Problem Set 3

BUAN/MIS 6356

Due: Monday, 2018-04-02-11:59pm

Deliverable:

an R source-code file named ps3.r

Question 1

Data

hprice1.csv contains data on 88 U.S. houses, their characteristics, and their prices at the time of sale.

Analysis

- Read the data hprice1.csv into a new variable: context1
- Run the following linear model using the 'lm' function. Store the result in: model1

$$\text{price}_i = \beta_0 + \beta_1 \text{bdrms}_i + \beta_2 \text{lotsize}_i + \beta_3 \text{sqrft}_i + e_i \tag{1}$$

- Compute the OLS significance test results and the White-corrected significance test (vcovHC) results for model1.
- Run the following linear model using the 'lm' function. Store the result in: model2

$$\ln[\text{price}_i] = \beta_0 + \beta_1 \text{bdrms}_i + \beta_2 \ln[\text{lotsize}_i] + \beta_3 \ln[\text{sqrft}_i] + e_i \tag{2}$$

- Compute the OLS significance test results and the White-corrected significance test results for model2.

Interpretations

a. Identify which variables are significant using the OLS test for model1.

b. Which variables are still significant after using the White-corrected significance test for model1?

c. Identify which variables are significant using the OLS test for model2.

d. Which variables are still significant after using the White-corrected significance test for model2?

e. Keeping these results in mind, what is the effect of taking logs on heteroskedasticity in the data?

Question 2

Data

beveridge.csv includes monthly observations on vacancy rates and unemployment rates for the U.S. from December 2000 through February 2012.

Analysis

- Read the data beveridge.csv into a new variable: context2
- Run the following linear model using the 'lm' function. Store the result in: model3

$$\text{urate}_t = \beta_0 + \beta_1 \text{vrate}_t + e_t \tag{3}$$

- Compute the OLS significance test results and the Newey-West-corrected significance test (5 lags; NeweyWest) results for model3.
- Perform the level and trend KPSS tests on urate_t and vrate_t (4 tests total).
- Perform the level and trend KPSS tests on Δurate_t and Δvrate_t (4 tests total).
- Perform the level and trend KPSS tests on $\Delta\left(\Delta\text{urate}_t\right)$ and $\Delta\left(\Delta\text{vrate}_t\right)$ (4 tests total).
- Run the following first-difference model using the 'lm' function. Store the result in: model4

$$\Delta\text{urate}_t = \beta_0 + \beta_1 \Delta\text{vrate}_t + e_t \tag{4}$$

- Compute the OLS significance test results and the Newey-West-corrected significance test (5 lags; NeweyWest) results for model4.

Interpretations

f. Do the OLS and NeweyWest significance tests show that the coefficient on the vanancy rate is significant or not (before we correct for stationarity)?

g. Based on the KPSS findings, which transformation/transformations should we apply to the unemployment rate before modeling?

h. Based on the KPSS findings, which transformation/transformations should we apply to the vacancy rate before modeling?

i. How have the significance tests changed from model3 to model4?

j. Which model better describes the data?

Question 3

Data

JTRAIN.csv has data on the scrap rates of various firms whether they were given a grant to increase their productivity. The scrap rate for a manufacturing firm is the number of defective items—products that must be discarded—out of every 100 produced. Thus, for a given number of items produced, a decrease in the scrap rate reflects higher worker productivity.

Analysis

- Read the data JTRAIN.csv into a new variable: context3
- Generate a new variable: d88

$$\text{d88}_t = \begin{cases} 1 & \text{if year}_t = 1988 \\ 0 & \text{otherwise} \end{cases}$$

- Generate a new variable: d89

$$\text{d89}_t = \begin{cases} 1 & \text{if year}_t = 1989 \\ 0 & \text{otherwise} \end{cases}$$

- Generate a new variable indicating whether or not the firm had a grant last year: $\text{grant}_{i,t-1}$

  [Hint: $\text{grant}_{i,t-1}$ is equal to zero when the year is 1987!]

- Run the following pooled linear model using the 'plm' function. Store the result in: model5

$$\ln[\text{scrap}_{it}] = \alpha + \beta_1 \text{d88}_t + \beta_2 \text{d89}_t + \beta_3 \text{grant}_{it} + \beta_4 \text{grant}_{i,t-1} + e_t \tag{5}$$

- Run the following fixed-effects model using the 'plm' function. Store the result in: model6

$$\ln[\text{scrap}_{it}] = \alpha_i + \beta_1 \text{d88}_t + \beta_2 \text{d89}_t + \beta_3 \text{grant}_{it} + \beta_4 \text{grant}_{i,t-1} + e_t \tag{6}$$

- Compute the OLS significance test results and the HAC-corrected significance test (Arellano) results for model6.

Interpretations

k. Interpret the estimated coefficient on grant_{it} in model5.

l. Interpret the estimated coefficient on $\text{grant}_{i,t-1}$ in model5.

m. How do you interpret the signs of β_3 and β_4?

n. Interpret the estimated coefficient on grant_{it} in model6.

o. Interpret the estimated coefficient on $\text{grant}_{i,t-1}$ in model6.

p. How do you interpret the signs of β_3 and β_4 now?

q. How do the significance results change from using the HAC (Arellano) significance results compared to OLS?