Title
Statistical Modeling of Marked Point Processes and (Ultra-)High Frequency Data

Author
Wen, Musen

Publication Date
2010

Peer reviewed|Thesis/dissertation

UNIVERSITY OF CALIFORNIA
RIVERSIDE

Statistical Modeling of Marked Point Processes and (Ultra-)High Frequency Data

A Dissertation submitted in partial satisfaction
of the requirements for the degree of

Doctor of Philosophy

in

Applied Statistics

by

Musen Wen

August 2010

Dissertation Committee:
Professor Keh-Shin Lii, Chairperson
Professor Subir Ghosh
Professor Gloria González-Rivera

The Dissertation of Musen Wen is approved:

Committee Chairperson

University of California, Riverside

Acknowledgements

I would like to express my gratitude to my advisor Professor Keh-Shin Lii for his advice, support and encouragement during my doctoral research. All the discussions, meetings, and great ideas are particular valuable in my whole life. I would like to thank Professor Subir Ghosh and Professor Gloria González-Rivera for their helpful discussions and serving on my dissertation committee. I would like to thank all the faculty members and graduate students in the department of Statistics for their wonderful teaching and friendship during the past few years. Also, I would like to thank the Graduate Division for the financial support, without which my dream to be a statistician would not come true. Finally, I would like to thank my parents, my sister and my wife, Jie Lian for their patience, support and love.

ABSTRACT OF THE DISSERTATION

Statistical Modeling of Marked Point Processes and (Ultra-)High Frequency Data

by

Musen Wen

Doctor of Philosophy, Graduate Program in Applied Statistics
University of California, Riverside, August 2010
Professor Keh-Shin Lii, Chairperson

The studies of stock transaction data, i.e., both the regularly-spaced *high frequency* data and the irregularly-spaced *ultra-high frequency* data, have been among the frontiers of modern financial data analysis. One of those data sets is the Trade and Quote (TAQ) data from the New York Stock Exchange (NYSE), which is a collection of all stock transaction information (e.g., the transaction date, time, prices and volumes, etc.) for every trading day. The analysis of the intraday transaction data still remains highly challenging today, especially on the statistical modeling aspects.

In this research, two new statistical modeling frameworks, namely, the Multi-Logit Mixture Autoregressive (MLMAR) models and the multivariate Mixture Transition Distribution (MMTD) models, are proposed respectively to handle above two types of financial data. The models are the univariate and multivariate generation of the MTD-type time series models.

The MLMAR time series model is a univariate time series model for the regularly-spaced intraday stock prices, which includes the exogenous information, such as the transaction volumes, the trading frequencies or any other market information, into the modeling framework. The MMTD model is a modeling framework for marked point processes in general, and ultra-high frequency transaction data in particular.

In both modeling frameworks, we solve a series of problems, which include the model specification, parameter estimation, prediction methodology and their applications to the stock transaction data. To show the capacity and advantage of the new models over the existing models, we also compare the new models with those benchmark models and show the new models' advantages in terms of either describing the underlying data generating process or prediction performance. For each class of time series model, potential extensions and related modeling issues are also discussed thereafter.

Contents

List of Tables

List of Figures

Chapter 1

Introduction

1.1 Ultra-high frequency data and issues

Since the seminal paper by Engle and Russell [26], the modeling and statistical analysis of financial transaction data have been among the frontiers of modern financial econometrics. One of the most famous transaction dataset is the Trade and Quote (TAQ) database from the New York Stock Exchange (NYSE).

Table 1.1 illustrates a typical record of the stock transaction data in the TAQ database. The first column is the symbol of the stock. In Table 1.1, it is a record for the IBM stock. The IBM stock transaction data has been widely used in the studies of ultra-high frequency data (a named first used by Engle). The second column is the date of the transactions. In our research, we mainly focus on intraday data, i.e., the transaction data from a particular trading day. The third column is one of the most interested quantities - the time for each transaction.

This can be seen as the *time stamp* when each transaction is made. The most distinguished feature of transaction data is that all transactions are irregularly-spaced in time. From the view of stochastic processes, the arrival of transactions is a point process. In the fourth and fifth column, the corresponding price and size/volume of the transaction are given. For different markets, the last column are quite different. These may generally include all other information related to the transactions.

Table 1.1: NYSE Trade and Quote (TAQ) database.

SYMBOL	DATE	TIME	PRICE	SIZE	OTHERS
IBM	20070103	8 : 24 : 40	97.4	100	-
IBM	20070103	8 : 25 : 23	97.4	400	-
IBM	20070103	8 : 25 : 47	97.4	100	-
IBM	20070103	8 : 29 : 11	97.4	100	-
IBM	20070103	8 : 29 : 11	97.4	300	-
IBM	20070103	8 : 32 : 40	97.5	300	-
IBM	20070103	8 : 34 : 11	97.4	1000	-
···	···	···	···	···	-

In practice, such datasets are also called *tick-by-tick* data, or simply *tick data*. Some distinguished features include irregularly-spaced transactions, the discreteness of price changes and intraday seasonality. For example, the empirical studies of stock market data show that the intraday data has seasonality. This happens because the transactions generally occur more frequently during the open and close hours and less frequently during the lunch time. This can be characterized by introducing a "U"-shape intensity function for the point process. In what follows, we take a look at these features in a bit more details.

The intraday prices and price changes are all discrete, because the smallest price change, as regulated by the stock exchanges, lives on a small collection of discrete values. For example, in stock market, the price change is a multiple of $1/16$ cent in 1990's. In recent years, the smallest jump of price is set to be $1/10$ cent. The smallest price change is thus called "a tick". For most of the intraday stock data, over 90% of the price changes fall into the categories of $\{-5, \cdots, -1, 0, 1, \cdots, 5\}$ multiple of one tick. The discreteness of the price changes causes a significant positive excess of kurtosis of the return distribution.

Trend or seasonality is also widely observed in the intraday transaction data. A typical "U"-shape pattern widely exists in transaction volumes series, frequencies of trades and the spread series, etc. A reverse "U"-shape pattern is generally observed in the transaction durations series. This implies that, in modeling long term ultra-high frequency data, these types of patterns should be captured firstly. In fact, people would generally deseasonalize the target time series using different methods. For example, [17] uses a second-order polynomial function to detrend the intraday data.

The most important features of intraday data are the dependence structures. These have been intensively studied in the past and could be found in many literatures in empirical finance and financial econometrics. Extensive introduction and investigation could be found in standard textbooks, such as [39] [59].

In financial econometrics, the focus of the studies is aiming at providing better estimation of volatility. On the other hand, in real practice, the modeling and prediction for such type of data are of more importance when people want to develop some useful trading strategies.

However, the modeling and prediction of intraday data are far more difficult. This is reflected by the fact that there exist only a few literature dealing with the modeling issues. The main focus of our research is to model the intraday transaction data. Before we study the new model, it will be helpful to review some existing models for the ultra-high frequency data, which is covered in next chapter.

Finally, notice that apart from some common features of the dataset, there exist many types of ultra-high frequency data. One important type is the ultra-high frequency data from the foreign exchanges (FX) market. In recent years, transaction data is also available in some derivatives markets. In our research, we focus on the transaction data from stock market only.

1.2 Outline

We outline the structure of the dissertation. In Chapter 2, we review some of the most successful models for ultra-high frequency transaction data.

In Chapter 3, we generalize the (univariate) Mixture Transition Distribution (MTD) model [47] and the Logistic Mixture Autoregressive model [70] to a new time series model, the Multi-logit Mixture Autoregressive (MLMAR) model. This new class of model is specifically built to model the high frequency intraday stock prices. We study the statistical properties of the new model, solve the estimation problem via an ECM algorithm, and investigate the prediction performance of the new model for the IBM stock intraday data.

In Chapter 4, we propose a new modeling framework for marked point processes in general and ultra-high frequency data in particular. The new time series model, i.e., the multivariate Mixture Transition Distribution model, is the multivariate extension of the MTD-type models and the bivariate MTD (BMTD) model [40]. In this new class of model, we discuss series of statistical modeling problems and use it to model the ultra-high frequency stock transaction data. We show that the new model outperform the benchmark BMTD model in capturing the underlying data generating processes.

In Chapter 5, we conclude by discussing some possible directions, existing issues and open problems.

Chapter 2

A Review of Models for Ultra-high Frequency Data

In this chapter, we discuss a few important models for ultra-high frequency financial data. It aims at providing a quick overview of the past development of the modeling of ultra-high frequency data. After the proposal of each benchmark model, there exist numerous follow-up and extended models. Such literature is so extensive that we are not able to list all of them. Instead, to illustrate the idea of modeling ultra-high frequency data, we focus on introducing a few benchmark models in this chapter.

2.1 ACD model

Engle and Russell [26] proposed a benchmark Autoregressive Conditional Duration (ACD) model for financial transaction data. Since then, it has been widely used and became a successful model. This also opens a door for the modern studies of financial transaction data. In the past decade, we have seen extensive development of new models under the general ACD modeling framework. A series of extended, modified or combined models have been actively explored and studied. Some of these models could be found in [5]. Now we introduce the benchmark models, following the introduction and notations as in [50].

Let t_i be the time stamp that a stock transaction is made. Denote $x_i = t_i - t_{i-1}$ to be the interval between two successive arrived transactions. We call x_i's *durations*. If we define Ψ_i to be the conditional expectation of the i^{th} duration, then we have the ACD model as follows.

Definition 2.1 *(Autoregressive conditional duration (ACD) model) The class of ACD*(q, p) *model for transaction durations* $\{x_i\}$ *is specified by*

$$x_i = \Psi_i \epsilon_i \tag{2.1}$$

$$\Psi_i = g(x_{i-1}, \cdots, x_{i-q}, \Psi_{i-1}, \cdots, \Psi_{i-p}) = \omega + \sum_{j=1}^{q} \alpha_j x_{i-j} + \sum_{j=1}^{p} \beta_j \Psi_{i-j} \tag{2.2}$$

where $\{\epsilon_i\}$*'s are i.i.d. sequences with* $E(\epsilon_i) = 1$*, and* $\omega > 0$*,* $\alpha_j, \beta_j \geq 0$*.*

To obtain the likelihood for ACD models, one should further specify the distribution of the error term $\{\epsilon_i\}$. Different specifications of the error distribution result in different classes of ACD-type models. Here we look at two simple ACD model as described in [26] [50]. The first one is the ACD model with exponential errors.

Example 2.1 *(Exponential ACD) A simple EACD*$(1, 1)$ *model could be written as*

$$x_i = \Psi_i \epsilon_i \tag{2.3}$$

$$\Psi_i = \omega + \alpha x_{i-1} + \beta \Psi_{i-1} \tag{2.4}$$

where $\{\epsilon_i\}$'s are i.i.d. exponentially distributed sequences with $E(\epsilon_i) = 1$, and $\omega > 0$, $\alpha, \beta \geq 0$.

The second example is the ACD model with Weibull errors, which is called Weibull ACD (WACD) model.

Example 2.2 *(Weibull ACD) A simple WACD$(1,1)$ model is given by*

$$x_i = \Psi_i \epsilon_i \tag{2.5}$$

$$\Psi_i = \omega + \alpha x_{i-1} + \beta \Psi_{i-1} \tag{2.6}$$

where $\{\epsilon_i\}$'s are i.i.d. Weibull$\left(\left(\Gamma(1+\frac{1}{\gamma}\right)^{-\gamma}, \gamma\right)$ distributed sequences, and $\omega > 0$, $\alpha, \beta \geq 0$.

Given any assumed distributions of the errors, the estimation of the model could be carried out by maximizing the log-likelihood function

$$l(\boldsymbol{\theta}; \mathbf{x}) = \sum_{i=1}^{n} \log f(x_i | \mathcal{I}_{i-1}; \boldsymbol{\theta}) \tag{2.7}$$

where $f(x_i | \mathcal{I}_{i-1}; \boldsymbol{\theta})$ is the conditional density for the duration x_i given the past; $\mathcal{I}_{i-1}$ is the past information up to time $i-1$ and $\boldsymbol{\theta}$ is the model parameter as given in (2.2). In other words, we obtain the estimates as

$$\hat{\boldsymbol{\theta}} = \arg\max_{\boldsymbol{\theta}} \sum_{i=1}^{n} \log f(x_i | \mathcal{I}_{i-1}; \boldsymbol{\theta}) \tag{2.8}$$

This could be easily done for simple cases, such as the EACD$(1,1)$ model (Example 2.1). For many other cases, one should generally turn to numerical optimizations.

Many extended and modified ACD type models have been developed and studied thereafter. Details of those models and related studies could be found in [5] [6] [22] [25] [48] [49] [50] [73].

2.2 Marked DSPP model

Another important modeling framework for stock transaction data is the class of marked doubly stochastic Poisson process (marked DSPP) model proposed by Rydberg and Shephard [63]. Essentially, the prices $\{P(t)\}$ evolve as follows.

$$P(t) = P(0) + \sum_{i=1}^{N(t)} Z_i \tag{2.9}$$

Here, $P(t)$ is the stock price at time t; $P(0)$ is the starting price; $N(t)$ is a counting process for the transaction arrivals; Z_i is the tick return (i.e., the price difference between two successive transactions). In fact, this pure jump process completely describes the price dynamics.

One drawback of this modeling framework is that a lot of (strong) assumptions are needed to build the model. For example, in [63] a simple MA(1) structure is assumed for tick return $\{Z_i\}$'s. Another example of the marked DSPP model proposed for ultra-high frequency financial data is by Centanni and Minozzo [16]. We refer to [17] for a detailed description of the model.

A key concern in proposing a specific marked DSPP is the tractability of the estimation schemes for the unobserved stochastic intensity. In Rydberg and Shephard's model [63], a particle filtering method is proposed to solve the estimation problem. While in Centanni and Minozzo's model [17], a reversible jump MCMC (RJMCMC) scheme is suggested based on the formulation of the particular type of intensity process used.

On the other hand, if our interest is in modeling the financial durations, this turns out to be a pure point process modeling problem. In fact, various types of point process models have been proposed for the transaction data, such as the self-exciting point process model and others [8]. Essentially, a successful model mostly relies on a proper type of the intensity process. Some related studies can be found in [5] [8].

2.3 Bivariate MTD model

Almost at the same time as the proposal of Centanni and Minozzo's model [17], Hassan and Lii [40] proposed a new framework for modeling more general marked point processes, with an application to model the transaction data, in particular, the durations and the volumes. In Hassan and Lii's bivariate Mixture Transition Distribution (BMTD) modeling framework, one significant advantage over the marked DSPP model is that in the BMTD model, no assumption of the independence between the marks and points is imposed. In fact, one of the key features for the BMTD model is its ability to model the dependence between marks and points.

This is a benchmark model that we will investigate in more details in later chapter, because the BMTD model is where we begin our research. For completeness, we describe it here briefly.

Definition 2.2 *(Bivariate mixture transition distribution (BMTD) model) A bivariate process* $\{(X_t, Y_t), t \in \mathbf{Z}\}$ *is generated by a BMTD model if the conditional distribution of* (X_t, Y_t) *given the past, evaluated at* (x_t, y_t)*, is given by*

$$F(x_t, y_t | x^{t-1}, y^{t-1}) = \sum_{j=1}^{p} \alpha_j F_j(x_t, y_t | x^{t-1}, y^{t-1}) \tag{2.10}$$

where $\sum_{j=1}^{p} \alpha_j = 1, \alpha_j > 0, j = 1, \cdots, p$*;* $F_j(x_t, y_t|x^{t-1}, y^{t-1})$ *is the conditional bivariate cumulative distribution function of* (X_t, Y_t) *given* $(X^{t-1}, Y^{t-1}) = (x^{t-1}, y^{t-1}) = ((x_1, y_1), \cdots, (x_{t-1}, y_{t-1}))$*, which is the past information up to time* $t - 1$.

The dependence structure of the marks and points relies on the particular class of bivariate distributions. For example, in modeling the transaction durations and volumes, a bivariate distribution with Gamma and Pareto marginals is used. We describe this case as follows.

In particular, let the bivariate sequences $\{(x_t, y_t)\}$ be the transaction durations and volumes. Then a BMTD model for this is given by

$$f(x_t, y_t|x^{t-1}, y^{t-1}) = \sum_{j=1}^{p} \alpha_j f_j(x_t, y_t|x^{t-1}, y^{t-1}) \tag{2.11}$$

where f_j takes the form

$$f_j(x_t, y_t|x^{t-1}, y^{t-1}) = \frac{x_t^{\gamma} e^{-x_t(1/\eta_j + y_t/\beta_j)}}{\eta_j^{\gamma}\Gamma(\gamma)\beta_j} \tag{2.12}$$

with built-in lag information given by

$$\eta_j = \theta_j x_{t-j} e^{-y_{t-j}}, \quad \eta_j, \theta_j, \beta_j > 0, \quad j = 1, \cdots, p \tag{2.13}$$

The BMTD model successfully captures some features like bursts, outliers and jumps in the ultra-high frequency data. The model outperforms some benchmark models in terms of the prediction performance. Following this idea, in Chapter 4 we start from the BMTD model and build up a more flexible modeling framework. Then, a series of related statistical modeling and forecasting issues are studied.

At this point, we shift to discuss another important aspect of the studies of ultra-high frequency data. It is on the volatility estimation from the ultra-high frequency data.

2.4 Volatility estimation from ultra-high frequency data

Combining the ACD model with a GARCH model for prices, Engle [25] proposed an ultra-high frequency measure of volatility. In the past decades, the volatility estimation from the ultra-high frequency data has been a hot topic in financial econometrics. In our studies, we will not focus on volatility estimation or forecasting using ultra-high frequency data. Instead of providing complete references, we mention only some seminal papers, i.e., some pioneer works done by Engle [25], Barndorff-Neilsen and Shephard [2] [3], etc.

At the end of this chapter, we make some remarks as follows.

Firstly, although the transaction dataset looks “simple” (at least from the first sight), the modeling of such data is extremely difficult and still very challenging so far. This should not be a surprise, since the transaction data are generally driven by very complicated background, such as economics situations, policies and the release of news, etc. In other words, they are driven by so many factors. In our research, we approach to such data set and propose statistical models for the transaction data at two different scales - the data at transaction level and the data at high frequency level.

Secondly, in modeling high frequency data or ultra-high frequency data, we adopt the “let data speak for themselves” approach. In other words, we observe the data and propose new statistical time series model to capture the underlying data generating processes. If the model successfully capture the features in the data, we hope to obtain much better prediction performance. Econometrics models generally consider more information or factors, thus the models may have intuitive meanings. However, the statistical time series modeling approach does not need to impose too much (economics) assumptions. Thus the statistical modeling is, in some sense, a "data-oriented" approach.

Lastly, one of the key motivations in modeling transaction data is to obtain better prediction performance. This is very useful in practice. Thus, whenever a new model is proposed, we investigate the model's prediction performance and compare them with some benchmark models.

Chapter 3

Multi-Logit Mixture Autoregressive Processes

In this chapter, we propose a new class of non-linear non-Gaussian time series model, the *Multi-logit Mixture Autoregressive (MLMAR)* model. We start with reviewing some key concept in time series analysis and some well known benchmark linear and non-linear time series models, which include the famous ARCH model [24] and Raftery's MTD model [62]. In Section 3.2, we propose the MLMAR model after a careful discussion of the motivations and modeling concerns for the high frequency intraday stock prices. In Section 3.3, we study the statistical properties of the new model and propose an ECM algorithm to solve the estimation problem. The prediction methodology and model selection criteria are also discussed. In Section 3.4, we fit the MLMAR model to the IBM intraday stock transaction data and compare its prediction performance with some benchmark models. In Section 3.5, as an interesting application of the new model to real trading practice, we develop a simple trading algorithm based on the forecasting capacity of the MLMAR model. Last, we conclude by pointing out some possible extensions and potential statistical modeling problems.

3.1 Introduction

In this section, we review some fundamental concept in time series analysis. The classical linear and non-linear time series models are briefly introduced thereafter. In particular, we take a close look at the Raftery's MTD model [62] and other MTD-type models, since our new time series model is also a specific MTD-type model.

3.1.1 ARIMA(p, d, q) processes

Since the proposal of the Box-Jenkins [12] approach in 1970s, linear ARIMA time series model has been a popular tool for analyzing most time series data. A modern account of this classic topic could be found in [14] or [37]. Notice that we interchange the notation of *processes* and *time series* frequently. The notation *time series* is used when a process is observed on discrete time index, while the notation *processes* is particularly reserved for continuous-time processes. When introducing the ARIMA model and related fundamental concepts in time series analysis, we follow the definitions and the notations as in [14].

The concept of stationarity plays an important role in linear time series analysis.

Definition 3.1 *(Autocovariance function (ACF)) If* $\{X_t, t \in T\}$ *is a processes such that* $Var(X_t) < \infty, \forall t \in T$*, then the autocovariance function* $\gamma_X(\cdot,\cdot)$ *of* $\{X_t, t \in T\}$ *is defined by*

$$\gamma_X(r,s) = Cov(X_r, X_s) = E\left[(X_r - EX_r)(X_s - EX_s)\right], \quad r,s \in T \tag{3.1}$$

Definition 3.2 *(Strict stationarity) A time series* $\{X_t, t \in \mathbf{Z}\}$ *is said to be strictly stationary if the joint distribution of* $(X_{t_1}, \cdots, X_{t_k})'$ *and* $(X_{t_1+h}, \cdots, X_{t_k+h})'$ *is the same, i.e.,*

$$(X_{t_1}, \cdots, X_{t_k})' \stackrel{d}{=} (X_{t_1+h}, \cdots, X_{t_k+h})', \quad \forall\{t_1, \cdots, t_k\}, k \in \mathbf{Z}^+, h \in \mathbf{Z} \tag{3.2}$$

However, the strict stationarity is a quite strong assumption. It is useful only in some theoretical studies. Instead, a weak form of stationarity, i.e., 2^{nd}-order stationarity, or covariance stationarity, is widely used and practically useful. If there is no ambiguity from the context, when a time series is said to be stationary, it generally refers to 2^{nd}-order stationarity.

Definition 3.3 *(2^{nd}-order stationarity) A time series* $\{X_t, t \in \mathbf{Z}\}$ *is said to be* 2^{nd}*-order stationary if*

$$E|X_t|^2 < \infty, \forall t \in \mathbf{Z} \tag{3.3a}$$

$$EX_t = m, \forall t \in \mathbf{Z} \tag{3.3b}$$

$$\gamma_X(r,s) = \gamma_X(r+t, s+t), \forall r,s,t \in \mathbf{Z} \tag{3.3c}$$

With the stationary concept in mind, we are ready to define the benchmark Autoregressive Moving Average (ARMA) time series model.

Definition 3.4 *(ARMA(p, q) process) A process $\{X_t, t \in \mathbf{Z}\}$ is said to be an ARMA(p, q) process if $\{X_t\}$ is stationary and*

$$X_t - \phi_1 X_{t-1} - \cdots - \phi_p X_{t-p} = Z_t + \theta_1 Z_{t-1} + \cdots + \theta_q Z_{t-q} \tag{3.4}$$

where $\{Z_t\}$ is white noise, i.e., $\{Z_t\} \sim WN(0, \sigma^2)$.

If we define the backward shift operator B such that $BX_t = X_{t-1}$ and $B^j(X_t) = X_{t-j}$, a more compact form for the ARMA(p, q) processes is given by

$$\phi(B)X_t = \theta(B)Z_t \tag{3.5}$$

where $\phi(z) = 1 - \phi_1 z - \cdots - \phi_p z^p$ and $\theta(z) = 1 + \theta_1 z + \cdots + \theta_q z^q$.

Two special classes of the ARMA(p, q) processes are the AR(p) and MA(q) processes:

Example 3.1 *(AR(p) process) If $\theta(z) \equiv 1$, then*

$$X_t - \phi_1 X_{t-1} - \cdots - \phi_p X_{t-p} = Z_t \tag{3.6}$$

is said to be an autoregressive (AR) process with order p.

Example 3.2 *(MA(q) process) If $\phi(z) \equiv 1$, then*

$$X_t = Z_t + \theta_1 Z_{t-1} + \cdots + \theta_q Z_{t-q} \tag{3.7}$$

is said to be a moving average (MA) process with order q.

The ARMA processes could be generalized to a class of important non-stationary processes, the Autoregressive-Integrated Moving Average (ARIMA) processes. It is defined as follows.

Definition 3.5 *(ARIMA(p, d, q) process) If d is a non-negative integer, then $\{X_t\}$ is said to be an ARIMA(p, d, q) process if $Y_t := (1 - B)^d X_t$ is a causal ARMA(p, q) process, i.e., $\{X_t\}$*

satisfies a differential equation of the form

$$\phi^*(B)X_t \equiv \phi(B)(1-B)^d X_t = \theta(B)Z_t, \quad \{Z_t\} \sim WN(0,\sigma^2) \tag{3.8}$$

For more details and a systematic treatment of the ARMA(p,q) and ARIMA(p,d,q) processes, we refer to standard text [14] or [37].

As the most widely accepted and used tools in modeling time series, the ARMA models have their own advantages, such as their relatively simple tractability, well developed estimation schemes and diagnostic tools, etc. Most importantly, the widely existing statistical softwares to implement the ARMA models greatly help maintaining the models' popularity. However, the ARMA models may fail to capture some important features in many real time series data, such as the conditional heteroscedasticity, the multi-modality in the conditional distribution and the regime switching behaviors, etc. These features mentioned are frequently observed in various types of financial time series. All these call for the innovations of nonlinear time series models. We will then review some of the popular nonlinear time series models.

