Topics in market microstructure

Zovko, I.

Citation for published version (APA):
Topics in Market Microstructure

Academisch Proefschrift

ter verkrijging van de graad van doctor
aan de Universiteit van Amsterdam
op gezag van de Rector Magnificus
prof. dr. D.C. van den Boom
ten overstaan van een door het college voor promoties
ingestelde commissie,
in het openbaar te verdedigen in de Agnietenkapel
op dinsdag 4 november 2008, te 12.00 uur

door

Ilija I. Zovko
geboren te Zagreb, Kroatië
Promotiecommissie:

Promotor: Prof. Dr. C.H. Hommes

Co-Promotor: Prof. Dr. J.D. Farmer

Overige laden: Prof. Dr. H.P. Boswijk

Prof. Dr. C.G.H. Diks

Prof. Dr. F.C.J.M. de Jong

Prof. Dr. T. Lux

Faculteit Economie en Bedrijfskunde
Universiteit van Amsterdam
Acknowledgements

The research in this thesis has been made possible by the contribution of numerous people with whom I have had the privilege to work with and learn from. First and foremost it was my thesis supervisor Prof. J. Doyne Farmer whose guidance and insight into market dynamics has been crucial for the choice of topics for the chapters of the thesis. Prof. Farmer is the leader of the Financial Markets group at the Santa Fe Institute (SFI), where I have had the privilege to do research.

I am also greatly indebted to Prof. Cars H. Hommes whose insightful suggestions and guidance was a bright light in at times a dark tunnel. Prof. Hommes is the director of the Center for Nonlinear Studies in Finance and Economics (CeNDEF) at the University of Amsterdam which I have had the privilege to be associated with and visit often.

The work presented in the thesis is a result of collaborations and informal discussions with many colleagues from the SFI and CeNDEF. Paolo Patelli, with whom I spent hours discussing and analysing market data, was a coauthor on the paper for the third chapter. Marcus Daniels, with whom I worked extensively throughout my first year at the SFI, cleaning up and understanding the London Stock Exchange data, was invaluable for building the computing infrastructure that allowed me and others at SFI to analyse the LSE market. I am grateful to Fabrizio Lillo for letting me steal some of his ideas over a glass of wine. I thank my colleagues Laszlo Gillemot, Szabolcz Mike, Mark Bieda, Makoto Nirei, Francois Ghoulimie and Yuzuru Sato for good fun and discussions.

I thank all the people at CeNDEF with whom I have had the privilege to exchange ideas and learn from during my visits to Amsterdam: Peter Boswijk, Cees Diks, Frank de Jong and Sebastiano Manzan.
I am indebted to all the people at the SFI whose help made my stay there possible and pleasant. Special thanks go to Laura Ware for her hospitality in homeless times and for her warm personality. Good friend and the responsible who made the computers at the SFI work, Nate Metheney. John Miller for his experienced academic advice, at times a “tough hand” and most importantly for showing me one example where game theory can be practical.\footnote{You can have a repeated game played by a number of flatmates in order to determine the allocation of rooms (which are of perceived different qualities) and the prices the tenants need to pay for them. The outcome of the game is an allocation of rooms and prices the tenants need to pay for them. After a few iterations of the game, one reaches a Nash equilibrium where none of the flatmates have an incentive to change his/her room given the price he/she is paying for it. It worked!} Rhonda Butler-Villa for helping in numerous issues with authorities (just the usual things...)

I also thank Constantino Tsallis for being a great role model; Ilya Peiros, Murray Gell-Mann, Goeffrey West, George Gummerman, Steve Lansing, Christopher Lantz and the late Sergei Starostin for the inspirational lectures and a general inspiring environment. I am grateful to Cormac McCarthy for pulling my car out of a ditch. Thanks go to people who I have had the privilege to meet and who took the time to answer my questions: Brian Arthur, J.P.Bouchaud, Sam Bowles, Michel Dacorogna, Raphael Douady, Thomas Lux, Soren Johansen, Rosario Mantegna and Ramon Marimon.

I am grateful to several institutions for providing the funding that allowed me to pursue academic studies without having to worry (for the most part) about the next lunch. The Santa Fe Institute and in particular to Prof. Farmer who offered me a stipend for my studies from his Markets grant. The Markets grant was provided by Credit Suisse-First Boston, McKinsey corporation, Bob Maxfield, Bill Miller. I am grateful to CeNDEF and Prof. Hommes, who kindly covered my lodging expenses for the durations of my stays in Amsterdam. I thank the SFI International program and its donors who supported a substantial part of my SFI stay. I am grateful to the Government of Croatia for the financial support they gave for my PhD studies which started with a Masters from the European University Institute in Florence, Italy.

In the end, needless to say, I am the most indebted to my parents for all reasons imaginable.
## 1 Introduction
1.1 The London Stock Exchange and the LSE data . . . . 2
   1.1.1 Trading day ........................................ 2
   1.1.2 Member firms ...................................... 4
   1.1.3 The LSE dataset, preparation and cleaning . . . 5
1.2 Summary of chapters .................................... 6

## 2 The power of patience: A behavioral regularity in limit order placement
2.1 Introduction ............................................. 9
2.2 Description of the London Stock Exchange data ....... 11
2.3 Properties of relative limit order prices .............. 12
   2.3.1 Unconditional distribution ......................... 12
   2.3.2 Time series properties ............................ 15
2.4 Volatility clustering .................................... 17
2.5 Conclusion ............................................. 19

## 3 The predictive power of zero intelligence in financial markets
3.1 Introduction ............................................. 21
   3.1.1 Continuous double auction .......................... 22
   3.1.2 Review of the model ................................ 24
   3.1.3 Predictions of the model ............................. 25
3.2 Testing the scaling laws ................................. 26
   3.2.1 Data ............................................... 26
   3.2.2 Testing procedure .................................. 27
   3.2.3 Spread ............................................. 28
   3.2.4 Price diffusion rate ............................... 28
3.3 Average market impact ................................. 29
3.4 Conclusions ............................................. 34
<table>
<thead>
<tr>
<th>Section</th>
<th>Title</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>3.5</td>
<td>Supplementary Material</td>
<td>37</td>
</tr>
<tr>
<td>3.5.1</td>
<td>Literature review</td>
<td>37</td>
</tr>
<tr>
<td>3.5.2</td>
<td>Dimensional analysis</td>
<td>40</td>
</tr>
<tr>
<td>3.5.3</td>
<td>The London Stock Exchange (LSE) data set</td>
<td>41</td>
</tr>
<tr>
<td>3.5.4</td>
<td>Opening auction, real order types, time</td>
<td>43</td>
</tr>
<tr>
<td>3.5.5</td>
<td>Measurement of model parameters</td>
<td>44</td>
</tr>
<tr>
<td>3.5.6</td>
<td>Estimating the errors for the regressions</td>
<td>48</td>
</tr>
<tr>
<td>3.5.7</td>
<td>Market impact</td>
<td>52</td>
</tr>
<tr>
<td>3.5.8</td>
<td>Extending the model</td>
<td>56</td>
</tr>
</tbody>
</table>

4 Correlation and clustering in the trading of the members of the LSE 59
4.1 Introduction .......................................................................... 59
  4.1.1 The LSE dataset ................................................................ 60
  4.1.2 Measuring correlations between strategies ...................... 60
4.2 Significance and structure in the correlation matrices .......... 64
  4.2.1 Density of the correlation matrix eigenvalue distribution ... 64
  4.2.2 Bootstrapping the largest eigenvalues ............................ 67
  4.2.3 Clustering of trading behaviour ..................................... 68
  4.2.4 Time persistence of correlations .................................... 71
4.3 Conclusions ........................................................................... 75

5 Market imbalances and stock ret.: heterogeneity of order sizes at the LSE 77
5.1 Introduction ........................................................................... 77
  5.2 Distribution of order sizes ................................................. 80
  5.3 Order size heterogen. and stock returns .............................. 82
  5.4 Conclusions ........................................................................... 94

6 Conclusions ............................................................................... 97
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-

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.
CHAPTER 1. INTRODUCTION

Once the opening auction is over, the market enters the continuous double auction phase. The possible uncleared orders from the auction are transferred 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 than the market order volume.

In case of an exceptionally large price move (more than 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 than 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 and 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 annual 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).
Chapter 2

The power of patience: A behavioral regularity in limit order placement


2.1 Introduction

Most modern financial markets are designed as a complex hybrid composed of a continuous double auction and an ‘upstairs’ trading mechanism serving the purpose of block trades. The double auction is believed to be the primary price discovery mechanism\(^1\). Limit orders, which specify both a quantity and a limit price (the worst acceptable price), are the liquidity providing mechanism for double auctions and the proper understanding of their submission process is important in the study of price formation.

\(^1\)According to the London Stock Exchange information bulletins (“SETS four years on - October 2001”, published by the London Stock Exchange), since the introduction of the SETS in 1997 to October 2001, the average percentage of trades in order book securities that have been executed at the price shown on the order book is 70% - 75%. Therefore SETS seems to serve as the primary price discovery mechanism in London.
CHAPTER 2. THE POWER OF PATIENCE: A BEHAVIORAL
REGULARITY IN LIMIT ORDER PLACEMENT

We study the relative limit price $\delta(t)$, the limit price in relation to the current best price. For buy orders $\delta(t) = b(t) - p(t)$, where $p$ is the limit price, $b$ is the best bid (highest buy limit price), and $t$ is the time when the order is placed. For sell orders $\delta(t) = p(t) - a(t)$, where $a$ is the best ask (lowest sell limit price). We find a striking regularity in the distribution of relative limit prices and we document clustering of order prices as seen by a slowly decaying autocorrelation function.

Biais, Hillion and Spatt (1995) study the limit order submission process on the Paris Bourse. They note that the number of orders placed up to five quotes away from the market decay monotonically but do not attempt to estimate the distribution or examine orders placed further then five best quotes. Our analysis looks at the price placement of limit orders across a much wider range of prices. Since placing orders out of the market carries execution and adverse selection risk, our work is relevant in understanding the fundamental dilemma of limit order placement: execution certainty vs. transaction costs (see, e.g., Cohen, et al. (1981); Harris (1997); Harris and Hasbrouck (1996); Holden and Chakravarty (1995); Kumar and Seppi (1992); Lo, et al. (2002)).

In addition to the above, our work relates to the literature on clustered volatility. It is well known that both asset prices and quotes display ARCH or GARCH effects (Engle (1982); Bollerslev (1986)), but the origins of these phenomena are not well understood. Explanations range from news clustering (Engle, et al. (1990)), macroeconomic origins (Campbell (1987); Glosten et al. (1993)) to microstructure effects (Lamoureux and Lastrapes (1990); Bollerslev and Domowitz (1991); Kavajecz and Odders-White (2001)). We provide empirical evidence that volatility feedback may in part be caused by limit order placement that in turn depends on past volatility levels.

This paper is organized as follows. Section II introduces the mechanics of limit order trading and describes the London Stock Exchange data we use. Section III presents our results on the distribution and time series properties of relative limit order prices. In section IV we examine the possible relationship of limit order prices and volatility which may lead to volatility clustering. Section V discusses and summarizes the result.

---

2 We have made a somewhat arbitrary choice in defining the “best price”. An obvious alternative would have been to choose the best ask as the reference price for buy orders, and the best bid as the reference price for sell orders. This would have the advantage that it would have automatically included orders placed inside the interval between the bid and ask (the spread), which are discarded in the present analysis. The choice of reference price does not seem to make a large difference in the tail; for large $\delta$ it leads to results that are essentially the same.
CHAPTER 2. THE POWER OF PATIENCE: A BEHAVIORAL REGULARITY IN LIMIT ORDER PLACEMENT

2.2 Description of the London Stock Exchange data

The limit order trading mechanism works as follows: As each new limit order arrives, it is matched against the queue of pre-existing limit orders, called the limit order book, to determine whether or not it results in any immediate transactions. At any given time there is a best buy price $b(t)$, and a best ask price $a(t)$. A sell order that crosses $b(t)$, or a buy order that crosses $a(t)$, results in at least one transaction. The matching for transactions is performed based on price and order of arrival. Thus matching begins with the order of the opposite sign that has the best price and arrived first, then proceeds to the order (if any) with the same price that arrived second, and so on, repeating for the next best price, etc. The matching process continues until the arriving order has either been entirely transacted, or until there are no orders of the opposite sign with prices that satisfy the arriving order’s limit price. Anything that is left over is stored in the limit order book.

On the London Stock Exchange, in addition to limit orders described above, traders can also submit crossing limit orders which result in immediate transactions while limiting market impact. Such crossing limit orders make up about 30% (in the example of Vodafone) of all limit orders and are more like market orders. In this paper we discard them and analyze only limit orders that enter the book. Of the analyzed orders 74% are submitted at the best quotes. Only 1% are submitted inside the spread (with $\delta < 0$), while the remaining 25% are submitted out of the market ($\delta > 0$). We investigate only limit orders with positive relative price $\delta > 0$ and refer to them in text simply as limit orders\(^3\).

The time period of the analysis is from August 1, 1998 to April 31, 2000. This data set contains many errors; we chose the names we analyze here from the several hundred that are traded on the exchange based on the ease of cleaning the data, trying to keep a reasonable balance between high and low volume stocks\(^4\). This left 50 different names, with a total of roughly seven million limit orders.

\(^3\)Even though limit orders placed in the spread are not numerous, they are very important in price formation. Dataset we use does not include enough events to provide statistically significant results. Our preliminary results indicate that orders placed in the spread behave qualitatively similar to orders placed out of the market, i.e., there are some indications of power law behavior in their limit price density towards the other side of the market.

\(^4\)The ticker symbols for the stocks in our sample are AIR, AL., ANL, AZN, BAA, BARC, BAY, BLT, BOC, BOOT, BPB, BSCl, BSY, BT.A, CCH, CCM, CS., CW., GLXO, HAS, HG., ICI, III, ISYS, LAND, LLOY, LMI, MKS, MNI, NPR, NU., PO., PRU, PSON, RB., RBOS, REED, RIO, RR., RTK, RTO, SB., SBRY, SHEL, SLP, TSCO, UNWS, UU., VOD, and WWH.
CHAPTER 2. THE POWER OF PATIENCE: A BEHAVIORAL REGULARITY IN LIMIT ORDER PLACEMENT

of which about two million are submitted out of the market ($\delta > 0$).

2.3 Properties of relative limit order prices

Choosing a relative limit price is a strategic decision that involves a tradeoff between patience and profit (e.g., Holden and Chakravarty (1995); Harris and Hasbrouck (1996); Sirri and Peterson (2002)). Consider, for example, a sell order; the story for buy orders is the same, interchanging “high” and “low”. An impatient seller will submit a limit order with a limit price well below $b(t)$, which will typically immediately result in a transaction. A seller of intermediate patience will submit an order with $p(t)$ a little greater than $b(t)$; this will not result in an immediate transaction, but will have high priority as new buy orders arrive. A very patient seller will submit an order with $p(t)$ much greater than $b(t)$. This order is unlikely to be executed soon, but it will trade at a good price if it does. A higher price is clearly desirable, but it comes at the cost of lowering the probability of trading – the higher the price, the lower the probability there will be a trade. The choice of limit price is a complex decision that depends on the goals of each agent. There are many factors that could affect the choice of limit price, such as the time horizon of the trading strategy. A priori it is not obvious that the unconditional distribution of limit prices should have any particular simple functional form.

2.3.1 Unconditional distribution

Figure (2.1) shows examples of the cumulative distribution for stocks with the largest and smallest number of limit orders. Each order is given the same weighting, regardless of the number of shares, and the distribution for each stock is normalized so that it sums to one. There is considerable variation in the sample distribution from stock to stock, but these plots nonetheless suggest that power law behavior for large $\delta$ is a reasonable hypothesis. This is somewhat clearer for the stocks with high order arrival rates. The low volume stocks show larger fluctuations, presumably because of their smaller sample sizes. Although there is a large number of events in each of these distributions, as we will show later, the samples are highly correlated, so that the effective number of independent samples is not nearly as large as it seems. To reduce the sampling errors we merge the data for all stocks, and estimate the sample distribution for the merged set using the method of ranks, as shown in figure (2.2). We fit the
Figure 2.1: (a) Cumulative distribution functions $P(\delta) = \text{Prob}\{x \geq \delta\}$ of relative limit price $\delta$ for both buy and sell orders for the 15 stocks with the largest number of limit orders during the period of the sample (those that have between 150,000 and 400,000 orders in the sample.) (b) Same for 15 stocks with the lowest number of limit orders, in the range 2,000 to 100,000. (To avoid overcrowding, we have averaged together nearby bins, which is why the plots appear to violate the normalization condition.)
resulting distribution to the functional form \(^5\)

\[
P(\delta) = \frac{A}{(x_0 + \delta)^\beta},
\]

\(^5\)The functional form we use to fit the distribution has to satisfy two requirements: it has to be a power law for large \(\delta\) and finite for \(\delta = 0\). A pure power law is either not integrable at 0 or at \(\infty\). If the functional form is to be interpreted as a probability density then it necessarily has to be truncated at one end. In our case the natural truncation point is 0. Clearly there is some arbitrariness in the choice of the exact form, but since we are mainly interested in the behavior for large \(\delta\), this functional form seems satisfactory.

\[A\] is set by the normalization, and is a simple function of \(x_0\) and \(\beta\). Fitting this to the entire sample (both buys and sells) gives \(x_0 = 7.01 \pm 0.05\), and \(\beta = 1.491 \pm 0.001\). Buys and sells gave similar values for the exponent, i.e. \(\beta = 1.49\) in both cases. Since these error bars based on goodness of fit are certainly overly optimistic, we also tested the stability of the results by fitting buys and sells separately on the first and last half of the sample, which gave values in the range \(1.47 < \beta < 1.52\). Furthermore, we checked whether there are significant differences in the estimated parameters for stocks with high vs. low order arrival rates. The results ranged from \(\beta = 1.5\)
CHAPTER 2. THE POWER OF PATIENCE: A BEHAVIORAL REGULARITY IN LIMIT ORDER PLACEMENT

for high to $\beta = 1.7$ for low arrival rates, but for the low arrival rate group we do not have high confidence in the estimate.

As one can see from the figure, the fit is reasonably good. The power law is a good approximation across more than two decades, for relative limit prices ranging from about $10 - 2000$ ticks. For British stocks ticks are measured either in pence, half pence, or quarter pence; in the former case, 2000 ticks corresponds to about twenty pounds. Given the low probability of execution for orders with such high relative limit prices this is quite surprising. (For Vodafone, for example, the highest relative limit price that eventually resulted in a transaction was 240 ticks). The value of the exponent $\beta \approx 1.5$ implies that the mean of the distribution exists, but its variance is formally infinite. Note that because normalized power law distributions are scale free, the asymptotic behavior does not depend on units, e.g. ticks vs. pounds. There appears to be a break in the power law at about 2000 ticks, with sell orders deviating above and buy orders deviating below. A break at roughly this point is expected for buy orders due to the fact that $p = 0$ places a lower bound on the limit price. For a stock trading at 10 pounds, for example, with a tick size of a half pence, 2000 ticks is the lowest possible relative limit price for a buy order. The reason for a corresponding break for sell orders is not so obvious, but in view of the extreme low probability of execution, is not surprising. It should also be kept in mind that the number of events in the extreme tail is very low, so this could also be a statistical fluctuation.

2.3.2 Time series properties

The time series of relative limit prices also has interesting temporal structure. This is apparent to the eye, as seen in figure (2.3b), which shows the average relative limit price $\delta$ in intervals of approximately 60 events for Barclays Bank. For reference, in figure (2.3a) we show the same series with the order of the events randomized. Comparing the two suggests that the large and small events are more clustered in the real series than in the shuffled series.

This temporal structure appears to be described by a slowly decaying autocorrelation function, as shown in figure (2.4).

