Topics in market microstructure

Zovko, I.

Chapter 1

Introduction

The topic of this thesis is Market microstructure. Market microstructure is an area of finance that studies the dynamics and processes through which investors' forecasts about future asset values are ultimately translated into the assets' current prices and trading volumes. The field encompasses also the study of trading rules which regulate the markets and constrain the actions of traders. In even broader terms, research directions that deal with the interrelation between institutional structure, strategic behavior, prices and welfare are all considered market microstructure.

The topics investigated in this thesis are also related to the field of Econophysics. Econophysics is a multidisciplinary field where ideas from physics and economics meet. Over the last 10 years or so, a large number of papers has been published in physics journals dealing with issues in economics and finance. Arguably, many of those ideas had already seen light in economics, but there are a number of ideas that contribute in an original way to finance and economic literature. A growing number of econophysics papers are now being published in more-and-more mainstream economics and finance journals.

Chapters 2 and 3 of this thesis are based on two papers[§†] and are related to the literature dealing with the microstructure of limit order books. Chapter 4 is based on a paper[‡] related to the litera-

[§] I. I. Zovko and J. D. Farmer. *The power of patience; a behavioral regularity in limit order placement.* Quantitative Finance, 2(5):387392, 2002.

[†] J. D. Farmer, P. Patelli, and I. Zovko. *The predictive power of zero intelligence in financial markets.* Proceedings of the National Academy of Sciences of the United States of America, 102(6):22549, Feb 8 2005.

[‡] I. I. Zovko and J. D. Farmer, *Correlations and clustering in the trading of members of the London Stock Exchange*, in Complexity, Metastability and Nonextensivity: An International Conference, S. Abe, T. Mie, H. Herrmann, P. Quarati, A. Rapisarda, and C. Tsallis, Eds., Springer, 2007.

ture on heterogeneous agent behavior in finance. The final chapter, Chapter 5, is related to agent heterogeneity, market microstructure, and information content of trades.

1.1 The London Stock Exchange and the LSE data

The research in this thesis is based on a dataset from the London Stock Exchange (LSE) from roughly 1998 to 2002. Some aspects of the LSE markets incurred changes since then and the information we give here refers to the period of the analysis. The LSE is one of the largest equity markets in the world facilitating trading in many British and international stocks. In the chapters however we use data only for a selection of British stocks from the primary listing as for those stocks the LSE is the primary market. Most of these stocks enter the FTSE 100 or FTSE 250 index.

The LSE market for the analysed stocks is a hybrid market with parallel trading in an open electronic limit order book (on-book) and a quotation block market (off-book). At the LSE, the on-book session is called the SETS (Stock Exchange Electronic Trading System), and the off-book session the SEAQ (Stock Exchange Automated Quotation System). The papers contained in the first two chapters of the thesis focus only on the limit order trade process and use only the on-book data. The last two chapters use also the off-book data and provide a comparison in some aspects of the two market designs.

1.1.1 Trading day

For the FTSE 100 stocks, the on-book trading session starts at 8:50 with a 10 minute opening auction. During the auction traders place orders to buy and sell but no execution takes place. Orders are differentiated by their execution priority. For example, limit orders are executed depending on their distance from the resulting clearing price while market orders take priority in execution. In case of insufficient volume for all the market orders to clear, priority is based on time of submission. At the end of the auction, a clearing price is calculated by a relatively complicated algorithm whose objective is basically to maximize the trade volume. It is this clearing price that is quoted in the papers as the *market opening price*. The exact time of the ending of the auction is random up to 30 seconds, i.e., the auction ends at 9:00 plus a random time interval less then 30 seconds.