3.1.2 Classical nonlinear time series models

In principle, as described in [66], the innovations of coming up new nonlinear time series models are *infinite*. In this section, we review some of the most important nonlinear time series models in history. We introduce these benchmark nonlinear time series models, following similar orders or notations as in [1] [28] [61] [66].

The first important class of nonlinear time series models is the Autoregressive Conditional Heteroscedastic (ARCH) model proposed by Engle [24]. This is undoubtedly the most important model in financial econometrics. It achieves huge success in modeling and forecasting volatility of financial time series. In recognition of the invention of the ARCH model, Engle was awarded the Nobel prize in Economics in 2003. Here, instead of using the original notation and definition, we follow the descriptions and notations as in [61].

Definition 3.6 *(ARCH(r) model) An autoregressive conditional heteroscedastic process of order r, i.e., ARCH(r), is defined as*

$$z_t = \sqrt{h_t}\epsilon_t, \quad h_t = \alpha_0 + \alpha_1 z_{t-1}^2 + \cdots + \alpha_r z_{t-r}^2 \tag{3.9}$$

where $\{\epsilon_t\} \sim IID(0,1)$, $\alpha_0 > 0$, $\alpha_i > 0$ $(i = 1, 2, \cdots, r)$.

In practice, the $\{\epsilon_t\}$'s are often specified to be standard Gaussian or Student-t distributed. The ARCH model keeps the stationarity but is able to capture the time-varying conditional variances. For the most simple ARCH(1) model, if the fourth moment of $\{z_t\}$ exists, the unconditional kurtosis of $\{z_t\}$ is given by

$$\frac{z_t^4}{[Var(z_t)]^2} = 3\frac{1-\alpha_1^2}{1-3\alpha_1^2} > 3 \tag{3.10}$$

Thus the ARCH model is able to capture the fat-tailed behavior which is widely observed in financial return series. Above properties still hold for general ARCH(r) models, although the formulations are much more complicated. Many other specifications of the conditional variances have been proposed since the first ARCH model. Among these, the Generalized ARCH (GARCH) model by Bollerslev [11] is the one that is widely used in practice.

Definition 3.7 *(GARCH(r, s) model) A process $\{z_t\}$ is a pure Generalized ARCH, GARCH(r, s), process if $\mu_t = E(z_t|F_{t-1}) = 0$ and*

$$z_t = \sqrt{h_t}\epsilon_t, \quad h_t = \alpha_0 + \sum_{i=1}^{r} \alpha_i z_{t-i}^2 + \sum_{j=1}^{s} \beta_j h_{t-j} \tag{3.11}$$

where $\{\epsilon_t\} \sim IID(0, 1)$, $\alpha_0 > 0$, $\alpha_i \geq 0$ $(i = 1, 2, \cdots, r)$; $\beta_j \geq 0$ $(j = 1, 2, \cdots, s)$; $\sum_{i=1}^{\max(r,s)}(\alpha_i + \beta_i) < 1$.

The GARCH models are able to capture the long run effect of the shocks. The ARCH(r) models that are used to model the volatility of asset returns are generally obtained with large order r (for example, $r = 8, 9$ or higher). However, the GARCH(r, s) models are able to capture the volatility process with much fewer parameters. It is not easy to identify of the orders of GARCH models. Thus, one generally fits the GARCH models with lower orders.

Autoregressive models have been extended in various ways to handle different types of nonlinear time series data. As emphasized in [66], in principle, the linear AR model can be generalized to a broader class of models, the *nonlinear autoregressive of order k with general noise* model [66], if there exists a mapping $f : \mathbf{R}^{k+1} \mapsto \mathbf{R}$ and

$$X_t = f(X_{t-1}, \cdots, X_{t-k}, \epsilon_t), \quad t \in \mathbf{Z} \tag{3.12}$$

where $\{\epsilon_t\}$ is a sequence of i.i.d. random variables. Further, it may assume that ϵ_t is also independent of $\{X_s, s < t\}$. These models could be further generalized by including exogenous variables up to time t. In what follows, we only briefly mention a few represented nonlinear time series models from these generalizations.

Definition 3.8 *(Self-exciting threshold autoregressive (SETAR) model [66]) A simple self-exciting threshold autoregressive, SETAR*$(l; k, \cdots, k)$ *model, where k is repeated l times, takes the form*

$$X_t = \phi_0^{(j)} + \sum_{i=1}^{k} \phi_i^{(j)} X_{t-i} + \epsilon_t^{(j)} \tag{3.13}$$

conditional on $X_{t-d} \in R_j$ $(j = 1, \cdots, l)$*, where d is the delay lag and* $\{R_j, j = 1, \cdots, l\}$ *forms a partition of* $\mathbf{R}$.

The model is proposed by Tong and further developed into a rich class of nonlinear models [65] [66] [67]. A further generation of the idea of the threshold AR model would be considering certain types of *smooth* transitions between these regimes. These are the Smooth Transition Regression (STR) models. The following Exponential Autoregressive (EXPAR) model is one special example.

Definition 3.9 *(Exponential autoregressive (EXPAR) model [1] [36]) A simple EXPAR model takes the form*

$$X_t = \sum_{i=1}^{p} \left\{ a_i + b_i \exp(-\gamma X_{t-1}^2) \right\} X_{t-i} + \epsilon_t \tag{3.14}$$

with $\gamma > 0$*,* a_i, b_i $(i = 1, \cdots, p)$ *are the coefficients.*

This is the model proposed in [36]. Notice that those STR models belong to a broader class of models, the Random Coefficient Autoregressive (RCA) models [57].

As we mention before, the innovations of coming up new nonlinear time series models are *infinite*. At the end of this subsection, we mention the class of bilinear models [33], following the notations as in [1]. For other nonlinear time series models, we refer to those classical works, such as [1] [28] [30] [66].

Definition 3.10 *(Bilinear model) A bilinear model take the form*

$$X_t = \sum_{i=1}^{p} \phi_i X_{t-i} + \sum_{j=1}^{q} \theta_j \epsilon_{t-j} + \sum_{i=1}^{P} \sum_{j=1}^{Q} \tau_{ij} X_{t-i} \epsilon_{t-j} + \epsilon_t \tag{3.15}$$

3.1.3 Raftery's MTD-type models

Now, we introduce a recent class of nonlinear non-Gaussian time series models, the MTD-type models. Since our new model is built upon the idea of MTD-type models, here we introduce most of the important MTD-type models developed so far.

Raftery [62] proposed a new model for high-order Markov Chain. The idea is that the conditional probability of observing $X_t = j_0$ given the past is a linear combination of the contributions from each of $X_{t-1}, \cdots, X_{t-l}$. More specifically, consider a random sequence $\{X_t \in \{1, 2, \cdots, m\}, t \in \mathbf{Z}\}$, the transition probability is given by

$$P(X_t = j_0 | X_{t-1} = j_1, \cdots, X_{t-l} = j_l) = \sum_{i=1}^{l} \lambda_i q_{j_0 j_i} \tag{3.16}$$

where $\sum_{i=1}^{l} \lambda_i = 1$ and $Q = \{q_{jk}\}$ is a non-negative $m \times m$ matrix with column sums equal to 1, and

$$0 \leq \sum_{i=1}^{l} \lambda_i q_{jk_i} \leq 1 \quad (j, k_1, \cdots, k_l = 1, \cdots, m). \tag{3.17}$$

The model can be generalized to model time series with continuous states. Following this idea, a series of important nonlinear non-Gaussian time series models have been proposed and studied in the past decade. The benchmark model is the Mixture Transition Distribution (MTD) model [47] described as follows.

Definition 3.11 *(Mixture transition distribution (MTD) model) A time series* $\{X_t, t \in \mathbf{Z}\}$ *is generated by a mixture transition distribution model if the conditional distribution of* X_t *given the past follows the form*

$$F(x_t|\mathcal{F}_{t-1}) = \sum_{i=1}^{p} \alpha_i G_i(x_t|x_{t-i}) \tag{3.18}$$

where $F(x_t|\mathcal{F}_{t-1})$ *is the conditional cumulative distribution function of* X_t *given the past, evaluated at* x_t*. Also,* $\sum_{i=1}^{p} \alpha_i = 1, \alpha_i \geq 0$ $(i = 1, \cdots, p)$*.* $G_i(\cdot|x)$ *is the conditional distribution for each value of* x*.*

In [47], $G_i(\cdot|\cdot)$ is specified as Gaussian distribution.

The MTD model explicitly specifies the data generating processes. It can be used to model a variety of nonlinear non-Gaussian features in many time series, such as the flat stretches, jumps and outliers. The following example shows a simple MTD model and its sample path.

Example 3.3 *(MTD model) A simple MTD model with* 2 *mixture components, up to 2-lags and constant variance for each mixture, could be specified by*

$$\begin{aligned} f(x_t|\mathcal{F}_{t-1}) &= \alpha_1\phi_1(x_t|x_{t-1}) + \alpha_2\phi_2(x_t|x_{t-2}) \\ &= 0.60\frac{1}{\sqrt{2\pi}\cdot 0.5}\exp\left(\frac{x_t - 0.88x_{t-1}}{2\cdot 0.5^2}\right) + 0.40\frac{1}{\sqrt{2\pi}\cdot 2}\exp\left(\frac{x_t - 0.94x_{t-2}}{2\cdot 2^2}\right) \end{aligned}$$

A simulated sample path from this specific example is illustrated in Figure 3.1.

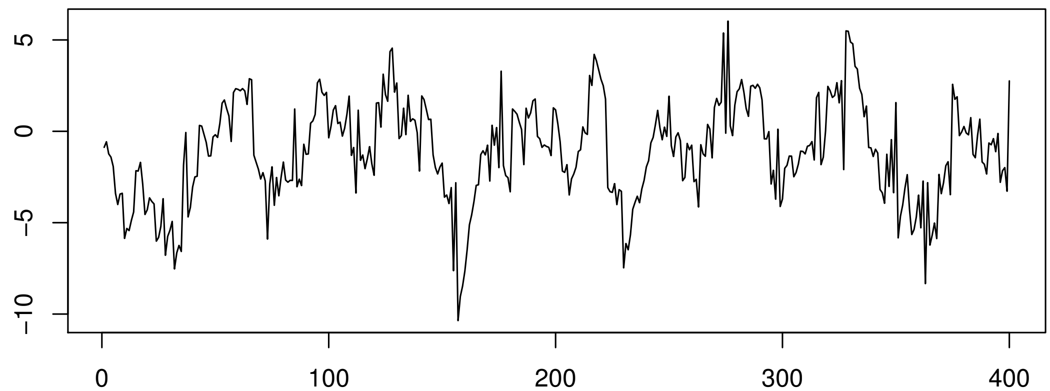

Figure 3.1: A simulated sample path from a MTD model.

Another important MTD-type model is the mixture autoregressive (MAR) model [71] described as follows.

Definition 3.12 *(Mixture autoregressive (MAR) model) A time series $\{X_t, t \in \mathbf{Z}\}$ is generated by a K-component MAR$(K; p_1, p_2, \cdots, p_K)$ model if*

$$F(x_t|\mathcal{F}_{t-1}) = \sum_{k=1}^{K} \alpha_k \Phi\left(\frac{x_t - \phi_{k0} - \phi_{k1}x_{t-1} - \cdots - \phi_{kp_k}x_{t-p_k}}{\sigma_k}\right) \tag{3.19}$$

where $F(x_t|\mathcal{F}_{t-1})$ is the conditional cumulative distribution function of X_t given the past, evaluated at x_t. $\Phi(\cdot)$ is the cumulative distribution function of standard Gaussian distribution and $\sum_{k=1}^{K} \alpha_k = 1, \alpha_k \geq 0, k = 1, \cdots, K$.

The MAR model provides much wider range of shape changing predictive distributions than the MTD model. Moreover, Wong and Li shows that the MAR model has the ability to handle cycles and conditional heteroscedasticity.

Notice that both MTD and MAR models assume constant variances in each mixture (Gaussian) density. When releasing such constraint, a useful class of time series model that is able to model similar ARCH effect may be obtained. This is the model to be described below [72].

Definition 3.13 *(Mixture autoregressive conditional heteroscedastic (MAR-ARCH) model)A time series $\{X_t, t \in \mathbf{Z}\}$ is generated by a MAR-ARCH$(K; p_1, p_2, \cdots, p_K; q_1, q_2, \cdots, q_K)$ model if*

$$F(x_t|\mathcal{F}_{t-1}) = \sum_{k=1}^{K} \alpha_k \Phi\left(\frac{e_{k,t}}{\sqrt{h_{k,t}}}\right) \tag{3.20}$$

with

$$e_{k,t} = x_t - \phi_{k0} - \phi_{k1}x_{t-1} - \cdots - \phi_{kp_k}x_{t-p_k}, \quad h_{k,t} = \beta_{k0} + \beta_{k1}e_{k,t-1}^2 + \cdots + \beta_{kq_k}e_{k,t-q_k}^2 \tag{3.21}$$

where $F(x_t|\mathcal{F}_{t-1})$ is the conditional cumulative distribution function of X_t given the past information $\mathcal{F}_{t-1}$, evaluated at x_t. $\Phi(\cdot)$ is the cumulative distribution function of standard Gaussian distribution and $\sum_{k=1}^{K} \alpha_k = 1, \alpha_k \geq 0$ $(k = 1, \cdots, K)$. To guarantee the non-negativity of conditional variance, $\beta_{k0} > 0$ $(k = 1, \cdots, K)$, and $\beta_{ki} \geq 0$ $(i = 1, \cdots, q_k; k = 1, \cdots, K)$.

Previous effort has been putting in parameterizing the conditional mean and conditional variance for each mixture component. There exist two other ways to generalize the idea to other MTD-type models. In what follows, we mention three models constructed by some new types of generalizations.

To model different types of time series data, it is necessary to switch to a certain distribution for the mixture components rather than staying with Gaussian distribution. For example, in modeling heavy tailed financial time series, Wong, Chan and Kam [69] show the great advantages by using Student-t distribution rather than Gaussian distribution to build a MTD-type model. Their model is described as follows.

Definition 3.14 *(Student t-mixture autoregressive (TMAR) model) A time series $\{X_t, t \in \mathbf{Z}\}$ is generated by a K-component TMAR$(K; p_1, p_2, \cdots, p_K)$ model if*

$$F(x_t|\mathcal{F}_{t-1}) = \sum_{k=1}^{K} \alpha_k F_{v_k}\left(\frac{x_t - \phi_{k0} - \phi_{k1}x_{t-1} - \cdots - \phi_{kp_k}x_{t-p_k}}{\sigma_k}\right) \tag{3.22}$$

where $\sum_{k=1}^{K} \alpha_k = 1, \alpha_k \geq 0$ $(k = 1, \cdots, K)$. $F(x_t|\mathcal{F}_{t-1})$ is the conditional cumulative distribution function of X_t given the past information, evaluated at x_t. $F_{v_k}(\cdot)$ is the cumulative distribution of the standardized Student t-distribution with v_k degrees of freedom.

With a Student-t distribution for each mixture component, the TMAR model is able to model the tail behavior of the conditional distributions.

Another type of generation is to consider taking the exogenous time series or variables into the model. For example, one can use exogenous variables to define the time changing weights of the mixtures. A model following this direction is described as follows [70].

Definition 3.15 *(Logistic mixture autoregressive (LMARX) model) The major time series of interest $\{Y_t, t \in \mathbf{Z}\}$, together with l time series of exogenous variables $\{X_{i,t}, i = 1, \cdots, l\}$ follows a LMARX model if*

$$F(y_t|\mathcal{F}_{t-1}, \Omega_t) = \sum_{k=1}^{2} \alpha_{k,t}\Phi\left(\frac{e_{k,t}}{\sigma_k}\right) \tag{3.23}$$

where

$$e_{k,t} = y_t - \mu_{k,t} = y_t - \phi_{k0} - \sum_{i=1}^{p_k} \phi_{ki}y_{t-i} - \sum_{i=1}^{l}\sum_{j=0}^{q_{ki}} \delta_{kij}x_{i,t-j} \tag{3.24}$$

$$\log(\alpha_{1,t}/\alpha_{2,t}) = \beta_0 + \beta_1 v_{1t} + \cdots + \beta_m v_{mt} \tag{3.25}$$

Here $F(y_t|\mathcal{F}_{t-1}, \Omega_t)$ *is the conditional c.d.f. of* Y_t *given the information in the sets* $\mathcal{F}_{t-1}$ *and* Ω_t *, evaluated at* y_t*;* $\mathcal{F}_{t-1} = \{y_{t-1}, y_{t-2}, \cdots\}$*;* $\Omega_t = \{x_{1,t}, x_{1,t-1}, \cdots; \cdots; x_{l,t}, x_{l,t-1}, \cdots\}$*;* $\Phi(\cdot)$ *is the cumulative distribution function of standard Gaussian distribution;* $\alpha_{k,t}$ *is the mixing proportion of the* k^{th} *component, with* $\alpha_{1,t} + \alpha_{2,t} = 1$*; and* $v_{it} \in \mathcal{F}_{t-1} \cup \Omega_t$ *for* $i = 1, \cdots, m$*.*

The LMARX model takes into account of the exogenous information in a way that exogenous variables control the time-varying weight for each mixture component. This also brings into rich predictive densities and may produce good prediction performance for some dataset.

All past developments of the MTD-type models have been focusing on univariate models. Hassan and Lii [40] generalize the MTD-type models to the bivariate situations, where their key motivation is to model marked point processes. We have described this model in Chapter 2. However, for completeness, we reproduce the BMTD model [40] here again.

Definition 3.16 *(Bivariate mixture transition distribution (BMTD) model) A bivariate process* $\{(X_t, Y_t), t \in \mathbf{Z}\}$ *is generated by the BMTD model if the conditional distribution of* (X_t, Y_t) *given the past, evaluated at* (x_t, y_t)*, can be written as*

$$F(x_t, y_t|x^{t-1}, y^{t-1}) = \sum_{j=1}^{p} \alpha_j F_j(x_t, y_t|x^{t-1}, y^{t-1}) \tag{3.26}$$

where $\sum_{j=1}^{p} \alpha_j = 1, \alpha_j > 0, j = 1, \cdots, p$*;* $F_j(x_t, y_t|x^{t-1}, y^{t-1})$ *is the conditional bivariate cumulative distribution function of* (X_t, Y_t) *given* $(X^{t-1}, Y^{t-1}) = (x^{t-1}, y^{t-1}) = ((x_1, y_1), \cdots, (x_{t-1}, y_{t-1}))$*, which is the past information up to time* $t - 1$*.*

The BMTD model is originally proposed to model marked point processes. Thus, the bivariate distributions proposed should have at least one non-negative marginal distribution. In fact, in [40] a specific class of bivariate distributions satisfying this constraint is proposed. For a bivariate random variable (X, Y), it follows the density form

$$f_{X,Y}(x,y) = Cx^{\delta+\gamma+1/\phi-1}\left|\frac{y-\mu}{\beta}\right|^{\delta} e^{-x^{\alpha}(\lambda+|y-\mu|^{\phi}/\beta^{\phi})}, \quad x > 0, -\infty < y < \infty \tag{3.27}$$

where α, ϕ, δ and γ are all positive shape parameters; β and λ are positive scale parameters; $\mu \in \mathbf{R}$ is a location parameter; C is the normalized constant.

3.2 A MLMAR model for intraday stock prices

After a brief review of the important MTD-type time series models, we now propose a new MTD-type model with direct motivation for the modeling of intraday high frequency stock prices. Thus, we will discuss the model always with stock prices data in mind. However, this is a more general time series modeling framework and it surely can be used to model many other types of time series met in practice.

In this section, we first discuss our motivation to propose the time series model with emphasis on empirical investigation of the intraday stock prices. After introducing the generalized Gaussian distribution (GGD), a key component of the new model, we propose the new MTD-type model and illustrate it with some examples.

3.2.1 Intraday prices and statistical modeling issues

Due to the fast development of information technology, the detailed information for *each* stock transaction can nowadays be fully recorded and released to public with some reasonable costs. One of such data sets is the daily *Trade and Quote (TAQ)* dataset from the New York Stock Exchange. In *TAQ* dataset, the time stamp for each transaction, the corresponding price, size, and description of the transaction conditions are recorded. For actively traded stocks, these result in a massive data set.

The easy access to the transaction data in real time cultivates an industry called *automatic* or *algorithmic* trading, where the trading signals and trading instructions are automatically generated by computers. As estimated in 2009, automatic trading accounts for about 73% of the all US equity trading volumes. It is undoubtedly that the success of the algorithmic trading relies mostly on the forecasting capacity from the statistical models used for the stock prices.

The tick-by-tick transaction data is the data set that we will model in next chapter. In this chapter, we want to model the high frequency data. Notice that there exists no strict definition of the high frequency data. Commonly, when the intraday data is sampled at time interval less than 1 minute, the resulted data will be referred as *high or very high frequency* data. This is an important playground for automatic trading and financial econometrics [9], [23], [31]. Figure 3.2 shows the sample paths of intraday IBM stock prices sampled at every 20 or 30 seconds. Although two paths seem to be similar, in financial studies they are treated differently.

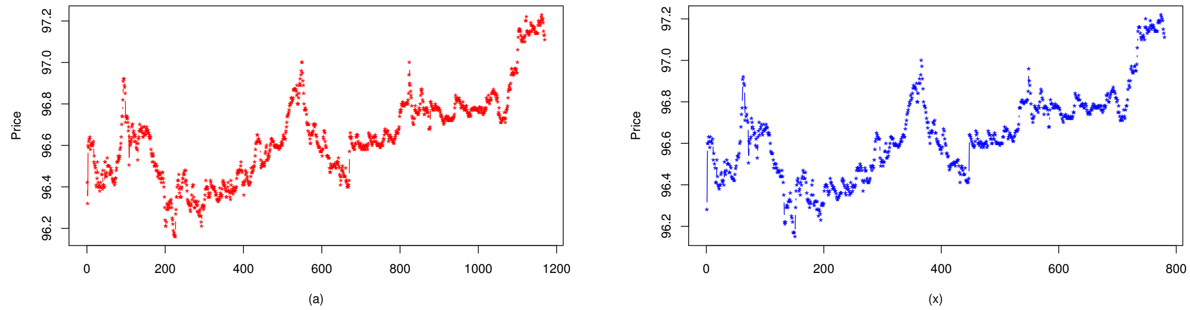

Figure 3.2: Intraday IBM stock prices (9:30AM - 4:00PM, 01/29/2007), sampled at every (a) 20 seconds and (b) 30 seconds.

Traditional market microstructure theory claims that the market *variables* can generally be categorized into several distinct groups. For example, in option market, the participants could be either hedgers, speculators or arbitragers; in stock market, traders can be informed traders and uninformed traders. Participants from different groups utilize different strategies and (should) illustrate different behaviors. Thus, a natural conjecture is that, when these reflect into the market data, the mixture of distributions models would generally obtain a better fit for the market data in many situations.

A review of the mixture of distributions model (MODM) in market microstructure theory can be found in [32]. Actually, in transaction data analysis, the mixture models are very successful in modeling i.i.d. returns [38]. Recently, within the ACD-type modeling framework, different mixture ACD models, such as [41] [48] [50] and the time varying mixture ACD model [49] obtain better fits for many different transaction data sets.

However, these models mainly focus on modeling the transaction durations. In this chapter, in order to build a model for high frequency stock prices, we target at constructing a

MTD-type model and, at the same time, taking into account of the other available information accompanied with the price series. These helpful information could be the transaction volumes, transaction frequencies and durations, etc. Above discussion would greatly help in understanding our new model.

The MLMAR model we propose for modeling high frequency data is a two-step further generalization of Wong and Li's Logistic Mixture Autoregressive (LMARX) [70] model, which consists of a mixture of *(only) two Gaussian* transition functions and allows the mixture proportions to change over time. The LMARX model has been applied to model the river flow data and the Canadian lynx data. However, in high frequency financial modeling, this model is not an ideal candidate. The reasons are given as follows.

Firstly, a model with two mixture components is not sufficient. When the market is very volatile, the conditional distribution of the current observation given the past may have three or even more modes. Figure 3.3 and Figure 3.4 illustrate some of the multimodal conditional density of the high frequency prices when conditioning on one-lag. As the sampling interval increases, the multi-modality of conditional density seems more common and significant. Thus, an extension of the LMARX model to allow *arbitrary k* time-varying mixture components will be well rewarded.

