Preparation Assignment

Machine Learning for Financial Markets

Financial Services and Innovation Lab (FSIL),
2023 Spring

Deadline for both tasks: Sunday, 26 January 2023, 11:55 PM

BUT! Suggested deadline for programming task: before course drop ends.

1 Research Idea

Write an A4 page research proposal on

1. broad topic (Corporate events or asset pricing)
2. Identify (possible) data source and variables
3. list 1-2 methods/tool that you want to use and/or want to learn about
4. propose a specific research question.

Less than an A4 page is okay, more than that is not. Research questions modify a lot during research so you do not need to worry about that, but the broad topic and data sources are needed, you have to start somewhere and it is a process of learning what can and what can you not do.

2 Programming Task

Note that this helps you to evaluate your fit to the team. If this task takes you more than 30 minutes then you shall expect that the workload will take much more for you as we calculate. This is not a problem but you need to be willing to invest even more of your time. You can use any resource you wish even if you can work together. The point of this exercise is to help you to understand what type of skills are expected.

Write a script or function in R or Python (alternatively Julia, MatLab).

1. Generate the following random variables:
   (a) $X_1 \sim N(0,1)$, where $N(0,1)$ is the standard normal distribution
   (b) $X_2 = \alpha + 0.2X_1 + \epsilon$, where $\epsilon \sim U(-1,1)$, where U is the uniform distribution with support of -1 and 1
   (c) $Y = \delta + 1.2X_1 + \eta$, where $\eta \sim exp(\lambda)$, where $exp(\lambda)$ is the exponential distribution
   (d) Use $N = 10,000$ realizations and set the seed thus your results are reproducible.
   (e) Set the parameters as follows: $\alpha = 3, \delta = -2, \lambda = 0.5$
2. Estimate the following model with OLS and report the coefficients and their standard errors
   (a) $Y_i = \delta + \beta_1 X_{1,i} + \beta_2 X_{2,i}$
3. Estimate the same model, but use a rolling window:
   (a) take the first hundred observations: $i = 1, \ldots, 100$, then from the second to hundred: $i = 2, \ldots, 101$ and do the same until $i = 9,900, \ldots, 10,000$
   (b) This should result in 1000 estimates for all three coefficients.
   (c) Plot the distributions of all coefficients and add the mean as a vertical line to your plot.