Quantitative Methods in High-Frequency Financial Econometrics: Modeling Univariate and Multivariate Time Series

Zur Erlangung des akademischen Grades
eines Doktors der Wirtschaftswissenschaften
(Dr. rer. pol.)

von der Fakultät für
Wirtschaftswissenschaften
der Universität Karlsruhe (TH)

genehmigte

DISSERTATION

von

M. Sc. Wei Sun

Tag der mündlichen Prüfung: 30.11.2007

Referent: Prof. Dr. Svetlozar T. Rachev

Korreferent: Prof. Dr. Andreas Geyer-Schulz

Karlsruhe (2007)

Contents

List of Figures

List of Tables

Acknowledgments

I would like to acknowledge many people for helping me during my doctoral work. I would especially like to thank my advisor, Prof. Svetlozar Rachev, for his guidance of my academic career from beginning. Throughout my doctoral work he encouraged me to develop independent thinking and research skills. He continually stimulated my analytical thinking and greatly assisted me with scientific research.

I am very grateful for my external advisor from Yale University, USA, Prof. Frank J. Fabozzi, for his generous time and commitment. He greatly assisted me with scientific reasoning and writing.

I am also very grateful for having an exceptional doctoral committee and wish to thank Prof. Andreas Geyer-Schulz, Prof. Detlef Seese, and Prof. Christian Hipp for their continual support, constructive suggestion, discussion, and encouragement.

I wish to thank Prof. Stefan Mittnik at University of Munich, Germany and Prof. Gennady Samorodnitsky at Cornell University, USA for their valuable remarks and suggestions. I'd like to thank Prof. Yongmiao Hong at Cornell University, USA for organizing the Summer School of Advanced Econometrics in 2006, where I updated my knowledge of frontier research in econometrics. I'd like to thank Prof. Michele Donato Cifarelli and Prof. Pietro Muliere at Bocconi University, Italy for organizing the Summer School of Probability and Statistics in 2007, where I improved my advanced mathematical skills by attending lectures and doing exercises.

I am extremely grateful for the assistance, generosity, and advice I received from Dr. Torsten Lüdecke at the German Karlsruher Kapitalmarktdatabank (KKMDB) at University of Karlsruhe, Germany. His assistance made it possible for me to complete the empirical work based on high-frequency data from German equity markets. I owe a same note of gratitude to Dr. Petko Kalve at Monash University, Australia for providing SIRCA data.

I extend many thanks to my colleagues and friends, especially, Prof. Kuno Egle, Prof.

Georg Bol, Dr. Christian Menn, Dr. Stefan Trück, Dr. Stoyan Stoyanov, Dr. Markus Höchstötter, Dr. Young Shin Kim, Sebastian Kring, Theda Schmidt, Thomas Plum, and two computer hiwis, Lyuben Atanasov and Jens Buechele. I would like to thank Renate Krüger and Fakultätsbibliothek staff for helping me maintain and organize books and other materials.

Finally, I'd like to thank my parents. I'm especially grateful to my girlfriend, Ye Chen, for her patience and for helping me keep my life in proper perspective and balance.

This research was supported by grants from the Deutschen Forschungsgemeinschaft (DFG) (RA 861/6 and RA 861/8-2) and Chinese Government Award for Outstanding Ph.D Students Abroad 2006 (No. 2006-180).

Preface

…Once upon a time, there lives an old master and his pupils. They are working on a skill called "dragon slay" at the academy. One day, a monster came to their country. This hideous monster attacked the helpless villagers, burned their houses, and ate their grazing cattle. People called this monster "the dragon". The master was called in by the king for a solution. "Your Majesty," said the master, "although I have work on 'dragon slay' for my whole life, till now I have never seen a real dragon. We should first know if this monster is a 'real dragon' as described by the books. From the books, I learned how to make a 'dragon-slay sword'. To make this sword, I need several special metals. The problem is that the books do not tell me the proportion of metals used to cast the sword. Fortunately, I know how to tell if the sword is a real 'dragon-slay sword' by using a touchstone. Unfortunately, this touchstone can be found only in the depth of the ocean. I have no idea about the exact location of it. Moreover, I am too old to fight with a dragon using the real 'dragon-slay sword'. We need to find someone who is able to use it."

——————— *Wei Sun*

Chapter 1

Introduction

A plethora of research articles on the topic of high frequency data has been observed, some of which are of high interest for financial economists and practitioners (such as investors and fund managers). The former group believes that high frequency provides the micro-perspective that was lacking from the traditional analysis so far and is able to decipher the puzzles of empirical finance, since the microstructure theory has unveiled patterns and relationships that could not be uncovered at daily or weekly frequencies. Traditional financial studies mainly consider the prevailing linear paradigm and efficient market hypothesis that severely distort the perspective of how markets effectively operate. Empirical evidence points to the realization that linearity and the efficient market hypothesis have less power in the explanation of phenomena observed from financial markets. The latter group, who has been equipped with the insights about the micro-perspective of financial markets and movements of speculative prices, could exploit them in asset allocation or establishing trading models. A new body of financial study has started emerging as a result of the availability of financial data at frequencies higher than daily. Such a new body of financial study in this dissertation is defined as *high frequency financial econometrics*.

Besides developments of high frequency data analysis in the academic field, growing international trade and the wide seeking for profit opportunities drive the integration and globalization of the markets, which induce increased uncertainty and volatility across global markets. It is generally agreed that the more global the markets become, the finer the frequency at which the analysis should be conducted. The reason is straightforward. Volatility brings risk and risk encourages the introduction of derivatives which, although designed to head off a disaster, can hold their own dangers (see, for instance, the recent collapse of subprime mortgage market). The boom in derivatives, globalization, electronic trading, and multiplication of markets transforms so much not only the markets themselves but also the way of investors manage their funds. Complex portfolios are often comprised of foreign exchange, equity, interest rates, and derivatives, and tend to diversify their risks over many geographical regions and different industries. Volatility grows with such global portfolios, which pushes portfolio managers to monitor and evaluate their assets at substantially high frequencies. Although long-term performance still remains the overriding objective, the yardstick of evaluation has actually become smaller with respect to real-time information from different markets.

Performance of assets is tightly linked to and evaluated through the risk perspective. Although the idea and practice of evaluating risk and reward simultaneously is not new to the empirical finance literature, what is novel are the newly developed methods for risk evaluation. Meantime, the growing complexity and automation of the markets require greater use of advanced quantitative methods, by which risk (i.e., volatility) can be accurately predicted. Good performance is rewarded only if it is the result of adhering to given risk limits that are relatively accurate and practical. As a result, the coherent risk measures are increasingly paid attention to with the aim of maintaining real-time risk control. Value at Risk (VaR) and risk management unit (RMU) methodologies that are used by practitioners, but neither their use nor their analytical sophistication has the possibility to be uniformly accepted. As a result, the New Basel Capital Accord (Basel II) places emphasis on flexibility and sensitivity of risk management. Consequently, market operators, from traders to funds managers, are increasingly paid more attention to the risks they undertake. It is obvious that there exist certain links between trading risks and real-time high frequency information, but the potential of exploiting such links is possible only by analyzing high frequency data with the help of technological advances (e.g., a powerful computer).

Still, the analysis of high frequency data retains different kind of challenges for financial market analysts and researchers. The contributions of this dissertation are threefold. First, challenges of employing advanced quantitative methods in the study of high frequency data and sought for solution by practitioners have been stated under a unified field of high frequency financial econometrics based on mining high frequency financial data. Second, more sophisticated models (i.e., ARMA-GARCH model with fractional Gaussian noise and fractional stable noise) are proposed for analyzing stylized facts of high frequency data. Third, sophisticated models (i.e., copula-based ARMA-GARCH models with fractional Gaussian/stable noise) are proposed for analyzing co-movement of international financial markets with different correlation styles (i.e., correlation with symmetric and asymmetric volatility).

In Chapter 2, high-frequency financial data are introduced. The major content of this chapter is to discuss topics about mining high-frequency data, which include information extraction and knowledge discovery from high-frequency data, practical issues of dealing with high frequency data, computational data mining methods, statistical data mining methods, and criteria for evaluating data mining methods.

In Chapter 3, high-frequency financial econometrics is defined. The research fields of high-frequency financial econometrics typically touched are discussed, which cover mechanisms in economic settings for financial markets, market price formation, study of transparency, liquidity, and volatility of financial markets, and pattern recognition and some stylized facts (i.e., random durations, heavy-tailed distributional properties, autocorrelation, seasonality, clustering, and long-range dependence).

In Chapter 4, long-range dependence effect is discussed in detail. Two self-similar processes, i.e., fractional Gaussian noise and fractional stable noise are discussed in this chapter.

In Chapter 5, I empirically investigate the return distribution of 27 German DAX stocks using high-frequency data under two separate assumptions regarding the return generation process: (1) It does not follow a Gaussian distribution and (2) it does not follow a random walk. In the empirical study, I develop the ARMA-GARCH model with one of the typical self-similar processes, fractional stable noise. I empirically compare this process with several alternative distributional assumptions in either fractal form or i.i.d. form (i.e., normal distribution, fractional Gaussian noise, generalized extreme value distribution, generalized Pareto distribution, and stable distribution) for modeling German equity market volatility. The empirical results suggest that fractional stable noise dominates these alternative distributional assumptions both in in-sample modeling and out-of-sample forecasting. My findings suggest that models based on fractional stable noise perform better than models based on the Gaussian random walk, the fractional Gaussian noise, and the non-Gaussian stable random walk.

Several studies that have investigated a few stocks have found that the spacing between consecutive financial transactions (referred to as trade duration) tend to exhibit long-range dependence, heavy tailedness, and clustering. When considering irregularly spaced tick-by-tick time series data, time durations should be modeled. In Chapter 6, I empirically investigate whether a larger sample of stocks exhibits those characteristics. Based on the modeling mechanism of self-similar processes, in this chapter I empirically compare a stable distribution with fractional stable noise with several alternative distributional assumptions (lognormal distribution, fractional Gaussian noise, exponential distribution, and Weibull distribution) in modeling trade duration data. The empirical results suggest that fractional stable noise and stable distribution dominate these alternative distributional assumptions. Comparing goodness of fit in modeling trade duration data for stable distribution and fractional stable noise based on a procedure using bootstrap methods, I find that empirically the autoregressive conditional duration model with stable distribution fits better than other combinations, while fractional stable noise itself fits better for the time series of trade duration.

Analyzing equity market co-movements is important for risk diversification of an international portfolio. Copulas have several advantages compared to the linear correlation measure in modeling co-movement. In Chapter 7, I introduce a copula ARMA-GARCH model for analyzing the co-movement of international equity markets. The model is implemented with an ARMA-GARCH model for the marginal distributions and a copula for the joint distribution. After goodness of fit testing, I find that the Student's t copula ARMA(1,1)-GARCH(1,1) model with fractional Gaussian noise is superior to alternative models investigated in my study where I model the simultaneous co-movement of nine international equity market indexes. This model is also suitable for capturing the long-range dependence and tail dependence observed in international equity markets.

In order to modeling asymmetric correlation, in Chapter 8, I introduce a skewed Student's t copula ARMA-GARCH model for analyzing the co-movement of indexes in German equity markets. The model is implemented with an ARMA-GARCH model for the marginal distributions and a skewed Student's t copula for the joint distribution. After goodness of fit testing, we find that the skewed Student's t copula ARMA(1,1)-GARCH(1,1) model with Lévy fractional stable noise is superior to alternative models investigated in this study when modeling the simultaneous co-movement of six German equity market indexes. This model is also suitable for capturing the long-range dependence, tail dependence, asymmetric correlation observed in German equity markets.

In Chapter 9, I perform three empirical studies in high-frequency financial econometrics, i.e., volatility prediction, computing Value at Risk (VaR), and short-term portfolio selection. In this chapter, I show that the neural network has the advantages in prediction. I compare several models in computing VaR and show that parametric models using stable distribution and fractional stable noise is better than those who use normal and fractional Gaussian noise. I propose a new methodology of selecting portfolios in a hypothetic real-time financial market, i.e., selecting stocks by their clustered risk measures. The result shows that accurately clustering the risk measures can generate a better portfolio performance.

I conclude this dissertation in Chapter 10 and point out future researches.

Chapter 2

Mining High-Frequency Financial Data

Many companies and scientific communities have been utilizing data mining technology as more and more success stories of data mining technology become known. Most banks and financial institutions offer a wide variety of financial services such as issuing bank checks, savings and loans, mortgages, customer transactions, investment services, credit cards, and derivatives. Financial data collected in the banking and financial industry are often relatively complete and informative that facilitates systematic data analysis and data mining to improve the management quality and competitiveness of financial institutions. In the banking industry, data mining is used heavily in the areas of modeling and predicting credit fraud, evaluating performance, controlling risks, analyzing profitability, and supporting direct-marketing campaigns. In the financial markets, data mining is used to forecast stock and commodity prices, to trade options and finanical derivaties, to rank bonds, to manage portfolio, and to analyze mergers and acquisitions. After the new Basel II Acord, data mining has been used for forecasting financial disasters. Due to the confidential policies, it is not easy to find reports about financial companies who use data mining in their bussiness. An easy way to imply the companies who use data mining technology in their bussiness is to look at the US Government Agency SEC[1] reports of some of the data mining companies who sell their produc. One finds some big customers such as Bank of America, First USA bank, Daiwa Securities, LBS Capital Management, and U.S. Bancorp. Kantardzic (2003) gives several examples of financial companies who utilizes data mining technology, for example, the "FAIS" system developed by the Financial Crimes Enforcement Network (FINCEN) of the US Treasury Department for detecting potential money-laundering activities from a large number of big cash transactions, the credit card-attrition model developed by the Mellon bank based on the IBM Intelligent Miner, and the data mining techniques

[1] Securities and Exchange Commission or SEC is a U.S. regulatory commission established by Congress in 1934 with primary responsibility for enforcing the federal securities laws and regulating the securities industry/stock market. More detialed information can be found from www.sec.gov.

used by Capital One Financial Croup based on Oracle datawarehouse.

With employing electronic trading and order routing systems in financial markets, an enormous quantity of trading data in electronic form is now available. A complete data set of transactions recorded and their associated characteristics such as transaction time, transaction price, posted bid/ask prices, and volumes are provided. These data are gathered at the ultimate frequency level in the financial markets and usually referred to as (ultra-)high-frequency data. Looming atop a wide variety of events in financial markets menace profiles of ever-growing mountains of high-frequency data after employing electronic trading system. These mountains grew as a result of great demand for efficient technology to generate, collect, and store such digital data. Unfortunately, the size of a high-frequency database usually is greater than the main memory of a normal PC. Without advanced computational techniques to help us analyze such accumulated datasets, we risk missing useful information that the data offer.

Faced with massive high-frequency data sets, people might find traditional approaches in statistics and pattern recognition fail to explain the data. For example, a statistical analysis package usually assumes data can be “loaded” into memory for future manipulation. But it is no longer true when we are dealing with large high-frequency data sets. Several problems have be pointed out, for example: What happens when the size of a dataset exceeds the capacity of a computer’s main memory? What happens if the database is on a remote server that does not allow a näive scan of the data? How do we sample effectively if we are not permitted to query for a stratified sample because the relevant fields are not indexed? What if the data set is in a multitude of tables and can only be accessed via some hierarchically structured set of fields? What if the relations are sparse (not all fields are defined or even applicable to any fixed subset of the data)? What can we do if the subsets of the database are collected at different scales without index for the differences? How do we identify a statistical model with a large number of variables? Therefore, new approaches, techniques, and solutions have to be developed to enable analysis based on large high-frequency databases. Data mining techniques may enable us in analyzing massive high-frequency databases with respect to solve the questions posted above.

Data mining has been defined as “the nontrivial extraction of implicit, previously unknown, and potentially useful information from data (see Frawley et al. 1992)” and “the science of extracting useful information from large data sets or databases (see Hand et al. 2001)”. Johnson and Wichern (2002) classified data mining problems into five categories, i.e., classification, prediction, association, clustering, and description. Data mining involves sorting through large amounts of data and picking out relevant information from them. It is increasingly used in business research and financial analysis. Data mining is a process requiring a sequence of steps:

(1) Define the problem and identify the objectives. (2) Gather and prepare the appropriate data. (3) Explore the data for suspected associations, unanticipated characteristics, and obvious anomalies with robust methods. (4) Clean the data and perform necessary variable transformation in the appropriate way. (5) Establish a rigth model right to explain the information discovered from the data.

What is the difference between data mining methods and statistical modeling? It is true that data mining methods and statistical modeling share a same subset. The major differences are (1) data mining methods are designed for massive data set but statistical modeling and inference is usually based on limited samples and (2) data mining methods are more robust (model-free or assumption-free) but statistical modeling are often model-based with some preliminary assumptions.

In this chapter, we are going to outline the potential data mining techniques used in high-frequency financial econometrics. The structure of this paper is: in Section 2, we introduce the basic structure of the high frequency financial database. In Section 3, we will talk about the research scope of high-frequency financial econometrics by introducing the latest relavent studies. In Section 4, some topics about information extraction and knowledge discovery will be discussed. Two principles of data mining methods, i.e., computational data mining and statistical data mining are introduced in Sections 5 and 6 respectively. In Section 7, several methods for evaluating data mining methods are reported. We summarize in Section 8.

2.1 High-Frequency Financial Data

There is no standardization of the term intra-daily data adopted by researchers in the market microstructure area. In this literature, there are several descriptions for intra-daily data. Hasbrouck (1996) mentions microstructure data or microstructure time series. Engle (2000) uses the term ultra-high frequency data to describe the ultimate level of disaggregation of financial time series. Alexander (2001) describes high-frequency data as real time tick data. Gourieroux and Jasiak (2001) use the expression tick-by-tick data while Tsay (2002) writes “high frequency data are observations taken at fine time intervals.” In this article, the term intra-daily data will be used interchangeably with the other terms used by market structure researchers as identified above.

As the full record of every movement in financial markets, intra-daily data offer the researcher or analyst a large sample size that increases statistical confidence. The data can reveal events in the financial market that are impossible to identify with low frequency data. While in some sense, intra-daily data might be regarded as the microscope used for studying financial markets, these data have broader interest in econometric modeling and market microstructure research.

Extremely large amounts of high frequency data have been generated by the expanding financial market with the implementation of powerful computer systems designed for electronic trading. The electronic trading systems fulfill vital tasks on stock markets such as maintaining an electronic order book ranked according to the goodness of price and time entry, entering new incoming buy and sell orders into the order book, matching automatically the buy and sell orders, reporting the recorded trade characteristics and releasing information either in real time or in historical data files (Gourieroux and Jasiak 2001). Performance of such tasks has been recorded in a continuous way that creates the high frequency data. Figure 2.1 shows one example of high frequency transaction data of German Münchener Rückversicherung on DAX[2].

843002	20040930	17:26:20.40	77.3500	676.00	V	DE0008430026
843002	20040930	17:26:20.63	77.3500	122.00	V	DE0008430026
843002	20040930	17:26:22.42	77.3500	474.00	V	DE0008430026
843002	20040930	17:26:59.05	77.3500	50.00	V	DE0008430026
843002	20040930	17:27:02.27	77.3500	150.00	V	DE0008430026
843002	20040930	17:27:37.05	77.3800	262.00	V	DE0008430026
843002	20040930	17:27:39.25	77.3900	115.00	V	DE0008430026
843002	20040930	17:27:39.75	77.4200	1505.00	V	DE0008430026
843002	20040930	17:27:46.73	77.5000	348.00	V	DE0008430026
843002	20040930	17:27:47.44	77.5000	1252.00	V	DE0008430026
843002	20040930	17:27:51.11	77.4300	115.00	V	DE0008430026
843002	20040930	17:27:53.78	77.5200	108.00	V	DE0008430026
843002	20040930	17:28:08.07	77.5100	150.00	V	DE0008430026
843002	20040930	17:28:50.17	77.4900	100.00	V	DE0008430026
843002	20040930	17:28:54.80	77.4900	100.00	V	DE0008430026
843002	20040930	17:29:16.51	77.5200	1657.00	V	DE0008430026
843002	20040930	17:29:37.09	77.5200	382.00	V	DE0008430026
843002	20040930	17:29:37.09	77.5500	618.00	V	DE0008430026
843002	20040930	17:29:52.58	77.5500	687.00	V	DE0008430026
843002	20040930	17:35:21.97	77.5600	94864.00	SA	DE0008430026

Figure 2.1: High Frequency Transaction Data of Münchener Rückversicherung on DAX. Source: KKMDB.

The first position consisting of 6 numbers stands for the stock trading code on German DAX. It is followed by the date in format of YYYYMMDD (Y=year, M=month, D=day). Next entry is the time stamp in format of HH:MM:SS.ss (H=hour, M=munite, S=second, s=second/100). After the time stamp, price is entered into the table with 11 digits. The following digits represent the traded volume. After the traded volume, there is the price indicator. The last column in the table is the international trading code. DE stands for Germany. There are essentially three time series arising from these records: trading prices, volumes and trading times which yield inter-trade durations. It has to be emphasized that data on transactions prices do not satisfy standard assumptions used in the theory of asset pricing since trading

[2]The data is from German Karlsruher Kapitalmarktdatabank (KKMDB)

prices do not exist in continuous time but are separated by unequal (irregularly spaced) time intervals, and for trading data, the transaction price is not unique but changeable.

Another type of high frequency data is the order book data. Figure 2.2 shows one example of the order book data. This display ranks the bids and asks of the 39 Nasdaq market makers and electronic communications networks that were providing quotes of Microsoft in October 5, 2001 at 11:26 Eastern time. The bids are ranked from highest to lowest, and the asks are from lowest to highest. It reports that Microsoft stock last traded at 55.97 dollar, 0.47 dollar down from the previous close, the market is currently 55.97 dollar bid and 55.98 dollar offered. The two rows near the bottom present order indications (OI) that are displayed for Bridge clients. The unit of size is one hundred shares.

us;MSFT Microsoft Corp *# Common Stock
\- 55.9700 DN 0.4700 (Q) 11.25 H 56.6400 L 54.9400 V 15,130,100

E/MM	TIME	-BID-	SIZE(H)	E/MM	TIME	-ASK-	SIZE(H)
(Q)	11.25 +	55.9700	9	(Q)	11.25	55.9800	19
INCA	11.25	55.9700	9	ISLD	11.25	55.9800	19
ISLD	11.25	55.9600	6	REDI	11.25	55.9800	2
BTRD	11.25	55.9500	6	INCA	11.25	55.9900	54
FCAP	11.24	55.9400	1	GSCO	11.25	55.9900	10
ARCA	11.25	55.9100	4	BRUT	11.25	55.9900	2
SBSH	11.25	55.8900	1	ARCA	11.25	55.9900	1
NITE	11.24	55.8800	29	SCHB	11.25	56	20
SCHB	11.25	55.8800	1	PERT	11.24	56	11
GSCO	11.25	55.8700	10	BEST	11.24	56	10
MLCO	11.23	55.8500	10	NITE	11.24	56	6
MSCO	11.25	55.8500	10	JEFF	11.10	56	1
REDI	11.25	55.8300	10	RAJA	11.18	56	1
MKXT	11.25	55.8100	5	ABNA	11.03	56.0300	1
MONT	11.19	55.8000	10	FBCO	11.24	56.0300	1

OI BUY : GSCO MLCO BEST
OI SELL : FBCO GSCO CANT

MSFT/q/Pg2 05-Oct-01 11:26 NYC (c)BRIDGE

Figure 2.2: Nasdaq Level II Quotation Montage for Microsoft Common Stock. Source: Harris (2003, page 15).

2.2 Information Extraction and Knowledge Discovery

Data mining is the most well-known application for information extraction and knowledge discovery, which describes the process of searching large volumes of data for patterns that can be considered knowledge about the data. In this section, several important concepts are introduced based on the process of information extraction and knowledge discovery.

2.2.1 Informative Data

Not all data can offer useful information for problem solving. Task based data mining only extracts information that can be extracted from large datasets, i.e., solving problems based only on mining informative data. Informative data has following features: precision, completeness, relevance, uniqueness, and coherence.

Precision

The principal requirement for the informative data is precision. Precision means the data values recorded conform to the real situation, i.e., the data tell nothing but the truth. Precise data should have a correctly reported value at the reported time. False definition of variables, misalignments, and data errors will significantly reduce the precision of the data and consequently reduce the informative reliability.

False definition of variables refers to give certain data values a wrong index, i.e., given $\{x_r\}$ and $\{y_s\}$, representing two sequences of measurements of the variable x and y with respect to characteristics r and s, false definition refers to illustrate the data for $\{x_r\}$ and $\{y_s\}$ with the name of $\{x_s\}$ and $\{y_r\}$.

Given $\{x_k\}$, the sequence of measurements of the variable x, which is ordered by the strictly increasing index sequency k, taking values from 1 to N inclusively. Misalignments means replacing $\{x_k\}$ with $\{x_{k'}\}$ where $\{k'\}$ with values from 1 to N is any sequence other than the desired index sequence $\{k\}$. Pearson (2005) gives four possible types of $\{k'\}$ for $\{k\}$, i.e., permutation, missing values with extension, duplication, and shift with duplication.

Dacorogna et al. (2000) separate the errors into two classes, i.e., human errors and system errors. Human errors are errors directly caused by human data contributors and system errors are caused by computer systems, their interaction, and failures. They point out that there are unintentional errors and intentional errors caused by human data contributors. Typing errors are the typical unintentional errors and dummy ticks produced just for technical testing is typically intentional.

Completeness

An extremely common anomaly in large datasets is missing data, which refers to data values that should be present in the dataset but for various reasons are absent. The common cause of missing data is measurement system failures, which can happen in manual or automated data collection systems. Two situations should be pointed out. One is the total system failure and the other is the practical system failure. Total system failure refers to the situation that all data values that should be recorded are actually missing at given measurement times, whereas practical system failure refers to some data values are recorded but other data values are not. Completeness generally requires that there is no missing data in the dataset at any given measurement time. Pearson (2005) points out the ignorable missing data and non-ignorable missing data. The ignorable missing data corresponds to the omission of a randomly selected subset of data values from a certain dataset. A random subset of missing data is only ignorable for statistical analysis under the further assumption of independence. Conversely, non-ignorable missing data corresponds to systematically missing data values from a dataset. He argues that the variability of results computed from N data values usually decreases with increasing N, therefore the effect of ignorable missing data is only an increase in this variability relative to the results obtained from a complete dataset. He also thinks that the consequence of non-ignorable missing data often leads to significant biases into the results. The ignorable missing data causes less negative effects when the data is sampled randomly and the number of the missing values is a small percentage of the whole sample size. Unfortunately, the missing values for high-frequency financial data are usually non-ignorable. Pearson (2005) provides an illustration based on the sample autocorrelation function to support the non-ignorability for time series data.

Relevance

Relevance is a criterion based on certain target problems. The data used for problem solving must be relevant to that problem. The higher the level of relevance between the data and the problem, the more efficient the solution that can be reached by data mining. It is often difficult to find the proxy variables or instrument variables in the analysis, particularly in inference, when the actually relevant variables cannot be obtained from the dataset. There are two main requirements for using a proxy: (1) the proxy or instrument variable must be correlated with the unavailable variables, and (2) the proxy or instrument must not add noise to the analysis (i.e., the proxy should not be correlated with other variables obtained for analysis and should not suffer from the same problem as the original predicting variable).

Uniqueness

Uniqueness requires the value for each unit for a given characteristic variable must be unique. For example, the transaction price of Volkswagen AG in Frankfurt Exchange at 16:18:16.32 in 16 September, 2004 is 32.55 Euro, this price must be unique. Dacorogna et al. (2001) introduce two types of data error in high-frequency data, i.e., repeated ticks and tick copying. Some data contributors let their computer repeat the last tick in more or less regular time intervals and some data contributors employ computers to copy and re-send the ticks of other contributors. Dacorogna et al. (2001) point out the harmfulness (1) if the old ticks are repeated thousands of times with high frequency, it obstructs the validation of the few good ticks, and (2) copying ticks can obstruct a clear identification.

Coherence

Coherence here refers to the dataset has no (1) misalignment for each observation under the same variable, (2) misdefined variables, (3) mislabeled samples, and (3) incoherent scaling for each variable.

Misalignments are data anomalies particularly observed in large datasets. Suppose $\{x_k^1\}$ and $\{x_k^2\}$ stand for two sequences of measurements of the variable x ordered by a strictly increasing index sequence k. The identity of the index series k usually has less meaning, but the fact that x_k^1 and x_k^2 have the same subscript value k means that these two data observations represent values of x^1 and x^2 under the same conditions, which means, the data values x_k^1 and x_k^2 are two distinct measurable attributes of the same variable. When the misalignment happens, the desired index sequence $\{k\}$ will be replaced by some other sequence $\{k'\}$. Pearson (2005) illustrates following example with many different forms of misalignment:

- (correct sequence)$\{k\} = \{1, 2, 3, 4, 5, 6, 7, 8, 9, 10\}$;
- (permutation)$\{k'\} = \{2, 7, 4, 5, 1, 9, 3, 6, 10, 8\}$;
- (missing values with extension)$\{k'\} = \{1, 2, 3, 4, 7, 8, 9, 10, 11, 12\}$;
- (duplication)$\{k'\} = \{1, 1, 2, 2, 3, 3, 4, 4, 5, 5\}$;
- (shift with duplication)$\{k'\} = \{1, 2, 3, 3, 4, 5, 6, 7, 8, 9\}$.

Misdefined variables are often met when analyzing large dataset whose components are gathered from different locations at different time. Suppose $\{x_k^i\}$, $i = 1, 2, 3, \ldots, N$ stand for i samples with the same measurements of a series of variable x_k ordered by a strictly increasing index sequence k, that is, the identity of the index series k has meaning now and $x_1, x_2, \ldots, x_k$ are elements in the variable space that describe the N samples. When the variables are misdefined, the desired index sequence $\{k\}$ will be replaced by some other sequence $\{k'\}$. The usually observed form of misdefined variable is variable interchange.

Mislabeled samples are the samples classified incorrectly. For example, the data sampled from the New York Stock Exchange is labeled as the data from Chicago, or the return data of

IBM is labeled as the returns of Microsoft. This kind of data anomaly can be avoided if the data manager takes good care of it. Incoherent scaling refers to the existence of different measures for the same variable. For example, there are some prices recorded in the return series.

2.2.2 Data Quality

In practice, massive datasets usually contain various types of anomalous records that might complicate the analysis problem and invalidate the results obtained with standard analysis procedures without indicating anything wrong. Particularly, the prevalence of outliers, missing data, and misalignments are the anomalies often encountered in information extraction and knowledge discovery from large datasets. Pearson (2005) points out two ideas focusing on dealing with imperfect data: data pretreatment and analytical validation. Data pretreatment is (1) to detect the outliers of various types, (2) to form the treatment strategies once the outliers are indicated, (3) to detect other types of data anomalies by using simple preliminary analyses, and (4) to detect and exclude the noninformative variables when analyzing the data. Analytical validation is to evaluate the assessment of results. The essential idea is to be utilized as "generalized sensitivity analysis (GSA)" based on the following logic: "A 'good' data analysis result should be insensitive to small changes in either the methods or the datasets on which the analysis is based (see Pearson 2005, page 177)". This is more or less connected with the statistical robustness which is going to be discussed later. The question naturally proposed here is that what is a "good" data characterization (see Pearson (2005))?

Pearson (2005) considers six alternative viewpoints from which "goodness" can be judged, i.e., (1) predictable qualitative behavior, (2) ease of interpretation, (3) appropriateness to the application, (4) historical acceptance, (5) availability, and (6) computational complexity. Pearson (2005) also points out that these criteria are often in conflict and it is important to balance them in practice. In addition, he argues that there is rarely a unique "best" approach to measure data quality, therefore, several different approaches should be compared in order to select a reasonable compromise to a particular task in practice.

2.2.3 Aggregation

Being a full record of transactions and their associated characteristics, intra-daily data represent the ultimate level of frequency at which trading data are collected. The salient feature of such data is that they are fundamentally irregularly spaced. It is necessary to distinguish intra-daily data from high frequency data because the former are irregularly spaced while the latter are sometimes spaced by aggregating to a finer fixed-time interval. In order to clarify the time interval, it is useful to refer to the data by its associated time interval. For example, if the raw intra-daily data have been aggregated to create an equally-spaced time series, say five minutes interval, then the return series is referred to as the “5 minutes intra-daily data”. In order to clarify the characteristic of the interval between data points, Dacorogna et al. (2001) propose employing a definition of homogeneous and inhomogeneous time series. The irregularly spaced time series is called an *inhomogeneous time series* while the equally spaced one is called a *homogeneous time series*. One can aggregate inhomogeneous time series data up to a specified fixed time interval in order to obtain a corresponding homogeneous version. Naturally, such aggregation will either lose information or create noise, or both. For example, if the observations in the original data are more than that in the aggregated one, some information will be lost. If the observations in the original data are much less than that in the aggregated one, noise will be generated. Data interpolation is required to create aggregated time series. Obviously, different interpolation methods lead to different results. This leads not only to the loss of information but also the creation of noise due to introducing errors in the interpolation process. Engle and Russel (1998) argue that in the aggregation of tick-by-tick data to some fixed time interval, if a short time interval is selected, there will be many intervals having no new information and, as a result, heteroskedasticity will be introduced; and if a wide interval is chosen, microstructure features might be missing. Therefore, it is reasonable to keep the data at the ultimate frequency level (see Aït-Sahalia et al. (2005)).

Standard econometric techniques are based on homogeneous time series analysis. Applying analytical methods of homogeneous time series to inhomogeneous time series may produce unreliability. That is the dichotomy in intra-daily data analysis: a researcher can retain the full information without creating noise but is challenged by the burden of technical complexity. Sometimes, it is not always necessary to retain the ultimate frequency level. In those instances, aggregating inhomogeneous intra-daily data to a relatively lower but still comparably higher frequency level of a homogeneous time series is needed. Wasserfallen and Zimmermann (1995) show two interpolation methods: linear interpolation and previous-tick interpolation. Given an inhomogeneous time series with times t_i and values $\varphi_i = \varphi(t_i)$, the index i identifies the irregularly spaced sequence. The target homogeneous time series is given at times $t_0 + j\Delta t$ with fixed time interval Δt starting at t_0. The index j identifies the regularly spaced sequence. The time $t_0 + j\Delta t$ is bounded by t_I and t_{I+1} as follows:

$$I = \max(\, i \,|t_i \leq t_0 + j\Delta t) \tag{2.0-1}$$

$$t_I \leq t_0 + j\Delta t > t_{I+1} \tag{2.0-2}$$

Data will be interpolated between t_I and t_{I+1}. The linear interpolation shows that

$$\varphi(t_0 + j\Delta t) \;=\; \varphi_I \;+\; \frac{t_0 + j\Delta t - t_I}{t_{I+1} - t_I}\,(\varphi_{I+1} - \varphi_I) \tag{2.0-3}$$

and previous-tick interpolation shows that

$$\varphi(t_0 + j\Delta t) \;=\; \varphi_I \tag{2.0-4}$$

Dacorogna et al. (2001) point out that linear interpolation relies on the future information whereas previous-tick interpolation is based on the information already known. Müller et al. (1990) suggest that linear interpolation is an appropriate method for independent and identically distributed (i.i.d.) increments stochastic processes.

More advanced techniques have been adopted by some researchers in order to find sufficient statistical properties of data but at the same time retaining the inhomogeniety of time series. Zumbach and Müller (2001), for example, propose a convolution operator to transform the original inhomogeneous time series to a new inhomogeneous time series in order to get more sophisticated quantities. Newly developed techniques such as the wavelet method have been adopted to analyze intra-daily data. For example, Gençay et al. (2001, 2002) employed a wavelet multiscaling method to remove intra-daily seasonality in five-minute intra-daily data of foreign exchange.

As mentioned in Section 2, intra-daily data exhibit daily patterns. Several methods of data adjusting have been adopted in empirical analysis in order to remove such patterns (see, Engle and Russell (1998), Veredas et al. (2002), Bauwens and Giot (2000, 2003), and Bauwens and Veredas (2004)).

2.2.4 Data Cleaning

In order to improve the quality of data, data cleaning is required to detect and remove errors and inconsistencies from the data set. Data quality problems are the result of misspelling during data entry, missing information, and other kinds of data invalidity. There are two types of error: human errors and computer system errors (see Dacorogna et al. (2001)). When multiple data sources must to be integrated, for instance, pooling the data in each trading day together for one year, the need for data cleaning increases significantly. The reason is that the sources often contain redundant data in different representations. High quality data analysis requires access to accurate and consistent data. Consequently, consolidating different data representations and eliminating duplicate information become necessary.

The object of data cleaning is the time series of transaction information. Usually, the transaction information is “quote” or “tick” information. Each quote or tick in the intra-daily data set contains a time stamp, an identification code, and variables of the tick, such as bid/ask price (and volume), trade price (and volume), and locations. Intra-daily data might contain several errors that should be specially treated. Decimal errors occur when the quoting software uses cache memory so that it fails to change a decimal digit of the quote. Test tick errors are caused by data managers' testing operation of sending artificial ticks to the system to check the sensitivity of recording. Repeated ticks are caused by data managers' test operation of letting the system repeat the last tick in the specified time intervals. Some errors occur when data managers copy the data or when a scaling problem occurs, see Dacorogna et al. (2001).

Coval and Shumway (2001) illustrate the existence of occasionally incorrect identification of the exact time. They use the data cleaning method to ensure that the time stamps on each tick were accurate and scaled to the second level. They introduce a method of summing up variables from 31 seconds past one minute to 30 seconds past the next minute to aggregate the tick to the minute level. Some detailed methods used in data cleaning are discussed in Dacorogna et al. (2001).

2.2.5 Data Snooping

Data snooping, sometimes being referred as data dredging or data fishing, is the inappropriate search for "statistically significant" relationships in large quantities of data. It is the "dark side" of data mining. This activity was explored as "specification searches", "multiple comparisons", or "overfitting" in statistics, but that term is now in widespread use with an essentially positive meaning, so the pejorative term *data dredging* is now used instead.

Conventional statistical procedure is first to formulate a research hypothesis, and collect relevant data, then carry out a statistical significance test to check if the results could be due to the effects of chance. Only the invariant patterns presented in the population can be regarded as statistically significant. The major point is that every hypothesis must be independently tested with observations that was not used in constructing the hypothesis. The reason is that every data set must contain some variant patterns that are not be present in the population investigated, or simply vanish when increasing the sample size sufficiently large. If the hypothesis is not tested on a different data set sampled from the same population, it is likely that the patterns found are not consistent patterns.

It is important to realize that significance tests do not protect against data snooping. When testing a data set on which the hypothesis is known to be true, the data set is no longer a representative data set but a biased sample, and any resulting significance levels are completely spurious. Ioannidis (2004) points out several corollaries leading to false research findings. For example, (1) the smaller the studies conducted, (2) the smaller the effect sizes, (3) the greater the number and the lesser the selection of tested relationships, and (4) the greater the flexibility in designs, definitions, outcomes, and analytical modes can lead to false research findings. Lo and MacKinlay (1990) indicate (1) tests of theory-motivated models, for which the empirical evidence is lacking, are likely to be biased least by data snooping and (2) tests of data-driven models, for which the theoretical motivation is lacking, are most susceptible to data snooping.

More emphasis should be placed on thinking and modeling, without blatant a *priori* data dredging and so much inference. There are many cases where the *post hoc* results are not misleading (not spurious) and exploration of the data is a useful thing to do. However, more caution is needed in accepting exploratory results as if they were somewhat confirmatory (see Lo and MacKinlay (1990)).

A variety of approaches to correct for the statistical effects of searching large model spaces is proposed in the literature. The major remedies are including:

1. *New data and cross-validation*: A common approach is to obtain new data or to divide an existing sample into two or more subsamples, using one subsample to select a small number of models and the other subsamples to obtain unbiased scores (see, for example, Kohavi 1995).

2. *Adjustments to significance tests*: To correct for the multiple comparison effects, several mathematical adjustments have been made to statistical significance tests, for example, Sidak and Bonferroni adjustments. These have been explored in detail for experimental design (see, for example, Hochberg and Tamhane 1987).

3. *Resampling and randomization methods*: Many of the most successful approaches are based on computationally intensive techniques such as resampling and randomization to increase the statistical significance (see, for example, White 1997).

2.2.6 Pattern Recognition

The task of pattern recognition in high-frequency data mining is to classify data (patterns) based on either a priori knowledge or on statistical information extracted from the patterns (see Duda et al. (2001)). The patterns to be classified are usually groups of measurements or observations, defining points in an appropriate multidimensional space. A complete pattern recognition system consists of (1) a sensor that gathers the observations to be classified or described, (2) a feature extraction process that computes numeric or symbolic information from the observations, and (3) a classification or description scheme that does the actual job of classifying or describing observations, relying on the extracted features. Duda et al. (2001) design a pattern recognition system with different operation components, i.e., sensing, segmentation and grouping, feature extraction, classification, and post-processing.

The input to a pattern recognition system is often some kind of a transducer. Duda et al. (2001) point out that the difficulty of the problem usually depends on the characteristics and limitations of the transducer. Sensing could be regarded as a design of the sensors that gathers the observations for pattern recognition. The key point is the quality of the sensors, i.e., the data quality.

The goal of segmentation and grouping is to simplify and/or change the representation of data into something that is more meaningful and easier to analyze. Segmentation and grouping is the process dividing a dataset into distinct subsets (segments/groups) that can be characterized in the same way or have similar features. Because each segment/group is fairly homogeneous in their characteristics, they are likely to react similarly to a given method for analyses.

Feature extraction involves simplifying the amount of resources required to describe a large set of data accurately. When performing an analysis of a complex dataset, one of the major problems is raised by the number of variables involved. Analyzing a large number of variables generally requires a large amount of memory and computation power. In practice, it is important to improve computation power by constructing combinations of the variables to get around these problems while still describing the data with sufficient accuracy (see Duda et al. (2001)). Duda et al. (2001, page 11) point out that "the traditional goal of the feature extractor is to characterize an object to be recognized by measurements whose values are very similar for objects in the same category, and very different for objects in different categories", which "leads to the idea of seeking distinguishing features that are invariant to irrelevant transformations of the input".

Duda et al. (2001) think that the major concern of pattern recognition is the design of a classifier which assigns the objective to a proper category by using the feature vector provided by a feature extractor in the stage of feature extraction. The abstraction provided by the feature vector representation of the input data enables such classification. The difficulty of the classification goes with the variability in the feature values for objects in the same category relative to the difference between feature values for objects in different categories. Advanced quantitative methods are needed for the classification process.

Post-processing is the stage in which the classification is evaluated. The simplest measure of classifier performance is the classification error rate, i.e., the percentage of new patterns that are assigned to the wrong category, see Duda et al. (2001). In general, a classification with minimum classification error rate is preferred. But sometimes the minimum error rate classification requires costs. It is important to balance the risk of high cost and minimum error rate.

2.3 Computational Data Mining

The major data mining methodologies which do not necessarily require a probability model are going to be discussed in this section. In the literature, many of these methodologies were developed from the fields of computer science and employed in solving data mining problems later. Because of their main origin in computer science, I agree with Giudici (2003), that calls the methods presented in this section "computational methods for data mining".

2.3.1 Cluster Analysis

Cluster analysis might be the most well-known descriptive data mining method. Tryon (1939) introduced the Cluster Analysis (CA) that encompasses different algorithms and methods for grouping objects with similarity into categories. Cluster analysis is an exploratory data analysis tool which aims at sorting different objects into groups in a way that the degree of association between two objects is maximal if they belong to the same group and minimal otherwise, see Tabachnick et al. (2000).

There are n observations with p values (variables) given by the following matrix:

$$C_{n\times p} = \begin{array}{c} \\ R_1 \\ R_2 \\ \vdots \\ R_n \end{array} \begin{array}{c} \begin{array}{cccc} c_1 & c_2 & \cdots & c_p \end{array} \\ \begin{bmatrix} c_{11} & c_{12} & \cdots & c_{1p} \\ c_{21} & c_{22} & \cdots & c_{2p} \\ \vdots & \vdots & & \vdots \\ c_{n1} & c_{n2} & \cdots & c_{np} \end{bmatrix} \end{array}$$

in which, R_n stands for the n-th observation, c_p for p-th variable, and $c_{ij}, (i = 1, 2, \cdots, n; j = 1, 2, \cdots, p)$ is the element of i-th observation and j-th variable. The objective of cluster analysis is to cluster the observations into internally homogeneous groups that are heterogeneous from group to group, i.e., externally separable among groups.

Two types of methods, hierarchical and non-hierarchical methods, are usually employed. Hierarchical methods allow us to generate a succession of groupings (also referred as partitions or clusters) with a number of clusters from 1 to n, starting from the simplest, where all observations are separated and belong to a unique cluster. The number of clusters is not preset and based on the nature of the data itself. Non-hierarchical methods allow us to separate the n units directly into a number of previously defined clusters (see Giudici 2003 and Johnson and Wichern 2002).

Giudici (2003) and Johnson and Wichern (2002) outline the hierarchical clustering algorithm as follows:

1. *Initialization*: Start with n clusters, each containing a single entity and an $n \times n$ symmetric matrix of distances $D = \{d_{ij}\}$.
2. *Selection*: The two "nearest" clusters are selected, in terms of the distance initially fixed, for example, the Euclidean distance.
3. *Updating*: the number of clusters is updated (to $n-1$) through the union, in a unique cluster, of the two groups selected in step 2. For example, if the two clusters selected in step 2 are R_t and R_s, in this step, R_t and R_s are merged as R_{st}. Update the entries in the distance matrix D by (1) deleting the rows and column corresponding to clusters R_t and R_s and (2) adding a row and column giving the distance between cluster R_{st} and the remaining clusters.
4. *Repetition*: Repeat Setps 2 and 3 a total of $n-1$ times. Record the identity of clusters that are merged and the levels (distance) at which the mergers take place.
5. *End*: Terminate the procedure when all the elements will be in a unique cluster.

There are some of different clustering methods applied in step 3. Some methods require only the distance matrix D and some require the distance matrix plus the original data matrix. The linkage methods require only the distance matrix and the method of the centroid and Ward's method require the data matrix as well as the distance matrix (see Giudici 2003 and Johnson and Wichern 2002).

Single linkage

The distance between two groups (R_{st} and any other cluster R_u) is defined as $d_{(rs),u} = \min\{d_{ru}, d_{su}\}$. Here the quantities d_{ru}, d_{su} are the distances between the nearest neighbors of clusters R_r and R_u and clusters R_s and R_u, respectively.

Complete linkage

The distance between two groups (R_{st} and any other cluster R_u) is defined as $d_{(rs),u} = \max\{d_{ru}, d_{su}\}$. Here the quantities d_{ru}, d_{su} are the distances between the most distant members of clusters R_r and R_u and clusters R_s and R_u, respectively.

Average linkage

The distance between two groups (R_{st} and any other cluster R_u) is defined as

$$d_{(rs),u} = (\sum_l \sum_m d_{lm})/N_{rs}N_u,$$

where d_{lm} is the distance between object l in the cluster (R_{rs}) and object m in the cluster R_u, and N_{rs} and N_u are the number of items in clusters R_{rs} and R_u, respectively.

Method of the centroid

The distance between two groups R_s and R_t having N_s and N_t elements respectively is defined as the distance between the respective centroids (for example, the means), μ_s and μ_t: $d_{rs} = d_{\mu_r \mu_s}$. The original data matrix is required in order to calculate the centroid of a group of observations. When processing the hierarchical clustering steps as discussed above, the distances with respect to the centroids of the two previous clusters will be replaced by the distances with respect to the centroid of the new cluster. The centroid of the new cluster is obtained from the weighted average, i.e., $(\mu_r N_r + \mu_s N_s)/(N_r + N_s)$. Giudici (2003) indicates that the average linkage method considers the average of the distance between the observations of each of the two groups, whereas the centroid method computes the centroid of each group then measures the distance between the centroids.

Ward's Method

Ward's method is to cluster the groups that have maximum internal cohesion and maximum external separation. This method does not require the distance matrix to be calculated from the original data matrix. The total deviance T of the p variables is divided in two parts: the deviance within the groups W and the deviance between the groups B, so $T = W + B$. Given a partition into g groups with respect to p variables, T, W, and B are defined as follows: (1) $T = \sum_{s=1}^{p} \sum_{r=1}^{n} (c_{rs} - \bar{c}_s)^2$, (2) $W = \sum_{k=1}^{g} W_k$ where W_k represents the deviance of the p variables in the kth gourp (number n_k and centroid $\bar{c}_k = [\bar{c}_{1k}, \cdots, \bar{c}_{pk}]'$), described by $W_k = \sum_{s=1}^{p} \sum_{r=1}^{n_k} (c_{rs} - \bar{c}_{sk})^2$, and (3) $B = \sum_{s=1}^{p} \sum_{k=1}^{g} n_k (\bar{c}_{sk} - \bar{c}_s)^2$. Groups are joined so that the increase in W is smaller and the increase in B is larger, which achieves the possibility of greatest internal cohesion and external separation. Giudici (2003) indicates that Ward's method can be interpreted as a variant of the centroid method without considering the distance matrix. Johnson and Wichern (2002) point out that Ward's method expects that the clusters of multivariate observations are to be roughly elliptically distributed.

The non-hierarchical methods of clustering obtain one partition of the n observations in g

groups ($g < n$) with a previous defined number of clusters g (g also can be determined as part of the clustering procedures). For any given value of g, based on which it is used to classify the n observations, a non-hierarchical algorithm classifies each of the observations only on the selection criterion (usually by means of an objective function). Since non-hierarchical methods do not need to determine the distance matrix and the original data matrix need not to be stored during the computer run, non-hierarchical methods can be applied to much larger data matrix than can hierarchical methods (see Johnson and Wichern 2002). In general, a non-hierarchical algorithm contains the following steps (see Giudici 2003):

1. Choose the number of groups g and an initial clustering of the n units in that number of groups.

2. Evaluate the "transfer" of each observation from the initial group to another group with the aim of maximizing the internal cohesion of the groups.

3. Repeat Step 2 until a stopping rule is satisfied.

The most well-known non-hierarchical clustering method is the k-means method (where k is the number of groups decided in advance, same as g in this Section) proposed by MacQueen (1967). This method is to assign each observation to the cluster having the nearest centroid (for example, the mean). The algorithm is given as follows:

1. Partition the observations into k initial clusters.

2. Proceed through the list of observations, assigning an observation to the cluster whose centroid is nearest[3]. Recompute the centroid for the cluster receiving the new observation and for the cluster losing the observation.

3. Repeat Step 2 until no more reassignments can be done.

Another way to process the k-means method is to start with a partition of all observations into k predefined groups in Step 1, and specify k initial centroids (so-called the "seed points" or simply "seed") and then proceed to Step 2. Johnson and Wichern (2002) indicate that the final assignment of observations to clusters will be dependent on the initial partition or the initially selected seed points. Experience shows that with the first reallocation step, most major changes in assignment occur. Discussions of other nonhierarchical clustering methods can be referred to Everitt (1993).

[3]Distance is usually computed using Euclidean distance with either standardized or unstandardized observations, see Johnson and Wichern (2002).

Johnson and Wichern (2002) list the major drawbacks of nonhierarchical clustering procedures based on the arguments of the preliminary k: (1) If two or more seed points inadvertently lie within a single cluster, their resulting clusters will be poorly differentiated. (2) The occurrence of an outlier might produce at least one group with very disperse observations. (3) By previously knowing the population consisting of k groups, the sampling method might be such that data from the group with less possibility to observe do not appear in the sample, therefore, the clusters are nonsensical. As empirical solution to reduce the negative influence of such drawbacks of the nonhierarchical clustering procedures, one can run the algorithm for several choices before finding the specified k.

2.3.2 K-Nearest-Neighbour Method

In data mining, the k-nearest neighbor method (k-NN) is used for classifying objects based on closest observations in the feature space. k-NN method is a type of *lazy learning*[4] where the function is only locally approximated and all computation is deferred until classification. It might be used for regression analysis.

The k-nearest neighbor algorithm is to minimize the distance between the query instance and the training samples in order to determine the K-nearest neighbors. After the K nearest neighbors are gathered, the simple majority of these K-nearest neighbors can be taken as the prediction of the query instance. Usually, the Euclidean distance or Mahalanobis distance are employed. The Euclidean distance between two points $P = (p_1, \dots, p_n)$ and $Q = (q_1, \dots, q_n)$ is defined as

$$D_E(P, Q) = \sqrt{\sum_{i=1}^{n} (p_i - q_i)^2}.$$

The Mahalanobis distance (see Mahalanobis 1936) from a group of values with mean $\mu = (\mu_1, \mu_2, \dots, \mu_p)^T$ and covariance matrix Σ for a multivariate vector $x = (x_1, x_2, \dots, x_p)^T$ is defined as

$$D_M(x) = \sqrt{(x - \mu)^T \Sigma^{-1} (x - \mu)}.$$

The Mahalanobis distance is based on correlations between variables by which different patterns can be identified and analyzed. It is a useful way of determining similarity/dissimilarity of

[4]In artificial intelligence, lazy learning is a learning method in which generalization beyond the training data is delayed until a query is made to the system. The opposed concept is *eager learning* that is a learning method in which the system tries to construct a general, input independent target function during training of the system.

variables. It differs from Euclidean distance in that it takes into account the correlations of the data set and is scale-invariant (i.e., not dependent on the scale of measurements). When the correlation structure of variables is unclear, the Mahalanobis distance measure is likely to be the most appropriate due to its scale-invariant property. If the covariance matrix is the identity matrix, the Mahalanobis distance reduces to the Euclidean distance (see Hair et al. (2006)).

The k-nearest neighbours algorithm needs to determine the value of parameter K (i.e., the number of nearest neighbors). The best choice of k depends solely on the data. In general, the larger the values of k are, the less the noise effect on the classification. Unfortunately, the larger the values of k are, the less the distinction between classes boundaries. There are various heuristic techniques that support the selection of a good k, for example, cross-validation. In addition, one special case where the class is predicted to be the class of the closest training sample (i.e., $k = 1$) is called the nearest neighbor algorithm (see Giudici 2003).

2.3.3 Neural Networks

A neural network (NN), sometimes called an "artificial neural network" (ANN), is a quantitative model based on biological neural networks. A neural network is an adaptive system that allows changes of its structure based on external or internal information that flows through the network when monitoring the data generating process. Opposed to the k-nearest neighbor method, a neural network is a type of *eager learning* method. In more practical terms neural networks are nonlinear statistical data modeling tools used to characterize highly complex and convoluted relationships between inputs and outputs or to find correlation patterns in financial data. In the financial markets, the NN models have been used for various tasks. Kantardzic (2003) reveals that Daiwa Securities, NEC Corporation, Carl & Associates, LBS Capital Management, Walkrich Investment Advisors, and O' Sallivan Brothers Investments are the financial companies who utilize neural network technology for data mining.

Like the linear and polynomial approximation methods, a neural network relates a set of input variables x_i, $i = 1, \ldots, k$ to a set of one or more output variables y_j, $j = 1, \ldots, k$. A neural networks are essentially mathematical models defining a function $f : X \rightarrow Y$. Each type of NN model corresponds to a class of such functions (see McNelis (2005)). The difference between a NN model and other approximation methods is that the NN take advantage of one or more hidden layers, in which the input variables are transformed by a special function known as a logistic or logsigmoid transformation, that is, the function $f(x)$ is a composition of other functions $g_i(x)$ that can further be defined as a composition of other functions. Functions

$f(x)$ and $g_i(x)$ are composed of a set of elementary computational units called neurons[5], which are connected together through weighted connections. These units are organized in layers so that every neuron in a layer is exclusively connected to the neurons of the proceeding layer and the subsequent layer. Every neuron represents an autonomous computational unit and receives inputs as a series of signals that dictate its activation. All the input signals reach the neuron simultaneously and the neurons can receive more than one input signal. Following the activation coming from input signals, the neurons produce the output signals. Every input signal is associated with a connection weight that determines the relative importance of the input signals in generating the final impulse transmitted by the neuron.

Formally, the algorithm mentioned above can be expressed as follows:

$$n_{k,t} = w_{k,0} + \sum_{i=1}^{i^*} w_{k,i}\, x_{i,t} \tag{2.0-5}$$

$$N_{k,t} = G(n_{k,t}) \tag{2.0-6}$$

$$y_t = \gamma_0 + \sum_{k=1}^{k^*} \gamma_k\, N_{k,t} \tag{2.0-7}$$

where $G(\cdot)$ represents the activation function and $N_{k,t}$ stands for the neurons. In these system, there are i^* input variables x and k^* neurons. A linear combination of these input variables observed at time t, i.e., $x_{i,t}$, $i = 1, \ldots, i^*$, with the coefficient vector (i.e., a set of input weights) $w_{k,i}$, $i = 1, \ldots, i^*$, and a constant term $w_{k,0}$ form the variable $n_{k,t}$. This variable $n_{k,t}$ is transformed by the activation function $G(\cdot)$ to a neuron $N_{k,t}$ at time (or observation) t. The set of k^* neurons at time (or observation) index t are combined in a linear way with the coefficient vector γ_k, $k = 1, \ldots, k^*$ and added to a constant term γ_0 to form the output value y_t at time t.

In defining a NN model, the activiation function $G(\cdot)$ is typically one of the elements to specify. Giudici (2003) summarizes three commonly employed types of activiation functions: linear, stepwise, and signoidal. A linear activation function is defined by

$$G(n_{k,t}) = \alpha + \beta\, n_{k,t}$$

where α and β are real constants. When $\alpha = 0$ and $\beta = 1$ a particular function called identity function, which is usually used for the model requiring the output of a neuron to be exactly equal to its level of activation, is defined. Giudici (2003) points out that due to the similarity between the linear activation function and the linear regression model, a linear regression model

[5]Sometimes such elementary computational units are called nodes, neurodes, units, or processing elements (PEs).

can be regarded as a simple type of neural network. A stepwise activation function is defined as

$$G(n_{k,t}) = \begin{cases} \alpha & n_{k,t} \geq \theta_k \\ \beta & n_{k,t} < \theta_k. \end{cases}$$

The activation function can assume only two values according to if or not the potential exceeds the threshold θ_k. When $\alpha = 0$, $\beta = 1$, and $\theta_k = 0$, the activation function is the sign function that takes value 0 if the potential is negative and value 1 if the potential is positive. A sigmoidal (or S-shaped) activation function is the most used one in practice. It is nonlinear and easily differentiable. It is defined by

$$G(n_{k,t}) = \frac{1}{1 + e^{-\alpha n_{k,t}}}$$

where α is a positive parameter that regulates the slope of the function. This kind of activation function produces only positive output and the domain of the function is $[0, 1]$.

The neurons of a NN model are organized in layers. There are three types of layers: input, output, and hidden. The input layer receives information only from the external information, i.e., an explanatory variable x_i. There is no calculation performed for the input layer. The input layer only transmits information to the next level. The output layer only produces the final results sent by the network to the outside of the system, i.e., response variable y_j. Between the input and output layer there can be one or more intermediate layers, called hidden layers since they are not directly connected with the external information. Giudici (2003) points out that the architecture of a NN model refers to the network's organization: (1) The number of layers. (2) The number of neurons belonging to each layer. (3) The manner in which the neurons are connected. (4) Direction of flow for the computation.

