UNIVERSITÄT DUISBURG-ESSEN
FAKULTÄT FÜR PHYSIK

DISSERTATION

Microstructures of Correlated Financial Markets

Von der Fakultät für Physik
der Universität Duisburg-Essen
genehmigte Dissertation
zur Erlangung des Grades
Dr. rer. nat.
von

M. Sc. Shanshan Wang
aus Hangzhou, China

Datum der Disputation: 11.10.2017
Gutachter: Prof. Dr. Thomas Guhr
Gutachterin: Prof. Dr. Antje Mahayni

Hiermit versichere ich, die vorliegende Dissertation selbstständig, ohne fremde Hilfe und ohne Benutzung anderer als den angegebenen Quellen angefertigt zu haben. Alle aus fremden Werken direkt oder indirekt übernommenen Stellen sind als solche gekennzeichnet. Die vorliegende Dissertation wurde in keinem anderen Promotionsverfahren eingereicht. Mit dieser Arbeit strebe ich die Erlangung des akademischen Grades Doktor der Naturwissenschaften (Dr. rer. nat.) an.

Ort, Datum	Shanshan Wang

List of publications

[1] Shanshan Wang, Rudi Schäfer, and Thomas Guhr, *Cross-response in correlated financial markets: individual stocks*, The European Physical Journal B, 89:105 (2016).

[2] Shanshan Wang, Rudi Schäfer, and Thomas Guhr, *Average cross-responses in correlated financial market*, The European Physical Journal B, 89:207 (2016).

[3] Shanshan Wang and Thomas Guhr, *Microscopic understanding of cross-responses between stocks: a two-component price impact model*, arXiv:1609.04890, submitted to Market Microstructure and Liquidity in September, 2016.

[4] Shanshan Wang, *Trading strategies for stock pairs regarding to the cross-impact cost*, arXiv:1701.03098, submitted to Physica A: Statistical Mechanics and its Applications in January, 2017.

[5] Shanshan Wang and Thomas Guhr, *Local fluctuation of the signed traded volume and the dependencies of demands: a copula analysis*, arXiv:1706.09240 (2017).

Author contributions

Here, I lay out my contributions to the publications and manuscripts mentioned above:

[1] The paper presents the price response of one stock to the trades of other stocks in a correlated market with empirical data. The project was supervised by R. Schäfer and T. Guhr. R. Schäfer and I developed the method of analysis. I carried out the data analysis. All authors contributed equally to analyzing the results. The text was written by T. Guhr and me.

[2] The paper extends the above empirical study [1] by performing different averages of cross-responses, and identifies active and passive cross-responses. The project was supervised by R. Schäfer and T. Guhr. R. Schäfer and I developed the method of analysis. I carried out the data analysis. All authors contributed equally to analyzing the results. The text was written by T. Guhr and me.

[3] The paper provides a price impact model with self- and cross-impacts to interpret the empirical cross-responses [2]. It also reveals the features for the price impacts depending on the traded volumes and on the time lag. The project was supervised by T. Guhr. I extended the model of Bouchaud et al. (2004) only with self-impacts to the one with self- and cross-impacts, and performed the empirical analysis and the theoretical computations. The text was written by T. Guhr and me.

[4] The paper extends the framework of trading strategies of Gatheral (2010) from single stocks to the two-dimensional case where the strategy for executions of two round-trip trades of two stocks is affected by the cross-impact. I carried out the extension of the model, the empirical analysis and the simulation of the trading strategy. The text was written by myself.

[5] The paper introduces the dependencies of demands, and discloses the influence of local fluctuations of the signed traded volumes on them. The project was supervised by T. Guhr. I did all empirical analysis, numerical simulations, and data fits. I also developed the method for analyzing the effect of local fluctuations on the dependencies of demands. The initial text was written by me and later edited by T. Guhr.

Acknowledgements

I want to express my deepest gratitude to my supervisor Thomas Guhr for giving me the great opportunity to work in his group on the interdisciplinary field of econophysics and for providing an outstanding environment to conduct research. His constant support and patient guidance broadened my scientific horizon and also improved my language competence.

I would like to thank my second advisor Rudi Schäfer who gave me much guidance in the analysis of empirical data. I am very grateful for his helpful advice and insightful discussions. I also would like to thank Thilo Schmitt for his great help in the programming and data processing.

I thank all current and former members of our group, especially Maram Akila, Desislava Chetalova, Yuriy Stepanov, Sebastian Krause, Andreas Mühlbacher, Martin Theissen, Daniel Wagner, Daniel Waltner, Tim Wirtz and Marcel Wollschläger for fruitful and extensive discussions.

I want to thank Daniel Waltner, Lijing Jin, Yu Chen, Hangfu Yang, Chao Shen for proofreading parts of this thesis and providing valuable feedback.

Furthermore, I want to say special thanks to my parents and friends for their support throughout these years.

Finally, I acknowledge the financial support from the China Scholarship Council and from the faculty of physics in University of Duisburg-Essen during my doctoral studies.

Abstract

In econophysics, physicists apply physical theories and methods to address economics problems. Due to an enormous amount of available data, financial markets can be statistically analyzed by physicists. The applied methods find their applications also in the context of other complex systems. In particular, with the development of the high-frequency trading, the market microstructure has gained growing attention. In this thesis, we will focus on the microstructure of financial markets, particularly on the correlation of order flow, the price impact and the dependence of demands.

We begin by developing a method to identify trade signs with a TAQ data set. With the identified trade signs, we carry out an analysis of empirical data for the price cross-response to trades and the cross-correlation of trade signs. To obtain a stable observation, we also average them. The average cross-correlation of trade signs turns out to be long memory. Meanwhile, the non-vanishing cross-response reflects non-Markovian features of prices. According to the average cross-responses, we identify the influencing and influenced stocks.

We then extend the price impact model of Bouchaud et al. (2004) to interpret our empirical results. The extended model contains the impacts of traded volumes, which are empirically revealed as power-law functions. The model also includes a self- and a cross-impact function of time lag. To quantify them, we propose a construction to fix the parameters and employ a diffusion function to corroborate the parameters. We thus quantify and interpret the price impacts in terms of the temporary and permanent components.

We further extend the framework of trading strategies of Gatheral (2010) from single stocks to the two-dimensional case. Thus, we can introduce the cross-impact to the strategy for executing two round-trip trades of two stocks. We apply the strategy to a specific case, in which we quantify the cross-impacts with empirical data and give a view of how the cross-impact affect the trading strategy.

We finally analyze the dependence of demands between stocks by a copula method. The empirical dependence of demands can be well described by a $\mathcal{K}$ copula density function. We also investigate how the large local fluctuations of the signed traded volumes affect the dependence of demands. Furthermore, we explore the asymmetries of tail dependencies of the copula density.

Zusammenfassung

In Wirtschaftsphysik konzentrieren sich Physiker auf die Anwendung physikalischer Theorien und Methoden zur Untersuchungen ökonomischer Probleme. Aufgrund des großen verfügbaren Datenvolumens können Finanzmärkte von Physikern mit statistischen Methoden untersucht werden. Diese Methoden finden ebenfalls Anwendung bei der Untersuchung anderer komplexer Systeme. Insbesondere mit dem Aufkommen des Hochfrequenzhandels, gewinnt die Mikrostruktur des Marktes wachsende Aufmerksamkeit. In dieser Arbeit konzentrieren wir uns auf die Mikrostruktur des Marktes, insbesondere auf die Korrelationen des Orderflusses, den Einfluss des Preises und die Abhängigkeit von der Nachfrage.

Wir beginnen mit der Entwicklung einer Methode zur Identifikation von Handelsvorzeichen mittels eines TAQ Datensatzes. Mit den identifizierten Handelsvorzeichen analysieren wir empirische Daten für die Kreuzantwort des Preises auf Handel und die Kreuzkorrelation der Handelsvorzeichen. Desweiteren mitteln wir diese, um sie zu stabilisieren. Diese gemittelten Kreuzkorrelationsfunktionen zeigen Langzeitkorrelationen. Die vorhandenen Kreuzantworten spiegeln die nicht markovschen Eigenschaften der Preise wider. Basierend auf den gemittelten Kreuzantworten identifizieren wir beeinflussende und beeinflusste Aktien.

Ausserdem erweitern wir das Preiseinflussmodell von Bouchaud et al. (2004) um unsere datenbasierten Ergebnisse zu interpretieren. Die erweiterten Modelle enthalten den Einfluss von Handelsvolumens, die sich aufgrund von Datenanalysen als Potenzgesetze identifizieren lassen. Das Modell enthält Selbst- und Kreuzeinflussfunktion der Zeitverzögerung. Um diese zu quantifizieren, schlagen wir eine Konstruktion vor, um die Parameter festzulegen und verwenden eine Diffusionsfunktion um die gewählten Parameter zu bestätigen. Wir quantifizieren und interpretieren damit den Einfluss auf die Preisentwicklung in permanenten und zeitabhängigen Komponenten.

Darüber hinaus erweitern wir das Modell der Handelsstategien von Gathernal (2010) vom ein- auf den zweidimensionalen Fall. Damit können wir Kreuzeinflussfunktionen in die Strategie zur Ausführung zweier Round-Trip-Geschäfte zweier Aktien einführen. Wir wenden die Strategie auf einen Spezialfall an, in dem wir die Kreuzeinflussfunktion mittels systembasierter Daten quantifizieren und erläutern, wie die Kreuzeinflussfunktion die Handelsstrategie beeinflusst.

Zuletzt untersuchen wir die Abhängigkeit der Nachfrage nach Aktien mit Hilfe einer Kopula-Methode. Die Abhängigkeit der Nachfrage kann datenbasiert gut beschrieben werden mittels eine $\mathcal{K}$ Kopula Dichte Funktion. Wir untersuchen auch wie starke lokale Schwankungen des Vorzeichens des Handelsvolumens die Abhängigkeit der Nachfrage beeinflussen. Ausserdem untersuchen wir die Asymmetrie der Abhängigkeit der Flügel der Kopula Dichte.

Contents

1 Introduction 1
1.1 Econophysics . . . 1
1.2 Financial markets . . . 3
1.2.1 NASDAQ stock market . . . 3
1.2.2 High-frequency trading . . . 3
1.3 Market microstructures . . . 4
1.3.1 Orders, order book and liquidity . . . 5
1.3.2 Continuous double auction . . . 6
1.3.3 Prices, returns and volatility . . . 8
1.3.4 Correlation of order flow . . . 9
1.3.5 Price impact and response . . . 10
1.3.6 Two-phase behavior in demands . . . 11
1.4 Outline of the thesis . . . 12

2 Data sets, time conventions and trade signs 15
2.1 Introduction . . . 15
2.2 Data sets . . . 15
2.2.1 TAQ data set . . . 16
2.2.2 TotalView-ITCH data set . . . 16
2.3 Time conventions . . . 18
2.4 Classification of trade signs . . . 19
2.5 Tests of sign classifications . . . 20
2.5.1 Obtaining empirical trade signs . . . 20
2.5.2 Tests for signs of single trades . . . 20
2.5.3 Tests for trade signs of one-second intervals . . . 22
2.6 Conclusions . . . 23

3 Cross-responses in correlated financial markets: individual stocks 25
3.1 Introduction . . . 25
3.2 Data description . . . 26
3.3 Cross-responses for pairs of stocks . . . 27
3.3.1 Cross-response functions . . . 27
3.3.2 Trade sign cross-correlators . . . 29
3.3.3 Influences of zero trade signs . . . 30

3.3.4 Cross-response noise . . . 32
3.4 Market response . . . 33
3.4.1 Market response structure . . . 33
3.4.2 Transient market impact . . . 34
3.4.3 Temporary violation of market efficiency . . . 35
3.4.4 Informed trades . . . 37
3.5 Comparisons of self- and cross-responses . . . 37
3.6 Conclusions . . . 40

4 Average cross-responses in correlated financial markets 43
4.1 Introduction . . . 43
4.2 Data description . . . 44
4.3 Average cross-responses of an individual stock . . . 44
4.3.1 Definitions . . . 44
4.3.2 Responses to the market . . . 45
4.3.3 Responses to economic sectors . . . 47
4.4 Influencing and influenced stocks from the viewpoint of average cross-responses 47
4.4.1 Identifying influencing and influenced stocks . . . 47
4.4.2 Relations of influencing stocks and trading frequency . . . 50
4.4.3 Role of trading frequency in response functions . . . 51
4.5 Comparisons of self- and average cross-responses . . . 52
4.6 Conclusions . . . 54

5 Microscopic understanding of cross-responses: a price impact model 57
5.1 Introduction . . . 57
5.2 Price impact model . . . 58
5.2.1 Setup of the model . . . 58
5.2.2 Simplifications of the model . . . 61
5.2.3 Scenario I: Cross-response related to trade sign cross-correlators . . 62
5.2.4 Scenario II: Cross-response related to trade sign self-correlators . . . 63
5.2.5 Scenario III: Cross-response related to both correlators . . . 64
5.3 Empirical analysis . . . 64
5.3.1 Data sets and definitions . . . 64
5.3.2 Properties of trade sign correlators . . . 65
5.3.3 Impacts of traded volumes . . . 67
5.4 A construction to quantify price impacts . . . 70
5.4.1 Impact function . . . 71
5.4.2 A construction to fix parameters . . . 73
5.5 Relation to correlated diffusion . . . 76
5.5.1 Price diffusion functions . . . 76
5.5.2 Correlated motion of prices . . . 78
5.5.3 Consistency of our model . . . 79
5.6 Price impacts of individual stocks . . . 81
5.7 Conclusions . . . 83

6 Trading strategies with cross-impact costs 85
6.1 Introduction 85
6.2 Model setup 86
6.2.1 Trade price 86
6.2.2 Costs of trading 88
6.2.3 A construction of trading strategies 91
6.3 Applications to a specific case 92
6.3.1 Data sets 92
6.3.2 Trade signs 92
6.3.3 Measurement for impacts of traded volumes 93
6.3.4 Measurement for cross-impacts of time lag 95
6.3.5 Computations and discussions of trading strategies 96
6.4 Conclusions 97

7 Local fluctuations of the signed traded volumes and the dependencies of demands: a copula analysis 101
7.1 Introduction 101
7.2 Data set and methods of analysis 102
7.2.1 Data set 102
7.2.2 Trade signs and demands 102
7.2.3 Copula densities 103
7.2.4 Empirical estimation of copula densities 104
7.3 Empirical results 106
7.3.1 Demand distributions 106
7.3.2 Copula densities 106
7.4 Comparison of two models with the empirical copula density 107
7.4.1 Bivariate $\mathcal{K}$ copula density 107
7.4.2 Gaussian copula density 109
7.4.3 Fits 109
7.5 Influence of local fluctuations on dependencies 109
7.5.1 Feasibility of our method 110
7.5.2 Influence on the dependence structure 111
7.5.3 Correlations induced by large local fluctuations 113
7.5.4 Influence on the asymmetries of tail dependencies 115
7.6 Conclusions 117

8 Conclusion and Outlook 119

Appendices 123

A Stock information 123
A.1 Trading information of 99 stocks 123
A.2 Lists of 496 stocks in S&P 500 index 124
A.3 Daily trading information of 31 stocks 126
A.4 Daily trading information of 100 stocks 126

B Error estimation 129

C Diffusion equation in two dimensions **131**

List of Figures **135**

List of Tables **139**

Bibliography **141**

Chapter 1

Introduction

1.1 Econophysics

In the past decades, some of physicists have been trying to apply theories and methods of physics to solve problems in economics and finance, resulting in a new interdisciplinary research field—econophysics. Such application can be traced back to the late 1700s when Adam Smith wrote his *Wealth of Nations* [150]. This book laid the foundations of economic thought in 1776. In the book, he introduced Newtonian ideas for causative forces in economics. Since then, financial markets have been recognized gradually as complex systems by physicists. As a new field, econophysics experiences many strong criticisms, for instance, physicists fail to build any valuable theory with explanatory power [32], until Benoit Mandelbrot found the first evidence for fat-tailed distributions in the early 1960s [106]. He showed that the fluctuations in cotton prices follow a distribution that differs from the Gaussian distribution. In other words, the price fluctuations departure from the process of random walks, which is proposed by Louis Bachelier in 1900 to explain the price fluctuations of security and commodity [12, 13]. More importantly, the finding of fat-tailed distributions suggests that the big fluctuations are the inherent characteristic of a normal market [14].

To describe a large amount of research work on problems in markets done by physicists, Eugene H. Stanley first proposed the term "econophysics" in 1995 [64]. He suggests that one way of success for econophysicists is focusing on data. The development of econophysics is indeed restricted by the amount of data. When Benoit Mandelbrot found the fat-tailed distributions, only a few thousand data points are available. Thirty years later, to analyze the scaling behavior of Standard & Poor's Index of the New York Stock Exchange (NYSE) over the six year period (1984–89), Rosario Mantegna and Eugene H. Stanley examined nearly five million data points [108]. Nowadays, the available data points are much more than 200 million and spans half a century [31]. Comparing with most other economic data existing in small, noisy data sets, the high-quality and long-term data from financial markets is regarded as the main reason to explain why physicists are so finance-centered [14].

In modern econophysics, the fat-tailed distribution mentioned above is one of stylized facts. The term "stylized facts" is originally introduced by the economist Nicholas Kaldor in 1961 [90]. He had an argument that any theory should start off with a summary of facts related to problems and any model should be capable of explaining relevantly stylized facts. He suggested that theorists should "concentrate on broad tendencies, ignoring individual details" [90]. More specifically, different individuals can be described by different parametric models, but stylized facts are statistical results of all individuals and thus capture the general, qualitative feature rather than specific features of individuals. Apart from the fat-tailed distribution [140, 142], a lot of stylized facts [26, 29, 36, 45, 46] common to a wide set of financial assets are also found out. It includes the absence of autocorrelations of returns [48, 123], volatility clustering [18, 47, 105], volume-volatility correlations [72, 128], power laws [23, 62], long memory [57, 98, 100], U shape [111], and so on.

By analyzing huge data, physicists have advanced establishing empirical facts about financial markets, but their contributions in economics and finance are far beyond that. How to interpret these empirical facts is another task of equal importance. To this end, physicists have developed many realistic models for markets. For instance, they apply the random matrix theory to financial correlations [131], develop copula approaches to analyze the statistical dependence between stocks [41, 165], employ spin models for trading decisions [22], use the model of Zero Intelligence Trading (ZIT) to simulate continuous double-auction trading [70], and so on. Among these models, the most striking ones for studying markets are agent-based models [37, 56, 96, 112], which originate in the Sante Fe Institute in the early 1980s [10, 11], and develop rapidly in the last two decades. In particular, an agent-based model is a computerized simulation of interactions of agents by following prescribed rules. Comparing with traditional equilibrium models, agent-based models are better to cope with nonlinear behaviors. Other extensively used models for exploring market microstructures are price impact models [27, 99]. These models describe how the price changes due to the effect of trades. The price impact is significant in both academic and industrial fields, since it is linked closely to the cost of trading [65] and the optimal executions of orders [5, 6, 65–67, 120].

As the techniques of data analysis in economics and finance develop pretty fast, physicists are able to identify the instability of markets by considering markets as complex systems [115, 116]. The market instability, on the one hand, reveals potential risks that investors may face, and on the other hand, suggests possible opportunities for investors to arbitrage. The latter, however, conflicts with a classical theory—Efficient Market Hypothesis (EMH) put forward by Fama in 1970 [54]. The EMH states that all available information is fully incorporated in an asset's price. As a result, any (statistical) arbitrage opportunity is absent. In view of the conflict between empirical facts and the classical theory, physicists have used their ways to examine the market efficiency. To some extent, they even have helped to improve the EMH, leading to an adaptive market hypothesis proposed by Andrew W. Lo in 2004 [102]. The adaptive market hypothesis (AMH) states the market efficiency from an evolutionary perspective, which not only reconciles the EMH with all its behavioral alternatives, but also incorporates arbitrage opportunities from time to time.

In short, we have shown that a part of physicists have devoted themselves to solving the problems of markets. Applying physics models to economics, they have obtained a lot of stylized facts. By doing so, they even have helped to improve some theories in economics, *e.g.*, the EMH. However, financial markets as complex systems have changed too much in past years, leaving many existed opinions to be updated and many problems to be solved. Finding new patterns of markets, disclosing market microstructures, revealing financial risks, and designing trading strategies are the main tasks that econophysicists devote themselves to carrying out. Fortunately, we now have a mass of available data in finance, as well as advanced techniques for data analysis. Benefiting from these, well will be able to better understand the real markets.

1.2 Financial markets

People who make transactions are called traders. A financial market is a place where traders gather together to buy or sell financial instruments. Here, the instruments include stocks, bonds, options, future contracts, forward contracts, foreign currency contracts, and other derivatives. The place may be a physical trading floor where traders meet and perform trades, *e.g.*, the New York Stock Exchange and the Chicago Mercantile Exchange; it also could be an electronic system linked with the internet. The electronic system facilitates traders in different places to submit orders for trading. Two typical electronic markets are NASDAQ and Euronext. Also, the financial instruments can be traded over the counter (OTC), *i.e.*, a trade is done between two parties without the supervision of an exchange. In this thesis, we will restrict ourselves to the stock markets, especially to the NASDAQ stock market. In the modern financial market, with the development of computerized trading, traders are in a race to make transactions faster. The trading at a fast speed results in the high-frequency trading. In the following, we will take a closer look on the NASDAQ stock market and the high-frequency trading.

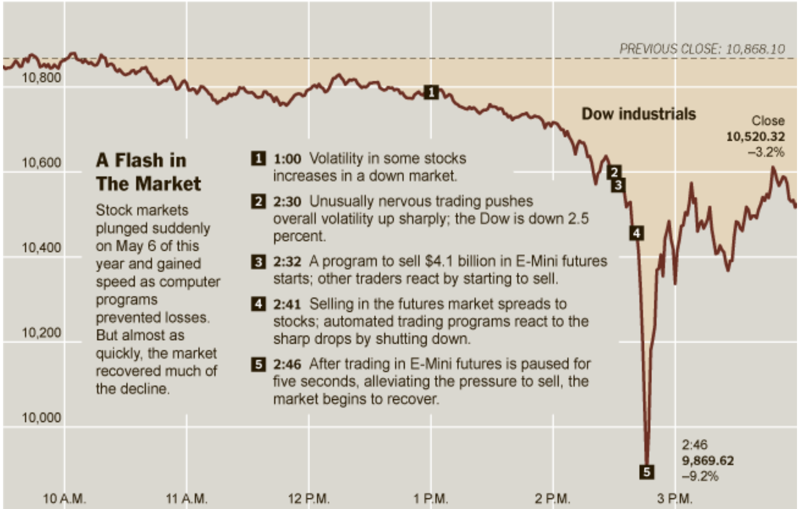
1.2.1 NASDAQ stock market

NASDAQ is the acronym of the National Association of Securities Dealers Automated Quotations and was found by National Association of Securities Dealers (NASD) in 1971. It originates from a quotation system and gradually turns to a publicly electric trading platform [117, 118]. Before the emerging of the electric trading platform, the traders had to call their brokers to make transactions. Such telephone-based trading presents inefficiency, which is noticeable during the market crash in October, 1987 [33, 145]. NASDAQ contains an automatic execution system for orders of up to 1000 shares [94]. This system, called the Small Order Execution System (SOES), guarantees the immediate automatic execution of eligible orders even under turbulent market conditions. As a consequence, NASDAQ computerized trading eliminates the inefficiency of trading.

There are three trading sessions in NASDAQ stock market, a pre-market session from 4:00 a.m. to 9:30 a.m., a regular trading session from 9:30 a.m. to 4:00 p.m., and an after market session from 4:00 p.m. to 8:00 p.m.. Thus, the data of quotes and order entry is available from 4:00 a.m. to 8:00 p.m. for each trading day. Moreover, up to 2017, more than 3000 companies are exchanged in NASDAQ stock market. These companies have a variety of levels of capitalization and cover 15 different industries. Comparing with other markets, NASDAQ stock market possesses richer data for intraday trading and various stocks. These trading information can be simultaneously broadcasted to millions of computer terminals worldwide and accessed equally by all NASDAQ participants.

1.2.2 High-frequency trading

The high-frequency trading refers to a type of algorithmic trading characterized by holding assets with very short time periods from milliseconds to minutes, probably hours. Most high-frequency trading strategies close the positions of assets by the end of each trading



Sources: Bloomberg (Dow industrials); Securities and Exchange Commission THE NEW YORK TIMES

Figure 1.1: The flash crash on May 6, 2010. The figure is taken from The New York Times [2].

day. Carried out by automated trading systems, the high-frequency trading seeks for profits from market liquidity imbalances, short-term pricing inefficiency, arbitrage strategies, and other possible strategies [4, 113, 152]. For instance, a basic high-frequency trading strategy is the execution of large orders chopped into pieces [26, 27, 42, 91, 136, 155]. In the last decade, the high-frequency trading develops quickly in security, foreign currency, and derivatives. For security exchange, as estimated by Tabb Group [74], the high-frequency trading in 2008 accounted for 52% of all traded volumes in United States and 21% in Europe. In 2010, this proportion increased to 56% in United States and to 38% in Europe [74].

The high-frequency trading provides a lot of liquidity for markets by market making [4, 113], but it also introduces risks to markets. A typical example is the flash crash on May 6, 2010, when the Dow Jones Industrial Average index fell by 9% but recovered within minutes [2, 93, 127, 163], shown in Fig. 1.1. For reducing such kind of risks, understanding the market microstructure is very necessary in addition to relevant policy required to regularize makets [113].

When identifying small changes in the quote stream, high-frequency traders rapidly open and close positions [4]. Therefore, the high-frequency trading has extremely high turnover rates, which leads to an incremental proportion of high-frequency trading in the total traded volume. Accordingly, the transaction costs push up to a large number. When the transaction costs are high enough, traders may fail to profit from trading. Thus, lowering the costs of trading is significant for traders to increase their profits. On the other hand, the high turnover rates result in plenty of order flow data with the resolution of one second or even one millisecond. Such a large number of data makes it possible for physicists to statistically analyze markets.

1.3 Market microstructures

Since the market crash in October, 1987, the field of market microstructures has considerably grown in size and importance. According to O'hara (1995), the market microstructure theory "focuses on how specific trading mechanisms affect the price formation process" [121]. Later, Harris (2003) defined the market microstructure as "the branch of financial economics that investigates trading and the organization of markets" [78]. Recently, a common definition given by Hasbrouck (2007) is that "market microstructure is the study of the trading mechanisms used for financial securities" [80]. In view of these definitions, the trading mechanisms are at the heart of market microstructures. The understanding of trading mechanisms are of obvious importance for many practical purposes, for instance, the quantifying of price impacts, the reduction of execution costs, the design of trading strategies, the organisation of markets and the lowering of financial risks [3, 94]. In addition to these areas, the study of market microstructures is of equal importance for exploring the role of information in the price discovery process, the interplay between liquidity taking and providing, and the implication for market efficiency [3, 94].

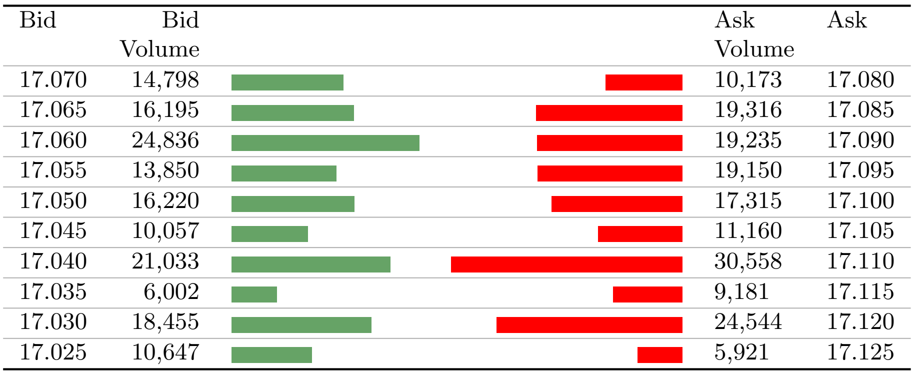
In the following, we will begin with a brief review of some basic concepts, and then introduce the mechanisms of price formation. We further overview the studies on the following key areas of market microstructures: correlations of order flow, price impacts, and the two-phase behavior in demands. For more details of market microstructures, the reader is referred to Refs. [3, 50, 78, 80, 121, 139].

1.3.1 Orders, order book and liquidity

In financial markets, the ultimate goal of traders is to make profits by the way of buying low and selling high. When they decide to perform a trade, they submit an order with trading information, including the price, volume, type, and other relevant conditions. The traded volume is the number of shares that traders intend to exchange. The trade type indicates that the volume is to be bought in or sold out, *i.e.,* buy or sell for short. All the trading information must be met for conducting a trade. The orders submitted by traders are classified into two major types, *i.e.*, market orders and limit orders. The market orders can be executed immediately at the current available price. The price for trading a market order is called the trade price. In contrast, the limit orders specify the worst price that traders can accept. Thus, not until the trade price reaches to the acceptable prices are the limit orders executed. If a limit order is placed at a price level that the trade price cannot arrive at or the limit order is cancelled by its trader in advance, the limit order would not be executed. Before execution or cancellation, the limit orders are stored in the order book as quotes, where buy (sell) limit orders are located at the side of bids (asks). If different limit orders have the same accepted price, they would be placed at the same price level. For a fixed price level, the available volume is the sum of the volumes of all those orders.

The order book, as shown in Fig. 1.3, is visible for all traders to guarantee that they have the same information of asks and bids. In the order book, the highest bid price of buy limit orders is regarded as the best bid, and the lowest ask price of sell limit orders is regarded as the best ask. The best bid and the best ask are also called the best quotes. A buy limit order with a price higher than or equal to the best ask or a sell limit order with a price lower than or equal to the best bid is marketable immediately. Such limit orders would not list in the order book. Thus, the best ask price is always larger than the best bid price. More precisely, at the moment of time t, the difference between the best ask $a(t)$ and the best bid $b(t)$ is the spread

$$s(t) = a(t) - b(t) \ , \tag{1.1}$$



Bid	Bid Volume			Ask Volume	Ask
17.070	14,798	🟩	🟥	10,173	17.080
17.065	16,195	🟩	🟥	19,316	17.085
17.060	24,836	🟩	🟥	19,235	17.090
17.055	13,850	🟩	🟥	19,150	17.095
17.050	16,220	🟩	🟥	17,315	17.100
17.045	10,057	🟩	🟥	11,160	17.105
17.040	21,033	🟩	🟥	30,558	17.110
17.035	6,002	🟩	🟥	9,181	17.115
17.030	18,455	🟩	🟥	24,544	17.120
17.025	10,647	🟩	🟥	5,921	17.125

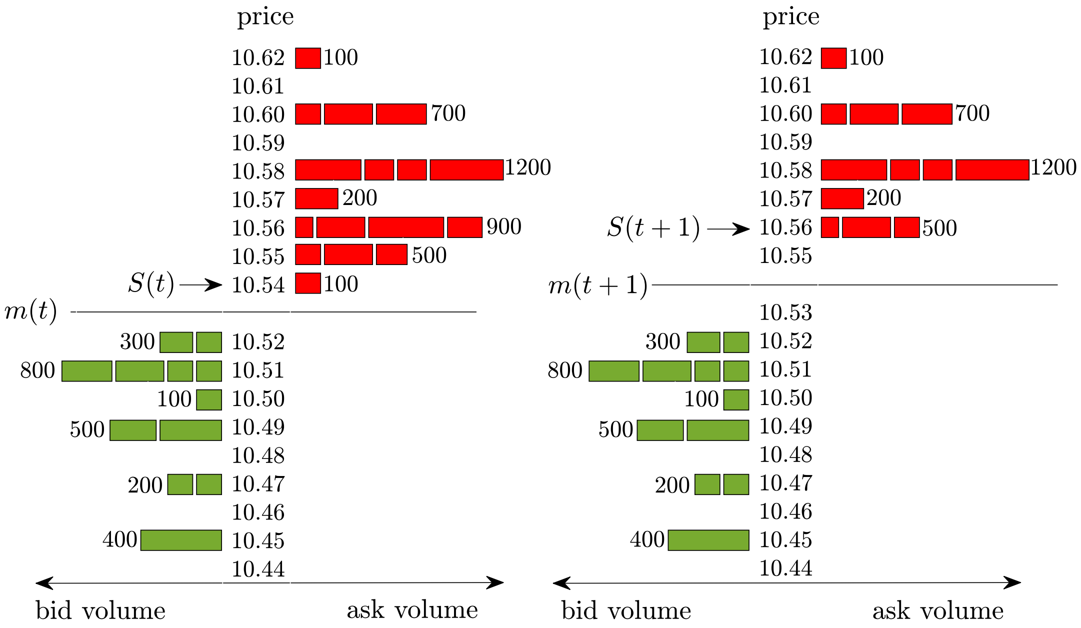
Figure 1.2: An example of order book. On the left hand side are the quotes to buy (bids), on the right side the quotes to sell (asks). The similar order books for different stocks can be found in Xetra stock exchange.

which is always positive. The bid-ask spread to some extent quantifies the market liquidity [9, 15, 44, 51, 139, 166], which characterizes the ability to trade an asset without significant change of the asset's price. In the stock market, a high liquid stock has a small spread, because the price levels near the best ask and the best bid are highly occupied by limit orders. In contrast, a low liquid stock has a bigger spread due to the low-density limit orders at the the price levels near the best quotes. As a result, a small market order can easily enlarge the spread of low liquid stocks. In fact, the spread measures the cost of an instantaneous round-trip of one share, *i.e.* a buy instantaneously followed by a sell. This cost for making transactions without time delay are also termed as the liquidity cost [15, 51, 166]. As a part of transaction costs, the liquidity cost is distinguished from the fixed brokerage commission. Another quantity to quantify the market liquidity is the market depth [44, 139], which is measured by the available volume at the best quote. If the available volume is high, the market depth is large and the best quote price is difficult to be shifted.

1.3.2 Continuous double auction

According to the characters of traders, the impatient traders favour submitting market orders for trading immediately at the available price, while the patient traders prefer to submit limit orders for a better price than the current trade price at the cost of time delay for execution. The market orders match against the limit orders of the opposite type (buy or sell) by the principle of primary price priority and secondary time priority. A market order leads to a buy (sell) trade for buying (selling) a given volume. The market orders will consume the volume at the best quotes for meeting the demands of traders. As a result, the best quote prices either remain unchanged or move to the next best prices.

Take an example. Fig. 1.3 shows a snap shot of the order book before and after executing a buy market order. As shown in the figure, there are 100 shares available at the best ask at the time t. If the market order is only for buying 100 shares, it will be fully executed with a trade price of 10.54 dollars and consume the volume at the best ask. However, if the market order is for buying 1000 shares, the case will be different. In particular, the market order will be executed partly at the price of 10.54 dollars for 100 shares. To fulfil the demand of 1000 shares, the market order will continue being executed



(a) order book at time t (b) order book at time $t+1$

Figure 1.3: Snap shot of the order book before and after a buy market order with a volume of 1000 shares is executed. At each price level, a bar stands for an limit order. The width of the bar indicates the volume that will be traded. We notice that both the trade price $S(t)$ and the midpoint price $m(t)$ move up after the execution of that market order.

at the price of 10.55 dollars for 500 shares and at the price of 10.56 dollars for the rest 400 shares, respectively. Consequently, this large market order changes the trade price from 10.54 dollars at the time t to 10.56 dollars at the time $t+1$. Such a price change due to a trade is termed as the price impact, which will be introduced in detail in Sec. 1.3.5.

Due to the incoming market orders, the trade price changes persistently. To make profits from the price difference between ask and bid, the traders submit limit orders. The submission of limit orders causes an anti-persistent change of the trade price. A balance between persistence and anti-persistence in the price change determines the final price which moves diffusively like a random walk [27]. Therefore, the interaction between market orders and limit orders leads to price formation. The detailed mechanism for price formation is referred to as the continuous double auction [35, 43, 59, 95, 151].

Since the trades for buying and selling occur at the best ask and the best bid, respectively, the trade price fluctuates between them. A better way to trace the price tendency is to look at the midpoint price, which is defined as the midpoint between the best ask $a(t)$ and the best bid $b(t)$,

$$m(t) = \frac{1}{2}\big[a(t) + b(t)\big] \ . \tag{1.2}$$

As mentioned above, the best ask price is either unchanged or pushed up by buy market orders, while the best bid price is either unchanged or dropped down by sell market orders. Accordingly, the midpoint price could be unaltered, and also could be raised by buy market orders or lowered by sell market orders.

1.3.3 Prices, returns and volatility

In 1990, Louis Bachelier proposed that the stock price has a nature of random walks [12, 13]. For instance, Fig. 1.1 (a) shows the time evolution of prices for Citigroup, Apple and Goldman Sachs. We find that the price evolution is rather irregular with strong fluctuations. In terms of stochastic processes, Bachelier [12, 13] modelled the stock price as Brownian motion, also known as Wiener process [164],

$$dS(t) = \mu dt + \sigma\eta\sqrt{dt} \ . \tag{1.3}$$

The equation above includes a deterministic part μdt and a stochastic part $\sigma\eta\sqrt{dt}$, where μ is a drift that measures the average growth of the random variable, η is a random variable that is independent of infinitesimal time step dt, and σ is the volatility of the price $S(t)$. Bachelier's model captures the randomness of prices, but it allows negative prices. In 1959, Osborne modelled the stock price as a geometric Brownian motion [122],

$$dS(t) = \mu S(t)dt + \sigma\eta S(t)\sqrt{dt} \ . \tag{1.4}$$

In this model, the prices are log-normally distributed and thus always positive, while the differences of logarithmic prices are normally distributed. By comparison, the geometric Brownian motion is much more realistic to describe stock prices, and has been extensively applied to economic modelling.

In Fig. 1.4 (a), we notice that the prices for the three stocks are quite different. Intuitively, the absolute prices $S(t)$ as well as the absolute price changes $\Delta S_k(t)$ are not suitable for statistical analysis of multiple stocks. To make the price changes of different stocks on the same footing, a typical way is to use relative price changes $r_k(t, \Delta t)$,

$$r_k(t, \Delta t) = \frac{\Delta S_k(t)}{S_k(t)} = \frac{S_k(t + \Delta t) - S_k(t)}{S_k(t)} \ . \tag{1.5}$$

Here, $r_k(t, \Delta t)$ is the so-called return, t is the initial time of the return and Δt is the return interval. As shown in Fig. 1.4 (b), the returns of the three stocks are comparable and fluctuate around zero. The positive and negative returns represent the profits and losses, respectively, while zero returns indicate neither profits nor losses. Also, we can express the returns with the exponential growth in prices, namely

$$\widetilde{r}_k(t, \Delta t) = \log S_k(t + \Delta t) - \log S_k(t) = \log \frac{S_k(t + \Delta t)}{S_k(t)} \ . \tag{1.6}$$

In fact, when the return interval Δt is small so that $\Delta S_k(t)/S_k(t) < 1$ holds, the logarithmic returns are approximately equivalent to the returns in Eq. (1.5), namely

$$\widetilde{r}_k(t, \Delta t) = \log\left(1 + \frac{\Delta S_k(t)}{S_k(t)}\right) \approx \frac{\Delta S_k(t)}{S_k(t)} = r_k(t, \Delta t) \ . \tag{1.7}$$

The variation of prices over time is measured by volatility. More importantly, the volatility can be used to quantify the risk of a stock k. As the volatility is unobservable on markets, it is often estimated by the standard deviation of returns,

$$\sigma_k = \sqrt{\left\langle r_k^2(t, \Delta t)\right\rangle_t - \left\langle r_k(t, \Delta t)\right\rangle_t^2} \ , \tag{1.8}$$

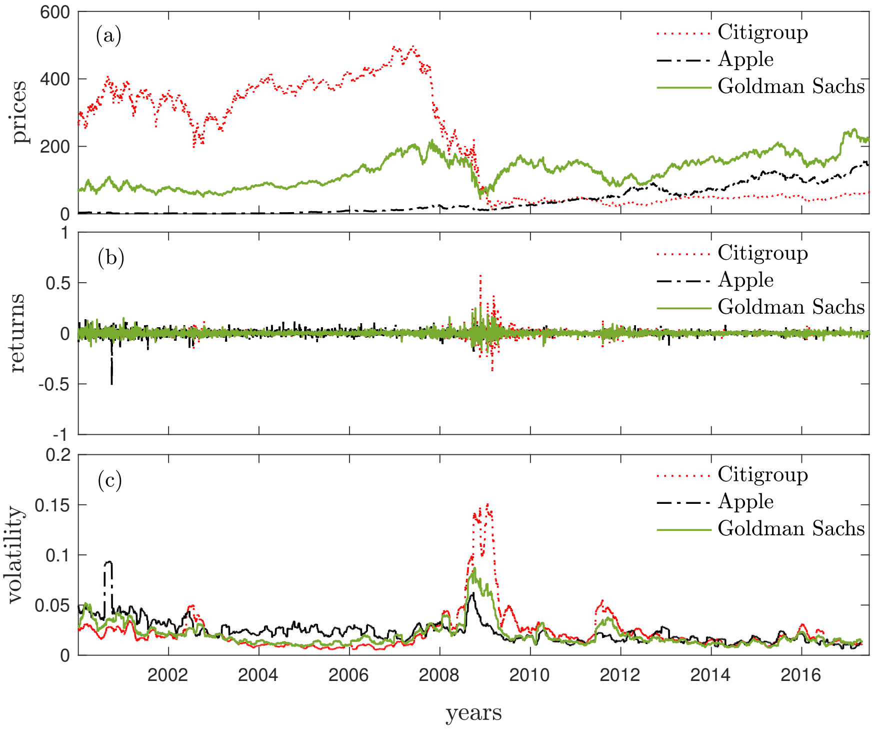


Figure 1.4: (a) Daily closing prices, (b) daily returns with $\Delta t = 1$ day, and (c) volatility estimated on a 40-days time window for Citigroup, Apple, and Goldman Sachs from January, 2000 to June, 2017.

where $\langle \cdots \rangle_t$ represents the average over all times t in a considered time window T. Thus, the average of returns can be written as

$$\big\langle r_k(t, \Delta t) \big\rangle_t = \frac{1}{T} \sum_{t=1}^{T} r_k(t, \Delta t) \ . \tag{1.9}$$

According to Eq. (1.8), the volatility depends on not only the return interval Δt, but also the estimation horizon T. The empirical studies [21, 144] showed that the volatility of returns is highly fluctuating rather than stationary. As seen from Fig. 1.4 (c), when we move the time window of $T = 40$ days, the volatility of returns for each stock fluctuates over time. Specifically, the volatility is high when the price changes dramatically over a short time interval, whereas the volatility is low when the price changes at a steady pace over a period of time. Furthermore, an early study [106] revealed that the volatilities are clustered and the autocorrelation of volatilities decays slowly with time. Such stylized fact is termed as the volatility clustering, which was first observed by Mandelbrot in 1963 [106].

1.3.4 Correlation of order flow

For a stock k, the order flow can be represented by a time series of trade signs $\varepsilon_k(t)$, where $\varepsilon_k(t) = +1$ stands for a buy trade and $\varepsilon_k(t) = -1$ stands for a sell trade. Thus, the correlation of order flow is expressed as

$$\Theta(\tau) = \Big\langle \varepsilon_k(t+\tau)\varepsilon_k(t) \Big\rangle_t \ . \tag{1.10}$$

In recent years, high auto-correlations of order flow have been found and analyzed [27, 55, 69, 98, 100, 114, 155]. The order flow exhibits a remarkable persistence that buy (sell) orders are often followed by more buy (sell) orders [27, 98]. Such behavior is normally termed as the long memory of order flow.

The long-memory correlation of order flow is probably due to order splitting [26, 27, 155] and herding behavior [155]. The order splitting refers to that a large order is chopped into pieces, which are executed one by one with the same trade sign. As introduced in Sec. 1.3.2, a large market order moves the trade price largely, leading to an extra cost for trading. Taking advantage of order splitting, traders can avoid large trading costs. The herding behavior in financial markets refers to that different traders place orders with the same trade sign. It could be yielded by public information or the imitations between traders. By comparison, Ref. [155] reveals that the persistence in order flow is mainly due to the order splitting rather than the herding behavior.

The long-memory correlation of order flow, described by a power-law function [27, 98], implies that the signs of future orders are highly predictable if we know the signs of past orders. As a result, we can predict the trade prices. This may lead to an inefficient market, unless this correlation is compensated by other quantities. In fact, this correlation is compensated by a bare impact function decaying with time [27]. Regardless of the decaying time of impacts, this correlation is also compensated by anti-correlated fluctuations in traded volumes and liquidity [98]. In consequence, the correlation of order flow is closely related to the price change caused by trades.

1.3.5 Price impact and response

The price change, on average, conditioned on initiating a trade with a given size and a given sign (buy or sell) is termed price impact [27, 58, 114]. A buyer-initiated (seller-initiated) trade refers to a market order for buying (selling). As market orders will consume the volume at the best quote, a buyer-initiated trade is expected to raise the price while a seller-initiated trade to lower the price. Thus, the price impact can be viewed as the price response to trades. Due to the price impact, a large market order may be executed at several different prices, rather than a constant price at which a trader initially intends to trade. Thus, the practical trade price may differ from the initial trade price. The difference between the two prices reflects the extra cost for trading, namely the price impact cost. This cost can largely affect the profits of traders. To reduce it, a lot of optimal execution strategies of orders have been proposed [5, 6, 65–67, 120]. In those strategies, the cost arising from price impacts is regarded as an important determinant. Either for such practical purpose or for academic exploring, an essential question centres on how to understand and quantify the price impact.

Through most studies, the mechanism of price impacts is obscure, but is often interpreted by three primary causes: short-run liquidity, traded volumes and information [38, 59, 61, 84, 89, 114, 162]. As introduced in Sec. 1.3.1, for a low liquid stock, the price levels near the best quote are less occupied by limit orders. As a result, only a few volumes near the best quote are available. To immediately meet large demands, that is, buying or selling for a large volume, traders have to lower their sell prices or raise their buy prices to obtain more available volumes [38]. Trade prices are thus moved to a far place before new limit orders coming. Hence, a lack of short-run liquidity will lead to the price change.

The volumes of market orders have an important impact on the trade price [99, 135]. Intuitively, large traded volumes will move the price to a large extent, leading to a large price impact. However, numerous studies [27, 59, 89, 114, 162] reveal that large price changes are, on average, not the result of large traded volumes. Due to the small liquidity, a large market order indeed shifts the price obviously, resulting in heavy tails of the return distributions [142]. No matter whether the large market orders are driven by valid information or not, they are always thought to carry specific information and thus signal traders to adjust their strategies. The traders, who intend to submit large market orders but would not like to be noticed by others, want to conceal their trading information. For this end, they employ the strategies such as splitting a large order into small pieces [26, 27, 155]. These small orders hidden by other random small orders move the price little by little without being detected. They, on the one hand, largely reduce the liquidity cost, and, on the other hand, generate the long-memory correlation of order flow [27, 98]. As a result, the small orders that are split from a large market order lead to the large price change on average [27].

The effect of information on the price has been noticed for a long time [61, 71, 75, 79, 89]. It can date back to 1970 when Fama proposed his famous theory—Efficient Market Hypothesis (EMH) [54]. The EMH states that all available information is processed and encoded in the current prices. It means that everyone has the equal information about the market so that any (statistical) arbitrage opportunity is absent. The available public information, for instance, the idiosyncratic news or market-wide news, affects the price temporarily, and does not generate large price change in total [89]. Besides, the private information is also incorporated in the price by trading. Comparing with the public information, the private information is more likely to induce the price impact, moving the price permanently [38].

1.3.6 Two-phase behavior in demands

For a trade, the volume of an individual stock is either bought in or sold out. To signal the trade direction, we use the signed traded volume, where the volume bought is positive and the volume sold is negative. In a certain time interval, the sum of all the signed traded volumes, *i.e.*, the difference between all bought-in volumes and all sold-out volumes, is termed the volume imbalance. It quantifies the demand which drives the buying and selling of the market. Recently, a two-phase behavior of demands in financial markets was discovered by Plerou, Gopikrishnan, and Stanley (PGS) [132]. The two-phase behavior, as shown in Fig. 1.5, is characterized by a transition from the unimodal distribution of volume imbalances to a bimodal one. Such transition is due to the change of local noise intensity, which is defined as the absolute value of fluctuations around the average of volume imbalances in a certain time interval. When the local noise intensity Σ is smaller than the critical value Σ_c, the distribution $P(\Omega|\Sigma)$ has only one peak. Once the local noise intensity Σ goes beyond the critical value Σ_c, the distribution $P(\Omega|\Sigma)$ shows up two peaks, which are located symmetrically around zero.

To find out the causes of the two-phase behavior, Matia and Yamasaki (MY) [110] estimated the volume imbalance and local noise intensity using Trades and Quotes data set and a numerical simulation, respectively. They found that the number of phases

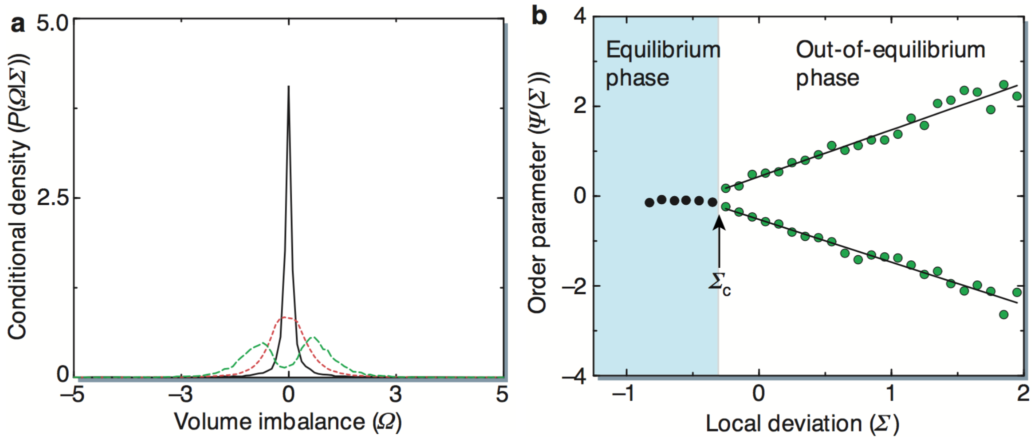


Figure 1.5: (a) Conditional density $P(\Omega|\Sigma)$ versus the volume imbalance Ω for varying local noise intensity Σ, (b) order parameter ψ, *i.e.*, positions of the maxima of $P(\Omega|\Sigma)$, versus Σ. The figures are taken from Ref. [132].

is related to the distribution of traded volumes. If the generated time series of traded volumes follows a Gaussian distribution, only one phase, *i.e.*, the unimodal distribution of volume imbalances, could be observed. In contrast, if this time series follows a fat-tailed distribution, two phases, *i.e.*, the unimodal and bimodal distributions of volume imbalances, could be observed. As a result, the two-phase behavior is more likely due to fat-tailed distributions of traded volumes. A similar conclusion is proposed by Potters and Bouchard (PB) [134]. They found that the local noise intensity is positively correlated with the magnitude of the volume imbalance, as long as traded volumes have fat tails. This correlation leads to the volume imbalances symmetrically distributing around zero. Thus, they argued that the two-phase behavior results from the statistical properties of traded volumes. In line with MY and PB's interpretations, PGS [133] further elucidated the significance of the two-phase behavior in terms of the price change. More specifically, large traded volumes, which imply large demands for buying or selling, will result in the large fluctuations of prices around the local equilibrium values.

In addition to above studies, the two-phase behavior is also interpreted by agent-based models [146, 147], minority games [167], herding models [167], and others [101]. Besides, the studies on the two-phase behavior have been extended from stock markets to future markets [85, 101], option markets [137] and financial indices [86].

1.4 Outline of the thesis

So far, we have had a basic knowledge about the financial markets and the market microstructures. Due to the revolutionary changes in the trading technology, the microstructures of financial markets have dramatically changed in the last decade. Numerous studies have devoted to the market microstructure, but the landscape of this field is not fully disclosed. Particularly, many studies are confined to single stocks, but the financial market is a rather complex system that companies as well as their stocks do not exist independently. On the other hand, due to the development of high-frequency trading, a gigantic amount of transaction data is available. With huge data, we thus can carry out the statistical analysis either for individual stocks or for the market as a whole. Further, it becomes possible to disclose latent stylized facts, accurately test some theories, and even construct more realistic models for practical applications. In this thesis, we will focus on the market microstructure by considering the financial market as a whole. In particular, we will restrict ourselves to the following key areas: correlations of order flow, price impacts, and the two-phase behavior in demands.

The thesis is organized as follows.

In Chapter 2, we introduce two data sets, a Trade and Quote (TAQ) data set and a TotalView-ITCH data set, where the TAQ data set is used for analyzing the empirical cross-responses in Chapters 3 and 4. We also set the time convention for all empirical analyses in this thesis. In addition, due to the low resolution of TAQ data set, it is difficult to employ the previous algorithm [97] to identify the signs for the trades during one-second intervals. We thus develop a method to figure out them with TAQ data set and test this method with TotalView-ITCH data set.

In Chapter 3, we empirically analyze the cross-response as well as the corresponding correlation of trade signs for pairs of stocks, and present a complete view of the response in the market as a whole. We also compare the self- and cross-responses in terms of time scales, price impacts, and trade sign correlations.

In Chapter 4, we perform the averages of cross-responses for an individual stock. By doing so, the drastic fluctuations at large time lags in the cross-responses are wiped out to some extent. Correspondingly, we also carry out the average of cross-correlations of trade signs. According to the average cross-responses, we further identify the influencing and influenced stocks and analyze the role of trading frequency in response functions. To find out the microscopic mechanism of cross-responses, once more, we compare the self- and average cross-responses.

In Chapter 5, we extend the price impact model of Bouchaud et al. [27]. The extended model aims to interpret the empirical results in Chapters 3 and 4. It includes a self-impact function and a cross-impact function. To quantify the self- and cross-price impacts that are difficult to be obtained from empirical data, we propose a construction to fix the parameters, and employ a price diffusion function to corroborate the parameters. By doing so, we can quantify and interpret the price impacts of individual stocks.

In Chapter 6, we extend the framework of trading strategies of Gatheral (2010) [65] from single stocks to two-dimensional cases. Our trading strategy is used for executing two round-trip trades of two stocks, where the cross-impact cost is an important determinant. We apply the model of the strategy to a pair of stocks and display how the cross-impacts affect the trading strategy.

In Chapter 7, we define the demand as the volume imbalance, *i.e.*, the sum of all the signed traded volumes in a fixed time interval, and analyze the dependence of demands between stocks using a copula method. Furthermore, we quantify the local fluctuation of the signed traded volumes by a local noise intensity, and investigate how the large local fluctuations affect the dependence of demands.

In Chapter 8, we conclude with all our findings and give a view of further study.

Chapter 2

Data sets, time conventions and trade signs

2.1 Introduction

Nowadays, data has been the core of many fields, including the econophysics. By statistical analyses of a large-scale of real-time data, a lot of stylized facts can be disclosed. The studies in this thesis begin with the analysis of empirical data, shown in Chapters 3 and 4. Besides, such empirical analyses as parts of studies are also included in other chapters. In general, the empirical results to some extent depend on how to process the data. However, the data processing is influenced by how the data is organized in files. For instance, the resolution of data determines which data points can be available; the type of data determines which information the data provides.

In financial markets, the real-time data from trading is recorded in data sets. To have a basic knowledge of data sets that we used, we give an introduction for them in Sec. 2.2. To facilitate the data processing in the following chapters, we set the time conventions in Sec. 2.3, where a physical time scale and a trading (event) time scale are discussed. Furthermore, making use of two different data sets, we develop a method to classify trade signs in Sec. 2.4 and test this method in Sec. 2.5. We conclude this chapter in Sec. 2.6. The classifications of trade signs in Secs. 2.4 and 2.5 refers to Ref. [161].

2.2 Data sets

The data sets used in this thesis include a Trade and Quote (TAQ) data set and a TotalView-ITCH data set. Our studies are mainly based on the TAQ data set. However, the disadvantage of this data set is at the resolution of transaction data, *i.e.*, the minimal distinguished time of transaction data. It is one second for the TAQ data set. For a higher resolution, we have to resort to the TotalView-ITCH data set, of which the timestamp is on the scale of milliseconds. In addition to the resolution, the two data sets differ from each other in terms of the following aspects.

