Modeling Multivariate Ultra-High-Frequency Financial Data by Monte Carlo Simulation Methods

by

TingTing Peng

Thesis submitted for the Doctor of Philosophy

in Economics and Finance

Supervisor: Marco Minozzo

Department of Economics

University of Verona

March 12, 2011

Acknowledgements

I owe a great deal of intellectual debt to my supervisor, Marco Minozzo, who deserves special thanks not only for his advising, but also for his unwavering support and his seemingly inexhaustible supply of good ideas. In addition, I thank our influential professors in Department of Economics, University of Verona, including Diego Lubian, Laura Magazzini, Federico Perali, Claudio Zoli, Angelo Zago, Luca Zarri, Eugenio Peluso. A particular thanks goes to Silvia Centanni, who opened my mind to this research topic and provided me with encouragement to continue this work. Sincere thanks go also to our coordinator, Giam Pietro Cipriani, who has been always supportive of this research and has been so helpful to solve the problem I met during these years.

I also want to thank my previous supervisor, Elisabetta Croci Angelini, for her continued support. Special Thanks are also due to Federico Zaniolo who was supportive and helpful by introducing me to his previous research, upon which I relied when dealing with ultra-high-frequency data. I also owe speical note of thanks to my colleagues and friends: Elisa Cavatorta, Francesco Andreoli, Dario Girardi, Silvia Domeneghetti, Andrea Bonfatti, Doriano Benedetti, Oli Chen, Tengfang Ling, Angelo Hu, Amy Lin, and Laura Visconti. Without their help, this work could not be finished in time.

Finally, I thank my parents, to whom this work is dedicated, for enabling and supporting my education and for providing me with the ‘stubborness gene’ necessary for graduate school.

TingTing Peng

Abstract

In this thesis, we propose a modeling framework for multivariate ultra-high-frequency financial data. The proposed models belong to the class of the doubly stochastic Poisson processes with marks which are characterized by the number of events in any time interval to be conditionally Poisson distributed, given another positive stochastic process called intensity. The key assumption of these models is that the intensities are specified through a latent common dynamic factor that jointly drives their common behavior. Assuming the intensities are unobservable, we propose a signal extraction (filtering) method based on the reversible jump Markov chain Monte Carlo algorithm. Our proposed filtering method allows to filter not only the intensities but also their specific and common components.

From an empirical stand point, on the basis of a comparison of real data with Monte Carlo simulated data, obtained under different assumptions for ticks (times and logreturns), based mainly on the behavior of the correlation between pairs of assets as a function of the sampling period (Epps effect), we found evidence for the existence of a single latent common factor responsible for the behavior observed in a set of assets from the Borsa di Milano.

Contents

List of Tables

List of Figures

Introduction

Ultra-high-frequency data (UHF), also known as tick-by-tick data, are available for most exchange markets of financial assets and also exist for stocks such as the New York Stock Exchange, the Paris Bourse, the London Stock Exchange or the Frankfurt Stock Exchange. By nature these data are irregularly spaced in time. Liquid markets generate hundreds or thousands of ticks per business day. Data vendors like Reuters transmit more than 275,000 prices per day for foreign exchange spot rates alone. In contrast to standard financial databases, which usually provide information on a daily or weekly basis (usually, the closing prices and traded volumes), UHF databases provide an enormous amount of information about the intraday behavior of financial markets. Indeed, they provide the finest possible information recording the time at which a market event, such as a trade or a quote update by a market maker or computer system, takes place together with its associated characteristics (for instance, the price and volume of the transaction) for all transactions or quotations. As a consequence, a systematic modeling of UHF data should be the primary object for researchers and practitioners who are interested in understanding financial markets.

In the econometrics literature, the availability of high-frequency, or intraday, data has opened an active field of research since the 1990s. Broadly speaking, there is just one main class of high-frequency models. The ARCH model of Engle (1982) and its generalized version (GARCH) by Bollerslev (1986) have been viewed as successful modeling frameworks for intraday data. In particular, they provide the basic framework for empirical studies of the market microstructure in applied econometric research. However, the main drawback behind this class of models is that they deal with regularly time-spaced data. With the advent of irregularly spaced UHF data, time transformations are needed to convert the original (irregularly spaced) data into regularly time-spaced data at a given frequency. Once the data have been sampled, the GARCH models can be applied. Obviously, in this manner, we will lose a considerable amount of information.

To overcome this loss of information, in the last decade models based on a marked point processes (MPPs) have been introduced for UHF data. In general, an MPP, denoted as $(T_i, Z_i)_{i\in\mathbb{N}}$, is a sequence of random times T_i (point process), representing the time instants at which the events (trades or quotations) occur, complemented by a sequence of random vectors Z_i, one for each random time, containing the values taken by some variables (such as changes of prices, bid-ask spreads or volumes), called the marks, describing the occurrence of the event. Since the random times T_i can take positive values in the real line, these processes can be used to characterized irregularly time-spaced data, overcoming the difficulties encountered with previous models. From a mathematical point of view, the UHF information relative to a given stock can be viewed as a trajectory of an MPP.

In the literature, two main classes of models based on MPPs have hitherto been pro-

posed: the autoregressive conditional duration (ACD) model introduced by Engle & Russell (1998), and the class of models based on doubly stochastic Poisson processes (DSPPs) with marks (see Frey, 2000, and Rydberg & Shephard, 2000). The basic idea behind the ACD model of Engle & Russell (1998) is to directly model the time intervals (also called durations) between market events such as the occurrence of a trade or a quote in a trading environment. A review of the original ACD model of Engle & Russell (1998) and of the logarithmic versions of it are provided in Chapter 2. In contrast, the second class of models provides a direct way to model the times of events rather than the durations. For example, Rydberg & Shephard (2000) proposed a model for the prices and times of trades based on marked DSPPs. Basically, the second class of models specifies the distribution of the number of events in any time interval by using another positive stochastic process λ called intensity. In particular, conditional on λ, the number of events in any given time interval is Poisson distributed, whereas the number of events in two disjoint time intervals are independent. More theoretical details about marked DSPPs are given in Chapter 2.

Both the previous two classes of UHF models, the ACD models and the models based on marked DSPPs, have been developed for univariate financial data. Since correlation measures for multivariate UHF data have direct applications in derivatives pricings, risk management and portfolio optimization, in the last years, a growing interest has undergone in modeling multivariate UHF data. The simplest way to model multivariate UHF data is to extend the existing univariate models. Given the assumption of the ACD models that the information set on which the duration is conditional cannot be updated within a duration spell, the extension of the ACD models encounters difficulty. This is because we need to update simultaneously the information set of different types of trading events that occur asynchronously for multivariate data. Alternatively, intensity-based DSPP models have been so far proposed such as the autoregressive conditional intensity (ACI) model of Russel (1999) in which he specifies a bivariate model of transaction arrival times and limit order submission arrival times (Chapter 2 provides a thorough review of ACI models). On this line of research, Hall & Hautsch (2004) estimated a bivariate ACI model for the arrival of buy and sell trades on a limit order book market, whereas Bauwens & Hautsch (2003) propose an extension of the ACI model which adds a latent Guassian autoregressive component to the log intensity.

Interestingly, many studies in the modeling of multivariate UHF financial data through point processes provide evidence for common movements and strong interdependencies among different assets. Russell (1999) found codependence in the processes of market order arrivals and limit order arrivals in NYSE trading. Bowsher (2002) provided evidence for a common pattern in trading intensities and price intensities based on NYSE transaction data. Spierdijk, Nijman, & van Soest (2002) showed significant co-movements in the trading intensities of US department store stocks, which are confirmed by Heinen & Rengifo (2003). Hall & Hautsch (2004, 2006) found co-movements in the arrival rates of market orders, limit orders and cancellations on the different sides of the market in Australian Stock Exchange. Furthermore, Bauwens & Hautsch (2006) provided significant evidence for the existence of a persistent common component that jointly drives the individual processes.

As mentioned above, correlation measures for multivariate UHF data play an important role in risk management, portfolio optimization, derivative pricing and so on. However, correlations calculated using a narrow time interval are heavily affected by the asynchronous feature of the data. Epps (1979) observed that ‘correlation among price changes

in common stocks of companies in one industry are found to decrease with the length of the interval for which the price changes are measured'. This feature of UHF data is the so-called 'Epps effect' which is a part of manifestation of the information aggregation process in stock markets. Trading in any given stock provides information signal and is particularly related to the news released for the company of interest. Indeed, there is a time lag between the moment in which the piece of news is released and the subsequent trading in any other relevant stock. More details about this topic are referring to Chapter 1.

In this thesis, we propose and apply a new type of multivariate dynamic intensity model, which allows to capture the common behavior in a set of assets. In particular, we consider a class of multivariate marked doubly stochastic Poisson processes in which the intensity of each marked DSPP is decomposed into two components: an individual specific component and a common component. This modeling framework allows to account for the asynchronicity in the arrivals of events in a set of assets. Our proposed model is detailed in Chapter 3. Our work extends the univariate model proposed by Centanni & Minozzo (2006) by allowing for a common component jointly driving all the intensities. Our model could be compared with the stochastic conditional intensity model (SCI) proposed by Bauwens & Hautsch (2006), which is based on an observation-driven component (observable) and a dynamic latent component (unobservable). The latent component in the SCI model is specified as a log-linear model based on a Gaussian autoregressive process of the first order, while the unobservable components in our model follow a shot noise process. Although the SCI model is different from our model, the formular shares with the latter in the presence of a common factor driving all the intensities of the univariate MPPs, that is, modeling the joint dynamic of the multivariate system.

For our model, we provide a filtering method (a signal extraction method) based on the reversible jump Markov chain Monte Carlo algorithm (RJMCMC) which has been introduced by Green (1995) in a Bayesian inferential context for the simulation of the posterior distribution. In our proposed latent factor model, we assume that the underlying intensities are unobservable, together with their individual and common components. Thus their conditional and unconditional moments cannot typically be computed analytically, and we must resort to Monte Carlo simulation methods. Among many others, Richardson & Green (1997), Dellaportas et al. (1998), Denison et al. (1997), Troughton & Godsill (1997), Insua & Müller (1998), Barbieri & O'Hagan (1996) and Huerta & West (1999) applied the reversible jump sampler to mixture models, variable selection, curve fitting, autoregressive models, neural networks, autoregressive moving average (ARMA) models and component structure in autoregressive (AR) models, respectively. In the financial literature, various filtering techniques, tailored to the particular situation at hand, have been proposed for models based on marked DSPPs. Frey & Runggaldier (2001) proposed a filtering technique based on the reference probability method for a model in which the intensity depends on the level of an unobserved Markovian state process. Rydberg & Shephard (2000) proposed a particle filtering method based on the auxiliary sampling importance resampling algorithm of Pitt & Shephard (1999) for a DSPP with unknown Markovian intensity. In this particle filtering technique, they needed to aggregate the time events in time intervals of fixed length.

In this thesis, we develop Green's RJMCMC algorithm to filter the unobservable intensities in our latent factor model. In general, the RJMCMC algorithm provides a simulated approximation of the conditional distribution of the intensity processes, over any fixed time

interval, given the observed trajectory of the price processes. Under our latent factor model, the most challenging task is to obtain, not only the filtering of the intensities, but also of their components. We construct two different RJMCMC algorithms for the filtering of the intensities of the model (and of their components) and show by an extensive simulation study that the performance of the first algorithm, although able to filter the intensities, is not as good as the second in the filtering of the components. Our proposed algorithms can be seen as an extension of the algorithm in Centanni & Minozzo (2006). More details of RJMCMC algorithm are given in Chapter 4.

Chapter 1

Multivariate ultra-high-frequency data

Modeling of multivariate ultra-high-frequency (UHF) financial data by using multivariate point processes has witnessed a growing interest in the recent literatures. Ultra-high-frequency data, also known as tick-by-tick data, are available for most exchange markets of financial assets and also exists for stocks such as the New York Stock Exchange, the Paris Bourse, the London Stock Exchange or the Frankfurt Stock Exchange. By nature these data are irregularly spaced in time. Liquid markets generate hundreds or thousands of ticks per business day. Data vendors like Reuters transmit more than 275,000 prices per day for foreign exchange spot rates alone. In contrast to the standard financial databases, which usually provide information on a daily or weekly basis (usually, the closing prices and traded volumes), UHF databases provide an enormous amount of information about intraday behavior of financial markets. Indeed, they provide the finest possible information recording the time at which a market event, such as a trade or a quote update by a market maker or computer system, takes place together with its associated characteristics (for instance, the price and volume of the transaction) for all transactions or quotations. As a consequence, a systematic modeling of the UHF data should be the primary object for researchers and practitioners who are interested in understanding financial markets. However, progress in modeling the data by multivariate case present difficulties. The approach adopted here overcomes the difficulties by using marked doubly stochastic Poisson processes (DSPPs) and provides a general framework for model specification and inference that would greatly facilitate the statistical analysis of these important data sets.

1.1 Market microstructure

Relating to the type of data available for researchers, the stylized facts and the models that are discovered and developed direct to these different type of samples, provides insight into the development of research, with increasing use of statistical tools and increasing interaction between statistically and economically equipped researchers. Taking account of frequency of data, the annual price can be considered as the data sample with the lowest frequency, while tick-by-tick data are the highest frequency available. Between those two extremes, monthly data, daily data and hourly data etc., are usually considered as standard frequency (or low frequency) data. Models have been identified at these different time scales. The ARCH model of Engle (1982) and its extensions (GARCH model) have

become well known models in the analysis of the standard frequency data. The autoregressive conditional duration (ACD) model proposed by Engle & Russell (1998) that deals with irregularly spaced data (tick-by-tick data, for instance). Essentially, the ACD model is to model the time intervals (so-called durations) between market events such as the occurrence of a trade or bid-ask quote. The ACD model is a duration model (as the name implied) which combines features of duration models with time series close to ARCH model.

Much of the motivation for the ACD model come from the market microstructure theory, where time between market events (duration) has had an important impact on the behavior of market participants. Then the high-frequency data open great possibilities to test market microstructure models. Over the past thirty years, there has been a considerable amount of research in market microstructure. Generally, this research has focused both on theoretical and on empirical models, on price driven and order driven markets, on liquidity, on the role of information in trading (more particularly the effects of information about trades and quotes on the functioning of the markets), on the importance of information asymmetry between traders and on the reaction between market participants. The theoretical models are based on Walrasian equilibrium and rational expectation models. O'Hara (1995), Biais, Foucault & Hillion (1997), Goodhart & O'Hara (1997) and more recently Madhavan (2000) provide excellent reviews of the informational role of prices in such theoretical models. But a growing number of studies make use of 'experiments' to investigate empirical market microstructure models, see Bloomfield & O'Hara (1999) for some recent development in this field.

1.1.1 Trading mechanism

In the simple trading mechanism, all agents are willing to trade an asset, where some agents want to buy the asset and others want to sell it, submit their supplies and demands to the market maker. The market maker aggregates the demands and supplies and then announces a first potential trading price. Observing this price, agents revise their supplies and demands, which leads to a revision of the price by the market maker. At the end of this process, a market clearing price is obtained, which equates aggregate supply and demand. Trades between buyers and sellers only take place when an equilibrium price has been reached.

Other existing trading mechanism deals in different way, which is the modern trading scheme. During the so-called preopening hours, agents submit their limit buy and sell orders to the computerized system. At the end of the preopening period, the opening price is determined by the computer as the price at which the maximum volume among all submitted orders can be traded.

1.1.2 Financial markets

There are mainly four distinguish structures in financial system: individuals (individuals, households, etc.), companies (business firms, companies, corporations, etc.), security markets (capital markets, money markets, etc.) and intermediate structures (banks, insurance companies, etc.).

The usual view of markets is as a place where buyers and sellers come together and trade at a common price, the price at which supply equals demand. Securities exchanges are often pointed out as excellent examples of markets that operate this way.

In general, markets can be defined as centralized or decentralized. The centralized markets refer to exchange or bourses, while the decentralized (over-the-counter) markets where individual participants directly transact without intermediary. Data from centralized markets are available from the same sources and sometimes directly from the exchanges. The recent shift from floor trading to electronic trading contribute to make this data more reliable and more easily available. Data from over-the-counter markets are collected and provided in real time by data providers such as Reuters, Bloomberg, or Bridge.

Note that, here, we make no attempt to list all types of financial markets. A nice review is provided by Dacorogna et al. (2001), another reference see also Bauwens & Giot (2001).

Spot markets

Spot markets are direct markets for primary assets, such as foreign exchange or equity. The assets are traded immediately at the time of transaction. Spot trading is the most original form of trading, but it has some disadvantages. The timing is not flexible, traders have to deal with the physical delivery of the traded assets (such as commodities) and the interest rate spot market is affected by the counterparty default risk. For this reason, derivative markets have become more favorable than spot markets in some cases.

The Foreign Exchange (FX) market is a major example of the spot market and is the largest financial market, which produces high-frequency data. This is due to the fact that both exchanged assets are currencies in the trading market. During the past twenty years, academic researchers, with the availability of data, have gained insight into the behavior of the FX markets through analyzing intraday data. And the analysis of intraday data leads to insights into the field of market microstructure.

Futures markets

Futures markets, such as most interest rates and commodities, have a higher liquidity and volume than the underlying spot markets and produce high-frequency data. A general description of futures markets will be given in the following.

Futures contracts are derivatives of an underlying asset, which can be defined as an agreement between two parties to buy or sell an asset at a certain time in the future for a certain price, see Hull (1993). At this expiry time, the underlying asset has to be delivered according to the contract, after which the contract no longer exists. The expiry dates are regularly scheduled, often in a quarterly sequence. The contract with the nearest expiry is called the first position, the following contract as the second position, and so on.

Most futures are traded at exchanges. Trading is typically geographically localized. There are rigidly defined opening and closing hours. Given that futures contracts are exchange traded and each transaction is recorded centrally, futures markets present a high price transparency. The historical data always includes tick-by-tick transaction prices and bid and ask quotes, and the flow orders from the clients of the exchange.

All clients buying or selling futures contracts are asked to put some money in a collateral account. This account covers the counterparty default risk of the exchange. If the

futures price move to the disfavor of a client, the amount of money on the collateral account may no longer cover the risk, and the exchange will require the client to increase it through a 'call for margin'. If the client fails to do this, the futures contract is terminated at its current market value.

Option markets

Option contract is a security which gives its holder the right to sell (buy) the underlying asset under the contracting condition. In terms of the expiration time, options are classified into two main types: European and American options. The expiration date of a European options is fixed, while the American options can be exercised at any time up to the expiration data. Most of the trading options are options of American type. Notice that the name 'European' and 'American', in reality, bear no relation to whether the trading takes place in Europe or in America.

The options providing the right to buy are referred to as call options, while a contract gives the right to sell, it is called put option.

1.1.3 Market liquidity

The markets can be classified as being order/price driven or centralized/decentralized, it is much more tricky to compare their characteristics and to assess their relative performance. The usual notion used in the existing market microstructure literature is to focus on the liquidity of trades. Liquidity can be defined as the ability to quickly buy or sell a large volume of shares with a minimal price impact. Trades offering greater liquidity are of course favored by investors as they allow a much more efficient allocation of capital. In particular, the ability to trade quickly is important for traders, as yet-to-be executed orders face the risk of unexpected movements in the underlying value of the asset.

Price driven markets

A price driven market relies on the existence of a market maker (designated person (usually employed by the exchange or by banks affiliated with the exchange) who has an obligation to quote firm bid-ask prices for a given asset. These bid-ask prices are valid up to a given number of shares. The market maker buys the asset at the bid price and sells the asset at the ask price.) to buy or sell the traded asset. The market maker buys the asset from the first trader and then sells it to the second trader. Meanwhile, the market maker provides liquidity to the market by posting quotes, standing ready to buy (at the bid price) and sell (at the ask price) the asset at any time. For instance, if a given market maker quotes a 100 euro bid for 100 shares and a 100.5 euro ask for 200 shares, a trader wishing to buy up to 200 shares can actually buy the shares at 100.5 euro.

As the market maker buys the asset from a trader and then, at a subsequent time, sells to another trader, he holds an unwanted inventory position for a certain time, which means he undertakes the potential risk. For instance, the company for which the market maker holds their shares could announce some bad news just after a large buy by the market maker. In this case, the market maker will only be able to sell the stock at a much lower price than the price he paid to buy the stock. To reward the market maker for running this risk and thus to

reward him for providing liquidity to the market, the exchange usually grants advantages to the market maker: 1) bid-ask spread, as the ask price is higher than the bid price, the market maker can benefit from buying and selling the stock immediately (or nearly so). 2) information advantage, the market maker usually has more information than normal traders. This is the case for the specialist at the New York Stock Exchange (NYSE), for instance, as the specialist has a full knowledge of the order book, while normal traders can only observe the best bid and ask prices.

Example (Trading shares in a price driven market) Assuming that we have two traders X and Y, and a market marker Z making the market for the stock of company A. X is willing to sell 100 shares of A and Y is going to buy 100 shares. In a price driven market, Z posts bid-ask quotes which are made available to all market participants, such as traders X and Y. Now suppose that the current bid price is 99 euro and that the current ask price is 100 euro. If trader X want to sell 100 shares, makes the trade at the bid price of 99 euro. Once the trade has been made, the market maker's inventory has gone up by 100 shares. Trader Y now buys the 100 shares at 100 euro, thereby, the market maker's inventory become zero, resulting in a net gain of 100 euro for the market maker.

On the other hand, if company A announces some bad news, which leads to a downward revision of the share price of A, after market maker buy 100 shares of asset A at bid price of 99 euro. As the news is public, no trader will be willing to buy the 100 shares at the prevailing ask price of 100 euro. At the same time, the market maker are not willing to undertake the risk (the situation goes worse) and would like to sell it as soon as possible. The only way to solve this problem is to reduce his ask price, say 95 euro, to attract buy orders. Suppose trader Y is willing to buy 100 shares at ask price of 95 euro, the market maker Z makes a net loss of 400 euro on this trade.

Order driven markets

For the order driven markets, it is also known as automated continuous double auctions, no market makers are involved in the trading process. Traders directly enter limit buy or sell orders in the order book (in an order driven market, the order book is usually managed by a centralized computer. In a price driven system, the order book is maintained by a market maker) which is a complete collection of limit buy and sell orders entered by traders. Trades are made when orders can be matched. Note that auction refers to an old trading mechanism where the auctioneer (can be the market maker) aggregates the demands (want to buy an asset) and supplies (want to sell an asset) of an asset, and announces a first potential trading price, then agents know and revise their supplies and demands, which leads to a revision of the price by the auctioneer. This trading proceeds repeatly until a market clearing price is obtained. In the order driven markets, the auctioneer is the centralized computer.

Example (Trading shares in an order driven market) Suppose that the current state of the order book is as given by the situation in Table 1.1. In the current state, five limit orders were entered into the order book. Two traders are willing to buy shares of 600 at 99 euro and 2500 at 99 euro, while three traders are going to sell with the lowest price at 100 euro of 2000 shares and the highest price at 103 of 400 shares. At this stage, a trade is not possible as the highest buy order is at 99 euro and the lowest sell order is at 100 euro.

Now suppose that a trader X enters a sell order for 1000 shares with the best limit price, the only solution is to sell 600 shares at the bid price 99 euro and 400 shares at the bid price 98 euro. The final state of the order book is given in Table 1.2. Although this example is very simple, it describes quite accurately how a limit order system works.

As detailed in the example, price priority is a very important feature of order book systems because it ensures that the most favorable trades are always executed first. Thus, for a trader wishing to sell (buy) shares, it ensures that his order will always be matched with the highest (lowest) limit buy (sell) orders. When several orders are competing for execution at similar prices, time priority is usually enforced, in the sense that orders first entered into the limit order book will have execution priority.

Table 1.1: Start state of the order book. Five limit orders were entered into the system: two traders are willing to buy shares of 600 and 2500 at 99 euro and 98 euro respectively; three traders are selling shares, with the lowest price at 100 euro (2000 shares) and the highest price at 103 euro (400 shares).

Shares	Limit buy (Euro)	Limit sell (Euro)	shares
600	99	100	2000
2500	98	101	1000
		103	400

Table 1.2: End state of the order book. A sell order for 1000 shares with limit buy order, at 99 euro for 600 shares and 98 euro for 400 shares.

Shares	Limit buy (Euro)	Limit sell (Euro)	shares
0	99	100	2000
2100	98	101	1000
		103	400

1.1.4 Information based models

Most market microstructure research concentrate on how price adjust to new information and how the trading mechanism affects asset prices. The first so-called information based model was introduced by Glosten & Milgrom (1985). In these models, traders and market makers (who are be assumed to be risk-neutral) do not have the same information regarding the value of the security they are trading. Under the perfect market assumption, new information would be instantly interpreted by all trading participants. In practical world, however, the underlying information is not known by all market participants. This suggests a different kind of traders in the market. Typically two kinds of traders are trading with market maker: informed and uninformed traders. The agent with superior knowledge regarding the value of asset referred to as informed trader, while agents without such information are considered as uninformed trader.

Uninformed traders do not have superior information (with respect to the market makers' information set) regarding the financial asset they are trading and they mainly trade for liquidity reasons. Informed traders, however, have superior information on the asset. It's reasonable to expect that the informed party will be more aggressive trading in the early periods taking the advantage of the private information. Because informed traders know the final value of the asset, they always try to benefit from their private information. They sell if they know bad news, buy if they know good news, and they trade as long as the price quoted by the market maker is different from the price implied by their private information.

The market maker, who is potentially confronted with both types of traders, does not know if he deals with an informed or an uninformed agent. When the incoming trade originate from a liquidity trader (uninformed agent), the market maker does not lose on the trade. However, if market maker is trading with informed trader, who trades because the quoted prices do not reflect the private information (either good or bad), in order to avoid this possible loss, the information model as introduced by Glosten & Milgrom (1985) specifies that the quoted prices are set equal to the market maker's conditional expected value of the asset given that the trader wishes to sell the asset to the market maker. As there is always the possibility that this sell order originates from the informed trader, the quoted bid price includes the possibility that bad news has occurred. The opposite is true for the ask price, which is set equal to the conditional expected value of the asset given that the trader wishes to buy the asset from the market maker. Because informed traders only buy when good news is to be released, the ask price must allow for this possibility. In both cases, the market maker assumes that the incoming trade is a signal of possible private information. An important implication of the model is that the bid-ask prices are different, with the ask price greater than the bid price. A positive spread arises, which compensates the market maker for the possible loss due to trading with the informed traders. Thus, the spread is the market maker's compensation for facing adverse selection in the order flow.

Much of modern research then focuses on asymmetric information model, see Easley & O'Hara (1987). The key assumption of asymmetric information model is that trade size (volume) convey the new pieces of information, that is, traders (informed and uninformed) can choose to submit a small order or a large one. They conclude that there should be order size dependency of the spread, as market makers should beware large size orders which probably originate from informed traders. Another information based model proposed by Easley & O'Hara (1992) focus on the role of time in the price adjustment process, that is, updating of the quotes. In this model, they argue that the time between trades, also called duration, conveys information. A long duration means that no new information (either good or bad) has been released. Thus the probability of dealing with an informed trader should be smaller than when a short duration is observed. Consequently, with a low probability of dealing with an informed trader, the quoted spread decreases. Another consequence of their model is that the release of news should lead to an increase in the trading intensity and this should imply more frequent revisions of the bid-ask prices posted by the market makers: "quotes converge to their strong form efficient values at exponential rate. Rates of convergence are increasing in the fraction of trades from the informed." (Easley & O'Hara, 1992).

1.2 Ultra-high-frequency data

Nowadays with the development of computer technology, tick-by-tick data becomes available for researchers in financial markets, which opens new direction for understanding of financial market efficiency. Ultra-high-frequency data, also known as tick-by-tick data, are direct information from markets. One logical unit of information is called a tick. This term originated from the language of practitioners and originally meant a number on a ticker tape, in a time before computers became an omnipresent tool. The term 'tick' is more neutral and general than the particular terms 'price', 'interest rate', 'quote' and so on. Whatever the quoted quantity, there is always a date and a time attached to every tick, a 'time stamp'. The sequence of time stamps in usually irregularly spaced. A large part of this thesis deal with the consequences of this fact.

The quoted quantities are often prices, but other information such transaction volume is also available from some markets, such as the foreign exchange market and the New York stock exchange. In this thesis, there is no attempt to analyze this variable, volume, rather than focus on price or logreturn.

In a way, transaction data are collected from an electronic limited order book for each stock. Incoming data are ranked according to price and time of entry and are continuously updated. Hence, the new records are the automatic match of the buy and sell orders, which generates a transaction (see Dacorogna et al.) and a transaction refers to a trade between a buyer and a seller of a volume of stocks at a given price. The analysis of intraday data also leads to get insight into the market microstructure where it is possible to study the behavior of intraday traders. Empirical results are reported for trading in *Borsa di Milano* for a three-week period from October 27, 2008 to November 14, 2008.

With these new data sets come new challenges associated with their analysis. Gathering the behavior of financial data and their returns is an important research activity. Without investigating the behavior of the data it is not possible to design models that can explain the data. Ultra-high-frequency data opened up a whole new field of exploration and brought to light some behaviors that could not be observed at low frequencies. In this Chapter we review the main characteristics of UHF financial data.

These characteristics can be grouped as the irregularly spaced in time, the discreteness of price changes, negative first-order autocorrelation of logreturns, positive cross-correlation between two time series and present so-called 'Epps effect'. We find a remarkable similarity between different assets from *Borsa di Milano*.

1.2.1 Data from Borsa di Milano

Before getting insight into characteristics of ultra-high-frequency data, the data we analyze through this thesis has to be explained. We use seven Italian banks traded in *Borsa di Milano* (S & P 500) for three-week period of October 27, 2008 to November 14, 2008. The seven Italian banks are Banco Popolare (POP), Mediobanca (MED), Banca Popolare di Milano (MIL), MPS Banca (MPS), Intesa SanPaolo Banca (ISP), UBI Banca (UBI) and Unicredit Banca (UCD). Thus, our research activity (involving empirical data) throughout this thesis is based on these candidate banks. In particular, computing auto- and cross-

correlation of logreturns with different time intervals, more detail refers to Section 1.3.3.

In the context of this thesis, ultra-high-frequency data is defined as the raw time series of prices. The time interval between transactions ranges from 0 seconds (several distinct trades recorded at the same time but not many) to 10 minutes. The trading hours in *Borsa di Milano* are 9:05 to 17:30 on weekdays. The computer recorder aggregates trades outside the opening hours of *Borsa di Milano* and for most of the banks of interest there is a noticeable number of transactions occurring from 9:00 to 17:35, which yields some sparse additional data. To avoid contaminating the data, in the sense that these outlying data may very different from those traded at the official time, the daily trading window we analyze including 8 hour and 20 minutes (equivalent to 30000 seconds) from 9:05 to 17:25.

1.2.2 Irregular time spacing

It has been long recognized that the sampling of financial time series plays a subtle but crucial role in determining their stochastic properties. In general, almost all of the financial data is point sampled, or discrete samples. Yet most of the studies published in the financial literature deal with time-averaged regularly spaced data via certain mechanism (here the mechanism refers to synchronization methods, linear interpolation and previous-tick interpolation, see details in Section 1.3.1). More generally, econometric problems are bound to arise when we ignore the fact that the statistical behavior of sampled data may be quite different from the behavior of the underlying stochastic process from which the sample was obtained. In particular, the advent of the ultra-high-frequency data makes this problem even more piercing and serious. On the other hand, it opens up the whole new field of exploration and bring to light some behaviors that could not be observed by standard econometric models.

Figure 1.1 illustrates the first 5 minutes' transaction logprice of seven Italian banks from *Borsa di Milano* on $27/10/2008$. The figure presented here will be the subject of several examples throughout this thesis. In Figure 1.1, the horizontal axis is the time of the day which is converted into digital number (in second), for instance, 0 denotes 9:05 a.m. (opening time) and 300 denotes 9:10 a.m.. The vertical axis is the logprice measured in Euros. Each asterisk indicates one transaction. It is clear from Figure 1.1 that each transaction not only irregularly spaced in time but also time varying of logprices. Basically, two facts can be summarized as follows. First, with univariate view on each asset, some transactions appear to occur almost contemporaneous while others may be a few seconds. Take Banco Popolare for instance, very short interarrival time (almost zero) of transaction for time between 0 to 30 second, whereas quite longer for time between 50 to 100 second. Second, with multivariate view on seven assets together, some assets are traded frequently, such as Intesa SanPaolo Banca and Unicredit Banca, exhibiting with frequency about 60 ticks in the first five minutes, while others are less frequently traded, such as Banco Popolare, Mediobanca and MPS Banca etc, exhibiting with frequency about 20 ticks in the first five minutes (almost three times less in contrast with the previous two banks). This fact can be better view by Table 1.3 which illustrates the summarized number of transactions and frequency per business day. Last but not the least, the transactions of seven Italian banks almost never coincide at the same time, this problem will refer to asynchronization, further detail defers to Section 1.3.

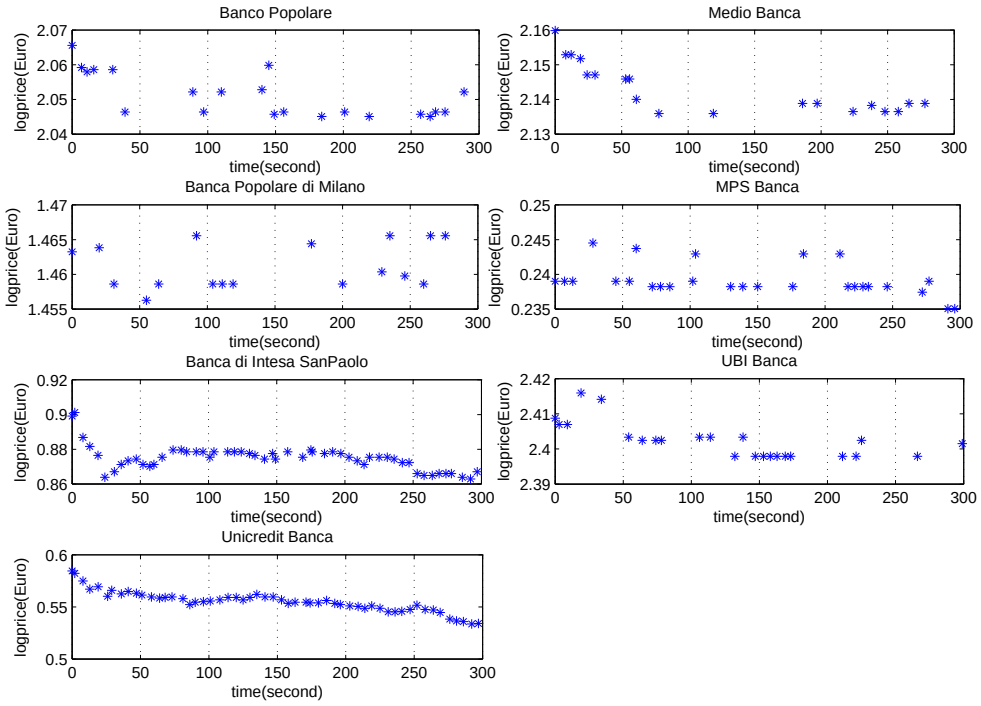

Figure 1.1: This figure illustrates the first 5 minutes' transaction logprice of seven Italian banks from *Borsa di Milano* after market open time (9:05 AM) on $27/10/2008$. The horizontal axis denotes the time of the day after market open which is measured in seconds. 0 indicates 9:05 AM, and 300 indicates 9:10 AM. The vertical axis indicates the logprice which is measured in Euros. The asterisks represent transactions, incorporating time and logprice.

Hence, ultra-high-frequency data is not only irregularly spaced in time serially but also asynchronized between different assets (see Figure 1.1). The problem becomes more complicated when one realizes that the rate of arrival of transaction data may vary over the course of the day, week. Such datasets have hitherto been hindered by the difficulties presented by multivariate case study.

1.2.3 Tick frequency

Ultra-high-requency data means a very large amount of data. The first 5 minutes' transaction as shown in Figure 1.1 gives an immediately evidence that the number of observations in (only) 5 minutes of liquid market, such as Unicredit Banca, is equivalent to the number of daily data within 3 months.

To manifest the enormous amounts of available ticks, Table 1.3 provides the summarized description of the total number of ticks and the average frequency per business day over three-week sampling period from October 27, 2008 to November 14, 2008 in *Borsa di Milano* database. Note here *business day* implies excluding of weekends, so the average frequency per business day is defined as the arithmetic average over fifteen days. On the largest market, Unicredit Banca feeds with almost 65000 ticks totally and more than 4000 ticks per business day; that is an average of almost 8 ticks per minute which can rise to 20 or more ticks per minute during the businest periods. On the other side, MPS banca, 14081 ticks totally and 938 ticks per business day are available; with an average of less than 2 ticks per minute.

Table 1.3: The total number and the average number of transactions for each asset from *Borsa di Milano*. The sampling period from October 27, 2008 to Nobember 14, 2008.

Asset	Total number of transactions	Frequency per business day
Banco Popolare	22993	1533
Mediobanca	14434	962
Banca popolare di Milano	14411	961
MPS Banca	14081	938
Intesa SanPaolo Banca	51626	3442
UBI Banca	15022	1002
Unicredit Banca	64444	4297

It is somehow more subtle but still quite important: the sampling data we analyzed through the whole thesis is three-week time period from October 27, 2008 to November 14, 2008 in the *Borsa di Milano*, where the time span is not large as many years or at least one year, but the total transaction here is definitely large enough for data analysis. Banco Poplare (not the most liquid asset), for instance, produced 22993 ticks during this sampling period. If we evaluate it in terms of daily, then it is essentially equivalent to 85 years' long time series. Hence, there is no problem for data analysis in such short term. On the other hand, it is rather convenient to collect ultra-high-frequency data in very short time.

Statistically, the higher the number of independently measured observations, the higher is the degree of freedom, which implies more precise estimators. The large amount of data allows us to distinguish between different models with a higher statistical precision. When diffferent models have to be ranked, the availablity of a few hundred thousand observations allow us to find beyond a doubt which model provides the best description of the data generating process (Müller et al., 1997).

1.2.4 Distributional properties of logreturns

In this section, we analyze the probability distribution of logreturns of ultra-high-frequency financial data. The probability distribution associates each movement size with a certain probability of occurrence. In the case of empirical data, the domain of possible logreturn values is divided into boxes, and one counts the frequency of occurrence in each box. One important issue in the case of ultra-high-frequency (tick-by-tick) data is that they are irregularly spaced in time as shown in the previous section. It is then interesting to examine empirically what kind of behavior is observed when logreturns are measured at different frequencies. Note that there is no privileged time interval at which the data and the generating process should be investigated.

Given the raw ultra-high-frequency data, which are irregularly spaced in time, if we want to construct a homogeneous time series, some interpolation methods must be applied. Here we use previous-tick interpolation method to homogenize the data (more details about previous-tick interpolation method are referred to Section 1.3.1). Then the logreturn can be calculated in the following formular:

$$R_{\Delta t}(t) = \ln P(t) - \ln P(t - \Delta t), \qquad t = \Delta t, 2\Delta t, 3\Delta t, \ldots. \tag{1.2.1}$$

where $P(t)$ denotes the price at time t, and Δt refers to equidistant time.

Now we apply Formular 1.2.1 to our empirical data which are seven Italian banks from *Borsa di Milano* sampled from October 27, 2008 to November 14, 2008 to calculate logreturns at different sampling intervals. Then we also compute the first four moments of logreturns at different sampling intervals ($\Delta t = 10$ seconds, 1 minute, 10 minutes and 1 hour) for the empirical data. The result is presented in Table 1.4.

Table 1.4: Moments of logreturn distribution for seven assets in *Borsa di Milano*. The sampling period is chosen from October 27, 2008 to November 14, 2008.

Asset	Time Interval	Mean	Variance	Skewness	Kurtosis
Banco Popolare	10 sec	8.35e–7	1.39e–6	5.57	311.61
(POP)	1 min	6.72e–6	5.43e–6	3.12	81.57
	10 min	6.38e–5	4.05e–5	1.45	18.55
	1 hr	4.01e–4	2.06e–4	0.15	4.53
Mediobanca	10 sec	1.02e–6	4.00e–7	2.95	193.05
(MED)	1 min	1.06e–6	1.85e–6	1.97	52.58
	10 min	8.64e–5	1.14e–5	0.86	14.35
	1 hr	4.51e–4	5.54e–5	0.43	4.15
Banca Popolare	10 sec	-5.34e–7	1.51e–6	-0.09	322.59
di Milano	1 min	-2.27e–6	6.71e–6	0.09	100.89
(MIL)	10 min	-3.21e–5	4.31e–5	0.08	17.95
	1 hr	3.35e–5	2.43e–4	0.24	4.91
MPS Banca	10 sec	2.95e–6	1.49e–6	3.50	198.96
(MPS)	1 min	1.71e–5	6.60e–6	1.50	59.19
	10 min	1.77e–4	4.90e–5	0.37	15.07
	1 hr	0.11e–2	2.81e–4	0.83	7.13
Intesa SanPaolo	10 sec	4.57e–8	1.87e–6	7.46	443.13
Banca	1 min	2.35e–6	8.42e–6	2.38	100.01
(ISP)	10 min	5.46e–5	9.15e–5	1.75	30.99
	1 hr	2.10e–4	5.12e–4	-0.15	5.25
UBI Banca	10 sec	2.14e–6	1.07e–6	1.21	166.35
(UBI)	1 min	1.33e–5	5.01e–6	2.08	74.07
	10 min	1.59e–4	3.89e–5	2.50	39.79
	1 hr	9.89e–4	1.96e–4	1.10	8.24
Unicredit Banca	10 sec	2.71e–6	1.85e–6	7.44	533.18
(UCD)	1 min	1.83e–5	9.32e–6	3.47	149.37
	10 min	2.19e–4	8.84e–5	1.70	19.70
	1 hr	0.12e–2	5.80e–4	-0.59	6.46

In Table 1.4 we present the empirically computed moments of the unconditional logreturn distribution measured at different sampling intervals. The data are seven Italian banks in *Borsa di Milano* for the period from October 27, 2008 to November 14, 2008. The means are close to zero, as compared to the standard deviations, and are slightly positive, except for Banca Popolare di Milano. One explanation for this phenomenon is that most of the assets in *Borsa di Milano* have experienced an upside market during this period, in contrast, Banca Popolare di Milano goes downside. On the other hand, the values of skewness are, except in very few cases, significantly greater than 1. For this, we can conclude that the empirical distribution of logreturn is almost positive skewed. Finally, for all time horizones, the empirically determined (excess) kurtosis exceeds the value zero, which is the theoretical value for the Guassian distribution. In case of the smallest sampling interval, 10 seconds, the kurtosis values are extremely high. While at intervals of 1 hour, the kurtosis is quite close to the Guanssian value. Another important feature from this table is that all of these seven Italian banks show one similar behavior, where mean and variance have an increasing function of sampling interval with range from 10 seconds to 1 hour, while the kurtosis plays in the opposite direction.

Table 1.4 suggests that the variance and the third moment are finite in the small sampling interval (corresponding to large number of samples) and that the fourth moment may not be finite. Some solid evidence in favor of these hypothesis is added by the tail index studies. Indeed, the larger the number of observations, the larger the empirically computed kurtosis (M. Dacorogna et al., 2001). At frequencies higher than 10 minutes, there seems to be some contradiction between the work of Goodhart & Figliuoli (1991), which claims that the fat tails start to decrease at these frequencies, and the paper of Bollerslev & Domowitz (1993), which gives some evidence of a still increasing fat-tailedness. M. Dacorogna et al. (2001) explained that both results hold depending on whether one uses the linear interpolation method or the previous-tick to obtain price values at fixed time intervals at such frequencies. However, the same result is obtained as Goodhart & Figliuoli (1991) in both our data by using previous-tick interpolation and M. Dacorogna et al. (2001) by using linear interpolation method, this suggests interpolation method does not impact on the statistical indices.

On the other hand, this contradiction is an example of the difficulty of making reliable analysis of prices values at frequencies higher than 10 minutes. The divergence of the fourth moment explains why obsolute values of the returns are often found to be the best choice of a definition of the volatility (see M. Dacorogna et al., 2001). Indeed, because the fourth moment of the distribution enters the computation of the autocorrelation function of the variance, the autocorrelation values will systematically decrease with a growing number of observations.

The original form of market price is tick-by-tick or, equivalently, discrete data. It is apparent as shown in Figure 1.1. In general, almost all of the economic data is discrete (Engle, 1998). Most of price changes of the UHF data fall in a small range. Without considering sampling interval, so no interpolation method is involved and logreturns are calculated based on its original time interval (not fixed). Thus the logreturn calculation is not followed with Equation 1.2.1 but

$$R_{t_i} = \ln P_{t_i} - \ln P_{t_{i-1}}, \tag{1.2.2}$$

where P_{t_i} denotes price at time t_i. Note that the different between t_i and t_{i-1} is not fixed.

Indeed, time t_{i-1} is the previous transaction time with respect to t_i, given the irregular spaced in time feature of the data, $||t_i - t_{i-1}||$ is thus a random variable.

The following two figures (Figure 1.2 and Figure 1.3) present the distributional property of logreturn from the raw data, where the logreturn computation is based on Equation 1.2.2. In some extent, it makes no sense to measure distribution of the logreturn in terms of time, saying, the majority of logreturn is around zero from Figure 1.3 without time index (5 minutes or 30 minutes).

Figure 1.2 illustrates stem plot of logreturns of seven assets (Italian banks) after getting rid of the overnight and opening transactions, the sampling period covers 15 business days of data spanning from $27/10/2008$ to $14/11/2008$ from *Borsa di Milano*. Most of the transaction prices are not very different from the previous one. Except some logreturns show a bit different from zero in each asset, take Unicredit Banca (UCD) for instance, about 12 stems quite outstanding with respect to the rest stems, the absolute value of these stems (logreturns) less than 0.1, the explaination of this behavior may be the impact of the public new announcement or private sector reveal such kind of information.

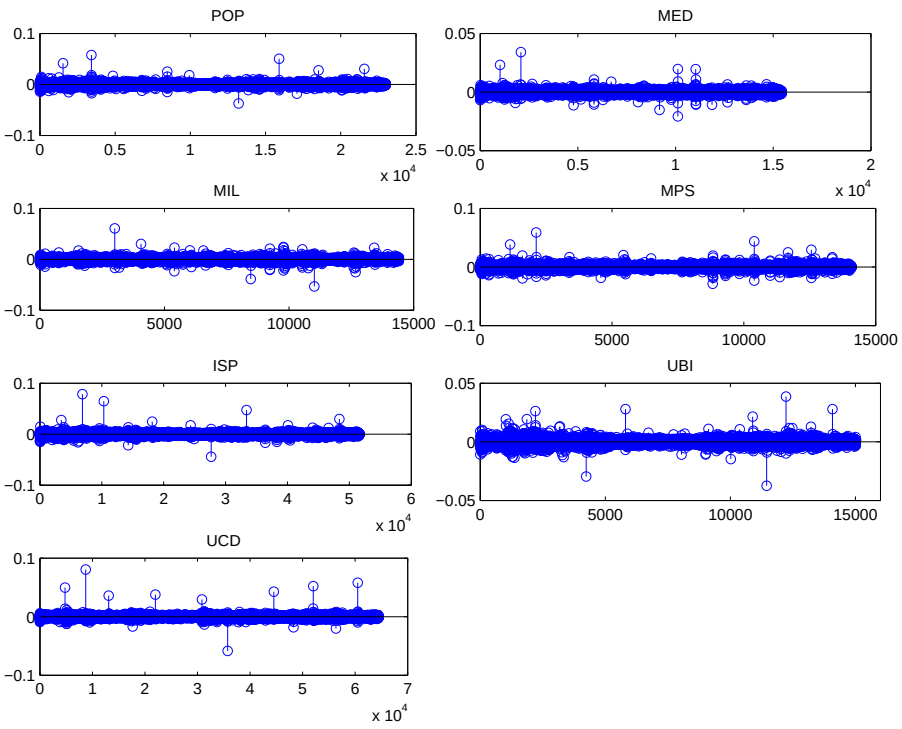

Figure 1.2: Stem plot of logprice change (logreturn) of seven Italian banks in *Borsa di Milano*. Most of the changing range focus on -0.05 and 0.05. The period is chosen from $27/10/2008$ to $14/11/2008$.

Figure 1.3 provides the histogram of logreturn of seven Italian banks in *Borsa di Milano* for the period from October 27, 2008 to November 14, 2008. The horizontal axis is measured in Euros. As we expected, all of these seven banks have the majority value of zero. Over 95% of logreturns lie between -0.01 and 0.01 for Banco Popolare (POP) and left skewed; The same focus interval for Banca Popolare di Milano (MIL), Intesa SanPaolo

Banca (ISP) and UBI Banca, but MIL and ISP are a bit right skewed; Over 95% of values lie between -0.005 to 0.005 symmetrically for Mediobanca (MED) and MPS Banca (MPS); Finally, Unicredit Banca (UCD) with over 98% of values around zero.

