Paneles heterogéneos

Gabriel Montes-Rojas

Paneles heterogéneos

- En los modelos previos, toda la heterogeneidad se modelaba con un intercepto que era específico de cada individuo, y autocorrelación y/o heteroscedasticidad de los errores.
- Paneles heterogéneos trata de modelos con pendientes heterogéneas (slope heterogeneity).
- En general se necesita que $T \to \infty$ para poder estimar cada ecuación por separado.

Paneles heterogéneos

Dos formas de modelar la heterogeneidad, coeficientes fijos y coeficientes aleatorios.

- Podriamos considerar que cada individuo tiene una estructura de coeficientes única, β_i.

  $$y_{it} = x_{it}\beta_i + u_{it}, i = 1, 2, ..., N; t = 1, 2, ..., T.$$

  Esto se puede modelar usando interacciones de las Xs con las dummies por individuo o estimando un modelo diferente para cada i.
  También podríamos tener que algunos coeficientes son específicos de cada individuo i y otros son comunes:

  $$y_{it} = x_{it}\beta_i + z_{it}\gamma + u_{it}, i = 1, 2, ..., N; t = 1, 2, ..., T.$$

  Si todos los coeficientes fueran específicos para cada individuo, el modelo es equivalente a MCO para cada i.
- Otra forma más simple de modelar heterogeneidad es con un **modelo de coeficientes aleatorios**. En este caso, los coeficientes mismos son variables aleatorias, que se asumen estacionarios (misma media y varianza).

SUR

Consideremos las siguientes M regresiones supuestamente no relacionadas ("seemingly unrelated regressions, SUR")

$$\boldsymbol{y}_i = \boldsymbol{X}_i \beta_i + \boldsymbol{u}_i, \quad i = 1, 2, \ldots, M,$$

donde $\boldsymbol{y}_i = (y_{i1}, y_{i1}, \ldots, y_{iT})'$ y $\boldsymbol{u}_i = (u_{i1}, u_{i1}, \ldots, u_{iT})'$ son vectores $T \times 1$, $\boldsymbol{X}_i$ es una matriz de $T \times K_i$ de K_i variables explicativas, y $\boldsymbol{\beta}_i$ es un vector $K_i \times 1$.
Supuesto 1: El vector $MT \times 1$ de errores tiene media condicional 0: $E(\boldsymbol{u}|\boldsymbol{X}_1, \boldsymbol{X}_2, \ldots, \boldsymbol{X}_M) = 0$.
Supuesto 2: Los errores no estan correlacionados entre observaciones: $E(\boldsymbol{u}_i \boldsymbol{u}_i'|\boldsymbol{X}_1, \boldsymbol{X}_2, \ldots, \boldsymbol{X}_M) = \sigma_{ii} \boldsymbol{I}_T$.
Supuesto 3: Distintos modelos pueden ser considerados para las correlaciones entre errores.

1. $E(\boldsymbol{u}_i \boldsymbol{u}_j') = 0, \ i \neq j$.
2. $E(\boldsymbol{u}_i \boldsymbol{u}_j') = \sigma_{ij} \boldsymbol{I}_T \neq 0, \ i \neq j$, pero $\boldsymbol{X}_i = \boldsymbol{X}_j$, para todo i, j.
3. $E(\boldsymbol{u}_i \boldsymbol{u}_j') = \sigma_{ij} \boldsymbol{I}_T \neq 0, \ i \neq j$, pero $\boldsymbol{X}_i \neq \boldsymbol{X}_j$, para al menos un par $i \neq j$.

SUR

El modelo se puede escribir como

$$\begin{pmatrix} \boldsymbol{y}_1 \\ \boldsymbol{y}_2 \\ \vdots \\ \boldsymbol{y}_M \end{pmatrix} = \begin{pmatrix} \boldsymbol{X}_1 & \mathbf{0} & \dots & \mathbf{0} \\ \mathbf{0} & \boldsymbol{X}_2 & \dots & \mathbf{0} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{0} & \mathbf{0} & \dots & \boldsymbol{X}_M \end{pmatrix} \begin{pmatrix} \beta_1 \\ \beta_2 \\ \vdots \\ \beta_M \end{pmatrix} + \begin{pmatrix} \boldsymbol{u}_1 \\ \boldsymbol{u}_2 \\ \vdots \\ \boldsymbol{u}_M \end{pmatrix}$$

$$= \boldsymbol{G}\beta + \boldsymbol{u},$$

donde $\boldsymbol{y}$, $\boldsymbol{G}$, β y $\boldsymbol{u}$ tienen dimensiones $MT \times 1$, $MT \times K$, $K \times 1$ y $MT \times 1$, respectivamente, y $K = \sum_{i=1}^{M} K_i$.

