ECONOMETRIA COD. 6089
SECONDO PROBLEM SET

Esercizio 1

Considerate il modello

$$y_i = \beta x_i + \varepsilon_i, \qquad i = 1, ..., 100$$

dove assumiamo che valgano tutte le assunzioni standard del modello di regressione lineare, tranne quella di omoschedasticità e

$$var(\varepsilon_i) = c \log x_i.$$

a) Costruite lo stimatore GLS di β per questo modello.
b) Costruite la sua matrice di varianza-covarianza.
c) Costruite il test di Breusch-Pagan per questo modello.

Esercizio 2

La seguente tabella riporta la regressione del salario, *wage*, sull'istruzione *educ*, sull'esperienza in livelli ed al quadrato e su una variabile dummy, *female* che assume valore uno in caso in cui la persona intervistata nel campione sia di sesso femminile. Sotto trovate anche le statistiche descrittive delle variabili della regressione:

$$wage = \beta_0 + \beta_1 educ + \beta_2 \exp er + \beta_3 female + \beta_4 \exp er^2 + \varepsilon$$

```
reg  wage educ exper female  expersq

      Source |       SS       df       MS              Number of obs =     526
-------------+------------------------------           F(  4,   521) =   70.17
       Model |  2506.95576     4  626.738939           Prob > F      =  0.0000
    Residual |  4653.45854   521  8.93178222           R-squared     =  0.3501
-------------+------------------------------           Adj R-squared =  0.3451
       Total |  7160.41429   525  13.6388844           Root MSE      =  2.9886

------------------------------------------------------------------------------
        wage |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
        educ |   .5562848   .0502875    11.06   0.000     .4574937    .6550759
       exper |   .2551276   .0348671     7.32   0.000     .1866302     .323625
      female |  -2.114035   .2625501    -8.05   0.000    -2.629822   -1.598248
     expersq |  -.0044396   .0007762    -5.72   0.000    -.0059645   -.0029148
       _cons |  -2.319204   .7388254    -3.14   0.002    -3.770647   -.8677608
------------------------------------------------------------------------------

sum  wage educ exper  expersq female

    Variable |       Obs        Mean    Std. Dev.       Min        Max
-------------+--------------------------------------------------------
        wage |       526    5.896103    3.693086        .53      24.98
        educ |       526    12.56274    2.769022          0         18
       exper |       526    17.01711    13.57216          1         51
     expersq |       526    473.4354    616.0448          1       2601
      female |       526    .4790875     .500038          0          1
```

a) Come interpretate il coefficiente associato ad *educ*?

b) Qual è potrebbe essere lo scopo di usare una variabile dummy tipo $female$?

c) Come interpretate il coefficiente stimato per la variabile? Scrivete i valori predetti di *wage* nel caso in cui la dummy sia uguale a 0 e nel caso sia uguale ad 1?

d) Se nel dataset ci fosse stata anche la dummy *male*, che assume valori pari ad 1 in caso di intervistato di sesso maschile, ci sarebbero stati problemi nell'includere questa dummy nella regressione precedente?

e) Come interpretate i coefficienti di *exper* ed $espersq$?

Esercizio 3

```
reg  colgpa hsperc sat

      Source |       SS       df       MS              Number of obs =    4137
-------------+------------------------------           F(  2,  4134) =  777.92
       Model |  490.606706     2  245.303353           Prob > F      =  0.0000
    Residual |  1303.58897  4134  .315333567           R-squared     =  0.2734
-------------+------------------------------           Adj R-squared =  0.2731
       Total |  1794.19567  4136  .433799728           Root MSE      =  .56155

------------------------------------------------------------------------------
      colgpa |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
      hsperc |  -.0135192   .0005495   -24.60   0.000    -.0145965    -.012442
         sat |   .0014762   .0000653    22.60   0.000     .0013482    .0016043
       _cons |   1.391757   .0715424    19.45   0.000     1.251495    1.532018
------------------------------------------------------------------------------
```

