

UNIVERSIDAD TORCUATO DI TELLA
MAESTRÍA EN ECONOMÍA Y ECONOMETRÍA
2022

Datos de Panel
Problem Set 3
Modelos de Datos de Panel Dinámicos

1. Considere la base de datos *mod.abdata.dta* que fue utilizada por Arellano y Bond en su famoso paper de 1991. Se trata de un panel de 140 empresas británicas encuestadas anualmente entre 1976 y 1984. El panel original no es balanceado, pero la versión para este ejercicio se trata de un panel balanceado de empresas con observaciones para exactamente 6 años entre 1977 y 1982. La variable que identifica la empresa es *id* y la variable que identifica el tiempo es *year*. La variable *n* es el empleo de la empresa. Luego, considere un modelo muy simplificado del siguiente tipo:

$$\ln n_{it} = \rho \ln n_{it-1} + \varepsilon_{it}$$

$$\varepsilon_{it} = c_i + \nu_{it}$$

$$E[c_i] = E[\nu_{it}] = E[c_i \nu_{it}] = 0$$

donde n_{it} es el empleo de la empresa i en el año t .

- a) Estime el modelo por OLS. ¿Qué sesgo esperarías encontrar y por qué?

Solution: Un problema inmediato que encontramos es que $\ln n_{it-1}$ está correlacionado con los efectos fijos que se encuentran en el término del error, lo cual da lugar al sesgo de paneles dinámicos. En general, bajo muchos supuestos, OLS sobrestima el valor real del parámetro ρ (ver Hsiao, p. 85).

- b) Estime el modelo usando efectos fijos (FE). ¿Permite la transformación *within* eliminar el sesgo de paneles dinámicos?

Solution: Ahora, el problema se encuentra en que $\ln \tilde{n}_{it-1} = \ln n_{it-1} - \ln \bar{n}_{i-1}$ está correlacionado con $\tilde{\nu}_{it} = \nu_{it} - \bar{\nu}_{i-1}$ aún cuando ν_{it} no tiene correlación serial. En particular, el término $\ln n_{it-1}$ correlaciona negativamente con $-(1/T - 1)\nu_{it-1}$ que se encuentra dentro de $\bar{\nu}_{i-1}$, mientras que, simétricamente, $-(1/T - 1)\ln n_{it}$ y ν_{it} también se encuentran correlacionados negativamente. Adicionalmente, hay otros pares de términos que correlacionan, pero su impacto es de segundo orden. Por lo tanto, notemos que en este caso el sesgo no surge por los c_i sino por el hecho de que, al ser dinámico el panel, estamos poniendo los ν_{it} en varios lados distintos, y eso es lo que genera el sesgo. Por último, cabe mencionar que Nickell mostró que este sesgo es siempre negativo si $\rho > 0$.

- c) Considere una transformación de diferencias finitas de primer orden del modelo. ¿Continúa siendo la variable dependiente rezagada potencialmente endógena?

Solution: A pesar de que los efectos fijos se van, la variable dependiente rezagada aún es potencialmente endógena ya que el término $\ln n_{it-1}$ en $\Delta \ln n_{it-1} = \ln n_{it-1} - \ln n_{it-2}$ está correlacionado con ν_{it-1} en $\Delta \nu_{it} = \nu_{it} - \nu_{it-1}$.

- d) Implemente el estimador de Anderson-Hsiao a partir del comando *ivregress* en Stata.

Solution: En este contexto, $\ln n_{it-2}$ surge como un instrumento candidato natural para $\Delta \ln n_{it-1}$ ya que matemáticamente están relacionados, pero $\ln n_{it-2}$ no está relacionado con el término de error $\Delta \nu_{it}$ siempre y cuando los ν_{it} no presenten correlación serial. Esto nos conduce al estimador propuesto por Anderson y Hsiao (1982). De esta forma, esta es la primera estimación consistente del modelo dados nuestros supuestos. Ahora bien, recordemos que 2SLS es eficiente bajo el supuesto de errores i.i.d. esféricos. Sin embargo, luego de tomar diferencias de primer orden, los errores $\Delta \nu_{it}$ pueden estar lejos de ser independientes. $\Delta \nu_{it}$ puede estar correlacionado con $\Delta \nu_{it-1}$ dado que comparten el término ν_{it-1} .

- e) Ahora, obtenga la estimación GMM de ρ utilizando todos los instrumentos posibles en niveles para el modelo en primeras diferencias. Para ello utilice el comando `xtabond2`.

Solution: El estimador GMM de Arellano-Bond (1991) aborda directamente el problema del inciso previo, modelando la estructura del error de forma más realista, lo que hace que sea asintóticamente más preciso y se comporte mejor en la práctica. Adicionalmente, al estar utilizando más condiciones de momentos, esto nos provee de mayor información, lo que nos permite obtener un estimador más eficiente que el del inciso previo.

- f) Obtenga la estimación de GMM de ρ utilizando todos los instrumentos posibles en niveles para el modelo en primeras diferencias e Δy_{it-1} como instrumento para el modelo en niveles.

