WORKSHOP

1. For OLS to be unbiased, the omitted variables must be uncorrelated with the included variables.

   In the last workshop, you estimated

   $$ac = \alpha + \beta q + \varepsilon \ .$$

   All factors that affect average costs other than output level are omitted (that is, are incorporated into ε).

   a. Name at least three factors that affect average costs other than output level.

   b. Which of these factors can be expected to be correlated with output level?

2. One factor, which you might have listed in question 1, is capacity. The capacity of a plant probably affects its average costs, because larger plants cost more to construct. Furthermore, the capacity of a plant is probably correlated with the output of the plant: plants with larger capacity generally produce larger output. Therefore, omitting capacity can cause bias in the estimated coefficient of q.

   a. Run a TSP program to see if q and k are correlated in your sample of plants.

   b. In the last step of the previous workshop, you ran a regression with k included as an explanatory variable. Compare the estimated coefficient of q in this regression that includes k with the previous regression that omitted k.

      i. Does the estimated coefficient of q change much when k is included?

      ii. Does the estimated coefficient of q get larger or smaller in magnitude when k is included?

      iii. Explain, in terms of real world causal processes, why the coefficient of q changes in the direction that it does when k is added.

3. Average cost is probably higher in areas with higher input prices (that is, where fuel, labor, and materials are most costly). Do you think that input prices are correlated with either q or k such that excluding them causes bias?

   You decided to include input prices. You construct an index of input prices. The index is calculated for each region and takes the values:

region 1	1.12
region 2	1.23
region 3	.87
region 4	.91

   Enter input prices into the regression by using these commands:

```
genr r = region;    #so you won't have to type
                #    region over and over
genr p = 1.12*(r.eq.1) + 1.23*(r.eq.2)
                +.87*(r.eq.3) + .91*(r.eq.4);
olsq ac c,q,k,p;
```

   Do the coefficients of q and k change much when p enters?

4. We learned that the variance of the OLS estimator is lower with larger samples. Run your regression with fewer observations and see whether the standard errors change. For example, omit the first 60 plants and run the regression on the remaining plants. This can be done by using

```
smpl 61 127;
olsq ac c,q,k,p;
```

5. We learned that the variance of the OLS estimator is lower when omitted variables have less influence. Compare your output from step 3 (the regression with q, k, and p as explanatory variables) with your output from the previous workshop with only q as an explanatory variable. Adding k and p reduces the influence of the omitted variables (because k and p are not part of the omitted variables when they are entered in the regression). But the standard error of the coefficient of q **rises** when k and p are added, contrary to our expectations. What is happening?