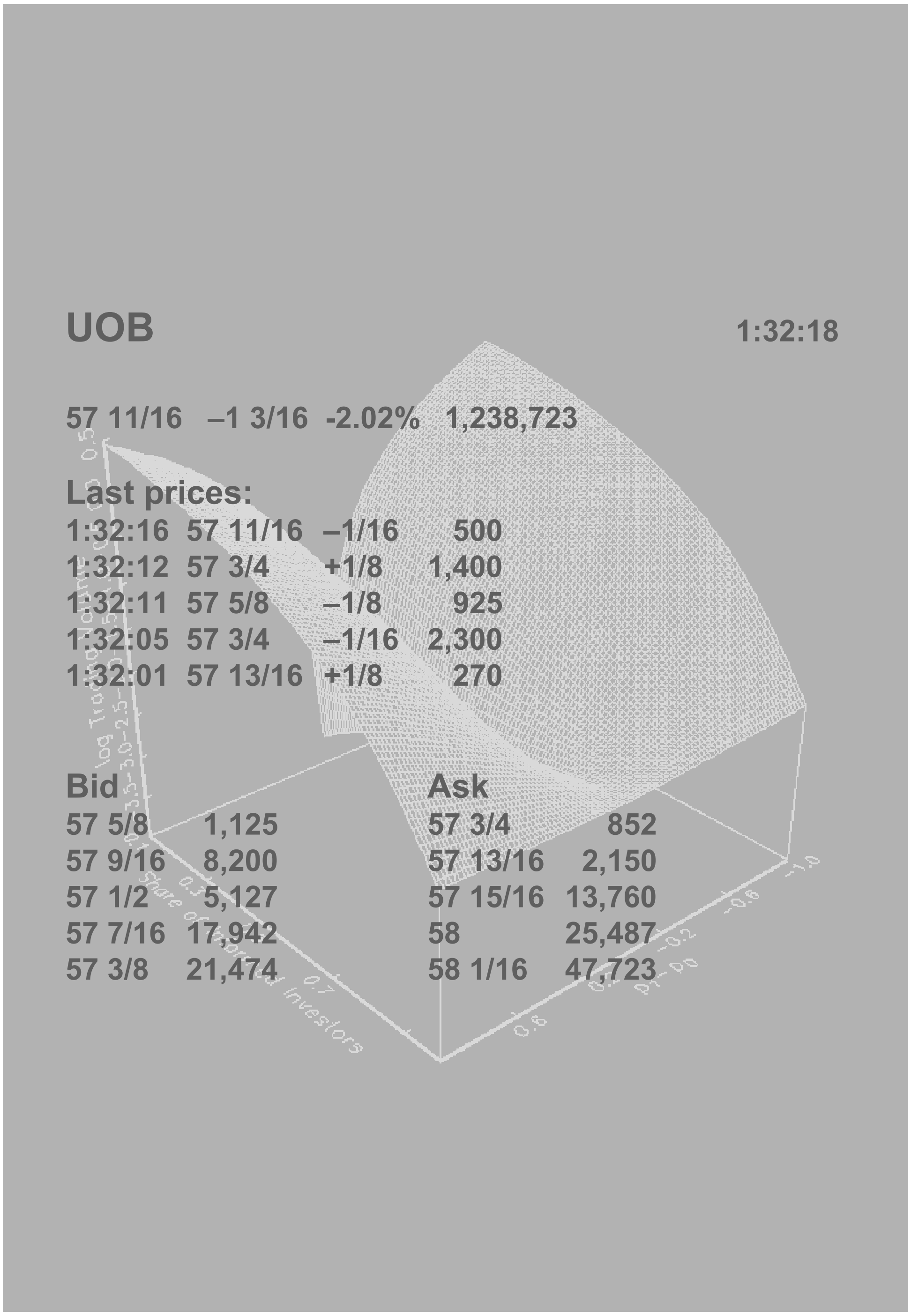
An Introduction to Market Microstructure Theory



Andreas Krause

An Introduction to Market Microstructure Theory

Andreas Krause

School of Management
University of Bath

Andreas Krause
University of Bath
School of Management
Claverton Down
Bath BA2 7AY
Great Britain

E-Mail: mnsak@bath.ac.uk

First draft
Version of June 21, 2005

Typeset using LaTeX

This is a preliminary draft of the book.
All comments are welcome.

Preface

There are three main fields in finance: asset pricing, corporate finance and market microstructure. *Asset pricing* encompasses not only the theories on the valuation of securities such as shares, corporate bonds, foreign exchange or derivatives, but also how prices may systematically deviate from these valuations. The field of *corporate finance* investigates financial decisions of companies such as investments, capital structure and dividend payouts. Finally *market microstructure* analyzes the effect trading rules have on market prices, it thus deals with the institutional setting of the trading process.

It is now widely acknowledged among academics as well as practitioners that the rules governing the trading on exchanges can have a significant impact on the success of a market. It is thus of importance to assess the trading rules of an exchange and how they affect the behavior of investors, e.g. through implicit and explicit trading costs. In addition competition between market participants as well as different market forms has to be taken into account. In traditional neoclassical models the trading process is assumed to be frictionless and instantaneous. As a consequence of this assumption the way trades are conducted in markets is not relevant for the outcome. Once this assumption is lifted, however, the trading rules become very much relevant for the outcome.

Over the last three decades models have been developed to investigate the impact trading rules have on market participants. This line of research has become known as *market microstructure theory*.[1]

[1] The term *market microstructure* has first been introduced by Garman (1976). According to Easley and O'Hara (1995, p. 357) it is

> "the study of the process and outcomes of exchanging assets under explicit trading rules."

Market microstructure theories are not interested in the fundamental value of the assets traded on an exchange, which is assumed to be determined exogenously, but to model the *price formation* in asset markets with a given fundamental value and how this is affected by the trading rules.

After some pioneering work by Demsetz (1968) and Bagehot (1971) the first models developed by Stoll (1978) focused on dealer markets and inventory effects that influence the price setting of market makers. Starting with Copeland and Galai (1983) for dealer markets and Kyle (1985) for auction markets, the attention changed from inventory effects to that of different information between market participants. These models dominate the literature since this time, addressing important questions like informational efficiency of prices and market liquidity. More recently increased attention has been paid to limit order markets and the order submission strategies of traders therein.

Structure of the book

The aim of this book is to provide an introduction to the main models used in market microstructure theory upon which much of the more advanced literature relies. A common topic throughout this text will be the presence of asymmetric information between market participants and other direct and indirect trading costs.

We will start in **chapter 1** with an overview of the trading rules commonly found in stock markets before the following three chapters then cover different market forms. **Chapter 2** focuses on auction markets and the effect asymmetric information between traders has on the evolution of the market price. Theories on dealer markets in **chapter 3** deal with the price setting behavior of market makers with and without the presence of asymmetric information between them and traders. In **chapter 4** limit order markets are considered and the optimal strategy for investors to submit limit orders is discussed; where models usually abstract from the presence of asymmetric information for simplicity. These theoretical chapters bring up many questions which can be investigated empirically, such as the relative importance of different aspects for the spread, informed trading or the relevance of market microstructure elements for daily asset returns. **Chapter 5** provides an overview of the key empirical methods used in market microstructure analysis.

While in most cases the market form is taken as given, the differences between them can be significant such that the choice of the market form is potentially important for the prices of traded assets. How the main market forms may evolve endogenously is modeled in **chapter 6**. The final **chapter 7** shows the importance of market microstructure for the prices of assets by linking its outcomes to asset pricing theory. An **outlook** towards more advanced and specialized topics in market microstructure concludes the main body of the text.

Besides a detailed overview of the regulation of the NASDAQ Stock Market in **appendix A**, the appendix provides a brief introduction to important mathematical methods in **appendix B** and key economic and financial concepts in **appendix C**. The reader unfamiliar with the presented ideas might wish to consult them prior to working through the main part of the text.

Throughout this book the emphasis is laid on theoretical models rather than empirical methods to estimate them. Readers more interested in empirical methods are referred to the original articles, many of which are mentioned throughout the book and which in many cases provide more details on these aspects. As many derivations of models are presented in more detail than the original articles, it allows the reader to fully appreciate the mathematical foundations of these models. It has however to be emphasized that it is of more relevance to understand the reasoning behind a result rather than its mathematical derivation.

Use of the book

This book is aimed at taught postgraduate students specializing in finance, although advanced undergraduate students might find much of the material accessible. It will also be a valuable introduction for beginning PhD students as a help in their exploration of the field. The detailed mathematical derivation of the models will provide them also with insights towards the problem of solving often complex models. Taught postgraduate students and their teachers will also find this feature of the book useful as it allows the class discussion to focus more on the interpretation of results rather than their derivation. From chapter 1 onwards the key readings at the beginning of each chapter represent the most important original articles in which the presented models have been developed.

In order to successfully use this book readers should be familiar with basic microeconomic theories as well as key financial concepts. Although appendix C provides a short introduction to these theories, readers without such knowledge are advised to consult basic books on microeconomics and finance in parallel. It is furthermore required that readers have been introduced to differentiation and integration as well as matrix algebra and stochastics. Although some slightly more advanced concepts are covered in appendix B, it does not serve as a substitute for such a background.

Each chapter features a set of ten **review questions** of varying difficulty which allow the reader to his knowledge. Although all these questions address problems in the respective chapters, some require additional reading and research before they can be answered properly. The aim of these questions is to enable the reader to test his own knowledge on the topics learnt. Chapters 2, 3, 4, 6 and 7 also feature an **application**, where the applications in chapters 2 to the first application in chapter 6 all build on each other. These applications require the reader to combine different and sometimes contrasting theories to solve a given problem. In most cases there will not be a single 'correct' answer but different foci will lend themselves towards producing very different conclusions. Much can be gained from discussing the applications in small groups to explore a range of opinions thoroughly using several market microstructure models where appropriate.

An appeal for help

The copy of the book you have in front of you is a first draft. Inevitably there will be a large number of spelling mistakes, grammatical errors and much more importantly mistakes in formulae, be it in forms of wrong indices, wrong symbols, missing brackets or even more serious mistakes. You can also expect the one or other cross-reference to be incorrect or references not being up-to-date.

Even more annoyingly to the reader than all these errors, some explanations given might not be as clear as they could be or in the worst case might even be wrong due some mix up during writing. You might also find that some topics are under-represented or over-represented and you would like to see a change.[2]

Please feel free to send me any comments you have, ideally with a suggestion for improvement but it is often also of value just to point out that you think there might be problem although you do not have a solution at hand. Sending me your comments and thoughts on this draft is an easy way to enter the hall of fame, aka getting your name mentioned in the acknowledgements in the final version.

Thank you very much for your help to improve this book.

June 2005 A.K.

Bath

[2] You also may just think that I should cite your paper when discussing an aspect of the theory.

Contents

Preface i

Contents vi

List of Figures x

List of Tables xiii

1. The organization of trading in stock markets 1
- 1.1 Definitions 2
- 1.2 The importance of organized stock markets 5
- 1.3 The different market forms 9
- 1.4 Market participants and order submission 17
- 1.5 Trading priority rules 20
- 1.6 Electronic trading mechanisms 22
- 1.7 Recent developments 25
- Review questions 28

2. Auction Markets 29
- 2.1 Auctions with informed investors and noise traders 31
  - 2.1.1 Auctions with a single informed investor 31
  - 2.1.2 Auctions with multiple informed investors 41
- 2.2 Auctions with strategic uninformed investors 46
  - 2.2.1 Markets with a single asset 46
  - 2.2.2 Auction markets with multiple assets 59
- 2.3 The informational content of trading volume 64
- 2.4 Explaining short-term movements of asset prices 70
- Review questions 72
- Application 72

3. Dealer Markets 75
- 3.1 Inventory-Based Models of Market Making 76
  - 3.1.1 The costs of market making 79
  - 3.1.2 Competitive price setting 87
  - 3.1.3 The price setting of a monopolistic market maker 93
- 3.2 Information-Based Models of Market Making 99
  - 3.2.1 Determination of adverse selection costs 102

3.2.2 Simultaneous trading at different stock exchanges 111
3.2.3 Comparing competitive and monopolistic market makers . 112
3.2.4 Explaining return patterns 114
3.3 Dealer markets with multiple assets 116
Review questions 122
Application 122

4. Limit Order Markets 125
4.1 Static models of limit order placement 127
4.2 Price dynamics with limit order trading 130
4.3 Dynamic models of limit order submissions 137
4.4 Informational efficiency with limit order trading 150
Review questions 152
Application 152

5. Empirical investigations of market microstructure models . . . 153
5.1 Estimating spread components 154
5.2 Estimating informed trading 161
5.3 Properties of daily returns 164
Review questions 167

6. The emergence of different market forms 169
6.1 Demand and Supply of Immediacy 171
6.2 Determination of the optimal market form 178
Review questions 190
Application 190

7. Asset Pricing with Trading Frictions 193
7.1 Required returns in the presence of bid-ask spreads 194
7.2 Compensation for adverse selection costs 198
7.3 The price impact of trading 202
7.4 Pricing liquidity risk 205
7.5 The equity premium puzzle 208
Review questions 211
Application 211

Outlook 213

Appendices 218

A. The NASDAQ Stock Market 221
A.1 History of the NASDAQ Stock Market 222
A.2 The organization of the NASD 230
A.3 Listing requirements 232
A.4 Registration as broker and market maker 235
A.5 Trading rules 236

A.6 Summary 240

B. Mathematical methods 243
B.1 Taylor series expansion 243
B.2 Dynamic programming 244
B.3 Constrained optimization 245
B.4 Conditional moments 247
B.4.1 Truncated distributions 247
B.4.2 Joint distributions 248
B.5 Stochastic processes 249
B.5.1 Definition 249
B.5.2 Itô's Lemma 250

C. Economic concepts 251
C.1 Utility theory 251
C.1.1 The expected utility hypothesis 251
C.1.2 Risk aversion 255
C.2 Portfolio selection theory 257
C.2.1 The mean-variance criterion 258
C.2.2 The Markowitz frontier 264
C.3 The Capital Asset Pricing Model 274
C.4 The Rational Expectations Approach 279
C.5 Bayesian learning 282

Bibliography 284

Author index 295

Index 298

List of Figures

1.1 Classification of market forms 11
1.2 Order flow in different market forms 18

2.1 Liquidity in the Kyle (1985) model 42
2.2 Liquidity with different number of auctions 47
2.3 Informativeness of prices with different number of auctions 47
2.4 Liquidity with different number of informed investors 48
2.5 Informativeness of prices with different number of informed investors 48
2.6 Market liquidity with varying number of uninformed investors . . 57
2.7 Market liquidity with varying number of informed investors 57
2.8 Market liquidity with different uncertainties about the liquidation value 58
2.9 Trading volume with different precision of information and changes of beliefs 69
2.10 Trading volume with different shares of informed investors and changes of beliefs 69

3.1 Demand and Supply in Dealer Markets 78
3.2 Costs of a market maker 80
3.3 Competitive price setting 87

4.1 The placement of limit orders 131
4.2 Order submission strategies 135
4.3 Influence of the trader composition and trade frequency on the spread 144
4.4 Influence of the trader composition and trade frequency on the waiting time 145
4.5 Influence of the trader composition and trade frequency on the duration 145
4.6 Influence of the tick size and waiting costs on the spread with $\theta = 0.7$ 146
4.7 Influence of the tick size and waiting costs on the waiting time with $\theta = 0.7$ 147
4.8 Influence of the tick size and waiting costs on the duration with $\theta = 0.7$ 147
4.9 Influence of the tick size and waiting costs on the spread with $\theta = 0.3$ 148
4.10 Influence of the tick size and waiting costs on the waiting time with $\theta = 0.3$ 148

4.11 Influence of the tick size and waiting costs on the duration with $\theta = 0.3$ . . . 149

5.1 Sequences of transaction prices with the initial price at the bid . . 157
5.2 Evolution of the order flow with informed trading . . . 162

A.1 Distribution of the number of market makers for NASDAQ securities in 1996 *Data: NASDAQ Factbook 1997* . . . 222
A.2 Daily US-Dollar trading volume on the NYSE and the NASDAQ . 228
A.3 Daily share trading volume on the NYSE and the NASDAQ . . . 229
A.4 Development of the Dow Jones Industrial Average and NASDAQ Composite Index *Data: Datastream* . . . 229
A.5 Organization of the NASD . . . 230

C.1 The Arrow-Pratt measure of risk aversion . . . 257
C.2 The mean-variance criterion . . . 259
C.3 The efficient frontier . . . 260
C.4 Determination of the optimal alternative . . . 263
C.5 Efficient portfolios with two assets . . . 267
C.6 Determination of the optimal portfolio with two assets . . . 268
C.7 Portfolio selection with three assets . . . 269
C.8 The optimal portfolio with $N > 2$ assets . . . 269
C.9 The optimal portfolio with a riskless asset . . . 271
C.10 Portfolio selection with short sales . . . 272
C.11 The Capital Asset Pricing Model . . . 278

List of Tables

1.1 Market forms of selected stock exchanges 14
1.2 Overview of alliances between major stock and futures exchanges 26

5.1 Parameter constellations for the spread components 157
5.2 Estimates of the spread components 160

A.1 Main historical events of the NASDAQ 227

Chapter 1

The organization of trading in stock markets

The main aim of this chapter is to introduce the reader to

- the different ways assets can be traded in markets
- the different types of orders that traders can use in asset markets
- the way orders are filled in asset markets
- recent developments in stock market trading

Key readings:

Ruben Lee: *What is an Exchange? The Automation, Management and Regulation of Financial Markets*, Oxford: Oxford University Press, 1998

When investigating the market microstructure it is essential to be aware about the way prices are determined during trading in real markets and the incentives of individuals to submit their orders. Although rules can vary significantly across markets, we can derive some general properties upon which markets base their actual rules.

The aim of this chapter is not to give a survey of all possible regulations and special forms found at different stock exchanges, but to explain the basic principles.[1] A special focus is laid on those aspects that have an impact on the price setting in these markets.[2]

After some preliminary definitions in chapter 1.1, the importance of stock markets for investors, companies and the society as a whole will be pointed out in chapter 1.2; chapter 1.3 describes the different forms of trading on an exchange and chapter 1.4 shows how orders are submitted to an exchange, who participates in trading and what forms of orders exist; according to which rules orders are executed is explained in chapter 1.5; chapter 1.6 deals with the influence of computerization on the organization of stock markets and trading; finally, chapter 1.7 points out some recent developments in financial markets.

1.1 Definitions

A *market* is a place where supply and demand for a good meet, i.e. potential sellers and buyers of a good submit their supply and demand schedules. This place has not to be a certain location,

> "... but the entire territory of which the parts are united by the relations of unrestricted commerce that prices there take the same level throughout with ease and rapidity."[3]

[1] Lee (1998) gives a very detailed and theoretically based overview of the current regulation of stock markets in the United States, including recent developments in electronic trading.

[2] An overview of the regulation of stock markets is given in Schwartz (1993, ch. 2 and 4) for the NYSE and the London Stock Exchange. A detailed description of the regulations in the United Kingdom, United States, Switzerland, France, Netherlands, Austria, and Japan can be found in Hopt et al. (1997, Part III) and in Hopt and Baum (1997) for Germany. A history of stock exchange regulations in Europe is given in Merkt (1997).

[3] Cournot (1838, p. 42).

To achieve this tendency towards an equal price, JEVONS (1911) emphasizes the importance of communication between market participants, i.e. the transmission of information, especially prices:

> "The traders may be spread over a whole town, or region of country, and yet make a market, if they are (...) in close communication with each other."[4]

MARSHALL states the characteristics a good must have to be traded in a market:[5]

- The good has to be *widely demanded*. A good that is not widely demanded cannot be traded frequently, hence there is only infrequent intercourse and too few prices that are needed to form a market as stated in the above definition of Cournot.
- The good has to be *homogeneous*.[6] If a good is not homogeneous, e.g. pieces of art or special machines, the pieces are not interchangeable with each other, hence they will fail to have a tendency towards equal prices as needed by the definition of a market.
- The good has to be *transferable* and *storable* at low costs, compared to its value. High costs of transferring the rights or storing the good would hinder the free intercourse wanted by Cournot to form a market. The gains that have to be made from trading the good to offset these costs would be too high.

Technological innovations reduced the costs of transfer and storage significantly in recent years, thereby increasing the number of goods that can be traded in markets. New information technologies enable traders to communicate at low costs over long distances, enlarging the region that can be viewed as a market.

[4] JEVONS (1911, p. 81).
[5] See MARSHALL (1920, pp. 141f.).
[6] Goods are homogeneous if they are interchangeable, i.e. one piece of a good has exactly the same characteristics as another.

Further improvements in standardization made more goods homogeneous, so that they can be traded in markets nowadays. The reduction in costs and standardization, besides further improvements, also increased the demand for many goods. Hence today more and more goods can be and are traded in markets.

Securities are rights that are chartered in a document. To execute or transfer this right the document has to be presented. *Financial securities* are securities, where the rights are a sequence of future cash flows. The future cash flow can consist of money (interest, dividends) or other financial securities (e.g. stocks or bonds). Most securities fit into one of the following categories: stocks, bonds, derivatives[7] or a combination of these. When addressing securities in economics, financial securities are principally being referred to. We will stick to this tradition in the remaining parts of this chapter.[8]

Securities are typically divided into smaller parts, each part representing a fraction of the whole, hence securities are homogeneous. They are transferable at very low costs and storage costs can be neglected.[9] As securities are also widely demanded as a means for investments, they fulfill all characteristics to be traded in a market.[10] These markets are called *security markets*.[11]

We have to distinguish two forms of security markets: primary and secondary markets. In *primary markets* new shares of a security are issued, i.e. for the first time sold to an investor,[12] while in *secondary markets* already issued shares of securities are traded among investors.

[7] A derivative is a security whose future cash flow depends on the value of another security.

[8] Another frequently used expression for financial securities is *asset*. In relation with prices and valuation this term is the most frequently used, while in relation with market regulations the term securities is more common. Therefore, starting in chapter 2, we will use the expression *asset* for securities.

[9] Many securities do not exist physically nowadays and if they do, they are stored at a clearing house, imposing very low storage costs. Formerly securities had to be handed over, but today the transfer usually is only booked into accounts kept at the clearing house.

[10] There are sometimes securities issued that are not widely demanded. They typically vanish after a short period of time.

[11] Often the term *financial markets* also refers only to security markets, although they in general encompass also rights that are no securities, such as bank loans or deposits. Often foreign exchange markets and even the commodity markets are included into the term financial markets, as the characteristics of these markets are very similar.

[12] We call investors all those, who are invested in securities or are interested in being so (potential investors).

If the amount of an existing security is increased, the new shares can be issued either by organizing a separate auction or by selling them in the open market, i.e. they are issued by placing a sell order from the issuer in the secondary market. In this case the issuer behaves like an investor. Although this form of issuing formally belongs to the primary market, it is referred to as an operation in the secondary market. Sometimes this form of increasing the amount of an existing security is also applied for issuing a new security, especially in OTC markets for derivatives.

In futures and options exchanges there does not exist a secondary market. All transactions take place in the primary market according to the definition above. If an investor buys a derivative, another investor has to issue a new unit of this security, an existing derivative cannot be bought. The outstanding amount of these derivatives is not fixed as in the case of stocks or bonds. If an investor wants to sell a derivative he has bought, he issues a derivative which exactly offsets the derivative he wants to sell.[13] Formally, all these operations must be placed into the category of primary markets. But since they have all the characteristics of a typical trading activity, they are assigned to secondary markets.

In most cases, security markets are being referred to the category of secondary markets. This convention will also be applied in this text. The remaining analysis will be concentrating on stock markets, but most concepts and findings can easily be adapted to other security markets.

1.2 The importance of organized stock markets

In most cases the time horizons of investors and the issuer of a security do not coincide. While companies have a need for capital of very long, even infinite

[13] For every derivative such an offsetting derivative exists because it can be found in two forms: as a long and as a short position. An investor is long if he has bought a security, he is short if has issued the security. In stock and bond markets investors usually are long and the company who has issued the security is short. However, there is the possibility of short sales by investors in many markets, so that only on average investors have to hold a long position. Adding a short and a long position of the same security exactly offsets the investor.

disposability, investors on the other hand may want to change their investments to adjust for new information, changed tastes, or liquidity needs. As the amount of outstanding shares is fixed (disregarding periodic increases and repurchases of capital) all shares have to be held by investors. To adjust their investments, investors have to trade them with each other. KEYNES (1936, p. 151) pointed this aspect out as follows:

> "In the absence of security markets, there is no object in frequently attempting to revalue an investment to which we are committed. But the Stock Exchange revalues many investments every day and the revaluations give a frequent opportunity to the individual (though not the community as a whole) to revise his commitments."

If no securities markets exist, it is very difficult to find another investor who wants to take the counterpart in a trade. The lack of communication between investors makes the search for a counterpart very costly.[14] When a counterpart is finally found (e.g. by advertising) the next problem is to find a price at which both are willing to trade. The valuation of a stock depends on the information an investor has, inconsistent information will make it difficult to find a suitable price.[15] Even if both can agree on a price, the trade may occur at a "wrong", i.e. informationally inefficient, price[16] when neither participant has precise information. This gives wrong incentives to investors, resulting in an inefficient allocation of resources.

To overcome these difficulties markets provide investors with two services: liquidity and the aggregation and revelation of information.

[14] Chapter 6.2 presents a formal model showing how these costs can give rise to the formation of organized markets.

[15] An overview of asset valuation can be found in nearly all textbooks on finance. Good synopsis are INGERSOLL (1987), DUFFIE (1996) and COCHRANE (2000), besides others.

[16] Prices are called *informationally efficient* if they fully reflect all available information. FAMA (1970) distinguishes three forms of efficiency, depending on the information available: weak, semi-strong and strong efficiency. Prices are *weakly efficient* if they reflect only information derived from previous prices or returns; if all publicly available information is reflected in the price, it is called *semi-strong efficient*. In cases where also private information available to any market participant is reflected, prices exhibit *strong efficiency*.

By providing *liquidity*, markets facilitate the exchange of assets between investors. As investors meet on markets, the costs of searching a counterpart for a trade is reduced significantly, a counterpart can easily be found by addressing the market. Competition between investors for a trade will further ensure that a better, informationally more efficient price will be charged. These reduced costs of trading will allow investors to adjust their investment decisions more frequently to their information and tastes, resulting in a more efficient allocation of resources.[17]

As markets generate prices, they can be used to reveal and *aggregate information* on a security.[18] Compared to other sources of information, prices can be observed at nearly no costs. Without having much additional costs, an investor can increase his information and in this way reduce the risk to trade at a disadvantageous price resulting from a lack of information. This further reduces his costs of trading.

By holding a portfolio that fits better his tastes and information, an investor reaches a higher level of utility. The reduced costs of adjusting his investment decisions increase his returns and hence the price he is willing to pay for an asset. This benefits also the issuers of assets as they can issue their assets at higher prices, reducing their costs of capital and increasing profits.[19] Increased profits give incentives for further investments and hence promote economic growth.

From the existence of security markets investors profit from higher returns and a better allocated portfolio, issuers of securities from lower costs of capital and higher profits and the society as a whole from a more efficient allocation of resources, higher investments and growth. Therefore everyone benefits from the existence of securities markets.

[17] Arrow (1964) stresses the importance of asset markets for an efficient allocation of risks between individuals.

[18] It was Hayek (1945) to point out the importance of prices as a source of information. How prices can aggregate and reveal information to investors is discussed in more detail in chapters 2 and 3.2.1.

[19] See Keynes (1930, Vol. II, p. 195).

A further reduction in the costs of trading can be achieved by applying uniform rules to the trading process. Security markets that apply a fixed set of rules governing trading are called *organized* security markets or an *exchange*.[20] The rules of an exchange should regulate[21]

The *admission* of stocks. The stocks to be traded on an exchange have to meet certain standards in order to guarantee a minimum of investor protection and ensure regular trading. These standards differ very much between exchanges.

The *access* to the market. It has to be determined who is allowed to trade directly in the market and what conditions have to be met for this access. By applying certain standards, the risk of a counterpart failing to fulfill his duties after a trade has been negotiated, can be reduced.

The *types of orders* that can be submitted. A standardization of the order types simplifies the trading process and the set of rules can be held small.

The *execution* of orders. Rules on the execution of an order include the rules to determine the prices at which a trade occurs and when trades are executed. These rules can avoid the problem of some investors gaining the advantage at the cost of others, e.g. as a consequence of personal links to other market participants.

The *clearing* of trades. A standardization of the clearing process, i.e. the way and time trades are settled, avoids a separate negotiation on this point. Most exchanges do not only set rules for the clearing process, but organize it by themselves through special clearing houses.

[20] A more formal definition of an exchange, based on the current legislation can be found in Lee (1998). Rudolph and Röhrl (1997, p. 168) also apply a similar definition of an exchange and point out that recent innovations, especially the computerization of exchanges (see also chapter 1.6), made it necessary to adapt the traditional legal definition of an exchange.

[21] Appendix A describes thee regulation of the NASDAQ Stock Market in more detail as an example of how these principles are applied in reality.

Besides the reduction of trading costs, the standardization of trading also makes prices more comparable to each other, both over time and between assets. The prices therewith can better aggregate and reveal information.

The rules imposed on the trading process are called the *market structure*.[22] They can be imposed by public regulation or self-regulation of the market. In most cases a combination of public regulation and self-regulation can be found.

The structure of an exchange will influence the costs of trading, which in turn will affect the price formation. We will therefore investigate the process of price formation in more detail in this chapter by giving a short overview of the basic market forms.

1.3 The different market forms

The *market form* describes the way trades occur in a market. In standard neoclassical theory the market form implicitly assumed is that of an *Walrasian auctioneer*. The auctioneer suggests a price to the investors. The investors[23] determine which amount they are willing to buy or to sell at the stated price and submit their decisions, called *orders*, to the auctioneer. If demand and supply do not exactly equal, no trade occurs. The auctioneer suggests a new price and the investors revise their decisions. This process continues until aggregate demand exactly equals aggregate supply. The price at which aggregate demand and supply are balanced, is called the *equilibrium price* or *market clearing price*. If the equilibrium price is found, all orders are executed in a single multilateral trade.

This process mostly is assumed to be finished in an instant of time, imposing no costs on investors. In reality, however, the revision of orders and the announcement of a new price would take considerable time and impose high costs, especially if many investors are trading on the market. Furthermore, it would be pure incident if a price could be found that clears the market exactly. Both,

[22] See O'Hara (1995, p. 1).

[23] In this section use the term *investor* for all market participants and assume that they directly interact at the exchange. However, in reality most investors have to use an agent for trading. This part of the structure of an exchange will further be discussed in the following sections.

prices and quantities are discrete, either by definition as no fractions of a stock can be traded or by convention, e.g. discrete prices. Additional rules have to be applied to determine the price with these small imbalances.

For these reasons a Walrasian auctioneer seldom exists in real markets, other market forms have been established over time.[24] There exists a wide variety of market forms around the globe, every exchange has its unique way of executing a trade. Also within an exchange there often exists different market segments, each having its own rules.[25] Despite these differences in detail, the market forms can be organized in one of six main market forms. Figure 1.1 shows the classification of market forms as will be used below.

In all market forms orders[26] can be submitted to the market at any time, what differs is the time and way these orders are executed. A market form which is very close to the concept of an Walrasian auctioneer is the *batch system* or (*periodic*) *call market*. In batch systems incoming orders are not executed immediately, but stored and executed in a multilateral trade at a predetermined point of time. The price that will be applied for this trade, is the price at which most orders can be executed, i.e. the price with the highest trading volume. We find batch systems in two forms: à la criée and par cassier.

Trading *à la criée* allows investors to revise their orders until the time of execution. To give additional information for the revision of orders, the price that would be applied if all orders were to be executed immediately, is continuously published.[27]

[24] The only examples of markets that use the concept of a Walrasian auctioneer are the London Gold Fixings at 10.30 am and 3 pm and the Frankfurt Foreign Exchange Fixing for selected currencies at 12 am that has been ceased to exist after December 30, 1998 with the introduction of the EURO.

[25] Different rules can often be found for frequently and infrequently traded stocks. The Frankfurt Stock Exchange with its official quotation (Amtliche Notierung), regulated unofficial market (Freiverkehr) and Neuer Markt is a good example for an exchange with different market segments applying different sets of rules.

[26] We find two forms of orders: orders that specify the worst price (highest price for a buy order, lowest price for a sell order) at which they can be executed (limit orders) and orders to be executed at any price (market orders). These order forms are presented in more detail in section 1.3.

[27] The term *à la criée* refers to the verbal order submission to the auctioneer that has been used at the Paris Stock Exchange prior to the introduction of electronic trading in 1986. How-

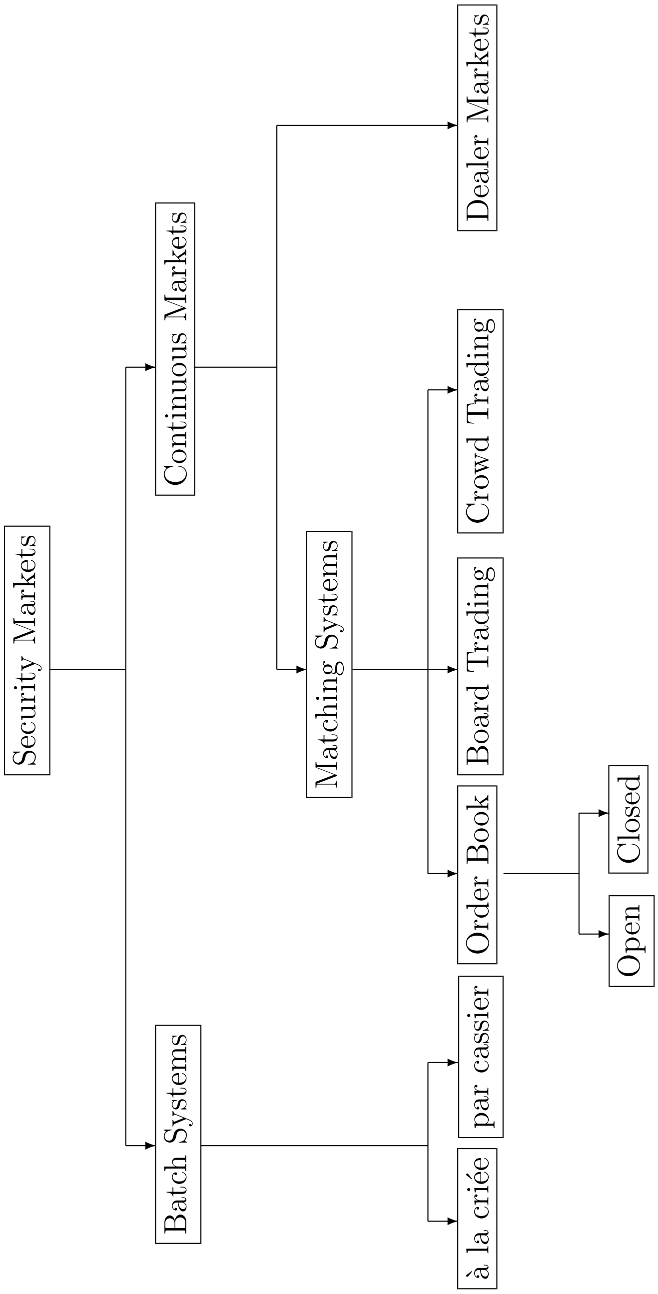


Fig. 1.1: Classification of market forms

Trading *par cassier* does not allow investors to revise their orders and typically the price that would be applied in case of immediate execution of all orders is not published.[28]

Batch trading can rarely be found in regular markets. It is only frequently used to determine the opening and sometimes the closing price of a trading day and to determine the price of some infrequently traded stocks.[29]

In *continuous markets* trades can occur not only at predetermined points of time, as in batch trading, but at any time two orders can be executed. For every submitted order it is immediately checked whether there exists another order in the market, such that these orders can be executed in a bilateral trade. If no such order exists, the order is stored and executed with the next matching order arriving in the market. We find two forms of continuous markets: dealer markets and matching systems.

In *dealer markets* special market participants, called *dealers*, have the obligation to "make the market", hence they are also called *market makers*. Every market maker has publicly to set prices at which he is willing to sell (*ask price*) and to buy (*bid price*) the security.[30,31] The bid and the ask prices have not to be, and will not be, equal, as we will see in chapter 3. At the stated price the market maker has to sell (buy) the security immediately from (to) any investor demanding this. The market maker trades on his own account, i.e. he forms the counterpart of the investor.

In *matching systems* no special market participants exist to form the counterpart by trading on their own account. The trades are only bilaterally executed between two investors. Three different forms of matching systems are known: order book systems, board trading and crowd trading.

ever, it is not relevant whether the order is submitted verbal or written, the important feature of this market form is the publication of the price and the possibility to revise orders. The name is only kept for historical reasons.

[28] Like à la criée the term *par cassier* is kept for historical reasons, as in this market form orders were submitted written. Important is only the impossibility to revise orders.

[29] See also table 1.1.

[30] When setting these prices, the market maker does not know whether the next order arriving at the market is a buy or a sell order, what its size is and when an order will arrive.

[31] The prices a market maker sets are also called *quotes* or *quoted prices*.

In *order book systems* all submitted orders are stored in an order book. If two orders cross, they are immediately executed and the price and volume of the trade are published. In some cases the order book keeper also acts as a market maker, like on the NYSE. A further distinction can be made whether the order book is *open* or *closed* to the public, i.e. if the investors can look into the order book or not. Mixed forms can also be found where the order book keeper can give some information, as on the Frankfurt Stock Exchange.

In *board trading* the prices at which investors are willing to trade are also entered into an order book, but only the best (highest) bid and the best (lowest) ask prices are published on the board, e.g. the two best prices on the Hong Kong Stock Exchange. If an order arrives in the market and accepts the best price stated on the board, they are immediately executed. The price at which the trade takes place is published on a separate board and the executed orders are cancelled from the order book. In this system the trade size is fixed to what is called a *lot*. This facilitates trading because the size of an order has not to be considered in matching them.

The third form of matching systems is *crowd trading*. Investors meet on the trading floor of the exchange and discuss the prices at which they are willing to conduct trades. If two investors agree upon a trade, their orders are executed and the price is published.

In another classification matching and batch systems are also called *auction markets*, to distinguish in this classification between matching and batch systems, matching systems are called *continuous auctions*.

As can be seen from table 1.1 every market form can be found on least at one of the leading stock exchanges.[32] There is no dominating market form to be found, what suggests that every market form has its advantages and disadvantages, although currently we observe a tendency towards order book systems.

[32] Only crowd trading cannot be found since the Swiss Exchange Zürich changed to an electronic trading platform in 1996. Of the more important exchanges nowadays only the London Metal Exchange applies crowd trading, but it is also planned to introduce electronic trading in the near future and hence change the market structure.

	Batch Systems		Continuous Markets			
			Matching Systems			
	à la criée	par cassier	Order Book	Board Trading	Crowd Trading	Dealer Markets
Frankfurt Stock Exchange						
Amtliche Notierung		opening	closed			
Freiverkehr		X				
Neuer Markt			closed			(X)
Hong Kong Stock Exchange				X		
London Stock Exchange						X
NASDAQ						X
NYSE		opening	closed			X
Tokyo Stock Exchange	opening		open			
Swiss Exchange Zürich			X			
Paris Stock Exchange			X			
Vienna Stock Exchange						X

Tab. 1.1: Market forms of selected stock exchanges

Batch systems have the advantage of collecting orders over a longer period of time. Large order imbalances that may occur over time, e.g. a large order arriving in the market, will have a smaller effect on prices than with an immediate execution of the order. Often an imbalance is reduced over time and the volatility of prices diminishes. If the trading is à la criée, investors can react to this imbalance as they can observe that the price would change significantly with execution. On the other hand, to revise an order imposes not only costs on investors, but also on the exchange. The order flow becomes difficult to handle and errors are more likely to occur than in trading par cassier. On many stock exchanges with matching systems all orders accumulated over night are cleared immediately at the beginning of the trading session in one multilateral trade at a single price. This is advantageous for the determination of opening prices. If they had to be executed in subsequent bilateral trades, these orders would hinder the handling of orders submitted at the beginning of the trading session.[33]

On the other hand, batch systems have the disadvantage that trades occur only a few times per day (once or twice). Investors have to wait a considerable time until they are able to trade the next time. A reaction to new information is not possible immediately, imposing waiting costs on investors. As less prices are available to investors, the aggregation and revelation of information through the price system cannot be assured as good as in continuous markets.[34]

The advantage of faster execution of an order and therewith reduced waiting costs is one of the main arguments for continuous markets. In dealer markets the market maker guarantees immediate execution of an order, but, as we will see in chapter 3, he will not provide this service for free. The fees charged by the market maker may counteract the advantages of this market form.

Matching systems impose no additional fees on investors[35], but therefore im-

[33] The NASDAQ has no call auction at the opening, hence trading volume is very high and the execution of orders submitted at the beginning of the trading hours takes a considerable time. Therefore they are currently considering to introduce a call auction at the opening.

[34] With trading à la criée prices are published continuously, but they will be biased up to short before the execution. In order to save costs, investors will not adjust their orders permanently, but only once just prior to the execution.

[35] We neglect here for simplicity any direct fees levied by exchanges, brokers and any taxes

mediate execution is not guaranteed as a matching order has to be found before an order can be executed, hence investors may also face waiting costs. Especially for infrequently traded stocks, the time until an order can be executed may be very long. This makes the differences to batch systems in terms of waiting costs less important, whereas the volatility of prices will in general be higher, as an order may have a substantial effect on prices. This favors batch systems for infrequently traded stocks, whereas for frequently traded stocks the impact of not too large orders on the price can be neglected, favoring matching systems as the result of more frequent trading.

Even for frequently traded stocks, where no considerable waiting costs exist, matching systems have the disadvantage that the price at which the trade will be executed, is not known in advance. It depends also on the order with which it is matched, whereas in dealer markets the investor knows the stated price of the market maker at which he will trade. This uncertainty on the price may favor dealer markets despite the fee market makers charge.

Crowd systems allow only for small trading volumes. The time to negotiate the price does not allow for too many investors trading. But this system ensures the best price in continuous markets, as all orders compete directly for a trade. Board trading is capable of handling larger trading volumes and order book systems are more flexible to handle orders of different sizes. With the ability of handling larger volumes due to advanced computerization, the costs of conducting a trade decreases for the exchange and hence for investors.

As can be seen from the above discussion, every market form has its advantages and disadvantages. The optimal market form depends on the nature of the stock, e.g. the frequency it is traded, and the personal tastes of the investors, e.g. their risk aversion or time preferences. A discussion of these aspects can be found in Angel (1997), where he suggests optimal trading rules for stocks of small companies.

In addition to the market forms presented above, many mixtures, like the

---

to be paid.

NYSE, and variations of these pure forms exist. Special trading forms, e.g. off-exchange trades for large orders (block trading) or special trading facilities for small orders, complete the list of market forms.

1.4 Market participants and order submission

In the last sections it has been assumed for the sake of simplicity that investors trade directly on the exchange. In reality, however, they have to use an agent, who trades for them on the exchange. This agent is called a *broker*.[36] A broker transmits the order he receives from an investor, his customer, to the exchange, where the order is treated according to the rules of the exchange. He does not trade on his own account. The broker also informs the investor about the execution and the applied price of his order and settles the accounts. He also is liable for fulfilling the trade, hence the counterpart risk is reduced significantly.

In dealer markets the broker has to identify the market maker[37] offering the best price and transmit the order to this market maker for immediate execution. In matching and batch systems the broker only transmits the order to a *match maker*[38] and waits for its execution. Figure 1.2 visualizes the way orders are submitted in different market forms.

Often the roles of market participants change. A market maker or broker may want to trade for his own account and hence by our definition become an investor, or on the NYSE the market maker also is keeper of the order book, i.e. match maker. Despite these mixtures of roles individual market participants have at a particular time, their activities will fit into one of the following categories: investor, broker, market maker or match maker.

[36] There are special conditions that must be met to act as broker. In our context these conditions are of no interest and are therefore omitted here. Appendix A.4 gives a detailed description of the conditions to be met for being granted access as broker to the NASDAQ.

[37] To become a market maker very strict conditions have to be fulfilled. In some markets only one market maker per asset is allowed, e.g. the *specialists* at the NYSE. Appendix A.4 gives a detailed description of the conditions to be met for being granted access as broker to the NASDAQ.

[38] A match maker is a person that stores the order, i.e. keeps the order book, and initiates the matching of the orders.

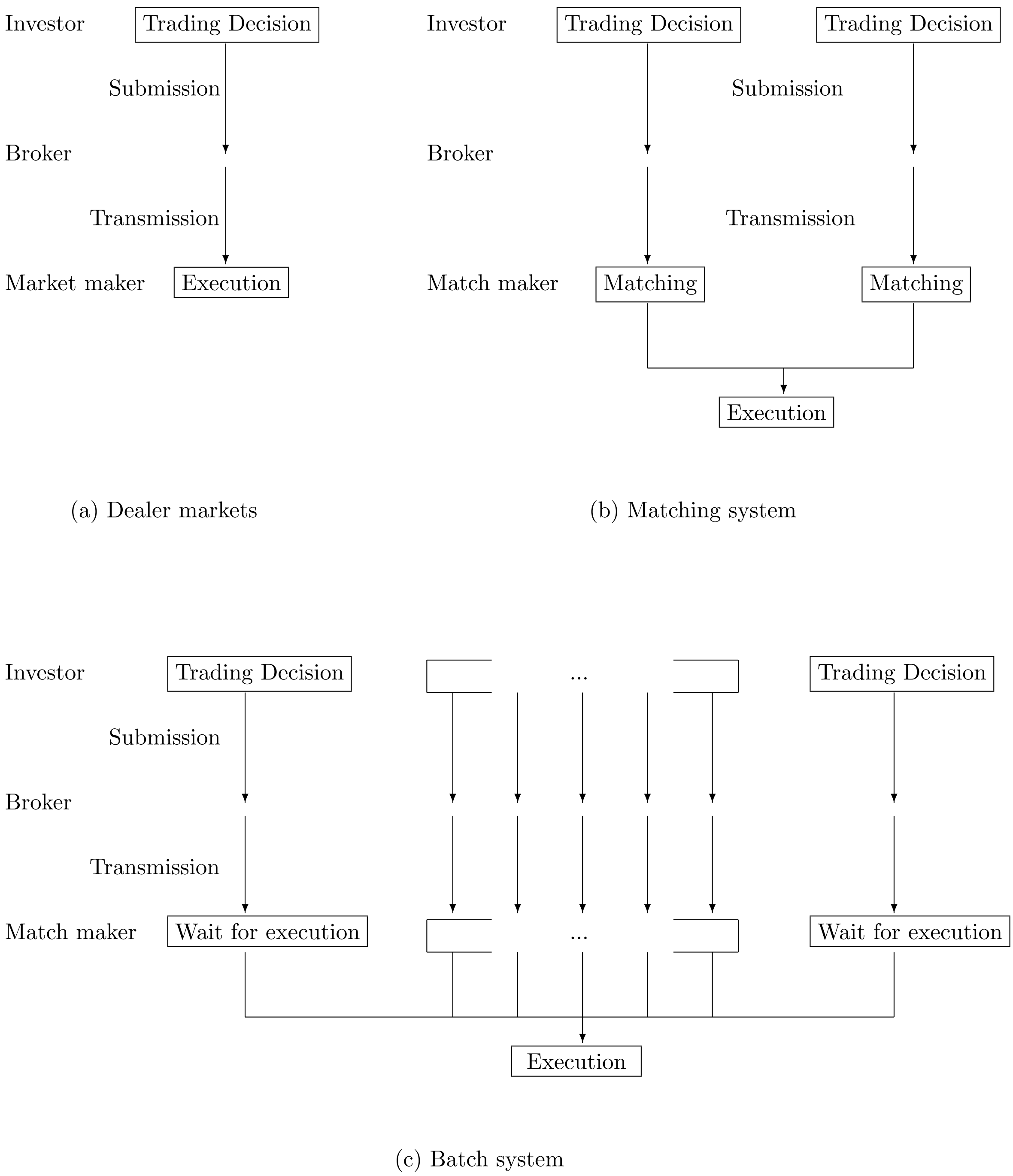


Fig. 1.2: Order flow in different market forms

The size of an order is not fixed, except in board trading. A market maker has to accept any order size at the stated price as long as the order is not too large. Large orders are normally divided into several smaller orders and executed over a longer time period, ranging between hours and months to avoid a significant influence on the price.

In most markets there exists a "normal" order size, as in board trading called a *lot*.[39] Orders with the size of a multiple of a lot can be divided into smaller, separate orders with the minimum size of one lot. These smaller orders will be executed separately with different matching orders or market makers at different points of time and prices. This splitting of large orders facilitates the execution of such an order and increases liquidity. If the orders were to be executed as a whole, it may be difficult to find a matching order with exactly the same size.

Orders with a size smaller than one lot (so called *odd-lots*)[40] are sometimes traded by special market makers or match makers, respectively, or a batch system is introduced for these small orders. The Fixing on the Frankfurt Stock Exchange at 12.00 noon is an example for such a batch system.

So far we only considered orders that were for execution at the best available price of the market. Such an order is called a *market order*. A market order will be executed at any price, in frequently traded stocks with continuous markets it normally will be executed within a short time after submission, as any offsetting order matches. In batch systems market orders are executed nearly with certainty.[41]

There exists another frequently used type of order, *limit order*. When submitting a limit order, the investor sets a maximum (minimum) price, the limit, at which he is willing to buy (sell) the security. Of course, he also will buy (sell)

[39] The typical lot on the Hong Kong Stock Exchange is 5,000 shares. In 1998 the Frankfurt Stock Exchange changed its lot size from 100 shares to 1 share. On the NASDAQ the lot sizes vary between 100 and 1000 shares, depending on the stocks.

[40] Orders that are larger than one lot, but are not a multiple of a lot are split into a part consisting of a multiple of a lot and a part with the odd-lot. They are then treated as different orders, see Keenan (1987, p. 23).

[41] There may be some special situations in which the execution is not guaranteed, but these situations have no practical relevance.

at lower (higher) prices. The execution of limit orders is not guaranteed, an offsetting order has to be found that fulfills these conditions. In most matching and batch systems the trade between two market orders is given priority over the trade of a limit and a market order. Therefore it can take a considerable time until a limit order can be executed. Limit orders will be canceled either at expiration, e.g. at the end of the day or upon withdrawal.

There exist many other order forms, but they are rarely applied in security markets. One order form that has been important in the past is the *stop loss order*. Stop loss orders are executed as a market sell order if the market price falls below a certain limit, otherwise the order is not executed. This type of order has been used in the past, when an investor had not the possibility to observe the market continuously or contact his broker immediately. The aim was to prevent greater losses in these cases. Nowadays it is possible to receive recent news on security prices and communicate without problems nearly all over the world and stop loss orders can rarely be found.[42]

1.5 Trading priority rules

We can frequently run into the situation where there is more than one order unexecuted in the market which matches an incoming order, e.g. two limit orders submitted which could be filled by an offsetting market order arriving at the market. In this case it is important to establish rules deciding which of these orders will be executed, called *trading priority rules.* They not only have an impact on the time an investor has to wait until his order is executed and hence his waiting costs, but also on the price applied.[43]

The most important rule is *price priority.* With price priority market orders

[42] Schwartz (1988, pp. 45 ff) also gives an overview is given of other order forms that are possible to submit to US stock exchanges. However, these order forms are only rarely found and for this reason not further considered here.

[43] Moulin (2000) provides an axiomatic treatment of priority rules. He shows that in each market there exists only a single optimal priority rule, however, his analysis does not allow him to determine this rule. Domowitz (1993) gives an overview of the priority rules applied on several important stock markets.

are executed first and only after all market orders have been executed, the limit order with the best price, i.e. lowest price for a buy order and highest price for a sell order, is executed, then the limit order with the second best price, and so forth. This rule ensures that securities can be bought at the lowest and be sold at the highest available price, reducing trading costs. Price priority is found as the first priority rule at all stock exchanges.

The price priority rule will in many cases not be sufficient to distinguish between all unexecuted orders in the market. It will often be found that more than one order has been submitted at the same price or is a market order, so that we need additional rules for choosing the order that is executed with a matching order. These rules are called *secondary trading priority rules.*

The most common rule is *time priority.* An order that has been transmitted earlier by the broker is executed before an order transmitted later.[44] Another rule that frequently can be found, often in combination with time priority, is *size priority.* With size priority a larger order is executed before a smaller order. This rule in combination with time priority can be found on the NYSE and since 1996 on the Toronto Stock Exchange, when the secondary priority rule was change from pure time priority.

In dealer markets the rule *public before dealer* is very important. If a limit order[45] submitted by an investor has the same limit as the price quoted by the market maker, the limit order is executed first.[46] Public before dealer is applied by the NASDAQ since 1997, while most other dealer market do not have this rule.

There are many other rules that have only minor importance in leading stock markets, such as *pro rata partial execution.* If at a certain price there is an

[44] The time at which an order is transmitted, in most cases at which it is entered into the computer system of the exchange, is measured in hundredth of seconds to ensure a clear distinction between all orders, also in times of high volume. In dealer markets similar rules can be established to determine the market maker executing an incoming order.

[45] Market orders have to be executed immediately either by the market maker or by a limit order. Consequently, there can be no unexecuted market orders.

[46] If more than one limit order is unexecuted at this price they are distinguished by another rule, e.g. time priority.

imbalance in the orders, i.e. not all orders can be executed on one side, all orders on the larger side are only partially executed with the same fraction. This rule may be applied in batch systems. The least complicated rule is the *random selection* of the order that is executed.

Not every rule can be applied in all market forms. There exists a wide variety of further rules, exceptions and modifications that are specific to certain stock exchange.

The rules and market forms presented here form only a part of the market structure. Additional rules not yet mentioned encompass maximum price change limits, lower transaction costs for certain groups of market participants, besides others. An exhaustive description of all possibilities to form a market structure lies beyond the scope of this chapter.[47]

1.6 Electronic trading mechanisms

In recent years exchanges have computerized more and more functions. At the beginning of this process brokers only used computers to facilitate their order handling, e.g. the settlement of trades with their clients, the supervision of order execution and the clearing of trades. Exchanges used computers only for displaying and storing publicized data and for the clearing process. Orders had to be transmitted in conventional ways, i.e. verbal or written, from the broker to the exchange. Market makers and match makers as well as brokers had to be physically present on the trading floor of the exchange.[48]

Later the computer was used to assist market makers and match makers in handling the order flow by ordering the orders according to the priority trading rules and displaying the relevant information. The orders had to be entered into the computers by the market makers and match makers themselves.

[47] Appendix A gives a more detailed overview of the market structure of the NASDAQ, including some, but not all, features omitted here. Rudolph and Röhrl (1997) give a more detailed overview of the economics and current state of stock exchange regulation.

[48] There existed also OTC markets that had no trading floor, but used a telephone for communication (telephone markets). On these markets in most cases only very infrequently traded stocks were listed, an exception has been the NASDAQ.

In 1973 the Frankfurt Stock Exchange had developed the first *automated order handling system*. It allowed brokers to transmit their orders electronically from their internal computer systems into the computer system of the exchange. The order was then automatically routed to the appropriate match maker. The execution of the orders still had to be done manually. Although the system has never been introduced in Frankfurt, other exchanges developed similar systems. These systems, e.g. the *DOT*-System of the NYSE implemented in 1976, were at the beginning only able to handle a small number of orders. For this reason the use was restricted to small orders. With time computer systems were able to handle more orders and the use was extended.[49]

The last step towards a *fully automated trading system* (or *electronic exchange*) was first taken in 1977 with the introduction of *CATS* on the Toronto Stock Exchange. It enabled not only the electronic transmission of orders, but without any interference was able to execute orders.[50] Similar to automated order handling systems introduced earlier, these systems were at the beginning only able to handle small amounts of orders and were therefore only used for the trade of infrequently traded stocks or the use was restricted to small orders. In 1991 the Frankfurt Stock Exchange introduced the *IBIS* trading system, where all major stocks could be traded.[51] But this trading system is not regarded as the official stock exchange, which still is a conventional exchange, it was initially established as a system for inter-bank trading. With the reduction of the minimum order size to 100 shares in 1998 this trading platform is now also open to the general public. The first official stock exchange to introduce a fully automated trading system for all traded securities and order sizes, including an electronic clearing of trades, was the Swiss Exchange in Zürich (EBS) in 1996. In the mean time, more and more exchanges have introduced an electronic exchange, at least for a

[49] The *SuperDOT 250*, implemented in November 1984 on the NYSE could handle a daily volume of 250 million shares. The capacity has been extended since then. Nowadays more than 1 billion shares can be traded on this system without problems. See Schwartz (1988, p. 27).

[50] In 1982 a similar system was introduced on the Tokyo Stock Exchange, see Schwartz (1988, p. 27), and 1986 on the Paris Stock Exchange, see Schwartz (1993, p. 90).

[51] A new version with the name *XETRA* is used since November 1997. Unser and Oehler (1998) provide a concise introduction into the features of this system.

part of their trading activities.[52]

In dealer markets, also with a fully automated trading system, a minimum of inference is needed. The system can automatically route the orders to the market maker quoting the best price, who then only has to confirm the trade and enter his new prices into the system. In matching or batch systems a match maker is no longer needed, his duties can be taken over entirely by the computer system.

With the introduction of a fully automated trading system, a trading floor is no longer needed. All market participants only have to be connected to the computer system of the exchange. Their locations have no importance, they can trade from any place around the globe. If market participants gain access from another country this is called *remote access* and enables twenty-four hour trading on the same exchange if the market participants are located in different time zones.[53]

Computerization *per se* does not change the market structure or forces to do so.[54] In many cases, however, the introduction of a new computer system is used to establish a new market structure to increase the efficiency of trading. In many cases order book systems are introduced, replacing or complementing market makers, like on the London Stock Exchange.

Nevertheless, computerization has had a great impact on trading: orders can be executed more accurately to the rules, as errors are reduced, they furthermore can be executed faster by a computer system, trading costs are lowered and trade information (volume and prices) can be transmitted faster. Investors are also able to react more quickly to changes in the market, as access to real-time data has become affordable to a wide public[55] and investors can submit their orders electronically to their broker, e.g. by using the internet.

[52] Especially the futures exchanges e.g. EUREX or LIFFE have introduced electronic trading.

[53] The GLOBEX system of the CME, introduced in 1992, was the first, and by now the only system that enables a twenty-four hour trading on the same exchange.

[54] As has been pointed out above, only exchanges with crowd trading are not able to preserve their market form, as this form needs the physical presence of the brokers on the trading floor to negotiate the price.

[55] Several brokers display real-time data for their customers on the internet for free. The costs of data providers also have decreased substantially in recent years due to competition from the internet.

Before the computerization of exchanges, a broker had to find the best price for an order, what in dealer markets with many competing market makers was a difficult task, because the quotes may change after every trade. In matching markets it was important to transmit the orders as soon as possible to the exchange before the prices changed too much. With a computerized exchange the best price is found by the computer system, the broker only has to transmit the order he receives - more and more electronically - from his client. His service is reduced to a pure transmission of the order. Due to this computerization and the increased global competition of brokers, broker fees have decreased significantly in the last years, reducing the costs to investors.[56]

As the broker no longer plays an active role, the submission of an order via a broker, who immediately passes this order without any interferences to the exchange, has the same effect as if the order would be directly submitted to the exchange, i.e. as if the investors were directly interacting.[57] Therefore the process of computerization is also called the *disintermediation* of financial markets.

1.7 Recent developments

Stock and futures exchanges are currently challenged by a number of changes taking place. First of all the computerization has increased the *competition* between exchanges. Due to the possibility of remote access in computerized markets, the physical location of an exchange close to the most important financial institutions has become less important. This development increased competition globally between exchanges for the listing of securities as well as for order flow.

In this competition trading costs are a very important factor. Therefore exchanges have to improve their market structure to reduce trading costs for investors. Increased fixed costs for developing, improving and maintaining the

[56] For a trade of about USD 10,000 several brokers offer fees of less than USD 10 when the order is placed using the internet, compared to fees of more than USD 100 not long ago and still applied by conventional brokers.

[57] The only reason, besides regulatory restrictions, that brokers are still used as intermediaires, is to reduce the counterpart risk of a trade as brokers guarantee to fulfill the trades of their customers.

Partners	**Year**	**Features**
DTB, SOFFEX	1998	common trading platform EUREX
NASDAQ, AMEX	1998	merger, separate markets
Paris, Zürich	1999	reciprocal access of members
EUREX, CBoT	1999	reciprocal access of members
LIFFE, CME	1999	reciprocal access of members
Paris, Lissabon	1999	reciprocal access of members
Paris, Brussels, Amsterdam	2000	merged to Euronext
Frankfurt, London	2000,2005	abandoned merger talks
NASDAQ, Instinet	2005	talks about merger
NYSE, Archipelago	2005	talk about merger

Tab. 1.2: Overview of alliances between major stock and futures exchanges

computer systems of an exchange with fast growing trading volumes lead to numerous *alliances* between exchanges, seeking economies of scale. The number of these alliances, cooperations and mergers are advancing very fast. Table 1.2 lists some of the most important alliances, which in most cases guarantee reciprocal access to the partners for all members of one exchange. In some cases a common listing of securities, a common trading platform or even a merger are planned. However, with exception of the EUREX trading platform,[58] Euronext and the NASDAQ/AMEX-merger none of the listed alliances is currently in operation. The agreed alliances were often quietly dissolved or let to no further cooperations between the exchanges after the burst of the internet boom in mid 2000 due to the reduced trading activity during that time period.

Typically stock exchanges are organized as non-profit organizations supported by its members, in general the financial institutions having direct access to the market. Growing investments into the computerization made it more and more difficult for exchanges to raise the necessary capital from its members. Therefore several exchanges discuss the transformation into *for-profit corporations* which would give them more flexibility in raising capital.[59]

[58] A brief description of the EUREX trading system is given in Schiller and Marek (2000).
[59] Such plans are under discussion at the NYSE, the NASDAQ, and the CBoT. The members of the LSE, Frankfurt Stock Exchange, Euronext and the CME already have approved such plans and the exchanges are quoted.

This development is further accelerated by the increasing number of *Electronic Communications Networks* (ECNs), which are private-owned trading platforms offering similar services as an exchange. These ECNs, which are mostly financed and operated by large financial institutions, have gained a substantial market share in trading NASDAQ securities.[60] ECNs have a gained a market share of more than a quarter in dollar trading volume of securities listed on the NASDAQ.[61] The importance of ECNs is witnessed in particular by the fact that both the NASDAQ and NYSE have both planned to merge with one of their main ECN competitors. We also have seen a small number of mergers between ECNs, e.g. Island and Instinet, confirming the general trend towards consolidation of exchnages.

A final development is the extension of *trading hours.* Especially institutional investors are demanding longer trading hours, which enable them to react faster on news they receive. On May 15, 2000 the Milano Stock Exchange extended trading 65 large stocks until 8.30 pm and the Frankfurt Stock Exchange extended their trading hours for all listed stocks until 8 pm on June 2, 2000.[62] The lack of demand in the aftermath of the internet bubble, however, let again to a reduction of the trading hours.

[60] The market share for securities listed on the NYSE can be neglected as regulation does generally not allow members of the NYSE to trade securities listed on the NYSE off exchange. Up to now in Europe only a single ECN concentrating on securities listed on the LSE, Tradepoint, exists, which has a negligible market share of less than 1% in trading volume.

[61] Gomber (2000) gives an overview of the requirements for a successful electronic trading system based on economic considerations.

[62] The traditional trading hours on the Frankfurt Stock Exchange have been from 10.30 am to 1.30 pm, currently most European markets operate from 9 am to 5 pm, the NYSE operates from 9.30 am to 4 pm. Extensions to 8 or 10 pm were planned by most leading stock exchanges, but after the burst of the internet bubble have never been implemented.

Review questions

1. What are the characteristics a good must have to be traded in a market?
2. What is the role of security markets in an economy?
3. What characterizes an exchange?
4. Compare call and auction markets.
5. What are the relative advantages and disadvantages of dealer markets and limit order markets?
6. What are the differences between market and limit orders?
7. Why do markets usually have fixed lot sizes?
8. Why is it important to have secondary priority rules?
9. What are the implications of computerized trading?
10. Why is there a tendency by exchanges to merge?

Chapter 2

Auction Markets

This chapter will introduce the reader to theories on auction markets. Particular attention will be paid to the implications of asymmetric information between investors. The main contents of this chapter evolves around

- the optimal behavior of informed investors
- the implications of this behavior for asset prices and trading volume
- the timing of trading

Key readings:

Albert S. Kyle: Continuous Auctions and Insider Trading, *Econometrica*, 53, 1315-1335, 1985

Craig W. Holden and Avanidhar Subrahmanyam: Long-Lived Private Information and Imperfect Competition, *Journal of Finance*, 47, 247-270, 1992

Matthew Spiegel and Avanidhar Subrahmanyam: Informed Speculation and Hedging in a Noncompetitive Securities Market, *Review of Financial Studies*, 5, 307-329, 1992

Lawrence Blume, David Easley and Maureen O'Hara: Market Statistics and Technical Analysis: The Role of Volume, *Journal of Finance*, 49, 153-181, 1994

This chapter will give an overview of the main contributions in market microstructure theory on order-driven or auction markets. Prices are not analyzed on a trade-by-trade basis, but aggregated over a given period of time, e.g. several hours of a trading day or an entire trading day. These models allow to explain short-term variations in price and trading volume, like the day-of-the-weak-anomaly.

We assume initially that a single asset is traded in T trading rounds by two groups of investors, informed and uninformed investors. After the final round of trading the asset is liquidated and the proceedings are consumed. The informed investors, called *insiders*,[1] receive a signal on the liquidation value of the asset before the first round of trading. Informed investors are assumed to be risk neutral. The uninformed investors do not receive a signal, hence their trades are not based on superior information, but on exogenous needs for liquidity (*noise traders*) as in the first two models to be presented or to rebalance their portfolios as a result of changing prices (*hedgers*) in the final model.

In each trading round (auction) investors can submit their orders to a match maker.[2] The match maker arranges the trades at a single price. This price has to be set such that it equals the expected liquidation value given the match makers information. The information the match maker has, as he is assumed to be uninformed, is the order imbalance of that auction. As the match maker does not know whether the orders submitted come from informed or uninformed investors (the use of brokers ensures anonymity of the investors to the match maker) more orders to buy than to sell could mean that informed investors received a signal indicating that the liquidation value is above the last observed price. But it could also be the result of pure incident. This price setting behavior is identical to the

[1] The term *insider* as used in this context has to be distinguished from the legal definition of an insider. Here an insider is an investor having acquired publicly available information. In the legal definition an insider has access to not yet publicly available information due to his position in a company, e.g. as member of the board.

[2] We allow only for market orders, the submission of limit orders, i.e. of demand schedules, is not allowed. When submitting an order the orders submitted by the other investors for this auction are not known to any investor, i.e. the choices have to be simultaneous and cannot be conditioned on the behavior of other investors.

efficient market hypothesis, where the price equals the expected value of the asset given a set of information.[3] For simplicity it is assumed that the match maker charges no fee for his service.

The next section presents the KYLE (1985) model. Although the model presented in section 2.1.1 is included as a special case in section 2.1.2 it is treated separately due to its outstanding position in the development of theories on auction markets. It assumes a single informed investor and a large number of noise traders. In section 2.1.2 this model is extended to include more than one informed investor. A last extension in 2.2 assumes that the uninformed investors are no longer noise traders, but risk averse hedgers maximizing their own utility. Section 2.3 shows the information trading volume reveals. Finally in 2.4 it is shown how the developed models can be used to explain some effects observed in stock markets. The presentation of the models have been adopted to use the same framework for all models.

2.1 Auctions with informed investors and noise traders

2.1.1 Auctions with a single informed investor

KYLE (1985) presents a model where a single informed investor trades a single asset together with N uninformed noise traders. It is assumed that the informed investor becomes to know the liquidation value v of the asset with certainty.[4] For uninformed investors the liquidation value is a random variable $\widetilde{v}$ that is normally distributed with initial mean p_0 and variance Σ_0:

$$\widetilde{v} \sim N\left(p_0, \Sigma_0\right). \tag{2.1}$$

[3] Although this behavior is very restrictive, it captures some of the behaviors in real stock exchanges applying auction markets. For example the Frankfurt Stock Exchange urges the match makers to rise the price if more buy than sell orders are in the market. See Deutsche Boerse Group (ed.): Rules and Regulations, Part 6, 3.22 and 3.3.1.2.

[4] The original models assume that the signal is not perfect, but that the liquidation value has a positive variance given this information. As informed investors are assumed to be risk neutral this remaining risk has not to be considered in the optimal behavior as maximizing expected profit and expected utility are equivalent.

Uninformed investors are assumed to trade for purely exogenous reasons, they do not maximize any objective function. Their order sizes, $\widetilde{u}^i$, are random variables that are assumed to be independently identically normally distributed with mean zero and variance $\sigma_{u^i}^2$.[5] They are independent of the order sizes of other uninformed investors, the behavior of the informed investor, independent over time and of $\widetilde{v}$:

$$\widetilde{u}^i \sim N\left(0, \sigma_{u^i}^2\right). \tag{2.2}$$

The total relevant order flow from uninformed investors is their order imbalance, all other orders can directly be matched:

$$\widetilde{u} = \sum_{i=1}^{N} \widetilde{u}^i \sim N\left(0, N\sigma_{u^i}^2\right) = N\left(0, \sigma_u^2\right). \tag{2.3}$$

We assume that T, N, $\sigma_{u^i}^2$, p_0 and Σ_0 are known by all market participants as well as the normal distribution of the variables.

The order size of the informed investor is denoted $\widetilde{x}$. The match maker cannot distinguish between orders from informed and uninformed investors, hence he only can observe the aggregated order flow $\widetilde{u}+\widetilde{x}$. This order flow is used to determine the price according to

$$\widetilde{p} = E[\widetilde{v}|u + x]. \tag{2.4}$$

Kyle (1985) first investigates the optimal behavior of the informed investor by choosing an optimal $\widetilde{x}$ in a single auction, i.e. for $T = 1$. The informed investor maximizes his expected profits from trading given his information on the liquidation value of the asset. His profits are

$$\widetilde{\pi} = (\widetilde{v} - \widetilde{p})\widetilde{x}. \tag{2.5}$$

Only linear equlibria are considered by Kyle (1985). This assumption gives rise

[5] In deriving the results the assumption of normality is central. Relaxing this assumption to the class of elliptical functions gives similar results, but further generalizations may give different results.

to the following pricing rules:

$$\widetilde{p} = \mu + \lambda(\widetilde{x} + \widetilde{u}), \tag{2.6}$$

$$\widetilde{x} = \alpha + \beta\widetilde{v}. \tag{2.7}$$

With (2.5) - (2.7) we get for the expected profits

$$\begin{aligned} E[\widetilde{\pi}|v] &= E[(\widetilde{v} - \mu - \lambda(\widetilde{x} + \widetilde{u}))\widetilde{x}|v] \\ &= (v - \mu - \lambda x)x. \end{aligned} \tag{2.8}$$

Maximizing (2.8) to find the optimal order size of the informed investor gives the following first order condition:

$$v - \mu - \lambda x - \lambda x = v - \mu - 2\lambda x = 0. \tag{2.9}$$

Rearranging results in

$$x = -\frac{\mu}{2\lambda} + \frac{v}{2\lambda}. \tag{2.10}$$

Comparing coefficients with (2.7) we get

$$\begin{aligned} \beta &= \frac{1}{2\lambda}, \\ \alpha &= -\frac{\mu}{2\lambda} = -\mu\beta. \end{aligned} \tag{2.11}$$

The second order condition for a maximum

$$-2\lambda < 0, \tag{2.12}$$

states that we only have to consider positive λ. Using (2.4) we get with (2.1) and (2.3) and the results of the conditional mean of jointly normally distributed random variables:

$$\begin{aligned} \widetilde{p} &= E[\widetilde{v}|\widetilde{x} + \widetilde{u}] = E[\widetilde{v}] + \frac{Cov[\widetilde{v}, \widetilde{x} + \widetilde{u}]}{Var[\widetilde{x} + \widetilde{u}]}(x + u - E[\widetilde{x} + \widetilde{u}]) \\ &= p_0 + \frac{Cov[\widetilde{v}, \alpha + \beta\widetilde{v} + \widetilde{u}]}{Var[\alpha + \beta\widetilde{v} + \widetilde{u}]}(\widetilde{x} + \widetilde{u} - E[\alpha + \beta\widetilde{v} + \widetilde{u}]) \\ &= p_0 + \frac{\beta Cov[\widetilde{v}, \widetilde{v}] + Cov[\widetilde{v}, \widetilde{u}]}{\beta^2 Var[\widetilde{v}] + Var[\widetilde{u}] + 2Cov[\widetilde{v}, \widetilde{u}]} \times \\ &\quad \times(\widetilde{x} + \widetilde{u} - \alpha - \beta E[\widetilde{v} - E[\widetilde{u}]]) \\ &= p_0 + \frac{\beta\Sigma_0}{\beta^2\Sigma_0 + \sigma_u^2}(\alpha + \beta p_0) + \frac{\beta\Sigma_0}{\beta^2\Sigma_0 + \sigma_u^2}(\widetilde{x} + \widetilde{u}). \end{aligned} \tag{2.13}$$

By comparing coefficients with (2.6) we see that

$$
\begin{aligned}
\lambda &= \frac{\beta\Sigma_0}{\beta^2\Sigma_0 + \sigma_u^2}, \\
\mu &= p_0 - \frac{\beta\Sigma_0}{\beta^2\Sigma_0 + \sigma_u^2}(\alpha + \beta p_0) = p_0 - \lambda(\alpha + \beta p_0).
\end{aligned}
\tag{2.14}
$$

Solving (2.11) and (2.14) we get with (2.12):

$$
\begin{aligned}
\beta &= \sqrt{\frac{\sigma_u^2}{\Sigma_0}}, \\
\lambda &= 2\sqrt{\frac{\Sigma_0}{\sigma_u^2}}, \\
\mu &= p_0, \\
\alpha &= -\beta p_0.
\end{aligned}
\tag{2.15}
$$

Therewith we can rewrite (2.6) and (2.7) as

$$p = p_0 + \lambda(x + u), \tag{2.16}$$

$$x = \beta(v - p_0). \tag{2.17}$$

The linear equilibrium exists and is unique. Nothing can be said about the existence of further nonlinear equilibria.

