MS&E 349 - Spring 2018
Financial Statistics

March 30, 2018

General Information

- **Class Meeting Times:** Tuesdays and Thursdays 10:30am-11:50am in room 380-380F
- **Instructor:** Markus Pelger, 312 Huang Engineering Center, mpelger@stanford.edu
- **Office hours:** Tuesdays 1:30-3:30pm in Huang 312
- **Target audience:** Ph.D. students.

Short Description

This Ph.D. course covers topics in financial statistics with a focus on current research. Topics will include time-series modeling, volatility modeling, high-frequency statistics, large dimensional factor modeling and estimation of continuous time processes. The course requires a strong background in statistics and mathematics and some knowledge of economics and finance. The evaluation will be based on group work, which includes two homework assignments and a final project.

Course Outline

Financial statistics is the intersection of statistical techniques and finance. Financial statistics provides a set of tools that are useful for modeling financial data and testing beliefs about how markets work and prices are formed. Conversely, new techniques in analyzing financial data can lead to empirical facts inconsistent with existing theories, begging for new models or investment strategies.

   The goal of this course is to introduce students to the frontiers of financial statistics and financial econometrics. The course will focus on the statistical theory of the estimation approaches, but we will also spent time covering a range of significant applications to the estimators. As we often have a large number of observations for financial data, we will develop an asymptotic inferential theory for our estimators.

   The course is not intended as an introduction class to financial statistics, but to lead Ph.D. students to the research frontiers of statistically analyzing financial data. It is not possible

within a quarter to cover all relevant topics in financial statistics in depth. Hence we will focus on a selected number of topics:

- Linear-time series models: ARIMA Models

- Volatility Modeling: GARCH

- General theory of extremum estimators: Probabilistic models, consistency, asymptotic normality, test statistics, maximum likelihood and general method of moments

- Linear factor modeling: Test statistics and estimation

- High-dimensional factor modeling: Curse of dimensionality, principal component analysis, random matrix theory and spiked covariance models, number of factors

- High-frequency statistics: Limit theorems, non-parametric volatility and jump estimation

- Machine-learning empirical asset pricing

Requirements

Grading: The weights for grading are

- Project 70%
- Homework 20%
- Participation 10%.

Both, the homework assignment and the project, should be done in groups. The group size will depend on the enrollment in the class, but the target number of students should be two per group. You should choose the same groups for the project and the homework. The homework assignments focus on the empirical implementation of the material covered in class. You have four weeks to complete each homework assignment. Please submit also the commented and executable code with each homework assignment. The homework will be graded on a 5 point scale.

The final project is one of the most important parts of the course. You have the choice between a number of paper that are at the research frontier in financial statistics. The goal is to prepare a 35 minutes presentation, lecture notes and an empirical application for the specific estimation method. Here are the details:

- There are more than 15 different papers to choose from. If you are interested in a paper that is not on the list, please talk with me. If it is appropriate for this class I am happy to include it. Each project will be based on one paper. Select your topic by April 24th.

  - The proposed list of papers is at the end of this syllabus.
  - Your contribution will consist of two elements: a presentation and a final report
  - Final report:
    * Prepare a set of lecture notes based on the paper. These lecture notes should be understandable for your classmates. There is no prescribed page number, but around 10 pages seems to be an appropriate number for summarizing the

paper. You should not just simply summarize some theorems, but also explain the reasoning behind them. It is not necessary to include all technicalities of the proofs, but you should provide the intuition for the proofs.

* Choose an empirical application for the method. I will help you to find the data to apply it to. Simulate an appropriate model and test your estimation code on the simulated data. Then apply your estimator to the data set and interpret the results. Include your simulation and empirical analysis in the final project report (this will be in addition to the 10 pages of the theory summary). Also include the commented code for the simulation and empirical analysis in an appendix to your report. Any programming language (Matlab, R, Python, etc.) is fine. For some papers it might be hard to find the appropriate data. In this case simulate an appropriate model and test the estimation.

* Submit your final report 4 days before the presentation day by email. I will distribute it to the class. All students are supposed to read the report before the presentation.

- Lecture:

* Hold an 35 minute presentation on your topic. This presentation should include the intuition of the theory behind the estimation approach and your empirical results. Every member of your group should be part of the presentation.

* Presentation slides need to be submitted by 5pm the day before the presentation. I will distribute the slides to the class.

- You have to meet with me as a group this quarter to discuss your final project.
- Active participation in class and during the student presentations will be rewarded.