One consequence of such a slowly decaying autocorrelation is the slow convergence of sample distributions to their limiting distribution. If we generate artificial IID data with equation (2.1) as its unconditional distribution, the sample distributions converge very quickly with only a few thousand points. In contrast for the real data, even for a stock with 200,000 points the sample distributions display large fluctuations. When we examine subsamples of the real data, the correlations in the deviations across subsamples are ob-
CHAPTER 2. THE POWER OF PATIENCE: A BEHAVIORAL REGULARITY IN LIMIT ORDER PLACEMENT

Figure 2.3: (a) Time series of randomly shuffled values of $\delta(t)$ for stock Barclays Bank. (b) True time series $\delta(t)$. (c) The absolute value of the change in the best price between each event in the $\delta(t)$ series.

Figure 2.4: The autocorrelation of the time series of relative limit prices $\delta$, averaged across all 50 stocks in the sample, and smoothed across different lags. This is computed in tick time, i.e., the x-axis indicates the number of events, rather than a fixed time.
previous and persist for long periods in time, even when there is no overlap in the subsamples. We believe that the slow convergence of the sample distributions is mainly due to the long range temporal dependence in the data.

2.4 Volatility clustering

To get some insight into the possible cause of the temporal correlations, we compare the time series of relative limit prices to the corresponding price volatility. The price volatility is measured as \( v(t) = |\log(b(t)/b(t-1))| \), where \( b(t) \) is the best bid for buy orders or the best ask for sell limit orders. We show a typical volatility series in figure (2.3c). One can see by eye that epochs of high limit price tend to coincide with epochs of high volatility.

To help understand the possible relation between volatility and relative limit price we calculate their cross-autocorrelation. This is defined as

\[
XCF(\tau) = \frac{\langle v(t-\tau)\delta(t) \rangle - \langle v(t) \rangle \langle \delta(t) \rangle}{\sigma_v \sigma_\delta},
\]  

(2.2)

where \( \langle \cdot \rangle \) denotes a sample average, and \( \sigma \) denotes the standard deviation. We first create a series of the average relative limit price and average volatility over 10 minute intervals. We then compute the cross-autocorrelation function and average over all stocks. The result is shown in figure (2.5).

We test the statistical significance of this result by testing against the null hypothesis that the volatility and relative limit price are uncorrelated. To do this we have to cope with the problem that the individual series are highly autocorrelated, as demonstrated in figure (2.4), and the 50 series for each stock also tend to be correlated to each other. To solve these problems, we construct samples of the null hypothesis using a technique introduced in Theiler, et al (1992). We compute the discrete Fourier transform of the relative limit price time series. We then randomly permute the phases of the series, and perform the inverse Fourier transform. This creates a realization of the null hypothesis, drawn from a distribution with the same unconditional distribution and the same autocorrelation function. Because we use the same random permutation of phases for each of the 50 series, we also preserve their correlation to each other. We then compute the cross autocorrelation function between each of the 50 surrogate limit price series and its corresponding true volatility series, and then average the results. We then repeat this experiment 300 times, which gives us a distribution of realizations of averaged sample cross-autocorrelation functions under the null
CHAPTER 2. THE POWER OF PATIENCE: A BEHAVIORAL REGULARITY IN LIMIT ORDER PLACEMENT

Figure 2.5: The cross autocorrelation of the time series of relative limit prices $\delta(t)$ and volatilities $v(t-\tau)$, averaged across all 50 stocks in the sample.

hypothesis. This procedure is more appropriate in this case than the standard moving block bootstrap, which requires choice of a timescale and will not work for a series such as this that does not have a characteristic timescale. The 2.5% and 97.5% quantile error bars at each lag are denoted by the two solid lines near zero in figure (2.5).

From this figure it is clear that there is indeed a strong contemporaneous correlation between volatility and relative limit price, and that the result is highly significant. Furthermore, there is some asymmetry in the cross-autocorrelation function; the peak occurs at a lag of one rather than zero, and there is more mass on the right than on the left. This suggests that there is some tendency for volatility to lead the relative limit price. This implies one of three things: (1) Volatility and limit price have a common cause, but this cause is for some reason felt later for the relative limit price; (2) the agents placing orders key off of volatility and correctly anticipate it; or, more plausibly, (3) volatility at least partially causes the relative limit price. Angel (1994) has suggested that volatility might affect limit order placement in this way.

Note that this suggests an interesting feedback loop: Holding other aspects of the order placement process constant, an increase in the average relative limit price will lower the depth in the limit or-
CHAPTER 2. THE POWER OF PATIENCE: A BEHAVIORAL REGULARITY IN LIMIT ORDER PLACEMENT

Order book at any particular price level, and therefore increase volatility. Since such a feedback loop is unstable, there are presumably nonlinear feedbacks of the opposite sign that eventually damp it. Nonetheless, such a feedback loop may potentially contribute to creating clustered volatility.

2.5 Conclusion

One of the most surprising aspects of the power law behavior of relative limit price is that traders place their orders so far away from the current price. As is evident in figure (2.2), orders occur with relative limit prices as large as 10,000 ticks (or 25 pounds for a stock with ticks in quarter pence). While we have taken some precautions to screen for errors, such as plotting the data and looking for unreasonable events, despite our best efforts, it is likely that there are still data errors remaining in this series. There appears to be a break in the merged unconditional distribution at about 2000 ticks; if this is statistically significant, it suggests that the very largest events may follow a different distribution than the rest of the sample, and might be dominated by data errors. Nonetheless, since we know that most of the smaller events are real, and since we see no break in the behavior until roughly $\delta \approx 2000$, errors are highly unlikely to be the cause of the power law behavior seen for $\delta < 2000$.

The conundrum of very large limit orders is compounded by consideration of the average waiting time for execution as a function of relative limit price. We intend to investigate the dependence of the waiting time on the limit price in the future, but since this requires tracking each limit order, the data analysis is more difficult. We have checked this for one stock, Vodafone, in which the largest relative limit price that resulted in an eventual trade was $\delta = 240$ ticks. Assuming other stocks behave similarly, this suggests that either traders are strongly over-optimistic about the probability of execution, or that the orders with large relative limit prices are placed for other reasons.

Since obtaining our results we have seen a recent preprint by Bouchaud et al. (2002) analyzing three stocks on the Paris Bourse over a period of a month. They also obtain a power law for $P(\delta)$, but they observe an exponent $\beta \approx 0.6$, in contrast to our value $\beta \approx 1.5$. We do not understand why there should be such a discrepancy in results. While they analyze only three stock-months of data, whereas we have analyzed roughly 1050 stock-months, their order arrival rates are roughly 20 times higher than ours, and their sample distributions appear to follow the power law scaling fairly well.

One possible explanation is the long-range correlation. Assum-
CHAPTER 2. THE POWER OF PATIENCE: A BEHAVIORAL REGULARITY IN LIMIT ORDER PLACEMENT

The Paris data show the same behavior we have observed, the decay in the autocorrelation is so slow that there may not be good convergence in a month, even with a large number of samples. The sample exponent $\beta$ based on one month samples may vary with time, even if the sample distributions appear to be well-converged. It is of course also possible that the French behave differently than the British, and that for some reason the French prefer to place orders much further from the midpoint.

Our original motivation for this work was to model price formation in the limit order book, as part of the research program for understanding the volatility and liquidity of markets outlined in Daniels, et al. (2001). $P(\delta)$ is important for price formation, since where limit orders are placed affects the depth of the limit order book and hence the diffusion rate of prices. The power law behavior observed here has important consequences for volatility and liquidity that will be described in a future paper.

Our results here are interesting for their own sake in terms of human psychology. They show how a striking regularity can emerge when human beings are confronted with a complicated decision problem. Why should the distribution of relative limit prices be a power law, and why should it decay with this particular exponent? Our results suggest that the volatility leads the relative limit price, indicating that traders probably use volatility as a signal when placing orders. This supports the obvious hypothesis that traders are reasonably aware of the volatility distribution when placing orders, an effect that may contribute to the phenomenon of clustered volatility. Plerou et al. (1999) have observed a power law for the unconditional distribution of price fluctuations. It seems that the power law for price fluctuations should be related to that of relative limit prices, but the precise nature and the cause of this relationship is not clear. The exponent for price fluctuations of individual companies reported by Plerou et al. is roughly 3, but the exponent we have measured here is roughly 1.5. Why these particular exponents? Makoto Nirei has suggested that if traders have power law utility functions, under the assumption that they optimize this utility, it is possible to derive an expression for $\beta$ in terms of the exponent of price fluctuations and the coefficient of risk aversion. However, this explanation is not fully satisfying, and more work is needed. At this point the underlying cause of the power law behavior of relative limit prices remains a mystery.
Chapter 3

The predictive power of zero intelligence in financial markets


3.1 Introduction

The traditional paradigm in economics is one of rational utility maximizing agents. Recognizing limitations in human cognition, economists have increasingly explored models in which agents have bounded rationality. We take this direction even further here by testing a model of trading in financial markets that drops agent rationality almost altogether. These results are particularly striking because the model predicts simple quantitative laws relating different properties of markets that are borne out well when tested against data.

While no one would dispute the fact that agents in financial markets behave strategically, and that for some purposes taking this into account is essential, we show in this paper that there are some problems where other factors may be more important. Previous work along these lines includes that of Becker (1962), who showed that
random agent behavior and a budget constraint is sufficient to guarantee the proper slope of supply and demand curves, and Gode and Sunder (1993), who demonstrated that if one replaces the students in a standard classroom economics experiment by zero-intelligence agents with a budget constraint, they perform surprisingly well. More specifically, the model we test here builds on earlier work on the double auction in financial economics (Mendelson, 1982; Cohen et al., 1985; Domowitz and Wang, 1994; Bollerslev et al., 1997) and physics (Bak et al., 1997; Eliezer and Kogan, 1998; Maslov, 2000; Slanina, 2001; Challet and Stinchcombe, 2001). (See also interesting subsequent work (Bouchaud et al., 2002, 2004)). The model makes the simple assumption that agents place orders to buy or sell at random (Daniels et al., 2003; Smith et al., 2003), subject to constraints imposed by current prices. While one might argue that tracking prices requires at least some intelligence, this is the minimal intelligence consistent with the assumptions of the model, which we will loosely refer to as “zero intelligence”. We show here that for certain problems such an approach can make surprisingly good quantitative predictions.

Another unusual aspect of the work presented here is the nature of the predictions that we test, which take the form of simple quantitative laws. These laws relate one set of market properties to another, placing restrictions on the allowed values of variables that are comparable to the ideal gas law of physics. They make quantitative predictions about magnitude and functional form, which are testable with only minimal auxiliary assumptions. This is in contrast to papers testing standard models based on rationality, which are typically forced to add strong auxiliary assumptions not contained in the original theoretical model, making the final results essentially qualitative. We present a brief review in the Supplementary Material (SM), Section 3.5.1.

### 3.1.1 Continuous double auction

The continuous double auction is the most widely used method of price formation in modern financial markets. The auction is called “double” because traders can submit orders both to buy and to sell and it is called “continuous” because they can do so at any time. Under the terminology we use here, an order that does not cross the opposite best price, and so does not result in an immediate transaction, is called a *limit order*. An example is a sell order with a higher price than any existing buy order. An order that does cross the opposite best price, and thus causes an immediate transaction, is called
Figure 3.1: A random process model of the continuous double auction. Stored limit orders are shown stacked along the price axis, with sell orders (supply) stacked above the axis at higher prices and buy orders (demand) stacked below the axis at lower prices. New sell limit orders are visualized as randomly falling down, and new buy orders as randomly “falling up”. New sell orders can be placed anywhere above the best buying price, and new buy orders anywhere below the best selling price. Limit orders can be removed spontaneously (e.g. because the agent changes her mind or the order expires) or they can be removed by market orders of the opposite type. This can result in changes in the best prices, which in turn alters the boundaries of the order placement process.

A market order\(^1\). Buy and sell limit orders accumulate in their respective queues, while buy and sell market orders cause transactions that remove limit orders. A limit order can also be removed from its queue by being cancelled, which can occur at any time. The lowest selling price offered at any point in time is called the best ask, \(a(t)\), and the highest buying price the best bid, \(b(t)\). The bid-ask spread \(s(t) \equiv a(t) - b(t)\) measures the gap between them. The best prices may change as new orders arrive or old orders are cancelled.

\(^1\)Real markets employ a host of different order types, which vary from market to market. However, by making appropriate decompositions (sometimes involving splitting an order into two pieces) it is always possible to break down the order flow into components that are effectively either market orders or limit orders.
3.1.2 Review of the model

The model that we test here (Daniels et al., 2003; Smith et al., 2003) was constructed to be the simplest possible sensible model of agent behavior in a continuous double auction. It assumes that two types of agents place orders randomly according to independent Poisson processes, as shown in Fig. 3.1. Impatient agents place market orders randomly with a Poisson rate of $\mu$ shares per unit time. Patient agents, in contrast, place limit orders randomly in both price and time. Buy limit orders are placed uniformly anywhere in the semi-infinite interval $-\infty < p < a(t)$, where $p$ is the logarithm of the price, and similarly sell limit orders are placed uniformly anywhere in $b(t) < p < \infty$. Both buying and selling limit orders arrive with same Poisson rate density $\alpha$ (measured in shares). Queued limit orders are cancelled according to a Poisson process, analogous to radioactive decay, with a fixed rate $\delta$ per unit time. To keep the model as simple as possible, there are equal rates for buying and selling, and all of these processes are independent except for indirect coupling through the boundary conditions, as explained below.

As new orders arrive they may alter the best prices $a(t)$ and $b(t)$, which in turn changes the boundary conditions for subsequent limit order placement. For example, the arrival of a buy limit order inside the spread will alter the best bid $b(t)$, which immediately alters the boundary condition for placing the next sell limit order. It is this feedback between order placement and price diffusion that makes this model interesting, and despite its apparent simplicity, very difficult to understand analytically. This model has been studied using simulation and with approximate analytic treatments based on mean field theory (Daniels et al., 2003; Smith et al., 2003).

Some readers may be puzzled by the use of a constant density over an infinite interval, which gives an infinite total arrival rate. The key is that the normalization is chosen to make the arrival rate in any given price interval finite. This is analogous to a model of snow falling and evaporating on an infinite plane: Though the total amount of snow arriving is infinite, the amount of snow falling in any given square during any given time is perfectly well-behaved. The situation here is much more complicated, due to the fact that market orders define a point removal process, and there are two kinds of “snow”, falling on overlapping and interacting intervals. Nonetheless, the basic trick of normalizing the density rather than the total is the same.
3.1.3 Predictions of the model

The rather radical assumption of a uniform limit order price density\(^2\) is made because it simplifies analysis, allowing the derivation of simple scaling laws relating the parameters to fundamental properties such as the average bid-ask spread. The mean value of the spread predicted based on a mean field theory analysis of the model (Daniels et al., 2003; Smith et al., 2003) is

\[
\hat{s} = (\mu/\alpha) f(\sigma \delta / \mu).
\]

(3.1)

The nondimensional ratio \(\epsilon \equiv \sigma \delta / \mu\) can be thought of as the ratio of removal by cancellation to removal by market orders, and plays an important role. \(f(\epsilon)\) is a slowly varying, monotonically increasing function that can be approximated (Smith et al., 2003) as \(f(\epsilon) = 0.28 + 1.86 \epsilon^{3/4}\). The scaling law above is reasonable in that it predicts that the spread increases when there are more market orders or cancellations (which remove stored limit orders), and decreases with more limit orders (which fill the spread in more quickly). The dependence on \(\mu/\alpha\) can be derived from dimensional analysis, under certain assumptions detailed in SM Section 3.5.2. However, the functional form of \(f(\epsilon)\) is not obvious. One of the predictions of the model, that to our knowledge has not been hypothesized elsewhere in the literature, is that the order size \(\sigma\) is an important determinant of the spread.

Another prediction of the model concerns the price diffusion rate, which drives the volatility of prices and is the primary determinant of financial risk. If we assume that prices make a random walk, then the diffusion rate measures the size and frequency of its increments. The variance \(V\) of a random walk grows as \(V(t) = Dt\), where \(D\) is the diffusion rate and \(t\) is time. This is the main free parameter in the Bachelier model of prices (Bachelier, 1964). While its value is essential for risk estimation and derivative pricing there is very little fundamental understanding of what actually determines it. In standard models it is often assumed to depend on “information arrival” (Clark, 1973), which has the disadvantage that it is impossible to measure directly. For our idealized model, numerical experiments indicate that the short term price diffusion rate is to a very good approximation given by the simple formula (Daniels et al., 2003; Smith et al., 2003)

\[
\hat{D} = k \mu^{5/2} \delta^{1/2} \sigma^{-1/2} \alpha^{-2},
\]

(3.2)

where \(k\) is a constant. This formula is reasonable in that it predicts that volatility increases with limit order removal (either by market

\(^2\)For an empirical investigation of the density of limit order placement see (Zovko and Farmer, 2002; Bouchaud et al., 2002)
orders or by cancellations) and decreases with limit order placement. The dependence on order size and the values of the scaling exponents are not so obvious. It has so far not been possible to derive this formula from theoretical considerations (though dimensional analysis was essential for guessing this functional form).

We would like to emphasize that the construction of the model and all the predictions derived from it were made prior to looking at the data. The model was constructed to be simple enough to be analytically tractable, and makes many strong assumptions. The assumption of random order placement leads to consequences that might be economically unreasonable in a rational setting, such as the existence of profit making opportunities. However, this is self-consistent with the assumption that the only intelligence the agents possess is the ability to mechanically adjust the prices of limit orders based on current best prices. Furthermore, simulations suggest that the arbitrage opportunities in this model are not risk-free, yielding only finite, risky profits.

A useful concept is that of liquidity, which in this context can be defined as the availability of standing limit orders that allow trading to take place. The impatient market order traders are liquidity demanders, and the limit order traders are liquidity providers. The use of a zero-intelligence agent model makes it possible to study the flow of liquidity in and out of the market, and to study its interaction with price formation. This has not been properly addressed by models that attempt to fully treat agent rationality. Abandoning the assumption of rationality gives the ability to focus modeling effort on other problems, such as those addressed here.

## 3.2 Testing the scaling laws

### 3.2.1 Data

We test this model with data from the electronic open limit order book of the London Stock Exchange (SETS), which includes about half of the total trading volume. We used data from eleven stocks in the period from August 1st 1998 to April 30th 2000, which includes 434 trading days and a total of roughly six million events. For all these stocks the number of total events exceeds 300,000 and was never less than 80 on any given day (where an event corresponds to an order placement or cancellation). Orders placed during the opening auction are removed to accomodate the fact that the model only applies for the continuous auction. See SM Section 3.5.3 for more details.

---

3. J. Rutt, J.D. Farmer, and J. Girard, work in progress.
3.2.2 Testing procedure

We test the model cross-sectionally over eleven stocks. For each stock we measure its average order flow rates and calculate the predicted average spread $\hat{\sigma}$ and diffusion rate $\hat{D}$ for that stock using equations 3.1 and 3.2. We then compare these predicted values to the actual values of the spread $\bar{\sigma}$ and diffusion rate $\bar{D}$ which we again measure from the data. The comparison is done via linear regressions of the predicted values against the actual, measured values. For a discussion see SM Sections 3.5.5 and 3.5.6.

Measurement of the parameters $\mu$ and $\sigma$ is straightforward: To measure $\mu$, for example, we simply compute the total number of shares of market orders and divide by time, or alternatively, we compute $\mu_t$ for each day and average; we get similar results in either case. However, a problem occurs in measuring the parameters $\alpha$ and $\delta$ due to the simplifying assumption of a uniform distribution of prices for limit order placement and a uniform cancellation rate. In the real data limit order placement and cancellation are concentrated near the best prices (Bouchaud et al., 2002; Zovko and Farmer, 2002). In order to cope with this we make an auxiliary assumption that order placement is uniform inside a price window around the best prices, and zero outside this window. We choose this price window $W$ to correspond to roughly 60% of limit orders away from the mid-price, and compute $\alpha$ by dividing the number of shares of limit orders placed inside this price window per unit time by $W$. We do this for each day and compute the average value of $\alpha$ for each stock. We similarly compute $\delta$ as the inverse of the average lifetime of orders cancelled inside the same price window $W$. See SM Section 3.5.5 for details.

