Datos en panel

Gabriel Montes-Rojas

3 tipos de estructuras de datos

- Corte transversal (Cross section) Muestra de individuos, hogares, firmas, países, etc. que se toman en un momento dado del tiempo.

  $\{y_i, x_i\}_{i=1}^{n}$, donde i representa individuos.

  Ej.: EPH de 2013, datos de PBI en 2013 para muchos países.
- Series de tiempo Muestra por varios periodos del mismo individuo, pais, firma, etc.

  $\{y_t, x_t\}_{t=1}^{T}$, donde t es tiempo.

  Ej.: Inflación en la Argentina.
- Datos en panel Combinación de las dos anteriores.

  $\{y_{it}, x_{it}\}_{i=1,t=1}^{n,T}$, donde i representa individuos y t tiempo.

  1. **Cortes transversales independientes**
  2. **Muestra longitudinal**

Cortes transversales independientes

- La estructura de datos en panel que es más frecuente son los cortes transversales independientes.
- Ej: Muchas encuestas de hogares no siguen los mismo hogares a lo largo del tiempo, estos cambian cada vez. La EPH contiene hogares que se renuevan cada vez y otros a los que se los sigue repetidamente.
- Supongamos dos periodos $t = 0, 1$, con n_0 individuos en $t = 0$ y n_1 en $t = 1$ (entonces tenemos un total de $n = n_0 + n_1$). Los datos son $\{y_{i0}, x_{i0}\}_{i=1}^{n_0}$, $\{y_{j1}, x_{j1}\}_{j=1}^{n_1}$

  Supongamos que corremos una regresión simple:

$$y_{ht} = \beta_0 + \beta_1 x_{ht} + u_{ht}, \;\; h = 1, 2, ..., (n_0 + n_1), t = 0, 1$$

¿Qué mide β_1?

Cortes transversales independientes

Supongamos que tanto y como x son mayores en $t = 1$ que en $t = 0$. ¡Entonces con una regresión simple estaríamos capturando el efecto del tiempo como un efecto de x en y!

Para ver esto definamos

$$u_{ht} = a_t + e_{ht}$$

donde a_t mide el efecto del tiempo mientras que e es un verdadero error. Entonces, $E(xu) = E(xa) + E(xe) = E(xa) \neq 0$.

$\Rightarrow$ ¿Cómo controlaríamos por el efecto de tiempo? Con variables dummy.

Cortes transversales independientes

Consideremos el siguiente modelo

$$y_{ht} = \beta_0 + \beta_1 x_{ht} + \delta_0 a_t + \delta_1 x_{ht} \times a_t + e_{ht},$$

$$h = 1, 2, ..., (n_0 + n_1), t = 0, 1$$

donde a_t es una variable dummy que toma valor 0 en t=0 y 1 en t=1.

- En este modelo no solo tenemos un efecto del tiempo sobre y, sino también de tiempo en la relación misma de x con y.
- El intercepto para $t = 0$ es β_0 pero para $t = 1$ es $\beta_0 + \delta_0$.
- La pendiente (efecto de un cambio de 1 unidad de x en y) para $t = 0$ es β_1 pero para $t = 1$ es $\beta_1 + \delta_1$.
- ¿Cómo contrastaría que la pendiente en $t = 0$ es la misma que en $t = 1$?

Modelo de componentes del error en una dirección

En un panel longitudinal el mismo individuo es observado a lo largo del tiempo.

$$y_{it} = \alpha + X_{it}'\beta + u_{it}$$

$i = 1, 2, ..., N$ es el índice de individuos (firmas, familias, hogares, países),
$t = 1, 2, ..., T$ es el índice de tiempo,
α es un escalar, β es $K \times 1$, X_{it} es la observación *it* de las K variables explicativas.
El error tiene esta estructura:

$$u_{it} = \mu_i + \nu_{it}$$

es el **error compuesto**. *En inglés:* **one-way error components model**.

- μ_i: efecto individual no observado que captura todos los factores constantes a lo largo del tiempo en y_{it},
- ν_{it}: errores idiosincráticos o shocks.

Modelo de componentes del error en dos direcciones

Consideremos el modelo

$$y_{it} = \alpha + X_{it}'\beta + u_{it}$$

$i = 1, 2, ..., N$ es el índice de individuos (firmas, familias, hogares, países),
$t = 1, 2, ..., T$ es el índice de tiempo,
α es un escalar, β es $K \times 1$, X_{it} es la observación itde las K variables explicativas.

$$u_{it} = \mu_i + \lambda_t + \nu_{it}$$

es el **error compuesto**. *En inglés:* **two-way error components model**.