Secondly, it is a widely accepted fact from empirical studies that the discreteness of price movements always induces a high degree of kurtosis for the high frequency data [27]. Figure 3.3 and 3.4 show some signs that the conditional distributions have much higher excess of kurtosis, i.e., some densities have very sharp "peaks". Thus, a better idea (within the MTD-

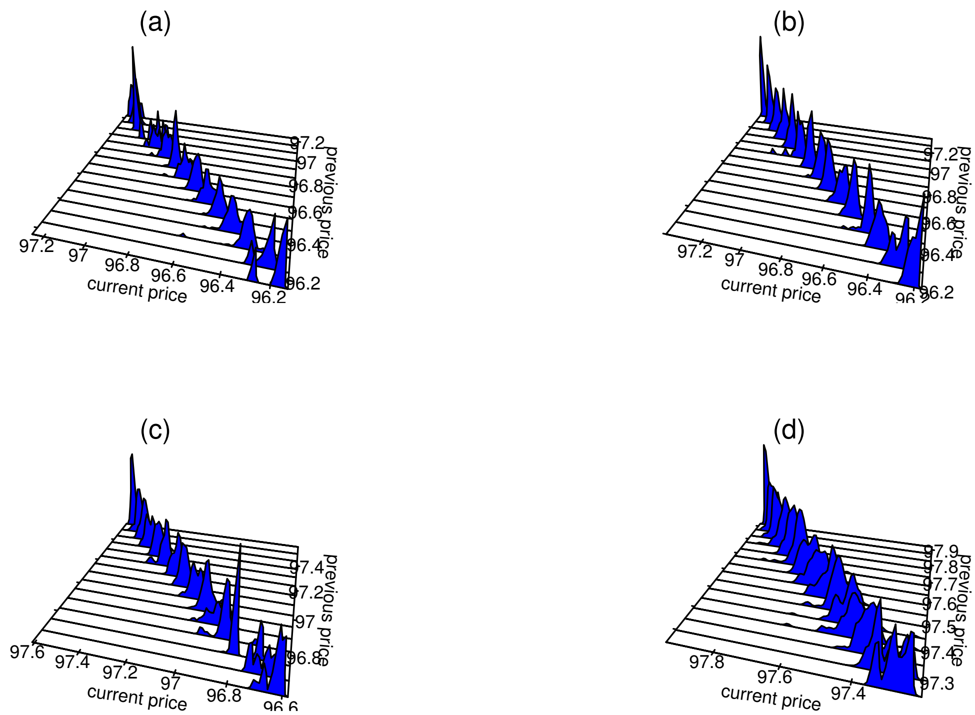

Figure 3.3: Conditional density of the intraday price. (a-d) for dates 01/22/2007 - 01/25/2007. (a)-(b) are sampled every 20 seconds; (c)-(d) are sampled every 30 seconds. The conditional density estimation and the optimal bandwidth chosen are via [4].

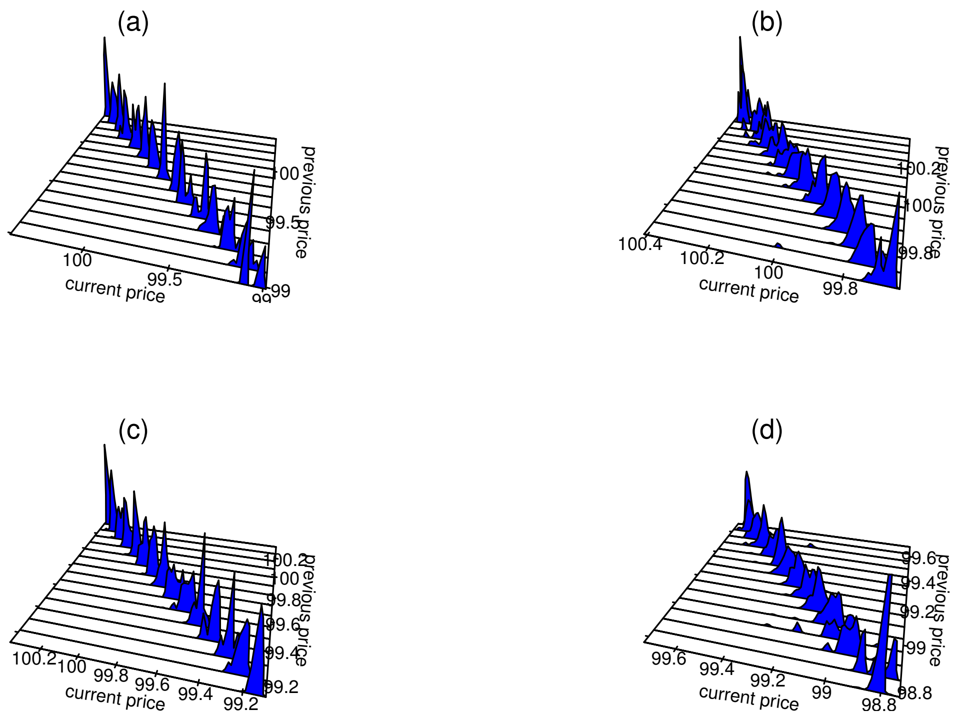

Figure 3.4: Conditional density of the intraday price. (a-d) for dates 02/05/2007 - 02/08/2007. (a)-(b) are sampled every 20 seconds; (c)-(d) are sampled every 30 seconds. The conditional density estimation and the optimal bandwidth chosen are via [4].

type modeling framework) is to select alternative types of distributions that are capable to capture such distributional features, instead of using the Gaussian distribution.

The generalized Gaussian distribution (GGD) turns out to be a good candidate for this. One of its nice properties is that beyond providing much greater flexibility (leptokurticity and platykurticity), the estimation for the resulted time series model still remains tractable, which we will see later. We will also see that such kind of model building could capture most of those conditional densities illustrated in Figure 3.3 and Figure 3.4.

Other models considering using alternative distributions could be found in [69], where Student t-mixture is used for modeling heavy tailed time series; and [40] , where a mixture of bivariate distributions are studied for tick-by-tick transaction data. Before introducing the new model, it is well deserved to take a look at an important distribution, i.e., the generalized Gaussian distribution (GGD).

3.2.2 Generalized Gaussian distribution (GGD)

The generalized Gaussian distribution (GGD) [53] has been used intensively in electronical engineering and computer science. In particular, it is a powerful tool in modeling non-Gaussian noise, where most of its successful applications are in signal processing and image processing [18] [21] [44]. It is a key building component in our new model. We briefly introduce some of its properties in this subsection. Here we follow the definition as in [53].

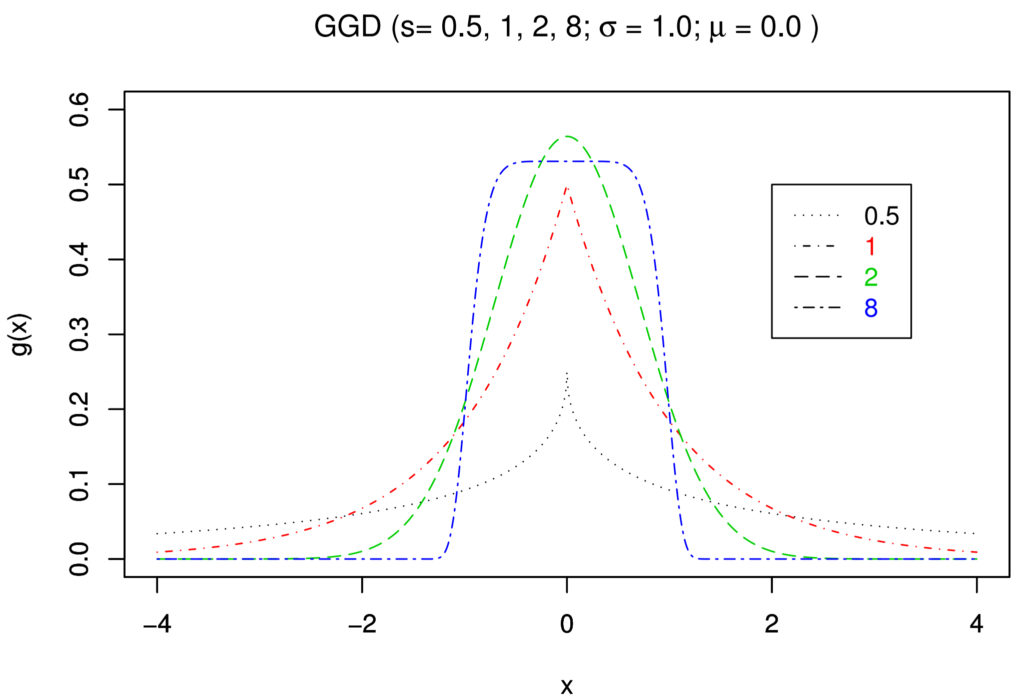

Figure 3.5: Generalized Gaussian density (GGD).

Definition 3.17 *(Generalized Gaussian distribution (GGD) [53]) A random variable X has a generalized Gaussian distribution (GGD) if its density has the form*

$$\tilde{\phi}(x;\mu,\sigma,s) = \frac{s}{2\sigma\,\Gamma(1/s)} \exp\left\{-\left|\frac{x-\mu}{\sigma}\right|^s\right\} \tag{3.28}$$

where μ, σ and s are the location, scale and shape parameter respectively. The density is symmetric at μ. By varying the shape parameter, we may obtain the Gaussian distribution and the Laplace distribution as its special cases. Some of the densities obtained by varying the shape parameter is plotted in Figure 3.5. We take a look at two examples as follows.

Example 3.4 *(Gaussian distribution) From the GGD, if we set $s = 2$, the density becomes*

$$\tilde{\phi}(x;\mu,\sigma) = \frac{1}{\sqrt{\pi}\sigma} \exp\left\{-\left|\frac{x-\mu}{\sigma}\right|^2\right\} \tag{3.29}$$

which is Gaussian distribution with mean μ and variance $\frac{\sigma^2}{2}$.

Example 3.5 *(Laplace distribution) From the GGD, if we set $s = 1$, the density becomes*

$$\tilde{\phi}(x;\mu,\sigma) = \frac{1}{2\sigma} \exp\left\{-\left|\frac{x-\mu}{\sigma}\right|\right\} \tag{3.30}$$

which is Laplace distribution with location parameter μ and scale parameter σ.

We quote two results on the moment properties of GGD from [53].

Proposition 3.1 *(Moments) If a random variable follows the generalized Gaussian distribution $\tilde{\phi}(x;\mu,\sigma,s)$, the n^{th}-order moments are given by*

$$E(X^n) = \frac{\mu^n \sum_{k=0}^{n} \left\{C_n^k(\frac{\sigma}{\mu})^k[1+(-1)^k]\Gamma((k+1)/s)\right\}}{2\Gamma(1/s)} \tag{3.31}$$

In particular,

$$E(X) = \mu; \quad E(X^2) = \mu^2 + \frac{\sigma^2\Gamma(3/s)}{\Gamma(1/s)}; \quad E(X^3) = \mu^3 + \frac{3\mu\sigma^2\Gamma(3/s)}{\Gamma(1/s)} \tag{3.32}$$

$$E(X^4) = \mu^4 + \frac{6\mu^2\sigma^2\Gamma(3/s)}{\Gamma(1/s)} + \frac{\sigma^4\Gamma(5/s)}{\Gamma(1/s)} \tag{3.33}$$

Proposition 3.2 *(Central Moments) If a random variable follows the generalized Gaussian distribution $\tilde{\phi}(x;\mu,\sigma,s)$, the n^{th}-order central moments are given by*

$$E\left[(X-\mu)^n\right] = \frac{\sigma^n\{1+(-1)^n\}\Gamma((n+1)/s)}{2\Gamma(1/s)} \tag{3.34}$$

In particular,

$$Var(X) = \frac{\sigma^2\Gamma(3/s)}{\Gamma(1/s)};\quad E\left[(X-\mu)^3\right] = 0;\quad E\left[(X-\mu)^4\right] = \frac{\sigma^2\Gamma(5/s)}{\Gamma(1/s)} \tag{3.35}$$

Moreover, skewness is 0 *and kurtosis are given by*

$$Kurt(X) = \frac{\Gamma(1/s)\Gamma(5/s)}{\Gamma^2(3/s)} \tag{3.36}$$

Above results will be useful in deriving the statistical properties of the new model. More detailed results on both GGD and its statistical properties could be found in [53] [55].

3.2.3 MLMAR model for intraday prices

After the discussion in previous sections, we are ready to propose the new MTD-type time series model. We generalize the MTD-type models [47] [69] [70] [71] to a new class, namely, the Multi-Logit Mixture Autoregressive (MLMAR) model.

Definition 3.18 *(***Multi-logit mixture autoregressive (MLMAR) model)** *The major time series of interest $\{Y_t, t\in\mathbf{Z}\}$, together with l exogenous time series $\{X_{j,t}, j=1,\cdots,l\}$ follows a MLMAR(k,p,q) model if*

$$f(y_t|\mathcal{F}_{t-1},\mathcal{I}_{t-1}) = \sum_{i=1}^{k}\alpha_{i,t}\tilde{\phi}\left(y_t;\mu_{i,t},\sigma_i,s_i\right) \tag{3.37}$$

where

$$\mu_{i,t} = \phi_{i1}y_{t-1} + \cdots + \phi_{ip}y_{t-p}, \quad i = 1, \cdots, k \tag{3.38}$$

$$\log(\alpha_{i,t}/\alpha_{1,t}) = \beta_{i,0} + \sum_{j=1}^{l}\sum_{r=1}^{q}\beta_{i,jr}x_{j,t-r}, \;\; i = 2, \cdots, k \tag{3.39}$$

$f(y_t|\mathcal{F}_{t-1}, \mathcal{I}_{t-1})$ *is the conditional density function of* Y_t *given the past information* $\mathcal{F}_{t-1}$ *and* $\mathcal{I}_{t-1}$ *, evaluated at* y_t*;* $\mathcal{F}_{t-1} = \{y_{t-1}, y_{t-2}, \cdots\}$*;*$\mathcal{I}_{t-1} = \{x_{1,t-1}, x_{1,t-2}, \cdots; \cdots; x_{l,t-1}, x_{l,t-2}\cdots\}$*;* $\tilde{\phi}(\cdot)$ *is the density of generalized Gaussian distribution;* $\alpha_{i,t}$ *is the* i^{th} *weight for the mixtures and* $\sum_{i=1}^{k}\alpha_{i,t} = 1$ *for all t.*

The MLMAR model is an extension of the LMARX [70] model. We have following remarks.

- The LMARX model has two mixture components; the MLMAR model allows for arbitrary k mixture components. MLMAR model is an extension of the LMARX model.

- The conditional density in each mixture of the LMARX model is limited to be Gaussian density; the MLMAR model considers far more flexible densities, while retaining the Gaussian density as one special case. Together with the generalization from two mixture components to k mixture components, the MLMAR model provides much richer predictive densities.

- In LMARX model, the past information set for exogenous time series is $\mathcal{I}_t$; in MLMAR model, we regularize and specify it as $\mathcal{I}_{t-1}$, which is more realistic. Notice that at time t, the exogenous information could only be observed up to time $t-1$ in most real situations. However, it is no doubt that the MLMAR model could also take $\mathcal{I}_t$ as the information set. This depends on real situations.

- In LMARX model, the parametrization of the conditional mean in each mixture component takes into the exogenous variables and the parametrization of mixture weights takes into account of $\{y_t\}$. It is no doubt that MLMAR model could be specified as this also. However, up to now, there is no convincing ways for the model selections or hypothesis tests of such arbitrary specification. Thus, in modeling high frequency data, we particularly specify a simplified model form of the MLMAR model, i.e., the exogenous variables control the weights only.

We give a general definition of the MLMAR model above. Since our main motivation for this new model is to model the high frequency intraday stock prices. For convenience, we rewrite the model specifically for the problem of interest. Thus, in what follows, we discuss the MLMAR model using the notations that are related to intraday data.

Suppose we observe the intraday stock prices $\{y_t, t = 1, 2 \cdots\}$, sampling at every T seconds; the total transaction volume during the time period $t-1$ and t, given by $v_t = \sum_{t-1<\tau_j\leq t} u_{\tau_j}$, where u_{τ_j} is the traded size at transaction time τ_j; the total numbers of trades happening during time $t-1$ and t, given by $n_t = \sum_j 1_{\{t-1<\tau_j\leq t\}}$, which is a measure of trading frequency. Figure 3.6 and Figure 3.7 plot two intraday data. Each figure includes the prices, transaction volumes and the total number of trades for each sample intervals (i.e., 20 seconds or 30 seconds).

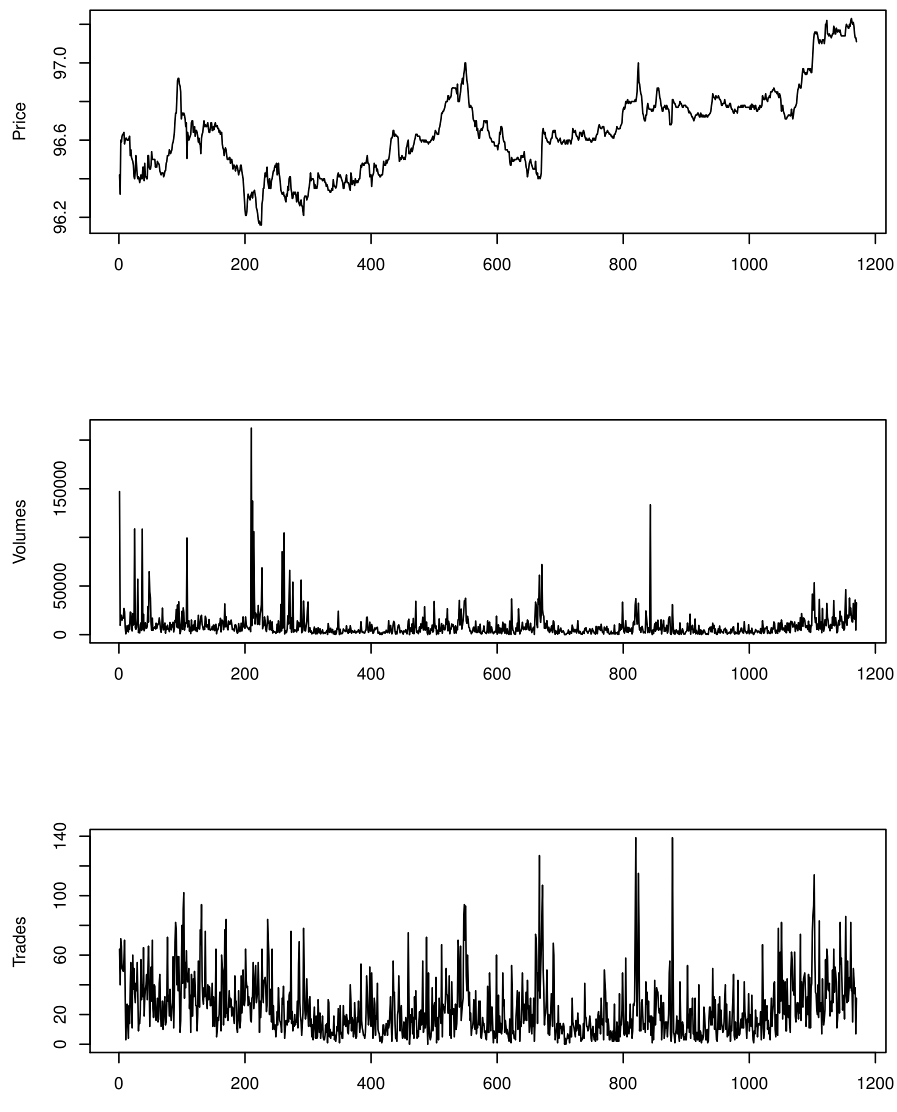

Figure 3.6: A plot of IBM stock intraday prices, volumes and number of transactions sampled every 20 seconds on 01/22/2007.

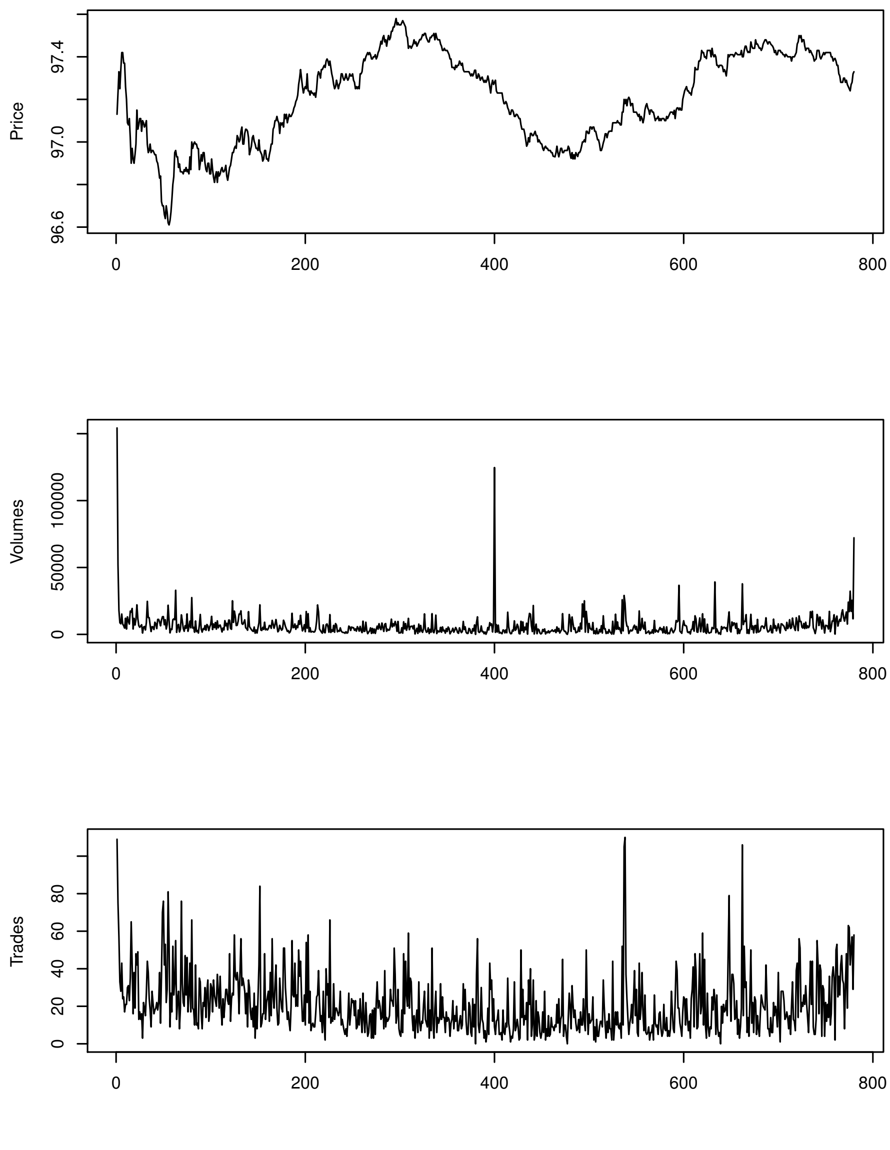

Figure 3.7: A plot of IBM stock intraday prices, volumes and number of transactions sampled every 30 seconds on 01/24/2007.

With previous notations, we reformulate the model in terms of modeling high frequency intraday prices.

Definition 3.19 (MLMAR(k, p, q) model for high frequency intraday prices) *An intraday price series $\{y_t, t = 1, 2, \cdots\}$ is said to follow a MLMAR(k, p, q) process if*

$$f(y_t|\mathcal{F}_{\mathbf{t-1}}, \mathcal{I}_{\mathbf{t-1}}) = \sum_{i=1}^{k} \alpha_{i,t} \frac{s_i}{2\sigma_i \Gamma(1/s_i)} \exp\left\{-\left|\frac{y_t - \mu_{i,t}}{\sigma_i}\right|^{s_i}\right\} \tag{3.40}$$

where

$$\mu_{i,t} = \sum_{j=1}^{p} \phi_{ij} y_{t-j} \quad (i = 1, \cdots, k)$$

$$\log\left(\frac{\alpha_{i,t}}{\alpha_{1,t}}\right) = \beta_{i,0} + \beta_{i,1} v_{t-1} + \cdots + \beta_{i,q} v_{t-q} + \beta_{i,q+1} n_{t-1} + \cdots + \beta_{i,2q} n_{t-q} = \boldsymbol{X}_t' \boldsymbol{\beta}_i, \quad i = 2, \cdots, k$$

where $\{v_t; t = 1, 2 \cdots\}$ and $\{n_t; t = 1, 2 \cdots\}$ are the total transaction volumes and the total number of transactions made in each fixed time interval respectively. In the equations that model the weights, $\mathbf{X}_t = (1, v_{t-1}, \cdots, v_{t-q}, n_{t-1}, \cdots, n_{t-q})'$ and $\boldsymbol{\beta}_i = (\beta_{i,0}, \beta_{i,1}, \cdots, \beta_{i,2q})'$.

In modeling high frequency intraday prices, we have following remarks:

- We want to emphasize that the *ultra-high frequency* data discussed in [26] considers the essential features of irregularly-spaced transaction time. In this chapter, the price series are sampled at fixed time interval, thus it belongs to the category of *high or very high frequency* data. However, since the exogenous variables include the transaction volumes and the transaction frequencies for the fixed time intervals (which are highly related with transaction arrivals), this model is also a type of *ultra-high frequency* data model with the prices as main focus.

- The choice of the exogenous variables or their functional transformations can be very flexible. For example, we can consider the largest transaction size traded in each fixed interval and the longest or average duration in each fixed interval. Both are important economic indicators that are related to liquidity. The exogenous variables can also be indicator variables, depend on whether the trades are buy-initiated or sell-initiated, etc. A model that includes these variables will be very useful in build up real statistical trading algorithm in practice.

- The shape parameter and the scale parameter of the generalized Gaussian distribution in the model could also be time dependent. However, we have not confirmed how much benefit could such kind of enlarged flexibilities bring in. Thus, in current stage, we consider these two sets of parameters as fixed scalars.

After introducing the new model, we would like to see a simple example.

