

Paneles dinámicos

Gabriel Montes-Rojas

Paneles dinámicos

Dos modelos que dan lugar a paneles dinámicos:

(1) Muestra longitudinal

$$y_{it} = x'_{it}\beta + \mu_i + \delta_t + v_{it}$$

$$AR(1) : v_{it} = \alpha v_{it-1} + \epsilon_{it}$$

o

$$MA(1) : v_{it} = \epsilon_{it} + \alpha\epsilon_{it}$$

$$i = 1, 2, \dots, N; t = 1, 2, \dots, T; |\alpha| < 1$$

(2) Paneles autorregresivos:

$$y_{it} = \alpha y_{it-1} + x'_{it}\beta + \mu_i + \delta_t + v_{it}$$

$$i = 1, 2, \dots, N; t = 1, 2, \dots, T; |\alpha| < 1$$

En el caso (1) nos interesa estimar la estructura de varianzas y covarianzas. En el (2) eliminar el sesgo dinámico.

Paneles dinámicos

Dos fuentes de persistencia:

- 1 Autocorrelación, persistencia dinámica, **state dependence**: α
Ej: costos de ajustes, costos hundidos, shocks económicos (terremotos, guerras).
 - 2 Efectos individuales, **individual heterogeneity**: μ
Ej: habilidad, tecnología, instituciones.
 - 3 Ver Bera, Sosa-Escudero y Yoon (2001), Zincenko, Sosa-Escudero, y Montes-Rojas, G.V. (2014).
- Comparación con VAR (Vector Autoregressive). Los paneles dinámicos son parsimoniosos.
 - Ventajas cuando T fija y se explota la variación en N .

Corto plazo vs. largo plazo

La estructura dinámica permite distinguir corto de largo plazo.

- β es el efecto contemporáneo **corto plazo** de x en y en t ;
- sin embargo el efecto en t tendrá un efecto en $t + 1$, $t + 2$, etc. por el rezago de la variable dependiente;
- el **efecto de largo plazo** se calcula como $\frac{\beta}{1-\alpha}$ (asumiendo $|\alpha| < 1$)

La ventaja de los paneles dinámicos es que permite tener un modelo de la **dinámica de ajuste**.

Muestras longitudinales con correlación serial

Los errores estándar están mal calculados...

¿Por qué es importante controlar por correlación serial?

- *Bertrand, M., Duflo, E., and Mullainathan, S. (2004). "How much should we trust differences-in-differences estimates?," Quarterly Journal of Economics 119(1), 249-275.*
- *Cameron, C., and Miller, D.L. (2015). "A practitioner's guide to cluster-robust inference," Journal of Human Resources 50(2), 317-372.*

Identificación

(Arellano, 2002, pp.58-60)

- Supongamos un panel corto con $T = 2$ pero N grande: $\{(y_{i1}, y_{i2})\}_{i=1}^N$.
- Modelo de **efectos aleatorios one-way (heterogeneidad pura)** con error $y_{it} = \mu_i + v_{it}$ tendría la siguiente estructura de varianzas y covarianzas:

$$\text{var}(y_{i1}) = \text{var}(y_{i2}) = \sigma_{\mu}^2 + \sigma_v^2$$

$$\text{cov}(y_{i1}, y_{i2}) = \sigma_{\mu}^2.$$

- Modelo **homogéneo con correlación serial** $y_{it} = \eta + v_{it}$ donde $v_{it} = \alpha v_{it-1} + \epsilon_{it}$, η constante, $\epsilon \sim iid(0, \sigma_{\epsilon}^2)$, $v_{i1} \sim iid(0, \sigma_v^2)$ con $\sigma_v^2 = \sigma_{\epsilon}^2 / (1 - \alpha^2)$ tendría la siguiente estructura de varianzas y covarianzas:

$$\text{var}(y_{i1}) = \text{var}(y_{i2}) = \sigma_v^2$$

$$\text{cov}(y_{i1}, y_{i2}) = \alpha \sigma_v^2.$$

Identificación

(Arellano, 2002, pp.58-60) $T = 2$ cont.

- En los dos modelos la autocorrelación $\frac{cov(y_{i1}, y_{i2})}{\sqrt{var(y_{i1})var(y_{i2})}}$ es constante... no se puede distinguir uno de otro.
- Un modelo combinado con **heterogeneidad** y **correlación serial**:

$$var(y_{i1}) = var(y_{i2}) = \sigma_{\mu}^2 + \sigma_v^2$$

$$cov(y_{i1}, y_{i2}) = \sigma_{\mu}^2 + \alpha\sigma_v^2.$$

Identificación

(Arellano, 2002, pp.58-60)

- Supongamos un panel corto con $T = 3$ pero N grande: $\{(y_{i1}, y_{i2}, y_{i3})\}_{i=1}^N$.
- Modelo de **efectos aleatorios one-way (heterogeneidad pura)** con error $y_{it} = \mu_i + v_{it}$ tendría la siguiente estructura de varianzas y covarianzas:

$$\text{var}(y_{i1}) = \text{var}(y_{i2}) = \text{var}(y_{i3}) = \sigma_{\mu}^2 + \sigma_v^2$$

$$\text{cov}(y_{i1}, y_{i2}) = \text{cov}(y_{i1}, y_{i3}) = \text{cov}(y_{i2}, y_{i3}) = \sigma_{\mu}^2.$$

Este modelo lo podemos llamar de **equicorrelación**, donde la correlación intra-cluster $IC = \sigma_{\mu}^2 / (\sigma_{\mu}^2 + \sigma_v^2)$.

- Modelo **homogéneo con correlación serial** $y_{it} = \eta + v_{it}$ donde $v_{it} = \alpha v_{it-1} + \epsilon_{it}$, η constante, $\epsilon \sim iid(0, \sigma_{\epsilon}^2)$, $v_{i1} \sim iid(0, \sigma_v^2)$ con $\sigma_v^2 = \sigma_{\epsilon}^2 / (1 - \alpha^2)$ tendría la siguiente estructura de varianzas y covarianzas:

$$\text{var}(y_{i1}) = \text{var}(y_{i2}) = \sigma_v^2$$

$$\text{cov}(y_{i1}, y_{i2}) = \alpha \sigma_v^2, \text{cov}(y_{i1}, y_{i3}) = \alpha^2 \sigma_v^2$$

Identificación

(Arellano, 2002, pp.58-60) $T = 3$ cont.

- Un modelo combinado con heterogeneidad y correlación serial:

$$\text{var}(y_{i1}) = \text{var}(y_{i2}) = \text{var}(y_{i3}) = \sigma_{\mu}^2 + \sigma_v^2$$

$$\text{cov}(y_{i1}, y_{i2}) = \text{cov}(y_{i2}, y_{i3}) = \sigma_{\mu}^2 + \alpha\sigma_v^2$$

pero

$$\text{cov}(y_{i1}, y_{i3}) = \sigma_{\mu}^2 + \alpha^2\sigma_v^2$$

- Este modelo está exactamente identificado y por lo tanto podemos distinguir uno de otro.

Identificación

(Arellano, 2002, pp.58-60) generalización

- En general, persistencia medida a través de la autocorrelación combina dos fuentes diferentes:

$$\rho = \frac{\sigma_{\mu}^2 + \alpha\sigma_{\nu}^2}{\sigma_{\mu}^2 + \sigma_{\nu}^2} = \alpha + \frac{(1-\alpha)\sigma_{\mu}^2}{\sigma_{\mu}^2 + \sigma_{\nu}^2} = \alpha + \frac{(1-\alpha)\lambda}{1+\lambda},$$

donde $\lambda = \sigma_{\mu}^2 / \sigma_{\nu}^2$.

