School of Electrical and Information Engineering

The University of Sydney

Computational models for stock market order submissions

Adam Blazejewski

A thesis submitted in fulfilment

of the requirements for the degree of

Doctor of Philosophy

March 2005

To my parents.

Abstract

The motivation for the research presented in this thesis stems from the recent availability of high frequency limit order book data, relative scarcity of studies employing such data, economic significance of transaction costs management, and a perceived potential of data mining for uncovering patterns and relationships not identified by the traditional top-down modelling approach. We analyse and build computational models for order submissions on the Australian Stock Exchange, an order-driven market with a public electronic limit order book. The focus of the thesis is on the trade implementation problem faced by a trader who wants to transact a buy or sell order of a certain size.

We use two approaches to build our models, top-down and bottom-up. The traditional, top-down approach is applied to develop an optimal order submission plan for an order which is too large to be traded immediately without a prohibitive price impact. We present an optimisation framework and some solutions for non-stationary and non-linear price impact and price impact risk. We find that our proposed transaction costs model produces fairly good forecasts of the variance of the execution shortfall. The second, bottom-up, or data mining, approach is employed for trade sign inference, where trade sign is defined as the side which initiates both a trade and the market order that triggered the trade. We are interested in an endogenous component of the order flow, as evidenced by the predictable relationship between trade sign and the variables used to infer it.

We want to discover the rules which govern the trade sign, and establish a connection between them and two empirically observed regularities in market order submissions, competition for order execution and transaction cost minimisation.

To achieve the above aims we first use exploratory analysis of trade and limit order book data. In particular, we conduct unsupervised clustering with the self-organising map technique. The visualisation of the transformed data reveals that buyer-initiated and seller-initiated trades form two distinct clusters. We then propose a local non-parametric trade sign inference model based on the k-nearest-neighbour classifier. The best k-nearest-neighbour classifier constructed by us requires only three predictor variables and achieves an average out-of-sample accuracy of 71.40% (SD=4.01%)[1], across all of the tested stocks. The best set of predictor variables found for the non-parametric model is subsequently used to develop a piecewise linear trade sign model. That model proves superior to the k-nearest-neighbour classifier, and achieves an average out-of-sample classification accuracy of 74.38% (SD=4.25%). The result is statistically significant, after adjusting for multiple comparisons.

The overall classification performance of the piecewise linear model indicates a strong dependence between trade sign and the three predictor variables, and provides evidence for the endogenous component in the order flow. Moreover, the rules for trade sign classification derived from the structure of the piecewise linear model reflect the two regularities observed in market order submissions, competition for order execution and transaction cost minimisation, and offer new insights into the relationship between them. The obtained results confirm the applicability and relevance of data mining for the analysis and modelling of stock market order submissions.

[1]SD denotes standard deviation.

Acknowledgements

I would first like to express my gratitude to my supervisor, Dr Richard Coggins, for leading me patiently yet persistently through the many ups and downs of my research. His contribution is particularly appreciated because he agreed to take me on at the crossroads of my candidature, in July 2001. I am also thankful for his belief in me when I ventured into a new field, computational finance. During the early stages of my PhD research I benefited greatly from the guidance of Professor Marwan Jabri, who was my supervisor between March 2000 and June 2001. Professor Mike Aitken was probably the most important influence on my path to knowledge in the field of finance. He introduced me to the world of stock exchanges, brokers, market surveillance, liquidity and transaction costs. Of equal, if not more importance, has been his role as the CEO of the Capital Market Cooperative Research Centre (CMCRC).

The CMCRC has created a collaborative environment where academics, PhD students, and practitioners from the industry can work together. The CMCRC and its industry partners provided the necessary computer equipment, financial datasets, and software that made my research possible. In particular, a unique dataset, comprising all trades and orders, and the full contents of the limit order book, as registered by the Australian Stock Exchange, was made available to me. Furthermore, I had an opportunity to spend three months on a project with ABN AMRO, which was a valuable experience. I also participated in two market microstructure courses organised by the CMCRC. The first course was conducted by Professor Jay Muthuswamy, while the second one was taught by

Dr Carole Comerton-Forde and Dr Elvis Jarnecic. I found both courses very informative, helpful in organising my research ideas, and useful in getting to know other people involved in the studies related to my own work.

Over the course of my candidature I received financial support in the form of the CMCRC scholarship, Australian Postgraduate Award scholarship, and the Norman I Price scholarship. These scholarships are gratefully acknowledged. They provided me with long term support that enabled the scientific pursuit of the problems and ideas presented in my thesis. I would also like to thank Professor Graham Partington, who made sure that the CMCRC scholarship programme worked very well.

I had the pleasure to work with and meet many dedicated people who in various ways helped to bring my research to completion. In particular, Fil Maisetti and Franc Carter offered a lot of support and advice when dealing with the hardware and software issues. Excellent assistance with proof-reading was provided by Mrs Inge Rogers. Many useful suggestions and constructive criticism were offered by the examiners of the thesis. To them and to other people whose names are not mentioned here I would like to say “Thank you”.

