

ECONOMETRÍA

APLICADA UTILIZANDO R.

PAPIME PE302513 LIBRO ELECTRÓNICO Y COMPLEMENTOS DIDÁCTICOS EN MEDIOS COMPUTACIONALES, PARA EL FORTALECIMIENTO DE LA ENSEÑANZA DE LA ECONOMETRÍA

Capítulo 14.

Modelos Panel: Aplicaciones en software R

Miguel Ángel Mendoza González



Introducción

- ❖ Los modelos panel normalmente se utilizan cuando el fenómeno económico, financiero, social, etc. que se está analizando tiene un componente de desagregación, corte transversal o sección cruzada y otro de series de tiempo (Cameron, C. y Trivedi, P., 2005; Frees, E., 2004; Greene, W., 1998; Hsiao, C., 2003; Wooldridge, J., 2002).
- ❖ Los modelos panel clásicos más usados son los de efecto común, efecto individual fijo y el de efecto individual aleatorio.



¿Cómo se elaboran?

Para sus aplicaciones es importante elaborar pruebas de hipótesis entre los supuestos de efectos comunes versus individuales, y entre los efectos individuales fijos o aleatorios. Así los supuestos econométricos básicos de varianza homoscedastica, normalidad, no-autocorrelación serial y/o contemporánea.



MODELO PANEL ESTÁTICO GENERAL

El modelo panel estático general tiene dos fuentes de heterogeneidad entre los elementos i de sección cruzada, por las constantes individuales μ_i y los parámetros de relación individual β_i entre la variable endógena $y_{i,t}$ y las exógenas $X_{i,t}$.

$$y_{i,t} = \mu_i + \beta_i X_{i,t} + \varepsilon_{i,t} \quad (1)$$

donde $i = 1, 2, 3, \dots, n$ es el identificador de los elementos de la sección cruzada y $t = 1, 2, 3, \dots, T$, el de tiempo.

- Con la especificación general se requieren los μ_i y los β_i , que son $i \times i = i^2$ parámetros a estimar al mismo tiempo. Aunque desde el punto de vista analítico es interesante identificar de manera individual las constantes y los parámetros de relación, la complicación aparece con la derivación del método de estimación y en la parte computacional. Por ello, es muy importante analizar modelos panel que **simplifiquen la cantidad de parámetros a estimar** y que sean analíticamente interesante.
- De la especificación general, se pueden aplicar restricciones a los parámetros que conlleva dos grupos de modelos. En la especificación general se establece que cada elemento de la sección cruzada de la variable endógena $y_{i,t}$, responde diferente a las variables exógenas $X_{i,t}$.

- La heterogeneidad de los efectos se identifican por β_i , pero al aplicar la restricción $\beta_1 = \dots = \beta_i = \beta$ se supone que cada elemento de la sección cruzada responde de la misma manera a la variables exógenas; esto es lo que se conoce como respuesta común. El modelo que resulta al aplicar las restricciones, tiene como única fuente de heterogeneidad a las constantes individuales identificadas por las μ_i ; como en la ecuación 2.

$$y_{i,t} = \mu_i + \beta X_{i,t} + \varepsilon_{i,t} \quad (2)$$

- El segundo grupo de modelos se obtiene al aplicar el siguiente nivel de restricciones a las constantes individuales $\mu_1 = \dots = \mu_i = \mu$; esto es lo que se conoce como efecto común. El nuevo modelo panel que resulta, es una especificación que supone homogeneidad en los elementos de la sección cruzada por condiciones iguales (μ) y respuesta igual (β) a las variables exógenas; ver ecuación 3.

$$y_{i,t} = \mu + \beta X_{i,t} + \varepsilon_{i,t} \quad (3)$$

¿Cuáles son lo supuestos económicos?

- En el modelo estático general como en las dos versiones con las restricciones de los parámetros, se requiere analizar los supuestos clásicos sobre los errores o innovaciones $\varepsilon_{i,t}$ de los modelos. Por default, se debe cumplir que la media de los errores por corte transversal y serie de tiempo es igual a cero, $E[\varepsilon_{i,t}] = 0$; que la varianza del modelo, dado las variables exógenas, sea constante para cada sección cruzada, pero puede ser diferente entre ellas, $E[\varepsilon_{i,t}^2 | x_{i,t}] = \sigma_i^2$; y, que no exista correlación serial ni contemporánea $E[\varepsilon_{i,t}\varepsilon_{j,s}] = 0$, con $t \neq s$ e $i \neq j$; esto supuestos se puede resumir en una matriz del siguiente tipo:

$$E[\varepsilon_i \varepsilon_i'] = \Omega = \begin{bmatrix} \varepsilon_1 \varepsilon_1' & \cdots & \varepsilon_1 \varepsilon_N' \\ \vdots & \ddots & \vdots \\ \varepsilon_N \varepsilon_1' & \cdots & \varepsilon_N \varepsilon_N' \end{bmatrix} \text{ donde } \varepsilon_i = [\varepsilon_{i,1} \ \varepsilon_{i,2} \ \dots \ \varepsilon_{i,t}]$$

- Si la matriz de varianzas y covarianzas cumple con los supuestos econométricos descritos anteriormente, entonces se puede escribir como $\Omega = \sigma^2 I_N \otimes I_T$, los estimadores son insesgados y eficientes, y por tanto el modelo se estima con mínimos cuadrados ordinarios (Greene, W., 1998).
- El supuesto de consistencia de los parámetros es relativo y depende de la comparación entre los modelos analizados.