Uninformed investors cannot observe the order flow, only the price that is set by the match maker. Using this information they can update their beliefs on the distribution of the liquidation value:

$$
\begin{aligned}
\Sigma_1 &= Var[\widetilde{v}|p] = Var[\widetilde{v}] - \frac{Cov[\widetilde{v},\widetilde{p}]^2}{Var[\widetilde{p}]} \\
&= \Sigma_0 - \frac{Cov[\widetilde{v}, p_0 + \lambda(\widetilde{x}+\widetilde{u})]^2}{Var[p_0 + \lambda(\widetilde{x}+\widetilde{u})]} \\
&= \Sigma_0 - \frac{\lambda^2 Cov[\widetilde{v}, \widetilde{x}+\widetilde{u}]^2}{\lambda^2 Var[\widetilde{x}+\widetilde{u}]} \\
&= \Sigma_0 - \frac{(Cov[\widetilde{v},\widetilde{x}] + Cov[\widetilde{v},\widetilde{u}])^2}{Var[\widetilde{x}] + Var[\widetilde{u} + 2Cov[\widetilde{x},\widetilde{u}]]} \\
&= \Sigma_0 - \frac{Cov[\widetilde{v}, \beta(\widetilde{v}-p_0)]^2}{Var[\beta(\widetilde{v}-p_0)] + \sigma_u^2 + 2Cov[\beta(\widetilde{v}-p_0),\widetilde{u}]} \\
&= \Sigma_0 - \frac{\beta^2 Cov[\widetilde{v},\widetilde{v}]^2}{\beta^2 Var[\widetilde{v}] + \sigma_u^2 + 2\beta Cov[\widetilde{v},\widetilde{u}]} \\
&= \Sigma_0 - \frac{\beta^2\Sigma_0^2}{\beta^2\Sigma_0 + \sigma_u^2} \\
&= \Sigma_0 - \frac{\frac{\sigma_u^2}{\Sigma_0}\Sigma_0^2}{\frac{\sigma_u^2}{\Sigma_0}\Sigma_0 + \sigma_u^2} \\
&= \Sigma_0 - \frac{\sigma_u^2\Sigma_0}{2\sigma_u^2} \\
&= \frac{1}{2}\Sigma_0,
\end{aligned} \tag{2.18}
$$

$$
\begin{aligned}
p_1 &= E[\widetilde{v}|p] = E[\widetilde{v}] + \frac{Cov[\widetilde{v},\widetilde{p}]}{Var[\widetilde{p}]}(p - E[\widetilde{p}]) \\
&= p_0 + \frac{Cov[\widetilde{v}, \beta(\widetilde{v}-p_0)]}{Var[\beta(\widetilde{v}-p_0)]}(p_0 + \lambda(x+u) - E[p_0 + \lambda(\widetilde{x}+\widetilde{u})]) \\
&= p_0 + \frac{\beta Cov[\widetilde{v},\widetilde{v}]}{\beta^2 Var[\widetilde{v}]}(\lambda(x + u - E[\widetilde{x}])) \\
&= p_0 + \frac{1}{\beta}\lambda(\beta(v-p_0) + u - E[\beta(\widetilde{v}-p_0)]) \\
&= p_0 + \lambda(v-p_0) + \frac{\lambda}{\beta}u - \lambda(E[\widetilde{v}-p_0]) \\
&= p_0 + \lambda(v-p_0) + 2\lambda^2 u \\
&= p + 2\lambda^2 u.
\end{aligned} \tag{2.19}
$$

As $E[\widetilde{u}|p] = 0$ we find that $E[\widetilde{p}_1|p] = p$, hence the posterior distribution of $\widetilde{v}$ is

$$
\widetilde{v}|p \sim N\left(p, \frac{1}{2}\Sigma_0\right). \tag{2.20}
$$

The variance the uninformed investors attribute to the liquidation value of the asset can be interpreted on how much information is incorporated into the price. A variance of zero has to be interpreted as perfect revelation of the information through prices, the closer this variance is to Σ_0 the less informative the price is. As by observing only the price the variance halves, it can be said that half of the information is incorporated into prices. The variance of the liquidation value we can view as a measure for the *informativeness of prices.*

λ measures the influence an additional unit of an order has on the price:

$$\frac{\partial p}{\partial(x+u)} = \lambda. \tag{2.21}$$

If we define a market as liquid if by placing an additional order the price does not change. Therefore λ measures the liquidity of a market, the closer it is to zero the more liquid the market is. Usually $1/\lambda$ is taken as a measure of liquidity as by this definition a larger value corresponds to higher liquidity.

From (2.15) we see that a market is more liquid the more liquidity traders are in the market, i.e. the larger N is, or the more their order sizes vary, i.e the larger $\sigma^2_{u^i}$ is.

This framework can now be extended to trades occurring in $1 < T < \infty$ auctions in a given time period $[0,1]$.This setting allows the informed investor to time his trades such that over time his expected profits are maximized. The order flow from an investor for auction $1 \leq t \leq T$ is denoted u^i_t for the order of an uninformed investor. The variance of the order flow from uninformed investors, $\sigma^2_{u^i,t}$, remains constant over the entire period, i.e.

$$\sigma^2_{u^i} = \sum_{t=1}^{T} \sigma^2_{u^i,t} = T\sigma^2_{u^i,t}. \tag{2.22}$$

Hence we have

$$u^i_t \sim N\left(0, \frac{1}{T}\sigma^2_{u^i}\right) \tag{2.23}$$

for all $t = 1, \ldots, T$. The total order flow from the informed investor in the first t auctions is denoted x_t and the order flow for a specific auction k Δx_k. Hence we have for all $t = 1, \dots, T$:

$$x_t = \sum_{k=1}^{t} \Delta x_k. \tag{2.24}$$

When again considering only linear equilibria we get in analogy to (2.6) and (2.7) for all $t = 1, \dots, T$:

$$\begin{aligned} \widetilde{p}_t &= \mu_t + \lambda_t(\Delta x_t + \widetilde{u}_t), && (2.25)\\ \Delta x_t &= \alpha_t + \beta_t v. && (2.26) \end{aligned}$$

The match maker sets his price again such that it equals the liquidation value of the asset given the order flows observed in the past:

$$p_t = E[\widetilde{v}|\Omega_t], \tag{2.27}$$

where $\Omega_t = \{\Delta x_1 + u_1, \dots, \Delta x_t + u_t\}$. The informed investor maximizes his expected profits by choosing the optimal order size given the liquidation value and past prices. His profits from the remaining auctions are given by[6]

$$\widetilde{\pi}_t = \sum_{k=t}^{T} (\widetilde{v} - p_k) x_k = (\widetilde{v} - p_t)\Delta x_t + \widetilde{\pi}_{t+1}. \tag{2.28}$$

We assume the expected profits to be quadratic in $v - p_t$:

$$E[\widetilde{\pi}_{t+1}|p_1, \dots, p_t, v] = \gamma_t(v - p_t)^2 + \delta_t. \tag{2.29}$$

Using (2.28) we get as the objective function of the informed investor:

$$\begin{aligned} E[\widetilde{\pi}_{t+1}|p_1, \dots, p_t, v] &= E[(\widetilde{v} - p_t)\Delta x_t + \widetilde{\pi}_{t+1}|p_1, \dots, p_t, v] && (2.30)\\ &= E[(\widetilde{v} - p_t)\Delta x_t|p_1, \dots, p_t, v]\\ &\quad + \gamma_t(v - p_t)^2 + \delta_t\\ &= E[(\widetilde{v} - \mu_t - \lambda_t(\Delta x_t + \widetilde{u}_t))\Delta x_t|p_1, \dots, p_t, v]\\ &\quad + \gamma_t(v - \mu_t - \lambda_t(\Delta x_t + u_t))^2 + \delta_t\\ &= (v - \mu_t - \lambda_t \Delta x_t)\Delta x_t + \delta_t + \gamma_t \lambda_t^2 \sigma_{u^i,t}^2\\ &\quad + \gamma_t(v - \mu_t - \lambda_t(\Delta x_t + u_t))^2. \end{aligned}$$

[6] It is worth noting that it is assumed here that future profits are not discounted to their present value.

Maximizing this expression for the optimal order size to submit in auction t gives the following first order condition:

$$\begin{aligned} (2.31) \qquad 0 &= v - \mu_t - \lambda_t \Delta x_t - \lambda_t \Delta x_t + 2\gamma_t (v - \mu_t - \lambda_t \Delta x_t)(-\lambda_t) \\ &= (v - \mu_t)(1 - 2\gamma_t \lambda_t) - 2\lambda_t \Delta x_t (1 - \gamma_t \lambda_t). \end{aligned}$$

Solving for Δx_t we get

$$(2.32) \qquad \Delta x_t = -\frac{1 - 2\gamma_t \lambda_t}{2\lambda_t (1 - \gamma_t \lambda_t)} \mu_t + \frac{1 - 2\gamma_t \lambda_t}{2\lambda_t (1 - \gamma_t \lambda_t)} v.$$

Comparing coefficients with (2.26) gives

$$(2.33) \qquad \beta_t = \frac{1 - 2\gamma_t \lambda_t}{2\lambda_t (1 - \gamma_t \lambda_t)},$$

$$(2.34) \qquad \alpha_t = -\frac{1 - 2\gamma_t \lambda_t}{2\lambda_t (1 - \gamma_t \lambda_t)} \mu_t = -\beta_t \mu_t.$$

The second order condition

$$(2.35) \qquad -2\lambda_t (1 - \gamma_t \lambda_t) < 0$$

implies that an equilibrium exists only if $\lambda_t (1 - \gamma_t \lambda_t) > 0$. Using (2.27) we get

$$\begin{aligned} (2.36) \qquad p_t &= E[\widetilde{v} | \Omega_t] \\ &= E[E[\widetilde{v} | \Omega_{t-1}] | \Delta x_t + u_t] \\ &= E[\widetilde{v} | \Omega_{t-1}] \\ &\quad + \frac{Cov[\widetilde{v}, \Delta \widetilde{x}_t + \widetilde{u}_t | \Omega_{t-1}]}{Var[\Delta \widetilde{x}_t + \widetilde{u}_t | \Omega_{t-1}]} (\Delta x_t + u_t - E[\Delta \widetilde{x}_t + \widetilde{u}_t | \Omega_{t-1}]) \\ &= p_{t-1} + \frac{Cov[\widetilde{v}, \alpha_t + \beta_t \widetilde{v} + \widetilde{u}_t | \Omega_{t-1}]}{Var[\alpha_t + \beta_t \widetilde{v} + \widetilde{u}_t | \Omega_{t-1}]} \times \\ &\quad \times (\Delta x_t + u_t - E[\alpha_t + \beta_t \widetilde{v} + \widetilde{u}_t | \Omega_{t-1}]) \\ &= p_{t-1} + \frac{\beta_t Cov[\widetilde{v}, \widetilde{v} | \Omega_{t-1}]}{\beta_t^2 Var[\widetilde{v} | \Omega_{t-1}] + \sigma^2_{u^i,t}} (\Delta \widetilde{x}_t + \widetilde{u}_t - \alpha_t - \beta_t p_{t-1}) \\ &= p_{t-1} + \frac{\beta_t \Sigma_{t-1}}{\beta_t^2 \Sigma_{t-1} + \sigma^2_{u^i,t}} (\Delta \widetilde{x}_t + \widetilde{u}_t - (\alpha_t + \beta_t p_{t-1})). \end{aligned}$$

Comparing coefficients with (2.25) we receive

$$\begin{aligned} (2.37) \qquad \lambda_t &= \frac{\beta_t \Sigma_{t-1}}{\beta_t^2 \Sigma_{t-1} + \sigma^2_{u^i,t}}, \\ \mu_t &= p_t - 1 - \frac{\beta_t \Sigma_{t-1}}{\beta_t^2 \Sigma_{t-1} + \sigma^2_{u^i,t}} (\alpha_t + \beta_t p_{t-1}) \\ &= p_{t-1} - \lambda_t (\alpha_t + \beta_t p_{t-1}). \end{aligned}$$

Solving (2.33) and (2.37) we get

$$\begin{aligned} \lambda_t &= \frac{\beta_t \Sigma_{t-1}}{\beta_t^2 \Sigma_{t-1} + \sigma^2_{u^i,t}}, \\ \beta_t &= \frac{1 - 2\lambda_t \gamma_t}{2\lambda_t (1 - \lambda_t \gamma_t)}, \\ \mu_t &= p_{t-1}, \\ \alpha_t &= -\beta_t p_{t-1}. \end{aligned} \tag{2.38}$$

Therewith we have found a unique linear equilibrium:

$$p_t = p_{t-1} + \lambda_t(\Delta x_t + u_t), \tag{2.39}$$

$$\Delta x_t = \beta_t (v - p_{t-1}). \tag{2.40}$$

By comparing the coefficients in (2.30) and (2.29) we see that

$$\delta_{t-1} = \delta_t + \gamma_t \lambda_t^2 \sigma^2_{u^i,t}, \tag{2.41}$$

$$\begin{aligned} \gamma_{t-1}(v - p_{t-1}) &= (v - \mu_t - \lambda_t \Delta x_t)\Delta x_t + \gamma_t (v - \mu_t - \lambda_t \Delta x_t)^2 \\ &= \left(v - p_{t-1} - \lambda_t \beta_t (v - p_{t-1})\right) \beta_t (v - p_{t-1}) \\ &\quad + \gamma_t \left(v - p_{t-1} - \lambda_t \beta_t (v - p_{t-1})\right)^2 \\ &= (v - p_{t-1})^2 (1 - \lambda_t \beta_t) \beta_t \\ &\quad + \gamma_t (v - p_{t-1})^2 (1 - \lambda_t \beta_t)^2 \\ &= (v - p_{t-1})^2 \left((1 - \lambda_t \beta_t)\beta_t + \gamma_t (1 - \lambda_t \beta_t)^2\right). \end{aligned} \tag{2.42}$$

Hence

$$
\begin{aligned}
(2.43) \qquad \gamma_{t-1} &= (1-\lambda_t\beta_t)\beta_t + \gamma_t(1-\lambda_t\beta_t) \\
&= \left(1-\lambda_t\frac{1-2\gamma_t\lambda_t}{2\lambda_t(1-\gamma_t\lambda_t)}\right)\beta_t + \left(1-\lambda_t\frac{1-2\gamma_t\lambda_t}{2\lambda_t(1-\gamma_t\lambda_t)}\right)^2\gamma_t \\
&= \frac{1}{2(1-\gamma_t\lambda_t)}\beta_t + \frac{1}{4(1-\gamma_t\lambda_t)^2}\gamma_t \\
&= \frac{2(1-\gamma_t\lambda_t)\beta_t + \gamma_t}{4(1-\gamma_t\lambda_t)^2} \\
&= \frac{2(1-\gamma_t\lambda_t)\frac{1-2\gamma_t\lambda_t}{2\lambda_t(1-\gamma_t\lambda_t} + \gamma_t}{4(1-\gamma_t\lambda_t)^2} \\
&= \frac{\frac{1-2\gamma_t\lambda_t}{\lambda_t} + \gamma_t}{4(1-\gamma_t\lambda_t)^2} \\
&= \frac{1-\gamma_t\lambda_t}{4\lambda_t(1-\gamma_t\lambda_t)^2} \\
&= \frac{1}{4\lambda_t(1-\gamma_t\lambda_t)}.
\end{aligned}
$$

The new variance of the liquidity value for the uninformed investors is determined as

$$
\begin{aligned}
(2.44) \qquad \Sigma_t &= Var[\widetilde{v}|p_t] \\
&= Var[\widetilde{v}|\Omega_t] \\
&= Var[\widetilde{v}|\Omega_{t-1}] - \frac{Cov[\widetilde{v}, \Delta\widetilde{x}_t + \widetilde{u}_t|\Omega_{t-1}]^2}{Var[\Delta\widetilde{x}_t + \widetilde{u}_t|\Omega_{t-1}]} \\
&= \Sigma_{t-1} - \frac{Cov[\widetilde{v}, \beta_t(\widetilde{v}-p_{t-1})|\Omega_{t-1}]^2}{Var[\beta_t(\widetilde{v}-p_{t-1}) + \sigma^2_{u^i,t}|\Omega_{t-1}]} \\
&= \Sigma_{t-1} - \frac{\beta_t^2\Sigma_{t-1}^2}{\beta_t^2\Sigma_{t-1} + \sigma^2_{u^i,t}} \\
&= \Sigma_{t-1} - \lambda_t\beta_t\Sigma_{t-1} \\
&= (1-\lambda_t\beta_t)\Sigma_{t-1}.
\end{aligned}
$$

Equations (2.38) - (2.41), (2.43) and (2.44) completely characterize the equilibrium. In order to avoid profits of the informed investor after the last auction, we have to impose the boundary condition $\delta_T = \gamma_T = 0$. It can now be shown that with these two conditions the equilibrium exists, i.e. (2.35) is fulfilled, and that it is a unique linear equilibrium.[7]

[7] See Kyle (1985, pp. 1325 f.) for a formal proof. Again, there can exist other, nonlinear

As we see from (2.37) $1 - \lambda_t \beta_t$ is always between zero and one, hence the variance of the liquidity value for the uninformed investors is strictly decreasing as long as $0 < \sigma^2_{u^i,t} < \infty$. Therewith the information is gradually incorporated into the price. An efficient market in the strong form is not achieved immediately, i.e. prices do not reveal fully the information, including the private information of the insider. They tend to become fully revealing over time. This behavior is the result of the profit maximization of the insider, he exploits his informational advantage over time by placing only small orders to hide his trades in the trades of noise traders.

Figure 2.1 illustrates the behavior of the liquidity parameter λ_t depending on the time and the number of auctions.[8] It can be seen that the liquidity is falling with time for all T, but as T increases λ_t becomes nearly constant over time. It is optimal for the informed investor is to hold λ_t constant, the more auctions there are, the better he can achieve this situation. Hence by placing an order the price is always equally affected, i.e. the costs of trading (adverse selection costs) are constant over time and prices adjust gradually.

2.1.2 Auctions with multiple informed investors

Holden and Subrahmanyam (1992) extended the model of Kyle (1985) to incorporate multiple informed investors. The only change that has to be made in the notation is to view the informed investors order flow, Δx_t, to be composed of the order flows of M informed investors, Δx_t^i. Denoting $\Delta \overline{x}_t^i$ the conjecture of informed investor i about the average order flow of the other informed investors, we have

$$\Delta x_t = \sum_{k=1}^{M} \Delta x_t^k = \Delta x_t^i + (M-1)\Delta \overline{x}_t^i. \tag{2.45}$$

With the other assumptions identical to Kyle (1985) the derivation follows exactly the steps already presented in the last subsection if we assume all informed

equilibria.

[8] A method to solve the dynamic equations has been proposed by Holden and Subrahmanyam (1992, pp. 253 f.).

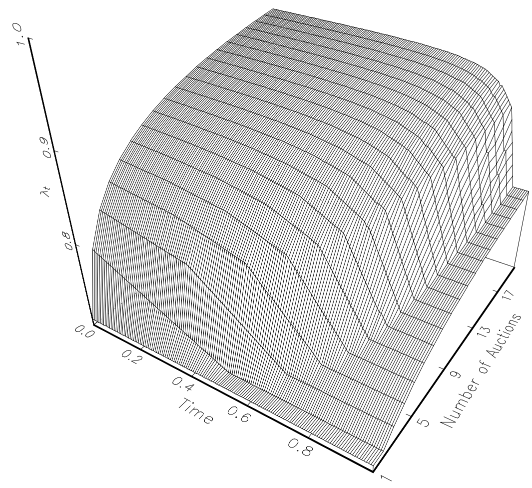


Fig. 2.1: Liquidity in the Kyle (1985) model

investors to be equal with respect to all characteristics. An informed investor maximizes in analogy to (2.30) with $\Omega_p = \{p_1, \ldots, p_t, v\}$:

$$
\begin{aligned}
(2.46)\quad E[\widetilde{\pi}_{t+1}|\Omega_p] &= E[(\widetilde{v} - p_t)\Delta x_t^i + \widetilde{\pi}_{t+1}|\Omega_p] \\
&= E[(\widetilde{v} - p_t)\Delta x_t^i|\Omega_p] + \gamma_t(v - p_t)^2 + \delta_t \\
&= E[(\widetilde{v} - \mu_t - \lambda_t(\Delta x_t^i + (M-1)\Delta\overline{x}_t^i + \widetilde{u}_t))\Delta x_t^i|\Omega_p] \\
&\quad +\gamma_t(v - \mu_t - \lambda_t(\Delta x_t^i + (M-1)\Delta\overline{x}_t^i + u_t))^2 + \delta_t \\
&= (v - \mu_t - \lambda_t\Delta x_t^i + (M-1)\Delta\overline{x}_t^i)\Delta x_t^i \\
&\quad +\gamma_t(v - \mu_t - \lambda_t(\Delta x_t^i + (M-1)\Delta\overline{x}_t^i + u_t))^2 \\
&\quad +\delta_t + \gamma_t\lambda_t^2\sigma_{u^i,t}^2.
\end{aligned}
$$

This gives the first order condition

$$
\begin{aligned}
(2.47)\quad 0 &= v - \mu_t - \lambda_t(\Delta x_t^i + (M-1)\Delta\overline{x}_t^i) - \lambda_t\Delta x_t^i \\
&\quad +2\gamma_t(v - \mu_t - \lambda_t(\Delta x_t^i + (M-1)\Delta\overline{x}_t^i))(-\lambda_t) \\
&= (v - \mu_t - \lambda_t(M-1)\Delta\overline{x}_t^i)(1 - 2\gamma_t\lambda_t) - 2\lambda_t\Delta x_t^i(1 - \gamma_t\lambda_t).
\end{aligned}
$$

As all informed investors are assumed to be equal, the only reasonable conjecture about the other informed investor's behavior is that they behave exactly in the same way, i.e. $\Delta\overline{x}_t^i = \Delta x_t^i$. Inserting this relation and solving for Δx_t^i gives the optimal order size for informed investors:

$$\Delta x_t^i = -\frac{1-2\gamma_t\lambda_t}{\lambda_t(1+M(1-2\gamma_t\lambda_t))}\mu_t + \frac{1-2\gamma_t\lambda_t}{\lambda_t(1+M(1-2\gamma_t\lambda_t))}v. \tag{2.48}$$

Comparing coefficients with the linear equilibrium from (2.26) we see that

$$\begin{aligned} \beta_t &= \frac{1-2\gamma_t\lambda_t}{\lambda_t(1+M(1-2\gamma_t\lambda_t))}, \\ \alpha_t &= -\frac{1-2\gamma_t\lambda_t}{\lambda_t(1+M(1-2\gamma_t\lambda_t))}\mu_t = -\beta_t\mu_t. \end{aligned} \tag{2.49}$$

the second order condition for a maximum

$$-2\lambda_t(1-\gamma_t\lambda_t) < 0 \tag{2.50}$$

remains unchanged from KYLE (1985).

From the price setting behavior of the match maker we get in analogy to (2.36):

$$\begin{aligned} p_t &= E[\widetilde{v}|\Omega_t] \\ &= E[E[\widetilde{v}|\Omega_{t-1}]|\Omega_t] \\ &= E[\widetilde{v}|\Omega_{t-1}] + \frac{Cov[\widetilde{v}, M\Delta\widetilde{x}_t^i + \widetilde{u}_t|\Omega_{t-1}]}{Var[M\Delta\widetilde{x}_t^i + \widetilde{u}_t|\Omega_{t-1}]} \times \\ &\quad \times(\Delta x_t + u_t - E[M\Delta\widetilde{x}_t^i + \widetilde{u}_t|\Omega_{t-1}]) \\ &= p_{t-1} + \frac{MCov[\widetilde{v}, \alpha_t + \beta_t\widetilde{v} + \widetilde{u}_t|\Omega_{t-1}]}{M^2Var[\alpha_t + \beta_t\widetilde{v} + \widetilde{u}_t|\Omega_{t-1}]} \times \\ &\quad \times(\Delta x_t + u_t - ME[\alpha_t + \beta_t\widetilde{v} + \widetilde{u}_t|\Omega_{t-1}]) \\ &= p_{t-1} + \frac{M\beta_t Cov[\widetilde{v}, \widetilde{v}|\Omega_{t-1}]}{M^2\beta_t^2 Var[\widetilde{v}|\Omega_{t-1}] + \sigma_{u^i,t}^2} \times \\ &\quad \times(\Delta\widetilde{x}_t + \widetilde{u}_t - M(\alpha_t - \beta_t p_{t-1})) \\ &= p_{t-1} + \frac{M\beta_t\Sigma_{t-1}}{M^2\beta_t^2\Sigma_{t-1} + \sigma_{u^i,t}^2}(\Delta\widetilde{x}_t + \widetilde{u}_t - M(\alpha_t + \beta_t p_{t-1})) \\ &= p_{t-1}\left(1 - M\beta_t\frac{M\beta_t\Sigma_{t-1}}{M^2\beta_t^2\Sigma_{t-1} + \sigma_{u^i,t}^2}\right) \\ &\quad + \frac{M\beta_t\Sigma_{t-1}}{M^2\beta_t^2\Sigma_{t-1} + \sigma_{u^i,t}^2}M\alpha_t + \frac{M\beta_t\Sigma_{t-1}}{M^2\beta_t^2\Sigma_{t-1} + \sigma_{u^i,t}^2}(\delta x_t + u_t). \end{aligned}$$

Comparing coefficients with (2.25) we get

$$
\begin{aligned}
\lambda_t &= \frac{M\beta_t\Sigma_{t-1}}{M^2\beta_t^2\Sigma_{t-1}+\sigma^2_{u^i,t}} \\
\mu_t &= p_{t-1}\left(1 - M\beta_t\frac{M\beta_t\Sigma_{t-1}}{M^2\beta_t^2\Sigma_{t-1}+\sigma^2_{u^i,t}}\right) \\
&\quad + \frac{M\beta_t\Sigma_{t-1}}{M^2\beta_t^2\Sigma_{t-1}+\sigma^2_{u^i,t}} M\alpha_t \\
&= p_{t-1}(1 - M\beta_t\lambda_t) - M\alpha_t\lambda_t,
\end{aligned}
\tag{2.51}
$$

which with (2.49) solves to

$$
\begin{aligned}
\lambda_t &= \frac{M\beta_t\Sigma_{t-1}}{M^2\beta_t^2\Sigma_{t-1}+\sigma^2_{u^i,t}}, \\
\beta_t &= \frac{1-2\gamma_t\lambda_t}{\lambda_t(1+M(1-2\gamma_t\lambda_t))}, \\
\mu_t &= p_{t-1}, \\
\alpha_t &= -\beta_t p_{t-1}.
\end{aligned}
\tag{2.52}
$$

The variance of the liquidation value for the uninformed investors from observing the prices is given by

$$
\begin{aligned}
\Sigma_t &= Var[\widetilde{v}|p_t] = Var[\widetilde{v}|\Omega_t] \\
&= Var[\widetilde{v}|\Omega_{t-1}] - \frac{Cov[\widetilde{v}, \Delta\widetilde{x}_t + \widetilde{u}_t|\Omega_{t-1}]^2}{Var[\Delta\widetilde{x}_t + \widetilde{u}_t|\Omega_{t-1}]} \\
&= \Sigma_{t-1} - \frac{M^2 Cov[\widetilde{v}, \Delta\widetilde{x}_t^i|\Omega_{t-1}]^2}{M^2 Var[\Delta\widetilde{x}_t^i|\Omega_{t-1}] + \sigma^2_{u^i,t}} \\
&= \Sigma_{t-1} - \frac{M^2 Cov[\widetilde{v}, \beta_t(\widetilde{v}-p_{t-1})|\Omega_{t-1}]^2}{M^2 Var[\beta_t(\widetilde{v}-p_{t-1}) + \sigma^2_{u^i,t}|\Omega_{t-1}]} \\
&= \Sigma_{t-1} - \frac{M^2\beta_t^2\Sigma_{t-1}^2}{M^2\beta_t^2\Sigma_{t-1}+\sigma^2_{u^i,t}} \\
&= \Sigma_{t-1} - M\lambda_t\beta_t\Sigma_{t-1} \\
&= (1 - M\lambda_t\beta_t)\Sigma_{t-1}.
\end{aligned}
\tag{2.53}
$$

As in KYLE (1985) the information is gradually incorporated into the price as the variance is strictly decreasing over time. From (2.46) we see that

$$\delta_{t-1} = \delta_t + \gamma_t \lambda_t \sigma^2_{u^i,t} \tag{2.54}$$

$$\begin{aligned}
\gamma_{t-1}(v-p_{t-1})^2 &= (v-\mu_t-\lambda_t(\Delta x_t^i+(M-1)\Delta\overline{x}_t^i))\Delta x_t^i \\
&\quad +\gamma_t(v-\mu_t-\lambda_t(\Delta x_t^i+(M-1)\Delta\overline{x}_t^i))^2 \\
&= (v-\mu_t-M\lambda_t\Delta x_t^i)\Delta x_t^i \\
&\quad +\gamma_t(v-\mu_t-M\lambda_t\Delta x_t^i)^2 \\
&= (v-\mu_t-M\lambda_t\beta_t(v-p_{t-1}))\beta_t(v-p_{t-1}) \\
&\quad +\gamma_t(v-\mu_t-M\lambda_t\beta_t(v-p_{t-1}))^2 \\
&= (v-p_{t-1})^2(1-M\lambda_t\beta_t)\beta_t \\
&\quad +\gamma_t(v-p_{t-1})^2(1-M\lambda_t\beta_t)^2 \\
&= (v-p_{t-1})^2((1-M\lambda_t\beta_t)\beta_t \\
&\quad +\gamma_t(1-M\lambda_t\beta_t)\beta_t)^2).
\end{aligned} \tag{2.55}$$

Hence we find that

$$\begin{aligned}
\gamma_{t-1} &= (1-M\lambda_t\beta_t)\beta_t+\gamma_t(1-M\lambda_t\beta_t)\beta_t)^2 \\
&= (1-M\lambda_t\frac{1-2\gamma_t\lambda_t}{\lambda_t(1+M(1-2\gamma_t\lambda_t))})\beta_t \\
&\quad +\gamma_t(1-M\lambda_t\frac{1-2\gamma_t\lambda_t}{\lambda_t(1+M(1-2\gamma_t\lambda_t))})^2 \\
&= \frac{1}{1+M(1-2\gamma_t\lambda_t)}\beta_t+\frac{1}{(1+M(1-2\gamma_t\lambda_t))^2}\gamma_t \\
&= \frac{(1+M(1-2\gamma_t\lambda_t))\beta_t+\gamma_t}{(1+M(1-2\gamma_t\lambda_t))^2} \\
&= \frac{(1+M(1-2\gamma_t\lambda_t))\frac{1-2\gamma_t\lambda_t}{\lambda_t(1+M(1-2\gamma_t\lambda_t))}+\gamma_t}{(1+M(1-2\gamma_t\lambda_t))^2} \\
&= \frac{1-2\gamma_t\lambda_t+\gamma_t\lambda_t}{\lambda_t(1+M(1-2\gamma_t\lambda_t))^2} \\
&= \frac{1-\gamma_t\lambda_t}{\lambda_t(1+M(1-2\gamma_t\lambda_t))^2}.
\end{aligned} \tag{2.56}$$

Holden and Subrahmanyam (1992, pp. 253 ff.) provide a method how to solve the dynamic system of (2.52), (2.53), (2.54) and (2.56). While a single informed investor uses its monopoly power to hold λ_t constant over time, competition between informed investors forces them to trade more aggressively on their information in the first auctions. In consequence much information will be revealed in these first auctions, resulting in a lower liquidity of the market and a more quickly dropping variance. Information is revealed much faster than with a single informed investor. As nearly all information has been revealed in the first auctions later trades are not very informative and λ_t fast drops near zero, i.e. the market becomes very liquid in later auctions. As the number of informed investors increases, information is revealed faster. In the limit as M reaches infinity, prices are fully revealing in an instant, i.e. with the first trade, which corresponds to the case of perfect competition.

With perfect competition therefore the market is efficient in the strong form, otherwise only in the semistrong form. If competition is too strong, profits from trading on this information are very low and may not cover the costs of acquiring information, although the prices will never be fully revealing as $\Sigma_t > 0$ in all auctions due to the noise of uninformed investors. This leads to the problem pointed out by Grossman and Stiglitz (1980) that markets cannot be fully revealing if information is costly. If the profits from information acquisition are too low to cover these costs, no investor will acquire information and prices are not informative. But this on the other hand gives incentives for investors to acquire information and make profits (monopolistic case), hence no equilibrium will exist.

Figures 2.2 - 2.5 illustrate the findings of this model.

2.2 Auctions with strategic uninformed investors

2.2.1 Markets with a single asset

Thus far we assumed that uninformed investors trade for exogenous reasons and do not respond to price changes, i.e. they were noise traders. We will now assume

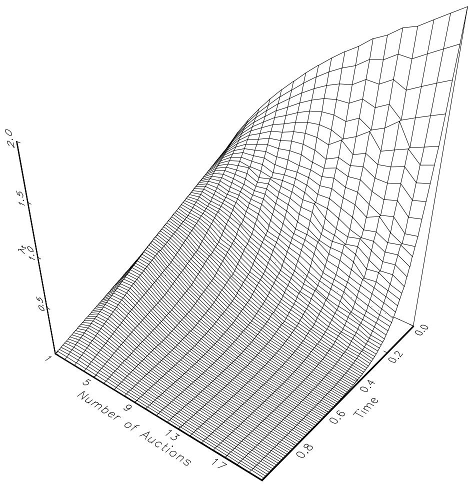


Fig. 2.2: Liquidity with different number of auctions

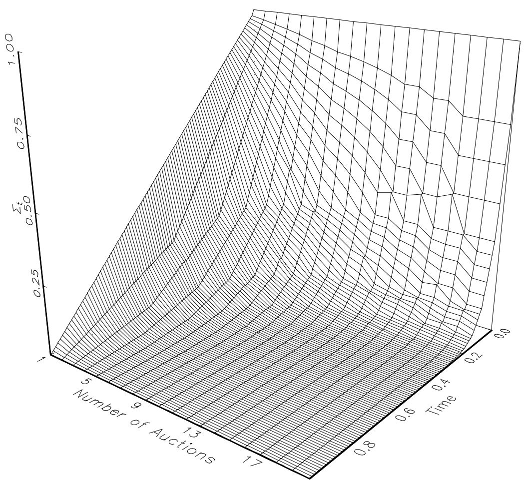


Fig. 2.3: Informativeness of prices with different number of auctions

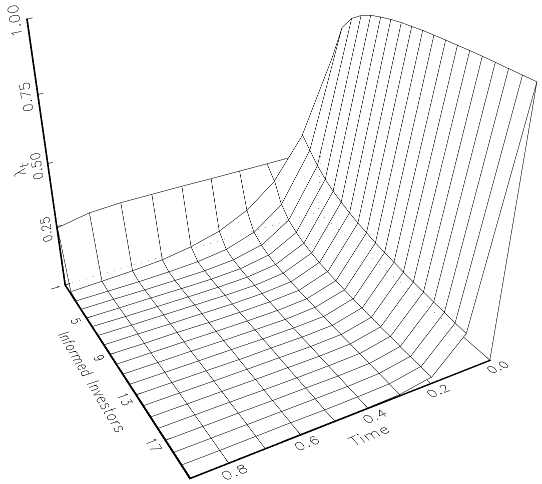


Fig. 2.4: Liquidity with different number of informed investors

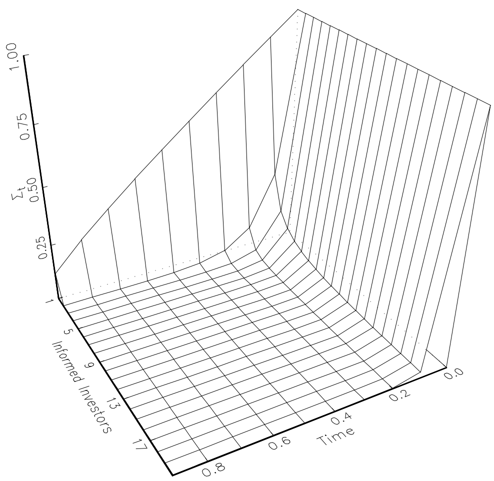


Fig. 2.5: Informativeness of prices with different number of informed investors

that they are risk averse investors holding a portfolio consisting of the risky asset and a riskless asset. We assume that the uninformed investors face a portfolio imbalance, w_j. This imbalance can be due to changed prices, new information or liquidity needs. They will not only trade this imbalance, but act to maximize their expected utility, i.e. they are hedgers.

We assume that the portfolio imbalance is a normally distributed random variable with mean zero and variance $\sigma_w^2 > 0$. The imbalance is assumed to be independent between investors and from any other relevant variable.[9]

$$w_j \sim N(0, \sigma_w^2) \tag{2.57}$$

for all $j = 1, \ldots, N$, where N denotes the number of uninformed investors.

There is only a single auction before the asset is liquidated. Again only linear equilibria are considered, the orders of uninformed investors are also linear in their portfolio imbalance:

$$u_j = \eta + \xi w_j \quad j = 1, \ldots, N, \tag{2.58}$$

$$x_i = \alpha + \beta v \quad i = 1, \ldots, M, \tag{2.59}$$

$$p = \mu + \lambda(x + u), \tag{2.60}$$

where $x = \sum_{i=1}^M x_i$ and $u = \sum_{j=1}^N u_j$.

The derivation of the equilibrium follows the same steps as before. The informed investors, assumed to be risk neutral, maximize their expected profits:

$$\begin{aligned} \widetilde{\pi}_i &= (\widetilde{v} - p)x_i = (\widetilde{v} - \mu - \lambda(\widetilde{x} + \widetilde{u}))\, x_i \\ &= \left(\widetilde{v} - \mu - \lambda\left(x_i + (M-1)\overline{x}_i + N\eta + \xi \sum_{j=1}^N \widetilde{w}_j\right)\right) x_i. \end{aligned} \tag{2.61}$$

$$E\left[\widetilde{\pi}_i | v\right] = (v - \mu - \lambda x_i - \lambda(M-1)\overline{x}_i - \lambda\eta N)\, x_i. \tag{2.62}$$

Maximizing this expression gives the following first order condition:

$$\begin{aligned} 0 &= v - \mu - \lambda x_i - \lambda(M-1)\overline{x}_i\lambda\eta N - \lambda x_i \\ &= v - \mu - 2\lambda x_i - \lambda(M-1)\overline{x}_i\lambda\eta N. \end{aligned} \tag{2.63}$$

[9] The model presented here follows Spiegel and Subrahmanyam (1992). Although they provide a more general framework, we restrict it here to make the results comparable to the models presented in the previous sections.

The second order condition $-2\lambda = 0$ states again the need for a positive solution for the liquidity parameter λ. With $\overline{x}_i = x_i$ as the only rational conjecture of an informed investor about the other informed investors trading decisions this solves to

$$x_i = -\frac{1}{(M+1)\lambda}(\mu + \lambda\eta N) + \frac{1}{(M-1)\lambda}. \tag{2.64}$$

By comparing with (2.59) we see that

$$\begin{aligned} \beta &= \frac{1}{\lambda(M+1)}, \\ \alpha &= -\frac{1}{(M+1)\lambda}(\mu + \lambda\eta N) = -\beta(\mu + \lambda\eta N). \end{aligned} \tag{2.65}$$

The price setting of the match maker gives us

$$\begin{aligned} p &= E\left[\widetilde{v}|x+u\right] \\ &= E[\widetilde{v}] + \frac{Cov[\widetilde{v}, \widetilde{x}+\widetilde{u}]}{Var[\widetilde{x}+\widetilde{u}]}(x + u - E[\widetilde{x}+\widetilde{u}]) \\ &= p_0 + \frac{Cov\left[\widetilde{v}, M(\alpha+\beta\widetilde{v}) + N\eta + \xi\sum_{j=1}^{N}\widetilde{w}_j\right]}{Var\left[M(\alpha+\beta\widetilde{v}) + N\eta + \xi\sum_{j=1}^{N}\widetilde{w}_j\right]} \times \\ &\quad \times \left(x + u - M(\alpha + \beta E[\widetilde{v}]) - N\eta - \xi\sum_{j=1}^{N}E[\widetilde{w}_j]\right) \\ &= p_0 + \frac{M\beta Cov[\widetilde{v},\widetilde{v}]}{M^2\beta^2 Var[\widetilde{v}] + \xi^2 N\sigma_w^2}(x + u - M(\alpha+\beta p_0) - N\eta) \\ &= p_0 + \frac{M\beta\Sigma_0}{M^2\beta^2\Sigma_0 + N\xi^2\sigma_w^2}(x + u - M(\alpha+\beta p_0) - N\eta) \\ &= p_0 - \frac{M\beta\Sigma_0}{M^2\beta^2\Sigma_0 + N\xi^2\sigma_w^2}M(\alpha+\beta p_0) + N\eta) \\ &\quad + \frac{M\beta\Sigma_0}{M^2\beta^2\Sigma_0 + N\xi^2\sigma_w^2}(x+u). \end{aligned} \tag{2.66}$$

Comparing coefficients with (2.60) results in

$$\begin{aligned} \lambda &= \frac{M\beta\Sigma_0}{M^2\beta^2\Sigma_0 + N\xi^2\sigma_w^2}, \\ \mu &= p_0 - \frac{M\beta\Sigma_0}{M^2\beta^2\Sigma_0 + N\xi^2\sigma_w^2}(M(\alpha+\beta p_0) + N\eta) \\ &= p_0 - \lambda(M(\alpha+\beta p_0) + N\eta). \end{aligned} \tag{2.67}$$

By inserting for β we get

$$\begin{aligned}\lambda &= \frac{\frac{M\Sigma_0}{\lambda(M+1)}}{\frac{M^2\Sigma_0}{\lambda^2(M+1)^2}+\xi^2 N\sigma_w^2} \\ &= \frac{M\Sigma_0(M+1)\lambda}{M^2\Sigma_0+(M+1)^2\lambda^2\xi^2 N\sigma_w^2}\end{aligned}$$

$$M^2\Sigma_0+(M+1)^2\lambda^2\xi^2 N\sigma_w^2 = M\Sigma_0(M+1)$$

$$\begin{aligned}\lambda^2 &= \frac{M(M+1)\Sigma_0-M^2\Sigma_0}{(M+1)^2\xi^2 N\sigma_w^2} \\ &= \frac{M\Sigma_0}{(M+1)^2\xi^2 N\sigma_w^2}.\end{aligned}$$

The second order condition implies λ to be positive, i.e. we find

$$\begin{aligned}\lambda &= \sqrt{\frac{\Sigma_0}{N\sigma_w^2}\frac{M}{(M+1)^2\xi^2}}, \\ \beta &= \sqrt{\frac{\sigma_w^2 N}{\Sigma_0}\frac{\xi^2}{M}}.\end{aligned} \tag{2.68}$$

With these results we get from (2.65)

$$\alpha = -\beta p_0, \tag{2.69}$$

$$\mu = p_0 - N\lambda\eta. \tag{2.70}$$

Uninformed investors maximize their expected utility from holding a portfolio. For simplicity we do not consider the utility derived from holding the optimal portfolio, but only the utility from a deviation from this portfolio. The deviation consists of the size of the portfolio imbalance, w_j and the amount traded, u_j adjusted by the price paid. The wealth of a deviation from the optimal portfolio is:

$$
\begin{aligned}
(2.71)\qquad \widetilde{W}_j &= \widetilde{v}(u_j + \widetilde{w}_j) - u_j p \\
&= \widetilde{v}(u_j + \widetilde{w}_j) - u_j(p_0 + \lambda(x+u) - N\lambda\eta) \\
&= v(u_j + w_j) - u_j \Bigg(p_0 - N\lambda\eta \\
&\quad +\lambda\Bigg(M\beta(\widetilde{v} - p_0) + u_j + (N-1)\eta + \xi \sum_{k\neq j} \widetilde{w}_j) \Bigg)\Bigg) \\
&= \widetilde{v}(u_j + \widetilde{w}_j) - u_j \Bigg(p_0 - N\lambda\eta + \lambda\beta M\widetilde{v} - \lambda\beta p_0 \\
&\quad +\lambda u_j - \lambda\eta N - \lambda\xi \sum_{k\neq j} \widetilde{w}_j \Bigg) \\
&= \widetilde{v}(u_j + \widetilde{w}_j) - u_j \Bigg(p_0(1 - \lambda\beta M) - N\lambda\eta + \lambda\beta M\widetilde{v} \\
&\quad +\lambda u_j - \lambda\eta N - \lambda\xi \sum_{k\neq j} \widetilde{w}_j \Bigg).
\end{aligned}
$$

Therewith we find

$$
\begin{aligned}
(2.72)\qquad E[\widetilde{W}_j|w_j] &= E[\widetilde{v}|w_j](w_j + u_j) - u_j \Bigg(p_0(1 - M\lambda\beta) \\
&\quad +M\lambda\beta E[\widetilde{v}|w_j] + \lambda u_j - N\lambda\eta - \lambda\xi \sum_{k\neq j} E[\widetilde{w}_k|w_j] \Bigg) \\
&= p_0(w_j + u_j) - u_j \left(p_0(1 - M\lambda\beta) + M\lambda\beta p_0 \right.\\
&\quad \left. +\lambda u_j - N\lambda\eta\right) \\
&= p_0 w_j + p_0 u_j - p_0 u_j - \lambda u_j^2 - N\lambda\eta u_j \\
&= -\lambda u_j^2 + p_0 w_j - N\lambda\eta u_j,
\end{aligned}
$$

$$
\begin{aligned}
(2.73)\qquad Var[\widetilde{W}_j|w_j] &= (w_j + v_j)^2\Sigma_0 \\
&\quad +u_j^2\left(M^2\lambda^2\beta^2\Sigma_0 + (N-1)\lambda^2\xi^2\sigma_w^2\right) \\
&\quad -2Cov[(u_j + w_j)v, Mu_j v\lambda\beta] \\
&= (w_j + v_j)^2\Sigma_0 \\
&\quad +u_j^2\left(M^2\lambda^2\beta^2\Sigma_0 + (N-1)\lambda^2\xi^2\sigma_w^2\right) \\
&\quad -2(u_j + w_j)\lambda\beta u_j M\Sigma_0.
\end{aligned}
$$

The expected utility of the uninformed investors is given by

$$E\left[U(W_j)|w_j\right] = U\left(E[W_j|w_j] - \frac{1}{2}zVar[W_j|w_j]\right). \tag{2.74}$$

Maximizing (2.74) after having inserted (2.72) and (2.73) gives the following first order condition:

$$\begin{aligned} 0 &= -2\lambda u_j - N\lambda\eta - \frac{1}{2}z\Bigg(2(w_j+u_j)\Sigma_0 \\ &\quad +2u_j\left(M^2\lambda^2\beta^2\Sigma_0 - 2M\lambda\beta\Sigma_0 + (N-1)\lambda^2\xi^2\sigma_w^2\right) \\ &\quad -2(w_jM\lambda\beta\Sigma_0 + 2Mu_j\lambda\beta\Sigma_0)\Bigg) \\ &= u_j\left(-2\lambda - z\Sigma_0\left(M^2\lambda^2\beta^2 - 2M\lambda\beta + 1 + (N-1)\lambda^2\xi^2\sigma_w^2\right)\right) \\ &\quad -zw_j\Sigma_0(1-M\lambda\beta) - N\lambda\eta \\ &= u_j\left(-2\lambda - z\Sigma_0(1-M\lambda\beta)^2 - z(N-1)\lambda^2\xi^2\sigma_w^2\right) \\ &\quad -zw_j\Sigma_0(1-M\lambda\beta) - N\lambda\eta. \end{aligned} \tag{2.75}$$

The second order condition for a maximum $-2\lambda - z\Sigma_0(1-M\lambda\beta)^2 - z(N-1)\lambda^2\xi^2\sigma_w^2 < 0$ is fulfilled as $\lambda > 0$ and all other terms are positive. Solving for the optimal trade size of uninformed investors gives

$$\begin{aligned} u_j &= -\frac{N\lambda\eta}{2\lambda + z\Sigma_0(1-M\lambda\beta)^2 + z(N-1)\lambda^2\xi^2\sigma_w^2} \\ &\quad -\frac{z\Sigma_0(1-M\lambda\beta)}{2\lambda + z\Sigma_0(1-M\lambda\beta)^2 + z(N-1)\lambda^2\xi^2\sigma_w^2}w_j. \end{aligned} \tag{2.76}$$

By comparing coefficients with (2.58) we see that

$$\begin{aligned} \xi &= -\frac{z\Sigma_0(1-M\lambda\beta)}{2\lambda + z\Sigma_0(1-M\lambda\beta)^2 + z(N-1)\lambda^2\xi^2\sigma_w^2}, \\ \eta &= -\frac{N\lambda\eta}{2\lambda + z\Sigma_0(1-M\lambda\beta)^2 + z(N-1)\lambda^2\xi^2\sigma_w^2}. \end{aligned} \tag{2.77}$$

As from the second order condition $\lambda > 0$ and the denominator is positive the last equation implies that

$$\eta = 0. \tag{2.78}$$

The first equation has to be solved for ξ:

$$z(N-1)\lambda^2\xi^3\sigma_w^2 + (z\Sigma_0(1-M\lambda\beta)^2 + 2\lambda)\xi + z\Sigma_0(1-M\lambda\beta) = 0 \tag{2.79}$$

As all coefficients are positive[10] the solution for ξ has to be negative. Inserting

[10] From (2.65) we know that $1 - M\lambda\beta = 1 - \frac{M}{M+1} = \frac{1}{M+1} > 0$.

for β and λ we get

$$
\begin{aligned}
(2.80)\qquad 0 &= z(N-1)\lambda^2\sigma_w^2\xi^3 + \left(z\Sigma_0\left(1-\frac{M}{M+1}\right)^2 + 2\lambda\right)\xi \\
&\quad + z\Sigma_0\left(1-\frac{M}{M+1}\right) \\
&= z\frac{\Sigma_0 M(N-1)\sigma_w^2}{N\sigma_w^2(M+1)^2\xi^2}\xi^3 \\
&\quad + \left(\frac{z\Sigma_0}{(M+1)^2} + 2\sqrt{\frac{\Sigma_0}{N\sigma_w^2}\frac{M}{(M+1)^2\xi^2}}\right)\xi + \frac{z\Sigma_0}{M+1} \\
&= z\Sigma_0\frac{M(N-1)+N}{N(M+1)^2}\xi + \frac{z\Sigma_0 - 2\sqrt{\frac{\Sigma_0 M}{\sigma_w^2 N}}}{M+1}.
\end{aligned}
$$

$$
(2.81)\qquad \xi = -\frac{\left(z\Sigma_0\sqrt{\sigma_w^2 N} - 2\sqrt{\Sigma_0 M}\right)N(M+1)}{\sqrt{\sigma_w^2 N}z\Sigma_0(M(N-1)+N)}.
$$

As ξ has to be negative, a linear equilibrium only exists if the numerator is positive as the denominator always is positive, i.e. if

$$
\begin{aligned}
z\Sigma_0\sqrt{\sigma_w^2 N} - 2\sqrt{\Sigma_0 M} &> 0, \\
(2.82)\qquad M &< \frac{1}{4}Nz^2\Sigma_0\sigma_w^2.
\end{aligned}
$$

The existence depends on a not too large fraction of informed investors in the market. The higher the risk aversion, uncertainty about the liquidation value, Σ_0, and dispersion of portfolio imbalances the more informed investors can participate in the market. This can be explained by the behavior of the uninformed investors, as in these cases their need to trade is higher and they are willing to make larger losses to rebalance their portfolios.

The equilibrium can be written as

$$
\begin{aligned}
(2.83)\qquad p &= p_0 + \lambda(x+u), \\
(2.84)\qquad x_i &= \beta(v-p_0), \\
(2.85)\qquad u_j &= \xi w_j,
\end{aligned}
$$

where λ, β and ξ are given by (2.68) and (2.81).

The variance of the liquidation value after observing the price is given by

$$
\begin{aligned}
\Sigma_1 &= Var[\widetilde{v}|p] = Var[\widetilde{v}] - \frac{Cov[\widetilde{v},\widetilde{p}]^2}{Var[\widetilde{p}]} \\
&= \Sigma_0 - \frac{Cov[\widetilde{v}, p_0 + \lambda(\widetilde{x}+\widetilde{u})]^2}{Var[p_0 + \lambda(\widetilde{x}+\widetilde{u})]} \\
&= \Sigma_0 - \frac{\lambda^2 Cov[\widetilde{v}, M\beta(\widetilde{v}-p_0) + \xi\sum_{j=1}^N \widetilde{w}_j]^2}{\lambda^2 Var[M\beta(\widetilde{v}-p_0) + \xi\sum_{j=1}^N \widetilde{w}_j]} \\
&= \Sigma_0 - \frac{M^2\beta^2\Sigma_0^2}{M^2\beta^2\Sigma_0^2 + \xi^2 N\sigma_w^2} \\
&= \frac{\xi^2 N\sigma_w^2\Sigma_0}{M^2\beta^2\Sigma_0^2 + \xi^2 N\sigma_w^2} \\
&= \frac{\xi^2 N\sigma_w^2\Sigma_0}{\frac{M\beta\Sigma_0}{\lambda}} \\
&= \frac{\lambda\xi^2 N\sigma_w^2}{M\beta} \\
&= \frac{\lambda\frac{\Sigma_0}{\sigma_w^2 N}\frac{M}{(M+1)^2\lambda^2}N\sigma_w^2}{M\beta} \\
&= \frac{\Sigma_0}{(M+1)^2\lambda\beta} \\
&= \frac{\Sigma_0}{M+1}.
\end{aligned}
\tag{2.86}
$$

The informativeness of the price does only depend on the number of informed investors. Competition between them to trade on their information results in a high revelation of information.

λ is decreasing in the risk aversion of uninformed investors. If they are more risk averse their wish to trade is larger, although they make losses from trading. This increased trading of uninformed investors enables informed investors to hide their trades better, the order flow becomes less informative. The same holds if the portfolio imbalance increases, i.e. σ_w^2 increases. The need to offset the imbalance at least partially is increased and trades of uniformed investors increase.

If the number uninformed investors increases the volatility of the price increases:

$$\begin{aligned}
Var[\widetilde{p}] &= Var[p_0 + \lambda(\widetilde{x} + \widetilde{u})] \\
&= \lambda^2 Var\left[M\beta(\widetilde{v} - p_0) + \xi \sum_{j=1}^{N} \widetilde{w}_j\right] \\
&= M^2\lambda^2\beta^2\Sigma_0 + \lambda^2\xi^2 N^2 \sigma_w^2.
\end{aligned} \tag{2.87}$$

As the price of the trade is not known at the time the order is submitted, the risk from trading is increased for uninformed investors. If they are not too risk averse this effect will be more than compensated by the additional orders from uninformed investors, i.e. adverse selection costs are smaller and λ will be decreasing in the risk aversion. If the uninformed investors are more risk averse, however, the reduced order from every single uninformed investor dominates the additional order flow generated by more uninformed investors in the market and λ increases with more uninformed investors. But if the number is sufficiently enlarged the additional order flow will dominate again and λ decreases. Figure 2.6 illustrates this finding.

By increasing the number of informed investors, competition between them is increased as has been shown in the previous section, resulting in an increased λ. On the other hand as we saw in (2.86) the informativeness of prices is increased, i.e. the adverse selection costs of the uniformed investors are reduced. If the uninformed investors are very risk averse this will induce them to trade more actively, thereby compensating the increased λ, which becomes decreasing. If they are less risk averse their increased trading cannot compensate the effect of competition and λ is increasing as illustrated in figure 2.7.

If the uncertainty about the liquidation value, Σ_0, increases, the adverse selection costs and the variance of the price will increase, hence uninformed investors will scale back their trades, this increases λ. On the other hand the benefits from offsetting an imbalance is increased. If the uninformed investors are very risk averse the first effect may dominate and λ will be increasing in Σ_0. If the uninformed investors are less risk averse the second effect will dominate for small uncertainties and λ decreases, if the uncertainty is too large, however, the first

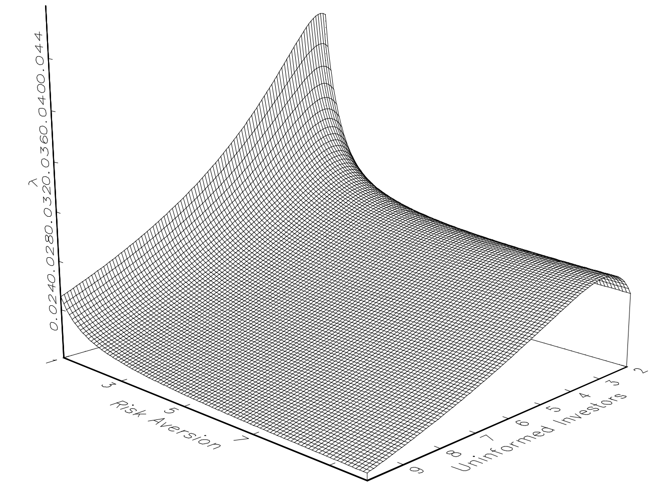


Fig. 2.6: Market liquidity with varying number of uninformed investors

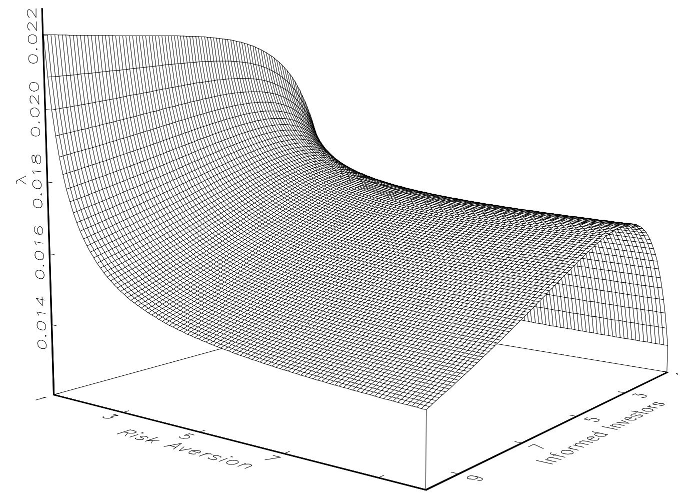


Fig. 2.7: Market liquidity with varying number of informed investors

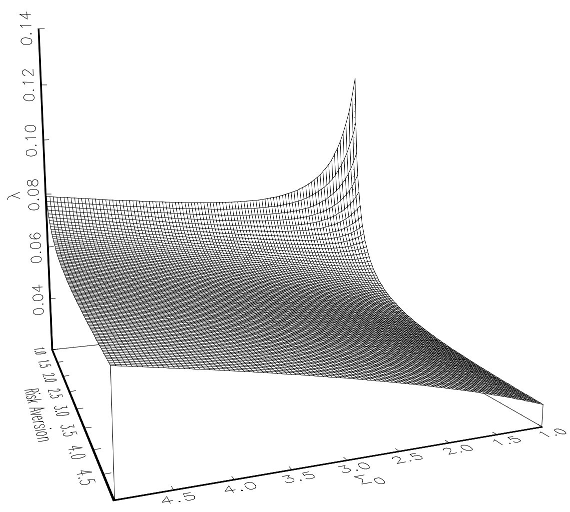


Fig. 2.8: Market liquidity with different uncertainties about the liquidation value

effect again will dominate and λ increses again. This behavior is illustrated in figure 2.8.[11]

Spiegel and Subrahmanyam (1992) provide a more general framework by allowing the signals the informed investors receive to be imperfect and differ between investors. Compared to the more restrictive version presented here no additional insights can be gained, whereas the tractability of the derivation is reduced significantly. Introducing risk averse informed investors does also not change the results derived here, but becomes in a general framework unhandable.[12]

A central element for calculating conditional moments and hence deriving the equilibrium is the assumption of normally distributed random variables. This assumption cannot be lifted without facing the problem that the results may

[11] These results are formally shown in Spiegel and Subrahmanyam (1992).

[12] Subrahmanyam (1991) presents a model with risk averse informed investors. He uses only noise traders, but no hedgers in his model. The results he derives are very similar to those with risk neutral informed investors, so that no new insights can be expected.

change. Foster and Viswanathan (1993) derive similar results within the more general class of elliptically contoured distributions instead of using the special case of a normal distribution, a generalization to other distributions cannot be made.

All models presented in this part considered only linear equilibria, nothing can be said about the existence and properties of non-linear equilibria. Kyle (1985, p. 1322) suspects that there exist no non-linear equilibrium, but he is not able to proof his suspicion.

There exist several extensions of these models, which we will not consider in detail here. Bondarenko (1999) relaxes the assumption of only a single match maker by investigating a model with a given number of match makers that are not restricted to quote prices such that they make zero expected profits, e.g. by applying strategic price setting to maximize their profits. He finds the results as derived above to hold with the number of match makers reaching infinity, i.e. if they behave competitively.

By adding another source of information on the value of the asset for the market maker, Jain and Mirman (1999) show that the match maker sets more informative prices and reduces the profits of the informed investor. Chau (1998) considers the trading of many pure noise traders, a single risk averse informed investor and a risk averse market maker. Although this model assumes a dealer rather than an auction market, it helps understanding the adjustment of prices to new information. In contrast to models with risk neutral market participants, he finds that risk aversion cause prices to overshoot before gradually adjusting to the new fundamental value.

2.2.2 Auction markets with multiple assets

Using the framework developed by Holden and Subrahmanyam (1992) for a single trade, Caballe and Krishnan (1994) extend the model to the more realistic case of investors being able to trade multiple assets.

Each investor receives a noisy signal of the liquidation values $\widetilde{v}$, such that[13]

$$\begin{aligned} \widetilde{s}_i &= \widetilde{v} + \widetilde{\varepsilon}_i, \\ \widetilde{\varepsilon}_i &\sim N\left(0, \Sigma_\varepsilon\right), \end{aligned} \tag{2.88}$$

where $\widetilde{s}_i = (\widetilde{s}_i^1, \ldots, \widetilde{s}_i^L)$ denotes the vector of signals investor i receives for each of the L assets, $\widetilde{v} = (\widetilde{v}^1, \ldots, \widetilde{v}^L)$ the vector of the liquidation values, $\widetilde{\varepsilon}_i = (\widetilde{\varepsilon}_i^1, \ldots, \widetilde{\varepsilon}_i^L)$ the vector of noise terms as received by investor i and Σ_ε the covariance matrix of noise terms. The liquidation value is multivariate normally distributed with mean $p_0 = (p_0^1, \ldots, p_0^L)$ and covariance matrix Σ_0:

$$\widetilde{v} \sim N(p_0, \Sigma_0). \tag{2.89}$$

The imbalance of orders from uninformed investors, $\widetilde{u} = (\widetilde{u}^1, \ldots, \widetilde{u}^L)$ is normally distributed with mean zero and covariance matrix Σ_u:

$$\widetilde{u} \sim N(0, \Sigma_u). \tag{2.90}$$

The match maker sets prices such that with $\widetilde{p} = (\widetilde{p}^1, \ldots, \widetilde{p}^L)$ denoting the vector of prices and the aggregate order flow from informed investors $\widetilde{x} = (\widetilde{x}^1, \ldots, \widetilde{x}^L)$, $\widetilde{x}^j = \sum_{i=1}^M \widetilde{x}_i^j$

$$\widetilde{p} = E[\widetilde{v} | \widetilde{x} + \widetilde{u}]. \tag{2.91}$$

The profits from trading for informed investors are the sum of the profits made from trading each asset:

$$\widetilde{\pi}_i = (\widetilde{v} - \widetilde{p})' \widetilde{x}_i. \tag{2.92}$$

As before the analysis is restricted to a linear equilibrium:

$$\widetilde{p} = \mu + \Lambda(\widetilde{x} + \widetilde{u}), \tag{2.93}$$

$$\widetilde{x}_i = \alpha + B(\widetilde{s}_i - p_0), \tag{2.94}$$

[13] With the exception of this assumption, where KYLE (1985) assumes perfect knowledge of the liquidation value, the assumptions are similar to those in KYLE (1985). It has to be noticed that all variables represent vectors or matrices of the considered assets.

where $\mu = (\mu_1, \ldots, \mu_L)'$ and $\alpha = (\alpha_1, \ldots, \alpha_L)'$ are vectors of constants and $\Lambda = (\lambda^{11}, \ldots, \lambda^{1L}; \ldots; \lambda^{L1}, \ldots, \lambda^{LL})$ and $B = (\beta^{11}, \ldots, \beta^{1L}; \ldots; \beta^{L1}, \ldots, \beta^{LL})$ are positive definite and nonsingular matrices of constants.

Informed investors maximize expected utility from trading, if we assume them to be risk neutral this is equivalent to maximizing expected profits. These are given by

$$
\begin{aligned}
(2.95) \qquad E[\widetilde{\pi}_i|s_i] &= E\left[(\widetilde{v}-\widetilde{p})'\widetilde{x}_i|s_i\right] \\
&= E\left[(\widetilde{v}-\mu-\Lambda(\widetilde{x}+\widetilde{u}))'x_i|s_i\right] \\
&= E\left[\left(\widetilde{v}-\mu-\Lambda\left(x_i+\sum_{j=1,j\neq i}^{M} x_j+\widetilde{u}\right)\right)' x_i|s_i\right] \\
&= E\Bigg[\Bigg(\widetilde{v}-\mu \\
&\quad -\Lambda\left(x_i+\sum_{j=1,j\neq i}^{M}(\alpha+B(\widetilde{s}_j-p_0)+\widetilde{u}\right)\Bigg)' x_i|s_i\Bigg].
\end{aligned}
$$

Differentiating gives us the first order condition

$$
\begin{aligned}
(2.96) \qquad 0 &= E[\widetilde{v}|s_i]-\mu-\Lambda x_i \\
&\quad -\Lambda\sum_{j=1,j\neq i}^{M}(\alpha+B(E[\widetilde{s}_j|s_i]-p_0)-\Lambda x_i \\
&= E[\widetilde{v}|s_i]-\mu-2\Lambda x_i \\
&\quad -\Lambda B\sum_{j=1,j\neq i}^{M}(E[\widetilde{s}_j|s_i]-p_0)-(M-1)\Lambda\alpha.
\end{aligned}
$$

The second order condition $-2\Lambda < 0$ is fulfilled by the assumption of Λ being positive definite. We further get

$$(2.97) \qquad E[\widetilde{v}|s_i] = E[\widetilde{s}_i-\widetilde{\varepsilon}_i|s_i] = s_i,$$

$$
\begin{aligned}
(2.98) \quad E[\widetilde{s}_j|s_i]-p_0 &= E[\widetilde{s}_j-p_0|s_i-p_0] \\
&= E[\widetilde{s}_j-p_0]+Cov[\widetilde{s}_j-p_0,\widetilde{s}_i-p_0]Var(\widetilde{s}_j-p_0)^{-1} \\
&\quad \times(s_i-p_0-E[\widetilde{s}_i-p_0]) \\
&= Cov[\widetilde{s}_j,\widetilde{s}_i]Var[\widetilde{s}_j]^{-1}(s_i-p_0) \\
&= \Sigma_s\Sigma_\varepsilon^{-1}(s_i-p_0),
\end{aligned}
$$

where Σ_s denotes the covariance matrix of the signals received. Inserting (2.97) and (2.98) into (2.96) we get

$$\begin{aligned} 0 &= s_i - \mu - (M-1)\Lambda\alpha \\ &\quad -\Lambda B \sum_{j=1, j\neq i}^{M} (\Sigma_s \Sigma_\varepsilon^{-1}(s_i - p_0)) - 2\Lambda x_i \\ &= s_i - \mu - (M-1)\Lambda\alpha \\ &\quad -(M-1)\Lambda B \Sigma_s \Sigma_\varepsilon^{-1}(s_i - p_0) - 2\Lambda x_i \\ &= (s_i - p_0) - (M-1)\Lambda\alpha \\ &\quad -(M-1)\Lambda B \Sigma_s \Sigma_\varepsilon^{-1}(s_i - p_0) + p_0 - \mu - 2\Lambda x_i. \end{aligned} \tag{2.99}$$

Solving for $2\Lambda x_i$ and inserting for x_i from (2.94) we get

$$\begin{aligned} \left(I - (M-1)\Lambda B \Sigma_s \Sigma_\varepsilon^{-1}\right)(s_i - p_0) &+ p_0 - \mu - (M-1)\Lambda\alpha \\ &= 2\Lambda\alpha + 2\Lambda B(s_i - p_0), \end{aligned} \tag{2.100}$$

where I denotes the identity matrix. By comparing coefficients we see that

$$p_0 - \mu - (M+1)\Lambda\alpha = 0, \tag{2.101}$$

$$I - (M-1)\Lambda B \Sigma_s \Sigma_\varepsilon^{-1} = 2\Lambda B. \tag{2.102}$$

From (2.91) we get

$$\begin{aligned} \widetilde{p} &= E[\widetilde{v}|x+u] \\ &= E[\widetilde{v}] + Cov[\widetilde{v}, \widetilde{x} + \widetilde{u}]Var[\widetilde{x} + \widetilde{u}]^{-1}(\widetilde{x} + \widetilde{u} - E[\widetilde{x} + \widetilde{u}]), \end{aligned} \tag{2.103}$$

where

$$E[\widetilde{v}] = p_0, \tag{2.104}$$

$$\begin{aligned} E[\widetilde{x} + \widetilde{u}] &= E[\widetilde{x}] + E[\widetilde{u}] \\ &= E\left[\sum_{i=1}^{M} \widetilde{x}_i\right] = E\left[\sum_{i=1}^{M} (\alpha + B(\widetilde{s}_i - p_0))\right] \\ &= M\alpha + B\sum_{i=1}^{M} E[\widetilde{s}_i - p_0] = M\alpha, \end{aligned} \tag{2.105}$$

$$(2.106) \quad \begin{aligned} Var[\widetilde{x}+\widetilde{u}] &= Var[\widetilde{x}] + Var[\widetilde{u}] \\ &= Var\left[\sum_{i=1}^{M} \widetilde{x}_i\right] + \Sigma_u \\ &= Var\left[\sum_{i=1}^{M} (\alpha + B(\widetilde{s}_i - p_0))\right] + \Sigma_u \\ &= MBVar[\widetilde{s}_i - p_0]B' \\ &\quad + M(M-1)BCov[\widetilde{s}_i - p_0, \widetilde{s}_j - p_0]B' + \Sigma_u \\ &= MB\Sigma_\varepsilon B' + M(M-1)B\Sigma_s B' + \Sigma_u, \end{aligned}$$

$$(2.107) \quad \begin{aligned} Cov[\widetilde{v}, \widetilde{x}+\widetilde{u}] &= Cov[\widetilde{v}, \widetilde{x}] + Cov[\widetilde{v}, \widetilde{u}] \\ &= Cov\left[\widetilde{s}_i - \widetilde{\varepsilon}_i, B\sum_{j=1}^{M} (\widetilde{s}_j - p_0)\right] \\ &= M\Sigma_\varepsilon B'. \end{aligned}$$

Inserting these relations into (2.103) we get

$$(2.108) \quad \widetilde{p} = p_0 + M\Sigma_\varepsilon B' \left[MB\Sigma_\varepsilon B' + M(M-1)B\Sigma_s B' + \Sigma_u\right]^{-1} (\widetilde{x} + \widetilde{u} - M\alpha).$$

Comparing coefficients with (2.93) gives

$$(2.109) \quad \Lambda = M\Sigma_\varepsilon B' \left[MB\Sigma_\varepsilon B' + M(M-1)B\Sigma_s B' + \Sigma_u\right]^{-1},$$

$$(2.110) \quad \mu = p_0 - M\Lambda\alpha.$$

we can now solve (2.101), (2.102), (2.109) and (2.110) for the unknown parameters μ, Λ, α and B. This calculation is conducted in CABALLE AND KRISHNAN (1994, pp. 701 f.). It turns out that

$$(2.111) \quad p = p_0 + \frac{\sqrt{M}}{2}\Gamma x,$$

$$(2.112) \quad x = \Theta(s_i - p_0),$$

where

$$\Gamma = \Sigma_u^{-\frac{1}{2}} H^{\frac{1}{2}} \Sigma_u^{-\frac{1}{2}}, \tag{2.113}$$

$$H = \Sigma_u^{\frac{1}{2}} G \Sigma_u^{\frac{1}{2}} \tag{2.114}$$

$$\begin{aligned} G = & \left[\Sigma_\varepsilon^{-1} + \frac{M-1}{2}\Sigma_\varepsilon^{-1}\Sigma_s\Sigma_\varepsilon^{-1}\right]^{-1} \\ & - \left[\frac{2}{M-1}\Sigma_s^{-1} + 2\Sigma_\varepsilon^{-1} + \frac{M-1}{2}\Sigma_\varepsilon^{-1}\Sigma_s\Sigma_\varepsilon^{-1}\right]^{-1}, \end{aligned} \tag{2.115}$$

$$\Theta = \frac{1}{\sqrt{M}\Gamma^{-1}\left[I + \frac{M-1}{2}\Sigma_s\Sigma_\varepsilon^{-1}\right]^{-1}}. \tag{2.116}$$

Although this result is very difficult to interpret due to its complexity, some general properties can be derived. It can be shown that Γ is positive definite and symmetric. This can be used to state that the order flow of the pth asset affects the price of the qth asset in the same degree as the order flow of the qth asset influences the price of the pth asset, i.e. the informativeness of trades in the other asset is equal for all assets.[14]

Further insights should be gained by analyzing the correlation structure in more detail, i.e. how different covariances in error terms or liquidity trade imbalances influence the covariance of the observed prices. The complexity of the result makes such an analysis very difficult to conduct in general. It should be noted that as

$$Var[\widetilde{v}|p] = \Sigma_0 - M\left[2\Sigma_\varepsilon^{-1} + (M-1)\Sigma_\varepsilon^{-1}\Sigma_s\Sigma_\varepsilon^{-1}\right]^{-1}, \tag{2.117}$$

the beliefs after observing prices can exhibit a completely different correlation structure than the initial beliefs.

2.3 The informational content of trading volume

Thus far prices have been used by uninformed investors to revise their beliefs of the fundamental value of the asset. Another variable is widely used in markets

[14] Caballe and Krishnan (1994, p.700) report that this result can be confirmed in empirical investigations.

as a source of information: trading volume. Using this additional information enables uninformed investors to learn the information more quickly. BLUME ET AL. (1994) provide a simple model of trading volume.

