Juan Menduiña

Trabajo Práctico N° 1:
Modelo de Regresión Lineal.

Ejercicio 1.

Utilizar la base de datos provista "cornwell.dta".

(a) *A partir de los datos de los siete años, y utilizando los logaritmos de todas las variables, estimar un modelo por POLS que relacione la tasa de crimen con prbarr, prbconv, prbpris, avgsen y polpc y que incluya un conjunto de dummies de año.*

POLS:

```
      Source |       SS           df       MS      Number of obs   =       630
-------------+----------------------------------   F(11, 618)      =     74.49
       Model |  117.644669        11  10.6949699   Prob > F        =    0.0000
    Residual |   88.735673       618  .143585231   R-squared       =    0.5700
-------------+----------------------------------   Adj R-squared   =    0.5624
       Total |  206.380342       629  .328108652   Root MSE        =    .37893

------------------------------------------------------------------------------
     lcrmrte | Coefficient  Std. err.      t    P>|t|     [95% conf. interval]
-------------+----------------------------------------------------------------
     lprbarr |  -.7195033   .0367657   -19.57   0.000    -.7917042   -.6473024
    lprbconv |  -.5456589   .0263683   -20.69   0.000    -.5974413   -.4938765
    lprbpris |   .2475521   .0672268     3.68   0.000     .1155314    .3795728
     lavgsen |  -.0867575   .0579205    -1.50   0.135    -.2005023    .0269872
      lpolpc |   .3659886   .0300252    12.19   0.000     .3070248    .4249525
         d82 |   .0051371    .057931     0.09   0.929    -.1086284    .1189026
         d83 |   -.043503   .0576243    -0.75   0.451    -.1566662    .0696601
         d84 |  -.1087542    .057923    -1.88   0.061     -.222504    .0049957
         d85 |  -.0780454   .0583244    -1.34   0.181    -.1925835    .0364928
         d86 |  -.0420791   .0578218    -0.73   0.467      -.15563    .0714719
         d87 |  -.0270426    .056899    -0.48   0.635    -.1387815    .0846963
       _cons |  -2.082293   .2516253    -8.28   0.000    -2.576438   -1.588149
------------------------------------------------------------------------------
```

(b) *Computar los errores estándar robustos a heteroscedasticidad arbitraria y a autocorrelación serial arbitraria.*

POLS (con errores estándar robustos):

```
Linear regression                               Number of obs     =        630
                                                F(11, 89)         =      37.19
                                                Prob > F          =     0.0000
                                                R-squared         =     0.5700
                                                Root MSE          =     .37893

                                (Std. err. adjusted for 90 clusters in county)
------------------------------------------------------------------------------
             |               Robust
     lcrmrte | Coefficient  std. err.      t    P>|t|     [95% conf. interval]
-------------+----------------------------------------------------------------
     lprbarr |  -.7195033   .1095979    -6.56   0.000    -.9372719   -.5017347
    lprbconv |  -.5456589   .0704368    -7.75   0.000    -.6856152   -.4057025
    lprbpris |   .2475521   .1088453     2.27   0.025     .0312787    .4638255
     lavgsen |  -.0867575   .1130321    -0.77   0.445    -.3113499    .1378348
      lpolpc |   .3659886    .121078     3.02   0.003     .1254092    .6065681
         d82 |   .0051371   .0367296     0.14   0.889    -.0678439    .0781181
         d83 |   -.043503    .033643    -1.29   0.199    -.1103509    .0233448
         d84 |  -.1087542   .0391758    -2.78   0.007    -.1865956   -.0309127
         d85 |  -.0780454   .0385625    -2.02   0.046    -.1546683   -.0014224
         d86 |  -.0420791   .0428788    -0.98   0.329    -.1272783    .0431201
         d87 |  -.0270426   .0381447    -0.71   0.480    -.1028353    .0487502
       _cons |  -2.082293   .8647054    -2.41   0.018    -3.800445   -.3641423
------------------------------------------------------------------------------
```

(c) *Implementar un contraste de Correlación Serial.*

Stata.

Se rechaza la hipótesis nula de no correlación serial.

(d) *Implementar un contraste de Heterocedasticidad.*

Stata.

Se rechaza la hipótesis nula de homocedasticidad.

