

# Project Proposals for MS&E 448

Spring Quarter 2018

Dr. Lisa Borland , TA: Enguerrand Horel

# 1 Build a High Frequency Price Movement Strategy

Students will have access to Tradeworx and Thesys data and simulator. Access order book data. Use this data to predict short term price movements:

- Can information in the order book be used to predict price movements?
- On a tick-by-tick time scale?
- On 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 (eg 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 counter-party venues with different liquidity profiles, response times, rejection rates, spreads, how do you optimally route your orders?
- Literature:
  - Machine Learning for Market Microstructure and High Frequency Trading* Michael Kearns and Yuriy Nevmyvaka <http://www.cis.upenn.edu/~mkearns/papers/Kearns>
  - A High Frequency Trade Execution Model for Supervised Learning*, Matthew Dixon arXiv:1710.03870v3 [q-fin:TR] Dec 2017
  - The Micro-Price: A High Frequency Estimator of Future Prices*, Stoikov, Sasha, (November 26, 2017). Available at SSRN: <https://ssrn.com/abstract=2970694> or <http://dx.doi.org/10.2139/ssrn.2970694>

Hands-on experience with what cutting-edge traders face in real life. Unique opportunity that this API is offered to students.

## 2 Build a classical statistical arbitrage strategy

- Data: Clean it, make sure adjusted for corporate actions etc.
- Build groups: sectors, clusters.
- Define residual returns.
- Predict residuals: O-U process or other statistical techniques, etc.
- Create a portfolio: Optimize for risk, transaction costs, liquidity etc. Convex linear optimization techniques. Can be intra day or daily.
- Simulate! Quantopian
- Data: Bloomberg, Quantopian, Thesys.
- Language: Python
- Literature:
  - Statistical Arbitrage in the U.S. Equities Market*, Avellaneda and Lee 2008.
  - Pairs trading and outranking: The multi-step-ahead forecasting case*, European Journal of Operational Research 207.3 (2010): 1702-1716, Nicolas Huck
  - High Frequency and Dynamic Pairs Trading Based on Statistical Arbitrage Using a Two-Stage Correlation and Cointegration Approach*, George J. Miao, International Journal of Economics and Finance; Vol. 6, No. 3; 2014

### 3 Build a novel cluster-based strategy

- Data: Clean it, make sure adjusted for corporate actions etc.
- Build your own factors (fundamental, statistical, clustering algos etc)
- Define residual returns.
- Predict residuals: O-U process or other statistical techniques, etc.
- Create a portfolio: Optimize for risk, transaction costs, liquidity etc. Convex linear optimization techniques. Can be intra day or daily.
- Simulate! Quantopian
- Data: Bloomberg, Quantopian, Thesys.
- Language: Python
- Literature:
  - Statistical Arbitrage in the U.S. Equities Market*, Avellaneda and Lee 2008.
  - Pairs trading and outranking: The multi-step-ahead forecasting case*, European Journal of Operational Research 207.3 (2010): 1702-1716, Nicolas Huck
  - High Frequency and Dynamic Pairs Trading Based on Statistical Arbitrage Using a Two-Stage Correlation and Cointegration Approach*, George J. Miao, International Journal of Economics and Finance; Vol. 6, No. 3; 2014

## 4 Fundamental signals (and Machine Learning) for stock price prediction

- Data: Quantopian, Estimize, Quandl, Bloomberg (Futures)
- Featureize and classify the data to find variables that are predictable of 1 month - 3 month returns
- Use novel machine learning techniques!
- Build a suite of predictors. You may use technical analysis-type signals
- Create forecasts over multiple horizons
- Build a portfolio that utilizes multi period optimization
- Take into account risk (correlations), transaction costs when doing portfolio optimization
- Convex optimization techniques
- Language: Python
- Simulate: Quantopian?
- Literature: Foundations of Technical Analysis: Computational Algorithms, Statistical Inference, and Empirical Implementation Andrew Lo, Harry Mamaysky, Jiang Wang <https://www.cis.upenn.edu/~mkearns/teaching/cis700/lo.p>
- Literature Quantopian blog posts, Ernie Chan's blog posts
- Literature: Prof. S Boyd's website
- Literature: Trend Following with Managed Futures: The Search for Crisis Alpha (Wiley Trading) 1st Edition by Alex Greyserman (Author), Kathryn Kaminski (Author)
- Literature: Chattopadhyay, Lyle, Wang Accounting Data, Market Values, and the Cross Section of Expected Returns Worldwide <http://papers.ssrn.com/>