Si usamos efectos aleatorios o fijos...

- Montes-Rojas, G. (2016). "An equicorrelation Moulton factor in the presence of arbitrary intra-cluster correlation," *Economics Letters* 145, 221-224.

Si usamos efectos aleatorios o fijos...

Supongamos el siguiente modelo

$$y_{it} = x_{it}\beta + v_{it},$$
$$v_{it} = \epsilon_{it} + \alpha\epsilon_{it-1}, 0 < \alpha < 1, \epsilon_{i0} = 0,$$
$$i = 1, 2, \dots, N, t = 1, 2, \dots, T.$$

Esto es una estructura MA(1) con $\epsilon \sim i.i.d.(0, \sigma_\epsilon^2)$.

Si usamos efectos aleatorios o fijos...

Definamos $\lambda_v^2 = \sigma_v^2(1 - \bar{\rho}_v)$, $\lambda_\mu^2 = \sigma_v^2 \bar{\rho}_v = 2/T \alpha \sigma_\epsilon^2$, $\bar{\rho}_v = \frac{2/T \alpha}{1 + \alpha^2}$ (note el factor $1/T$), y $\lambda_\mu^2 + \lambda_v^2 = \sigma_v^2 = (1 + \alpha^2) \sigma_\epsilon^2$.

Definamos la correlación intra-cluster

$$IC = \bar{\rho}_v = \frac{\lambda_\mu^2}{\lambda_\mu^2 + \lambda_v^2} = \frac{2/T \alpha}{(1 + \alpha^2)}$$

Nuestro objetivo es calcular la varianza del estimador de OLS, $V(\hat{\beta}|x)$:

$$\hat{\beta} = \frac{\sum_{i=1}^N \sum_{t=1}^T y_{it} x_{it}}{\sum_{i=1}^N \sum_{t=1}^T x_{it}^2}$$

Si usamos efectos aleatorios o fijos...

$V_0 := V_0(\hat{\beta}|x)$ es la varianza OLS estándar:

$$V_0 = \frac{(1 + \alpha^2)\sigma_\epsilon^2 \sum_{i=1}^N \sum_{t=1}^T x_{it}^2}{\left(\sum_{i=1}^N \sum_{t=1}^T x_{it}^2\right)^2} = \frac{(\lambda_\mu^2 + \lambda_\nu^2) \sum_{i=1}^N \sum_{t=1}^T x_{it}^2}{\left(\sum_{i=1}^N \sum_{t=1}^T x_{it}^2\right)^2}.$$

$V_1 := V_1(\hat{\beta}|x)$ es la varianza correcta:

$$\begin{aligned} V_1 &= \frac{(1 + \alpha^2)\sigma_\epsilon^2 \sum_{i=1}^N \sum_{t=1}^T x_{it}^2 + 2\alpha\sigma_\epsilon^2 \sum_{i=1}^N \sum_{t=2}^T x_{it}x_{it-1}}{\left(\sum_{i=1}^N \sum_{t=1}^T x_{it}^2\right)^2} \\ &= \frac{(\lambda_\mu^2 + \lambda_\nu^2) \sum_{i=1}^N \sum_{t=1}^T x_{it}^2 + \lambda_\mu^2 T \sum_{i=1}^N \sum_{t=2}^T x_{it}x_{it-1}}{\left(\sum_{i=1}^N \sum_{t=1}^T x_{it}^2\right)^2}. \end{aligned}$$

$V_2 := V_2(\hat{\beta}|x)$ es la varianza de un modelo de equicorrelación:

$$V_2 = \frac{(\lambda_\mu^2 + \lambda_\nu^2) \sum_{i=1}^N \sum_{t=1}^T x_{it}^2 + 2\lambda_\mu^2 \sum_{i=1}^N \sum_{t=1}^{T-1} \sum_{j=t+1}^T x_{it}x_{ij}}{\left(\sum_{i=1}^N \sum_{t=1}^T x_{it}^2\right)^2}.$$

Si usamos efectos aleatorios o fijos...

- Definamos $\rho_x^{(1)} = \frac{1}{N(T-1)} \sum_{i=1}^N \sum_{t=2}^T x_{it} x_{it-1}$ y $\rho_x^{(T)} = \frac{2}{NT(T-1)} \sum_{i=1}^N \sum_{t=1}^{T-1} \sum_{j=t+1}^T x_{it} x_{ij}$.
- El factor de Moulton (1986,1987,1990) es (usar OLS cuando hay correlación entre errores)

$$\frac{V_1}{V_0} = 1 + IC(T-1)\rho_x^{(1)}.$$

Si asumimos que $IC > 0$ y $\rho_x^{(1)} > 0$ entonces OLS subestima la varianza (resultado esperado), o sea, $V_1 \geq V_0$.

- Ahora si comparamos

$$\frac{V_1}{V_2} = \frac{1 + IC(T-1)\rho_x^{(1)}}{1 + IC(T-1)\rho_x^{(T)}}.$$

Si asumimos que $IC > 0$ y $\rho_x^{(1)} > \rho_x^{(T)}$ entonces RE subestima la varianza, o sea, $V_1 \geq V_2$.

Contrastes para efectos aleatorios vs. correlación serial

- Baltagi, B. H., and Li, Q. (1991). "A joint test for serial correlation and random individual effects." *Statistics and Probability Letters* 11, 277-280.
 - Baltagi, B. H., and Li, Q. (1995). "Testing AR(1) against MA(1) disturbances in an error component model." *Journal of Econometrics* 68, 133-151.
 - Bera, A., Sosa-Escudero, W., and Yoon, M. (2001). "Tests for the error component model in the presence of local misspecification." *Journal of Econometrics* 101, 1-23.
 - Bera, A., and Sosa-Escudero, W. (2008). "Tests for unbalanced error-components models under local misspecification." *STATA Journal* 8(1), 68-78.
- http://www.stata-journal.com/sjpdf.html?articlenum=sg164_1

Estimación

(Baltagi, 2008, cap.5, pp.92-94, Baltagi y Wu, 1999)

- Consideremos un modelo

$$y_{it} = x'_{it}\beta + \mu_i + \delta_t + v_{it},$$

$$AR(1) : v_{it} = \alpha v_{it-1} + \epsilon_{it},$$

$$\epsilon \sim i.i.d.(0, \sigma_\epsilon^2), \mu \sim i.i.d.(0, \sigma_\mu^2),$$

$$v_{i0} \sim iid(0, \sigma_v^2), \sigma_v^2 = \sigma_\epsilon^2 / (1 - \alpha^2),$$

$$i = 1, 2, \dots, N; t = 1, 2, \dots, T; |\alpha| < 1$$

- La estrategia es: (1) obtener un estimador del parámetro de AR(1); (2) transformar los residuos para que queden sin correlación; (3) estimar la varianza

Estimación

(Baltagi, 2008, cap.5, pp.92-94, Baltagi y Wu, 1999)

- 1 Construir los residuos de un modelo simple de FE, $\{\tilde{v}\}_{it}$.
- 2 Estimar

$$\tilde{\alpha} = \frac{NT}{N(T-1)} \frac{\sum_{i=1}^N \sum_{t=2}^T \tilde{v}_{it} \tilde{v}_{it-1}}{\sum_{i=1}^N \sum_{t=1}^T \tilde{v}_{it}^2}.$$

- 3 Eliminar AR(1) aplicando la transformación C_α a los residuos:

$$C_\alpha = \begin{bmatrix} (1 - \alpha^2)^{1/2} & 0 & 0 & \cdots & 0 & 0 & 0 \\ -\alpha & 1 & 0 & \cdots & 0 & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & -\alpha & 1 & 0 \\ 0 & 0 & 0 & \cdots & 0 & -\alpha & 1 \end{bmatrix}.$$