(e) *Asumir que se cumple el supuesto de exogeneidad estricta y que u_{it} sigue un proceso AR(1). Computar el estimador de FGLS siguiendo el enfoque de Prais-Winsten. Una descripción del procedimiento se puede encontrar en Wooldridge (2010), sección 7.8.6. Observación: GLS necesita exogeneidad estricta para conseguir estimadores consistentes.*

FGLS:

```
      Source |       SS           df       MS      Number of obs   =       630
-------------+----------------------------------   F(12, 618)      =   2050.54
       Model |  885.585523        12  73.7987936   Prob > F        =    0.0000
    Residual |  22.2417551       618  .035989895   R-squared       =    0.9755
-------------+----------------------------------   Adj R-squared   =    0.9750
       Total |  907.827278       630  1.44099568   Root MSE        =    .18971

------------------------------------------------------------------------------
 tilde_lcrmrte | Coefficient  Std. err.      t    P>|t|     [95% conf. interval]
---------------+----------------------------------------------------------------
 tilde_lprbarr |   -.481208   .0333124   -14.45   0.000    -.5466271   -.4157888
tilde_lprbconv |  -.3353095   .0209135   -16.03   0.000    -.3763796   -.2942395
tilde_lprbpris |  -.1624321   .0339271    -4.79   0.000    -.2290585   -.0958058
 tilde_lavgsen |  -.0203981   .0289633    -0.70   0.482    -.0772766    .0364804
  tilde_lpolpc |   .3806954   .0298461    12.76   0.000     .3220834    .4393074
     tilde_d82 |   .0120433   .0222954     0.54   0.589    -.0317405    .0558272
     tilde_d83 |  -.0721363   .0288915    -2.50   0.013    -.1288737   -.0153989
     tilde_d84 |  -.1092092   .0333946    -3.27   0.001    -.1747898   -.0436286
     tilde_d85 |  -.1018016   .0364716    -2.79   0.005    -.1734249   -.0301784
     tilde_d86 |  -.0775719   .0381852    -2.03   0.043    -.1525605   -.0025834
     tilde_d87 |  -.0395482   .0394024    -1.00   0.316    -.1169271    .0378307
    tilde_ones |  -2.027131   .2099692    -9.65   0.000    -2.439471   -1.614792
------------------------------------------------------------------------------
```

(f) *Computar los errores estándar robustos a heteroscedasticidad arbitraria y a autocorrelación serial arbitraria para el modelo con las variables transformadas del inciso previo. Sugerencia de Wooldridge: "… If we have any doubts about the homoskedasticity assumption, or whether the AR(1) assumption sufficiently captures the serial dependence, we can just apply the usual fully robust variance matrix and associated statistics to pooled OLS on the transformed variables. This allows us to probably obtain an estimator more efficient than POLS (on the original data) but also guards against the rather simple structure we imposed on Ω. Of course, failure of strict exogeneity generally causes the Prais-Winsten estimator of β to be inconsistent."*

FGLS (con errores estándar robustos):

```
Linear regression                               Number of obs     =        630
                                                F(12, 89)         =     837.00
                                                Prob > F          =     0.0000
                                                R-squared         =     0.9755
                                                Root MSE          =     .18971

                                  (Std. err. adjusted for 90 clusters in county)
--------------------------------------------------------------------------------
               |               Robust
 tilde_lcrmrte | Coefficient  std. err.      t    P>|t|     [95% conf. interval]
---------------+----------------------------------------------------------------
 tilde_lprbarr |   -.481208   .0718373    -6.70   0.000    -.6239471   -.3384689
tilde_lprbconv |  -.3353095   .0440331    -7.61   0.000    -.4228023   -.2478168
tilde_lprbpris |  -.1624321   .0500207    -3.25   0.002    -.2618222   -.0630421
 tilde_lavgsen |  -.0203981   .0258077    -0.79   0.431    -.0716774    .0308813
  tilde_lpolpc |   .3806954    .106039     3.59   0.001     .1699982    .5913925
     tilde_d82 |   .0120433    .015186     0.79   0.430    -.0181309    .0422176
     tilde_d83 |  -.0721363     .01852    -3.90   0.000    -.1089352   -.0353374
     tilde_d84 |  -.1092092   .0215974    -5.06   0.000    -.1521229   -.0662956
     tilde_d85 |  -.1018016   .0244324    -4.17   0.000    -.1503483    -.053255
     tilde_d86 |  -.0775719   .0236838    -3.28   0.002    -.1246312   -.0305126
     tilde_d87 |  -.0395482   .0252489    -1.57   0.121    -.0897173    .0106209
    tilde_ones |  -2.027131   .7465981    -2.72   0.008    -3.510606   -.5436569
--------------------------------------------------------------------------------
```

Ejercicio 2.