The laws that we describe here do not make temporal predictions, but rather are restrictions of state variables. The ideal gas law, $PV = RT$, provides a good analogy. It predicts that pressure $P$, volume $V$, and temperature $T$ are constrained – any two of them determines the third. The gas constant $R$ is the only free parameter. In very much the same way, we are testing two relations between properties of order flows and properties of prices. We are not attempting to predict the temporal behavior of the order flows, only trying to see whether the restrictions between order flows and prices predicted by the model are valid. It is important to emphasize that while $\mu$, $\alpha$, $\delta$, and $\sigma$ can be viewed as free parameters of the model, they are not free parameters in the test of the model. Rather they are now variables, like $P$, $V$, and $T$ in the ideal gas law. The only free parameter is the price window $W$. We chose $W = 0.6$ as a prior; it turns out that it is also roughly the value that maximizes the goodness of fit, however, varying $W$ does not change the goodness of fit substantially.
CHAPTER 3. THE PREDICTIVE POWER OF ZERO INTELLIGENCE IN FINANCIAL MARKETS

3.2.3 Spread

To test Equation 3.1, we measure the average spread $\bar{s}$ across the full time period for each stock, and compare to the predicted average spread $\hat{s}$ based on order flows. Spread is measured as the average of $\log b(t) - \log a(t)$ (recall that in the model $p$ represents log price). The spread is measured after each event, with each event given equal weight. The opening auction is excluded.

To test our hypothesis that the predicted and actual values coincide, we perform a regression of the form $\log \bar{s} = A \log \hat{s} + B$. We took logarithms to do the regression because the spread is positive and the log of the spread is approximately normally distributed\(^4\). We use $A$ and $B$ for hypothesis testing. Based on the model we predict that the comparison should yield a straight line with $A = 1$ and $B = 0$. However, because of the degree of freedom in choosing the price interval $W$ as described above, the value of $B$ is somewhat arbitrary; varying $W$ through reasonable values changes $B$ significantly, with much less effect on $A$.

The least squares regression, shown together with the data comparing the predictions to the actual values in Fig. 3.2, gives $A = 0.99 \pm 0.10$ and $B = 0.06 \pm 0.29$. We thus strongly reject the null hypothesis that $A = 0$, indicating that the predictions are far better than random. More importantly, we are unable to reject the null hypothesis that $A = 1$. The regression has $R^2 = 0.96$, so the model explains most of the variance. Note that because of long-memory effects and cross-correlations between stocks the errors in the regression are larger than they would be for IID data (see SM Section 3.5.6).

3.2.4 Price diffusion rate

As for the spread, we compare the predicted price diffusion rate based on order flows to the actual price diffusion rate $\bar{D}$ for each stock averaged over the 21 month period, and regress the logarithm of the predicted vs. actual values, as shown in Fig. 3.3. See SM Section 3.5.5 for details on the diffusion rate and its estimation.

The regression gives $A = 1.33 \pm 0.25$ and $B = 2.43 \pm 1.75$. Thus, we again strongly reject the null hypothesis that $A = 0$. We are still unable to reject the null hypothesis that $A = 1$ with 95% confidence, though there is some suggestion that the real values increase faster than the predicted values. In any case, the predictions are at least a good approximation. Although the results are not as good as for the

\(^4\) An alternative would have been to take logarithms of each event and average the logarithms. We instead regard this as a test of the cross-sectional averages, and take logarithms of the cross-sectional values.
spread, $R^2 = 0.76$, so the model still explains most of the variance.

### 3.3 Average market impact

Market impact is practically important because it is the dominant source of transaction costs for large trades, and conceptually important because it provides a convenient probe of the revealed supply and demand functions in the limit order book (see SM Section 3.5.7). When a market order of size $\omega$ arrives, if it removes all limit orders at the best bid or ask it will immediately change the midpoint price $m \equiv (a + b)/2$. We define the average market impact function $\phi$ in terms of the instantaneous logarithmic midpoint price shift $\Delta p$ conditioned on order size, $\phi(\omega) = E[\Delta p | \omega]$. $\Delta p$ is the difference between the price just before a market order arrives and the price just after it arrives (before any other events).

A long-standing mystery about market impact is that it is a
highly concave function of $\omega$ (Hausman et al., 1992; Farmer, 1996; Torre, 1997; Kempf and Korn, 1999; Plerou et al., 2002a; Bouchaud et al., 2002; Lillo et al., 2003; Gabaix et al., 2003a). This is unexpected since simple arguments would suggest that because of the multiplicative nature of returns, market impact should grow at least linearly (Smith et al., 2003). We know of no model that explains this. The model we are testing here predicts a concave average market impact function, with the concavity becoming more pronounced for small values of $\epsilon = \sigma \delta / \mu$. Intuitively, the concavity is due to the fact that limit orders near the best price are removed by transactions more rapidly than those far from the best price. As a result the average density of stored limit orders in the book increases moving away from the midpoint. An increase in density of limit orders implies a decreased price response to a market order of given size, resulting in a concave market impact function.

While the predictions of the model are qualitatively correct, from

Figure 3.3: Regressions of predicted values based on order flow using equation 3.2 vs. actual values for the logarithm of the price diffusion rate. The dots show the average predicted and actual value for each stock averaged over the full 21 month time period. The solid line is a regression; the dashed line is the diagonal, with $A = 1$ and $B = 0$. 
a quantitative point of view the model predicts a larger variation with \( \epsilon \) than what we actually observe. Nonetheless, the model is still quantitatively useful for understanding market impact, as described below.

A surprising regularity of the average market impact function is uncovered by simply plotting the data in the non-dimensional coordinates dictated by the model, as shown in Fig. 3.4. If we view market impact in standard dimensional units, such as British Pounds or shares, there is large variability from stock to stock; the story becomes much simpler in non-dimensional units.

When we plot the average market impact in standard dimensional coordinates, the behavior is highly variable from stock to stock. For example, in Fig. 3.4(b) we plot the average market impact \( \phi(\omega) = E[\Delta p|\omega] \) as a function of the order size \( \omega \) in units of British Pounds. We do this by binning together events with similar \( \omega \) and plotting this vs. the corresponding mean price impact \( \Delta p \) for each bin. The result varies widely from stock to stock. We have explored a variety of other ways for renormalizing the order size, as described in SM Section 3.5.7, but they all give similar results.

Plotting the data in non-dimensional units tells a simpler story. To do this we normalize the price shift and order size by appropriate dimensional scale factors based on the daily order flow rates. For the derivation of the non-dimensional coordinates used to do this see SM Section 3.5.2. This transforms the standard coordinates to non-dimensional coordinates as\(^5\) \( \Delta p \rightarrow \Delta p \cdot \alpha_t / \mu_t \) and \( \omega \rightarrow \omega \cdot \delta_t / \mu_t \), where \( \alpha_t, \mu_t, \) and \( \delta_t \) are the average parameters for day \( t \). The data collapses onto roughly a single curve, as shown in Fig. 3.4(a). The variations from stock to stock are quite small; on average the corresponding bins for each stock deviate from each other by about 8%, roughly the size of the statistical sampling error. We do not find that this variation is statistically significant, through we should also say that such tests are complicated by the long-memory property of these time series and cross correlations between stocks, so that we do not consider the results fully reliable (see SM Section 3.5.6). In contrast, using standard dimensional coordinates the differences are easily shown to be highly statistically significant. This collapse illustrates that the non-dimensional coordinates dictated by the model provide substantial explanatory power: We can understand how the average market impact varies from stock to stock by a simple transformation of coordinates. Plotting in double logarithmic scale shows

\[ [\Delta p \cdot \alpha / \mu] = \text{price} \cdot \frac{\text{shares}/(\text{price} \cdot \text{time})}{\text{shares/time}} \tag{3.3} \]

is without dimensions, i.e. the units cancel out.

---

\(^5\)Notice that \( \Delta p \) has units of price but
that the curve of the collapse is roughly a power law of the form \( \omega^{0.25} \) (see SM Section 3.5.7). This provides a more fundamental explanation for the empirically constructed collapse of average market impact for the New York Stock Exchange found earlier (Lillo et al., 2003).
Figure 3.4: The average market impact as a function of the mean order size. In (a) the price differences and order sizes for each transaction are normalized by the non-dimensional coordinates dictated by the model, computed on a daily basis. Most of the stocks collapse onto a single curve; there are a few that deviate, but the deviations are sufficiently small that given the long-memory nature of the data and the cross-correlations between stocks, it is difficult to determine whether these deviations are statistically significant. This means that we understand the behavior of the market impact as it varies from stock to stock by a simple transformation of coordinates. In (b), for comparison we plot the order size in units of British pounds against the average logarithmic price shift.
3.4 Conclusions

We have shown that the model we have presented here does a good job of predicting the average spread, and a decent job of predicting the price diffusion rate. Also, by simply plotting the data in non-dimensional coordinates we get a better understanding of the regularities of market impact. These results are remarkable because the underlying model largely drops agent rationality, instead focusing all its attention on the problem of understanding the constraints imposed by the continuous double auction.

It is worth comparing our results to those of previous empirical work. For example, Hasbrouck and Saar (Hasbrouck and Saar, 2002) find a positive correlation between volatility and the ratio of market orders to limit orders. They perform regressions of this ratio against volatility and several other dependent variables, and obtain goodness of fits similar to ours (except that in their case volatility was one of several independent variables, whereas in our case it was the dependent variable). They then discuss the results in terms of their consistency with three effects that one would expect from agent rationality. For example, one such effect is called “market order certainty”: When prices are more volatile, market orders become more attractive to a risk-averse rational agent, and so the fraction of market orders should increase. The observed positive correlations are consistent with this.

The model we test here offers an alternative explanation that does not depend on strategic choice. We also predict a positive correlation between volatility and the fraction of market orders (see equation 3.2) but for a different reason: An increase in the rate of market order submission reduces liquidity and thus increases price volatility. We certainly believe that agents respond in important ways to changing market conditions such as volatility, and indeed we have demonstrated this in previous work (Zovko and Farmer, 2002). Nonetheless, we argue that it is also necessary to understand the impact of agents’ actions on market conditions. By carefully treating the feedback in both directions between price formation and limit order pricing under minimal assumptions of rationality, this model provides a null hypothesis against which claims of rational behavior can be measured.

An important feature of the model we test here is its parsimony and falsifiability. Our model makes simultaneous quantitative predictions about volatility, spread, and market impact. We postulate specific functional forms for the relation between order flows and spread and volatility; while there are multiple variables involved, there is only one free parameter. Rationality-based theories, in contrast, rarely make predictions about magnitude or functional form,
and as a result their predictions are harder to test. Such tests generally require stronger auxiliary assumptions, such as imposed functional forms with multiple free parameters. Empirical studies that test such models often test only the sign of such effects, which often have a variety of alternative explanations. Our model makes sharper predictions, and is consequently more testable (Ziliak and McCloskey, 2004).

The approach taken here succeeds in part because it is less ambitious than a standard rationality-based model. This can be viewed as a divide and conquer strategy. Rather than attempting to explain the properties of the market from fundamental assumptions about utility maximization by individual agents, we divide the problem into two parts. The first (easier) problem, addressed here, is that of understanding the characteristics of the market given the order flows. The second (harder) problem, which remains to be investigated, is that of explaining why order flow varies as it does. Explaining variations in order flow involves behavioral and/or strategic issues that are likely to be much more difficult to understand. It is always desirable to solve easier problems first.

The model succeeds in part by reducing the problem to the measurement of the right variables. By measuring the rate of market order placement vs. limit order placement, and the rate of order cancellation, we are able to measure how patient or impatient traders are. The model makes quantitative predictions about how this affects other market properties. The agreement with the model indicates that patience is an important determinant of market behavior. Variations in patience might be explained by a rationality-based explanation in terms of information arrival, or a behavioral-based explanation driven by emotional response, but in either case it suggests that patience is a key factor.

These results have several practical implications. For market practitioners, understanding the spread and the market impact function is very useful for estimating transaction costs and for developing algorithms that minimize their effect. For regulators they suggest that it may be possible to make prices less volatile and lower transaction costs, if this is desired, by creating incentives for limit orders and disincentives for market orders. These scaling laws might also be used to detect anomalies, e.g. a higher than expected spread might be due to improper market maker behavior.

This is part of a broader research program that might be somewhat humorously characterized as the “low-intelligence” approach: We begin with minimally intelligent agents to get a good benchmark of the effect of market institutions, and once this benchmark is well-understood, add more intelligence, moving toward market efficiency. We thus start from almost zero rationality and work our way up, in
contrast to the canonical approach of starting from perfect rationality and working down.

The model we test here was constructed before looking at the data (Daniels et al., 2003; Smith et al., 2003), and was designed to be as simple as possible for analytic analysis. A more realistic (but necessarily more complicated) model would more closely mimic the properties of real order flows, which are price dependent and strongly correlated both in time and across price levels, or might incorporate elements of the strategic interactions of agents. An improved model would hopefully be able to capture more features of the data than those we have studied here. We know there are ways in which the current model is inappropriate, e.g., it allows arbitrage opportunities that do not exist in the real market. Nonetheless, as we have shown above, this extremely simple model does a good job of explaining some important properties of markets. For further discussion see SM Section 3.5.8.

How is it conceivable to successfully model a situation in which we know that agents engage in clever strategic behavior in terms of a model that completely neglects this? Perhaps a telephone exchange provides a good analogy: Even though each customer has a perfectly good reason for picking up the phone, communications engineers design exchanges by assuming they do so at random. Similarly, there are situations in markets where rational behavior can be treated in aggregate as though it were noise. The question is whether rational effects are more important or less important than stochastic effects. Rational effects are clearly important in determining overall price levels, but they may be dominated by random fluctuations in determining volatility. We do not mean to claim that market participants are unintelligent: Indeed, one of the virtues of this model is that it provides a benchmark to separate properties that are driven by the statistical mechanics of the market institution from those that are driven by the strategic behavior of agents. It suggests that institutions strongly shape our behavior, so that some of the properties of markets may depend more on the structure of institutions than on the rationality of individuals.
3.5 Supplementary Material

3.5.1 Literature review

We will begin by reviewing what we will call the "standard literature", which includes both purely empirical studies, as well as theoretical studies predicated on rationality models. We will then review the literature on random process models of the continuous double auction, which is more closely related to the model we test here.

Standard literature

The market microstructure literature focusing on the understanding of spread, volatility and market impact in financial markets is theoretically and empirically extensive. The theoretical analyses traditionally use the underlying paradigm of rational agents. Models of spread, starting with Demsetz (1968); Tinic (1972); Stoll (1978); Amihud and Mendelson (1980); Ho and Stoll (1981), have examined the possible determinants of spreads as a result of rational, utility-maximizing problem faced by the market makers. Models providing insight into the utility-maximizing response of agents to other various measures of market conditions such as volatility are for example Lo (2002) who investigate a simple model in which the log stock price is modeled as a Brownian motion diffusion process. Provided agents prefer a lower expected execution time, their model predicts a positive relationship between volatility and limit order placement. Copeland and Galai (1983); Glosten and Milgrom (1985); Easley and O’Hara (1987); Glosten (1995); Foucault (1999); Easley et al. (2001) examine asymmetric information effects on order placement. Andersen (1996) modifies the Glosten and Milgrom (1985) model with the stochastic volatility and information flow perspective. Other models of trading in limit order markets include Cohen et al. (1981); Angel (1994); Harris (1998); Chakravarty and Holden (1995); Seppi (1997); Rock (1990); Parlour and Seppi (2003); Parlour (1998); Foucault et al. (2001); Domowitz and Wang (1994).


Empirical research in volatility was initiated by statistical descriptions of the volatility process (Engle (1982); Bollerslev (1986), Bollerslev et al. (1992) for a survey), but has grown increasingly ambitious with multivariate structural models of the interaction between volatility and other economic variables. Positive correlation
of daily volume and volatility has been documented in Clark (1973); Epps and Epps (1976); Tauchen and Pitts (1983). Volume has been entered into ARCH specifications by Lamoureux and Lastrapes (1994). Other empirical investigations of volatility determinants and consequences include Gallant et al. (1992); Andersen (1996); Blume et al. (1991); Reiss and Werner (1994); Fleming et al. (2001); Hasbrouck (2003). A comprehensive study of the joint distribution of returns and volume is done by Gallant et al. (1992, 1993).

Random process models of continuous double auction

There are two independent lines of prior work, one in the financial economics literature, and the other in the physics literature. The models in the economics literature are directed toward econometrics and treat the order process as static. In contrast, the models in the physics literature are mostly conceptual toy models, but they allow the order process to react to changes in prices, and are thus fully dynamic. Our model bridges this gap. This is explained in more detail below.

The first model of this type that we are aware of in the economics literature was due to Mendelson (1982), who modeled random order placement with periodic clearing. Cohen et al. (1985) developed a model of a continuous auction, modeling limit orders, market orders, and order cancellation as Poisson processes. However, they only allowed limit orders at two fixed prices, buy orders at the best bid, and sell orders at the best ask. This assumption allowed them to use standard results from queuing theory to compute properties such as the expected number of stored limit orders, the expected time to execution, and the relative probability of execution vs. cancellation. Domowitz and Wang (1994) extended this to multiple price levels by assuming arbitrary order placement and cancellation processes (which can take on any value at each price level). They assume that these processes are fixed in time, and do not respond to changes in the best bid or ask. This allows them to derive the distribution of the spread, transaction prices, and waiting times for execution. This model was tested by Bollerslev et al. (1997) on three weeks of data for the Deutschemark/U.S. Dollar exchange rate. They showed that it does a good job of predicting the distribution of the spread. However, since the prices are pinned, the model does not make a prediction about price diffusion, and this also creates errors in the predictions of the spread and stored supply and demand.

The models in the physics literature, which appear to have been developed independently, differ in that they address price dynamics. That is, they incorporate the feedback between order placement and price formation, allowing the order placement process to change
in response to changes in prices. These models have mainly been conceptual toy models designed to understand the anomalous diffusion properties of prices (a property that all of these models fail to reproduce, as explained later). This line of work begins with a paper by Bak et al. (1997) which was developed by Eliezer and Kogan (1998) and by Tang and Tian (1999). They assume that limit orders are placed at a fixed distance from the midpoint, and that the limit prices of these orders are then randomly shuffled until they result in transactions. It is the random shuffling that causes price diffusion. This assumption, which we feel is unrealistic, was made to take advantage of the analogy to a standard reaction-diffusion model in the physics literature. Maslov (2000) introduced an alternative model that was solved analytically in the mean-field limit by Slanina (2001). Each order is randomly chosen to be either a buy or a sell with equal probability, and either a limit order or a market order with equal probability. If a limit order, it is randomly placed within a fixed distance of the current price. Both the Bak et al. model and that of Maslov result in anomalous price diffusion, in the sense that the Hurst exponent $H = 1/4$ (in contrast to standard diffusion, which has $H = 1/2$, or real prices which tend to have $H > 1/2$). In addition, the Maslov model unrealistically requires equal probabilities for limit and market order placement, otherwise the inventory of stored limit orders either goes to zero or grows without bound. A model adding a Poisson order cancellation process was proposed by Challet and Stinchcombe (2001), and independently by Daniels et al. (2003). Challet and Stinchcombe showed that this results in $H = 1/4$ for short times, but asymptotically gives $H = 1/2$. The Challet and Stinchcombe model, which posits an arbitrary, unspecified function for the relative position of limit order placement, is quite similar to that of Domowitz and Wang (1994), but allows for the possibility of order placement responding to price movement.

The model we test here was introduced by Daniels et al. (2003). Like other physics models, it treats the feedback between order placement and price movement. It has the advantage that it is defined in terms of five scalar parameters, and so is parsimonious and can easily be tested against real data. Its simplicity enables a dimensional analysis, which gives approximate predictions about many of the properties of the model. Perhaps most important is the use to which the model is put: With the exception of reference (Eliezer and Kogan, 1998), work in the physics literature has focused almost entirely on the anomalous diffusion of prices. While interesting and important for refining risk calculations, from a practical point of view this is a second-order effect. In contrast, the model studied here focuses on first order effects of primary interest to market participants, such as the bid-ask spread, volatility, depth profile, price impact, and the
probability and time to fill an order. It demonstrates how dimensional analysis becomes a useful tool in an economic setting, and the analysis done in Daniels et al. and Smith et al. develops mean field theories to understand many relevant market properties. Many of the important properties of the model can be stated in terms of simple scaling relations in terms of the five parameters.

