Last Name: _____ First Name: _____

M.I.T. ID# _____

15.450 Final Exam

Professor Hui Chen

Spring 2020

Please read the following instructions carefully.

- Please write your name and MIT ID number on the first page.
- This exam lasts 90 minutes. You can take the exam anytime between 12:00 AM EDT and 10:00 PM EDT on 5/12/20.
- Submit the electronic version of your answers on Canvas by 10:00 PM EDT. If you are taking pictures of a hand-written exam, make sure to include all the pages.
- You can use the lecture notes and/or the textbooks listed on our syllabus during the exam. Do not consult anyone or use any on-line material.
- Always explain your answers and show your work, but try to be concise. Answers without explanations will receive no credit. Wrong answers with partially correct work may receive partial credit.

I hereby certify that I answered the questions in this exam without consulting anyone or any on-line material.

Signature: _____ Date: _____

1. Explain the main difference(s) between AIC and BIC for model selection.

2. Provide an example of a consistent but biased estimator, and an example of an inconsistent but unbiased estimator.

3. **True or false:** In an event study, one should try to specify the estimation window to be as long as possible in order to estimate the coefficients of the benchmark model accurately. (Please explain your answer.)

4. **True or false:** One weakness of the value-at-risk (VaR) measure is that it requires returns to be normally distributed. (Please explain your answer.)

5. You are building a time-series model to predict the daily return volatility of the S&P 500 ETF (ticker 'SPY').

   (a) How would you measure the variance of daily return σ_t^2 of SPY using intra-day data (e.g. 5-minute returns)?

   (b) Suppose you have constructed 10 potential features, x_{it} for $i = 1, \ldots, 10$, and you would like to construct a linear model to predict the (log) return variance on the next day:

$$\ln \sigma_{t+1}^2 = a + \sum_i b_i x_{it} + \epsilon_{t+1}.$$

   Explain how you would build the model using best subset selection.

6. Explain the meaning of the bias-variance trade-off in statistical learning. In particular, explain the meanings of the "bias" and "variance".

7. When would we prefer to use LASSO over ridge regression to fit a linear model?

8. You are helping a fintech startup building a model to predict the default risk of mortgage borrowers. You have collected a large dataset on borrower characteristics (such as income and leverage), market information (such as house prices and interest rates), and loan performances (default or non-default).

   (a) Explain how you would apply the KNN method to predict whether a borrower will default in one year.

   (b) What advantages and disadvantages does KNN have compared to a logit model?