Prerequisites

The course requires a strong background in statistics and mathematics and some knowledge of economics and finance.

The prerequisite of this course is knowledge of probability and stochastic processes at the level of MS&E 321 or 322 or similar and of statistics on the level of STATS 200 or similar. Knowledge of financial engineering at the level of MS&E 245A, MS&E 245B, MATH 180, MATH 240 or FINANCE 622 is highly desirable. Knowledge of some programming language or mathematical/statistical software package such as MATLAB, Mathematica, R or Python is required.

In view of the aforementioned prerequisites, you are expected to be familiar with the following concepts/terminology:

- Large-sample theory for least-squares regression
- Maximum likelihood estimation
- Point and interval estimation, tests of hypotheses
- Returns of financial assets, stock returns, fixed-income securities

- Sample space, events, probability axioms, basic rules of probability, independence, basic counting arguments, conditional probability, Bayes Theorem, random variable,
- probability density function (pdf), cumulative distribution function (cdf), expected value, moments, moment generating function, variance, standard deviation, covariance.
- Markov's inequality, convergence in probability, (weak) law of large numbers.
- Convergence in distribution, central limit theorem, confidence intervals.
- Geometry and algebra of vectors, matrix operations, determinants, (linear) subspaces of R^n, eigenvalues and eigenvectors, orthogonality.
- Limits, continuity, derivatives, integrals, mean value theorem, fundamental theorem of calculus, Taylor series, uniform convergence.
- Brownian motion and Poisson processes
- Markov chains

I have prepared some review material to refresh your knowledge for some of the above topics. Please talk with me if you are not sure if you meet the above requirements.

Course Material

Unfortunately, there is not Ph.D. level textbook for financial statistics. The course readings are assigned to each topic based on book chapters, papers and notes.

The following textbooks are good introductory master level textbooks:

- Tsay: Analysis of Financial Time Series
- Ruppert and Matteson: Statistics and Data Analysis for Financial Engineering
- Fan: The Elements of Financial Econometrics
- Lai: Statistical Models and Methods for Financial Markets
- Carmona: Statistical Analysis of Financial Data in R
- Campbell, Lo and MacKinlay: The Econometrics of Financial Markets

These are advanced graduate textbooks that cover some of the topics in this course

- Singleton: Empirical Dynamic Asset Pricing
- Aït Sahalia and Hansen: Handbook of Financial Econometrics, vol I and II
- Newey and McFadden: Handbook of Econometrics, Large Sample Estimation and Hypothesis Testing
- Aït Sahalia and Jacod: High-Frequency Financial Econometrics
- Bai, Yao and Zheng: Large Sample Covariance Matrices and High-Dimensional Data Analysis
- Cochrane: Asset Pricing
- Hamilton: Time Series Analysis

Detailed Course Outline and Reading List

The readings are categorized into required readings and recommended additional readings:

1. **Review**
   Required readings:
   - Review: Introduction to Financial Data
   - Review: Fixed Income Securities

   Recommended readings:
   - Review Lecture Notes: Regression and Asymptotics
   - Review Lecture Notes: Introduction to Levy processes with applications in finance

2. **Time-Series**
   Required readings:
   - Lecture Notes: Time-Series
   - Lecture Slides: Linear Time Series

   Recommended readings:
   - Tsay (Analysis of Financial Time Series 2002): Chapter 2
   - Cochrane (Time Series for Macroeconomics and Finance 2005): Chapters 1-6

3. **Volatility**
   Required readings:
   - Lecture Slides: Volatility
   - Andersen, Bollerslev and Diebold (Handbook of Financial Econometrics 2010): Parametric and Nonparametric Volatility Measurement

   Recommended readings:
   - Tsay (Analysis of Financial Time Series 2002): Chapter 3
   - Fan (The Elements of Financial Econometrics 2014): Chapter 3

4. **Extremum Estimation**
   Required readings:
   - Newey and McFadden: Large Sample Estimation and Hypothesis Testing

   Recommended readings:
   - Singleton (Empirical Dynamic Asset Pricing): Chapters 1-4
   - Hansen (Econometrica 1982): Large Sample Properties of Generalized Method of Moments Estimators

5. **Linear Factor Modeling**
   Required readings:
   - Lecture Slides: Linear Factor Modeling
   - Jagannathan, Skoulakis and Wang (Handbook of Financial Econometrics 2010): The Analysis of the Cross-Section of Security Returns

   Recommended readings:

- Cochrane (Asset Pricing 2000): Chapter 12

6. **High-Dimensional Factor Modeling**
   Required readings:

   - Bai and Ng (Foundations and Trends in Econometrics 2008): Large Dimensional Factor Analysis
   - Bai, J. (Econometrica 2003). Inferential Theory for Factor Models of Large Dimensions
   - Lettau, M. and Pelger, M. (2018). Estimating Latent Asset-Pricing Factors
   - Lettau, M. and Pelger, M. (2018). Factors that Fit the Time Series and Cross-Section of Stock Returns
   - Pelger, M. and Xiong (2018). State-Varying Factor Models of Large Dimensions
   - Pelger, M. and Xiong (2018). Interpretable Proximate Factors for Large Dimensions

   Recommended readings:

   - Tsay (Analysis of Financial Time Series 2002): Chapter 11
   - Ruppert and Matteson (Statistics and Data Analysis for Financial Engineering 2015): Chapter 18
   - Fan, Liao and Mincheva (Journal of the Royal Statistical Society 2013): Large covariance estimation by thresholding principal orthogonal complements
   - Bai, Yao and Zheng (Large Sample Covariance Matrices and High-Dimensional Data Analysis 2015): Chapter 1-12
   - Benaych-Georges and Nadakuditi (Advances in Mathematics 2011): Eigenvalues and eigenvectors of low rank perturbation
   - Ahn and Horenstein (Econometrica 2013): Eigenvalue Ratio Test for the Number of Factors
   - Bai (Econometrica 2003): Inferential Theory for Factor Models of Large Dimensions
   - Bai and Ng (Econometrica 2002): Determining the number of factors in approximate factor models
   - Onatksi Determining the Number of Factors from Empirical Distribution of Eigenvalues
   - Bai, Yao and Zheng (Large Sample Covariance Matrices and High-Dimensional Data Analysis 2015): Chapter 11

7. **High-Frequency Statistics**
   Required readings:

   - Podolskij and Vetter (2009): Understanding limit theorems for semimartingales: a short survey
   - Aït-Sahalia and Jacod (Journal of Economic Literature 2012): Analyzing the Spectrum of Asset Returns: Jump and Volatility Components in High Frequency Data
   - Lee and Mykland (Review of Financial Studies 2008): Jumps in Financial Markets: A New Nonparametric Test and Jump Dynamics

– Pelger (2018): Large-dimensional factor modeling based on high-frequency observations
– Pelger (2018): Understanding Systematic Risk: A High-Frequency Approach

Recommended readings:

– Barndorff-Nielsen and Shephard (Journal of Financial Econometrics 2006): Econometrics of Testing for Jumps in Financial Economics Using Bipower Variation
– Aït-Sahalia and Jacod (High-Frequency Financial Econometrics 2014): Chapters 1-4, 6-8, 10
– Christensen, Oomen, Podolskij (Journal of Financial Economics 2014): Fact or friction: Jumps at ultra high frequency
– Aït-Sahalia and Jacod (The Annals of Statistics 2009): Testing for Jumps in a Discretely Observed Process

Papers for Final Project (Choose one paper)

1. Large-dimensional factor modeling

   – Fan, J., Liao, Y. and Wang, W. (2016). Projected Principal Component Analysis in Factor Models. Annals of Statistics
   – Bai, J. and Ng, S. (2017). Principal Components and Regularized Estimation of Factor Models. Working paper
   – Fan, J., Liao, Y. and Micheva, M. (2013). Large Covariance Estimation by Thresholding Principal Orthogonal Complements. Journal of Royal Statistical Society
   – Fan, J., Xue, L. and Yao, J. (2017). Sufficient forecasting using factor models. Journal of Econometrics

2. Testing of asset-pricing models

   – Chen, L. and Pelger, M. (2018). Testing Asset Pricing Models for Many Assets. Working paper
   – Bryzgalova, S. (2016). Spurious Factors in Linear Asset Pricing Models. Working paper
   – Kelly, B. and S. Pruitt (2015). The three-pass regression filter: A new approach to forecasting using many predictors. Journal of Econometrics
   – Giglio, S. W. and Xiu, D. (2016). Inference on risk premia in the presence of omitted factors.
   – Fan, J., Liao, Y. and Yao, J. (2015). Power Enhancement in High Dimensional Cross-Sectional Tests. Econometrica

3. High-dimensional asset pricing

   – Kelly, B., S. Pruitt, and Y. Su (2017). Characteristics are risk exposures. Working paper
   – Feng, G, Giglio, S. W. and Xiu, D. (2017). Taming the Factor Zoo. Working paper