Different information flows lead to different types of network. The NN models can by divided into to types based on the information flow: feedforward networks and recurrent networks. In the feedforward network, the information moves in only one direction forwardly from the input layer through the hidden layer and to the output layer. There are no cycles or loops in such a network. Equations (2.0-5)-(2.0-7) actually describe a feedforward network. Contrary to feedforward networks, recurrent networks are models with bi-directional information flow. This network allows the neurons to depend not only on the input variable x_i but also on their own lagged values $n_{k,t-p}$ at order p. McNelis (2005) shows that the recurrent network builds "memory" in the evolution of the neurons. Replace Equation (2.0-5) by the following Equation (2.0-8) and the system of a recurrent network is formed,

$$n_{k,t} = w_{k,0} + \sum_{i=1}^{i^*} w_{k,i}\, x_{i,t} + \sum_{k=1}^{k^*} \phi_k\, n_{k,t-p} \tag{2.0-8}$$

McNelis (2005) points out that this recurrent network has an indirect feedback effect from the lagged unsquashed neurons to the current neurons, not a direct feedback from lagged neurons to the level of output.

The neural networks has the possibility of learning which attracts the most interest in data mining. A NN model modifies its interconnection weights by means of a set of learning (training) samples. The learning process leads to parameters of a network and these parameters implicitly store knowledge extracted from the data. More general, given a specific task to solve and a class of functions F, learning is a process of using observations (learning/training samples) to identify $f^* \in F$ that solves the task optimally. This requires defining a *cost function* $C : F \rightarrow \Re$ such that, for the optimal solution f^*, $C(f^*) \leq C(f)\ \forall f \in F$, that is the optimal solution has the minimal cost. Since the NN models learn from data, the cost function must be a function of the observations. (see, for example, Kantardzic 2003).

The cost function C is an important concept for measuring of how far away we are from an optimal solution to the target problem to be solved. Learning algorithms search through the whole solution space to identify a function minimizing the cost function. Three major learning paradigms corresponding to a particular task are listed in the literature, i.e., supervised learning, unsupervised learning, and reinforcement learning (see, for example Giudici 2003). In supervised learning, a set of example pairs $(x, y), x \in X, y \in Y$ are given to find a function f in the feasible class of functions matcheing the examples. In other words, the mapping implied by the data is inferred by minimizing a cost function which is related to the mismatch between the mapping and the data and implicitly contains prior knowledge about the problem domain. Classification (or pattern recognition) and regression are two types of supervised learning. Unsupervised learning is a method of fitting a model to observations. It is distinguished from supervised learning by the fact that there is no a priori output. In unsupervised learning, a data set of input objects is collected. Unsupervised learning then typically treats inputs as random variables. Typically, clustering, data compression, and density estimation fall within the paradigm of unsupervised learning. In reinforcement learning, data x is usually generated by an agent's interactions given a certain environment. At each point in time t, the agent performs an action y_t and the environment generates an observation x_t and an instantaneous cost c_t, according to certain unknown dynamics. The task is to find the rule that dominates the selecting actions of agents that minimize the measurable long-term cost or the expected cumulative cost. The dynamics of the environment and the long-term cost for each rule can be estimated if they are not available.

Training a neural network model is to select one model from the set of feasible models that minimizes the cost criterion. Numerous algorithms are available for training neural network models based on the straightforward application of optimization methods and statistical estimation. For example, evolutionary methods, simulated annealing, expectation-maximization, and non-parametric methods are the methods commonly used for training neural networks. Most of the algorithms are based on the form of gradient descent. This can be done by taking the partial derivative of the cost function with respect to the network parameters then changing those parameters in a gradient-related direction (see, Giudici 2003, Kantardzic 2003, and McNelis 2005).

2.3.4 Wavelet Analysis

Wavelets are mathematical tools used to divide a given function into different frequency components and study each component with a resolution that matches its scale (see Gençay et al. (2002)). A wavelet transform is the representation of a function by wavelets. The wavelets are scaled and translated copies (known as "child wavelets") of a finite-length or fast-decaying oscillating waveform (known as the "mother wavelet"). The wavelet transform utilizes the mother wavelet function and translates it to capture characteristics that are local in time and local in frequency. The resulting time-frequency partition corresponding to the wavelet transform is (1) long in time when capturing low-frequency events, thus good frequency resolution for such events can be obtained, and (2) long in frequency when capturing high-frequency events, thus good time resolution for such events can be obtained. Wavelet transforms have advantages over traditional Fourier transforms for representing functions that have discontinuities and sharp peaks. They intelligently adapt themselves to capture characteristics through a wide range of frequencies (see, for example, Gençay et al. 2002 and Percival and Walden 2000).

Compared with the Fourier transform, by which signals are represented as a sum of sinusoids, the wavelet transform is localized in both time and frequency whereas the standard Fourier transform is only localized in frequency. Someone might argue that the Short-time Fourier transform (STFT) is also time and frequency localized, but the STFT has the limitation of providing uniform time resolution for all frequencies[6]. Using multiresolution analysis the wavelets can reach a better signal representation. This is a very important feature of a wavelet transform, particularly in mining high-frequency financial data. Gençay et al. (2002) states several advantages by using wavelet methods: (1) Wavelet methods provide a natural platform to deal with the time-varying features in most real-world financial/economic time series data, which can avoid the stationarity assumption for a process. (2) Wavelet methods provide an easy vehicle to study the multiresolution properties of the time series data. The wavelet trans-

[6]Wavelet transforms provide high time resolution and low frequency resolution for high frequencies and high frequency resolution and low time resolution for low frequencies.

form decomposes a process into different time horizons (scales) by differentiating seasonalities, revealing jumps and volatility clustering, and identifying local and global dynamic properties of the process at these time scales. (3) Wavelet methods provide a convenient way of dissolving the correlation structure of a process across time scales.

Broadly speaking, there are two main types of wavelets. One is the continuous wavelet transform (CWT) which can deal with the time series defined over the entire real axis, and the other is the discrete wavelet transforms (DWT) which can deal with the time series defined over a range of integers (see Percival and Walden 2000). The CWT is a function of two variables $W(u,s)$ and is obtained by projecting the function of interest $x(t)$ into a particular wavelet ψ via

$$W(u,s) \quad = \quad \int_{-\infty}^{\infty} x(t)\, \psi_{u,s}(t)\, dt,$$

where

$$\psi_{u,s}(t) \quad = \quad \frac{1}{\sqrt{s}}\, \psi(\frac{t-u}{s})$$

is the translated (by u) and dilated (by s) version of the original wavelet function. The translation of a wavelet function $\psi(t-u)$ shifts its range u units to the right and a dilation of the function $\psi(t/s)$ expands its range by a multiplicative factor s. The resulting wavelet coefficients are a function of location parameters and scale parameters. The DWT is defined by the following equation:

$$W(u,s) \quad = \quad \sum_u \sum_s \langle x, \psi_{u,s} \rangle\, \psi_{u,s}(t),$$

where

$$\psi_{u,s}(t) \quad = \quad a^{-u/2}\, \psi(a^{-u} t - sb)$$

and $a > 1, b > 0$ (see Gençay et al. 2002).

2.3.5 Other Methods

In this section, other methods (i.e., decision trees and decision rules, association rules, genetic algorithms, and fuzzy inference) commonly used for computational data mining are briefly introduced.

Decision Trees and Decision Rules

Decision trees and decision rules are data mining methodologies applied for classification problems. In general, classification is a learning process of a function that maps a data into predefined classes. Every classification based on inductive-learning is to find a classifier that can predict the class for a given sample. For a decision tree representation, each interior node corresponds to an attribute; an outgoing branch represents a possible value of that attribute (node). A leaf stands for a possible value of the target variable given the values of the variables represented by the path from the root. Splitting the source set into subsets following an attribute value test makes one learn the tree. A recursive manner is adopted in repeating this process on each derived subset. The recursion will be stopped when splitting is either non-feasible, or a singular classification can be applied to each element of the derived subset. In order to improve the classification rate, a random forest classifier of a number of decision trees is used. Trees combinate several mathematical and computing techniques in the processes of description, categorization, and generalization of a given data set (see Kantardzic (2003)).

Association Rules

Data mining tasks usually require the analysis of data whose inherent relationship might be obscured by the data quantity and dimension. Being an unsupervised learning method, association rules determine relationships among a set of items in a database. Numerous applications of association rules involve categorical data with which normal cluster analysis loses its power. Rule-mining methods are also available for numerical data. Association rule inference involves finding sets of items which satisfy specific criteria, and in turn are used to infer the rules themselves. Three basic algorithms, i.e., Apriori, sampling, and partitioning, are often discussed in practise. More detailed discussion about association rules can be found in Kantardzic (2003) and others.

Genetic Algorithms

When dealing with problems that involve a large number of attributes, a search space growing in a combinatorially explosive manner will be imposed by such high dimensionality. Furthermore, an increased number of training samples are typically required to generate reliable results. The traditional machine learning algorithms tends to perform poorly due to the increased search space. Genetic algorithms aim to find good solutions to large-scale optimization problems through a unique combination of stochastic and directed search techniques. Genetic algorithms can deal with expansive search spaces and can help to reduce the burden of dimensionality. As stochastic optimization methods genetic algorithms have first been introduced by Holland

(1975) starting as computer simulations of natural genetic systems. Genetic algorithms encode each point in a solution space into a string named a chromosome and change the string during the execution. The string structures in the chromosomes undergo operations similarly to the natural evolution process in order to obtain increasingly optimal solutions. The quality of the solution depends on a goodness-of-fit value that relates to the objective function of the optimization problem. Genetic algorithms can be used for prediction, clustering, and association rule inference. For each of these uses, the method assumes a starting model then iteratively refined to find the optimal model for a given application. Holland (1975), Langdon and Poli (2002), and Haupt (2004) provide reference in details.

2.4 Statistical Data Mining

Statistical data mining refers to the use of robust statistical methods (i.e., statistical modeling and statistical inference) in data mining. It covers almost the whole area of statistical research. In this section, a partial introduction of common methods in mining high-frequency financial data is given focusing on statistical visualization, standard statistical models, and nonparametric methods for density estimation.

2.4.1 Robustness

Optimality and stability are two mutually complementary characteristics of any statistical procedure. In a parametric formulation, linearity of regression, stochastic independence, homoscedasticity, non-multicollinearity, and normality of errors, which are typically assumed provide access to standard statistical tools for drawing conclusions about parameters of interest. There are many cases, particularly in finance, where such regularity assumptions cannot be guaranteed to be tenable. Under this situation, the behavior of many optimal decisions is not stable to small deviations from prior assumptions. For example, some optimal procedures based on the least squares method are unstable and perform poorly under small deviations from the normality assumption (see Rachev and Mittnik (2000)). Therefore, plausible departures from model-assumptions of classical statistical procedures have been paid more attention, which leads to robust statistics.

Box (1953) first uses the term of Robustness. Box and Anderson (1955) argue that good statistical methods should be insensitive to changes without involving the parameters to be estimated or the hypothesis to be tested. It should be effective in being sensitive to the changes of parameters to be estimated or the hypothesis to be tested. This idea illustrates the natural implications of insensitivity to model departures and sensitivity to good performance under the model. Based on finding robust alternatives of the classical methods in practice, Tukey (1962) points out the direction in statistical studies called exploratory or probability-free data analysis. Roughly speaking, robustness means optimality of methods used for data analysis and stability of statistical inference under the variations of the accepted models. For high-frequency financial econometrics, I consider robustness based on Box (1953) and Tukey (1962) at two levels, i.e., robustness in mining high-frequency financial data and robustness in modeling with high-frequency financial data.

2.4.2 Visualization

Data visualization is a method of using computer graphics to present data in innovative ways by which obscured abstract or complex information become obvious and easy to understand with the help of images or animations. Practical application of data visualization involves selecting, rearranging, transforming, and representing abstract data in a specific form for exploration, particularly by means of computer graphics. Basic functions for generating complex images from abstract data can be realized by computer graphics. A large variety of techniques are available in data visualization. Choosing a specific technique should be determined by the type of data and the dimension of the data to be visualized. The generally discussed issues contain volume visualization for volumetric data, large data set visualization for finance, informative visualization for text libraries, and statistical visualization for data analysis and understanding. Many visualization techniques are designed to deal with multidimensional multivariate data sets. Rao et al. (2005) provide a reference for data mining and data visualization and Giudici (2003) discusses graphical models in statistical data mining.

2.4.3 Standard Models

The purpose of this section is to briefly discuss issues about high-frequency data modeling without going into the details of actual model specifications[7]. It is generally agreed that the nature of data and the purpose of modeling greatly affect the type of model being developed. There are two major purposes of high-frequency data models, i.e., predicting the timing of market transactions and predicting the volatility for hedging. Ghysels et al. (1998) define two broad categories, i.e., (1) market transactions models and (2) financial time series models. For the tick-by-tick data, the first category includes models based on point processes, durations and hazard models. For the intradaily equally spaced data prices (volumn) and number of transactions, the first category includes models based on subordinated processes

[7]The details will be discussed in Chapter 3

with transactions-based time deformation models and the second category contains unequally spaced ARCH and time deformation stochastic volatility models. For the intradaily equally spaced data prices (volumn) but no transactions information, the first category includes latent variable models with subordinated processes and the second category includes standard ARCH and stochastic volatility models.

Ghysels et al. (1998) argue that it is important to note the difference between market transactions models and the broad class of volatility models (i.e., ARCH and SV models). They state "the latter are a relatively uniform class involving the same fundamental structure which may not necessarily lend itself easily to temporal aggregation" while market transactions models "depending on the specific context and purpose one encounters point processes counting transaction events, data, duration, and hazard models, continuous time jump processes, among many others" which "clearly do not share the same parametric structure and typically little is known about their temporal aggregation features".

2.4.4 Nonparametric Methods

Nonparametric methods have been widely used in data mining. The main application of nonparametric methods is for density estimation. The kernel estimator is a basic method used to estimate density, see Silverman (1986). If the random variable X has density $f(x)$, then

$$f(x) \;=\; \lim_{a\to 0}\frac{1}{2a}P\left(x-a<X<x+a\right)$$

By counting the proportion of sampling observations falling in the interval of $(x-a, x+a)$, the probability $P\left(x-a<X<x+a\right)$ can be estimated for any given a. Defining kernel function K for

$$\int_{-\infty}^{\infty} K(x)d(x) \;=\; 1$$

in which, $K(x)$ usually but not always is regarded as a symmetric probability density function, for example, normal density. The kernel estimator is defined by

$$\tilde{f}(x) \;=\; \frac{1}{na}\sum_{i=1}^{n} K\left(\frac{x-X_i}{a}\right)$$

where a is the window width and n is sample size. The kernel estimator can be looked as a sum of bumps placed at the observations X_i. The kernel function $K(x)$ determines the shape of the bumps and the window width a determines the width of the bumps.

For evaluating the quality of estimation, the mean square error (abbreviated MSE) has been defined as:

$$MSE_x(\tilde{f}) = E\left(\tilde{f}(x)-f(x)\right)^2 = \left(E\tilde{f}(x)-f(x)\right)^2 + var\left(\tilde{f}(x)\right)$$

Meanwhile, to measure the global closeness of fit of $\tilde{f}(x)$ to $f(x)$ by integrating the MSE over x, the mean integrated square error (MISE) is defined as follows:

$$MISE_x(\tilde{f}) = E\int \left(\tilde{f}(x) - f(x)\right)^2 dx = \int \left(E\tilde{f}(x) - f(x)\right)^2 dx + \int var\left(\tilde{f}(x)\right) dx$$

Given a symmetric kernel function K, $\int tK(t)dt = 0$ and $\int t^2K(t)dt = k_2 \neq 0$, Silverman (1986) shows the approximation of MISE is:

$$\frac{1}{4}a^4k_2^2\int f''(x)^2dx \;+\; \frac{1}{na}\int K(t)^2dt$$

It is clear that the bias in the estimation of $f(x)$ depends on the window width. The optimal window width a_{opt} can be chosen by minimizing the MISE. Silverman (1986) shows that

$$a_{opt} = n^{-1/5}\,k_2^{-2/5}\left(\int f''(x)^2dx\right)^{-1/5}\left(\int K(t)^2dt\right)^{1/5},$$

for which the optimal solution is given by Epanechnikov kernel $K_E(x)$:

$$K_E(x) = \begin{cases} \frac{3}{4\sqrt{5}}(1 - \frac{x^2}{5}) & -\sqrt{5} \leq x \leq \sqrt{5} \\ 0 & \text{else.} \end{cases}$$

A slight drawback suffered by the kernel estimator is its inefficiency with long-tailed distributions. Since across the whole sample, the window width is fixed, a good degree of smoothing over the center of the distribution will often leave spurious noise on the tails, see Silverman (1986) and Dowd (2005). Silverman (1986) offers some solutions such as the nearest neighbor method and the variable kernel method. For the nearest neighbor method, the window width placed on an observation depends on the distance between that observation and its nearest neighbors. For the variable kernel estimator, the density $f(x)$ is estimated under:

$$\tilde{f}(x) \;=\; \frac{1}{n}\sum_{i=1}^{n}\frac{1}{ah_{i,k}}K\left(\frac{x - X_i}{ah_{i,k}}\right)$$

where $h_{i,k}$ is the distance between X_i and the kth nearest of the other data points. The window width of the kernel placed on X_i is proportional to $h_{i,k}$, therefore flatter kernels will be placed on more sparse data.

2.5 Evaluation of Data Mining Methods

The principle task of data mining in high-frequency financial econometrics is to build the right model right to explore the information contained in massive data. To choose the right model and build the model right requires some sequential procedures (forward, backward, and stepwise) that allow such a model to be chosen through a sequence of pairwise comparisons. All this suggests the need for a systematic study to compare and evaluate statistical models for data mining. In this section, the most important methods that are frequently used will be reviewed.

2.5.1 Criteria Based on Statistical Goodness-of-fit Techniques

Distance between statistical models

Letting $F_s(x)$ denote the empirical sample distribution and $\tilde{F}(x)$ the estimated distribution function, several distance measures are defined as follows:

Anderson-Darling (AD) distance

$$AD = \sup_{x\in\Re} \frac{|F_s(x) - \tilde{F}(x)|}{\sqrt{\tilde{F}(x)(1-\tilde{F}(x))}}.$$

Chi-squared (D_{χ^2}) distance

$$D_{\chi^2} = \sum_x \frac{\left(F_s(x) - \tilde{F}(x)\right)^2}{\tilde{F}(x)}.$$

Cramer Von Mises (CVM) distance

$$CVM = \int_{-\infty}^{\infty} \left(F_s(x) - \tilde{F}(x)\right)^2 d\tilde{F}(x)$$

Entropic (E) distance

$$E = \sum_x F_s(x) \log \frac{F_s(x)}{\tilde{F}(x)}.$$

Kolmogorov-Smirnov (KS) distance

$$KS = \sup_{x\in\Re} |F_s(x) - \tilde{F}(x)|.$$

Kuiper (K) distance

$$K = \sup_{x\in\Re}\left(F_s(x) - \tilde{F}(x)\right) + \sup_{x\in\Re}\left(\tilde{F}(x) - F_s(x)\right).$$

Discrepancy of a statistical models

Discrepancy is another statistical measure for goodness of fit. Assuming f represents the unknown density of the population, and $g = \psi_\theta$ be a family of density functions (with a vector of I parameters θ) that approximates it, the discrepancy of g and f based on the Euclidean distance can be defined as follows:

$$\Delta(f, \psi_\theta) = \sum_{i=1}^{n} (f(x_i) - \psi_\theta(x_i))^2$$

This discrepancy, which is a function of parameters θ, measures the total squared error made by replacing f by g. The discrepancy also can be calculated with the distances introduced in the previous section. If the true population density f is known, the discrepancy simply determines how well the f can be approximated by g. Then the discrepancy of g (due to the parametric approximation) can be obtained as the discrepancy between the unknown probabilistic model and the best parametric model ψ^*_θ,

$$\Delta(f, \psi^*_\theta) = \sum_{i=1}^{n} (f(x_i) - \psi^*_\theta(x_i))^2$$

Unfortunately, f is not always known, then the best parametric model cannot be identified. But, the sample estimation, denoted as $\psi_{\hat{\theta}}$, for which the I parameters can be estimated from the data, can be used to substitute f. Then the discrepancy of g (due to the estimation process) turns to be

$$\Delta(\psi_{\hat{\theta}}, \psi^*_\theta) = \sum_{i=1}^{n} (\psi_{\hat{\theta}}(x_i) - \psi^*_\theta(x_i))^2.$$

It is better to choose a family of density where the model has a rich number of parameters in order to get a relatively good approximation. The discrepancy due to approximation turns out to be smaller for more complex models with a large number of parameters. But the sample estimates obtained with a more complex model tend to overfit the data, therefore the discrepancy due to estimation will increase (see Giudici 2003). The total discrepancy measuring the discrepancy between f and $\psi_{\hat{\theta}}$ takes both these factors (discrepancy from parametric approximation and discrepancy from estimation) into account. It is defined as follows,

$$\Delta(f, \psi_{\hat{\theta}}) = \sum_{i=1}^{n} (f(x_i) - \psi_{\hat{\theta}}(x_i))^2.$$

Giudici (2003) argues that the best model to approximate f will be the model $\psi_{\hat{\theta}}$ that minimizes the total discrepancy, since minimization of the discrepancy from the parametric approximation favors complex models, which are more adaptable to the data, whereas minimization of the discrepancy from estimation favors the simple models, which are more stable when faced with variations in the observed sample. He also points out that the total discrepancy cannot be calculated in practice since f is unknown, then the total expected discrepancy obtained with respect to the sample density can be used. Then the problems turn to be finding an appropriate estimator of the total expected discrepancy. Such an estimator is known as minimum discrepancy estimator.

One discrepancy often used for model evaluation, particularly based on the evaluation of information loss, is the Kullback-Leibler discrepancy (see Kullback and Leibler 1951) defined as follows:

$$\Delta_{KL}(f, \psi_{\hat{\theta}}) = \sum_{i=1}^{n} f(x_i) \log\Big(\frac{f(x_i)}{\psi_{\hat{\theta}}(x_i)}\Big).$$

The Kullback-Leibler discrepancy derives from the entropy distance rather than Eucliedean distance. The best model under the Kullback-leibler discrepancy is the one with a minimal loss of information from the true unknown f (see Burnham and Anderson 1998).

2.5.2 Criteria Based on Score Functions

When using the maximum likelihood estimation (MLE) methods, the sore function (or simply score) is the partial derivative with respect to some parameter θ of the logarithm of the likelihood function. The expected value of the score is zero. If one wants to repeatedly sample from some distribution, and repeatedly compute the score with the true parameter θ, the mean value of the score will tend to zero as the repeated samples go to infinity. Two criteria based on the score function are the Akaike information criterion (AIC) and the Schwartz information criterion (SIC).

Akaike information criterion (AIC)

Akaike (1974) proposes a measure grounded in the concept of entropy as

$$AIC = -2 \log \ell(\hat{\theta}; x_1, \cdots, x_n) + 2q$$

where $\log \ell(\hat{\theta}; x_1, \cdots, x_n)$ is the logarithm of the likelihood function calculated in the maximum likelihood parameter estimation and q is the number of parameters in the model.

Schwartz information criterion (SIC)

Schwartz (1978) proposes a statistical criterion for model selection defined as

$$SIC = -2 \log \ell(\hat{\theta}; x_1, \cdots, x_n) + q \log n$$

where $\log \ell(\hat{\theta}; x_1, \cdots, x_n)$ is the logarithm of the likelihood function calculated in the maximum likelihood parameter estimation, q is the number of parameters in the model, and n is the sample size.

Criteria based on score function offer a high generality of application, complete ordering, and simple calculation. Their main disadvantage is that they do not define threshold levels to choose one model rather than another (see Giudici 2003 and Zucchini 2000).

2.5.3 Criteria Based on Loss Functions

Since the main problem dealt with by data mining is to reduce the uncertainties in the risk factors, which are correlated and not necessarily causal, the best model is the one that leads to the lowest loss. The receiver characteristic (ROC) curve measures the predictive accuracy of a model based on the cost-benefit analysis of loss, see, for example, Giudici (2003), Lasko et al. (2005), Spackman (1989), and Zweig and Campbell (1993).

To plot the ROC curve, a two-class prediction problem (binary classification), in which the observations of a validation data set are classified to either a positive (p) or a negative (n) category, four possible categories are classified:

1. Observations predicted as events and effectively such (called true positive), i.e., a prediction is p and the actual value is also p.

2. Observations predicted as events and effectively non-events (called false positive), i.e., a prediction is p but the actual value is n.

3. Observations predicted as non-events and effectively events (called false negative), i.e., the prediction outcome is n while the actual value is p.

4. Observations predicted as non-events and effectively such (called true negative), i.e., both the prediction outcome and the actual value are n.

To draw the ROC curve, only the true positive rate (i.e., the TPR, which determines a diagnostic test performance on classifying positive instances correctly among all positive samples available) and the false positive rate (i.e., the FPR, which determines the number of incorrect positive results while they are actually negative among all negative samples available) are needed. The ROC space is defined in the Cartesian plane by FPR and TPR as x and y axes respectively. Such a space depicts relative trade-offs between benefits (true positive) and costs (false positive). The proportion of events predicted as such is called sensitivity, and the proportion of non-events predicted as such is called specificity. Then the ROC curve is sometimes called the sensitivity-specificity plot, since TPR is equivalent to sensitivity and FPR is equal to 1 minus specificity. Each prediction result or one case of a confusion matrix (where the main diagonal elements show the number of observations that have been classified correctly for each class and the off-diagonal elements indicate the number of observations that have been incorrectly classified) corresponds to one point in the ROC space.

The best possible prediction method yields a point at the upper left corner in the ROC space. The point of $(0, 1)$ in the ROC space is also called a perfect classification, corresponding to 100% sensitivity (all true positives are found) and 100% specificity (no false positives are found). A completely random guess would give a point along a 45° line (i.e., the no-discrimination line) going through the left bottom to the top right corners. The ROC space is divided by the 45° line in areas of good or bad classification/diagnostic. Points lying above the diagonal line bespeak good modeling results, while points lying below this line point out worse modeling results. Giudici (2003) points out that the ROC curve will always lie above the 45° line and the area between the curve and the line can be calculated. The larger the area, the better the model.

2.5.4 Criteria Based on Bayesian Methods

Bayesian criteria provide coherent statistical measure by combining the deviance differences and the scoring function. In the Bayesian derivation, each model is given a score that corresponds to the posterior probability of the model. Then a model turns to be a discrete random variable that takes values on the candidate models space. The Bayesian criteria provide a complete ordering of the models and can be used to compare non-nested models and the models belonging to different classes. The model that maximizes this posterior probability will be chosen. Bernardo and Smith (1994) and Rachev et al. (2007) give more information about Bayesian theory and selection criteria for Bayesian models.

There are several Bayesian approaches. One approach is through Bayes factors. The posterior probability of a model (M) given data vector x_i, $i = 1, \ldots, n$, $P(M|x_i)$ is given by Bayes' theorem:

$$P(M|x_i) = \frac{P(x_i|M)P(M)}{P(x_i)}$$

The key data-dependent term $P(x_i|M)$ is a likelihood, and is sometimes called the evidence for model M; evaluating it correctly is the key to Bayesian model comparison. To calculate this, the parameters of the model must be integrated out, that is,

$$P(M \mid x_i) = \int P(x_i \mid \theta, M) \, P(\theta \mid M) \, d\theta$$

where $P(\theta \mid M)$ is the prior distribution of the parameters given the considered model M. Such integration sometimes is a difficult task. Markov Chain Monte Carlo (MCMC) provides a successful but computationally intensive way to approximate such integration problems (see Rachev et al. (2007)).

Given a model selection problem in which two models M_1 and M_2 are compared on the basis of a data vector x_i. The Bayes factor K is given by

$$K = \frac{P(x_i \mid M_1)}{P(x_i \mid M_2)} = \frac{\int P(\theta_1 \mid M_1) P(x_i \mid \theta_1, M_1)\, d\theta_1}{\int P(\theta_2 \mid M_2) P(x_i \mid \theta_2, M_2)\, d\theta_2}$$

where $P(x_i \mid M_{1,2})$ is called the marginal likelihood for model $M_{1,2}$ and the models $M_{1,2}$ will be parameterized by vectors of parameters $\theta_{1,2}$ respectively. The logarithm of K is called the weight of evidence given by x_i for M_1 over M_2. A value of $K > 1$ shows that the data indicate that M_1 is more strongly supported by the data under consideration than M_2.

2.5.5 Criteria Based on Computational Methods

Computationally intensive model selection criteria have been introduced after using widely spread computational methods in data mining. These criteria are generally applicable to all the models (i.e., parametric and nonparametric models). A possible problem with these criteria is that they require intensive computation, thus a long time to design and implement is needed. In this section, two commonly used criteria, the cross-validation and bootstrapping, are introduced.

The cross-validation criterion

Cross-validation (rotation estimation) is a statistical procedure of dividing the sample into subsamples (usually two subsamples) such that the analysis (such as fit a model) is initially performed on a single subsample (called training sample), while the other subsamples (called validation or testing samples) are used for confirming and validating the initial analysis (see Ron (1995)). Using this criterion the performance between two or more models can be compared by evaluating an appropriate discrepancy function on the validation or testing samples.

One problem with the cross-validation criterion is how to decide the size of validation or testing samples. Another problem is that if the validation sample is used to choose a model, the results obtained are not real measurements of the model's performance that can be compared with the measurements obtained from other models since the validation sample is in fact used to construct the model (see Giudici (2003)).

To solve the problem pointed out above, several common types of cross-validation have been introduced. Holdout validation tries to avoid crossing over the data. Choosing observations randomly from the initial sample to form the validation data, and the remaining observations are kept as the training data. In general, less than a third of the initial sample is used for forming validation data. For example, Giudici (2003) suggests the proportions of 75% and 25% are used for the training and validation samples respectively. K-fold cross-validation method suggests to divide the initial sample into K subsamples with equal size. The model is fitted k times, leaving out one of the subsets each time for calculating a validation rate. The cross-validation process is then repeated K times (the folds), with each of the K subsamples used exactly once as the validation data. The final error is the arithmetic average of the errors obtained. Leave-one-out cross-validation (i.e., the jackknife method) involves using a single observation from the original sample as the validation data, and the remaining data as the training sample. The disadvantage of these validation methods is that they retrain the model several times and require intensive computation (see Duda et al. (2001), Giudici (2003), and Ron (1995)).

The bootstrap criterion

Efron (1979) introduces the bootstrap method based on the idea of reproducing the "real" distribution of the population with a resampling of the observed data. Application of this method stems from the assumption that the observed sample is in fact a population, for which we can calculate the sample density. That is, the choice for an approximation of the population distribution is the empirical distribution from the observed data. It can be used for constructing hypothesis tests, particularly, when the population distribution is unknown (see, for example, Sun et al. (2007b)). It can be used as an alternative to inference based on parametric assumptions when those assumptions are in doubt (see Davison and Hinkley (1997)).

The advantage of bootstrapping is its simplicity when compared with analytical methods. The idea is straightforward to employ the bootstrap to derive estimates of standard errors and confidence intervals for these estimators of the underlying distribution with complex parameters, such as proportions, percentile points, correlation coefficients, and odds ratio. The disadvantage of bootstrapping is that it has asymptotical consistence only under some conditions, it does not guarantee consistence for general finite sample and tends to be overly optimistic. Bootstrap method usually requires intensive computation particularly when the large number of bootstrap samples is required.

Two situations should be taken into account in data mining. When the empirical distribution is unclear, smooth bootstrapping is used. For this situation, a small amount of (usually normally distributed) zero-centered random noise is added onto each resampled observation, which is equivalent to sampling from a kernel density estimate of the data. When the empirical distribution can be described by a parametric model (often fitting the data by the maximum likelihood estimation), the bootstrapping samples of random numbers can be drawn from such a parametric model.

Bootstrap methods are often used in regression analysis. The resampling of residuals is one of the common approaches. The bootstrap procedure is:

1. Fit the model and retain the fitted values $\hat{\theta}_i$ and the residuals ϵ_i, $i = 1, \ldots, n$.
2. For each pair, (x_i, y_i), in which x_i is the explanatory variables, add a randomly resampled residual, ϵ_i to the response variable y_i
3. Refit the model and retain the estimated parameters.
4. Repeat steps 2 and 3 many times.

This procedure retains the information in the explanatory variables. However, a question arises as to which residuals to resample, i.e., residuals follow a parametric model or not. Residuals following no parametric model are one option, another is studentized residuals (in linear regression). Whilst there are arguments in favor of using studentized residuals, in practice it often makes little difference and it is easy to run both and compare the results against each other. Rachev et al. (2007) illustrate several applications of bootstrap methods in financical analysis.

Chapter 3

High-Frequency Financial Econometrics

"Financial econometrics is the econometrics of financial markets. It is a quest for models that describe financial time series such as prices, returns, interest rates, financial ratios, defaults, and so on (see Rachev et al. 2007)". The financial equivalent of the laws of physics, financial econometrics represents the quantitative, mathematical laws of financial markets. Following the above definition, I say *high-frequency financial econometrics* is the econometrics based on the data gathered at the ultimate frequency level of financial markets. In the study of high-frequency financial econometrics, the time series of prices, returns, interest rates, financial ratios, defaults, and others are all observed at the tick-by-tick level. I also call such time series the *high-frequency time series*. Similarly, I describe the financial theory (especially, the market microstructure theory) that investigates financial phenomena at intra-daily or accumulated intra-daily level as the *high-frequency finance*. In this sense, high-frequency financial econometrics is the quantitative, mathematical laws of tick-by-tick financial markets. Although currently high-frequency finance is focusing on the market microstructure study due to the limitation of data, high-frequency financial econometrics has a broader scope of research driven by the increasing demand of financial decisions made at intra-daily scale. For example, day-traders measure risk and select the optimal portfolio within one day or even within a couple of hours, for which standard financial theories based on low-frequency data cannot provide efficient solutions. As a growing area, high-frequency financial econometrics can offer solutions with quantitative techniques developed from tick-by-tick financial data. High-frequency financial econometrics typically touches on one or more of the following aspects of financial markets.

3.1 Mechanisms in Economic Settings

In the study of market quality, much of the debate in the literature, according to Madhavan (2000), centers on floor versus electronic markets, and auctions versus dealer systems. Using intra-daily data for Bund future contracts, Frank and Hess (2000) compare the electronic system of the DTB (Deutsche Terminbörse) and the floor trading of the LIFFE (London International Financial Futures Exchange). They find that the floor trading turns out to be more attractive in periods of high information intensity, high volatility, high volume, and high trading frequency. The reason they offer for this finding is that market participants infer more information from observing actual trading behavior. Coval and Shumway (2001) confirm the claim that market participants gather all possible information rather than only relying on easily observable data, say, past prices, in determining their trade values. They suggest that subtle but non-transaction signals play important roles: The electronic exchanges lose information that can be observed in a face-to-face exchange setting. Based on the co-existence of the floor and electronic trading system in the German stock market, Theissen (2002) argues that when employing intra-daily data, both systems contribute to the price discovery process almost equally. Based on testing whether the upstairs intermediation can lower adverse selection cost, Smith et al. (2001) report that the upstairs market is complementary in supplying liquidity. Bessembinder and Venkataraman (2004) show that the upstairs market does a good job of complementing the electronic exchange, because the upstairs market is efficient for many large trades and block-sized trades.

By confirming that market structure does impact the incorporation of news into market prices, Masulis and Shivakumar (2002) compare the speed of adjustment on the NYSE and Nasdaq. They demonstrate that there are faster price adjustments to new information on the Nasdaq. Weston (2002) confirms that the electronic trading system has improved the liquidity of the Nasdaq. Boehmer et al. (2005) suggest that since the electronic trading reveals more information, market quality could be increased by exposing such information in the electronic trading system.

Huang and Stoll (2001) point out that tick size, bid/ask spread, and market depth are not independent from market structure. They are linked to market structure. It is necessary to take account of tick size, bid/ask spread, and market depth when analyzing the market structure. Numerous studies investigate how market structure impacts price discovery and trading costs in different securities markets. Chung and Van Ness (2001) show that after introducing new order handling rules in the Nasdaq, bid/ask spreads decreased, confirming that market structure has a significant effect on trading costs and the price forming process.

3.2 Formation of Market Price

How are prices determined in the financial market? As Madhavan (2000) notes, studying the market maker is the logical starting point for any analysis of market price determination. In market microstructure studies, inventory and asymmetric information are the factors that influence the price movement. Naik and Yadav (2003) examine whether equivalent inventories or ordinary inventories dominate the process of trading and pricing decisions by dealer firms. They find that ordinary inventories play the main role consistent with the decentralized nature of market making. Bollen et al. (2004) propose a new model based on intra-daily data to understand and measure the determinants of bid/ask spreads of market makers. Using the vector autoregressive model, Dufour and Engle (2000) find that time durations between transaction impact the process of price formation. Engle and Lunde (2003) develop a bivariate model of trades and quotes for NYSE traded stocks. They find that high trade arrival rates, large volume per trade, and wide bid/ask spreads can be regarded as information flow variables that revise prices. They suggest that if such information is flowing, prices respond to trades more quickly.

3.3 Transparency of the Market

Market transparency is the ability of market participants to observe information about the trading process, see O'Hara (1995). Referring to the time of trade, Harris (2003) defines *ex ante* (pre-trade) transparency and *ex post* (post trade) transparency. Comparing market transparency is complicated by the lack of a criterion that can be used to judge the superiority of one trading system over another such as floor market versus electronic market, anonymous trading versus disclosure trading, and auction system versus dealer system. According to Madhavan (2000), there is broad agreement of the influence of market transparency. Market transparency does affect informative order flow and the process of price discovery. Madhavan also points out that while partial disclosure will improve liquidity and reduce trading costs, complete disclosure will reduce liquidity, a situation Harris (2003) refers to as the "ambivalence" from the viewpoint of a trader's psychology. By investigating the OpenBook[1] in NYSE, Boehmer *et al.* (2005) find that there is a higher cancel rate and a short time-to-cancel of limit orders in the book, suggesting that traders attempt to manipulate the exposure of their trades. By confirming market design does impact the trading strategy of investors, they support increasing pre-trade transparency and suggest that market quality can be enhanced by greater transparency of the limit order book.

[1] OpenBook was introduced in January 2002 allowing traders off the NYSE floor exchange to find each price in real time for all listed securities. Before OpenBook, only best bid/ask could be observed.

3.4 Liquidity of the Market

A topic of debate in finance is the meaning of liquidity. Some market participants refer to liquidity as the ability to convert an asset into cash quickly, some define liquidity in terms of low transaction cost, and some think high transaction activity is liquidity, see Easley and O'Hara (2003) and Schwartz and Francioni (2004). Schwartz and Francioni point out that a better approach for defining liquidity should be based on the attributes of liquidity. They define the depth, breadth, and resiliency as the dimensions of liquidity, while Harris (2003) identifies immediacy, width, and depth as the relevant attributes of liquidity. In the microstructure literature, the bid/ask spread proxies for liquidity, see Easley and O'Hara (2003). Schwartz and Francioni (2004) show that liquidity can be approximated by the trade frequency of an asset traded in the market. The frequency can be measured by the magnitude of short-term price fluctuation for such an asset. Chordia *et al.* (2002) find that order imbalances affect liquidity and returns at the aggregate market level. They suggest using order imbalance as a proxy for liquidity.

Using bid/ask spread as a proxy for liquidity, Chung *et al.* (2001) compare the bid/ask spread on the NYSE and the Nasdaq. They find that the average NYSE specialist spreads are significant smaller than the Nasdaq specialist spreads. Huang and Stoll (2001) report that dealer markets have relative higher spreads than auction markets. By comparing spreads in different markets, they found that the spreads on the London Stock Exchange are larger than that for the same stocks listed on the NYSE and the spreads in the Nasdaq are larger than that on stocks listed on the NYSE. By checking the growth of electronic communication networks (ECNs) in Nasdaq, Weston (2002) confirms that the electronic trading system has the ability to improve the liquidity on the Nasdaq. Kalay *et al.* (2004) report that the opening is more liquid than the continuous trading stage. For small price changes and small quantities, there is a less elastic supply curve than demand curve.

From the perspective of liquidity providers, researchers usually use the order book data. Coughenour and Deil (2002) catagorize the specialist firms on the NYSE into two types: owner-specialist firms and employee-specialist firms. By investigating the influence of these two types of liquidity providers, they show that with similar trading costs, the owner-specialist firms have a greater frequency of large trades and have a greater incentive to reduce adverse selection costs. For employee-specialist firms, the stocks traded exhibit less sensitivity between change in quoted depth and quoted spreads. Meantime, these stocks show price stability at the opening. Peterson and Sirri (2002) investigate order submission strategies and find that limit orders perform worse than market orders involving the trading costs. But investors still prefer limit orders, which suggests that individual investors are less able to choose an optimal trading strategy.

3.5 Volatility of the Market

Volatility is one of the most important risk metrics. Harris (2003) defines it as the tendency for prices to change unexpectedly. Volatility could be regarded as the market reaction to news reflected by price changes . He distinguishes fundamental volatility and transitory volatility. *Fundamental volatility* is caused by endogenous variables which determine the value of trading instruments. *Transitory volatility* is due to the trading activity of uninformed traders.

Volatility is not constant over the trading stage, but changes over time. Transitory volatility might occur in a very short time period before it converts to its fundamental value. The intra-daily data reflects both fundamental volatility and transitory volatility with sufficient statistical significance.

Engle (2000) adopts the GARCH model to the irregularly spaced ultra-high frequency data. Letting d_i be the duration between two successive transactions and r_i the return between transactions $i-1$ and i, then the conditional variance per transaction is:

$$V_{i-1}(r_i|d_i) = \phi_i \tag{3.0-1}$$

and the conditional volatility per unit of time is defined as,

$$V_{i-1}(\frac{r_i}{\sqrt{d_i}}|d_i) = \sigma_i^2 \tag{3.0-2}$$

Then the connection between equations (3.0-1) and (3.0-2) can be established by $\phi_i = d_i\sigma_i^2$. The predicted variance conditional on past returns and durations is $E_{i-1}(\phi_i) = E_{i-1}(d_i\sigma_i^2)$. Using ARMA(1,1) with innovations ε_i, the series of return per unit of time is

$$\frac{r_i}{\sqrt{d_i}} = a\,\frac{r_{i-1}}{\sqrt{d_{i-1}}} + \varepsilon_i + b\,\varepsilon_{i-1} \tag{3.0-3}$$

If the current duration contains no information, the simple GARCH specification is used, and then

$$\sigma_i^2 = \omega + \alpha\,\varepsilon_{i-1}^2 + \beta\,\sigma_{i-1}^2 \tag{3.0-4}$$

If durations are informative, Engle (2000) proposes an autoregressive conditional duration (ACD) model to define the expected durations. If the ACD model is

$$\psi_i^2 = h + m\,d_{i-1} + n\,\psi_{i-1}^2 \tag{3.0-5}$$

then the ultra-high frequency GARCH model is expressed as:

$$\sigma_i^2 = \omega + \alpha\,\varepsilon_{i-1}^2 + \beta\,\sigma_{i-1}^2 + \gamma_1\,d_i^{-1} + \gamma_2\,\frac{d_i}{\psi_i} + \gamma_3\,\xi_{i-1} + \gamma_4\,\psi_i^{-1} \tag{3.0-6}$$

where ξ_{i-1} is the long-run volatility computed by exponentially smoothing r^2/d with a parameter 0.995 such that

$$\psi_i = 0.005\left(\frac{r_{i-1}^2}{d_{i-1}}\right) + 0.995\,\psi_{i-1} \tag{3.0-7}$$

Alexander (2001) points out that a number of studies have shown that the aggregation properties of GARCH models are not straightforward. The persistence in volatility seems to be lower when it is measured using intra-day data than when measured using daily or weekly data. For example, fitting a GARCH(1,1) to daily data would yield a sum of autoregressive parameter and moving average parameter estimates that is greater than the sum of that estimated from fitting the same GARCH process by using 2-day returns. In the heterogeneous ARCH (HARCH) model proposed by Müller et al. (1997), a modified process is introduced so that the squared returns can be taken at different frequencies. The return of r_t of the HARCH(n) process is defined as follows,

$$\begin{aligned} r_t &= h_t\,\varepsilon_t \\ h_t^2 &= \alpha_0 + \sum_{j=1}^{n} \alpha_j\,(\sum_{i=1}^{j} r_{t-i})^2 \end{aligned} \tag{3.0-8}$$

where ε_t is an i.i.d. random variable with zero mean and unit variance, $\alpha_0 > 0$, $\alpha_n > 0$, $\alpha_j \geq 0$, for $j = 1, 2, ..., n-1$. The equation for the variance h_t^2 is a linear combination of the squares of aggregated returns. Aggregated returns may extend over some long intervals from a time point in the distant past up to time $t-1$. The HARCH process belongs to the wide ARCH family but differs from all other ARCH-type processes in the unique property of considering the volatilities of returns measured over different interval sizes. Dacorogna et al. (2001) generalized the HARCH process. In equation (3.0-9), all returns considered by the variance equation are observed over the recent interval ending at time $t-1$. This strong limitation is justified by observing it empirically, but a more general formula of the process with observation intervals ending in the past before $t-1$ can be shown as follows:

$$\begin{aligned} r_t &= h_t\,\varepsilon_t \\ h_t^2 &= \alpha_0 + \sum_{j=1}^{n}\sum_{k=1}^{j} \alpha_{jk}\,(\sum_{i=k}^{j} r_{t-i})^2 + \sum_{i=1}^{q} b_i\,h_{t-i}^2 \end{aligned} \tag{3.0-9}$$

where

$$\begin{aligned} \alpha_0 > 0,\ \alpha_{jk} \geq 0, \quad &for \quad j = 1, 2, ..., n;\ k = 1, 2, ..., j; \\ b_j \geq 0 \quad &for \quad i = 1, 2, ..., q. \end{aligned} \tag{3.0-10}$$

The generalized process equation considers all returns between any pair of two time points in the period between $t-n$ and $t-1$. It covers the case of HARCH (all $\alpha_{jk}=0$ except some α_{j1}), as well as that of ARCH and GARCH (all $\alpha_{jk}=0$ except some α_{jj}).

From historical data, realized volatility can be computed. Dacorogna et al. (2001) show the realized volatility as:

$$\upsilon(t_i) = \left[\frac{1}{n}\sum_{j=1}^{n} |r(\Delta t; t_{i-n+j})^p|\right]^{1/p} \tag{3.0-11}$$

where n is the number of return observations, and r stands for the returns in the regularly spaced time intervals. Δt is the return interval. Taylor and Xu (1997), Andersen and Bollerslev (1998), and Giot and Laurent (2004), among others, show that summing up intra-daily squared returns can estimate the daily realized volatility. Given that a trading day can be divided into n equally spaced time intervals, and if $r_{i,t}$ denotes the intra-daily return of the ith interval of day t, the daily volatility for day t can be expressed as:

$$\left[\sum_{i=1}^{n} r_{i,t}\right]^2 = \sum_{i=1}^{n} r_{i,t}^2 + 2\sum_{i=1}^{n}\sum_{j=i+1}^{n} r_{j,t}\, r_{j-i,t} \tag{3.0-12}$$

Andersen et al. (2001a) show that if the returns have a zero mean and are uncorrelated, $\sum_{i=1}^{n} r_{i,t}^2$ is a consistent and an unbiased estimator of the daily variance. $[\sum_{i=1}^{n} r_{i,t}]^2$ is called the *daily realized volatility* since all squared returns on the right side of equation (16) can be observed when intra-daily data sampled over an equally spaced time interval are available. This is one method for modeling daily volatility using intra-daily data, and a method that has been generalized by Andersen et al. (2001a, 2001b). Another method is to estimate the intra-daily duration model on trade durations for a given asset. It is observed that longer durations lead to lower volatility and shorter durations lead to higher volatility and durations are informative, see Engle (2000) and Dufour and Engle (2000). Gerhard and Hautsch (2002) estimate daily volatility for an intra-daily duration model at which irregularly time spaced volatility has been used at the aggregated level.

There are several articles that provide a detailed discussion of modeling volatility, see, for example, Andersen et al. (2005a, 2005b, 2006). Modeling and forecasting volatility based on intra-daily data has attracted a lot of research interest. Andersen et al. (2001) improve the inference procedures for using intra-daily data forecasts. Bollerslev and Wright (2001) propose a method to model volatility dynamics by fitting an autoregressive model to log-squared, squared or absolute returns of intra-daily data. They show that when working with intra-daily data, using a simple autoregressive model can provide a better prediction for forecasting future volatility than standard GARCH or exponential GARCH (EGARCH) models. They suggest that intra-daily data can be easily used to generate superior daily volatility forecasts. Blair et al. (2001) offer evidence that intra-daily returns provide much more accurate forecasts of realized volatility than daily returns. In the context of stochastic volatility models, Barndorff-Nielsen and Shephard (2002) investigate the statistical properties of realized volatility. Corsi et al. (2005) study the time-varying volatility of realized volatility. Using a GARCH diffusion process, Martens (2001) and Marten et al. (2002) point out that using intra-daily data can improve the out-of-sample daily volatility forecasts. Bollen and Inder (2002) use the vector autoregressive heteroskedasticity and autocorrelation (VARHAC) estimator to estimate daily realized volatility from intra-daily data. Andersen et al. (2003) propose a modeling framework for integrating intra-daily data to predict daily return volatility and return distribution. Thomakos and Wang (2003) investigate the statistical properties of daily realized volatility of futures contracts generated from intra-daily data. Using 5-minute intra-daily foreign exchange data, Morana and Beltratti (2004) illustrate the existence of structural breaks and long memory in the realized volatility process. Fleming et al. (2003) indicate that the economic value of the realized volatility approach is substantial and deliver several economic benefits for investment decision making. Andersen et al. (2005a) summarize the parametric and non-parametric methods used in volatility estimation and forecasting.

3.6 Pattern Recognition and Stylized Facts

The task of pattern recognition in high-frequency data mining is to classify data (patterns) based on either a priori knowledge or on statistical information extracted from the patterns. The patterns to be classified are usually groups of measurements or observations, defining points in an appropriate multidimensional space. A complete pattern recognition system consists of a sensor that gathers the observations to be classified or described; a feature extraction process that computes numeric or symbolic information from the observations; and a classification or description scheme that does the actual job of classifying or describing observations, relying on the extracted features. Duda et al. (2001) designed a pattern recognition system with different operation components, i.e., sensing, segmentation and grouping, feature extraction, classification, and post-processing.

The input to a pattern recognition system is often some kind of a transducer. Duda et al. (2001) pointed out that the difficulty of the problem usually depends on the characteristics and limitations of the transducer. Sensing could be regarded as a design of the sensors that gathers the observations for pattern recognition. The key point is quality of the sensors, i.e., the data quality.

The goal of segmentation and grouping is to simplify and/or change the representation of data into something that is more meaningful and easier to analyze. Segmentation and grouping is the process dividing a dataset into distinct subsets (segments/groups) that can be characterized in the same way or have similar features. Because each segment/group is fairly homogeneous in their characteristics, they are likely to feedback similarly to a given method for analysis.

Feature extraction involves simplifying the amount of resources required to describe a large set of data accurately. When performing analysis of complex data, one of the major problems stems from the number of variables involved. Analysis with a large number of variables generally requires a large amount of memory and computation power. In practise, it is important to improve computation power by constructing combinations of the variables to get around these problems while still describing the data with sufficient accuracy. Duda et al. (2001, page 11) point out that "the traditional goal of the feature extractor is to characterize an object to be recognized by measurements whose values are very similar for objects in the same category, and very different for objects in different categories", which "leads to the idea of seeking distinguishing features that are invariant to irrelevant transformations of the input".

Duda et al. (2001) think that the major concern of pattern recognition is the design of a classifier which assigns the object to a proper category by using the feature vector provided by a feature extractor in the stage of feature extraction. The abstraction provided by the feature vector representation of the input data enables such a classification. The difficulty of the classification goes with the variability in the feature values for objects in the same category relative to the difference between feature values for objects in different categories. Advanced quantitative methods are needed for the classification process.

Post-processing is the stage in which the classification is evaluated. The simplest measure of classifier performance is the classification error rate, i.e., the percentage of new patterns that are assigned to the wrong category, see Duda et al. (2001). In general, a classification with minimum classification error rate is preferred. But sometimes the minimum error rate classification requires intensive computations. It is important to balance the risk of high cost and minimum error rate.

In financial data analysis, an important task is to identify the statistical properties of the target data set, which shares similar task for pattern recognition. Those statistical properties are referred to as *stylized factors*. Stylized factors offer building blocks for further modeling in such a way so as to encompass these statistical properties. Being a full record of market transactions, intra-daily data have properties that have been observed. In this section, some stylized facts of intra-daily data will be reviewed.

3.6.1 Random Durations

For intra-daily data, irregularly spaced time intervals between successive observations is the salient feature compared with classical time series data (see, for example, Engle (2000) and Ghysels (2000)). Given time points equally spaced along the time line, for intra-daily data, at one time point, there might be no observation or several observations. In intra-day data, observations arrive at random time. This causes the duration between two successive observations not to be constant in a trading day. For a time series, if the time space (duration) is constant over time, it is an equally spaced time series or *homogeneous* time series. If the duration varies throught time, it is an unequally spaced time series, i.e., the *inhomogeneous* time series, see Dacorogna et al. (2001).

3.6.2 Distributional Properties of Returns

Many techniques in modern finance rely heavily on the assumption that the random variables under investigation follow a Gaussian distribution. However, time series observed in finance often deviate from the Gaussian model, in that their marginal distributions are found to possess heavy tails and are possibly asymmetric. Bollerslev et al. (1992) mention that intra-daily data exhibit fatter tails in the unconditional return distributions and Dacorogna et al. (2001) confirm the exhibition of heavy tails in intra-daily return data. In such situations, the appropriateness of the commonly adopted normal distribution assumption for returns is highly questionable in research involving intra-daily data. It is often argued that financial asset returns are the cumulative outcome of a vast number of pieces of information and individual decisions arriving almost continuously in time. Hence, in the presence of heavy tails, it is natural to assume that they are approximately governed by a non-Gaussian stable distribution, see Rachev and Mittnik (2000), and Rachev et al. (2005). Marinelli et al. (2000) first model the heavy tailedness in intra-daily data. Mittnik et al. (2002) point out that other leptokurtic distributions, including Student's t, Weibull, and hyperbolic, lack the attractive central limit property. Sun et al. (2006a) confirm the findings of heavy tailedness in intra-daily data. Wood et al. (1985) present evidence that stock returns are not independently and identically distributed. They find that the distributions are different for the return series in the first 30 minutes of the trading day, at the market close of the trading day, and during the remainder of the trading day.

3.6.3 Autocorrelation

The study by Wood et al. (1985), one of the earliest studies employing intra-daily data, finds that the trading day return series is non-stationary and is characterized by a low-order autoregressive process. For the volatility of asset returns, autocorrelation has been documented in the literature (see among others, Engle (1982), Baillie and Bollerslev (1991), Bollerslev et al. (1992), Hasbrouck (1996), and Bollerslev and Wright (2001)). The existence of negative first-order autocorrelation of returns at higher frequency, which disappears once the price formation is over, has been reported in both the foreign exchange market and equity market. The explanation for the finding of negative autocorrelation of stock returns observed by researchers is due to by what is termed the bid-ask bounce. According to the bid-ask bounce explanation, the probability of a trade executing at the bid price and then being followed by a trade executing at the ask price is higher than a trade at the bid price followed by another trade at the bid price (see Alexander (2001), Dacorogna et al. (2001), and Gourieroux and Jasiak (2001)).

3.6.4 Seasonality

Many intra-daily data display seasonality. By seasonality, I mean periodic fluctuations. Daily patterns in the trading day have been found in the markets for different types of financial assets (see, Jain and Joh (1988), McInish and Wood (1991), Bollerslev and Domowitz (1993), Engle and Russel (1998), Andersen and Bollerslev (1997), Bollerslev et al. (2000), and Veredas et al. (2002)). One such trading pattern is the well-known "U-shape" pattern of daily trading (see, Wood et al. (1985), Ghysels (2000), and Gourieroux and Jasiak (2001)). This trading pattern refers to the observation in a trading day that trade intensity is high at the beginning and at the end of the day, and trading durations increase and peak during lunch time. As a result, return volatility exhibits a U-shape where the two peaks are the beginning and the end of a trading day, with the bottom approximately during the lunch period. Hong and Wang (2000) confirm the U-shape patterns in the mean and volatility of returns over a trading day. They find that around the close and open there exist higher trading activity, the returns of open-to-open being more volatile than that of close-to-close. Besides the intra-day pattern, there exists a day-of-week pattern evidenced by both lower returns and higher volatility on Monday.

3.6.5 Clustering

Many financial time series display volatility clustering. It is observed that large returns are followed by more large returns and small returns by more small returns. Equity, commodity, and foreign exchange markets often exhibit volatility clustering at higher frequency. Volatility clustering becomes pronounced in intra-daily data (see, for example, Alexander (2001), Haas et al. (2004)). Besides volatility clustering, intra-daily data exhibit quote clustering and duration clustering. Quote or price clustering is the preference for some quote/prices over others. Duration clustering means that the long and short durations tend to occur in clusters (see, for example, Bauwens and Giot (2001), Chung and Van Ness (2004), Engle and Russell (1998), Feng et al. (2004), Huang and Stoll (2001), and Sun et al. (2006b)).

3.6.6 Long-range Dependence

Long-range dependence or *long memory* (sometimes also referred to as *strong dependence* or *persistence*) denotes the property of time series to exhibit persistent behavior. (A more precise mathematical definition will be provided in Section 5.) It is generally believed that when the sampling frequency increases for financial returns, long-range dependence will be more significant. Baillie (1996) discusses long memory processes and fractional integration in econometrics. Marinelli et al. (2000) propose subordinated modeling to capture long-range dependence and heavy tailedness. Several researchers, focusing both on theoretical and empirical issues, discuss long-range dependence. Doukhan et al. (2003), Robinson (2003), and Teyssiére and Kirman (2006) provide an overview of the important contributions to this area of research. Investigating the stocks comprising the German DAX, Sun et al. (2006a) confirm that long-range dependence does exist in intra-daily return data. I provide a more detailed discussion of long-range dependence in Chapter 4.

Chapter 4

Long Range Dependence and Fractal Processes

Long-range dependence is the dependence structure across long time periods. As stated earlier, it denotes the property of a time series to exhibit persistent behavior, i.e., a significant dependence between very distant observations and a pole in the neighborhood of the zero frequency of their spectrum. In the time domain, if $\{X_t, t \in T\}$ exhibits long-range dependence, its autocovariance function $\gamma(k)$ has the property of $\sum |\gamma(k)| = \infty$, where k measures the distance between two observations, i.e., the order of lags. In the frequency domain, if $\{X_t, t \in T\}$ exhibits long-range dependence, its spectral density $f(\lambda)$ $(-\pi < \lambda < \pi)$ has a "pole" at frequency zero, i.e., $f(0) = 1/2\pi \sum_{k=-\infty}^{\infty} \gamma(k) = \infty$.

4.1 Estimation and Detection of LRD in the Time Domain

4.1.1 The Rescaled Adjusted Range Approach

The rescaled adjusted range method, denoted by R/S, was proposed by Hurst (1951) and discussed in detail in Mandelbrot and Wallis (1969), Mandelbrot (1975), Mandelbrot and Taqqu (1979), and Beran (1994). For a time series, $\{X_t, t \geq 1\}$, let $Y_T = \sum_{t=1}^{T} X_t$ and

$$S^2(t,k) = \frac{1}{k} \sum_{i=t+1}^{t+k} (X_i - \overline{X}_{t,k})^2 \tag{4.0-1}$$

where $\overline{X}_{t,k} = k^{-1} \sum_{i=t+1}^{t+k} X_i$, then define the adjusted range

$$R(t,k) = \max_{0 \leq i \leq k} \Big[Y_{t+i} - Y_t - \frac{i}{k}(Y_{t+k} - Y_t) \Big]$$

$$- \min_{0 \le i \le k} \Big[Y_{t+i} - Y_t - \frac{i}{k}(Y_{t+k} - Y_t) \Big] \tag{4.0-2}$$

the standardized ratio $R(t,k)/S(t,k)$ is the rescaled adjusted range, i.e., the R/S statistic. Hurst observed that for a large k based on the Nile River data, $\log E[R/S] \approx a + H \log k$ with $H > 0.5$. To determine H by using the R/S statistic, I can do the following:

1. Divide the time series of length N into K blocks.

2. For each lag t, starting at points $t_i = iT/K + 1$, compute $R(k_i, t)/S(k_i, t)$, $i = 1, 2, \cdots$, for all possible k such that $t_i + k \le N$.