SUR

Los Casos 1 y 2 son equivalentes a estimar una regresión por individuo.
Supongamos el Caso 3, $E(\boldsymbol{u}_i\boldsymbol{u}_j') = \sigma_{ij}\boldsymbol{I}_T \neq 0$, $i \neq j$, pero $\boldsymbol{X}_i \neq \boldsymbol{X}_j$, para al menos un par $i \neq j$.

$$E(\boldsymbol{u}_i\boldsymbol{u}_j'|\boldsymbol{X}_1, \boldsymbol{X}_2, \ldots, \boldsymbol{X}_M) = \boldsymbol{\Omega} = \boldsymbol{\Sigma} \otimes \boldsymbol{I}_T,$$

donde $\boldsymbol{\Sigma} = \{\sigma_{ij}\}$ es una matriz simétrica $M \times M$. En particular,

$$\boldsymbol{\Omega} = \boldsymbol{\Sigma} \otimes \boldsymbol{I}_T = \begin{pmatrix} \sigma_{11}\boldsymbol{I}_T & \sigma_{12}\boldsymbol{I}_T & \ldots & \sigma_{1M}\boldsymbol{I}_T \\ \sigma_{12}\boldsymbol{I}_T & \sigma_{22}\boldsymbol{I}_T & \ldots & \sigma_{2M}\boldsymbol{I}_T \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_{1M}\boldsymbol{I}_T & \sigma_{2M}\boldsymbol{I}_T & \ldots & \sigma_{MM}\boldsymbol{I}_T \end{pmatrix},$$

y también $\boldsymbol{\Omega}^{-1} = \boldsymbol{\Sigma}^{-1} \otimes \boldsymbol{I}_T$.
Entonces podemos plantear el MCG,

$$\beta_{MCG} = \left[\boldsymbol{G}'(\boldsymbol{\Sigma}^{-1} \otimes \boldsymbol{I}_T)\boldsymbol{G}\right]^{-1} \boldsymbol{G}'(\boldsymbol{\Sigma}^{-1} \otimes \boldsymbol{I}_T)\boldsymbol{y},$$

$$Var(\beta_{MCG}) = \left[\boldsymbol{G}'(\boldsymbol{\Sigma}^{-1} \otimes \boldsymbol{I}_T)\boldsymbol{G}\right]^{-1}$$

En la práctica $\hat{\sigma}_{ij} = \frac{\hat{\boldsymbol{u}}_i\hat{\boldsymbol{u}}_j}{T}$, $i, j = 1, 2, \ldots, M$, con los residuos MCO, y se arma $\hat{\boldsymbol{\Sigma}}$.

Paneles heterogéneos

Modelo de Swamy (1970)

$$y_{it} = \boldsymbol{x}_{it}\boldsymbol{\beta}_i + u_{it}, i = 1, 2, ..., N; t = 1, 2, ..., T,$$

$$\boldsymbol{\beta}_i = \boldsymbol{\beta} + \boldsymbol{\eta}_i,\ E(\boldsymbol{\eta}_i) = 0,\ E(\boldsymbol{x}_{it}\boldsymbol{\eta}_i) = 0, u_{it} \sim iid(0, \sigma_i^2)$$

$$E(\boldsymbol{\eta}_i\boldsymbol{\eta}_j') = \begin{cases} \boldsymbol{\Omega}_\eta, & si \quad i = j, \\ \boldsymbol{0}, & si \quad i \neq j. \end{cases}$$

- En este modelo tenemos K variables explicativas y K parámetros $\boldsymbol{\beta}$ a estimar (un vector $K \times 1$).
- η es una variable aleatoria de dimensión $K \times 1$ (la cantidad de variables explicativas).
- En este modelo la aleatoriedad es invariante en el tiempo.
- Nos interesa la media β y la función de varianzas-covarianzas de los parámetros.

