AMERICAN UNIVERSITY - CENTRAL ASIA 31.10.2022
Econometrics I (ECO 608)
Nurgul Tilenbaeva

Seminar 5

Endogeneity (Part I)

1 Specification Error

1. Import the Stata data file `"hprice1"` from the e-course platform.
2. Describe the dataset to get a better sense of what the data is about.
3. Obtain summary statistics of each variable.
4. Run the linear regression given by

   $price = \beta_0 + \beta_1 lotsize + \beta_2 sqrft + \beta_3 bdrms + u$

   where

   `price`=house price, in thousands of US dollars;
   `lotsize`=size of lot in square feet;
   `sqrft`=size of house in square feet;
   `bdrms`=number of bedrooms.

   (a) Obtain the fitted values $\hat{y}$ from estimating the model.
   (b) Run the linear regression given by

   $price = \beta_0 + \beta_1 lotsize + \beta_2 sqrft + \beta_3 bdrms + \delta_1 \hat{y}^2 + \delta_2 \hat{y}^3 + \delta_3 \hat{y}^4 + u$

   (c) Test the null hypothesis $H_0 : \delta_1 = 0, \delta_2 = 0, \delta_3 = 0$ against the alternative $H_1 : H_0$ *is not true* at `5% significance level`. Make the calculations by hand using the corresponding table. What do the results of hypothesis-testing suggest?
   (d) Check your results by running a RESET test in Stata.
5. Run the linear regression given by

   $log(price) = \beta_0 + \beta_1 log(lotsize) + \beta_2 log(sqrft) + \beta_3 bdrms + u$

   (a) Obtain the fitted values $\tilde{y}$ from estimating the model.

(b) Run the linear regression given by

$log(price) = \beta_0 + \beta_1 log(lotsize) + \beta_2 log(sqrft) + \beta_3 bdrms + \delta_1 \tilde{y}^2 + \delta_2 \tilde{y}^3 + \delta_3 \tilde{y}^4 + u$

(c) Test the null hypothesis $H_0 : \delta_1 = 0, \delta_2 = 0, \delta_3 = 0$ against the alternative $H_1 : H_0$ *is not true* at `5% significance level`. Make the calculations by hand using the corresponding table. What do the results of hypothesis-testing suggest?

(d) Check your results by running a RESET test in Stata.

(e) Based on RESET tests for the models in (4) and (5), which model specification is preferred?

6. Now we would like to test the model

$price = \beta_0 + \beta_1 lotsize + \beta_2 sqrft + \beta_3 bdrms + u$

against the model

$price = \beta_0 + \beta_1 log(lotsize) + \beta_2 log(sqrft) + \beta_3 log(bdrms) + u$

and vice versa.

Run the linear regression of a comprehensive model given by

$price = \gamma_0 + \gamma_1 lotsize + \gamma_2 sqrft + \gamma_3 bdrms + \gamma_4 log(lotsize) + \gamma_5 log(sqrft) + \gamma_6 log(bdrms) + u$

(a) Test the null hypothesis $H_0 : \gamma_4 = 0, \gamma_5 = 0, \gamma_6 = 0$ against the alternative $H_1 : H_0$ *is not true* at `5% significance level` as a test of the first model. Make the calculations by hand using the corresponding table. What do the results of hypothesis-testing suggest?

(b) Check your results by running an F test in Stata.

(c) Test the null hypothesis $H_0 : \gamma_1 = 0, \gamma_2 = 0, \gamma_3 = 0$ against the alternative $H_1 : H_0$ *is not true* at `5% significance level` as a test of the second model. Make the calculations by hand using the corresponding table. What do the results of hypothesis-testing suggest?

(d) Check your results by running an F test in Stata.

(e) Based on your results, which model in (6) you would prefer? Which test describes the procedure in (6)?

7. Now use the **Davidson-MacKinnon test** to choose among the models in (6).

   (a) First, run the linear regression given by

   $$price = \beta_0 + \beta_1 log(lotsize) + \beta_2 log(sqrft) + \beta_3 log(bdrms) + u$$

   (b) Obtain the fitted values $\bar{y}$ from estimating the model.

   (c) Run the linear regression given by

   $$price = \beta_0 + \beta_1 lotsize + \beta_2 sqrft + \beta_3 bdrms + \theta_1 \bar{y} + u$$

   (d) Test the null hypothesis $H_0 : \theta_1 = 0$ against the alternative $H_1 : \theta_1 \neq 0$ at `5% significance level`. What do the results of hypothesis-testing suggest?

   (e) Run the linear regression given by

   $$price = \beta_0 + \beta_1 lotsize + \beta_2 sqrft + \beta_3 bdrms + u$$

   (f) Obtain the fitted values $\dot{y}$ from estimating the model.

   (g) Run the linear regression given by

   $$price = \beta_0 + \beta_1 log(lotsize) + \beta_2 log(sqrft) + \beta_3 log(bdrms) + \theta_1 \dot{y} + u$$

   (h) Test the null hypothesis $H_0 : \theta_1 = 0$ against the alternative $H_1 : \theta_1 \neq 0$ at `5% significance level`. What do the results of hypothesis-testing suggest?

   (i) Based on your results, which model in (6) you would prefer?