3. Plot its logarithm against the logarithm of k. This plot is sometimes called the *pox plot* for the R/S statistic.

4. The parameter H is the estimated slope of the line in the pox plot.

The R/S method requires cutting off both the low and high end of the plot to make reliable estimates. The low end of the plot stands for the short-range dependence in the time series and there are too few points on the high end. In the literature it is also argued that the R/S cannot provide the confidence intervals for the estimates and cannot discriminate slight LRD from no LRD. Compared with other methods, the R/S approach is less efficient. When the time series is non-stationary and departs from the normal distribution, this method is not robust, see Lo (1991), Taqqu et al. (1995), and Taqqu and Teverovsky (1998).

Lo modifies the R/S approach and proposes a test procedure for the null hypothesis of no LRD. In Lo's method, he suggests using a weighted sum of autocovariance for S instead of the sample standard deviation to normalize R. Meantime, his modification suggests not considering multiple lags but only using the length N of the series, i.e.,

$$\begin{aligned} S_q^2(N) &= \frac{1}{N}\sum_{j=1}^{N}(X_j - \overline{X}_N)^2 \\ &+ \frac{2}{N}\sum_{j=1}^{q}\omega_j(q)\Big(\sum_{i=j+1}^{N}(X_i - \overline{X}_N)(X_{i-j} - \overline{X}_N)\Big) \end{aligned} \tag{4.0-3}$$

where $\overline{X}_N$ denotes the sample mean of the time series, and $\omega_j(q) := 1 - \frac{j}{q+1}$, $q < N$. I can use the following term to represent $S_q(N)$ by adding the weighted sample autocovariances to the sample variance, i.e.,

$$S_q^2(N) = S^2 + 2\sum_{j=1}^{q}\omega_j(q)\hat{\gamma}_j \tag{4.0-4}$$

where $\hat{\gamma}_j$ are the sample autocovariances. Lo shows that the distribution of the statistic

$$V_q(N) := \frac{N^{-1/2}R(N)}{S_q(N)} \tag{4.0-5}$$

is asymptotic to

$$W_1 = \max_{0\leq t\leq 1} W_0(t) - \min_{0\leq t\leq 1} W_0(t) \tag{4.0-6}$$

where W_0 is the standard Brownian bridge. This fact allows the computation of a 95% confidence interval for W_1. Thus, Lo uses the interval $[0.809, 1.862]$ as the asymptotic 95% acceptance region of the null hypothesis of no LRD.

Since Lo only provides the method to test if LRD is present or not without suggesting an estimator of H, Teverovsky et al. (1999) modified Lo's method to get an estimator of H. They suggest using V_q with a wide range of values of q, and then ploting the estimates as was done for the pox plot.

4.1.2 ARFIMA Model

Conventional analysis of time series under the stationarity assumption typically relies on the standard integrated autoregressive moving average model, i.e. ARIMA model of following form:

$$\alpha(L)(1-L)^d X_t = \beta(L)\varepsilon_t \tag{4.0-7}$$

where, $\varepsilon_t \sim (0, \sigma^2)$, and $\alpha(L)$ is the autoregressive polynomials in the lag operator L such that $\alpha(L) = 1 - \alpha_1(L)-, ..., -\alpha_p(L)^p$. $\beta(L)$ is the moving average polynomials in the lag operator L such that $\beta(L) = 1 + \beta_1(L)-, ..., -\beta_q(L)^q$. All roots of $\alpha(L)$ and $\beta(L)$ lie outside the unit circle. Granger and Joyeux (1980) and Hosking (1981) generalize d to a non-integer value by the fractional differencing operator defined by

$$(1-L)^d = \sum_{k=0}^{\infty} \frac{\Gamma(k-d)L^k}{\Gamma(-d)\Gamma(k+1)} \tag{4.0-8}$$

where $\Gamma(\cdot)$ is the gamma function. That is the Autoregressive Fractional Integrated Moving Average (ARFIMA) model allows a non-integer value of d. The ARFIMA model has the ability to capture significant dependence between distant observations compared with the ARIMA model. Hosking (1981) shows that the autocorrelation function $\rho(k)$ of an ARFIMA process has a slower hyperbolic decay pattern, that is $\rho(k) \sim k^{2d-1}$, with $d < 0.5$ when $k \to \infty$ and the autocorrelation of ARIMA decay follows an exponential pattern, that is, $\rho(k) \sim r^k$, with $r \in (0, 1)$ when $k \to \infty$. The memory of the time series is captured by d, therefore the existence of long memory can be tested based on the statistical significance of the fractional differencing parameter d.

For the first difference of the series Y_t, $Y_t = (1-L)X_t$, the equation

$$(1-L)^d X_t = \alpha^{-1}(L)\beta(L)\varepsilon_t = u_t \tag{4.0-9}$$

can be used to estimate d. In this case, the Hurst index is $1/2 + d$.

4.1.3 Variance-Type Method

Teverovsky and Taqqu (1995), Taqqu et al. (1995), and Taqqu and Teverovsky (1996) discuss the variance-type methods for estimating the Hurst index: the aggregated variance method and the differenced variance method.

For the *aggregated variance method*, the variance of X is of order N^{2H-2} suggesting:

1. For an integer m between 2 and $N/2$, divide a given series of length N into blocks of length m, and compute the sample mean over each k-th block.

   $$\overline{X}_k^{(m)} := \frac{1}{m} \sum_{t=(k-1)m+1}^{km} X_t$$

   where $k = 1, 2, \cdots, [N/m]$.

2. For each m, compute the sample variance of $\overline{X}_k^{(m)}$,

   $$\overline{s}_m^2 := \frac{1}{[N/m]-1} \sum_{k=1}^{[N/m]} (\overline{X}_k^{(m)} - \overline{X})^2.$$

3. Plot $\log s_m^2$ against $\log m$.

4. For large values of m, the result should be a straight line with a slope of $2H-2$. Then the slope can be estimated by fitting a least-squares line in the log-log plot. If the series has no long-range dependence and finite variance, then $H = 0.5$ and the slope of the fitted line is -1.

There are two types of non-stationarity, one is jumps in the mean and the other is slowly declining trends. Teverovsky and Taqqu (1995) distinguish these from long-range dependence by using the *differenced variance method*. They difference the variance and study the sample variance $\overline{s}_{m_i+1}^2 - \overline{s}_{m_i}^2$, where m_i are the successive values of m as defined above. Using it together with the original aggregated variance method, the difference variance method can detect the presence of the two types of non-stationary effects mentioned above.

Another method involving the variance is the variance of residuals method introduced by Peng et al. (1994). Similar to the aggregated variance method, the series is divided into blocks with size of m. Next, the partial sums are computed within each block, i.e.,

$$Y(k)^{(m)} := \sum_{t=(k-1)m+1}^{km} X_t$$

where $k = 1, 2, \cdots, [N/m]$. A regression line $a + bk$ is fitted to the partial sums within each block, and the sample variance of the residuals $\overline{s}^{(m)}$ is computed. Taqqu et al. (1995) prove that in the Gaussian case, the variance of residuals is proportional to m^{2H}. By plotting $\log \overline{s}^{(m)}$ against $\log m$, the slope, i.e., $2H$, can be estimated (see Taqqu and Teverovsky (1996) for more details).

4.1.4 Absolute Moments Method

The *absolute moments method* is a generalization of the aggregated variance method. Using this method

1. For an integer m between 2 and $N/2$, divide a given series of length N into blocks of length m, and compute the sample mean over each k-th block.

$$\overline{X}_k^{(m)} := \frac{1}{m} \sum_{t=(k-1)m+1}^{km} X_t$$

   where $k = 1, 2, \cdots, [N/m]$.

2. For each m, compute the n-th absolute moment of $\overline{X}_k^{(m)}$,

$$AM_n^{(m)} = \frac{1}{[N/m]-1} \sum_{k=1}^{[N/m]} |\overline{X}_k^{(m)} - \overline{X}|^n.$$

3. The $AM_n^{(m)}$ is asymptotically proportional to $m^{n(H-1)}$.

4. Plot $\log AM_n^{(m)}$ against $\log m$.

5. For large values of m, the result should be a straight line with a slope of $n(H-1)$. Then the slope can be estimated by fitting a least-squares line in the log-log plot. If the series has no long-range dependence and finite variance, then $H = 0.5$ and the slope of the fitted line is $-n/2$.

Similar to the absolute moments method, Higuchi (1988) suggests the fractal dimension method, see Taqqu and Teverovsky (1998). The difference between these two methods is that the absolute moments method (when $n = 1$) uses a moving window to compute the aggregated series, while the fractal dimension method uses the non-intersecting blocks. The fractal dimension method requires intensive computation and increases accuracy in shorter time series. Taqqu et al. (1995), and Taqqu and Teverovsky (1998) discuss these two methods in detail.

4.2 Estimation and Detection of LRD in the Frequency Domain

4.2.1 Periodogram Method

Geweke and Porter-Hudak (1983) introduced a semi-nonparametric procedure to test long memory based on the slope of spectral density around the angular frequency $\omega = 0$. For the periodogram of X_t at frequency ω_j, i.e., $g(\omega)$, which is defined as follows

$$g(\omega) = \frac{1}{2\pi T}\left|\sum_{t=1}^{T} e^{it\omega}(X_t - \overline{X})\right|^2 \tag{4.0-10}$$

the differencing parameter d can be consistently estimated by the regression

$$\ln g(\omega) = c - d\ln(4\sin^2(\frac{\omega_j}{2})) + \eta_j, \quad j = 1, 2, ..., n \tag{4.0-11}$$

where $\omega_j = 2\pi j/T, (j = 1, 2, \cdots, T-1)$ denotes the Fourier frequencies of the sample, T is the sample size, and $n = f(T) << T$ is the number of Fourier frequencies included in the spectral regression. As Geweke and Porter-Hudak (1983) show, the slope of the line in log-log plot is $1 - 2H$.

Extensions and improvements to the periodogram method, for example, the continuous periodogram method and the averaged (cumulative) periodogram method, have been discussed in the literature, see, for example, Robinson (1995a), Taqqu et al. (1995), Taqqu and Teverovsky (1998), Moulines and Soulier (1999), and Hurvich and Brodski (2001).

4.2.2 Whittle-Type Methods

The *Whittle estimator* is the extension of the periodogram method. If the time series X_t follows a Gaussian distribution, the Gaussian maximum likelihood estimate (MLE) might have optimal asymptotic statistical properties and can be used for approximation, see Whittle (1951) and Hannan (1973). The periodogram of X_t at frequency ω_j is defined as $g(\omega)$,

$$g(\omega) = \frac{1}{2\pi T} \left| \sum_{t=1}^{T} X_t e^{it\omega} \right|^2 \tag{4.0-12}$$

which is an estimator of the spectral density. It is evaluated at the Fourier frequencies $\omega_j = 2\pi jT$.

Beran (1994) shows that the following equation is an approximation to the Gaussian likelihood in terms of the periodogram $I(\cdot)$,

$$L_W(\theta) = -\frac{1}{2\pi} \sum_{j=1}^{[T/2]} \log f_\theta(\omega_j) + \frac{I_N(\omega_j)}{f_\theta(\omega_j)}$$

and the Whittle estimator is found by minimizing it for a given parametric spectral density $f_\theta(\omega)$. The Gaussian likelihood can be replaced by different approximations without affecting first-order limit distributional characteristics. Robinson (2003) shows that the estimates which maximize such approximations are all $\sqrt{n}$-consistent and of the same limit normal distribution as the Gaussian MLE. Fox and Taqqu (1986) show the Whittle estimate $\hat{H}$ of H is asymptotically normal with rate of convergence $T^{1/2}$ and the asymptotic distribution of $\sqrt{T}(\hat{H} - H)$ is Gaussian. Relaxing the Gaussian assumption, Giraitis and Surgailis (1990) discuss the properties of the Whittle estimate $\hat{H}$ of H.

Robinson (1995b) develops the *local Whittle method* and Taqqu and Teverovsky (1998) provide further discussion about it. The local Whittle method is a semi-parametric estimator. It only specifies the parametric form of the spectral density with ω approaching zero. It assumes that

$$f_{c,H}(\omega) = c\,\omega^{1-2H}$$

for frequencies ω close to the origin. One estimate minimizes

$$\sum_{j=1}^{m} \log f_{c,H}(\omega_j) + \frac{I_T(\omega_j)}{f_{c,H}(\omega_j)}$$

with respect to c and H for some $m < [T/2]$, T being the length of the data.

Taqqu and Teverovsky (1998) introduce the *aggregated Whittle method* which provides a robust Whittle estimator without considering exact parametric information about the spectral density. It can be used for longer time series. This method suggests aggregating the data to create a shorter series.

$$X_i := \frac{1}{m} \sum_{t=m(i-1)+1}^{mt} X_t$$

If the aggregation level of m is high enough and long-range dependence occurs, then the new series will approach a fractional Gaussian noise. In the finite variance case, the Whittle estimator can increase the estimation accuracy with an underlying fractional Gaussian noise assumption.

4.3 Econometric Modeling of LRD

Several econometric models have been extended to describe long-range dependence, for example, extending the ARMA model to the ARFIMA model discussed in the previous section. In this section, I introduce four types of extensions, the GARCH-type extension, the stochastic volatility type extension, the unit root type extension, and the regime switching type extension.

4.3.1 GARCH-Type Extension

Robinson (1991) suggests extending the GARCH model by using fractional differences in order to accommodate the existence of long-range dependence. The fractional differencing operator is defined as in equation (4.0-7) by Baillie et al. (1996). The fractionally integrated GARCH (FIGARCH) model is then

$$(1-L)^d\beta(L)(h_t^2-\mu)=\alpha(L)(r_t^2-\mu) \tag{4.0-13}$$

For a well-defined process, the parameters α_j, β_j, and d are constrained. Then the coefficients θ_j are all nonnegative in

$$\begin{aligned} h_t^2 &= \mu + (1-L)^{-d}\,\alpha(L)\,\beta^{-1}(L)\,r_t^2 \\ &= \mu + \sum_{j=0}^{\infty}\theta(r_{t-1-j}^2-\mu) \end{aligned} \tag{4.0-14}$$

Implied by equation (4.0-14), the parameters α_j and β_j are constrained as in the standard GARCH model. This also implies that the parameter d is constrained to be positive. Breidt et al. (1998) argue that the r_t in equation (4.0-14) is not covariance stationary and the autocovariance function of r_t is not defined. Bollerslev and Mikkelsen (1996) formulate a fractionally integrated EGARCH model of the following form

$$\log h_t^2=\mu_t + \theta(L)\phi(L)^{-1}(1-L)^{-d}g(\epsilon_{t-1}) \tag{4.0-15}$$

where $\theta(z)=1+\theta_1 z+,\cdots,+\theta_p z^p$ for $|z|\le 1$ is an autoregressive polynomial, and $\phi(z)=1-\phi_1 z-,\cdots,-\phi_p z^p$ is a moving average polynomial and $\phi(z)$ has no roots in common with $\theta(z)$. The fractional integrated EGARCH model gives a strictly stationary and ergodic process. Nelson (1991) shows that $(\log h_t^2-\mu_t)$ is covariance stationary and $d<0.5$.

4.3.2 Stochastic Volatility Type Extension

A long-range dependence stochastic volatility model is discussed by Breidt et al. (1998). The stochastic volatility model is defined by

$$r_t = h_t \epsilon_t, \quad h_t = h \exp(u_t/2),$$

where u_t is independent of ϵ_t, ϵ_t is independent and identically distributed (i.i.d.) with mean zero and variance one. u_t in a simple long-range dependence model can be defined as

$$(1-L)^d u_t = \eta_t$$

where η_t follows i.i.d. normal distribution with zero mean and variance σ_η^2, and $d \in (-0.5, 0.5)$. For long-range dependence, u_t can be expressed as an ARFIMA (p, d, q) process, defined as

$$(1-L)^d \phi(L) u_t = \theta(L) \eta_t$$

where η_t follows i.i.d. normal distribution with zero mean and variance σ_η^2.

4.3.3 Unit Root Type Extension

Robinson (1994) considers the following model that nests a unit root model in order to grasp the effect of long-range dependence:

$$\phi(L) r_t = \epsilon_t, \quad t \geq 1, \tag{4.0-16}$$

$$r_t = 0, \quad t \leq 0, \tag{4.0-17}$$

where ϵ_t is an $I(0)$ (i.e., integrated process of order zero) with parametric autocorrelation and

$$\phi(L) = (1-L)^{d_1} (1+L)^{d_2} \prod_{j=3}^{n} (1 - 2\cos\omega_j L + L^2)^{d_j} \tag{4.0-18}$$

where ω_j are given distinct real numbers in $(0, \pi)$, and the d_j, $0 \leq j \leq n$, are arbitrary real numbers. This model also covers seasonal and cyclical components. Velasco and Robinson (2000) propose the following model

$$(1-L)^s r_t = \epsilon_t, \quad t \geq 1, \tag{4.0-19}$$

$$r_t = 0, \quad t \leq 0, \tag{4.0-20}$$

$$(1-L)^{d-s} \epsilon_t = u_t, \quad t = 0, \pm 1, \cdots, \tag{4.0-21}$$

where s is the integer part of $d + 1/2$ and u_t is a parametric $I(0)$ process. ϵ_t is a stationary $I(d-s)$ process. Marinucci and Robinson (1999) discuss the difference between the two models given by equations (4.0-16)-(4.0-17) and equations (4.0-19)-(4.0-21) with respect to the two definitions of nonstationarity $I(d)$ processes.

4.3.4 Regime Switching Type Extension

Diebold and Inoue (2001) show that long-range dependence models and regime switching models are intimately related in several circumstances, including a simple mixture model, stochastic permanent break model, and Markov-switching model. They demonstrate that with suitably adapted time varying transition probabilities these regime switching models can generate an autocovariance structure. This autocovariance structure is similar to the fractionally integrated processes. Banerjee and Urga (2005) provide an overview of the recent development in the studies of modeling regime switching and long-range dependence.

Haldrup and Nielsen (2006) propose a regime switching multiplicative seasonal ARFIMA model that accommodates both fractional integration and regime switching simultaneously. The model is:

$$A_{s_t}(L)\,(1-\alpha_{s_t}L^{24})(1-L)^{d_{s_t}}(y_t-\mu_{s_t}) = \epsilon_{s_t,t}, \quad \epsilon_{s_t,t}\sim N(0,\sigma^2_{s_t})$$

where $A_{s_t}(L)$ is an eighth order lag polynomial capturing the within-the-day effects. The polynomial $(1-\alpha_{s_t}L^{24})$ stands for a daily quasi-difference filter, $s_t = 0, 1$ denotes the regime determined by a Markov chain with transition probabilities, d_{s_t} stands for the order of difference, and μ is the mean.

4.4 Fractal Processes and Long-Range Dependence

Fractal processes (self-similar processes) are tightly connected with the analysis of long-range dependence. Self-similar processes are invariant in distribution with respect to changes of time and space scale. The scaling coefficient or self-similarity index is a non-negative number denoted by H, the Hurst parameter. If $\{X(t+h)-X(h), t\in T\} \stackrel{d}{=} \{X(t)-X(0), t\in T\}$ [1] for all $h\in T$, the real-valued process $\{X(t), t\in T\}$ has stationary increments. Samorodnisky and Taqqu (1994) provide a succinct expression of self-similarity: $\{X(at), t\in T\} \stackrel{d}{=} \{a^H X(t), t\in T\}$. The process $\{X(t), t\in T\}$ is called *H-sssi* if it is self-similar with index H and has stationary increments. Long-range dependence processes are asymptotically second-order self-similar (see, Willinger et al. (1998)).

4.4.1 Specification of the Fractal Processes

Lamperti (1962) first introduced semi-stable processes (which we nowadays call self-similar processes). Let T be either $R, R_+ = \{t : t \geq 0\}$ or $\{t : t > 0\}$. Then the real-valued

[1] "$\stackrel{d}{=}$" means equality in distribution.

process $\{X(t), t \in T\}$ is self-similar with Hurst index $H > 0$ (*H-ss*) for any $a > 0$ and $d \geq 1$, $t_1, t_2, ..., t_d \in T$, satisfying:

$$\Big(X(at_1), X(at_2), ..., X(at_d)\Big) \stackrel{d}{=} \Big(a^H X(t_1), a^H X(t_2), ...a^H X(t_d)\Big). \tag{4.0-22}$$

Fractional Gaussian Noise

For a given $H \in (0,1)$ there is basically a single Gaussian *H-sssi* process, namely fractional Brownian motion (fBm) that was first introduced by Kolmogorov (1940). Mandelbrot and Wallis (1968) and Taqqu (2003) clarify the definition of fBm as a Gaussian *H-sssi* process $\{B_H(t)\}_{t \in R}$ with $0 < H < 1$. Mandelbrot and van Ness (1968) defined the stochastic representation

$$B_H(t) := \frac{1}{\Gamma(H+\frac{1}{2})}\Big(\int_{-\infty}^{0}[(t-s)^{H-\frac{1}{2}} - (-s)^{H-\frac{1}{2}}]dB(s) + \int_0^t (t-s)^{H-\frac{1}{2}}dB(s)\Big) \tag{4.0-23}$$

where $\Gamma(\cdot)$ represents the Gamma function:

$$\Gamma(a) := \int_0^\infty x^{a-1}e^{-x}dx$$

and $0 < H < 1$ is the Hurst parameter. The integrator B is the ordinary Brownian motion. The main difference between fractional Brownian motion and ordinary Brownian motion is that the increments in Brownian motion are independent while in fractional Brownian motion they are dependent. As to the fractional Brownian motion, Samorodnitsky and Taqqu (1994) define its increments $\{Y_j, j \in Z\}$ as fractional Gaussian noise (fGn), which is, for $j = 0, \pm1, \pm2, ...$, $Y_j = B_H(j-1) - B_H(j)$.

Fractional Stable Noise

Fractional Brownian motion can capture the effect of long-range dependence, but has less power to capture heavy tailedness. The existence of abrupt discontinuities in financial data, combined with the empirical observation of sample excess kurtosis and unstable variance, confirms the stable Paretian hypothesis first identified by Mandelbrot (1963, 1983). It is natural to introduce stable Paretian distribution in self-similar processes in order to capture both long-range dependence and heavy tailedness. Samorodinitsky and Taqqu (1994) introduce the α-stable *H-sssi* processes $\{X(t), t \in R\}$ with $0 < \alpha < 2$. If $0 < \alpha < 1$, the Hurst parameter values are $H \in (0, 1/\alpha]$ and if $1 < \alpha < 2$, the Hurst parameter values are $H \in (0, 1]$. There are many different extensions of fractional Brownian motion to the stable distribution. The most commonly used is the *linear fractional stable motion* (also called *linear fractional Lévy motion*), $\{L_{\alpha,H}(a,b;t), t \in (-\infty,\infty)\}$, which is defined by Samorodinitsky and Taqqu (1994) as follows:

$$L_{\alpha,H}(a,b;t) := \int_{-\infty}^{\infty} f_{\alpha,H}(a,b;t,x)M(dx) \tag{4.0-24}$$

where

$$\begin{aligned} f_{\alpha,H}(a,b;t,x) := & \, a\Big((t-x)_+^{H-\frac{1}{\alpha}} - (-x)_+^{H-\frac{1}{\alpha}}\Big) \\ & +b\Big((t-x)_-^{H-\frac{1}{\alpha}} - (-x)_-^{H-\frac{1}{\alpha}}\Big) \end{aligned} \tag{4.0-25}$$

and where $x_+ = \max(x,0), x_- = \min(x,0)$, a, b are real constants, $|a|+|b| > 1$, $0 < \alpha < 2$, $0 < H < 1$, $H \neq 1/\alpha$, and M is an α-stable random measure on R with Lebesgue control measure and skewness intensity $\beta(x)$, $x \in (-\infty,\infty)$ satisfying: $\beta(\cdot) = 0$ if $\alpha = 1$. They define linear fractional stable noises expressed by $Y(t)$, and $Y(t) = X_t - X_{t-1}$,

$$\begin{aligned} Y(t) &= L_{\alpha,H}(a,b;t) - L_{\alpha,H}(a,b;t-1) \\ &= \int_R \Big(a\Big[(t-x)_+^{H-\frac{1}{\alpha}} - (t-1-x)_+^{H-\frac{1}{\alpha}}\Big] \\ &+ b\Big[(t-x)_-^{H-\frac{1}{\alpha}} - (t-1-x)_-^{H-\frac{1}{\alpha}}\Big]\Big) M(dx) \end{aligned} \tag{4.0-26}$$

where $L_{\alpha,H}(a,b;t)$ is a linear fractional stable motion defined by equation (4.0-24), and M is a stable random measure with Lebesgue control measure given $0 < \alpha < 2$. In this chapter, if there is no special indication, the fractional stable noise (fsn) is generated from a linear fractional stable motion.

Some properties of these processes have been discussed in Mandelbrot and Van Ness (1968), Maejima (1983), Maejima and Rachev (1987), Manfields et al. (2001), Rachev and Mittnik (2000), Rachev and Samorodnitsky (2001), Samorodnitsky (1994, 1996, 1998), and Samorodinitsky and Taqqu (1994).

4.4.2 Estimation of Fractal Processes

Estimating the Self-Similarity Parameter in Fractional Gaussian Noise

Beren (1994) discusses the Whittle estimation (which I discussed earlier) of the self-similarity parameter. For fractional Gaussian noise, Y_t, let $f(\lambda; H)$ denote the power spectrum of Y_t after being normalized to have variance 1 and let $I(\lambda)$ the periodogram of Y_t, that is

$$I(\lambda) = \frac{1}{2\pi N} \left| \sum_{t=1}^{N} Y_t \, e^{i\,t\,\lambda} \right|^2 \tag{4.0-27}$$

The Whittle estimator of H is obtained by finding $\hat{H}$ that minimizes

$$g(\hat{H}) = \int_{-\pi}^{\pi} \frac{I(\lambda)}{f(\lambda; \hat{H})} d\lambda \tag{4.0-28}$$

Estimating the Self-Similarity Parameter in FSN

Stoev et al. (2002) proposed the least-squares (LS) estimator of the Hurst index based on the finite impulse response transformation (FIRT) and wavelet transform coefficients of the fractional stable motion. A FIRT is a filter $v = (v_0, v_1, ..., v_p)$ of real numbers $v_t \in \Re, t = 1, 2, ..., p$, and length $p+1$. It is defined for X_t by

$$T_{n,t} = \sum_{i=0}^{p} v_i \, X_{n(i+t)} \tag{4.0-29}$$

where $n \geq 1$ and $t \in N$. The $T_{n,t}$ are the FIRT coefficients of X_t, that is, the FIRT coefficients of the fractional stable motion. The indices n and t can be interpreted as "scale" and "location". If $\sum_{i=0}^{p} i^r v_i = 0$, for $r = 0, ..., q-1$, but $\sum_{i=0}^{p} i^q v_i \neq 0$, the filter v_i can be said to have $q \geq 1$ zero moments. If $\{T_{n,t}, n \geq 1, t \in N\}$ are the FIRT coefficients of fractional stable motion with the filter v_i that have at least one zero moment, Stoev et al. (2002) prove the following two properties of $T_{n,t}$: (1) $T_{n,t+h} \stackrel{d}{=} T_{n,t}$, and (2) $T_{n,t} \stackrel{d}{=} n^H T_{1,t}$, where $h, t \in N$ and $n \geq 1$. I suppose that $T_{n,t}$ are available for the fixed scales n_j $j = 1, ..., m$ and locations $t = 0, ..., M_j - 1$ at the scale n_j, since only a finite number, say M_j, of the FIRT coefficients are available at the scale n_j.

By using these properties, I have

$$E \log |T_{n_j,0}| = H \log n_j + E \log |T_{1,0}| \tag{4.0-30}$$

The left-hand side of this equation can be approximated by

$$Y_{\log}(M_j) = \frac{1}{M_j} \sum_{t=0}^{M_j-1} \log |T_{n_j,t}| \tag{4.0-31}$$

Then I get

$$\begin{pmatrix} Y_{\log}(M_1) \\ \vdots \\ Y_{\log}(M_m) \end{pmatrix} = \begin{pmatrix} \log n_1 & 1 \\ \vdots & \vdots \\ \log n_m & 1 \end{pmatrix} \begin{pmatrix} H \\ E \log |T_{1,0}| \end{pmatrix} + \begin{pmatrix} \sqrt{M}\Big(Y_{\log}(M_1) - E \log |T_{n_1,0}|\Big) \\ \vdots \\ \sqrt{M}\Big(Y_{\log}(M_m) - E \log |T_{n_m,0}|\Big) \end{pmatrix} \tag{4.0-32}$$

In short, I can express the above equation as follows

$$Y = X\theta + \frac{1}{\sqrt{M}}\varepsilon \tag{4.0-33}$$

Equation (4.0-33) shows that the self-similarity parameter H can be estimated by a standard linear regression of the vector Y against the matrix X. Stoev et al. (2002) explain this procedure.

Estimating the Parameters of Stable Paretian Distribution

The stable Paretian distribution requires four parameters for complete description: an index of stability $\alpha \in (0, 2]$ (also called the tail index), a skewness parameter $\beta \in [-1, 1]$, a scale parameter $\gamma > 0$, and a location parameter $\zeta \in \Re$. There is unfortunately no closed-form expression for the density function and distribution function of a stable distribution. Rachev and Mittnik (2000) give the definition of the stable distribution: A random variable X is said to have a stable distribution if there are parameters $0 < \alpha \leq 2$, $-1 \leq \beta \leq 1$, $\gamma \geq 0$ and real ζ such that its characteristic function has the following form:

$$E\exp(i\theta X) = \begin{cases} \exp\{-\gamma^\alpha|\theta|^\alpha(1 - i\beta(\text{sign}\theta)\tan\frac{\pi\alpha}{2}) + i\zeta\theta\} & if \quad \alpha \neq 1 \\ \exp\{-\gamma|\theta|(1 + i\beta\frac{2}{\pi}(\text{sign}\theta)\ln|\theta|) + i\zeta\theta\} & if \quad \alpha = 1 \end{cases} \tag{4.0-34}$$

and,

$$\text{sign}\,\theta = \begin{cases} 1 & if \quad \theta > 0 \\ 0 & if \quad \theta = 0 \\ -1 & if \quad \theta < 0 \end{cases} \tag{4.0-35}$$

Stable density is not only supported for all of $(-\infty, +\infty)$, but also for a half line. For $0 < \alpha < 1$ and $\beta = 1$ or $\beta = -1$, the stable density is only for a half line.

In order to estimate the parameters of the stable distribution, the maximum likelihood estimator given in Rachev and Mittnik (2000) has been employed. Given N observations, $X = (X_1, X_2, \cdots, X_N)'$ for the positive half line. The log-likelihood function is of the form

$$\ln(\alpha, \lambda; X) = N\ln\lambda + N\ln\alpha + (\alpha - 1)\sum_{i=1}^{N}\ln X_i - \lambda\sum_{i=1}^{N}X_i^\alpha \tag{4.0-36}$$

which can be maximized using, for example, a Newton-Raphson algorithm. It follows from the first-order condition,

$$\lambda = N\left(\sum_{i=1}^{N}X_i^\alpha\right)^{-1} \tag{4.0-37}$$

that the optimization problem can be reduced to finding the value for α which maximizes the concentrated likelihood

$$\ln^*(\alpha; X) = \ln\alpha + \alpha\nu - \ln\left(\sum_{i=1}^{N} X_i^{\alpha}\right) \tag{4.0-38}$$

where $\nu = N^{-1}\Sigma_{i=1}^{N}\ln X_i$. The information matrix evaluated at the maximum likelihood estimates, denoted by $I(\hat{\alpha}, \hat{\lambda})$, is given by

$$I(\hat{\alpha}, \hat{\lambda}) = \begin{pmatrix} N\hat{\alpha}^{-2} & \sum_{i=1}^{N} X_i^{\hat{\alpha}}\ln X_i \\ \sum_{i=1}^{N} X_i^{\hat{\alpha}}\ln X_i & N\hat{\lambda}^{-2} \end{pmatrix}$$

It can be shown that, under fairly mild condition, the maximum likelihood estimates $\hat{\alpha}$ and $\hat{\lambda}$ are consistent and have asymptotically a multivariate normal distribution with mean $(\alpha, \lambda)'$ (see Rachev and Mittnik (2000)).

Other methods for estimating the parameters of a stable distribution (i.e., the method of moments based on the characteristic function, the regression-type method, and the fast Fourier transform method) are discussed in Stoyanov and Racheva-Iotova (2004a, 2004b, 2004c).

4.4.3 Simulation of Fractal Processes

Simulation of Fractional Gaussian Noise

Paxson (1997) provides a method to generate the fractional Gaussian noise by using the Discrete Fourier Transform of the spectral density. Bardet et al. (2003) describe a concrete simulation procedure based on this method that overcomes some of the implementation issues encountered in practice. The procedure is:

1. Choose an even integer M. Define the vector of the Fourier frequencies $\Omega = (\theta_1, ..., \theta_{M/2})$, where $\theta_t = 2\pi t/M$ and compute the vector $F = f_H(\theta_1), ..., f_H(\theta_{M/2})$, where

   $$f_H(\theta) = \frac{1}{\pi}\sin(\pi H)\Gamma(2H+1)(1-\cos\theta)\sum_{t\in\aleph}|2\pi t + \theta|^{-2H-1}$$

   $f_H(\theta)$ is the spectral density of fGn.

2. Generate $M/2$ i.i.d exponential $Exp(1)$ random variables $E_1, ..., E_{M/2}$ and $M/2$ i.i.d uniform $U[0,1]$ random variables $U_1, ..., U_{M/2}$.

3. Compute $Z_t = \exp(2i\pi U_t)\sqrt{F_t E_t}$, for $t = 1, ..., M/2$.

4. Form the M-vector: $\tilde{Z} = (0, Z_1, ..., Z_{(M/2)-1}, Z_{M/2}, \overline{Z}_{(M/2)-1}, ..., \overline{Z}_1)$.

5. Compute the inverse FFT of the complex Z to obtain the simulated sample path.

Simulation of Fractional Stable Noise

Replacing the integral in equation (4.0-26) with a Riemann sum, Stoev and Taqqu (2004) generate the approximation of fractional stable noise. They introduce parameters $n, N \in \aleph$, then express the fractional stable noise $Y(t)$ as

$$Y_{n,N}(t) := \sum_{j=1}^{nN} \Big((\frac{j}{n})_+^{H-1/\alpha} - (\frac{j}{n} - 1)_+^{H-1/\alpha} \Big) L_{\alpha,n}(nt - j) \tag{4.0-39}$$

where $L_{\alpha,n}(t) := M_\alpha((j+1)/n) - M_\alpha(j/n)$, $j \in \Re$. The parameter n is the mesh size and the parameter M is the cut-off of the kernel function.

Stoev and Taqqu (2003) describe an efficient approximation involving the Fast Fourier Transformation (FFT) algorithm for $Y_{n,N}(t)$. Consider the moving average process $Z(m)$, $m \in \aleph$,

$$Z(m) := \sum_{j=1}^{nM} g_{H,n}(j) L_\alpha(m - j) \tag{4.0-40}$$

where

$$g_{H,n}(j) := \Big((\frac{j}{n})_+^{H-1/\alpha} - (\frac{j}{n} - 1)_+^{H-1/\alpha} \Big) n^{-1/\alpha} \tag{4.0-41}$$

and where $L_\alpha(j)$ is the series of i.i.d standard stable Paretian random variables. Since $L_{\alpha,n}(j) \stackrel{d}{=} n^{-1/\alpha} L_\alpha(j), j \in \Re$, equation (4.0-40) and (4.0-41) imply $Y_{n,N}(t) \stackrel{d}{=} Z(nt)$, for $t = 1, ..., T$. Then, the computing is moved to focus on the moving average series $Z(m)$, $m = 1, ..., nT$. Let $\tilde{L}_\alpha(j)$ be the $n(N+T)$-periodic with $\tilde{L}_\alpha(j) := L_\alpha(j)$, for $j = 1, ..., n(N+T)$ and let $\tilde{g}_{H,n}(j) := g_{H,n}(j)$, for $j = 1, ..., nN$; $\tilde{g}_{H,n}(j) := 0$, for $j = nN+1, ..., n(N+T)$, then

$$\{Z(m)\}_{m=1}^{nT} \stackrel{d}{=} \Big\{ \sum_{j=1}^{n(N+T)} \tilde{g}_{H,n}(j) \tilde{L}_\alpha(n - j) \Big\}_{m=1}^{nT} \tag{4.0-42}$$

because for all $m = 1, ..., nT$, the summation in equation (4.0-40) involves only $L_\alpha(j)$ with indices j in the range $-nN \leq j \leq nT - 1$. Using a circular convolution of the two $n(N+T)$-periodic series $\tilde{g}_{H,n}$ and $\tilde{L}_\alpha$ computed by using their Discrete Fourier Transforms (DFT), the variables $Z(n)$, $m = 1, ..., nT$ (i.e., the fractional stable noise) can be generated.

4.4.4 Implications of Fractal Processes

Fractal processes have been applied to the study of computer networks. Leland et al. (1994) and Willinger et al. (1997) employ fractal processes in modeling Ethernet traffic. Feldmann et al. (1998) discuss the fractal processes in the measurement of TCP/IP and ATM WAN traffic.

Paxson and Floyd (1995), Paxson (1997), and Feldmann et al. (1998) discuss the characteristics of self-similarity in wide-area traffic with respect to the fractal processes. Crovella and Bestavros (1997) provide evidence of self-similarity in world-wide web traffic by means of fractal processes modeling. An extensive bibliographical review of the research in the area of network traffic and network performance involving the fractal processes and self-similarity is provided by Willinger et al. (1996). Sahinoglu and Tekinay (1999) survey studies on the self-similarity phenomenon in multimedia traffic and its implications in network performance.

Baillie (1996) provides a survey of the major econometric research on long-range dependence processes, fractional integration, and applications in economics and finance. Bhansali and Kokoszka (2006) review recent research on long-range dependence time series. Theoretical and empirical research on long-range dependence in economics and finance is provided by Robinson (2003), Rangarajan and Ding (2006), and Teyssiére and Kirman (2006).

Based on the modeling mechanism of fractal processes, Sun et al. (2006) empirically compare fractional stable noise with several alternative distributional assumptions in either fractal form or i.i.d. form (i.e., normal distribution, fractional Gaussian noise, generalized extreme value distribution, generalized Pareto distribution, and stable distribution) for modeling returns of major German stocks. The empirical results suggest that fractional stable noise dominates these alternative distributional assumptions both in in-sample modeling and out-of-sample forecasting. This finding suggests that the model built on non-Gaussian non-random walk (fractional stable noise) performs better than those models based on either the Gaussian random walk, the Gaussian non-random walk, or the non-Gaussian random walk.

Chapter 5

Modeling Univariate High-Frequency Time Series I

5.1 Introduction

Because return volatility estimates are key inputs in valuation modeling and trading strategies, considerable research in the financial econometrics literature has been devoted to return volatility modeling. The preponderance of empirical evidence from financial markets throughout the world fails to support the hypothesis that returns follow a Gaussian random walk.[1] In addition to the empirical evidence, there are theoretical arguments that have been put forth for rejecting both the Gaussian assumption and the random walk assumption. One of the most compelling arguments against the Gaussian random walk assumption is that markets exhibit a fractal structure. That is, markets exhibit a geometrical structure with self-similarity when scaled (see, Mandelbrot (1963, 1997)). As to this point, the normal distribution assumption and the random walk assumption cannot both be simultaneously valid for describing financial markets. It seems that the only way to explain fractal scaling is to abandon either the Gaussian hypothesis or the random walk hypothesis. By abandoning the Gaussian hypothesis, researchers end up with stable Paretian distributions.[2] The normal distribution is a special case with finite variance (details are discussed in Rachev and Mittnik (2000)). The implication of rejecting the random walk hypothesis is that researchers must accept that returns in financial

[1]See, Fama (1963, 1965), Mandelbrot (1963, 1997), and Rachev and Mittnik (2000).

[2]To distinguish between a Gaussian and non-Gaussian stable distribution, the latter is usually referred to as stable Paretian distribution or Lévy stable distribution. Referring to it as a stable Paretian distribution highlights the fact that the tails of the non-Gaussian stable density have Pareto power-type decay; referring to it as a Lévy stable distribution recognizes the pioneering works by Paul Lévy in characterizing the non-Gaussian stable laws (see Rachev and Mittnik (2000)).

markets are not independent but instead exhibit trends. Markets prone to trending have been characterized by long-range dependence and volatility clustering. Samorodnisky and Taqqu (1994) demonstrate that the properties of some self-similar processes can be used to model financial markets that are characterized as being non-Gaussian and non-random walk. Such financial markets have been stylized by long-range dependence, volatility clustering, and heavy tailedness.

Long-range dependence or long memory denotes the property of a time series to exhibit persistent behavior, i.e., a significant dependence between very distant observations and a pole in the neighborhood of the zero frequency of its spectrum.[3] Long-range dependence time series typically exhibit self-similarity. The stochastic processes with self-similarity are invariant in distribution with respect to changes of time and space scale. The scaling coefficient or self-similarity index is a non-negative number denoted by H, the Hurst parameter. If $\{X(t+h) - X(h), t \in T\} \stackrel{d}{=} \{X(t) - X(0), t \in T\}$ for all $h \in T$, the real-valued process $\{X(t), t \in T\}$ has stationary increments. A succinct expression of self-similarity is $\{X(at), t \in T\} \stackrel{d}{=} \{a^H X(t), t \in T\}$. The process $\{X(t), t \in T\}$ is called *H-sssi* if it is self-similar with index H and has stationary increments (see, Samorodnisky and Taqqu (1994) and Doukhan et al. (2003)).

In modeling return volatility, long-range dependence, volatility clustering, and heavy tailedness should be treated simultaneously in order to obtain more accurate predictions. Rachev and Mittnik (2000) note that for modeling financial data, not only does model structure play an important role, but distributional assumptions influence the modeling accuracy. A distribution that is rich enough to encompass those stylized facts exhibited in return data is the stable distribution. Fama (1963), Mittnik and Rachev (1993a, 1993b), Rachev (2003), and Rachev *et al.* (2005) have demonstrated the advantages of stable distributions in financial modeling. Moreover, Taqqu and Samorodnitsky (1994), Rachev and Mittnik (2000), Rachev and Samorodnitsky (2001), Doukhan *et al.* (2003), and Racheva and Samorodnitsky (2003) have reported that long-range dependence, self-similar processes, and stable distribution are very closely related.

In this chapter, I empirically investigate the return distribution of 27 German DAX stocks using intra-daily data under two separate assumptions regarding the return generation process

[3] Baillie (1996) provides a survey of the major econometric research on long-range dependence processes, fractional integration, and applications in economics and finance. Doukhan et al. (2003) and Robinson (2003) provide a comprehensive review of the studies on long-range dependence. Bhansali and Kokoszka (2006) review recent research on long-range dependence time series. Recent theoretical and empirical research on long-range dependence in economics and finance is provided by Rangarajan and Ding (2006) and Teyssiére and Kirman (2006). Sun et al. (2007a) provide a review of long-range dependence research based on using intra-daily data.

(1) it does not follow a Gaussian distribution and (2) it does not follow a random walk. In the empirical study, I develop the ARMA-GARCH model based on these assumptions. Abandoning the Gaussian hypothesis, I analyze the data by employing an ARMA-GARCH model with several independent and identically distributed (i.i.d.) residuals following a normal distribution, stable distribution, generalized extreme value distribution, and generalized Pareto distribution. When I desert the random walk hypothesis but maintain the Gaussian hypothesis, I utilize an ARMA-GARCH model with fractional Gaussian noise. The ARMA-GARCH model with fractional stable noise is used when I drop the assumptions that the return distribution is Gaussian and follows a random walk. Using several goodness of fit criteria for evaluating both in-sample simulation and out-of-sample forecasting for a sample with 6,600 observations, I find that the ARMA-GARCH model with fractional stable noise outperforms the other models investigated. This finding suggests that such a model can capture the stylized facts better without considering either the Gaussian or random walk hypotheses. In other words, this result supports the hypotheses that return distributions in financial markets are better characterized as fractals rather than Gaussian random walks.

The organization of this chapter is as follows. In Section 5.2, I introduce two self-similar processes: fractional Gaussian noise and fractional stable noise. The method for estimating the parameters in the underlying process is introduced in Section 5.3. In Section 5.4, methods for simulating fractional Gaussian noise and fractional stable noise are explained. The empirical results are reported in Section 5.5, where I compare the goodness of fit for both in-sample simulation and out-of-sample forecasting based on several criteria for the ARMA-GARCH model with fractional stable noise and with other distributions. I summarize the conclusions in Section 5.6.

5.2 Specification of the self-similar processes

Lamperti (1962) first introduced the semi-stable processes (which we today refer to as self-similar processes). Let T be either R, $R_+ = \{t : t \geq 0\}$ or $\{t : t > 0\}$. The real-valued process $\{X(t), t \in T\}$ has stationary increments if $X(t+a) - X(a)$ has the same finite-dimensional distributions for all $a \geq 0$ and $t \geq 0$. Then the real-valued process $\{X(t), t \in T\}$ is self-similar with exponent of self-similarity H for any $a > 0$, and $d \geq 1$, $t_1, t_2, ..., t_d \in T$, satisfying:

$$\Big(X(at_1), X(at_2), ..., X(at_d)\Big) \stackrel{d}{=} \Big(a^H X(t_1), a^H X(t_2), ...a^H X(t_d)\Big). \tag{5.0-1}$$

5.2.1 Fractional Gaussian noise

For a given $H \in (0,1)$, there is basically a single Gaussian *H-sssi* [4] process, namely fractional Brownian motion (fBm), first introduced by Kolmogorov (1940). Mandelbrot and Wallis (1968) and Taqqu (2003) define fBm as a Gaussian *H-sssi* process $\{B_H(t)\}_{t\in R}$ with $0 < H < 1$. Mandelbrot and van Ness (1968) define the stochastic representation

$$B_H(t) := \frac{1}{\Gamma(H+\frac{1}{2})}\Big(\int_{-\infty}^{0}[(t-s)^{H-\frac{1}{2}}-(-s)^{H-\frac{1}{2}}]dB(s)+\int_0^t(t-s)^{H-\frac{1}{2}}dB(s)\Big), \quad (5.0\text{-}2)$$

where $\Gamma(\cdot)$ represents the Gamma function:

$$\Gamma(a) := \int_0^\infty x^{a-1}e^{-x}dx,$$

and $0 < H < 1$ is the Hurst parameter. The integrator B is ordinary Brownian motion. The principal difference between fractional Brownian motion and ordinary Brownian motion is that the increments in Brownian motion are independent while in fractional Brownian motion they are dependent. For fractional Brownian motion, Samorodnitsky and Taqqu (1994) define its increments $\{Y_j, j \in Z\}$ as fractional Gaussian noise (fGn), which is, for $j = 0, \pm1, \pm2, ...$, $Y_j = B_H(j-1) - B_H(j)$.

5.2.2 Fractional stable noise

While fractional Brownian motion can capture the effect of long-range dependence, it has less power to capture heavy tailedness. The existence of abrupt discontinuities in financial data, combined with the empirical observation of sample excess kurtosis and unstable variance, confirms the stable Paretian hypothesis identified by Mandelbrot (1963, 1983). It is natural to introduce the stable Paretian distribution in self-similar processes in order to capture both long-range dependence and heavy tailedness. Samorodinitsky and Taqqu (1994) introduce the α-stable *H-sssi* processes $\{X(t), t \in R\}$ with $0 < \alpha < 2$. If $0 < \alpha < 1$, the exponent of self-similarity are $H \in (0, 1/\alpha]$ and if $1 < \alpha < 2$, the exponent of self-similarity are $H \in (0,1)$. In addition, Cohen and Samorodnitsky (2006) show that with exponent $H' = 1 + H(1/\alpha - 1)$, process $\{X(t), t \in R\}$ is a well-defined symmetric α-stable ($S\alpha S$) process. It has stationary increments and is self-similar. They show that (1) for $0 < \alpha < 1$, a family of H'-*sssi* $S\alpha S$ processes with $H' \in (1, 1/\alpha)$ is obtained, (2) for $1 < \alpha < 2$, a family of H'-*sssi* $S\alpha S$ processes with $H' \in (1/\alpha, 1)$ is obtained, and (3) for $\alpha = 1$, a family of *1-sssi* $S\alpha S$ processes is obtained.

[4] The abbreviation of "*sssi*" means self-similar stationary increments, if the exponent of self-similarity H is to be emphasized, then "*H-sssi*" is adopted.

There are many different extensions of fractional Brownian motion to the stable distribution. The most commonly used is linear fractional stable motion (also called linear fractional Lévy motion), $\{L_{\alpha,H}(a,b;t), t \in (-\infty,\infty)\}$, which Samorodinitsky and Taqqu (1994) define as

$$L_{\alpha,H}(a,b;t) := \int_{-\infty}^{\infty} f_{\alpha,H}(a,b;t,x)M(dx), \tag{5.0-3}$$

where

$$f_{\alpha,H}(a,b;t,x) := a\Big((t-x)_+^{H-\frac{1}{\alpha}} - (-x)_+^{H-\frac{1}{\alpha}}\Big) + b\Big((t-x)_-^{H-\frac{1}{\alpha}} - (-x)_-^{H-\frac{1}{\alpha}}\Big), \tag{5.0-4}$$

where $x_+ = \max(x,0), x_- = \min(x,0)$, and a, b are real constants. $|a|+|b| > 1$, $0 < \alpha < 2$, $0 < H < 1$, $H \neq 1/\alpha$, and M is an α-stable random measure on R with Lebesgue control measure and skewness intensity $\beta(x)$, $x \in (-\infty,\infty)$ satisfying: $\beta(\cdot) = 0$ if $\alpha = 1$. They define linear fractional stable noises expressed by $Y(t)$, and $Y(t) = X_t - X_{t-1}$,

$$\begin{aligned} Y(t) &= L_{\alpha,H}(a,b;t) - L_{\alpha,H}(a,b;t-1) \\ &= \int_R \Big(a\Big[(t-x)_+^{H-\frac{1}{\alpha}} - (t-1-x)_+^{H-\frac{1}{\alpha}}\Big] \\ &+ b\Big[(t-x)_-^{H-\frac{1}{\alpha}} - (t-1-x)_-^{H-\frac{1}{\alpha}}\Big]\Big)M(dx), \end{aligned} \tag{5.0-5}$$

where $L_{\alpha,H}(a,b;t)$ is a linear fractional stable motion defined by equation (5.0-3), and M is a stable random measure with Lebesgue control measure given $0 < \alpha < 2$. Samorodinitsky and Taqqu (1994) show that the kernel $f_{\alpha,H}(a,b;t,x)$ is d-self-similar with $d = H - 1/\alpha$ when $L_{\alpha,H}(a,b;t)$ is $1/\alpha$-self-similar. This implies $H = d + 1/\alpha$ (see Taqqu and Teverovsky (1998) and Weron et al. (2005)).[5] In this chapter, if there is no special indication, the fractional stable noise (fsn) is generated from a linear fractional stable motion.

5.3 Empirical analysis

The empirical analysis involves comparing the performance of six ARMA-GARCH models with different kinds of residuals (i.e., residuals with forms of white noise, fractional Gaussian noise, fractional stable noise, stable distribution, generalized Pareto distribution, and generalized extreme value distribution). The analysis is performed for in-sample simulation and out-of-sample forecasting of the German DAX stock returns based on four goodness of fit criteria: the Kolmogorov-Smirnov distance, the Anderson-Darling distance, the Cramer Von Mises distance, and the Kuiper distance.

[5] Some properties of these processes have been discussed in Mandelbrot and Van Ness (1968), Maejima (1983), Maejima and Rachev (1987), Manfields *et al.* (2001), Rachev and Mittnik (2000), Rachev and Samorodnitsky (2001), Racheva and Samorodnitsky (2003), Samorodnitsky (1994, 1996, 1998), Samorodinitsky and Taqqu (1994), and Cohen and Samorodnitsky (2006).

5.3.1 Data and Methodology

Recent research suggests that high-frequency data better reflect the market microstructure and increase the level of statistical significance (for example, Bollerslev and Wright (2000) and Dacorogna *et at* (2000)). In this study, I investigate the high-frequency data at 1-minute frequency for 27 German DAX component stocks[6] from January 7, 2002 to December 19, 2003. I calculate the stock returns by

$$y_{i,t} = \log\Big(\frac{P_{i,t}}{P_{i,t-1}}\Big). \tag{5.0-6}$$

I will let N ($N = 220,050$) denote the length of the sample. The sub-sample series used for the in-sample analysis are randomly selected by a moving window with length T. Replacement is allowed in the sampling. Letting T_F denote the length of the forecasting series, I perform one-week ahead out-of-sample forecasting ($1 \leq T \leq T + T_F \leq N$). In the empirical analysis, sub-sample length (i.e., the window length) of $T = 10,000$ (approximately one month) was chosen for the in-sample simulation and $T_F = 2,250$ (approximately one week) for the out-of-sample forecasting. A total of 6,600 sub-samples (200 sub-samples for each stock index) were randomly created.

I define the ARMA-GARCH model for the conditional mean equation as:

$$y_t = \alpha_0 + \sum_{i=1}^{r} \alpha_i \, y_{t-i} + \varepsilon_t + \sum_{j=1}^{m} \beta_j \varepsilon_{t-j}. \tag{5.0-7}$$

Let $\varepsilon_t = \sigma_t \, u_t$, where the conditional variance of the innovations, σ_t^2, is by definition

$$Var_{t-1}(y_t) = E_{t-1}(\varepsilon_t^2) = \sigma_t^2, \tag{5.0-8}$$

and u_t is i.i.d white noise. The general GARCH(p,q) process for the conditional variance of the innovation is then

$$\sigma_t^2 = \kappa + \sum_{i=1}^{p} \gamma_i \, \sigma_{t-i}^2 + \sum_{j=1}^{q} \theta_j \, \varepsilon_{t-j}^2. \tag{5.0-9}$$

In this analysis, ARMA(1,1)-GARCH(1,1) has been parameterized with different kinds of u_t, i.e. white noise, fractional Gaussian noise, fractional stable noise, stable distribution, generalized Pareto distribution, and generalized extreme value distribution.

[6]The data are from German Karlsruher Kapitalmarktdatabank (KKMDB). The DAX index consists of 30 stocks and the composition of the index changes every year. The database I developed included data from January 2002 to January 2004 for stocks that remained in the index over the entire period. Only 27 stocks satisfied that requirement and they are the ones used in this study.

The Kolmogorov-Smirnov (KS) distance, the Anderson-Darling (AD) distance, the Kuiper (K) distance, and the Cramer Von Mises (CVM) distance are used as the criterion for the goodness of fit testing. Letting $F_s(x)$ denote the empirical sample distribution and $\tilde{F}(x)$ the estimated distribution function, these measures are defined as follows:

$$KS = \sup_{x\in\Re} |F_s(x) - \tilde{F}(x)|, \tag{5.0-10}$$

$$AD = \sup_{x\in\Re} \frac{|F_s(x) - \tilde{F}(x)|}{\sqrt{\tilde{F}(x)(1-\tilde{F}(x))}}, \tag{5.0-11}$$

$$K = \sup_{x\in\Re}\Big(F_s(x) - \tilde{F}(x)\Big) + \sup_{x\in\Re}\Big(\tilde{F}(x) - F_s(x)\Big), \tag{5.0-12}$$

and

$$CVM = \int_{-\infty}^{\infty}\Big(F_s(x) - \tilde{F}(x)\Big)^2 d\tilde{F}(x). \tag{5.0-13}$$

The major disadvantage of KS statistics is that it tends to be more sensitive near the center of the distribution than at the tails. AD statistics can overcome this. The reliability of testing the empirical distribution increases with the help of these two statistics, with KS distance focusing on the deviations around the median of the distribution and AD distance on the discrepancies in the tails (see Rachev and Mittnik (2000)).

5.3.2 Preliminary Test

Table 5.1 shows the descriptive statistics of the returns of the 27 DAX stocks in this study. From the statistics reported in this table, it can be seen that excess kurtosis exists. Figure 5.1 shows the *Q-Q* plot of some stock returns. Notice that a concave departure from the straight line (exponential distribution) in the *Q-Q* plot is an indication of a heavy-tailed distribution (whereas a convex departure shows a light-tailed distribution).

The Hurst index $H \in (0,1)$ is the index of self-similarity. For Gaussian processes with stationary increments, when

1. $H \in (0, 0.5)$, the increments of a process tend to have opposite signs and thus are more zigzagging due to the negative covariance.

2. $H \in (0.5, 1)$, the covariance between these two increments is positive and less zigzagging of the process.

3. $H = 0.5$, the covariance between this two increments is zero.

This can be restated as following: If the Hurst index is

1. less than 0.5, the process displays “anti-persistence” (i.e., positive excess return is more likely to be reversed and the performance in the next period is likely to be below the average, or in the contrary, negative excess return is more likely to be reversed and the performance in the next period is likely to be above the average).

2. greater than 0.5, the process displays “persistence” (i.e., positive excess return or negative excess return is more likely to be continued and the performance in the next period is likely to be the same as that in the current period).

3. equal to 0.5, the process displays no memory (i.e., the performance in the next period has equal probability to be below or above the performance in the current period).

For fractional stable processes, if the process has the index α $(0 < \alpha < 2)$, when $H = 1/\alpha$ which corresponds to a process with independent increments, this process has no memory. When $H > 1/\alpha$, the process displays long-range dependence and when $H < 1/\alpha$, the process displays negative dependence. In addition, long-range dependence is only possible when $\alpha > 1$, since $H \in (0, 1)$ (see, Samorodnitsky and Taqqu (1994)).

In order to check long-range dependence in stock returns, I use the methods introduced in Sections 5.3.1 and 5.3.2 to estimate the Hurst index under the Gaussian and stable assumptions. I employed the MLE method explained in Section 5.3.3 to estimate the stable parameter. The results, reported in Table 5.1, indicate that the Hurst index does not have an estimated value of 0.5 if fractional Gaussian noise is assumed. This suggests the occurence of either long memory or short memory under the Gaussian assumption [7]. In Table 1, I can observe both fluctuation and long memory under the non-Gaussian stable assumption.

Engle (1982) proposes a Lagrange-multiplier test for the ARCH phenomenon. A test statistic for ARCH of lag order q is given by

$$X_q \equiv nR_q^2$$

where R_q^2 is the non-centered goodness-of-fit coefficient of a qth order autoregression of the squared residuals taken from the original regression

$$\hat{u}_t^2 = \omega_0 + \omega_1 \hat{u}_{t-1}^2 + \omega_2 \hat{u}_{t-2}^2 + \cdots + \omega_q \hat{u}_{t-q}^2 + e_t, \tag{5.0-14}$$

where $\hat{u}$ is the residual in the original regression equation. Under the null hypothesis of the residuals of the original model being normally i.i.d., the ARCH statistic of lag order q follows a χ^2 distribution with q degrees of freedom:

$$\lim_{n\to\infty} X_q \sim \chi_q^2.$$

[7] There are various extensions of the self-similarity property for generalized random processes, see Dobrushin (1979).

