ECON 3150/4150, Spring term 2014. Lecture 7

The multivariate regression model (I)

Ragnar Nymoen

University of Oslo

4 February 2014

References to Lecture 7 and 8

- **SW**
  - Ch. 6
- **BN**
  - Kap 7.1-7.8

Omitted variable bias I

- Assume that we estimate a simple regression model

$$Y_i = \beta_0 + \beta_1 X_{1i} + \varepsilon_i \tag{1}$$

by OLS, but that another regression model is

$$Y_i = \beta_0^* + \beta_1^* X_{1i} + \beta_2^* X_{2i} + \varepsilon_i^* \tag{2}$$

where X_{2i} is a second regressor and ε_i^* has classical properties conditional on both X_{1i} and X_{2i}, in particular:

$$E(\varepsilon_i^* \mid X_{1i}) = 0 \forall i$$
$$E(\varepsilon_i^* \mid X_{2i}) = 0 \forall i$$

Omitted variable bias II

and

$$E(Y_i \mid X_{1i}, X_{2i}) = \beta_0^* + \beta_1^* X_{1i} + \beta_2^* X_{2i}$$

i.e., linearity of the conditional expectation function.

- We want to evaluate the OLS estimator $\hat{\beta}_1$, of (1), in the light of (2).
- First, show that

$$\hat{\beta}_1 = \beta_1^* + \beta_2^* \frac{\sum_{i=1}^n (X_{1i} - \bar{X}_1) X_{2i}}{\sum_{i=1}^n (X_{1i} - \bar{X}_1)^2} + \frac{\sum_{i=1}^n (X_{1i} - \bar{X}_1) \varepsilon_i^*}{\sum_{i=1}^n (X_{1i} - \bar{X}_1)^2}$$

Omitted variable bias III

- If X_{1i} and X_{2i} have a well defined joint probability function (after Y_i has been marginalized out from $f(Y_i, X_{1i}, X_{2i})$) we can use "plim algebra" (Slutsky's theorem in Lecture 2) to show that

$$plim(\hat{\beta}_1) = \beta_1^* + \beta_2^* \tau_{12} \quad (3)$$

$$\tau_{12} = \frac{cov(X_1, X_2)}{\sigma_{X_1}^2} = \rho_{X_1 X_2} \frac{\sigma_{X_2}}{\sigma_{X_1}} \quad (4)$$

  i.e. the population regression coefficient between X_2 and X_1!

- We can also show (with no extra or stronger assumptions) that:

$$E(\hat{\beta}_1) = \beta_1^* + \beta_2^* \tau_{12} \quad (5)$$

Omitted variable bias IV

- Returning to (3) and (4): These expressions are more precise than eq. (6.1) in SW, because "$\rho_{Xu} = 0$" in a regression model under quite mild assumptions.
- The point is that although $E(\varepsilon_i \mid X_{1i}) = 0$, there is nothing in the specification of the simple regression model that implies $E(\varepsilon_i \mid X_{2i}) = 0$.
- And it is this possibility: That Y is affected by other variables than X_1, that underlies the omitted variable bias.

Omitted variable bias V

- In summary (assume that causality is settled from theory, to keep that issue out of the way for the time being):
  - With simple regression we are in general not estimating the partial effect of a change in X_1
  - For that to be true, we must have *orthogonal regressors* ($\tau_{12} = 0$), or that Y is independent of X_2 ($\beta_2^* = 0$).
  - In general therefore, simple regression gives us the gross (or "total") effect on Y of a change in X_1.

Bi- and multivariate regression I

- In general, the theory of multiple regression covers a large number of regressors, denoted k, some of them random other deterministic.
- Still, a lot can be learned by studying the case of $k = 2$ in detail first.
- We follow that route, and mention the extension to the case of $k > 2$ at the end, without going into the matrix algebra needed for the general case.

