Project Proposals for MS&E 448

Spring Quarter 2021

Dr. Lisa Borland

1 Build a High Frequency Strategy

Description

The goal of this project is to design and implement a trading strategy based on limit order book data, i.e. high frequency stocks data, that exploits the specific structure of such datasets.

Initial questions

- What is a good time scale to consider: tick-by-tick or a larger time-scale (for example, can an integrated order book profile predict anything over longer horizons)?
- Use machine learning techniques or impose a fundamental relationship
- Discuss and analyze execution tactics (e.g. if you are aggressing, can you really get that price? How much slippage do you expect? Adverse selection?)
- Given the nature of your alpha signal, and how you expect the price to move immediately after entering an order, what is the best execution strategy to optimize the probability of fill in such a way that you minimize market impact and avoid adverse selection?
- If you can make money getting the mid price, but lose if you have to pay the spread, can you get around this by executing cleverly?
- Given multiple potential counterparty venues with different liquidity profiles, response times, rejection rates, spreads, how do you optimally route your orders?
- Alternatively to playing the role of a profit-maximizing trader, see if you can take the perspective of a market maker and implement an optimal market making strategy.

Data and coding environment

Students will have access to MayStreet limit order book data and high frequency simulator.

References

- Michael Kearns and Yuriy Nevmyvaka. Machine learning for market microstructure and high frequency trading. *High Frequency Trading: New Realities for Traders, Markets, and Regulators, Risk Books*, 2013
- Sasha Stoikov. The micro-price: A high frequency estimator of future prices. *SSRN-id2970694*, 2018

2 Build a statistical arbitrage strategy

Description

The goal of this project is to design and implement a statistical arbitrage trading strategy. This implies identifying pairs or groups of stocks that are expected to behave similarly and use this to exploit their prices differences.

Initial questions

- Do you want to consider pairs of stocks or create clusters of similar stocks?
- How can you find similar stocks?
- What technique can you use to predict residuals returns: O-U process or other statistical techniques?
- How can you use your signal to create a portfolio that optimize for risk, transaction costs, liquidity etc?

Data and coding environment

Refer to the Guide to Financial Data in the Data section of the website. Once you will have designed a strategy, you can simulate its performance in a finance simulator such as cvxportfolio, or write your own simple backtester.

References

- Marco Avellaneda and Jeong-Hyun Lee. Statistical arbitrage in the us equities market. *Quantitative Finance*, 10(7):761–782, 2010
- Nicolas Huck. Pairs trading and outranking: The multi-step-ahead forecasting case. *European Journal of Operational Research*, 207(3):1702–1716, 2010
- George J Miao. High frequency and dynamic pairs trading based on statistical arbitrage using a two-stage correlation and cointegration approach. *International Journal of Economics and Finance*, 6(3):96, 2014
- Alexander Lipton and Marcos Lopez de Prado. A closed-form solution for optimal mean-reverting trading strategies. *arXiv preprint arXiv:2003.10502*, 2020

3 Build an index arbitrage strategy

Description

The goal of this project is to design and implement an index arbitrage trading strategy which is a variant of statistical arbitrage. Index arbitrage can be done in various different ways. One way is to identifying spread between spot prices of stocks constituting of an index and the related index futures contract. Another way, is to measure within an index, the degree of reactivity of each stock to the market impact. Spread between leaders and laggers can then be exploited.

Data and coding environment

Refer to the Guide to Financial Data in the Data section of the website. Once you will have designed a strategy, you can simulate its performance in a finance simulator such as cvxportfolio, or write your own simple backtester. Students can also have access to MayStreet limit order book data and high frequency simulator.

References

- Y Peter Chung. A transactions data test of stock index futures market efficiency and index arbitrage profitability. *The Journal of Finance*, 46(5):1791–1809, 1991

- James Richard Cummings and Alex Frino. Index arbitrage and the pricing relationship between australian stock index futures and their underlying shares. *Accounting & Finance*, 51(3):661–683, 2011

- Nicolas Huth and Frédéric Abergel. High frequency lead/lag relationships—empirical facts. *Journal of Empirical Finance*, 26:41–58, 2014

- Chester Curme, Michele Tumminello, Rosario N Mantegna, H Eugene Stanley, and Dror Y Kenett. Emergence of statistically validated financial intraday lead-lag relationships. *Quantitative Finance*, 15(8):1375–1386, 2015

4 Crypto-currencies Price Prediction, Perhaps Using News and Social Networks Data

Description

The goal of this project is to design and implement a model that achieves good price predictions of a chosen crypto-currency. Predictions can be made using past prices information in addition to news or social networks data such as tweet sentiments and volume or google trend data.