En este ejercicio, se examinará un modelo para el costo total de producción en la industria aeronáutica a modo de ilustrar una aplicación de un modelo heterocedástico por grupos. Considerar la siguiente función de costos:

$ln\, cost_{jt} = \beta_1 + \beta_2\, ln\, output_{jt} + \beta_3\, load\, factor_{jt} + \beta_4\, ln\, fuel\, price_{jt} + \delta_2\, Firm_2 + \delta_3\, Firm_3 + \delta_4\, Firm_4 + \delta_5\, Firm_5 + \delta_6\, Firm_6 + \varepsilon_{jt}.$

(a) *Utilizar la base de datos provista "greene97.dta", la cual contiene datos para seis compañías áreas observadas, anualmente, durante 15 años. Estimar la ecuación por POLS.*

```
      Source |       SS           df       MS      Number of obs   =        90
-------------+----------------------------------   F(8, 81)        =   3935.79
       Model |   113.74827         8  14.2185338   Prob > F        =    0.0000
    Residual |  .292622888        81  .003612628   R-squared       =    0.9974
-------------+----------------------------------   Adj R-squared   =    0.9972
       Total |  114.040893        89  1.28135835   Root MSE        =    .06011

------------------------------------------------------------------------------
          lc | Coefficient  Std. err.      t    P>|t|     [95% conf. interval]
-------------+----------------------------------------------------------------
          lq |   .9192845   .0298901    30.76   0.000     .8598126    .9787565
          lf |  -1.070396     .20169    -5.31   0.000    -1.471696   -.6690961
         lpf |   .4174918   .0151991    27.47   0.000     .3872503    .4477333
             |
          id |
          2  |  -.0412359    .025184    -1.64   0.105    -.0913441    .0088722
          3  |  -.2089211   .0427986    -4.88   0.000     -.294077   -.1237653
          4  |   .1845557   .0607527     3.04   0.003     .0636769    .3054345
          5  |   .0240547   .0799041     0.30   0.764    -.1349293    .1830387
          6  |   .0870617   .0841995     1.03   0.304     -.080469    .2545924
             |
       _cons |   9.705942    .193124    50.26   0.000     9.321686     10.0902
------------------------------------------------------------------------------
```

(b) *Ahora, asumir que, dentro de cada compañía área, se tiene que:*

Var $[\varepsilon_{jt} \mid x_{jt}] = \sigma_j^2$, *t= 1, ... , T.*

Por lo tanto, si las varianzas fueran conocidas, el estimador GLS sería:

$$\hat{\beta} = [\sum_{j=1}^{N} \frac{1}{\sigma_j^2} X_j' X_j]^{-1} \sum_{j=1}^{N} \frac{1}{\sigma_j^2} X_j' y_j,$$

donde X_j *es una matriz TxK. Sin embargo, en este caso práctico, las varianzas son desconocidas. Luego, se solicita computar el estimador de FGLS a través de los siguientes métodos:*

(i) *Estimar el modelo calculando el estimador necesario para la varianza específica de la compañía áreas a partir de los residuos de OLS, es decir,* $\hat{\sigma}_j^2 = \frac{e_j' e_j}{n_j}$.

```
      Source |       SS           df       MS      Number of obs   =        90
-------------+----------------------------------   F(8, 81)        =   5526.83
       Model |  118.222298         8  14.7777873   Prob > F        =    0.0000
    Residual |  .216579991        81  .002673827   R-squared       =    0.9982
-------------+----------------------------------   Adj R-squared   =    0.9980
       Total |  118.438878        89  1.33077391   Root MSE        =    .05171

------------------------------------------------------------------------------
          lc | Coefficient  Std. err.      t    P>|t|     [95% conf. interval]
-------------+----------------------------------------------------------------
          lq |    .925765   .0267809    34.57   0.000     .8724795    .9790506
          lf |  -1.216307   .1855858    -6.55   0.000    -1.585565   -.8470495
         lpf |   .4056077   .0125488    32.32   0.000     .3806395    .4305758
             |
          id |
          2  |   -.046026   .0237611    -1.94   0.056    -.0933031    .0012511
          3  |  -.2020985   .0361494    -5.59   0.000    -.2740246   -.1301725
          4  |   .1905462   .0551602     3.45   0.001     .0807946    .3002977
          5  |   .0371723   .0704438     0.53   0.599    -.1029887    .1773334
          6  |    .094588   .0743639     1.27   0.207    -.0533728    .2425488
             |
       _cons |   9.942316   .1622899    61.26   0.000      9.61941    10.26522
------------------------------------------------------------------------------
```