Equivalente a

$$\tilde{a}_{ijt} = \begin{cases} (1 - \tilde{\alpha}^2)^{1/2} a_{ijt} & \text{if } t = 1 \\ (1 - \tilde{\alpha}^2)^{1/2} \left[\left(\frac{1}{1 - \tilde{\alpha}^2} \right)^{1/2} a_{ijt} - \left(\frac{\tilde{\alpha}^2}{1 - \tilde{\alpha}^2} \right)^{1/2} a_{ijt-1} \right] & \text{if } t > 1 \end{cases}.$$

- 4 Construir decomposición espectral (no tan fácil)

Estimación

Podemos entonces resumir las siguientes estrategias para obtener errores estándar correctos:

- Los errores estándar de OLS son incorrectos porque no toman en cuenta la correlación intra-cluster.
- Los errores estándar de RE son incorrectos porque asumen que la correlación es constante intra-cluster (equicorrelación).
- Habría entonces que identificar la estructura correcta de la matriz correlaciones de los errores.
 - 1 Una posibilidad es estimar la autocorrelación (AR o MA), luego los efectos aleatorios.
 - 2 Otra alternativa es usar clusters y errores robustos.
- Angrist y Pischke (2009, ch.8) “The clustered variance estimator [...] is consistent as the number of groups gets large under any within-group correlation structure.” (p.313)
- Wooldridge (2010) recomienda implementar estimador RE, que es probable que sea más eficiente que OLS, pero “to make the variance estimator of the random effects robust to arbitrary heteroskedasticity and within-group correlation” (p.867).

Estimación

- 1 En STATA ver xtregar: <http://www.stata.com/manuals13/xtxtregar.pdf>
- 2 En STATA ver opción cluster y/o robust después de cualquier estimador.

Paneles autorregresivos

Paneles autorregresivos

(caracterización general de Arellano, cap.6)

- Supongamos un modelo donde $\{(y_{i0}, y_{i1}, \dots, y_{iT}, \mu_i), i = 1, 2, \dots, N\}$ es una muestra aleatoria con segundos momentos finitos, tal que

$$y_{it} = \alpha y_{it-1} + \mu_i + v_{it}, t = 1, 2, \dots, T, |\alpha| < 1.$$

- Supuesto B1:

$$E(v_{it} | y_i^{t-1}, \mu_i) = 0$$

donde $y_i^{t-1} = (y_{i0}, y_{i1}, \dots, y_{i(t-1)})$ y $E(\mu_i) = 0$, $\text{var}(\mu_i) = \sigma_\mu^2$.

Este supuesto implica que $E(v_{it} v_{i(t-j)} | y_i^{t-1}, \mu_i) = 0, j > 0$ con lo que los errores no están autocorrelacionados.

- Supuesto B2: Homoscedasticidad condicional:

$$E(v_{it}^2 | y_i^{t-1}, \mu_i) = \sigma_t^2$$

- Supuesto B3: Homoscedasticidad de series de tiempo (no condicional):

$$E(v_{it}^2) = \sigma_v^2$$

Paneles autorregresivos

(caracterización general de Arellano, cap.6)

- Supuesto B4: (estacionariedad en la media)

$$E(y_{i0}|\mu_i) = \frac{\mu_i}{1 - \alpha}$$

en cuyo caso $E(y_{it}|\mu_i) = E(y_{i0}|\mu_i)$ (media de steady state) para todo t .

- Supuesto B5: (estacionariedad en la media)

$$\text{var}(y_{i0}|\mu_i) = \frac{\sigma_v^2}{1 - \alpha}$$

- Calcular las covarianzas

$$\text{cov}(y_{it}, y_{i(t-j)}|\mu_i) = \alpha^{2t-j} \text{var}(y_{i0}|\mu_i) + \alpha^j \left(\sum_{s=0}^{t-j-1} \alpha^{2s} \right) \sigma_v^2, j \geq 0.$$

Sesgo de paneles dinámicos, OLS

Para obtener el sesgo del estimador OLS

$$\begin{aligned}\hat{\alpha}_{OLS} &= \frac{\sum_{i=1}^N \sum_{t=1}^T y_{i,t} y_{i,t-1}}{\sum_{i=1}^N \sum_{t=1}^T y_{i,t-1}^2} \\ &= \alpha + \frac{\sum_{i=1}^N \sum_{t=1}^T (\mu_i + v_{it}) y_{i,t-1}}{\sum_{i=1}^N \sum_{t=1}^T y_{i,t-1}^2}\end{aligned}$$

Sesgo de paneles dinámicos, OLS

Calculamos el límite en probabilidad de numerador y denominador (Hsiao, 2003, p.74)

$$\begin{aligned}
 & \text{plim}_{N \rightarrow \infty} \frac{1}{NT} \sum_{i=1}^N \sum_{t=1}^T (\mu_i + v_{it}) y_{i,t-1} \\
 &= \frac{1}{T} \frac{1 - \alpha^T}{1 - \alpha} \text{Cov}(\mu_i, y_{i0}) + \frac{1}{T} \frac{\sigma_\mu^2}{(1 - \alpha)^2} [(T - 1) - T\alpha + \alpha^T] \\
 & \text{plim}_{N \rightarrow \infty} \frac{1}{NT} \sum_{i=1}^N \sum_{t=1}^T y_{i,t-1}^2 \\
 &= \frac{1}{T} \frac{1 - \alpha^{2T}}{(1 - \alpha^2)} \frac{\sum_{i=1}^N y_{i0}^2}{N} + \frac{1}{T} \frac{\sigma_\mu^2}{(1 - \alpha)^2} \left[T - 2 \frac{1 - \alpha^T}{1 - \alpha} + \frac{1 - \alpha^{2T}}{1 - \alpha^2} \right] \\
 & \quad + \frac{1}{T} \frac{2}{(1 - \alpha)} \left[\frac{1 - \alpha^T}{1 - \alpha} - \frac{1 - \alpha^{2T}}{1 - \alpha^2} \right] \text{Cov}(\mu_i, y_{i0}) \\
 & \quad + \frac{1}{T} \frac{\sigma_v^2}{(1 - \alpha)^2} [(T - 1) - T\alpha^2 + \alpha^{2T}]
 \end{aligned}$$

Sesgo de paneles dinámicos, OLS

- En general $\hat{\alpha}_{OLS} \geq \alpha$, OLS sobrestima el valor real, bajo muchos supuestos. En particular asumiendo que $Cov(\mu_i, y_{i0}) > 0$.
- El resultado se mantiene con otras variables de control.
- No es tan claro con otro orden de rezagos.

Sesgo de paneles dinámicos, Nickell (1981)

El estimador de efectos fijos producirá la transformación **within** de todas las variables.
En particular,

$$\tilde{y}_{i,t-1} = y_{i,t-1} - \bar{y}_{i,-1}$$

donde $\bar{y}_{i,-1} = \sum_{t=1}^T y_{i,t-1} / T$

El principal problema es que $\tilde{y}_{i,t-1}$ estará correlacionado con $(u_{i,t-1} - \bar{u}_i)$ aún si u_{it} no tiene correlación serial.

Nickell, S. (1981) "Biases in Dynamic Models with Fixed Effects," Econometrica, 49, 1417-1426.