2.2.1 TAQ data set

The Trade and Quote (TAQ) data set, obtained from New York Stock Exchange, contains intraday transaction data for all the stocks listed on the New York Stock Exchange, American Stock Exchange, Nasdaq National Market System and SmallCap issues. For each stock, the intraday transaction data is recorded in two individual files, a trade file and a quote file.

The trade file lists the information of all successive transactions of market orders, as shown in Table 2.1. In our studies, we focus on the dates, time, prices, volumes, ignoring other quantities in Table 2.1, *i.e.*, G127, CORR and COND. Here, G127 is an indicator that combines "G", Rule 127, and stopped stock trade, CORR is a correlation indicator, and COND indicates a sale condition [1]. In each trading day, the time intervals between two continuous trades are different. They may be a few minutes or several seconds or even several milliseconds. Thus, during an one-second interval, more than one trades may occur and are recorded in the trade file. On the other hand, all the quote information for successive buy and sell limit orders is stored in the quote file, shown in Table 2.2. Here, we ignore the information of Mode which indicates the quote condition [1]. In Table 2.2, "Ask" and "Bid" refer to the best ask price and the best bid price, respectively. Meanwhile, "Ask size" and "Bid size" are the volumes at the best ask and the best bid. The volume in the trade file is in the unit of shares, but the volume in the quote file is in the unit of lots. To make them in the same unit, we have to rescale the quote volume by 1 lot=100 shares. It is worth mentioning that there may be several quotes or no quote during an one-second interval.

For the TAQ data set, the resolution of transaction information is one second. During the time interval of one second, if there are several trades recorded in the trade file and several quotes recorded in the quote file, it is difficult to distinguish the chronological order of trades and quotes in two individual files. In this case, it is also unlikely to match each trade with the preceding quote. Moreover, the recorded transaction information can be traced from 7:00:00 to 20:00:00 in Eastern Standard Time (EST), because the TAQ data set contains the information of pre- and post-market trades. Here, the pre-market trades occur before the opening of the market at 9:30:00 EST, and the post-market trades occur after the closing of the market at 16:00:00 EST.

In general, during the opening and closing time, the market exhibits large fluctuations due to the effects of overnight information and artifacts. When using the TAQ data set, we always ignore the first and last ten minutes of the market trading time in this thesis. In addition, the transaction information in the TAQ data set comes from different exchanges, but we only use the information from NASDAQ stock market in the year of 2008.

2.2.2 TotalView-ITCH data set

The TotalView-ITCH data set has a higher resolution than the TAQ data set. The timestamp of intraday transaction data can be accurate to one millisecond, as shown in Table 2.3 for Microsoft Corp. on Jan. 14th, 2008. The time in the table records the number of milliseconds after the midnight, converted by

$$\text{time} = \Big[(h \text{ hours} \times 3600 + m \text{ minutes} \times 60 + s \text{ seconds}) \times 1000 + x\Big] \text{ milliseconds} \ . \tag{2.1}$$

The TotalView-ITCH data set contains the information of displayed orders. The executions of most non-displayed orders, such as market orders, are not recorded in this data

Table 2.1: An example of Microsoft Corp. from the trade file in the TAQ data set

Date	Time	Price (dollars)	Size (shares)	G127	CORR	COND
2008-01-09	12:28:30	34.30	200	0	0	@F
2008-01-09	12:28:32	34.29	200	0	0	
2008-01-09	12:28:32	34.29	600	0	0	
2008-01-09	12:28:37	34.30	700	0	0	
2008-01-09	12:28:37	34.30	250	0	0	@F
2008-01-09	12:28:40	34.30	100	0	0	
2008-01-09	12:28:40	34.30	100	0	0	
2008-01-09	12:28:40	34.30	700	0	0	
2008-01-09	12:28:40	34.30	100	0	0	
2008-01-09	12:28:40	34.30	200	0	0	

Table 2.2: An example of Microsoft Corp. from the quote file in the TAQ data set

Date	Time	Ask (dollars)	Bid (dollars)	Ask size (lots)	Bid size (lots)	Mode
2008-01-09	12:28:36	34.2900	34.3000	20	91	12
2008-01-09	12:28:37	34.2900	34.3000	13	91	12
2008-01-09	12:28:37	34.2900	34.3000	10	91	12
2008-01-09	12:28:37	34.2900	34.3000	10	92	12
2008-01-09	12:28:37	34.2900	34.3000	12	92	12
2008-01-09	12:28:38	34.2900	34.3000	14	92	12
2008-01-09	12:28:40	34.2900	34.3000	23	92	12
2008-01-09	12:28:40	34.2900	34.3000	24	92	12
2008-01-09	12:28:40	34.2900	34.3000	5	92	12
2008-01-09	12:28:40	34.2900	34.3000	6	92	12

Table 2.3: An example of Microsoft Corp. from TotalView-ITCH data set

Time	Ticker	Order ID	Type	Shares	Price	MPID	X
50543186	MSFT	173842357	S	400	342900		Q
50543311	MSFT	173835197	F	0	0	JPMS	Q
50543311	MSFT	173835199	F	0	0		Q
50543311	MSFT	173835227	E	300	0		Q
50543312	MSFT	173843928	S	500	342700		Q
50543314	MSFT	173843952	B	100	342500		Q
50543314	MSFT	173843958	B	100	342500	JPMS	Q
50543692	MSFT	173026550	D	0	0		Q
50543693	MSFT	173847899	S	500	342700		Q
50543693	MSFT	173847900	S	1800	342700		Q

set. Each displayed order contains a message, given in the fourth column of Table 2.3, where the type of the message is marked by B, S, E, C, F, D, X or T. Here, B indicates that a buy limit order is added to the order book, while S indicates a sell limit order is added. When an order with the message of S or B is issued, it will be given a new and unique ID, as listed in the third column of Table 2.3. Meanwhile, the volume and the price of this order are recorded in the fifth and sixth columns of the data set, respectively. If this order is then executed or cancelled in part, the message will be marked by E or C. The executed or cancelled volume is displayed in the data set, but the corresponding price is shown as zero. If this order is executed or cancelled in full, the type of the message turns to F or D. As the order will not be available any more in this case, the volume and the price are displayed as zero. In the TotalView-ITCH data set, it is worth paying attention to the price that is in the unit of 10^{-4} dollars. The type of X represents the cross event with a bulk volume and the type of T indicates the execution of non-displayed orders. For the orders marked by X and T, it is difficult to identify their trade signs. Thus, these two types of orders are often excluded in our studies. In addition, the specifications of eight types of messages are given by TradingPhysics.com [156], where we obtain the TotalView-ITCH data set.

In our study, we focus on the transaction information of the first six columns in the TotalView-ITCH data set, where the ticker denotes the symbol of a stock. The last two columns, MPID and X, are ignored. Here, MPID represents the market participant ID associated with the transaction. However, not every transaction is given a market participant ID. Due to a lack of specifications for the last column, X is a unknown quantity.

In contrast with the TAQ data set, the TotalView-ITCH data set provides the detailed information of order flow with a higher resolution of data. However, it does not contain the information of the best quotes and the trades, which can be obtained conveniently from the TAQ data set. Thus, our empirical analysis in the following chapters are mainly based on the TAQ data set. The TotalView-ITCH data set will be used to test the accuracy of the trade sign classification in Sec. 2.5.

2.3 Time conventions

The empirical studies in the following chapters will centre on the price response, *i.e.*, the average price change caused by trades. Here, the trade is indicated by the sign of a trade, which is identified from the trade information. The price changes is represented by the change of the midpoint prices before and after this trade. The midpoint price can be obtained from quote information. For the TAQ data set, the trades and quotes are recorded in two individual files. Previous studies on the price self-response were carried out on a trading time scale [24, 27, 28, 52, 69, 98, 99, 154]. The trading time scale, *i.e.* the event time scale, regards every transaction as a timestamp. Due to the resolution of one second in the TAQ data set, it is difficult to match each trade with the preceding quote during an one-second interval. Thus, the studies in Refs. [98, 99] treated all the trades in the same time interval as a single trade by lumping together all these trades, and then considered one lumped trade as one trading time step. Such treatment is possible for single stocks, as the trading time is unique for each stock. When considering the price cross-response across different stocks, however, the trading time varies greatly in different stocks.

In view of this, we carry out the empirical cross-responses on a physical time scale instead of the trading time scale. The physical time scale maps the transaction information into a real and discrete time axis with the unit time as the time step. We set the unit time to one second, *i.e.*, the minimal distinguished time in the TAQ data set. When reviewing the trading time scale, we find that two continuous trades as two continuous trading time steps may occur with a time interval of several seconds or dozens of minutes. Superficially, the price is impacted by the last second trade, regardless of the last trade that leads to the instantaneous price change. In fact, the price may be impacted not only by the last second trade but also by other factors, such as news, during the long time interval. In contrast, if we use the physical time scale, we are able to trace the real-time price response.

2.4 Classification of trade signs

A widely used approach to classify trade signs was put forward by Lee and Ready [97]. By comparing the trade price with the preceding midpoint price, the approach can correctly classify at least 85% of all trades. A mathematical form for the trade sign can be expressed by

$$\varepsilon(t;n) = \mathrm{sgn}\big(S(t;n) - m(t;n)\big) \ , \tag{2.2}$$

where the sign $\varepsilon(t;n)$ of n-th trade in the time interval t is worked out by a sign function sgn of the trade price $S(t;n)$ and the preceding midpoint price $m(t;n)$. For the TAQ data set, it is unlikely to match each trade with the preceding quote during the one-second intervals. Another approach for sign classification was proposed by Holthausen et al. [83]. They defined a trade as buyer-initiated if the trade is executed at a price above the prior price, and as seller-initiated if the trade occurs at a price below the prior one. It is expressed mathematically as

$$\varepsilon(t;n) = \mathrm{sgn}\big(S(t;n) - S(t;n-1)\big), \quad \text{only for} \ \ S(t;n) \neq S(t;n-1) \ . \tag{2.3}$$

Since they did not classify the trades which have the same price as the prior one, this approach has the accuracy of 52.8% for sign classification.

In view of the difficulty of Lee and Ready's approach and the incompleteness of Holthausen et al.'s approach, we therefore propose an approach that further extends the one of Holthausen et al.. We first classify the sign for each trade by comparing the current and the prior price,

$$\varepsilon(t;n) = \begin{cases} \mathrm{sgn}\big(S(t;n) - S(t;n-1)\big) & , \ \text{if} \ \ S(t;n) \neq S(t;n-1), \\ \varepsilon(t;n-1) & , \ \text{otherwise.} \end{cases} \tag{2.4}$$

If the current price is higher (lower) than the prior price, the trade sign is defined as $+1$ (-1). Here, we also classify the trades that are ignored by Holthausen et al.. If two consecutive trades having the same trading direction together did not exhaust the available volume at the best quote, the prices of both trades would be the same. Regarding to this, we set the same sign for two consecutive trades that have the same price, *i.e.* the second line of Eq. (2.4).

During the time interval t, the number of trades is denoted by $N(t)$, and the individual trades carried out are numbered by $n = 1, ..., N(t)$. Therefore, we define the trade sign for each time interval of one second by

$$\varepsilon(t) = \begin{cases} \mathrm{sgn}\left(\sum\limits_{n=1}^{N(t)} \varepsilon(t;n)\right) & , \ \text{if} \quad N(t) > 0 \ , \\ 0 & , \ \text{if} \quad N(t) = 0 \ . \end{cases} \tag{2.5}$$

Here, if more than one trade occur in the one-second interval t, we average all the trade signs in this interval. Formally, the first line of Eq. (2.5) also includes the case of $N(t) = 1$. As a result, $\varepsilon(t) = +1$ implies that a majority of trades in the time interval t were triggered by buy market orders, whereas $\varepsilon(t) = -1$ indicates a majority of sell market orders. We have $\varepsilon(t) = 0$ whenever trades did not take place in the time interval t or there was a balance of buy and sell market orders in this interval.

2.5 Tests of sign classifications

Since the TAQ data set does not provide the information of trade types, *i.e.*, buy or sell, we proposed an approach to classify trade signs in the last section. On the other hand, the TotalView-ITCH data set lists the information of order types, from which we can derive the empirical sign of each trade. Therefore, we resort to the TotalView-ITCH data set to test the accuracy of our approach. In Sec. 2.5.1, we introduce how to obtain the empirical trade signs from the TotalView-ITCH data set. We then test the approach for classifying the signs of single trades in Sec. 2.5.2 and the signs of aggregated trades in one-second intervals in Sec. 2.5.3.

2.5.1 Obtaining empirical trade signs

In the TotalView-ITCH data set, the executed orders can be classified as non-displayed orders and displayed orders in the order book. The executed non-displayed orders taking the message of T corresponds to hidden trades. The trade signs of hidden trades are difficult to be identified. To test the accuracy of the trade sign classification, we thus ignore the execution of non-displayed orders. For the displayed orders, by following the order message of E and F, the executed limit orders can be sifted out. Here, we regard an execution of one limit order as a transaction. Through the unique order ID, the limit order can be traced back to the previous information since this order was released in the order book. We thus obtain the execution information of the limit order, such as the type of buy or sell, the volume and the price. When a market order to buy (sell) a certain volume matches with a limit order to sell (buy) the same volume at a price offered by the limit order, a trade occurs. Therefore, as a counterpart of the executed limit order, the market order has the opposite trade type but the same traded volume and trade price comparing with those of the limit order executed simultaneously. By matching, we can derive the trade information of a market order if we know the information of the executed limit order correspondingly. The trade signs of market orders, inferred by this way, are called the empirical trade signs.

2.5.2 Tests for signs of single trades

To test the approach of the sign classification, we use the intraday data of Apple Inc. (AAPL), Goldman Sachs Group (GS) and Exxon Mobil Corp. (XOM), obtained from TotalView-ITCH data set for NASDAQ stock market. For each stock, we randomly select two trading days in 2008. Following the order messages of B and S, all the limit order from 9:40:00 to 15:50:00 EST can be identified as to buy or to sell, where the executed limit orders are identified as buyer- or seller-initiated. We name these trades as the identified trades. As shown in Table 2.4, due to the high-frequency trading, there are more than 10000 trades for each stock in each trading day. The number of identified trades reflects the total number of the available empirical trade signs. For a trade, if the empirical trade sign is the same as the sign $\varepsilon(t;n)$ worked out from the theoretical approach in Eq. (2.4), there is a match. The number of matches divided by the number of identified trades gives the accuracy of the trade sign classification. We carry out the accuracy with respect to all the consecutive trades. For six tested samples, the average accuracy of the sign classification is equal to 85%, listed in Table 2.4. To visualize the accuracy or the ratio of matches, we show an example of AAPL during one minute at January 7th, 2008 in Fig. 2.1

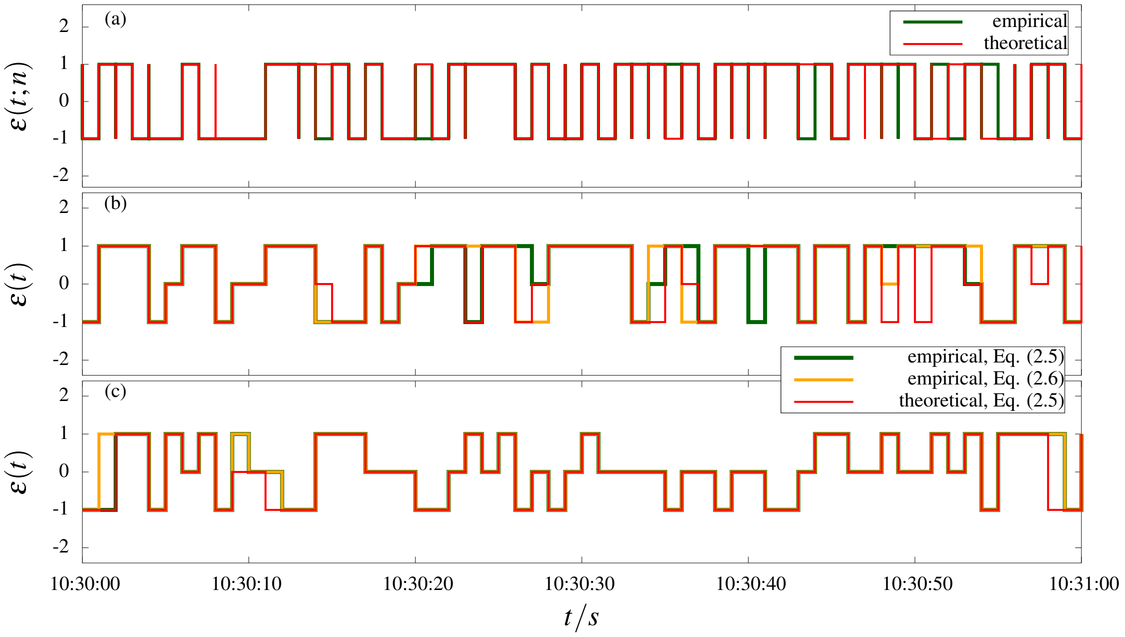


Figure 2.1: Comparisons between empirical and theoretical trade signs versus the physical time for AAPL during one minute. (a) The sign comparison of every trade for AAPL on January 7th, 2008. (b) The sign comparisons of every second for AAPL on January 7th, 2008 with the worst accuracy difference of 5% out of six samples. (c) The sign comparisons of every second for AAPL on June 2nd, 2008. It shows a typical accuracy difference of 2%.

Table 2.4: Accuracy of trade sign classification

Stock Date	AAPL 20080107	AAPL 20080602	GS 20081007	GS 20081210	XOM 20080211	XOM 20080804	6 samples (average)
Sign classification for single trades[1]							
Number of identified limit orders	745020	407843	150532	199224	544451	596882	
Number of identified trades	120287	52691	19942	17902	38455	59580	
Number of matches	103635	47748	16668	15454	30478	49921	
Accuracy of the classification	0.86	0.91	0.84	0.86	0.79	0.84	0.85
Sign classification for one-second intervals [1] [2]							
Total number of identified signs	17115	12180	8283	6853	8782	9590	
Number of matches for Eq.(2.5)	13956	10636	6801	5784	6516	7777	
Accuracy for Eq.(2.5)	0.82	0.87	0.82	0.84	0.74	0.81	0.82
Number of matches for Eq.(2.6)	13256	10302	6715	5690	6446	7603	
Accuracy for Eq.(2.6)	0.77	0.85	0.81	0.83	0.73	0.79	0.80
Total number of $\varepsilon(t) = 0$ found empirically [1]							
Using Eq.(2.5)	6000	10515	14218	15512	13719	12866	12138
Using Eq.(2.6)	5343	10186	14051	15426	13571	12731	11885

[1] The intraday trading time is set from 9:40:00 to 15:50:00 EST (total 22200 seconds).

[2] The simultaneous occurrences of $\varepsilon(t) = 0$ in three kinds of trade signs are excluded.

(a), where the theoretical trade signs nicely match with the empirical ones.

2.5.3 Tests for trade signs of one-second intervals

We also test the trade sign $\varepsilon(t)$ of every second, defined by Eq. (2.5). The trade sign reveals the number imbalance of buy and sell trades in the one-second interval. With the signs of single trades calculated by Eq. (2.4), we work out the trade sign of every second by Eq. (2.5), namely the theoretical trade sign. Correspondingly, with the empirical signs of single trades, we evaluate the aggregated trade signs for one second by the sign of the number imbalance of trades in Eq. (2.5) and by the sign of the volume imbalance of trades [63, 129, 130], given by

$$\varepsilon(t) = \begin{cases} \mathrm{sgn}\left(\sum\limits_{n=1}^{N(t)} \varepsilon(t;n) v(t;n)\right) & , \quad \text{if} \quad N(t) > 0 \ , \\ 0 & , \quad \text{if} \quad N(t) = 0 \ . \end{cases} \tag{2.6}$$

Here, $v(t;n)$ is the traded volume of n-th trade in the time interval t. If a balance of the volumes bought and sold in the one-second interval t was absent, the trade sign $\varepsilon(t)$ resulting from Eq. (2.6) is $+1$ or -1. $\varepsilon(t) = +1$ implies buyer-initiated market orders in t, whereas $\varepsilon(t) = -1$ means seller-initiated market orders. If there was a balance of the volumes bought and sold by market orders, we have $\varepsilon(t) = 0$. In addition, $\varepsilon(t) = 0$ also stems from the absence of trades in t. To make clear how the trade signs are compared, we show a diagram in Fig. 2.2.

As we consider the intraday trading time from 9:40:00 to 15:50:00 EST, there are total 22200 one-second intervals for each trading day. However, in case the accuracy of the sign classification is affected by an excess of zero trade signs, the simultaneous occurrences of $\varepsilon(t) = 0$ in three kinds of trade signs are excluded. Here, three kinds of trade signs include the theoretical signs aggregated by the number imbalance of trades in the one-second interval, the empirical signs aggregated by the number imbalance of trades in the one-second interval, and the empirical signs aggregated by the volume imbalance of trades in the one-second interval. Thus, the remaining trade signs for one-second intervals are used to test the accuracy of the classification Eq. (2.5). Those remaining signs are named as identified signs. As shown in Table 2.4, for each tested stock in each trading day, more

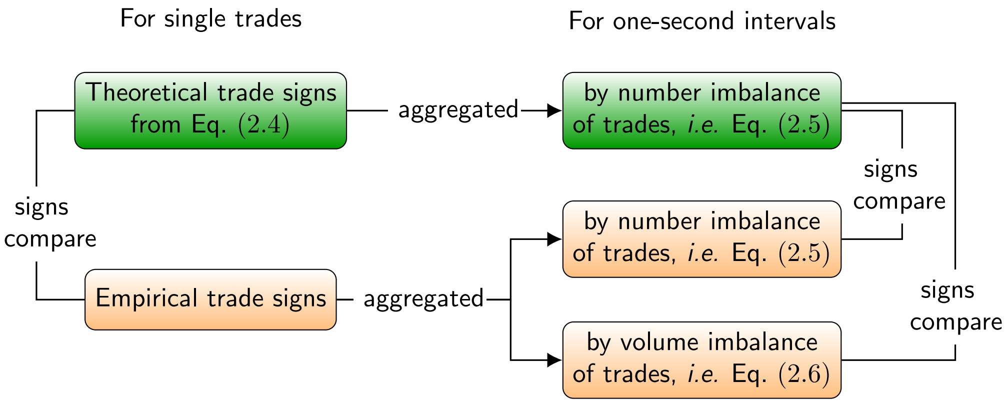

For single trades
For one-second intervals
Theoretical trade signs from Eq. (2.4)
aggregated
by number imbalance of trades, i.e. Eq. (2.5)
signs compare
by number imbalance of trades, i.e. Eq. (2.5)
signs compare
Empirical trade signs
aggregated
by volume imbalance of trades, i.e. Eq. (2.6)
signs compare

Figure 2.2: A diagram for illustrating the comparisons of trade signs

than 6000 identified signs are used to compare with the theoretical signs arising from Eq. (2.5). We count the number of matches if the empirical and theoretical signs for the same time interval are the same. Again, we work out the accuracy of the sign classification through the way that divides the number of matches by the identified signs.

Shown in Table 2.4, the comparison between the theoretical and empirical trade signs, both aggregated from the number imbalance of trades in one-second intervals with Eq. (2.5), manifests an average accuracy of 82% for the trade sign classification over six samples. Very closely, the comparison between the theoretical and empirical trade signs, aggregated from the number imbalance and the volume imbalance of trades in one-second intervals with Eq. (2.5) and Eq. (2.6), respectively, exhibits an average accuracy of 80% over six samples. The two scenarios imply a high correctness for Eq. (2.5) to classify trade signs. For these two scenarios, there is a difference of 2% between the average accuracies. It demonstrates that the different ways to aggregate trades affect sightly the trade signs for one-second intervals. Fig. 2.1 (b) and (c) show the comparisons of three kinds of trade signs for AAPL in one minute. In contrast to other five samples, AAPL on January 7th, 2008 has the worst accuracy difference of 5% between the two scenarios, shown in Fig. 2.1 (b). Even so, the matches of signs still can be found most of the time. Typically, AAPL on January 2nd, 2008 shows the general case with the accuracy difference of 2%, shown in Fig. 2.1 (c), where the three kinds of trade signs match with each other rather well.

2.6 Conclusions

In this chapter, we first introduced two data sets, a TAQ data set and a TotalView-ITCH data set. The TAQ data set has the resolution of one second for transaction data. The transaction data contains the best quote information and the trade information, which are recorded in two individual files for each stock. The TotalView-ITCH data set lists all the order flow information with eight types of messages, submission to buy (B), submission to sell (S), execution in part (E), execution in full (F), cancellation in part (C), cancellation in full (D), cross event with a bulk volume (X), and execution of non-displayed orders (T). As the TotalView-ITCH data set does not provide the information of the best quotes and trades, our empirical studies in the following chapters will use the TAQ data set. In view of our studies across different stocks instead of in single stocks, we considered to carry out the empirical studies on a physical time scale with the time step of one second.

In contrast with the TotalView-ITCH data set, the TAQ data set lacks of the information of trade types, *i.e.*, buy and sell. On the other hand, the previous approaches are not available in our case to classify the trade signs. Therefore, we proposed an approach of the sign classification for single trades and for one-second time intervals, respectively. Further, we tested this approach with empirical trade signs identified by the TotalView-ITCH data set. As a result, this approach can correctly classify 85% signs for single trades and 82% signs for one-second intervals, where the trades are aggregated by the way of the number imbalance. If we aggregate the trades by the volume imbalance in one-second intervals, the accuracy for classifying the signs of one-second intervals is 80%. The accuracy difference of 2% demonstrates that the different ways to aggregate trades have a slight effect on the signs for one-second intervals.

Chapter 3

Cross-responses in correlated financial markets: individual stocks

3.1 Introduction

In the continuous order-driven markets, the price is formed under the mechanism of continuous double auction. A buy market order drives the price up by matching against the sell limit orders in the best ask, while a sell market order drops the price down by matching against the buy limit orders in the best bid. The price change due to a market order or more explicitly a trade refers to the price impact or the price response to trades. Recently, it has received considerable attention [27, 58, 69, 99, 114, 135]. Such price change will lead to an extra cost for trading, *i.e.*, the liquidity cost [15, 51, 166]. To reduce the liquidity cost, a basic strategy is to split a large order into small pieces and executes them one by one over a long time [26, 27, 155]. As a result, the order splitting yields a long-memory correlation of order flow [27, 55, 69, 98, 100, 114, 155], which implies that buy (sell) orders are often followed by more buy (sell) orders. If other factors could not suppress the persistence of order flow, the trade sign in a later time would be predicted. As the price change is related to the correlation of order flow, the future price could be predicted as well. However, the predictability in prices obviously departures from the Efficient Market Hypothesis (EMH) [54] that rules out any arbitrage opportunities. This conflict is reconciled with a time-decaying impact propagator, proposed in Ref. [27]. To recover the market efficiency, the time scale for the impact decaying is crucial, on which, first, the relevant new information arrives and, second, the prices change. Many studies have devoted to the price response to trades, but they are restricted to one single stock, *i.e.*, to the self-response. In this chapter, we go beyond this by investigating the price response between stocks, namely the cross-response.

The chapter is organized as follows. In Sec. 3.2, we describe the data set and the data processing. In Sec. 3.3, with a large-scale of real-time data, we carry out the cross-response as well as the corresponding correlation of trade signs for pairs of stocks. In sec. 3.4, we present a complete view of the response in the market as a whole, identifying several structural characteristics, discussing the efficiency of markets, and shedding light on the price impact across different stocks. To be clear about the cross-response, in Sec. 3.5, we compare the self- and cross-response in terms of time scales, price impacts, and correlations of trade signs. We give our conclusions in Sec. 3.6. This chapter refers to Ref. [161].

3.2 Data description

The study in this chapter is based on the TAQ data set, introduced specifically in Chapter 2, with the transaction data over the whole year of 2008. All the stocks we used are from NASDAQ stock market. In Sec. 3.3, to study the price response across different stocks, we select six companies from three different economic sectors in 2008, listed in Table 3.1. In Sec. 3.4, for the whole market response, we pick out the first ten stocks with the largest average market capitalization from each economic sector of S&P 500 index in 2008, whereas only nine stocks were available for the sector of telecommunications services in that year. Here, the market capitalization is defined as the trade price multiplied with the traded volume, and the average is performed over every trade during the year 2008. We list the selected 99 stocks in Appendix A.1.

When studying the price response across a pair of stocks i and j, we only consider the trading days that both stocks have trades. The setting is for the reason that a lack of trades for either stock in one trading day would not impact on or be impacted by the other stock in the same trading day. As explained in Sec. 2.3 in detail, due to the non-synchronized trading time for different stocks, we employ a real, physical time scale with the time step of one second rather than the trading time scale. Therefore, for every one-second interval, there is an observation. Every observation includes two important quantities of the price cross-response, the midpoint price and the trade sign. We obtain the trade sign in each time interval by the approach of the sign classification, Eq. (2.4) and Eq. (2.5). The midpoint price results from the best quotes. In a time interval, there may be more than one midpoint price; there also may be absent of midpoint prices. For the former, we consider the last midpoint price in the time interval $t-1$ as the prior midpoint price of the trade in t. For the latter, since nothing was triggered to update the best quote, the best quote keeps the same as the previous one. Thus, the time interval that lacks of midpoint prices can be filled by the last available midpoint price before this time interval.

In the sequel, we consider eight pairs of stocks listed in Table 3.1, four within the same economic sector and four across different economic sectors. All the quantities referring to a particular stock carry its index i and referring to a pair of stocks carry two such indices.

Table 3.1: Company information

Company	Symbol	Sector
Apple Inc.	AAPL	Information technology
Microsoft Corp.	MSFT	Information technology
Goldman Sachs Group	GS	Financials
JPMorgan Chase	JPM	Financials
Exxon Mobil Corp.	XOM	Energy
Chevron Corp.	CVX	Energy

3.3 Cross-responses for pairs of stocks

To study the mutual dependences between stocks, we introduce a price cross-response function as well as a trade sign cross-correlator for pairs of stocks in Secs. 3.3.1 and 3.3.2, respectively. Accordingly, we display empirical results of them in these sections. It turns out that the definitions in- and excluding zero trade signs make a difference. We therefore analyze and compare the two definitions in Sec. 3.3.3. In Sec. 3.3.4, we further discuss the response noise, which provides a effective and meaningful lagged time for the response of stock pairs.

3.3.1 Cross-response functions

To measure how a buy or sell trade of one stock indexed by j at time t influences the price of another stock indexed by i at a later time $t+\tau$, we introduce a cross-response function, where we employ the logarithmic price difference or the log-returns with a time lag τ to represent the price change. Via the midpoint price $m_i(t)$, the price change at a given time t is defined by

$$r_i(t,\tau) \ = \ \log m_i(t+\tau) - \log m_i(t) \ = \ \log \frac{m_i(t+\tau)}{m_i(t)} \ . \tag{3.1}$$

It is important to keep in mind the one-second accuracy for time t. To acquire the statistical significance, the cross-response function is the time average of the product of time-lagged returns and trade signs for stocks i and j, respectively, expressed as

$$R_{ij}(\tau) \ = \ \Big\langle r_i(t,\tau)\varepsilon_j(t)\Big\rangle_t \ . \tag{3.2}$$

Here, $\langle ...\rangle_t$ represents the average over all the time t. Eq. (3.2) contains two possible and meaningful definitions, including and excluding the trade signs $\varepsilon_j(t)=0$. The two definitions affect directly on the normalization of results. More specifically, in contrast with the case without zero trade signs $\varepsilon_j(t)=0$, the inclusion of zero trade signs increases the total number of events, which determines the normalization constant for the average of results. However, the events with zero trade signs do not yield any contribution to the cross-response due to $r_i(t,\tau)\varepsilon_j(t)=0$. Therefore, the two definitions exhibit the different statistical significance. The cross-response including $\varepsilon_j(t)=0$ measures the price impact of market orders of stock j diluted by the time without trading, while the cross-response excluding $\varepsilon_j(t)=0$ purely measures the price impact of market orders, ignoring the influence of non-trading time. We will discuss this issue further and give more details in Sec. 3.3.3.

With the Eq. (3.2), we carry out the empirical cross-responses for different stock pairs $(i,\ j)$, shown in Figs. 3.1 and 3.2 for in- and excluding trade signs $\varepsilon_j(t)=0$, respectively. For the response versus the time lag, an increase to a maximum is always followed by a decrease. It implies that the trend of price changes is eventually reversed. This trend does not depend on the stocks of a pair coming form the same economic sector or from different sectors. As an intuitive understanding, the paired stocks from the same sector face the similar systematic risks, yielding a stronger cross-response than the stocks from different sectors. However, the strong cross-response may also occur between the paired stocks from different sectors, such as (GS, AAPL). This may arise from how investors assemble their portfolios, regardless of a variety of reasons. To disperse the investment risks, the

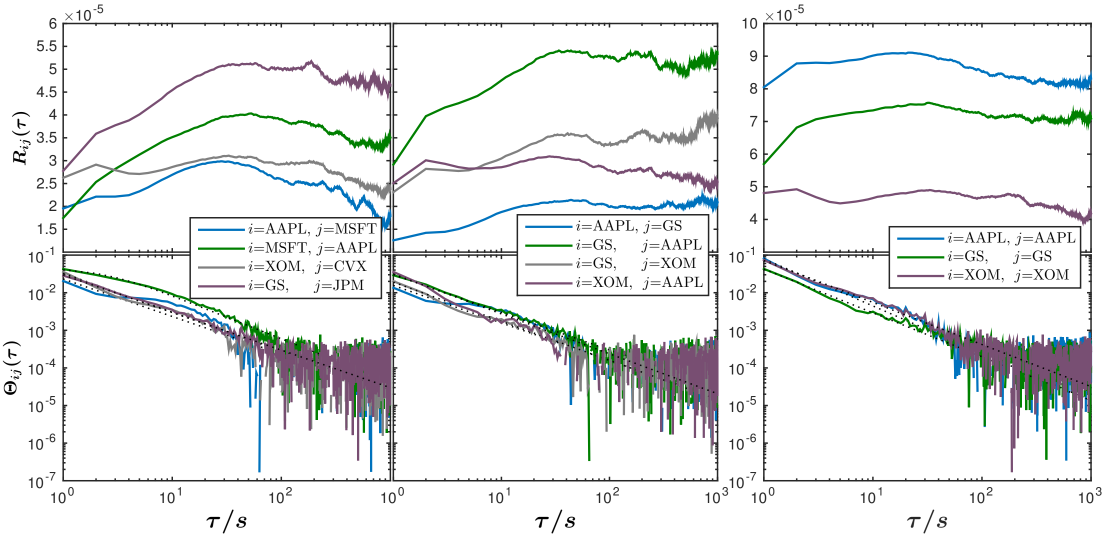


Figure 3.1: Cross-response functions $R_{ij}(\tau)$ including $\varepsilon_j(t) = 0$ in 2008 versus time lag τ on a logarithmic scale (top panels). Corresponding trade sign cross-correlators $\Theta_{ij}(\tau)$ for different stock pairs on a doubly logarithmic scale, fit as dotted lines (bottom panels). The stock pairs in the first column of panels are from the same economic sectors, and in second column of panels are from the different economic sectors. The third column of panels are the self-responses and sign self-correlators to be compared with cross-responses and sign cross-correlators.

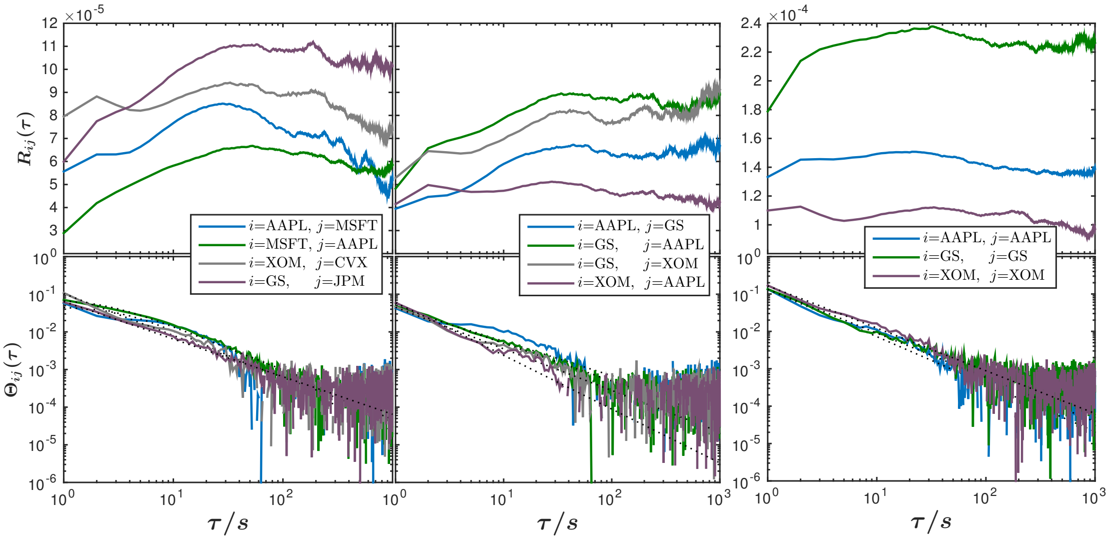


Figure 3.2: Cross-response functions $R_{ij}(\tau)$ excluding $\varepsilon_j(t) = 0$ in 2008 versus time lag τ on a logarithmic scale (top panels). Corresponding trade sign cross-correlators $\Theta_{ij}(\tau)$ for different stock pairs on a doubly logarithmic scale, fit as black dotted lines (bottom panels). The stock pairs in the first column of panels are from the same economic sectors, and in second column of panels are from the different economic sectors. The third column of panels are the self-responses and sign self-correlators to be compared with cross-responses and sign cross-correlators.

investors often lump together the stocks with little correlation and different economic risks in a portfolio. The gradual execution of trades for such a portfolio induces the price cross-response to the trades between different stocks. The trend of price reversion is also independent of in- or excluding zero trade signs $\varepsilon_j(t) = 0$. For the two cases, the general trends are quite similar. Only the overall amplitude is changed.

Here, we notice that when $\tau = 1$, the cross-response $R_{ij}(1)$ measures the impact of trades during one second. This is rather different from the instantaneous impact of one trade in single stocks. In our study, the cross-response is carried out on a physical time scale rather than a trading time scale, so that the time step is one second instead of one trade. Moreover, different from the self-response, the cross-response links different stocks, where the trades of one stock cannot immediately shift the price of another stock by consuming the volume of the later stock in the order book. The price impact between stocks is more likely to be induced by other mechanisms. For instance, the transmitted trading information may indirectly move the price of impacted stocks by affecting the placements and cancellations of limit orders, or even the executions of market orders. Due to the high frequency trading, the trading information can be detected very fast and the trades can be executed at the level of milliseconds. Therefore, the one-second interval is sufficient to identify the trading information or other effects from another stock.

3.3.2 Trade sign cross-correlators

In view of the self-response mainly induced by sign self-correlations [27], for the cross-response we introduce a trade sign cross-correlator

$$\Theta_{ij}(\tau) \;=\; \Big\langle \varepsilon_i(t+\tau)\varepsilon_j(t) \Big\rangle_t \tag{3.3}$$

in terms of the time lag τ. We once more distinguish the two possible definitions, including and excluding $\varepsilon_j(t) = 0$. It turns out that the differences are negligible. The sum of the product of trade signs between stocks is unchanged. Only the total number of trades enlarges or shrinks the average values, but they do not change the main feature of decaying. Therefore, the inclusion of $\varepsilon_j(t) = 0$ decreases the amplitude of sign cross-correlation, as shown in Figs. 3.1 and 3.2. This can eliminate the suspicion of ones about that the inclusion of zero trade signs induces the strong sign correlation.

In Figs. 3.1 and 3.2, the non-zero sign correlation across stocks turns out to be fitted well by a power law

$$\Theta_{ij}(\tau) = \frac{\vartheta_{ij}}{\Big(1+(\tau/\tau_{ij}^{(0)})^2\Big)^{\gamma_{ij}/2}} \ . \tag{3.4}$$

To estimate the fit error, we use the normalized χ_{ij}^2, seen in Appendix B. Thus, for eight stock pairs analyzed, the parameters of the best fit as well as the values χ_{ij}^2 of the errors are listed in Table 3.2. In contrast to the sign self-correlation with a long memory on the trading time scale [27, 98], the sign cross-correlations for most of the stocks exhibit a short memory with the exponents $\gamma_{ij} \leq 1$. Usually, the exponents smaller than unity [17] indicates a long memory for the decaying. The short-memory sign cross-correlation implies that the price change persists shortly and the price is reversed at a small time lag. To compare the difference between the cases in- and excluding $\varepsilon_j(t) = 0$, it is instructive to look at the parameters $\tau_{ij}^{(0)}$, which measure the decay time scale of sign cross-correlations. For most of the stock pairs, the case excluding $\varepsilon_j(t) = 0$ has a longer decay time scale

Table 3.2: Fit parameters and normalized χ_{ij}^2 for the trade sign cross-correlators.

sign	stock i	stock j	ϑ_{ij}		$\tau_{ij}^{(0)}$ [s]		γ_{ij}		χ_{ij}^2 ($\times 10^{-6}$)	
correlators			inc. 0	exc. 0	inc. 0	exc. 0	inc. 0	exc. 0	inc. 0	exc. 0
cross	AAPL	MSFT	0.46	0.05	0.05	3.46	1.00	1.35	0.23	1.52
	MSFT	AAPL	0.04	0.07	2.34	2.34	1.15	1.15	0.10	0.27
	XOM	CVX	0.61	0.67	0.06	0.21	1.04	1.16	0.07	0.52
	GS	JPM	0.45	0.48	0.07	0.13	1.00	1.00	0.04	0.18
	AAPL	GS	0.46	0.28	0.03	0.14	1.00	0.91	0.11	0.99
	GS	AAPL	0.49	0.49	0.06	0.10	1.00	1.00	0.05	0.13
	GS	XOM	0.61	0.73	0.04	0.08	1.04	1.10	0.04	0.20
	XOM	AAPL	0.76	0.29	0.05	0.34	1.09	1.42	0.12	0.18
self	AAPL	AAPL	0.60	0.96	0.21	0.21	1.27	1.27	0.18	0.50
	GS	GS	0.54	0.71	0.12	0.25	1.17	1.18	0.04	0.44
	XOM	XOM	0.54	0.89	0.17	0.23	1.12	1.14	0.09	0.49

than the one including $\varepsilon_j(t) = 0$. It illustrates that the absence of trading or the balance of buy and sell market orders accelerates the decay of sign cross-correlations.

In addition, we notice the large fluctuations of sign cross-correlators at larger time lags τ. They are partly due to the limited statistics. In particular, if the time lag τ is large, the time lags τ for different times t will overlap with each other. The large overlaps lead to the poor statistics when averaging the correlation over the time t.

3.3.3 Influences of zero trade signs

In our study, the physical time scale projects the transaction data to every time interval of one second. If there is an imbalance of buy and sell market orders in a time interval, the trade sign of this interval is either $+1$ or -1. If there is no trade or there is a balance of buy and sell market orders in a time interval, it is necessary or even inevitable to introduce the zero trade sign for this time interval.

For each trading day, on average, more than half of the total physical time features zero trade signs, listed in Table 2.4. For those non-trading time intervals, the price is not affected by trades so that the price response is zero at that time. After averaging the response over entire time, the effect of absence of trades is involved, diluting the impact of the trade itself. For instance, the impacts from the last trade one minute ago and from the last trade five hours ago, respectively, on the price at this moment are very different. For the latter, the current price may be influenced more significantly by the news during that five hours than by the last trade five hours ago. The time with or without trading is a non-trivial feature in the transaction data and contains much information on the trading activity. Especially, for whatever economic or other reasons, the time without trading reflects the disagreement of traders on the price. In view of this, the inclusion of zero trade signs in the cross-response function is reasonable.

Alternatively, we rule out all the zero trade signs from the cross-response function. This means the impact of trades is unaltered by the time without trading. The alternative choice implies the sign for one trade is fixed and independent of the time period τ_0 without trading. However, the empirical result reveals a possible reversion of the trade sign after τ_0, as shown in Fig. 3.3 for five successive trading days of AAPL in 2008. The $p_s(\tau_0)$ and $p_d(\tau_0)$ in Fig. 3.3 are the probabilities of finding the same and different trade signs, respectively, before and after the time period τ_0 without trading, where we have $p_s(\tau_0) + p_d(\tau_0) = 1$. When this period τ_0 enlarges, the probability of finding the same sign falls down slowly,

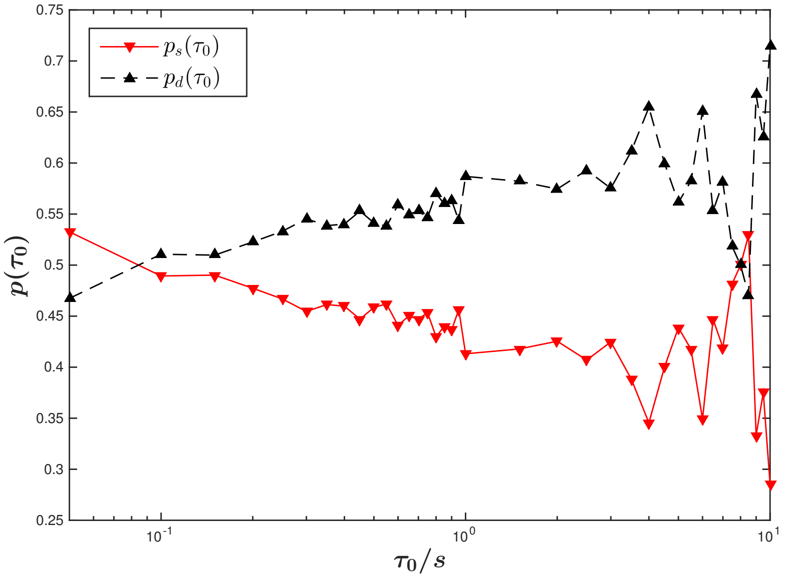


Figure 3.3: Probabilities $p_s(\tau_0)$ and $p_d(\tau_0)$ for the change of trade signs versus the time τ_0 without trading. The intraday data used stems from five successive trading days of AAPL in 2008. The strong fluctuations at large τ_0 are due to the limited statistics.

while the probability of finding the different signs raises up gradually. Thus, for a long time without trading, it is unlikely to keep the trade sign unchanged. Likewise, it is not possible to maintain the same impact of trades after a long time period without trading. From this perspective, the exclusion of zero trade signs may introduce a bias. Although we incline towards the case including zero trade signs, we display the empirical results of both cases for comparisons.

When there is a balance of buy and sell market orders in time interval t, the signs of buy and sell market orders cancel each other out. However, whether or not the behaviour of the trading itself causes the cross-response is unknown. To quantify the effect of zero trade signs, we use $R_{ij}^{(\text{inc. }0)}(\tau)$ and $R_{ij}^{(\text{exc. }0)}(\tau)$ to represent the cross-response in- and excluding $\varepsilon(t)=0$, respectively. Thus, the response $R_{ij}^{(\text{only }0)}(\tau)$ stemming from $\varepsilon(t)=0$ can be quantified by the difference of the two kinds of cross-responses,

$$R_{ij}^{(\text{only }0)}(\tau) = R_{ij}^{(\text{inc. }0)}(\tau) - R_{ij}^{(\text{exc. }0)}(\tau) \ . \tag{3.5}$$

Here, we do not distinguish the two sources of $\varepsilon(t)=0$, *i.e.*, a lack of trading and a balance of buy and sell market orders. Analogously, the sign cross-correlator for $\varepsilon(t)=0$ can be measured by

$$\Theta_{ij}^{(\text{only }0)}(\tau) = \Theta_{ij}^{(\text{inc. }0)}(\tau) - \Theta_{ij}^{(\text{exc. }0)}(\tau) \ . \tag{3.6}$$

As shown in Fig. 3.4, both the cross-response and the sign cross-correlator for $\varepsilon(t)=0$ exhibit negative values, which implies the inclusion of zero trade signs will weaken the impact of trades. In contrast, the exclusion of the zero trade signs enlarges this impact. In this sense, the results with the inclusion of $\varepsilon(t)=0$ gives a conservative estimation for the impact of trades and the price cross-response. In addition, the non-linear response with a reversed trend in the case of $\varepsilon(t)=0$ further corroborates our analysis on the influence of the time without trading.

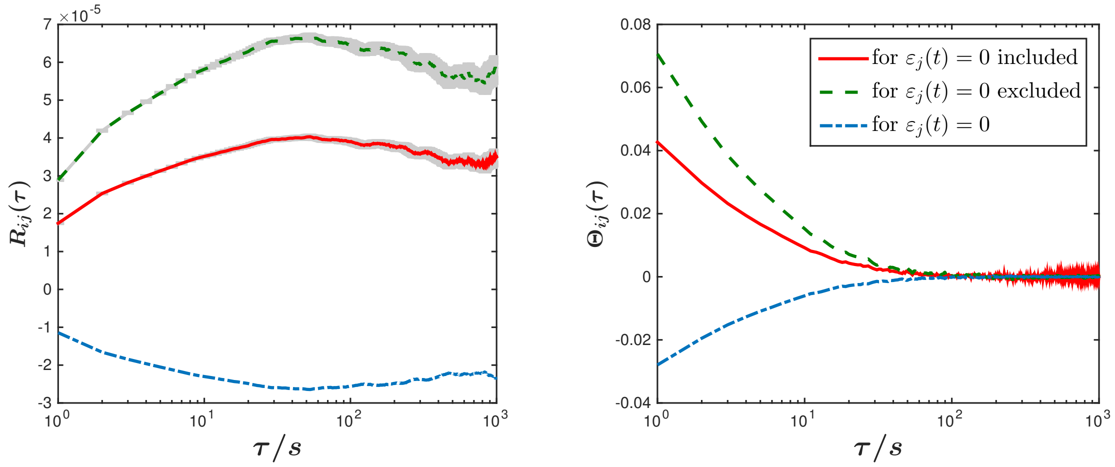


Figure 3.4: The three cross-responses (left) and sign cross-correlators (right) of stock pair (MSFT, AAPL) versus the physical time lag τ: in- and excluding $\varepsilon(t) = 0$ as well as only for $\varepsilon(t) = 0$. The shaded regions indicate the standard error bars.

3.3.4 Cross-response noise

In Figs. 3.1 and 3.2, an increase of the cross-response is shown up after the decreasing back at a large time lag τ close to 1000 seconds. Correspondingly, the strong fluctuations in the decaying of sign cross-correlators are visible. We attribute these behaviour to the noise. The identification of the noise helps us to find out the effective lagged period for the cross-response as well as the sign cross-correlators. To this end, we introduce an estimator $\xi_{ij}(\tau)$ for cross-response noise. The procedure of the estimation is as follows. We use a running integer number $T_{ij}^{(c)}$ to label the common trading days that the trades of stocks i and j took place. Then we separate our data into two sets, for days with odd and even numbers, respectively. With the averages over odd or even days, we work out the corresponding cross-responses $R_{ij}^{(k)}(\tau)$, where $k = 1,\ 2$. The two results should be very close to the cross-response $R_{ij}(\tau)$ averaged over all days. Otherwise, the results are influenced by the noise. Therefore, the estimator $\xi_{ij}(\tau)$ for the response noise is defined as some kind of normalized Euclidian distance in terms of the time lag τ,

$$\xi_{ij}(\tau) = \frac{1}{|R_{ij}(\tau)|}\sqrt{\frac{1}{2}\sum_{k=1}^{2}\Big(R_{ij}^{(k)}(\tau) - R_{ij}(\tau)\Big)^2}\ . \tag{3.7}$$

Fig. 3.5 displays the empirical results of the response noise during the year 2008, where all the cross-responses include the zero trade signs $\varepsilon_j(t) = 0$. Obviously, for time lag smaller than about 120 seconds, most stock pairs keep a low response noise with a value below 0.06. When increasing the time lag, the noise raises and the instability of cross-responses also grows. The strongest noise here could reach to a value of more than 0.25 for time lags tending towards 1000 seconds. This evidence visibly explains the upwards trends after reversing back for the cross-response. As to a deep interpretation for the response noise, it should be traced back to the cross-correlator of trade signs. When the sign cross-correlator becomes weak at the large time lag, other factors dominate to move the price, leading to a strong fluctuation of cross-responses. Certainly, the limited statistics is possible to blur the picture, as only 22200 seconds of effective trading time are

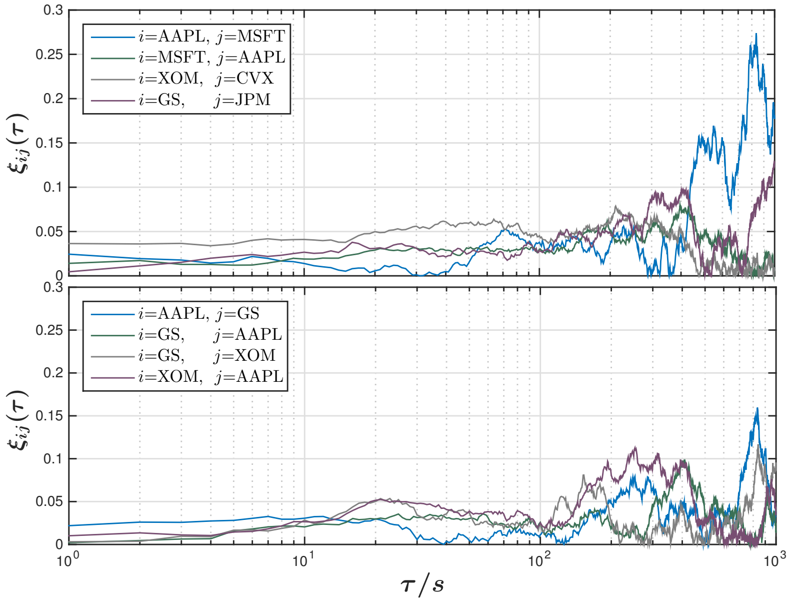


Figure 3.5: Cross-response noise $\xi_{ij}(\tau)$ for different stock pairs during the year 2008 versus the time lag τ measured on a logarithmic scale. Stocks of a pair are from the same economic sector (top); stocks of a pair are from different sectors (bottom).

used for each trading day. Overall, the analysis of noises clearly quantifies the effective and meaningful lagged period of cross-responses. Meanwhile, the cross-response at large time lags is little sense for stock pairs.

3.4 Market response

To explore the cross-response for the whole market, we begin with the market response structures in- and excluding zero trade signs in Sec. 3.4.1. We then discuss the possible causes for the transient impact of trades in Sec. 3.4.2. In terms of the transient impact, we explain the temporary deviation of market efficiency in Sec. 3.4.3. We further disclose the informed trades for the whole market in Sec. 3.4.4.

3.4.1 Market response structure

So far, the price cross-response function and the trade sign correlator have given us a kind of microscopic information for pairs of stocks. However, due to different constitutions of stocks, such information could be distributed among different pairs. Whether the information among different stock pairs cancels each other out or generates a particular information for the market is obscure. To make clear how the trading of individual stocks influences the market as a whole, we introduce the market response by a matrix $\rho(\tau)$ whose entries are normalized response functions at a given time lag τ,

$$\rho_{ij}(\tau) = \frac{R_{ij}(\tau)}{\max\left(|R_{ij}(\tau)|\right)} \,. \tag{3.8}$$

Here, the denominator is the maximum over all stock pairs (i, j) at each given time lag τ. The diagonal elements in the matrix are the self-responses, and the off-diagonal elements are the cross-responses. Due to different arguments, *i.e.*, the returns and the trade signs, in each entry, the response matrix is not symmetric, $\rho_{ij}(\tau) \neq \rho_{ji}(\tau)$ where $i \neq j$. Thus, the non-symmetric response matrix is distinguished from the symmetric matrix of Pearson correlation coefficients.