Paneles heterogéneos

Modelo de Hsiao (1974,1975)

$$y_{it} = \boldsymbol{x}_{it}\boldsymbol{\beta}_{it} + u_{it}, i = 1, 2, ..., N; t = 1, 2, ..., T$$

$$\boldsymbol{\beta}_{it} = \boldsymbol{\beta} + \boldsymbol{\eta}_i + \boldsymbol{\lambda}_t$$

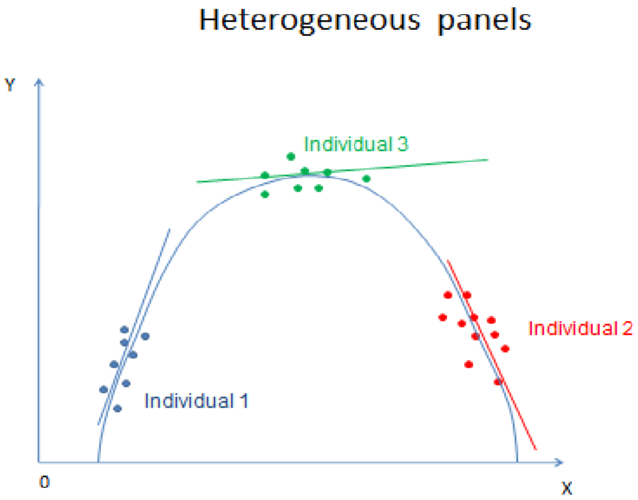
$$E(\boldsymbol{\eta}_i) = 0, E(\boldsymbol{\lambda}_t) = 0, E(\boldsymbol{\eta}_i'\boldsymbol{\lambda}_t) = 0, u_{it} \sim iid(0, \sigma_i^2)$$

$$E(\boldsymbol{x}_{it}\boldsymbol{\eta}_i) = 0, E(\boldsymbol{x}_{it}\boldsymbol{\lambda}_t) = 0$$

$$E(\boldsymbol{\eta}_i\boldsymbol{\eta}_j') = \begin{cases} \boldsymbol{\Omega}_\eta, & si \quad i = j, \\ \boldsymbol{0}, & si \quad i \neq j. \end{cases}$$

$$E(\boldsymbol{\lambda}_t\boldsymbol{\lambda}_h') = \begin{cases} \boldsymbol{\Omega}_\lambda, & si \quad t = h, \\ \boldsymbol{0}, & si \quad t \neq h. \end{cases}$$

Paneles heterogéneos



Consecuencias de obviar heterogeneidad

- Supongamos el siguiente modelo generador de los datos:

$$y_{it} = \mu_i + \beta_i x_{it} + u_{it}, i = 1, 2, ..., N; t = 1, 2, ..., T,$$

$$u_{it} \sim iid(0, \sigma_u^2),$$

$$\beta_i = \beta + \eta_i.$$

Este es un modelo de efectos fijos con heterogeneidad en las pendientes.

- Ahora supongamos que se estima el modelo

$$y_{it} = \alpha_i + \delta_x x_{it} + \delta_z z_{it} + \nu_{it}, i = 1, 2, ..., N; t = 1, 2, ..., T,$$

donde z es un regresor adicional que el investigador cree que es necesario, cuando en realidad no lo es.

- $\nu_{it} \sim iid(0, \sigma_\nu^2)$, distribuido independientemente de $w_{it} = (x_{it}, z_{it})$ para todo i, t. w_i sigue un proceso estacionario en covarianza con $\Omega_i = \begin{bmatrix} \omega_{ixx} & \omega_{ixz} \\ \omega_{izx} & \omega_{izz} \end{bmatrix}$.
- Consideremos el estimador FE

$$plim_{N,T\to\infty}\hat{\delta}_{FE} = \left[plim \frac{1}{NT}\sum_{i=1}^{N} W_i' Q_T W_i\right]^{-1}\left[plim \frac{1}{NT}\sum_{i=1}^{N} W_i' Q_T (X_i\beta_i)\right],$$

donde $Q_T = I_T - \frac{1}{T}\iota_T\iota_T'$ (¿qué transformación hace esta matriz?)