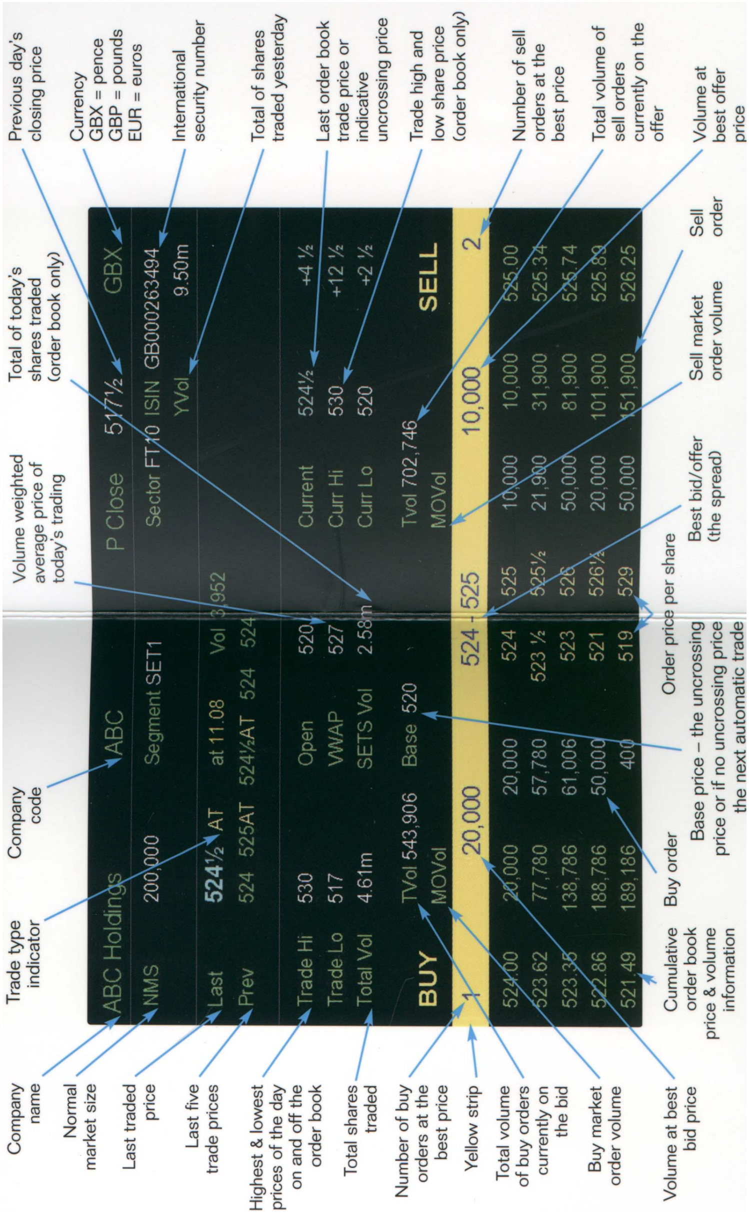

Figure 1.1: An example of a trading screen for the SETS (Stock Exchange Electronic Trading System), the electronic open limit order book of the London Stock Exchange. In this screen the traders are displayed, among some general stock information such as the previous day closing price and volume, also the current best bid and ask prices and volumes and also up to 5 levels of bid and ask limit orders in the limit order book.

Once the opening auction is over, the market enters the continuous double auction phase. The possible uncleared orders from the auction are transfered to the order book. The continuous double auction is the main trading phase of the market. Traders may continuously submit orders to buy or sell and possible trades are cleared instantly. The main two types of orders are *limit orders* and *market orders*. A sell (buy) limit order is an offer to sell (buy) a specific amount of shares at a specified price or higher (lower). These offers are recorded by the system and stored in the *limit order book*. Traders can cancel the limit orders they have submitted to the book at any time, unless they have resulted in a transaction. A sell (buy) market order is an order to sell (buy) a certain quantity of shares at the best currently available prices in the order book. A large market order can transact against multiple limit orders and at multiple prices if the volume at the best price is smaller then the market order volume.

In case of an exceptionally large price move (more then 10%-20% difference from the last transaction price), the trading system suspends trading and enters an auction period identical to the opening auction. This suspension of trading allows traders to process potentially new information.

The trading session ends at 16:30 with yet another auction period. The clearing price of this auction is quoted as the market closing price. At 16:40 plus a random interval less then 30 seconds the trading session closes.

During the on-book session, traders can also trade on the off-book market. The off-book market is an electronic quotation market ant the traders are ultimately arranged over phone. The off-book market is intended for large block trades.

1.1.2 Member firms

The member firms of the exchange are any firm that pays an anual fee to the LSE and participates in trading by directly sending orders to the exchange. The member firms typically are investment banks and hedge funds. Each member firm has to be registered with a *Clearing house* which guarantees its obligations resulting from trading. Some firms which are large investment banks are typically their own clearing house.

In the LSE data we used each member firm is assigned an *institutional code* which associates each market event to a particular member firm. The institutional codes do not reveal the real identify of an institution and in addition are scrambled at the turn of each month, across stocks and across markets. A colleague at the SFI, Marcus Daniels, was able to partially unscramble the codes across months. He used the fact that if an institution has an order standing in the orderbook at the turn of the month, and since each order has a unique identifier, one can observe with which institutional code the order is associated with prior to the scrambling and afterwards. In this way we can link the old code and the new code an institution is given at the recoding. This method works well for active institutions since they are likely to have an order standing in the book, but less well (if at all) for inactive ones. The institutional codes are used in Chapters 4 and 5.

1.1.3 The LSE dataset, preparation and cleaning

Many months of work have been spent to understand and prepare the dataset used in the thesis and ultimately by the entire Santa Fe Institute finance group. The candidate is very indebted to Marcus Daniels without whom it would have been impossible to make the data usable. The LSE data was purchased by Prof. Farmer as the basis for his Markets project and it came with no documentation and in a completely raw format on over a 100 CDs. It took a long time to understand the organization of the data and the functioning of the LSE.

