Instrumentos óptimos en el análisis de datos de panel con una aplicación a los determinantes del crecimiento económico

Manuel Arellano

VI Encuentro de Economía Aplicada
Granada, 5 de junio de 2003

Plan de la presentación:

1. La cuestión causal en el análisis empírico

1.1. Regresión múltiple
1.2. Variables instrumentales
1.3. Efectos fijos

2. Modelización de instrumentos óptimos

2.1 GMM
2.2 Variables instrumentales con regresión restringida
2.3 Ilustración empírica: VAR de empleo y salarios
2.4 Resultados de simulaciones

3. Crecimiento y tasas de convergencia entre países

1. La cuestión causal en el análisis empírico

- Queremos medir el efecto causal de una variable sobre otra. El problema es que en ausencia de datos experimentales la asociación medida por una regresión de y sobre x no necesariamente capta un efecto causal.

- Supongamos por simplicidad que x es binaria. La regresión es
$$y = x\gamma + \varepsilon$$
donde ε es tal que su correlación con x es nula y
$$\gamma = E(y \mid x = 1) - E(y \mid x = 0).$$

- Si representamos la relación causal por
$$y = x\beta + u,$$
en el argot econométrico se dice que x es "endógena" si está correlacionada con u, en cuyo caso $\beta \neq \gamma$ y $u \neq \varepsilon$.

- En un ejemplo concreto esto quiere decir lo siguiente. Supongamos que y=lluvia, x=ir con paraguas, u=otros determinantes de y. Aquí el efecto causal es $\beta = 0$ y $u = y$. Tenemos causalidad inversa.

- Este ejemplo es banal pero ilustra el problema. Si tuviéramos datos experimentales podríamos comprobar empiricamente lo que ya sabemos gracias al sentido común.

- Otros ejemplos: función de producción, ecuación de demanda o evaluación de un programa de formación.

- Si el error u tiene poca variación, el margen de correlación entre x y u es escaso, pero con frecuencia no es así.
- Los efectos causales son importantes para la evaluación de políticas porque nos permiten responder a preguntas del tipo ¿cuál fue o cual sería el efecto de?
- ¿Cómo medir efectos causales? La respuesta tradicional de la econometría incluye: regresión múltiple, variables instrumentales y efectos fijos.

1.1. Regresión múltiple

- Calcular la regresión para submuestras con distintos valores de las características w.
- La idea es que aunque pueda haber correlación entre x y u en el conjunto de la población, quizá no la haya por subgrupos definidos por un valor dado de w.
- En la regresión lineal consideramos
$$y = x\gamma_1 + w\gamma_2 + \varepsilon,$$
con el objetivo de que γ_1 se pueda interpretar de forma causal.

1.2. Variables instrumentales

- Una variable z correlacionada con x pero no con u que carece de un efecto directo sobre y.

- Ejemplo: En la población de mujeres con dos o más hijos sea y=participación laboral, x=tener dos o más hijos, u= oportunidades laborales (despidos, ascensos), z=tener dos hijos del mismo sexo. Empiricamente se observa que
$$E(x \mid z=1) > E(x \mid z=0).$$
Obtenemos
$$\beta = \frac{E(y \mid z=1) - E(y \mid z=0)}{E(x \mid z=1) - E(x \mid z=0)}.$$

- ¿Cual es la intuición? Se atribuye el efecto de z sobre la participación a x ya que no se espera que el sexo de los hijos per se tenga un efecto directo. Formalmente:
$$\begin{aligned} E(y \mid z=1) &= \beta E(x \mid z=1) + E(u) \\ E(y \mid z=0) &= \beta E(x \mid z=0) + E(u) \end{aligned}$$

- En la práctica el problema es encontrar instrumentos que sean convincentes a priori, pero incluso teniéndolos puede haber problemas de tipo estadístico.

- En series temporales es frecuente utilizar retardos como instrumentos, pero si los errores son persistentes no está claro que los retardos sean mucho más exógenos que las variables sin retardar.

