ECON-122
Introduction to Econometrics

Agnieszka Postepska

July 18th, 2011

- remember that we are making a statement about how x affect y holding everything else constant
- hardly ever is one regressor enough to satisfy the ceteris paribus requirement
- with only one x the assumption we are making is that all other factors that affect y are uncorrelated with x - highly unlikely
- we would like to control for other factors that affect y in the regression equation
- as multiple regression can accommodate many explanatory variable, we can make statements about causality (as in our example with vitamins - if we control for social status and we still find effect of vitamins on GPA we can claim that we have found causal effect)

Example to motivate multiple regression

- suppose we are interested in the effect of education (edu) and experience (exp) on wages
- we can either build two simple regression models:

$$wage_i = \beta_0 + \beta_1 edu_i + u_i$$
$$wage_i = \alpha_0 + \alpha_1 exp_i + v_i$$

- or we can build one model with two explanatory variable

$$wage_i = \beta_0 + \beta_1 edu_i + \beta_3 exp + \epsilon_i$$

- which formulation is better and why?

Model with k independent variables

- no need to stop with just two - we want to be able to control for as many factors as we want
- general multiple linear regression model can be written in the population as follows:

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + ... + \beta_k x_k + u$$

- now we have $k + 1$ unknown parameters to estimate
- u is still the disturbance (error term) that contains all other factors that affect y other than x_1, x_2, x_3 etc
- once again - remember that no matter how many variables $x's$ we include in our model, there will always be factors we cannot control for

Mechanics of OLS with k independent variables

Objective function for two independent variables becomes:

$$\min_{\hat{\beta}_0,\hat{\beta}_1,\hat{\beta}_2} \sum_i (y_i - \hat{\beta}_0 - \hat{\beta}_1 x_{i1} - \hat{\beta}_2 x_{i2})^2$$

And for k independent variables:

$$\min_{\hat{\beta}_0,\hat{\beta}_1,\ldots,\hat{\beta}_k} \sum_i (y_i - \hat{\beta}_0 - \hat{\beta}_1 x_{i1} - \hat{\beta}_2 x_{i2} - \ldots - \hat{\beta}_k x_{ik})^2$$

Derivations of FOC on the blackboard...

Introduction to matrix algebra on the blackboard

The model

Let us summarize what we have done on the blackboard:
The Multiple Regression Model is of the following form (thick font represents vectors and matrices)

$$\underbrace{\mathbf{y}}_{nx1} = \underbrace{\overbrace{\mathbf{X}}^{nx(k+1)} \underbrace{\hat{\beta}}_{(k+1)x1} + \overbrace{\mathbf{u}}_{nx1}}_{nx1}$$

And as we showed on the blackboard:
OLS estimator is given by:

$$\hat{\beta}_{\mathbf{OLS}} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y}_{(k+1)x1}$$