Learning-To-Rank Stocks for Portfolio Rebalancing Strategies

Lluís Navarro Girbés

València, Julio 2022

KSCHOOL

**Modeling a Score Function for Stock Ranking and Wealth Management**

Main Purpose

▶ To Improve Indexed Trading (ETFs)

▶ Previous Financial Literature, classical methods (Non-ML) and heuristical prediction. "Screening Rules and Portfolio Performance", 2019 (NAJEF). US Stocks Dataset

▶ Main Differences with Previous Literature/Research:

Quantitative Finance Use of Alternative Risk/Performance Measures (**PMs**) as predictors.
ML for inferring the most efficient intertemporal rankings supplied by Learning-To-Rank (**LTR**) algorithms.

Data Sience/ML **LTR** application out of the usual scope *e-commerce, web search engines, entertainment*.
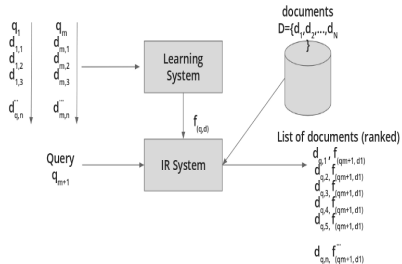
KSCHOOL
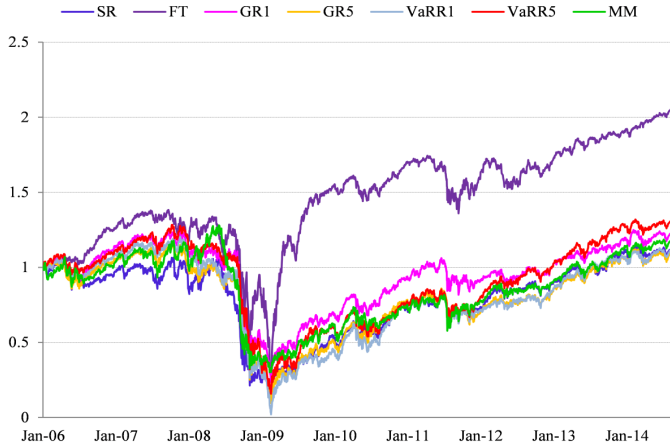
(a) Cumulative Wealth of Several **PM**-based Heuristical Strategies
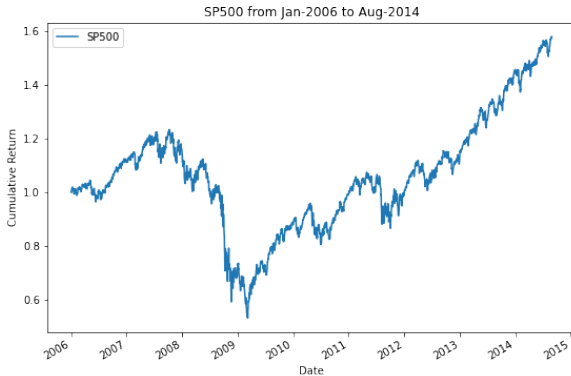
(b) Cumulative Wealth of (replicating Strategy) SP500
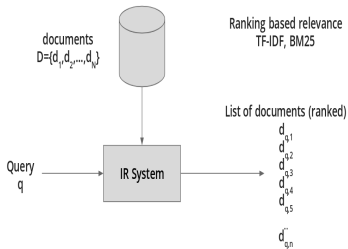
(c) Query (Rebalance Event), Documents (Stocks), Rankings

(d) LTR execution flow

KSCHOOL

The Dataset

- ▶ Daily **end-of-day quotes**, **market capitalizations** and **other** financial information publicly available information of the stocks composing S&P500 (EUROSTOXX600).
- ▶ **Maximum Time-Series Period**: The maximum possible avoiding delisting effects. *Estimated:* January 2005-Now
- ▶ **Source and Methodology:** Yahoo Finance public API (Python library `yfinance`, MIT licence) and Web Scrapping of `Y! Finance`, `Wikipedia` (`marketscreener`) websites for stock lists.
- ▶ ICYMI: `nasdaqdatalink`, `alphavantage` APIs. E.g. `quandl` (`nasdaqdatalink`) free database `WIKI\PRICES` w/ 1,000 stock historical series.
- ▶ **Remark:** `Kaggle` datasets already exist. Avoided from the beginning.

