Capítulo 1
INTRODUCCIÓN AL ANÁLISIS ECONOMÉTRICO CON DATOS DE PANEL

Manuel Arellano *
Centro de Estudios Monetarios y Financieros (CEMFI)

Prólogo

Conocí a Gonzalo Mato en 1986 en la Universidad de Oxford. Gonzalo se disponía a iniciar su trabajo sobre la demanda de inversión utilizando datos de empresas españolas, bajo la supervisión de Steve Nickell. Cuando llegó a Oxford, Gonzalo ya tenía experiencia en el uso de técnicas econométricas para datos de panel, y además un gran interés —que yo compartía— por profundizar sus conocimientos en este campo. Desde el principio, las conversaciones con Gonzalo fueron enormemente fructíferas. De hecho, el primer artículo que publiqué en inglés fue la respuesta a una pregunta de Gonzalo. El interés de Gonzalo por los métodos econométricos no era teórico o formal, sino que trataba de ver si las técnicas le podían ser de utilidad en sus trabajos, los cuales siempre respondían a un objetivo bien definido. Yo creo que Gonzalo siempre tuvo claro lo que tenía que hacer, y esa claridad de fondo le permitía ser original, emprendedor y, a veces, hasta osado. Recuerdo que no tuvo ningún inconveniente en presentar, al mes de su llegada a Oxford, un seminario sobre los determinantes de la concentración industrial en España. Más allá de las afinidades académicas, Gonzalo era un amigo con el que me entendía realmente bien, con el que se operaba la «química» de la que hablan en Inglaterra. La penúltima vez que lo vi fue en Venecia, gracias a una conferencia sobre el análisis de datos de panel. Nos despedimos por la noche en un *vaporetto* y me dijo que se volvía a

* Agradezco los comentarios de Víctor Aguirregabiria, César Alonso, Olympia Bover, Dolores Collado y Juan José Dolado, así como la excelente labor mecanográfica de Carmina Arellano.

Madrid sin asistir a la última sesión porque añoraba mucho a sus hijos Jaime y Marta.

Este capítulo, como el resto del libro, está dedicado a la memoria de Gonzalo Mato. El objeto del trabajo es presentar una introducción al análisis econométrico con datos de panel y se basa en unas sesiones organizadas por el Servicio de Estudios del Banco de España en diciembre de 1991. Se intenta poner énfasis en las ideas clave, contraponiéndolas a otras situaciones en econometría, y se han minimizado los argumentos formales y desarrollos matemáticos.

1. Introducción: regresiones simples *versus* regresiones múltiples

Se habla de datos de panel cuando tenemos observaciones repetidas a lo largo del tiempo para una muestra de unidades individuales. Podemos decir que para una variable y_{it} tenemos $i = 1 \dots N$ individuos observados a lo largo de $t = 1 \dots T$ períodos de tiempo (i puede referirse a países, regiones, industrias, empresas o familias, entre otros).

Empezaremos repasando las diferencias entre coeficientes de regresión simple y coeficientes de regresión múltiple (o entre coeficientes alternativos de regresión múltiple basados en un conjunto de regresores y en un subconjunto de los mismos, respectivamente). Esto se hace por dos motivos:

1) A menudo el análisis de regresión se motiva en términos del ajuste de líneas o planos por mínimos cuadrados a nubes de puntos. Por el contrario, aquí queremos poner énfasis en la idea de regresión como media condicionada (o como aproximación lineal a una media condicionada). Esto es, una regresión describe cómo cambia la media para distintos subgrupos de la población en la forma especificada por los valores de los regresores.
2) La ventaja principal de los datos de panel es que nos permiten estimar coeficientes de regresión múltiple que no se podrían estimar con datos de corte transversal o con datos de series temporales.

Consideremos las dos regresiones siguientes:

$$E(y \mid x_1, x_2) = \beta_0 + \beta_1 x_1 + \beta_2 x_2$$
$$E(y \mid x_1) = \gamma_0 + \gamma_1 x_1.$$

Para simplificar la presentación, supongamos que x_1, x_2 son binarias. Por ejemplo,

y = número de horas anuales trabajadas por una mujer casada

$$x_1 = \text{salario de la mujer} = \begin{cases} 0 \text{ salario bajo} \\ 1 \text{ salario alto} \end{cases}$$

$$x_2 = \text{ingresos del marido} = \begin{cases} 0 \text{ ingresos bajos} \\ 1 \text{ ingresos altos} \end{cases}$$

En este caso, β_0, β_1, β_2 especifican las horas trabajadas en promedio por cada uno de cuatro subgrupos de la población:

	marido pobre	marido rico			
Salario bajo	β_0	$\beta_0 + \beta_2$	=	$\bar{y}_{00}$	$\bar{y}_{01}$
Salario alto	$\beta_0 + \beta_1$	$\beta_0 + \beta_1 + \beta_2$		$\bar{y}_{10}$	$\bar{y}_{11}$

En la regresión simple, γ_0 y γ_1 especifican las horas trabajadas en promedio para dos subgrupos de la población:

Salario bajo	γ_0
Salario alto	$\gamma_0 + \gamma_1$

(En el caso en que x_1, x_2 no sean binarias pero discretas con soporte finito el análisis es similar pero el número de celdas se multiplica.)

Sea π_0 la proporción de maridos ricos en el grupo de mujeres con salario bajo [con lo cual $(1 - \pi_0)$ es la proporción de maridos pobres en ese mismo grupo], y sea $p_1 = \pi_0 + \pi_1$ la proporción de maridos ricos en el grupo de mujeres con salario alto (π_1 denota la diferencia entre p_1 y π_0).

