Modelos básicos de 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. **Muestras longitudinales**

Modelos

En un panel longitudinal el mismo individuo es observado a lo largo del tiempo.
$i = 1, 2, ..., N$ es el índice de individuos (firmas, familias, hogares, países),
$t = 1, 2, ..., T$ es el índice de tiempo.
Supongamos los siquientes modelos:

1. $y_{it} = \alpha + X'_{it}\beta + u_{it}$, α es un escalar, β es un vector $K \times 1$, X_{it} es la observación it de las K variables explicativas. **Pooled OLS**.
2. $y_{it} = \alpha_{it} + X'_{it}\beta_{it} + u_{it}$, α_{it} es un escalar espec'ifico para cada it, β_{it} es un vector $K \times 1$ específico para cada it, X_{it} es la observación it de las K variables explicativas. **Modelo imposible de identificar, igual cantidad de parámetros que de observaciones**.
3. $y_{it} = \alpha_i + X'_{it}\beta_i + u_{it}$, α_i es un escalar específico para cada i, β_i es un vector $K \times 1$ específico para cada i, X_{it} es la observación it de las K variables explicativas. **SUREG, Paneles heterogéneos**.
4. $y_{it} = \alpha_t + X'_{it}\beta_t + u_{it}$, α_t es un escalar específico para cada t, β_t es un vector $K \times 1$ específico para cada t, X_{it} es la observación it de las K variables explicativas.

One-way error components model

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

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.

One-way error components model

En notación matricial

$$y = \alpha\iota_{NT} + X\beta + u = Z\delta + u,$$

donde y y u son vectores $NT \times 1$, X es una matriz $NT \times K$, $Z = [\iota_{NT}', X']'$, $\delta' = [\alpha, \beta']$ y ι_{NT} es un vector de 1s con dimensión $NT \times 1$

$$u = Z_\mu \mu + \nu$$

donde $Z_\mu = I_N \otimes \iota_T$ es una matriz $NT \times N$ de 1s y 0s, $\otimes$ es el producto de Kronecker, ι_T es un vector de 1s de dimensión T, $\mu = [\mu_1, ..., \mu_N]'$ es un vector $N \times 1$ que contiene los efectos individuales, y ν es un vector $NT \times 1$ con los errores. De esta manera $Z_\mu \mu$ es un vector también $NT \times 1$.
Notar que Z_μ es una matriz de dummies para todos los individuos $i = 1, 2, ..., N$. Es decir, contiene una dummy para cada i.

Two-way error components model

Consideremos el modelo

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

α es un escalar, β es $K \times 1$, X_{it} es la observación *it* de 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.

Two-way error components model

En notación matricial

$$u = Z_\mu \mu + Z_\lambda \lambda + \nu,$$

donde $Z_\lambda = \iota_N \otimes I_T$ es una matriz $NT \times T$ de 1s y 0s, ι_N es un vector de 1s de dimensión N, $\lambda = [\lambda_1, ..., \lambda_T]'$ es un vector $T \times 1$ que contiene los efectos temporales.

Paneles multi-dimensionales

- Los modelos anteriores se pueden extender a una mayor cantidad de dimensiones.
- Ejemplos de 3 dimensiones (i, j, t):
  1. Modelos de comercio exterior podrían tener efectos por exportador, importador, tiempo, etc.
  2. Modelos de inmigración podrían tener efectos por país de origen, país de destino, año de inmigración, etc.

  En ambos casos podríamos tener efectos para cada elemento,

  $$u_{ijt} = \mu_i + \alpha_j + \lambda_t + \nu_{ijt},$$

  o efectos bilaterales, donde nos importan los efectos de a pares,

  $$u_{ijt} = \mu_{ij} + \lambda_t + \nu_{ijt}$$

- Un caso particular es el de los modelos anidados. Por ejemplo, si evaluamos políticas educativas, tendríamos un efecto por alumno (i), que pertenece a una determinada clase (j), de una escuela (k) a lo largo del tiempo (t). Así podríamos pensar en una estructura de errores:

  $$u_{ijkt} = \mu_i + \alpha_j + \gamma_k + \lambda_t + \nu_{ijkt}$$

- Ver el libro de Laszlo Matyas (`http://www.metrixmdp.eu`)