*El estimador de efectos fijos (within) es sesgado de orden $O(1/T)$, o sea, el sesgo desaparece solo cuando $T \rightarrow \infty$. Entonces habrá un gran sesgo para **T chico** (incluso cuando N sea grande).*

Este problema también se aplica a los modelos de primeras diferencias. La razón es que $\Delta y_{i,t-1} = y_{i,t-1} - y_{i,t-2}$ está correlacionado con $\Delta u_{i,t} = u_{i,t} - u_{i,t-1}$.

Sesgo de paneles dinámicos, Nickell (1981)

Para obtener el sesgo del estimador de efectos fijos debemos usar

$$\hat{\alpha}_{FE} = \frac{\sum_{i=1}^N y'_{i,-1} Q_{\mu} y_i}{\sum_{i=1}^N y'_{i,-1} Q_{\mu} y_{i,-1}}$$

donde $y_{i,-1} = \{y_{iT-1}, y_{iT-2}, \dots, y_{i0}\}$, FE: within-group = efectos fijos.
Para cualquier i

$$\begin{aligned} E(y'_{i,-1} Q_{\mu} y_i) &= \alpha E(y'_{i,-1} Q_{\mu} y_{i,-1}) + E(y'_{i,-1} Q_{\mu} v_i) \\ &= \alpha E(y'_{i,-1} Q_{\mu} y_{i,-1}) + E\left(\sum_{t=1}^T [y_{it-1} (v_{it} - \bar{v}_i)]\right) \end{aligned}$$

Sesgo de paneles dinámicos, Nickell (1981)

El sesgo aparece porque

$$E \left(\sum_{t=1}^T [y_{it-1}(v_{it} - \bar{v}_i)] \right) = -E \left(\sum_{t=1}^T [y_{it-1}\bar{v}_i] \right) \neq 0$$

Notemos que

$$\begin{aligned} y_{it-1} &= \alpha y_{it-2} + \mu_i + v_{it-1} = \alpha^2 y_{it-3} + (1 + \alpha)\mu_i + v_{it-1} + \alpha v_{it-2} \\ &= \left(\sum_{j=0}^{t-2} \alpha^j \right) \mu_i + \sum_{j=0}^{t-2} \alpha^j v_{it-1-j} + \alpha^{t-1} y_{i0} \end{aligned}$$

También,

$$\bar{v}_i = \frac{1}{T} \sum_{t=1}^T v_{it} = \frac{1}{T} (v_{iT} + v_{iT-1} + \dots + v_{it-1} + v_{it-2} + \dots + v_{i1})$$

$$E[y_{it-1}\bar{v}_i] = \frac{1}{T} \sigma_v^2 (1 + \alpha + \dots + \alpha^{t-2})$$

Sesgo de paneles dinámicos, Nickell (1981)

Nickell (1981) encuentra que

$$E(y'_{i,-1} Q v_i) = -\sigma_v^2 h_T(\alpha)$$

donde $h_T(\alpha) = \frac{1}{1-\alpha} \left[1 - \frac{1}{T} \left(\frac{1-\alpha^T}{1-\alpha} \right) \right]$. Además,

$$E(y'_{i,-1} Q y_{i,-1}) = \frac{\sigma_v^2 (T-1)}{(1-\alpha^2)} \left(1 - \frac{2\alpha h_T(\alpha)}{(T-1)} \right)$$

Entonces,

$$plim_{N \rightarrow \infty} (\hat{\alpha}_{FE} - \alpha) = -\frac{(1-\alpha^2) h_T(\alpha)}{(T-1)} \left(1 - \frac{2\alpha h_T(\alpha)}{(T-1)} \right)^{-1} = O(1/T)$$

Cuando $T = 2$, $plim_{N \rightarrow \infty} (\hat{\alpha}_{FE} - \alpha) = -\frac{1+\alpha}{2}$

Sesgo de paneles dinámicos, Nickell (1981)

- Consideremos ahora el estimador FD (primeras diferencias):
 $\Delta y_{it} = \alpha \Delta y_{it-1} + \Delta v_{it}$, $|\alpha| < 1$. Notemos que $\Delta y_{it-1} = \alpha \Delta y_{it-2} + \Delta v_{it-1}$.
- Además, $plim_{N \rightarrow \infty} \hat{\alpha}_{FD} = \frac{E(\Delta y_{it-1} \Delta y_{it})}{E(\Delta y_{it-1})^2}$.
- $E(\Delta y_{it-1} \Delta y_{it}) = \alpha E[(\Delta y_{it-1})^2] + E(\Delta v_{it-1} \Delta v_{it}) = \alpha E[(\Delta y_{it-1})^2] - \sigma_v^2$
 (usando $E(\Delta y_{it-2} \Delta v_{it}) = 0$)
- $E[(\Delta y_{it-1})^2] = \alpha^2 E[(\Delta y_{it-2})^2] + 2\alpha E(\Delta y_{it-2} \Delta v_{it-1}) + E[(\Delta v_{it-1})^2]$
 Entonces, asumiendo estacionariedad (ej. $E[(\Delta y_{it-1})^2] = E[(\Delta y_{it-2})^2]$)

$$E[(\Delta y_{it-1})^2] = \frac{2\sigma_v^2(1-\alpha)}{1-\alpha^2} = \frac{2\sigma_v^2}{1+\alpha}$$
- Entonces, $plim_{N \rightarrow \infty} (\hat{\alpha}_{FD} - \alpha) = -\frac{(1+\alpha)}{2}$. Notar que no depende de T .

Magnitud del sesgo dinámico para estimadores de efectos fijos

$$y_{it} = \alpha y_{it-1} + \mu_i + v_{it}, |\alpha| < 1.$$

T	α							
	0	0.1	0.25	0.5	0.75	0.9	0.95	0.99
2	-0.500	-0.550	-0.625	-0.750	-0.875	-0.950	-0.975	-0.995
3	-0.333	-0.373	-0.433	-0.536	-0.642	-0.706	-0.728	-0.746
5	-0.200	-0.224	-0.261	-0.331	-0.411	-0.463	-0.481	-0.496
10	-0.100	-0.111	-0.129	-0.162	-0.207	-0.243	-0.257	-0.270
15	-0.067	-0.074	-0.085	-0.106	-0.135	-0.162	-0.174	-0.185
20	-0.050	-0.055	-0.063	-0.078	-0.099	-0.120	-0.130	-0.140

Sesgo asintótico de paneles dinámicos

- Aún cuando $T \rightarrow \infty$, puede haber un sesgo si también $N \rightarrow \infty$
- Si $N/T \rightarrow 0$ entonces no hay sesgo de paneles dinámicos. La intuición es que para cada i se puede correr una regresión individual.
- (Alvarez and Arellano, 2003) Pero para $N/T \rightarrow k < \infty$ y $N/T^3 \rightarrow 0$,

$$\sqrt{NT} [\hat{\alpha}_{FE} - (\alpha - 1/T(1 + \alpha))] \xrightarrow{d} N(0, (1 - \alpha^2))$$

- Nickel (1981) propone usar el factor de ajuste $-1/T(1 + \alpha)$ para T grande.

Anderson y Hsiao (1981)

Anderson, T.W. and Hsiao, C. (1981) "Estimation of Dynamic Models with Error Components," *Journal of the American Statistical Association*, 76, 598-606.

