Market Microstructure and Algorithmic Trading Final Paper

Start Assignment

- Due Wednesday by 11:59pm
- Points 50
- Submitting a file upload
- File Types ipynb, pdf, zip, and html
- Available Nov 17 at 6:30pm - Dec 17 at 11:59pm

The final paper requires you to use the NYSE TAQ Level 1, CME Level 2, Crypto Level I, Crypto Level II, SpiderRock and/or LOBSTER data to perform original research. Build your own Jupyter notebook using the techniques learned throughout the course as well as machine learning techniques you've learned throughout the program.

Requirements

- This project must be performed and submitted in **groups of 4-5 students** (a combination of 2 homework teams with the possibility of an extra student who did not partner for the homework assignments).
- You must submit **three files**: the **Jupyter notebook (.ipynb)**, an **HTML (.html)** file which is generated by exporting your Jupyter notebook and a **PDF (.pdf) report** (not to exceed 6 pages) with your write-up. These may be combined into a **single ZIP (.zip) file**.
- Workbooks **must execute without errors from beginning to end within 15 minutes to receive full credit**. To help with this, you may submit an extra model files that contains a trained model object. Note that **ONNX** (https://onnx.ai/onnx/) is a more secure format for **model files** (https://scikit-learn.org/stable/model_persistence.html) than pickle.
- Please specify any extra anaconda packages that are required to run your analysis.
- Your submission should include a **clear statement of the topic** you are investigating, a **thorough description of the steps of your analysis**, and an **explanation** and **interpretation** of your results.
- Your submission must include a **minimum of two relevant tables and formulas** to help elucidate your analysis.
- You must also generate **fully-labeled charts** to help make your presentation clear.
- Your submission must provide at **least one reference**, and possibly more, relating to your analysis (e.g., where you got the idea for your analysis or related work on the same topic).
- The write-up should be a **polished report, rather than a collage** compiled from each contributor.
- If your project involves training a predictive model on a training sample, **be sure your hold-out testing sample is after your training sample**. For example, you should use **TimeSeriesSplit** (https://scikit-learn.org/dev/modules/generated/sklearn.model_selection.TimeSeriesSplit.html) to perform K-Fold cross validation.

Suggestions

You are not limited to single-security analysis and can therefore look for lead-lag effects across securities, if relevant, for the question you are investigating. Similarly, you are not limited to single-exchange analysis and can therefore look for latency arbitrage opportunities between venues.

Demonstrate your understanding of the concepts introduced in this course by commenting qualitatively (if not quantitatively) on how market impact would effect your analysis.

You may refer to **recent papers** (https://www.connectedpapers.com/main/6def0370e9733d233991d862483249e556501972/Microstructure-in-the-Machine-Age/graph) and the below list for ideas, or create your own. Extra points will be given for innovative work.

Ideas

- Mean-variance analysis of an optimal order-execution algorithm (**Almgren and Chriss** (https://papers.ssrn.com/sol3/papers.cfm?abstract_id=53501), or with conditional information)
- Analysis of intraday data using machine learning techniques. i.e.: **"The Conduits of Price Discovery: A Machine Learning Approach". Amy Kwan, Richard Philip, Andriy, Shkilko** (https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3710491)
- Cross-market latency: how do orders on one market affect other markets for the same security

