Econometría Avanzada

Modelos para Datos Panel II

Manuel Fernández

14 de octubre de 2021

Temas de la clase

1. Panel dinámico
2. Primeras diferencias - IV (Anderson-Hsiao)
3. Arellano-Bond
4. Efectos fijos - *TWFE*
5. Ashenfelter and Krueger (1994)
6. Errores estándar *cluster-robust*

Sección

1. Panel dinámico
2. Primeras diferencias - IV (Anderson-Hsiao)
3. Arellano-Bond
4. Efectos fijos - *TWFE*
5. Ashenfelter and Krueger (1994)
6. Errores estándar *cluster-robust*

Modelo con variable dependiente rezagada

- Modelo a estimar:

$$y_{it} = \alpha y_{it-1} + \mathbf{x}_{it}'\beta + \underbrace{\eta_i + \epsilon_{it}}_{e_{it}}, \tag{1}$$

- y_{it-1} es la variable dependiente rezagada ("*state dependence*").
- Asumimos $|\alpha| < 1$: el modelo es estacionario.
- Estructura de los datos: $\{(y_{it}, \mathbf{x}_{it}) : i = 1, ..., N \quad \text{y} \quad t = 1, ..., T\}$.
- La primera observación para cada unidad i es

$$y_{i2} = \alpha y_{i1} + \mathbf{x}_{i2}'\beta + \eta_i + \epsilon_{i2}. \tag{2}$$

- Tenemos $T - 1$ ecuaciones en niveles para cada unidad i.

Estimadores alternativos

- En el contexto de modelos para datos panel, hemos considerado:
  1. *Pooled MCO*.
  2. Estimador de efectos fijos (transformación *within*).
  3. Estimador de primeras diferencias.
  4. Estimador de primeras diferencias - IV.
- Vamos a mostrar que entre estas alternativas, solo la última es (potencialmente) aplicable cuando hay variable dependiente rezagada.

Descartando *pooled MCO*

- Modelo a estimar:

$$y_{it} = \alpha y_{it-1} + \mathbf{x}'_{it}\beta + \underbrace{\eta_i + \epsilon_{it}}_{e_{it}},$$

- Por construcción, $\mathbb{E}[y_{it-1}\eta_i] \neq 0$:

$$\mathbb{E}[y_{it-1}\eta_i] = \mathbb{E}\left[\underbrace{(\alpha y_{it-2} + \mathbf{x}'_{it-1}\beta + \eta_i + \epsilon_{it-1})}_{y_{it-1}}\eta_i\right] \neq 0. \tag{3}$$

- η_i es parte del proceso que genera y_{it-1}.
- Estimadores *pooled MCO* (y efectos aleatorios) quedan **descartados**.

Descartando estimador FE

- Modelo a estimar:

$$y_{it} = \alpha y_{it-1} + \mathbf{x}_{it}'\beta + \underbrace{\eta_i + \epsilon_{it}}_{e_{it}},$$

2. Por construcción, $\mathbb{E}[y_{it-1}\epsilon_{it-1}] \neq 0$:

$$\mathbb{E}[y_{it-1}\epsilon_{it-1}] = \mathbb{E}\left[\underbrace{(\alpha y_{it-2} + \mathbf{x}_{it-1}'\beta + \eta_i + \epsilon_{it-1})}_{y_{it-1}}\epsilon_{it-1}\right] \neq 0. \tag{4}$$

- ϵ_{it-1} es parte del proceso que genera y_{it-1}.
- **Cualquier transformación** que implique suponer exogeneidad estricta respecto a ϵ_{it} **queda descartada**, incluyendo la transformación *within*.

Descartando el estimador FD

- Empezamos con el modelo simplificado:

$$y_{it} = \alpha y_{it-1} + \eta_i + \epsilon_{it}, \tag{5}$$

- El modelo en primeras diferencias es:

$$\Delta y_{it} = \alpha \Delta y_{it-1} + \Delta \epsilon_{it}, \tag{6}$$

- El estimador MCO del modelo transformado es consistente si

$$\mathbb{E}[\Delta y_{it-1} \Delta \epsilon_{it}] = \mathbb{E}\left[(y_{it-1} - y_{it-2})(\epsilon_{it} - \epsilon_{it-1})\right] = 0.$$

- Dado que $\mathbb{E}[y_{it-1}\epsilon_{it-1}] \neq 0$, el estimador no es consistente.
- **La transformación en primeras diferencias tampoco funciona**.