- μ_i: efecto individual no observado que captura todos los factores constantes a lo largo del tiempo en y_{it},
- λ_t: efecto temporal no observado que captura todos los factores constantes a lo de los individuos en y_{it},
- ν_{it}: errores idiosincráticos o shocks.

Inconsistencia de MCO

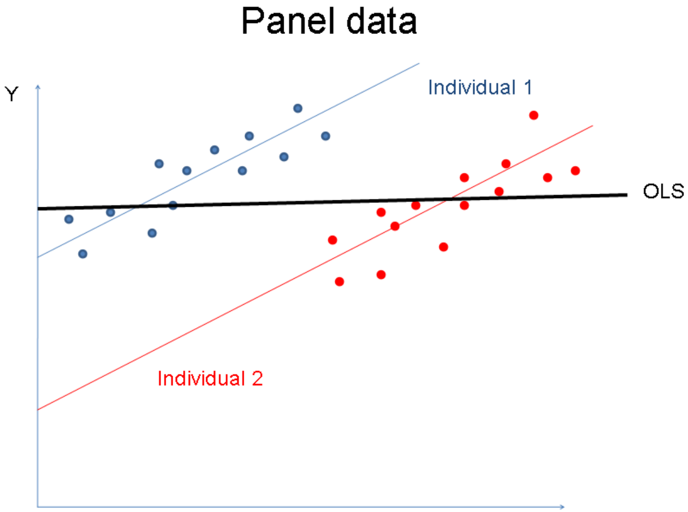
Si $cov(x_{it}, \mu_i) \neq 0 \Rightarrow$ MCO es sesgado e inconsistente. ¿Por qué?
Supongamos un modelo de regresión simple:

$$y_{it} = \beta_0 + \beta_1 x_{it} + \mu_i + \nu_{it} \; i = 1, 2, ..., N, t = 1, 2, ..., T,$$

donde $E[\nu_{it}|x_{it}] = 0, \forall i, t$.
Probar:

1. $\hat{\beta}_1^{MCO} = \frac{\sum_{i=1}^N \sum_{t=1}^T (x_{it}-\bar{x})(y_{it}-\bar{y})}{\sum_{i=1}^N \sum_{t=1}^T (x_{it}-\bar{x})^2} \xrightarrow{p} \beta_1 + \frac{cov(x_{it}, \mu_i)}{var x_{it}}$ cuando $N, T \to \infty$.
2. $E[\hat{\beta}_1^{MCO}|X] = \beta_1 + \frac{\sum_{i=1}^N \sum_{t=1}^T E[\mu_i|X](x_{it}-\bar{x})}{\sum_{i=1}^N \sum_{t=1}^T (x_{it}-\bar{x})^2}$.

Inconsistencia de MCO



Los siguientes estimadores se proponen como soluciones a este problema:

- **Estimador en primeras diferencias**
- **Efectos fijos**

Estimador en primeras diferencias (first differences,FD)

$$y_{it} = \beta_0 + \beta_1 x_{it} + \mu_i + \nu_{it}$$

$$y_{it-1} = \beta_0 + \beta_1 x_{it-1} + \mu_i + \nu_{it-1}$$

$$\Rightarrow \Delta y_{it} = \beta_1 \Delta x_{it} + \Delta \nu_{it}$$

donde Δ es el operador de diferencias, $\Delta y_{it} = y_{it} - y_{it-1}$.
Definamos $\hat{\beta}_1^{FD} = \frac{\sum_{i=1}^{N}\sum_{t=1}^{T}\Delta x_{it}\Delta y_{it}}{\sum_{i=1}^{N}\sum_{t=1}^{T}\Delta x_{it}^2}$. Probar que es un estimador insesgado y consistente cuando $N, T \rightarrow \infty$.

¡Lo que pasa es que μ_i desaparece, entonces se acaba el problema!

Efectos fijos (fixed-effects,FE)

$$y_{it} = \beta_0 + \beta_1 x_{it} + \mu_i + \nu_{it}$$

$$\bar{y}_i = \beta_0 + \beta_1 \bar{x}_i + \bar{\mu}_i + \bar{\nu}_i$$

$$\Rightarrow \tilde{y}_{it} = \beta_1 \tilde{x}_{it} + \tilde{\nu}_{i,t}$$

donde ˜ es una transformación que se aplica a cada invididuo (se usa en inglés **within** transformation), $\tilde{y}_{it} = y_{it} - \bar{y}_i$, $\bar{y}_{it} = T^{-1}\sum_{t=1}^{T} y_{it}$.

