Introduction to Quantitative Research

Disclaimer

This workshop is for educational purposes only and does not constitute an offer to sell, a solicitation to buy, or a recommendation for any security; nor does it constitute an offer to provide investment advisory or other services by the speakers. Nothing contained herein constitutes investment advice or offers any opinion with respect to the suitability of any security and any views expressed herein should not be taken as advice to buy, sell, or hold any security or as an endorsement of any security or company. The speakers are not responsible for the losses incurred due to the buying and selling of securities.

Outline

- Looking for Ideas
- Framing a question for research
- Research structure
- Collecting the right data
- Financial Machine Learning
- Some common pitfalls
- Useful resources of research

Finding Ideas

The way to inspired for ideas is to read research papers (mostly academic and better if peer reviewed), peer reviewed journals and research articles. Some sources are listed below -

Research paper libraries

- SSRN: https://www.ssrn.com/
- arXiv: https://arxiv.org/
- RePEc: http://repec.org/

Journals

- The Journal of Finance
- Journal of Portfolio Management
- Journal of Financial Data Science

Newsletter

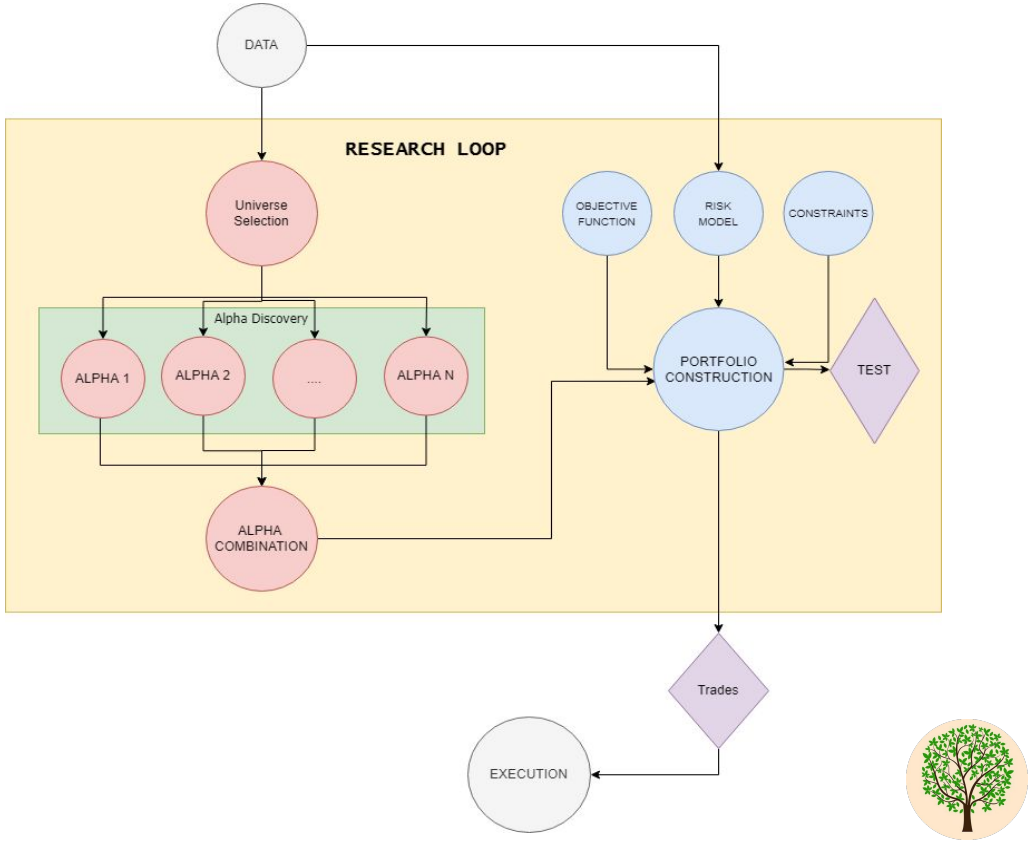
- Paper Digest: https://www.paperdigest.org/
- Quantocracy: https://quantocracy.com/

Research Question

- A research question is a question used for frame the main hypothesis of your research for the answer you are seeking for.
- A good research question is the key for doing quantitative research. Without a clear research question the focus of the research might shift and may not lead to anything useful and indeed may lead to some false discovery.
- Vague question must be avoided at all cost e.g. “Does momentum strategy work ?”.
- A research question must clear, precise and provide specific information for a reader to clearly understand it e.g.: “Did cross-sectional momentum on the top 10 equities by market cap over last 5 years out-performed the market ?”.