They assume the fundamental value to be normally distributed with mean p_0 and variance σ^2:

$$\widetilde{p} \sim N(p_0, \sigma^2). \tag{2.118}$$

A share of γ of the total N investors are assumed to be informed, but unlike in the previous models their information is not perfect. They only observe a noisy signal of the fundamental value:[15]

$$\widetilde{\psi}_1^i = \widetilde{p}^* + \widetilde{\varepsilon}_1^i, \tag{2.119}$$

where

$$\widetilde{p}^* \sim N(p, \sigma_p^2) \tag{2.120}$$

is a common signal. There remains uncertainty about the transformation of the information into the fundamental value. This information is not observed purely, i.e. there is some noise with

$$\widetilde{\varepsilon}_1^i \sim N(0, \widetilde{\sigma}_{\varepsilon_1}^2). \tag{2.121}$$

The error variance $\widetilde{\sigma}_{\varepsilon_1}^2$ is itself a random variable, i.e. the quality of information varies randomly over time. The informed investors know the realization of $\widetilde{\sigma}_{\varepsilon_1}^2$ when they receive their information, uninformed investors only know its distribution. Uninformed investors are assumed to receive also the signal $\widetilde{p}^*$ but with another error term $\widetilde{\varepsilon}_2$:

$$\begin{aligned} \widetilde{\psi}_2^i &= \widetilde{p}^* + \widetilde{\varepsilon}_2^i, \\ \widetilde{\varepsilon}_1^i &\sim N(0, \widetilde{\sigma}_{\varepsilon_2}^2), \end{aligned} \tag{2.122}$$

[15] This modification is necessary as otherwise uninformed investors would be able to deduct the information completely by only observing price and volume. In this case informed investors would not be able to make profits from their information and if information is costly there would be no incentives to become informed. If no one is informed it would be profitable to be informed and no equilibrium exists. See GROSSMAN AND STIGLITZ (1980) for more details on this problem.

where $\widetilde{\sigma}^2_{\varepsilon_2}$ is known to all investors, informed and uninformed. With the assumption that $\widetilde{\sigma}^2_{\varepsilon_2} > \widetilde{\sigma}^2_{\varepsilon_1}$ the information of informed investors is more precise, that is why they are called informed. We therewith have for the distribution of the true value $\widetilde{\psi}_i$ of group i, if we assume the errors terms to be independent of all other relevant variables:

$$\widetilde{\psi}_i \sim N(p, \sigma_p^2 + \widetilde{\sigma}^2_{\varepsilon_i}). \tag{2.123}$$

The demand of the investors for the asset has been derived by DIAMOND AND VERRECCHIA (1981) and BROWN AND JENNINGS (1989). When we restrict the equilibrium to be linear we receive with risk averse investors j the demand for every individual in groups 1 (informed investors) and 2 (uninformed investors) as

$$\begin{aligned} d_1^j &= \frac{p_0 - p_1}{z\sigma^2} + \frac{\psi_1^j - p_1}{z(\sigma_p^2 + \sigma_{\varepsilon_1}^2)}, \\ d_2^j &= \frac{p_0 - p_1}{z\sigma^2} + \frac{\psi_2^j - p_1}{z(\sigma_p^2 + \sigma_{\varepsilon_2}^2)}, \end{aligned} \tag{2.124}$$

where z denotes the Arrow-Pratt measure of risk aversion and p_1 the price applied. The total demand of all investors has to equal zero as long as the amount of the asset is fixed to achieve an equilibrium. Aggregation over all individuals gives us after dividing by N and multiplying by z:

$$\begin{aligned} 0 &= \frac{1}{N}\left(\sum_{i=1}^{\gamma N}\left(\frac{p_0 - p_1}{\sigma^2} + \frac{\psi_1^j - p_1}{\sigma_p^2 + \sigma_{\varepsilon_1}^2}\right)\right. \\ &\quad \left. + \sum_{i=\gamma N+1}^{N}\left(\frac{p_0 - p_1}{\sigma^2} + \frac{\psi_2^j - p_1}{\sigma_p^2 + \sigma_{\varepsilon_2}^2}\right)\right) \\ &= \frac{p_0 - p_1}{\sigma^2} + \frac{1}{N}\left(\frac{1}{\sigma_p^2 + \sigma_{\varepsilon_1}^2}\sum_{1=1}^{\gamma N}(\psi_1^i - p_1)\right. \\ &\quad \left. + \frac{1}{\sigma_p^2 + \sigma_{\varepsilon_2}^2}\sum_{i=\gamma N+1}^{N}(\psi_2^i - p_1)\right) \\ &= \frac{p_0 - p_1}{\sigma^2} + \frac{\gamma(\overline{\psi}_1 - p_1)}{\sigma_p^2 + \sigma_{\varepsilon_1}^2} + \frac{(1-\gamma)(\overline{\psi}_2 - p_1)}{\sigma_p^2 + \sigma_{\varepsilon_2}^2}, \end{aligned} \tag{2.125}$$

where $\overline{\psi}_1 = \frac{1}{N}\sum_{i=1}^{\gamma N}\psi_1^i$ and $\overline{\psi}_2 = \frac{1}{N}\sum_{i=\gamma N+1}^{N}\psi_2^i$ denote the average realization of the information. Solving for p_1 gives

$$p_1 = \frac{\frac{p_0}{\sigma^2} + \frac{\gamma}{\sigma_p^2+\sigma_{\varepsilon_1}^2}\overline{\psi}_1 + \frac{1-\gamma}{\sigma_p^2+\sigma_{\varepsilon_2}^2}\overline{\psi}_2}{\frac{1}{\sigma^2} + \frac{\gamma}{\sigma_p^2+\sigma_{\varepsilon_1}^2} + \frac{1-\gamma}{\sigma_p^2+\sigma_{\varepsilon_2}^2}}. \tag{2.126}$$

If we let $N \to \infty$ then the law of large numbers can be used to show that $\overline{\psi}_j \to p$ and (2.126) becomes

$$p_1 = \frac{\frac{p_0}{\sigma^2} + \left(\frac{\gamma}{\sigma_p^2+\sigma_{\varepsilon_1}^2} + \frac{1-\gamma}{\sigma_p^2+\sigma_{\varepsilon_2}^2}\right)p}{\frac{1}{\sigma^2} + \frac{\gamma}{\sigma_p^2+\sigma_{\varepsilon_1}^2} + \frac{1-\gamma}{\sigma_p^2+\sigma_{\varepsilon_2}^2}}. \tag{2.127}$$

The total trading volume is the absolute value of the demands given in (2.124). To avoid double counting this has to be divided by two. Hence the trading volume, adjusted to per capita average volume is with $N \to \infty$:

$$\begin{aligned} V &= \frac{1}{2}\frac{1}{N}\left(\sum_{i=1}^{\gamma N}|d_1^i| + \sum_{i=\gamma N+1}^{N}|d_2^i|\right) \\ &= \frac{\gamma}{2}\frac{1}{\gamma N}\sum_{i=1}^{\gamma N}|d_1^i| + \frac{1-\gamma}{2}\frac{1}{(1-\gamma)N}\sum_{i=\gamma N+1}^{N}|d_2^i| \\ &\to \frac{\gamma}{2}E[|d_1|] + \frac{1-\gamma}{2}E[|d_2|]. \end{aligned} \tag{2.128}$$

Inserting from (2.124) and (2.127) gives the expression for the volume, which is explicitly stated in Blume et al. (1994, p. 165), but not reproduced here. The evaluation is best conducted with numerical examples.[16]

If the information received by informed investors is only of low precision and suggests only a small deviation from prior beliefs, i.e. $\sigma_{\varepsilon_1}^2$ is relative large compared to $\sigma_{\varepsilon_2}^2$, informed investors have not much confidence in their information. Although their beliefs are widely spread they do not trade much on their information, hence trading volume will be relatively small. If the precision increases, i.e. $\sigma_{\varepsilon_1}^2$ decreases, they become more confident in their information and if their beliefs are wide enough dispersed, trading volume increases. For $\sigma_{\varepsilon_1}^2 = \sigma_p^2$ the trading volume reaches its maximum. If the precision of information is further increased, their confidence also increases, but on the other hand the dispersion of

[16] The results of the illustrations below can be shown to be valid in all economically relevant situations.

beliefs is reduced, they do not find many informed investors to trade with, they are forced to trade nearly only with uninformed investors, hence trading volume reduces.

If the information received suggests a large deviation from the prior belief of p_0, the rebalancing of the portfolios will dominate, even if the precision is not large. The trade volume will increase with the precision of the information, even if the dispersion of beliefs is reduced.

The larger the signal suggests the deviation from prior beliefs is the more need the investors to rebalance their portfolios, this need increases with the precision of information. As p_1 denotes also the new belief of the informed investors, $p_1 - p_0$ denotes the change in belief. Therewith the volume should be V-shaped in the change of beliefs, where the V becomes more pronounced the more precise the information is. The lowest trading volume should be found at $p_0 = p_1$.

Figure 2.9 visualizes these findings using the expression for trading volume provided by Blume et al. (1994). Virtually all empirical investigations on price changes and volume find V-shaped pattern between the price change and volume. With this simple model this pattern can be explained.

Figure 2.10 shows furthermore that the V-shape does not only depend on the precision of information, but also on the dissipation of information, i.e. the share of informed investors γ. The more investors are informed the less pronounced the V-shape is until it vanishes if all investors are informed. This pattern can be explained with the fact that with a price change due to new information only few investors can expect to make a profits as their beliefs deviate from the current price. Hence the trading volume does not respond so sensitive to price changes. With only minor price changes, or equivalently changes in beliefs, the volume is increasing for $\sigma^2_{\varepsilon_1} > \sigma^2_p$ and decreasing for $\sigma^2_{\varepsilon_1} < \sigma^2_p$ due to the effect of rebalancing portfolios as described above.

While prices reveal information about the magnitude of a signal, i.e. the change in beliefs of the informed investors, trading volume reveals information about the precision and dissemination of information. By observing prices and

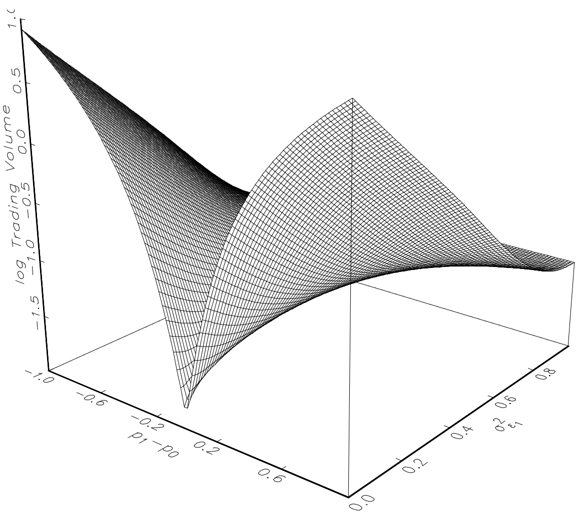


Fig. 2.9: Trading volume with different precision of information and changes of beliefs

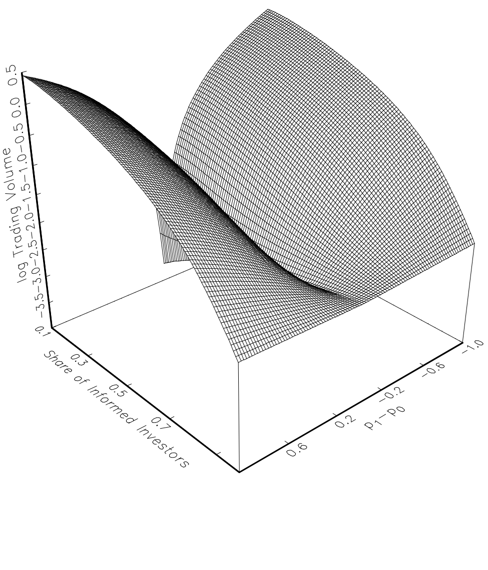


Fig. 2.10: Trading volume with different shares of informed investors and changes of beliefs

trading volume an investor can find out the beliefs of the informed investors and their changes by observing prices and the persistence of the movement, i.e. the confidence of informed investors by observing trading volume. This can be observed in markets where large price changes in association with low trading volume in most cases are viewed as being not very persistent, whereas the same price change with a large trading volume is viewed as persistent by market observers.

2.4 Explaining short-term movements of asset prices

In recent years many empirical studies have been conducted to detect patterns in asset markets. Many such patterns have been found between trading days as well as within a trading day. Various explanations and models have been offered to explain the found behavior. In this subsection we will not cover this literature in detail, but concentrate on two widely cited contributions, whose results can be explained intuitively with the models presented above, such that there is no need to develop the models that are actually chosen by the authors in detail.

Foster and Viswanathan (1990) use the above derived framework to explain the variation of asset price volatility and trading volume within a trading week. Trading takes place typically from Monday to Friday, while they assume that information is acquired at every day of the week, i.e. from Monday to Sunday, but portfolio imbalances are not aggregated over the weekend. Therewith informed investors should have a larger informational advantage from trading at Mondays, on the other hand adverse selection costs are highest on Mondays. Hence uninformed investors will not trade so actively on this day. If we assume further that a part of the information becomes public by other sources than prices, e.g. publications in newspapers, rumors, over the following trading days, informed investors are forced trade on their information also on Mondays in order to make profits, hence the fraction of informed trades will be relative high on these days and trading volume should be low due to the reduced participation of uninformed investors. A high fraction of informed trades implies a large change in the price, hence we should expect to find a larger volatility on Mondays. Empirical investigations support this effect as suggested by this model.

Due to adverse selection costs uninformed investors always make losses when trading with informed investors. The more uninformed investors are in the market, the more these costs are distributed among uninformed investors, reducing costs of trading for a single uninformed investor. It would therefore be a beneficial strategy for uninformed investors to concentrate their trading in a certain trading round. Also informed investors would be better able to hide their trades and would be able to trade more and hence make more profits in these trading rounds. It is reasonable to assume that for most investors it is of no importance at which time of the day they trade as the settlement of the orders depends only on the day of trading and not on the time of trading within a day. Admati and Pfleiderer (1988) show that with such assumptions trading will be concentrated at certain points of time in the trading day, provided competition between informed investors is large enough and adverse selection costs do not increase too much by the presence of more informed investors. Again the presence of many informed and uninformed investors increases the volatility of prices in those times of higher trading activity, a result that can be confirmed in empirical studies.

The two models shortly presented above can explain some of the effects observed in asset market regarding volatility and trading volume. However, they cannot explain observed patterns in returns. Admati and Pfleiderer (1989) provide a model how to explain such patterns, but they have to introduce another rule how the match maker determines prices. They assume the match maker to set prices strategically, what leaves the framework of the models presented above.

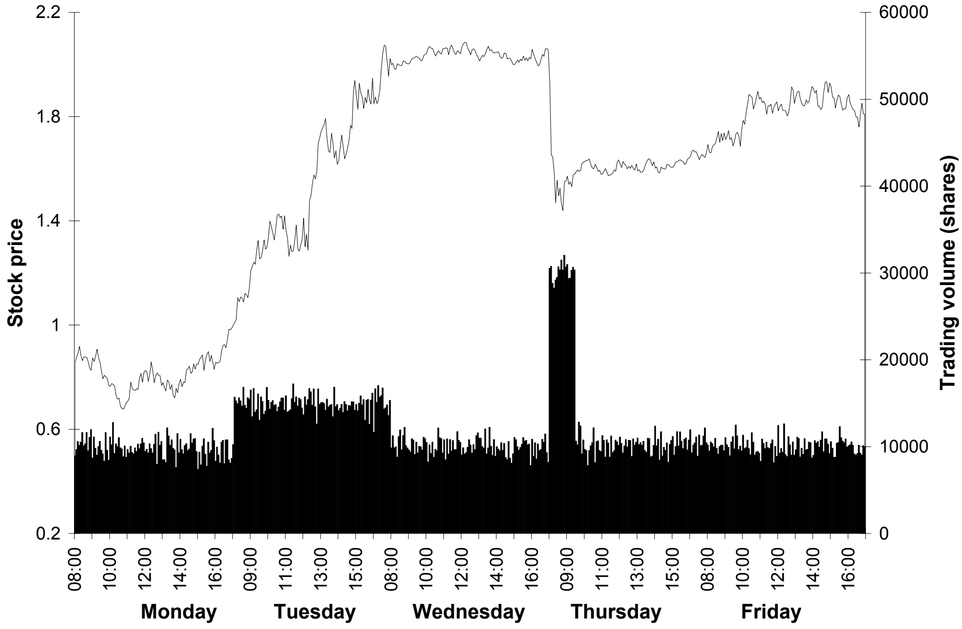
Review questions

1. In the single period model of Kyle (1985), why is not all information included into the price after trading?
2. Why is liquidity increasing over time in the multi-period version of the model of Kyle (1985)?
3. How does the informational efficiency of the market evolve over time in Kyle (1985)?
4. Why do a larger number of traders increase the informational efficiency of prices in Kyle (1985)?
5. A larger number of auctions increases the efficiency of the market. Why?
6. How can you explain the increasing liquidity of the market for more risk averse uninformed investors?
7. Why is the liquidity not monotone decreasing in the number of risk averse uninformed traders?
8. Explain the V-shaped trading volume in the change of belief.
9. What role does the quality of information play for trading volume?
10. Why do investors cluster their trading?

Application

In the morning of Wednesday, 21 November 2004 a rumor spread in the market about an imminent approach for a takeover of Nanobot Inc. by its larger rival TechAppliance Inc. at a significant premium. TechAppliance Inc. made it publicly known on Thursday, 22 November 2004 at 12.00 noon that it was not planning a takeover bid in Nanobot Inc. As usual in such cases the market regulator took a closer look at the trading activity during these days, in particular as the stock had seen significant price movements.

For a preliminary analysis the regulator has produced a chart of stock prices (represented by the line) and trading volumes (represented by the bars) in five minute intervals. This chart is reproduced below. Given the theories presented in this chapter, what can be inferred from this on the trading activity, particularly during Tuesday and Thursday morning?



Stock price
Trading volume (shares)
2.2
1.8
1.4
1
0.6
0.2
60000
50000
40000
30000
20000
10000
0
08:00
10:00
12:00
14:00
16:00
09:00
11:00
13:00
15:00
08:00
10:00
12:00
14:00
16:00
09:00
11:00
13:00
15:00
08:00
10:00
12:00
14:00
16:00
Monday
Tuesday
Wednesday
Thursday
Friday

Chapter 3

Dealer Markets

This chapter will introduce the main theories on dealer markets. The focus in this chapter will be on the determination of the spread in such markets, and the main contents will cover

- costs of market making without asymmetric information
- competitive and monopolistic spreads
- adverse selection costs in markets with asymmetric information
- the interaction of multiple assets and their effects on the spread

Key readings:

Hans R. Stoll: The Supply of Dealer Services in Securities Markets, *Journal of Finance*, 33, 1133-1151, 1978

Thomas Ho and Hans R. Stoll: Optimal Dealer Pricing under Transactions and Return Uncertainty, *Journal of Financial Economics*, 9, 47-73, 1981

Lawrence R. Glosten and Paul R. Milgrom: Bid, Ask and Transaction Prices in a Specialist Market with Heterogenously Informed Traders, *Journal of Financial Economics*, 14, 71-100, 1985

Thomas Gehrig and Matthew Jackson: Bid-Ask Spreads with Indirect Competition Among Specialists, *Journal of Financial Markets*, 1, 89-119, 1998

The last chapter analyzed how prices adjust to new information, therefore the order flow has been aggregated over a given period of time. The match maker had no other role than to determine the equilibrium price, he played no active role in the trading process. In the following sections we will now investigate the behavior of prices on a trade-by-trade basis. We therefore introduce a market maker into the trading process, replacing the match maker. The market maker directly influences the prices by quoting prices at which he is willing to buy and sell the asset. But in general he also has no active role in the trading process as he does not initiate or actively search for a trade, but waits for an order to arrive and then clears this order at the stated price on his own account. It is intuitively clear that transaction prices are not only influenced by the orders, but also by the behavior of the market maker.

Two main groups of theories modeling the behavior of market makers have evolved, inventory and information-based models. Information-based models are close to the models of auction markets, they assume a risk-neutral market maker and two groups of investors, informed and uninformed investors. Inventory-based models assume a risk averse market maker and all investors have the same information and agree on the implications of this information on the fundamental value, i.e. they agree on the fundamental value. While in information-based models trades can be motivated by exploiting informational advantages (informed investors) and need for liquidity (uninformed investors, or noise traders), in inventory-based models need for liquidity is the only source of trade. The coming section analyzes inventory-based models and the following section information-based models.

3.1 Inventory-Based Models of Market Making

A market maker provides the service of enabling an investor to trade immediately at a given price by acting as counterpart of the order. He then waits for another order offsetting his position. Therewith he takes the risk of not knowing when and at which price he can offset this position. A trade that is typically conducted between two investors is divided into two parts that occur at different prices at different times. Not knowing when and at which price an offsetting order arrives imposes costs on the market maker that he has to cover by quoting different prices at which he willing to buy (bid price) and to sell (ask price) the asset. As the investor has not to bear these risks, he is willing to pay for this service by accepting a less favorable price for the trade until the costs imposed by the market maker equals his costs from waiting for an offsetting order and trading at an unknown price (waiting costs).[1]

We now have to make some assumptions on the the determination of the order flow.[2] At first we assume that all investors place their orders independently of each other. Each investor either submits a buy or a sell order, but not both. Therewith the order submissions to buy and to sell are independent of each other. The need for trading is exogenously given by a liquidity event, which determines the waiting costs of an investor. This liquidity event is assumed to be a random variable (hence waiting costs are a random variable) that is independent and identical distributed between investors. If the waiting costs are higher than the costs imposed by the market maker,[3] he will submit his order to the market maker, otherwise he will wait for an offsetting order by himself. The higher the costs for trading with the market maker, the less orders are submitted. We can now aggregate all orders (separately for buy and sell orders) and see that the orders submitted in a given period of time follow a Poisson process or, equivalently, that the probability of an order arriving is Poisson distributed.[4]

A Poisson distribution is characterized by the order arrival rate λ. Let λ_a denote the order arrival rate for buy orders (for trades at the ask) and λ_b for sell orders (for a trade at the bid). Let further denote p^* the fundamental value all investors agree on, p_a the ask and p_b the bid price. The costs the market maker

[1] Demsetz (1968, p. 37) was the first to introduce the concept of waiting costs into literature. His pathbreaking article lead the way to the following literature on market microstructure theory.

[2] These assumptions have first explicitly been stated by Garman (1976, pp. 258 ff.).

[3] The costs are the difference between the value of the asset and the quoted price.

[4] To see this it has to be noted that submitting an order or not is a binomial variable, by aggregating binomial variables they converge to a Poisson process.

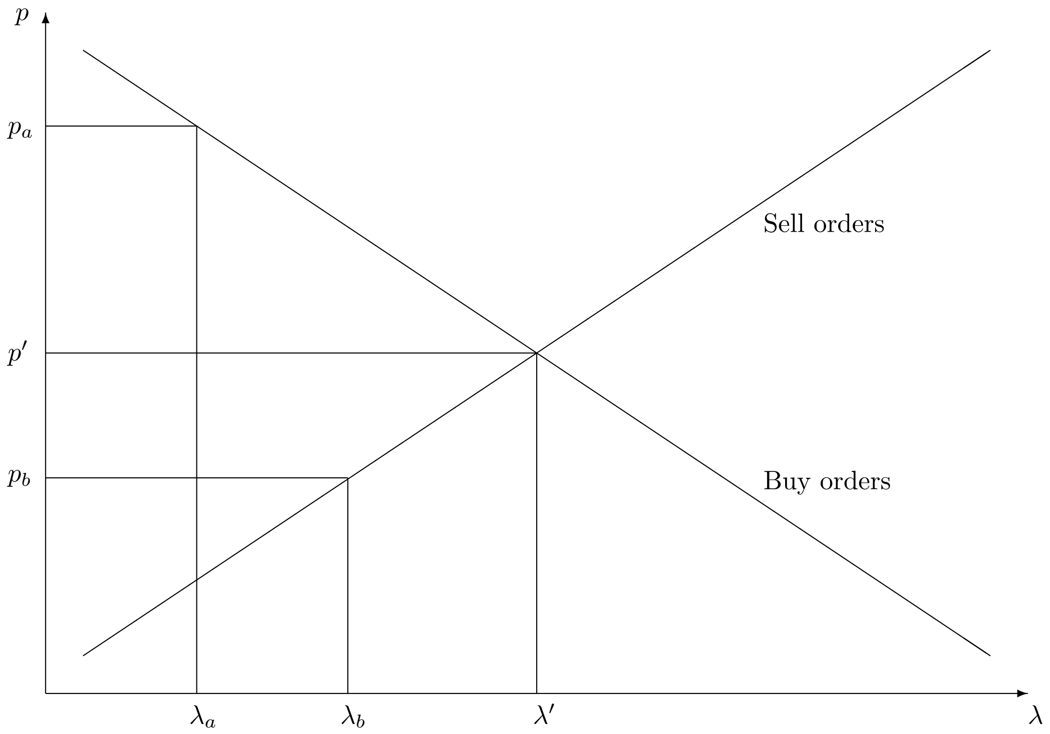


Fig. 3.1: Demand and Supply in Dealer Markets

imposes on the investors are $p_a - p^*$ and $p^* - p_b$.[5] Therewith we find that

$$\frac{\partial \lambda_b}{\partial p_b} > 0, \qquad \frac{\partial \lambda_a}{\partial p_a} < 0,$$
$$\frac{\partial \lambda_b}{\partial p_a} = 0, \qquad \frac{\partial \lambda_a}{\partial p_b} = 0. \tag{3.1}$$

Figure 3.1 visualizes these findings. At (p',λ') we have a stochastic equilibrium of individual order arrival rates.[6] However by choosing $p_a = p_b = p'$ the market maker would not be able to cover his costs. He has to choose two prices on the curves such that $p_a > p_b$, e.g. those marked in figure 3.1. As in this case $\lambda_b > \lambda_a$, it is more likely that a sell order arrives next in the market, i.e. the market maker expects to increase his position in the asset by offsetting the order. Similarly he will choose $\lambda_b < \lambda_a$ to decrease his position in the asset and $\lambda_a = \lambda_b$ if he does not want his position to change.

Before determining the costs of market making it is necessary to characterize

[5] It has already been pointed out that $p_b \leq p_a$.

[6] The equilibrium is stochastic as only expected demand and supply equal by having the same order arrival rates. The realized demand and supply may not equal.

the trading process and the behavior of the market maker in more detail. We assume that only a single trade per period of time is submitted to the market maker, either a buy or a sell order.[7] When quoting prices at which he is willing to buy and sell the asset he does not know whether the next order will be a buy or a sell order.

We consider an economy with only a single risky asset that is traded with the market maker and a riskless asset bearing no interest, e.g. money. Assume further that the market maker has an optimal portfolio consisting of the single risky asset and the riskless asset, chosen according to portfolio selection theory. Any deviations from this optimal portfolio are denoted as *inventory*.[8] After a single trading round we assume the risky asset to be liquidated at the fundamental value.

The market maker is assumed to be risk averse and maximize his expected utility of terminal wealth that occurs after a single round of trading by setting optimal bid and ask prices.

3.1.1 The costs of market making

The first to provide a model how to determine the costs of a market maker was Stoll (1978). Suppose that the market maker holds his optimal portfolio, denoted E in figure 3.2. By accepting a trade the portfolio actually held deviates from his optimal portfolio, suppose it is located at E', i.e. a buy order arrives at the market and the share of the asset in the portfolio is reduced, while the share of the riskless asset is increased. The utility level from holding portfolio E' instead of the optimal portfolio E decreases from U_0 to U_1. This difference in utility are the costs that the market maker faces from his service. He has to be compensated for this loss in utility, which is done by holding a larger portfolio,

[7] We could also allow for many orders to be submitted and concentrate on the order imbalance receiving the same results. In this sense we can interpret this situation as an auction market presented above, where the order imbalance has not been served and now is served by the market maker.

[8] As the optimal portfolio is fixed we can concentrate our analysis on the inventory that is a linear transformation of the entire portfolio held.

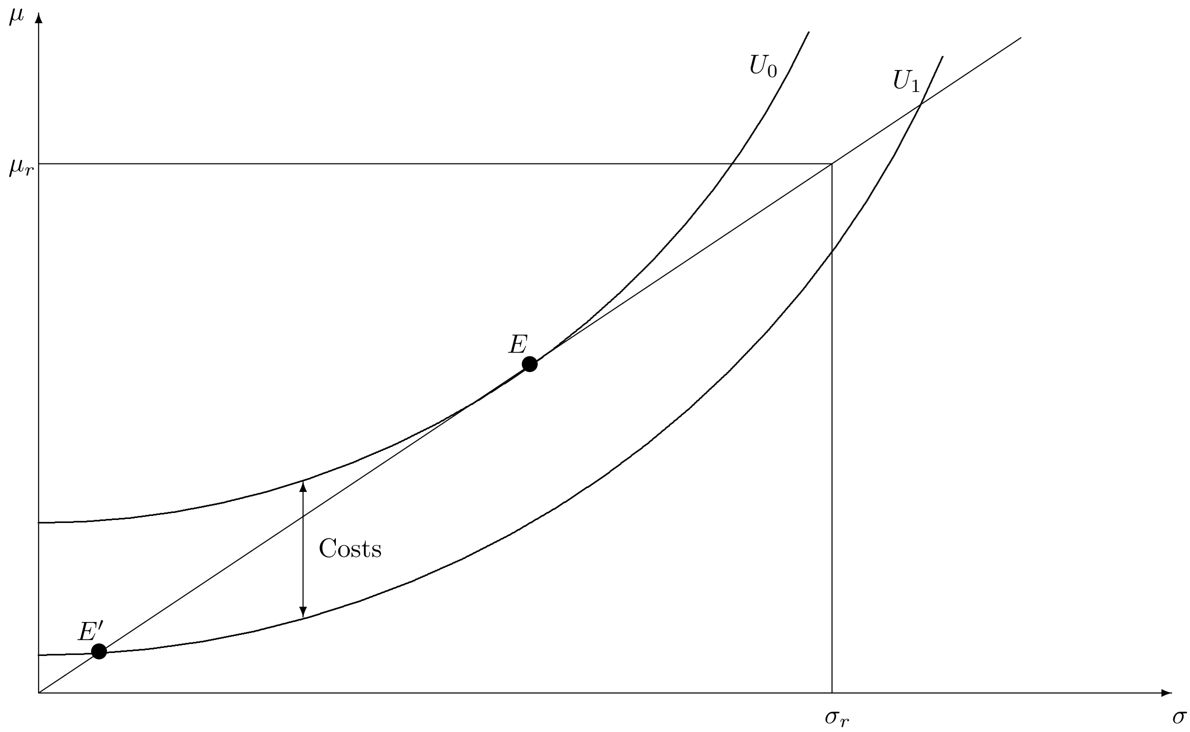


Fig. 3.2: Costs of a market maker

i.e. a higher total wealth. For holding a non-optimal portfolio he is compensated by an increase in the level of holdings.

It is also obvious from figure 3.2 that, if the market maker does currently not hold his optimal portfolio, i.e. his inventory is nonzero, he may gain in utility from accepting an order. Suppose as an example that he holds portfolio E', be accepting a sell order of the same size as before the buy order, he would gain utility by reaching portfolio E again, hence his costs would be negative. In this case p_b would lie above p^* in figure 3.1.

We only assume these costs to occur which are also called *inventory costs*. Other costs, e.g. for order processing, are not included here, but they can easily incorporated into this framework without changing the argument significantly.[9] Early contributions to market microstructure theory were concerned with the

[9] Stoll (1978, pp. 1144 ff.) also provides a more general framework by assuming more than a single risky asset to be in the market. But as he also assumes the market maker to act as market maker only for a single asset the results obtained are identical to those in this restricted version, they only come with more notational complications.

ability of the market maker to deliver the asset or the money. By allowing no short sales of the market maker he faces the risk of running out of stock, of assets as well as of money.[10] Although such a situation cannot be ruled out by any model, it has been found to be not relevant in practice. By allowing short sales a bankruptcy of the market maker can be avoided and these concerns have not to be considered. As also the results of the models focusing on the threat of bankruptcy give similar results, models of inventory costs have attracted more attention in the literature.

For bearing the above described inventory costs the market maker has to be compensated, i.e. his expected utility of terminal wealth from holding the initial portfolio and from holding the new portfolio after having accepted a trade have to be equal. Let $\widetilde{W}^*$ denote the terminal wealth without a trade and $\widetilde{W}$ the terminal wealth after having accepted a trade, then the costs are determined such that

$$E\left[U(\widetilde{W}^*)\right] = E\left[U(\widetilde{W})\right]. \tag{3.2}$$

The initial portfolio has not necessarily to be the optimal portfolio, it is the optimal portfolio plus the inventory of the market maker, which can be either positive or negative. Let k denote the fraction of total wealth that is invested into the risky asset in the optimal portfolio, the optimal holding of the risky asset then is kW_0, where W_0 denotes the initial total wealth. The total amount of the risky asset actually held is kW_0+I, with I denoting the inventory. With $\widetilde{R}$ denoting the return of the fundamental value of the risky asset, the final wealth of the portfolio is

$$\begin{aligned} \widetilde{W}^* &= W_0 + (kW_0 + I)\widetilde{R} \\ &= W_0\left(1 + \left(k + \frac{I}{W_0}\right)\widetilde{R}\right). \end{aligned} \tag{3.3}$$

With $\mu = E[\widetilde{R}]$ and $\sigma^2 = Var[\widetilde{R}]$ we get

$$E[\widetilde{W}^*] = W_0\left(1 + \left(k + \frac{I}{W_0}\right)\mu\right), \tag{3.4}$$

$$Var[\widetilde{W}^*] = W_0^2\left(k + \frac{I}{W_0}\right)^2\sigma^2. \tag{3.5}$$

[10] See e.g. Garman (1976) or Amihud and Mendelson (1980).

Approximating $U(\widetilde{W}^*)$ by a second order Taylor series around $E[\widetilde{W}^*]$ we get

$$\begin{aligned}(3.6)\qquad E\left[U(\widetilde{W}^*)\right] &= E\left[U(E[\widetilde{W}^*]) + U'(E[\widetilde{W}^*])(\widetilde{W}^* - E[\widetilde{W}^*])\right. \\ &\quad \left. + \frac{1}{2}U''(E[\widetilde{W}^*])(\widetilde{W}^* - E[\widetilde{W}^*])^2\right] \\ &= U\left(E[\widetilde{W}^*] + \frac{1}{2}U''(E[\widetilde{W}^*])Var[\widetilde{W}^*]\right).\end{aligned}$$

Let Q denote the trade size, measured in value not in numbers of assets traded, where $Q > 0$ for a sell order and $Q < 0$ for a buy order. We further denote C as the costs of the market maker to conduct a trade of size Q, transformed from utility into value. We then have for the terminal wealth with accepting a trade:

$$\begin{aligned}(3.7)\qquad \widetilde{W} &= W_0 + (kW_0 + I)\widetilde{R} + Q(1+\widetilde{R}) - (Q - C) \\ &= W_0\left(1 + \left(k + \frac{I}{W_0}\right)\widetilde{R}\right) + Q(1+\widetilde{R}) - (Q - C),\end{aligned}$$

where the first term denotes the part of terminal wealth that arises from holding the initial portfolio, the second term the part that has been affected by the change in inventory due to accepting an order of size Q and the last term the benefits in money from conducting this trade.

From (3.7) we get with assuming the order size Q to be known[11]

$$\begin{aligned}(3.8)\qquad E[\widetilde{W}] &= W_0\left(1 + \left(k + \frac{I}{W_0}\right)\mu\right) + Q(1+\mu) - (Q - C) \\ &= W_0\left(1 + \left(k + \frac{I}{W_0}\right)\mu\right) + Q\mu + C, \\ (3.9)\qquad Var[\widetilde{W}] &= W_0^2\left(k + \frac{I}{W_0}\right)^2\sigma^2 + Q^2\sigma^2 + 2W_0\left(k + \frac{I}{W_0}\right)Q\sigma^2 \\ &= \sigma^2\left(W_0\left(k + \frac{I}{W_0}\right) + Q\right)^2.\end{aligned}$$

[11] If we do not know the order size, but only its expected value and variance the results do not change, instead of the order size only its expected values have to be used if we assume the order size to be independent of other factors, e.g. costs.

Approximating $U(\widetilde{W})$ by a second order Taylor series around $E[\widetilde{W}]$ gives

$$
\begin{aligned}
(3.10) \qquad E\left[U(\widetilde{W})\right] &= E\left[U(E[\widetilde{W}]) + U'(E[\widetilde{W}])(\widetilde{W} - E[\widetilde{W}]) \right. \\
&\quad \left. + \frac{1}{2}U''(E[\widetilde{W}])(\widetilde{W} - E[\widetilde{W}])^2 \right] \\
&= U\left(E[\widetilde{W}] + \frac{1}{2}U''(E[\widetilde{W}])Var[\widetilde{W}] \right).
\end{aligned}
$$

If Q is relatively small compared to the initial wealth of the market maker we can assume that

$$
\begin{aligned}
(3.11) \qquad U'(E[\widetilde{W}^*]) &= U'(E[\widetilde{W}]), \\
(3.12) \qquad U''(E[\widetilde{W}^*]) &= U''(E[\widetilde{W}]).
\end{aligned}
$$

The mean-value theorem states that there exists a $\widetilde{W}'$ between $\widetilde{W}$ and $\widetilde{W}^*$ such that

$$\frac{U(E[\widetilde{W}]) - U(E[\widetilde{W}^*])}{E[\widetilde{W}] - E[\widetilde{W}^*]} = U'(E[\widetilde{W}']).$$

With (3.11) we can rewrite this as

$$(3.13) \qquad \frac{U(E[\widetilde{W}]) - U(E[\widetilde{W}^*])}{U'(E[\widetilde{W}^*])} = E[\widetilde{W}] - E[\widetilde{W}^*].$$

By inserting (3.6) and (3.10) into (3.2) we get with (3.12) after rearranging

$$U(E[\widetilde{W}]) - U(E[\widetilde{W}^*]) = \frac{1}{2}U''(E[\widetilde{W}^*])(Var[\widetilde{W}^*] - Var[\widetilde{W}]).$$

With z as the Arrow-Pratt measure of absolute risk aversion we get after dividing by $U'(E[\widetilde{W}^*])$ and inserting (3.4), (3.5), (3.8), (3.10) and (3.13):

$$(3.14) \qquad Q\mu + C = \frac{1}{2}z\sigma^2 \left(Q^2 + 2QW_0 \left(k + \frac{I}{W_0} \right) \right).$$

For the optimal portfolio, i.e. $I = 0$, we see from (3.3) that $\widetilde{W}^* = W_0(1 + k\widetilde{R})$, hence

$$
\begin{aligned}
(3.15) \qquad E[U(\widetilde{W}^*)] &= U\left(E[\widetilde{W}^*] - \frac{1}{2}zVar[\widetilde{W}^*] \right) \\
&= U\left(W_0(1 + k\mu) - \frac{1}{2}zW_0^2k^2\sigma^2 \right)
\end{aligned}
$$

such that

$$\frac{\partial U}{\partial \mu} = W_0 k U'(.), \tag{3.16}$$

$$\frac{\partial U}{\partial \sigma^2} = -z W_0^2 k^2 U'(.). \tag{3.17}$$

By totally differentiating (3.15) we get

$$\begin{aligned} dE[U(\widetilde{W}^*)] &= \frac{\partial U}{\partial \mu} d\mu + \frac{\partial U}{\partial \sigma^2} d\sigma^2 \\ &= (W_0 k d\mu - z W_0^2 k^2 d\sigma^2) U'(.). \end{aligned} \tag{3.18}$$

Setting (3.18) equal to zero we get the slope of the indifference curve at the optimal portfolio:

$$\frac{d\mu}{d\sigma^2} = z W_0 k. \tag{3.19}$$

The slope of the security market line is known from portfolio selection theory to be

$$\frac{d\mu}{d\sigma^2} = \frac{\mu}{\sigma^2}. \tag{3.20}$$

As we know from portfolio selection theory these two slopes have to be identical, hence we get from these two relations after rearranging:

$$k = \frac{\mu}{z W_0 \sigma^2}. \tag{3.21}$$

Inserting (3.21) into (3.14) we get

$$\begin{aligned} C &= -Q\mu + \frac{1}{2} z\sigma^2 \left(Q^2 + 2QW_0 \left(\frac{\mu}{zW_0\sigma^2} + \frac{I}{W_0} \right) \right) \\ &= -Q\mu + \frac{1}{2} z\sigma^2 Q^2 + Q\mu + z\sigma^2 QI \\ &= z\sigma^2 \left(\frac{1}{2} Q^2 + QI \right). \end{aligned} \tag{3.22}$$

This is the expression for the costs a market maker faces when accepting orders of size Q. The costs depend on a characteristic of the asset, the variance of the fundamental value of the asset σ^2, a characteristic of the trade, the trade size Q, and two characteristics of the market maker, his risk aversion z and his inventory position I.

It is more natural to assume the trade size to be always positive. Define $Q' = |Q|$ as the trade size and we get the costs for a trade at the ask, C_a, and at the bid, C_b, by

$$C_b = z\sigma^2 \left(\frac{1}{2}Q^2 + Q'I\right), \tag{3.23}$$

$$C_a = z\sigma^2 \left(\frac{1}{2}Q^2 - Q'I\right). \tag{3.24}$$

The relative costs are given by

$$c_b = \frac{C_b}{Q'} = z\sigma^2 \left(\frac{1}{2}Q' + I\right), \tag{3.25}$$

$$c_a = \frac{C_a}{Q'} = z\sigma^2 \left(\frac{1}{2}Q' - I\right). \tag{3.26}$$

When buying the asset the market maker reduces the price at which he is willing to buy compared to the fundamental value and increases it when selling the asset. Therefore the *reservation prices* of a market maker for the entire trade of size Q' are given by

$$p_b = p^* - C_b, \tag{3.27}$$

$$p_a = p^* + C_a. \tag{3.28}$$

The difference between the bid and the ask price is denoted the *spread*, s. From (3.23) - (3.28) we get

$$s = p_a - p_b = C_a + C_b = z\sigma^2 Q^2. \tag{3.29}$$

The spread does not depend on the inventory of the market maker, but only on his risk aversion, the trade size, and the variance of the fundamental value of the asset.[12] Unless the market maker is risk neutral, i.e. $z = 0$ or the asset is riskless, i.e. $\sigma^2 = 0$, the spread will always be positive. We therewith have verified the result already earlier stated that

$$p_a > p_b. \tag{3.30}$$

[12] The reservation prices for trading a single unit of the asset, as usually published, are given

The inventory does not influence the spread, but only the level of prices. It can easily be verified that if $I \neq 0$ the spread is not located symmetrically around p^*. If the inventory is sufficiently large the costs can even become negative, implying that $p_b > p^*$ or $p_a < p^*$.

These results derived by Stoll (1978) made use of the assumption that the market maker does know the trade size before quoting the price or that he is allowed to quote different prices for every trade size. In reality, however, the market maker does not know the trade size, he has to accept any trade at a single stated price up to a certain limit.[13] This adds another uncertainty to the market maker, the order size. It can be shown that the results above do not change if the order size is independent of the costs and expected return of the asset. The order size becomes a random variable and therefore instead of Q' and Q^2 we have to insert $E[Q']$ and $E[Q^2]$ into the above derived formulas. No further insight can be gained from this generalization.

The costs and reservation prices derived here are those that occur if the market maker has a time horizon of a single trade, i.e. he makes his consideration on a trade-by-trade basis. Such a short-term behavior can be justified if there is a fierce competition between market makers.

If the market maker has a time horizon longer than a single trade the costs are reduced. By accepting a trade with which the inventory position becomes less favorable there exists the chance that in one of the next trades an offsetting order arrives, what reduces his costs. Nevertheless the influences of the above parameters on the costs do not change significantly.

---

by

$$\begin{aligned} p_b &= p^*(1 - c_b), \\ p_a &= p^*(1 + c_a). \end{aligned}$$

and the spread is

$$s = p_a - p_b = p^*(c_a + c_b) = p^* z \sigma^2 Q'.$$

In this representation we can see that the higher the fundamental value of the asset is, the higher we expect the spread to be.

[13] At the NASDAQ Rule 4613 requires a market maker has to accept orders of at least an equivalent of USD 50,000-100,000, depending on several characteristics of the asset and the market.

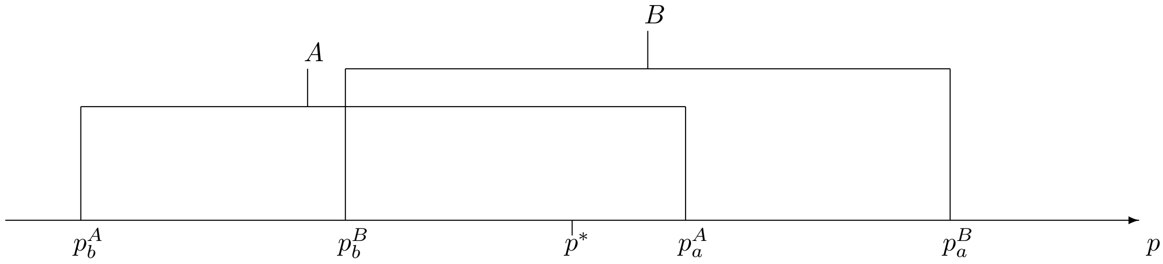


Fig. 3.3: Competitive price setting

After having derived the costs and reservation prices of a market maker we will in the following subsections discuss the price setting of market makers, first under competition and then for a monopolistic market maker.

3.1.2 Competitive price setting

Let us for simplicity assume throughout this subsection that all market makers have the same risk aversion and the order size is fixed to Q', i.e. the spread quoted by all market makers is identical. If all market makers have the same inventory we can easily see from (3.23) - (3.28) that they all have the same costs and reservation prices. Hence, if they act competitively they all will quote their reservation prices. Quoting a lower bid or a higher ask price would make this market maker not to offer the best price and hence he is excluded from the order flow by the assumption of strict price priority and would not receive a trade and thereby make no profits, like in the case when he quotes his reservation price. Quoting a higher bid or a lower ask price would bring him a loss, so that he will not quote such prices.

After a single trade however, the inventory position of the market maker executing the trade will change and thereby the costs and reservation prices. Ho and Stoll (1980) provide a framework how competitive market makers set their quotes in such a situation.

Figure 3.3 illustrates the situation where two market makers, A and B, are in the market and have different costs. Suppose the price setting of the bid price, the market maker with the highest bid price receives an incoming order, if both quote the same price it is assigned to one of them randomly. Market maker B has the lower costs for a trade at the bid, but by quoting his reservation price he would make no profits. If he lowers his quote he would be able to make a profit. As the costs of market maker A do not allow him to quote a higher price than p_a^A he cannot prevent him from doing so, market maker B still has the best price in the market and will receive all orders that arrive at the market until he quotes a price just a fraction above p_a^A.[14] As quoting a price just a fraction above p_a^A gives him the highest profits, he will quote this price. By quoting the same price, p_a^A, he would have to share the order flow with the other market maker, reducing his expected profits.[15] For simplicity we neglect the fraction that has to be quoted above p_a^A and say that he will quote p_a^A, only keeping in mind this arbitrary small fraction.

As is obvious with the same risk aversion it is not possible for a market maker to quote the best price as well on the ask as the bid side, unless all market makers have the same costs and quote their reservation prices. An inventory position allowing him to have smaller costs on one side of the trade gives rise to larger costs on the other side as can easily be seen from (3.23) and (3.24). Generalizing the example with two market makers to M market makers, we derive easily in a similar way the result that the market maker with the lowest costs for this side of the trade receives the order. He quotes the reservation price (minus a fraction) of the market maker with the second lowest costs. If we define C^1 and C^2 as the costs of the market maker with the lowest and second lowest costs, the profits

[14] We assume here that prices can be set continuously for simplicity. In case of discrete prices he would have to quote the last discrete price above p_a^A but below p_a^B. We also rule out the possibility that the profit maximum lies between p_a^A and p_a^B as λ_b is falling with p_b. In this case we get a behavior comparable to a monopolistic market maker to be treated in 3.1.3.

[15] The possibility that both market makers collude to make higher profits by quoting prices that are below the reservation prices of both market makers is neglected here.

from a trade are given by

$$
\begin{aligned}
\pi_b^1 &= C_b^1 - C_b^2 \\
&= z\sigma^2\left(\frac{1}{2}Q^2 + Q'I^1\right) - z\sigma^2\left(\frac{1}{2}Q^2 + Q'I^2\right) \\
&= z\sigma^2 Q'\left(I^1 - I^2\right), \\
\pi_a^1 &= C_a^1 - C_a^2 \\
&= z\sigma^2\left(\frac{1}{2}Q^2 - Q'I^1\right) - z\sigma^2\left(\frac{1}{2}Q^2 - Q'I^2\right) \\
&= z\sigma^2 Q'\left(I^2 - I^1\right).
\end{aligned}
\tag{3.31}
$$

Therewith the spread observed in the market (*market spread*) is the difference between the reservation prices of the market makers with the second lowest costs on each side of the trade. Following Ho and Stoll (1983) we will investigate how the market spread is related to the reservation spread. Until now we assumed the market makers to be passive, i.e. to wait for an order arriving at the market, to offset any undesirable inventory position. Another possibility for the market makers would be to initiate a trade by themselves with another market maker (*interdealer trading*).

The expected utility to wait for an offsetting order to arrive at the market is for the market maker with the lowest costs (the other market makers will not receive a trade)

$$
E[U(W_1)] = E[U(W_0)] + U'(W_0)\lambda_a(C_a^2 - C_a^1), \tag{3.32}
$$

where W_1 denotes the wealth of the market maker in the next period, W_0 the initial wealth, λ_a the probability that an order arrives at the market in the next period and $C_a^1 - C_a^2$ the gain from the trade according to (3.31). This gain in wealth is transformed into utility by multiplying with the first derivative of the utility function by the concept of a first order Taylor series approximation.

For interdealer trading we assume that it takes place at the beginning of a period and that the market maker has enough time to update his quotes for his new inventory and then waits for orders to arrive at the market. We further assume that the market maker with the second lowest costs quotes his reservation price.

The expected utility of wealth after interdealer trading, W_0', is composed of the expected utility from his initial wealth, W_0, and an adjustment for the fee he has to pay the other market maker, $\varrho = C_b^2$, and the gain from offsetting his inventory, $-C_a^1$:

$$E[U(W_0')] = E[U(W_0)] + U'(W_0)(-C_a^1 - \varrho). \tag{3.33}$$

After this interdealer trade his inventory has changed from I^1 to $I^1 - Q$ and therewith costs have changed from C_a^1 to $C_a^{1'}$. The expected utility from a trade is now

$$E[U(W_1')] = E[U(W_0')] + U'(W_0)(C_a^2 - C_a^{1'})\lambda_a. \tag{3.34}$$

If we assume W_0 and W_0' to differ not too much such that $U'(W_0) = U'(W_0')$ we get from (3.33) and (3.34):

$$E[U(W_1')] = E[U(W_0)] + U'(W_0)\left((C_a^2 - C_a^{1'})\lambda_a - C_a^1 - \varrho\right). \tag{3.35}$$

For choosing interdealer trading rather than waiting for an offsetting order it is necessary that $E[U(W_1')] > E[U(W_1)]$. Inserting from (3.32) and (3.35) we get after rearranging:

$$\lambda_a(C_a^2 - C_a^{1'}) < \lambda_a(C_a^2 - C_a^1) - C_a^1 - \varrho.$$

Inserting for C_a^1 and $C_a^{1'}$ we get after eliminating C_a^2:

$$\begin{aligned} -\varrho &> \lambda_a\left(-z\sigma^2\left(\frac{1}{2}Q^2 - Q'I_1\right) + z\sigma^2\left(\frac{1}{2}Q^2 - Q'(I_1 - Q')\right)\right) \\ &\quad + z\sigma^2\left(\frac{1}{2}Q^2 - Q'I_1\right) \\ &= z\sigma^2\left(Q^2\lambda_a + \frac{1}{2}Q^2 - I_1Q'\right), \end{aligned}$$

$$\varrho < z\sigma^2\left(I_1Q' - Q^2\left(\lambda_a + \frac{1}{2}\right)\right). \tag{3.36}$$

If there are only two market makers the market maker who wants to trade with his only competitor will be charged the fee that equals the second best reservation price, i.e. his own reservation price at the bid. Inserting from (3.23) for ϱ we get

$$z\sigma^2 \left(I_1 Q' + \frac{1}{2} Q^2 \right) < z\sigma^2 \left(I_1 Q' - Q^2 \left(\lambda_a + \frac{1}{2} \right) \right)$$

which solves for

$$\lambda_a > 1.$$

As λ_a is the probability that an order arrives at the market, it cannot exceed 1, hence with two market makers no interdealer trading occurs.

With more than two market makers we get with $\varrho = C_b^2 = z\sigma^2 \left(I_2 Q' + \frac{1}{2} Q^2 \right)$ from (3.36):

$$z\sigma^2 \left(I_2 Q' + \frac{1}{2} Q^2 \right) < z\sigma^2 \left(I_1 Q' - Q^2 \left(\lambda_a + \frac{1}{2} \right) \right),$$

which solves for

$$I_1 - I_2 > Q'(1 + \lambda_a), \tag{3.37}$$

i.e. if the difference in the inventory between the best and the second best market maker is large enough, interdealer trading is induced. For the second best market maker we find $\lambda_a = 0$ as he will never serve a trade, therewith he induces a trade only if $I_1 - I_2 > Q'$. For the best market maker we find $0 \leq \lambda_a \leq 1$, hence for him the divergence has to be even larger before he initiates a trade as he can hope to offset his inventory by an incoming order. If $I_1 - I_2 > 2Q'$ he will always induce an interdealer trade.

If there is the situation that for all market makers $I_1 - I_2 < Q'(1 + \lambda_a)$, i.e. no market maker wants to induce an interdealer trade, there will be no interdealer trading in the future provided that the parameters Q' and λ_a do not change. To see this suppose that all market makers have the same inventory, an order arriving can lead to a deviation of Q' for this market maker from the others, no market maker wants to induce a trade. Further orders arriving at the market affect only the inventories of the market makers with the most deviating inventories as they quote the best prices. They can therewith only narrow the divergence and no interdealer trading takes place.

The market spread, s_M, is the difference between the best quoted prices on each side of the trade. The best quoted prices are the reservation prices of the second best market makers, hence we have

$$\begin{aligned} s_M &= C_a^2 + C_b^2 \\ &= z\sigma^2\left(\frac{1}{2}Q^2 - I_1 Q'\right) + z\sigma^2\left(I_2 Q' + \frac{1}{2}Q^2\right) \\ &= z\sigma^2 Q^2 + z\sigma^2 Q'(I_2 - I_1) \\ &= s + z\sigma^2 Q'(I_2 - I_1). \end{aligned} \tag{3.38}$$

With only two market makers it has been shown that no interdealer trading occurs, therewith $I_2 - I_1 \leq Q'$, hence we find that

$$s \leq s_M \leq 2s. \tag{3.39}$$

With three market makers, one of them is the best on the ask, one at the bid and the third is the second best on both sides, hence $I_2 = I_1$ and we get

$$s_M = s. \tag{3.40}$$

With more than three market makers, the second best market maker on the bid side will have a lower inventory than the second best at the ask side, i.e. $I_2 \leq I_1$. As the difference is allowed to be maximal Q' we find that with more than three market makers

$$0 \leq s_M \leq s. \tag{3.41}$$

Incoming orders have the tendency to balance the inventory of the market makers as always the market maker with the most deviating inventory serves the order. Therewith we have the tendency of $I_2 - I_1$ to converge to zero and the market spread to converge to the reservation spread. If the inventories are quite similar another order increases the divergence and the process of convergence can start again.

Allowing the trade size to vary, adds many more possibilities how the spread can behave over time, but the general finding of the mechanism is not changed.

As all market makers have to pose their quotes for a period simultaneously, it has implicitly been assumed that market makers know the inventory positions of their competitors and can therewith calculate their reservation prices and set their prices accordingly, especially the reservation price of the second best market maker has to be known. By observing past prices of the other market makers it is also possible to determine the inventory position by inverting equation (3.23) and (3.24) without initially knowing their inventory position.

If other factors also influence the quoted prices and reservation prices, the exact inventory position cannot so easily be determined. It may be impossible to infer the exact reservation prices. If the market makers can infer only the probability distribution of the reservation prices instead of the the exact reservation prices, Biais (1993) has shown that the bid and ask quotes, on average, are identical to those quoted with full knowledge of the other market makers' reservation prices. He further finds that in this case the spread and the quoted prices are more volatile than with full knowledge.

After having investigated the competitive price setting, we now turn to the price setting of a monopolistic market maker.

3.1.3 The price setting of a monopolistic market maker

In several markets, e.g. the NYSE, a single market maker is granted a monopoly in providing his services. The main focus of a monopolistic market maker is not to cover his costs, but to maximize his expected utility of terminal wealth by choosing optimal bid and ask prices. Ho and Stoll (1981) provide a model, how a monopolistic market maker sets his prices.

As the demand for the service of the market maker will play an important role, at first we will model this side in more detail. The order arrival rates λ_a and λ_b can be interpreted as the probability that an order arrives at the market within a given time period $[t, t+1[$. Approximating by a first order Taylor series around zero gives us

$$\lambda_a(p_a) = \lambda_a(0) + \frac{\partial \lambda_a(0)}{\partial p_a} p_a, \tag{3.42}$$

$$\lambda_b(p_b) = \lambda_b(0) + \frac{\partial \lambda_b(0)}{\partial p_b} p_b. \tag{3.43}$$

If we denote the quoted prices to be the fundamental value of the asset adjusted by the fees the market maker charges, x_a and x_b, respectively, we get in analogy to (3.27) and (3.28):

$$p_a = p^* + x_a, \tag{3.44}$$

$$p_b = p^* - x_b. \tag{3.45}$$

Inserting these relations into (3.42) and (3.43) we get

$$\lambda_a(p_a) = \lambda_a(0) + \frac{\partial \lambda_a(0)}{\partial p_a} p^* + \frac{\partial \lambda_a(0)}{\partial p_a} x_a, \tag{3.46}$$

$$\lambda_b(p_b) = \lambda_b(0) + \frac{\partial \lambda_b(0)}{\partial p_b} p^* + \frac{\partial \lambda_b(0)}{\partial p_b} x_b. \tag{3.47}$$

The first two terms can be interpreted as the demand of the market maker if he would charge no costs, i.e. it is a measure for the size of the liquidity event investors face. We will denote these terms by α_a and α_b, respectively. The last term can be interpreted as an adjustment in the demand due to the sensitivity of the demand to fees charged, the absolute values of this sensitivities will be denoted β_a and β_b, respectively:

$$\lambda_a(p_a) = \alpha_a - \beta_a x_a, \tag{3.48}$$

$$\lambda_b(p_b) = \alpha_b - \beta_b x_b. \tag{3.49}$$

We assume that there are T time periods in which trading can take place, at every point of time $t \in [0, T]$ the market maker chooses bid and ask fees that are optimal in the sense that they maximize his expected utility of terminal wealth. At time T the asset is liquidated at the fundamental value and the proceedings are consumed. As before the total wealth of the market maker consists of two components, his inventory and money:[16]

$$W_t = I_t + M_t \tag{3.50}$$

for all $t = 1, \dots, T$. The inventory changes with the rate of return and with trading. By assuming again a fixed trade size Q' we get

$$\Delta I_{t+1} = \widetilde{R}_{t+1} I_t \Delta t + p^* Q_b + p^* Q_a, \tag{3.51}$$

where Q_b equals Q' if a trade at the bid occurs and is zero otherwise. Q_a equals Q' if a trade at the ask occurs and is zero otherwise. Money holdings change with

$$\Delta M_{t+1} = (p^* - x_b) Q_b + (p^* + x_a) Q_a = p_b Q_b + p_a Q_a. \tag{3.52}$$

Let $\widetilde{W}$ denote the terminal wealth of the market maker, he then has to maximize $E[U(\widetilde{W})]$ by choosing optimal fees x_a and x_b. We can define a performance function J as

$$J(t, M_t, I_t) = \max_{x_a, x_b} E[U(\widetilde{W}) | t, M_t, I_t]. \tag{3.53}$$

After the last trade at time $t = T$ has taken place the portfolio is liquidated. Therefore the market maker faces no uncertainty and (3.53) has to fulfill the boundary restriction that

$$J(T, M_T, I_T) = U(W_T). \tag{3.54}$$

With optimal fees no further increase in J can be achieved by changing fees, i.e.

$$\max_{x_a, x_b} dJ(t, M_t, I_t) = 0. \tag{3.55}$$

From the principle of optimality in dynamic programming we can use the fundamental recurrence relation and rewrite (3.53) as

$$\begin{aligned} J(t, M_t, I_t) &= \max_{x_a, x_b} \{ L(t, M_t, I_t, x_a, x_b) \Delta t \\ &\quad + J(t + \Delta t, M_t + \Delta M_{t+1}, I_t + \Delta I_{t+1}) \}, \end{aligned} \tag{3.56}$$

[16] Ho and Stoll (1981) also consider the optimal portfolio as part of the wealth. But as he is compensated for the risks associated with holding the optimal portfolio by the market it is not necessary to take into account this part of his wealth, but only deviations.

where $L(t, M_t, I_t, x_a, x_b)$ denotes the expected gain in utility in the current period and J the performance in the remaining periods. The expected gain in utility in the present period consists of the expected gain from a trading at the bid and at the ask, what can be interpreted as the difference in the performance functions with and without a trade:

$$\begin{aligned} (3.57)\ L(t, M_t, I_t, x_a, x_b) &= \lambda_a \left(J(t, M_t + p_a Q', I_t - Q') - J(t, M_t, I_t)\right) \\ &\quad +\lambda_b \left(J(t, M_t - p_b Q', I_t + Q') - J(t, M_t, I_t)\right). \end{aligned}$$

We can further approximate $J(t + \Delta t, M_t + \Delta M_{t+1}, I_t + \Delta I_{t+1})$ by a first order Taylor series around (t, M_t, I_t):

$$\begin{aligned} (3.58) \qquad J(t + \Delta t, M_t \Delta M_{t+1}, I_t + \Delta I_{t+1}) &= J(t, M_t, I_t) + J_t \Delta t \\ &\quad +J_M \Delta M_{t+1} + J_I \Delta I_{t+1}, \end{aligned}$$

where the subscripts denote the partial derivatives with respect to this variable evaluated at the appropriate point. Inserting (3.57) and (3.58) into (3.56) we get after dividing by Δt and rearranging:

$$\begin{aligned} (3.59) \qquad -J_t &= \max_{x_a, x_b} \left\{ J_I \frac{\Delta I_{t+1}}{\Delta t} + J_M \frac{\Delta M_{t+1}}{\Delta t} \right. \\ &\quad +\lambda_a \left(J(t, M_t + p_a Q', I_t - Q') - J(t, M_t, I_t)\right) \\ &\quad \left. +\lambda_b \left(J(t, M_t - p_b Q', I_t + Q') - J(t, M_t, I_t)\right) \right\}. \end{aligned}$$

If we let $\Delta t \to 0$, we see that

$$(3.60) \qquad \frac{\Delta I_{t+1}}{\Delta t} \to \frac{\partial I_t}{\partial t},$$

$$(3.61) \qquad \frac{\Delta M_{t+1}}{\Delta t} \to \frac{\partial M_t}{\partial t} = 0.$$

As with Δt also λ_a and λ_b converge to zero, the probability of a trade arriving within a very short period of time also approaches zero. We now assume that the return of the risky asset in a given period of time follows an Itô process:

$$(3.62) \qquad \widetilde{R}_{t+1} = \mu dt + \sigma dz$$

with dz denoting a standard Wiener process. Therewith we have from (3.51) with $\Delta t \to 0$:

$$dI_t = \mu I dt + \sigma I dz. \tag{3.63}$$

From Itô's lemma we get

$$dJ(t, M_t, I_t) = J_I dI_t + J_t dt + \frac{1}{2} J_{II} (dI_t)^2, \tag{3.64}$$

where

$$(dI_t)^2 = \sigma^2 I_t^2 dt, \tag{3.65}$$

$$J_I dI_t = \mu I_t J_I dt. \tag{3.66}$$

As time is no own variable in the model we find

$$J_t dt = 0. \tag{3.67}$$

Inserting (3.65) - (3.67) into (3.64) gives after dividing by dt:

$$J_t = \mu I_t J_I + \frac{1}{2} J_{II} \sigma^2 I_t^2. \tag{3.68}$$

Using these results we can rewrite (3.58) as

$$\begin{aligned} -J_t &= \mu I_t J_I + \frac{1}{2} J_{II} \sigma^2 I_t^2 \\ &\quad + \max_{x_a, x_b} \{ \lambda_a (J(t, M_t + (p^* + x_a) Q', I_t - Q') - J(t, M_t, I_t)) \\ &\quad + \lambda_b (J(t, M_t - (p^* - x_b) Q', I_t + Q') - J(t, M_t, I_t)) \} . \end{aligned} \tag{3.69}$$

We denote τ as the remaining time or time horizon, i.e. $\tau = T - t.$. Therewith we have

$$J_t = J_\tau \frac{\partial \tau}{\partial t} = -J_\tau. \tag{3.70}$$

Define further $LJ = \mu I_t + \frac{1}{2}\sigma^2 I_t^2 J_{II}$, $BJ = J(t, M_t + p^* Q', I_t - Q')$, $SJ = J(t, M_t - p^* Q', I_t + Q')$ and $J = J(t, M_t, I_t)$. Using this notation we can approximate

$J(t, M_t + (p^* + x_a)Q', I_t - Q')$ and $J(t, M_t - (p^* + x_b)Q', I_t + Q')$ by a first order Taylor series around $(t, M_t + p^*Q', I_t - Q')$ and $(t, M_t - p^*Q', I_t + Q')$, respectively:

$$\begin{aligned}
J(t, M_t + p_aQ', I_t - Q') &= J(t, M_t + p^*Q', I_t - Q') + J_M x_a Q' \\
&= BJ + J_M x_a Q',
\end{aligned} \tag{3.71}$$

$$\begin{aligned}
J(t, M_t - p_bQ', I_t + Q') &= J(t, M_t - p^*Q', I_t + Q') + J_M x_b Q' \\
&= SJ + J_M x_a Q'.
\end{aligned} \tag{3.72}$$

Therewith (3.69) can be written as

$$\begin{aligned}
J_\tau = \; & LJ + \max_{x_a, x_b} \{\lambda_a(BJ - J) + \lambda_b(SJ - J) \\
& + \lambda_a x_a Q' B J_M + \lambda_b x_b Q' S J_M\}.
\end{aligned} \tag{3.73}$$

Inserting for λ_a and λ_b we can conduct this maximization and obtain the following first order conditions for the optimal fees x_a and x_b:

$$\begin{aligned}
\alpha_a - 2\beta_a x_a Q' B J_M - \beta_a(BJ - J) &= 0, \\
\alpha_b - 2\beta_b x_b Q' S J_M - \beta_b(SJ - J) &= 0.
\end{aligned} \tag{3.74}$$

The second order condition for a maximum can be shown to be fulfilled as the Hesse-Matrix is negative definite. Solving for the optimal fees we get:

$$\begin{aligned}
x_a &= \frac{\alpha_a}{2\beta_a} + \frac{J - BJ}{2Q'BJ_M}, \\
x_b &= \frac{\alpha_b}{2\beta_b} + \frac{J - SJ}{2Q'SJ_M}.
\end{aligned} \tag{3.75}$$

The first term can easily be verified to be the optimal fee at which benefits are maximized, provided that a predetermined trade occurs with certainty. The second term adjusts the fees for the risk in the return of the asset and the uncertainty of the order flow. To derive an explicit expression for the fees we have to find a solution for J. Ho and Stoll (1981) provide a separate mathematical appendix where they show how to derive a solution for J by further Taylor series approximations. We do not track down this long and tedious derivation, but only state the results they achieve if the time horizon is not too long.

The fees a monopolistic market maker charges to compensate for the risks he faces, i.e. the second terms in (3.75), depend positively on his risk aversion, the variance of the fundamental value and the trade size. The expressions are similar to those obtained when deriving the costs of market making. As a fourth factor the profits he can make influence his fees, the larger the expected profits are, the higher fees he charges. These profits depend on the size of the liquidity event (α_a and α_b) and the sensitivity to a higher fee (β_a and β_b). A final factor is the time horizon. The longer the time horizon the higher the fees. With a longer time horizon the market maker faces a higher total risk from holding the inventory, this effect offsets the benefits from the possibility of an offsetting order arriving at the market.

The formula provided in the appendix to Ho and Stoll (1981) gives no further substantial insight. The exact relations are subject to many approximations, such that it has to be handled with care, qualitative reasonings seem more appropriate than exact interpretations.

In comparing the results of the price setting for a monopolistic market maker and competitive market makers it can be seen that the main difference lies in the fact that a monopolistic market maker is not only compensated for the risk directly connected to his inventory, but also for the risks of the future order flow, i.e. for the risk of an unfavorable shift in his inventory. As this risk increases the longer his time horizon is, the higher the fee he charges.

3.2 Information-Based Models of Market Making

The last section focused on the influence of inventory costs on bid and ask prices and the spread. It has been assumed that all investors and market makers agree on the fundamental value of the asset, i.e. have equal information. In this section we will take up the line already laid in section 2 by assuming that there exist two groups of investors, informed and uninformed investors. Unlike in the models building on Kyle (1985) we will not aggregate the order flow over a given period of time, but examine the behavior on a trade-by-trade basis by introducing a market maker instead of a match maker.

The basis for the information-based models of market making has been laid by Bagehot (1971) with his distinction of market and trading gains. A *market gain* arises from the price change and dividends of an asset in a given time period. An investor can realize the market gain by holding the asset during the entire time period without trading. Let p_1 denote the price of the asset at the end of the period and p_0 at the beginning, then with assuming that no dividends are paid, the market gain is

$$\Delta p = p_1 - p_0. \tag{3.76}$$

The market gain is the same for all investors holding the asset at the beginning and at the end of the period. Let us now assume that all investors hold the same amount of assets at the beginning and at the end of the period.[17] Investors can either only hold the asset over the entire period or they can trade with each other. The gain associated with this trading activity is the *trading gain*, denoted f^i. The total gain of investor i is

$$\pi^i = \Delta p + f^i. \tag{3.77}$$

The total gain of all investors can only be the market gain by comparing the wealth at the beginning and at the end of the period, i.e. with N investors it is

$$\pi = \sum_{i=1}^{N} \pi^i = \sum_{i=1}^{N} \Delta p = N\Delta p. \tag{3.78}$$

By aggregating (3.77) over all investors we get

$$\pi = \sum_{i=1}^{N} \pi^i = \sum_{i=1}^{N} (\Delta p + f^i) = N\Delta p + \sum_{i=1}^{N} f^i. \tag{3.79}$$

Inserting from (3.78) we receive

$$\sum_{i=1}^{N} f^i = 0, \tag{3.80}$$

[17] If some investors decide to liquidate their position in the asset there has to be found another investor taking his position as the number of outstanding shares is fixed. We therefore could aggregate these investors such that the assumption is fulfilled.

i.e. the total trading gains are zero. For every investor making trading profits there must be another making a loss. If there are two groups of investors, informed and uninformed, where the uninformed investors have to trade for exogenous reasons, the informed investors will only trade if they expect to profit from trading, while the uninformed are forced to trade even if they make trading losses.

With a market maker acting as counterpart in every trade, the informed investors will only trade with him if he sets a price at which they make an expected profit, otherwise they would refuse to trade. Hence, if we allow only for a single trade per period of time, we see from (3.80) that as there are only two market participants involved, the market maker will make a loss from trading with an informed investor. He will try to reduce the loss he receives from trading with an informed investor by quoting a less favorable price than what he thinks the fundamental value is, but he will never be able to make a gain. This remaining loss he has to offset from another source. This other source are uninformed investors, he has to charge a price to them such that he makes a gain from trading with them and they make a loss.

By the anonymity of the two investors provided by a broker, the market maker does not know whether the investor he trades with is informed or uninformed. We will see below that this problem of loosing to one and gaining from the other group gives rise to the spread. The losses to informed investors are also called *adverse selection costs*.

Bagehot (1971) further states that market makers will be uninformed. They observe the order flow and try to balance the buy and sell orders by quoting appropriate prices. By observing the order flow they aggregate the information available in the market as well as the errors. Errors cancel out by the law of large numbers if the number of informed investors is sufficiently high and they make no systematic errors. If the market maker would become informed and quote his prices according to his own information, he faces the risk of relying on a large error in his considerations, causing him a large loss. Therefore he will not invest his wealth to become informed.[18]

3.2.1 Determination of adverse selection costs

The first to formalize the idea of Bagehot (1971) were Copeland and Galai (1983). They provide a simple framework in which only a single trade takes place before the information is fully revealed to all market participants, e.g. by liquidation of the asset at the fundamental value. The liquidation value of the asset is a random variable, $\widetilde{p}$, which has a known distribution F and an expected value of $E[\widetilde{p}] = p_0$. It is assumed that informed investors know the exact liquidation value, p^*, before trading takes place. Uninformed investors trade for exogenous reasons, but their demands depend on the fee charged by the market maker, the larger the fee, the smaller their demand. The market maker is assumed to be risk neutral, i.e. he faces no inventory costs.

Trading takes place as follows: At the beginning informed investors become to know the liquidation value and the market maker sets his prices. Knowing these prices informed and uninformed investors place their orders. Only one of these orders will be served by the market maker, this order is chosen randomly. The probability that the order chosen is from an informed investor is γ_I. This probability depends on the prices he quotes. The larger the fee from the point of an uninformed investor, the less uninformed investors submit orders and hence γ_I increases. An informed investor will submit a buy order if $p_a < p^*$ and a sell order if $p^* < p_b$, if $p_b < p^* < p_a$ he will not submit an order and all orders in the market are from uninformed investors. As the market maker does not know p^*, he does not know what his loss is, he only knows that he makes a loss. His expected profit from trading with an informed investor is

$$E[\pi_I] = \int_{p_a}^{\infty} (p_a - p)dF(p) + \int_{0}^{p_b} (p - p_b)dF(p) \leq 0. \tag{3.81}$$

[18] Although the assumption of uninformed market makers is common in dealer markets, it can reasonably be assumed that market makers can get private information from other sources, such that he will not be completely uninformed. Calcagno and Lovo (1998) provide a model where market makers are at least partially informed and have different information. However, we consider this line of research not further at this point.