Data and coding environment

Refer to the Guide to Financial Data in the Data section of the website. Historical tweets related to a specific topic such as a crypto-currency can be obtained using the following python package. Once you will have designed a strategy, you can simulate its performance by writing your own simple backtester.

References

- Fan Fang, Carmine Ventre, Michail Basios, Hoiliong Kong, Leslie Kanthan, Lingbo Li, David Martinez-Regoband, and Fan Wu. Cryptocurrency trading: A comprehensive survey. *arXiv preprint arXiv:2003.11352*, 2020

- Sean McNally, Jason Roche, and Simon Caton. Predicting the price of bitcoin using machine learning. In *2018 26th Euromicro International Conference on Parallel, Distributed and Network-based Processing (PDP)*, pages 339–343. IEEE, 2018

- Young Bin Kim, Jun Gi Kim, Wook Kim, Jae Ho Im, Tae Hyeong Kim, Shin Jin Kang, and Chang Hun Kim. Predicting fluctuations in cryptocurrency transactions based on user comments and replies. *PloS one*, 11(8):e0161197, 2016

- Vytautas Karalevicius, Niels Degrande, and Jochen De Weerdt. Using sentiment analysis to predict interday bitcoin price movements. *The Journal of Risk Finance*, 19(1):56–75, 2018

- Muhammad Amjad and Devavrat Shah. Trading bitcoin and online time series prediction. In *NIPS 2016 Time Series Workshop*, pages 1–15, 2017

5 Short term Spot FX Price Prediction using Quotes from different Liquidity Providers

Description

The FX market is highly fragmented. There is no real notion of an order book as in equities. Each participant receives quotes from the counter-parties that they have a relationship with. At any given time, quotes from these different liquidity providers with show quotes at bid and ask that are different from each other. The LPs can also skew the mid price quote (based on their own view, or for risk management purposes.) The goal of this project is to build a hypothetical "quote book" from the data, analyze the behaviors of the the different LPs, and see if there is any information that can be detected exploiting the different views of the LPs that can predict short term price movement. Modeling techniques can include any statistical or machine learning technique. We include as references a couple of methods using reinforcement learning.

Data and coding environment

FX data from the integral platform are available on the class website in the Data section. Once you will have designed a strategy, you can simulate its performance by writing your own simple backtester.

References

- Martin D. Gould, Mason A. Porter, and Sam D. Howison. Quasi-centralized limit order books. *arXiv:1502.00680v2*, 2016

- Michael AH Dempster and Vasco Leemans. An automated fx trading system using adaptive reinforcement learning. *Expert Systems with Applications*, 30(3):543–552, 2006

- Yue Deng, Feng Bao, Youyong Kong, Zhiquan Ren, and Qionghai Dai. Deep direct reinforcement learning for financial signal representation and trading. *IEEE transactions on neural networks and learning systems*, 28(3):653–664, 2017

- Martin D. Gould, Mason A. Porter, and Sam D. Howison. Quasi-centralized limit order books. *arXiv:1502.00680v2*, 2016

6 Improving on Classical Anomalies by Combining Them and perhaps adding external sentiment data

Description

Well-know market anomalies (low volatility, size, quality) exist. The goal of the project is to find alpha based on these, and to optimally combine the different signals to construct a superior portfolio. You should see if applying smarter prediction techniques (eg smarter volatility prediction for the low volatility if you can use smarter statistical/machine learning or clustering techniques either to improve on the simple anomalies or to better predict when they will work. Additionally you can include sentiment data.

Data and coding environment

Refer to the Guide to Financial Data in the Data section of the website.Once you will have designed a strategy, you can simulate its performance by writing your own simple backtester.

References

- Jean-Philippe Bouchaud Pierre Blanc, Rémy Chicheportiche. The fine structure of volatility feedback ii: overnight and intra-day effects. *arXiv:1309.5806*, 2014

- A. Beveratos G. Simon L. Laloux M. Potters J.-P. Bouchaud S. Ciliberti, Y. Lempérière. Deconstructing the low-vol anomaly. *arXiv:1510.01679*, 2015

- Guillaume Simon Yves Lempérière Jean-Philippe Bouchaud Stefano Ciliberti, Emmanuel Sérié. The "size premium" in equity markets: Where is the risk? *arXiv:1708.006449*, 2017

- Augustin Landier Guillaume Simon Jean-Philippe Bouchaud, Stefano Ciliberti and David Thesmar. The excess returns of "quality" stocks: A behavioral anomaly. *arXiv:1601.04478v1*, 2016

7 Trend-Following Strategies in Futures Markets

Description

Trend following of futures is a classical strategy that has worked for decades. Rather than prediction, it involves quickly detecting when a trend has started, and managing when to exit a trade. The goal of this project is to replicate and improve on the basic ideas, using more advanced statistical / machine learning techniques in conjunction with exploiting correlation structures within and across asset classes.