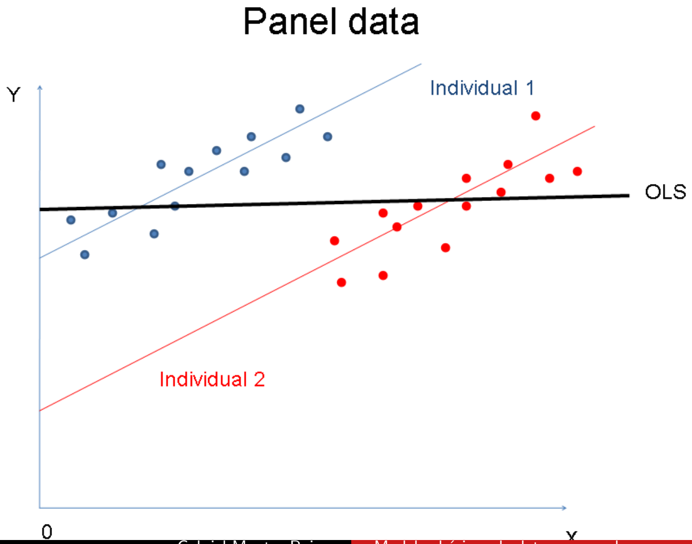
Inconsistencia de OLS

Si $cov(x_{it}, \mu_i) \neq 0 \Rightarrow$ OLS 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. $\hat{\beta}_1^{OLS} = \frac{\sum_{i=1}^N \sum_{t=1}^T (x_{it}-\bar{x})(y_{it}-\bar{y})}{\sum_{i=1}^N \sum_{t=1}^T (x_{it}-\bar{x})^2} = \beta_1 + \frac{\sum_{i=1}^N T\mu_i(\bar{x}_i-\bar{x})+\sum_{i=1}^N \sum_{t=1}^T (x_{it}-\bar{x})(\nu_{it}-\bar{\nu})}{\sum_{i=1}^N \sum_{t=1}^T (x_{it}-\bar{x})^2}$
$\xrightarrow{p} \beta_1 + \frac{cov(x_{it},\mu_i)}{var x_{it}}$ cuando $N, T \to \infty$ y donde $\bar{x}_i = \frac{1}{T}\sum_{t=1}^T x_{it}$.
2. $E[\hat{\beta}_1^{OLS}|X] = \beta_1 + \frac{\sum_{i=1}^N \sum_{t=1}^T E[\mu_i|X](x_{it}-\bar{x})}{\sum_{i=1}^N \sum_{t=1}^T (x_{it}-\bar{x})^2}$.

Inconsistencia de OLS



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 $N, T \to \infty$.

¡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)

- En notación matricial

$$y = \alpha\iota_{NT} + X\beta + Z_\mu\mu + \nu = Z\delta + Z_\mu\mu + \nu$$

- Definamos $P_\mu = Z_\mu(Z'_\mu Z_\mu)^{-1}Z'_\mu$ como la matriz de proyección en Z_μ, el subespacio de dummies por individuo. $Z_\mu Z'_\mu = I_N \otimes J_T$. Entonces $P_\mu = Z_\mu(Z'_\mu Z_\mu)^{-1}Z'_\mu = I_N \otimes \overline{J_T}$ donde $\overline{J_T} = J_T/T$.
- P_μ es una matriz que promedia las observaciones para los individuos. Así $P_\mu y$ tiene elementos $\bar{y}_i = 1/T\sum_{t=1}^{T} y_{it}$ repetido T veces para cada i.
- Probar que P_μ es simétrica ($P'_\mu = P_\mu$) e idempotente ($P_\mu \times P_\mu = P_\mu$) por lo que $rank(P_\mu) = tr(P_\mu) = N$.

Efectos fijos (fixed-effects,FE)

- Definamos también el complemento $Q_\mu = I_{NT} - P_\mu$. Q_μ se define entonces como las desviaciones con respecto a las medias (en inglés within-group operator): $\tilde{y}_{it} = y_{it} - \bar{y}_i$.
- Probar que Q_μ es simétrica e idempotente, y que $P_\mu Q_\mu = 0_{NT}$, por lo que $rank(Q_\mu) = tr(Q_\mu) = N(T-1)$.
- Entonces,

$$Q_\mu y = Q_\mu X\beta + Q_\mu \nu, \Rightarrow \hat{\beta}_{FE} = (X'Q_\mu X)^{-1} X'Q_\mu y$$

- ¿Qué tipo de variables quedan excluidas?