Definamos $\hat{\beta}_1^{FE} = \frac{\sum_{i=1}^{N}\sum_{t=1}^{T}\tilde{x}_{it}\tilde{y}_{it}}{\sum_{i=1}^{N}\sum_{t=1}^{T}\tilde{x}_{it}^2}$. Probar que es un estimador insesgado y consistente cuando N, $T \to \infty$.

¡Otra vez deseparece μ_i! (porque $\bar{\mu}_i = T^{-1}\sum_{t=1}^{T}\mu_i = T^{-1}T\mu_i = \mu_i$)

Efectos fijos (fixed-effects,FE)

En los estimadores de efectos fijos μ_i *se asumen como* **parámetros fijos a ser estimados** *y* $\nu_{it} \sim IID(0, \sigma_\nu^2)$. $x_{it} \perp\!\!\!\perp \nu_{it}, \forall i, t$.
Otro modo de ver los modelos FE es $E(\nu_{it}|x_{it}) \neq 0$ *pero que* $E(\nu_{it}|x_{it}, \mu_i) = 0$.
Entonces necesitamos estimar μ_i.

Efectos fijos (fixed-effects,FE)

Notar que $\hat{\beta}_{FE}$ es equivalente a un modelo MCO con una dummy para cada individuo i, $\hat{\beta}_{LSDV}$.

La prueba es una aplicación del Teorema de **Frisch-Waugh-Lovell**. Supongamos el modelo

$$\boldsymbol{y} = \boldsymbol{X}_1\beta_1 + \boldsymbol{X}_2\beta_2 + \boldsymbol{u}$$

con estimadores MCO $\hat{\boldsymbol{\beta}} = [\hat{\boldsymbol{\beta}}_1\hat{\boldsymbol{\beta}}_2]$.
El teorema muestra que

$$\hat{\boldsymbol{\beta}}_1 = (\boldsymbol{X}_1'\boldsymbol{M}_2\boldsymbol{X}_1)^{-1}\boldsymbol{X}_1'\boldsymbol{M}_2\boldsymbol{y}$$

Si tomamos que X_2 es el conjunto de las **dummies** por individuo y X_1 son las variables de control, se puede demostrar que la matriz $\boldsymbol{M}_2$ realiza la transformación **within** de las variables. Entonces nos queda una regresión MCO de $\boldsymbol{M}_2\boldsymbol{y}$ en $\boldsymbol{M}_2\boldsymbol{X}_1$.

Efectos fijos (fixed-effects,FE)

Probar que el estimador FE es un promedio ponderado de los estimadores MCO individuales. Es decir, $\hat{\beta}_{FE} = \sum_{i=1}^{N} \omega_i \hat{\beta}^i_{MCO}$ donde $\omega_i \geq 0, i = 1, 2, \ldots, N$ es un ponderador tal que $\sum_{i=1}^{N} \omega_i = 1$ y $\hat{\beta}^i_{MCO}$ son los estimadores de MCO de cada individuo i por separado.

Resolvamos el caso para la regresión simple. $y_{it} = \beta_0 + \beta_1 x_{it} + \mu_i + \nu_{it}$.
Supongamos que corremos una regresión para cada i por separado, usando solamente las T observaciones. Entonces, tenemos $\hat{\beta}_1^{MCO,i} = \frac{\sum_{t=1}^{T}(x_{it}-\bar{x}_i)(y_{it}-\bar{y}_i)}{\sum_{t=1}^{T}(x_{it}-\bar{x}_i)^2}$.
Ahora, el estimador FE es

$$\hat{\beta}_1^{FE} = \frac{\sum_{i=1}^{N}\sum_{t=1}^{T}(x_{it}-\bar{x}_i)(y_{it}-\bar{y}_i)}{\sum_{i=1}^{N}\sum_{t=1}^{T}(x_{it}-\bar{x}_i)^2} = \frac{\sum_{i=1}^{N}\sum_{t=1}^{T}\frac{\sum_{t=1}^{T}(x_{it}-\bar{x}_i)^2}{\sum_{t=1}^{T}(x_{it}-\bar{x}_i)^2}(x_{it}-\bar{x}_i)(y_{it}-\bar{y}_i)}{\sum_{i=1}^{N}\sum_{t=1}^{T}(x_{it}-\bar{x}_i)^2}$$

$$= \frac{\sum_{i=1}^{N}\sum_{t=1}^{T}\left(\sum_{t=1}^{T}(x_{it}-\bar{x}_i)^2\right)\hat{\beta}_1^{MCO,i}}{\sum_{i=1}^{N}\sum_{t=1}^{T}(x_{it}-\bar{x}_i)^2} = \sum_{i=1}^{N}\omega_i\hat{\beta}_1^{MCO,i},$$

donde $\omega_i = \frac{\sum_{t=1}^{T}(x_{it}-\bar{x}_i)^2}{\sum_{i=1}^{N}\sum_{t=1}^{T}(x_{it}-\bar{x}_i)^2}$.

