Lin-log

```
gretl version 2017d
Current session: 2018-03-28 13:00

# PS6.1 for Example 6.1
? open data4-1

Read datafile C:Files-1.gdt
periodicity: 1, maxobs: 14
observations range: 1 to 14

Listing 5 variables:
  0) const     1) price     2) sqft      3) bedrms    4) baths

? logs sqft bedrms baths
Listing 8 variables:
  0) const       1) price       2) sqft        3) bedrms      4) baths
  5) l_sqft      6) l_bedrms    7) l_baths

# learn about the logs command
? help logs

logs

Argument:   varlist

The natural log of each of the series in varlist is obtained and the
result
stored in a new series with the prefix l_ ("el" underscore). For
example,
"logs x y" creates the new variables l_x = ln(x) and l_y = ln(y).

Menu path:    /Add/Logs of selected variables

# "best" linear model
? ols price 0 sqft
```

```
Model 1: OLS, using observations 1-14
Dependent variable: price

             coefficient   std. error   t-ratio   p-value
  --------------------------------------------------------
  const       52,3509      37,2855       1,404    0,1857
  sqft         0,138750     0,0187329    7,407    8,20e-06 ***

Mean dependent var   317,4929   S.D. dependent var   88,49816
Sum squared resid    18273,57   S.E. of regression   39,02304
R-squared            0,820522   Adjusted R-squared   0,805565
F(1, 12)             54,86051   P-value(F)           8,20e-06
Log-likelihood      -70,08421   Akaike criterion     144,1684
Schwarz criterion    145,4465   Hannan-Quinn         144,0501

# estimate linear-log model
? ols price 0 l_sqft l_bedrms l_baths

Model 2: OLS, using observations 1-14
Dependent variable: price

             coefficient   std. error   t-ratio    p-value
  --------------------------------------------------------
  const      -1773,14       372,290     -4,763     0,0008  ***
  l_sqft       303,709       58,6762     5,176     0,0004  ***
  l_bedrms    -143,249       91,1029    -1,572     0,1469
  l_baths       -8,65825     95,1471    -0,09100   0,9293

Mean dependent var   317,4929   S.D. dependent var   88,49816
Sum squared resid    14969,35   S.E. of regression   38,69024
R-squared            0,852975   Adjusted R-squared   0,808868
F(3, 10)             19,33856   P-value(F)           0,000174
Log-likelihood      -68,68806   Akaike criterion     145,3761
Schwarz criterion    147,9323   Hannan-Quinn         145,1395

Excluding the constant, p-value was highest for variable 7 (l_baths)

# omit l_baths
? omit l_baths
Test on Model 2:

  Null hypothesis: the regression parameter is zero for l_baths
  Test statistic: F(1, 10) = 0,00828074, p-value 0,929291
  Omitting variables improved 3 of 3 information criteria.
```

```
Model 3: OLS, using observations 1-14
Dependent variable: price

             coefficient   std. error   t-ratio   p-value
  --------------------------------------------------------
  const       −1749,97      259,141     −6,753    3,14e-05 ***
  l_sqft        299,972      39,9758     7,504    1,19e-05 ***
  l_bedrms     −145,094      84,7188    −1,713    0,1148

Mean dependent var   317,4929   S.D. dependent var   88,49816
Sum squared resid    14981,74   S.E. of regression   36,90497
R-squared            0,852853   Adjusted R-squared   0,826099
F(2, 11)             31,87767   P-value(F)           0,000026
Log-likelihood      −68,69385   Akaike criterion     143,3877
Schwarz criterion    145,3049   Hannan-Quinn         143,2102

# omit l_bedrms also
? omit l_bedrms
Test on Model 3:

  Null hypothesis: the regression parameter is zero for l_bedrms
  Test statistic: F(1, 11) = 2,93319, p-value 0,114784
  Omitting variables improved 0 of 3 information criteria.

Model 4: OLS, using observations 1-14
Dependent variable: price

             coefficient   std. error   t-ratio   p-value
  --------------------------------------------------------
  const       −1660,81      273,543     −6,071    5,57e-05 ***
  l_sqft        263,316      36,3816     7,238    1,03e-05 ***

Mean dependent var   317,4929   S.D. dependent var   88,49816
Sum squared resid    18976,68   S.E. of regression   39,76670
R-squared            0,813616   Adjusted R-squared   0,798084
F(1, 12)             52,38323   P-value(F)           0,000010
Log-likelihood      −70,34850   Akaike criterion     144,6970
Schwarz criterion    145,9751   Hannan-Quinn         144,5787
```

RESET

```
gretl version 2017d
Current session: 2018-03-29 11:13

# PS6.12 for Ramsey's RESET test - Example 6.9
? open data6-1

Read datafile C:Files-1.gdt
periodicity: 1, maxobs: 20
observations range: 1 to 20

Listing 4 variables:
  0) const       1) UNITCOST    2) OUTPUT      3) INPCOST

? ols UNITCOST const OUTPUT INPCOST

Model 1: OLS, using observations 1-20
Dependent variable: UNITCOST

             coefficient   std. error   t-ratio   p-value
  --------------------------------------------------------
  const       5,19878      0,565756      9,189    5,27e-08 ***
  OUTPUT     −0,0351553    0,00417800   −8,414    1,82e-07 ***
  INPCOST     0,0207658    0,00294897    7,042    1,98e-06 ***

Mean dependent var   4,882000   S.D. dependent var   0,774016
Sum squared resid    0,982611   S.E. of regression   0,240418
R-squared            0,913677   Adjusted R-squared   0,903521
F(2, 17)             89,96711   P-value(F)           9,06e-10
Log-likelihood       1,753976   Akaike criterion     2,492048
Schwarz criterion    5,479244   Hannan-Quinn         3,075180

