Quantitative Methods in Finance

Tutorial, Part 8:
Regression models with dummy explanatory variables.

Example 1: Based on quarterly data on profits and sales of industrial enterprises in the United States in the period 1965q1–1970q4 (data file `profit.dta`; both variables are expressed in billion USD at constant prices) estimate the following regression function:

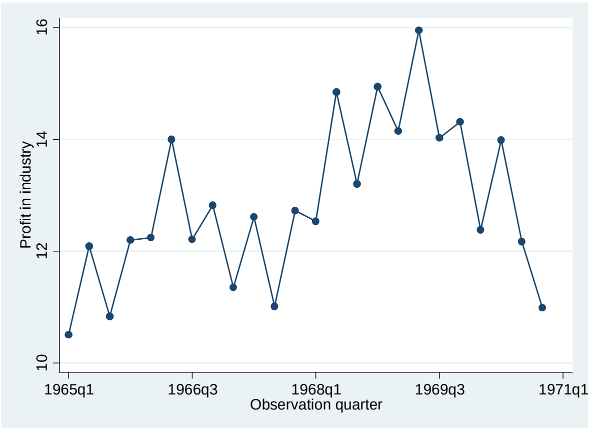
$$PROFIT_t = \beta_1 + \beta_2 SALES + u_t$$

and analyze the potential impact of the seasonal component on the link between the (numeric) variables by using dummy explanatory variables.

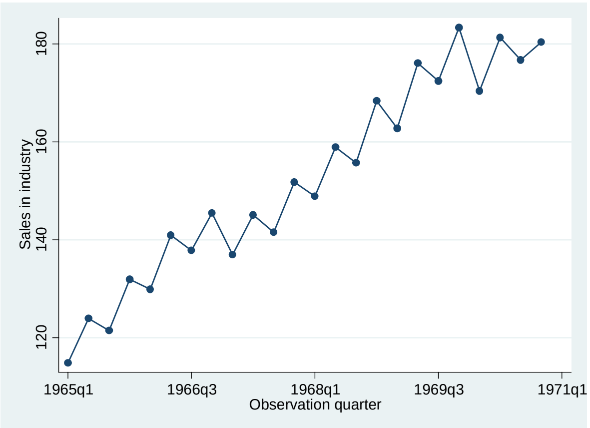
Computer printout of the results in Stata:

```
. tsset quarter
        time variable:  quarter, 1965q1 to 1970q4
                delta:  1 quarter
```

```
. twoway connected profit quarter
```




```
. twoway connected sales quarter
```




```
. gen q=quarter(dofq(quarter))

. tabulate q, gen(d)

          q |      Freq.     Percent        Cum.
------------+-----------------------------------
          1 |          6       25.00       25.00
          2 |          6       25.00       50.00
          3 |          6       25.00       75.00
          4 |          6       25.00      100.00
------------+-----------------------------------
      Total |         24      100.00
```

```
. list quarter profit sales q d1-d4

     +----------------------------------------------------------+
     | quarter   profit     sales   q   d1   d2   d3   d4 |
     |----------------------------------------------------------|
  1. |  1965q1   10.503   114.862   1    1    0    0    0 |
  2. |  1965q2   12.092   123.968   2    0    1    0    0 |
  3. |  1965q3   10.834   121.454   3    0    0    1    0 |
  4. |  1965q4   12.201   131.917   4    0    0    0    1 |
  5. |  1966q1   12.245   129.911   1    1    0    0    0 |
     |----------------------------------------------------------|
  6. |  1966q2   14.001   140.976   2    0    1    0    0 |
  7. |  1966q3   12.213   137.828   3    0    0    1    0 |
  8. |  1966q4    12.82   145.465   4    0    0    0    1 |
  9. |  1967q1   11.349   136.989   1    1    0    0    0 |
 10. |  1967q2   12.615   145.126   2    0    1    0    0 |
     |----------------------------------------------------------|
 11. |  1967q3   11.014   141.536   3    0    0    1    0 |
 12. |  1967q4    12.73   151.776   4    0    0    0    1 |
 13. |  1968q1   12.539   148.862   1    1    0    0    0 |
 14. |  1968q2   14.849   158.913   2    0    1    0    0 |
 15. |  1968q3   13.203   155.727   3    0    0    1    0 |
     |----------------------------------------------------------|
 16. |  1968q4   14.947   168.409   4    0    0    0    1 |
 17. |  1969q1   14.151   162.781   1    1    0    0    0 |
 18. |  1969q2   15.949   176.057   2    0    1    0    0 |
 19. |  1969q3   14.024   172.419   3    0    0    1    0 |
 20. |  1969q4   14.315   183.327   4    0    0    0    1 |
     |----------------------------------------------------------|
 21. |  1970q1   12.381   170.415   1    1    0    0    0 |
 22. |  1970q2   13.991   181.313   2    0    1    0    0 |
 23. |  1970q3   12.174   176.712   3    0    0    1    0 |
 24. |  1970q4   10.985    180.37   4    0    0    0    1 |
     +----------------------------------------------------------+

. regress profit sales

      Source |       SS       df       MS              Number of obs =      24
-------------+------------------------------           F(  1,    22) =   11.70
       Model |  16.4051235     1  16.4051235           Prob > F      =  0.0024
    Residual |  30.8438618    22  1.40199372           R-squared     =  0.3472
-------------+------------------------------           Adj R-squared =  0.3175
       Total |  47.2489853    23  2.05430371           Root MSE      =  1.1841

------------------------------------------------------------------------------
      profit |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
       sales |   .0409541   .0119724     3.42   0.002     .0161249    .0657832
       _cons |   6.597974   1.840287     3.59   0.002     2.781452     10.4145
------------------------------------------------------------------------------

. regress profit sales d1-d4
note: d4 omitted because of collinearity

      Source |       SS       df       MS              Number of obs =      24
-------------+------------------------------           F(  4,    19) =    5.26
       Model |  24.8290561     4  6.20726403           Prob > F      =  0.0050
    Residual |  22.4199291    19  1.17999627           R-squared     =  0.5255
-------------+------------------------------           Adj R-squared =  0.4256
       Total |  47.2489853    23  2.05430371           Root MSE      =  1.0863
```

```
------------------------------------------------------------------------------
      profit |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
       sales |   .0382462   .0114809     3.33   0.004     .0142163     .062276
          d1 |  -.1838562   .6542925    -0.28   0.782    -1.553306    1.185594
          d2 |   1.139036   .6307096     1.81   0.087    -.1810546    2.459126
          d3 |  -.4016617   .6361179    -0.63   0.535    -1.733072    .9297484
          d4 |  (omitted)
       _cons |   6.872219   1.892072     3.63   0.002     2.912067    10.83237
------------------------------------------------------------------------------
```

