Time Series Analysis for Finance

MAP 5117, Spring 2021 MWF 5:00-5:50 pm

Instructor: Enrique Villamor.

Office: DM 414B.

Phone: 305-785-3238.

Email: villamor@fiu.edu.

Office hours: MWF 3:30 pm – 5:00 pm

Prerequisites:

Working knowledge of Stochastic Calculus, Differential Equations, Linear Algebra, Basic Statistics and Probability.

Textbook: Advances in Financial machine Learning by Marcos Lopez de Prado

Content:

Machine learning (ML) is changing virtually every aspect of our lives. Today ML algorithms accomplish tasks that until recently only expert humans could perform. As it relates to finance, this is the most exciting time to adopt a disruptive technology that will transform how everyone invests for generations. This book explains scientifically sound ML tools that have worked for me over the course of two decades, and have helped me to manage large pools of funds for some of the most demanding institutional investors.

Books about investments largely fall in one of two categories. On one hand we find books written by authors who have not practiced what they teach. They contain extremely elegant mathematics that describes a world that does not exist. Just because a theorem is true in a logical sense does not mean it is true in a physical sense. On the other hand we find books written by authors who offer explanations absent of any rigorous academic theory. They misuse mathematical tools to describe actual observations. Their models are overfit and fail when implemented. Academic investigation and publication are divorced from practical application to financial markets, and many applications in the trading/investment world are not grounded in proper science.

A first motivation for writing this book is to cross the proverbial divide that separates academia and the industry. I have been on both sides of the rift, and I understand how difficult it is to cross it and how easy it is to get entrenched on one side. Virtue is in the balance. This book will not advocate a theory merely because of its mathematical beauty, and will not propose a solution just because it appears to work. My goal is to transmit the kind of knowledge that only comes from experience, formalized in a rigorous manner.

A second motivation is inspired by the desire that finance serves a purpose. Over the years some of my articles, published in academic journals and newspapers, have expressed my displeasure with the current role that finance plays in our society. Investors are lured to gamble their wealth on wild hunches originated by charlatans and encouraged by mass media. One day in the near future, ML will dominate finance, science will curtail guessing, and investing will not mean gambling. I would like the reader to play a part in that revolution.

A third motivation is that many investors fail to grasp the complexity of ML applications to investments. This seems to be particularly true for discretionary rms moving into the "quantamental" space. I am afraid their high expectations will not be met, not because ML failed, but because they used ML incorrectly. Over the coming years, many rms will invest with off-the-shelf ML algorithms, directly imported from academia or Silicon Valley, and my forecast is that they will lose money (to better ML solutions). Beating the wisdom of the crowds is harder than recognizing faces or driving cars. With this book my hope is that you will learn how to solve some of the challenges that make finance a particularly difficult playground for ML, like backtest overfitting. Financial ML is a subject in its own right, related to but separate from standard ML, and this book unravels it for you.

This book disentangles a web of interconnected topics and presents them in an ordered fashion. Each chapter assumes that you have read the previous ones. Part 1 will help you structure your financial data in a way that is amenable to ML algorithms. Part 2 discusses how to do research with ML algorithms on that data. Here the emphasis is on doing research and making an actual discovery through a scientific process, as opposed to searching aimlessly until some serendipitous (likely false) result pops up. Part 3 explains how to backtest your discovery and evaluate the prob- ability that it is false.

These three parts give an overview of the entire process, from data analysis to model research to discovery evaluation.

Chapter 1: Financial Machine Learning as a Distinct Subject

PART 1: DATA ANALYSIS

Chapter 2: Financial Data Structures

Chapter 3: Labeling

Chapter 4: Sample Weights

Chapter **5: Fractionally Differentiated Features**

PART 2: MODELLING

Chapter 6: Ensemble Methods

Chapter 7: Cross-Validation in Finance

Chapter **8: Feature Importance**

Chapter 9: Hyper-Parameter Tuning with Cross-Validation

PART 3: BACKTESTING

Chapter 10: Bet Sizing

Chapter 11: The Dangers of Backtesting

Chapter 12: Backtesting through Cross-Validation

Chapter 13: Backtesting on Synthetic Data

Chapter 14: Backtest Statistics

Chapter 15: Understanding Strategy Risk

Chapter 16: Machine Learning Asset Allocation

Homeworks, exams, grades:

There will be 6 or 7 homework assignments, and a final exam. Grades will be based on the HW (1/2) and the final (1/2). Collaboration on HW is encouraged (homeworks are not exams), but students must write up and turn in their solutions individually. If you work closely with another student (or someone else, or you use other resources such as something on the web) please identify your collaborators and/or sources on your solution sheet. HW may be turned in late only by securing permission before it is due; no credit will be given for HW turned in after a solution sheet has been distributed. The HW is important not only because it counts as part of the grade, but also because it helps you master the material (and therefore helps you do well on the exam). The final will be closed-book, but you may bring two 8.5 × 11 1 sheets of notes (both sides, any font). Requests to take a makeup exam must be made in advance, and will not be granted for matters of personal convenience.

Semester plan:

Weeks 1-3: Part 1. Data Analysis. Weeks 4-7: Part 2: Modeling. Weeks 8-16: Part 3: Backtesting.