Efectos fijos (fixed-effects,FE)

- Podemos plantear $\hat{\mu}_i = \bar{y}_i - \hat{\beta}_{FE}\bar{x}_i$ como el estimador de los "efectos fijos".
- Si $T \to \infty$, $(\hat{\mu}_{FE}, \hat{\beta}_{FE})$ son estimadores consistentes e insesgados.
- Pero si T está fijo y $N \to \infty$, sólo $\hat{\beta}_{FE}$ es consistente, aunque ambos son insesgados. El problema es conocido como el problema de parámetros incidentales de Neyman y Scott (1948).
- Contraste de efectos fijos: correr el modelo LSDV, y contrastar conjuntamente por la significatividad de las dummies.

Efectos fijos (fixed-effects,FE)

Usar efectos fijos tiene algunas desventajas:

- Hay un costo en comparación con MCO. La transformación within es equivalente a estimar una dummy para cada individuo. O sea estimar N parámetros adicionales. Más parámetros significa menos presición.
- Entonces eso afecta los grados de libertad y por ende la precisión de lo que estimamos. Los grados de libertad son $NT - N - K$
- También la transformación within (y first-differences) elimina TODO aquello que esta fijo para cada individuo. Entonces no se puede medir por ejemplo el efecto de SEXO, para individuos, o CONTINENTE para países.

Efectos aleatorios (random-effects,RE)

Un modelo alternativo es el de efectos aleatorios. Tiene un supuesto MUY importante y restrictivo: $cov(X, \mu) = 0$.
Consideremos el modelo

$$y_{it} = \beta_0 + \beta_1 x_{it} + u_{it} = \beta_0 + \beta_1 x_{it} + \mu_i + \nu_{it}.$$

En este caso hay **correlación serial**:
$Cov(u_{it}, u_{is}) = Cov(\mu_i + \nu_{it}, \mu_i + \nu_{is}) = Var(\mu_i) \neq 0$ para $t \neq s$. El estimador de efectos aleatorios (RE) tiene en cuenta esta particularidad y produce un estimador eficiente GLS:

$$\hat{\beta}_{RE} = (X'\Omega^{-1}X)^{-1}(X'\Omega^{-1}y),$$

donde $\Omega = E(uu')$ es la matriz de varianzas-covarianzas de los errores compuestos.
Una de las ventajas de RE es que se pueden reincorporar variables fijas por individuos:

$$y_{it} = \beta_0 + \beta_1 x_{it} + \beta_2 z_i + \mu_i + \nu_{it},$$

donde z captura todas las variables que estan fijas para cada individuo y que se pueden observar.

Efectos aleatorios (random-effects,RE)

En el modelo RE $\mu_i \sim iid(0, \sigma_\mu^2)$, $\nu_{it} \sim iid(0, \sigma_\nu^2)$, $\mu_i \perp\!\!\!\perp \nu_{it}$. *Además tenemos* $X_{it} \perp\!\!\!\perp \mu_i$ *y* $X_{it} \perp\!\!\!\perp \nu_{it}$ *para todo i y t.*

Definamos a Ω como la matriz de varianzas y covarianzas de todos los errores, que tiene esta estructura:

$$
\begin{aligned}
cov(u_{it}, u_{js}) &= \sigma_\mu^2 + \sigma_\nu^2 && para\ i = j, t = s \\
&= \sigma_\mu^2 && para\ i = j, t \neq s \\
&= 0 && para\ i \neq j
\end{aligned}
$$

Otra forma de verlo es como un modelo de **equicorrelación** intra cluster:

$$
\begin{aligned}
correl(u_{it}, u_{js}) &= 1 && para\ i = j, t = s \\
&= \sigma_\mu^2 / (\sigma_\mu^2 + \sigma_\nu^2) := \rho && para\ i = j, t \neq s \\
&= 0 && para\ i \neq j, t \neq s
\end{aligned}
$$