I use Engle's test to check whether the ARCH effect occurs. In Table 5.2, I report the test statistics and the critical values to reject the null hypothesis that there is no ARCH effect at different lag orders. It is clear from the results reported in the table that an ARCH effect is exhibited in the return time series studied.

I use the Ljung-Box-Pierce Q-statistic based on the autocorrelation function to test serial correlation (i.e., the memory effect). The Q-statistic is

$$Q :\sim \chi^2_m = N(N+2)\sum_{k=1}^{m}\frac{\rho_k^2}{N-k}, \tag{5.0-15}$$

where N denotes the sample size, m the number of autocorrelation lags included in the statistic, and ρ_k the sample autocorrelation at a lag of order k which is

$$\rho_k = \frac{\sum_{t=1}^{N-k} y_t y_{t+k}}{\sum_{t=1}^{N} y_t^2}. \tag{5.0-16}$$

Ljung and Box (1978) show that the Q-statistic is following an asymptotic χ^2 distribution.

The Ljung-Box-Pierce test results reported in Table 5.3 indicate that the hypothesis that there is no serial correlation can be rejected at different lags. I find that the memory effect occurs in the returns for each stock from a lag of 10 minutes to a lag of one month. In order to see when the memory effect vanishes, I compare the Q-statistic with its corresponding critical value. When the quotient of the Q-statistic and the corresponding critical value is less than 1, I cannot reject the null hypothesis that there is no serial correlation. The results reported in Table 5.4 show that all the stock returns exhibit serial correlation even after a half year. In 8 months, the memory effect vanishes for 8 stocks and in 10 months, the memory effect vanishes for 19 stocks. From Table 5.4, I see that the decay of autocorrelation is slow.

5.3.3 Results

The AD, KS, CVM, and Kuiper statistics were calculated for the six candidate distributional assumptions. The results of the descriptive statistics of the computed values for the four criteria for the in-sample study are reported in Table 5.5. As can be seen from this table, the ARMA-GARCH with fractional stable noise model exhibits a smaller mean value for all criteria than the other five models. That is, for the in-sample study, the ARMA-GARCH with fractional stable noise model has the best performance. I also perform one week ahead out-of-sample forecasting for stock returns. The results for the descriptive statistics of the computed four criteria, reported in Table 5.6, indicate that the ARMA-GARCH with fractional stable noise model exhibits a smaller mean value for all criteria than the other five models. This suggests that the ARMA-GARCH with fractional stable noise is better at forecasting than the other models studied.

5.4 Conclusions

There is considerable interest in the modeling of market volatility. Most models assume that residuals are independent and identically distributed and follow the Gaussian distribution. But the overwhelming empirical evidence does not support the hypothesis that financial asset returns can be characterized as Gaussian random walks. There are a number of arguments against both the Gaussian assumption and the random walk assumption. One of the most compelling arguments against the Gaussian random walk is that there exist fractals in financal markets. In this chapter, I empirically investigate 27 German DAX stocks sampled over two years at one minute frequency level under three separate assumptions regarding the return generation process (1) it does not follow a Gaussian distribution, (2) it does not follow a random walk, and (3) it does not follow a Gaussian random walk. When I model non-Gaussian random walk, I employ one of the self-similar processes (i.e., fractional stable noise) to capture the fractal structure in financial markets.

In this empirical analysis, I investigate the ARMA-GARCH model with six different forms of residuals in both fractal forms (i.e., fractional stable noise and fractional Gaussian noise) and i.i.d. forms (i.e., stable distribution, white noise, generalized Pareto distribution, and generalized extreme value distribution) for the modeling volatility of 27 German stocks. In-sample (one month) simulation and out-of-sample (one week) forecasting were empirically investigated. By using parameters estimated from the empirical series, I simulate an in-sample series for each stock return and one-step ahead forecasting series with these six modeling structures. Then I compared the goodness of fit for the generated series to the sampled series by adopting four criteria for the goodness of fit test (the Kolmogorov-Smirnov distance, Anderson-Darling distance, Cramer von Mises distance, and Kuiper distance). Based on a comparison of these criteria, the empirical evidence shows that the ARMA-GARCH model with fractional stable noise demonstrates a better performance in modeling volatility than other models. My results also indicate that there exists a fractal structure in financial markets and the i.i.d. assumption in modeling is inappropriate for characterizing financial asset returns.

The empirical evidence suggests that stocks exhibit three characteristics: long-range dependence, heavy tailedness, and volatility clustering. Many studies have found that the stable distribution is a better description of financial returns because it can capture heavy tailedness and has a close relationship with long-range dependence. As a self-similar process, fractional stable noise can capture the reported stylized facts in financial return data. This finding should be taken into account in modeling volatility so that a more accurate prediction might be realized by well-defined functional models.

Table 5.1: Statistical characteristics of stocks in the study.

	mean	std	skewness	kurtosis	max	min	$\tilde{\alpha}$	$\tilde{H}_{FGN}$	$\tilde{H}_{fsn}$
Adidas	1.83E-07	0.0012	0.4041	94.9171	0.0670	-0.0327	0.7903	0.4791	1.3168
Allianz	-3.01E-06	0.0015	-0.5477	212.8600	0.0949	-0.0905	1.4222	0.5203	0.5616
BASF	2.65E-07	0.0012	1.2349	121.8300	0.0863	-0.0341	1.2247	0.4548	0.4589
BAYER	-1.29E-06	0.0015	0.8794	85.5980	0.0877	-0.0521	1.3998	0.5137	0.4947
BMW	-1.27E-07	0.0012	0.1213	58.6840	0.0459	-0.0463	1.2010	0.4946	0.4873
Commerzbank	-4.40E-07	0.0018	0.2352	48.3320	0.0539	-0.0577	1.1694	0.5337	0.5186
Daimler	-7.83E-07	0.0013	-0.2877	56.4390	0.0469	-0.0597	1.4160	0.4628	0.4621
Dt.Bank	-5.72E-07	0.0012	-0.6959	68.6110	0.0339	-0.0571	1.3938	0.5332	0.5082
Dt.Post	2.16E-07	0.0015	-0.3476	41.8600	0.0393	-0.0512	0.8724	0.4258	1.0786
Dt. Telecom	-8.79E-07	0.0015	-0.8822	100.8200	0.0557	-0.0913	1.8337	0.5293	0.5190
Eon	-4.01E-07	0.0011	-0.2422	26.6700	0.0234	-0.0334	1.2434	0.4214	0.4381
Fressenius	-7.16E-07	0.0017	5.0585	646.8212	0.1982	-0.0447	0.8294	0.5411	1.2408
Henkel	3.64E-08	0.0011	1.9018	261.9911	0.0977	-0.0389	0.7562	0.4115	1.2339
Hypovereinsbank	-1.92E-06	0.0019	-2.3888	236.3100	0.0694	-0.1475	1.2688	0.5469	0.5304
Infineon	-2.36E-06	0.0020	-0.7092	145.4146	0.0707	-0.1016	1.6018	0.5351	0.5451
Linde	-1.01E-07	0.0013	-0.2952	44.1141	0.0340	-0.0392	0.7968	0.4716	1.2696
Lufthansa	-4.10E-07	0.0016	0.5289	79.2296	0.0821	-0.0541	1.2694	0.4647	0.4647
Man	3.80E-08	0.0016	0.3009	65.3610	0.0754	-0.0461	0.7645	0.5372	1.3134
Metro	2.66E-06	0.0015	0.1496	63.0310	0.0477	-0.0455	1.1293	0.4806	0.5084
Muechenerrueck	-3.52E-06	0.0015	-0.4968	105.6100	0.07250	-0.0748	1.3203	0.5842	0.5559
RWE	-9.71E-07	0.0013	-0.1861	27.6080	0.0353	-0.0355	1.2270	0.3960	0.4400
SAP	3.54E-06	0.0011	-0.1718	128.7802	0.0577	-0.0427	1.3693	0.5481	0.5485
Schering	-2.45E-07	0.0012	-0.3756	64.1430	0.0299	-0.0553	1.1411	0.4299	0.4271
Siemens	2.74E-06	0.0011	0.1990	49.7927	0.0379	-0.0350	1.4426	0.5251	0.5275
ThyssenKrupp	2.18E-06	0.0015	-0.1627	34.7571	0.0373	-0.0419	1.1064	0.5019	0.5126
Tui	-1.59E-06	0.0019	0.0812	145.4800	0.1253	-0.1032	0.8074	0.4965	1.1906
Volkswagen	-5.18E-07	0.0013	-0.5010	60.3170	0.0502	-0.0584	1.1807	0.4908	0.5042

Table 5.2: ARCH-test for different lags at $\alpha = 0.05$.

$q =$	$1min$	$2min$	$5min$	$10min$	$15min$	$20min$	$25min$	$30min$	$60min$
Adidas	4403.1	6357.5	6530.4	6772.9	6821.2	6836.3	6894.1	6901.5	6986.5
Allianz	24.9	35.8	62.6	111.6	120.2	124.1	126.7	129.1	136.3
BASF	52.6	58.6	70.3	83.0	88.1	92.6	95.0	97.2	110.3
BAYER	2590.8	2595.1	2739.4	4511.9	4837.8	4840.4	4853.3	4881.6	5068.2
BMW	1211.3	1618.1	1780.2	1941.7	2030.9	2096.5	2129.6	2154.1	2262.1
Commerzbank	2595.7	2831.3	3693.9	4046.8	4109.8	4136.0	4181.2	4196.0	4264.1
Daimler	307.0	374.5	402.4	475.8	504.5	522.2	537.0	552.8	578.2
Dt.Bank	1172.6	1304.8	1504.4	1786.6	1911.8	1992.4	2073.5	2130.8	2325.2
Dt.Post	1552.0	3241.9	3390.5	3542.5	3601.2	3631.2	3661.7	3678.2	3702.1
Dt. Telecom	319.7	364.3	418.0	488.8	537.9	569.6	597.1	625.8	693.7
Eon	5050.6	5534.0	5896.9	6418.7	6668.7	6796.2	6940.1	7028	7251.6
Fressenius	9701.1	11503.0	12842.1	12981.0	13044.0	13083.0	13111.0	13124.0	13211.0
Henkel	131.2	139.6	148.1	185.4	194.2	196.0	197.4	198.9	211.5
Hypovereinsbank	75.9	90.2	100.9	121.6	127.9	154.2	154.6	163.1	179.7
Infineon	2477.1	2995.4	3056.6	3087.6	3101	3104.4	3113.1	3117.4	3128.6
Linde	3526.6	5312.5	5564.4	5797.4	5896.7	6048.9	6102.1	6146.3	6192.0
Lufthansa	1311.3	1605.0	3466.5	3519.5	3590.8	3626.7	3643.2	3649.3	3688.6
Man	3302.3	3890.1	4360.1	4530.6	4564.7	4572.9	4590.3	4600.9	4628.0
Metro	3654.2	4173.9	5454.1	5510.7	5551.1	5582.3	5597.6	5615.3	5652.3
Muechenerrueck	361.4	493.5	620.2	652.4	674.5	688.4	697.2	708.9	793.5
RWE	1747.6	2354.2	3049.4	3653.3	3875.2	3983.9	4042.6	4098.7	4284.3
SAP	351.0	377.9	410.2	448.5	466.4	483.8	496.7	502.7	515.4
Schering	2059.0	2335.1	2616.9	2763.7	2828.9	2865.3	2910.5	2937	2956.6
Siemens	732.8	896.3	1045.9	1178.5	1339.0	1398.2	1475.8	1513.6	1614.2
ThyssenKrupp	2237.2	3127.1	3399.1	3566.4	3671.6	3794	3824.2	3852.9	3892.2
Tui	154.4	251.8	299.5	331.4	346.1	358.9	366.9	374.6	395.7
Volkswagen	2682.0	3367.0	4020.8	4806.4	4906.4	4980.2	5020.3	5036.7	5114.0
Critical value	3.8415	5.9915	11.0705	18.307	24.996	31.4104	37.6525	43.773	67.5048

Table 5.3: Ljung-Box-Pierce Q-test statistic for different lags at $\alpha = 0.05$.

$k =$	$10min$	$30min$	$1hr$	$2hr$	$4hr$	$1day$	$1wk$	$2wk$	$1mon$
Adidas	2016.7	2083.8	2131.5	2194.1	2321.3	2579.8	4910.9	7783.8	15872.0
Allianz	309.4	371.7	421.8	514.8	694.8	987.5	3923.6	7490.2	18300.0
BASF	624.7	665.5	708.2	800.2	961.6	1319.7	4014.8	7174.5	16150.0
BAYER	572.4	650.2	745.0	865.1	1136.3	1506.8	4453.2	7758.6	17859.0
BMW	653.2	698.7	741.5	849.2	983.3	1279.4	3855.4	6854.6	15903.0
Commerzbank	1990.2	2013.7	2067.1	2166.1	2289.2	2613.1	5127.9	8440.4	17903.0
Daimler	1324.3	1376.4	1428.9	1479.6	1618.1	1895.0	4461.3	7311.3	15993.0
Dt.Bank	710.6	760.1	830.1	933.4	1080.0	1453.6	3993.9	7076.4	16047.0
Dt.Post	2030.7	2096.6	2132.0	2187.1	2310.3	2579.8	5042.4	7870.4	16261.0
Dt. Telecom	2031.5	2084.9	2170.6	2289.2	2436.1	2794.6	5652.3	9147.6	19491.0
Eon	822.6	866.2	924.4	1020.4	1147.9	1515.8	4351.1	7554.7	17160.0
Fressenius	2565.9	2640.8	2680.1	2765.3	2885.1	3154.0	5705.7	8722.2	17796.0
Henkel	1742.1	1823.5	1878.5	1959.9	2130.6	2477.5	5168.2	8111.4	16844.0
Hypovereinsbank	910.2	959.8	1046.2	1138.5	1277.7	1586.7	3931.8	6924.0	16185.0
Infineon	1454.0	1518.1	1555.4	1619.4	1789.7	2071.6	4489.3	7482.8	16435.0
Linde	1048.7	1097.6	1136.9	1203.0	1371.1	1662.1	3986.9	6690.9	14994.0
Lufthansa	2491.6	2544.4	2590.2	2659.1	2823.4	3087.4	5568.6	8543.5	17758.0
Man	1817.7	1866.1	1913.5	1991.6	2134.2	2419.6	4778.6	7874.1	16652
Metro	1339.1	1400.2	1426.3	1506.9	1692.5	1954.3	4319.2	7230.2	16405.5
Muechenerrueck	220.2	280.9	332.8	409.6	555.2	880.1	3708.7	7020.0	16338.1
RWE	942.2	989.4	1051.8	1146.8	1303.7	1626.3	4683.1	8025.7	17937.3
SAP	484.2	511.3	554.4	629.7	748.9	1060.7	3254.8	6141.3	14647.2
Schering	1421.1	1451.2	1486.2	1579.2	1746.6	2006.5	4599.5	7511.9	15945.0
Siemens	776.5	850.9	929.6	1003.3	1158.3	1481.5	4221.6	7596.7	16957.0
ThyssenKrupp	2623.9	2693.6	2733.4	2813.3	2959.5	3206.6	5407.0	8185.1	16618.0
Tui	1561.9	1599.7	1633.4	1734.4	1873.2	2166.8	4732.4	7751.7	17175.0
Volkswagen	369.7	394.97	459.9	526.6	674.9	987.9	3598.3	6671.3	15722.0
Critical value	18.3	43.8	79.1	146.6	277.1	532.1	2515.1	4962.3	12256.0

Table 5.4: Ljung-Box-Pierce Q-test statistic compared with corresponding critical values for different lags at $\alpha = 0.05$.

$k =$	$1day$	$5days$	$10days$	$1mon$	$2mon$	$4mon$	$6mon$	$8mon$	$10mon$
Adidas	4.8487	1.9526	1.5686	1.2951	1.1898	1.0947	1.0515	1.0239	1.0299
Allianz	1.8560	1.5600	1.5094	1.4932	1.3902	1.2348	1.0749	0.9917	0.9557
BASF	2.4802	1.5963	1.4458	1.3177	1.2552	1.1445	1.0452	0.9959	0.9838
BAYER	2.8319	1.7706	1.5635	1.4572	1.3652	1.2013	1.0657	1.0012	0.9773
BMW	2.4045	1.5329	1.3813	1.2976	1.2269	1.1416	1.0524	1.0108	0.9924
Commerzbank	4.9112	2.0389	1.7009	1.4607	1.3198	1.1917	1.0869	1.0351	1.0239
Daimler	3.5616	1.7738	1.4734	1.3049	1.2272	1.1339	1.0510	1.0076	0.9989
Dt.Bank	2.7320	1.5880	1.4260	1.3094	1.2349	1.1262	1.0277	0.9864	0.9787
Dt.Post	4.8486	2.0049	1.5860	1.3268	1.2082	1.0977	1.0330	1.0174	1.0270
Dt. Telecom	5.2522	2.2474	1.8434	1.5903	1.4769	1.2730	1.1243	1.0352	0.9920
Eon	2.8489	1.7300	1.5224	1.4001	1.3238	1.1677	1.0313	0.9758	0.9704
Fressenius	5.9277	2.2686	1.7577	1.4520	1.3370	1.2179	1.1164	1.0531	1.0139
Henkel	4.6563	2.0549	1.6346	1.3743	1.2953	1.1638	1.0696	1.0328	1.0240
Hypovereinsbank	2.9821	1.5633	1.3953	1.3206	1.2731	1.1032	1.0071	0.9755	0.9655
Infineon	3.8934	1.7850	1.5079	1.3410	1.2748	1.1783	1.0882	1.0239	0.9898
Linde	3.1239	1.5852	1.3483	1.2234	1.1806	1.1272	1.0777	1.0231	0.9914
Lufthansa	5.8025	2.2141	1.7217	1.4489	1.3122	1.2106	1.1114	1.0540	1.0301
Man	4.5475	1.9000	1.5868	1.3587	1.2489	1.1460	1.0739	1.0315	1.0143
Metro	3.6729	1.7173	1.4570	1.3385	1.2598	1.1772	1.0847	1.0289	0.9925
Muechenerrueck	1.6540	1.4746	1.4147	1.3331	1.2643	1.1484	1.0282	0.9816	0.9721
RWE	3.0566	1.8620	1.6173	1.4636	1.3523	1.2060	1.0769	1.0169	0.9907
SAP	1.9935	1.2941	1.2376	1.1951	1.1414	1.0983	1.0388	1.0089	0.9942
Schering	3.7711	1.8288	1.5138	1.3010	1.2165	1.1067	1.0378	1.0153	1.0121
Siemens	2.7844	1.6785	1.5309	1.3836	1.2824	1.1324	1.0167	0.9690	0.9722
ThyssenKrupp	6.0266	2.1498	1.6495	1.3559	1.2374	1.1373	1.0737	1.0488	1.0392
Tui	4.0724	1.8816	1.5621	1.4013	1.3218	1.2191	1.1020	1.0404	0.9927
Volkswagen	1.8567	1.4307	1.3444	1.2828	1.2223	1.1262	1.0303	0.9896	0.9771

Table 5.5: Summary of in-sample goodness of fit statistics for different models.

a. AD-statistic	AD_{mean}	AD_{std}	AD_{median}	AD_{max}	AD_{min}	AD_{range}
ARMA-GARCH-fGn	46.6768	54.3660	13.7335	55.8541	21.6282	34.2259
ARMA-GARCH-fsn	44.1625	53.2522	15.2382	64.6001	1.3917	63.2084
ARMA-GARCH-nor	46.7177	54.3751	13.7480	58.5747	21.0690	37.5057
ARMA-GARCH-sta	45.4108	53.5900	14.3638	95.2204	2.9886	92.2318
ARMA-GARCH-gev	46.6401	54.2656	13.7441	60.2271	21.0914	39.1357
ARMA-GARCH-gpd	51.2203	54.4755	20.2070	109.5363	3.5018	106.0244
b. KS-statistic	KS_{mean}	KS_{std}	KS_{median}	KS_{max}	KS_{min}	KS_{range}
ARMA-GARCH-fGn	0.4998	0.4992	0.0034	0.5285	0.4887	0.0398
ARMA-GARCH-fsn	0.4938	0.4965	0.0261	0.9725	0.2745	0.6980
ARMA-GARCH-nor	0.5003	0.4994	0.0043	0.5455	0.4893	0.0562
ARMA-GARCH-sta	0.5089	0.4974	0.0622	0.9725	0.3910	0.5815
ARMA-GARCH-gev	0.5000	0.4989	0.0057	0.5775	0.4825	0.0950
ARMA-GARCH-gpd	0.5698	0.5198	0.1335	1.0000	0.4165	0.5835
c. CVM-statistic	CVM_{mean}	CVM_{std}	CVM_{median}	CVM_{max}	CVM_{min}	CVM_{range}
ARMA-GARCH-fGn	449.5326	517.0503	203.7237	896.6917	82.9310	813.7607
ARMA-GARCH-fsn	445.2936	515.2956	202.2463	1473.6038	34.4169	1439.1868
ARMA-GARCH-nor	449.6701	517.3712	203.7739	889.0445	82.9532	806.0913
ARMA-GARCH-sta	454.7954	516.7259	230.4694	2985.6218	57.5066	2928.1152
ARMA-GARCH-gev	449.2438	517.0805	203.8510	886.1477	83.0288	803.1189
ARMA-GARCH-gpd	524.3399	521.2781	370.3320	1978.9581	52.5637	1926.3945
d. Kuiper-statistic	$Kuiper_{mean}$	$Kuiper_{std}$	$Kuiper_{median}$	$Kuiper_{max}$	$Kuiper_{min}$	$Kuiper_{range}$
ARMA-GARCH-fGn	0.9931	0.9937	0.0029	0.9985	0.9757	0.0227
ARMA-GARCH-fsn	0.9693	0.9862	0.0473	0.9985	0.5125	0.4860
ARMA-GARCH-nor	0.9931	0.9938	0.0029	0.9990	0.9750	0.0240
ARMA-GARCH-sta	0.9796	0.9877	0.0224	0.9990	0.6550	0.3440
ARMA-GARCH-gev	0.9913	0.9925	0.0048	0.9990	0.9570	0.0420
ARMA-GARCH-gpd	0.9696	0.9773	0.0287	1.0000	0.6505	0.3495

Table 5.6: Goodness of fit statistics for out-of-sample one week forecasting of different models.

a. AD-statistic	AD_{mean}	AD_{std}	AD_{median}	AD_{max}	AD_{min}	AD_{range}
ARMA-GARCH-fGn	30.1821	22.5228	13.6283	55.2241	21.5834	33.6407
ARMA-GARCH-fsn	27.6038	22.4110	12.2880	68.1149	1.2129	66.9021
ARMA-GARCH-nor	30.1927	22.5228	13.6421	59.2046	21.1361	38.0684
ARMA-GARCH-sta	28.8034	22.3886	13.0021	101.7023	2.6941	99.0082
ARMA-GARCH-gev	30.1205	22.5005	13.5541	59.8893	20.7111	39.1782
ARMA-GARCH-gpd	32.3273	23.9319	15.3931	108.9876	4.1084	104.8792
b. KS-statistic	KS_{mean}	KS_{std}	KS_{median}	KS_{max}	KS_{min}	KS_{range}
ARMA-GARCH-fGn	0.5018	0.5006	0.0049	0.5375	0.4905	0.0470
ARMA-GARCH-fsn	0.4965	0.4985	0.0278	0.9555	0.2820	0.6734
ARMA-GARCH-nor	0.5020	0.5010	0.0055	0.5615	0.4880	0.0735
ARMA-GARCH-sta	0.5105	0.4990	0.0617	0.9653	0.4049	0.5603
ARMA-GARCH-gev	0.5018	0.5005	0.0064	0.5705	0.4846	0.0858
ARMA-GARCH-gpd	0.5700	0.5210	0.1333	1.0000	0.4049	0.5951
c. CVM-statistic	CVM_{mean}	CVM_{std}	CVM_{median}	CVM_{max}	CVM_{min}	CVM_{range}
ARMA-GARCH-fGn	226.5630	92.4072	245.5856	950.0136	82.5743	867.4392
ARMA-GARCH-fsn	223.6907	91.6806	244.8920	1873.2304	34.0265	1839.2039
ARMA-GARCH-nor	226.6043	92.4969	245.5955	948.5973	82.7246	865.8726
ARMA-GARCH-sta	229.7204	92.0018	264.0086	2852.7912	77.7136	2775.0774
ARMA-GARCH-gev	225.5252	92.2819	243.0575	933.6036	82.4613	851.1423
ARMA-GARCH-gpd	248.8990	94.0770	282.0017	1929.6210	55.7156	1873.9054
d. Kuiper-statistic	$Kuiper_{mean}$	$Kuiper_{std}$	$Kuiper_{median}$	$Kuiper_{max}$	$Kuiper_{min}$	$Kuiper_{range}$
ARMA-GARCH-fGn	0.9935	0.9940	0.0032	1.0000	0.9715	0.0285
ARMA-GARCH-fsn	0.9698	0.9870	0.0481	0.9990	0.5362	0.4627
ARMA-GARCH-nor	0.9935	0.9940	0.0032	1.0000	0.9725	0.0275
ARMA-GARCH-sta	0.9801	0.9885	0.0231	0.9995	0.6876	0.3118
ARMA-GARCH-gev	0.9918	0.9930	0.0052	0.9995	0.9592	0.0402
ARMA-GARCH-gpd	0.9703	0.9780	0.0299	1.0000	0.6425	0.3575

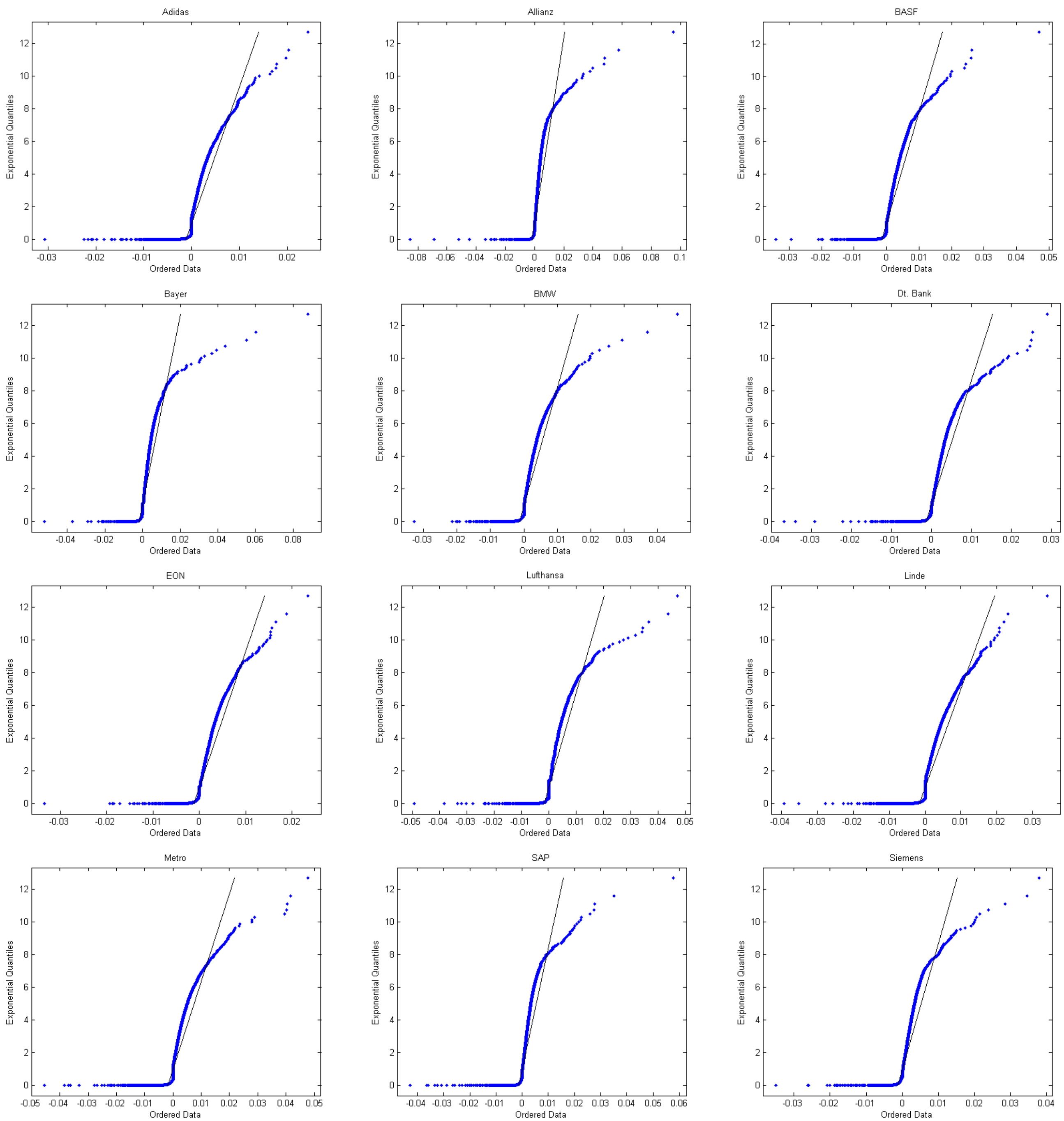

Figure 5.1: Q-Q plot of the returns for the stocks in study.

Chapter 6

Modeling Univariate High-Frequency Time Series II

6.1 Introduction

There is considerable interest in the information content and implications of the time between consecutive financial transactions (referred to as trade duration) for trading strategies and intra-day risk management. Market microstructure theory, supported by empirical evidence, suggests that the spacing between trades be treated as a variable to be explained or predicted since time carries information and closely correlates with price volatility (see, Bauwens and Veredas (2004), Diamond and Varrecchia (1987), Engle (2000), Engle and Russell (1998), Hasbrouck (1996), and O'Hara (1995)). Empirically it has been found that returns and volatility directly interact with trade duration and trade order size; and, short duration moves price more than long duration across stocks and across time (see, Manganelli (2005) and Furfine (2007)). Several studies (for example, Bauwens and Giot (2000), Dufour and Engle (2000), Engle and Russell (1998), and Jasiak (1998)) report that trade duration tends to exhibit long-range dependence (i.e., trade duration tends to be persistent), heavy tailedness (i.e., extremely short or long trade duration can be often observed), and clustering (i.e., short duration follows short duration and long duration follows long duration).

These findings raise two questions in studying the information content of trade duration that I address in this chapter:

1. Can single stochastic processes which capture long-range dependence and heavy tailedness be used in modeling trade duration data ?

2. Can a relatively "powerful" distributional assumption in a relatively "simple" functional

structure be used for efficient modeling of trade duration data?

It is necessary to treat long-range dependence, heavy tailedness, and clustering simultaneously in order to obtain more accurate predictions. Rachev and Mittnik (2000) note that for modeling financial data, not only does model structure play an important role, but distributional assumptions influence the modeling accuracy. The stable Paretian distribution[1] can be used to capture characteristics of trade duration since it is rich enough to encompass those stylized facts in such data, such as non-Guassian, heavy tails, long-range dependence, and clustering. Other researchers have shown the advantages of stable distributions in financial modeling (see, Fama (1963), Mittnik and Rachev (1993), Rachev and Mittnik (2000), and Sun et al. (2007b)). Meanwhile several studies have reported that long-range dependence, self-similar processes, and stable distribution are very closely related (see, Doukhan et al. (2003), Rachev and Mittnik (2000), Rachev and Samorodnitsky (2001), Samorodnitsky and Taqqu (1994), and Sun et al. (2007b)).

The Hurst index is also used to model long-range dependence, see Hurst (1951, 1955). It quantifies the degree of long-range dependence and measures the self-similarity scaling. Fortunately, one type of self-similar process can possess the Hurst index and stable distribution together (i.e., can capture both long-range dependence and heavy tailedness). This kind of stochastic process is a fractional stable noise generated from fractional stable motion (see, Samorodnitsky and Taqqu (1994)). Therefore, single stochastic processes can capture long-range dependence and heavy tailedness, answering the first question posed above. Based on estimating intensity of point processes, an autoregressive conditional duration (ACD) model is proposed by Engle and Russell (1998) for modeling trade duration with intertemporal correlation. The ACD model is a joint approach combining transition analysis and Engle's autoregressive conditional heteroscedasticity (ARCH) model. The motivation behind the ACD and the ARCH models is that in financial market events tend to occur in clusters. If fractional stable noise can be subordinated into the functional structure of the ACD model, then the second question can be answered.

In order to answer the two questions posed above, this chapter introduces fractional stable noise as the single stochastic processes to model trade duration. In the empirical analysis of this chapter, fractional stable noise is subordinated to the ACD model to model trade duration.

[1]To distinguish between Gaussian and non-Gaussian stable distribution, the latter is usually named stable Paretian distribution or Lévy stable distribution. Referring to it as a stable Paretian distribution highlights the fact that the tails of the non-Gaussian stable density have Pareto power-type decay and Lévy stable is the recognition of pioneering works done by Paul Lévy to the characterization of non-Gaussian stable laws (see Rachev and Mittnik (2000)).

As to self-similar processes, other single stochastic processes, such as fractional Gaussian noise which captures long-range dependence, are also introduced as an alternative. Since the stable distribution itself can capture heavy tailedness and long-range dependence, I propose it as an alternative distribution that can better explain trade duration. In the empirical analysis, stable distribution is also subordinated to the ACD model. Some other distributions that are often used in modeling trade duration, such as lognormal distribution, exponential distribution and Weibull distribution, have been selected as alternative distributional assumptions in order to compare goodness of fit with the stable distribution and fractional stable noise. Utilizing two test statistics usually used to evaluate model performance under heavy-tailed assumptions, I examine trade duration for a sample of stocks to compare which distributional assumption fits better. By applying a newly developed test procedure that I formulate, based on a bootstrap method, I obtain empirical results that suggest the fractional stable noise and stable distribution dominate these alternative assumptions with high statistical significance. Comparing goodness of fit in the modeling of trade duration data for stable distribution and fractional stable noise, the empirical results indicate that the ACD model with stable distribution fits better than other combinations, while fractional stable noise itself fits better for the time series of trade duration.

The contributions of this chapter to the microstructure finance literature are threefold. The first, and the major contribution, is that I am the first to employ fractal models in the study of roughness[2] of trade duration. I present new applications of fractal processes to model trade duration movements with a large sample. this study confirms the advantages of fractal models in analyzing roughness of trade duration compared to other models, since "the proper language of the theory of roughness in nature and culture is fractal geometry" (see Mandelbrot (2005, p.193)). Second, I extend the ACD model with fractal models and find results that are consistent with other studies that have shown that trade duration is informative and correlated with market volatility. Fractal processes have been shown to be superior to other models in modeling equity market intra-daily return dynamics (see Sun et al. (2007a)). The stock returns and their durations exhibit similar movement captured by fractal models, implying a certain inter-relationship between market volatility and information contained in trade duration. This is partially explained by the use of the ACD-GARCH model in the study of ultra-high-frequency data (see Ghysels and Jasiak (1998)). Third, trade durations of different stocks exhibit dependence and such dependence is possibly caused by new information revealed from trade intensity

[2]Roughness refers to an intricate, highly irregular appearance on all resolutions. It could be thought of as the synonym of irregularity, see Mandelbrot (1997). Davies and Hall (1999) discuss the surface roughness and point out that fractal methods partition the characteristics of surface roughness into two parts, "one of them scale invariant and the other governed by propeties of scale" (see Davies and Hall (1999, p.5)).

(see Simonsen (2007)). In this study, I provide some evidence to support the dependence exhibited in trade durations because of the distributional similarity observed from this data with the help of fractal processes.

This chapter is organized as follows. A brief review of point processes and several trade duration models based on estimating intensity of such processes is provided in Section 6.2. In Section 6.3, I introduce the empirical method based on trade duration data for 18 of the component stocks of the Dow Jones. I report the empirical results in Section 6.4. In this section, I compare the goodness of fit of models used in the empirical study with help of the procedure of bootstrap methods I developed. I summarize the conclusions in Section 6.5.

6.2 Point processes in modeling durations

Given a probability space $(\Omega, \mathcal{A}, \mathrm{P})$, a family of random variables $(X_t)_{t\in\mathcal{T}}$ on Ω with values in some set M, (i.e., for all $t \in \mathcal{T}$ and $\mathcal{T}$ is some index set), $X_t : (\Omega, \mathcal{A}) \to (\mathrm{M}, \mathcal{B})$ is defined as a stochastic process with index set $\mathcal{T}$ and state space M. A sequence $(T_n)_{n\in N}$ of positive real random variables is a point process if $T_n(\omega) < T_{n+1}(\omega)$ for all $\omega \in \Omega$ and all $n \in N$; and $\lim_{n\to\infty} T_n(\omega) = \infty$, for all $\omega \in \Omega$. T_n is called the nth arrival time and $T_n = \sum_{i=1}^{n} \tau_i$; and $\tau_n = T_n - T_{n-1}$ (where $\tau_1 = T_1$) is called the nth *waiting time* (*duration*) for the *point process*.

A stochastic process $(N_t)_{t\in[0,\infty)}$ is a *counting process* if: $N_t : (\Omega, \mathcal{A}) \to (\mathrm{N}_0, \mathrm{P}(\mathrm{N}_0))$ for all $t \geq 0$, $N_0 \equiv 0$; $N_s(\omega) \leq N_t(\omega)$, for all $0 \leq s < t$ and all $\omega \in \Omega$; $\lim_{s\to t, s<t} N_s(\omega) = N_t(\omega)$, for all $t \geq 0$ and all $\omega \in \Omega$; $N_t(\omega) - \lim_{s\to t, s<t} N_s(\omega) \in (0, 1)$, for all $t > 0$ and all $\omega \in \Omega$; and $\lim_{t\to\infty} N_t(\omega) = \infty$, for all $\omega \in \Omega$. A point process $(T_n)_{n\in N}$ corresponds to a counting process $(N_t)_{t\in[0,\infty)}$ and vice versa, i.e.,

$$N_t(\omega) = |\{n \in N : T_n(\omega) \leq t\}|$$

for all $\omega \in \Omega$ and all $t \geq 0$. For all $\omega \in \Omega$ and all $n \in N$,

$$T_n(\omega) = \min\{t \geq 0 : N_t(\omega) = n\}$$

The mean value function of the counting process is $m(t) = E(N_t)$, for $t \geq 0$, and $m : [0, \infty) \to [0, \infty), m(0) = 0$. m is an increasing and a right continuous function with $\lim_{t\to\infty} m(t) = \infty$. If the mean value function is differentiable at $t > 0$, then the first-order derivative is called *the intensity of the counting process*. Defining $\lambda(t) = \mathrm{d}m(t)/\mathrm{d}t$,

$$\lim_{\Delta t\to 0, \Delta t\neq 0} \frac{1}{|\Delta t|} P\left(|N_{t+\Delta t} - N_t| = 1\right) = \lambda(t)$$

and

$$\lim_{\Delta t \to 0, \Delta t \neq 0} \frac{1}{|\Delta t|} P\left(|N_{t+\Delta t} - N_t| \geq 2\right) = 0$$

From the viewpoint of the point processes literature (for example, Daley and Vere-Jones (2003)), ultra-high frequency financial data can be described as marked point processes; that is, the state space M is a product space of $R^2 \otimes \mathcal{M}$ where $\mathcal{M}$ is the mark space. The ultra-high frequency transaction data contain two types of processes: time of transactions and events observed at the time of the transaction (see Engle (2000) and Engle and Lunde (2003)). Those events can be identified or described by marks, such as trade prices, posted bid and ask price, and volume. The amount of time between events is the duration. The intensity is used to characterize the point processes and is defined as the expected number of events per time increment considered as a function of time. In survival analysis, the intensity equals the hazard rate. For n durations, $d_1, d_2, ..., d_n$, which are sampled from a population with density function f and corresponding cumulative distribution function F, the survival function $S(t)$ is:

$$S(t) = P[d_i > t] = 1 - F(t)$$

and the intensity or hazard rate $\lambda(t)$ is:

$$\lambda(t) = \lim_{\Delta t \to 0} \frac{P[t < d_i \leq t + \Delta t \mid d_i > t]}{\Delta t}$$

The survival function and the density function can be obtained from the intensity,

$$\lambda(t) = \frac{f(t)}{S(t)} = \frac{-\mathrm{d}\log(S(t))}{\mathrm{d}t}$$

Several models have been proposed to model durations by estimating the intensity[3]. The favored models in the literature are the autoregressive conditional duration (ACD) model proposed by Engle and Russell (1998), the stochastic conditional duration (SCD) model by Bauwens and Veredas (2004), and the stochastic volatility duration (SVD) model by Ghysels et al. (2004). The ACD model expresses the conditional expectation of duration as a linear function of past durations and past conditional expectation. The disturbance is specified as an exponential distribution and as an extension of the Weibull distribution. The SCD model assumes that a latent variable drives the movement of durations. Then the expected durations in the SCD model are treated as the observed durations driven by a latent variable. The SVD model tries to capture the mean and variance of durations. The SCD and SVD models are

[3] See recent reviews of Bauwens and Hautsch (2007) and Sun et al. (2007c).

mixed distributions models. The SCD model combines Weibull and gamma distributions while in the SVD model the durations are expressed as independently and exponentially distributed with a gamma heterogeneity. Extensions to these models have been suggested in the literature. Jasiak (1998) offers the fractional integrated ACD model, Gramming and Maurer (1999) replace the Weibull distribution by the Burr distribution, Bauwens and Giot (2000) propose the logarithmic ACD model, and Zhang et al. (2001) introduce the threshold ACD model. Bauwens and Giot (2003) propose an asymmetric ACD model; Feng et al. (2004) propose a linear non-Gaussian state-space version of the SCD model to capture the leverage effect of the expected durations. Simonsen (2007) extends the ACD model to examine the dependence between durations. Table 6.2 summarizes these studies.

The ACD(m,n) model specified in Engle and Russell (1998) is

$$d_i = \psi_i \, \varepsilon_i$$

$$\psi_i = \omega \; + \; \sum_{j=0}^{m} \alpha_j \, d_{i-j} \; + \; \sum_{j=0}^{n} \beta_j \, \psi_{i-j}$$

where ε_i are the i.i.d. innovations.

Bauwens and Giot (2000) give the logarithmic version of the ACD model as follows

$$d_i = e^{\psi_i} \, \varepsilon_i$$

Two possible specifications of conditional durations are

$$\psi_i = \omega \; + \; \sum_{j=0}^{m} \alpha_j \log d_{i-j} \; + \; \sum_{j=0}^{n} \beta_j \, \psi_{i-j}$$

and

$$\psi_i = \omega \; + \; \sum_{j=0}^{m} \alpha_j \log \varepsilon_{i-j} \; + \; \sum_{j=0}^{n} \beta_j \, \psi_{i-j}$$

Zhang et al. (2001) extend the conditional duration to a switching-regime version. Defining $L_q = [l_{q-1}, l_q)$, and $q = 1, 2, ..., Q$ for a positive integer Q, where $-\infty = l_0 < l_1 < ... < l_q = +\infty$ are the threshold values, d_i follows a q-regime threshold ACD (TACD(m,n)) model; that is:

$$\psi_i = \omega^{(q)} \; + \; \sum_{j=0}^{m} \alpha_j^{(q)} \, d_{i-j} \; + \; \sum_{j=0}^{n} \beta_j^{(q)} \, \psi_{i-j}$$

For example, if there is a threshold value l_h and $0 < h < q$, the TACD(1,1) model can be expressed as following:

$$\psi_i = \begin{cases} \omega_1 \; + \; \alpha_1 \, d_{i-1} \; + \; \beta_1 \, \psi_{i-1} & if \quad 0 < d_{i-1} \leq l_h \\ \omega_2 \; + \; \alpha_2 \, d_{i-1} \; + \; \beta_2 \, \psi_{i-1} & if \quad l_h < d_{i-1} < \infty \end{cases}$$

The threshold l_h determines the regime boundaries. Fernandes and Gramming (2005) propose nonparametric tests for ACD models and suggested practical application for estimation of intraday volatility patterns.

The SCD model given by Bauwens and Veredas (2004) takes the following form:

$$d_i = \Psi_i \, \varepsilon_i$$

where

$$\Psi_i = e^{\psi_i}$$

$$\psi_i = \omega \, + \, \beta \, \psi_{i-1} \, + \, u_i$$

in which $|\beta| < 1$, and u_i is independently normally distributed with zero mean and variance σ^2. Denoting I_{i-1} the information set before d_i, $u_i|I_{i-1} \sim N(0, \sigma^2)$, $\varepsilon_i|I_{i-1}$ follows some distribution with positive support, and u_i is independent of $\varepsilon_j|I_{i-1}$ for any i and j.

Ghysels et al. (2004) proposed a SVD model by assuming that durations are independently and exponentially distributed with gamma heterogeneity. More explicitly, the model can be expressed as:

$$d_i = \frac{U_i}{cV_i}$$

where U_i and V_i are two independent variables with distributions Γ(1,1) (i.e. exponential) and $\Gamma(a, a)$ respectively. Then this expression can be transferred with suitable nonlinear transformations to the expression with Gaussian factors:

$$d_i = \frac{\Psi(1, \Phi(F_1))}{c\Psi(a, \Phi(F_2))} \; = \; \frac{H(1, F_1)}{cH(1, F_2)}$$

where F_1 and F_2 are i.i.d standard normal variables, Φ is the cdf of the standard normal distribution, and $\Psi(a, .)$ is the quantile function of the $\Gamma(a, a)$ distribution.

6.3 Empirical study

In this section, I report the empirical tests that investigate the goodness of fit of several candidate distribution assumptions in modeling roughness of trade duration.

6.3.1 The data

Ultra-high frequency data of 18 Dow Jones index component stocks based on NYSE trading for year 2003 are examined.[4] The companies in the sample are listed in Table 6.1. The sample is considerably larger than other studies that have investigated trade duration. Table 6.2 lists those studies and the stocks included in each one. Note that for the studies that include U.S. stocks, IBM is included in 7 of 11 studies and because the sample size is small, IBM constitutes a major part of those studies. IBM is included in this study also.

The trade durations were calculated for regular trading hours (i.e., overnight trading was not considered). Consistent with Engle and Russell (1998) and Ghysels et al. (2004), open trades are deleted in order to avoid effects induced by the opening auction. Therefore trade durations are considered only from 10:00 to 16:00. In the dataset, I observe many consecutive zero durations, which imply the existence of multiple transactions within a second. I aggregate these intra-second transactions within a second (see Engle and Russell (1998)).

The panels in Figure 6.1 illustrate several sampled trade duration series. These plots show data characteristics that are consistent with the data patterns reported in the literature. For the trade durations of each stock in the study, sample period runs were performed from January 4, 2003 to December 31, 2003. I will let N denote the length of the sample, sub-sample series that have been randomly selected by a moving window with length T ($1 \leq T \leq N$). Replacement is allowed in the sampling. Stoev and Taqqu (2004) suggest that $2^{14} - 6,000 = 10,384$ is the optimal length for a fractional stable noise series to be simulated efficiently. Therefore, in the empirical analysis, sub-sample length (i.e., the window length) of $T = 10,384$ was chosen. A total of 684 sub-samples were randomly created.

6.3.2 The methodology of finding the best model

In the empirical study, I simulate a series for each distributional assumption with and without subordinating them into the ACD(1,1) structure. Then I compare the goodness of fit of the simulated together with the originial trade duration series.

The ACD model can be defined as follows:

$$d_i = \psi_i u_i,$$

[4]The data from were provided by The Securities Industry Research Center of Asia-Pacific in Australia. The Dow Jones index consists of 30 stocks. The whole database I developed included data from 1996 to 2003 for stocks that remained in the index over the entire period. Only 18 stocks satisfied that requirement and I use the data of 2003 in this study.

and

$$\psi_i^2 \;=\; \kappa + \sum_{t=1}^{p} \gamma_i \, d_{i-t} \;+\; \sum_{j=1}^{q} \theta_j \, \psi_{i-j}^2,$$

where d_i are the durations and u_i are i.i.d innovations that can be calculated from d_i/ψ_i. I define

$$\tilde{u_i} \;=\; \frac{d_i}{\hat{\psi}_i},$$

where $\hat{\psi}_i$ is the estimation of ψ_i. In the empirical analysis, an ACD(1,1) model structure is adopted. The objective is to check the statistical characteristics exhibited by trade duration d_i and the error term $\tilde{u_i}$ in ACD(1,1) structure. I simulate d_i and $\tilde{u_i}$ with the ACD(1,1) structure based on the parameters estimated from the empirical series. Then I test the goodness of fit between the empirical series and the simulated series. Six candidate distributional assumptions — lognormal distribution, stable distribution, exponential distribution, Weibull distribution, fractional Gaussian noise, and fractional stable noise[5] — are analyzed for estimation, simulation, and testing. Trade durations are positive numbers, therefore the stable distribution, fractional Gaussian noise, and fractional stable noise are defined on positive supports correspondingly.

6.4 Results

6.4.1 Preliminary Tests

The descriptive statistics of the trade duration data in this study are presented in Table 6.1. The second column shows the number of observations for each stock in this study. The mean, maximum, and minimum of trade duration for each stock are shown in Table 6.1. In this dataset, General Electric and IBM are the most frequently traded stocks. It can be seen that trade durations exhibit diurnal patterns, especially, long duration can be observed at lunch time (see Engle (2000)). The maximum value of trade duration is reported in the fifth column in Table 6.1. In this dataset, I removed the overnight durations.

Engle (1982) proposes a Lagrange-multiplier test for the ARCH phenomenon. A test statistic of lag order q is given by

$$X_q \equiv nR_q^2,$$

[5]The specifications of stable distribution and fractal processes are shown in the Appendix. Further information about stable distribution and fractal processes can be found in Rachev and Mittnik (2000), Sun et al. (2007a) and references therein.

where R_q^2 is the non-centered goodness-of-fit coefficient of a qth order autoregression of the squared residuals taken from the original regression

$$\hat{u}_i^2 = \omega_0 + \omega_1 \hat{u}_{i-1}^2 + \omega_2 \hat{u}_{i-2}^2 + \cdots + \omega_q \hat{u}_{i-q}^2 + e_i,$$

where $\hat{u}$ is the residual in the original regression equation. Under the null hypothesis of the residuals of the original model being normally i.i.d., the ARCH statistic of lag order q follows a χ^2 distribution with q degrees of freedom:

$$\lim_{n\to\infty} X_q \sim \chi_q^2.$$

The test statistics and critical values of Engle-test are presented in Table 6.3, in which I can reject the null hypothesis that there is no ARCH effect at different lag levels for the duration increments. It is clear that an ARCH effect is exhibited in these data. This implies that large durations are followed by large duration and small duration are followed by small duration. This result reveals the phenomenon that when there is information in the market, speed of trading is faster and fast trades are close together. Since trade duration can reflect the information flows in the market, I can explain this result as follows. When there is information, typically there exists a higher proportion of information traders. These traders trade with higher speed and fast trades cluster together. In contrast, when there is less information, the speed of trades slows down till the new information is captured. (The empirical evidence can be found in Dufour and Engle (2000), Engle and Russell (1998), and Engle (2000)).

The Hurst index $H \in (0, 1)$ usually serves as the measure of the tendency of a process and stands for the self-similarity index in Gaussian stochastic processes. It can be somewhat explained by considering the covariance of two consecutive increments. When $H \in (0, 0.5)$, the increments of a process tend to have opposite signs and thus are more zigzagging due to the negative covariance; when $H \in (0.5, 1)$, the covariance between these two increments is positive and less zigzagging of the process; when $H = 0.5$, the covariance between this two increments is zero. It can be stated as follows: If the Hurst index is less than 0.5, the process displays "anti-persistence" which means that the positive excess return is more likely to be reversed and the performance in the next period is likely to be below the average, or on the contrary, the negative excess return is more likely to be reversed and the performance in the next period is likely to be above the average. If the Hurst index is greater than 0.5, the process displays "persistence" which means that the positive excess return or the negative excess return is more likely to be continued and the performance in the next period is likely to be the same as that in the current period. If the Hurst index is equal to 0.5, the process displays no memory, which means the performance in the next period has equal probability to be below and above the performance in the current period. From Table 6.1, I find that the Hurst index has no value of 0.5, which indicates that the memory effect occurs in my samples.

The Hurst index for non-Gaussian stable processes has different bounds for "persistence" and "anti-persistence". For tail index $\alpha \in (0, 2)$, when $H \in (0, 1/\alpha)$, the processes exhibit "anti-persistence", and when $H \in (1/\alpha, 1)$, the processes exhibit "persistence". There is no long-range dependence when $\alpha \in (0, 1]$ because the Hurst index is bounded in the interval $(0, 1)$. When $H = 1/\alpha$, depending on the value of α the processes exhibit either no memory or long-range dependence.[6] From Table 6.1, I find that the Hurst index has no value of $1/\alpha$. Therefore, I find that long-range dependence occurs in my samples.

My results indicate that the Hurst index has no value of 0.5 or $1/\alpha$. As described by Mandelbrot (2005), such patterns are the "wild" forms of roughness which require considering fractal models. The results reported in Table 1 confirm Mandelbrot's statement that "the proper language of the theory of roughness in nature and culture is fractal geometry" (see Mandelbrot (2005, p.193)).

I use the Ljung-Box-Pierce Q-statistic based on the autocorrelation function to test the serial correlation (i.e., the memory effect). The Q-statistic is given as follows:

$$Q :\sim \chi^2_m = N(N+2) \sum_{k=1}^{m} \frac{\rho_k^2}{N-k},$$

where, N denotes the sample size, m the number of autocorrelation lags included in the statistic, and ρ_k the sample autocorrelation at lag order k which is

$$\rho_k = \frac{\sum_{t=1}^{N-k} d_i \, d_{i+k}}{\sum_{t=1}^{N} d_i^2}.$$

The Q-statistic follows an asymptotic χ^2 distribution with m degrees of freedom.

The null hypothesis that there is no serial correlation can be rejected at different lags based on the results reported in Table 6.4. These results are similar to the empirical results in Simonsen (2007). This table implies that the memory effect occurs in each duration series. In order to see when the memory effect vanishes, I compare the Q-statistic with its corresponding critical value. When the quotation of the Q-statistic and the corresponding critical value are less than 1, I cannot reject the null hypothesis that there is no serial correlation. I also shows such quotations in Table 6.4. From this table, all the trade durations exhibit serial correlation. After 500 lags, the memory effect vanishes for 7 stocks and after 1500 lags, the memory effect vanishes for 13 stocks. From the ratios in Table 6.4, I find that the speed of autocorrelation decay is declining, which confirms the effect of long-range dependence.

[6] A detailed discussion see Samorodnitsky and Taqqu (1994) and Cohen and Samorodnitsky (2006).

6.4.2 Goodness of fit test

The Kolmogorov-Smirnov distance (KS) and the Anderson-Darling distance (AD) proposed by Rachev and Mittnik (2000) are used as the criterion for the goodness of fit testing. They are defined as following:

$$KS = \sup_{x\in\Re}|F_s(x) - \tilde{F}(x)|,$$

and

$$AD = \sup_{x\in\Re}\frac{|F_s(x) - \tilde{F}(x)|}{\sqrt{\tilde{F}(x)(1-\tilde{F}(x))}},$$

where $F_s(x)$ denotes the empirical sample distribution and $\tilde{F}(x)$ is the estimated distribution function. The major disadvantage of the KS is that it tends to be more sensitive near the center of the distribution than at the tails. But AD statistic can overcome this. The reliability of testing the empirical distribution will be increased with the help of these two statistics, with the KS distance focusing on the deviations around the median of the distribution and the AD distance on the discrepancies in the tails.

The AD and KS statistics are calculated for the six candidate distributional assumptions. Table 6.5 reports the descriptive statistics of the computed AD and KS statistics. From Table 6.5, fractional stable noise and stable distribution exhibit a smaller mean value for the AD and KS statistics in comparison with the other four distributions. Figure 6.2 shows the boxplot of AD statistics of $\tilde{u}_t$ for the six alternative distributional assumptions investigated. Figure 6.3 shows the boxplot of AD statistics for d_t. Figure 6.4 shows the boxplot of KS statistics of $\tilde{u}_t$ for the six alternative distributional assumptions. Figure 6.5 shows the boxplot of KS statistics of d_t. These figures show that fractional stable noise and stable distribution have a small value of AD and KS statistics, confirming the results reported in Table 6.5. These results indicate that with or without an ACD(1,1) model structure, the fractional stable noise and the stable distribution perform better than the other four tested distributional assumptions based on the criterion for goodness of fit testing.

From Figures 6.2 to 6.5, I can see that the fractional stable noise and the stable distribution fit $\tilde{u}_t$ and d_t better than other distributional assumptions. In order to empirically examine my conjecture, I formulate a statistical test procedure. Because I know that smaller AD and KS statistics mean better goodness of fit, in my test I am going to statistically test how significantly "smaller" AD and KS statistics are. The hypothesis test is:

$$H_0 : \mu_{criterion1} - \mu_{criterion2} \geq 0$$
$$H_1 : \mu_{criterion1} - \mu_{criterion2} < 0$$

where $\mu_{criterion}$ is the mean value of AD or KS statistics of the candidate distributional assumptions investigated. The distributions of AD and KS values are unknown. All AD or KS values are expressed as i.i.d. random variables $X_1, X_2, ..., X_n$, each with distribution function $F_X(\cdot|\theta)$. A $100(1-\alpha)\%$ upper confidence bound (UCB) is defined as $U(X_1, X_2, ..., X_n)$ for a function of $h(\theta)$ if for every θ,

$$P_\theta\left(h(\theta) \leq U(X_1, X_2, ..., X_n)\right) \geq 1-\alpha$$

and $(-\infty, U(X_1, X_2, ..., X_n)]$ is the $100(1-\alpha)\%$ upper confidence interval for $h(\theta)$. Similarly, $L(X_1,\ X_2,\ ..., X_n)$ is a $100(1-\alpha)\%$ lower confidence bound (LCB) for the function $h(\theta)$ for every θ

$$P_\theta\left(h(\theta) \geq L(X_1, X_2, ..., X_n)\right) \geq 1-\alpha$$

and $[L(X_1, X_2, ..., X_n), +\infty)$ is the $100(1-\alpha)\%$ lower confidence interval for $h(\theta)$.

As hypothesis testing and confidence intervals are dual concepts, the hypothesis testing is in fact evaluated by following test rules; that is, (1) if UCB is less than zero, H_0 can be rejected, (2) if LCB is greater than zero, H_0 cannot be rejected, and (3) if H_0 is greater than LCB but at the same time less than UCB, there is no statistically significant conclusion. Employing the bootstrap method introduced in DiCiccio and Efron (1996), the 99% bootstrap confidence intervals are reported in Table 6.6. From this table, at a high confidence level, fractional stable noise and stable distribution are more suitable to modeling trade duration data with or without support of an ACD(1,1) structure.

In comparing the fractional stable noise and stable distribution, it is unclear as to whether the fractional stable noise is better than the stable distribution or vice versa. I compare the supporting cases for the fractional stable noise and stable distribution. I present my comparison result in Table 6.7.

The fractional stable noise has a slightly better support than stable distribution as a single process in modeling trade durations. Since the fractional stable noise belongs to fractal processes, this result confirms the superiority of fractal processes in modeling roughness exhibited in trade duration (see Mandelbrot (2005) and references therein). The stable distribution has a greater number of supporting cases in comparison to the fractional stable noise in modeling duration data with an ACD(1,1) structure. The stable distribution has advantages in capturing heavy-tailedness in observations. As I mentioned above, trade durations reflect the information flows in the market. When there is information in the market, a higher proportion of information traders trade in the market with a higher speed of trading. Similarly, when there is less or no information in the market, a higher proportion of traders trade in the market with a lower speed of trading. Therefore, extreme events (i.e., very short durations and very long durations) are often observed in the market, causing the distribution of durations to have heavy-tails. The ACD model with a stable distribution captures simultaneously both extreme events and clusters of extreme events driven by information in the market.