Model specification

The model can be specified by the linear relationship

$$Y_i = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \varepsilon_i \;\; i = 1, 2, \ldots, n \tag{6}$$

and the set assumptions:

a. X_{ji} $(j = 1, 2)$, $(i = 1, 2, \ldots, n)$ can be deterministic or random
For a deterministic variable we assume that at least two values of the variables are distinct. For random Xs, we assume $var(X_{ji}) = \sigma^2_{X_j} > 0$ $(j = 1, 2)$, $\rho^2_{X_1 X_2} < 1$ and no excess kurtosis in the distributions

b. $E(\varepsilon_i) = 0, \; \forall \; i$

c. $var(\varepsilon_i) = \sigma^2, \; \forall \; i$

d. $cov(\varepsilon_i, \varepsilon_j) = 0, \; \forall \; i \neq j,$

e. β_0, β_1 and σ^2 are constant parameters

For the purpose of statistical inference for finite n:

f. $\varepsilon_i \sim IIN(0, \sigma^2)$.

Comments to specification I

- **a.** is formulated to accommodate different data types (random/deterministic, continuous/binary)
- $\rho^2_{X_1 X_2} < 1$ is a way of saying that the two random variables are truly separate variables.
- In many presentations, you will find an assumption about "absence of exact linear relationships between the variables often called absence of *exact collinearity*. But this can only occur for the case for deterministic variables, and would be an example of "bad model specification", e.g., specifying X_{2i} as a variables with the number 100 as the value for all i. (An example of the "dummy-variable fallacy/pit-fall").

Comments to specification II

- For random variables, we can of course be unlucky and draw a sample where $r^2_{X_1 X_2}$ is very high. But this "*near exact collinearity*" is *a property of the sample*, not of the regression model
- **b.-d. and f**. These are the same as in the case with one variable. Since we want a model formulation that allows random explanatory variables they should be interpreted as conditional on $X_{1i} = x_{1i}$ and $X_{2i} = x_{2i}$. The explicit conditioning has been omitted to save notation.

OLS estimation

Nothing new here: Choose the estimates that minimize

$$S(\beta_0,\beta_1,\beta_2) = \sum_{i=1}^{n}(Y_i - \beta_0 - \beta_1 X_{1i} - \beta_2 X_{2i})^2 \qquad (7)$$

or, equivalently, using the same type of re-specification as with simple regression

$$S(\alpha,\beta_1,\beta_2) = \sum_{i=1}^{n}(Y_i - \alpha - \beta_1(X_{1i} - \overline{X}_1) - \beta_2(X_{2i} - \overline{X}_2))^2$$

where

$$\alpha = \beta_0 + \beta_1 \overline{X}_1 + \beta_2 \overline{X}_2$$

- **Rest of derivation** in class

OLS estimates I

By solving the 1oc for minimum of $S(\alpha,\beta_1,\beta_2)$ (or $S(\beta_0,\beta_1,\beta_2)$, we obtain, for $\hat{\beta}_1$ and $\hat{\beta}_2$:

$$\hat{\beta}_1 = \frac{\hat{\sigma}^2_{X_2}\hat{\sigma}_{Y,X_1} - \hat{\sigma}_{Y,X_2}\hat{\sigma}_{X_1,X_2}}{\hat{\sigma}^2_{X_1}\hat{\sigma}^2_{X_2} - \hat{\sigma}^2_{X_1,X_2}} \tag{8}$$

$$\hat{\beta}_2 = \frac{\hat{\sigma}^2_{X_1}\hat{\sigma}_{Y,X_2} - \hat{\sigma}_{Y,X_1}\hat{\sigma}_{X_1,X_2}}{\hat{\sigma}^2_{X_1}\hat{\sigma}^2_{X_2} - \hat{\sigma}^2_{X_1,X_2}} \tag{9}$$

where $\hat{\sigma}^2_{X_j}(j = 1, 2)$, $\hat{\sigma}_{Y,X_j}$ $(j = 1, 2)$ and $\hat{\sigma}_{X_1,X_2}$ are empirical variances and covariances.
The estimates for the two versions of the intercepts become:

$$\hat{\beta}_0 = \bar{Y} + \hat{\beta}_1\overline{X}_1 + \hat{\beta}_2\overline{X}_2$$
$$\hat{\alpha} = \bar{Y}$$