Efectos aleatorios (random-effects,RE)

Más en detalles, tenemos $\Omega =$

$$E\begin{bmatrix}
(\mu_1+\nu_{11})^2 & (\mu_1+\nu_{11})(\mu_1+\nu_{12}) & \cdots & (\mu_1+\nu_{11})(\mu_1+\nu_{1T}) & (\mu_1+\nu_{11})(\mu_2+\nu_{21}) & (\mu_1+\nu_{11})(\mu_2+\nu_{22}) & \cdots & (\mu_1+\nu_{11})(\mu_2+\nu_{2T}) & \cdots \\
(\mu_1+\nu_{12})(\mu_1+\nu_{11}) & (\mu_1+\nu_{12})^2 & \cdots & (\mu_1+\nu_{12})(\mu_1+\nu_{1T}) & (\mu_1+\nu_{12})(\mu_2+\nu_{21}) & (\mu_1+\nu_{12})(\mu_2+\nu_{22}) & \cdots & (\mu_1+\nu_{12})(\mu_2+\nu_{2T}) & \cdots \\
\vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\
(\mu_1+\nu_{1T})(\mu_1+\nu_{11}) & (\mu_1+\nu_{1T})(\mu_1+\nu_{12}) & \cdots & (\mu_1+\nu_{1T})^2 & (\mu_1+\nu_{1T})(\mu_2+\nu_{21}) & (\mu_1+\nu_{1T})(\mu_2+\nu_{22}) & \cdots & (\mu_1+\nu_{1T})(\mu_2+\nu_{2T}) & \cdots \\
(\mu_2+\nu_{21})(mu_1+\nu_{11}) & (\mu_2+\nu_{21})(\mu_1+\nu_{12}) & \cdots & (\mu_2+\nu_{21})(\mu_1+\nu_{1T}) & (\mu_2+\nu_{21})^2 & (\mu_2+\nu_{21})(\mu_2+\nu_{22}) & \cdots & (\mu_2+\nu_{21})(\mu_2+\nu_{2T}) & \cdots \\
(\mu_2+\nu_{22})(mu_1+\nu_{11}) & (\mu_2+\nu_{22})(\mu_1+\nu_{12}) & \cdots & (\mu_2+\nu_{22})(\mu_1+\nu_{1T}) & (\mu_2+\nu_{22})(\mu_2+\nu_{21}) & (\mu_2+\nu_{22})^2 & \cdots & (\mu_2+\nu_{22})(\mu_2+\nu_{2T}) & \cdots \\
\vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\
(\mu_2+\nu_{2T})(mu_1+\nu_{11}) & (\mu_2+\nu_{2T})(\mu_1+\nu_{12}) & \cdots & (\mu_2+\nu_{2T})(\mu_1+\nu_{1T}) & (\mu_2+\nu_{2T})(\mu_2+\nu_{21}) & (\mu_2+\nu_{2T})(\mu_2+\nu_{22}) & \cdots & (\mu_2+\nu_{2T})^2 & \cdots \\
\vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\
(\mu_N+\nu_{N1})(mu_1+\nu_{11}) & (\mu_N+\nu_{N1})(\mu_1+\nu_{12}) & \cdots & (\mu_N+\nu_{N1})(\mu_1+\nu_{1T}) & (\mu_N+\nu_{N1})(\mu_2+\nu_{21}) & (\mu_N+\nu_{N1})(\mu_2+\nu_{22}) & \cdots & (\mu_N+\nu_{N1})(\mu_2+\nu_{2T}) & \cdots \\
(\mu_N+\nu_{N2})(mu_1+\nu_{11}) & (\mu_N+\nu_{N2})(\mu_1+\nu_{12}) & \cdots & (\mu_N+\nu_{N2})(\mu_1+\nu_{1T}) & (\mu_N+\nu_{N2})(\mu_2+\nu_{21}) & (\mu_N+\nu_{N2})(\mu_2+\nu_{22}) & \cdots & (\mu_N+\nu_{N2})(\mu_2+\nu_{2T}) & \cdots \\
\vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\
(\mu_N+\nu_{NT})(mu_1+\nu_{11}) & (\mu_N+\nu_{NT})(\mu_1+\nu_{12}) & \cdots & (\mu_N+\nu_{NT})(\mu_1+\nu_{1T}) & (\mu_N+\nu_{NT})(\mu_2+\nu_{21}) & (\mu_N+\nu_{NT})(\mu_2+\nu_{22}) & \cdots & (\mu_N+\nu_{NT})(\mu_2+\nu_{2T}) & \cdots \\
\vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots & \ddots
\end{bmatrix}$$