Efectos fijos (fixed-effects,FE)

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

La prueba es una aplicación del Teorema de **Frisch-Waugh-Lovell**.
Nota: $\hat{\beta}_{FE} = (\sum_{i=1}^{N}(X_i'(I_T - \bar{J}_T)X_i))^{-1}\sum_{i=1}^{N}(X_i'(I_T - \bar{J}_T)y_i)$.

Teorema de Frisch-Waugh-Lovell

Consideremos los siguientes modelos de regresión:

$$A.\ y = \beta_1 x_1 + \beta_2 x_2 + ... + \beta_K x_K + u$$

$$B.\ y = \gamma_2 x_2 + \gamma_3 x_3 + ... + \gamma_K x_K + v$$

$$C.\ x_1 = \delta_2 x_2 + \delta_3 x_3 + ... + \delta_K x_K + e$$

- Computar los residuos del modelo B ($\hat{v}$) y C ($\hat{e}$).
- Correr la siguiente regresión auxiliar: $\hat{v} = \alpha_0 + \alpha_1 \hat{e} + \textit{residuo}$
- Chequear que $\hat{\alpha}_1 = \hat{\beta}_1$
- ¿Cómo se interpreta este resultado?

Teorema de Frisch-Waugh-Lovell

En notación matricial

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

y consideremos $\hat{\beta} = [\hat{\beta}_1 \;\; \hat{\beta}_2]'$.
Ahora construyamos $\boldsymbol{M}_2\boldsymbol{y} = \boldsymbol{M}_2\boldsymbol{X}_1\beta_1 + \boldsymbol{M}_2\boldsymbol{u}$ donde $\boldsymbol{M}_2$ es la proyección residual de $\boldsymbol{X}_2$. El teorema de FWL muestra que

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

Prueba: Consideremos $\boldsymbol{y} = \boldsymbol{P}_{\boldsymbol{X}}\boldsymbol{y} + \boldsymbol{M}_{\boldsymbol{X}}\boldsymbol{y} = \boldsymbol{X}_1\hat{\beta}_1 + \boldsymbol{X}_2\hat{\beta}_2 + \boldsymbol{M}_{\boldsymbol{X}}\boldsymbol{y}$. Multipliquemos ambos lados por $\boldsymbol{X}_1'\boldsymbol{M}_2$ y obtenemos $\boldsymbol{X}_1'\boldsymbol{M}_2\boldsymbol{y} = \boldsymbol{X}_1'\boldsymbol{M}_2\boldsymbol{X}_1\hat{\beta}_1$. (Usamos el resultado $\boldsymbol{M}_2\boldsymbol{X}_2 = \boldsymbol{0}$ y $\boldsymbol{M}_{\boldsymbol{X}}\boldsymbol{M}_2\boldsymbol{X}_1 = \boldsymbol{0}$). Resolver por $\hat{\beta}_1$,

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

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 $\boldsymbol{M}_2 = \boldsymbol{Q}_\mu$ realiza la transformación within de las variables. Entonces nos queda una regresión MCO de $\boldsymbol{M}_2\boldsymbol{y}$ en $\boldsymbol{M}_2\boldsymbol{X}_1$.

Algebra de MCO (nota para la slide anterior)

Proyección ortogonal: *Definamos la matriz* $\boldsymbol{P_X} \equiv \boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'$ *como la matriz que proyecta y en el espacio generado por* $\boldsymbol{X}$. *Tenemos que* $\hat{\boldsymbol{y}} = \boldsymbol{P_X y}$, *los valores predecidos.*

Proyección residual: *Definamos la matriz* $\boldsymbol{M_X} \equiv \boldsymbol{I_N} - \boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'$ *como la proyección de y en el complemento del espacio de* $\boldsymbol{X}$. *Notemos que* $\hat{\boldsymbol{u}} = \boldsymbol{M_X y}$, *residuos de la regresión.*

Notar que $\boldsymbol{P_X M_X} = \boldsymbol{0}$. Además ambas matrices son idempotentes: $\boldsymbol{P_X P_X} = \boldsymbol{P_X}, \boldsymbol{M_X M_X} = \boldsymbol{M_X}$.