Fig. 3.6 shows the 99×99 matrices of the market responses across 99 stocks, listed in Appendix A.1, in the year 2008. The stocks in the matrices are ordered according to the economic sectors. For revealing the information of the time evolution, we consider the different time lags τ =1, 2, 60, 300, 1800, 7200 s for the market response. Visible in Fig. 3.6, for a fixed time lag, the price change of one stock is always affected by the trades of all others, and vice versa. With the evolution of the time lag, the intensity of the responses distributed in the market raises up first and falls down then. In line with the cross-responses for pairs of stocks, we also check the market responses in- and excluding zero trade signs $\varepsilon_j(t) = 0$. For both cases, regardless of the similar trend of time evolution, the microstructures of market response are quite different at each time lag. In the case including $\varepsilon_j(t) = 0$, the matrices feature striking patterns of strips grouped by sectors, such as the visibly strong strips in information technology (IT) sector. This market patten, reflecting the distribution of price impacts of trades, keeps stable over time. In contrast, the case excluding $\varepsilon_j(t) = 0$ displays a relatively homogeneous distribution of the impacts across different stocks. This patten is not changed too much with time increasing. In addition, for case either in- or excluding $\varepsilon_j(t) = 0$, a dependence of sectors is visible in the market response at each time lag. For example, the responses of utilities (U), financials (F) and energy (E) to their own sectors present the different strengths. Even in the same sector, the block responses for the cases in- and excluding $\varepsilon_j(t) = 0$ may be different, such as utilities (U) at $\tau = 60$ s.

Therefore, an abundance of information in the market microstructure can be identified by cross-responses. How to understand these information and what is the significance of these information for the whole market? To answer these questions, we further explore the nature of the market in the following sections.

3.4.2 Transient market impact

The price reversion with the time lag in Figs. 3.1 and 3.2 and the market response of the time evolution in Fig. 3.6 manifest a transient impact of trades on the price. However, the trade of one stock cannot directly influence the liquidity of another stock to move this stock's price. The transient impact, either for individual stocks or for the whole market, must be due to other mechanisms. We therefore discuss the possible causes of the transient impact, partly transferring a line of causes put forward for self-responses [155] to our case of cross-responses, although we cannot provide the sufficient evidence to support the following statements currently.

In single stocks, the splitting of large orders is one strategy for traders to hide their trading intentions and prevent the large impact cost from the low liquidity market. As a result, a sequence of small trades with the same type, buy or sell, induces the correlation of trade signs. The order splitting could exist in different stocks. When two such sequences of small trades separately from two individual stocks partly overlap each other on the physical time axis, a cross-correlation of trade signs will be generated, leading to a price response to trades between stocks, or say, a cross-impact of trades on the price. The partial overlapping of sequences may be due to the economic correlation reflecting on the return time series or purely be coincidental. In any case, since the self-impact of trades

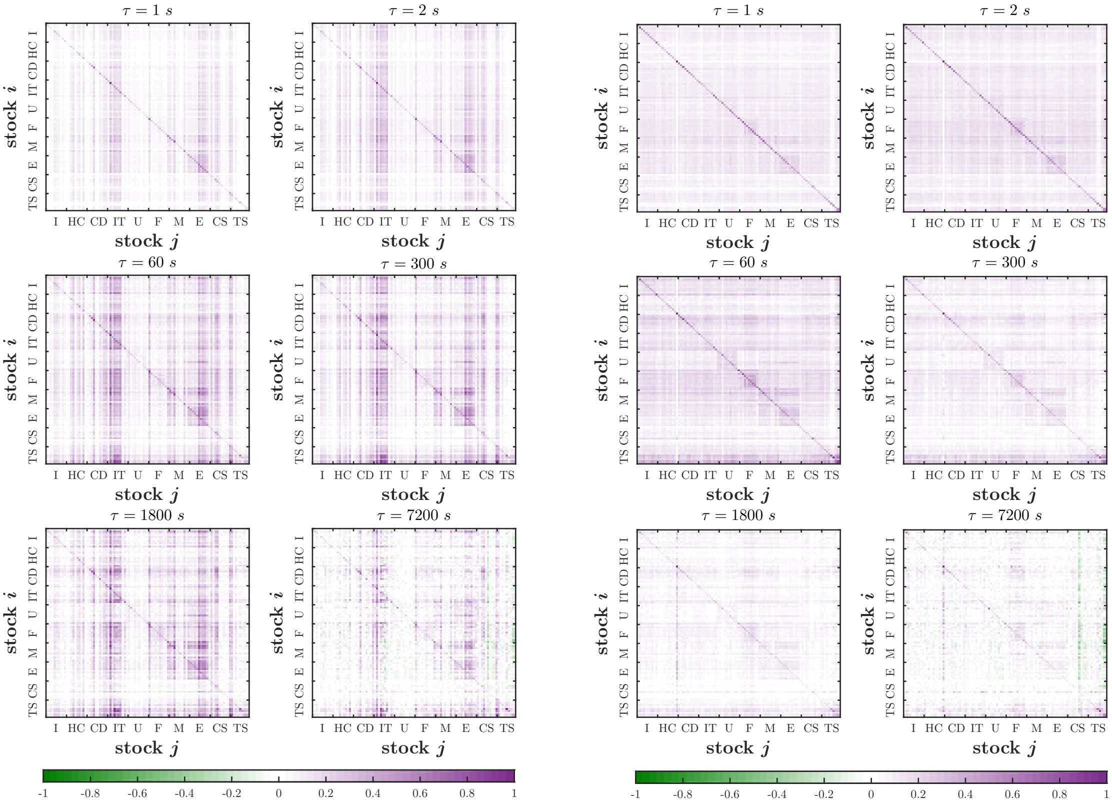


Figure 3.6: Matrices of market response with entries $\rho_{ij}(\tau)$ for $i, j = 1, \ldots, 99$ at different time lags $\tau = 1, 2, 60, 300, 1800, 7200$ s in the year 2008. The stocks pairs (i, j) belong to the sectors industrials (I), health care (HC), consumer discretionary (CD), information technology (IT), utilities (U), financials (F), materials (M), energy (E), consumer staples (CS), and telecommunications services (TS). The responses in the first two columns of panels include $\varepsilon_j(t) = 0$, while in the last two columns $\varepsilon_j(t) = 0$ is excluded.

in single stocks, related the sign self-correlation, is transient, the cross-impact of trades between stocks indirectly induced by the same way should be transient as well. Another possible cause we should not ignore is the behavior of traders. The overreaction to the trading information, *e.g.* herding behavior, prompts traders to extend their activities to other stocks which they did not trade previously, giving occasion to the price change of these stocks. The overreaction persists shortly. After the traders calming down and taking up again their previous trading patterns, less trades occur to these stock so that to impact slightly on their price. Here, regarding to different stocks, the traders act as the distributors of the trading information.

3.4.3 Temporary violation of market efficiency

The non-vanishing price response either in single stocks [27] or across different stocks shown in Fig. 3.6 reveals a non-Markovian feature in the price. The Efficient Market Hypothesis (EMH) [54], however, states that the price encodes all the available information, hinting the absence of arbitrage opportunities, or to be more specific, a zero response on average. The conflict seems to imply a violation of EMH. Actually, the transient impact in Fig. 3.6 demonstrates that the market efficiency is shortage on short time scales, but restored on longer time scales. The temporary violation of EMH is related to the behavior of some traders, which as discussed in Ref. [27] will alert other traders. No matter whether or not the trading is driven by valid information, the alerted traders act similar to arbitrageurs reverting the price until a state compatible with the EMH is reached. The process for the restoration of efficiency thus takes some time. To support this interpretation, we give the following analysis with a clear evidence shown in Fig. 3.7.

We treat the market incorporating all the information as a whole, which washes out the random fluctuations from the cross-responses by a self-averaging process,

$$\overline{R}(\tau) \;=\; \langle\langle R_{ij}(\tau)\rangle_j\rangle_i \;, \tag{3.9}$$

where $i = j$ is excluded. The doubly averaged response functions (3.9) with a more statistical significance are displayed in Fig. 3.7. Caused by a small part of potentially informed traders, an increasing response trend followed by a decreasing trend for the whole market shows up. In particular, the decreasing trend owes to the reverting action of the alerted traders. Compared to a stock pair, the decay of the cross-response for the market takes longer time, which partly arises from the noise reduction in the self-averaging process. Essentially, the whole market needs more time to digest all potential information than one individual stock. As the trend reversion lasts about three hours, the restoration of efficiency for the whole market, combined with the impact of trades vanishing, presents a rather slow process.

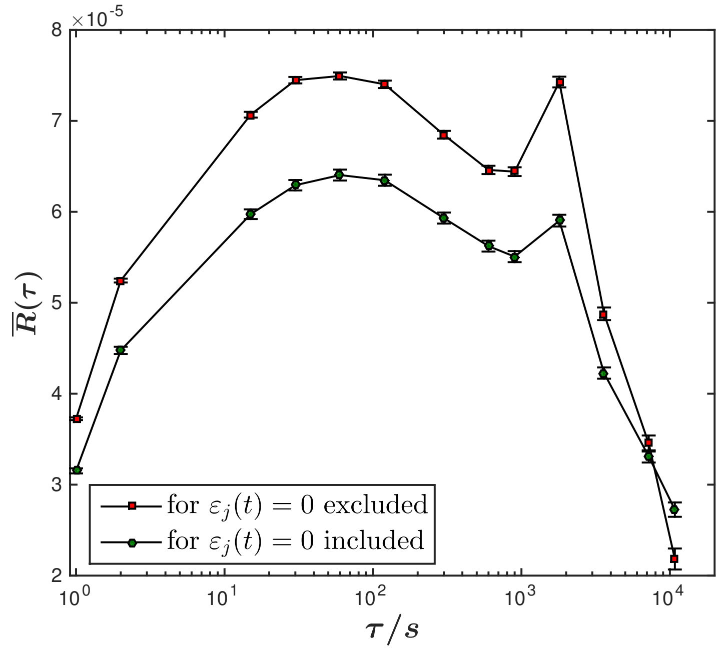


Figure 3.7: Doubly averaged response functions $\overline{R}(\tau)$ in- and excluding $\varepsilon_j(t) = 0$ for the whole market in 2008 versus time lag τ on a logarithmic scale. The error bars indicate the standard errors. For better comparison, the doubly average response function for $\varepsilon_j(t) = 0$ included and its error bars are is scaled up by a factor of six.

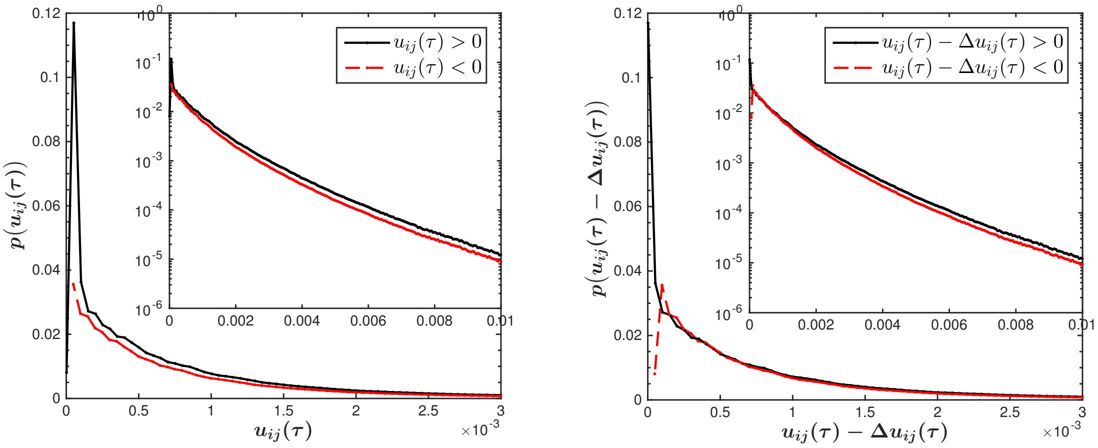


Figure 3.8: Probability distribution of the signed returns $u_{ij}(\tau) = r_i(t,\tau)\varepsilon_j(t)$ for the whole market in 2008, excluding $\varepsilon_j(t) = 0$, for $\tau = 30$ s: unshifted (left) and shifted (right) by $-\Delta u_{ij}(\tau)$ with $\Delta u_{ij}(30) = 5 \times 10^{-5}$. The distributions for negative arguments (dashed lines) are folded back to the positive regions. A logarithmic scale is used in the insets.

3.4.4 Informed trades

Generalizing the analysis of Ref. [27] for individual stocks, we examine the possible informed trades by carrying out the probability distribution of signed returns,

$$u_{ij}(\tau) = r_i(t,\tau)\varepsilon_j(t) \tag{3.10}$$

for the whole market in 2008. Here we rule out the case of $\varepsilon_j(t) = 0$. An example of the distribution at $\tau = 30$ s is shown in Fig. 3.8, where to highlight the small asymmetry the distribution of negative arguments are folded back to the positive regions. Carrying over the line of reasoning for individual stocks in Ref. [26] to our analysis, a horizontal shift of $\Delta u_{ij} = 5 \times 10^{-5}$ to symmetrize the distribution reveals a non-zero average impact from all trades, while the prevailed asymmetry between the tailed distributions hints that some trades are informed. For the former, the value of $\Delta u_{ij} = 5 \times 10^{-5}$ is comparable with the value of $\overline{R}(30) \sim 7 \times 10^{-5}$ for the doubly averaged response function. For the latter, it discloses and proves the existence of informed trades.

3.5 Comparisons of self- and cross-responses

Since the cross-response we have studied so far on a physical time scale is very different from the self-response in previous studies on a trading time scale, a comparison between both cases is therefore necessary to provide a view of the difference and to deepen the understanding of the price impact. To this end, it is worth to recall a crucial quantity, i.e., the trade sign $\varepsilon(t)$, which represents an aggregated sign of all trades in time interval t for our study, instead of the sign of t-th trade often used for previous studies. The distinctive definitions of the trade sign, due to the employment of different time scales, differ the measurements for the response as well as the trade sign correlation. Certainly, the measurement across two or more stocks also give rise to a distinguished impact mechanism from the one in single stocks. In the following, we will discuss these possible differences in detail.

- Different measurements for responses

  The trading time axis treats one trade as one time step. On this scale, the response with the lag of one time step measures the instantaneous impact of a single trade. When considering the lag of two time steps, the response comprises the price impact from the second last trade. Between the last and the second last trade, there is a gap of real physical time, which could be one second or five hours. For the larger time gap, the price may be influenced strongly by the prevailed news instead of the second last trade. Although this lagged response is an average result over all the second last trades, it loses sight of the time gap. So is the case for the lag of more than two time steps. Despite the exposure of the instantaneous impact for a single trade and a reasonable explanation for the additional impact from such time gap, the measurement on the trading time scale may introduce a bias for the impact from most lead trades. Differently, the physical time axis considers the equal time interval as one time step. With one-second interval on this scale in our case, the response lagged by one time step measures the impact of trades during one second, either in single stocks or across different stocks. For every lagged physical time, we always can find out a price response that comprises the impact from trades during the given time interval. Thus, the measurement on this scale is absent of the distortion from the uncertain time gap between two successive trades, although it cannot disclose the instantaneous impact of a single trade as on the trading time scale.

- Different impact ways and causes

  Regardless of the measurements on different time scales, the ways how trades impact on the price in single stocks are the same. By consuming the volume in the best quote, *i.e.*, influencing the market depth, the trades shift the price instantaneously to a worse level, leading to the price self-response. The price impact by this way originates from the market liquidity, referring to the market's ability to buy or sell an assent quickly without changing the price of the assent significantly. The lower the liquidity, the greater the price is impacted, and the more the extra trading cost should be paid. On the trading time scale, in other words, the price impact is on account of the liquidity cost that how much the traders should pay for a trade except for the conventional brokerage commission. This price impact could take place in absence of the informed trades, which has been demonstrated by Ref. [27]. On the physical time scale, however, as the one-second time interval facilitates the informational transmission and detection for high frequency trading, the price impact is partly due to the liquidity cost from uninformed trades and partly due to other causes from informed trades. The latter cause is obvious at a very short time lag but disappears at a long time lag. We show the corresponding evidence in Fig. 3.9.

Figure 3.9 shows the probability distribution of signed returns of AAPL in 2008, carried out on the physical time scale and at different time lags, excluding the zero trade signs $\varepsilon_j(t) = 0$. To highlight the small asymmetry, the distribution of negative arguments is once more folded back to the positive region. A stable horizontal shifts $\Delta u_{ij}(\tau) = 5.15 \times 10^{-5}$ for all the time lags are shown around the zero vertical axis, implying a price impact from uninformed trades. The asymmetry of tailed distributions, reflecting the imbalance of buy and sell orders, is accessible at small time lags $\tau = 1$, 15, 120 s but invisible at the large lag $\tau = 300$ s. We therefore infer that informed trades are triggered by the insufficient liquidity at a very short time, but are wiped out when the liquidity is restored by submitting more limit orders at a long time. Some sharp

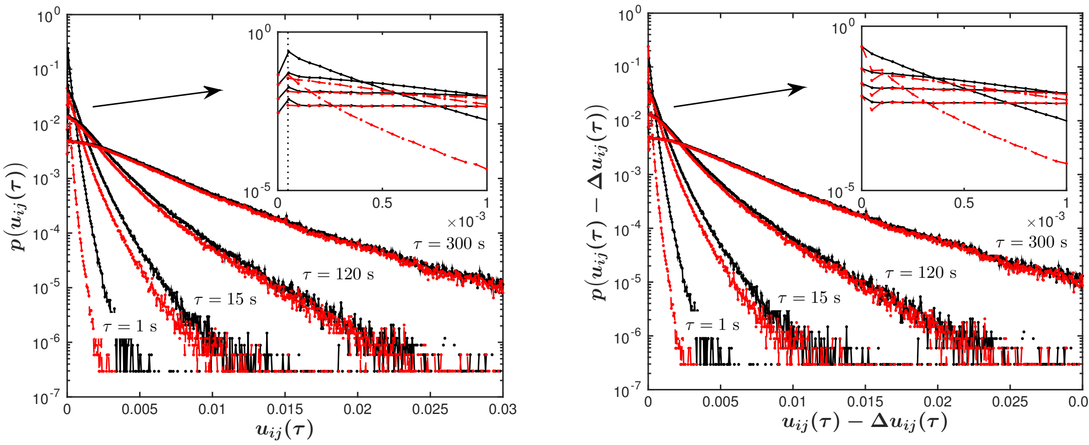


Figure 3.9: Probability distribution of signed return $u_{ij}(\tau)$ for physical time scale, where both i and j are AAPL in 2008: unshifted (left) and shifted by $-\Delta u_{ij}(\tau)$ (right). Here, $\Delta u_{ij}(\tau) = 5.15 \times 10^{-5}$ with $\tau = 1$, 15, 120, and 300 s, respectively. The negative parts of the distributions (dashed lines) are folded back to the positive regions. The insets are the amplified probability distributions near zero signed return.

bends for the distribution in the inset of Fig. 3.9 (right) is due to the folding procedure.

On the physical time scale, however, the way that the trades of one stock impact on the price of another stock is very different from the case in single stocks, as the trades of one stock cannot remove the volume in the best quote of another stock. This price cross-impact is more likely to occur through an indirect way, such as trading information spread, which influences the traders' actions. The price shift corresponding to the cross-impact of trades induces the price cross-response. Figs. 3.1 and 3.2 show that the cross-responses are mostly weaker than the self-responses on the same time scale. Put differently, the cross-impact between stocks is relative weak compared to the self-impact in single stocks. Instead of working on the volume directly as the self-impact, the cross-impact indirectly via a way of trading information may face to a loss of effective information during the spread, as the competing information, *e.g.*, other trades or relevant incoming news, may replace the trading information from the impacting stock. Comparing Fig. 3.8 with Fig. 3.9, we find the price cross-impact, same as the self-impact in the same scale, is induced both by uninformed trades and informed trades.

- Different properties and causes of trade sign correlations

Since a buy market order pushes the price up, whereas a sell market order drops the price down, the price shift accompanies with the trading type, indicated by the trade sign. When considering the price response to a trade, the trade signs in the price shift and the trade, respectively, produce a correlation. The origin of price response, therefore, should be traced back to the correlation of trade signs. If the trade shifts the price permanently or not is related to the performance of the trade sign correlation.

The quantitative analysis in Ref. [27] has shown a power-law decay of time lag for the trade sign correlation, coinciding with our analysis. Based on the trading time scale in Ref. [27], the trade sign correlation in single stocks exhibits a long memory with the exponent $0 < \gamma < 1$ in the power-law function. The long-memory property is more likely to be due to the order splitting, *i.e.*, the order with a large volume is fragmented into slices with small volumes. The traders use this way to hide their trading intentions and dwindle down the liquidity cost triggered immediately by the large order. Differently, on the physical time scale, either for sign self-correlation or for sign cross-correlation, we cannot seize up a long memory no matter whether the zero trade signs are included or not. Instead, a short memory with the exponent $1 < \gamma_{ij} < 2$ in the power-law function turns up, seen in Table 3.2. The differentiated memory property in our case results from the implementation of the physical time scale with the equal time intervals, which probably suppress the influence of multiple fragmented orders in one time interval but magnify the effect of a random trade in another time interval.

In spite of the memory property, on the physical time scale, the trade sign self-correlation as the one in trading time scale is very probable to stem from the order splitting. For the trade sign cross-correlation, it has several possible causes. When two large orders separately for two stocks are fragmented to trade, the coincident overlap on the physical time axis may introduce this sign cross-correlation. The actions of order splitting could be from different traders, but also be from the same trader for different stocks if this trader was executing a portfolio. As we already stated, the one-second time interval helps to spread the trading information. If a lot of high frequency traders detect such information, a herding behavior, referring to that some traders imitate the actions of other traders to buy or sell specific stocks regardless of their own trading pattens, could be induced to yield the sign cross-correlation.

So far, we have compared the difference between the self- and cross-responses on different time scales, and further give the corresponding analyses. Although some of the analyses have not been supported by empirical evidences, they at least provide clues to dig out the potential factor for future studies. To clarify the differences concisely, we provide a synopsis of the comparison in Table 3.3.

3.6 Conclusions

We extended the study of price self-responses in single stocks to price cross-responses in a whole correlated market. The price cross-response to trades between different stocks, as a function of the time lag, performs to be an increase followed by a decrease. This price reversion not only between two stocks but also across the whole market reveals a transient price impact from trades, which accounts for the temporary violation of market efficiency, *i.e.*, the market being inefficient at a short time but efficient at a long time. The temporary distortion of market efficiency roots from potentially informed traders who shift the price deviating from a foundational value. However, some traders, who might be interpreted as arbitrageurs, drive the price to reverse and thereby help to restore the efficiency. Pictorially speaking, the market needs time to react to the distortion of efficiency.

By presenting the response matrices, we provided a view of the market microstructure. The grouped strips pattens in the response matrices carried out by including zero trade signs are remarkable and stable in time. These features are also distinguished by sectors. The market response therefore provides much information about how the trade of one stock affects the prices of other stocks, and how this stock's own price is influenced by the trades of other stocks.

Table 3.3: Comparisons of self- and cross-responses

time scales	trading time scale	physical time scale	
responses	self-response	self-response	cross-response
trade sign $\varepsilon(t)$	sign of t-th trade	aggregated sign in time t	aggregated sign in time t
$R_{ii}(1)$ or $R_{ij}(1)$	instantaneous impact	one-second impact	one-second impact
impact causes	uninformed trades	uniformed trades and informed trades	uniformed trades and informed trades
impact ways	change prices by consuming the volumes in the order book directly	change prices both by directly consuming the volumes in the order book and indirectly affecting placements or cancellations of limit orders	change prices by indirectly affecting placements or cancellations of limit orders or even executions of market orders
trade sign correlations	self-correlation	self-correlation	cross-correlation
properties of sign correlations	long memory	short memory	short memory
causes of sign correlations	main: order splitting	main: order splitting	probable: order splitting, herding behavior, portfolios, et al.

Since the correlation of trade signs is closely related to the price response, we thus carried out the corresponding sign correlation for the cross-responses by a defined correlator. We find a power-law fashion for the cross-correlator, coinciding with the function for the sign self-correlation. However, different from the long-memory self-correlation on a trading time scale, the sign cross-correlation for stock pairs reveals a short-memory decaying process on a physical time scale. Here, the different time scales are partly responsible for the difference of memory properties. The rest responsibility should be owed to different correlation causes.

Regarding to the employment of a physical time scale rather than a trading time scale used in previous studies, and the price response across different stocks instead of in single stocks, we compared the differences between them. In contrast with the self-response on the trading time scale only due to uninformed trades, we find the self- and the cross-responses on the physical time scale are induced both by the uninformed and informed trades, as the one-second time interval on this time scale facilitate to spread the trading information, which could be detected and processed by high frequency traders. For the responses in single stocks and between different stocks, we find the ways that the price is impacted by trades are very different. The impact in single stocks works directly on the volumes to move the price, but between different stocks it shifts the price by an indirect way.

Chapter 4

Average cross-responses in correlated financial markets

4.1 Introduction

In the last chapter, we have found non-vanishing price responses across different stocks, revealing a non-Markovian feature. However, the cross-response for a pair of stocks strongly suffers from the noise and exhibits drastic fluctuations at large time lags. Correspondingly, due to the noise effect, the cross-correlation of trade signs turns out to be short memory instead of long memory. In contrast with the impact in single stocks, the impact of a given stock on other stocks may depend on various factors, for instance, the economic dependencies of companies, the grouping of investments in portfolios, and the capital turnover for buying in some stocks at the cost of selling out other stocks and so on. The complicated connection among stocks suggests that the price of one stock is affected not only by one stock but also by many related stocks. For a given stock, a better view of the price response is to average the cross-responses between different stocks and the given stock. By doing so, it to some extent smooths out the drastic fluctuations in both the price response and the sign cross-correlation at large time lags. With the reduction of noise, the more interesting observations can be expected. In this chapter, we therefore introduce the average cross-response of an individual stock to the whole market and to different economic sectors. In this setting, we also discuss the influences of individual stocks on other stocks.

The chapter is organized as follows. In Sec. 4.2, we describe the data used in this study. In Sec. 4.3, we carry out the average cross-responses of an individual stock, including the cross-response to the whole market and to different economic sectors. Meanwhile, we work out the average cross-correlations of trade signs. In Sec. 4.4, making use of the average cross-responses, we identify the influencing and influenced stocks, analyze the relations between influencing stocks and trading frequency, and discuss the role of trading frequency in response functions. In Sec. 4.5, by discussing the impact mechanisms in details, we compare the self- and average cross-response. We come to conclusions in Sec. 4.6. This chapter refers to Ref. [160].

4.2 Data description

Following the study in the last chapter, we again use the TAQ data set for NASDAQ stock market in 2008. The trading time is set to be from 9:40:00 to 15:50:00 EST for the days that the two stocks of a pair are traded. In Sec. 4.3, for the average cross-response of individual stocks, we choose Apple Inc. (AAPL), Goldman Sachs Group (GS) and Exxon Mobil Corp. (XOM) as the sample stocks. For one individual stock, the averages of cross-responses are carried out over other available 495 stocks in S&P 500 index or over other available stocks in a given economic sector, where the self-response of the stocks are always omitted. To analyze the influencing and influenced stocks according to the average cross-response in Sec. 4.4, we select the 99 stocks listed in Appendix A.1, which contains ten economic sectors and ten or nine stocks in each sector.

4.3 Average cross-responses of an individual stock

We introduce the passive and active cross-response functions as well as corresponding sign correlators in Sec. 4.3.1. We then carry out and discuss the two kinds of average responses to the whole market in Sec. 4.3.2 and to different economic sectors in Sec. 4.3.3, where we further analyze the property of trade sign cross-correlators.

4.3.1 Definitions

In correlated financial markets, the price of one stock might be influenced by multiple stocks. Meanwhile, the trades of this stock also might impact on the prices of multiple stocks. Since the definition (3.2) of the cross-response is not symmetric, we therefore can perform two conceptually different averages,

$$R_i^{(p)}(\tau) = \left\langle R_{ij}(\tau) \right\rangle_j \quad \text{and} \quad R_j^{(a)}(\tau) = \left\langle R_{ij}(\tau) \right\rangle_i \,, \tag{4.1}$$

which are named as passive and active cross-response functions, respectively. Importantly, we exclude the self-response for the stock pair $(i,\, i)$ or $(j,\, j)$ in these averages. The passive cross-response function $R_i^{(p)}(\tau)$ measures how the price of stock i changes due to the trades of all other stocks, while the active cross-response function $R_j^{(a)}(\tau)$ quantifies which effect the trades of stock j have on the average price of all other stocks. In particular, the former includes the collective impact from multiple stocks, and the latter contains the individual impact on the average price of multiple stocks. To reduce the fluctuation of trade sign correlations at large time lags, we also average the cross-correlator of trade signs in Eq. (3.3) over different stock pairs. Thus, corresponding to the passive and active cross-responses, we introduce

$$\Theta_i^{(p)}(\tau) = \left\langle \Theta_{ij}(\tau) \right\rangle_j \quad \text{and} \quad \Theta_j^{(a)}(\tau) = \left\langle \Theta_{ij}(\tau) \right\rangle_i \,, \tag{4.2}$$

as passive and active cross-correlators of trade signs. Because the time lag only enters the trade sign with index i, the average correlators here are not symmetric either.

Due to a lack of trades on the physical time scale, we introduced the zero trade signs $\varepsilon_j(t) = 0$ to empirical calculations in Chapter 3. The empirical results show a dependence on whether or not we include the zero trade signs in calculations. If including, the empirical results contain the effect of the time between two successive trades, which could be very

Table 4.1: Fit parameters and errors χ_i^2 or χ_j^2 for the average trade sign cross-correlators.

Sign cross-correlators	stock i, j	ϑ_i or ϑ_j		$\tau_i^{(0)}$ or $\tau_j^{(0)}$ [s]		γ_i or γ_j		χ_i^2 or χ_j^2 $(\times 10^{-6})$	
		inc. 0	exc. 0	inc. 0	exc. 0	inc. 0	exc. 0	inc. 0	exc. 0
$\Theta_i^{(p)}(\tau)$	AAPL	0.01	0.05	0.47	0.88	0.68	0.73	0.07	4.59
	GS	0.03	0.22	0.23	0.20	0.92	0.90	0.01	0.38
	XOM	0.27	0.83	0.06	0.12	1.32	1.33	0.02	1.20
$\Theta_j^{(a)}(\tau)$	AAPL	0.02	0.03	1.44	1.44	0.90	0.91	0.03	0.08
	GS	0.01	0.03	1.31	1.27	0.85	0.83	0.02	0.18
	XOM	0.02	0.03	0.55	1.08	0.71	0.95	0.11	0.08

short or very long. Otherwise, the results only reflect the impact of trades. In particular, the inclusion of zero trade signs diminishes the cross-response strength, but it does not change the trend of price reversion with the time lag. For a given time lag, the inclusion of zero trade signs changes the market response structure by reshaping the feature in the response matrix. In this chapter, as the averages for the responses and sign correlations are performed over different stocks, the uncertainty of the effect from the zero trade signs leads us to look at both cases.

4.3.2 Responses to the market

By averaging the cross-response and sign cross-correlator functions over other 495 stocks in S&P500 index, we carry out the empirical analysis for the sample stocks AAPL, GS and XOM. Their empirical results are presented in Fig. 4.1, where the two cases in- and excluding $\varepsilon_j(t) = 0$ are distinguished. Instead of at every second, we only calculate the results at several time lags. For each stock, we have checked that the results averaged over 495 stocks are similar to those averaged over other 98 stocks listed in Appendix A.1.

Figure 4.1 shows the passive and active cross-responses versus the time lag. For each stock, the passive one reverses faster than the active one. Compared with the persistent active response for hundreds of seconds, the passive response only persists for dozens of seconds before reversing with sizeable volatility. This difference is due to the fact that the price change in one stock is easer to be detected than the price change dispersed over different stocks. By extending the previous interpretations based on the study of single stocks, we can expound our view about the price dynamic on short time scales in terms of the passive cross-responses. When the price goes up, less market orders will be emitted to buy and more limit orders to sell. Thus, the price reverses [84] without a need to evoke new information as a cause. Moreover, such liquidity induced mean reversion attracts more buyers, which motivates liquidity providers to raise the price again. This process running over and over again causes the volatility during the decline of the passive cross-response. Considering the market as a whole, the mean reversion accentuates the short-run price volatility, which is consistent with the single-stock analysis [25, 77]. As the active cross-response disperses the impact of trades over the prices of different stocks, it is conceivable that this process takes longer time than the passive case. Accompanied with this dispersion, the volatility is spread out.

Shown in Fig. 4.1 , when excluding $\varepsilon_j(t) = 0$, the strength of the passive and active cross-responses are comparable. However, when including $\varepsilon_j(t) = 0$, the active ones are much stronger than the passive ones. The different response strengths are due to a part of stocks that are able to influence other stocks strongly. If looking again the matrices of the

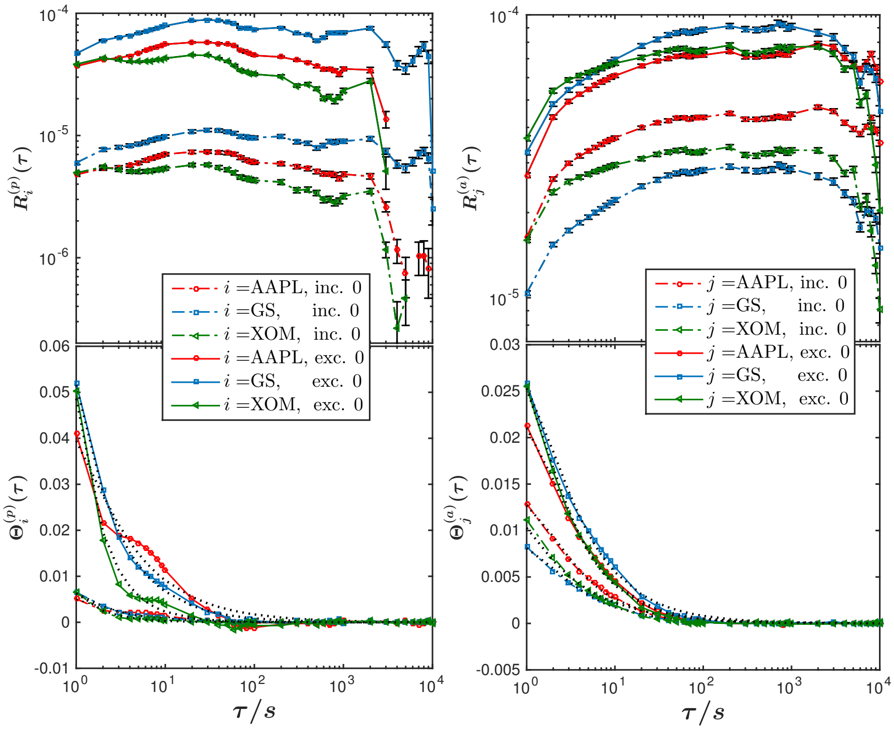


Figure 4.1: Passive and active cross-response functions $R_i^{(p)}(\tau)$ and $R_j^{(a)}(\tau)$ for $i, j = \mathrm{AAPL}, \mathrm{GS}, \mathrm{XOM}$ in the year 2008 versus time lag τ on a logarithmic scale (top, left and right). Corresponding passive and active trade sign cross-correlators $\Theta_i^{(p)}(\tau)$ and $\Theta_j^{(a)}(\tau)$, fit as dotted lines (bottom, left and right). The error bars indicate the standard errors.

market response in Fig. 3.6, we can find out these stocks signalled by the vertical stripes, such as the stocks in the IT sector. As the vertical stripes are much more pronounced than the corresponding horizontal ones, the active cross-response of an individual stock, resulting from averaging the cross-response function over the stocks along the vertical column, is much stronger the passive one by averaging over the stocks along the horizontal row. Briefly, the abilities for a specific stock to influence on or be influenced by other stocks are different.

To analyze the average cross-correlators of trade signs in Fig. 4.1, we fit the empirical results with the power-law function (3.4). The fitting parameters and errors are listed in Table 4.1. Remarkably, the exponents with $0 < \gamma_i < 1$ and $0 < \gamma_j < 1$ in the power law indicate the long-memory property for the average sign cross-correlators. Recalling the short-memory property in the sign cross-correlators of stock pairs, we thus infer that the price change related to the trade sign cross-correlations can accumulate to persist over longer time. By comparing Table 3.2 with Table 4.1, we find the two decay time scales $\tau_{ij}^{(0)}$ for the cases in- and excluding $\varepsilon_j(t) = 0$ are effectively enhanced and become comparable after averaging. The variations are attributed to the reduction of noise, which slows down the decaying process of sign correlations.

4.3.3 Responses to economic sectors

From the matrices of the market response in Fig. 3.6, we find that there are quite different responses between a given stock and the stocks from different economic sectors, regardless of the given stock impacted by other stocks or impacting on other stocks. If we treat each economic sector as a whole, how a given stock responds to each sector? For different sectors, do the yielded responses show the similar behaviour or not? To this end, we work out the average cross-responses to ten economic sectors in S&P 500 index for the sample stocks, *i.e.*, AAPL, GS and XOM.

The empirical results of the passive and active cross-responses are displayed in Figs. 4.2 and 4.3, respectively, where the calculations in- and excluding the zero trade signs $\varepsilon_j(t) = 0$ are distinguished. Once more, we observe the price reversion for all the cases. Besides, the sector dependence of the response is also remarkable. We will go into detail about this. For passive cross-responses, the three sample stocks have in common that they are all affected strongly by the trades within their own sectors, which is not surprising because of the same economic effects. For active cross-responses, a time-dependent clustering across different sectors is visible. In particular, the trades of AAPL and GS have a significant impact on the stocks from financials (F), but a lesser one on utilities (U), health care (HC) and consumer staples (CS). Since the latter three sectors serve the needs of daily life, they are more stable than other sectors and also much difficult to be impacted by the stocks unrelated to them. In addition, the trades of XOM are more likely to influence energy (E), the sector it belongs to, but less impact on health care (HC) and consumer staples (CS). Here, compared with health care (HC) and consumer staples (CS), utilities (U) are rather strongly coupled to energy (E) in an economic sense.

4.4 Influencing and influenced stocks from the viewpoint of average cross-responses

According to the active and passive cross-responses, we identify the influencing and influenced stocks in Sec. 4.4.1. In view of the large change in the ranks of influencing stocks in- and excluding zero trade signs, we thus discuss the relations between influencing stocks and the trading frequency in Sec. 4.4.2. Further, we analyze the role of trading frequency in the average cross-response functions in Sec. 4.4.3.

4.4.1 Identifying influencing and influenced stocks

Considering different measurements for the passive and active cross-responses, we thus classify the stocks as the influencing and influenced stocks. The influenced stocks with large passive cross-responses are easily impacted by trades of other stocks, whereas the influencing stocks with strong active cross-responses have profound impacts on prices of other stocks. To identify the two kinds of stocks, we rank the 99 stocks listed in Appendix A.1 according to the numerical values of responses, normalized by the Eq. (3.8) at a given time lag τ. The first fifteen stocks with the largest average cross-responses at $\tau = 1$, 2, 60, 300 s are shown in Figs. 4.4 and 4.5, where we show the ranks for the cases in- and excluding the zero trade signs $\varepsilon_j(t) = 0$, respectively.

Either including $\varepsilon_j(t) = 0$ in the passive cross-response or not, the rank of strongly influenced stocks keeps stable in general, especially for FTR, X, CF, NUE and S. Since the

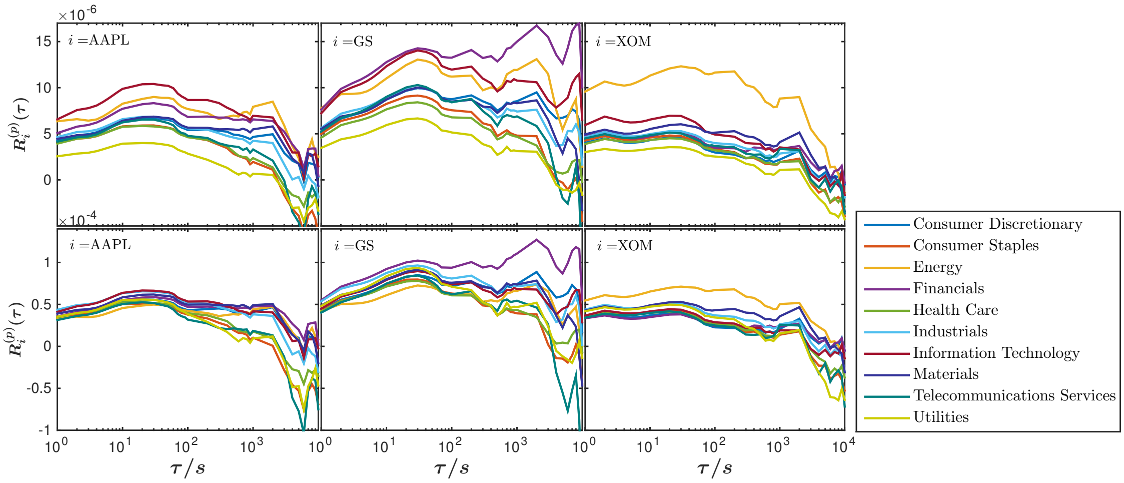


Figure 4.2: Passive cross-response functions $R_i^{(p)}(\tau)$ of the stocks i = AAPL, GS, XOM to ten different economic sectors in the year 2008 versus time lag τ on a logarithmic scale: in- and excluding $\varepsilon_j(t) = 0$ in the top and bottom panels, respectively.

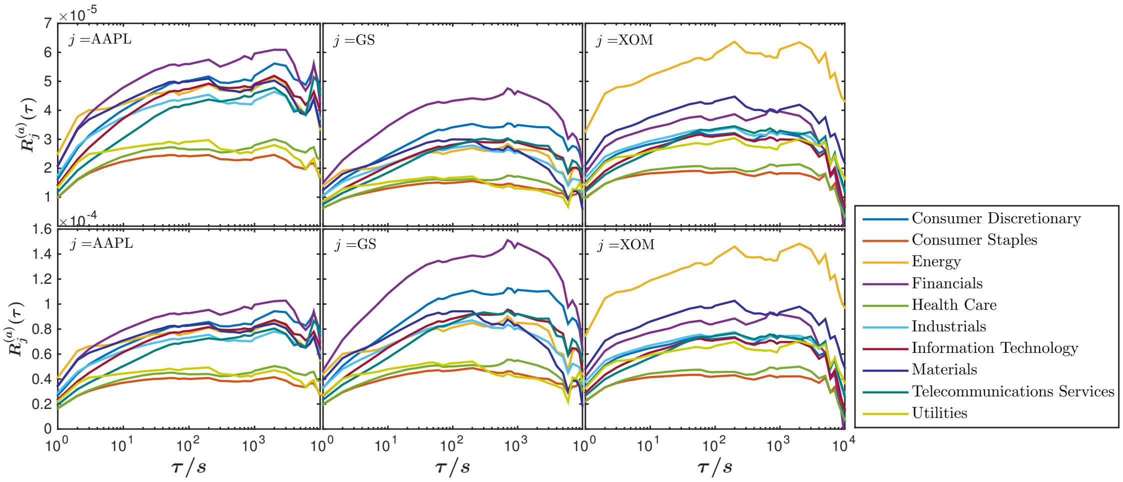


Figure 4.3: Active cross-response functions $R_j^{(a)}(\tau)$ of the stocks j = AAPL, GS, XOM to ten different economic sectors in the year 2008 versus time lag τ on a logarithmic scale: in- and excluding $\varepsilon_j(t) = 0$ in the top and bottom panels, respectively.

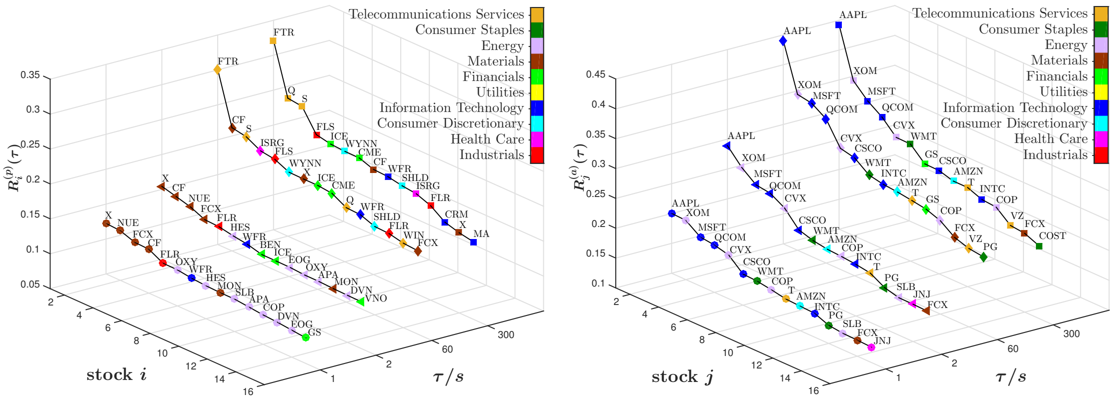


Figure 4.4: The first fifteen stocks with the strongest passive (top) and active (bottom) cross-response functions $R_i^{(p)}(\tau)$ and $R_j^{(a)}(\tau)$ versus stock index i or j and time lags $\tau = 1$ s ($\circ$), 2 s ($\triangleleft$), 60 s ($\diamond$), and 300 s ($\square$). The cross-response functions include $\varepsilon_j(t) = 0$. The ordinates of top and bottom graphs extend over different intervals.

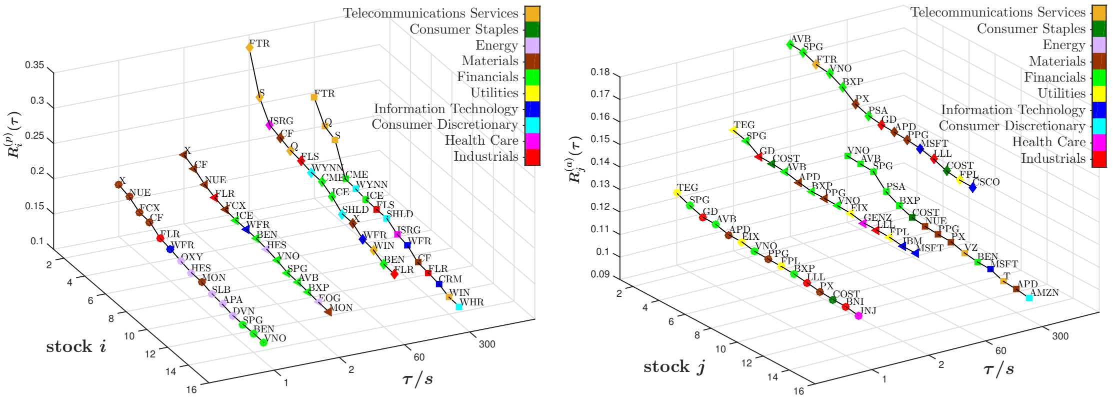


Figure 4.5: The first fifteen stocks with the strongest passive (top) and active (bottom) cross-response functions $R_i^{(p)}(\tau)$ and $R_j^{(a)}(\tau)$ versus stock index i or j and time lags $\tau = 1$ s ($\circ$), 2 s ($\triangleleft$), 60 s ($\diamond$), and 300 s ($\square$). The cross-response functions exclude $\varepsilon_j(t) = 0$. The ordinates of top and bottom graphs extend over different intervals.

effect of zero trade signs is diminished across different stocks, it alters the rank of influenced stocks very slightly. However, the effect of zero trade signs is spread to different stocks by the active cross-responses. Including $\varepsilon_j(t) = 0$ or not thus greatly changes the rank of strongly influencing stocks. When including $\varepsilon_j(t) = 0$, the rank reveals that most of influencing stocks come from information technology (IT) even if we shift the time lag. This conforms to the stable response structure in Fig. 3.6. When excluding $\varepsilon_j(t) = 0$, more stocks from financial (F) occupy the top positions. Even some stocks from utilities (U), which are absent from the top rank in the former case, are identified as the influencing stocks in this case.

4.4.2 Relations of influencing stocks and trading frequency

On a physical time scale, the zero trade sign indicates a lack of trades or a balance of buy and sell trades in a time interval. If many zero trade signs are included, it hints less time intervals having an imbalance of trades. Here, we introduce the trading frequency as the proportion of time intervals having an imbalance of trades to the total considered time intervals. Therefore, the effect of zero trade signs on the average cross-responses converts to the role of trading frequency in these responses, which further leads to the different ranks of influencing and influenced stocks. Due to the evident alteration in the rank of influencing stocks, we thus explore how the trading frequency affects the influencing stocks.

As a surrogate of the trading frequency, we consider the average daily number of trades. The daily number of trades is set to one per second if there is an imbalance of buy and sell trades in one-second intervals, or otherwise to zero per second. The intraday trading time is limited to be from 9:40:00 to 15:50:00 EST. Thus the total time intervals for each stock in each day are fixed. We use the 99 stocks listed in Appendix A.1 again for analysis. Fig. 4.6 shows the dependence of the active cross-response in- and excluding $\varepsilon_j(t) = 0$, respectively, on the average daily number of trades. We find a linear dependence for the case including $\varepsilon_j(t) = 0$, but a random relation for the case excluding $\varepsilon_j(t) = 0$. The linear dependence is accompanied by a high correlation between the response and the number of trades. It implies that the influencing stocks are exactly the stocks with the high trading frequency if we consider the effect of time by including zero trade signs. Pictorially speaking, the impact of trades from the frequently traded stocks is very probable to be rapidly followed by another impact. Before the previous impact vanishing, the new impact continues pushing the price, leading to a persistent change in the price. However, the impact of trades from infrequently traded stocks is lessened by the time period without any trade or with a balance of trades. This is so, because the price during this period may be affected by other information rather than trades, and may deviate from the previous price trend far away. If we ignore the effect of time by excluding zero trade signs, the linear dependence disappears. No matter how many trades occur in a trading day, the active cross-response at each time lag floats in a limited range. Even so, the strongly influencing stocks are more likely to be identified as those stocks with the daily number of trades lower than 2000.

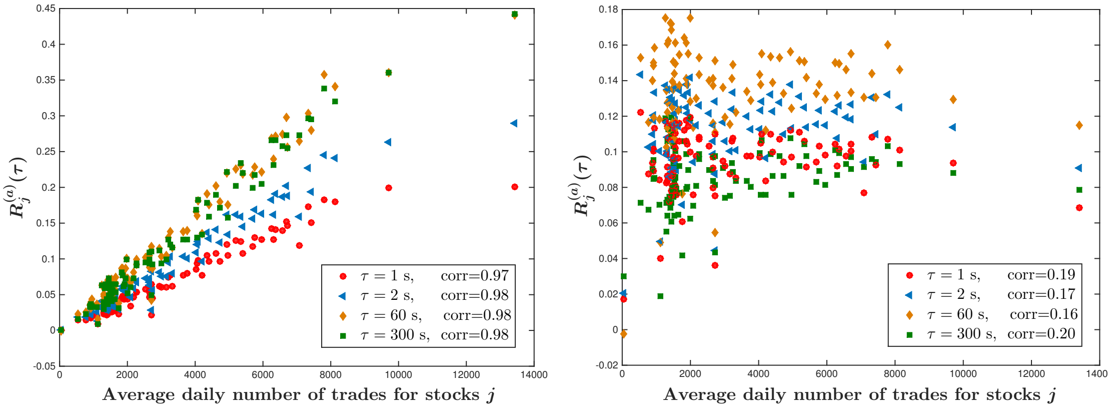


Figure 4.6: Relation between active cross-response and average daily number of trades for the stocks j. The Pearson correlation coefficient of these two quantities is denoted by corr. The number of daily trades is at most one per second from 9:40 to 15:50 New York local time. The cross-response functions in- and exclude $\varepsilon_j(t) = 0$ in the left and the right panels, respectively.

4.4.3 Role of trading frequency in response functions

To our knowledge, the asynchronous trading induces spurious lead-lag correlations, and the returns of frequently traded stocks generally lead to those of infrequently traded stocks [103, 104]. For instance, the incoming news influences the frequently traded stocks first, and then the infrequently traded stocks. In Sec. 4.4.2, the high correlation between the active cross-response including zero trade signs and the average daily number of trades might be misinterpreted, so that the influence that the stocks exert is owed to the lead-lag correlation of returns rather than the impact of trades. Here, we have to clarify that this lead-lag correlation is not fit to explain our case. What we focus on is how a lack or a balance of trades affects the impact degree of trades. Suppose a buy trade of stock j occurs at time t, which pushes the price of stock i up. Between time t and a later time $t + \tau$, the news comes in and triggers sell trades of stock i. The triggered sell trades shift the price to deviate from the previous price trend, debilitating the impact of that buy trade. To which degree is the impact of stock j on the stock i weakened due to the news between time t and $t + \tau$?

To answer the above question, we recall the definitions of the cross-response in- and excluding $\varepsilon_j(t) = 0$,

$$R_{ij}^{(\text{inc. }0)}(\tau) = \frac{\sum_{t=1}^{T_j+T_{j;n}} r_i(t,\tau)\varepsilon_j(t)}{T_j + T_{j;n}} \ , \tag{4.3}$$

$$R_{ij}^{(\text{exc. }0)}(\tau) = \frac{\sum_{t=1}^{T_j} r_i(t,\tau)\varepsilon_j(t)}{T_j} \ . \tag{4.4}$$

For stock j, T_j and $T_{j;n}$ are the total trading time of stock j and the total time without any trade or with a balance of buy and sell trades. If trading does not take place or there is a buy-sell balance, the products $r_i(t,\tau)\varepsilon_j(t)$ vanish. Thus, the numerators in Eqs. (4.3) and (4.4) are the same, but the denominators, i.e., the normalization constants, are different. Hence, we have

$$R_{ij}^{(\text{inc. }0)}(\tau) = f_j R_{ij}^{(\text{exc. }0)}(\tau) \ , \tag{4.5}$$

where the trading frequency defined in Sec. 4.4.2 can be written in a mathematical formula

$$f_j = \frac{T_j}{T_j + T_{j;n}} \ . \tag{4.6}$$

In parts of the literature, the term frequency is used in a colloquial sense. For example, a frequently traded stock is a very often traded stock. Our definition (4.6) is consistent with that, but quantifies it by f_j, $0 \leq f_j \leq 1$. The trading frequency is thus inversely proportional to time $T_{j;n}$. In addition, f_j also can be regarded as the probability for trades to occur unevenly on the time scale of one second. In our case, f_j rescales the degree of the impact: the higher the trading frequency f_j is, the less the impact of trades is weakened by time, and the more the remaining impact has on average after a time lag. Since the trading frequency is no more than one, the cross-response including $\varepsilon_j(t) = 0$ can never be stronger than the one excluding $\varepsilon_j(t) = 0$.

In view of rather different features of passive and active cross-responses, we further look at how the trading frequency affects them. For passive cross-response, we have

$$\begin{aligned} R_i^{(p,\ \text{inc. } 0)}(\tau) &= \frac{1}{k}\sum_{j=1}^{k} R_{ij}^{(\text{inc. } 0)}(\tau) \\ &= \frac{1}{k}\sum_{j=1}^{k} f_j R_{ij}^{(\text{exc. } 0)}(\tau) \ , \end{aligned} \tag{4.7}$$

where k is the total number of stocks j. By weighing the response excluding $\varepsilon_j(t) = 0$, the trading frequency links the two cases of passive cross-responses, *i.e.*, in- and excluding $\varepsilon_j(t) = 0$. The weighed term is summed over different stocks j to yield the passive cross-response including $\varepsilon_j(t) = 0$. This average greatly reduces the effect of trading frequency, so that the ranks of influenced stocks based on the passive cross-responses does not change too much. For active cross-response, we find

$$\begin{aligned} R_j^{(a,\ \text{inc. } 0)}(\tau) &= \frac{1}{k}\sum_{i=1}^{k} R_{ij}^{(\text{inc. } 0)}(\tau) \\ &= \frac{1}{k}\sum_{i=1}^{k} f_j R_{ij}^{(\text{exc. } 0)}(\tau) \\ &= f_j R_j^{(a,\ \text{exc. } 0)}(\tau) \ . \end{aligned} \tag{4.8}$$

Here, the trading frequency f_j leads the active cross-response including $\varepsilon_j(t) = 0$ to be proportional to the one excluding $\varepsilon_j(t) = 0$. This explains the strong linear correlation between the active cross-response including $\varepsilon_j(t) = 0$ and the average daily number of trades, no matter whether or not there is a correlation between the case excluding $\varepsilon_j(t) = 0$ and the number of trades.