- How quickly does the limit order book (LOB) replenish itself
- Predict whether the next order is a buy or sell
- Predict future volume or order imbalance over the rest of the trading day
- Intraday cross-correlation across a cross-section of stocks or futures contracts. i.e. **"Common Factors in Prices, Order Flows, and Liquidity". Joel Hasbrouck, Duane Seppi (https://www.sciencedirect.com/science/article/abs/pii/S0304405X0000091X)**
  - Signed order flow for marketable orders
  - Signed order flow for BBO limit orders
  - BBO spreads
  - LOB depth
  - LOB depth imbalance
  - Time-series of Kyle's λ
- How do different order types affect prices?
  - $\Delta m_t = \mu + \lambda x_t + \epsilon_t$ (where x_t are signed market orders MOs or limit orders LO at the BBO or even behind the BBO)
- Does liquidity look different for predictable and unpredictable orders? Measures of predictability include:
  - Orders on even seconds, minutes, hours
  - Orders where $|x_t - \mathbb{E}_{t-1}[x_t]|$ is small vs large given $\mathbb{E}_{t-1}[x_t]$ is estimated by a time-series regression
- How does the price impact of a particular order type change over time? And what features of the environment induce those changes?
  - Model a time-varying price-impact coefficient $\lambda_t = \lambda(z)$
    i.e. if $\lambda_t = \lambda_1 z_{1,t-1} + \lambda_2 z_{2,t-1}$
    the Kyle equation $\Delta m_t = \mu + \lambda_t x_t + e_t$ becomes $\Delta m_t = \mu + \lambda_1 (z_{1,t-1} x_t) + \lambda_2 (z_{2,t-1} x_t) + e_t$
  - Possible explanatory variables z:
    - Intraday time buckets
    - Prior order-flow imbalance interacted with sign of current order
    - Time duration elapsed since the LO or last MO
    - Volume of prior LO "churn"

Market Microstructure and Algorithmic Trading Final Paper

Criteria	Ratings					Pts
Kdb+ Knowledge How well does the research demonstrate mastery of Kdb+	**4 pts** **Advanced** Implemented new Kdb+ code to perform innovative analytics	**3 pts** **Above Average** Implemented new Kdb+ code to perform analytics	**2 pts** **Average** Used Kdb+ code from class to analyze data	**1 pts** **Below Average** Used kdb+ to download data	**0 pts** **Insufficient** No use of Kdb+	4 pts
Tables and Graphs How well does the research use tables and graphs to enhance the interpretability of the research	**4 pts** **Advanced** Presents at least 2 tables and 2 graph (with at least 1 custom designed) supporting the results of the analysis	**3 pts** **Above Average** Presents at least 1 tables and 2 graph (with at least 1 custom designed) supporting the results of the analysis	**2 pts** **Average** Presents at least 1 library-generated table and 1 graph supporting the results of the analysis	**1 pts** **Minimal** Presents at least 1 table and 1 graph	**0 pts** **Insufficient** No charts	4 pts
Nature and Purpose How well is the nature and purpose of the investigation described	**4 pts** **Advanced** Describes motivation for research, ties topic into published papers, and describes how results improve/extend/contrast with existing research	**3 pts** **Above Average** Describes motivation for research and ties topic into published papers, but does not describe how results improve/extend/contrast with existing research	**2 pts** **Average** Describes motivation for research and describes how results improve/extend/contrast with existing research but does not tie topic into published papers	**1 pts** **Below Average** Describes motivation for research, but does not tie into published papers	**0 pts** **Insufficient** Does not describe motivation of investigation	4 pts
Demonstrates findings with equations	**4 pts** **Advanced** Equations are provided for all features and the model is supported with a detailed derivation	**3 pts** **Above Average** Equations are provided for all features but the model is only described in words	**2 pts** **Average** More than one equation is provided which describes either the model or the features used in the reserach	**1 pts** **Below Average** A single equation is provided	**0 pts** **Insufficient** No equations are provided	4 pts
Writing Quality How clearly does the paper present the research results	**4 pts** **Academic** Academic writing without grammatical/spelling errors	**3 pts** **Above Average** Colloquial writing without grammatical/spelling errors	**2 pts** **Average** Writing with minimal grammatical/spelling errors	**1 pts** **Below Average** Academic writing with many grammatical/spelling errors	**0 pts** **Insufficient** Colloquial writing with many grammatical/spelling errors	4 pts
References and Sources How well does the paper tie the research topic into existing research	**4 pts** **Advanced** Provides at least two references and makes connection with research topic	**3 pts** **Above Average** Provides one references and makes connection with research topic	**2 pts** **Average** Provides at least two references but makes no connection with research topic	**1 pts** **Below Average** Provides one reference but makes no connection of research topic	**0 pts** **Insufficient** Not Included	4 pts