If he trades with an uninformed investor the agreed value of the asset is p_0, he will make a profit of $p_a - p_0$ from a trade at the ask and of $p_0 - p_b$ from a trade at the bid. Uninformed investors can either buy or sell the asset, we assign a probability of γ_a for a trade at the ask and γ_b for a trade at the bid, such that $\gamma_a + \gamma_b = 1$. These probabilities depend on the fees charged by the market maker, or equivalently by the prices he sets, such that

$$\begin{aligned} &\frac{\partial \gamma_U}{\partial p_a} > 0, &\qquad &\frac{\partial \gamma_U}{\partial p_b} < 0, \\ &\frac{\partial \gamma_a}{\partial p_a} < 0, &\qquad &\frac{\partial \gamma_b}{\partial p_b} > 0, \\ &\frac{\partial \gamma_a}{\partial p_b} = 0, &\qquad &\frac{\partial \gamma_b}{\partial p_a} = 0. \end{aligned} \tag{3.82}$$

The expected profits from trading with an uninformed investor are therewith

$$E[\pi_U] = \gamma_a(p_a - p_0) + (1 - \gamma_a)(p_0 - p_b) \geq 0. \tag{3.83}$$

The total expected profits of the market maker are

$$E[\pi] = \gamma_I E[\pi_I] + (1 - \gamma_I)E[\pi_U]. \tag{3.84}$$

The costs or reservation prices are obtained by setting $E[\pi] = 0$. A monopolistic market maker would maximize (3.84).

Solving the problem to determine the adverse selection costs would rely on many assumptions and is therefore not conducted here. Nevertheless we can use (3.84) to derive some implications of the model. It can be shown that we always have a positive spread if $\gamma_I > 0$, i.e. if there are informed investors. Suppose that

$p_a = p_b = p'$, we then get from (3.81):

$$
\begin{aligned}
(3.85)\quad E[\pi_I] &= \int_{p'}^{\infty}(p'-p)dF(p) + \int_0^{p'}(p-p')dF(p) \\
&= p'\int_{p'}^{\infty}dF(p) - \int_{p'}^{\infty}pdF(p) + \int_0^{p'}pdF(p) - p'\int_0^{p'}dF(p) \\
&= p' - p'\int_0^{p'}dF(p) - E[p] + \int_0^{p'}pdF(p) \\
&\quad + \int_0^{p'}pdF(p) - p'\int_0^{p'}dF(p) - p'\int_0^{p'}dF(p) \\
&= p' - p_0 + 2\int_0^{p'}(p-p')dF(p).
\end{aligned}
$$

As $\int_0^{p'}(p-p')dF(p) < 0$ we get for $p' < p_0$ that $E[\pi_I] < 0$ and as also $\int_0^{p'}(p-p')dF(p) < p_0 - p'$ we have for $p' \geq p_0$ $E[\pi_I] < p_0 - p' < 0$. Hence we always have $E[\pi_I] < 0$. From (3.83) we get

$$
(3.86)\qquad E[\pi_U] = \gamma_a(p'-p_0) + (1-\gamma_a)(p_0-p') = (p'-p_0)(2\gamma_a - 1).
$$

If $p' < p_0$ we need $\gamma_a < \frac{1}{2}$ for $E[\pi_U] > 0$, what is necessary to compensate for the loss from trading with informed investors. This would imply that the probability of a trade at the bid is higher than a trade at the ask, although the fee for a trade at the bid is positive while it is negative for a trade at the ask. If we reasonably assume that the probabilities for a trade are equal for the same fee this contradicts (3.82), hence we cannot set $p' < p_0$. With the same argumentation the case $p' > p_0$ can be ruled out. Hence the only solution is to set $p' = p_0$, for which $E[\pi_U] = 0$. Inserting these findings into (3.84) we get

$$
(3.87)\qquad E[\pi] = \gamma_I E[\pi_I] < 0.
$$

As a market maker will not accept to make a loss the spread always has to be positive if $\gamma_I > 0$.[19]

Further we can see from (3.84) that the more informed investors are present, i.e. the higher γ_I is, the larger the first term becomes. To compensate the losses

[19] A negative spread has earlier been pointed out to be not possible by arguments of arbitrage.

from a higher probability of trading with informed investors a larger spread has to be quoted.[20]

If the uncertainty about the liquidation value increases, i.e. more probability is put on the tails of F, e.g. with a larger variance, we can see from (3.81) that the expected losses from trading with informed investors increase. This is compensated by the market maker with quoting a larger spread.

The results found are very similar to those of KYLE (1985) in the case of a single auction. As with the Kyle model it would therefore be interesting to extend this static model to a dynamic model by allowing more than a single trade in a given time period before the asset is liquidated. This generalization has been undertaken by the model of GLOSTEN AND MILGROM (1985).

With only a single trade before the asset is liquidated there is no need to exploit information from the order flow. If there are more trades, information from the order flow will be used by the market maker to minimize his losses from trading with an informed investor through updating his beliefs.

The market maker uses all information available to him from former trades, denoted Ω_t. He knows that an informed investor only trades if the quoted price is above (trade at the bid) or below (trade at the ask) the liquidation value, while an uninformed investor trades independent of the quotes and fundamental value. We only assume that the frequency with which he trades is sensitive to the fee the market maker charges. His belief on the fundamental value equals that of the market maker as both are uninformed, but the market maker can observe the former order flows.

The probability that a buy or a sell order arrives at the market are given by

$$
\begin{aligned}
(3.88) \qquad Pr(\text{Buy order}|\Omega_t) &= \gamma_I Pr(p^* > p_a^t|\Omega_t) + (1-\gamma_I)\gamma_a, \\
(3.89) \qquad Pr(\text{Sell order}|\Omega_t) &= \gamma_I Pr(p^* < p_b^t|\Omega_t) + (1-\gamma_I)(1-\gamma_a).
\end{aligned}
$$

[20] In general it cannot be determined how this larger spread is achieved, by increasing only the ask, decreasing only the bid price or a combination of these. ADMATI AND PFLEIDERER (1988) provide an interesting strategy of the market makers in setting their prices as will be presented in section 3.2.4.

Let us denote the order form by o_t, where

$$(3.90)\qquad o_t = \begin{cases} 1 & \text{if the order is a buy order} \\ -1 & \text{if the order is a sell order} \end{cases},$$

then $\Omega_t = \{o_1, \ldots, o_{t-1}\}$. The expected profit of the market maker is given by

$$(3.91)\qquad \begin{aligned} E[\pi_t|\Omega_t] &= (p_a^t - E[p|\Omega_t, o_t = 1])Pr(o_t = 1|\Omega_t) \\ &\quad +(E[p|\Omega_t, o_t = 1] - p_b^t)Pr(o_t = -1|\Omega_t). \end{aligned}$$

In order to determine the costs of market making his expected profits have to equal zero. He can achieve this either by setting both prices so that he makes no profit on either side or he can set the prices so that he makes a profit on one side of the trade and a loss on the other. If we assume competitive market makers the profits on one side of the trade will be deteriorated by other market makers undercutting the price by applying another price strategy, hence the second alternative is ruled out as all market makers have the same costs. Therewith the bid and ask prices are determined as follows:

$$(3.92)\qquad \begin{aligned} p_a^t &= E[p|\Omega_t, o_t = 1], \\ p_b^t &= E[p|\Omega_t, o_t = -1]. \end{aligned}$$

With Bayes rule and (3.88) and (3.89) we get

$$(3.93)\qquad \begin{aligned} Pr(p^* > p^t) &= Pr(p^* > p_a^t|\Omega_t)Pr(o_t = 1|p^* > p_a^t, \Omega_t) \times \\ &\quad \times \Big(Pr(p^* > p_a^t|\Omega_t)Pr(o_t = 1|p^* > p_a^t, \Omega_t) \\ &\quad +Pr(p^* \leq p_a^t|\Omega_t)Pr(o_t = 1|p^* \leq p_a^t, \Omega_t) \Big)^{-1} \\ &= Pr(p^* > p_a^t|\Omega_t)(\gamma_I + (1-\gamma_I)\gamma_a) \times \\ &\quad \Big(\times Pr(p^* > p_a^t|\Omega_t)(\gamma_I + (1-\gamma_I)\gamma_a) \\ &\quad +(1 - Pr(p^* > p_a^t|\Omega_t))(1-\gamma_I)\gamma_a \Big)^{-1} \\ &= Pr(p^* > p_a^t|\Omega_t)\frac{\gamma_I + (1-\gamma_I)\gamma_a}{Pr(p^* > p_a^t|\Omega_t)\gamma_I + (1-\gamma_I)\gamma_a} \\ &\quad \begin{cases} = & Pr(p^* > p_a^t|\Omega_t) \quad \text{if} \quad \gamma_I = 0 \\ > & Pr(p^* > p_a^t|\Omega_t) \quad \text{if} \quad \gamma_I > 0 \end{cases}. \end{aligned}$$

As long as there are informed investors a buy order increases the probability that the fundamental value is above the ask. A similar result can be obtained for a trade at the bid, it increases the probability that the fundamental value is below the bid. With this result we get

$$
\begin{aligned}
(3.94) \qquad p_a^t &= E[p_a^t|o_t = 1] = E[E[p|\Omega_t, o_t = 1]|o_t = 1] \\
&= E[p|\Omega_t, o_t = 1] \\
& \quad \begin{cases} = & E[p|\Omega_t] \quad \text{if} \quad \gamma_I = 0 \\ > & E[p|\Omega_t] \quad \text{if} \quad \gamma_I > 0 \end{cases},
\end{aligned}
$$

i.e. as long as there is the threat of adverse selection the ask will exceed the fundamental value assigned by the market maker. The same result can be obtained for the bid such that

$$
\begin{aligned}
(3.95) \qquad p_a^t &= E[p|\Omega_t] = p_b^t \quad \text{if} \quad \gamma_I = 0, \\
p_a^t &> E[p|\Omega_t] > p_b^t \quad \text{if} \quad \gamma_I > 0.
\end{aligned}
$$

If there is the possibility of trading with an informed investor the spread will always be positive due to adverse selection costs. As we can easily see, (3.93) is increasing in the fraction of informed investors, γ_I. Therewith from (3.94) we see that p_a^t increases and analogue p_b^t decreases, hence the spread increases in γ_I.

Let p^t denote the price at which a trade occurs at time t, i.e. $p^t = p_a^t$ if $o_t = 1$ and $p^t = p_b^t$ if $o_t = -1$. By using (3.92) we get

$$
(3.96) \qquad E[p^{t+1}|\Omega_t] = E[E[p|\Omega_t, o_t]|\Omega_t] = E[p|\Omega_t] = E[p|\Omega_{t-1}, o_t] = p^t.
$$

This implies that expected price changes between two subsequent trades are independent of each other if the fundamental value does not change. These findings are in contrast to the behavior with inventory costs where price changes are negatively correlated.

From (3.92), (3.94) and (3.95) we get

$$
\begin{aligned}
p_a^{t+1} &= E[p|\Omega_{t+1}, o_{t+1} = 1] \\
&= E[E[p|\Omega_{t+1}]|o_{t+1} = 1] \\
&= E[E[p|\Omega_t, o_t]|o_{t+1} = 1] \\
&= \begin{cases} E[p_a^t|o_{t+1} = 1] & \text{if } \ o_t = 1 \\ E[p_b^t|o_{t+1} = 1] & \text{if } \ o_t = -1 \end{cases} \\
&> \begin{cases} E[p|\Omega_{t+1}] & \text{if } \ o_t = 1 \\ E[p|\Omega_{t+1}] & \text{if } \ o_t = -1 \end{cases} \\
&= \begin{cases} p_a^t & \text{if } \ o_t = 1 \\ p_b^t & \text{if } \ o_t = -1 \end{cases} \\
p_b^{t+1} &= E[p|\Omega_{t+1}, o_{t+1} = -1] \\
&= E[E[p|\Omega_{t+1}]|o_{t+1} = -1] \\
&= E[E[p|\Omega_t, o_t]|o_{t+1} = -1] \\
&= \begin{cases} E[p_a^t|o_{t+1} = -1] & \text{if } \ o_t = 1 \\ E[p_b^t|o_{t+1} = -1] & \text{if } \ o_t = -1 \end{cases} \\
&< \begin{cases} E[p|\Omega_{t+1}] & \text{if } \ o_t = 1 \\ E[p|\Omega_{t+1}] & \text{if } \ o_t = -1 \end{cases} \\
&= \begin{cases} p_a^t & \text{if } \ o_t = 1 \\ p_b^t & \text{if } \ o_t = -1 \end{cases} .
\end{aligned}
\tag{3.97}
$$

If the previous transaction has taken place at the ask (bid) the new ask (bid) price is higher (lower) than the previous ask (bid) price. We see further that always $p_a^{t+1} > p_b^t$ and $p_b^{t+1} < p_a^t$, i.e. the prices will never be revised so much that both prices are outside the former spread. This price behavior can easily be explained by looking at (3.92). The former ask and bid prices were the best guess of the market maker given the transactions he was waiting for, hence from (3.95) we see that $p_a^{t+1} > E[p|\Omega_{t+1}] = p^t > p_b^{t+1}$.

Define the spread as $s_t = p_a^t - p_b^t$ and $\xi_t = \frac{1}{Pr(o_t=1|\Omega_t)Pr(o_t=-1|\Omega_t)}$, $\overline{\xi} = \frac{1}{T}\sum_{t=1}^{T}\xi_t$,

$\overline{s} = \frac{1}{T}\sum_{t=1}^{T} s_t$ and $e_t = E[(p^t - p^{t-1})^2|\Omega_t]$. We then find with $p^0 = E[p]$

$$\begin{aligned} Var[p] &\geq Var[p^T] = Var\left[\sum_{t=1}^{T}(p^t - p^{t-1})\right] \\ &= \sum_{t=1}^{T} Var[p^t - p^{t-1}] = \sum_{t=1}^{T} E[(p^t - p^{t-1})^2] \\ &= E\left[\sum_{t=1}^{T} E[(p^t - p^{t-1})^2|\Omega_t]\right] \\ &= E\left[\sum_{t=1}^{T} e_t\right]. \end{aligned} \tag{3.98}$$

We further know that the bid and ask prices are increased following a trade at the ask in the previous period and decreased following a trade at the bid. Therewith we get

$$\begin{aligned} e_t &= (p_a^{t-1} - p_b^t)^2 Pr(o_{t-1} = 1|\Omega_{t-1}) Pr(o_t = -1|\Omega_{t-1}) \\ &\quad + (p_b^{t-1} - p_a^t)^2 Pr(o_{t-1} = -1|\Omega_{t-1}) Pr(o_t = 1|\Omega_{t-1}) \\ &\quad + (p_a^{t-1} - p_a^t)^2 Pr(o_{t-1} = 1|\Omega_{t-1}) Pr(o_t = 1|\Omega_{t-1}) \\ &\quad + (p_b^{t-1} - p_b^t)^2 Pr(o_{t-1} = -1|\Omega_{t-1}) Pr(o_t = -1|\Omega_{t-1}) \\ &\geq (p_a^t - p_b^t)^2 Pr(o_t = 1|\Omega_t) Pr(o_t = -1|\Omega_t) \\ &= \frac{s_t^2}{\xi_t}. \end{aligned} \tag{3.99}$$

Therewith we have found that

$$e_t \xi_t \geq s_t^2. \tag{3.100}$$

Aggregating over all T trades we get

$$\sum_{t=1}^{T} e_t \xi_t \geq \sum_{t=1}^{T} s_t^2. \tag{3.101}$$

With the Cauchy-Schwartz inequality this can be transformed into

$$\sum_{t=1}^{T} e_t \sum_{t=1}^{T} \xi_t \geq \left(\sum_{t=1}^{T} s_t\right)^2. \tag{3.102}$$

Hence we get with the above definitions

$$\sum_{t=1}^{T} e_t \geq \frac{T^2 \overline{s}^2}{T\overline{\xi}} = \frac{T\overline{s}^2}{\overline{\xi}}, \tag{3.103}$$

and by taking expectations we get with (3.98):

$$E[\overline{s}^2] \leq \frac{\overline{\xi} Var[p]}{T}. \tag{3.104}$$

If $\overline{\xi}$ is bounded[21] we find the average squared spread to be independent of any trade patterns. We can use (3.104) to derive a relationship between the average spread and the average trading volume in a given period of time. As we have assumed that within a given period of time T trades of a fixed size occur, a large T implies a high trading volume. As can be seen from (3.104) and as ξ_t and $Var[p]$ are independent of T we expect a small average spread.

With a large number of trades the information of informed investors is revealed much faster to the market maker than with only few trades. This reduces the adverse selection costs and hence the spread. The initial adverse selection costs at the first trades, represented by $Var[p]$ increase the average spread.

As we saw in the Holden and Subrahmanyam (1992) model, trading will be especially intense shortly after information has been released to informed investors, in our case the early trades of the time period, causing adverse selection costs to increase with γ_I in the first trades. The result obtained here can explain why spreads in actively traded stocks are smaller than in less actively traded stocks.

Krishnan (1992) showed that the Kyle (1985) and Glosten and Milgrom (1985) models are equivalent if we restrict the Kyle (1985) model slightly. Although Kyle (1985) allows for no market makers and no spread, the prices at which trades occur turn out to be identical if we allow only for fixed order imbalances of the same size as an order size here. In this case the price determined by the match maker equals the bid and ask price of the market maker. The properties of the prices are identical in both cases.

Although the model of Glosten and Milgrom (1985) model gave further insights into the behavior of bid and ask prices and the spread, no explicit formula

[21] For $\overline{\xi}$ to be bounded $\frac{1}{T}\sum_{t=1}^{T}\xi_t$ has to converge, what is achieved if the market maker sets prices always low enough such that both, $Pr(o_t = 1|\Omega_t)$ and $Pr(o_t = -1|\Omega_t)$ are bounded away from zero, i.e. if he can assure a sufficient large trading of uninformed investors.

could be derived directly, only the equivalence with the model of KYLE (1985) and its further extensions, such formulas can easily be provided.

3.2.2 Simultaneous trading at different stock exchanges

The result that competitive market makers set prices equal to the expected fundamental value given the order flow is due to the assumption of Bertrand competition between market makers. Although not explicitly stated in the models presented thus far, it has been assumed that investors route their entire order flow to a single market maker quoting the most favorable price. As market makers quoting less favorable prices do not participate in the order flow, they make zero profits, ignoring fixed costs. Hence Bertrand competition requires market makers to quote prices such that they make expected profits of zero, which has been shown in the previous section to imply quotes that equal market maker's inference about the fundamental value.

In the models considered the assumption of strict price priority ensured the entire order flow to be routed to a single market maker quoting the best price for the entire order. As has been shown that the quoted prices become less favorable the larger the order size is, it can be profitable for investors to split their orders between market makers to reduce the order size submitted to each market maker and therewith receive more favorable prices.[22] For this reason also market makers quoting not the most favorable price will receive a fraction of the order flow, the less favorable the quotes are, the smaller this fraction will be. The optimal splitting of the order flow will give an investor equal costs for trading with every market maker.

Splitting the order flow requires that investors are able to route their orders directly to different market makers, violating strict price priority. A possibility to achieve such a situation is the assumption that an asset is traded at several stock exchanges and that trading at these stock exchanges can take place simultane-

[22] It is common to assume that market makers condition their quotes only on the order flow they receive, but not on the order flow the other market makers receive, hence such a splitting of the order flow would not influence the quoted prices.

ously. Given this possibility to split the order flow Bernhardt and Hughson (1997) show that competitive market makers are able to make positive expected profits by quoting less competitive prices.

The intuition behind their result is that when quoting not the most favorable price, market makers still will receive a fraction of the order flow, whereas if investors are not allowed to split the order flow they would receive no orders. This situation enables market makers, as formally shown in Bernhardt and Hughson (1997), to quote less competitive prices, i.e. higher ask and lower bid prices, and make positive expected profits. These less favorable prices increase the trading costs for investors. They further show that prices become more competitive by increasing the number of competing market makers, in the limiting case, with an infinite number of competing market makers, the quoted prices are competitive.[23]

Furthermore Dennert (1993) points out that as a result of splitting the order flow the share of informed trades increases. Therewith adverse selection costs increase as shown in the previous subsection, what increases the trading costs further.

With these two effects due to the possibility to split the order flow, e.g. by trading simultaneously at different stock exchanges, competition among market makers not necessarily reduces trading costs as usually expected.

3.2.3 Comparing competitive and monopolistic market makers

If the number of informed investors is increased, it has been shown above that the adverse selection costs increase and therewith the fee charged by the market maker. If we assume that uninformed investors demand responds to the fee charged, their demand reduces further, the share of informed investors increases and again the fee increases. If the adverse selection costs are high enough, the

[23] Bernhardt and Hughson (1997) also show that a linear equilibrium as derived in Kyle (1985) only exists with enabling investors to split the order flow, if the demand of liquidity traders for trading are price elastic.

demand of uninformed investors nearly vanishes and the market maker cannot offset his losses from trading with the informed investors. He has to quote such high fees that the market breaks down.

The market breakdown as the result of too much informed investors is due to the need of the market maker to achieve zero expected profits. It has been shown in HOLDEN AND SUBRAHMANYAM (1992) that especially shortly after new information is available, informed investors trade very actively on this information. Afterwards the information is revealed through prices, the trades of informed investors reduce and with it the adverse selection costs. However if at the beginning the market breaks down, prices never will be able to reveal information and the threat remains *ad infinitum*.

GLOSTEN (1989) derives a model where, with further assumptions like a normally distributed liquidation value, conditions are derived under which the market breaks down as the result of too high adverse selection costs with competitively acting market makers. These results are in line with the argumentation thus far and are therefore not presented here in more detail.

To overcome the problem of a market breakdown he proposes to establish a monopolistic market maker. As such a market maker maximizes his total profits over a given time period, he not necessarily has to avoid losses from every trade, he can be compensated for losses by larger profits in future trades. GLOSTEN (1989) shows that the optimal price setting strategy of a risk neutral monopolistic market maker is to set an average price that enables him to make extraordinary profits in normal trading environments and compensates him for losses in times of high adverse selection costs.

By incurring a loss shortly after new information has become available to informed investors he keeps the market open and learns the information through orders he receives. After having learnt this information sufficiently well, his adverse selection costs are reduced and he can make larger profits compensating him for the incurred loss. By shutting the market down he would not be able to learn the information and hence would make no profits.

As long as the adverse selection costs are higher than a certain threshold, a monopolistic market maker quotes prices more favorable to investors than a competitive market maker would be able to do. Only if adverse selection costs are low, competing market makers charge a smaller fee. This result is different to the result obtained for inventory-based models, where a monopolistic market maker always quotes less favorable prices than competitive market makers.

We should therefore find competing market makers in markets with only small adverse selection costs, i.e. only few informed investors, while for markets with high adverse selection costs, i.e. many informed investors, a monopolistic market maker would be preferred.

The NASDAQ has a system of competing market makers, while the NYSE has a monopolistic market maker (specialist). The many analysts following the companies at the NYSE make adverse selection costs much higher at the NYSE than at the NASDAQ, where especially the information of the informed investors are much less precise than at the NYSE, as the companies mostly work in a more dynamic environment. This would be a rationale for the NYSE to have a monopolistic market maker and the NASDAQ to have competing market makers.

3.2.4 Explaining return patterns

There exists significant evidence of systematic patterns in asset returns, as well within a trading day as across trading days. The most prominent of this effect is the Monday-effect, where the returns on Mondays are negative on average. Admati and Pfleiderer (1989) present a model that is able to explain such a pattern.

In section 2.3 the model of Admati and Pfleiderer (1988) has been presented, that showed the concentration of trading in a few periods. This pattern in trading volume arises endogenously out of the model as uninformed investors want to reduce their adverse selection costs from trading with informed investors. Nevertheless such behavior was not able to explain patterns in returns.

By introducing a market maker we are able to explain such patterns in expected returns. As the uninformed investors react sensitive to the fee charged by the market maker, quoting a high fee reduces the trading of uninformed investors and increases adverse selection costs. If the market maker quotes a low fee his profits are small, but also his adverse selection costs are reduced by the increased number of uninformed investors trading.

By quoting a very low fee on one side of the trade, e.g. the ask side, and a high fee on the other side of the trade, in our example the bid side, the market maker will induce many uninformed trades on the ask side and virtually a breakdown of the market on the bid side. Although this price setting behavior causes a large order imbalance, the ask price does not change significantly after every trade as most trades are from uninformed investors. On the bid side nearly no uninformed investors trade due to the high costs, nearly all trades come from informed investors, hence an order submitted to the bid side is much more informative than at the ask side. The high fee charged and the information revelation reduces the profits of informed investors from trading on their information. If it is known that the market maker will reverse his strategy in the next period of time, e.g. the next trading day, by quoting low fees at the bid and high fees at the ask side, an informed investor will prefer to wait until this period as his profits are increased, provided that his information is not revealed beforehand by other means. The market maker will have to reverse his strategy in order to offset his large inventory position he had to acquire as a result of the large order imbalance in the first period.

With this quoting strategy the market maker does not only induce a pattern in trading volume, but also a pattern in the types of trades that occur. In the example above the market maker induces more trades at the ask in the first and more trades at the bid in the second period. Admati and Pfleiderer (1989) show that such a strategy is an equilibrium for a monopolistic market maker as well as for competitive market makers. They furthermore show that the concentration of trading at the bid and at the ask under certain assumptions do not concentrate in the same period like in Admati and Pfleiderer (1988).

If the trading at the ask concentrates in a period and the trading at the bid in the next period, it is more likely to observe an ask price in the first and a bid price in the second period. Therewith the expected return can easily be shown to be negative between these two periods. If in the next period trading again concentrates at the ask, the expected return will be positive.

The model presented here shows that patterns in expected returns can arise, but not how they are timed, neither that they have to be timed equally by all market makers. This timing has to be explained by exogenous behavior of investors. The Monday-effect suggests with this model a concentration of buy orders on Fridays and a concentration of sell orders on Mondays. As over the weekend no trading is possible and there are many investment funds that have to invest money newly acquired, they invest it on Fridays. Over the past years there has been a massive flow into funds, such that they have to invest, i.e. buy. Their need to buy can be justified by the frequent benchmarking with an index, as the closing of an index is taken as their benchmark, they best can track the index by trading near the close, i.e. the end of the day and the end of the week. This causes a concentration of trading at the ask on Fridays, resulting in an expected loss on Mondays as well as for the beginning of the next trading day, an effect that also is confirmed empirically.

3.3 Dealer markets with multiple assets

We assume a market with $L > 1$ different risky assets. Each asset is assigned to a single market maker, who is granted a monopoly to act as market maker for this asset. This market structure can be found, e.g. at the NYSE, where these market makers are called specialists. We allow each specialist to make the market for more than a single asset, i.e. be specialist in more than one asset.[24] We have $K \leq L$ specialists. The specialists are assumed to be risk neutral, i.e. they face no inventory costs. It is further assumed that no private information is

[24] Hagerty (1991) allows every specialist only to make the market in one asset, so that the framework presented here is a generalization of her setting.

in the market to avoid the problem of adverse selection costs. The aim of these assumptions is to concentrate on the effect of execution of monopoly power in this environment. In such a market the spread of competitive market makers would be zero as we know from sections 3.1 and 3.2.

Following Gehrig and Jackson (1998) we have two groups of investors, indexed $i = \{1, 2\}$. Non-optimal endowments in their portfolio holdings induce them to trade, i.e. trading takes place as the result of a liquidity event. One group buys an asset and the other sells it. Let the endowment of investors in group i of asset j be denoted e_j^i and $e^i = (e_1^i, \ldots, e_L^i)$ is the vector of endowments. The total supply is $e = e^1 + e^2$.

Specialist k controls a subset L_k of the L assets. Specialists set prices such that the market clears, i.e. demand equals supply. Such prices exist if we assume the specialist to know the endowments of the investors and all other relevant variables.

The risky assets are traded in a single round of trading, after the trade has occurred they are liquidated at the fundamental value. The fundamental value of asset j is denoted $\widetilde{v}_j$ and has a mean a mean of μ_j and a covariance matrix Σ, which we assume to be positive definite and non singular.

A risk averse investor will determine the demand such that he maximizes his expected utility of terminal wealth, which is given by

$$\widetilde{W}^i = \sum_{j=1}^{L} \left(x_j^i \widetilde{v}_j + p_{b_j} \max(e_j^i - x_j^i, 0) + p_{a_j} \min(e_j^i - x_j^i, 0)) \right), \tag{3.105}$$

where x_j^i denotes the demand of investor i for asset j, p_{b_j} the bid and p_{a_j} the ask price of asset j.The final terms denote the change in money holding as a result of trading and the first term the return from holding the asset. With a competitive market maker we have shown earlier that in this environment $p_{b_j} = p_{a_j} = p_j^c$.

Inserting this we get

$$
\begin{aligned}
(3.106) \qquad \widetilde{W}^i &= \sum_{j=1}^{L} \left(x_j^i \widetilde{v}_j + p_j^c (e_j^i - x_j^i) \right), \\
(3.107) \qquad E[\widetilde{W}^i] &= \sum_{j=1}^{L} \left(x_j^i \mu_j + p_j^c (e_j^i - x_j^i \right) = x^{i\,\prime} \mu + p^{c\,\prime} (e^i - x^i), \\
(3.108) \qquad Var[\widetilde{W}^i] &= \sum_{j=1}^{L} (x_j^i)^2 \Sigma_{jj} + \sum_{j=1}^{L} \sum_{k=1, k \neq j}^{L} x_j^i x_k^i \Sigma_{jk} \\
&= x^{i\,\prime} \Sigma x^i.
\end{aligned}
$$

The expected utility is then given by

$$
(3.109) \qquad E[U(\widetilde{W}^i)] = U(E[\widetilde{W}^i] - \frac{1}{2} z Var[\widetilde{W}^i]),
$$

inserting and differentiating gives the first order condition for a maximum:

$$
(3.110) \qquad \mu - p^c - z \Sigma x^i = 0.
$$

Solving for the optimal demand we get

$$
(3.111) \qquad x^i = \frac{1}{z} \Sigma^{-1} (\mu - p^c).
$$

The second order condition $-z\Sigma < 0$ is fulfilled with the assumption of a positive definite covariance matrix. The market clearing condition $x^1 + x^2 = e$ gives us the competitive price as

$$
(3.112) \qquad p^c = \mu - \frac{1}{2} z \Sigma e.
$$

Let $\theta^i = x^i - e^i$ denote the vector of trades investor i conducts. If the specialist has market power he is able to charge different prices for the two groups of investors, denoted p^i. Therewith (3.111) becomes

$$
(3.113) \qquad x^i = \frac{1}{z} \Sigma^{-1} (\mu - p^i).
$$

Market clearing requires that $\theta^1 + \theta^2 = 0$:

$$
\begin{aligned}
(3.114) \qquad 0 &= \theta^1 + \theta^2 = x^1 + x^2 - e \\
&= \frac{1}{z} \Sigma^{-1} (2\mu - p^1 - p^2) - e.
\end{aligned}
$$

Solving (3.112) for e and inserting into (3.114) gives

$$\frac{1}{z}\Sigma^{-1}(2\mu - p^1 - p^2) - \frac{2}{z}\Sigma^{-1}(\mu - p^c) = \frac{1}{z}\Sigma^{-1}(2p^c - p^1 - p^2) = 0, \tag{3.115}$$

what requires

$$p^1 - p^c = p^c - p^2. \tag{3.116}$$

The specialist as being risk neutral maximizes his total profits. With R_j^l denoting the revenues from asset j for specialist l we get with the market clearing condition $\theta^1 + \theta^2 = 0$ and (3.116):

$$R_j^l = \theta_j^1 p_j^1 + \theta_j^2 p_j^2 = \theta_j^1 p_j^1 - \theta_j^1(2p_j^c - p_j^1) = 2\theta_j^1(p_j^1 - p_j^c). \tag{3.117}$$

The total revenues of specialist l are given by

$$R^l = \sum_{j\epsilon L_l} R_j^l = 2\sum_{j\epsilon L_l} \theta_j^1(p_j^1 - p_j^c) = 2\sum_{j\epsilon L_l}(x_j^1 - e_j^1)(p_j^1 - p_j^c). \tag{3.118}$$

Differentiating with respect to p_j^1 for all $j \in L_l$ we get with (3.113) the first order condition

$$\frac{1}{z}\left(\Sigma^{-1}\right)_j'(\mu - p^1) - e_j^1 - \frac{1}{z}\sum_{i\epsilon L_l}\Sigma_{ij}^{-1}(p_i^1 - p_i^c) = 0, \tag{3.119}$$

where $\left(\Sigma^{-1}\right)_j$ denotes that jth row of Σ^{-1} and Σ_{ij}^{-1} the (i,j)th element of Σ^{-1}. Defining

$$A_{ij} = \begin{cases} 2\Sigma_{ij}^{-1} & \text{if } \; i = j \; \text{ or } \; i,j \in L_l \\ \Sigma_{ij}^{-1} & \text{else} \end{cases} \tag{3.120}$$

we can rewrite (3.119) in vector form as

$$\frac{1}{z}\Sigma^{-1}(\mu - p^c) - e^1 - \frac{1}{z}A(p^1 - p^c) = 0, \tag{3.121}$$

where from (3.111) and (3.113)

$$\mu - p^c = z\Sigma x^1 = z\Sigma\frac{1}{z}\Sigma^{-1}(\mu - p^1) = \mu - p^1. \tag{3.122}$$

Solving (3.121) for $p^1 - p^c$ we get

$$p^1 - p^c = A^{-1}\left(\Sigma^{-1}(\mu - p^c) - ze^1\right). \tag{3.123}$$

As from (3.112) we get

$$e = \frac{2}{z}\Sigma^{-1}(\mu - p^c) \tag{3.124}$$

we find

$$\begin{aligned} p^1 - p^c &= \frac{1}{2}zA^{-1}\left(\frac{2}{z}\Sigma^{-1}(\mu - p^c) - 2e^1\right) \\ &= \frac{1}{2}zA^{-1}(e - 2e^1) \\ &= \frac{1}{2}zA^{-1}(e^2 - e^1). \end{aligned} \tag{3.125}$$

If an investor of group 1 buys the asset, i.e. $e_j^1 < e_j^2$, we find that $p_j^1 > p_j^c > p_j^2$ and we can interpret p_j^1 as the ask price and p_j^2 as the bid price, as this is the price at which an investor of group 2 will sell the asset. If an investor of group 2 buys the asset the relations change. The spread is given by

$$s = |p^1 - p^2| = z|A^{-1}(e^2 - e^1)| > 0. \tag{3.126}$$

The bid and ask prices are symmetric around p^c as can be seen from (3.116), because the market maker has the same market power on both sides of the trade. The spread is increasing in the risk aversion of the investors, as with a higher risk aversion demand reacts less elastic to price changes, because the risk of holding a non-optimal portfolio has a larger influence on the expected utility than trading costs. In the same manner differences in endowment, i.e. the portfolio imbalance, increases the spread. The exact influence on the spread depends on the covariances of the assets, that form A. If two assets are very similar, i.e. their correlation is close to 1, the investor can trade the other asset instead and receive only a small reduction in expected utility (substitution). This indirect competition between assets (if they are not assigned to the same market maker) forces the specialists to compete against each other. This reduces their market power and hence the spread.

Using this result Gehrig and Jackson (1998) derive some further results by analyzing the correlation structure of the assets and the resulting spreads in more detail. It turns out that for assets with a high correlation of the fundamental value, spreads are lower if they are assigned to different market makers. This is the result of the mentioned indirect competition. With a negative correlation the assets become complements and they cannot be traded for each other, which increases the market power of the specialists. Lower spreads result if such assets are assigned to the same market maker.

Comparing the situation with two specialists each being assigned to one asset and a single market maker being responsible for both assets, the endowments of the investors have to be considered. If the two groups of investors are both well engaged in the assets and only want to rebalance their portfolio, it turns out that a single market maker quotes a lower spread for positively correlated assets, while two specialists would give lower spreads for negatively correlated assets.

If the investors however want to change their overall engagement in the asset market without changing the composition of their portfolios the relation exactly reverses. For positively correlated assets two specialists would give lower spreads and a single specialist for negatively correlated assets.

For the optimal allocation of the responsibilities of the specialists it is not only important to take into account the correlation structure of the assets, but also the motives of trading, portfolio rebalancing of well engaged investors or a change in the engagement in the asset market.

It can be preferable to assign a monopoly in market making for several assets to the same market maker in order to prevent indirect competition, while in other situations this indirect competition reduces the spread.

We can summarize these findings by stating that if assets trade as substitutes the spread is lower if the assets are assigned to different market makers, while for assets trading as complements the reverse is true.

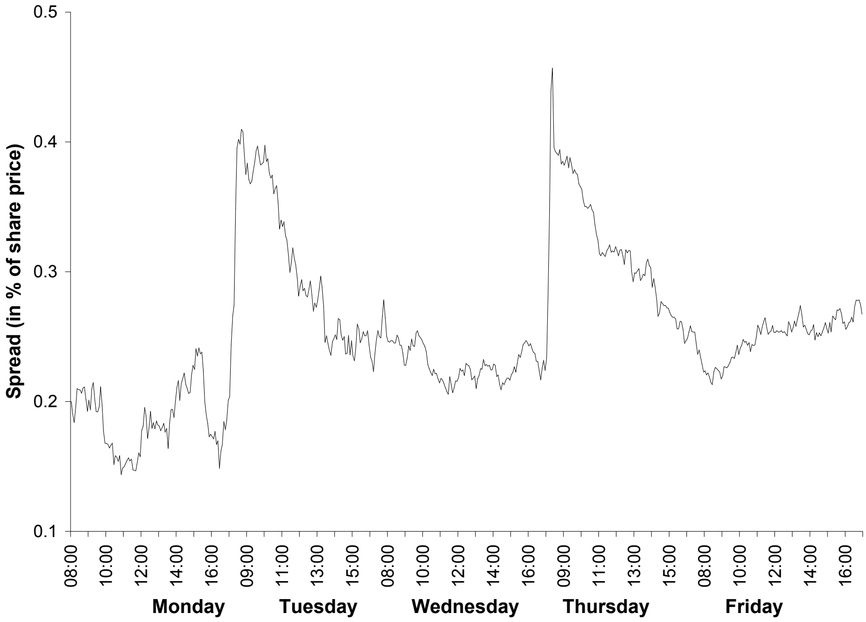
Review questions

1. Why should the spread increase with the risk of the asset?
2. Why will the best quoted spread under competition always be larger than the costs of market making?
3. What is the importance of the time horizon of a monopolistic market maker?
4. Why do uninformed traders always make losses in the presence of informed traders?
5. Why does in the presence of asymmetric information the spread decrease with trading volume?
6. Compare the model on adverse selection with KYLE (1985).
7. Why can market makers quote non-competitive prices when investors can split their orders?
8. Under which conditions are monopolistic market makers preferable to competitive market makers?
9. How can market makers induce clustering of trading?
10. When does allocating two assets to the same market maker result in a lower spread than allocating it to two different market makers?

Application

Consider the same case as in the application of chapter 2 on page 72. After the results of this initial investigation the market regulator sought to consider additional evidence and compiled a graph with the average spread in each five-minute interval, which is reproduced below.

Given the theories on dealer markets detailed above and the details of the case presented in chapter 2, what can be said about the development of the spread and its causes and how does this development fit into the analysis of the case provided previously?


0.5
0.4
0.3
0.2
0.1
Spread (in % of share price)
08:00
10:00
12:00
14:00
16:00
09:00
11:00
13:00
15:00
08:00
10:00
12:00
14:00
16:00
09:00
11:00
13:00
15:00
08:00
10:00
12:00
14:00
16:00
Monday
Tuesday
Wednesday
Thursday
Friday

Chapter 4

Limit Order Markets

In this chapter we will focus on the submission strategies of limit order traders. Our emphasis will be to understand the trade-off between the price a trader receives and the waiting time until the order is filled. The main aspects in this chapter will be

- existence of a spread in competitive markets
- price properties in limit order trading
- dynamic limit order submission strategies
- the informational efficiency of limit order markets

Key readings:

Kalman J. Cohen, Steven F. Maier, Robert A. Schwartz and Daniel K. Whitcomb: Transactions Costs, Order Placement Strategy, and Existence of the Bid-Ask Spread, *Journal of Political Economy*, 89, 287-305, 1981

Christine A. Parlour: Price Dynamics in Limit Order Markets, *Review of Financial Studies*, 11, 789-816, 1998

Thierry Foucault, Odean Kadan and Eugene Kandel: Limit order book as a market for liquidity, *Review of Financial Studies*, 18, *forthcoming*, 2005

In all models presented so far investors were only allowed to submit market orders to a match maker or market maker. The market maker has been the sole provider of immediacy. As has already been mentioned in section 1.4 many different order forms, beside market orders, exist, the most important being the limit order.[1] Limit orders can be viewed as an alternative to market makers for providing immediacy. Like the quotes of a market maker it enables an investor to trade at a fixed price, the limit price, with certainty. For such an investor there is no difference between a limit order and the quotes of a market maker. We can therefore interpret the quotes of a market maker also as a limit order.

Despite these similarities between market makers and limit orders, the concepts of inventory costs and adverse selection costs, cannot easily be applied to investors submitting limit orders. The most important difference between a market maker and a limit order trader is that the market maker is obliged to quote both bid and ask prices, while the limit order trader is free to submit either a limit order, a market order or not to trade at all. He therefore will only submit a limit order if this is the most profitable alternative for him. Despite these differences their similarity for investors submitting market orders suggests that limit orders are competitors to the quotes of market makers and that they will influence each other. In this section it shall be investigated what the consequences are from the introduction of limit orders.

Throughout this section we assume strict price priority not only between quotes of market makers, but also between their quotes and limit orders and within limit orders, i.e. an incoming market order is executed at the best available price, a limit order or the quote of a market maker.[2]

---

[1] The importance of limit orders even exceeds the of market orders. In the second half of 1993 62% of all orders submitted to the SuperDOT system of the NYSE were limit orders, furthermore the market maker has only been involved in 17% of all transactions. See Chakravarty and Holden (1995, pp. 213 ff.).

[2] Not all exchanges apply such strict rules. The NASDAQ goes even a step further to enhance competition between limit orders and market makers in their Rule 2110 and interpretation IM-2110-2 by forcing market makers to give limit orders priority at the same price as the quotes of a market maker.

4.1 Static models of limit order placement

Despite their importance in trading, limit orders have only recently attracted more attention. One of the few early exemptions are Cohen et al. (1981), who investigate the optimal order placement decision of an investor and its implications for the spread.

They do not distinguish between limit orders and quotes of a market maker, as only the decision of an investor is considered. It is assumed that investors are free to submit an order at every time, the order can either be a market or a limit order for a fixed size of trade. The arrival of market orders is assumed to follow a Poisson process with order arrival rate λ, i.e. per period of time the expected number of market orders equals λ. A limit order can be submitted at any price, it cancels if the limit order has to execute a market order. The investor is then free to submit a new order, either a limit or market order.[3]

If the limit order is not executed it remains valid for the whole period of time, e.g. a trading day and cancels automatically afterwards. All limit orders are assumed to be published, i.e. the limit order book is open. Investors see not only the best prices available, but also all other prices that can occur.

We assume now an investor who wants to buy the asset.[4] He can do so either by submitting a market order that executes with certainty, i.e. probability one, at the best available price p_a. The alternative he has now is to submit a limit bid order. When submitting a limit bid order, the execution of this order is not guaranteed as it has to wait for a market sell order to arrive at the market. Before such an order executes there must be no other limit bid order offering a more favorable price. This uncertainty of execution imposes costs on the investor, which earlier have already considered as waiting costs, reducing his expected utility from trading by a limit order. But on the other hand he will be able to trade at a more favorable price.

---

[3] We can introduce a market maker to ensure at least one limit order on both sides of the trades.

[4] The same considerations can be made for an investor selling the asset.

It is obvious from previous sections, the higher he sets his limit bid price p_b^L, the higher the probability of execution, ϕ. If he sets a limit bid price above p_a his order will execute with probability one as investors could make arbitrage profits by selling at p_b^L and buying at the lower price p_a. This negative spread cannot happen as the investor would be able to buy the asset at a lower price by submitting a market order. If p_b^L increases from below to p_a the probability of execution increases, but does not converge to one. This can easily be shown as below.

The number of trades in a time period in which the limit order remains valid, $N(\lambda)$, is finite as long as $\lambda < \infty$. If the probability of an order arriving at the market never equals zero, despite a very high fee charged, the probability that all $N(\lambda)$ trades are at the ask is strictly positive, hence the limit bid order will never execute. It is not necessary that all trades have to occur at the ask. As the change in the ask price upwards due to the many trades at the ask, causes the market maker also to revise his bid quotes upwards as the result of inventory costs or update of believes. This may result in bid quotes exceeding the limit bid price even if it has initially been the best available bid price. In the same manner other investors can place limit orders that offer a higher price. Hence the probability of execution is always smaller than one, even if the limit bid is very close to the ask, while for $p_b^L \geq p_a$ the order executes with certainty. We have found a probability jump in execution at $p = p_a$:

$$\lim_{p_b^L \to p_a} \phi(p_b^L, \lambda) < 1. \tag{4.1}$$

If more orders arrive at the market, i.e. if λ is increased, the probability that the order will not execute is reduced as can easily be seen from the above example. For $\lambda < \lambda'$ we have for all $p_b^L < p_a$:

$$\phi(p_b^L, \lambda) < \phi(p_b^L, \lambda'). \tag{4.2}$$

When $\lambda \to \infty$ the probability of non-execution of the limit order goes to zero, i.e. for all $p_b^L < p_a$ we have

$$\lim_{\lambda \to \infty} \phi(p_b^L, \lambda) = 1, \tag{4.3}$$

the jump in probability vanishes. A continuous trade, i.e. $\lambda = \infty$, requires the investors to revise their portfolios at every instant of time. This is a reasonable assumption as long as trading is costless, if trading imposes costs investors will not revise their portfolios too often as this would exceed the benefits. Hence in the presence of trading costs (other than the fee charged by the market maker), like costs for submitting orders, we find that $\lambda < \infty$ and a probability jump can be found at $p = p_a$.

The investor would never submit a limit bid order such that $p_b^L > p_a$ as has been stated above. The expected utility of submitting such an order would be falling and its maximum would be at $p_b^L = p_a$, i.e. by submitting a market order.

Assuming the utility function to be continuous, the expected utility will make a jump downwards by lowering p_b^L below p_a as a result of the probability jump in execution. Lowering p_b^L further increases the profits from trading at the stated prices, but also the probability of execution reduces as the fee charged increases. Depending on the slope of the utility curves and the sensitivity of investors to changes in the fee the expected utility can either increase or decrease, also a maximum or minimum at some point is possible. In general no shape can be predicted, Cohen et al. (1981) assume that the probability of execution at first reduces only slowly, increasing expected utility, and then it decreases faster, decreasing also expected utility. There exists some point where the expected utility reaches its maximum.

If p_b^L reaches p_b, i.e. the best available bid price, the expected utility jumps again downwards, if submitting a limit order at the same price the order flow has to shared, hence the probability of execution is reduced by its share in the order flow. If the limit order is submitted at prices even lower as p_b at first the limit orders with a better bid price have to be executed before this order is executed. By lowering the bid price further and further, an increasing number of limit orders have to be executed before. Depending on the utility function, the sensitivity of investors to changes in fees and the distribution of other limit bid orders the expected utility may be falling or increasing.

Figure 4.1 presents an example how the expected utility may look like. In the first panel the investor will choose to submit a limit bid order at $p_b^L = p_b'$. The spread is reduced from s to s'. If the expected utility at $p_b^L = p_a$ is higher, as in the second panel, the investor would submit a market buy order.[5]

The jump in expected utility at $p = p_a$ prevents the spread from converging to zero with an increased number of limit orders submitted. There always exists a price below p_a such that the expected utility from submitting a market order is higher, hence the spread will never sink below a certain level. The spread that can be achieved depends on the size of the probability jump at p_a. With a low λ this jump will be larger, confirming the empirical finding that more frequently traded assets have smaller spreads than less actively traded assets. For this result no adverse selection costs are needed as in the previous models to explain this result.

Although this analysis of Cohen et al. (1981) is very intuitive, it faces several problems for the analysis of the behavior of spreads. For determination of the probability of execution dynamic aspects have to be taken into account. In every N trades new limit orders can be submitted, reducing the execution probability of existing limit orders by offering a more favorable price. Also adverse selection costs may become a severe problem if the limit order cannot be withdrawn. The next section addresses some of these aspects.

4.2 Price dynamics with limit order trading

The order submission strategy does not only depend on the unexecuted limit orders in the market, but also on the expectations of future order submissions during the remaining trading rounds. Limit orders submitted at more favorable prices will have priority and therefore reduce the probability of execution. But also limit orders submitted at the other side of the trade or at less favorable prices influence the execution probability as the submitted orders are not market

[5] In all cases it has to be compared with the expected utility from submitting no order at all. If the expected utility from submitting no order is higher, the investor will not submit an order as he has no obligation to do so.

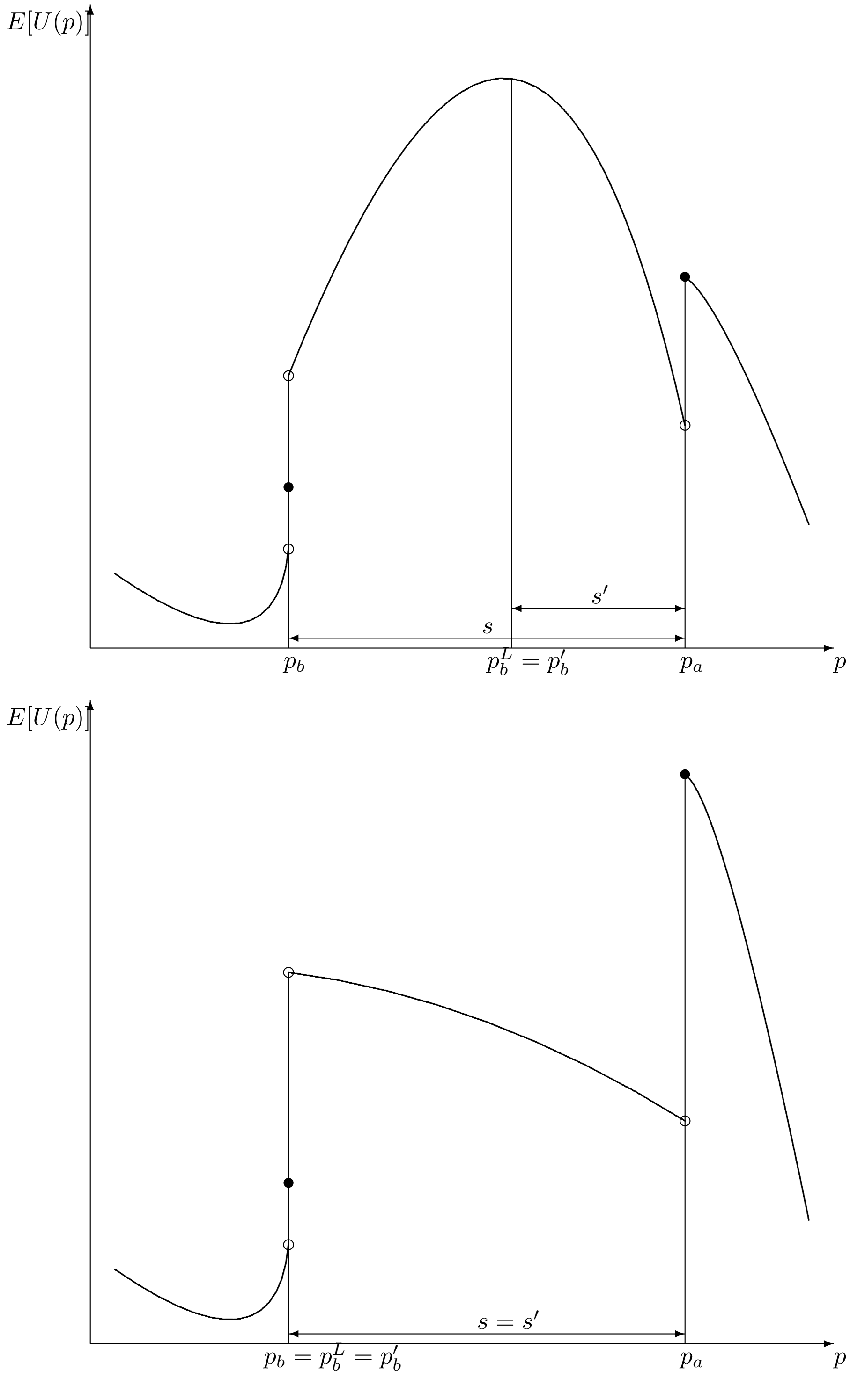


Fig. 4.1: The placement of limit orders

orders that have to be executed against a limit order. Most models of limit order trading presented in the literature thus far are static and therefore do not allow to model these dynamic aspects. The first to model the order submission strategy in a dynamic environment is the recent work by PARLOUR (1998), although the dynamic aspects are incomplete as no equilibrium submission strategies are considered. She models not only the order submission strategy, but also its implications for the movement of prices, on which we concentrate here.

The time structure of the model consists of two periods. Every investor has an initial endowment of the asset which is optimal for him. In the first period he can choose to trade the asset at given bid and ask prices, p_b and p_a, respectively. He can buy the asset and finance this by reducing his consumption in period 1, C_1, by the price he is charged for the asset, or he can sell the asset and consume the amount received. In period 2 the asset is liquidated at the fundamental value p^*, which is known to all market participants, hence there is no informational asymmetry, adverse selection or even uncertainty about this value. All market participants have to reverse their trades of period 1 by reducing their consumption in period 2, C_2, by p^* if they have sold the asset or to increase consumption by the same amount if they have bought the asset before. Hence the asset is a mean to delay or advance consumption between the two time periods. The trade size is assumed to be fixed.

The investors are assumed to be risk neutral with a utility function

$$U(C_1, C_2) = C_1 + \beta C_2, \tag{4.4}$$

where $\beta \geq 0$ is a parameter that denotes the preferences for consumption in periods 1 and 2.[6] For $\beta < 1$ consumption period 1 is preferred and for $\beta > 1$ consumption in period 2. The more β deviates from 1 the more urgent is the need for trading in period 1 as we will see. If $\beta \approx 1$ the investors are very patient with trading. In another interpretation βC_2 is the value the investor assigns to consumption in period 2, as this consumption consists of the asset, it is the value

[6] In terms of microeconomics it is the rate of substitution between consumption in these two periods as $\frac{\partial C_1}{\partial C_2} = \beta$.

the investor assigns to the asset in period 1.

Investors differ only in their value for β, i.e. their time preferences or the value they assign to the asset,[7] they know their own β and they know the distribution F of the β of the other investors.

Trading takes place only in period 1, there is a fixed number of $T \geq 1$ trades in this period. For every trade an investor is chosen randomly to arrive at the market. Which investor, i.e. which β will be selected, depends on the distribution function F.[8] The chosen investor has to decide whether he wants to submit a market buy, limit buy, limit sell, market sell order or not to trade at all.

If an investor submits a limit order he is not free to choose his limit price. The limit price is assumed to be fixed at p_a for a limit ask order and at p_b for a limit bid order, which are the current bid and ask prices. This strong assumption is made to isolate the effect limit orders have on the behavior of prices. The absence of any inventory or adverse selection costs prevent the bid and ask prices from having any dynamics due to the behavior of a market or match maker. But it also restricts the competition of limit orders through the price. The orders submitted in this way are executed with arriving market orders according to time priority, i.e. all orders that were submitted in earlier trading rounds will be executed first. It is obvious that the more limit orders are already unexecuted in the limit order book, the less likely is the execution of a newly submitted limit order. With a finite number of trading rounds the probability of execution will always be lower than one.

An investor submitting a market order receives the amount of p_b in period 1 and has to repurchase the asset in period 2 at p^*. When submitting a limit sell order he receives the amount p_a in period 1 and has to repurchase the asset in period 2 also at p^*, but only if his limit order executes, what happens with probability π^s. When submitting a market buy order he pays p_a in period 1 and

[7] It would not be correct to assign different beliefs on the value of the asset and that its value revealed to the investors in period 2. The value is not a random variable as every investor knows his own β and here therefore receives exactly the amount he expects.

[8] We assume that there are much more investors than trading rounds, so that we do not have to care about an investor to be chosen twice to trade.

receives p^* in period 2. With a limit buy order he pays only p_b in period 1 and also receives p^* in period 2, provided that his limit order executes, what has a probability of π^b. If the limit orders do not execute, he will receive no payment and also has to make no payments, the same situation if he decides not to trade at all. The expected utility is therewith given by

$$E[U(C_1, C_2)] = \begin{cases} p_b - \beta p^* & \text{market sell order} \\ \pi^s(p_a - \beta p^*) & \text{limit sell order} \\ 0 & \text{no order} \\ \pi^b(-p_b + \beta p^*) & \text{limit buy order} \\ -p_b + \beta p^* & \text{market buy order} \end{cases} . \tag{4.5}$$

The optimal strategy is to maximize (4.5). Comparing market and limit sell orders gives the condition for preferring a market sell order if $p_b - \beta p^* > \pi^s(p_a - \beta p^*)$ or

$$\beta_M^s < \frac{p_a}{p^*} - \frac{p_a - p_b}{p^*(1 - \pi^s)} < \frac{p_b}{p^*}. \tag{4.6}$$

If β_M^s turns out to be negative then we have to set $\beta_M^s = 0$. Hence if $\beta < \beta_M^s$ submitting a market sell order is preferred to submitting a limit sell order. Comparing a limit sell order with not trading gives $\pi^s(p_a - \beta p^*) > 0$ or

$$\beta_L^s < \frac{p_a}{p^*}, \tag{4.7}$$

if $\beta < \beta_L^s$ submitting a limit order will be preferred to not trading. The submission of a limit buy order is preferred to not trading if $\pi^b(-p_b + \beta p^*) > 0$ or

$$\beta_L^b > \frac{p_b}{p^*}. \tag{4.8}$$

If $\beta < \beta_L^b$ then not to submit an order is preferred to submitting a limit buy order. For preferring a limit buy to a market buy order we need $\pi^b(-p_b + \beta p^*) > -p_a + \beta p^*$ or

$$\beta_M^b > \frac{p_b}{p^*} + \frac{\pi_a - \pi_b}{pi^*(1 - \pi^b)} > \frac{p_a}{p^*}. \tag{4.9}$$

The final comparison that has to be made is for comparing a limit sell and a limit buy order as they are both preferred to the other alternatives for $\beta_L^b < \beta < \beta_L^s$.

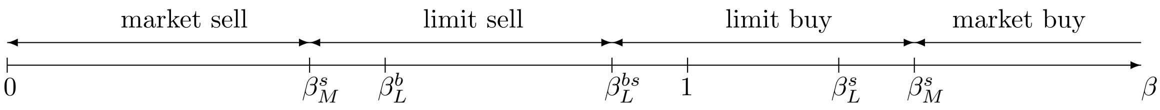


Fig. 4.2: Order submission strategies

The condition that a limit sell order is preferred, $\pi^s(p_a - \beta p^*) > \pi^b(-p_b + \beta p^*)$, solves to

$$\beta_L^{bs} < \frac{p_a}{p^*} - \frac{\pi^b}{\pi^b + \pi^s}\frac{p_a - p_b}{p^*}. \tag{4.10}$$

By the transitivity of preferences all alternatives can now be clearly ordered. Figure 4.2 illustrates the order submission strategy. For extreme values of β the investors prefer to submit market orders to ensure consumption in their preferred period. If β is close enough to one the need to consume in the preferred period is less important than the possible gain from receiving a more favorable price, hence limit orders are submitted. In all cases the alternative not to trade is dominated.

The probability of execution of a limit order, π^b and π^s, plays a central role like in Cohen et al. (1981). Not only the past order submissions, which are known to all investors from the open order book, are of importance, but also the future behavior of the investors. As the distribution of β is known, we could in general use (4.7) - (4.10) to determine the probabilities for the different order types to be submitted in the remaining trading rounds. Unfortunately the important parameters β_M^s and β_M^b depend themselves on these probabilities. Therefore Parlour (1998) proposes an indirect approach to address this problem.

Let b_t^B and b_t^A denote the number of unexecuted limit buy and sell orders in the market. As the execution probability depends on these parameters, also β_M^s and β_M^b will depend on them. Parlour (1998, pp. 809 ff.) gives a detailed proof that

$$\begin{aligned}
\beta_M^s(b_t^B, b_t^A) &\geq \beta_M^s(b_t^B, b_t^A - 1), \\
\beta_M^s(b_t^B, b_t^A) &\geq \beta_M^s(b_t^B + 1, b_t^A), \\
\beta_M^s(b_t^B, b_t^A) &\leq \beta_M^s(b_t^B + 1, b_t^A + 1),
\end{aligned} \tag{4.11}$$

$$\begin{aligned}
\beta_M^b(b_t^B, b_t^A) &\leq \beta_M^b(b_t^B - 1, b_t^A), \\
\beta_M^b(b_t^B, b_t^A) &\leq \beta_M^b(b_t^B, b_t^A + 1), \\
\beta_M^b(b_t^B, b_t^A) &\geq \beta_M^b(b_t^B + 1, b_t^A - 1).
\end{aligned} \tag{4.12}$$

By noting that

$$\begin{aligned}
\frac{\partial \beta_M^s}{\partial \pi^s} &\leq 0, \\
\frac{\partial \beta_M^b}{\partial \pi^b} &\geq 0
\end{aligned} \tag{4.13}$$

as can easily be seen by differentiating (4.6) and (4.9), we can relate the execution probabilities to the number of outstanding limit orders. Surprisingly the execution probabilities depend on both sides of the limit order book and not only on the side at which the order executes.

For observing a transaction at the ask price a market buy order has to be submitted. A market buy order is submitted if the investor chosen has a β such that $\beta > \beta_M^b$, which will be observed with probability $1 - F(\beta_M^B)$. With p_t denoting the price observed in trading round t and o_t the order type at time t we get

$$Pr(p_{t+1} = p_a | o_t) = 1 - F(\beta_M^s(b_{t+1}^B, b_{t+1}^A) | o_t). \tag{4.14}$$

If in trading round t also a market order has been submitted, we find that $b_{t+1}^B = b_t^B$ and $b_{t+1}^A = b_t^A - 1$. With (4.11) we get

$$\beta_M^s(b_{t+1}^B, b_{t+1}^A) = \beta_M^s(b_t^B, b_t^A - 1) \leq \beta_M^s(b_t^B, b_t^A). \tag{4.15}$$

If the order has been a market sell order we have $b_{t+1}^B = b_t^B - 1$ and $b_{t+1}^A = b_t^A$. Again with (4.11) we get

$$\beta_M^s(b_{t+1}^B, b_{t+1}^A) = \beta_M^s(b_t^B - 1, b_t^A) \geq \beta_M^s(b_t^B, b_t^A). \tag{4.16}$$

Obviously the expression $\beta_M^s(b_{t+1}^B, b_{t+1}^A)$ in equation (4.15) is smaller than in equation (4.16). By the monotonicity of the distribution function we have

$$F(\beta_M^s(b_{t+1}^B, b_{t+1}^A) | p_t = p_a) \leq F(\beta_M^s(b_{t+1}^B, b_{t+1}^A) | p_t = p_b). \tag{4.17}$$

Inserting into (4.14) gives the results that

$$Pr(p_{t+1} = p_a | p_t = p_a) \geq Pr(p_{t+1} = p_a | p_t = p_b). \tag{4.18}$$

$$Pr(p_{t+1} = p_b | p_t = p_a) \leq Pr(p_{t+1} = p_b | p_t = p_b), \tag{4.19}$$

It turns out that it is more likely to observe subsequent trades both at the bid or at the ask rather than a change of the side of the trade. We now can calculate the covariance of subsequent price changes. It will remain negative, but as the probability that the price change equals zero is increased owing to (4.18) and (4.19), we get

$$Cov(\Delta p_{t+1}, \Delta p_t) \geq -\frac{s^2}{4}. \tag{4.20}$$

4.3 Dynamic models of limit order submissions

In order to obtain a fully dynamic equilibrium model of limit order submissions we have to take into account that orders submitted later but at a more favorable price are filled first. Hence any costs associated with such prolonged waiting has to be taken into account when submitting an order. Foucault et al. (2005) provide such a model which abstracts from the problem of a fundamental value changing over time or asymmetric information. This allows them to focus purely on the dynamics of the trading process itself.

Suppose risk neutral buyers and sellers arrive at the market alternating in fixed time intervals of length Δt.[9] Each trader submits either a market or a limit order upon arrival in the market. There are two groups of traders, patient and impatient traders. The only difference between these two trader groups is the waiting costs they face when their order is not filled immediately. Patient traders face waiting costs of $c \geq 0$ per time period and impatient traders of $c' > c$. The fraction of patient traders is fixed at θ.

Let each trader i have a valuation of the asset at V_i and the current bid and ask prices prior to the submission of the order are denoted P_b and P_a, respectively.

[9] Changing the requirement that buyers and seller arrive alternating at the market and introducing stochastic time intervals between order submissions does not change the derived results significantly.

With a tick size of d the price a buyer pays is $P_b^i = P_a - z_i d$ and a seller receives $P_a^i = P_b + z_i d$, where z_i denotes the spread expressed in the number of tick sizes after he has submitted his order. The case of $z_i = 0$ corresponds to a market order and $z_i > 0$ to the submission of a limit order, where we assume that limit orders have to reduce the prevailing spread.[10] Thus the expected profits for a buyer and seller will be

$$\begin{aligned} \Pi_i^B(z_i) &= V_i - P_b^i - c_i t(z_i) = (V - P_a) + z_i d - c_i t(z_i), \\ \Pi_i^B(z_i) &= P_a^i - V_i - c_i t(z_i) = (P_b - V_i) + z_i d - c_i t(z_i), \end{aligned} \tag{4.21}$$

where $t(z_i)$ denotes the expected time until the order is filled. As the current bid and ask prices are given when submitting the order the trader seeks to maximize

$$\pi_i(z_i) = z_i d - c_i t(z_i). \tag{4.22}$$

The equilibrium obviously requires the expected waiting time for all possible limit prices to be determined. While it is obvious that $t(0) = 0$ from the fact that market orders are filled immediately, for $z_i > 0$ this will be more difficult to determine.

Suppose that the trader submits a limit order when the spread is s. We are showing later that under certain conditions the trader will only submit limit orders that improve the spread, i.e. $z_i < s$. If the subsequent order to a limit order is a market order the spread reverts back from z_i to s, while for another limit order the spread becomes $z_j < z_i$. This procedure continues until we have $z_k = 1$, in which case the following order has to be a market order. Hence once an order is cleared the spread will return to what it has been before the order was submitted.

Thus the expected waiting time for an order submitted at a spread of z_i will be the time until a market order arrives multiplied the probability of this happening and the expected time it takes to clear any better limit orders submitted later with a lower limit price. Let $p_{z'}(z)$ denote the probability that, given a spread of

[10] Below we will derive conditions for which this assumption is an equilibrium outcome.

z, an order with a spread of z' is submitted. In this case we can write

$$t(z) = p_0(z)\Delta t + \sum_{z'=1}^{z-1} p_{z'}(z)\left(\Delta t + t(z') + t(z)\right), \tag{4.23}$$

and furthermore $t(1) = \Delta t$. The first term denotes the time for a market order to arrive and the second term for better limit orders to be cleared. In the brackets the first term accounts for the time until the next order arrives in the market, second term is the time this subsequent order takes to be filled, after which we face the original problem again. As we know that $\sum_{z'=0}^{z-1} p_{z'}(z) = 1$, we find that

$$t(z) = p_0(z)\Delta t + \sum_{z'=1}^{z-1} p_{z'}(z)t(z') + t(z)\left(1 - p_0(z)\right),$$

which easily solves for

$$t(z) = \frac{\Delta t + \sum_{z'=1}^{z-1} p_{z'}(z)t(z')}{p_0(z)}. \tag{4.24}$$

Let us define the equilibrium spreads such that $s_1 < s_2 < \ldots < s_N$, where $s_1 = \frac{c_k t(s_1)}{d} = \frac{c_k \Delta t}{d}$ as the lowest spread giving a positive profit for a limit order as defined in (4.22) and noting that for s_1 we only observe the submission of limit limit orders by patient traders. Assume finally that we never observe a spread exceeding s_N.

We can indeed show that impatient traders always submit market orders while patient traders submit limit orders if $s \geq s_2$. Thus the probability of observing a market order is $1 - \theta$, hence $p_0(z) = 1 - \theta$. As a patient trader will submit a limit order such that the new spread is just one equilibrium spread smaller than the existing spread, we observe that $p_{s_{i-1}}(s_i) = \theta$ and for any $k > 1$ we have $p_{s_{i-k}}(s_i) = 0$. We can therefore rewrite (4.24) as

$$t(s_i) = \frac{\Delta t + \theta t(s_{i-1})}{1 - \theta}. \tag{4.25}$$

Shifting by one spread we obtain that $t(s_{i+1}) = \frac{\Delta t + \theta t(s_i)}{1-\theta}$, which yields

$$t(s_{i+1}) - t(s_i) = \frac{\theta}{1 - \theta}\left(t(s_i) - t(s_{i-1})\right). \tag{4.26}$$

As $t(s_1) = \Delta t$ and from (4.24) we have $t(s_2) = \Delta t \frac{1+\theta}{1-\theta}$ we get from iterating (4.26) that

$$t(s_i) = \Delta t \left(1 + 2 \sum_{k=1}^{N-1} \left(\frac{\theta}{1-\theta}\right)^k\right). \tag{4.27}$$

Having now established the expected waiting times for each equilibrium spread, it remains to be determined what these equilibrium spreads are. Suppose the spread is currently s_{i+1}, then the limit order submitted by a patient trader is s_i rather then s_{i-1}, thus it is $\pi_i(s_i) \geq \pi_i(s_{i-1})$, which after inserting from (4.22) becomes

$$s_i - s_{i-1} \geq (t(s_i) - t(s_{i-1})) \frac{c}{d}. \tag{4.28}$$

The trader also does not submit an order $s_i - 1$ when the spread is s_i, but prefers s_{i-1}, unless the two are identical, thus $\pi_i(s_{i-1}) \geq \pi_i(s_i - 1)$ or equivalently

$$s_i - s_{i-1} \leq (t(s_i) - t(s_i - 1)) \frac{c}{d} + 1. \tag{4.29}$$

In equilibrium it must thus be either $t(s_i - 1) = t(s_i)$ or $t(s_i - 1) = t(s_{i-1})$ as these are the only two equilibrium spreads. As furthermore we see from (4.27) that $t(s_i) > t(s_{i-1})$, we can rewrite (4.29) as

$$s_i - s_{i-1} < (t(s_i) - t(s_{i-1})) \frac{c}{d} + 1. \tag{4.30}$$

Combining this expression with (4.28) we see that the spread improvement Δs_i between equilibrium spreads is given by

$$\Delta s_i = s_i - s_{i-1} = \left\lceil (t(s_i) - t(s_{i-1})) \frac{c}{d} \right\rceil = \left\lceil 2 \left(\frac{\theta}{1-\theta}\right)^{i-1} \Delta t \frac{c}{d} \right\rceil, \tag{4.31}$$

where the last equality follows from (4.27) and we note that the improvement must be in whole ticks, indicated by the ceiling function $\lceil \cdot \rceil$. Noting that $s_1 = \left\lceil \frac{c}{d} \Delta t \right\rceil$, we obviously obtain that

$$s_i = s_1 + \sum_{k=2}^{n} \Delta s_n. \tag{4.32}$$

It is worth noting at this stage that the spread improvements can be larger than a single tick, i.e. $\Delta s_i > 1$ for sufficiently large waiting costs, low tick sizes or infrequent trading. In all these cases the waiting costs outweigh the less favorable price received. Finally we note that the spread improvement increases the larger the spread is if $\theta > \frac{1}{2}$, i.e. we have many patient traders. BIAIS ET AL. (1995) confirm the existence of limit orders improving the spread by more than a single tick.

The spread improvement can also be used to evaluate the resiliency of the market. Suppose a liquidity shock reaches the market causing a large number of market orders to be submitted. Naturally the spread widens as limit orders get filled and are not replaced. The market will return to the original spread only after the same number of patient limit order traders arrive at the market. The spread prevalent at any point of time until the order book is refilled to its previous state will thus be larger and by how much the spread is increased will depend on the spread improvement. Thus the larger the spread improvement the quicker the market spread will return to its previous level.

We are now in a position to show that all limit orders are spread improving. Suppose that a trader could also submit limit order at the current spread level and let us denote the expected waiting time by t'. Thus our assertion requires that

$$
\begin{aligned}
s_{i-1}d - ct(s_{i-1}) &\geq s_i d - ct'(s_i), \\
(s_i - s_{i-1})d &\leq c\left(t'(s_i) - t(s_{i-1})\right).
\end{aligned}
\tag{4.33}
$$

With the limit order being submitted later we have by time priority that the first order at that price has to be filled first, thus $t'(s_i) > t(s_i)$. After this order being filled there are at least two more orders that have to be filled before the order finally can be filled. The next trade will be on the same side of the trade and only a market order does not increase the order book and thus the waiting time. After that at least another market order is required to fill the order. We have thereby an additional expected waiting time of at least $2(1-\theta)\Delta t$. In case of a limit order being submitted subsequently, at least an additional two more trades are required and the expected waiting time increases by at least $4\left(1-(1-\theta)^2\right)\Delta t$. Hence

$$\begin{aligned} t'(s_i) &\geq t(s_i) + 2(1-\theta)\Delta t + 4\left(1-(1-\theta)^2\right)\Delta t \\ &= t(s_i) + 2\left(1+\theta(2-\theta)\right)\Delta t. \end{aligned} \tag{4.34}$$

Inserting this result into (4.33) we obtain that

$$(s_i - s_{i-1})d - c\left(t(s_i) - t(s_{i-1})\right) \leq 2c\Delta t\left(1+\theta(2-\theta)\right). \tag{4.35}$$

Noting the expression for $\Delta s_i = s_i - s_{i-1}$ from (2.11) we see immediately that

$$d \leq 2c\Delta t\left(1+\theta(2-\theta)\right). \tag{4.36}$$

Thus for a sufficiently small tick size we always observe spread improving limit orders. But we also see that for very active markets, i.e. a small Δt, this is unlikely to be fulfilled and we would observe queuing of limit orders at the inside spread as often observed in real markets. A small tick size allows to avoid queuing at low costs given that the required spread improvement can be very small indeed.