# generate estimated Y and its powers
? genr yhat1 = UNITCOST-$uhat
Generated series yhat1 (ID 4)
? genr yhat2 = yhat1*yhat1
Generated series yhat2 (ID 5)
? genr yhat3 = yhat1*yhat2
Generated series yhat3 (ID 6)
? genr yhat4 = yhat2*yhat2
Generated series yhat4 (ID 7)
? list
Listing 8 variables:
  0) const       1) UNITCOST    2) OUTPUT      3) INPCOST     4) yhat1
  5) yhat2       6) yhat3       7) yhat4
```

```
# estimated model with new variables
? ols UNITCOST const OUTPUT INPCOST yhat2 yhat3 yhat4

Model 2: OLS, using observations 1-20
Dependent variable: UNITCOST

             coefficient   std. error   t-ratio   p-value
  --------------------------------------------------------
  const      224,118       626,756       0,3576   0,7260
  OUTPUT      −1,91640       5,54662    −0,3455   0,7348
  INPCOST      1,13619       3,27698     0,3467   0,7340
  yhat2      −17,6124       47,8037     −0,3684   0,7181
  yhat3        2,49635       6,38096     0,3912   0,7015
  yhat4       −0,129121      0,316727   −0,4077   0,6897

Mean dependent var   4,882000   S.D. dependent var   0,774016
Sum squared resid    0,545155   S.E. of regression   0,197331
R-squared            0,952108   Adjusted R-squared   0,935003
F(5, 14)             55,66441   P-value(F)           9,65e-09
Log-likelihood       7,645399   Akaike criterion    −3,290799
Schwarz criterion    2,683595   Hannan-Quinn        −2,124534

Excluding the constant, p-value was highest for variable 2 (OUTPUT)

? omit yhat2 yhat3 yhat4
Test on Model 2:

  Null hypothesis: the regression parameters are zero for the variables
    yhat2, yhat3, yhat4
  Test statistic: F(3, 14) = 3,74473, p-value 0,0364069
  Omitting variables improved 0 of 3 information criteria.

Model 3: OLS, using observations 1-20
Dependent variable: UNITCOST

             coefficient   std. error   t-ratio   p-value
  ---------------------------------------------------------
  const       5,19878      0,565756      9,189    5,27e-08 ***
  OUTPUT     −0,0351553    0,00417800   −8,414    1,82e-07 ***
  INPCOST     0,0207658    0,00294897    7,042    1,98e-06 ***

Mean dependent var   4,882000   S.D. dependent var   0,774016
Sum squared resid    0,982611   S.E. of regression   0,240418
R-squared            0,913677   Adjusted R-squared   0,903521
F(2, 17)             89,96711   P-value(F)           9,06e-10
```

```
Log-likelihood       1,753976   Akaike criterion     2,492048
Schwarz criterion    5,479244   Hannan-Quinn         3,075180

# Repeat example with the square of OUTPUT added to the model
? square OUTPUT
Listing 9 variables:
  0) const        1) UNITCOST     2) OUTPUT       3) INPCOST
  4) yhat1        5) yhat2        6) yhat3        7) yhat4
  8) sq_OUTPUT

? ols UNITCOST const OUTPUT INPCOST sq_OUTPUT

Model 4: OLS, using observations 1-20
Dependent variable: UNITCOST

              coefficient    std. error    t-ratio    p-value
  --------------------------------------------------------------
  const       10,5223        0,736531      14,29     1,58e-010 ***
  OUTPUT      −0,174473      0,0180692     −9,656    4,46e-08  ***
  INPCOST      0,0201678     0,00139534    14,45     1,33e-010 ***
  sq_OUTPUT    0,000894804   0,000115360    7,757    8,26e-07  ***

Mean dependent var   4,882000   S.D. dependent var   0,774016
Sum squared resid    0,206417   S.E. of regression   0,113583
R-squared            0,981866   Adjusted R-squared   0,978466
F(3, 16)             288,7748   P-value(F)           3,87e-14
Log-likelihood       17,35712   Akaike criterion    −26,71424
Schwarz criterion   −22,73132   Hannan-Quinn        −25,93674

? genr yhat5 = UNITCOST-$uhat
Generated series yhat5 (ID 9)
? genr yhat6 = yhat5*yhat5
Generated series yhat6 (ID 10)
? genr yhat7 = yhat5*yhat6
Generated series yhat7 (ID 11)
? genr yhat8 = yhat6*yhat6
Generated series yhat8 (ID 12)
? list
Listing 13 variables:
  0) const        1) UNITCOST     2) OUTPUT       3) INPCOST
  4) yhat1        5) yhat2        6) yhat3        7) yhat4
  8) sq_OUTPUT    9) yhat5       10) yhat6       11) yhat7
 12) yhat8

? ols UNITCOST const OUTPUT INPCOST sq_OUTPUT yhat6 yhat7 yhat8
```

```
Model 5: OLS, using observations 1-20
Dependent variable: UNITCOST

              coefficient    std. error    t-ratio   p-value
  ----------------------------------------------------------
  const       −99,0282       386,002       −0,2565   0,8015
  OUTPUT        1,91694        7,26511      0,2639   0,7960
  INPCOST      −0,224111       0,840193    −0,2667   0,7939
  sq_OUTPUT    −0,00980809     0,0372521   −0,2633   0,7965
  yhat6         3,10725       12,3045       0,2525   0,8046
  yhat7        −0,337696       1,59740     −0,2114   0,8359
  yhat8         0,0129103      0,0769299    0,1678   0,8693

Mean dependent var   4,882000   S.D. dependent var   0,774016
Sum squared resid    0,185731   S.E. of regression   0,119528
R-squared            0,983683   Adjusted R-squared   0,976153
F(6, 13)             130,6220   P-value(F)           7,47e-11
Log-likelihood       18,41310   Akaike criterion    −22,82621
Schwarz criterion   −15,85608   Hannan-Quinn        −21,46557

Excluding the constant, p-value was highest for variable 12 (yhat8)

? omit yhat6 yhat7 yhat8
Test on Model 5:

  Null hypothesis: the regression parameters are zero for the variables
    yhat6, yhat7, yhat8
  Test statistic: F(3, 13) = 0,482626, p-value 0,700039
  Omitting variables improved 3 of 3 information criteria.