De esta manera podemos relacionar las medias de los dos cuadros anteriores de la forma siguiente:

$$\gamma_0 = \beta_0(1 - \pi_0) + (\beta_0 + \beta_2)\pi_0 = \beta_0 + \beta_2\pi_0$$

$$\gamma_0 + \gamma_1 = (\beta_0 + \beta_1)(1 - p_1) + (\beta_0 + \beta_1 + \beta_2)p_1 = \beta_0 + \beta_1 + \beta_2 p_1 =$$
$$= (\beta_0 + \beta_2\pi_0) + (\beta_1 + \beta_2\pi_1).$$

Por tanto [1],

$$\gamma_1 = (\beta_1 + \beta_2\pi_1).$$

Empíricamente se observa que $\pi_0 < p_1$, o lo que es lo mismo, $\pi_1 > 0$. Por otra parte esperamos que $\beta_1 > 0$ (nótese que $\beta_1 = \bar{y}_{10} - \bar{y}_{00} = \bar{y}_{11} - \bar{y}_{01}$ es la diferencia de las horas trabajadas en promedio entre las mujeres de salario alto y las de salario bajo para cada uno de los subgrupos), y también que $\beta_2 < 0$ ($\beta_2 = \bar{y}_{01} - \bar{y}_{00} = \bar{y}_{11} - \bar{y}_{10}$). Sin embargo, el signo de γ_1 es incierto y puede ocurrir que $\gamma_1 < 0$ (en el caso en que $\beta_1 < -\beta_2\pi_1$).

Aunque γ_1 puede tener interés estadístico (esto es, saber si trabajan más las mujeres con salario alto, o viceversa), por lo general muchas veces el interés principal se centra en β_1 (esto es, si aumentamos el salario efectivo de todas las mujeres, con independencia de si sus maridos son ricos o pobres, ¿cuál es el efecto sobre las horas trabajadas?).

En cada caso, si disponemos de una muestra aleatoria con observaciones sobre y, x_1, x_2, o bien de muestras aleatorias para cada uno de los subgrupos, podemos estimar los coeficientes β y γ calculando las medias muestrales para cada uno de los subgrupos.

2. El control de la heterogeneidad inobservable constante

Una de las ventajas más importantes de los datos de panel con respecto a otros tipos de datos es que nos permiten controlar diferencias inobservables.

Supongamos que estamos interesados en la regresión lineal

$$E(y \mid x_1, x_2) = \beta_0 + \beta_1 x_1 + \beta_2 x_2.$$

Como refleja la discusión anterior, si tenemos una muestra de corte transversal con observaciones de y, x_1, x_2 podemos obtener estimaciones

[1] Esto es,

$E(x_2 \mid x_1) = \pi_0 + \pi_1 x_1$ en donde tenemos:

$$E(x_2 \mid x_1 = 0) = \text{Prob}\ (x_2 = 1 \mid x_1 = 0) = \pi_0$$
$$E(x_2 \mid x_1 = 1) = \text{Prob}\ (x_2 = 1 \mid x_1 = 1) = \pi_0 + \pi_1.$$

Por tanto,

$$\begin{aligned} E(y \mid x_1) &= E(E(y \mid x_1, x_2) \mid x_1) = \beta_0 + \beta_1 x_1 + \beta_2 E(x_2 \mid x_1) = \\ &= (\beta_0 + \beta_2\pi_0) + (\beta_1 + \beta_2\pi_1)x_1 = \gamma_0 + \gamma_1 x_1 \end{aligned}$$

consistentes de β_0, β_1, β_2. Supongamos que x_2 no se observa. Con un solo corte transversal ya no podemos estimar β_1 (consistentemente), excepto si $\beta_2 = 0$ o $\mathrm{Cov}(x_1, x_2) = 0$, esto es, si $\pi_1 = 0$ en el ejemplo anterior.

Sin embargo, si para cada individuo en la muestra tenemos dos o más observaciones temporales, bajo determinadas condiciones puede ser posible estimar (consistentemente) β_1, aunque no se observe x_2 [y $\beta_2 \neq 0$, $\mathrm{Cov}(x_1, x_2) \neq 0$]. Consideremos el caso en que $T = 2$ y supongamos que

$$E(y_1 \mid x) = \beta_0 + \beta_1 x_{11} + \beta_2 x_{21}$$
$$E(y_2 \mid x) = \beta_0 + \beta_1 x_{12} + \beta_2 x_{22},$$

en donde

$$x = (x_{11}, x_{12}, x_{21}, x_{22})'.$$

Nótese que estamos suponiendo que los $\beta's$ son los mismos para los dos períodos y que la regresión de cada período sólo depende de los valores contemporáneos de x (estos supuestos no son esenciales para el argumento que vamos a presentar, pero si prescindimos de ellos tendríamos modelos más complicados).

Supongamos además que $x_{21} = x_{22} = x_2$ (esto es, x_2 tiene variación de corte transversal pero es constante en el tiempo para un individuo determinado). Entonces tenemos que

$$E(y_2 - y_1 \mid x) = \beta_1(x_{12} - x_{11}).$$

Esto es [2],

$$\beta_1 = \frac{E[(x_{12} - x_{11})(y_2 - y_1)]}{E[(x_{12} - x_{11})^2]},$$

---

[2] Nótese que si tenemos $E(y \mid x) = \beta x$ entonces,

$$E(h(x)\,(y - \beta x)) = E\,(h(x)E(y - \beta x \mid x)) = 0$$

para cualquier función $h(x)$. Por lo cual β satisface

$$\beta = E(h(x)y)/E(h(x)x).$$

(En la sección 4 se retoma esta discusión.)

Si $E(.\mid x)$ denota un predictor lineal óptimo en lugar de una esperanza condicionada, este resultado sigue siendo válido para $h(x) = x$. De hecho, las conclusiones básicas de este trabajo permanecen inalteradas si reemplazamos medias condicionadas por predictores lineales en cada caso.

que se puede estimar (consistentemente) reemplazando esperanzas por medias muestrales, supuesta la disponibilidad de un panel con dos observaciones de tiempo para cada individuo.