Efectos aleatorios (random-effects,RE)

Más en detalles, tenemos $\Omega =$

$$\begin{bmatrix}
\sigma_\mu^2+\sigma_\nu^2 & \sigma_\mu^2 & \dots & \sigma_\mu^2 & 0 & 0 & \dots & 0 & \dots & 0 & 0 & \dots & 0 \\
\sigma_\nu^2 & \sigma_\mu^2+\sigma_\nu^2 & \dots & \sigma_\mu^2 & 0 & 0 & \dots & 0 & \dots & 0 & 0 & \dots & 0 \\
\vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots & \\
\sigma_\nu^2 & \sigma_\mu^2 & \ddots & \sigma_\mu^2+\sigma_\nu^2 & 0 & 0 & \dots & 0 & \dots & 0 & 0 & \dots & 0 \\
0 & 0 & \dots & 0 & \sigma_\mu^2+\sigma_\nu^2 & \sigma_\mu^2 & \dots & \sigma_\mu^2 & \dots & 0 & 0 & \dots & 0 \\
0 & 0 & \dots & 0 & \sigma_\mu^2 & \sigma_\mu^2+\sigma_\nu^2 & \dots & \sigma_\mu^2 & \dots & 0 & 0 & \dots & 0 \\
\vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots & \\
0 & 0 & \dots & 0 & \sigma_\mu^2 & \sigma_\mu^2 & \ddots & \sigma_\mu^2+\sigma_\nu^2 & \dots & 0 & 0 & \dots & 0 \\
0 & 0 & \dots & 0 & 0 & 0 & \dots & 0 & \dots & \sigma_\mu^2+\sigma_\nu^2 & \sigma_\mu^2 & \dots & \sigma_\mu^2 \\
0 & 0 & \dots & 0 & 0 & 0 & \dots & 0 & \dots & \sigma_\mu^2 & \sigma_\mu^2+\sigma_\nu^2 & \dots & \sigma_\mu^2 \\
\vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots & \\
0 & 0 & \dots & 0 & 0 & 0 & \dots & 0 & \dots & \sigma_\mu^2 & \sigma_\mu^2 & \ddots & \sigma_\mu^2+\sigma_\nu^2
\end{bmatrix}$$

Notar la estructura de **diagonal por bloques**.

Efectos aleatorios (random-effects,RE)

¿Cuáles serían las consecuancias de ignorar la correlación serial?
Supongamos el siguiente modelo de regresión univariada: $y_{it} = \beta x_{it} + \mu_i + \nu_{it}$, donde $\mu_i \sim iid(0, \sigma_\mu^2)$, $\nu_{it} \sim iid(0, \sigma_\nu^2)$, $\mu_i \perp\!\!\!\perp \nu_{it}$. Además tenemos $x_{it} \perp\!\!\!\perp \mu_i$ y $x_{it} \perp\!\!\!\perp \nu_{it}$ para todo i y t.

- Por un lado se puede probar que el estimador OLS $\hat{\beta}_{OLS} = \frac{\sum_{i=1}^N \sum_{t=1}^T y_{it} x_{it}}{\sum_{i=1}^N \sum_{t=1}^T x_{it}^2}$ es insesgado y consistente.
- Sin embargo, la varianza de OLS estimada (bajo los supuestos de independencia de los errores que es lo que estimaríamos al ignorar la correlación entre observaciones del mismo individuo) es

$$var(\hat{\beta}_{OLS}.est) = (\sigma_\mu^2 + \sigma_\nu^2) \frac{1}{\sum_{i=1}^N \sum_{t=1}^T x_{it}^2},$$

pero la varianza correcta es

$$var(\hat{\beta}_{OLS}) = (\sigma_\mu^2 + \sigma_\nu^2) \frac{1}{\sum_{i=1}^N \sum_{t=1}^T x_{it}^2} + \sigma_\mu^2 \frac{\sum_{i=1}^N \sum_{t=1}^T \sum_{s=1, t\neq s}^T x_{it} x_{is}}{(\sum_{i=1}^N \sum_{t=1}^T x_{it}^2)^2}.$$

- El resultado es que los errores estándar que estima STATA por default son incorrectos, por lo que toda la inferencia es incorrecta.
- Esta diferencia se llama el "factor de Moulton". Notar que depende de la **correlación intra-individuo** de las Xs. En particular si las Xs no están relacionadas entre sí para un mismo individuo, o sea $Cov(x_{it}, x_{is}) = 0, t \neq s$, entonces la varianza estándar es correcta.