Sección

Modelo de primeras diferencias - IV

- Modelo a estimar:

$$\Delta y_{it} = \alpha \Delta y_{it-1} + \Delta \epsilon_{it}.$$

- Si y_{it-1} es predeterminada, y_{it-2} es un buen candidato para ser variable instrumental de Δy_{it-1}:

  1. $\mathbb{E}[y_{it-2}\Delta y_{it-1}] = \mathbb{E}\left[y_{it-2}(y_{it-1} - y_{it-2})\right] \neq 0$ (relevancia)
  2. $\mathbb{E}[y_{it-2}\Delta \epsilon_{it}] = \mathbb{E}\left[y_{it-2}(\epsilon_{it} - \epsilon_{it-1})\right] = 0$ (exogeneidad)

- En este caso, la estimación por **variables instrumentales resuelve el problema de endogeneidad**.
- El estimador fue propuesto inicialmente por Anderson and Hsiao (1981).

Modelo de primeras diferencias - IV

- Considere el modelo más general:

$$y_{it} = \alpha y_{it-1} + \mathbf{x}_{it}'\beta + \eta_i + \epsilon_{it}.$$

- Suponga que y_{it-1} es predeterminada.
- Diferentes posibilidades dependiendo de los supuestos que hagamos sobre $\mathbb{E}\left[\mathbf{x}_{it}\underbrace{(\eta_i + \epsilon_{it})}_{e_{it}}\right]$.
    1. $\mathbf{x}_{it}$ puede estar o no correlacionado con η_i.
    2. $\mathbf{x}_{it}$ puede ser endógena, predeterminada o estrictamente exógena con respecto a ϵ_{it}.

Modelo de primeras diferencias - IV

- Considere $\mathbb{E}[x_{it}\eta_i] \neq 0$ y $\mathbb{E}[x_{is}\epsilon_{it}] = 0 \;\; \forall s \leq t$. **¿En qué caso estamos?**
- Modelo en primeras diferencias

$$\Delta y_{it} = \alpha \Delta y_{it-1} + \Delta \mathbf{x}_{it}' \boldsymbol{\beta} + \Delta \epsilon_{it},$$

- Dado que y_{it-1} es predeterminada, podemos instrumentar Δy_{it-1} con y_{it-2}
- Dado que $\mathbf{x}_{it}$ es predeterminado, podemos instrumentar $\Delta \mathbf{x}_{it}$ con $\mathbf{x}_{it-1}$.

Sección

Arellano-Bond

- Con y_{it-1} y $\mathbf{x}_{it}$ predeterminadas, los instrumentos propuestos son y_{it-2} y $\mathbf{x}_{it-1}$ respectivamente.
- Sin embargo, dependiendo del número de periodos, **hay muchos otros posibles instrumentos**.
- El número de instrumentos posibles puede ser muy "grande".
- **Pregunta:** ¿Cuál de los posibles instrumentos debemos usar para cada variable endógena?
- Arellano and Bond (1991) muestran que usar todos los instrumentos disponibles resulta en un estimador **más eficiente**. **Intuición:** se "explota" toda la información disponible.

Arellano-Bond y GMM

1. Para cada periodo hay un número diferente de instrumentos (el número de rezagos cambia).
2. Al incluir más rezagos como instrumentos podemos perder más observaciones (e.g. la primera observación para cada unidad sólo tiene un rezago posible).
3. Los instrumentos para un periodo pueden no ser exógenos para otros periodos. Por ejemplo y_{i1}, y_{i2} y y_{i3} son posibles instrumentos para Δy_{i4}, pero no para Δy_{i3} o Δy_{i2}.

- Implicaciones:
    1. La matriz de instrumentos no toma la misma forma que trabajamos en el estimador de MC2E.
    2. Es más conveniente usar un método alternativo de estimación: **Método Generalizado de Momentos** (GMM).

Sección

Efectos fijos - modelo *TWFE*

- La posibilidad de controlar por efectos fijos es más general de lo que hemos trabajado. Por ejemplo, considere el modelo:

$$y_{it} = \mathbf{x}'_{it}\beta + e_{it}, \tag{7}$$

- Pero ahora suponga que el error estructural toma la forma:

$$e_{it} = \phi_t + \eta_i + \epsilon_{it}. \tag{8}$$

- En palabras, el error estructural e_{it} ahora contiene tres partes:
  1. ϕ_t: Un componente que es específico al periodo, pero **constante entre observaciones** i.
  2. η_i: un componente que es específico a la unidad de observación i, pero que es **constante en el tiempo**
  3. ϵ_{it}: Un componente aleatorio con variación tanto en i como en t.

