Second test in Introductory Econometrics

Robert M. Kunst

December 11, 2017

1. Conceptual issues [12 points]

   (a) Consider a multiple linear regression model that is assumed to follow the assumptions MLR.1–MLR.6:

   $$y_i = \beta_0 + \beta_1 x_{1,i} + \beta_2 x_{2,i} + u_i, \quad u_i \sim N(0, \sigma^2)$$

   Please classify the elements of this model into the three categories 'parameters','observed variables','unobserved variables', approximately as follows: 'Parameters: ...; Observed variables: ...; Unobserved variables: ...';

   (b) Which of the conditions MLR.1–MLR.6 are required for the BLUE property, i.e. which of them are the Gauss-Markov assumptions?

   (c) Suppose now that the model is really a model for time-indexed data, but that it is static. List the Gauss-Markov conditions for this type of model.

   (d) When is a regression model called dynamic?

2. The attached software output represents two regression specifications that both attempt at explaining female labor force participation (WLFP) across the U.S. states. Potential explanatory variables are YF, median earnings by females, EDUC, the percentage of female high-school graduates, UE, the unemployment rate, WH, the percentage of 'white' people, MR, the marriage rate, DR, the divorce rate, and URB, the percentage of urban population. We generally assume that the Gauss-Markov conditions hold, including the normal distribution of errors. [16 points]

   (a) Following the first regression, you see an F–test reported (the command "test" following the "reg" regression). The numerator degrees of freedom have been replaced by a question mark. What would they be for this test (a number, please)?

   (b) What is the correctly formulated null hypothesis for this hypothesis test, using our β_j notation? Can this H_0 be rejected at the 10% level? What about the 1% level? Does this result support the second regression specification or the first one?

   (c) You also see values of information criteria printed after each regression (the command "estat ic"). Which of the two specifications is supported by the AIC criterion? Which is supported by the BIC criterion?

   (d) What would be an argument for keeping the regressor variable URB, even though the more parsimonious specification appears to be generally preferred?

```
. reg wlfp yf educ ue wh dr mr urb
Source   |       SS       df       MS            Number of obs =      50
-------------+------------------------------     ***********************
Model    |  609.420767     7  87.0601096         ***********************
Residual |  222.054386    42   5.2870092         R-squared     =  0.7329
-------------+------------------------------     Adj R-squared =  0.6884
Total    |  831.475154    49  16.9688807         Root MSE      =  2.2993

------------------------------------------------------------------------------
wlfp     |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
yf       |   .0045002   .0008736     5.15   0.000     .0027373    .0062631
educ     |   .2771559   .0569967     4.86   0.000     .1621319    .3921798
ue       |  -1.121567   .2698184    -4.16   0.000    -1.666083   -.5770513
wh       |  -.1285665   .0351162    -3.66   0.001    -.1994339    -.057699
dr       |   .2284444   .1719451     1.33   0.191    -.1185548    .5754437
mr       |  -.2256207   .1539693    -1.47   0.150    -.5363434     .085102
urb      |  -.0693069   .0301506    -2.30   0.027    -.1301532   -.0084606
_cons    |    50.9947   10.43746     4.89   0.000     29.93105    72.05834
------------------------------------------------------------------------------

. test dr mr urb
( 1)  dr = 0
( 2)  mr = 0
( 3)  urb = 0

F(  ?,    42) =    2.02
Prob > F =    0.1261

. estat ic
-----------------------------------------------------------------------------
Model    |    Obs    ll(null)   ll(model)     df          AIC         BIC
-------------+---------------------------------------------------------------
.        |     50   -141.2264   -108.2194      8     232.4388     247.735
-----------------------------------------------------------------------------

. reg wlfp yf educ ue wh
Source   |       SS       df       MS            Number of obs =      50
-------------+------------------------------     ***********************
Model    |  577.431962     4  144.357991         ***********************
Residual |  254.043192    45  5.64540426         R-squared     =  0.6945
-------------+------------------------------     Adj R-squared =  0.6673
Total    |  831.475154    49  16.9688807         Root MSE      =   2.376

------------------------------------------------------------------------------
wlfp     |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
yf       |   .0044547   .0007586     5.87   0.000     .0029268    .0059826
educ     |   .2270847   .0505901     4.49   0.000      .125191    .3289785
ue       |  -1.048123   .2738994    -3.83   0.000    -1.599785   -.4964616
wh       |  -.1128279   .0351666    -3.21   0.002    -.1836571   -.0419987
_cons    |   36.48408   4.385885     8.32   0.000     27.65045     45.3177
------------------------------------------------------------------------------

. estat ic
-----------------------------------------------------------------------------
Model    |    Obs    ll(null)   ll(model)     df          AIC         BIC
-------------+---------------------------------------------------------------
.        |     50   -141.2264    -111.584      5     233.1679     242.728
-----------------------------------------------------------------------------
```

3. Heteroskedasticity. [7 points]

   (a) Suppose there is heteroskedasticity in a simple regression model $y_i = \beta_0 + \beta_1 x_i + u_i$, thus violating one Gauss-Markov condition, while all other such conditions hold. Will OLS be unbiased and consistent?

   (b) Suppose the form of this heteroskedasticity is known as $\sigma_i^2 = \gamma^2 x_i^2$. How can you use the idea of 'weighted least squares' to specify a regression that fulfills all Gauss-Markov conditions and can be estimated efficiently by OLS? Provide the dependent variable and the two regressor variables for this model.