- Freyberger, J., Neuhierl, A. and Weber, M, (2017) Dissecting Characteristics Nonparametrically. Working paper
- Kozak, S., Nagel, S. and Santosh, S. (2017). Shrinking the Cross-Section. Working paper

4. High-frequency statistics

- Aï-Sahalia, Y. and Xiu, D. (2017). Using Principal Component Analysis to Estimate a High Dimensional Factor Model with High-Frequency Data. Journal of Econometrics
- Li, J., Tauchen, G., Lin, H. and Todorov, V. (2018) Rank Tests at Jump Events. Journal of Business and Economic Statistics.
- Andersen, T. G., Fusari, N. and Todorov, A. (2017). The Pricing of Short-Term Market Risk: Evidence from Weekly Options

5. Covariance matrix estimation

- Fan, J., Han, F., Liu, H., and Vickers, B. (2016). Robust Inference of Risks of Large Portfolios. Journal of Econometrics
- Fan, J., Wang, W, and Zhong, Y. (2017). Robust Covariance Estimation for Approximate Factor Models. Journal of Econometrics
- Ledoit, O. and Wolf, M. (2004). A Well-Conditioned Estimator for Large-Dimensional Covariance Matrices. Journal of Multivariate Analysis,

Preliminary Timetable

Date	Day	Week	Topic	Due Dates
4/3/17	Tue	1	Introduction	
4/5/17	Thu	1	Review Linear Time-Series Models	
4/10/17	Tue	2	Volatility Modeling: GARCH	
4/12/17	Thu	2	Extremum Estimators: Probability Model	Selection of group
4/17/17	Tue	3	Extremum Estimators: Consistency	
4/19/17	Thu	3	Extremum Estimators: Asymptotic Normality	
4/24/17	Tue	4	Extremum Estimators: Test Statistics	Selection of project
4/26/17	Thu	4	Linear Factor Modeling	
5/1/17	Tue	5	High-Dimensional Statistical Factor Modeling: Strong Factors	
5/3/17	Thu	5	High-Dimensional Statistical Factor Modeling: Weak Factors	
5/8/17	Tue	6	High-Dimensional Statistical Factor Modeling: Number of Factors	HW 1 due
5/10/17	Thu	6	High-Frequency Statistics: Limit Theorems	
5/15/17	Tue	7	High-Frequency Statistics: Jumps	
5/17/17	Thu	7	Large Dimensional High-Frequency Statistics	
5/22/17	Tue	8	Student presentation	
5/24/17	Thu	8	Student presentation	
5/29/17	Tue	9	Student presentation	
5/31/17	Thu	9	Student presentation	
6/5/17	Tue	10	Research Discussion	HW 2 due

Learning Outcomes

1. Students should be able to derive the statistical properties for the most important estimators in financial statistics. Based on these techniques students should be in the position to develop their own new estimators and derive their properties.
2. Students should be able to weigh the advantages and disadvantages of different estimation approaches and choose the appropriate technique for an application.
3. Students should become familiar with working with financial data.
4. Students should be able to learn new estimation methods on their own by reading the relevant literature and to present their results in oral and written form.

Related Courses

If you are looking for an introductory class to financial statistics I would recommend the following classes. They focus more on discrete time-series modeling, while we will have a stronger focus on continuous-time models.

– STATS 240: Statistical Methods in Finance
– STATS 244: Quantitative Trading: Algorithms, Data and Optimization
– STATS 207: Introduction to Time Series Analysis
– ECON 271: Intermediate Econometrics I
– ECON 272: Intermediate Econometrics III

The following Ph.D. courses are complementary to this class and deal with related problems:

– FINANCE 632: Empirical Dynamic Asset Pricing
– MGTECON 603 and 604: Econometric Methods I and II
– ECON 273: Advanced Econometrics I
– ECON 274: Advanced Econometrics II
– ECON 276: Computational Econometrics
– MS&E 347: Credit Risk: Modeling and Management

Honor Code

We expect the students to create an environment in which honor code violation is not to be tolerated.

Students with documented disabilities

Students who have a physical or mental impairment that may necessitate an academic accommodation or the use of auxiliary aids and services in a class must initiate the request

with the Disability Resource Center (DRC). The DRC will evaluate the request along with the required documentation, recommend appropriate accommodations, and prepare a verification letter dated in the current academic term in which the request is being made. Please contact the DRC as soon as possible; timely notice is needed to arrange for appropriate accommodations. The DRC is located at 123 Meyer Library (phone 723-1066 Voice; 725-1067 TTY).