We can now finally investigate the time between two orders being filled, i.e. the time between two trades taking place. Let the first order being filled have a spread of s_i. We would thus require to know the expected number of traders arriving in the market until the next trade takes place. The next trade takes place only once a market order has been submitted, thus an impatient trader arrives at the market. The probability to observe m patient traders is given by

$$Prob(N_i = m) = \theta^{m-1}(1-\theta), \tag{4.37}$$

thus the expected value is

$$E[N_i] = \sum_{m=1}^{i} m\theta^{m-1}(1-\theta) + (i+1)\theta^i = \frac{1-\theta^{i+1}}{1-\theta}, \tag{4.38}$$

where a maximum of i limit orders are possible before the spread is so small that all investors submit a market order, which is captured by the final term. Thus the duration becomes

$$D_i = \frac{1-\theta^{i+1}}{1-\theta}\Delta t. \tag{4.39}$$

We see that the duration increases with the spread and the fraction of patient traders, besides the trading activity.

As we know which spreads are to be found in equilibrium, it would be of relevance to establish in which frequency we find which spread and finally to determine the average spread, average waiting time and duration to analyze the properties of the market further.

Suppose the current spread is s_i, then the following two scenarios can happen. A market order arrives with probability $1-\theta$ and the spread increases to s_{i+1} or with probability θ a new limit order is submitted and the spread reduces to s_{i-1}. If $i=1$ only market orders can be submitted and if $i=N$ the spread does not increase further with a market order but remains constant. The transition matrix becomes

$$\Lambda = \begin{bmatrix} 0 & 1 & 0 & 0 & \cdots & 0 & 0 & 0 \\ \theta & 0 & 1-\theta & 0 & \cdots & 0 & 0 & 0 \\ 0 & \theta & 0 & 1-\theta & \cdots & 0 & 0 & 0 \\ 0 & 0 & \theta & 0 & \cdots & 0 & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & 0 & \cdots & 0 & 1-\theta & 0 \\ 0 & 0 & 0 & 0 & \cdots & \theta & 0 & 1-\theta \\ 0 & 0 & 0 & 0 & \cdots & 0 & \theta & 1-\theta \end{bmatrix}. \tag{4.40}$$

With the probability of observing a spread of s_i denoted u_i and its vector u, a stable equilibrium requires that after a trade the probabilities do not change, thus $\Lambda u = u$. With the usual restriction that $\sum_{i=1}^N u_i = 1$ we obtain that

$$\begin{aligned} u_1 &= \frac{\theta^{N-1}}{\theta^{N-1} + \sum_{j=1}^N \theta^{N-j}(1-\theta)^{j-2}}, \\ u_i &= \frac{\theta^{N-i}(1-\theta)^{i-2}}{\theta^{N-1} + \sum_{j=1}^N \theta^{N-j}(1-\theta)^{j-2}}. \end{aligned} \tag{4.41}$$

We see that $u_i = u_j \left(\frac{\theta}{1-\theta}\right)^{j-i}$ for $i, j \geq 2$ and $u_2 = \frac{u_1}{\theta}$. For $\theta = \frac{1}{2}$ we obtain a uniform distribution of spreads across the range, apart from the smallest equilibrium spread. For $\theta < \frac{1}{2}$, i.e. a market dominated by impatient traders, the

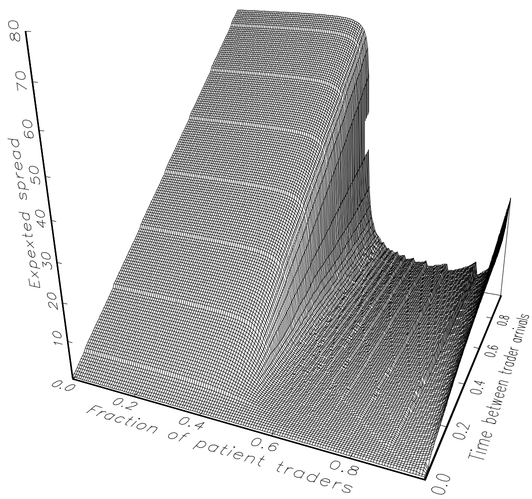


Fig. 4.3: Influence of the trader composition and trade frequency on the spread

probability of high spreads is increasing due to the liquidity demands of these traders while for $\theta > \frac{1}{2}$ the dominance of patient traders leads to less likely large spreads as they liquidity supply outstrips liquidity demand.

Rather than using conditional spreads, durations and waiting times which are difficult to assess empirically , it is straightforward to evaluate the expected spread, duration and waiting times. Figures 4.3-4.11 show how these properties are affected by selected variables. We observe typically two different behaviors for $\theta < \frac{1}{2}$ and $\theta > \frac{1}{2}$, i.e. markets dominated by patient and impatient traders, respectively.

Firstly we observe that for $\theta < \frac{1}{2}$ the average spread is not really affected by the trader composition unless the fraction of impatient traders is very close to $\frac{1}{2}$. The impatient traders begin then to become important and quickly fill the submitted limit orders, thus waiting times are becoming quite short too. With the waiting costs thus being small, the spreads set by limit order traders, i.e. patient traders, can become quite small. As their fraction of the market increases, the

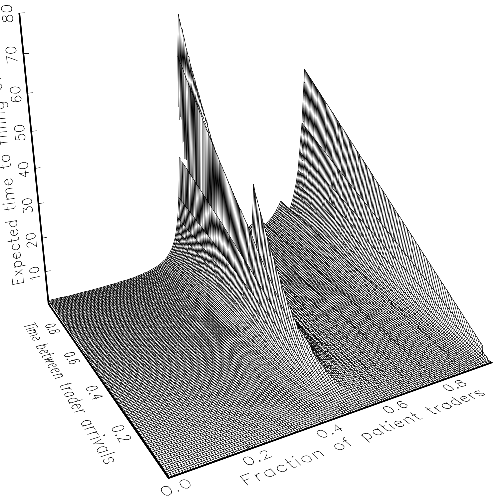


Fig. 4.4: Influence of the trader composition and trade frequency on the waiting time

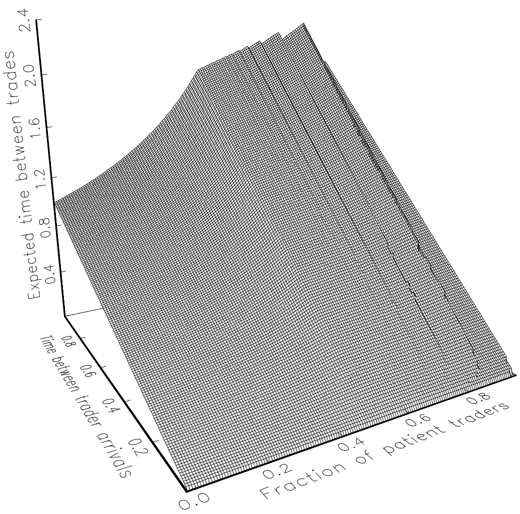


Fig. 4.5: Influence of the trader composition and trade frequency on the duration

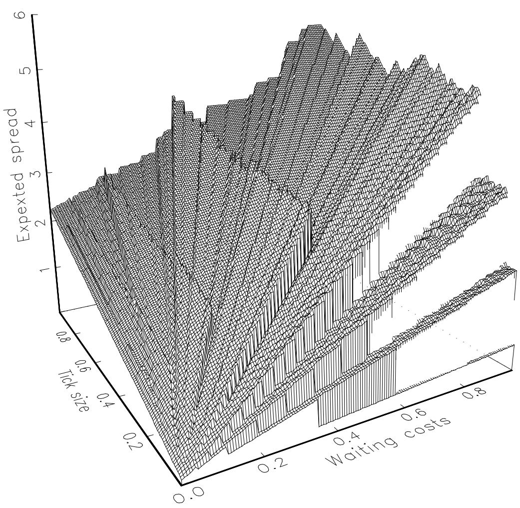


Fig. 4.6: Influence of the tick size and waiting costs on the spread with $\theta = 0.7$

increased competition reduces the spread which is balanced against the reduced number of impatient traders. If however the time interval between trader arrivals increases , the total waiting costs increase. This induces patient traders to seek larger benefits and increasing the spread at which they post consequently. As impatient traders are not much affected by this behavior, orders are still filled within the same time as before. Obviously the time between trades, the duration, is increasing in the time between trader arrivals and with more patient traders as market orders become less common.

For $\theta > \frac{1}{2}$ the picture is however, very different. The increased number of patient traders submitting limit orders reduces the spread and consequently the time to fill orders as more market orders are submitted once the spread is reduced sufficiently. Only once the fraction of patient traders becomes very large, the lack of market orders becomes relevant and the time to fill an order increases. The increased total waiting costs require a larger incentive for patient traders to submit limit orders, thus increasing the expected spread. As these two effects

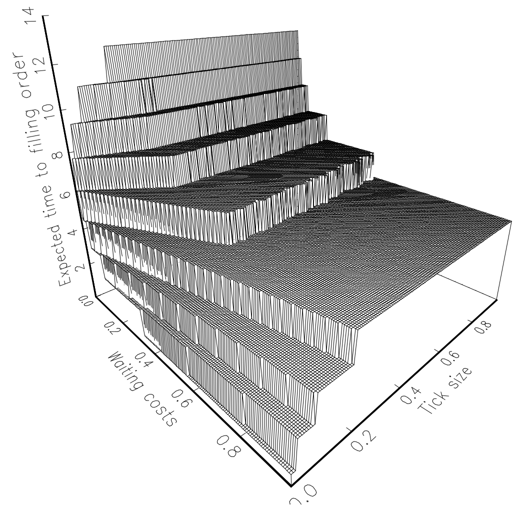


Fig. 4.7: Influence of the tick size and waiting costs on the waiting time with $\theta = 0.7$

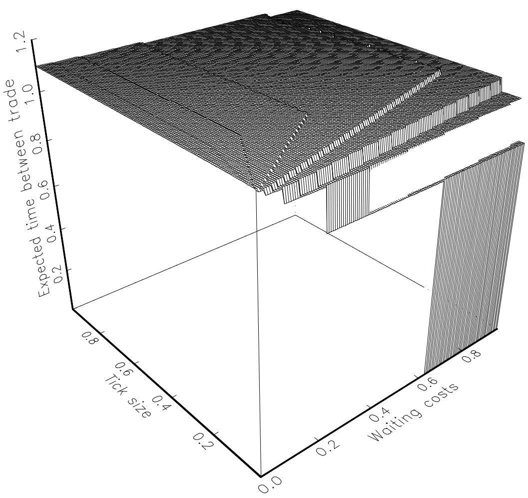


Fig. 4.8: Influence of the tick size and waiting costs on the duration with $\theta = 0.7$

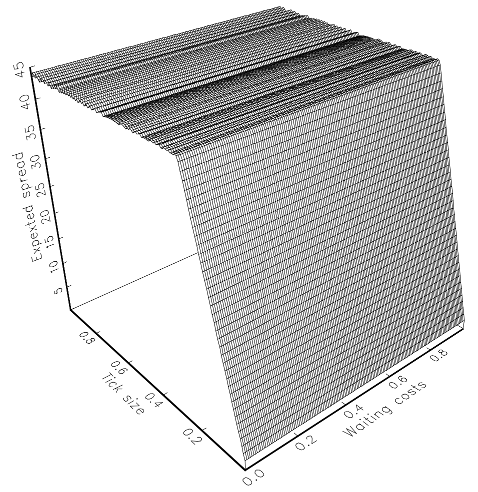


Fig. 4.9: Influence of the tick size and waiting costs on the spread with $\theta = 0.3$

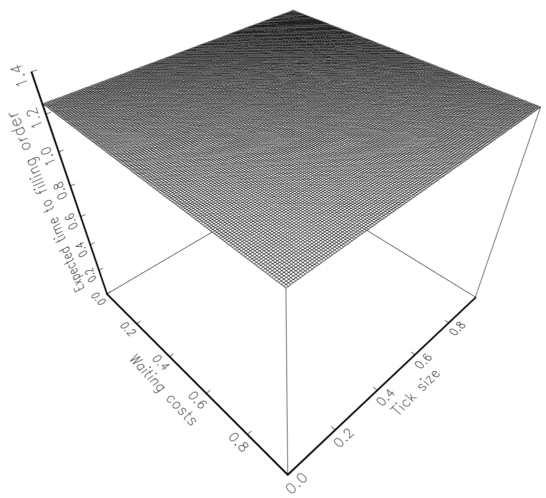


Fig. 4.10: Influence of the tick size and waiting costs on the waiting time with $\theta = 0.3$

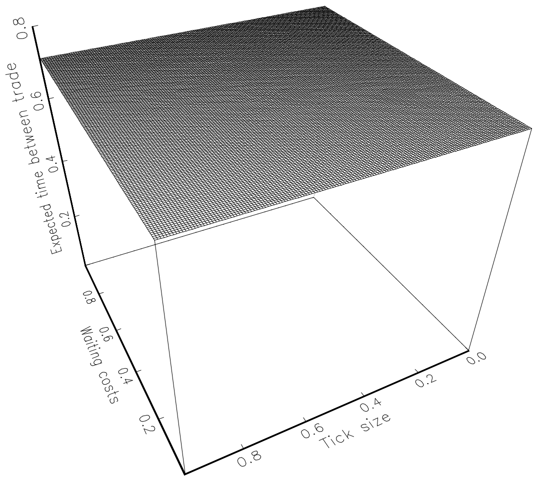


Fig. 4.11: Influence of the tick size and waiting costs on the duration with $\theta = 0.3$

balance, the duration between trades remains largely unaffected.

We can finally observe that the waiting costs and tick size are of limited relevance in market dominated by impatient traders as their considerations are irrelevant with respect to these variables. Only in markets dominated by patient traders do the costs of undercutting the spread become relevant and increase the spread while having no big effect on the duration. The increased spread then makes the waiting time until an order is filled longer.

Empirical work testing this or similar models have not yet been conducted, but the results reported in Biais et al. (1997), Cho and Nelling (2000), Lo et al. (2004) and Hollifield et al. (2004) all find that a limit order submitted at a larger spread takes longer to fill, in accordance with our results and Lo et al. (2004) also report that increased trading reduces this time for which the above model provides weak evidence.

The model presented here provides a first stepping stone towards a fully dynamic model of limit order submissions. Clear drawbacks in this model are the absence of asymmetric information or any other form of uncertainty about the fundamental value. Also the requirement that traders do not queue was excluded in condition (4.30), but we see immediately that for even modestly liquid assets the spread has to be unrealistically small. As we observe such queuing in reality, it is save to assume that this condition is not fulfilled. Rosu (2003) provides an alternative model to explain this observation while Luckock (2001) does not account for endogenous demand.

4.4 Informational efficiency with limit order trading

A formal model of the informational efficiency of prices allowing for limit order trading has been proposed by Brown and Zhang (1997). As their model returns to the batch framework of chapter 2 we do not derive their results explicitly, but give the intuition behind their findings.

If only market orders are allowed to be submitted to the market, informed investors face the risk that due to a large imbalance in the orders of uninformed investors the price is less favorable than expected, they may even make losses. Hence informed investors will trade much less aggressive and prices do not reveal so much information.

If informed investors and risk averse hedgers are allowed to submit limit orders to protect themselves from a too unfavorable shift in prices, both groups will trade more actively. As informed investors can rule out losses from trading, uninformed investors still face the risk of making a loss as they do not have access to the same information. Therefore informed investors will trade much more aggressively than uninformed investors to exploit their informational advantage.

If all informed investors had perfect knowledge of the liquidation value as assumed in section 2, the match maker would observe a bundling of limit orders at this value and would be able to deduct the true value with certainty from observing the limit order flow and by quoting this price the match maker would hinder the informed investors from making profits. As there are no possible profits from being informed, there are no incentives to become informed, hence with the results of Grossman and Stiglitz (1980) no equilibrium would exist.

To avoid such a situation it is assumed that the informed investors do not have perfect knowledge of the fundamental value, but observe only a noisy signal. These different observations of the information results in different limit prices and the match maker cannot directly deduct the fundamental value of the asset.

The more aggressive trades of informed investors, nevertheless, make it easier to deduct the fundamental value of the asset and with the possibility to submit limit orders, prices become more efficient than if only market orders were allowed to be submitted.

Review questions

1. Why does the spread in limit order markets never collapse?
2. What aspects of the limit order price do traders have to consider when making their decision?
3. How do investors decide whether to submit a market or a limit order?
4. What is the impact of a more filled order book on the order submission strategy?
5. What are the properties of transaction prices in limit order markets?
6. Why is the average spread very low when the market consists of many patient traders?
7. Explain the influence of the tick size and the spread and waiting time.
8. Why is the fraction of informed traders so important for the properties of the market?
9. Why do limit order potentially reveal more information than market orders?
10. Why can markets never become fully informationally efficient?

Application

In light of the theories presented in this chapter, re-examine your analysis of the spread in the applications presented in chapters 2 and 3 on pages 72 and 122. Is there an alternative explanation for the observed pattern and can the two explanations be reconciled?

Chapter 5

Empirical investigations of market microstructure models

This chapter will provide a short introduction to empirical methods in market microstructure research. The main topics of this chapter will comprise of

- estimating the spread components
- estimating informed trading
- estimating the impact on daily returns

Key readings:

Hans R. Stoll: Inferring the Components of the Bid-Ask spread: Theory and Empirical Tests, *Journal of Finance*, 44, 115-134, 1989

David Easley, Nicholas M. Kiefer, Maureen O'Hara and John B. Paperman: Liquidity, information and infrequently traded stocks, *Journal of Finance*, 51, 1405-1436, 1996.

Andreas Krause: Inventory effects on daily returns in financial markets, *International Journal of Theoretical and Applied Finance*, 6, 739-765, 2003

We thus far considered mostly theoretical results of the models investigated, besides the application of the results to explain some empirical observations. In this chapter we want to establish how important market microstructure effects are empirically. For this reason we firstly consider ways to estimate the size of inventory and adverse selection costs in dealer markets and informed trading before then continuing to investigate the implications for daily returns.

5.1 Estimating spread components

In the preceding chapters we identified two components of the spread in dealer markets, s: *inventory costs*, c_I, and *adverse selection costs*, c_A. Thus far ignored have been *order processing costs*, c_O, which arise from maintaining the infrastructure of a market maker, e.g. his computer system and staff, which are mostly fixed costs, as well as variable costs arising from conducting a trade, like fees charged by the exchange. We assume that all costs include normal profits as has been determined previously. We then find a final component of the spread, which is mostly neglected in the literature, *excess profits*, c_π.

The spread can therewith be written as the sum of its components:

$$s = c_I + c_A + c_O + c_\pi. \tag{5.1}$$

In this section we will briefly review some contributions on the estimation of these components. No investigation addresses excess profits, for which reason we neglect this spread component in the remainder.[1]

At first we will briefly consider the estimation technique as introduced by Stoll (1989) and then show the results of several empirical investigations.

There exists a large variety of estimation techniques for the components of the spread. Many of these techniques concentrate on the determination of the

[1] Although the literature usually uses the described techniques to estimate the spread components in dealer markets, the liquidity suppliers, i.e. limit order traders, will face similar costs and we can reasonably conclude that the results will also hold in auction markets without a market maker.

adverse selection component and do not further distinguish between inventory and order processing costs. We will here concentrate on the approach developed by STOLL (1989), which explicitly uses the results of market microstructure models as described before.

We will at first investigate the properties of prices if only one of the components is present and finally combine these properties into a single framework used for estimation. The model of STOLL (1989) requires to observe every single trade in an asset to conduct the estimation.[2] We assume for simplicity that the first observed trade has taken place at the bid, results for the first trade at the ask can be derived in the same manner.

At first we investigate the case that only order processing costs are present, i.e. $s = c_O$, which are supposed to be constant. With the fundamental value not changing over time the bid and ask price are the same throughout the sampling period, hence the price change is either zero if the next trade is also at the bid or s if the next trade is at the ask. We assume furthermore that prices are set such that the market clears on average, hence the probability for a trade at the bid and at the ask are both .5. We then have for the change of the transaction price at time t, $\Delta\widetilde{p}_t = \widetilde{p}_t - \widetilde{p}_{t-1}$:

$$\Delta\widetilde{p}_t = \begin{cases} s & \text{with probability } .5 \\ 0 & \text{with probability } .5 \end{cases} . \tag{5.2}$$

When only inventory costs are present, i.e. $s = c_I$, we know from the inventory based models of market making that after a trade at the bid both, the bid and ask prices, decrease to adjust for the larger inventory, while the spread is held constant. The linearity of inventory costs in inventory and the symmetry of the price changes implies the price to fall by $.5s$. To see this, notice that the spread compensates the market maker for his costs to trade at the bid and at the ask. With the symmetry the costs of trading on a single side of the market are $.5s$.

[2] As such data have only been available recently, early contributions, like ROLL (1984), developed models using daily data to estimate the components. A major problem many of them address is to determine the spread as quoted by market makers, as such data also have not readily been available. We do not consider these models here in more detail as it has become a standard nowadays to use data on transaction basis.

For the next transaction these costs do not change relative to the new inventory because of the linearity, hence costs also change by $.5s$ and therewith prices. The probabilities for a trade at the bid and ask, respectively, are no longer equal. The market maker decreases the ask price and hence the costs for investors of trading at the ask, while he increases these costs for a trade at the bid, the probability of receiving such a trade reduces, hence

$$\Delta\widetilde{p}_t = \begin{cases} .5s & \text{with probability } .5 < \gamma < 1 \\ -.5s & \text{with probability } 0 < 1-\gamma < .5 \end{cases} . \tag{5.3}$$

A similar argumentation we can use in the presence of solely adverse selection costs, i.e. $s = c_A$. As we know from information based models of market making prices are set such that they are the expected fundamental value in case a transaction takes place at this side of the market. Assuming adverse selection costs do not to change over time and therewith a constant spread is applied, prices also change by $.5s$. When assuming the market to clear on average we find

$$\Delta\widetilde{p}_t = \begin{cases} .5s & \text{with probability } .5 \\ -.5s & \text{with probability } .5 \end{cases} . \tag{5.4}$$

Let us now define a $\delta \in [0,1]$ such that $(1-\delta)s$ measures the price change if the side of the trade changes in subsequent trades, $-\delta s$ then is the price change if the side does not change. Let further $\gamma \in [0,1]$ measure the probability that the side of the trade changes. Using this notation we can rewrite equations (5.2) - (5.4) as

$$\Delta\widetilde{p}_t = \begin{cases} (1-\delta)s & \text{with probability } \gamma \\ -\delta s & \text{with probability } 1-\gamma \end{cases} . \tag{5.5}$$

The sequence of possible transaction prices together with their transition probabilities are shown in figure 5.1 for convenience, where the superscripts ^{a} and ^{b}, respectively, denote the ask and bid price. Table 5.1 exhibits the parameter constellations for the different spread components as identified above.

The expected price change between two transactions with the initial trade being at the bid is given by

$$E[\Delta\widetilde{p}_t] = (1-\delta)s\gamma + (-\delta s)(1-\gamma) = (\gamma - \delta)s. \tag{5.6}$$

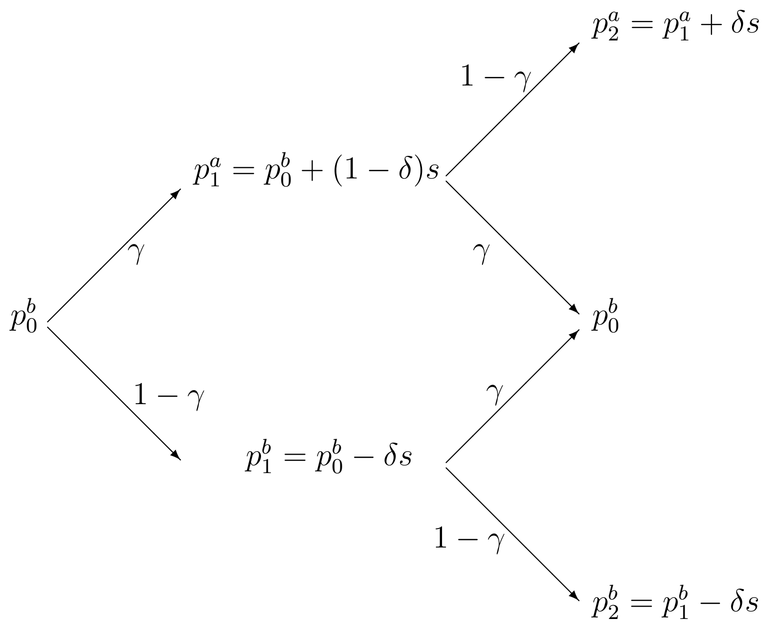


Fig. 5.1: Sequences of transaction prices with the initial price at the bid

Spread component	δ	γ
Order processing costs	0	$\frac{1}{2}$
Adverse selection costs	$\frac{1}{2}$	$\frac{1}{2}$
Inventory costs	$\frac{1}{2}$	$\frac{1}{2} < \gamma < 1$

Tab. 5.1: Parameter constellations for the spread components

Due to the symmetry, the expected price change with the initial trade at the ask is given by $-(\gamma-\delta)s$ as easily can be shown. We can now define the *realized spread* as the gross revenue made by the market maker from a round trip, i.e. a trade at the bid followed by a trade ask and vice versa. In this case the market maker has the same inventory position and belief on the fundamental value of the asset before and after these two trades have taken place. The realized spread therewith can be calculated as the expected price change after a trade at the bid has taken place, minus the expected price change after a trade at the ask has taken place, hence we find the realized spread, s_r, to be

$$s_r = 2(\gamma-\delta)s. \tag{5.7}$$

As we see from table 5.1 the realized spread is zero with only adverse selection costs being present, while in the presence of order processing and inventory costs we find $0 < s_r \leq s$. Hence we can use the realized spread to determine the part of the spread due to adverse selection costs. However, at first we have to determine the parameters γ and δ.

If the probability of the initial trade to take place at the bid is as likely as taking place at the ask, we see that the unconditional expected price change is zero. We then can determine the first order autocovariance of transaction prices with help of the sequences denoted in figure 5.1 to be

$$\begin{aligned}
\rho_T &= Cov[\Delta\widetilde{p}_t, \Delta\widetilde{p}_{t-1}] \\
&= E[\Delta\widetilde{p}_t\Delta\widetilde{p}_{t-1}] \\
&= \frac{1}{2}\left[(1-\delta)s\gamma\delta s(1-\gamma) + (1-\delta)s\gamma(-(1-\delta))s\gamma\right. \\
&\quad \left.+(1-\gamma)(-\delta s)(1-\gamma)(-\delta s) + (1-\gamma)(-\delta s)(1-\delta)s\gamma\right] \\
&\quad +\frac{1}{2}\left[(1-\delta)s\gamma\delta s(1-\gamma) + (1-\delta)s\gamma(-(1-\delta))s\gamma\right. \\
&\quad \left.+(1-\gamma)(-\delta s)(1-\gamma)(-\delta s) + (1-\gamma)(-\delta s)(1-\delta)s\gamma\right] \\
&= \left[\delta^2(1-\gamma)^2 - \gamma^2(1-\delta)^2\right]s^2 \\
&= \left[\delta^2(1-2\gamma) - \gamma^2(1-2\delta)\right]s^2.
\end{aligned} \tag{5.8}$$

In the same manner we can derive the first order serial covariance of the bid price:[3]

$$\begin{aligned}
(5.9) \qquad \rho_Q &= Cov[\Delta\widetilde{p}_t^b, \Delta\widetilde{p}_{t-1}^b] \\
&= E[\Delta\widetilde{p}_t^b \Delta\widetilde{p}_{t-1}^b] \\
&= \frac{1}{2}\left[-\delta s(1-\gamma)(-\delta s)(1-\gamma) + (-\delta s)(1-\gamma)(1-\delta)s\gamma\right] \\
&\quad +\frac{1}{2}\left[\delta s(1-\delta)s(1-\gamma) + \delta s\gamma(-\delta s)\gamma\right] \\
&= \left[\delta^2(1-\gamma)^2 - \delta^2\gamma^2\right] s^2 \\
&= \delta^2(1-2\gamma)s^2.
\end{aligned}$$

We can now estimate these covariances from the data sample and determine the parameters γ and δ accordingly. The nonlinearities implied by (5.8) and (5.9) cause severe biases in small sample sizes as reported by Brooks and Masson (1996). Therefore large sample sizes are needed for conducting this estimation.

Having estimated the parameters γ and δ we can finally determine the spread components. The share of the adverse selection component is given by[4]

$$(5.10) \qquad s_A = \frac{s - s_r}{s} = 1 - 2(\gamma - \delta).$$

Using the realized spread we can determine the inventory cost component with the results from table 5.1 by inserting the relevant parameter values into (5.7) as

$$(5.11) \qquad s_I = 2\gamma - 1.$$

The remainder is the order processing component:

$$(5.12) \qquad s_O = 1 - 2\delta.$$

Besides these estimations several other techniques are used in empirical investigations, in many cases modifications of this estimator. A common approach in most techniques is to decompose the spread into different components and use

[3] For the ask price it can be shown to yield the same result such that it is of no interest which quoted price is observed.

[4] We here directly define the share a cost component has at the spread as the spread component. We could instead also define the component in absolute values without changing the meaning of the results.

Authors	Exchange	Period	s_A	s_I	s_O
Glosten and Harris (1988)	NYSE	Dec 1981	.35	.65	
Stoll (1989)	NASDAQ/NMS	Oct-Dec 1984	.43	.10	.47
George et al. (1991)	NASDAQ	1983-1987	.09	.91	
Affleck-Graves et al. (1994)	NYSE/AMEX	Mar/Apr 1985	.59	.29	.12
Affleck-Graves et al. (1994)	NASDAQ	Apr 1985	.35	.24	.41
Schmidt and Treske (1996)	Frankfurt	Feb-Sep 1995	.22	.55	.23
Porter and Weaver (1996)	NYSE	1990	.02	.98	
Porter and Weaver (1996)	AMEX	1990	.16	.84	
Porter and Weaver (1996)	NASDAQ	1990	-.01	1.01	
Huang and Stoll (1997)	DJIA	1992	.10	.29	.61

Tab. 5.2: Estimates of the spread components

the properties of each component as a reference point to determine the influence of this component on the entire spread. We will present the results of several empirical investigations in the next section.

A large number of empirical investigations are reported in the literature addressing the composition of the spread, some are reported in table 5.2. The evidence shows large differences between the authors, an observation that cannot only be attributed to the investigation of different assets in different periods of time. Several authors, e.g. Huang and Stoll (1997), use different estimation techniques for the same data set and derive results that differ substantially. Neal and Wheatley (1995) compare two widely applied estimation techniques, those of Glosten and Harris (1988) and George et al. (1991) for various stocks and mutual funds and show large differences in the estimates of adverse selection components, with the estimates of Glosten and Harris (1988) being smaller throughout.

We will therefore not interpret the results found in the literature in too much detail. Despite all problems one faces when estimating the spread components, it can be seen that adverse selection components form a considerable part of the spread, estimates for many investigations, including several not included in the table, ranging from about a third to a half of the observed spread. Other investigations, e.g. Hansch et al. (1998) and Snell and Tonks (1999) find significant evidence of inventory costs, although they do not estimate the share of the spread to be attributed to these costs.

In chapters 2.3 and 3.2.1 we showed that a larger share of informed investors increases trading volume and that adverse selection costs increase. We should therefore expect to find a positive relation between trading volume and adverse selection costs. Indeed, Glosten and Harris (1988) report the adverse selection costs and therewith the entire spread, to be increasing in trading volume. A similar result is reported by Lin et al. (1995) and Laux (1993).

We may summarize these empirical results shortly by stating that support is found for all spread components to be relevant in markets and adverse selection costs increase with trading volume. However, determining the effect of each component on the total spread is very difficult and subject to the model applied for estimation. Up to now there exists no generally accepted framework for the estimation of spread components and therewith no reliable estimation of the spread components.

5.2 Estimating informed trading

Apart from evaluating the origins and composition of the spread it is also of interest to determine the amount of informed trading in the market. As usually not only the motivations for a trade are not known, but databases do not even disclose the identity of a trader, it is not possible to make direct inferences. However, Easley et al. (1996) developed a simple model that uses the number of trades at the bid and ask to determine the probability of informed trading (PIN).

Suppose a market consists of two types of traders, informed traders and uninformed traders. Uninformed traders are arriving at the market following a Poisson process, independently for buy and sell orders.[5] Suppose that in any time period, e.g. a trading day, there are ε orders submitted by uninformed investors, for sell as well as buy orders. Informed investors only trade when they

[5] A Poisson process states that the probability of x orders arriving at the market in a given period of time to be $Prob(x \text{ orders arriving}) = \frac{\lambda^x e^{-x}}{x!}$, where λ denotes the expected number of arrivals, i.e. $E[x] = \lambda$.

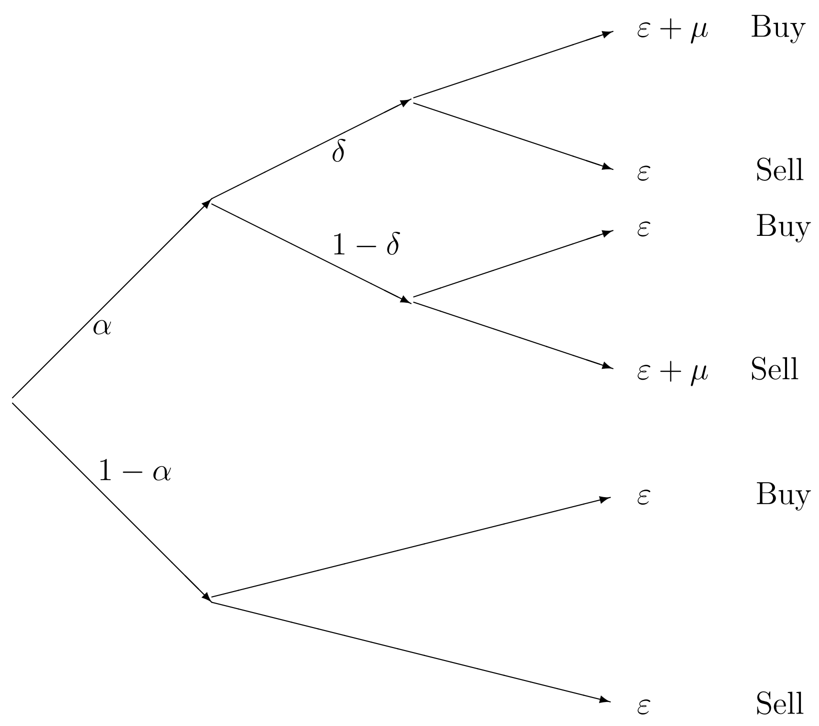


Fig. 5.2: Evolution of the order flow with informed trading

receive new information, which we assume to be happening with probability α. Upon receiving good news they will buy the asset and upon receiving bad news they will sell the asset. Good news arrive with probability δ and the order arrival rate of informed investors in that case is μ. Figure 5.2 illustrates this process.

We easily see that the expected number of orders is given by

$$E[v] = \alpha\left(\delta(\varepsilon+\varepsilon+\mu) + (1-\delta)(\varepsilon+\varepsilon+\mu)\right) + (1-\alpha)(\varepsilon+\varepsilon) = \alpha\mu + 2\varepsilon, \tag{5.13}$$

and the informed trading is given by

$$E[v_I] = \alpha\left(\delta\mu + (1-\delta)\mu\right) = \alpha\mu. \tag{5.14}$$

Hence the fraction of informed trading is given by

$$PI = \frac{\alpha\mu}{\alpha\mu + 2\varepsilon}, \tag{5.15}$$

which is obviously also the probability of informed trading for each individual trade. In an empirical investigation we now simply have to estimate the parameters α, μ and ε. To achieve this we use the assumption that order arrivals follow an independent Poisson process. In the absence of any information the probability of observing exactly B buy and S sell orders[6] is given by

$$Prob_N(B,S) = Prob_N(B)Prob_N(S) = e^{-\varepsilon}\frac{\varepsilon^B}{B!}e^{-\varepsilon}\frac{\varepsilon^S}{S!} = e^{-2\varepsilon}\frac{\varepsilon^{B+S}}{B!S!}. \tag{5.16}$$

In the case of good and bad news we obtain

$$\begin{aligned} Prob_G(B,S) &= Prob_G(B)Prob_G(S) = e^{-(\mu+2\varepsilon)}\frac{(\mu+\varepsilon)^B\varepsilon^S}{B!S!}, \\ Prob_B(B,S) &= Prob_B(B)Prob_B(S) = e^{-(\mu+2\varepsilon)}\frac{(\mu+\varepsilon)^S\varepsilon^B}{B!S!}. \end{aligned} \tag{5.17}$$

Hence the total probability which we can observe is given by

(5.18)

$$Prob(B,S) = (1-\alpha)e^{-2\varepsilon}\frac{\varepsilon^{B+S}}{B!S!} + \alpha\delta e^{-(\mu+2\varepsilon)}\frac{(\mu+\varepsilon)^B\varepsilon^S}{B!S!} + \alpha(1-\delta)e^{-(\mu+2\varepsilon)}\frac{(\mu+\varepsilon)^S\varepsilon^B}{B!S!}.$$

We can now use this probability function to estimate the required parameters. Investigations of trading on the NYSE by Easley et al. (1996), Chung and Li (2003) and Chung et al. (2005) show that the PIN is about 14% with more liquid stocks having a lower and less liquid stocks a higher fraction of informed trading. Similar results are obtained in Hanousek and Podpiera (2002) and Němeček and Hanousek (2002) for the Prague Stock Exchange with a PIN of about 30% and Ma and Yang (2002) find also 14% for Taiwan.

The model as presented here is very restrictive as it assumes firstly that the order arrival rates remain constant over time which is not a realistic assertion as pointed out by Easley et al. (2001). Lei and Wu (2001) and Lei and Wu (2005) allow for time varying order arrival rates as well as different rates for buy and sell orders. This specification leads to slightly higher estimates of the PIN of about 22%. Venter and de Jongh (2002) also extend the model to allow uninformed liquidity traders reacting to the fact of information arrival without changing the outcome significantly.

In order to estimate our model above, trades have to be classified as initiated by a buyer or a seller. Obviously simply looking at the bid and ask prices as a

[6] More precisely we have to evaluate buyer and seller initiated trades, respectively. In simple terms, a buyer (seller) initiated trade is one which was conducted as the result of buy (sell) order being subnitted.

guide is insufficient as a seller might submit a low ask price in order to obtain a quick trade. Ellis et al. (2000) point out that common ways to determine the initiation of trades fail in 20-25% of cases. Grammig and Theissen (2002) point out that in this case the PIN will be biased downwards.

5.3 Properties of daily returns

As this section will show, microstructure elements do not only affect prices between subsequent trades but also affect the properties of daily and to a much lesser degree extend to weekly and even monthly returns. Building on the simple inventory based model of market making, Krause (2003) derives such properties.

Suppose a dealer market with a single competitive market maker in which investors trade without the ability to use limit orders. The fundamental value of the asset is assumed to be common knowledge and between subsequent trading days the log-price evolves as

$$P_t^* = P_{t-1}^* + \varepsilon_t, \tag{5.19}$$

where $\varepsilon_t \sim iidN(0, \sigma^2)$. As derived in chapter 3.1 the prices quoted by the market maker will depend on his inventory position at the trade τ on trading day t, this adjustment is denoted $\eta_{t,\tau}$. The mid-price quote of the market maker will be

$$P_{t,\tau}^M = P_t^* + \eta_{t,\tau}, \tag{5.20}$$

and the bid and ask prices are then given as

$$\begin{aligned} p_{t,\tau}^b &= P_{t,\tau}^M - \frac{1}{2} s_{t,\tau}, \\ p_{t,\tau}^a &= P_{t,\tau}^M + \frac{1}{2} s_{t,\tau}. \end{aligned} \tag{5.21}$$

Using the result from chapter 3.1, denoting the trade size with $Q_{t,\tau}$, measured in nominal value, and an inventory with the value of $I_{t,\tau}$ we obtain from the equations above that

$$\eta_{t,\tau} = z\sigma^2 \left(\frac{1}{2} Q_{t,\tau} - I_{t,\tau} \right) - \frac{1}{2} s_{t,\tau}, \tag{5.22}$$

which becomes after inserting for $s_{t,\tau}$

$$\eta_{t,\tau} = -\frac{1}{2} z\sigma^2 I_{t,\tau}. \tag{5.23}$$

Define $\alpha = z\sigma^2$ as the inventory effect and noting that $I_{t,\tau} = I_{t,\tau-1} - \chi_{t,\tau-1} Q_{t,\tau-1}$, where

$$\chi_{t,\tau} = \begin{cases} +1 & \text{if} \quad \text{the order is a sell order} \\ 0 & \text{if} \quad \text{no order arrives} \\ -1 & \text{if} \quad \text{the order is a buy order} \end{cases}, \tag{5.24}$$

and we easily get that

$$\eta_{t,\tau} = \eta_{t,\tau-1} + \alpha \chi_{t,\tau-1} Q_{t,\tau-1}. \tag{5.25}$$

The final assumption necessary is regarding the behavior of investors. Suppose they are pure liquidity traders and their trading is affected only by the prices charged by the market maker. Their trading depends on the costs imposed on them by the market maker which is defined as $C^a_{t,\tau} = P^a_{t,\tau} - P^*_t$ and $C^b_{t,\tau} = P^*_t - P^b_{t,\tau}$ for trades at the bid and ask, respectively. The probability of trading is then for $x = a, b$ given by

$$\lambda^x_{t,\tau} = \frac{1}{1 + e^{C^x_{t,\tau}}}. \tag{5.26}$$

Any *ex-ante* expectations would imply that on average $C^x_{t,\tau} = \frac{1}{2} s_{t,\tau}$. Let us assume for simplicity that the trade size is fixed and thus the spread is constant and simply denoted by s. When we investigate the daily return we have to focus on the final trade of the day which determines the relevant price. The above model can now be used to derive properties of the moments of the daily price returns. KRAUSE (2003) shows that for daily returns we obtain approximately the following result:

$$\begin{aligned} Var[\Delta P_t] &= 2\alpha + \sigma^2, \\ Cov[\Delta P_t, \Delta P_{t-1}] &= -\alpha, \\ Var[\Delta P_t^2] &= 9.44\alpha^2 + 8\alpha\sigma^2, \\ Cov[\Delta P_t^2, \Delta P_{t-1}^2] &= 2\alpha^2. \end{aligned} \tag{5.27}$$

We firstly note from this result that the volatility of the asset price, $\sqrt{Var[\Delta P_t]}$, is higher than the volatility of the fundamental value, σ. KRAUSE (2003) estimates the price volatility to be about three times the volatility of the fundamental value. Secondly, returns should exhibit a negative first order autocorrelation, a result not necessarily confirmed empirically. Thirdly, returns exhibit heteroscedasticity as observed empirically and finally the volatility is positively correlated over time, also in accordance with empirical evidence.

Thus we can conclude that the model provides some very realistic results, apart from the negative autocorrelation of returns. The introduction of long-lived private information would however easily introduce a positive autocorrelation. Informed traders spreading their trades over several days to optimally exploit their information, would easily cause this change in combination with uninformed traders revising their expectations on the fundamental value.

The model by ROLL (1984) can be seen as a special case which neglects inventory effects by setting $\alpha = 0$ and assuming a constant fundamental value, i.e. $\sigma^2 = 0$. Setting $\lambda = \frac{1}{2}$ by proposing that traders are not elastic to trading costs, we recover the result that $Var[\Delta P_t] = \frac{1}{2}s^2$ and $Cov[\Delta P_t, \Delta P_{t-1}] = -\frac{1}{4}s^2$ and thus $Corr[\Delta P_t, \Delta P_{t-1}] = -\frac{1}{2}$.

Review questions

1. Why do we need transaction data for estimating microstructure models?
2. What can be said about the composition of the spread in stock markets?
3. Why is the adverse selection component of the spread not a measure of informed trading?
4. What is the difference between observing a transaction at the ask and a buyer initiated trade?
5. Why can informed trading not be estimated directly from the trades?
6. What can be said about the amount of informed trading in stock markets?
7. What causes the negative autocorrelation in daily returns?
8. Why is the daily volatility of assets higher than the volatility of the fundamental value?
9. What causes heteroskedasticity in daily returns?
10. How is informed trading likely to affect the properties of daily returns?

Chapter 6

The emergence of different market forms

The aim of this chapter is to show under which conditions different market forms emerge endogenously. The focus will be on the number of market makers that might emerge and the emergence of dealer or limit order markets. The main aspects considered in this chapter will be

- the decision by market participants to act as market makers
- the relevance of search costs in markets
- the decision to become a match maker

Key readings:

Sanford J. Grossman and Merton H. Miller: Liquidity and Market Structure, *Journal of Finance*, 43, 617-637, 1988

A. Yavas: Market Makers versus Match Makers, *Journal of Financial Intermediation*, 2, 33-58, 1992

Chapter 1.3 gave an overview of the different market forms and their advantages and disadvantages from the view of an investor. It was intuitively shown that investors may demand different market forms, depending on their preferences. These different market forms also have to be supplied, i.e. there have to be match makers and market makers. In this section we will derive conditions under which auction and dealer markets arise in an unregulated market.

In both market forms, auction and dealer markets, we find specialized market participants, match makers and market makers. Contrary to investors they do not actively trade in the market, they only react to the demand of investors.

A match maker receives an order[1] from an investor and has to find an offsetting order in the market. When he has found such an order he arranges the trade and determines the price. For these efforts he charges a fee to both investors. At the time investors submit their orders to the match maker neither the price at which the trade is conducted nor the time the trade occurs is known. Investors face the risk that prices may change significantly in the mean time and the trading decisions may not longer be optimal.

A market maker, on the other hand, publishes prices at which he is willing to buy and sell the asset on his own account from any investor. At the time the market maker publishes these prices, he does not know whether the next order arriving will be a buy or a sell order, but any order is immediately executed at the stated prices. Investors know the price they will receive upon submitting an order and also the time of trading. They face no risk after having submitted the order. This risk is now taken by the market maker, he does not know when another, offsetting order will arrive at the market and at which price he will be able to trade. This risk imposes costs on the market maker for which he has to be compensated. Typically a market maker is not allowed to charge any fee to investors, therefore he has to incorporate his costs into the prices he charges. He will quote a higher price for selling the asset to an investor (*ask price*) and a lower price for buying the asset from an investor (*bid price*). As has been shown in

[1] Unless otherwise stated only market orders will be considered throughout this chapter.

sections 3.1 and 3.2 the ask price will always exceed the bid price, the difference between these two prices is called the *spread*.

The possibility for investors to trade at a fixed price without any delay is called *immediacy*. The willingness to trade with a market maker is the *demand* for immediacy and the willingness to act as market maker is the *supply* of immediacy.[2]

The following subsection will provide an equilibrium of immediacy before in 6.2 a model is presented to determine the equilibrium market form.

6.1 Demand and Supply of Immediacy

Grossman and Miller (1988) assume a market with N participants, a single risky asset[3] and a riskless asset with a zero rate of return.[4] Of the N market participants $n < N$ face an exogenously given liquidity event, i.e. their holding of the risky asset is no longer optimal and they have to trade in order to rebalance their portfolio by trading the risky against the riskless asset. What causes this liquidity event is of no importance here, we could easily assume the liquidity event to be caused by an exogenous need for changing the investment position, but also informational asymmetries to induce trading. The other $N - n$ market participants face no need for changing their portfolios. The n market participants facing a liquidity event are called *investors*.

We assume that there exist three time periods. At the beginning of period 1 the liquidity event occurs and at the end of this period a first round of trading takes place. At the beginning of period 2 new information is released on the fundamental value of the asset and afterwards a second round of trading occurs. In period 3 the fundamental value is fully revealed and the portfolios are liquidated at the fundamental value without further trading.

[2] Another form of supplying immediacy is by submitting limit orders. As the effect of this form of immediacy is not important in this setting, limit orders are not considered here. A formal treatment of limit orders is given in chapter 4.

[3] This single asset can also be interpreted as the optimal risky portfolio as determined by portfolio theory.

[4] The assumption of a zero return can be justified by using the return of the riskless asset as a normalization, i.e. all returns have to be interpreted as excess returns.

The price formation has to be such that after the second trading round all market participants hold their optimal portfolio, i.e. all market participants have adjusted their portfolios according to the liquidity event and the new information released in period 2.

Investors can offset their liquidity event either by trading in period 1, in period 2 or in both periods. Assume without loss of generality that they trade either in period 1 or in period 2, but not in both periods. When an investor decides to trade depends on his exogenously given preferences and the costs in each period. Assume that n_1 investors trade in period 1 and n_2 investors in period 2, where $n_1 + n_2 = n$. Let the number of assets that have to be traded as a result of the liquidity event by investor i be denoted x_0^i, where with a positive amount the investor buys additional units of the asset and with a negative amount he sells the asset in exchange for the riskless asset. The sum of all orders has to equal zero because the overall holding of the asset remains constant:

$$\sum_{i=1}^{n} x_0^i = 0. \tag{6.1}$$

But if some investors decide to trade in period 1 and others in period 2, the orders arriving in each period will not necessarily be balanced. In order to clear the market there have to be some market participants facing no liquidity event that are willing to take offsetting positions enabling investors to trade. These market participants do not have to trade as they face no liquidity event, but they can take the offsetting positions in period 1 and can these positions offset by trading again with the remaining investors in period 2. Those market participants that take voluntary offsetting position in period 1 are called *market makers*. Let there be $M \leq N - n$ market makers.

With X_0^1 denoting the order imbalance in period 1 we define

$$X_0^1 \equiv \sum_{i=1}^{n_1} x_0^i. \tag{6.2}$$

Market makers have only to be concerned with the net order imbalance as all other orders can immediately be offset in the same period by matching the orders directly and only the excess is routed to a market maker. Hence they are only concerned about trades in one direction, either they buy or they sell the asset. As market makers have to quote fixed prices at which they are willing to trade, but only trades at one side occur, it has not to be distinguished between bid and ask prices. The price of the asset in period t, P_t, has to be interpreted as a bid or ask price, depending on the net order flow.

We assume that all market participants maximize their expected utility of final, i.e. period 3, wealth. Denote the wealth of investor i in period t by W_t^i, the amount of the riskless asset he holds by B_t^i, the units of risky assets held by q_t^i. Let further W_0^i be the initial wealth of an investor before the liquidity event occurs. We therewith can determine the final wealth as follows:

$$\begin{aligned} W_1^i &= W_0^i + P_1 x_0^i \\ &= B_1^i + P_1 q_1^i, \end{aligned} \tag{6.3}$$

$$\begin{aligned} W_2^i &= B_1^i + P_2 q_1^i \\ &= B_2^i + P_2 q_2^i, \end{aligned} \tag{6.4}$$

$$W_3^i = B_2^i + P_3 q_2^i. \tag{6.5}$$

Inserting $B_2^i = B_1^i + P_2(q_1^i - q_2^i)$ and $B_1^i = W_0^i - P_1(q_1^i - x_0^i)$ from manipulating (6.3) and (6.4), (6.5) becomes

$$\begin{aligned} W_3^i &= B_1^i + P_2(q_1^i - q_2^i) + P_3 q_2^i \\ &= W_0^i + P_1(x_0^i - q_1^i) + P_2(q_1^i - q_2^i) + P_3 q_2^i \\ &= W_0^i + P_1(x_0^i - q_1^i) + P_2(q_1^i - x_0^i) - P_2(q_1^i - x_0^i - q_1^i + q_2^i) \\ &\quad + P_3(q_2^i - x_0^i + x_0^i) \\ &= W_0^i + (P_2 - P_1)(q_1^i - x_0^i) + (P_3 - P_2)(q_2^i - x_0^i) + P_3 x_0^i. \end{aligned} \tag{6.6}$$

Define the excess demand over the liquidity event in period $t = 1, 2$ by investor i as

$$\xi_t^i = q_t^i - x_0^i. \tag{6.7}$$

Replacing (6.7) in (6.6) gives

$$W_3^i = W_0^i + (P_2 - P_1)\xi_1^i + (P_3 - P_2)\xi_2^i + P_3 x_0^i. \tag{6.8}$$

By using the approximation of the expected utility we get with risk aversion z^i:

$$E[U(W_3^i)] = U\left(E[W_3^i] - \frac{1}{2}z^i Var[W_3^i]\right). \tag{6.9}$$

By using all available information, Ω_2, at time $t = 2$ we get

$$\begin{aligned} E[U(W_3^i)|\Omega_2] \quad = \quad U\Bigg(& W_0^i + (P_2 - P_1)(q_1^i - x_0^i) \\ & +(E[P_3|\Omega_2] - P_2)(q_2^i - x_0^i) + E[P_3|\Omega_2]x_0^i \\ & -\frac{1}{2}z^i(\xi_2^i + x_0^i)^2 Var[P_3|\Omega_2]\Bigg). \end{aligned} \tag{6.10}$$

Maximizing (6.10) for the optimal excess demand in period 2, ξ_2^i, gives the following first order condition:

$$\left(E[P_3|\Omega_2] - P_2 - z^i(\xi_2^i + x_0^i)Var[P_3|\Omega_2]\right)U'(.) = 0. \tag{6.11}$$

With $U' > 0$ and $U'' < 0$ as required for risk averters the second order condition can easily be shown to be fulfilled. Solving for ξ_2^i gives the optimal excess demand in period 2 as

$$\xi_2^i = \frac{E[P_3|\Omega_2] - P_2}{z^i Var[P_3|\Omega_2]} - x_0^i. \tag{6.12}$$

As only for investors we find that $x_0^i \neq 0$, for all other market participants the excess demand simplifies to

$$\xi_2^i = \frac{E[P_3|\Omega_2] - P_2}{z^i Var[P_3|\Omega_2]}, \tag{6.13}$$

where only market makers trade, hence this demand is only valid for market makers. For market participants that are neither investors nor market makers we find $\xi_2^i = 0$. Assuming all market participants to have the same risk aversion, i.e. $z^i = z$ for all $i = 1, \ldots, N$,[5] we can aggregate (6.12) and (6.13) and get with

[5] Assuming equal risk aversion does not change the arguments to be derived below. We could easily proceed with different degrees of risk aversion at the cost of additional notation.

(6.1):[6]

$$\xi_2^I = \sum_{i=1}^{n} \xi_2^i = n\frac{E[P_3|\Omega_2] - P_2}{zVar[P_3|\Omega_2]}, \tag{6.14}$$

$$\xi_2^M = \sum_{i=1}^{M} \xi_2^i = M\frac{E[P_3|\Omega_2] - P_2}{zVar[P_3|\Omega_2]}. \tag{6.15}$$

Market clearing in period 2 requires

$$0 = \xi_2^I + \xi_2^M = (n+M)\frac{E[P_3|\Omega_2] - P_2}{zVar[P_3|\Omega_2]}, \tag{6.16}$$

which implies

$$E[P_3|\Omega_2] = P_2 \tag{6.17}$$

unless $M = n = 0$.

By inserting (6.17) into (6.12) and (6.13) we receive

$$\xi_2^i = -x_0^i \qquad \text{for investors,} \tag{6.18}$$

$$\xi_2^i = 0 \qquad \text{for market makers.} \tag{6.19}$$

With using (6.17) - (6.19) it is now possible to determine the optimal excess demand in period 1 by returning to (6.9) and evaluationg this expression using all available information from period 1, Ω_1:

$$\begin{aligned} E[U(W_3^i)|\Omega_1] &= U\Big(W_0^i + (E[P_2|\Omega_1] - P_1)\xi_1^i \\ &\quad +(E[P_3|\Omega_1] - E[P_2|\Omega_1])\xi_2^i + E[P_2|\Omega_1]x_0^i \\ &\quad -\frac{1}{2}z^i(\xi_1^i + x_0^i)^2 Var[E[P_3|\Omega_2]|\Omega_1]\Big) \\ &= U\Big(W_0^i + (E[P_3|\Omega_1] - P_1)\xi_1^i + E[P_3|\Omega_1]x_0^i \\ &\quad -\frac{1}{2}z^i(\xi_1^i + x_0^i)^2 Var[E[P_3|\Omega_2]|\Omega_1]\Big). \end{aligned} \tag{6.20}$$

The first order condition for a maximum is

$$\left(E[P_3|\Omega_1] - P_1 - z^i(\xi_1^i + x_0^i)Var[E[P_3|\Omega_2]\Omega_1]\right)U'(.) = 0. \tag{6.21}$$

[6] It is necessary to aggregate over all investors in period 2 and not only over those trading in period 2, because the order imbalance of the investors having traded in period 1 has to be offset in period 2.

The second order condition can easily be shown to be fulfilled. Solving for ξ_1^i gives the optimal excess demand in period 1:

$$\xi_1^i = \frac{E[P_3|\Omega_1] - P_1}{z^i Var[E[P_3|\Omega_2]|\Omega_1]} - x_0^i. \tag{6.22}$$

For market makers this again reduces to

$$\xi_1^i = \frac{E[P_3|\Omega_1] - P_1}{z^i Var[E[P_3|\Omega_2]|\Omega_1]}. \tag{6.23}$$

Aggregating over all investors and market makers we get with (6.2) and the assumption of equal risk aversion for all market participants:

$$\xi_1^I = n_1 \frac{E[P_3|\Omega_1] - P_1}{z Var[E[P_3|\Omega_2]|\Omega_1]} - X_0^1, \tag{6.24}$$

$$\xi_1^M = \frac{E[P_3|\Omega_1] - P_1}{z Var[E[P_3|\Omega_2]|\Omega_1]}. \tag{6.25}$$

Market clearing in period 1 requires

$$0 = \xi_1^I + \xi_1^M = (n_1 + M)\frac{E[P_3|\Omega_1] - P_1}{z^i Var[E[P_3|\Omega_2]|\Omega_1]} - X_0^1. \tag{6.26}$$

The rate of return of the market makers from their activity is given by

$$r = \frac{P_2 - P_1}{P_1} = \frac{E[P_3|\Omega_2] - P_1}{P_1}. \tag{6.27}$$

With (6.26) the expected value and variance of this return is given by

$$Var[r|\Omega_1] = \frac{1}{P_1^2} Var[E[P_3|\Omega_2]|\Omega_1], \tag{6.28}$$

$$\begin{aligned} E[r|\Omega_1] &= \frac{E[E[P_3|\Omega_2]|\Omega_1] - P_1}{P_1} = \frac{E[P_3|\Omega_1] - P_1}{P_1} \\ &= \frac{X_0^1 z}{P_1(n_1 + M)} Var[E[P_3|\Omega_2]|\Omega_1] \\ &= \frac{P_1 X_0^1 z}{n_1 + M} Var[r|\Omega_1]. \end{aligned} \tag{6.29}$$

To derive the equilibrium number of market makers, i.e. the supply of immediacy, suppose that there exists a fixed cost of C for becoming a market maker, e.g. costs of the back office. The expected utility from not participating in the market for all market participants not facing a liquidity event is $E[U(W_0^i)|\Omega_1]$. The profits from acting as market maker are $(P_2 - P_1)\xi_1^i$, hence the expected utility is $E[U(W_0^i - C + (P_2 - P_1)\xi_1^i)|\Omega_1]$.

In equilibrium the expected utility from acting as market maker and not participating in the market have to be equal:

$$E[U(W_0^i)|\Omega_1] = E[U(W_0^i - C + (P_2 - P_1)\xi_1^i)|\Omega_1]. \tag{6.30}$$

As the risk from holding the initial portfolio is equal in both cases it can be neglected in the further analysis and only the additional risk that arises from acting as market maker has to be considered.

Inserting (6.23) and (6.26) we get from (6.30):

$$\begin{aligned}
E[U(W_0^i)|\Omega_1] &= E\left[U\left(W_0^i - C + (P_2 - P_1)\frac{E[P_3|\Omega_1] - P_1}{zVar[E[P_3|\Omega_2]|\Omega_1]}\right)\middle|\,\Omega_1\right] \\
&= E\left[U\left(W_0^i - C + (P_2 - P_1)\frac{X_0^1}{n_1 + M}\right)\middle|\,\Omega_1\right] \\
&= U\left(W_0^i - C + (E[P_2|\Omega_1] - P_1)\frac{X_0^1}{n_1 + M} - \frac{1}{2}z\left(\frac{X_0^1}{n_1 + M}\right)^2 Var[P_2|\Omega_1]\right).
\end{aligned}$$

By comparing coefficients we get

$$\begin{aligned}
C &= (E[P_2|\Omega_1] - P_1)\frac{X_0^1}{n_1 + M} - \frac{1}{2}z\left(\frac{X_0^1}{n_1 + M}\right)^2 Var[P_2|\Omega_1] \\
&= (E[E[P_3|\Omega_2]|\Omega_1] - P_1)\frac{X_0^1}{n_1 + M} \\
&\quad - \frac{1}{2}z\left(\frac{X_0^1}{n_1 + M}\right)^2 Var[P_2|\Omega_1] \\
&= (E[P_3|\Omega_1] - P_1)\frac{X_0^1}{n_1 + M} - \frac{1}{2}z\left(\frac{X_0^1}{n_1 + M}\right)^2 Var[P_2|\Omega_1] \\
&= \frac{P_1 X_0^1}{n_1 + M}E[r|\Omega_1] - \frac{1}{2}z\left(\frac{X_0^1}{n_1 + M}\right)^2 Var[P_2|\Omega_1] \\
&= z\left(\frac{X_0^1}{n_1 + M}\right)^2 P_1^2 Var[r|\Omega_1] - \frac{1}{2}z\left(\frac{X_0^1}{n_1 + M}\right)^2 Var[P_2|\Omega_1] \\
&= z\left(\frac{X_0^1}{n_1 + M}\right)^2 Var[P_2|\Omega_1] - \frac{1}{2}z\left(\frac{X_0^1}{n_1 + M}\right)^2 Var[P_2|\Omega_1] \\
&= \frac{1}{2}z\left(\frac{X_0^1}{n_1 + M}\right)^2 Var[P_2|\Omega_1].
\end{aligned} \tag{6.31}$$

Solving for M as the number of market makers gives the optimal supply of immediacy:

$$M = |X_0^1|\sqrt{\frac{zVar[P_2|\Omega_1]}{2C}} - n_1. \tag{6.32}$$

The number of market makers is larger the lower the costs of market making, the higher the order imbalance, the higher the risk aversion and uncertainty about the price in period 2 and the lower the number of investors trading in the first period. For a high number of market makers it would be optimal that only a few investors were trading in the first period having a large order imbalance.[7]

When determining the optimal number of market makers, the restriction $M \leq N - n$ has further to be taken into account. In this framework immediacy is provided because of the possible profits that can be derived from a temporary order imbalance. How large this order imbalance is and how many investors decide to trade in period 1 also depends on the costs charged by market makers for trading in the first period, i.e. the demand for immediacy. The higher these costs, the lower demand will be.

6.2 Determination of the optimal market form

The last section presented a model to determine the number of market makers endogenously. It was assumed that only the order imbalance in periods 1 and 2 are offset by market makers. This implies that the total order flow at first is matched and only those orders that could not be matched in this process are routed to market makers. Implicitly therewith it has been assumed that this matching process is costless and takes place in an instant. In reality, however, this matching process will be costly. Searching for an offsetting order is time intensive and it is not ensured that an offsetting order can be found in a given time period, even if such an order exists. These costs can give rise to the emergence of a specialized market participant, the match maker. He can be used by investors

[7] Grossman and Miller (1988, pp. 626) derive a similar result by assuming that the order imbalance is not known to the market makers in advance. For solving this problem they have to assume an exponential utility function.

to match their orders. As many investors choose this match maker, he can more easily find an offsetting order. Unlike the market maker he does not offset the order by trading on his own account, he only matches two orders of investors. The match maker, as well as the market maker, makes it more easy to execute an order, hence they both facilitate trading and provide liquidity, what has been identified in chapter 1.2 to be one reason for the emergence of markets.

In chapter 6.1 it has been assumed that a market participant not facing a liquidity event only can choose to become a market maker or not to participate in the market. In this section a model will be presented in which a single market participant can choose to become a match maker or market maker.[8] Yavas (1992) provides a model how this market participant chooses between becoming a market maker or a match maker and hence whether an auction or a dealer market emerges.

We only have two groups of market participants, besides the match maker or market maker. One group assigns a value of P_1 and the other of P_2 to the asset. All market participants are risk neutral, implying that maximizing expected utility is equivalent to maximizing expected profits. Every investor knows the value he assigns to the asset, but not the value the other groups assigns. Their value is a random variable whose distribution function $F_i(P_i)$ is known to all market participants.

There is not a fixed buyer and a fixed seller group. If the price used for a transaction is below the value assigned to the asset, it is bought by the investor, otherwise it is sold.[9] If two investors meet they reveal their values of the asset. The joint surplus of a trade, $|P_1 - P_2|$, is assumed to be shared equally, each investor receiving $\frac{1}{2}|P_1 - P_2|$.[10] For the following we assume without loss of generality that $P_1 > P_2$.

[8] The lack of competition between market makers or match makers does not effect the model to be presented here. We can similarly assume that all market participants have to make the same decision, i.e. that we are considering a group of market participants acting competitively.

[9] If the two groups assign the same value to the asset no transaction occurs, but for simplicity we neglect this case without changing results.

[10] Yavas (1992, p. 36) shows that other divisions of the surplus do not affect the main results.

Let further $0 \leq \theta \leq 1$ denote the probability at which two investors of different groups meet. This probability depends on the search intensity of the two investors, A_1 and A_2, i.e. $\theta = \theta(A_1, A_2)$. We assume that

$$\begin{aligned} \frac{\partial \theta(A_1, A_2)}{\partial A_i} &> 0, \\ \frac{\partial^2 \theta(A_1, A_2)}{\partial A_i^2} &< 0 \qquad i = 1, 2. \end{aligned} \tag{6.33}$$

The costs of searching for investor i are denoted $C_i(A_i)$, where

$$\begin{aligned} \frac{\partial C_i(A_i)}{\partial A_i} &> 0 \qquad i = 1, 2, \\ \frac{\partial^2 C_i(A_i)}{\partial A_i^2} &> 0 \qquad i = 1, 2, \\ \frac{\partial C_i(A_i)}{\partial A_{3-i}} &= 0 \qquad i = 1, 2, \\ \frac{\partial C_i(A_i)}{\partial A_i} &> \frac{\partial \theta(A_1, A_2)}{\partial A_i} \qquad i = 1, 2, \\ C_i(0) &= 0. \end{aligned} \tag{6.34}$$

The functions $\theta(A_1, A_2)$ and $C_i(A_i)$ are assumed to be known by all market participants. As for an investor the value his trading partner assigns to the asset is random, we get the expected surplus from a trade in the absence of a market or match maker:

$$S_i(P_i) = \int_0^\infty \frac{1}{2} |P_i - \widetilde{P}_{3-i}| dF_{3-i}(\widetilde{P}_{3-i}). \tag{6.35}$$

The search intensity of the other group of investors is not known to investors of the other group, hence they have to be conjectured, this value is denoted A_{3-i}^0. The expected profits from searching are

$$\pi_i(A_i, P_i, A_{3-i}^0) = \theta(A_i, A_{3-i}^0) \int_0^\infty \frac{1}{2} |P_i - \widetilde{P}_{3-i}| dF_{3-i}(\widetilde{P}_{3-i}) - C_i(A_i). \tag{6.36}$$

Maximizing expected profits by choosing an optimal search intensity gives the following first order condition:

$$(6.37)\quad \begin{aligned} \frac{\partial \pi_i}{\partial A_i} &= \frac{\partial \theta(A_i, A_{3-i}^0)}{\partial A_i} \int_0^\infty \frac{1}{2}|P_i - \widetilde{P}_{3-i}| dF_{3-i}(\widetilde{P}_{3-i}) - \frac{\partial C_i(A_i)}{\partial A_i} \\ &= \frac{\partial \theta(A_i, A_{3-i}^0)}{\partial A_i} S_i(P_i) - \frac{\partial C_i(A_i)}{\partial A_i} \\ &= 0. \end{aligned}$$

The second order condition for a maximum can easily be shown to be fulfilled if $\frac{\partial^2 C_i(A_i)}{\partial A_i^2} > \frac{\partial^2 \theta(A_i, A_{3-i}^0)}{\partial A_i^2} S_i(P_i)$. The term on the left side is positive from the assumption in (6.34), the first term on the right side is negative from (6.33) and the second term positive as can be seen from its definition in (6.35), hence the left side is negative. The second order condition is always fulfilled with the assumptions stated above.

As we assume all functions to be common knowledge, investors can correctly infer the search intensity of the investors of the other group by solving their equation (6.37). Inserting this result, they can then solve (6.37) to find their own optimal search activity.

So far we have not introduced a match maker or market maker, the investors were supposed to search for each other on their own behalf. Without introducing such an intermediaire this search equilibrium is subject to two inefficiencies. The first inefficiency is that investors may not meet, although they would like to trade with each other. The second inefficiency is that the probability that investors meet, depends on the search activities of both groups, hence each group produces positive externalities to the other group. Therefore the search activity in equilibrium will be lower than in the social optimum.

These inefficiencies enable match or market makers to be established in the market. It will first be analyzed how investors behave in such markets and then these two market forms are compared.

We first analyze the case where the intermediaire is a market maker. Like the investors we assume him to be risk neutral. He is also assumed to know the distributions of the fundamental values each group of investors assigns to the asset, but not the valuation itself. Furthermore he knows the cost functions for searching and the probability function of meeting upon searching.

At the beginning of the period a market maker quotes his prices at which he is willing to sell the asset, P_a, and at which he is willing to buy the asset, P_b, where $P_a \geq P_b$.[11] These prices become known to all investors at no costs. They then have to decide whether they want to search for an offsetting order on their own or whether they want to trade directly with the market maker at the stated prices. By searching on their own investors hope to receive a better price than the quotes of the market maker. If the search is not successful, i.e. they find no offsetting order, or upon meeting the price is less favorable than the quotes of the market maker, they can yet trade with the market maker at the same stated prices. But as the search process takes time this trade will take place at the end of the period, hence the surplus achieved has to be discounted by a rate of ρ.[12]

If $P_i > P_a$ the investor will buy the asset from the market maker and if $P_i < P_b$ he will sell it to him. If $P_b > P_i > P_a$ the investor would make a loss from trading with the market maker and hence will not trade with him.

The surplus from trading with the market maker, $P_i - P_a$ and $P_b - P_i$, respectively, is now not divided as the market maker quotes firm prices. In the presence of a market maker the surplus from a direct trade between investors has to be divided as follows: To prevent the investors to refuse the trade and trade with the market maker instead, they have at first to receive the discounted surplus they would get from trading with him afterwards. After both investors have received this compensation for the surplus they could receive from trading with the market maker afterwards, the remaining surplus is divided equally between them. If the surplus from a direct trade with each other cannot give this minimum compensation, it is refused. Therewith they only trade with each other

[11] If $P_a < P_b$ an investor could buy the asset at P_a from the market maker and sell it to him in an instant of time at P_b making a profit, while the market maker would make a loss. To prevent this arbitrage it is required that $P_a \geq P_b$. See chapters 3.1 and 3.2 for more sophisticated models of market making.

[12] This discount factor could also be interpreted as an adoption of utility to the risk that prices have changed unfavorably in the mean time if we assume risk averse market participants.

if

$$P_2 - P_1 \geq \rho\left(\max\{0, P_1 - P_a, P_b - P_1\} + \max\{0, P_2 - P_a, P_b - P_2\}\right). \tag{6.38}$$

Without an intermediaire the reservation prices of the investors have been their valuation of the asset. With the existence of a market maker these reservation prices change. For trading with each other the price of an investor to buy the asset from another has to be reduced by the surplus he could earn from trading with the market maker. Similarly for the investor selling the asset the reservation price is increased by this amount. The reservation prices become

$$\begin{aligned} P_1^r &= P_1 - \rho \max\{0, P_1 - P_a, P_b - P_1\}, \\ P_2^r &= P_2 + \rho \max\{0, P_2 - P_a, P_b - P_2\}. \end{aligned} \tag{6.39}$$

The negotiated price will always be between these two reservation prices. If $P_1^r \geq P_2^r$ we see that

$$P_1 \geq P_1^r \geq P_2^r \geq P_2. \tag{6.40}$$

The prices are allowed to change only in a smaller interval by subsequent trades with the presence of a market maker, hence the price dispersion is reduced.

