation and Estimation of Dynamic Models Using Panel Data», *Journal of Econometrics, 18,* págs. 47-82.

Arellano, M. (1988): «An Alternative Transformation for Fixed Effects Models with Predetermined Variables», Applied Economics Discussion Paper, *57,* Oxford.

Arellano, M. (1989a): «An Efficient GLS Estimator of Triangular Models with Covariance Restrictions», *Journal of Econometrics, 42,* págs. 267-273.

Arellano, M. (1989b): «On the Efficient Estimation of Simultaneous Equations with Covariance Restrictions», *Journal of Econometrics, 42,* págs. 247-265.

Arellano, M. (1989c): «A Note on the Anderson-Hsiao Estimator for Panel Data», *Economics Letters, 31,* págs. 337-341.

Arellano, M. (1990): «Testing for Autocorrelation in Dynamic Random Effects Models», *Review of Economic Studies, 57,* págs. 127-134.

Arellano, M. y Bond, S. (1988a): «Some Tests of Specification for Panel Data: Monte Carlo Evidence and an Application to Employment Equations», Applied Economics Discussion Paper 55, Oxford.

Arellano, M. y Bond, S. (1988b): «Dynamic Panel Data Estimation Using DPD - A Guide for Users», Institute for Fiscal Studies Working Paper 88/15, London.

Arellano, M. y Bover, O. (1989): «Another Look at the Instrumental-Variable Estimation of Error-Components Models», London School of Economics, mimeo.

Ashenfelter, O.; Deaton, A. y Solon, G. (1986): «Collecting Panel Data in Developing Countries: Does It Make Sense?», LSMS WP 23, The World Bank, Washington.

Avery, R. B.; Hansen, L. P. y Hotz, V. J. (1983): «Multiperiod Probit Models and Orthogonality Condition Estimation», *International Economic Review, 24,* págs. 21-35.

Balestra, P. y Nerlove, M. (1966): «Pooling Cross Section and Time Series Data in the Estimation of a Dynamic Model: The Demand for Natural Gas», *Econometrica, 34,* págs. 585-612.

Bhargava, A. y Sargan, J. D. (1983): «Estimating Dynamic Random Effects Models from Panel Data Covering Short Time Periods», *Econometrica, 51,* págs. 1635-1659.

Bover, O. (1989): «Estimating Intertemporal Labour Supply Elasticities Using Structural Models», *Economic Journal, 99,* págs. 1026-1039.

Bover, O. y Arellano, M. (1988): «Estimating Dynamic Limited Dependent Variable Models from Panel Data with and Application to Female Labour Supply, Nuffield College, Oxford.

Bowden, R. J. (1989): *Statistical Games and Human Affairs: The View from Within,* Cambridge University Press, New York.

Breusch, T. S.; Mizon, G. E. y Schmidt, P. (1989): «Efficient Estimation Using Panel Data», *Econometrica, 57,* págs. 695-700.

Browning, M. J.; Deaton, A. S. y Irish, M. (1985): «A Profitable Approach to Labour Supply and Commodity Demands over the Life-Cycle», *Econometrica, 53,* págs. 503-544.

Butler, J. S. y Moffitt, R. (1982): «A Computationally Efficient Quadrature Procedure for the One-Factor Multinomial Probit Model», *Econometrica, 50,* págs. 761-764.

Chamberlain, G. (1980): «Analysis of Covariance with Qualitative Data», *Review of Economic Studies, 47,* págs. 225-238.

Chamberlain, G. (1982): «Multivariate Regression Models for Panel Data», *Journal of Econometrics, 18,* págs. 5-46.

Chamberlain, G. (1984): «Panel Data», in Z. Griliches and M. D. Intrilligator (eds.), *Handbook of Econometrics,* vol. II, Elsevier Science.

Chowdhury, G. y Nickell, S. (1985): «Hourly Earnings in the United States: Another Look at Unionization, Schooling, Sickness, and Unemployment Using PSID Data», *Journal of Labor Economics, 3,* págs. 38-69.

Cox, D. R. (1988): «Some Aspects of Conditional and Asymptotic Inference: A Review», *Sankhya.*

Cox, D. R. y Reid, N. (1987): «Parameter Orthogonality and Approximate Conditional Inference», *Journal of the Royal Statistical Society,* Series B, *49,* págs. 1-39.

Cruddas, A. M.; Reid, N. y Cox, D. R. (1989): «A Time Series Illustration of Approximate Conditional Likelihood», *Biometrika, 76,* págs. 231-237.