- Notar que tanto los estimadores FE como FD no son consistentes para T finito.
- Anderson y Hsiao (1981) proponen usar un estimador VI (variables instrumentales) para obtener estimadores consistentes. Su principal idea se resume en:

- 1 Usar primeras diferencias FD del modelo para cancelar μ_i . Sin embargo, este estimador es inconsistente...
- 2 Entonces usar $\Delta y_{i,t-2}$ o $y_{i,t-2}$ como variable instrumental para $\Delta y_{i,t-1}$.
- 3 ¿Por qué?
 1. $E[y_{i,t-2}\Delta v_{i,t}] = E[y_{i,t-2}(v_{i,t} - v_{i,t-1})] = 0$
 2. $E[y_{i,t-2}\Delta y_{i,t-1}] = E[y_{i,t-2}(y_{i,t-1} - y_{i,t-2})] = -\sigma_v^2(1 - \alpha) \left(\frac{1 - \alpha^{2(t-1)}}{1 - \alpha^2} \right) \neq 0$
[Probar que funciona para $\Delta y_{i,t-2}$.]

¿Cómo funciona IV? (revisión)

- Consideremos el modelo $y = X\beta + u$ donde $\text{cov}(X, u) \neq 0$, o sea endogeneidad.

$$\begin{aligned} \text{plim } \hat{\beta}_{OLS} &= E[X'X]^{-1}E[X'y] = E[X'X]^{-1}E[X'(X\beta + u)] \\ &= \beta + E[X'X]^{-1}E[X'u] \neq \beta \end{aligned}$$

- $E[X'X]^{-1}E[X'u]$ es el sesgo asintótico de OLS con endogeneidad.
- Supongamos que Z es una variable instrumental. Tiene que satisfacer dos condiciones:
 - $E[Z'u] = 0$, exogeneidad.
 - $E[Z'X] \neq 0$, correlación con la variable endógena.

$$\text{Entonces } \text{plim } \hat{\beta}_{IV} = E[Z'X]^{-1}E[Z'y] = E[Z'X]^{-1}(E[Z'X]\beta + E[Z'u]) = \beta.$$

Anderson and Hsiao (1981)

En el modelo simple, $y_{it} = \alpha y_{it-1} + \mu_i + v_{it}$, $|\alpha| < 1$, usando $y_{i,t-2}$ como IV,

$$plim \hat{\alpha}_{IV} = \frac{E(y_{it-2} \Delta y_{it})}{E(y_{it-2} \Delta y_{it-1})}$$

de donde

$$E(y_{it-2} \Delta y_{it}) = \alpha E(y_{it-2} \Delta y_{it-1}) + E(y_{it-2} \Delta v_{it})$$

La clave está en que está correlacionado con $\Delta y_{i,t-1}$ pero no con $\Delta u_{i,t}$.

- 1 $E[y_{i,t-2} \Delta v_{i,t}] = E[(\alpha y_{i,t-3} + v_{i,t-3})(v_{i,t} - v_{i,t-1})] = 0$ (asumiendo que no hay correlación serial en los errores);
- 2 $E[y_{i,t-2} \Delta y_{i,t-1}] = E[y_{i,t-2}(y_{i,t-1} - y_{i,t-2})] \neq 0$.

Arellano, M. and Bond, S. (1991) "Some tests of specification for panel data: Monte Carlo evidence and an application to employment equations," Review of Economic Studies, 58, 277–297.

Blundell, R. and Bond, S. (1998) "Initial conditions and moment restrictions in dynamic panel data models," Journal of Econometrics, 87, 115–43.

Arellano and Bond (1991)

- Consideremos el modelo en primeras diferencias

$$\Delta y_{it} = \alpha \Delta y_{it-1} + \Delta v_{it} = y_{it} - y_{it-1} = \alpha(y_{it-1} - y_{it-2}) + (v_{it} - v_{it-1})$$

- Para $t=3$, $y_{i3} - y_{i2} = \alpha(y_{i2} - y_{i1}) + (v_{i3} - v_{i2})$. Notemos que y_{i1} es un instrumento válido ya que está correlacionado con $(y_{i3} - y_{i2})$ pero no con $(v_{i3} - v_{i2})$. (siempre que los errores no están autocorrelacionados)
- Para $t=4$, $y_{i4} - y_{i3} = \alpha(y_{i3} - y_{i2}) + (v_{i4} - v_{i3})$. Notemos que y_{i2} y y_{i1} son instrumentos válidos para $(y_{i3} - y_{i2})$ (no están correlacionados con $(v_{i4} - v_{i3})$).
- Podemos continuar de esta manera construyendo instrumentos adicionales para cada periodo, tal que para T , el conjunto de instrumentos es $(y_{i1}, y_{i2}, \dots, y_{iT-2})$.
- Entonces tenemos $t - 2$ condiciones de momento a explotar para cada t .
- En total, bajo el supuesto B1 de Arellano (2002) tenemos $T(T - 1)/2$ restricciones de momentos a explotar.
- Más condiciones de momento implica que IV simple (Anderson-Hsiao) es ineficiente, se puede construir un estimador GMM mejor.

Estimadores GMM

(ver notas optativas sobre estimadores M)

- GMM (Generalized Method of Moments) es un método para generar estimadores a partir de condiciones de momento. Supongamos que $g(Z, \theta)$ es una condición de momento tal que para θ_0 satisface $E[g(Z, \theta)] = 0$ si y solo si $\theta = \theta_0$, para variables aleatorias Z (por ejemplo $Z = (X, Y)$).
- $\hat{\theta}_{GMM} = \underset{\theta}{\operatorname{argmin}} \left[n^{-1} \sum_{i=1}^N g(z_i, \theta) \right]' \Omega \left[n^{-1} \sum_{i=1}^N g(z_i, \theta) \right]$ para una matriz Ω definida positiva.
- La matriz Ω determina la eficiencia del estimador. La óptima es $\Omega = E [g(Z, \theta_0)g(Z, \theta_0)']^{-1}$.

Arellano and Bond (1991)

- Construyamos Z como el conjunto de IV. $Z'\Delta y = Z'(\Delta y_{-1})\alpha + Z'\Delta v$

$$Z_i = \begin{bmatrix} [y_{i1}] & 0 & \cdots & 0 \\ 0 & [y_{i1}, y_{i2}] & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & [y_{i1}, \dots, y_{i,T-2}] \end{bmatrix}$$

- $\hat{\alpha} = [(\Delta y_{-1})' Z \Omega Z' (\Delta y_{-1})]^{-1} [(\Delta y_{-1})' Z \Omega Z' (\Delta y)]$

Arellano and Bond (1991)

- Estimador GLS en una etapa (one-step estimator): $\hat{\Omega} = \sum_{i=1}^N Z_i' G Z_i$ para una matriz G tal que $E(\Delta v_i \Delta v_i') = \sigma_v^2 G$. En particular se usa

$$\hat{\Omega} = \sum_{i=1}^N Z_i' D D' Z_i$$

donde $D = \begin{bmatrix} -1 & 1 & 0 & \cdots & 0 \\ 0 & -1 & 1 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots & \\ 0 & 0 & 0 & \cdots & 1 \end{bmatrix}$ es la matriz $(T-1) \times T$ de operador

en primeras diferencias.

- Estimadore GMM en dos etapas (two-step estimator):
 $\hat{\Omega} = \sum_{i=1}^N Z_i' (\Delta \hat{v}_i) (\Delta \hat{v}_i)' Z_i$ para una estimación preliminar (ej. one-step) de \hat{v} .

Arellano and Bond (1991)

El modelo también admite otras variables exógenas x que pueden estar relacionadas con μ_i , entonces x_s son instrumentos válidos.