4.5 Comparisons of self- and average cross-responses

In Sec. 3.5, we have compared and analyzed the differences of the self- and cross-responses, where the time scale affects the responses as well as the sign correlations significantly. In this section, since the averages of responses improve parts of the properties, such as a longer

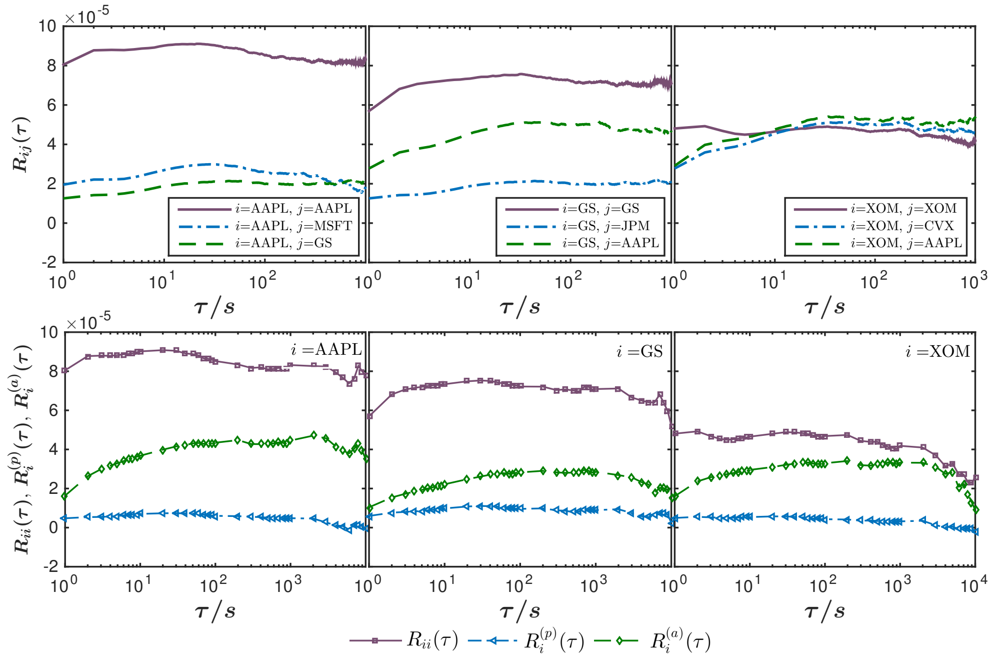


Figure 4.7: Comparisons of the self-responses for AAPL, GS and XOM with cross-responses to different stocks (top) versus the time lag τ. Comparisons of self-responses, passive cross-responses, and active cross-responses for the same stocks (bottom) versus the time lag τ. The zero trade signs are included. The scale on the horizontal axes is logarithmic.

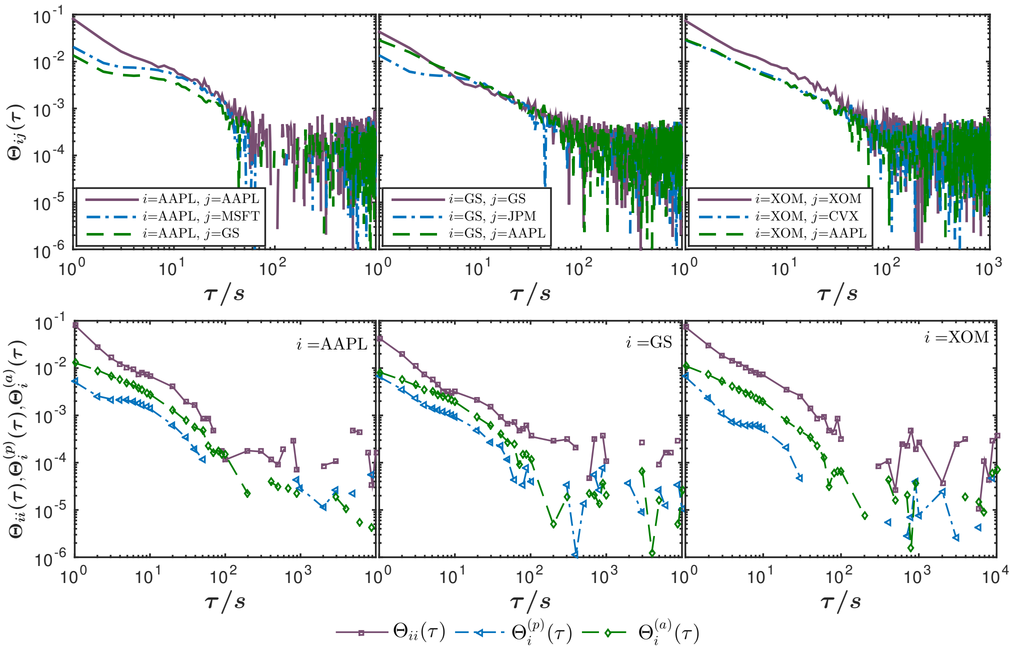


Figure 4.8: Comparisons of the trade sign self-correlators for AAPL, GS and XOM with cross-correlators with different stocks (top) versus the time lag τ. Comparisons of the self-correlators, passive cross-correlators, and active cross-correlators for the same stocks (bottom) versus the time lag τ. The zero trade signs are included. The scales on all axes are logarithmic. Negative values are suppressed.

persistence for price change, we will extend the comparisons to the responses within and without averaging on a physical time scale. Figs. 4.7 and 4.8 display the dependence of each response as well as the corresponding correlation of trade signs on the time lag. Here, AAPL, GS and XOM as the sample stocks are analyzed, and the influence of trading frequency is taken into account by including the zero trade signs.

To be more specific, Fig. 4.7 shows the comparison between the self- and cross-responses as well as the average cross-responses, *i.e.*, passive and active cross-responses. Typically, the self-response of one stock is stronger than the cross-response of this stock paired with other stocks. However, we notice an exception of XOM which is contrary to the general case. Comparing the self-responses with the average cross-responses, we find that the general case is aways hold, *i.e.*, the self-response of one stock is stronger than the active and passive cross-responses of this stock. Above exception of XOM implies an instability in the cross-responses for stock pairs, whereas the reduction of noise by averaging the cross-responses greatly improves the stability and persistency. We therefore have reasons to believe that it is convenient to look at the self- and cross-responses on a short time scale, but for investigating the stability, persistence or even the efficiency of the market as a whole, the average quantities give useful information on a longer time scale.

Corresponding to Fig. 4.7, Fig. 4.8 depicts the trade sign self- and cross-correlators as well as the average cross-correlators. The difference of the decaying correlators for single and paired stocks can be distinguished at the time lag below 10 seconds. With the increasing of the time lag, the different correlators are difficult to be distinguished, except for the cases of XOM. The study of the self-response [27] proposed a “bare” impact function with a power-law decay, which describes the impact mechanism and restrains the amplification effect due to the time-accumulated correlations of trade signs. Therefore, the price self-response is described by the impact function and the trade sign correlation together. If the impact mechanisms in self- and cross-responses are the same, the trade sign correlation should behave in a similar way to the response, as the two quantities are in proportion. However, the inconsistent behaviours of the responses and the sign correlations for XOM reveal different impact mechanisms for self- and cross-responses.

Furthermore, the passive cross-response contains the collective impact of different stocks on an individual stock. The collective impact is difficult to be observed, but it can be reflected indirectly by the behavior of trade sign correlators with a qualitative analysis. In the second row of Fig. 4.8, the sign correlators decay with the time lag in an orderly fashion that the self-correlators are always larger than the active and passive ones. This orderly fashion suggests a stronger stability of the sign correlators in the averages than in stock pairs. For the case of averages, the different sign correlators can be distinguished until 100 seconds. Comparing with the sign cross-correlators for stock pairs, we find a longer persistence for the average sign cross-correlators. Either the stronger stability or the longer persistence in the average sign correlators is consistent with the performance of the average cross-responses.

4.6 Conclusions

We introduced average cross-response functions, a passive and an active one, which measures how much the price of an individual stock changes due to the collective impact of trades from multiple stocks, and which effect the trade of one individual stock has on the average price of multiple stocks, respectively. The empirical results show very different features for the two kinds of responses. The passive one reverses quickly before declining in a volatile way, while the active one reverses at a longer time with less volatility. The difference arises from that the price change in one stock easily alerts the market participants, but the price change dispersed over different stocks is difficult to be detected. The average cross-responses allow for more statistically significant statements than the cross-response for individual stock pairs. In addition, we introduced the corresponding active and passive cross-correlators of trade signs. Remarkably, we found that the short memory of sign cross-correlation in individual stock pairs turns to be long memory after averaging. The long memory implies a long persistence either for the trade sign correlation or for the price change.

By comparing the cross-responses in- and excluding zero trade signs, we found that the time with a lack of trades or a balance of buy and sell trades lowers the impact of trades, leading to a weak cross-response. Here, the trading frequency plays a crucial role in rescaling the cross-response. According to the response degree, we therefore identified the different ranks for influenced and influencing stocks, especially for the latter. Furthermore, by comparing the self-responses with various cross-responses, we found that the self- and cross-responses result from different impact mechanisms. It is convenient to look at the cross-responses for stock pairs at a short time scale, but the averages performed over different stock pairs give more useful and interesting information at a longer time scale. It helps to investigate the stability, persistence or the efficiency of the market as a whole.

So far, we presented an empirical study and restricted ourselves to those interpretations that are possible. Certainly, model-based interpretations are called for, and we will present them in Chapter 5.

Chapter 5

Microscopic understanding of cross-responses: a price impact model

5.1 Introduction

In Chapters 3 and 4, we showed that the trades have an impact on the price either for pairs of individual stocks or for suitable averages over stock pairs. We gave the possible interpretations for this stylized fact, but to further understand the relation between prices and trades, those interpretations are obviously restricted. For this reason, the model-based interpretations are called for. In previous studies, many models are put forward to interpret the price impact from trades. For a single trade, an early model assumes a linear dependence of the impact on the traded volume [95], but the later empirical studies found that this dependence is nonlinear and can be described by a power-law [8, 60, 99], by a square-root [63, 130, 153] or by a logarithm [135]. In terms of time lags, the price impact for a single trade is found to be transient, *i.e.*, decaying with time, which is described by a propagator in a price impact model [27]. The price impact models mentioned above are all confined to single stocks. However, only very few models [30, 81, 124] are constructed to shed light on the price impact across different stocks, namely cross-impact. Upon completion of this study, a multivariate propagator model for cross-impacts is presented [16].

Due to a lack of available models to account for our empirical results of cross-responses, we therefore aim to extend the price impact model of Bouchaud *et al.* [27]. While this model contains one impact function for single stocks, we propose a price impact model for correlated markets which contains two impact functions depending on the time lag, one for the impact in single stocks and the other one for the impact across different stocks, *i.e.*, a self-impact function and a cross-impact function. The two impact functions result from the short-run liquidity of the stock itself and the trading information of other stocks, respectively. We demonstrate that the cross-response function of prices is indeed related to both, the self- and the cross-correlations of trade signs, as empirically analyzed in Chapters 3 and 4.

The chapter is organized as follows. In Sec. 5.2, we construct the price impact model with two impact functions depending on the time lag, *i.e.*, a self-impact function and a cross-impact function. We thereby obtain the average cross-response functions of individual stocks for three scenarios: the cross responses related to sign cross-correlations, to sign self-correlations, and to both, respectively. In Sec. 5.3, we resort to empirical data to analyze and calibrate the memory properties of the trade sign self- and cross-correlators, and to determine the relation between the average cross-responses and the traded volumes. In Sec. 5.4, due to the difficulty to quantify the price impacts using empirical data, we propose a construction to fix the parameters in the impact functions. In Sec. 5.5, we introduce a diffusion function that describes the correlated motion of different stock prices, and use it to corroborate the parameters. In Sec. 5.6, we quantify and interpret the price impacts of individual stocks in detail. We give our conclusions in Sec. 5.7. This chapter refers to Ref. [158]. In the following, we use the original text from Ref. [158].

5.2 Price impact model

We setup our model for the response functions in a market of correlated stocks. In Sec. 5.2.1, we construct the price impact model between two stocks to model the cross-response functions for stock pairs. In Sec. 5.2.2, we reduce the complexity of the response functions to be averaged by defining them per share and also by restricting ourselves to three scenarios. We obtain the passive and active averaged response functions for the three scenarios in Secs. 5.2.3, 5.2.4 and 5.2.5, respectively.

5.2.1 Setup of the model

Suppose a buy market order having a trade sign $\varepsilon_i(t) = +1$ with a unsigned volume $v_i(t)$ larger than that at the best ask was executed at initial time t. It is impossible to immediately issue new sell limit orders at the best ask in order to consume the volumes of the market order. In other words, there is an insufficient short-run liquidity. The buy market order therefore moves the initial trade price to a higher price instantaneously. Here, both the traded volume and the trade price are for the same stock i. The change of the trade price is reflected in the movement of the midpoint price, and the direction of the movement is indicated by the trade sign, *e.g.* in the present case $\varepsilon_i(t) = +1$ for the price raising. In this study we use the logarithmic midpoint price to replace the trade price. The price impact from the traded volume is always non-negative either for a buy or sell market order of the same stock. It is denoted by $f_i(v_i(t))$. However, the price impact cannot persist all the time, as the new incoming limit orders enlarge the liquidity. We say that a stock is liquid if there are many shares which can be sold or bought without time delay and with little impact on the stock price. The liquidity can be estimated by looking at the bid-ask spread, *i.e.* at the difference of the best bid and the best ask prices [15, 51]. Hence, if there is enough volume available at the new best ask with a price smaller than the one of last trade, the price in the following trades reverses. In view of the influence of the short-run liquidity, a price impact function $G_{ii}(\tau)$ versus time lag τ for a single trade is used to modulate the degree of the price impact due to the volumes traded. Furthermore, all other sources that indirectly cause the price change, such as the new information, are described by a random variable $\eta_{ii}(t)$. Hence, using discrete time, we have the trade price after the time step of length $\tau = 1$,

$$\log m_i(t+1) = \log m_i(t) + G_{ii}(1) f_i\big(v_i(t)\big)\varepsilon_i(t) + \eta_{ii}(t). \tag{5.1}$$

For the next time step, the trade price is not only influenced by the trade at the first time step, but also affected by the trade at the initial time t with the remnant impact modulated by $G_{ii}(2)$ with $\tau = 2$,

$$\begin{aligned}\log m_i(t+2) &= G_{ii}(1) f_i\big(v_i(t+1)\big)\varepsilon_i(t+1) + \eta_{ii}(t+1) \\ &+ G_{ii}(2) f_i\big(v_i(t)\big)\varepsilon_i(t) + \eta_{ii}(t) \\ &+ \log m_i(t) \ . \end{aligned} \tag{5.2}$$

Now suppose infinitely many trades were executed before time t, each of these trades has an impact on the trade price at time t. Accounting for the past price at time $-\infty$, we obtain the trade price at time t by constructing a superposition model, where all the price impacts from past trades are summed up,

$$\begin{aligned}\log m_i(t) &= \sum_{t'<t} G_{ii}(t-t') f_i\big(v_i(t')\big)\varepsilon_i(t') + \sum_{t'<t} \eta_{ii}(t') \\ &+ \log m_i(-\infty) \ . \end{aligned} \tag{5.3}$$

We notice the sum over the random variables $\eta_{ii}(t')$. The prototype of this model was proposed in Ref. [27]. It describes the price impact from past trades, focusing on the same stock only. Here, we go beyond this and by comprising the trades from the same stock as well as from the other stocks.

When considering a trade from another stock j with trade sign $\varepsilon_j(t')$, the trade also produces a price impact $g_i(v_j(t'))$ different from the one in the impacted stock i. As the trades of stock j do not consume volumes directly from the order book of stock i, we attribute the price impact $g_i(v_j(t'))$ to transmission of trading information. We emphasize that the trading information in our model only contains trade directions, *i.e.* buy and sell, and traded volumes of market orders, rather than other information, such as private information and relevant news which will later on be modelled by random variables. Due to the latter competing information, the price impact from volume traded for stock j cannot remain unchanged. Hence, to scale how the price impact depends on the time lag, we employ a price impact function $G_{ij}(\tau)$ for a single trade. To distinguish these two types of impact functions, we refer to $G_{ii}(\tau)$ as to the self-impact function of the stock i, while we refer to $G_{ij}(\tau)$ as to the cross-impact function between impacted stock i and impacting stock j. Moreover, we use random variables $\eta_{ij}(t')$ to model all the above mentioned sources belonging to stock j that may cause price change of stock i. All the random variables $\eta_{ii}(t')$ and $\eta_{ij}(t')$ are assumed to be independent of trade signs and to not show autocorrelations in time. Thus, we arrive at the following model

$$\begin{aligned}\log m_i(t) &= \sum_{t'<t} \Big[G_{ii}(t-t') f_i\big(v_i(t')\big)\varepsilon_i(t') + \eta_{ii}(t') \Big] \\ &+ \sum_{t'<t} \Big[G_{ij}(t-t') g_i\big(v_j(t')\big)\varepsilon_j(t') + \eta_{ij}(t') \Big] \\ &+ \log m_i(-\infty) \end{aligned} \tag{5.4}$$

for the impacts of trades from different stocks.

As a consequence of the trade superposition model (5.4), the price change of stock i resulting from Eq. (3.1) comprises two components. The first one is due to the short-run liquidity of stock i itself,

$$\begin{aligned}
r_{ii}^{(L)}(t,\tau) &= \sum_{t\leq t'<t+\tau} G_{ii}(t+\tau-t')f_i\big(v_i(t')\big)\varepsilon_i(t') \\
&+ \sum_{t'<t} \Big[G_{ii}(t+\tau-t') - G_{ii}(t-t')\Big] f_i\big(v_i(t')\big)\varepsilon_i(t') \\
&+ \sum_{t\leq t'<t+\tau} \eta_{ii}(t') \ .
\end{aligned} \tag{5.5}$$

As explained at the beginning of Sec. 5.2.2, the influence of the short-run liquidity is described by the self-impact $G_{ii}(\tau)$. Apart from the impact of traded volume, $G_{ii}(\tau)$ can be regarded as the impact of a single trade of stock i on its own price after the time τ. The second contribution results from the trading information transmitted from stock j to stock i,

$$\begin{aligned}
r_{ij}^{(I)}(t,\tau) &= \sum_{t\leq t'<t+\tau} G_{ij}(t+\tau-t')g_i\big(v_j(t')\big)\varepsilon_j(t') \\
&+ \sum_{t'<t} \Big[G_{ij}(t+\tau-t') - G_{ij}(t-t')\Big] g_i\big(v_j(t')\big)\varepsilon_j(t') \\
&+ \sum_{t\leq t'<t+\tau} \eta_{ij}(t') \ .
\end{aligned} \tag{5.6}$$

Here, the cross-impact function $G_{ij}(\tau)$ plays the rôle of information propagator between stocks i and j. It describes the impact of a single trade of stock j on the price of stock i after time τ, without taking the impact of traded volume of stock j into account. The sum of these two components

$$r_{ij}(t,\tau) = r_{ii}^{(L)}(t,\tau) + r_{ij}^{(I)}(t,\tau) \tag{5.7}$$

constitutes the total price change of stock i due to effects of stock i and of another stock j.

Employing the definition (3.2), we now calculate the time average response functions. For the two components of the price change of stock i we obtain

$$\begin{aligned}
R_{ij}^{(C)}(\tau) &= \Big\langle r_{ii}^{(L)}(t,\tau)\varepsilon_j(t)\Big\rangle_t \\
&= \sum_{t\leq t'<t+\tau} G_{ii}(t+\tau-t')\left\langle f_i\big(v_i(t')\big)\right\rangle_t \Theta_{ij}(t'-t) \\
&+ \sum_{t'<t} \Big[G_{ii}(t+\tau-t') - G_{ii}(t-t')\Big] \left\langle f_i\big(v_i(t')\big)\right\rangle_t \Theta_{ji}(t-t') \ ,
\end{aligned} \tag{5.8}$$

$$\begin{aligned}
R_{ij}^{(S)}(\tau) &= \Big\langle r_{ij}^{(I)}(t,\tau)\varepsilon_j(t)\Big\rangle_t \\
&= \sum_{t\leq t'<t+\tau} G_{ij}(t+\tau-t')\left\langle g_i\big(v_j(t')\big)\right\rangle_t \Theta_{jj}(t'-t) \\
&+ \sum_{t'<t} \Big[G_{ij}(t+\tau-t') - G_{ij}(t-t')\Big] \left\langle g_i\big(v_j(t')\big)\right\rangle_t \Theta_{jj}(t-t') \ .
\end{aligned} \tag{5.9}$$

The superscripts (C) and (S) refer to the response functions due to cross- and self-correlations of trade signs, respectively. Thus, the price response of stock i to the trades of stock j contains both contributions,

$$R_{ij}(\tau) = R_{ij}^{(C)}(\tau) + R_{ij}^{(S)}(\tau) \ . \tag{5.10}$$

Here we assume that both, the self- and the cross-correlator of trade signs, are independent of the impacts of traded volumes, *i.e.* of $f_i\big(v_i(t')\big)$ and $g_i\big(v_j(t')\big)$. The average impacts of traded volumes over all times t do not depend on the traded time t'. We notice that Eq. (5.8) only contains the trade sign cross-correlators $\Theta_{ij}(\tau)$ and $\Theta_{ji}(\tau)$, while Eq. (5.9) only involves the trade sign self-correlators $\Theta_{jj}(\tau)$.

5.2.2 Simplifications of the model

We want to perform two kinds of averages of the response functions with respect to the stock indices. To this end, we have a more detailed look at the functions (5.8) and (5.9) for a stock pair (i, j) and we notice that the contribution of the traded volumes to the price response is independent of the time lag. This time independence means that the time-dependent information of the response is only included in the impact functions and the trade sign correlators. Since we focus on how the price changes respond to the trades on a certain time scale, it is reasonable to reduce the complexity of each component by dividing the average impact of traded volumes, after averaging the response components for an individual stock i over different j, *i.e.*, over the second index,

$$R_{i,0}^{(p,C)}(\tau) \ = \ \frac{\Big\langle R_{ij}^{(C)}(\tau)\Big\rangle_j}{\Big\langle f_i(v_i)\Big\rangle_t} \qquad \text{and} \qquad R_{i,0}^{(p,S)}(\tau) \ = \ \frac{\Big\langle R_{ij}^{(S)}(\tau)\Big\rangle_j}{\Big\langle g_i(v_j)\Big\rangle_{t,j}} \ . \tag{5.11}$$

We emphasize that the total passive response per share $R_{i,0}^{(p)}(\tau)$ to be defined later on is not simply the sum of the above two functions $R_{i,0}^{(p,C)}(\tau)$ and $R_{i,0}^{(p,S)}(\tau)$ which measure the passive responses related to the cross- and self-correlators of trade signs, respectively. Likewise, for the active response per share we define the two contributions by averaging over different j, now being the first index,

$$R_{i,0}^{(a,C)}(\tau) \ = \ \frac{\Big\langle R_{ji}^{(C)}(\tau)\Big\rangle_j}{\Big\langle f_j(v_j)\Big\rangle_{t,j}} \qquad \text{and} \qquad R_{i,0}^{(a,S)}(\tau) \ = \ \frac{\Big\langle R_{ji}^{(S)}(\tau)\Big\rangle_j}{\Big\langle g_j(v_i)\Big\rangle_{t,j}} \ . \tag{5.12}$$

Again, the total active response $R_{i,0}^{(a)}(\tau)$ to be defined below is not simply the sum. The impact functions of traded volumes are distinguished by the individual stock with the subscript i of the passive response with the superscript (p) or the active response with (a). As the average impact of traded volumes is unrelated to the traded time, we omit the argument t' of the traded volumes. It is worth mentioning that those average impact functions of traded volumes as denominators in Eqs. (5.11) and (5.12) are quite different. For the passive response of stock i, $\langle f_i(v_i)\rangle_t$ and $\langle g_i(v_j)\rangle_{t,j}$ quantify the impacts of traded volumes, respectively, from stock i and stocks j on the price of stock i on average. For the active response of stock i, $\langle f_j(v_j)\rangle_{t,j}$ and $\langle g_j(v_i)\rangle_{t,j}$ measure how the traded volumes of stocks j and stock i, respectively, influence the average price of stocks j.

To further clarify the mechanisms in our model, we now consider it for the following three scenarios.

Scenario I *The cross-impact of trading information from other stocks is very weak, which allows us to set* $G_{ij}(\tau) \to 0$. *Therefore, the price cross-response only comes from the cross-correlators of trade signs* $\Theta_{ij}(\tau)$.

Scenario II *The cross-correlator of trade signs* $\Theta_{ij}(\tau)$ *is small enough to be ignored. Therefore, the cross-response only comes from the self-correlator of trade signs, while the cross-impact* $G_{ij}(\tau)$ *transmitting trading information between stocks is important.*

Scenario III *Both the self- and the cross-correlators of trade signs are responsible for the price cross-response with non-negligible self- and cross-impacts.*

The average response functions and average impact functions for the three scenarios are discussed in detail in Secs. 5.2.3, 5.2.4 and 5.2.5, respectively. For convenience and to avoid a cumbersome notation, we set the time t at which every trade is executed to zero, $t = 0$.

5.2.3 Scenario I: Cross-response related to trade sign cross-correlators

When the cross-impact function approaches zero, the response component $R_{ij}^{(S)}(\tau)$ related to the trade sign self-correlators vanishes, and only $R_{ij}^{(C)}(\tau)$ related to the cross-correlators remains. Hence, the price response across stocks is rooted in the sign cross-correlation. As for the price change, the self-impact function cannot be neglected, although it does not contribute to the correlations across stocks. From Eqs. (5.8), (5.11) and (5.12), the passive and active response functions per share follow as

$$R_{i,0}^{(p)}(\tau) = \sum_{0 \leq t < \tau} G_{ii}(\tau - t)\Theta_i^{(p)}(t) + \sum_{t<0} \Big[G_{ii}(\tau - t) - G_{ii}(-t)\Big]\Theta_i^{(a)}(-t) \ , \tag{5.13}$$

$$R_{i,0}^{(a)}(\tau) = \sum_{0 \leq t < \tau} \langle G_{jj}(\tau - t)\rangle_j \Theta_i^{(a)}(t) + \sum_{t<0} \Big[\langle G_{jj}(\tau - t)\rangle_j - \langle G_{jj}(-t)\rangle_j\Big]\Theta_i^{(p)}(-t) \ . \tag{5.14}$$

When performing averages over stock indices, the trade sign correlators and impact functions are always assumed to be independent of each other. Therefore, the passive and active trade sign cross-correlators, *i.e.* $\Theta_i^{(p)}(\tau)$ and $\Theta_i^{(a)}(\tau)$, appear in Eqs. (5.13) and (5.14).

To facilitate a comparison with the theoretical impact functions resulting from simulations to be discussed in Sec. 5.4, we transform the average response functions further. In a first step, by substituting τ' for the time intervals $\tau - t$ and $-t$ in impact functions of Eq. (5.13) and Eq. (5.14), the passive and active response functions become

$$R_{i,0}^{(p)}(\tau) = \sum_{\tau'=1}^{\infty} A_i^{(p)}(\tau, \tau') G_{ii}(\tau') \ , \tag{5.15}$$

$$R_{i,0}^{(a)}(\tau) = \sum_{\tau'=1}^{\infty} A_i^{(a)}(\tau, \tau') \langle G_{jj}(\tau')\rangle_j \ , \tag{5.16}$$

where we introduce

$$A_i^{(p)}(\tau,\tau') = \begin{cases} \Theta_i^{(p)}(\tau-\tau') - \Theta_i^{(a)}(\tau') & , \text{ if } 0 < \tau' \leq \tau \leq \infty \ , \\ \Theta_i^{(a)}(\tau'-\tau) - \Theta_i^{(a)}(\tau') & , \text{ if } 0 < \tau < \tau' \leq \infty \ , \end{cases} \tag{5.17}$$

$$A_i^{(a)}(\tau,\tau') = \begin{cases} \Theta_i^{(a)}(\tau-\tau') - \Theta_i^{(p)}(\tau') & , \text{ if } 0 < \tau' \leq \tau \leq \infty \ , \\ \Theta_i^{(p)}(\tau'-\tau) - \Theta_i^{(p)}(\tau') & , \text{ if } 0 < \tau < \tau' \leq \infty \ . \end{cases} \tag{5.18}$$

Equations (5.17) and (5.18) guarantee the positivity of the time lags in the passive and active trade sign correlators. The second step of the transformation is to employ a matrix notation. As we use discretized time, the quantities for different time lags τ or τ' can be treated as elements of average response vectors $R_{i,0}^{(p)}$ and $R_{i,0}^{(a)}$, impact vectors G_{ii} and $\langle G_{jj}\rangle_j$, and sign correlation matrices $A_i^{(p)}$ and $A_i^{(a)}$. We arrive at the rather concise expressions

$$R_{i,0}^{(p)} = A_i^{(p)} G_{ii} \qquad \text{and} \qquad R_{i,0}^{(a)} = A_i^{(a)} \langle G_{jj}\rangle_j \ , \tag{5.19}$$

which may be inverted,

$$G_{ii} = [A_i^{(p)}]^{-1} R_{i,0}^{(p)} \qquad \text{and} \qquad \langle G_{jj}\rangle_j = [A_i^{(a)}]^{-1} R_{i,0}^{(a)} \ . \tag{5.20}$$

These expressions render it possible to calculate the empirical impact functions from the empirically found responses per share and trade sign correlators. As the above vectors and matrices have infinite dimensions, we use a large cut-off T_{cut} of 3000 seconds for calculations.

5.2.4 Scenario II: Cross-response related to trade sign self-correlators

According to Eq. (5.9), the information propagator $G_{ij}(\tau)$ which transmits the trading information revealed by the self-correlators of trade signs across stocks is crucial in this scenario. Interestingly, the trade sign self-correlator not only relates to the self-response in single stocks [27, 98] but also to the cross-response between stocks. When considering different stocks, we group the impacts of trading information either coming from different stocks or transmitted to different stocks into an individual impact function, *i.e.* $G_i^{(p)}(\tau)$ for the former and $G_i^{(a)}(\tau)$ for the latter. Here, $G_i^{(p)}(\tau)$ is the price impact of stock i due to all single trades of different stocks, while $G_i^{(a)}(\tau)$ is the impact of a single trade of stock i on the average price of different stocks. We refer to $G_i^{(p)}(\tau)$ and $G_i^{(a)}(\tau)$ as to the passive and active impact function of stock i, respectively. Hence, from Eqs. (5.9), (5.11) and (5.12), we calculate

$$R_{i,0}^{(p)}(\tau) = \sum_{0\leq t<\tau} G_i^{(p)}(\tau-t)\,\langle\Theta_{jj}(t)\rangle_j + \sum_{t<0}\Big[G_i^{(p)}(\tau-t) - G_i^{(p)}(-t)\Big]\,\langle\Theta_{jj}(-t)\rangle_j \ , \tag{5.21}$$

$$R_{i,0}^{(a)}(\tau) = \sum_{0\leq t<\tau} G_i^{(a)}(\tau-t)\Theta_{ii}(t) + \sum_{t<0}\Big[G_i^{(a)}(\tau-t) - G_i^{(a)}(-t)\Big]\Theta_{ii}(-t) \ , \tag{5.22}$$

as the passive and active response functions per share.

To obtain the empirical impact functions, transformations similar to the ones from Eq. (5.15) to Eq. (5.20) are carried out. This yields

$$G_i^{(p)} = [\langle A_{jj} \rangle_j]^{-1} R_{i,0}^{(p)} \qquad \text{and} \qquad G_i^{(a)} = [A_{ii}]^{-1} R_{i,0}^{(a)} \ , \tag{5.23}$$

with the matrix elements

$$A_{ii}(\tau, \tau') = \begin{cases} \Theta_{ii}(\tau - \tau') - \Theta_{ii}(\tau') & , \text{ if } \ 0 < \tau' \leq \tau \leq \infty \ , \\ \Theta_{ii}(\tau' - \tau) - \Theta_{ii}(\tau') & , \text{ if } \ 0 < \tau < \tau' \leq \infty \ . \end{cases} \tag{5.24}$$

The matrix elements $\langle A_{jj} \rangle_j$ are defined analogously. Again, the infinity ∞ in Eq. (5.24) will be cut off by a large time T_{cut} of 3000 seconds for calculations. Therefore, Eqs. (5.20) and (5.23) reveal the empirical price impacts depending on the time lag τ.

5.2.5 Scenario III: Cross-response related to both correlators

We now take into account both response components that were individually studied in Scenarios I and II. Neither the self- nor the cross-impacts as propagators of single trades can be neglected, both contribute to the price change. We underline once more that both, the cross-impact and the trade sign cross-correlator, generate responses across different stocks. When taking different stocks into account, the impact functions either becomes the average self-impact functions or enters active and passive response functions. Compared to Scenarios I and II, the average response function here describes the price response to trades between stocks completely, regardless of the complexity and hence depending on numerous parameters. By making use of the average response functions per share obtained in Scenarios I and II, we find the average response functions

$$R_i^{(p)}(\tau) = R_{i,0}^{(p,C)}(\tau) \big\langle f_i(v_i) \big\rangle_t + R_{i,0}^{(p,S)}(\tau) \big\langle g_i(v_j) \big\rangle_{t,j} \ , \tag{5.25}$$

$$R_i^{(a)}(\tau) = R_{i,0}^{(a,C)}(\tau) \big\langle f_j(v_j) \big\rangle_{t,j} + R_{i,0}^{(a,S)}(\tau) \big\langle g_j(v_i) \big\rangle_{t,j} \ . \tag{5.26}$$

The passive response functions, $R_{i,0}^{(p,C)}(\tau)$ and $R_{i,0}^{(p,S)}(\tau)$ have the same forms as in Eqs. (5.13) and (5.21), respectively. Similarly, the active response functions $R_{i,0}^{(a,C)}(\tau)$ and $R_{i,0}^{(a,S)}(\tau)$ have the same forms as in Eqs. (5.14) and (5.22), respectively.

5.3 Empirical analysis

We analyze the memory properties of the trade sign correlators and the relation between the responses and the traded volumes. In Sec. 5.3.1, we discuss the data set and introduce some definitions, *e.g.* trade sign and time scale. In Sec. 5.3.2, we check the memory properties of the self- and cross-correlators of trade signs for 31 stocks. In Sec. 5.3.3, we analyze the impacts of traded volumes for passive and active cross-responses of individual stocks.

5.3.1 Data sets and definitions

The empirical analysis employs the Trades and Quotes (TAQ) data set. Among all the markets included in the TAQ data set, we only use the data from NASDAQ stock market in 2008, since NASDAQ is a purely electronic market. To investigate the average response across different stocks, we choose the first 31 stocks from S&P500 index (see Appendix A.3) with the largest average number of daily trades. As in Refs. [160, 161], we use the physical time instead of the trading time which is convenient when considering self-responses in single stocks [27, 98]. However, when looking at different stocks which each have their own trading time, we found that the physical time is the better choice. As our data has a one-second resolution, it is only meaningful to define the number of daily trades as none or one per second from 9:40 to 15:50 New York local time on the physical time scale, even though more than one trade or quote can occur in this second. To determine the sign of every trade in the one-second interval, we cannot employ the approach of comparing the trades price with the preceding midpoint price in the best quote [97], since trades and quotes data are listed in two individual files without sufficiently short time stamps to specify the preceding midpoint price of the trade. Instead, we employ our approach [161]. If there are $N(t)$ trades in the time interval labeled by t, then the trades are numbered $n = 1, ..., N(t)$ and the corresponding prices are denoted $S(t; n)$. For two consecutive trades in the interval t, the sign of the price change is defined as

$$\varepsilon(t;n) = \begin{cases} \mathrm{sgn}\big(S(t;n) - S(t;n-1)\big) & , \quad \text{if} \quad S(t;n) \neq S(t;n-1) \ , \\ \varepsilon(t;n-1) & , \quad \text{otherwise} \ . \end{cases} \tag{5.27}$$

According to Eq. (5.27), a buy market order with the trade sign $\varepsilon(t; n) = +1$ is executed if the trade price raises, while a sell market order with $\varepsilon(t; n) = -1$ is executed if the trade price falls. If the trade price is unchanged, the trade sign is set to be the same as the preceding one, because the two consecutive trades with the same trading direction did not exhaust the available volume at the best price. If there are more than one trade in the interval t, these trades are aggregated yielding a single trade sign for t,

$$\varepsilon(t) = \begin{cases} \mathrm{sgn}\left(\sum\limits_{n=1}^{N(t)} \varepsilon(t;n)\right) & , \quad \text{if} \quad N(t) > 0 \ , \\ 0 & , \quad \text{if} \quad N(t) = 0 \ . \end{cases} \tag{5.28}$$

The case $N(t) = 1$ is included. If the majority of trades in second t was triggered by buy (or sell) market orders, then $\varepsilon(t) = +1$ (or -1). If trading did not take place or if there was a balance of buy and sell market orders in the second t, the trade sign is set to $\varepsilon(t) = 0$.

We only consider those days for a stock pair (i, j) in which trading took place in both stocks. In each such day, the trading time is limited from 9:40 to 15:50 of New York local time, which avoids overnight effects and any artifacts due to opening and closing of the market.

5.3.2 Properties of trade sign correlators

For both, the self- and the cross-correlators of trade signs appearing in the response functions in Sec. 5.2, an empirical check of their memories is called for: For the long-memory sign correlation, a buy (sell) market order is more likely to be followed by other buy (sell) market orders. The price thus changes persistently. For the short-memory sign correlation, a buy (sell) market order is not as often followed by other buy (sell) market orders. Thus, the price is more likely to quickly reverse. Previous studies have found long memory in individual stocks, making the trade sign self-correlator slowly decay in a slow power-law fashion [26, 27, 98]. One way to characterize the long-memory process [17] is to use the covariance function $Y(\tau)$. In the present case we may identify this object with the trade-sign correlator. In general, the process under consideration has long memory, if in the limit $\tau \to \infty$, the covariance function has the form

$$Y(\tau) \sim \tau^{-\gamma} L(\tau) \ , \tag{5.29}$$

where $0 < \gamma < 1$. The function $L(\tau)$ has to be slowly varying at $\tau \to \infty$ [53], implying

$$\lim_{\tau \to \infty} \frac{L(\alpha\tau)}{L(\tau)} = 1 \ , \quad \text{for all } \alpha \ . \tag{5.30}$$

This asymptotic characterization ignores the correlation at any smaller time lag. The exponent γ determines the rate of decay of the correlation rather than their absolute size and thus also whether the integrated correlation function remains finite. Even a small correlation can generate a long-memory process, characterized by the exponent γ. The smaller γ, the longer the memory. In financial markets, the exponent γ is often measured via a power-law function of the trade sign correlator,

$$\Theta_{ij}(\tau) \simeq \frac{\vartheta_{ij}}{\tau^{\gamma}} \quad \text{for large}\, \tau \ . \tag{5.31}$$

The constant ϑ_{ij} as well as the exponent γ are fit parameters [160, 161]. In the case $i = j$, the above function is the trade sign self-correlator, while for $i \neq j$ it is the cross-correlator. A more refined functional dependence is not needed, as we are only interested in the long-memory properties.

For the sign self-correlators on the trading time scale, Lillo and Farmer (LF) [98] found $\gamma = 0.6$ by analyzing 20 highly capitalized stocks traded in the London Stock Exchange. Bouchaud *et al.* [26] measured a value of γ ranging from 0.2 to 0.7 for the Paris Stock Exchange, *e.g.* $\gamma \approx 0.2$ for France-Telecom, and $\gamma \approx 0.67$ for Total. Here, we will work out the sign self-correlations on the physical time scale of 31 individual stocks. The results are listed in Appendix A.3. In these stocks, 71% show long-memory with $\gamma < 1$, and the rest, 29%, show short-memory with $\gamma \geq 1$. It is not surprising to find short-memory self-correlators in individual stocks. Similar results were obtained in Ref. [52]. The short-memory is due to a balance of long-memory positive and negative correlations. This might also be an explanation for our findings for the self-correlations. However, the positive correlation dominates in the first 10000 seconds. Thus, we tend more to the our explanation put forward in Ref. [161]. In the one-second time intervals, several trades with the same trading direction may occur, they are aggregated to yield one trade sign. As only the net effect of these trades matters, not their individual effects, they have the same overall effect as if just one trade occurred in this one-second interval. We also work out the average self-correlation of trade signs $\langle \Theta_{ii}(\tau) \rangle_i$. The 31 stocks considered exhibit a long-memory with $\gamma = 0.87$, unaffected by the short-memory of a small part of stocks.

In previous analyses [160, 161], we provided considerable evidence that the short-memory of sign cross-correlation is converted into long-memory when averaging across the market. We are thus led to check the memory property for the average cross-correlators of trade signs for each individual stocks across other 30 stocks. We find that the passive cross-correlator $\Theta_i^{(p)}$ of 77.4% of the stocks show long-memory with the γ ranging from 0.62 to 1, and the active cross-correlator $\Theta_i^{(a)}$ of all stocks exhibits long-memory with γ ranging from 0.75 to 1, see Appendix A.3.

5.3.3 Impacts of traded volumes

According to Eqs (5.8) and (5.9), the traded volumes contribute to the price response. We assume that the impact of traded volumes is independent of the time lag τ. In previous studies, Lillo, Farmer and Mantegna [99] have shown that the impact can be described by a concave function on the trading time scale. More specifically, it is a power law, $R(v) \sim v^{\delta}$. Similar studies have been put forward [92, 129] for time-aggregated volumes. In the study of Potters *et al.* [135], who analyzed stocks traded at the Pairs Bourse and NASDAQ, also on trading time scale, a logarithmic impact, $R(v) \sim \log(v)$, was found. To the best of our knowledge, studies of the price impact of traded volumes on the physical time scale are so far lacking, but will be provided here. For later comparison, we first work out the impact of traded volumes in individual stocks on the physical time scale. We refer to the aggregation of all traded volumes in a one-second interval as traded volume. To put all stocks on roughly the same footing, we normalize the traded volume of each stock at time t by dividing its average traded volume in that year, 2008,

$$v_i(t) = \frac{T \sum_{n=1}^{N(t)} v_i(t;n)}{\sum_{t=1}^{T} \sum_{n=1}^{N(t)} v_i(t;n)} \ , \tag{5.32}$$

where T denotes the total trading time, *i.e.* the days of trading in both stocks of a pair multiplied by 22200 seconds in each day of 2008. By binning the traded volumes, we obtain the price response as a function of the traded volumes.

Figure 5.1 shows the relation between the price self-response and the traded volumes $\langle R_{ii}(v_i, \tau=1) \rangle_i$ for individual stocks at time lag $\tau = 1$, averaged over the 31 stocks listed in Appendix A.3. The fit of power-law and logarithm functions to the empirical results indicates the relation is more in line with the former, $R(v) \sim v^{\delta}$ with an exponent $\delta = 0.51$. The exponent value is consistent with previous studies on the trading time scale. Lillo, Farmer and Mantegna [99] found $\delta \sim 0.5$ for small traded volumes and $\delta \sim 0.2$ for large traded volumes in the stocks from New York Stock Exchange in 1995. In another study of Lillo and Farmer, $\delta = 0.3$ resulted for Vodafone [98], one of the five highly capitalized stocks in the London Stock Exchange. The $\delta = 0.51$ in our case strongly corroborates the square-root impact function of traded volumes in single stocks, as found in Refs. [63, 130, 153].

Figure 5.2 displays the dependences of average cross-responses for different stocks i on the traded volumes at $\tau = 1$. The stocks i are C, MSFT, INTC, CSCO and BAC, the first five stocks with the largest average daily traded volumes among all stocks we studied. Here, four different dependences are discussed: How does the passive cross-responses $\langle R_{ij}^{(C)}(\tau) \rangle_j$ and $\langle R_{ij}^{(S)}(\tau) \rangle_j$ of the stock i depend on average on the traded volumes of stock i itself (Scenario I, see Fig. 5.2 a) and of the other stocks j (Scenario II, see Fig. 5.2 c), respectively? — How does the active cross-responses of stock i, *i.e.* $\langle R_{ji}^{(C)}(\tau) \rangle_j$ and $\langle R_{ji}^{(S)}(\tau) \rangle_j$ of stock i depend on average on the traded volumes of other stocks j (Scenario I, see Fig. 5.2 b) and of stock i itself (Scenario II, see Fig. 5.2 d), respectively? — In contrast to the average self-response versus traded volumes in Fig. 5.1, the average cross-responses raise for small traded volumes but decay for large ones. The nonlinear dependence complicates the impact function of traded volumes, even though the average cross-response as well as the average self-response of each stock i for traded volumes smaller than their average can be fitted by a power law.

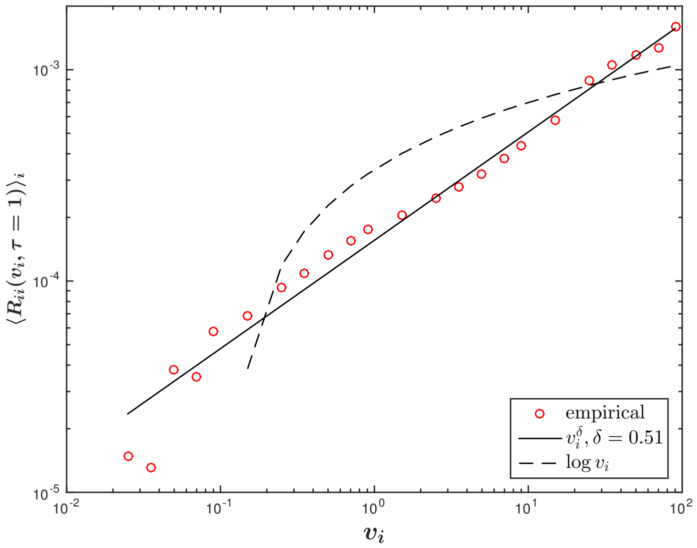


Figure 5.1: The average self-response over 31 stocks listed in Appendix A.3 versus traded volumes at time lag $\tau = 1$ on a doubly logarithmic scale. The open circles represent the empirical results, the solid line represents the power-law fit, and the dash line represents the natural logarithmical fit.

To make the analysis feasible, we focus on the region of traded volumes which affect the average cross-response considerably. As an example, we show in Fig. 5.3 the probability density distributions of traded volumes for MSFT and for other 30 stocks paired with MSFT. We conclude that the small volumes are traded frequently, while the large volumes are not. Not surprisingly, it is easier to analyze the self- or cross-correlators of trade signs in frequently traded stocks. As the power-law dependence in Fig. 5.2 levels off around the average of traded volumes, the ratios of the number of trades with volumes smaller than their average among the total trades are counted. It reaches 73% of the traded volume smaller than the average for MSFT and to 71% for the other 30 stocks paired with MSFT. The high proportions narrow down our study to trades below the average traded volume, described well by the power law instead of a complex nonlinear relation. In the power-law function of traded volumes, the exponent δ indicates the strength of price impact. As visible in Fig. 5.2, the δ ranges from 0.27 to 1.02 for the average cross-responses. To be more specific, the comparison of the δ in Fig. 5.2 a) and c) reveals that the stock price is more likely to be influenced by the traded volumes of the stock itself rather than by those of the other stocks. Put differently, the passive response of stock i depends strongly on the traded volumes of stock i itself, but weakly on the volumes of other stocks j. For the active response of stock i, the impact of traded volumes from stock i itself differs across different i, see Fig. 5.2 d), while this impact from other stocks j basically keeps stable whatever the stock i might be, see Fig. 5.2 b). This is so, because neither the average price changes nor the traded volumes of stocks j, over which we average, vary too much for different stocks i.

The analyses in Figs. 5.2 and 5.3 yield a power-law impact function of traded volumes for the region of volumes smaller than their average. Hence, for passive and active cross-responses of MSFT to the other 30 stocks considered, we arrive at the following

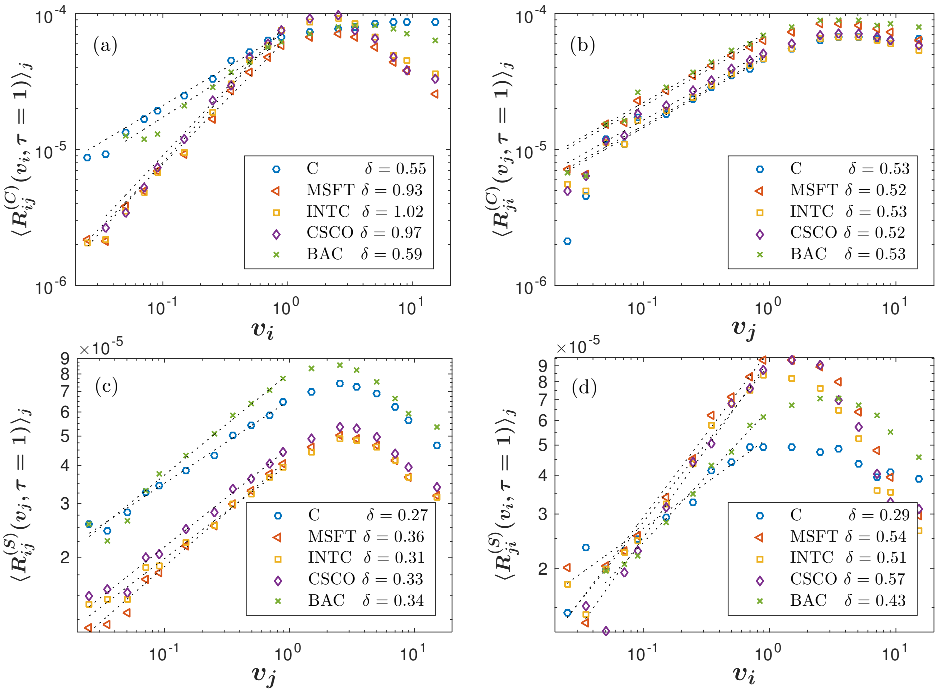


Figure 5.2: The dependence of the average cross-responses on the traded volumes at time lag $\tau = 1$ on a doubly logarithmic scale. (a) The passive response $\langle R_{ij}^{(C)}(v_i,\tau = 1)\rangle_j$ of stock i for Scenario I. (b) The active response $\langle R_{ji}^{(C)}(v_j,\tau = 1)\rangle_j$ of stock i for Scenario I. (c) The passive response $\langle R_{ij}^{(S)}(v_j,\tau = 1)\rangle_j$ of stock i for Scenario II. (d) The active response $\langle R_{ji}^{(S)}(v_i,\tau = 1)\rangle_j$ of stock i for Scenario II. Here, the stocks i are C, MSFT, INTC, CSCO and BAC, respectively. The other 30 stocks j are listed in Appendix A.3. The markers represent the empirical results and the dot lines represent the power-law fits with the exponents δ.

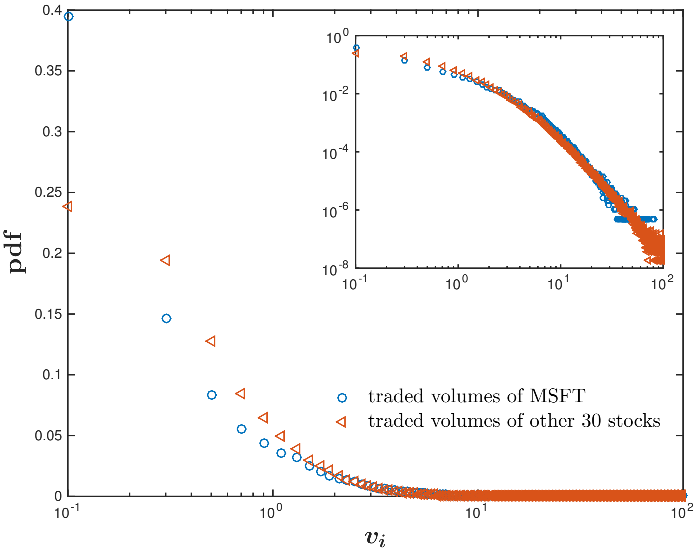


Figure 5.3: The probability density distribution of traded volumes for MSFT and other 30 stocks on a logarithmic scale. MSFT and the other 30 stocks are listed in Appendix A.3. The insert is the probability density distribution of traded volumes on a doubly logarithmic scale.

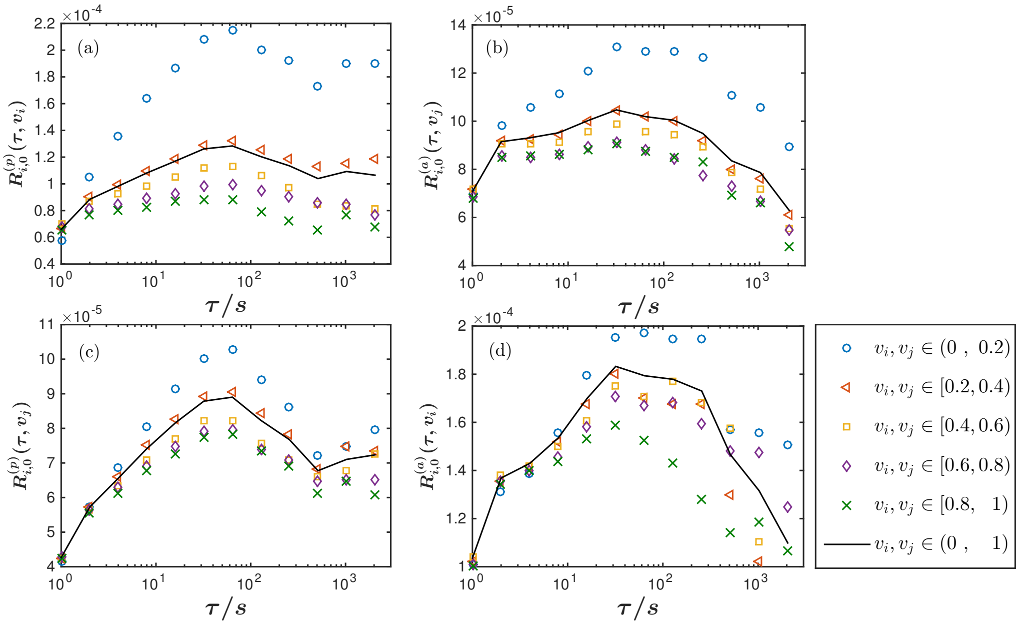


Figure 5.4: The dependence of average responses per share on the time lag τ in each region of traded volumes on a logarithmic scale. (a) The passive responses per share $R^{(p)}_{i,0}(\tau, v_i)$ of stock i in the regions of traded volumes v_i for Scenario I. (b) The active responses per share $R^{(a)}_{i,0}(\tau, v_j)$ of stock i in the regions of traded volumes v_j of other stocks j for Scenario I. (c) The passive responses per share $R^{(p)}_{i,0}(\tau, v_j)$ of stock i in the regions of the traded volumes v_j of other stocks j for Scenario II. (d) The active responses per share $R^{(a)}_{i,0}(\tau, v_i)$ of stock i in the regions of traded volumes v_i for Scenario II. The regions of v_i and v_j are (0, 0.2), [0.2, 0.4), [0.4, 0.6), [0.6, 0.8), [0.8, 1), and (0, 1), respectively. Here, the stock i is MSFT and the other stocks j are listed in Appendix A.3.

approximations of the average impact functions,

$$\langle f_i(v_i)\rangle_t \approx 0.28 \qquad \text{and} \qquad \langle f_j(v_j)\rangle_{t,j} \approx 0.56 \qquad \text{in Scenario I,} \tag{5.33}$$

$$\langle g_i(v_j)\rangle_{t,j} \approx 0.66 \qquad \text{and} \qquad \langle g_j(v_i)\rangle_{t,j} \approx 0.43 \qquad \text{in Scenario II.} \tag{5.34}$$

which are independent of time lag τ. Dividing those constants leads to the passive and active cross-responses per share, defined in Eqs. (5.11) and (5.12), respectively. Figure 5.4 shows the average cross-responses per share for MSFT versus time lag τ in each bin of the traded volumes. As seen, small traded volumes have stronger relative responses than large ones. The small traded volumes are more likely due to the fragmentation of large orders, by which the traders try to conceal their trading intention. The consecutive trades of small orders do not only prompt the self-correlator of trade signs in individual stocks [27], but may also lead to the cross-correlator of trade signs between stocks when trying to split two large orders of stocks in the same portfolio.