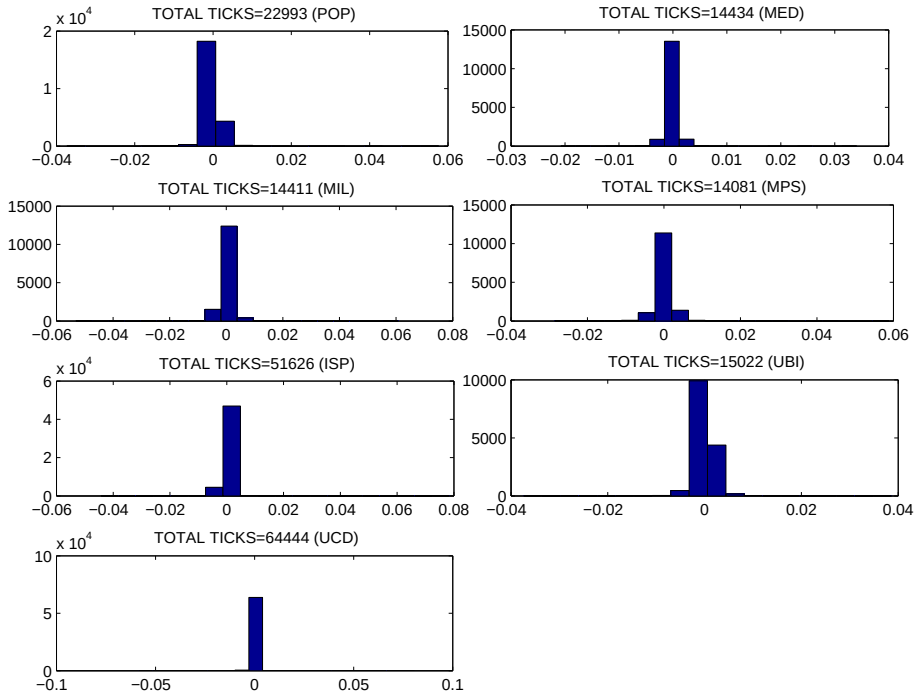

Figure 1.3: Histogram of logreturn of seven assets in *Borsa di Milano*. The majority of the value are around zero. The period is chosen from $27/10/2008$ to $14/11/2008$.

1.2.5 Negative first-order autocorrelation of logreturns

Goodhart (1989) and Goodhart & Figliuoli (1991) first reported the existence of negative first-order autocorrelation of returns at the highest frequency data, or tick-by-tick data, in which the autocorrelation swift back to zero when time lag becomes larger. The autocorrelation function of a stochastic quantity reveals at the same time serially dependence and periodicity. The autocorrelation function signals a periodic pattern by peaking at lags that are integer multiples of the particular period.

In Figure 1.4, the autocorrelation coefficient of logreturns is ploted against its time lags. The logreturns are computed using the raw data, in other words, all of these correlation are not calculated in the fixed time interval (Equation 1.2.1), but according to the natural data where they are irregularly spaced in time (Equation 1.2.2). For instance, the negative autocorrelation of Mediobanca is evaluated between two successive ticks (t_i and t_{i-1}) and there is no sense to say how many seconds last for this negative autocorrelation. Hence, the horizontal axis 600 (the maximum) indicates the 601^{th} transaction, but not the 600^{th} second. There is significant autocorrelation for all of seven banks in the first-order. Here the first-order in the sense that the negative autocorrelation is observed only at the first lag (two successive transactions). For longer lags, the autocorrelations mainly lie within the

95% confidence interval of an identical and indipendent Gaussian distribution. Besides, the negative first-order autocorrelation implies that high logreturn after low logreturn, or vice versa. One explaination of this negative relationship between two successive price changes in the same asset is undoubtedly the persistence for the short periods of a similar effect that exists among changes in price from one transaction to the next.

Table 1.5 lists out part of true value of the autocorrelation coefficient against its lags. Note that the lag mentioned here refers to previous transaction, for example, when lag is equal to 2, it indicates the second counter backward transaction, or meaningfully, the autocorrelation with lag 2 is given by $\rho_R(t_i, t_{i-2})$, as shown in Equation 3.4.22. It is thus no fixed time window is involved, that is, take the previous example, lag is equal to 2, it does not mean 2 seconds or 2 minutes, it can be any time length. In fact,it only depends on the inter-arrival time of transactions which are random. What we are interested here is the first-order autocorrelation of logreturns, the longer autocorrelations then become less important here. All of the first lag autocorrelation coefficient for the seven banks are out of the 95% confidence interval of i.i.d. Gaussian distribution. MPS Banca shows the strongest linear dependence in the first lag with coefficient value -0.2931, while Unicredit Banca (UCD) displays the least linear dependence in the first-order time lag with coefficient value of -0.1646. The second least value fall in Intesa SanPaolo Banca with coefficient of -0.1966, whereas Banco Popolare (POP) (-0.2622), Mediobanca (MED) (-0.2726) and Banca Popolare di Milano (MIL) (-0.2545) share similare correlation coefficient up or down -0.26. Essentially, these autocorrelations are evaluated using the autocorrelation coefficient

$$\rho_R(t_i, t_{i-1}) = \frac{\sum_{i=1}^{n} \left[R(t_i) - \bar{R}\right] \left[R(t_{i-1}) - \bar{R}\right]}{\sqrt{\sum_{i=1}^{n} \left[R(t_i) - \bar{R}\right]^2 \sum_{i=1}^{n} \left[R(t_{i-1}) - \bar{R}\right]^2}}, \tag{1.2.3}$$

with

$$\bar{R} = \frac{1}{n} \sum_{i=1}^{n} R(t_i), \tag{1.2.4}$$

where t_i is the ith transaction time and $R(t_i)$ denotes the corresponding value of logreturn, and n is the total number of transactions. Note that $t_0 = 0$ and $R_{t_0} = 0$.

1.2.6 Seasonality

Intraday financial data typically contain very strong periodic patterns, such as volatility clustering or persistence. In volatility clustering, large changes tend to follow large changes, and small changes tend to follow small changes. The changes from one period to the next are typically of unpredictable sign (positive or negative). Large disturbances, positive or negative, become part of the information set used to construct the variance forcast of the next period's disturbance. In this way, large shocks of either sign can persist and influence volatility forecasts for several periods. Volatility clustering, which is a type of heteroscedasticity, accounts for information is available to traders at different rate. The informed agents wish to trade larger quantities as soon as get the private information, a rational market maker will consider large orders as evidence of trading by informed paties and will adjust beliefs and prices accordingly. As Admati & Pfleiderer(1988) suggested that liquidity traders (uninformed agent) prefer to trade when the market is very liquid in

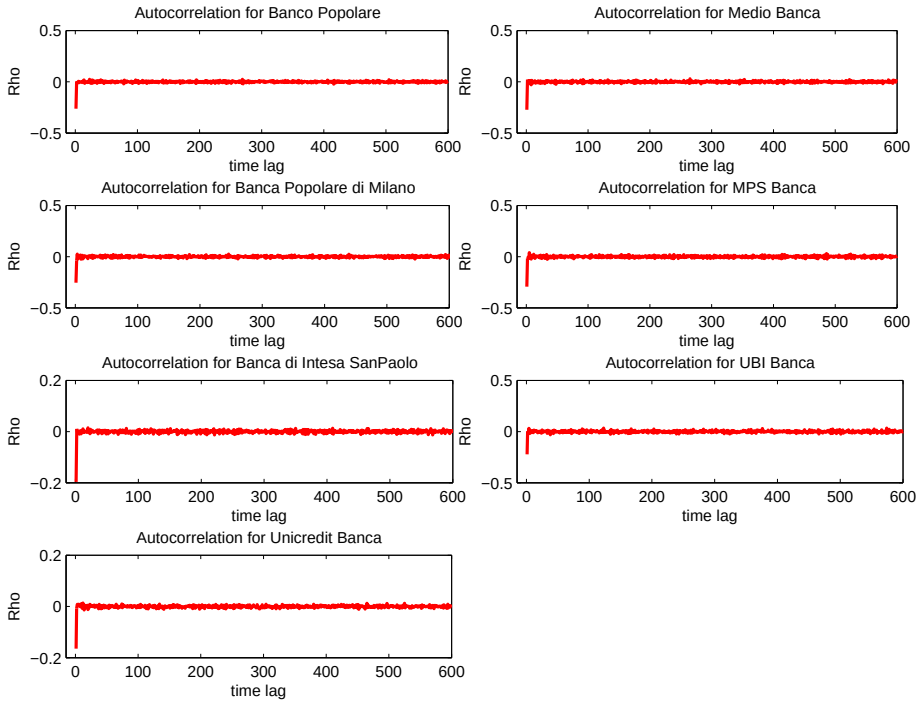

Figure 1.4: The autocorrelation of logreturns. This figure gives the autocorrelation of logreturns against time lags (measured in seconds) for seven banks in *Borsa di Milano*. The sampling period contains 15 business day from $27/10/2008$ to $14/11/2008$. Without any exclusion, all of them displays negative first order autocorrelation.

Table 1.5: Autocorrelation coefficient computed at different time lags. The samples are chosen from *Borsa di Milano* including three-week period from 27/10/2008 and 14/11/2008.

Time lag	POP	MED	MIL	MPS	ISP	UBI	UCD
1	-0.2622	-0.2726	-0.2545	-0.2931	-0.1966	-0.2222	-0.1646
2	0.0043	0.0221	-0.0385	-0.0167	0.0104	0.0029	-0.0079
3	-0.015	0.0059	0.0247	-0.006	-0.0018	-0.0237	0.0081
4	-0.0042	-0.0088	-0.022	-0.017	0.0005	0.0314	0.0042
5	0.0057	0.0106	0.0046	-0.0396	0.001	-0.0168	0.009
6	-0.0088	-0.0191	0	-0.0089	-0.0102	-0.0105	0.0049
7	-0.0026	0.0127	0.0086	0.0166	-0.007	-0.0121	0.0001
8	-0.001	-0.0296	0.0156	-0.024	0.0036	0.004	-0.009
9	0.0014	0.0142	-0.0257	0.0087	-0.0092	0.0115	0.0105
10	0.0006	-0.0159	-0.0045	0.0046	-0.004	-0.0013	0.0024
.	.	.	.	.	.	.	.
.	.	.	.	.	.	.	.
.	.	.	.	.	.	.	.

order to minimize transaction costs. It seems also ture for the informed party since a trade needs two parties and it is easy to trade in the tick market. Furthermore, Easley & O'Hara (1992) proposed a framework in which prices efficiently incorporate all the available information.

For most stock markets volatility, the frequency of trades and the bid-ask spreads all typically exhibit a U-shape pattern over the course of the day. References such as Ghysels & Jasiak (1995), Andersen & Bollerslev (1997), and Hasbrouck (1999) all found such intradaily seasonalities in the stock markes. Volatility is systematically high near the opening hour and follows a decrease, which is in turn followed by an increase of volatility just prior to the closing. Ballocchi et al. (1999) study the Eurofutures markets and find the expected intraday seasonality. For all contracts traded on London International Financial Futures Exchange (LIFFE) the hourly tick activity displays the U-shape with its minimum around 11 a.m. to 1 p.m. (GMT) and a clustering of activity around the beginning and the end of the trading day. Engle & Russell (1998) first to study the time between trades, or durations, which tend to be shortest near the opening and just before the closing.

Figure 1.5 presents a periodic pattern of durations for seven Italian banks in *Borsa di Milano*. The sampling period covers 3 weeks from 27/10/2008 to 14/11/2008. The day (weekday) is subdivided into 17 intervals from 9:05 to 17:25 with 16 half-hourly intervals and one 20 minutes as the last subinteval (from 17:05 to 17:25). Roughly, $\Delta t = 0.5$ hour is chosen as a sampling interval. The horizontal axis denotes the time of day and vertical axis denotes duration time measured in seconds. Each bin indicates half-hourly interval except the last bin instead of 20 minutes. The height of each bin indicates the mean value of duration during this half-hourly interval (20 minutes for the last bin). For example, the mean duration of Banco Popolare (POP) from 9:35 to 10:05 (the second bin of the upper left panel as shown in Figure 1.5) is around 18 seconds.

From Figure 1.5, it apparently shows periodic pattern of intrady durations, lower duration near the opening and just before the closing of the market, while higher duration is around 13:05 and 14:05 for most of banks (POP, MED, MIL, MPS and UBI). Two panels, ISP-duration and UCD-duration, as shown in Figure 1.5, which are very different from the others. The durations of Intesa SanPaolo (ISP) during the day are quite stable (around 10 seconds) except the peak time during 11:35 to 12:05 with 45 seconds. One explaination for this phenomenon can be that there is some public new announcement or private sector reveal such kind of information around 11:30, which lends the traders stop trading for a while. Similarly, the peak time of the durations of Unicredit banca (UCD) is different from 13:05 to 14:05 but around 9:35 and 10:05. Moreover, its duration near the closing time (17:05 to 17:25) even higher than midday. Similar explaination such as news release can be one reason for this phenomenon.

An analysis of the average numbers of ticks as a function of daytime give an idea about the market activity during the day. They are counts of original transactions from order book, though biased by our data supplier. Figure 1.6 improve the knowledge of the intraday studies since most of the intraday studies are dealing with intraday duration. They show, for example, that event the least active hour, 13:05 to 14:05 (noon break), contains more than 25 ticks for all of these seven Italian banks from *Borsa di Milano*, a sufficient quantity for a meaningful analysis. In Figure 1.6, the day (weekday) is subdivided into 17 subintervals from 9:05 to 17:25 with 16 of 30-minute intervals and one 20 minutes as the last subinterval (from 17:05 to 17:25). So the first 16 bins in the figure represent half-

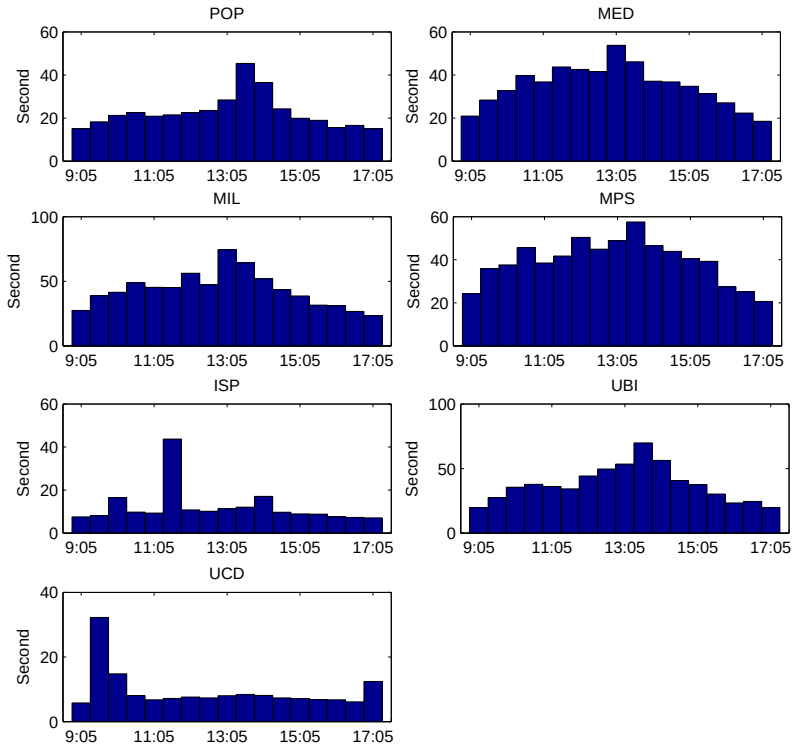

Figure 1.5: This figure illustrates a periodic pattern of durations for seven Italian banks in *Borsa di Milano*. A sampling interval of $\Delta t = 0.5$ hour is chosen. The day (weekday) is subdivided into 17 intervals from 9:05 to 17:25 with 16 half-hourly intervals and one 20 minutes as the last subinterval (from 17:05 to 17:25). The horizontal axis denotes the time of day and vertical axis denotes average duration time measured in seconds. Each bin indicates half-hourly interval except the last bin with 20 minutes. The height of each bin indicates the mean value of duration during this half-hourly interval (20 minutes for the last bin). The sampling period covers 3 weeks from 27/10/2008 to 14/11/2008.

hourly time interval and the last bin means the 20-minute time interval. The height of each bin represents the frequency of data. As shown from Figure 1.6, the last bin of each panel is lower than we expected. Recall that the last subinterval contains only 20 minutes while the others are 30 minutes, if we multiple 1.5 of that height, we will obtain a 'smile' shape for each panel. For the most frequent data, Unicredit bank (UCD), such kind of smile shape is still observed but with less obvious with respect to the others. The reason is simple that day time does not impact on the trading activity (average number of tick) on average based on half-hourly time interval for the frequently traded asset. While most of the banks (POP, MED, MIL, MPS, ISP and UBI) are highly traded near the market opening time and follows less activity in the subsequent time, the least active time period is around 13:05 to 14:05 which is due to lunch break, then start to active again until the closing time. The sampling period covers 3 weeks from 27/10/2008 to 14/11/2008.

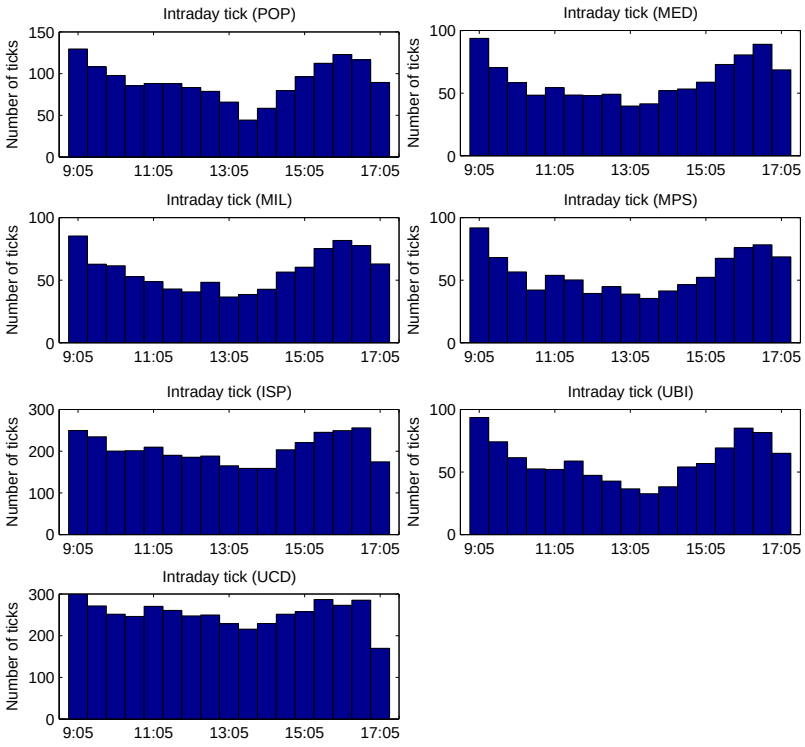

Figure 1.6: This figure illustrates a periodic pattern of intraday distribution of the tick frequency for seven Italian banks in *Borsa di Milano*. A sampling interval of $\Delta t = 0.5$ hour is chosen. The day (weekday) is subdivided into 17 intervals from 9:05 to 17:25 with 16 half-hourly intervals and one 20 minutes as the last subinterval (from 17:05 to 17:25). Each observation of the analyzed variable is made in one of these subintervals. Note that the last bin (time from 17:05 to 17:25) of each panel is lower than we expected. This is due to the smaller time interval with 20 minutes while the other bins have 30 minutes. The sample pattern does not account for weekends. The sampling period covers 3 weeks from 27/10/2008 to 14/11/2008.

Intraday volatility in terms of mean absolute returns is plotted against time of day (from 9:05 to 17:25) in Figure 1.7 for seven Italian banks in *Borsa di Milano*. Again, the day

(weekday) is subdivided into 17 subintervals from 9:05 to 17:25 with 16 of 30-minute intervals and one 20 minutes as the last subinterval (from 17:05 to 17:25). The horizontal axis is measured in the time of day; the vertical axis is absolute returns which need to be multiple of 0.001. The sampling period covers 3 weeks from 27/10/2008 to 14/11/2008. The figures indicate distinctly uneven intraday volatility patterns. The intraday volatility is higher near the opening time, say from 9:05 to 10:05, than the rest of time of day. Indeed, there is no common time that these seven banks have the lowest volatility. Moreover, unlike the Foreign Exchange market, stock exchanges and money market exchanges where the shape of the seasonality present so-called U-shape, what we observe from the data could be referred to 'opposite-J' or the 'first-half-U' shape. Such kind of result may due to some market condition during that period (sampling period from 27/10/2008 to 14/11/2008), such as significant economic announcement.

The seasonality in terms of volatility also found for USD-DEM (Foreign Exchange rate) but the conventional U-shape is not presented, as investigated by Dacorogna et al. (2001). In that case, the day is subdiveded into 24 hours from 0:00-1:00 to 23:00-24:00 (GMT) and a sampling interval is 1 hour. The explaination of volatility pattern (non-U-shape) that they consider the structure of the world market, which consists of three main parts with different time zones: America, Europe, and East Asia. The patterns are characterized by the weight of the market, such as the main daily maximum volatility occurs when both the Amerian and the European markets are active, a higher volatility for the USD-JPY when the East Asian markets are active.

The natural way to analyze seasonal heteroskedasticity in the form of weekly volatility patterns is through the intraweek statistics. Table 1.6 shows the average number of ticks for each day of week for the POP, MED, MIL, MPS, ISP, UBI and UCD from *Borsa di Milano* including three weeks period from October 27, 2008, to November 14, 2008. In general, lower activity in Monday and Friday and higher activity in the rest of weekdays.

Table 1.6: Average number of ticks for each day of the week for the POP, MED, MIL, MPS, ISP, UBI and UCD from *Borsa di Milano*. The sampling period is from October 27, 2008, to November 14, 2008.

Day of the week	POP	MED	MIL	MPS	ISP	UBI	UCD
Monday	1396	910	942	931	3011	834	4249
Tuesday	1678	1078	944	946	3574	1087	4120
Wednesday	1513	1114	1237	1105	3764	987	4525
Thursday	1598	1167	861	914	3507	970	4358
Friday	1480	876	821	799	3353	1129	4230

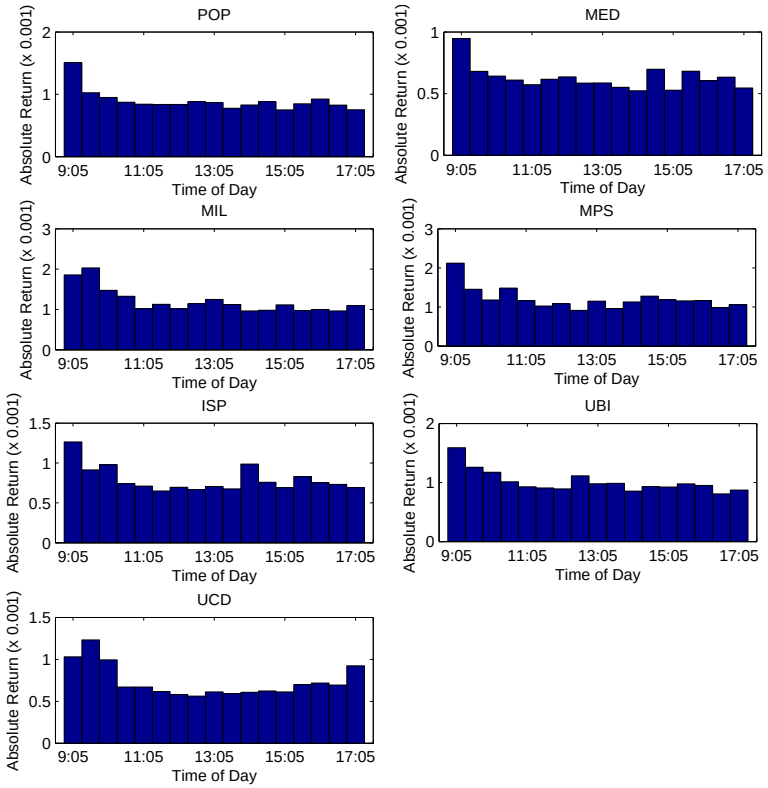

Figure 1.7: This figure illustrates a periodic pattern of intraday distribution of the absolute return for seven Italian banks in *Borsa di Milano*. A sampling interval of $\Delta t = 0.5$ hour is chosen. The day (weekday) is subdivided into 17 intervals from 9:05 to 17:25 with 16 half-hourly intervals and one 20 minutes as the last subinterval (from 17:05 to 17:25). Each observation of the analyzed variable is made in one of these subintervals. The sample pattern does not account for weekends. The sampling period covers 3 weeks from 27/10/2008 to 14/11/2008.

1.3 Multivariate ultra-high-frequency data

In describing the multivariate time series of ultra-high-frequency data, an important challenge is to study correlation between transactions which arrive inherently in irregular time intervals. The standard econometric techniques are based on fixed time interval analysis, where the assumption is that the calculated correlation represents the true linear relationship between two time series over time. However, for the irregularly-spaced tansaction data, which may arrive hourly or even few seconds, the most challenge is the asynchronization among different series. This is clearly shown on Figure 1.1. There is a natural inclination for econometrician to aggregate transaction data to some fixed time interval.

Nevertheless, if a short time interval is chosen, there will be many intervals with no new information and heteroskedasticity of a particular form will be introduced into the data. On the other hand, if a large interval is chosen, the microstructure features of the data will be lost. Consequencly, making the choice of an 'optimal' interval seems very difficult. For stocks, market activity is higher near the open and the close time and less active in the middle of the day (lunch break). For currency markets, it's more complicated due to different time zones and business holidays. But there are clear periods of high and low activity as markets arround the world open and close time. More details refer to Section 1.2.6. One more abstruse thing is that some assets are generally in low frequency but might suddenly exhibit very high activity. This may suggest to some observable event such as a news release or to an unobservable event captured by some specialists. In statistical point of view, we can consider such behavior of financial data as a stochastic process, in particular, we consider a doubly stochastic Poisson process (see Chapter 2). In short, there is no privileged time interval at which the data should be investigated. We thus study dynamic correlations of transaction data, that is, we evaluate the correlations as a function of sampling interval. This links to so-called 'Epps effect' (Epps, 1979) which highlights the dependence of stock return cross-correlations on the sampling frequency of the financial data.

In time series analysis, equally spaced time series are called homogeneous, while unequally spaced time series are called inhomogeneous. When considering the spacing of data in time, a discussion of the time scale is necessary. Many time series of daily data in finance, for example, have only five observations per week; there are no observations on Saturdays and Sundays. Such a time series is homogeneous only if using a special 'business time' scale, which omits weekends (and holidays). The term 'homogeneous' and 'inhomogeneous' have to be understood in the context of the chosen time scale. Inhomogeneous time series by themselves are conceptually simple, no fixed or random time scales.

In order to give out an explicit imagination, Figure 1.8 illustrates a graphical example of market activity of two assets. The horizontal black arrow indicates time axis on which each cross mark denotes transaction time. The corresponding logprice is shown by vertical blue line. The vertical dash line represents the time grid such that the time is cutten into equally spaced time scale. The red arrow takes the last transaction logprice to the fixed time scale (previous-tick interpolation, formal definition see the following). From Figure 1.8, the first asset is more frequent than the second one and asynchronicity of two assets is obvious, standard Pearson correlation measure plays no role in this case. For instance, it is difficult to say the correlation is measured between $\log P_2$ (the first asset) and $\log P_1$ (the

second asset) or between $\log P_2$ (the first asset) and $\log P_2$ (the second asset), since they are not synchronized.

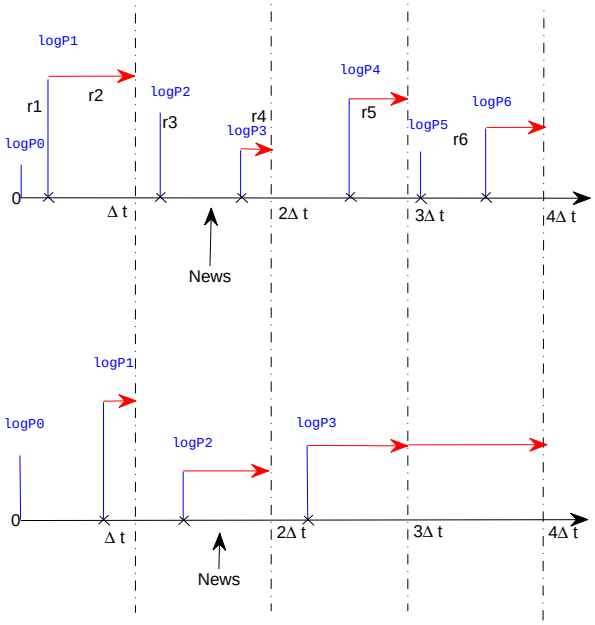

Figure 1.8: A graphical example of market activity of two assets. The cross marks on each axis indicate transaction times and the associated vertical solid lines indicate logprices. The vertical dash lines indicate the fixed time scale. The arrows take the last transaction logprice to the fixed time scales.

To overcome the irregular temporal structure, in the literature, there are mainly two methods: first, select an interpolation scheme (i.e. linear, previous-tick, etc.) and a sampling interval and then homogenising the actual price series by imputing values at the equidistant points where original values may not exist. Second, adapt a Fourier method which can be directly applied to the raw time series to obtain correlation statistics. See further reference given by Barucci and Renò (2002), Renò (2003).

In this thesis we adapt interpolation based method, in particular, previous-tick interpolation, to compute correlations from ultra-high-frequency data. As part of the analysis we also endeavour to characterise the Epps effect and show that trading synchronicity and stock liquidity are relevant factors.

In what follows, two interpolation methods will be introduced, both conceptually and computationally, such that the raw time series (inhomogeneous) is then transformed to a homogeneous time series which is suitable to work with. Especially, standard techniques of time series analysis can be applied again.

1.3.1 Synchronization by interpolation methods

Ultra-high-frequency data are inhomogeneous time series. For most methods, these raw time series are not suitable to work with, because standard techniques of time series analysis require equally spaced (homogeneous) data point. This problem can be overcome by

selecting an interpolation schema (i.e. previous-tick, linear interpolation etc.) which will be introduced in the following.

To homogenizing the raw time series with irregularly spaced in time by interpolation methods, a reference time scale should be first considered, say Δt. Let the inhomogeneous series with times t_j and corresponding value $Z_j = Z(t_j)$ where the index j denotes the irregularly spaced sequence of the raw data. By utilizing an interpolation method, we construct a homogeneous time series with values at time $t_0 + i\Delta t$, equally spaced by Δt, rooted at a time t_0. The index i refers to the homogeneous series.

The two most important interpolation methods are previous-tick interpolation and linear interpolation. For the previous-tick interpolation, it takes the most recent value before the sampling interval. It is defined by griding the time axis into fixed time interval as $t_0 + i\Delta t$, the corresponding value for these artificial time interval take the most recent value of the raw data. More formally,

$$Z(t_0 + i\Delta t) = Z_{j'}. \tag{1.3.1}$$

with

$$j' = \max\{j | t_j \leq t_0 + i\Delta t\},\ t_{j'} \leq t_0 + i\Delta t < t_{j'+1}. \tag{1.3.2}$$

where j' denotes the most recent time index before time interval $t_0 + i\Delta t$.

For the linear interpolation, as the name implies, this method is applied to interpolate two points by a line. For instance, if we want to interpolate time between $t_{j'}$ and $t_{j'+1}$, the value of time $t_0 + i\Delta t$ is given by

$$Z(t_0 + i\Delta t) = Z_{j'} + \frac{t_0 + i\Delta t - t_{j'}}{t_{j'+1} - t_{j'}}(Z_{j'+1} - Z_{j'}). \tag{1.3.3}$$

Figure 1.9 illustrates a graphic example for two interpolation methods. Each stem presents a transaction which contains its occurrence time and the associated price. To interpret the interpolation with above formula, here we set $\Delta t = 5$, so the sychronized time series will be measured in 5-second time interval with the index i. $t_0 = 0$ and j' is the most recent time index of time $t_0 + i\Delta t$ (formally, see Equaation 1.3.2), in Figure 1.9, the arrow can be interpreted as the previous-tick interpolation taking the most recent price $Z_{j'}$ for the fixed time scale $t_0 + i\Delta t$, and the red line represents the linear interpolation on which the yellow rectangular indicates the new value according to Equation 1.3.3.

The only difference between these two interpolation methods is the later one is definitely artificial in the sense that there does not exist such kind of value in the data except the two successive tick with the same value, that is, $Z_{j'+1} = Z_{j'}$. Both methods have their merits and disadvantages. Previous-tick method respect causality as it exclusively uses information already known at previous time, whereas linear interpolation uses infromation from time $t_{j'+1}$, which lies in the future of time $t_0 + i\Delta t$. When using previous-tick interpolation over a gap (a long period of missing data) in the raw data, a spurious jump of Z may be observed at the end of the gap, which may spoil a statistical analysis of extreme returns of Z. In case of linear interpolation, on one hand, some of the original data (which fall between the fixed time scales) is removed from the computations whilst on the other hand we use generated data point that does not exist and this can distort the data. Throughout this thesis, we choose previous-tick interpolation both for empirical data analysis (in this Chapter) and for simulation study (see Chapter 3).

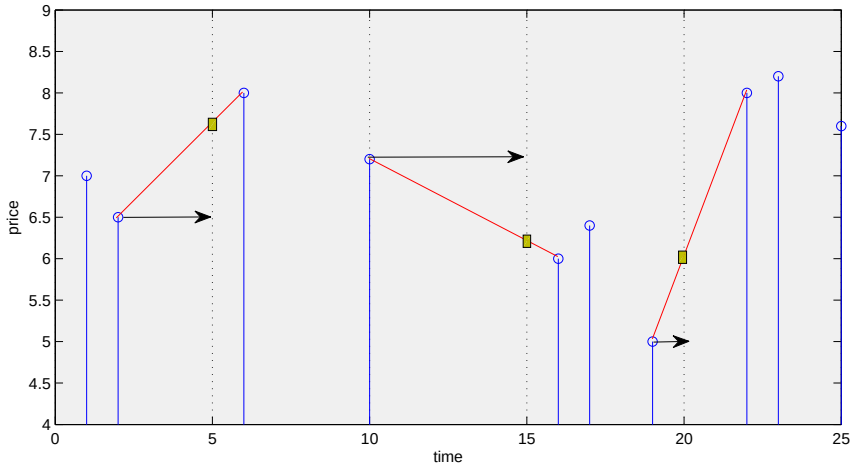

Figure 1.9: A graphic example of interpolation. Each stem indicates one transaction which contains its occurrence time and the associated price. The arrow represents the previous-tick interpolation and the red line represents the linear interpolation on which the yellow rectangular is the new value for time $t_0 + i\Delta t$ according to Equation 1.3.3.

It is worth to stress that previous-tick interpolation is, in fact, a process of logreturn aggregation. To explain this fact, we illustrate an example as shown in Figure 1.10, in which it shows how the logreturns were aggregated by the previous-tick sychronization. The arrow indicates the time horizon on which the cross mark represents transaction (tick) time, and the corresponding bue line stands for logprice. Between each cross mark, r_i, $i = 1, 2 \ldots, 6$, denote the logreturns for the corresponding time intervals, where $r_i = \log P_i - \log P_{i-1}$, $i = 1, 2 \ldots, 6$. Let the two vertical dash lines as the time grid line so that time horizone is divided into fixed time intervals, denoting Δt. The red arrow indicates the previous-tick interpolation as already shown in the previous figure (see Figure 1.9). Now consider the first fixed time interval, the logreturn of this time interval is then given by,

$$\begin{aligned} r_{\Delta t} &= \log P_3 - \log P_0 \\ &= \log P_3 - \log P_2 + \log P_2 - \log P_1 + \log P_1 - \log P_0 \\ &= r_3 + r_2 + r_1. \end{aligned}$$

Thus, previous-tick interpolation is in line with econometric model that aggregate the logreturns to some fixed time interval. In this way, the standard correlation measurement can be applied again. Again, however, if a short time interval is chosen, there will be many intervals with no new information and heteroskedasticity of a particular form will be introduced into the data. On the other hand, if a large interval is chosen, the microstructure features of the data will be lost. As discussed in the beginning of this section, there is no priviledged sampling interval, hence we need to consider a dynamic sampling interval, which is relevant to Epps effect (see Section 1.3.3).

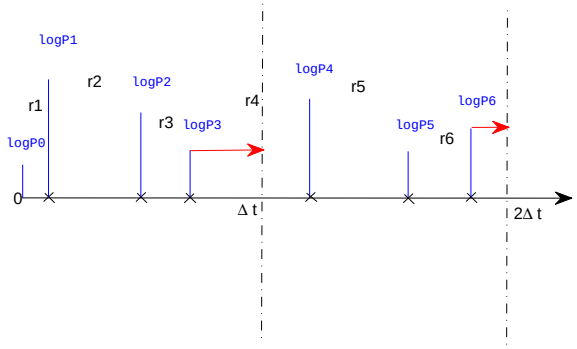

Figure 1.10: An example for previous-tick synchronization.

1.3.2 Correlation measures

Analyzing the correlation and covariates between financial time series is a standard technique used in empirical finance. Often this is estimated quantatively using the linear correlation coefficient, which is a basic mearsurement of the linear dependence or independence between variables. The popularity of this measure stems from its simple definition, practical ease of use, and its straightforward results, which are easily interpreted, scale free, and directly comparable.

On the other hand, the linear correlation coefficient calculation greatly discards the time variables. Because the correlation coefficient calculation only require the data input with the same length. No matter what is the subscript (time) of each variable, hourly or secondly. The variables of two time series and their covariance are constructed either with the assumption of being constant or as a type of average value if value changes are recognized. It is generally accepted that correlations in financial time series vary over time and are even subject to correlation 'breakdown' or large changes in correlation in critical periods. In the following, we investigate the correlation as a function of time interval using ultra-high-frequency data.

Although the calculation of the correlation coefficient is well defined and rather simple, a number of unresolved issues exist with respect to application of the rule and interpretation of results in the ultra-high-frequency data. The impact of time series data frequency on correlations should also be clearly established, this especially relevant as ultra-high-frequency data becomes more widely available. Previous authors have demonstrated a dramatic decrease in correlation as data frequency enters the intra-hour level, for both stock (Epps, 1979) and foreign exchange returns (see Guillaume et al., 1997; Low et al., 1996). This discussion attemps to characterize and investigate more deeply the Epps effect in a number of financial time series.

Lagged correlation is a more powerful tool to investigate the relation between two time series. The lagged correlation function considers the two time series not only simultaneously (at lag 0) but also with a time shift. The correlation coefficient ρ_τ is measured and plotted against the value of the lag. Lagged correlation reveals causal relation and information flow structures in the sense of Granger causality, if two time series are generated on the

basis of a synchronous information flow, they would have a symmetric lagged correlation function, $\rho_\tau = \rho_{-\tau}$. The symmetry would be voilated only by insignificant small, purely stochastic deviations. As soon as the deviations between ρ_τ and $\rho_{-\tau}$ become significant, there is asymmetry in the information flow.

Covolatility

The calculation of correlation coefficients is straightforward but some inconvenience is inherently introduced by its simple definition, especially, homogeneous time seres requirement. However, the problem requires more careful treatment at ultra-high-frequency data where one cannot dictate the time or number of observations. One often faces the main problem when estimating correlation between two high-frequency time series. It involves correlating two time series of inherently different frequencies. If the two time series are both regular with respect to data arrival intervals but of different frequencies, one might create from them two equally spaced, homogeneous time series, which both have frequencies equal to the lesser frequent of the two. This easy situation does not occur very often, though.

It is more common to face with time series such as stock in *Borsa di Milano* and foreign exchange (FX) rates where data frequency can vary from very few quotes to hundreds of quotes per hour. Then we would like to ask, what is the best way to measure the dependence between two time series with different frequencies and activity peaks and valleys at completely different daytimes? Ideally, one would prefer the correlation calculation to be updated more often when more information exists and less often when it does not exist.

One way to do this is to introduce a time scale that compresses physical time, if there is no information and to expand it when it exists, as proposed by Dacorogna et al. (2001). This method has been found useful for a number of applications, but is time-consuming to implement in practice. Moreover, it has the multivariate problem of two time series for which it would needs a common time scale. The other way is to use some form of data interpolation.

Formulation of an adjusted correlation measure

An extension of the standard correlation measure is proposed in Dacorogna et al. (2001) by incorporating a 'covolatility weighting' for the time series. The weight has the role of emphasising periods where trading has a noticeable effect on asset prices. The idea is to develop a measure of correlation where information exists and to avoid updating our measure where data do not exist. This implies that a lower data frequency or data gaps in one time series may limit the use of another one, and the unavoidable price to pay is a certain loss of statistical significance.

Let X and Y be two asset price time series which have been homogenised and synchronised to a time step Δt, the time length of the trading period is T and $T = n\Delta t$. The standard linear correlation coefficient is a measure of correlation between two time series X_i and Y_i and is defined as follows:

$$\rho(x_i, y_i) = \frac{\sum_{i=1}^{n}(X_i - \bar{X})(Y_i - \bar{Y})}{\sqrt{\sum_{i=1}^{n}(X_i - \bar{X})^2 \sum_{i=1}^{n}(Y_i - \bar{Y})^2}}, \tag{1.3.5}$$

with the sample means

$$\bar{X} = \sum_{i=1}^{n} \frac{X_i}{n} \quad \text{and} \quad \bar{Y} = \sum_{i=1}^{n} \frac{Y_i}{n}. \tag{1.3.6}$$

Correlation values ρ are unitless and may range from -1 (perfectly anticorrelated) to 1 (perfectly correlated). A vaule of zero indicates two uncorrelated series. Note that the two variables X_i and Y_i are usually returns of two financial assets. More often researches assume that the mean value of returns equal to zero.

To estimate the local covolatility for each of these observations, it is defined by further dividing each time span Δt over which X_i and Y_i are calculated into m equal subintervals from which subreturn values, $\breve{X}_j$ and $\breve{Y}_j$, can be obtained as corresponding returns on a smaller time scale $\Delta\breve{t}$. This redefined time series now consists of $\breve{n} = T/\Delta\breve{t}$ equally spaced return observations where $\Delta t \equiv m\Delta\breve{t}$.

For each of the previous coarse returns, X_i and Y_i, there exists a corresponding estimation of covolatility between the two homogeneous time series of returns

$$\omega_i(\breve{X}_i, \breve{Y}_i, \breve{t}) = \sum_{j=1}^{m} \left(\left|\breve{X}_{i,m-j} - \left\langle \breve{X}_m \right\rangle\right| \cdot \left|\breve{Y}_{i,m-j} - \left\langle \breve{Y}_m \right\rangle\right|\right)^{\alpha}, \tag{1.3.7}$$

where

$$\left\langle \breve{X}_m \right\rangle = \sum_{j=1}^{m} \frac{\breve{X}_{i,m-j}}{m} \quad \text{and} \quad \left\langle \breve{Y}_m \right\rangle = \sum_{j=1}^{m} \frac{\breve{Y}_{i,m-j}}{m}. \tag{1.3.8}$$

The most common choice for α is 0.5, though this can be investigated as a way to magnify or demagnify the weight given to farther outlying return values. By setting $\alpha = 0$, in Equation 1.3.7, $\omega_i = m$. This result reduces the following Equation 1.3.9 to the standard correlation coefficient. If $\omega_i = 0$, it implies that returns derived from interpolated (linear or previous-tick) prices existing outside of region of interest, Δt.

In Dacorogna et al.(2001), they specify an adjusted correlation calculation by inserting weights in all the sums of covolatility calculation:

$$\breve{\rho}(X_i, Y_i, \omega_i) = \frac{\sum_{i=1}^{T/\Delta t}(X_i - \bar{X})(Y_i - \bar{Y})\omega_i}{\sqrt{\sum_{i=1}^{T/\Delta t}[(X_i - \bar{X})^2\omega_i]}\sqrt{\sum_{i=1}^{T/\Delta t}[(Y_i - \bar{Y})^2\omega_i]}}, \tag{1.3.9}$$

where the sample means $\bar{X}$ and $\bar{Y}$ are not as defined in Equation 1.3.6, but are given by

$$\bar{X} = \frac{\sum_{i=1}^{T/\Delta t}(X_i \cdot \omega_i)}{\sum_{i=1}^{T/\Delta t}\omega_i} \quad \text{and} \quad \bar{Y} = \frac{\sum_{i=1}^{T/\Delta t}(Y_i \cdot \omega_i)}{\sum_{i=1}^{T/\Delta t}\omega_i}. \tag{1.3.10}$$

Note that, as in Equation 1.3.5, X_i and Y_i in Equation 1.3.9 are logarithmic returns taken over the same time period Δt. These coarse return values can be defined as the sum of the finer return values $\breve{X}_{i,m-j}$, that is,

$$X_i = \sum_{j=1}^{m} \breve{X}_{i,m-j}. \tag{1.3.11}$$

The performance of covolatility adjusted correlation measure proposed by Dacorogna et al. (2001) is tested by synthetic Monte Carlo simulation. Two independently and identically Normal distributed random time series, A_i and B_i, with mean zero and standard deviation $\sigma = 0.01$, sample size $n = 10,000$. A third series, C_i, can then be formed as a linear combination of the previous two, A_i and B_i, and is given by

$$C_{i=1}^{n} = lA_{i=1}^{n} + (1-l)B_{i=1}^{n} \tag{1.3.12}$$

where the constant l is a parameter satisfying $0 \leq l \leq 1$. Equation 1.3.12 defines the new series C_i has a controllable correlation to the original data series A_i.

Building another new series of returns, D_i, by linear interpolation method (see Equation 1.3.3). In practice, each consisting of 50 price observations were then deleted and replaced by prices linearly interpolated from the price bracketing the deleted sections. The distance between these artificial data gaps also consisted of 50 observations, creating an alternating series of original data patches followed by data gaps filled with linearly interpolated prices.

Table 1.7: Results of a Monte Carlo simulation of correlations. Comparing the covolatility adjusted linear correlation $\breve{\rho}$ to the standard linear correlation ρ, both applied to synthetic time series. The series D_i is like C_i (defined in Equation 1.3.12), but regularly spaced sections of the data are replaced by linearly interpolated data. Details are described in the text. Note the similarity of the second column $\rho(A_i, C_i)$ and the fourth column $\breve{\rho}(A_i, D_i)$.

l Equation 1.3.12	$\rho(A_i, C_i)$ Equation 1.3.5	$\rho(A_i, D_i)$ Equation 1.3.5	$\breve{\rho}(A_i, D_i)$ Equation 1.3.9
0.0	0.00	0.00	0.00
0.1	0.12	0.10	0.12
0.2	0.23	0.15	0.22
0.3	0.38	0.28	0.38
0.4	0.52	0.40	0.51
0.5	0.69	0.51	0.69
0.6	0.83	0.62	0.82
0.7	0.92	0.67	0.91
0.8	0.97	0.72	0.95
0.9	0.99	0.74	0.97
1.0	1.00	0.74	0.99

Results of comparison to the standard linear correlation calculation are shown in Table 1.7. A comparison of the seconod columns $\rho(A_i, C_i)$ and fourth column $\breve{\rho}(A_i, D_i)$ shows that the covolatility adjusted correlation measure described by Equation 1.3.9 successfully approximates the original standard linear correlation between distributions A and C before some data patches were replaced by linearly interpolated values.

As we have described before, the most peculiar feature of ultra-high-frequency data is that all transaction data are inherently irregularly spaced in time, which leads the standard correlation study of time series become useless. Because the standard correlation calculation is based on regularly spaced time series. However, if interpolation methods are applied

to the raw data, then correlation between any two of UHF data will be obtain through the standard way. But the resulting 'correlation' has already lost its original meaning. For example, correlation coefficient ρ is the correlation between tick i (asset 1) and tick j (asset 2), or between tick $i-1$ (asset 1) and tick j (asset 2), $i = 1, 2, \ldots, 6$, $j = 1, 2, 3$ (see Figure 1.8). Here tick refers to cross mark as shown in Figure 1.8. For this reason, we put quote mark on the term correlation. Whenever asynchronous data involved, the standard correlation must be viewed in different way.

The impact of time series data frequency on correlations should be clearly established. This is especially relevant as higher frequency data becomes more widely availalbe and more often used in order to improve statistical tools. Often measuring the dependence or independence of financial time series is estimated quantatively using the linear correlation coefficient (Pearson correlation coefficient), which is a standard measurement of the dependence between variables of interest. The popularity of this measure stems from its simple definition and its straightforward results, which is easily interpreted and directly comparable. A problem arises when the two time series of unregularly spaced tick-by-tick data have different frequencies which involve with trading time may or may not overlap.

It is generally accepted that correlations in financial time series vary over time and are even subject to large changes in correlation in critical periods. To do such kind of correlation adjustment, one would prefer the correlation calculation to be updated more often when more information arrives and less often when there is less information. However, this method is time-consuming to implement in practice.

