Final Review

General Tips

- Review the interview and analytical questions from the problem sets carefully. Based on past exams, it is likely that some of these questions will show up on the exam. You might want to write your answer down for each of these on your cheat sheet.
- For the questions that involve coding, you probably won't be asked to write Python code with pencil and paper, but you should be able to explain the methodology. For example, the question might describe a setup (e.g. automated mortgage loan approval system) and ask you to explain how to build a model.

Basics

- What is an estimator?
- What is a random sample?

General Estimation Methods

- Method of moments
- Maximum likelihood
- GMM
- Bayes Estimation (MAP)
- GMM
- You might be asked to solve for an estimator in closed form if the problem is sufficiently simple (e.g. MLE for Bernoulli random variables). You need to understand how to formulate the problem (e.g. write down the likelihood function) and maximize (e.g. take the first-order condition and solve)

- You should write down the asymptotic distribution for these estimators on your cheat sheets

Consistent and Unbiased Estimators

- What does it mean for an estimator to be unbiased / consistent? You should know both the mathematical definition and how to explain these two concepts intuitively
- What's an example of a consistent but biased estimator?
- What's an example of an unbiased but inconsistent estimator?
- How can you show that an estimator is consistent? (Typically, you need to use the law of large numbers or use the results in class regarding general estimation methods)
- How can you show that an estimator is unbiased? (Typically, you need to do algebra using the properties of expectations)

Linear Regression Model

- OLS estimator
- Asymptotic distribution of OLS
- What key assumption do you need to show that OLS is unbiased? How can you prove that OLS is unbiased using this assumption?
- Gauss-Markov: what are the assumptions? what does it mean for an estimator to be BLUE?
- not BLUE does not imply biased
- Omitted variable bias

Event study

- You should be able to explain how to design an event study
- Size of the estimation window: tradeoff between reducing the variance of the estimated abnormal returns and issues related to non-stationarity

- How to deal with multiple firms announcing earnings on the same day?

Financial Time Series

- Stationarity (mathematical definition + intuition)
- Mean-reversion
- AR process
- MA process
- Random walk
- How to compute forecasts?
- How to compute standard errors of forecasts?
- How to compute autocovariance and autocorrelation?
- How to test for a unit root?
- How to deal with seasonality?

Estimators for classification models

- Logistic Regression
- KNN
- When do you want to use each?
- Confusion matrix, type I error, type II error and total error rate
- Choice of cutoff

Bias Variance Trade-off

- It is very likely that you'll get asked about this!
- Mathematical decomposition of MSE into bias, variance, and irreducible error
- Can you explain intuitively what bias is?
- Can you explain intuitively what variance is?

- If your model has high bias, what do you want to do?
- If you model has high variance, what do you want to do?
- Methods for regularizing linear models: Ridge and LASSO (what's the difference, when do you want to use one versus the other?)

Model Selection

- Metrics: R-squared, adjusted R-squared, AIC, BIC, accuracy, MSE, ROC curve
- What is the formula for each of these metrics?
- How do you use them?
- What is the difference between them?
- Methods: Best subset selection, Cross-validation for IID data and for time series data, Train test split, forward stepwise selection

Decision Trees

- Can you explain intuitively how decision trees are built?
- Can you explain the difference between random forest, boosting, and bagging?