In spite of the fact that the data is collected automatically (electronically) by the trading system in real time and is of high quality, there are numerous problems and missing data in the dataset. For example, there are days where prices in the orderflow are missing. The problems in the data are probably due to either system failures or upgrades as is with any real-life computer system. Luckily the data contained much redundancy. For example, from the on-book orderflow, one can reconstruct the book and transaction prices and volumes. These reconstructed transactions and volumes can then be verified against transaction or spread data contained in separate files.

To deal with these problems the dataset we use in our publications has been extensively cleaned and processed from it's original format. The basic idea behind the dataset cleaning is that the data be self-consistent. What we mean by this is the following: the order flow itself is sufficient to reconstruct the order book, the spread and all transaction information. On the other hand, the transactions and the spreads are provided in the data in addition to the orderflow. Therefore, one can reconstruct the book passing through the orderflow and reconstruct the spreads and transaction prices. One then compares this reconstructed information to the one provided in the data files. It is possible in many cases to infer missing data in the orderflow by requiring that the reconstructed spreads and transactions coincide with the provided data files.

In the end, we end with a large dataset containing the complete on-book orderflow, i.e., the volumes and prices of all limit orders, market orders and cancellations, time-stamped to the second. From the orderflow one can then calculate the transaction prices and volumes. The off-book data contains only transaction volumes and prices. It is also time-stamped, but by traders and brokers recording the deals.

1.2 Summary of chapters

In Chapter 2 (Zovko and Farmer, 2002) we demonstrate a striking regularity in the way people place limit orders in financial markets. Merging the data from 50 stocks traded on the LSE, we demonstrate that for both buy and sell orders, the unconditional cumulative distribution of relative limit prices decays roughly as a power law with exponent approximately -1.5. This behavior spans more than two decades, ranging from a few ticks to about 2000 ticks. The relative limit price is defined as the difference between the limit price and the best market price available at the moment in units of ticks, which is the minimal price increment at the market. We also find that the time series of relative limit prices show interesting temporal structure, characterized by an autocorrelation function that asymptotically decays as $C(\tau) \sim \tau^{-0.4}$. Furthermore, relative limit price levels are positively correlated with and are led by price volatility. We speculate that this feedback may potentially contribute to clustered volatility.

In Chapter 3 (Farmer et al., 2005) we turn our attention to market design and investigate a situation where the constraints imposed by market institutions may dominate strategic behavior of agents. We use the LSE limit order book data to test a simple model in which minimally intelligent agents place orders to trade at random. The model treats the statistical mechanics of order placement, price formation, and the accumulation of revealed supply and demand within the context of the continuous double auction, and yields simple laws relating order arrival rates to statistical properties of the market. We test the validity of these laws in explaining the cross-sectional variation for eleven stocks. The model explains 96% of the variance of the gap between the best buying and selling prices (the spread), and 76% of the variance of the price diffusion rate, with only one free parameter. We also study the market impact function, describing the response of quoted prices to the arrival of new orders. The non-dimensional coordinates dictated by the model approximately collapse data from different stocks onto a single curve. This chapter demonstrates the existence of simple laws relating prices to order flows, and in a broader context, it suggests that there are circumstances where the strategic behavior of agents may be dominated by other considerations.

Chapter 4 (Zovko and Farmer, 2007) analyzes correlations in patterns of trading of different member firms of the LSE. The collection of strategies associated with a member firm is defined by the sequence of signs of net volume traded by that firm in hour intervals. Using several methods we show that there are significant and persistent correlations between firms. In addition, the correlations are structured into correlated and anti-correlated groups. Clustering techniques using the correlations as a distance metric reveal a meaningful clustering structure with two groups of firms trading in opposite directions.

In the final Chapter 5, we show evidence that the heterogeneity of trade order sizes plays a role in price formation in addition to the signed order flow. Heterogenous composition of order sizes (e.g., almost all trade volume concentrated in one order) on the buy (bid) side, unless balanced by a similarly heterogenous sell (ask) side of the market, produces an imbalance which drives prices up, and vice versa. This effect is preset on both daily and hourly timescales. We show that a quotation market design (off-book or upstairs market), as opposed to a limit order design (on-book or downstairs market), helps limit the price impact of large orders causing the heterogeneity but does not remove it completely. In addition, the impact of a large order is limited in case the trading is done against similarly large orders, regardless of the market design. This fact seems to be at odds with the interpretation of information content of trades, and we propose it may be more liquidity that determines the impact of an order.

The heterogeneity of order sizes present at the market seems to be a consequence of the fat-tailed distribution of order sizes: for the on-book market with a tail exponent equal to 3, for the off-book market equal to 3/2 (tail exponents are for the cumulative distribution).