RGC Ref. No.:
UGC/FDS16/B10/17

(please insert ref. above)

RESEARCH GRANTS COUNCIL
COMPETITIVE RESEARCH FUNDING SCHEMES FOR
THE LOCAL SELF-FINANCING DEGREE SECTOR

FACULTY DEVELOPMENT SCHEME (FDS)

Completion Report
(for completed projects only)

Submission Deadlines: 1. *Auditor's report with unspent balance, if any: within **six** months of the approved project completion date.*
2. *Completion report: within **12** months of the approved project completion date.*

Part A: The Project and Investigator(s)

1.   Project Title

Quantifying the impact of limit order book transparency on trading profitability

2.   Investigator(s) and Academic Department(s) / Unit(s) Involved

Research Team	Name / Post	Unit / Department / Institution
Principal Investigator	Samuel Ping-Man Choi / Assistant Professor	Lee Shau Kee School of Business and Administration, / Hong Kong Metropolitan University
Co-Investigator	Franklin Sze-Sing Lam / Associate Professor	Lee Shau Kee School of Business and Administration, / Hong Kong Metropolitan University
Co-Investigator	Hie-Yiin HUNG / Senior Lecturer (Retired in January 2021)	Lee Shau Kee School of Business and Administration, / Hong Kong Metropolitan University

3.   Project Duration

	Original	Revised	Date of RGC / Institution Approval *(must be quoted)*
Project Start Date	1 January 2018	NA	
Project Completion Date	31 December 2019	31 December 2020	September 2019
Duration *(in month)*	24	36	September 2019
Deadline for Submission of Completion Report	31 December 2020	31 December 2021	September 2019

Part B: The Final Report

5.   Project Objectives

5.1   Objectives as per original application

1.   To measure the trading profitability under various degrees of LOB transparency.

2. To evaluate the impact of LOB transparency on the market design and the performance of AT and HFT strategies.

3. To identify the key LOB features that can benefit AT and HFT strategies.

5.2   Revised objectives

Date of approval from the RGC: N.A.

Reasons for the change:

5.3   Realisation of the objectives
       (Maximum 1 page; please state how and to what extent the project objectives have been achieved; give reasons for under-achievements and outline attempts to overcome problems, if any)

1. To measure the trading profitability under various degrees of LOB transparency.

We have measured three degrees of LOB transparency: anonymous LOB, broker IDs and LOB with broker ID. In terms of trading profitability, three models achieved, on average, 73.94%, 58.89%, and 70.50% correct trade decisions respectively.

2. To evaluate the impact of LOB transparency on the market design and the performance of AT and HFT strategies.

Our empirical result strongly suggests that the broker ID queues in the LOB consist of significant information content for price prediction and it provides insights for regulators to determine the appropriate degree of LOB transparency so as to enable a fair market design for all investors

3. To identify the key LOB features that can benefit AT and HFT strategies.
The identified key LOB features are: top 5 bid and ask prices and their order volumes in the LOB, top 10 broker IDs and the mid-price movement for each tick.

5.4   Summary of objectives addressed to date

Objectives *(as per 5.1/5.2 above)*	Addressed *(please tick)*	Percentage Achieved *(please estimate)*
1. To measure the trading profitability under various degrees of LOB transparency	✔	100%
2. To evaluate the impact of LOB transparency on the market design and the performance of AT and HFT strategies	✔	100%
3. To identify the key LOB features that can benefit AT and HFT strategies.	✔	100%

6.   Research Outcome

6.1   Major findings and research outcome
(Maximum 1 page; please make reference to Part C where necessary)

Our research proposes adopting deep learning as a new instrument to measure the information content of broker ID queues in the limit order book (LOB). With deep learning, it becomes feasible to quantitatively evaluate the effect of anonymity by comparing the prediction accuracy of price movement using the LOB data with and without broker identities.

With a huge Hong Kong stock LOB dataset, our empirical result indicates that using only the broker ID queues in the LOB can achieve, on average, 85.13% overall prediction accuracy comparing to using the anonymous LOB information for prediction horizon k=20. However, the performance differences diminish as k increases, and both models perform virtually the same when k=100. This outcome is unintuitive since the broker ID dataset contains no information related to price and volume, nor any information regarding the institutional brokers and retail brokers. It implies that the broker ID queues in the LOB comprise significant information content which is valuable to traders. Our study also provides a measurable value of the broker information in the bid-ask queue for the market data provider to determine an appropriate price for the level 2 data. It should also be noted that the prediction accuracy reflects only the lower-bound, rather than the actual, value of the LOB feature for a given learning model. The prediction performance could increase if a more suitable learning model can be identified and adopted for the LOB dataset.

The contributions of our research are two folds. First, a machine-learning based tool for finance researchers is proposed to quantitatively measure the price predictability of LOB features, and an empirical study is conducted on the impact of LOB transparency on traders' profitability. Second, the empirical result strongly suggests that the broker ID queues in the LOB consist of significant information content for price prediction and it provides insights for regulators to determine the appropriate degree of LOB transparency so that a fair market for all investors is guaranteed.

6.2   Potential for further development of the research and the proposed course of action
(Maximum half a page)

Several research questions remained to be answered. For instance, how can broker ID sequences capture the price movement characteristics? Do institutional brokers or retail brokers in LOB offer more significant information with respect to the short-term price movement under different market trends? Can investors gain advantages on trading profitability with more LOB information?