Subsequent to reference Daniels et al. (2003), Bouchaud et al. (2002) demonstrated that they can derive a simple equation for the depth profile, by making the assumption that prices execute a random walk and introducing an additional free parameter. In this paper we show how to do this from first principles without introducing a free parameter. Chiarella and Iori (2002) have numerically studied fundamentalists and technical traders placing limit orders; a talk on this work by Giulia Iori in part inspired this model.

3.5.2 Dimensional analysis

Dimensional analysis can be used to simplify the study of this model and to make some approximate predictions about several of its properties. For a good reference on dimensional analysis see Barenblatt (1987).

There are three fundamental dimensional quantities in this model: shares, price, and time. There are five parameters. Because they have independent dimensions, when the dimensional constraints between the parameters are taken into account, this leaves only two independent degrees of freedom. It turns out that the order flow rates \( \mu, \alpha, \) and \( \delta \) are more important than the discreteness parameters \( \sigma \) and \( dp \), in the sense that the properties of the model are much more sensitive to variations in the order flow rates than they are to variations in \( \sigma \) or \( dp \). It therefore natural to construct non-dimensional units based on the order flow parameters alone. There are unique combinations of the three order flow rates with units of shares, price, and time. This gives characteristic scales for price, shares, and time, that are unique up to a constant: the characteristic number of shares \( N_c = \mu/\delta \), the characteristic price interval \( p_c = \mu/\alpha \), and the characteristic timescale \( t_c = 1/\delta \). These are the unique combinations of these three parameters with the correct dimensions.

These characteristic scales can be used to define non-dimensional coordinates based on the order flow rates. These are \( \hat{p} = p/p_c \) for price, \( \hat{N} = N/N_c \) for shares, and \( \hat{t} = t/t_c \) for time. The use of non-dimensional coordinates has the great advantage that it reduces the number of degrees of freedom from five to two. That is, instead of five independent parameters, we only have two independent parameters, which makes the model easier to understand.

The two irreducible degrees of freedom are naturally discussed in
terms of non-dimensional versions of the discreteness parameters. A non-dimensional scale parameter based on order size is constructed by dividing the typical order size $\sigma$ (with dimensions of shares) by the characteristic number of shares $N_c$. This gives the non-dimensional parameter $\epsilon \equiv \sigma/N_c = \delta \sigma/\mu$, which characterizes the granularity of the order flow. A non-dimensional scale parameter based on tick size is constructed by dividing the tick size $dp$ by the characteristic price, i.e. $dp/p_c = \alpha dp/\mu$. The usefulness of this is that the properties of the model only depend on the two non-dimensional parameters, $\epsilon$ and $dp/p_c$: Any variations of the parameters $\mu$, $\alpha$, and $\delta$ that keep these two non-dimensional parameters constant gives exactly the same market properties. One of the interesting results that emerges from analysis of the model is that the effect of the granularity parameter $\epsilon$ is generally much more important than the tick size $dp/p_c$. For a more detailed discussion, see reference Smith et al. (2003).

While we have investigated numerically the effect of varying the tick size in ref. Smith et al. (2003), for the purposes of comparing to data, here we simply take the limit $dp \to 0$, which provides a reasonable approximation.

### 3.5.3 The London Stock Exchange (LSE) data set

The London Stock Exchange is composed of two parts, the electronic open limit order book, and the upstairs quotation market, which is used to facilitate large block trades. During the time period of our dataset 40% to 50% of total volume was routed through the electronic order book and the rest through the upstairs market. It is believed that the limit order book is the dominant price formation mechanism of the London Stock Exchange: about 75% of upstairs trades happen between the current best prices in the order book (LSEbulletin, 2001). Our analysis involves only the data from the electronic order book. We chose to study this data set because we have a complete record of every action taken by every participating institution, allowing us to measure the order flows and cancellations and estimate all of the necessary parameters of our model.

We used data from the time period August 1st 1998 - April 30th 2000, which includes a total of 434 trading days and roughly six million events. We chose 11 stocks each having the property that the number of total number of events exceeds 300,000 and was never less than 80 on any given day. Some statistics about the order flow for each stock are given in table 3.1.

The trading day of the LSE starts at 7:50 with a roughly 10 minute long opening auction period (during the later part of the dataset the auction end time varies randomly by 30 seconds). Dur-
CHAPTER 3. THE PREDICTIVE POWER OF ZERO INTELLIGENCE IN FINANCIAL MARKETS

<table>
<thead>
<tr>
<th>stock ticker</th>
<th>num. events (1000s)</th>
<th>average (per day)</th>
<th>limit (1000s)</th>
<th>market (1000s)</th>
<th>deletions (1000s)</th>
<th>eff. limit (shares)</th>
<th>eff. market (shares)</th>
<th># days</th>
</tr>
</thead>
<tbody>
<tr>
<td>AZN</td>
<td>608</td>
<td>1405</td>
<td>128</td>
<td>188</td>
<td>4,921</td>
<td>4,967</td>
<td>292</td>
<td>429</td>
</tr>
<tr>
<td>BARC</td>
<td>571</td>
<td>1318</td>
<td>128</td>
<td>172</td>
<td>6,406</td>
<td>7,370</td>
<td>271</td>
<td>433</td>
</tr>
<tr>
<td>CW.</td>
<td>511</td>
<td>1184</td>
<td>134</td>
<td>134</td>
<td>11,151</td>
<td>12,671</td>
<td>300</td>
<td>434</td>
</tr>
<tr>
<td>GLXO</td>
<td>814</td>
<td>1885</td>
<td>225</td>
<td>225</td>
<td>8,927</td>
<td>13,846</td>
<td>200</td>
<td>434</td>
</tr>
<tr>
<td>LLOY</td>
<td>644</td>
<td>1485</td>
<td>159</td>
<td>184</td>
<td>11,376</td>
<td>12,097</td>
<td>184</td>
<td>432</td>
</tr>
<tr>
<td>PRU</td>
<td>314</td>
<td>884</td>
<td>57</td>
<td>104</td>
<td>9,502</td>
<td>12,097</td>
<td>94</td>
<td>432</td>
</tr>
<tr>
<td>RTR</td>
<td>408</td>
<td>978</td>
<td>94</td>
<td>195</td>
<td>8,597</td>
<td>16,483</td>
<td>195</td>
<td>431</td>
</tr>
<tr>
<td>SB.</td>
<td>665</td>
<td>1526</td>
<td>128</td>
<td>112</td>
<td>9,965</td>
<td>16,483</td>
<td>152</td>
<td>426</td>
</tr>
<tr>
<td>SHEL</td>
<td>592</td>
<td>1367</td>
<td>156</td>
<td>170</td>
<td>13,589</td>
<td>16,483</td>
<td>156</td>
<td>429</td>
</tr>
<tr>
<td>VOD</td>
<td>940</td>
<td>2161</td>
<td>296</td>
<td>207</td>
<td>30,133</td>
<td>38,550</td>
<td>296</td>
<td>434</td>
</tr>
</tbody>
</table>

Table 3.1: Summary statistics for stocks in the dataset. Fields from left to right: stock ticker symbol, total number of events (effective market orders + effective limit orders + order cancellations) in thousands, average number of events in a trading day, number of effective limit orders in thousands, number of effective market orders in thousands, number of order deletions in thousands, average limit order size in shares, average market order size in shares, number of trading days in the sample.
ing this time orders accumulate without transactions; then a clearing price for the opening auction is calculated, and all opening transactions take place at this price. Following the opening at 8:00 the market runs continuously, with orders matched according to price and time priority, until the market closes at 16:30. In the earlier part of the dataset, until September 22nd 1999, the market opening hour was 9:00. During the period we study there have been some minor modifications of the opening auction mechanism, but since we discard the opening auction data anyway this is not relevant.

Some stocks in our sample (VOD for example) have had stock price splits and tick price changes during the period of our sample. We take splits into account by transforming stock sizes and prices to pre-split values. In any case, since all measured quantities are in logarithmic units, of the form $\log(p_1) - \log(p_2)$, the absolute price scale drops out. Our theory predicts that the tick size should change some of the quantities of interest, such as the bid-ask spread, but the predicted changes are small enough in comparison with the effect of other parameters that we simply ignore them (and base our predictions on the limit where the tick size is zero). Since granularity is much more important than tick size, this seems to be a good approximation.

3.5.4 Opening auction, real order types, time

Since the model does not take the opening auction into account, we simply neglect orders leading up to the opening auction, and base all our measurements on the remaining part of the trading day, when the auction is continuous.

In order to treat simply and in a unified manner the diverse types of orders traders can submit in a real market (for example, crossing limit orders, market orders with limiting price, ‘fill-or-kill, execute & eliminate) we use redefinitions based on whether an order results in an immediate transaction, in which case we call it an effective market order, or whether it leaves a limit order sitting in the book, in which case we call it an effective limit order. Marketable limit orders (also called crossing limit orders) are limit orders that cross the opposing best price, and so result in at least a partial transaction. The portion of the order that results in an immediate transaction is counted as an effective market order, while the non-transacted part (if any) is counted as an effective limit order. Orders that do not result in a transaction and do not leave a limit order in the book, such as for example, failed fill-or-kill orders, are ignored altogether. These have no effect on prices, and in any case, make up only a very small fraction of the order flow, typically less than 1%. Note that we drop the term “effective”, so that e.g. “market order” means “effective market order”.

43
A limit order can be removed from the book for many reasons, e.g. because the agent changes her mind, because a time specified when the order was placed has been reached, or because of the institutionally-mandated 30 day limit on order duration. We will lump all of these together, and simply refer to them as “cancellations”.

Our measure of time is based on the number of events, i.e., the time elapsed during a given period is just the total number of events, including effective market order placements, effective limit order placements, and cancellations. We call this event time. Price intervals are computed as the difference in the logarithm of prices, which is consistent with the model, in which all prices are assumed to be logarithmic in order to assure their positiveness.

3.5.5 Measurement of model parameters

We test the predictions of the model against real data cross-sectionally on eleven stocks. The parameters of the model are stated in terms of order arrival rates, cancellation rate, order size, and our dataset allows us to compute for each stock the average values of these parameters. As we explain here, these average rates are calculated as means of the daily values weighted by the daily number of events. An alternative would have been to calculate the mean values of the parameters over the entire 2 year period for each stock. While this works well for the parameters $\mu$ and $\sigma$, it does not work as well for $\alpha$ and $\delta$, as explained below.

Measuring $\mu$ and $\sigma$

The parameter $\mu_t$, which characterizes the average market order arrival rate on day $t$ is just the ratio of the number of shares of effective market orders (for both buy and sell orders) to the number of events during the trading day. Thus for $\mu$ it makes no difference whether we measure it across the whole period, or take a weighted average of daily values. This is also true for the average order size $\sigma_t$. One complication in measuring $\sigma$ is that the model assumes that the average size for limit orders and market orders is the same, whereas for the real data this is not strictly true. Nonetheless, as seen in Table 3.1, although the limit order size tends to be a bit larger than the market order size, it is still a fairly good approximation to take them to be the same. For the purposes of this analysis we use the limit order size to measure order size, based on theoretical arguments that this is more important than the market order size. In any case, this does not make a significant difference in the results.
Measuring $\alpha$ and $\delta$

The measuring of the cancellation rate $\delta_t$ and the limit order rate density $\alpha_t$ is more complicated, due to the highly simplified assumptions we made for the model. In contrast to our assumption of a constant density for placement of limit orders across the entire logarithmic price axis, real limit order placement is highly concentrated near the best prices. Roughly $2/3$ of all orders are placed either at the best price or inside the spread. Outside the spread the density of limit order placement falls off as a power law as a function of the distance from the best prices (Bouchaud et al., 2002; Zovko and Farmer, 2002). In addition, we have assumed a constant cancellation rate, whereas in reality orders placed near the best prices tend to be cancelled much faster than orders placed far from the best prices. We cope with these problems by introducing an auxiliary assumption. Basically, we assume that order placement is constant inside an interval, and is zero outside that interval. This is described in more detail below.

In order to estimate the limit order rate density for day $t$, $\alpha_t$, we make an empirical estimate of the distribution of the relative price for effective limit order placement on each day. For buy orders we define the relative price as $\Delta = m - p$, where $p$ is the logarithm of the limit price and $m$ is the logarithm of the midquote price. Similarly for sell orders, $\Delta = p - m$. We then somewhat arbitrarily choose $Q_{t,\text{lower}}$ as the 2 percentile of the density of $\Delta$ corresponding to the limit orders arriving on day $t$, and $Q_{t,\text{upper}}$ as the 60 percentile of $\Delta$. Assuming constant density within this range, we calculate $\alpha_t$ as $\alpha_t = L / (Q_{t,\text{upper}} - Q_{t,\text{lower}})$ where $L$ is the total number of shares of effective limit orders within the price interval $(Q_{t,\text{lower}}, Q_{t,\text{upper}})$ on day $t$. The choice of $Q_{t,\text{upper}}$ is made in a compromise to include as much data as possible for statistical stability, but not so much as to include orders that are unlikely to ever be executed, and therefore unlikely to have any effect on prices.

Similarly, to cope with the fact that in reality the average cancellation rate $\delta$ decreases (Bouchaud et al., 2002) with the relative price $\Delta$, whereas in the model $\delta$ is assumed to be constant, we base our estimate for $\delta$ only on canceled limit orders within the range of the same relative price boundaries $(Q_{t,\text{lower}}, Q_{t,\text{upper}})$ defined above. We do this to be consistent in our choice of which orders are assumed to contribute significantly to price formation (orders closer to the best prices contribute more than orders that are further away). We then measure $\delta_t$, the cancellation rate on day $t$, as the inverse of the average lifetime of a canceled limit order in the above price range. Lifetime is measured in terms of number of events happening between the introduction of the order and its subsequent cancellation.
The parameter $Q_t^{\text{upper}}$ is referred to as $W$ in the main text. In other subsequent studies (to be reported elsewhere) we are able to set the parameter ($Q_t^{\text{lower}} = 0$, and to compute $\Delta$ relative to the opposite best price rather than the midprice, with negligible differences in the results. The difference is that in the later studies we have a cleaner dataset. In this dataset there are some points that are clearly outliers, and it was convenient to introduce a lower cutoff for outlier removal. Thus, we do not feel that ($Q_t^{\text{lower}}$ is an important parameter for this analysis, and we have not discussed it in the text (where we have limited space).

The use of this procedure dictates that it is better to choose an average of daily parameters rather than computing average parameters based on ratios of values for the whole period. Because the width of the interval over which orders are placed varies significantly in time. Moment-by-moment orderbook reconstruction makes it clear that the properties of the market tend to be relatively stationary during each day, changing more dramatically overnight. The order-flows on different days can be rather dissimilar. This non-stationarity of the order flow means that $\delta$ and $\alpha$ parameters calculation would perform poorly if we attempted to use an average price interval over the whole period. This would have the result that on some days we might count only a small fraction of the orderflow, excluding many orders that were important for price formation, while on other days we would include almost all the orders, including many that were not very relevant for price formation. This problem makes it natural to use daily averages of parameters.

This introduces the worry that daily variations in $W$ might be an important predictive variable, above and beyond its effect on changing $\alpha$ (which is consistent with the model). There is a tendency for the value of $W$ on a given day to track the spread, due to regularities in order placement, and therefore to automatically have some correlation with the spread. We have done several studies, which will be reported in a future work, testing the importance of this effect. These show that while daily variations in $W$ do give additional predictability for the spread, other aspects of the model are substantially responsible for these results.

**Measuring the price diffusion rate**

The measurement of the price diffusion rate requires some discussion. We measure the intraday price diffusion by computing the mid-point price variance $V(\tau) = \text{Var}\{m(t+\tau) - m(t)\}$, for different time scales $\tau$. The averaging over $t$ includes all events that change the mid-point price. The plot of $V(\tau)$ against $\tau$ is called a diffusion curve and for an IID random walk is a straight line with slope $D$, the diffusion coefficient.
In our case, the computation of $D$ is as follows: For each day we compute the diffusion curve. In this way we avoid the overnight price changes which would bias our estimate. To the daily diffusion curve we then fit a straight line $V(\tau) = D \tau$ using least squares weighted by the square root of the number of observations for each value of $V(\tau)$. In fitting a straight line we are assuming IID mid-point price movement which is relatively well born out in the data. For an example of see Fig. 3.5. Averaging the daily diffusion rates we obtain the full sample estimate of the stock diffusion. We weigh the averaging by the number of events in each day. One must bear in mind that the price diffusion rate from day to day has substantial correlations, as illustrated in Fig. 3.6.

Figure 3.5: Illustration of the procedure for measuring the price diffusion rate for Vodafone (VOD) on August 4th, 1998. On the $x$ axis we plot the time $\tau$ in units of ticks, and on the $y$ axis the variance of mid-price diffusion $V(\tau)$. According to the hypothesis that mid-price diffusion is an uncorrelated Gaussian random walk, the plot should obey $V(\tau) = D \tau$. To cope with the fact that points with larger values of $\tau$ have fewer independent intervals and are less statistically significant, we use a weighted regression to compute the slope $D$. 
Figure 3.6: Time series (top) and autocorrelation function (bottom) for daily price diffusion rate $D_t$ for Vodafone. Because of long-memory effects and the short length of the series, the long-lag coefficients are poorly determined; the figure is just to demonstrate that the correlations are quite large.

3.5.6 Estimating the errors for the regressions

The error bars presented in the text are based on a bootstrapping method. It may at first seem that the proper method would be to simply use White’s heteroskedasticity consistent estimators, however, we are driven to use this method for two reasons.

First, within each stock the daily values of the dependent variables display slowly decaying positive autocorrelation functions. Averaging the daily values to get an estimate of the stock-specific average may seem to remedy the autocorrelation problem. However, the autocorrelation is very persistent, almost to the scale of the length of our dataset, and the variables may indeed have long memory.\(^6\) This

---

\(^6\)It has recently been shown that order sign, order volume and liquidity as reflected by volume at the best price, are long-memory processes (Bouchaud
makes us suspicious about using standard statistics.

A second reason in using a bootstrapping method for inference is the fact that, in addition to possibly being long memory, the daily values of the variables are cross-correlated across stocks. (A high volatility in one stock on a particular day is likely to be associated with high volatility in other stocks.) These two reasons lead us to believe that using standard or White’s estimators would underestimate the regression errors.

The method we use is inspired by the variance plot method described in Beran (1994), Section 4.4. We divide the sample into blocks, apply the regression to each block, and then study the scaling of the deviation in the results as the blocks are made longer to coincide with the full sample. We divide the $N$ daily data points for each stock into $m$ disjoint blocks, each containing $n$ adjacent days, so that $n \approx N/m$. We use the same partition for each stock, so that corresponding blocks for each stock are contemporaneous. We perform an independent regression on each of the $m$ blocks, and calculate the mean $M_m$ and standard deviation $\sigma_m$ of the $m$ slope parameters $A_i$ and intercept parameters $B_i$, $i = 1, \ldots, m$. We then vary $m$ and study the scaling as shown in Figs 3.7 and 3.8.

Figs 3.7(a) and (b) illustrate this procedure for the spread, and Figs 3.8(a) and (b) illustrate this for the price diffusion rate. Similarly, panels (c) and (d) in each figure show the mean and standard deviation for the intercept and slope as a function of the number of bins. As expected, the standard deviations of the estimates decreases as $n$ increases. The logarithm of the standard deviation for the intercept and slope as a function of log $n$ is shown in panels (e) and (f). For IID normally distributed data we expect a line with slope $\gamma = -1/2$; instead we observe $\gamma > -1/2$. For example for the spread $\gamma \approx -0.19$. $|\gamma| < 1/2$ is an indication that this is a long memory process.

