Making sense of Econometrics: Basics
Lecture 7: Model Specification

Hany Abdel-Latif & Anita Staneva

Egypt Scholars Economic Society

November 22, 2014

Assignment & feedback

enter classroom at

        http://b.socrative.com/login/student/

room name c28efb78

Outline

1. **Model Specification**
   - meaning
   - underfitting
   - overfitting
   - detecting specification errors

what is model specification

- concerns with two sets of choices
  - the set of variables included in a model
  - the functional form of the relationships we specify

- we will deal with the first of these choices

i underfitting:
  - omitting a relevant variable

ii overfitting
  - inclusion of an irrelevant variable

omitting relevant variable

- may occur due lack of data availability, oversight or ignorance
  - example:
    the omission of consumers' income or the price of a competing or substitute good in a demand function
- either way, this may lead to omitted variable bias (specification bias)

omitted variable bias

- imagine that the true causes of Y can be represented as

$$Y_i = \beta_1 + \beta_2 X_{2i} + \beta_3 X_{3i} + u_i$$

- now, if X_3 is omitted, the equation becomes instead

$$Y_i = \beta_1 + \beta_2 X_{2i} + \epsilon_i$$

where

$$\epsilon_i = u_i + \beta_3 X_{3i}$$

consequences of omitted variable

- this model violates our assumption that $E(\epsilon_i) = 0$
  - $E(u_i + \beta_3 X_3) \neq 0$
- this assumption is required in order to show that $\hat{\beta}$ is unbiased estimator of β
- by excluding X_3, we risk biasing our coefficients for X_2
- if the coefficient is biased, then the standard error is also biased
- then confidence interval and hypothesis testing procedures are unreliable

irrelevant variables

- may occur due to
  - inadvertent inclusion of theoretically unimportant variables
  - the researcher is unsure as to the importance of an explanatory variable
    (particularly, if theory is not very well developed)
  - including everything in the model on which data is available, hoping to include those variables which should be included

irrelevant variables

- assume that the true regression specification is

$$Y_i = \beta_1 + \beta_2 X_{2i} + u_i$$

- but the researcher for some reason includes an extra variable

$$Y_i = \beta_1 + \beta_2 X_{2i} + \beta_3 X_{3i} + v_i$$

- the misspecified equation's error term then becomes

$$v_i = u_i - \beta_3 X_{3i}$$

consequences of irrelevant variables

- the inclusion of an irrelevant variable will not cause bias
  - since the true coefficient of the irrelevant variable is zero
- the OLS estimators will also be consistent
- However, the inclusion of irrelevant variable will increase the variance of the estimated coefficient

  - this increased variance will tend to decrease the absolute magnitude of their t-ratios
  - this results in increasing the likelihood of incorrectly excluding a relevant variable

detecting omitted variables

- examination of residuals:
  - if a variable is omitted, its influence is implicitly absorbed by the error term
  - the behaviour of the residuals will mirror that of the omitted variable

detecting irrelevant variables

- tests for variable significance
  - if theory dictates that all these explanatory variables should be included in the model, then include them even if they are individually insignificant

  - if theory is not definitive regarding a particular variable, a researcher may include it in order to avoid omitted variable bias
- the significance of variables can be tested using the usual t-test and F-test
  - restricted least squares RLS (see lecture 3 pp. 42-45)

note of caution

- RLS for model specification test should follow a " general to specific", rather than "specific to general" approach
- commence with the most general model suggested by economic theory
- perform test of irrelevant variables in order to determine an appropriate specification

good model

- data admissible

- consistent with the theory

- uses regressors not correlated with error term

- exhibit parameter constancy

- exhibit data coherency

- encompassing