Some methods for approximating a homogeneous time series from unevenly spaced, tick-by-tick data involve some form of data imputation. Methods of imputating data vary in complexity and effectiveness and most have been found to be beneficial (methods of imputation to get homogeneous time series provide easy calculation or understand) under some set of conditions and assumptions. Nevertheless, all forms of imputation rely on a model, and a standard supposition is that critical characteristics of the data do not change between in-sample and out-of-sample periods. In recent literature, B.Tòth & J.Kertesz (2008) investigated into decomposition of the cross-correlations for tick-by-tick data. They intended to connect the cross-correlation on a certain time scale to lagged autocorrelations and cross-correlations on smaller time scales with previous-tick synchronization approach. Dacorogna et al. (2001) proposed a method similar to the standard correlation coefficient with a weight factor that depends on the joint volatility of the time series. This method also requires simultaneous time series and linear interpolation is implemented.

1.3.3 Epps effect

Epps (1979) reported empirical evidence that stock correlation decreases when sampling frequency increases. In Epps (1979), it investigated comovements in Stock Prices during very short periods in response to the normal flow of economic and political news during the course of the trading day. And Epps (1979) observed that "correlations among price changes in common stocks of companies in one industry are found to decrease with the length of the interval for which the price changes are measured". This phenomenon is so called Epps effect for acknowledgement of the first identifiable author to thoroughly document it. In fact, this feature of ultra-high-frequency data is in part a manifestation of information aggregation process in the stock market. For this explaination, we defer to Section 1.3.4.

Considering nonoverlapping time intervals, if the two covariance-stationary processes, x_t and y_t, have zero mean, then the correlation between x_t and y_t equals to the nonoverlapping sums, $X_{tk} = \sum_{i=0}^{k-1} x_{tk-i}$ and $Y_{tk} = \sum_{i=0}^{k-1} y_{tk-i}$. It is easily seen that the correlation between X_{tk} and Y_{tk} equals that between x_t and y_t if

$$\begin{aligned}
\mathrm{E}(x_{t-i}y_{t-j}) &= \sigma_{xy},\ i = j \\
&= 0,\ i \neq j \\
\mathrm{E}(x_{t-i}x_{t-j}) &= \sigma_{x^2},\ i = j \\
&= 0,\ i \neq j \\
\mathrm{E}(y_{t-i}y_{t-j}) &= \sigma_{y^2},\ i = j \\
&= 0,\ i \neq j
\end{aligned}$$

However, Epps (1979) observed that the correlations decrease significantly as the interval declines from three hours. Note that logreturns for intervals longer than 10 minutes are merely nonoverlapping sums of logreturns for 10-minute intervals. It is controversial with the hypothesis that price-change series are stationary and zero correlations between price changes at different periods, namely, nonstationary price change and existence of either(or both) autocorrelations between price changes for one stock in different periods or lagged cross-correlations between price changes of different stocks in different periods.

Whereas most of the recent papers suggest that there are two main factors causing Epps effect: the first one is a possible lead-lag effect between stock returns, the other one is the asychronicity of ticks in case of different stocks. For example, Zebedee (2001) argues that the Epps effect is mainly due to the lead-lag relationship. On the other hand, Lundin et al. (1999) claim that different assets play different roles at different frequecies, so that it is not possible to recover the same correlation at different time scales. Moreover, Lundin et al. (1999) find a significant inverse relation between correlation and activity: the more an asset is traded, less evident is the Epps effect. This implies that the synchronicity explains the correlation decrease at higher frequencies. Renó (2003) also confirms that Epps effect mostly due to synchronicity by using Monte Carlo simulation method. Similarly, he found an inverse relation between trading activity and the correlation drop.

Figure 1.11 illustrates a graphical example of Epps-curve. The behavior of correlation coefficient, as a function of the sampling frequency, shows unstable relation between two

assets. Enter certain sampling frequency (time scale=20 in Figure 1.11), there is a dramatical decrease of correlation coefficient as sampling frequency increase. But when sampling frequency greater than that certain sampling frequency (time scale=20 in Figure 1.11), correlation coefficient reaches to its stable level.

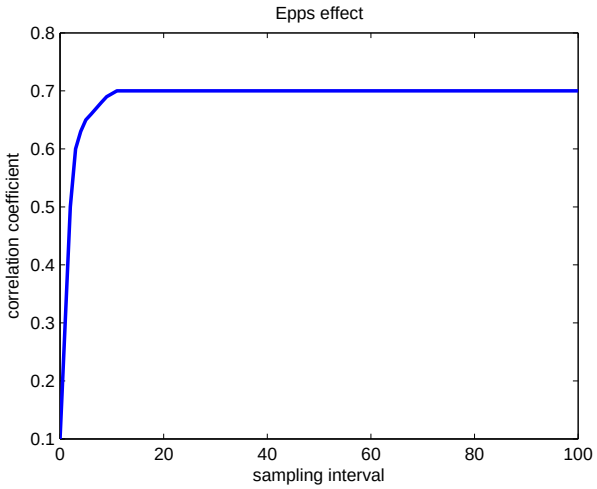

Figure 1.11: A graphical example of Epps effect. It shows the correlation coefficient as a function of sampling interval. Note the time scale refers to sampling interval.

Decomposition of correlations

Considering two assets A and B, the price $P^A(t)$ and $P^B(t)$ are already synchronized by previous-tick interpolation with reference time step Δt, so $t = i\Delta$, $i = 0, 1, \ldots, N$, and $T = N\Delta t$. Then the logreturn will be denoted by

$$R^A_{\Delta t}(t) = \log P^A(t) - \log P^A(t - \Delta t). \tag{1.3.13}$$

The general Pearson correlation measure with time lag τ is defined by

$$C^{A/B}_{\Delta t}(\tau) = \frac{\left\langle R^A_{\Delta t}(t) R^B_{\Delta t}(t+\tau)\right\rangle - \left\langle R^A_{\Delta t}(t)\right\rangle \left\langle R^B_{\Delta t}(t+\tau)\right\rangle}{\sigma^A \sigma^B}, \tag{1.3.14}$$

where

$$\langle R_{\Delta t}(t)\rangle = \frac{1}{T - \Delta t} \sum_{i=\Delta t}^{T} R_{\Delta t}(i), \qquad \sigma = \sqrt{\langle R_{\Delta t}(t)^2\rangle - \langle R_{\Delta t}(t)\rangle^2}, \tag{1.3.15}$$

and T is the total length of time. Obviously, when time lag $\tau = 0$, $\rho^{A/B}_{\Delta t} \equiv C^{A/B}_{\Delta t}(\tau = 0)$.

Now let us consider a smaller sampling step, Δt_0, and $\Delta t = n\Delta t_0$ with n being a positive integer, the change in the measured quantity in the time window Δt is merely the sum of changes of shorter, non-overlapping time window Δt_0:

$$R_{\Delta t}(t) = \sum_{s=1}^{n} R_{\Delta t_0}(t - \Delta t + s\Delta t_0). \tag{1.3.16}$$

Using this result, we can rewrite the component of Equation 1.3.14 in terms of smaller sampling step Δt_0,

$$\begin{aligned}
\left\langle R^A_{\Delta t}(t) R^B_{\Delta t}(t+\tau) \right\rangle &= \frac{1}{T-\Delta t} \sum_{i=\Delta t}^{T} R^A_{\Delta t}(i) R^B_{\Delta t}(i) \\
&= \frac{1}{T-\Delta t} \sum_{i=\Delta t}^{T} \Big(\sum_{s=1}^{n} R^A_{\Delta t_0}(t-\Delta t+s\Delta t_0) \Big) \\
&\quad \Big(\sum_{s=1}^{n} R^B_{\Delta t_0}(t-\Delta t+s\Delta t_0) \Big) \\
&= \sum_{s=1}^{n} \sum_{q=1}^{n} \Big\langle R^A_{\Delta t_0}(t-\Delta t+s\Delta t_0) R^B_{\Delta t_0}(t-\Delta t+s\Delta t_0) \Big) \Big\rangle \\
&= \sum_{x=-n+1}^{n-1} (n-|x|) \left\langle R^A_{\Delta t_0}(t) R^B_{\Delta t_0}(t+x\Delta t_0) \right\rangle . \qquad (1.3.17)
\end{aligned}$$

Similarly,

$$\left\langle R^A_{\Delta t}(t)^2 \right\rangle = \sum_{x=-n+1}^{n-1} (n-|x|) \left\langle R^A_{\Delta t_0}(t) R^A_{\Delta t_0}(t+x\Delta t_0) \right\rangle , \qquad (1.3.18)$$

$$\left\langle R^B_{\Delta t}(t)^2 \right\rangle = \sum_{x=-n+1}^{n-1} (n-|x|) \left\langle R^B_{\Delta t_0}(t) R^B_{\Delta t_0}(t+x\Delta t_0) \right\rangle . \qquad (1.3.19)$$

Consequently, we can obtain the following relationship between correlations on the two different time scales

$$\begin{aligned}
\rho^{A/B}_{\Delta t} = \Big(& \textstyle\sum_{x=-n+1}^{n-1} (n-|x|) \left\langle R^A_{\Delta t_0}(t) R^B_{\Delta t_0}(t+x\Delta t_0) \right\rangle \\
& -n^2 \left\langle R^A_{\Delta t_0}(t) \right\rangle \left\langle R^B_{\Delta t_0}(t+x\Delta t_0) \right\rangle \Big) \times \\
\Big(& \textstyle\sum_{x=-n+1}^{n-1} (n-|x|) \left\langle R^A_{\Delta t_0}(t) R^A_{\Delta t_0}(t+x\Delta t_0) \right\rangle \\
& -n^2 \left\langle R^A_{\Delta t_0}(t) \right\rangle^2 \Big)^{-1/2} \times \\
\Big(& \textstyle\sum_{x=-n+1}^{n-1} (n-|x|) \left\langle R^A_{\Delta t_0}(t) R^B_{\Delta t_0}(t+x\Delta t_0) \right\rangle \\
& -n^2 \left\langle R^B_{\Delta t_0}(t) \right\rangle^2 \Big)^{-1/2} . \qquad (1.3.20)
\end{aligned}$$

Moreover, according to Tóth & Kertósz (2007), Equation 1.3.20 can be rewritten as

$$\begin{aligned}
\rho^{A/B}_{\Delta t} = & \Big(\textstyle\sum_{x=-n+1}^{n-1} (n-|x|) f^{A/B}_{\Delta t_0}(x\Delta t_0) \Big) \times \\
& \Big(\textstyle\sum_{x=-n+1}^{n-1} (n-|x|) f^{A/A}_{\Delta t_0}(x\Delta t_0) \Big)^{-1/2} \times \qquad (1.3.21) \\
& \Big(\textstyle\sum_{x=-n+1}^{n-1} (n-|x|) f^{B/B}_{\Delta t_0}(x\Delta t_0) \Big)^{-1/2} \rho^{A/B}_{\Delta t_0} \times
\end{aligned}$$

where

$$f_{\Delta t_0}^{A/B}(x\Delta t_0) = \frac{\left\langle R_{\Delta t_0}^{A}(t) R_{\Delta t_0}^{B}(t + x\Delta t_0) \right\rangle}{\left\langle R_{\Delta t_0}^{A}(t) R_{\Delta t_0}^{B}(t) \right\rangle}. \tag{1.3.22}$$

and similarly for $f_{\Delta t_0}^{A/A}(x\Delta t_0)$ and $f_{\Delta t_0}^{B/B}(x\Delta t_0)$, defined both for postive and negative value of x.

Thus, the correlation coefficient for any sampling time step, Δt, can be expressed by knowing the coefficient on a shorter sampling interval, Δt_0 given that Δt is multiple of Δt_0.

Empirical cross-correlation of financial data from *Borsa di Milano*

In this subsection, we present empirical results of the highly traded assets from *Borsa di Milano* (S & P 500). The sampling period covers three-week time from October 27, 2008 to November 14, 2008. The daily trading window we analyze including 8 hours and 20 minutes (equivalent to 30000 seconds) from 9:05 to 17:25. Seven Italian banks – Banco Popolare (POP), Mediobanca (MED), Banca Popolare di Milano (MIL), MPS Banca (MPS), Intesa SanPaolo Banca (ISP), UBI Banca (UBI) and Unicredit Banca (UCD)–are chosen as sources of data for study of comovements between (among) them. Transaction data are collected from an electronic limited order book of *Borsa di Milano* (S & P 500).

In Figure 1.12 through 1.18, the linear correlation coefficients were calculated as function of sampling intervals and illustrated in graphs. Note that previous-tick interpolation is applied to synchronize the raw data. The postitive cross-correlation between two assets are observed in Figure 1.12 through 1.18, although logreturns for Banco Popolare seem only weakly related to those for the other six banks, the correlations among Mediobanca (MED), Banca Popolare di Milano (MIL), MPS Banca (MPS), Intesa SanPaolo Banca (ISP), UBI Banca (UBI) and Unicredit Banca (UCD) are fairly large and stable when the logreturns pertain to sampling intervals of three hours or longer. This phenomenon is so called Epps effect as described in the beginning of this section.

The correlation coefficient between Banco Poplare and the rest six banks shown in Figure 1.12 appear largely volatile when sampling interval greater than 10000 seconds (about 3 hours), especially for POP-MPS, POP-MIL, POP-ISP, and POP-UCD. But POP-MED presents quite stable relationship as sampling interval increase. The Epps effect is replicated here with small sampling interval window, enter 2000 seconds (about half hour). On the other hand, the correlation coefficient between Medio banca and the rest six banks shown in Figure 1.13 display stablity as sampling interval become larger, around 0.65. Epps effect is also observed as the sampling interval enter 2000 seconds (about half hour).

In Figure 1.14, Banca Popolare di Milano have the highest correlation with UBI banca, almost 0.9, while about 0.7 with the rest banks, except Banco Popolare (very volatile as sampling interval greater than 3 hours). The similar resluts can be found in Figure 1.15 through Figure 1.18. Note that correlation calculated with lower data frequency are not simply an average of those calculated with higher data frequencies, this can be partially explained by Epps effect that correlation decrease dramatically enter certain sampling interval and maintain quite stable level after pass that certain sampling interval.

Finally, one question should be posed here is that, since the correlation coefficient only measures the linear dependence, two banks are really have linear relationship but not other relations, such as quadratic or exponential, for instance. Having virtually the same

statistical significance for correlation calculations as shown in Figure 1.12 through 1.18, we also illustrate scatterplot of synchronized logreturns (by previous-tick synchronization) of two banks (ISP-UCD) against different sampling intervals, as shown in Figure 1.19. The linear relationship between Intesa SanPaolo (ISP) and Unicredit Banca (UCD) exhibits the same result as shown in Figure 1.16 (left bottom panel). This implies that correlation calculations as shown in Figure 1.12 through 1.18 have statistical significance. Again, the correlation increase as sampling interval increase verifies the Epps effect.

The Epps effect has been widely associated with non-synchronous trading, when fresh observations of transactions prices do not arise simultaneouly across markets, but are separated by a few seconds, for instance. If non-synchronous trading is the source of the Epps effect, the most challenge is to explicit the account of the pure jump nature of price process.

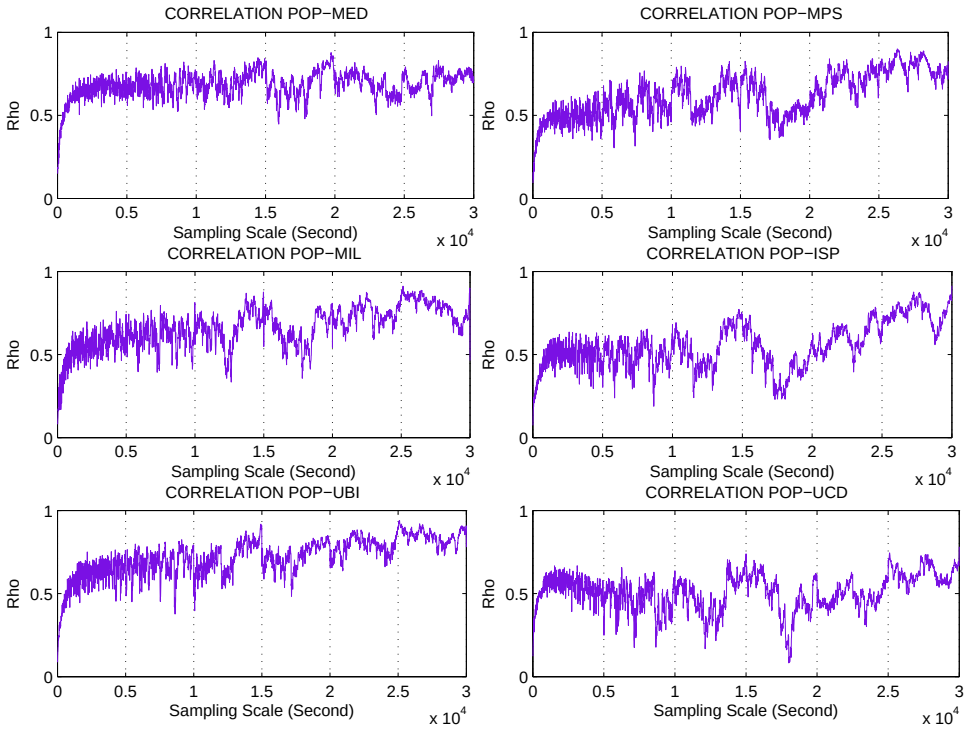

Figure 1.12: The correlation (ρ) of logreturns between Banco Popolare and Mediobanca, Banca Popolare di Milano, MPS Banca, Intesa SanPaolo Banca, UBI Banca and Unicredit Banca, respectively. Previous-tick interpolation is applied. The horizontal axis indicates the sampling interval measured in seconds. The vertical axis indicates the values taken by the correlation coefficient. The sampling period covers three weeks from $27/10/2008$ to $14/11/2008$ in *Borsa di Milano*.

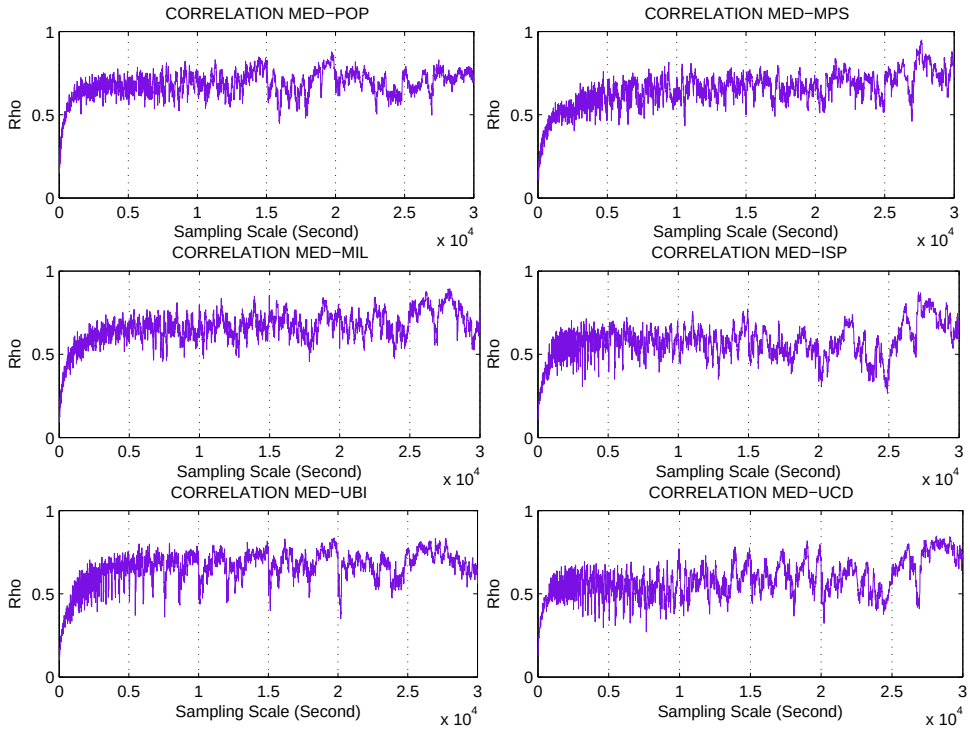

Figure 1.13: The correlation (ρ) of logreturns between Mediobanca and Banco Popolare, Banca Popolare di Milano, MPS Banca, Intesa SanPaolo Banca, UBI Banca and Unicredit Banca, respectively. Previous-tick interpolation is applied. The horizontal axis indicates the sampling interval measured in seconds. The vertical axis indicates the values taken by the correlation coefficient. The sampling data includes there-week time period from $27/10/2008$ to $14/11/2008$ in *Borsa di Milano*.

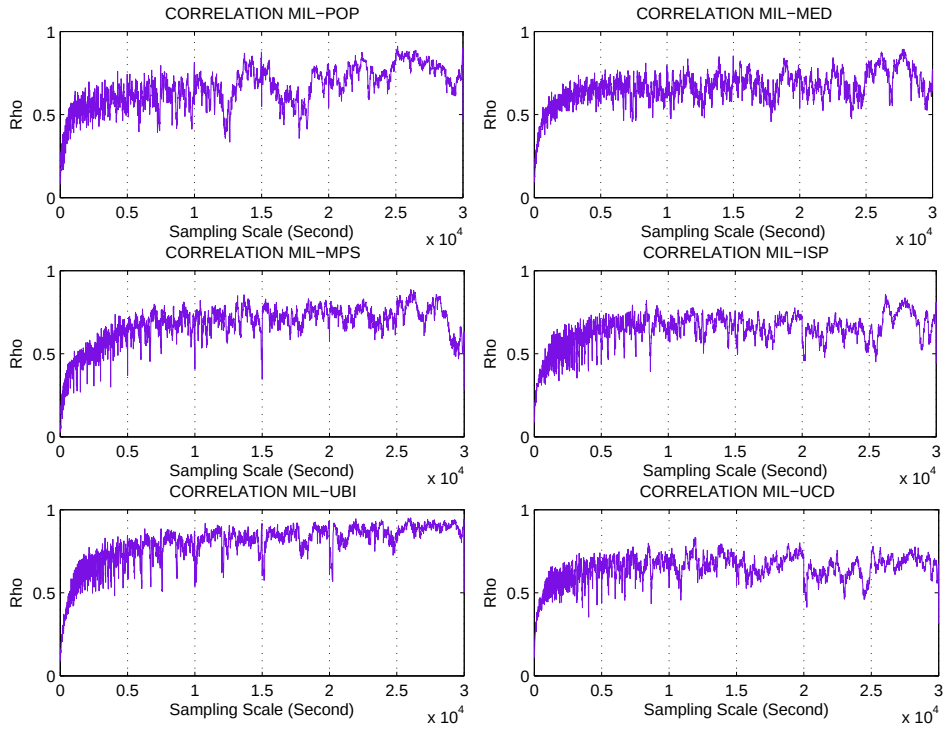

Figure 1.14: The correlation (ρ) of logreturns between Banca Popolare di Milano and Banco Popolare, Mediobanca, MPS Banca, Intesa SanPaolo Banca, UBI Banca and Unicredit Banca, respectively. Previous-tick interpolation is applied. The horizontal axis indicates the sampling interval measured in seconds. The vertical axis indicates the values taken by the correlation coefficient. The sampling period covers three weeks from $27/10/2008$ to $14/11/2008$ in *Borsa di Milano*.

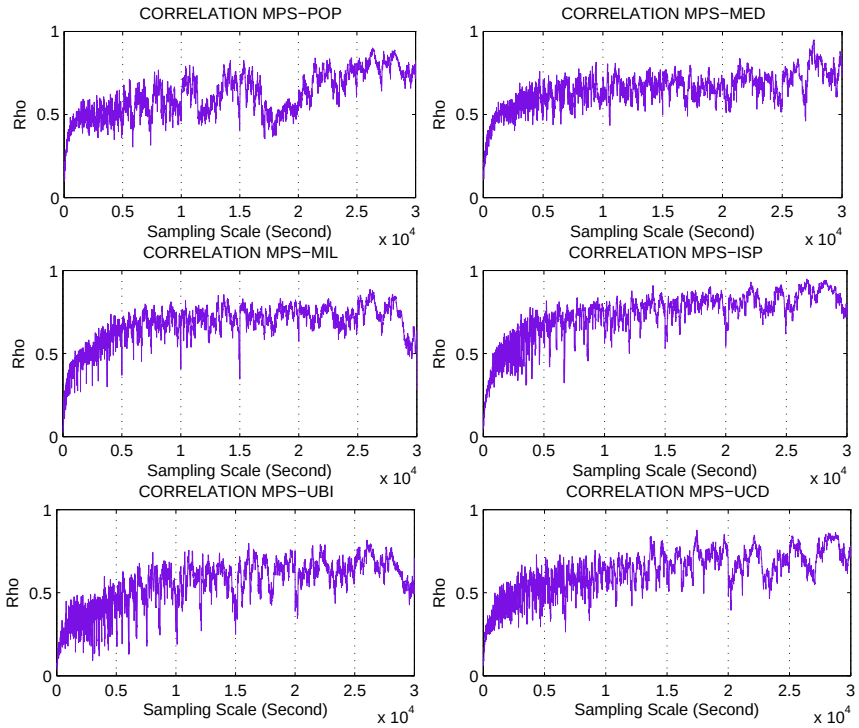

Figure 1.15: The correlation (ρ) of logreturns between MPS Banca and Banco Popolare, Mediobanca, Banca Popolare di Milano, Intesa SanPaolo Banca, UBI Banca and Unicredit Banca, respectively. Previous-tick interpolation is applied. The horizontal axis indicates the sampling interval measured in seconds. The vertical axis indicates the values taken by the correlation coefficient. The sampling period includes 15 business days from $27/10/2008$ to $14/11/2008$ in *Borsa di Milano*.

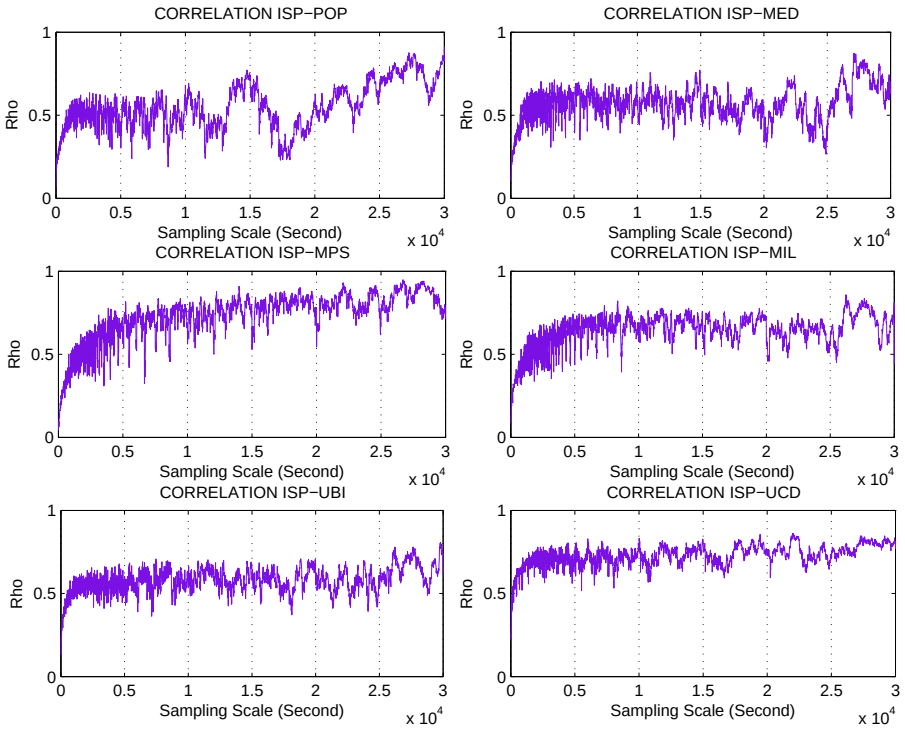

Figure 1.16: The correlation (ρ) of logreturns between Intesa SanPaolo Banca and Banco Popolare, Mediobanca, Banca Popolare di Milano, MPS Banca, UBI Banca and Unicredit Banca, respectively. Previous-tick interpolation is applied. The horizontal axis indicates the sampling interval measured in seconds. The vertical axis indicates the values taken by the correlation coefficient. The sampling data were chosen from *Borsa di Milano* with three-week period from $27/10/2008$ to $14/11/2008$.

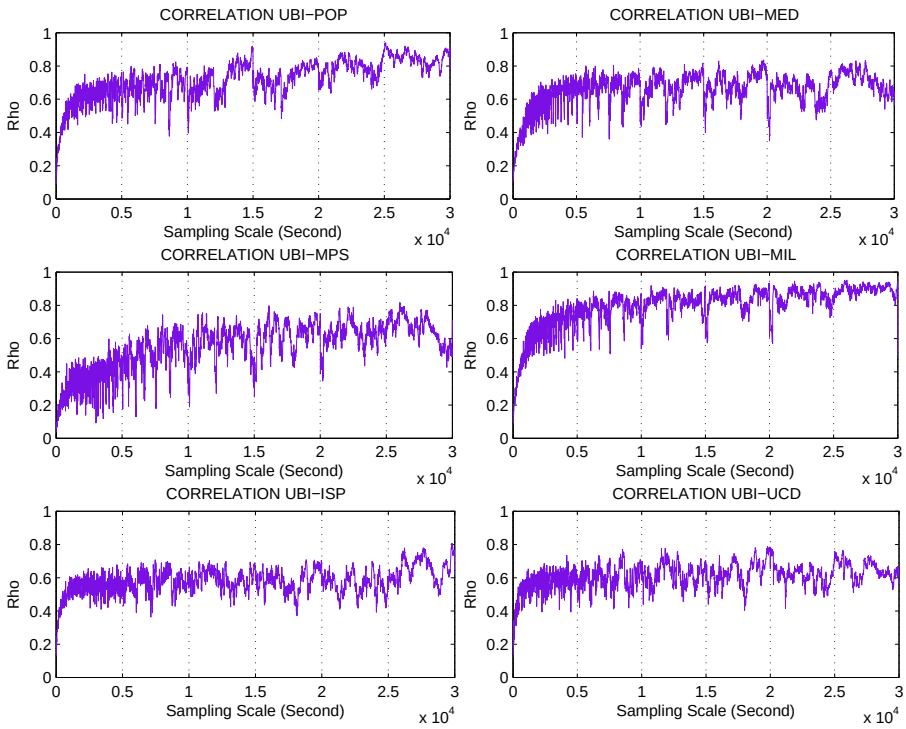

Figure 1.17: The correlation (ρ) of logreturns between UBI Banca and Banco Popolare, Mediobanca, Banca Poplare di Milano, MPS Banca, Intesa SanPaolo Banca and Unicredit Banca, respectively. Previous-tick interpolation is applied. The horizontal axis indicates the sampling interval measured in seconds. The vertical axis indicates the values taken by the correlation coefficient. The sampling data covers three-week time period from $27/10/2008$ to $14/11/2008$ in *Borsa di Milano*.

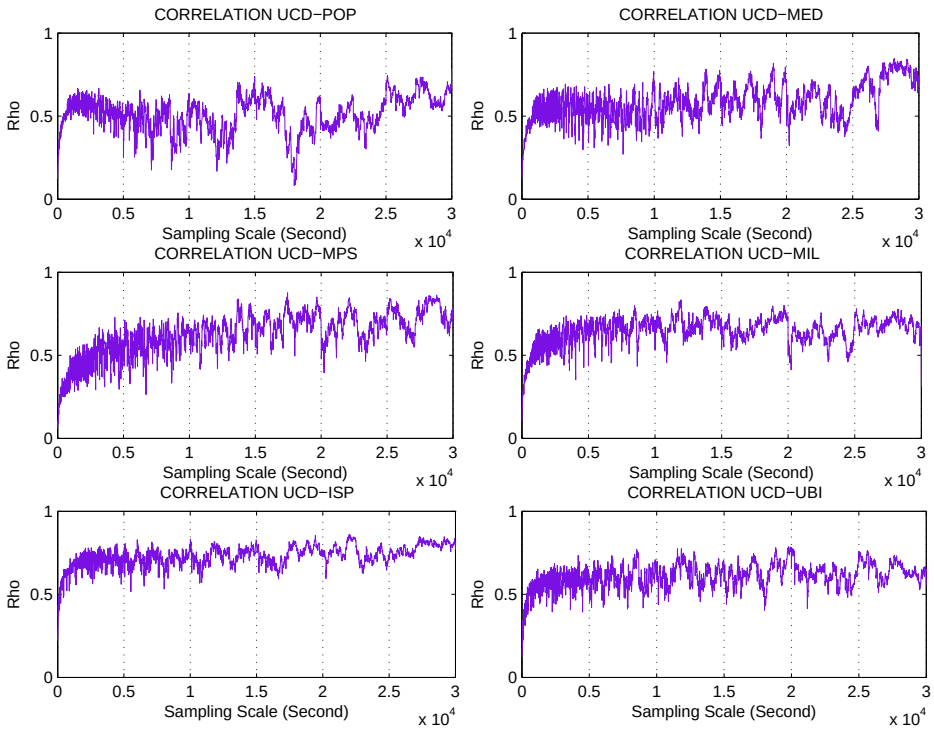

Figure 1.18: The correlation (ρ) of logreturns between Unicredit Banca and Banco Popolare, Mediobanca, Banca Popolare di Milano, MPS Banca, Intesa SanPaolo Banca and UBI Banca, respectively. Previous-tick interpolation is applied. The horizontal axis indicates the sampling interval measured in seconds. The vertical axis indicates the values taken by the correlation coefficient. The sampling data includes 15 business days from $27/10/2008$ to $14/11/2008$ in *Borsa di Milano*.

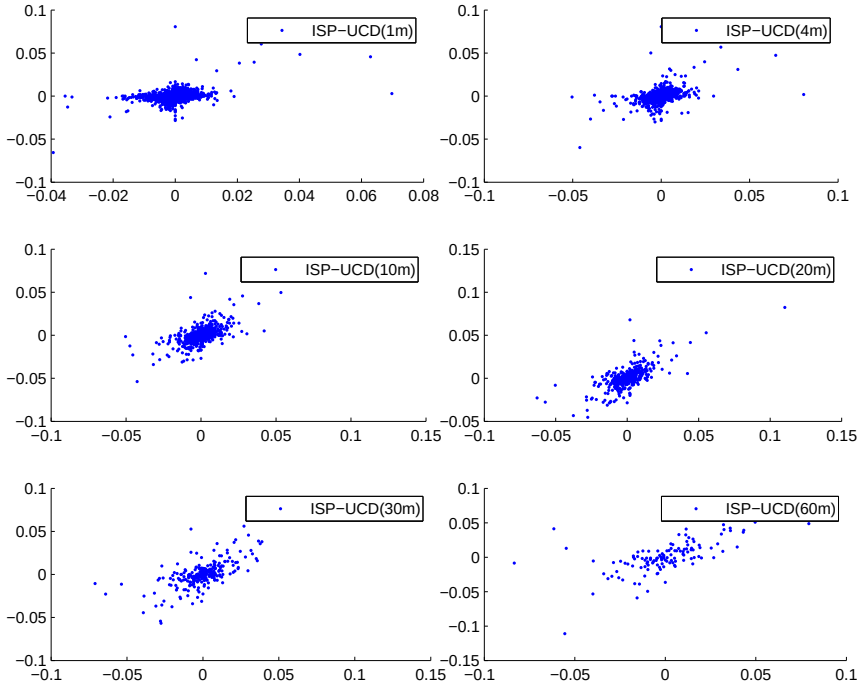

Figure 1.19: The scatterplots of logreturns between Intesa SanPaolo Banca (ISP) and Unicredit banca (UCD) with different sampling intervals. Previous-tick interpolation is applied. The sampling data includes 15 business days from $27/10/2008$ to $14/11/2008$ in *Borsa di Milano.*

1.3.4 Possible causes of Epps effect

In Epps' paper (1979), he claimed that stock correlations decrease dramatically when sampling interval decrease, enter intra-hour. This phenomenon has been observed in several markets, see for example Bonanno et al. (2001) and Zebedee (2001) for stock price, Lundin et al. (1999) and Muthuswamy et al. (2001) for foreign exchange rates.

Considerable work have done to understand the cause of Epps effect, in the literature, two main statistical features of the data may produce this effect: asynchronous trading and lead-lag relationships. Zebedee (2001) argues that the Epps effect is mainly due to the lead-lag relationship, claiming correlation is switching to other nearby time intervals as sampling frequency increases. On the other hand, Lundin et al. (1999) find a significant inverse relation between correlation and activity: the more an asset is traded, less evident is the Epps effect. This implies that asynchronicity is explaining the correlation decrease at higher frequencies. But it worth to remark that non-synchronous trading itself could be a source of spurious lead-lag relations, see Chan (1992, 1993).

In addition, it is sometimes suggested that the Epps effect may depend upon the fact that correlations are lagged, so that when reducing the sampling frequency under time scales comparable to this lag, the correlation measurements turn out to be lower, see B. Tóth et al. (2007) for example. However, Renó (2003) and Lo & Mackinlay (1990) investigated the impact of asynchronous data on covariance measurement, especially, Renó (2003) show that non-synchronicity and lags in correlation have a substantial effect on Epps phenomenon although it is mostly due to asynchronicity.

In fact, it is not difficult to understand the non-synchronicity produce Epps effect. Suppose that the returns to stock A and B are temporally independent but asset A trades more frequently than B. If the news affecting the aggregate stock market and if it arrives near the close of the market on one day, then it is more likely that A's end of day price will reflect this information than B's simply because B may not trade after the news arrives. Of course, asset A will respond to this information eventually but the fact that it responds with a lag induces spurious cross-autocorrelation between the closing price A and B. Then the correlation in daily data is smaller than the one in larger time scale, say, 2-day.

However, Epps effect is dealing with intra-day data. Given the previous macroscopic explaination, it is easier to understand when shrink the sampling interval, the result is the same. Here, we also illustrate a graphic example to show asynchronous data causing Epps effect. In Figure 1.8, suppose that the returns to stock A and B are temporally independent but asset A trades more frequently than B. When news arrive during the day, stock A reveals it sooner and reacts with $\log P_3^A$ as shown in Figure 1.8, whereas stock B responds with $\log P_3^B$ with a delay. After the previous-tick interpolation (the vertical dash line as the time grid and the red arrow as previous-tick interpolation) with sampling interval equals to Δt, $\log P_3^A$ falls in the second subinterval and $\log P_3^B$ falls in the third one. If take this as common situation, especially, when taking non-overlapping sampling interval, then correlation will be presented with a lag. In other words, the concurrent correlation of returns will be zero.

On the other hand, if larger sampling interval is considered, the correlation should be increased. This is because previous-tick interpolation is information aggregation procedure, as described in Section 1.3.1. For instance, if we take $3\Delta t$ as sampling interval (see Figure 1.8), then the news is revealed by both two stocks enter the sampling interval ($3\Delta t$),

this will result the increased correlation. Extremely, when sampling interval large enough such that most of the information can be revealed by both stocks, then correlation reach the stable level even if the sampling interval still increase. This exactly what we observe from the empirical data that correlation increase as sampling interval increase then reach quite stable level, see Section 1.3.3.

It seems previous-tick interpolation have some effect on Epps effect, but asynchronicity is the original reason. There is time lag between the moment the news appear and subsequent trading in any other related stocks. This is caused by market participants who need a certain amount of time to interpret the news and adjust their trading strategies accordingly. In particular, it can take up to two hours for the stock return correlation of companies in the same industry to reach a stable level.

Epps effect, in certain extent, also manifests the information aggregation process in the stock market. Correlations among price changes in common stocks of companies in one industry are found to decrease as increasing the sampling frequency for which the price changes are measured. From our assumption that market traders holding diverging views on the information they observed. As a consequence, in a very small time interval, there is a time lag between the moment the news arrives and subsequent trading in any other related stocks result in smaller correlation; in a certain large time interval, market traders are almost homogeneous in terms of interpretation of the news, the the correlation between different (related) assets increase and become stable. Here stable refers to the correlation coefficient remain in a certain range (very small) even the sampling interval increases. This is because of the characteristic of news, arriving with a suprise and decay as time pass by, but all of the market participants will know it at the end of the day (in sense of certain amount of time).

Chapter 2

Marked point processes

2.1 Introduction

Point processes are used to model intervals between events. An important historical example is given by renewal theory, which could be defined in a narrow sense as the study of the sequence of intervals between successive replacements of a component that is liable to failure and is replaced by a new component every time a failure occurs. Alternatively, point processes can be derived by counting the numbers of events in given intervals. In this latter approach, the machinery of discret distributions plays a central role.

The Poisson process, which takes its name from the Poisson distribution originally studied by Siméon-Denis Poisson, may be taken as a starting point for the study of point processes. The first discussions of the counting problem known to us are by Seidel (1876) and Abbé (1879), who treated the occurrence of thunderstorms and the number of blood cells in haemocytometer squares, respectively, and both apparently independently of Poisson's work. Lyon & Thoma (1881), based on Abbé's data, and Student (1907) gave further discussions of the blood cell problem, the latter paper being famous as one of the earliest applications of the chi-square goodness-of fit text. Shortly afterward, the Poisson process arose in a very important context. Erlang (1909) derived the Poisson distribution for the number of incoming calls to telephone trunking systems by supposing the number of calls in disjoint intervals to be independent and considering the limiting behavior when the interval of observation is divided into an increasing number of equal subintervals. This effectively reproduces the Poisson distribution as the limit of the binomial.

Many new applications were introduced and existing fields of application were extended and deepened. On the queueing theory side, a paper of foundamental importance is Connie Palm's (1943) study of intensity fluctuation of a general theory of the input stream to the detailed analysis of particular telephone trunking system. Three of his themes, in particular, were important for the future of point processes. The first is the systematic description of properties of a renewal process, as a first generalization of the Poisson process as input to a service system. The notion of a regeneration point, a time instant at which the system reverts to a specified state with the property that the future evolution is independent of how the state was reached, has proved exceptionally fruitful in many different applications. In Palm's terminology, the Poisson process is characterized by the property that every instant is a regeneration point, whereas for a general renewal process only those instants at which a new interval is started from regeneration points. hence, he called a

Poisson process a process without aftereffects and a renewal process a process with limited aftereffects. Another important idea was his realization that two types of distribution function: Poisson process and renewal process are important in describing a stationary point process. The distribution of the time to the next event from an arbitrary origin and the distribution of the time to the next event from and arbitrary event of the process. The relations between the two sets of distribution are given by a set of equations now commonly called the Palm-khinchin equations. A third important contribution was his (incomplete) proof of the first limit theorem for point processes, namely, that superpostion of a large number of independent sparse renewal processes leads to a Poisson process in the limit.

All these ideas have led to important further development. In applications, these ideas have been useful not only in queueing theory but also in the study of level-crossing problems. Here the pioneering work was due to Rice (1944) and McFadden (1956, 1958). More rigorous treatments, using some of the Palm-Khinchin theory, were given by Leadbetter and other writers (see e.g., Leadbetter, 1972; the monographs by Cramér & Leadbetter, 1967 and Leadbetter, Lindgren & Rootzen, 1983).

On statistical side, Cox's (1955) paper contained seeds leading to the treatment of many statistical questions concerning data generated by point processes and discussing various models, including the important class of doubly stochastic Poisson processes. A further range of techniques was introduced by Bartlett (1963), who showed how to adapt methods of time series analysis to a point process context and brought together a variety of different models. This work is extended to processes in higher dimensions in a second paper (Bartlett, 1964). Lewis (1964) used similar techniques to discuss the instants of failure of a computer. The subsequent nonograph by Cox & Lewis (1966) was a further important development that, perhaps for the first time, showed clearly the wide range of applications of point processes as well as extending many of the probabilistic and statistical asperts of such processes.

Perhaps the most important development was the rapid growth of interest in point processes in connumications engineering (see e.g., Snyder, 1975). It is a remarkable fact that in nature, for example in nerve systems, the transfer of information is more often effected by pulse signals than by continous signals. This fact seems to be associated with the high signal / noise ratios that it is possible to achieve by thses means,, for the same reason, pulse techniques are becoming increasingly important in communication applications. For such processes, just as for continuous processes, it is meaningful to pose questions concerning the prediction, interpolation, and estimation of signals, and the detection of signals against background noise. Since the signals are intrinsically nonnegative, the distributions cannot be Gaussian, so linear models are not in general appropriate. Thus, the development of a suitable theory for point processes is closely linked to the development of nonlinear techniques in other branches of stochastic process theory. As in applications to processes of diffusion type, martingale methods provide a powerful tool in the discussion of these problems, yielding, for example, structual information about the process and its likelihood function as well as more technical convergence results. Amongst books, developments in this area were surveyed in Liptser & Shiryayev (1974; English translation, 1977, 1978; 2nd ed. 2000), Brémaud (1981) and Jacobsen (1982).

In other new field of applications, such as spatial point processes, or spatial point patterns as they are often called, have become a burgeoning subject in their own right. The many fields of application include envirnmental studies, ecology, geography, astrophisics,

fisheries and forestry, as well as substantially new topics, such as image processing and spatial epidemic theory. Ripley (1981) and Diggle (1983) discuss both models and statistical procedures, while Cressie (1991) gives a broad overview with the emphasis an applications in biology and ecology. Image processing is discussed in the now classical work of Serra (1982). The broad-ranging set of papers in Barndorff-Nielsen et al. (1998) covers many of these applications and associated theory.

Time, space-time and market space-time point processes have continued to recerive considerable attensiton. As well as in the earlier applications to queueing theory, reliability, and electrical engineering, they have found important uses in geophysics, neurophysiology, finance and economics. Snyder & Miller (1991) describe some of the more recent applications in medical fields. Extrem-value ideas in finance are discussed, from a rather different of view than in Leadbetter et al. (1983) and Resnick (1987), in Embrechts et al. (1997).

The growing range of applications has led to an upsurge of interest in inference problems for point process models. Many of the texts referred to above devote a substantial part of their discussion to the practicle implementation of inference procedures. General principles of inference for point processes are treated in the text by Liptser & Shiryayev (1974; English translation, 1977, 1978; 2nd ed. 2000), Brémnd (1981) and in Kutoyants (1980, 1984), Karr (1986, 2nd ed. 1991) and Kutoyants (1998).

Theoretical aspects have also continued to flourish, particularly in the connections with statistical mechanics and stochastic geometry. Including Kingman's (1993) discussion of the Poisson process and Last & Brandt's (1995) exposition of marked point processes. Branching processes in higher-dimensional spaces exhibits many remarkable characteristics, some of which are outlined in Dawson et al. (2000).

2.2 The general theory of processes on the real line

Let us consider a real-valued stochastic processes on the positive time axis, $\{X_t(\omega)\} = \{X(t,\omega)\} = \{X(t)\}$, $t \in (0,\infty) \equiv \mathbb{R}_+$. In general theory, a stochastic process $\{X_t(\omega)\}$ can be regard as an index family of random variables on a common probability space $(\Omega, \mathcal{F}, \mathsf{P})$, with index as a function on the product space $\mathbb{R}_+ \times \Omega$. The stochastic process $X : \mathbb{R}_+ \times \Omega \mapsto \mathcal{B}(\mathbb{R}_+) \otimes \mathcal{F}$ is measurable when this mapping is measurable, that is, for all $A \in \mathcal{B}(\mathbb{R})$,

$$\{(t,\omega) : X(t,\omega) \in A\} \in \mathcal{B}(\mathbb{R}_+) \otimes \mathcal{F}, \tag{2.2.1}$$

where the right-hand side denotes the product σ-algebra of the two σ-algebras there. As a consequence of this measurability theorem, $X(\cdot,\omega) : \mathbb{R}_+ \mapsto \mathbb{R}$ is measurable almost surely, while for measurable functions $h : \mathbb{R} \mapsto \mathbb{R}$,

$$Y(\omega) \equiv \int_{\mathbb{R}_+} h(X_t(\omega))\mathrm{d}t.$$

is a random variable provided the integral exists. A stochastic process on $\mathbb{R}_+$, if defined merely as an index family of random variables on a common probability space, is necessarily measurable if, for example, the trajectories are either continuous almost surely or monotonic and right-continuous almost surely.

For the evolution of a stochastic process, we observe $\{X_s(\omega) : 0 < s \leq t\}$ for some (unknown) ω and finite time interval $(0, t]$. It is then natural to consider the σ-algebra

$$\mathcal{F}_t^{(X)} \equiv \sigma\{X_s(\omega) : 0 < s \leq t\}.$$

generated by all possible such evolutions. Obviously,

$$\mathcal{F}_s^{(X)} \subseteq \mathcal{F}_t^{(X)}$$

for $0 < s < t < \infty$. In general, we may also have some foreknowledge of the process X, and this is denoted by σ-algebra $\mathcal{F}_0$. An expanding family $\zeta = \{\mathcal{F}_t : 0 \leq t < \infty\}$ of sub-σ-algebras of $\mathcal{F}$ is called a filtration or history, and it is natural to assume that those histories incorporate information on the process X. For this reason, if a random variable $X_t(\omega)$ is $\mathcal{F}_t$-measurable (all t), then X is ζ-adapted to the filtration $\mathcal{F}_t$, $0 \leq t < \infty$.

