Problem Set 3

Buan 6340

Due: Fri 2019-07-20 at 11:59pm

Programming Questions

For these questions, you are going to need to create a simulation of linear modeling:

$$y_i = \beta_0 + x_i'\beta + e_i$$

where $x_i \sim iid\mathcal{N}_r\left(\mu_x, \Sigma_x\right)$ and $e_i \sim iid\mathcal{N}\left(0, \sigma^2\right)$ for $i = 1, \ldots, n$. You need to specify the various parameters in this simulation.

Question 1 In your simulation, estimate the mean squared prediction error using cross-validation. Compare your cross-validation MSPE to the actual MSPE. How does your estimation improve based on different choices made in the cross-validation procedure?

Question 2 In your simulation of linear modeling, perform t-tests on one of the βs using bootstrapping and compare them to the classical t-tests. How does your estimation improve based on the different choices made in the bootstrapping procedure?

Data Questions

Question 3 Using the data in wooldridge.db, extract the table for "jtrain". Analyze the jtrain panel data set using python. In particular, model the variable 'scrap' using the 'grant' variable. You can find information about these variables in the "jtrain_labels" table. Make sure that you at the very least do both a pooled OLS model and a fixed-effects model.