This method can be used to extrapolate the error for $m = 1$, i.e. the full sample. This is illustrated in panels (e) and (f) in each figure. The inaccuracy in these error bars is evident in the unevenness of the scaling. This is particularly true for the price diffusion rate. To get a feeling for the accuracy of the error bars, we estimate the standard deviation for the scaling regression assuming standard error, and repeat the extrapolation for the one standard deviation positive and negative deviations of the regression lines, as shown in panels (e) and (f) of Figs 3.7 and 3.8. The results are summarized in Table 3.2.

One of the effects that is evident in Figs 3.7(c-d) and 3.8(c-d) is that the slope coefficients tend to decrease as $m$ increases. We believe this is due to the autocorrelation bias discussed in Section (3.5.6).
Figure 3.7: Subsample analysis of regression of predicted vs. actual spread. To get a better feeling for the true errors in this estimation (as opposed to standard errors which are certainly too small), we divide the data into subsamples (using the same temporal period for each stock) and apply the regression to each subsample. (a) (top left) shows the results for the intercept, and (b) (top right) shows the results for the slope. In both cases we see that progressing from right to left, as the subsamples increase in size, the estimates become tighter. (c) and (d) (next row) shows the mean and standard deviation for the intercept and slope. We observe a systematic tendency for the mean to increase as the number of bins decreases. (e) and (f) show the logarithm of the standard deviations of the estimates against log $n$, the number of each points in the subsample. The line is a regression based on binnings ranging from $m = N$ to $m = 10$ (lower values of $m$ tend to produce unreliable standard deviations). The estimated error bar is obtained by extrapolating to $n = N$. To test the accuracy of the error bar, the dashed lines are one standard deviation variations on the regression, whose intercepts with the $n = N$ vertical line produce high and low estimates.
CHAPTER 3. THE PREDICTIVE POWER OF ZERO INTELLIGENCE IN FINANCIAL MARKETS

Figure 3.8: Subsample analysis of regression of predicted vs. actual price diffusion (see Fig. 6), similar to the previous figure for the spread. The scaling of the errors is much less regular than it is for the spread, so the error bars are less accurate.

<table>
<thead>
<tr>
<th>regression</th>
<th>estimated</th>
<th>standard</th>
<th>bootstrap</th>
<th>low</th>
<th>high</th>
</tr>
</thead>
<tbody>
<tr>
<td>spread intercept</td>
<td>0.06</td>
<td>0.21</td>
<td>0.29</td>
<td>0.25</td>
<td>0.33</td>
</tr>
<tr>
<td>spread slope</td>
<td>0.99</td>
<td>0.08</td>
<td>0.10</td>
<td>0.09</td>
<td>0.11</td>
</tr>
<tr>
<td>diffusion intercept</td>
<td>2.43</td>
<td>1.22</td>
<td>1.76</td>
<td>1.57</td>
<td>1.97</td>
</tr>
<tr>
<td>diffusion slope</td>
<td>1.33</td>
<td>0.19</td>
<td>0.25</td>
<td>0.23</td>
<td>0.29</td>
</tr>
</tbody>
</table>

Table 3.2: A summary of the bootstrap error analysis described in the text. The columns are (left to right) the estimated value of the parameter, the standard error from the cross sectional regression in Fig. 6, the one standard deviation error bar estimated by the bootstrapping method, and the one standard deviation low and high values for the extrapolation, as shown in Figs 3.7(e-f) and 3.8(e-f).
3.5.7 Market impact

Relation of market impact to supply and demand schedules

The market impact function is closely related to the more familiar notions of supply and demand. We have chosen to measure average market impact in this paper rather than average relative supply and demand for reasons of convenience. Measuring the average relative supply and demand requires reconstructing the limit order book at each instant, which is both time consuming and error prone. The average market impact function, in contrast, can be measured based on a time series of orders and best bid and ask prices.

At any instant in time the stored queue of sell limit orders reveals the quantity available for sale at each price, thus showing the supply, and the stored buy orders similarly show the revealed demand. The price shift caused by a market order of a given size depends on the stored supply or demand through a moment expansion (Smith et al., 2003). Thus, the collapse of the market impact function reflects a corresponding property of supply and demand. Normally one would assume that supply and demand are functions of human production and desire; the results we have presented here suggest that on a short timescale in financial markets their form is dictated by the dynamical interaction of order accumulation, removal by market orders and cancellation, and price diffusion.

Alternative market impact collapse plots

We have demonstrated a good collapse of the market impact using nondimensional units. However, in deciding what “good” means, one should compare this to the best alternatives available. We compare to three such alternatives. In figure 3.9, the top left panel shows the collapse when using non-dimensional units derived from the model (repeated from the main text). The top right panel shows the average market impact when we instead normalise the order size by its sample mean. Order size is measured in units of shares and market impact is in log price difference. The bottom left panel attempts to take into account daily variations of trading volume, normalizing the order size by the average order size for that stock on that day. In the bottom right panel we use trade price to normalize the order sizes which are now in monetary units (British Pounds). We visually see that none of the alternative rescalings comes close to the collapse we obtain when using non-dimensional units; because of the much greater dispersion, the error bars in each case are much larger.
Figure 3.9: Market impact collapse under 4 kinds of axis rescaling. In each case we plot a normalised version of the order size on the horizontal axis vs. a (possibly normalised) average market impact $\log(p_{t+1}) - \log(p_t)$ on the vertical axis. (a) (top left) collapse using non-dimensional units based on the model; (b) (top right) order size is normalised by its mean value for the sample. (c) (bottom right) order size is normalised the average daily volume. (d) (bottom right) Order size is multiplied by the current best midpoint price, making the horizontal axis the monetary value of the trade.
CHAPTER 3. THE PREDICTIVE POWER OF ZERO INTELLIGENCE IN FINANCIAL MARKETS

Error analysis for market impact

Assigning error bars to the average market impact is difficult because the absolute price changes $\Delta p$ have a slowly decaying positive autocorrelation function. This may be a long-memory process, although this is not as obvious as it is for other properties of the market, such as the volume and sign of orders (Bouchaud et al., 2004; Lillo and Farmer, 2003). The signed price changes $\Delta p$ have an autocorrelation function that rapidly decays to zero, but to compute market impact we sort the values into bins, and all the values in the bin have the same sign. One might have supposed that because the points entering a given bin are not sequential in time, the correlation would be sufficiently low that this might not be a problem. However, the autocorrelation is sufficiently strong that its effect is still significant, particularly for smaller market impacts, and must be taken into account.

To cope with this we assign error bars to each bin using the variance plot method described in, for example (Beran, 1994), Section 4.4. This is a more straightforward version of the method discussed in Section (3.5.6). The sample of size $N = 434$ is divided into $m$ subsamples of $n$ points adjacent in time. We compute the mean for each subsample, vary $n$, and compute the standard deviation of the means across the $m = N/n$ subsamples. We then make use of theorem 2.2 from (Beran, 1994) that states that the error in the $n$ sample mean of a long-memory process is $\hat{e} = \sigma n^{-\gamma}$, where $\gamma$ is a positive coefficient related to the Hurst exponent and $\sigma$ is the standard deviation. By plotting the standard deviation of the $m$ estimated intercepts as a function of $n$ we estimate $\gamma$ and extrapolate to $n = \text{sample length}$ to get an estimate of the error in the full sample mean. An example of an error scaling plot for one of the bins of the market impact is given in Fig. 3.10.

A central question about Fig. 3.9 is whether the data for different stocks collapse onto a single curve, or whether there are statistically significant idiosyncratic variations from stock to stock. From the results presented in Fig. 3.9 this is not completely clear. Most of the stocks collapse onto the curve for the pooled data (or the pooled data set with themselves removed). There are a few that appear to make statistically significant variations, at least if we assume that the mean value of the bins for different order size levels are independent. However, they are most definitely not independent, and this non-independence is difficult to model. In any case, the variations are always fairly small, not much larger than the error bars. Thus the collapse gives at least a good approximate understanding of the market impact, even if there are some small idiosyncratic variations it does not capture.
CHAPTER 3. THE PREDICTIVE POWER OF ZERO INTELLIGENCE IN FINANCIAL MARKETS

Figure 3.10: The variance plot procedure used to determine error bars for mean market impact conditional on order size. The horizontal axis $n$ denotes the number of points in the $m$ different samples, and the vertical axis is the standard deviation of the $m$ sample means. We estimate the error of the full sample mean by extrapolating $n$ to the full sample length.

**Market impact in log-log coordinates**

If we fit a function of the form $\phi(\omega) = K \omega^\beta$ to the market impact curve, we get $\beta = 0.26 \pm 0.02$ for buy orders and $\beta = 0.23 \pm 0.02$ for sell orders, as shown in Fig. 3.11. The functional form of the market impact we observe here is not in agreement with a recent theory by Gabaix et al. (2003a), which predicts $\beta = 0.5$. While the error bars given are standard errors, and are certainly too optimistic, it is nonetheless quite clear that the data are inconsistent with $\beta = 1/2$, as discussed in Ref. (Farmer and Lillo, 2004). This relates to an interesting debate: The theory for average market impact put forth by Gabaix et al. follows traditional thinking in economics, and postulates that agents optimize their behavior to maximize profits,
while the theory we test here assumes that they behave randomly, and that the form of the average market impact function is dictated by the statistical mechanics of price formation.

Figure 3.11: The average market impact vs. order size plotted on log-log scale. The upper left and right panels show buy and sell orders in non-dimensional coordinates; the fitted line has slope $\beta = 0.26 \pm 0.02$ for buy orders and $\beta = 0.23 \pm 0.02$ for sell orders. In contrast, the lower panels show the same thing in dimensional units, using British pounds to measure order size. Though the exponents are similar, the scatter between different stocks is much greater.

### 3.5.8 Extending the model

In the interest of full disclosure, and as a stimulus for future work, in this section we detail the ways in which the current model does not accurately match the data, and sketch possible improvements. This model was intended to describe a few average statistical properties of the market, some of which it describes surprisingly well. However, there are several aspects that it does not describe well. Fixing the problems with this models require a more sophisticated model of order flow, including a more realistic model of price dependence in order placement and cancellations (Bouchaud et al., 2002; Zovko
and Farmer, 2002), long-memory properties (Bouchaud et al., 2004; Lillo and Farmer, 2003) and the relationship of the different components of the order flow to each other. This is a much harder problem, and is likely to require a more complicated model. Members of our group are actively working on this problem. While this will certainly have many advantages over the current model, it will also have the disadvantage of introducing more free parameters and thereby complicating the scaling laws (and making the possibility for analytic results more remote).

One of the major ways in which this model is not realistic concerns price diffusion. Real price increments are roughly white, i.e., they are roughly uncorrelated. One might naively think that under IID Poisson order flow, price increments should also be IID. However, due to the coupling of boundary conditions for the buy market order/sell limit order process to those of the sell market order/buy limit order process, this is not the case. Because of the fact that supply and demand tend to build (i.e. the depth of standing limit orders increases) as one moves away from the center of the book, price reversals are more common than price changes in the same direction. As a result, the price increments generated by this model are more anti-correlated than those of real price series. This has an interesting consequence: If we add the assumption of market efficiency, and assume that real price increments must be white, it implies that real order flow should be positively autocorrelated in order to compensate for the anticorrelations induced by the continuous double auction. This has indeed subsequently been observed to be the case (Bouchaud et al., 2004; Lillo and Farmer, 2003).

One of the side effects of this anticorrelation of prices is that it implies that there exist arbitrage opportunities that can be taken advantage of by an intelligent agent. A separate study of these arbitrage opportunities makes it clear that they are not risk-free in the sense usually used in economics. That is, taking advantage of them necessarily involves taking risk, and they do not permit arbitrarily large profits – returns decrease with the size of investment and eventually go to zero. Exploring the nature of these arbitrage opportunities, and the effect that exploiting them has on prices is one of the directions in which this model can be improved (one that is being actively explored). However, we do not feel that the existence of such arbitrage opportunities (which in our opinion mimic those of real markets) presents a serious problem for the purposes for which we are using this model.

In the following bullets we summarize the main directions in which members of our group are working to improve this model.

- *Price diffusion.* The variance of real prices obeys the relation-
CHAPTER 3. THE PREDICTIVE POWER OF ZERO INTELLIGENCE IN FINANCIAL MARKETS

ship $\sigma^2(\tau) = D\tau^{2H}$ to a good approximation for all values of $\tau$, where $\sigma^2(\tau)$ is the variance of price changes or returns computed on timescale $\tau$. The Hurst exponent $H$ is close to and typically a little greater than 0.5. In contrast, under Poisson order flow, as already discussed above, due to the dynamics of the double continuous auction price formation process, prices make a strongly anti-correlated random walk. This means that the function $\sigma^2(\tau)$ is nonlinear. Asymptotically $H = 0.5$, but for shorter times $H < 0.5$. Alternatively, one can characterize this in terms of a timescale-dependent diffusion rate $D(\tau)$, so that the variance of prices increases as $\sigma^2(\tau) = D(\tau)\tau$. Refs. (Daniels et al., 2003; Smith et al., 2003) showed that the limits $\tau \rightarrow 0$ and $\tau \rightarrow \infty$ obey well-defined scaling relationships in terms of the parameters of the model. In particular, $D(0) \sim \mu^2\delta/\alpha^2\epsilon^{-1/2}$, and $D(\infty) \sim \mu^2\delta/\alpha^2\epsilon^{1/2}$. Interestingly, and for reasons we do not fully understand, the prediction of the short term diffusion rate $D(0)$ does a good job of matching the real data, as we have shown here, while $D(\infty)$ does a much poorer job.

- **Market efficiency.** The question of market efficiency is closely related to price diffusion. The anti-correlations mentioned above imply a market inefficiency. We are investigating the addition of “low-intelligence” agents to correct this problem.

- **Correlations in spread and price diffusion.** We have already discussed in Section (3.5.6) the problems that the autocorrelations in spread and price diffusion create for comparing the theory to the model on a daily scale. This is related to the fact that this model does not correctly capture either the fat tails of price fluctuations or the long-memory of volatility.

- **Lack of dependence on granularity parameter.** In Section (3.5.7) we discuss the fact that the model predicts more variation with the granularity parameter than we observe. Apparently the Poisson-based non-dimensional coordinates work even better than one would expect. This suggests that there is some underlying simplicity in the real data that we have not fully captured in the model.

Although in this paper we are stressing the fact that we can make a useful theory out of zero-intelligence agents, we are certainly not trying to claim that intelligence doesn’t play an important role in what financial agents do. Indeed, one of the virtues of this model is that it provides a benchmark to separate properties that are driven by the statistical mechanics of the market institution from those that are driven by conditional intelligent behavior.
Correlations and clustering in the trading of members of the London Stock Exchange


4.1 Introduction

The aim of this paper is to examine the heterogeneity of the trading strategies associated with different members of the London Stock Exchange (LSE). This is made possible by a dataset that includes codes identifying which member of the exchange placed each order. While we don’t know who the member actually is, we can link together the trading orders placed by the same member. Member firms can be large investment banks, in which case the order-flow associated with the code will be an aggregation of various strategies used by the bank and its clients, or at the other extreme it can be a single hedge fund. Thus, while we cannot identify patterns of trading at the level of individual trading strategies, we can test to see whether there is heterogeneity in the collections of strategies associated with different members of the exchange. For convenience we will refer to
the collection of strategies associated with a given member of the exchange as simply a strategy and the member of the exchange as simply an institution, but it should be borne in mind throughout that “a strategy” is typically a collection of strategies, which may reflect the actions of several different institutions, and thus may internally be quite heterogeneous.

We define a strategy by its actions, i.e. by the net trading of an institution as a function of time. If the net volume traded by an institution in a period of time is positive there is an net imbalance of buying, and conversely, if the net volume is negative there is a net imbalance of selling. We test to see whether two strategies are similar in terms of their correlations in the times when they are net buyers and when they are net sellers.

Two studies similar in spirit to this one are (Lillo et al., 2007; Vaglica et al., 2007) in which the authors analyse trading strategies using data from the Spanish Stock Exchange. A number of related studies analysing market correlations can be found in (Laloux et al., 1999a,b; Plerou et al., 2002b, 1999; Potters et al., 2005; Bonanno et al., 2000; Mantegna, 1999; Onnela et al., 2003).

4.1.1 The LSE dataset

The LSE is a hybrid market with two trading mechanisms operating in parallel. One is called the on-book or “downstairs” market and operates as an anonymous electronic order book employing the standard continuous double auction. The other is called the off-book or “upstairs” market and is a bilateral exchange where trades are arranged via telephone. We analyse the two markets separately.

The market is open from 8:00 to 16:30, but for this analysis we discarded data from the first hour (8:00 – 9:00) and last half hour of trading (16:00 – 16:30) in order to avoid possible opening and closing effects.

We base the analysis on four stocks, Vodafone Group (VOD, telecommunications), AstraZeneca (AZN, pharmaceuticals), Lloyds TSB (LLOY, insurance) and Anglo American (AAL, mining). We chose VOD as it is one of the most liquid stocks on the LSE. LLOY and AZN are examples of frequently traded liquid stocks, and AAL is a low volume illiquid stock.

4.1.2 Measuring correlations between strategies

The institution codes we use in this analysis are re-scrambled by the exchange each month for privacy reasons\(^1\). This naturally divides

\(^1\)However, we have found a way to track institutions’ trading on the on-book market for longer time periods. We use this fact in a subsequent part of the
the dataset into monthly intervals which we treat as independent samples. The data spans from September 1998 to May 2001, so we have 32 samples for each stock.

In order to define the trading strategies we further divide the monthly samples into hourly intervals. We believe that one hour is a reasonable choice, capturing short time scale intraday variations, but also providing some averaging to reduce noise. For each of these hour intervals and for each institution individually, we calculate the net traded volume in monetary units (British Pounds). Net volume is total buy volume minus total sell volume. We then assign to each institution and hour interval a +1, -1 or 0 describing it’s strategy in that interval. If the net volume in an interval is positive (the institution in that period is a net buyer) we assign it the value +1. If the institution’s net volume in the interval is negative (the institution is a net seller) we assign it the value -1. If the institution is not active within the interval we assign it the value 0. We discard institutions that are not active for more than 1/3 of the time.

Three examples of trading strategies are shown in figure 4.1. The examples show cumulated trading strategies for three institutions trading Vodafone on-book in the month of November 2000.

In the end we obtain for each month of trading a set of time series representing the net trading direction for each institution $x_i(t)$, which can be organized in a ’strategy matrix’ $M$ with dimensions $N \times T$, where $N$ is the number of active institutions and $T$ is the number of hour intervals in that month. The number of active institutions varies monthly and between stocks. Typical value of $N$ for the on-book market for liquid stocks is around $N \sim 70$, and for less liquid stocks $N \sim 40$. For the off-book market, the numbers are 1.5 to 2 times larger. The number of hourly intervals depends on the number of working days in a month, and is around $T = 7 \times 20 = 140$.

Given the monthly strategy matrices $M$ we then construct the $N \times N$ monthly correlation matrices between the institutions’ strategies. A color example of a correlation matrix for off-book trading in Vodafone in November 2000 is given in the top panel of figure 4.2. Dark colors represent high absolute correlations, with red positive and blue negative. Since the ordering of institutions is arbitrary we use the ordering suggested by a clustering algorithm as explained later in the text. It is visually suggestive that the correlations are not random: Some groups of institutions are strongly anticorrelated with the rest while in turn being positively correlated among themselves.