5.4 A construction to quantify price impacts

With numerical simulations and data comparisons, we try to quantify the price impacts. The construction sketched in the sequel should not be viewed as a fit. The number of

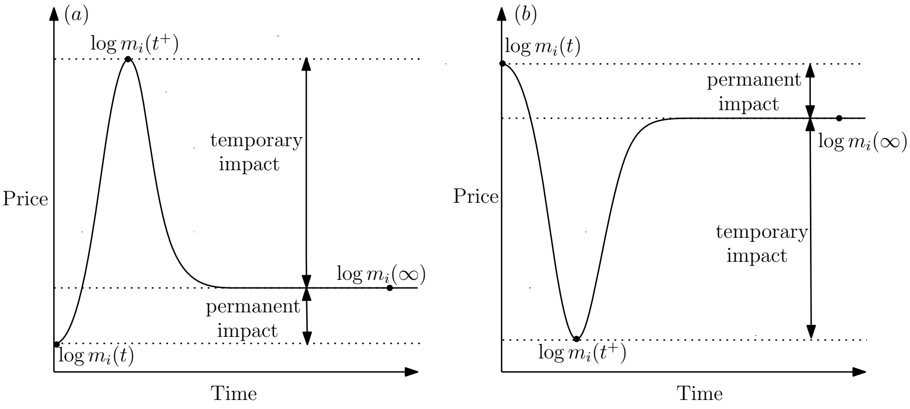


Figure 5.5: The sketch of the dependence of the price impact on the time. (a) the price impact of a buyer-initiated market order. (b) the price impact of a seller-initiated market order.

parameters is forbiddingly large. The purpose of the construction is to show that consistent and reasonable forms of the impact functions lie within the parameter domains. We choose a functional form for the impact and discuss the temporary and permanent impact components in Sec. 5.4.1. We then construct a strategy to determine the parameters for the impact functions in Sec. 5.4.2.

5.4.1 Impact function

The long-memory of the average sign correlations with $\gamma < 1$ leads to persistence in the price change. If other factors did not contribute to the price change, the price would be predictable and arbitrage would be possible on long time scales. As this is not only inconsistent with the EMH [54] and also in general unrealistic, the price at some larger time lag τ has to reverse. In a previous study [161], we argued that the EMH is not valid on short time scales, but restored on longer ones. The price impact thus has to decay, implying the analogous behavior for the impact function in the response functions. Mathematically speaking, as the sign correlators $\Theta_{ij}(\tau)$ decay as a power-law function (5.31), the integrals of the sign correlators over the time lag will increase according to $\tau^{1-\gamma}$ if $\gamma < 1$. When considering constant impact functions, the response functions (5.8) and (5.9) will also increase with $\tau^{1-\gamma}$. For $\tau \to \infty$, the response functions will tend to be infinite. A decaying impact function can outmaneuver this divergence. For the small part of stocks with $\gamma > 1$, the response functions (5.8) and (5.9) with $\tau^{1-\gamma}$ will not tend to infinity, and the divergence problem does not emerge.

A single trade can affect the stock price in different ways and with different strengths. It is thus highly unlikely to reverse the price exactly to the previous one while the price impact decays to zero. As the final price differs, in general, from the initial one, the price impact in our model comprises two components, a temporary impact and a permanent impact [88]. The temporary impact is measured by the difference between the instantaneous price $\log m_i(t^+)$ and the final price $\log m_i(\infty)$, while the permanent impact is measured by the difference between the final price $\log m_i(\infty)$ and the initial price $\log m_i(t)$. Here, the instantaneous price is induced by a buy or sell market order immediately. This is shown in Fig. 5.5.

The memory properties of the average sign correlator as well as the impact components require the impact function to include two parts,

$$G(\tau) = \frac{\Gamma_0}{\left[1 + \left(\frac{\tau}{\tau_0}\right)^2\right]^{\beta/2}} + \Gamma \ , \tag{5.35}$$

an algebraically decaying term with exponent β and with overall strength Γ_0 over the time scale τ_0 as well as a constant term Γ. Here, $G(\tau)$ is a general impact function, neither restricted to a special stock nor to an impact type. It stands for the self-impact function $G_{ii}(\tau)$, cross-impact function $G_{ij}(\tau)$, passive impact function $G_i^{(p)}(\tau)$ and active impact function $G_i^{(a)}(\tau)$, as all of them follow a power-law. In Eq. (5.35), the decaying term describes the temporary impact component, and converges to the price change with average sign correlations of long-memory. The constant term provides the permanent impact component, including the possibility of average sign correlations of short-memory. In our price impact model, we use an self- and a cross-impact function. Hence, for individual stocks, the temporary self-impact comes from the short-run liquidity cost. The difficulty to immediately find willing buyers or sellers induces a price concession from the initial price to the instantaneous price which yields more available volumes for trading [38, 83]. The permanent self-impact results from private information, which is subsequently incorporated in the new equilibrium price [38, 83]. The two self-impacts are measured in Ref. [8]. Across stocks, the temporary cross-impact is attributed to the transmission of trading information which, however, is always weakened by competing information. As the strategy traders may benefit much more from this trading information, a permanent cross-impact can result.

According to Eq. (5.35), the temporary and permanent impact components can be quantified from the impact of a single trade. For instance, if there was a buy market order of stock i at initial time t with the volume $v_i(t)$. The instantaneous price of stock i for Scenario I is

$$\begin{aligned} \log m_i(t^+) &= \log m_i(t) + G_{ii}(t^+ - t) f_i\big(v_i(t)\big) \\ &= \log m_i(t) + G_{ii}(0^+) f_i\big(v_i(t)\big) \ , \end{aligned} \tag{5.36}$$

where the superscript + indicates a time increment smaller than the distance to the next trade. After the time τ, the restored liquidity due to the new coming limit orders make the price reverse to

$$\log m_i(t+\tau) = \log m_i(t) + G_{ii}(\tau) f_i\big(v_i(t)\big) \ . \tag{5.37}$$

At $\tau \to \infty$ the price change will approach the limit

$$\begin{aligned} \log m_i(\infty) - \log m_i(t) &= G_{ii}(\infty) f_i\big(v_i(t)\big) \\ &= \Gamma f_i\big(v_i(t)\big) \ . \end{aligned} \tag{5.38}$$

This is the permanent impact and Γ measures the permanent impact per share of stock i,

$$\Gamma = \frac{\log m_i(\infty) - \log m_i(t)}{f_i\big(v_i(t)\big)} \ . \tag{5.39}$$

Furthermore, the price reversion of stock i occurs according to

$$\begin{aligned} \log m_i(t^+) - \log m_i(\infty) &= [G_{ii}(t^+ - t) - G_{ii}(\infty)] f_i\big(v_i(t)\big) \\ &= \Gamma_0 f_i\big(v_i(t)\big) \ , \end{aligned} \tag{5.40}$$

where we assume that the instantaneous impact function $G_{ii}(t^+ - t)$ equals the initial one, $G_{ii}(t^+ - t) = G_{ii}(0)$. Equation (5.40) states the temporary impact. Correspondingly, Γ_0 measures the temporary impact per share of the stock i,

$$\Gamma_0 = \frac{\log m_i(t^+) - \log m_i(\infty)}{f_i\big(v_i(t)\big)} \ . \tag{5.41}$$

Analogously, consider Scenario II and suppose the price change of stock i is triggered by a buy market order of stock j. The cross-impact function $G_{ij}(\tau)$ as well as the traded volumes of stock j contribute, such that Γ_0 and Γ measure the temporary and permanent impacts per share of stock j, respectively.

5.4.2 A construction to fix parameters

All impact functions (5.35) have the same form, but the details are encoded in the parameters, *i.e.*, the permanent impact component per share Γ, the temporary impact component per share Γ_0, the time scale τ_0 and the rate β of the decay. Unfortunately, we cannot determine these impacts directly from empirical data nor observe the latent characteristic of each impact depending on the time lag. To address this problem, we resort to the response functions in Sec. 5.2, which comprises the impact functions of time lag, the impact functions of traded volumes and the trade-sign correlators. The last two types of functions can be determined individually from empirical data, because, as analyzed in Sec. 5.3, the parameters ϑ_{ij} and γ in the trade-sign correlators (5.31) can be obtained by fitting to the empirical correlators, and the average impact functions of traded volumes are approximately constant. Hence, only four parameters in the impact functions (5.35) are to be determined.

Rather than an exact fit or simulation, we use the following strategy which is merely a construction using consistency arguments to look for appropriate parameters. To begin with, we restrict the response functions to Scenario I and to Scenario II, respectively, where we give initial values that are possible or close to ideal values for the four parameters in the response functions. By adjusting the values of the parameters in 10^6 iterations to reach a minimal error between the empirical responses per share and the theoretical ones, we find best suited values for the parameters in the Scenarios I and II, respectively. The comparisons of empirical and theoretical results for price responses and impacts are shown in Fig. 5.6, where the empirical impacts in Scenarios I and II result from Eqs. (5.20) and (5.23), respectively. The fit parameters and the normalized errors χ^2 [19], see Appendix B, to present the goodness of fits are listed in Table 5.1.

In a typical market situation, both Scenarios I and II have been accounted for. There are, however, arbitrarily many ways of doing this which all would leave us with too many parameters to be fitted. We thus have no other choice than resorting to an ad hoc method. It turned out that the interpolating ansatz

$$R_i^{(p)}(\tau) = wR_{i,0}^{(p,C)}\big|_{\mathrm{I}}(\tau)\big\langle f_i(v_i)\big\rangle_t + R_{i,0}^{(p,S)}(\tau)\big\langle g_i(v_j)\big\rangle_{t,j} \ , \tag{5.42}$$

$$R_i^{(a)}(\tau) = wR_{i,0}^{(a,C)}\big|_{\mathrm{I}}(\tau)\big\langle f_j(v_j)\big\rangle_{t,j} + R_{i,0}^{(a,S)}(\tau)\big\langle g_j(v_i)\big\rangle_{t,j} \ , \tag{5.43}$$

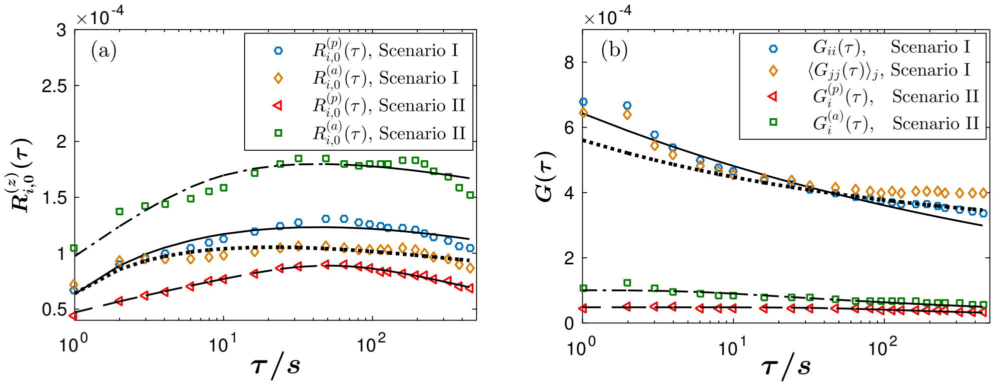


Figure 5.6: (a) The average cross-responses $R^{(z)}_{i,0}(\tau)$ per share versus the time lag τ in Scenario I and II, where the superscript z stands for either a or p, indicating active and passive cross-responses, respectively. (b) The impact functions $G(\tau)$ versus the time lag τ in Scenario I and II, where the $G(\tau)$ stands for the self-impact function $G_{ii}(\tau)$ of stock i, the average self-impact function $\langle G_{jj}(\tau)\rangle_j$ of other stocks j, the passive impact function $G^{(p)}_i(\tau)$ and the active impact function $G^{(a)}_i(\tau)$ of stock i, respectively. The stock i is MSFT and the other stocks j are listed in Appendix A.3.

with a weight w with $0 < w < 1$ yields good results. Here $R^{(p,C)}_{i,0}\big|_{\mathrm{I}}(\tau)$ and $R^{(a,C)}_{i,0}\big|_{\mathrm{I}}(\tau)$ are the passive and active response per share, respectively, of Scenario I, and $R^{(p,S)}_{i,0}(\tau)$ and $R^{(a,S)}_{i,0}(\tau)$ are the passive and active response per share, respectively, of Scenario III, for which the parameters will be fitted. We proceed as follows: we choose the values $w = 0.10,\ 0.30,\ 0.50,\ 0.70$ and 0.90 and for each given w we try to fit the parameters entering $R^{(p,S)}_{i,0}(\tau)$ and $R^{(a,S)}_{i,0}(\tau)$. The results are shown in Fig. 5.7 and 5.8 and in Table 5.2. Once more, we emphasize that the construction to quantify the response functions does not exclude that ansatzes other than (5.42) and (5.43) or an altogether different approach might yield comparable or even better results. Our point is only to show that our model is capable of reproducing the data with appropriate parameter values.

The weight determines the proportion of self-impacts in the average cross-responses, giving rise to different features of the price change. To find out the most appropriate weight w and to test how reasonable the parameters for Scenario III are, we need a different function depending on the same parameters. Therefore, in the third step, we approach this issue with the price diffusion function, introduced in Sec. 5.5.

Table 5.1: The fit parameters and errors for Scenarios I and II

Scenario	response function	impact function	Γ $(\times 10^{-6})$	Γ_0 $(\times 10^{-4})$	τ_0 [s]	β	χ^2_{R0} $(\times 10^{-6})$	χ^2_G $(\times 10^{-6})$
I	$R^{(p)}_{i,0}(\tau)$	$G_{ii}(\tau)$	0.0001	10.24	0.02	0.13	1.14	6.93
	$R^{(a)}_{i,0}(\tau)$	$\langle G_{jj}(\tau)\rangle_j$	260.90	20.97	0.00008	0.21	1.04	10.74
II	$R^{(p)}_{i,0}(\tau)$	$G^{(p)}_i(\tau)$	1.58	0.47	43.25	0.18	0.26	0.38
	$R^{(a)}_{i,0}(\tau)$	$G^{(a)}_i(\tau)$	0	1.01	5.73	0.16	1.90	1.73

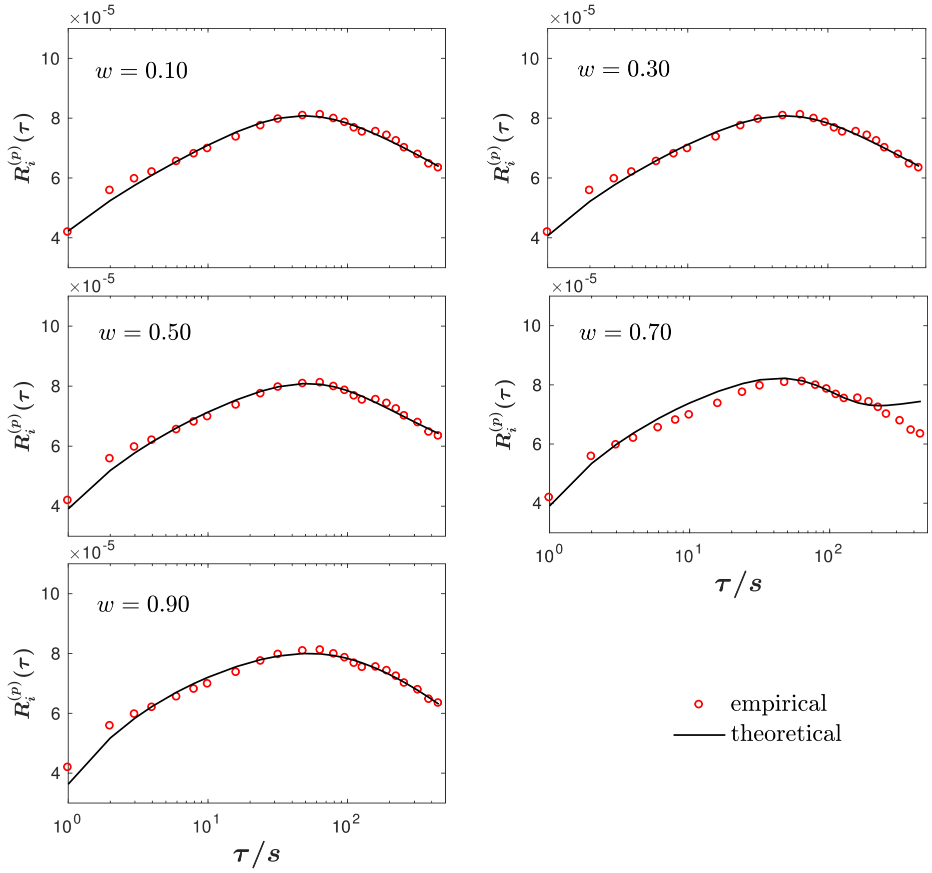


Figure 5.7: The passive responses versus the time lag τ in Scenario III with the weights $w = 0.10,\ 0.30,\ 0.50,\ 0.70$ and 0.90, respectively.

Table 5.2: The fit parameters and errors for Scenario III

response function	impact function	w	Γ	Γ_0 ($\times 10^{-4}$)	τ_0 [s]	β	χ_R^2 ($\times 10^{-6}$)
$R_i^{(p)}(\tau)$	$G_i^{(p)}(\tau)$	0.10	0	0.59	43.49	0.18	0.29
		0.30	0	0.42	51.55	0.25	0.30
		0.50	0	0.25	70.87	0.49	0.35
		0.70	0	0.11	270.68	15.41	0.98
		0.90	0.03	-274.43	59.68	-0.0002	0.45
$R_i^{(a)}(\tau)$	$G_i^{(a)}(\tau)$	0.10	0	3.49	0.011	0.15	1.05
		0.30	0	2.89	0.008	0.17	1.09
		0.50	0	2.57	0.004	0.19	1.14
		0.70	0	2.90	0.002	0.28	1.23
		0.90	0.33	-3292.80	823.19	-0.00008	1.03

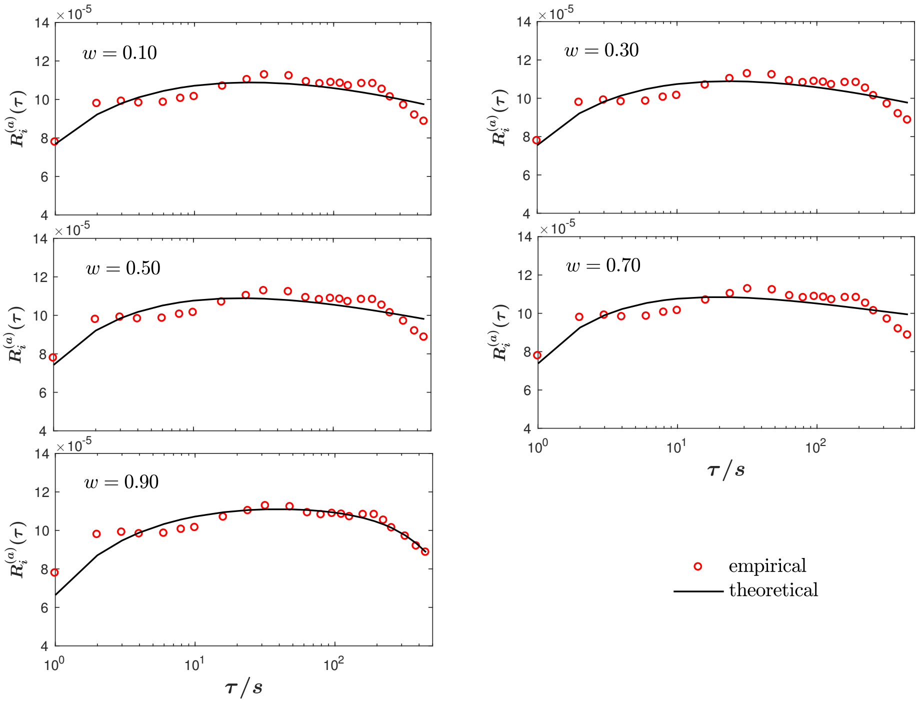


Figure 5.8: The active responses versus the time lag τ in Scenario III with the weights $w = 0.10,\ 0.30,\ 0.50,\ 0.70$ and 0.90, respectively.

5.5 Relation to correlated diffusion

We study how the self- and cross-impacts are related to the correlated diffusion of prices. We define the diffusion function in Sec. 5.5.1, and then discuss the time-lag dependence for three stochastic processes modelling the correlated motion of different stock prices in Sec. 5.5.2. We further analyze the stochastic processes depending on the weight of self-impacts, and compare the empirical and theoretical diffusions to corroborate the choice of parameters in Sec. 5.5.3.

5.5.1 Price diffusion functions

We begin with briefly recalling some properties of diffusion in two dimensions. Consider a particle moving in a flat two-dimensional space with coordinates (x, y). The particle position is changed by random increments u and v in the directions x and y, respectively. We introduce the probability $p(x, y|t)dxdy$ to find the particle in the area element $dxdy$ at time t with the joint probability density $p(x, y|t)$. The partial differential equation for the joint probability density reads

$$\frac{\partial p(x,y|t)}{\partial t} = \frac{\langle u^2\rangle}{2\tau}\frac{\partial^2 p(x,y|t)}{\partial x^2} + \frac{\langle v^2\rangle}{2\tau}\frac{\partial^2 p(x,y|t)}{\partial y^2} + \frac{\langle uv\rangle}{\tau}\frac{\partial}{\partial x}\frac{\partial}{\partial y}p(x,y|t)\ . \tag{5.44}$$

For the convenience of the reader, details are given in Appendix C. The angular brackets indicate averages over the distribution of u and v. We emphasize the presence of the last term in Eq. (5.44) which is non-zero if the random increments are not independent. Hence, the probability density $p(x, y|t)$ does not factorize. In general, the diffusion equation can be written as [76],

$$\frac{\partial p(\vec{r}|t)}{\partial t} = \nabla \cdot \left(\hat{D} \nabla p(\vec{r}|t) \right) \ , \tag{5.45}$$

where $\hat{D}$ is a symmetric diffusion tensor in homogeneous and anisotropic media,

$$\hat{D} = \left[\begin{array}{cc} \hat{D}_{xx} & \hat{D}_{xy} \\ \hat{D}_{yx} & \hat{D}_{yy} \end{array} \right] \ . \tag{5.46}$$

This tensor is not space dependent and the two-dimensional diffusion equation becomes

$$\frac{\partial p(x,y|t)}{\partial t} = \hat{D}_{xx} \frac{\partial^2 p(x,y|t)}{\partial x^2} + \hat{D}_{yy} \frac{\partial^2 p(x,y|t)}{\partial y^2} + 2\hat{D}_{xy} \frac{\partial}{\partial x} \frac{\partial}{\partial y} p(x,y|t) \ , \tag{5.47}$$

with $\hat{D}_{xy} = \hat{D}_{yx}$. Equations (5.44) and (5.47) coincide and allow for the identification

$$\langle u^2 \rangle = 2\hat{D}_{xx}\tau \ , \quad \langle v^2 \rangle = 2\hat{D}_{yy}\tau \ , \quad \text{and} \quad \langle uv \rangle = 2\hat{D}_{xy}\tau \ . \tag{5.48}$$

In another terminology, the elements of the diffusion tensor $\hat{D}$ are constants in the case of Brownian motion, see details in Appendix C.

We apply the results (5.48) to the motion of two stocks with indices i and j and obtain the price diffusion function for these two different stocks,

$$D_{ij}(\tau) = \Big\langle r_{ij}(t,\tau)\, r_{ji}(t,\tau) \Big\rangle_t \ . \tag{5.49}$$

This diffusion function can be positive, negative or zero. To accumulate statistics, it is helpful to carry out an additional average over the stock index j which defines the quantity

$$\langle D_i \rangle(\tau) = \Big\langle r_{ij}(t,\tau)\, r_{ji}(t,\tau) \Big\rangle_{t,j} \ , \tag{5.50}$$

where $r_{ij}(t,\tau)$ and $r_{ji}(t,\tau)$ are the logarithmic midpoint price changes of stocks i and j, respectively. They can be calculated by Eq. (3.1). We notice that the diffusion functions in (5.48) read $2\hat{D}_{xx}\tau$ and so on, *i.e.* they are linear functions in time, while the diffusion coefficients $\hat{D}_{xx}$, etc, are constants. To test the simulated results for all the scenarios, we employ the price diffusion function, which reflects the price fluctuations with time lag τ. For each stock, due to different causes, *i.e.* the short-run liquidity from the stock itself and the trading information from other stocks, the price change contains two components, *i.e.* Eqs. (5.5) and (5.6). Hence, the price diffusion function can be decomposed into four individual sub-functions for different combinations of the components,

$$\langle D_i \rangle(\tau) = \langle D_i \rangle^{(LL)}(\tau) + \langle D_i \rangle^{(II)}(\tau) + \langle D_i \rangle^{(LI)}(\tau) + \langle D_i \rangle^{(IL)}(\tau) \ . \tag{5.51}$$

In view of Eq. (5.7), we define the diffusion functions $\langle D_i \rangle^{(XY)}(\tau)$ with (XY) indexed as (LL), (II), (LI), and (IL) in the following way

$$\begin{aligned} \langle D_i \rangle^{(LL)}(\tau) &= \Big\langle r_{ii}^{(L)}(t,\tau) r_{jj}^{(L)}(t,\tau) \Big\rangle_{t,j} \ , \\ \langle D_i \rangle^{(II)}(\tau) &= \Big\langle r_{ij}^{(I)}(t,\tau) r_{ji}^{(I)}(t,\tau) \Big\rangle_{t,j} \ , \\ \langle D_i \rangle^{(LI)}(\tau) &= \Big\langle r_{ii}^{(L)}(t,\tau) r_{ji}^{(I)}(t,\tau) \Big\rangle_{t,j} \ , \\ \langle D_i \rangle^{(IL)}(\tau) &= \Big\langle r_{ij}^{(I)}(t,\tau) r_{jj}^{(L)}(t,\tau) \Big\rangle_{t,j} \ . \end{aligned} \tag{5.52}$$

With Eqs. (5.5) and (5.6), we can cast all diffusion functions into to a unified expression,

$$\begin{aligned}
\langle D_i \rangle^{(XY)}(\tau) &= \sum_{t \le t' < t+\tau} G_1(t+\tau-t') G_2(t+\tau-t') \Theta_1(0) V \\
&+ \sum_{t' < t} \Big[G_1(t+\tau-t') - G_1(t-t') \Big] \Big[G_2(t+\tau-t') - G_2(t-t') \Big] \Theta_1(0) V \\
&+ \langle \Delta_i \rangle^{(XY)}(\tau) V + \tau D_\eta^{(XY)} \ , \qquad (5.53)
\end{aligned}$$

where the numbers V are averages of products of the traded volumes, see Table 5.3. The noise contributions $D_\eta^{(XY)}$, which are assumed to be constants, stem form the random price fluctuations. Moreover, $\langle \Delta_i \rangle^{(XY)}(\tau)$ is the contribution induced by the correlation between the impact functions and the sign correlators, it is given by

$$\begin{aligned}
\langle \Delta_i \rangle^{(XY)}(\tau) &= \sum_{t \le t' < t'' < t+\tau} G_1(t+\tau-t') G_2(t+\tau-t'') \Theta_2(t''-t') \\
&+ \sum_{t \le t'' < t' < t+\tau} G_1(t+\tau-t') G_2(t+\tau-t'') \Theta_1(t'-t'') \\
&+ \sum_{t' < t'' < t} \Big[G_1(t+\tau-t') - G_1(t-t') \Big] \Big[G_2(t+\tau-t'') - G_2(t-t'') \Big] \Theta_2(t''-t') \\
&+ \sum_{t'' < t' < t} \Big[G_1(t+\tau-t') - G_1(t-t') \Big] \Big[G_2(t+\tau-t'') - G_2(t-t'') \Big] \Theta_1(t'-t'') \\
&+ \sum_{t \le t'' < t+\tau} \sum_{t' < t} \Big[G_1(t+\tau-t') - G_1(t-t') \Big] G_2(t+\tau-t'') \Theta_2(t''-t') \\
&+ \sum_{t \le t' < t+\tau} \sum_{t'' < t} G_1(t+\tau-t') \Big[G_2(t+\tau-t'') - G_2(t-t'') \Big] \Theta_1(t'-t'') \ . \qquad (5.54)
\end{aligned}$$

Table 5.3 summarizes all appearing quantities appearing in the above Eqs. (5.53) and (5.54). Equation (5.51) describes the price diffusion for Scenario III. As for Scenarios I and II, the price diffusions

$$\langle D_i \rangle(\tau) = \begin{cases} \langle D_i \rangle^{(LL)}(\tau) & \text{(Scenario I)} \\ \langle D_i \rangle^{(II)}(\tau) & \text{(Scenario II)} \end{cases} , \qquad (5.55)$$

result from only one component of price change.

5.5.2 Correlated motion of prices

In individual stocks, the price stochastic process is often be interpreted as either be normal diffusion, super-diffusion or sub-diffusion [82]. For normal diffusion, the coefficients, such

Table 5.3: The quantities in diffusion functions

$\langle D_i \rangle^{(XY)}(\tau)$	V	$G_1(\tau)$	$G_2(\tau)$	$\Theta_1(\tau)$	$\Theta_2(\tau)$	$D_\eta^{(XY)}$
$\langle D_i \rangle^{(LL)}(\tau)$	$\langle f_i(v_i) f_j(v_j) \rangle_j \approx 0.179$	$G_{ii}(\tau)$	$\langle G_{jj}(\tau) \rangle_j$	$\Theta_i^{(p)}(\tau)$	$\Theta_i^{(a)}(\tau)$	$D_\eta^{(LL)}$
$\langle D_i \rangle^{(II)}(\tau)$	$\langle g_i(v_j) g_j(v_i) \rangle_j \approx 0.308$	$G_i^{(p)}(\tau)$	$G_i^{(a)}(\tau)$	$\Theta_i^{(a)}(\tau)$	$\Theta_i^{(p)}(\tau)$	$D_\eta^{(II)}$
$\langle D_i \rangle^{(LI)}(\tau)$	$\langle f_i(v_i) g_j(v_i) \rangle_j \approx 0.208$	$G_{ii}(\tau)$	$G_i^{(a)}(\tau)$	$\Theta_{ii}(\tau)$	$\Theta_{ii}(\tau)$	$D_\eta^{(LI)}$
$\langle D_i \rangle^{(IL)}(\tau)$	$\langle g_i(v_j) f_j(v_j) \rangle_j \approx 0.416$	$G_i^{(p)}(\tau)$	$\langle G_{jj}(\tau) \rangle_j$	$\langle \Theta_{jj}(\tau) \rangle_j$	$\langle \Theta_{jj}(\tau) \rangle_j$	$D_\eta^{(IL)}$

as $\hat{D}_{xx}$ or $\hat{D}_{yy}$ in Eq. (5.48), are constant. If the price changes persistently, we have super-diffusion, in which, *e.g.*, a high price is more likely to be followed by another high price. If this process continues for a long time, the price is eventually likely to be higher than the initial one. However, if the process persists only for a short time and then reverses, it is said to change anti-persistently, and typically sub-diffusion occurs. A high price is likely to be followed by a low price, and *vice versa*. This can last for a long time. In mathematical terms, super- and sub-diffusion are characterized by a non-linear time-dependence of the diffusion function. The three processes can be associated with the Hurst exponent H [73], which is used to measure long-memory process for the auto-correlation of the time series [98, 109],

$$D_{ii}(\tau) = \left\langle r_{ii}^2(t,\tau) \right\rangle_t \sim \tau^{2H} \ . \tag{5.56}$$

Here, $H = 1/2$ indicates normal diffusion, while $H > 1/2$ and $H < 1/2$ correspond to super- and sub-diffusion, respectively [27, 109].

Across different stocks, we transfer this way of analysis. The diffusion function $\langle D_i \rangle(\tau)$ characterizes the correlated motion of the prices. Hence, we introduce an expression analogous to the above one,

$$\langle D_i \rangle(\tau) \sim \tau^{2\lambda} \tag{5.57}$$

with a new exponent λ, which is not necessarily equal to H. This is related to some studies of fractional Brownian motion [107, 126]. For an empirical analysis, it is useful to divide out a linear time dependence according to

$$\sqrt{|\langle D_i \rangle(\tau)|/\tau} \sim \tau^{\lambda - \frac{1}{2}} \ . \tag{5.58}$$

As we are here only interested in the time behavior, we use an absolute value to prevent this expression from being imaginary in case of a negative diffusion function. For normal diffusion with $\lambda = 1/2$, we obtain the constant diffusion coefficient. If $\lambda > 1/2$, the function (5.58) is an increasing function of τ. Thus, compared to the normal diffusion, the correlations increase in time and — if $\langle D_i \rangle(\tau) > 0$ — a high price of one stock becomes more likely to be followed by a high price of another stock. This is super-diffusion for correlated stocks. In contrast, if $\lambda < 1/2$, the function (5.58) decreases with τ and the correlation decays as compared to the normal diffusion. This is sub-diffusion for correlated stocks.

5.5.3 Consistency of our model

With the parameters in Scenario III, we work out the theoretical price diffusion according to Eq. (5.51), where the total noise contribution, *i.e.* the sum over the $D_\eta^{(XY)}$, is set to 1×10^{-8} for all cases. The comparisons between empirical and theoretical price diffusions are shown in Fig. 5.9. Theoretical result largely depends on the weight w, presenting distinct stochastic processes. For $w < 0.50$, due to the small proportion of self-impacts, the cross-impacts dominate in the price change, resulting in a super-diffusive process for the motion of prices in the first 500 seconds. The overestimated cross-impacts fail to reverse the price effectively, and thus provide opportunities of arbitrage, which violates the EMH [54]. For $w > 0.50$, the process transforms from super-diffusion at first to sub-diffusion later on, as a higher proportion of self-impacts quickly prevents prices from moving in a correlated and persistent manner. The motion of prices in opposite directions

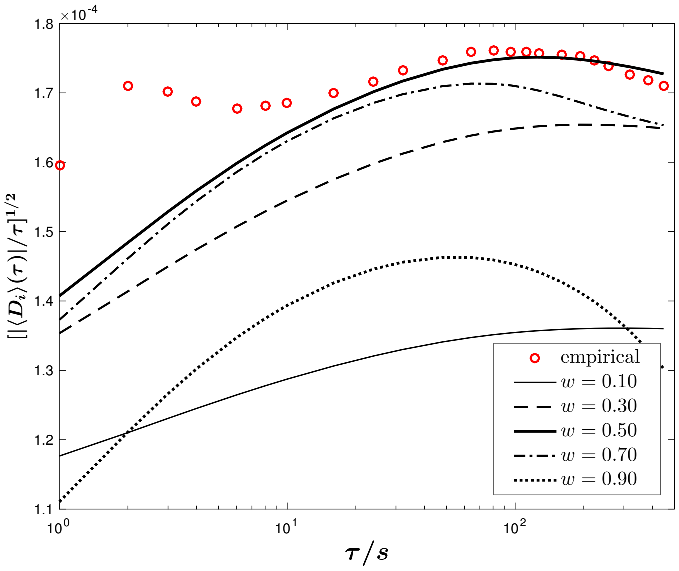


Figure 5.9: The comparison of the theoretical results of $[|\langle D_i\rangle(\tau)|/\tau]^{1/2}$ in Scenario III, where the weights are $w = 0.10,\ 0.30,\ 0.50,\ 0.70$ and 0.90, respectively. These theoretical results are also compared with the empirical result. The random fluctuation is set to $\sum_{(XY)} D_\eta^{(XY)} = 1 \times 10^{-8}$ in each case, where (XY) is indexed as (LL), (II), (LI), and (IL) for Scenario III.

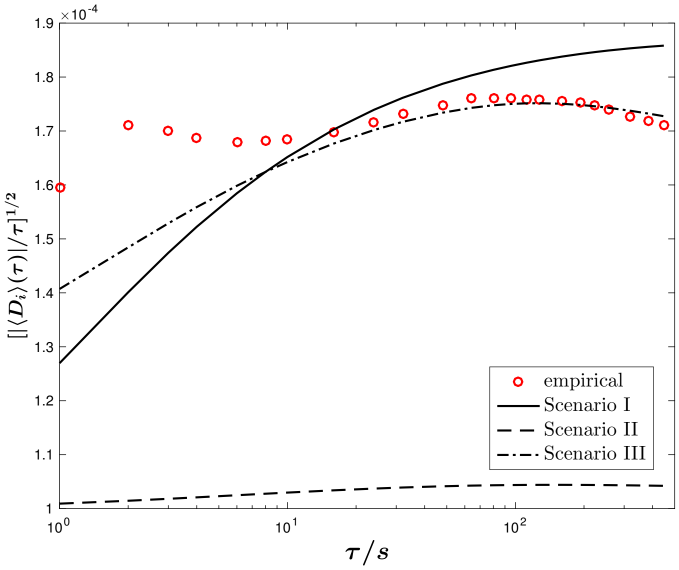


Figure 5.10: The comparison of the theoretical results of $[|\langle D_i\rangle(\tau)|/\tau]^{1/2}$ in three scenarios, where for Scenario III, we only show the theoretical result with the weight $w = 0.5$. These theoretical results are also compared with the empirical result. The random fluctuation is set to $\sum_{(XY)} D_\eta^{(XY)} = 1 \times 10^{-8}$ in each scenario, where (XY) is indexed as (LL) for Scenario I, as (II) for Scenario II, and as (LL), (II), (LI), and (IL) for Scenario III.

opens possible opportunities of arbitrage as well. The consistency between empirical and theoretical diffusions lead us to favor the weight $w = 0.50$, especially at larger time lags. In this case, the diffusion coefficient is approximately constant, implying a normal diffusion that the price movement cannot be predicted. Also, we compare the price diffusion in the extreme cases, *i.e.*, Scenarios I and II with only self-impacts or cross-impacts. As shown in Fig. 5.10, the two scenarios all exhibit super-diffusive behaviors. In contrast, the better agreement for $w = 0.5$ in Scenario III lets us conclude that the average cross-responses can indeed be described by the two response components, namely one including the sign self-correlators together with cross-impacts, the other one the sign cross-correlators together with self-impacts. The results are also supported by the small errors χ_D^2 of the price diffusion, listed in Table 5.4. The above interpretation is also consistent with the line of arguing in Ref. [27] where only the self-responses were addressed.

5.6 Price impacts of individual stocks

The price impacts of individual stocks i include the self-impact $G_{ii}(\tau)$ as well as the cross-impacts. For the latter we introduced a passive impact $G_i^{(p)}(\tau)$ and an active impact $G_i^{(a)}(\tau)$. The difficulty to obtain the empirical impacts leads us to describe them with the function (5.35). Using consistency arguments in Secs. 5.4 and 5.5, we are able to fix the parameters. Thus, the resulting impact function is only a possible one, not excluding others with similar properties. Taking MSFT as an example, with the parameters in Scenario III and the weight $w = 0.50$, we are able to provide the price impacts versus time lag, as shown in Fig. 5.11. The resulting parameters for the impact functions are listed in Table 5.5.

Table 5.5 shows the temporary and permanent components, Γ_0 and Γ in the self-impact of MSFT, measuring the impact of a single trade of the stock on its own price after a time τ. In our model the self-impact is due to the short-run liquidity in general, however, the existence of the above two components requires more explanation. This is reminiscent of Ref. [61], where the price impact is separated into a mechanical impact and an informational impact. The mechanical impact of a market order is referred to as the change of future prices without any future change in decision making. The average mechanical impact decays to zero monotonically in time in a power-law fashion, similar to the temporary component in our self-impact. The informational impact is the remainder after the mechanical impact being removed. It grows with time and approaches a constant value, just as the permanent component in our self-impact. Thus, a line of reasoning put forward in Ref. [61] can be partly transferred to our case. As the incoming limit orders following the instantaneous price change offer more liquidity for the market and reverse the trade price towards the previous price, the temporary component is still induced by the short-run liquidity, but the reversed final price is less likely to exactly be the previous one. Therefore, the induced permanent component as well as the informational impact may result from private information. If individual agents possess private information to

Table 5.4: The fit errors of $[|\langle D_i\rangle(\tau)|/\tau]^{1/2}$ in three scenarios

Scenario	I	II	III				
			$w = 0.10$	$w = 0.30$	$w = 0.50$	$w = 0.70$	$w = 0.90$
$\chi_D^2(\times 10^{-5})$	0.28	1.43	0.85	0.31	0.17	0.21	0.73

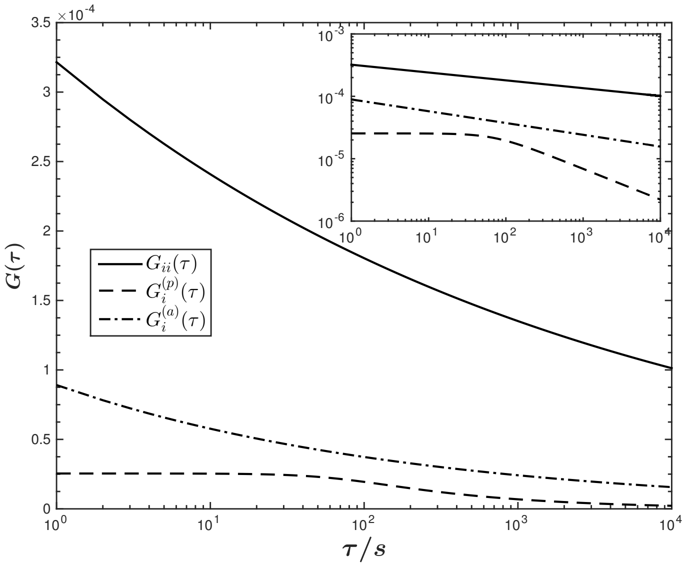


Figure 5.11: The dependence of price impacts $G_{ii}(\tau)$, $G_i^{(p)}(\tau)$ and $G_i^{(a)}(\tau)$ on the time lag τ on a logarithmic scale. The price impacts are calculated with the parameters in Scenario III and the weight $w = 0.50$. The insert is the dependence of price impacts on the time lag on a doubly logarithmic scale.

trade, a price change due to the trading affects will emerge. Other intelligent agents will then adjust their market expectations based partly on the private information and partly on all available public information. The private information is made public via the trade price, visible in the permanent component of the self-impact. However, in contrast to the temporary component of the order of 10^{-4}, the permanent component is very small, of the order of 10^{-11}.

For the cross-impact of MSFT, the permanent component is absent. Hence, either passive or active impacts only contain the temporary component. We recall that the passive impact is the price change of an individual stock i induced by single trades of different stocks and the active impact is the average price change of different stocks triggered by a single trade of stock i. The cross-impact accounts for the trading information transmitted across stocks. We stress once more that the trading information indicates the trading directions (buy and sell) and trading volumes of other stocks, unrelated to private information, news and so on, which are viewed as competing information. Because of the interference of this competing information, the influence of trading information can neither remain for a long time nor be as strong as the self-impact. Moreover, the difficulty for the traders to distinguish useful and useless trading information makes a permanent

Table 5.5: The parameters of impact functions in Scenario III with $w = 0.50$

impact function	Γ $(\times 10^{-10})$	Γ_0 $(\times 10^{-4})$	τ_0 [s]	β
$G_{ii}(\tau)$	0.5	5.12	0.025	0.13
$G_i^{(p)}(\tau)$	0	0.25	70.873	0.49
$G_i^{(a)}(\tau)$	0	2.57	0.004	0.19

impact unlikely.

In addition, although the passive impact shows weak temporary component, it decays very slowly with the decay time scale of $\tau_0 = 70.873$ s. The decay of the impact at about 70 s is visible in Fig. 5.11. It is a consequence of the reduction of the noise induced by random trades, as the passive impact extracts the trading information from multiple stocks. In contract, the trading information from only one stock leads to small decay time scales either for self-impact or for active impact.

5.7 Conclusions

We put forward a price impact model for the average cross-response functions for individual stocks. It comprises two impact functions, *i.e.* a self-impact function and a cross-impact function. We introduced and studied three scenarios, namely cross-responses exclusively due to the trade-sign cross-correlators (Scenario I), or to the trade-sign self-correlators (Scenario II), or to both (Scenario III), respectively. Thereby, we managed to greatly reduce the complexity of the problem, and facilitated the determination of the model parameters.

In the empirical analysis we demonstrated that, for most stocks, the self-correlators and average cross-correlators of trade signs have a long memory, which provides a strong support for setting up the impact function of the time lag. The empirical analysis also revealed power-law relations between the average cross-responses and the traded volumes that are smaller than their average. The relations hold regardless of the passive or active cross-responses and regardless of the traded volumes of the impacted or impacting stock. To further explore the parameter space of our model, we defined the average cross-responses per share, which are the average cross-responses divided by the average impact functions of traded volumes. Our empirical analysis manifests that the smaller the volumes, the larger are the responses per share. This indicates the fragmentation of large orders.

Using the average cross-response functions and the price diffusion functions, we gave a construction to fix the parameters for the impact functions. The results indicate that there are two components present in the response functions. One contains the self-impacts that suppress the amplification effects due to sign cross-correlators, the other one contains the cross-impacts that suppress the amplification effects due to sign self-correlators.

As an example, we studied the price impacts of MSFT. The self-impact includes temporary and permanent components. The temporary component, as a decaying power-law function, is the result of short-run liquidity, while the permanent component, approaching a constant, is due to private information. However, the permanent component is rather small compared to the temporary one. The cross-impacts, separated into an active impact and a passive impact, only contain the temporary component. It comes from the trading information transmitted across stocks. In our study, the trading information is limited to the trading directions, *i.e.* buy and sell, and traded volumes of impacting stocks. The interference of competing information weakens the influence of the trading information. Consequently, the cross-impacts are neither as strong as the self-impact nor persistent permanently.

Chapter 6

Trading strategies with cross-impact costs: an application of cross-responses

6.1 Introduction

The price change due to a market order will lead to an extra cost for trading. In particular, large market orders will consume the volumes at several price levels near the best quote and shift the trade price to a higher or a lower level immediately. To reduce the transaction cost, traders are more likely to execute a sequence of small orders, which are split from a large order. By doing so, it largely lowers the cost from the price impact. On the basis of the order splitting, a lot of optimal trading strategies are proposed [5–7, 49, 65, 66, 120]. Most of these strategies, however, focus on the self-impact cost in single stocks, ignoring the cross-impact cost between stocks. Only very few of execution strategies take the cross-impact into account [7, 34, 143]. In particular, Almgren and Chriss (2001) considered the optimal execution for portfolio transactions by minimizing a combination of volatility risk and transaction costs, where the permanent and temporary impacts were set to be linear in the rate of trading [7]. Cartea *et al.* (2015) constructed an optimal execution strategy for liquidating a basket of assets whose price were co-integrated [34]. In their model, they assumed linear temporary and permanent price impacts in the speed of trading as well. Recently, Schneider and Lillo (2016) extended the framework of trading strategies from Gatheral (2010) [65] to multiple assets [143]. They also discussed the possible constraints on the shape and size of the cross-impact. On condition of absence of dynamical arbitrage, they found the cross-impact must be odd and linear function of trading intensity and symmetric for a pair of stocks. However, empirical studies show a nonlinear impact function of order size either in single stocks or across multiple stocks [99, 135, 158].

Our study aims to construct a trading strategy regarding to the cross-impact cost, and reveal the influence of cross-impacts on the trading strategy. We thus extend the framework of trading strategies of Gatheral (2010) from single stocks to the two-dimensional case, where the trade price is generated by the price impact model from Ref. [158], also see Chapter 5. Our trading strategy for a pair of stocks results from three parameters, *i.e.*, the rate of trading for each stock, and the ratio of trading periods of two stocks. By minimizing the cross-impact cost, we thus can obtain an optimal trading strategy. According to the empirical analysis, we employ a power-law impact functions of time lag and of order size to calculate the costs for a specific case.

This chapter is organized as follows. In Sec. 6.2, making use of the price impact model with both self- and cross-impacts, we construct a trading strategy in terms of the cross-impact cost. In Sec. 6.3, we apply the trading strategy to a pair of stocks and calculate the cost depending on the tree parameters, where we quantify the cross-impacts of traded volumes and of time lag for the need of computation. We conclude our results in Sec. 6.4. This chapter refers to Ref. [157]. In the following, I use the original text from Ref. [157].

6.2 Model setup

In Sec. 6.2.1, we introduce the price impact model with both self- and cross-impacts, and transform the discrete model into a continuous one. In Sec. 6.2.2, we discuss the cost of trading and derive in detail the function for cross-impact costs. In Sec. 6.2.3, We construct a trading strategy with three free parameters.

6.2.1 Trade price

In financial markets, a buy market order will raise or maintain the stock price, while a sell market order will drop or maintain the stock price. The price change on average due to a buyer-initiated or a seller-initiated trade refers to the price impact. This impact can be propagated to the price in a future time. Therefore, the stock price is the result of the accumulation of impacts from all past trades. The impact can be classified as a self-impact and a cross-impact [158]. The self-impact is related to the trades from the impacted stock itself. As the insufficient volumes at the best ask or bid cannot fulfil the large demand of market orders in a short term, leading to a lack of the short-run liquidity, the traded price has to be conceded to a higher ask price or a lower bid price. This instantaneous price change for stock i at time t thus results from the impact of traded volumes $f_i\big(v_i(t)\big)$. However, the price change is not fixed with time [27]. When new limit orders come into the order book, the price is reversed towards previous one gradually. Such price change due to the restoration of liquidity in a long term is characterized by a self-impact function $G_{ii}(\tau)$. Differently, the cross-impact [16, 160, 161] across different stocks is more likely due to the trading information containing the traded volumes and trade signs rather than other information, because the trade of one stock cannot consume the volume of another stock directly showing in the order book. By the transmission of trading information, the stock j has an impact $g_i\big(v_j(t)\big)$ on the stock i. With the time increasing, more and more competitive information, such as news, covers the trading information and weakens the impact of stock j gradually. This decaying process is depicted by a cross-impact function $G_{ij}(\tau)$.

Therefore, taking into account the price impacts from two different stocks i and j, the logarithmic midpoint price of stock i at time t can be expressed [158] as

$$\begin{aligned}
\log m_i(t) &= \sum_{t'<t}\Big[G_{ii}(t-t')f_i\big(v_i(t')\big)\varepsilon_i(t')+\eta_{ii}(t')\Big] \\
&+ \sum_{t'<t}\Big[G_{ij}(t-t')g_i\big(v_j(t')\big)\varepsilon_j(t')+\eta_{ij}(t')\Big] \\
&+ \log m_i(-\infty) \ . \qquad (6.1)
\end{aligned}$$

Here, $m_i(t)$ is the midpoint of the best ask price and the best bid price at time t. $\varepsilon_i(t)$ and $\varepsilon_j(t)$ are the trade signs of stocks i and j, respectively. $\varepsilon_i(t) = +1$ means a buy market order of stock i at time t, while $\varepsilon_i(t) = -1$ means a sell market order. $\varepsilon_i(t) = 0$ represents a lack of trading or a balance of buy and sell market orders at t. The trade signs in Eq. (6.1) clearly indicate the directions of price changes, a buy for price increasing and a sell for price decreasing. Apart from the causes already described by the impact functions, *i.e.*, $f_i\big(v_i(t)\big)$, $g_i\big(v_j(t)\big)$, $G_{ii}(\tau)$, $G_{ij}(\tau)$, all remaining causes of the price change arising from stock i and stock j are modelled by the random variables $\eta_{ii}(t)$ and $\eta_{ij}(t)$, respectively.

In Eq. (6.1), the impact functions of traded volumes $f_i\big(v_i(t)\big)$ and $g_i\big(v_j(t)\big)$ describe the unsigned price changes caused by the unsigned volumes $v_i(t)$ and $v_j(t)$ of market orders. That means buying in and selling out the same volume have the same strength of impact on the stock i, but the impact raising or dropping the price is determined by the terms $f_i\big(v_i(t)\big)\varepsilon_i(t)$ and $g_i\big(v_j(t)\big)\varepsilon_j(t)$. To facilitate the calculation, we merge the trade signs into unsigned volumes and the unsigned impact functions by the following way,

$$\begin{aligned} f_i\big(v_i(t)\big)\varepsilon_i(t) &\longrightarrow \tilde{f}_i\big(\nu_i(t)\big) \ , \\ g_i\big(v_j(t)\big)\varepsilon_j(t) &\longrightarrow \tilde{g}_i\big(\nu_j(t)\big) \ , \end{aligned} \tag{6.2}$$

where $\tilde{f}_i\big(\nu_i(t)\big)$ and $\tilde{g}_i\big(\nu_j(t)\big)$ are signed impact functions of signed volumes $\nu_i(t)$ and $\nu_j(t)$. Thus, when selling out the volume $\nu_i(t)$ with $\nu_i(t) > 0$, the price either changes $\tilde{f}_i\big(-\nu_i(t)\big)$ or changes $-\tilde{f}_i\big(\nu_i(t)\big)$, *i.e.*, the negative price change of buying in the same volume. This also meets the case of $\nu_i(t) < 0$. As a result, we have

$$\tilde{f}_i\big(-\nu_i(t)\big) = -\tilde{f}_i\big(\nu_i(t)\big) \ . \tag{6.3}$$

Analogously,

$$\tilde{g}_i\big(-\nu_j(t)\big) = -\tilde{g}_i\big(\nu_j(t)\big) \ . \tag{6.4}$$

With the substitution of Eq. (6.2), the price impact model (6.1) is revised as

$$\begin{aligned} \log m_i(t) &= \sum_{t'<t} \Big[G_{ii}(t-t')\tilde{f}_i\big(\delta\nu_i(t')\big) + \eta_{ii}(t')\Big] \\ &+ \sum_{t'<t} \Big[G_{ij}(t-t')\tilde{g}_i\big(\delta\nu_j(t')\big) + \eta_{ij}(t')\Big] \\ &+ \log m_i(-\infty) \ , \end{aligned} \tag{6.5}$$

where $\delta\nu_i(t')$ and $\delta\nu_j(t')$ are the signed traded volumes in each time interval $\delta t'$. In unit time interval $\delta t''$, the signed traded volumes are the rates of trading,

$$\begin{aligned} \dot{\nu}_i(t'') &= \frac{\delta\nu_i(t'')}{\delta t''} = \frac{\delta\nu_i(t')}{\delta t'} \ , \\ \dot{\nu}_j(t'') &= \frac{\delta\nu_j(t'')}{\delta t''} = \frac{\delta\nu_j(t')}{\delta t'} \ . \end{aligned} \tag{6.6}$$

The positive rates of trading are for buy market orders, while the negative rates for sell market orders. Considering the limit case that $\delta t' \to \delta t''$, we transform the discrete time process of the price into a continuous process,

$$\begin{aligned}\log m_i(t) &= \int_{-\infty}^{t} G_{ii}(t-t')\tilde{f}_i\big(\dot{\nu}_i(t')\big)dt' + \int_{-\infty}^{t} \eta_{ii}(t')dt' \\ &+ \int_{-\infty}^{t} G_{ij}(t-t')\tilde{g}_i\big(\dot{\nu}_j(t')\big)dt' + \int_{-\infty}^{t} \eta_{ij}(t')dt' \\ &+ \log m_i(-\infty) \ . \end{aligned} \tag{6.7}$$

The continuous process of price is on a physical time scale, rather than a trade or an event time scale that considers a trade or an event as a time stamp. For linear impact functions of traded volumes, Eq. (6.7) is hold. For nonlinear ones, however, Eq. (6.7) will return an approximate value for the price.