- *Instrumentos múltiples.* Si tenemos varios z_j que a priori satisfacen $Cov\left(z_j, u\right) = 0$

$$\begin{aligned} y &= \beta x + u \\ Cov\left(z_j, y\right) &= \beta Cov\left(z_j, x\right) + Cov\left(z_j, u\right) \end{aligned}$$

y de aquí

$$\beta = \frac{Cov\left(z_j, y\right)}{Cov\left(z_j, x\right)}.$$

- Se comprueba que es óptimo combinar los instrumentos utilizando un predictor de x:

$$\widehat{x} = \pi_1 z_1 + ... + \pi_r z_r$$
$$\beta = \frac{Cov\left(\widehat{x}, y\right)}{Cov\left(\widehat{x}, x\right)}.$$

- ¿Cual es el problema de tener muchos instrumentos? Morir de éxito. Si usamos tantos instrumentos como observaciones, muestralmente $\widehat{x} = x$ y estamos de vuelta en la regresión.

- Es crucial la diferencia entre suponer conocidos los coeficientes π_j y estimarlos.

- Si r es grande en relación al tamaño muestral los estimadores de VI no son fiables, especialmente si individualmente el poder predictivo de cada instrumento es bajo.

1.3. Efectos fijos

- Si tenemos observaciones para los mismos individuos en dos periodos distintos y pensamos que

$$\begin{aligned} y_{i1} &= \beta x_{i1} + u_{i1} \\ y_{i2} &= \beta x_{i2} + u_{i2} \\ u_{it} &= \eta_i + v_{it} \quad (t = 1, 2) \end{aligned}$$

  donde la parte de u correlacionada con x es constante en el tiempo, la regresión en diferencias está libre del problema de endogeneidad:

$$y_{i2} - y_{i1} = \beta \left(x_{i2} - x_{i1} \right) + \left(v_{i2} - v_{i1} \right).$$

- Variables predeterminadas. Se plantea el problema de que x_{i2} sea independiente de v_{i2} pero no de v_{i1}. Ejemplos:

  (a) x=número de hijos, v=*shocks* laborales;

  (b) $x_2 = y_1$, en este caso hay correlación por construcción.

- En ejemplos como estos las diferencias sirven para eliminar una fuente de endogeneidad (los efectos fijos) y los retardos (x_{i1}) nos sirven como instrumentos.

- La conclusión es que los datos de panel además de eliminar efectos fijos generan multitud de instrumentos internos (cuyo número depende del de olas T).

2. Modelización de instrumentos óptimos

- Modelo de efectos fijos con errores que satisfacen restricciones secuenciales:

$$y_{it} = x_{it}\beta + \eta_i + \varepsilon_{it}$$
$$E(\varepsilon_{it}|z_{i1}, ..., z_{it}) = 0$$
$$(t = 1, ..., T; i = 1, ..., N),$$

- Si una variable explicativa x_{kit} is predeterminada para $\varepsilon_{i(t+j)}$, entonces $x_{ki(t-j)}$ es parte de z_{it}.
- x_{kit} puede ser completamente endógena en el sentido de no ser predeterminada para ningún adelanto de ε.
- z_{it} puede contener instrumentos externos que no son parte de x_{it} o sus retardos.

Ejemplos:

- Ecuación de un modelo VAR con efectos individuales.
- Modelo de regresión de ajuste parcial con regresores predeterminados.
- Relación estructural entre variables endógenas. Ejemplos: (i) Crecimiento. (ii) Consumo de los hogares.

2.1 GMM

- Con datos de panel podemos considerar

$$\begin{aligned} \widehat{\Delta x_{i2}} &= \pi_{21} z_{i1} \\ \widehat{\Delta x_{i3}} &= \pi_{31} z_{i1} + \pi_{32} z_{i2} \\ &\vdots \\ \widehat{\Delta x_{iT}} &= \pi_{T1} z_{i1} + ... + \pi_{T(T-1)} z_{i(T-1)}. \end{aligned}$$

- Los estimadores GMM utilizan $\widehat{\Delta x_{it}}$ como instrumento. El problema es que el número de coeficientes de primera etapa puede resultar grande incluso en micropaneles. ¿Qué ocurre si T y N son pequeños?