Solution: Aquí lo que estamos haciendo es implementar el estimador GMM de Blundell-Bond (1998).

- g) Repita las estimaciones de los incisos e) y f) incluyendo efectos fijos de tiempo.

Solution: En la estimaciones que vimos hasta aquí, se rechaza la hipótesis nula del test de Sargan, el cual sabemos que se puede pensar como un test de especificación del modelo. Esto implica que el *Data Generating Process* no es el modelo simple que vimos hasta aquí. Por lo tanto, la idea al incluir efectos fijos de tiempo es ir “mejorando” el modelo de a poco.

2. En este ejercicio se ilustrará el hecho de que los estimadores de Arellano-Bond y de Blundell-Bond pueden extenderse en forma directa a modelos que incluyan regresores estrictamente exógenos y regresores secuencialmente exógenos.

En su paper original, Arellano y Bond modelaron el empleo de las empresas (n) utilizando un modelo de ajuste parcial para reflejar los costos de contratación y despido, incluyendo dos rezagos de la variable empleo. Otras variables incluidas fueron el nivel salarial actual y el rezagado (w), el stock de capital actual, rezagado

una y dos veces (k) y la producción agregada actual, rezagada una y dos veces en el sector de la empresa (ys). Todas las variables se expresan en logaritmos. También se incluye un conjunto de variables *dummy* de tiempo.

- a) Estime el modelo por OLS. Compute los errores estándar robustos a heterocedasticidad y correlación serial.
- b) Estime el modelo por FE. Compute los errores estándar robustos a heterocedasticidad y correlación serial.

Solution: Sabemos que bajo ciertos supuestos, en el modelo de regresión la variable dependiente rezagada está correlacionada positivamente con el error, lo que hace que la estimación por OLS de su coeficiente esté sesgada hacia arriba. Por otra parte, en la estimación de efectos fijos, se subestima el verdadero valor del coeficiente que acompaña al primer rezago de n debido al signo negativo en el período $t - 1$ en el error transformado. Por lo tanto, dadas las direcciones opuestas del sesgo presente en estas estimaciones, una estimación consistente para el verdadero parámetro debería estar entre estos valores.

- c) Implemente el estimador de Anderson-Hsiao usando n_{it-2} como instrumento.

Solution: Aunque esta estimación debería ser consistente, los resultados obtenidos no son los esperados. El coeficiente del rezago de n está fuera de los límites de sus correspondientes estimaciones de OLS y FE, y, es mucho mayor que la unidad, una condición de estabilidad. Además, vemos que la estimación del error estándar es muy alta, y el coeficiente no es estadísticamente significativo a los niveles convencionales.

- d) Estime la ecuación de empleo usando el estimador de Arellano-Bond. Asuma que la única endogeneidad presente es en el rezago de la variable dependiente.

Solution: Notamos que el coeficiente de la variable dependiente rezagas se encuentra ahora dentro del rango de estabilidad. Sin embargo, en ambas estimaciones encontramos que no es estadísticamente significativo.

- e) Ahora, considere como hicieron Blundell y Bond (1998) que los salarios y el stock de capital no deben tomarse como estrictamente exógenos en este contexto (como se hizo en los modelos anteriores). Reestime el modelo usando el estimador de A-B y considerando a los salarios y el stock de capital como regresores secuencialmente exógenos.

Solution: Ahora, los resultados de la estimación en un paso y en dos pasos parecen razonables.

- f) Adicionalmente, Blundell y Bond (1998) eliminan de su modelo los rezagos más largos (de dos períodos) del empleo y el capital, y prescinden del nivel de producto agregado sectorial. Considerando esta cuestión, compute el estimador de Blundell-Bond.

Solution: El resultado de la estimación está dentro de los valores esperados.

3. Cuando hay muchos instrumentos, surgen dos problemas principales:

- Sobreestimación (*overfitting*) de la variable endógena
- Mala estimación de la matriz de pesos W

En estos casos, se proponen las siguientes soluciones:

- a) Probar diferentes especificaciones de IV recortando el número de rezagos en la matriz de instrumentos \mathbf{Z} .
- b) Colapsar/combinar instrumentos. Se modifica la matriz de instrumentos para el individuo i :

$$\mathbf{Z}_i = \begin{bmatrix} y_{i1} & 0 & 0 & \cdots \\ y_{i1} & y_{i2} & 0 & \cdots \\ y_{i3} & y_{i2} & y_{i1} & \cdots \\ \vdots & \vdots & \vdots & \ddots \end{bmatrix}$$

Si el modelo funciona debería dar resultados similares con distintos instrumentos.

Retome el ejercicio 2.e) para ver una aplicación de esta cuestión. Estime el modelo de empleo restringiendo el máximo rezago a 3 y 4 períodos. Por último, estime el modelo colapsando instrumentos. Analice si los resultados obtenidos son robustos.

4. Considere nuevamente el modelo del primer ejercicio. Obtenga el estimador LSDVC propuesto por Kiviet (1995) a partir del comando *xtlsdvc*. Luego, estime la matriz de varianzas y covarianzas de los coeficientes de Kiviet siguiendo el procedimiento explicado en clase.