Denotemos $\eta = \beta_2 x_2$. A la variable η la denominamos «efecto individual» o «efecto permanente» y representa diferencias inobservables potencialmente correlacionadas con x_{11} y x_{12} sobre las que estamos interesados en condicionar, para poder estimar el coeficiente de regresión múltiple β_1. Nótese que β_1 es el coeficiente de regresión múltiple, no sólo con respecto a x_2, sino con respecto a todas las variables constantes en el tiempo.

Ejemplos

Nótese que hasta ahora habíamos omitido el subíndice de individuos para simplificar la notación. A continuación vamos a considerar tres modelos distintos que dan lugar a ecuaciones del tipo

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

1) Función de producción Cobb-Douglas de un producto agrícola (Mundlak)

y_{it} = logaritmo de la producción
x_{it} = logaritmo de un *input* (trabajo)
η_i = calidad del suelo (*input* constante a lo largo del tiempo)
v_{it} = diferencias aleatorias de y_{it} con respecto a $E(y_{1t} \mid x_i, \eta_i)$ que representan un *input* aleatorio (lluvia) fuera del control del productor.

Supongamos que η_i es conocido por el productor pero no por el económetra. Si el productor maximiza beneficios esperados es fácil ver que la demanda de trabajo x_{it} estará correlacionada con η_i (véase Chamberlain, 1984).

2) Oferta de trabajo intertemporal

y_{it} = horas trabajadas
x_{it} = salarios
η_i = función de la utilidad marginal de la riqueza, que a su vez depende de los salarios futuros, activos y tipos de interés, de nuevo inobservable y correlacionada con x_{it}.

β = elasticidad intertemporal del salario (mide el efecto de un cambio a corto plazo en los salarios sobre la oferta de trabajo manteniendo constante el perfil de salarios futuros —un parámetro de importancia macroeconómica—. Véase MaCurdy, 1981).

3) Ingresos individuales y rendimientos de la educación

y_{it} = logaritmo de los ingresos
x_{it} = x_i = educación (años de)
η_i = «habilidad», que habitualmente se supone correlacionada con el nivel de educación
β = mide los «rendimientos» de la educación

En este caso cuando diferenciamos la ecuación también desaparece $x_{it}\beta$. De hecho, los datos de panel no son tan útiles en este caso como cuando se puede explotar la variación temporal para separar la variación permanente de corte transversal (véase Griliches, 1977).

La relevancia práctica de la discusión precedente sobre la heterogeneidad inobservable se ha demostrado repetidas veces: hay muchas aplicaciones en las que los resultados de regresiones en niveles y en diferencias o desviaciones son claramente distintos. La interpretación de estas discrepancias depende de cada caso, pero a menudo sugiere que en las regresiones de corte transversal el supuesto de *ceteris paribus* no se satisface debido a que los regresores están correlacionados con características individuales inobservables incluidas en los términos de perturbación.

Medición econométrica *versus* problemas de predicción

La discusión anterior revela que las ventajas vistas por ahora aparecen básicamente en el contexto de *problemas de medición econométrica*, en contraposición a *problemas de predicción*. Esta distinción es importante. Incluyendo efectos individuales conseguimos estimar ciertos coeficientes a costa de dejar una parte de la regresión sin modelizar. El «R^2» que obtenemos en la población viene dado por

$$R^2 = \beta^2 \frac{Var(x)}{Var(y)},$$

que puede ser muy bajo si la varianza de η es grande.

Podríamos obtener modelos alternativos con R^2 más altos fácilmente.

Por ejemplo, incluyendo regresores constantes en el tiempo z_i que «explicaran» en parte el efecto individual:

$$E(y_{it} \mid x_i, z_i) = \delta_0 + \delta_1 x_{it} + \delta_2 z_i$$

o bien modelos autorregresivos, por ejemplo,

$$E(y_{it} \mid y_{it-1}, x_i) = \alpha_0 + \alpha_1 y_{i(t-1)} + \alpha_2 x_{it}$$

Sin embargo, en general estos modelos, aunque más apropiados para problemas de predicción (por ejemplo, detección del fraude fiscal, en el primer caso), no serían útiles para estimar el coeficiente β de la discusión anterior.

3. Modelos estáticos: discusión general

Consideremos el sistema de T ecuaciones

$$E(y_{it} \mid x_i, \eta_i) = x_{it}'\beta + \eta_i \qquad (t = 1 \ldots T),$$

en donde

$$x_i = \begin{pmatrix} x_{i1} \\ \vdots \\ x_{iT} \end{pmatrix}, \qquad \beta = \begin{pmatrix} \beta_1 \\ \vdots \\ \beta_k \end{pmatrix}, \qquad x_{it} = \begin{pmatrix} x_{1it} \\ \vdots \\ x_{kit} \end{pmatrix}.$$