Model 6: OLS, using observations 1-20
Dependent variable: UNITCOST

              coefficient    std. error    t-ratio    p-value
  ------------------------------------------------------------
  const       10,5223        0,736531      14,29     1,58e-010 ***
  OUTPUT      −0,174473      0,0180692     −9,656    4,46e-08  ***
  INPCOST      0,0201678     0,00139534    14,45     1,33e-010 ***
  sq_OUTPUT    0,000894804   0,000115360    7,757    8,26e-07  ***

Mean dependent var   4,882000   S.D. dependent var   0,774016
Sum squared resid    0,206417   S.E. of regression   0,113583
R-squared            0,981866   Adjusted R-squared   0,978466
F(3, 16)             288,7748   P-value(F)           3,87e-14
Log-likelihood       17,35712   Akaike criterion    −26,71424
Schwarz criterion   −22,73132   Hannan-Quinn        −25,93674
```

POLINOMIÁLIS

RESET előtte

```
RESET test for specification (squares and cubes)
Test statistic: F = 5,859717,
with p-value = P(F(2,15) > 5,85972) = 0,0132

RESET test for specification (squares only)
Test statistic: F = 11,677220,
with p-value = P(F(1,16) > 11,6772) = 0,00353

RESET test for specification (cubes only)
Test statistic: F = 11,364674,
with p-value = P(F(1,16) > 11,3647) = 0,00389
```

A modell:

```
gretl version 2017d
Current session: 2018-03-28 13:14

# PS6.2 for Example 6.2
? open data6-1

Read datafile C:Files-1.gdt
periodicity: 1, maxobs: 20
observations range: 1 to 20

Listing 4 variables:
  0) const       1) UNITCOST    2) OUTPUT      3) INPCOST

? square OUTPUT INPCOST
Listing 6 variables:
  0) const         1) UNITCOST      2) OUTPUT        3) INPCOST
  4) sq_OUTPUT     5) sq_INPCOST

# estimate models
? ols UNITCOST 0 OUTPUT sq_OUTPUT INPCOST sq_INPCOST

Model 1: OLS, using observations 1-20
Dependent variable: UNITCOST
```

```
             coefficient    std. error    t-ratio   p-value
  -----------------------------------------------------------
  const        10,3907        1,33070        7,808    1,16e-06 ***
  OUTPUT       −0,173953      0,0191452     −9,086    1,74e-07 ***
  sq_OUTPUT     0,000891366   0,000122458    7,279    2,70e-06 ***
  INPCOST       0,0221617     0,0166131      1,334    0,2021
  sq_INPCOST   −8,56983e-06   7,11345e-05   −0,1205   0,9057

Mean dependent var   4,882000   S.D. dependent var   0,774016
Sum squared resid    0,206217   S.E. of regression   0,117251
R-squared            0,981884   Adjusted R-squared   0,977053
F(4, 15)             203,2449   P-value(F)           7,21e-13
Log-likelihood       17,36679   Akaike criterion    −24,73359
Schwarz criterion   −19,75493   Hannan-Quinn        −23,76170

Excluding the constant, p-value was highest for variable 5 (sq_INPCOST)

? omit sq_INPCOST
Test on Model 1:

  Null hypothesis: the regression parameter is zero for sq_INPCOST
  Test statistic: F(1, 15) = 0,0145139, p-value 0,905707
  Omitting variables improved 3 of 3 information criteria.

Model 2: OLS, using observations 1-20
Dependent variable: UNITCOST

            coefficient    std. error    t-ratio    p-value
  -----------------------------------------------------------
  const       10,5223        0,736531      14,29     1,58e-010 ***
  OUTPUT      −0,174473      0,0180692     −9,656    4,46e-08  ***
  sq_OUTPUT    0,000894804   0,000115360    7,757    8,26e-07  ***
  INPCOST      0,0201678     0,00139534    14,45     1,33e-010 ***

Mean dependent var   4,882000   S.D. dependent var   0,774016
Sum squared resid    0,206417   S.E. of regression   0,113583
R-squared            0,981866   Adjusted R-squared   0,978466
F(3, 16)             288,7748   P-value(F)           3,87e-14
Log-likelihood       17,35712   Akaike criterion    −26,71424
Schwarz criterion   −22,73132   Hannan-Quinn        −25,93674

# retrieve regression coefficients
? genr b1 = $coeff(0)
Generated scalar b1 = 10,5223
? genr b2 = $coeff(OUTPUT)
```

```
Generated scalar b2 = -0,174473
? genr b3 = $coeff(sq_OUTPUT)
Generated scalar b3 = 0,000894804
? genr b4 = $coeff(INPCOST)
Generated scalar b4 = 0,0201678
# generate estimated averages to graph the average cost function
? genr time
Generated series time (ID 6)
? genr x = 40 + (time*5)
Generated series x (ID 7)
? genr INPCOST1 = 80
Generated scalar INPCOST1 = 80
? genr INPCOST2 = 115
Generated scalar INPCOST2 = 115
? genr INPCOST3 = 150
Generated scalar INPCOST3 = 150
? genr yhat1 = b1+(b2*x) + (b3*x*x) + (b4*INPCOST1)
Generated series yhat1 (ID 8)
? genr yhat2 = b1+(b2*x) + (b3*x*x) + (b4*INPCOST2)
Generated series yhat2 (ID 9)
? genr yhat3 = b1+(b2*x) + (b3*x*x) + (b4*INPCOST3)
Generated series yhat3 (ID 10)
? print -o x yhat1 yhat2 yhat3

                x         yhat1         yhat2         yhat3

 1             45      6,096434      6,802308      7,508181