The additional surplus from trading directly with each other instead with the market maker is $\frac{1}{2}(P_1^r - P_2^r)$ if $P_1^r > P_2^r$, otherwise no trade occurs. There also does not occur a trade if the investors do not meet. They will not meet if one of them decides to trade with the market maker directly instead of searching first, i.e. his search activity is zero ($A_i = 0$). Therefore the probability distribution of the other investors valuation changes to

$$F_i^D(P_i) = \begin{cases} F_i(P_i) & \text{if } A_i > 0 \\ 0 & \text{if } A_i = 0 \end{cases}. \tag{6.41}$$

The expected surplus from a trade with each other now becomes by using (6.39):

$$S_i^D(P_i, P_a, P_b) = \int_0^\infty \frac{1}{2} \max\{0, P_1^r - P_2^r\} dF_{3-i}^D(P_{3-i}) \tag{6.42}$$

As we see from (6.40) and (6.41) that $P_1^r - P_2^r \leq P_1 - P_2$ and $F_i^D(P_i) \leq F_i(P_i)$, it is obvious that the expected surplus from a trade with each other is reduced in the presence of a market maker:

$$S_i(P_i) \geq S_i^D(P_i, P_a, P_b). \tag{6.43}$$

The expected profits are determined by the expected surplus from trading with each other, the costs of searching,[13] and the expected surplus from trading with the market maker:

$$\begin{aligned} &\pi_i(A_i, P_i, A_{3-i}^0, P_a, P_b) \\ &= \int_0^\infty \frac{1}{2}\theta(A_i, A_{3-i}^0) \max\{0, P_1^r - P_2^r\} dF_{3-i}^D(P_{3-i}) \\ &-C_i(A_i) + (1-\theta(A_i, A_{3-i}^0))\rho \max\{0, P_i - P_a, P_p - P_i\}. \end{aligned} \tag{6.44}$$

Maximizing to determine the optimal search activity, A_i, gives the following first order condition:[14]

$$\begin{aligned} \frac{\partial \pi_i}{\partial A_i} &= \frac{\partial \theta(A_i, A_{3-i}^0)}{\partial A_i} S_i^D(P_i, P_a, P_b) - \frac{\partial C_i(A_i)}{\partial A_i} \\ &\quad - \frac{\partial \theta(A_i, A_{3-i}^0)}{\partial A_i} \rho \max\{0, P_i - P_a, P_b - P_i\} \\ &= \frac{\partial \theta(A_i, A_{3-i}^0)}{\partial A_i} \left(S_i^D(P_i, P_a, P_b) - \rho \max\{0, P_i - P_a, P_b - P_i\}\right) \\ &\quad - \frac{\partial C_i(A_i)}{\partial A_i} \\ &= 0. \end{aligned} \tag{6.45}$$

By inspection of (6.43) we see that the term in brackets has to be smaller than $S_i(P_i)$, but still is positive. Let us for notational simplicity rewrite (6.37) and (6.45) as

$$\theta' S_i - C_i' = 0, \tag{6.46}$$

$$\theta_D' S_i^{D'} - C_D' = 0, \tag{6.47}$$

where $S_i^{D'} = S_i^D(P_i, P_a, P_b) - \rho \max\{0, P_i - P_a, P_b - P_i\} \leq S_i$. Solving for S_i and

[13] These costs also have to be beared if no trade occurs as the investors do not meet or find it more profitable to trade with the market maker later, they are sunk costs.

[14] The second order condition can be proofed to be fulfilled in the same manner as in (6.37).

$S_i^{D'}$ we get

$$S_i = \frac{C_i'}{\theta'},$$
$$S_i^{D'} = \frac{C_D'}{\theta_D'}.$$

Hence we have

$$\frac{C_i'}{\theta'} \geq \frac{C_D'}{\theta_D'}. \tag{6.48}$$

From (6.34) we can find a k and a k_D such that $C_i' = k\theta'$ and $C_D' = k_D\theta_D'$, where because of (6.33) and (6.34) k and k_D are strictly increasing in A_i. Inserting these relations into (6.48) we get

$$k \geq k_D. \tag{6.49}$$

Hence the optimal search activity is smaller in the presence of a market maker than without. This is the result of the possibility to trade with the market maker instead of only trading directly with other investors.

The optimal search intensity can be determined using (6.45). Based on these interferences of the optimal search activities the market maker can set his prices optimal to maximize his expected profits. YAVAS (1992, pp. 42 ff.) shows that an overall equilibrium exists and assigns the problems of multiple equilibria that are of no importance for our purpose. It can easily be shown that the expected profits of the investors are higher in the presence of a market maker, hence it is beneficial for investors to have a market maker. A market maker will also make profits from his activity and therefore there will be market participants acting as market maker.

We now turn to the case where the intermediaire has decided to become a match maker. Like in the presence of a market maker investors can either decide to search for an offsetting order by themselves, facing the risk that they do not meet, or use the match maker where we assume that he has the certainty that the order will be executed.[15] If they decide to search and do not find an offsetting

[15] We could also assume that the probability of execution is significantly higher by using the match maker without changing the argument.

order, they can afterwards turn to the match maker for execution and receive the discounted surplus. For his service the match maker charges a fee from both investors whose orders he matches, the fee is a fraction c of the transaction price. The price is determined by the match maker such that the surplus is equally distributed between the investors.

The reservation prices for a direct trade with each other are again the valuations of the asset as no other market participant offers a better price. The match maker only matches the orders and offers no fixed price. The surplus from trading will $P_1 - P_2$, that will be divided equally. When trading with the help of a match maker the surplus however first has to cover the fees that have to be paid to the match maker. If the surplus does not cover these costs no trade will occur. With P denoting the transaction price, we get

$$\begin{aligned} P_1 - P &\geq cP, \\ P - P_2 &\geq cP, \end{aligned} \tag{6.50}$$

which implies

$$\begin{aligned} P &\leq P_1^r = \frac{P_1}{1+c} \leq P_1, \\ P &\geq P_2^r = \frac{P_2}{1-c} \geq P_2, \end{aligned} \tag{6.51}$$

where P_i^r denote the reservation prices. A trade through the match maker will only occur at prices between P_2^r and P_1^r, with $P_1^r > P_2^r$. As this interval is smaller than $[P_2, P_1]$ the price dispersion is reduced like in the presence of a market maker. The surplus from trading with a match maker is given by

$$\begin{aligned} P_1 - P_2 - c(P_1 + P_2) &= (1+c)P_1^r - (1-c)P_2^r \\ &\quad -c(1+c)P_1^r + (1-c)P_2^r \\ &= P_1^r(1+c)(1-c) - P_2^r(1+c)(1-c) \\ &= (1-c^2)(P_1^r - P_2^r) \\ &> 0. \end{aligned} \tag{6.52}$$

If $P_2^r > P_1^r$ the surplus would be negative and no trade would occur. Like in the presence of a market maker investors can also directly use the match maker instead of searching for an offsetting order by themselves, i.e. $A_i = 0$. In this case an investor searching an offsetting order cannot find a counterpart and his probability distribution changes to

$$F_i^M(P_i) = \begin{cases} F_i(P_i) & \text{if } A_i > 0 \\ 0 & \text{if } A_i = 0 \end{cases} . \tag{6.53}$$

The expected surpluses from trading directly with each other and by using the match maker are given by

$$S_i(P_i) = \int_0^\infty \frac{1}{2}|P_i - P_{3-i}|dF_{3-i}^M(P_{3-i}), \tag{6.54}$$

$$S_i^M(P_i, c) = \int_0^\infty \frac{1}{2}\max\{0, |P_i - P_{3-i}| - c(P_1 + P_2)\}dF_{3-i}^M(P_{3-i}). \tag{6.55}$$

Therewith the expected profit of an investor is

$$\begin{aligned} \pi_i^M(P_i, A_i, A_{3-i}^M, c) &= \theta(A_i, A_{3-i}^0)S_i(P_i) \\ &\quad +(1 - \theta(A_i, A_{3-i}^0))\rho S_i^M(p_i, c) - C_i(A_i). \end{aligned} \tag{6.56}$$

Maximizing the expected profits to find the optimal search activity gives the following first order condition:

$$\begin{aligned} \frac{\partial \pi_i}{\partial A_i} &= \frac{\partial \theta(A_i, A_{3-i}^0)}{\partial A_i}S_i(P_i) - \frac{\partial C_i(A_i)}{\partial A_i} \\ &\quad - \frac{\partial \theta(A_i, A_{3-i}^0)}{\partial A_i}\rho S_i^M(P_i, c) \\ &= \frac{\partial \theta(A_i, A_{3-i}^0)}{\partial A_i}(S_i(P_i) - \rho S_i^M(P_i, c)) - \frac{\partial C_i(A_i)}{\partial A_i} \\ &= 0. \end{aligned} \tag{6.57}$$

As from inspection of (6.54) and (6.55) we see that $S_i(P_i) \geq S_i^M(P_i, c) \geq \rho S_i^M(P_i, c) \geq 0$, it can easily be verified that $0 \leq S_i(P_i) - \rho S_i^M(P_i, c) \leq S_i(P_i)$. This is in line with the results in the presence of a market maker, the second order condition is fulfilled and it turns out that the search activities are reduced in the presence of a match maker. The optimal search activity can be derived from (6.57) and given these results that match maker can determine his optimal fee. Yavas (1992, pp. 50 f.) again shows the existence of such an equilibrium. As in the presence of a market maker the expected profits of the investors are higher with the presence of a match maker, hence it is beneficial for investors to have a match maker. A match maker will also make profits from his activity and therefore there will be market participants acting as match maker.

Thus far it has only been shown that a market maker as well as a match maker are beneficial for investors and that both forms will be provided as they make a profit from their activities. Nothing has been said about which form will be preferred. We therefore now turn to the choice a market participant will make, whether to become a market maker or a match maker. The decision is made on the basis of the expected returns one can earn from these two activities. Yavas (1992) provides some conditions for this choice.[16]

A market participant will decide to become a *market maker* if the value he assigns to the asset differs largely from that of the investors, i.e. if his valuation is at the tail of the distribution of the investors. By offering a very favorable price to both investor groups he still makes a sufficient profit from his activity with a high probability as it is very unlikely that two investors will find it more profitable to trade with each other even if they meet. If search is very efficient, i.e. the costs are relatively small, two investors searching are very likely to meet, but they in most cases will not trade with each other as the market maker can offer a more favorable price and even with very efficient search the market maker will make a considerable profit.

However, if the valuation of the asset does not differ much between investors and intermediaire and search is inefficient, i.e. imposes high costs, he will choose to become a *match maker*. As market maker he would in many cases not offer a more favorable price to the investors than if they trade directly with each other, so that he is unlikely to serve an order, hence his expected profits are low. Whereas a match maker charges fees from both sides independent of the price he charges giving him higher profits.

These conditions to become a market maker or a match maker may change over time and therefore he may want to revise his decision. As a market should

[16] The long and tedious proofs are not presented here to limit space, but they are provided in detail by Yavas (1992, pp. 51 ff.).

apply the same rules over a longer period of time to give investors stable environments to trade and enable the price mechanism to work properly, frequent changes of the market forms should not be allowed. When taking into account that a decision binds him over a given period the above criteria have to be made in a dynamic setting, taking into account changes that are expected. The general idea is not affected, it has to be taken the (discounted) profits over the time period he is bound.

The decision in reality often is restricted by rules stating that all assets in a certain market have to be traded with the same rules. For the determination of the optimal market form in this case the costs and benefits have to aggregated. A certain variety can be achieved by defining market segments that can be traded according to different rules. The assets are then traded into one of these segments according to their characteristics.

Review questions

1. Why are some traders willing to act as market makers?
2. What drives the number of market makers?
3. Who of the traders is likely to become a market maker?
4. Why is searching inefficient?
5. What are the types of inefficiency in search?
6. How can intermediaries overcome the inefficiencies in search?
7. Why is the search of traders reduced in the presence of an intermediary?
8. What is the main difference between a match and a market maker?
9. What drives the decision whether an intermediary becomes a market or match maker?
10. How can a stock exchange decide on its optimal market form?

Application

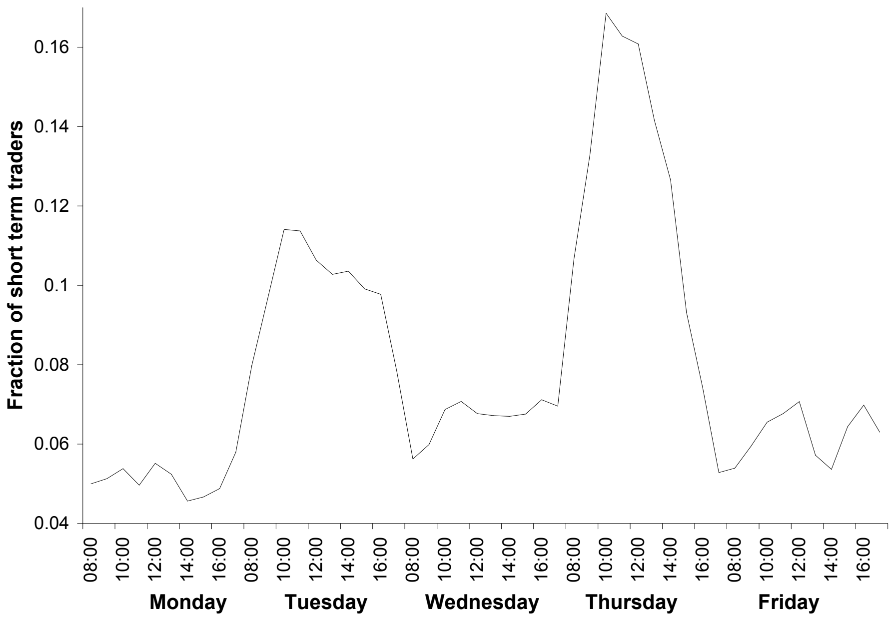
Application 1

On continuation of the application in chapters 2, 3 and 4 the regulator has produced additional information on the trading activity in the market. He has analyzed the fraction of traders who within an hour buy and sell the stock at least once, the graph is shown below.

What can be said about the evolution of the market structure and how does this fit within the wider context of the case as presented and analyzed before?

Application 2

The state of Galaria has set up a stock exchange after embarking on a programme of economic liberalizations in the mid-1990s. Since then the stock market has developed considerably and the number of listed stocks has increased to 534 last month with a total market capitalization of the equivalent of $85bn, equal to nearly 26% of GDP.



The stocks listed on the stock exchange fall mainly into two categories. Firstly large formerly state-owned companies that have been privatized and whose shares are widely held by private investors. Holdings are in most cases small, although a few investors, mostly investment funds, have accumulated significant stakes in these companies. The companies in this category are mostly utilities and a few well established mining and oil exploration companies.

The second group of companies listed can best be described as small and medium-sized technology companies. Their origin lies in the drive of the state to become a leader in computer technology and they are to a large degree engaged in computer and chip manufacturing. Only recently have they become engaged also in the production of LCD and plasma screens, mostly as subcontractors to the leading companies in this sector. Another characteristic of these stocks is that while they are much more actively traded than the former state-owned companies, their volatility is also significantly higher. In contrast to the formerly state-owned companies their prospects are subject to very contrasting views in the investment community.

Thus far the stock exchange has relied on an open out cry mechanism on the floor of the stock exchange. But repeatedly this trading mechanism has been criticized for not being able to cope with the steadily increasing trading volume. In response to these criticisms the government has in cooperation with the stock exchange set up a task force to review the market structure of the stock exchange.

Given the information you have been given above, what would your recommendation to the task force be on the best market structure?

Chapter 7

Asset Pricing with Trading Frictions

This chapter will provide an overview of the influence market microstructure elements have on asset prices. The emphasis will be on the equilibrium rate of return an asset has to generate in the market. We will investigate

- the impact of the bid-ask spread
- the impact of adverse selection
- the impact of imperfect liquidity in the market
- the fraction of the equity premium these trading frictions can explain

Key readings:

Yakov Amihud and Haim Mendelson: Asset Pricing and the Bid-Ask Spread, *Journal of Financial Economics*, 17, 223-249, 1986

Maureen O'Hara (2003): Liquidity and price discovery, *Journal of Finance*, 58, 1335-1354, 2003

W.J. Breen, L. S. Hodrick and R. A. Korajczyk: Predicting equity liquidity, *Management Science*, 48, 470-483, 2003

V. V. Acharya and Lasse Pedersen: Asset pricing with liquidity risk. Mimeo, London Business School, 2004

The models considered thus far focused exclusively on the trading decisions of individual investors and the behavior of special intermediaries present in the market. The costs faced by market makers and other traders, such as inventory costs or adverse selection costs were borne indirectly by investors through the bid-ask spread or the price impact of their trade, i.e. illiquidity. These trading costs of investors will reduce their net returns from holding the asset. Assuming them to be exogenously given and the required rate of return determined by the risk of the asset, the market returns have to be higher to compensate for these costs. Therefore we can reasonably argue that the market returns of assets will depend on microstructure elements as identified above.

The benefits of including market microstructure elements into the determinants of asset returns and thus asset prices is not only restricted to a better explanation of observed prices but can directly be used by companies to evaluate the optimal market form for their listing. Hence it will not only be investors seeking to trade in a market structure that reduces their trading costs but companies will seek to list on exchanges whose microstructure ensures the highest possible price for their shares.

The main sources of transaction costs for investors are the bid-ask spread and the threat that as an uninformed investors he will make losses when trading with an informed investor (adverse selection), a problem similar to that of the market maker. The final element of the trading costs is the illiquidity of the market through which an investor affects the price when trading. In this chapter we will address each of these costs in turn and investigate how these costs affect asset prices.

7.1 Required returns in the presence of bid-ask spreads

Let us consider a market in which $M > 1$ investors can trade $N + 1$ assets. Investors differ only in the time horizons of their investments, denoted T_j for $j = 1, 2, \ldots, M$. The assets differ in the spread that is exogenously imposed on investors, s_i for $i = 1, 2, \ldots, N$. Furthermore there exists an asset for which no spread is charged, i.e. $s_0 = 0$.

To avoid any effect arising from the uncertainty of prices in the future, Amihud and Mendelson (1986a) assume that the price of all assets remains constant for all time horizons of investors with the ask price at P_i and the bid price at $P_i(1 - s_i)$ for all $i = 0, 1, 2, \ldots, N$. With each asset paying a fixed dividend of d_i in each time period the return for investor j from investing into asset i is for a time horizon of T_j given by

$$r_{ij} = \frac{1}{T_j} \frac{P_i(1 - s_i) + d_i T_j - P_i}{P_i} = \frac{d_i}{P_i} - \frac{s_i}{T_j}, \tag{7.1}$$

thus consisting of the gross return adjusted by the costs of the spread for each unit of time the investment is held. The return of the asset in the market will now be determined by those investors willing to pay the highest price, thus requiring the lowest return. Conversely, any investor will seek to hold that asset which provides him with the highest return.

Consider now two assets, l and m, which are held by different investors, p and q. With the foresaid we obviously require that $r_{lp} \geq r_{mp}$ and $r_{lq} \leq r_{mq}$. Inserting from (7.1) we obtain the conditions that

$$\begin{aligned} \frac{d_l}{P_l} - \frac{s_l}{T_p} &\geq \frac{d_m}{P_m} - \frac{s_m}{T_p}, \\ \frac{d_l}{P_l} - \frac{s_l}{T_q} &\leq \frac{d_m}{P_m} - \frac{s_m}{T_q}. \end{aligned} \tag{7.2}$$

We can now easily combine these two inequalities to obtain

$$(s_l - s_m)(T_p - T_q) \geq 0. \tag{7.3}$$

Thus if $s_l > s_m$ we require that $T_p > T_q$, which means that assets with a larger spread are held by investors with longer investment horizons. This result obtained here has been confirmed empirically by Atkins and Dyl (1990).

The market return will be the minimum return any investor requires as his demand will be served first due to him willing to pay the highest price:

$$r_i = \min_{j=1,2,\ldots,M} \left(r_{ij} + \frac{s_i}{T_j} \right). \tag{7.4}$$

We see that this return is increasing in the spread and as with the spread the time horizon increases this increase becomes less and less pronounced, i.e. r_i is concave in the spread.[1]

Using the asset with a spread as a benchmark, we see that the asset price is decreasing with the spread. Capitalizing the dividend at the respective market returns we obtain that $P_0 = \frac{d_0}{r_0}$ and $P_i = \frac{d_i}{r_i}$. By assuming that $d_0 = d_i$ for simplicity we get

$$\frac{P_i}{P_0} = \frac{r_0}{r_i}, \tag{7.5}$$

which is decreasing in the spread as we argued above that r_i was increasing in the spread. In an alternative interpretation, we can determine that the price of the asset is reduced by the present value of future trading costs:

$$\frac{d_i}{P_i} = r_i = \min_{j=1,2,\ldots,M} \left(r_{ij} + \frac{s_i}{T_j} \right), \tag{7.6}$$

$$P_i = \max_{j=1,2,\ldots,M} \left(\frac{d_i}{r_{ij} + \frac{s_i}{T_j}} \right), \tag{7.7}$$

$$P_i = \frac{d_i}{r_i} - P_i \frac{\frac{s_i}{T_j}}{r_i}. \tag{7.8}$$

We can thus conclude this analysis by stating that a higher spread increases the required gross rate of return companies have to achieve which in turn reduces the value of shares. This increases the costs of equity to companies and provides them an incentive to seek a listing at a stock exchange that is able to provide the lowest spread as is discussed in Amihud and Mendelson (1988). Another aspect apparent from the above analysis is that the required gross rate of return is also increasing the shorter the time horizon of investors is. Here we find incentives for companies to attract more long-term investors as this actually reduces the cost of equity.

Empirical evidence overwhelmingly support the finding that a higher bid-ask spread increases the required rate of return. Amihud and Mendelson (1986a)

[1] It can be shown that in equilibrium the increase in T_j does not lead to the adjustment term to reduce, although with $r_i = \frac{d_i}{P_i}$ the price will also change.

find evidence that an increase in the spread by one percentage point increases the return by about 2.5% p.a., a similar result is obtained in Amihud and Mendelson (1986b), Amihud and Mendelson (1989) and Amihud and Mendelson (1991) with additional evidence for the concavity of the relationship between the spread and the return. Eleswarapu (1997) also finds a positive relationship but points out that this result is mostly driven by a strong relationship in January. This seasonality is confirmed in Rubio and Tapia (1998) for the Spanish market.

The proposed effect on the value of assets is also confirmed in Amihud and Mendelson (1991) reporting that a doubling of the spread causes prices to drop by about 6%. For the Swiss stock exchange and the NASDAQ Loderer and Roth (2003) find that a spread of 1% causes a discount of the asset of about 9.4% relative to an asset trading without a spread. The average discount they find on the Swiss exchange is 12% and 28% for the NASDAQ. Mixed evidence on the effect of the spread can be obtained by investigating a change of exchange listings. While Baker and Edelman (1992) report a positive effect of a reduced spread on the asset value, the results of Kadlec and McConnell (1994) and Barclay et al. (1998) are inconclusive. Finally Benveniste et al. (2001) report a premium of 12-22% for liquid real estate investment trusts compared to illiquid vehicles trading at a significant spread.

Other models using a transaction cost approach to trading yield similar results to those obtained above, we could easily extend the model to include other relevant costs for investors such as broker fees. In particular Lo et al. (2004) develop a model which also shows a concave premium for higher transaction costs and a discount of the asset price increasing with trading costs. This discount is very low in Constantinides (1986) using a similar model, but this can be attributed to the fact that he only allows for very infrequent trading of the asset, thus requiring no significant compensation for trading costs. This is similar to Vayanos (1998) where the time horizon is endogenously determined. Other examples of model variations are Huang (2003) who also considers borrowing as an alternative to selling or Swan (2002) using a similar model to that used above but allowing for endogenous trading.

7.2 Compensation for adverse selection costs

The spread is not the only cost faced by investors in the market. As has been pointed out in chapter 3.2, informed traders always achieve an expected profit at the expense of uninformed traders. These losses have to be compensated for to the uninformed traders or they would not be willing to participate in the market. This compensation is not included in the spread as the spread only compensates the market maker. In this section we present a simplified model first developed in Easley et al. (2001) and O'Hara (2003) which allows us to determine the premium investors require for adverse selection.

Suppose a market with a single asset and a risk free asset B paying a fixed return of r. The risky asset is traded in a single trading round after which it is distributed to its holders at the fundamental value v. This fundamental value is not known to traders at the time of trading, but its distribution is common knowledge: $v \sim N\left(\overline{v}, \sigma_v^2\right)$. The market has three types of traders, noise traders, uninformed strategic traders and informed strategic traders. A fraction λ of the strategic traders are informed and receive a common signal $s \sim N\left(v, \sigma_s^2\right)$ about the true value of the asset.

Each trader i has an initial wealth $W_i = B_i + q_i p$, consisting of the risk free asset B_i and an investment of q_i units into the risky asset at price p. The wealth in the period after trading becomes

$$\begin{aligned} W_i' &= (1+r)B_i + q_i v \\ &= (1+r)(W_i - q_i p) + q_i v \\ &= (1+r)W_i + q_i(v - (1+r)p). \end{aligned} \tag{7.9}$$

Traders are risk averse with a common absolute risk aversion z and seek to maximize the expected utility using their information set Ω_i:

$$\begin{aligned}(7.10) \qquad E\left[U(W_i')|\Omega_i\right] &\approx U\left(E\left[W_i'|\Omega_i\right] - \frac{1}{2}zVar\left[W_i'|\Omega_i\right]\right) \\ &= U\left((1+r)W_i + q_i\left(E\left[v|\Omega_i\right] - (1+r)p\right)\right. \\ &\quad \left. -\frac{1}{2}zq_i^2 Var\left[v|\Omega_i\right]\right).\end{aligned}$$

The optimal demand can easily be derived as

$$(7.11) \qquad q_i = \frac{E\left[v|\Omega_i\right] - (1+r)p}{zVar\left[v|\Omega_i\right]}.$$

With informed traders receiving the common signal s we can use Bayes' theorem such that

$$\begin{aligned}(7.12) \qquad v_I &= E\left[v|\Omega_i\right] = \frac{\frac{\overline{v}}{\sigma_v^2} + \frac{s}{\sigma_s^2}}{\frac{1}{\sigma_v^2} + \frac{1}{\sigma_s^2}} = \frac{\overline{v}\sigma_s^2 + s\sigma_v^2}{\sigma_v^2 + \sigma_s^2}, \\ (7.13) \qquad \sigma_I^2 &= Var\left[v|\Omega_i\right] = \frac{\sigma_v^2\sigma_s^2}{\sigma_v^2 + \sigma_s^2}.\end{aligned}$$

Thus the demand of informed investors is determined as

$$(7.14) \qquad q_I = \frac{\overline{v}\sigma_s^2 + s\sigma_v^2 - (1+r)(\sigma_v^2 + \sigma_s^2)p}{z\sigma_v^2\sigma_s^2}.$$

Uninformed investors do not receive the signal directly, but can infer some information from the prevailing price. Suppose that

$$(7.15) \qquad p = \alpha\overline{v} + \beta s + \gamma q + \delta x,$$

where $q = \sum_i q_i + u$, with $u \sim N(0, \sigma_u^2)$ the demand of noise traders, is the total demand of all traders and x the total supply of the asset in the market. Define

$$(7.16) \qquad \theta = \frac{p - \alpha\overline{v} + (\gamma - \delta)q}{\beta} = s + \frac{\delta}{\beta}(x - q),$$

which can be inferred from the available information by uninformed traders. We easily see that in equilibrium with $E[q] = x$ we obtain

$$\begin{aligned}(7.17) \qquad E[\theta] &= E[s] + \frac{\delta}{\beta}(x - E[q]) = v, \\ Var[\theta] &= Var[s] + \frac{\delta^2}{\beta^2}Var[q] = \sigma_s^2 + \frac{\delta^2}{\beta^2}\sigma_u^2 = \sigma_\theta^2.\end{aligned}$$

Using θ as a signal the uninformed traders form their beliefs as follows:

$$v_U = E[v|\theta] = \frac{\frac{\overline{v}}{\sigma_v^2} + \frac{\theta}{\sigma_\theta^2}}{\frac{1}{\sigma_v^2} + \frac{1}{\sigma_\theta^2}} = \frac{\overline{v}\sigma_\theta^2 + \theta\sigma_v^2}{\sigma_v^2 + \sigma_\theta^2}, \tag{7.18}$$

$$\sigma_U^2 = Var[v|\theta] = \frac{\sigma_v^2\sigma_\theta^2}{\sigma_v^2 + \sigma_\theta^2}, \tag{7.19}$$

giving rise to the demand of uninformed investors of

$$q_U = \frac{\overline{v}\sigma_\theta^2 + \theta\sigma_v^2 - (1+r)(\sigma_v^2 + \sigma_\theta^2)p}{z\sigma_v^2\sigma_\theta^2}. \tag{7.20}$$

In equilibrium we obviously require that demand and supply equal, thus

$$\lambda q_I + (1-\lambda)q_U + E[u] = \lambda q_I + (1-\lambda)q_U = x. \tag{7.21}$$

Inserting from (7.14), (7.16) and (7.20) we easily obtain that

$$\begin{aligned} 0 = {} & \frac{1}{z\sigma_v^2}\overline{v} + \frac{\lambda\sigma_\theta^2 + (1-\lambda)\sigma_s^2}{z\sigma_s^2\sigma_\theta^2}s - \frac{\delta}{\beta}\frac{1-\lambda}{\sigma_\theta^2}q + \left(\frac{\delta}{\beta}\frac{1-\lambda}{z\sigma_\theta^2} - 1\right)x \\ & -(1+r)\frac{\sigma_s^2\sigma_\theta^2 + \lambda\sigma_v^2\sigma_\theta^2 + (1-\lambda)\sigma_v^2\sigma_s^2}{z\sigma_v^2\sigma_s^2\sigma_\theta^2}p. \end{aligned} \tag{7.22}$$

By comparing coefficients we see immediately that from (7.15)

$$\frac{\delta}{\beta} = \frac{\frac{\delta}{\beta}\frac{1-\lambda}{z\sigma_\theta^2} - 1}{\frac{\lambda\sigma_\theta^2 + (1-\lambda)\sigma_s^2}{z\sigma_s^2\sigma_\theta^2}} = \sigma_s^2\frac{\frac{\delta}{\beta}(1-\lambda) - z\sigma_\theta^2}{\lambda\sigma_\theta^2 + (1-\lambda)\sigma_s^2}, \tag{7.23}$$

which solves for

$$\frac{\delta}{\beta} = \frac{z\sigma_s^2}{\lambda} \tag{7.24}$$

and implying that

$$\sigma_\theta^2 = \sigma_s^2 + \frac{z^2}{\lambda^2}\sigma_s^4 + \sigma_u^2. \tag{7.25}$$

Now define

$$\begin{aligned} \varepsilon &= (1+r)\frac{\sigma_s^2\sigma_\theta^2 + \lambda\sigma_v^2\sigma_\theta^2 + (1-\lambda)\sigma_v^2\sigma_s^2}{\sigma_v^2\sigma_s^2\sigma_\theta^2}, \\ \alpha &= \frac{1}{\sigma_v^2\varepsilon}, \\ \beta &= \frac{\lambda\sigma_\theta^2 + (1-\lambda)\sigma_s^2}{\sigma_s^2\sigma_\theta^2\varepsilon}, \\ \gamma &= -z\frac{\sigma_s^2(1-\lambda)}{\sigma_\theta\varepsilon\lambda}, \\ \delta &= z\frac{\sigma_s^2(1-\lambda) - \sigma_\theta^2\lambda}{\sigma_\theta^2\varepsilon\lambda} \end{aligned} \tag{7.26}$$

as the coefficients for (7.15). We know that ex ante $E[s] = \overline{v}$ and $E[q] = x$ from the equilibrium condition, hence with the above parameters we have

$$
\begin{aligned}
E[p] &= \alpha\overline{v} + \beta E[s] + \gamma E[g] + \delta x \\
&= (\alpha + \beta)\overline{v} + (\gamma + \delta)x \\
&= \frac{\overline{v}}{1+r} - \frac{z}{\varepsilon}x, \\
E[v_U] &= \frac{\overline{v}\sigma_\theta^2 + E[\theta]\sigma_v^2}{\sigma_v^2 + \sigma_\theta^2} = \overline{v}.
\end{aligned}
\tag{7.27}
$$

Thus we obtain the expected return on this asset as

$$
\mu = \frac{E[v_U] - E[p]}{E[p]} = \frac{r\varepsilon\overline{v} + z(1+r)x}{\varepsilon\overline{v} - z(1+r)x}.
\tag{7.28}
$$

It is now straightforward, although tedious, to show that the expected return is increasing in the risk of the asset, σ_v^2, as well as the variance of the signal, σ_s^2, and noise trading σ_u^2. While the increasing returns for the risk of the asset are a direct consequence of the compensation risk averse traders require for holding the asset, the total compensation also includes the risk arising from the imperfect information. A less precise signal causes informed traders to trade more cautiously and thus the price reveals less information, increasing the risk of uninformed investors. Similarly does increased noise trading cause more uncertainty about the price emerging in equilibrium, thus increasing the risk to all traders.

Of more interest to determine the effect of adverse selection is the behavior of the expected return as the fraction of informed traders, λ, changes. We can derive that the expected return is indeed decreasing in λ. Obviously there cannot be any adverse selection if $\lambda = 1$, i.e. all information is public, a case which serves as a useful benchmark. With $\lambda < 1$ the uninformed investors make an expected loss when trading and as the price tends to reveal less and less information this risk increases, resulting in an a higher expected return to compensate for this effect. We can show that the adverse selection premium, using the case of $\lambda = 1$ as a benchmark, is concave in λ.

In an alternative approach to this problem, Gârleanu and Pedersen (2004) come to similar conclusions although their model focuses on distortions arising from adverse selection as the result of prices not fully reflecting all available information. In their approach traders may in the future make decisions based on prices that are not fully revealing and thus decisions might be suboptimal. The resulting utility loss has to be compensated by higher expected returns and obviously the more severe potential distortions are, the higher this return has to be, similar to the model we used above.

Empirical evidence confirms the relationship between adverse selection and expected returns. Easley et al. (2002) find that for every 10% of informed trading, used as a proxy for adverse selection, the expected return increases by 2.5% p.a. An earlier investigation by Brennan and Subrahmanyam (1996) also showed a positive relationship between adverse selection and expected returns. Evaluating the listing choice of companies, Baruch and Saar (2004) find evidence that companies seek listing at an exchange such that adverse selection costs are reduced, reducing their costs of equity and increasing company value. Thus empirical research confirms the results obtained here.

7.3 The price impact of trading

With prices essentially determined by supply and demand, it is apparent that a trader buying an asset increases demand and a trader selling increases supply of the asset in the market. With standard economic theory it is thus reasonable to assume that these trading decisions affect prices, even without the implications arising from asymmetric information. Empirical evidence in Breen et al. (2002) confirms this assertion by finding that a trade of 0.1% of outstanding shares affects the returns by about 2.65% in the direction of the trade, at least temporarily. It has to be noted that these price effects are not necessarily associated with information, but merely reflect the illiquidity of the market.

Suppose now that when executing an order of size x the price P_0 changes to $\widehat{P}_0$ as follows:

$$\frac{\widehat{P}_0 - P_0}{P_0} = \lambda x P_0, \tag{7.29}$$

where λ is the measure of illiquidity and xP_0 the trading volume.

Suppose now a risk averse investor with absolute risk aversion z invests into this asset for T time periods which has an expected return of μ per time period and a variance of σ^2:

$$\frac{\widehat{P}_T - P_0}{P_0} = \mu T + \varepsilon_T, \tag{7.30}$$

with $\varepsilon_T \sim N(0, \sigma^2 T)$. Assuming that the investment is liquidated at the end of the investment horizon at the true value P_T we can now determine the trader's optimal investment strategy. The investor has to decide whether to invest his initial wealth W_0 into riskless bonds, B, which are perfectly liquid or into x shares, thus with paying $\widehat{P}_0$ for shares worth P_0 we have

$$W_0 = (B - x\widehat{P}_0) + xP_0. \tag{7.31}$$

At liquidation we have with inserting from (7.29) and (7.30) and a risk free rate of r:

$$\begin{aligned} W_T &= (B - x\widehat{P}_0)(1 + rT) + xP_T \\ &= B(1 + rT) + xP_0\left((\mu - r)T - (1 + rT)\lambda x P_0 + \varepsilon_T\right), \end{aligned} \tag{7.32}$$

hence

$$\begin{aligned} E[W_T] &= B(1 + rT) + xP_0\left((\mu - r)T - (1 + rT)\lambda x P_0\right), \\ Var[W_T] &= x^2 P_0^2 \sigma^2 T. \end{aligned} \tag{7.33}$$

Making the usual approximation that the expected utility with absolute risk aversion z is $E\left[U(W_T)\right] = U\left(E[W_T] - \frac{1}{2} z Var[W_T]\right)$,we obtain the optimal investment amount as

$$xP_0 = \frac{(\mu - r)T}{z\sigma^2 T + 2\lambda(1 + rT)}. \tag{7.34}$$

Inserting this result back into (7.33) we get the expected utility as

$$E\left[U(W_T)\right] = B(1 + rT) + \frac{1}{2}\frac{(\mu - r)^2 T^2}{z\sigma^2 T + 2\lambda(1 + rT)}. \tag{7.35}$$

Let us consider a perfectly liquid asset, i.e. an asset with $\lambda = 0$, whose expected utility would be

$$E\left[U(W_T')\right] = B(1 + rT) + \frac{1}{2}\frac{(\mu' - r)^2 T}{z\sigma^2}. \tag{7.36}$$

From the derivation of the CAPM we know that in equilibrium the expected return of such an asset would follow[2]

$$\mu' = r + z\sigma^2, \tag{7.37}$$

hence (7.36) reduces to

$$E\left[U(W_T')\right] = B(1 + rT) + \frac{1}{2}z\sigma^2 T. \tag{7.38}$$

In order to compensate for the illiquidity of the asset, the return should be higher than for the perfectly liquid asset. This compensation should be such that the expected utility of holding them is equal, thus requiring that

$$z\sigma^2 = \frac{(\mu - r)^2 T^2}{z\sigma^2 T + 2\lambda(1 + rT)}, \tag{7.39}$$

which solves for

$$\mu = r + \sqrt{\frac{z\sigma^2\left(z\sigma^2 T + s\lambda(1 + rT)\right)}{T}}. \tag{7.40}$$

The liquidity premium is then given by

$$\psi = \mu - \mu' = \sqrt{\frac{z\sigma^2\left(z\sigma^2 T + s\lambda(1 + rT)\right)}{T}} - z\sigma^2. \tag{7.41}$$

Using partial derivatives we can easily obtain that the required liquidity premium is increasing in the illiquidity λ as should be expected. With increasing illiquidity demand for the asset is reduced as we see from equation (7.34), costs are not increasing linearly. This effect causes the liquidity premium to be concave in the illiquidity, a similar argument to the clientele effect in the model by Amihud and Mendelson (1986a)

[2] As we only have a single asset in our market, the covariance of the asset with the market of course becomes the variance of the asset.

As with a longer time horizon the costs are spread, the liquidity premium is actually reducing. An increasing risk aversion of the investor as well as a higher volatility of the asset the liquidity premium increases, despite the lower demand for the asset. However the effect on the liquidity premium is very small and can be neglected in most cases.

Assuming that a trade of 0.1% is the average trade size, we see from (7.29) and using the result in Breen et al. (2002) that $\lambda = 0.0265$. With this result we can calculate that the liquidity premium for a time horizon of 1 year is about 2.2% p.a., for 3 months it is 5.9% p.a. and for 1 month even 12.2% p.a. for an average stock. Such a liquidity premium is very much in line with empirical evidence as shown in Pástor and Stambaugh (2003).

Other models have been developed which provide very similar results to those presented here. Pereira and Zhang (2004) use the same framework but allow for dynamic trading strategies. Their analysis also includes an uncertain investment horizon and investors are subject to another price impact when selling their holdings at the end of their time horizon. Huang (2003) investigates uncertain investment horizons but focuses his analysis on the presence of borrowing constraints to finance consumption rather than having to sell the asset. Despite these differences his results are compatible with the above model.

7.4 Pricing liquidity risk

The models presented thus far all implicitly assumed that the costs investors face remain constant over time. Empirical evidence in Chordia et al. (2000), Huberman and Halka (2000) and Hasbrouck and Seppi (2001) suggest that various measures of liquidity do not only vary over time but are also varying systematically across different assets. Thus costs are varying and it is reasonable to propose that investors require a compensation for assuming this additional risk as shown in Acharya and Pedersen (2004).

Suppose a risk averse investor with absolute risk aversion z trading N assets which have a return of r_i, proportional liquidity costs of c_i and the covariance two assets i and j is $\sigma_{ij} = Cov[r_i - c_i, r_j - c_j]$, interpreted as the covariance of the net return. A trader will maximize the expected utility of his holding in these assets by choosing optimal weights for each of the assets, x_i, subject to the constraint that $\sum_{i=1}^{N} x_i = 1$.

$$E\left[U(r_P)\right] = U\left(E[r_P] - \frac{1}{2} z Var[r_P]\right), \tag{7.42}$$

where r_P denotes the net return on the portfolio of the trader. We have obviously that $E[r_P] = E\left[\sum_{i=1}^{N}(r_i - c_i)x_i\right] = \sum_{i=1}^{N}(\mu_i - \overline{c}_i)x_i$, with $\mu_i = E[r_i]$ and $\overline{c}_i = E[c_i]$. Furthermore it is that $Var[r_p] = \sum_{i=1}^{N}\sum_{j=1}^{N} x_i x_j \sigma_{ij}$. The Lagrange approach thus yields the objective function

$$L = (\mu_i - \overline{c}_i)x_i - \frac{1}{2} z \sum_{i=1}^{N}\sum_{j=1}^{N} x_i x_j \sigma_{ij} + \lambda(1 - \sum_{i=1}^{N} x_i), \tag{7.43}$$

from which we obtain the optimal solution that

$$\mu_i = \lambda + \overline{c}_i + z\sigma_{iP}, \tag{7.44}$$

where $\sigma_{iP} = Cov[r_i - c_i, r_P - c_P] = \sum_{j=1}^{N} x_j Cov[r_i - c_i, r_j - c_j] = \sum_{j=1}^{N} x_j \sigma_{ij}$. Suppose that there exists an asset k that is risk free in its gross return and trades at zero costs, hence as $\overline{c}_k = 0$ and $\sigma_{kP} = 0$ we obtain that $\mu_k = \lambda$. More commonly this risk-free rate is denoted r, thus $\lambda = r$ and (7.44) becomes

$$\mu_i = r + \overline{c}_i + z\sigma_{iP}. \tag{7.45}$$

With the portfolio of assets being a linear combination of the individual assets, the expected return of the portfolio becomes

$$\mu_P = r + \overline{c}_P + z\sigma_P^2. \tag{7.46}$$

With no differences between investors, the portfolio held will be identical for all investors. In equilibrium the portfolio held will thus have to be equal to the market portfolio, denoted by the subscript M. Using this notation and inserting (7.46) into (7.45) we obtain

$$\mu_i = r + \overline{c}_i + (\mu_M - \overline{c}_M - r)\frac{\sigma_{iM}}{\sigma_M^2}, \tag{7.47}$$

which on first sight looks very similar to the traditional CAPM, but on closer inspection the covariance term shows a significant difference:

$$\begin{aligned}(7.48)\qquad \sigma_{iM} &= Cov[r_i - c_i, r_M - c_M] \\ &= Cov[r_i, r_M] + Cov[c_i, c_M] - Cov[r_i, c_M] - Cov[r_M, c_i].\end{aligned}$$

It is apparent that only the first term, $Cov[r_i, r_M]$, is found in the CAPM while the three other terms are associated with the liquidity risk of the assets.

The second term reflects the commonality in liquidity as found empirically in Chordia et al. (2000), Huberman and Halka (2000) and Hasbrouck and Seppi (2001). The rationale for requiring an additional return for this risk is that when the covariance is high, an investor wishing to buy or sell an asset and is faced with high liquidity costs, he is very likely to find high costs also in other assets that otherwise might have served as suitable substitutes to trade in order to avoid these high costs. This liquidity risk has to be compensated.

When stocks are reacting sensitively to changes in the market liquidity, as measured by the third term, this reduces the required returns as any illiquidity in the market as this effect is already captured in the first term. Empirical evidence shows support for this relationship as shown in Pástor and Stambaugh (2003).

If the market is not performing well, investors are usually more willing to invest into liquid assets whose return requirements are consequently reduced as captured in the final term above. Here investors seem to value the opportunity to react to developments in the market at no great costs, thus it act as an insurance against liquidity shortfalls.

As we also find that

$$(7.49)\qquad \sigma_M^2 = Var[r_M - c_M] = Var[r_M] + Var[c_M] - 2Cov[r_M, c_M]$$

we see immediately that the first term in (7.48) does not recapture the traditional β of the CAPM, but the total risk to investors, including the liquidity risk.

Empirical work on the pricing of liquidity risk by Hasbrouck and Seppi (2001) suggests that it can explain about two thirds of the return differences across stocks, while Martínez et al. (2003) find no evidence of liquidity risk to be priced in the Spanish market. Acharya and Pedersen (2004) report a total contribution of liquidity risk to the expected return of 1.1% p.a., of which the commonality (the second term) contributes only 0.08% p.a., the sensitivity (the third term) also only 0.16% p.a. The most important contribution the the returns comes from the liquidity insurance (the final term) which contributes 0.82% p.a., thus becoming the most important liquidity risk for investors. The average liquidity costs are estimated at $\overline{c}_i = 3.5\%$ p.a. to give a total of 4.6% p.a. attributable to liquidity effects. Similarly Porter (2003) reports a liquidity premium between 2 and 5% p.a.

7.5 The equity premium puzzle

The return on equity in the United States has exceeded that of Treasury Bills, the proxy for the risk free rate, by about 6% p.a. over the last century. Such an equity premium has been pointed out by Mehra and Prescott (1985) to be inconsistent in traditional models with the preferences of investors as found by other empirical work. In order to require an equity premium of 6% p.a. investors have to be extremely risk averse with an absolute risk aversion of about 20, while empirical evidence gained in other settings suggests an absolute risk aversion of no more than 10 and in many cases even below 3. Also the introduction of fixed transaction costs for trading stocks does not alleviate this problem, as the costs would have to be prohibitively high in order to justify the observed returns. Standard economic models seem thus not be able to explain this "puzzle". Realistic parameter constellations suggest a risk premium of only 0.35% p.a. as pointed out in Mehra and Prescott (1985).

As Kocherlakota (1996) mentions, the puzzle remains robust against modifications of the underlying theory. He shows that modifying preferences, for example, does not resolve the puzzle. The introduction of market frictions is, however, seen as a possible solution to this problems as

> "...the models will be stronger if they explicitly take into account

the informational problems that lead to trading frictions."[3]

As the previous sections have explicitly investigated the effects such trading frictions have on the required return of assets, it would be reasonable to summarize the evidence and see how much they may be able to contribute to resolving the equity premium puzzle.

As we have seen that a spread of 1% increases the required return by about 2.5% p.a. and noting that the average effective spread for liquid stocks on the NYSE is about 0.2%, we obtain that these trading frictions contribute about 0.5% p.a. to the equity premium. Furthermore traders face adverse selection costs in the market;[4] evidence suggests that about 12% of trades are conducted by informed traders for liquid stocks on the NYSE, thus contributing about 3% p.a. to the equity premium. We furthermore found that the price impact contributes about 2% p.a., but we have to take into account that a part of the price impact will be due to informational asymmetries rather than genuine illiquidity. From Chung and Wei (2005) we can infer this amount is approximately 1.5% p.a., thus leaving us with 0.5% p.a. from illiquidity. Finally we found the liquidity risk to contribute about 1% p.a. to the equity premium.

Adding these components up we obtain that trade frictions contribute about 5% p.a. to the equity premium. Together with the 0.35% p.a. to compensate for the risk of the asset, these data suggest a total risk premium of 5.35% p.a., relatively close to the observed 6% p.a. As the liquidity of market has increased in recent years, the estimate for the liquidity risk is likely to be too low for the more distant past and would therefore underestimate the total risk premium it is able to explain.

This result has, however, to be taken with caution. In particular, no joint estimation has been conducted and any correlations between factor identified can easily overestimate the total effect. Although this analysis is no evidence that the equity premium puzzle has been solved, it provides some strong evidence that

[3] Kocherlakota (1996, p.66).

[4] Although adverse selection costs are included in the spread, they are the adverse selection costs facing market makers rather than investors.

trading frictions can indeed explain a substantial fraction of this puzzle.

Review questions

1. Why is the time horizon important for the size of the premium in the presence of a bid-ask spread?
2. Why is the expected return concave in the bid-ask spread?
3. Why do investors need to be compensated for adverse selection?
4. Why is the precision of information important for the adverse selection premium?
5. How is the liquidity premium determined?
6. Why is the liquidity premium not linearly increasing in the illiquidity?
7. What are the components of liquidity risk?
8. Why does the liquidity premium also compensate for a covariance between returns and trading costs?
9. How well can trade frictions explain the equity premium of stocks?
10. Why is it not appropriate to sum up the components of the equity premium as determined in individual investigations?

Application

You observe that the excess returns of stocks vary considerably across markets. Five different markets have been chosen and a variety of characteristics have been recorded for the leading index in that market. These characteristics are shown in the table below. Can the data given explain the different excess returns in the markets?

	GOB40	SRX50	HGSE25	UGR30	PBX100
Excess returns (p.a.)	0.147	0.103	0.076	0.119	0.091
Volatility (p.a.)	0.347	0.233	0.172	0.305	0.251
Volatility of cash flows (p.a.)	0.081	0.064	0.123	0.106	0.0096
β in world portfolio	1.34	1.29	1.41	1.17	1.22
Trading volume ($ bn p.a.)	107	375	275	462	327
Market capitalization ($ bn)	51	540	1021	527	616
Average spread (%)	1.52	1.34	0.81	0.48	1.24
Institutional trading (%)	24	39	74	56	93

Outlook

This book provided an overview of the basic models used in market microstructure theory for auction, dealer and limit order markets. We also saw the rationale for the different market forms as well as the relationship between the market microstructure and asset pricing.

Despite a vast amount of literature covering these aspects alone, we have not been able to consider many other trading rules that can have a substantial impact on the price formation process. Although it is well beyond the scope of this book to provide an exhaustive review of all these aspects, it is nevertheless worth mentioning a number of these trading rules such that the reader is aware of their relevance.

Competition between market makers When investigating dealer markets we had assumed that market makers either behave competitive or are granted a monopoly. In reality, however, they are in many cases in direct or indirect competition with each other, giving rise to partial market power either through monopolistic competition (market makers for different stocks) or implicit collusion (market makers for the same stocks). Dutta and Madhavan (1997) have shown that non-competitive spreads are likely to be observed in this case. This work also gave rise to a widespread discussion on the collusion of market makers to raise spreads by avoiding certain ticks systematically, Christie and Schultz (1994) initiated this debate with their findings on the avoidance of odd-eighth by market makers on the NASDAQ.

Competition between market makers and limit orders Throughout this book we have assumed that we operate either a pure dealer market where traders are only allowed to trade with market makers via market orders or a pure limit order market in which no market maker was present. In contrast, we observe in most cases that dealer markets allow for the submission of limit orders. Consequently limit orders are competing with market makers for order flow and it is likely that this competition will affect the resulting behavior of market makers and limit order traders. SEPPI (1997) as well as BONDARENKO AND SUNG (2003) model such a market structure.

Order preferencing arrangements In many markets such as the NASDAQ, brokers reach an agreement with a specific market maker to route their entire business towards them; the market makers in return agree to match the best available price. Brokers receive a payment for each order thus received. With a certain order flow guaranteed these arrangements will easily impact in the degree of competition between market makers and thus on the spread quoted as well as the market depth. BLOOMFIELD AND O'HARA (1998) investigate this impact in more detail.

Market entry in dealer markets In market with multiple competing market makers, such as the NASDAQ, the way new market makers can enter the market will have an impact on the degree of competition between market makers with all its consequences for the trading process as WAHAL (1997) and KRAUSE (2005) show.

Market transparency Markets differ substantially on the amount of information that is available to market participants. Differences include the knowledge of the order book beyond the best available bid and ask prices, the anonymity of quotes and of traders. Increased information about these aspects would without doubt increase the informational efficiency of the market and will have implications for the behavior of market participants as shown by BOARD AND SUTCLIFFE (2000), DE FRUTOSA AND MANZANO (2005) and FOUCAULT ET AL. (2004).

Opening and closing of trading In many cases the opening and closing prices on stock exchanges are determined differently from the process employed throughout the trading day. The way the opening prices are determined can be of relevance for subsequent trading as it does not only set a first benchmark for the price but also affects the holdings of investors. Similarly does the closing procedure affect the behavior prior to this. Bacidore and Lipson (2001) as well as Ellul et al. (2003) investigate this aspect in more detail.

Tick size Prices in markets have to be quoted at discrete prices on a given grid. If we allow this grid to change, e.g. by reducing the tick size this will have direct impacts on the price setting behavior of all market participants and traders, as already seen in chapter 4.3. More detailed models are presented in Cordella and Foucault (1999), Harris (1994) and Bourghelle and Declerck (2004).

Block trading In many cases stock exchanges have special trading facilities for large orders to which different trading rules apply. Such trades obviously provide information to all market participants and the way they conducted and how much information becomes available will affect the normal trading process. Models in this area include Seppi (), Booth et al. (2002) and Saar (2001).

Market fragmentation In a similar way to block trading for large orders do many stock exchanges provide a different trading facility for small orders, often through electronic trading. As larger orders can be split into a number of small order to be routed via that trading facility, the presence of fragmented markets is likely to alter the behavior of market participants and thus affect prices as Chowdhry and Nanda (1991) and Madhavan (1995) show.

Priority rules We usually assumed that trading was firstly conducted using strict price priority as the first priority rule, followed by time priority. In many cases, however, stock exchanges employ more complex priority rules on the secondary priority rule, such as elements of size priority. With the order in which orders are executed, it is obvious that prices are affected. MOULIN (2000) as well as ANGEL AND WEAVER (1998) investigate this aspect further.

Price limits and trading halts Many stock exchanges employ circuit breakers which prevent stocks from changing more than a predefined amount. Once reaching these limits the implications range from a suspension of trading for a short time to not allowing prices to go beyond the price limit for the remaining trading session. The presence of such price limits will not only be relevant once the price limit is reached, but also prior to that as the prospect of the stock reaching the price limit and the subsequent inability to trade will naturally change the behavior and thus prices. Investigations of this topic are found in CHAN ET AL. (2005) and EDELEN AND GERVAIS (2003).

Ownership of stock exchanges Stock exchanges have traditionally been mutual organizations dominated by brokers and dealers, only recently have they incorporated and are now in some cases themselves listed on the stock exchange as a public company. Such changes in the ownership of stock exchanges affects the incentives they have in changing the trading rules and are thus indirectly of relevance for the price formation. COUGHENOUR AND DELI (2002), PIRRONG (1999), PIRRONG (2001) and HART AND MOORE (1996) provide some insights into these topics.

Competition between stock exchanges Many stocks are listed at several stock exchanges and traded at the same time. Consequently there will be competition between the different stock exchanges to attract trading volume in these stocks, see e.g. the model in PARLOUR AND SEPPI (2003). Furthermore stock exchanges will also be competing to attract stocks for listing in the first place, as modeled in FOUCAULT AND PARLOUR (2004).

> But recent developments did not only see increased competition between stock exchanges, but also increased cooperations and alliances as investigated by ARNOLD ET AL. (1999). All these aspects will affect the behavior of market participants and therefore affect price formation.

In order to fully understand the operation of a stock exchange all of the aspects mentioned above have to be investigated. Using the models developed in these areas as well as the empirical evidence collected will help to build a market structure that is most suitable for the stock exchange. The details of the market structure will be affected by a large number of factors such as the characteristics of the companies listed as well as the characteristics of the investors. We can expect that in most cases a compromise between different aspects will have to be reached as it is quite unlikely that all factors will lead to the same optimal solution.

Using the provided references the reader is actively encouraged to investigate some of the trading rules mentioned here in more detail to deepen his understanding of the trading process. This is particularly valuable as in many cases the optimal rule is not what intuitively would be expected. Many competing aspects often give rise to market structures that would not be thought of as being optimal.

Appendices

Appendix A

The NASDAQ Stock Market

In chapter 1 possible market structures have been described, this appendix will give an overview of the structure of a specific stock market, the *NASDAQ Stock Market*. Although the *New York Stock Exchange* (*NYSE*) is the dominant exchange not only of the United States but of the entire world,[1] it faces fierce competition, especially from the NASDAQ Stock Market.[2]

Increased public attention has been paid in recent years to this market as many companies operating in fast growing sectors like information technology, biotechnology or telecommunications are listed on the NASDAQ. While the number of companies listed on the NYSE grew only slowly in the past years, the NASDAQ was able to attract a much larger number of companies. An increasing number of non-US companies consider to be listed on the NASDAQ rather than on the NYSE.[3] This increased importance of the NASDAQ also resulted in widened attention of NASDAQ trading rules in the academic literature, especially after the findings of Christie and Schultz (1994) on implicit collusion among NASDAQ market makers.

The NASDAQ is a dealer market with many market makers competing for the order flow of a specific security. For most securities there are between 3 and 15

[1] When mentioning *"Wall Street"* this mostly is referred to the NYSE, which is located at 11 Wall Street in New York, but it is also used as a synonym for any stock exchange in the United States.

[2] The importance can be seen from the wide dispersion of computers displaying NASDAQ quotes. In 1996 300,613 such computers were operated within the United States and 37,846 in other countries. The countries with the largest number of these computers were Canada (16007 computers), Switzerland (6731) and the United Kingdom (5984), see The NASDAQ Stock Market, Inc. (1997, p. 32).

[3] 418 foreign companies have been listed on the NASDAQ in 1997 compared to 343 on the NYSE.

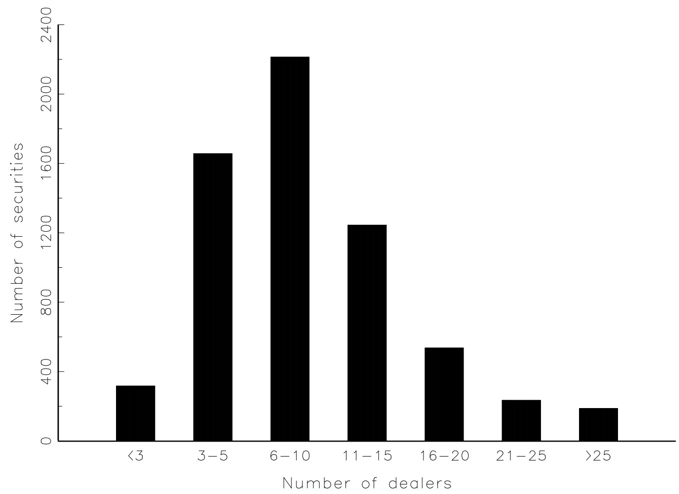


Fig. A.1: Distribution of the number of market makers for NASDAQ securities in 1996 *Data: NASDAQ Factbook 1997*

market makers, with an average of about 10 market makers per security. Every market maker makes the market for nearly 100 securities on average. Figure A.1 shows the distribution of the number of market makers per security.

A.1 History of the NASDAQ Stock Market[4]

With the Securities Exchange Act of 1934 every registered securities exchange was allowed to issue their own rules for admission to the market, listing of securities and trading within the framework of the Act. Exchanges were established as self-regulatory organizations. The Act did not encompass the trading taking place in securities not listed on a securities exchange, i.e. traded on the *over-the-counter market* (*OTC market*).

The Maloney Act of 1938 amended the Securities Exchange Act to provide also a framework for OTC transactions. It allowed the establishment of national

[4] This section follows Smith et al. (1998, pp. 3-14) unless otherwise stated.

securities associations that would issue guidelines for OTC trading and serve as *self-regulatory organizations* for their members. The only such association ever founded was the *National Association of Securities Dealers, Inc.* (*NASD*) in 1939. As the Act allowed members of such associations to discriminate against non-members in trading, the number of members increased from a portion of 22% of all firms engaged in securities trading in 1939 to 83% in 1982. Those firms not being members of the NASD where only regulated by the Securities and Exchange Act, while members were also subject to the rules of the association. In order to simplify and standardize supervision, in 1983 all firms engaged in the OTC markets were forced to join a national securities association. As the NASD was the only such association, nearly all companies trading securities became members, with the exception of those few only trading on a registered exchange, the NYSE, Amex or one of the five regional exchanges.[5]

A characteristic of OTC markets is that market participants are not centralized on a trading floor like the NYSE, but that they are dispersed all over the country or even the entire world. A transaction before the start of computerization typically occurred as follows: an investor submitted an order to a broker, the broker then tried to find out the market maker quoting the best price.[6] As a medium of communication in most cases the telephone had been used, the broker had to phone the market makers and ask for their quotes.[7] The broker traded with the market maker quoting the best price on his own account at the stated price. The customers had been charged a mark-up on this price to cover the expenses of the broker. The prices of the market makers were not published real-time due to the lack of any technologies enabling this, quotes (typically average or closing quotes) were published the following day in printed bulletins only

[5] The remaining regional exchanges in the United States are: Boston Stock Exchange, Chicago Stock Exchange, Cincinnati Stock Exchange, Pacific Stock Exchange, and Philadelphia Stock Exchange.

[6] OTC markets are typically organized as dealer markets. By having established a market maker it is easy to find a counterpart that otherwise would hardly be found as a result of decentralization.

[7] Often it happened that market makers and brokers were identical and the search process therefore was simplified, but it remained difficult to determine whether there existed a more favorable price from another market maker.

available to brokers and market makers.

Such a trading mechanism has the disadvantage that finding the best quoted price is very difficult, time intensive and may not be found, such that transactions occurred at less favorable prices. Transaction costs of trading will be high due to the lack of transparency in such a market. Furthermore the reaction of investors to new information was difficult as the informativeness of prices has been low. These inefficiencies of OTC markets made them not very attractive for investors and companies, hence for a long time they were no meaningful competitors to registered exchanges. Often for small companies it was the only possibility to raise new equity by being traded on OTC markets, most of them applied to be listed on an exchange when they fulfilled the listing requirements.

Improvements in telecommunications and computer technologies during the 1960's enabled quotes to be disseminated faster. In 1966 the NASD began to consider an automated quotation system that would allow real-time quotes to be displayed on screens connected with a central computer. The market makers would have been able to enter their quotes and the best quotes were to be displayed on screens, mentioning the market maker quoting it. With this information a broker could directly address the market maker quoting the best price and was no longer forced to call all market makers in order to get this information. Under the name *National Association of Securities Dealers Automated Quotation System* (*NASDAQ*) this system was put into operation on February 8, 1971 by linking about 500 market makers, a large number of brokers and even more interested parties, like investment consultants, to a central computer. Three different levels of service had been established: The *level 1 service* allowed to follow the market by observing the best available quotes. This service was designed for investment consultants and the public. With a *level 2 service*, available for institutional investors and brokers, it was possible to observe not only the best quotes, but all quotes of the market makers. The names of the market makers quoting the prices were also displayed. With *level 3 service* market makers were able to enter their quotes.

To be listed on the NASDAQ, companies had to meet minimal requirements with respect to size and corporate governance, otherwise they were not included into this new system.[8] Trading securities listed on the NASDAQ changed significantly. The determination of the best available price had become much easier and the transparency of the market increased. Also the settlement between brokers and their customers changed, brokers did no longer charge their customers a mark-up on the price they received from the market maker, but charged the same price and instead used commission fees to cover their costs.

Initially the NASDAQ had been designed only to disseminate information on quotes, information on trades having occurred could not be obtained. Improved computer technologies, however, allowed to provide these information for 40 of the most active securities in 1982. In due time more securities that met requirements more restrictive than being listed on the NASDAQ were incorporated into this new service, called *NASDAQ National Market System* (*NASDAQ/NMS*).[9] In 1983 682 securities were listed on the NMS, 2587 in 1990 and 4371 in 1996. Those securities not included into the NMS are mostly small and infrequently traded. These *regular NASDAQ*[10] securities were traded only with quote information until 1992, when information on trades had also been added.

By launching the *Small Order Execution System* (*SOES*) in 1984 the NASDAQ became a trading platform rather than only a tool for information dissemination of quotes and trades. The SOES enabled orders to be automatically routed to the market maker quoting the best price. The market maker only had to confirm the execution of orders by pushing a button, a confirmation of order execution is sent to the broker electronically without further personal interference. Originally participation in the SOES was voluntary for market makers, the use was restricted to NMS securities and order sizes of 500 shares or below. In 1985 all securities were included and the maximum order size for NMS securities raised to 1000 shares. During the crash of 1987 market makers were difficult to

[8] These requirements are stated in appendix A.3.
[9] In 1993 renamed into *NASDAQ National Market* (*NNM*).
[10] In 1993 this tier of the NASDAQ has been renamed into *SmallCap Market.*

reach by phone and many orders could not executed within an acceptable time. In reaction to this experience, participation in the SOES for NMS securities became mandatory for all market makers in 1988.

The same experience during the crash of 1987 led to the development of the *Order Confirmation Transaction Service* (*OCT*) in 1988,[11] where orders could be submitted electronically to a specific market maker instead of using the phone. By pushing a button to confirm the execution of the order, this system enables faster execution of trades, hence larger trading volumes can be processed than by using the phone only.

Also in 1988 the *Advanced Computerized Execution System* (*ACES*) has been introduced. Participation in this system is voluntary for market makers and brokers. It allows orders to be automatically routed to the best participating market maker and the execution is again confirmed only by pushing a button. Unlike in the SOES, the order size is not restricted by the system. Every market maker participating in this system has to negotiate with one or more brokers up to which order size he is willing to execute the orders at the stated prices.[12] He can negotiate different order sizes with different brokers and for different securities. A negotiation with all brokers is not necessary. Between large brokers and market makers similar private systems exist, especially in cases where brokers and market makers are employed by the same financial institution.

Since these developments the systems have continually been improved to be easier to handle and to be able to conduct an increasing number of trades. A new system called *NASDAQ Order Delivery and Execution System* (*NODES*) is currently awaiting approval by the Securities and Exchange Commission. It aims to replace and improve the current SOES and SelectNet.[13]

The progress in trading transparency and increased standards in regulation

[11] An improved system has been introduced in 1990 under the name *SelectNet*.

[12] As will be presented in section A.5 the trading rules require the quotes to be valid for a minimum order size. This system enables market makers and brokers to negotiate a higher order size bilaterally.

[13] See *Research Matters* 1(2), 1998, p. 4, published by the NASD Economic Research Department.

Year	**Event**
1939	Foundation of NASD
1971	NASDAQ starts operation as a quote dissemination system
1982	Introduction of a two tier market with dissemination of trade information for the National Market
1984	SOES launched with mandatory participation
1988	OCT introduced ACES introduced SOES becomes mandatory for National Market securities
1992	Dissemination of trade information for the SmallCap Market

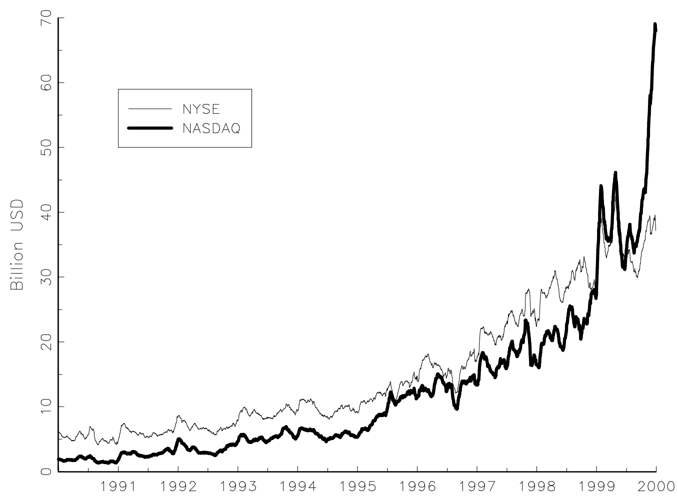
Tab. A.1: Main historical events of the NASDAQ

has made the former OTC market comparable to a securities exchange, consequently in most minds it is regarded as an exchange. Table A.1 summarizes the main events in the history of the NASDAQ. The NASDAQ is today widely accepted by investors and companies as a market comparable to the NYSE. It has become the largest market of the world in dollar and share trading volume ahead of the NYSE and the second largest in market capitalization, just behind the NYSE. In recent years it has significantly catched up with the NYSE and in many respects surpassed it.

Most recently the NASDAQ Composite Index outperformed the Dow Jones Industrial Average Index, which mostly consists of stock listed on the NYSE.[14] In combination with more attention being paid to internet and biotechnology stocks, which are mostly listed on the NASDAQ, the market received more and more interest from the general public. Figures A.2 to A.4 illustrate these recent developments.

Many large companies, although fulfilling the requirements to be listed on the NYSE, such as Microsoft or Intel, remain to be listed on the NASDAQ and also many foreign companies decide to be listed on the NASDAQ rather than on the NYSE. This gives evidence that the NYSE and NASDAQ have become equal competitors. Recent improvements in the transparency of the markets are

[14] Only in late 1999 the Dow Jones Index included large companies listed on the NASDAQ, like Microsoft or Cisco Systems.



Data: NYSE and NASDAQ

Fig. A.2: Daily US-Dollar trading volume on the NYSE and the NASDAQ (20 day moving average) *Data: NYSE and NASDAQ*

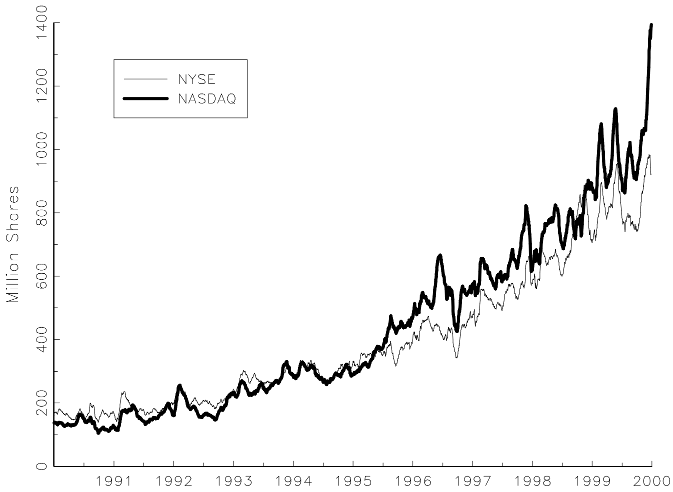


Fig. A.3: Daily share trading volume on the NYSE and the NASDAQ (20 day moving average) *Data: NYSE and NASDAQ*

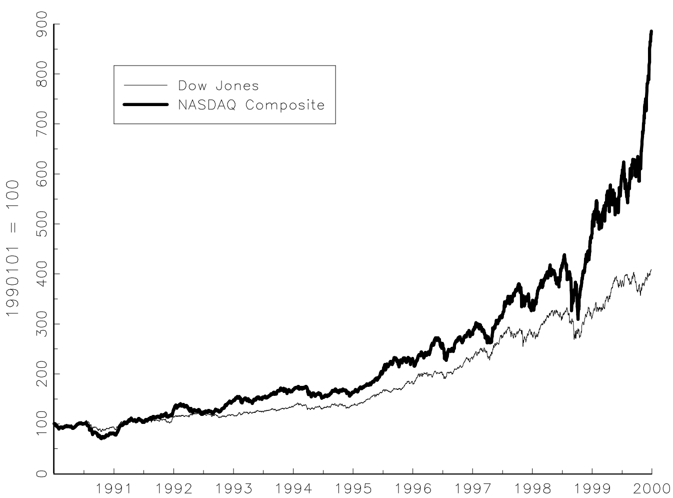


Fig. A.4: Development of the Dow Jones Industrial Average and NASDAQ Composite Index *Data: Datastream*

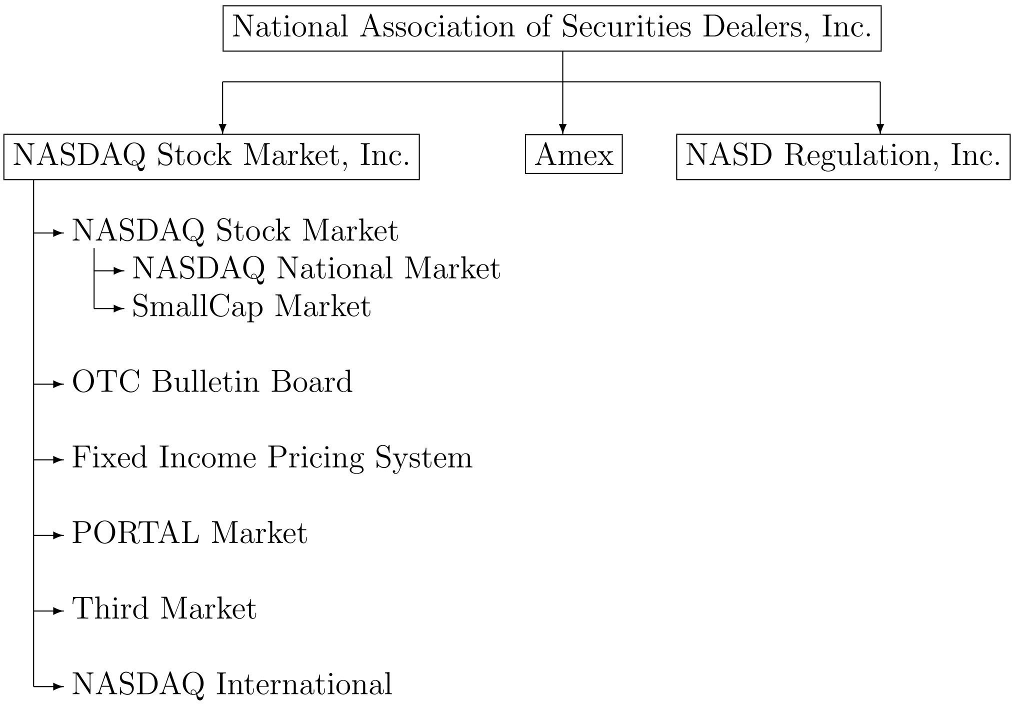


Fig. A.5: Organization of the NASD

attributed to the competition for trading volume and the listing of companies.

A.2 The organization of the NASD[15]

The *NASD* is a self-regulatory organization supported by its members, brokers and market makers trading on OTC markets. The NASD itself only operates departments that serve the organization as a whole, such as economic research, human resources or finance. All operations are conducted by three subsidiaries, the NASDAQ Stock Market, Inc.[16], the American Stock Exchange (Amex) and the NASD Regulation, Inc. Figure A.5 shows the organizational structure of the NASD.

The *NASDAQ Stock Market, Inc.* operates the different OTC markets. It develops and maintains the computer and telecommunications networks used for market operations and develops new trading systems. It further promotes

[15] This section is based on Smith et al. (1998, pp. 14-20).
[16] The NASD currently considers plans for a going public of the NASDAQ Stock Market, Inc.

the listing of securities and investments into NASDAQ securities, e.g. through sponsoring or seminars.

The NASDAQ Stock Market, Inc. runs several OTC markets, most prominent is the *NASDAQ Stock Market* with its two tiers, the NNM and the SmallCap Market. These markets are mostly referred to as NASDAQ, a convention that will also be used in this work. The other markets are of less importance and therefore receive only limited attention. The *OTC Bulletin Board* (*OTCBB*) is a pure quotation system for securities not listed on the NASDAQ or any exchange. No trade information is displayed and no trades can be conducted or initiated through this system. The *Fixed Income Pricing System* (*FIPS*) is a quotation and trade information system for about 50 of the most actively traded high yield corporate bonds (rated BB+ or lower by Standard & Poor's). The *Private Offerings, Resales and Trading Through Automated Linkages* (*PORTAL*) *Market* allows private placements of securities to be better allocated by publishing information on prices and the securities themselves. This market is restricted to institutional investors with an investment of at least USD 100 Mio. in security markets. It also provides a platform for trading those securities, but this possibility is rarely used. In the *Third Market* securities listed on a registered exchange can be traded off the exchange by using certain facilities of the NASDAQ and applying similar rules. The attempt to offer trading in NNM securities and selected foreign securities at European trading hours at *NASDAQ International* operating in exactly the same way as the NASDAQ Stock Market using its computer facilities, has not generated much interest thus far. Currently the NASDAQ is planning to expand their activity to Japan, Canada and Europe through building up new trading platforms and seeking cooperations with established stock exchanges and private trading platforms.

