

All questions and sub-questions are mandatory.

- Answers should be typed in where possible. In any case be aware that answers need to be CLEARLY readable in order to be marked.
- Part of the Python code used in class to clean and pre-process LOB data can be used.
- Answers will be checked for plagiarism (also with lecture notes and slides).
- Answers will be marked for scientific quality and clarity of explanations.
- Submit ONE pdf file with all answers and screenshots of code snapshots (for question 1).
- Maximum Word limits are strict (no tolerance is applied).
- The coursework guidance folder contains LOB data to be used in question 1 and the research paper on which part of question 3 is based on ("Measuring and Modeling Execution Cost and

Assignment Project Exam Help

<https://eduassistpro.github.io/>

Add WeChat edu_assist_pro

1. Empirical properties of LOBs [40 marks]

For this question please consider the LOB data attached for 5 stocks, namely AMZN, GOOG, INTC, AAPL, MSFT. To perform the analysis and produce the plots you can use a programming language of your choice (e.g., Python, Matlab).

- a. Compute and plot the mid-price $m(t)$ and its moving average over 10 minutes for the ticker APPL.
- b. Calculate and plot the log-returns versus time for GOOG.
- c. Within the `message` file of the AMZN LOB data consider only the visible executions events. Consider the trade signs of the visible executions (ϵ_t) and write a function to compute the lagged autocorrelation $C_\epsilon(\tau)$ as a function of the lag $\tau = 0, \dots, 150$. Briefly discuss the observed statistical properties of the order flow.
[Max 150 words]
- d. What are the two types of order flow?
- e. For **each** stock, compute the *relative tick size*. Explain the procedure and comment on the results obtained.
- f. For **each** stock, calculate and plot the size of the *first gap* over time and its time average over all events. Explain the procedure and comment on the results obtained.
- g. For **each** stock, calculate the spread averaged over all events.
- h. Produce a scatter plot of the average spread and average gap calculated for all stocks and comment on the results obtained.

[Question 2 cont. on next page]

2. Propagator models [30 Marks]

a. Provide a definition of price impact. [Max 300 words]

b. One of the measures of price impact is the following:

$$R(\ell) = \langle (\epsilon_k \cdot (x_{k+\ell} - x_k)) \rangle_k$$

where x is the mid-price (computed just before and immediately after the arrival of an LOB event, for instance a market order arrival) and ϵ_k is the sign of the k -th trade.

Discuss the observed behaviour of the response function $R(\ell)$ in the context of:

- (i) the impact of individual trades (market orders, limit orders and cancellations).
- (ii) the dependence of the response function $R(\ell)$ (and price impact) on the order size.
- (iii) the effect of non-symmetric liquidity provision.

c. Another rel

return r is the return and ϵ is the trade sign (the average is computed over time). u-
lations it can be written as a function of t.

d. Which are the main classes of so-called propagator models introduced to reproduce and estimate price impact? Briefly discuss the mathematical setup and the main assumptions. [Max 500 words]

e. Briefly describe how the propagator can be estimated from the empirical data. [Max 300 words]

[Question 3 cont. over page]

3. Empirical cost analysis [30 marks]

- a. Discuss the type of execution costs that investors face when submitting their orders and how different types of cost affect different types of market participants. [**Max 200 words**]
- b. Describe the different choices available to traders in the context of an execution of multiple trades. [**Max 200 words**]

Consider the paper “Measuring and Modeling Execution Cost and Risk” for the following questions. The paper examines transaction costs from an empirical point of view. To answer the following questions you can refer to specific plots or tables included in the paper.

- c. Describe the features of the dataset used in the paper and explain the data cleaning process. [**Max 300 words**]
- d. Describe the type of risk used to assess the risk.
- e. Discuss in what lies the novelty of the paper and the main result. [**Max 500 words**]
- f. What conclusions can be drawn in terms of the relationship between market conditions (e.g., spread and liquidity), type of order (e.g., sizes) and transaction cost and risk? [**Max 500 words**]