Adopting the special notation

$$\mathcal{H} = \{\mathcal{F}_t^{(X)} : 0 \leq t \leq \infty\} \equiv \{\mathcal{H}_t : 0 \leq t < \infty\},$$

where $\mathcal{F}_o^{(X)} = \liminf_{t>0} \mathcal{F}_t^{(X)} = \{\emptyset, \Omega\}$ and $\mathcal{F}_\infty^{(X)} = \bigcap_{t>0} \mathcal{F}_t^{(X)}$, and call $\mathcal{H}$ the internal, minimal, or natural history of the process X, both of these last two names reflecting the fact that $\mathcal{H}$ is the smallest family of nested σ-algebras to which X is adapted. Any history of the form $\{\mathcal{F}_o \vee \mathcal{H}_t : 0 \leq t \leq \infty\}$ is called an intrinsic history.

Suppose X is measurable and ζ-adapted. An apparently stronger condition to impose on X is that of progressive measurability with respect to ζ, meaning that for every $t \in \mathbb{R}_+$ and any $A \in \mathcal{B}(\mathbb{R})$,

$$\{(s, \omega) : 0 < s \leq t, X_s(\omega) \in A\} \in \mathcal{B}((0, t]) \times \mathcal{F}_t. \tag{2.2.2}$$

Notice that Formular 2.2.2 is more restrictive on X than Formular 2.2.1. Furthermore, Formular 2.2.2 implies Formular 2.2.1, the converse is not quite true. What can be shown, however, is that given any measurable ζ-adapted $\mathbb{R}$-valued process X, we can find an ζ-progressively measurable process Y which is measurable and ζ-adapted satisfy

$$\Pr\{\omega : X_t(\omega) = Y_t(\omega)\} = 1 \quad \text{for all } t$$

2.3 Basic properties of the Poisson process

The archetypal point processes are the Poisson and renewal processes. Before introducing more complex and general theory of point processes, we would like to give an account of some elementary properties of Poisson process in this section.

In the following, we shall first give a few equivalent definitions of the Poisson Process.

Assuming that a counting process $\{N(t), t \geq 0\}$ is a stochastic process which counts the number of events that have occurred up to time t. Obviously, $N(t)$ is non-negative and integer-valued for all $t \geq 0$. Furthermore, $C(t)$ is non-decreasing in t. $N(t) - N(s)$ equals the number of events in the time interval $(s, t]$, $s < t$.

$N(t)$ could, for instance, denote the number of arrivals of customers at a railway station in $(0, t]$, or the number of calls to a telephone call-center during period of $(0, t]$, or

the number of accidents on a particular highway in time interval $(0, t]$. A Poisson process is a stochastic counting process that has a desirable additional properties that the number of events in any disjoint intervals are independent (*'independent increments'*) and that the number of events in any given interval depends only on the length of that interval, and not on its particular position in time (*'stationary increments'*). In the case of the arrivals at the railway station, the staionarity assumption is clearly not fulfilled. There will be many more arrivals between 5 P.M. and 6 P.M. than between, say, 5 A.M. and 6 A.M.. However, restricting oneself to subsequent working days between 5 P.M. and 6 P.M. does allow one to use the stationary increments assumption. Similarly, the independent increments assumption may be violated in some cases, for instance, the number of births of animals in a particular zoo in $(0, t]$, but it will be a reasonably accurate representation of reality in many cases.

Nevertheless, these two properties are extremely important account for mathematical point of view. An extensive discussion of stochastic process with stationary and independent increments can refer to Feller (1966). In what follows, we make an additional assumption such that the counting processes (with stationary and independent increments) are Poisson processes.

Definition 2.3.1 *A Poisson process* $\{N(t), t \geq 0\}$ *is a counting process with the following additional properties:*

(i) *N(0)=0;*

(ii) *The process has stationary and independent increments;*

(iii) $\Pr(N(\Delta t) = 1) = \lambda \Delta t + o(\Delta t)$ *and* $\Pr(N(\Delta t) \geq 2) = o(\Delta t), \Delta t \to 0$*, for some* $\lambda > 0$*.*

Note that the last property of Definition 2.3.2 states the probability of a single event is approximately proportional to the length of that small interval Δt but to have two or more events in such small interval is impossible. Thus Δt is somehow indicating the maximum length (time interval) for occurring one event.

An equivalent definition is

Definition 2.3.2 *A Poisson process* $\{N(t), t \geq 0\}$ *is a counting process with the following additional properties:*

(i) *N(0)=0;*

(ii) *The process has stationary and independent increments.*

(iii) $\Pr(N(t) = k) = e^{-\lambda t} \frac{(\lambda t)^k}{k!}$, $k = 0, 1, \ldots$.

Note that the last property of Defination 2.3.2 states that the number of events in any interval of length t is Poisson distributed with mean λt, where λ is called the rate of the Poisson process.

Now consider the relation, following directly from the last property of Defination 2.3.2, that in case of $k = 0$,

$$\Pr(N(t) = 0) = e^{-\lambda t}. \tag{2.3.1}$$

is the probability of occurring no events in an interval of length t. This may also be interpreted as the probability of the first event occur on the interval with length t. In other words, it gives nothing other than the survivor function for the length of this interval. Equation 2.3.1 therefore shows that the interval under consideration has an exponential distribution.

The following definition is relevant to exponential distribution.

Definition 2.3.3 *A Poisson process $\{N(t), t \geq 0\}$ is a counting process with the following additional properties:*

(i) *N(0)=0;*

(ii) *The only changes in the process are unit jumps upward. The intervals between jumps are independently exponentially distributed random variables with mean $1/\lambda, \lambda > 0$.*

To understand this, consider Equation 2.3.1 in Definition 2.3.2, the probability of the first occurrence (ξ) is greater than t (suvivor function), that is,

$$\Pr(\xi > t) = \Pr(N(t) = 0) = e^{-\lambda t}.$$

Conversely, the probability that an event occurs during t units of time is given by

$$\Pr(\xi \leq t) = 1 - \Pr(\xi > t) = 1 - e^{-\lambda t}. \tag{2.3.2}$$

Differentiating Equation 2.3.2 with respect to t, we can obtain the probability density function of exponential distribution $f(t) = \lambda e^{-\lambda t}$. For the mean of exponential distribution, we have

$$\begin{aligned} \mathrm{E}(\xi) &= \int_0^{+\infty} t\lambda e^{-\lambda t}\mathrm{d}t = -\int_0^{+\infty} t\mathrm{d}e^{-\lambda t} \\ &= \int_0^{+\infty} e^{-\lambda t}\mathrm{d}t = \frac{1}{\lambda}. \end{aligned} \tag{2.3.3}$$

Note $\int_0^{+\infty} \lambda e^{-\lambda t}\mathrm{d}t = 1$ and $\lim_{t\to+\infty} te^{-\lambda t} = 0$ (L'Hospital).

The most peculiar feature of Poisson process is that the mean and variance are equal and that both are proportional to the length of the interval, as shown in Definition 2.3.2. It readily follows that the Probability Generating Function of $N(t)$ is given by $\mathrm{E}(z^{N(t)}) = \sum_{k=0}^{\infty} z^k \Pr(N(t) = k) = \exp\{\lambda(1 - z)t\}$. Differentiation yields $\mathrm{E}(N(t)) = \lambda t$ and $\mathrm{E}(N(t)(N(t) - 1)) = (\lambda t)^2$, and thus $\mathrm{Var}(N(t)) = \lambda t$. Hence, the mean and variance of the number of events occurring in the interval $(0, t]$ are given by

$$\mathrm{E}(N(t)) = \lambda t = \mathrm{Var}(N(t)). \tag{2.3.4}$$

where the constant λ can be interpreted as the mean rate or mean density of events of the process. It also coincides with the intensity of the process as defined in the remainder sections.

Relation between the Poisson process and the exponential distribution

There is an intimate relation between the Poisson process and the exponential distribution, as is already being revealed by Definition 2.3.3. Now we go somewhat deeper into this relation.

From the independence property, we can extend Equation 2.3.1 to the distribution of the time interval between any two consecutive points of the process, given a point in $(-\Delta, 0]$, has the same exponential form, which is independent of Δ, is therefore the limiting form of this conditional distribution as $\Delta \to 0$. When such a unique limiting form exists, it can be identified with the distribution of the time interval between two arbitrary points of the process. Similarly, considering the limiting forms of more complicated joint distributions, the successive intervals are independently distributed as well as having exponential distribution.

On the other side, the particular interval containing the origin is not exponentially distributed. Indeed, its distribution has an Erlang (or gamma) distribution with density $\lambda^2 t e^{-\lambda t}$. This result has been referred to as the 'waiting-time paradox', because it describes the predicament of a passenger arriving at a bus stop when the bus service follows a Poisson pattern. The intuitive explanation is that since the position of the origin (the passengers' arrival) is unrelated to the process governing the buses, it may be treated as effectively uniform over any given time interval.

To verify this result, let t_k, $k = 1, 2, \ldots$, denotes the time from the origin $t_0 = 0$ to the kth point of the process to the right of the origin. Then we have the following identical events

$$\{t_k > t\} = \{N(0, t] < k\}. \tag{2.3.5}$$

Thus, in particular, they have the same probability function, but the probability of the event on the right of Equation 2.3.5 is given directly by Definition 2.3.2, so that we have

$$\Pr\{t_k > t\} = \Pr\{N(0, t] < k\} = \sum_{j=0}^{k-1} \frac{(\lambda t)^j}{j!} e^{\lambda t}. \tag{2.3.6}$$

Differentiating this expression, which gives the survivor function for the time to the kth point, we obtain the corresponding density function

$$f_k(t) = \frac{\lambda^k t^{k-1}}{(k-1)!} e^{\lambda t}, \tag{2.3.7}$$

which is an Erlang density function. Since the time to the kth event can be considered as the sum of the lengths of the k random intervals $(t_0, t_1], (t_1, t_2], \ldots, (t_{k-t}, t_k]$, which are independently and exponentially distributed as shown above, this provides an indirect proof of the result that the sum of k independent exponential random variables has the Erlang distribution.

Relation between the Poisson process and the uniform distribution

In this subsection, we discuss a property of the Poisson process that often is very useful in applications. If exactly one event of a Poisson process has occurred in $(0, t]$, then the time of that occurrence is uniformly distributed on $(0, t)$. The informal explanation is that,

because of the stationary and independent increments, each subinterval of equal length in $(0,t)$ has the same probability to contain that event. The formal derivation is:

$$\begin{aligned}\Pr(t_1 \le s | N(t)=1) &= \frac{\Pr(\text{one event in}(0,s], \text{no event in}(s,t])}{\Pr(N(t)=1)} \\ &= \frac{\Pr(N(s)=1)\Pr(N(t-s)=0)}{\Pr(N(t)=1)} \\ &= \frac{[e^{-\lambda s}\lambda s][e^{-\lambda(t-s)}]}{e^{-\lambda t}\lambda t} \\ &= \frac{s}{t}, \qquad\qquad 0 \le s \le t.\end{aligned}$$

More generally, the following can be proved for a Poisson process: If $N(t)=n$, then the events times $t_1, t_2, \cdots, t_n$ are distributed like the order statistics of n independent random variables that are uniformly distributed on $(0,t)$. The property that a Poisson arrival 'is just as likely to occur in any interval' has proved to be extremely useful in, for instance, queueing theory. In queueing terms this property states that an outside observer, arriving to a queue according to a Poisson process, sees the system as if it were in steady state, that is, the number of customers seen by an arriving customer has the same distribution as the steady-state number of customers.

Relation between the Poisson process and the binomial ditribution

Another important result is worth stressing is that the conditional distributions for the Poisson process are corresponding to a binomial distribution for the number in the subinterval $(0,t]$, given the number in the larger interval $(0,T]$

$$\begin{aligned}\Pr\{N(0,t]=k | N(0,T]=N\} &= \frac{\Pr\{N(0,t]=k, N(t,T]=N-k\}}{\Pr\{N(0,T]=N\}} \\ &= \binom{N}{k}(p_{t,T})^k(1-p_{t,T})^{N-k}, \qquad (2.3.8)\end{aligned}$$

where $p_{t,T}=t/T$, representing a binomial distribution. This property states that given that N events occured in $(0,T]$, the probability that k of them occured in $(0,t]$ is given by the binomial distribution with parameters N and 'success' probability $p_{t,T}=t/T$.

Finally, one simple but important extension to Poisson process with time-varying rate $\lambda(t)$, known as nonhomogeneous or inhomogeneous Poisson process. The process can be defined exactly as above three definitions, with the quantities $\lambda(t)=\int_0^t \lambda \mathrm{d}s$. Thus, the joint distributions are still Poisson, and the independence property still holds.

2.3.1 The general Poisson process

To release the above stationary Poisson process, we obtain more general Poisson process. We suppose that $N(A)$, the number of points in the set A, is defined and finite for every bounded set A in the Borel σ-field $\mathcal{B}(\mathcal{X}) \equiv \mathcal{B}_{\mathcal{X}}$ generated by the open sphere of $\mathcal{X}$, where $\mathcal{X}$ denotes a complete separable metric space $\mathcal{X}$. The Poisson process can then be defined

by assuming that there exists a boundedly finite Borel measure $\Lambda(\cdot)$ such that for every finite family of disjoint bounded Borel sets $\{A_i, i = 1, \ldots, k\}$

$$\Pr(N(A_i) = n_i, i = 1, \ldots, k) = \prod_{i=1}^{k} \frac{[\Lambda(A_i)]^{n_i}}{n_i!} e^{\Lambda(A_i)}. \tag{2.3.9}$$

The measure $\Lambda(\cdot)$ is called the parameter measure of the process. Note that Equation 2.3.9 embraces nontrivial increase in generality because the parameter measure may have both a discrete or atomic component and a continuous singular component.

In this general setting, we first clarify the role of the discrete component of $\Lambda(\cdot)$. Suppose that $\Lambda(\cdot)$ has an atom of mass λ_0 at the point x_0. Since the single-point set $\{x_0\}$ is a Borel set, it follows at once from Equation 2.3.9 that $N\{x_0\} \equiv N(\{x_0\})$ must have a Poisson distribution with parameter λ_0. Any point x_0, therefore, with the property $\Pr(N\{x_0\}) > 0$ is a fixed atom of the process. Thus, we conclude that every atom of $\Lambda(\cdot)$ is a fixed atom of $N(\cdot)$. Conversely, if x_0 is a fixed atom of $N(\cdot)$, then $N\{x_0\}$ must have a Poisson distribution with nonzero parameter λ_0. From this, it follows that x_0 is an atom of $\Lambda(\cdot)$ with mass λ_0.

Note that any point of the process is an atom of its particular realization. For a given point x_0, representing a fixed atom of the process, there must be positive probability of it recurring over a whole family of realizations. Thereby, the fixed atoms relate to the probability structure of the process, not to the structure of individual realizations.

In the Poisson case, the fixed atoms are also the key to the problem of orderliness. Let us see the following theorem first.

Theorem 2.3.1 *The Poisson defined by Equation 2.3.9 is orderly if and only if it has no fixed atoms; equivalently, if and only if the parameter measure has no discrete component.*

When $\mathcal{X}$ is the real line, the distribution function $F_\Lambda(x) \equiv \Lambda(0, x]$ is continuous if and only if Λ has no discrete component, so in this case Λ itself could be called continuous.

Broadly speaking, this general Poisson process refers to inhomogeneous Poisson process, where the measure $\Lambda(\cdot)$ is not constant but a time dependent quantity. An important example of an inhomogeneous Poisson process is the doubly stochastic Poisson process. For this inhomogeneous Poisson process, the measure $\Lambda(\cdot)$ varies stochastically. An introduction of the doubly stochastic Poisson process is given in Section 2.4.3.

2.4 Simple results for stationary point processes on the line

In this section, we give an account of some of the distinctive aspects of stationary point processes on the line. It is intuitively reasonable for some aspects, in fact, it provides a rigorous basis for more complex processes without the burden of too much mathematical detail.

A point process on the line may be taken as modeling the occurrence of some phenomenon at the time events $\{t_i\}$ with i in some suitable index set. For such a process, there are four equivalent descriptions of the sample paths:

(i) counting measures;

(ii) nondecreasing integer-valued step functions;

(iii) sequences of points;

(iv) sequences of intervals.

In describing a point process as a counting measure, it does not matter that the process is on the real line. However, for the three other methods of describing the point process, the order properties of the reals are used in an essential way.

As described in introduction, counting measure is the most common way to figure point processes. To make this notion precise, take any subset A of the real line and let $N(A)$ denote the number of occurrences of the process in the set A, that is, let $N(A)$ be the number of indices i for which t_i lies in A; more formally,

$$N(A) = \{i : t_i \in A\}. \tag{2.4.1}$$

Thus $N(A)$ is nonnegative integer-valued (possible ∞).

For any disjoint sets $A_1, \ldots, A_m$, we have

$$N(\cup_{i=1}^{m} A_i) = \sum_{i=1}^{m} N(A_i), \qquad \text{for mutually disjoint} \quad A_1, \ldots, A_m. \tag{2.4.2}$$

The notion of stationarity of a point process appears to be a simple matter, it means that the distribution of the number of points lying in an interval depends on its length but not its location (as introuced in Section 2.3), that is

$$p_k(x) \equiv \Pr(n(t, t+x] = k) \qquad (x > 0, k = 0, 1, \ldots)$$

depends on the length x but no the location t.

Definition 2.4.1 *A point process is stationary when for every $m = 1, 2, \ldots$ and all bounded Borel subsets $A_1, \ldots, A_m$ of the real line, the joint distribution of*

$$\{N(A_1 + t), \ldots, N(A_m + t)\}$$

does not depende on t ($-\infty < t < \infty$).

In the case when the point process is defined only on the positive half-line, the sets A_i, $i = 1, \ldots, m$, must be Borel subsets of $(0, \infty)$ and we require $t > 0$. Another intuitation is that the intervals $\{y_i\}$ should be stationary as well, where $y_i = t_i - t_{i-1}$, the definition is given in the following.

Definition 2.4.2 *A point process is interval stationary when for every $m = 1, 2, \ldots$ and all integers $i_1, \ldots, i_m$, the joint distribution of $\{y_{i_1+k}, \ldots, y_{i_m+k}\}$ does not depend on k ($k = 0, \pm 1, \ldots$).*

Note that this definition makes no reference to the point t_0 to complete the specification of a sample path. It is most natural to set $t_0 = 0$. Such processes may then be regarded as a generalization of renewal processes in that the intervals between occurrences, instead of being mutually independent and identically distributed, constitute merely a stationary process.

A natural way of measuring the average density of points of a point process is via its mean, or in the case of a stationary point process, its mean density is defined as

$$m = \mathrm{E}(N(0,1]). \tag{2.4.3}$$

For larger interval $(0, x]$, by defining the function

$$M(x) = \mathrm{E}(N(0,x]), \tag{2.4.4}$$

it satisfies Cauchy's functional equation and obtains

$$M(x) = M(1)x = mx * 1cm(0 \leq x < \infty) \tag{2.4.5}$$

In fact, it is a consequence of the additivity properties of $N(\cdot)$ as in Equation 2.4.2. For $x, y \geq 0$,

$$\begin{aligned} M(x+y) &= \mathrm{E}(N(0,x+y]) = \mathrm{E}(N(0,x] + N(x,x+y]) \\ &= \mathrm{E}(N(0,x]) + \mathrm{E}(N(x,x+y]) = \mathrm{E}(N(0,x]) + \mathrm{E}(N(0,y]) \\ &= M(x) + M(y). \end{aligned}$$

There is another natural way to measure the rate of occurrence of points of a stationary point process, due to originally to Khinchin(1955).

Proposition 2.4.1 *For a stationary point process, the limit*

$$\lambda = \lim_{h \to 0} \frac{\Pr\{N(0,h] > 0\}}{h}. \tag{2.4.6}$$

exists, though it may be infinite.

The parameter λ is called the intensity of the point process, for when it is finite, it makes sense to rewrite Equation 2.4.6 as

$$\begin{aligned} \Pr(N(x,x+h] > 0) &= \Pr\{\text{there is at least one point in } (x,x+h]\} \\ &= \lambda h + o(h) \qquad (h \to 0). \end{aligned} \tag{2.4.7}$$

These two measures of the 'rate' of a stationary point process coincide when the point process has the following property.

Definition 2.4.3 *A point process is simple when* $\Pr(N(\{t\}) = 0 \text{ or } 1 \text{ for all } t) = 1$.

From this definition, a point process is called simple if, with probability 1, all its points are distinct. It simply means that all the pairs (t_i, m_i) are distinct, while simplicity of the ground process requires that no two points of N occur simultaneously. Daley (1974) called this sample path property almost sure orderliness to contrast it with the following ananlytic property due to Khinchin (1955).

Definition 2.4.4 *A simple stationary point process is orderly when*

$$\Pr\{N(0,h] > 1\} = o(h) \qquad (h \to 0). \tag{2.4.8}$$

It is worth stressing that stationarity plays no role in the definition of a simple point process, nor does it matter whether the point process is defined on the real line or even a Euclidean space. While orderliness can be defined for point processes that either are nonstationary or are on some space different from the real line.

Definition 2.4.3 and 2.4.4 coincide with a consequence of the following prosposition, given that for stationary point processes with finite intensity.

Proposition 2.4.2 *A simple stationary point process of finite intensity is orderly.*

Proof For any positive integer n, $\mathrm{E}(N(0,1]) = n\mathrm{E}(N(0,n^{-1}])$ by simple stationarity, so

$$\begin{aligned} m &= \mathrm{E}(N(0,1]) = n\sum_{j=1}^{\infty} \Pr(N(0,n^{-1}] \geq j) \\ &\geq n\varphi(n^{-1}) + n\Pr\{N(0,n^{-1}] > 1\}. \end{aligned} \tag{2.4.9}$$

Being simple stationary, Proposition 2.4.1 applies, so $n\varphi(n^{-1}) \to \lambda$ as $n \to \infty$. Moreover, $n\Pr(N(0,n^{-1}] > 1) \to 0$ as $n \to \infty$, which by Equation 2.4.8 in Definition 2.4.4 is the same as orderliness.

2.4.1 Compound Poisson processes

A limitation of the Poisson process is that the jumps are always of unit size. A stochastic process $X(t)$ is called a Compound Poisson Process if it can be represented by

$$X(t) = \sum_{i=0}^{N(t)} Y_i, \qquad t \geq 0, \tag{2.4.10}$$

where $N(t)$ is a Poisson process and $Y_0, Y_1, \ldots, Y_{N(t)}$ are independent, identically distributed random variables that are also independent of $N(t)$. $X(t)$ could, for example, represent the accumulated workload input into a queueing system in $(0,t]$: Customers arrive according to a Poisson process $N(t)$ and the ith customer requires a service need time of length Y_i. Alternatively, the Poisson arrival process might represent the number of insurance claims in $(0,t]$, while the Y_i represent independent claim sizes. $X(t)$ is then the total amount of monetary claims up to time t.

It is easily seen that, if the rate of the Poisson process equals λ and the Y_i have a common Laplace-Stieltjes transform $\vartheta(u) = \mathrm{E}(e^{-sY_i})$, then

$$\mathrm{E}(e^{-sY_i}) = e^{-\lambda(1-\vartheta)t}. \tag{2.4.11}$$

Differentiation then readily yields that $\mathrm{E}(X(t)) = \lambda t\mathrm{E}(Y_1)$ and $\mathrm{Var}(X(t)) = \lambda t\mathrm{E}(Y_1^2)$. Compound Poisson processes are an important subclass of Lévy processes. We refer to J. Bertoin (1996) for a detailed account of the theory of Lévy processes.

2.4.2 Renewal processes

The renewal process has been the subject of much study, both as a model in many fields of application (see e.g. Cox, 1962; Cox & Lewis, 1996 and Cox & Isham, 1980) and as a source of important theoretical problems. As a topic so important and with such far-reaching applications that it can hardly be omitted.

Let $Y_0, Y_1, Y_2, \ldots$ be independent identically distributed nonnegative random variables, and define the partial sums

$$H_0 = 0, \quad H_n = H_{n-1} + Y_n = Y_1 + \cdots + Y_n \qquad (n = 1, 2, \ldots). \tag{2.4.12}$$

For Borel subsets A of $(0, \infty)$, we attempt to define the counting measure of a point process by setting

$$N(A) = \#\{n : H_n \in A\}. \tag{2.4.13}$$

Note that $N(A)$ may not be finite, because subset A may not be bounded. In other words, the right-hand side of Equation 2.4.13 is finite almost surely whenever the subset A is bounded, thus justifying the definition of Equation 2.4.13. The process is so defined is the (ordinary) renewal process.

Orderliness of the process here means $H_{n+1} > H_n$ for $n = 0, 1, \ldots$; that is, $Y_n > 0$ for all $n \geq 0$ with probability 1. Consider the lifetime distribution, the probability that $Y_n > 0$ for $n = 0, 1, \ldots, N-1$ is equal to $[\Pr\{Y_0 > 0\}]^N \to 0$ as $N \to \infty$, given $\{Y_n\}$ are independent and identical distributed, unless $\Pr\{Y_0 > 0\} = 1$. Thus, the process is orderly if and only if $\Pr\{Y_0 > 0\} = 1$; that is, if and only if the lifetime distribution has nonzero mass at the origin.

Taking the expectations of Equation 2.4.13 yields the renewal measure

$$U(A) = \mathrm{E}(\#\{n : H_n \in A, n = 0, 1, \ldots\}) = \mathrm{E}(N(A)), \tag{2.4.14}$$

this equation remains valid even if subset A includes the origin. $U(A)$ is just the first moment or expectation measure of $N(\cdot)$.

By denoting $F(\cdot)$ as the common lifetime distribution and F^{k*} as its $k-$fold convolution which is thus the distribution function for H_k. The quantity most commonly studies is the cumulative function, called the renewal function,

$$U(x) \equiv U([0, x]) = 1 + \sum_{k=1}^{\infty} F^{k*}(x) \qquad (x \geq 0). \tag{2.4.15}$$

Note that $U(x)$ is finite in the sense that $U(x) < \infty$ for all $x > 0$. Taking Laplace-Stieltjes transforms in Equation 2.4.15, we have, if the real part of θ is greater than 0,

$$\chi(\theta) \equiv \int_0^\infty e^{-\theta x} \mathrm{d}U(x) = \sum_{k=0}^{\infty} (\varphi(\theta))^k = \frac{1}{1-\varphi(\theta)}, \tag{2.4.16}$$

where $\varphi(\theta) = \int_0^\infty e^{-\theta x} \mathrm{d}F(x)$. Equivalently,

$$\varphi(\theta) = 1 - \frac{1}{\chi(\theta)}.$$

which shows (using the uniqueness theorem for Laplace-Stieltjes transforms) that U determines F uniquely and hence that there is a one-to-one correspondence between lifetime distributions F and renewal functions U.

Rewrite Equation 2.4.15, we have for $x > 0$

$$U(x) = 1 + \int_0^x U(x-y) \mathrm{d}F(y), \tag{2.4.17}$$

this equation can be viewed as the most important special case of the general renewal equation

$$Z(x) = z(x) + \int_0^x Z(x-y) \mathrm{d}F(y) \qquad (x > 0), \tag{2.4.18}$$

where the solution function Z is generated by the initial function z. If the initial function $z(x)$ is measurable and bounded on finite intervals, one solution to Equation 2.4.18 is given by

$$Z_0(x) = z(x) + \sum_{k=1}^{\infty} \int_0^x z(x-y) \mathrm{d}F^{k*}(y) = \int_0^x z(x-y) \mathrm{d}U(y). \tag{2.4.19}$$

Using the monotonicity of the relation $z \to Z_0$, we have that, if $z > 0$, Equation 2.4.19 is the minimal nonnegative solution to the general renewal equation 2.4.18. Indeed, it is considerably true, for if $z(x)$ is merely measurable and bounded on finite intervals, the difference $D(x)$ between any two solutions of Equation 2.4.18 with the same property satisfies

$$D(x) = \int_0^x D(x-y) \mathrm{d}F^{k*}(y), \qquad \text{for each} \qquad k = 1, 2, \ldots, \tag{2.4.20}$$

thus, $D(x) \equiv 0$ from the fact that $F^{k*}(x) \to 0$ as $k \to \infty$ and the assumed boundedeness of D. It can be summarized as the following lemma.

Lemma 2.4.1 *When $z(x)$ is measurable and bounded on finite intervals, the general renewal equation 2.4.18 has a unique measurable solution that is also bounded on finite intervals, and it is given by 2.4.19. In particular, $U(x)$ is the unique monotonic and finite-valued solution of 2.4.17.*

Example (Exponential intervals) The lack of memory property of the exponential distribution bequeaths on the renewal process that it generates the additional independence properties of the Poisson process. Assuming that

$$F(x) = 1 - e^{\lambda x}, \qquad (\lambda > 0, 0 \leq x < \infty).$$

The renewal function for the corresponding Poisson process is $U(x) = 1 + \lambda x$ according to the integral equation in 2.4.17.

Finally, it is worth noting that the renewal density has uniform marginals, corresponding to the fact that each marginal process is Poisson process.

2.4.3 Doubly stochastic Poisson processes

The doubly stochastic Poisson process, also known as Cox process, is a generalization of the Poisson process (see Section 2.3.1 for more details). It is obtained by randomizing the parameter measure in a Poisson process.

Definition 2.4.5 (Daley and Vere-Jones, 2003) *Let ξ be a random measure on $\mathcal{X}$. A point process N on $\mathcal{X}$ is a doubly stochastic Poisson process directed by ξ when, conditional on ξ, realizations of N are those of a Poisson process $N(\cdot|\xi)$ on $\mathcal{X}$ with parameter measure ξ.*

From this definition, the probabilities in the Poisson process $N(\cdot|\xi)$ are measurable functions of ξ. In the sense that the realization of the number of points in the set A, $N(A|\xi)$, is n, we have

$$\Pr(N(A) = n|\xi) = \frac{(\xi(A))^n e^{-\xi(A)}}{n!},$$

which is a measurable function of ξ.

The finite-dimensional distributions are easily obtained in terms of the distribution of the underlying directing measure ξ. Again, assume the realization of the number of points in the set A, $N(A|\xi)$, is n, we have

$$\Pr(N(A) = k) = \mathrm{E}\left(\frac{(\xi(A))^n}{n!} e^{-\xi(A)}\right) = \int_0^\infty \frac{x^n}{n!} e^{-x} F_A(\mathrm{d}x),$$

where F_A is the distribution function for the random measure $(\xi(A))$.

Regarding the relation between the doubly stochastic Poisson process and the renewal process, Kingman (1964) showed that any stationary doubly stochastic Poisson process is also a stationary renewal process which must be directed by the stationary version of the random measure described. Here stationary refers to translate invariant or homogeneous.

Moreover, the doubly stochastic Poisson process can be viewed as a special case of a cluster point process which will be shown in the next section (Section 2.4.4).

2.4.4 Cluster processes

Cluster processes form are of the most important and widely used models in point process studies, whether applied or theoretical. They are natural models for the locations of objects in the plane or in three-dimensional space, in a remarkable range of contexts: for instance, plants, molecules, human settlements, stars, and earthquake epicentres, etc. Along the tiem axis, they have been used to model photoelectric emissions, volcano eruptions, arrivals and departures at queueing systems, nerve signals and many other phenomena. The cluster mechanism is also a natural way to describe the locations of individuals from consecutive generations of a branching process, an application with unexpectedly rich mathematical structure as well as its obvious practical applications.

Two components are involved in cluster process: one is the locations of clusters and the other is the locations of elements within a cluster. To model the cluster locations, we suppose there is given a process N_c of cluster centres $\{y_i\}$ (often unobserved), whose generic realization consists of the points $\{y_i\} \subset \mathcal{Y}$, more often, $\mathcal{Y} = \mathcal{X}$. On the other hand, to model the cluster elements, we specify a countable family of point processes $N(\cdot|y_i)$ indexed by the cluster centres $\{y_i\}$. The centres y_i act as the germs (= ancesters in the branching process context) for the clusters they generate. In general, it is supposed that there are no special features attatching to the points of a given cluster that would allow them to be distinguished from the points in some other cluster. More formally, we have the following definition.

Definition 2.4.6 *N is a cluster process on the complete separable metric space (c.s.m.s) $\mathcal{X}$, with centre process N_c on the c.s.m.s $\mathcal{Y}$ and the component processes the measurable family of point processes $\{N(\cdot|y) : y \in \mathcal{Y}\}$, when for every bounded $A \in \mathcal{B}_{\mathcal{X}}$,*

$$N(A) = \int_{\mathcal{Y}} N(A|y) N_c(\mathrm{d}y) = \sum_{y_i \in N_c(\cdot)} N(A|y_i) < \infty, \qquad \textit{almost surely}. \tag{2.4.21}$$

The definition of a cluster process requires the superposition of the clusters to be almost surely boundedly finite. However, there is no requirement in general that the individual clusters must themselves be almost surely finite, that is, the condition $N(\mathcal{X}|y) < \infty$ almost surely is not necessary, but it is a natural constraint in many examples. A general cluster random measure can be introduced in the same way by allowing the component processes to be random measures.

For simplicity, we assume that the component processes to be mutually independent. We shall then speak of the component processes as coming from an independent measurable family and thereby defining an independent cluster process. Hence, it is to be understood that multiple independent copies of $N(\cdot|y)$ are taken when $N_c\{y\} > 1$. If $\mathcal{Y} = \mathcal{X}$ (that is, the cluster centre process and the component process are all defined on the same space $\mathcal{X}$ and $\mathcal{X}$ admits translations), then the further constraint that the translated components $N(A - y|y)$ are identically distributed may be added, thus producing a natural candidate for a stationary version of the process.

Conditions for the existence of the resultant point process are not so easily obtained as for the Cox process, even though the superposition of the cluster member processes involves only operations that are clearly measurable. The difficulty revolves around the

finiteness requirement enbodied in Equation 2.4.21. The number of clusters that are potentially able to contribute points to a given bounded set soars as the dimension of the state space increases, imposing delicate constraints that have to be met by any proposed existence theorem. For independent cluster processes, the finiteness condition can be rephrased somewhat more formally as follows.

Lemma 2.4.2 *An independent cluster process exists if and only if, for any bounded set* $A \in \mathcal{B}_{\mathcal{X}}$*,*

$$\int_{\mathcal{Y}} p_A(y) N_c(\mathrm{d}y) = \sum_{y_i \in N_c} p_A(y_i) < \infty, \qquad \textit{almost surely}. \tag{2.4.22}$$

where $p_A(y) = \Pr\{N(A|y) > 0\}$ *for* $y \in \mathcal{Y}$ *and* $A \in \mathcal{B}_{\mathcal{X}}$*, and* Π *is the probability measure for the process of cluster centres.*

Proof The sum of Equation 2.4.22 is required to converge almost surely as part of the definition of a cluster process. The converse, for given N_c, is an application of the second Borel-Cantelli lemma to the sequence of events

$$E_i = \{\text{cluster } i \text{ contributes at least one point to the set } A\}.$$

When the components of the process are stationary, that is, their cluster centre process is stationary and the distribution of the cluster members depends on only on their positions relative to the cluster centre, a simple sufficient condition for the resultant cluster process to exist is that the mean cluster size be finite.

The moments are easier to handle. Taking expectations conditional on the cluster centres yields

$$\mathrm{E}[N(A)|N_c] = \sum_{y_i \in N_c} M_1(A|y_i) = \int_{\mathcal{Y}} M_1(A|y) N_c(\mathrm{d}y),$$

where $M_1(\cdot|y)$ denotes the expectation measure of the cluster member process with centre at y, assuming this latter exists. From the assumption that the cluster member processes form a measurable family, it follows also taht whenever $M_1(A|y)$ exists, it defines a measurable kernel (a measure in A for each y and a measurable function of y for each fixed Borel set $A \in \mathcal{B}_{\mathcal{X}}$). Then we can take expections with respect to the cluster centre process to obtain

$$\mathrm{E}[N(A)] = \int_{\mathcal{Y}} M_1(A|y) M^c(\mathrm{d}y), \tag{2.4.23}$$

finite or infinite, where $M^c(\cdot) = \mathrm{E}[N_c(\cdot)]$ is the expectation measure for the process of cluster centres. From this representation, it is clear that the first moment measure of the resultant process exists if and only if the integral in Equation 2.4.23 is finit for all bounded Borel sets A.

One class of cluster processes, which is frequently used in application and plays a critical role in the theory, is specifying a Poisson process as the cluster centres. This class also refers to a Poisson cluster process. The basic existence and moment results for Poisson cluster process are summarized in the following proposition.

Proposition 2.4.3 *Suppose that the cluster centre process is Poisson with parameter measure $\mu_c(\cdot)$ and that the cluster member processes form an independent measurable family. Then, using the notation above,*

(i) *a necessary and sufficient condition for the existence of the resultant process is the convergence for each bounded $A \in \mathcal{B}_{\mathcal{X}}$ of the integrals*

$$\int_{\mathcal{Y}} \Pr_A(y)\mu_c(\mathrm{d}y); \tag{2.4.24}$$

(ii) *when the process exists, its probability generating function is given by the expression*

$$G[h] = \exp\big(-\int_{\mathcal{F}}(1 - G_m[h|y])\mu_c(\mathrm{d}y)\big). \tag{2.4.25}$$

Proof Since $\mathrm{E}[N_c(\mathrm{d}y)] = \mu_c(\mathrm{d}y)$ for a Poisson cluster process, condition 2.4.24 implies the convergence of 2.4.22 almost surely and hence the existence of the process. If the process exists, then for $h \in \mathcal{V}(\mathcal{X})$, under the independent cluster assumptions, we have

$$\begin{aligned} G[h] = \mathrm{E}(G[h|N_c]) &= \mathrm{E}\big[\exp\big(-\int_{\mathcal{Y}}(-\log G_m[h|y])N_c(\mathrm{d}y)\big)\big] \\ &= G_c[G_m[h|\cdot]], \end{aligned} \tag{2.4.26}$$

where $G_m[h|y]$ for $h \in \mathcal{V}(\mathcal{X})$ is the probability generating function of $N(\cdot|y)$, and

$$G[h|N_c] = \prod_{y_i \in N_c} G_m[h|y_i] = \exp\big[-\int_{\mathcal{Y}}(-\log G_m[h|y])N_c(\mathrm{d}y)\big] \tag{2.4.27}$$

is the conditional probability generating function of N given N_c. Since for $\bar{h} \in \mathcal{V}(\mathcal{X})$, $G_c[\bar{h}] = \exp(-\int[1 - \bar{h}]\mu_c(\mathrm{d}y))$, Equation 2.4.25 is just the appropriate special form of Equation 2.4.26. The measurable family requirements of the family of probability generating function for the cluster centres follow from the initial assumptions for the process. Thus, the probability generating function representation is valid whenever the cluster process exists.

2.5 Marked point processes

In many stochastic process models, point processes arise not as the primary object of study but as a component of a more complex model. Especially, the point processes are the component that carries the information about the locations in time or space of objects that may themselves have a stochastic structure and stochastic dependency relations. In many situations, such components are characterized by both a location and a weight or other distinguishing attribute. In terms of point process theory, many such models can be subsumed under the heading of marked point processes. In this section, we provide an initial study of marked point processes, particularly those with links to the doubly stochastic Poisson processes (or Cox processes) and cluster processes in the preceding sections.

The class of marked point processes is much richer than might at first glance. This is due to the great variety of forms that can be taken by the marks and the variety of dependence relations that can exist between the marks themselves and their locations. When $\mathcal{X} = \mathbb{R}$, for instance, many noticeable results can be obtained by taking the mark at an event t_i to represent some feature from the history of the process up to T_i, see Matthes (1978). Figure 2.1 illustrates a graphical example of marked point processes $(T_i, Z_i)_{i\in\mathbb{N}}$. Notice that T_i can be continuous in the sense that T_i can locate any time in a given time interval.

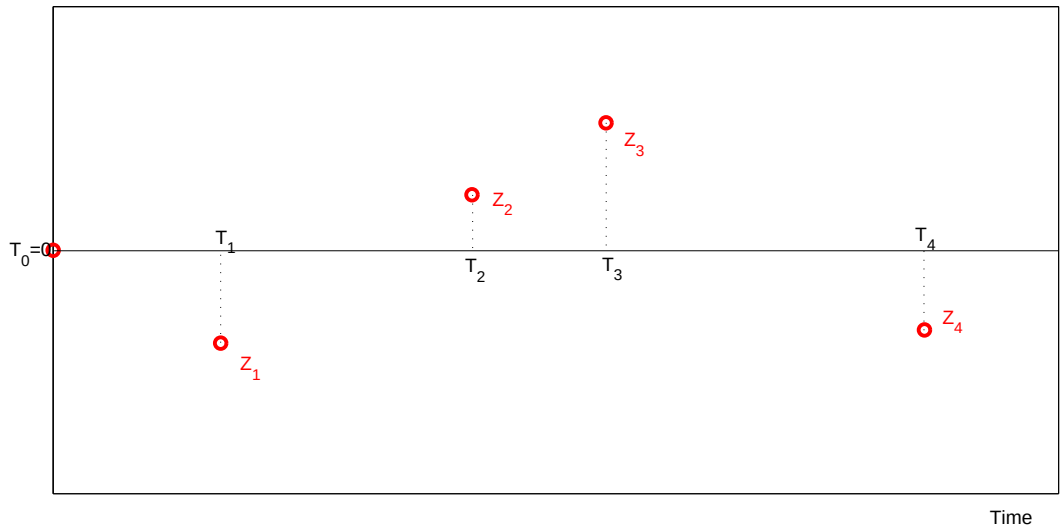

Figure 2.1: This figure shows an example of marked point processes $(T_i, Z_i)_{i\in\mathbb{N}}$. $\{T_i\}$ is the point process and $\{Z_i\}$ is the sequence of associated marks.

In general, marked point processes can be defined in the following two classes of point processes:

Definition 2.5.1 *Let φ be a marked point process,*

1. *φ is simple if the ground process (Given a measurable space $(\mathcal{S}, \mathcal{B})$, for marked point processes, the space $\mathcal{S}$ is a product space $\mathcal{G} \otimes \mathcal{M}$, where $\mathcal{G}$ is called the ground space and $\mathcal{M}$ the mark space), denoted as N_t, is simple.*

2. *φ on $\mathcal{X} = \mathbb{R}^d$ is stationary (homogeneous) if the probability structure of the process is invariant under shifts in $\mathcal{X}$.*

Given a measurable space $(\mathcal{S}, \mathcal{B})$, a point process N is a non-negative, integer-valued random measure defined on $\mathcal{B}$; that is, a measurable mapping from a probability space $(\Omega, \mathcal{F}, \mathsf{P})$ into the space $N_{\mathcal{S}}$ of integer-valued, σ-finite measures on $(\mathcal{S}, \mathcal{B})$. For any $B \in \mathcal{B}$, $N(B)$represents the number of points in B. Following Daley & Vere-Jones (1988), we shall assume that $\mathcal{S}$ is a complete separable metric space, with $\mathcal{B}$ the σ-algebra of Borel sets in $\mathcal{S}$. While in general $N(B)$ is permitted to take the value $+\infty$, it is required to be a.s. finite for compact B.

For marked point processes, the space $\mathcal{S}$ is a product space $\mathcal{G} \otimes \mathcal{M}$, where $\mathcal{G}$ is called the ground space and $\mathcal{M}$ the mark space. For instance, if $\mathcal{S} = \mathcal{G} \otimes \mathcal{M}$ where $\mathcal{G}$ is the real line, the process N also refers to as a temporal marked point process.

The structure of a (univariate) marked point process may be expressed in a variety of ways. If the ground process N_t is not necessarily simple, it can be thought of as a cluster process in which the cluster centres T_i are the distinct locations in $\mathcal{X}$ and the cluster members are all pairs in $\mathcal{X} \times \mathcal{M}$ of the form (T_i, Z_{ij}), where the Z_{ij} are the marks of the points with common locations T_i. By suitably redefining the marks, any marked point process on $\mathcal{X}$ can be represented as a marked point process on $\mathcal{X}$ for which the ground process N_t is simple. Thus, all the pairs (T_i, Z_i) are distinct, while simplicity of the ground process requires that no two points occur simultaneously. For many applications, though not for all, we may therefore assume that the marked point processes we encounter are simple.

The next pair of definitions characterize two important types of independence relating to the mark structure of marked point processes.

Definition 2.5.2 (Independent marks and unpredictable marks) *Let the marked point processes $\varphi = (T_i, Z_i)_{i\in\mathbb{N}}$ on $\mathcal{X} \times \mathcal{M}$ be given.*

1. *φ has independent marks if, given the ground process $N_t = \#\{i : T_i \leq t\}$, the $\{Z_i\}$ are mutually independent random variables such that the distribution of Z_i depends only on the corresponding location t_i.*

2. *For $\mathcal{X} = \mathbb{R}$, φ has unpredictable marks if the distribution of the mark at T_i is independent of locations and marks (T_j, Z_j) for which $T_j < T_i$.*

The most common case of the marked point processes with independent marks occurs when the Z_i are independent and identical distributed. Similarly, the most common case of a process with unpredictable marks occurs when the marks are conditionally independent and identical distributed, given the past of the process (but the marks may influence the future of N_t).

The next proposition outlines the basic structure of processes with independent marks, introducing in particular the mark kernel K at a specified location. K is a $\mathcal{G}_0 \otimes \mathbb{R}^+$-measurable stochastic kernel from $(\Omega \times \mathbb{R}^+)$ to $\mathbb{R}$, that is a mapping such that $K(\cdot, \cdot, B)$ is measurable for all $B \in \mathcal{B}(\mathbb{R})$ that lies between 0 and 1, and $K(\omega, t, \mathbb{R}) = 1$ for all $(\omega, t) \in \Omega \times \mathbb{R}^+$.

Proposition 2.5.1 (Structure of marked point processes with independent marks) *Let φ be a marked point process with independent marks.*

1. *The probability structure of φ is completely defined by the distribution of the ground process N_t and the mark kernel $\{K(M|t) : M \in \mathcal{B}(\mathbb{R}), t \in \mathcal{X}\}$, representing the conditional distribution of the mark, given the location t.*

2. *The moment measure M_k of order k for φ exists if and only if the corresponding moment measure M_k^g exists for the ground process N_t, in which case*

$$M_k(\mathrm{d}T_1 \times \cdots \times \mathrm{d}T_k \times \mathrm{d}m_1 \times \cdots \times \mathrm{d}m_k) = M_k^g(\mathrm{d}T_1 \times \cdots \times \mathrm{d}T_k) \prod_{i=1}^{k} K(\mathrm{d}Z_i|T_i). \tag{2.5.1}$$

3. A random measure for some bounded set $A \in \mathcal{B}_{\mathcal{X}}$ can be defined by

$$\epsilon(A) = \int_{A\times\mathcal{M}} mN(\mathrm{d}t \times \mathrm{d}m) = \sum_{T_i \in A} Z_i. \tag{2.5.2}$$

where $Z \in \mathcal{M}$. If $Z_i = Z$ almost surely for all i, then $\epsilon(A) = ZN_t(A)$.

(For the proof see Daley & Vere-Jones (pag.196–197)).

An important special case arises when the ground process N_t is Poisson. We can refer such a process as a compound Poisson process. In fact, it extends the compound Poisson process introduced in Section 2.4.1, where $\mathcal{M} = \mathbb{Z}_+$. For this (generalized) compound Poisson process, the marks often represent a weight associated with the point, such as a monetary value in financial applications, an energy or seismic moment in seismology, a weight or volume in forestry or geological prospecting, and so on. In such cases, ϵ, as defined in Equation 2.5.2, measures the total value, energy, weight, volume, etc., accumulating within a certain time interval or region.

Lemma 2.5.1 *A compund Poisson process that has mark kernel K, and for which the Poisson ground process N_t has intensity measure $\lambda(\cdot)$, is equivalent to a Poisson process on the product space $\mathcal{X} \times \mathcal{M}$ with intensity measure $\Lambda(\mathrm{d}t \times \mathrm{d}Z) = \lambda(\mathrm{d}t)K(\mathrm{d}Z|t)$.*

(For the proof see Daley & Vere-Jones (pag.199)). The next model, the Hawkes process,

figures widely in applications of point processes to seismology, neurophysiology, epidemiology, and reliability. It is also an important model from the theoretical point of view and will figure repeatedly in later sections of this paper. One reason for its versatility and popularity is that it combines in the one model both a cluster process representation and a simple conditional intensity representation, which is moreover linear. It comes closest to fulfilling, for point processes, the kind of role that the autoregressive model plays for conventional time series. However, the class of processes that can be approximated by Hawkes processes is more restricted than the class of time series models that can be approximated by autoregressive models. In particular, its representation as a cluster process means that the Hawkes process can only be used in situations that are overdispersed relative to the Poisson model.