Data and coding environment

Refer to the Guide to Financial Data in the Data section of the website.

References

- Lasse H. Pedersen Brian Hurst, Yao Hua Ooi. A century of evidence on trend-following investing. *https://www.aqr.com/Insights/Research/Journal-Article/A-Century-of-Evidence-on-Trend-Following-Investing*, 2014

- P. Seager M. Potters J. P. Bouchaud Y. Lempérière, C. Deremble. Two centuries of trend following. *https://arxiv.org/pdf/1404.3274.pdf*, 2014

8 Options volatility trading strategy

Description

The challenge will be to come up with volatility predictions, absolute or relative value, utilizing at the money options or nearby strikes.

- Create signals.
- Put together a portfolio.
- Discuss hedging and risk management.
- Discuss Execution issues and ways in which the backtest could deviate in real life.

Data and coding environment

Refer to the Guide to Financial Data in the Data section of the website.

References

- Wolfgang Karl Härdle and Elena Silyakova. Implied basket correlation dynamics. *https://arxiv.org/pdf/1404.3274.pdf*, 2020
- Ruslan Goyenko and Chengyu Zhang. The joint cross section of option and stock returns predictability with big data and machine learning. *Available at SSRN*, 2020

9 Calibrating an agent-based model on real stock market behavior

Description

- Negative feedback (range bound market)
- Positive Feedback (a trending market, exponential growth, bubbles).
- Toy model: agent based, different market participants
- See if you can calibrate the toy model to real data, maybe find regimes of positive and negative feedback.
- Can you build a trading strategy based on this.

Data and coding environment

Data: Daily or intraday, US stocks. Refer to the Guide to Financial Data in the Data section of the website.

References

- TT Chen, B Zheng, Y Li, and XF Jiang. New approaches in agent-based modeling of complex financial systems. *arXiv preprint arXiv:1703.06840*, 2017
- Thomas Lux and Michele Marchesi. Volatility clustering in financial markets: a microsimulation of interacting agents. *International journal of theoretical and applied finance*, 3(04):675–702, 2000

10 Project X

Description

Students may propose their own project idea.

11 (Newly added) Trading Algorithm for J-Curve Shifts in VC-backed Tech Companies

Description

This project is in partnership with Catarina Capital, and students who choose this project will collaborate with the company throughout the quarter.

The project seeks to identify and exploit the J-curved evolution of operating cash flows in companies backed by VCs. The main criterion to use is the historical curve of Income from Continuing Operations (ICO), as well as its relationship to indexes focused on cash burn/gen. from different tech assets. According to the project proposal, looking into ICO provides a robust way to forecast stock up & down price behaviors for most assets, especially for VC-backed ones. The goal of the project is to refine this idea and create a quantitative algorithm that automatizes the identification of J-Curve shifts and the conception of a dynamic tech-stocks portfolio to be invested.

For more info on this project, please refer to the document uploaded on Canvas that outlines the proposal by Catarina Capital.

References

[1] Muhammad Amjad and Devavrat Shah. Trading bitcoin and online time series prediction. In *NIPS 2016 Time Series Workshop*, pages 1–15, 2017.

[2] Marco Avellaneda and Jeong-Hyun Lee. Statistical arbitrage in the us equities market. *Quantitative Finance*, 10(7):761–782, 2010.

[3] Lasse H. Pedersen Brian Hurst, Yao Hua Ooi. A century of evidence on trend-following investing. *https://www.aqr.com/Insights/Research/Journal-Article/A-Century-of-Evidence-on-Trend-Following-Investing*, 2014.

[4] TT Chen, B Zheng, Y Li, and XF Jiang. New approaches in agent-based modeling of complex financial systems. *arXiv preprint arXiv:1703.06840*, 2017.

[5] Y Peter Chung. A transactions data test of stock index futures market efficiency and index arbitrage profitability. *The Journal of Finance*, 46(5):1791–1809, 1991.

[6] James Richard Cummings and Alex Frino. Index arbitrage and the pricing relationship between australian stock index futures and their underlying shares. *Accounting & Finance*, 51(3):661–683, 2011.

