
Datos de Panel
Problem Set 0
Repaso OLS, GLS & FGLS
Lectura y resumen de datos de panel en Stata

1. Considere el siguiente modelo de regresión:

$$ltotexp_i = \beta_0 + \beta_1 suppins_i + \beta_2 phylim_i + \beta_3 actlim_i + \beta_4 totchr_i + \beta_5 age_i + \beta_6 female_i + \beta_7 income_i + u_i, \quad i = 1, \dots, N$$

- Use la base de datos “mus03data.dta”, la cual contiene datos de corte transversal de gastos médicos, para estimar la ecuación por OLS usando comandos de matrices en Stata. Adicionalmente, reporte los errores estándar usuales de OLS y los estadísticos t asociados.
- Utilice el comando `regress` para verificar los resultados obtenidos.
- Implemente un test de significatividad individual para `totchr`.
- Implemente un test de significatividad conjunta para todas las variables del modelo, excluyendo el intercepto.

Solution: *Inciso (a)* Tenemos que reconstruir las expresiones usuales de OLS:

$$\begin{aligned}\hat{\beta} &= (X'X)^{-1}X'y \\ \hat{\Sigma}_{\hat{\beta}} &= s^2(X'X)^{-1} \\ s.e.(\hat{\beta}_j) &= \left(\hat{\Sigma}_{\hat{\beta}}(j, j)\right)^{\frac{1}{2}} \\ t_j &= \hat{\beta}_j / s.e.(\hat{\beta}_j)\end{aligned}$$

Opción 1: Usar el comando `matrix accum`

Opción 2: Usar el comando `mkmat`

2. En este ejercicio vamos a aprender cómo setear los datos como panel en Stata y cómo generar estadísticas descriptivas del panel. Adicionalmente, veremos cómo convertir los datos de *wide form* a *long form* y cómo generar un panel para simulaciones.
- Utilice la base `mus08psidextract.dta` y describa la base de datos de la manera usual y como un panel.
 - Utilice la base `pigweights.dta`. Los datos se encuentran en formato *wide*. Utilice el comando `reshape` para llevarlos a formato *long*. Luego, describa la base de la misma forma que en el inciso (a).
 - Genere un panel de 5000 observaciones con 10 períodos temporales y 500 unidades en el corte transversal. El panel debe estar en formato *long*. Genere observaciones de $x_{it} \sim \mathcal{N}(0, 1)$, $u_{it} \sim \mathcal{N}(0, 1)$ y además $y_{it} = 1 + x_{it} + u_{it}$. Estime por POLS.

Solution: Ver: “PS 0 - Ex3.do”.

3. Considere el siguiente modelo:

$$y_i = \beta_1 + \beta_2 x_{2i} + \beta_3 x_{3i} + u_i, \quad i = 1, \dots, N$$

$$u_i = \sqrt{\exp(-1 + 0,2 \cdot x_{2i})} \cdot \varepsilon_i, \quad i = 1, \dots, N$$

con $\beta_1 = 1$, $\beta_2 = 1$, $\beta_3 = 1$, $x_2 \sim \mathcal{N}(0, 25)$, $x_3 \sim \mathcal{N}(0, 25)$ y $\varepsilon \sim \mathcal{N}(0, 25)$ Luego, el error u es heterocedástico con una varianza condicional igual a $25 \cdot \exp(-1 + 0,2 \cdot x_2)$.

- Genere 1000 muestras de $N=10$ observaciones a partir del modelo presentado. Para cada muestra estime por OLS, GLS y FGLS los parámetros del modelo y realice un test de hipótesis para contrastar que $H_0 : \beta_3 = 1$. Reporte tamaño del test al 1 %. Adicionalmente, reporte la media, mediana y desvío estándar de las estimaciones de β_1 , β_2 y β_3 .
- Repita el punto anterior con N igual a 20, 30, 100, 200 y 500.
- Describa detalladamente las propiedades de muestra finita de FGLS de acuerdo a lo que observó de los puntos anteriores.

Solution: Sea $E[uu' | X] = \sigma_u^2 \Omega$, donde Ω es definida positiva, por lo que su inversa también lo es. Por lo tanto, es posible encontrar una matriz no singular P tal que:

$$\Omega^{-1} = P'P$$

En este caso particular, tenemos que:

$$\Omega^{-1} = \begin{bmatrix} \frac{1}{\exp(-1+0,2 \cdot x_{21})} & 0 & \dots & 0 \\ \vdots & \frac{1}{\exp(-1+0,2 \cdot x_{22})} & \dots & \vdots \\ \vdots & \dots & \ddots & \vdots \\ 0 & \dots & \dots & \frac{1}{\exp(-1+0,2 \cdot x_{2N})} \end{bmatrix}$$

Luego,

$$P = \begin{bmatrix} \frac{1}{\sqrt{\exp(-1+0,2 \cdot x_{21})}} & 0 & \dots & 0 \\ \vdots & \frac{1}{\sqrt{\exp(-1+0,2 \cdot x_{22})}} & \dots & \vdots \\ \vdots & \dots & \ddots & \vdots \\ 0 & \dots & \dots & \frac{1}{\sqrt{\exp(-1+0,2 \cdot x_{2N})}} \end{bmatrix}$$

Premultiplicamos al modelo por P obtenemos que:

$$y_* = X_* \beta + u_*$$

donde $y_* = Py$, $X_* = PX$ y $u_* = Pu$. Luego, el estimador de MCG consiste en estimar el modelo transformado por OLS:

$$\begin{aligned} \hat{\beta} &= (X_*' X_*)^{-1} X_*' y_* \\ &= (X' \Omega^{-1} X)^{-1} X' \Omega^{-1} y \end{aligned}$$

En este caso, Ω es conocida. Ahora, supongamos que no la conocemos. Entonces, podemos utilizar el método de FGLS que utiliza un estimador consistente para Ω . FGLS es asintóticamente equivalente a MCG. En este caso, el procedimiento consiste en:

- Estimar el modelo por OLS. Guardarse los residuos.
- Generar el logaritmo de los residuos al cuadrado, $\ln(\hat{u}^2)$. Regresarlos contra la variable x_2 :

$$\ln(\hat{u}_i^2) = \gamma_1 + \gamma_2 x_{2i} + \nu_i$$

Esto asegura que la varianza estimada sea positiva

- c) Generar la predicción del logaritmo de la varianza σ_i^2 . Transformarlos tomándole la exponencial para obtener $\hat{\sigma}_i^2 = \exp(\hat{\gamma}_1 + \hat{\gamma}_2 x_{2i})$
- d) Regresar $y_i/\hat{\sigma}_i$ sobre $x_i/\hat{\sigma}_i$.