A formal test of significance involving the t-test cannot be used as

\footnote{Since the data assumes only three distinct values (0,1 and -1) Pearson and Spearman correlations are equivalent.}
CHAPTER 4. CORRELATION AND CLUSTERING IN THE TRADING OF THE MEMBERS OF THE LSE

Figure 4.1: Three examples of institutions’ strategies for on-book trading in Vodafone. We plot the cumulative sum of the (+1, -1 and 0) indicators of hourly net trading volume within a month. The three institutions were not chosen randomly but rather to illustrate three very different trading styles. Institution 2598 appears to be building up a position, institution 3733 could be acting as a market maker, while 3463 seems to be a mix of the two. In reality, only a small number of institutions show strong autocorrelations in their strategy (such as the top and bottom institutions in the plot) and do not have such suggestive cumulative plots.
Figure 4.2: Example of a rearranged correlation matrix for off-book market trading in Vodafone in November 2000. The ordering of institutions is based on the result of the clustering algorithm explained in section 4.2.3. Red colors represent positive correlations between institutions’ strategies, blue represent negative correlations, and darker colors are larger correlations. The dendrogram resulting from the clustering is shown below the correlation matrix. For visual clarity the cluster is cut at height 1.4.
it assumes normally distributed disturbances, whereas we have discrete ternary values. Later in the text we use a bootstrap approach to test the significance. Now, however, we test the significance of the correlation coefficients using a standard algorithm as in ref. (Best and Roberts, 1975). The algorithm calculates the approximate tail probabilities for Spearman’s correlation coefficient $\rho$. Its precision unfortunately degrades when there are ties in the data, which is the case here. With this caveat in mind, as a preliminary test, we find that, for example, for on-book trading in Vodafone for the month of May 2000, 10.3% of all correlation coefficients are significant at the 5% level. Averaging over all stocks and months, the average percentage of significant coefficients for on-book trading is $10.5\% \pm 0.4\%$, while for off-book trading it is $20.7\% \pm 1.7\%$. Both of these averages are substantially higher than the 5% we would expect randomly with a 5% acceptance level of the test.

### 4.2 Significance and structure in the correlation matrices

The preliminary result of the previous section that some correlation coefficients are non-random is further corroborated by testing for non-random structure in the correlation matrices. The hypothesis that there is structure in the correlation matrices contains the weaker hypothesis that some coefficients are statistically significant.

The test for structure in the matrices would involve multiple joint tests for the significance of the coefficients. An alternative method, however, is to examine the eigenvalue spectrum of the correlation matrices. Intuitively, one can understand the relation between the two tests by remembering that eigenvalues $\lambda$ are roots of the characteristic equation $\det(A - \lambda I) = 0$, and that the determinant is a sum of permutations of products of the matrix elements $\det(A) = \sum_{\pi} \epsilon_{\pi} \Pi_{\pi} a_{\pi}$, where $\pi$ are the permutations and $\epsilon_{\pi}$ is the Levi-Civita antisymmetric tensor. On the other hand the test is directly related to principal component analysis, as the eigenvalues of the correlation matrix determine the principal components.

The existence of empirical eigenvalues larger than the values expected from the null implies that there is structure in the correlation matrices and the coefficients are significant.

#### 4.2.1 Density of the correlation matrix eigenvalue distribution

For a set of infinite length uncorrelated time series all eigenvalues of the corresponding correlation matrix (which in this case would be
diagonal) are equal to 1. For finite length time series, however, even if the underlying generating processes are completely uncorrelated, the eigenvalues will not exactly be equal to one - there will be some scatter around one. This scattering is described by a result from random matrix theory (Laloux et al., 1999a,b; Potters et al., 2005). For $N$ random uncorrelated variables, each of length $T$, in the limit $T \to \infty$ and $N \to \infty$ while keeping the ratio $Q = T/N \geq 1$ fixed, the density of eigenvalues $p(\lambda)$ of the correlation matrix is given by the functional form

$$p(\lambda) = \frac{Q}{2\pi\sigma^2} \frac{\sqrt{(\lambda_{\text{max}} - \lambda)(\lambda - \lambda_{\text{min}})}}{\lambda},$$

$$\lambda_{\text{min}}^{\text{max}} = \sigma^2(1 + 1/Q \pm 2\sqrt{1/Q}).$$

(4.1)

$\sigma^2$ is the variance of the time series and $\lambda \in [\lambda_{\text{min}}, \lambda_{\text{max}}]$. Apart from being a limiting result, this expression is derived for Gaussian series. As it turns out the Gaussian assumption is not critical, at least not for the right limit $\lambda_{\text{max}}$, which is the one of interest for this study. We show in a subsequent section a simulation result confirming this observation.

A further consideration is the fact that the parameters $Q = T/N$ and $\sigma$ change every month$^3$, as both the number of hour intervals and the number of institutions vary. Consequently the predicted eigenvalue density under the null changes from month to month. In principle we should construct a separate test for each month, comparing the eigenvalues of a particular month with the null distribution using the appropriate value of $Q = T/N$ and $\sigma$. However, monthly $Q$ and $\sigma$ do not vary too much, and the variation does not change the functional form of the null hypothesis substantially. In view of the fact that Eq. 4.1 is valid only in the limit in any case, we pool eigenvalues for all months together, construct a density estimate and compare it with the null using the monthly averages of $Q$ and $\sigma$.

Figure 4.3 shows the empirical eigenvalue density compared with the expected density under the null for the stock Vodafone. The top figure shows on-book trading, while the bottom figure shows off-book trading. In both markets there are a number of eigenvalues larger than the cutoff $\lambda_{\text{max}}$ and not consistent with uncorrelated time-series. The eigenvalues are larger in the off-book market because the correlation matrices are larger (there are more traders active in the off-book market than the on-book market). The largest eigenvalue in the off-book market is 5 times the noise cutoff whereas it is only 2 times the cutoff in the on-book market. Similar results are found on other stocks as well.

---

$^3$ $\sigma$ is calculated mechanically using the standard formula, as if the time series had a continuous density function rather than discrete ternary values.
Figure 4.3: Empirical density of eigenvalues of the correlation matrices (red) compared to the theoretical density for a random matrix (blue). We see that there are many eigenvalues not consistent with the hypothesis of a random matrix.
4.2.2 Bootstrapping the largest eigenvalues

The weaknesses of the parametric eigenvalue test can be remedied by focusing only on the largest eigenvalues and making a bootstrap test. For each month we construct a realization of the null hypothesis by randomly shuffling buy and sell periods for each institution. In this way we obtain a bootstrapped strategy matrix $M$ in which the institutions’ strategies are uncorrelated, while preserving the number of buying, selling or inactivity periods for each institution. The shuffling therefore preserves the marginal correlations between strategies, meaning that long term (monthly or longer) correlations between institutions are not altered.

The shuffling also removes serial correlation in each institution’s strategy. For most institutions this is not a problem because they do not display autocorrelations in the first place. However, for the group of institutions that do show autocorrelated strategies, this can be an issue.

From the bootstrapped strategy matrix we calculate the correlation matrix and the eigenvalues. This is repeated for each month separately 1000 times.

As already noted, instead of looking at the significance of all empirical eigenvalues, we focus only on the largest two eigenvalues for each month. Correspondingly, we compare them with the null distribution of the two largest eigenvalues for each month: From each of the 1000 simulated correlation matrices, we keep only the largest two. We are therefore comparing the empirical largest two eigenvalues with an ensemble of largest eigenvalues from the 1000 simulated correlation matrices that correspond to the null appropriate to that month. In this way the variations of $Q$ and $\sigma$, as well as small sample properties, are taken into account in the test.

Figure 4.4 shows the results for all 32 months of trading in Vodafone. Again, the top figure shows the monthly eigenvalues for the on-book market while the bottom figure for the off-book market. The largest empirical monthly eigenvalues are shown as blue points. They are to be compared with the blue vertical error bars which represent the width of the distribution of the maximum eigenvalues under the null. The error bars are centered at the median and represent two standard deviations of the underlying distribution. Since the distribution is relatively close to a normal, the width represents about 96% of the density mass. The analogous red symbols show the second largest eigenvalue for each month. We first note that the median of the distribution of the maximum eigenvalue under the random null fluctuates roughly between 2.4 and 2.5. These values are not so different from $\lambda^{max} = 2.5$ which we used in the parametric test. It even seems that in small samples and with ternary
data, the tendency of $\lambda^{\text{max}}$ is to decline as the number of points used decreases, further strengthening the parametric test. The same conclusion can be drawn by looking at the off-book market. The largest eigenvalue is significant in all months on both markets. Corroborating the parametric test, the largest eigenvalues on the off-book market are relatively further away from the corresponding null than for the on-book market, confirming the observation that the correlations are stronger for off-book trading. However, being stronger, they are perhaps of a more simple nature: The second largest eigenvalue is almost never significant for off-book trading, while on the on-book market it is quite often significant.

4.2.3 Clustering of trading behaviour

The existence of significant eigenvalues allows us to use the correlation matrix as a distance measure in the attempt to classify institutions into groups of similar or dissimilar trading patterns. We apply clustering techniques using a metric chosen so that two strongly correlated institutions are ‘close’ and anti-correlated institutions are ‘far away’. A functional form fulfilling this requirement and satisfying the properties of being a metric is (Bonanno et al., 2000)

$$d_{i,j} = \sqrt{2 \cdot (1 - \rho_{i,j})},$$

(4.2)

where $\rho_{i,j}$ is the correlation coefficient between strategies $i$ and $j$. We have tried several reasonable modifications to this form but without obvious differences in the results. Ultimately the choice of this metric is influenced by the fact that it has been successfully used in other studies (Bonanno et al., 2000). We use complete linkage clustering, in which the distance between two clusters is calculated as the maximum distance between its members. We also tried using minimum distance (called “single linkage clustering”), which produced clusters similar to minimal spanning trees but without obvious benefits

The first benefit of creating a clustering is to rearrange the columns of the correlation matrix according to cluster membership. In the top part of figure 4.2 we already showed the rearranged correlation matrix for off-book trading in Vodafone for May 2000. In the bottom part is the corresponding dendrogram. In the correlation matrix one notices a highly correlated large group of institutions as the red block of the matrix. One also notices a smaller number of institutions with strategies that are anti-correlated with the large group. These institutions in turn are correlated among themselves. Finally, to the right part of the matrix there is a group of institutions that is weakly

\footnote{We have also constructed minimal spanning trees from the data but without an obvious interpretation.}
Figure 4.4: Largest eigenvalues of the correlation matrix over the 32 months for the stock Vodafone. The top figure is for on-book trading, the bottom for off-book trading. Blue points represent the largest empirical eigenvalues and are to be compared with the blue error bars which denote the null hypothesis of no correlation. Red points are the second largest eigenvalues and are to be compared with the red error bars. The error bars are centered at the median and correspond to two standard deviations of the distribution of largest monthly eigenvalues under the null
The correlated and anti-correlated groups of institutions are easily identifiable, however, for this month, the clustering algorithm does not properly classify the institutions at the top clustering level. We have added lines to help guide the eye to perhaps a better clustering than the algorithm came up with. It seems that the leftmost group of correlated institutions should have been clustered together with the rest of the correlated institutions.
correlated with both of the previous two. These basic observations are also confirmed in the clustering dendrogram - the dendrogram is plotted so that the institutions in the correlation matrix correspond to the institutions in the dendrogram. Cutting the dendrogram at height 1.7 for example, we recover the two main clusters consisting of the correlated red and the anti-correlated blue institutions. Cutting the dendrogram at a finer level, say just below 1.6, we also recover the weakly correlated cluster of institutions. The structure of the dendrogram below 1.4 is suppressed for clarity of the figure, as those levels of detail are noisy. For other months and other stocks we observe very similar patterns. The top clustering level typically will classify institutions as a larger correlated group and a smaller anti-correlated group.

The clustering for the on-book market is similar, though weaker. Figure 4.5 shows the correlation matrix and the clustering dendrogram for the on-book trading for the same month and stock as the example we showed earlier in figure 4.2. We again see correlated and anti-correlated groups of institutions, as well as the weakly correlated group. The clustering algorithm in this case, however, does not select the correlated and anti-correlated groups at the top level of the clustering, selecting rather the weakly correlated group in one cluster and the other two in the other. At a finer level of clustering (lower height in the dendrogram) the three groups are clustered separately. The clustering algorithm and the distance metric we currently use may not be optimal in selecting the institutions into clusters, but there is indication that the clustering makes sense. In any case, the existence of clusters of institutions based on the correlation in their strategies suggests that it may be possible to develop a taxonomy of trading strategies.

### 4.2.4 Time persistence of correlations

Time persistence, when it is possible to investigate it, offers a fairly robust and strong test for spuriousness. If a correlation is spurious it is not likely to persist in time. In contrast, if the correlations are persistent than the clusters of institutions also persist in time. As noted before, the LSE rescrambles the codes assigned to the institutions at the turn of each month. It is therefore not possible to simply track the correlations between institutions in time. Fortunately, there is a partial solution to this problem. By exploiting other information in the dataset we are able to unscramble the codes over a few months in a row for some institutions. Unfortunately, the method works only for trading on the on-book market and typically
Table 4.1: Regression results of equation 4.3 for correlations between institutions for two consecutive months. Significant slope coefficients show that if two institutions’ strategies were correlated in one month, they are likely to be correlated in the next one as well. The table does not contain the off-book market because we cannot reconstruct institution codes for the off-book market in the same way as we can for the on-book market. The ± values are the standard error of the coefficient estimate and the values in the parenthesis are the standard p-values.

<table>
<thead>
<tr>
<th>Stock</th>
<th>Intercept</th>
<th>Slope</th>
<th>$R^2$</th>
</tr>
</thead>
<tbody>
<tr>
<td>AAL</td>
<td>-0.010 ± 0.004 (0.02)</td>
<td>0.25 ± 0.04 (0.00)</td>
<td>0.061</td>
</tr>
<tr>
<td>AZN</td>
<td>-0.01 ± 0.003 (0.00)</td>
<td>0.14 ± 0.03 (0.00)</td>
<td>0.019</td>
</tr>
<tr>
<td>LLOY</td>
<td>0.003 ± 0.003 (0.28)</td>
<td>0.23 ± 0.02 (0.00)</td>
<td>0.053</td>
</tr>
<tr>
<td>VOD</td>
<td>0.008 ± 0.001 (0.00)</td>
<td>0.17 ± 0.01 (0.00)</td>
<td>0.029</td>
</tr>
</tbody>
</table>

Given the problems with tracking institutions in time we focus only on persistence up to two months. To form a dataset we seek all pairs of institution codes that are present at the market for two months in a row. For all such pairs we compare the correlation in the first of the two months $c_1$ to the correlation in the second of the two months $c_2$. If the correlation between two institutions was high in the first month, we estimate how likely is it that it will be high in the second month as well by calibrating a simple linear regression

$$c_2 = \alpha + \beta \cdot c_1 + \epsilon,$$

assuming $\epsilon$ to be i.i.d. Gaussian. For the stock VOD we identify

---

5In the LSE data we use each order submitted to the limit order book is assigned a unique identifier. This identifier allows us to track an order in the book and all that happens to it during its history. If at the turn of the month (the scrambling period) an institution has an order sitting in the book, we can connect the institution codes associated with the order before and after the scrambling. For example, if an order coded AT82F31E13 was submitted to the book on the 31st by institution 2331, and that same order was then canceled on the 1st by institution 4142, we know that the institution that was 2331 was recoded as 4142. This typically allows us to link the codes for most active institutions for many months in a row, and in several cases even for the entire 32 month period. The LSE has indicated that they do not mind us doing this, and has since provided us with the information we need to unscramble all the codes.
7246 linkable consecutive pairs, for AZN 1623, for LLOY 1930 and for AAL 640. All the regressions are well specified - the residuals are roughly normal and i.i.d. The regression results for the on-book market are summarized in table 4.1. All stocks show significant and positive slope coefficients with \( R^2 \) around 5%. Correlated institutions tend to stay correlated, though the relationship is not strong.

Another sign of persistence is if an institution gets consistently clustered in a given cluster. If two institutions tend to be clustered in a given cluster more often than random then we can infer that the cluster is meaningful. For this purpose we must have a way to distinguish the clusters by some property. A visual examination of many correlation matrices and dendrograms makes it clear that it is often the case that the two top level clusters are typically of quite different sizes. It seems natural to call them the *majority* and the *minority* cluster. Even though it was not always the case, the number of members in the two top clusters differed by a large number more often than not. Acknowledging that this may not be a very robust distinguishing feature, we choose it as a simple means to distinguish the main clusters.

The probability that an institution would randomly be clustered in the minority a given number of times is analogous to throwing a biased coin the same number of times, with the bias being proportional to the ratio of the sizes of the two clusters. If the probability for being in the minority was a constant \( p \) throughout the \( K \) months, the expected number of times \( x \) an institution would randomly end up in the minority would be described by a binomial distribution

\[
B(x, p, K) = \binom{x}{K} p^x (1 - p)^{K-x}.
\] (4.4)

In our case, however, the probability of being in the minority is not a constant, but varies monthly with the number of active institutions and the size of the minority. If the size of the minority is half the total number of institutions, the probability of ending in the minority by chance is 1/2. If the size of the minority is very small compared to the number of total institutions, the probability of ending in that cluster by chance is consequently very small. Denoting by \( \nu_k \) the number of active institutions in month \( k \) and by \( \mu_k \) the number of institutions in the minority cluster, then the probability for an institution to be in minority for month \( k \) by chance is \( p_k = \mu_k / \nu_k \). The expected number of times for an institution to be in the minority by chance is then

\[
P(x, p_k, K) = \prod_{k \in \text{min}} p_k \cdot \prod_{k \in \text{maj}} (1 - p_k),
\] (4.5)
where \( k \) indexes the months in which the institution was in the majority or minority. A further complication is that not all institutions are active on the same months, so that the probability density differs from institution to institution: Depending on which months the institution was active, the above product picks out the corresponding probabilities \( p_k \). Because of this complication we calculate the probability density for each institution through a simulation. We simply pick out the months the institution was active, for each month draw a trial randomly according to \( p_k \), and calculate the number of times the trial was successful, i.e., that the institution ended up in the minority. Repeating this many times we get the full distribution function for the number of times the institution can end up in the minority at random for each institution.

Table 4.2: Result of the test on minority members for on-book trading in Vodafone. In bold are institutions whose behavior is not consistent with the hypothesis of random behavior.

<table>
<thead>
<tr>
<th>Inst. code</th>
<th>Times in minority</th>
<th>Out of possible</th>
<th>Prob. of non-random behavior</th>
</tr>
</thead>
<tbody>
<tr>
<td>3265</td>
<td>16</td>
<td>32</td>
<td>0.99</td>
</tr>
<tr>
<td>2548</td>
<td>7</td>
<td>32</td>
<td>0.14</td>
</tr>
<tr>
<td>2575</td>
<td>6</td>
<td>32</td>
<td>0.07</td>
</tr>
<tr>
<td>2533</td>
<td>3</td>
<td>19</td>
<td>0.11</td>
</tr>
<tr>
<td>2040</td>
<td>14</td>
<td>31</td>
<td>0.97</td>
</tr>
<tr>
<td>1720</td>
<td>9</td>
<td>20</td>
<td>0.93</td>
</tr>
<tr>
<td>1876</td>
<td>5</td>
<td>14</td>
<td>0.73</td>
</tr>
<tr>
<td>2688</td>
<td>8</td>
<td>30</td>
<td>0.34</td>
</tr>
<tr>
<td>1776</td>
<td>11</td>
<td>22</td>
<td>0.99</td>
</tr>
<tr>
<td>2086</td>
<td>9</td>
<td>23</td>
<td>0.86</td>
</tr>
<tr>
<td>0867</td>
<td>10</td>
<td>22</td>
<td>0.95</td>
</tr>
<tr>
<td>2995</td>
<td>12</td>
<td>20</td>
<td>1.00</td>
</tr>
<tr>
<td>2569</td>
<td>7</td>
<td>21</td>
<td>0.64</td>
</tr>
</tbody>
</table>

Similarly, because we are using the institution codes over intervals of more than one month, we can perform this test only for institutions on the on-book market. We limit the test to the stock Vodafone and apply it only on institutions that we can track for more than 12 out of the 32 months. This results in 13 institutions on which we base the test. For other stocks we are not able to track institutions for long periods and the power of the test would be weak.

The results for the 13 institutions are given in table 4.2. The leftmost column is the institution code, followed by the number of times that institution has been in the minority. The column named 'Out
of possible’ counts the number of months an institution has been present in the market - it is the maximum number of times it could have been in the minority. Finally, the rightmost column gives one minus the probability that the institution could have randomly been so many times in the minority. We choose to display one minus the probability as it represents the probability of accepting the hypothesis that the behavior of that institution is not consistent with the random null hypothesis. Most institutions have quite high probabilities of non-random behavior and in bold we select the institutions which pass the test at the 5% level. Out of 13 institutions, 5 of them have been in the minority cluster more often than they would have been just by chance at the 5% acceptance level. This is substantially higher than the expected number of 0.65 out of 13 tested at this acceptance level.