The *American Stock Exchange* (*Amex*) has become a subsidiary of the NASD since their merger has come into effect on October 30, 1998. It is a registered exchange and is operated independently of the other markets.

The *NASD Regulation, Inc.* (*NASDR*), founded in 1996, has overtaken all regulatory affairs that have formerly been conducted directly by the NASD.[17] Besides defining rules for trading on NASDAQ markets, it also supervises the compliance to these rules and has the authority to sanction violations. If laws have been violated it informs the legal authorities and cooperates in investigations. The NASDR further administers written tests to qualify securities professionals and registers them.

A.3 Listing requirements

To be listed on the NASDAQ Stock Market, a company has to meet certain criteria regarding corporate governance, public disclosure, and size. The aim of these criteria is to ensure a minimum of investor protection and to enable an orderly trading process.

According to Rule 4310 all companies have to meet the following qualitative criteria:[18]

- at least two independent directors on the board,
- an independent internal audit committee,
- an independent public accountant auditing the company,
- an annual report to be distributed to all shareholders,
- an annual meeting of shareholders,
- important corporate actions have to be approved by the shareholders,
- a quorum of at least $^1/_3$ of all outstanding shares for all decisions of the shareholders,
- prompt disclosure of information through the media that affect the value of the shares.

[17] As SCHULTZ (2000) points out, it was the Christie-Schultz debate that forced the NASD to give its regulatory division more autonomy by founding a separate subsidiary.

[18] These criteria can be adapted for foreign companies to meet the regulatory framework in their country of residence.

Additionally to these qualitative criteria, companies have to meet certain quantitative standards. To be listed on the SmallCap Market Rule 4310 requires the companies to meet these minimal criteria:

- net tangible assets: USD 4 Mio.,
- market capitalization: USD 50 Mio.,
  or net income: USD 750,000,
  or operating history: 1 year,
- number of publicly held shares: 1 Mio.,
- round lot shareholders[19]: 300,
- minimum bid price: USD 4,
- number of market makers for the security: 3.

To be listed on the NNM Rule 4420 requires that one of the following three standards has to be met:

- Standard 1:
  - net tangible assets: USD 6 Mio.,
  - pretax income: USD 1 Mio.,
  - number of publicly held shares: 1.1 Mio.,
  - market value of publicly held shares: USD 8 Mio.,
  - round lot shareholders: 400,
  - minimum bid price: USD 5,
  - number of market makers: 3.
- Standard 2:
  - net tangible assets: USD 18 Mio.,

[19] Shareholders holding at least shares of one trading lot, in most cases 500 shares.

- operating history: 2 years,
- number of publicly held shares: 1.1 Mio.,
- market value of publicly held shares: USD 18 Mio.,
- round lot shareholders: 400,
- minimum bid price: USD 5,
- number of market makers: 3.

• Standard 3:

- market capitalization: USD 75 Mio.,
  or total assets: USD 75 Mio. and total revenue: USD 75 Mio.,
- number of publicly held shares: 1.1 Mio.,
- market value of publicly held shares: USD 20 Mio.,
- round lot shareholders: 400,
- minimum bid price: USD 5,
- number of market makers: 4.

Once a company is listed, it can fall short of these quantitative criteria, but not of the qualitative criteria. In order to maintain the listing similar, less restrictive criteria have to be met according to Rules 4310 and 4450.

Additional to these criteria, companies have to pay an entry fee for being listed and an annual fee to maintain the listing. These fees depend on the market on which the company is listed and its size. The entry fee for a listing on the SmallCap Market is between USD 5,000 and USD 10,000, for a listing on the NNM between USD 5,000 and USD 50,000. The annual fees are USD 4,000 for the first security of a company listed on the SmallCap Market and USD 1,000 for each additional security. In the NNM this fee varies between USD 5,250 and USD 20,000.[20]

[20] See Rules 4510 and 4520.

A.4 Registration as broker and market maker

There exist two prerequisites to register as broker or market maker. The first concerns the capital requirements to ensure those market participants to conduct their duties without facing the threat of bankruptcy. These prerequisites are regulated by the *Securities and Exchange Commission* (*SEC*). The other prerequisites refer to their qualifications and are regulated by the NASD.

A *broker* has to maintain a net capital[21] of at least USD 100,000. A *market maker* needs a net capital of USD 2,500 for each security he makes the market in.[22] Additionally, a minimum of USD 100,000 and a maximum of USD 1 Mio. applies. In most cases brokers and market makers are companies rather than individuals, whose business is conducted by employees. In this case the company as a whole has to fulfill these requirements, it has not to be fulfilled for every single employee acting as market maker or broker.[23]

The Securities Exchange Act requires every broker and market maker acting on OTC markets, like the NASDAQ Stock Market, to be member of a national securities association, hence they have to be member of the NASD. Not only the brokerage companies and companies acting as market makers have to be registered, but the by-laws of the NASD require every employee of those companies who is involved in brokerage or market making activities to become a member. While companies are registered with approval of the SEC, their personnel has to prove their qualifications to become members. Without being registered as member, no individual is allowed to conduct businesses related to brokerage or market making.

The by-laws of the NASD require members to have an appropriate qualification to conduct the business they are assigned to.[24] According to Rules 1021, 1031 and 1041 these qualifications have to be shown by passing a qualification

[21] Net capital is the net worth adjusted for unrealized profits and losses, subordinated loans and similar.

[22] For securities with a market value of less than USD 5 per share the requirement is USD 1,000.

[23] See SEC Rule 15c3-1.

[24] See Article III, Section 2 of the by-laws of the NASD.

examination conducted by the NASDR. These examinations are designed to explore the qualifications for a specific duty. If the duties of a person change, it can be necessary to pass another examination for his new duties. After having passed the examination the person is registered as member and allowed to conduct the business he has been assigned to. Furthermore, Rule 1120 requires registered members to follow certain continuing education requirements.

When registered as broker, one is free to act as broker for all securities listed on the NASDAQ Stock Market. Being registered as market maker also allows to become market maker in every listed security, it is only necessary to register for the securities one wants to make a market in. Market making can begin the next trading day.[25] With registration as market maker for a specific security it is the obligation to quote always a price at which one is willing to buy and sell the security.[26]

To withdraw the registration as market maker for a specific security follows a similar process. A request to withdraw the registration for this security becomes effective the next trading day. The only restriction faced upon withdrawal is that it is not allowed to register again as market maker for the same security during the next 20 trading days.[27] This free market entry and exit in most cases leads to more than two market makers being registered for a security. Furthermore, it is the aim to prevent registered market makers from making extraordinary profits through their activities by imposing the threat of new market makers entering.

A.5 Trading rules

A market maker registered for a security has the obligation to quote prices both for buying and selling the security from the public during the trading hours from 9.30 am to 4 pm.[28] The monthly average spread a market maker quotes for a security must not exceed 150% of the average spread of all market makers for

[25] See Rule 4611.
[26] See Rule 4613. Section A.5 describes the trading rules in more detail.
[27] See Rule 4620.
[28] See Rules 4613 and 4617. Extending trading hours to 8 or 10 pm is currently considered by the NASDAQ.

this security. Rule 2440 and Interpretation IM-2440 require the market makers not to charge a too large spread. As a guideline a maximum spread of 5% is mentioned in this rule, depending on the circumstances, e.g. market conditions and characteristics of the security. Larger spreads can be justifiable, but also a spread of 5% may be viewed as too large by supervisors of the NASDR, forcing market makers to reduce their spread. However, no fixed rule can be applied to determine the maximum spread, it is subject to interpretations by the NASDR.

The quotes are further restricted by tick sizes, the increments have to be multiples of USD $^1/_{32}$ for securities with bid prices below USD 10 and USD $^1/_{16}$ for those above.[29] Quotes have to be firm, i.e. upon request the market maker has to trade at least at the stated prices, but he is free to choose a more favorable price for the transaction.[30] The obligation of a firm quote is only waived for a short period of time to enable the market maker an update of his quotes after having executed an order.[31]

Furthermore, quotes have to be valid at least for a normal trading size, a lot of 100 shares. The number of shares a market maker is willing to trade at the quotes are displayed on the screen next to their quote. Rule 4613 requires the minimum trade sizes for which the quotes have to be valid to be larger than 100 shares under certain conditions. For securities listed on the SmallCap Market the minimum trade sizes is 500 shares if the average daily non-block volume[32] exceeds 1000 shares or the bid price is below USD 10.

For securities listed on the NNM these limits are:

- 1000 shares if
  - the average daily non-block volume is above 3000 shares,

[29] Changing to a decimal system is considered for late 2000. Through 1997 the tick size has been USD $^1/_8$ for securities with a bid price above USD 10 and USD $^1/_{16}$ for securities with a bid price of USD 10 or below.

[30] When choosing a more favorable price he is not restricted to the tick sizes in determining the price he charges. The tick sizes only apply to quotes, not to transaction prices. More favorable prices than those quoted are frequently observed as a result of preferencing arrangements, which are described below.

[31] See Rule 3320 and interpretation IM-3320.

[32] For the definition of block trades see below.

  - the bid price is below USD 100, and
  - there are at least 3 market makers for this security.
- 500 shares if
  - the average daily non-block volume is above 1000 shares,
  - the bid price is below USD 150, and
  - there are at least 2 market makers for this security.
- 200 shares if
  - the average daily non-block volume is below 1000 shares,
  - the bid price is below USD 250, and
  - there are at least 2 market makers for this security.

Orders that are larger than the market makers are willing to accept, can be broken into parts of at least a lot and be executed like several smaller orders. This may result in different prices applied for each part and the parts may be executed by different market makers. Limit orders may also be executed in parts of at least a lot with offsetting orders. To avoid partial execution the order has to be specially marked as a *All-or-None order* by the investor.

Trades of 10,000 shares and above are called *block trades*. Such trades are subject to special treatment. They can either be traded through market or limit orders as a whole or be broken into several smaller orders within the normal trading procedure. Investors face the risk of influencing the price significantly in an unfavorable way through the placement of such an order. For this reason such orders are usually traded in a special market (*upstairs market*) for separate negotiation with other block trades.

An order arriving on the market has to be executed at the best available price (*price priority*), i.e. the market maker quoting the most favorable price has to execute the order at the stated or an even more favorable price.[33] If several

[33] See Rule 2320.

market makers quote the same price, the market maker executing the order can be chosen without restrictions by the brokers. Preferencing arrangements, as described below, are applied in most of these cases to determine the routing of the order flow.

Price priority and interpretation IM-2110-2 of Rule 2110 give the guidelines for handling *limit orders*. Limit orders can be accepted by market makers, but they do not have to be. By accepting a limit order, the market maker has to follow the established rules. He must not trade ahead of a limit order he has received, i.e. is not allowed to execute an offsetting order on his own account at the same or a less favorable price than the limit has been set (*public before dealer*). A market maker can immediately execute limit orders on his own account or can route them to other market makers. To enhance the transparency of the market, SEC Rule 11Ac1-4 requires unexecuted limit orders to be displayed in the quotes of a market maker. If the limit order has the best available price, also its size has to be displayed. The obligation is only waived for orders below 100 and above 10,000 shares and if it must not be partially executed (All-or-None orders).

The aim of these rules on the handling of limit orders is to guarantee a maximum of transparency and enhance competition further by allowing limit orders directly to compete with the quotes of market makers.[34]

As has been stated above, *preferencing* in most cases determines the market maker who executes an incoming order in the case where several market makers quote the best price. With preferencing a broker routes his entire order flow to one specific market maker, provided he quotes the best available price and price priority can be applied.[35] The two main reasons for such a behavior are either

[34] Through 1994 limit orders were interpreted to be offers of investors to trade with a market maker at the stated price. Hence market makers could trade with other investors on their own accounts at less favorable prices, i.e. higher ask and lower bid prices. Limit orders were only executed against quotes of the market makers and not orders from other investors, consequently they also have not been published. From 1994 onwards, market makers were still allowed to trade ahead of limit orders, but only if they quoted the same or a more favorable price, limit orders had not to be published. In 1997 the current regulation has been introduced. Allegations of market makers colluding on wide spreads in 1994 lead to these changes in the handing of limit orders to enhance competition.

[35] In many cases preferencing arrangements require that the entire order flow is routed to a specific market maker, regardless of his current quotes. To fulfill the requirement of price

internalization or payment-for-order-flow.

In many cases companies act both as broker and market maker. In this case vertical integration results in preferencing, what is also called *internalization*. To maximize profits, the brokerage department has to route all orders to the own market makers if they quote the best available prices.

A broker may also be willing to route his order flow to a specific market maker because he receives a payment from him (*payment-for-order-flow*). This payment can either be in form of cash, or the market maker charges a more favorable price to the broker than his quote. The broker can either receive this difference to the quoted price by charging his customer the quoted price or he can forward the whole or a part of this surplus to his customer and gain a competitive advantage over other brokers, either by charging a more favorable price or by reducing his commission fees. Other forms of payments can also include various services, e.g. research reports on companies or conducting the clearing process. Payments typically have a value between USD .01 and USD .02 for each share.

Those market makers and brokers participating in ACES or similar private arrangements are very likely in preferencing arrangements. Preferencing adds another source of competition between market makers, besides price competition they also compete for trading volume.

A.6 Summary

The NASDAQ has grown out of a telecommunications network for disseminating quote information to a network having all features of an exchange. This development has been made possible by improvements in computer and telecommunications technologies. Differences to registered computerized exchanges are only minor nowadays, so that the NASDAQ is mostly referred to as an exchange, although it misses this formal status and is "only" an OTC market.

The NASDAQ is characterized by the competition of market makers for the

priority for investors, the market makers in turn guarantee to charge only the best available price, i.e. if necessary they improve their quotes. Such arrangements are also called *price matching arrangements*.

best price and for trading volume. If admitted as market maker to the NASDAQ, there are virtually no entry barriers for market making in a specific security. This allows for hit-and-run competition, ensuring no extraordinary profits to be made by market makers, hence low spreads should be expected.[36] Together with rules ensuring high standards of transparency for the market, trading costs should be low. Other stock exchanges, like the NYSE, have to establish a much more complex set of rules to abandon the use of market power that arises as the result of high entry barriers or a lack of competition by granting monopolies of market making.

Listing requirements of the NASDAQ are much less restrictive compared to those of other stock exchanges, e.g. the NYSE, as it is designed for small companies. These small, in many cases highly innovative companies operating in fast growing industries, made the NASDAQ well known to be a market for high-tech stocks. High market transparency and a seemingly competitive trading environment induced many companies having grown to sizes that qualify for a listing on the NYSE to remain their listing on the NASDAQ. It has become the major competitor of the NYSE for the listing of companies and the NASDAQ Composite Index is one of the most important indices of the world, having received increased attention in recent years.

[36] This result can best be described with the theory of *contestable markets*. It states that the threat of new market participants entering in case of excess profits, forces market incumbents to charge competitive prices. The absence of entry barriers (legal restrictions, sunk costs) are the prerequisites for a contestable market. Baumol et al. (1988) provide a detailed overview of the theory of contestable markets. However, implicit collusion between market makers will enable them, despite these competitive forces, to quote noncompetitive prices and receive excess profits.

Appendix B

Mathematical methods

The aim of this appendix is to provide a short introduction to selected topics which are used in the main parts of this books. Necessarily the coverage is not comprehensive and may at times be incomplete for the sake of simplicity. For a more detailed overview the reader is referred to specialist literature on the topics. Generally it is assumed that the reader is familiar with the basics of analysis, algebra and statistics.

B.1 Taylor series expansion

Let $f : \mathcal{D} \mapsto \mathbb{R}$ be $(n+1)$-times differentiable on $\mathcal{D}$ with $\mathcal{D} \subset \mathbb{R}$.[1] We then can define a *Taylor polynome* by

$$\begin{aligned} T_n : \quad & \mathcal{D} \mapsto \mathbb{R} \\ & x \mapsto T_n \equiv \sum_{k=0}^{n} \frac{f^{(k)}(a)}{k!}(x-a)^k, \end{aligned} \tag{B.1}$$

where $a \in \mathcal{D}$ and $n \geq 0$ and $f^{(k)}$ denotes the kth derivative of f.

Let there exist an open interval $\mathcal{D}_0 \subset \mathcal{D}$, $a \in \mathcal{D}$ and two constants $c > 0$ and $M > 0$ such that for all $k \geq 0$ and $x \in \mathcal{D}_0$

$$\left| f^{(k)}(x) \right| \leq c k! M^k. \tag{B.2}$$

If this condition is fulfilled, we can expand f into a Taylor series:

$$f(x) = \sum_{k=0}^{\infty} \frac{f^{(k)}(a)}{k!}(x-a)^k \tag{B.3}$$

[1] We can derive similar results for multidimensional functions $f : \mathcal{D} \mapsto \mathbb{R}^M$ with $\mathcal{D} \subset \mathbb{R}^N$.

for all $x \in \mathcal{D}_0$ with $|x-a| < \frac{1}{M}$.

As higher order terms become arbitrarily small, we can approximate f by

(B.4) $$f(x) = \sum_{k=0}^{n} \frac{f^{(k)}(a)}{k!}(x-a)^k$$

and call this an *expansion of f into a nth order Taylor series around a.* In practice it is always assumed that the conditions for a Taylor series expansion are met.

B.2 Dynamic programming[2]

When individuals have not only to make decisions that are optimal at a given point of time by choosing an optimal value of the *control variables*, u, but that these values have to be chosen optimal over a certain period of time, they have to determine an optimal *time path* for these variables. In many cases an additional problem arises, that the environment determining the outcome of this optimization problem changes through a changing *state variable*, x, which may be influenced by the control variables.

We define a function $I(x,u,t)$ which measures the payoff at a certain point of time, t. The aim of the individual now is to maximize the payoffs he receives over time, i.e. the *control problem* is given by[3]

(B.5) $$\max_{u(t)} J = \int_{t_0}^{t_1} I(x,u,t)dt,$$

where t_0 and t_1 denote the starting and end point of the considerations. The state variable changes according to the differential equation

(B.6) $$\frac{\partial x}{\partial t} = f(x,u,t).$$

We define the solution to the control problem by $J^*(x,t)$ and call this the *optimal performance function.*

[2] This section is based on Intriligator (1971).
[3] In dynamic programming it is usually assumed that future payoffs are not discounted to their present value.

The *principle of optimality* now requires that regardless of the current state the remaining decisions have to be optimal. Therewith at point $t+\Delta t$ with state $x+\Delta x$ the optimal performance function has to be $J^*(x+\Delta x, t+\Delta t)$. We can now write the optimal performance function as

$$J^*(x,t) = \max_{u(t)}\{I(x,u,t)\Delta t + J^*(x+\Delta x, t+\Delta t)\}, \tag{B.7}$$

which is known as the *fundamental recurrence relation*. Here $I(x,u,t)\Delta t$ denotes the payoff in the interval $]t, t+\Delta t]$. Approximating the second term in brackets by a first order Taylor series around (x,t), we get

$$J^*(x+\Delta x, t+\Delta t) = J^*(x,t) + \frac{\partial J^*}{\partial x}\Delta x + \frac{\partial J^*}{\partial t}\Delta t, \tag{B.8}$$

which gives after inserting into (B.7):

$$0 = \max_{u(t)}\{I(x,u,t)\Delta t + \frac{\partial J^*}{\partial x}\Delta x + \frac{\partial J^*}{\partial t}\Delta t\}. \tag{B.9}$$

Dividing by Δt and taking the limit $\Delta t \to 0$ we get with

$$\lim_{\Delta t \to 0}\frac{\Delta x}{\Delta t} = \frac{\partial x}{\partial t} = f(x,u,t) \tag{B.10}$$

$$-\frac{\partial J^*}{\partial t} = \max_{u(t)}\left\{I(u,x,t) + \frac{\partial J^*}{\partial x}f(x,u,t)\right\}. \tag{B.11}$$

This partial differential equation is known as the *Bellman equation*. Solving this equation will give the optimal performance function, given boundary conditions.

B.3 Constrained optimization

We want to find the optimal solution to a problem by maximizing (or minimizing, which works similar to the method demonstrated here) an objective function. In many cases however the individual conducting this optimization faces one or more constraints, such as a limited budget to buy goods or assets. Suppose we face the following problem:

$$\begin{aligned} \max_{(x,y)\in\mathcal{R}} \quad & f(x,y) \\ s.t. \quad & g(x,y) = 0. \end{aligned} \tag{B.12}$$

If the constraint is originally that $h(x,y)=c$, we simply define $g(x,y)=h(x,y)-c$ and obtain the above constraint. With this setting we can now define the Lagrangian function as

$$\max_{(x,y,\lambda)\in\mathcal{R}\times\mathbb{R}} f(x,y)+\lambda g(x,y), \tag{B.13}$$

where λ is the Lagrangian multiplier. This equation is now equivalent to the initial problem (B.12) as easily can be seen. We solve this maximization problem by analyzing the following equations simultaneously:

$$\begin{aligned} f_x+\lambda g_x &= 0, \\ f_y+\lambda g_y &= 0, \\ g(x,y) &= 0, \end{aligned} \tag{B.14}$$

where the indices x and y indicate the partial derivative of the function with respect to this variable. We see that the final condition recovers our constraint. The second order condition for this optimization problem is then derived from the Hesse-matrix being negative definite:

$$H(x,y)=\begin{bmatrix} f_{xx}-\lambda g_{xx} & f_{xy}-\lambda g_{xy} & g_x \\ f_{xy}-\lambda g_{xy} & f_{yy}-\lambda g_{yy} & g_y \\ g_x & g_y & 0 \end{bmatrix}. \tag{B.15}$$

The Lagrange multipliers also have an economic interpretation. Consider the initial constraint that $h(x,y)=c$, we then can write that $\frac{\partial f(x,y)}{\partial c}=f_x\frac{\partial x}{\partial c}+f_y\frac{\partial y}{\partial c}$. Using the optimality conditions in (B.14) and noting that $g_x=h_x$ and $g_y=h_y$ we obtain

$$\begin{aligned} \frac{\partial f(x,y)}{\partial c} &= \lambda\left(h_x\frac{\partial x}{\partial c}+h_y\frac{\partial y}{\partial c}\right) \\ &= \lambda, \end{aligned} \tag{B.16}$$

where the final equation follows from the fact that $h(x,y)=c$ implies $h_x\frac{\partial x}{\partial c}+h_y\frac{\partial y}{\partial c}=1$.

We can thus interpret the value of λ as the shadow value of the constraint, i.e. by how much the objective function would change if we changed the constraint marginally. In cases where the constraint is a budget constraint, we can interpret λ as the marginal utility of this budget.

B.4 Conditional moments

B.4.1 Truncated distributions

We have a random variable x which has a distribution $f(x)$. Let us assume that we know that the value of x is above a threshold t. We are interested in the distribution of x given the information that $x > t$. Using Bayes' rule we immediately see that this distribution $f(x|x > t)$ is given by

$$f(x|x>t) = \frac{f(x)}{Prob(x>t)}. \tag{B.17}$$

This function is also referred to as the truncated distribution. The moments of this distribution are now simply given by

$$\begin{aligned} E[x|x>t] &= \int_t^{+\infty} x f(x|x>t)dx, \\ Var[x|x>t] &= \int_t^{+\infty} (x - E[x|x>t])^2 f(x|x>t)dx. \end{aligned} \tag{B.18}$$

A particularly simple case can be obtained for the normal distribution $f(x) = N(\mu, \sigma^2)$. In this case we can show that

$$f(x|x>t) = \frac{\frac{1}{\sigma}\phi\left(\frac{x-\mu}{\sigma}\right)}{1-\Phi\left(\frac{t-\mu}{\sigma}\right)}, \tag{B.19}$$

where $\phi(\cdot)$ denotes the standard normal distribution and $\Phi(\cdot)$ its cumulative distribution. The moments of this truncated normal distribution can be found as

$$\begin{aligned} E[x|x>t] &= \mu + \sigma\frac{\phi\left(\frac{t-\mu}{\sigma}\right)}{1-\Phi\left(\frac{t-\mu}{\sigma}\right)}, \\ Var[x|x>t] &= \sigma^2\left(1 - \frac{\phi\left(\frac{t-\mu}{\sigma}\right)}{1-\Phi\left(\frac{t-\mu}{\sigma}\right)}\left(\frac{\phi\left(\frac{t-\mu}{\sigma}\right)}{1-\Phi\left(\frac{t-\mu}{\sigma}\right)} - \frac{t-\mu}{\sigma}\right)\right). \end{aligned} \tag{B.20}$$

For other forms of truncation, such as $f(x|x < t)$, $f(x|t' < x < t)$ as well as other distribution functions, see JOHNSON AND KOTZ (1970) for a comprehensive overview.

B.4.2 Joint distributions

Suppose we have two random variables x and y who have a joint distribution $f(x,y)$. The marginal distributions are defined as

$$
\begin{aligned}
f_x(x) &= \int_y f(x,t)dt, \\
f_y(y) &= \int_x f(t,y)dt.
\end{aligned}
\tag{B.21}
$$

They are thus the probability distributions in only one variable. Suppose now that we know the value of one of the two variables, say x. We now want to obtain the probability distribution of y conditional on the value of x. We can use Bayes' theorem to obtain that in that case the distribution function is given by

$$
f(y|x) = \frac{f(x,y)}{f_x(x)}. \tag{B.22}
$$

This expression serves as the new density function, the conditional mean and variance are obviously given by

$$
\begin{aligned}
E[y|x] &= \int_y yf(y|x)dy, \\
Var[y|x] &= \int_y (y - E[y|x])^2 f(y|x)dy.
\end{aligned}
\tag{B.23}
$$

This expression is in general very difficult to evaluate as the marginal distributions can be very different from the original joint distribution. The exception to this is the multi-variate normal distribution, whose marginal distributions are normal and for whose conditional moments we have some very simple formulae.

Let $\mu = (\mu_x, \mu_y)$ be the vector of expected values of the two variables x and y, σ_x^2 and σ_y^2 their variances and σ_{xy} their covariance. If x and y have a joint normal distribution with mean μ and covariance matrix $\Sigma = \begin{bmatrix} \sigma_x^2 & \sigma_{xy} \\ \sigma_{xy} & \sigma_y^2 \end{bmatrix}$, $f(x,y) = N(\mu, \Sigma)$, we have as the marginal distributions $f_x(x) = N(\mu_x, \sigma_x^2)$ and $f_y(y) = N(\mu_y, \sigma_y^2)$ and the conditional distribution is given by $f(y|x) = N\left(\mu_y + \frac{\sigma_{xy}}{\sigma_x^2}(x - \mu_x), \sigma_y^2 - \frac{\sigma_{xy}^2}{\sigma_x^2}\right)$.

The conditional moments are thus given by

$$E[y|x] = \mu_y + \frac{\sigma_{xy}}{\sigma_x^2}(x-\mu_x), \qquad \text{(B.24)}$$
$$Var[y|x] = \sigma_y^2 - \frac{\sigma_{xy}^2}{\sigma_x^2}.$$

We use such distributions for example in cases where investors are not knowing the fundamental value of an asset, y, but receive a signal x which is correlated with the fundamental value. This signal is then used to update the information the investor has on the fundamental value in the way shown above.

For other joint distributions or other forms of conditioning, JOHNSON AND KOTZ (1970) provide a comprehensive overview.

B.5 Stochastic processes

B.5.1 Definition

Let us consider a variable that changes randomly over time investigate this variable at fixed intervals of length Δt. The easiest form the random variable x_t can evolve is given by

$$x_{t+\Delta t} = x_t + \mu\Delta t + \varepsilon_t', \qquad \text{(B.25)}$$

where $\varepsilon_t' \sim iidN(0,\sigma^2\Delta t)$. We can now define $\Delta x_t = x_{t+\Delta t} - x_t$ and the above equation becomes $\Delta x_t = \mu\Delta t + \varepsilon_t$. If we define a standard normally distributed variable $\varepsilon_t \sim iidN(0,1)$ equation (B.25) can be rewritten as

$$\Delta x_t = \mu\Delta t + \varepsilon\sigma\sqrt{\Delta t}. \qquad \text{(B.26)}$$

Letting Δt becoming very small, hence investigating the variables at ever smaller periods of time, will allow us to obtain the values in continuous time and we write this as

$$dx = \mu dt + \sigma\varepsilon\sqrt{dt}, \qquad \text{(B.27)}$$

which is also known as the *Brown-Wiener process*. With $\mu = 0$ $\sigma = 1$ we obtain the standard Wiener process

$$dz = \varepsilon\sqrt{dt} \qquad \text{(B.28)}$$

and it is quite common to write (B.27) as

$$dx = \mu dt + \sigma dz. \tag{B.29}$$

Although this stochastic process is the most widely used, there exist a wide variety of stochastic processes which have been widely investigated. TODOROVIC (1992) provides an introduction to the topic.

B.5.2 Itô's Lemma

One of the most fundamental results on stochastic processes is *Itô's lemma.* It essentially states how to differentiate a function $F(x,t)$ that contains a stochastic process, x. Using the Taylor series expansion of F we obtain

$$\begin{aligned} F(x+\Delta x, t+\Delta t) &\approx F(x,t) + \frac{\partial F}{\partial x}\Delta x + \frac{\partial F}{\partial t}\Delta t \\ &+\frac{1}{2}\left(\frac{\partial^2 F}{\partial x^2}\left(\Delta x\right)^2 + 2\frac{\partial^2 F}{\partial x \partial t}\Delta x \Delta t + \frac{\partial^2 F}{\partial t^2}\left(\Delta t\right)^2\right). \end{aligned} \tag{B.30}$$

In letting Δt becoming very small we obtain as above

$$dF(x,t) \approx \frac{\partial F}{\partial x}dx + \frac{\partial F}{\partial t}dt + \frac{1}{2}\left(\frac{\partial^2 F}{\partial x^2}\left(dx\right)^2 + 2\frac{\partial^2 F}{\partial x \partial t}dxdt + \frac{\partial^2 F}{\partial t^2}\left(dt\right)^2\right). \tag{B.31}$$

With $dx = \mu dt + \sigma dz$ we obtain easily that

$$dF(x,t) = \frac{\partial F}{\partial x}dx + \frac{\partial F}{\partial t}dt + \frac{1}{2}\frac{\partial^2 F}{\partial x^2}\sigma^2 dt, \tag{B.32}$$

which is known as *Itô's lemma.* To obtain the last equation the important assumption is made that as $\Delta t \to 0$, all terms with a higher than linear order will vanish.

Appendix C

Economic concepts

This appendix provides a brief overview of key economic concepts used implicitly or explicitly in this book. It does usually not provide a critique of the concepts presented nor does it allow for alternative approaches to the problems being highlighted. For a more comprehensive coverage, the reader is referred to specialist literature on the subjects. In compiling this appendix it was assumed that the reader is familiar is basic microeconomic theories.

C.1 Utility theory

Since *John von Neumann* and *Oskar Morgenstern* introduced the expected utility hypothesis in 1944, it has become the most popular criterion for modeling decisions under risk. This appendix will give a brief introduction into the reasoning behind expected utility and its implications for risk aversion.

C.1.1 The expected utility hypothesis

The value, and therewith the returns of assets depend on their future cash flows. These future cash flows usually cannot be predicted with certainty by investors, they are a random variables, hence returns are also random variables. Investment decisions therewith have to be made under risk.[1] In their work von Neumann

[1] According to Knight (1921, pp. 197ff.) a decision has to be made under *risk* if the outcome is not known with certainty, but the possible outcomes and the probabilities of each outcome are known. The probabilities can either be assigned by objective or subjective functions. Keynes (1936, p. 68) defines risk as the possibility of the actual outcome to be different from the expected outcome. In contrast, under *uncertainty* the probabilities of each outcome are not known or even not all possible outcomes are known. Cymbalista (1998) provides an approach of asset valuation under uncertainty. In this work we only consider decisions to be made under

and Morgenstern (1953) presented a criterion to make an optimal decision if five axioms are fulfilled.[2]

Let $A = \{a_1, \dots, a_N\}$ be the set of all possible alternatives an individual can choose between,[3] $S = \{s_1, \dots, s_M\}$ all states that affect the outcome,[4] and $C = \{c_{11}, \dots, c_{1M}; \dots; c_{N1}, \dots, c_{NM}\}$ the outcomes, where c_{ij} is the outcome if state s_j occurs and alternative a_i has been chosen.

Axiom 1. *A is completely ordered.*

A set is completely ordered if it is complete, i.e. we have either $a_i \succeq a_j$ or $a_j \succeq a_i$ for all $i, j = 1, \dots, N$ and "$\succeq$" denotes the preference. The set has further to be transitive, i.e. if $a_i \succeq a_j$ and $a_j \succeq a_k$ we then have $a_i \succeq a_k$. Axiom 1 ensures that all alternatives can be compared with each other and are ordered consistently.[5]

To state the remaining axioms we have to introduce some notations. Let any alternative be denoted as a lottery, where each outcome c_{ij} has a probability of p_{ij}. We can write alternative i as

$$a_i = [p_{i1}c_{i1}, \dots, p_{iM}c_{iM}] \text{ , where } \sum_{j=1}^{M} p_{ij} = 1 \text{ for all } i = 1, \dots, N.$$

Axiom 2 (Decomposition of compound lotteries). *If the outcome of a lottery is itself a lottery (compound lottery), the first lottery can be decomposed into its final outcomes:*

Let $a_i = [p_{i1}b_{i1}, \dots, p_{iM}b_{iM}]$ *and* $b_{ij} = [q_{ij1}c_1, \dots, q_{ijL}c_L]$. *With* $p^*_{ik} = \sum_{l=1}^{M} p_{il}q_{ilk}$[6] *we have* $[p_{i1}b_{i1}, \dots, p_{iM}b_{iM}] \sim [p^*_{i1}c_1, \dots, p^*_{iM}c_L]$

risk.

[2] Many different ways to present these axioms can be found in the literature. We here follow the version of Levy and Sarnat (1972, p. 202)

[3] As for $N = 1$ there is no decision to make for the individual it is required that $N \geq 2$.

[4] As with $M = 1$ the outcome can be predicted with certainty we need $M \geq 2$ possible states.

[5] The transitivity ensures consistent decisions of individuals. It is equivalent with the usual assumption in microeconomics that indifference curves do not cross.

[6] This representation of joint probabilities assumes that the two lotteries are independent. If the two lotteries where not independent the formula has to be changed, but the results remain valid. It is also assumed throughout this appendix that there is no joy of gambling, i.e. that there is no gain in utility from being exposed to uncertainty.

Axiom 3 (Composition of compound lotteries). *If an individual is indifferent between two lotteries, they can be interchanged into a compound lottery:*

If $a_i = [p_{i1}b_{i1}, \ldots, p_{ij}b_{ij}, \ldots, p_{iM}b_{iM}]$ *and* $b_{ij} \sim [q_{ij1}c_1, \ldots, p_{ijL}c_L]$ *then* $a_i \sim [p_{i1}b_{i1}, \ldots, p_{ij}[q_{ij1}c_1, \ldots, p_{ijL}c_L], \ldots, p_{iM}b_{iM}]$.

These two axioms ensure that lotteries can be decomposed into their most basic elements (axiom 2) and that more complex lotteries can be build up from their basic elements (axiom 3).

Axiom 4 (Monotonicity). *If two lotteries have the same two possible outcomes, then the lottery is preferred that has the higher probability on the more preferred outcome:*

Let $a_i = [p_{i1}c_1, p_{i2}c_2]$ *and* $b_i = [q_{i1}c_1, q_{i2}c_2]$ *with* $c_1 \succ c_2$, *if* $p_{i1} > q_{i1}$ *then* $a_i \succ b_i$.

Given the same possible outcomes this axiom ensures the preference relation"$\succ$" to be a monotone transformation of the relation "$>$" between probabilities.

Axiom 5 (Continuity). *Let* a_i, b_i *and* c_i *be lotteries. If* $a_i \succ b_i$ *and* $b_i \succ c_i$ *then there exists a lottery* d_i *such that* $d_i = [p_1 a_i, p_2 c_i] \sim b_i$.

This axiom ensures the mapping from the preference relation "$\succ$" to the probability relation "$>$" to be continuous.

The validity of these axioms is widely accepted in the literature. Other axioms have been proposed, but the results from these axioms are identical to those to be derived in an instant.

Given these assumptions the following theorem can be derived, where U denotes the utility function.

Theorem 1 (Expected utility principle). *An alternative* a_i *will be preferred to an alternative* a_j *if and only if the expected utility of the former is larger, i.e*

$$a_i \succ a_j \Leftrightarrow E\left[U\left(a_i\right)\right] > E\left[U\left(a_j\right)\right].$$

Proof. Define a lottery $a_i = [p_{i1}c_1, \ldots, p_{iM}c_M]$, where without loss of generality $c_1 \succ c_2 \succ \cdots \succ c_M$. Such an order is ensured to exist by axiom 1.

Using axiom 5 we know that there exists a lottery such that

$$c_i \sim [u_{i1}c_1, u_{i2}c_M] = [u_i c_1, (1 - u_i)c_M] \equiv c_i^*.$$

We can now use axiom 3 to substitute c_i by c_i^* in a_i:

$$a_i \sim [p_{i1}c_1^*, \ldots, p_{iM}c_M^*].$$

This alternative only has two possible outcomes: c_1 and c_M. By applying axiom 2 we get

$$a_i \sim [p_i c_1, (1 - p_i)c_M]$$

with $p_i = \sum_{j=1}^{M} u_{ij}p_{ij}$, what is the definition of the expected value for discrete random variables: $p_i = E[u_i]$.[7]

The same manipulations as before can be made for another alternative a_j, resulting in

$$a_j \sim [p_j c_1, (1 - p_j)c_M]$$

with $p_j = E[u_j]$. If $a_i \succ a_j$ then we find with axiom 4 that, as $c_1 \succ c_M$:

$$p_i > p_j.$$

The numbers u_{ij} we call the utility of alternative a_i if state s_j occurs. The interpretation as utility can be justified as follows: If $c_i \succ c_j$ then axiom 4 implies that $u_i > u_j$, we can use u_i to index the preference of the outcome, i.e. a higher u implies preference for this alternative and vice versa.

Therewith we have shown that $a_i \succ a_j$ is equivalent to $E\left[U\left(a_i\right)\right] > E\left[U\left(a_j\right)\right]$. $\square$

The criterion to choose between two alternatives, is to take that alternative with the highest expected utility. To apply this criterion the utility function has to be known. As in most cases we do not know the utility function, it is necessary to analyze this criterion further to derive a more handable criterion.

[7] The extension to continuous random variables is straightforward by replacing the probabilities with densities.

C.1.2 Risk aversion

> "Individuals are *risk averse* if they always prefer to receive a fixed payment to a random payment of equal expected value."[8]

From many empirical investigations it is known that individuals are risk averse, where the degree of risk aversion differs widely between individuals.[9] The Arrow-Pratt measure is the most widely used concept to measure this risk aversion. We will derive this measure following PRATT (1964), a similar measure has independently also been developed by ARROW (1963).

With the definition of risk aversion above, an individual prefers to receive a fixed payment of $E[x]$ to a random payment of x. To make the individual indifferent between a fixed payment and a random payment, there exists a number π, called *risk premium*, such that he is indifferent between receiving $E[x]-\pi$ and x. By applying the expected utility principle we see that the expected utility of these two payments has to be equal:

$$E\left[U\left(x\right)\right] = E\left[U\left(E\left[x\right]-\pi\right)\right] = U\left(E\left[x\right]-\pi\right). \tag{C.1}$$

The term $E\left[x\right]-\pi$ is also called the *cash equivalent* of x. Approximating the left side by a second order Taylor series expansion around $E\left[x\right]$ we get

$$\begin{aligned} E\left[U\left(x\right)\right] &= E\Bigg[U\left(E\left[x\right]\right) + U'(E[x])(x-E[x]) \\ &\qquad + \frac{1}{2}U''(E[x])(x-E[x])^2\Bigg] \\ &= U(E[x]) + U'(E[x])E[x-E[x]] \\ &\qquad + \frac{1}{2}U''(E[x])E[(x-E[x])^2] \\ &= U(E[x]) + \frac{1}{2}U''(E[x])Var[x]. \end{aligned} \tag{C.2}$$

where $U^{(n)}(E[x])$ denotes the nth derivative of U with respect to its argument evaluated at $E[x]$. In a similar way we can approximate the right side by a first

[8] DUMAS AND ALLAZ (1996, p. 30), emphasize added.

[9] Despite this clear evidence for risk aversion, many economic theories assume that individuals are risk neutral. Prominent examples in finance are the information-based models of market making (see section 3.2) and several asset pricing theories.

order Taylor series around $E[x]$ and get

$$U(E[x]-\pi) = U(E[x]) + U'(E[x])\pi. \tag{C.3}$$

Inserting (C.2) and (C.3) into (C.1) and solving for the risk premium π we get

$$\pi = \frac{1}{2}\left(-\frac{U''(E[x])}{U'(E[x])}\right) Var(x). \tag{C.4}$$

Pratt (1964) now defines

$$z = -\frac{U''(E[x])}{U'(E[x])} \tag{C.5}$$

as the *absolute local risk aversion*. This can be justified by noting that the risk premia has to be larger the more risk averse an individual is and the higher the risk. The risk is measured by the variance of x, $Var[x]$,[10] hence the other term in (C.4) can be interpreted as risk aversion. Defining $\sigma^2 = Var[x]$ we get by inserting (C.5) into (C.4):

$$\pi = \frac{1}{2} z\sigma^2. \tag{C.6}$$

If we assume that individuals are risk averse, we need $\pi > 0$, implying $z > 0$. It is reasonable to assume positive marginal utility, i.e. $U'(E[x]) > 0$, which implies that $U''(E[x]) < 0$. This relation is also known as the first Gossen law and states the saturation effect. The assumption of risk aversion is therefore in line with the standard assumptions of microeconomic theory.

The conditions $U'(E[x]) > 0$ and $U''(E[x]) < 0$ imply a concave utility function. The concavity of the function (radius) is determined by the risk aversion.[11]

Figure C.1 visualizes this finding for the simple case of two possible outcomes, x_1 and x_2, having equal probability of occurrence.

[10] A justification to use the variance as a measure of risk is given in appendix C.2.

[11] For risk neutral individuals the risk premium, and hence the risk aversion, is zero, resulting in a zero second derivative of U, the utility function has to be linear. For risk loving individuals the risk premium and the risk aversion are negative, the second derivative of the utility function has to be positive, hence it is convex.

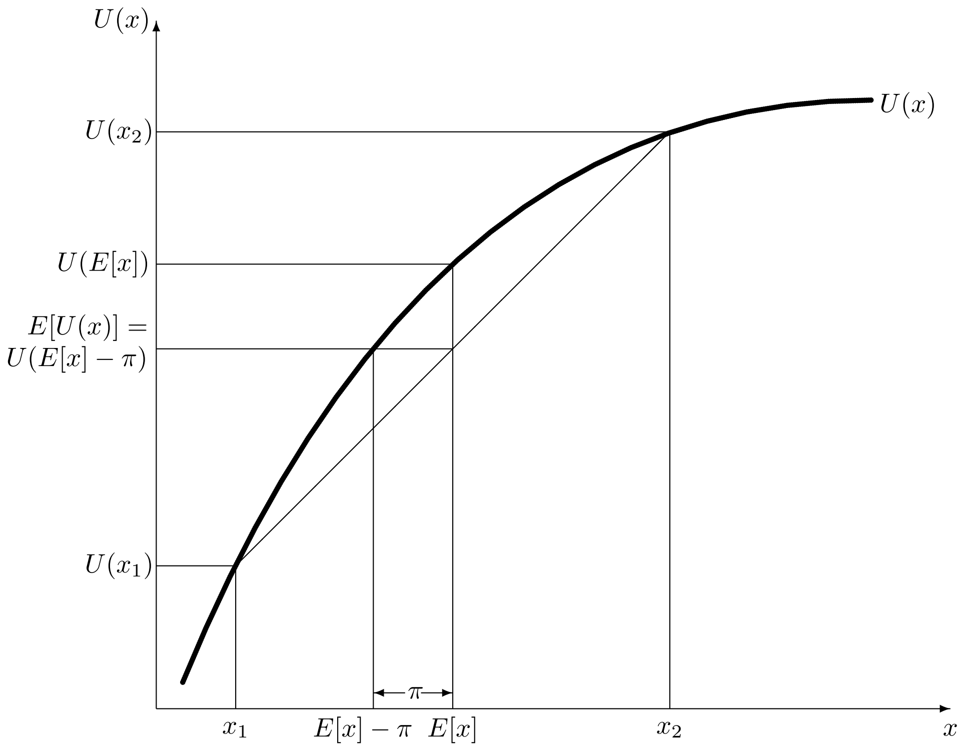


Fig. C.1: The Arrow-Pratt measure of risk aversion

C.2 Portfolio selection theory

When considering to invest into asset markets, an investor has to make three decisions:

- the amount he wants to invest into the asset market,
- determine the assets he wants to invest in,
- determine the amount he wants to invest into each selected asset.

This appendix describes a method how to make these decisions and find an *optimal portfolio*.[12] Such a portfolio

> "... is more than a long list of good stocks and bonds. It is a balanced whole, providing the investor with protections and opportunities with respect to a wide range of contingencies. The investor should build toward an integrated portfolio which best suites his needs."[13]

[12] A *portfolio* is the entirety of all investments of an individual.
[13] Markowitz (1959, p.3).

For this reason the associated theory is called *portfolio selection theory* or short *portfolio theory*, rather than asset selection. The portfolio selection theory has been developed by MARKOWITZ (1959), TOBIN (1958) and TOBIN (1966). Although the concepts employed in their theory have much been criticized for capturing the reality only poorly, it has been the starting point for many asset pricing models and up to date there has been developed no widely accepted alternative.

C.2.1 The mean-variance criterion

Even by using the Arrow-Pratt measure of risk aversion, the utility function has to be known to determine the first and second derivative for basing a decision on the expected utility concept. Preferable would be a criterion that uses only observable variables instead of individual utility functions. For this purpose many criteria have been proposed,[14] the most widely used is the mean-variance criterion. Although it also is not able to determine the optimal decision, it restricts the alternatives to choose between by using the utility function.

The *mean-variance criterion* is the most popular criterion not only in finance. The reason is first that it is easy to apply and has some convenient properties in terms of moments of a distribution and secondly by the use of this criterion in the basic works on portfolio selection by MARKOWITZ (1959), TOBIN (1958), and TOBIN (1966). Consequently, theories basing on their work, like the Capital Asset Pricing Model, also apply the mean-variance criterion, which by this mean became the most widely used criterion in finance.

It has the advantage that only two moments of the distribution of outcomes, mean and variance, have to be determined, whereas other criteria make use of the whole distribution. The outcome is characterized by its expected value, the mean, and its risk, measured by the variance of outcomes.[15]

[14] See LEVY AND SARNAT (1972, ch. VII and ch. IX) for an overview of these criteria.

[15] One of the main critics of the mean-variance criterion starts with the assumption that risk can be measured by the variance. Many empirical investigations have shown that the variance is not an appropriate measure of risk. Many other risk measures have been proposed, see BRACHINGER AND WEBER (1997) for an overview, but these measures have the disadvantage

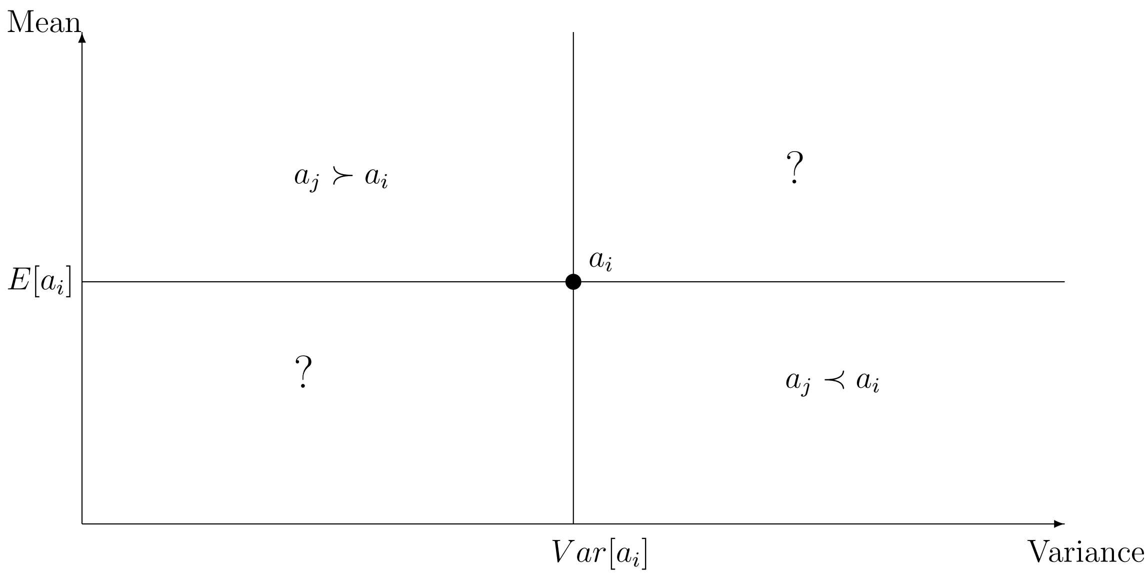


Fig. C.2: The mean-variance criterion

The mean-variance criterion is defined as

$$a_i \succeq a_j \Leftrightarrow \begin{cases} Var[a_i] < Var[a_j] \quad \text{and} \quad E[a_i] \geq E[a_j] \\ \text{or} \\ Var[a_i] \leq Var[a_j] \quad \text{and} \quad E[a_i] > E[a_j] \end{cases} . \tag{C.7}$$

It is a necessary, although not sufficient, condition to prefer a_i over a_j that $Var[a_i] \leq Var[a_j]$ and $E[a_i] \geq E[a_j]$. An alternative is preferred over another if it has a smaller risk (variance) and a larger mean. Nothing can in general be said about the preferences if $Var[a_i] > Var[a_j]$ and $E[a_i] > E[a_j]$, other decision rules have to be applied.

In figure C.2 an alternative in the upper left and lower right areas can be compared to a_i by using the mean-variance criterion, while in the areas marked by "?" nothing can be said about the preferences. If we assume all alternatives to lie in a compact and convex set in the (μ, σ^2)-plane,[16] all alternatives that are not dominated by another alternative according to the mean-variance criterion lie on a line at the upper left of the set of alternatives. In figure C.3 this is illustrated

of being less easily computable and difficult to implement as a criterion. In more recent models higher moments, such as skewness and kurtosis are also incorporated to cover the distribution in more detail.

[16] We will see that this condition is fulfilled in the case of portfolio selection for all relevant portfolios.

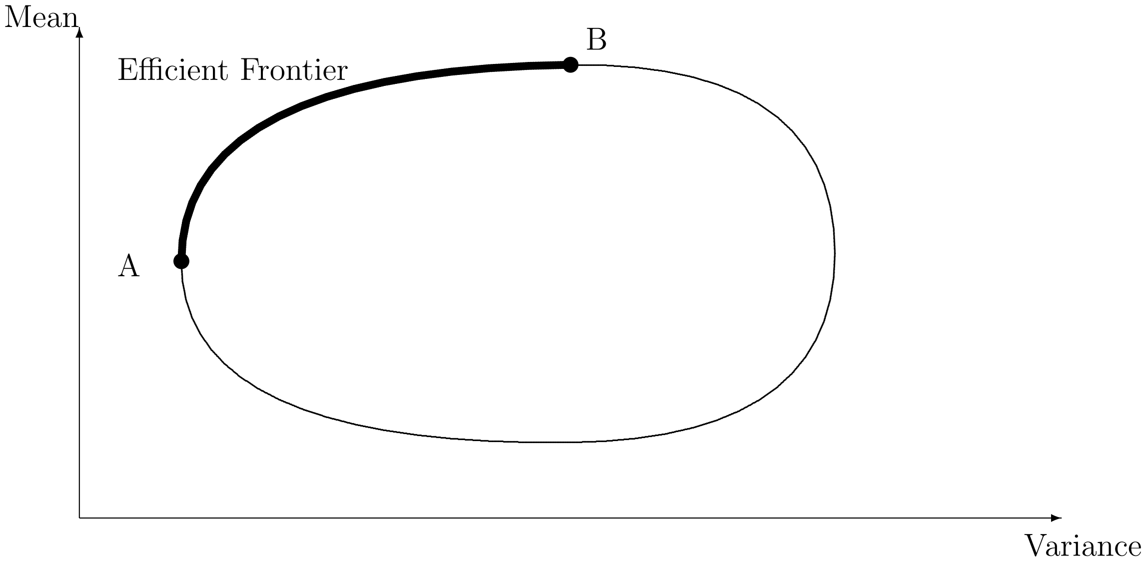


Fig. C.3: The efficient frontier

where all alternatives are located in the oval. The undominated alternatives are represented by the bold line between points A and B. All alternatives that are not dominated by another alternative are called *efficient* and all efficient alternatives form the *efficient frontier*. Without having additional information, e.g. the utility function, between efficient alternatives cannot be distinguished.

The mean-variance criterion can be shown to be not optimal in general, i.e. the true preferences are not always reflected by the results of this criterion.[17] If the utility function is quadratic, we will show that the mean-variance criterion always reflects the true preferences.

Instead of defining the utility function by a term like $y = b_0 + b_1 x + b_2 x^2$ we can without loss of generality normalize the function by choosing $b_0 = 0$ and $b_1 = 1$.[18]

[17] Levy and Sarnat (1972, pp. 310 f.) also provide a generalization of the mean-variance criterion that is always optimal. As this criterion cannot be handled so easily, it is rarely applied and therefore not further considered here.

[18] The concept of expected utility implies that the utility function is only determined up to a positive linear transformation. This allows to apply the transformation $y \to \frac{y-b_0}{b_1}$ to achieve the normalization.

The utility function and its derivatives are therefore given by

$$U(x) = x + bx^2, \tag{C.8}$$

$$U'(x) = 1 + 2bx, \tag{C.9}$$

$$U''(x) = 2b. \tag{C.10}$$

According to (C.5) the Arrow-Pratt measure of risk aversion turns out to be

$$z = -\frac{2b}{1 + 2bE[x]}. \tag{C.11}$$

If we concentrate on risk averse individuals and assume positive marginal utility, (C.11) implies that

$$b < 0. \tag{C.12}$$

But if $b < 0$ we see from (C.9) that the marginal utility is only positive if

$$E[x] < -\frac{1}{2b}. \tag{C.13}$$

For large expected values the marginal utility can become negative. This unreasonable result can only be ruled out if the risk aversion is sufficiently small.[19]

If we define $E[x] = \mu$ and $Var[x] = \sigma^2$ we can write the expected utility as

$$E[U(x)] = E\left[x + bx^2\right] = \mu + bE\left[x^2\right] = \mu + b\left(\mu^2 + \sigma^2\right). \tag{C.14}$$

The indifference curves are obtained by totally differentiating both sides:

$$dE[U(x)] = (1 + 2b\mu)d\mu + 2\sigma d\sigma = 0. \tag{C.15}$$

The slope of the indifference curve in the (μ,σ)-plane is obtained by rearranging (C.15):[20]

$$\frac{d\mu}{d\sigma} = -\frac{2b\sigma}{1 + 2b\mu} = z\sigma > 0, \tag{C.16}$$

[19] This restriction on the expected value is another argument often brought forward against the use of a quadratic utility function and hence the mean-variance criterion. Another argument is that the risk aversion increases with the expected outcome: $\frac{\partial z}{\partial E[x]} = \frac{4b^2}{(1+2bE[x])^2} > 0$, which contradicts empirical findings. Moreover in many theoretical models a constant risk aversion is assumed, which has been shown by PRATT (1964) to imply an exponential utility function. If the expected outcome does not vary too much, constant risk aversion can be approximated by using a quadratic utility function.

[20] Instead of using the variance as a measure of risk, it is more common to use its square root, the standard deviation. As the square root is a monotone transformation, the results are not changed by this manipulation.

i.e. for risk averse investors the indifference curves have a positive slope in the (μ,σ)-plane.

The equation of the indifference curve is obtained by solving (C.14) for μ:

$$\begin{aligned} E\left[U\left(x\right)\right] &= \mu + b\mu^2 + b\sigma^2 \\ \mu^2 + \frac{1}{b}\mu + \sigma^2 &= \frac{E\left[U\left(x\right)\right]}{b} \\ \left(\mu + \frac{1}{2b}\right)^2 + \sigma^2 &= \frac{1}{b}E\left[U\left(x\right)\right] + \frac{1}{4b^2}. \end{aligned} \tag{C.17}$$

Defining $r^* = -\frac{1}{2b}$ as the expected outcome that must not be exceeded for the marginal utility to be positive according to equation (C.13), we can rewrite the equation for the indifference curves as

$$(\mu - r^*) + \sigma^2 = -2r^* E\left[U\left(x\right)\right] + r^{*2} \equiv R^2, \tag{C.18}$$

which is the equation of a circle with center $\mu = r^*$, $\sigma = 0$ and radius R.[21] With this indifference curve, which has as the only parameter a term linked to the risk aversion, it is now possible to determine the optimal alternative out of the efficient alternatives, that is located at the point where the efficient frontier is tangential to the indifference curve. Figure C.4 shows the determination of the optimal alternative C.

We will now show that with a quadratic utility function the mean-variance criterion is optimal.[22] We assume two alternatives with $\mu_i = E\left[a_i\right] > E\left[a_j\right] = \mu_j$. Let further $\sigma_i^2 = Var\left[a_i\right]$ and $\sigma_j^2 = Var\left[a_j\right]$. If $a_i \succ a_j$ it has to be shown that

$$E\left[U\left(a_i\right)\right] > E\left[U\left(a_j\right)\right].$$

Substituting the utility functions gives

$$\mu_i + b\mu_i^2 + b\sigma_i^2 > \mu_j + b\mu_j^2 + b\sigma_j^2,$$

[21] The results that the indifference curves are circles gives rise to another objection against the use of a quadratic utility function. An individual with a quadratic utility function should be indifferent between an expected outcome of $r^* + v$ and $r^* - v$ for any given σ. From the mean-variance criterion (C.7) we know that for a given σ the alternative with the higher expected outcome will be preferred. In practice this problem is overcome by using only the lower right sector of the circle.

[22] Such a proof is given e.g. in Levy and Sarnat (1972, pp. 385 ff.).

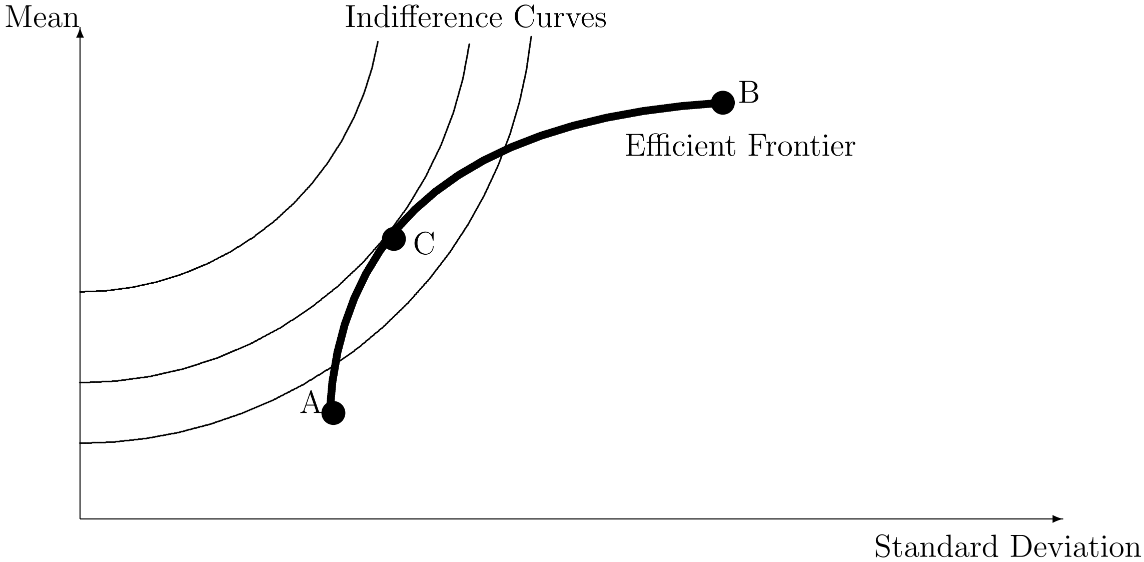


Fig. C.4: Determination of the optimal alternative

$$\mu_i - \mu_j + b(\mu_i^2 - \mu_j^2) + b(\sigma_i^2 - \sigma_j^2) = (\mu_i - \mu_j)\,[1 + b(\mu_i + \mu_j)] + b(\sigma_i^2 - \sigma_j^2) > 0.$$

Dividing by $-2b > 0$ gives us

$$(\mu_i - \mu_j)\left[-\frac{1}{2b} - \frac{\mu_i + \mu_j}{2}\right] - \frac{\sigma_i^2 - \sigma_j^2}{2} > 0. \tag{C.19}$$

From (C.13) we know that $-\frac{1}{2b} > \mu_i$ and $-\frac{1}{2b} > \mu_j$, hence we find that

$$-\frac{1}{2b} > \frac{\mu_i + \mu_j}{2} \tag{C.20}$$

With the assumption that $\mu_i > \mu_j$ and as $b < 0$ the first term in (C.19) is positive. If now $\sigma_i^2 \leq \sigma_j^2$ as proposed by the mean-variance criterion, (C.19) is fulfilled and we have shown that it represents the true preferences.

If $\sigma_i^2 > \sigma_j^2$ in general nothing can be said which alternative will be preferred. For $\mu_i = \mu_j$ we need $\sigma_i^2 < \sigma_j^2$ in order to prefer a_i over a_j. This is exactly the statement made by the mean-variance criterion in (C.13). Therewith it has been shown that in the case of a quadratic utility function the mean-variance criterion is optimal, i.e. represents the true preferences.[23]

[23] A quadratic utility function is not only a sufficient condition for the optimality of the mean-variance criterion, but also a necessary condition. This is known in the literature as the Schneeweiss-Theorem, see SCHNEEWEISS (1967).

C.2.2 The Markowitz frontier

The portfolio selection theory is based on several assumptions:[24]

- no transaction costs and taxes,
- assets are indefinitely divisible,
- each investor can invest into every asset without restrictions,
- investors maximize expected utility by using the mean-variance criterion,
- prices are given and cannot be influenced by investors (competitive prices),
- the model is static, i.e. only a single time period is considered.

Some of these assumptions, like the absence of transaction costs and taxes have been lifted by more recent contributions without giving fundamentally new insights.

In portfolio selection theory the different alternatives to choose between are the compositions of the portfolios, i.e. the weight each asset has.[25] Assume an investor has to choose between $N > 1$ assets, assigning a weight of x_i to each asset. The expected return of each asset is denoted μ_i and the variance of the returns by $\sigma_i^2 > 0$ for all $i = 1, \ldots, N$.[26] The covariances between two assets i and j will be denoted σ_{ij}.

The weights of the assets an investor holds, have to sum up to one and are assumed to be positive as we do not allow for short sales at this stage:

$$\sum_{i=1}^{N} x_i = 1, \qquad \text{(C.21)}$$
$$x_i \geq 0, \; i = 1, \ldots, N.$$

[24] See Lintner (1965a, p. 15).

[25] The decision which portfolio is optimal does not depend on total wealth for a given constant risk aversion, hence it can be analyzed by dealing with weights only.

[26] Instead of investigating final expected wealth and its variance after a given period of time (the time horizon), we can use the expected return and variances of returns as they are a positive linear transformation of the wealth. As has been noted above, the decision is not influenced by such a transformation when using expected utility.

For the moment assume that there are only $N = 2$ assets. The characteristics of each asset can be represented as a point in the (μ,σ)-plane. We then can derive the location of any portfolio in the (μ,σ)-plane by combining these two assets.

The expected return and the variance of the return of the portfolio is given by

$$\mu_p = x_1\mu_1 + x_2\mu_2 = \mu_2 + x_1(\mu_1 - \mu_2), \tag{C.22}$$

$$\begin{aligned} \sigma_p^2 &= x_1^2\sigma_1^2 + x_2^2\sigma_2^2 + 2x_1x_2\sigma_{12} \\ &= \sigma_2^2 + x_1^2(\sigma_1^2 + \sigma_2^2 - 2\sigma_1\sigma_2\rho_{12}) + 2x_1(\sigma_1\sigma_2\rho_{12} - \sigma_2^2), \end{aligned} \tag{C.23}$$

where $\rho_{12} = \frac{\sigma_{12}}{\sigma_1\sigma_2}$ denotes the correlation of the two assets.

The portfolio with the lowest risk is obtained by minimizing (C.23). The first order condition is

$$\frac{\partial\sigma_p^2}{\partial x_1} = 2x_1(\sigma_1^2 + \sigma_2^2 - 2\sigma_1\sigma_2\rho_{12}) + 2(\sigma_1\sigma_2\rho_{12} - \sigma_2^2) = 0.$$

The second order condition for a minimum is fulfilled unless $\sigma_1 = \sigma_2$ and $\rho_{12} \neq 1$:

$$\frac{\partial^2\sigma_p^2}{\partial x_1^2} = 2(\sigma_1^2 + \sigma_2^2 - 2\sigma_1\sigma_2\rho_{12}) > 2(\sigma_1 - \sigma_2)^2 > 0$$

Solving the first order condition gives the weights in the *minimum risk portfolio* (*MRP*):

$$x_1^{MRP} = \frac{\sigma_2^2 - \sigma_1\sigma_2\rho_{12}}{\sigma_1^2 + \sigma_2^2 - 2\sigma_1\sigma_2\rho_{12}}. \tag{C.24}$$

The minimum variance can be obtained by inserting (C.24) into (C.23):

$$\begin{aligned} \sigma_{MRP}^2 &= \sigma_2^2 + \frac{(\sigma_2^2 - \sigma_1\sigma_2\rho_{12})^2}{\sigma_1^2 + \sigma_2^2 - 2\sigma_1\sigma_2\rho_{12}} - 2\frac{(\sigma_2^2 - \sigma_1\sigma_2\rho_{12})^2}{\sigma_1^2 + \sigma_2^2 - 2\sigma_1\sigma_2\rho_{12}} \\ &= \sigma_2^2 - \frac{(\sigma_2^2 - \sigma_1\sigma_2\rho_{12})^2}{\sigma_1^2 + \sigma_2^2 - 2\sigma_1\sigma_2\rho_{12}} \\ &= \frac{\sigma_1^2\sigma_2^2(1 - \rho_{12}^2)}{\sigma_1^2 + \sigma_2^2 - 2\sigma_1\sigma_2\rho_{12}}. \end{aligned} \tag{C.25}$$

If the returns of the two assets are uncorrelated ($\rho_{12} = 0$), then (C.25) reduces to

$$\sigma_{MRP}^2 = \frac{\sigma_1^2\sigma_2^2}{\sigma_1^2 + \sigma_2^2}. \tag{C.26}$$

This variance is smaller than the variance of any of these two assets.[27] By holding an appropriate portfolio, the variance, and hence the risk, can be reduced, whereas the expected return lies between the expected returns of the two assets.

With perfectly negative correlated assets ($\rho_{12} = -1$) we find that

$$\sigma^2_{MRP} = 0 \tag{C.27}$$

and the risk can be eliminated from the portfolio.

In the case of perfectly correlated assets ($\rho_{12} = 1$) the minimum variance is the variance of the asset with the lower variance:

$$\sigma^2_{MRP} = \begin{cases} \sigma_1^2 & \text{if} \quad \sigma_1^2 \leq \sigma_2^2 \\ \sigma_2^2 & \text{if} \quad \sigma_1^2 > \sigma_2^2 \end{cases} . \tag{C.28}$$

We can derive a general expression for the mean-variance relation:

$$\begin{aligned} \sigma_p^2 - \sigma^2_{MRP} &= \sigma_2^2 + x_1^2(\sigma_1^2 + \sigma_2^2 - 2\sigma_1\sigma_2\rho_{12}) + 2x_1(\sigma_1\sigma_2\rho_{12} - \sigma_2^2) \\ &\quad - \frac{\sigma_1^2\sigma_2^2(1-\rho_{12}^2)}{\sigma_1^2 + \sigma_2^2 - 2\sigma_1\sigma_2\rho_{12}} \\ &= (x_1 - x_1^{MRP})^2(\sigma_1^2 + \sigma_2^2 - 2\sigma_1\sigma_2\rho_{12}). \end{aligned} \tag{C.29}$$

With μ_{MRP} denoting the expected return of the minimum risk portfolio, we find that

$$\mu_p - \mu_{MRP} = (x_1 - x_1^{MRP})(\mu_1 - \mu_2). \tag{C.30}$$

Solving (C.30) for $x_1 - x_1^{MRP}$ and inserting into (C.29) we obtain after rearranging:

$$(\mu_p - \mu_{MPR})^2 = \frac{\sigma_p^2 - \sigma^2_{MRP}}{\sigma_1^2 + \sigma_2^2 - 2\sigma_1\sigma_2\rho_{12}}(\mu_1 - \mu_2)^2. \tag{C.31}$$

This equation represents a hyperbola with axes

$$\begin{aligned} \mu_p &= \mu_{MRP} + \frac{\mu_1 - \mu_2}{\sqrt{\sigma_1^2 + \sigma_2^2 - 2\sigma_1\sigma_2\rho_{12}}}\sigma_p, \\ \mu_p &= \mu_{MRP} - \frac{\mu_1 - \mu_2}{\sqrt{\sigma_1^2 + \sigma_2^2 - 2\sigma_1\sigma_2\rho_{12}}}\sigma_p. \end{aligned}$$

[27] Suppose $\sigma^2_{MRP} = \frac{\sigma_1^2\sigma_2^2}{\sigma_1^2+\sigma_2^2} > \sigma_2^2$, this would imply that $\sigma_1^2 > \sigma_1^2 + \sigma_2^2$ and hence $\sigma_2^2 < 0$, which contradicts the assumption that $\sigma_2^2 > 0$. A similar argument can be used to show that $\sigma^2_{MRP} < \sigma_1^2$.

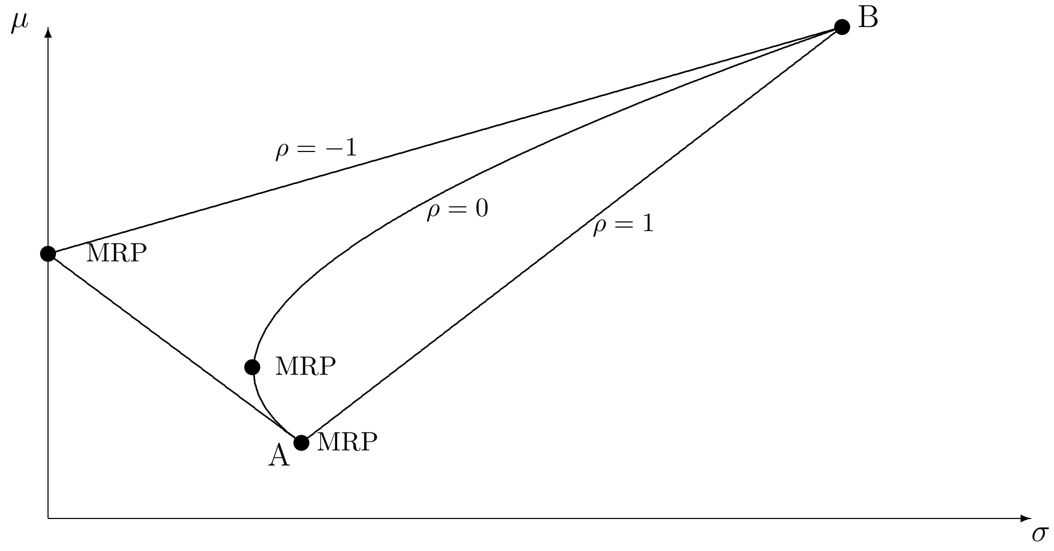


Fig. C.5: Efficient portfolios with two assets

The efficient portfolios lie on the upper branch of this hyperbola, i.e. above the minimum risk portfolio.[28] Figure C.5 shows the efficient portfolios for different correlations. It can easily be shown that in the case of perfect positive correlation the efficient portfolios are located on a straight line connecting the two assets, in case of perfectly negative correlation on straight lines connecting the assets with the minimum risk portfolio. Between efficient portfolios can only be distinguished by using the utility function. Figure C.6 adds the indifference curve to the opportunity locus and determines the location of the *optimal portfolio* (OP). The location of the optimal portfolio depends on the risk aversion of the investor, the more risk averse the investor is the more close the optimal portfolio will be located to the minimum risk portfolio.

It is now possible to introduce a third asset. In a similar way hyperbolas can be deducted representing all combinations of this asset with one of the other two. Furthermore we can view any portfolio consisting of the two other assets as a single new asset and can combine it in the same manner with the third asset. Figure C.7 illustrates this situation. All achievable portfolio combinations are

[28] The efficient frontier is also called *opportunity locus*.