Several researchers find evidence that fractal processes or the stable distribution are better in modeling market volatility (see, Rachev and Mittnik (2000) and Sun et al. (2007a)). As we know, price adjustments are closely connected with information revealed from the market. When there is information, the proportion of information traders tends to be higher and the speed of price adjustment tends to be faster. Extreme events can be observed both from trade durations and price adjustment. In my study, I show that the fractional stable noise and stable distribution can capture the trade duration movements well. This result explains why fractional stable noise and stable distribution exhibit superiority in modeling price adjustments because of the tight relationship between the speed of price adjustment and information revealed in trade duration. My models capture the situation of fast price adjustments and trade in clusters.

In addition, by analyzing a large sample as I have done in my study, I can detect evidence that some trade durations of different stocks exhibit similar movements, see Figure 6.1. For example, the panels of Eastman Kodak Co. and Merck & Co. and the panels of General Electric and IBM exihibit similar co-movement patterns. The findings of Simonsen (2007) reveal that there is a certain dependence in trade durations of stocks. The empirical results partially support that finding. In this study, most cases provide support that the ACD model with stable distribution is preferred. In this study, the trade duration dynamics exhibit a movement/co-movement which can be captured by the same model. This might lead us to find further evidence of the dependence in my sample with relatively robust tests.

6.5 Conclusions

The empirical research with very few stocks have demonstrated that trade duration data exhibit three characteristics: long-range dependence, heavy tailedness, and clustering. In this chapter, I investigate the presence of these characteristics using a larger number of stocks and investigate whether for modeling trade duration data: (1) a single stochastic process capturing long-range dependence and heavy tailedness and; (2) a relatively powerful distributional assumption in a relatively simple functional structure can be used.

To examine these issues, I introduce fractional stable noise and fractional Gaussian noise to capture long-range dependence and heavy tailedness in modeling the trade duration. In the empirical analysis, I investigate six distributional assumptions (fractional stable noise, fractional Gaussian noise, stable distribution, lognormal distribution, exponential distribution, and Weibull distribution) for modeling the trade duration for 18 Dow Jones index component stocks. By using parameters estimated from the empirical series, I simulate a series for each distributional assumption with and without subordinating them into an ACD(1,1) structure. Then I compare the goodness of fit for these generated series to the empirical series by adopting two test criteria for testing heavy-tailed distributions, the Kolmogorov-Smirnov and Anderson-Darling statistics. A test procedure is formulated based on a bootstrap method, and it is used in order to obtain empirical results.

The above test procedure yields empirical evidence which shows that the stable distribution and fractional stable noise are better in modeling trade duration than the exponential distribution, lognormal distribution, Weibull distribution, and fractional Gaussian noise. The results indicate that residuals from the ACD(1,1) model are more likely to be described by a stable distribution and trade durations exhibit the features of fractional stable noise. That is, stable distribution subordinated with an ACD(1,1) structure and fractional stable noise, demonstrate superior performance in the modeling of trade duration.

My results are consistent with the general findings in the literature on trade duration: informative and short trade durations move prices more than long trade duration. Based on the models I employ, my findings also explain the relationship between price adjustment and information revealed in trade duration. In addition, my results confirm the advantage of fractal models in the study of roughness in trade duration and partially provide evidence for duration dependence.

I argue that it is critical that the findings reported in this chapter be taken into account in modeling trade duration. Many studies have found that stable distribution is a better description of financial data because it can capture heavy tailedness and has a close relationship with long-range dependence. As a self-similar process, fractional stable noise can capture almost all reported stylized facts in financial return data, such as heavy tailedness, long memory, non-Gaussian characters, and clustering. Therefore, if fractional stable noise and stable distribution can be properly employed in financial modeling, more accurate prediction might be realized by well-defined functional models.

Table 6.1: Statistical characteristics of trade duration in 2003 for 18 stocks

Stock	$size$	$mean$	Min	Max	$Hurst_{\tilde{u}_t}$	$Hurst_{d_t}$	$\alpha_{\tilde{u}_t}$	α_{d_t}
Alcoa Inc.	511817	10.8858	0	6194	0.6139	0.7202	1.3291	1.1022
American Express	819489	6.9582	0	6148	0.6379	0.5774	1.5673	1.3316
Caterpillar	571251	10.1333	0	6139	0.6932	0.6289	1.3475	1.0286
E.I.DuPont de Nemours	715515	7.8020	0	6114	0.6107	0.6379	1.3832	1.1930
Walt Disney	813090	6.8123	0	6134	0.6936	0.6943	1.3827	1.2792
Eastman Kodak Co.	475512	11.9756	0	6145	0.6445	0.7108	1.4295	1.1715
General Electric	1188851	4.8295	0	6196	0.4814	0.4725	1.4833	1.4413
General Motors	684082	8.4509	0	6147	0.4647	0.5679	1.2921	1.1606
IBM	986153	5.8425	0	6143	0.5186	0.4952	1.6492	1.2792
Int.Paper Company	606742	9.1403	0	6174	0.6195	0.7110	1.3677	1.0742
Coca-Cola Co.	695948	8.1239	0	6145	0.5676	0.5763	1.3470	1.1606
McDonalds	615302	9.0091	0	6165	0.5548	0.6555	1.3144	1.2437
3M Co.	739258	7.7422	0	6142	0.5819	0.5948	1.3705	1.2349
Altria Group	846140	6.7920	0	6149	0.5396	0.5842	1.3825	1.1282
Merck & Co.	875457	6.6009	0	6135	0.5861	0.5252	1.4886	1.1715
Procter & Gamble	780633	7.2657	0	6166	0.6853	0.5704	1.4437	1.2792
AT&T Inc.	553896	10.1818	0	6161	0.6280	0.6894	1.4133	1.1542
United Technologies	657286	8.5779	0	6161	0.6051	0.6544	1.3176	1.1282

Table 6.2: Studies of Trade Duration and Stocks Included in Sample

Study	Exchange	Stocks	Model distribution(s)
Bauwens (2005)	Tokyo Stock Exchange	Nippon Steel, Sony, Tokyo Electric, Toyota	Generalized gamma
Bauwens and Veredas (2004)	New York Stock Exchange	Boeing, Coca Cola, Walt Disney, Exxon	Weibull
Bauwens and Giot (2000)	New York Stock Exchange	Boeing, Walt Disney, IBM	Exponential, Weibull
Bauwens and Giot (2003)	New York Stock Exchange	Walt Disney, IBM	Weibull
Bauwens et al. (2004)	New York Stock Exchange	Boeing, Coca Cola, Walt Disney, Exxon	Burr, Exponential,
			Generalized gamma, Weibull
Engle and Russell (1998)	New York Stock Exchange	IBM	Exponential, Weibull
Engle and Lunde (2003)	New York Stock Exchange	Bank-American, Walt Disney, General Motors	Exponential
		Federal National Mortagage, McDonald's,	
		Monsanto, Procter & Gamble, Schlumberger	
Feng et al. (2004)	New York Stock Exchange	Boeing, Coca Cola, IBM	Exponential, log-Weibull, log-Gamma
Fernandes and Gramming (2005)	New York Stock Exchange	Exxon	Exponential, Weibull, Burr
			Generalized gamma
Fernandes and Gramming (2006)	New York Stock Exchange	Boeing, Coca Cola, Walt Disney, Exxon, IBM	Burr
Ghysels et al. (2004)	Paris Stock Exchange	Alctel	Exponential, Gamma
Jasiak (1998)	Paris Stock Exchange	Alctel	Weibull
	New York Stock Exchange	IBM	
Zhang et al. (2001)	New York Stock Exchange	IBM	Generalized gamma

Table 6.3: Engle-test for different lags of increments series generated from trade duration data. Test statistic and critical value are presented.

Stock	$lag1$	$lag2$	$lag3$	$lag5$	$lag10$	$lag15$	$lag25$	$lag50$
Alcoa Inc.	25020	33327	37483	41644	45423	46829	48008	48894
American Express	24976	33297	37441	41573	45281	46619	47668	48292
Caterpillar	24977	33322	37467	41629	45410	46814	47993	48880
E.I.DuPont de Nemours	24995	33329	37484	41647	45436	46851	48046	48970
Walt Disney	24999	33332	37490	41654	45439	46853	48045	48965
Eastman Kodak Co.	24967	33315	37466	41632	45417	46826	48012	48911
General Electric	24990	33328	37488	41653	45435	46848	48036	48945
General Motors	24987	33316	37463	41616	45377	46765	47909	48717
IBM	24967	33290	37430	41559	45256	46583	47614	48204
Int. Paper Company	24985	33325	37480	41645	45429	46841	48029	48936
Coca-Cola Co.	24981	33322	37480	41647	45430	46843	48036	48956
McDonalds	24978	33319	37479	41643	45426	46835	48019	48916
3M Co.	25138	33485	37667	41858	45667	47087	48284	49206
Altria Group	24990	33326	37484	41647	45429	46840	48026	48929
Merck & Co.	24988	33325	37481	41648	45437	46852	48049	48978
Procter & Gamble	25015	33332	37479	41639	45421	46833	48023	48937
AT&T Inc.	24975	33320	37475	41642	45430	46843	48035	48953
United Technologies	24973	33318	37473	41639	45427	46839	48031	48950
Critical Value	3.8415	5.9915	7.8147	11.0700	18.3070	24.9960	37.6520	67.5050

Table 6.4: Ljung-Box-Pierce Q-test statistic for different lags at α=0.05. The italic numbers show the ratios of Q-test statistics compared with corresponding critical values.

stock	lag10	lag20	lag50	lag100	lag200	lag500	lag1000	lag1500
Alcoa Inc.	760.3	931.7	1379.5	2041.0	3224.5	5912.9	9897.8	13327.0
	41.5	*29.7*	*20.4*	*16.4*	*13.8*	*10.7*	*9.2*	*8.4*
American Express	330.5	336.2	344.5	352.8	372.8	405.6	596.7	691.3
	18.1	*10.7*	*5.1*	*2.8*	*1.6*	*0.7*	*0.6*	*0.4*
Caterpillar	413.9	446.1	516.6	623.5	814.5	1329.9	2093.9	2899.9
	22.6	*14.2*	*7.6*	*5.0*	*3.4*	*2.4*	*1.9*	*1.8*
E.I.DuPont de Nemours	371.6	401.4	455.6	527.1	644.4	840.9	1184.5	1376.5
	20.3	*12.7*	*6.7*	*4.2*	*2.7*	*1.5*	*1.1*	*0.8*
Walt Disney	294.5	319.2	374.8	452.1	606.5	936.2	1475.7	1620.3
	16.1	*10.1*	*5.5*	*3.6*	*2.5*	*1.6*	*1.3*	*1.0*
Eastman Kodak Co.	157.9	177.6	204.4	257.0	351.2	654.6	1078.8	1530.8
	8.6	*5.6*	*3.0*	*2.0*	*1.5*	*1.1*	*1.1*	*0.9*
General Electric	386.8	391.1	393.1	394.8	397.4	406.6	601.1	627.5
	21.1	*12.4*	*5.8*	*3.1*	*1.6*	*0.7*	*0.5*	*0.3*
General Motors	350.8	356.2	365.7	373.3	396.8	513.5	584.7	832.1
	19.1	*11.3*	*5.4*	*3.0*	*1.6*	*0.9*	*0.5*	*0.5*
IBM	159.9	165.7	169.7	174.0	181.2	195.7	357.1	418.0
	8.7	*5.2*	*2.5*	*1.3*	*0.7*	*0.3*	*0.3*	*0.2*
Int. Paper Company	344.3	400.2	519.7	698.0	986.6	1424.1	1846.4	2299.0
	18.8	*12.7*	*7.6*	*5.6*	*4.2*	*2.5*	*1.7*	*1.4*
Coca-Cola Co.	244.2	253.4	266.0	293.4	318.8	366.1	620.3	758.9
	13.3	*8.0*	*3.9*	*2.3*	*1.3*	*0.6*	*0.5*	*0.4*
McDonalds	233.2	256.1	292.2	346.9	442.6	899.3	1016.4	1343.8
	12.7	*8.1*	*4.3*	*2.7*	*1.8*	*1.6*	*0.9*	*0.8*
3M Co.	515.3	528.2	551.2	579.0	631.5	750.7	1044.4	1340.2
	28.1	*16.8*	*8.1*	*4.6*	*2.6*	*1.3*	*0.9*	*0.8*
Altria Group	279.1	284.2	292.6	299.0	319.1	357.7	556.0	617.1
	15.2	*9.0*	*4.3*	*2.4*	*1.3*	*0.6*	*0.5*	*0.3*
Merck & Co.	262.9	268.0	276.1	280.0	285.9	303.0	461.6	535.6
	14.3	*8.5*	*4.0*	*2.2*	*1.2*	*0.5*	*0.4*	*0.3*
Procter & Gamble	517.6	522.9	533.5	541.1	553.0	580.6	774.7	824.3
	28.2	*16.6*	*7.9*	*4.3*	*2.3*	*1.1*	*0.7*	*0.5*
AT&T Inc.	280.2	319.1	398.1	517.9	691.1	990.2	1370.2	1709.4
	15.3	*10.1*	*5.8*	*4.1*	*2.9*	*1.7*	*1.2*	*1.1*
United Technologies	327.1	356.2	417.2	491.2	621.1	834.9	1130.4	1389.9
	17.8	*11.3*	*6.1*	*3.9*	*2.6*	*1.5*	*1.1*	*0.8*
Critical Value	18.3	31.4	67.5	124.3	233.9	553.1	1074.7	1591.2

Table 6.5: Summary of KS statistics for alternative distributional assumptions, " * " indicates the test for d_t, otherwise for $\tilde{u}_t$. Mean, median, standard deviation ("std"), maximum value ("max"), minimum value ("min") and range of AD and KS statistics are presented in this table.

	AD_{mean}	AD_{median}	AD_{std}	AD_{max}	AD_{min}	AD_{range}	AD^*_{mean}	AD^*_{median}	AD^*_{std}	AD^*_{max}	AD^*_{min}	AD^*_{range}
FGN	2.7056	1.7220	2.7171	15.2900	0.1732	15.1170	4.4772	3.9688	3.0290	19.3230	0.2423	19.0810
fsn	0.4468	0.4207	0.1809	1.1384	0.1066	1.0318	0.4574	0.4307	0.1836	1.2071	0.1014	1.1057
lognormal	2.6100	1.6285	2.7004	15.4960	0.1652	15.3310	4.3202	3.8918	3.0646	19.6470	0.2407	19.4070
stable	0.4460	0.4233	0.1787	1.1384	0.1122	1.0262	0.4583	0.4372	0.1835	1.0795	0.1060	0.9736
exponential	1.0553	1.0516	0.0198	1.1521	0.9719	0.1802	1.0572	1.0526	0.0174	1.1246	1.0254	0.0993
weibull	1.0487	1.0472	0.0220	1.1103	0.9471	0.1632	1.0572	1.0526	0.0174	1.1246	1.0254	0.0993

	KS_{mean}	KS_{median}	KS_{std}	KS_{max}	KS_{min}	KS_{range}	KS^*_{mean}	KS^*_{median}	KS^*_{std}	KS^*_{max}	KS^*_{min}	KS^*_{range}
FGN	0.2615	0.2901	0.0866	0.3955	0.0644	0.3311	0.3157	0.3359	0.0715	0.4054	0.0902	0.3152
fsn	0.0381	0.0361	0.0101	0.0887	0.0197	0.0690	0.0493	0.0484	0.0097	0.0920	0.0288	0.0632
lognormal	0.2590	0.2816	0.0858	0.3984	0.0613	0.3371	0.3129	0.3334	0.0717	0.4030	0.0891	0.3139
stable	0.0378	0.0366	0.0099	0.0861	0.0195	0.0666	0.0495	0.0484	0.0101	0.0923	0.0281	0.0641
exponential	0.5265	0.5249	0.0091	0.5579	0.4856	0.0723	0.5278	0.5257	0.0081	0.5586	0.5126	0.0459
weibull	0.5236	0.5230	0.0105	0.5522	0.4729	0.0793	0.5278	0.5257	0.0081	0.5586	0.5126	0.0459

Table 6.6: Bootstrap 99% confidence intervals for mean of differences in AD and KS statistics, " * " indicates statistics for d_t, otherwise for $\tilde{u}_t$. "FGN" stands for fractional Gaussian noise, "fsn" stands for fractional stable noise, "exp" stands for exponential distribution, "wbl" stands for Weibull distribution.

T:	AD	AD^*	KS	KS^*
$E(T_{stable} - T_{FGN})$	(-2.5223, -1.9851)	(-4.3235 , -3.7124)	(-0.2318 , -0.2157)	(-0.2734 , -0.2594)
$E(T_{stable} - T_{lognormal})$	(-2.4281, -1.8893)	(-4.1642 , -3.5481)	(-0.2292 , -0.2132)	(-0.2706 , -0.2565)
$E(T_{stable} - T_{exp})$	(-0.6278 , -0.5913)	(-0.6175 , -0.5803)	(-0.4895 , -0.4878)	(-0.4793 , -0.4774)
$E(T_{stable} - T_{wbl})$	(-0.6211 , -0.5850)	(-0.6175 , -0.5804)	(-0.4868 , -0.4848)	(-0.4793 , -0.4773)
$E(T_{fsn} - T_{FGN})$	(-2.5236 , -1.9774)	(-4.3204 , -3.7077)	(-0.2315 , -0.2154)	(-0.2736 , -0.2597)
$E(T_{fsn} - T_{lognormal})$	(-2.4265 , -1.8888)	(-4.1638 , -3.5485)	(-0.2290 , -0.2130)	(-0.2708 , -0.2568)
$E(T_{fsn} - T_{exp})$	(-0.6274 , -0.5904)	(-0.6182 , -0.5815)	(-0.4893 , -0.4875)	(-0.4795 , -0.4776)
$E(T_{fsn} - T_{wbl})$	(-0.6203 , -0.5838)	(-0.6185 , -0.5815)	(-0.4866 , -0.4845)	(-0.4795 , -0.4776)
$E(T_{fsn} - T_{stable})$	(-0.0057 , 0.0074)	(-0.0077 , 0.0059)	(-0.0002 , 0.0007)	(-0.0007 , 0.0002)

Table 6.7: Supporting cases comparison of goodness of fit for fractional stable noise and stable distribution based on AD and KS statistics. Symbol " * " indicates the test for d_t, otherwise the test is for $\tilde{u}_t$. Symbol " $\succ$" means being preferred and "$\sim$" means indifference. Numbers shows the supporting cases to the statement in the first column and the number in parentheses give the proportion of supporting cases in the whole sample.

	AD	AD^*	KS	KS^*
$fsn \succ stable$	327	345	327	362
	(47.81%)	(50.44%)	(47.81%)	(52.93%)
$stable \succ fsn$	344	328	351	318
	(50.29%)	(47.95%)	(51.32 %)	(46.49%)
$fsn \sim stable$	13	11	6	4
	(1.90%)	(1.61 %)	(0.87%)	(0.58%)

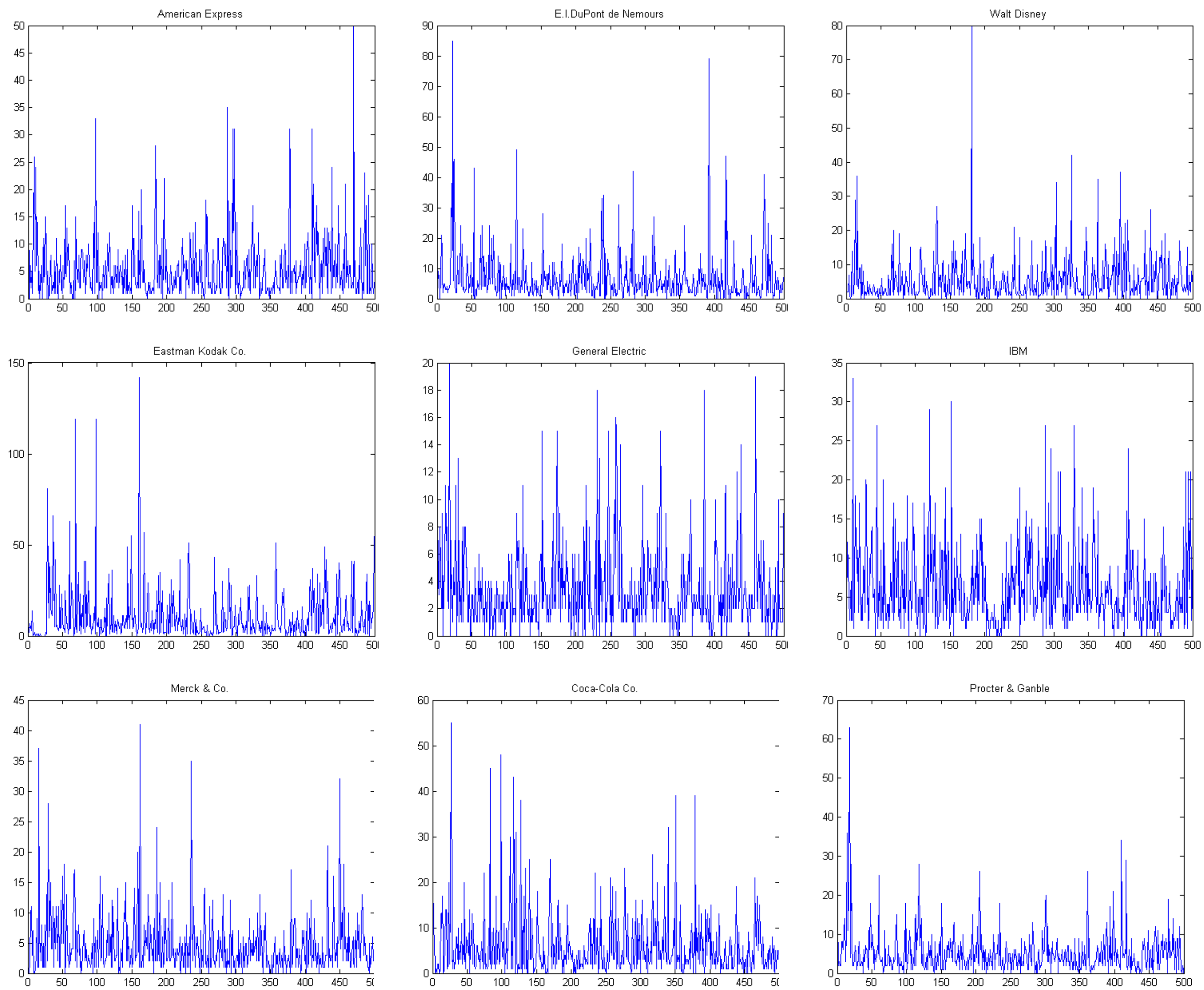

Figure 6.1: Plot of trade duration for several stocks.

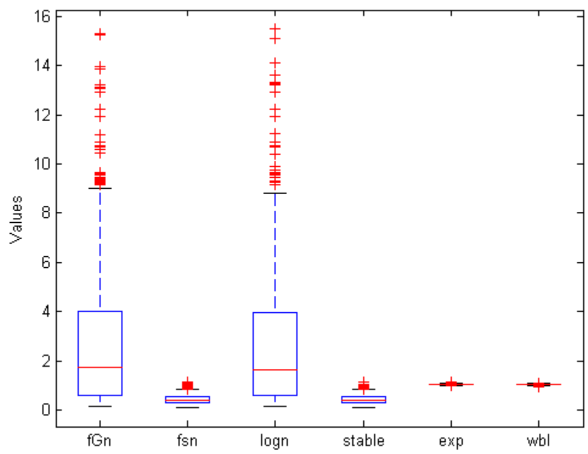

Figure 6.2: Boxplot of AD statistics for $\tilde{u}_t$ in alternative distributional assumptions.

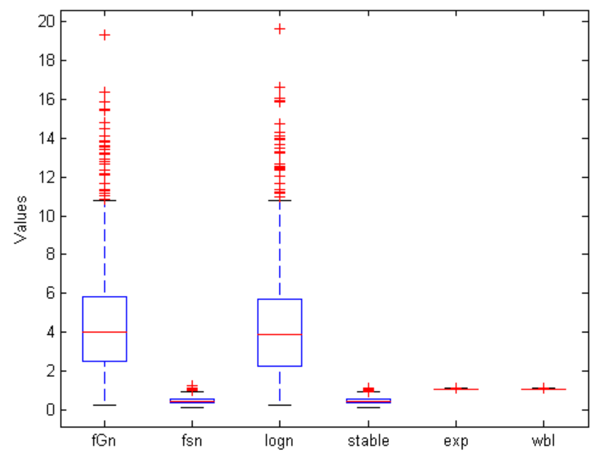

Figure 6.3: Boxplot of AD* statistics for d_t in alternative distributional assumptions.

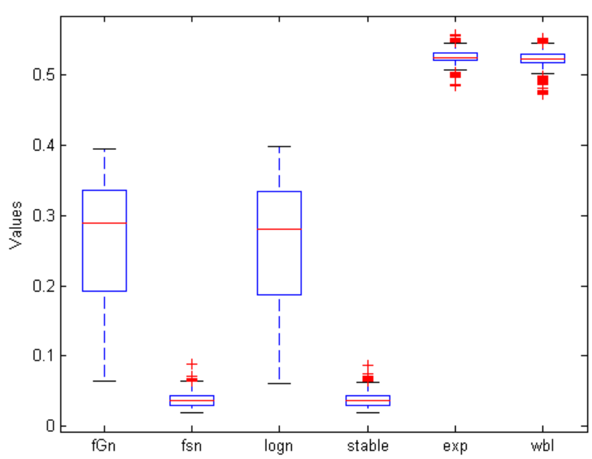

Figure 6.4: Boxplot of KS statistics for $\tilde{u}_t$ in alternative distributional assumptions.

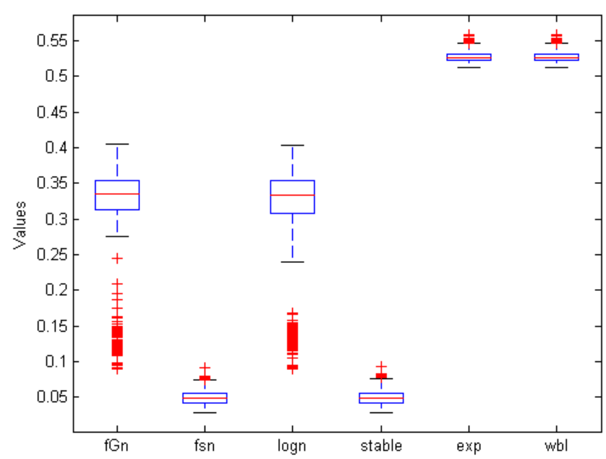

Figure 6.5: Boxplot of KS* statistics for d_t in alternative distributional assumptions.

Chapter 7

Modeling Multivariate High-Frequency Time Series I

7.1 Introduction

The co-movement of world equity markets is often used as a barometer of economic globalization and financial integration. Analyzing such co-movements is important for risk diversification of an international portfolio. The source of co-movement of international equity markets is the volatility-in-correlation effects found by Andersen et al. (2001) in individual stock returns and by Solnik et al. (1996) in international equity index returns. In fact, volatility-in-correlation effect could be explained by the tail dependence of the underlying assets, which shows extreme events happening simultaneously. Co-movement, volatility-in-correlation, and tail dependence in a sense are interrelated when analyzing the dependence structure of international equity markets. It is also found that correlations between consecutive returns decay slowly, that is, long-range dependence in returns is exhibited. Co-movement reflects intercorrelation between underlying asset returns (or returns in different markets) and long-range dependence exhibits autocorrelation within a single asset return (or return of a single market). Therefore, when analyzing international equity markets, I face two dependence structures: the correlation within a single market and the correlation between several markets.

When dealing with the dependence (i.e. long-range dependence) of a single market, I should take other stylized factors into account such as volatility clustering and distributional heavy-tails. It is necessary to treat long-range dependence, volatility clustering, and heavy-tailedness simultaneously in order to obtain more accurate predictions of market volatility. Rachev and Mittnik (2000) note that for modeling financial data, not only does model structure play an important role, but distributional assumptions influence modeling accuracy. The stable Paretian distribution can be used to capture characteristics of financial data since it is rich enough to encompass those stylized facts. Other researchers have shown the advantages of stable distributions in financial modeling (see, Fama (1963), Mittnik and Rachev (1993), Rachev (2003), and Rachev et al. (2005)). Several studies have reported that long-range dependence, self-similar processes, and stable distribution are very closely related (see, Taqqu and Samorodnitsky (1994), Rachev and Mittnik (2000), Rachev and Samorodnitsky (2001), Doukhan et al.. (2003), and Racheva and Samorodnitsky (2003)). Long-range dependence processes are asymptotically second-order self-similar (see Willinger et al. (1998)). Second-order self-similarity describes the property that the correlation structure of a process is preserved irrespective of time scaling. Although self-similarity and long-range dependence are different concepts, in the case of second-order self-similarity, long-range dependence implies self-similarity and vice versa. As to this point, it is natural to employ specified self-similar processes in the study of the within-market dependence together with capturing volatility clustering and heavy-tailedness.

When dealing with the dependent structure between several markets, the usual linear correlation is often applied. But the usual linear correlation is not a satisfactory measure of the dependence among global equity markets for several reasons (see Embrechts et al. (2003), Rachev et al. (2005), and Sun et al. (2007d)). First, when the variance of returns in those markets turns out to be infinite, that is, extreme events are frequently observed, the linear correlation between these markets is undefined. Second, linear correlation assumes that both marginal and joint distributions of returns in these markets are elliptical. In real-world markets, this assumption is unwarranted. Third, the linear correlation is not invariant under nonlinear strictly increasing transformations, implying that the return might be uncorrelated whereas the prices are correlated or vice versa. Fourth, linear correlation only measures the degree of dependence but does not clearly discover the structure of dependence. It has been widely observed that market crashes or financial crises often occur in different countries at about the same time period even when the correlation among those markets is fairly low. The structure of dependence also influences the diversification benefit gained based on a linear correlation measure. Embrechts et al. (2003) and Rachev et al. (2005) illustrate the drawbacks of using linear correlation to analyze dependency. A more prevalent approach which overcomes the disadvantages of linear correlation is to model dependency by using copulas. With the copula method, the nature of dependence that can be modeled is more general and the dependence of extreme events can be considered.

Based on a copula-ARMA-GARCH modeling structure for stock market indexes from nine different countries, in this chapter I compare several candidate specifications using simulation methods. In my modeling structure, the marginal distribution captures the long-range dependence, heavy tails, and volatility clustering simultaneously in order to obtain more accurate predictions, and these marginal distributions are connected by a specified copula. The empirical results indicate that the Student's t copula and ARMA-GARCH model with fractional Gaussian noise dominate the alternative models tested in this study.

I organized this chapter as follows: A brief introduction of copula considering tail dependence is provided in Section 7.2. The data and empirical methodology I used in my study are described in Section 7.3. In Section 7.4, I specify two self-similar processes: fractional Gaussian noise and fractional stable noise. Methods for estimating the parameters of underlying self-similar processes are introduced. In Section 7.5, the simulation methods applied in my empirical study are introduced. The empirical results based on high-frequency data at 1-minute level for nine international stock market indexes are reported in Section 7.6. In that section, I compare the goodness of fit for both the marginal distribution and joint distribution. I summarize my conclusions in Section 7.7.

7.2 Unconditional copulas and tail dependence

In this Section, I show the definition of unconditional copulas and tail dependence. The test for tail dependence applied in this chapter is also introduced.

7.2.1 Definition of unconditional copulas and tail dependence

Copulas enable the dependence structure to be extracted from both the joint distribution function and the marginal distribution functions. From a mathematical viewpoint, a copula function C is a probability distribution function on the n-dimensional hypercube. Sklar (1959) has shown that any multivariate probability distribution function F_Y of some random vector $Y = (Y_1, ..., Y_n)$ can be represented with the help of a copula function C of the following form:

$$
\begin{aligned}
F_Y(y_1, ..., y_n) &= P(Y_1 \leq y_1, \ldots, Y_n \leq y_n) \\
&= C(P(Y_1 \leq y_1), \ldots, P(Y_n \leq y_n)) \\
&= C(F_{Y_1}(y_1), \ldots, F_{Y_n}(y_n)) \qquad (7.0\text{-}1)
\end{aligned}
$$

where F_{Y_i}, $i = 1, \ldots, n$ denote the marginal distribution functions of the random variables, Y_i, $i = 1, \ldots, n$.

When the variables are continuous, the density c associated with the copula is given by:

$$
c(F_{Y_1}(y_1), \ldots, F_{Y_n}(y_n)) = \frac{\partial^n C(F_{Y_1}(y_1), \ldots, F_{Y_n}(y_n))}{\partial F_{Y_1}(y_1), \ldots, \partial F_{Y_n}(y_n)}. \qquad (7.0\text{-}2)
$$

The density function f_Y corresponding to the n-variate distribution function F_Y is

$$f_Y(y_1, ..., y_n) = c(F_{Y_1}(y_1), \ldots, F_{Y_n}(y_n)) \prod_{i=n}^{n} f_{Y_i}(y_i), \tag{7.0-3}$$

where f_{Y_i}, $i = 1, \ldots, n$ is the density function of F_{Y_i}, $i = 1, \ldots, n$ (see, Joe (1997), Cherubini et al. (2004), and Nelsen (2006)).

Two commonly used unconditional copulas are the unconditional Gaussian copula and unconditional Student's t copula. The unconditional Gaussian copula and unconditional Student's t copula are specified in this section. The multivariate version of these two copulas are given as follows. Let ρ be the correlation matrix which is a symmetric, positive definite matrix with unit diagonal, and Φ_ρ the standardized multivariate normal distribution with correlation matrix ρ. The unconditional multivariate Gaussian copula is then

$$C(u_1, \ldots, u_n; \rho) = \Phi_\rho\Big(\Phi^{-1}(u_1), \ldots, \Phi^{-1}(u_n)\Big),$$

and the corresponding density is

$$c(u_1, \ldots, u_n; \rho) = \frac{1}{|\rho|^{1/2}} \exp\Big(-\frac{1}{2}\lambda^T(\rho^{-1} - I)\lambda\Big),$$

where $\lambda = (\Phi^{-1}(u_1), \ldots, \Phi^{-1}(u_n))^T$ and $u_i, i = 1, 2, \ldots, n$ are the margins.

The unconditional (standardized) multivariate Student's t copula $T_{\rho,\nu}$ can be expressed as

$$T_{\rho,\nu}(u_1, \ldots, u_n; \rho) = t_{\rho,\nu}\Big(t_\nu^{-1}(u_1), \ldots, t_\nu^{-1}(u_n)\Big),$$

where $t_{\rho,\nu}$ is the standardized multivariate Student's t distribution with correlation matrix ρ and ν degrees of freedom and t_ν^{-1} is the inverse of the univariate cumulative density function (c.d.f) of Student's t with ν degrees of freedom. The density of the unconditional multivariate Student's t copula is

$$c_{\rho,\nu}(u_1, \ldots, u_n; \rho) = \frac{\Gamma(\frac{\nu+n}{2})}{\Gamma(\frac{\nu}{2})|\rho|^{1/2}} \Big(\frac{\Gamma(\frac{\nu}{2})}{\Gamma(\frac{\nu+1}{2})}\Big)^n \left(\frac{\Big(1 + \frac{1}{\nu}\lambda^T\rho^{-1}\lambda\Big)^{-\frac{\nu+n}{2}}}{\prod_{j=1}^n \Big(1 + \frac{\lambda_j^2}{\nu}\Big)^{-\frac{\nu+1}{2}}}\right),$$

where $\lambda_j = t_\nu^{-1}(u_j)$ and $u_j, j = 1, 2, \ldots, n$ are the margins.

In financial data, we can observe that extreme events happen simultaneously for different assets. In a time interval, several assets might exhibit extreme values. *Tail dependence* reflects the dependence structure between extreme events. It turns out that tail dependence is a copula property. Letting $(Y_1, Y_2)^T$ be a vector of continuous random variables with marginal distribution functions F_1, F_2, then the coefficient of the upper tail dependence of $(Y_1, Y_2)^T$ is

$$\lambda_U = \lim_{u \to 1} P\Big(Y_2 > F_2^{-1}(u) | Y_1 > F_1^{-1}(u)\Big), \tag{7.0-4}$$

and the coefficient of the lower tail dependence of $(Y_1, Y_2)^T$ is

$$\lambda_L = \lim_{u \to 0} P\Big(Y_2 < F_2^{-1}(u) | Y_1 < F_1^{-1}(u)\Big). \tag{7.0-5}$$

If $\lambda_U > 0$, there exists upper tail dependence and the positive extreme values can be observed simultaneously. If $\lambda_L > 0$, there exists lower tail dependence and the negative extreme values can be observed simultaneously. Embrechts et al. (2003) introduce some coefficients of tail dependence of different copulas.

Empirical studies have failed to support the assumption that return data follow a Gaussian distribution. Instead, return data often exhibit excess kurtosis and heavy tails. Nor is the multivariate Gaussian distribution warranted either when studying multi-dimensional return data because of tail dependence among them. The Student's t copula and Clayton copulas can characterize tail dependence in multi-dimensional return data but the Gaussian copula can not. In Clayton copulas, only Joe-Clayton copula can capture both the lower and upper tail dependence and the Joe-Clayton copula allows for asymmetric tail dependence. Unfortunately, the Joe-Clayton copula cannot be implemented for multi-dimensional return data because of computational difficulties, while the Student's t copula can be implemented to capture tail dependence in multi-dimensional return data (see, Embrechts et al. (2003), Patton (2005), and Nelsen (2006)).

7.2.2 Test of tail dependence

Asymmetric correlations have been observed by several studies (see for example, Ang and Chen (2002), Login and Solnik (2001), Hu (2006), and Hong et al. (2006)) that show stocks tend to have greater correlations with market when the market goes down than it goes up. If the symmetry of stock returns is rejected by certain tests, then the data cannot be modelled by any symmetric distribution and correspondingly the symmetric copulas are not significantly reliable for the multivariate correlation modeling.

Following Ang and Chen (2002) and Login and Solnik (2000), based on tail dependence defined by equation (7.0-4) and (7.0-5), Hong et al. (2006) propose a test procedure for exceedance correlation. I adopt this procedure in this chapter to test exceedance correlation. Let X_{1t}, X_{2t} be the returns of two returns in period t, and q the exceedance level. Then define

$$\rho^+(q) = \text{corr}(X_{1t}, X_{2t} | X_{1t} > q, X_{2t} > q)$$

and

$$\rho^-(q) = \text{corr}(X_{1t}, X_{2t} | X_{1t} < -q, X_{2t} < -q).$$

The null hypothesis of a symmetric correlation is $H_0 : \rho^+(q) = \rho^-(q)$ for all $q > 0$. As Hong et al. (2003) point out if the null hypothsis is true, the following $m \times 1$ difference vector

$$\hat{\rho}^+ - \hat{\rho}^- = [\hat{\rho}^+(q_1) - \hat{\rho}^-(q_1), ..., \hat{\rho}^+(q_m) - \hat{\rho}^-(q_m)]'$$

must be close to zero, where $q_1, q_2, ..., q_m$ are m chosen exceedance levels. Hong et al. (2006) show that under the null hypothesis of symmetry and certain regularity conditions, the vector $\hat{\rho}^+ - \hat{\rho}^-$ has an asymptotic normal distribution with zero mean and a positive definite variance-covariance matrix Ω. In order to estimate Ω to obtain a feasible test statistic, the sample means and variances of the two conditional series are computed:

$$\hat{\mu}_1^+(q) = \frac{1}{T_q^+} \sum_{t=1}^{T} X_{1t} 1(X_{1t} > q, X_{2t} > q)$$
$$\hat{\mu}_2^+(q) = \frac{1}{T_q^+} \sum_{t=1}^{T} X_{1t} 1(X_{1t} > q, X_{2t} > q)$$

$$\hat{\sigma}_1^+(q)^2 = \frac{1}{T_q^+ - 1} \sum_{t=1}^{T} [X_{1t} - \hat{\mu}_1^+(q)]^2 \, 1(X_{1t} > q, X_{2t} > q)$$
$$\hat{\sigma}_2^+(q)^2 = \frac{1}{T_q^+ - 1} \sum_{t=1}^{T} [X_{2t} - \hat{\mu}_2^+(q)]^2 \, 1(X_{1t} > q, X_{2t} > q)$$

where T_q^+ is the number of the observations for both X_{1t} and X_{2t} are larger than q and $1(\cdot)$ is the indicator function. The conditional correlation $\hat{\rho}^+(q)$ is then

$$\hat{\rho}^+(q) = \frac{1}{T_q^+ - 1} \sum_{t=1}^{T} [Y_{1t}(q) Y_{2t}(q)] \, 1(X_{1t} > q, X_{2t} > q)$$

where

$$Y_{1t}(q) = \frac{X_{1t} - \hat{\mu}_1^+(q)}{\hat{\sigma}_1^+(q)}$$
$$Y_{2t}(q) = \frac{X_{2t} - \hat{\mu}_2^+(q)}{\hat{\sigma}_2^+(q)}$$

and similarly the conditional correlation $\hat{\rho}^-(q)$ can be derived.

Hong et al. (2006) show that under general conditions, a consistent estimator of Ω is:

$$\hat{\Omega} = \sum_{l=1-T}^{T-1} k(l/p) \hat{\gamma}_l$$

where $\hat{\gamma}_l$ is an $N \times N$ with (i, j)-th element

$$\hat{\gamma}_l(q_i, q_j) = \frac{1}{T} \sum_{t=|l|+1}^{T} \delta_t(q_i) \delta_{t-l}(q_j)$$

$$\delta_t(q) = \frac{T}{T_q^+}[Y_{1t}^+(q)Y_{2t}^+(q)]1(X_{1t} > q, X_{2t} > q) - \frac{T}{T_q^-}[Y_{1t}^-(q)Y_{2t}^-(q)]1(X_{1t} < -q, X_{2t} < -q)$$

$k(\cdot)$ is a kernel function assigning weights to each lag of order l, and p is the smoothing parameter. Then the test statistic for the null hypothesis of symmetry is

$$J_\rho = T(\hat{\rho}^+ - \hat{\rho}^-)'\hat{\Omega}^{-1}(\hat{\rho}^+ - \hat{\rho}^-).$$

With choosing a suitable smoothing parameter p, Hong et al. (2006) provide that the symmetry test has a simple asymptotic χ^2 distribution with m degrees of freedom.

7.3 Data and empirical methodology

7.3.1 Data

In previous studies of the co-movement of international equity markets, low-frequency data have been examined. Because stock indexes change their composition quite often over time, it is difficult to find the impact of these changes in composition when analyzing the return history of stock indexes using low-frequency data. Dacorogna et al. (2001) calls this phenomenon the "breakdown of the permanence hypothesis". In order to overcome this problem, I use high-frequency data in this study.

Employing high-frequency data has several advantages compared with low-frequency data. First, with a very large amount of observations, high-frequency data offers a higher level of statistical significance. Second, high-frequency data are gathered at a low level of aggregation, thereby capturing the heterogeneity of players in financial markets. These players should be properly modeled in order to make valid inferences about market movements. Low-frequency data, say daily or weekly data, aggregate the heterogeneity in a smoothing way. As a result, many of the movements in the same direction are strengthened and those in the opposite direction cancelled in the process of aggregation. The aggregated series generally show smoother behavior than their components. The relationships between the observations in these aggregated series often exhibit greater smoothness than their components. For example, a curve exhibiting a one-week market movement based on daily return data might be a line with a couple of nodes. The smooth line segment veils the intra-daily fluctuation of the market. But high-frequency data can reflect such intra-daily fluctuations and the intra-daily co-movement can be taken into account. Third, using high-frequency data in analyzing the co-movement of international equity markets can consider both microstructure effects and macroeconomic factors. This is because information contained in high-frequency data can be resolved into a higher frequency part (i.e., the intra-daily fluctuation) and a lower frequency part (i.e., low-frequency smoothness). The information provided by the higher frequency part mirrors the microstructure effect of the equity markets and the information in the lower frequency part shows the smoothed trend that is usually influenced by macroeconomic factors in these markets.

Standard econometric techniques are based on homogeneous time series analysis. If a researcher uses analytic methods of homogeneous time series for inhomogeneous time series, the reliability of the results will be doubtful. Aggregating inhomogeneous tick-by-tick data to the equally spaced (homogeneous) time series is needed. Engle and Russell (1998) argue that for aggregating tick-by-tick data to a fixed time interval, if a short time interval is chosen, there will be many intervals in which there is no new information, and if choosing a wide interval, micro-structure features might be missing. Aït-Sahalia (2005) suggests keeping the data at the ultimate frequency level. In my empirical study, intra-daily data, which I refer to as the high-frequency data in this chapter, at 1-minute level were aggregated from tick-by-tick data to investigate the co-movement of international equity markets.

The high-frequency data of the nine international stock indexes listed in Table 1 from January 8, 2002 to December 31, 2003 were aggregated to the 1-minute frequency level. The aggregation algorithm is based on the linear interpolation introduced by Wasserfallen and Zimmermann (1995). That is, given an inhomogeneous series with times t_i and values $\varphi_i = \varphi(t_i)$, the index i identifies the irregularly spaced sequence. The target homogeneous time series is given at times $t_0 + j\Delta t$ with fixed time interval Δt starting at t_0. The index j identifies the regularly spaced sequence. The time $t_0 + j\Delta t$ is bounded by two times t_i of the irregularly spaced series, $I = \max(i \,|t_i \leq t_0 + j\Delta t)$ and $t_I \leq t_0 + j\Delta t > t_{I+1}$. Data are interpolated between t_I and t_{I+1}. The linear interpolation shows that

$$\varphi(t_0 + j\Delta t) \;=\; \varphi_I \;+\; \frac{t_0 + j\Delta t - t_I}{t_{I+1} - t_I}\,(\varphi_{I+1} - \varphi_I). \tag{7.0-6}$$

Dacorogna et al. (2001) pointed out that linear interpolation relies on the future of time and Müller et al. (1990) suggests that linear interpolation is an appropriate method for stochastic processes with independent and identically distributed (i.i.d.) increments.

Empirical evidence has shown seasonality in high-frequency data. In order to remove such disturbances, several methods of data adjusting have been adopted in modeling. Engle and Russell (1998) and other researchers adopt several methods to adjust the seasonal effect in the data. In my study, seasonality is treated as one type of self-similarity. Consequently, it is not necessary to adjust for the seasonal effect in the data.

7.3.2 Empirical methodology

To investigate the co-movement of international stock markets, I use the market index for each country as the proxy for the market movement and propose the copula ARMA-GARCH model. This model is implemented with an ARMA-GARCH model for the marginal distributions and a copula for the joint distribution. Six GARCH models with different kinds of residuals (i.e., residuals with forms of white noise, fractional Gaussian noise, fractional stable noise, stable distribution, generalized Pareto distribution, and generalized extreme value distribution) for the marginal distributions are simulated. After goodness of fit testing, I use the best goodness of fit model for the marginal distributions with Gaussian copula and Student's t copula for the joint distribution to simulate the returns on the equity indexes. Then, the models will be tested with several goodness of fit test methods for a large dataset.

I define the ARMA-GARCH model for the conditional mean equation as:

$$y_t = \alpha_0 + \sum_{i=1}^{r} \alpha_i \, y_{t-i} + \varepsilon_t + \sum_{j=1}^{m} \beta_j \varepsilon_{t-j}. \tag{7.0-7}$$

Let $\varepsilon_t = \sigma_t \, u_t$, where the conditional variance of the innovations, σ_t^2, is by definition

$$Var_{t-1}(y_t) = E_{t-1}(\varepsilon_t^2) = \sigma_t^2. \tag{7.0-8}$$

The general GARCH(p,q) processes for the conditional variance of the innovation is then

$$\sigma_t^2 = \kappa + \sum_{i=1}^{p} \gamma_i \, \sigma_{t-i}^2 + \sum_{j=1}^{q} \theta_j \, \varepsilon_{t-j}^2. \tag{7.0-9}$$

Since $\varepsilon_t = \sigma_t \, u_t$, u_t could be calculated from ε_t / σ_t. Defining

$$\tilde{u_t} = \frac{\varepsilon_t^s}{\hat{\sigma_t}}, \tag{7.0-10}$$

where ε_t^s is estimated from the sample and $\hat{\sigma_t}$ is the estimation of σ_t. In my study, ARMA(1,1)-GARCH(1,1) are parameterized as marginal distributions with different kinds of u_t (i.e., normal distribution, fractional Gaussian noise, fractional stable noise, stable distribution, generalized Pareto distribution, and generalized extreme value distribution).

From the goodness of fit testing for marginal distributions, I find the best fit model. Then taking the best fit model as marginal distributions for each stock index return, I simulate a multivariate Gaussian copula and Student's t copula for the dependence structure of the nine stock index returns. The simulation method adopted is introduced in Section 7.4.

Correlations reported in several studies (see for example, Ang and Chen (2002), Login and Solnik (2001), and Hong et al. (2006)) show that stocks tend to have greater correlations during periods when the market goes down than during periods of rising prices. If symmetry of stock returns is rejected by certain tests, then the data cannot be modelled by any symmetric distribution and correspondingly the symmetric copulas are not significantly reliable for multivariate correlation modeling. I compute the J_ρ statistic (that follows the asymptotic χ^2 distribution) proposed by Hong et al. (2006) to test the null hypothesis of symmetric correlation in my dataset.

The Kolmogorov-Smirnov distance (KS) and the Anderson-Darling distance (AD) suggested by Rachev and Mittnik (2000) and the Cramer Von Mises distance (CVM)[1] are used as the criterion for the goodness of fit testing.

7.4 Analysis of the marginal distribution

As mentioned in the previous section, I apply ARMA(1,1)-GARCH(1,1) with alternative distributions for residuals u_t in my empirical study to model the marginal distribution of each equity market return. The reason I only consider ARMA-GARCH at lag order 1 is to make my analysis satisfy the stationary conditions for the processes I adopted for the marginal distribution (see Mittnik et al. (2001)). In addition, I use 1-minute level of high frequency data: higher lag order contains less information because of the long memory effect (see Sun et al. (2007b)). The key point of the marginal distributions is to empower the residuals u_t to capture the stylized factors, such as, long-range dependence and heavy-tailedness. One of the powerful forms of u_t is the self-similar process. Two specified self-similar processes applied in the empirical study are the fractional Gaussian noise and the fractional stable noise. The reason why such self-similar processes are powerful is that they impose an index on quantifying the degree of long-range dependence and measuring self-similarity. In this section, based on the residuals u_t in ARMA(1,1)-GARCH(1,1) model, I introduce how to estimate the parameters of the two

[1] Specifically, these criterion are defined as follows:

$$KS = \sup_{x\in\Re}\left|F_s(x) - \tilde{F}(x)\right|,$$

$$AD = \sup_{x\in\Re}\frac{\left|F_s(x) - \tilde{F}(x)\right|}{\sqrt{\tilde{F}(x)(1-\tilde{F}(x))}},$$

and

$$CVM = \int_{-\infty}^{\infty}\left(F_s(x) - \tilde{F}(x)\right)^2 d\tilde{F}(x),$$

where $F_s(x)$ denotes the empirical sample distribution and $\tilde{F}(x)$ is the estimated distribution function.

specified self-similar processes for u_t.

7.4.1 The self-similarity parameter

Self-similarity is defined by Samorodnitsky and Taqqu (1994) as follows. Let T be either $R, R_+ = \{t : t \geq 0\}$ or $\{t : t > 0\}$. The real-valued process $\{X(t), t \in T\}$ is self-similar with Hurst index $H > 0$, if for any $a > 0$ and $d \geq 1$, $t_1, t_2, ..., t_d \in T$, satisfying:

$$\Big(X(at_1), X(at_2), ..., X(at_d)\Big) \stackrel{d}{=} \Big(a^H X(t_1), a^H X(t_2), ...a^H X(t_d)\Big). \tag{7.0-11}$$

The Hurst index H plays a key role in such processes to capture long-range dependence.

Let $\phi(k)$ denote the kth-order autocovariance function of $\{X(t), t \in T\}$, for $0 < H < 1$, and $H \neq 0.5$, $\phi(k) \sim H(2H-1)k^{2(H-1)}$ holds. $\{X(t), t \in T\}$ is called long-range dependence if $\sum_{k=1}^{\infty} \phi(k) = +\infty$. $\{X(t), t \in T\}$ is called short-range dependence if $\sum_{k=1}^{\infty} \phi(k) < +\infty$. If $0 < H < 0.5$, then $\sum_{k=1}^{\infty} \phi(k) \sim \sum_{k=1}^{\infty} H(2H-1)k^{2(H-1)}$. Note that in this case, $2(H-1) < -1$, $\sum_{k=1}^{\infty} k^{2(H-1)} < +\infty$. Thus $X(t)$ exhibits short-range dependence. If $0.5 < H < 1$, then $2(H-1) > -1$, thus $\sum_{k=1}^{\infty} k^{2(H-1)} = +\infty$, $X(t)$ shows long-range dependence. For all $k \geq 1$, if $H = 0.5$, the autocovariance is zero and $X(t)$ is a random walk; if $H = 1$, then $\phi(k) = 1$ and I have the degenerate situation which shows no memory effect and the process is not autocorrelated at any lag; and if $H > 1$, then $\phi(k) > 1$ and that is impossible.

Several methods for estimating the Hurst index have been proposed (see, Beran (1994)). Applying the method of calculating the R/S statistic proposed by Hurst (1951), I estimated the Hurst index of the nine international equity index returns and show the results in Table 7.1.

7.4.2 Specification of the self-similar processses

In this section, specification of two self-similar processes used in my empirical study, fractional Gaussian noise and fractional stable noise, are introduced. Samorodnitsky and Taqqu (1994) clarified the definition of FBM as a Gaussian process having self-similarity index H and stationary increments (see also Mandelbrot and Wallis (1968)). Mandelbrot and van Ness (1968) defined the stochastic representation

$$B_H(t) := \frac{1}{\Gamma(H+\frac{1}{2})}\Big(\int_{-\infty}^{0}[(t-s)^{H-\frac{1}{2}} - (-s)^{H-\frac{1}{2}}]dB(s) + \int_0^t (t-s)^{H-\frac{1}{2}} dB(s)\Big), \tag{7.0-12}$$

where $\Gamma(\cdot)$ represents the Gamma function:

$$\Gamma(a) := \int_0^{\infty} x^{a-1} e^{-x} dx,$$

and $0 < H < 1$ is the Hurst parameter. The integrator B is ordinary Brownian motion. The main difference between fractional Brownian motion and ordinary Brownian motion is that the increments in Brownian motion are independent while in fractional Brownian motion they are dependent. As for the fractional Brownian motion, Samorodnitsky and Taqqu (1994) define its increments $\{Y_j, j \in Z\}$ as fractional Gaussian noise (FGN), which is for $j = 0, \pm 1, \pm 2, ...,$ $Y_j = B_H(j-1) - B_H(j)$.

Fractional Brownian motion can capture the effect of long-range dependence, but with less power to capture the heavy tailedness. The existence of abrupt discontinuities in return data, combined with the empirical observation of sample excess kurtosis and unstable variance, suggests that the return series can best be described by a stable Paretian distribution (see, Mandelbrot (1963, 1983)). It is natural to introduce the stable Paretian distribution in self-similar processes in order to capture both long-range dependence and heavy tailedness. Samorodinitsky and Taqqu (1994) introduce the α-stable *H-sssi* processes $\{X(t), t \in R\}$ with $0 < \alpha < 2$. If $0 < \alpha < 1$, the values of the Hurst index are $H \in (0, 1/\alpha]$ and if $1 < \alpha < 2$, the values of the Hurst index are $H \in (0, 1]$. There are several extensions of fractional Brownian motion to the stable distribution. The most commonly used is the linear fractional stable motion (also called linear fractional Lévy motion), which is defined by Samorodinitsky and Taqqu (1994).[2] In this chapter, if there is no special indication, fractional stable noise (fsn) is generated from linear fractional stable motion.

7.4.3 Estimation of the self-similarity parameter

Beran (1994) discusses the approximate maximum likelihood estimator (MLE)[3] of the self-similarity parameter. For fractional Gaussian noise, Y_t, let $f(\lambda; H)$ denote the power spectrum of Y after being normalized to have variance 1 and let $I(\lambda)$ denote the periodogram of Y_t; that is,

$$I(\lambda) = \frac{1}{2\pi N} \left| \sum_{t=1}^{N} Y_t \, e^{i\,t\,\lambda} \right|^2 . \tag{7.0-13}$$

The MLE of H is to find $\hat{H}$ that minimizes

$$g(\hat{H}) = \int_{-\pi}^{\pi} \frac{I(\lambda)}{f(\lambda; \hat{H})} d\lambda. \tag{7.0-14}$$

[2]Some properties of these processes have been discussed in Mandelbrot and Van Ness (1968), Maejima and Rachev (1987), Manfields et al.. (2001), Rachev and Mittnik (2000), Rachev and Samorodnitsky (2001), Samorodnitsky (1994), and Samorodinitsky and Taqqu (1994).

[3]It is also called the Whittle estimator.

Stoev et al. (2002) proposed the least-squares (LS) estimator for the Hurst index based on the finite impulse response transformation (FIRT) and wavelet transform coefficients of the fractional stable motion. FIRT is a filter $v = (v_0, v1, ..., v_p)$ of real numbers $v_t \in \Re, t = 1, ..., p$, and length $p + 1$. It is defined for X_t by

$$T_{n,t} = \sum_{i=0}^{p} v_i \, X_{n(i+t)}, \tag{7.0-15}$$

where $n \geq 1$ and $t \in N$. The $T_{n,t}$ are the FIRT coefficients of X_t (i.e., the FIRT coefficients of the fractional stable motion). The indices n and t can be explained as "scale" and "location". If $\sum_{i=0}^{p} i^r v_i = 0$, for $r = 0, ..., q-1$, but $\sum_{i=0}^{p} i^q v_i \neq 0$, the filter v_i can be said to have $q \geq 1$ zero moments. If $\{T_{n,t}, n \geq 1, t \in N\}$ is the FIRT coefficients of fractional stable motion with the filter v_i that has at least one zero moment, Stoev et al. (2002) prove the following properties of $T_{n,t}$: (1) $T_{n,t+h} \stackrel{d}{=} T_{n,t}$, and (2) $T_{n,t} \stackrel{d}{=} n^H T_{1,t}$, where $h, t \in N$, $n \geq 1$. I assume that $T_{n,t}$ are available for the fixed scales n_j $j = 1, ..., m$ and locations $t = 0, ..., M_j - 1$ at the scale n_j, since only a finite number, say M_j, of the FIRT coefficients are available at the scale n_j. By using these two properties, I have

$$E \log |T_{n_j,0}| = H \log n_j + E \log |T_{1,0}|. \tag{7.0-16}$$

The left-hand side of this equation can be approximated by

$$Y_{\log}(M_j) = \frac{1}{M_j} \sum_{t=0}^{M_j-1} \log |T_{n_j,t}|. \tag{7.0-17}$$

Then I obtain

$$\begin{pmatrix} Y_{\log}(M_1) \\ \vdots \\ Y_{\log}(M_m) \end{pmatrix} = \begin{pmatrix} \log n_1 & 1 \\ \vdots & \vdots \\ \log n_m & 1 \end{pmatrix} \begin{pmatrix} H \\ E \log |T_{1,0}| \end{pmatrix} + \begin{pmatrix} \sqrt{M}\Big(Y_{\log}(M_1) - E \log |T_{n_1,0}|\Big) \\ \vdots \\ \sqrt{M}\Big(Y_{\log}(M_m) - E \log |T_{n_m,0}|\Big) \end{pmatrix}. \tag{7.0-18}$$

I can express equation (7.0-18) as follows

$$Y = X\theta + \frac{1}{\sqrt{M}} \varepsilon, \tag{7.0-19}$$

where ε is the vector showing the difference between $\sqrt{M} Y_{\log}(M_m)$ and $\sqrt{M} E(\log |T_{n_m,0}|)$. Equation (7.0-19) shows that the self-similarity parameter H can be estimated by a standard linear regression of the vector Y against the matrix X. Stoev et al. (2002) provide the details for implementing such a procedure.

