

Problem Set 3 Solutions

BUAN 6356

Due: Tuesday, 2017-10-24-11:59pm

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

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

- 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

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

- 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.

$$\text{lotsize } (p = 0.001823) \text{ \& } \text{sqrft } (p = 1.658e - 14)$$

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

$$\text{only sqrft } (p = 0.003406)$$

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

$$\text{lotsize } (p = 3.307e - 05) \text{ \& } \text{sqrft } (p = 5.006e - 11)$$

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

$$\text{lotsize } (p = 0.002243) \text{ \& } \text{sqrft } (p = 1.298e - 07)$$

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

They mitigate the heteroskedasticity.

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

$$\text{urate}_t = \beta_0 + \beta_1 \text{vrates}_t + e_t \quad (3)$$

- 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 vrates_t (4 tests total).
- Perform the level and trend KPSS tests on Δurate_t and Δvrates_t (4 tests total).
- Perform the level and trend KPSS tests on $\Delta(\Delta \text{urate}_t)$ and $\Delta(\Delta \text{vrates}_t)$ (4 tests total).
- Run the following first-difference model using the 'lm' function. Store the result in: model4

$$\Delta \text{urate}_t = \beta_0 + \beta_1 \Delta \text{vrates}_t + e_t \quad (4)$$

- 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 vacancy rate is significant or not (before we correct for nonstationarity)?

Yes ($p < 2.2e - 16$) and Yes ($p < 2.2e - 16$)

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

After first-differencing, urate_t is level stationary.

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

After first-differencing, vrates_t is level stationary.

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

It is insignificant ($p_{ols} = 0.79745, p_{nw} = 0.7342$)

j. Which model better describes the data?

We can only use linear models when the series is stationary.
So model4 is the better model.

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

$$d88_t = \begin{cases} 1 & \text{if year}_t = 1988 \\ 0 & \text{otherwise} \end{cases}$$

- Generate a new variable: d89

$$d89_t = \begin{cases} 1 & \text{if year}_t = 1989 \\ 0 & \text{otherwise} \end{cases}$$

- Generate a new variable indicating whether or not the firm had a grant last year: grant_{i,t-1}
[Hint: 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

$$\ln[\text{scrap}_{it}] = \alpha + \beta_1 d88_t + \beta_2 d89_t + \beta_3 \text{grant}_{it} + \beta_4 \text{grant}_{i,t-1} + e_t \quad (5)$$

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

$$\ln[\text{scrap}_{it}] = \alpha_i + \beta_1 d88_t + \beta_2 d89_t + \beta_3 \text{grant}_{it} + \beta_4 \text{grant}_{i,t-1} + e_t \quad (6)$$

- 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.

Getting a grant this period is associated with a 0.20% increase in the scrap rate.

- l. Interpret the estimated coefficient on grant_{i,t-1} in model5.

Getting a last period is associated with a 0.04% increase in the scrap rate.

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

The funding agency chose bad (high scrap rate) firms to get the grants.

Hence grants are positively correlated with the scrap rate.

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

Getting a grant this period is associated with a 0.25% decrease in the scrap rate.

- o. Interpret the estimated coefficient on grant_{i,t-1} in model6.

Getting a last period is associated with a 0.42% decrease in the scrap rate.

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

For a given firm, getting a grant is going to decrease the overall scrap rate.

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

Grant_{i,t-1} is significant in the OLS results at 5% but insignificant in the HAC model.