In introducing the model, Hawkes (1971, 1972) stressed the linear representation aspect from which the term 'self-exciting derives'. Here we derive its cluster process representation, following Hawkes & Oakes (1974), mainly because this approach leads directly to extensions in higher dimensional spaces but also because it simplifies study of the model.

Example (Hawkes process: a self-exciting process) The points $\{x_i\}$ of a Hawkes process are of two types: 'immigrants' without extant parents in the process, and 'offspring' that are produced by existing points. An evolutionary construction of the points is as follows. Immigrants $\{y_j\}$, say, arrive according to a Poisson process at constant rate λ_c, while the offspring arise as elements of a finite Poisson process that is associated with some point already constructed. Any point of the process, located at x', say, has the potential to produce further points whose locations are those of a (finite) Poisson process with intensity measure $\lambda(A-x')$; we assume that $\lambda(\cdot)$ has total mass $\nu = \lambda(\cdot) < 1$ and that all these finite

Poisson processes are mutually independent and, given the point that generates them, identically distributed and independent of the immigrant process as well. Consequently, each immigrant has the potential to produce descendants whose numbers in successive generations constitute a Galton.Watson branching process with Poisson offspring distribution whose mean is ν. Since $\nu < 1$, this branching process is subcritical and therefore of finite total size with mean $1/(1-\nu) < \infty$ if we include the initial immigrant member. Regard the totality of all progeny of a given immigrant point y_j as a cluster; then the totality of all such immigrant points and their clusters constitutes a Hawkes process.

An important task is to find conditions that ensure the existence of a stationary Hawkes process. Since the immigrant process is stationary, a sufficient condition is that the mean cluster size be finite.

The most interesting extensions appear when we drop the assumption of completely independent marks and consider ways in which either the marks can influence the distribution of marks, or both. Using the Hawkes process of Example [Hawkes Process] as below illustrates some of the many possible issues that can arise.

Example (Marked Hawkes process) Marked versions of the Hawkes process of Example (Hawkes Process) are best known from Hawkes (1971, 1972), who considered the multivariate case in detail, with an application in Hawkes & Adamopoulos (1973), though Kerstan (1964) considered them at length. We consider here the case of unpredictable marks.

In extending the Hawkes process of Example (Hawkes Process) to marked point processes $\varphi = (T_i, Z_i)$, we interpret the marks Z_i as the 'type' of an individual in a multitype branching process. Recall that, in the branching process interpretation, points in a Hawkes process are either 'immigrants' without parents or 'offspring' of another point in the process. This model then incorporates the following assumptions:

(i)immigrants arrive according to a compound Poisson process $N(\mathrm{d}y \times \mathrm{d}Z)$ with constant rate μ_c and fixed mark distribution $K(\mathrm{d}Z)$;

(ii) each individual in the process, whether an immigrant or not, has the potential to act as an ancestor and thereby yield first-generation offspring according to an ordinary Poisson process with intensity measure $\mu(\mathrm{d}u|Z) = \varphi(Z)\mu(\mathrm{d}u)$ that depends only on the mark k of the ancestor event and the distance u of the offspring from the ancestor; and

(iii) the marks of the offspring from an independent and identical distribution sequence with the same distribution function K as the immigrants.

The factor $\varphi(Z)$ determines the relative average sizes of families with different marks, while the measure $\mu(\cdot)$ determines how the family members are spread out along the time axis. For a stable process, $\mu(\mathcal{X})$ must be finite, and for the sake of definiteness, we assume that $\mu(\mathcal{X}) = 1$ so that $\varphi(Z)$ becomes the expected number of direct offspring with mark Z.

In principle, the analysis of such a process requires the general theory of multiple type branching processes with a continuous range of types. However, the assumption of independent and identical distributed marks (that is, offspring types) greatly simplifies the analysis. Indeed, the assumptions above imply that the ground process N_t for this marked point process can be described as an ordinary Hawkes process with immigration rate μ_c

and infectivity measure

$$\mu_g(\mathrm{d}u) = \eta\mu(\mathrm{d}u), \qquad \text{where} \quad \eta = \mathrm{E}[\varphi(Z)] = \int_{\mathcal{M}} \varphi(Z)K(\mathrm{d}k) < \infty, \tag{2.5.3}$$

where μ_g denotes the intensity measure for the ground process. If then $\eta < 1$, the total number of offspring is almost surely finite with finite mean $1/(1-\eta)$ so that the ground process is well defined and has a stationary version. Since the overall process may itself be regarded as a Poisson cluster process taking its value in $\mathcal{X} \times \mathcal{M}$, which implies that the overall process has a well-defined stationary version. More formally, we give the following proposition:

Proposition 2.5.2 *Using the notion above, sufficient conditions for the existence of a stationary version of the marked Hawkes process with unpredictable marks are:*

1. *the intensity measure $\mu(\cdot)$ is totally finite (and then taken to be a probability measure);*
2. $\eta = \mathrm{E}[\varphi(Z)] < 1$.

Many variations and extensions of this model are possible. The most simple case is linear form for the conditional intensity, which leads itself to various types of generalization. The mark can be expanded to include a spatial as well as size component, as for the spatial ETAS model described below. On the other hand, the assumption of unpredictable marks can also be weakened in several ways, for example by allowing the distributions of the marks of the offspring to depend on either the mark of the ancestor or the offspring's distance from the ancestor, or both.

If the branching structure is critical rather than subcritical (that is, $\eta = 1$), further types of behavior can occur. For example, if the infectivity function is sufficient long-tailed, Brémaud & Massoulié (2001) provides examples of stationary Hawkes processes without immigration (that is, of a Hawkes process whose cluster overlap at such large distance that the process maintains a stationary regime.).

Example (Ordinary and spatial ETAS models) Ogata (1988) introduced the ETAS (Epidemic Type After-Shock) model to describe earthquake occurrence, following earlier applications of the Hawkes model to this context by Hawkes & Adamopoulos (1973) and Vere-Jones & Ozaki (1982). It corresponds to the special case of the marked Hawkes process where $\mathcal{X} = \mathcal{M} = \mathbb{R}$, the T_i are the occurrence times of the earthquakes and the m_i as their magnitudes, and the following specific assumptions are made:

$$\begin{aligned}
\varphi(m) &= Ae^{\alpha(k_{k0})} I_{\{k>k_0\}}(k), \\
\mu(\mathrm{d}u) &= \frac{D}{(c+u)^{1+p}} I_{\{u>0\}}(u)\mathrm{d}u, \\
K(\mathrm{d}k) &= \beta e^{-\beta(k-k_0)} I_{\{k>k_0\}}(k)\mathrm{d}k.
\end{aligned}$$

These assumptions are dictated largely by seimological considerations: thus, the mark distribution cited above corresponds to the Gutenberg-Richter frequency-magnitude law,

while the power-law form for μ follows the empirical Omori law for aftershock sequences. The free parameters are β, α, c, A and $4p$. $D = pc^p$ is a normalizing constant chosen to ensure $\int_0^\infty \mu(\mathrm{d}u) = 1$.

In this case, sufficient conditions for a stationary process are that

$$p > 0, \quad \beta > \alpha \quad \text{and} \quad \eta = A\beta/(\beta - \alpha) < 1.$$

The last condition in particular is physically somewhat unrealistic since it is well known that the frequency-magnitude distribution cannot retain the pure exponental form indefinitely, but must drop to zero much more quickly for very large magnitudes.

2.5.1 Marked doubly stochastic Poisson processes

This section introduces the marked doubly stochastic Poisson Process (marked DSPP). In order to simplify the exposition, the univariate marked DSPP will be described before extending to the multivariate case. Textbook treatments of the martigale-based, intensity theory of point process are given by Bremaud (1981) and Karr (1991).

The doubly stochastic Poisson process, also known as Cox process, for recognition of its appearance in a seminal paper of Cox (1955), is obtained by randomizing the parameter measure in a Poisson process. It is therefore the generalization of a Poisson process. Under doubly stochastic Poisson process, the intensity function is assumed to be stochastic, which provides flexibility that intensity not only depend on time but also allowing it to be stochastic process. Thus, the doubly stochasitic Poisson process can be viewed as a two step randomisation procedure. A process $\lambda(t)$ is used to generate another process $N(t)$ by acting as its intensity. That is, $N(t)$ is a Poisson process conditional on $\lambda(t)$ which itself is a stochastic process. In particular, if $\lambda(t)$ is deterministic then $N(t)$ is a Poisson process.

Many alternative definitions of a doubly stochastic Poisson process can be given. We will present the one adopted by Last & Brandt (1995).

Definition 2.5.3 (Last and Brandt, 1995) *Given a probability space $(\Omega, \mathcal{F}, \mathsf{P})$ with information structure $\mathcal{F}$. The information structure $\mathcal{F}$ is the filtration, denoted by $\{\mathcal{F}_t\}_{t\in\mathbb{R}^+}$, representing the information available at time t. An φ adapted to the filtration $\{\mathcal{F}_t\}_{t\in\mathbb{R}^+}$ is a doubly stochastic Poisson process if there exists a $\mathcal{F}_0$-measurable random measure ν on $\mathbb{R}^+ \times \mathbb{R}$ such that*

$$\mathsf{P}\Big(\mu\big((s,t]\times A\big) = k \Big| \mathcal{F}_s\Big) = \frac{\Big(\upsilon\big((s,t]\times A\big)\Big)^k}{k!} e^{-\upsilon\left((s,t]\times A\right)}, \tag{2.5.4}$$

almost surely, for every $A \in \mathcal{B}(\mathbb{R})$, where μ denotes the counting measure associated to the MPP φ, that is,

$$\mu\big(\omega, (0,t]\times A\big) = \sum_{i=1}^{N_t} \mathbf{1}_{\{Z_i \in A\}}.$$

Equation 2.5.4 gives us

$$\mathrm{E}(e^{ix\mu_{s,t}}|\mathcal{F}_s) = \exp\{(e^{ix} - 1)\upsilon_{s,t}\}, \tag{2.5.5}$$

where $\mu_{s,t} = \mu\big(\omega, (s,t] \times A\big)$ and $\upsilon_{s,t} = \upsilon\big((s,t] \times A\big)$. Moreover, from Equation 2.5.4 we can easily find that

$$\mathrm{E}(\theta^{\mu_{s,t}}) = \mathrm{E}\big(\exp\{-(1-\theta)(\upsilon_{0,t} - \upsilon_{0,s})\}\big). \tag{2.5.6}$$

Equation 2.5.6 suggests that the problem of finding the distribution of doubly stochastic Poisson processes φ, the marked point processes, is equivalent to the problem of finding the distribution of $\upsilon_{0,t}$. It means that we just have to find the probability generating function of φ to retrieve the moment generating function of $\upsilon_{0,t}$ and vice versa.

In particular, let $\{\mathcal{F}_\sqcup\}_{\sqcup \in \mathcal{R}^+}$ be a complete right continuous filtration, $\mathbf{\Phi} = (T_i, Z_i)_{i \in \mathbf{N}^+}$ be the adapted MPP, then also define a counting process denoted by $\mathbf{N}$, $\mathbf{N}_t = \sum_{i \geq 1} 1(T_i \leq t)$. Assuming that Φ is a marked DSPP with respect to a filtration $\{\mathbf{G}_t\}_{t \in R^+}$, where $\mathbf{G}_t$ has the form $\mathcal{G}_t = \mathcal{G}_0 \vee \mathcal{F}_t$, with compensator $\nu(\mathrm{d}t, \mathrm{d}u) = \lambda_t \mathrm{d}t \mathcal{K}(t, \mathrm{d}u)$, where intensity λ is a $\mathcal{G}_t$-measurable stochastic process and $\mathbf{K}$ is a $\mathcal{G}_0 \otimes \mathcal{B}(\mathbb{R}^+)$-measurable stochastic kernel from $(\omega \times \mathbb{R}^+)$ to $\mathcal{R}$ (i.e. a mapping such that $\mathbf{K}(\cdot, \cdot, B)$ is measurable for all $B \in \mathcal{B}(\mathbb{R})$ and $\mathbf{K}(\omega, t, \mathbb{R}) = 1$ for all $(\omega, t) \in \Omega \times \mathbb{R}^+$; see Last & Brandt 1995, App. A2).

These assumptions imply that DSPPs are point processes in which the number of jumps $N_t - N_s$ in any time interval $(s,t]$ is Poisson distributed, given another positive stochastic process λ, called *intensity*, formally,

$$\Pr\big(N_t - N_s = k \big| \mathcal{G}_s\big) = \frac{\Big(\int_s^t \lambda_u \mathrm{d}u\Big)^k}{k!} \exp\bigg(-\int_s^t \lambda_u \mathrm{d}u\bigg).$$

Moreover, given λ, the number of jumps in disjoint time intervals are independent with means equal to the integral of the intensity in the intervals. For the distribution of the time of the next event, we have

$$\begin{aligned} \Pr(T_{N_t+1} > t \big| \mathcal{F}_s) &= \mathrm{E}\big(\Pr(N_t - N_s = 0 | \mathcal{G}_s) \big| \mathcal{F}_s\big) \\ &= \mathrm{E}\bigg(\exp\bigg(-\int_s^t \lambda_u \mathrm{d}u\bigg) \bigg| \mathcal{F}_s\bigg). \end{aligned}$$

In addition, the conditional distribution of the mark of the next event, given the time of the next event, a realization of the process λ, and the past realization of the MPP, is determined by the stochastic kernel $\mathbf{K}$, that is,

$$\Pr\big(Z_{N_s+1 \in B} \big| \mathcal{G}_s, T_{N_s+1}\big) = \int_B \mathbf{K}(T_{N_s+1}, \mathrm{d}z),$$

for all $B \in \mathcal{B}$.

Proposition 2.5.3 *Considering the moment generating function for the number of counts arriving in a time interval of duration T for a cluster process and show that this information yields results that are mathematically equivalent to those obtained from the stochastic rate approach.*

Suppose that the events from a primary homogeneous Poisson point process $A(t)$ with constant rate ν occur at times $\{\tau_j\}$, indexed by j, where j rangs from 0 to ∞. Each primary event yields a random number of secondary events occurring at random delay times after

the primary event. Let the number of secondary events occurring within the specified counting time window $(0, T]$, denoted by the random function $B(\tau_j)$. Since the secondary events are the result of an inhomogeneous Poisson process, $B(\tau_j)$ is Poisson distributed with

$$\mathrm{E}(B(\tau_j)|\tau_j)_B = \int_0^T X_j e^{k(t-\tau_j)} \mathrm{d}t, \tag{2.5.7}$$

and with the corresponding moment generating function

$$M_{B(\tau_j)}(s)|\tau_j = \exp[(e^s - 1)\mathrm{E}(B(\tau_j)|\tau_j)_B]. \tag{2.5.8}$$

where the subscript of expectation $\mathrm{E}(\cdot)_B$, B, represents expection over the distribution of the subscripted variable B. From Equation 2.5.7, we have

$$\mathrm{E}(B(\tau_j)|\tau_j)_B = X_j(-1/k)[e^{-k(T-\tau_j)} - e^{k\tau_j}]$$

.

The number of events n occurring in the interval $[0, T]$ is the sum of all the secondary events indexed by their respective primary events

$$n = \sum_j B(\tau_j).$$

Similarly, the integrated rate at any time t may be expressed as a sum of the integrated impulse response functions

$$C(t) = \sum_j X_j e^{-k(t-\tau_j)}.$$

The moment generating function for the number of counts n becomes

$$\begin{aligned}
M_n(s) &= \mathrm{E}\big(\exp(-sn)\big)_n \\
&= \mathrm{E}\bigg(\exp(-s\sum_j B(\tau_j))\bigg)_n \\
&= \mathrm{E}\bigg(\prod_j \exp(-sB(\tau_j))\bigg)_n \\
&= \mathrm{E}\bigg(\prod_j \mathrm{E}(\exp(sB(\tau_j))|\tau_j)_B\bigg)_{\{\tau_j\}} \\
&= \mathrm{E}\bigg(\prod_j M_{B(\tau_j)}(s)|\tau_j\bigg)_{\{\tau_j\}} \\
&= \mathrm{E}\bigg(\prod_j \exp((e^{-s}-1)E(B(\tau_j)|\tau_j))\bigg)_{\{\tau_j\}} \\
&= \mathrm{E}\bigg(\exp\bigg((e^{-s}-1)\sum_j\bigg(\int_0^T X_j e^{-k(t-\tau_j)}\mathrm{d}t\bigg)\bigg)\bigg) \\
&= \mathrm{E}\big(\exp((e^{-s}-1)C)\big)_C \\
&= M_C\big(1-e^{-s}\big).
\end{aligned} \tag{2.5.9}$$

Thus the number of counts occurring in an arbitrary interval of length T calculated within the framework of a point process is a Poisson transformation of the integrated rate C.

Models for multivariate intensity functions have been developed in many disciplines. In economics, models have been used a type of multivariate point process providing multiple events in the analysis of unemployment. Lancaster (1990) provides a good review of this model. In finance, Engle & Lunde (1999) propose a bivarite model for transaction data. Bauwens & Hautsch (2006) have used multivarite point processes to study trading of financial market. And in seismology, Ogata, Akaike, and Katsura (1982) have modeled earth quake occurrence as a multivariate point process.

These models, however, are somehow constrained in the sense that one process must be specified as the 'driving process' of the other, it turns out that two processes are not treated symmetrically. And we have to make a choice as to which process drives the other. This is quite objective then. We adopt different approach to avoid such kind of decision by generalizing univariate model. The Univariate DSPP can readily be extended to the multivariate models by introducing a common component that capture the comovement in the intensities.

2.5.2 Multivariate marked DSPP

We consider a K dimensional multivariate point process. Each point process consists of a strictly increasing, stochastic set of arrival times, denoted by $\{t_i^k\}_{i\in\{1,2,\dots\ n_k\}}$, $k = 1,\dots,K$. It is also convenient to introduce counting process associated with k^{th} point process $N^k(t)$, where $N^k(t) = \sum_{i\geq 1} 1(t_i^k \leq t)$. This is simply the number of events that have occurred on the k^{th} by time t. The counting process is useful for indexing the arrival times in a multivariate context since for any time, there will likely be a different number of events will have occurred for each process. Let $t_0^k < t_1^k < t_2^k < \ldots < t_i^k < \cdots < t_{n_k}^k$ denote the arrival times associated with the $k^{\text{th}}(k = 1,\dots,K)$ point process. At time t the most recent arrival time will be denoted by $t_{N_k(t)}^k$. Associated with $N^k(t)$ arrival time of k^{th} point process is a set of characteristics denoted by the vector $Z_{N_k(t)}^k$, which is so called marks. Then the double sequence $\{t_i^k, Z_i^k\}_{k=1,\dots,K}$ is called multivariate marked point processes.

Let $\{t_i\}_{i\in\{1,2,\dots,n\}}$ with $n = \sum_{k=1}^K n_k$ be the sequence of event arrival times of the simple pooled point process that pools the K individual point process and is assumed to be orderly, that is, $0 = t_0 < t_1 < \ldots < t_n$, more details refer to Section 3.4.1 in Chapter refchap:simulation. Correspondingly define $\tilde{N}(t) = \sum_{i\geq 1} 1(t_i \leq t)$ as the counting function of the pooled point process $\{t_i\}_{i\in\{1,2,\dots,n\}}$.

By denoting $\mathcal{F}_t$ as the information set of the pooled process up to t, $N^k(t)$ is assumed to be adapted to $\mathcal{F}_t$. Note that sufficient conditions for the intensity to be $\mathcal{F}_t$-predictable are that the sample paths of the process are left continuous and have right-continuous, that is, càglà path, and that intensity is adapted to $\mathcal{F}_t$. That is, the common information set $\mathcal{F}_t$ upon which each intensity is conditioned is updated continously as new information arrives. Engle & Russel (1998) propose a model for dependent point process based on a particular information set. The ACD model for process k is specified conditional on a sub-sigma field $\mathcal{F}_{t^k} \subseteq \mathcal{F}_t$ consisting of the marginal history, or the internal information of $N^k(t)$ and the associated marks.

The conditional probability that an event of type k occurs in the next instant is given by the conditional intensity fucntion. More formally, the k^{th} marked point process is defined by the $\mathcal{F}_t$-conditional intensity function

$$\lambda^k(t|\mathcal{F}_t) = \lim_{\delta t \to 0} \frac{\Pr(N^k(t+\delta t) - N^k(t) > 0|\mathcal{F}_t)}{\delta t} \tag{2.5.10}$$

where $\lambda^k(t|\mathcal{F}_t)$ is non-negative.

2.6 Univariate point processes in finance

In this section, we review duration-based and intensity-based models of financial point processes. In Section 2.6.1, we outline the class of autoregressive conditonal duration models (ACD), which is by far the most used class in the literature on financial point processes. In Section 2.6.2, we survey different autoregressive conditional intensity models (ACI). Specifying a continuous-time setting on the basis of the conditional intensity function leading to a particularly powerful modeling framework for multivariate processes.

2.6.1 Dynamic duration models

In this section, we discuss a class of dynamic high-frequency duration models, following the introduction of the autoregressive conditional duration (ACD) model by Engel & Russell (1998). In section 2.6.1, we provide a thorough review of the original ACD model of Engel & Russell (1998) and of logarithmic versions of this model, which are by far the most used class in the literature on financial point processes. The logarithmic ACD model has been first introduced in Bauwens & Giot (2000), which is extended by considering several distributions for the error term.

Autoregressive conditional duration models (ACD)

Let x_i be the inter-event duration which is standardised by a seasonality function $s(t_i)$, that is, $x_i = (t_i - t_{i-1})/s(t_i)$. Assuming that the seasonality function $s(t_i)$ is a spline function capturing time-of-day or day-of-week effects. Specially, $s(t_i)$ is specified according to a linear or cubic spline function and is estimated separately in a first step yielding seasonality adjusted duration x_i.

Engle & Russell (1998) model the duration $\{x_i\}_{i=1,2,\ldots,n}$ in terms of a multiplicative error structure and assume:

$$x_i = \psi_i \epsilon_i, \tag{2.6.1}$$

where ψ_i denotes a function of the past durations, and ϵ_i denotes an independent identical distributed random varible. It is assumed that

$$\mathrm{E}(\epsilon_i) = 1, \tag{2.6.2}$$

Then ψ_i corresponds to the conditional duration mean with $\psi_i = \mathrm{E}(x_i|\mathcal{F}_{t_{i-1}})$.

In terms of intensity function, the ACD model can be rewritten as

$$\lambda(t|\mathcal{F}_t) = \lambda_\epsilon\left(\frac{x(t)}{\psi_{\breve{N}_t+1}}\right)\frac{1}{\psi_{\breve{N}_t+1}}, \tag{2.6.3}$$

where λ_ϵ denotes the hazard function of the ACD error term.

By defining the conditional duration ψ_i as a function of the information set $\mathcal{F}_{t_{i-1}}$, denoted as ψ. Using autoregressive moving average structure with order (p, q), provides

$$\psi_i = \psi(\psi_{i-1}, \ldots, \psi_{i-q}, x_{i-1}, \ldots, x_{i-p}). \tag{2.6.4}$$

For simplicity, we consider the case when $p = q = 1$.

There are basically three modeling frameworks of ACD model put forward in the literature by specifying the conditional duration ψ_i. The first model refers to linear ACD model, which is akin to the generalized autoregressive conditional heteroskedasticity (GARCH) model with order 1. But the conditional duration ψ_i is restricted to be positive. To relax this limitation, logarithmic ACD model is proposed by Bauwens & Giot (2000), which refers to the second modeling framework. Finally, a more flexible specification of the conditonal duration in terms of a power transformation leading to the third modeling framework, so-called augmented ACD model.

Linear ACD model. Let the condition duration function ψ_i as follows:

$$\psi_i = \omega + \beta\psi_{i-1} + \alpha x_{i-1}. \tag{2.6.5}$$

Since ψ_i must be positive, set the restrictions $\omega > 0$, $\alpha > 0$ and $\beta > 0$. It is also assumed that $\beta = 0$ if $\alpha = 0$, otherwise β is not identified. Furthermore, the stochastic process defined by 2.6.1, 2.6.2 and 2.6.5 is covariance-stationary if

$$(\alpha + \beta)^2 - \alpha^2\sigma^2 < 0, \tag{2.6.6}$$

where $\sigma^2 = \text{Var}(\epsilon_i) < 0$, and has the following moments and autocorrelations:

(1) $\text{E}(x_i) = \mu_x = \omega/(1 - \alpha - \beta)$,

(2) $\text{Var}(x_i) = \sigma^2 = \mu_x^2\sigma^2\frac{1-\beta^2-2\alpha\beta}{1-(\alpha+\beta)^2-\alpha^2\sigma^2}$,

(3) $\rho_1 = \frac{\alpha(1-\beta^2-\alpha\beta)}{1-\beta^2-2\alpha\beta}$ and $\rho_n = (\alpha + \beta)\rho_{n-1}$ for $n > 2$.

The condition 2.6.6 ensures the existence of the variance. These results are akin to those for the GARCH(1,1) zero-mean process. They can be generalized to ACD(p, q) process with $p, q > 1$.

One limitation of linear ACD model is that it is difficult to allow ψ_i to depend on functions of covariates without violating the positivity restriction. For this reason, Bauwens & Giot(2000) propose a class of logarithmic ACD models.

Logarithmic ACD model. The conditional duration is assumed as

$$\ln \psi_i = \omega + \beta \ln \psi_{i-} + \alpha g(\epsilon_{i-1}), \tag{2.6.7}$$

where $g(\epsilon_{i-1})$ is either $\ln \epsilon_{i-1}$ or ϵ_{i-1}. One advantage of this setting is able to augment ψ_i by function of covariates, further detail refers to Bauwens & Giot (2000). Analogically, the stochastic process defined by 2.6.1, 2.6.2 and 2.6.7 is covariance-stationary if

$$\beta < 1, \mathrm{E}(\epsilon_i \exp\{\alpha g(\epsilon_i)\}) < \infty,\ \mathrm{E}(\exp\{2\alpha g(\epsilon_i)\}) < \infty. \tag{2.6.8}$$

The process has the following moments and autocorrelations:

Augmented ACD model. A more flexible specification of the conditional duration is suggested by Fernandes & Grammig (2006) yielding augmented ACD model. In this case, the conditional duration ψ_i is specified via a power transformation, formally,

$$\psi_i^{\delta_1} = \omega + \beta \psi_{i-1}^{\delta_1} + \alpha \psi_{i-1}^{\delta_1} [|\epsilon_{i-1} - \xi| - \rho(\epsilon_{i-1} - \xi)]^{\delta_2}, \tag{2.6.9}$$

where $\delta_1 > 0$, $\delta_2 > 0$, ξ and ρ are model parameters. The crucial factor characterizing augmented ACD model is the exposition $[|\epsilon_{i-1} - \xi| - \rho(\epsilon_{i-1} - \xi)]^{\delta_2}$, which allows a wide variety of shapes of the curve tracing the impact of ϵ_{i-1} on ψ_i for a given value of ψ_{i-1} and parameters α, β and ω. The parameter ξ and ρ shift and rotation parameters, respectively. If $\xi = \rho = 0$, the linear ACD model is obtained by setting $\delta_1 = \delta_2 = 1$, the logarithmic ACD model by setting δ_1 and δ_2 approach to zero in case that $g(\epsilon_{i-1}) = \ln \epsilon_i$ while letting δ_1 tend to zero and $\delta_2 = 1$ in case $g(\epsilon_{i-1}) = \epsilon_i$.

The flexibility in ACD model can also be achieved in different ways. Dufour & Engle (2000) propose exponential ACD model providing an even more general augmented ACD model by specifying the term $[|\epsilon_{i-1} - \xi| - \rho(\epsilon_{i-1} - \xi)]^{\delta_2}$. While Zhang et al. (2001) present a threshold ACD model, wherein the ACD equation and the error distribution change according to a threshold variable such as the previous duration. Basically, the threshold ACD model is defined as

$$xi_i = \psi_i \epsilon_i^{(m)}, \psi_i = \omega^{(m)} + \beta^{(m)} \psi_{i-1} + \alpha^{(m)} x_{i-1}. \tag{2.6.10}$$

where $x_i \in [r_{m-1}, r_m)$, and $0 = r_0 < r_1 < \ldots < r_M = \infty$ are the threshold parameters. The superscript (m) indicates that the distribution or the model parameters can vary with the regime operation at observation i. This model can be viewed as a mixture ACD model. Another flexible model is proposed by Meitz & Tersvirta (2006) by generalizing linear and logarithic ACD models, leading to smooth transition ACD model. Some strict condition is set for this model, such as strict stationary, ergodicity and existence of moments. The idea of the smooth transition ACD model is to let the expected duration depend on its past duration nonlinearly.

ACD and related models have a variety of applications in financial markets. Specially, ACD models can be used to estimate and predict the intra-day volatility of returns from the intensity of price durations, which is particularly well suited for irregularly spaced time series data. As shown by Engle & Russell (1998), a price intensity is closely related to the instantaneous price change (return) volatility. Bauwens & Giot (2003) model the direction of the price change between two consecutive trades yielding a competing risk model, in

which the direction of the price change is triggered by a Bernoulli process. Prigent et al. (2001) use a similar model for option pricing. Russell & Engle (2005) develop an autoregressive conditonal multinomial model to simultaneously model the time between trades and the dynamic evolution of (discrete) price changes.

2.6.2 Dynamic intensity models

In this section, we review the most important types of dynamic intensity models which are applied to model financial point processes. In Section 2.6.2 we present a class of Hawkes processes. In Section 2.6.2, we survey autoregressive intensity models. Marked doubly stochastic Poisson process is discussed in Section 2.5.1.

Hawkes Processes

Hawkes processes figure widely in applications of point processes to seismology, neurophysiology, epidemiology, and reliability. See e.g. Vere-Jones (1970), Vere-Jones & Ozaki (1982), and Ogata (1988) among others. Bowsher (2007) was the first applying Hawkes models to financial point processes.

Hawkes (1971, 1972) proposed the linear representation aspect from which the term 'self-exciting' derives, where the intensity is driven by a weighted function of the time distance to previous points of the process. A general class of univariate Hawkes processes is given by

$$\lambda(t) = \varphi(\mu(t)) + \sum_{t_i<t} \omega(t - t_i)), \tag{2.6.11}$$

where φ denotes a possibly nonlinear function, $\mu(t)$ is a deterministic function of time, and $\omega(s)$ denotes a weight function.

If φ is a positive function, we obtain the class of nonlinear Hawkes processes considered by Brémaud & Massoulié (1996). In this case, $\mu(t)$ and $\omega(t)$ can take negative values since the transformation $\varphi(\cdot)$ preserves the non-negativity of the process. Such a specification is useful whenever the intensity may be negatively affected by the process history or covariates. For instance, in the context of financial duration processes, $\mu(t)$ can be parameterized as a function of covariates. Stability conditions for nonlinear Hawkes processes are derived by Brémaud & Massoulié (1996). For the special case where $\varphi(\cdot)$ is a linear function, we obtain the class of linear Hawkes processes originally considered by Hawkes (1971). They are analytically and computationally more tractable than their nonlinear counterparts, however, they require $\mu(t) > 0$ and $\omega(t) > 0$ in order to ensure non-negativity.

As pointed out by Hawkes & Oakes (1974), linear self-exciting processes can be viewed as clusters of Poisson processes. The points $\{t_i\}$ of a Hawkes process are of two types: 'immigrants' without extant parents in the process, and 'offspring' that are produced by existing points. The immigrants arrive according to a Poisson process and define the centers of so-called Poisson clusters (see Section 2.4.4). While the offspring arise as elements of a finite Poisson process that is associated with some point already constructed. If we condition on the arrival time, t_i, say, of an immigrant, then independently of the previous history, t_i is the center of a Poisson process, $\iota(t)$, of offspring on (t_i, ∞) with intensity function $\lambda_i(t) = \lambda(t - t_i)$, where λ is a non-negative function. The process $\iota(t_i)$ defines

the first generation offspring process with respect to t_i. Furthermore, if we condition on the process $\iota(t_i)$, then each of the events in $\iota(t_i)$, say, t_j, generates a Poisson process with intensity measure $\lambda_j(t) = \lambda(t-t_j)$. These independent Poisson processes build the second generation of offspring with respect to t_i. Similarly, further generations arise. The set of all offspring points arising from one immigrant are called a Poisson cluster, then the totality of all such immigrant points and their clusters constitutes a Hawkes process.

The immigrants and offsprings can be referred to as 'main shocks' and 'after shocks' respectively. This admits an interesting interpretation which is useful not only in seismology but also in high-frequency finance. Bowsher (2007), Hautsch (2004) and Large (2007) illustrate that Hawkes processes capture the dynamics of financial point processes remarkably well. This indicates that the cluster structure implied by the self-exciting nature of Hawkes processes seem to be a reasonable description of the timing structure of events on financial markets.

As suggested by Hawkes (1971), $\omega(t)$ in Equation 2.6.11 can be parameterized as

$$\omega(t) = \sum_{j=1}^{P} \alpha_j e^{-\beta_j t}, \tag{2.6.12}$$

where $\alpha_j \geq 0$, $\beta_j > 0$ for $j = 1, \ldots, P$ are model parameters, and P denotes the order of the process and is selected exogenously. The parameters α_j are scale parameters, whereas β_j drive the strength of the time decay. For $P > 1$, the intensity is driven by the superposition of different exponentially decaying weighted sums of the backward constraint $\beta_1 > \ldots > \beta_P$.

To ensure the existence of a stationary Hawkes process, it requires $0 < \int_0^\infty \omega(s)ds < 1$ and this is ensured only for $\sum_{j=1}^{P} \alpha_j/\beta_j < 1$ (see Hawkes,1971).

Bowsher (2007) proposes a generalization of the Hawkes model which allows to model point processes that are interrupted by time periods where no activity takes place. In high-frequency financial time series these effects occur because of trading breaks due to nights, weekends and holidays. In order to account for such effects, Bowsher proposes to remove all non-activity periods and to contanenate consecutive activity periods by a spill-over function.

Autoregressive conditonal intensity models (ACI)

Rather than build models based on durations (see Section 2.6.1), Russell (1999) propose autoregressive conditional intensity models (ACI), which directly modeling the instantaneous arrival rates (intensities) leading to a class of dynamic intensity model applied to financial transactions data. Intensity-based models overcome the problem of ACD models that it is difficult to model the expected duration in multivariate context. In the autoregressive conditional intensity models (ACI) proposed by Russell (1999), where intensity function is directly modeled as an autoregressive process which is updated by past realizations of the integrated intensity.

By denoting multivariate intensities $\lambda(t) = (\lambda^1(t), \lambda^2(t), \ldots, \lambda^K(t))'$, Russell (1999) specify each element $\lambda^k(t)$ in terms of a proportional intensity structure given by

$$\lambda^k(t) = \exp(\varphi^k_{\breve{N}(t)+1})\lambda^k_0(t)s^k(t),\ k = 1, 2, \ldots, K, \tag{2.6.13}$$

where $\varphi_{\breve{N}(t)+1}$ allows to capture the dynamic structure, $\lambda : 0^k(t)$ is a baseline intensity component capturing the deterministic evolution of the intensity changes between two consecutive points, and $s^k(t)$ denotes a deterministic function of time t capturing deterministic patterns exhibiting intraday financial data, seasonality, for instance. In their application they specify $s^k(t)$ by a piecewise linear spline. The function $\varphi_{\breve{N}(t)+1}$ is indexed by the left-continuous counting function and is update instantaneously after the new arrival event. Therefore, φ_i is time invariant over $(t_{i-1}, t_i]$, while the evolution of the intensity function between two consecutive arrival times is specified by $\lambda_0^k(t)$ and $s^k(t)$. Finally, the non-negativity of the intensity process is obtained by the exponential form of ψ_i.

Russell (1999) specify ψ_i as

$$\psi_i = \sum_{k=1}^{K} \left(a^k \epsilon_{i-1} + b^k \psi_{i-1} \right) y_{i-1}^k, \tag{2.6.14}$$

where ϵ_i is independent identical distributed unit exponential random variable given by

$$\epsilon_i = \sum_{k=1}^{K} \left(1 - \int_{t^k_{N^k(t_i)-1}}^{t^k_{N^k(t_i)}} \lambda^k(u) \mathrm{d}u \right) y_i^k, \tag{2.6.15}$$

where y_i^k defines an indicator variable that takes on the value one if the i^{th} point of the pooled process is the type k and zero otherwise. a^k is a $(K \times 1)$ parameter vector and b^k is a $(K \times K)$ matrix of persistence parameters. So, if the most recent arrival was of type k then $\epsilon_{\breve{N}_t}$ represents the innovation associated with the event arrival of process k, yielding $\epsilon_{\breve{N}_t} = \epsilon^k_{N^k(t)}$ as shown in equation 2.6.15.

The fundamental principle of the ACI model is that at each event t_i all K process are updated by the realization of the integrated intensity with respect to the most recent process. The innovation term ϵ_i plays an important role in ACI models, Bowsher (2007) specifies the ACI innovation term as $\tilde{\epsilon}_i = 1 - \wedge(t_{i-1}, t_i)$, where $\wedge(t_{i-1}, t_i) = \sum_{k=1}^{K} \wedge^k(t_{i-1}, t_i)$ denotes the integrated intensity of the pooled process computed between the two most recent points. Then, following the argument shown above, $\tilde{\epsilon_i}$ is a independent identical distributed innovation term with zero mean. Since the regime-switching nature of the persistence matrix the derivation of stationary conditions is difficult. However, a sufficient (but not necessary) condition is that the eigenvalues of the matrix b^k for all $k = 1, 2, \ldots, K$ lie inside of the unit circle.

A point process on $[t_0, \infty)$ is said to evolve with after-effect, where if for any $t > t_0$ the realization of events on $[t_0, \infty)$ does depend on the sequence of events in the interval $[t_0, t)$. A point process that evolves with such after-effects can be described using the conditional intensity function which specifies the instantaneous probability of the event arrival conditional upon filtration of event arrival times. In formular, the conditional intensity is given by:

$$\lambda(t; \mathcal{F}_t) = \lim_{\Delta t \to 0} \frac{\Pr\{\mathbf{N}(t + \Delta t) - \mathbf{N}(t) > 0 | \mathcal{F}_t\}}{\Delta t}. \tag{2.6.16}$$

The conditional intensity function associate with any single waiting time has traditionally been called a hazard function in the econometrics literature.

2.7 Multivariate point processes in finance

Russell (1999) extends the work of Engle & Russell (1998) and propose a new econometric model for multivariate transactions data. The data are treated as a bivariate marked point process with dependent arrival rates. In this multivariate context it is difficult to model the expected duration which is the foundamental element of the ACD model. The problem is that it is rather difficult to model the distribution of a duration when new information can arrive within a duration. Thus, Russell (1999) build models directly on the instantaneous arrival rates, or intensities, to assess how quickly information in the transaction process impacts the price through quote adjustments.

Basically, Russell (1999) proposed a model for the instantaneous arrival rate given the multivariate filtration of arrival times and associated marks. The proposed model also allows for marks, or characteristics, associated with the arrival times, to influence futher arrival rates.

Considering a K dimensional multivariate point process, denoted by $\{t_i^k\}_{i\in\{1,\ldots,n^k\}}$, $k = 1,\ldots,K$. Each point process consists of a strictly increasing, stochastic set of arrival times, $t_0^k < t_1^k < \cdots < t_{n^k}^k$. By denoting $N^k(t)$ as a counting process associated with the kth point process, where $N^k(t) \equiv \sum_{i\geq 1} 1_{t_i^k \leq t}$. This is simply the number of events that have occured on the kth process by time t. The counting process $N^k(t)$ is useful for indexing arrival times in a multivariate point process since for any time t, there will likely be a different number of events will have occured for each process.

At time t the most recent arrival time will be denoted by $t_{N_k(t)}^k$. Associated with the $N^k(t)$ arrival time of the kth point process is a set of characteristics denoted by the vector $Z_{N_k(t)}^k$. $Z_{N_k(t)}^k$ is called a vector of marks with dimension L_k. The process $(N^k(t), Z_{N_k(t)}^k)$ is called a marked point process.

Furthermore, let $\{t_i\}_{i\in\{1,2,\ldots,n\}}$ with $n = \sum_{k=1}^K n^k$, be the sequence of event arrival times of the simple pooled or superposed point process, that pools the K individual point processes. Then assume that the pooled point process $\{t_i\}_{i\in\{1,2,\ldots,n\}}$ is orderly in the sense that $0 < t_1 < t_2 < \ldots < t_n$. Note that if the pooled process is orderly then the marginal process must also be orderly.

Let $\mathcal{F}_t$ denote the filtration of the pooled process $\{t_i\}_{i\in\{1,2,\ldots,n\}}$, $N^k(t)$ is associated to be adapted to $\mathcal{F}_t$. Engle & Russell (1998) propose a model for dependent point processes based on a special filtration, in particular, the filtration $\mathcal{F}_t$ is unchanged between arrival times. The ACD model for process k is specified conditional on a sub-sigma field $\mathcal{F}_t^k \subseteq \mathcal{F}_t$ consisting of the marginal history, or the internal filtration of $N^k(t)$ and the associated marks. That is, the information set consists of past arrival times and past marks of the marginal series $\{t_i^k\}_{i\in\{1,2,\ldots,n^k\}}$, so that new information was assumed to arrive only at past arrival times of events of type k.

For the univariate case, see Engle & Russell (1998), it was natural to consider parameterizations for the waiting time or duarion until the next event conditional on information available at the start of the waiting time, given the filtration is unchanged between arrival times. However, for the multivariate point process, the filtration $\mathcal{F}_t$ it is very difficult to parameterize the model in terms of the conditional distribution of waiting times. In particular, a full characterization would require specification of the joint density of the distribution of the next arrival time, and the complete path of information set $\mathcal{F}_t$ (or covariate) over the waiting time. A more natural way that does not require joint modeling of the complete path

of the information set or covariate $\mathcal{F}_t$ is obtained by directly specifying a model for the instantaneous probability that an event of type k occurs in the next instant given information available at time t.

The conditional probability that an event of type k occurs in the next instant is given by the conditional intensity function. More formally, the conditional intensity function associated with the kth marked point process is defined as

$$\lambda^k(t|\mathcal{F}_t) = \lim_{\Delta t \to 0} \frac{\Pr(N^k(t+\Delta t) - N^k(t)) > 0|\mathcal{F}_t}{\Delta t}, \tag{2.7.1}$$

where $\lambda^k(t|\mathcal{F}_t)$ is a non-negative intensity function. Equation 2.7.1 is simply the probability per unit time that an event of type k occurs in the next instant.

Models for multivariate intensity functions have been used in a variety of disciplines. For example, Daley (1968), Brown (1970) and others, have used multivariate point processes to study queues. These processes may be bivariate when a model for inputs and outputs are considered. Ogata, Akaike, & Katsura (1982) have modeled earthquake occurrences as a multivariate point process. In economics, competing risks models have been used in the analysis of unemployment spell, striks and recessions. These models can be viewed as a type of multivariate point process provided multiple events are observed for a single agent. More recently, Engle & Lunde (1999) propose a bivarite model for transaction data. In this model, it is assumed that one process is designated as the driving process of the other, thus the two processes are not treated symmetrically.

In the following, we present two multivariate point process models, one is generalised Hawkes model, originated from Hawkes (1971, 1972), in particular, we outline the generalised Hawkes processes interpreted in terms of financial point processes by Bowsher (2006); the other is stochastic conditional intensity processes proposed by Bauwens & Hautsch (2006).

Generalise Hawkes model

The generalised Hawkes model is a class of model for multivariate event data which are specified via the vector conditional intensity (see Equation 2.7.1). Literally, it is an extension of Hawkes processes, which is outlined in Section 2.6.2.

Following Bowsher (2006), the intensity is defined recursively in terms of the levels of the stochastic component of the intensity at the end of the previous trading day. In particular, the process is specifed by the $\mathcal{F}_t$-conditional intensity

$$\lambda(t) = \mu(t) + \sum_{j=1}^{k} \tilde{\lambda}_j(t), \ \forall t \geq 0, \tag{2.7.2}$$

where $\mu(t)$ is a positive, deterministic functions, $\tilde{\lambda}_j(0) = 0$, and

$$\tilde{\lambda}_j(t) = \pi_i \tilde{\lambda}_j(\tau_{d-1}) e^{\rho_j(t-\tau_{d-1})} + \int_{[\tau_{d-1},t)} \alpha_j e^{-\beta_j(t-u)} \mathrm{d}N(u), \tag{2.7.3}$$

for $\tau_{d-1} < t \leq \tau_d$, where τ_{d-1} denotes the end of the $(d-1)^{th}$ trading day, $\alpha_j \geq 0$, $\beta_j > 0$, $\pi_j > 0$ and $\rho_j > 0$. Note that the data is concatenated by excluding the non-trading period,

in particular, the real half-line is partitioned as $(0,\infty) = (0,\tau_1] \cup (\tau_1,\tau_2] \cup \cdots \cup (\tau_{d-1},\tau_d] \cup \cdots$.

The conditional intensity of generalised Hawkes processes is thus the sum of a deterministic component $\mu(t)$, and K stochastic components $(\tilde{\lambda}_j(t))_{j=1}^K$. Equation 2.7.3 expresses each $\tilde{\lambda}_j(t)$ as the sum of the exponentially-damped value of $\pi_j \cdot \tilde{\lambda}_j(\tau_{d-1})$, where $\tilde{\lambda}_j(\tau_{d-1})$ is the level of the j^{th} component at the end of the previous trading day, and the contributions of events occurring prior to time t on day d.

The first term of Equation 2.7.3 is referred to as the j^{th} intensity 'spillover effect' between trading days. By evaluating the second term in Equation 2.7.3 yielding

$$\sum_{i:\tau_{d-1}\leq s_i<t} \Big[\alpha_j \exp\{\beta_j(t-s_i)\}\Big],$$

where the 'response function' $[\alpha_j \exp\{\beta_j(t-s_i)\}]$ gives the j^{th} impact at time t of the event with occurrence time s_i. The term $[\alpha_j \exp\{\beta_j(t-s_i)\}]$ gives the j^{th} has sample paths that are left-continuous, jumping up by an amount α_j in response to the occurrence of an event and then dacaying until the occurrence of the next event. The events are thus self-exciting in the sense that their occurrence increases the intensity for future events, resulting in their 'clustering' which is a noticable feature of market event data.

Stochastic conditional intensity processes

The second multivariate point process model introducing here is proposed by Bauwens & Hautsch (2006). The basic idea of this model is that the intensity function is driven by observation-driven dynamics as well as latent dynamics. The conditional intensity function given the observation process history is thus not deterministic but stochastic and follows a dynamic process.

The $\mathcal{F}_t$-intensity function $\lambda(t|\mathcal{F}_t)$ completely characterizes the evolution of point process in dependence of the history $\mathcal{F}_t$. The key assumption made in Bauwens & Hautsch (2006) is that the conditioning set $\mathcal{F}_t$ consisting of not only observable process history but also unobservable one. $\lambda(t|\mathcal{F}_t)$ is thus depending on observable characteristics as well as on unobservable ones. For the unobservable factor in term of point process, it leads to the class of doubly stochastic Poisson processes. The standard doubly stochastic Poisson process is characterized by the intensity process $\lambda(t|\mathcal{F}_t^*)$, where $\mathcal{F}_t^*$ denotes the history of some unobserved process up to time t. Then define the information set $\mathcal{F}_t$ as $\mathcal{F}_t = \sigma(\mathcal{F}_t^0 \cup \mathcal{F}_t^*)$, consisting of an observable conditioning set $\mathcal{F}_t^0$ and the history $\mathcal{F}_t^*$ of an unobservable factor $\lambda^*(t)$.

The basic stochastic conditional intensity model for k-type intensity process is given by

$$\lambda^*(t|\mathcal{F}_t) \equiv \lim_{\Delta\to 0} \frac{\Pr(N^k(t+\Delta) - N^k(t) > 0|\mathcal{F}_t)}{\Delta} = \lambda^{0,k}(t)\big[\lambda^*_{\breve{N}(t)+1}\big]^{\sigma_k^*} \qquad (2.7.4)$$

where $\lambda^*_{\breve{N}(t)+1} \equiv \lambda^*(t_{\breve{N}(t)+1})$ denotes a common latent component that depends on its past history $\mathcal{F}^*$ and is updated at each point of the pooled process $\{t_i\}_{i\in\{1,\ldots,n\}}$. $\breve{N}(t) \equiv \#\{i : t_i < t\}$ denotes the left-continuous counting functions of the pooled process $\{t_i\}_{i\in\{1,\ldots,n\}}$.

The coefficient σ_k^* is a process-specific scaling parameter driving the latent factor's impact on the k-type intensity component. The process-specific function $\lambda^{0,k} \equiv \lambda^{0,k}(t|\mathcal{F}_t^0)$ denotes a conditionally deterministic idiosyncratic k-type intensity component given the observable history, $\mathcal{F}_t^0$. Note that each component of intensity by Equation 2.7.4, $\lambda^{0,k}(t)$ and $\lambda^*_{\breve{N}(t)+1}$, is positive-valued random variable. Of course, the resulting intensity $\lambda^*(t|\mathcal{F}_t)$ is positive too.