Las variables explicativas *exógenas* se pueden separar en dos tipos:

- Las variables **predeterminadas** donde $E[x_{it}v_{is}] \neq 0$ para $s < t$ pero $E[x_{it}v_{is}] = 0$ para $s \geq t$. Entonces $[x'_{i1}, x'_{i2}, \dots, x'_{i(s-1)}]$ son instrumentos válidos para la ecuación en diferencias en el periodo s .
- Por ejemplo, x'_{i1} y x'_{i2} son instrumentos válidos para $y_{i3} - y_{i2} = \alpha(y_{i2} - y_{i1}) + (x'_{i3} - x'_{i2})\beta + (v_{i3} - v_{i2})$. Tenemos entonces, la matriz de instrumentos Z . $Z'\Delta y = Z'(\Delta y_{-1})\alpha + Z'(\Delta X)\beta + Z'\Delta v$

$$Z_i = \begin{bmatrix} [y_{i1}, x'_{i1}, x'_{i2}] & 0 & \dots & 0 \\ 0 & [y_{i1}, y_{i2}, x'_{i1}, x'_{i2}, x'_{i3}] & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & [y_{i1}, \dots, y_{i,T-2}, x'_{i1}, \dots, x'_{iT}] \end{bmatrix}$$

- Las variables **estrictamente exógenas** se pueden usar $[x'_{i1}, x'_{i2}, \dots, x'_{iT}]$ como un elemento más de la diagonal de Z_i .
- La metodología de Arellano-Bond se puede usar para la presencia de otras variables endógenas. Es decir, la estructura de rezagos en paneles dinámicos sirve para resolver casos de endogeneidad general.

Arellano and Bond (1991)

- El estimador de Arellano-Bond depende de la ausencia de correlación serial de segundo orden en los errores de la ecuación en diferencias. O sea, $E[\Delta v_{it} \Delta v_{it-2}] = 0$. AB proponen un contraste para eso que debe ser chequeado. Bajo la hipótesis nula los errores no tienen autocorrelación AR(2) (o mayor) o siguen un paseo aleatorio. ¿Qué hacer si este supuesto se rechaza?
- Arellano-Bond también proponen un contraste de Sargan-Hansen (overidentifying restrictions):

$$m = \Delta \hat{v}' Z \left[\sum_{i=1}^N Z_i' \Delta(\hat{v}_i) (\Delta(\hat{v}_i))' Z_i \right] Z' \Delta \hat{v} \sim \chi^2_{\#rest - \#par}$$

Contraste para la validez de los instrumentos

Contraste de Sargan-Hansen

Un supuesto crucial de GMM es que los instrumentos son válidos si no están correlacionados con los residuos. Supongamos que queremos armar un contraste para la validez de los instrumentos.

Requerimiento importante: Necesitamos **más** variables instrumentales que variables endógenas. Tomemos el modelo

$$y_1 = \beta_0 + \beta_1 y_2 + \beta_2 z_1 + \beta_3 z_2 + u$$

donde y_2 es (potencialmente) endógena; z_1 and z_2 son variables explicativas exógenas; z_3 and z_4 son IV.

- 1 Supongamos que en el modelo anterior usamos 2SLS con z_3 como la **única** variable instrumental.
- 2 Computar $\hat{u}_3 = y_1 - \hat{\beta}_0 - \hat{\beta}_1 y_2 - \hat{\beta}_2 z_1 - \hat{\beta}_3 z_2$.
- 3 Correr la regresión auxiliar $\hat{u}_3 = \delta_0 + \hat{\delta}_1 z_1 + \hat{\delta}_2 z_2 + \delta_4 z_4$.
- 4 Chequer la significancia de z_4 .
- 5 Esto nos da un contraste válido para la validez de z_4 como IV. Pero tenemos que asumir que z_3 es una IV válida.

Contraste para la validez de los instrumentos

Contraste de Sargan-Hansen

- 1 Si tenemos más IVs que variables endógenas, entonces el modelo está sobre-identificado (over-identified).
- 2 Consideremos H_0 : todas las IVs son exógenas. Si rechazamos entonces alguna de las IVs es endógena.
- 3 Estimar el modelo con todas las IVs usando 2SLS. Obtener los residuos \hat{u} .
- 4 Correr la regresión de \hat{u} en TODAS las variables exógenas (IVs, X exógenas, constante).
- 5 Computar $NR_u^2 \stackrel{a}{\sim} \chi_{L-K}^2$, donde R_u^2 es el de la última regresión.

Contraste para la validez de los instrumentos

Contraste de Sargan-Hansen

- Supongamos que \hat{e} son los residuos del modelo de 2SLS.
- Si el modelo está sobreidentificado entonces $Z'\hat{e}$ no es exactamente 0, y $\frac{1}{\sqrt{N}}Z'\hat{e} \xrightarrow{d} N(0, V_{Z-e})$.
- Entonces podemos armar $\frac{1}{\sqrt{N}}Z'\hat{e}V_{Z-e}^{-1}\frac{1}{\sqrt{n}}Z'\hat{e}$, donde V_{Z-e} es la varianza de $Z'e$.
- Los contrastes de Sargan y Hansen difieren en como V_{Z_e} es estimada.
- Contraste de Sargan: asume homoscedasticidad. Usa estimadores one-step: $V_{Z_e} = Z'Z$.
- Contraste de Hansen: permite heteroscedasticidad, usa two-step.

Ahn and Schmidt (1995)

- Ahn y Schmidt (1995) muestran que con supuestos no muy fuertes hay condiciones adicionales que se pueden usar.
- Arellano-Bond usan $T(T-1)/2$ condiciones de momento usando $E(y_{is}\Delta v_{it}) = 0, t = 2, 3, \dots, T, s = 0, \dots, t-2$.
- Sin embargo también están disponibles las condiciones $E(v_{iT}\Delta v_{it}) = 0, t = 2, \dots, T-1$ usando el supuesto de varianza constante $E(v_{it}^2) = \sigma_i^2, t = 1, 2, \dots, T$
- Hay entonces un conjunto de $T(T-1)/2 + (T-1)$ condiciones de momento.

Blundell and Bond (1998)

- Blundell y Bond (1998) encuentran que el estimador de Anderson-Hsiao cumple con la siguiente propiedad. Supongamos que $T = 3$ tal que α está exactamente identificado de $E(y_{i1}\Delta v_{i3}) = 0$.
- En este caso la primera etapa de 2SLS se obtiene de una regresión de Δy_{i2} en y_{i1} : $\Delta y_{i2} = (\alpha - 1)y_{i1} + \mu_i + v_{i2}$ (asumiendo $E(v_{i1} = 0)$). Entonces,

$$plim(\hat{\alpha} - 1) = (\alpha - 1) \frac{c}{c + \sigma_{\mu}^2 / \sigma_v^2},$$

donde $c = (1 - \alpha) / (1 + \alpha)$. El estimador es sesgado hacia abajo (cuanto más grande es $\sigma_{\mu}^2 / \sigma_v^2$).

- También el contraste F se vuelve $F = \frac{(\sigma_v^2 c)^2}{\sigma_{\mu}^2 + \sigma_v^2 c} \rightarrow 0$ cuando $\alpha \rightarrow 1$.
 $E[y_{i,t-2}\Delta y_{i,t-1}] = E[y_{i,t-2}(y_{i,t-1} - y_{i,t-2})] = -\sigma_v^2(1 - \alpha) \left(\frac{1 - \alpha^{2(t-1)}}{1 - \alpha^2} \right) \neq 0$
- Blundell-Bond entonces buscan usar condiciones adicionales: rezagos en diferencias para la ecuación en niveles. El supuesto necesario es que:

$$E\left[\left(y_{i1} - \frac{\mu_i}{1 - \alpha}\right)\mu_i\right] = 0$$

tal que y converge a su media $\frac{\mu_i}{1 - \alpha}$ para cada individuo a partir de $t = 2$.