7.4.4 The parameters of a stable Non-Gaussian distribution

A stable distribution requires four parameters for complete description: an index of stability $\alpha \in (0, 2]$ also called the tail index, a skewness parameter $\beta \in [-1, 1]$, a scale parameter $\gamma > 0$, and a location parameter $\zeta \in \Re$. There is unfortunately no closed-form expression for the density function and distribution function of a stable distribution. Rachev and Mittnik (2000) give the definition for the stable distribution: A random variable X is said to have a stable distribution if there are parameters $0 < \alpha \leq 2$, $-1 \leq \alpha \leq 1$, $\gamma \geq 0$ and ζ real such that its characteristic function has the following form:

$$E\exp(i\theta X) = \begin{cases} \exp\{-\gamma^\alpha|\theta|^\alpha(1 - i\beta(\text{sign}\theta)\tan\frac{\pi\alpha}{2}) + i\zeta\theta\}, & if \quad \alpha \neq 1 \\ \exp\{-\gamma|\theta|(1 + i\beta\frac{2}{\pi}(\text{sign}\theta)\ln|\theta|) + i\zeta\theta\}, & if \quad \alpha = 1 \end{cases} \tag{7.0-20}$$

and,

$$\text{sign}\theta = \begin{cases} 1, & if \quad \theta > 0 \\ 0, & if \quad \theta = 0 \\ -1, & if \quad \theta < 0 \end{cases} \tag{7.0-21}$$

For $0 < \alpha < 1$ and $\beta = 1$ or $\beta = -1$, the stable density is only for a half line.

In order to estimate the parameters of a stable distribution, I use the ML method given in Rachev and Mittnik (2000). Given N observations, $X = (X_1, X_2, \cdots, X_N)'$ for the positive half line, the log-likelihood function is of the form

$$\ln(\alpha, \lambda; X) = N\ln\lambda + N\ln\alpha + (\alpha - 1)\sum_{i=1}^{N}\ln X_i - \lambda\sum_{i=1}^{N}X_i^\alpha, \tag{7.0-22}$$

which can be maximized using, for example, a Newton-Raphson algorithm. It follows from the first-order condition,

$$\lambda = N\left(\sum_{i=1}^{N}X_i^\alpha\right)^{-1} \tag{7.0-23}$$

that the optimization problem can be reduced to finding the value for α which maximizes the concentrated likelihood

$$\ln^*(\alpha; X) = \ln\alpha + \alpha\nu - \ln\left(\sum_{i=1}^{N}X_i^\alpha\right), \tag{7.0-24}$$

where $\nu = N^{-1}\Sigma_{i=1}^{N}\ln X_i$.

The information matrix evaluated at the maximum likelihood estimates, denoted by $I(\hat{\alpha}, \hat{\lambda})$, is given by

$$I(\hat{\alpha}, \hat{\lambda}) = \begin{pmatrix} N\hat{\alpha}^{-2} & \sum_{i=1}^{N} X_i^{\hat{\alpha}} \ln X_i \\ \sum_{i=1}^{N} X_i^{\hat{\alpha}} \ln X_i & N\hat{\lambda}^{-2} \end{pmatrix}.$$

It can be shown that under fairly mild conditions, the maximum likelihood estimates $\hat{\alpha}$ and $\hat{\lambda}$ are consistent and have asymptotically a multivariate normal distribution with mean $(\alpha, \lambda)'$(see Rachev and Mittnik (2000)).[4]

7.5 Simulating the co-movement of international equity markets

7.5.1 Simulation of the marginal distribution

Paxson (1997) gives a method to generate the fractional Gaussian noise by using the Discrete Fourier Transform of the spectral density. Bardet et al. (2003) give a concrete simulation procedure based on this method with respect to alleviating some of the problems faced in practice. The procedure is:

1. Choose an even integer M. Define the vector of the Fourier frequencies $\Omega = (\theta_1, ..., \theta_{M/2})$, where $\theta_t = 2\pi t/M$ and compute the vector $F = f_H(\theta_1), ..., f_H(\theta_{M/2})$, where

   $$f_H(\theta) = \frac{1}{\pi}\sin(\pi H)\Gamma(2H+1)(1-\cos\theta)\sum_{t\in\aleph}|2\pi t + \theta|^{-2H-1}$$

   $f_H(\theta)$ is the spectral density of FGN.

2. Generate $M/2$ i.i.d exponential $Exp(1)$ random variables $E_1, ..., E_{M/2}$ and $M/2$ i.i.d uniform $U[0,1]$ random variables $U_1, ..., U_{M/2}$.

3. Compute $Z_t = \exp(2i\pi U_t)\sqrt{F_t E_t}$, for $t = 1, ..., M/2$.

4. Form the M-vector: $\tilde{Z} = (0, Z_1, ...Z_{(M/2)-1}, Z_{M/2}, \overline{Z}_{(M/2)-1}, ..., \overline{Z}_1)$.

5. Compute the inverse fast Fourier transform of the complex Z to obtain the simulated sample path.

[4] Other methods for estimating the parameters of a stable distribution (i.e., the method of moments based on the characteristic function, the regression-type method, and the fast Fourier transform method) are discussed in Stoyanov and Racheva-Iotova (2004).

Stoev and Taqqu (2004) generate the approximation of fractional stable noise. They introduce parameters $n, N \in \aleph$, and let the fractional stable noise $Y(t)$ be expressed as

$$Y_{n,N}(t) := \sum_{j=1}^{nN} \Big((\frac{j}{n})_+^{H-1/\alpha} - (\frac{j}{n} - 1)_+^{H-1/\alpha} \Big) L_{\alpha,n}(nt - j), \tag{7.0-25}$$

where $L_{\alpha,n}(t) := M_\alpha((j+1)/n) - M_\alpha(j/n)$, $j \in \Re$. The parameter n is the mesh size and the parameter M is the cut-off of the kernel function. Stoev and Taqqu (2004) describe an efficient approximation involving the fast Fourier transform algorithm for $Y_{n,N}(t)$. Consider the moving average process $Z(m)$, $m \in \aleph$,

$$Z(m) := \sum_{j=1}^{nM} g_{H,n}(j) L_\alpha(m - j), \tag{7.0-26}$$

where

$$g_{H,n}(j) := \Big((\frac{j}{n})_+^{H-1/\alpha} - (\frac{j}{n} - 1)_+^{H-1/\alpha} \Big) n^{-1/\alpha}, \tag{7.0-27}$$

and where $L_\alpha(j)$ is the series of i.i.d standard stable Paretian random variables. Since $L_{\alpha,n}(j) \stackrel{d}{=} n^{-1/\alpha} L_\alpha(j), j \in \Re$, equations (7.0-25) and (7.0-26) imply $Y_{n,N}(t) \stackrel{d}{=} Z(nt)$, for $t = 1, ..., T$. Then, the computing of $Y_{n,N}(t)$ is transferred to focus on the moving average series $Z(m)$, $m = 1, ..., nT$. Let $\tilde{L}_\alpha(j)$ be the $n(N+T)$-periodic with $\tilde{L}_\alpha(j) := L_\alpha(j)$, for $j = 1, ..., n(N+T)$ and let $\tilde{g}_{H,n}(j) := g_{H,n}(j)$, for $j = 1, ..., nN$; $\tilde{g}_{H,n}(j) := 0$, for $j = nN+1, ..., n(N+T)$. Then

$$\{Z(m)\}_{m=1}^{nT} \stackrel{d}{=} \Big\{ \sum_{j=1}^{n(N+T)} \tilde{g}_{H,n}(j) \tilde{L}_\alpha(n - j) \Big\}_{m=1}^{nT}, \tag{7.0-28}$$

because for all $m = 1, ..., nT$, the summation in equation (7.0-26) involves only $L_\alpha(j)$ with indices j in the range $-nN \leq j \leq nT - 1$. Using a circular convolution of the two $n(N+T)$-periodic series $\tilde{g}_{H,n}$ and $\tilde{L}_\alpha$ computed by using the Stoev-Taqqu discrete Fourier transform, the variables $Z(n)$, $m = 1, ..., nT$ (i.e., the fractional stable noise), can be generated.

7.5.2 Simulation of the multi-dimensional copulas

Embrechts et al. (2003) suggest a simulation method for the n-dimension Gaussian copula and Student's t copula. For the Gaussian copula, the algorithm is:

1. Find the Cholesky decomposition A of the correlation matrix ρ.

2. Simulate n independent random variates $y_1, \ldots, y_n$ from $\mathcal{N}(0,1)$.

3. Set $\mathbf{z} = A\mathbf{y}$.

4. Set $u_i = \Phi(z_i)$ for $i = 1, \ldots, n$.

5. $(u_1, \ldots, u_n)^T \sim C_\rho^N$.

For the Student's t copula, the algorithm is:

1. Find the Cholesky decomposition A of the correlation matrix ρ.

2. Simulate n independent random variates $y_1, \ldots, y_n$ from $\mathcal{N}(0, 1)$.

3. Simulate a random variate α from χ_ν^2 independent of $y_1, \ldots, y_n$.

4. Set $\mathbf{z} = A\mathbf{y}$.

5. Set $\mathbf{x} = \frac{\sqrt{\nu}}{\sqrt{\alpha}}\mathbf{y}$.

6. Set $u_i = t_\nu(x_i)$ for $i = 1, \ldots, n$.

7. $(u_1, \ldots, u_n)^T \sim C_{\nu,\rho}^t$.

These algorithms have been adopted in Section 6 for the empirical research in this chapter.

7.6 Empirical results

Table 7.1 shows the descriptive statistics for the nine international stock indexes in my study. All returns for the indexes used in this study are calculated as

$$y_{i,t} = 100 \times \log\Big(\frac{P_{i,t}}{P_{i,t-1}}\Big).$$

From the statistics reported in this table, it can be seen that excess kurtosis exists. Figure 7.1 shows the movement of the nine stock indexes. From this figure, the co-movement can be observed.

For the return of each stock index in my study, I denote N as the sample length, sub-sample series that have been randomly selected by a moving window with length T $(1 \leq T \leq N)$. Replacement is allowed in the sampling. In the empirical analysis, sub-sample length (i.e., the window length) of $T = 1$ month was chosen. A total of 1,800 (200 sub-samples for each stock index) sub-samples were randomly created.

Engle (1982) proposes a Lagrange-multiplier test for the ARCH phenomenon. Table 7.2 shows the test statistics and the critical values to reject the null hypothesis that there is no

ARCH effect at different lag levels. It is clear from the results reported in the table that an ARCH effect is exhibited in these return series.

I use the Ljung-Box-Pierce Q-statistic based on the autocorrelation function to test for serial correlation (i.e., the memory effect). Table 7.3 shows that the null hypothesis that there is no serial correlation can be rejected at different lags. The table shows that the memory effect occurs for each index return series. In order to see when the memory effect vanishes, I compare the Q-statistic with its corresponding critical value. When the quotient of the Q-statistic and the corresponding critical value are less than 1, I cannot reject the null hypothesis that there is no serial correlation. From Table 7.3, I find that the quotient of the Q-statistic to its corresponding critical value exceeds unity. I can therefore reject the null hypothesis that there is no serial correlation and can say that long-range dependence is exhibited by my dataset.

Table 7.4 reports the parameters estimated from the ARMA(1,1)-GARCH(1,1) assuming that residuals are identically and independently normally distributed with zero mean and unit variance. Based on equation (7.0-9), I generate the empirical residuals. The descriptive statistic of the empirical residuals $\tilde{u}_t$ is shown in Table 7.5. The results reported in the table make it clear that excess kurtosis still exists and the residuals do not follow i.i.d. $N(0,1)$ distribution.

Table 7.6 shows the parameters estimated for empirical residuals $\tilde{u}_t$ based on the methods introduced in Section 7.4. As I mentioned in that section, the Hurst index for non-Gaussian stable processes has different bounds for "persistence" and "anti-persistence". For tail index $\alpha \in (0, 2)$, when $H \in (0, 1/\alpha)$, the processes exhibit "anti-persistence", and when $H \in (1/\alpha, 1)$, the processes exhibit "persistence". There is no long-range dependence when $\alpha \in (0, 1]$ because the Hurst index is bounded in the interval $(0, 1)$. When $H = 1/\alpha$, depending on the value of α, the processes exhibit either no memory or long-range dependence. From Tables 7.1 and 7.6, I find that the Hurst index has no value that is equal to $1/\alpha$. Therefore, I find that long-range dependence occurs in my dataset.

The AD and KS statistics were calculated for the six candidate distributional assumptions. Table 7.7 reports the descriptive statistics of the computed AD, KS, and CVM statistics. As can be seen in the table, ARMA-GARCH with a fractional Gaussian noise model exhibits a smaller mean value for the AD, KS, and CVM statistics than the other five models. Figure 7.2 shows the boxplot of AD statistics for the six alternative ARMA-GARCH models investigated. Figure 7.3 shows the boxplot of KS statistics and Figure 7.4 shows the boxplot of CVM statistics. ARMA-GARCH with fractional Gaussian noise model exhibits smaller mean and less outliers, demonstrating the advantage of this model.

I applied the exceedance correlation test of Hong et al. (2006). The test statistic J_ρ and corresponding p value for rejecting the null hypothesis of symmetric correlation are illustrated in Table 8 and Table 7.9. For the test result reported in Table 7.8, I chose q (i.e., the exceedance level) at 0.8 quantile while Table 7.9 reports the test result at exceedance level of 0.95 quantile. At the exceedance level of 0.7 quantile, two pairs of correlation are not symmetric (i.e., DAX/STOXX with a p value of 0.0458 and FCHI/HSI with a p value of 0.0323) if I set the confidence level at $\alpha = 0.05$. However, at the same confidence level, the tests at exceedance level of 0.95 quantile do not reject the null hypothesis of symmetric correlation for all pairs tested. This test implies that in my sample, tail dependence exists in a symmetric way for the nine indexes from international equity markets. Therefore, the copulas based on the symmetric correlation assumption are still valid.

As can be seen from Table 1, the index returns clearly do not follow the Gaussian distribution. Stable parameters in this table also exhibit the non-Gaussian characteristic since the stable parameter is equal to 2 for the Gaussian case. The heavy tailedness can be easily observed in the data. Accordingly it seems that the application of heavy-tailed distributions should perform better than the Gaussian distribution or fractional Gaussian noise. The empirical result found by simulating the marginal distribution indicates that the fractional Gaussian noise subordinated in the ARMA-GARCH model fits better than other alternatives. I believe that there are two reasons for this. First, the heavy tailendness of index returns stems from the heavy tailedness of each index component stock. The equity market indexes aggregate those stocks that exhibit heavy tailedness. The aggregation of heavy-tailed distributions is asymptotically self-similar, and the fractional Gaussian noise is a typical stochastic process with self-similarity. The heavy-tailedness effect is considered in the self-similarity. Second, after the aggregation, although equity market indexes exhibit heavy tailedness, the influence of such effect (non-Gaussian and heavy tailedness) in the movement of the market index is weak compared to long-range dependence and volatility clustering.

From the goodness of fit testing for marginal distributions, I find the best fit model is ARMA-GARCH with a fractional Gaussian noise model. Then taking ARMA-GARCH with a fractional Gaussian noise model as marginal distributions for each stock index return, I simulate multivariate Gaussian copula (in this empirical study, 9 dimensions for the data) and Student's t copula for the dependence structure of these nine index returns. Table 7.8 shows the descriptive statistics of the computed AD, KS, and CVM statistics for 200 sub-sample matrices with nine marginal distributions as the column vectors. In this table, the Student's t copula exhibits smaller mean values for the computed AD, KS, and CVM statistics than the Gaussian copula, indicating the better performance. My results agree with those of Marshal and Zeevi (2002) who investigated the extreme co-movements in equity markets and found that the Gaussian copula is not sufficient to model equity return dependence while the student t copula performs better.

7.7 Conclusion

There is considerable interest in the co-movement of international equity markets. The linear correlation measure is not satisfactory to discover the dependence structure between equity markets. With several advantages, copulas are regarded as the ideal measure to model both the degree and structure of dependence. Some works are based on the bivariate co-movement. In this chapter, I use the copula ARMA-GARCH model to capture the multivariate co-movement among the international equity markets in this study.

In this empirical analysis, I investigate a ARMA-GARCH model with six forms for the residuals (fractional stable noise, fractional Gaussian noise, stable distribution, white noise, generalized Pareto distribution, and generalized extreme value distribution) for modeling the marginal distribution for the nine international equity market indexes. By using parameters estimated from the empirical series, I simulate a series for each index returns with these six different modeling structures. Then I compare the goodness of fit for these generated series to the empirical series by adopting three criteria for the goodness of fit test: the Kolmogorov-Smirnov distance, the Anderson-Darling distance, and the Cramer von Mises distance. Based on a comparison of these criteria, the empirical evidence shows that the ARMA-GARCH model with fractional Gaussian noise demonstrates better performance in modeling marginal distributions.

Using an ARMA-GARCH model with fractional Gaussian noise, I simultaneously simulate the nine index returns with both the Gaussian copula and Student's t copula. By using the same criteria of goodness of fit test in comparing marginal distributions, I find that the Student's t copula is better than the Gaussian copula when modeling the multivariate co-movement of these nine equity markets. The reason is that Student's t copula can capture the tail dependence among these index returns for both positive and negative extreme values, while the Gaussian copula cannot.

The findings reported in this chapter should be taken into account in modeling the co-movement of global equity markets for several reasons. First, using multi-dimensional copulas rather than bivariate copulas can reveal the simultaneous co-movement of several markets. Second, when modeling the marginal distribution of each market's returns, my model can capture long-range dependence, heavy tails, and volatility clustering simultaneously. Third, using high-frequency data, the impact of both macroeconomic factors and microstructure effects on each market can be considered. The model reveals that similar factors impact the co-movement of international markets and that investors' behaviors in each market are similar, especially their reactions in each market to world news are similar. With this model, more accurate prediction is possible for the simultaneous co-movement of several equity markets.

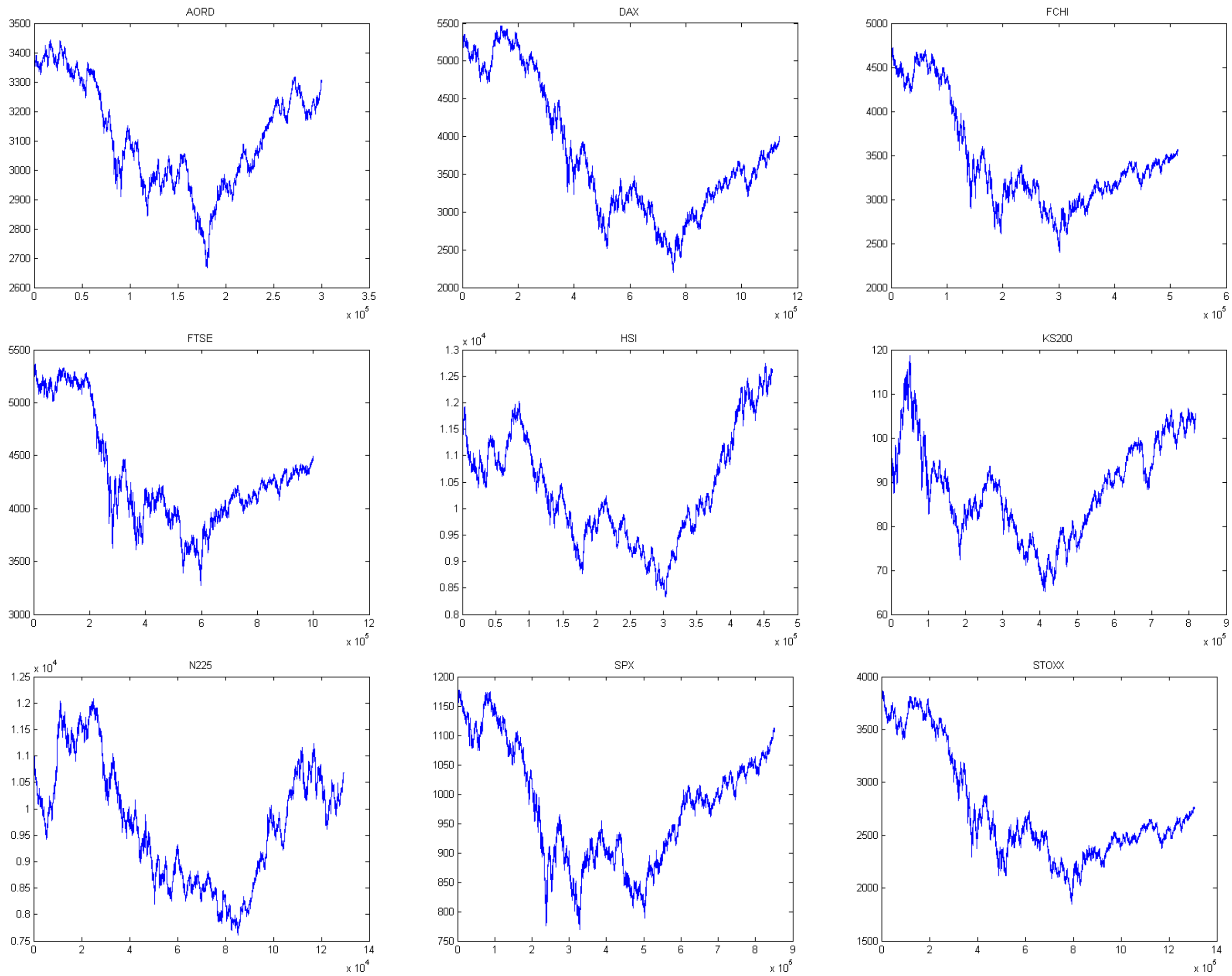

Figure 7.1: Plot of Index Movements

Table 7.1: Summary of the statistical characteristics of nine index returns.

	location	mean	std	kurtosis	skewness	mininum	maximum	Hurst	α
AORD	Austrilia	-5.38E-06	0.0213	1903.7000	-0.8007	-2.3168	2.3168	0.5182	1.3864
DAX	Germany	-2.31E-05	0.0371	448.7800	-0.4372	-4.4347	2.9733	0.4978	1.2222
FCHI	France	-5.08E-05	0.0506	732.5400	0.7020	-3.9898	3.9050	0.5065	1.3253
FTSE	UK	-1.53E-05	0.0222	846.8900	0.3852	-2.2313	2.7460	0.5110	1.3270
HSI	China	2.19E-05	0.0486	235.6000	1.3178	-2.3099	2.6161	0.5410	1.4970
KS200	South Korea	2.25E-05	0.0439	1404.1000	4.9465	-3.7141	4.3014	0.5106	1.2961
N225	Japan	3.31E-06	0.0759	121.5700	1.7980	-2.0686	3.4282	0.4821	1.2411
SPX	US	-3.76E-06	0.0169	4478.7000	-14.7220	-3.6071	2.3110	0.5313	1.1819
STOXX	Switzerland	-2.45E-05	0.0247	619.6600	-2.3390	-3.4999	2.7500	0.5010	1.1720

Table 7.2: Result of the ARCH-test for different lags at $\alpha = 0.05$.

	lag1	*lag2*	*lag5*	*lag10*	*lag15*	*lag20*	*lag25*	*lag30*
AORD	37152	37669	39206	39257	39265	39346	39356	39355
DAX	3910	4103	4565	6678	6745	6898	6914	6932
FCHI	10	12	19	23	29	60	69	75
FTSE	12854	12913	14395	14465	14512	15008	15472	15483
HSI	21	28	37	41	44	47	48	49
KS200	8	21	38	39	39	39	39	39
N225	66	73	633	642	644	646	647	647
SPX	7	11	16	17	17	17	17	18
STOXX	2342	3763	4831	4961	5005	5035	5060	5087
Critical Value	3.8415	5.9915	11.0700	18.3070	24.9960	31.4100	37.6520	43.7730

Table 7.3: The Ljung-Box-Pierce Q-test statistic for different lags at $\alpha = 0.05$.

	$10-min$	$30-min$	$1hour$	$2hours$	$4hours$	$1day$	$1week$	$1month$
AORD	3202	3276	3375	3516	4245	6493	16943	51385
DAX	43937	44127	44375	44698	45360	47259	60129	105330
FCHI	8220	8461	8737	90051	97211	12711	26178	70851
FTSE	34008	35637	35944	36465	37205	38854	53398	100850
HSI	10490	10851	11162	11619	12324	13623	22593	55679
KS200	1547	1687	1807	2019	2785	3890	15770	52254
N225	2748	2836	2931	3166	3769	4945	13352	43790
SPX	25670	257030	258100	258550	259790	262420	283190	349850
STOXX	167900	168360	168770	169160	1701405	172290	189380	246550
Critical Value	55.7585	146.5673	277.1376	532.0754	1033.1928	2023.0522	9829.0489	38856.9694

Table 7.4: Estimated parameters of the AMAR(1,1)-GARCH(1,1) model with residuals following normal distribution with zero mean and unit variance. Numbers in parentheses are the standard errors. These parameters are used in the empirical simulation.

	α_0	α_1	β_1	κ	γ_1	θ_1
AORD	3.9724E-07	-0.1952	0.1136	4.6260E-009	0.6486	0.3438
	(9.5130E-08)	*(1.1634E-11)*	*(1.1618E-11)*	*(3.3215E-12)*	*(1.2023E-12)*	*(1.6527E-11)*
DAX	-1.9424E-07	0.5559	-0.3766	1.3826E-008	0.6558	0.3442
	(3.5419E-08)	*(1.1621E-12)*	*(1.3573E-12)*	*(1.7137E-12)*	*(3.4729E-12)*	*(2.2729E-12)*
FCHI	-9.5146E-08	0.5869	-0.4720	2.7586E-08	0.8135	0.1227
	(1.4999E-07)	*(1.0717E-06)*	*(1.2618E-6)*	*(8.4707E-12)*	*(2.7037E-05)*	*(2.1210E-05)*
FTSE	-1.2392E-07	0.8232	-0.7568	6.5616E-09	0.5987	0.2812
	(3.3959E-08)	*(3.0773E-13)*	*(2.3373E-13)*	*(3.7753E-12)*	*(1.2205E-12)*	*(1.2504E-13)*
HSI	-9.8445E-10	0.5154	-0.6893	2.8409E-08	0.6931	0.2610
	(5.5655E-08)	*(2.5658E-04)*	*(3.3054E-04)*	*(4.7913E-11)*	*(3.3037E-04)*	*(3.2846E-04)*
KS200	4.3005E-06	0.0075	-0.2692	1.9238E-08	0.6582	0.3418
	(2.4262E-08)	*(8.6562E-06)*	*(9.6637E-06)*	*(3.0034E-12)*	*(2.5391E-05)*	*(2.1966E-05)*
N225	-4.8124E-06	0.4782	-0.2905	6.6660E-08	0.6170	0.3766
	(4.5034E-07)	*(8.2554E-04)*	*(7.5777E-04)*	*(2.5256E-10)*	*(8.2544E-04)*	*(8.0406E-04)*
SPX	2.5689E-07	0.5548	-0.0262	2.4089E-09	0.8386	0.0619
	(3.1323E-08)	*(3.1166E-12)*	*(2.4869E-15)*	*(4.4352E-14)*	*(2.7531E-11)*	*(5.2184E-11)*
STOXX	1.3438E-07	0.6101	-0.3995	5.8155E-09	0.6677	0.2631
	(7.3838E-08)	*(2.9345E-12)*	*(1.3513E-12)*	*(1.3865E-13)*	*(2.2687E-11)*	*(2.4868E-11)*

Table 7.5: Summary of the empirical $\tilde{u}_t$.

	mean	*variance*	*kurtosis*	*skewness*
AORD	-0.0027	0.9085	326.1807	2.5378
DAX	0.0017	0.9384	1034.6394	-2.2134
FCHI	-3.6688E-05	0.9196	1032.1787	2.0182
FTSE	0.0038	0.9337	84.3404	-0.5075
HSI	0.0023	0.9522	147.8326	0.7201
KS200	-0.0170	1.0982	1379.5626	6.9291
N225	0.0103	1.0463	74.6847	0.6791
SPX	-0.0026	1.0008	14390.3118	-40.7157
STOXX	-0.0011	0.9487	6051.9163	-6.9826

Table 7.6: Parameters estimated from the empirical $\tilde{u}_t$.

	$Hurst_{FGN}$	$Hurst_{fsn}$	α	β	γ	ζ
AORD	0.5366	0.5791	1.8777	-0.2610	0.4765	0.0063
DAX	0.5619	0.5492	1.3976	-0.0103	0.3765	-0.0026
FCHI	0.5476	0.5107	1.4688	-0.0153	0.3508	-0.0039
FTSE	0.5544	0.5112	1.3538	-0.0126	0.4032	0.0014
HSI	0.4588	0.6376	1.0914	-0.0016	0.2550	-9.8442E-04
KS200	0.5168	0.5594	1.4387	0.0147	0.3869	-0.0195
N225	0.5554	0.5260	1.3326	-0.0234	0.3629	0.0021
SPX	0.5278	0.5149	1.3533	-0.0113	0.3535	-0.0073
STOXX	0.5787	0.5604	1.3847	-5.3461E-04	0.3852	-0.0019

Table 7.7: Summary of the AD, KS and CVM statistics for alternative models for marginal distribution. Mean, median, standard deviation ("std"), maximum value ("max"), minimum value ("min") and range of the AD, KS and CVM statistics are presented in this table. "FGN" stands for fractional Gaussian noise, "fsn" for fractional stable noise, "normal" for white noise, "stable" for stable distribution, "gev" for generalized extreme value distribution, and "gpd" for generalized Pareto distribution.

	AD_{mean}	AD_{median}	AD_{std}	AD_{max}	AD_{min}	AD_{range}
$ARMA-GARCH_{FGN}$	54.8381	54.8220	0.3222	56.2561	53.5083	2.7482
$ARMA-GARCH_{fsn}$	54.8459	54.7861	0.7213	63.4861	51.7917	11.6952
$ARMA-GARCH_{normal}$	55.0686	54.9050	0.8139	66.1060	53.9640	12.1425
$ARMA-GARCH_{stable}$	55.7993	55.1920	2.2074	77.7650	52.8670	24.9051
$ARMA-GARCH_{gev}$	55.4836	55.2240	1.2978	73.6110	53.4447	20.1677
$ARMA-GARCH_{gpd}$	76.7643	70.2010	18.1530	109.5412	45.5832	63.9523
	KS_{mean}	KS_{median}	KS_{std}	KS_{max}	KS_{min}	KS_{range}
$ARMA-GARCH_{FGN}$	0.5017	0.5012	0.0025	0.5136	0.4966	0.0170
$ARMA-GARCH_{fsn}$	0.5032	0.5016	0.0058	0.5855	0.4945	0.0910
$ARMA-GARCH_{normal}$	0.5039	0.5020	0.0075	0.6035	0.4965	0.1070
$ARMA-GARCH_{stable}$	0.5116	0.5053	0.0202	0.7103	0.4948	0.2155
$ARMA-GARCH_{gev}$	0.5079	0.5052	0.0121	0.6721	0.4967	0.1754
$ARMA-GARCH_{gpd}$	0.7059	0.6476	0.1646	1.0000	0.4309	0.5691
	CVM_{mean}	CVM_{median}	CVM_{std}	CVM_{max}	CVM_{min}	CVM_{range}
$ARMA-GARCH_{FGN}$	502.4011	500.1612	5.4108	535.6220	498.5100	37.1120
$ARMA-GARCH_{fsn}$	502.4176	500.2272	5.6007	545.4330	497.8300	47.6120
$ARMA-GARCH_{normal}$	502.9730	500.3900	6.3452	568.2810	498.6210	69.6620
$ARMA-GARCH_{stable}$	505.6970	500.9900	16.8300	773.8670	497.6620	276.1900
$ARMA-GARCH_{gev}$	503.8650	500.8452	10.4230	684.7800	498.2801	186.4900
$ARMA-GARCH_{gpd}$	910.7413	627.1045	551.4111	1999.7110	388.6400	1611.1001

Table 7.8: J_ρ statistic of testing exceedence correlation at quantile=0.8. p values of rejecting the null hypothesis of symmetric correlation are reported in parentheses.

	DAX	FCHI	FTSE	HSI	KS200	N225	SPX	STOXX
AORD	1.0058	1.2757	2.3229	0.0894	0.7097	0.0610	0.5043	0.4164
	(0.3158)	(0.2586)	(0.1274)	(0.7649)	(0.3995)	(0.8048)	(0.4775)	(0.5187)
DAX		0.001	2.9011	2.161	0.7393	0.8153	0.0376	3.9866
		(0.9743)	(0.0885)	(0.1415)	(0.3898)	(0.3665)	(0.8461)	(0.0458)
FCHI			0.343	4.5819	3.1019	1.6173	0.1942	0.9737
			(0.5580)	(0.0323)	(0.0781)	(0.2034)	(0.6594)	(0.3237)
FTSE				0.1115	2.2976	0.2142	0.0644	0.1000
				(0.7384)	(0.1295)	(0.6434)	(0.7995)	(0.7517)
HSI					0.6317	0.6457	0.495	0.1771
					(0.4267)	(0.4216)	(0.4816)	(0.6738)
KS200						0.5464	0.0819	0.3178
						(0.4597)	(0.7746)	(0.5728)
N225							8.13E-05	0.242
							(0.9928)	(0.6226)
SPX								0.1221
								(0.7266)

Table 7.9: J_ρ statistic of testing exceedence correlation at quantile=0.95. p values of rejecting the null hypothesis of symmetric correlation are reported in parentheses.

	DAX	FCHI	FTSE	HSI	KS200	N225	SPX	STOXX
AORD	2.5350	1.2649	0.4065	0.9127	1.4148	0.3300	0.0256	0.7522
	(0.1113)	(0.2607)	(0.5237)	(0.3393)	(0.2342)	(0.5656)	(0.8726)	(0.3857)
DAX		0.5586	0.3197	0.3560	0.7390	0.0086	0.1135	1.5363
		(0.4548)	(0.5717)	(0.5506)	(0.3899)	(0.9258)	(0.7360)	(0.2151)
FCHI			1.8125	0.7246	0.0236	1.3454	0.6177	3.3126
			(0.1782)	(0.3946)	(0.8778)	(0.2460)	(0.4318)	(0.0680)
FTSE				0.9908	1.2734	1.1343	0.919	0.7235
				(0.3195)	(0.2591)	(0.2868)	(0.3377)	(0.3949)
HSI					1.2354	2.2382	0.4741	0.3233
					(0.2663)	(0.1346)	(0.4910)	(0.5695)
KS200						3.4007	1.6541	0.0108
						(0.0651)	(0.1983)	(0.9170)
N225							0.5622	0.0927
							(0.4533)	(0.7606)
SPX								0.5598
								(0.4543)

Table 7.10: Summary of the AD, KS and CVM statistics for alternative models for joint distribution. Mean, median, standard deviation ("std"), maximum value ("max"), minimum value ("min") and range of the AD, KS and CVM statistics are presented in this table.

	AD_{mean}	AD_{median}	AD_{std}	AD_{max}	AD_{min}	AD_{range}
Gaussian copula	0.9241	0.9374	0.0338	0.9718	0.8370	0.1348
Student's t copula	0.9237	0.9362	0.0340	0.9716	0.8382	0.1334
	KS_{mean}	KS_{median}	KS_{std}	KS_{max}	KS_{min}	KS_{range}
Gaussian copula	48.4519	55.5841	16.3230	67.9456	9.6306	58.3150
Student's t copula	48.4470	55.5060	16.3190	67.9740	9.9158	58.0580
	CVM_{mean}	CVM_{median}	CVM_{std}	CVM_{max}	CVM_{min}	CVM_{range}
Gaussian copula	785.6190	798.7101	24.5134	817.5083	729.5811	87.9272
Student's t copula	785.2964	798.1155	24.6323	817.9235	728.6673	89.2562

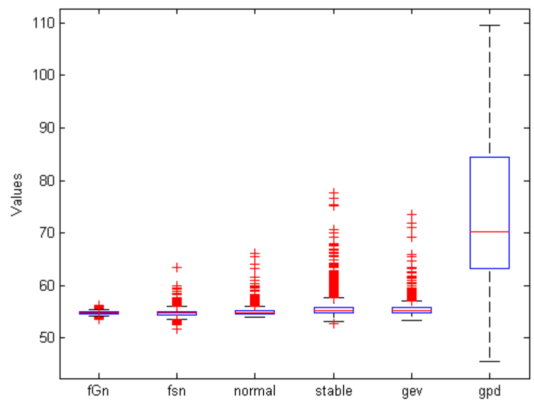

Figure 7.2: Boxplot of AD statistics of modeling marginal distribution with alternative residual distributions.

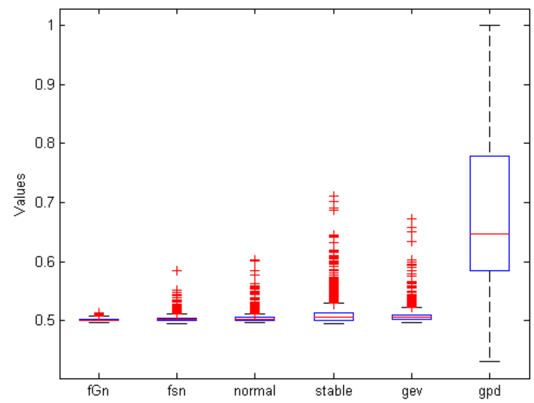

Figure 7.3: Boxplot of KS statistics of modeling marginal distribution with alternative residual distributions.

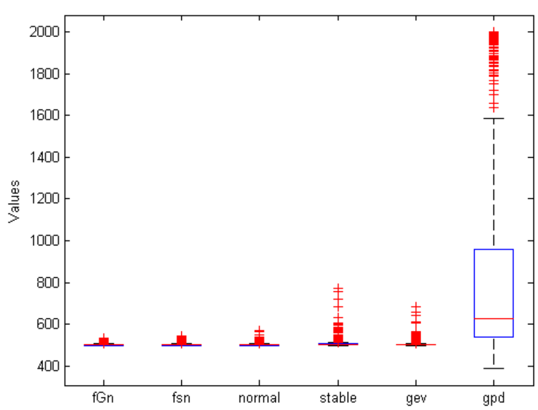

Figure 7.4: Boxplot of CVM statistics of modeling marginal distribution with alternative residual distributions.

Chapter 8

Modeling Multivariate High-Frequency Time Series II

8.1 Introduction

Correlation is a measure of dependence between random variables that is commonly used in finance, particularly for measuring risk diversification potential in portfolio management. Sun et al. (2007b) points out that when analyzing an equity market, two dependence structures are encountered: the correlation within a single asset and the correlation between several assets. However the term "correlation" is very often incorrectly used to mean any notion of dependence. Actually correlation is one particular measure of dependence among many. In a world of spherical and elliptical distributions, it is the accepted measure. Because financial theories and risk management analysis rely crucially on the dependence structure of assets, the limitations of correlation as a measure of the dependence among random variables requires that we seek a more reliable measure of dependence. Copulas are alternative measures that can overcome the limitations of correlation.

When dealing with the dependence structure between several assets, the usual linear correlation is often applied. But the usual linear correlation is not a satisfactory measure of the dependence among several assets in the equity market. First, when the variance of returns in those assets turns out to be infinite, that is, extreme events are frequently observed, the linear correlation between these assets is undefined. Second, linear correlation assumes that both marginal and joint distributions of returns in these assets are elliptical. In real-world markets, this assumption is unwarranted. Third, the linear correlation is not invariant under nonlinear strictly increasing transformations, implying that the return might be uncorrelated whereas the prices are correlated or vice versa. Fourth, linear correlation only measures the degree of dependence but does not clearly discover the structure of dependence. It has been widely observed that the price of several assets drops at about the same time period even when the correlation among those assets is fairly low. In real-world markets this is due to the problem of liquidity in markets and the difficulty of financing positions in the repo (repurchase agreements) market. The structure of dependence also influences the diversification benefit gained based on a linear correlation measure. Embrechts et al. (2003) and Rachev et al. (2005) illustrate the drawbacks of using linear correlation to analyze dependency. A more prevalent approach which overcomes the disadvantages of linear correlation is to model dependency by using copulas. With the copula method, the nature of dependence that can be modeled is more general and the dependence of extreme events can be considered.

Two commonly used unconditional copulas are the unconditional Gaussian copula and unconditional Student's t copula[1]. Empirical studies have failed to support the assumption that return data follow a Gaussian distribution[2]. Instead, return data often exhibit excess kurtosis and heavy tails. Nor is the multivariate Gaussian distribution warranted either when studying multi-dimensional return data because of tail dependence among them. The Student's t copula and Clayton copulas can characterize tail dependence in multi-dimensional return data but the Gaussian copula cannot. In Clayton copulas, only the Joe-Clayton copula can capture both the lower and upper tail dependence and allows for asymmetric tail dependence. Unfortunately, due to computational difficulties, the Joe-Clayton copula cannot be implemented for multi-dimensional return data; however, the Student's t copula can be implemented to capture tail dependence in multi-dimensional return data (see, Embrechts et al. (2003), Patton (2005), and Nelsen (2006)). Although the Student's t copula can capture the tail dependence in multi-dimensional return data, it is only good for modeling symmetric correlations or tail dependence. Several studies point out that stocks tend to have greater correlations in a bearish market than in a bullish market (see, for example, Ang and Chen 2002, Login and Solnik 2001, and Hong et al. 2006). The Joe-Clayton copula is a traditional tool to cope with asymmetric dependence structure but it possesses the major deficiency mentioned above. In order to model the asymmetric tail dependence, I employ the skewed Student's t copula (also see Demarta and McNeil 2005) in this study.

When dealing with the dependence (i.e., long-range dependence) of a single asset, we should take other stylized factors into account such as volatility clustering and distributional heavy tails. It is necessary to treat long-range dependence, volatility clustering, and heavy tailedness simultaneously in order to obtain more accurate predictions of market volatility. Rachev and

[1]Specification of these two copulas is given in the Appendix.

[2]For a summary of the empirical evidence, see Rachev et al. (2005).

Mittnik (2000) note that for modeling financial data, not only does model structure play an important role, but distribution assumptions influence modeling accuracy. Historically, the normal distribution has been used despite the mounting evidence that the returns of financial assets do not follow the normal law. Empirically asset returns have been found to be skewed and characterized by kurtosis exceeding that of the normal distribution. Consequently, a flexible distribution is needed to deal with such stylized facts. Moreover, in order to model the dynamic behavior of financial returns through time, not only is a more flexible static distribution needed, but also more flexible stochastic processes. Such processes must have independent and stationary increments based on a more general distribution than the normal distribution. To ensure independent and stationary increments, the required distribution has to be infinitely divisible (see Rachev and Mittnik 2000 and Schoutens 2003).

Lévy distributions/processes provide an ideal solution for replacing the normal distribution with sophisticated infinitely divisibility. For example, the Lévy stable distribution can be used to capture characteristics of financial data since it is rich enough to encompass the stylized facts noted earlier. Other researchers have shown the advantages of Lévy stable distributions in financial modeling (see, Fama (1963), Mittnik and Rachev (1993), Rachev (2003), and Rachev et al. (2005)). Several other examples are reported in the literature, such as the Lévy process with Variance Gamma distributed increments proposed by Madan and Seneta (1990) and the NIG Lévy process proposed by Barndorff-Nielsen (1998). Sun et al. (2007a) report the Lévy fractional stable process has advantages in modeling ultra-high-frequency data because this process has self-similarity which describes the property that the correlation structure of a process is preserved irrespective of time scaling (also see Taqqu and Samorodnitsky 1994, Rachev and Mittnik 2000, Rachev and Samorodnitsky 2001, Doukhan et al. 2003, and Willinger et al. (1998)). Therefore, it is natural to employ specific Lévy processes in the study of the within-asset dependence together with capturing volatility clustering and heavy-tailedness.

Based on a copula-ARMA-GARCH modeling structure for six indexes in the German equity market, I compare several candidate specifications using simulation methods. In the modeling structure, the marginal distribution captures the long-range dependence, heavy tails, and volatility clustering simultaneously in order to obtain more accurate predictions (for example, utilizing Lévy processes in the marginals), and these marginal distributions are connected by a specified copula. The empirical results indicate that the skewed Student's t copula and ARMA-GARCH model with Lévy fractional stable noise dominate the alternative models tested in this study.

I organized this chapter as follows. A brief introduction of skewed Student's t copula is provided in Section 8.2. In Section 8.3, I specify three Lévy family models investigated in my study (i.e., Lévy stable distribution, fractional Gaussian noise, and Lévy fractional stable noise) utilized in modeling the marginal distribution for each index. The study's data and empirical methodology are described in Section 8.4 and the empirical results based on high-frequency data at 1-minute level for six German equity market indexes are reported. In that section, I report the comparison of the goodness of fit for both the marginal distribution and joint distribution. I summarize my conclusions in Section 8.5.

8.2 Skewed Student's *t* Copula

Copulas enable the dependence structure to be extracted from both the joint distribution function and the marginal distribution functions. From a mathematical viewpoint, a copula function C is a probability distribution function on the n-dimensional hypercube. Sklar (1959) has shown that any multivariate probability distribution function F_Y of some random vector $Y = (Y_1, ..., Y_n)$ can be represented with the help of a copula function C of the following form:

$$\begin{aligned} F_Y(y_1, ..., y_n) &= P(Y_1 \leq y_1, \ldots, Y_n \leq y_n) \\ &= C(P(Y_1 \leq y_1), \ldots, P(Y_n \leq y_n)) \\ &= C(F_{Y_1}(y_1), \ldots, F_{Y_n}(y_n)) \end{aligned}$$

where F_{Y_i}, $i = 1, \ldots, n$ denote the marginal distribution functions of the random variables, Y_i, $i = 1, \ldots, n$.

When the variables are continuous, the density c associated with the copula is given by:

$$c(F_{Y_1}(y_1), \ldots, F_{Y_n}(y_n)) = \frac{\partial^n C(F_{Y_1}(y_1), \ldots, F_{Y_n}(y_n))}{\partial F_{Y_1}(y_1), \ldots, \partial F_{Y_n}(y_n)}.$$

The density function f_Y corresponding to the n-variate distribution function F_Y is

$$f_Y(y_1, ..., y_n) = c(F_{Y_1}(y_1), \ldots, F_{Y_n}(y_n)) \prod_{i=n}^{n} f_{Y_i}(y_i),$$

where f_{Y_i}, $i = 1, \ldots, n$ is the density function of F_{Y_i}, $i = 1, \ldots, n$ (see, Joe (1997), Cherubini et al. (2004), and Nelsen (2006)).

8.2.1 Multivariate skewed Student's *t* distribution

The form of the multivariate skewed Student's t distribution I am using is defined through the following stochastic representation[3]:

[3]For more information, see Section 12.7 in Rachev and Mittnik (2000) and Demarta and McNeil (2005).

$$X := \mu + \gamma W + Z\sqrt{W} \tag{8.0-1}$$

where $W \in IG(\nu/2, \nu/2)$ and $Z \in N(0, \Sigma)$, Z is independent of W, $\gamma = (\gamma_1, \ldots, \gamma_n)$ is a n-dimensional vector accounting for the skewness, $\mu = (\mu_1, \ldots, \mu_n)$ is a n-dimensional location parameter vector, and ν is the degrees of freedom. I denote this distribution by $X \in t_n(\nu, \mu, \Sigma, \gamma)$. The notation $IG(\nu/2, \nu/2)$ stands for the inverse gamma distribution with parameters $\nu/2$. Thus, W is a one-dimensional random variable and Z is a random vector having a zero-mean multivariate normal distribution with covariance matrix

$$\Sigma = \begin{pmatrix} \sigma_{11} & \sigma_{12} & \ldots & \sigma_{1n} \\ \sigma_{21} & \sigma_{22} & \ldots & \sigma_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_{n1} & \sigma_{n2} & \ldots & \sigma_{nn} \end{pmatrix}.$$

The multivariate skewed Student's t distribution allows for the closed-form expression of its density.

$$f_X(x) = \frac{aK_{(\nu+n)/2}\left(\sqrt{(\nu + (x-\mu)'\Sigma^{-1}(x-\mu))\gamma'\Sigma^{-1}\gamma}\right)\exp((x-\mu)'\Sigma^{-1}\gamma)}{(\nu + (x-\mu)'\Sigma^{-1}(x-\mu))\gamma'\Sigma^{-1}\gamma)^{-\frac{\nu+n}{4}}\left(1 + \frac{(x-\mu)'\Sigma^{-1}(x-\mu)}{\nu}\right)^{\frac{\nu+n}{n}}}$$

where $x \in R^n$, K is the modified Bessel function of the third kind, and

$$a = \frac{2^{\frac{2-\nu-n}{2}}}{\Gamma(\nu/2)(\pi\nu)^{n/2}\sqrt{|\Sigma|}}.$$

The skewed Student's t copula is defined as the copula of the multivariate distribution of X. Therefore, the copula function is

$$C(u_1, \ldots, u_n) = F_X(F_1^{-1}(u_1), \ldots, F_n^{-1}(u_n))$$

where F_X is the multivariate distribution function of X and $F_k^{-1}(u_k)$, $k = 1, n$ is the inverse c.d.f of the k-th marginal of X. That is, $F_X(x)$ has the density $f_X(x)$ defined above and the density function $f_k(x)$ of each marginal is

$$f_k(x) = \frac{aK_{(\nu+1)/2}\left(\sqrt{\left(\nu + \frac{(x-\mu_k)^2}{\sigma_{kk}}\right)\frac{\gamma_k^2}{\sigma_{kk}}}\right)\exp\left((x-\mu_k)\frac{\gamma_k}{\sigma_{kk}}\right)}{\left(\left(\nu + \frac{(x-\mu_k)^2}{\sigma_{kk}}\right)\frac{\gamma_k^2}{\sigma_{kk}}\right)^{-\frac{\nu+1}{4}}\left(1 + \frac{(x-\mu_k)^2}{\nu\sigma_{kk}}\right)^{\nu+1}}, \quad x \in R \tag{8.0-2}$$

where σ_{kk} is the k-th diagonal element in the matrix Σ.

In the sequel, I assume that there are n risk variables with historical data $X_{T\times n} = (X_1, \ldots, X_n)$. The algorithm for estimating the parameters and generating the scenarios are given in the next section. I describe not only the generation of simulations from the fitted copula but also how it is combined with one-dimensional assumptions for the marginals.

8.2.2 Simulation Algorithm

The simulation algorithm of the multivariate skewed Student's t distribution involves the following steps:

1. I first use the maximum likelihood estimation (MLE) method to obtain the estimated parameters $(\nu, \hat{\mu}_i, \hat{\sigma}_i, \hat{\gamma}_i)$ for $i = 1, n$. No flexibility is lost by fixing the degrees of freedom; the symmetric case appears when $\hat{\gamma} = 0$. Estimate the matrix Σ of the multivariate skewed Student's t distribution by the following formula

$$\hat{\Sigma} = \left(\mathrm{cov}(X) - \frac{2\nu^2}{(\nu-2)^2(\nu-4)}\hat{\gamma}\hat{\gamma}'\right)\frac{\nu-2}{2}$$

   where ν is the degree of freedom and $\hat{\gamma} = (\hat{\gamma}_1, \ldots, \hat{\gamma}_n)$.

2. Draw N independent n-dimensional vectors from the multivariate skewed Student's t distribution using the stochastic representation (1) with the already fitted parameters $(\nu, \hat{\mu}, \hat{\Sigma}, \hat{\gamma})$. The result from that step is a N-by-n matrix $S = \{S_{ij}\}$ with simulations.

   (a) Draw N independent n-dimensional vectors from the multivariate normal distribution $N(0, \hat{\Sigma})$.

   (b) Draw N independent random numbers from the inverse gamma distribution with parameters $IG(\nu/2, \nu/2)$.

   (c) Obtain final simulations by formula (8.0-1) using the estimated parameter values.

3. Transform simulations S to uniform simulations U using the sample distribution function of the marginals. Denote by $\hat{F}_k(x)$ the sample c.d.f. of the k-th marginal,

$$\hat{F}_k(x) = \frac{1}{N}\sum_{j=1}^{N} I\{S_{jk} \le x\}$$

   where $I\{A\}$ stands for the indicator function of the set A. Then

$$U_{jk} = \hat{F}_k(S_{jk}), \quad j = 1, N \quad k = 1, n \tag{8.0-3}$$

Ideally, instead of using the sample c.d.f., one could use the c.d.f.

$$F_k(x) = \int_{-\infty}^{x} f_X(t)dt$$

where the density $f_X(t)$ is given in (8.0-2). While this approach is more accurate, it is not analytically tractable.

In order to combine the copula with different one-dimensional marginals, two more steps are needed:

1. Fit marginal distribution parameters (depends on corresponding one-dimensional model) for all risk variables. In effect, I have a set of parameters for each variable.
2. Once I have obtained the uniform scenarios $U = \{U_{jk}\}$ in (8.0-3), I transform them by the inverse of the fitted one-dimensional distribution functions,

$$R_{jk} = G_k^{-1}(U_{jk}), \quad j = 1, N \quad k = 1, n$$

If no closed-form expressions are available, one can resort to the inverse of the sample c.d.f.

8.3 Lévy Processes with Specifications

Lévy processes have become increasingly popular in mathematical finance because they can describe the observed reality of financial markets in a more accurate way. They capture jumps, heavy-tails, and skewness observed from "real" asset price processes. Meanwhile, Lévy processes provide us the appropriate option pricing framework to model implied volatilities across strike and across maturities with respect to the "risk-neutral" assumption. In this section, I introduce the definition of Lévy processes as well as one specific form (the Lévy fractional stable motion) and two extensions of infinitely divisible distributions (the Lévy stable distribution and fractional Brownian motion) that I apply in this study.

8.3.1 Lévy processses

Suppose $\phi(u)$ is the characteristic function of a distribution. If for every positive interger n, $\phi(u)$ is also the n-th power of a characteristic function, this distribution is said to be infinitely divisible. A stochastic process $X = (X_t)_{t\geq 0}$ can be defined for every such an infinitely divisible distribution. For this stochastic process $X = (X_t)_{t\geq 0}$ on $(\Omega, \mathcal{F}, \mathrm{P})$ to be called a Lévy process, the following five conditions (see Sato (1999)) have to be satisfied:

1. X_0=0 almost surely.
2. X has independent increments: given $0 < t_1 < t_2 < \cdots < t_n$, the random variables $X_{t_1}, X_{t_2} - X_{t_1}, \cdots, X_{t_n} - X_{t_{n-1}}$ are independent.
3. X has stationary increment: for $t \geq 0$, the distribution of $X_{t+s} - X_s$ does not depend on $s \geq 0$.
4. X is stochastically continuous: $\forall\, t \geq 0$ and $\varepsilon > 0$, $\lim_{s \to t} P[(X_s - X_t) > \varepsilon] = 0$.
5. X is right continuous and has left limits (càdlàg).

The cumulative characteristic function $\psi(u) = \log \phi(u)$ must satisfy the Lévy-Khintchine formula given as follows:

$$\psi(u) = i\gamma u - \frac{\sigma^2}{2}u^2 + \int_{-\infty}^{+\infty} \Big(\exp(iux) - 1 - iux\mathbf{1}_{\{|x|<1\}}\Big) v(dx)$$

where $\gamma \in R, \sigma^2 \geq 0$ and v is a measure on $R\backslash\{0\}$ with

$$\int_{-\infty}^{+\infty} \inf\{1, x^2\} v(dx) = \int_{-\infty}^{+\infty} (1 \wedge x^2) v(dx) < \infty.$$

As to this case, the infinitely divisible distribution has a Lévy triplet $[\gamma, \sigma^2, v(dx)]$ and v is called the Lévy measure of X (see Sato 1999 for more general reference).

8.3.2 Lévy Stable Distribution

The Lévy stable distribution (sometimes referred to as α-stable distribution) has four parameters for complete description: an index of stability $\alpha \in (0, 2]$ (also called the tail index), a skewness parameter $\beta \in [-1, 1]$, a scale parameter $\gamma > 0$, and a location parameter $\zeta \in \Re$. There is unfortunately no closed-form expression for the density function and distribution function of a Lévy stable distribution. Rachev and Mittnik (2000) give the definition for the Lévy stable distribution: A random variable X is said to have a Lévy stable distribution if there are parameters $0 < \alpha \leq 2$, $-1 \leq \beta \leq 1$, $\gamma \geq 0$ and ζ real such that its characteristic function has the following form:

$$E\exp(i\theta X) = \begin{cases} \exp\{-\gamma^\alpha |\theta|^\alpha (1 - i\beta(\text{sign}\theta)\tan\frac{\pi\alpha}{2}) + i\zeta\theta\}, & if \quad \alpha \neq 1 \\ \exp\{-\gamma|\theta|(1 + i\beta\frac{2}{\pi}(\text{sign}\theta)\ln|\theta|) + i\zeta\theta\}, & if \quad \alpha = 1 \end{cases} \tag{8.0-4}$$

and,

$$\text{sign}\theta = \begin{cases} 1, & if \quad \theta > 0 \\ 0, & if \quad \theta = 0 \\ -1, & if \quad \theta < 0 \end{cases} \tag{8.0-5}$$

For $0 < \alpha < 1$ and $\beta = 1$ or $\beta = -1$, the stable density is only for a half line.

8.3.3 Fractional Brownian Motion

For a given $H \in (0,1)$, there is basically a single Gaussian *H-sssi* [4] process, namely fractional Brownian motion (fBm), first introduced by Kolmogorov (1940). Mandelbrot and Wallis (1968) and Taqqu (2003) define fBm as a Gaussian *H-sssi* process $\{B_H(t)\}_{t\in R}$ with $0 < H < 1$. Mandelbrot and van Ness (1968) define the stochastic representation

$$B_H(t) := \frac{1}{\Gamma(H+\frac{1}{2})}\Big(\int_{-\infty}^{0}[(t-s)^{H-\frac{1}{2}} - (-s)^{H-\frac{1}{2}}]dB(s) + \int_0^t (t-s)^{H-\frac{1}{2}}dB(s)\Big), \quad (8.0\text{-}6)$$

where $\Gamma(\cdot)$ represents the Gamma function:

$$\Gamma(a) := \int_0^\infty x^{a-1}e^{-x}dx,$$

and $0 < H < 1$ is the Hurst parameter. The integrator B is ordinary Brownian motion. The principal difference between fractional Brownian motion and ordinary Brownian motion is that the increments in Brownian motion are independent while in fractional Brownian motion they are dependent. For fractional Brownian motion, Samorodnitsky and Taqqu (1994) define its increments $\{Y_j, j \in Z\}$ as fractional Gaussian noise (fGn), which is, for $j = 0, \pm 1, \pm 2, ...$, $Y_j = B_H(j-1) - B_H(j)$.