 2             50      5,649102      6,354976      7,060849
 3             55      5,246511      5,952384      6,658258
 4             60      4,888660      5,594533      6,300407
 5             65      4,575549      5,281422      5,987296
 6             70      4,307178      5,013051      5,718925
 7             75      4,083547      4,789421      5,495294
 8             80      3,904657      4,610531      5,316404
 9             85      3,770507      4,476380      5,182254
10             90      3,681097      4,386971      5,092844
11             95      3,636427      4,342301      5,048174
12            100      3,636498      4,342371      5,048245
13            105      3,681309      4,387182      5,093055
14            110      3,770860      4,476733      5,182606
15            115      3,905151      4,611024      5,316898
16            120      4,084182      4,790055      5,495929
17            125      4,307954      5,013827      5,719700
18            130      4,576465      5,282339      5,988212
19            135      4,889717      5,595591      6,301464
20            140      5,247710      5,953583      6,659456
```

RESET utána

```
RESET test for specification (squares and cubes)
Test statistic: F = 0,762810,
with p-value = P(F(2,14) > 0,76281) = 0,485

RESET test for specification (squares only)
Test statistic: F = 0,178377,
with p-value = P(F(1,15) > 0,178377) = 0,679

RESET test for specification (cubes only)
Test statistic: F = 0,244728,
with p-value = P(F(1,15) > 0,244728) = 0,628
```

Késleltetések

```
gretl version 2017d
Current session: 2018-03-28 11:59

# PS6.5 for Example 6.4
? open data6-3

Read datafile C:Files-3.gdt
periodicity: 1, maxobs: 42
observations range: 1948 to 1989

Listing 4 variables:
  0) const    1) Year     2) Cons     3) DI

? genr C1 = Cons(-1)
Generated series C1 (ID 4)
# print C1 to see how it is stored internally
? print Cons C1

Cons:
Full data range: 1948 - 1989 (n = 42)

1858,00  1877,00  1917,00  1892,00  1887,00  1963,00  2037,00  2117,00
2126,00  2160,00  2203,00  2285,00  2354,00  2386,00  2418,00  2513,00
2573,00  2595,00  2627,00  2676,00  2738,00  2742,00  2813,00  2889,00
3062,00  3216,00  3169,00  3157,00  3171,00  3159,00  3338,00  3478,00
3477,00  3478,00  3516,00  3676,00  3727,00  3849,00  4082,00  4286,00
4584,00  4744,00

C1:
Full data range: 1948 - 1989 (n = 41)

     NA  1858,00  1877,00  1917,00  1892,00  1887,00  1963,00  2037,00
2117,00  2126,00  2160,00  2203,00  2285,00  2354,00  2386,00  2418,00
2513,00  2573,00  2595,00  2627,00  2676,00  2738,00  2742,00  2813,00
2889,00  3062,00  3216,00  3169,00  3157,00  3171,00  3159,00  3338,00
3478,00  3477,00  3478,00  3516,00  3676,00  3727,00  3849,00  4082,00
4286,00  4584,00

# generate DY(t) - DY(t-1) using the diff function
? genr DiffI = diff(DI)
Generated series DiffI (ID 5)
# print DI and DiffI to see how they are stored internally
? print DI DiffI
```

```
DI:
Full data range: 1948 - 1989 (n = 42)

1875,00  1909,00  1957,00  1932,00  1964,00  2051,00  2110,00  2201,00
2248,00  2273,00  2298,00  2402,00  2538,00  2623,00  2628,00  2724,00
2820,00  2859,00  2907,00  2934,00  2971,00  2986,00  3093,00  3117,00
3369,00  3576,00  3547,00  3564,00  3551,00  3476,00  3736,00  3947,00
4003,00  3975,00  3969,00  4076,00  4165,00  4263,00  4444,00  4589,00
4847,00  5084,00

DiffI:
Full data range: 1948 - 1989 (n = 41)

      NA   34,0000   48,0000  -25,0000   32,0000   87,0000   59,0000
 91,0000   47,0000   25,0000   25,0000   104,000   136,000   85,0000
 5,00000   96,0000   96,0000   39,0000   48,0000   27,0000   37,0000
 15,0000   107,000   24,0000   252,000   207,000  -29,0000   17,0000
-13,0000  -75,0000   260,000   211,000   56,0000  -28,0000  -6,00000
 107,000   89,0000   98,0000   181,000   145,000   258,000   237,000

# note that ols starts in 1949 because C1 and DiffI are
# undefined for 1948
? ols Cons 0 C1 DiffI

Model 1: OLS, using observations 1949-1989 (T = 41)
Dependent variable: Cons

             coefficient   std. error    t-ratio    p-value
  ----------------------------------------------------------
  const      −46,8020      22,5661        −2,074   0,0449    **
  C1           1,02188      0,00830664   123,0     4,85e-051 ***
  DiffI        0,705785     0,0710598      9,932   4,12e-012 ***

Mean dependent var   2901,390   S.D. dependent var   764,0013
Sum squared resid    45965,44   S.E. of regression   34,77955
R-squared            0,998031   Adjusted R-squared   0,997928
F(2, 38)             9631,956   P-value(F)           3,89e-52
Log-likelihood      −202,1290   Akaike criterion     410,2580
Schwarz criterion    415,3987   Hannan-Quinn         412,1299
rho                  0,208728   Durbin-Watson        1,535694
```

KÉSLELTETÉS 2.