- *Alternativas desde una perspectiva de serie temporal.* Una posibilidad es considerar

$$\widehat{\Delta x_{it}} = \pi z_{i(t-1)} \ \ (t = 2, ..., T),$$

  o bien

$$\begin{aligned} \widehat{\Delta x_{i3}} &= \pi_1 z_{i1} + \pi_2 z_{i2} \\ &\vdots \\ \widehat{\Delta x_{iT}} &= \pi_1 z_{i(T-2)} + \pi_2 z_{i(T-1)}. \end{aligned}$$

- Un problema es que se pierden cortes transversales enteros. Otro es que si z contiene efectos individuales todos los retardos son predictores de las diferencias.

- En la práctica se utilizan los métodos GMM pero a menudo los resultados no son fiables. Lo veremos en una simulación para empresas.

2.2 Variables instrumentales con regresión restringida

- La propuesta aquí es utilizar como instrumento

$$\begin{aligned}\widehat{\Delta x_{i2}} &= \pi_{21}(\gamma)\, z_{i1} \\ \widehat{\Delta x_{i3}} &= \pi_{31}(\gamma)\, z_{i1} + \pi_{32}(\gamma)\, z_{i2} \\ &\vdots \\ \widehat{\Delta x_{iT}} &= \pi_{T1}(\gamma)\, z_{i1} + ... + \pi_{T(T-1)}(\gamma)\, z_{i(T-1)}\end{aligned}$$

  donde γ son los coeficientes de un modelo parsimonioso para las variables instrumentales (Arellano, 2003).

- Hay dos niveles de supuestos: los sustantivos de los que depende la consistencia y los auxiliares de los que depende la eficiencia (como en MC2E, frente a posibles predictores no lineales por ejemplo).

- Cuando T/N no es despreciable, GMM tiene un sesgo que no desaparece asintoticamente, lo que cuestiona formalmente la legitimidad del estimador (Alvarez–Arellano).

- Si T es grande SIV$\simeq$ST–IV; si T es pequeño SIV$\simeq$GMM, pero en casos intermedios la diferencia con GMM puede ser enorme en la práctica como veremos.

- Como modelos para los instrumentos consideramos VAR con efectos individuales y condiciones iniciales no restringidas.

- También desarrollamos métodos de estimación de las regresiones restringidas y de los errores estándar.

2.3 Ilustración empírica: VAR de empleo y salarios

- Estimamos un modelo VAR de empleo y salarios utilizando datos de panel de empresas.
- En primer lugar consideramos un panel completo de 738 empresas manufactureras españolas, de las que se dispone de observaciones para 1983-1990.
- En segundo lugar, un panel más largo de 385 empresas de la misma fuente (Central de Balances, Banco de España) para las que disponemos de 14 años, también desde 1983. El tamaño medio de las empresas del segundo panel es más del doble que las del primero.
- Estimamos un modelo VAR(1) para los logaritmos del empleo y los salarios (n_{it} y w_{it} respectivamente). Ambas ecuaciones incluyen efectos de tiempo y de empresa.
- Los efectos temporales se eliminan antes de la estimación tomando los datos en desviaciones respecto a las medias de corte transversal de cada año.
- La Tabla 1 contiene los resultados.