## 5 Calibrating an agent-based model on real stock market behavior

- 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: Daily or intraday, US stocks.
- Literature:  
New approaches in agent-based modeling of complex financial systems  
March 4, 2017 T. T. Chen<sup>1,2</sup>, B. Zheng<sup>1,2</sup>, Y. Li<sup>1,2</sup>, X. F. Jiang<sup>1,2</sup>

## 6 Build a market making strategy

- Data: Tradeworx and Thesys, cutting edge access to high frequency data from 13 exchanges, microsecond resolution, and simulator.
- Place an order to sell above the market price
- Place an order to buy below the market price
- Make money on the bid-ask spread
- Problems:
  - You are competing with others.
  - The market is moving
  - You have to manage your inventory risk
  - Will you lay-off some risk in correlated markets?
- What other problems are there?

Literature:

*A stochastic model for order book dynamics* Rama Cont, Sasha Stoikov, Rishi Talleja <http://www.columbia.edu/~ww2040/orderbook.pdf>

*High-frequency trading in a limit order book* Marco Avellaneda and Sasha Stoikov, *Quantitative Finance*, Vol. 8, No. 3, April 2008, 217-224

*Machine Learning for Market Microstructure and High Frequency Trading*

Michael Kearns and Yuriy Nevmyvaka <http://www.cis.upenn.edu/~mkearns/papers/KearnsNevmyvaka.pdf>

*A High Frequency Trade Execution Model for Supervised Learning*, Matthew Dixon arXiv:1710.03870v3 [q-fin:TR] Dec 2017

## 7 Uncovering causal relationships among stock moves

- Data: Quantopian, Bloomberg intra day top of book, or Thesys
- Method: Consider a time series of quotes on a set of US large cap stocks. The time series contains the best bid and best ask price across a variety of markets and their timestamp. We want to:
- Build and fit a model of causal relationships between quotes events.
- Interpret the results to cluster the stocks in 'communities', identify their leaders and laggards.
- Design a trading algorithm that uses that information

Literature:

Ranking Causal Influence of Financial Markets via Directed Information Graphs Theo Diamandis, Yonathan Murin, and Andrea Goldsmith Department of Electrical Engineering, Stanford University, Stanford CA, USA ARXIV:SUBMIT/2138416 [eess.SP] Jan 2018

## 8 Add on project: Sharpe ratio, Sortino, what summary statistic to use to best predict out of sample performance

- Data: Quantopian, Thesys
- Given a strategy, what is the best measure to predict out of sample returns?
- How to avoid overfitting? How much Data do you need?
- Explore these measures while designing your own optimal strategy
- Or develop measures or machine learning techniques to optimally select which strategies will perform (given a set of strategy returns which you come up with based on past data).
- Literature:
  - The probability of backtest overfitting* Bailey et al, <http://papers.ssrn.com/>
  - Generalization in Adaptive Data Analysis and Holdout Reuse*, Dwork et al, <http://arxiv.org/pdf/1506.02629.pdf>
  - Data-snooping, technical trading rule performance, and the bootstrap*, Sullivan, Ryan, Allan Timmermann, and Halbert White. *The journal of Finance* 54.5 (1999): 1647-1691

## 9 Options Volatility Trading

The challenge will be to come up with volatility predictions, absolute or relative value, utilizing at the money options or nearby strikes. Stand-alone project that does not utilize Thesys or Quantopian.

- Data: Bloomberg, Stanford Data Sets - You must collect and clean options data yourselves.
- Predict volatility based on eg the dynamics of the underlying
- 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.
- Optional: Use an extension of Black-Scholes theory to incorporate fat tails and skew, create a flat volatility surface, explore signals in that representation.

## 10 Project X

Students may submit their own proposals!

- Must be well formulated
- You may clone ideas from Quantopian platform but you must reference these
- Your own work has to be substantially different if cloned

## 11 Reading Materials etc.

- Papers will be posted on the class website
- Links to papers may be supplied
- Papers can be read regardless of your project (cross pollination)
- Main work will be done in the iPython notebook environment