Efectos fijos (fixed-effects,FE)

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

- Notar que $var(\hat{\beta}_{FE}) = \sigma_\nu^2 (X' Q_\mu X)^{-1}$.
- Podemos plantear $\hat{\mu}_i = \bar{y}_i - \hat{\beta}_{FE} \bar{x}_i$ o $\hat{\mu}_{FE} = P_\mu y - \hat{\beta}_{FE} P_\mu X$ 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)-two-way

Para el modelo two-way tenemos $Z_\lambda Z'_\lambda = J_N \otimes I_T$. Entonces $P_\lambda = Z_\lambda (Z'_\lambda Z_\lambda)^{-1} Z'_\lambda = \overline{J_N} \otimes I_T$ donde $\overline{J_N} = J_N / N$.
Definamos la proyección residual

$$Q_{\mu\lambda} = E_N \otimes E_T = I_N \otimes I_T - I_N \otimes \overline{J_T} + \overline{J_N} \otimes I_T + \overline{J_N} \otimes \overline{J_T}$$

donde $E_N = I_N - \overline{J_N}$ y $E_T = I_T - \overline{J_T}$. Esta transformación elimina los efectos de μ_i y λ_t simultáneamente. La matriz $Q_{\mu\lambda}$ hace la siguiente transformación: $y_{it} - \bar{y}_i - \bar{y}_t + \bar{y}$. Entonces,

$$Q_{\mu\lambda} y = Q_{\mu\lambda} X\beta + Q_{\mu\lambda}\nu, \Rightarrow \hat{\beta}_{FE-2} = (X' Q_{\mu\lambda} X)^{-1} X' Q_{\mu\lambda} y$$

¿Qué tipo de variables quedan excluidas?

Efectos fijos (fixed-effects,FE)

Usar efectos fijos tiene algunas desventajas:

- Hay un costo en comparación con OLS sin controlar por las dummies por individuo. La transformación within es equivalente a estimar una dummy para cada individuo. O sea estimar $N-1$ parámetros adicionales (en el modelo one-way). Más parámetros significa menos precisión.
- Entonces eso afecta los grados de libertad y por ende la precisión de lo que estimamos. Los grados de libertad son $NT - N - K$, comparado con $NT - 1 - K$
- 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 fijos (fixed-effects,FE)

Comparando FE con FD (Baltagi, p.17):

- FE es más eficiente que FD cuando $\nu_{it} \sim iid(0, \sigma_\nu^2)$.
- FD es más eficiente que FE cuando ν_{it} es un paseo aleatorio (random walk).

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.

$$\Omega = E(uu') = Z_\mu E(\mu\mu')Z_\mu' + E(\nu\nu') = \sigma_\mu^2(I_N \otimes J_T) + \sigma_\nu^2(I_N \otimes I_T).$$

Ω 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 $\mu_i \sim iid(0, \sigma^2_\mu)$, $\nu_{it} \sim iid(0, \sigma^2_\nu)$, $\mu_i \perp\!\!\!\perp \nu_{it}$. 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 $\hat{\beta}_{OLS} = \frac{\sum_{i=1}^{N}\sum_{t=1}^{T} y_{it}x_{it}}{\sum_{i=1}^{N}\sum_{t=1}^{T} x^2_{it}}$ 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

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

  pero la varianza correcta es

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

- 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,

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

Efectos aleatorios (random-effects,RE)

Para estimarlo vamos a tratar el problema como GLS (generalized least squares) donde se estima Ω.

- Siguiendo a Wansbeek y Kapteyn (1982,1983), Baltagi p.18, reemplacemos J_T por $T\overline{J_T}$, y I_T por $(E_T + \overline{J_T})$, donde por definición $E_T = I_T - \overline{J_T}$. Entonces reagrupando términos tenemos

$$\Omega = T\sigma_\mu^2(I_N \otimes \overline{J_T}) + \sigma_\nu^2(I_N \otimes E_T) + \sigma_\nu^2(I_N \otimes \overline{J_T})$$

$$= (T\sigma_\mu^2 + \sigma_\nu^2)(I_N \otimes \overline{J_T}) + \sigma_\nu^2(I_N \otimes E_T) = \sigma_1^2 P_\mu + \sigma_\nu^2 Q_\mu$$

  donde $\sigma_1^2 = (T\sigma_\mu^2 + \sigma_\nu^2)$.

