Second test in Introductory Econometrics

Robert M. Kunst

December 7, 2016

1. Gauss-Markov etc. [12 points]

   (a) Please list the assumptions that are required in the *multiple* linear regression model for the BLUE property (linear efficiency) of the OLS estimator (Gauss-Markov conditions).

   (b) Which additional assumption is needed for the exact F– and t–distributions under the null of the F– and t–test statistics?

   (c) Which of the Gauss-Markov conditions is unimportant for the consistency property of the OLS estimator? When do we call an estimator consistent?

2. The attached software output represents two attempts at explaining cigarette consumption in a cross section of randomly selected persons across the US. Potential explanatory variables are educ, an indicator of the person's education, cigpric, the local cigarette price, white, an ethnic dummy, age, the person's age, income, the person's income, and restaurn, a dummy for smoking bans in local restaurants (this is historical data, now all restaurants ban smoking). We generally assume that the Gauss-Markov conditions hold, including the normal distribution of errors (at least approximately). [14 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 The numerator degrees of freedom have been replaced by a question mark.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 5% level? What about the 10% level? Intuitively, 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?

   (d) Following the second specification, somebody keeps the residuals from that regression and regresses them on all regressors of the first specification. Why does this somebody do such a thing? Maybe he/she wants to run another test? What is the asymptotic distribution of this test statistic under its null?

```
. reg cigs educ cigpric white age income restaurn

  Source |       SS       df       MS                  Number of obs =     807
---------+------------------------------                F(  6,   800) =    2.59
   Model |  2889.70751     6  481.617918                Prob > F      =  0.0172
Residual |  148863.975   800  186.079969                R-squared     =  0.0190
---------+------------------------------                Adj R-squared =  0.0117
   Total |  151753.683   806  188.280003                Root MSE      =  13.641

------------------------------------------------------------------------------
    cigs |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
---------+--------------------------------------------------------------------
    educ |  -.3680446   .1692771    -2.17   0.030    -.7003243   -.0357648
 cigpric |    .004311   .1026571     0.04   0.967    -.1971981       .20582
   white |  -.1123855   1.480575    -0.08   0.940    -3.018657    2.793886
     age |  -.0439658   .0287385    -1.53   0.126    -.1003775    .0124459
  income |   .0001306    .000056     2.33   0.020     .0000206    .0002405
restaurn |  -2.990576   1.137075    -2.63   0.009    -5.222579   -.7585731
   _cons |   13.14521    6.76139     1.94   0.052    -.1269484    26.41737
------------------------------------------------------------------------------

. estat ic
-----------------------------------------------------------------------------
Model |    Obs    ll(null)   ll(model)     df          AIC         BIC
------+----------------------------------------------------------------------
    . |    807   -3258.088    -3250.33      7     6514.661    6547.514
-----------------------------------------------------------------------------

. test cigpric white age

( 1)  cigpric = 0
( 2)  white = 0
( 3)  age = 0

F(  ?,   800) =    0.78
Prob > F =    0.5050

. reg cigs educ income restaurn

  Source |       SS       df       MS                  Number of obs =     807
---------+------------------------------                F(  3,   803) =    4.40
   Model |  2453.99514     3   817.99838                Prob > F      =  0.0044
Residual |  149299.688   803  185.927382                R-squared     =  0.0162
---------+------------------------------                Adj R-squared =  0.0125
   Total |  151753.683   806  188.280003                Root MSE      =  13.636

------------------------------------------------------------------------------
    cigs |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
---------+--------------------------------------------------------------------
    educ |  -.3243288   .1667578    -1.94   0.052    -.6516614    .0030037
  income |   .0001308    .000056     2.34   0.020     .0000209    .0002406
restaurn |  -2.926764    1.12031    -2.61   0.009    -5.125846   -.7276817
   _cons |   10.92859   2.056444     5.31   0.000     6.891949    14.96523
------------------------------------------------------------------------------

. estat ic
-----------------------------------------------------------------------------
Model |    Obs    ll(null)   ll(model)     df          AIC         BIC
------+----------------------------------------------------------------------
    . |    807   -3258.088    -3251.51      4     6511.019    6529.793
-----------------------------------------------------------------------------
```

3. Heteroskedasticity. [9 points]

   (a) Suppose there is heteroskedasticity in a regression model, thus violating one Gauss-Markov condition, while all other such conditions hold. Will OLS be unbiased and consistent?

   (b) Suppose there is homoskedasticity in a regression model, and moreover two regressors have a high correlation of around 0.9. Will OLS still be unbiased and consistent? Will all standard errors be correct?

   (c) Suppose there is heteroskedasticity in a regression model, all other Gauss-Markov conditions hold, and you use the robust White-Eicker standard errors. Will OLS be linear efficient (BLUE)?