Quantitative Methods in Finance

Tutorial, Part 6:
Model diagnostics. Normality of the disturbances and multicollinearity.

Example 1: For a given country we collected data for a period of 15 years on personal consumption (*PC*), wages (*WAGE*), social benefits (*SOB*), earnings from abroad (*ABR*), and special budgetary income (*SBI*). We estimated the following function of personal consumption:

$$\widehat{PC} = 0.46 + 0.78WAGE + 0.22SOB + 0.02ABR - 0.16SBI$$
$$t: \quad (4.5) \quad (1.2) \qquad\quad (0.4) \qquad (1.0) \qquad (-0.2)$$

We also estimated the regression model among the explanatory variables and found that 95 percent of the variability of wages is explained by the variation of social benefits, earnings from abroad, and special budgetary income.

a) Consider the issue of multicollinearity in case of the estimated consumption function.
b) What kind of consumption function would you estimate to at least somewhat alleviate the problems caused by multicollinearity?

■

Example 2: The agricultural institute estimated the wine harvest per hectare of vineyard (*HPV*) for the past thirty years, depending on age of the vineyard (*AGE)*, the number of sunny days per year (*SD*), the amount of precipitation (*PRE*), and the average temperature in the third quarter of the year (*T3*):

$$HPV_t = \beta_1 + \beta_2 PRE_t + \beta_3 AGE_t + \beta_4 SD_t + \beta_5 T3_t + u_t \,.$$

Since all the estimated partial regression coefficients of the model above were statistically insignificant and the determination coefficient extremely high (0.9885), they decided to estimate the regression model between the explanatory variables:

$$PRE_t = \beta_1 + \beta_2 AGE_t + \beta_3 SD_t + \beta_4 T3_t + u_t \,.$$

They found that the sum of squares of the observed amount of precipitation, $\sum PRE_t^2$, equals 528, the sum of squared residuals is 14.5, and the average annual amount of precipitation equals 2.

Explain what they wanted to verify on the basis of the regression model among the explanatory variables, and what was the obtained conclusion. Based on this conclusion, what should they do in order to get reliable estimates of the effect of individual factors on the wine harvest per hectare?

■

Example 3: Based on data for 27 houses in a U.S. district, which were sold in 1977 (file `housing1.dta`), we want to estimate how the price of a house (*PRICE*; in 1,000 USD) is affected by its size (*SIZE*; in 1,000 square feet), age (*AGE*; in years), the number of rooms (*ROOMS*), and the number of bedrooms (*BEDROOM*).

Analyze the problem of multicollinearity. What kind of a regression model would you be estimating to overcome the problems caused by multicollinearity?

For the proposed function of the house price, verify if the assumption of a normal distribution of the disturbances *u* is fulfilled.

Computer printout of the results in Stata:

```
. regress price size age rooms bedrooms

      Source |       SS       df       MS              Number of obs =      27
-------------+------------------------------           F(  4,    22) =   42.74
       Model |  4716.28866     4  1179.07216           Prob > F      =  0.0000
    Residual |  606.914165    22  27.5870075           R-squared     =  0.8860
-------------+------------------------------           Adj R-squared =  0.8653
       Total |  5323.20282    26   204.73857           Root MSE      =  5.2523

------------------------------------------------------------------------------
       price |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
        size |   .0223931   .0037252     6.01   0.000     .0146674    .0301188
         age |  -.1488602   .0816217    -1.82   0.082    -.3181331    .0204128
       rooms |    1.42822   2.637175     0.54   0.594    -4.040945    6.897386
    bedrooms |  -1.735464   3.911259    -0.44   0.662     -9.84692    6.375991
       _cons |    6.21797   7.723658     0.81   0.429    -9.799916    22.23586
------------------------------------------------------------------------------

. pwcorr size age rooms bedrooms, sig

             |     size      age    rooms bedrooms
-------------+------------------------------------
        size |   1.0000
             |
             |
         age |  -0.1781   1.0000
             |   0.3742
             |
       rooms |   0.8457   0.0043   1.0000
             |   0.0000   0.9829
             |
    bedrooms |   0.7934   0.1026   0.9244   1.0000
             |   0.0000   0.6105   0.0000
             |

. collin size age rooms bedrooms, corr
(obs=27)

  Collinearity Diagnostics

                        SQRT                   R-
  Variable      VIF     VIF    Tolerance    Squared
----------------------------------------------------
      size      4.08    2.02    0.2453      0.7547
       age      1.24    1.11    0.8064      0.1936
```

```
    rooms      8.98    3.00    0.1113      0.8887
 bedrooms      7.56    2.75    0.1323      0.8677
----------------------------------------------------
 Mean VIF      5.46

                           Cond
        Eigenval          Index
---------------------------------
    1     2.7110          1.0000
    2     1.0483          1.6081
    3     0.1729          3.9598
    4     0.0678          6.3235
---------------------------------
 Condition Number         6.3235
 Eigenvalues & Cond Index computed from deviation sscp (no intercept)
 Det(correlation matrix)    0.0333
```