- Esta es la descomposición espectral de Ω donde σ_1^2 (de multiplicidad N) y σ_ν^2 (de multiplicidad $N(T-1)$) son las raíces características de Ω.
- Esto permite hallar $\Omega^r = \sigma_1^{2r} P + \sigma_\nu^{2r} Q$ para todo r (en particular $r = -1$). [¿Por qué? Probar que es cierto.]

Efectos aleatorios (random-effects,RE)

- Entonces, surgen los siguientes estimadores

$$\hat{\sigma}_1^2 = \frac{u' P_\mu u}{tr(P_\mu)} = T \sum_{i=1}^{N} \bar{u}_i^2 / N$$

$$\hat{\sigma}_\nu^2 = \frac{u' Q_\mu u}{tr(Q_\mu)} = \frac{\sum_{t=1}^{T} \sum_{i=1}^{N} (u_{it} - \bar{u}_i)^2}{N(T-1)}$$

- Notar que u no es observado... hay diferentes alternativas, todas ellas consistentes: (i) usar los residuos OLS, (ii) los residuos de FE, y (iii) otras.

Efectos aleatorios (random-effects,RE)

El estimador RE es entonces

$$\begin{aligned}\hat{\beta}_{RE} &= [(X'Q_\mu X/\sigma_\nu^2) + (X'(P_\mu - \bar{J}_{NT})X/\sigma_1^2)]^{-1}[(X'Q_\mu y/\sigma_\nu^2) + (X'(P_\mu - \bar{J}_{NT})y/\sigma_1^2)] \\ &= [W_{XX} + \phi^2 B_{XX}]^{-1}[W_{Xy} + \phi^2 B_{Xy}]\end{aligned}$$

donde $W_{XX} = X'Q_\mu X$, $B_{XX} = X'(P_\mu - \bar{J}_{NT})X$, $\phi^2 = \sigma_\nu^2/\sigma_1^2$. También $var(\hat{\beta}_{RE}) = \sigma_\nu^2[W_{XX} + \phi^2 B_{XX}]^{-1}$.

Por otro lado $\hat{\beta}_{FE} = W_{XX}^{-1}W_{Xy}$ y $\hat{\beta}_{BE} = B_{XX}^{-1}B_{Xy}$, FE: fixed-effects-within; BE: between (regresión de los promedios intra-individuo solamente). Entonces, $\hat{\beta}_{RE} = W_1\hat{\beta}_{FE} + W_2\hat{\beta}_{BE}$ donde $W_1 = [W_{XX} + \phi^2 B_{XX}]^{-1}W_{XX}$ y $W_2 = [W_{XX} + \phi^2 B_{XX}]^{-1}\phi^2 B_{XX} = I - W_1$, es decir, RE es un promedio ponderado de FE y BE.

Nota: Análisis cuando $T \to \infty$, $N \to \infty$, $(\sigma_\nu^2, \sigma_\mu^2) \to 0, \infty$.

Efectos aleatorios (random-effects,RE)-two-way

En el modelo RE two-way $\mu_i \sim iid(0, \sigma^2_\mu)$, $\lambda_t \sim iid(0, \sigma^2_\lambda)$, $\nu_{it} \sim iid(0, \sigma^2_\nu)$, $\mu_i \perp\!\!\!\perp \nu_{it}$, $\lambda_i \perp\!\!\!\perp \nu_{it}$. *Además tenemos* $X_{it} \perp\!\!\!\perp \mu_i$, $X_{it} \perp\!\!\!\perp \lambda_t$ *y* $X_{it} \perp\!\!\!\perp \nu_{it}$ *para todo* i *y* t.

$$\Omega = E(uu') = Z_\mu E(\mu\mu')Z'_\mu + Z_\mu E(\lambda\lambda')Z'_\lambda + E(\nu\nu') =$$

$$\sigma^2_\mu(I_N \otimes J_T) + \sigma^2_\lambda(J_N \otimes I_T) + \sigma^2_\nu(I_N \otimes I_T).$$

Ω tiene esta estructura:

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

Efectos aleatorios (random-effects,RE)-two-way

Más en detalles, tenemos $\Omega = E \ldots$

$$
\left[\begin{array}{cccccccc}
(\mu_1+\lambda_1+\nu_{11})^2 & (\mu_1+\lambda_1+\nu_{11})(\mu_1+\lambda_2+\nu_{12}) & \cdots & (\mu_1+\lambda_1+\nu_{11})(\mu_1+\lambda_T+\nu_{1T}) & (\mu_1+\lambda_1+\nu_{11})(\mu_2+\lambda_1+\nu_{21}) & (\mu_1+\lambda_1+\nu_{11})(\mu_2+\lambda_2+\nu_{22}) & \cdots & (\mu_1+\lambda_1+\nu_{11})(\mu_2+\lambda_T+\nu_{2T}) \\
(\mu_1+\lambda_2+\nu_{12})(\mu_1+\lambda_1+\nu_{11}) & (\mu_1+\lambda_2+\nu_{12})^2 & \cdots & (\mu_1+\lambda_2+\nu_{12})(\mu_1+\lambda_T+\nu_{1T}) & (\mu_1+\lambda_2+\nu_{12})(\mu_2+\lambda_1+\nu_{21}) & (\mu_1+\lambda_2+\nu_{12})(\mu_2+\lambda_2+\nu_{22}) & \cdots & (\mu_1+\lambda_2+\nu_{12})(\mu_2+\lambda_T+\nu_{2T}) \\
\vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots \\
(\mu_1+\lambda_T+\nu_{1T})(\mu_1+\lambda_1+\nu_{11}) & (\mu_1+\lambda_T+\nu_{1T})(\mu_1+\lambda_1+\nu_{12}) & \cdots & (\mu_1+\lambda_T+\nu_{1T})^2 & (\mu_1+\lambda_T+\nu_{1T})(\mu_2+\lambda_1+\nu_{21}) & (\mu_1+\lambda_T+\nu_{1T})(\mu_2+\lambda_2+\nu_{22}) & \cdots & (\mu_1+\lambda_T+\nu_{1T})(\mu_2+\lambda_T+\nu_{2T}) \\
(\mu_2+\lambda_1+\nu_{21})(mu_1++\lambda_1\nu_{11}) & (\mu_2+\lambda_1+\nu_{21})(\mu_1+\lambda_2+\nu_{12}) & \cdots & (\mu_2+\lambda_1+\nu_{21})(\mu_1+\lambda_T+\nu_{1T}) & (\mu_2+\lambda_1+\nu_{21})^2 & (\mu_2+\lambda_1+\nu_{21})(\mu_2+\lambda_2\nu_{22}) & \cdots & (\mu_2+\lambda_1+\nu_{21})(\mu_2+\lambda_T+\nu_{2T}) \\
(\mu_2+\lambda_2+\nu_{22})(mu_1+\lambda_1+\nu_{11}) & (\mu_2+\lambda_2+\nu_{22})(\mu_1+\lambda_2+\nu_{12}) & \cdots & (\mu_2+\lambda_2+\nu_{22})(\mu_1+\lambda_T+\nu_{1T}) & (\mu_2+\lambda_2+\nu_{22})(\mu_2+\lambda_1+\nu_{21}) & (\mu_2+\lambda_2+\nu_{22})^2 & \cdots & (\mu_2+\lambda_2+\nu_{22})(\mu_2+\lambda_T+\nu_{2T}) \\
\vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots \\
(\mu_2+\lambda_T+\nu_{2T})(mu_1+\lambda_1+\nu_{11}) & (\mu_2+\lambda_T+\nu_{2T})(\mu_1+\lambda_2+\nu_{12}) & \cdots & (\mu_2+\lambda_T+\nu_{2T})(\mu_1+\lambda_T+\nu_{1T}) & (\mu_2+\lambda_T+\nu_{2T})(\mu_2+\lambda_1+\nu_{21}) & (\mu_+\lambda_T+\nu_{2T})(\mu_+\lambda_2+\nu_{22}) & \cdots & (\mu_2+\lambda_T+\nu_{2T})^2 \\
\vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots \\
(\mu_N+\lambda_1+\nu_{N1})(mu_1+\lambda_1+\nu_{11}) & (\mu_N+\lambda_1+\nu_{N1})(\mu_1+\lambda_2+\nu_{12}) & \cdots & (\mu_N+\lambda_1+\nu_{N1})(\mu_1+\lambda_T+\nu_{1T}) & (\mu_N+\lambda_1+\nu_{N1})(\mu_2+\lambda_1+\nu_{21}) & (\mu_N+\lambda_1+\nu_{N1})(\mu_2+\lambda_2+\nu_{22}) & \cdots & (\mu_N+\lambda_1+\nu_{N1})(\mu_2+\lambda_T+\nu_{2T}) \\
(\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_+\lambda_2+\nu_{N2})(\mu_2+\lambda_1+\nu_{21}) & (\mu_N+\lambda_2+\nu_{N2})(\mu_2+\lambda_2+\nu_{22}) & \cdots & (\mu_N+\lambda_2+\nu_{N2})(\mu_2+\lambda_T+\nu_{2T}) \\
\vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots \\
(\mu_N+\lambda_T+\nu_{NT})(mu_1+\lambda_1+\nu_{11}) & (\mu_N+\lambda_T+\nu_{NT})(\mu_+\lambda_2+\nu_{12}) & \cdots & (\mu_N+\lambda_T+\nu_{NT})(\mu_1+\lambda_T+\nu_{1T}) & (\mu_N+\lambda_T+\nu_{NT})(\mu_2+\lambda_1+\nu_{21}) & (\mu_N+\lambda_T+\nu_{NT})(\mu_+\lambda_2+\nu_{22}) & \cdots & (\mu_N+\lambda_T+\nu_{NT})(\mu_2+\lambda_T+\nu_{2T}) \\
\vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots
\end{array}\right.
$$