Blundell and Bond (1998)

- Con el supuesto adicional $E[\Delta v_{it} \mu_i] = 0$ tenemos $T - 2$ restricciones adicionales

$$E[\Delta y_{it}(\mu_i + v_{it})] = 0, t = 3, 4, \dots, T.$$

- Particularidad: $\Delta y_{i,t-1}$ se usan como instrumentos en **niveles**.
- Blundell-Bond desarrollan un estimador GMM que toma las condiciones de Arellano-Bond junto con las $T - 2$ condiciones adicionales.
- El estimador se llama **System GMM**: estima simultáneamente dos ecuaciones, una en niveles y otra en diferencias, usando diferentes instrumentos para cada una.
- Gana eficiencia para T chico.

Motivación: Roodman (2009,p136)

136

Bulletin

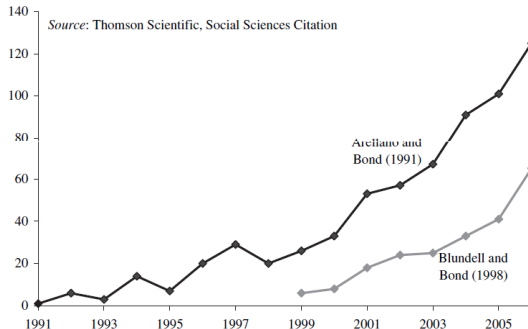


Figure 1. Citations of Arellano and Bond (1991) and Blundell and Bond (1998) per year, 1991–2006

Proliferación de instrumentos: muchos instrumentos

Arellano-Bond usan $T(T - 1)/2$ condiciones de momento usando $E(y_{is}\Delta v_{it}) = 0$, $t = 2, 3, \dots, T$, $s = 0, \dots, t - 2$. Alvarez y Arellano (2003) muestran que a menos que $\lim(T/N) = 0$, el estimador tendrá un sesgo de orden $O(1/N)$. Hay dos problemas principales con muchos instrumentos:

- 1 Sobreestimación (overfitting)
- 2 Mala estimación de la matriz de pesos Ω .

Proliferación de instrumentos: muchos instrumentos

- Cuando el nro de IV crece, entonces sobreestima (overfits) la variable endógena. Ejemplo: Si $\#IV = NT$ (tantas observaciones como IV), entonces en la primera etapa de 2SLS $R^2 = 1$, y por construcción $OLS = 2SLS$.
- Arellano (2003) muestra que el sesgo de overfitting es $O(j/N)$, j es el nro de instrumentos para variables pre-determinadas, y $O(jT/N)$ para variables endógenas.

Proliferación de instrumentos: muchos instrumentos

- Cuando el nro de IV crece, la matriz de pesos Ω se estima de forma imprecisa.
- La matriz óptima Ω depende de $var(Z'\nu)$.
- Esta es cuadrática en el nro de IV, lo que significa cuártica en T , o sea $O(T^4)$.
- Se requieren momentos segundos de $Z'\nu$, que implica momentos cuartos de las distribuciones subyacentes (Hayashi, 2000, p.215).
 - 1 La matriz estimada puede ser singular.
 - 2 Sesgo hacia abajo en los errores estándar: muestra más significancia de la que verdaderamente hay. El problema es que no se toma en cuenta que Ω es estimada.
 - 3 El contraste de Sargan-Hansen (para validez de IV) es no creíble. En general, un alto p-valor del contraste se usa como validez de los IV. Pero el paper de Bowsher (2000) muestra que muchos IV vuelve el test de Sargan-Hansen poco creíble.

"A few calculations were made by the author on the order of magnitude of the errors involved in this approximation. They were found to be proportional to the number of instrumental variables, so that, if the asymptotic approximations are to be used, this number must be small." (Sargan, 1958)

Proliferación de instrumentos: muchos instrumentos

- Soluciones propuestas:

- 1 Regla: El nro de IV no debería ser mayor que N . Es decir, mantener el nro. de IV cercano al nro de individuos. [Nota: Roodman (2009) en realidad dice que esto no tiene ningún valor estadístico.]
- 2 Probar diferentes especificaciones de IV. El problema de IV no afecta la consistencia de GMM, entonces en muestras chicas es importante analizar su comportamiento. Si el modelo funciona debería dar resultados similares con distintos IV. Probar cortar los rezagos en 1 o 2, que es equivalente a poner 0 en la matriz Z .
- 3 Colapsar/combinar instrumentos. Ventaja que toda la información se usa (todos los IVs). Es equivalente a imponer que ciertos conjuntos de parámetros tienen el mismo coeficiente. Busca minimizar $\sum_t y_{i,t-l} \Delta v_{it}$ para cada l en vez de $y_{i,t-l} \Delta v_{it}$ para cada t, l .

$$Z_i = \begin{bmatrix} y_{i1} & 0 & 0 & \cdots \\ y_{i2} & y_{i1} & 0 & \cdots \\ y_{i3} & y_{i2} & y_{i1} & \cdots \\ \vdots & \vdots & \vdots & \ddots \end{bmatrix} \begin{bmatrix} \Delta y_{i2} \\ \Delta y_{i3} \\ \Delta y_{i4} \\ \vdots \end{bmatrix}$$

Simulation: Roodman (2009,p150)

150

Bulletin

TABLE 1
Simulation results for $N = 100$, $\alpha = 0.8$, varying ρ , T , and instrument set

Time dimension (T)	$\rho = 0.0$ (system GMM valid)				$\rho = 0.9$ (system GMM invalid)			
	5	10	15	20	5	10	15	20
Full instrument set								
Instrument count	10	45	105	190	10	45	105	190
Estimated α mean	0.86	0.84	0.83	0.82	1.03	0.94	0.87	0.84
Estimated α standard dev.	0.12	0.06	0.04	0.03	0.05	0.04	0.03	0.02
Hansen JP average	0.459	0.410	0.608	1.000	0.314	0.107	0.603	1.000
Lag 1 only								
Instrument count	7	17	27	37	7	17	27	37
Estimated α mean	0.87	0.85	0.84	0.85	1.02	0.94	0.91	0.89
Estimated α standard dev.	0.12	0.07	0.06	0.04	0.06	0.04	0.03	0.03
Hansen JP average	0.469	0.462	0.455	0.438	0.284	0.123	0.150	0.199
Collapsed								
Instrument count	5	10	15	20	5	10	15	20
Estimated α mean	0.80	0.80	0.80	0.80	1.04	0.92	0.83	0.80
Estimated α standard dev.	0.17	0.08	0.05	0.04	0.05	0.06	0.06	0.04
Hansen JP average	0.488	0.488	0.467	0.468	0.267	0.060	0.105	0.199
Lag 1 only, collapsed								
Instrument count	3	3	3	3	3	3	3	3
Estimated α mean	0.81	0.80	0.80	0.80	1.04	0.98	0.90	0.84
Estimated α standard dev.	0.18	0.09	0.06	0.05	0.06	0.06	0.08	0.07
Hansen JP average	0.510	0.513	0.488	0.501	0.205	0.018	0.012	0.017
Observations	300	800	1,300	1,800	300	800	1,300	1,800

Note: All instrument sets include the constant term.



www.elsevier.com/locate/worlddev

World Development Vol. 44, pp. 77–87, 2013
© 2012 Elsevier Ltd. All rights reserved.
0305-750X/\$ - see front matter

<http://dx.doi.org/10.1016/j.worlddev.2012.12.006>

Can Poor Countries Lobby for More US Bilateral Aid?