Efectos aleatorios (random-effects,RE)

Prueba de la diapositiva anterior:
Consideremos la varianza condicional en x. Esto simplifica las pruebas.

$$var(\hat{\beta}_{OLS}) = var\left(\frac{\sum_{i=1}^{N}\sum_{t=1}^{T} x_{it}y_{it}}{\sum_{i=1}^{N}\sum_{t=1}^{T} x_{it}^2}\right) = var\left(\frac{\sum_{i=1}^{N}\sum_{t=1}^{T} x_{it}(\beta x_{it} + \mu_i + \nu_{it})}{\sum_{i=1}^{N}\sum_{t=1}^{T} x_{it}^2}\right) = \frac{var\left(\sum_{i=1}^{N}\sum_{t=1}^{T} \mu_i x_{it} + \sum_{i=1}^{N}\sum_{t=1}^{T} \nu_{it}x_{it}\right)}{(\sum_{i=1}^{N}\sum_{t=1}^{T} x_{it}^2)^2}$$

Ahora hay que considerar que la varianza de la suma es la suma de las varianzas más todas las covarianzas. Como se asume independencia entre individuos $i = 1, 2, ..., N$, y también entre μ_i y ν_{it}, podemos escribir

$$= \frac{\sum_{i=1}^{N} var\left(\sum_{t=1}^{T} \mu_i x_{it}\right) + \sum_{i=1}^{N} var\left(\sum_{t=1}^{T} \nu_{it}x_{it}\right)}{(\sum_{i=1}^{N}\sum_{t=1}^{T} x_{it}^2)^2}$$

El punto central aparece ahora. Por un lado, $var\left(\sum_{t=1}^{T} \nu_{it}x_{it}\right) = \sum_{t=1}^{T} var\left(\nu_{it}x_{it}\right) = \sum_{t=1}^{T} \sigma_\nu^2 x_{it}^2$, ya que se asume que no hay correlación serial en ν_{it}. Por otro lado,
$var\left(\sum_{t=1}^{T} \mu_i x_{it}\right) = \sum_{t=1}^{T} var(\mu_i)x_{it}^2 + \sum_{t=1}^{T}\sum_{s=1,s\neq t}^{T} cov(\mu_i x_{it}, \mu_i x_{is})$, ya que aquí hay covarianzas que no son cero. En particular, $cov(\mu_i x_{it}, \mu_i x_{is}) = var(\mu_i)x_{it}x_{is} = \sigma_\mu^2 x_{it}x_{is}$.
Entonces, usando estos resultados tenemos

$$var(\hat{\beta}_{OLS}) = (\sigma_\mu^2 + \sigma_\nu^2)\frac{1}{\sum_{i=1}^{N}\sum_{t=1}^{T} x_{it}^2} + \sigma_\mu^2 \frac{\sum_{i=1}^{N}\sum_{t=1}^{T}\sum_{s=1,t\neq s}^{T} x_{it}x_{is}}{(\sum_{i=1}^{N}\sum_{t=1}^{T} x_{it}^2)^2}.$$

Efectos aleatorios (random-effects,RE)

Análisis de la diapositiva anterior:
El término adicional a considerar es $\hat{C}_X = \frac{1}{NT(T-1)} \sum_{i=1}^{N} \sum_{t=1}^{T} \sum_{s=1,t\neq s}^{T} x_{it}x_{is}$.

- Notar que $\hat{C}_X$ lo podemos definir como un estimador de la covarianza intra-individuo de las Xs.
- Si $\hat{C}_X = 0$, entonces la varianza con los supuestos de Gauss-Markov es correcta. Esto implica que no solo hay que considerar la covarianza intra-individuo o **intra-cluster**en los errores, sino también de las Xs.

Tomemos ahora $\hat{V}_X = \frac{1}{NT} \sum_{i=1}^{N} \sum_{t=1}^{T} x_{it}^2$.

- Notar que $\hat{V}_X$ lo podemos definir como un estimador de la varianza intra-individuo de las Xs.
- Si $\hat{V}_X > 0$, tiene que haber variabilidad en las Xs.

Ahora definimos $\hat{\rho}_X = \frac{\hat{C}_X}{\hat{V}_X}$, como la correlación intra-individuo o **intra-cluster** de las Xs.