Equivalentemente podemos escribir

$$E(\Delta y_{it} \mid x_i, \eta_i) = \Delta x_{it}'\beta = E(\Delta y_{it} \mid x_i) \quad (t = 2, \ldots, T)$$

con $\Delta y_{it} = y_{it} - y_{i(t-1)}$

junto con

$$E(\bar{y}_i \mid x_i, \eta_i) = \bar{x}_i'\beta + \eta_i,$$

en donde

$$\bar{y}_i = (1/T)\Sigma_t y_{it}, \quad \text{etc.}$$

Sea v_{it} la desviación de y_{it} con respecto a su media condicionada dados x_i y η_i

$$v_{it} = y_{it} - E(y_{it} \mid x_i, \eta_i).$$

Si estas desviaciones son «clásicas» en el sentido de que $v_{it} \mid x_i, \eta_i \sim iid(0, \sigma^2)$, las desviaciones en las ecuaciones en primeras diferencias $\Delta y_{it} - \Delta x_{it}'\beta$ estarán autocorrelacionadas. Una transformación lineal de las ecuaciones en primeras diferencias que elimina la autocorrelación es la siguiente:

$$E(y_{it}^* \mid x_i, \eta_i) = x_{it}^{*\prime}\beta \qquad t = 1 \ldots (T-1),$$

en donde

$$y_{it}^* = c_t\left(y_{it} - \frac{1}{(T-t)}(y_{i(t+1)} + \ldots + y_{iT})\right)$$

con

$$c_t^2 = (T-t)/(T-t+1).$$

A la transformación y_{it}^* se le llama desviaciones ortogonales (véase Arellano y Bover, 1990a). Por tanto, el sistema de T ecuaciones se puede escribir de la forma siguiente:

$$\begin{aligned} y_i^* &= X_i^*\beta + v_i^* \\ \bar{y}_i &= \bar{x}_i'\beta + \eta_i + \bar{v}_i, \end{aligned}$$

en donde

$$y_i^* = (y_{i1}^* \ldots y_{i(T-1)}^*)', \; X_i^* = (x_{i1}^* \ldots x_{i(T-1)}^*)', \text{ etc.}$$

Sea

$$u_i^+ = \begin{pmatrix} v_i^* \\ \eta_i + \bar{v}_i \end{pmatrix}$$

y supongamos que $\mathrm{Var}(\eta_i \mid x_i) = \sigma_\eta^2$. Se puede comprobar que

$$\mathrm{Var}(u_i^+ \mid x_i) = \sigma^2 \begin{pmatrix} I_{T-1} & 0 \\ 0 & \dfrac{1}{\theta^2 T} \end{pmatrix}$$

con

$$\theta^2 = \frac{\sigma^2}{\sigma^2 + T\sigma_\eta^2}.$$

Obsérvese que la media condicionada de y_i^* dados x_i y η_i coincide con la media condicionada de y_i^* dado sólo x_i, puesto que la primera es independiente del valor que tome η_i.

$$E(y_i^* \mid x_i, \eta_i) = E(y_i^* \mid x_i) = X_i^* \beta,$$

pero, sin embargo, en general

$$E(\bar{y}_i \mid x_i, \eta_i) = \bar{x}_i' \beta + \eta_i \neq E(\bar{y}_i \mid x_i).$$

A continuación introduzcamos la siguiente notación para la media de $\bar{y}_i$ dado x_i

$$E(\bar{y}_i \mid x_i) = \bar{x}_i' \gamma.$$

[Nótese que estamos suponiendo que $E(\bar{y}_i \mid x_i) = E(\bar{y}_i \mid \bar{x}_i)$.]

Utilizando de nuevo el argumento de la nota 2 aplicado a vectores de coeficientes, se puede comprobar que β y γ satisfacen en la población las siguientes ecuaciones:

$$\beta = [E(X_i^{*\prime} X_i^*)]^{-1} E(X_i^{*\prime} y_i^*)$$
$$\gamma = [E(\bar{x}_i \bar{x}_i')]^{-1} E(\bar{x}_i \bar{y}_i).$$

Por otra parte, sabemos que $\gamma = \beta$ si $E(\eta_i \mid x_i) = 0$, esto es, si los efectos individuales son independientes en media de los regresores observables. Tenemos, pues, al menos tres estimadores distintos utilizando el principio de analogía [3]: el estimador «intra-grupos» (*within-groups*)

$$\hat{\beta}_{WG} = (X^{*\prime} X^*)^{-1} X^{*\prime} y^*,$$

el estimador «entre-grupos» (*between-groups*)

$$\hat{\gamma}_{BG} = (\bar{X}' \bar{X})^{-1} \bar{X}' \bar{y}$$

y el estimador de mínimos cuadrados generalizados (MCG)

$$\hat{\beta}_{MGC} = (X^{*\prime} X^* + \hat{\theta}^2 T \bar{X}' \bar{X})^{-1} (X^{*\prime} y^* + \hat{\theta}^2 T \bar{X}' \bar{y}).$$

---

[3] El principio de analogía es una regla natural para seleccionar estimadores. Un parámetro de la población es una característica definida en la distribución de la población. Para estimarlo utilizamos la misma característica pero definida en la muestra (véase Goldberger, 1991, y Manski, 1988).

El estimador MCG es mínimos cuadrados ponderados aplicado al sistema completo bajo el supuesto de que $\gamma = \beta$ y $\hat{\theta}^2$ es un estimador preliminar consistente de θ^2. En la notación anterior tenemos

$$X^{*\prime}X^{*} = \sum_{i=1}^{N} X_i^{*\prime}X_i^{*}, \; \bar{X}'\bar{X} = \sum_{i=1}^{n} \bar{x}_i\bar{x}_i', \text{ etc.}$$

(véase Arellano, 1993).

Por otra parte, las varianzas respectivas son:

$$V_{WG} = \mathrm{Var}(\hat{\beta}_{WG}) = \sigma^2(X^{*\prime}X^{*})^{-1}$$
$$V_{BG} = \mathrm{Var}(\hat{\gamma}_{BG}) = \sigma^2(\theta^2 T\bar{X}'\bar{X})^{-1}$$
$$V_{MCG} = \mathrm{Var}(\hat{\beta}_{MCG}) = \sigma^2(X^{*\prime}X^{*} + \theta^2 T\bar{X}'\bar{X})^{-1}.$$

Estas varianzas satisfacen la siguiente relación:

$$V_{MCG}^{-1} = V_{WG}^{-1} + V_{BG}^{-1},$$

por lo que se puede comprobar que $\hat{\beta}_{MCG}$ es una media ponderada de $\hat{\beta}_{WG}$ y de $\hat{\gamma}_{BG}$.