Absence of perfect sample collinearity I

It is clear that (8) for $\hat{\beta}_1$ and (9) for $\hat{\beta}_2$ require

$$M := \hat{\sigma}^2_{X_1}\hat{\sigma}^2_{X_2} - \hat{\sigma}^2_{X_1,X_2} = \hat{\sigma}^2_{X_1}\hat{\sigma}^2_{X_2}(1 - r^2_{X_1X_2}) > 0$$

Cannot have perfect empirical correlation between the two regressors. Must have:

$$\hat{\sigma}^2_{X_1} > 0, \text{ and } \hat{\sigma}^2_{X_2} > 0 \text{ and } r^2_{X_1X_2} < 1 \iff -1 < r_{X_1X_2} < 1$$

- If any one of these conditions should fail, we have *exact (or perfect) collinearity*.
- Absence of perfect collinearity is a requirement about the nature of the sample.

Absence of perfect sample collinearity II

- The case of $r_{X_1X_2} = 0$ also has a name. It is called *perfect orthogonality*. It does not create any problems in (8) or (9).
- In practice, the relevant case is $-1 < r_{X_1X_2} < 1$, i.e. a *degree of collinearity* (not perfect)

Expectation I

- Conditional on the values of X_1 and X_2, $\hat{\beta}_1$ is still a random variable because ε_i and Y_i are random variables.
- In that interpretation $\hat{\beta}_1$, $\hat{\beta}_2$, and $\hat{\beta}_0$ are *estimators* and we want to know their expectation, variance, and whether they are consistent or not.
- Start by considering $E(\hat{\beta}_1 \mid X_1, X_2)$, i.e., conditional on all the values of the two regressors.

Expectation II

- Write $\hat{\beta}_1$ as

$$\hat{\beta}_1 = \frac{\left(\hat{\sigma}^2_{X_2}\hat{\sigma}_{Y,X_1} - \hat{\sigma}_{Y,X_2}\hat{\sigma}_{X_1,X_2}\right)}{M}$$

then $E(\hat{\beta}_1 \mid X_1, X_2)$ becomes

$$E(\hat{\beta}_1 \mid X_1, X_2) = \frac{\hat{\sigma}^2_{X_2}}{M} E(\hat{\sigma}_{Y,X_1} \mid X_1, X_2) - \frac{\hat{\sigma}_{X_1,X_2}}{M} E(\hat{\sigma}_{Y,X_2} \mid X_1, X_2) \tag{10}$$

- Evaluate this in class, in order to show that

$$E(\hat{\beta}_j) = E\left[E(\hat{\beta}_j \mid X_1, X_2)\right] = \beta_j,\ j = 1, 2 \tag{11}$$

since $E(\varepsilon_i \mid X_1, X_2) = 0\ \forall\ i$ is generic for the regression model.

Variance I

Find that (under the classical assumptions of the model):

$$var(\hat{\beta}_j \mid X_1, X_2) = \frac{\sigma^2}{n\hat{\sigma}^2_{X_j}\left[1 - r^2_{X_1,X_2}\right]}, j = 1, 2 \qquad (12)$$

and this also holds unconditionally.

- The BLUE property of the OLS estimators extends to the multivariate case (will no show)
- The variance (12) is low in samples that are informative about the “separate contributions” from X_1 and X_2:
  - $\hat{\sigma}^2_{X_j}$ high
  - $r^2_{X_1,X_2}$ low

Variance II

- $var(\hat{\beta}_j)$ is lowest when $r^2_{X_1,X_2} = 0$, the regressors are orthogonal

Covariance I

In many applications need to know $cov(\hat{\beta}_1, \hat{\beta}_2)$.
It is easiest to find by starting from the second normal equation

$$\hat{\beta}_1 \hat{\sigma}^2_{X_1} + \hat{\beta}_2 \hat{\sigma}_{X_1,X_2} = \hat{\sigma}_{YX_1}$$