Criteria	Ratings					Pts
Look-ahead Bias How well is the time-series nature of the data handled during the model creation and testing process (may include validation set if cross-validation is used)	**4 pts** **Advanced** Features are generated with strictly historical data and training sample(s) are prior to hold-out validation and testing sample(s) -- or analysis was based on contemporaneous (not predictive) data	**3 pts** **Above Average** Features are generated with strictly historical data and training sample(s) are prior to testing sample(s) but not cross-validation sample(s)	**2 pts** **Average** Features are generated with strictly historical data and training sample(s) are prior to cross-validation sample(s) but not testing sample(s)	**1 pts** **Below Average** Features are generated with strictly historical data but training sample(s) are prior to testing sample(s) and cross-validation sample(s)	**0 pts** **Insufficient** Features are generated using concurrent/future data and training sample(s) are not prior to testing sample(s) or cross-validation sample(s)	4 pts
Market Impact Analysis	**4 pts** **Advanced** Includes market impact estimates in profitability analysis or demonstrates how the analysis which directly models market impact can be used to influence trading decisions	**3 pts** **Above Average** Discusses how market impact would influence profitability or how the analysis which directly models market impact can be used to influence trading decisions	**2 pts** **Average** Discusses market impact in general and how it relates to the research topic -- or topic does not model price changes	**1 pts** **Below Average** Topic discusses price changes and/or trading strategies and discusses market impact in general	**0 pts** **Insufficient** Topic discusses price changes and/or trading strategies but does not include any discussion on market impact	4 pts
Notebook Quality How well is the research documented	**4 pts** **Advanced** Notebook includes description of each step along with formulas, and execution completes successfully within 15 minutes	**3 pts** **Above Average** Notebook includes description of each step and execution completes successfully within 15 minutes	**2 pts** **Average** Notebook includes description of each step along with formulas, but execution does not complete successfully within 15 minutes	**1 pts** **Below Average** Notebook includes description of each calculation but execution does not complete successfully within 15 minutes	**0 pts** **Insufficient** Notebook only includes code, and execution does not complete successfully within 15 minutes	4 pts
Conclusion How well does the paper summarize the research results and suggest further improvements	**4 pts** **Advanced** Summarizes research and findings, connects with prior research and discusses future extensions which can be made	**3 pts** **Above Average** Summarizes research and findings, and connects with prior research but does not discuss future extensions which can be made	**2 pts** **Average** Summarizes research and findings, and discusses future extensions which can be made but does not connect with prior research	**1 pts** **Below Average** Summarizes research and findings but does not connect with prior research or discuss future extensions which can be made	**0 pts** **Insufficient** Does not summarize research and/or findings	4 pts
Innovation How well does the paper innovate by introducing new ideas or	**5 pts** **Advanced** Advances market microstructure research by introducing new findings or theory		**3 pts** **Average** Extends or generalizes findings from an existing paper		**0 pts** **Insufficient** Merely implements an existing paper or fails to articulate/investigate new concepts.	5 pts

Criteria	Ratings					Pts
extending existing research						
Depth of Analysis How clearly does the paper present the tradeoffs between a simple model and interpretability (complex ML models often have uninterpretable results)	**5 pts** **Academic** Provides in-depth reasoning for model selection and interprets the results in transparent and clearly understandable terms	**4 pts** **Above Average** Briefly explains reasons for model selection and explains the model results in understandable terms	**3 pts** **Average** Interprets the model results in transparent and clearly understandable terms but does not explain reason for model selection	**2 pts** **Below Average** Explains reasons for model selection but does not interpret model results in transparent and clearly understandable terms	**0 pts** **Insufficient** Does not explain reasons for model selection or interpret model results	5 pts
						Total Points: 50