Example 3.6 *(MLMAR(3,2,2)) A simple MLMAR(3,2,2) model for the high frequency data can be specified as:*

$$f(y_t|\mathcal{F}_{\mathbf{t-1}}, \mathcal{I}_{\mathbf{t-1}}) = \sum_{i=1}^{3} \alpha_{i,t}\tilde{\phi}(y_t; \mu_{i,t}, \sigma_i, s_i) \tag{3.41}$$

where $\tilde{\phi}(x; \mu, \sigma, s)$ is the generalized Gaussian density with location μ, scale parameter σ and shape parameter s. A particular form of the MLMAR(3,2,2) model could be

$$\mu_{1,t} = 1.000y_{t-1},\ \sigma_1^2 = 1.5,\ s_1 = 2.0; \tag{3.42}$$

$$\mu_{2,t} = 0.7021y_{t-1} + 0.2979y_{t-2},\ \sigma_2^2 = 1.0,\ s_2 = 1.0; \tag{3.43}$$

$$\mu_{3,t} = 1.1087y_{t-1} - 0.1087y_{t-2},\ \sigma_3^2 = 2.0,\ s_3 = 2.0; \tag{3.44}$$

$$log\frac{\alpha_{2,t}}{\alpha_{1,t}} = 0.68 + 0.3v_{t-1} + 0.2v_{t-2} + 0.4n_{t-1} - 0.5n_{t-2}; \tag{3.45}$$

$$log\frac{\alpha_{3,t}}{\alpha_{1,t}} = 0.88 + 0.2v_{t-1} + 0.1v_{t-2} + 0.4n_{t-1} + 0.3n_{t-2}. \tag{3.46}$$

Notice that in this simple example, the exogenous variables are the volumes and the transaction frequencies only.

3.3 Statistical inference

In this section, we study some statistical properties of the model. Essentially, the new class of MLMAR model is a time series model for general non-stationary and nonlinear time series. However, under very special conditions, we can obtain certain stationary conditions as in [47] [68] or [71].

3.3.1 Statistical properties

From the model specification, we can show that the conditional mean is given by

$$E(y_t|\mathcal{F}_{t-1}, \mathcal{I}_{t-1}) = \sum_{i=1}^{k} \alpha_{i,t}\mu_{i,t} = \sum_{i=1}^{k}\sum_{j=1}^{p} \alpha_{i,t}\phi_{ij}y_{t-j} \tag{3.47}$$

One advantage of the model lies in its capability in capturing time changing conditional variance, which is given by

$$\mathrm{var}(y_t|\mathcal{F}_{t-1},\mathcal{I}_{t-1}) = \sum_{i=1}^{k}\alpha_{i,t}c_i\sigma_i^2 + \sum_{i=1}^{k}\alpha_{i,t}\left(\sum_{j=1}^{p}\phi_{ij}y_{t-j}\right)^2 - \left(\sum_{i=1}^{k}\sum_{j=1}^{p}\alpha_{i,t}\phi_{ij}y_{t-j}\right)^2 \tag{3.48}$$

where $c_i = \frac{\Gamma(3/s_i)}{\Gamma(1/s_i)}$. The model has great flexibilities in capturing unimodal and multi-modal conditional distributions of the prices.

Notice that the MLMAR model is a model for non-stationary time series. Only under very special situations, we may obtain stationary time series. The non-stationarity is largely due to the exogenous time series $\{X_{j,t}, j = 1, \cdots, l\}$. If we take no account of the exogenous time series and only consider the major time series Y_t, we have similar weak stationary conditions for the MLMAR model. Since all mixture weights are assumed to be constant, the stationary conditions for the MTD-type model with a mixture of generalized Gaussian distributions are exactly the same as that for a model with mixtures of Gaussian distributions. Thus, follow [47] [68] or [71], we have two theorems.

Theorem 3.1 *(*1^{st} **order stationarity of MLMAR(k,p,0)** *[68]) For a MLMAR(k,p,0) process, a necessary and sufficient condition for the process* $\{Y_t\}$ *to be* 1^{st} *order stationary is all roots of the equation*

$$1 - \sum_{j=1}^{p}\left(\sum_{i=1}^{k}\alpha_i\phi_{ij}\right)z^{-j} = 0$$

lie inside the unit circle.

Theorem 3.2 *(2^{nd}* **order stationarity of MLMAR(k,1,0)** *[68]) For a MLMAR(k,1,0) process, which is 1^{st} order stationary, a necessary and sufficient condition for the processes $\{Y_t\}$ to be 2^{nd} order stationary if*

$$\left|\sum_{i=1}^{k} \alpha_i \phi_{i1}^2\right| < 1$$

The proof of the above theorems for the generalized Gaussian distributions situations is quite similar to the proof for the Gaussian case [47] and [68], apart from a few constants involving the shape parameters.

3.3.2 Parameter estimation: an ECM algorithm

In this section, we discuss the model estimation problem. Since the problem involves mixtures, the EM algorithm originally proposed by Demspter, Laird and Rubin [20] turns out to be a natural candidate to obtain the maximum likelihood estimates. We propose here an ECM algorithm [52] to solve the estimation problem, where quite complicated optimizations in the M-Step of the ECM algorithm are involved. The efficiency of the ECM algorithm is considered and evaluated via intensive simulation studies.

Recall that the MLMAR(k, p, q) model is given by

$$\begin{aligned} f\left(y_t|\mathcal{F}_{\mathbf{t-1}}, \mathcal{I}_{\mathbf{t-1}}\right) &= \sum_{i=1}^{k} \alpha_{i,t} \tilde{\phi}\left(y_t; \mu_{i,t}, \sigma_i, s_i\right) \\ &= \sum_{i=1}^{k} \alpha_{i,t} \frac{s_i}{2\sigma_i \Gamma(1/s_i)} \exp\left\{-\left|\frac{y_t - \mu_{i,t}}{\sigma_i}\right|^{s_i}\right\} \\ &= \sum_{i=1}^{k} \alpha_{i,t} \frac{s_i}{2\sigma_i \Gamma(1/s_i)} \exp\left\{-\left|\frac{e_{i,t}}{\sigma_i}\right|^{s_i}\right\} \qquad (3.49) \end{aligned}$$

where

$$e_{i,t} = y_t - \mu_{i,t} = y_t - \sum_{j=1}^{p} \phi_{ij} y_{t-j} = y_t - \mathbf{Y}_t' \boldsymbol{\phi}_i;\ \mathbf{Y}_t = \left(y_{t-1}, y_{t-2}, \cdots, y_{t-p}\right)';\ \boldsymbol{\phi}_i = \left(\phi_{i1}, \cdots, \phi_{ip}\right)'$$

$$\log\left(\frac{\alpha_{i,t}}{\alpha_{1,t}}\right) = \beta_{i,0} + \beta_{i,1} v_{t-1} + \cdots + \beta_{i,q} v_{t-q} + \beta_{i,q+1} n_{t-1} + \cdots + \beta_{i,2q} n_{t-q} = \boldsymbol{X}_t' \boldsymbol{\beta}_i \ \ (i = 2, \cdots, k)$$

The conditional log-likelihood for the observed data is given by

$$\begin{aligned} l^{**} &= \sum_{t=p+1}^{n} \log f\left(y_t | \mathcal{F}_{\mathbf{t-1}}, \mathcal{I}_{\mathbf{t-1}}\right) \\ &= \sum_{t=p+1}^{n} \log\left\{\sum_{i=1}^{k} \alpha_{i,t} \frac{s_i}{2\sigma_i \Gamma(1/s_i)} \exp\left\{-\left|\frac{e_{i,t}}{\sigma_i}\right|^{s_i}\right\}\right\} \end{aligned} \tag{3.50}$$

To find the maximum likelihood estimates directly by maximizing this log-likeliood is practically intractable, even numerically. However, a proposed Expectation-Conditional Maximization (ECM) [52] algorithm enables us to obtain the m.l.e. for this problem iteratively.

Suppose our observations are $\{y_t, (v_t, n_t)'\}_{t=1}^n$, where y_t is the observed transaction price at time t; and $(v_t, n_t)'$ is the exogenous information flow as previously discussed. We define

$$\mathbf{Z} = (\mathbf{Z}_1, \cdots, \mathbf{Z}_n) = \begin{pmatrix} z_{1,1} & \cdots & z_{1,t} & \cdots & z_{1,n} \\ \vdots & & \vdots & & \vdots \\ z_{k,1} & \cdots & z_{k,t} & \cdots & z_{k,n} \end{pmatrix} \tag{3.51}$$

be the unobserved random variables, such that at each time t, $z_{j,t} = 1$ $(j = 1, 2, \cdots, k)$ if y_t is generated from the j^{th} component of the conditional distribution and $z_{j,t} = 0$ otherwise.

Then the conditional log-likelihood for the complete data is given by

$$l = \sum_{t=p+1}^{n} \sum_{i=1}^{k} \left(Z_{i,t}\log(\alpha_{i,t}) + Z_{i,t}\log(s_i) - Z_{i,t}\log\Gamma(1/s_i) - Z_{i,t}\log(\sigma_i) - Z_{i,t}\left|\frac{e_{i,t}}{\sigma_i}\right|^{s_i} \right) \tag{3.52}$$

The ECM algorithm for the MLMAR model is derived as follows.

E-Step: Let the conditional expectation of the i^{th} component of $\mathbf{Z}_t$ be $\tilde{z}_{i,t}$, then

$$\tilde{z}_{i,t} = \frac{\alpha_{i,t}\tilde{\phi}(y_t; \mu_{i,t}, \sigma_i, s_i)}{\sum_{j=1}^{k} \alpha_{j,t}\tilde{\phi}(y_t; \mu_{j,t}, \sigma_j, s_j)} \tag{3.53}$$

CM-Steps: Notice that the likelihood involving the mixture proportions is

$$\begin{aligned} l^* &= \sum_{t=p+1}^{n} \sum_{i=1}^{k} \tilde{z}_{i,t}\log\left(\frac{\alpha_{i,t}}{\alpha_{1,t}}\alpha_{1,t}\right) = \sum_{t=p+1}^{n} \sum_{i=1}^{k} \tilde{z}_{i,t}\left(\log\frac{\alpha_{i,t}}{\alpha_{1,t}} + \log\alpha_{1,t}\right) \\ &= \sum_{t=p+1}^{n} \left(\sum_{i=2}^{k} \tilde{z}_{i,t}\boldsymbol{\beta}_i'\boldsymbol{X}_t - \left(\sum_{i=1}^{k}\tilde{z}_{i,t}\right) \cdot \log\left(1 + \sum_{j=2}^{k}\boldsymbol{\beta}_j'\boldsymbol{X}_t\right)\right) \end{aligned} \tag{3.54}$$

Taking partial derivatives with respect to $\boldsymbol{\beta}_i$ $(i = 2, \cdots, k)$ yields,

$$\frac{\partial l}{\partial \boldsymbol{\beta}_i} = \frac{\partial l^*}{\partial \boldsymbol{\beta}_i} = \sum_{t=p+1}^{n} \left(\tilde{z}_{i,t} - \frac{\sum_{i=1}^{k} \tilde{z}_{i,t}}{1 + \sum_{j=2}^{k} \boldsymbol{\beta}_j'\boldsymbol{X}_t} \right) \boldsymbol{X}_t \tag{3.55}$$

$$\frac{\partial^2 l}{\partial \boldsymbol{\beta}_i \partial \boldsymbol{\beta}_i^T} = \sum_{t=p+1}^{n} \frac{\sum_{i=1}^{k} \tilde{z}_{i,t}}{\left(1 + \sum_{j=2}^{k} \boldsymbol{\beta}_j'\boldsymbol{X}_t\right)^2} \boldsymbol{X}_t\boldsymbol{X}_t' \tag{3.56}$$

Similarly, other partial derivatives with respect to σ_i, s_i and the $\boldsymbol{\phi}_i$'s are given as follows.

$$\frac{\partial l}{\partial \sigma_i} = \sum_{t=p+1}^{n} \frac{-\tilde{z}_{i,t}}{\sigma_i} - \sum_{t=p+1}^{n} \frac{\tilde{z}_{i,t}|e_{i,t}|^{s_i} s_i}{\sigma_i^{s_i+1}} \tag{3.57}$$

$$\frac{\partial l}{\partial s_i} = \sum_{t=p+1}^{n} \frac{\tilde{z}_{i,t}}{s_i} + \tilde{z}_{i,t}\frac{1}{\Gamma(1/s_i)s_i^2} - \tilde{z}_{i,t}\left|\frac{e_{i,t}}{\sigma_i}\right|^{s_i}\log\left|\frac{e_{i,t}}{\sigma_i}\right| \tag{3.58}$$

$$\frac{\partial^2 l^*}{\partial s_i^2} = \sum_{t=p+1}^{n} \frac{-\tilde{z}_{i,t}}{s_i^2} \left\{ 1 + \frac{2}{s_i \Gamma(1/s_i)} - \frac{1}{\Gamma'(1/s_i) s_i^2} \right\} - \tilde{z}_{i,t} \left(\log |\frac{e_{i,t}}{\sigma_i}| \right)^2 |\frac{e_{i,t}}{\sigma_i}|^{s_i} \tag{3.59}$$

$$\frac{\partial l}{\partial \boldsymbol{\phi}_i} = \sum_{t=p+1}^{n} \sum_{i=1}^{k} \frac{\tilde{z}_{i,t} s_i}{\sigma_i^{s_i}} \left\{ 1_{\{e_{i,t}>0\}} \left(y_t - \boldsymbol{\phi}_i' \boldsymbol{Y}_t \right)^{s_i - 1} - 1_{\{e_{i,t}<0\}} \left(-y_t + \boldsymbol{\phi}_i' \boldsymbol{Y}_t \right)^{s_i - 1} \right\} \boldsymbol{Y}_t \tag{3.60}$$

$$\frac{\partial^2 l}{\partial \boldsymbol{\phi}_i \boldsymbol{\phi}_i^T} = - \sum_{t=p+1}^{n} \sum_{i=1}^{k} \frac{\tilde{z}_{i,t} s_i}{\sigma_i^{s_i}} \left\{ 1_{\{e_{i,t}>0\}} (s_i - 1) \left(y_t - \boldsymbol{\phi}_i' \boldsymbol{Y}_t \right)^{s_i - 2} + 1_{\{e_{i,t}<0\}} (s_i - 1) \left(-y_t + \boldsymbol{\phi}_i' \boldsymbol{Y}_t \right)^{s_i - 2} \right\} \boldsymbol{Y}_t \boldsymbol{Y}_t' \tag{3.61}$$

Then starting with $\boldsymbol{\beta}^{(0)} = \left(\boldsymbol{\beta}_2^{(0)\prime}, \cdots, \boldsymbol{\beta}_k^{(0)\prime} \right)'$, $\boldsymbol{\phi}^{(0)} = \left(\boldsymbol{\phi}_1^{(0)\prime}, \cdots, \boldsymbol{\phi}_k^{(0)\prime} \right)'$, $\sigma_i^{(0)}$ $(i = 1, \cdots, k)$ and $s_i^{(0)}$ $(i = 1, \cdots, k)$, for the **CM-steps** we have the iteratively updates given as follows

$$\boldsymbol{\beta}_i^{(m+1)} = \boldsymbol{\beta}_i^{(m)} - \left\{ \frac{ \left(\sum_{i=1}^{k} \tilde{z}_{i,t} \right) \boldsymbol{X}_t \boldsymbol{X}_t' }{ \left(1 + \sum_{j=2}^{k} \boldsymbol{\beta}_j^{(m)\prime} \boldsymbol{X}_t \right)^2 } \right\}^{-1} \left\{ \sum_{t=p+1}^{n} \left(\tilde{z}_{i,t} - \frac{ \sum_{i=1}^{k} \tilde{z}_{i,t} }{ 1 + \sum_{j=2}^{k} \boldsymbol{\beta}_j^{(m)\prime} \boldsymbol{X}_t } \right) \boldsymbol{X}_t \right\} \tag{3.62}$$

The update of $\sigma_i^{(m+1)}$ is

$$\sigma_i^{(m+1)} = \left\{ \left(\sum_{t=p+1}^{n} \tilde{z}_{i,t} |e_{i,t}^{(m)}|^{s_i^{(m)}} s_i^{(m)} \right) \left(\sum_{t=p+1}^{n} \tilde{z}_{i,t} \right)^{-1} \right\}^{1/s_i^{(m)}} \tag{3.63}$$

The update of $s_i^{(m+1)}$ is

$$s_i^{(m+1)} = s_i^{(m)} - \left\{ \frac{\partial^2 l}{\partial s_i^2} \right\}^{-1} \Big|_{s_i^{(m)}, e_{i,t}^{(m)}, \sigma_i^{(m+1)}} \left\{ \frac{\partial l}{\partial s_i} \right\} \Big|_{s_i^{(m)}, e_{i,t}^{(m)}, \sigma_i^{(m+1)}} \tag{3.64}$$

At last, the update of $\boldsymbol{\phi}_i^{(m+1)}$ is

$$\boldsymbol{\phi}_i^{(m+1)} = \boldsymbol{\phi}_i^{(m)} - \left\{ \frac{\partial^2 l}{\partial \boldsymbol{\phi}_i \boldsymbol{\phi}_i^T} \right\}^{-1} \Big|_{s_i^{(m+1)}, e_{i,t}^{(m)}, \sigma_i^{(m+1)}} \left\{ \frac{\partial l}{\partial \boldsymbol{\phi}_i} \right\} \Big|_{s_i^{(m+1)}, e_{i,t}^{(m)}, \sigma_i^{(m+1)}} \tag{3.65}$$

Then the ECM algorithm runs the **E-step** and the **CM-steps** iteratively until converge. Since the ECM algorithm belongs to the generalized EM algorithm, it preserves the increas-

Table 3.1: Empirical means and standard deviations(SD) of the parameter estimates via ECM algorithm for simulated data.

$\boldsymbol{s}$	mean(sd)	$\boldsymbol{\sigma}^2$	mean(sd)	$\boldsymbol{\phi}_1$	mean(sd)	$\boldsymbol{\phi}_2$	mean(sd)
2.0000	2.0515 (0.19700)	1.5000	1.7327 (0.0923)	1.0000	0.9492 (0.0207)	0.7021	0.7089 (0.0191)
1.0000	1.0001 (0.03280)	1.0000	1.0065 (0.0622)	0.0000	0.0507 (0.0207)	0.2979	0.2911 (0.0190)
2.0000	1.9863 (0.08916)	2.0000	1.9893 (0.0469)				

$\boldsymbol{\phi}_3$	mean(sd)	$\boldsymbol{\beta}_1$	mean(sd)	$\boldsymbol{\beta}_2$	mean(sd)
1.1087	1.1134 (0.0142)	0.6800	0.6056 (0.3855)	0.8800	0.9484 (0.3566)
-0.1087	-0.1135 (0.0143)	0.3000	0.3319 (0.0572)	0.2000	0.2097 (0.0675)
		0.2000	0.2331 (0.0908)	0.1000	0.1312 (0.0957)
		0.4000	0.4219 (0.0664)	0.4000	0.3945 (0.0735)
		-0.5000	-0.5357 (0.0775)	0.3000	0.3119 (0.0985)

ing likelihood property. Thus, if the likelihood is bounded above, then the ECM algorithm surely converges to the maximized likelihood estimates given enough iterations.

A simulation study

To investigate the performance of the ECM algorithm, we simulate 100 sample paths from the MLMAR(3,2,2) model as in Example 3.6. Each sample path has 5000 successive observations. In particular, the exogenous information series are simulated with two AR(1) processes: $X_{1,t} = 0.7X_{1,t-1} + \epsilon_{1,t}$ and $X_{2,t} = 0.6X_{2,t-1} + \epsilon_{2,t}$ with $\epsilon_{1,t}$ and $\epsilon_{2,t}$ are both Gaussian white noise with mean 0, but with variance 3 and 4 respectively.

The ECM algorithm is carried out. The empirical means and standard deviations of the estimates are summarized in Table 3.1. The result shows that the ECM algorithm works pretty well, where the means are very closed to the true parameters with relatively small standard deviations.

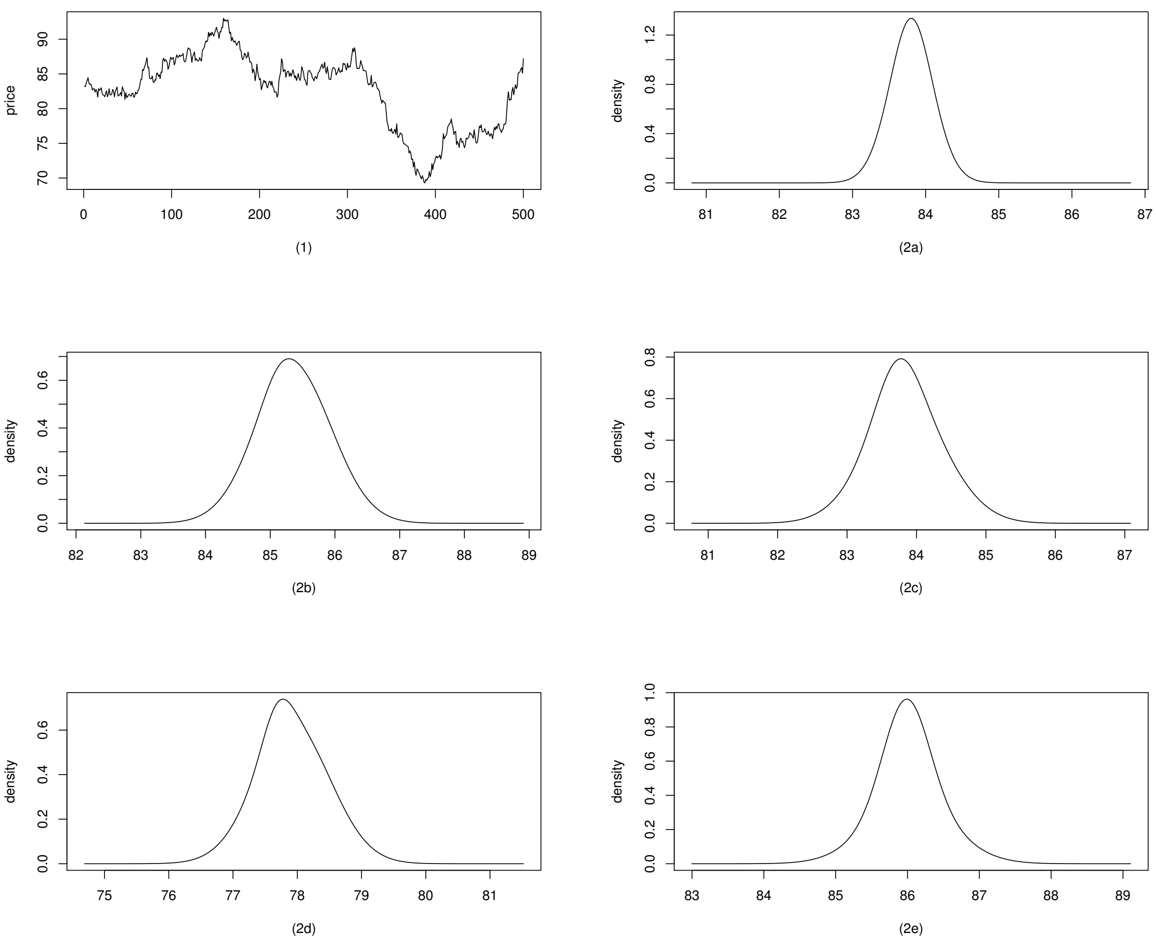

Figure 3.8: (1) Simulated MLMAR(3,2,2) processes with three Gaussian components; (2a)-(2e) conditional densities.

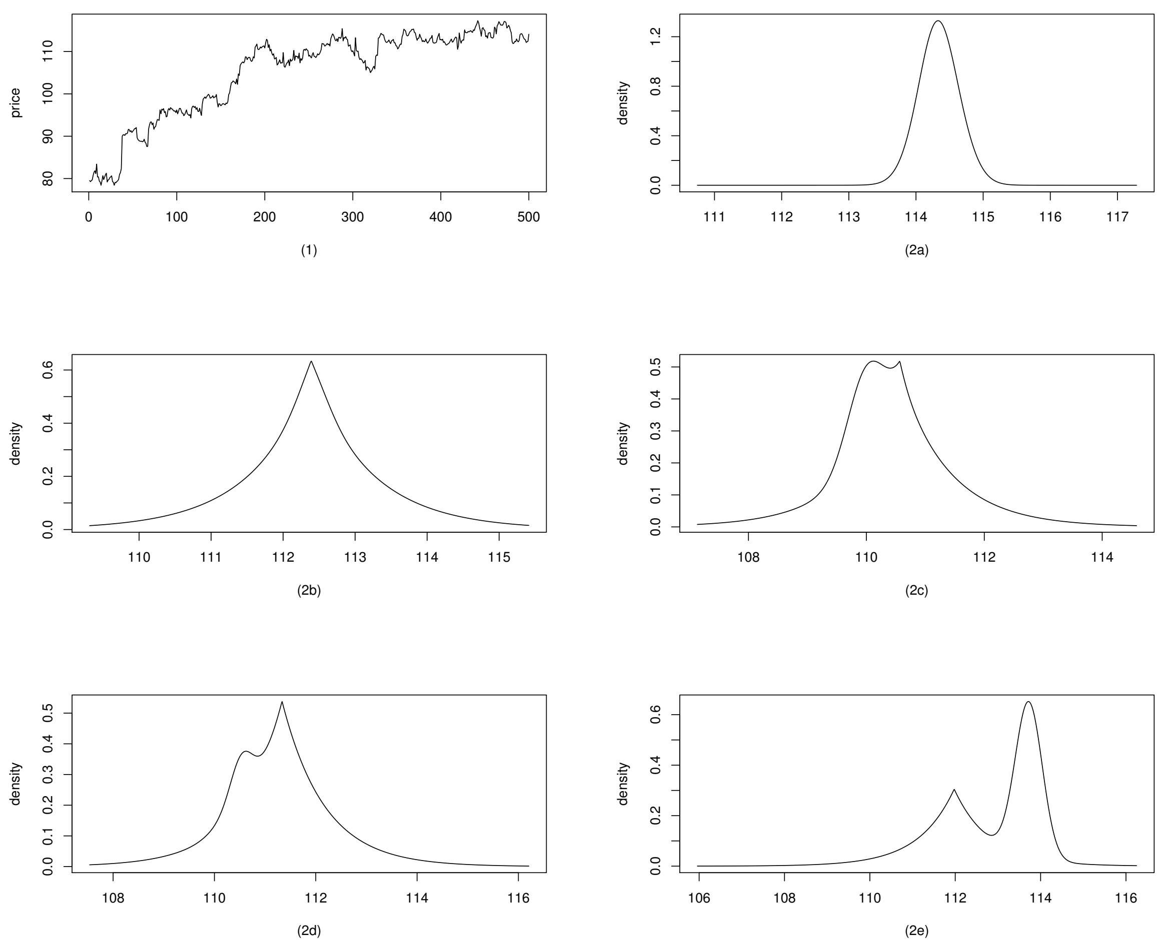

Figure 3.9: (1) Simulated MLMAR(3,2,2) processes with two Gaussian components and one Laplace component; (2a)-(2e) conditional densities.