To identify the latent process λ_i^*, a distributional assumption is imposed such that λ_i^* is conditionally independent and identical (i.i.d.) lognormally distributed, that is,

$$\ln \lambda_i^* | \mathcal{F}^*_{t_{i-1}} \sim \text{i.i.d.} \qquad N(m_i^*, 1). \tag{2.7.5}$$

It is assumed that λ_i^* has left-continuous sample paths with right-hand limits. Therefore, it is indexed by the left-continuous counting function, that is, it is updated instantaneously after the occurrence of t_{i-1} and remains constant until t_i (inclusive). For the conditional mean process, it is assumed as an autoregressive model with order one (AR(1)) and is given by

$$\ln \lambda_i^* = a^* \ln \lambda_{i-1}^* + u_i^*,\ u_i^* \sim \text{i.i.d.} \qquad N(0,1). \tag{2.7.6}$$

By defining $\lambda_i^{k*} \equiv \ln \lambda_i^*$ as the (log) latent component's influence on the k-type intensity component. By substituting it into Equation 2.7.6, we obtain

$$\lambda_i^{k*} = a^* \lambda_{i-1}^{k*} + \sigma_k^* u_i^*.$$

Note that σ_k^* can be negative. Hence, the latent component can simultaneously increase one component while decreasing another one. However, because of the symmetry of the assumed distribution of $\ln \lambda_i^*$, the signs of the scaling parameters σ_k^* are not identified. For instance, it is not distinguishable between the cases $\sigma_1^* > 0, \sigma_2^* < 0$ and $\sigma_1^* < 0, \sigma_2^* > 0$. For this reason, it is restricted to impose an identification assumption to the sign of one of the coefficients σ_k^*. Thereby, it is sufficient to identify the sign of the remaining parameters σ_j^*, for $j \neq k$.

In Bauwens & Hautsch (2006), the basic model can be extended by specifying two parameters, one is process-specific scaling parameters σ_k^*, the other is autoregressive parameter a^*. First, let the process-specific scaling parameters σ_k^* be time-varying allows conditional heterskedasticity in the latent process. An example of conditional heteroskedasticity could be intradaily seasonality associated with deterministic fluctuations of the overall information and activity flow that could be driven by institutional settings, like the opening of other related markets. A natural specification could be to index σ_k^* itself by the counting function and parameterize it in terms of a linear spline function

$$\sigma_{k,i}^* = \sigma_k^* \left\{ 1 + \sum_{m=1}^{M} v_m^* 1_{\{\tau(t_i) > \bar{\tau}_m\}} [\tau(t_i) - \bar{\tau}_m] \right\}. \tag{2.7.7}$$

where $\tau(t_i)$ denotes the time of day at t_i and $\bar{\tau}_m$, $m = 1, \ldots, M$, denote exogenously given knots dividing the trading day into subperiods and v_m^* the corresponding coefficients of the spline function.

The second method for model extensiton is to specify the autoregressive parameter a^* depending on the time elapsed since the last event. The motivation of this extension is

that the latent factor is observable only at irregularly spaced points in time. In Bauwens & Hautsch (2006), a threshold model is proposed for regime-switching latent dynamics, more formally,

$$\ln \lambda_i^* = a_r^* 1_{\{\bar{x}_{r-1} < x_{\check{N}(t)} \leq \bar{x}_r\}} \ln \lambda_{i-1}^* + u_i^*, \quad r = 1, \ldots, R, \tag{2.7.8}$$

where $x_i = t_i - t_{i-1}$, $\bar{x}_r$ denotes the thresholds (with $\bar{x}_0 = 0$ and $\bar{x}_R =$) which is assumed as exogenous, and a_r^* are the regime-dependent latent autoregressive parameters.

On the other side, the observation-driven component $\lambda^{0,k}(t)$ is specified as an autoregressive conditional intensity (ACI) model, that is,

$$\lambda^{0,k}(t) = \Psi^k_{\check{N}(t)+1} \lambda_0^k(t) h^k(t), \quad k = 1, \ldots, K, \tag{2.7.9}$$

where Ψ_i^k captures the dynamic structure and possible exogenous variable, $\lambda_o^k(t)$ is a baseline intensity component, and $h^k(t)$ is a deterministic function to capture seasonality, for instance.

And the baseline intensity component,$\lambda_0^k(t)$, is parameterized as a product of Burr-type hazard functions, that is,

$$\lambda_0^k(t) = \exp(\omega^k) \prod_{r=1}^{K} \frac{x^r(t)^{p_r^k - 1}}{1 + \kappa_r^k x^r(t)^{p_r^k}}, \quad p_r^k > 0, \kappa_r^k \geq 0, \tag{2.7.10}$$

where $x^r(t) \equiv t - t^r_{\check{N}^r(t)}$ denotes the backward recurrente time associated with the r^{th} process. The parameters p_r^k and κ_r^k determine the shape of $\lambda_0^k(t)$ between two k-type events as a deterministic function of the times elapsed since the most recent events in all K processes. A special case occurs when the kth process depends only on its own backward recurrence time, in which case $p_r^k = 1$ and $\kappa_r^k = 0, \forall r \neq k$.

To ensure the positivity of $\lambda^{0,k}(t)$, the dynamic component Ψ_i^k is specified as an exponential transformation of a stochastic process and of explanatory variables z_i, more formally,

$$\Psi_i^k = \exp(\tilde{\Psi}_i^k + z'_{i-1}\eta^k), \tag{2.7.11}$$

where η^k are process-specific parameters associated with covariates observed at the most recent point. The process $\tilde{\Psi}_i^k$ is a left-continuous dynamic process that is updated instantaneously after the occurrence of t_{i-1} and does not change until t_i. In particular, it is assumed as a vectorial autoregressive moving average process,

$$\tilde{\Psi}_i^k = \sum_{k=1}^{K} (A^k \epsilon_{i-1} + B\tilde{\Psi}_{i-1}^k) y_{i-1}^k, \tag{2.7.12}$$

where ϵ_i is a (scalar) innovation term, $A^r = \{\alpha_j^k\}$ is a $(K \times 1)$ coefficient vector reflecting the impact of ϵ_i on the intensity of the K processes when the previous point $(t_{\check{N}(t)})$ was of type k, and $B = \{\beta_{ij}\}$ denotes a (K) coefficient matrix. Finally, y_i^k is an indicator variable that takes on the value 1 if the ith point of the pooled processes is of type k.

To update the autoregressive process 2.7.12 by new information, as suggested by Russell (1999), it is to specify the innovation variable in terms of the integrated intensity

$$\Lambda^k(t_{i-1}^k, t_i^k) \equiv \int_{t_{i-1}^k}^{t_i^k} \lambda^k(u; \mathcal{F}_u) \mathrm{d}u. \tag{2.7.13}$$

Using the multivariate random time change theorem proven by Brown & Nair (1988), it can be shown that under fairly weak regularity conditions, the processes $\Lambda^k(0, t_i^k), i = 1, \ldots, n^k, k = 1, \ldots, K$, are independent Poisson processes with unit intensity. Consequently, the integrated intensity function $\Lambda^k(t_{i-1}^k, t_i^k)$ corresponds to the increment of a unit Poisson process and thus has standard independent and identical exponential distribution, that is,

$$\Lambda^k(t_{i-1}^k, t_i^k) = \sum_{j=N(t_{i-1}^k)}^{N(t_i^k)-1} \Lambda^k(t_j, t_{j+1}) \sim \text{ i.i.d. Exp(1).} \tag{2.7.14}$$

In order to compute the innnovation term based on the observable history of the process, L. Bauwens & N. Hautsch (2006) propose to spesify it in terms of the integrated observation-driven intensity component, which is given by

$$\Lambda^{0,k}(t_{i-1}^k, t_i^k) \equiv \sum_{j=N(t_{i-1}^k)}^{N(t_i^k)-1} \int_{t_j}^{t_{j+1}} \lambda^{0,k}(u)\mathrm{d}u = \sum_{j=N(t_{i-1}^k)}^{N(t_i^k)-1} (\lambda_j^*)^{-\sigma_k^*} \Lambda^k(t_j, t_{j+1}). \tag{2.7.15}$$

Then $\Lambda^{0,k}(t_{i-1}^k, t_i^k)$ corresponds to the sum of (piecewise) integrated k-type intensities that are observed through the duration spell and are standardized by the corresponding (scale) realizations of the latent component, $(\lambda^*)^{-\sigma_k^*}$, during that spell. A particular assumption is made for the SCI innovation ϵ_i by using $\Lambda^{0,k}(\cdot)$,

$$\epsilon_i = \sum_{k=1}^{K} \big\{ -\gamma - \ln \Lambda^{0,k}\big[t_{N^k(t_i)-1}^k, t_{N^k(t_i)}^k\big] \big\} y_i^k, \tag{2.7.16}$$

where $\gamma = 0.5772$ denotes Euler's constant. Thus, ϵ_i is a function of the integral over the observation-driven intensity component, computed over the time between the two most recent points of the process that has been observed at t_i. Therefore, it is a function of the past intervent waiting time, the past intensity, and the past realizations of the latent component. The major advantage of this specification is that ϵ_i can be computed exclusively based on past observables. This leads to a separation between the observation-driven and the parameter-driven components of the model.

Note that according to the innovation specification (Equation 2.7.16) ϵ_i depends on lagged values of the latent factor λ_i^*. Therefore, λ_i^* influences the intensity process $\lambda^k(t; \mathcal{F}_t)$ in two ways: first, it affects $\lambda^k(t; \mathcal{F}_t)$ contemporaneously (according to Equation 2.7.5); second, by lagged realizations which interact with ϵ_{i-1}. For this reason, the latent factor causes cross-autocorrelations not only between $\lambda^k(t; \mathcal{F}_t)$ but also between the individual observation-driven components $\lambda^{0,k}(t), k = 1, \ldots, K$.

Chapter 3

A modeling framework for bivariate logreturns

3.1 Introduction

Correlations between logreturns of different financial assets play an important role in fields such as risk managment and portfolio allocation. In this Chapter we study, by Monte Carlo simulation, the correlation implied by five modeling frameworks for logreturn processes. The five frameworks are (1) returns generated from independent Wiener process; (2) concurrently correlated Wiener process; (3) bivariate autoregressive with first lag (Bi-AR(1)); (4) Bi-AR(1) by specializing zero cross-correlation; and (5) independent Bi-AR(1). In particular, we study the effect of the sampling interval on the correlation between pairs of logreturns, and compare the simulated correlations with the correlations calculated on the empirical data. Here, the empirical data refers to seven Italian banks from *Borsa di Milano*: Banco Popolare (POP), Mediobanca (MED), Banca Poplare di Milano (MIL), MPS Banca (MPS), Intesa San Paolo Banca (ISP), UBI Banca (UBI) and Unicredit Banca (UCD). The sampling peirod covers 15 business days from October 27, 2008 to Nobember 14, 2008.

For the time series, the standard econometric model based on the unit time period, which is simple but is not an appropriate despcription of ultra-high-frequency data that embed unequally spaced in time. A more intuitive way to model unequally spaced data is bound to marked point processes (MPPs), see Chapter 2 for details. Basically, the marked point process can be seen as sequences of pairs $\varphi = (T_i, Z_i)_{i\in\mathbb{N}}$, where T_i refers to the time of an event and Z_i indicates the associated mark (a graphic example is shown in Figure 2.1 of Chapter 2). This way requires to model both the times of trades as well as the logreturns. In the recent literature, models based on marked point processes have been extensively studied in the field of finance. In this context, we classify these models into two main classes. The first class is univariate models based on the MPP, such as Engle & Russell (1998), Zhang, Russell & Tsay (2000), and Bauwens & Giot (2001), which are based on autoregressive conditional duration models (ACD); whereas Rydberg & Shephard (2000), Frey (2000), Frey & Runggaldier (2001), and Centanni & Minozzo (2006) proposed a univariate model based on doubly stochastic Poisson process. The second class is multivariate models based on the MPP, such as Bowsher (2002), Hall & Hautsch (2004), Hall & Hautsch (2006), Hautsch (2005), and Bauwens & Hautsch (2006), where the model of interest (not only driven by one factor but also others) is a complex structure under some

specializations.

On the other hand, there has been a number of empirical studies that documented the existence of strong commonalities and co-movements in individual trading characteristics as well as across the markets. Based on this fact, we would like to construct a multivariate model to capture co-movements in arrival rates of the UHF data by extending the univariate model proposed by Centanni & Minozzo (2006).

In this chapter, therefore, we assume the transaction times of the UHF follow model based on doubly stochastic Poisson processes (DSPPs), the distribution of the number of events in any time interval is characterized through another positive stochastic process λ called intensity, a structural model in which the intensity function is completely characterized by another MPP (as we shall see Chapter 2). The key assumption of our proposed model for timing is that each intensity is consisting of two components. One is the common dynamic component which jointly and contemporaneously drive the underlying intensity, the other is the stochastic individual component of the underlying intensity which is characterized by its own specific MPP. Notice that both of two components themselves are intensities of the DSPPs.

The reminder of this chapter is organised in the following way. In Section 3.2, we present the intensity model. The simulation procedure of the intensity is illustrated in Section 3.3. Section 3.4 outlines five modeling frameworks for logreturns and discusses statistical properties of each framework. Finally, Section 3.5 provides Monte Carlo simulation results of five modeling frameworks and discuss the implied correlations of logreturns.

3.2 Latent factor model of the intensity

In financial markets, news or information often arrives with a surprise. This leads to a jump in stock prices at random in time. Quite often this effect partially vanishes as time passes by. In particular, Cox et al. (1985) propose a model in terms of stochastic noise effect. The idea behind is that some unexpected event occurs due to various reasons and, as time passes by, further information may arise so that the market comes back to more reasonable levels. Here we pick up this idea and assume that stock prices incorporate shot-noise effects, namely there exists sudden jumps in the stock price but with exponential decay as time passes by.

To explore the behavior of the jump time, stochastic point process is paricularly stressed. Stochastic point process is useful for describing phenomena that occur at random points in time or space. In general, the randomness is exhibited in varying times between events, varying numbers of events in any time interval. The most familiar stochastic point process is perhaps the univariate homogeneous Poisson point process. It is characterized by a constant rate (λ), which is the number of events expected to happen in a unit time interval. Note that unit time interval can be secondly, hourly or daily, it depends on time measurement. One important property of the homogeneous Poisson process is that it is memoryless, namely, having full knowledge of the past and present yields no information about the behavior of the process in very recent future. This is due to the memoryless of exponential distribution of interval between events.

However, financial time series exhibite some persistence pattern, such as autocorrelation, seasonality and data clustering ect. Memoryless of the homogeneous Poisson point

process then become a great drawback for describing financial time series. The alternative stochastic point process is non-homogeneous point process, which do not share the memoryless property. In particular, non-homogeneous point process is directed that the intensity is not constant but a random process. An important example of a non-homogeneous point process is to allow the intensity to be specified by a real-valued, non-negative (stationary) random process λ_t with internal history $\mathcal{F}_t$ leading to the class of double stochastic Poisson process (DSPP), also called Cox process.

Doubly stochastic Poisson process is specified via the conditional intensity. The information set upon which intensity is conditioned is updated continuously as new information arrives, thus allowing any market event (or transaction) to have immediate impact on the intensity as they occur in continuous time. When the intensity is conditioned on the natural filtration of the point process, the approach is mathematically equivalent to the duration-based model but is intuitively more appealing. The duration-based model is proposed by Engle & Russell (1998) by specifying the time between price changes of fixed size, with the name Autoregressive Conditional Duration (ACD) model. However, it is difficult to model multivariate conditional durations, because the information set cannot be updated within a duration spell as a basic assumption in ACD model. For multivariate transaction data, the information set has to be updated continuously to allow time-varying covariates or event arrivals in other point process, for instance. Thus, duration models are typically used in univariate framework.

Therefore, for the multivariate data analysis, we adopt a different approach in which the model is specified via the conditional intensity rather than build models based on durations. We propose a new model for the instantaneous arrival rate (intensity) given the multivariate filtration of arrival times and associated marks. In particular, this modeling is preferential as to account for the asynchronous event arrivals in a multivariate framework. In the econometrics literature, intensity-based modeling has hitherto utilised the Autoregressive Conditional Intensity (ACI) model proposed by Russel (1999) in which he specify a bivariate model of transaction arrival times and limit order submission arrival times and its extensions. Hall & Hautsch (2004) estimated a bivarite ACI model for the arrival of buy and sell trades on a limit order book market. Bauwens & Hautsch (2003) propose an extension of the ACI model which adds a latent, Guassian autoregressive component to the log intensity. The core factor of all of these ACI models is specified in 'event time', with the consequence that understanding the properties of the continuous time conditional intensity process, or equivalently the distribution of the Point process.

Now suppose K intensities are of interest, we assume the intensities are following a factor model. Essentially, the K intensities are breaking up into $K+1$ separate components or factors. One is referring to a common component (factor), the other K components (factors) are associated with the individual specific characters. For simplicity, we consider a linear combination of two components, more formally,

$$\tilde{\lambda}_t^{0k} = \lambda_t^{(k)} + a_k \lambda_t^{(0)}, \qquad k = 1, 2, \ldots, K. \tag{3.2.1}$$

where $\lambda_t^{(k)}$ are the specific component of intensity $\tilde{\lambda}_t^{0k}$ evaluated at time t; $\lambda_t^{(0)}$ is the common component of all K intensities; the coefficient a_k are real parameters which drive the common component $\lambda_t(0)$ impact on the k specific component $\lambda_t^{(k)}$ contemporaneously.

This proposed model is fully parametric and can be seen as the counterpart of the stochastic volatility model (Taylor, 1982) or the stochastic conditional duration model

(Bauwens & Veredas, 2004), which are motivated by the mixture-of-distribution hypothesis. In these models, the process dynamics are only driven by the dynamics of the latent component, whereas our model is parameterized in terms of two components: not only a common latent component but also a individual (latent) one.

For each component $\lambda_t^{(i)}$, we particularly consider a subclass of marked doubly stochastic Poisson process by specifying the underlying intensity as shot noise intensity proposed by Cox & Isham (1980, 1986). Since the common factor intensity model concerns on multivariate process, some notations of the model will be consistent with marked doubly stochastic Poisson process as described in Section 2.5.2 of Chapter 2.

Given a probability space $(\Omega, \mathcal{F}, \mathsf{P})$ and a complete right continuous filtration $\{\mathcal{F}_t\}_{t\geq 0}$. Considering $K+1$ adapted marked point process (MPP) $\Phi^{(i)} = (t_j^i, Z_j^i)_{j\in\mathbb{N}}$, where $\{t^i\}$ are positive and strictly increasing random variables, that is, $t_j^i < t_{j+1}^i$, and $\{Z^i\}$ are $\mathbb{R}$-valued random variables, and $i = 0, 1, \ldots, K$. By denoting $N_t^{(i)}$ as counting process, $N_t^{(i)} = \#\{j : t_j^i \leq t\}$. The counting process is useful for indexing the arrival times in a multivariate context since for any time t, there will likely be a different number of events will have occurred for each process. In particular, we assume the marked point process $\Phi^{(i)}$ is characterized by the number of events in any disjoint interval are independent and Poisson distributed, given intensity $\lambda_t^{(i)}$. The conditional distribution of number of events in $(s, t]$ is given by

$$\Pr\left(N_t^{(i)} - N_s^{(i)} = n^{(i)} \middle| \lambda_t^{(i)}\right) = \frac{1}{n^{(i)}!}\left(\int_s^t \lambda_u^{(i)} \mathrm{d}u\right)^{n^{(i)}} \exp\left(\int_s^t \lambda_u^{(i)} \mathrm{d}u\right) \tag{3.2.2}$$

Furthermore, we specify the intensity $\lambda_t^{(i)}$, at which Equation 3.2.2 conditioned, by following shot noise process (Cox & Isham, 1980). Adopt the idea of Cox & Isham (1980, 1986), the intensity function is completely characterized by another MPP $\Phi^{'(i)} = (\tau_j^{(i)}, X_j^{(i)})_{j\in\mathbb{N}}$, which has a finite number of points in bounded intervals (it is nonexplosive) and for which $\tau_0^{(i)} 0$. Given our factor intensity model, each component ($K+1$ components in our model) is associated with its specific MPP $\Phi^{'(i)}$, so $K+1$ marked point processes should be considered thereafter. For simplicity, we assume that all of $K+1$ MPP are mutually independent, that is, $\Phi^{'(i)}$ is independent of $\Phi^{'(j)}$, for any $i \neq j$, $i, j = 0, 1, \ldots, K$. This implies the each component $\lambda_t^{(i)}$ is independent from another.

A shot noise intensity $\lambda_t^{(i)}$ (Cox & Isham 1980) can be expressed as

$$\lambda_t^{(i)} = \lambda_0^{(i)} e^{(-\kappa_i t)} + \sum_{j=0}^{N_t^{'(i)}} X_j^{(i)} \cdot e^{(-\kappa_i (t - \tau_j^{(i)}))},\ t \geq 0 \tag{3.2.3}$$

where:

$\lambda_0^{(i)}$ the initial value of $\lambda_t^{(i)}$;

κ_i the magnitude of exponential decay, where $\kappa_i > 0$;

$N_t^{'(i)}$ the number of jumps, $N_t^{'(i)} = \#\{j : \tau_j^{(i)} \leq t\}$;

$X_j^{(i)}$ the jump size, where $\mathrm{E}(X_j^{(i)}) < \infty$, $X_0^{(i)} = 0$, and $X_j^{(i)} > 0$ for all $j > 0$;

$\tau_j^{(i)}$ the time at which jump happens, where $0 < \tau_j^{(i)} \leq T$ and $\tau_0^{(i)} = 0$.

Note that $\lambda_t^{(i)}$ is a positive stochastic process, $i = 0, 1, \ldots, K$. For the jump times $\{\tau_j^{(i)}\}$ and jump sizes $\{X_j^{(i)}\}$ of the intensity $\lambda_t^{(i)}$, the shot noise intensity implies that the conditional distribution of $\tau_j^{(i)}$, given $(\tau_m^{(i)}, X_m^{(i)})$, $m = 1, \ldots, j-1$, and the conditional distribution of $X_j^{(i)}$, given $\tau_m^{(i)}$, $m = 1, \ldots, j$, and $X_m^{(i)}$, $m = 1, \ldots, j-1$ has an analytic form. Moreover, $N_t^{'(i)}$ is a Poisson process with constant rate and $X_j^{(i)}$ are independent and identical distributed random variables which are also independent from $\tau_j^{(i)}$.

In this modeling framework, a number of models can be specified under different hypothesis on the frequency and magnitude of these jumps. In this thesis, we will specify our model as follows:

- the initial value $\lambda_0^{(i)}$ of the intensity process $(\lambda_t^{(i)})$ have Gamma distribution with scale parameter ν_i/k_i and rate parameter γ_i (so $\mathrm{E}(\lambda_0^{(i)}) = \nu_i/k_i\gamma_i$);
- the number of news arrivals $N^{'(i)}$ are Poisson processes with constant intensity ν_i;
- the jump size $X_j^{(i)}$ have independent and identical exponential distribution with mean $1/\gamma_i$;
- the news inter arrival time $\tau_j^{(i)} - \tau_{j-1}^{(i)}$ are independently exponentially distributed with mean $1/\nu_i$.

A graphic example is illustrated in Figure 3.1, which is based on the above assumptions. The resulting figure shows that the intensity is entirely move up by the jumps and tail off exponentially until the next jump arrives.

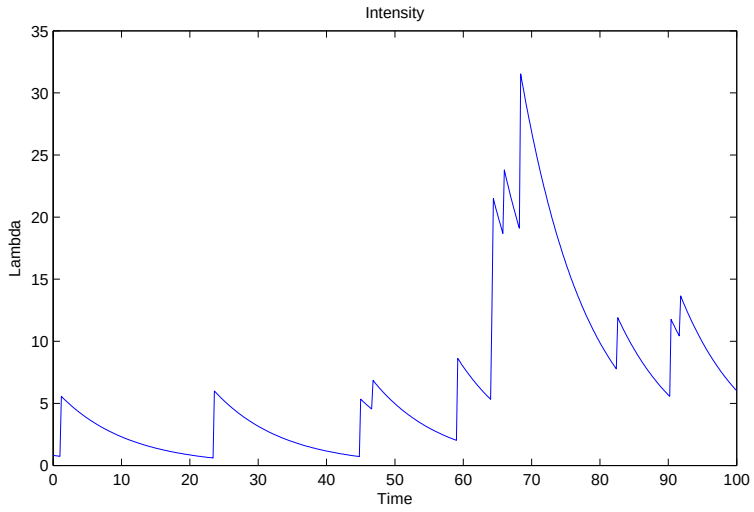

Figure 3.1: A shot noise intensity process with parameters $\nu = 0.1$, $\kappa = 0.1$, $\gamma = 0.2$ (Equation 3.2.3).

In terms of economics, the shot noise intensity can fully capture the characteristics of the impact of information on the financial market. As we know, information often arrives with a surprise which leads to a jump in stock prices at random in time, then this effect partially vanishes as time passes by (news will be not as news any more). The shot noise intensity then can be interpreted as information process, the jump times can be interpreted as news arriving time and the jump sizes can be the importance of the news. When an important news arrives into the market, a big jump moves up the intensity because many transactions are made (corresponding with high frequency) caused by this piece of news. Of course, this news will be revealed eventually and vanishes as time passes by.

If information process follows shot noise intensity, some implications we need to address. Basically, shot noise intensity implies that information arrives with constant rate ν, this produces homogeneous Poisson events (jump times), news arriving times τ_j. Each news has its importance (jump size) X_j, this lends to different transaction frequencies λ. All of this results marked DSPP events for the transactions. A schematical illustration is shown in Figure 3.2.

Hence, marked doubly stochastic Poisson processes can be viewed as two stage stochastic process. Figure 3.2 illustrates marked DSPP as a two-stage stochastic process. The first stage is a homogeneous Poisson process with constant rate ν. Its output becomes the input to stochastic intensity function by Equation 3.2.3, which then produces shot noise as its output. This shot-noise becomes the time varying rate for the last stage, a second Poisson process. The resulting point process is not homogeneous, but rather exhibits the variations of the shot-noise driving process. Thus the two forms of randomness inherent in the DSPP are explicitly separated into two Poisson processes.

In this thesis, we will focus on the bivariate case of intensity. Therefore, taking $K = 2$ in Formular 3.2.1, we have

$$\begin{cases} \tilde{\lambda}_t^{01} = \lambda_t^{(1)} + a_1\lambda_t^{(0)}, \\ \tilde{\lambda}_t^{02} = \lambda_t^{(2)} + a_2\lambda_t^{(0)}, \end{cases} \tag{3.2.4}$$

where $\lambda_t^{(1)}$ and $\lambda_t^{(2)}$ are two specific components, $\lambda_t^{(0)}$ is a common component, and coefficient a_1 (a_2) is a scalar parameter which drives the common component $\lambda_t^{(0)}$ impact on the individual intensities $\lambda_t^{(1)}$ $(\lambda_t^{(2)})$ contemporaneously.

3.3 Simulation procedure

In this section, we schematically illustrate simulation procedure of bivariate stochastic intensity. Because of the limitation of the size of the figure, we divide the simulation procedure into three parts, including Figure 3.3, Figure 3.4 and Figure 3.5. Here we repeat the output of previous figure as the beginning of following one.

Figure 3.3 illustrates the first part of the simulation procedure of bivariate stochastic intensity. The first stage is to simulate three independent homogenous Poisson processes with constant rate v_1, v_0, and v_2 respectively. The output (times) becomes the input to the following procedure. First, each time corresponds to a jump size. Then the number of times serves as the input for generating jump size, where the initial value ($t_0^{(i)} = 0$) of jump size is assumed as Gamma distributed with shape parameter v_i/k_i and scale parameter r_i, and the rest jump size follow independent exponential distribution with parameter $1/r_i$

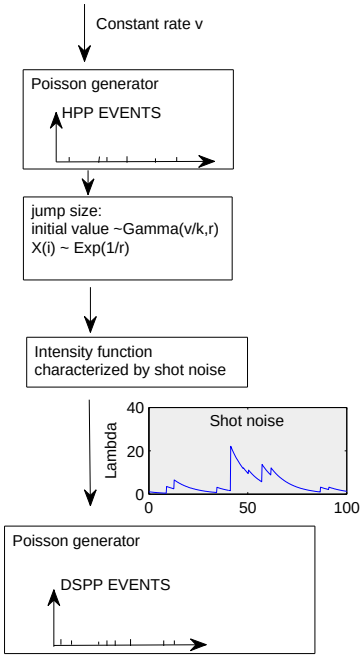

Figure 3.2: A schematical description of marked DSPP. A primary homogeneous Poisson point process $\varphi(t)$ with constant rate ν serves as the input to the impulse response function. The continuous-time stochastic process at the output of this impulse response function is shot noise, which serves as the random rate for another Poisson process whose output is arrival time of event.

for $i = 0, 1, 2$ (as shown in the middle figure of Figure 3.3). Second, all the times and corresponding jump size becomes the input of shot noise intensity function (see Equation 3.2.3), which produce the time varying rate (intensity) $\lambda_t^{(i)}$.

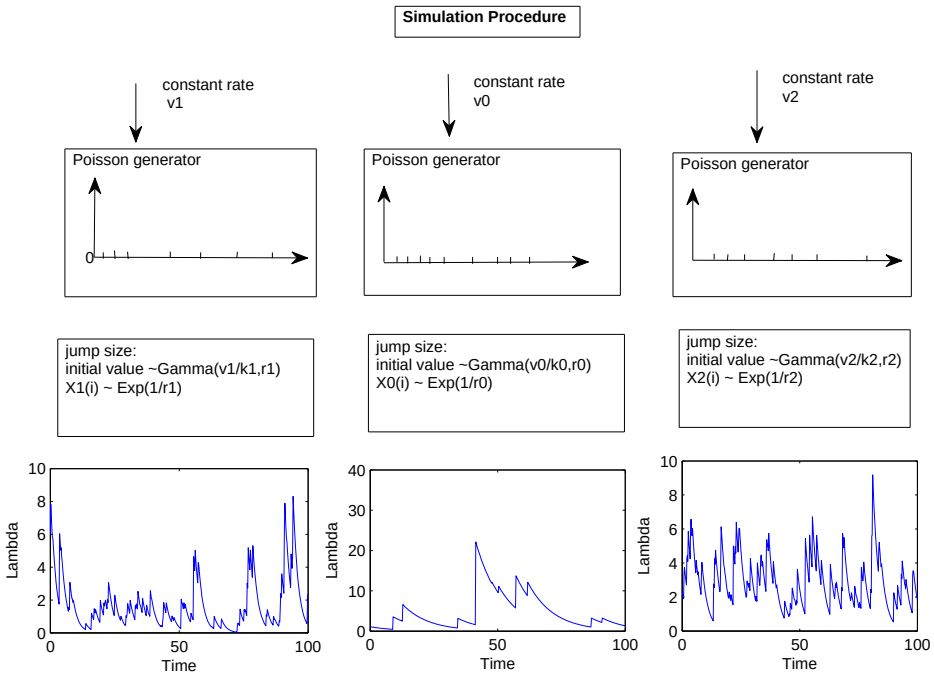

Figure 3.3: Simulation procedure. Three independent constant rate v_1, v_0, v_2 serves as the input to a homogeneous Poisson generator (upper figures). The number of times as the output of each homogeneous Poisson process generator serves as input of the jump size (middle figures). The initial value of jump size follows Gamma distribution with shape parameter v_i/k_i and scale parameter r_i and the rest values of jump size follow independent exponential distribution with parameter $1/r_i$ for $i = 0, 1, 2$. All the times (upper figure) and corresponding jump size (middle figure) serve as the input of shot noise intensity function (Equation 3.2.3), which produce random rate (intensity) $\lambda_t^{(i)}$.

After generating three independent shot noise intensity processes, $\lambda_t^{(1)}$, $\lambda_t^{(0)}$ and $\lambda_t^{(2)}$, according to Equation 3.2.3. As the model described (see Equations 3.2.4), the intensity is consisting of two components, one is individual intensity ($\lambda_t^{(1)}$ and $\lambda_t^{(2)}$) and the other is common intensity component ($\lambda_t^{(0)}$), with coefficients a_1 and a_2 which are process-specific scaling parameters leading the common component's impact on the individual intensity component. The bottom figures of Figure 3.4 present time points generating by homogeneous Poisson process with constant rate d_1 (left bottom of Figure 3.4) and d_2 (right bottom Figure 3.4), which are maxmum values of the underlying intensities λ_t^{01} (left middle of Figure 3.4) and λ_t^{02} (right middle of Figure 3.4) for given time T ($T = 100$ in the simulation procedure figure).

To simulate time points according to the underlying stochastic shot noise intensity which is time-varying process, the thinning method is applied here. Basically, the thin-

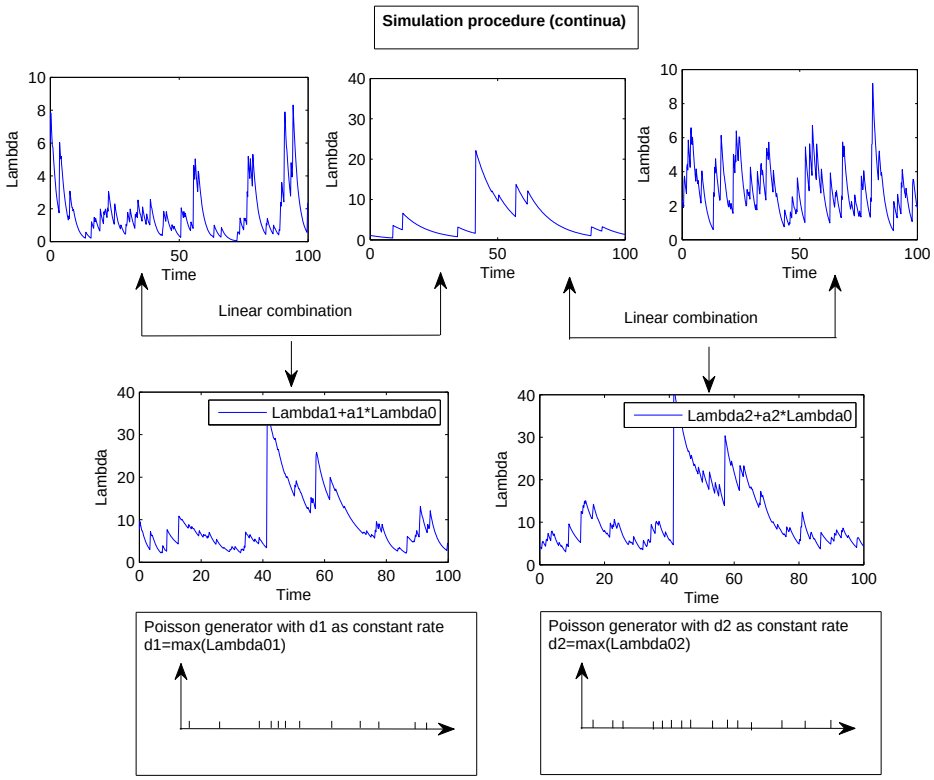

Figure 3.4: Simulation procedure (continue). Three independent shot noise intensity processes as the output of the previous procedure (Figure 3.3) become the component of a linear combination. $\lambda_t^{(1)}$ (upper left) and $\lambda_t^{(2)}$ (upper right) are individual components while $\lambda_t^{(0)}$ (upper middle) is the common component. The middle figures provide the resulting stochastic (shot noise) intensity process after the linear combination. a_1 and a_2 are constant parameters providing the weight of common component $\lambda_t^{(0)}$ on the whole intensities, λ_t^{01} and λ_t^{02} (middle figure). Take the maxmum value of intensity, λ_t^{01} and λ_t^{02}, denoted by d_1 and d_2 respectively. Then d_1 and d_2 serve as the input (constant rate) of homogeneous Poisson generator (bottom figure).

ning method can be summarized as follows:

Given a simulated trajectory $\lambda_{[0,T]}$ of the intensity process, the thinning method is given by the following steps (Last & Brandt 1995, Prop. 9.1.7, p. 278):

1. Generate a random number n of points from a Poisson process with constant intensity $d = \max\{t \in [0,T]\lambda_{[0,T]}\}$. A sequence $t_1, t_2, \cdots, t_n$ is obtained by simulating independent interarrival times w_i from an exponential distribution with mean $1/d$ and by setting $t_i = t_{i-1} + w_i$ until T is reached.

2. Generate u_i from uniform distribution $\mathbf{U}(0,1)$.

3. If $\lambda_{[0,T]}(t_i)/d > u_i$ then keep point t_i, otherwise delete t_i.

4. Repeat steps 2 and 3 for $i = 1, 2, \ldots, n$.

The output (times) of Figure 3.4 are homogeneous Poisson points ($\{t_i^1\}$ and $\{t_i^2\}$), the thinning method providing a way to randomize the data such that the resulting time points (accepted time points) are mimic from the underlying stochastic shot noise intensities. As for the thinning method, there are basically two stages. First, generate a value u from a standard uniform distribution $\mathbf{U}(0,1)$; second, comparing the value $\lambda^{01}_{t_i^1}/d_1$ ($\lambda^{01}_{t_i^2}/d_2$) with u, if the former is greater than the latter, then t_i^1 (t_i^2) is accepted; otherwise, t_i^1 (t_i^2) is removed (as shown in the middle figures of Figure 3.5). As for the last stage of the simulation procedure, the time pooling provides a sequence of time points orderly composed of the time points from underlying stochastic shot noise intensities, λ_t^{01} and λ_t^{02}.

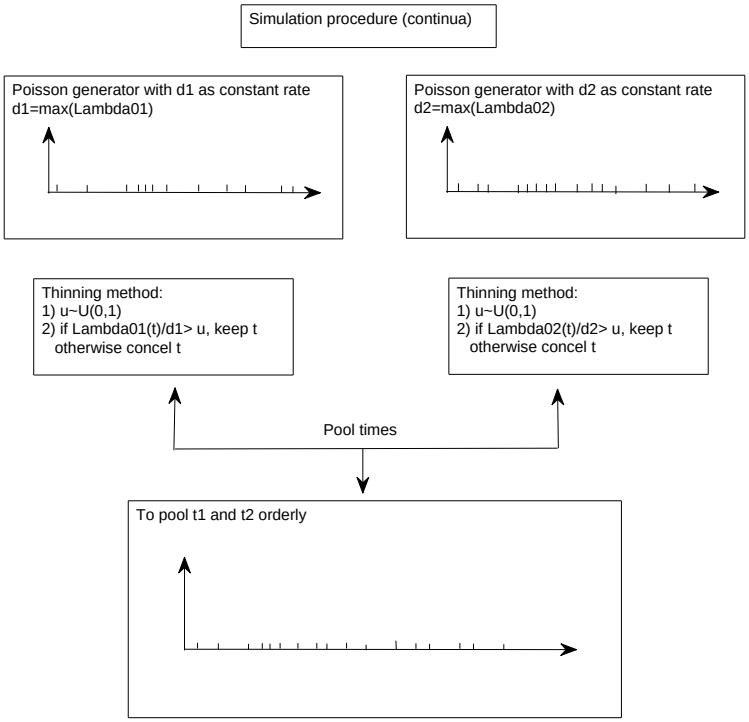

Figure 3.5: Simulation procedure (continue). Given the time points from homogeneous Poisson process with constant rate d_1 and d_2 (upper figures), thinning method is applied such that inhomogeneous Poisson processes are obtained. The middle figures outline the thinning method: firstly, generate a value u from a standard uniform distribution $\mathbf{U}(0,1)$; second, comparing the value $\lambda^{01}_{t^1_i}/d_1$ ($\lambda^{01}_{t^2_i}/d_2$) with u, if the former is greater than the latter, then t^1_i (t^2_i) is accepted; otherwise, t^1_i (t^2_i) is canceled. The last stage pool all the times (accepted after thinning method) orderly.

3.4 Five modeling frameworks for logreturn processes

In the following, we will present five modeling frameworks for logreturn processes: (1) returns generated from independent Wiener process; (2) concurrently correlated Wiener process; (3) bivariate autoregressive with first lag (Bi-AR(1)); (4) Bi-AR(1) by specializing zero cross-correlation; and (5) independent Bi-AR(1). The aim is to compare these five frameworks to see which one fits the empirical data better from the point of view of the implied correlations, which are among the basic characteristics that we are trying to model.

3.4.1 Pooled times

The problem with simulating the bivariate logreturn processes is non-synchronicity of two time series. Basically, the generated data neither coincide at the same time nor have the same number of transactions for two series. To tackle this problem, some kind of techniques must be introduced. Here we adopt the method used in Bauwens & Hautsch(2006), where they pool two transaction times orderly. Let t denote physical time and let $\{t_i\}_{i=1,\ldots,N}$, $N = n^1 + n^2$, denote pooled time, where n^1 is the number of transaction times for asset one and n^2 is the number of transactions for asset two. Precisely, $0 < t_1 < t_2 < \ldots < t_N$, which is pooling two sequences of transaction times orderly. In the sense that if j^{th} transaction time of asset one, denoted as t_j^1, is ordered after t_m, then $t_{m+1} = t_j^1$. $\{t_j^1\}_{j=1,\ldots,n^1}$ and $\{t_j^2\}_{j=1,\ldots,n^2}$ denote the transaction times for asset one and asset two, respectively.

Figure 3.6 illustrates a graphic example for pooled times. The ticks on the first arrow represent the transaction times for asset one while that on the second arrow indicate the transaction times for asset two. All the ticks on the third arrow show the pooled times by collecting both transaction times of asset one and of asset two orderly. The vertical dashed line displays the connection between individual transaction time and the pooled one. Note that the pooled times do not need to be equally spaced.

Accordingly, such pooled times can be treated as the 'transaction times' for both two assets. And the 'trades' they made coincide at the same time, though we do not require that they are equally spaced. Indeed, this is a way to generate a more general Wiener process (see Section 3.4.3 and Section 3.4.4) in the sense that is not measuring in one unit time but in varied time. Given the pooled times, it is convenient to simulate the bivariate return process according to the model of interest.

Though the pooled time approach solve the problem of simulating asynchronous time series, one trade-off is create some arrival times, perhaps a lot, at which no transaction are made. For example, the pooled time $\{t_i\}_{i=1,2,\ldots,8}$, in Figure 3.6, gives asset one (the first axes in Figure 3.6) three additional arrival times, that is, $tt1, tt2, tt3$. We call the arrival times, $tt1, tt2, tt3$, in side of asset one, as 'artificial times' and the rest, $t1, t2, t3, t4, t5$, as 'true times'. Since the logreturn generating work are based on the pooled time, there will be a lot of 'artificial' logreturns involved consequently. In order to keep the simulated logreturns genuine, at least the number of transaction times should be the same as original one, n^1 for asset one and n^2 for asset two, but not N as a whole for both of these two assets. Then, our task is to keep the 'true' logreturns and get rid of those 'artifical' logreturns and this can be handled in Matlab. We call this procedure as data matching.

Notice that the rest 'true' logreturns after the data matching, the number of logreturns for each asset match with those original arrival times. Moreover, the non-synchronicity

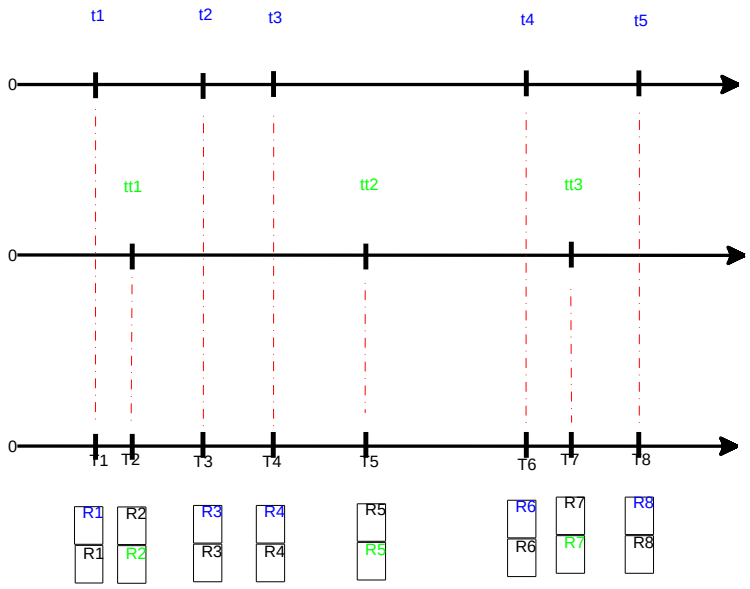

Figure 3.6: A graphic example for the term ‘pooled times’. The first axis indicates the transaction time flow for asset one; the second axis indicates the transaction time flow for asset two; The third axis represents the pooled times by collecting the transaction times from the asset one (the first axis) and the asset two (the second axis) orderly. The veritical dashed line shows the ordered transaction time associated with original transaction flow.

property of the logreturns is restored.

3.4.2 Definition of logreturn

Up to now, the term 'logreturn' has not been defined. To avoid confusion with other version of 'logreturn' defined in testbook or other papers, the term 'logreturn' throughout this article will be defined as the following form

$$R_k(t_i) = \log S_k(t_i) - \log S_k(t_{i-1}), \tag{3.4.1}$$

where $r_k(t_i)$ is the logreturn of asset k at time t_i and $S_k(t_i)$ denotes the transaction price of asset k at time t_i, $k = 1, 2$. Note that here $t_i - t_{i-1}$ is real valued and not discrete as in many discrete time modeling frameworks.

From Equation 3.4.1, the logreturn $R_k(t)$ can also be interpreted as the log-price increment or log-price movement.

3.4.3 Independent Wiener process

The first modeling framework is based on the Wiener process. Let $R_k(t_i)$ be the logreturn of k^{th} asset at time t_i, $k = 1, 2$, $i = 1, 2, \ldots, N$, where N is the number of pooled times. The bivariate logreturn processes is given by:

$$\begin{cases} R_1(t_i) = \mu_1 + \xi_1(t_i), \\ R_2(t_i) = \mu_2 + \xi_2(t_i), \end{cases} \tag{3.4.2}$$

where $\xi^k_{t_j} \sim \mathbf{N}\big(0, (t^k_j - t^k_{j-1})\sigma_{kk}\big)$ and μ_k is the mean of R^k, $k = 1, 2$.

The basic idea of Wiener process is that it has independent increment and Normal distribution with mean zero and variance proportional to inter-arrival time. At this point, logreturn process is correlated with the arriving rate (intensity). If inter-arrival time is small, corresponding to high intensity (almost surely), then the logreturn will have low variance. On the other hand, if the underlying intensity is pretty low, corresponding to long inter-arrival time(almost surely), then the variance of the logreturn responds to a high level according to the property of Wiener process, and great fluctuation of the logreturns will be observed.

Given Equation 3.4.2, the logreturns $R_k(t)$ is totally characterized by the error term $\xi_k(t)$, which follows independent Wiener process, $k = 1, 2$. More formally, the error term $\boldsymbol{\xi}(t_i) = (\xi_1(t_i), \xi_2(t_i))'$ is given by

$$\boldsymbol{\xi}(t_i) = \begin{pmatrix} \xi_1(t_i) \\ \xi_2(t_i) \end{pmatrix} \sim \mathbf{N}\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, (t_i - t_{i-1}) \begin{pmatrix} \sigma_{11} & 0 \\ 0 & \sigma_{22} \end{pmatrix} \right), \tag{3.4.3}$$

where $\sigma_{11} > 0$, $\sigma_{22} > 0$ are parameters. Note that the positive definite requirement of the covariance matrix $\Sigma = (t_i - t_{i-1}) \begin{pmatrix} \sigma_{11} & 0 \\ 0 & \sigma_{22} \end{pmatrix}$ is satisfied in this case, because all the elements of the diagonal matrix Σ are positive. The flexibility of this model is characterized by setting the parameters σ_{11} and σ_{22} which can be different from one. Besides, variance of $\xi_1(t_i)$ ($\xi_2(t_i)$) depends on time increment $(t_i - t_{i-1})$. This is a basic property of (standard) Wiener process. Notice that $(t_i - t_{i-1})$ is real valued but not discrete as we claimed before.

The logreturns $R_k(t_i)$, $k = 1, 2$, are therefore independent normally distributed with mean μ_k, variance $(t_i - t_{i-1})\sigma_{kk}$, and covariance equal to zero. Note that the covariance between $R_k(t_i)$ and $R_k(t_{i+1})$ is zero. This is due to the independence of ξ_{kt}. Specially, for $0 = t_0 < t_1 < \ldots < t_N$, $\xi_{t_{i-1}}$ and ξ_{t_i} are independent, $i = 1, 2, \ldots, N$.