Efectos fijos - modelo *TWFE*

- El modelo a estimar es:

$$y_{it} = \mathbf{x}'_{it}\boldsymbol{\beta} + \phi_t + \eta_i + \epsilon_{it} \tag{9}$$

- El modelo lo podemos estimar usando las mismas técnicas que hemos trabajado (e.g. diferentes transformaciones).
- Para el efecto fijo de tiempo, usualmente lo trabajamos incluyendo **dummies específicas a cada periodo** (e.g. año, mes, día,...).
- Una vez eliminamos (controlamos por) los efectos fijos, la consistencia del estimador del vector de parámetros β va a depender $\mathbb{E}[\mathbf{x}_{it}\epsilon_{it}]$.

Ejemplo: Ashenfelter and Krueger (1994)

Ashenfelter and Krueger (1994) Estimates of the Economic Return to Schooling from a New Sample of Twins. *American Economic Review.*

Pregunta de investigación: cuantificar los retornos a la educación.

Problemas:

1. Endogeneidad por variable omitida (e.g. habilidad, características familiares)
2. Error de medición (e.g. educación es auto-reportada)

Estrategia de identificación:

1. Comparar gemelos idénticos (monocigóticos) con diferentes niveles de educación $\implies$ controlar por características genéticas y del entorno familiar.
2. Instrumentar los años de educación auto-reportados de un gemelo i con los años de educación que el gemelo $-i$ dice que tiene su hermano.

Ejemplo: Ashenfelter and Krueger (1994)

- Modelo para cada gemelo:

$$
\begin{aligned}
y_{i1} &= \mathbf{w}_i'\gamma + \mathbf{x}_{i1}'\beta + \eta_i + \epsilon_{i1} \\
y_{i2} &= \mathbf{w}_i'\gamma + \mathbf{x}_{i2}'\beta + \eta_i + \epsilon_{i2}
\end{aligned}
\tag{10}
$$

1. y_{ij}: logaritmo del salario para cada gemelo $j \in \{1, 2\}$.
2. $\mathbf{w}_i$: variables **observadas** con variación a nivel familia, pero constante entre los gemelos (e.g. raza, edad).
3. $\mathbf{x}_{ij}$: variables con variación a nivel de cada gemelo (e.g. **educación**, experiencia laboral, estado civil).
4. η_i: variables **no observadas** con variación a nivel familia, pero constante entre los gemelos (e.g. habilidad natural/genética, características familiares).

$$
\begin{aligned}
y_{i1} - y_{i2} &= (\mathbf{x}_{i1}' - \mathbf{x}_{i2}')\beta + \epsilon_{i1} - \epsilon_{i2} \\
y_i^* &= \mathbf{x}_i'^*\beta + \epsilon_i^*.
\end{aligned}
\tag{11}
$$

Ejemplo Ashenfelter and Krueger (1994)

TABLE 3—ORDINARY LEAST-SQUARES (OLS), GENERALIZED LEAST-SQUARES (GLS), INSTRUMENTAL-VARIABLES (IV), AND FIXED-EFFECTS ESTIMATES OF LOG WAGE EQUATIONS FOR IDENTICAL TWINS[a]

Variable	OLS (i)	GLS (ii)	GLS (iii)	IV[a] (iv)	First difference (v)	First difference by IV (vi)
Own education	0.084 (0.014)	0.087 (0.015)	0.088 (0.015)	0.116 (0.030)	0.092 (0.024)	0.167 (0.043)
Sibling's education	—	—	−0.007 (0.015)	−0.037 (0.029)	—	—
Age	0.088 (0.019)	0.090 (0.023)	0.090 (0.023)	0.088 (0.019)	—	—
Age squared (÷100)	−0.087 (0.023)	−0.089 (0.028)	−0.090 (0.029)	−0.087 (0.024)	—	—
Male	0.204 (0.063)	0.204 (0.077)	0.206 (0.077)	0.206 (0.064)	—	—
White	−0.410 (0.127)	−0.417 (0.143)	−0.424 (0.144)	−0.428 (0.128)	—	—
Sample size:	298	298	298	298	149	149
R^2:	0.260	0.219	0.219	—	0.092	—

Notes: Each equation also includes an intercept term. Numbers in parentheses are estimated standard errors.

[a]Own education and sibling's education are instrumented for using each sibling's report of the other sibling's education as instruments.