Efectos fijos *versus* efectos aleatorios

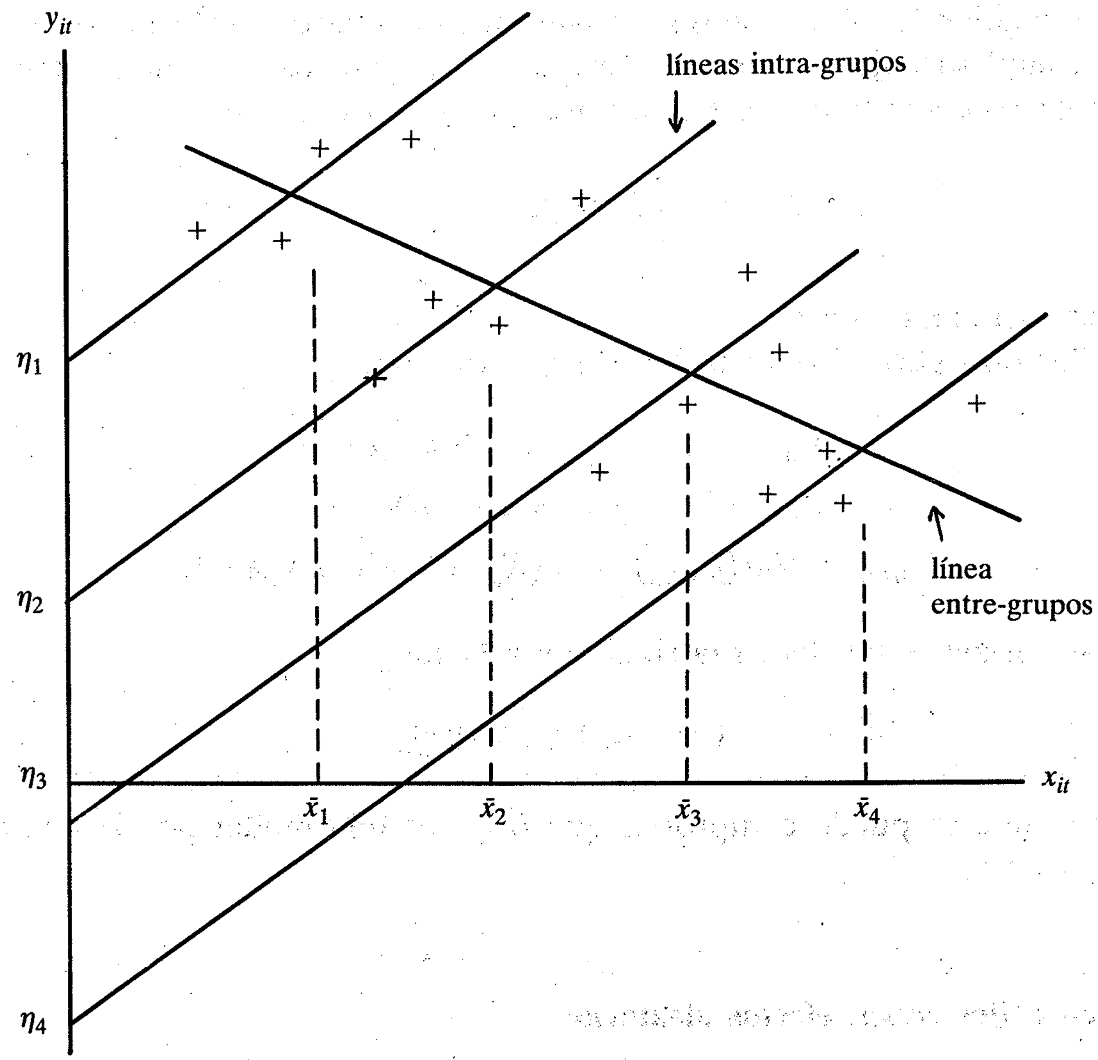
El estimador intragrupos se puede interpretar de formas alternativas:

(i) Como la regresión de $y_{it} - \bar{y}_i$ sobre $x_{it} - \bar{x}_i$
(ii) Como la regresión de y_{it} sobre x_{it} y variables ficticias individuales (efectos «fijos»).

El estimador $\hat{\beta}_{MCG}$ también se puede interpretar como MCG en un modelo de regresión con un error «clásico» compuesto: $u_{it} = \eta_i + v_{it}$ (efectos «aleatorios» no correlacionados). Véase el apéndice para una descripción del álgebra de estos dos modelos.

¿Cómo escogemos entre efectos «fijos» y efectos «aleatorios»? El problema no es si los efectos son fijos o aleatorios. De hecho, como revela la discusión anterior, los efectos pueden siempre considerarse aleatorios sin pérdida de generalidad. El problema es si los efectos están correlacionados con x_i o no, esto es, si la restricción $\beta = \gamma$ se cumple o no. En la interpretación de efectos fijos del estimador intragrupos, las realizaciones de los efectos en la muestra son tratadas como parámetros a estimar.

La situación se puede apreciar más claramente con el siguiente gráfico:



En este caso hay una correlación negativa entre η_i y $\bar{x}_i$. Incluso si los $\eta's$ son «fijos» en el sentido de que el muestreo es estratificado, en la población conjunta de $\eta's$ y $x's$ —a partir de la que se hace el muestreo— habrá una correlación negativa entre los dos.