8.3.4 Lévy Stable Motion

While fractional Brownian motion can capture the effect of long-range dependence, it has less power to capture heavy tailedness. The existence of abrupt discontinuities in financial data, combined with the empirical observation of sample excess kurtosis and unstable variance, confirms the stable Paretian hypothesis identified by Mandelbrot (1963, 1983). It is natural to introduce the stable Paretian distribution in self-similar processes in order to capture both long-range dependence and heavy tailedness. Samorodinitsky and Taqqu (1994) discuss the α-stable *H-sssi* processes $\{X(t), t \in R\}$ with $0 < \alpha < 2$. If $0 < \alpha < 1$, the exponent of self-similarity is $H \in (0, 1/\alpha]$ and if $1 < \alpha < 2$, the exponent of self-similarity is $H \in (0,1)$. In addition, Cohen and Samorodnitsky (2006) show that with exponent $H' = 1 + H(1/\alpha - 1)$, the process $\{X(t), t \in R\}$ is a well-defined symmetric α-stable ($S\alpha S$) process. It has stationary

[4] The abbreviation of "*sssi*" means self-similar stationary increments, if the exponent of self-similarity H is to be emphasized, then "*H-sssi*" is adopted. Lamperti (1962) first introduced the semi-stable processes (which we today refer to as self-similar processes). Let T be either R, $R_+ = \{t : t \geq 0\}$ or $\{t : t > 0\}$. The real-valued process $\{X(t), t \in T\}$ has stationary increments if $X(t+a) - X(a)$ has the same finite-dimensional distrituions for all $a \geq 0$ and $t \geq 0$. Then the real-valued process $\{X(t), t \in T\}$ is self-similar with exponent of self-similarity H for any $a > 0$, and $d \geq 1$, $t_1, t_2, ..., t_d \in T$, satisfying: $(X(at_1), X(at_2), ..., X(at_d)) \stackrel{d}{=} (a^H X(t_1), a^H X(t_2), ...a^H X(t_d))$.

increments and is self-similar. They show that (1) for $0 < \alpha < 1$, a family of H'-*sssi* $S\alpha S$ processes with $H' \in (1, 1/\alpha)$ is obtained, (2) for $1 < \alpha < 2$, a family of H'-*sssi* $S\alpha S$ processes with $H' \in (1/\alpha, 1)$ is obtained, and (3) for $\alpha = 1$, a family of *1-sssi* $S\alpha S$ processes is obtained.

There are many extensions of fractional Brownian motion to the Lévy stable distribution. The most commonly used is linear fractional Lévy motion (also called linear fractional stable motion), $\{L_{\alpha,H}(a,b;t), t \in (-\infty,\infty)\}$, which Samorodinitsky and Taqqu (1994) define as

$$L_{\alpha,H}(a,b;t) := \int_{-\infty}^{\infty} f_{\alpha,H}(a,b;t,x)M(dx), \tag{8.0-7}$$

where

$$f_{\alpha,H}(a,b;t,x) := a\Big((t-x)_+^{H-\frac{1}{\alpha}} - (-x)_+^{H-\frac{1}{\alpha}}\Big) + b\Big((t-x)_-^{H-\frac{1}{\alpha}} - (-x)_-^{H-\frac{1}{\alpha}}\Big), \tag{8.0-8}$$

where $x_+ = \max(x, 0), x_- = \min(x, 0)$ and a, b are real constants. $|a| + |b| > 1$, $0 < \alpha < 2$, $0 < H < 1$, $H \neq 1/\alpha$, and M is an α-stable random measure on R with Lebesgue control measure and skewness intensity $\beta(x)$, $x \in (-\infty, \infty)$ satisfying: $\beta(\cdot) = 0$ if $\alpha = 1$. They define linear fractional stable noises expressed by $Y(t)$, and $Y(t) = X_t - X_{t-1}$,

$$\begin{aligned} Y(t) &= L_{\alpha,H}(a,b;t) - L_{\alpha,H}(a,b;t-1) \\ &= \int_R \Big(a\Big[(t-x)_+^{H-\frac{1}{\alpha}} - (t-1-x)_+^{H-\frac{1}{\alpha}}\Big] \\ &+ b\Big[(t-x)_-^{H-\frac{1}{\alpha}} - (t-1-x)_-^{H-\frac{1}{\alpha}}\Big]\Big)M(dx), \end{aligned} \tag{8.0-9}$$

where $L_{\alpha,H}(a,b;t)$ is a linear fractional stable motion defined by equation (8.0-8), and M is a stable random measure with Lebesgue control measure given $0 < \alpha < 2$. Samorodinitsky and Taqqu (1994) show that the kernel $f_{\alpha,H}(a,b;t,x)$ is d-self-similar with $d = H - 1/\alpha$ when $L_{\alpha,H}(a,b;t)$ is $1/\alpha$-self-similar. This implies $H = d + 1/\alpha$ (see Taqqu and Teverovsky (1998) and Weron et al. (2005)).[5] In this chapter, if there is no special indication, the fractional stable noise (fsn) is generated from a linear fractional Lévy motion.

[5] Some properties of these processes have been discussed in Mandelbrot and Van Ness (1968), Maejima (1983), Maejima and Rachev (1987), Manfields et al. (2001), Rachev and Mittnik (2000), Rachev and Samorodnitsky (2001), Racheva and Samorodnitsky (2003), Samorodnitsky (1994, 1996, 1998), Samorodinitsky and Taqqu (1994), and Cohen and Samorodnitsky (2006).

8.4 Data and empirical methodology

8.4.1 Data

In my study, I consider six German equity market indexes: DAX, HDAX, MDAX, Midcaps, SDAX, and TecDAX). DAX is the most commonly cited stock index for companies listed on the Frankfurt Stock Exchange. The composition of the DAX is the 30 largest and most liquid stocks traded on that exchange. The HDAX includes 110 companies and represents a broader index covering all sectors and the shares of the largest companies listed in Prime Standard[6]. The MDAX includes the 50 companies from classic sectors that rank immediately below the companies included in the DAX index. The company's size is based on the order book volume and market capitalization. The SDAX is the small caps index for 50 smaller companies in Germany, which in terms of order book volume and market capitalization rank directly below the MDAX shares. Midcaps is the short name of Midcap Market Index which consists of 80 German companies trading on the Frankfurt Stock Exchange. The TecDAX stock index tracks the performance of the 30 largest German companies from the technology sector. The companies rank in terms of order book turnover and market capitalization below those included in the DAX. [7]

In previous studies of the co-movement in equity markets, low-frequency data are usually examined. Because stock indexes change their composition quite often over time, it is difficult to find the impact of these changes in composition when analyzing the return history of stock indexes using low-frequency data. Dacorogna et al. (2001) calls this phenomenon the "breakdown of the permanence hypothesis". In order to overcome this problem, I use high-frequency data in my study.

The high-frequency data of the six indexes in the German equity markets listed in Table 8.1 from January 2 to September 30, 2006 were aggregated to the 1-minute frequency level[8]. The aggregation algorithm is based on the linear interpolation introduced by Wasserfallen and Zimmermann (1995). That is, given an inhomogeneous series with times t_i and values $\varphi_i = \varphi(t_i)$, the index i identifies the irregularly spaced sequence. The target homogeneous time

[6]Prime Standard is a market segment of the German Stock Exchange that lists German companies which comply with international transparency standards including quarterly reporting in German and English, application of international accounting standards (IFRS/IAS or US-GAAP), publication of a financial calendar, staging of at least one analyst conference a year, and ad-hoc disclosure also in English. All companies listed in the DAX, MDAX, TecDAX, and SDAX have to belong to Prime Standard.

[7]More detailed information is available from www.deutsche-boerse.com.

[8]Starting in 2006, DAX indexes are calculated at every second. In my original dataset, DAX, MDAX, and TecDAX are sampled at one-second level.

series is given at times $t_0 + j\Delta t$ with fixed time interval Δt starting at t_0. The index j identifies the regularly spaced sequence. The time $t_0 + j\Delta t$ is bounded by two times t_i of the irregularly spaced series, $I = \max(i \,|t_i \leq t_0 + j\Delta t)$ and $t_I \leq t_0 + j\Delta t > t_{I+1}$. Data are interpolated between t_I and t_{I+1}. The linear interpolation shows that

$$\varphi(t_0 + j\Delta t) = \varphi_I + \frac{t_0 + j\Delta t - t_I}{t_{I+1} - t_I}(\varphi_{I+1} - \varphi_I). \tag{8.0-10}$$

Dacorogna et al. (2001) pointed out that linear interpolation relies on the future of time and Müller et al. (1990) suggests that linear interpolation is an appropriate method for stochastic processes with independent and identically distributed (i.i.d.) increments. Empirical evidence has shown the seasonality in high-frequency data. In order to remove such disturbance, several methods of data adjusting have been adopted in modeling. Engle and Russell (1998) and other researchers adopt several methods to adjust the seasonal effect in data. In my study, seasonality is treated as one type of self-similarity. Consequently, it is not necessary to adjust for the seasonal effect in the data.

8.4.2 Empirical methodology

To investigate the co-movement of different indexes in German equity markets, I propose the copula ARMA-GARCH model. This model is implemented with an ARMA-GARCH model for the marginal distributions and one copula for the joint distribution. Six GARCH models with different kinds of residuals (i.e., residuals with forms of white noise, fractional Gaussian noise, Lévy fractional stable noise, Lévy stable distribution, generalized Pareto distribution, and generalized extreme value distribution) for the marginal distributions are estimated. After goodness of fit testing, I use the best goodness of fit model for the marginal distributions with multi-dimesional Gaussian copula, Student's t copula, and skewed Student's t copula for the joint distribution to simulate the returns on the equity indexes. Then, the models will be tested with several goodness of fit test methods for a large dataset.

I define the ARMA-GARCH model for the conditional mean equation as:

$$y_t = \alpha_0 + \sum_{i=1}^{r} \alpha_i \, y_{t-i} + \varepsilon_t + \sum_{j=1}^{m} \beta_j \varepsilon_{t-j}.$$

Let $\varepsilon_t = \sigma_t \, u_t$, where the conditional variance of the innovations, σ_t^2, is by definition

$$Var_{t-1}(y_t) = E_{t-1}(\varepsilon_t^2) = \sigma_t^2.$$

The general GARCH(p,q) processes for the conditional variance of the innovation is then

$$\sigma_t^2 = \kappa + \sum_{i=1}^{p} \gamma_i \, \sigma_{t-i}^2 + \sum_{j=1}^{q} \theta_j \, \varepsilon_{t-j}^2.$$

Since $\varepsilon_t = \sigma_t u_t$, u_t could be calculated from ε_t/σ_t. Defining

$$\tilde{u_t} = \frac{\varepsilon_t^s}{\hat{\sigma_t}}, \tag{8.0-11}$$

where ε_t^s is estimated from the sample and $\hat{\sigma_t}$ is the estimation of σ_t. In my study, ARMA(1,1)-GARCH(1,1) are parameterized as marginal distributions with different kinds of u_t (i.e., normal distribution, fractional Gaussian noise, fractional stable noise, Lévy stable distribution, generalized Pareto distribution, and generalized extreme value distribution).

I will let N ($N = 98,999$) denote the length of the sample. The sub-sample series used for the in-sample analysis are randomly selected by a moving window with length T ($1 \leq T \leq N$). Replacement is allowed in the sampling. Letting T_F denote the length of the forecasting series, I perform one-week ahead out-of-sample forecasting ($1 \leq T \leq T + T_F \leq N$). In the empirical analysis, sub-sample length (i.e., the window length) of $T = 10,000$ (approximately one month) was chosen for the in-sample simulation and $T_F = 2,250$ (approximately one week) for the out-of-sample forecasting. A total of 300 sub-samples (50 sub-samples for each stock index) were randomly created.

From the goodness of fit testing for marginal distributions, I find the best fit model. Then taking the best fit model as marginal distributions for each stock index return, I simulate a multivariate Gaussian copula, Student's t copula, and skewed Student's t copula for the dependence structure of the nine stock index returns. The simulation method adopted was explained in Section 8.2.

The Kolmogorov-Smirnov distance (KS), the Anderson-Darling distance (AD), the Cramer Von Mises distance (CVM), and the Kuiper distance (K) are used as the criterion for the goodness of fit testing. Specifically, these criterion are defined as follows:

$$KS = \sup_{x \in \Re} |F_s(x) - \tilde{F}(x)|,$$

$$AD = \sup_{x \in \Re} \frac{|F_s(x) - \tilde{F}(x)|}{\sqrt{\tilde{F}(x)(1 - \tilde{F}(x))}},$$

$$CVM = \int_{-\infty}^{\infty} \Big(F_s(x) - \tilde{F}(x)\Big)^2 d\tilde{F}(x),$$

and

$$K = \sup_{x \in \Re} \Big(F_s(x) - \tilde{F}(x)\Big) + \sup_{x \in \Re} \Big(\tilde{F}(x) - F_s(x)\Big).$$

where $F_s(x)$ denotes the empirical sample distribution and $\tilde{F}(x)$ is the estimated distribution function.

8.4.3 Empirical results

Table 8.1 reports the descriptive statistics for the six German equity market indexes in my study. All returns for the indexes used in this study are calculated as

$$y_{i,t} = 100 \times \log\Big(\frac{P_{i,t}}{P_{i,t-1}}\Big).$$

From the statistics reported in this table, it can be seen that excess kurtosis exists. Figure 8.1 shows the movement of the six stock indexes at its own level. From this figure, the co-movement of indexes at different levels can be observed. Figure 8.2 shows the return plots for each index; from the figure I can observe that although each index has a different magnitude of return, the extreme values occur almost at same time.

Engle (1982) proposes a Lagrange-multiplier test for the ARCH phenomenon. A test statistic for ARCH of lag order q is given by

$$X_q \equiv nR_q^2,$$

where R_q^2 is the non-centered goodness-of-fit coefficient of a qth-order autoregression of the squared residuals taken from the original regression

$$\hat{u}_t^2 = \omega_0 + \omega_1 \hat{u}_{t-1}^2 + \omega_2 \hat{u}_{t-2}^2 + \cdots + \omega_q \hat{u}_{t-q}^2 + e_t,$$

The $\hat{u}$ in the above equation is the residual in the original regression equation. Under the null hypothesis of the residuals of the original model being normally i.i.d., the ARCH statistic of lag order q follows a χ^2 distribution with q degree of freedom:

$$\lim_{n\to\infty} X_q \sim \chi_q^2.$$

Table 8.2 shows the test statistics and the critical values to reject the null hypothesis that there is no ARCH effect at different lag levels for index returns. It is clear from the results reported in this table that an ARCH effect is exhibited in these return series.

I use the Ljung-Box-Pierce Q-statistic based on the autocorrelation function to test for serial correlation (i.e., the memory effect). The Q-statistic is given as follows:

$$Q :\sim \chi_m^2 = N(N+2) \sum_{k=1}^{m} \frac{\rho_k^2}{N-k},$$

where N denotes the sample size, m the number of autocorrelation lags included in the statistic, and ρ_k the sample autocorrelation at lag order k which is

$$\rho_k = \frac{\sum_{t=1}^{N-k} y_t y_{t+k}}{\sum_{t=1}^{N} y_t^2}.$$

Ljung and Box (1978) show that the Q-statistic follows an asymptotic χ^2 distribution.

The results reported in Table 8.3 show that the null hypothesis that there is no serial correlation can be rejected at different lags for index returns. The table shows that the memory effect occurs for each index return series. In order to see when the memory effect vanishes, I compare the Q-statistic with its corresponding critical value. When the quotient of the Q-statistic and the corresponding critical value are less than 1, I cannot reject the null hypothesis that there is no serial correlation. From Table 8.3, I find that the quotient of the Q-statistic to its corresponding critical value exceeds unity. I can therefore reject the null hypothesis that there is no serial correlation and can say that long-range dependence is exhibited by my dataset.

Table 8.1 also shows the parameters estimated for index returns based on the methods introduced in Section 8.4. The Hurst index $H \in (0, 1)$ is the index of self-similarity. For Gaussian processes with stationary increments, when

1. $H \in (0, 0.5)$, the increments of a process tend to have opposite signs and thus are more zigzagging due to the negative covariance.

2. $H \in (0.5, 1)$, the covariance between these two increments is positive and less zigzagging of the process.

3. $H = 0.5$, the covariance between this two increments is zero.

This can be restated as following: If the Hurst index is

1. less than 0.5, the process displays “anti-persistence” (i.e., positive excess return is more likely to be reversed and the performance in the next period is likely to be below the average, or in the contrary, negative excess return is more likely to be reversed and the performance in the next period is likely to be above the average).

2. greater than 0.5, the process displays “persistence” (i.e., positive excess return or negative excess return is more likely to be continued and the performance in the next period is likely to be the same as that in the current period).

3. equal to 0.5, the process displays no memory (i.e., the performance in the next period has equal probability to be below and above the performance in the current period).

For Lévy fractional stable process, if the process has the index α ($0 < \alpha < 2$) when $H = 1/\alpha$ which corresponds to a process with independent increments, then I say this process has no memory. When $H > 1/\alpha$, the process displays long-range dependence and when $H < 1/\alpha$, the process displays negative dependence. In addition, long-range dependence is only possible when $\alpha > 1$, since $H \in (0, 1)$ (see, Samorodnitsky and Taqqu (1994)).

In order to check long-range dependence in stock returns, I use the methods introduced by Sun et al. (2007a) to estimate the Hurst index under the Gaussian and stable assumptions. I employed the MLE method explained by Rachev and Mittnik (2000) to estimate the stable parameter. The results, reported in Table 1, indicate that the Hurst index does not have an estimated value of 0.5 if fractional Gaussian noise is assumed. This suggests the occurence of either long memory or short memory under the Gaussian assumption [9]. In Table 8.1, I can observe both fluctuation and long memory under the non-Gaussian stable assumption.

Table 8.4 reports the parameters estimated from the ARMA(1,1)-GARCH(1,1) assuming that residuals are identically and independently normally distributed with zero mean and unit variance. Based on equation (11), I generate the empirical residuals. The descriptive statistic of the empirical residuals $\tilde{u}_t$ is shown in Table 8.5. The results reported in the table make it clear that excess kurtosis still exists and the residuals do not follow i.i.d. $N(0,1)$ distribution.

Table 8.5 reports the descriptive statistics for $\tilde{u}_t$. I could expect that $\tilde{u}_t$ follows an i.i.d. standard normal distribution. Unfortunately, in my study, $\tilde{u}_t$ is not i.i.d. standard normal residuals. They have excess kurtosis, which means heavy tails still exist. Meantime, I could also observe long-range dependence in $\tilde{u}_t$ by simply checking the value of Hurst index and α. I use the Ljung-Box-Pierce Q-test again for testing long-range dependence. Table 8.6 reports the result. This table shows that the memory effect occurs for each $\tilde{u}_t$ series (although for DAX and HDAX, the memory effect vanishes after a 1 hour lag). The results in Table 8.6 confirm the information offered by the value of the Hurst index and α in Table 8.5. From Table 8.7, I can see that there is no ARCH effect in $\tilde{u}_t$. From Tables 8.5 and 8.6, I can confirm the assumption of long-range dependence and heavy tailedness in $\tilde{u}_t$. Therefore using fractional processes (i.e., fractional stable noise and fractional Gaussian noise) and heavy-tailed distribution (i.e., Lévy stable distribution, generalized Pareto distribution, and generalized extreme value distribution) in the simulation of $\tilde{u}_t$ is reasonable.

[9] There are various extensions of the self-similarity property for generalized random processes, see Dobrushin (1979).

The AD, KS, CVM, and K statistics were calculated for the six candidate distributional assumptions in ARMA(1,1)-GARCH(1,1) model. Table 8 reports the mean value of each goodness of fit statistic. As can be seen in the table, ARMA-GARCH with a fractional stable noise model exhibits a smaller mean value both in the in-sample simulation and out-of-sample forecasting for all goodness of fit statistics I used than the other five models.

I use the Gaussian copula, Student t copula, and skewed Student t copula for the joint distribution of six equity indexes. I employ each of the six candidate models once as marginal distribution with the Gaussian copula or t copula for in-sample simulation and out-of-sample forecasting. The mean value of each goodness of fit statistics for each marginal and joint distribution are reported. Table 8 reports the results of the Gaussian copula with these candidate models, Table 8.9 reports the results of the Student t copula as a joint distribution, and Table 8.10 reports the results of the skewed Student t copula as a joint distribution.

The smaller the value of the four criterion distances, the better the performance of the model. I can find that in both in-sample and out-of-sample studies, the ARMA-GARCH model with Lévy fractional stable noice has minimal values of criterion distances for all three joint distributions. This coincides with the result reported in Sun et al. (2007a). I report the summary statistics of my total sample by groups of each criteria with respect to different models in Table 8.11. I can see that for the criteria values of each model with the Gaussian copula, the in-sample study is smaller than that of the Student's t coupla, which means the Gaussian copula is better than the Student's t copula for in-sample study. But for the out-of-sample forecasting, the Student's t copula is a little better. For the skewed Student's t copula, all of the four criteria values significantly decrece compared with the Gaussian and Student's t copula. This implies that the six German indexes exhibit asymmetric correlations, confirming the intuition that volatility is higher in downside markets than in upside markets observed in Figures 8.1 and 8.2.

In order to confirm this conclusion implied from Table 11, I use Mahalanobis distance (see Mahalanobis (1936)) to check if the skewed Student's t copula is significantly better than other two copulas. The Mahalanobis distance from a group of values with mean $\mu = (\mu_1, \mu_2, \dots, \mu_p)^T$ and covariance matrix Σ for a multivariate vector $x = (x_1, x_2, \dots, x_p)^T$ is defined as as

$$D_M(x) = \sqrt{(x-\mu)^T \, \Sigma^{-1} \, (x-\mu)}.$$

It is based on correlations between variables by which different patterns can be identified and analyzed. It is a useful way of determining similarity/dissimilarity of variables. It differs from Euclidean distance in that it takes into account the correlations of the data set and is scale-invariant (i.e., not dependent on the scale of measurements). When the correlation structure of variables is unclear, the Mahalanobis distance measure is likely to be the most appropriate due to its scale-invariant property. If the covariance matrix is the identity matrix, the Mahalanobis distance reduces to the Euclidean distance.

I report the results in Table 8.12. The Mahalanobis distance tells us that the level of goodness of fit for the Gaussian copula and Student's t copula is lower than that of the skewed Student's t copula in both in-sample and out-of-sample studies. For the in-sample case, the Mahalanobis distance shows that the skewed Student's t copula is better than the Gaussian coupla and the Student's t copula (because of $4.3623 < 4.7630$ and $4.3802 < 4.7630$). Similar results can be found in the out-of-sample forecasting that the skewed Student's t copula dominates other two copulas (because $4.3098 < 4.5399$ and $4.4174 < 4.5399$). The Mahalanobis distance confirms the conclusion drawn directly from Table 8 to Table 12: The six German markets comove at intra-daily level with asymmetric correlations. These results agree with the results reported in previous studies, for example, Ang and Chen (2002) and Login and Solnik (2001).

8.5 Conclusions

There is considerable interest in the co-movement of assets in equity markets. The linear correlation measure is not a satisfactory measure for discovering the dependence structure between equity assets. With several advantages, copulas are regarded as the ideal measure to model both the degree and structure of dependence. While some papers (for example, Hong et al. (2006)) are based on the bivariate co-movement, in this chapter I adopt the copula ARMA-GARCH model using intra-daily data to capture the multi-dimensional co-movement among six indexes in the German equity markets.

In my empirical analysis, I investigate an ARMA-GARCH model with six forms for the residuals (Lévy fractional stable noise, fractional Gaussian noise, Lévy stable distribution, white noise, generalized Pareto distribution, and generalized extreme value distribution) for modeling the marginal distribution for six German equity market indexes. By using parameters estimated from the empirical return series, I simulate a return series for each index with these six different modeling structures. Then I compare the goodness of fit for these generated series to the empirical series by utilizing three criteria for the goodness of fit test: the Kolmogorov-Smirnov distance, the Anderson-Darling distance, the Cramer von Mises distance, and the Kuiper distance. Based on a comparison of these criteria, the empirical evidence shows that the ARMA-GARCH model with fractional stable noise demonstrates better performance in modeling marginal distributions.

Using an ARMA-GARCH model with six candidate residuals, I simultaneously simulate the six index returns with the Gaussian copula, Student's t copula, and skewed Student's t copula. By using the same goodness of fit tests in comparing marginal distributions, I find that the skewed Student's t copula with Lévy fractional stable ARMA-GARCH model is the best one when modeling the multivariate co-movement of the six German equity market indexes investigated. The reason is that the skewed Student's t copula can capture the tail dependence among these index returns for both positive and negative extreme values, especially when the returns are asymmetric. With my model, I find that the comovement of equity markets studied is asymmetric at the intra-daily level. This finding is consistent with those reported in other studies.

The findings reported in this chapter should be taken into account in modeling the co-movement of assets in an equity market for several reasons. First, using multi-dimensional copulas rather than bivariate copulas can reveal the simultaneous co-movement of several assets. Second, when modeling the marginal distribution of each market return, my model can capture long-range dependence, heavy tails, and volatility clustering simultaneously. Third, using high-frequency data, the impact of both macroeconomic factors and microstructure effects on asset return can be considered. The model reveals that similar macroeconomic factors impact the co-movement of different sectors in German equity markets and the feedback from the markets are also similar.

Table 8.1: Summary of the statistical characteristics of six index returns.

	mean	variance	kurtosis	skewness	minimum	maximum	α	H_{FGN}	H_{fsn}
DAX	0.0103	0.0017	129.1500	-1.0195	-1.6531	1.3618	1.2750	0.4999	0.7843
HDAX	0.0105	0.0015	235.5500	-0.5942	-1.7048	1.3866	1.2639	0.5103	0.8014
MDAX	0.0155	0.0010	847.3900	-5.3287	-2.2902	1.8328	1.5078	0.5289	0.6922
Midcap	0.2636	1.0723	838.5400	-4.7981	-81.8860	48.8030	1.6419	0.53158	0.6406
SDAX	0.0145	0.0008	118.4800	0.1410	-1.0037	1.2468	1.4187	0.6270	0.8319
TDAX	0.0100	0.0023	265.6300	-0.5213	-2.2653	2.5694	1.4595	0.5875	0.7727

Table 8.2: Result of the ARCH-test of returns for different lags at $\alpha = 0.05$.

lag	$1min$	$5min$	$10min$	$2hours$	$1day$	$1week$	$2weeks$	$1month$
DAX	273.7	457.6	484.6	491.0	497.9	509.4	521.2	552.0
HDAX	20.2	79.4	94.2	95.8	97.6	102.26	105.8	119.5
MDAX	4.9	128.6	132.9	134.1	134.7	135.8	136.29	137.9
Midcap	20.0	948.8	956.6	957.5	957.8	958.9	959.4	961.1
SDAX	168.5	595.9	680.3	696.6	718.4	734.8	745.0	769.9
TDAX	63.5	630.3	688.4	696.7	706.2	712.4	713.3	727.5
Critical Value	3.8	11.0	18.3	24.9	31.4	37.6	43.7	79.0

Table 8.3: Result of the Ljung-Box-Pierce Q-test of returns for different lags at $\alpha = 0.05$.

lag	$10min$	$30min$	$1hour$	$2hours$	$1day$	$1week$	$2weeks$	$1month$
DAX	221.4	265.2	304.4	405.9	793.1	3228.1	6273.6	12294.0
HDAX	157.0	207.7	245.4	329.7	748.4	3388.3	7277.0	13452.0
MDAX	2247.7	2511.3	2670.2	2774.9	3190.6	6513.9	11190.0	19164.0
Midcap	3466.8	3887.9	4098.9	4200.0	4631.1	7994.1	12982	20160.0
SDAX	205.6	354.2	497.5	653.4	1042.1	3298.5	6370.5	12448.0
TDAX	696.5	837.8	900.7	968.1	1298.6	3722.1	6750.4	12626.0
Critical Value	18.3	43.7	79.0	146.5	532.0	2515.1	4962.3	9829.0

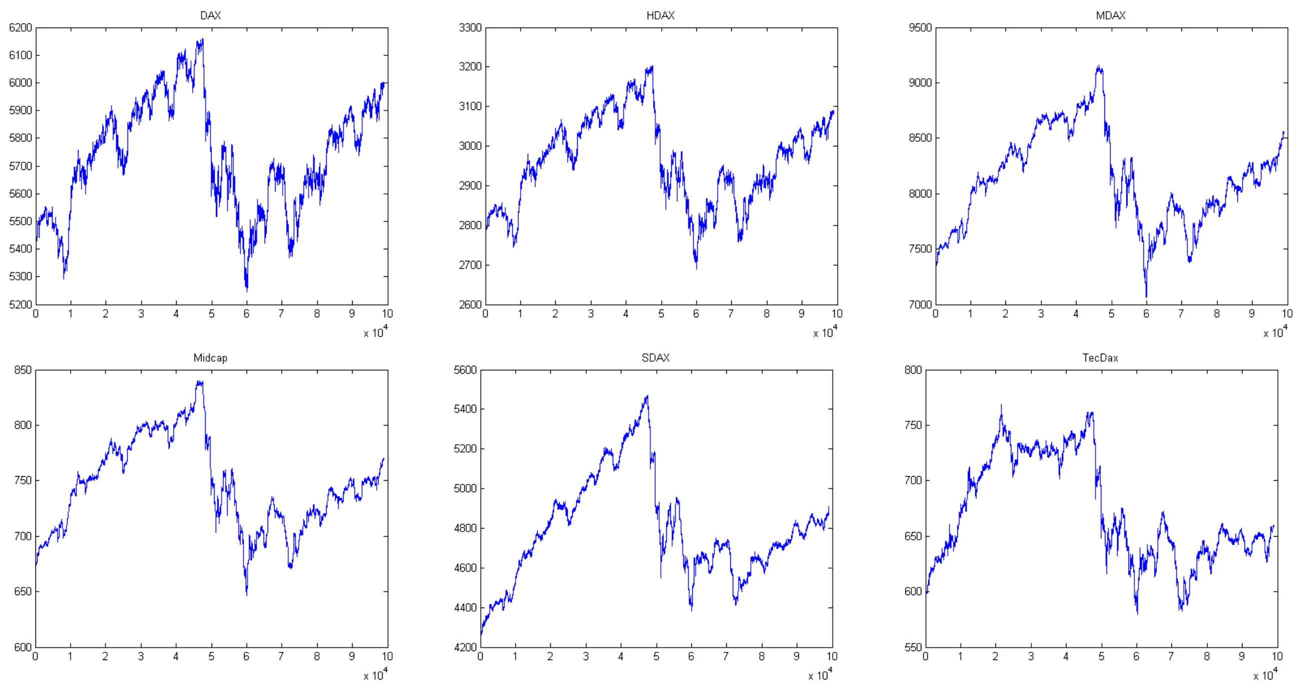

Figure 8.1: Plot of index movements

Table 8.4: Estimated parameters of the AMAR(1,1)-GARCH(1,1) model with residuals following normal distribution with zero mean and unit variance. Numbers in parentheses are the standard errors. These parameters are used in the empirical simulation.

	α_0	α_1	β_1	κ	γ_1	θ_1
DAX	9.01E-05	0.2768	-0.2543	7.39E-06	0.9727	0.0272
	(*5.93E-05*)	(*0.1146*)	(*0.1146*)	(*2.92E-08*)	(*8.82E-05*)	(*1.37E-04*)
HDAX	1.94E-04	-0.9001	0.8991	6.70E-06	0.9755	0.0244
	(*1.35E-04*)	(*5.67E-04*)	(*5.65E-04*)	(*2.21E-08*)	(*6.64E-05*)	(*9.45E-05*)
MDAX	6.02E-06	0.9701	-0.9397	9.89E-06	0.9602	0.0397
	(*3.51E-06*)	(*0.0020*)	(*0.0027*)	(*2.98E-08*)	(*5.63E-05*)	(*7.92E-05*)
Midcap	3.28E-06	0.9772	-0.9437	1.02E-05	0.9582	0.0417
	(*3.01E-06*)	(*0.0013*)	(*0.0014*)	(*2.63E-08*)	(*7.46E-05*)	(*1.00E-04*)
SDAX	9.44E-06	0.5972	-0.6292	5.71E-06	0.9634	0.0340
	(*2.54E-05*)	(*0.0465*)	(*0.0450*)	(*5.16E-08*)	(*1.60E-04*)	(*1.99E-04*)
TDAX	2.07E-05	0.3745	-0.4041	3.06E-05	0.9589	0.0335
	(*6.75E-05*)	(*0.0597*)	(*0.0589*)	(*1.44E-07*)	(*1.38E-04*)	(*1.49E-04*)

Table 8.5: Summary of the statistical characteristics of $\tilde{u}_t$ for six index returns.

	mean	variance	kurtosis	skewness	minimum	maximum	α	H_{FGN}	H_{fsn}
DAX	0.0002	1.0128	239.2300	-2.1231	-57.7260	37.2870	1.4247	0.5004	0.7022
HDAX	0.0019	1.0408	275.4600	-0.6858	-51.0230	40.6071	1.4021	0.5296	0.7429
MDAX	0.0013	1.0724	1151.9000	-6.9930	-84.2930	63.9842	1.6287	0.5010	0.6150
Midcap	0.0026	1.0723	838.5400	-4.7981	-81.8860	48.8033	1.6419	0.5319	0.6406
SDAX	0.0077	1.0000	117.1100	0.5478	-42.3140	41.4521	1.4461	0.6493	0.8408
TDAX	0.0016	1.0000	493.6200	0.3715	-58.7710	66.6430	1.5464	0.6261	0.7728

Table 8.6: Result of the ARCH-test of $\tilde{u}_t$ for different lags at $\alpha = 0.05$.

lag	$1min$	$5min$	$10min$	$2hours$	$1day$	$1week$	$2weeks$	$1month$
DAX	0.0589	0.4059	0.9996	1.9868	2.6270	3.0881	3.4956	5.8496
HDAX	0.0025	0.4342	1.9824	3.3537	4.0287	4.5703	5.1865	7.9064
MDAX	0.01489	0.0720	0.2269	0.3361	0.3685	0.3996	0.4382	0.7116
Midcap	0.0426	0.2289	0.9114	1.1740	1.2205	1.2641	1.3359	1.9323
SDAX	0.03849	0.9951	9.3913	12.1210	12.5150	12.7620	13.2190	15.9760
TDAX	0.0002	0.8199	1.0287	1.3896	1.4333	1.4926	1.5703	2.0892
Critical Value	3.8415	11.07	18.307	24.996	31.41	37.652	43.773	79.082

Table 8.7: The Ljung-Box-Pierce Q-test statistic of $\tilde{u}_t$ for different lags at $\alpha = 0.05$.

lag	$10min$	$30min$	$1hour$	$2hours$	$1day$	$1week$	$2weeks$	$1month$
DAX	18.4	44.2	66.7	109.0	405.2	2366.8	4785.2	9665.7
HDAX	43.1	67.0	83.6	127.6	451.0	2312.5	4988.6	9677.1
MDAX	53.9	94.1	110.6	151.8	471.3	2894.6	6058.8	12091.0
Midcap	56.5	110.3	131.3	178.6	488.9	2377.5	6177.3	11862.0
SDAX	18.6	54.1	108.0	197.2	571.3	2507.2	5042	10180.0
TDAX	62.0	142.5	221.9	268.6	569.4	2521.4	5086.7	10606.0
Critical Value	18.3	43.7	79.1	146.6	532.1	2515.1	4962.3	9829.0

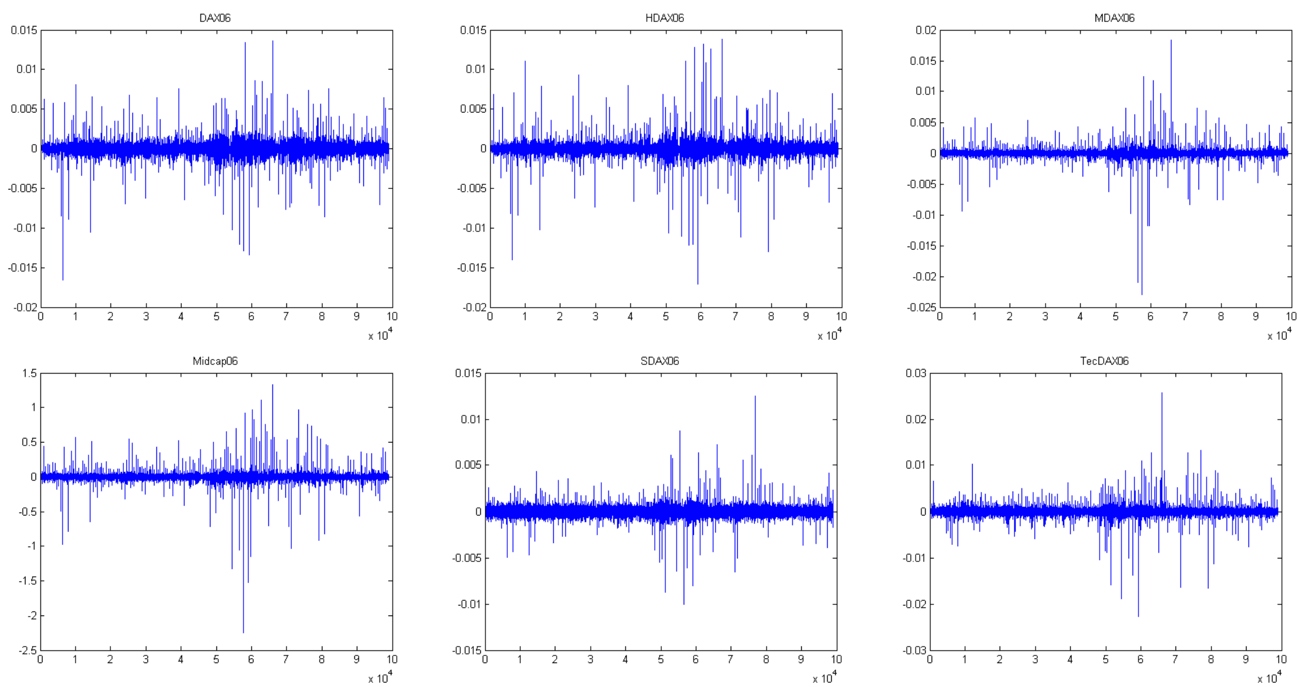

Figure 8.2: Plot of index return.

Table 8.8: Mean of the in-sample and out-of-sample goodness of fit statistic for alternative models with joint distribution (Gaussian copula).

			In-sample							Out-of-sample		
	FGN	FSN	Normal	Stable	GEV	GPD	FGN	FSN	Normal	Stable	GEV	G
KS	0.5060	0.5048	0.5061	0.5120	0.5086	0.8831	0.5074	0.5073	0.5084	0.5139	0.5091	0
AD	32.9751	32.8757	32.9751	33.3603	33.1234	57.4691	24.0592	24.0571	24.1089	24.3672	24.1423	41
CVM	177.3689	177.3087	177.3719	177.8170	177.4900	561.5001	93.9545	93.9241	93.9594	94.2369	93.9836	29
K	0.9982	0.9977	0.9983	0.9977	0.9979	0.9979	0.9983	0.9979	0.9986	0.9979	0.9981	0

Table 8.9: Mean of the in-sample and out-of-sample goodness of fit statistic for alternative models with joint distribution (t-copula).

			In-sample							Out-of-sample		
	FGN	FSN	Normal	Stable	GEV	GPD	FGN	FSN	Normal	Stable	GEV	G
KS	0.5049	0.5053	0.5062	0.5140	0.5085	0.8834	0.5080	0.5069	0.5075	0.5163	0.5108	0
AD	32.9006	32.9230	32.9776	33.4938	33.1240	57.4796	24.0882	24.0333	24.0658	24.4880	24.2235	41
CVM	177.3417	177.3131	177.3719	177.9565	177.5208	561.7679	93.9563	93.9064	93.9477	94.3381	94.0597	29
K	0.9983	0.9976	0.9983	0.9975	0.9979	0.9981	0.9984	0.9977	0.9985	0.9978	0.9980	0

Table 8.10: Mean of the in-sample and out-of-sample goodness of fit statistic for alternative models with joint distribution (Skewed t-copula).

			In-sample							Out-of-sample		
	FGN	FSN	Normal	Stable	GEV	GPD	FGN	FSN	Normal	Stable	GEV	G
KS	0.5013	0.4993	0.5017	0.5075	0.5046	0.8570	0.5038	0.5018	0.5054	0.5088	0.5068	0
AD	32.1926	31.7879	32.2002	32.3578	32.1734	54.0610	23.6406	23.4495	23.6880	23.8110	23.7328	40
CVM	176.3376	175.6668	176.3228	176.2896	176.2375	537.6912	93.3896	93.0380	93.4694	93.4583	93.4071	28
K	0.9902	0.9862	0.9901	0.9861	0.9883	0.9893	0.9909	0.9873	0.9911	0.9875	0.9896	0

Table 8.11: Summary statistics by groups of each creteria with respect to different models.

In-sample	FGN	FSN	Normal	Stable	GEV	GPD
Gaussian copula	52.9620	52.9217	52.9628	53.1717	53.0300	155.2125
Student t copula	52.9364	52.9347	52.9635	53.2405	53.0378	155.2822
Skewed t copula	52.5054	52.2350	52.5037	52.5352	52.4759	148.3996
Out-of-sample	FGN	FSN	Normal	Stable	GEV	GPD
Gaussian copula	29.8779	29.8716	29.8938	30.0290	29.9083	85.2602
Student t copula	29.8877	29.8611	29.8799	30.0851	29.9480	85.2583
Skewed t copula	29.6312	29.4941	29.6635	29.6914	29.6591	81.6776

Table 8.12: The matrix of Mahalanobis distances between each pair of group means of four creteria for in-sample and out-of-sample analysis. The numbers in parenthesis are the values of out-of-sample analysis.

	Gaussian copula	Student t copula	Skewed t copula
Gaussian copula	0	4.3098 (4.3623)	4.5399 (4.7630)
Student t copula	4.3098 (4.3623)	0	4.4174 (4.3802)
Skewed t copula	4.5399 (4.7630)	4.4174 (4.3802)	0

Chapter 9

Empirical Studies in High-Frequency Financial Econometrics

In this chapter, three empirical studies are shown based on the tasks of high-frequency financial econometrics. The first empirical study is forecasting of market volatility using Neural Network method introduced in chapter 2. The second empirical study is computing Value at Risk based on the ARMA(1,1)-GARCH(1,1) with different types of residuals (i.e., standard normal, stable distribution, fractional Gaussian noise, and fractional stable noise). The third empirical study is selecting a portfolio with respect to short-term trading based on high-frequency data.

9.1 Volatility Prediction

Here I apply the feedforward network regression to the intra-daily DAX index in 2006. The training set has a size of 40,000 observations and prediction set has a size of 10,000. The feedforward network model has 10 inputs and 10 hidden units for the price, i.e.,

$$p_t = f(p_{t-1}, p_{t-2}, \ldots, p_{t-10}, \theta) + \varepsilon_t$$

where ε_t is distributed with zero mean and variance σ^2. The feedforward network for the returns are

$$r_t = f(r_{t-1}, r_{t-2}, \ldots, r_{t-10}, \theta) + \varepsilon_t$$

where $r_t = \log(p_t) - \log(p_{t-1})$, ε_t is distributed with zero mean and variance σ^2. The function $f(\cdot)$ can be explicitly written as follows:

$$\begin{aligned} y_t &= f(x_t, \theta) + \varepsilon_t && (9.0\text{-}1) \\ &= s\Big(\beta_0 + \sum_{j=1}^{q} \beta_i\, g(\alpha_{i0} + \sum_{j=1}^{n} \alpha_{ij} x_{j,t})\Big) + \varepsilon. \end{aligned}$$

where $s(\cdot)$ and $g(\cdot)$ are known activation functions, and $q = 10$, $n = 10$ in this study.

Given the above network structure and the chosen funcional form of $s(\cdot)$ and $g(\cdot)$, a major empirical issue is then to estimate the unknown parameters θ with a sample of the dataset. A recursive estimation methodology called backpropogation is applied. In backpropogation, the starting point is a random weight θ vector that is updated according to

$$\hat{\theta}_{t+1} = \hat{\theta}_t + \gamma \nabla f(x_t, \hat{\theta}_t)\Big(y_t - f(x_t, \hat{\theta}_t)\Big) \tag{9.0-2}$$

where $\nabla f(x_t, \hat{\theta}_t)$ is the column gradient vector of $f(\cdot)$ with respect to $\hat{\theta}_t$, and γ is the parameter controlling the learning rate. White (1989) derived the statistical properties for this estimator by imposing appropriate conditions on the learning rate and definition of $s(\cdot)$ and $g(\cdot)$. For large samples and real-time applications, a recursive Newton algorithm is applied since it allows for adaptive estimation (see Kuan and White (1994)). The form of this recursive Newton algorithm is

$$\begin{aligned} \hat{\theta}_{t+1} &= \hat{\theta}_t + \gamma_t \hat{\psi}_t^{-1} \nabla f(x_t, \hat{\theta}_t)\Big(y_t - f(x_t, \hat{\theta}_t)\Big) \\ \hat{\psi}_{t+1} &= \hat{\psi}_t + \gamma_t\Big(\nabla f(x_t, \hat{\theta}_t) \nabla f(x_t, \hat{\theta}_t)^T - \hat{\psi}_t\Big), \end{aligned} \tag{9.0-3}$$

where $\hat{\psi}_t$ is an estimated, approximate Newton direction matrix and γ_t is a sequence of learning rates of order $1/t$. The estimation of the feedforward network regression is carried out by nonlinear least squares as follows

$$\min_{\theta} L(\theta) = \sum_{t=1}^{N}\Big(y_t - f(x_t, \theta)\Big)^2.$$

The price predictions are presented in Figure 9.1 and return predictions in Figure 9.2. In both training sets 40,000 observations are used and in the prediction, totally 10,000 one-step-ahead predictions are computed. The minimal mean squared prediction errors are 0,0024 and 0.0144 for Figure 9.1 and 9.2 respectively. It shows that this feedforward neural network can reach a relatively good prediction.

9.2 Computation of Value at Risk

Value at risk (VaR) is a risk measure popularly used in risk management. Given $\alpha \in (0, 1]$, R is the random gain and loss of an investment in a certain period, VaR of a random variable R at level of α is the absolute value of the worst loss not to be exceeded with a probability of at least α, more formally, if the α-quantile of R is $q_\alpha(R) = \inf\{r \in \Re : P[R \leq r] \geq \alpha\}$, the VaR at confidence level α of R is $VaR_\alpha(R) = q_\alpha(-R)$, see Dowd (2005).

The parametric approach of VaR estimation is based on the assumption that the financial returns R_t is a function of two components μ_t and ε_t. That is $R_t = f(\mu_t, \varepsilon_t)$. R_t can be regarded as a function of ε_t conditional on a given μ_t, typically this function takes a simple linear form $R_t = \mu_t + \varepsilon_t = \mu_t + \sigma_t u_t$. Usually μ_t is called the location component and σ_t is called the scale component. u_t is a i.i.d random variables which follows the probability density function f_u. The VaR based on information up to time t is

$$V\tilde{a}R_t := q_\alpha(R_t) = -\tilde{\mu}_t + \tilde{\sigma}\, q_\alpha(u) \tag{9.0-4}$$

where $q_\alpha(u)$ is the α-quantile implied by f_u.

Unconditional parametric approaches set μ_t and σ_t as constants, therefore the returns R_t are i.i.d random variables with density $\sigma^{-1} f_u(\sigma^{-1}(R_t - \mu))$. Conditional parametric approaches set location component and scale component as functions not constants. A typical time-varying conditional location setting is the ARMA(r,m) processes, i.e., the conditional mean equation is:

$$\mu_t = \alpha_0 + \sum_{i=1}^{r} \alpha_i R_{t-i} + \sum_{j=1}^{m} \beta_j \varepsilon_{t-j} \tag{9.0-5}$$

and typical time-varying conditional variance setting is GARCH(p,q) processes, that is,

$$\sigma_t^2 = \kappa + \sum_{i=1}^{p} \gamma_i \sigma_{t-i}^2 + \sum_{j=1}^{q} \theta_j \varepsilon_{t-j}^2 \tag{9.0-6}$$

Backtesting is the usual method to evaluate the VaR estimators and its forecasting quality. It can be performed for in-sample estimation evaluation and for out-of-sample interval forecasting evaluation. The backtesting is based on the indicator function I_t which is defined as $I_t(\alpha) = I(r_t < -VaR_t(\alpha))$. The indicator function shows violations of the quantiles of the loss distribution. Then the process $\{I_t\}_{t \in T}$ is a process of i.i.d Bernoulli variables with violation probability $1 - \alpha$. Christoffersen (1998) shows that evaluating the accuracy of VaR can be reduced to checking whether the number of violations is correct on average and the pattern of violations is consistent with i.i.d processes. In another word, an accurate VaR measure should satisfy both the unconditional coverage property and independent property. The unconditional coverage property means that the probability of realization of a loss in excess of the estimated $VaR_t(\alpha)$ must be exactly α%, that is, $P(I_t(\alpha) = 1) = \alpha$. The independent property means that previous VaR violations do not presage a future VaR violations.

In this study, I use high frequency DAX index data sampled at the 1-minute level in 2004 and 2005. The residuals in the parametric ARMA(1,1)-GARCH(1,1) are normal distribution, stable distribution, fractional stable noise, and fraction Gaussian noise. Table 9.1 shows the result of VaR computed by the above mentioned models for the in-sample estimation. The empirical VaR is computed by a nonparametric kernel estimator. Table 9.2 shows the difference of parametric ARMA(1,1)-GARCH(1,1) with different residuals (i.e., normal distribution, stable distribution, fractional stable noise, and fraction Gaussian noise) comparing with the empirical VaR value. From Table 9.2, the results show for the 95% VaR, that ARMA(1,1)-GARCH(1,1) model with fractional stable noise has minimal absolute distance to the empirical value. But the ARMA(1,1)-GARCH(1,1) model with standard normal residuals performs better when computing 99% VaR than other alternatives.

In order to test the prediction power of the computed VaR value. I test six different predictions with different size of training and prediction sets:

1. training period is 6 months (236160 data points) and one-step-ahead forecast for 6 months, 3 months, 1 month, 1 week, 1 day and 1 hour;

2. training period is 3 months (124800 data points) and one-step-ahead forecast for 3 months, 1 month, 1 week, 1 day and 1 hour;

3. training period is 1 month (40320 data points) and one-step-ahead forecast for 1 month, 1 week, 1 day and 1 hour;

4. training period is 1 week (9600 data points) and one-step-ahead forecast for 1 week, 1 day and 1 hour;

5. training period is 1 day (1920 data points) and one-step-ahead forecast for 1 day and 1 hour;

6. training period is 1 hour(240 data points) and one-step-ahead forecast for 1 hour.

Table 9.3 and 9.4 report the results for computing 95% VaR and 99% respectively. For the prediction of 95% VaR, ARMA(1,1)-GARCH(1,1) with fractional stable noise performs better than the other alternatives, but for the prediction of 99% VaR, although all values computed seem conservative, the ARMA(1,1)-GARCH(1,1) with standard normal distribution has better performance among others.

9.3 Portfolio Management

Jegadeesh and Titman (1993) report that equity returns exhibit short-term continuation. In their study, a momentum strategy that sorting firms by their previous returns over the past

3-12 months then holding the stocks with the best prior performance and short selling those with worst prior performance, has the possibility of generating an excess return of about one percent per month for US stocks. Following their findings, numerous researchers have been motivated to study momentum in other markets and other sample periods, such as Jegadeesh and Titman (1993, 2001), Grundy and Martin (2001), Lewellen (2002). Meanwhile, literature document that equity returns are negatively autocorrelated and their prices tend to revert to their trend line over long time horizon, see De Bondt and Thaler (1985, 1987) and others. As most important study, De Bondt and Thaler (1985) show an overreact effect that investors are overly optimistic about recent winners and overly pessimistic about recent losers. The research they did is based on the time horizon of 36 months of two portfolio returns of 35 New York Stock Exchange common stocks, for which two portfolios have been formed, one consists of past extreme winners over the prior three years and another consists of past extreme losers during the same time period. They show that a contrarian investment strategy that buying worst performance stocks and short selling best performance shocks over three to five years can also generate excess returns over the next three to five years. Their result of loser's portfolio outperforming and winner's portfolio under-performing has attracted other researchers to test mean reversion and to investigate profitability of contrarian strategies such as Chopra et al. (1992), Lakonishok et al. (1994), and Balvers et al. (2004).

Balvers et al. (2004) point out that most previous researchers study momentum and mean reversion separately, and the reason may be that a process of continuation leads to profitable momentum investment strategies and a process of mean reversion leads to profitable contrarian investment strategies. There is no direct contradiction in the profitability of both momentum and contrarian investment strategies since momentum strategies work for a sorting period ranging from 1 month (or more commonly 3 months) to 12 months and similar 1 (or 3) to 12 months holding period, while contrarian strategies typically work for a sorting and holding period from 3 to 5 years. The result shows the connection between return continuation for a time period up to 12 months and mean reversion for a time period about 3 to 5 years. The overreact hypothesis of De Bondt and Thaler (1985, 1987) and the behavioral finance theories of Barberis et al. (1998) and Hong and Stein (1999) imply the observed pattern of momentum in a short time period and mean reversion in a long time period. Balvers et al. (2004) demonstrate that momentum and mean reversion can simultaneously occur to the same set of assets. By using data for equity market indices for 18 mature markets, they illustrate that momentum and mean reversion are negatively correlated.

Jegadeesh (1990) found profitability of contrarian strategies for a very short time period: from one week to one month. Lee and Swaminathan (2000) referred an intermediate period of momentum from about one month up to one year. It is not clear for the definition of the exact time horizon of momentum and mean-reversion. The reason of the existing non-unified definition of time horizon scaling might be that in different literature different frequency data have been employed. Therefore, the observations are exhibited differently. In the past long time, the opinion has been held that short-term price fluctuation could be regarded as irrelevant noise and not worthwhile to collect and investigate, while recent discoveries in finance demonstrate that short-term observations or high frequency data has high information density and indicates long-term trends. With the development of modern computer techniques, analyzing high frequency data is possible. More and more researchers believe that high frequency data can reflect the changing of financial market more accurately than low frequency data can. Sahalia et al. (2005) confirm to use the sampled data as highest frequency as possible.

Biglova et al. (2004) and Rachev et al. (2007) show that momentum profitability can be significantly generated by using suitable criteria to select the component stocks for winner's and loser's portfolio. Martin et al. (2003) proposed conditional value at risk measures (STARR ratio and R-ratio) with respect to the stable distributed loss function. Empirical studies showed that these ratios perform better than other generally used ratios, see Biglova et al. (2004) and Rachev et al. (2005a).

Selecting rules

Basically, momentum strategies attempt to generate profit by selling stocks or portfolio with lowest past returns and buying stocks or portfolio with highest past returns. As it has been mentioned in the literature, momentum effect and mean reversion are connecting, that means past stocks with highest returns have the possibility to be worse performing in the future while stocks with lowest returns have the possibility to be better performing. Since stock return is tightly connected with its risk, past cumulative returns as the selecting rule for momentum strategy is not adequate since it neglected the risk imposed in the returns. It is necessary to consider the risk or component with risk measures as rules of selecting stocks to form the momentum portfolios. Biglova et al. (2004) and Rachev et al. (2007) show that momentum profitability can be significantly generated by using suitable criteria to select the component stocks for winner's and loser's portfolio. I propose two different types of selecting rules to rank stock returns, one is the univariate dimension rule which is based on a single reward-risk measure, such as, the Sharpe ratio, the Rachev ratios, and the STARR ratio. Another is multivariate dimension rule that combines several different reward-risk measures to install a multivariate risk measure structure to clarify the stock returns' performance.

There are several selecting rules based on single reward risk ratio. If r_j is the return series, then the cumulative return is just the summation. The Sharpe ratio is,

$$\gamma(r_j) = \frac{E(r_j\ -\ r_f)}{\sigma_{r_j - r_f}} \tag{9.0-7}$$

where r_f stands for risk free rate of asset return. Leland (1999) states that when the distribution of asset returns is not normal, Sharpe ratio is not completely reliable.

Besides standard deviation, value at risk (VaR) is another risk measure. Given $\alpha \in (0, 1]$, R is random gain and loss of an investment in certain period, VaR of a random variable R at level of α is the absolute value of the worst loss not to be exceeded with a probability of at least α, more formally, if α-quantile of R is $q_\alpha(R) = inf\{r \in \Re : P[R \leq r] \geq \alpha\}$, the VaR at confidence level α of R is $VaR_\alpha(R) = q_\alpha(-R)$. Arztner et al (1999) proposed the concept of a coherent risk measure. A risk measure is called coherent if it is monotonous, positively homogeneous, translation invariant and sub additive. As to this point standard deviation and VaR are not coherent risk measures. But Rockafellar and Uryasev (2002) show the conditional VaR or expected tail loss (ETL) is coherent and can be expressed as $CVaR_\alpha(-R) = E(-R| - R > VaR_\alpha(R))$. That is, for a loss $-R$ with $E(|-R|) < \infty$ and distribution function F_{-R}, the conditional value at risk at confidence level $\alpha \in (0, 1)$ is defined as

$$CVaR_\alpha(-R) = \frac{1}{1-\alpha}\int_\alpha^1 q_u(F_{-R})du = \frac{1}{1-\alpha}\int_\alpha^1 VaR_u(F_{-R})du \tag{9.0-8}$$

where $q_u(F_{-R})$ is the quantile function of F_{-R}.

The STARR ratio proposed by Martin et al. (2003) is the ratio of the expected excess return and its conditional value at risk or expected tail loss for stable loss function:

$$\gamma(r_j) = \frac{E(r_j - r_j)}{E(-r_j| - r_j > VaR_{\alpha 100\%}(r_j - r_f)} \tag{9.0-9}$$

The Rachev ratio, R-ratio(α, β) is,

$$\gamma(r_j) = \frac{E(-r_j| - r_j > VaR_{(1-\alpha)100\%}(r_f - r_j)}{E(-r_j| - r_j > VaR_{(1-\beta)100\%}(r_j - r_f)} \tag{9.0-10}$$

where $\alpha \in [0, 1]$ and $\beta \in [0, 1]$ and the the loss function is stable distributed.

I propose a selecting rule based on clustering different R-ratios as the multivariate dimension rule. After clustering R-ratios, the clusters are ranked from low to high. Portfolio formation is done based on the rank of R-ratio clusters. Buying winner's portfolio and selling loser's portfolio will generate the profits.

The trading strategy

Three assumptions are made for employing trading strategies tested in this empirical study, i.e.,

1. The market is sufficiently liquid. One can buy and sell stocks immediately after one makes the decision.

2. There is no transaction costs.

3. The execution of an order for selling or buying can be finished within 1 minute.

Setting Δt as the time unit for trading, for instance, in this study, Δt equals one week. t_0 is the starting time of holding period. Defining $t_0 - m\Delta t$ is the ranking period, during this period, each DAX component stock is ranked by certain criterion. The five highest return stocks based on the criterion are named as top five, and the five lowest return stocks are named as bottom five. The equally weighted portfolio of top five is named as winner's portfolio and equally weighted portfolio of bottom five is the loser's portfolio. $t_0 + n\Delta t$ is the holding period after t_0. Momentum strategies are going to exploit the continuation in stock returns. Jegadeesh and Titman (1993) gave the trading strategy of buying the winner's portfolio of highest past returns and selling loser's portfolio of lowest past returns. This study adopts this strategy but only to buy the winner's portfolio. Three different timing strategies have been used for the ranking and holding period, i.e., equal ranking and holding period ($m = n$), longer ranking period with short holding period ($m > n$) and short ranking period with longer holding period ($m < n$). In this study, the ranking and holding periods goes from 1 week to 5 weeks. By using the selecting rule based on the clustering of R-ratios, hierarchical and non-hierarchical clustering methods are both considered. In the hierarchical clustering, the winner's portfolio and loser's portfolio both contain five stocks, but for the non-hierarchical clustering, the number of stocks contained in the winner's and the loser's portfolio is solely depending on the partitioning of the returns.