Osservate l'output della regressione:

$$\mathrm{col}\,gpa = \boldsymbol{\beta}_0 + \boldsymbol{\beta}_1 hesperc + \boldsymbol{\beta}_2 sat + \boldsymbol{\varepsilon}$$

dove col*gpa* misura su una scala da 0 a 4 i voti degli studenti di un college americano, *hesperc* esprime in termini percentili la promozione alla scuola superiore (*hesperc*=5 significa che lo studente era nel 5% dei bravi della sua classe) e *sat* rappresenta il punteggio di un test attitudinale per l'ingresso in università, in cui vengono testate le abilità matematiche e verbali.

a) Ecco l'output del test di Ramsey-Reset. Spiegate il suo utilizzo e commentate il risultato ottenuto.

```
estat ovtest

Ramsey RESET test using powers of the fitted values of colgpa
       Ho:  model has no omitted variables
                F(3, 4131) =     40.74
                  Prob > F =      0.0000
```

b) Introducendo la dummy $female$ (come visto nell'esercizio precedente), cosa possiamo capire dalla stima?

```
reg  colgpa hsperc sat female

      Source |       SS       df       MS              Number of obs =    4137
-------------+------------------------------           F(  3,  4133) =  549.93
       Model |  511.874076     3  170.624692           Prob > F      =  0.0000
    Residual |   1282.3216  4133  .310264118           R-squared     =  0.2853
-------------+------------------------------           Adj R-squared =  0.2848
       Total |  1794.19567  4136  .433799728           Root MSE      =  .55701

------------------------------------------------------------------------------
      colgpa |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
      hsperc |   -.012594   .0005564   -22.64   0.000    -.0136848   -.0115032
         sat |    .001589   .0000662    24.00   0.000     .0014593    .0017188
      female |   .1489008   .0179848     8.28   0.000     .1136409    .1841607
       _cons |   1.190771   .0750024    15.88   0.000     1.043726    1.337816
```

c) Se avessimo generato una variabile interagita tra *sat* e $female$ e svolto la seguente regressione, come avremmo interpretato la variabile dummy? Scrivete il valore predetto di colgpa quando $female = 1$ e quando $female = 0$.

```
g satf=sat*female

. reg  colgpa hsperc sat satf

      Source |       SS       df       MS              Number of obs =    4137
-------------+------------------------------           F(  3,  4133) =  550.13
       Model |  512.005618     3  170.668539           Prob > F      =  0.0000
    Residual |  1282.19006  4133   .31023229           R-squared     =  0.2854
-------------+------------------------------           Adj R-squared =  0.2848
       Total |  1794.19567  4136  .433799728           Root MSE      =  .55699

------------------------------------------------------------------------------
      colgpa |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
      hsperc |  -.0126411   .0005552   -22.77   0.000    -.0137295   -.0115527
         sat |   .0015305   .0000651    23.51   0.000     .0014028    .0016581
        satf |   .0001443   .0000174     8.31   0.000     .0001103    .0001784
       _cons |    1.25364    .072884    17.20   0.000     1.110748    1.396532
------------------------------------------------------------------------------
```

Esercizio 4

Consideriamo il seguente modello:

$$y = \beta_0 + \beta_1 x + \varepsilon$$

in formato matriciale:

$$\mathbf{y} = X\boldsymbol{\beta} + \boldsymbol{\varepsilon} \tag{1}$$

Ma il vero modello è

$$y = \beta_0 + \beta_1 x + \beta_2 z + u$$

in formato matriciale:

$$\mathbf{y} = X\boldsymbol{\beta} + Z\boldsymbol{\gamma} + \boldsymbol{\varepsilon} \tag{2}$$

a) Se stimiamo il modello 1, invece del modello 2, in che problema incorriamo. Lo stimatore OLS sarà ancora non distorto?

b) Se il modello vero fosse stato il modello 1 e quello stimato il modello 2, in che problema incorriamo. Lo stimatore OLS di $\widehat{\beta}$ sarà ancora non distorto?