Entonces,

$$var(\hat{\beta}_{OLS}) = \frac{(\sigma_\mu^2 + \sigma_v^2)}{NT\hat{V}_X} + \frac{(T-1)}{NT\hat{V}_X}\sigma_\mu^2\hat{\rho}_X = var(\hat{\beta}_{OLS}.est) + \frac{(T-1)}{NT\hat{V}_X}\sigma_\mu^2\hat{\rho}_X$$

Contraste de Hausman

- Un contraste de la validez de efectos aleatorios es

$$H_0 : Cov(X, \mu) = 0$$

$$H_A : Cov(X, \mu) \neq 0$$

La hipótesis nula es que las variables X no están relacionadas con los efectos individuales, μ.

- Para evaluar este contraste se compara $\hat{\beta}_{FE}$ con $\hat{\beta}_{RE}$. El estimador FE es siempre **consistente**. Sin embargo, el de RE es válido si las Xs no estan correlacionadas con μ.

$$H_0 : \beta_{FE} - \beta_{RE} = 0$$

$$H_A : \beta_{FE} - \beta_{RE} \neq 0$$

- Entonces, si la diferencia es aproximandamente 0 no hay evidencia suficiente para que H_0 sea rechazada. Si por el contrario $\hat{\beta}_{FE} - \hat{\beta}_{RE} \neq 0$ entonces sí hay que rechazar la nula. Este es el llamado **contraste de Hausman**.

Contraste de Hausman

- Notemos que el estimador RE es GLS, entonces es el de menor varianza. Por lo tanto $var(\hat{\beta}_{FE}) - var(\hat{\beta}_{RE})$ es una matriz definida positiva. Para demostrarlo notemos que tanto W_{XX} como B_{XX} son matrices definidas positivas (formas cuadráticas), mientras que $\phi^2 > 0$. Entonces $[W_{XX} + \phi^2 B_{XX}] - W_{XX}$ es definida positiva, por lo que $[W_{XX} + \phi^2 B_{XX}]^{-1} - W_{XX}^{-1}$ es definida negativa, o $var(\hat{\beta}_{RE}) - var(\hat{\beta}_{FE})$ es definida negativa.
- Se puede probar que $var(\hat{\beta}_{FE} - \hat{\beta}_{RE}) = var(\hat{\beta}_{FE}) - var(\hat{\beta}_{RE})$.
- Entonces, definamos

$$H = (\hat{\beta}_{FE} - \hat{\beta}_{RE})'[var(\hat{\beta}_{FE}) - var(\hat{\beta}_{RE})]^{-1}(\hat{\beta}_{FE} - \hat{\beta}_{RE})$$

- El estadístico H una distribución χ^2_K bajo la nula. Se rechaza H_0 si $H > \chi^2_K(1-\alpha)$ (valor crítico).

¿Cómo implementar paneles en STATA?

- Los datos hay que organizarlos:

id	tiempo	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}

  Es muy importante que no haya valores repetidos.
- Primero STATA tiene que identificar que se trata de datos en paneles. Para eso se nececita una variable numérica, ej. `id`, que identifica el individuo. Luego, `iis id`
- Pero si se tiene una muestra longitudinal con una estructura de series de tiempo, con una variable `tiempo`,
  `tsset id tiempo`
- Nota: la variable de tiempo tiene que ser en números discretos consecutivos. O sea, $t = -2, -1, 0, 1, ...$

¿Cómo implementar paneles en STATA?

- Entonces estamos listos para usar datos en paneles:
- `reg D.y D.x1 D.x2 D.x3` (modelo en diferencias)
- `xtreg y x1 x2 x3, fe` (modelo de efectos fijos)
- `xi: reg y x1 x2 x3 i.id` (lo mismo pero implementado "a mano" con dummies para id) [Nota: comparar con una regresión de las variables transformadas within. ¿Cúal sería el problema con este modelo?]
- `xtreg y x1 x2 x3, re` (modelo de efectos aleatorios)
- `xtreg y x1 x2 x3, be` (modelo between)
- `xi: xtreg y x1 x2 x3 i.tiempo, fe` (modelo de efectos fijos, two-way)

¿Cómo implementar paneles en STATA?

- Contraste de Hausman test

```
xtreg y x1 x2 x3, fe
est store fe
xtreg y x1 x2 x3, re
est store re
hausman fe re
```

Ejemplos en la web

- http://fmwww.bc.edu/gstat/examples/wooldridge/wooldridge13.html
- http://fmwww.bc.edu/gstat/examples/wooldridge/wooldridge14.html
- https://www.stata.com/manuals13/xtxtreg.pdf