```
gretl version 2017d
Current session: 2018-03-28 12:28

#  PS6.6 for Application Section 6.7
? open data3-3

Read datafile C:Files-3.gdt
periodicity: 1, maxobs: 34
observations range: 1960 to 1993

Listing 4 variables:
  0) const      1) YEAR       2) PATENTS    3) R_D

? ols PATENTS 0 R_D

Model 1: OLS, using observations 1960-1993 (T = 34)
Dependent variable: PATENTS

             coefficient   std. error   t-ratio    p-value
  ---------------------------------------------------------
  const       34,5711      6,35787       5,438    5,56e-06  ***
  R_D          0,791935    0,0567036    13,97     3,64e-015 ***

Mean dependent var   119,2382   S.D. dependent var   29,30583
Sum squared resid    3994,300   S.E. of regression   11,17237
R-squared            0,859065   Adjusted R-squared   0,854661
F(1, 32)             195,0551   P-value(F)           3,64e-15
Log-likelihood      -129,2704   Akaike criterion     262,5408
Schwarz criterion    265,5935   Hannan-Quinn         263,5818
rho                  0,945182   Durbin-Watson        0,233951

# generate lagged variables
? genr R_D1 = R_D(-1)
Generated series R_D1 (ID 4)
? genr R_D2 = R_D(-2)
Generated series R_D2 (ID 5)
? genr R_D3 = R_D(-3)
Generated series R_D3 (ID 6)
? genr R_D4 = R_D(-4)
Generated series R_D4 (ID 7)
# print them to see that missing observations are denoted by -999
? print R_D R_D1 R_D2 R_D3 R_D4
```

```
R_D:
Full data range: 1960 - 1993 (n = 34)

57,9400  60,5900  64,4400  70,6600  76,8300  80,0000  84,8200  86,8400
88,8100  88,2800  85,2900  83,1800  85,0700  86,7200  85,4500  83,4100
87,4400  90,1100  94,5000  99,2800  103,640  108,770  113,960  121,720
133,330  144,780  148,390  150,900  154,360  157,190  161,860  164,540
166,700  165,200

R_D1:
Full data range: 1960 - 1993 (n = 33)

     NA  57,9400  60,5900  64,4400  70,6600  76,8300  80,0000  84,8200
86,8400  88,8100  88,2800  85,2900  83,1800  85,0700  86,7200  85,4500
83,4100  87,4400  90,1100  94,5000  99,2800  103,640  108,770  113,960
121,720  133,330  144,780  148,390  150,900  154,360  157,190  161,860
164,540  166,700

R_D2:
Full data range: 1960 - 1993 (n = 32)

     NA       NA  57,9400  60,5900  64,4400  70,6600  76,8300  80,0000
84,8200  86,8400  88,8100  88,2800  85,2900  83,1800  85,0700  86,7200
85,4500  83,4100  87,4400  90,1100  94,5000  99,2800  103,640  108,770
113,960  121,720  133,330  144,780  148,390  150,900  154,360  157,190
161,860  164,540

R_D3:
Full data range: 1960 - 1993 (n = 31)

     NA       NA       NA  57,9400  60,5900  64,4400  70,6600  76,8300
80,0000  84,8200  86,8400  88,8100  88,2800  85,2900  83,1800  85,0700
86,7200  85,4500  83,4100  87,4400  90,1100  94,5000  99,2800  103,640
108,770  113,960  121,720  133,330  144,780  148,390  150,900  154,360
157,190  161,860

R_D4:
Full data range: 1960 - 1993 (n = 30)

     NA       NA       NA       NA  57,9400  60,5900  64,4400  70,6600
76,8300  80,0000  84,8200  86,8400  88,8100  88,2800  85,2900  83,1800
85,0700  86,7200  85,4500  83,4100  87,4400  90,1100  94,5000  99,2800
103,640  108,770  113,960  121,720  133,330  144,780  148,390  150,900
154,360  157,190
```

```
# generate square terms
? square R_D R_D1 R_D2 R_D3 R_D4
Listing 13 variables:
  0) const      1) YEAR       2) PATENTS    3) R_D        4) R_D1
  5) R_D2       6) R_D3       7) R_D4       8) sq_R_D     9) sq_R_D1
 10) sq_R_D2   11) sq_R_D3   12) sq_R_D4

# reset sample range because lagged variables are undefined for the
# period 1960-63
? smpl 1964 1993
Full data range: 1960 - 1993 (n = 34)
Current sample: 1964 - 1993 (n = 30)

# estimate the most general urestricted model
? ols PATENTS const R_D R_D1 R_D2 R_D3 R_D4 sq_R_D sq_R_D1 sq_R_D2
sq_R_D3 \par  sq_R_D4

Model 2: OLS, using observations 1964-1993 (T = 30)
Dependent variable: PATENTS

             coefficient    std. error     t-ratio     p-value
  -------------------------------------------------------------
  const      85,3526        22,1027        3,862       0,0011  ***
  R_D        -0,0476812      1,12511      -0,04238     0,9666
  R_D1        0,603316       2,05619       0,2934      0,7724
  R_D2        0,000179397    2,18503       8,210e-05   0,9999
  R_D3       -0,586882       2,05219      -0,2860      0,7780
  R_D4       -0,183709       1,09938      -0,1671      0,8691
  sq_R_D     -0,000732587    0,00489771   -0,1496      0,8827
  sq_R_D1    -0,00175375     0,00890019   -0,1970      0,8459
  sq_R_D2     0,00173631     0,00982210    0,1768      0,8616
  sq_R_D3    -0,000756426    0,00923801   -0,08188     0,9356
  sq_R_D4     0,00714386     0,00508520    1,405       0,1762

Mean dependent var   123,3300   S.D. dependent var   28,79514
Sum squared resid    223,3789   S.E. of regression   3,428817
R-squared            0,990710   Adjusted R-squared   0,985821
F(10, 19)            202,6257   P-value(F)           4,08e-17
Log-likelihood      -72,68324   Akaike criterion     167,3665
Schwarz criterion    182,7796   Hannan-Quinn         172,2973
rho                  0,101051   Durbin-Watson        1,797425

Excluding the constant, p-value was highest for variable 5 (R_D2)

# omit variables with p-values above 0.9
? omit 3 5 11
```

```
Test on Model 2:

  Null hypothesis: the regression parameters are zero for the variables
    R_D, R_D2, sq_R_D3
  Test statistic: F(3, 19) = 0,00695718, p-value 0,999173
  Omitting variables improved 3 of 3 information criteria.