Deaton, A. (1985): «Panel Data from Time Series of Cross-Sections», *Journal of Econometrics, 30,* págs. 109-126.

Gourieroux, C.; Monfort, A. y Trognon, A. (1984): «Pseudo Maximum Likelihood Methods: Applications to Poisson Models», *Econometrica, 52,* págs. 701-720.

Gourieroux, C.; Monfort, A. y Trognon, A. (1985): «Moindres Carrés Asymptotiques», *Annales de l'Inséé, 58,* págs. 91-122.

Griliches, Z. (1977): «Estimating the Returns to Schooling: Some Econometric Problems», *Econometrica, 45,* págs. 1-22.

Griliches, Z. y Hausman, J. A. (1986): «Errors in Variables in Panel Data», *Journal of Econometrics, 31,* págs. 93-118.

Hajivassiliou, V. A. y McFadden, D. L. (1989): «Country Heterogeneity and External Debt Crises: Estimation by the Method of Simulated Moments», mimeo.

Hall, B. H. (1987): «The Relationship Between Firm Size and Firm Growth in the US Manufacturing Sector», *Journal of Industrial Economics, 35,* págs. 583-606.

Hall, R. y Mishkin, F. (1982): «The Sensitivity of Consumption to Transitory Income: Estimates from Panel Data on Households», *Econometrica, 50,* págs. 461-481.

Hansen, L. P. (1982): «Large Sample Properties of Generalized Method of Moments. Estimators», *Econometrica, 50,* págs. 1029-1054.

Hansen, L. P. y Singleton, K. J. (1982): «Generalized Instrumental Variables Estimation of Nonlinear Rational Expectations Models», *Econometrica, 50,* págs. 1269-1286.

Hausman, J. A. (1978): «Specification Tests in Econometrics», *Econometrica, 46,* págs. 1251-1272.

Hausman, J. A.; Hall, B. H. y Griliches, Z. (1984): «Econometric Models for Count Data with an Application to the Patents - R & D Relationship», *Econometrica, 52,* págs. 909-938.

Hausman, J. A. y Taylor, W. E. (1981): «Panel Data and Unobservable Individual Effects», *Econometrica, 49,* págs. 1377-1398.

Hausman, J. A. y Wise, D. A. (1979): «Attrition Bias in Experimental and Panel Data: The Gary Income Maintenance Experiment», *Econometrica, 47,* págs. 455-473.

Hayashi, F. y Inoue, T. (1989): «The Firm Growth-Q Relation with Multiple Capital Goods: Theory and Evidence from Panel Data on Japanese Firms», mimeo.

Heckman, J. J. (1974): «Shadow Prices, Market Wages, and Labor Supply», *Econometrica, 42,* págs. 679-693.

Heckman, J. J. (1976): «The Common Structure of Statistical Models of Truncation, Sample Selection and Limited Dependent Variables and a Simple Estimator for Such Models», *Annals of Economic and Social Measurement, 5,* págs. 475-492.

Heckman, J. J. (1979): «Sample Selection Bias as a Specification Error», *Econometrica, 47,* págs. 153-161.

Heckman, J. J. (1981a); «Statistical Models for Discrete Panel Data», in C. F. Manski y

D. McFadden (eds.): *Structural Analysis of Discrete Data with Econometric Applications,* MIT Press, Cambridge, Mass.

Heckman, J. J. (1981b): «The Incidental Parameters Problem and the Problem of Initial Conditions in Estimating a Discrete Time-Discrete Data Stochastic Process», in C. F. Manski y D. McFadden (eds.): *Structural Analysis of Discrete Data with Econometric Applications,* MIT Press, Cambibrdge, Mass.

Heckman, J. J. y MaCurdy, T. E. (1980): «A Life-Cycle Model of Female Labour Supply», *Review of Economic Studies, 47,* págs. 47-74.

Hotz, V. J. y Miller, R. A. (1888): «An Empirical Analysis of Life Cycle Fertility and Female Labor Supply», *Econometrica, 56,* págs. 91-118.

Hotz, V. J.; Kydland, F. E. y Sedlacek, G. L. (1988): «Intertemporal Preferences and Labor Supply», *Econometrica, 56,* págs. 335-360.

Holtz-Eakin, D. (1988): «Testing for Individual Effects in Autoregressive Models», *Journal of Econometrics, 39,* págs. 297-307.