Consecuencias de obviar heterogeneidad

- Si $\beta_i = \beta$ entonces $plim_{N,T\to\infty}\hat{\delta}_{FE} = (\beta', 0')'$.
- Si los paneles son heterogéneos

$$plim \frac{1}{NT}\sum_{i=1}^{N} W_i' Q(X_i\eta_i) = plim \begin{pmatrix} \frac{1}{NT}\sum_{i=1}^{N} X_i' Q(X_i\eta_i) \\ \frac{1}{NT}\sum_{i=1}^{N} Z_i' Q(X_i\eta_i) \end{pmatrix}$$

- Entonces necesitamos que $\frac{1}{NT}\sum_{i=1}^{N}\omega_{ixx}\eta_i \xrightarrow{p} 0$ y $\frac{1}{NT}\sum_{i=1}^{N}\omega_{ixz}\eta_i \xrightarrow{p} 0$. En otras palabras, no tiene que haber correlación sistemática entre los β_i y los **momentos segundos** de los regresores.
- $$plim\hat{\delta}_{x,FE} - \beta = \frac{cov(\omega_{ixx},\eta_i)E(\omega_{izz}) - E(\omega_{ixz})cov(\omega_{ixz},\eta_i)}{E(\omega_{ixx})E(\omega_{izz}) - (E(\omega_{ixz}))^2}$$

  $$plim\hat{\delta}_{z,FE} = \frac{cov(\omega_{ixz},\eta_i)E(\omega_{ixx}) - E(\omega_{ixz})cov(\omega_{ixx},\eta_i)}{E(\omega_{ixx})E(\omega_{izz}) - (E(\omega_{ixz}))^2}$$
- Supongamos el modelo $y_{it} = \alpha_i + \delta_x x_{it} + \delta_z x_{it}^2 + v_{it}$. En este caso en general no podemos obtener que FE estima bien a menos que β_i sea proporcional a x_{it}.

Estimador de Swamy (1970)

Consideremos el modelo de Swamy (1970) expuesto anteriormente.

- En el modelo de Swamy, la matriz de varianzas-covarianzas es

$$\mathbf{\Sigma} = E(\boldsymbol{\nu}\boldsymbol{\nu}') = \begin{bmatrix} \mathbf{\Sigma}_1 & 0 & \dots & 0 \\ 0 & \mathbf{\Sigma}_2 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \mathbf{\Sigma}_N \end{bmatrix}$$

  donde $\nu_{it} = u_{it} + \boldsymbol{x}_{it}\boldsymbol{\eta}_i$ y $\mathbf{\Sigma}_i = \sigma_i^2 I_T + \boldsymbol{X}_i \Omega_\eta \boldsymbol{X}_i'$ para $i = 1, 2, ..., N$.
- Se propone entonces el estimador

$$\hat{\beta}_{SW} = (\boldsymbol{X}'\Sigma\boldsymbol{X})^{-1}(\boldsymbol{X}'\Sigma\boldsymbol{y}) = [\sum_{i=1}^{N} \boldsymbol{X}_i'\Sigma_i^{-1}\boldsymbol{X}_i]^{-1} \sum_{i=1}^{N} \boldsymbol{X}_i'\Sigma_i^{-1}\boldsymbol{y}_i$$

Estimador de Swamy (1970)

- Si $\boldsymbol{\Omega}_\eta$ es no singular,

$$\boldsymbol{\Sigma}_i^{-1} = \frac{\boldsymbol{I}_T}{\sigma_i^2} - \frac{\boldsymbol{X}_i}{\sigma_i^2}\left(\frac{\boldsymbol{X}_i'\boldsymbol{X}_i}{\sigma_i^2} + \boldsymbol{\Omega}_\eta^{-1}\right)^{-1}\frac{\boldsymbol{X}_i'}{\sigma_i^2} = \frac{\boldsymbol{I}_T}{\sigma_i^2} - \frac{\boldsymbol{X}_i\boldsymbol{\Omega}_\eta}{\sigma_i^2}\left(\boldsymbol{I}_K + \frac{\boldsymbol{X}_i'\boldsymbol{X}_i}{\sigma_i^2}\boldsymbol{\Omega}_\eta\right)^{-1}\frac{\boldsymbol{X}_i'}{\sigma_i^2}$$