Model 3: OLS, using observations 1964-1993 (T = 30)
Dependent variable: PATENTS

             coefficient    std. error    t-ratio   p-value
  ---------------------------------------------------------
  const      84,8409        19,0579        4,452    0,0002  ***
  R_D1        0,604292       0,635069      0,9515   0,3517
  R_D3       −0,735232       0,523331     −1,405    0,1740
  R_D4       −0,0744638      0,513419     −0,1450   0,8860
  sq_R_D     −0,000949134    0,00115128   −0,8244   0,4186
  sq_R_D1    −0,00169303     0,00341377   −0,4959   0,6249
  sq_R_D2     0,00162815     0,00253810    0,6415   0,5278
  sq_R_D4     0,00660987     0,00196462    3,364    0,0028  ***

Mean dependent var   123,3300   S.D. dependent var   28,79514
Sum squared resid    223,6243   S.E. of regression   3,188219
R-squared            0,990700   Adjusted R-squared   0,987741
F(7, 22)             334,7991   P-value(F)           7,67e-21
Log-likelihood      −72,69970   Akaike criterion     161,3994
Schwarz criterion    172,6090   Hannan-Quinn         164,9854
rho                  0,094485   Durbin-Watson        1,810383

Excluding the constant, p-value was highest for variable 7 (R_D4)

# omit variables with p-values above 0.5
? omit 7 9 10
Test on Model 3:

  Null hypothesis: the regression parameters are zero for the variables
    R_D4, sq_R_D1, sq_R_D2
  Test statistic: F(3, 22) = 0,324242, p-value 0,807788
  Omitting variables improved 3 of 3 information criteria.
```

```
Model 4: OLS, using observations 1964-1993 (T = 30)
Dependent variable: PATENTS

             coefficient    std. error    t-ratio    p-value
  ----------------------------------------------------------
  const      82,8545       12,0355         6,884    3,24e-07  ***
  R_D1        0,477055      0,327782       1,455    0,1580
  R_D3       −0,637010      0,238843      −2,667    0,0132    **
  sq_R_D     −0,00114637    0,000999965   −1,146    0,2625
  sq_R_D4     0,00651885    0,000678426    9,609    7,13e-010 ***

Mean dependent var   123,3300   S.D. dependent var   28,79514
Sum squared resid    233,5118   S.E. of regression   3,056218
R-squared            0,990289   Adjusted R-squared   0,988735
F(4, 25)             637,3376   P-value(F)           9,28e-25
Log-likelihood      −73,34868   Akaike criterion     156,6974
Schwarz criterion    163,7034   Hannan-Quinn         158,9386
rho                  0,077923   Durbin-Watson        1,843677

Excluding the constant, p-value was highest for variable 8 (sq_R_D)

# omit one at a time
? omit 8
Test on Model 4:

  Null hypothesis: the regression parameter is zero for sq_R_D
  Test statistic: F(1, 25) = 1,31426, p-value 0,262479
  Omitting variables improved 3 of 3 information criteria.

Model 5: OLS, using observations 1964-1993 (T = 30)
Dependent variable: PATENTS

             coefficient   std. error    t-ratio    p-value
  ---------------------------------------------------------
  const      94,2474       6,83005       13,80     1,79e-013 ***
  R_D1        0,120004     0,102788       1,167    0,2536
  R_D3       −0,481437     0,197731      −2,435    0,0221    **
  sq_R_D4     0,00614607   0,000599009   10,26     1,24e-010 ***

Mean dependent var   123,3300   S.D. dependent var   28,79514
Sum squared resid    245,7876   S.E. of regression   3,074633
R-squared            0,989778   Adjusted R-squared   0,988599
F(3, 26)             839,2019   P-value(F)           5,54e-26
Log-likelihood      −74,11721   Akaike criterion     156,2344
Schwarz criterion    161,8392   Hannan-Quinn         158,0274
rho                  0,131734   Durbin-Watson        1,728149
```

```
Excluding the constant, p-value was highest for variable 4 (R_D1)

? omit 4
Test on Model 5:

  Null hypothesis: the regression parameter is zero for R_D1
  Test statistic: F(1, 26) = 1,36302, p-value 0,253611
  Omitting variables improved 3 of 3 information criteria.

