Juan Menduiña

Trabajo Práctico N° 0:
Repaso OLS, GLS y FGLS. Lectura y Resumen de Datos de Panel en Stata.

Ejercicio 1.

Considerar el siguiente modelo de regresión:

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

(a) *Usar 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, reportar los errores estándar usuales de OLS y los estadísticos t asociados.*

	beta	se	t
suppins	.25564276	.04622641	5.5302312
phylim	.30205979	.05697091	5.3020003
actlim	.35600541	.06211178	5.7316894
totchr	.37582014	.01842273	20.399812
age	.00380163	.00365613	1.039797
female	-.08432753	.0455442	-1.8515536
income	.00254982	.0010194	2.5013046
_cons	6.7037374	.27675999	24.222206

(b) *Utilizar el comando regress para verificar los resultados obtenidos.*

```
      Source |       SS           df       MS      Number of obs   =     2,955
-------------+----------------------------------   F(7, 2947)      =    124.98
       Model |  1264.72124         7  180.674463   Prob > F        =    0.0000
    Residual |  4260.16814     2,947  1.44559489   R-squared       =    0.2289
-------------+----------------------------------   Adj R-squared   =    0.2271
       Total |  5524.88938     2,954  1.87030785   Root MSE        =    1.2023

------------------------------------------------------------------------------
     ltotexp | Coefficient  Std. err.      t    P>|t|     [95% conf. interval]
-------------+----------------------------------------------------------------
     suppins |   .2556428   .0462264     5.53   0.000     .1650034    .3462821
      phylim |   .3020598   .0569709     5.30   0.000      .190353    .4137666
      actlim |   .3560054   .0621118     5.73   0.000     .2342185    .4777923
      totchr |   .3758201   .0184227    20.40   0.000     .3396974    .4119429
         age |   .0038016   .0036561     1.04   0.299    -.0033672    .0109705
      female |  -.0843275   .0455442    -1.85   0.064    -.1736292    .0049741
      income |   .0025498   .0010194     2.50   0.012      .000551    .0045486
       _cons |   6.703737     .27676    24.22   0.000     6.161075      7.2464
------------------------------------------------------------------------------
```

(c) *Implementar un test de significatividad individual para totchr.*

```
(1)  totchr = 0

     F(  1,  2947) =  416.15
          Prob > F =    0.0000
```

Por lo tanto, se puede observar que, con un nivel de significancia del 1%, estos datos aportan evidencia suficiente para indicar que la variable *totchr* es estadísticamente significativa.

(d) *Implementar un test de significatividad conjunta para todas las variables del modelo, excluyendo el intercepto.*

```
(1)  suppins = 0
(2)  phylim = 0
(3)  actlim = 0
(4)  totchr = 0
(5)  age = 0
(6)  female = 0
(7)  income = 0

     F(  7,  2947) =  124.98
          Prob > F =    0.0000
```

Por lo tanto, con un nivel de significancia del 1%, estos datos aportan evidencia suficiente para indicar que las variables del modelo, en conjunto, son estadísticamente significativas.

Ejercicio 2.

En este ejercicio, se va a aprender cómo setear los datos como panel en Stata y cómo generar estadísticas descriptivas del panel. Adicionalmente, se verá cómo convertir los datos de wide form a long form y cómo generar un panel para simulaciones.

(a) *Utilizar la base "mus08psidextract.dta" y describir la base de datos de la manera usual y como un panel.*

Stata.

(b) *Utilizar la base "pigweights.dta". Los datos se encuentran en formato wide. Utilizar el comando reshape para llevarlos a formato long. Luego, describir la base de la misma forma que en el inciso (a).*

Stata.

(c) *Generar un panel de 5000 observaciones con 10 períodos temporales y 500 unidades en el corte transversal. El panel debe estar en formato long. Generar 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}*. Estimar por POLS.*

```
      Source |       SS           df       MS      Number of obs   =     5,000
-------------+----------------------------------   F(1, 4998)      =   4972.58
       Model |  5006.33835         1  5006.33835   Prob > F        =    0.0000
    Residual |  5031.92865     4,998  1.00678844   R-squared       =    0.4987
-------------+----------------------------------   Adj R-squared   =    0.4986
       Total |   10038.267     4,999  2.00805501   Root MSE        =    1.0034

------------------------------------------------------------------------------
           y | Coefficient  Std. err.      t    P>|t|     [95% conf. interval]
-------------+----------------------------------------------------------------
           x |    1.01052   .0143303    70.52   0.000     .9824264    1.038614
       _cons |    1.02376   .0141901    72.15   0.000     .9959417    1.051579
------------------------------------------------------------------------------
```

Ejercicio 3.

Considerar el siguiente modelo:

$y_i = \beta_1 + \beta_2 x_{2i} + \beta_3 x_{3i} + u_i$, *i= 1, … , N.*
$u_i = \sqrt{e^{(-1+0,2x_{2i})}}\ \varepsilon_i$, *i= 1, … , N.*

con β_1= 1, β_2= 1, β_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 un varianza condicional igual a* $25e^{(-1+0,2x_2)}$.