. regress profit sales d2-d4

```
      Source |       SS       df       MS              Number of obs =      24
-------------+------------------------------           F(  4,    19) =    5.26
       Model |  24.8290561     4  6.20726403           Prob > F      =  0.0050
    Residual |  22.4199291    19  1.17999627           R-squared     =  0.5255
-------------+------------------------------           Adj R-squared =  0.4256
       Total |  47.2489853    23  2.05430371           Root MSE      =  1.0863

------------------------------------------------------------------------------
      profit |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
       sales |   .0382462   .0114809     3.33   0.004     .0142163     .062276
          d2 |   1.322892   .6384745     2.07   0.052    -.0134505    2.659234
          d3 |  -.2178055   .6322552    -0.34   0.734    -1.541131     1.10552
          d4 |   .1838562   .6542925     0.28   0.782    -1.185594    1.553306
       _cons |   6.688363   1.711366     3.91   0.001     3.106432    10.27029
------------------------------------------------------------------------------
```

. test d3=d4=0

```
 ( 1)  d3 - d4 = 0
 ( 2)  d3 = 0

       F(  2,    19) =    0.20
            Prob > F =    0.8197
```

. regress profit sales d2

```
      Source |       SS       df       MS              Number of obs =      24
-------------+------------------------------           F(  2,    21) =   11.17
       Model |    24.35495     2   12.177475           Prob > F      =  0.0005
    Residual |  22.8940352    21  1.09019215           R-squared     =  0.5155
-------------+------------------------------           Adj R-squared =  0.4693
       Total |  47.2489853    23  2.05430371           Root MSE      =  1.0441

------------------------------------------------------------------------------
      profit |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
       sales |   .0393105    .010575     3.72   0.001     .0173186    .0613024
          d2 |   1.331353   .4930214     2.70   0.013     .3060584    2.356647
       _cons |   6.515581   1.623084     4.01   0.001     3.140194    9.890968
------------------------------------------------------------------------------
```

■

Example 2: Based on the data for 150 industrial enterprises we were estimating the following production function: $\ln Q_i = \beta_1 + \beta_2 \ln L_i + \beta_3 \ln K_i + u_i$.

We divided the enterprises according to their ownership and got the following results:

- state ownership: $\widehat{\ln Q} = 3.63 + 0.53 \ln L + 0.48 \ln K$;
- mixed ownership: $\widehat{\ln Q} = 3.24 + 0.18 \ln L + 0.68 \ln K$;
- private ownership: $\widehat{\ln Q} = 4.73 + 0.23 \ln L + 0.55 \ln K$.

Based on the estimated production functions, write out a single regression model with explicit values of regression coefficients that will allow for a simultaneous estimation of the above three production functions.

Example 3: For 100 four-person households, we gathered data on their expenditures for food (*EXP* in 1,000 monetary units) and disposable income (*INCOME* in 1,000 monetary units). We estimated the following regression functions:

$$\widehat{EXP} = 43.523 + 0.175 INCOME \qquad R^2 = 0.452;$$

$$\begin{array}{lllll} \widehat{EXP} = 31.251 + & 4.317 D1 + & 13.251 D2 + & 0.163 INCOME & \\ se(b_j):\ (5.44) & (2.15) & (4.25) & (0.05) & R^2 = 0.642. \end{array}$$

We divided the households into three groups: agricultural, mixed, and non-agricultural. We also defined two dummy variables: *D1* equals 1, if the household is mixed, and *D2* equals 1, if the household is non-agricultural.

a) Interpret the estimated regression coefficients in both regression functions.
b) Check the assumption that the type of household does not affect the expenditures.
c) Can we claim that at a given level of income, the average expenditures for food in mixed households are higher than that of agricultural households?
d) Explain why we used only two dummy variables, although we divided households into three groups.