Efectos aleatorios (random-effects,RE)-two-way

Más en detalles, tenemos $\Omega =$

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

Notar la estructura de **diagonal por bloques** (depende de como se mire).

Efectos aleatorios (random-effects,RE)-two-way

Para estimarlo vamos a tratar el problema como GLS (generalized least squares) donde se estima Ω.

- Baltagi p.37, reemplazar J_N por $N\overline{J_N}$, I_N por $E_N + \overline{J_N}$, J_T por $T\overline{J_T}$, I_T por $E_T + \overline{J_T}$,entonces

$$\Omega = \sum_{j=1}^{4} \eta_j Q_j$$

  donde $\eta_1 = \sigma_\nu^2$, $\eta_2 = T\sigma_\mu^2 + \sigma_\nu^2$, $\eta_3 = N\sigma_\lambda^2 + \sigma_\nu^2$, y $\eta_4 = T\sigma_\mu^2 + N\sigma_\nu^2 + \sigma_\nu^2$, $Q_1 = E_N \otimes E_T$, $Q_2 = E_N \otimes \overline{J_T}$, $Q_3 = \overline{J_N} \otimes E_T$, y $Q_4 = \overline{J_N} \otimes \overline{J_T}$. Esta es la descomposición espectral de Ω donde η_j son las raíces características. Cada Q_j es simétrica e idempotente.

- Entonces, surgen los siguientes estimadores

$$\hat{\eta}_j = \frac{u' Q_j u}{tr(Q_j)}, j = 1, 2, 3, 4$$

Efectos aleatorios (random-effects,RE)-two-way

El estimador RE es entonces

$$\begin{aligned}\hat{\beta}_{RE} \quad &= [(X'Q_1X/\sigma_\nu^2) + (X'Q_2X/\eta_2) + (X'Q_3X/\eta_3)]^{-1}] \\ &\quad [(X'Q_1y/\sigma_\nu^2) + (X'Q_2y/\eta_2) + (X'Q_3y/\eta_3)] \\ &= [W_{XX} + \phi_2^2 B_{XX} + \phi_3^2 C_{XX}]^{-1}[W_{Xy} + \phi_2^2 B_{Xy} + \phi_3^2 C_{Xy}]\end{aligned}$$

donde $W_{XX} = X'Q_1X$, $B_{XX} = X'Q_2X$, $C_{XX} = X'Q_3X$, $\phi_j^2 = \sigma_\nu^2/\eta_j, j = 2, 3$. También $var(\hat{\beta}_{RE}) = \sigma_\nu^2[W_{XX} + \phi_2^2 B_{XX} + \phi_3^2 C_{XX}]^{-1}$.

Entonces, $\hat{\beta}_{RE} = W_1\hat{\beta}_{FE} + W_2\hat{\beta}_{BE1} + W_3\hat{\beta}_{BE2}$ donde $W_1 = [W_{XX} + \phi_2^2 B_{XX} + \phi_3^2 C_{XX}]^{-1}W_{XX}$, $W_2 = [W_{XX} + \phi_2^2 B_{XX} + \phi_3^2 C_{XX}]^{-1}\phi_2^2 B_{XX}$, y $W_3 = [W_{XX} + \phi_2^2 B_{XX} + \phi_3^2 C_{XX}]^{-1}\phi_2^2 C_{XX}$.