Holtz-Eakin, D.; Newey, W. y Rosen, H. S. (1988): «Estimating Vector Autoregressions with Panel Data», *Econometrica, 56,* págs. 1371-1395.

Hsiao, C. (1985): «Benefits and Limitations of Panel Data», *Econometric Review, 4 (1),* págs. 121-174.

Hsiao, C. (1986): *Analysis of Panel Data,* Cambridge University Press.

Joreskog, K. y Goldberger, A. (1975): «Estimation of a Model with Multiple Indicators and Multiple Causes of a Single Latent Variable Model», *Journal of the American Statistical Association, 70,* págs. 631-639.

Kiefer, N. M. (1980): «Estimation of Fixed Effect Models for Time Series of Cross-Sections with Arbitrary Intertemporal Covariance», *Journal of Econometrics, 14,* págs. 195-202.

Kiefer, N. M. (1989): «The ET Interview: Arthur S. Goldberger», *Econometric Theory, 5,* págs. 133-160.

MaCurdy, T. E. (1981): «An Empirical Model of Labor Supply in a Life-Cycle Setting», *Journal of Political Economy, 89,* págs. 1059-1085.

Manski, C. (1985): «Semiparametric Analysis of Discrete Response: Asymptotic Properties of the Maximum Score Estimator», *Journal of Econometrics, 27,* págs. 313-333.

Manski, C. (1987): «Semiparametric Analysis of Random Effects Linear Models from Binary Panel Data», *Econometrica, 55,* págs. 357-362.

Mauleón, I. (1987): «Problemas Prácticos en el Tratamiento Econométrico de Datos Cross-Section», *Investigaciones Económicas,* vol. XI, *núm. 1,* págs. 41-94.

McCullagh, P. y Nelder, J. A. (1983): *Generalized Linear Models,* Chapman and Hall, London.

McFadden, D. (1989): «A Method of Simulated Moments for Estimation of Multinomial Probits without Numerical Integration», *Econometrica, 57,* págs. 995-1026.

Mundlak. Y. (1978): «On the Pooling of Time Series and Cross Section Data», *Econometrica, 46,* págs. 69-85.

Neyman, J. y Scott, E. L. (1948): «Consistent Estimates Based on Partially Consistent Observations», *Econometrica, 16,* págs. 1-32.

Nickell, S. (1981): «Biases in Dynamic Models with Fixed Effects», *Econometrica, 49,* págs. 1417-1426.

Pakes, A. y Ericson, R. (1987): «Empirical Implications of Alternative Models of Firm Dynamics», SSRI DP 8803, University of Wisconsin.

Robinson, P. M. (1987): «Asymptotically Efficient Estimation in the Presence of Heteroskedasticity of Unknown Form», *Econometrica, 55,* págs. 875-891.

Sargan, J. D. (1958); «The Estimation of Economic Relationships Using Instrumental Variables», *Econometrica, 26,* págs. 393-415.

Sargan, J. D. (1980); «Some Tests of Dynamic Specification for a Single Equation», *Econometrica, 48,* págs. 879-897.

Sargan, J. D. (1988a): «Testing for Misspecification after Estimating Using Instrumental Variables», en E. Maasoumi (ed.): *Contributions to Econometrics: John Denis Sargan,* vol. 1, Cambridge University Press.

Sargan, J. D. (1988b): *Lectures on Advanced Econometric Theory,* Blackwell, Oxford.

Stoker, T. M. (1986): «Simple Tests of Distributional Effects on Macroeconomic Equations», *Journal of Political Economy, 94,* págs. 763-795.

Tobin, J. (1958): «Estimation of Relationships for Limited Dependent Variables», *Econometrica, 26,* págs. 24-36.

White, H. (1980): «A Heteroskedasticity-Consistent Covariance Matrix Estimator and a Direct Test for Heteroskedasticity», *Econometrica, 48,* págs. 817-838.

Zeldes, S. P. (1989): «Consumption and Liquidity Constraints: An Empirical Investigation», *Journal of Political Economy, 97,* págs. 305-346.

Abstract

This article presents a synthesis of the methods that are available for the econometric analysis of panel data in a unified framework. In particular, we analyse the form in which the properties of the various estimators depend on the assumptions about explanatory variables, permanent unobservable effects and disturbance terms. Firstly we study both static and dynamic linear models with individual effects. Subsequently the analysis is extended to limited dependent variable models with individual effects and dynamic responses.

Recepción del original, octubre de 1989
Versión final, diciembre de 1989