Sección

1. Panel dinámico
2. Primeras diferencias - IV (Anderson-Hsiao)
3. Arellano-Bond
4. Efectos fijos - *TWFE*
5. Ashenfelter and Krueger (1994)
6. Errores estándar *cluster-robust*

Errores estándar - estructura de grupo

- En los casos en los que tenemos datos con una **estructura de grupo**, es muy factible que tengamos correlación serial por construcción.
  - Por ejemplo, los estudiantes dentro de un mismo colegio van a tener resultados que están correlacionados porque comparten un mismo ambiente de estudio (profesores, instalaciones, etc).
- En **modelos panel** donde observamos las mismas unidades en diferentes momentos del tiempo, es muy probable que el error estructural esté correlacionado en el tiempo.
- Si no estimamos bien los errores estándar de los parámetros de interés, no vamos a poder hacer **inferencia estadística válida**.

Errores estándar *cluster-robust*

- La solución más común es calcular los errores estándar que sean robustos a este tipo de correlación serial de grupo (*cluster-robust*).
- Suponga que tenemos $g = 1, ..., G$ grupos (clusters) que podemos identificar en los datos.

$$y_{ig} = \mathbf{x}'_{ig}\beta + \epsilon_{ig}. \tag{12}$$

- Asumimos que $\mathbb{E}[\epsilon_{ig}\epsilon_{jg'}|\mathbf{x}_{ig}, \mathbf{x}_{jg'}] = 0$ a menos que $g = g'$.
- **Intuitivamente**, vamos a permitir que exista correlación serial entre miembros de un mismo grupo o *cluster* (*intraclass correlation*), pero independencia entre los grupos o *clusters*.
- Si asumimos de forma incorrecta que no hay correlación serial dentro de los grupos podemos subestimar los errores estándar $\implies$ intervalos de confianza y p-valores más pequeños.

Errores estándar *cluster-robust*

- Juntando las observaciones para cada *cluster*:

$$\boldsymbol{y_g} = \boldsymbol{X_g}\beta + \boldsymbol{\epsilon_g}, \tag{13}$$

  donde $\boldsymbol{y_g}$ y $\boldsymbol{\epsilon_g}$ son vectores de $N_g \times 1$; $\boldsymbol{X_g}$ es una matriz de $N_g \times K$; y hay N_g observaciones por cada *cluster*.

- Juntando las observaciones para los G *clusters*:

$$\mathbf{y} = \mathbf{X}\beta + \boldsymbol{\epsilon}, \tag{14}$$

- El estimador MCO es:

$$\begin{aligned}\hat{\beta} &= (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y} \\ &= \Big(\sum_g \boldsymbol{X_g'X_g}\Big)^{-1}\Big(\sum_g \boldsymbol{X_g'y_g}\Big)\end{aligned} \tag{15}$$

Errores estándar *cluster-robust*

- La varianza condicional de $\hat{\beta}$ es:

$$Var[\hat{\beta}|\mathbf{X}] = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{B}(\mathbf{X}'\mathbf{X})^{-1}, \tag{16}$$

donde

$$\mathbf{B} = \mathbf{X}' Var[\epsilon|\mathbf{X}]\mathbf{X}. \tag{17}$$

- Por el supuesto de independencia entre clusters, el término **B** es:

$$\begin{aligned}\mathbf{B} &= \sum_g \boldsymbol{X_g'}\,\mathbb{E}[\boldsymbol{\epsilon_g \epsilon_g'}|\boldsymbol{X_g}]\boldsymbol{X_g} \\ &= \sum_{g=1}^{G}\sum_{i=1}^{N_g}\sum_{j=1}^{N_g} \mathbf{x}_{ig}\mathbf{x}'_{jg}\omega_{ig,jg},\end{aligned} \tag{18}$$

donde $\omega_{ig,jg} = \mathbb{E}[\epsilon_{ig}\epsilon_{jg}|\boldsymbol{X_g}]$ es la covarianza entre las observaciones *ig* y *jg*

Errores estándar *cluster-robust* - algunas consideraciones

1. Como una **práctica informal**, se suele requerir al menos 42 *clusters* para usar este método. Si no hay suficientes *clusters*, se necesita aplicar otros métodos (e.g. *bootstrap*).
2. Es posible definir *clusters* en diferentes dimensiones: espacial, temporal, etc.
3. El supuesto de no correlación serial entre *clusters* es importante. En la práctica, es el investigador el que define a que nivel definimos los *clusters*.