6.2.2 Costs of trading

A trading strategy $\Pi_i = \{\nu_i(t)\}$ is referred to a round-trip trade [65] of stock i if the total bought-in volume is the same as the total sold-out volume in a trading period $\mathbb{T}_i$, which can be expressed as

$$\int_0^{\mathbb{T}_i} \dot{\nu}_i(t)dt = 0 \tag{6.8}$$

with the rate of trading $\dot{\nu}_i(t)$. During a trading period $\mathbb{T}_i$, the cost of trading is the expected cost of a sequence of small trades,

$$\Omega_i(\Pi_i) = E\Big[\int_0^{\mathbb{T}_i} \dot{\nu}_i(t)\big(\log m_i(t) - \log m_i(0)\big)dt\Big] \ . \tag{6.9}$$

Here, we use the difference of the logarithmic midpoint prices, *i.e.* the price return, to represent the price change. Due to the trades randomly initiated by buy and sell market orders, the trade price fluctuates between the best ask and the best bid. In contrast, the midpoint price between the best ask and best bid is better to indicate the price trend, as it is raised by a buy market order and lowered by a sell market order. To avoid the dramatic price shifting, the small trades in a round trip are restricted to exchange the volume that is less than the average. The cost of trading in Eq. (6.9) can be separated into the cost induced by the self-impact

$$\Omega_{ii}(\Pi_i) = \int_0^{\mathbb{T}_i} \dot{\nu}_i(t)dt \int_0^{t} G_{ii}(t-t')\tilde{f}_i\big(\dot{\nu}_i(t')\big)dt' \ , \tag{6.10}$$

and the cost induced by the cross-impact

$$\Omega_{ij}(\Pi_i) = \int_0^{\mathbb{T}_i} \dot{\nu}_i(t)dt \int_0^{t} G_{ij}(t-t')\tilde{g}_i\big(\dot{\nu}_j(t')\big)dt' \ . \tag{6.11}$$

Since the price impacts of all the buy and sell trades in a round trip are averaged to greatly lower the effects of random variables, the costs induced by $\eta_{ii}(t')$ and $\eta_{ij}(t')$ are ignored in above two equations. The total cost of stock i is the sum of the two parts,

$$\Omega_i(\Pi_i) = \Omega_{ii}(\Pi_i) + \Omega_{ij}(\Pi_i) \ . \tag{6.12}$$

For the paired stock j with a trading period $\mathbb{T}_j$, the cost of trading is analogously given by

$$\Omega_j(\Pi_j) = \Omega_{jj}(\Pi_j) + \Omega_{ji}(\Pi_j) \ . \tag{6.13}$$

The two round-trip trades of the stocks i and j start with a time difference Δt, where to keep the model as simple as possible, we only consider the limit cases, $\Delta t = 0$ and $\Delta t \to \infty$. The former means there is an overlap in the execution periods of the two round-trip trades, leading to the self-impact accompanied with the nontrivial cross-impact. The latter implies the two round-trip trades are executed individually without any overlap in time so that only the self-impact is present in each round-trip trade. If the costs arising from the self-impacts in above two cases are the same, the case with $\Delta t = 0$ has an extra cost induced by the cross-impacts,

$$\Omega_c(\Pi_i, \Pi_j) = \Omega_{ij}(\Pi_i) + \Omega_{ji}(\Pi_j) \ . \tag{6.14}$$

Therefore, the cross-impact cost $\Omega_c(\Pi_i, \Pi_j)$ determines the optimal execution of the two round-trip trades. If $\Omega_c(\Pi_i, \Pi_j) \geqslant 0$, executing the two round-trip trades individually with $\Delta t \to \infty$ is preferred to eliminate the extra cost from the cross-impacts. If $\Omega_c(\Pi_i, \Pi_j) < 0$, executing the two round-trip trades with $\Delta t = 0$ contributes to reduce the trading costs or even to profit from the possible opportunities of arbitrage. The cross-impact cost $\Omega_c(\Pi_i, \Pi_j)$ are detailed as follows.

For stock i, during the trading period $\mathbb{T}_i$, the volumes are bought in within the first θ_i period by a rate $\dot{v}_i^{(\mathrm{in})}(t)$, and then are sold out totally in the remaining time $(1-\theta_i)\mathbb{T}_i$ by a rate $-\dot{v}_i^{(\mathrm{out})}(t)$, where θ_i is a scaling factor of the bought-in time during the trading period. Analogously for stock j with all the quantities indexed by j instead of i. Here, the rates $\dot{v}_i^{(\mathrm{in})}(t)$, $\dot{v}_j^{(\mathrm{in})}(t)$, $\dot{v}_i^{(\mathrm{out})}(t)$ and $\dot{v}_j^{(\mathrm{out})}(t)$ are always positive values. To trace the rates of trading in different time regions, the constant rates are denoted as a function of time t or t' in the following integrals. Furthermore, to reduce the complexity of the integrals, the trading strategies are distributed in the following time regions.

(I) $0 \leqslant \theta_j \mathbb{T}_j \leqslant \theta_i \mathbb{T}_i \leqslant \mathbb{T}_j$

The transformation from buying in to selling out the stock i occurs during the period of stock j being sold out. Thus, the cross-impact costs in Eq. (6.14) can be expanded as

$$\begin{aligned}
\Omega_{ij}(\Pi_i) \ = \ & \int_0^{\theta_i \mathbb{T}_i} \dot{v}_i^{(\mathrm{in})}(t) dt \int_0^{\theta_j \mathbb{T}_j} G_{ij}(t-t') \tilde{g}_i\big(\dot{v}_j^{(\mathrm{in})}(t')\big) dt' \\
+ \ & \int_0^{\theta_i \mathbb{T}_i} \dot{v}_i^{(\mathrm{in})}(t) dt \int_{\theta_j \mathbb{T}_j}^{t} G_{ij}(t-t') \tilde{g}_i\big(-\dot{v}_j^{(\mathrm{out})}(t')\big) dt' \\
+ \ & \int_{\theta_i \mathbb{T}_i}^{\mathbb{T}_i} \big(-\dot{v}_i^{(\mathrm{out})}(t)\big) dt \int_0^{\theta_j \mathbb{T}_j} G_{ij}(t-t') \tilde{g}_i\big(\dot{v}_j^{(\mathrm{in})}(t')\big) dt' \\
+ \ & \int_{\theta_i \mathbb{T}_i}^{\mathbb{T}_i} \big(-\dot{v}_i^{(\mathrm{out})}(t)\big) dt \int_{\theta_j \mathbb{T}_j}^{t} G_{ij}(t-t') \tilde{g}_i\big(-\dot{v}_j^{(\mathrm{out})}(t')\big) dt' \ ,
\end{aligned} \tag{6.15}$$

$$\begin{aligned}
\Omega_{ji}(\Pi_j) &= \int_0^{\theta_j\mathbb{T}_j} \dot{v}_j^{(\mathrm{in})}(t)dt \int_0^t G_{ji}(t-t')\tilde{g}_j\big(\dot{v}_i^{(\mathrm{in})}(t')\big)dt' \\
&+ \int_{\theta_j\mathbb{T}_j}^{\mathbb{T}_j} \big(-\dot{v}_j^{(\mathrm{out})}(t)\big)dt \int_0^{\theta_i\mathbb{T}_i} G_{ji}(t-t')\tilde{g}_j\big(\dot{v}_i^{(\mathrm{in})}(t')\big)dt' \\
&+ \int_{\theta_j\mathbb{T}_j}^{\mathbb{T}_j} \big(-\dot{v}_j^{(\mathrm{out})}(t)\big)dt \int_{\theta_i\mathbb{T}_i}^t G_{ji}(t-t')\tilde{g}_j\big(-\dot{v}_i^{(\mathrm{out})}(t')\big)dt' \ . \qquad (6.16)
\end{aligned}$$

Here, due to the lag effect of cross-impacts, it is possible that the upper limit of integrals of selling out the volume of one stock goes beyond the trading period of this stock. However, the lag effect after finishing the process of buying in the volume is quickly covered by the effect of selling out the volume, it will not influence on the upper limit of integrals of buying in the volume. The cases in other time regions also have the similar treatment for the intergrals.

(II) $0 \leqslant \theta_j\mathbb{T}_j \leqslant \mathbb{T}_j \leqslant \theta_i\mathbb{T}_i$

Before emptying all the bought-in volumes of stock i, a round-trip trade of stock j has been fully executed. The cost $\Omega_{ij}(\Pi_i)$ has the same expression as Eq. (6.15). The cost $\Omega_{ji}(\Pi_j)$ is given by

$$\begin{aligned}
\Omega_{ji}(\Pi_j) &= \int_0^{\theta_j\mathbb{T}_j} \dot{v}_j^{(\mathrm{in})}(t)dt \int_0^t G_{ji}(t-t')\tilde{g}_j\big(\dot{v}_i^{(\mathrm{in})}(t')\big)dt' \\
&+ \int_{\theta_j\mathbb{T}_j}^{\mathbb{T}_j} \big(-\dot{v}_j^{(\mathrm{out})}(t)\big)dt \int_0^t G_{ji}(t-t')\tilde{g}_j\big(\dot{v}_i^{(\mathrm{in})}(t')\big)dt' \ . \qquad (6.17)
\end{aligned}$$

(III) $0 \leqslant \theta_i\mathbb{T}_i \leqslant \theta_j\mathbb{T}_j \leqslant \mathbb{T}_i$

The transformation from buying in to selling out the stock j occurs during the period of emptying all bought-in volumes of stock i. Thus, we have the cross-impact costs

$$\begin{aligned}
\Omega_{ij}(\Pi_i) &= \int_0^{\theta_i\mathbb{T}_i} \dot{v}_i^{(\mathrm{in})}(t)dt \int_0^t G_{ij}(t-t')\tilde{g}_i\big(\dot{v}_j^{(\mathrm{in})}(t')\big)dt' \\
&+ \int_{\theta_i\mathbb{T}_i}^{\mathbb{T}_i} \big(-\dot{v}_i^{(\mathrm{out})}(t)\big)dt \int_0^{\theta_j\mathbb{T}_j} G_{ij}(t-t')\tilde{g}_i\big(\dot{v}_j^{(\mathrm{in})}(t')\big)dt' \\
&+ \int_{\theta_i\mathbb{T}_i}^{\mathbb{T}_i} \big(-\dot{v}_i^{(\mathrm{out})}(t)\big)dt \int_{\theta_j\mathbb{T}_j}^t G_{ij}(t-t')\tilde{g}_i\big(-\dot{v}_j^{(\mathrm{out})}(t')\big)dt' \ , \qquad (6.18)
\end{aligned}$$

$$\begin{aligned}
\Omega_{ji}(\Pi_j) &= \int_0^{\theta_j\mathbb{T}_j} \dot{v}_j^{(\mathrm{in})}(t)dt \int_0^{\theta_i\mathbb{T}_i} G_{ji}(t-t')\tilde{g}_j\big(\dot{v}_i^{(\mathrm{in})}(t')\big)dt' \\
&+ \int_0^{\theta_j\mathbb{T}_j} \dot{v}_j^{(\mathrm{in})}(t)dt \int_{\theta_i\mathbb{T}_i}^t G_{ji}(t-t')\tilde{g}_j\big(-\dot{v}_i^{(\mathrm{out})}(t')\big)dt' \\
&+ \int_{\theta_j\mathbb{T}_j}^{\mathbb{T}_j} \big(-\dot{v}_j^{(\mathrm{out})}(t)\big)dt \int_0^{\theta_i\mathbb{T}_i} G_{ji}(t-t')\tilde{g}_j\big(\dot{v}_i^{(\mathrm{in})}(t')\big)dt' \\
&+ \int_{\theta_j\mathbb{T}_j}^{\mathbb{T}_j} \big(-\dot{v}_j^{(\mathrm{out})}(t)\big)dt \int_{\theta_i\mathbb{T}_i}^t G_{ji}(t-t')\tilde{g}_j\big(-\dot{v}_i^{(\mathrm{out})}(t')\big)dt' \ . \qquad (6.19)
\end{aligned}$$

(IV) $0 \leqslant \theta_i \mathbb{T}_i \leqslant \mathbb{T}_i \leqslant \theta_j \mathbb{T}_j$

Before all the bought-in volumes of stock j being sold out, the execution of the round-trip trade of stock i has finished. Thus, the cross-impact cost $\Omega_{ij}(\Pi_i)$ is

$$\begin{aligned}\Omega_{ij}(\Pi_i) &= \int_0^{\theta_i\mathbb{T}_i} \dot{v}_i^{(\mathrm{in})}(t)dt \int_0^t G_{ij}(t-t')\tilde{g}_i\big(\dot{v}_j^{(\mathrm{in})}(t')\big)dt' \\ &+ \int_{\theta_i\mathbb{T}_i}^{\mathbb{T}_i} \big(-\dot{v}_i^{(\mathrm{out})}(t)\big)dt \int_0^t G_{ij}(t-t')\tilde{g}_i\big(\dot{v}_j^{(\mathrm{in})}(t')\big)dt' \ . \end{aligned} \tag{6.20}$$

The expression of the cost $\Omega_{ji}(\Pi_j)$ is the same as Eq. (6.19).

6.2.3 A construction of trading strategies

A round-trip trade ends up when the net volume is zero, leading to

$$\dot{v}_i^{(\mathrm{in})}\theta_i\mathbb{T}_i - \dot{v}_i^{(\mathrm{out})}(1-\theta_i)\mathbb{T}_i = 0 \ , \tag{6.21}$$

$$\dot{v}_j^{(\mathrm{in})}\theta_j\mathbb{T}_j - \dot{v}_j^{(\mathrm{out})}(1-\theta_j)\mathbb{T}_j = 0 \ . \tag{6.22}$$

Setting $\dot{v}_i$ and $\dot{v}_j$ as the sums of bought-in rates and sold-out rates for stocks i and j, respectively,

$$\dot{v}_i = \dot{v}_i^{(\mathrm{in})} + \dot{v}_i^{(\mathrm{out})} \ , \tag{6.23}$$

$$\dot{v}_j = \dot{v}_j^{(\mathrm{in})} + \dot{v}_j^{(\mathrm{out})} \ , \tag{6.24}$$

the bought-in and sold-out rates can be denoted as

$$\dot{v}_i^{(\mathrm{in})} = \kappa_i\dot{v}_i \qquad \text{and} \qquad \dot{v}_i^{(\mathrm{out})} = (1-\kappa_i)\dot{v}_i \tag{6.25}$$

for stock i, and

$$\dot{v}_j^{(\mathrm{in})} = \kappa_j\dot{v}_j \qquad \text{and} \qquad \dot{v}_j^{(\mathrm{out})} = (1-\kappa_j)\dot{v}_j \tag{6.26}$$

for stock j, where the scaling factors of the bought-in rates κ_i and κ_j are bound to

$$0 < \kappa_i < 1 \qquad \text{and} \qquad 0 < \kappa_j < 1 \ . \tag{6.27}$$

According to Eqs. (6.21)—(6.26), the scaling factors of bought-in time θ_i and θ_j can be replaced by

$$\theta_i = 1-\kappa_i \qquad \text{and} \qquad \theta_j = 1-\kappa_j \ . \tag{6.28}$$

To connecting the stock i with the stock j, we introduce $\zeta_{\mathbb{T}}$, which links the trading periods of two stocks,

$$\zeta_{\mathbb{T}} = \frac{\mathbb{T}_i}{\mathbb{T}_j} \ , \tag{6.29}$$

and ζ_v, which combines the total bought-in (or sold-out) volumes v_i and v_j of two stocks,

$$\zeta_v = \frac{v_i}{v_j} \ . \tag{6.30}$$

By making use of the definition (6.6), the sums of bought-in and sold-out rates of the two stocks have the ratio,

$$\frac{\dot{v}_i}{\dot{v}_j} = \frac{\zeta_v}{\zeta_{\mathbb{T}}}\frac{(1-\kappa_j)\kappa_j}{(1-\kappa_i)\kappa_i} \ . \tag{6.31}$$

Therefore, to execute two round-trip trades of stocks i and j, we need to preset the ratio ζ_v of the total bought-in volumes, the bought-in rate $\dot{v}_i^{(\mathrm{in})}$ and the trading period $\mathbb{T}_i$ according to the practical demand. With three free parameters κ_i, κ_j and $\zeta_{\mathbb{T}}$, we then can work out the remaining quantities by Eqs. (6.23)—(6.31), including the sold-out rate $\dot{v}_i^{(\mathrm{out})}$ and $\dot{v}_j^{(\mathrm{out})}$, the bought-in rate $\dot{v}_j^{(\mathrm{in})}$, the trading period $\mathbb{T}_j$, and the time for buying in and selling out each stock. As a result, a trading strategy is determined by the set of $\{\kappa_i, \kappa_j, \zeta_{\mathbb{T}}\}$, where the optimal trading strategy is conditioned on the minimal cost of cross-impacts

$$\Omega_c(\Pi_i, \Pi_j) = \min\big\{\Omega_c\big(\kappa_i, \kappa_j, \zeta_{\mathbb{T}}\big)\big\} \ . \tag{6.32}$$

6.3 Applications to a specific case

The cost functions in Eqs. (6.15)—(6.20) contain the impact functions of time lag and of traded volumes. However, these impact functions have not been determined yet. Although Ref. [158] gives the functional form for them, the parameters in the functions depend on the specific stocks. To result in a feasible trading strategy in terms of the cross-impact cost for a specific case, it is necessary to measure the price impacts. Therefore, in Sec. 6.3.1, we introduce the data set used for the empirical measurement. In Sec. 6.3.2, we describe the algorithm for classifying the trade signs, which is crucial for measuring the price impacts. In Sec. 6.3.3, we quantify the impacts of traded volumes, fitted by a power law. In Sec. 6.3.4, with the help of the cross-response functions and the self-correlators of trade signs, we measure the cross-impacts of time lag between two stocks. In Sec. 6.3.5, using the fitted and preset parameters, we carry out and discuss the trading strategies with respect to the cross-impact costs.

6.3.1 Data sets

We apply our trading strategy to a specific pair of stocks, Apple Inc. (AAPL) and Microsoft Corp. (MSFT), where AAPL is indexed by i and MSFT is indexed by j in the following. We use the Trades and Quotes (TAQ) data set, where the data of two stocks comes from the NASDAQ stock market in 2008. For a given stock in each year, the TAQ data set contains a trade file recording all the information of each trade and a quote file recording all the information of each quote. The information of trades and quotes has the resolution of one second. However, more than one trade or quote may be found in the TAQ data set on the time scale smaller than one second. In addition, we only consider the trading days that AAPL and MSFT all have trades so as to have the cross-impacts during the intraday trading time. To avoid the dramatic fluctuation of prices due to any artifact at the opening and closing of the market, we exclude the data in the first and the last ten minutes of trading.

6.3.2 Trade signs

The trade sign plays a crucial role in measuring the price impacts from empirical data. Since the TAQ data set lacks of the information about the trade type (buy or sell) or the trade sign, a method to identify the trade signs is required. One representative algorithm put forward by Lee and Ready [97] is to compare the trade price with the preceding midpoint price. However, it is difficult to employ this algorithm to identify the signs for the trades during the one-second interval, because we cannot match those trades with their preceding midpoint prices without a higher resolution of the TAQ data set. In view of this, we resort to the algorithm described in our previous study [161]. The sign $\varepsilon(t;n)$ of n-th trade in the time interval t results from the sign of price change if the prices $S(t;n)$ and $S(t;n-1)$ for two consecutive trades are different, or otherwise from the preceding trade sign,

$$\varepsilon(t;n) = \begin{cases} \operatorname{sgn}\big(S(t;n) - S(t;n-1)\big) & , \quad \text{if} \quad S(t;n) \neq S(t;n-1) \ , \\ \varepsilon(t;n-1) & , \quad \text{otherwise} \ . \end{cases} \tag{6.33}$$

It is worth to mention that the trade price $S(t;n)$, found directly from the trade file of the TAQ data set, differs from the midpoint price $m(t)$, which is obtained from the last quote prior to the time interval t in the quote file. Moreover, the trade sign $\varepsilon(t)$ for the time interval of one second is defined as

$$\varepsilon(t) = \begin{cases} \operatorname{sgn}\left(\sum\limits_{n=1}^{N(t)} \varepsilon(t;n)\right) & , \quad \text{if} \quad N(t) > 0 \ , \\ 0 & , \quad \text{if} \quad N(t) = 0 \ . \end{cases} \tag{6.34}$$

That is a sign function of the sum of the trade signs $\varepsilon(t;n)$ in time interval t if there were trades in this interval. Otherwise, the absence of trading in t leads $\varepsilon(t)$ to be zero. Same as the $\varepsilon(t)$ in Eq. (6.1), the sign $\varepsilon(t)$ here indicates the trade type of market orders. $\varepsilon(t) = +1$ ($\varepsilon(t) = -1$) means a majority of buy (sell) market orders in time interval t, and $\varepsilon(t) = 0$ means a lack of trading or a balance of buy and sell market orders in this interval. The tests of this algorithm using the TotalView-ITCH data set, carried out in Ref. [161], reveals the average accuracy of 85% for Eq. (6.33) and of 82% for Eq. (6.34) to identify the trade signs.

6.3.3 Measurement for impacts of traded volumes

The traded volume in this study refers to the aggregation of all the traded volumes in the time interval t. To put different stocks in the same footing, the traded volumes of each stock are normalized by dividing the average of traded volumes over a whole year,

$$v(t) = \frac{T \sum_{n=1}^{N(t)} v(t;n)}{\sum_{t=1}^{T} \sum_{n=1}^{N(t)} v(t;n)} \ , \tag{6.35}$$

where $v(t;n)$ is the volume of the n-th trade in the time interval t, $N(t)$ is the number of trades in t, and T is the total time intervals for trading during a whole year. Thus, $v(t) < 1$ indicates that the traded volumes are smaller than their average. Conditioned on the unsigned volumes $v_j(t)$, the price change of stock i, on average, due to the trades of stock j, *i.e.*, the price cross-response, is given [135, 158] by

$$R_{ij}(v_j, \tau) = \Big\langle r_i(t,\tau)\varepsilon_j(t) \Big| v_j(t) \Big\rangle_t \ , \tag{6.36}$$

where $\langle \cdots \rangle_t$ means the average over all the time t, and the price change $r_i(t,\tau)$ at time t with a time lag τ is defined as the difference of logarithmic midpoint prices,

$$r_i(t,\tau) = \log m_i(t+\tau) - \log m_i(t) = \log \frac{m_i(t+\tau)}{m_i(t)} \ . \tag{6.37}$$

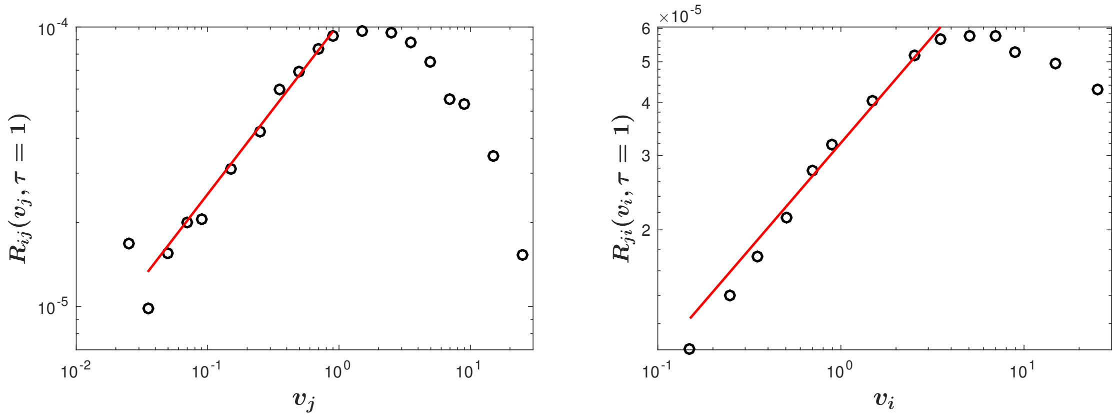


Figure 6.1: Empirical (circle) and fitted (line) results of impact functions of traded volumes. Stock i is AAPL, and stock j is MSFT.

Since the influence of traded volumes is independent of the time lag, Eq. (6.36) can be approximately decomposed into

$$R_{ij}(v_j, \tau) \approx R_{ij}(\tau) g_i(v_j) \ , \tag{6.38}$$

where

$$R_{ij}(\tau) = \Big\langle r_i(t,\tau)\varepsilon_j(t) \Big\rangle_t \tag{6.39}$$

is the price cross-response depending on the time lag, and $g_i(v_j)$ is the impact function of traded volumes. For the average price change of stock j induced by stock i, analogously we have,

$$R_{ji}(v_i, \tau) \approx R_{ji}(\tau) g_j(v_i) \ . \tag{6.40}$$

Using the empirical data of AAPL and MSFT, we carry out the dependence of price changes on the traded volumes with $\tau = 1$, as shown in Fig. 6.1. Coinciding with Ref. [158], the dependencies for small traded volumes are fitted well by a power law,

$$g_i(v_j) = v_j^{\delta_{ij}} \qquad \text{and} \qquad g_j(v_i) = v_i^{\delta_{ji}} \ , \tag{6.41}$$

where the parameters δ_{ij} and δ_{ji} for AAPL and MSFT, respectively, are listed in Table 6.1. To make the trading strategy feasible, we limit the volume of each trade in strategies to be smaller than the average.

We notice that the traded volumes and the impact functions in Eq. (6.41) are all unsigned. With the positive rates of trading to buy a stock, the signed impact functions of traded volumes in Eqs. (6.15)—(6.20) are the same as the unsigned ones, as shown in Eq. (6.41). With the negative rates of trading to sell a stock, according to Eq. (6.4), the signed impact functions turn to unsigned ones by the following way,

$$\tilde{g}_i\big(-\dot{v}_j^{(\mathrm{out})}(t')\big) = -\tilde{g}_i\big(\dot{v}_j^{(\mathrm{out})}(t')\big) = -g_i\big(\dot{v}_j^{(\mathrm{out})}(t')\big) \ , \tag{6.42}$$

$$\tilde{g}_j\big(-\dot{v}_i^{(\mathrm{out})}(t')\big) = -\tilde{g}_j\big(\dot{v}_i^{(\mathrm{out})}(t')\big) = -g_j\big(\dot{v}_i^{(\mathrm{out})}(t')\big) \ . \tag{6.43}$$

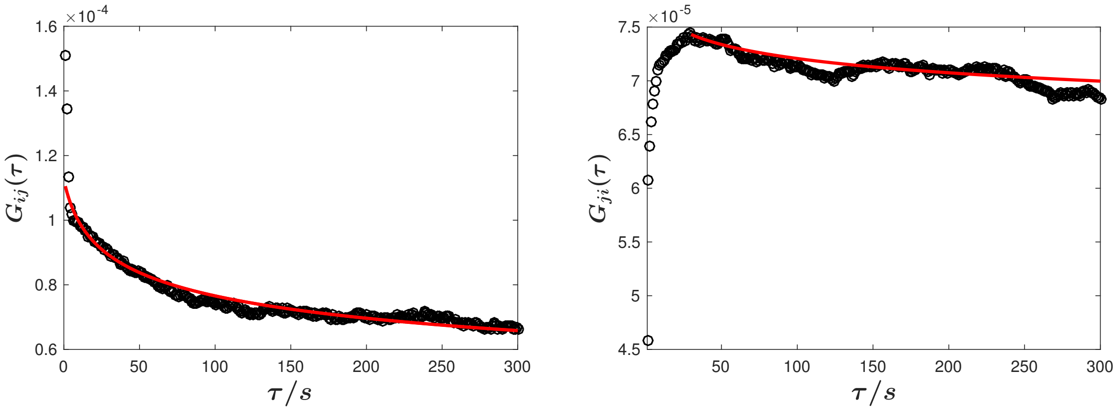


Figure 6.2: Empirical (circle) and fitted (line) results of impact functions of time lag. Stock i is AAPL, and stock j is MSFT.

6.3.4 Measurement for cross-impacts of time lag

The price cross-response comprises two components [158]. One arises from the self-impacts and is related to the cross-correlators of trade signs. The other one results from the cross-impacts and is related to the self-correlators of trade signs. Here, we focus on the response component containing the cross-impacts, which is given by

$$\begin{aligned} R_{ij}^{(S)}(\tau) \;=\; & \sum_{0\leq t<\tau} G_{ij}(\tau-t)\left\langle g_i\big(v_j(t)\big)\right\rangle_t \Theta_{jj}(t) \\ & + \sum_{t<0}\Big[G_{ij}(\tau-t)-G_{ij}(-t)\Big]\left\langle g_i\big(v_j(t)\big)\right\rangle_t \Theta_{jj}(-t) \ . \end{aligned} \tag{6.44}$$

The superscript (S) in the response function indicates the response component related to the the self-correlator $\Theta_{jj}(\tau)$, defined as

$$\Theta_{jj}(\tau) = \Big\langle \varepsilon_j(t+\tau)\varepsilon_j(t) \Big\rangle_t \ . \tag{6.45}$$

In Eq. (6.44), by replacing $\tau - t$ or $-t$ with τ' in each cross-impact function of time lag,

$$\begin{aligned} R_{ij}^{(S)}(\tau) \;=\; & \sum_{0<\tau'\leq\tau} G_{ij}(\tau')\left\langle g_i\big(v_j(t)\big)\right\rangle_t \Theta_{jj}(\tau-\tau') \\ & + \sum_{\tau'>\tau} G_{ij}(\tau')\left\langle g_i\big(v_j(t)\big)\right\rangle_t \Theta_{jj}(-\tau+\tau') \\ & - \sum_{\tau'>0} G_{ij}(\tau')\left\langle g_i\big(v_j(t)\big)\right\rangle_t \Theta_{jj}(\tau') \ , \end{aligned} \tag{6.46}$$

Table 6.1: Parameters for impact functions

δ_{ij}	δ_{ji}	$\langle g_i(v_j(t))\rangle_t$	$\langle g_j(v_i(t))\rangle_t$	$\Gamma_{0,ij}$	$\Gamma_{0,ji}$	$\tau_{0,ij}$	$\tau_{0,ji}$	β_{ij}	β_{ji}
0.61	0.50	0.40	0.60	1.13×10^{-4}	0.79×10^{-4}	7.34	4.75	0.14	0.03

and using the symmetric property of sign self-correlators $\Theta_{jj}(\tau)=\Theta_{jj}(-\tau)$, we have

$$\frac{R_{ij}^{(S)}(\tau)}{\big\langle g_i\big(v_j(t)\big)\big\rangle_t}=\sum_{\tau'=1}^{\infty}A_{jj}(\tau,\tau')G_{ij}(\tau') \ , \tag{6.47}$$

where

$$A_{jj}(\tau,\tau')=\Theta_{jj}(\tau-\tau')-\Theta_{jj}(\tau') \ . \tag{6.48}$$

The component $R_{ij}^{(S)}(\tau)$ is the cross-response $R_{ij}(\tau)$ weighted by a quantity w_i with $0<w_i<1$. Therefore, the cross-impact of time lag entering the impact matrix G_{ij} can be quantified from empirical data by

$$G_{ij}=\frac{w_i}{\big\langle g_i\big(v_j(t)\big)\big\rangle_t}A_{jj}^{-1}R_{ij} \ , \tag{6.49}$$

where A_{jj} is the matrix of sign correlators with the elements worked out by Eq. (6.48) and R_{ij} is the response matrix with the elements $R_{ij}(\tau)$, defined by Eq. (6.39). Analogously for stock j, we have

$$G_{ji}=\frac{w_j}{\big\langle g_j\big(v_i(t)\big)\big\rangle_t}A_{ii}^{-1}R_{ji} \ . \tag{6.50}$$

Although we can estimate the weight w_i and w_j by a complicated method, as introduced in Ref. [158], to facilitate the calculation, we assume $w_i=w_j$ and further normalize the cross-impact of time lag by w_i or w_j. By this way, it also normalizes the cost of trading according to Eqs. (6.15)—(6.20), but it does not change the sign of the cost, used to distinguish the profit from the cost.

Using Eqs. (6.49) and (6.50), we work out the empirical cross-impacts of time lag between AAPL and MSFT, shown in Fig. 6.2 with circles. To obtain the cross-impacts in the first 300 seconds, we replace the ∞ in Eq. (6.47) by a large cut-off of 3000 seconds. Due to the fluctuations of sign self-correlators and of cross-responses, the cross-impacts in small time lags are unstable. We thus extract the empirical results with stably decaying for parameter fits. To fit to empirical data, here, we employ simplified power-law functions instead of the complicated functional form in Ref. [158],

$$G_{ij}(\tau)=\frac{\Gamma_{0,ij}}{\Big(1+\frac{\tau}{\tau_{0,ij}}\Big)^{\beta_{ij}}} \qquad \text{and} \qquad G_{ji}(\tau)=\frac{\Gamma_{0,ji}}{\Big(1+\frac{\tau}{\tau_{0,ji}}\Big)^{\beta_{ji}}} \ , \tag{6.51}$$

where $\tau_{0,ij}$ and $\tau_{0,ji}$ are the time scales having the positive values, β_{ij} and β_{ji} are the rates of decaying, and $\Gamma_{0,ij}$ and $\Gamma_{0,ji}$ are the temporary impact components per share. The fitted values of these parameters are listed in Table 6.1.

6.3.5 Computations and discussions of trading strategies

To obtain the trading strategy, we consider to totally buy in the same volume for AAPL and MSFT, such that $\zeta_v=1$. Further, we set the trading period of AAPL as $\mathbb{T}_i=1$ unit of time. For one unit of time, we plan to buy in 0.1 times average traded volume of AAPL, resulting in $\dot{v}_i^{(\mathrm{in})}=0.1$. With these preset values and fitted parameters listed in Table 6.1, we carry out the trading strategies $\{\kappa_i,\kappa_j,\zeta_{\mathbb{T}}\}$ in four time regions using the cost function (6.14). The four time regions lead to the three free parameters in the strategies bound to the conditions,

$$\begin{cases} \frac{1-\kappa_j}{1-\kappa_i} \leqslant \zeta_{\mathbb{T}} \leqslant \frac{1}{1-\kappa_i} & \text{(region I)} \ , \\ \zeta_{\mathbb{T}} \geqslant \frac{1}{1-\kappa_i} & \text{(region II)} \ , \\ 1-\kappa_j \leqslant \zeta_{\mathbb{T}} \leqslant \frac{1-\kappa_j}{1-\kappa_i} & \text{(region III)} \ , \\ 0 < \zeta_{\mathbb{T}} \leqslant 1-\kappa_j & \text{(region IV)} \ . \end{cases} \tag{6.52}$$

Due to the boundary conditions, with a given ratio of trading periods $\zeta_{\mathbb{T}}$, one may not obtain the numerical solution for the cost function (6.14).

As examples, here, we consider the ratio of trading periods $\zeta_{\mathbb{T}} = 0.5$, 1, and 2, meaning the trading period of AAPL is the half of, the same as, and the twice of the trading period of MSFT, respectively. Leaving out the cases without numerical solutions, the cross-impact costs depending on the scaling factors of bought-in rates κ_i and κ_j are displayed in Fig. 6.3. In terms of the costs, the trading strategies can be classified as two types, one with the non-negative cost and the other one with the negative cost. For the non-negative cost, it is better to execute the two round-trip trades individually without any overlap in time, *i.e.* $\Delta t \to \infty$ so as to circumvent the extra cost. If the two round-trip trades inevitably start at the same time, by using the strategy $\{\kappa_i, \kappa_j, \zeta_{\mathbb{T}}\}$ with the minimal cross-impact costs, one can lower the total cost of trading. Taking $\zeta_{\mathbb{T}} = 2$ as an example, for the two round-trip trades starting at the same time, we find that the minimal positive cost in region III is at the position of the maximal κ_i and κ_j. It suggests quickly buying in AAPL and MSFT and then slowly selling out them can lower the total cost of trading to some extent. On the other hand, the presence of the negative cost is possible, especially at a small time scale when the market has not reached to an efficient state [158]. Such case can be seen when $\zeta_{\mathbb{T}} = 1$ in regions I and III. The negative cost of trading implies the possible opportunities of arbitrage or a reduction of the total cost for trading. In particular, by minimizing the cross-impact cost to obtain an optimal execution strategy $\{\kappa_i, \kappa_j, \zeta_{\mathbb{T}}\}$, one can maximize the possibility of arbitrage.

6.4 Conclusions

We extend the framework of trading strategies for single stocks [65] to a pair of stocks. For one stock, to lower the execution cost from price self-impacts, traders favour to submit a sequence of small trades. A round trip for buying in and selling out a sequence of small trades is termed a round-trip trade. By considering the executions of two round-trip trades from different stocks, we construct a trading strategy $\{\kappa_i, \kappa_j, \zeta_{\mathbb{T}}\}$, which can be described by the trading rates κ_i and κ_j of the paired stocks and the ratio of their trading periods $\zeta_{\mathbb{T}}$. By minimizing the cross-impact cost, one can obtain the optimal execution strategy for the two round-trip trades.

We apply our trading strategy to a pair of stocks, AAPL and MSFT. To determine the impact functions in the strategy, we measure the cross-impacts of time lag and of traded volumes using the empirical data. By numerical computation with the fitted parameters and the preset values, we picture the trading strategy in terms of the cross-impact cost. The positive cost suggests that the individual executions of two round-trip trades without

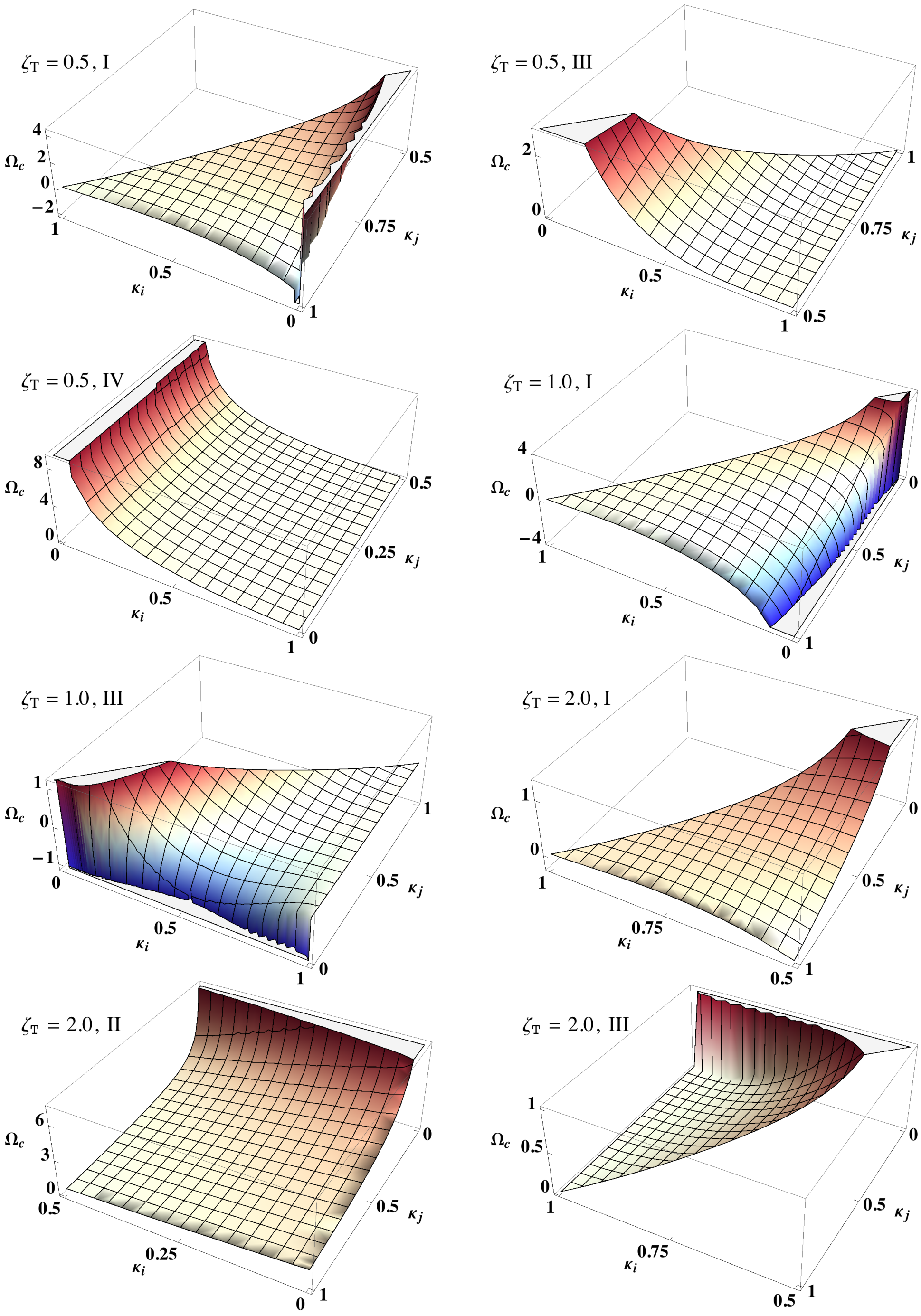


Figure 6.3: Trading strategies with respect to the costs Ω_c. The increasing of positive and negative costs is displayed by the colour from white to dark red and to dark blue, respectively. Zero cost is indicated by white. To view clearly, the directions and ranges of axes are adjusted for specific cases. The costs of trading Ω_c are all rescaled by multiplying 10^6.

any overlap in time can circumvent the extra cost. The negative cost implies that the two round-trip trades starting at the same time lead to the possible opportunities of arbitrage or a reduction of the total cost for trading. The different ways for order execution reveal the influence of cross-impacts on the optimal trading strategy. Certainly, an improved strategy with respect to the cross-impact cost is called for, but it is beyond this study.

Chapter 7

Local fluctuations of the signed traded volumes and the dependencies of demands: a copula analysis

7.1 Introduction

In the previous chapters, we found that the traded volume is crucial for price impacts and optimal execution of orders, where we considered the traded volume and the trade sign individually. In this chapter, we will introduce the trade sign into the traded volume, leading to the signed traded volume. In a certain time interval, the sum of all the signed traded volumes is the volume imbalance, which indicates the demands of stocks at that time interval. A positive volume imbalance implies that the demand for a stock is larger than the supply, whereas a negative volume imbalance implies an opposite case. As introduced in Chapter 1.3.6, conditioned on the local noise intensity, the distribution of demands exhibits a two-phase behavior [132], that is, a unimodal distribution transforms to a bimodal distribution with the increase of the local noise intensity. A lot of studies [85, 86, 101, 110, 132–134, 137, 146, 147, 167] have been devoted to the two-phase behavior, but only the statistical properties of individual stocks are taken into account in those studies. Here, we want to look at the statistical dependence of demands across different stocks, and figure out the influence of the local noise intensity on this dependence structure.

To this end, we employ copulas, which was first introduced by Sklar in 1959 [148, 149]. In copulas, all marginal distributions are mapped to uniform distributions. The joint distribution density is then measured as a function of the corresponding quantiles in the resulting uniform distributions. Consequently, the copula separates the pure statistical dependence of random variables from the marginal probability distributions. In contrast with the correlation coefficient, it is better to capture the nonlinear dependence of two random variables. Due to this advantage, the copula has become an important, standard tool for directly modelling and analyzing the statistical dependencies of different systems. An overview of applications of copulas in finance refers to Refs. [68, 125].

Recently, a $\mathcal{K}$ copula density function was proposed to describe the fat-tailed dependence [41, 165]. The $\mathcal{K}$ copula density is based on a multivariate distribution in terms of a modified Bessel function of the second kind. This distribution is a result of the Random Matrix Model for the non-stationarity of financial data [40, 140]. It turns out to account for the empirical multivariate distributions of returns with fat tails rather well [39, 40, 140, 141]. In this study, by comparing the $\mathcal{K}$ copula density with the Gaussian copula density, we will show that the $\mathcal{K}$ copula density provides a good description for the empirical dependence of demands. With the copula density, we also investigate the asymmetry of the tail dependencies of demands, and further demonstrate the influence of the large local noise intensity on the dependence structure.

The chapter is organized as follows. In Sec. 7.2, we introduce the data set and the methods of analysis, including the definition of the demands, the concept of the copula density, and the estimation method of the empirical copula density. In Sec. 7.3, we show the empirical results for the demand distributions of individual stocks, the empirical copula density between stocks and the tail asymmetries of the empirical copula density. In Sec. 7.4, we fit the empirical copula density by a bivariate $\mathcal{K}$ copula density function and a Gaussian copula density function, and compare the two fit results. In Sec. 7.5, we investigate the influences of local fluctuations on the dependence of demands and on the tail asymmetries of dependencies. We conclude and discuss our results in Sec. 7.6. This chapter refers to Ref. [159]. In the following, we use the original text from Ref. [159].

7.2 Data set and methods of analysis

We present our data set in Sec. 7.2.1, and give basic definitions of trade signs in Sec. 7.2.2. Although there are detailed presentations on copulas in the statistics literature [41, 87, 119, 165], we give a short sketch of the concept for the convenience of the reader in Sec. 7.2.3. In Sec. 7.2.4, we illustrate and discuss how the empirical copula densities are estimated.

7.2.1 Data set

The stocks are from NASDAQ stock market in the year 2008, where all successive transactions and quotes of those stocks are recorded in Trades and Quotes (TAQ) data set. To avoid overnight effects and the drastic fluctuations at the opening and closing of the market, we exclude the trades occurring in the first and the last 10 minutes of the trading time for each day. For a stock pair, only the common trading days are taken into account for calculating the copula densities, because the dependence between stocks is absent when either stock does not trade. In Sec. 7.2.2, to calculate the conditional probability density distributions, we use 496 available stocks from S&P 500 index in 2008. For the empirical copula densities to be evaluated in Secs. 7.3 and 7.5, we select the first 100 stocks, listed in Appendix A.4, with the largest average number of daily trades among that 496 stocks. The number of daily trades, also excluding the ones in the first and the last 10 minutes of the daily trading time, is averaged over a whole year for each stock.

7.2.2 Trade signs and demands

In a time interval labeled t, various trades with running number n may occur with corresponding prices $S(t;n)$. Each such trade in the TAQ data set can be classified as either buyer-initiated or seller-initiated [160, 161] by

$$\varepsilon(t;n) = \begin{cases} \mathrm{sgn}\big(S(t;n) - S(t;n-1)\big) & , \quad \text{if } S(t;n) \neq S(t;n-1), \\ \varepsilon(t;n-1) & , \quad \text{otherwise}, \end{cases} \tag{7.1}$$

where $\varepsilon(t;n)$ represents the sign of n-th trade in a time interval. A trade is identified as buyer-initiated if $\varepsilon(t;n) = 1$, and a seller-initiated if $\varepsilon(t;n) = -1$. Zero values for $\varepsilon(t;n)$ are absent, because we do not aggregate the trade signs in a physical time interval as in our previous studies [160, 161]. It is worth mentioning that due to the resolution of one second in the TAQ data set, the algorithm of Lee and Ready [97] cannot be used to classify the trades occurring in a time interval of one second. Instead, Eq. (7.1) is designed to classify continuous trades in smaller time scale than one second, too.

The demand can be quantified as the volume imbalance, *i.e.* the difference between all bought-in volumes and all sold-out volumes in a time interval t,

$$\nu(t) = \sum_{n=1}^{N_{\mathrm{trades}}(t)} v(t;n)\varepsilon(t;n) \ . \tag{7.2}$$

Here, $N_{\mathrm{trades}}(t)$ denotes the number of trades in time interval t, and $v(t;n)$ is the unsigned volume for n-th trade in t. To have, at the same time, many trades in the time intervals t and a long time series $\nu(t)$ in each trading day, we use time intervals t of one minute.

7.2.3 Copula densities

Let $F_{kl}(x_1, x_2)$ be a joint cumulative distribution of the random variables x_1 and x_2 with marginal cumulative distributions $F_k(x_1)$ and $F_l(x_2)$, respectively. According to Sklar's theorem[119], there exists a copula $\mathrm{Cop}_{kl}(q_1, q_2)$ for all quantiles $q_1, q_2 \in [0, 1]$ satisfying

$$F_{kl}(x_1, x_2) = \mathrm{Cop}_{kl}\big(F_k(x_1), F_l(x_2)\big) \ . \tag{7.3}$$

In terms of the probability density function $f_k(x_1)$ of the random variable x_1, the marginal cumulative distribution function $F_k(x_1)$ can be expressed as,

$$F_k(x_1) = \int_{-\infty}^{x_1} f_k(s) ds \ , \tag{7.4}$$

and analogously for $F_l(x_2)$. The inverse cumulative distribution function $F_k^{-1}(\cdot)$ is known as the quantile function. We thus have

$$q_1 = F_k(x_1) \quad \text{and} \quad x_1 = F_k^{-1}(q_1) \ , \tag{7.5}$$

and accordingly for $q_2 = F_l(x_2)$. Hence, using Eq. (7.3) , the copula can be expressed as the cumulative joint distribution of the quantiles,

$$\mathrm{Cop}_{kl}(q_1, q_2) = F_{kl}\big(F_k^{-1}(q_1), F_l^{-1}(q_2)\big) \ . \tag{7.6}$$

Thus, the dependence structure of random variables is separated from the marginal probability distributions. In other words, the pure dependence structure is measured independently of the particular marginal distribution. The copula density is given as the two-fold derivative

$$\mathrm{cop}_{kl}(q_1, q_2) = \frac{\partial^2}{\partial q_1 \partial q_2} \mathrm{Cop}_{kl}(q_1, q_2) \tag{7.7}$$

with respect to the quantiles.

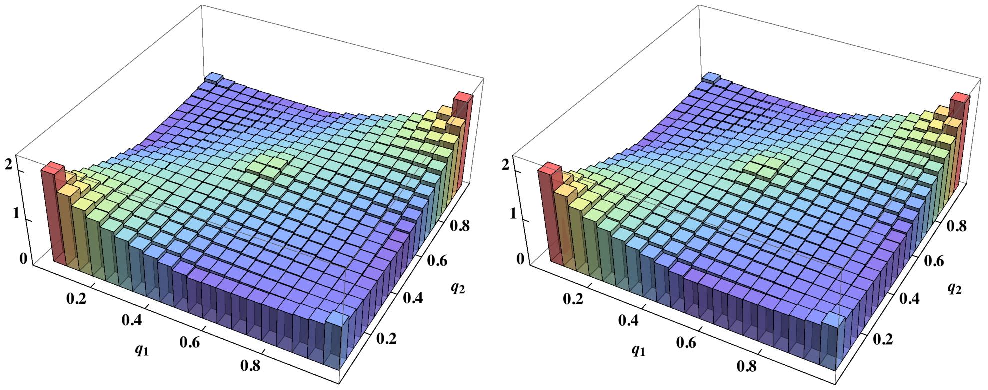


Figure 7.1: The empirical copula density $\mathrm{cop}(q_1, q_2)$ of volume imbalances averaged over 4950 stock pairs $(k,\, l)$. Left: the order of stocks is preset; right: the order of stocks is shuffled.

7.2.4 Empirical estimation of copula densities

To estimate the empirical pairwise copula densities of demands, we first map all observations of volume imbalances $\nu_k(t)$ from stock k to a uniformly distributed time series $q_1(t)$ by

$$q_1(t) = F_k(\nu_k(t)) = \frac{1}{T}\sum_{\tau=1}^{T}\Theta\big(\nu_k(t) - \nu_k(\tau)\big) - \frac{1}{2T} \ , \tag{7.8}$$

where $\Theta(\cdot)$ is the Heaviside step function, and T is the length of the time series. The volume imbalance $\nu_k(t)$ is defined in Eq. (7.2). To arrive at generic results, we average over all $L(L-1)/2$ stock pairs,

$$\mathrm{cop}(q_1, q_2) = \frac{2}{L(L-1)}\sum_{k=1}^{L-1}\sum_{l=k+1}^{L}\mathrm{cop}_{kl}(q_1, q_2) \ , \tag{7.9}$$

where $\mathrm{cop}_{kl}(q_1, q_2)$ is a histogram over two dimensions. The bin size of all these histograms is $\Delta q_1 = \Delta q_2 = 1/20$. Following Refs. [41, 165], we do not use a symmetrized definition of the averaged copula.

One might argue that the empirical copula densities should be averaged over $L(L-1)$ stock pairs by

$$\mathrm{cop}(q_1, q_2) = \frac{1}{L(L-1)}\sum_{k=1}^{L-1}\sum_{l=k+1}^{L}\Big(\mathrm{cop}_{kl}(q_1, q_2) + \mathrm{cop}_{lk}(q_1, q_2)\Big) \ , \tag{7.10}$$

which would make the averaged copula densities independent of the order of two stocks in a pair. To clarify the reasons for the choice of definition (7.9), we first point out that the order of stocks in a pair will not influence the averaged copula densities largely, as we consider the average of copula densities over a large amount of stock pairs. The purpose of averaging is to wash out the individual features of specific stock pairs and to reveal the generic ones.

When the number of stocks tends to the infinity, *i.e.* $L \to \infty$, the definitions (7.10) and (7.9) are equivalent. We calculate the empirical copula density of volume imbalances with the first definition (7.9), as shown in Fig. 7.1. Here, to facilitate the calculation, we

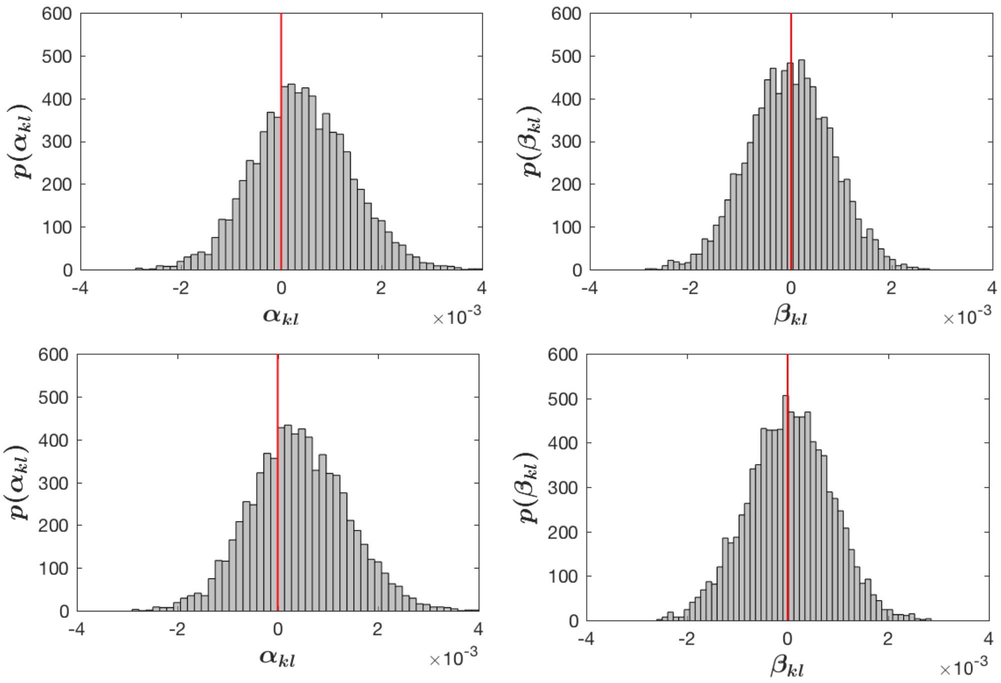


Figure 7.2: Histograms of asymmetry values of copula densities for positive dependence $p(\alpha_{kl})$ (left) and for negative dependence $p(\beta_{kl})$ (right) with 4950 stock pairs $(k,\ l)$. Top: the order of stocks is preset; bottom: the order of stocks is shuffled. All the histograms are normalized to one.

replace $L \to \infty$ by $L = 100$, and total 4950 stock pairs are used. To have a better view of the dependence structure, the tail asymmetries of the copula density are characterized by two quantities, α_{kl} and β_{kl},

$$\alpha_{kl} = \int\limits_{0.8}^{1} dq_1 \int\limits_{0.8}^{1} dq_2 \ \mathrm{cop}_{kl}(q_1, q_2) - \int\limits_{0}^{0.2} dq_1 \int\limits_{0}^{0.2} dq_2 \ \mathrm{cop}_{kl}(q_1, q_2) \ , \tag{7.11}$$

$$\beta_{kl} = \int\limits_{0}^{0.2} dq_1 \int\limits_{0.8}^{1} dq_2 \ \mathrm{cop}_{kl}(q_1, q_2) - \int\limits_{0.8}^{1} dq_1 \int\limits_{0}^{0.2} dq_2 \ \mathrm{cop}_{kl}(q_1, q_2) \ , \tag{7.12}$$

i.e., we look into the corners of size 0.2 times 0.2 in the (q_1, q_2) plane. Thus, α_{kl} describes the asymmetry of positive dependence. A shift away from zero in the histogram of α_{kl} can be seen in Fig. 7.2. However, the asymmetry of the negative dependence, indicated by β_{kl}, is more significant. An overall symmetric distribution around zero for β_{kl} can be found in Fig. 7.2. That implies the averaged $\mathrm{cop}_{kl}(q_1, q_2)$ over 4950 stock pairs $(k,\ l)$ is equivalent to the averaged $\mathrm{cop}_{lk}(q_1, q_2)$ over 4950 stock pairs $(l,\ k)$. Hence the two definitions (7.9) and (7.10) are equivalent for all practical purposes.