For the first and second questions, further investigation is required to identify what exact information of broker ID that can contribute to the high prediction accuracy. We conjecture that the broker IDs appearing in the LOB provide sufficient evidence to distinguish among various categories of brokers such as institutional brokers and retail brokers. In addition, the change of broker sequence in the LOB may indicate the short-term momentum of a stock. For the last question, it would be interesting to explore further whether the prediction performance can be boost with the complete LOB dataset (i.e., levels 1 to 4 data). In theory, more information should provide better prediction performance. However, our pilot study shows that a straightforward mixture of the anonymous LOB and broker IDs for the above deep learning model does not yield any significant improvement on the prediction accuracy. In fact, the prediction performance of using both bid-ask prices and broker ID results in roughly 2-3% worse in accuracy when normalizing broker ID with occurrence frequency. There are two main reasons for the unsatisfactory result. First, the collected LOB data is incomplete; it contains only an aggregated order volume for a specific price instead of an order for each individual broker. Second, broker ID and bid-ask prices are two distinctive data types (the former is categorical while the latter is numerical), and thus directly combining both in

one model results in disappointing performance. A new model architecture is thus required to judiciously combine the anonymous LOB and broker IDs in order to achieve higher prediction results. In the future, further investigation effort should be spent towards these directions.

7. Layman's Summary
(Describe <u>in layman's language</u> the nature, significance and value of the research project, in no more than 200 words)

Limit order books (LOBs) have been widely adopted as a trading mechanism in global securities markets and the degree of LOB transparency is one of the most studied topics in market design. Our research aims to analyze the importance of broker IDs in the LOB with respect to the price movement predictability through a deep learning model. Our result indicates that the prediction accuracy of anonymous LOB and broker ID models are fairly high, ranging from 57.63% to 68.70% and 53.70% to 59.39% respectively. When comparing the performance of both datasets, it is surprising that the overall prediction performance solely based on broker ID dataset can reach, on average, 85.13% that of the anonymous LOB dataset. The contributions of this study are two folds. First, a machine-learning based tool for finance researchers is proposed to quantitatively measure the price predictability of LOB features, and an empirical study is conducted on the impact of LOB transparency on traders' profitability. Second, the empirical result strongly suggests that the broker ID queues in the LOB contain significant information for price prediction and it provides insights for regulators to determine the appropriate degree of LOB transparency to maintain a fair market for all investors.

Part C: Research Output

8. **Peer-Reviewed Journal Publication(s) Arising <u>Directly</u> From This Research Project**
    (Please attach a copy of the publication and/or the letter of acceptance if not yet submitted in the previous progress report(s). All listed publications must acknowledge RGC's funding support by quoting the specific grant reference.)

The Latest Status of Publications				Author(s) *(denote the corresponding author with an asterisk*)*	Title and Journal / Book *(with the volume, pages and other necessary publishing details specified)*	Submitted to RGC *(indicate the year ending of the relevant progress report)*	Attached to this Report *(Yes or No)*	Acknowl-edged the Support of RGC *(Yes or No)*	Accessible from the Institutional Repository *(Yes or No)*
Year of Publication	Year of Acceptance *(For paper accepted but not yet published)*	Under Review	Under Preparation *(optional)*						
2021					Analyzing the Importance of Broker Identities in the Limit Order Book Through Deep Learning. Big Data. 2021 Apr;9(2):89-99. doi: 10.1089/big.2020.0053. Epub 2020 Nov 17. PMID: 33202194.	No	Yes [Attachment 1]	Yes	Yes

9. **Recognized International Conference(s) In Which Paper(s) Related To This Research Project Was / Were Delivered**
    (Please attach a copy of each conference abstract)

Month / Year / Place	Title	Conference Name	Submitted to RGC *(indicate the year ending of the relevant progress report)*	Attached to this Report *(Yes or No)*	Acknowledged the Support of RGC *(Yes or No)*	Accessible from the Institutional Repository *(Yes or No)*
2019 / October / London, UK	Measuring the Price Predictability of Broker Identities in the Limit Order Book – a Deep Learning Approach	ICBIB 2019: International Conference on Big Data in Business	No	Yes [Attachment 2]	Yes	Yes

10. **Whether Research Experience And New Knowledge Has Been Transferred / Has Contributed To Teaching And Learning**
(Please elaborate)

The research findings have been shared with undergraduate students in FINT B337F

(Algorithmic and High-Frequency Trading).

11. **Student(s) Trained**
(Please attach a copy of the title page of the thesis)

Name	Degree Registered for	Date of Registration	Date of Thesis Submission / Graduation
NA			

12. **Other Impact**
(e.g. award of patents or prizes, collaboration with other research institutions, technology transfer, teaching enhancement, etc.)

NIL

13. **Statistics on Research Outputs**

	Peer-reviewed Journal Publications	Conference Papers	Scholarly Books, Monographs and Chapters	Patents Awarded	Other Research Outputs (please specify)	
No. of outputs arising directly from this research project	1	1	0	0	Type	No.
					NIL	

14. Public Access Of Completion Report
(Please specify the information, if any, that cannot be provided for public access and give the reasons.)

Information that Cannot Be Provided for Public Access	Reasons
NA	