(para la última igualdad no necesitamos que sea no singular)

Estimador de Swamy (1970)

- El estimador se puede escribir como:

$$\overline{\hat{\beta}}_{SW} = \sum_{i=1}^{N} \boldsymbol{R}_i \hat{\beta}_i$$

donde $\boldsymbol{R}_i = [\sum_{i=1}^{N} (\boldsymbol{\Omega}_\eta + \boldsymbol{\Sigma}_{\hat{\beta}_i})^{-1}]^{-1} (\boldsymbol{\Omega}_\eta + \boldsymbol{\Sigma}_{\hat{\beta}_i})^{-1}$, donde $\hat{\boldsymbol{\beta}}_i = (\boldsymbol{X}_i' \boldsymbol{X}_i)^{-1} \boldsymbol{X}_i' \boldsymbol{y}_i$ y $\boldsymbol{\Sigma}_{\hat{\beta}_i} = \sigma_i^2 (\boldsymbol{X}_i' \boldsymbol{X}_i)^{-1}$ son los estimadores y sus varianzas, respectivamente, para cada i.

- $Var(\overline{\hat{\beta}}_{SW}) = [\sum_{i=1}^{N} (\boldsymbol{\Omega}_\eta + \boldsymbol{\Sigma}_{\hat{\beta}_i})^{-1}]^{-1}$.
- Para estimar la varianza necesitamos un estimador de $\boldsymbol{\Omega}_\eta$. Se propone:

$$\hat{\boldsymbol{\Omega}}_\eta = \frac{1}{N-1} \sum_{i=1}^{N} (\hat{\boldsymbol{\beta}}_i - \hat{\boldsymbol{\beta}}_{MG})(\hat{\boldsymbol{\beta}}_i - \hat{\boldsymbol{\beta}}_{MG})'$$

donde $\hat{\boldsymbol{\beta}}_{MG}$ se define abajo como el estimador de *mean-group*.

Estimador media de grupos (MG estimator)

- Una alternativa es el mean-group (MG)

$$\hat{\beta}_{MG} = \frac{1}{N}\sum_{i=1}^{N}\hat{\beta}_i$$

donde $\hat{\beta}_i = (\boldsymbol{X}_i'\boldsymbol{X}_i)^{-1}\boldsymbol{X}_i'\boldsymbol{y}_i$ son los estimadores OLS de cada i.

- Notar que $\hat{\beta}_i = \boldsymbol{\beta} + \boldsymbol{\eta}_i + \zeta_i$, donde $\zeta_i = (\boldsymbol{X}_i'\boldsymbol{X}_i)^{-1}\boldsymbol{X}_i'\boldsymbol{u}_i$, entonces $\hat{\beta}_{MG} = \boldsymbol{\beta} + \bar{\boldsymbol{\eta}} + \bar{\zeta}$ con $\bar{\boldsymbol{\eta}} = \frac{1}{N}\sum_{i=1}^{N}\boldsymbol{\eta}_i$ y $\bar{\zeta} = \frac{1}{N}\sum_{i=1}^{N}\zeta_i$.
- $var(\hat{\beta}_{MG}) = var(\bar{\boldsymbol{\eta}}) + var(\bar{\zeta}) = \frac{1}{N}\boldsymbol{\Omega}_\eta + \frac{1}{N^2}\sum_{i=1}^{N}\sigma_i^2 E\left[T(\boldsymbol{X}_i'\boldsymbol{X}_i)^{-1}\right]$

Contrastes para homogeneidad vs. heterogeneidad

- Dadas las potenciales consecuencias negativas de obviar la heterogeneidad es importante chequear por la misma.
- Los test de heterogeneidad se pueden interpretar como de *poolability*, o sea, si se pueden "juntar" los modelos en una misma función lineal.
- Supongamos el siguiente modelo generador de los datos:

$$y_{it} = \mu_i + x_{it}\boldsymbol{\beta}_i + u_{it}, i = 1, 2, ..., N; t = 1, 2, ..., T$$

$$u_{it} \sim iid(0, \sigma_i^2)$$

- La hipótesis nula de interés es

$$H_0 : \boldsymbol{\beta}_i = \boldsymbol{\beta}, \forall i$$

  y la alternativa

$$H_A : \boldsymbol{\beta}_i \neq \boldsymbol{\beta}_j, i \neq j$$

- Notar que los contrastes de homogeneidad se refieren a las pendientes $\boldsymbol{\beta}$, y en general no se contrasta por el intercepto.