4.3 Conclusions

We have shown that even very crude definitions of institutions’ strategies defined on intervals of an hour period produce significant and persistent correlations. On the off-book market these correlations are organized in a way that there is typically a small group of institutions anti-correlated with a larger second group. The strategies within the two groups are correlated. Clustering analysis also clearly reveals this structure. The volume transacted by the smaller group, typically containing no more than 15 institutions on Vodafone, accounts for about half of the total trading volume. The larger group, typically of around 80 institutions on Vodafone, transacts the remaining half of the total trade volume. This is an indication that the smaller group can be identified as the group of dealers on the off-book market. They provide liquidity for the larger group of institutions and their strategies are anti-correlated: the dealers buy when the other institutions are selling and vice versa. The single large monthly eigenvalue in the off-book market is related to this basic dynamics.

Contrary to the off-book market, the on-book market does not display only one large eigenvalue. There are typically one or two significant eigenvalues for each month. The eigenvalues are relatively smaller and the correlations not as strong. Still, we are able to identify the basic clustering structure seen on the off-book market, namely a small and large group of anti-correlated strategies. However, the volume traded by the small cluster does not seem to equal the volume of the large cluster. The dynamics seems to be more complicated. The largest eigenvalue may still be related to transactions between the two clusters of institutions, however the occasional
second largest eigenvalue suggests that there is more complicated dynamics taking place.

These results suggest that trading on the LSE is a relatively structured process in the aspect of trading strategies. At a given time, there are groups of institutions all trading in the same direction, with other groups trading in the opposite direction, providing liquidity.

It is important to stress that what we have conveniently labeled a “strategy” is more typically a collection of strategies all being executed by the same member of the exchange. From this point of view it is particularly remarkable that we observe heterogeneity, as it depends on the tendency of certain types of strategies to execute through particular members of the exchange (or in some cases that pure strategies take the expense to purchase their own membership). One expects that if we were able to observe actual account level information we would see much cleaner and stronger similarities and differences between strategies.
Chapter 5

Market imbalances and stock returns: heterogeneity of order sizes at the London Stock Exchange

5.1 Introduction

The connection of trade volume and stock returns has been well established (Andersen, 1996; Brock and LeBaron, 1996; Weber and Rosenow, 2004; Tauchen and Pitts, 1983; Chordia and Swaminathan, 2000). Large price moves are associated with large trade volume. On a finer level of detail, it has been shown that price moves are also driven by the properties of order flow (Plerou et al., 2002a; Berger et al., 2006; Carlson and Lo, 2006; Evans and Lyons, 2002; Payne, 2003; Farmer et al., 2005; Kaniel et al., 2008; Kumar and Lee, 2005; Gopikrishnan et al., 2000; Gabaix et al., 2003b; Maslov and Mills, 2001; Solomon and Richmond, 2001). For example, the number of buyer or seller initiated trades and volumes (signed order flow) has influence on price returns (Plerou et al., 2002a). In this paper we show that, in addition to the orderflow, the heterogeneity in order sizes plays a role in price formation. We investigate the pressures exhibited on the price by the trading of large orders either matched or unmatched by large orders on the other side of the market.¹

¹Throughout the paper, when referring to the two sides of the market, we mean the bid and ask sides of the market, not the sellers and buyers of financial services.
Table 5.1: Basic facts and trading statistics for the analysed stocks. From left to right the columns are: stock ticker symbol, company name, industry, market capitalization in 2005, median daily number of trades in the analysed period for the on-book and off-book; median daily volume for the two markets and total; median daily number of trading firms.
On the other hand, the effect of volume moving the price has been interpreted in the context of informed and uninformed trading, as in for example Madhavan and Cheng (1997) and Smith et al. (2001). It is said that trades based on information (informed trades) move the price, while liquidity based trades (uninformed trades) do not. Our analysis shows evidence that it may be more appropriate to consider the lack of liquidity, not informed trading, that moves prices.

Using codes identifying member firms in a dataset from the London Stock Exchange (LSE), we disaggregate the daily and hourly trade volume into quantities bought and sold by individual member firms and call this the member firms’ size in the given interval. Unfortunately, codes in the LSE data identify only the aggregate order flow of a member firm, not individual orders. To obtain the actual orders from aggregate data, Vaglica et. al. (Vaglica et al., 2007) use an algorithm detecting “patches”. We take a different approach and analyse the data on two different timescales, hourly and daily. Given the indication that a firm’s trading is typically dominated by one large order at any given time (Vaglica et al., 2007), we assume that at one of the timescales the firms’ trades are corresponding to the firms’ orders and hope to detect the influence of order properties on at least one of the timescales. As we will show, the results are significant on both timescales and draw similar conclusions.

The order sizes are found to be extremely heterogenous (Vaglica et al., 2007), seemingly reflecting the heterogeneity of investor sizes (Zipf, 1949; Pushkin and Aref, 2004; Gabaix et al., 2006). Since large order sizes can incur the cost of significant market impact, different market mechanisms are designed to help limit this cost. As the LSE is a hybrid market with concurrent trading through an anonymous limit order book (on-book) and through a dealership market (off-book), by comparing them we show that indeed there are different impacts associated with different mechanisms.

The on-book market of the LSE largely uses the continuous double auction, while the off-book is an electronic quotation market where trading is ultimately done via phone. The member firms are institutions entitled to trade on the LSE. For example, large investment banks and hedge funds which directly send orders to the LSE are member firms. Investment banks can act either as a broker or can trade for their own account, while hedge funds normally trade for their own account.

We base the analysis on 12 stocks with ticker symbols AAL, RTR, CGNU, TSCO, PRU, DGE, LLOY, AZN, SHEL, HSBA, BPA, VOD. The period of the analysis is from 2002 to 2004 comprising 32 months of trading which corresponds to 674 trading days. Some basic facts and trading statistics for the analysed stocks is given in table 5.1. The markets are open from 8:00 to 16:30 but we discard data from
the first and last half hour of trading to avoid possible opening or closing anomalies. The member firm codes that make this analysis possible do not identify the firms by name and are scrambled monthly and across stocks. This makes it impossible to track member firms in time and investigate the properties of order sizes for individual firms with good statistics.

## 5.2 Distribution of order sizes

Apart from being interesting in the context of market heterogeneity, the distribution of order sizes (the volume a firm has bought or sold in a day) is important in its own right. On the one hand it may be helpful to the literature on heterogenous agent based models Boswijk et al. (2007); Hommes (2006), on the other hand some rational expectation models Gabaix et al. (2006) use this distribution as an input of the agents optimization problem.

However, total trade volume has grown over the years and it is not a-priori clear that the order size distribution is a stable distribution. We find that on the LSE the increase in total trade volume was paralleled by the increase in the number of firms trading. For on-book trading, the number of firms trading in a low-activity stock has increased from about 30 to about 40. For a high-activity stock the
numbers increased from about 70 to about 90. For off-book trading, the numbers are approximately 50% higher. Therefore, the typical daily volume traded by a firm seems to have remained more or less constant. For high-activity stocks (eg., BPA, VOD), the median daily volume traded by a firm on-book is about 5–7 million Pounds, while for low-activity stocks (eg., AAL, CGNU, DGE) it is about 1–2 million Pounds throughout the sample. The numbers on the off-book market are lower, from 0.8-1 million for high activity stocks to 0.3-0.5 million for low activity stocks.

Therefore, it seems reasonable to decompose the total daily volume $V_t$ on day $t$ as

$$V_t = \sum_{i=1}^{N_t} v_{t,i},$$

where $N_t$ is the number of firms trading on day $t$ and $v_{t,i}$ is the daily firm trade volume. Decompositions are done each day separately for the sell and buy side of the market.

As noted, we consider $v_{t,i}$ to correspond the firm’s order size. At either the daily or hourly timescale, we believe this is a reasonable assumption. However, it is a caveat to keep in mind.

Apart from the different means from stock to stock, the distribution of order sizes $v_{t,i}$ seems to be stable over time. When normalized by dividing with the mean size for each stock, the densities seem the same across stocks. In the insets of figure 5.1 we show the density functions of normalized order sizes for all stocks in the two markets. The curves collapse allowing us to pool the data to produce the density estimates in the main figures.

The main graphs of figure 5.1 represent the estimated density function\(^2\) of normalized order sizes pooled across all stocks. The left panel represents on-book trading, while the right represents off-book trading. Both densities show power-law behaviour in the tails with exponents around 3 for on-book and 3/2 for off-book trading, which were obtained by fitting a power-law to the tails of the distribution for values larger than 10. Looking at the Hill plots in figure 5.2 we see that the power law behavior in the on-book indeed seems to be valid for orders larger than 10 times the average size. For the off-book market the threshold is not as clear.

Disaggregating the volume on hourly intervals we obtain strikingly similar distributions and the same exponents. This remains to be investigated in the future. As a further speculation, the functional forms of the densities are very similar in shape to the Tsallis q-Gaussian (for the on-book market) and double q-Gaussian (for the off-book market) (Tsallis et al., 2003).

---

\(^2\)Actually shown is one minus the cumulative distribution function (CDF).
Figure 5.2: Maximum likelihood estimate of the Hill exponent for the on-book (left) and off-book (right) markets. Panels display Hill plots which show the value of the Hill exponent as a function of either the order statistics (the n-th largest value beyond which power-law behavior starts) or the threshold (the corresponding value of the variable).

The model of Gabaix et al. (2006) predicts the tail exponent of optimal order size based on a rational agent minimizing impact to be equal to 3/2. Their argument depends on using first order risk aversion. If one measures risk aversion using the variance, the exponent becomes 3 instead. In addition, their model rests on the assumption that investor sizes are distributed according to a power law, while new evidence suggests the distribution may be more close to a log-normal (Schwarzkopf and Farmer, 2008). In any case, the empirical exponents found in several studies are not in complete agreement. Vaglica et al. (2007) have found the exponent to be around 2 for the Spanish stock market. Gabaix et al. (2003b) and Gopikrishnan et al. (2000) quote for the transaction volumes of the Paris Bourse and the NYSE/NASDAQ exponents of 3/2. Lillo et al. (2005), using transaction volumes from the same LSE dataset we use, found the exponents to be around 2.9 for the on-book market and 1.6 for the off-book market.

5.3 Order size heterogeneity and stock returns

Since the order size distribution is a fat tailed distribution, the likelihood that a disproportionally large order be present on the market on a given day is not negligible. In such a situation the composition
of member firm order sizes is very heterogenous and extremely uneven. For example, it may happen that out of the total of 80 firms, the largest 2 firms account for 90\% of the sell volume, while the remaining 78 account for only 10\% of the sell volume. Similar size heterogeneity may or may not happen on the buy side of the market forcing the two large sellers to either transact with many counter parties to fill their order or to transact with a small number of large counter parties.

The terminology homogenous and heterogenous is inspired by the notion that a heterogenous partition of firm order sizes is a situation where there are great differences in the order sizes traded by member firms. A homogenous order partition is a situation when the order sizes traded by member firms are similar. An alternative choice of words would be to call a heterogenous composition of firms a concentrated composition, since most of the trade volume is concentrated in a few number of large firms. A homogenous composition would then be labeled as a diluted, or equal composition, where most firms are trading equal volumes.

The interplay of the two trading sides possibly determines demand and supply pressures and may ultimately influence the price move. However, it is not a-priori clear what kind of price pressure the above example situation may produce. In the setting of a financial market it may be plausible that a large seller will have to sell at a discount to a large number of buyers. In a traditional market for goods however, a seller which is the only source of a good usually is considered to have monopoly power over small buyers and may sell with a premium. We empirically address this question by quantifying the heterogeneity of firms’ order sizes and investigating its’ influence on price returns.

We use two similar approaches to quantify the size heterogeneity. One approach, rooted in statistical mechanics, uses the entropy, the other, rooted more in economics, uses the Gini index. Both approaches give identical results in terms of interpretation.

To calculate the entropy of a partition of the total volume for the sell (buy) side, we denote by $w_i$ the fraction of the total volume sold (bought) by firm $i$. We then calculate the entropy as

$$ E_s = \sum_{i=1}^{n} w_i \log(1/w_i), \quad i \in \text{firms}, $$

where by $E_s$ we denote the sell side and analogously by $E_b$ the buy side entropy. $n$ is the number of firms on the side of the market in question. The imbalance between the two sides is naturally formed as the difference of entropies. We assign to the sell side the positive
sign and define the entropy imbalance as
\begin{equation}
\delta E = E_s - E_b.
\end{equation}

For a heterogenous volume partition the entropy is low\(^3\), therefore
a positive value of the difference denotes that the buy side of
the market is more heterogenous than the sell side. Alternatively, a
large negative value of the difference means the sell side is more
heterogenous and can be interpreted as the presence of a sell order
on the market, unmatched by a large buy order.

To calculate the Gini index of a partition, we use the expression (Dixon et al., 1987; Sen, 1973)
\begin{equation}
G = \frac{1}{2n} \sum_{i=1}^{n} \sum_{j=1}^{n} |w_i - w_j|.
\end{equation}

The expression formulates the Gini index as the mean relative difference
between all pairs of orders. If the partition is extremely
uneven (one firm trading all the volume and the other firms trading
infinitesimal amounts), the Gini index approaches one. If the part-
tition is perfectly even (all firms trading equal amounts) the Gini
index is zero.

In terms of the Gini indices we define the imbalance as
\begin{equation}
\delta G = \frac{G_b - G_s}{G_b + G_s},
\end{equation}
denoting the Gini index of the sell side of the market by \(G_s\) and
the Gini index of the buy side of the market by \(G_b\). With this
construction, similar to the entropy, the Gini imbalance is positive
when buy orders are more heterogeneous, and is negative when the
sell orders are more heterogeneous. The imbalance is zero when both
sides of the market are similarly fragmented.

The difference between the two measures is that the entropy, in
addition to the information about the partition of the total volume,
contains also information about the number of firms. The larger the
number of firms among which the volume is distributed, the larger
the entropy.

Both the entropy and Gini measures are approximately normally
distributed due to the sums in their definition and the central limit
theorem. In terms of the time structure, both indices on a given
side of the market are substantially serially correlated, however the
imbalance variables are i.i.d.

We address the question of influence of the order size heterogeneity
imbalance on the returns in terms of linear regressions. However,

\(^3\)In the terminology we adopt, a uniform partition of total trade volume among
the member firms is a homogeneous partition.
in addition to the size imbalance, we also include the imbalances in the signed order flow - it has been shown, for example in (Plerou et al., 2002a), that signed order flow has explanatory power for price returns. We include the imbalance in the signed number of trades $T$, signed transaction volume $V$ and numbers of buying and selling firms $N$. All imbalance variables are constructed as

$$
\delta X = \frac{X_b - X_s}{X_b + X_s},
$$

where $X$ stands for any of the three variables. Finally, each imbalance variable is normalized by dividing with the standard deviation of the variable for the corresponding stock.
For the on-book market we have a full trading record (all limit orders, market orders and cancelations), and can therefore determine the sign of a trade precisely. However, for the off-book market this is not possible as the trades are negotiated via phone. To infer the trade sign we adopt a common algorithm where the trades are labeled as buyer or seller initiated depending on where the transaction price is in respect to the current bid and ask prices (Lee and Ready, 1991). The transactions close to the bid (i.e., below the mid-price) are labeled as seller initiated and the ones close to the ask (i.e., above the mid-price) as buyer initiated.

Our main result concerns the price returns and its relation to market imbalances. Other market variables such as the spread and volatility, though very important for markets, can not be analysed with a simple regression model as they are strongly correlated in time; this is left for future work.

The price returns are calculated as the log difference of the volume weighted average price (VWAP) of the last 10% and first 10% of trades of the day or hour, depending on the time scale of the analysis. The returns are normalized by the stock specific standard deviation. We also performed the analysis with market adjusted returns (subtracting from the non-normalized stock returns the FTSE100 returns) and obtained basically the same results.

As the results for the Gini imbalance and the entropy imbalance confirm each other, we show only the analysis and results for the entropy imbalance. We estimate the following regression of the price returns $\delta P$

$$\delta P_t = \alpha \cdot \delta E_t + \beta \cdot \delta V_t + \gamma \cdot \delta T_t + \tau \delta N_t + \epsilon_t, \quad (5.7)$$

where $\delta E_t$, $\delta V_t$, $\delta T_t$ and $\delta N_t$ are the previously described entropy, volume, number of trades and number of firms imbalance measures. To be able to compare the numerical values of the regression estimates between the hourly and daily regressions, we premultiply the hourly returns by the number of trading hours in a day (9 hours). This is to bring the scale of hourly returns to the scale of daily returns. The correlations, joint dependencies and histograms for the five variables are shown in figure 5.3. Some of the explanatory variables, such as signed trades and signed volume, are strongly correlated. This may lead to instabilities in coefficient estimates for those variables and we need to keep this in mind when interpreting results.

The results for the on- and off-book markets, as well as for the daily and hourly returns are collected in table II. Apart from the value of the coefficient, its error and p-value, we list also $R^2_s$ and $R^2_p$, $R^2_s$ is the value of R-square of a regression with only the selected variable, and no others, included. It is equal to the square root of the absolute value of the correlation between the variable and the
Table 5.2: Regression results showing the significance of the market imbalance variables on price returns. Columns from left to right are estimated coefficient, its error and in the parenthesis the p-value of the test that the coefficient is zero assuming normal statistics; $R_s^2$ is the value of $R^2$ in a regression where only the selected variable is present in the regression. It expresses how much the variable on its own (solo) explains price returns. Final column $R_p^2$ is the partial $R^2$ of the selected variable. It expresses how much the variable explains price returns above the other three variables. We show separate results for the on- and off-book market, as well as for the daily and hourly returns. The constant in the regressions is not reported.
price returns. The column labeled $R^2_p$ is the partial R-square defined as $R^2_p = (R^2_2 - R^2_1)(1 - R^2_1)$ where $R^2_2$ is the R-square of a regression with all variables included, and $R^2_1$ is the R-square of a regression with all but the selected variable. $R^2_p$ is the partial contribution to the R-square due to the selected variable.

As we see from the table, the imbalances play a significant role in determining the price move on both markets and both time scales. All imbalance variables, except for the number of trades in the daily analysis, are significant beyond 1% level. However, signed trades are (statistically) insignificant only when analysed together with other orderflow variables due to the strong cross-correlations. In terms of economic significance, by looking at $R^2_s$ and $R^2_p$, the most important contribution seems to come from the signed volume imbalance, followed by the number of firms and entropy in the daily analysis, and the number of trades for the hourly analysis.

As expected, a large contribution to the price returns comes from the orderflow. Surprisingly however, one of the orderflow variables, the signed volume, changes sign between the two markets. We have performed reality checks by separating the data by stocks and by years (the data of each stock we separated in 3 one-year periods), and in all those subsamples found the same sign reversal. This suggests that an increase in the buyer initiated volume on the on-book market drives the price up, but a similar increase in buyer initiated volume on the off-book market drives the price down (analogously for the sell volume). The reason for this sign reversal seems to lie in cross market trading. The correlation of the volume imbalance between the two markets is significantly negative (about -0.15), implying that an increase in buyer initiated volume on the on-book volume is linked to an increase in the seller initiated volume on the off-book market (and vice versa). This effect is very suggestive of market making: for example, a firm buying on the on-book market and selling on the off-book market. Unfortunately, with our dataset we can not investigate further this behaviour as the member firm codes in the on- and off-book markets can not be linked.

However, an important point that is implied by this sign reversal is that it seems that the on-book market is responsible for the majority of price discovery on the LSE. The sign of the volume imbalance for the on-book market is positive, implying the intuitive behaviour that an increase in buyer initiated volume is linked with an increase in price. In contrast, an increase in buyer initiated volume on the off-book market leads to a decrease in price. We suspect that this is in fact due to the interaction of the two markets and that the increase of buyer-initiated volume on the off-book market is related to the increase of seller-initiated volume on the on-book market explaining the decrease in price. As we see here, one needs
CHAPTER 5. MARKET IMBALANCES AND STOCK RET.: HETEROGENEITY OF ORDER SIZES AT THE LSE

to be very careful to include all sources of price adjustment forces when explaining price returns. This question is very interesting in its own right; unfortunately we leave it for further research.