Tabla 1
Empleo y salarios: Modelo VAR
Datos de panel de empresas españolas

	WG	GMM	PML	SIV
	$N = 738, T = 8$			
	Ecuación de empleo			
$n_{i(t-1)}$	0.71	0.86	1.00	0.93
	(0.03)	(0.06)		(0.07)
$w_{i(t-1)}$	0.08	0.12	0.08	0.14
	(0.03)	(0.07)		(0.08)
	Ecuación de salarios			
$n_{i(t-1)}$	0.06	-0.03	0.01	-0.02
	(0.02)	(0.08)		(0.08)
$w_{i(t-1)}$	0.44	0.29	0.68	0.32
	(0.03)	(0.10)		(0.10)
	$N = 385, T = 14$			
	Ecuación de empleo			
$n_{i(t-1)}$	0.86	0.83	0.995	0.90
	(0.05)	(0.05)		(0.05)
$w_{i(t-1)}$	0.26	0.29	0.28	0.30
	(0.09)	(0.13)		(0.15)
	Ecuación de salarios			
$n_{i(t-1)}$	0.01	-0.09	0.03	-0.15
	(0.02)	(0.04)		(0.05)
$w_{i(t-1)}$	0.45	0.34	0.62	0.32
	(0.07)	(0.10)		(0.12)

Datos en desviaciones respecto a las medias de corte transversal.

2.4 Simulaciones de Monte Carlo

- A continuación llevamos a cabo un ejercicio de simulación calibrado al panel de empresas anterior.
- El diseño se escogió de la representación de ajuste parcial del modelo de la forma siguiente:

$$\begin{aligned} y_{it} &= 1 + 0.8y_{i(t-1)} - 0.5x_{it} + 0.3x_{i(t-1)} + \eta_i + v_{it} \\ x_{it} &= 0.5 + 0.3x_{i(t-1)} + \xi_i + \varepsilon_{it}. \end{aligned}$$

- Todas las variables inobservables son normales *iid* (tanto en i como en t) con media cero y $\sigma_v^2 = \sigma_\varepsilon^2 = 0.01$, $Corr\,(v_{it}, \varepsilon_{it}) = 0$, $\sigma_\eta^2 = \sigma_\xi^2 = 0.09$, y $Corr\,(\eta_i, \xi_i) = 0.6$.
- En consecuencia, la varianza de los efectos fijos es 9 veces la de los errores aleatorios.
- No hay realimentación de retardos de y a x, y el efecto de largo plazo de x sobre y es la unidad.
- Las observaciones iniciales fueron generadas a partir de la distribución de estado estacionario del proceso.

- El VAR correspondiente es

$$\begin{pmatrix} y_{it} \\ x_{it} \end{pmatrix} = \begin{pmatrix} 0.75 \\ 0.5 \end{pmatrix} + \begin{pmatrix} 0.8 & 0.15 \\ 0 & 0.3 \end{pmatrix} \begin{pmatrix} y_{i(t-1)} \\ x_{i(t-1)} \end{pmatrix} + \begin{pmatrix} c_i + e_{it} \\ \xi_i + \varepsilon_{it} \end{pmatrix}$$

  donde $e_{it} = v_{it} - 0.5\varepsilon_{it}$, $c_i = \eta_i - 0.5\xi_i$,

$$Var \begin{pmatrix} e_{it} \\ \varepsilon_{it} \end{pmatrix} = \Omega = \begin{pmatrix} 0.0125 & -0.005 \\ -0.005 & 0.01 \end{pmatrix}$$

$$Var \begin{pmatrix} c_i \\ \xi_i \end{pmatrix} = \Omega_\eta = \begin{pmatrix} 0.0585 & 0.009 \\ 0.009 & 0.09 \end{pmatrix}.$$

- Las correlaciones implicadas son $Corr\,(e_{it}, \varepsilon_{it}) = -0.447$ y $Corr\,(c_i, \xi_i) = 0.124$.

- En la Tabla 2 presentamos medianas y errores absolutos medianos de los estimadores WG, GMM, PML y SIV para $\{N = 738, T = 8\}$ y $\{N = 385, T = 14\}$.

- En otro experimento con $N = 738$, $T = 8$ generamos datos con una tendencia en las varianzas. La especificación en este caso es

$$\sigma^2_{vt} = \sigma^2_{\varepsilon t} = 0.01 + 0.001t,$$

  con $t = 0$ seleccionado de forma que la secuencia resultante de varianzas esté comprendida entre .005 y .012. Estos resultados también están en la Tabla 2.