Preliminar Exploratory Analysis

- ▶ Significant **non-gaussianity** for all stock returns are found: this fact motivates the use of Alternative **PMs** as predictors.
- ▶ Individual stocks: Serial correlation for higher moments r_t^n, uncorrelated raw returns r_t.
- ▶ Strong cross-correlation between some different assets
- ▶ Some assets behave "very similar".

KSCHOOL

Targets

- ▶ Show if **PMs** (individual or simple combinations) are relevant for Portfolio Construction and Wealth Management.
- ▶ **Predict Scores** and then **Rankings of Relevant Assets** in order to design dynamic Portfolio Rebalancing Strategies based on these relevances.
- ▶ Construct a Frontend showing the Realised (track-record) and predicted metrics (financial or data-science driven), etc.
  In sum, **develop a KID (Key Information Dashboard)** summaryzing all the relevant (future and realized) financials including portfolio selection.

State-Of-The-Art

Finance:

- ▶ Very scarce works.
- ▶ Features based on a simple combination of Technical Indicators (rolling simple stats of returns $f(\text{prices})$).
- ▶ New Sentiment indicators (as features/predictors).

Data Science:

- ▶ Out of the Original Scope: A Novel or Less-Known application.

KSCHOOL

Main technical bullets

Data Acquisition and Wrangling : `yfinance (nasdaqdatalink, alphavantage), pandas`

Exploratory (Time-Series) Data Analysis : `statmodels, arima, pd, matplotlib, sns`

Supervised LTR (Modeling) :

- ▶ *Pairwise*: RANKNET (`keras-tf`) Deep Learning, LAMBDAMART (`XGBoost`)
- ▶ *Listwise*: LISTNET, LISTMLE (`tf`), Deep Learning
- ▶ *All In*: `pytorchltr (torch)`, Deep Learning

Frontend : Financial Dashboard with Key Information (KID) (`streamlit`)

Expected Results

- ▶ Significative gain over Heuristical-based Portfolio Rebalancing
- ▶ A ML-based recommender supported by sofisticated PMs (provided by Risk Management literature) as features (not technical indicators).
- ▶ A nice/clear KID to help invest/hedge decisions: Equity Desk (superhedging, CIB recommendations), Private Banking (retail recommendations)

KSCHOOL

Disaster Recovery Plan

Data:
- Insufficient data or too many delisting/additions (survivorship bias): SP500 → STOXX600 → WIKI\PRICES (and change bench) or `alphavantage` API.
- Last Resource: Use `Kaggle` dataset (updated daily) with a complete analysis of the survivors on the time sample given.

Modeling: Empirical Results not conclusive (or **LTR** not suitable for the problem).
- Try **R-t-R** (regress then rank).
- **Techs:** LSTM, CNN (`torch`, `tf`) (RNN, Deep Learning)
- Last Resource: Regress or Classify (Forecast) SP500 (index) or representative stocks.

KSCHOOL

Mandatory update and/or reschedule each end-of-week

Time Schedule

Data: Jul-25/Aug-12 (3w)
- ► Managing Delisting/Addtions,
- ► EDA raw-returns,
- ► Features (**PMs**) construction, analysis (dimensionality reduction)

Modeling, Part I (MVP): Aug-16/Sep-09 (3w, 1 week-out)
- ► Understand LTR basics (algorithm, metrics).
- ► Accurate mapping of the [Queries,Relevant Documents] to the Intertemporal Portfolio problem: [Rebalance Events,Relevant Assets]
- ► Implement RANKNET
- ► Empirical Analysis. Conclusions. KID Prototype. Memo guidelines.

Modeling, Part II (Generalization): Sep-12/Oct-07 (4w)
- ► Rest of the LTR Implementations (pairwise & listwise).
- ► Empirical comparison with benchmark (random/naive baselines)
- ► KID generalisation. Memo final skeleton

Final Stage, KID (Frontend) Oct-10/Nov-11 (5w)
- ► Design-Test final KID.
- ► Memo/Presentation

KSCHOOL

León et al. (2019), Screening Rules and Portfolio Performance. *North American Journal of Economics and Finance*

León et al. (2020), Modeling asset returns under time-varying semi-nonparametric distributions, *Journal of Banking and Finance*.

Poh et al., (2022). Enhancing Cross-Sectional Currency Strategies by Context-Aware Learning to Rank with Self-Attention. *The Journal of Financial Data Science*

Song, et al., (2017). Stock Portfolio Selection Using Learning-to-Rank Algorithms with News Sentiments, *Neurocomputing*

Figure: La Persistència de la Memòria.