Quantitative Methods in Finance

Tutorial, Part 3:
Calculation and interpretation of the parameters of an estimated regression model.

Example 1: In the example from the previous tutorial, we were estimating, based on the quantities stated below for 10 observations, the regression coefficients in a linear regression model of the type $y_i = \beta_1 + \beta_2 x_{2i} + \beta_3 x_{3i} + u_i$:

$$\overline{y} = 2.4; \quad \overline{x}_2 = 0.4; \quad \overline{x}_3 = 0.8; \quad \sum y_i^2 = 184; \quad \sum x_{2i}x_{3i} = -2;$$
$$\sum x_{2i}^2 = 10; \quad \sum x_{3i}^2 = 12; \quad \sum y_i x_{2i} = 40; \quad \sum y_i x_{3i} = -2.$$

Find out whether the regression coefficients are statistically significant. Calculate the relevant measures of adequacy (reliability) of the sample regression model, and establish whether the model as a whole explains the variability of our dependent variable satisfactorily.

Write down the estimated regression model with all the parameters calculated above. Also, calculate the elasticity estimates, and explain the differences between regression coefficients and elasticities for a linear regression model.

■

Example 2: We decided to analyse the housing market in a particular municipality. For this purpose, we gathered data on a sample of 30 apartments that are for sale (the data are provided in Stata Data file `housing.dta`, while the programming code is given in Stata Do file `housing-commands.do`). We have data available on the price of apartment (*PRICE*; in mil. monetary units), size (*M2*; in square meters), age (*AGE*; in years), and distance of apartment from the city centre (*DIS*; in kilometres).

Estimate the linear multiple regression model, where you investigate the effects of the above three explanatory variables on the price of apartment, and interpret the results. Also, calculate the fitted values and the residuals of the model, and write them out together with the observed values of the dependent variable.

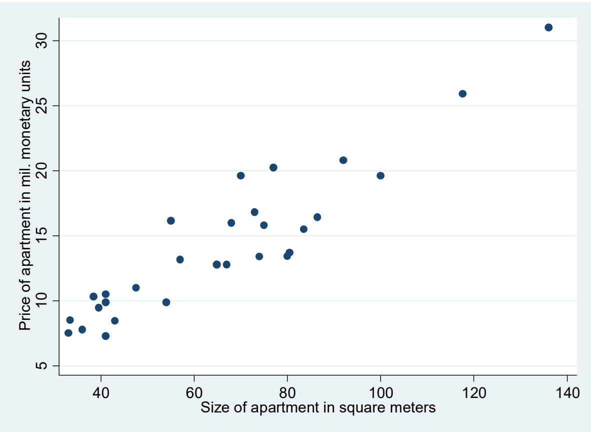
Computer printout of the results in Stata:

```
. describe

Contains data from stanovanje.dta
  obs:            30
 vars:             4                          DD Mmm YYYY HH:MM
 size:           330
```

```
-----------------------------------------------------------------------------------
              storage  display  value
variable name   type   format   label   variable label
-----------------------------------------------------------------------------------
price           float  %8.0g            Price of apartment in mil. monetary units
m2              float  %8.0g            Size of apartment in square meters
dis             byte   %8.0g            Distance of apartment from the city centre
age             byte   %8.0g            Age of apartment in years
-----------------------------------------------------------------------------------
Sorted by:
```

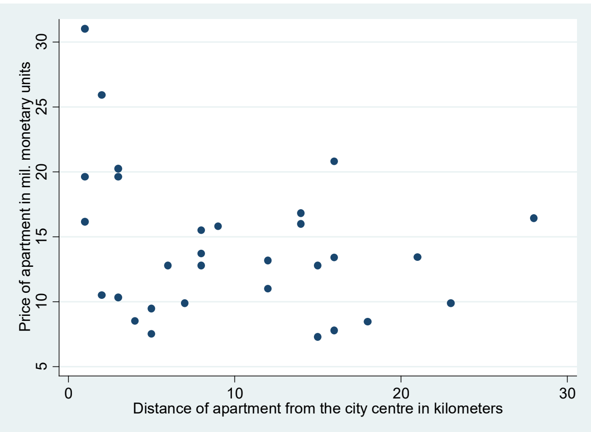
```
. sum

    Variable |       Obs        Mean    Std. Dev.       Min        Max
-------------+--------------------------------------------------------
           n |        30        15.5    8.803408          1         30
       price |        30    14.21717❶   5.528791❷      7.3         31
          m2 |        30       65.65❸   25.23567         33        136
         dis |        30    9.866667    7.233701          1         28
         age |        30    17.03333    10.49953          1         45
```