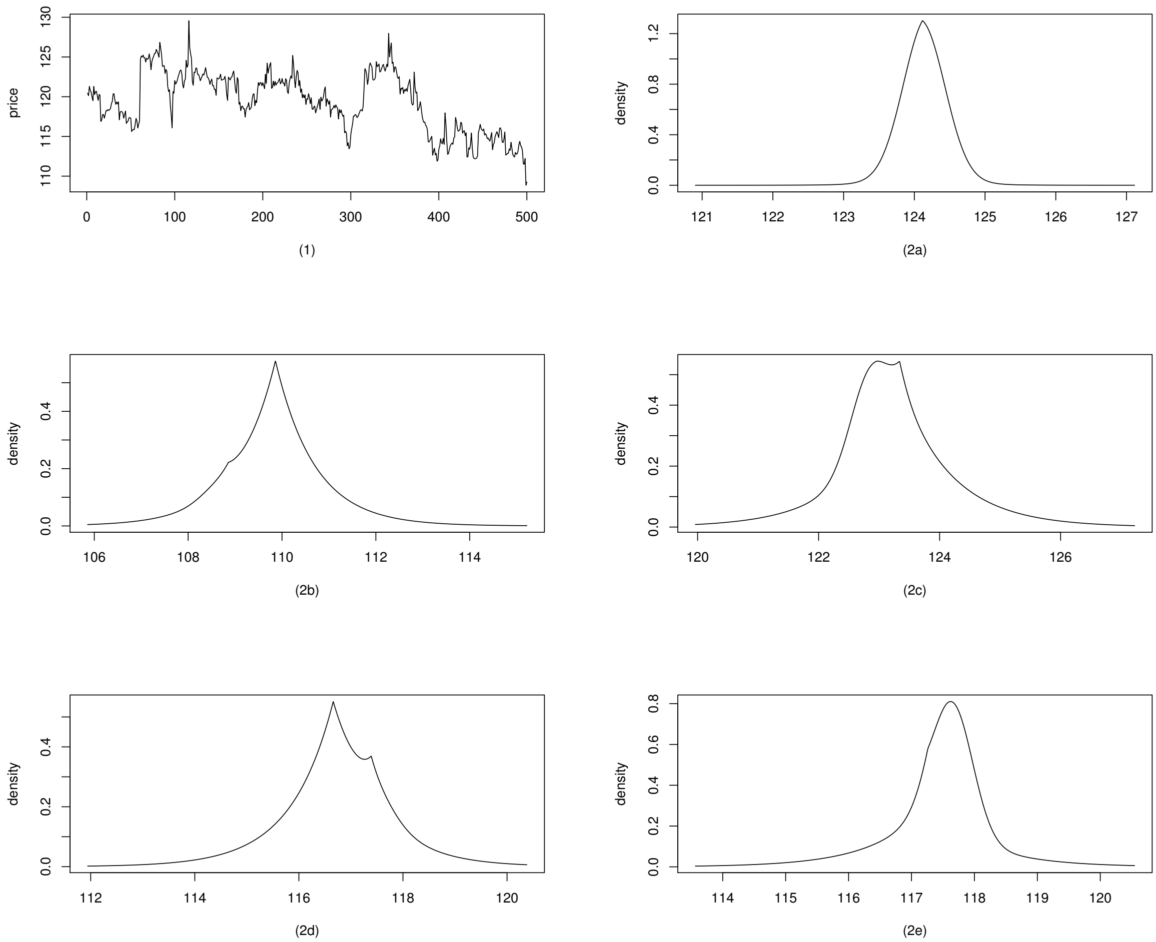

Figure 3.10: (1) Simulated MLMAR(3,2,2) processes with two Laplace components and one Gaussian component; (2a)-(2e) conditional densities.

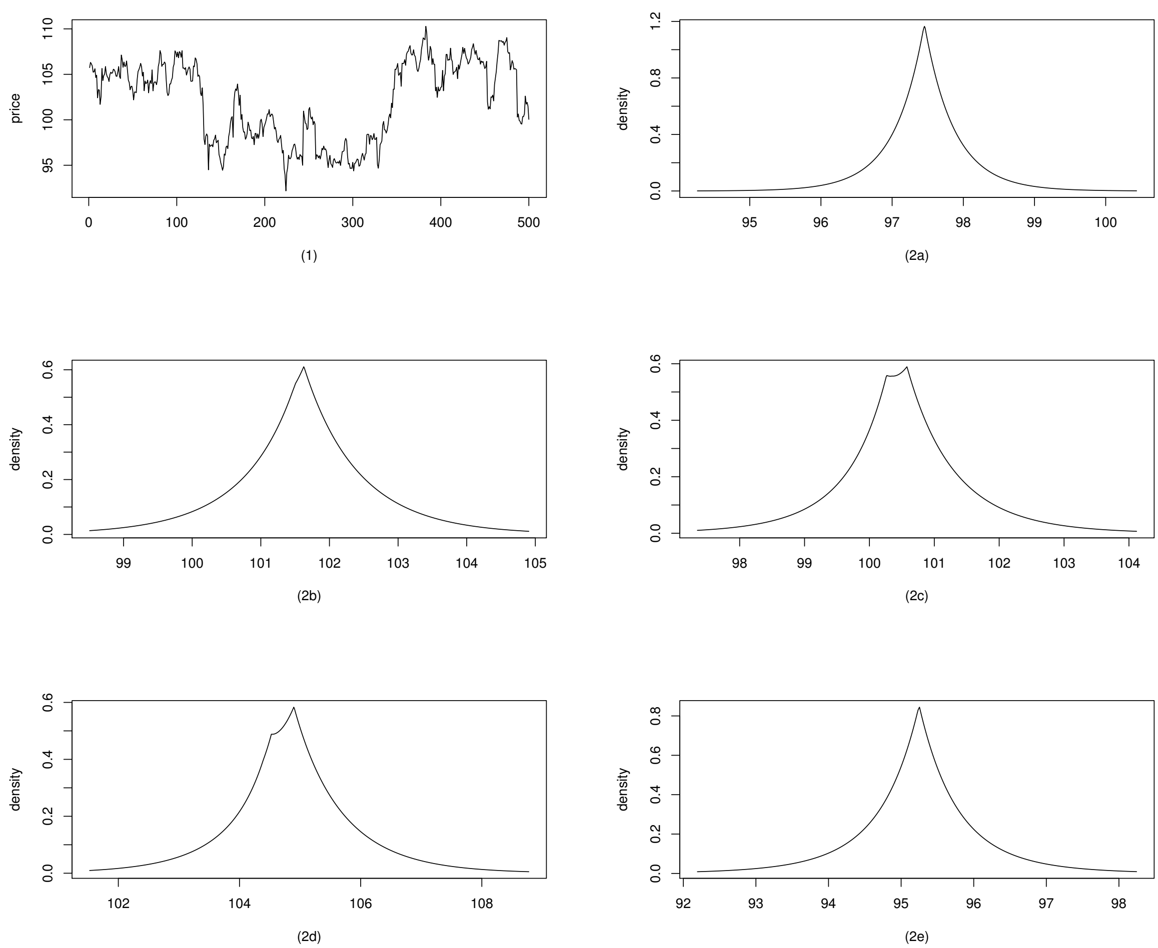

Figure 3.11: (1) Simulated MLMAR(3,2,2) processes with three Laplace components; (2a)-(2e) conditional densities.

3.3.3 Prediction and model selection

One advantage for the MLMAR model is its ability to obtain multi-modal predictive distributions. When the interest is in the point forecast, the highest density point from the predictive distribution may be used. If we want to construct a prediction confidence interval, a highest density region is preferred. In such situation, the predictive highest density region may be comprised of several disjoint intervals.

In practice, the point forecast could be used to get a single best predictive price. A fully visualized density forecast could be used for risk accessment. Figure 3.8 - Figure 3.11 illustrate some of the realized predictive densities from the MLMAR models. All parameter settings are similar to those in the Example 3.6, except that the conditional distributions for the mixture components are different; the scale parameters are $\sigma_1^2 = 0.1$, $\sigma_2^2 = 0.7$ and $\sigma_3^2 = 0.2$; and the exogenous informations are given by two AR(1) process: $X_{1,t} = 0.7X_{1,t-1} + \epsilon_{1,t}$ and $X_{2,t} = 0.6X_{2,t-1} + \epsilon_{2,t}$ with $\epsilon_{1,t}$ and $\epsilon_{2,t}$ are both Gaussian white noise with mean 0, but with variance 3 and 4 respectively.

Figure 3.8 illustrates a sample path and some conditional densities facilitated by a model whose all three mixture components are Gaussian distributions. This is similar to the case as the LMARX model [70].

Figure 3.9 illustrates a sample path and some conditional densities facilitated by a model whose first two mixture components are Gaussian distributions and last one is Laplace distribution.

Figure 3.10 illustrates a sample path and some conditional densities facilitated by a model whose first two mixture components are Laplace distributions and last one is Gaussian distribution.

Figure 3.11 illustrates a sample path and some conditional densities facilitated by a model whose all three mixture components are Laplace distributions.

From above discussions and observations, we could see very clearly that the MLMAR model provides far more flexible and richer conditional densities than the LMARX model. Now, so many types of predictive densities are possible. These may include asymmetric densities, heavy or light tails densities, densities with multi-modality and leptokurtic densities, etc. All these good properties are facilitated by using mixtures of generalized Gaussian distributions.

Last, we discuss the model selection problem. For most MTD-type models, a widely accepted and used model selection criteria is the BIC value. The most recent development for the model selection problem is due to [54], where the situation of a mixture of Gaussian distributions is studied. For the MLMAR model, we would suggest using BIC as the general model selection criteria, although better criteria for such model may exist and will be well deserved for further studies.

3.4 Modeling IBM intraday stock prices

In this section, we apply the MLMAR model to real intraday stock transaction data. We pick a piece of IBM intraday stock transaction data from the TAQ data base. The date we select here is the data on 01/22/2007. Notice that there exist some transactions before and after the standard trading hours. To be unified, we select the data that strictly falls between the nominal trading hours, i.e., 9:30AM - 4:00PM.

The dataset comprises the intraday stock prices $\{y_t, t = 1, 2, \cdots\}$, sampling at every $T =$ 20 seconds. This results in totally 1170 observations. The exogenous information include the transaction volumes and frequencies of transactions. The total transaction volume (in 10,000 shares) during the time period $t-1$ and t is $v_t = \sum_{t-1<\tau_j\leq t} u_{\tau_j}$, where u_{τ_j} is the traded size at transaction time τ_j. A measure of the transaction speed/frequency is the average number of trades per second during time interval $t-1$ and t, given by $n_t = \frac{1}{20}\sum_{\tau_j} 1_{\{t-1<\tau_j\leq t\}}$. Notice that here, n_t is the average number of trades rather than the total number of trades, which we used when introducing the new modeling framework.

We fit a combination of the MLMAR(k, p, q) models to the intraday data with $k = 1, \cdots, 4$; $p = 1, \cdots, 4$; and $q = 1, \cdots, 4$. The best MLMAR model with the minimum BIC value turns out to be a MLMAR(2, 3, 2) model given by

$$f(y_t|\mathcal{F}_{\mathbf{t-1}}, \mathcal{I}_{\mathbf{t-1}}) = \alpha_{1,t}\tilde{\phi}(y_t; \mu_{1,t}, \sigma_1, s_1) + \alpha_{2,t}\tilde{\phi}(y_t; \mu_{2,t}, \sigma_2, s_2) \tag{3.66}$$

where

$$\mu_{1,t} = 0.9326y_{t-1} + 0.0382y_{t-2} + 0.0272y_{t-3}, \ \sigma_1^2 = 0.02, \ s_1 = 1.67;$$

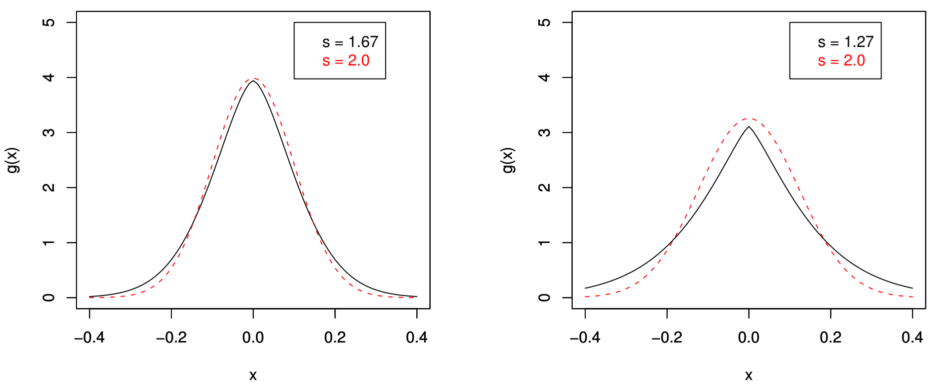

Figure 3.12: A comparison of the predictive densities of the fitted model with Gaussian density (with common scale parameters). (a) for the first mixture component (s=1.67); (b) for the second mixture component (s=1.27).

$$\mu_{2,t} = 0.8707y_{t-1} - 0.0448y_{t-2} + 0.1742y_{t-3},\ \sigma_2^2 = 0.03,\ s_2 = 1.27;$$

$$\log\frac{\alpha_{2,t}}{\alpha_{1,t}} = -1.6780 - 0.0828v_{t-1} + 0.2975v_{t-2} + 0.7906n_{t-1} + 0.2018n_{t-2}.$$

Figure 3.12 shows a comparision of the predictive densities (given by above model) with Gaussian densities. The fitted GGD indicates positive excess of kurtosis in this case, since the shape parameters are both less than 2.

We want to compare the prediction performance of the fitted MLMAR(2, 3, 2) model with the standard univariate ARIMA model. The best fitted ARIMA model for the high frequency prices is given by a ARIMA(2, 1, 4) model

$$(1-0.6514B+0.9534B^2)(1-B)y_t = (1-0.7534B+0.9408B^2-0.0148B^3-0.0850B^4)z_t \quad (3.67)$$

where B is backward shift operator and $\{z_t\}$ is are white noise $\text{WN}(0, \sigma_z^2)$.

We also fit the LMARX model to the real data, which is a special case of the MLMAR model with a mixture of two Gaussian transition distributions.

Table 3.2: Prediction mean square error for the IBM high frequency intraday prices.

Sample Frequency	MLMAR (Model)	LMARX (Model)	ARIMA (Model)
20s	0.000604 (2,3,2)	0.000603 (2,3,2)	0.000672 (5,1,5)
30s	0.000927 (4,2,3)	0.000931 (2,2,2)	0.001044 (3,1,5)
45s	0.001513 (2,2,2)	0.001530 (2,2,2)	0.001668 (1,1,2)
60s	0.001985 (2,2,2)	0.001998 (2,2,2)	0.001954 (2,1,2)

In high frequency trading, the one-step ahead prediction is of the key concern in building up the trading signals. Thus, we calculate the one-step ahead prediction mean square error for all fitted models.

We fit the models to the data sampling every 20, 30, 45 and 60 seconds, and calculate the one-step ahead prediction mean square errors. Results are summarized in Table 3.2. Overall, we see that the prediction performance of the MLMAR model are very attractive. For the data sampled every 30, 45 and 60 seconds, the overall prediction accuracy is improved by 25% of 1-tick (i.e., USD 1 cents) than the LMARX models . However, for the time series with sampling frequency to be 60 seconds, the linear ARIMA model seems doing a better job. This gives us a hint that for the high frequency intraday data, the best models may depend on the sampling frequencies.

3.5 Application: simplified high frequency trading

One of the most important motivations to model the high frequency data is to facilitate automated trading. In recent years, the orders are able to be placed and executed tens or even hundreds of times per day for the same stock, FX, or other financial products that can be traded via electronic systems. As an interesting application of the MLMAR model to stock trading, we discuss a simplified example. We first provide the trading definition and criteria. Then we investigate the algorithm's performance in generating profit.

To set up a high frequency trading algorithm, let $F_t \in \{-1, 1\}$ be the position (long/short) taken at time t. Neutral positions are not allowed so that the"trader" is always in the market. During each trading, the realized price return is given by

$$r_t = y_t - y_{t-1} \tag{3.68}$$

Also, a standard practice is that a fixed amount c is invested in each trade with a transaction cost δ for each transaction unit. Then the cumulative profit at time T, P_T, is given by

$$P_T = \sum_{t=1}^{T} R_t = \sum_{t=1}^{T} c \cdot (F_{t-1} r_t - \delta |F_t - F_{t-1}|) \tag{3.69}$$

where R_t is the realized return taken into the consideration of transaction cost. For highly liquid market, the transaction cost δ would be very small. To gain an insight for the performance of the MLMAR model in this simplified trading example, we calculate the cumulative profit by dropping the term that involves transaction cost and setting $c = 1000$ share of the stocks.

The data is sampled every 30 seconds. Thus, for a single stock, it results in 780 observations of prices $\{y_t\}_{t=1}^{780}$ and related exogenous informations. The first 450 observations are used to fit the MLMAR(k, p, q) time series model; the remaining 330 observations are reserved for out of sample forecast and testing the trading algorithm. Suppose the one-step

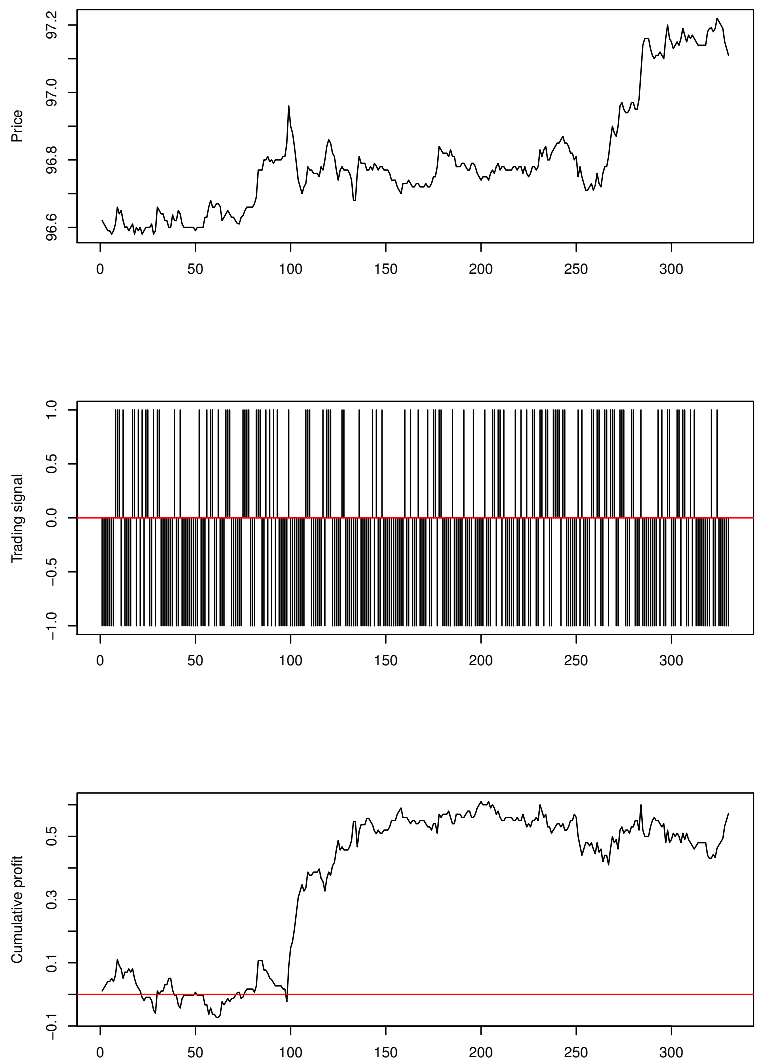

Figure 3.13: High frequency trading experiment based on the out-of-sample forecast of ML-MAR model. Training with first 450 observation on prices, cumulative volumes and frequencies of transactions; forecasting with remaining 330 observations.

forecast of the price y_t is $\hat{y}_t$. Then a straightforward trading signals (long/short) would be $F_{t-1} = 1$ if $\hat{y}_t > y_{t-1}$, and $F_{t-1} = -1$ if $\hat{y}_t < y_{t-1}$.

Figure 3.13 shows the result of the cumulative profit via the trading based on MLMAR mode for the date 01/22/2007. The algorithm trading performance is very interesting and attractive. However, we notice that this is a much simplified case for real trading, which excludes transaction cost and the uncertainty of obtaining the target executed prices one wish to get. Thus, the real performance based on the MLMAR model would be different. In this example, it mainly illustrates the out-of-sample prediction performance of the model.

3.6 Conclusion and discussion

In this chapter, with an aim at modeling high frequency stock prices from the TAQ dataset, we proposed a class of Multi-Logit Mixture Autoregressive model. Related statistical modeling and estimation problem are discussed. The model is used to fit the high frequency stock prices series and a potential profitable trading algorithm based on the new model is derived and illustrated. There exist possible extensions of the model. For example, one may choose different types of functions to define the time changing mixture proportions.

For the estimation problem, the requirement of large sample size may turn out to be a potential problem for practical data modeling. This issue may be alleviated by using other estimation methods. The final goal of modeling the high frequency data is to put them into practice. To be a potential successful algorithm trading model, the MLMAR model should be extended and enriched in ways such as adding more (important) exogenous variables that represent the current market conditions. However, all these are tractable from the statistical modeling aspects. The difficulty is from the financial interpretations of new variables.

Chapter 4

Multivariate MTD Framework for Marked Point Processes

In Chapter 3, we discussed the high frequency intraday data, i.e., the intraday prices sampled at fixed time interval ($h = 20$ seconds, for example). Such kind of data is indeed an aggregation of the TAQ data, which undoubtedly loses certain useful information. The most distinguished feature of the *ultra-high frequency data* is that the transactions are irregularly-spaced in time [26]. The ultra-high frequency data is essentially a type of marked point process (MPP).

In this chapter, the marked point process data sets, in particular the ultra-high frequency data or tick-by-tick data, are the main subjects that we are going to model. In Section 4.1, we briefly review the past research and motivate our new model. We also introduce a useful statistical technique for constructing multivariate distributions. It is called the *Copula*. Most of the nice features of our new model naturally originate from the good properties of Copula. In Section 4.2, we propose the multivariate MTD (MMTD) model and discuss its relationship with existing models. In Section 4.3, we study the estimation schemes and discuss the prediction issues. In Section 4.4, we illustrate the usefulness of the new modeling framework via two numerical examples - one with simulated data and one with real ultra-high frequency data. We conclude in Section 4.5.

4.1 Introduction

4.1.1 Review

We have introduced the MTD-type models in Section 3.1 of Chapter 3. In this section, we briefly review some literature that represent the recent development. At this moment, we particularly focus on the motivation aspects for the proposal of our new model.

Le, Martin and Raftery proposed a univariate Mixture Transition Distribution (MTD) time series model [47]. In their model, a time series $\{X_t : t = 1, 2, \cdots\}$ is generated from the MTD model if

$$F\left(x_t|\mathcal{F}_{t-1}\right) = \sum_{i=1}^{p} \omega_i G_i\left(x_t|x_{t-i}\right)$$

such that $\sum_{i=1}^{p} \omega_i = 1, \omega_i > 0,\ i = 1, 2, \cdots, p;\ F\left(x_t|\mathcal{F}_{t-1}\right)$ is the conditional cumulative distribution function (cdf) of X_t evaluated at x_t, given the past; $G_i(\cdot|x),\ i = 1, 2, \cdots, p$, is a Gaussian cumulative distribution function for each value of x.

