

Preface

Statistical tools play pivotal role in modern finance. This course aims at demonstrating fundamental statistical techniques that are widely applied in financial data analysis. We begin with a conventional, relatively easy but critically important start of exploring data analysis in Chapter 1, and then move to linear regression models, which includes method, theory and diagnostics in Chapter 2. Factor models and PCA are addressed in Chapter 3. Statistical analysis of portfolio theory and CAPM models are the contents of Chapters 4-5. In the second half of the semester, more advanced topics on parametric, Bayesian methods and nonparametric models will be taught in Chapters 6-7. We end with the statistical methods in risk management (Chapter 8). Depending on the pace of the teaching, we may continue with other topics, such as statistical analysis of trading.

Since our program offers courses on time series and financial derivatives, this course purposely avoided these subjects, though they are extremely important in statistics and finance.

As expected, the examples in this course are related with financial markets. R is the statistical software used in this course. Though the lecture does not teach R coding, students are required to practice data analysis with R.

The course covers both statistics and finance topics. We tend to put slightly more weight on statistical methods than on financial models.

Other than the lecture notes, we use two popular textbooks: "Statistics and Data Analysis for Financial Engineering" by D. Ruppert (RU) and "Statistical Models and Methods for Financial Markets" by T.L. Lai and H. Xing (LX). They are adopted by the master programs in financial engineering in many universities including Cornell and Stanford. All the topics in course are covered in these two books. We try to select topics and levels that are best fit to our students.

Some general words of advice:

- A. "It's tough to make predictions, especially about the future." (quoting Yogi Berra).
- B. "In God we trust, everything else bring data" (quoting Edwards Deming).
- C. "All models are wrong, some are useful" (quoting George Box).
- D. "Occam's Razor" or "Simplicity is beauty" or "Everything else the same, bring the simplest model". Vote for parsimonious models or bet on sparsity.
- E. Financial data are *generally* non-stationary, and non-normal as well. As a result, except for cases such as limit order book data, there exists considerable risk of generalization.
- F. Trade-off between bias and variance. This is one rule of thumb in model selection. Always keeping in mind of the risk of overfitting.
- G. Develop intimate feeling about ... data.