. scatter price m2



. scatter price age

. scatter price dis



```
. regress price m2 age dis

      Source |       SS       df       MS              Number of obs =      30
-------------+------------------------------           F(  3,    26) =   85.83❻
       Model |  805.156707     3  268.385569           Prob > F      =  0.0000❼
    Residual |  81.3017384⓰ 26  3.12698994❺           R-squared     =  0.9083❽
-------------+------------------------------           Adj R-squared =  0.8977❾
       Total |  886.458445⓱ 29  30.5675326           Root MSE      =  1.7683❹

------------------------------------------------------------------------------
       price |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
          m2 |   .2051172❿  .0137199    14.95⓫ 0.000⓬   .1769156    .2333188
         age |  -.0486017⓭  .0328782    -1.48   0.151    -.1161838    .0189804
         dis |  -.1635526⓮  .0455379    -3.59   0.001    -.2571572   -.0699481
       _cons |   3.192791⓯ 1.083781     2.95   0.007     .9650468    5.420536
------------------------------------------------------------------------------

. estat vce

Covariance matrix of coefficients of regress model

        e(V) |         m2          age          dis        _cons
-------------+------------------------------------------------
          m2 |  .00018824
         age | -.00013912    .00108098
         dis |  .00004938   -.00002756     .0020737
       _cons | -.01047515   -.00900722   -.02323306    1.1745823

. estat ic

-----------------------------------------------------------------------------
       Model |    Obs    ll(null)   ll(model)     df          AIC         BIC
-------------+---------------------------------------------------------------
           . |     30   -93.35871   -57.52271⓲    4     123.0454⓳  128.6502⓴
-----------------------------------------------------------------------------
               Note:  N=Obs used in calculating BIC; see [R] BIC note

. predict e, resid
. predict pricehat, xb

. list price pricehat e, mean

     +-------------------------------+
     |  price   pricehat           e |
     |-------------------------------|
  1. |  13.16   12.43582    .7241774 |
  2. |     16   14.80242    1.197579 |
  3. |   15.5   18.18543   -2.685427 |
  4. |   20.8   19.39813    1.401871 |
  5. |   13.7   16.83079   -3.130794 |
     |-------------------------------|
  6. |   8.46   8.242655    .2173457 |
  7. | 16.815   14.80737    2.007628 |
  8. |    7.5   8.560676   -1.060675 |

 ...

 29. |   10.5   11.12969   -.6296857 |
 30. | 12.784   13.60394   -.8199354 |
     |-------------------------------|
Mean | 14.217   14.21717    9.20e-09 |
     +-------------------------------+
```

■

Solution of example 2

Interpretation of some of the model parameters:

- $s_e = 1.76833$ (❹): standard error of regression:

Estimate of the standard deviation of the disturbances, measuring the deviations of observed values of the apartment price from its fitted values, obtained based on the multiple regression model, is equal to 1.768 mil. monetary units.

- $KV = \frac{s_e}{\bar{y}} = \frac{1.7683}{14.2172} = 0.124$: estimate of the coefficient of variation:

Coefficient of variation is a relative measure of the adequacy of a regression model. Its estimate amounts to 0.124, meaning that the standard error of regression represents 12.4 percent of the mean apartment price.

- Value of the F–statistic and the corresponding p–value for testing the null hypothesis:

$$H_0 : \beta_j = 0; \quad \forall j = 2,3,4$$
$$H_1 : \beta_j \neq 0; \quad \exists j = 2,3,4$$

Value of the F–statistic for testing the null hypothesis that all partial regression coefficients are (simultaneously) equal to 0 amounts to 85.83 (❻), while the exact level of statistical significance is p = 0.000 (❼). Thus we can reject the null hypothesis and conclude that at least one partial regression coefficient is statistically significantly different from 0. We can conclude, based on sample data, that the regression model as a whole explains the variability of apartment price satisfactorily, and exhibits a linear relationship among the variables.

- $R^2 = 0.9083$ (❽): estimate of the multiple determination coefficient:

Estimate of the multiple determination coefficient is equal to 0.9083, which means that almost 91 percent of the variability in apartment price is explained by the linear relationship between apartment price and its size, age and distance from the city centre.

- $b_2 = 0.205117$ (❶⓿): estimate of the (first) partial regression coefficient:

If the apartment size increases by one square meter, the apartment price increases on average by 205,117 monetary units, keeping the remaining two explanatory variables, i.e. apartment age and its distance from the city centre, unchanged.

- Value of the t–statistic and the corresponding p–value for testing: $H_0 : \beta_2 = 0$

Value of the t–statistic for testing the null hypothesis that the second partial regression coefficient is equal to 0 amounts to 14.95 (❶❶), while the exact level of statistical significance is p = 0.000 (❶❷). Thus, we can reject the null hypothesis and conclude that the partial regression coefficient β_2 is statistically significantly different from 0. We can conclude, based on sample data, that apartment size affects its price.

- Estimate of the (first) partial elasticity at the means:

$$\hat{\varepsilon}_{\bar{y},\bar{x}_2} = b_2 \frac{\bar{x}_2}{\bar{y}} = 0.20512 \cdot \frac{65.65}{14.2172} = 0.947; \qquad \bar{y} = 14.2172 \qquad \bar{x}_2 = 65.65$$

If the mean apartment size increases by one percent, the mean apartment price increases by 0.947 percent, keeping the remaining two explanatory variables, i.e. apartment age and its distance from the city centre, unchanged.

- Log-likelihood value of the model (⓲):

$$\ln L = -\frac{n}{2}\left[\ln(2\pi) + \ln\left(\frac{RSS}{n}\right) + 1\right] = -57.52$$

- Akaike information criterion (⓳):

$$AIC = -2\ln L + 2k = 123.05$$

- Schwarz criterion or Bayesian information criterion (⓴):

$$SC = BIC = -2\ln L + k\ln(n) = 128.65$$

Representation of the regression model:

$$PRICE_i = \beta_1 + \beta_2 M2_i + \beta_3 AGE_i + \beta_4 DIS_i + u_i$$

$$\begin{array}{lcccc} \widehat{PRICE}_i = & 3.19279 & +\,0.20512 M2_i & -\,0.04860 AGE_i & -\,0.16355 DIS_i \\ p: & (0.007) & (0.000) & (0.151) & (0.001) \end{array}$$

$$\begin{array}{cccccc} n = 30 & R^2 = 0.9083 & \bar{R}^2 = 0.8977 & s_e = 1.76833 & KV = 0.124 & F = 85.83 \\ & & & & & (0.000) \end{array}$$

■