Research Structure

- Research is highly dependent on quality and type of data. “Garbage In Garbage Out ”
- The research loop starts with selecting an universe of securities.
- The alpha discovered from various ideas are then combined using a model.
- Then portfolios are constructed using some risk model and constraints.
- Finally, the portfolio is backtested and iterative research is conducted following some protocols to avoid overfitting.



Data

- Data is the fuel to quantitative models and the high quality data is crucial for success and doing quantitative research.
- Data can have missing values, error in recording, fault in cleaning or issues with aggregation.
- There might also be other issues with the dataset altogether i.e. survivorship bias in the universe selection or having low frequency data which might restrict the backtest to have a narrow view.
- Handling outliers in the data is nuisance, removing dataset will introduce bias in research and keeping it may lead to skewness in analysis. Some methods that scientists prefer are - Winsorization, Mean/Median Imputation etc.
- Check these videos for more details:
  https://www.youtube.com/watch?v=bN5aZLp0gvw&ab_channel=Quantopian
  https://www.youtube.com/watch?v=ZME_61rPNVY&ab_channel=Hudson%26Thames

Financial ML : Preparing the data

- Financial data comes in many shapes and forms and it can be classified into four types: Fundamental data, Market data, financial analytics and alternative data.

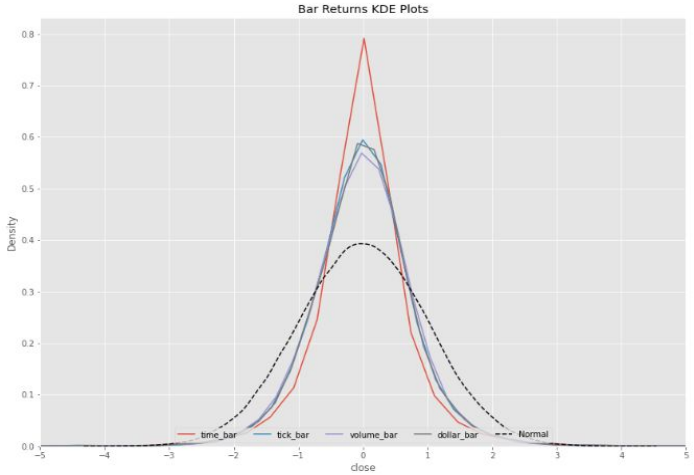
- For applying machine learning models on financial data we need to structure it to extract valuable information. Finance practitioner often tend to use a table representation which they refer as Bars.

- Time bars are the most common type of bars that is used widely but something called information driven bars seems give better result while using with machine learning models and has better statistical properties.

TABLE 2.1 The Four Essential Types of Financial Data

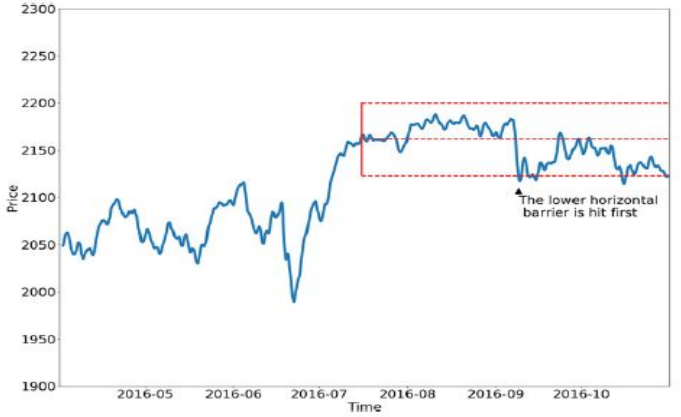
Fundamental Data	Market Data	Analytics	Alternative Data
• Assets • Liabilities • Sales • Costs/earnings • Macro variables • . . .	• Price/yield/implied volatility • Volume • Dividend/coupons • Open interest • Quotes/cancellations • Aggressor side • . . .	• Analyst recommendations • Credit ratings • Earnings expectations • News sentiment • . . .	• Satellite/CCTV images • Google searches • Twitter/chats • Metadata • . . .