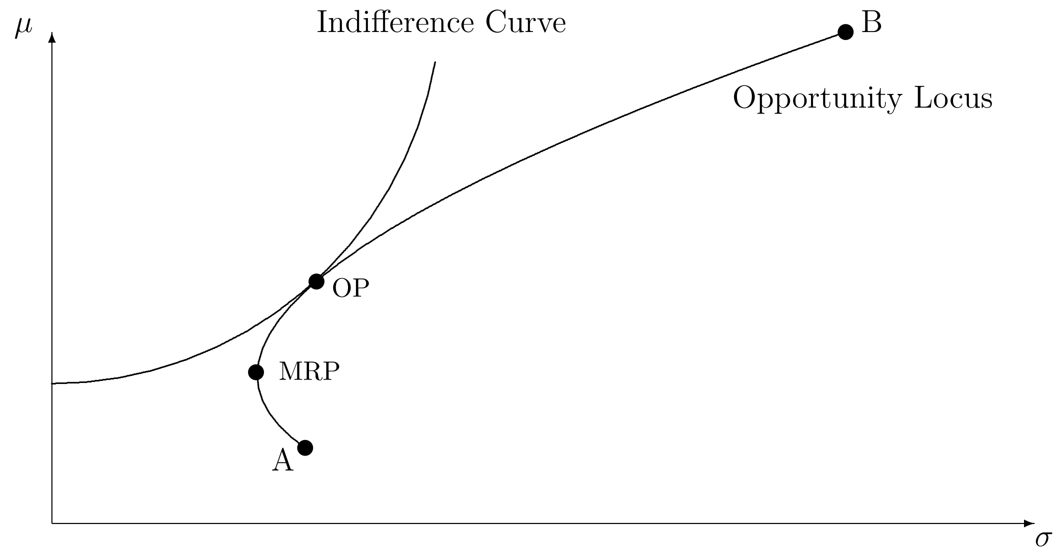


Fig. C.6: Determination of the optimal portfolio with two assets

now located in the area bordered by the bold line connecting points A and C, where the bold line encircling the different hyperbolas is the new opportunity locus.

This concept can be generalized to $N > 3$ assets in the same manner. All achievable assets will be located in an area and the efficient frontier will be a hyperbola. Using the utility function the optimal portfolio can be determined in a similar way as in the case of two assets as shown in figure C.8. If an asset is added, the area of achievable portfolios is enlarged and encompasses the initial area. This can simply be shown by stating that the new achievable portfolios encompass also the portfolios assigning a weight of zero to the new asset. With a weight of zero these portfolios are identical to the initially achievable portfolios. To these portfolios those have to be added assigning a non-zero weight to the new asset. Therefore the efficient frontier moves further outward to the upper left. By adding new assets the utility can be increased.

Thus far it has been assumed that $\sigma_i^2 > 0$, i.e. all assets were risky. It is also possible to introduce a riskless asset, e.g. a government bond, with a variance of zero. Define the return of the riskless asset by r, then in the case of two assets

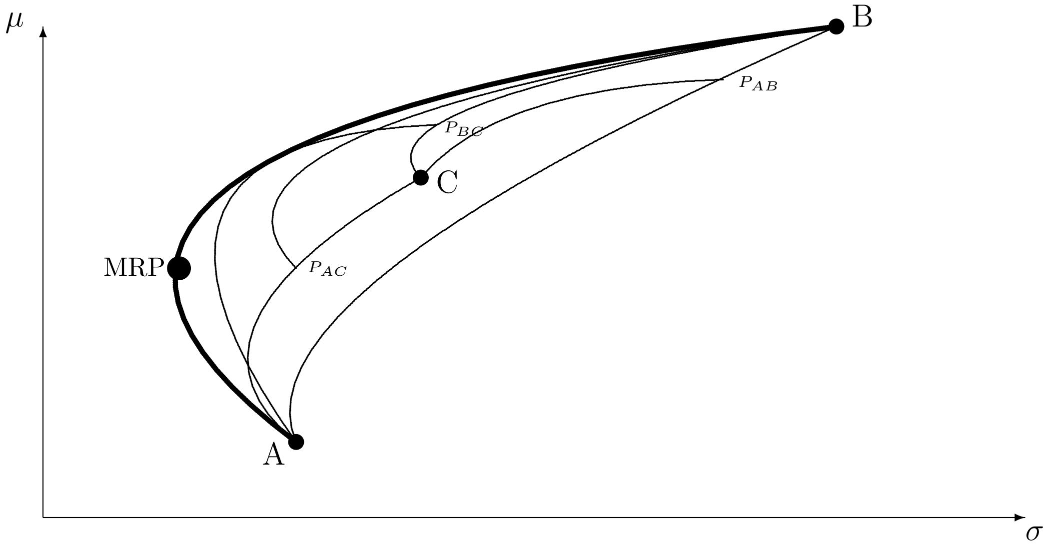


Fig. C.7: Portfolio selection with three assets

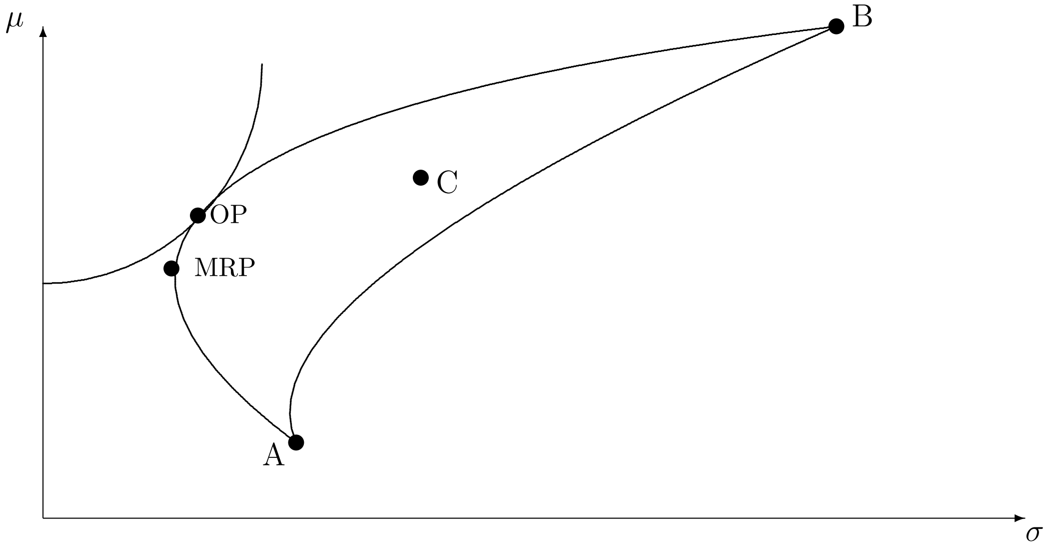


Fig. C.8: The optimal portfolio with $N > 2$ assets we get from (C.22) and (C.23):

$$\mu_p = x_1\mu_1 + x_2 r = r + x_1(\mu_1 - r), \tag{C.32}$$

$$\sigma_p^2 = x_1^2\sigma_1^2. \tag{C.33}$$

Solving (C.33) for x_1 and inserting into (C.32) gives

$$\mu_p = r + \frac{\sigma_p}{\sigma_1}(\mu_1 - r) = r + \frac{\mu_1 - r}{\sigma_1}\sigma_p. \tag{C.34}$$

The expected return of the portfolio is linear in the variance of the portfolio return, i.e. the hyperbola reduces to a straight line from the location of the riskless asset, $(0, r)$, to the location of the asset. In the case of many risky assets we can combine every portfolio of risky assets with the riskless asset and obtain all achievable portfolios. As shown in figure C.9 all achievable portfolios are located between the two straight lines, the upper representing the efficient frontier. There exists a portfolio consisting only of risky assets that is located on the efficient frontier. It is the portfolio consisting only of the risky assets at which the efficient frontiers with and without a riskless asset are tangential.[29] This portfolio is called the *optimal risky portfolio* (*ORP*). The efficient frontier with a riskless asset is also called the *capital market line*.

All efficient portfolios are located on the capital market line, consequently they are a combination of the riskless asset and the optimal risky portfolio. The optimal portfolio can be obtained in the usual way by introducing the indifference curves. As the optimal portfolio always is located on the capital market line, it consists of the risky asset and the optimal risky portfolio. Which weight is assigned to each depends on the risk aversion of the investor, the more risk averse he is the more weight he will put on the riskless asset. The weights of the optimal risky portfolio do not depend on the risk aversion of the investor. The decision process can therefore be separated into two steps, the determination of the optimal risky portfolio and then the determination of the optimal portfolio

[29] It is also possible that no tangential point exists, in this case a boundary solution exists and the risky portfolio consists only of a single risky asset.

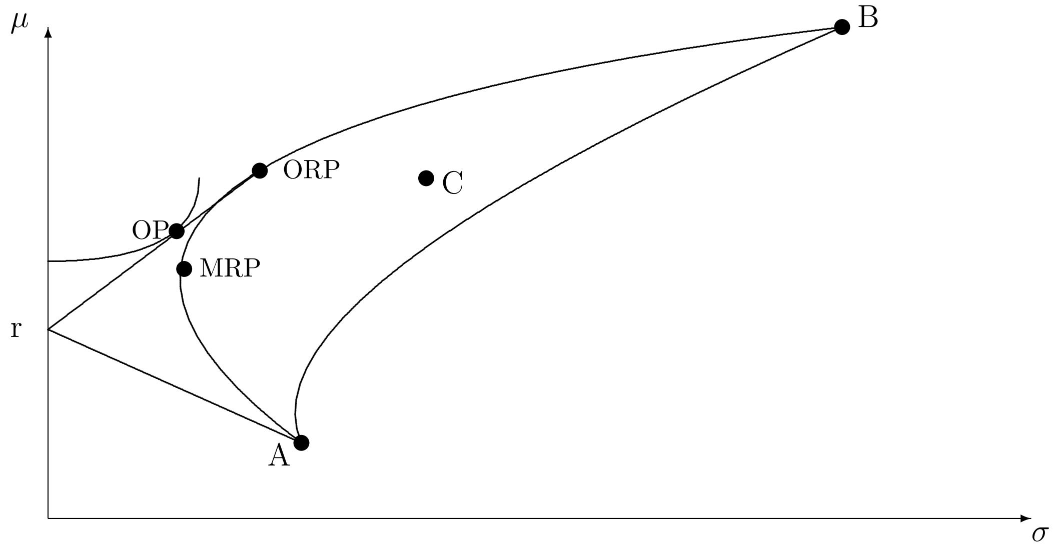


Fig. C.9: The optimal portfolio with a riskless asset

as a combination of the ORP with the riskless asset. As this result has first been presented by TOBIN (1958) it is also called the *Tobin separation theorem*.[30]

So far we have assumed that $x_i \geq 0$ for all $i = 1, \ldots, N$. If we allow now some x_i to be negative, the possibilities to form portfolios is extended. An asset with $x_i < 0$ means that the asset is sold short, i.e. it is sold without having owned it before. This situation can be viewed as a credit that has not been given and has not to be repaid in money (unless money is the asset), but in the asset. The assets can be the riskless asset or the risky assets, in the former case the short sale is an ordinary credit. It is assumed that credits can be obtained at the same conditions (interest rate or expected return and risk) as investing in the asset.

By allowing short sales the efficient frontier of the risky portfolios further moves to the upper left as new possibilities to form portfolios are added by lifting the restriction that the weights must be non-negative. Therewith the capital market line becomes steeper and the utility of the optimal portfolio increases.

[30] For investors being less risk averse it is possible that the optimal portfolio is located on the part of the efficient frontier above the ORP, in this case the optimal portfolio does not contain the riskless asset and assigns different weights to the risky assets compared to the ORP. Therefore in general the Tobin separation theorem does only hold with the inclusion of short sales, as described in the next paragraph.

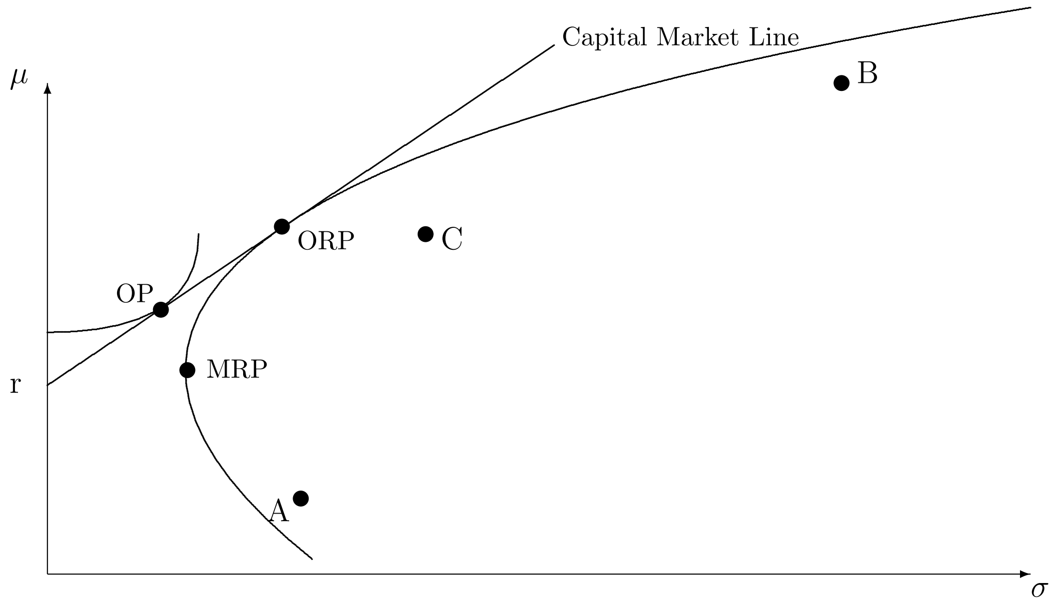


Fig. C.10: Portfolio selection with short sales

Figure C.10 illustrates this case. The Capital Market Line extends beyond the ORP and therewith the optimal portfolio will always be a combination of the riskless asset and the ORP. The Tobin separation theorem applies in all cases, independent of the degree of risk aversion. If the ORP is located above the ORP, the riskless asset is sold short and a larger fraction of the optimal portfolio consists of the ORP.

In applying the portfolio theory to determine the optimal portfolio several problems are faced:

- determination of the risk aversion of the investor,
- determination of the expected returns, variances and covariances of the assets,
- computation of the efficient frontier and the optimal portfolio.

There exists no objective way to determine the risk aversion of an investor, most investors are only able to give a qualitative measure of their risk aversion, if at all. The transformation into a quantitative measure is an unsolved, but for the determination of the optimal portfolio critical problem. It is important for the allocation between the riskless asset and the optimal risky portfolio.

Expected returns, variances and covariances can be obtained from estimates based on past data. But there is no guarantee that these results are reasonable for the future. It is also possible to use other methods to determine these moments, e.g. by using subjective beliefs. The determination of these moments are critical for the determination of the optimal risky portfolio.

To determine the efficient frontier and the optimal portfolio non-trivial numerical optimization routines have to be applied.[31] Advances in computer facilities and the availability of these routines do not impose a threat anymore as it has done in former years.

When having solved the above mentioned problems, the portfolio theory does allow to answer the questions raised at the beginning of this appendix:

- the share to be invested into risky assets is determined by the optimal portfolio,
- the assets to invest in are those included in the optimal risky portfolio,
- the shares to invest in each selected asset are given by the weights of the optimal risky portfolio.

The portfolio theory has developed a method how to allocate resources optimal. Although mostly only financial assets are included, other assets like human capital, real estate and others can easily be included, although it is even more difficult to determine their characteristics.

A shortcoming of the portfolio theory is that it is a static model. It determines the optimal portfolio at a given date. If the time horizon is longer than one period, the prices of assets change over time, and therewith the weights of the assets in the initial portfolio change. Even if the expected returns, variances and covariances do not change, this requires to rebalance the portfolio every period.

[31] For a detailed description of the mathematical concepts to solve these problems see Markowitz (1959) and Aschinger (1990).

As assets with a high realized return enlarge their weight, they have partially to be sold to buy assets which had a low return (sell the winners, buy the losers). In a dynamic model other strategies have been shown to achieve a higher expected utility for investors, but due to the static nature of the model such strategies cannot be included in this framework.

C.3 The Capital Asset Pricing Model

In this section we will derive the *Capital Asset Pricing Model* (*CAPM*) that is the most prominent model in asset pricing. Sharpe (1964) and Lintner (1965a) developed the CAPM independent of each other by using portfolio theory to establish a market equilibrium.

The basis of the CAPM is portfolio theory with a riskless asset and unlimited short sales. We do not consider only the decision of a single investor, but aggregate them to determine a market equilibrium. In portfolio theory the price of an asset was exogeneously given and could not be influenced by any investor. Given this price he formed his beliefs on the probability distribution.[32] The beliefs were allowed to vary between investors.

In the CAOM asset prices (or equivalently expected asset returns) will no longer be exogenously given, but be an equilibrium of the market. From basic finance theory we know that the current price affects the expected returns and vice versa. Given future expected dividends and assuming that markets are efficient, i.e. that the prices of assets equal their fundamental value, a high current price results in a low expected return in the next period and a low current price in a high expected return. In the same way in order to expect a high return, the price has to be low and for a low expected return a high price is needed. This equivalence of price and return allows us to concentrate either on prices or on expected returns. Convention in the academic literature requires us to focus on expected returns.

[32] As portfolio theory makes use of the mean-variance criterion it is sufficient to form beliefs only about means, variances and covariances instead of the entire distribution.

Additionally to the assumptions already stated for portfolio theory, we have to add that all investors have the same beliefs on the probability distribution of all assets, i.e. agree on the expected returns, variances and covariances.[33] If all investors agree on the characteristics of an asset the optimal risky portfolio will be equal for all investors, even if they differ in their preferences (risk aversion). Because all assets have to be held by the investors the share each asset has in the optimal risky portfolio has to be equal to its share of the market value of all assets. The optimal risky portfolio has to be the market portfolio. Moreover all assets have to be marketable, i.e. all assets must be traded and there are no other investment opportunities not included into the model.

Every investor j ($j = 1, \ldots, M$) maximizes his expected utility by choosing an optimal portfolio, i.e. choosing optimal weights for each asset:

$$
\begin{aligned}
\text{(C.35)} \qquad \max_{\{x_i\}_{i=1}^N} E\left[U^j(R_p)\right] &= \max_{\{x_i\}_{i=1}^N} U^j\left(\mu_p - \frac{1}{2} z_j \sigma_p^2\right) \\
&= \max_{\{x_i\}_{i=1}^N} \left(\mu_p - \frac{1}{2} z_j \sigma_p^2\right) \\
&= \max_{\{x_i\}_{i=1}^N} \left(\sum_{i=1}^N x_i \mu_i - \frac{1}{2} z_j \sum_{k=1}^N \sum_{i=1}^N x_i x_k \sigma_{ik}\right)
\end{aligned}
$$

for all $j = 1, \ldots, M$ with the restriction $\sum_{i=1}^N x_i = 1$.

The Lagrange function for solving this problem can easily be obtained as

$$
\text{(C.36)} \qquad L_j = \sum_{i=1}^N x_i \mu_i - \frac{1}{2} z_j \sum_{k=1}^N \sum_{i=1}^N x_i x_k \sigma_{ik} + \lambda \left(1 - \sum_{i=1}^N x_i\right).
$$

The first order conditions for a maximum are given by

$$
\begin{aligned}
\text{(C.37)} \qquad \frac{\partial L_j}{\partial x_i} &= \mu_i - z_j \sum_{k=1}^N x_k \sigma_{ik} - \lambda = 0, \qquad i = 1, \ldots, N, \\
\text{(C.38)} \qquad \frac{\partial L_j}{\partial \lambda} &= 1 - \sum_{i=1}^N x_i = 0
\end{aligned}
$$

[33] Lintner (1965b) calls this assumption *idealized uncertainty*. Sharpe (1970, pp. 104 ff.) also considers different beliefs. As the main line of argument does not change, these complications are not further considered here. Several assumptions made in portfolio theory can also be lifted without changing the results significantly. Black (1972) restricts short sales and Sharpe (1970, pp. 110 ff.) applies different interest rates for borrowing and lending the riskless asset.

for all $j = 1, \dots, M$.[34]

Solving the above equations for μ_i gives

$$\text{(C.39)} \qquad \begin{aligned} \mu_i &= \lambda + a_j \sum_{k=1}^{N} x_k \sigma_{ik} = \lambda + z_j Cov\left[R_i, \sum_{k=1}^{N} x_k R_k\right] \\ &= \lambda + z_j Cov\left[R_i, R_p\right] = \lambda + z_j \sigma_{ip}. \end{aligned}$$

With $\sigma_{ip} = 0$ we find that $\mu_i = \lambda$, hence we can interpret λ as the expected return of an asset which is uncorrelated with the market portfolio. As the riskless asset is uncorrelated with any portfolio, we can interpret λ as the risk free rate of return r:

$$\text{(C.40)} \qquad \mu_i = r + z_j \sigma_{ip}.$$

From (C.40) we see that the expected return depends linearly on the covariance of the asset with the market portfolio. The covariance can be interpreted as a measure of risk for an individual asset (*covariance risk*). Initially we used the variance as a measure of risk, but as has been shown in the last section the risk of an individual asset can be reduced by holding a portfolio. The risk that cannot be reduced further by diversification is called *systematic risk*, whereas the diversifiable risk is called *unsystematic risk*. The total risk of an asset consists of the variation of the market as a whole (systematic risk) and an asset specific risk (unsystematic risk). As the unsystematic risk can be avoided by diversification it is not compensated by the market, efficient portfolios therefore only have systematic and no unsystematic risk.

The covariance of an asset can be also interpreted as the part of the systematic risk that arises from an individual asset:

$$\begin{aligned} \sum_{i=1}^{N} x_i \sigma_{ip} &= \sum_{i=1}^{N} x_i Cov[R_i, R_p] = Cov\left[\sum_{i=1}^{N} x_i R_i, R_p\right] \\ &= Cov[R_p, R_p] = Var[R_p] = \sigma_p^2. \end{aligned}$$

Equation (C.40) is valid for all assets and hence for any portfolio, as the equation for a portfolio can be obtained by multiplying with the appropriate weights and

[34] The second order condition for a maximum can be shown to be fulfilled, due to space limitations this proof is not presented here.

then summing them up, so that we can apply this equation also to the market portfolio, which is also the optimal risky portfolio:

$$\mu_p = r + z_j \sigma_p^2. \tag{C.41}$$

Solving for z_j and inserting into (C.40) gives us the usual formulation of the CAPM:

$$\mu_i = r + (\mu_p - r)\frac{\sigma_{ip}}{\sigma_p^2}. \tag{C.42}$$

Defining $\beta_i = \frac{\sigma_{ip}}{\sigma_p^2}$ we can rewrite (C.42) as

$$\mu_i = r + (\mu_p - r)\beta_i.^{35} \tag{C.43}$$

β_i represents the relative risk of the asset i (σ_{ip}) to the market risk (σ_p^2). The beta for the market portfolio is easily shown to be 1. We find a linear relation between the expected return and the relative risk of an asset. This relation is independent of the preferences of the investors (z_j), provided that the mean-variance criterion is applied and that the utility function is quadratic. This equilibrium line is called the *Capital Market Line* (*CML*). Figure C.11 illustrates this relation.

For the risk an investor takes he is compensated by the amount of $\mu_p - r$ per unit of risk, the total amount $(\mu_p - r)\beta_i$ is called the *risk premium* or the *market price of risk*. The risk free rate of return r may be interpreted as the price for time. It is the compensation for not consuming the amount in the current period, but wait until the next period.[36]

Equation (C.43) presents a formula for the expected return given the interest rate, r, the beta and the expected return on the market portfolio, μ_p. However, the expected return of the market portfolio is not exogenous as it is a weighted average of the expected returns of the individual assets. In this formulation only relative expected returns can be determined, the level of expected returns, i.e. the market risk premium, is not determined by the CAPM. Although we can

[35] It also can frequently be found that only excess returns over the risk free rate of return are considered. by defining $\mu_j' = \mu_j - r$ (C.43) becomes $\mu_i' = \mu_p' \beta_i$.

[36] This in opposition to the interpretation of KEYNES (1936, pp. 165) who interpreted the interest rate as a compensation for giving up liquidity, i.e. binding resources over time.

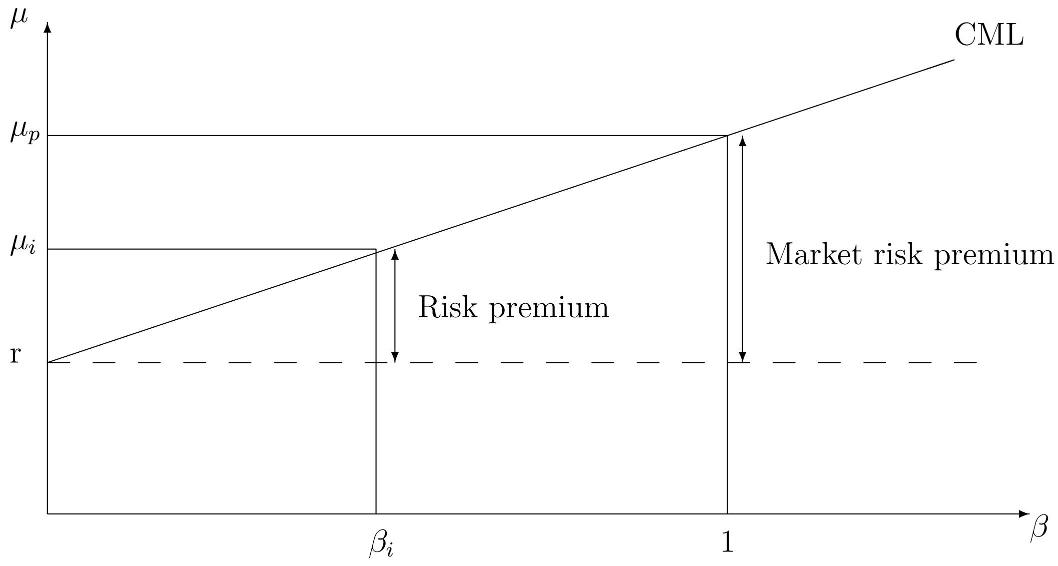


Fig. C.11: The Capital Asset Pricing Model

reasonably assume μ_p to be given when investigating a small capitalized asset, we have to determine μ_p endogenously.

Thus far we only considered the market portfolio, but an investor will in general not hold the market portfolio (optimal risky portfolio). As we saw in section C.2 the optimal portfolio will be a combination of the market portfolio and the riskless asset, where the shares will vary among investors. We can now add another restriction to our model. The return on the optimal risky portfolio has to be such that the market for the riskless assets has to be in equilibrium. The amounts of riskless assets lent and borrowed have to be equal:[37]

$$\sum_{j=1}^{M} x_{jr} = 0, \tag{C.44}$$

where x_{jr} denotes the demand of the jth investor for the riskless asset. With this additional market to be in equilibrium it is possible to determine μ_p endogenously, whiose value will depend on the risk aversion of investors. We herewith have found an equilibrium in the expected returns.

Nevertheless this result remains to determine only relative prices. The risk free

[37] Alternatively a fixed supply of the riskless asset can be assumed without changing the argument.

rate r is not determined endogenously. Although is can reasonably be assumed that r can substantially be influenced by monetary policy, especially for longer time horizons it is not given.

C.4 The Rational Expectations Approach

This appendix introduces the concept of rational expectations. The literature on this subject has grown rapidly since it was first introduced by MUTH (1961). Therefore here only the basic concept is presented, it is not the aim to give an overview of the entire theory. The first applications of rational expectations could be found in macroeconomics in the 1970's, most prominently by the Nobel laureate Robert E. Lucas. Nowadays rational expectations are used in nearly every field of economics, especially in financial markets theory.[38]

Using his individual information, every investor will form his expectations on a certain economic variable which is not known with certainty, e.g. the fundamental value of an asset and act on this individual expectation. These expectations and actions transform into prices, where the price realized may be different from the expectations of all investors. In order to analyze how expectations and the therewith associated actions transform into prices it is necessary to model the way expectations are formed.

When introducing *rational expectations* MUTH (1961, p. 316) proposed that expectations

> "... are essentially the same as the predictions of the relevant economic theory."

He explains this definition of rational expectations by stating that these expectations tend to be distributed for a given information set about the distribution of the theory, i.e. on average the mean of the expectations equals the forecast of the theory. This implicitly states that all investors agree on the theory, i.e. on the structure of the economy as well as on the parameters. The assumption that

[38] SHEFFRIN (1996) presents several examples of using rational expectations in different fields of economics.

rational expectations produce no systematic errors, implies that investors do not only have to agree on the theory, but that the theory also has to be true.

Rational expectations are represented by conditional expectations. Let X_t be the variable to be forecasted and Ω_{t-1} the information available, then with X_t^e representing the expectation of X_t we define rational expectations as

$$X_t^e = E\left[X_t | \Omega_{t-1}\right]. \tag{C.45}$$

That there exists no systematic error in these expectations gives

$$E\left[X_t - X_t^e | \Omega_{t-1}\right] = 0, \tag{C.46}$$

$$E\left[(X_t - X_t^e) X_t | \Omega_{t-1}\right] = 0, \tag{C.47}$$

i.e. the expected forecast error is zero and uncorrelated with the true value of the variable. Many examples show that this in general only holds if the theory is correct.

A *narrow version* of the definition given by MUTH (1961) has been presented by LUCAS AND PRESCOTT (1971). They require not only the mean of the expectations and the predictions of the theory to coincide, but they must have the same probability distribution.

A *weak version* assumes that information is costly and therefore every individual only acquires information until his marginal costs and benefits equal. Only this information acquired is used to form rational expectations, not all available information. It also allows for rational expectations that are not based on the correct theory. Investors then form their expectations "as if" they knew the correct model. These expectations in general will be biased, i.e. not fulfill equations (C.46) and (C.47).[39]

Rational expectations have much been criticized for the very restrictive assumptions, especially in the narrow version. The assumption that all individuals know the true model of the economy has been the main target, but also the assumption that all available information has to be taken into account. The weak

[39] PESARAN (1987, p.23) also points out that in such situations self-fulfilling prophecies can occur until the model is identified to be wrong and a sudden change in the model applied may correct the situation.

version has also been attacked as by its definition it allows to define every form of expectations to be called rational. Especially how investors could learn the true model of the economy remains an unsolved problem.

A more general criticism on rational choice theory, and therewith rational expectations, in most cases addresses the fact that individuals are not only rational but in many situations behave emotional or imitate others. These behavioral approaches have attracted increased attention in recent years, especially in finance, where they are used to explain several of features of asset prices that cannot be explained using rational behavior.[40] In most cases models using rational choice make use of very restrictive assumptions to be able to derive results by using sophisticated mathematical methods.[41] These assumptions make predictions of behavior in actual markets very difficult and in many cases they fail. Here BEED AND BEED (2000) question the usefulness of rational choice theory and its contribution to the advancement of economic knowledge in general. However, at present there has been developed no more powerful tool to address economic problems.

By allowing investors to learn some aspects of the economy, the narrow version can be weakened partly. We can allow investors to learn some parameters of the economy, e.g. the beta in the CAPM. If we assume that all investors know the structure of the economy, but not the parameters, it can be shown that they will learn the true parameters over time, i.e. their expectations converge to the narrow version. The most widely used concept of learning is Bayesian learning to be presented in the next section.

Despite these modifications it remains an unsolved problem how investors learn the true structure of the economy. As no other theory on expectation formation exists that provides a better explanation of economic phenomena, rational expectations are widely used in economics.

[40] See THALER (1993), SHLEIFER (2000) or HIRSCHLEIFER (2001) for an overview.

[41] Another objective against the current economic theory and its dominating rational choice theory is the concentration on self defined and very abstract problems rather than on "real world problems" that concern the society as FREY (2000, 25 f.) points out.

C.5 Bayesian learning

We assume investors to know the structure of the economy, but not the parameters in the model. In a first step they form beliefs about these parameters, e.g. by assigning a random number or another reasoning. Every investor can have different beliefs. Based on these first beliefs they form their expectations and act accordingly. When they see the realization of the outcome, they realize that their expectations were not correct. As they know the structure of the economy, the only source of these deviations can be the values of the parameters assigned. Hence they want to change their beliefs about these parameters. This process continues until expectations and realizations coincide and the beliefs are correct.

If there are random variables in the economy, the most widely used method to model changes in beliefs uses Bayes rule and hence is called *Bayesian learning*.[42]

We know from probability theory that for discrete random variables with Prob denoting the probability:[43]

$$\text{(C.48)}\ \text{Prob}[\Omega_t] = \text{Prob}[\Omega_t|a = a_t]\text{Prob}[a = a_t] + \text{Prob}[\Omega_t|a \neq a_t]\text{Prob}[a \neq a_t],$$

$$\text{(C.49)}\ \text{Prob}[a = a_t|\Omega_t] = \frac{\text{Prob}[a = a_t, \Omega_t]}{\text{Prob}[\Omega_t]},$$

$$\text{(C.50)}\ \text{Prob}[\Omega_t|a = a_t] = \frac{\text{Prob}[a = a_t, \Omega_t]}{\text{Prob}[a = a_t]},$$

where a is the true parameter and a_t the current belief of the value of this parameter. Combining relations (C.48) - (C.50) gives

(C.51)

$$\text{Prob}[a = a_t|\Omega_t] = \frac{\text{Prob}[\Omega_t|a = a_t]\text{Prob}[a = a_t]}{\text{Prob}[\Omega_t|a = a_t]\text{Prob}[a = a_t] + \text{Prob}[\Omega_t|a \neq a_t]\text{Prob}[a \neq a_t]},$$

which is also known as *Bayes' rule* or *Bayes' theorem.* The probability that $a = a_t$ in the last period, $\text{Prob}[a = a_t]$, is called the *prior belief.* On this belief the investor based his decision. Given this belief for all possible values the entire

[42] If no random variables are present, other models of learning have to be used. In some cases it may be possible to solve the equations directly for the parameters and obtain the true values in a single step. As in most models random variables are incorporated we concentrate on this case here.

[43] For continuous random variables the argument does not change. Instead of probabilities densities have to be used.

distribution is known. In the next period he learns the realization of the process, that forms part of his new information set, Ω_t. Based on this new information he changes his belief that $a = a_t$ to $\text{Prob}[a = a_t|\Omega_t]$ according to equation (C.51), his *posterior belief*. For the next period these beliefs are his prior beliefs, which he updates on the new information received in the next period. Applying this relation for all possible a_t we receive the distribution and can calculate the relevant parameters, e.g. mean and variance.

By using Bayes' rule to change beliefs it can be shown that the beliefs converge to the true values of the parameters. Hence expectations converge to rational expectations in the narrow sense. This property makes Bayesian learning attractive to use in rational expectation models. What remains an unsolved problem is how to learn the true structure of the economy.

Bibliography

Acharya, V. V. and Pedersen, L. H. (2004): Asset pricing with liquidity risk. Mimeo, London Business School.

Admati, A. R. and Pfleiderer, P. (1988): A Theory of Intraday Patterns: Volume and Price Variability. In: Review of Financial Studies, 1, 3–40.

Admati, A. R. and Pfleiderer, P. (1989): Divide and Conquer: A Theory of Intraday and Day-of-the-Week Mean Effects. In: Review of Financial Studies, 2, 189–223.

Affleck-Graves, J., Hedge, S. P. and Miller, R. E. (1994): Trading Mechanisms and the Components of the Bid-Ask Spread. In: Journal of Finance, 49, 1471–1488.

Amihud, Y. and Mendelson, H. (1980): Dealership Market: Market Making with Inventory. In: Journal of Financial Economics, 8, 31–53.

Amihud, Y. and Mendelson, H. (1986a): Asset Pricing and the Bid-Ask Spread. In: Journal of Financial Economics, 17, 223–249.

Amihud, Y. and Mendelson, H. (1986b): Liquidity and stock returns. In: Financial Analysts Journal, 43–48.

Amihud, Y. and Mendelson, H. (1988): Liquidity and asset prices: Financial management implications. In: Financial Management, 5–15.

Amihud, Y. and Mendelson, H. (1989): The effect of beta, bid-ask spread, residual risk and size on stock returns. In: Journal of Finance, 44, 479–486.

Amihud, Y. and Mendelson, H. (1991): Liquidity, asset prices and financial policy. In: Financial Analysts Journal, 56–66.

Angel, J. J. (1997): How Best to Supply Liquidity to a Small-Capitalization Securities Market. Mimeo, Georgetown University.

Angel, J. J. and Weaver, D. G. (1998): Priority Rules !. Mimeo, Georgetown University.

Arnold, T., Hersch, P., Mulherin, J. H. and Netter, J. (1999): Merging markets. In: Journal of Finance, 54, 1083–1107.

Arrow, K. J. (1963): The Theory of Risk Aversion, chapter 9, 147–171. In: Arrow (1984).

Arrow, K. J. (1964): The Role of Securities in the Optimal Allocation of Risk-bearing. In: Review of Economic Studies, 31, 91–96.

Arrow, K. J. (1984): Collected Papers of Kenneth J. Arrow: Individual Choice under Certainty and Uncertainty. Oxford: Basil Blackwell.

Aschinger, G. (1990): General Remarks on Portfolio Theory and its Application. Working Paper Nr. 158, Insitute for Economic and Social Sciences, University of Fribourg (Switzerland).

Atkins, A. B. and Dyl, E. A. (1990): Price reversals, bid-ask spreads, and market efficiency. In: Journal of Financial and Quantitative Analysis, 25, 535–547.

Bacidore, J. and Lipson, M. L. (2001): The effects of opening and closing procedures on the nyse and nasdaq. Mimeo, University of Georgia.

Bagehot, W. (1971): The Only Game in Town. In: Financial Analysts Journal, 27, 12–14, 22.

Baker, H. K. and Edelman, R. B. (1992): Amex-tp-nyse transfers, market microstructure, and shareholder wealth. In: Financial Management, 60–72.

Barclay, M. J., Kandel, E. and Marx, L. M. (1998): The effects of transaction costs on stock prices and trading volume. In: Journal of Financial Intermediation, 7, 130150.

Baruch, S. and Saar, G. (2004): Asset returns and the listing choice of firms. Mimeo, University of Utah.

Baumol, W. J., Panzar, J. C. and Willig, R. D. (1988): Contestable Markets and the Theory of Industry Structure. Revised edition, San Diego, CA: Harcourt Brace Jovanovich Publishers.

Beed, C. and Beed, C. (2000): Intellectual Progress and Academic Economics: Rational Choice and Game Theory. In: Journal of Postkeynesian Economics, 23, 163–185.

Benveniste, L., Cappozza, D. R. and Seguin, P. J. (2001): The value of liquidity. In: Real Estate Economics, 29, 633–660.

Bernhardt, D. and Hughson, E. (1997): Splitting Orders. In: Review of Financial Studies, 10, 69–101.

Biais, B. (1993): Price Formation and Equilibrium Liquidity in Fragmented and Centralized Markets. In: Journal of Finance, 58, 157–185.

Biais, B., Foucault, T. and Salani, F. (1995): Implicit Collusion on Wide Spreads. Working Paper No. 153, Finance and Banking Discussion Paper Series, Universitat Pompeu Fabra.

Biais, B., Hillion, P. and Spatt, C. (1997): An empirical analysis of the limit order book and the orde flow on the paris bourse. In: Journal of Finance, 50, 1655–1689.

Black, F. (1972): Capital Market Equilibrium with Restricted Borrowing. In: Journal of Business, 45, 444–455.

Bloomfield, R. and O'Hara, M. (1998): Does order preferencing matter?. In: Journal of Financial Economics, 50, 3–37.

Blume, L., Easley, D. and O'Hara, M. (1994): Market Statistics and Technical Analysis: The Role of Volume. In: Journal of Finance, 49, 153–181.

Board, J. and Sutcliffe, C. (2000): The proof of the pudding: The effects of increased trade transparency in the london stock exchange. In: Journal of Business Finance & Accounting, 27, 887–909.

Bondarenko, O. (1999): Competing Market Makers, Liquidity Provision, and Bid-Ask Spread. Mimeo, University of Illinois at Chicago.

Bondarenko, O. and Sung, J. (2003): Specialist participation and limit orders. In: Journal of Financial Markets, 6, 539–571.

Booth, G. G., Lin, J.-C., Martikainen, T. and Tse, Y. (2002): Trading and pricing in upstairs and downstairs stock markets. In: Review of Financial Studies, (15), 1111–1135.

Bourghelle, D. and Declerck, F. (2004): Why markets should not necessarily reduce the tick size. In: Journal of Banking and Finance, 28, 373–398.

Brachinger, H.-W. and Weber, M. (1997): Risk as a Primitive: A Survey of Measures of Perceived Risk. In: OR Spektrum, 19, 235–250.

Breen, W. J., Hodrick, L. S. and Korajczyk, R. A. (2002): Predicting equity liquidity. In: Management Science, 48, 470483.

Brennan, M. J. and Subrahmanyam, A. (1996): Market Microstructure and Asset Pricing: On the Compensation for Illiquidity in Stock Returns. In: Journal of Financial Economics, 41, 441–464.

Brooks, R. and Masson, J. (1996): Performance of Stoll's Spread Component Estimator: Evidence from Simulations, Time Series, and Cross-Sectional Data. In: Journal of Financial Research, 19, 459–476.

Brown, D. P. and Jennings, R. H. (1989): On Technical Analysis. In: Review of Financial Studies, 2, 527–551.

Brown, D. P. and Zhang, Z. M. (1997): Market Orders and Market Efficiency. In: Journal of Finance, 52, 277–308.

Caballe, J. and Krishnan, M. (1994): Imperfect Competition in a Multi-Security Market with Risk Neutrality. In: Econometrica, 62, 595–604.

Calcagno, R. and Lovo, S. M. (1998): Bid-Ask Price Competition with Asymmetric Information between Market Makers. CORE Discussion Paper 9816, Universite de Catholique de Louvain.

Chakravarty, S. and Holden, C. W. (1995): An Integrated Model of Market and Limit Orders. In: Journal of Financial Intermediation, 4, 213–241.

Chan, S. H., Kim, K. A. and Rhee, S. G. (2005): Price limit performance: evidence from transactions data and the limit order book. In: Journal of Empirical Finance, 12, 269–290.

Chau, M. (1998): Dynamic Trading and Market-Making with Inventory Costs and Private Information. Mimeo, ESSEC, France.

Cho, J.-W. and Nelling, E. (2000): The probability of limit order execution. In: Financial Analysts Journal, 28–33.

Chordia, T., Roll, R. and Subrahmanyam, A. (2000): Commonality in Liquidity. In: Journal of Financial Economics, 56, 3–28.

Chowdhry, B. and Nanda, V. (1991): Multimarket trading and market liquidity. In: Review of Financial Studies, 4, 483–511.

Christie, W. G. and Schultz, P. H. (1994): Why Do NASDAQ Market Makers Avoid Odd-eighth Quotes?. In: Journal of Finance, 49, 1813–1840.

Chung, K. H. and Li, M. (2003): A statistical model of a limit order market. In: Financial Review, 38, 257–272.

Chung, K. H., Li, M. and McInish, T. H. (2005): Information-based trading, price impact of trades, and trade autocorrelation. In: Journal of Banking & Finance, 29, 16451669.

Chung, S. and Wei, P. (2005): The relationship between bidask spreads and holding periods: The case of chinese a and b shares. In: Global Finance Journal, 15, 239 249.

Cochrane, J. (2000): Asset Pricing. Chicago: Graduate School of Business, University of Chicago.

Cohen, K. J., Maier, S. F., Schwartz, R. A. and Whitcomb, D. K. (1981): Transactions Costs, Order Placement Strategy, and Existence of the Bid-Ask Spread. In: Journal of Political Economy, 89, 287–305.

Cohen, K. J., Maier, S. F., Schwartz, R. A. and Whitcomb, D. K. (1986): The Microstructure of Securities Markets. Englewood Cliffs, NJ: Prentice-Hall.

Constantinides, G. M. (1986): Capital market equilibrium with transaction costs. In: Journal of Political Economy, 94, 842–862.

Copeland, T. E. and Galai, D. (1983): Information Effects on the Bid-Ask-Spread. In: Journal of Finance, 38, 1457–1469.

Cordella, T. and Foucault, T. (1999): Minimum Price Variations, Time Priority, and Quote Dynamics. In: Journal of Financial Intermediation, 8, 141–173.

Coughenour, J. F. and Deli, D. N. (2002): Liquidity provision and the organizational form of nyse specialist firms. In: Journal of Finance, 52, 841–869.

Cournot, A. (1963 (orig. 1838)): The Mathematical Principles of the Theory of Wealth. Homewood, Ill.: Richard D. Irwin, Inc.

Cymbalista, F. (1998): Zur Unmöglichkeit rationaler Bewertung unter Unsicherheit - Eine monetär-keynesianische Kritik der Diskussion um die Markteffizienzthese. Studien zur Monetären Ökonomie. Marburg: Metropolis Verlag.

de Frutosa, M. A. and Manzano, C. (2005): Trade disclosure and price dispersion. In: Journal of Financial Markets, 8, 183–216.

Demsetz, H. (1968): The Cost of Transacting. In: Quarterly Journal of Economics, 82, 33–53.

Dennert, J. (1993): Price Competition between Market Makers. In: Review of Economic Studies, 60, 735–751.

Diamond, D. W. and Verrecchia, R. E. (1981): Information Aggregation in a Noisy Rational Expectations Economy. In: Journal of Financial Economics, 9, 221–235.

Domowitz, I. (1993): A Taxonomy of Automated Trade Execution Systems. In: Journal of International Money and Finance, 12, 607–631.

Duffie, D. (1996): Dynamic Asset Pricing. 2nd edition, Princeton, NJ: Princeton University Press.

Dumas, B. and Allaz, B. (1996): Financial Securities: Market Equilibrium and Pricing Methods. London: Chapman & Hall.

Dutta, P. K. and Madhavan, A. (1997): Competition and Collusion in Dealer Markets. In: Journal of Finance, 52, 245–276.

Easley, D., Engle, R. F., O'Hara, M. and Wu, L. (2001): Time-varying arrival rates of informed and uninformed trades. Mimeo, Cornell University.

Easley, D., Hvidkjaer, S. and O'Hara, M. (2002): Is information risk a determinant of asset returns?. In: Journal of Finance, 57, 2185–2221.

Easley, D., Kiefer, N. M., O'Hara, M. and Paperman, J. B. (1996): Liquidity, information and infrequently traded stocks. In: Journal of Finance, 51, 1405–1436.

Easley, D. and O'Hara, M. (1995): Market Microstructure, chapter 12, 357–383. In: Jarrow et al. (1995).

Edelen, R. and Gervais, S. (2003): The role of trading halts in monitoring a specialist market. In: Review of Financial Studies, 16, 263–300.

Eleswarapu, V. R. (1997): Cost of transating and expected reurns in the nasdaq market. In: Journal of Finance, 52, 2113–2127.

Ellis, K., Michaelt, R. and O'Hara, M. (2000): The accuracy of trade classification rules: Evidence from nasdaq. In: Journal of Financial and Quantitative Analysis, 35, 529–551.

Ellul, A., Shin, H. S. and Tonks, I. (2003): How to open and close the market: Lessons from the london stock exchange. Mimeo, Indiana University.

Fama, E. F. (1970): Efficient Capital Markets: A Review of Theory and Empirical Work. In: Journal of Finance, 25, 383–423.

Foster, F. D. and Viswanathan, S. (1990): A Theory of Interday Variations in Volume, Variance, and Trading Costs in Securities Markets. In: Review of Financial Studies, 3, 593–624.

FOSTER, F. D. AND VISWANATHAN, S. (1993): The Effect of Public Information and Competition on Trading Volume and Price Volatility. In: Review of Financial Studies, 6, 23–56.

FOUCAULT, T., KADAN, O. AND KANDEL, E. (2005): Limit order book as a market for liquidity. In: Review of Financial Studies, 18, forthcoming.

FOUCAULT, T., MOINAS, S. AND THEISSEN, E. (2004): Does anonymity matter in electronic limit order markets?. Mimeo, HEC, School of Management, Paris.

FOUCAULT, T. AND PARLOUR, C. A. (2004): Competition for listings. In: RAND Journal of Economics, 35, 329–355.

FREY, B. S. (2000): Was bewirkt die Volkswirtschaftslehre?. In: Perspektiven der Wirtschaftspolitik, 1(1), 5–33.

GÂRLEANU, N. AND PEDERSEN, L. H. (2004): Adverse selection and the required return. In: Review of Financial Studies, 17, 643–665.

GARMAN, M. B. (1976): Market Microstructure. In: Journal of Financial Economics, 3, 257–275.

GEHRIG, T. AND JACKSON, M. (1998): Bid-Ask Spreads with Indirect Competition Among Specialists. In: Journal of Financial Markets, 1, 89–119.

GEORGE, T. J., KAUL, G. AND NIMALENDRAN, M. (1991): Estimation of the Bid-Ask Spread and its Components: A New Approach. In: Review of Financial Studies, 4, 623–656.

GLOSTEN, L. R. (1989): Insider Trading, Liquidity, and the Role of the Monopolist Specialist. In: Journal of Business, 62, 211–235.

GLOSTEN, L. R. AND HARRIS, L. E. (1988): Estimating the Components of the Bid/Ask Spread. In: Journal of Financial Economics, 21, 123–142.

GLOSTEN, L. R. AND MILGROM, P. R. (1985): Bid, Ask and Transaction Prices in a Specialist Market with Heterogenously Informed Traders. In: Journal of Financial Economics, 14, 71–100.

GOMBER, P. (2000): Elektronische Handelssysteme. Innovative Konzepte und Technologien im Wertpapierhandel. Heidelberg: Physica-Verlag.

GRAMMIG, J. AND THEISSEN, E. (2002): Estimating the probability of informed trading - does trade misclassification matter?. Mimeo, University of St Gallen.

GROSSMAN, S. J. AND MILLER, M. H. (1988): Liquidity and Market Structure. In: Journal of Finance, 43, 617–637.

GROSSMAN, S. J. AND STIGLITZ, J. E. (1980): On the Impossibility of Informationally Efficient Markets. In: American Economic Review, 70, 393–408.

HAGERTY, K. (1991): Equilibrium Bid-Ask Spreads in Markets with Multiple Assets. In: Review of Economic Studies, 58, 237–257.

HANOUSEK, J. AND PODPIERA, R. (2002): Information-driven trading at the prague stock exchange. In: Economics of Transition, 10, 757–769.

HANSCH, O., NAIK, N. Y. AND VISWANATHAN, S. (1998): Do Inventories Matter in Dealership Markets? Evidence from the London Stock Exchange. In: Journal of Finance, 53, 1623–1656.

HARRIS, L. E. (1994): Minimum Price Variations, Discrete Bid-Ask Spreads, and Quotation Sizes. In: Review of Financial Studies, 7, 149–178.

HART, O. AND MOORE, J. (1996): The governance of exchanges: Members cooperatives versus outside ownership. In: Oxford Review of Economic Policy, 12, 53–69.

HASBROUCK, J. AND SEPPI, D. J. (2001): Common factors in prices, order flows, and liquidity. In: Journal of Financial Economics, 59, 383–411.

HAYEK, F. A. (1945): The Use of Knowledge in Society. In: American Economic Review, 35, 519–530.

HIRSCHLEIFER, D. (2001): Investor psychology and asset pricing. In: Journal of Finance, 56, 1533–1597.

HO, T. AND STOLL, H. R. (1980): On Dealer Markets under Competition. In: Journal of Finance, 35, 259–268.

HO, T. AND STOLL, H. R. (1981): Optimal Dealer Pricing under Transactions and Return Uncertainty. In: Journal of Financial Economics, 9, 47–73.

HO, T. AND STOLL, H. R. (1983): The Dynamics of Dealer Markets under Competition. In: Journal of Finance, 38, 1053–1074.

HOLDEN, C. W. AND SUBRAHMANYAM, A. (1992): Long-Lived Private Information and Imperfect Competition. In: Journal of Finance, 47, 247–270.

HOLLIFIELD, B., MILLER, R. A. AND SANDÅS, P. (2004): Empirical analysis of limit order markets. In: Review of Economic Studies, 71, 1027–1063.

HOPT, K. J. AND BAUM, H. (1997): Börsenrechtsreform in Deutschland, chapter 3, 287–467. In: HOPT ET AL. (1997).

HOPT, K. J., RUDOLPH, B. AND BAUM, H., (eds.) (1997): Börsenreform - Eine ökonomische, rechtsvergleichende und rechtspolitische Untersuchung. Stuttgart: Schäffer-Poeschel Verlag.

HUANG, M. (2003): Liquidity shocks and equilibrium liquidity premia. In: Journal of Economic Theory, 109, 104–129.

HUANG, R. D. AND STOLL, H. R. (1997): The Components of the Bid-Ask Spread: A General Approach. In: Review of Financial Studies, 10, 995–1034.

HUBERMAN, G. AND HALKA, D. (2000): Systematic liquidity. In: Journal of Financial Research, 24, 161–178.

INGERSOLL, J. E. (1987): Theory of Financial Decision Making. Studies in Financial Economics. Savage, MD: Rowman & Littlefield.

INTRILIGATOR, M. D. (1971): Mathematical Optimization and Economic Theory, chapter 13, 326–343. Englewwod Cliffs, N. J., USA: Prentice Hall.

JAIN, N. AND MIRMAN, L. J. (1999): Insider Trading with Correlated Signals. In: Economics Letters, 65, 105–113.

JARROW, R. A., MAKSIMOVIC, V. AND ZIEMBA, W. T., (eds.) (1995): Finance. Handbooks in Operations Research and Management Science. Amsterdam: Elsevier.

JEVONS, W. S. (1924 (orig. 1911)): Die Theorie der Politischen Ökonomie. Jena: Verlag von Gustav Fischer.

JOHNSON, N. AND KOTZ, S. (1970): Distributions in Statistics, 4 vols. New York: Houghton Miffin.

KADLEC, G. B. AND MCCONNELL, J. J. (1994): The effect of ¡arket segmenttation and illiquidity on market prices: Evidence from exchange listings. In: Journal of Finance, 49, 611–636.

KEENAN, W. M. (1987): The Securities Industry: Securities Trading and Investment Banking, chapter 9. Wiley Professional Banking and Finance Series. 6th edition, New York: John Wiley & Sons.

KEYNES, J. M. (1930): A Treatise on Money. London.

Keynes, J. M. (1936): The General Theory of Employment, Interest, and Money. London.

Knight, F. H. (1957 (Orig. 1921)): Risk, Uncertainty and Profit. New York: Kelly & Millman, Inc.

Kocherlakota, N. R. (1996): The equity premium: Its still a puzzle. In: Journal of Economic Literature, 34, 4271.

Krause, A. (2003): Inventory effects on daily returns in financial markets. In: International Journal of Theoretical and Applied Finance, 6, 739–765.

Krause, A. (2005): Optimal stock allocation in specialist markets. In: Research in Economics, 59, 23–39.

Krishnan, M. (1992): An Equivalence between the Kyle (1985) and the Glosten-Milgrom (1985) Models. In: Economics Letters, 40, 333–338.

Kyle, A. S. (1985): Continuous Auctions and Insider Trading. In: Econometrica, 53, 1315–1335.

Laux, P. A. (1993): Trade Sizes and Theories of the Bid-Ask Spread. In: Journal of Financial Research, 16, 237–249.

Lee, R. (1998): What is an Exchange? The Automation, Management and Regulation of Financial Markets. Oxford: Oxford University Press.

Lei, Q. and Wu, G. (2001): The behavior of uninformed investors and time-varying informed trading activities. Mimeo, University of Michigan Business School.

Lei, Q. and Wu, G. (2005): Time-varying informed and uninformed trading activities. In: Journal of Financial Markets, 8, 153181.

Levy, H. and Sarnat, M. (1972): Investment and Portfolio Analysis. New York: John Wiley & Sons, Inc.

Lin, J.-C., Sanger, G. C. and Booth, G. G. (1995): Trade Size and Components of the Bid-Ask Spread. In: Review of Financial Studies, 8, 1153–1183.

Lintner, J. (1965a): Security Prices, Risk, and Maximal Gains from Diversification. In: Journal of Finance, 20, 587–615.

Lintner, J. (1965b): The Valuation of Risk assets and the selection of Risky Investments in Stock Portfolios and Capital Budgets. In: Review of Economics and Statistics, 47, 13–37.

Lo, A. W., Mamaysky, H. and Wang, J. (2004): Asset prices and trading volume under fixed transactions costs. In: Journal of Political Economy, 112, 1054–1090.

Loderer, C. and Roth, L. (2003): The pricing discount for limited liquidity: evidence from swx swiss exchange and the nasdaq. In: Journal of Empirical Finance, 12, 239 268.

Lucas, R. E. and Prescott, E. C. (1971): Investment under Uncertainty. In: Econometrica, 39, 659–681.

Luckock, H. (2001): A statistical model of a limit order market. Mimeo, University of Sydney.

Ma, T. and Yang, J. (2002): Measuring the probability of informed trading in a call auction market and a comprehensive analysis on the determinants of informed trading. Mimeo, Department of Finance, National Sun Yat-sen University, Kaohsiung, Taiwan.

Madhavan, A. (1995): Consolidation, fragmentation, and the disclosure of trading information. In: Review of Financial Studies, 8, 579–603.

Markowitz, H. M. (1959): Portfolio Selection: Efficient Diversification of Investments. New Haven, CT: Yale University Press.

Marshall, A. (1997, orig. 8th. ed. 1920): Principles of Economics. Amherst, N.Y.: Prometheus Books.

Martínez, M. A., Nieto, B., Rubio, G. and Tapia, M. (2003): Asset pricing and systematic liquidity risk: An empirical investigation of the spanish stock market. Mimeo, Universidad Carlos III de Madrid.

Mehra, R. and Prescott, E. C. (1985): The equity premium: A puzzle. In: Journal of Monetary Economics, 15, 145–161.

Merkt, H. (1997): Zur Entwicklung des deutschen Börsenrechts von den Anfängen bis zum Zweiten Finanzmarktförderungsgesetz, 17–141. In: Hopt et al. (1997).

Moulin, H. (2000): Priority Rules and Other Asymmetric Rationing Methods. In: Econometrica, 68, 643–684.

Muth, J. F. (1961): Rational Expectations and the Theory of Price Movements. In: Econometrica, 29, 315–335.

Neal, R. and Wheatley, S. (1995): How Reliable are Adverse Selection Models of the Bid-Ask Spread?. Federal Reserve Bank of Kansas City Research Working Paper RWP 95-02.

Němeček, L. and Hanousek, J. (2002): Market structure, liquidity, and information based trading at the prague stock exchange. In: Emerging Markets Review, 3, 293–305.

O'Hara, M. (1995): Market Microstructure Theory. Cambridge, Mass.: Blackwell Publishers.

O'Hara, M. (2003): Liquidity and price discovery. In: Journal of Finance, 58, 1335–1354.

Parlour, C. A. (1998): Price Dynamics in Limit Order Markets. In: Review of Financial Studies, 11, 789–816.

Parlour, C. A. and Seppi, D. J. (2003): Liquidity-based competition for order flow. In: Review of Financial Studies, 16, 301–343.

Pástor, L. and Stambaugh, R. F. (2003): Liquidity risk and expected stock returns. In: Journal of Political Economy, 111, 642–685.

Pereira, J. P. and Zhang, H. H. (2004): The liquidity premium in a dynamic model with price impact. Mimeo, University of North Carolina.

Pesaran, M. H. (1987): The Limits to Rational Expectations. Oxford: Basil Blackwell Ltd.

Pirrong, C. (1999): The Organization of Financial Exchange Markets: Theory and Evidence. In: Journal of Financial Markets, 2, 329–357.

Pirrong, C. (2001): The macrostructure of electronic financial markets. Mimeo, Oklahoma State University.

Porter, D. C. and Weaver, D. G. (1996): Estimating Bid-Ask Spread Components: Specialist versus Multiple Market Maker Systems. In: Review of Quantitative Finance and Accounting, 6, 167–180.

Porter, R. B. (2003): Measuring market liquidity. Mimeo, University of Florida.

Pratt, J. W. (1964): Risk Aversion in the Small and in the Large. In: Econometrica, 32, 122–136.

Roll, R. (1984): A Simple Implicit Measure of the Effective Bid-Ask Spread in an Efficient Market. In: Journal of Finance, 39, 1127–1139.

Rosu, I. (2003): A dynamic model of the limit order book. Mimeo, MIT.

Rubio, G. and Tapia, M. (1998): The liquidity premium in equidity pricing under a continuous auction system. In: Eurpean Journal of Finance, 4, 1–28.

Rudolph, B. and Röhrl, H. (1997): Grundfragen der Börsenorganisation aus ökonomischer Sicht, chapter 2, 143–285. In: Hopt et al. (1997).

SAAR, G. (2001): Price impact asymmetry of block trades: An institutional trading explanation. In: Review of Financial Studies, 14, 1153–1181.

SCHILLER, B. AND MAREK, M. (2000): DTB + SOFFEX = EUREX. In: WiSt, (4), 219–221.

SCHMIDT, H. AND TRESKE, K. (1996): Komponenten der Geld-Brief-Spanne am Deutschen Aktienmarkt. In: Zeitschrift für Betriebswirtschaft, 66, 1033–1056.

SCHNEEWEISS, H. (1967): Entscheidungskriterien bei Risiko. Hamburg: Springer Verlag.

SCHULTZ, P. H. (2000): Regulatory and Legal Pressures and the Costs of NASDAQ Trading. Mimeo, University of Notre-Dame.

SCHWARTZ, R. A. (1988): Equity Markets: Structure, Trading, and Performance. New York: Harper & Row.

SCHWARTZ, R. A. (1993): Reshaping the Equity Markets: A Guide For the 1990s. Homewood, Ill.: Business One Irwin.

SEPPI, D. J.. Equilibrium block trading and asymmetric information. In: Journal of Finance.

SEPPI, D. J. (1997): Liquidity provision with limit orders and a strategic specialist. In: Review of Financial Studies, 10, 103–150.

SHARPE, W. F. (1964): Capital Asset Prices: A Theory of Market equilibrium Under Conditions of Risk. In: Journal of Finance, 19, 425–442.

SHARPE, W. F. (1970): Portfolio Theory and Capital Markets. New York: McGraw-Hill Book Co.

SHEFFRIN, S. M. (1996): Rational Expectations. Cambridge: Cambridge University Press.

SHLEIFER, A. (2000): Inefficient Markets - An Introduction to Behavorial Finance. Oxford: Oxford University Press.

SMITH, J. W., SELWAY, J. P. AND MCCORMICK, D. T. (1998): The Nasdaq Stock Market: Historical Background and Current Operation. NASD Working Paper 98-01.

SNELL, A. AND TONKS, I. (1999): Measuring Microstructure Effects for Less-Liquid Stocks in a Dealer Market. University of Bristol, Mimeo.

SPIEGEL, M. AND SUBRAHMANYAM, A. (1992): Informed Speculation and Hedging in a Noncompetitive Securities Market. In: Review of Financial Studies, 5, 307–329.

STOLL, H. R. (1978): The Supply of Dealer Services in Securities Markets. In: Journal of Finance, 33, 1133–1151.

STOLL, H. R. (1989): Inferring the Components of the Bid-Ask spread: Theory and Empirical Tests. In: Journal of Finance, 44, 115–134.

SUBRAHMANYAM, A. (1991): Risk Aversion, Market Liquidity, and Price Efficiency. In: Review of Financial Studies, 4, 417–441.

SWAN, P. L. (2002): Can "illiquidity" explain the equity premium puzzle? the value of endogenous market trading. Mimeo, University of New South Wales.

THALER, R. H., (ed.) (1993): Advances in Behavioral Finance. New York: Russel Sage Foundation.

THE NASDAQ STOCK MARKET, INC., (ed.) (1997): The NASDAQ Stock Market 1997 Factbook. New York.

TOBIN, J. (1958): Liquidity Preference as a Behavior Towards Risk. In: Review of Economic Studies, 25, 65–86.

TOBIN, J. (1966): The Theory of Portfolio Selection. In: F. H. Hahn and F. P. R. Brechling: The Theory of Interest Rates, New York, 3–51.

Todorovic, P. (1992): An Introduction to Stochastic Processes and Their Applications. New York: Springer Verlag.

Unser, M. and Oehler, A. (1998): Das Börsenhandelssystem XETRA. In: WiSt, (9), 463–468.

Vayanos, D. (1998): Transaction costs and asset prices: A dynamic equilibrium model. In: Review of Financial Studies, 11, 1–58.

Venter, J. H. and de Jongh, D. C. J. (2002): Extending the ekop model to estimate the probability of informed trading. Mimeo, Centre for Business Mathematics and Informatics, North-West University, South Africa.

von Neumann, J. and Morgenstern, O. (1953): The Theory of Games and Economic Behavior. Princeton: Princeton University Press.

Wahal, S. (1997): Entry, Exit, Market Makers, and the Bid-Ask Spread. In: Review of Financial Studies, 10, 871–901.

Yavas, A. (1992): Market Makers versus Match Makers. In: Journal of Financial Intermediation, 2, 33–58.

Author index

Acharya, Viral V. 205, 208
Admati, Anat R. 71, 105, 114, 115
Affleck-Graves, John 160
Allaz, Blaise 255
Amihud, Yakov 81, 195–197, 204
Angel, James J. 16, 216
Arnold, Tom 217
Arrow, Kenneth J. 7, 255, 285
Aschinger, Gerhard 273
Atkins, Allen B. 195

Bacidore, Jeff 215
Bagehot, Walter ii, 100–102
Baker, H. Kent 197
Barclay, Michael J. 197
Baruch, Shmuel 202
Baum, Harald 2
Baumol, William J. 241
Beed, Cara 281
Beed, Clive 281
Benveniste, Lawrence 197
Bernhardt, Dan 112
Biais, Bruno 93, 141, 149
Black, Fisher 275
Bloomfield, Robert 214
Blume, Lawrence 65, 67, 68
Board, John 214
Bondarenko, Oleg 59, 214
Booth, G. Geoffrey 161, 215
Bourghelle, David 215
Brachinger, Hans-Wolfgang 258
Breen, William J. 202, 205
Brennan, Michael J. 202
Brooks, Raymond 159
Brown, David P. 66, 150

Caballe, Jordi 59, 63, 64
Calcagno, Riccardo 102
Cappozza, Dennis R. 197
Chakravarty, Sugato 126
Chan, Soon Huat 216
Chau, Minh 59
Cho, Jin-Wan 149
Chordia, Tarun 205, 207
Chowdhry, Bhagwan 215
Christie, William G. 213, 221
Chung, Kee H. 163
Chung, Shifei 209
Cochrane, John 6
Cohen, Kalman J. 127, 129, 130, 135
Constantinides, George M. 197
Copeland, Thomas E. ii, 102
Cordella, Tito 215
Coughenour, Jay F. 216
Cournot, Augustin 2
Cymbalista, Flávia 251

de Frutosa, M. Angeles 214
de Jongh, D. C. J. 163
Declerck, Fany 215
Deli, Daniel N. 216
Demsetz, Harold ii, 77
Dennert, Jürgen 112
Diamond, Douglas W. 66
Domowitz, Ian 20
Duffie, Darrel 6
Dumas, Bernard 255
Dutta, Prajit K. 213
Dyl, Edward A. 195

Easley, David i, 65, 67, 68, 161, 163, 198, 202

Edelen, Roger 216
Edelman, Richard B. 197
Eleswarapu, Venkat R. 197
Ellis, Katrina 164
Ellul, Andrew 215
Engle, Robert F. 163, 198

Fama, Eugene F. 6
Foster, F. Douglas 59, 70
Foucault, Thierry 137, 141, 214–216
Frey, Bruno S. 281

Galai, Dan ii, 102
Gârleanu, Nicolae 201
Garman, Mark B. i, 77, 81
Gehrig, Thomas 117, 120
George, Thomas J. 160
Gervais, Simon 216
Glosten, Lawrence R. 105, 110, 113, 160, 161
Gomber, Peter 27
Grammig, Joachim 164
Grossman, Sanford J. 46, 65, 151, 171, 178

Hagerty, Kathleen 116
Halka, Dominika 205, 207
Hanousek, Jan 163
Hansch, Oliver 160
Harris, Lawrence E. 160, 161, 215
Hart, Oliver 216
Hasbrouck, Joel 205, 207
Hayek, Friedrich August 7
Hedge, Shantaram P. 160
Hersch, Philip 217
Hillion, Pierre 149
Hirschleifer, David 281
Ho, Thomas 87, 89, 93, 95, 98, 99
Hodrick, Laurie Simon 202, 205
Holden, Craig W. 41, 45, 59, 110, 113, 126
Hollifield, Burton 149
Hopt, Klaus J. 2
Huang, Ming 197, 205
Huang, Roger D. 160
Huberman, Gur 205, 207
Hughson, Eric 112
Hvidkjaer, Soeren 202

Ingersoll, Jonathan E. 6
Intriligator, Michael D. 244

Jackson, Matthew 117, 120
Jain, Neelam 59
Jennings, Robert H. 66
Jevons, William Stanley 3
Johnson, N. 247, 249

Kadan, Ohad 137
Kadlec, Gegrory B. 197
Kandel, Eugene 137, 197
Kaul, Gautam 160
Keenan, W. Michael 19
Keynes, John Maynard 6, 7, 251, 277
Kiefer, Nicholas M. 161, 163
Kim, Kenneth A. 216
Knight, Frank H. 251
Kocherlakota, Narayana R. 208, 209
Korajczyk, Robert A. 202, 205
Kotz, S. 247, 249
Krause, Andreas 164–166, 214
Krishnan, Murugappa 59, 63, 64, 110
Kyle, Albert S. ii, 31, 32, 40, 41, 43, 44, 59, 60, 72, 99, 105, 110–112, 122

Laux, Paul A. 161
Lee, Ruben 2, 8
Lei, Qin 163
Levy, Haim 252, 258, 260, 262
Li, Mingsheng 163
Lin, Ji-Chai 161, 215
Lintner, John 264, 274, 275
Lipson, Marc L. 215
Lo, Andrew W. 149, 197
Loderer, Claudio 197
Lovo, Stefano M. 102
Lucas, Robert E. 280
Luckock, Hugh 150

Ma, Tai 163
Madhavan, Ananth 213, 215
Maier, Steven F. 127, 129, 130, 135
Mamaysky, Harry 149, 197

Manzano, Carolina 214
Marek, Michael 26
Markowitz, Harry M. 257, 258, 273
Marshall, Alfred 3
Martikainen, Teppo 215
Martínez, Miguel A. 207
Marx, Leslie M. 197
Masson, Jean 159
McConnell, John J. 197
McCormick, D. Timothy 222, 230
McInish, Thomas H. 163
Mehra, Rajnish 208
Mendelson, Haim 81, 195–197, 204
Merkt, Hanno 2
Michaelt, Roni 164
Milgrom, Paul R. 105, 110
Miller, Merton H. 171, 178
Miller, Robert A. 149
Miller, Robert E. 160
Mirman, Leonard J. 59
Moinas, Sophie 214
Moore, John 216
Morgenstern, Oskar 252
Moulin, Hervé 20, 216
Mulherin, J. Harold 217
Muth, John F. 279, 280

Naik, Narayan Y. 160
Nanda, Vikram 215
Neal, Robert 160
Nelling, Edward 149
Netter, Jeffry 217
Nieto, Belén 207
Nimalendran, M. 160
Němeček, Libor 163

Oehler, Andreas 23
O'Hara, Maureen i, 9, 65, 67, 68, 161, 163, 164, 198, 202, 214

Panzar, John C. 241
Paperman, Joseph B. 161, 163
Parlour, Christine A. 132, 135, 216
Pástor, Luboš 205, 207
Pedersen, Lasse Heje 201, 205, 208
Pereira, Joao Pedro 205
Pesaran, M. Hashem 280
Pfleiderer, Paul 71, 105, 114, 115
Pirrong, Craig 216
Podpiera, Richard 163
Porter, David C. 160
Porter, R. Burt 208
Pratt, John W. 255, 256, 261
Prescott, Edward C. 208, 280

Rhee, S. Ghon 216
Röhrl, Heiner 8, 22
Roll, Richard 155, 166, 205, 207
Rosu, Ioanid 150
Roth, Lukas 197
Rubio, Gonzalo 197, 207
Rudolph, Bernd 8, 22

Saar, Gideon 202, 215
Salani, François 141
Sandås, Patrick 149
Sanger, Gary C. 161
Sarnat, Marshall 252, 258, 260, 262
Schiller, Bettina 26
Schmidt, Hartmut 160
Schneeweiss, H. 263
Schultz, Paul H. 213, 221, 232
Schwartz, Robert A. 2, 20, 23, 127, 129, 130, 135
Seguin, Paul J. 197
Selway, James P. 222, 230
Seppi, Duane J. 205, 207, 214–216
Sharpe, William F. 274, 275
Sheffrin, Steven M. 279
Shin, Hyun Song 215
Shleifer, Andrei 281
Smith, Jeffrey W. 222, 230
Snell, Andy 160
Spatt, Chester 149
Spiegel, Matthew 49, 58
Stambaugh, Robert F. 205, 207
Stiglitz, Joseph E. 46, 65, 151
Stoll, Hans R. ii, 79, 80, 86, 87, 89, 93, 95, 98, 99, 154, 155, 160
Subrahmanyam, Avanidhar 49, 58, 202, 205, 207

Subrahmanyam, Avinadhar 41, 45, 59, 110, 113
Sung, Jaeyoung 214
Sutcliffe, Charles 214
Swan, Peter L. 197

Tapia, Mikel 197, 207
Theissen, Erik 164, 214
Tobin, James 258, 271
Todorovic, Petar 250
Tonks, Ian 160, 215
Treske, Kai 160
Tse, Yiuman 215

Unser, Matthias 23

Vayanos, Dimitri 197
Venter, J. H. 163
Verrecchia, Robert E. 66
Viswanathan, S. 59, 70, 160
von Neumann, John 252

Wahal, Sunil 214
Wang, Jiang 149, 197
Weaver, Daniel G. 160, 216
Weber, Marianne 258
Wei, Peihwang 209
Wheatley, Simon 160
Whitcomb, David K. 127, 129, 130, 135
Willig, Robert D. 241
Wu, Guojun 163
Wu, Liuren 163, 198

Yang, Josephine 163
Yavas, Abdullah 179, 185, 187, 188

Zhang, Harold H. 205
Zhang, Zhi Ming 150

Index

hhh, i

jjj, i

This book explains how the way assets are traded in markets affect their prices. It explores how asymmetric information between market participants cause informed traders to behave strategically in order to maximize the profits they make from their information. This behavior causes information to be incorporated into the price gradually. Also in the absence of asymmetric information will some traders behave strategically in order to minimize any trading costs, e.g. through the submission of optimal limit orders. These are only a few examples of the way trading rules in markets affect behavior and thus the prices emerging.

Andreas Krause is Lecturer in Finance at the School of Management of the University of Bath, Great Britain. He obtained a PhD in Finance from the University of Fribourg (Switzerland) with his dissertation on strategic behavior of market makers in dealer markets. He has published several papers on market microstructure theory and extensive teaching experience at all undergraduate and postgraduate levels.

Price £25