Problem Set 4: F Tests & Functional Form

These questions are intended to help you become familiar with the principals and concepts that underlie the F tests for goodness of fit and forecasting stability and get used to estimating and interpreting models in logarithmic form

1. Read in the data set *prisons.dta* from the course website. This contains monthly observations on the number of UK prison inmates over time.

a) list the data so you can see what the raw data look like
b) draw the scatter graph (in Stata) of prison inmates over time
c) run a regression of the number of prisoners on a simple (monthly) time trend – use the variable time_a - for the period January 1990-Decmber 2000 inclusive

$$\text{Prisoners} = b_0 + b_1\text{Trend} + u$$

Interpret your results
What does the F test of goodness of fit tell you about how well the model fits the data?

d) Forecast the expected number of prisoners and construct the 95% confidence intervals for the observations beginning January 2001 and ending in December 2006

e) Do the Chow test of forecast stability to test formally whether the model forecasts well out of sample

f) Repeat parts b-e using the unemployment rate (*urate*) as the explanatory variable instead of the simple time trend

$$\text{Prisoners} = b_0 + b_1\text{Urate} + u$$

What happens? Why? (hint look at the tests for goodness of fit in each model)

2. Graph the following functional forms

$$Y = \beta_0 + \beta_1(1/X) \qquad \text{with } \beta_1>0 \text{ and with } \beta_1<0$$

If $\hat{y} = 5 - 0.5(1/X)$

Interpret the estimated impact on y of a unit increase in X when a) X=1
b) X=10

3. The following output is based on a regression of regional house prices (measured in £000) on average regional incomes, (measured in £000), over the period 1970-2000 in each of 10 regions

$$\hat{price} = 10 + 0.3 Income$$

$R^2 = 0.50$

$$\log(\hat{price}) = 9.0 + 0.5 \log(Income)$$

$R^2 = 0.60$

$$\log(\hat{price}) = 8.5 + 0.04 Income$$

$R^2 = 0.40$

$$\hat{price} = 11 + 20.0 \log(Income)$$

$R^2 = 0.70$

Interpret the effects of the estimated coefficients in all 4 cases
Graph the fitted lines

Which equations are directly comparable? Which model gives the best fit?

4. The following output is taken from a regression of the level of hourly pay, (measured in £) on years of experience

```
. reg hourpay xper if _n<201
      Source |       SS       df       MS              Number of obs =     200
-------------+------------------------------           F(  1,   198) =    4.87
       Model |  102.923238     1  102.923238           Prob > F      =  0.0285
    Residual |  4183.77445   198   21.130174           R-squared     =  0.0240
-------------+------------------------------           Adj R-squared =  0.0191
       Total |  4286.69769   199  21.5411944           Root MSE      =  4.5968
------------------------------------------------------------------------------
     hourpay |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
        xper |   .0592086   .0268275     2.21   0.028     .0063044    .1121128
       _cons |   7.388821   .6251459    11.82   0.000     6.156022    8.621619

. reg loghw xper if _n<201
      Source |       SS       df       MS              Number of obs =     200
-------------+------------------------------           F(  1,   198) =    9.61
       Model |  2.54039667     1  2.54039667           Prob > F      =  0.0022
    Residual |  52.3213262   198  .264249122           R-squared     =  0.0463
-------------+------------------------------           Adj R-squared =  0.0415
       Total |  54.8617229   199   .27568705           Root MSE      =  .51405
------------------------------------------------------------------------------
       loghw |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
        xper |   .0093021   .0030001     3.10   0.002     .0033858    .0152183
       _cons |   1.830032   .0699096    26.18   0.000      1.69217    1.967895
```

```
. reg hadj xper if _n<201
      Source |       SS       df       MS              Number of obs =     200
-------------+------------------------------           F(  1,   198) =    4.87
       Model |  1.82877085     1  1.82877085           Prob > F      =  0.0285
    Residual |  74.3385589   198  .375447267           R-squared     =  0.0240
-------------+------------------------------           Adj R-squared =  0.0191
       Total |  76.1673298   199  .382750401           Root MSE      =  .61274
------------------------------------------------------------------------------
        hadj |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
        xper |   .0078924    .003576     2.21   0.028     .0008404    .0149444
       _cons |   .9849134   .0833306    11.82   0.000     .8205841    1.149243

. reg lhadj xper if _n<201
      Source |       SS       df       MS              Number of obs =     200
-------------+------------------------------           F(  1,   198) =    9.61
       Model |  2.54039656     1  2.54039656           Prob > F      =  0.0022
    Residual |  52.3213255   198  .264249119           R-squared     =  0.0463
-------------+------------------------------           Adj R-squared =  0.0415
       Total |  54.8617221   199  .275687046           Root MSE      =  .51405
------------------------------------------------------------------------------
       lhadj |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
        xper |   .0093021   .0030001     3.10   0.002     .0033858    .0152183
       _cons |  -.1851372   .0699096    -2.65   0.009        -.323   -.0472743
```

Do the Box-Cox test of functional form in order to decide which specification fits best

5. The following output is taken from the distribution of OLS residuals

```
. su reshat1 if _n<201, det

                          Residuals
-------------------------------------------------------------
      Percentiles      Smallest
 1%    -6.323713      -6.448029
 5%    -5.227664      -6.414042
10%    -4.487434      -6.233384       Obs                 200
25%    -3.299453       -6.18045       Sum of Wgt.         200

50%    -.9010892                      Mean          -1.44e-10
                        Largest       Std. Dev.      4.585193
75%     1.951943       16.25964
90%      5.31512       17.23212       Variance       21.02399
95%     9.095832       17.61862       Skewness        1.63479
99%     17.42537       20.11699       Kurtosis        6.75094

. su reshat2 if _n<201, det

                          Residuals
-------------------------------------------------------------
      Percentiles      Smallest
 1%    -1.169948      -1.427379
 5%     -.743401      -1.216464
10%    -.6195743      -1.123431       Obs                 200
25%    -.3532593      -1.113894       Sum of Wgt.         200
```

```
50%    -.0013355                      Mean           -1.52e-09
                       Largest        Std. Dev.        .491033
75%     .3477869      1.147901
90%     .5949472      1.218912        Variance        .2411134
95%     .8182689      1.228457        Skewness        .0166853
99%     1.223685       1.29072        Kurtosis        2.992205
```

Do the Jarque-Bera test for normality in the OLS residuals for both equations.
Which model do you prefer?

6. Read in the data set gdpuk_ps4.dta from the course website

Use stata to get a scatter plot of nominal and real gdp over time. What patterns do you see in the data?

Now estimate the growth rate of nominal gdp and real gdp by choosing the appropriate functional form for your simple OLS regressions.

7. Now read in the data set ps4data.dta from the course website

Do an OLS regression of a) hourly wage on age b) the log of hourly wage on age. Interpret your results.

Now do an OLS regression of a) hourly wage on the log of age b) the log of hourly wage on the log of age. Interpret your results.