Esercizio 5

Il seguente output riporta il risultato della regressione:

$$lprice = \beta_0 + \beta_1 lproptax + \varepsilon$$

dove il logaritmo del prezzo di un appartamento, *lprice*, è regredito su il logaritmo della tassa di proprietà del quartiere dove è ubicato l'appartamento, *lproptax*

```
reg  lprice  lproptax

      Source |       SS       df       MS              Number of obs =     506
-------------+------------------------------           F(  1,   504) =  224.71
       Model |  26.0826748     1  26.0826748           Prob > F      =  0.0000
    Residual |  58.4995961   504  .116070627           R-squared     =  0.3084
-------------+------------------------------           Adj R-squared =  0.3070
       Total |  84.5822709   505  .167489645           Root MSE      =  .34069

------------------------------------------------------------------------------
      lprice |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
    lproptax |  -.5733679   .0382489   -14.99   0.000    -.6485147   -.4982211
       _cons |   13.34193   .2273745    58.68   0.000     12.89522    13.78865
------------------------------------------------------------------------------

. sum  lprice lproptax

    Variable |       Obs        Mean    Std. Dev.       Min        Max
-------------+--------------------------------------------------------
      lprice |       506    9.941057     .409255   8.517193    10.8198
    lproptax |       506    5.931405    .3963666   5.231109   6.566672
sum  price proptax

    Variable |       Obs        Mean    Std. Dev.       Min        Max
-------------+--------------------------------------------------------
       price |       506    22511.51    9208.856       5000      50001
     proptax |       506    40.82372    16.85371       18.7       71.1
```

a) Commentate l'effetto di *lproptax* su *lprice*.

b) Come potete ricavare l'effetto di *proptax* su *price* a partire dai risultati riportati?

c) Qual è l'utilità di considerare variabili espresse in logaritmi?

Esercizio 6

Considerate la stessa regressione di prima, l'unica differenza è che il regressore e la variabile dipendente sono in livelli e non in logaritmi.

$$price = \beta_0 + \beta_1 proptax + \varepsilon$$

```
reg  price proptax

      Source |       SS       df       MS              Number of obs =     506
-------------+------------------------------           F(  1,   504) =  140.64
       Model |  9.3433e+09     1  9.3433e+09           Prob > F      =  0.0000
    Residual |  3.3482e+10   504  66433096.2           R-squared     =  0.2182
-------------+------------------------------           Adj R-squared =  0.2166
       Total |  4.2826e+10   505    84803032           Root MSE      =  8150.7

------------------------------------------------------------------------------
       price |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
     proptax |  -255.2159   21.52043   -11.86   0.000    -297.4967   -212.9351
       _cons |   32930.37   950.3317    34.65   0.000     31063.27    34797.47
------------------------------------------------------------------------------
```

a) Definite la pendenza e l'elasticità in questo caso, riferite al punto medio.

b) Ipotizzate di avere la variabile dipendente espressa in logaritmi e la variabile indipendente espressa in livelli:

$$lprice = \beta_0 + \beta_1 proptax + \varepsilon$$

```
reg  lprice proptax

      Source |       SS       df       MS              Number of obs =     506
-------------+------------------------------           F(  1,   504) =  229.87
       Model |   26.493825     1   26.493825           Prob > F      =  0.0000
    Residual |  58.0884459   504  .115254853           R-squared     =  0.3132
-------------+------------------------------           Adj R-squared =  0.3119
       Total |  84.5822709   505  .167489645           Root MSE      =  .33949

------------------------------------------------------------------------------
      lprice |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
     proptax |  -.0135904   .0008964   -15.16   0.000    -.0153515   -.0118293
       _cons |   10.49587   .0395834   265.16   0.000      10.4181    10.57364
------------------------------------------------------------------------------
```

Definite la pendenza e l'elasticità in questo caso, riferite al punto medio.

c) Ipotizzate di avere ora la variabile dipendente in livelli e la variabile indipentente in logaritmi:

$$price = \beta_0 + \beta_1 lproptax + \varepsilon$$

```
reg  price lproptax

      Source |       SS       df       MS              Number of obs =     506
-------------+------------------------------           F(  1,   504) =  145.39
       Model |  9.5881e+09     1  9.5881e+09           Prob > F      =  0.0000
    Residual |  3.3237e+10   504  65947363.2           R-squared     =  0.2239
-------------+------------------------------           Adj R-squared =  0.2223
       Total |  4.2826e+10   505    84803032           Root MSE      =  8120.8

------------------------------------------------------------------------------
       price |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
    lproptax |  -10993.16   911.7085   -12.06   0.000    -12784.38   -9201.946
       _cons |   87716.41    5419.75    16.18   0.000     77068.33     98364.5
------------------------------------------------------------------------------
```

Definite la pendenza e l'elasticità in questo caso, riferite al punto medio.