By introducing more complicated parametrizations in above G_i's, a series of extended univariate time series models have been developed and found their applications in modeling a quite broad range of financial data. For example, they are able to model the daily stock prices [71], interest rate [45], S&P 500 stock index [72] and the U.S. annual consumer price inflation data [10], etc. The MTD-type models are particularly helpful in capturing features like bursts, flat stretches, heteroscedasticity and outliers.

Recently, we see quite extensive development of the MTD-type models. For example, in [51], the construction of stationary MTD models is considered; in [46], the MTD models within the Bayesian context are studied. Another direction is to extend the model to the bivariate situation. In [40], the univariate MTD model is generalized to the bivariate case. It mainly relies on the proposal of a particular class of bivariate distributions. In [7] a multivariate (Gaussian) MTD model under the mixed multivariate Gaussian situations is proposed and used to model bivariate stock price series.

The motivation of the multivariate MTD (MMTD) model in this Chapter is twofold. Firstly, in most existing literature, the models are exclusively built under the Gaussian distribution framework. This is no longer suitable for the time series derived from marked point process data, whose duration series are subjected to non-negative constrain. Moreover, for financial transaction durations, the dynamic range may be from 1 second to hundreds of seconds. Thus, moving away from Gaussian construction of the MTD-type models is quite necessary.

Secondly, in modeling marked point processes, the marks and points comprise a multivariate time series. Thus, it is necessary to consider building up a multivariate model for such kind of data. In this chapter, we generalize the univariate MTD-type models to the multivariate case and extend beyond the Gaussian distribution.

We observe that the multivariate MTD model to be discussed later provides simple, flexible and powerful ways to model marked point processes data. Moreover, it has the potential to model high-dimensional marks and allows flexible choices of dependence structures between the marks and the point process.

4.1.2 Copula and dependence

Before we get into the details of the new multivariate MTD model, we introduce a statistical technique that is used to build up multivariate distributions - the *Copula*. Copula is a useful tool to construct multivariate distributions with various dependence structures. One nice property of Copula is that one can model the marginal distributions and the dependence separately. The popularity of Copula in financial modeling starts from the 1990s, when it was used to model joint default risk for a portfolio.

In this subsection, we review some important definitions and properties of Copula, following Nelsen's [56] standard introduction and notations. Extensive applications of Copula in financial modeling can be found in [19].

Definition 4.1 *A Copula function $\mathscr{C}$ is a mapping $\mathscr{C}: I^2 \to I$, where $I = [0, 1]$, with the following properties*

- *For every $u, v \in I$*

$$\mathscr{C}(u, 0) = \mathscr{C}(0, v) = 0; \tag{4.1}$$

- *For every $u, v \in I$,*

$$\mathscr{C}(u, 1) = u; \quad \mathscr{C}(1, v) = v; \tag{4.2}$$

- *For every $u_1, u_2, v_1, v_2 \in I$ such that $u_1 \leq u_2$, $v_1 \leq v_2$,*

$$\mathscr{C}(u_2, v_2) - \mathscr{C}(u_2, v_1) - \mathscr{C}(u_1, v_2) + \mathscr{C}(u_1, v_1) \geq 0 . \tag{4.3}$$

The usefulness of Copula is based on a fundamental theorem by Sklar, which we called the *Sklar's fundamental theorem* for Copula.

Theorem 4.1 *(Sklar's Fundamental Theorem) Let H be the joint distribution function with marginals F and G. Then there exists a Copula $\mathscr{C}$ such that for all (x, y) in R^2,*

$$H(x, y) = \mathscr{C}\left(F(x), G(y)\right) \tag{4.4}$$

If F and G are continuous, then $\mathscr{C}$ is unique; Otherwise, $\mathscr{C}$ is uniquely determined by RanF(range of F) $\times$ RanG(range of G). Conversely, if $\mathscr{C}$ is a Copula and F and G are distribution functions, then the function H determined by (4.4) is a joint distribution function with marginals F and G.

By taking a partial derivatives on both hand sides of (4.4), the following corollary could be obtained.

Corollary 4.2 *For continuous random vector, the Copula density,* $c_\theta(u, v) := \frac{\partial^2 \mathscr{C}_\theta(u,v)}{\partial u \partial v}$ *is related to the density of the distribution H, denoted as h, by*

$$h(x, y) = c\left(F(x), G(y)\right) f(x) \cdot g(y). \tag{4.5}$$

where $f(x)$ *and* $F(x)$*(or* $g(y)$ *and* $G(y)$*)are the PDF and CDF for the random variable X(or Y).*

In many practical applications, the Copula family called "*Archimedean Copulas*" receives much attention because of its capacity to capture various dependence structures and many other nice properties. In our examples of new MMTD models, we illustrate them mostly with Archimedean Copulas. In what follows, we review their definitions following [56].

Definition 4.2 *Let* ϕ *be a continuous, strictly decreasing function,* $\phi\colon I \to [0, \infty]$*, such that* $\phi(1) = 0$*. The* pseudo-inverse *of* ϕ *is the function* $\phi^{-1}\colon [0, \infty] \to I$ *given by*

$$\phi^{[-1]}(t) = \begin{cases} \phi^{-1}(t) & \text{if } 0 \le t \le \phi(0); \\ 0 & \text{if } \phi(0) \le t \le \infty. \end{cases} \tag{4.6}$$

Lemma 4.3 *Let* ϕ *and* $\phi^{[-1]}(t)$ *defined as above. Let* $\mathscr{C} : I^2 \to I$ *given by*

$$\mathscr{C}(u, v) = \phi^{[-1]}\left(\phi(u) + \phi(v)\right). \tag{4.7}$$

Then $\mathscr{C}$ *satisfies the boundary conditions (4.1) and (4.2).*

Lemma 4.4 *Let* ϕ *and* $\phi^{[-1]}(t)$ *defined as above and let* $\mathscr{C} : I^2 \to I$ *defined as in the previous Lemma. Then* $\mathscr{C}$ *satisfies (4.3) if and only if for all* $v \in I$*, whenever* $u_1 \le u_2$*,*

$$\mathscr{C}(u_2, v) - \mathscr{C}(u_1, v) \le u_2 - u_1. \tag{4.8}$$

Theorem 4.5 *Let ϕ and $\phi^{[-1]}(t)$ and $\mathscr{C}: I^2 \to I$ defined as previous. Then the function $\mathscr{C}$ is a Copula if and only if ϕ is convex.*

Definition 4.3 *Copula of the form (4.7), provided that ϕ is convex, are called Archimedean Copulas. The function ϕ is called generator of the Copula. If $\phi(0) = \infty$, we say ϕ is a strict generator. In this case, we have $\phi^{[-1]}(t) = \phi^{-1}(t)$.*

More details, such as the proof of above lemmas and theorem, could be found in [56]. We take a look at a special Copula, the *product Copula*.

Example 4.1 *Let $\phi(t) = -ln(t)$ for any $t \in (0, 1]$. Because $\phi(0) = \infty$, thus it is strict. In this case, we can easily verify $\phi^{[-1]}(t) = \phi^{-1}(t) = e^{-t}$. The constructed Copula is thus given by*

$$\mathscr{C}(u, v) = e^{\{-[-ln(u)-ln(v)]\}} = uv. \tag{4.9}$$

This is called the product Copula. Further, it's trivial to show that bivariate random variable constructed using product Copula is independent.

Now, we turn to the dependence concept for Copula. Notice that the merit of the Copula largely relies on its ability to model various dependence structures. For Copula, a widely used scale-invariant measure of associations is the *Kendall*'s τ, given as follows.

Definition 4.4 *Kendall's τ is defined as the probability of concordance subtracted the probability of discordance,*

$$\tau = \tau_{X,Y} = P\{(X_1 - X_2)(Y_1 - Y_2) > 0\} - P\{(X_1 - X_2)(Y_1 - Y_2) < 0\}. \tag{4.10}$$

where (X_1, Y_1) and (X_2, Y_2) are independent identically distributed random vectors.

For the dependence (in terms of Kendall's τ) of the Archimedean Copulas, we have the following theorem [56].

Theorem 4.6 *Let X and Y be random variables with an Archimedean Copula $\mathscr{C}$ generated by ϕ. The Kendall's τ for X and Y, $\tau_{\mathscr{C}}$, is given by*

$$\tau_{\mathscr{C}} = 1 + 4 \int_0^1 \frac{\phi(t)}{\phi'(t)} dt. \tag{4.11}$$

Above brief review of Copula will greatly help in understanding the construction of our new multivariate MTD model given later. Now, we turn to another important issue of Copula.

In fact, there exist hundreds of Copulas proposed so far. Given so many Copulas, which Copula shall we choose? In practice, one generally fits a bunch of different Copulas and uses criterias such as AIC, BIC, or goodness-of-fit tests to select the best fitted Copula. Thus, choosing a Copula that describes the dependence structure properly would be an important issue and is well deserved for further investigations. In current research, our main emphasis is in the demonstration of the utility of the proposed multivariate MTD model. Thus we would limit our scope to the Archimedean Copulas. And in particular, we pick the *Clayton* Copula as our main illustration of some models. Other types of Copulas could be explored in the future.

4.2 The multivariate MTD model

Now, we introduce the multivariate MTD (MMTD) model for general multivariate time series, and for marked point processes data in particular. The general modeling framework is given as follows.

4.2.1 The model: a model for marked point process

Definition 4.5 *(Multivariate mixture transition distribution (MMTD) model) A time series* $\{\mathbf{x_t} \in \boldsymbol{R}^k, t = 1, 2, \cdots\}$ *is generated from a k-variate, p-component multivariate Mixture Transition Distribution (MMTD) model, if the conditional distribution of the observation at time t, namely,* $\mathbf{x_t} = (x_{1t}, \cdots, x_{kt})'$*, given the past is specified by*

$$F\left(\mathbf{x_t}|\mathbf{x^{t-1}}\right) = \sum_{i=1}^{p} \omega_i \mathscr{C}_i\left(\mathbf{x_t}|\mathbf{x^{t-1}}\right), \quad s.t. \sum_{i=1}^{p} \omega_i = 1; \quad \omega_i > 0, \quad i = 1, 2, \cdots, p \tag{4.12}$$

where

$$\mathscr{C}_i\left(\mathbf{x_t}|\mathbf{x^{t-1}}\right) = \mathscr{C}_{i,\boldsymbol{\theta}_i}\left(F_{i1}\left(x_{1t}|\mathbf{x^{t-1}}\right), \cdots, F_{ik}\left(x_{kt}|\mathbf{x^{t-1}}\right); \Theta_i\right) \tag{4.13}$$

is a certain k-variate Copula.

In (4.12), $F\left(\mathbf{x_t}|\mathbf{x^{t-1}}\right)$ is the conditional cumulative distribution function (cdf) of $\mathbf{X_t}$ evaluated at $\mathbf{x_t}$ given the past $\mathbf{X}^{t-1} = \mathbf{x}^{t-1} = (\mathbf{x}_1, \cdots, \mathbf{x}_{t-1})$; the multivariate distribution $\mathscr{C}_i\left(\mathbf{x_t}|\mathbf{x^{t-1}}\right)$ is a conditional cdf for the i^{th} mixture, which is to be constructed via Copula. In (4.13), $\mathscr{C}_{i,\boldsymbol{\theta}_i}(\cdot, \cdots, \cdot)$ is the i^{th} Copula with parameters $\boldsymbol{\theta}_i = (\theta_{i1}, \cdots, \theta_{iq_i})$; $F_{ij}\left(x_{jt}|\mathbf{x^{t-1}}\right), j = 1, 2, \cdots, k$ specifies the conditional distribution for the j^{th}-variate at time t given the past, within the i^{th} mixture; Θ_i represents a collection of all the parameters involving into the $F_{ij}\left(x_{jt}|\mathbf{x^{t-1}}\right)$'s,

$j = 1, 2, \cdots, k$. The Sklar's Fundamental Theorem (4.4) guarantees (4.13) to be well defined multivariate distributions.

This definition of the multivariate MTD model is very general. In application we expect that some simple models with certain unified structure are used. Thus, we consider the cases with following simplifications. Some examples will be given later.

s1. Same type of Copula for all mixture components, i.e., $\mathscr{C}_{i,\boldsymbol{\theta}_i}$'s are from a certain Copula and $\boldsymbol{\theta}_i = \boldsymbol{\theta}$ for all i's. Further, we focus on Copula from the one-parameter Archimedean Copulas, i.e., $\boldsymbol{\theta}_i = \theta$, which is a scalar parameter.

s2. In (4.13), let $F_{ij}\left(x_{jt}|\mathbf{x^{t-1}}\right)$ (for fixed $j = 1, 2, \cdots, k$) has the same class of univariate distribution over all $i = 1, 2, \cdots, p$. For example, $F_{i1}\left(x_{jt}|\mathbf{x^{t-1}}\right)$'s, $i = 1, 2, \cdots, p$ may all be Gamma distribution.

When specifying models from (4.12) and (4.13), it is convenient to take three successive steps — specify the marginal distribution; choose a Copula and specify the build-in lag information.

In particular, when $k = 2$ the model can be written as

$$F\left(x_t, y_t|x^{t-1}, y^{t-1}\right) = \sum_{i=1}^{p} \omega_i \mathscr{C}_\theta\left(F_{i1}\left(x_t|x^{t-1}, y^{t-1}\right), F_{i2}\left(y_t|x^{t-1}, y^{t-1}\right); \Theta_i\right) \tag{4.14}$$

where $(x^{t-1}, y^{t-1}) = \{(x_j, y_j), j = 1, \cdots, t-1\}$ is the past information up to time $t-1$.

If all probability density functions exist, we may rewrite (4.14) in a density form, which is convenient and useful in future computations.

$$f\left(x_t, y_t|x^{t-1}, y^{t-1}\right) = \sum_{i=1}^{p} \omega_i f_i(x_t, y_t|x^{t-1}, y^{t-1}; \theta, \Theta_i) \tag{4.15}$$

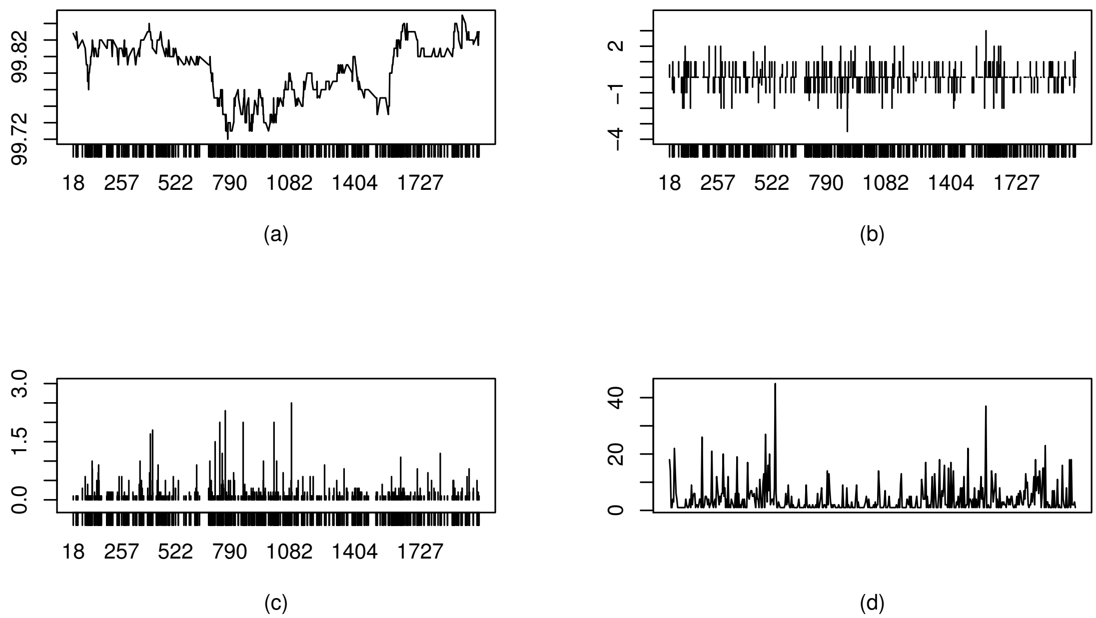

Figure 4.1: IBM stock transaction data - first 500 transactions starting from 12:30PM (where time stamp here is zero) on 02/06/2007. (a) transaction prices (in US dollars) versus time stamp; (b) (normalized) tick-returns (i.e., returns divided by minimum tick size 0.1) versus time stamp; (c) transaction volumes (in thousand shares) versus time stamp; (d) inter-trades durations (in seconds).

with

$$\begin{aligned} f_i(x_t, y_t | x^{t-1}, y^{t-1}; \theta, \Theta_i) &= f_{i1}(x_t | x^{t-1}, y^{t-1}) f_{i2}(y_t | x^{t-1}, y^{t-1}) \cdot \\ &\quad c_\theta \Big(F_{i1}(x_t | x^{t-1}, y^{t-1}), F_{i2}(y_t | x^{t-1}, y^{t-1}) \Big) \end{aligned} \tag{4.16}$$

where $c_\theta(u, v) = \frac{\partial^2 \mathscr{C}_\theta(u,v)}{\partial u \partial v}$ is the density function of Copula $\mathscr{C}_\theta(u, v)$. f_{i1}, f_{i2} are the density of F_{i1} and F_{i2} respectively. Notice that the parameters Θ_i appeared in (4.15) is implicit on the right hand side of equation (4.16).

In what follows, we discuss why the multivariate MTD model proposed will be a good candidate in modeling marked point processes. The ultra-high frequency financial transaction

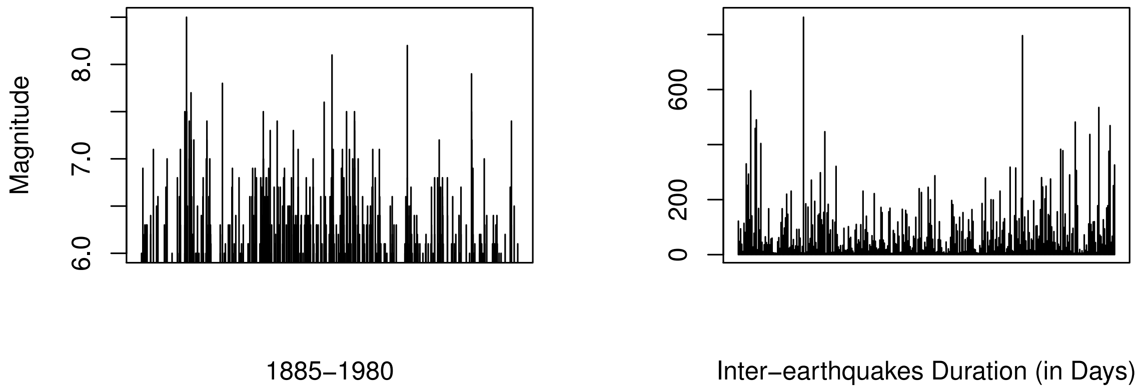

Figure 4.2: Earthquake record (1885-1980, magnitudes ≥ 6.0) in an offshore region east of Honshu and south of Hokkaido [58].

data is just a particular type of marked point process data. The most distinguished feature of such data is the irregularily-spaced transaction time.

To facilitate the discussion, we illustrate all ideas within certain real data context. Let T_i, $i = 0, 1, 2, ...$ be the time when the i^{th} transaction occurs (with $T_0 = 0$). We denote $X_{1i} = T_i - T_{i-1}$ to be the duration between two successive transactions. Let $(X_{2i}, \cdots, X_{ki})'$ represents the $k - 1$ dimensional marks at time T_i, which can be the associated transaction returns, traded volumes or other recorded quantities. Then, the transaction data can be fully represented by the sequences $\{\mathbf{X_t} = (X_{1t}, \cdots, X_{kt})'\}_{t=1}^{N}$.

We say that the financial transaction data is generated from a multivariate MTD model if the conditional distribution of $\mathbf{X_t}$ evaluated at $\mathbf{x_t}$ given the past follows (4.12) and (4.13). A natural constraint for such specification is that the first component of $\mathbf{X_t}$ (i.e., durations) should have non-negative support.

Figure 4.1 displays the typical multivariate time series from the transaction data set. It is a record of IBM stock transactions made on the New York Stock Exchange (NYSE) for a period of about 2000 seconds (i.e., about half an hour). We see the irregularly-spaced transaction time (see Figure 4.1 (a), (b) and (c); where in (d) the derived inter-trades durations are directly plotted) and the very frequent bursts or outliers for both transaction volumes and the tick-returns series (i.e., the difference between successive prices). Moreover, we notice that the inter-trades durations also display very similar features (bursts, jumps and outliers, etc.) and even show some sign of clustering effect.

An amazing fact is that for *most* marked point processes data met in practice, there exist many common features as those mentioned for the transaction data. The earthquake data is another example, which we will see later. For such kind of marked point processes data, to our best knowledge, very few statistical models are built so far to capture the underlying data generating processes. In particular, for the transaction data, very few statistical research is on describing the transaction dynamics simultaneously for the marks and points. Hassan and Lii [40] propose a bivariate MTD model as a first trial.

As mentioned before, apart from the transaction data, other data sets such as the earthquake data (see [13], [58] and Figure 4.2), sea wave data (see [40]) and precipitation data (see [34]), etc. all share very similar nonlinear non-Gaussian features mentioned above. A unified modeling framework to model such kind of marked point processes data will be very useful.

The univariate (Gaussian) MTD-type models have been proved to be able to capture such kinds of features. However, notice that for most data sets mentioned above, the marks and durations all belong to positive interval $(0, +\infty)$. Further, the marks and the points should be considered simultaneously. In other words, they are multivariate time series. Thus it is necessary to generalize the univariate MTD-type models to the multivariate situations as in (4.12) and (4.13). We will see later that many marked point processes data, especially the transaction data, can be easily modeled via the multivariate MTD models. Figure 4.3 shows the simulated transaction data via a specific multivariate MTD model.

In the remaining of this section, to illustrate the multivariate MTD modeling, we specify two types of models. They are for $k = 2$ (bivariate series) and $k = 3$ (trivariate series) respectively. In these examples, the multivariate MTD models all move beyond the Gaussian distribution and bring in different dependence structures by using different Copulas. We will see their realizations and applications in later sections.

Model A1 When $k = 2$, we can specify the i^{th} component in (4.16) as follows,

$$f_{i1}(x_t|x^{t-1}, y^{t-1}) = \frac{1}{\Gamma(\alpha_{i1})\beta_{i1}^{\alpha_{i1}}} x_t^{\alpha_{i1}-1} e^{-x_t/\beta_{i1}}, \quad x_t > 0. \tag{4.17}$$

with $\alpha_{i1} > 0$ as a scalar; $\beta_{i1} = \beta_{i1}^{*} e^{-(x_{t-i}+y_{t-i})}$, $i = 1, 2, \cdots, p$ as the built-in lag information. In modeling financial transaction data, the first variate in this bivariate sequence is generally used to represent the inter-trades duration. Similar, we specify

$$f_{i2}(y_t|x^{t-1}, y^{t-1}) = \frac{1}{\Gamma(\alpha_{i2})\beta_{i2}^{\alpha_{i2}}} y_t^{\alpha_{i2}-1} e^{-y_t/\beta_{i2}}, \quad y_t > 0. \tag{4.18}$$

with $\alpha_{i2} > 0$ as a scalar; $\beta_{i2} = \beta_{i2}^{*} e^{-\sqrt{x_{t-i}}}$, $i = 1, 2, \cdots, p$. This second variate is used to represent transaction volumes. Then, $\Theta_i = \left(\alpha_{i1}, \beta_{i1}^{*}, \alpha_{i2}, \beta_{i2}^{*}\right)$, $i = 1, 2, \cdots, p$ are model parameters.

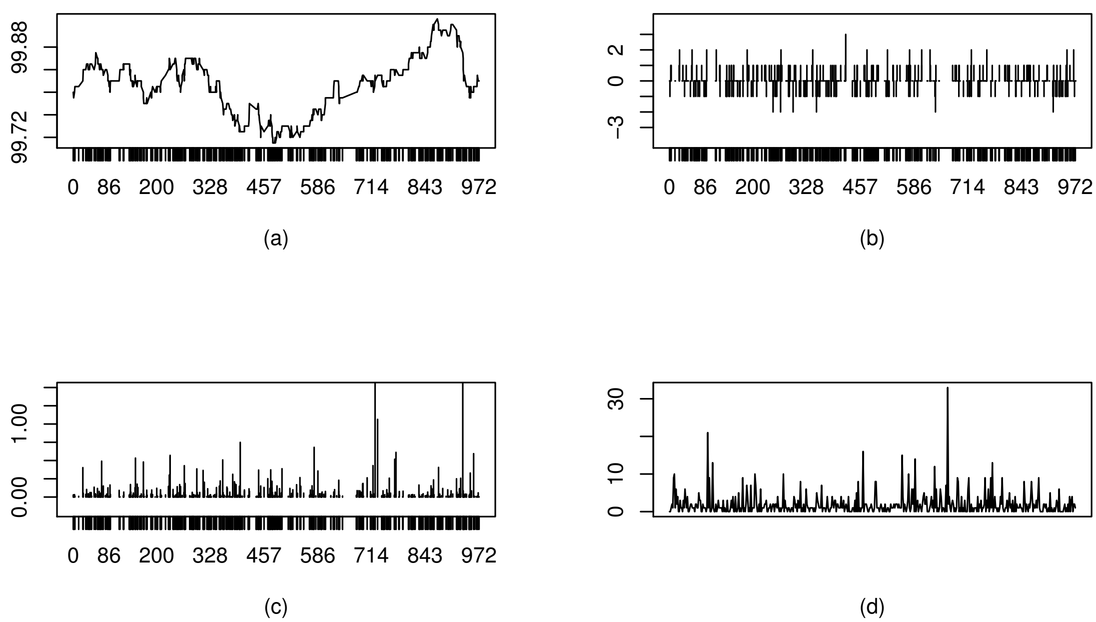

Figure 4.3: Simulated stock transaction data (500 observations of a period around 1000 seconds, i.e., 15 minutes) with Model B. (a) transaction prices versus time stamp; (b) (normalized) tick-returns versus time stamp; (c) transaction volumes (in thousand shares) versus time stamp; (d) inter-trades durations (in seconds).