Contrastes para homogeneidad vs. heterogeneidad

- F-test de $K(N-1)$ restricciones. Este contraste funciona para N fijo y $T \to \infty$ y $\sigma_i^2 = \sigma^2$ (homoscedasticidad):

$$F = \frac{N(T-K-1)}{K(N-1)} \frac{RSSR - USSR}{USSR}$$

  Este test corresponde a correr una regresión de efectos fijos con las interacciones de cada i con cada variable de control $\boldsymbol{X}$, excepto una, más la X sin interactuar, y luego chequear que todas las interacciones son cero. En general tiene un resultado pobre si N es grande y/o la varianza es heteroscedástica.

- **Contraste de Hausman.** Comparar $\hat{\beta}_{MG}$ y $\hat{\beta}_{FE}$. Para poder usar este contraste se necesita:
  (a) Bajo la hipótesis nula, FE es más eficiente. ¿Por qué? Porque estima sólo un parámetro en ves de N.
  (b) Bajo la alternativa $\hat{\beta}_{MG} - \hat{\beta}_{FE}$ debería tender a un vector no cero.

Contrastes para homogeneidad vs. heterogeneidad

- **Contraste de Swamy (1970).**

$$\hat{S} = \sum_{i=1}^{N} (\hat{\beta}_i - \hat{\beta}_{WFE})' \frac{\boldsymbol{X}_i' \boldsymbol{Q}_T \boldsymbol{X}_i}{\hat{\sigma}_i^2} (\hat{\beta}_i - \hat{\beta}_{WFE})$$

donde $\hat{\sigma}_i^2 = \frac{1}{T-K-1}(\boldsymbol{y}_i - \boldsymbol{X}_i\hat{\beta}_{WFE})'\boldsymbol{Q}_T(\boldsymbol{y}_i - \boldsymbol{X}_i\hat{\beta}_{WFE})$ y
$\hat{\beta}_{WFE} = (\sum_{i=1}^{N}(\boldsymbol{X}_i'\boldsymbol{Q}_T\boldsymbol{X}_i)/\hat{\sigma}_i^2)^{-1}\sum_{i=1}^{N}(\boldsymbol{X}_i'\boldsymbol{Q}_T\boldsymbol{y}_i)/\hat{\sigma}_i^2$ donde $\boldsymbol{Q}_T = \boldsymbol{I}_T - \overline{\boldsymbol{J}}_T$.
Cuando N está fijo y $T \to \infty$, bajo la hipótesis nula, $\hat{S} \xrightarrow{d} \chi^2_{K(N-1)}$.
La intuición de este contraste es que nosotros no conocemos el valor de β bajo la hipótesis nula. Sin embargo, si este fuera uno se podría estimar con FE. Entonces podríamos compara todas las estimaciones de β_i con esta referencia. Si esa diferencia no es muy grande, entonces podríamos aceptar la nula.

- **Contraste de Pesaran y Yamagata (2008).**
Este contraste parte del de Swamy, pero estandariza el estadístico $\hat{S}$ usando su distribución chi cuadrado.
Reemplazar $\hat{\sigma}_i^2$ por $\tilde{\sigma}_i^2 = \frac{1}{T-1}(\boldsymbol{y}_i - \boldsymbol{X}_i\hat{\beta}_{FE})'\boldsymbol{Q}_T(\boldsymbol{y}_i - \boldsymbol{X}_i\hat{\beta}_{FE})$ y usar

$$\Delta = \sqrt{N}\left(\frac{N^{-1}S - K}{\sqrt{2K}}\right).$$

Este contraste funciona para $N, T \to \infty$.

STATA

- Ver comando `https://www.stata.com/manuals13/xtxtrc.pdf`
- Modelo de coeficientes aleatorios de Swamy

  ```
  xtrc y x1 x2 x3, beta
  ```
- Ejemplo:

  ```
  webuse invest2, clear
  xtrc invest market stock, beta
  ```