- Finalmente, en la Tabla 3 presentamos simulaciones adicionales para $\{N = 200, T = 8\}$,$\{N = 100, T = 6\}$, y $\{N = 50, T = 15\}$.

- En todos los casos llevamos a cabo 1000 replicaciones.

Tabla 2

Simulaciones de Monte Carlo para el modelo VAR

	WG		GMM		PML		SIV	
	mediana	mae	mediana	mae	mediana	mae	mediana	mae
				$N = 738, T = 8$				
a_{11}	0.48	0.32	0.72	0.08	0.80	0.02	0.80	0.05
a_{12}	0.09	0.06	0.13	0.03	0.15	0.02	0.15	0.03
a_{21}	0.03	0.03	0.01	0.04	0.00	0.01	−0.01	0.05
a_{22}	0.12	0.18	0.29	0.03	0.30	0.01	0.30	0.03
				$N = 385, T = 14$				
a_{11}	0.63	0.17	0.74	0.06	0.80	0.01	0.80	0.03
a_{12}	0.12	0.03	0.13	0.02	0.15	0.01	0.15	0.02
a_{21}	0.02	0.02	0.01	0.02	−0.00	0.01	−0.00	0.02
a_{22}	0.21	0.09	0.29	0.02	0.30	0.01	0.30	0.02
				$N = 738, T = 8$				
				Datos con tendencia en varianza				
a_{11}	0.50	0.30	0.70	0.11	0.86	0.06	0.81	0.07
a_{12}	0.11	0.04	0.12	0.04	0.17	0.02	0.16	0.04
a_{21}	0.03	0.03	0.01	0.05	−0.01	0.01	−0.01	0.06
a_{22}	0.11	0.19	0.29	0.03	0.31	0.02	0.30	0.03

$a_{11} = 0.8, a_{12} = 0.15, a_{21} = 0, a_{22} = 0.3$

1000 replicaciones. *mae* es error absoluto mediano.

Tabla 3

Simulaciones de Monte Carlo para el modelo VAR

	WG		GMM		PML		SIV	
	mediana	mae	mediana	mae	mediana	mae	mediana	mae
				$N = 200, T = 8$				
a_{11}	0.48	0.32	0.59	0.21	0.80	0.05	0.80	0.09
a_{12}	0.09	0.06	0.08	0.07	0.15	0.03	0.16	0.06
a_{21}	0.03	0.03	0.03	0.07	−0.00	0.02	0.00	0.09
a_{22}	0.12	0.18	0.27	0.05	0.30	0.02	0.30	0.05
				$N = 100, T = 6$				
a_{11}	0.35	0.45	0.40	0.40	0.80	0.10	0.80	0.22
a_{12}	0.08	0.07	0.04	0.13	0.15	0.06	0.15	0.12
a_{21}	0.04	0.05	0.04	0.14	−0.00	0.05	−0.01	0.19
a_{22}	0.04	0.26	0.22	0.10	0.30	0.05	0.29	0.11
				$N = 50, T = 15$				
a_{11}	0.64	0.16	0.62	0.18	0.80	0.04	0.80	0.06
a_{12}	0.12	0.04	0.10	0.06	0.15	0.04	0.15	0.05
a_{21}	0.02	0.03	0.02	0.04	0.00	0.02	0.00	0.06
a_{22}	0.22	0.09	0.24	0.06	0.30	0.03	0.30	0.04

$a_{11} = 0.8, a_{12} = 0.15, a_{21} = 0, a_{22} = 0.3$

1000 replicaciones. *mae* es error absoluto mediano.