- Similar to the discussion in [40], the built-in lag information can be arbitrary in principle so that very broad dependence, linear or non-linear, can be explored and constructed.

- On the other hand, the particular specifications of built-in lag information are also subjected to domain knowledge, model stability and the estimation concerns, etc. Similar arguments also apply to the construction of other models.

Last, we specify the *Clayton* Copula in (4.16), whose density is given by

$$c_\theta(u, v) = (1 + \theta)u^{-1-\theta}v^{-1-\theta}\left(-1 + \frac{1}{u^\theta} + \frac{1}{v^\theta}\right)^{-2-\frac{1}{\theta}}, \quad \theta \in (-1, \infty)\backslash\{0\}. \tag{4.19}$$

Model A1 will be used to study the efficiency of the estimation procedures that will be discussed later.

Model A2 Similar to Model A1, we modify the specification of the built-in lag information (through the parameterizations of β_{i1} and β_{i2}) as follows,

$$\beta_{i1} = \beta^*_{i1}\left(\frac{1 + y^{-1}_{t-i}e^{-y_{t-i}}}{1 + y^{1/3}_{t-i}x^{-1}_{t-i}e^{-x_{t-i}}}\right), \;\; \beta_{i2} = \beta^*_{i2}\left(\frac{1 + x^{-1}_{t-i}e^{-x_{t-i}}}{1 + x^{1/3}_{t-i}y^{-1}_{t-i}e^{-y_{t-i}}}\right), \;\; i = 1, 2, ..., p. \tag{4.20}$$

where we obtain a slightly different model, which we will use to model the real tick-by-tick transaction data.

Model B We consider a trivariate ($k = 3$) time series. This is because in modeling the transaction data, not only the inter-trades durations and transaction volumes but also the tick-returns are all of great interest. We specify in (4.12) using *Gaussian* Copula as follows,

$$\mathscr{C}_i\left(\mathbf{x_t}|\mathbf{x^{t-1}}\right) = \Phi_\Sigma\left(\Phi^{-1}\left(F_{i1}\left(x_{1t}|\mathbf{x^{t-1}}\right)\right), \Phi^{-1}\left(F_{i2}\left(x_{2t}|\mathbf{x^{t-1}}\right)\right), \Phi^{-1}\left(F_{i3}\left(x_{3t}|\mathbf{x^{t-1}}\right)\right)\right) \tag{4.21}$$

where Φ_Σ is the standardized multivariate Gaussian distribution with correlation matrix $\Sigma_{3\times 3}$; Φ^{-1} is the inverse of the univariate standard Gaussian cumulative distribution function Φ. For $i = 1, 2, \cdots, p$, we specify $f_{i1}(x_t|x^{t-1}, y^{t-1}, z^{t-1})$ and $f_{i2}(y_t|x^{t-1}, y^{t-1}, z^{t-1})$ similar to the specifications in Model A1, where the component $\{x_t\}$ is for durations and $\{y_t\}$ for transaction volumes. Finally, let $\{z_t\}$ represents the tick-returns,

$$f_{i3}(z_t|x^{t-1}, y^{t-1}, z^{t-1}) = \frac{1}{\sqrt{2\pi}\sigma_{i3}} \exp\left(\frac{z_t - \mu_{i3}}{2\sigma_{i3}^2}\right) . \quad (4.22)$$

We may specify the built-in lag information $\sigma_{i3} = \sigma_{i3}^* e^{-\sqrt{x_{t-i}}}$, for example. This very simple model will be used to simulate the transaction processes with two dimensional marks, i.e., transaction volumes $\{y_t\}$ and tick returns $\{z_t\}$.

4.2.2 Relationship with existing models

Due to the fact that the use of *product Copula* ($\mathscr{C}(u, v) = uv$) implies independence, the multivariate MTD model can be decomposed into independent univariate MTD-type models. Its relationship with most existing MTD-types models could be summarized in the following two results.

- The multivariate MTD model can be decomposed into univariate MTD-type models to obtain the MTD model by Le, Martin and Raftery [47] and all other extended models, such as Berchtold's [10] and Wong and Li's [71] [72], etc.

- The multivariate (Gaussian) MTD model by Bauwens, Hafner and Rombouts [7] can be constructed from the multivariate MTD modeling framework.

We verify above arguments by providing following two examples. At the same time, they also serve as examples of the MMTD models, in addition to those previously constructed models. On the other hand, we see that the MTD model, its extended models and the recent multivariate (Gaussian) MTD models [7] can be treated as special cases of the multivariate MTD models.

Example 4.2 *(A decomposition) Consider the bivariate case given in (4.16). Let's specify the following construction for the* i^{th} *component:*

$$f_{i1}(x_t|x^{t-1}, y^{t-1}) = \frac{1}{\sqrt{2\pi}\sigma_i} \exp\left(\frac{x_t - \phi_i x_{t-i}}{2\sigma_i^2}\right) \tag{4.23}$$

Similar construction applies to f_{i2}*. Moreover, in (4.16) we specify a product Copula, whose density function is given by* $c_\theta(u, v) = 1$*. Via this specific construction, the univariate time series* $\{X_t\}$ *and* $\{Y_t\}$ *are independent. Both follow the benchmark MTD model as described in [47].*

We may generalize this by specifying the i^{th} component in more complicated way. For example,

$$f_{i1}(x_t|x^{t-1}, y^{t-1}) = \frac{1}{\sqrt{2\pi}\sigma_i} \exp\left(\frac{x_t - \phi_{i0} - \sum_{j=1}^{p_i} \phi_{ij} x_{t-j}}{2\sigma_i^2}\right) \tag{4.24}$$

with similar construction applies to f_{i2}. The product Copula is used again. We then decompose the multivariate MTD model to obtain the Mixture Autoregressive (MAR) model as described in [71].

Example 4.3 *(Multivariate (Gaussian) MTD model) Bauwens, Hafner and Rombouts proposed a multivariate (Gaussian) MTD model [7] to model multivariate stock returns series.*

In their model, the dynamics of the k-dimensional vector time series $\{\mathbf{x_t} = (x_{1t}, \ldots, x_{kt})'\}_{t=1}^{N}$ *are specified by*

$$f\left(\mathbf{x_t}|\mathbf{x^{t-1}}\right) = \sum_{i=1}^{p} \omega_i f_i\left(\mathbf{x_t}|\boldsymbol{\mu}_i, \boldsymbol{\Sigma}_{it}\right) \tag{4.25}$$

where $f_i\left(\mathbf{x_t}|\boldsymbol{\mu}_i, \boldsymbol{\Sigma}_{it}\right)$ *are multivariate Gaussian densities with constant mean vector* $\boldsymbol{\mu}_i = \{\mu_{i1}, \cdots, \mu_{ik}\}$ *and time changing variance-covariance matrix* $\boldsymbol{\Sigma}_{it}$.

Within the general multivariate MTD modeling framework, for the i^{th} *component in the mixture, we specify in (4.12) with Gaussian Copula, i.e.,* $\mathscr{C}_i(\mathbf{u}; \boldsymbol{\Sigma}_{it}) = \Phi_{\boldsymbol{\Sigma}_{\mathbf{it}}}\left(\Phi^{-1}(u_1), \ldots, \Phi^{-1}(u_n)\right)$, *where* $\Phi_{\boldsymbol{\Sigma}_{it}}$ *is the multivariate Gaussian distribution with zero mean vector and time changing correlation matrix* $\boldsymbol{\Sigma}_{it}$ *and the* Φ^{-1} *is the inverse of the standard univariate Gaussian distribution function* Φ. *Further, we may let each marginal to be Gaussian distributed with mean from corresponding element of* $\boldsymbol{\mu}_i$ *and variance from the diagonal element of* Σ_{it}. *Then, with such specification, we reconstruct the multivariate (Gaussian) MTD model from the multivariate MTD modeling framework, due to the fact that "Gaussian Copula generates the multivariate standard Gaussian distribution if and only if the margins are standard Gaussian." . We also mention that the restriction (s1) mentioned before has to be dropped here, since model (4.25) takes time-varying Copula parameters,* Σ_{it}.

From previous discussions, we see that the multivariate MTD model may be seen as a natural multivariate extension of the univariate MTD-type models.

4.3 Parameter estimation and predictions

4.3.1 An EM algorithm

Generally, for the statistical models related to mixtures, a good candidate for parameter estimation is the Expectation Maximization (EM) algoritm [20]. For the estimation for MMTD model here, however, a direct EM algorithm will not work efficiently, if not impossible. This is largely due to the extremely complicated likelihood for all types of MMTD models constructed via Copula.

We propose an EM algorithm for the MMTD model. The new EM algorithm proposed here has a refined two-step optimization within the M-step. To simplify our notation, we derive and illustrate the new EM algorithm for the bivariate model as described in (4.15) and (4.16). It is obvious that the algorithm could be carried over to higher-dimensional cases without any difficulties.

Suppose the observed marked point process data $\{(x_t, y_t)\}_{t=1}^{n}$ is generated from model (4.15)-(4.16). We may define the "unobserved" random variables as follows,

$$\mathbf{Z} = (\mathbf{Z}_1, \mathbf{Z}_2, ..., \mathbf{Z}_n) = \begin{pmatrix} z_{1_1} & z_{2_1} & \dots & z_{n_1} \\ \dots & \dots & \dots & \dots \\ z_{1_p} & z_{2_p} & \dots & z_{n_p} \end{pmatrix} \tag{4.26}$$

where $\mathbf{Z}_t = (z_{t_1}, z_{t_2}, ..., z_{t_p})',\ t = 1, \dots, n$ is an indicator random variable, such that z_{t_j} equals to 1 if (x_t, y_t) is generated from the j^{th} component, and zero otherwise.

It can be shown that the probability that a certain observation comes from the j^{th} mixture component is given by,

$$P\left\{\mathbf{Z}_t = \left(z_{t_1}, z_{t_2}, \ldots, z_{t_j}, \ldots, z_{t_p}\right)' = \left(0, 0, ..., 1_{(j^{th})}, ..., 0\right)'\right\} = \omega_j, \quad j = 1, 2, ..., p. \tag{4.27}$$

Under the previous setup, we obtain the (conditional) log-likelihood for the complete data given as follows,

$$l(\mathbf{\Theta}|(\mathbf{X}, \mathbf{Y}), \mathbf{Z}) = \sum_{t=p+1}^{n}\sum_{j=1}^{p} z_{t_j} log\ \omega_j + \sum_{t=p+1}^{n}\sum_{j=1}^{p} z_{t_j} log\ f_j(x_t, y_t|x^{t-1}, y^{t-1}; \theta, \Theta_X, \Theta_Y) \tag{4.28}$$

In (4.28), $\mathbf{\Theta} = \left\{\left(\omega_1, \omega_2, ..., \omega_p\right), \Theta_X, \Theta_Y, \theta\right\}$ are the model parameters, where θ is the parameter for Copula and Θ_X and Θ_Y are the parameters being involved in the built-in lag information parametrization for $\{X_t\}$ and $\{Y_t\}$ respectively. For example, in Model A1, $\Theta_X = \left(\alpha_{i1}, \beta^*_{i1}; i = 1, 2, \cdots, p\right)$ and $\Theta_Y = \left(\alpha_{i2}, \beta^*_{i2}; i = 1, 2, \cdots, p\right)$. $(\mathbf{X}, \mathbf{Y}) = \{(x_t, y_t)\}_{t=1}^n$ are the observed data; $f_j(\cdot)$ is of the same form as in (4.16).

The EM algorithm runs iteratively via the following two steps:

E-Step let $\mathbf{\Theta}^{(m)}$ denote the estimates from the m^{th} iteration ($m = 1, 2, \cdots$). The conditional expectation of the j^{th} component of Z_t, $\tilde{z}_{t_j}$, given the observed data is

$$\tilde{z}_{t_j} = \frac{\omega_j f_j(x_t, y_t|x^{t-1}, y^{t-1}; \mathbf{\Theta}^{(m)})}{\sum_{i=1}^{p} \omega_i f_i(x_t, y_t|x^{t-1}, y^{t-1}; \mathbf{\Theta}^{(m)})}, \quad j = 1, 2, \cdots, p. \tag{4.29}$$

M-Step Given the the "missing data" $\{\tilde{\mathbf{z}}_t\}$, the M-Step aims to maximize the complete (conditional) log-likelihood function l over $\mathbf{\Theta}$. In (4.28), by taking the partial derivatives with respect to ω_j, we obtain a direct estimator,

$$\hat{\omega}_j = \frac{\sum_{t=p+1}^{n} \tilde{z}_{t_j}}{\sum_{t=p+1}^{n}\sum_{i=1}^{p} \tilde{z}_{t_i}}, \quad j = 1, 2, \cdots, p. \tag{4.30}$$

However, the full density form (4.15) and (4.16) would be extremely complicated, since the Copula density $c_\theta(u, v)$ itself would be complicated in general. Thus it is not feasible to optimize over the parameter space $(\Theta_X, \Theta_Y, \theta)$ simultaneously. Instead, A two-step optimization procedure within the M-step is proposed instead.

$\mathbf{M}^{(1)}$-step Maximize over each marginal, i.e.,

$$\hat{\Theta}_X = \arg\max_{\Theta_X} \left\{ \sum_{t=p+1}^{n} \sum_{j=1}^{p} \tilde{z}_{t_j} log f_{j1} \left(x_t | x^{t-1}, y^{t-1}; \Theta_X \right) \right\} \tag{4.31}$$

and

$$\hat{\Theta}_Y = \arg\max_{\Theta_Y} \left\{ \sum_{t=p+1}^{n} \sum_{j=1}^{p} \tilde{z}_{t_j} log f_{j2} \left(y_t | x^{t-1}, y^{t-1}; \Theta_Y \right) \right\} \tag{4.32}$$

$\mathbf{M}^{(2)}$-step Given the estimates from the $\mathbf{M}^{(1)}$-step, maximize with respect to the Copula parameter θ, where in our example is a scalar parameter.

$$\hat{\theta} = \arg\max_{\theta} \left\{ \sum_{t=p+1}^{n} \sum_{j=1}^{p} \tilde{z}_{t_j} log \left(c_\theta \left(F_{j1}(x_t | x^{t-1}, y^{t-1}; \hat{\Theta}_X), F_{j2}(y_t | x^{t-1}, y^{t-1}; \hat{\Theta}_Y) \right) \right) \right\} \tag{4.33}$$

Lastly, updated the parameter estimates $\mathbf{\Theta}^{(m+1)} = \left\{ \left(\hat{\omega}_1, \cdots, \hat{\omega}_p \right), \hat{\Theta}_X, \hat{\Theta}_Y, \hat{\theta} \right\}$ and run the E-Step and M-Step iteratively.

Above outlines an EM algorithm for the estimation of multivariate MTD models. We have following remarks.

- For each optimization problem involved, Newton-Raphson or other optimization schemes would suffice to find the optimal parameter values. For the initial values used in the EM algorithm, equal weights for ω's can be used. The parameters Θ_X and Θ_Y can be estimated initially by method of moments.

Table 4.1: Empirical means and standard deviations(SD) of the parameter estimates via EM algorithm for simulated data.

ω	Mean(sd)	α_{i1}	Mean(sd)	β^*_{i1}	Mean(sd)
0.40	0.4057(0.0217)	1.0	1.0065(0.0589)	0.3	0.3060(0.0269)
0.60	0.5943(0.0217)	2.0	2.0192(0.1200)	0.2	0.1981(0.0139)
α_{i2}	Mean(sd)	β^*_{i2}	Mean(sd)	θ	Mean(sd)
3.0	3.0137(0.2251)	0.2	0.2026(0.0200)	8.0	8.3218(0.3959)
2.0	1.9968(0.1228)	0.4	0.4026(0.0292)		

- The two-step optimization procedure within the M-Step is similar to the Inference Functions for Margins (IFM) method proposed by [43] for i.i.d. samples. This procedure is one of the widely used estimation methods for Copula. Our experiments show that the convergence of the proposed EM algorithm is quite fast in general.

- Most importantly, the proposed EM algorithm could be easily generalized to the cases when $k > 2$, as in the general MMTD model (4.12). For example, when $k = 3$, we need to consider the processes $\{X_t, (Y_t, Z_t)\}_{t=1}^n$, where X_t is the duration and (Y_t, Z_t) are the associated two-dimensional marks. With the new EM algorithm, only one optimization term is needed to add in the $\mathbf{M}^{(1)}$-step. Thus the complexity of the algorithm remains the same! We adopt this EM algorithm as a general scheme for the estimation problems in the MMTD modeling framework.

4.3.2 Simulation studies

Simulation studies are carried out to evaluate the performance of the proposed EM algorithm. We simulate 100 realizations from Model A1 ($p = 2$). Each sample path has 1000 observations. We run the EM algorithm iteratively until convergence. The parameters specified to generate the processes, the sample mean and standard deviation of the estimated parameters are summarized in Table 4.1. The result shows that the EM algorithm works very well. The estimated means are very closed to the true parameters with relatively small standard deviations.

We point out that for small value of the Copula parameter θ that characterizes the dependence, the estimate is not as precise as those for large θ's. For example, if we use $\theta = 2.5$ (and other parameters remain the same) to generate the processes, the empirical mean for the estimated θ value is 3.0518, with a standard deviation 0.2204. The estimate is out of the range of two standard deviation from the true value. Thus, it seems there might be room for further improvement. However, for *Clayton* Copula whose generator is $\phi(t) = \frac{1}{\theta}\left(t^{-\theta} - 1\right)$, the *Kendall*'s τ is given by $\tau(\theta) = \frac{\theta}{2+\theta}$. We see that $\left|\frac{\tau(3.0)-\tau(2.5)}{\tau(2.5)}\right| = 8\%$. Thus, the relative difference that describes the dependence, in terms of *Kendal*'s τ, is indeed quite small.

4.3.3 Prediction and visualization

For the prediction of marked point processes, an important concern is the *joint* prediction of marks and points. This is very meaningful for such data sets. For example, in the earthquake data (see Figure 4.2), given the past earthquake history, people are equally interested in both the time (i.e., point) and the magnitude (i.e., mark) for next earthquake. Moreover, in the context of MTD modeling, it is a well known fact that the conditional mean is no longer a proper predictor (see [72]). This is because the predictive distribution may be multimodal.

In view of above two reasons, we thus consider the concept of *Highest Density Region* (HDR) for the joint prediction problem. We suggest using the predictive HDR as the joint prediction region for the multivariate MTD models. In one-step ahead prediction, given the conditional distribution from the MMTD model, a simple scheme is to sample large enough i.i.d. observations from the given distribution and adapt those available multivariate density estimation schemes, such as [42], to construct the predictive HDR contour for next observation.

An illustration of the above discussed method will be given later by an example for the real transaction data. For the predictive HDR, a full predictive density can also be visualized via the estimation schemes [42]. This helps to provide detailed density information (see Figure 4.5).

The predictive HDR approach works efficiently for cases with $k \leq 3$, which are the major interest for most practical problems. As the dimension of the MMTD model increases, the computational time increases dramatically.

4.3.4 Model evaluation: empirical coverage

Notice that in univariate MTD-type models [47] [70] [71], an evaluation criteria for the model in terms of describing the data generating processes is the *empirical coverage*. We generalize this concept to the multivariate situations. The discussion in this subsection paves the way for the evaluation of different models for real transaction data.

For univariate time series, we define the empirical coverages of the prediction intervals (PI) to be the percentages of the data that falls within them. Thus, if the model adequately describes the underlying data generating process, the empirical coverages of those PI's should be close to the nominal coverages (see [47]). However, in the multivariate cases, or in particular, under the settings of bivariate time series, we generalize this concept to the prediction highest density regions (PHDR). Notice that, since the model involves mixture of distributions, the PHDR might be comprised of several disjoint regions.

To construct the prediction HDR given by the MMTD models, a formulation of the regions is generally analytically intractable. However, due to the easy simulation schemes for most Copulas, we can construct the prediction HDR via Monte Carlo sampling. Then our last problem turns to how to judge whether a data point falls into the prediction HDR or not. This could be done via the density quantile approach as described below.

We first sample enough i.i.d. bivariate observations (via the predictive distribution) to construct the $1 - \alpha$ density quantile [42]. To evaluate the empirical coverage of the PHDR, we compare the predictive density evaluated at each (x_t, y_t), i.e., $\tilde{f}\left(x_t, y_t | x^{t-1}, y^{t-1}\right)$, and the corresponding estimated $1 - \alpha$ predictive density quantile $\hat{f}_{1-\alpha}$. If the former is greater than the later, then we may conclude that the true observation is within the PHDR, and vice versa.

We report some simulation results that we carry out to investigate the efficiency of the methodology proposed above. In the simulation studies, we generate 30 sample paths from a bivariate time series $\{(x_t, y_t)\}_{t=1}^{1000}$ specified by Model A1. We estimate the model parameters and then calculate the prediction HDR and the empirical coverages. The result is summarized in Table 4.2. At the same time, we calculate the empirical coverage from the best fitted BMTD model for the simulated data (which is generated from the MMTD model). The result is given in In Table 4.3. Clearly, we see that the empirical coverages from the estimated MMTD model are very close to the nominal probabilities, while the fitted BMTD model fails to capture the underlying data generating processes. The empirical coverages for fitted BMTD model are far away from the nominal probabilities.

Table 4.2: Empirical coverages of one-step ahead PHDR from the fitted multivariate MTD model.

nominal probability	empirical coverage	(standard deviation)
0.4	0.36	(0.0197)
0.5	0.49	(0.0193)
0.6	0.61	(0.0205)
0.7	0.73	(0.0153)
0.8	0.83	(0.0100)
0.9	0.90	(0.0054)

Table 4.3: Empirical coverages of one-step ahead PHDR from the fitted BMTD model.

nominal probability	empirical coverage	(standard deviation)
0.4	0.53	(0.0185)
0.5	0.63	(0.0198)
0.6	0.71	(0.0201)
0.7	0.79	(0.0175)
0.8	0.86	(0.0126)
0.9	0.92	(0.0101)

Table 4.4: Parameters used to simulate a transaction data series with transaction durations, volumes and tick-returns.

ω	$(\omega_1, \omega_2, \omega_3)$ (0.40 , 0.30, 0.30)	α_1	$(\alpha_{11}, \alpha_{21}, \alpha_{31})$ (1.25, 2.0, 1.20)	α_2	$(\alpha_{12}, \alpha_{22}, \alpha_{32})$ (1.50, 1.00, 1.20)	β_1	$(\beta_{11}, \beta_{21}, \beta_{31})$ (0.15, 0.07, 0.14)
β_2	$(\beta_{12}, \beta_{22}, \beta_{32})$ (0.01, 0.10, 0.05)	μ_3	$(\mu_{13}, \mu_{23}, \mu_{33})$ (0, -1.0, 1.0)	σ_3^*	$(\sigma_{13}^*, \sigma_{23}^*, \sigma_{33}^*)$ (0.20, 0.40, 0.80)	Σ	$\begin{pmatrix} 1 & -.4 & -.4 \\ -.4 & 1 & -.4 \\ -.4 & -.4 & 1 \end{pmatrix}$

4.4 Numerical examples

In this section, to illustrate the utilities of the proposed MMTD model, we provide two numerical examples. In the first example, we simulate a MMTD process with two dimensional marks and show the similarities between its sample path and the real transaction data. In the second example, we fit the MMTD models for the real IBM transaction data and compare them with the benchmark bivariate MTD [40] models.

4.4.1 A simulated example

The multivariate MTD model provides a way to model the marked point processes with high-dimensional marks, which is beyond the capability from the univariate MTD-type models and the BMTD model. In this example, we simulate a marked point process from Model B, where the vector (X_t, Y_t, Z_t) corresponds to the duration, transaction volume and tick return. The parameters used to generate the transaction data are given in Table 4.4. Notice that this is a model with $p = 3$ and a simple Σ.

A realization from Model B is plotted in Figure 4.3. The prices, which start with initial price 99.80 and follow the simulated tick-returns, are constructed and plotted. Notice that, since the (normalized) tick returns are discrete, we round the simulated series to their nearest integer values.

Compare Figure 4.1 and Figure 4.3, we see that the model is capable to capture certain nonlinear features, such as frequent burst and flat stretches, etc. in the real transaction data. A careful evaluation of the modeling will be discussed in next example with real transaction data sets.

4.4.2 Modeling IBM tick-by-tick data

The data

In this example, we are going to model real tick-by-tick transaction data made on the New York Stock Exchange (NYSE). Three pieces of short time trading data (on Feb 6, 15 and 22 of 2007) for IBM stock are extracted from the TAQ database. Each piece of data is comprised of a full record of all transactions made during 12:30PM to 1:30PM. In Figure 4.1, only the first half of the Feb 6 data is plotted for easy visualization purpose.