It is worth stressing that the logreturn model based on the independent Wiener process nests two important properties. First, two sequencies of logreturns, $\{R_1(t)\}$ and $\{R_2(t)\}$, are concurrently independent. This implies that logreturns $R_1(t)$ and $R_2(t)$ are independent at the same time t, $t \in \{t_1, \ldots, t_N\}$. The second property is both of the two logreturns, $\{R_1(t)\}$ and $\{R_2(t)\}$, are independent serially. In the sense that each sequence, $R_1(t)$ for instance, $R_1(t_i)$ is independent from $R_1(t_j)$ for any $i \neq j$, $i, j = 1 \ldots, N$. This simple framework is worthwhile to discussing, because, in some extent, it verifies whether the common intensity would impact on the correlation of the logreturns through the time depending variance.

3.4.4 Concurrently correlated Wiener process

The second modeling framework differs from the first one in that we introduce correlation into Wiener process. The form of logreturn process is the same as Equation 3.4.2, let the error term $\boldsymbol{\xi}(t_i) = (\xi_1(t_i), \xi_2(t_i))$ be independent normally distributed with mean zero and covariance matrix $\Sigma = \begin{pmatrix} \sigma_{11}\sigma_{12} \\ \sigma_{21}\sigma_{22} \end{pmatrix}$. More formally,

$$\boldsymbol{\xi}(t_i) = \begin{pmatrix} \xi_1(t_i) \\ \xi_2(t_i) \end{pmatrix} \sim N\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, (t_i - t_{i-1}) \begin{pmatrix} \sigma_{11} & \sigma_{12} \\ \sigma_{21} & \sigma_{22} \end{pmatrix} \right), \tag{3.4.4}$$

where $\sigma_{12} = \sigma_{21} \neq 0$, $\sigma_{11} > 0$, $\sigma_{22} > 0$ are parameters.

The idea of this modeling framework is to let two sequences of logreturns $\{R_1(t)\}$ and $\{R_2(t)\}$ be concurrently correlated but still serially independent, $\{t_i\}_{i \in \{1,\ldots,N\}}$ refers to pooled times. From Equation 3.4.2 and Equation 3.4.4, the logreturn process $r_k(t)$ have independent Normal distribution with mean μ_k and variance $\sigma_{kk}(t_i - t_{i-1})$, $k = 1, 2$ and $t_0 = 0$. the only difference between this model and the first one is that the covariance between two series $R_1(t_i)$ ($R_2(t_i)$) and $r_2(t_i)$ ($R_1(t_i)$) is $\sigma_{12}(t_i - t_{i-1})$ ($\sigma_{21}(t_i - t_{i-1})$). So the concurrent correlation between two series $\{R_1(t)\}$ and $\{R_2(t)\}$ depends on the inter-arrival time $(t_i - t_{i-1})$. High correlation is corresponding to long inter-arrival time, while low correlation responds to short inter-arrival time. But this is not the case, since the variance is also proportional to the inter-arrival time. The correlation coefficient (ρ) is then an appropriate measurement and $\rho = \sigma_{12}/\sqrt{\sigma_{11} \times \sigma_{22}}$. Thereby, correlation between two series $\{R_1(t)\}$ and $\{R_2(t)\}$ has nothing to do with the inter-arrival time.

One important property nests in this modeling framework is, therefore, that the inter-arrival time or intensity plays a special role in the fluctuation of each logreturn process through variance of ξ_1 (ξ_2) but there is no place for the correlation between two logreturn series. The correlation between $\{R_1(t)\}$ and $\{R_2(t)\}$ is fixed by the parameters σ_{11}, σ_{22} and σ_{12}. While higher intensity corresponds to lower fluctuation of each logreturn process, vice versa.

3.4.5 Bivariate autoregressive process of the first order

The third modeling framework of logreturns is quite different from the previous two models in which the logreturns are constructed based on the Wiener process. In what follows, we will introduce a simple vector model which is particularly useful for modeling asset returns, it is so called vector autoregressive (VAR) model. A multivaiate time series $\mathbf{r}(t_i)$ is a VAR process of order 1, or VAR(1) for short, if follows the model

$$\mathbf{R}(t_i) = \boldsymbol{\beta}_0 + \boldsymbol{\beta}_1 \mathbf{R}(t_{i-1}) + \boldsymbol{\varepsilon}(t_i), \tag{3.4.5}$$

where $\boldsymbol{\beta}_0$ is a k dimensional vector, $\boldsymbol{\beta}_1$ is a $k \times k$ matrix, and $\{\boldsymbol{\varepsilon}(t_i)\}$ is a sequence of serially uncorrelated random vectors with mean zero and covariance matrix Σ. Again, the time sequence $\{t_i\}_{i\in\{0,1,\ldots,N\}}$ refers to pooled times. In practice, the covariance matrix Σ is required to be postive definite; otherwise, the dimension of $\mathbf{R}(t_i)$ can be reduced. Moreover, it is often assumed that $\boldsymbol{\varepsilon}(t_i)$ has multivairiate nomal distribution. Without specification, we also assume that $\boldsymbol{\varepsilon}(t_i)$ is normal distributed.

The coefficient matrix $\boldsymbol{\beta}_1$ measure the dynamic dependence of $\mathbf{R}(t)$. The diagonal elements of matrix $\boldsymbol{\beta}_1$ measure the autocorrelation coefficient for each asset $\{R_i(t)\}$, $i = 1, \ldots, K$, that is, the coefficient of linear dependence on its most previous time. While the off-diagonal elements of matrix $\boldsymbol{\beta}_1$ measure auto-cross-correlation coefficient, namely, β_{ij}, for example, denotes the coefficient of linear dependence of i^{th} asset at time t, $R_i(t)$, on j^{th} asset at time $t-1$, $R_j(t-1)$, where $i, j = 1, \ldots, K$.

We also assume that k eigenvalues of $\boldsymbol{\beta}_1$ are less than one to avoid the explosion of $\mathbf{R}(t_i)$. This is because $\mathbf{R}(t_i)$ can be expressed as an explicit function of $\boldsymbol{\varepsilon}(t_i)$ with coefficient $\boldsymbol{\beta}_1$.

For the concurrent-cross-correlation, it is implicitly supported by the covariance matrix Σ. Taking bivariate case for instance, the linear dependence between $\{R_1(t)\}$ and $\{R_2(t)\}$ is given by the off-diagonal elements σ_{12} and σ_{21} of the covariance matrix Σ of $\boldsymbol{\varepsilon}_t$. If $\sigma_{12} = \sigma_{21} = 0$, then there is non concurrent linear relationship between the two series, $\{R_1(t)\}$ and $\{R_2(t)\}$; otherwise, the two time series $\{R_1(t)\}$ and $\{R_2(t)\}$ are concurrently positively correlated when $\sigma_{12} = \sigma_{21} > 0$, whereas negatively correlated when $\sigma_{12} = \sigma_{21} < 0$.

Property of VAR(1)

Assume that VAR(1) model is weakly stationary, taking expectation of Equation 3.4.5 , and let $E(\boldsymbol{\varepsilon}_t) = 0$, yielding:

$$\mathrm{E}(\mathbf{R}_t) = \boldsymbol{\beta}_0 + \boldsymbol{\beta}_1 \mathrm{E}(\mathbf{R}_{t-1}). \tag{3.4.6}$$

Since $\mathrm{E}(R_t)$ is time-invariant, we have

$$\boldsymbol{\mu} \equiv \mathrm{E}(\mathbf{R}_t) = \boldsymbol{\beta}_0 + \boldsymbol{\beta}_1 \boldsymbol{\mu}, \tag{3.4.7}$$

Shifting $\boldsymbol{\beta}_1 \boldsymbol{\mu}$ to the left and collecting $\boldsymbol{\mu}$, we obtain

$$(I - \boldsymbol{\beta}_1)\boldsymbol{\mu} = \boldsymbol{\beta}_0, \tag{3.4.8}$$

where I is the $K \times K$ identity matrix.

Moving $(I - \boldsymbol{\beta}_1)$ to the right part, we have

$$\boldsymbol{\mu} = (I - \boldsymbol{\beta}_1)^{-1}\boldsymbol{\beta}_0. \tag{3.4.9}$$

It is worth stressing that the sufficient condition for Equation 3.4.9 is requiring that the matrix $I - \boldsymbol{\beta}_1$ is nonsingular. This can be satisfied by our assumption that all the eigenvalue of $\boldsymbol{\beta}_1$ is less than one. In fact, such specification for $\boldsymbol{\beta}_1$ is sufficient and necessary condition to obtain weakly stationary distribution of return process $\mathbf{R}_t$. The following setting gives out an explaination in terms of formular.

Taking $\boldsymbol{\beta}_0 = (I - \boldsymbol{\beta}_1)\boldsymbol{\mu}$ into Equation 3.4.5, yielding:

$$\mathbf{R}_t - \boldsymbol{\mu} = \boldsymbol{\beta}_1(\mathbf{R}_{t-1} - \boldsymbol{\mu}) + \boldsymbol{\varepsilon}_t. \tag{3.4.10}$$

Set $\hat{\mathbf{R}}_t = \mathbf{R}_t - \boldsymbol{\mu}$, then VAR(1) can be rewritten as:

$$\hat{\mathbf{R}}_t = \boldsymbol{\beta}_1\hat{\mathbf{R}}_{t-1} + \boldsymbol{\varepsilon}_t. \tag{3.4.11}$$

Now repeat this substitution, yielding:

$$\begin{aligned}
\hat{R}_t &= \boldsymbol{\beta}_1(\boldsymbol{\beta}_1\hat{R}_{t-1} + \boldsymbol{\varepsilon}_{t-1}) + \boldsymbol{\varepsilon}_t \\
&= \boldsymbol{\beta}_1^2\hat{R}_{t-2} + \boldsymbol{\beta}_1\boldsymbol{\varepsilon}_{t-1} + \boldsymbol{\varepsilon}_t \\
&= \boldsymbol{\beta}_1^2(\hat{R}_{t-3}\boldsymbol{\beta}_1 + \boldsymbol{\varepsilon}_{t-2}) + \boldsymbol{\beta}_1\boldsymbol{\varepsilon}_{t-1} + \boldsymbol{\varepsilon}_t \\
&= \boldsymbol{\beta}_1^3\hat{R}_{t-3} + \boldsymbol{\beta}_1^2\hat{R}_{t-2} + \boldsymbol{\beta}_1\boldsymbol{\varepsilon}_{t-1} + \boldsymbol{\varepsilon}_t \\
&= \dots \\
&= \boldsymbol{\varepsilon}_t + \boldsymbol{\beta}_1\boldsymbol{\varepsilon}_{t-1} + \boldsymbol{\beta}_1^2\boldsymbol{\varepsilon}_{t-2} + \cdots + \boldsymbol{\beta}_1^l\boldsymbol{\varepsilon}_{t-l} + \boldsymbol{\beta}_1^{l+1}\hat{R}_{t-l-1}.
\end{aligned}$$

Repeat this iteration infinitely, we will get return process $\mathbf{R}_t$ in terms of noise term $\boldsymbol{\varepsilon}_{t-i}$, $i = 0, 1, 2, \dots$ and is given by

$$\hat{R}_t = \boldsymbol{\varepsilon}_t + \boldsymbol{\beta}_1\boldsymbol{\varepsilon}_{t-1} + \boldsymbol{\beta}_1^2\boldsymbol{\varepsilon}_{t-2} + \cdots. \tag{3.4.12}$$

This Equation 3.4.12 implies several characteristics of VAR(1) model for return process:

1. Since $\boldsymbol{\varepsilon}_t$ is serially uncorrelated, $\text{Cov}(\boldsymbol{\varepsilon}_t, \mathbf{R}_{t-l}) = 0$, for all $l > 0$ and $\text{Cov}(\boldsymbol{\varepsilon}_t, \mathbf{R}_t) = \Sigma$, therefore,$\boldsymbol{\varepsilon}_t$ refers to shock or innovation of return series at time t. $\mathbf{R}_t$ dependents on the past shock $\mathbf{R}_{t-j}$ with coefficient matrix $\boldsymbol{\beta}_1^j$, in order to make such dependence to be meaningful, $\boldsymbol{\beta}_1^j$ must converge to 0 as j approach to ∞. This means that the K eigenvalues of $\boldsymbol{\beta}_1$ must be less than 1, otherwise, $\boldsymbol{\beta}_1^j$ will either explode or converge to a nonzero matrix as j approach to ∞. In fact, the requirement of all eigenvalue of $\boldsymbol{\beta}_1$ are less than 1 is necessary and sufficient condition for weak stationary of R_t.

2. Based on Equation 3.4.12, if we take the expectation of $\hat{\mathbf{R}} = \mathbf{R}_t - \boldsymbol{\mu}$ and its transpose, we obtain

$$\begin{aligned}
\text{Cov}(R_t) \equiv \Gamma &= E(\mathbf{R}_t - \boldsymbol{\mu})(\mathbf{R}_t - \boldsymbol{\mu})' = E(\hat{\mathbf{R}}_t\hat{\mathbf{R}}_t') \\
&= \Sigma + \boldsymbol{\beta}_1\Sigma\boldsymbol{\beta}_1' + \boldsymbol{\beta}_1^2\Sigma\boldsymbol{\beta}_1^{2'} + \cdots = \sum_{j=0}^{\infty}\boldsymbol{\beta}_1^j\Sigma\boldsymbol{\beta}_1^{j'}
\end{aligned}$$

   Note that in case $j = 0$, we assume $\boldsymbol{\beta}_1^0 = I$, which is a $K \times K$ identity matrix.

3. Consider the autoregression with lag l, the covariance is then

$$\begin{aligned}\mathrm{E}(\mathbf{R}_t-\boldsymbol{\mu})(\mathbf{R}_{t-l}-\boldsymbol{\mu})' &= \mathrm{E}(\hat{\mathbf{R}}_t\hat{\mathbf{R}}'_{t-l})\\ &= \mathrm{E}(\boldsymbol{\beta}_1\hat{R}_{t-1}+\boldsymbol{\varepsilon}_t, \hat{R_{t-l}}')\\ &= \boldsymbol{\beta}_1\mathrm{E}(\hat{R}_{t-1},\hat{R}_{t-l}), l>0\end{aligned}$$

therefore,

$$\Gamma_l = \boldsymbol{\beta}_1\Gamma_{l-1}, \tag{3.4.13}$$

where Γ_j is the lag-j cross covariance matrix of R_t.

If repeat this substitution, we obtain

$$\Gamma_l = \boldsymbol{\beta}_1^l\Gamma_0. \tag{3.4.14}$$

VAR(1) model can be considered as the generalized model of return process. Any specific assumption attached to this model, then different framework will be obtained. In general, one would prefer to give speical attention on matrix $\boldsymbol{\beta}_1$ or noise term $\boldsymbol{\varepsilon}_t$. Consider, for example, GARCH(1,1) attempting to specialize the error term $\boldsymbol{\varepsilon}_t$ and obtaining remarkable result of interest. In the following, we will focus on the specializition of matrix $\boldsymbol{\beta}_1$. If the off-diagonal element of matrix $\boldsymbol{\beta}_1$, β_{12} and β_{21}, set to zero then VAR(1)model reduced to concurrently correlated AR(1). In extreme case, set matrix $\boldsymbol{\beta}_1 = 0$ so that the multiplicative return process just concurrently correlated without any lagged autocorrelation or cross-correlation.

Now, we consider a bivaiate case, that is, $k = 2$, $\mathbf{R}_{t_i} = [R_1(t_i), R_2(t_i)]'$, and $\boldsymbol{\varepsilon} = [\epsilon_1(t_i), \epsilon_2(t_i)]'$, denoted as Bi-AR(1). Essentially, Bi-AR(1) model can be expressed in the following two equations:

$$\begin{cases} R_1(t_i) = \beta_{01} + \beta_{11}R_1(t_{i-1}) + \beta_{12}R_2(t_{i-1}) + \epsilon_1(t_i), \\ R_2(t_i) = \beta_{02} + \beta_{21}R_1(t_{i-1}) + \beta_{22}R_2(t_{i-1}) + \epsilon_2(t_i), \end{cases} \tag{3.4.15}$$

where β_{ij} is the (i,j)th element of matrix $\boldsymbol{\beta}_1$ and β_{0i} is the i^{th} element of vector $\boldsymbol{\beta}_0$. Based on the first equation of Equation 3.4.15, β_{12} indicates the linear dependence of $R_1(t_i)$ on $R_2(t_{i-1})$ in the presence of $R_1(t_{i-1})$. Therefore, β_{12} can be considered as the conditional effect of $R_2(t_{i-1})$ on $R_1(t_i)$ given $R_1(t_{i-1})$. Specially, if $\beta_{12} = 0$, then $R_1(t_i)$ does not depend on $R_2(t_{i-1})$, and the model shows that $R_1(t_i)$ only depends on its own past. Similarly, if $\beta_{21} = 0$, then the second equation of Equation 3.4.15 shows that $R_2(t_i)$ does not depend on $R_1(t_{i-1})$ given $R_2(t_{i-1})$. This typical model will be addressed in detail in Section 3.4.6 and Section 3.4.7.

In general, the coefficient matrix $\boldsymbol{\beta}_1$ measures the dynamic dependence of $\mathbf{R}(t_i)$. Now consider Equation 3.4.15, if $\beta_{12} = 0$ and $\beta_{21} \neq 0$, then there is no directional relationship from $R_1 t_i$ to $R_2(t_i)$ while there exist linear dependence of $R_2(t_i)$ on $R_1(t_{i-1})$ given $R_2(t_{i-1})$. Second, if $\beta_{12} \neq 0$ and $\beta_{21} \neq 0$, then there is a feedback relationship between the two series $\{R_1(t)\}_{t\in\{t_0,t_1,\ldots,t_N\}}$ and $\{R_2(t)\}_{t\in\{t_0,t_1,\ldots,t_N\}}$. However, if $\beta_{12} = \beta_{21} = 0$, then $R_1(t)$ and $R_2(t)$ are uncoupled, they only depend on their own past.

On the other hand, the concurrent relationship between $R_1(t_i)$ and $R_2(t_i)$ is shown by the off-diagonal element σ_{12} of the covariance matrix Σ of $\boldsymbol{\varepsilon}(t_i)$, where Σ is given by

$$\Sigma = \begin{pmatrix} \sigma_{11} & \sigma_{12} \\ \sigma_{21} & \sigma_{22} \end{pmatrix}.$$

If $\sigma_{12} = \sigma_{21} = 0$ then there is no concurrent linear relationship between the two series, more details concerning on this argument refer to Section 3.4.7.

One point is somehow subtle but still quite important is that Bi-AR(1) model has weaker dependence on intensity with respect to the Wiener process. This is because the only factor relevant to intensity displays in Bi-AR(1) model is the number of the pooled times N. This leads to one crucial shotcoming that the number of the pooled times N actually given by the averaged quantity of the intensity process, which could not specify the process of the underlying intensity. In particular, when the intensity follows shot noise process in which the underlying intensity is a stochastic process and incorporates several jumps. Note that the subscript t_i, in fact, can be written as usual t, we use the note t_i for two reasons. One is just want to refer the time to pooled times; the other which is usually omitted is that the time measurement is not in terms of one unit time, that is, $t_i - t_{i-1}$ is not necessary to be one.

After the previous description of Bi-AR(1), it is convenient to summarize what matters in this section about Bi-AR(1). Basically, there are two factors determine the underlying model. One is $\boldsymbol{\beta}_1$ which characterizes the relationship of autocorrelation and cross-correlation between two time series $\{R_1(t)\}$ and $\{R_2(t)\}$. The other is Σ which relates to the correlation between two time series $\{R_1(t)\}$ and $\{R_2(t)\}$ concurrently or not. Here we concern on two time series in more complicated situation where $\{R_1(t)\}$ and $\{R_2(t)\}$ are not only autocorrelated but also cross-correlated, moreover, they are concurrenlty correlated. So $\beta_{ij} \neq 0$, $\sigma_{ij} \neq 0$, for all $i, j = 1, 2$. We specify two time series $\{R_1(t)\}$ and $\{R_2(t)\}$ without cross-correlation but concurrently correlated, that is, $\beta_{12} = \beta_{21} = 0$, $\beta_{ii} \neq 0$ and $\beta_{ij} \neq 0$, for all $i, j = 1, 2$, in Section 3.4.6. Finally, in the section 3.4.7 we restrict Bi-AR(1) model to very speical case that two time series $\sigma_{ij} \neq 0$, for all $i, j = 1, 2$ are just autocorrelated, that is, $\beta_{12} = \beta_{21} = 0$, $\beta_{ii} \neq 0$ and $\sigma_{12} = \sigma_{21} =$, $\sigma_{ii} \neq 0$, for all $i = 1, 2$.

3.4.6 Bi-AR(1) with zero cross-correlation

The fourth modeling framework is based on the Bi-AR(1) as described in Section 3.4.5 by specifying the off-diagonal elements of $\boldsymbol{\beta}_1$, that is, β_{12} and β_{21}, equal to zero. So two time series $\{R_1(t)\}$ and $\{R_2(t)\}$ are autocorrelated per se and concurrently correlated between them.

To ensure the two eigenvalues of $\boldsymbol{\beta}_1$ be less than one and greater than zero, we assume $0 < \beta_{kk} < 1$. This assumption is sufficient and necessary condition for guarantee the these two eigenvalues be less than one and greater than zero. As a matter of fact, the determinant of $\boldsymbol{\beta}_1$ which is $\beta_{11} \cdot \beta_{22}$. Based on the algebraic theory, the determinant of $\boldsymbol{\beta}_1$ is equivalent to the product of two eigenvalues of $\boldsymbol{\beta}_1$, say η_1 and η_2, so $\eta_1 \cdot \eta_2 = \beta_{11} \cdot \beta_{22}$. Under the assumption that $0 < \beta_{11} < 1$ and $0 < \beta_{22} < 1$, we can obtain $0 < \beta_{11} \cdot \beta_{22} < 1$, therefore, $0 < \eta_1 \cdot \eta_2 = \beta_{11} \cdot \beta_{22} < 1$. Notice that it does not mean that $\eta_1 = \beta_{11}$ and $\eta_2 = \beta_{22}$.

As shown in previous section, we obtain the cross covariance matrix of $\mathbf{R}_{t_i}$ as follows:

$$\boldsymbol{\Gamma}_l \equiv [\Gamma_{ij}(l)] = \boldsymbol{\beta}_1^l \boldsymbol{\Gamma}_0, \tag{3.4.16}$$

where $\mathbf{\Gamma}_l$ is the lag l cross covariance matrix of $\mathbf{R}(t_i)$; $\mathbf{\Gamma}_0$ is the covariance matrix of $\mathbf{R}(t_i)$ which is given by:

$$\mathbf{\Gamma}_0 = \Sigma + \boldsymbol{\beta}_1\Sigma\boldsymbol{\beta}_1' + \boldsymbol{\beta}_1\Sigma\boldsymbol{\beta}_1' + \boldsymbol{\beta}_1^2\Sigma\boldsymbol{\beta}_1'^2 + \ldots = \sum_{j=0}^{\infty}\boldsymbol{\beta}_1^j\Sigma\boldsymbol{\beta}_1'^j. \tag{3.4.17}$$

From Equation 3.4.17, the bivariate case can be expressed:

$$\begin{aligned}
&\Gamma_0 = \textstyle\sum_{j=0}^{\infty}\boldsymbol{\beta}_1^j\Sigma\boldsymbol{\beta}_1'^j \\
&= \begin{pmatrix} \sigma_{11} & \sigma_{12} \\ \sigma_{21} & \sigma_{22} \end{pmatrix} + \begin{pmatrix} \beta_{11} & 0 \\ 0 & \beta_{22} \end{pmatrix}\begin{pmatrix} \sigma_{11} & \sigma_{12} \\ \sigma_{21} & \sigma_{22} \end{pmatrix}\begin{pmatrix} \beta_{11} & 0 \\ 0 & \beta_{22} \end{pmatrix} \\
&+ \begin{pmatrix} \beta_{11}^2 & 0 \\ 0 & \beta_{22}^2 \end{pmatrix}\begin{pmatrix} \sigma_{11} & \sigma_{12} \\ \sigma_{21} & \sigma_{22} \end{pmatrix}\begin{pmatrix} \beta_{11}^2 & 0 \\ 0 & \beta_{22}^2 \end{pmatrix} + \cdots \\
&= \begin{pmatrix} \sigma_{11} + \beta_{11}^2\sigma_{11}^2 + \beta_{11}^4\sigma_{11} + \cdots & \sigma_{12} + \beta_{11}\beta_{22}\sigma_{12}^2 + \beta_{11}^2\beta_{22}^2\sigma_{12} + \cdots \\ \sigma_{21} + \beta_{11}\beta_{22}\sigma_{21}^2 + \beta_{11}^2\beta_{22}^2\sigma_{21} + \cdots & \sigma_{22} + \beta_{22}^2\sigma_{22}^2 + \beta_{22}^4\sigma_{22} + \cdots \end{pmatrix} \\
&= \begin{pmatrix} \sigma_{11}/(1-\beta_{11}^2) & \sigma_{12}/(1-\beta_{11}\beta_{22}) \\ \sigma_{21}/(1-\beta_{11}\beta_{22}) & \sigma_{22}/(1-\beta_{22}^2) \end{pmatrix}.
\end{aligned}$$

Note that $1 + \beta_{11}^2 + \beta_{11}^4 + \cdots = 1/(1-\beta_{11}^2)$.

For the lag l covariance, according to Equation 3.4.16, we have

$$\begin{aligned}
\mathbf{\Gamma}_l &= \boldsymbol{\beta}^l\mathbf{\Gamma}_0 \\
&= \begin{pmatrix} \beta_{11}^l & 0 \\ 0 & \beta_{22}^l \end{pmatrix}\begin{pmatrix} \sigma_{11}/(1-\beta_{11}^2) & \sigma_{12}/(1-\beta_{11}\beta_{22}) \\ \sigma_{21}/(1-\beta_{11}\beta_{22}) & \sigma_{22}/(1-\beta_{22}^2) \end{pmatrix} \qquad (3.4.18) \\
&= \begin{pmatrix} \sigma_{11}\beta_{11}^l/(1-\beta_{11}^2) & \sigma_{12}\beta_{11}^2/(1-\beta_{11}\beta_{22}) \\ \sigma_{21}\beta_{22}^l/(1-\beta_{11}\beta_{22}) & \sigma_{22}\beta_{22}^l/(1-\beta_{22}^2) \end{pmatrix}, \qquad (3.4.19)
\end{aligned}$$

where $\sigma_{11}\beta_{11}^l/(1-\beta_{11}^2)$ is the covariance of $\{R_1(t)\}$ with lag l; similarly, $\sigma_{22}\beta_{22}^l/(1-\beta_{22}^2)$ is the covariance of $\{R_2(t)\}$ with lag l; $\sigma_{12}\beta_{11}^2/(1-\beta_{11}\beta_{22})$ is covariance between $\{R_1(t)\}$ and $\{R_2(t)\}$ with lag l; and $\sigma_{12}\beta_{22}^2/(1-\beta_{11}\beta_{22})$ is covariance between $\{R_2(t)\}$ and $\{R_1(t)\}$ with lag l.

To see the linear dependence of $R_1(t_i)$ on $R_2(t_{i+l})$, it is convenient to compute the lag l cross-correlation matrix of $\mathbf{r}(t_i)$, which is defined as

$$\boldsymbol{\rho}_l \equiv [\rho_{ij}(0)] = \mathbf{D}^{-1}\mathbf{\Gamma}_l\mathbf{D}^{-1}, \tag{3.4.20}$$

where $\mathbf{D}$ is the diagonal matrix of standard deviations of the individual series $\mathbf{R}_k(t_i)$, $k = 1, 2$. In other words, $\mathbf{D} = \{\sqrt{\Gamma_{11}(0)}, \sqrt{\Gamma_{22}(0)}\}$, where $\sqrt{\Gamma_{11}(0)}$ and $\sqrt{\Gamma_{22}(0)}$ are the off-diagonal element of matrix $\mathbf{\Gamma}_0$.

Based on Equation 3.4.18, the explicit expression of the off-diagonal elements of $\boldsymbol{\rho}_l$, which measures the cross-correlation of $\mathbf{R}(t_i)$ with lag l, is given by:

$$\begin{aligned}
\rho_{ij}(l) &= \frac{\Gamma_{ij}(l)}{\sqrt{\Gamma_{11}(0)\Gamma_{22}(0)}} \\
&= \frac{\text{Cov}(R_1(t_i), R_2(t_{i-l}))}{\sqrt{\text{Var}(R_1(t_i))}\sqrt{\text{Var}(R_2(t_{i-l}))}} \\
&= \frac{\sigma_{12}\beta_{ii}^l/(1-\beta_{11}\beta_{22})}{\sqrt{\frac{\sigma_{11}}{1-\beta_{11}^2}}\sqrt{\frac{\sigma_{22}}{1-\beta_{22}^2}}} \\
&= \frac{\sqrt{1-\beta_{11}^2}\sqrt{1-\beta_{22}^2}}{(1-\beta_{11}\beta_{22})}\frac{\sigma_{12}}{\sqrt{\sigma_{11}\sigma_{22}}}\beta_{ii}^l \\
&= A\beta_{ii}^l, \qquad (3.4.21)
\end{aligned}$$

where $A = (\sqrt{1-\beta_{11}^2}\sqrt{1-\beta_{22}^2}/(1-\beta_{11}\beta_{22}))(\sigma_{12}/\sqrt{\sigma_{11}\sigma_{22}})$, $i \neq j$, $i, j = 1, 2$.

While the autocorrelation of $\mathbf{R}(t_i)$ with lag l can be found:

$$\begin{aligned}
\rho_{ii}(l) &= \frac{\Gamma_{ii}(l)}{\sqrt{\Gamma_{11}(0)\Gamma_{22}(0)}} \\
&= \frac{\text{Cov}(R_1(t_i), R_2(t_{i-l}))}{\sqrt{\text{Var}(R_1(t_i))}\sqrt{\text{Var}(R_2(t_i))}} \\
&= \frac{\sigma_{ii}\beta_{ii}^l/(1-\beta_{11}^2)}{\sqrt{\sigma_{11}/(1-\beta_{11}^2)}\sqrt{\sigma_{22}/(1-\beta_{22}^2)}} \\
&= \frac{\sqrt{1-\beta_{11}^2}\sqrt{1-\beta_{22}^2}}{(1-\beta_{ii}^2)}\frac{\sigma_{12}}{\sqrt{\sigma_{11}\sigma_{22}}}\beta_{ii}^l \\
&= B\beta_{ii}^l, \qquad (3.4.22)
\end{aligned}$$

where $B = (\sqrt{1-\beta_{11}^2}\sqrt{1-\beta_{22}^2}/(1-\beta_{11}^2))(\sigma_{12}/\sqrt{\sigma_{11}\sigma_{22}})$, $i = 1, 2$.

When $l > 0$, the correlation coefficient in Equation 3.4.21 measures the linear dependence between $R_1(t_i)$ and $R_2(t_{i-l})$, which occured prior to time t_i. Generally, if $\rho_{ij}(l) \neq 0$ and $l > 0$, it means that the series $R_j(t_i)$ 'leads' the series $R_i(t_i)$ at lag l. Analogously, $\rho_{ji}(l)$ measures the linear dependence between $R_j(t_i)$ and $R_i(t_{i-l})$, and we say that the series $R_i(t_i)$ 'leads' the series $R_j(t_i)$ at lag l if $\rho_{ji}(l) \neq 0$ and $l > 0$. Equation 3.4.22 shows that the diagonal element $\rho_{ii}(l)$ is simply the lag l autocorrelation coefficient of $r_i(t_i)$.

Some important properties of the correlation matrix $\boldsymbol{\rho}(l)$ with $l > 0$ can be summarized as follows. First, the cross-correlation at lag l, that is, the correlation coefficient $\rho_{ij}(l)$ decreases as l increases, behaving in this respect, as the auto-correlation at lag l, $\rho_{ii}(l)$. This is due to the fact that the first derivative of $\rho_{ij}(l)$ with respect to l, where $i, j = 1, 2$, is negative. For the first derivative of cross-correlation with lag l ($\rho_{ij}(l)$ in Equation 3.4.21), we obtain

$$\frac{\partial \rho_{ij}(l)}{\partial l} = A\beta_{ii}^l \ln(\beta_{ii}), \qquad (3.4.23)$$

where $i, j = 1, 2$.

For the first derivative of auto-correlation with lag l, $\rho_{ii}(l)$ (in Equation 3.4.22), we have

$$\frac{\partial \rho_{ij}(l)}{\partial l} = B\beta_{ii}^{l}\ln(\beta_{ii}), \tag{3.4.24}$$

where $i = 1, 2$.

Given the assumption of $0 < \beta_{ii} < 1$, we have $0 < \beta_{ii}^{l} < 1$ and $\ln(\beta_{ii}) < 0$. Since A and B are positive, the first derivative of cross- and auto-correlation with lag l (in Equation 3.4.23 and 3.4.24) are negative. This leads to the cross-correlation between $R_1(t_i)$ and $R_2(t_{i-l})$ and the auto-correlation between $R_k(t_i)$ and $R_k(t_{i-l})$ as decreasing function of lag l. This phenomenon has been widely observed in the empirical financial data, including seven Italian banks from *Borsa di Milano* in case of ultra-high-frequency data, see Chapter 1 for more details.

Second, in general, $\rho_{12}(l) \neq \rho_{21}(l)$, or as a matter of fact, $\beta_{11} \neq \beta_{22}$. If $\beta_{11} = \beta_{22}$, the logreturn model becomes very specific in the sense that two time series depend on their past with the same coefficient. In order to keep the model in more general way, we would rather to assume that $\beta_{11} \neq \beta_{22}$. Furthermore, the two correlation coefficients $\rho_{12}(l)$ and $\rho_{21}(l)$ measure different linear relationships between $\{R_1(t)\}$ and $\{R_2(t)\}$ with lag l. Therefore, $\mathbf{\Gamma}_l$ and $\boldsymbol{\rho}_l$ are in general not symmetric.

Thirdly, from $\text{Cov}(R_1(t_i), R_2(t_{i-l})) = \text{Cov}(R_2(t_{i-l}), R_1(t_i))$ and by the weak stationary assumption, we get

$$\text{Cov}(R_2(t_{i-l}), R_1(t_i)) = \text{Cov}(R_2(t_i), R_1(t_{i+l})) = \text{Cov}(R_2(t_i), R_1(t_{i-(-l)}))$$

Consequently, $\Gamma_{12}(l) = \Gamma_{21}(-l)$. Since $\Gamma_{ij}(-l)$ is the (j, i)th element of the matrix $\mathbf{\Gamma}_{-l}$, we have $\mathbf{\Gamma}_l = \mathbf{\Gamma}'_{-l}$ and with the same reason $\boldsymbol{\rho}_l = \boldsymbol{\rho}'_{-l}$. Nevertheless, unlike the univariate case, $\rho_l \neq \rho_{-l}$ for a general vector time series when $l > 0$. Because $\boldsymbol{\rho}_l = \boldsymbol{\rho}'_{-l}$ it suffices in practice to consider the cross-correlation matrix $\boldsymbol{\rho}_l$ for $l \geq 0$.

3.4.7 Independent Bi-AR(1)

The simplest case for Bi-AR(1) model is that two time series $\{R_1(t)\}$ and $\{R_2(t)\}$ are just autocorrelated per se. So there is no relationship between the underlying two time series. Basically, the model can be described separetly. Therefore, Bi-AR(1) can be rewritten as

$$\begin{cases} R_1(t_i) = \beta_{01} + \beta_{11}R_1(t_{i-1}) + \epsilon_1(t_i), \\ R_2(t_i) = \beta_{02} + \beta_{22}R_2(t_{i-1}) + \epsilon_2(t_i), \end{cases} \tag{3.4.25}$$

where $\beta_{kk} \neq 0$ and ϵ_k with mean zero and variance $\sigma_{kk} \neq 0$, $k = 1, 2$. And

$$\text{Cov}(\epsilon_1(t_i), \epsilon_2(t_i)) = \sigma_{12} = \sigma_{21}0.$$

In this case, the underlying two time series $\{R_1(t)\}$ and $\{R_2(t)\}$ are mutually independent. The cross-correlation between these two logreturns $\{R_1(t)\}$ and $\{R_2(t)\}$ is therefore zero. On the other hand, the auto-correlation of $\{R_k(t)\}$ can refer to Equation 3.4.22, to see the velocity of change of the auto-correlation coefficient $\rho_{kk}(l)$ with respect to the time lag l, the first derivative of $\rho_{kk}(l)$ with respect to l becomes the first place and displays in a standard way. The form keep the same as Equation 3.4.24. Consequently, the same conclusion will be drawed that the auto-correlation of $\{R_k(t)\}$ is a decrease function of the time

lag l, in particular, the speed of decay equals to β_{kk}^l. The smaller the β_{kk}, the sharper of the curve approach to horizontal line. This means the weaker dependence on its own past, the smaller auto-correlation be obtained. Second, the longer time lag l, that is, the greater l, the nearer of the curve to the horizontal line. The interpretation of this fact can be stated that the time serie $\{R_1(t)\}$ has quite strong linear dependence on its most rescent past, but approachs zero when concerns on its long past. Note that the asssumption of $0 < \beta_{kk} < 1$ to ensure the weak stationary of $\{R_1(t)\}$ here still need to be held.

3.5 Model comparison by Monte Carlo simulation methods

Simulation studies play a key role in nowadays scientific research. When uncertainty becomes normal, it is natural to treat the substances subject to random fluctuations. These substances can not be described by an exact mathematical rule but only through the use of probability statements, namely stochastic subject which matters of statistics. So whenever one chooses to simulate a stochastic subject, it will be based on probability distribution. From a statistical point of view, stochastic subject can be regarded as a function of random variables.

In this section, we implement the above five modeling framework by Monte Carlo simulation methods. As addressed before, generally, in the study of complicated stochastic models, simulations of model play an important part for three reasons. They enable one to: (1) understand the basic statistical characteristics of the model; (2) generate data in order to evaluate estimation techniques for the model parameters; and (3) derive predictions to test the performance of the model. Here we take the first advantage of the simulation study of model, that is, to understand the basic characteristics of the models.

This simulation study will direct to answer the following two questions: first, which modeling framework of logreturns fit the UHF data best on the criterion of Epps curve? Second, does common factor intensity model affect Epps effect?

To answer the second question, we present two frameworks for time generating process: one is based on two independent doubly stochastic Poisson process (DSPP), the other is concerning on common factor model which is also based on the doubly stochastic Poisson process. The only difference between these two frameworks is that the common component of the intensity is introduced in the second one. The aim of this comparison is to show whether the underlying intensity affects the correlation between logreturns.

3.5.1 Independent DSPP

As introduced in Section 3.2, each intensity $\lambda_t^{(i)}$ is assumed to be a shot noise process (Cox & Isham 1980, 1986), an explicit expression is shown as follows:

$$\lambda_t^{(i)} = \lambda_0^{(i)} e^{(-\kappa_i t)} + \sum_{j=0}^{N_t'^{(i)}} X_j^{(i)} \cdot e^{(-\kappa_i (t - \tau_j^{(i)}))}, \quad t \geq 0, \tag{3.5.1}$$

where:

$\lambda_0^{(i)}$ the initial value of the process $\lambda_t^{(i)}$ is Gamma distributed with shape parameter ν_i/κ_i and rate parameter γ_i;

$X_j^{(i)}$ the jump size has independent and identical exponential distribution with mean $1/\gamma_i$;

$\tau_j^{(i)}$ the time at which jump happens, which is assumed as Poissson process with constant intensity ν_i.

Notice that two intensities are independent with $i = 1, 2$. Figure 3.7 illustrates the simulation result based on the parameters $\nu_1 = 0.2$, $\kappa_1 = 0.2$, $\gamma_1 = 0.2$, $\nu_2 = 0.15$, $\kappa_2 = 0.15$, $\gamma_2 = 0.3$, under the DSPP model (Equation 4.4.2). The upper two figures show the intensity process $\tilde{\lambda}_t^1$ (on the left), $\tilde{\lambda}_t^2$ (on the right); the bottom two figures show the corresponding event times based on the obove intensities (see Figure 3.7).

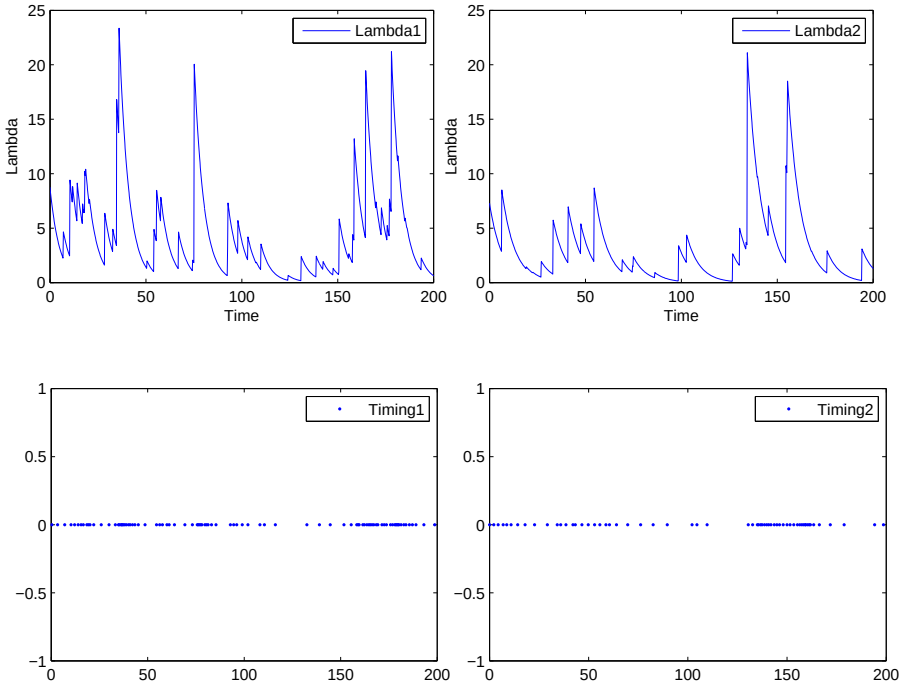

Figure 3.7: Two independent intensities and corresponding event times. The intensity $\tilde{\lambda}_t^1$ is shown on the left top; the other intensity $\tilde{\lambda}_t^2$ is ploted on the right top; the corresponding event times based on the intensity $\tilde{\lambda}_t^1$ is shown on the bottom left; the event times based on the intensity $\tilde{\lambda}_t^2$ is presented on the right bottom. Results are based on the parameters $\nu_1 = 0.2$, $\kappa_1 = 0.2$, $\gamma_1 = 0.2$, $\nu_2 = 0.15$, $\kappa_2 = 0.15$, $\gamma_2 = 0.3$ (Equation 4.4.2).

3.5.2 Common factor model based on DSPP

Consider a simple factor model for the intensity. Basically, the intensities $\tilde{\lambda}_t^{01}$ and $\tilde{\lambda}_t^{02}$ are assumed as a linear combination of the components $\lambda_t^{(0)}$, $\lambda_t^{(1)}$, $\lambda_t^{(2)}$. The form is given by

$$\begin{cases} \tilde{\lambda}_t^{01} = \lambda_t^{(1)} + a_1 \lambda_t^{(0)}, \\ \tilde{\lambda}_t^{02} = \lambda_t^{(2)} + a_2 \lambda_t^{(0)}. \end{cases} \tag{3.5.2}$$

Figure 3.8 illustrates the common factor model of intensity and the resulting point process (or event times) after the linear combination of two components. Setting model parameters $\nu_0 = 0.1$, $\kappa_0 = 0.1$, $\gamma_0 = 0.2$, $\nu_1 = 0.2$, $\kappa_1 = 0.2$, $\gamma_1 = 0.2$, $\nu_2 = 0.15$, $\kappa_2 = 0.15$, $\gamma_2 = 0.3$, $a_1 = a_2 = 1$. Each component follows DSPP process (Equation

4.4.2), where the rate of the Poisson process is the shot-noise-driven process, more details refer to Section 3.2. In Figure 3.8, the common component $\lambda_t^{(0)}$ is ploted on the top, the specific component $\lambda_t^{(1)}$ is on the left middle and $\lambda_t^{(2)}$ is on the right middle. Whereas the left bottom shows the event times which is the realization of the underlying intensity $\tilde{\lambda}_t^{01}$, a linear combination of $\lambda_t^{(0)}$ (on the top) and $\lambda_t^{(1)}$ (on the left middle). Similarly, the right bottom shows the event times which is the realization of the underlying intensity $\tilde{\lambda}_t^{02}$, a linear combination of $\lambda_t^{(0)}$ (on the top) and $\lambda_t^{(2)}$ (on the right middle).

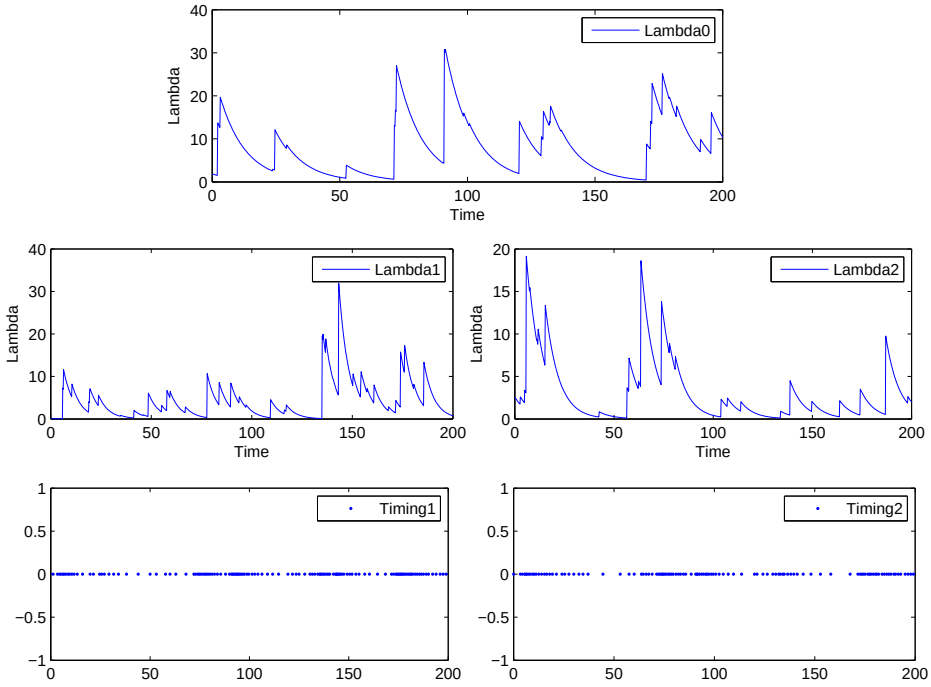

Figure 3.8: Common factor model of the intensities and corresponding event times. The common component $\lambda_t^{(0)}$ is shown on the top; the specific component $\lambda_t^{(1)}$ is on the left middle; the other specific component $\lambda_t^{(2)}$ is ploted on the right middle; the event times based on the intensity $\tilde{\lambda}_t^{01}$ is shown on the bottom left; the event times based on the intensity $\tilde{\lambda}_t^{02}$ is presented on the right bottom. Results are based on the parameters $\nu_0 = 0.1$, $\kappa_0 = 0.1$, $\gamma_0 = 0.2$, $\nu_1 = 0.2$, $\kappa_1 = 0.2$, $\gamma_1 = 0.2$, $\nu_2 = 0.15$, $\kappa_2 = 0.15$, $\gamma_2 = 0.3$ (Equation 3.5.2 and Equation 4.4.2).

To answer the first question, we compare the correlation coefficient function of sampling interval for these five modeling frameworks of logreturns. The simulation result is shown on Figure. Note that the same timing (common factor model) for these five models.

In the following, we implement five modeling frameworks as described in Section 3.4 by Monte Carlo simulation. For each framework, we compare two time generating models: independent DSPP and common factor model.

3.5.3 Independent Wiener process

For Independent Wiener process model, the simulation work will be based on Equation 3.4.2 as described in the previous section. Assume that $\xi_1(t_i)$ and $\xi_2(t_i)$ are normally distributed with mean zero and variance $t_i - t_{i-1}$. Moveover, these two time series $\xi_1(t_i)$ and $\xi_2(t_i)$ are independent from each other given the time $\{t_i\}$, where $\{t_i\}$ are the pooled times (see Section 3.4.1).

Therefore, the pooled times $\{t_i\}$ plays an important role in the logreturn generating process based on independent Wiener model. Accounting for the times, it is necessary to recall the time generating process. As introduced in Section 3.2, intensity-based common factor model for time generating process whose essential block is presenting one common intensity. It is then interesting that the common intensity whether or not introduce some sort of correlation into the logreturn generating process when it is based on the independent Wiener process model. Thereby, the Comparison of two models of time generating process, independent DSPP (in Section 3.5.1) and common factor model (in Section 3.5.2), provides full knowledge about intereaction between model of logreturn and of time generating.