(ii) *Estimar el modelo tratándolo como una forma del modelo de heteroscedasticidad multiplicativa de Harvey (1976). Utilizar el procedimiento en dos etapas.*

```
Heteroskedastic linear regression               Number of obs     =         90
Two-step GLS estimation
                                                Wald chi2(8)      =   36250.22
                                                Prob > chi2       =     0.0000

------------------------------------------------------------------------------
          lc | Coefficient  Std. err.      z    P>|z|     [95% conf. interval]
-------------+----------------------------------------------------------------
lc           |
          lq |    .932333   .0295289    31.57   0.000     .8744574    .9902086
          lf |  -1.115165   .1991174    -5.60   0.000    -1.505428   -.7249023
         lpf |   .4086271   .0141468    28.88   0.000     .3808999    .4363543
             |
          id |
          2  |  -.0387054   .0242462    -1.60   0.110    -.0862271    .0088163
          3  |  -.1929047   .0407071    -4.74   0.000    -.2726892   -.1131203
          4  |   .2082512   .0589261     3.53   0.000     .0927583    .3237442
          5  |   .0572051   .0780464     0.73   0.464    -.0957631    .2101733
          6  |   .1207862   .0825116     1.46   0.143    -.0409335    .2825059
             |
       _cons |   9.841375   .1768449    55.65   0.000     9.494765    10.18798
-------------+----------------------------------------------------------------
lnsigma2     |
          id |
          2  |   .9333314   .8111556     1.15   0.250    -.6565043    2.523167
          3  |    .575379   .8111556     0.71   0.478    -1.014457    2.165215
          4  |    .639489   .8111556     0.79   0.430    -.9503466    2.229325
          5  |   .6042102   .8111556     0.74   0.456    -.9856255    2.194046
          6  |   .7988952   .8111556     0.98   0.325    -.7909405    2.388731
             |
       _cons |  -6.213752   .5735736   -10.83   0.000    -7.337935   -5.089568
------------------------------------------------------------------------------
Wald test of lnsigma2=0: chi2(5) = 1.56                   Prob > chi2 = 0.9063
```

(c) *Comparar los resultados obtenidos en el inciso (b).*

Los resultados obtenidos en el inciso (b) son semejantes en cuanto a valores estimados de los parámetros y a significatividad estadística.

Ejercicio 3.

Considerar la siguiente ecuación de salarios:

$$y_{jt} = \beta_0 + \beta_1 x_{jt} + u_{jt}, \; j = 1, 2, \dots, N; \; t = 1, 2 \qquad (1)$$

donde $\beta_0 = \beta_1 = 1$, $u_j \sim \mathcal{N}(0, \Omega)$, $\Omega = \begin{bmatrix} 1 & 0 \\ 0 & 4 \end{bmatrix}$ *y* $x_{jt} \sim U[1, 20]$.

Generar 1000 muestras de N= 5 observaciones de corte transversal a partir del modelo (1). Para cada muestra, estimar por FGLS los parámetros del modelo y realizar un test de hipótesis para contrastar que H_0: $\beta_1 = 1$. *Reportar tamaño del test al 1% y el poder del test cuando* $\beta_1 = 0{,}8$. *Luego, repetir el procedimiento con N= 500. ¿Se aprecia algún cambio en el tamaño y/o en el poder del test ante el incremento de N?*

```
                  N_5  N_500
  tam_test_1      2.7      1
poder_tes~08     33.2    100
```

Por lo tanto, se puede observar que, ante el incremento de N, el tamaño del test tiende al nivel de significación del 1% y el poder del test tiende al 100%.