We are particularly interested in modeling the transaction volumes and inter-trades durations. Among all quantities recorded in the TAQ database, the volumes and durations are two most important proxies for the stock market liquidity [59], which is an important risk measure for high frequency trading.

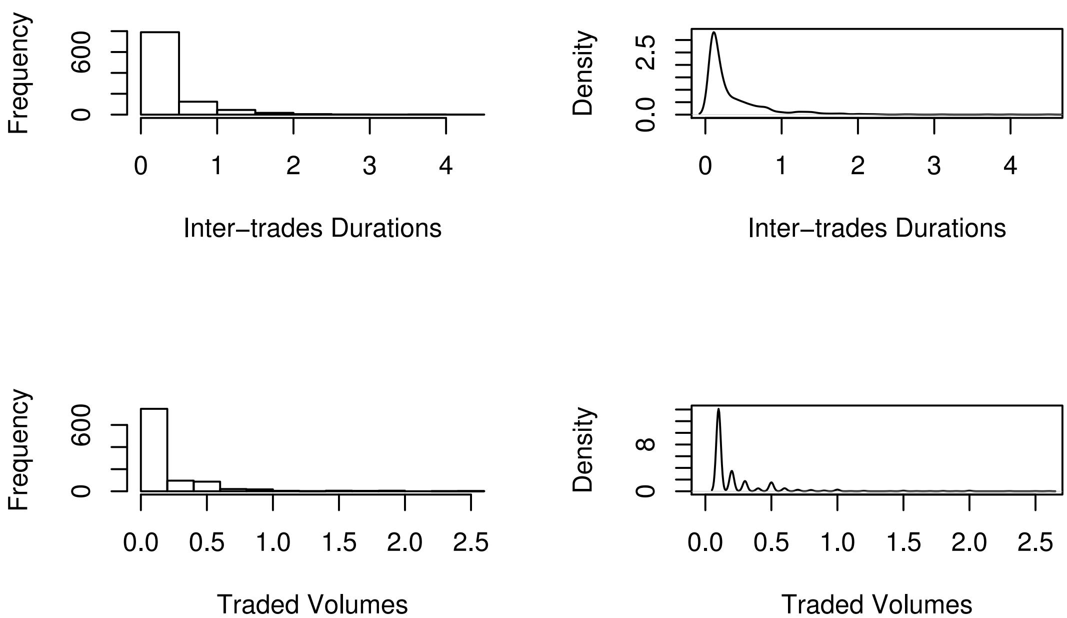

Figure 4.4: Unconditional marginal frequency histograms and smoothed densities estimated for inter-trades durations (in tens of seconds) and transaction volumes (in thousand of shares) of the IBM stock recorded from 12:30PM to 1:30PM on 02/06/2007.

Since the smallest time unit recorded is one second, there may exist "simultaneous" transactions (this is the case when the trading intensity is super high). [15] provides a comprehensive discussion of the data handling concerns for tick-by-tick transaction data. One natural way is to take the first recorded ones as the marked point processes observations. We particularly select the data from time 12:30PM to 1:30PM, a period when the trading activity is relatively much lower than other periods of the day so that there are few "simultaneous" transactions. Other natural ways may be choosing the last ones of those simultaneous observations or adding up the volumes (that being traded at the same time) and considering them as one transaction, etc.

Table 4.5: Estimations of the multivariate MTD and BMTD for the IBM Feb 06 data.

	$\hat{\boldsymbol{\omega}}$	$\hat{\boldsymbol{\alpha}}_1$	$\hat{\boldsymbol{\beta}}_1^*$	$\hat{\boldsymbol{\alpha}}_2$	$\hat{\boldsymbol{\beta}}_2^*$	$\hat{\theta}$
multivariate MTD	0.5229	0.8019	0.3153	0.6425	0.6574	0.8272
	0.4771	1.5656	0.0930	1.1664	0.1749	
	$\hat{\boldsymbol{\alpha}}$	$\hat{\boldsymbol{\beta}}$	$\hat{\boldsymbol{\theta}}$			
BMTD	0.2143	0.2464	5.4398			
	0.4700	0.0376	1.7319			
	0.3157	0.0362	1.6919			

Fitted Models

In Figure 4.4, the frequency histogram and the smoothed (unconditional) density are plotted for both the inter-trades durations and the traded volumes of the IBM TAQ data on Feb 6 (12:30-1:30PM). This might give us a hint that a mixture of Gamma distributions for both durations ([50], etc.) and volumes might be appropriate.

We fit Model A2 to the February 6 IBM stock transaction data. For the built-in lag information construction, we take into account of the model stability and simplicity. Surely, these can be other types of built-in lag information, linear or nonlinear. The domain knowledge, where in this example the market microstructure theory and empirical findings ([59], [35], etc.), would play a significant role.

We use the Bayes Information Criteria (BIC) to select the order of the models,

$$BIC = -2 \cdot \log(maximized\ likelihood) + K \cdot \log(n), \tag{4.34}$$

where K is the number of independent parameters to be estimated and n is the sample size.

The estimation is carried out via the proposed EM algorithm in Section 4.3. In terms of BIC values, the best fitted multivariate MTD model for the February 6 data is of order $p = 2$.

We also fit the BMTD [40] model to the February 6 data, the best model is of order $p = 3$. Estimation results for both models are given in Table 4.5.

Moreover, univariate ARIMA models are also fitted to the inter-trades durations and traded volumes individually. The best fitted model for the demeaned inter-trades durations is given by a ARIMA(1,0,1) model,

$$X_t - 0.9397X_{t-1} = Z_t - 0.8788Z_{t-1} \tag{4.35}$$

where Z_t is white noise series with variance 0.2016.

The best fitted model for the demeaned traded volumes series is given by a ARIMA(1,0,0) model,

$$X_t - 0.0061X_{t-1} = Z_t \tag{4.36}$$

where Z_t is white noise series with variance 0.08547.

By plotting a simulated series from both ARIMA models for the transaction data, we see that such kinds of linear models are no longer suitable since they can easily violate the nonnegativity requirement for both durations and volumes.

Thus, we would further compare the multivariate MTD models with the BMTD models. Table 4.6 summarizes a comparison between these two models used to fit all three pieces of IBM stock transaction data. In terms of the maximized log-likelihood and the BIC values, the preference for the multivariate MTD model over the BMTD model is quite strong.

Table 4.6: Model comparison for the IBM transaction data.

Data	Model	k	log-likelihood	BIC
Feb 6	BMTD(p=3)	9	-1275	2612
(12:30-1:30PM)	MMTD(p=2)	11	-349	774
Feb 15	BMTD(p=3)	9	-998	2056
(12:30-1:30PM)	MMTD(p=3)	16	-475	1056
Feb 22	BMTD(p=3)	9	-1079	2220
(12:30-1:30PM)	MMTD(p=2)	11	-77	230

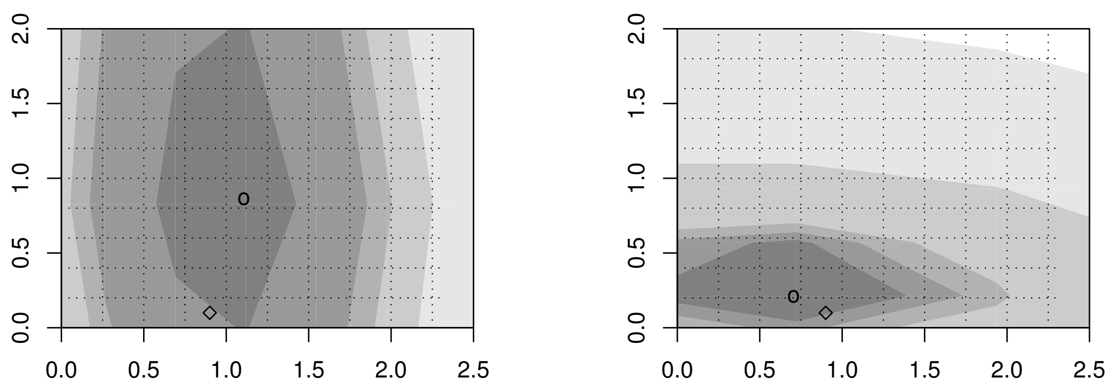

Figure 4.5: One-step ahead ($1-\alpha = 10\%,\ 30\%,\ 50\%,\ 70\%$) PHDR for an observation of the IBM Data (Left: BMTD model; Right: Multivariate MTD model). y-axis represents volumes (in thousand of shares); x-axis represents durations (in tens of seconds); the $\diamond$ represents the true observation; "o" represents the estimated highest density point; the shaded areas are the corresponding estimated HDR via [42].

Prediction & empirical coverage

Figure 4.5 displays a typical one-step ahead PHDR for a particular time point of the IBM Feb 6 data. We see that the true observation (denoted as $\diamond$) falls into the 30% prediction HDR for both models. A careful comparison shows that the area of the PHDR (10%, 30% HDR, for example) in the multivariate MTD models is much smaller than that in the BMTD model. The estimated highest density point in the multivariate MTD is much closer to the true observation than that of the BMTD model.

An overall prediction is studied by looking at the empirical coverages of the one-step ahead prediction for both BMTD and multivariate MTD models. We first sample enough i.i.d. bivariate observations (via the predictive distribution) to construct the $1 - \alpha$ density quantile [42]. To evaluate the empirical coverage of the PHDR, we compare the predictive density evaluated at each (x_t, y_t), i.e., $\tilde{f}\left(x_t, y_t | x^{t-1}, y^{t-1}\right)$ and the corresponding estimated $1 - \alpha$ predictive density quantile $\hat{f}_{1-\alpha}$. If the former is greater than the later, then we shall conclude that the true observation is within the PHDR, and vice versa.

We repeat the estimation of the empirical coverages of predicted HDR (i.e., the proportion of the data that falls into the nominal predicted HDR) for 50 times. If the model fits the data well, then the empirical coverages should be very close to the nominal probability coverages [47]. Table 4.7 shows the mean and standard deviation of each estimated empirical coverage. A comparison of the empirical coverages for these two models is given.

The results from these three transaction data sets show that the multivariate MTD model provide more accurate prediction than the BMTD model, in a sense that the multivariate

Table 4.7: Empirical coverages of the $(1-\alpha)$ PHDR's for the IBM transaction data (12:30PM - 1:30PM on 02/06/2007, 02/15/2007 and 02/22/2007).

Data	$1-\alpha$	.40	.50	.60	.70	.80	.90	.95
Feb 06	BMTD	0.34	0.55	0.68	0.78	0.86	0.93	.96
12:30-1:30PM	(sd)	(0.0046)	(0.0032)	(0.0034)	(0.002)	(0.0019)	(0.0020)	(0.0014)
	MMTD	0.38	0.51	0.62	0.74	0.84	0.90	0.94
	(sd)	(0.0029)	(0.0020)	(0.0026)	(0.0027)	(0.0021)	(0.0016)	(0.0011)
Feb 15	BMTD	0.40	0.55	0.67	0.77	0.85	0.94	0.97
12:30-1:30PM	(sd)	(0.0035)	(0.0033)	(0.0031)	(0.0031)	(0.0014)	(0.0015)	(0.0015)
	MMTD	0.42	0.53	0.65	0.73	0.82	0.89	0.92
	(sd)	(0.0033)	(0.0032)	(0.0037)	(0.0021)	(0.0019)	(0.0016)	(0.0016)
Feb 22	BMTD	0.40	0.56	0.68	0.78	0.86	0.94	0.97
12:30-1:30PM	(sd)	(0.0046)	(0.0033)	(0.0028)	(0.0028)	(0.0025)	(0.0009)	(0.0012)
	MMTD	0.37	0.47	0.60	0.72	0.82	0.90	0.93
	(sd)	(0.0026)	(0.0032)	(0.0034)	(0.0029)	(0.0020)	(0.0014)	(0.0014)

MTD predictive HDR has empirical coverages much closer to the nominal coverages than the BMTD model in general. We conclude that the multivariate MTD model captures the underlying data generating process quite successfully.

4.5 Conclusion and discussion

In this chapter, we proposed a new time series modeling framework by generalizing the univariate MTD-type models and the BMTD model to the multivariate cases. The multivariate MTD models can be used to model the marked point processes in general and the financial transaction data in particular. When bivariate series are considered, it provides an alternative construction to the BMTD model [40]. Further, an EM algorithm is proposed to solve the general estimation problem. Within the multivariate MTD modeling framework, we have very flexible choices of the building components, the parameterizations and the dependence structures between marks and the points. For further development, autoregressive types of parameterizations, time varying Copula parameters and discrete marginals (so as to model tick-returns, for example) may be considered. The MMTD modeling framework proposed will be very useful in modeling various kinds of marked point processes data sets we meet in practice.

Chapter 5

Future work

We have generalized the MTD-type models to two new time series models, the MLMAR mode and the MMTD model. We applied the models to the financial transaction data at two different levels - the *high frequency* and *ultra-high frequency* levels. The models are shown to be successful in terms of either describing the underlying data generating processes or prediction performance. However, there exist some statistical modeling issues and open problems to be further investigated in the future.

Firstly, in both MLMAR and MMTD modeling frameworks, how to specify a model, i.e., the model selection beyond the BIC/AIC criteria, still remains as an open problem. Notice that very recently, Naik, Shi and Tsai [54] has brought up this issue and tried to solve it for the Gaussian situation. Moreover, for the MMTD model, further research on how to selecting a suitable Copula should be a well deserved task. This may be done via certain tests of hypothesis.

Secondly, we may consider building the MMTD model with discrete marginals. This will be very meaningful. Since most time series data we met in practice are discrete. For example, the tick returns (i.e., the price difference between two successive transactions) in the transaction data sets take only a handful of integer values. To model the prices, we also need to consider discrete marginal for the MMTD models. Moreover, it will be very interesting to model marked point processes with high-dimensional marks. For example, we may consider the earthquake location, which is determined by its latitude and longitude, as two dimensional mark.

Lastly, the MMTD model indeed provides a general modeling framework for all marked point processes. Thus, it will be very interesting to see whether other types of marked point processes data (for example, earthquake data) can be modeled and forecasted accurately. This is undoubtedly a highly rewarding task.

Bibliography

[1] A. Amendola and C. Francq. Concepts and tools for nonlinear time series modelling. Technical report, MPRA Paper No. 16668, 2009.

[2] O. E. Barndorff-Neilsen and N. Shephard. Econometric analysis of realized volatility and its use in estimating stochastic volatility models. *Journal of the Royal Statistical Society: Series B*, 64:253–280, 2002.

[3] O. E. Barndorff-Neilsen and N. Shephard. Econometric analysis of realized covariation: high frequency based covariance, regression, and correlation in financial economics. *Econometrica*, 72:885–925, 2004.

[4] D. M. Bashtannyk and R. J. Hyndman. Bandwidth selection for kernel conditional density estimation. *Computational Statistics and Data Analysis*, 36:279–298, 2001.

[5] L. Bauwens and P. Giot. *Econometric modelling of stock market intraday activity*. Kluwer Academic Publishers, 2001.

[6] L. Bauwens and P. Giot. Asymmetric acd models: introducing price information in acd models. *Empirical Economics*, 28:1–23, 2003.

[7] L. Bauwens, C. M. Hafner, and J. V. K. Rombouts. Multivariate mixed normal conditional heteroskedasticity. *Computational Statistics and Data Analysis*, 51:3551–3566, 2007.

[8] L. Bauwens and N. Hautsch. Modelling financial high frequency data using point processes. *Handbook of Financial Time Series (Ed. Andersen, et al.)*, pages 953–979, 2009.

[9] L. Bauwens, W. Pohlmeier, and D. Veredas, editors. *High frequency financial econometrics: recent developments*. Physica-Verlag, 2007.

[10] A. Berchtold. Mixture transition distribution (mtd) modeling of heteroscedastic time series. *Computational Statistics and Data Analysis*, 41:399–411, 2003.

[11] T. Bollerslev. Generalized autoregressive conditional heteroskedasticity. *Journal of Econometrics Reviews*, 31:307–327, 1986.

[12] G. E. Box, G. M. Jenkins, and G. C. Reinsel. *Time Series Analysis: Forecasting and Control (4th Ed.)*. Wiley, 2008.

[13] D. Brillinger. Time series, point processes, and hybrids. *Canadian Journal of Statistics*, 22:177–206, 1994.

[14] P. J. Brockwell and R. A. Davis. *Time Series: Theory and Methods (2nd Ed.)*. Springer, 1991.

[15] C. T. Brownlees and G. M. Gallo. Financial econometric analysis at ultra-high frequency: Data handling concerns. *Computational Statistics and Data Analysis*, 51:2232–2245, 2006.

[16] S. Centanni and M. Minozzo. Estimation and ltering by reversible jump mcmc for a doubly stochastic poisson model for ultra-high-frequency nancial data. *Statistical Modelling*, 6:97–118, 2006.

[17] S. Centanni and M. Minozzo. A monte carlo approach to filtering for a class of marked doubly stochastic poisson processes. *Journal of the American Statistical Association*, 101, 2006.

[18] F. Chapeau-Blondeau and A. Monir. Numerical evaluation of the lambert w function and application to generation of generalized gaussian noise with exponent 1/2. *IEEE TRANSACTIONS ON SIGNAL PROCESSING*, 50:2160–2165, 2002.

[19] U. Cherubini, E. Luciano, and W. Vecchiato. *Copula Methods in Finance*. Wiley, 2004.

[20] A. P. Demspter, N. M. Laird, and D. B. Rubin. Maximum likelihood from incomplete data via the em algorithm. *Journal of the Royal Statistical Society: Series B*, 39:1–38, 1977.

[21] M. N. Do and M. Vetterli. Wavelet-based texture retrieval using generalized gaussian density and kullbackleibler distance. *IEEE TRANSACTIONS ON IMAGE PROCESSING*, 11:146–158, 2002.

[22] A. Dufour and R. F. Engle. Time and the price impact of a trade. *Journal of Finance*, 55:2467–2498, 2000.

[23] C. L. Dunis and B. Zhou, editors. *Nonlinear modelling of high frequency financial time series*. Wiley, 1998.

[24] R. F. Engle. Autoregressive conditional heteroscedasticity with estimates of the variance of united kingdom inflations. *Econometrica*, 50:987–1007, 1982.

[25] R. F. Engle. The econometrics of ultra-high frequency data. *Econometrica*, 68:1–22, 2000.

[26] R. F. Engle and J. R. Russell. Autoregressive conditional duration: A new model for irregularly spaced transaction data. *Econometrica*, 66:1127–1162, 1998.

[27] R. F. Engle and J. R. Russell. Analysis of high frequency financial data. *Handbook of Financial Econometrics (Ed. Ait-Sahalia, Y. and L. P. Hansen, L. P.)*, 2004.

[28] J. Fan and Q. Yao. *Nonlinear Time Series - Nonparametric and Parametric Methods*. Springer-Verlag, 2003.

[29] S. K. Fan and Y. Lin. A fast estimation method for the generalized gaussian mixture distribution on complex images. *Computer Vision and Image Understanding*, 113:839–853, 2009.

[30] J. Gao. *Nonlinear Time Series: Semiparametric and Nonparametric Methods*. Chapman & Hall / CRC, 2007.

[31] C. A. E. Goodhart and L. Figliuoli. Every minute counts in financial markets. *Journal of International Money and Finance*, 10:23–52, 1991.

[32] C. A. E. Goodhart and M. O'Hara. High frequency data in financial markets: issues and applications. *Journal of Empirical Finance*, 4:73–114, 1997.

[33] C. W. J. Granger and A. P. Andersen. *An introduction to bilinear time series models*. Vanenhoek and Ruprecht, 1978.

[34] P. Guttorp. *Stochastic modeling of scientific data*. Chapman and Hall, 1995.

[35] C. M. Hafner. Durations, volume and the prediction of nancial returns in transaction time. *Quantitative Finance*, 5:145–152, 2005.

[36] V. Haggan and T. Ozaki. Modeling nonlinear random vibrations using an ampitude-dependent autoregressive time series model. *Biometrika*, 68:189–196, 1981.

[37] J. D. Hamilton. *Time Series Analysis*. Princeton University Press, 1994.

[38] L. Harris. Transaction data tests of the mixture of distribution hypothesis. *Journal of Financial and Quantitative Analysis*, 22:127–141, 1987.

[39] J. Hasbrouck. *Empirical Market MIcrostructure*. Oxford University Press, 2007.

[40] M. Y. Hassan and K. S. Lii. Modeling marked point processes via bivariate mixture transition distribution models. *Journal of the American Statistical Association*, 101:1241–1252, 2006.

[41] R. Hujer and S. Vuletic. Econometric analysis of financial trade processes by discrete mixture diuration models. *Journal of Economic Dynamics and Control*, 31:635–667, 2007.

[42] R. J. Hyndman. Computing and graphing highest density region. *The American Statistician*, 50:120–126, 1996.

[43] H. Joe and J. Xu. The estimation method of inference functions for margins for multivariate models. Technical report, University of British Columbia, 1996.

[44] K. Kokkinakis and A. K. Nandi. Exponent parameter estimation for generalized gaussian probability density functions with application to speech modeling. *Signal Processing*, 85:1852–1858, 2005.

[45] M. Lanne and P. Saikkonen. Modeling the u.s. short-term interest rate by mixture autoregressive processes. *Journal of Financial Econometrics*, 1:96–125, 2003.

[46] J. W. Lau and M. K. P. So. Bayesian mixture of autoregressive models. *Computational Statistics and Data Analysis*, 53:38–60, 2008.

[47] N.D. Le, D. Martin, and A. E. Raftery. Modeling flat stretches, bursts, and outliers in time series using mixture transition distribution models. *Journal of the American Statistical Association*, 91:1504–1515, 1996.

[48] G. Luca and G. Gallo. Mixture processes for financial intradaily durations. *Studies in Nonlinear Dynamics and Econometrics*, 8, 2004.

[49] G. Luca and G. Gallo. Time-varying mixing weights in mixture autoregressive conditional duration models. *Econometrics Reviews*, 28:102–120, 2009.

[50] G. D. Luca and P. Zuccolotto. Finite and infinite mixture for financial durations. *METRON - International Journal of Statistics*, LXI:431–455, 2003.

[51] R. H. Mena and S. G. Walker. Stationary mixture transition distribution models via predictive distributions. *Journal of Statistical Planning and Inference*, 137:3103–3112, 2007.

[52] X. Meng and D. Rubin. Maximum likelihood estimation via the ecm algorithm: a general framework. *Biometrika*, 80:267–278, 1993.

[53] S. Nadarajah. A generalized normal distribution. *Journal of Applied Statistics*, 32:685–694, 2005.

[54] Naik, Shi, and Tsai. Extending the akaike information criterion to mixture regression models. *Journal of the American Statistical Association*, 102:244–254, 2007.

[55] M. Nardon and P. Pianca. Simulation techniques for generalized gaussian densities. Technical report, University of Venice, 2006.

[56] R. B. Nelsen. *An introduction to copulas (2nd Ed.)*. Springer, 2006.

[57] D. F. Nicholls and B. G. Quinn. *Random coefficient autoregressive models: an introduction*. Springer-Verlag, 1982.

[58] Y. Ogata. Statistical models for earthquake occurrence and residual analysis for point processes. *Journal of the American Statistical Association*, 83:9–27, 1988.

[59] M. O'Hara. *Market microstructure theory*. Wiley-Blackwell, 1998.

[60] R. C. A. Oomen. Statistical models for high frequency security prices. In *EFA 2003 Annual Conference Paper No. 105*, 2002.

[61] D. Pena, G. C. Tiao, and R. S. Tsay, editors. *A course in time series analysis*. Wiley, 2001.

[62] A. E. Raftery. A model for high-order markov chain. *Journal of the Royal Statistical Society: Series B*, 47:528–539, 1985.

[63] T. H. Rydberg and N. Shephard. A modelling framework for the prices and times of trades made on the new york stock exchange. Technical report, Nuffield College, Oxford University, 1999.

[64] T. H. Rydberg and N. Shephard. Dynamics of trade-by-trade price movements: decomposition and models. *Journal of Financial Econometrics*, 1:2–25, 2003.

[65] H. Tong. *Threshold models in nonlinear time series analysis*. Springer-Verlag, 1983.

[66] H. Tong. *Non-linear Time Series - A Dynamical System Approach*. Oxford University Press, 1990.

[67] H. Tong and K. S. Lim. Threshold autoregression, limit cycles and cyclical data (with discussion). *Journal of the Royal Statistical Society: Series B*, 42:245–292, 1980.

[68] C. S. Wong. *Statistical inference for some nonlinear time series models*. PhD thesis, Unversity of Hong Kong, 1998.

[69] C. S. Wong, W. S. Chan, and P. L. Kam. A student t-mixture autoregressive model with applications to heavy-tailed financial data. *Biometrika*, 96:751–760, 2009.

[70] C. S. Wong and W. K. Li. On a logistic mixture autoregressive model. *Biometrika*, 88:833–846, 2001.

[71] C.S. Wong and W.K. Li. On a mixture autoregressive model. *Journal of the Royal Statistical Society: Series B*, 62:95–115, 2000.

[72] C.S. Wong and W.K. Li. On a mixture autoregressive conditional heteroscedastic model. *Journal of the American Statistical Association*, 96, 2001.

[73] M. Y. Zhang, J. R. Russell, and R. S. Tsay. A nonlinear autoregressive conditional duration model with applications to fnancial transaction data. *Journal of Econometrics*, 104:179–207, 2001.