In terms of the statistical specification of the model, residuals are i.i.d. and very close to normal. All the explanatory variables are exogenous to the model. Furthermore, we have also confirmed that the statistical errors under normality assumptions estimated in the model are correct by a bootstrap test. By shuffling the price returns and keeping all other variables intact, we obtain a realization of the null hypothesis where all the explanatory variables are correlated but are uncorrelated with the returns. Repeating the shuffling 1000 times and estimating the model on the bootstrapped data, we get a distribution of the coefficients under the null. The standard deviation of the estimates and the p-values obtained in this way coincide with the theoretical values shown in the table. Estimating recursive and sliding window regressions are all in line with the overall estimates. Looking at the regressions for each stock individually also leads to the same conclusions.

The overall $R^2$ on the on-book market is 32% for the daily analysis and 26% for hourly analysis. For the off-book market it is substantially lower at 7% for daily and 0% for hourly regressions. This suggests that the quotation market design (the off-book) substantially helps in limiting the price impact of order heterogeneity, but does not remove it completely. Finally, from the signs of the coefficient estimates we can summarize several facts.

1. The excess of buyer initiated over seller initiated volume drives prices up on the on-book market and down on the off-book market (and vice versa for the excess of seller initiated volume over buyer initiated volume).

2. Excess heterogeneity of buy order sizes seems to drive the price up; excess heterogeneity of sell orders seems to drive it down.

3. An excess in the number of buying firms pushes the price down, an excess of selling firms drives it up.

4. The number of transactions only plays a role at the hourly scale, not on the daily scale.

At this point it may be interesting to compare our regression results to the usual information based paradigm, such as for example in Keim and Madhavan (1996); Madhavan and Cheng (1997) and Smith et al. (2001). They also regress the price returns to signed volume (among other things) and compare the different effects on an on-book and off-book market, but differ in their interpretation of the price impact – or its lack of – by using concepts of informed and
uninformed trading. The regression results are nevertheless readily comparable.

The main difference in the regressions is their model specification which is parametrized for the comparison of the two markets. They estimate the effects using only one regression, where we use a separate regression for each market. By using only one regression equation, they are estimating the price effect of signed volume on both markets by two parameters. One is the overall effect of the signed volume (ignoring if the trade was coming from the on- or off-book market) and the other which is the correction to the first parameter for trades coming from the off-book market. They find a positive coefficient for the overall parameter, and a negative coefficient for the correction parameter, which they interpret as evidence that the off-book trades are less informed then the on-book trades. Indeed, we also find both a smaller coefficient and a substantially smaller $R^2$ for the off-book market trades. Our interpretation though is different, as we are saying that the off-book trades have a smaller price impact due to liquidity.

The issue of sign reversal of the volume effect is also not at odds with their results. Firstly, in the estimates of Smith et al. (2001) the magnitude of the correction parameter can easily be greater then the main effect parameter, effectively making the volume effect on the off-book market negative, and in one of their samples Madhavan and Cheng (1997) in fact report this. Another fact which may raise questions about the stability of their estimates is the disturbing asymmetry in the buyer and seller initiated trades that the studies find (they find the price effect for seller initiated trades, but no effect for buyer initiated trades). Asymmetry is always either very good, in which case it points to fundamental flaws in the current understanding of the problem, or is very bad, in which case it points to the problem of the analysis. The bottom line being, the reasons for accepting asymmetries where we do not expect them have to be very well grounded. Developed financial markets, such as all markets analysed in this discussion, to the authors’ knowledge do not show much evidence for the buyer/seller asymmetry (apart from trivial sign changes).

The information paradigm is also at odds with our following result. From “Fact 2” we can conclude that when a large order is transacted against multiple small orders it incurs a higher cost of price impact. In the information interpretation, this would suggest that large orders carry more information than small orders put together. This in itself can be interesting as it suggests that firms placing large orders are more ‘informed’ then the many small firms they transact with. The question that naturally arises in this context is what is the impact of a large order when it is transacted with
a comparably large order. To investigate this we create four dummy variables, each indexing one of the four market situations of interest. Variable D1 indexes instances when the sell side is heterogenous and the buy side homogenous, meaning there are a few large sell orders and many small buy orders. D2 indexes instances with the sell side homogenous and buy side heterogenous, the opposite of D1. D3 indexes instances with both sides being homogenous, i.e., many small orders on both sides. Finally D4 indexes situations with both sides being heterogenous, i.e., few large orders on both sides of the market. In order for a market side to be labeled as heterogenous the entropy needs to be smaller than the 25th quantile of the entropy density. A homogenous market side corresponds to the entropy being larger than the 75th quantile. In table 5.3 we display the results for a regression of daily price returns against signed orderflow and the four dummy variables. As expected, the order flow variables are significant and of the same sign as observed before. Of the four conditional means corresponding to the dummy variables, only D1 and D2 are significantly different from zero. This means that only situations with one side of the market being heterogenous leads to price impact. If both sides of the market are similarly fragmented, i.e., large sell orders trading with large buy orders, there is no large price impact. This suggests another fact:

5. Transacting of large buy orders against large sell orders does not substantially move the prices (and vice versa).

Using the informed vs. uninformed trading paradigm in this context, which of the oppositely trading large firms was correctly informed and which one was wrong? If both were transacting with small firms, they would incur a positive price impact, implying that their order was “informed”, however when transacting with a comparably large firm, there is no impact. Is it possible that somehow their different information sets cancel each other when confronted in the market? It is not obvious that such questions have an answer. To the authors this is suggestive that it may be liquidity and its absence that limits price impact of large trades, not the fact that large trades are in some ways more informed.

As a last corroboration of the significance of the entropy imbalance, we estimate partial regressions controlling for the influence of the orderflow. We first regress the returns $\delta P_t$ and the entropy imbalance $\delta E_t$ on the orderflow:

$$
\delta P_t = \alpha_1 \cdot \delta V_t + \beta_1 \cdot \delta N_t + \gamma_1 \cdot \delta T_t + \epsilon_{1,t},
$$

$$
\delta E_t = \alpha_2 \cdot \delta V_t + \beta_2 \cdot \delta N_t + \gamma_2 \cdot \delta T_t + \epsilon_{2,t},
$$

(5.8)

To estimate the entropy density, we merge the sell and buy sides as they seem identical.

---

4To estimate the entropy density, we merge the sell and buy sides as they seem identical.
<table>
<thead>
<tr>
<th>Daily</th>
<th>Coef.</th>
<th>Error</th>
<th>p-val</th>
</tr>
</thead>
<tbody>
<tr>
<td>on-book market</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Intercept</td>
<td>-0.07</td>
<td>0.02</td>
<td>0.00</td>
</tr>
<tr>
<td>Signed volume, $V$</td>
<td>0.40</td>
<td>0.01</td>
<td>0.00</td>
</tr>
<tr>
<td>No. firms, $N$</td>
<td>-0.27</td>
<td>0.01</td>
<td>0.00</td>
</tr>
<tr>
<td>No. signed trades, $T$</td>
<td>-0.01</td>
<td>0.25</td>
<td>0.52</td>
</tr>
<tr>
<td>D1 (large sell orders, small buy orders)</td>
<td>-0.19</td>
<td>0.04</td>
<td>0.00</td>
</tr>
<tr>
<td>D2 (large buy orders, small sell orders)</td>
<td>0.13</td>
<td>0.04</td>
<td>0.00</td>
</tr>
<tr>
<td>D3 (small sell orders, small buy orders)</td>
<td>-0.00</td>
<td>0.86</td>
<td>0.15</td>
</tr>
<tr>
<td>D4 (large sell orders, large buy orders)</td>
<td>0.03</td>
<td>0.41</td>
<td>0.53</td>
</tr>
<tr>
<td>off-book market</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Intercept</td>
<td>-0.064</td>
<td>0.011</td>
<td>0.00</td>
</tr>
<tr>
<td>Signed volume, $V$</td>
<td>-0.109</td>
<td>0.006</td>
<td>0.00</td>
</tr>
<tr>
<td>No. firms, $N$</td>
<td>-0.071</td>
<td>0.008</td>
<td>0.00</td>
</tr>
<tr>
<td>No. signed trades, $T$</td>
<td>-0.008</td>
<td>0.008</td>
<td>0.52</td>
</tr>
<tr>
<td>D1 (large sell orders, small buy orders)</td>
<td>-0.109</td>
<td>0.003</td>
<td>0.00</td>
</tr>
<tr>
<td>D2 (large buy orders, small sell orders)</td>
<td>0.13</td>
<td>0.04</td>
<td>0.00</td>
</tr>
<tr>
<td>D3 (small sell orders, small buy orders)</td>
<td>-0.00</td>
<td>0.86</td>
<td>0.15</td>
</tr>
<tr>
<td>D4 (large sell orders, large buy orders)</td>
<td>0.03</td>
<td>0.41</td>
<td>0.53</td>
</tr>
</tbody>
</table>

Table 5.3: Regression results showing the effect of various market imbalances and its effect on the price returns. The dummy variables D1–D4 index market heterogeneity situations, for example the trading of few large sell orders with a large number of small buy orders. Details are in the text.
and then regress the residuals $\delta P_t'$ and $\delta E_t'$ obtained from the above regressions on each other

$$
\delta P_t' = \alpha \cdot \delta E_t' + \epsilon_{3,t}.
$$

(5.9)

In this way we remove the linear effect of the orderflow on both the price returns and the entropy imbalance. The entropy imbalance still remains significant: On the on-book market $\alpha = 0.20 \pm 0.01$, p-val = 0 and $R^2 = 3\%$. On the off-book market $\alpha = 0.039 \pm 0.008$ and p-val = 0. The on-book effect of the orderflow corrected entropy imbalance on the orderflow corrected price returns can also be seen in figure 5.4. The small points are the scatterplot of the variables while the larger points are the conditional means, binned on the x-axis.
5.4 Conclusions

We have shown that order size heterogeneity plays a role in price formation in addition to the signed order flow. Large orders which make up most of the trade volume in a given interval will incur the cost of market impact unless they transact with similarly large orders on the other side of the market. Alternatively, firms can limit the impact of large orders by trading in the off-book market. Trading in the off-book market however is not anonymous, which may be a concern in some situations.

The effect of size heterogeneity seems to be present on both daily and hourly time scales, though its economic importance on the hourly scale is very small. This large difference in order sizes can be attributed to the density of order sizes which is a fat tailed distribution with a tail exponent around 3/2 on the off-book market and 3 on the on-book market of the LSE and almost identical for the hourly and daily scale.

The fact that a large order moves the price when transacted with many small opposite orders, and does not move the price when transacted with a similarly large opposite order, is difficult to reconcile with the paradigm of information content of trades. For example, both the buy and sell large orders would have been informed if they transacted with multiple small orders (as they would move the price), but if they transact against each other they end up with no information content (as the price does not move). Did the information cancel out? It seems more likely that it is the limited liquidity that characterizes a situation of trading with multiple small orders which produces the price impact of a large order.

The interpretational differences between lack of liquidity and information content of trades is ultimately irrelevant. Neither interpretation really answers the question of why the price moves or not. In one interpretation, we sweep the question ab-initio under the rug by saying that the trades are informed or uninformed. As long as we do not have an explanation of which trades are informed beforehand, there is no way to know if a trade will or will not suffer price impact. On the other hand, in the interpretation using liquidity, we also do not provide a definitive answer as long as we have no explanation of liquidity fluctuations. However, many models, including the one tacitly implied here, have provided some insight into a possible mechanistic explanation of liquidity. Empirical behavioral finance studies may on the other hand provide insight for models of information content of trades.

In this paper, we have only touched on some aspects of the interaction of the member firms and market mechanisms in determining the price returns. Many other interesting things remain to be inves-
We have observed a curious fact that the entropy imbalance in the on-book market seems to be leading the imbalance in the off-book market. A shift in the entropy imbalance in the on-book market is substantially correlated with the shift in the imbalance in the off-book market up to three days later. This may be a signature of a firm first trying to transact its large order anonymously in the on-book market creating the imbalance there for two-three days, and then if it failed, resorting to the off-book market, where it then creates the imbalance there. This and other things seem interesting for future research.
Chapter 6

Conclusions

Unifying ideas behind the four chapters comprising this thesis are both market microstructure and an agent based view of trading in financial markets.

In the first chapter we have shown that there exist some properties of aggregate trader behavior that follow simple statistical regularities. The fact that one can describe the unconditional behaviour in limit order price placement of many rational agents as a simple statistical law is quite interesting. One could argue that the unconditional distribution of limit order prices for a single trader can be described by a distribution whose properties depend on the characteristics of the trader (size, trading horizon, strategy, risk preferences, etc.) However, it is not apriori clear why should a mixture distribution of all traders’ limit prices converge to a stable fat-tailed distribution.

We also scratched the surface of possible conditional dependencies by showing the relationship between volatility and limit order price placement. We propose the possibility that the feedback between the limit order prices and the transaction prices contribute to the effect of clustered volatility.

In the second chapter we use the zero-intelligence paradigm to investigate the extent to which market design influences the outcome of the trading proces. We assumed that the empirical orderflow was coming from a zero-intelligent trader and was describable as three unconditional statistical processes which are independent from each other: market order placement (rate; number of market orders per unit time), limit order placement (rate and price; number of limit orders per unit time and per unit price) and cancellation (rate; number of orders canceled per unit time). We then estimated the parameters of these proceses from market data. Using simple dimensional anal-
yis from Daniels et al. (2003) we predict the value of the spread, volatility and market impact from the estimated process parameters. Interestingly the predicted values were very close to the empirical values observed cross-sectionally over 11 stocks investigated. This led us to speculate that some market variables are more influenced by the constraints imposed on trading actions and market design, than by the strategic interaction of traders. In fact, the virtue of the zero-intelligence model is that it can be used as a benchmark against which one can analyse strategic interaction of traders.

In chapter three, we analysed a possible signature of strategic interaction of agents by looking at the correlations in the buying and selling periods between member firms. The trading of each firm was divided in hourly periods and assigned a +1 for each hour interval the firm was a net buyer, and a -1 for each hour interval when the firm was a net seller. Even with this very simple characterisation of a firm’s strategy, we see interesting patterns between firms. The correlation matrix of firm’s strategies is not random and is persistent in time. Firms which are correlated in one month tend to be correlated in the next one. We found that the clustering brought by the correlations in strategies separates the firms in a buying and a selling group. The number of firms in the two groups are typically very uneven, with one group being much larger than the other. Furthermore, in spite of the fact that sometimes the small group is the buying group and sometimes the selling group, we found that some firms are more often than at random classified in the smaller group.

In the final chapter, we are turning to the investigation of agent heterogeneity and its influence on the market. We contribute to the literature on return-volume relation by analysing the influence of member firm order sizes\(^1\) on the price process. We find that, in addition to the signed orderflow, the composition of the order sizes on the bid and ask sides of the market determines the price return. Larger heterogeneity of firms (e.g., large fraction of the total volume concentrated in one order) on the bid side, all other things kept equal, has a positive price impact, and vice-versa for the ask side. Said differently, large orders when transacted with multiple small orders adversely move the price. In contrast we also find that large orders when transacted with large orders, do not move the price. This seems to be at odds with the interpretation of informed vs. uninformed trades, and we propose that it may be more likely it is the lack of liquidity associated with small orders that moves the price.

---

\(^1\)Firm order size is the number of Pounds a firm trades during a day or hour, depending on the aggregation scale.


BIBLIOGRAPHY


Deze kandidaat dissertatie is een verzameling van vier papieren die zich bezighouden met kwesties in de microstructuur van de handel op de aandelenmarkten.

In hoofdstuk 2 tonen we een opvallende regelmaat in de manier waarop mensen plaats limiet orders in financiële markten. Samen-voegen van de gegevens uit 50 aandelen, tonen wij aan dat voor beide kopen en verkopen orders, de onvoorwaardelijke cumulatieve verdeling van de relatieve prijzen limiet\(^2\) vergaat ongeveer als een macht met recht exponent ongeveer \(-1.5\). Dit gedrag omspant meer dan twee decennia, varierend van een paar tikken op ongeveer 2000 teken. Wij vinden ook dat de tijd reeks van relatieve limiet prijzen tonen interessante temporele structuur, gekenmerkt door een autocorreleerachtige functie dat asymptotically vergaat als \(C(\tau) \sim \tau^{-0.4}\). Bovendien, de relatieve prijsniveau’s zijn positief gecorreleerd met en worden geleid door prijsschommelingen. Wij speculeren dat deze feedback mogelijk kan aan geclusterde volatiliteit bijdragen.

In hoofdstuk 3 draaien we onze aandacht om de markt ontwerpen. We onderzoeken van een situatie waarin de beperkingen opgelegd door de markt instellingen, kunnen domineren strategisch gedrag van agenten.

Wij maken gebruik van de LSE beperken orderportefeuille gegevens voor het testen van een eenvoudig model waarin minimaal intelligente agenten orders plaatsen voor de handel zo maar. Het model behandelt de statistische mechanica van de bestelling plaatsen, de prijsvorming en de accumulatie van het licht van vraag en aanbod in het kader van de continue dubbele veiling, en de opbrengst een-

\(^2\)We definieren de relatieve limiet prijs als het verschil tussen de maximum prijs en de beste marktprijs beschikbaar op het moment.
voudige wetten betreffende order aankomst tarieven aan de statistische eigenschappen van de markt. We testen de geldigheid van deze wetten in het verklaren van de cross-sectionele variatie voor elf voorraden. Het model verklaart 96% van de variantie van de kloof tussen de beste aankoop en verkoop van de prijzen (de verspreiding), en 76% van de variantie van de prijs verspreiding, met slechts een vrije parameter. We bestuderen ook de markt effect functie, een beschrijving van de reactie van de prijzen op de komst van nieuwe orders. De non-dimensionale cordinaten ingegeven door het model ongeveer ineenstorting gegevens uit verschillende bestanden op een curve. In dit hoofdstuk wordt getoond het bestaan van eenvoudige wetten betreffende de prijzen te bestellen stromen, en in een bredere context, omdat het suggereert dat er omstandigheden waarin het strategisch gedrag van agenten kan worden gedomineerd door andere overwegingen.

Hoofdstuk 4 analyseert correlaties in de patronen van de handel van de verschillende leden van de LSE. De verzameling van strategieën in verband met een instelling wordt bepaald door de volgorde van de tekens van de netto volume verhandeld door die instelling in intervallen uur. Met behulp van verschillende methodes kunnen we aantonen dat er significante en aanhoudende correlaties tussen de instellingen zijn. Daarnaast zijn de correlaties zijn gestructureerd in gecorreleerd en anti-gecorreleerd groepen. Clustering technieken met behulp van de correlaties als een afstand statistieken blijken een zinvolle clustering structuur met twee groepen van instellingen van de handel in tegengestelde richting.

In het laatste hoofdstuk 5 bewijzen we dat de heterogeniteit van de handel om maten speelt een rol in de vorming van de prijs, naast de ondertekende bestelbon stroming. Heterogene samenstelling van de bestelling op de maten kopen (bid) kant, tenzij evenwichtig door een even heterogene verkopen (vragen) kant van de markt, produceert een onevenwichtigheid die drijft de prijzen omhoog, en vice versa. Dit effect is ingesteld op zowel per dag en per uur tijdschema. We tonen aan dat een citaat markt ontwerp (off-boek), in tegenstelling tot een limietorder ontwerp (on-boek), helpt beperken de prijs impact van grote orders, waardoor de heterogeniteit, maar niet verwijderen. Het effect is ook beperkt in het geval de handel wordt gedaan tegen vergelijkbare grote orders onafhankelijk van de markt te ontwerpen. De heterogeniteit lijkt een gevolg van de vet-staart distributie van orde maten: voor de on-boek markt met een staart exponent gelijk is aan 3, voor de off-boek markt die gelijk is aan 3/2 (staart exponenten zijn voor de cumulatieve distributie).