[7] Chester Curme, Michele Tumminello, Rosario N Mantegna, H Eugene Stanley, and Dror Y Kenett. Emergence of statistically validated financial intraday lead-lag relationships. *Quantitative Finance*, 15(8):1375–1386, 2015.

[8] Michael AH Dempster and Vasco Leemans. An automated fx trading system using adaptive reinforcement learning. *Expert Systems with Applications*, 30(3):543–552, 2006.

[9] Yue Deng, Feng Bao, Youyong Kong, Zhiquan Ren, and Qionghai Dai. Deep direct reinforcement learning for financial signal representation and trading. *IEEE transactions on neural networks and learning systems*, 28(3):653–664, 2017.

[10] Fan Fang, Carmine Ventre, Michail Basios, Hoiliong Kong, Leslie Kanthan, Lingbo Li, David Martinez-Regoband, and Fan Wu. Cryptocurrency trading: A comprehensive survey. *arXiv preprint arXiv:2003.11352*, 2020.

[11] Martin D. Gould, Mason A. Porter, and Sam D. Howison. Quasi-centralized limit order books. *arXiv:1502.00680v2*, 2016.

[12] Ruslan Goyenko and Chengyu Zhang. The joint cross section of option and stock returns predictability with big data and machine learning. *Available at SSRN*, 2020.

[13] Wolfgang Karl Härdle and Elena Silyakova. Implied basket correlation dynamics. *https://arxiv.org/pdf/1404.3274.pdf*, 2020.

[14] Nicolas Huck. Pairs trading and outranking: The multi-step-ahead forecasting case. *European Journal of Operational Research*, 207(3):1702–1716, 2010.

[15] Nicolas Huth and Frédéric Abergel. High frequency lead/lag relationships—empirical facts. *Journal of Empirical Finance*, 26:41–58, 2014.

[16] Augustin Landier Guillaume Simon Jean-Philippe Bouchaud, Stefano Ciliberti and David Thesmar. The excess returns of "quality" stocks: A behavioral anomaly. *arXiv:1601.04478v1*, 2016.

[17] Vytautas Karalevicius, Niels Degrande, and Jochen De Weerdt. Using sentiment analysis to predict interday bitcoin price movements. *The Journal of Risk Finance*, 19(1):56–75, 2018.

[18] Michael Kearns and Yuriy Nevmyvaka. Machine learning for market microstructure and high frequency trading. *High Frequency Trading: New Realities for Traders, Markets, and Regulators, Risk Books*, 2013.

[19] Young Bin Kim, Jun Gi Kim, Wook Kim, Jae Ho Im, Tae Hyeong Kim, Shin Jin Kang, and Chang Hun Kim. Predicting fluctuations in cryptocurrency transactions based on user comments and replies. *PloS one*, 11(8):e0161197, 2016.

[20] Alexander Lipton and Marcos Lopez de Prado. A closed-form solution for optimal mean-reverting trading strategies. *arXiv preprint arXiv:2003.10502*, 2020.

[21] Thomas Lux and Michele Marchesi. Volatility clustering in financial markets: a microsimulation of interacting agents. *International journal of theoretical and applied finance*, 3(04):675–702, 2000.

[22] Sean McNally, Jason Roche, and Simon Caton. Predicting the price of bitcoin using machine learning. In *2018 26th Euromicro International Conference on Parallel, Distributed and Network-based Processing (PDP)*, pages 339–343. IEEE, 2018.

[23] George J Miao. High frequency and dynamic pairs trading based on statistical arbitrage using a two-stage correlation and cointegration approach. *International Journal of Economics and Finance*, 6(3):96, 2014.

[24] Jean-Philippe Bouchaud Pierre Blanc, Rémy Chicheportiche. The fine structure of volatility feedback ii: overnight and intra-day effects. *arXiv:1309.5806*, 2014.

[25] A. Beveratos G. Simon L. Laloux M. Potters J.-P. Bouchaud S. Ciliberti, Y. Lempérière. Deconstructing the low-vol anomaly. *arXiv:1510.01679*, 2015.

[26] Guillaume Simon Yves Lempérière Jean-Philippe Bouchaud Stefano Ciliberti, Emmanuel Sérié. The "size premium" in equity markets: Where is the risk? *arXiv:1708.006449*, 2017.

[27] Sasha Stoikov. The micro-price: A high frequency estimator of future prices. *SSRN-id2970694*, 2018.

[28] P. Seager M. Potters J. P. Bouchaud Y. Lempérière, C. Deremble. Two centuries of trend following. *https://arxiv.org/pdf/1404.3274.pdf*, 2014.