GABRIEL V. MONTES-ROJAS*

City University London, UK

Summary. — This article explores if countries can lobby the US government for the allocation of US bilateral foreign aid. We consider an informational lobby model where lobbying has two effects. First, a direct effect by informing US policymakers about their countries' needs. Second, an indirect effect on policymakers by informing them about common interests in economic or geopolitical terms. The lobbyist thus influences the decisions about the allocation of aid resources. We estimate the effect of the recipient country's lobbying agents in obtaining foreign aid. The econometric results show that lobbying positively affects the amount of bilateral aid received.
© 2012 Elsevier Ltd. All rights reserved.

Key words — foreign aid, lobbying, interest groups

1. INTRODUCTION

Bilateral and multilateral aid is increasingly selective and allocated by donors on the basis of objective criteria. Three-fourths of aid agencies, including Denmark, Norway, Sweden, the United Kingdom, Ireland, and the Netherlands,

(2000) find evidence that the direction of foreign aid is dictated by political and strategic considerations, rather than by the economic needs or performance of the recipients. Alesina and Weder (2002) analyze whether corrupt countries receive more aid. Chong, Gradstein, and Calderon (2001) analyze the effect of foreign aid on inequality and poverty.

University, Department of Government (<http://dvn.iq.harvard.edu/dvn/dv/Voeten/faces/study/StudyPage.xhtml?studyId=38311&versionNumber=1&tab=files>). We compute the proportion of votes where each country coincides with the US in the General Assembly on an annual basis.

Recipient countries' needs are captured by aggregate data for all disasters, all countries on an annual basis, estimated damage costs (US\$), from the Emergency Events Database (<http://www.emdat.be/>). We transform the data to real terms and computed the log value (imputing a value of 1 for the non-existent log of 0). This variable is now defined as *Disaster*.

4. ECONOMETRIC RESULTS

(a) *Econometric models*

Our interest lies in evaluating the link between foreign lobbying and foreign aid for the period 1997–2009. Consider a panel data model of the form

$$\ln(Aid_{it}) = \beta \ln(Lobby_{it-1}) + \gamma X_{it} + \mu_i + \delta_t + \epsilon_{it}, \quad (1)$$

where i denotes country, t year, Aid foreign aid, $Lobby$ represents the FARA lobbying variable, X a set of additional control variables, and (μ, δ, ϵ) an error components model with country- and time-specific effects. Country fixed-effects are in-

ter persistence is significant. Second, a recent application of the effect of FARA lobbying in a related context by Gawande *et al.* (2009) proposes to use a dynamic specification to account for the fact that lobbying has both a short-run and long-run effect. Therefore, the proposed dynamic model is

$$\ln(Aid_{it}) = \alpha \ln(Aid_{it-1}) + \beta \ln(Lobby_{it-1}) + \gamma X_{it} + \mu_i + \delta_t + \epsilon_{it}. \quad (2)$$

The long-run effect of *Lobby* on *Aid* is $\frac{\beta}{1-\alpha}$.

In dynamic panel data models with unobserved effects, the treatment of the initial observations is an important theoretical and practical problem. As is well known, the usual within estimator is inconsistent, and can be badly biased. We thus follow the Anderson and Hsiao (1981) and Arellano and Bond (1991) strategy by taking first order differences and using lagged values of the dependent variable and other covariates in levels to instrument the autoregressive dependent variable. These instruments are also valid for other potential endogenous variables. Thus we also use instruments for the lobbying variable while we consider that all other covariates (population, GDP, year dummies) are exogenous. In particular, we implement the Blundell and Bond (1998) System GMM estimator that incorporates information from the levels regression instrumented with lagged differences and has better bias properties. The validity of this econometric method depends on the

Table 3. Effect of lobbying on total US assistance: different GMM estimators

Model	GDPpc ≤ US\$5,000			GDPpc ≤ US\$10,000		
	(1) Collapsed	(2) 1 lag	(3) Full	(4) Collapsed	(5) 1 lag	(6) Full
$\ln(Aid_{i,t-1})$	0.408*** (0.133)	0.464*** (0.103)	0.216** (0.104)	0.417*** (0.130)	0.437*** (0.0952)	0.220** (0.0930)
$\ln(Lobby_{i,t-1})$	0.0408* (0.0243)	0.0614** (0.0240)	0.0646** (0.0279)	0.0439** (0.0211)	0.0666*** (0.0214)	0.0608** (0.0263)
$\ln(GDP_{i,t-1})$	-0.263*** (0.0976)	-0.288*** (0.103)	-0.377*** (0.140)	-0.348*** (0.0912)	-0.377*** (0.0939)	-0.466*** (0.119)
$\ln(Pop_{i,t-1})$	0.695*** (0.160)	0.645*** (0.145)	0.941*** (0.173)	0.763*** (0.174)	0.748*** (0.146)	1.025*** (0.159)
Observations	1,500	1,500	1,500	1,812	1,812	1,812
Number of countries	117	117	117	141	141	141
Arellano-Bond AR(2) stat	-1.099	-0.984	-1.508	-0.755	-0.690	-1.217
AR(2) p-value	0.272	0.325	0.132	0.450	0.490	0.224
Hansen stat	30.05	55.43	102.1	31.18	52.72	130.7
Hansen p-value	0.148	0.137	1.000	0.118	0.200	0.980
#IV	25	47	168	25	47	168

Notes: Robust standard errors in parentheses. Dependent variable $\ln(Aid_{i,t})$. All specifications include year dummies. Collapsed is the Roodman (2009) collapsed instruments System GMM estimator. 1 lag is the System GMM estimator where only 1 lag is used in the GMM generation of instruments. Full is the Blundell and Bond (1998) System GMM estimator with all possible instruments.

* $p < 0.1$.
 ** $p < 0.05$.
 *** $p < 0.01$.



How to do it in STATA?

- First, you need to set-up STATA to identify that you have a panel data structure. For that, you need a numerical variable, say ID, that identifies individuals, regions, etc. Then, type

```
iis ID
```

Alternatively, if you have both a cross-sectional and a time-series structure (say given by TIME)

Organise your data

IND	TIME	YVAR	XVAR
1	1	y ₁₁	x ₁₁
1	2	y ₁₂	x ₁₂
1	3	y ₁₃	x ₁₃
2	1	y ₂₁	x ₂₁
2	2	y ₂₂	x ₂₂
2	3	y ₂₃	x ₂₃

- It is very important that there are no repeated values, i.e. you cannot have twice individual=1 and time=2

- Type

```
tsset IND TIME
```

this recognizes the variable IND as *i*, the individual indicator and TIME as *t*, the time indicator

How to do it in STATA?

- Now you are ready to estimate panel data models:

```
xtreg y x1 x2 x3, fe (FE model)
```

```
xi: reg y x1 x2 x3 i.idvar (same, but including dummies by idvar)
```

```
xtreg y x1 x2 x3, re (RE model)
```

How to do it in STATA?

- Hausman test

```
xtreg y x1 x2 x3, fe (FE model)
est store fe (keep info about FE model)
xtreg y x1 x2 x3, re (RE model)
est store re (keep info about RE model)
hausman fe re (Hausman test)
```

How to do it in STATA?

First-differences model

`tsset idvar timevar` (to get differences, lags, etc. you need the `tsset`)

`reg D.y D.x1 D.x2 D.x3` (to run in FD)

`reg y L.y x1 x2 x3` (OLS with lag of the dependent variable)

`xtreg y L.y x1 x2 x3, fe` (FE with lag of the dependent variable)

Arellano-Bond estimator

There are 3 different alternative commands for Arellano-Bond and System-GMM:

- 1 By default, `xtabond`
`http://www.stata.com/help.cgi?xtabond`
- 2 `xtabond2` as given by Roodman (2009). You need to install it:
`ssc install xtabond2, all replace`
- 3 Another alternative
`http://www.stata.com/help.cgi?xtdpdsys`