Results

Table 9.5 reports the performance of trading based on portfolios selected by single measures. Generally, the R-ratio has a better performance. This result coincides with previous studies. Table 9.6 reports the performance of trading based on the portfolio selected by clustering R-ratios. Silhouette values evaluate the quality of clustering, that is, the higher the value is, the better the performance given by the clusters. The cluster of type 5 has the highest silhouette value and performance of the trading based on portfolios selected by this type of clusters is the best (0.8417) in this study. It shows that, given a correct clustering of R-ratios, the highest feasible profit can be reached.

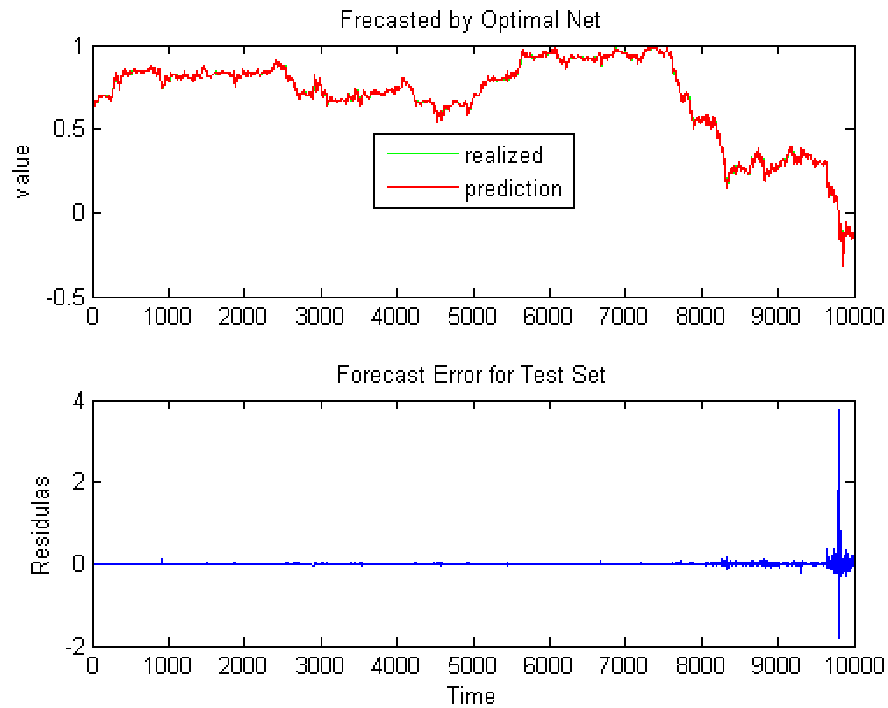

Figure 9.1: The intra-daily DAX 2006 predictions with a single-layer feedforward network model. The feedforward network model has 10 hidden units and 10 past prices.

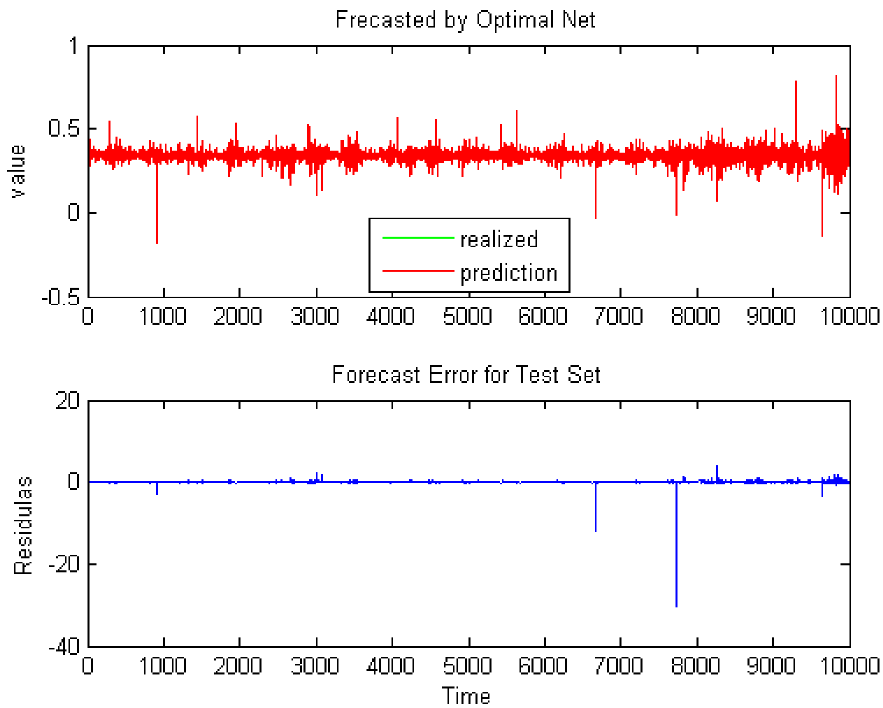

Figure 9.2: The intra-daily DAX 2006 predictions with a single-layer feedforward network model. The feedforward network model has 10 hidden units and 10 past returns.

Table 9.1: VaR values calculated by Kernel estimator (empirical) and ARMA(1,1)-GARCH(1,1) with different residuals (i.e., normal, stable, fractional stable noise, and fractional Gaussian noise).

			95%					99%		
	Empirical	Normal	stable	FSN	FGN	Empirical	Normal	stable	FSN	FGN
6 M	1.4018	1.8682	1.3874	1.3929	1.8727	2.6607	2.6382	3.7350	3.6972	2.6419
3 M	1.5218	2.0601	1.5090	1.4778	2.0544	2.8501	2.9267	3.9461	3.8371	2.9178
1 M	1.5402	2.3234	1.5562	1.6189	2.3158	2.9148	3.2939	4.3943	4.0179	3.2580
1 W	1.4901	1.8864	1.3882	1.4224	1.9109	2.8036	2.6445	4.4273	3.5853	2.6159
1 D	1.3861	1.7735	1.5159	1.4412	1.8549	2.4768	2.6127	4.5385	4.4081	2.8099
1 H	1.6819	1.5163	1.5085	1.7600	1.9079	2.7741	2.6503	5.6604	6.4754	2.9549

Table 9.2: Difference between VaR values calculated by Kernel estimator (empirical) and ARMA(1,1)-GARCH(1,1) with different residuals (i.e., normal, stable, fractional stable noise, and fractional Gaussian noise).

		95%				99%		
	Normal	stable	FSN	FGN	Normal	stable	FSN	FGN
6 M	-0.4044	0.0144	0.0089	-0.4709	0.0225	-1.0743	-1.0365	0.0188
3 M	-0.5383	0.0128	0.0440	-0.5326	-0.0766	-1.0960	-0.9870	-0.0677
1 M	-0.7832	-0.0160	-0.0787	-0.7756	-0.3791	-1.4795	-1.1031	-0.3432
1 W	-0.3963	0.1019	0.0677	-0.4208	0.1591	-1.6237	-0.7817	0.1877
1 D	-0.3874	-0.1298	-0.0551	-0.4688	-0.1359	-2.0617	-1.9313	-0.3331
1 H	0.1656	0.1734	-0.0781	-0.2260	0.1238	-2.8863	-3.7013	-0.1808
$\lvert \sum \rvert$	2.6752	0.4483	0.3325	2.8947	0.8970	10.2215	9.5409	1.1313

Table 9.3: Violations of VaR (95%) computed from ARMA(1,1)-GARCH(1,1) with different residuals.

	6 months	(236160)	3 months	(124800)	1 month	(40320)	1 week	(9600)	1 day	(1920)	1 hour	(240)
	Normal	FSN	Normal	FSN	Normal	FSN	Normal	FSN	Normal	FSN	Normal	FSN
1 hour	13	25	12	25	4	11	13	21	8	16	8	6
(240)	*0.0542*	*0.1042*	*0.0500*	*0.1042*	*0.0167*	*0.0458*	*0.0542*	*0.0875*	*0.0333*	*0.0667*	*0.0333*	*0.0250*
1 day	35	71	49	105	33	73	47	110	59	95		
(1920)	*0.0182*	*0.0370*	*0.0255*	*0.0547*	*0.0172*	*0.0380*	*0.0245*	*0.0573*	*0.0307*	*0.0495*		
1 week	117	246	185	425	187	458	260	545				
(9600)	*0.0122*	*0.0256*	*0.0193*	*0.0443*	*0.0195*	*0.0477*	*0.0271*	*0.0568*				
1 month	545	1124	686	1602	661	1624						
(40320)	*0.0135*	*0.0279*	*0.0170*	*0.0397*	*0.0164*	*0.0403*						
3 months	1572	3165	1810	4222								
(124800)	*0.0126*	*0.0254*	*0.0145*	*0.0338*								
6 months	2744	5538										
(236160)	*0.0116*	*0.0235*										
	Stable	FGN	Stable	FGN	Stable	FGN	Stable	FGN	Stable	FGN	Stable	FGN
1 hour	25	14	25	14	13	4	18	13	15	8	6	6
(240)	*0.1042*	*0.0583*	*0.1042*	*0.0583*	*0.0542*	*0.0167*	*0.0750*	*0.0542*	*0.0625*	*0.0333*	*0.0250*	*0.0250*
1 day	71	37	98	50	90	33	96	47	93	56		
(1920)	*0.0370*	*0.0193*	*0.0510*	*0.0260*	*0.0469*	*0.0172*	*0.0500*	*0.0245*	*0.0484*	*0.0292*		
1 week	244	113	408	182	542	180	484	260				
(9600)	*0.0254*	*0.0118*	*0.0425*	*0.0190*	*0.0565*	*0.0188*	*0.0504*	*0.0271*				
1 month	1146	556	1608	684	1955	647						
(40320)	*0.0284*	*0.0138*	*0.0399*	*0.0170*	*0.0485*	*0.0160*						
3 months	3259	1588	4221	1801								
(124800)	*0.0261*	*0.0127*	*0.0338*	*0.0144*								
6 months	5728	2781										
(236160)	*0.0243*	*0.0118*										

Table 9.4: Violations of VaR (99%) computed from ARMA(1,1)-GARCH(1,1) with different residuals.

	6 months	(236160)	3 months	(124800)	1 month	(40320)	1 week	(9600)	1 day	(1920)	1 hour	(240)
	Normal	FSN	Normal	FSN	Normal	FSN	Normal	FSN	Normal	FSN	Normal	FSN
1 hour	2	0	5	2	1	1	9	2	4	1	3	0
(240)	*0.0083*	*0.0000*	*0.0208*	*0.0083*	*0.0042*	*0.0042*	*0.0375*	*0.0083*	*0.0167*	*0.0042*	*0.0125*	*0.0000*
1 day	13	4	17	10	12	7	24	5	22	4		
(1920)	*0.0068*	*0.0021*	*0.0089*	*0.0052*	*0.0063*	*0.0036*	*0.0125*	*0.0026*	*0.0115*	*0.0021*		
1 week	40	10	62	25	69	36	114	30				
(9600)	*0.0042*	*0.0010*	*0.0065*	*0.0026*	*0.0072*	*0.0038*	*0.0119*	*0.0031*				
1 month	198	48	241	93	257	120						
(40320)	*0.0049*	*0.0012*	*0.0060*	*0.0023*	*0.0064*	*0.0030*						
3 months	593	188	637	268								
(124800)	*0.0048*	*0.0015*	*0.0051*	*0.0021*								
6 months	1055	354										
(236160)	*0.0045*	*0.0015*										
	Stable	FGN	Stable	FGN	Stable	FGN	Stable	FGN	Stable	FGN	Stable	FGN
1 hour	0	2	2	5	1	1	1	9	1	2	0	2
(240)	*0.0000*	*0.0089*	*0.0083*	*0.0208*	*0.0042*	*0.0042*	*0.0042*	*0.0375*	*0.0042*	*0.0083*	*0.0000*	*0.0083*
1 day	4	13	11	18	6	11	2	25	4	20		
(1920)	*0.0021*	*0.0068*	*0.0057*	*0.0094*	*0.0031*	*0.0057*	*0.0010*	*0.0130*	*0.0021*	*0.0104*		
1 week	11	40	24	65	35	69	21	112				
(9600)	*0.0011*	*0.0042*	*0.0025*	*0.0068*	*0.0036*	*0.0072*	*0.0022*	*0.0117*				
1 month	58	197	90	241	133	248						
(40320)	*0.0014*	*0.0049*	*0.0022*	*0.0060*	*0.0028*	*0.0062*						
3 months	213	591	260	651								
(124800)	*0.0017*	*0.0047*	*0.0021*	*0.0052*								
6 months	388	1044										
(236160)	*0.0016*	*0.0044*										

Table 9.5: Performance of trading based on the portfolios selected with single measure.

week	$E(r_p)$ mean	$E(r_p)$ Sharpe	$E(r_p)$ STARR95%	$E(r_p)$ R(0.01,0.01)	$E(r_p)$ R(0.05,0.01)	$E(r_p)$ R(0.01,0.05)	$E(r_p)$ R(0.05,0.05)	$E(r_p)$ R(0.1,0.1)
m=1, n=1	-0.1080	-0.1004	-0.1080	-0.0267	-0.0182	-0.0396	-0.0215	-0.0759
m=2, n=2	0.0485	0.0508	0.0485	0.0436	0.0521	0.0104	0.0167	0.0436
m=3, n=3	0.0311	0.0311	0.0311	0.0312	0.0312	0.0311	0.0243	0.0230
m=2, n=1	0.1144	0.0651	0.1144	0.0876	0.1131	0.0262	0.0753	0.2137
m=3, n=1	-0.0421	-0.0421	-0.0421	-0.0241	-0.0241	0.0685	0.1057	-0.0482
m=4, n=1	0.0489	0.0573	0.0915	0.1445	0.1412	0.1389	0.0987	0.1346
m=5, n=1	0.0277	0.0072	0.0277	0.0458	0.0458	-0.0371	-0.0260	0.0458
m=3, n=2	0.0909	0.0909	0.0909	0.1079	0.1079	0.0828	0.1006	0.0866
m=4, n=2	0.0726	0.0726	0.0726	0.0726	0.0801	0.0607	0.0705	0.0726
m=1 , n=2	0.0412	0.0448	0.0412	0.0783	0.0929	0.0658	0.0778	0.0491
m=1, n=3	-0.0107	-0.0091	0.0311	-0.0080	-0.0006	0.0033	-0.0020	-0.0136
m=1, n=4	0.0154	0.0165	0.0154	0.0325	0.0272	0.0237	0.0272	0.0166
m=1, n=5	0.0148	0.0108	0.0148	0.0257	0.0226	0.0346	0.0221	0.0110
m=2, n=3	0.0282	0.0282	0.0282	0.0343	0.0438	0.0189	0.0343	0.0343
m=2, n=4	0.0086	0.0086	0.0086	0.0086	0.0086	-0.0063	0.0086	0.0086
$\sum_n E(r_p)$	0.3817	0.3327	0.4665	0.6541	0.7238	0.3446	0.6126	0.6025

Table 9.6: Performance of trading based on the portfolios selected by clustered R-ratios.

	week	$E(r_p)$ cluster type 1	$E(r_p)$ cluster type 2	$E(r_p)$ cluster type 3	$E(r_p)$ cluster type 4	$E(r_p)$ cluster type 5
m=1,	*n=1*	0.0356	0.0484	0.0432	0.0921	0.1183
m=2,	*n=2*	0.0191	0.0510	0.0051	0.0242	0.0205
m=3,	*n=3*	0.0057	0.0057	0.0057	0.0113	0.0357
m=2,	*n=1*	-0.0075	0.0787	0.0262	0.0147	0.0655
m=3,	*n=1*	0.0537	-0.0436	0.0843	0.0457	0.0848
m=4,	*n=1*	0.2294	0.2105	0.1755	0.1002	0.1963
m=5,	*n=1*	-0.0176	-0.0176	0.0065	0.0057	0.0065
m=3,	*n=2*	0.0820	0.0737	0.0737	0.0960	0.1141
m=4,	*n=2*	0.0216	0.0216	0.0750	0.1002	0.0689
m=1 ,	*n=2*	0.0793	0.1752	0.0995	0.0358	0.1273
m=1,	*n=3*	-0.0355	-0.0420	-0.0254	-0.0238	-0.0190
m=1,	*n=4*	0.0026	0.0095	0.0095	0.0045	0.0045
m=1,	*n=5*	0.0190	0.0190	0.0190	-0.0045	-0.0024
m=2,	*n=3*	0.0135	0.0029	0.0002	0.0178	0.0118
m=2,	*n=4*	0.0051	0.0084	0.0084	0.0187	0.0086
	$\sum_n E(r_p)$	**0.5065**	**0.4851**	**0.6070**	**0.6215**	**0.8417**
	(Silhouette values)	(0.6420)	(0.6372)	(0.6863)	(0.6874)	(0.7392)

Chapter 10

Conclusion

In this dissertation, I point out several aspects of studies for high-frequency financial econometrics. I proposed several models with respect to univariate and multivariate high-frequency financial time series data. These models show the advantages of using self-similar processes and copulas. Recently, we have experienced an explosion in the number of financial studies based on high-frequency data, but the existing literature left several problems unsolved. Using self-similar processes and copulas can fill up some topics missing in the literature. However, based on the viewpoint of data mining for high-frequency financial econometrics, several future investigations need to be done while in this dissertation they are only briefly mentioned. The empirical studies proposed in Chapter 9 indeed highlight three major directions for future investigation, that is, (1) improving the prediction method based on more robust statistical models, (2) finding models for risk management, and (3) proposing short-term or day trading models based on more realistic situations.

The models proposed in this dissertation can be used to identify arbitrage opportunities. Someone argues that "Indeed, the arbitrage opportunities found in empirical studies have become smaller and smaller, due to constant progress in speeds of both communication devices and computers." This statement has been discussed in different occasions. For example, in behavioral finance, people believe that the arbitrage opportunities can be captured from market anomalies. But after such arbitrage opportunities have been pointed out by academia, no one can duplicate it in the real market. The conclusions are[1] (1), once the arbitrage opportunities have been pointed out, everyone tries to duplicate it, then it cannot be used any more, and, (2) the techniques used to find these arbitrage opportunities by academia are not robust. At the high-frequency level, the market is heterogenous, which means that certain arbitrage

[1] see Chapters 15, 17, and 18 in Handbook of the economics of finance II, edited by George M. Constantinides, Milton Harris, and Rene M. Stulz, Elsevier, 2003.

opportunities might be hedged out due to the heterogenity.

It can be accepted (as I believe) that certain arbitrage opportunities pointed out by academia might be realized to generate profit, if and only if the following conditions are satisfied:

- No one else in the market is using same strategy to take advantage of arbitrage opportunities as you do.
- The techniques used for finding arbitrage opportunities are robust enough.
- If such techniques are not robust enough, you might have the possibility to take advantage of the arbitrage opportunities only if the market satisfies your assumptions with which you applied the techniques to find such arbitrage opportunities.

It is also important to keep in mind that the following three variables i.e., model, algorithm, and data, dominate the quality of an econometric study.

There is an ongoing methodological debate about model specification. Kennedy (2003) points out that "Until about the mid-1970s, econometricians were too busy doing econometrics to worry about the principles that were or should be guiding empirical research". He thinks that the econometric profession began to examine model specification problems with a critical eye after people found failure of large-scale econometric model and gaps between how econometrics was taught and how it was applied by practitioners. Pagan (1987, 1995) stylizes three main approaches to the specification problem with respect to considerable risk of oversimplification, i.e., average economic regression (AER), Test-Test-Test (TTT), and fragility analysis (FA). AER refers to the researcher begins with a specification from assumed-known theories, which is characterized as proceedings from a simple model and "testing up" to specify a more general model. Comparing with AER, TTT is the other way that can be characterized as "testing down" from a general model to a more specific model. FA considers a range of "key" variables and undertakes an "extreme bounds analysis" after identifying a general family of models in which the coefficients of the key variables are estimated using all feasable combinations of included/excluded "doubtful" variables. Lütkepohl (2007) discusses general-to-specific (gets) modeling and specific-to-general (spec) modeling based on the empirical work related to cointegrated systems. Most researchers might agree with that all models are imperfect representations of the phenomenon of being models, due to the essence of simplification by filtering many aspects of the "real world", and occasionally, these neglected aspects may turn out to be important. Although such controversy has never been resolved, some principles have been generally accepted to guide model specification (see, for example, Kennedy (2003)). These principles are summarized as follows: (1) Economic theory should be the foundation of and guiding force in model specification; (2) Tests should be applied for misspecifications; and (3) Models can be used either for policy evaluation or forecasting.

With the help of large computer systems, econometric/statistical inference and forecasting can be realized expectedly. Unfortunately, numerical methods required for solving problems involved in inference and forecasting are limited. Such algorithmic limitation reduces the accuracy of computation. Approximation and simulation methods are not merely sought out by financial economists but also by researchers from mathematics and computer science. Given a "good" and sophisticated model, a limited algorithm applied to get the perspective result will weaken the power of this model.

In Chapter 2, I have mentioned several aspects for using high-frequency financial data. Good data quality will positively support the analysis. Mining data with appropriate technologies is the process of discovering hidden, actionable, and meaningful patterns, profiles, and trends. Only informative data can provide a solid foundation of analysis.

All econometric/statistical analysis (with respect to model, algorithm, and data) rely explicitly or implicitly on a number of assumptions, which generally aim at formalizing what has been known about the model and data, and at the same time aim at making the resulting model manageable from the theoretical and computational viewpoints. There have been attempts to justify such assumptions, although such attempts are easily proven inadequate. For example, normal (Gaussian) distribution, is the case I just mentioned. Therefore, robustness is considered ultimately in econometric/statistical analysis. The robust approach to statistical modeling and data analysis aims at deriving methodologies that produce reliable and accurate sought out solutions by considering less unrealistic assumptions. High-frequency financial econometrics is based on the analysis of high frequency financial data that is possessed of a higher level of uncertainty than that of low frequency data. In order to reduce risk caused by the uncertainties, robust methods should be applied in future investigations.

Bibliography

[1] Adams, G., G. McQueen, R. Wood (2004) The effects of inflation news on intra-daily stock returns. Journal of Business 77(3): 547-574.

[2] Aït-Sahalia, Y., P. Mykland, L. Zhang (2005) How often to sample a continuous-time process in the presence of market microstructure noise. Review of Financial Studies 18(2): 351-416.

[3] Akaike, H. (1974) A new look at the statistical model identification. IEEE Transactions on Automatic Control 19(6): 716-723.

[4] Alexander, C. (2001) Market Models: A Guide to Financial Data Analysis. John Wiley & Sons, New York.

[5] Andersen, T., T. Bollerslev (1997) Intraday periodicity and volatility persistence in financial markets. Journal of Empirical Finance 4(2-3): 115-158.

[6] Andersen, T., T. Bollerslev (1998) Answering the skeptics: yes, standard volatility models do provide accurate forecasts. International Economic Review 39(4): 885-905

[7] Andersen, T., T. Bollerslev, A. Das (2001) Variance-ratio statistics and high-frequency data: testing for changes in intraday volatility patterns. Journal of Finance 59(1): 305-327.

[8] Andersen, T., T. Bollerslev, F. Diebold, H. Ebens (2001) The distribution of realized stock return volatility. Journal of Financial Economics 61(1): 43-76.

[9] Andersen, T., T. Bollerslev, F. Diebold, P. Labys (2000) Exchange rate returns standardized by realized volatility are (nearly) Gaussian. Multinational Finance Journal 4(3-4): 159-179.

[10] Andersen, T., T. Bollerslev, F. Diebold, P. Labys (2001) The distribution of realized exchange rate volatility. Journal of the American Statistical Association 96(453): 42-55.

[11] Andersen, T., T. Bollerslev, F. Diebold, P. Labys (2003) Modeling and forecasting realized volatility. Econometrica 71(2): 579-626.

[12] Andersen, T., T. Bollerslev, F. Diebold (2005) Parametric and nonparametric measurements of volatility. In: Aït-Sahalia Y, Hansen L (eds) Handbook of Financial Econometrics. North Holland Press, Amsterdam, forthcoming.

[13] Andersen, T., T. Bollerslev, P. Christoffersen, F. Diebold (2006) Volatility and correlation forecasting. In: G. Elliott, C. Granger, A. Timmermann (eds.) Handbook of Economic Forecasting. North Holland Press, Amsterdam.

[14] Andersen, T., T. Bollerslev, P. Christoffersen, F. Diebold (2005) Practical volatility and correlation modeling for financial market risk management. In: M. Carey, R. Stulz (eds) Risks of Financial Institutions. University of Chicago Press.

[15] Artzner, P., F. Delbaen, J. M. Eber, D. Heath (2000) Coherent measures of risk. Mathematical Finance 9(3): 203-228.

[16] Baillie, R. (1996) Long memory processes and fractional integration in econometrics. Journal of Econometrics 73(1): 5-59.

[17] Baillie, R., T. Bollerslev (1991) Intra-day and inter-market volatility in foreign exchange rates. Review of Economic Studies 58(3): 565-585.

[18] Baillie, R., T. Bollerslev, H. Mikkelsen (1996) Fractional integrated generalized autoregressive conditional heteroskedasticity. Journal of Econometrics 74(1): 3-30.

[19] Balvers, R., Y. Wu, E. Gilliland (2000) Mean reversion across national stock markets and parametric contrarian investment strategies. Journal of Finance 55(2): 745-772.

[20] Banerjee, A., G. Urga (2005) Modeling structural breaks, long memory and stock market volatility: an overview. Journal of Econometrics 129(1-2): 1-34.

[21] Barberis, N., A. Shleifer, R. Vishny (1998) A model of investor sentiment. Journal of Financial Economics 49(3): 307-343.

[22] Bardet, J., G. Lang, G. Oppenheim, A. Philippe, M. Taqqu (2003) Generators of long-range dependent processes: A survey. In: P. Doulkhan, G. Oppenheim, M. Taqqu (eds.) Theory and Applications of Long-range Dependence. Birkhäuser, Boston.

[23] Barndorff-Nielsen, O., N. Shephard (2002) Econometric analysis of realized volatility and its use in estimating stochastic volatility models. Journal of Royal Statistical Society B 64(2): 253-280.

[24] Bauwens, L., D. Veredas (2004) The stochastic conditional duration model: a latent variable model for the analysis of financial durations. Jouranl of Econometrics 119(2): 381-412.

[25] Bauwens, L., P. Giot (2000) The logarithmic ACD model: An application to the bid-ask quote process of three NYSE stocks. Annales d'Economie et de Statistique 60: 117-149.

[26] Bauwens, L., P. Giot (2001) Econometric Modelling of Stock Market Intraday Activity. Kluwer Academic Publishers, Boston.

[27] Bauwens, L., P. Giot (2003) Asymmetric ACD models: Introducing price information in ACD models. Empirical Economics 28(4): 709-731.

[28] Bauwens, L., N. Hautsch (2006) Modelling financial high frequency data using point processes. In Handbook of Financial Time Series. Springer-Verlag, forthcoming.

[29] Bernardo, J. M., A. Smith (1994) Bayesian Theory. Wiley: New York.

[30] Beltratti, A., C. Morana (1999). Computing value at risk with high frequency data. Journal of Empirical Finance 6(5): 431-455.

[31] Beran, J. (1994) Statistics for Long-Memory Processes. Chapman & Hall, New York.

[32] Bessembinder, H., K. Venkataraman (2004) Does an electronic stock exchange need an upstairs market? Journal of Financial Economics 73(1): 3-36.

[33] Bhansali, R., P. Kokoszaka (2006) Prediction of long-memory time series: a tutorial review In: G. Rangarajan, M. Ding (eds.) Processes with Long-Range Correlations: Theory and Applications. Springer, New York.

[34] Biglova, A., S. Ortobelli, S. Rachev, S. Stoyanov (2004) Comparison among different approaches for risk estimation in portfolio theory. Journal of Portfolio Management (2004-Fall).

[35] Blair, B., S. Poon, S. Taylor (2001) Forecasting S& P 100 volatility: The incremental information content of implied volatilities and high-frequency index returns. Journal of Econometrics 105(1): 5-26.

[36] Bloomfield, R., M. O'Hara (1999) Market transparency: Who wins and who loses? Review of Financial Studies 12(1): 5-35.

[37] Bloomfield, R., M. O'Hara (2000) Can transparent markets survive? Journal of Financial Economics 55(3): 425-459.

[38] Boehmer, E., G. Saar, L. Yu (2005) Lifting the veil: An analysis of pre-trade transparency at the NYSE. Journal of Finance 60(2): 783-815.

[39] Bollen, B., B. Inder (2002) Estimating daily volatility in financial markets utilizing intraday data. Journal of Empirical Finance 9(5): 551-562.

[40] Bollen, N., T. Smith, R. Whaley (2004) Modeling the bid/ask spread: Measuring the inventory holding premium. Journal of Financial Economics 72(1): 97-141.

[41] Bollerslev, T., R. Chou, K. Kroner (1992) ARCH modeling in finance: A review of the theory and empirical evidence. Journal of Econometrics 52(1-2): 783-815.

[42] Bollerslev, T., I. Domowitz (1993) Trading patterns and prices in the interbank foreign exchange market. Journal of Finance 48(4): 1421-1443.

[43] Bollerslev, T., H. Mikkelsen (1996) Modelling and pricing long-memory in stock market volatility. Journal of Econometrics 73(1): 154-184.

[44] Bollerslev, T., J. Cai, F. Song (2000) Intraday periodicity, long memory volatility and macroeconomic announcement effects in the US treasury bond market. Journal of Empirical Finance 7: 37-55.

[45] Bollerslev, T., J. Wright (2001) High-frequency data, frequency domain inference and volatility forecasting. The Review of Economics and Statistics 83(4): 596-602.

[46] Bollerslev, T., Y. Zhang (2003) Measuring and modeling systematic risk in factor pricing models using high-frequency data. Journal of Empirical Finance 10(5): 533-558.

[47] Box, G. E. P. (1953) Non-normality and tests of variances. Biometrika 40(3-4): 318-335.

[48] Box, G. E. P., S. L. Anderson (1955) Permutation theory in the derivation of robust criteria and the study of departures from assumption. Journal of Royal Statistics Sociaty (B) 17: 1-34.

[49] Breidt, F., N. Crato, P. de Lima (1998) The detection and estimation of long memory in stochastic volatility. Journal of Econometrics 83(1-2): 325-348.

[50] Burnham, K. P., A. R. Anderson (1998) Model Selection and Inference: A Practical Information-Theoretic Approach. Springer-Verlag: New York.

[51] Chopra, N., J. Lakonishok, J. Ritter (1992) Measuring abnormal performance: Do stocks overreact? Journal of Financial Economics 31(2): 235-268.

[52] Chordia, T., R. Roll, A. Subrahmanyam (2002) Order imbalance, liquidity and market returns. Journal of Financial Economics 65(1): 111-130.

[53] Christoffersen, P., (1998) Evaluating interval forecasts. International Economic Review 39(4): 841-862.

[54] Chung, K., B. Van Ness (2001) Order handling rules, tick size and the intraday pattern of bid-ask spreads for Nasdaq stocks. Journal of Financial Markets 4(2): 143-161.

[55] Corsi, F., U. Kretschmer, S. Mittnik, C. Pigorsch (2005) The volatility of realized volatility. CFS Working Paper No.2005/33.

[56] Coughenour, J., D. Deli (2002) Liquidity provision and the organizational form of NYSE specialist firms. Journal of Finance 57(2): 841-869.

[57] Coval, J., T. Shumway (2001) Is sound just noise? Journal of Finance 56(5): 1887-1910.

[58] Crovella, M., A. Bestavros (1997) Self-similarity in World Wide Web traffic: Evidence and possible causes. IEEE/ACM Transactions on Networking 5(6): 835-846.

[59] Dacorogna, M., R. Gençay, U. Müller, R. Olsen, O. Pictet (2001) An Introduction of High-Frequency Finance. Academic Press, San Diego.

[60] Davison, A. C., D. V. Hinkley (1997) Bootstrap Methods and Their Application. Cambridge University Press: Cambridge.

[61] De Bondt, W., R. Thaler (1985) Does the stock market overreacts? Journal of Finance 40(3): 793-805.

[62] De Bondt W, Thaler R (1987) Further evidence on investor overreaction and stock market seasonality. Journal of Finance 42(3): 557-581

[63] Diebold, F., A. Inoue (2001) Long memory and regime switching. Journal of Econometrics 105(1): 131-159.

[64] Doukhan, P., G. Oppenheim, M. Taqqu (2003) (eds) Theory and Applications of Long-Range Dependence. Birkhäuser, Boston.

[65] Dowd, K., (2005) Measuring Market Risk (2nd Edition). John Wiley & Sons: West Sussex.

[66] Duda, R. O., P. E. Hart, D. G. Stork (2001) Pattern Classification. Wiley: New York.

[67] Dufour, A., R. Engle (2000) Time and the price impact of a trade. Journal of Finance 55(6): 2467-2498.

[68] Easley, D., M. O'Hara (2003) Microstructure and asset pricing. In: G. Constantinides, M. Harris, R. Stulz (eds.) Handbook of the Economics of Finance. Elsevier, North Holland.

[69] Efron, B. (1979) Bootstrap methods: another look at the Jackknife. The Annals of Statistics 7(1): 126.

[70] Engle, R. (1982) Autoregressive conditional heteroskedasticity with estimates of the variance of U.K. inflation. Econometrica 50(4): 987-1007.

[71] Engle, R. (2000) The econometrics of ultra-high frequency data. Econometrica 68(1): 1-22.

[72] Engle, R., J. Russell (1998) Autoregressive conditional duration: A new model for irregularly spaced transaction data. Econometrica 66(5): 1127-1162.

[73] Engle, R., J. Russell (2004) Analysis of high frequency data. In: Y. Aït-Sahalia, L. Hansen (eds.) Handbook of Financial Econometrics. North Holland Press, Amsterdam.

[74] Everitt, B. S. (1993) Cluster Analysis (3rd edition). Edward Arnold: London.

[75] Faust, J., J. Rogers, E. Swanson, J. Wright (2003) Identifying the effects of monetary policy shocks on exchange rates using high frequency data. Journal of European Economic Association 1(5): 1031-1057.

[76] Feldmann, A., A. Gilbert, W. Willinger (1998) Data networks as cascades: investing the multifractal nature of Internet WAN traffic. ACM Computer Communication Review 28(4): 42-55.

[77] Feldmann, A., A. Gilbert, W. Willinger, T. Kurtz (1998) The changing nature of network traffic: scaling phenomena. ACM Computer Communication Review 28(2): 5-29.

[78] Feng, D., G. Jiang, P. Song (2004) Stochastic conditional duration models with "leverage effect" for financial transaction data. Journal of Financial Econometrics 2(3): 390-433.

[79] Fernandes, M., J. Gramming (2005) Nonparametric specification tests for conditional duration models. Journal of Econometrics 127(1): 35-68.

[80] Fleming, J., C. Kirby, B. Ostdiek (2003) The economic value of volatility timing using "realized" volatility. Journal of Financial Economics 67(3): 473-509.

[81] Franke, G., D. Hess (2000) Information diffusion in electronic and floor trading. Journal of Empirical Finance 7(5): 455-478.

[82] Gençay, R., F. Selçuk, B. Whitcher (2001) Differentiating intraday seasonalities through wavelet multi-scaling. Physica A 289(3-4): 543-556.

[83] Gençay, R., F. Selçuk, B. Whitcher (2002) An Introduction to Wavelets and Other Filtering Methods in Finance and Economics. Elsevier, San Diego.

[84] Gençay, R., G. Ballocchi, M. Dacorogna, R. Olsen, O. Pictet (2002) Real-time trading models and the statistical properties of foreign exchange rates. International Economic Review 43(2): 463-491.

[85] Gerhard, F., N. Hautsch (2002) Volatility estimation on the basis of price intensities. Journal of Empirical Finance 9(1): 57-89.

[86] Geweke, J., S. Porter-Hudak (1983) The estimation and application of long memory time series models. Journal of Time Series Analysis 4: 221-238.

[87] Ghysels, E. (2000) Some econometric recipes for high-frequency data cooking. Journal of Business and Economic Statistics 18(2): 154-163.

[88] Ghysels, E., C. Gouriéroux, J. Jasiak (2004) Stochastic volatility duration models. Journal of Econometrics 119(2): 413-433.

[89] Ghysels, E., C. Gouriéroux, J. Jasiak (1998) High frequency financial time series data: some stylized facts and models of stochastic volatility. In C. Dunis, B. Zhou (eds.), Nonlinear Modeling of High Frequency Financial Time Series, 127-159.

[90] Giot, P., S. Laurent (2004) Modeling daily Value-at-Risk using realized volatility and ARCH type models. Journal of Empirical Finance 11(3): 379-398.

[91] Giraitis, L., D. Surgailis (1990) A central limit theorem for quadratic forms in strongly dependent random variables and its application to asymptotical normality of Whittle's estimate. Probability Theory and Related Fields 86(1): 87-104.

[92] Giudici, P. (2003) Applied Data Mining. Wiley: West Sussex.

[93] Everitt, B. S. (1993) Cluster Analysis (3rd edition). Edward Arnold: London.

[94] Gleason, K., I. Mathur, M. Peterson (2004) Analysis of intraday herding behavior among the sector ETFs. Journal of Empirical Finance 11(5): 681-694.

[95] Goodhart, C., M. O'Hara (1997) High frequency data in financial market: issues and applications. Journal of Empirical Finance 4(2-3): 73-114.

[96] Goodhart, C., R. Payne (2000) (ed.) The Foreign Exchange Market: Empirical Studies with High-Frequency Data. Palgrave, Macmillan.

[97] Gouriéroux, C., J. Jasiak (2001) Financial Econometrics: Problems, Models and Methods. Princetion University Press, Princeton and Oxford.

[98] Grammig, J., K. Maurer (1999) Non-monotonic hazard functions and the autoregressive conditional duration model. The Econometrics Journal 3(1): 16-38.

[99] Grundy, B., J. Martin (2001) Understanding the nature and the risks and the sources of the rewards to momentum investing. Review of Financial Studies 14(1): 29-78.

[100] Gwilym, O., C. Sutcliffe (1999) High-Frequency Financial Market Data. Risk Books.

[101] Haas, M., S. Mittnik, M. Paolella (2004) Mixed normal conditional heteroskedasticity. Journal of Financial Econometrics 2(2): 211-250.

[102] Hair, J. F., W. Black, B. Babin, R. Anderson, R. Tatham (2006) Multivariate Data Analysis (6th edition). Pearson Prentice Hall, New Jersey.

[103] Haldrup, N., M. Nielsen (2006) A regime switching long memory model for electricity prices. Journal of Econometrics 135(1-2): 349-376.

[104] Hannan, E. (1973) The asymptotic theory of linear time series models. Journal of Applied Probability 10(1): 130-145.

[105] Harris, L. (2003) Trading and exchanges. Oxford university press, Oxford.

[106] Hasbrouck, J. (1996) Modeling microstructure time series. In: G. Maddala, C. Rao (eds.) Statistical Methods in Finance (Handbook of Statistics, Volume 14). North-Holland, Amsterdam.

[107] haupt Haupt, R. L., S. E. Haupt (2004) Practical Genetic Algorithms. Wiley, New York.

[108] Higuchi, T. (1988) Approach to an irregular time series on the basis of the fractal theory. Physica D 31(2): 277-283.

[109] Hochberg, H., A. Tamhane (1987) Multiple Comparison Procedures. Wiley.

[110] Holland, J. (1975) Adaption in Natural and Artificial Systems. University of Michigan Press, Ann Arbor.

[111] Hong, H., J. Stein (1999) A unified theory of under reaction, momentum trading and overreaction in asset markets. Journal of Finance 54(6): 2143-2184.

[112] Hong, H., J. Wang (2000) Trading and returns under periodic market closures. Journal of Finance 55(1): 297-354.

[113] Hotchkiss, E., T. Ronen (2002) The informational efficiency of the corporate bond market: An intraday analysis. The Review of Financial Studies 15(5): 1325-1354.

[114] Huang, R., H. Stoll (2001) Tick size, bid-ask spreads, and market structure. Journal of Financial and Quantitative Analysis 36(4): 503-522.

[115] Hurst, H. (1951) Long-term storage capacity of reservoirs. Transactions of the American Society of Civil Engineers 116(2447): 770-808.

[116] Hurvich, C., J. Brodsky (2001) Broadband semiparametric estimation of the memory parameter of a long memory time series using fractional exponential models. Journal of Time Series Analysis 22(2): 221-249.

[117] Ioannidis, J. (2005) Why most published research findings are false. PloS Med 2(8): e124.

[118] Jain, P., G. Joh (1988) The dependence between hourly prices and trading volume. Journal of Financial and Quantitative Analysis 23(3): 269-284.

[119] Jasiak, J. (1998) Persistence in intertrade durations. Finance 19: 166-195.

[120] Jegadeesh, N., S. Titman (1993) Returns to buying winners and selling losers: Implication for stock market efficiency. The Journal of Finance 48(1): 65-91.

[121] Jegadeesh, N., S. Titman (2001) Profitability of momentum strategies: An evaluation of alternative explanations. Journal of Finance 56(2): 699-720.

[122] Johnson, R., D. Wichern (2002) Applied Multivariate Statistical Analysis (Fifth Edition). Prentice-Hall, Inc: Upper Saddle River.

[123] Kalay, A., O. Sade, A. Wohl (2004) Measuring stock illiquidity: an investigation of the demand and supply schedules at the TASE. Journal of Financial Economics 74(3): 461-486.

[124] Kantardzic, M. (2003) Data Mining: Concepts, Models, Methods, and Algorithms. IEEE Press: Piscataway, NJ.

[125] Kennedy, P. (2003) A guide to econometrics (5th edition). Blackwell, Oxford.

[126] Kohavi, R. (1995) A study of cross-validation and bootstrap for accuracy estimation and model selection. Proceedings of the Fourteenth International Joint Conference on Artificial Intelligence 2(12): 11371143.

[127] Kolmogoroff, A. (1940) Wienersche Spiralen und einige andere interessante Kurven im Hilbertschen Raum. Comptes Rendus (Doklady) Acad. Sci. URSS (N.S.) 26: 115-118.

[128] Kryzanowski, L., H. Zhang (2002) Intraday market price integration for shares cross-listed internationally. Journal of Financial and Quantitative Analysis 37(2): 243-269.

[129] Kuan, C. M., H. White (1994) Artificial neural network: An econometric perspective. Econometric Reviews 13(1): 1-91.

[130] Kullback, S., R. A. Leibler (1951) On information and sufficiency. Annals of Mathematical Statistics 22(1): 79-86.

[131] Lakonishok, J., A. Shleifer, R. Vishny (1994) Contrarian investment: extrapolation and risk. Journal of Finance 49(5): 1541-1578.

[132] Lasko, T. A., J. G. Bhagwat, K. H. Zou, L. Ohno-Machado (2005) The use of receiver operating characteristic curves in biomedical informatics. Journal of Biomedical Informatics 38(5): 404-415.

[133] Lamperti, J. (1962) Semi-stable stochastic Processes. Transactions of the American Mathematical Society 104(1): 62-78.

[134] Langdon, W. B., R. Poli (2002) Foundations of Genetic Programming. Springer, New York.

[135] Lee, C., B. Swaminathan (2000) Price Momentum and Trading Volume. Journal of Finance 55(5): 2017-2069.

[136] Leland, H. E. (1999) Beyond mean-variance: performance measurement in a non-symmetrical world. Financial Analyst Journal 55(1): 27-36.

[137] Leland, W., M. Taqqu, W. Willinger, D. Wilson (1994) On the self-similar nature of Ethernet traffic. IEEE/ACM Transactions on Networking 2: 1-15.

[138] Lewellen, J. (2002) Momentum and autocorrelation in stock returns. Review of Financial Studies 15(2): 533-563.

[139] Lo, A., A. C. MacKinlay (1990) Data snooping biases in tests of financial asset pricing models. The Review of Financial Studies 3(3): 431-467.

[140] Lo, A. (1991) Long-term memory in stock market prices. Econometrica 5985): 1279-1313.

[141] Longstaff, F., A. Wang (2004) Electricity forward prices: A high-frequency empirical analysis. Journal of Finance 59(4): 1877-1900.

[142] Lütkepohl, H. (2007) General-to-specific or specific-to-general modelling? An opinion on current econometric terminology. Journal of Econometrics 136(1): 319-324.

[143] MacQueen, J. B. (1967) Some methods for classification and analysis of multivariate observations. Proceedings of 5th Berkeley Symposium on Mathematical Statistics and Probability 1: 281-297.

[144] Madhavan, A. (2000) Market Microstructure: A survey. Journal of Financial Markets 3(3): 205-258.

[145] Maejima, M. (1983) On a class of self-similar processes. Zeitschrift für Vahrscheinlichkeitstheorie und verwandte Gebiete 62: 235-245.

[146] Maejima, M., S. Rachev (1987) An ideal metric and the rate of convergence to a self-similar process. Annals of Probability 15(2): 708-727.

[147] Mahalanobis, P. C. (1936) On the generalised distance in statistics. Proceedings of the National Institute of Science of India 12: 49-55.

[148] Mandelbrot, B. (1963) New methods in statistical economics. Journal of Political Economy 71(5): 421-440.

[149] Mandelbrot, B. (1975) Limit theorems on the self-normalized range of weakly and strongly dependent processes. Zeitschrift für Wahrscheinlichkeitstheoie und verwandte Gebiete 31: 271-285.

[150] Mandelbrot, B. (1983) The Fractal Geometry of Nature. Freeman, San Francisco.

[151] Mandelbrot, B., J. Wallis (1969) Computer experiments with fractional Gaussian noises. Water Resources Research 5: 228-267.

[152] Mandelbrot, B., M. Taqqu (1979) Robust R/S analysis of long-run serial correlation. Bulletin of the International Statistical Institute 48: 59-104.

[153] Mansfield, P., S. Rachev, G. Samorodnitsky (2001) Long strange segments of a stochastic process and long-range dependence. The Annals of Applied Probability 11(3): 878-921.

[154] Marinelli, C., S. Rachev, R. Roll, H. Göppl (2000) Subordinated stock price models: heavy tails and long-range dependence in the high-frequency Deutsche Bank price record. In: G. Bol, G. Nakhaeizadeh, K. Vollmer (eds.) Datamining and Computational Finance. Physica-Verlag, Heidelberg.

[155] Marinucci, D., P. Robinson (1999) Alternative forms of fractional Brownian motion. Journal of Statistical Planning and Inference 80(1-2): 111-122.

[156] Martens, M. (2001) Forecasting daily exchange rate volatility using intraday returns. Journal of International Money and Finance 20(1): 1-23.

[157] Martens, M., Y. Chang, S. Taylor (2002) A comparison of seasonal adjustment methods when forecasting intraday volatility. The Journal of Financial Research 25(2): 283-299.

[158] Martin, D., S. Rachev, F. Siboulet (2003) Phi-alpha optimal portfolios and extreme risk management. Wilmott Magazine of Finance November: 70-83.

[159] Masulis, R., L. Shivakumar (2002) Does market structure affect immediacy of stock price responses to news? Journal of Financial and Quantitative Analysis 37(4): 617-648.

[160] McInish, T., R. Wood (1992) An analysis of intradaily patterns in bid/ask spreads for NYSE stocks. Journal of Finance 47(2): 753-746.

[161] McNelis, P. D. (2005) Neural Networks in Finance: Gaining Predictive Edge in the Market. Elsevier: Amsterdam.

[162] Mittnik, S., M. Paolella, S. Rachev (2002) Stationary of stable power-GARCH processes. Journal of Econometrics 106(1): 97-107.

[163] Morana, C., A. Beltratti (2004) Structural change and long-range dependence in volatility of exchange rate: either, neither or both? Journal of Empirical Finance 11(5): 629-658.

[164] Moulines, E., P. Soulier (1999) Broadband log-periodogram regression of time series with long-range dependence. Annals of Statistics 27(4): 1415-1439.

[165] Müller, U., M. Dacorogna, R. Olsen, O. Pictet, M. Schwarz, C. Morgenegg (1990) Statistical study of foreign exchange rates, empirical evidence of a price change scaling law, and intraday analysis. Journal Banking and Finance 14(6): 1189-1208.

[166] Müller, U., M. Dacorogna, R. Dave, R. Olsen, O. Pictet, J. von Weizsäcker (1997) Volatilities of different time resolutions: analyzing thedynamics of market components. Journal of Empirical Finance 4(2-3): 213-239.

[167] Naik, N., P. Yadav (2003) Do dealer firms manage inventory on a stock-by-stock or a portfolio basis? Journal of Financial Economics 69(1): 325-353.

[168] Neilsen, B., N. Shephard (2004) Econometric analysis of realized covariation: high frequency based covariance, regression, and correlation in financial economics. Econometrica 72(3): 885-925.

[169] Nelsen, R. B. (2006) An Introduction to Copulas. Springer: New York.

[170] Nelson, D. (1991) Conditional heteroskedasticity in asset returns: a new approach. Econometrica 59(2): 347-370.

[171] O'Hara, M. (1995) Market Microstructure Theory. Blackwell, Cambridge.

[172] Pagan, A. (1987) Three econometric methodologies: a critical appraisal. Journal of Economic Surveys 1(1-2): 3-23.

[173] Pagan, A. (1995) Three econometric methodologies: an update. In L. Oxley et al. (eds.) Surveys in Econometrics. Oxford: Basil Blackwell, 30-41.

[174] Paxson, V. (1997) Fast, approximation synthesis of fractional Gaussian noise for generating self-similar network traffic. Computer Communications Review 27(5): 5-18.

[175] Pearson, R. K. (2005) Mining Imperfect Data: Dealing with Contamination and Incomplete Records. SIAM, Philadelphia.

[176] Peng, C., S. Buldyrev, M. Simons, H. Stanley, A. Goldberger (1994) Mosaic organization of DNA nucleotides. Physical Review E 49: 1685-1689.

[177] Percival, D. B., A. T. Walden (2000) Wavelet Methods for Time Series Analysis. Cambridge University Press: Cambridge.

[178] Peterson, M., E. Sirri (2002) Order submission strategy and the curious case of marketable limit orders. Journal of Financial and Quantitative Analysis 37(2): 221-241.

[179] Rachev, S., S. Mittnik (2000) Stable Paretian Models in Finance. Wiley, New York.

[180] Rachev, S., G. Samorodnitsky (2001) Long strange segments in a long range dependent moving average. Stochastic Processes and their Applications 93(1): 119-148.

[181] Rachev, S., C. Menn, F. Fabozzi (2005) Fat-Tailed and Skewed Asset Return Distributions. Wiley, New Jersey.

[182] Rachev, S., S. Mittnik, F. Fabozzi, S. Focardi, T. Jasic (2007) Financial Econometrics: From Basics to Advanced Modelling Techniques. Wiley: Hoboken New Jersey.

[183] Rachev, S., J. Hsu, B. Bagasheva, F. Fabozzi (2007a) Bayesian Methods in Finance. Wiley: Hoboken New Jersey.

[184] Rachev, S., T. Jasic, S. Stoyanov, F. Fabozzi (2007) Momentum strategies based on reward-risk stock selection criteria. Journal of Banking and Finance 31(8): 2325-2346.

[185] Rangarajan, G., M. Ding (2006) (ed) Processes with Long-Range Correlations: Theory and Applications. Springer, New York.

[186] Rao, C. R., E. Wegman, J. Solka (2005) (ed) Handbook of Statistics 24: Data Mining and Data Visualization. Elsevier, Amsterdam.

[187] Robinson, P. (1991) Testing for strong serial correlation and dynamic conditional heteroskedasticity in multiple regression. Journal of Econometrics 47(1): 67-84.

[188] Robinson, P. (1994) Efficient tests of nonstationary hypotheses. Journal of the American Statistical Association 89(428): 1420-1437.

[189] Robinson, P. (1995a) Log-periodogram regression of time series with long range dependence. The Annals of Statistics 23(3): 1048-1072.

[190] Robinson, P. (1995b) Gaussian semiparametric estimation of long range dependence. The Annals of Statistics 23(5): 1630-1661.

[191] Robinson, P. (2003) (ed) Time Series with Long Memory. Oxford University Press, New York.

[192] Rockafellar, R. T., S. Uryasev (2002) Conditional Value-at-Risk for general loss distribution. Journal of Banking and Finance 26(7): 1143-1471.

[193] Ron, K. (1995) A study of cross-validation and bootstrap for accuracy estimation and model selection. Proceedings of the Fourteenth International Joint Conference on Artificial Intelligence 2(12): 11371143.

[194] Samorodnitsky, G. (1994) Possible sample paths of self-similar alpha-stable processes. Statistics and Probability Letters 19(3): 233-237.

[195] Samorodnitsky, G. (1996) A class of shot noise models for financial applications. In: C. Heyde, Y. Prohorov, R. Pyke, S. Rachev (eds) Proceeding of Athens International Conference on Applied Probability and Time Series. Volume 1: Applied Probability. Springer, Berlin.

[196] Samorodnitsky, G. (1998) Lower tails of self-similar stable processes. Bernoulli 4(1): 127-142.

[197] Samorodnitsky G, Taqqu M (1994) Stable Non-Gaussian Random Processes: Stochastic Models with Infinite Variance. Chapman & Hall/CRC, Boca Raton

[198] Schwartz, R., R. Francioni (2004) Equity Markets in Action. John Wiley & Sons, Hoboken.s

[199] Schwarz, G. (1978) Estimating the dimension of a model. Annals of Statistics 6(2): 461-464.

[200] Silverman, B. W. (1986) Density Estimation for Statistics and Data Analysis. Chapman & Hall: New York.

[201] Smith, B., D. Turnbull, R. White (2001) Upstairs market for principal agency trades: analysis of adverse information and price effects. Journal of Finance 56(5): 1723-1746.

[202] Spackman, K. A. (1989) Signal detection theory: Valuable tools for evaluating inductive learning. Proceedings of the Sixth International Workshop on Machine Learning: 160163, San Mateo, CA: Morgan Kaufman.

[203] Stoev, S., V. Pipiras, M. Taqqu (2002) Estimation of the self-similarity parameter in linear fractional stable motion. Signal Processing 82(12): 1873-1901.

[204] Stoev, S., M. Taqqu (2004) Simulation methods for linear fractional stable motion and FARIMA using the Fast Fourier Transform. Fractals 12(1): 95-121.

[205] Stoyanov, S., B. Racheva-Iotova (2004a) Univariate stable laws in the field of finance—approximation of density and distribution functions. Journal of Concrete and Applicable Mathematics 2(1): 38-57.

[206] Stoyanov, S., B. Racheva-Iotova (2004b) Univariate stable laws in the field of finance—parameter estimation. Journal of Concrete and Applicable Mathematics 2(4): 24-49.

[207] Stoyanov, S., B. Racheva-Iotova (2004c) Numerical methods for stable modeling in financial risk management. In: S. Rachev (ed) Handbook of Computational and Numerical Methods. Birkhäuser, Boston.

[208] Spulber, D. (1999) Market Microstructure: Intermediaries and the Theory of the Firm. Cambridge University Press, Cambridge.

[209] Sun, W., S. Rachev, F. Fabozzi (2007a) Fratals or i.i.d.: evidence of long-range dependence and heavy tailedness form modeling German equity market volatility. Journal of Economics and Business 59(6): 575-595.

[210] Sun, W., S. Rachev, F. Fabozzi, P. Kalev (2007b) Fratals in duration: capturing long-range dependence and heavy tailedness in modeling trade duration. Annals of Finance, forthcoming.

[211] Sun, W., S. Rachev, F. Fabozzi (2007c) Long-Range Dependence, Fractal Processes, and Intra-daily Data. In: D. Seese, C. Weinhardt, F. Schlottmann (eds.) Handbook on Information Technology in Finance. Springer, forthcoming.

[212] Sun, W., S. Rachev, F. Fabozzi, P. Kalev (2007d) A New Approach to Modeling Co-movement of International Equity Markets: Evidence of Unconditional Copula-Based Simulation of Tail Dependence. Empirical Economics, forthcoming.

[213] Taqqu, M., V. Teverovsky (1998) Estimating long-range dependence in finite and infinite variance sereis. In: R. Adler, R. Feldman, M. Taqqu (eds) A Practical Guide to Heavy Tails. Birkhäuser, Boston.

[214] Taqqu, M., V. Teverovsky, W. Willinger (1995) Estimators for long-range dependence: an empirical study. Fractals 3(4): 785-798.

[215] Taylor, S., X. Xu (1997) The incremental volatility information in one million foreign exchange quotations. Journal of Empirical Finance 4(4): 317-340.

[216] Teverovsky, V., M. Taqqu (1995) Testing for long-range dependence in the presence of shifting means or a slowly declining trend using a variance-type estimator. Preprint.

[217] Teverovsky, V., M. Taqqu, W. Willinger (1999) A critical look at Lo's modified R/S statistic. Journal of Statistical Planning and Inference 80(1-2): 211-227.

[218] Teyssiére, G., A. Kirman (2006) (eds) Long Memory in Economics. Springer, Berlin.

[219] Theissen, E. (2002) Price discovery in floor and screen trading systems. Journal of Empirical Finance 9(4): 455-474.

[220] Thomakos, D., T. Wang (2003) Realized volatility in the futures markets. Journal of Empirical Finance 10(3): 321-353.

[221] Tsay, R. (2002) Analysis of Financial Time Series. John Wiley & Sons, New York.

[222] Tukey, J. W. (1962) The future of data analysis. Annals of Mathematical Statistics 33(1):1-67.

[223] Velasco, C., P. Robinson (2000) Whittle pseudo-maximum likelihood estimation for nonstationary time series. Journal of the American Statistical Association 95(452): 1229-1243.

[224] Veredas, D., J. Rodrìguez-Poo, A. Espasa (2002) On the (intraday) seasonality of a financial point process: a semiparametric approach. Working Paper, CORE DP 2002/23, Université catholique de Louvain.

[225] Wasserfallen, W., H. Zimmermann (1985) The behavior of intradaily exchange rates. Journal of Banking and Finance 9(1): 55-72.

[226] Weston, J. (2002) Electronic communication networks and liquidity on the Nasdaq. Journal of Financial Services Research 22(1-2s): 125-139.

[227] White, H. (1989) Some asymptotic results for learning in single hidden-layer feedforward network models. Journal of the American Statistical Association 84(408): 1003-1013.

[228] Whittle, R. (1951) Hypothesis Testing in Time Series Analysis. Uppsala, Almqvist.

[229] Willinger, W., M. Taqqu, A. Erramilli (1996) A bibliographical guide to self-similar traffic and performance modeling for modern high-speed networks. Stochastic Networks: Theory and Applications, Royal Statistical Society Lecture Notes Series, Vol. 4. Oxford University Press, Oxford.

[230] Willinger, W., M. Taqqu, R. Sherman, D. Wilson (1997) Self-similarity through high-variability: statistical analysis of Ethernet LAN traffic at the source level. IEEE/ACM Transactions on Networking 5(1): 71-86.

[231] Willinger, W., V. Paxson, M. Taqqu (1998) Self-similarity and heavy tails: structural modeling of network traffic. In: R. Adler, R. Feldman, M. Taqqu (eds) A Practical Guide to Heavy Tails. Birkhäuser, Boston, 27-53.

[232] Wood, R., T. McInish, J. Ord (1985) An investigation of transaction data for NYSE stocks. Journal of Finance 40(3): 723-739.

[233] Zhang, M., J. Russell, R. Tsay (2001) A nonlinear autoregressive conditional duration model with applications to financial transaction data. Journal of Econometrics 104(1): 179-207.

[234] Zucchini, W. (2000) An introduction to model selection. Journal of Mathematical Psychology 44(1): 41-61.

[235] Zumbach, G., U. Müller (2001) Operators on inhomogeneous time series. International Journal of Theoretical and Applied Finance 4(1): 147-178.

[236] Zweig, M. H., G. Campbell (1993) Receiver-operating characteristic (ROC) plots: a fundamental evaluation tool in clinical medicine. Clinical chemistry 39(8): 561-577.