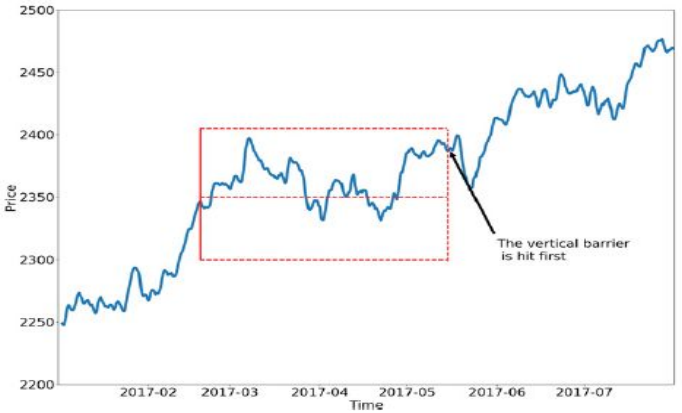
Source: Advances in Financial Machine Learning



Financial ML : Labelling the data

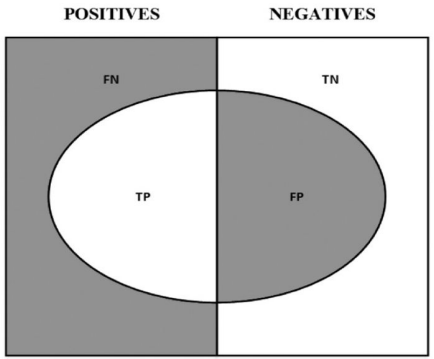
- Now that we have prepared our dataset, in order to run a supervised learning algorithm we need some sort of labels or values that can be predicted on unseen features samples.
- These labels can have a particular view of the market i.e. either the price or the direction of the market or outcome of a bet.
- It is difficult to predict the price due to inherent nature of the market and assumptions of machine learning algorithm so literatures suggest that it is better focus of predicting the direction and magnitude of the move separately rather trying to predict the price.
- A technique suggested by Lopez de Prado [2018] is called *Triple Barrier Method*.

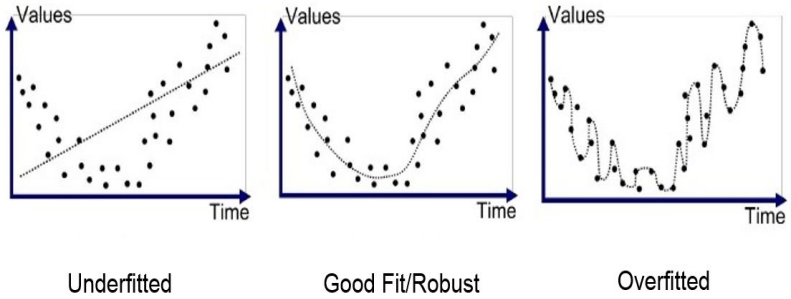




Financial ML : Modelling

- It is crucial to perform a proper modelling techniques to train the dataset on. First we need to a machine learning algorithm to performing regression or classification task depending on the labels.
- The goal modelling is to minimize the three source of error that ML models generally suffer from i.e. Bias, Variance and Noise.
- The method that suggested in Advance in Financial Machine Learning by Lopez de Prado [2018] is to have a model for predicting the sides of a bet which have a high-recall (relatively low precision) and an meta-model on top to decide the size of the bet by improving the precision of the earlier model.
- Techniques like Bagging is preferred over Boosting because boosting tends to address the problem of underfitting while bagging address overfitting. In finance generally overfitting is an issue.