The difference of the two definitions (7.9) and (7.10) lies in whether or not the order of stocks influences the dependence structure of average copula density. We therefore shuffle the order of stocks and recalculate the copula density with definition (7.9). The results in Figs. 7.1 and 7.2 do not change too much compared to the original ones with preset order of stocks as listed in Appendix A.4.

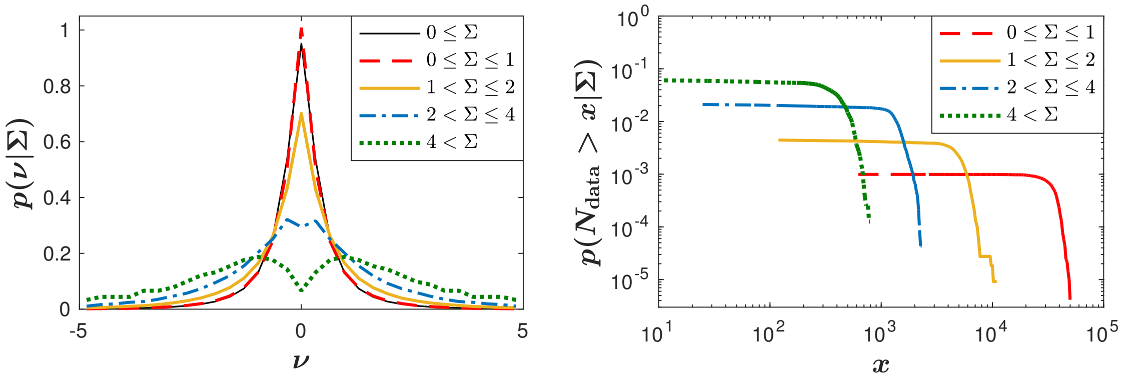


Figure 7.3: Left: the probability density distributions of volume imbalance conditioned on the local noise intensity $p(\nu|\Sigma)$, right: the cumulative probability density distributions of the numbers of data points conditioned on the local noise intensity $p(N_{\mathrm{data}} > x|\Sigma)$ versus variable x.

We thus employ the definition (7.9) to average the empirical copula densities rather than the definition (7.10).

7.3 Empirical results

In Sec. 7.3.1, We examine the effect of the local noise intensity on the marginal distribution of volume imbalances. In Sec. 7.3.2, we show the empirical copula density and discuss the asymmetry of tailed dependence of the copula.

7.3.1 Demand distributions

It is useful to introduce the local noise intensity [132],

$$\Sigma(t) = \left\langle \left| v(t;n)\varepsilon(t;n) - \left\langle v(t;n)\varepsilon(t;n) \right\rangle_n \right| \right\rangle_n \ , \tag{7.13}$$

which can be understand as the amount of fluctuations around the local average of volume imbalance in a time interval.

We investigate the two-phase behavior by examining the distribution of the volume imbalance conditioned on the local noise intensity, $p(\nu|\Sigma)$, as shown in Fig. 7.3. The distributions are found for altogether 496 stocks in S&P 500. To include different stocks on equal footing, we scale out the volatilities. For larger $\Sigma(t)$, the transition from a unimodal distribution to a bimodal distribution appears. Especially, when the $\Sigma(t) > 4$, the bimodal distribution is obvious. These large local noise intensity and bimodal distributions are exactly what we are interested in when looking at the copula density of demands between stocks conditioned on the local noise intensity.

7.3.2 Copula densities

From the empirical copula densities of volume imbalances in Fig. 7.1, strong dependencies of large demands between stocks can be inferred, positive as well as negative ones. The positive demands mean that the buyer-initiated trades dominate in the market. The negative demands correspond to supplies of volumes, *i.e.*, seller-initiated trades dominate.

Thus, either the large supplies or the large demands between stocks exhibit strong, positive dependencies. In contrast, the dependencies between large supply of one stock and large demand of another stock, *i.e.* the negative dependencies, also exist, but are not as pronounced.

As we have seen in section 7.3.2, the negative dependencies are almost symmetric for the average copula densities, but the positive dependencies are not. Once more, the asymmetry of the α_{kl} distribution in in Fig. 7.1 is important, as it implies a stronger dependence of demands than of supplies. To further quantify the asymmetry of distributions, we introduce the skewness, defined as

$$\text{skewness} = \frac{\langle (x-\mu)^3 \rangle}{\sigma^3} \ , \tag{7.14}$$

where μ is the mean of x, and σ is the standard deviation of x. Here, x stands for α_{kl} and β_{kl}, respectively. We thus measure the skewness of the distributions, listed in Table 7.1, where the one for α_{kl} is 0.0977. This suggests that from a large trade of one stock, it is more likely to find similar trades of other stocks, where the possibility of buy trades is higher than the possibility of sell trades. When the traded volumes are much larger than the market depth, these buy trades will push the prices up [133, 160, 161]. In financial markets, the persistent raising of prices of most stocks indicates a bull market. Consequently, the asymmetry of positive dependencies suggests the traders are more optimistic expecting a bull market.

7.4 Comparison of two models with the empirical copula density

We fit the empirical copula density with two functions, a bivariate $\mathcal{K}$ copula density function and a Gaussian copula density function. Since the two copula density functions are discussed in Refs. [41, 165], we only shortly introduce them in Secs. 7.4.1 and 7.4.2, respectively. We then compare them with the empirical results in Sec. 7.4.3.

7.4.1 Bivariate $\mathcal{K}$ copula density

A K component vector $\mathbf{r} = \big(r_1, ..., r_K\big)$ with elements r_k, $k = 1, ..., K$, normalized to zero mean and unit variance, follows a multivariate $\mathcal{K}$ distribution [40, 140], if its probability

Table 7.1: The skewness of distribution of asymmetries

for positive dependence	$p(\alpha_{kl})$ preset	$p(\alpha_{kl})$ shuffled	$p(\alpha_{kl}^{(\mathrm{ss})})$	$p(\alpha_{kl}^{(\mathrm{ll})})$	$p(\alpha_{kl}^{(\mathrm{sl})})$	$p(\alpha_{kl}^{(\mathrm{ls})})$
	0.0977	0.0977	0.0665	0.1247	-0.0351	0.0008
for negative dependence	$p(\beta_{kl})$ preset	$p(\beta_{kl})$ shuffled	$p(\beta_{kl}^{(\mathrm{ss})})$	$p(\beta_{kl}^{(\mathrm{ll})})$	$p(\beta_{kl}^{(\mathrm{sl})})$	$p(\beta_{kl}^{(\mathrm{ls})})$
	-0.0319	-0.0257	-0.0373	-0.0473	-0.0186	-0.0140

density is given by

$$\begin{aligned}\langle g\rangle(\mathbf{r}|C,N) &= \frac{1}{2^{N/2+1}\Gamma(N/2)\sqrt{\det(2\pi C/N)}}\frac{\mathcal{K}_{(K-N)/2}\left(\sqrt{N\mathbf{r}^{\dagger}C^{-1}\mathbf{r}}\right)}{\sqrt{N\mathbf{r}^{\dagger}C^{-1}\mathbf{r}}^{(K-N)/2}} \\ . &= \frac{1}{(2\pi)^{K}\Gamma(N/2)\sqrt{\det C}}\int\limits_{0}^{\infty} dz z^{\frac{N}{2}-1}e^{-z}\sqrt{\frac{\pi N}{z}}^{K}\exp\left(-\frac{N}{4z}\mathbf{r}^{\dagger}C^{-1}\mathbf{r}\right) .\end{aligned} \tag{7.15}$$

The notation $\langle g\rangle$ indicates that this distribution results from a random matrix average to model non-stationary, *i.e.*, fluctuating covariance or correlation matrices with a mean value C. The parameter N measures the strength of these fluctuations, $1/N$ can be viewed as the corresponding variance. $\mathcal{K}_m$ is the modified Bessel function of the second kind of order m. In the present content $\mathbf{r}$ is a vector of returns, which are time series $r_k = r_k(t), t = 1, \cdots, T$. The distribution (7.16) is assumed to hold for each time t. It is worth mentioning that the parameter N is different from $N_{\text{trades}}(t)$ in Eq. (7.2), which represents the number of trades in the time interval t. In the bivariate case $K = 2$, the joint pdf of the $\mathcal{K}$ distribution reads,

$$f(x_1,x_2) = \frac{1}{\Gamma(N/2)}\int\limits_{0}^{\infty} dz z^{\frac{N}{2}-1}e^{-z}\frac{N}{4\pi z}\frac{1}{\sqrt{1-c^2}}\exp\left(-\frac{N}{4z}\frac{x_1^2-2cx_1x_2+x_2^2}{1-c^2}\right) , \tag{7.16}$$

with the correlation matrix

$$C = \left[\begin{array}{cc} 1 & c \\ c & 1 \end{array}\right] , \tag{7.17}$$

which only depends on one correlation coefficient c. By integrating $f(x_1, x_2)$ over the whole range of x_2, we can obtain the marginal distribution density,

$$\begin{aligned} f_k(x_1) &= \int\limits_{-\infty}^{\infty} dx_2 f(x_1,x_2) \\ &= \frac{1}{\Gamma(N/2)}\int\limits_{0}^{\infty} dz z^{\frac{N}{2}-1}e^{-z}\sqrt{\frac{N}{4\pi z}}exp\left(-\frac{N}{4z}x_1^2\right) ,\end{aligned} \tag{7.18}$$

and analogously for $f_l(x_2)$. Further, the integral of the probability density function yields the marginal cumulative distribution,

$$\begin{aligned} F_k(x_1) &= \int\limits_{-\infty}^{x_1} d\xi f_k(\xi) \\ &= \frac{1}{\Gamma(N/2)}\int\limits_{0}^{\infty} dz z^{\frac{N}{2}-1}e^{-z}\int\limits_{-\infty}^{x_1} d\xi\sqrt{\frac{N}{4\pi z}}exp\left(-\frac{N}{4z}\xi^2\right) ,\end{aligned} \tag{7.19}$$

and $F_l(x_2)$ accordingly. With Eqs. (7.6) and (7.7), the copula density function can be derived as

$$\text{cop}^{\mathcal{K}}_{c,N}(q_1,q_2) = \frac{f\big(F_k^{-1}(q_1),F_l^{-1}(q_2)\big)}{f_k\big(F_k^{-1}(q_1)\big)f_l\big(F_l^{-1}(q_2)\big)} . \tag{7.20}$$

A more detailed discussion of the bivariate $\mathcal{K}$ copula is given in Ref. [41].

7.4.2 Gaussian copula density

Here, one assumes that the random variables x_1 and x_2, normalized to zero mean and unit variance, follow a bivariate normal distribution with a correlation coefficient c. The bivariate cumulative normal distribution of x_1 and x_2 is given by

$$F(x_1,x_2) = \int_{-\infty}^{x_1}\int_{-\infty}^{x_2} \frac{1}{2\pi\sqrt{1-c^2}}\exp\left(-\frac{y_1^2+y_2^2-2cy_1y_2}{2(1-c^2)}\right)dy_2dy_1 \ . \tag{7.21}$$

Hence, the marginal cumulative normal distribution of x_1 is

$$F_k(x_1) = \int_{-\infty}^{x_1} \frac{1}{\sqrt{2\pi}}\exp\left(-\frac{y_1^2}{2}\right)dy_1 \ , \tag{7.22}$$

and analogously for $F_l(x_2)$. Using Eqs. (7.21), (7.22) and (7.6), we find an explicit expression of the Gaussian copula density

$$\begin{aligned}\mathrm{cop}_c^G(q_1,q_2) &= \frac{\partial^2}{\partial q_1\partial q_2}F\big(F_k^{-1}(q_1),F_l^{-1}(q_2)\big)\\ &= \frac{1}{\sqrt{1-c^2}}\exp\left(-\frac{c^2F_k^{-1}(q_1)^2+c^2F_l^{-1}(q_2)^2-2cF_k^{-1}(q_1)F_l^{-1}(q_2)}{2(1-c^2)}\right)\end{aligned} \tag{7.23}$$

by carrying out the partial derivatives in Eq. (7.7).

7.4.3 Fits

To fit the empirical copula, we first work out the average correlation coefficient $\bar{c} = 0.10$ by averaging over $L(L-1)/2$ stock pairs for the $L = 100$ corresponding to 100 stocks listed in Appendix A.4. In the $\mathcal{K}$ copula density function in Eq. (7.20), the correlation coefficient c is replaced by $\bar{c}$. Thus, in Eq. (7.20), only the free parameter N needs to be fitted. By minimizing the squared difference between the empirical copula density and the model copula density, we find $N = 6.72$. With the same $\bar{c}$, we also carry out this comparison using the Gaussian copula density function in Eq. (7.23). To quantify the goodness of fit, we work out the difference between the empirical copula density and the model copula density. Fig. 7.4 shows the two fits and the difference between data and model. The Gaussian copula density differs from the empirical one by a large extent. In particular, the tailed dependencies are poorly captured. In contrast, the $\mathcal{K}$ copula density exhibits a good fit to the empirical result, as it works much better for the tails. This supports previous studies in which the $\mathcal{K}$ distribution was found to give good descriptions of multivariate data subject to nonstationarities [39, 40, 140, 141].

7.5 Influence of local fluctuations on dependencies

In Sec. 7.5.1, We discuss a method to analyze the conditional copula density. In Sec. 7.5.2, we define copula densities conditioned on the local noise intensity and discuss the influence of large local fluctuations on the dependence of volume imbalances between stocks. In Sec. 7.5.3, we give an explanation of this influence with respect to the cross-correlation of volume imbalances. In Sec. 7.5.4, we investigate the influence of large local fluctuations on the asymmetries of tailed dependencies.

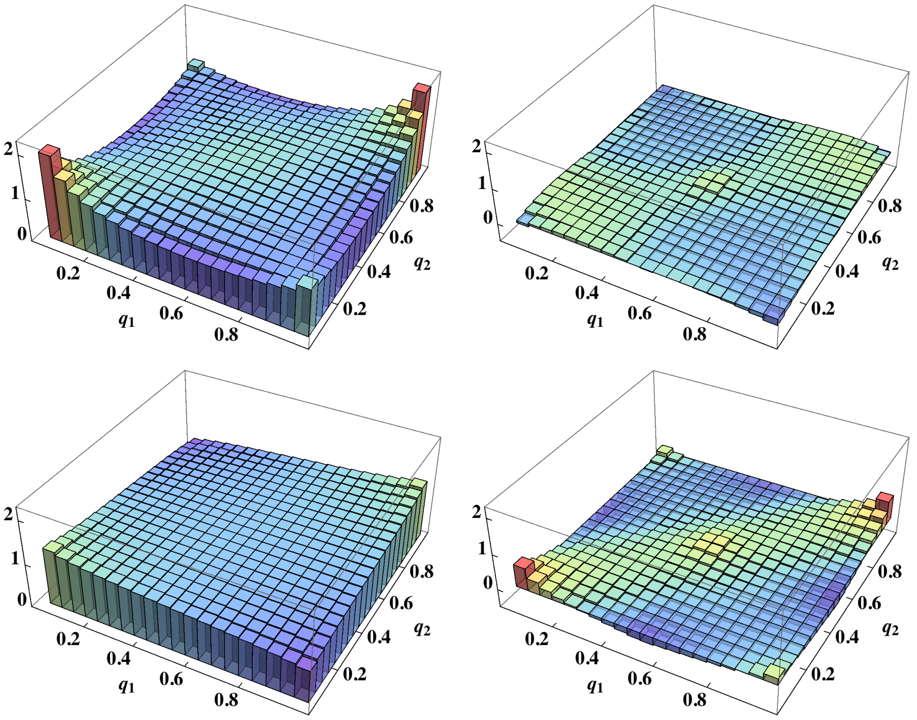


Figure 7.4: $\mathcal{K}$ copula density $\mathrm{cop}^{\mathcal{K}}_{\bar{c},N}(q_1, q_2)$ with $\bar{c} = 0.10$ and $N = 6.72$ (left, top). The error between the empirical copula density and the $\mathcal{K}$ copula density, defined as $\mathrm{cop}(q_1, q_2) - \mathrm{cop}^{\mathcal{K}}_{\bar{c},N}(q_1, q_2)$, (right, top). Gaussian copula density $\mathrm{cop}^{G}_{\bar{c}}(q_1, q_2)$ with $\bar{c} = 0.10$ (left, bottom). The error between the empirical copula density and the Gaussian copula density, defined as $\mathrm{cop}(q_1, q_2) - \mathrm{cop}^{G}_{\bar{c}}(q_1, q_2)$, (right, bottom).

7.5.1 Feasibility of our method

We work out the cumulative probability densities of the numbers of data points N_{data} for the four ranges of noises in Fig. 7.3. A data point gives a volume imbalance as well as a corresponding local noise intensity in the time interval of one minute. As seen in Fig. 7.3, for larger numbers of data points, it is less possible to observe the bimodal distribution. In particular, for some stocks, the bimodal distribution with $\Sigma > 4$ results from only several dozens of data points. When considering the conditional dependencies of demands between two individual stocks that have bimodal marginal distributions, however, these data points are not sufficient to have access to the better statistical property. We thus employ the following method to measure the influence of large local noise intensity. First, we work out the conditional copula density, excluding 50 data points with the largest local noise intensity from both stocks or either stock of a pair. We find little difference between the copula densities excluding 10, 50 and 100 such data points, respectively. However, due to data points that result in a unimodal distribution, enlarging the number of such data points to more than 100 will make the copula density different. Next, we subtract that conditional copula density from the corresponding unconditional one including all data points. The difference between them is the part induced by the large local noise intensity.

7.5.2 Influence on the dependence structure

We now condition the empirical copula densities on the local noise intensity Σ. The conditional copula densities are worked out by excluding the first 50 data points with the largest or smallest local noise intensity. The exclusion of data points with extremely small local noise intensity is to rule out the construed effect that the large change of dependence structure is randomly induced by excluding any kind of data points. Let $\Sigma_{k,\max}$ denote the minimum of the first 50 data points with the extremely large local noise intensity for stock k, and $\Sigma_{k,\min}$ the maximum of the first 50 data points with extremely small local noise intensity for this stock. We write the conditional copula densities as

$$\begin{aligned}
\mathrm{cop}^{(\mathrm{ss})}(q_1,q_2) &= \mathrm{cop}\left(q_1,q_2\middle|\Sigma_k<\Sigma_{k,\max},\Sigma_l<\Sigma_{l,\max}\right) \ , \\
\mathrm{cop}^{(\mathrm{ll})}(q_1,q_2) &= \mathrm{cop}\left(q_1,q_2\middle|\Sigma_k>\Sigma_{k,\min},\Sigma_l>\Sigma_{l,\min}\right) \ , \\
\mathrm{cop}^{(\mathrm{sl})}(q_1,q_2) &= \mathrm{cop}\left(q_1,q_2\middle|\Sigma_k<\Sigma_{k,\max},\Sigma_l>\Sigma_{l,\min}\right) \ , \\
\mathrm{cop}^{(\mathrm{ls})}(q_1,q_2) &= \mathrm{cop}\left(q_1,q_2\middle|\Sigma_k>\Sigma_{k,\min},\Sigma_l<\Sigma_{l,\max}\right) \ .
\end{aligned} \tag{7.24}$$

Here, Σ_k and Σ_l are the local noise intensity for stock k and stock l, respectively. Furthermore, $\mathrm{cop}^{(\mathrm{ss})}(q_1,q_2)$ indicates that the copula density results from the quantiles q_1 and q_2 with small local noise intensity, while $\mathrm{cop}^{(\mathrm{ll})}(q_1,q_2)$ represents the opposite case. Similarly, $\mathrm{cop}^{(\mathrm{sl})}(q_1,q_2)$ denotes the copula density from the quantiles q_1 with small local noise intensity and the quantiles q_2 with the large local noise intensity, and vice versa for $\mathrm{cop}^{(\mathrm{ls})}(q_1,q_2)$. We show the four types of conditional copula densities in Fig. 7.5. Only $\mathrm{cop}^{(\mathrm{ll})}(q_1,q_2)$ reveals strongly positive dependencies, the dependencies in the other copula densities are nearly uniform at the corners and the centres.

To study the influence of large local fluctuations, indicated by the local noise intensity, we look at the difference between the unconditional and the conditional copula densities,

$$\begin{aligned}
\Delta\mathrm{cop}^{(\mathrm{ll})}(q_1,q_2) &= \mathrm{cop}(q_1,q_2)-\mathrm{cop}^{(\mathrm{ss})}(q_1,q_2) \ , \\
\Delta\mathrm{cop}^{(\mathrm{ss})}(q_1,q_2) &= \mathrm{cop}(q_1,q_2)-\mathrm{cop}^{(\mathrm{ll})}(q_1,q_2) \ , \\
\Delta\mathrm{cop}^{(\mathrm{ls})}(q_1,q_2) &= \mathrm{cop}(q_1,q_2)-\mathrm{cop}^{(\mathrm{sl})}(q_1,q_2) \ , \\
\Delta\mathrm{cop}^{(\mathrm{sl})}(q_1,q_2) &= \mathrm{cop}(q_1,q_2)-\mathrm{cop}^{(\mathrm{ls})}(q_1,q_2) \ .
\end{aligned} \tag{7.25}$$

The unconditional copula density $\mathrm{cop}(q_1,q_2)$ is worked out with all data points. As shown in Fig. 7.6, the extremely small local fluctuations from two stocks have a very slight effect on the positive dependencies of the copula density. This effect is quantified by $\Delta\mathrm{cop}^{(\mathrm{ss})}(q_1,q_2)$. However, the extremely large local fluctuations present in either stock or both stocks of a pair not only enhance the positive dependencies, but also suppress the negative dependencies of the copula densities. The degrees of enhancing and suppressing are measured by $\Delta\mathrm{cop}^{(\mathrm{sl})}(q_1,q_2)$, $\Delta\mathrm{cop}^{(\mathrm{ls})}(q_1,q_2)$, and $\Delta\mathrm{cop}^{(\mathrm{ll})}(q_1,q_2)$, respectively. Comparing the effects of the extremely large and small local fluctuations, we find that the construed effect is absent in the change of dependencies due to large local fluctuations. In the copula densities, the lower corner along the positive diagonal, corresponding to the negative volume imbalances, reveals the dependence of supplies between stocks, while the upper corner along the positive diagonal, corresponding to the positive volume imbalances, reveals the dependence of demands. Combining Figs. 7.5 and 7.6, we find that the extremely large local fluctuations in either stock of a pair are crucial to prompt the strong dependence between demands or supplies. A possible interpretation might be as

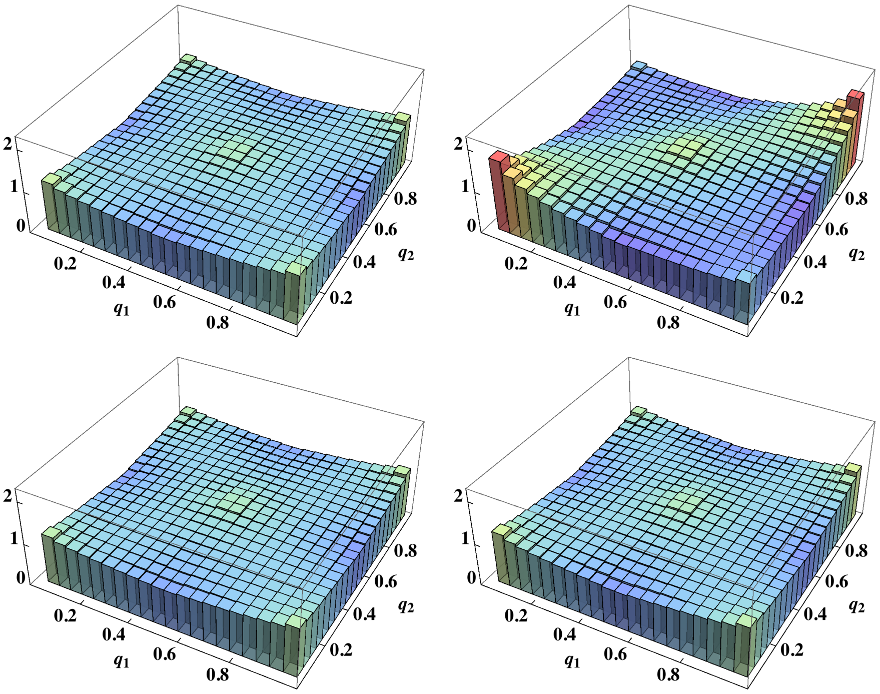


Figure 7.5: Empirical copula densities conditioned on the local noise intensity $\mathrm{cop}^{(\mathrm{ss})}(q_1, q_2)$ (left, top), $\mathrm{cop}^{(\mathrm{ll})}(q_1, q_2)$ (right, top), $\mathrm{cop}^{(\mathrm{sl})}(q_1, q_2)$ (left, bottom), and $\mathrm{cop}^{(\mathrm{ls})}(q_1, q_2)$ (right, bottom).

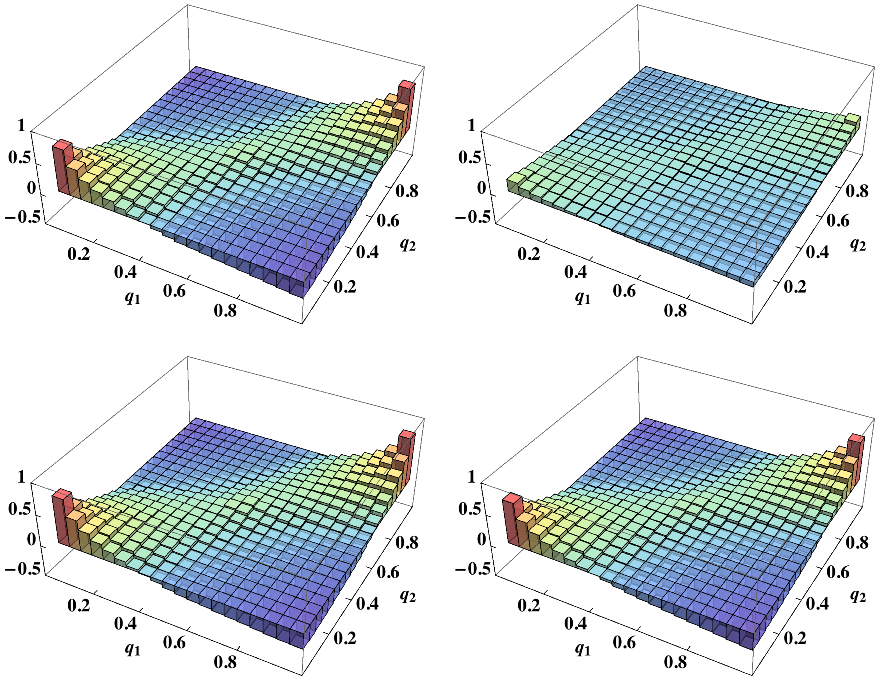


Figure 7.6: The influences of the local noise intensity on the copula density $\Delta\mathrm{cop}^{(\mathrm{ll})}(q_1, q_2)$ (left, top), $\Delta\mathrm{cop}^{(\mathrm{ss})}(q_1, q_2)$ (right, top), $\Delta\mathrm{cop}^{(\mathrm{ls})}(q_1, q_2)$ (left, bottom), and $\Delta\mathrm{cop}^{(\mathrm{sl})}(q_1, q_2)$ (right, bottom).

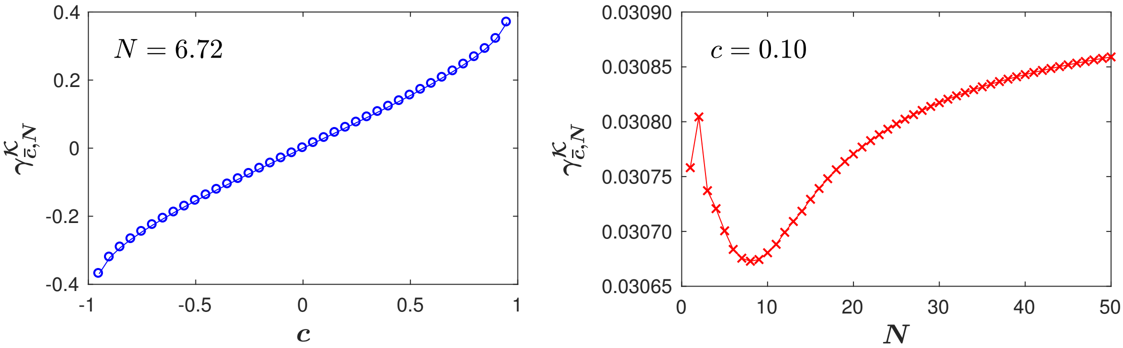


Figure 7.7: Left: the dependence of $\gamma^{\mathcal{K}}_{\bar{c},N}$ on the correlation coefficient c, where $N = 6.72$; right: the dependence of $\gamma^{\mathcal{K}}_{\bar{c},N}$ on the parameter N, where $c = 0.10$. Here, $\gamma^{\mathcal{K}}_{\bar{c},N}$ is the difference between positive and negative dependencies for the bivariate $\mathcal{K}$ copula density. The ranges of vertical axes for two subgraphs are different.

following: An extremely large trade may either be random or include useful information. In any case, the extremely large trade pushes the price up for a large demand or drops the price down for a large supply. Increase of the price may induce higher expectation for the raising of the price or induce herding behavior [20, 138] of trading, leading more volumes to be bought. Analogously, drop of the price leads to more volumes being sold. Due to the correlations between stocks [129], the effect of a large demand or supply of one stock is very likely to spread to another stock and induce the similar behavior for the volumes to be bought or sold. The presence of such large trades in both stocks of a pair causes, on the one hand, large local fluctuations in the stocks, and, on the other hand, mutual dependence of demands or supplies in the considered stock pair.

7.5.3 Correlations induced by large local fluctuations

As shown in the top row of Fig. 7.5, the change due to large local fluctuations are mainly visible in the positive and negative corners of the copula density. To quantify how such fluctuations affect the dependence structure, we define the difference between positive and negative dependencies of demands for a stock pair $(k,\, l)$ as

$$\begin{aligned}\gamma_{kl} \quad &= \quad \Bigg(\int\limits_{0.8}^{1} dq_1 \int\limits_{0.8}^{1} dq_2 \ \mathrm{cop}_{kl}(q_1, q_2) + \int\limits_{0}^{0.2} dq_1 \int\limits_{0}^{0.2} dq_2 \ \mathrm{cop}_{kl}(q_1, q_2)\Bigg) \\ &\quad - \Bigg(\int\limits_{0}^{0.2} dq_1 \int\limits_{0.8}^{1} dq_2 \ \mathrm{cop}_{kl}(q_1, q_2) + \int\limits_{0.8}^{1} dq_1 \int\limits_{0}^{0.2} dq_2 \ \mathrm{cop}_{kl}(q_1, q_2)\Bigg) \,, \qquad (7.26)\end{aligned}$$

where the terms in the brackets do not coincide with α_{kl} and β_{kl} in Eqs. (7.11) and (7.12). However, the amount of data points is not large enough to empirically analyze γ_{kl}. Rather, we resort to the $\mathcal{K}$ copula density (7.20) which, as we have shown, describes the data well. Hence, we replace $\mathrm{cop}_{kl}(q_1, q_2)$ by $\mathrm{cop}^{\mathcal{K}}_{\bar{c},N}(q_1, q_2)$ in definition (7.20) with $\gamma^{\mathcal{K}}_{\bar{c},N}$ instead of γ_{kl}. Using Eq. (7.20), we calculate $\gamma^{\mathcal{K}}_{\bar{c},N}$ as a function of c for a given N and vice versa, respectively, as shown in Fig. 7.7. The two given values $N = 6.72$ and $c = 0.10$ are from the fit to the empirical copula. We find that the difference between positive and negative dependencies of demands is drastically affected by the correlation coefficient c rather than by the parameter N. This leads us to dissect the cross-correlation of the volume imbalance between stocks k and l,

$$\begin{aligned}\mathrm{corr}\big(\nu_k(t),\nu_l(t)\big) &= \big\langle \nu_k(t)\nu_l(t)\big\rangle_t \\ &= \big\langle p_k^+(t)|\nu_k(t)|p_l^+(t)|\nu_l(t)| + p_k^-(t)|\nu_k(t)|p_l^-(t)|\nu_l(t)| \\ &\quad -p_k^+(t)|\nu_k(t)|p_l^-(t)|\nu_l(t)| - p_k^-(t)|\nu_k(t)|p_l^+(t)|\nu_l(t)|\big\rangle_t \\ &= \big\langle P_{kl}(t)|\nu_k(t)||\nu_l(t)|\big\rangle_t \ . \end{aligned} \tag{7.27}$$

Here, $p_k^+(t)$ is the probability that a surplus of volumes is bought for stock k in time interval t, corresponding to the positive volume imbalances of stock k, and $p_k^-(t)$ is the probability that a surplus of volumes is sold, corresponding to the negative volume imbalances. Importantly, we have $p_k^+(t) + p_k^-(t) = 1$. The quantity $P_{kl}(t)$ introduced in Eq. (7.27) can be written as

$$\begin{aligned}P_{kl}(t) &= p_k^+(t)p_l^+(t) + p_k^-(t)p_l^-(t) - p_k^+(t)p_l^-(t) - p_k^-(t)p_l^+(t) \\ &= 4p_k^+(t)p_l^+(t) - 2p_k^+(t) - 2p_l^+(t) + 1 \ , \end{aligned} \tag{7.28}$$

and may be interpreted as effective weight referring to the volume imbalances of both stocks at each time step t. The value of $P_{kl}(t)$ is bound between -1 and 1.

In Ref. [134], Potters and Bouchaud (BP) have demonstrated that the local noise intensity

$$\widetilde{\Sigma}(t) = \Big\langle \big(v(t;n)\varepsilon(t;n) - \langle v(t;n)\varepsilon(t;n)\rangle_n\big)^2\Big\rangle_n \ , \tag{7.29}$$

and the square of volume imbalances are positively correlated,

$$\begin{aligned}\Big\langle \widetilde{\Sigma}(t)\nu^2(t)\Big\rangle &= (N_{\mathrm{trades}}-1)\Big(\big\langle v^4(t;n)\big\rangle - 3\big\langle v^2(t;n)\big\rangle^2\Big) \\ &\quad +(1-\frac{3}{N_{\mathrm{trades}}})\sum_{n_i\neq n_j=1}^{N_{\mathrm{trades}}}\big\langle v^2(t;n_i)v^2(t;n_j)\big\rangle \ , \end{aligned} \tag{7.30}$$

if the traded volumes have fat tails, *i.e.*, $\big\langle v^4(t;n)\big\rangle > 3\big\langle v^2(t;n)\big\rangle^2$, and/or are positively correlated, *i.e.*, $\big\langle v^2(t;n_i)v^2(t;n_j)\big\rangle \geq 0$. They neglect the fluctuation of the number of trades $N_{\mathrm{trades}} = N_{\mathrm{trades}}(t)$. Using their conclusion in our case, we have

$$\Sigma(t) \sim |\nu(t)| \tag{7.31}$$

for fat-tailed traded volumes. Thus, the correlation of the volume imbalance in Eq. (7.27) is approximately

$$\mathrm{corr}\big(\nu_k(t),\nu_l(t)\big) \sim \big\langle P_{kl}(t)\Sigma_k(t)\Sigma_l(t)\big\rangle_t \ . \tag{7.32}$$

We analyze the dependence of $P_{kl}(t)$ on $p_k^+(t)$ and $p_l^+(t)$, see Fig. 7.8. For very small volume imbalances, the probability of a surplus of volumes bought is very close to the one of a surplus of volumes sold in time t, *i.e.* $p_k^+(t) \approx p_k^-(t) \approx 0.5$. For this case, $P_{kl}(t)$ tends to zero, as seen in Fig. 7.8. Accordingly, the correlation of volume imbalances goes towards zero according to Eq. (7.32). A correlation coefficient around zero indicates that the positive dependencies on the copulas are comparable to the negative ones as shown

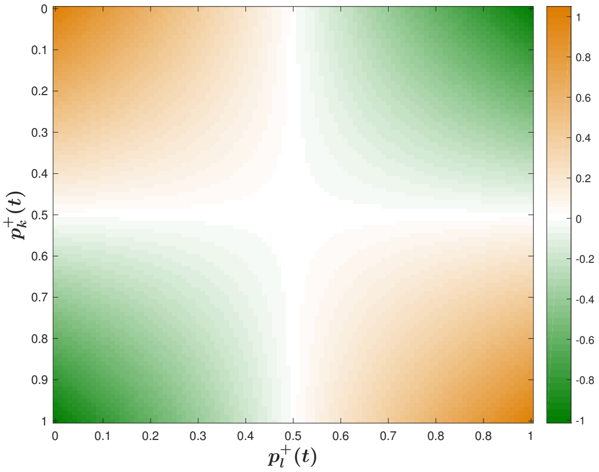


Figure 7.8: The contour of $P_{kl}(t)$ depending on $p_k^+(t)$ and $p_l^+(t)$. The value of $P_{kl}(t)$ is indicated by the color.

in Fig. 7.7. Consequently, the very small local fluctuations, positively correlated with the absolute values of the very small volume imbalances, result in a similar strength of dependencies in the four corners of the copula density, see Fig. 7.5. In contrast, the very large volume imbalances, corresponding to the very large local fluctuations according to Eq. (7.31), imply a high probability for most of the traded volumes being bought or sold. When both stocks k and l have very large volume imbalances, we find a rather high effective weight $P_{kl}(t)$ at the four corners of Fig. 7.8. As a result, the very large local fluctuations in both stocks together with a high value of $P_{kl}(t)$ lead to a considerable correlation of volume imbalances. This correlation turns out to be positive, as the positive dependencies prevailing over negative ones result in a positive asymmetry γ_{kl}, corresponding to a positive correlation in Fig. 7.7.

7.5.4 Influence on the asymmetries of tail dependencies

In Sec. 7.3, we quantified and analyzed the asymmetries of tail dependencies between stocks. Here, we want to find out how the large local fluctuations act on the tail asymmetries in the copula density, characterized by α_{kl} and β_{kl} for positive and negative dependencies, respectively. We work out the distributions of α_{kl} and β_{kl} for four conditional copula densities, defined in Eq. (7.24), and show the results in Figs. 7.9 and 7.10. For the negative dependencies in Fig. 7.10, the overall asymmetries are not pronounced in the four distributions $p(\beta_{kl}^{(\mathrm{ss})})$, $p(\beta_{kl}^{(\mathrm{ll})})$, $p(\beta_{kl}^{(\mathrm{sl})})$, $p(\beta_{kl}^{(\mathrm{ls})})$. Their skewness in Table 7.1 is relatively small and changes slightly compared to the skewness of the distributions of α_{kl}. For the positive dependencies, the overall asymmetries depend on the local fluctuations, see Fig. 7.9. If both stocks of a pair have small local fluctuations, the skewness of the distribution $p(\alpha_{kl}^{(\mathrm{ss})})$ is 0.0665, which is smaller than the value of 0.0977 in the unconditional copula density, defined in Eq. (7.9). If both stocks have large local fluctuations, an overall right shift of the distribution $p(\beta_{kl}^{(\mathrm{ll})})$ shows up with a skewness of 0.1247. If one stock has large local fluctuations and the other one has small local fluctuations, independently of the symmetry or asymmetry of $p(\beta_{kl}^{(\mathrm{sl})})$ and $p(\beta_{kl}^{(\mathrm{ls})})$, we find a very small skewness for

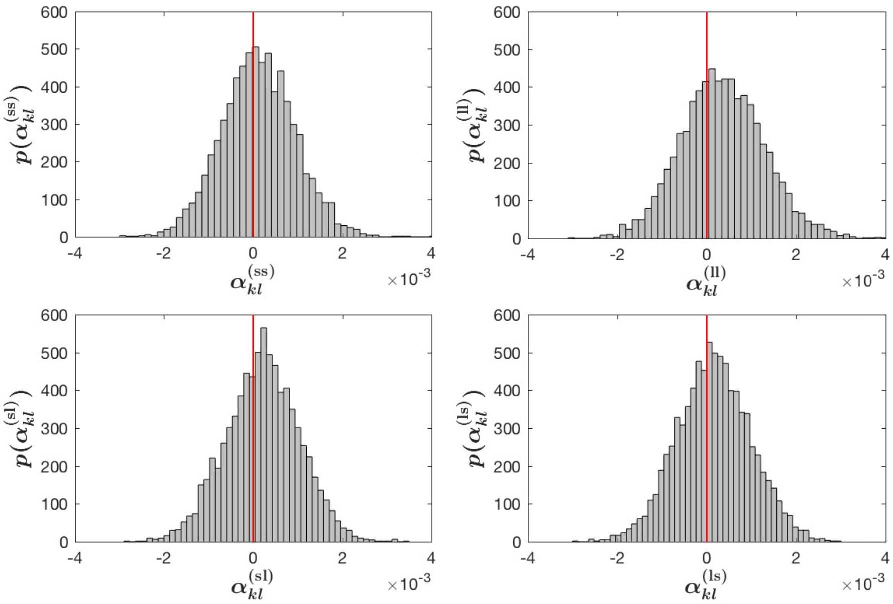


Figure 7.9: Histograms of asymmetry values of 4950 stock pairs $(k,\ l)$ for positive dependence $p(\alpha_{kl}^{(\mathrm{ss})})$, $p(\alpha_{kl}^{(\mathrm{ll})})$, $p(\alpha_{kl}^{(\mathrm{sl})})$, and $p(\alpha_{kl}^{(\mathrm{ls})})$, corresponding to the copula densities $\mathrm{cop}^{(\mathrm{ss})}(q_1, q_2)$, $\mathrm{cop}^{(\mathrm{ll})}(q_1, q_2)$, $\mathrm{cop}^{(\mathrm{sl})}(q_1, q_2)$, and $\mathrm{cop}^{(\mathrm{ls})}(q_1, q_2)$, respectively. All the histograms are normalized to one.

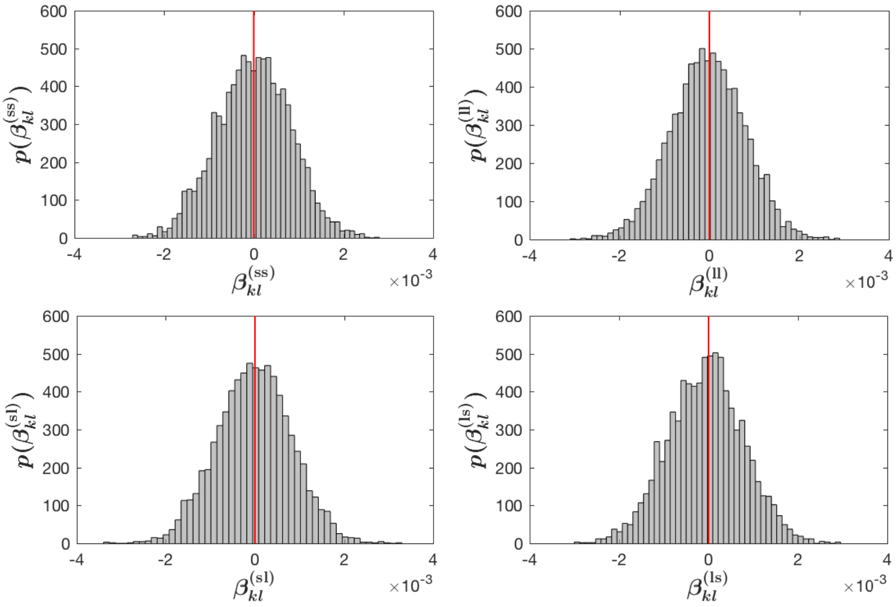


Figure 7.10: Histograms of asymmetry values of 4950 stock pairs $(k,\ l)$ for negative dependence $p(\beta_{kl}^{(\mathrm{ss})})$, $p(\beta_{kl}^{(\mathrm{ll})})$, $p(\beta_{kl}^{(\mathrm{sl})})$, and $p(\beta_{kl}^{(\mathrm{ls})})$, corresponding to the copula densities $\mathrm{cop}^{(\mathrm{ss})}(q_1, q_2)$, $\mathrm{cop}^{(\mathrm{ll})}(q_1, q_2)$, $\mathrm{cop}^{(\mathrm{sl})}(q_1, q_2)$, and $\mathrm{cop}^{(\mathrm{ls})}(q_1, q_2)$, respectively. All the histograms are normalized to one.

them. Among the four distributions for positive dependence, only $p(\alpha_{kl}^{(\mathrm{ll})})$ exhibits sizeable right shift and a positive fat tail, implying that the large local fluctuations in both stocks contribute to the dependence of demands more than the one of supplies, probably indicating a bull market.

7.6 Conclusions

We investigated the influence of large local fluctuations on the dependence of demands between stocks. The demand is quantified by the volume imbalance, where the positive demand is due to a surplus of volumes bought, while the negative demand is the supply if a surplus of the volumes is sold. We employed copulas to study the dependence of demands, and found stronger positive dependencies than negative ones. Hence, if the demand for one stock is large, it is likely to find large demand for other stocks as well. The situation is analogous for supplies. The bivariate $\mathcal{K}$ copula density function describes the empirical copula density better than the Gaussian one, especially the fat-tailed dependencies. The bivariate $\mathcal{K}$ copula density function follows from a random matrix model and only depends on two parameters, an average correlation coefficient c and a parameter N measuring the strength of the fluctuations of the correlations.

We discussed the empirical copula densities conditioned on the local noise intensities, and found that the extremely large local fluctuations from both stocks of a pair strengthen the positive dependencies of demands but weaken the negative ones. We attributed this interesting feature to the cross-correlation of volume imbalances between stocks, which in turn is related to large local fluctuations and signs of the volume imbalances. We uncover that the larger the local fluctuations, the stronger is the cross-correlation of volume imbalances, and the bigger is the difference between positive and negative dependencies of demands in the copula densities.

We also looked at the asymmetries of tail dependencies of demands. They are not pronounced for negative dependencies but sizeable for the positive ones. For the latter, the large local fluctuations cause a shift from zero to the right in the distribution of the asymmetries. We therefore conclude that large local fluctuations influence the dependence of demands more than the dependence of supplies, probably reflecting a bull market with persistent increase of prices in the markets.

Chapter 8

Conclusion and Outlook

In the last two decades, the market microstructure has gained growing attention. Due to a gigantic amount of available data, it is possible to statistically analyze financial markets and reveal the market microstructure. In particular, with the development of high-frequency trading, the study of the market microstructure is of obvious importance for practical purposes, for instance, price impacts, execution costs, and market risks. In this thesis, we began with the analysis of empirical data and then constructed proper models to interpret the empirical results. As an extension, we also showed a basic application for our findings. Different from some previous studies, we considered the financial market as a whole, and focused on the relation between stocks, rather than single stocks.

We first developed a method to classify trade signs with the TAQ data set. This method effectively classifies the signs for the trades recorded in the TAQ data set during the time intervals of one second and the sign for the aggregated trade during each time interval. By tests, this method has an average accuracy of 85% for the former and of 82% for the latter. With the classified trade signs, we carried out the cross-response as well as the cross-correlation of trade signs for pairs of stocks. The empirical analysis is based on a physical time scale that considers each time interval as a time step. Due to strong fluctuations at large time lags, the cross-correlation of trade signs turns out to be short memory for stock pairs. Meanwhile, the non-vanishing cross-response reflects the non-Markovian features of prices.

To reduce the noise and have a stable observation, we performed the averages of cross-responses for an individual stock. By performing conceptually different averages, we obtained active and passive cross-responses. The active cross-response measures which effect the trades of one stock have on the average price of other stocks, while the passive one quantifies how the price of one stock changes due to the trades of other stocks. Interestingly, the two types of average cross-responses show different characteristics. The active one lasts for a longer time before the price reversion than the passive one. Furthermore, we observed the average cross-response for the market as whole. The non-zero response implies that the market lacks of efficiency at a short time scale but restores the efficiency at a longer time scale. We also carried out the average cross-correlation of trade signs. Since the strong fluctuations at the large time lags are effectively wished out, the average cross-correlation of trade signs turns out to be long memory. That means a buy (sell) market order is more likely to be followed by more buy (sell) market orders.

To interpret our empirical results, we extended the price impact model of Bouchaud et al. [27]. In our model, the price is impacted by traded volumes from two stocks at a given time step. The impacts increase instantaneously to maxima and then decay with time. We used the impact functions of traded volumes to describe the instantaneous process. The empirical results manifest that the impact functions of traded volumes follow a power law. We characterized the decaying process by a self- and a cross-impact function of time lag. The self-impact function comes from the stock itself and the cross-impact function comes from a different stock. Due to the difficulty to quantify the price impacts depending on the time lag from empirical data, we thus proposed a construction to fix the parameters for theses impact functions. Further, we employed a diffusion function to corroborate the parameters. According to the quantified price impacts, we found that the price self-impact contains a strong temporary component and a very weak permanent component. In contrast, we found that the active and passive impacts, corresponding to the active and passive response functions, respectively, only have the temporary component.

The price change due to a market order will lead to an extra cost for trading. To reduce the transaction cost, we considered a trading strategy for executing two round-trip trades of two stocks, where the cross-impacts between two stocks were introduced. Our strategy is extended from the framework of the trading strategy of Gatheral (2010) for single stocks [65], and it depends on three parameters, *i.e.*, the ratio of trading periods of two stocks and the rate of trading of each stock. By minimizing the cross-impact cost, the strategy can arrive at the optimization. We applied the strategy to a pair of stocks. With the cross-impacts quantified by the empirical data, we displayed the effect of the cross-impacts on the trading strategy. The positive cross-impact cost suggests that traders can circumvent the extra cost if they execute two round-trip trades without any overlap in time. The negative cross-impact cost implies that finding opportunities of arbitrage or reducing the total transaction cost is possible if traders start the two round-trip trades at the same time.

The traded volumes are crucial for price impacts and optimal executions of orders. We introduced the trade sign into the traded volume, leading to the signed traded volume. In a certain time interval, the sum of all the signed traded volumes is the volume imbalance, which measures the demand of one stock in that time interval. We investigated the dependence of demands between stocks using copulas, and demonstrated that the empirical dependence can be well described by a bivariate $\mathcal{K}$ copula density function. Also, we used the local noise intensity to quantify the local fluctuations, and analyzed the effect of the large local fluctuations on the dependencies of demands. The empirical results show that the extremely large local fluctuations from both stocks of a pair strength the positive dependencies of demands but weaken the negative ones. We attributed this interesting feature to the cross-correlation of volume imbalances between stocks. Furthermore, by analyzing the asymmetries of tail dependencies of demands, we found that the large local fluctuations influence the dependencies of positive demands more significantly than the dependencies of negative demands.

By analyzing a large-scale of real-time data, we found many interesting features for the correlated financial markets. We also used appropriate models to interpret them, but the model-based interpretations are not totally confined to this thesis. For better understanding the features, more proper models from a new perspective are expected. We introduced the cross-impact to a trading strategy. To highlight the role of cross-impacts, we restricted ourselves to considering two stocks. Typically, a portfolio contains many different stocks. We thus expect to construct a strategy in terms of the cross-impact for more than two stocks. Furthermore, it is also expected to employ an alternative framework of trading strategies, such as the Markowitz theory for portfolio optimization. In general, a linear combination of risk elements determines a portfolio. Introducing the cross-impact in the Markowitz theory will result in an extra risk, which will affect the portfolio.

Appendix A

Stock information

A.1 Trading information of 99 stocks

In Chapter 3, we evaluated the market response for the 99 stocks from ten economic sectors: industrials (I), health care (HC), consumer discretionary (CD), information technology (IT), utilities (U), financials (F), materials (M), energy (E), consumer staples (CS), and telecommunications services (TS) as listed in Table A.1. These stocks are also used for the analyses of influential and influenced stocks in Chapter 4. The acronym AMC in Table A.1 stands for averaged market capitalization.

Table A.1: Information of 99 stocks from ten economic sectors

Industrials (I)

Symbol	Company	AMC
FLR	Fluor Corp. (New)	14414.4
LMT	Lockheed Martin Corp.	12857.8
FLS	Flowserve Corporation	12670.2
PCP	Precision Castparts	12447.0
LLL	L-3 Communications Holdings	12170.8
UNP	Union Pacific	11920.9
BNI	Burlington Northern Santa Fe C	11837.5
FDX	FedEx Corporation	10574.7
GWW	Grainger (W.W.) Inc.	10416.8
GD	General Dynamics	10035.6

Financials (F)

Symbol	Company	AMC
CME	CME Group Inc.	49222.9
GS	Goldman Sachs Group	21524.3
ICE	Intercontinental Exchange Inc.	14615.3
AVB	AvalonBay Communities	11081.6
BEN	Franklin Resources	10966.2
BXP	Boston Properties	10893.0
SPG	Simon Property Group Inc	10862.4
VNO	Vornado Realty Trust	10802.3
PSA	Public Storage	10147.9
MTB	M&T Bank Corp.	9920.2

Health Care (HC)

Symbol	Company	AMC
ISRG	Intuitive Surgical Inc.	31355.9
BCR	Bard (C.R.) Inc.	11362.7
BDX	Becton Dickinson	10298.4
GENZ	Genzyme Corp.	9728.8
JNJ	Johnson & Johnson	9682.6
LH	Laboratory Corp. of America Holding	9035.7
ESRX	Express Scripts	8864.6
CELG	Celgene Corp.	8783.1
ZMH	Zimmer Holdings	8681.7
AMGN	Amgen	8543.0

Materials (M)

Symbol	Company	AMC
X	United States Steel Corp.	15937.7
MON	Monsanto Co.	14662.6
CF	CF Industries Holdings Inc	14075.5
FCX	Freeport-McMoran Cp & Gld	11735.7
APD	Air Products & Chemicals	10246.4
PX	Praxair Inc.	10234.5
VMC	Vulcan Materials	8700.4
ROH	Rohm & Haas	8527.1
NUE	Nucor Corp.	7997.4
PPG	PPG Industries	7336.7

Table A.1: (continued)

Consumer Discretionary (CD)

Symbol	Company	AMC
WPO	Washington Post	61856.1
AZO	AutoZone Inc.	14463.7
SHLD	Sears Holdings Corporation	11759.2
WYNN	Wynn Resorts Ltd.	11507.9
AMZN	Amazon Corp.	10939.2
WHR	Whirlpool Corp.	9501.9
VFC	V.F. Corp.	9051.2
APOL	Apollo Group	8495.8
NKE	NIKE Inc.	8149.5
MCD	McDonald's Corp.	8025.6

Energy (E)

Symbol	Company	AMC
RIG	Transocean Inc. (New)	16409.5
APA	Apache Corp.	13981.9
EOG	EOG Resources	13095.0
DVN	Devon Energy Corp.	12499.7
HES	Hess Corporation	11990.4
XOM	Exxon Mobil Corp.	11460.3
SLB	Schlumberger Ltd.	11241.1
CVX	Chevron Corp.	11100.0
COP	ConocoPhillips	10215.3
OXY	Occidental Petroleum	9758.4

Information Technology (IT)

Symbol	Company	AMC
GOOG	Google Inc.	62971.6
MA	Mastercard Inc.	28287.8
AAPL	Apple Inc.	22104.1
IBM	International Bus. Machines	15424.9
MSFT	Microsoft Corp.	10845.1
CSCO	Cisco Systems	8731.4
INTC	Intel Corp.	8385.8
QCOM	QUALCOMM Inc.	7739.4
CRM	Salesforce Com Inc.	7691.9
WFR	MEMC Electronic Materials	7392.8

Consumer Staples (CS)

Symbol	Company	AMC
BUD	Anheuser-Busch	9780.6
PG	Procter & Gamble	9711.5
CL	Colgate-Palmolive	9549.2
COST	Costco Co.	9545.9
WMT	Wal-Mart Stores	9325.7
PEP	PepsiCo Inc.	9180.7
LO	Lorillard Inc.	8919.0
UST	UST Inc.	8433.1
GIS	General Mills	8243.3
KMB	Kimberly-Clark	8069.5

Utilities (U)

Symbol	Company	AMC
ETR	Entergy Corp.	12798.7
EXC	Exelon Corp.	9738.8
CEG	Constellation Energy Group	9061.5
FE	FirstEnergy Corp.	8689.4
FPL	FPL Group	7742.8
SRE	Sempra Energy	6940.6
STR	Questar Corp.	6520.4
TEG	Integrys Energy Group Inc.	5978.4
EIX	Edison Int'l	5877.5
AYE	Allegheny Energy	5864.9

Telecommunications Services (TS)

Symbol	Company	AMC
T	AT&T Inc.	6336.2
VZ	Verizon Communications	5732.5
EQ	Embarq Corporation	5318.7
AMT	American Tower Corp.	5195.6
CTL	Century Telephone	4333.8
S	Sprint Nextel Corp.	2533.7
Q	Qwest Communications Int	2201.3
WIN	Windstream Corporation	2089.1
FTR	Frontier Communications	1580.9

A.2 Lists of 496 stocks in S&P 500 index

In Chapter 4, the averages of cross-responses are performed over 496 stocks. These stocks are from NASDAQ stock market and belong to S&P 500 index in 2008. In the following, we list all the symbols of these stocks, which are classified into ten sectors.