El contraste de la restricción $\gamma = \beta$

El sistema no restringido se puede escribir

$$\begin{pmatrix} y_i^* \\ \bar{y}_i \end{pmatrix} = \begin{pmatrix} X_i^* & 0 \\ \bar{x}_i' & \bar{x}_i' \end{pmatrix} \begin{pmatrix} \beta \\ \delta \end{pmatrix} + \begin{pmatrix} v_i^* \\ \bar{u}_i \end{pmatrix},$$

en donde $\delta = \gamma - \beta$. Si estimamos este sistema por MCO obtenemos $\hat{\beta}_{WG}$ y $\hat{\delta} = \hat{\gamma}_{BG} - \hat{\beta}_{WG}$. A continuación podemos calcular el contraste de la *ji*-cuadrado (Wald) de la hipótesis nula H_0: $\delta = 0$:

$$h = (\hat{\gamma}_{BG} - \hat{\beta}_{WG})'(\hat{V}_{WG} + \hat{V}_{BG})^{-1}(\hat{\gamma}_{BG} - \hat{\beta}_{WG}).$$

Se puede comprobar que h también se puede escribir

$$h = (\hat{\beta}_{GLS} - \hat{\beta}_{WG})'(\hat{V}_{WG} - \hat{V}_{GLS})^{-1}(\hat{\beta}_{GLS} - \hat{\beta}_{WG}),$$

que es la forma de un *contraste de Hausman* (véanse Hausman, 1978, y Hausman y Taylor, 1981). Por otra parte, si los errores son heterocedásticos se pueden calcular versiones robustas de errores estándar y contrastes.

4. Variables instrumentales y el método generalizado de momentos

En preparación para el tratamiento de modelos dinámicos con efectos individuales, en esta sección presentamos la formulación general de los tres modelos básicos que utilizaremos: modelos de regresión, de variables instrumentales y modelos generalizados de momentos.

Regresión

Un modelo de regresión lineal especifica una media condicionada de la forma:

$$E(y \mid x) = \beta_0 + \beta_1 x.$$

Sea u la desviación de y con respecto a su media:

$$u = y - \beta_0 - \beta_1 x.$$

Por tanto,

$$E(u \mid x) = 0.$$

Este resultado implica:

1) $E(u) = 0$
2) $\text{Cov}(u, x) = E(u\, x) = E[E(u \mid x)\, x] = 0.$

Esto es, u y x no están correlacionados. A partir de (2) podemos despejar β_1:

$$\mathrm{Cov}(y - \beta_0 - \beta_1 x, x) = 0$$
$$\mathrm{Cov}(y,x) - \beta_1 \,\mathrm{Var}(x) = 0$$
$$\beta_1 = \frac{\mathrm{Cov}(y,x)}{\mathrm{Var}(x)}\,.$$

A continuación, utilizando (1) obtenemos β_0:

$$\beta_0 = E(y) - \frac{\mathrm{Cov}(y,x)}{\mathrm{Var}(x)} E(x),$$

Dado que β_0 y β_1 satisfacen las ecuaciones anteriores en la población, es natural estimarlos utilizando las mismas expresiones a nivel muestral.

Si x es un vector $kx1$, el resultado es similar:

$$\beta_1 = [\mathrm{Var}(x)]^{-1}\,\mathrm{Cov}(y,x),$$

pero intentaremos ceñirnos al caso $k = 1$ para simplificar la exposición.

Variables instrumentales

Un modelo lineal de variables instrumentales especifica una relación lineal entre y y x de la forma

$$E(y - \gamma_0 - \gamma_1 x \mid z) = 0.$$

Hay muchos ejemplos en econometría de modelos de este tipo (errores en las variables, simultaneidad, modelos dinámicos con autocorrelación, etc.).

Los modelos dinámicos de datos de panel que estudiaremos en la sección siguiente son modelos de este tipo, a diferencia de los modelos estáticos vistos anteriormente, que eran modelos de regresión.

Veamos en primer lugar cómo están definidas γ_0 y γ_1 en este modelo. Sea $v = y - \gamma_0 - \gamma_1 x$;

tenemos

$$E(v \mid z) = 0,$$

que implica, como en el caso anterior,

1) $E(v) = 0.$
2) $\mathrm{Cov}(v, z) = E(z\, v) = 0.$

Resolviendo γ_1 en (2):

$$\mathrm{Cov}(y,z) - \gamma_1 \,\mathrm{Cov}(x,z) = 0$$

$$\gamma_1 = \frac{\mathrm{Cov}(y,z)}{\mathrm{Cov}(x,z)} \,.$$

Igualmente a partir de (1)

$$\gamma_0 = E(y) - \frac{\mathrm{Cov}(y,z)}{\mathrm{Cov}(x,z)} E(x).$$

Nótese que en general $\gamma_1 \neq \beta_1$: Se trata de dos modelos distintos, ambos correctamente definidos en la población. La elección entre uno u otro depende exclusivamente de si estamos interesados en β_1 o en γ_1 por motivos económicos, no estadísticos. (Naturalmente, si el objetivo es minimizar la varianza del error siempre escogeremos β_0, β_1 con preferencia a γ_0, γ_1.)

Igualmente la estimación de γ_1 se lleva a cabo reemplazando covarianzas poblacionales por covarianzas muestrales en su definición, y de forma similar para γ_0.

Modelos generalizados de momentos

Los modelos anteriores se pueden llamar «problemas de momentos» porque los parámetros están definidos por funciones de momentos (medias, varianzas y covarianzas) de la distribución conjunta de las variables consideradas en la población.

Ahora vamos a considerar modelos generalizados de momentos en el sentido de que implican que los parámetros satisfacen más de una ecuación de momentos en la población [4].