Financial ML : Asset Allocation

- A small percentage of the data content is signal, which is mostly suppressed by arbitrage force. ML techniques can help denoising the covariance matrix without giving up the little signal it contains.

- Denoised covariance matrix can be useful in deriving the linear relationship but for non-linear relationships between assets clustering algorithms are used.

- Techniques like mean-variance optimization are not stable due its dependence on covariance matrix which can be noisy, and ML methods can help correct that.

- ML based asset allocation algorithms like HRP (Hierarchical Risk Parity) show the evidences outperform the benchmark CLA(Critical Line Algorithm).

Some common pitfalls

- **Forecasting the market**: Market is a dynamic system that changes its states depending on the supply and demand. Most of the statistical methods and machine learning techniques has the base assumption that the data is stationary, samples are IID and from a normal distribution. Market data does qualify even a single requirement, although there are techniques to solve each of these problem the dynamic nature of the market doesn't seem to be fit for forecasting.

- **Backtest for validation:** Backtest shouldn't considered as a research tool rather a sanity check of the theory. It is must to develop a theory (or hypothesis) first than work on proving that or contradicting using feature importance analysis which is robust to overfitting.

- **All weather strategies:** Try to avoid researching for that holy grail strategy which you think should work on all market regimes because likelihood of the existence of such strategy is rather slim and discovery is most likely be a false positive. Rather focus on strategies that can perform optimally under a specific regime like mean-reversion or market sell-off.

Some useful resources

- Connected Papers: https://www.connectedpapers.com/
- TOPICS IN MATHEMATICS WITH APPLICATIONS IN FINANCE : https://ocw.mit.edu/courses/18-s096-topics-in-mathematics-with-applications-in-finance-fall-2013/
- Quantconnect: https://www.quantconnect.com/
- Quant Research: https://quantresearch.org/
- ML-quant: https://www.ml-quant.com/
- Quantopian github: https://github.com/quantopian
- Man-Oxford: https://www.oxford-man.ox.ac.uk/publications/
- JP Morgan AI Research: https://www.jpmorgan.com/insights/technology/artificial-intelligence/ai-research-publications

Reference

- https://hudsonthames.org/best-research-practices-for-your-quantitative-finance-research-group/
- https://alpaca.markets/learn/alternative-bars-02/
- https://alpaca.markets/learn/alternative-bars-03/
- Lopez de Prado. 2018. Advances in Financial Machine Learning
- Lopez de Prado. 2020. Machine Learning for Asset Managers.
- Matthew F. Dixon, Igor Halperin and Paul Bilokon. 2020. Machine Learning in Finance: From Theory to Practice
- Cerniglia, J. A., Fabozzi, F. J., & Kolm, P. N. (2016). Best Practices in Research for Quantitative Equity Strategies. The Journal of Portfolio Management, 42(5), 135–143. doi:10.3905/jpm.2016.42.5.135

Reference

- López de Prado, Marcos and López de Prado, Marcos, The 7 Reasons Most Machine Learning Funds Fail (Presentation Slides) (September 2, 2017). Available at SSRN: https://ssrn.com/abstract=3031282 or http://dx.doi.org/10.2139/ssrn.3031282

- López de Prado, Marcos and López de Prado, Marcos and Lipton, Alex, Three Quant Lessons from COVID-19 (Presentation Slides) (March 27, 2020). Available at SSRN: https://ssrn.com/abstract=3562025 or http://dx.doi.org/10.2139/ssrn.3562025

- López de Prado, Marcos and López de Prado, Marcos, Backtesting (May 14, 2015). Available at SSRN: https://ssrn.com/abstract=2606462 or http://dx.doi.org/10.2139/ssrn.2606462

QNA

Task for Day 5

Implementing a research paper

- Find a research paper that explains a trading strategy of your choice and implement the strategy.
- Create a Jupyter Notebook (or Google Colab) Notebook and make sure to explain the strategy using markdown.
- Perform a Backtest on the strategy using suitable dataset as per the paper.
- Mention the papers and other resources in the reference.
- Last date for submission : 25-05-2022
- Submission Form : https://forms.gle/EPF6UoasjCtuxxNb6