Model 6: OLS, using observations 1964-1993 (T = 30)
Dependent variable: PATENTS

             coefficient   std. error    t-ratio    p-value
  ---------------------------------------------------------
  const      91,3464       6,40463       14,26     4,33e-014 ***
  R_D3       -0,295068     0,117463      -2,512    0,0183    **
  sq_R_D4     0,00585565   0,000548564   10,67     3,44e-011 ***

Mean dependent var   123,3300   S.D. dependent var   28,79514
Sum squared resid    258,6727   S.E. of regression   3,095233
R-squared            0,989242   Adjusted R-squared   0,988446
F(2, 27)             1241,430   P-value(F)           2,68e-27
Log-likelihood      -74,88365   Akaike criterion     155,7673
Schwarz criterion    159,9709   Hannan-Quinn         157,1121
rho                  0,165750   Durbin-Watson        1,665191

# compute predicted values
? genr pred = PATENTS - $uhat
Generated series pred (ID 13)
# save residuals
? genr error = $uhat
Generated series error (ID 14)
# compute absolute % error
? genr abspcerr = 100*abs(error)/PATENTS
Generated series abspcerr (ID 15)
# print values as in a table with the -o flag
? print -o R_D PATENTS pred error abspcerr

              R_D      PATENTS         pred        error     abspcerr

1964        76,83         93,2      93,1259     0,074083     0,079488
1965        80,00        100,4      93,8292     6,570807     6,544629
1966        84,82         93,5      94,8126    -1,312580     1,403829
1967        86,84         93,0      97,9126    -4,912644     5,282413
1968        88,81         98,7     102,3060    -3,605998     3,653494
```

1969	88,28	104,4	103,7949	0,605085	0,579583
1970	85,29	109,4	107,8508	1,549204	1,416092
1971	83,18	111,1	109,3000	1,800022	1,620182
1972	85,07	105,3	111,4826	-6,182597	5,871412
1973	86,72	109,6	111,8153	-2,215251	2,021214
1974	85,45	107,4	109,3989	-1,998912	1,861185
1975	83,41	108,0	106,7597	1,240283	1,148410
1976	87,44	110,0	108,1349	1,865092	1,695538
1977	90,11	109,0	110,1694	-1,169448	1,072888
1978	94,50	109,3	109,4910	-0,191014	0,174761
1979	99,28	108,9	106,2848	2,615230	2,401497
1980	103,64	113,0	109,5287	3,471297	3,071944
1981	108,77	114,5	111,0093	3,490723	3,048666
1982	113,96	118,4	114,3445	4,055506	3,425258
1983	121,72	112,4	118,4819	-6,081902	5,410945
1984	133,33	120,6	122,1489	-1,548881	1,284313
1985	144,78	127,1	126,9982	0,101834	0,080121
1986	148,39	133,0	131,4774	1,522611	1,144820
1987	150,90	139,8	138,7609	1,039085	0,743265
1988	154,36	151,9	152,7217	-0,821732	0,540969
1989	157,19	166,3	170,3030	-4,003030	2,407114
1990	161,86	176,7	175,7597	0,940300	0,532145
1991	164,54	178,4	179,1376	-0,737635	0,413472
1992	166,70	187,2	184,4873	2,712672	1,449077
1993	165,20	189,4	188,2722	1,127791	0,595455

DUMMY

```
gretl version 2017d
Current session: 2018-03-28 14:06

# PS7.3, to duplicate Table 7.2
? open data7-3

Read datafile C:Files-3.gdt
periodicity: 1, maxobs: 14
observations range: 1 to 14

Listing 8 variables:
  0) const      1) price      2) sqft       3) bedrms     4) baths
  5) pool       6) famroom    7) firepl

? ols price const sqft

Model 1: OLS, using observations 1-14
Dependent variable: price

             coefficient   std. error   t-ratio   p-value
  --------------------------------------------------------
  const       52,3509      37,2855       1,404    0,1857
  sqft         0,138750     0,0187329    7,407    8,20e-06 ***

Mean dependent var   317,4929   S.D. dependent var   88,49816
Sum squared resid    18273,57   S.E. of regression   39,02304
R-squared            0,820522   Adjusted R-squared   0,805565
F(1, 12)             54,86051   P-value(F)           8,20e-06
Log-likelihood      −70,08421   Akaike criterion     144,1684
Schwarz criterion    145,4465   Hannan-Quinn         144,0501
```

```
# This is model E in Table 7.2
? ols price const sqft bedrms baths pool famroom firepl

Model 2: OLS, using observations 1-14
Dependent variable: price

             coefficient   std. error    t-ratio    p-value
  ---------------------------------------------------------
  const       39,0571      89,5397       0,4362     0,6758
  sqft         0,146551     0,0301014    4,869      0,0018  ***
  bedrms      -7,04553     28,7363      -0,2452     0,8134
  baths       -0,263691    41,4547      -0,006361   0,9951
  pool        53,1958      22,0635       2,411      0,0467  **
  famroom    -21,3447      42,8734      -0,4979     0,6338
  firepl      26,1880      53,8454       0,4864     0,6416

Mean dependent var   317,4929   S.D. dependent var   88,49816
Sum squared resid    9010,244   S.E. of regression   35,87726
R-squared            0,911504   Adjusted R-squared   0,835650
F(6, 7)              12,01657   P-value(F)           0,002213
Log-likelihood      -65,13456   Akaike criterion     144,2691
Schwarz criterion    148,7425   Hannan-Quinn         143,8550

Excluding the constant, p-value was highest for variable 4 (baths)

# omit variable with highest pvalue, one at a time
? omit baths
Test on Model 2:

  Null hypothesis: the regression parameter is zero for baths
  Test statistic: F(1, 7) = 4,04618e-005, p-value 0,995102
  Omitting variables improved 3 of 3 information criteria.

Model 3: OLS, using observations 1-14
Dependent variable: price

             coefficient   std. error   t-ratio   p-value
  --------------------------------------------------------
  const       38,9067      80,7830       0,4816   0,6430
  sqft         0,146419     0,0203633    7,190    9,33e-05 ***
  bedrms      -7,11121     25,0850      -0,2835   0,7840
  pool        53,2086      20,5529       2,589    0,0322   **
  famroom    -21,4518      36,8807      -0,5817   0,5768
  firepl      26,3052      47,3275       0,5558   0,5935
```

```
Mean dependent var   317,4929   S.D. dependent var   88,49816
Sum squared resid    9010,296   S.E. of regression   33,56020
R-squared            0,911503   Adjusted R-squared   0,856193
F(5, 8)              16,47976   P-value(F)           0,000496
Log-likelihood      -65,13460   Akaike criterion     142,2692
Schwarz criterion    146,1035   Hannan-Quinn         141,9143

Excluding the constant, p-value was highest for variable 3 (bedrms)

? omit bedrms
Test on Model 3:

  Null hypothesis: the regression parameter is zero for bedrms
  Test statistic: F(1, 8) = 0,0803636, p-value 0,784003
  Omitting variables improved 3 of 3 information criteria.