Supongamos

$$E(y - \gamma_0 - \gamma_1 x \mid z,w) = 0.$$

Siguiendo el análisis anterior, γ_1 continúa satisfaciendo

$$\gamma_1 = \frac{\mathrm{Cov}(y,z)}{\mathrm{Cov}(x,z)} \,,$$

---

[4] De hecho, el problema anterior también satisface $E(vh(z)) = 0$ para todo $h(z)$ excepto en el caso en que $E(. \mid z)$ represente un predictor lineal óptimo.

pero ahora además

$$\gamma_1 = \frac{\mathrm{Cov}(y,w)}{\mathrm{Cov}(x,w)},$$

y de forma similar para γ_0.

Sin embargo, surge el problema de que en una muestra los dos cocientes anteriores no coincidirán debido a errores muestrales, aunque sean iguales en la población. Por tanto, vamos a utilizar como estimador de γ_1 (y de γ_0) una media ponderada de ambos cocientes muestrales, con ponderaciones acordes a sus varianzas relativas.

Estos estimadores, definidos como medias ponderadas de varios estimadores simples de variables instrumentales, se llaman estimadores del método generalizado de momentos (MGM).

Nótese que en este caso podemos contrastar si la diferencia entre los dos cocientes anteriores en la muestra es significativamente distinta de cero. A este tipo de inferencias se le denomina contrastes de «restricciones de sobreidentificación» (véase Sargan, 1988).

5. Modelos autorregresivos con efectos individuales

Vamos a considerar modelos para datos de panel del tipo

$$y_{it} = \alpha_1 y_{i(t-1)} + \ldots + \alpha_p y_{i(t-p)} + \eta_i + v_{it}.$$

Obsérvese que este modelo combina dos formas distintas de dependencia en y_{it}: autorregresiva y heterogeneidad permanente. En la práctica puede resultar difícil separar estos dos efectos porque ambos inducen un comportamiento temporal de y_{it} similar. Sin embargo, muchos modelos económicos de comportamiento individual incluyen estos dos ingredientes y por ello puede tener interés considerar formas reducidas con estas características. Si y_{it} es un vector de variables y η_i un vector de efectos individuales tendríamos un modelo «VAR» con efectos individuales.

La forma natural de interpretar esta ecuación es como una especificación de la media condicionada de y_{it} dado $y_{i(t-1)}, \ldots, y_{i1}, \eta_i$:

$$E(y_{it} \mid y_{i(t-1)}, \ldots, y_{i1}, \eta_i) = \alpha_1 y_{i(t-1)} + \ldots + \alpha_p y_{i(t-p)} + \eta_i;$$

por tanto v_{it} es la desviación de y_{it} con respecto a la media anterior.

Empecemos considerando el caso más simple: $T = 3$, $p = 1$; tenemos eliminando las *i's* por el momento:

$$E(y_2 \mid y_1, \eta) = \alpha y_1 + \eta \, .$$
$$E(y_3 \mid y_2, y_1, \eta) = \alpha y_2 + \eta .$$

La observación más importante que debemos hacer es que no podemos eliminar η diferenciando como en el modelo estático porque ahora los conjuntos en que condicionamos en cada ecuación son distintos.

Sin embargo,

$$E(y_3 \mid y_1, \eta) = \alpha E(y_2 \mid y_1, \eta) + \eta .$$

Por tanto, tenemos:

$$E(y_2 - \alpha y_1 \mid y_1, \eta) = \eta$$
$$E(y_3 - \alpha y_2 \mid y_1, \eta) = \eta$$

y restando

$$E(\Delta y_3 - \alpha \Delta y_2 \mid y_1, \eta) = 0;$$

por consiguiente también

$$E(\Delta y_3 - \alpha \Delta y_2 \mid y_1) = 0;$$

con lo que obtenemos un modelo de variables instrumentales en el que α está definido como

$$\alpha = \frac{\text{Cov}(y_1, \Delta y_3)}{\text{Cov}(y_1, \Delta y_2)} \, .$$

Nótese que este modelo es distinto del modelo de regresión:

$$E(\Delta y_3 \mid \Delta y_2) = \delta \Delta y_2,$$

en donde δ se define

$$\delta = \frac{\text{Cov}(\Delta y_2, \Delta y_3)}{\text{Var}(\Delta y_2)} \, .$$

Es fácil comprobar que bajo supuestos clásicos

$$\delta = \alpha - \frac{(1 + \alpha)}{2} < \alpha.$$

Por tanto tomar primeras diferencias y estimar por MCO proporcionará estimaciones inconsistentes de α (y consistentes de δ). Otro modelo distinto es

$$E(y_3 \mid y_2) = \gamma y_2,$$

para el que

$$\gamma = \frac{\mathrm{Cov}(y_2, y_3)}{\mathrm{Var}(y_2)} .$$

En este caso habitualmente $\gamma > \alpha$ (véase Arellano-Bover, 1990b).

La situación es la misma si $T > 3$ y si en lugar de regresiones en primeras diferencias consideramos regresiones intra-grupos. Nickell (1981) obtuvo que el coeficiente de regresión simple en la población de un modelo autorregresivo en desviaciones con respecto a las medias era igual a

$$\delta_N = \alpha - \frac{(1+\alpha)h}{(T-1)}\left(1 - \frac{2\alpha h}{(T-1)(1-\alpha)}\right)^{-1},$$

en donde

$$h = 1 - \frac{(1-\alpha^T)}{T(1-\alpha)} .$$

El cuadro siguiente (tomado de Arellano y Bover, 1990b) muestra los sesgos para distintos valores de α y T.

CUADRO I. *El sesgo de Nickell.*

$(T+1)\diagdown\alpha$	0,05	0,5	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

Nótese que el sesgo es de orden $1/T$ (sin embargo, incluso con $T = 15$ el sesgo es del 22 por ciento cuando $\alpha = 0{,}5$). Otras características importantes son que si $\alpha > 0$ el sesgo siempre es negativo y que el sesgo no tiende a cero cuando $\alpha \to 0$.