Last but not least, I am very grateful to my parents for introducing me to the world of science, and to my wife, Yining, for she is always there for me and her smile fills my soul with joy.

Statement of originality

The research presented in this thesis was conducted at the School of Electrical and Information Engineering, The University of Sydney, and at the premises of the industry partner, Capital Markets Cooperative Research Centre (CMCRC), between July 2001 and March 2005 under the supervision of Dr Richard Coggins. I declare that I am the sole author of the thesis. The only exception is chapter 3 which was a collaborative effort with Dr Richard Coggins, with each co-author contributing equally. Whenever the pronoun 'we' is used it refers to the author only, except for chapter 3. The entire thesis is an original work, unless otherwise indicated or referenced. The specific original contributions are as follows:

1. Estimation procedures for the parameters used in trading plan optimisation, including σ, η, h_{v0}, ϵ, β, f_{v0}, α, and the order volume as shown in Figure 3.1, have been derived empirically by the author (Chapter 3 [56]).

2. All formulae and their derivations in chapter 3 have been adapted from referenced sources or independently developed by Dr Richard Coggins. The analytical formulae presented in section 3.2, and in particular the analytic optimal trading plan solution given by formula 3.11, have been derived by the author and Dr Richard Coggins (Chapter 3).

3. The idea, of applying the self-organising map (SOM) technique for unsupervised clustering of multi-dimensional trade level data leading to the observation that trade sign forms two clusters in the SOM, is that of the author (Chapter 4 [34]).

4. The idea of applying a k-nearest-neighbour classifier for the non-parametric classification of trade sign is that of the author (Chapter 5 [35]).

5. The set of candidate predictor variables **V**, the set of constraints **C** for combining candidate predictor variables into variable sets, and the discovery of the best combinations of the predictor variable set and the training interval length, for the non-parametric classification of trade sign are due to the author (Chapter 5 [35]).

6. The two dimensional projection of the space of three predictor variables and trade sign, shown in Figure 6.1, and the discovery of the six distinct regions in that projection are due to the author (Chapter 6 [36]).

7. The idea of applying a piecewise linear classifier for the parametric classification of trade sign, the classifier's formulae 6.1 and 6.2, and the interpretation of the classification accuracy within the six regions of the piecewise linear classifier as a reflection of two behavioural regularities, competition for order execution and transaction cost minimisation, are due to the author (Chapter 6 [36]).

8. All Alice code developed in the SMARTS® system for data extraction, data pre-processing, and other tasks, and all Matlab® code, except that written by Dr Richard Coggins for the calculation of optimal trading plans in chapter 3, developed for data pre-processing, classifier training and testing, data visualisation, statistical processing, and other tasks, and the trading simulation code developed in gcc on Unix for measuring the performance of various pre-computed trading strategies, unless otherwise noted, is due to the author (Chapters 3 to 6 [34–36, 56]).

I hereby declare that the work presented in this thesis has not been previously submitted for the award of a degree or a diploma at The University of Sydney or elsewhere.

Adam Blazejewski

29 March, 2005

List of research papers

The work presented in this thesis has also been reported by the author in the following papers:

Reviewed journal papers

1. Blazejewski, A., Coggins, R., 2005. A local non-parametric model for trade sign inference. Physica A: Statistical Mechanics and its Applications 348 (15 March), 481–495.
   Thesis chapter 5, Ref. [35].

Journal papers under review

1. Blazejewski, A., Coggins, R., 2005. A piecewise linear model for trade sign inference. To be resubmitted to Physica A: Statistical Mechanics and its Applications.
   Thesis chapter 6, Ref. [36].

Reviewed conference papers

1. Blazejewski, A., Coggins, R., 2004. Application of self-organising maps to clustering of high-frequency financial data. In: Purvis, M. (Ed.), Proceedings of the Australasian Workshop on Data Mining and Web Intelligence (DMWI2004), 2004, Dunedin, New Zealand. Vol. 32 of CRPIT. Australian

Computer Society, Inc., pp. 85–90.
Thesis chapter 4, Ref. [34].

2. Coggins, R., Blazejewski, A., Aitken, M., 2003. Optimal trade execution of equities in a limit order market. In: Proceedings of the IEEE International Conference on Computational Intelligence for Financial Engineering (CIFEr'03), 2003, Hong Kong. Hong Kong, pp. 371–378.
Thesis chapter 3, Ref. [56].

Reviewed conference presentations

1. Blazejewski, A., Coggins, R., Aitken, M., 2003. Dynamic non-parametric model for trade direction forecasting. Presented at the 16th Australasian Finance & Banking Conference, 2003, Sydney, Australia.
Thesis chapter 5, Ref. [37].

Contents

List of Figures

List of Tables