Model 4: OLS, using observations 1-14
Dependent variable: price

             coefficient   std. error   t-ratio   p-value
  --------------------------------------------------------
  const       18,2693      33,1797       0,5506   0,5953
  sqft         0,144969     0,0186760    7,762    2,81e-05 ***
  pool        55,1793      18,3267       3,011    0,0147   **
  famroom    -19,7763      34,4940      -0,5733   0,5805
  firepl      21,0357      41,2408       0,5101   0,6223

Mean dependent var   317,4929   S.D. dependent var   88,49816
Sum squared resid    9100,809   S.E. of regression   31,79939
R-squared            0,910614   Adjusted R-squared   0,870887
F(4, 9)              22,92180   P-value(F)           0,000097
Log-likelihood      -65,20457   Akaike criterion     140,4091
Schwarz criterion    143,6044   Hannan-Quinn         140,1134

Excluding the constant, p-value was highest for variable 7 (firepl)

? omit firepl
Test on Model 4:

  Null hypothesis: the regression parameter is zero for firepl
  Test statistic: F(1, 9) = 0,260173, p-value 0,622275
  Omitting variables improved 3 of 3 information criteria.

Model 5: OLS, using observations 1-14
Dependent variable: price
```

```
             coefficient   std. error   t-ratio   p-value
  --------------------------------------------------------
  const       22,7361      30,7966       0,7383   0,4773
  sqft         0,147290     0,0174302    8,450    7,27e-06 ***
  pool        53,3693      17,3020       3,085    0,0116   **
  famroom     -7,33610     23,4731      -0,3125   0,7611

Mean dependent var   317,4929   S.D. dependent var   88,49816
Sum squared resid    9363,895   S.E. of regression   30,60048
R-squared            0,908030   Adjusted R-squared   0,880439
F(3, 10)             32,91049   P-value(F)           0,000017
Log-likelihood      -65,40405   Akaike criterion     138,8081
Schwarz criterion    141,3643   Hannan-Quinn         138,5715

Excluding the constant, p-value was highest for variable 6 (famroom)

# Model F
? omit famroom
Test on Model 5:

  Null hypothesis: the regression parameter is zero for famroom
  Test statistic: F(1, 10) = 0,0976769, p-value 0,761055
  Omitting variables improved 3 of 3 information criteria.

Model 6: OLS, using observations 1-14
Dependent variable: price

             coefficient   std. error   t-ratio   p-value
  --------------------------------------------------------
  const       22,6728      29,5058       0,7684   0,4584
  sqft         0,144415     0,0141849   10,18     6,18e-07 ***
  pool        52,7898      16,4817       3,203    0,0084   ***

Mean dependent var   317,4929   S.D. dependent var   88,49816
Sum squared resid    9455,359   S.E. of regression   29,31856
R-squared            0,907132   Adjusted R-squared   0,890247
F(2, 11)             53,72383   P-value(F)           2,11e-06
Log-likelihood      -65,47210   Akaike criterion     136,9442
Schwarz criterion    138,8614   Hannan-Quinn         136,7667

# rerun full model for Wald test
? ols price const sqft bedrms baths pool famroom firepl

Model 7: OLS, using observations 1-14
Dependent variable: price
```

```
             coefficient   std. error    t-ratio    p-value
  ---------------------------------------------------------
  const       39,0571      89,5397       0,4362     0,6758
  sqft         0,146551     0,0301014    4,869      0,0018  ***
  bedrms      -7,04553     28,7363      -0,2452     0,8134
  baths       -0,263691    41,4547      -0,006361   0,9951
  pool        53,1958      22,0635       2,411      0,0467  **
  famroom    -21,3447      42,8734      -0,4979     0,6338
  firepl      26,1880      53,8454       0,4864     0,6416

Mean dependent var   317,4929   S.D. dependent var   88,49816
Sum squared resid    9010,244   S.E. of regression   35,87726
R-squared            0,911504   Adjusted R-squared   0,835650
F(6, 7)              12,01657   P-value(F)           0,002213
Log-likelihood      -65,13456   Akaike criterion     144,2691
Schwarz criterion    148,7425   Hannan-Quinn         143,8550

Excluding the constant, p-value was highest for variable 4 (baths)

# Wald F-statistic for Models E and F is provided
? omit bedrms baths famroom firepl
Test on Model 7:

  Null hypothesis: the regression parameters are zero for the variables
    bedrms, baths, famroom, firepl
  Test statistic: F(4, 7) = 0,0864517, p-value 0,983881
  Omitting variables improved 3 of 3 information criteria.

Model 8: OLS, using observations 1-14
Dependent variable: price

             coefficient   std. error   t-ratio   p-value
  --------------------------------------------------------
  const       22,6728      29,5058       0,7684   0,4584
  sqft         0,144415     0,0141849   10,18     6,18e-07 ***
  pool        52,7898      16,4817       3,203    0,0084   ***

Mean dependent var   317,4929   S.D. dependent var   88,49816
Sum squared resid    9455,359   S.E. of regression   29,31856
R-squared            0,907132   Adjusted R-squared   0,890247
F(2, 11)             53,72383   P-value(F)           2,11e-06
Log-likelihood      -65,47210   Akaike criterion     136,9442
Schwarz criterion    138,8614   Hannan-Quinn         136,7667
```