Esercizio 7

Consideriamo il seguente modello:

$$C_i = c_0 + c_1 Yd_i + \varepsilon_i$$

C=consumo; c_0= consumo autonomo; c_1=propensione marginale del consumo al reddito disponibile; Yd=reddito disponibile.

Sono stati intervistati N=50 consumatori, di cui n1=20 consumatori europei ed n2=30 consumatori americani.

a) Scrivete un modello in cui si tengano conto delle differenze tra i due gruppi di consumatori (sia per il consumo autonomo che per la propensione marginale del consumo al reddito), usando una variabile dummy. Usate il parametro δ come coefficiente nel caso dell'interazione con la variabile dummy.

b) Scrivete il modello nel caso $D = 0$

c) Scrivete il modello nel caso $D = 1$

d) Scrivete la statistica del Chow test nel caso in cui testiamo l'ipotesi nulla:

$$H_0 : \delta_0 = \delta_1 = 0$$

ovvero che non ci sia un cambiamento tra i due gruppi di consumatori.

e) In questo caso abbiamo i seguenti valori nella stima del modello completo e ristretto:

MODELLO COMPLETO

$$\begin{aligned} \widehat{C} &= 8 + 0.8 * Yd \\ i &= 1, ..., n_1 \end{aligned}$$

con $RSS = 200$

$$\begin{aligned} \widehat{C} &= 20 + 0.2 * Yd \\ i &= n_1 + 1, ..., N \end{aligned}$$

con $RSS = 150$
con $RSS\ totale\ del\ modello\ completo = 200 + 150 = 350$

MODELLO RISTRETTO

$$\widehat{C} = 10 + 0.4 * Yd$$

con $RSS = 500$

Calcolate la statistica F. Consideriamo un valore critico, in questo caso , pari a 3.18. Possiamo rifiutare l'ipotesi nulla?

Esercizio 8

a) Provate a scrivere il modello completo nel caso in cui vogliate spiegare la differenza nel prezzo di un biglietto aereo tra due compagnie, Alfa e Beta, considerando come regressori il costo del carburante ed il costo del servizio di catering a bordo, oltre ai costi fissi di manuntenzione che rappresentano la costante.
b) Che test potete utilizzare per testare tale differenza?
c) Considerate la stima del modello completo:

```
. reg  prezzo carburante catering

      Source |       SS       df       MS              Number of obs =    1149
-------------+------------------------------           F(  2,  1146) =  483.48
       Model |  2647166.85     2  1323583.42           Prob > F      =  0.0000
    Residual |  3137332.44  1146  2737.63738           R-squared     =  0.4576
-------------+------------------------------           Adj R-squared =  0.4567
       Total |  5784499.28  1148  5038.76244           Root MSE      =  52.322

------------------------------------------------------------------------------
      prezzo |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
  carburante |     .08822   .0029673    29.73   0.000     .0823981    .0940419
    catering |    73.7979   9.343216     7.90   0.000     55.46617    92.12962
       _cons |   42.26237   7.916568     5.34   0.000     26.72978    57.79496
------------------------------------------------------------------------------
reg  prezzo carburante catering

      Source |       SS       df       MS              Number of obs =    1149
-------------+------------------------------           F(  2,  1146) =  417.06
       Model |  2686377.72     2  1343188.86           Prob > F      =  0.0000
    Residual |  3690823.28  1146  3220.61368           R-squared     =  0.4212
-------------+------------------------------           Adj R-squared =  0.4202
       Total |  6377200.99  1148  5555.05313           Root MSE      =   56.75

------------------------------------------------------------------------------
      prezzo |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
  carburante |   .0899637   .0032647    27.56   0.000     .0835582    .0963691
    catering |   79.34504   10.20206     7.78   0.000     59.32823    99.36186
       _cons |   40.93595   8.548038     4.79   0.000     24.16439    57.70752
------------------------------------------------------------------------------
```