. regress price size age

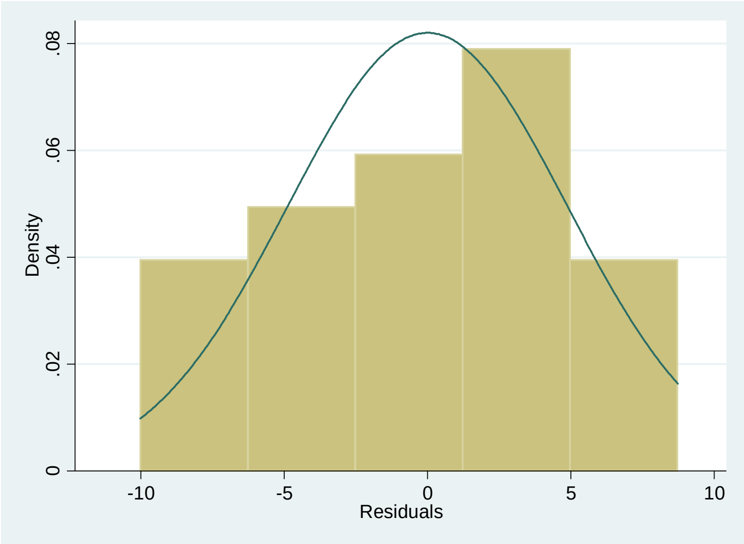
```
      Source |       SS       df       MS              Number of obs =      27
-------------+------------------------------           F(  2,    24) =   91.85
       Model |  4708.12215     2  2354.06107           Prob > F      =  0.0000
    Residual |  615.080677    24  25.6283615           R-squared     =  0.8845
-------------+------------------------------           Adj R-squared =  0.8748
       Total |  5323.20282    26   204.73857           Root MSE      =  5.0624

------------------------------------------------------------------------------
       price |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
        size |   .0231243   .0018073    12.79   0.000     .0193942    .0268544
         age |  -.1523485   .0717955    -2.12   0.044    -.3005271   -.0041699
       _cons |   9.093323   4.222131     2.15   0.042     .3792732    17.80737
------------------------------------------------------------------------------
```

. collin size age, corr

```
(obs=27)

  Collinearity Diagnostics

                        SQRT                   R-
  Variable      VIF     VIF    Tolerance    Squared
----------------------------------------------------
      size      1.03    1.02    0.9683      0.0317
       age      1.03    1.02    0.9683      0.0317
----------------------------------------------------
  Mean VIF      1.03

                           Cond
        Eigenval          Index
---------------------------------
    1     1.1781          1.0000
    2     0.8219          1.1972
---------------------------------
 Condition Number         1.1972
 Eigenvalues & Cond Index computed from deviation sscp (no intercept)
 Det(correlation matrix)    0.9683
```

. predict eprice, resid

```
. histogram eprice, normal
(bin=5, start=-10.017778, width=3.7491447)
```




```
. sum eprice, detail

                          Residuals
-------------------------------------------------------------
      Percentiles      Smallest
 1%    -10.01778      -10.01778
 5%    -8.911816      -8.911816
10%    -6.921759      -6.921759       Obs                  27
25%    -3.455669      -6.515256       Sum of Wgt.          27

50%     .9022909                      Mean          -1.88e-08
                        Largest       Std. Dev.      4.863841
75%     3.624211       5.456469
90%     6.491363       6.491363       Variance       23.65695
95%     6.566592       6.566592       Skewness      -.3191895
99%     8.727945       8.727945       Kurtosis       2.349854

. return list

scalars:
                  r(N) =  27
              r(sum_w) =  27
               r(mean) =  -1.87644252071e-08
                r(Var) =  23.65694875107343
                 r(sd) =  4.863840946317368
           r(skewness) =  -.3191895093658222
           r(kurtosis) =  2.349854454971561
                r(sum) =  -5.06639480591e-07
                r(min) =  -10.01777839660645
                r(max) =  8.727945327758789
                 r(p1) =  -10.01777839660645
                 r(p5) =  -8.911815643310547
                r(p10) =  -6.921758651733398
                r(p25) =  -3.455668926239014
                r(p50) =  .9022908806800842
                r(p75) =  3.624211072921753
                r(p90) =  6.491362571716309
                r(p95) =  6.566591739654541
                r(p99) =  8.727945327758789
```

```
. scalar obs=r(N)
. scalar sk=r(skewness)
. scalar ku=r(kurtosis)

. scalar jb=obs*(sk^2/6 + (ku-3)^2/24)
. display jb
.93399413

. display chi2tail(2,jb)
.62688193

. jb6 eprice
Jarque-Bera normality test:   .934 Chi(2)  .6269
Jarque-Bera test for Ho: normality: (eprice)
```

■