- Consumer Discretionary (CD)

  ADM, AVP, BUD, CAG, CCE, CL, CLX, COST, CPB, CVS, DF, DPS, EL, GIS, HNZ, HSY, K, KFT, KMB, KO, KR, LO, MKC, MO, PBG, PEP, PG, PM, RAI, SJM, SLE, STZ, SVU, SWY, SYY, TAP, TSN, UST, WAG, WFMI, WMT

- ConsumerStaples (CS)

  ADM, AVP, BUD, CAG, CCE, CL, CLX, COST, CPB, CVS, DF, DPS, EL, GIS, HNZ, HSY, K, KFT, KMB, KO, KR, LO, MKC, MO, PBG, PEP, PG, PM, RAI, SJM, SLE, STZ, SVU, SWY, SYY, TAP, TSN, UST, WAG, WFMI, WMT

- Energy (E)

  APA, APC, BHI, BJS, BTU, CAM, CHK, CNX, COG, COP, CVX, DVN, EOG, EP, ESV, HAL, HES, MEE, MRO, MUR, NBL, NBR, NE, NOV, OXY, PXD, RDC, RIG, RRC, SE, SII, SLB, SUN, SWN, TSO, VLO, WFT, WMB, XOM, XTO

- Financials (F)

  ACAS, AFL, AIG, AIV, AIZ, ALL, AMP, AOC, AVB, AXP, BAC, BBT, BEN, BK, BXP, C, CB, CBG, CINF, CIT, CMA, CME, COF, DDR, DFS, EQR, ETFC, FHN, FII, FITB, GNW, GS, HBAN, HCBK, HCP, HIG, HST, ICE, IVZ, JNS, JPM, KEY, KIM, L, LM, LNC, LUK, MBI, MCO, MER, MET, MI, MMC, MS, MTB, NCC, NDAQ, NTRS, NYX, PBCT, PCL, PFG, PGR, PLD, PNC, PRU, PSA, RF, SCHW, SLM, SOV, SPG, STI, STT, TMK, TROW, TRV, UNM, USB, VNO, WB, WFC, XL, ZION

- Health Care (HC)

  ABC, ABI, ABT, AET, AGN, AMGN, BAX, BCR, BDX, BIIB, BMY, BRL, BSX, CAH, CELG, CI, COV, CVH, DGX, DVA, ESRX, FRX, GENZ, GILD, HSP, HUM, ISRG, JNJ, KG, LH, LLY, MCK, MDT, MHS, MIL, MRK, MYL, PDCO, PFE, PKI, RX, SGP, STJ, SYK, THC, TMO, UNH, VAR, WAT, WLP, WPI, WYE, XRAY, ZMH

- Industrials (I)

  AVY, AW, BA, BNI, CAT, CBE, CHRW, CMI, COL, CSX, CTAS, DE, DHR, DOV, EFX, EMR, ETN, EXPD, FAST, FDX, FLR, FLS, GD, GE, GR, GWW, HON, IR, ITT, ITW, JEC, LLL, LMT, LUV, MAS, MMM, MTW, MWW, NOC, NSC, PBI, PCAR, PCP, PH, PLL, R, RHI, ROK, RRD, RTN, TXT, TYC, UNP, UPS, UTX, WMI

- Information Technology (IT)

  A, AAPL, ACS, ADBE, ADI, ADP, ADSK, AKAM, ALTR, AMAT, AMD, APH, BMC, BRCM, CA, CIEN, CPWR, CRM, CSC, CSCO, CTSH, CTXS, CVG, DELL, EBAY, EMC, ERTS, FIS, FISV, GLW, GOOG, HPQ, HRS, IBM, INTC, INTU, JAVA, JBL, JDSU, JNPR, KLAC, LLTC, LSI, LXK, MA, MCHP, MOLX, MOT, MSFT, MU, NOVL, NSM, NTAP, NVDA, NVLS, ORCL, PAYX, QCOM, QLGC, SNDK, SYMC, TDC, TEL, TER, TLAB, TSS, TXN, VRSN, WFR, WU, XLNX, XRX, YHOO

- Materials (M)

  AA, AKS, APD, ATI, BLL, BMS, CF, DD, DOW, ECL, EMN, FCX, IFF, IP, MON, MWV, NEM, NUE, PPG, PTV, PX, ROH, SEE, SIAL, TIE, VMC, WY, X

- Telecommunications Services (TS)

  AMT, CTL, EQ, FTR, Q, S, T, VZ, WIN

- Utilities (U)

  AEE, AEP, AES, AYE, CEG, CMS, CNP, D, DTE, DUK, DYN, ED, EIX, ETR, EXC, FE, FPL, GAS, NI, PCG, PEG, PGN, PNW, POM, PPL, SO, SRE, STR, TE, TEG, WEC, XEL

A.3 Daily trading information of 31 stocks

In Chapter 5, the 31 investigated stocks listed in Table A.2 are ranked and selected according to the average number of daily trades for each stock in 2008. Here, we have a trade if the trade sign in a time interval of one second is non-zero. We exclude the first and the last ten minutes from the intraday trading time. Except for the information of economic sector and the average number of daily trades for each stock, the other information, *i.e.* the average daily traded volume, and the exponents γ for the trade sign self-correlator $\Theta_{ii}(\tau)$, the passive cross-correlator $\Theta_i^{(p)}(\tau)$ and the active cross-correlator $\Theta_i^{(a)}(\tau)$, are also listed in Table A.2.

Table A.2: The average daily trading information and the γ values for each stock

Symbol	Sector	Average number of daily trades	Average daily traded volume ($\times 10^6$)	γ		
				for $\Theta_{ii}(\tau)$	for $\Theta_i^{(p)}(\tau)$	for $\Theta_i^{(a)}(\tau)$
AAPL	Information Technology	13415	13.27	1.36	0.71	0.83
JPM	Financials	10284	12.62	1.07	1.06	0.81
XOM	Energy	9708	7.79	1.19	1.50	0.95
BAC	Financials	9599	18.08	0.92	0.89	0.79
WFC	Financials	9040	12.43	0.90	0.88	0.81
MER	Financials	8823	9.20	0.97	0.96	0.79
C	Financials	8297	30.48	0.78	0.68	0.75
QCOM	Information Technology	8132	8.43	0.84	0.87	0.87
MS	Financials	7860	7.04	1.00	0.92	0.79
MSFT	Information Technology	7794	30.39	0.70	0.70	0.84
WMT	Consumer Staples	7438	6.28	0.91	0.94	0.88
CVX	Energy	7331	3.65	1.30	1.51	0.99
GS	Financials	7073	3.46	1.23	0.94	0.79
WB	Financials	6856	14.41	0.78	0.66	0.75
COP	Energy	6712	3.23	1.07	1.24	0.95
CSCO	Information Technology	6697	22.41	0.69	0.67	0.83
CHK	Energy	6603	4.50	0.94	0.92	0.91
INTC	Information Technology	6567	25.61	0.65	0.62	0.81
GE	Industrials	6475	16.67	0.78	0.73	0.83
HAL	Energy	6455	4.43	0.90	0.94	0.92
AMZN	Consumer Discretionary	6371	3.66	0.98	0.86	0.88
FCX	Materials	6308	3.08	1.11	1.11	0.91
T	Telecommunications Services	6239	6.64	0.83	0.84	0.86
USB	Financials	6078	4.49	0.88	0.88	0.82
HPQ	Information Technology	6056	4.47	0.86	0.88	0.88
AXP	Financials	6046	3.42	0.96	0.97	0.85
SLB	Energy	5952	2.56	1.20	1.30	0.98
AIG	Financials	5928	12.38	0.82	0.76	0.77
GILD	Health Care	5851	3.27	0.81	0.92	0.89
PG	Consumer Staples	5765	3.45	0.93	1.12	0.92
ORCL	Information Technology	5696	14.94	0.66	0.65	0.84

A.4 Daily trading information of 100 stocks

In Chapter 7, we select 100 stocks for the copula analysis. With the TAQ data set, we calculate the average number of daily trades for 496 available stocks from S&P 500 index in 2008. Here, we define a trade if the total volume, *i.e.*, the sum of all the unsigned volumes, in a time interval of one second is non-zero. We exclude the first and the last ten minutes from the intraday trading time. The first 100 stocks with the largest average number of daily trades are listed in Table A.3, which records in detail the information of symbols, economic sectors and the average numbers of daily trades for each stock.

Table A.3: The first 100 stocks with the largest average number of daily trades

Stocks	Sectors	Numbers	Stocks	Sectors	Numbers
C	Financials	98990.8	AMGN	HealthCare	21715.6
BAC	Financials	90648.7	SPLS	ConsumerDiscretionary	21528.3
AAPL	InformationTechnology	86242.5	SBUX	ConsumerDiscretionary	21420.5
MSFT	InformationTechnology	80399.8	GILD	HealthCare	20880.6
JPM	Financials	75825.8	FCX	Materials	20762.8
WFC	Financials	68118.3	SYMC	InformationTechnology	20490.1
INTC	InformationTechnology	63849.0	NCC	Financials	20029.7
GE	Industrials	61435.8	GLW	InformationTechnology	19865.2
CSCO	InformationTechnology	60952.6	DIS	ConsumerDiscretionary	19754.3
WB	Financials	60803.0	ADBE	InformationTechnology	19180.9
XOM	Energy	56978.8	TGT	ConsumerDiscretionary	19068.0
MER	Financials	55616.2	KO	ConsumerStaples	18812.8
AIG	Financials	48129.0	VLO	Energy	18770.7
QCOM	InformationTechnology	47234.5	F	ConsumerDiscretionary	18741.8
ORCL	InformationTechnology	45197.8	SLB	Energy	18712.4
MS	Financials	42930.3	SNDK	InformationTechnology	18464.7
YHOO	InformationTechnology	39279.2	ALTR	InformationTechnology	18444.2
WMT	ConsumerStaples	37852.5	XLNX	InformationTechnology	18187.2
DELL	InformationTechnology	36807.0	BMY	HealthCare	17949.7
T	TelecommunicationsServices	36013.4	SGP	HealthCare	17947.4
CMCSA	ConsumerDiscretionary	35446.6	DTV	ConsumerDiscretionary	17933.3
PFE	HealthCare	31997.7	RF	Financials	17566.6
NVDA	InformationTechnology	31618.3	MOT	InformationTechnology	17293.7
AMAT	InformationTechnology	31156.6	HCBK	Financials	17177.7
HD	ConsumerDiscretionary	30661.2	NTAP	InformationTechnology	17017.8
HAL	Energy	30160.6	XTO	Energy	16897.1
HPQ	InformationTechnology	29049.2	GOOG	InformationTechnology	16870.1
CHK	Energy	28869.6	MO	ConsumerStaples	16818.3
USB	Financials	28501.7	CVS	ConsumerStaples	16129.3
BRCM	InformationTechnology	28333.2	JAVA	InformationTechnology	15962.9
CVX	Energy	28211.5	BBBY	ConsumerDiscretionary	15786.4
EMC	InformationTechnology	27682.0	BK	Financials	15589.9
EBAY	InformationTechnology	27589.1	LLTC	InformationTechnology	15450.5
SCHW	Financials	25703.3	WFT	Energy	15316.3
AA	Materials	25148.4	MU	InformationTechnology	14973.1
TXN	InformationTechnology	24315.1	HBAN	Financials	14899.5
GS	Financials	24113.1	MCD	ConsumerDiscretionary	14896.6
COP	Energy	24010.7	COST	ConsumerStaples	14812.1
PG	ConsumerStaples	23998.2	UNH	HealthCare	14685.2
VZ	TelecommunicationsServices	23540.2	DOW	Materials	14684.0
AXP	Financials	23508.4	NBR	Energy	14642.2
AMZN	ConsumerDiscretionary	23213.6	COF	Financials	14577.6
FITB	Financials	23105.8	KFT	ConsumerStaples	14542.3
JNPR	InformationTechnology	22952.1	AMD	InformationTechnology	14516.8
GM	ConsumerDiscretionary	22379.9	GPS	ConsumerDiscretionary	14501.0
TWX	ConsumerDiscretionary	22075.0	OXY	Energy	14166.3
LOW	ConsumerDiscretionary	21933.8	CAT	Industrials	14003.3
JNJ	HealthCare	21906.1	M	ConsumerDiscretionary	13884.9

Table A.3: (continued)

Stocks	Sectors	Numbers	Stocks	Sectors	Numbers
MRK	HealthCare	21903.3	DD	Materials	13859.4
S	TelecommunicationsServices	21724.9	DHI	ConsumerDiscretionary	13810.6

Appendix **B**

Error estimation

Suppose we measured or numerically simulated a set of M data points $y(\tau_m)$ at positions $\tau_m,\ m = 1, \ldots, M$. We want to describe the data with a function $f(\tau)$ by fitting its M_P parameters. To assess the quality of the fit, the normalized χ^2 [19]

$$\chi^2 \ = \frac{1}{M - M_P} \sum_{m=1}^{M} \big(f(\tau_m) - y(\tau_m)\big)^2 \tag{B.1}$$

is used. Here, $M - M_P$ is referred to as the number of degrees of freedom. In our case, we have $M = 1000$, $M_P = 3$ for the fitting of trade sign cross–correlators in stock pairs, and $M = 34$.

Appendix C

Diffusion equation in two dimensions

Consider a particle moving in a flat two–dimensional space with coordinates (x, y) at time t. After a small and fixed time τ, this particle moves to the position $(x + u, y + v)$. The random increments u and v in the direction x and y, respectively, can be positive or negative and satisfies a normalized and symmetric distribution of marginal probability density,

$$\int_{-\infty}^{+\infty} q(u)du = 1 \qquad \text{and} \qquad q(-u) = q(u) \ , \tag{C.1}$$

$$\int_{-\infty}^{+\infty} q(v)dv = 1 \qquad \text{and} \qquad q(-v) = q(v) \ . \tag{C.2}$$

Their joint probability density distribution is also normalized to unity between $-\infty$ and $+\infty$,

$$\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} q(u,v)dudv = 1 \ . \tag{C.3}$$

According to the Eqs .(C.1)–(C.3), the random increments have following properties,

$$\langle uv\rangle = \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} uvq(u,v)dudv \ , \tag{C.4}$$

$$\langle u^n\rangle = \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} u^n q(u,v)dudv = \int_{-\infty}^{+\infty} u^n q(u)du \ , \tag{C.5}$$

$$\langle v^n\rangle = \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} v^n q(u,v)dudv = \int_{-\infty}^{+\infty} v^n q(u)du \ . \tag{C.6}$$

As the particle moves without any external driving force, the positive and negative increments in each direction have equal probability, which lead to

$$\langle u\rangle = 0 \qquad \text{and} \qquad \langle v\rangle = 0 \ . \tag{C.7}$$

We introduce the probability $p(x, y|t)dxdy$ to find the particle in the area element $dxdy$ at time t with the joint probability density $p(x, y|t)$. Now suppose the particle moves to the position (x', y') at time t. After the time increment τ, the probability density to find the particle in the new position (x, y) is

$$\begin{aligned} & p(x, y|t+\tau) \\ &= \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} dx'dy' p(x', y'|t) \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} dudv q(u, v)\delta(x-(x'+u))\delta(y-(y'+v)) \\ &= \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} dudv q(u, v) p(x-u, y-v|t) \ , \end{aligned} \tag{C.8}$$

where the δ functions $\delta(x-(x'+u))\delta(y-(y'+v))$ as the proper filter fix the x and y to $x'+u$ and $y'+v$, respectively. The random increments u and v during the time increment τ are assumed to be all very small. Thus the Eq. (C.8) can be derived as,

$$\begin{aligned} & p(x, y|t+\tau) + \tau\frac{\partial p(x, y|t)}{\partial t} \\ &= \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} dudv q(u, v)\left\{ p(x, y|t) - u\frac{\partial p(x, y|t)}{\partial x} - v\frac{\partial p(x, y|t)}{\partial y} \right. \\ & \left. + \frac{1}{2}\left[u^2\frac{\partial^2}{\partial x^2}p(x, y|t) + v^2\frac{\partial^2}{\partial y^2}p(x, y|t) + 2uv\frac{\partial}{\partial x}\frac{\partial}{\partial y}p(x, y|t)\right]\right\} \ . \end{aligned} \tag{C.9}$$

Employing the Eqs. (C.1)–(C.7), the last equation (C.9) becomes

$$\begin{aligned} \frac{\partial p(x, y|t)}{\partial t} &= \frac{\partial^2 p(x, y|t)}{\partial x^2}\frac{1}{2\tau}\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} u^2 q(u, v) dudv \\ &+ \frac{\partial^2 p(x, y|t)}{\partial y^2}\frac{1}{2\tau}\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} v^2 q(u, v) dudv \\ &+ \frac{\partial}{\partial x}\frac{\partial}{\partial y}p(x, y|t)\frac{1}{\tau}\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} uv q(u, v) dudv \qquad \text{(C.10)} \\ &= \frac{\langle u^2\rangle}{2\tau}\frac{\partial^2 p(x, y|t)}{\partial x^2} + \frac{\langle v^2\rangle}{2\tau}\frac{\partial^2 p(x, y|t)}{\partial y^2} + \frac{\langle uv\rangle}{\tau}\frac{\partial}{\partial x}\frac{\partial}{\partial y}p(x, y|t) \ . \end{aligned}$$

Here, the angular brackets indicate averages over the distribution of u and v. The last term in Eq. (C.11) is non–zero if the random increments are not independent. In general, to find the particle at the time t at the position $\vec{r}$, the diffusion equation can be written as [76]

$$\frac{\partial p(\vec{r}|t)}{\partial t} = \nabla\cdot\left(\hat{D}\nabla p(\vec{r}|t)\right) \ . \tag{C.11}$$

In homogeneous and anisotropic media, the diffusion tensor $\hat{D}$ is symmetric and depends on the direction. For a flat two–dimensional space, it is given by

$$\hat{D} = \left[\begin{array}{cc} \hat{D}_{xx} & \hat{D}_{xy} \\ \hat{D}_{yx} & \hat{D}_{yy} \end{array}\right] \ , \tag{C.12}$$

where $\hat{D}_{xy} = \hat{D}_{yx}$. Thus, the two–dimensional diffusion equation (C.11) turns into

$$\begin{aligned} \frac{\partial p(x,y|t)}{\partial t} &= \left[\begin{array}{cc} \frac{\partial}{\partial x} & \frac{\partial}{\partial y} \end{array} \right] \left[\begin{array}{cc} \hat{D}_{xx} & \hat{D}_{xy} \\ \hat{D}_{yx} & \hat{D}_{yy} \end{array} \right] \left[\begin{array}{c} \frac{\partial}{\partial x} p(x,y|t) \\ \frac{\partial}{\partial y} p(x,y|t) \end{array} \right] \\ &= \hat{D}_{xx} \frac{\partial^2 p(x,y|t)}{\partial x^2} + \hat{D}_{yy} \frac{\partial^2 p(x,y|t)}{\partial y^2} + 2\hat{D}_{xy} \frac{\partial}{\partial x} \frac{\partial}{\partial y} p(x,y|t) \ . \end{aligned} \tag{C.13}$$

Equations (C.11) and (C.13) coincide and allow for the identification,

$$\langle u^2 \rangle = 2\hat{D}_{xx}\tau \ , \quad \langle v^2 \rangle = 2\hat{D}_{yy}\tau \ , \quad \text{and} \quad \langle uv \rangle = 2\hat{D}_{xy}\tau \ . \tag{C.14}$$

For the Brownian motion, the diffusion coefficients $\hat{D}_{xx}$, $\hat{D}_{yy}$ and $\hat{D}_{xy}$ are constant.

List of Figures

1.1 The flash crash on May 6, 2010. 4
1.2 An example of order book. On the left hand side are the quotes to buy (bids), on the right side the quotes to sell (asks) 6
1.3 Snap shot of the order book before and after a buy market order with a volume of 1000 shares is executed 7
1.4 (a) Daily closing prices, (b) daily returns with $\Delta t = 1$ day, and (c) volatility estimated on a 40-days time window for Citigroup, Apple, and Goldman Sachs from January, 2000 to June, 2017. 9
1.5 (a) Conditional density $P(\Omega|\Sigma)$ versus the volume imbalance Ω for varying local noise intensity Σ, (b) order parameter ψ, *i.e.*, positions of the maxima of $P(\Omega|\Sigma)$, versus Σ. The figures are taken from Ref. [132]. 12

2.1 Comparisons between empirical and theoretical trade signs versus the physical time for AAPL during one minute. 21
2.2 A diagram for illustrating the comparisons of trade signs 22

3.1 Cross-response functions $R_{ij}(\tau)$ including $\varepsilon_j(t) = 0$ in 2008 versus time lag τ on a logarithmic scale (top panels). Corresponding trade sign cross-correlators $\Theta_{ij}(\tau)$ for different stock pairs on a doubly logarithmic scale, fit as dotted lines (bottom panels) 28
3.2 Cross-response functions $R_{ij}(\tau)$ excluding $\varepsilon_j(t) = 0$ in 2008 versus time lag τ on a logarithmic scale. Corresponding trade sign cross-correlators $\Theta_{ij}(\tau)$ for different stock pairs on a doubly logarithmic scale, fit as black dotted lines. 28
3.3 Probabilities $p_s(\tau_0)$ and $p_d(\tau_0)$ for the change of trade sign versus the time τ_0 without trading. 31
3.4 The three cross-responses and sign cross-correlators of stock pair (MSFT, AAPL) versus the physical time lag τ 32
3.5 Cross-response noise $\xi_{ij}(\tau)$ for different stock pairs during the year 2008 versus the time lag τ measured on a logarithmic scale. 33
3.6 Matrices of market response with entries $\rho_{ij}(\tau)$ for $i, j = 1, \ldots, 99$ at different time lags $\tau = 1, 2, 60, 300, 1800, 7200$ s in the year 2008. 35
3.7 Doubly averaged response functions $\overline{R}(\tau)$ in- and excluding $\varepsilon_j(t) = 0$ for the whole market in 2008 versus time lag τ on a logarithmic scale. 36

3.8 Probability distribution of the signed returns $u_{ij}(\tau) = r_i(t,\tau)\varepsilon_j(t)$ for the whole market in 2008, excluding $\varepsilon_j(t) = 0$, for $\tau = 30$ s 37

3.9 Probability distribution of signed return $u_{ij}(\tau)$ for physical time scale, where both i and j are AAPL in 2008 39

4.1 Passive and active cross-response functions $R_i^{(p)}(\tau)$ and $R_j^{(a)}(\tau)$ for $i, j =$ AAPL, GS, XOM in the year 2008 versus time lag τ on a logarithmic scale. Corresponding passive and active trade sign cross-correlators $\Theta_i^{(p)}(\tau)$ and $\Theta_j^{(a)}(\tau)$, fit as dotted lines). 46

4.2 Passive cross-response functions $R_i^{(p)}(\tau)$ of the stocks $i =$ AAPL, GS, XOM to ten different economic sectors in the year 2008 versus time lag τ on a logarithmic scale. 48

4.3 Active cross-response functions $R_j^{(a)}(\tau)$ of the stocks $j =$ AAPL, GS, XOM to ten different economic sectors in the year 2008 versus time lag τ on a logarithmic scale 48

4.4 The first fifteen stocks with the strongest passive and active cross-response functions $R_i^{(p)}(\tau)$ and $R_j^{(a)}(\tau)$ versus stock index i or j and time lags $\tau = 1$ s, 2 s, 60 s, and 300 s. The cross-response functions include $\varepsilon_j(t) = 0$. 49

4.5 The first fifteen stocks with the strongest passive and active cross-response functions $R_i^{(p)}(\tau)$ and $R_j^{(a)}(\tau)$ versus stock index i or j and time lags $\tau = 1$ s, 2 s, 60 s, and 300 s. The cross-response functions exclude $\varepsilon_j(t) = 0$. 49

4.6 Relation between active cross-response and average daily number of trades for the stocks j. 51

4.7 Comparisons of the self-responses for AAPL, GS and XOM with cross-responses to different stocks versus the time lag τ. Comparisons of self-responses, passive cross-responses, and active cross-responses for the same stocks versus the time lag τ. 53

4.8 Comparisons of the trade sign self-correlators for AAPL, GS and XOM with cross-correlators with different stocks versus the time lag τ. Comparisons of the self-correlators, passive cross-correlators, and active cross-correlators for the same stocks versus the time lag τ. 53

5.1 The average self-response over 31 stocks versus traded volumes at time lag $\tau = 1$ on a doubly logarithmic scale 68

5.2 The dependence of the average cross-responses on the traded volumes at time lag $\tau = 1$ on a doubly logarithmic scale 69

5.3 The probability density distribution of traded volumes for MSFT and other 30 stocks on a logarithmic scale 69

5.4 The dependence of average responses per share on the time lag τ in each region of traded volumes on a logarithmic scale 70

5.5 The sketch of the dependence of the price impact on the time 71

5.6 (a)The average cross-responses per share versus the time lag τ in Scenario I and I. (b) The impact functions versus the time lag τ in Scenario I and II 74

5.7 The passive responses versus the time lag τ in Scenario III with the weights $w = 0.10,\ 0.30,\ 0.50,\ 0.70$ and 0.90, respectively. 75

5.8 The active responses versus the time lag τ in Scenario III with the weights $w = 0.10,\ 0.30,\ 0.50,\ 0.70$ and 0.90, respectively. 76

5.9 The comparison of the theoretical results of the diffusion coefficient in Scenario III . . . 80
5.10 The comparison of the theoretical results of the diffusion coefficient in three scenarios . . . 80
5.11 The dependence of price impacts $G_{ii}(\tau)$, $G_i^{(p)}(\tau)$ and $G_i^{(a)}(\tau)$ on the time lag τ on a logarithmic scale . . . 82

6.1 Empirical and fitted results of impact functions of traded volumes . . . 94
6.2 Empirical and fitted results of impact functions of time lag . . . 95
6.3 Trading strategies with respect to the costs Ω_c . . . 98

7.1 The empirical copula density $\mathrm{cop}(q_1, q_2)$ of volume imbalances averaged over 4950 stock pairs $(k,\, l)$. . . . 104
7.2 Histograms of asymmetry values of copula densities for positive dependence $p(\alpha_{kl})$ (left) and for negative dependence $p(\beta_{kl})$ (right) with 4950 stock pairs $(k,\, l)$ . . . 105
7.3 The probability density distributions of volume imbalance conditioned on the local noise intensity (left). The cumulative probability density distributions of the numbers of data points conditioned on the local noise intensity (right) . . . 106
7.4 $\mathcal{K}$ copula density $\mathrm{cop}^{\mathcal{K}}_{\bar{c},N}(q_1, q_2)$ with $\bar{c} = 0.10$ and $N = 6.72$ (left, top). The error between the empirical copula density and the $\mathcal{K}$ copula density (right, top). Gaussian copula density $\mathrm{cop}^{G}_{\bar{c}}(q_1, q_2)$ with $\bar{c} = 0.10$ (left, bottom). The error between the empirical copula density and the Gaussian copula density (right, bottom). . . . 110
7.5 Empirical copula densities conditioned on the local noise intensity . . . 112
7.6 The influences of the local noise intensity on the copula density . . . 112
7.7 The dependence of $\gamma^{\mathcal{K}}_{\bar{c},N}$ on the correlation coefficient c (left). The dependence of $\gamma^{\mathcal{K}}_{\bar{c},N}$ on the parameter N (right) . . . 113
7.8 The contour of $P_{kl}(t)$ depending on $p_k^+(t)$ and $p_l^+(t)$. . . . 115
7.9 Histograms of asymmetry values of 4950 stock pairs $(k,\, l)$ for positive dependence . . . 116
7.10 Histograms of asymmetry values of 4950 stock pairs $(k,\, l)$ for negative dependence . . . 116

List of Tables

2.1 An example of Microsoft Corp. from the trade file in the TAQ data set . . 17
2.2 An example of Microsoft Corp. from the quote file in the TAQ data set . . 17
2.3 An example of Microsoft Corp. from TotalView-ITCH data set 17
2.4 Accuracy of trade sign classification 21

3.1 Company information 26
3.2 Fit parameters and normalized χ^2_{ij} for the trade sign cross-correlators. . . . 30
3.3 Comparisons of self- and cross-responses 41

4.1 Fit parameters and errors χ^2_i or χ^2_j for the average trade sign cross-correlators. 45

5.1 The fit parameters and errors for Scenarios I and II 74
5.2 The fit parameters and errors for Scenario III 75
5.3 The quantities in diffusion functions 78
5.4 The fit errors of $[|\langle D_i\rangle(\tau)|/\tau]^{1/2}$ in three scenarios 81
5.5 The parameters of impact functions in Scenario III with $w = 0.50$ 82

6.1 Parameters for impact functions 95

7.1 The skewness of distribution of asymmetries 107

A.1 Information of 99 stocks from ten economic sectors 123
A.2 The average daily trading information and the γ values for each stock . . . 126
A.3 The first 100 stocks with the largest average number of daily trades 127

Bibliography

[1] TAQ 3 User's Guide. *New York Stock Exchange, Inc.*, Version 1.1.9, 2008.

[2] A flash in the market. *The New York Times*, Oct. 2, 2010.

[3] F. Abergel, J.-P. Bouchaud, T. Foucault, M. Rosenbaum, and C.-A. Lehalle, editors. *Market microstructure: confronting many viewpoints.* John Wiley & Sons, 2012.

[4] I. Aldridge. *High-frequency trading: a practical guide to algorithmic strategies and trading systems*, volume 459. John Wiley & Sons, 2009.

[5] A. Alfonsi and J. I. Acevedo. Optimal execution and price manipulations in time-varying limit order books. *Applied Mathematical Finance*, 21(3):201–237, 2014.

[6] A. Alfonsi and P. Blanc. Dynamic optimal execution in a mixed-market-impact Hawkes price model. *Finance and Stochastics*, 20(1):183–218, 2016.

[7] R. Almgren and N. Chriss. Optimal execution of portfolio transactions. *Journal of Risk*, 3:5–40, 2001.

[8] R. Almgren, C. Thum, E. Hauptmann, and H. Li. Direct estimation of equity market impact. *Risk*, 18(7):58–62, 2005.

[9] Y. Amihud and H. Mendelson. Asset pricing and the bid-ask spread. *Journal of Financial Economics*, 17(2):223–249, 1986.

[10] P. W. Anderson, K. Arrow, and D. Pines, editors. *The economy as an evolving complex system.* Addison-Wesley, 1988.

[11] W. B. Arthur, S. N. Durlauf, and D. A. Lane, editors. *The economy as an evolving complex system II*, volume 28. Addison-Wesley, 1997.

[12] L. Bachelier. Théorie de la spéculation. *Annales Scientifiques de l'École Normale Superieure*, III(17):21–86, 1900.

[13] L. Bachelier. *Louis Bachelier's theory of speculation: the origins of modern finance.* Princeton University Press, 2011.

[14] P. Ball. Econophysics: culture crash. *Nature*, 441(7094):686–688, 2006.

[15] S. M. Bartram, F. Fehle, and D. G. Shrider. Does adverse selection affect bid–ask spreads for options? *Journal of Futures Markets*, 28(5):417–437, 2008.

[16] M. Benzaquen, I. Mastromatteo, Z. Eisler, and J.-P. Bouchaud. Dissecting cross-impact on stock markets: An empirical analysis. *Journal of Statistical Mechanics: Theory and Experiment*, 2017(2):023406, 2017.

[17] J. Beran. *Statistics for Long-Memory Processes*, volume 61. CRC Press, 1994.

[18] J. M. Berger and B. Mandelbrot. A new model for error clustering in telephone circuits. *IBM Journal of Research and Development*, 7(3):224–236, 1963.

[19] P. R. Bevington and D. K. Robinson. *Data Reduction and Error Analysis for the Physical Sciences.* McGraw–Hill, New York, 2003.

[20] S. Bikhchandani and S. Sharma. Herd behavior in financial markets. *IMF Economic Review*, 47(3):279–310, 2000.

[21] F. Black. Studies of stock price volatility changes. In *Proceedings of the 1976 Meetings of the American Statistical Association, Business and Economics Statistics Section*, pages 177–181, 1976.

[22] S. Bornholdt. Expectation bubbles in a spin model of markets: Intermittency from frustration across scales. *International Journal of Modern Physics C*, 12(05):667–674, 2001.

[23] J.-P. Bouchaud. Power laws in economics and finance: some ideas from physics. *Quantitative Finance*, 1(1):105–112, 2001.

[24] J.-P. Bouchaud. The subtle nature of financial random walks. *Chaos: An Interdisciplinary Journal of Nonlinear Science*, 15(2):026104, 2005.

[25] J.-P. Bouchaud. Price impact. In *Encyclopedia of quantitative finance.* John Wiley & Sons, 2010.

[26] J.-P. Bouchaud, J. D. Farmer, and F. Lillo. How markets slowly digest changes in supply and demand. In T. Hens and K. R. Schenk-Hoppé, editors, *Handbook of Financial Markets: Dynamics and Evolution*, pages 57–160. North-Holland, Elsevier, 2009.

[27] J.-P. Bouchaud, Y. Gefen, M. Potters, and M. Wyart. Fluctuations and response in financial markets: the subtle nature of ‘random’ price changes. *Quantitative Finance*, 4(2):176–190, 2004.

[28] J.-P. Bouchaud, J. Kockelkoren, and M. Potters. Random walks, liquidity molasses and critical response in financial markets. *Quantitative Finance*, 6(02):115–123, 2006.

[29] J.-P. Bouchaud and M. Potters. *Theory of Financial Risk and Derivative Pricing: From Statistical Physics to Risk Management.* Cambridge University Press, 2003.

[30] A. Boulatov, T. Hendershott, and D. Livdan. Informed trading and portfolio returns. *Review of Economic Studies*, 80(1):35–72, 2013.

[31] M. Buchanan. It's a (stylized) fact! *Nature Physics*, 8(1):3–3, 01 2012.

[32] M. Buchanan. What has econophysics ever done for us? *Nature Physics*, 9(6):317–317, 06 2013.

[33] M. A. Carlson. A brief history of the 1987 stock market crash with a discussion of the federal reserve response. *SSRN: https://ssrn.com/abstract=982615*, 2007.

[34] Á. Cartea, L. Gan, and S. Jaimungal. Liquidating Baskets of Co-Moving Assets. *SSRN: https://ssrn.com/abstract=2681309*, 2015.

[35] T. N. Cason and D. Friedman. Price formation in double auction markets. *Journal of Economic Dynamics and Control*, 20(8):1307–1337, 1996.

[36] A. Chakraborti, I. M. Toke, M. Patriarca, and F. Abergel. Econophysics review: I. Empirical facts. *Quantitative Finance*, 11(7):991–1012, 2011.

[37] A. Chakraborti, I. M. Toke, M. Patriarca, and F. Abergel. Econophysics review: II. Agent-based models. *Quantitative Finance*, 11(7):1013–1041, 2011.

[38] L. K. Chan and J. Lakonishok. Institutional trades and intraday stock price behavior. *Journal of Financial Economics*, 33(2):173–199, 1993.

[39] D. Chetalova, R. Schäfer, and T. Guhr. Zooming into market states. *Journal of Statistical Mechanics: Theory and Experiment*, 2015(1):P01029, 2015.

[40] D. Chetalova, T. A. Schmitt, R. Schäfer, and T. Guhr. Portfolio return distributions: sample statistics with stochastic correlations. *International Journal of Theoretical and Applied Finance*, 18(02):1550012, 2015.

[41] D. Chetalova, M. Wollschläger, and R. Schäfer. Dependence structure of market states. *Journal of Statistical Mechanics: Theory and Experiment*, 2015(8):P08012, 2015.

[42] C. Chiang and W.-C. Chiang. Reducing inventory costs by order splitting in the sole sourcing environment. *Journal of the Operational Research Society*, 47(3):446–456, 1996.

[43] C. Chiarella and G. Iori. A simulation analysis of the microstructure of double auction markets. *Quantitative Finance*, 2(5):346–353, 2002.

[44] T. Chordia, R. Roll, and A. Subrahmanyam. Commonality in liquidity. *Journal of Financial Economics*, 56(1):3–28, 2000.

[45] T. Chordia, R. Roll, and A. Subrahmanyam. Order imbalance, liquidity, and market returns. *Journal of Financial Economics*, 65(1):111–130, 2002.

[46] R. Cont. Empirical properties of asset returns: stylized facts and statistical issues. *Quantitative Finance*, 1(2):223–236, 2001.

[47] R. Cont. Volatility clustering in financial markets: empirical facts and agent-based models. In *Long memory in economics*, pages 289–309. Springer, 2007.

[48] R. Cont, M. Potters, and J.-P. Bouchaud. Scaling in stock market data: stable laws and beyond. *Scale Invariance and Beyond*, 7:75–85, 1997.

[49] G. Curato, J. Gatheral, and F. Lillo. Optimal execution with non-linear transient market impact. *Quantitative Finance*, 17(1):41–54, 2017.

[50] F. De Jong and B. Rindi. *The microstructure of financial markets.* Cambridge University Press, 2009.

[51] H. Demsetz. The Cost of Transacting. *The Quarterly Journal of Economics*, 82(1):33–53, 1968.

[52] Z. Eisler, J.-P. Bouchaud, and J. Kockelkoren. The price impact of order book events: market orders, limit orders and cancellations. *Quantitative Finance*, 12(9):1395–1419, 2012.

[53] P. Embrechts, C. Kluppelberg, and T. Mikosch. Modelling extremal events. *British Actuarial Journal*, 5(2):465–465, 1999.

[54] E. F. Fama. Efficient capital markets: A review of theory and empirical work. *The Journal of Finance*, 25(2):383–417, 1970.

[55] J. Farmer, A. Gerig, F. Lillo, and H. Waelbroeck. The market impact of large trading orders: Correlated order flow, asymmetric liquidity, and efficient prices. *Working paper*, 2009.

[56] J. D. Farmer and D. Foley. The economy needs agent-based modelling. *Nature*, 460(7256):685–686, 2009.

[57] J. D. Farmer, A. Gerig, F. Lillo, and S. Mike. Market efficiency and the long-memory of supply and demand: Is price impact variable and permanent or fixed and temporary? *Quantitative Finance*, 6(02):107–112, 2006.

[58] J. D. Farmer, A. Gerig, F. Lillo, and H. Waelbroeck. How efficiency shapes market impact. *Quantitative Finance*, 13(11):1743–1758, 2013.

[59] J. D. Farmer, L. Gillemot, F. Lillo, S. Mike, and A. Sen. What really causes large price changes? *Quantitative Finance*, 4(4):383–397, 2004.

[60] J. D. Farmer and F. Lillo. On the origin of power-law tails in price fluctuations. *Quantitative Finance*, 4(1):7–11, 2004.

[61] J. D. Farmer and N. Zamani. Mechanical vs. informational components of price impact. *The European Physical Journal B*, 55(2):189–200, 2007.

[62] X. Gabaix. Power laws in economics and finance. *Annual Review of Economics*, 1(1):255–294, 2009.

[63] X. Gabaix, P. Gopikrishnan, V. Plerou, and H. E. Stanley. A theory of power-law distributions in financial market fluctuations. *Nature*, 423(6937):267–270, 2003.

[64] K. Gangopadhyay. Interview with Eugene H. Stanley. *IIM Kozhikode Society & Management Review*, 2(2):73–78, 2013.

[65] J. Gatheral. No-dynamic-arbitrage and market impact. *Quantitative Finance*, 10(7):749–759, 2010.

[66] J. Gatheral and A. Schied. Dynamical models of market impact and algorithms for order execution. In J.-P. Fouque and J. A. Langsam, editors, *HANDBOOK ON SYSTEMIC RISK*, pages 579–599. Cambridge, 2013.

[67] J. Gatheral, A. Schied, and A. Slynko. Transient linear price impact and Fredholm integral equations. *Mathematical Finance*, 22(3):445–474, 2012.

[68] C. Genest, M. Gendron, and M. Bourdeau-Brien. The advent of copulas in finance. *The European Journal of Finance*, 15(7-8):609–618, 2009.

[69] A. Gerig. A theory for market impact: How order flow affects stock price. *arXiv:0804.3818*, 2008.

[70] D. K. Gode and S. Sunder. Allocative efficiency of markets with zero-intelligence traders: market as a partial substitute for individual rationality. *Journal of Political Economy*, 101(1):119–37, 1993.

[71] C. Gomes and H. Waelbroeck. Is market impact a measure of the information value of trades? Market response to liquidity vs. informed metaorders. *Quantitative Finance*, 15(5):773–793, 2015.

[72] P. Gopikrishnan, V. Plerou, X. Gabaix, and H. E. Stanley. Statistical properties of share volume traded in financial markets. *Physical Review E*, 62(4):R4493, 2000.

[73] A. Górski, S. Drożdż, and J. Speth. Financial multifractality and its subtleties: an example of DAX. *Physica A: Statistical Mechanics and its Applications*, 316(1):496–510, 2002.

[74] J. Grant. High-frequency trading: Up against a bandsaw. *Financial Times*, September 2, 2010.

[75] S. J. Grossman and J. E. Stiglitz. Information and competitive price systems. *The American Economic Review*, pages 246–253, 1976.

[76] D. Gupta. *Diffusion processes in advanced technological materials.* Springer Science & Business Media, 2010.

[77] P. Handa, R. A. Schwartz, and A. Tiwari. The ecology of an order-driven market. *The Journal of Portfolio Management*, 24(2):47–55, 1998.

[78] L. Harris. *Trading and exchanges: Market microstructure for practitioners.* Oxford University Press, USA, 2003.

[79] J. Hasbrouck. Trades, quotes, inventories, and information. *Journal of Financial Economics*, 22(2):229–252, 1988.

[80] J. Hasbrouck. *Empirical market microstructure: The institutions, economics, and econometrics of securities trading.* Oxford University Press, 2007.

[81] J. Hasbrouck and D. J. Seppi. Common factors in prices, order flows, and liquidity. *Journal of financial Economics*, 59(3):383–411, 2001.

[82] S. Havlin and D. Ben-Avraham. Diffusion in disordered media. *Advances in Physics*, 51(1):187–292, 2002.

[83] R. W. Holthausen, R. W. Leftwich, and D. Mayers. The effect of large block transactions on security prices: A cross-sectional analysis. *Journal of Financial Economics*, 19(2):237–267, 1987.

[84] C. Hopman. Do supply and demand drive stock prices? *Quantitative Finance*, 7(1):37–53, 2007.

[85] K. Hwang, J. Kang, and D. Ryu. Phase-transition behavior in the emerging market: Evidence from the KOSPI200 futures market. *International Review of Financial Analysis*, 19(1):35–46, 2010.

[86] S.-M. Jiang, S.-M. Cai, T. Zhou, and P.-L. Zhou. Note on two-phase phenomena in financial markets. *Chinese Physics Letters*, 25(6):2319, 2008.

[87] H. Joe. *Multivariate models and multivariate dependence concepts.* CRC Press, 1997.

[88] B. Johnson. *Algorithmic Trading & DMA: An introduction to direct access trading strategies.* 4Myeloma Press, 2010.

[89] A. Joulin, A. Lefevre, D. Grunberg, and J.-P. Bouchaud. Stock price jumps: news and volume play a minor role. *arXiv:0803.1769*, 2008.

[90] N. Kaldor. Capital accumulation and economic growth. In *The theory of capital*, pages 177–222. Springer, 1961.

[91] P. Kelle and E. A. Silver. Safety stock reduction by order splitting. *Naval Research Logistics (NRL)*, 37(5):725–743, 1990.

[92] A. Kempf and O. Korn. Market depth and order size. *Journal of Financial Markets*, 2(1):29–48, 1999.

[93] A. Kirilenko, A. S. Kyle, M. Samadi, and T. Tuzun. The flash crash: High-frequency trading in an electronic market. *The Journal of Finance*, 72(3):967–998, 2017.

[94] C. Krishnamurti. Introduction to market microstructure. In R. Vishwanath and C. Krishnamurti, editors, *Investment management: A modern guide to security analysis and stock selection.* Springer, 2009.

[95] A. S. Kyle. Continuous auctions and insider trading. *Econometrica: Journal of the Econometric Society*, 53(6):1315–1336, 1985.

[96] B. LeBaron. Agent-based computational finance. *Handbook of Computational Economics*, 2:1187–1233, 2006.

[97] C. Lee and M. J. Ready. Inferring trade direction from intraday data. *The Journal of Finance*, 46(2):733–746, 1991.

[98] F. Lillo and J. D. Farmer. The long memory of the efficient market. *Studies in Nonlinear Dynamics & Econometrics*, 8(3), 2004.

[99] F. Lillo, J. D. Farmer, and R. N. Mantegna. Econophysics: Master curve for price-impact function. *Nature*, 421(6919):129–130, 2003.

[100] F. Lillo, S. Mike, and J. D. Farmer. Theory for long memory in supply and demand. *Physical Review E*, 71(6):066122, 2005.

[101] G. Lim, S. Y. Kim, K. Kim, D.-I. Lee, and S.-B. Park. Dynamical mechanism of two-phase phenomena in financial markets. *Physica A: Statistical Mechanics and its Applications*, 386(1):253–258, 2007.

[102] A. W. Lo. The adaptive markets hypothesis. *The Journal of Portfolio Management*, 30(5):15–29, 2004.

[103] A. W. Lo and A. C. MacKinlay. An econometric analysis of nonsynchronous trading. *Journal of Econometrics*, 45(1):181–211, 1990.

[104] A. W. Lo and A. C. MacKinlay. When are contrarian profits due to stock market overreaction? *The Review of Financial Studies*, 3(2):175–205, 1990.

[105] T. Lux and M. Marchesi. Volatility clustering in financial markets: a microsimulation of interacting agents. *International Journal of Theoretical and Applied Finance*, 3(04):675–702, 2000.

[106] B. Mandelbrot. The variation of certain speculative prices. *The Journal of Business*, 36(4):394–419, 1963.

[107] B. B. Mandelbrot and J. W. Van Ness. Fractional Brownian motions, fractional noises and applications. *SIAM Review*, 10(4):422–437, 1968.

[108] R. N. Mantegna and H. E. Stanley. Scaling behaviour in the dynamics of an economic index. *Nature*, 376(6535):46, 1995.

[109] I. Mastromatteo, B. Tóth, and J.-P. Bouchaud. Agent-based models for latent liquidity and concave price impact. *Physical Review E*, 89(4):042805, 2014.

[110] K. Matia and K. Yamasaki. Statistical properties of demand fluctuation in the financial market. *Quantitative Finance*, 5(6):513–517, 2005.

[111] T. H. McInish and R. A. Wood. An analysis of intraday patterns in bid/ask spreads for nyse stocks. *The Journal of Finance*, 47(2):753–764, 1992.

[112] S. Mike and J. D. Farmer. An empirical behavioral model of liquidity and volatility. *Journal of Economic Dynamics and Control*, 32(1):200–234, 2008.

[113] R. S. Miller and G. Shorter. High frequency trading: Overview of recent developments. *Washington, DC: Congressional Research Service*, CRS Report for Congress, R44443, 2016.

[114] E. Moro, J. Vicente, L. G. Moyano, A. Gerig, J. D. Farmer, G. Vaglica, F. Lillo, and R. N. Mantegna. Market impact and trading profile of hidden orders in stock markets. *Physical Review E*, 80(6):066102, 2009.

[115] M. C. Münnix, T. Shimada, R. Schäfer, F. Leyvraz, T. H. Seligman, T. Guhr, and H. E. Stanley. Identifying States of a Financial Market. *Scientific Reports*, 2, 2012.

[116] N. Musmeci, T. Aste, and T. Di Matteo. Interplay between past market correlation structure changes and future volatility outbursts. *Scientific Reports*, 6, 2016.

[117] NASDAQ. https://en.wikipedia.org/wiki/nasdaq.

[118] NASDAQ's story. http://business.nasdaq.com/discover/nasdaq-story.

[119] R. B. Nelsen. *An introduction to copulas*. Springer Science & Business Media, 2007.

[120] A. A. Obizhaeva and J. Wang. Optimal trading strategy and supply/demand dynamics. *Journal of Financial Markets*, 16(1):1–32, 2013.

[121] M. O'hara. *Market microstructure theory*, volume 108. Blackwell Cambridge, MA, 1995.

[122] M. F. M. Osborne. Brownian motion in the stock market. *Operations Research*, 7:145–173, 1959.

[123] A. Pagan. The econometrics of financial markets. *Journal of Empirical Finance*, 3(1):15–102, 1996.

[124] P. Pasquariello and C. Vega. Strategic cross-trading in the US stock market. *Review of Finance*, 19:229–282, 2013.

[125] A. J. Patton. A review of copula models for economic time series. *Journal of Multivariate Analysis*, 110:4–18, 2012.

[126] E. Perrin, R. Harba, C. Berzin-Joseph, I. Iribarren, and A. Bonami. nth-order fractional Brownian motion and fractional Gaussian noises. *IEEE Transactions on Signal Processing*, 49(5):1049–1059, 2001.

[127] M. Phillips. Nasdaq: Here's our timeline of the flash crash. *The Wall Street Journal*, May 11, 2010.

[128] V. Plerou, P. Gopikrishnan, L. A. N. Amaral, X. Gabaix, and H. E. Stanley. Economic fluctuations and anomalous diffusion. *Physical Review E*, 62(3):R3023, 2000.

[129] V. Plerou, P. Gopikrishnan, X. Gabaix, and H. E. Stanley. Quantifying stock-price response to demand fluctuations. *Physical Review E*, 66(2):027104, 2002.

[130] V. Plerou, P. Gopikrishnan, X. Gabaix, and H. E. Stanley. On the origin of power-law fluctuations in stock prices. *Quantitative Finance*, 4(1):C11–C15, 2004.

[131] V. Plerou, P. Gopikrishnan, B. Rosenow, L. A. N. Amaral, T. Guhr, and H. E. Stanley. Random matrix approach to cross-correlations in financial data. *Physical Review E*, 65(6):066126, 2002.

[132] V. Plerou, P. Gopikrishnan, and H. E. Stanley. Econophysics: Two-phase behaviour of financial markets. *Nature*, 421(6919):130–130, 2003.

[133] V. Plerou, P. Gopikrishnan, and H. E. Stanley. Two phase behaviour and the distribution of volume. *Quantitative Finance*, 5(6):519–521, 2005.

[134] M. Potters and J.-P. Bouchaud. Comment on:"Two-phase behaviour of financial markets". *arXiv preprint cond-mat/0304514*, 2003.

[135] M. Potters and J.-P. Bouchaud. More statistical properties of order books and price impact. *Physica A: Statistical Mechanics and its Applications*, 324(1):133–140, 2003.

[136] D. Ryu. The effectiveness of the order-splitting strategy: an analysis of unique data. *Applied Economics Letters*, 19(6):541–549, 2012.

[137] D. Ryu. What types of investors generate the two-phase phenomenon? *Physica A: Statistical Mechanics and its Applications*, 392(23):5939–5946, 2013.

[138] D. S. Scharfstein and J. C. Stein. Herd behavior and investment. *The American Economic Review*, 80(3):465–479, 1990.

[139] A. B. Schmidt. *Financial markets and trading: an introduction to market microstructure and trading strategies*, volume 637. John Wiley & Sons, 2011.

[140] T. A. Schmitt, D. Chetalova, R. Schäfer, and T. Guhr. Non-stationarity in financial time series: Generic features and tail behavior. *EPL (Europhysics Letters)*, 103(5):58003, 2013.

[141] T. A. Schmitt, D. Chetalova, R. Schäfer, and T. Guhr. Credit risk and the instability of the financial system: An ensemble approach. *EPL (Europhysics Letters)*, 105(3):38004, 2014.

[142] T. A. Schmitt, R. Schäfer, M. C. Münnix, and T. Guhr. Microscopic understanding of heavy-tailed return distributions in an agent-based model. *EPL (Europhysics Letters)*, 100(3):38005, 2012.

[143] M. Schneider and F. Lillo. Cross-impact and no-dynamic-arbitrage. *SSRN: https://ssrn.com/abstract=2889029*, 2016.

[144] G. W. Schwert. Why does stock market volatility change over time? *The Journal of Finance*, 44(5):1115–1153, 1989.

[145] R. J. Shiller. Investor behavior in the october 1987 stock market crash: Survey evidence. *NBER Working Paper*, 2446, 1987.

[146] S. Sinha and S. Raghavendra. Phase transition and pattern formation in a model of collective choice dynamics. *SFI Working Paper*, 2004-09-028.

[147] S. Sinha and S. Raghavendra. Emergence of two-phase behavior in markets through interaction and learning in agents with bounded rationality. In *Practical Fruits of Econophysics*, pages 200–204. Springer, 2006.

[148] A. Sklar. Fonctions de répartition à n dimensions et leurs marges. *Publications de l'Institut de Statistique de l'Université de Paris*, 8:229–231, 1959.

[149] A. Sklar. Random variables, joint distribution functions, and copulas. *Kybernetika*, 9(6):449–460, 1973.

[150] A. Smith. *Wealth of nations*. JSTOR, 2005.

[151] E. Smith, J. D. Farmer, L. Gillemot, and S. Krishnamurthy. Statistical theory of the continuous double auction. *Quantitative Finance*, 3(6):481–514, 2003.

[152] L. Tabb, R. Iati, and A. Sussman. Us equity high frequency trading: Strategies, sizing and market structure. *TABB Group Report*, 2009.

[153] N. Torre. BARRA market Impact model handbook. *BARRA Inc., Berkeley*, 1997.

[154] B. Toth, Z. Eisler, F. Lillo, J. Kockelkoren, J.-P. Bouchaud, and J. D. Farmer. How does the market react to your order flow? *Quantitative Finance*, 12(7):1015–1024, 2012.

[155] B. Tóth, I. Palit, F. Lillo, and J. D. Farmer. Why is equity order flow so persistent? *Journal of Economic Dynamics and Control*, 51:218–239, 2015.

[156] Tradingphysics. http://www.tradingphysics.com.

[157] S. Wang. Trading strategies for stock pairs regarding to the cross–impact cost. *arXiv:1701.03098*, 2017.

[158] S. Wang and T. Guhr. Microscopic understanding of cross-responses between stocks: a two-component price impact model. *arXiv:1609.04890*, 2016.

[159] S. Wang and T. Guhr. Local fluctuations of the signed traded volumes and the dependencies of demands: a copula analysis. *arXiv:1706.09240*, 2017.

[160] S. Wang, R. Schäfer, and T. Guhr. Average cross-responses in correlated financial market. *The European Physical Journal B*, 89:207, 2016.

[161] S. Wang, R. Schäfer, and T. Guhr. Cross-response in correlated financial markets: individual stocks. *The European Physical Journal B*, 89:105, 2016.

[162] P. Weber and B. Rosenow. Large stock price changes: volume or liquidity? *Quantitative Finance*, 6(1):7–14, 2006.

[163] J. Whitman. The markets' wild ride. *Financial Post*, May 6, 2010.

[164] N. Wiener. Differential space. *Journal of Mathematics and Physics*, 2:131–174, 1923.

[165] M. Wollschläger and R. Schäfer. Impact of non-stationarity on estimating and modeling empirical copulas of daily stock returns. *Journal of Risk*, 19(1):1–23, 2016.

[166] M. Wyart, J.-P. Bouchaud, J. Kockelkoren, M. Potters, and M. Vettorazzo. Relation between bid–ask spread, impact and volatility in order-driven markets. *Quantitative Finance*, 8(1):41–57, 2008.

[167] B. Zheng, T. Qiu, and F. Ren. Two-phase phenomena, minority games, and herding models. *Physical Review E*, 69(4):046115, 2004.