El modelo de panel con efectos comunes (pooled OLS estimator)

- La especificación tipo pool impone restricciones a los parámetros individuales, al establecer que una constante común ($\mu_1 = \dots = \mu_i = \mu$) y efecto común ($\beta_1 = \dots = \beta_i = \beta$) con respecto a las variables exógenas, como la ecuación 3. El estimador *pooled OLS* se obtiene al apilar (stacking) los datos sobre i y t con NT observaciones y aplicando OLS. Si el modelo está correctamente especificado y las variables exógenas no están correlacionados con los errores, entonces se puede estimar consistentemente.

- En otras palabras, si se cumple $Cov = [\varepsilon_{it}, x_{it}] = 0$ entonces $N \rightarrow \infty$ o $T \rightarrow \infty$ son suficientes para la consistencia. El estimador *pooled OLS* es inconsistente si el modelo apropiado es el de efectos fijos, debido a que las constantes individuales que no se incluyeron en el modelo *pool* están correlacionadas con las variables exógenas.

El modelo de panel con efectos fijos en constante (estimador within)

- La restricción que se elimina con el modelo de efectos fijos es que existe una constante individual para cada elemento de la sección cruzada (μ_i). Desde el punto de vista de los estimadores, el estimador *within* a diferencia del *pooled OLS*, explora las características de los datos panel y mide la asociación entre las desviaciones entre los elementos de las variables exógenas desde sus valores promedio en el tiempo y las desviaciones entre los elementos de la variable endógena desde su valor promedio en el tiempo. Los pasos para la estimación consiste en:

- En primer lugar se comienza con el modelo de efectos individuales en constante, en el cual se prueba el caso de $\mu_i = \mu$
- Entonces se toma el valor promedio en el tiempo

$$\bar{y}_i = \mu + \beta \bar{X}_i + \bar{\varepsilon}_i$$

- Al modelo de efectos individuales en constante se le resta el modelo promedio en el tiempo, con el cual se obtiene el estimador *within*

$$y_{i,t} - \bar{y}_i = \mu_i + \beta [X_{i,t} - \bar{X}_i] + \varepsilon_{i,t} - \bar{\varepsilon}_i$$

Modelo Panel con efectos aleatorios en constante (*feasible GLS estimator*)

- En el modelo de efectos aleatorios, se asume que la constante individual tiene una distribución con media μ y una desviación estándar σ_μ , que junto con los errores o innovaciones $\varepsilon_{it} \sim [0, \sigma_\varepsilon^2]$ configuran las dos partes aleatorias o probabilísticas del modelo panel con efectos aleatorios.
- Al estimador que se utiliza se le conoce como el estimador de mínimos cuadrados generalizado factible (*feasible GLS estimator*), que puede calcularse al estimar con mínimos cuadrados ordinario el siguiente modelo transformado

$$y_{i,t} - \hat{\lambda} \bar{y}_i = (1 - \hat{\lambda}) \mu_i + \beta [X_{i,t} - \hat{\lambda} \bar{X}_i] + v_{i,t}$$

Donde $v_{i,t} = (1 - \hat{\lambda}) \mu_i + (\varepsilon_{i,t} - \bar{\varepsilon}_i)$ es iid asintóticamente y $\hat{\lambda} = 1 - \frac{\sigma_\varepsilon}{\sqrt{\sigma_\varepsilon^2 + T \sigma_\mu^2}}$. Notar que $\hat{\lambda} = 0$

corresponde a pooled OLS, $\hat{\lambda} = 1$ corresponde al estimador *within* y cuando $T \rightarrow \infty$ entonces $\hat{\lambda} \rightarrow 1$. El estimador para β es básicamente en dos etapas.

- El estimador de efectos aleatorios es totalmente eficiente dentro del supuesto del mismo estimador, aunque la eficiencia la gana en realidad al compararse con el estimador *pooled* y es inconsistente si el modelo de efectos fijos es el correcto.

ELECCIÓN DE MODELOS ALTERNATIVOS

El procedimiento de elección de la especificación de la constante del modelo panel con los estimadores *pooled*, efectos fijos (*within*) o efectos aleatorios (*feasible GLS estimator*), es el siguiente.

Modelo de efectos individuales (IE) versus el modelo Pool

- En sentido estricto, se tiene que comparar los modelos de efectos individuales (fijos y aleatorias) con respecto al modelo pool. Sin embargo, es tradición comparar en esta primera fase el modelo de efectos fijos con el modelo pool, para comparar la eficiencia del primero.

- *Prueba pooling*

Para ello, se utiliza una prueba de restricción de parámetros entre los dos modelos y se analizan las hipótesis:

$$H_n: \forall \mu_i = 0$$

$$H_a: \mu_1 \neq 0, \dots, \mu_i \neq 0$$

Para analizar las hipótesis, se utiliza un estadístico $\lambda^2(k)$ con lo k grados de libertad definidos por la cantidad de efectos individuales; a esta prueba se le conoce como *pooling*.

Modelo de efectos aleatorios (EA) versus el modelo de efectos fijos (EF)

- En el caso de que el modelo de *efectos fijos* sea eficiente en comparación con el modelo *pooled*, entonces se puede analizar si el modelo de efectos aleatorios es eficiente en comparación con el modelo de efectos fijos. Para probar la consistencia del modelo panel con efectos aleatorios, se utiliza la prueba de *Hausman*.
- *Prueba de Hausman*

Las hipótesis que se utiliza para analizar la consistencia se resume en:

H_n : Estimador EA es consistente con respecto al estimado EF

H_a : Estimador EF es consistente con respecto al estimador EA

El estadístico para probar se define como:

$$\lambda^2(k) : H = [\beta_{EF} - \beta_{EA}]^T [Cov(\beta_{EF}) - Cov(\beta_{EA})]^{-1} [\beta_{EF} - \beta_{EA}]$$

Donde β es el vector de coeficientes compuesto con los parámetros $[\beta \ \gamma]$ del modelo, Cov es la matriz de varianza-covarianza y k es el número de coeficientes