El estimador intragrupos es aconsejable cuando las aproximaciones para T grande son suficientemente buenas.

Continuando con el argumento anterior para el modelo autorregresivo, si $T = 4$ tenemos una ecuación adicional

$$E(y_4 \mid y_3, y_2, y_1, \eta) = \alpha y_3 + \eta,$$

que tomando esperanzas condicionadas a (y_2, y_1, η) proporciona

$$E(y_4 - \alpha y_3 \mid y_2, y_1, \eta) = \eta$$

y restando de la ecuación anterior

$$E(\Delta y_4 - \alpha \Delta y_3 \mid y_1, y_2) = 0.$$

Nótese que con $T = 4$ ya tenemos un problema de estimación de momentos generalizados, puesto que α además de satisfacer el cociente anterior también satisface

$$\alpha = \frac{\mathrm{Cov}(\Delta y_4, y_1)}{\mathrm{Cov}(\Delta y_3, y_1)} = \frac{\mathrm{Cov}(\Delta y_4, y_2)}{\mathrm{Cov}(\Delta y_3, y_2)} .$$

En general, para T observaciones temporales tenemos $(T - 2)$ ecuaciones de variables instrumentales del tipo

$$E(\Delta y_t - \alpha \Delta y_{t-1} \mid y_{t-2}, \ldots, y_1) = 0 \quad (t = 3, \ldots, T).$$

Esto es, ecuaciones con distintos instrumentos válidos en cada una de ellas. Nótese que incluso con valores moderados de T el número de restricciones de sobreidentificación será grande.

Las ponderaciones óptimas en el cálculo de los estimadores MGM serán distintas según se suponga que la varianza de $\Delta y_t - \alpha \Delta y_{t-1} \mid y_{t-2}, \ldots, y_1$ es constante o no.

Nótese que la elección de ponderaciones afectará la precisión del estimador pero no su consistencia (sea cual sea el esquema de ponderaciones, tenemos una media de cocientes de momentos que coinciden con α en la población, y que por tanto se aproximan a α a medida que N crece).

Usando la terminología de Arellano y Bond (1991) podemos distinguir dos tipos de estimadores MGM: el estimador MGM de «una etapa», que es óptimo en el caso homocedástico aunque podemos calcular estimaciones de sus varianzas robustas al caso heterocedástico, y el estimador MGM en «dos etapas», que es óptimo en general. No obstante, debe tenerse en cuenta que estos resultados no son más que aproximaciones asintóticas para muestras con un gran número de individuos.

Modelos autorregresivos con variables explicativas adicionales

En este contexto es importante distinguir los modelos con variables predeterminadas de los modelos que contienen variables exógenas en sentido estricto.

A) *Variables predeterminadas*

El modelo es

$$E(y_t \mid y_{t-1}, \ldots, y_1, x_t, x_{t-1}, \ldots, x_1, \eta) = \alpha y_{t-1} + \beta x_t + \eta$$

(la discusión es la misma si el modelo incluye retardos de x_t y/o retardos adicionales de y_t).

Para $T = 3$ tenemos, siguiendo el argumento anterior,

$$E(\Delta y_3 - \alpha \Delta y_2 - \beta \Delta x_3 \mid y_1, x_2, x_1) = 0$$

y en general

$$E(\Delta y_t - \alpha \Delta y_{t-1} - \beta \Delta x_t \mid y_{t-2}, \ldots, y_1, x_{t-1}, \ldots, x_1) = 0 \quad (t = 3, \ldots, T).$$

La expresión anterior clarifica los instrumentos válidos para cada ecuación en primeras diferencias.

Nótese que en los modelos autorregresivos que hemos visto hasta ahora, las desviaciones de y_t con respecto a la media condicionada no están correlacionadas con $y_{t-1}, \ldots, y_1$ y por tanto estarán no autocorrelacionadas.

Si, por ejemplo, queremos permitir errores de media móvil de primer orden, consideraremos la siguiente especificación:

$$E(y_t - \alpha y_{t-1} - \beta x_t \mid y_{t-2}, \ldots, y_1, x_{t-1}, \ldots, x_1, \eta) = \eta.$$

B) *Variables exógenas en sentido estricto*

El modelo en este caso es de la forma

$$E(y_t - \alpha y_{t-1} \mid x_1, \ldots, x_T, \eta) = \beta x_t + \eta.$$

Con lo cual tenemos

$$E(\Delta y_t - \alpha \Delta y_{t-1} - \beta \Delta x_t \mid x_1, \ldots, x_T) = 0.$$

Esto es, permitimos autocorrelación arbitraria en los errores pero necesitamos para ello una variable estrictamente exógena (tenemos los mismos instrumentos para todas las ecuaciones).

Obviamente, todo tipo de combinaciones de los modelos anteriores son posibles, y se pueden utilizar contrastes de restricciones de sobreidentificación para discriminar entre modelos alternativos.

Comentarios finales

(i) Exogeneidad estricta *versus* ausencia de autocorrelación.
Hay contextos en los que hay razones *a priori* para esperar errores aleatorios: algunos modelos de expectativas racionales, VAR's, etc. En el resto de los casos la identificación de parámetros en base a supuestos auxiliares de ausencia de autocorrelación es conflictiva. Por otra parte, es a menudo difícil argumentar la exogeneidad estricta de variables económicas en las aplicaciones.

(ii) ¿Funcionan estos estimadores en la práctica?
Surgen problemas de excesiva imprecisión con valores de T muy pequeños (como 2 ó 3). También hay problemas de identificación en modelos con raíces autorregresivas unitarias o próximas a la unidad. Sin embargo, estas técnicas son a menudo muy útiles, como lo demuestra su utilización creciente en el trabajo aplicado.