e la stima del modello ristretto:

```
reg  prezzo carburante catering

      Source |       SS       df       MS              Number of obs =    2298
-------------+------------------------------           F(  2,  2295) =  896.69
       Model |  5470133.56     2  2735066.78           Prob > F      =  0.0000
    Residual |   7000151.5  2295  3050.17494           R-squared     =  0.4387
-------------+------------------------------           Adj R-squared =  0.4382
       Total |  12470285.1  2297   5428.9443           Root MSE      =  55.228

------------------------------------------------------------------------------
      prezzo |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
  carburante |     .08986   .0022122    40.62   0.000      .085522    .0941981
    catering |   76.69622   6.896794    11.12   0.000     63.17162    90.22082
       _cons |   38.33582   5.831495     6.57   0.000     26.90027    49.77137
------------------------------------------------------------------------------
```

Calcolate la statistica del test precedentemente descritto, coi dati degli output.

Esercizio 9

Considerate il modello
$Y_i = X_i\beta + u_i$
$var\,(u_i) = \sigma_1^2$ per i dispari
$var\,(u_i) = \sigma_2^2$ per i pari
con σ_1^2 e σ_2^2 note.

a) Come trasformereste il modello in modo da avere errori con varianza costante?

b) Qual è la varianza dello stimatore OLS? Quale è la varianza dello stimatore OLS del modello trasformato?

Esercizio 10

Considerate il seguente modello:

$$wage = \beta_0 + \beta_1 educ + \beta_2 \exp er + \beta_3 \exp er^2 + \epsilon$$

dove $wage$ rappresenta il salario percepito da un lavoratore, $educ$ sono gli anni di istruzione, $exper$ gli anni di esperienza lavorativa ed $exper^2$ sono gli anni di esperienza al quadrato.

a) Qual è l'impatto di $\exp er$ sulla variabile $wage$?

b) Quale sarà il segno di β_3?

Ecco il modello stimato in STATA

```
reg wage educ exper expersq

      Source |       SS       df       MS              Number of obs =     526
-------------+------------------------------           F(  3,   522) =   64.11
       Model |  1927.87673     3  642.625576           Prob > F      =  0.0000
    Residual |  5232.53756   522  10.0240183           R-squared     =  0.2692
-------------+------------------------------           Adj R-squared =  0.2650
       Total |  7160.41429   525  13.6388844           Root MSE      =  3.1661

------------------------------------------------------------------------------
        wage |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
        educ |   .5953429   .0530251    11.23   0.000      .4911741    .6995118
       exper |    .268287   .0368969     7.27   0.000      .1958023    .3407717
     expersq |  -.0046123    .000822    -5.61   0.000     -.006227   -.0029975
       _cons |   -3.96489   .7521526    -5.27   0.000     -5.442508   -2.487272
------------------------------------------------------------------------------
```

c) Calcolate l'impatto del'esperienza sul salario, considerando che $exper = 1$

d) Scrivete l'ipotesi nulla per testare che il coefficiente dell'istruzione sia pari a 0.6. Scrivete l'ipotesi alternativa nel caso in cui il coefficiente è maggiore di 0.6 e nel caso in cui è diverso a 0.6. Testate il test a due code.

e) Scrivete l'ipotesi nulla in cui testate $\beta_2 + \beta_3 = 0.25$. Scrivete il test t che potete implementare in questo caso. Avete tutti i dati nella tabella di STATA?

f) Scrivete l'ipotesi nulla in cui testate $\beta_2 = -\beta_3$. Scrivete il test t che potete implementare in questo caso. Avete tutti i dati nella tabella di STATA? C'è un modo per riscrivere il modello in modo da testare $H0$ guardando direttamente alla significatività di un coefficiente di regressione?