When we take (conditional) variance on both sides, we get

$$\hat{\sigma}^4_{X_1} var(\hat{\beta}_1) + \hat{\sigma}^2_{X_1 X_2} var(\hat{\beta}_2) + 2\hat{\sigma}^2_{X_1} \hat{\sigma}_{X_1,X_2} cov\left(\hat{\beta}_1, \hat{\beta}_2\right) = \frac{1}{n^2} var(\hat{\sigma}_{YX_1})$$

The rhs we have from before:

$$n^{-2} var(\hat{\sigma}_{YX_1}) = n^{-2} \frac{\sigma^2}{n} \hat{\sigma}^2_{X_1} = n^{-1} \sigma^2 \hat{\sigma}^2_{X_1}$$

Covariance II

Insertion of expressions for $var(\hat{\beta}_1)$ and $var(\hat{\beta}_2)$, solving for $cov\left(\hat{\beta}_1, \hat{\beta}_2\right)$ gives

$$cov\left(\hat{\beta}_1, \hat{\beta}_2\right) = -\frac{\sigma^2}{n}\frac{\hat{\sigma}_{X_1 X_2}}{M}$$

Algebra details in note on the web-page.

Consistency of estimators I

Show for $\hat{\beta}_1$

$$\text{plim}\left(\hat{\beta}_1\right) = \text{plim}\left(\frac{\left(\hat{\sigma}^2_{X_2}\hat{\sigma}_{Y,X_1} - \hat{\sigma}_{Y,X_2}\hat{\sigma}_{X_1,X_2}\right)}{M}\right)$$

$$= \frac{\text{plim}\left(\hat{\sigma}^2_{X_2}\right)\text{plim}(\hat{\sigma}_{Y,X_1}) - \text{plim}(\hat{\sigma}_{Y,X_2})\,\text{plim}(\hat{\sigma}_{X_1,X_2})}{\text{plim}\, M}$$

Based on the assumptions of the regression model:

$$\text{plim}(\hat{\sigma}^2_{X_j}) = \sigma^2_{X_j}\ j = 1, 2$$

$$\text{plim}(\hat{\sigma}_{X_1,X_2}) = \sigma_{X_1 X_2}$$

$$\text{plim}\, M = \sigma^2_{X_1}\sigma^2_{X_2} - \sigma^2_{X_1,X_2}$$

Consistency of estimators II

$$\begin{aligned} \text{plim}(\hat{\sigma}_{Y,X_1}) &= \beta_1\sigma^2_{X_1} + \beta_2\sigma_{X_1X_2} + \text{plim}\left[\frac{1}{n}\sum_{i=1}^{n}\varepsilon_i(X_{1i} - \bar{X}_1)\right] \\ &= \beta_1\sigma^2_{X_1} + \beta_2\sigma_{X_1X_2} \\ \text{plim}(\hat{\sigma}_{Y,X_2}) &= \beta_1\sigma_{X_1X_2} + \beta_2\sigma^2_{X_2} \end{aligned}$$

$$\begin{aligned} \text{plim}\left(\hat{\beta}_1\right) &= \frac{\sigma^2_{X_2}\left[\beta_1\sigma^2_{X_1} + \beta_2\sigma_{X_1X_2}\right] - \left[\beta_1\sigma_{X_1X_2} + \beta_2\sigma^2_{X_2}\right]\sigma_{X_1,X_2}}{\sigma^2_{X_1}\sigma^2_{X_2} - \sigma^2_{X_1,X_2}} \\ &= \frac{\beta_1(\sigma^2_{X_2}\sigma^2_{X_1} - \sigma^2_{X_1X_2}) + \beta_2\sigma^2_{X_2}\sigma_{X_1X_2} - \beta_2\sigma^2_{X_2}\sigma_{X_1,X_2}}{\sigma^2_{X_1}\sigma^2_{X_2} - \sigma^2_{X_1,X_2}} \\ &= \beta_1 \end{aligned}$$

The OLS estimators $\hat{\beta}_0$ and $\hat{\beta}_2$ are also consistent