3. Crecimiento y tasas de convergencia entre países

- Finalmente consideramos un modelo de Solow de los determinantes del crecimiento del tipo estudiado por Caselli, Esquivel y Lefort (1996). La ecuación es

$$y_{it} = \alpha y_{i(t-1)} + s_{i(t-1)}\gamma + f'_{i(t-1)}\delta + \eta_i + \varepsilon_{it}$$
$$E\left(\varepsilon_{it} \mid y_i^{t-1}, s_i^{t-1}, f_i^{t-2}\right) = 0.$$

- El intervalo temporal es 5 años, y_{it} es log PIB per-capita, $f_{i(t-1)}$ es un vector de variables flujo que contiene las tasas de inversión y de crecimiento de la población, y $s_{i(t-1)}$ es una variable *stock* que mide la tasa de escolarización secundaria.
- En este caso $z_{it} = \left(y_{i(t-1)}, s_{i(t-1)}, f_{i(t-2)}\right)$.
- Utilizando GMM, Caselli et al. obtuvieron una estimación sorprendentemente grande de la tasa de convergencia de alrededor del 10%. Este resultado contradecía ampliamente las estimaciones previas de corte transversal de Barro y coautores con tasas de convergencia del 2-3%.
- Caselli et al. argumentaron que las estimaciones anteriores estaban sesgadas debido a no tener en cuenta los efectos de país con regresores predeterminados.
- La preocupación es que sus estimaciones tengan un sesgo negativo de muestra finita en el coeficiente autorregresivo del PIB que se traduzca en un sesgo positivo en las tasas de convergencia (Bond, Hoeffler y Temple, 2001).

- Obtuvimos los datos de Caselli et al. y reestimamos un modelo de Solow aumentado con el nuevo estimador de variables instrumentales basado en las regresiones restringidas.

- Encontramos una tasa de convergencia más pequeña de un 4%, aunque estimada con poca precisión (Tabla 4).

- No obstante, desde un punto de vista sustantivo es al menos tan importante explicar la heterogeneidad en niveles de renta de estado estacionario puesta de manifiesto por esta literatura como obtener buenas estimaciones del coeficiente de convergencia.

- Utilizamos datos similares a Caselli et al. y Bond et al. La muestra es la misma que la de la Tabla 2 de Bond et al. excepto por la exclusión de 9 países y 17 observaciones a fín de disponer de un panel completo.

Tabla 4
Modelo de Solow aumentado
$N = 92, T = 5$

	OLS	WG	GMM	SIV
$(1+\beta)$	0.947	0.680	0.698	0.808
(s.e.)	(0.017)	(0.057)	(0.107)	(0.248)
$\ln(enr_t)$	0.035	-0.049	-0.140	-0.114
(s.e.)	(0.014)	(0.029)	(0.066)	(0.229)
$\ln(s_t)$	0.081	0.138	0.144	0.090
(s.e.)	(0.017)	(0.039)	(0.055)	(0.075)
$\ln(n_t + g + d)$	-0.094	-0.033	0.230	0.227
(s.e.)	(0.053)	(0.152)	(0.339)	(1.384)
λ implicada	0.011	0.077	0.072	0.043
(s.e.)	(0.004)	(0.017)	(0.031)	(0.062)

Datos en desviaciones respecto a las medias de corte transversal.

Datos para intervalos de 5 años 1960-1985.

s. e. robustos a heterosc. y autocorrelación.

n = tasa de crecimiento de la población; g = tasa de cambio tecnológico;

d = tasa de depreciación del capital físico ($g + d = 0.05$);

s = tasa de ahorro; enr = tasa de escolarización secundaria.

Referencias

- Arellano, M. (2003): “Modelling Optimal Instrumental Variables for Dynamic Panel Data Models”, Documento de Trabajo 0310, CEMFI, Madrid (disponible en la dirección: ftp://ftp.cemfi.es/pdf/papers/ma/siv2003.pdf).
- Bond, S. R., A. Hoeffler, y J. Temple (2001): “GMM Estimation of Empirical Growth Models”, CEPR Discussion Paper 3048, Londres.
- Caselli, F., G. Esquivel, y F. Lefort (1996): “Reopening the Convergence Debate: A New Look at Cross-Country Growth Empirics”, *Journal of Economic Growth*, 1, 363-389.