(a) *Generar 1000 muestras de N= 10 observaciones a partir del modelo presentado. Para cada muestra, estimar por OLS, GLS y FGLS los parámetros del modelo y realizar un test de hipótesis para contrastar que* H_0*:* β_3*= 1. Reportar tamaño del test al 1%. Adicionalmente, reportar la media, mediana y desvío estándar de las estimaciones de* β_1, β_2 *y* β_3.

(b) *Repetir el inciso anterior con N igual a 20, 30, 100, 200 y 500.*

	N_10	N_20	N_30	N_100	N_200	N_500
tam_test_1~s	.8	1.1	.8	1	.8	.8
media_b1_ols	1.0054159	1.0233711	.9913776	1.0073724	.99101667	1.0004401
mediana_b1~s	1.0432544	1.0316013	1.011929	1.000255	1.0050259	1.0013884
de_b1_ols	1.4418954	.90375325	.72169432	.39574087	.28065068	.18146336
media_b2_ols	.99360726	1.0100861	1.0053778	1.003707	.99668205	1.0001666
mediana_b2~s	1.001702	1.0022839	1.0072177	1.001808	.99718451	1.001999
de_b2_ols	.37958122	.24111343	.19670063	.11009015	.07716763	.05011949
media_b3_ols	1.0118828	.99105405	.98908491	1.0008243	.99812821	1.0005964
mediana_b3~s	1.0206553	1.0012873	.99137709	1.0031048	.99734056	1.0004818
de_b3_ols	.30943622	.18576126	.15294122	.07752204	.05484899	.03532914
tam_test_1~s	.8	1.6	.7	.8	.7	1.3
media_b1_gls	1.0280305	1.0098569	.99142743	1.0028527	.99508303	1.0002454
mediana_b1~s	1.0521936	1.0215993	.99141711	.97967514	.99639186	1.0007963
de_b1_gls	1.2535653	.78859652	.6326124	.34452127	.24216176	.15728003
media_b2_gls	1.0041153	.99966412	1.0056483	1.0009635	.99825912	1.0002415
mediana_b2~s	1.0016103	1.0011113	1.0057288	.99995628	.99866092	1.0011777
de_b2_gls	.27698551	.14838535	.10750797	.05216017	.03670627	.02233929
media_b3_gls	.99808488	.99177559	.99704105	1.0014807	.99794066	1.0011461
mediana_b3~s	1.0007759	.99349907	1.0007703	1.0030935	.99855053	1.0003986
de_b3_gls	.23968415	.13024611	.10009673	.05074781	.03363409	.02125601
tam_test_1~s	2.6	1.9	1.1	1.3	.8	1.1
media_b1_f~s	1.01992	1.01678	.9940748	1.0072039	.99260909	1.0000234
mediana_b1~s	1.0103357	1.023284	.98835871	.98473564	.99415511	.99863401
de_b1_fgls	1.3995429	.84752793	.66857393	.35431677	.24580091	.15830694
media_b2_f~s	1.0032334	1.0017987	1.0081194	1.0025778	.99772954	1.000189
mediana_b2~s	1.0093035	1.0049713	1.0057506	1.0009448	.99776992	1.0012414
de_b2_fgls	.33065581	.18431274	.13396221	.0558778	.0377256	.02264737
media_b3_f~s	1.006726	.99314471	.99556178	1.0007816	.99824046	1.0011263
mediana_b3~s	1.0098851	.99663675	.99534097	1.0025634	.99817607	1.0005233
de_b3_fgls	.29018109	.14444993	.11601262	.05232108	.03433107	.02134511

(c) *Describir, detalladamente, las propiedades de muestra finita de FGLS de acuerdo a lo que se observó de los puntos anteriores.*

Las propiedades de muestra finita de FGLS, de acuerdo a lo que se observó de los puntos anteriores, son:

- Sesgo: El estimador FGLS puede estar sesgado si el modelo subyacente no se especifica correctamente o si la estructura de correlación verdadera en los datos se impone de manera incorrecta. Sin embargo, a medida que el tamaño de muestra aumenta, este sesgo tiende a disminuir.
- Eficiencia relativa: La eficiencia relativa del estimador FGLS en comparación con otros estimadores (como el estimador OLS) puede variar dependiendo de la estructura de correlación verdadera en los datos y del modelo. Sin embargo, en algunos casos, el estimador FGLS puede proporcionar estimaciones más precisas que el estimador OLS, especialmente cuando la estructura de correlación de los errores es ignorada por las estimaciones por OLS.
- Varianza finita: La varianza del estimador FGLS depende del tamaño de muestra y de la estructura de correlación verdadera en los datos. A diferencia de las propiedades asintóticas, en muestras finitas, la varianza del estimador FGLS puede no converger a la varianza asintótica y puede ser mayor o menor dependiendo de las características específicas de los datos y del modelo.
- Robustez: El estimador FGLS puede ser más robusto que otros estimadores en presencia de violaciones de los supuestos de homocedasticidad y de correlación serial en los datos. Esto significa que el estimador FGLS puede proporcionar estimaciones más precisas, incluso cuando los supuestos clásicos no se cumplan completamente.