Applying Data and Analytics in Private Markets Investing

Lecturer: Rafael Nicolas Fermin Cota

Session : Semester 1

Description

It is becoming increasingly common for private equity firms to collect huge amounts of data over time. The incoming data is diverse - balance sheets, income statements, statements of cash flow, operating metrics, debt schedules, budgets vs. actual, segment data, month/quarterly/annual data points, etc. The challenge right now is to manage portfolio companies' data efficiently and make that data ready for forecasting, but before a single estimate can be calculated, there are a number of steps that need to be taken. These steps include but are not limited to: (i) collecting and structuring the data, (ii) cleaning and standardizing the data, (iii) linking the data with other data sets, (iv) generating aggregations or new data attributes, (v) storing it in a database, and (vi) making it available for investment decisions. In this course, we will look at some R packages and methods that have been developed to handle the analysis of large collections of time-series. In class I use a simple prescription: Think big, start small, and grow fast. Students will start by collecting data from a portfolio of companies and store it in a easy-to-access format. Once they have collected all the data from databases and spreadsheets, they will undertake a process of building, testing, and iterating forecasting models, then rapidly scaling them to achieve maximum performance.

Course Outline

Module 1: Wrangling

The aim for this module is to teach students different tidy data techniques to clean, structure, aggregate, and preprocess time-series data for forecasting and machine learning prediction. The foundation for tidy data management is the tidyverse[1], a collection of libraries, that work in harmony, are built for scalability, and are taught in this module. In class, I also introduce students to the dbplyr package[2], which is the database backend for dplyr that allows students to easily use remote database tables as if they are in-memory data frames by automatically converting dplyr code into SQL.

Module 2: Forecasting

The dramatic increase in the availability of large collections of time series raises the need for developing reliable efficient and automatic algorithms for forecasting. Selecting the most appropriate forecast model or a combination of models to use in forecasting is a challenging task. Such problems typically lack algebraic expressions. It is not possible to calculate derivative information, and the problem may exhibit uncertainty or noise. Expert knowledge is required. Even with the necessary knowledge

[1] https://www.tidyverse.org/
[2] https://dbplyr.tidyverse.org/

and skills, success is not guaranteed. Since prior expert knowledge is often expensive, not always readily available, and subject to bias and personal preferences, meta-learning (e.g., time series features) can serve as a promising complement to this form of advice through the automatic accumulation of experience based on the performance of multiple applications of a learning system. In this module, we use the modeltime[3] workflow for forecast model selection using time series features.

Sample Lectures/Projects

[1] Airbnb Data Analysis for Singapore:
https://rpubs.com/rafael_nicolas/airbnb_singapore_data_analysis

[2] Treasure trove of data science recipes to help students achieve operational excellence, and develop competitive edge in the private equity industry:
https://rpubs.com/rafael_nicolas/nus_pe_ds_course

[3] Prediction for at least N time series and at least one model for the M5 Walmart competition: https://www.linkedin.com/posts/rnfc_dba4761m5finalproject-activity-6875925555452153856-MCOz

Optional Reading List

[1] R for Data Science: https://r4ds.had.co.nz/

[2] Spreadsheet Munging Strategies: https://nacnudus.github.io/spreadsheet-munging-strategies/

[3] Tidy Modeling with R: https://www.tmwr.org/

[4] Feature Engineering and Selection: https://bookdown.org/max/FES/

[5] Forecasting: Principles and Practice: https://otexts.com/fpp3/

Assessment

Continuous Assessment:

Class Contribution	20%
Individual Assignments	40%
Group Project	40%

[3] https://cran.r-project.org/web/packages/modeltime/vignettes/getting-started-with-modeltime.html