[Probar que en este caso también el estimador es un promedio ponderado de tres estimadores. Análisis cuando $T \to \infty$, $N \to \infty$, $(\sigma_\nu^2, \sigma_\mu^2, \sigma_\lambda^2) \to 0, \infty$.]

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).

Contraste de Mundlak

(Hsiao, 2003, sec. 3.2; Wooldridge, 2012, sec. 10.7.3)

- Mundlak (1978) propone usar un modelo de regresión que contenga los promedios de individuos i de las variables que varían en i y t:

$$y_{it} = \beta_0 + \beta_1 x_{it} + \beta_2 z_i + \beta_3 \bar{x}_i + \mu_i + \nu_{it}$$

- Entonces un contraste de $H_0 : \beta_3 = 0$ es un contraste por la validez de RE.

Contrastes para efectos aleatorios

Breusch-Pagan (1980)

- Breusch, T.S. and Pagan, A. (1980). "The Lagrange Multiplier Test and its Applications to Model Specification in Econometrics." *Review of Economic Studies* 47(1), 239–253.
(ver Baltagi, pp-63-65)
Supongamos la función de logaritmo de verosimilitud

$$L(\beta,\theta) = constant - \frac{1}{2} log|\Omega| - \frac{1}{2} u'\Omega^{-1} u,$$

donde $u = y - X\beta$, $\theta' = (\sigma^2_\mu, \sigma^2_\delta, \sigma^2_\nu)$, y

$$\Omega = \sigma^2_\mu (I_N \otimes J_T) + \sigma^2_\delta (J_N \otimes I_T) + \sigma^2_\nu (I_N \otimes I_T).$$

Las hipótesis de interés son

- $H_0^{\sigma^2_\mu, \sigma^2_\delta} : \sigma^2_\mu = \sigma^2_\delta = 0$. Contraste por la ausencia de efectos aleatorios en ambos niveles.
- $H_0^{\sigma^2_\mu} : \sigma^2_\mu = 0$. Contraste por la ausencia de efectos aleatorios al nivel individual.
- $H_0^{\sigma^2_\delta} : \sigma^2_\delta = 0$. Contraste por la ausencia de efectos aleatorios al nivel temporal.

Contrastes para efectos aleatorios
Breusch-Pagan (1980)

La matriz de información es diagonal por bloques en cuanto a β y θ. Es implica que sólo tenemos que considerar las derivadas con respecto a θ.
Consideremos las derivadas

$$\partial L/\partial \theta_r = \frac{1}{2} tr[\Omega^{-1}\partial\Omega/\partial\theta_r] + \frac{1}{2}[u'\Omega^{-1}\partial\Omega/\partial\theta_r\Omega^{-1}u],$$

para $r = 1, 2, 3$.
La matriz de información tiene elementos

$$J_{rs}(\theta) = E[\frac{\partial^2 L}{\theta_r\theta_s}] = \frac{1}{2} tr[\Omega^{-1}\partial\Omega/\partial\theta_r\Omega^{-1}\partial\Omega/\partial\theta_s].$$

El contraste de Breusch-Pagan es un contraste LM (multiplicadores de Lagrange), entonces slo necesita estimar bajo la hipoótesis nula conjunta $\hat{\sigma}_\nu^2 = \hat{u}'\hat{u}/NT$, $\hat{u}$ los residuos OLS. El contraste es

$$LM = \hat{D}'\hat{J}^{-1}\hat{D}$$

donde D son los scores (derivadas). $LM \sim \chi_2^2$ bajo la hipótesis nula $H_0^{\sigma_\mu^2,\sigma_\delta^2}$. Por otro lado $LM = LM_1 + LM_2$ donde LM_1 y LM_2 son los contrastes LM marginales para $H_0^{\sigma_\mu^2}$ y $H_0^{\sigma_\delta^2}$, respectivamente.

¿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. En el ejemplo se representa un panel balanceado. Podra ser desbalanceado si tuviera gaps (ej., la obs. $i = 1, t = 2$ no está).

- 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, $t = -2, -1, 0, 1, ...$ o $t = 1980, 1981, 1982....$

¿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, sólo con `tsset`)
- `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?]
- `areg y x1 x2 x3, abs(id)` (no reporta las dummies)
- `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
```