Model based on independent DSPP and independent Wiener process

Firstly, we consider the time generating process based on the independent DSPP (in Section 3.5.1), the simulated times are shown on the bottom two figures in Figure 3.7. Then we pool these two sequences of times orderly as shown in Figure 3.6. Now the remaining task is to generate two logreturn processes, which will be based on the same underlying pooled times. In principle, to simulate logreturns which are independent wiener processes is based on Equation 3.4.2 and Equation 3.4.3.

After generate bivariate logreturn processes, as stated before, we calculate the corresponding log-prices for the pooled times. The data matching work will be proceeded so that the number of event times for each time serie coincide with the original time generating model. Finally, previous-tick interpolation should be applied for calculating logreturns based on different sampling scales given the asynchronous data, this is because the data matching work restore the data to the inherent asynchronicity.

Figure 3.9 shows six scatterplots of logreturns $\{R_1(t)\}$ and $\{R_2(t)\}$ based on different sampling scales. Note that the grid of previous-tick interpolation as shown in Figure 3.9 also refers to sampling scale. All the scatterplots of Figure 3.9 show no significant linear dependence between two time series $\{R_1(t)\}$ and $\{R_2(t)\}$. The same effect is better viewed from the top figure of Figure 3.11 where the correlation coefficients are plotted as a function of the sampling scale and where the same data are used in Figure 3.9 and Figure 3.11 (top).

Note that the grid, 0.2, 1, etc, shows the sampling scale with respect to the simulated data, which is different from the physical time scale, 10 seconds, 1 mimute, for instance. Of course, the simulated data can be converted into physical time scale, but our concern here is about the linear relationship with respect to sampling scale. No matter which scale is utilized, the important thing is to evaluate the linear relationship as a function of its adaptive scale. The same for the following scatterplots.

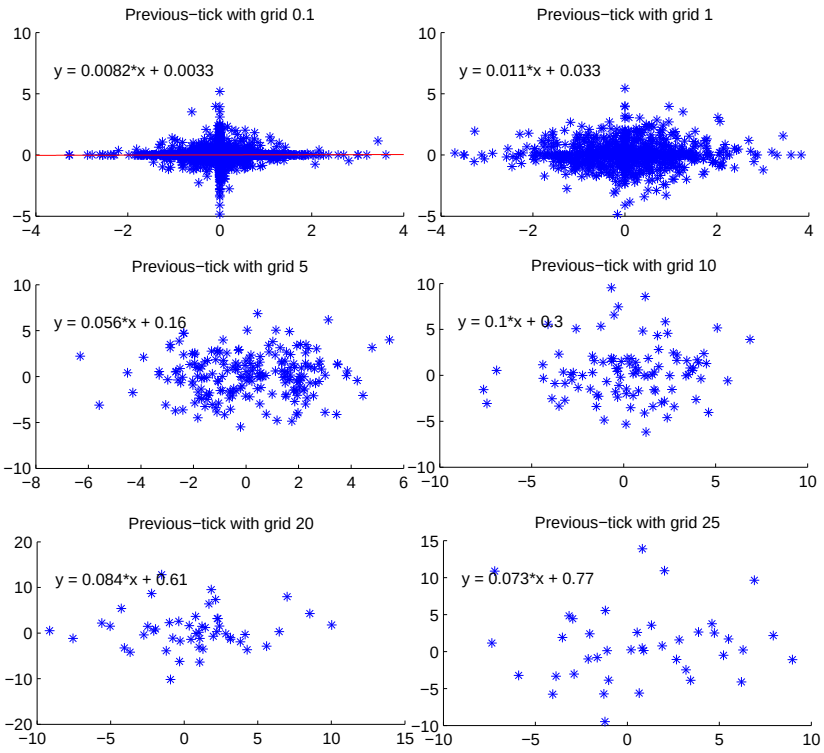

Figure 3.9: Some scatterplots of logreturns $\{R_1(t)\}$ and $\{R_2(t)\}$ based on different grids of previous-tick interpolation. The left top figure shows the scatterplot based on the grid with 0.1 (previous-tick); the right top figure shows the scatterplot with grid 1 (previous-tick). The left middle figure based on the grid with 5 (previous-tick); the right middle figure based on the grid with 10. The left bottom shows the scatterplot with grid 20 (previous-tick); the right bottom shows the scatterplot based on the grid 25 (previous-tick). The plots also provide the linear regression equation as shown on the left shoulder of each figure. This result is based on independent Wiener processes for the logreturn generating process and independent DSPP for time generating.

Model based on a common factor and independent Wiener process

Secondly, in case that the time generating process based on the common factor model (see Section 3.5.2), graphically, the simulated times are presented on the bottom figures in Figure 3.8. Given the underlying two sequences of the simulated times, in practice, we pool them orderly as shown in Figure 3.6. Accordingly, we go on the simulation work about logreturn processes which are based on the pooled times. Since we generate some false logreturns (a trade-off with asynchronization), perhaps a lot, time matching work should be carried out then.

Now, the inherent asynchronicity for two time series would turn out again. However, the asynchronicity for this time is not so disagreeable, indeed, it is closer to the real data which are asynchronous originally, this makes the simulation work meaningful. Moreover, this procedure keeps the number of logreturns in line with the time generating process based on the intensity. To show the correlation (linear dependence) between two series, scatterplot then provides the visual strength of relationship between two variables, and the previous-tick interpolation will be applied (see Figure 1.9) such that the log-prices are equally spaced. This is because the data input for the scatterplot are two time series with equal (i.e., homogeneous) spacing between ticks.

The resulting linear dependence between two time series $\{R_1(t)\}$ and $\{R_2(t)\}$ based on different sampling scales is presented with scatterplots in Figure 3.10. Note that the grid of previous-tick interpolation as shown in Figure 3.10 is the same as sampling scale. As seen from Figure 3.10, there is no significant linear relationship between two time series $\{R_1(t)\}$ and $\{R_2(t)\}$. The slope of each linear regression as shown on the left shoulder of each panel is oscillating around zero. The same effect is better viewed in the bottom figure of Figure 3.11 where correlation coefficients are plotted as a function of the sampling scale. Note that the same data are used in Figure 3.10 and Figure 3.11 (bottom).

Correlation versus sampling scale

To better gauge the linear dependence between two time series, the correlation coefficient measurement should be taken on the first place. When the standard Pearson correlation coefficient is calculated between two time series, the assumption is that this quantity is evaluated for equally spaced data. For the ultra-high-frequency financial time series, however, this method cannot be applied directly, since they are irregularly spaced in time and asynchronous for any two time series. What matters here is the simulation progress which is based on the characteristics of the ultra-high-frequency data, the generated logreturn processes thereby possess the feature of irregular spaced in time and non-synchronicity when two time series are involved. Section 3.3 provides the data generating procedure about irregularly time-spaced data.

Given the simulated logreturns, the dependence between these two series varying with the sampling scale is shown in the scatterplots, Figure 3.9 and Figure 3.10. From the shape of the scatterplots in Figure 3.9 and Figure 3.10, it confirms that the involved two time series have linear relationship but not other kind of relationship, such as quadratic, cubic, non-linear etc. The limit of the scatterplot, that measuring in one fix time scale, goes against understanding the linear dependence varying with different sampling scale in a complex way. We thus calculate correlations literally as a function of sampling scale for

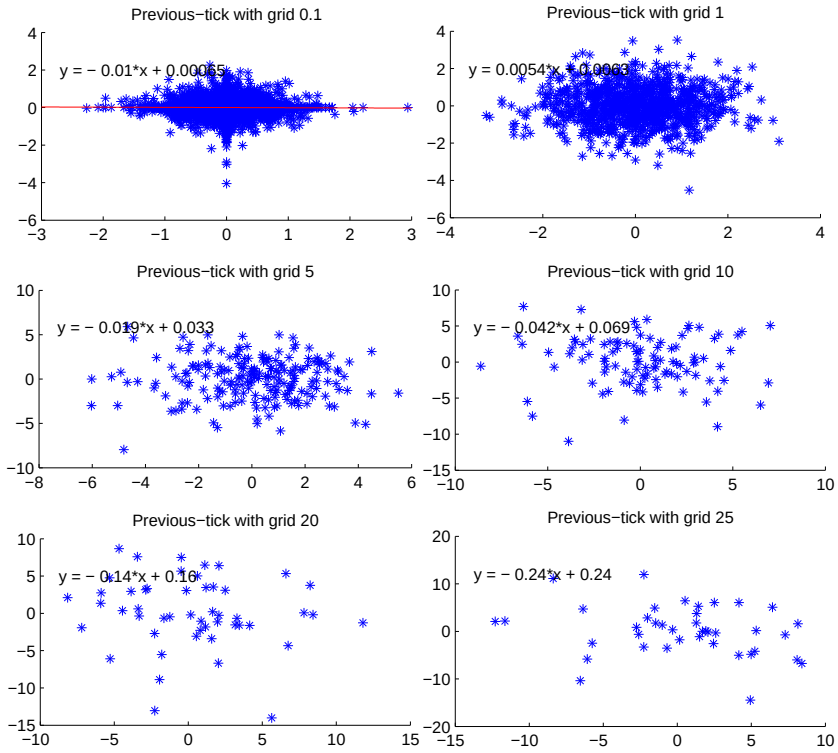

Figure 3.10: Some scatterplots of logreturns $\{R_1(t)\}$ and $\{R_2(t)\}$ based on different grids of previous-tick interpolation. The left top figure shows the scatterplot based on the grid 0.1 (previous-tick); the right top figure shows the scatterplot with grid 1 (previous-tick). The left middle figure is based on the grid 5 (previous-tick); the right middle figure is based on the grid 10 (previous-tick). The left bottom shows the scatterplot with grid 20 (previous-tick); the right bottom shows the scatterplot based on the grid 20 (previous-tick). Each panel also provides the linear regression equation as shown on the left shoulder. This result is based on the concurrently correlated Wiener process for the logreturn generating process and the common factor intensity model for time generating.

two time series in an effort to better understand the timing and level of change that can occur.

Results of correlation coefficient calculation versus sampling scale are shown in Figure 3.11. Two panels display the correlation coefficient of logreturns based on the underlying two different time generating model. The upper panel shows the result based on independent DSPP for time generating process and independent Wiener process for logreturn simulation. The bottom panel shows the result based on common factor model for time generating process and independent Wiener process for logreturn simulation. Note that the sampling scale refers to previous-tick sampling interval.

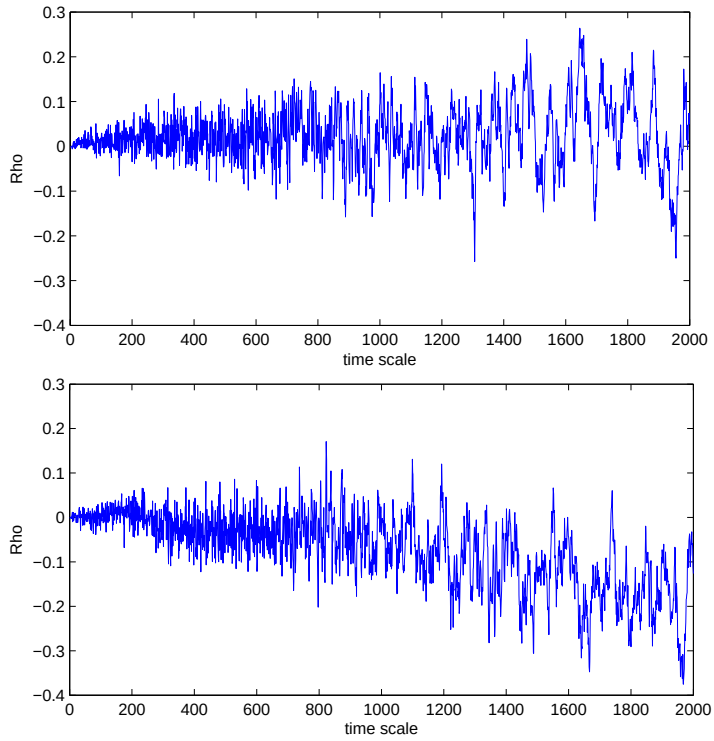

Figure 3.11: The linear correlation coefficients evaluated as a function of the sampling scale. The time generating process is based on the independent DSPP and the logreturn process is according to the independent Wiener process (top panel). The time generating process is based on the common factor model and the logreturn process is based on the independent Wiener process (bottom panel).

From Figure 3.11, the correlation coefficient of logreturns, which is based on independent Wiener process, is oscilating around zero for both two time generating processes. Another common feature for both two figures as shown in Figure 3.11 is that the range of oscilation is increasing as sampling scale increases. The explaination for this phenomenon can be the less data involved in correlation calcualtion when the sampling scale become larger. On the other hand, the difference in the behaviors of the two figures is somehow subtle but still worth to point out: the slop of the upper figure (see Figure 3.11) is zero while the bottom figure (see Figure 3.11) inclines to negative.

3.5.4 Concurrently correlated Wiener process

As for the correlated Wiener process model, the simulation work will be based on Equation 3.4.2 as described in the previous section. Assume that $\xi_1(t_i)$ and $\xi_2(t_i)$ are normally distributed with mean zero and variance $t_i - t_{i-1}$. Moveover, these two time series $\xi_1(t_i)$ and $\xi_2(t_i)$ are concurrently correlated with covariance $\sigma_{12}(t_i - t_{i-1})$, more details refer to Section 3.4.4, where $\{t_i\}$ are the pooled times. the positive definiteness of Σ (the covariance matrix of $\xi_1(t_i)$ and $\xi_2(t_i)$) can be ensured by setting σ_{12} as a positive real number. More precisely, σ_{12} should be greater than zero and less than one, that is, $0 < \sigma_{12} < 1$. Because the correlation coefficient (rho) in this case is equal to σ_{12} which should be greater than -1 and less than 1. Here we assume $\sigma_{12} = 0.9$ for the simulation, so the covariance matrix $\Sigma = (t_i - t_{i-1})\binom{1 \quad 0.9}{0.9 \quad 1}$.

Model based on independent DSPP and concurrently correlated Wiener process

We propose two methods for time generating as describe before. Firstly, we consider the time generating process based on the independent DSPP (in Section 3.5.1), the simulated times are shown on the bottom two figures in Figure 3.7. Then we pool these two sequences of times orderly as shown in Figure 3.6. This leads the way for simulating logreturn processes, which require the same length of each component of the bivariate logreturn processes.

After generate bivariate logreturn processes, data matching work is proceeded so that the number of event times for each time serie is the same as that of original time generating model based on the intensity process. Then, previous-tick interpolation should be applied given the asynchronous data, this is because the data matching work restore the data to the inherent asynchronicity.

Figure 3.12 shows some scatterplots of logreturns $\{R_1(t)\}$ and $\{R_2(t)\}$ based on different sampling scales. Note that the grid of the previous-tick interpolation also refers to sampling scale. Several results can be drawn from this figure. First, the scatterplots present linear relationship between two time series $\{R_1(t)\}$ and $\{R_2(t)\}$. Second, when compare the slope of linear regression equation as shown on the left shoulder of each panel with increasing sampling scale, it is easy to find that linear dependence between two logreturns $\{R_1(t)\}$ and $\{R_2(t)\}$ increase as sampling scale increase, then remain at quite stable level. In Figure 3.12, the slope is almost zero when sampling interval equals to 0.2 (left top panel), then with dramatic increase (0.73) when sampling interval is 5 (left middle panel), while no big difference between sampling interval 20 (left bottom panel) and 30 (right bottom panel), 1.1 and 1 respectively. The same effect is better viewed on the top figure of Figure 3.14 where the correlation coefficients are plotted as the function of the sampling scale.

Third, the slope of linear regression equation is not the same as the correlation coefficient generally, $\rho = k \cdot \sigma_{R_1}/\sigma_{R_2}$, where k is the slope of the linear regression and σ_{R_1} is the variance of logreturn $\mathbf{R}_1$, σ_{R_2} is the variance of logreturn $\mathbf{R}_2$. If two time series $\{R_1(t)\}$ and $\{R_2(t)\}$ have the same variance, that is, $\sigma_{R_1} = \sigma_{R_2}$, then the slope is equivalent to the correlation coefficient. In our case, we set parameters $\sigma_{R_1} = \sigma_{R_2} = 1$, so the slope of the linear regression can be treated as the estimator of the correlation coefficient. Hence, 0.04 as the slope of the linear regression equation on the left top figure (in Figure 3.12) means $\hat{\rho} = 0.04$ when the sampling scale is 0.2; While $\hat{\rho} = 0.92$ if the sampling scale is 15 (right

middle panel). It is interesting that $\hat{\rho} = 1.1 > 1$ in case of grid 20 (left bottom panel), one explaination is that it may over-estimate. However, it is important to be clear in mind that there is only one true correlation coefficient of the model, that is, $\rho = 0.9$, which do not depend on the sampling scale.

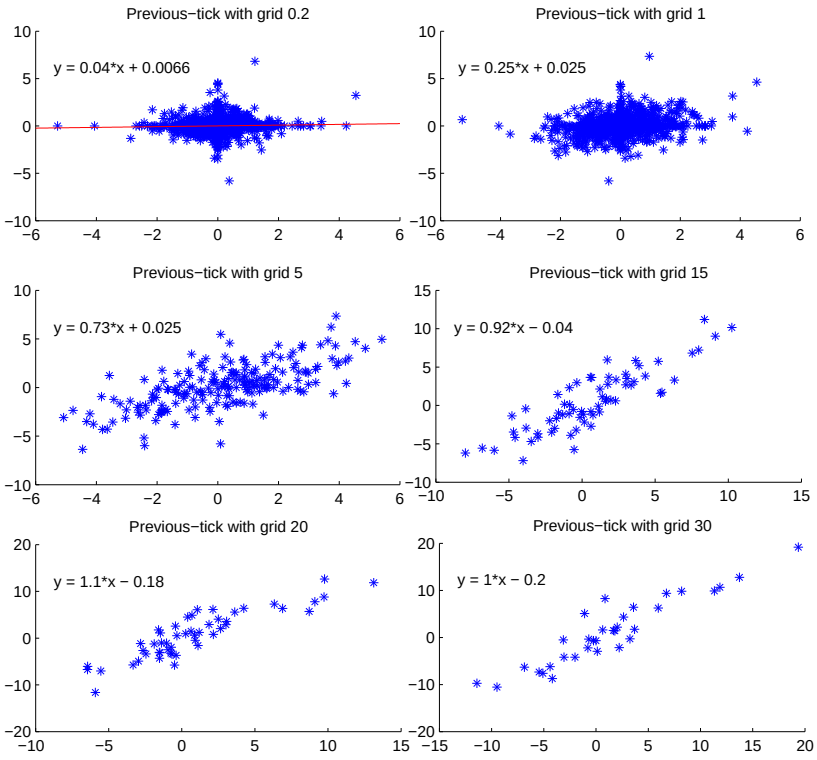

Figure 3.12: Some scatterplots of logreturns $\{R_1(t)\}$ and $\{R_2(t)\}$ based on different sampling scales. The previous-tick interpolation is applied. The left top figure shows the scatterplot based on the grid 0.2 (previous-tick); the right top figure shows the scatterplot with grid 1 (previous-tick). The left middle figure is based on the grid 5 (previous-tick); the right middle figure is based on the grid 15 (previous-tick). The left bottom shows the scatterplot with grid 20 (previous-tick); the right bottom shows the scatterplot based on the grid 30(previous-tick). The linear regression equation is provided on the left shoulder of each plot. This result is based on concurrently correlated Wiener process for the logreturn generating processes and independent DSPP for time generating process.

Model based on a common factor and concurrently correlated Wiener process

Secondly, in case that the time generating process based on the common factor model (see Section 3.5.2), graphically, the simulated times are presented on the bottom figures in Figure 3.8. Given the two sequences of the simulated times, in practice, we pool them orderly as shown in Figure 3.6. Then, we go on the simulation work for logreturn processes which are based on the pooled times. Since we generate some false logreturns (a trade-off with asynchronization), time matching work should be carried out then.

Now, the inherent asynchronicity for two time series turn out again. However, the asynchronicity for this time is not so disagreeable as before, indeed, it is closer to the real data which are asynchronous. Moreover, the number of logreturns is in line with the time generating process based on the intensity. In order to present scatterplot, which is a way to show the correlation between two series, the previous-tick interpolation will be applied (see Figure 1.9). This is because the scatterplot dealing with bivariate vectors that have the same length.

Figure 3.13 provides six scatterplots of logreturns $\{R_1(t)\}$ and $\{R_2(t)\}$ according to different sampling scales. Note that the grid of the previous-tick interpolation is the same as sampling scale. First of all, the plots present strong linear dependence between two time series $\{R_1(t)\}$ and $\{R_2(t)\}$. Second, the correlation increase as sampling increase then remain at quite stable level, about 0.9 as shown in Figure 3.13. The same phenomenon is better viewed on the bottom figure of Figure 3.14. Finally, the slope of linear regression equation, as shown on the left shoulder of each panel, can be different from the correlation coefficient in general. But in this modeling framework, the slope of linear regression equals to the correlation coefficient. This is due to the underlying two time series do share the same variance $(t_i - t_{i-1})$ by setting parameters $\sigma_{11} = \sigma_{22} = 1$. Generally, $\rho = k \cdot \sigma_{R_1}/\sigma_{R_2}$, where k is the slope of the linear regression and σ_{R_1} is the variance of logreturn $\mathbf{R}_1$, σ_{R_2} is the variance of logreturn $\mathbf{R}_2$.

In some extent, the slope k can be treated as an estimator of correlation coefficient ρ, denoted as $\hat{\rho}$. If this is the case, then $\hat{\rho} = 0.41$ when sampling interval is 0.2 (left top figure); $\hat{\rho} = 0.74$ in case sampling interval is 1 (right top figure); and $\hat{\rho} = 0.9$ for sampling interval 5 (left middle figure); $\hat{\rho} = 0.95$ for sampling interval 15 (right middle figure). However, it is important to be clear in mind that one unique correlation coefficient is assumed in the model, that is, $\rho = 0.9$.

By comparing two figures presented in Figure 3.13 and Figure 3.12, where the model based on different time generating process, we find that the model based on common factor intensity model (Figure 3.13) better present Epps effect and the stable level (0.92) is much closer to the true one than the model based on independent DSPP. A more complex view refers to correlation coefficient as the function of sampling interval (see Figure 3.14), the same effect (Epps effect) is obtained but the bottom figure of Figure 3.14 displays much more dramatic increase when sampling interval enter 5 than that of the top figure.

Correlation versus sampling scale

To better gauge the linear dependence between two time series, the correlation coefficient measurement will be used. When the standard Pearson correlation coefficient is calculated between two time series, the assumption is that this quantity is evaluated for equally spaced data. For the case of ultra-high frequency financial time series this is invalide, since they are irregularly spaced in time and asynchronous for any two time series. What matters here is the simulation progress is based on the character of the ultra-high frequency data, the generated logreturn processes thereby possess the feature of irregular spaced in time and non-synchronicity when two time series are involved. This issu is addressed in the previous part concerning on time generating context.

Given the simulated logreturns, the dependence between these two series varying with the sampling scale is shown in the scatterplots. From the shape of the scatterplots, it

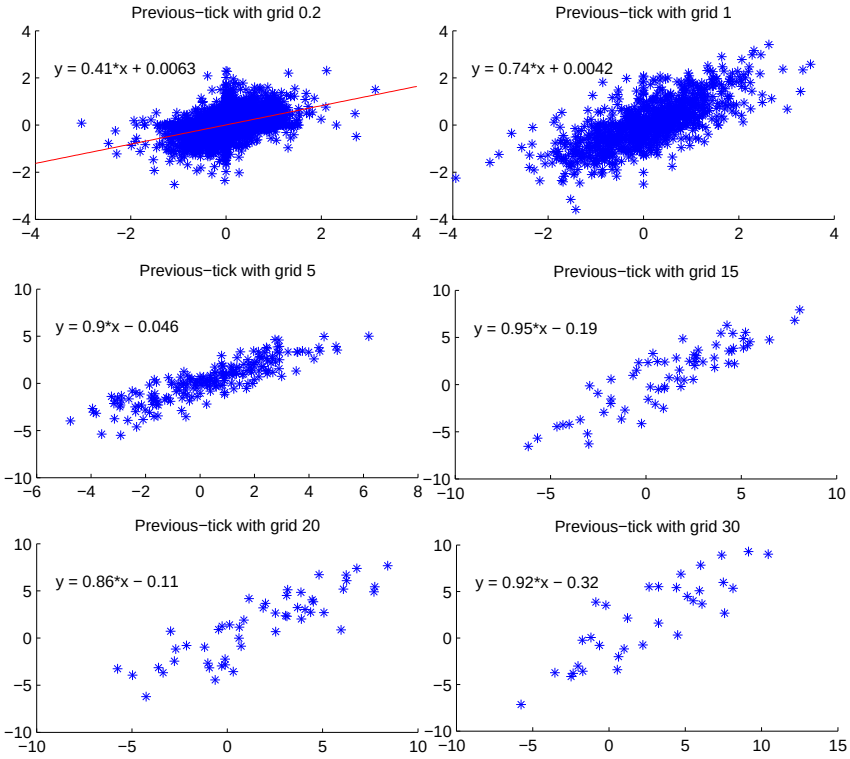

Figure 3.13: Some scatterplots of logreturns $\{R_1(t)\}$ and $\{R_2(t)\}$ based on different sampling scales. The left top figure shows the scatterplot based on the grid with 2 (previous-tick); the right top figure shows the scatterplot with grid equals to 6 (previous-tick); the left bottom shows the scatterplot with grid equals to 15 (previous-tick); the right bottom shows the scatterplot based on the grid equals to 30. We also provide the linear regression equation as shown on the left shoulder of each plot. This result is based on the concurrently correlated Wiener process for the logreturn generating processes and the common factor intensity model for time generating.

confirms that the involved two time series have linear relationship but not other kind of relationship, such as quadratic, cubic, non-linear etc. The limite of the scatterplot, that measuring in one fix time scale, goes against understanding the linear dependence varying with different sampling scale in a complex way. We thus calculate correlations literally as a function of sampling scale for two time series in an effort to better understand the timing and level of change that can occur.

Results of correlation coefficient calcualtion versus sampling scale are shown in Figure 3.14. Two panels display the correlation coefficient of logreturns based on the underlying two different time generating model. The upper panel shows the result based on independent DSPP for time generating process and concurrently correlated Wiener process for logreturn simulation. The bottom panel shows the result based on common factor model for time generating process and concurrently correlated Wiener process for logreturn simulation. Note that the sampling scale refers to previous-take sampling interval.

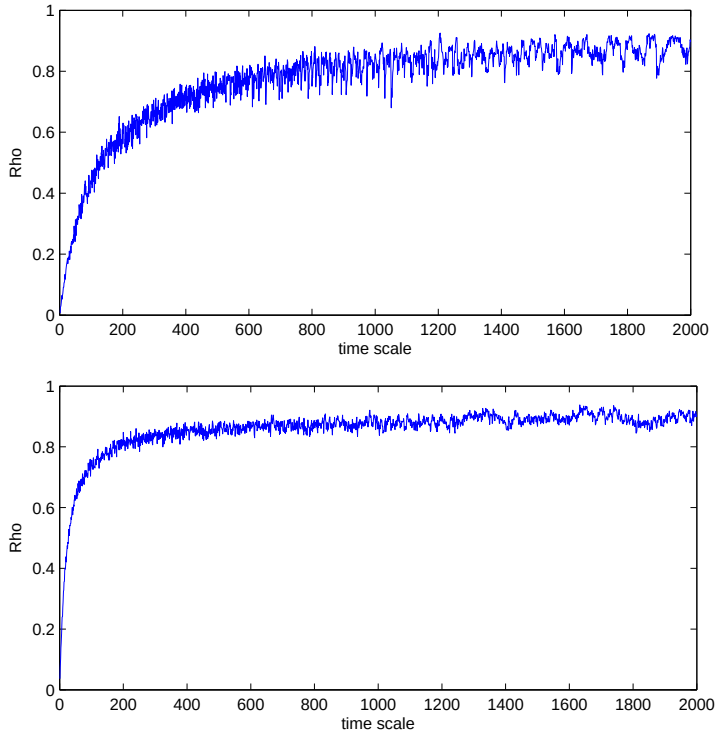

Figure 3.14: The linear correlation coefficients evaluated as a function of the sampling scale. The time generating process is based on the independent DSPP and the logreturn process is according to the concurrently correlated Wiener process (top panel). The time generating process is based on the common factor model and the logreturn process is based on the concurrently correlated Wiener process (bottom panel).

Figure 3.14 show a dramatic decrease in correlation as the sampling scale of logreturns become small for both two figures. This phenomenon is referring to Epps (1979) effect after the first identifiable author to thoroughly document it, more details concerning on this topic refer to Chapter 1. The same effect is also observed in the scatterplots (see Figure 3.12 and Figure 3.13). Basically, the correlation coefficients increase as the sampling scale in-

crease under certain time interval and then remain quite stable after certain sampling scale. For the case of the upper figure in Figure 3.14, the correlation increase as the sampling scale increase enter 1000, after 1000 the correlation coefficient become quite stable with the level of 0.8. On the other hand, for the bottom figure in Figure 3.14, the sampling scale is much smaller, at the time scale only equals to 200, to reach the stable level of correlation coeficient 0.8, with respect to the upper figure. This implies the sampling scale for the stability of correlation coefficient is smaller under the common factor model for the time generating process than under the one without the common intensity or independent DSPP. Note that the horizontal axes, time scale, in Figure 3.14 have nothing to do with physical time, like measured in seconds or minutes. It is a way to grid the time horizon $T = 1000$, providing sampling interval for the simulated data. It is also worth noting that the stable level of correlation coefficient is different from the real one under the model. As assumed in the underlying model, the correlation coefficient is equal to 0.9 which is different with 0.8 as shown in both two figures. Nevertheless, they are quite close.

3.5.5 Bivariate autoregressive process of the first order

A model based on the bivariate autoregressive process of order 1 (Bi-AR(1)), the simulation work will be according to Formular 3.4.15 as described in the Section 3.4.5. In principle, this model dealing with two time series $\{R_1(t)\}$ and $\{R_2(t)\}$ which are not only auto-correlated with parameters $\beta_{11} \neq 0$ (for $\{R_1(t)\}$) and $\beta_{22} \neq 0$ (for $\{R_2(t)\}$), but also cross-correlated with parameters $\beta_{12} \neq 0$ (linear dependence of $\{R_1(t)\}$ on $\{R_2(t)\}$) and $\beta_{21} \neq 0$ (linear dependence of $\{R_2(t)\}$ on $\{R_1(t)\}$), besides, they are also concurrently correlated through the white noise $\epsilon_1(t)$ and $\epsilon_2(t)$. Note that here white noise mean random variables are normally distributed with mean 0 and covariance matrix Σ. It is worth to stress that the off-diagonal elements of the covariance matrix Σ are different from zero if they are concurrently correlated. In the underlying model, $\sigma_{12} = \sigma_{21} \neq 0$, therefore.

For the simulation, we assume the Bi-AR(1) as follows:

$$\begin{cases} R_1(t_i) = -0.5R_1(t_{i-1}) + 0.2R_2(t_{i-1}) + \epsilon_1(t_i), \\ R_2(t_i) = -0.6R_2(t_{i-1}) - 0.3R_1(t_{i-1}) + \epsilon_2(t_i), \end{cases} \tag{3.5.3}$$

where the covariance matrix of $\epsilon_1(t_i)$ and $\epsilon_2(t_i)$ is $\Sigma = \binom{0.1 \quad 0.1}{0.1 \quad 0.2}$.

In this specific modeling framework, it implies that $\boldsymbol{\beta}_0 = [0, 0]'$ and $\boldsymbol{\beta}_1 = \binom{-0.5 \quad 0.2}{-0.3 \quad -0.6}$ (see Formular 3.4.15). So the two time series $\{R_1(t)\}$ and $\{R_2(t)\}$ are auto-correlated with coefficient -0.5 for $\{R_1(t)\}$ and -0.6 for $\{R_2(t)\}$, and cross-correlated with coefficient 0.2 for $R_1(t_i)$ on $R_2(t_{i-1})$ and -0.3 for $R_2(t_i)$ on $R_1(t_{i-1})$. Moreover, they are also concurrently correlated through the white noise $\epsilon_1(t_i)$ and $\epsilon_2(t_i)$ with covariance matrix $\Sigma = \binom{0.1 \quad 0.1}{0.1 \quad 0.2}$. The negative auto-correlation of logreturns can be interpreted that the movement of prices present wave shape, so it moves up and down and so forth. The cross-correlation between two logreturns can be interpreted as the lead-lag relationship, $R_1(t_i)$ depends on $R_2(t_{i-1})$ with coefficient 0.2 means $\{R_1(t)\}$ leads $\{R_2(t)\}$ while $R_2(t_i)$ depends on $R_1(t_{i-1})$ with coefficient -0.3 shows $\{R_2(t)\}$ depends on $\{R_2(t)\}$ with a lag.

According to Equation 3.4.17, the concurrent covariance matrix $\boldsymbol{\Gamma}_0$ is $\binom{0.1282 \quad 0.0962}{0.0962 \quad 0.3846}$ and it is easy to compute the correlation coefficient which is equal to $\rho = 0.433$. For the auto- and cross-correlation with lag-1, according to Equation 3.4.18, we have $\boldsymbol{\Gamma}_1 =$

$\begin{pmatrix} -0.0449 & 0.0288 \\ -0.0962 & -0.2596 \end{pmatrix}$. $\rho_{12}(1) = 0.13$, $\rho_{21}(1) = -0.43$, $\rho_{11}(1) = -0.2$ and $\rho_{22} = -0.56$, according to Equation 3.4.21 and Equation 3.4.22.

Model based on independent DSPP and Bi-AR(1)

In what follows, we will implement the underlying modeling framework by simulation in that two time generating processes are considered. Firstly, we consider the time generating process based on the independent DSPP (in Section 3.5.1), the simulated times are shown on the bottom two figures in Figure 3.7. Then we pool these two sequences of times orderly as shown in Figure 3.6. Now the remaining part is to simulate two logreturn processes, in principle, this can be obtained by data generating according to Equation 3.5.3, given the underlying pooled times.

After generate bivariate logreturn processes, as before, data matching work must be used so that the number of event times for each time serie is the same as that of original time generating model based on the intensity process. Since the intensity based time generating suffers asynchronous problem for two time series, some sort of interpolation should be applied, here we use previous-tick interpolation. Then the correlation measurement based on any sampling scale can proceed, though it is different from the true one.

Figure 3.15 illustrates some scatterplots of logreturns $\{R_1(t)\}$ and $\{R_2(t)\}$ based on different sampling scales. Note that the grid of the previous-tick interpolation is the same as sampling scale. Several results can be obtained. First of all, the plots present significant linear dependence between two time series $\{R_1(t)\}$ and $\{R_2(t)\}$. Second, the correlation increase as sampling increase then remain at quite stable level, about 0.7 as shown in Figure 3.15. The slope is almost zero (0.053) when sampling interval equals to 0.2 (left top panel), then increase steadily as the sampling interval increase, 0.32 for grid 1 (left top panel), 0.58 for grid 4 (left middle panel) and 0.65 for grid 10 (right middle panel). It peaks at the level of 0.73 with grid 15 (left bottom panel) and decreases when sampling interval is 20. The same effect is better viewed on the top figure of Figure 3.17 where the correlation coefficients are plotted as the function of the sampling scale.

Finally, the slope of linear regression equation, as shown on the left shoulder of each panel, can be different from the correlation coefficient in general. $\rho = k \cdot \sigma_{R_1}/\sigma_{R_2}$, where k is the slope of the linear regression and σ_{R_1} is the variance of logreturn $\mathbf{R}_1$, σ_{R_2} is the variance of logreturn $\mathbf{R}_2$. In some extent, the slope k product with ratio $\sigma_{R_1}/\sigma_{R_2}$ can be treated as an estimator of correlation coefficient ρ, denoted as $\hat{\rho}$. If this is the case, then $\hat{\rho} = 0.0175$ when sampling interval is 0.2 (left top figure); $\hat{\rho} = 0.11$ in case sampling interval is 1 (right top figure); and $\hat{\rho} = 0.191$ for sampling interval 4 (left middle figure); $\hat{\rho} = 0.215$ for sampling interval 10 (right middle figure) etc. However, it is important to be clear in mind that one unique correlation coefficient is assumed in the model, that is, $\rho = 0.433$.

Model based on a common factor and Bi-AR(1)

Secondly, in case that the time generating process based on the common factor model (see Section 3.5.2), graphically, the simulated times are presented on the bottom figures in Figure 3.8. Given the two sequences of the simulated times, in practice, we pool them orderly as shown in Figure 3.6. Then we generate the logreturn processes according to

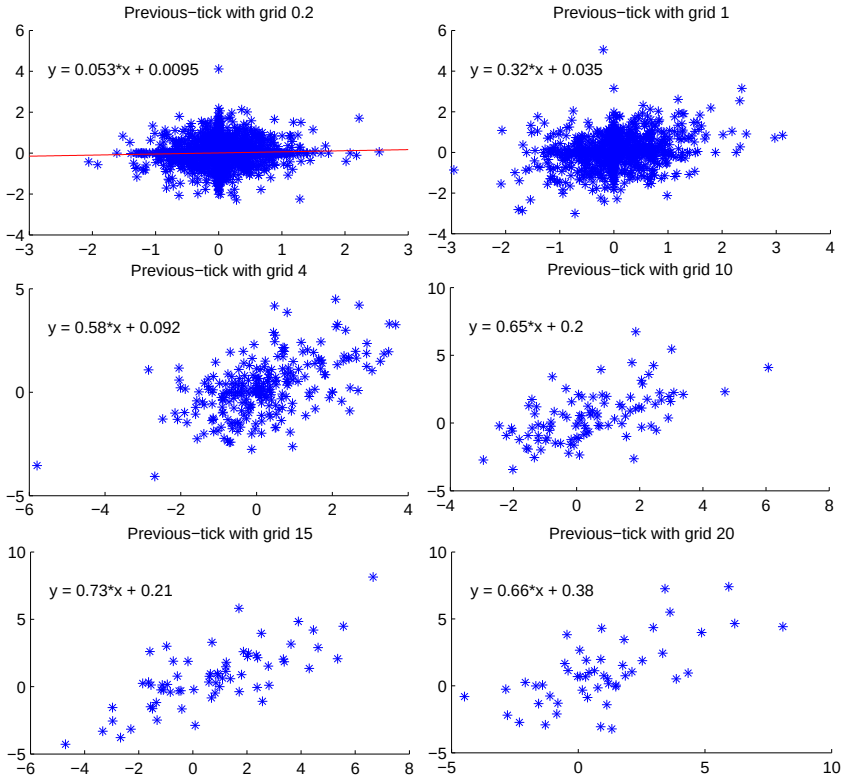

Figure 3.15: Some scatterplots of logreturns $\{R_1(t)\}$ and $\{R_2(t)\}$ based on different sampling scales. Here the previous-tick interpolation is applied. The left top figure shows the scatterplot based on the grid with 0.2 (previous-tick); the right top figure shows the scatterplot with grid equals to 1 (previous-tick). The left middle with grid 4 (previous-tick); the right middle with grid 10 (previous-tick). The left bottom with grid 15 (previous-tick); the right bottom with grid 20 (previous-tick). We also provide the linear regression equation as shown on the left shoulder of each plot. This result is based on independent Wiener processes for the logreturns generating processes and independent DSPP for time generating.

Equation 3.5.3 and make use of the underlying pooled times. Since we generate some false logreturns (a trade-off with asynchronization), time matching work should be carried out then.

When the time matching work is done, the inherent asynchronicity for two time series is turned out again. However, the asynchronicity for this time is not so disagreeable as before, indeed, it is even closer to the real data which are asynchronous. Moreover, the number of logreturns is in line with the time generating process based on the intensity. In order to present scatterplot, which is a way to show the correlation between two series, some sort of interpolation must be used, here we apply previous-tick interpolation(see Figure 1.9). This is because the scatterplot dealing with bivariate vectors that have the same length.

Figure 3.16 provides some scatterplots of logreturns $\{R_1(t)\}$ and $\{R_2(t)\}$ according to different sampling scales. Here the grid of previous-tick interpolation (as shown in figures) also refers to sampling scale. Similar results are obtained as the model based on independent DSPP. First, the plots present significant linear dependence between two time series $\{R_1(t)\}$ and $\{R_2(t)\}$. Second, the correlation increase as sampling increase then remain at quite stable level, about 0.7 as shown in Figure 3.16. The slope is much greater than the previous one (see Figure 3.15) when sampling interval equals to 0.2 (left top panel), 1 (right top panel), 4 (left middle panel) and 10 (right middle panel). While it peaks at the level of 0.74 with grid 15 (left bottom panel) and decreases when sampling interval is 20, which is similar as the previous model (see Figure 3.15). The same effect is better viewed in Figure 3.17 where the correlation coefficients are plotted as the function of the sampling scale.

Finally, the slope of linear regression equation, as shown on the left shoulder of each panel, can be different from the correlation coefficient in general. $\rho = k \cdot \sigma_{R_1}/\sigma_{R_2}$, where k is the slope of the linear regression and σ_{R_1} is the variance of logreturn $\mathbf{R}_1$, σ_{R_2} is the variance of logreturn $\mathbf{R}_2$. In some extent, the slope k product with ratio $\sigma_{R_1}/\sigma_{R_2}$ can be treated as an estimator of correlation coefficient ρ, denoted as $\hat{\rho}$. If this is the case, then $\hat{\rho} = 0.112$ when sampling interval is 0.2 (left top figure); $\hat{\rho} = 0.2$ in case sampling interval is 1 (right top figure); and $\hat{\rho} = 0.228$ for sampling interval 4 (left middle figure); $\hat{\rho} = 0.244$ for sampling interval 10 (right middle figure) etc. However, it is important to be clear in mind that one unique correlation coefficient is assumed in the model, that is, $\rho = 0.433$.

Correlation versus sampling scale

To better gauge the linear dependence between two time series, the correlation coefficient measurement takes the first place. When the standard Pearson correlation coefficient is calculated between two time series, the assumption is that this quantity is evaluated for equally spaced data. For the case of ultra-high frequency financial time series this is invalide, since they are irregularly spaced in time and asynchronous for any two time series. What matters here is the simulation progress is based on the character of the ultra-high frequency data, the generated logreturn processes thereby possess the feature of irregular spaced in time and non-synchronicity when two time series are involved. This issu is addressed in the previous part concerning on time generating context.

Given the simulated logreturns, the dependence between these two series varying with the sampling scale is shown in the scatterplots. From the shape of the scatterplots, it confirms that the involved two time series have linear relationship but not other kind of

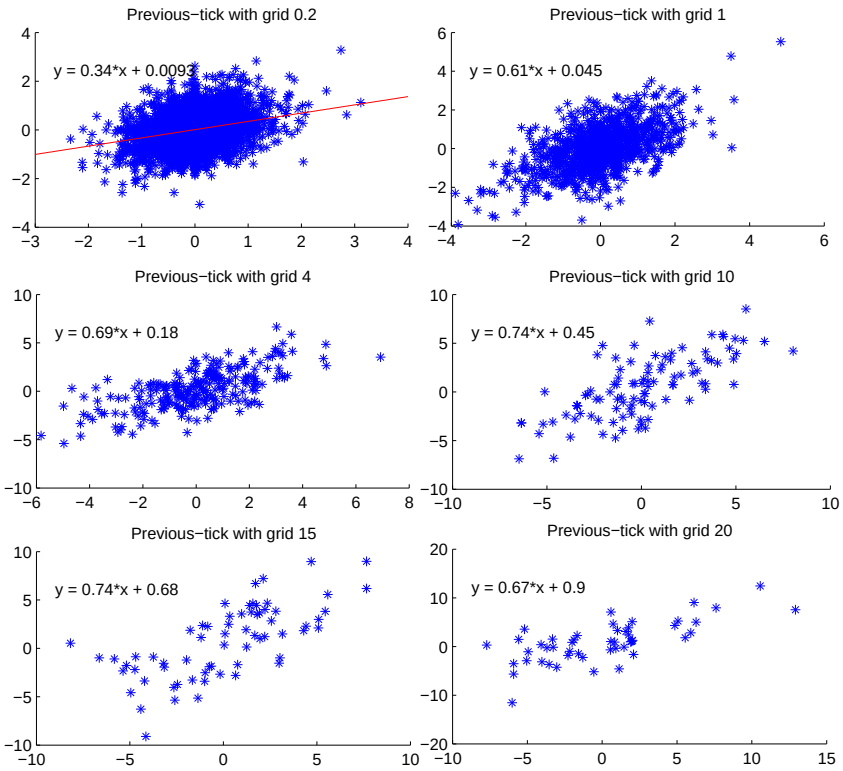

Figure 3.16: Some scatterplots of logreturns $\{R_1(t)\}$ and $\{R_2(t)\}$ based on different sampling scales. Here the previous-tick interpolation is applied. The left top figure shows the scatterplot based on the grid with 0.2 (previous-tick); the right top figure shows the scatterplot with grid equals to 1 (previous-tick). The left middle with grid of 4 (previou-tick); the right middle with grid 10 (previous-tick). The left bottom with grid 15 (previous-tick); the right bottom with grid of 20 (previous-tick). We also provide the linear regression equation as shown on the left shoulder of each plot. This result is based on the concurrently correlated Wiener process for the logreturn generating processes and the common factor intensity model for time generating.

relationship, such as quadratic, cubic, non-linear etc. The limite of the scatterplot, that measuring in one fix time scale, goes against understanding the linear dependence varying with different sampling scale in a complex way. We thus calculate correlations literally as a function of sampling scale for two time series in an effort to better understand the timing and level of change that can occur.

Results of correlation coefficient calcualtion versus sampling scale are shown in Figure 3.17. Two panels display the correlation coefficient of logreturns based on the underlying two different time generating model. The upper panel shows the result based on independent DSPP for time generating process and VAR(1) for logreturn simulation. The bottom panel shows the result based on common factor model for time generating process and VAR(1) for logreturn simulation. Note that the sampling scale refers to previous-take sampling interval.

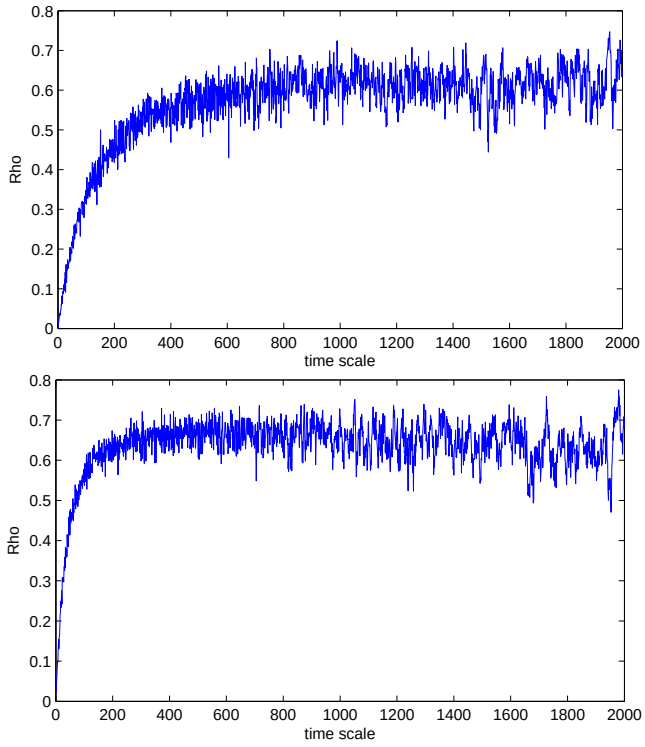

Figure 3.17: The linear correlation coefficients evaluated as a function of the sampling scale. The time generating process is based on the independent DSPP and the logreturn process is according to VAR(1) (top panel). The time generating process is based on the common factor model and the logreturn process is based on VAR(1) (bottom panel).

Figure 3.17 show a dramatic decrease in correlation as the sampling scale of logreturns become small for both two figures. This phenomenon is referring to Epps (1979) effect after the first identifiable author to thoroughly document it, more details concerning on this topic refer to Chapter 1. The same effect is also observed in the scatterplots (see Figure 3.15 and Figure 3.16). Basically, the correlation coefficients increase as the sampling scale increase under certain time interval and then remain quite stable after certain sampling scale. For the case of the upper figure in Figure 3.17, the correlation increase as the sampling

scale increase enter 1000, after 1000 the correlation coefficient become quite stable with the level of 0.8. On the other hand, for the bottom figure in Figure 3.17, the sampling scale is much smaller, at the time scale only equals to 200, to reach the stable level of correlation coeficient 0.8, with respect to the upper figure. This implies the sampling scale for the stability of correlation coefficient is smaller under the common factor model for the time generating process than under the one without the common intensity or independent DSPP. Note that the horizontal axes, time scale, in Figure 3.17 have nothing to do with physical time, like measured in seconds or minutes. It is a way to grid the time horizon $T = 1000$, providing sampling interval for the simulated data. It is also worth noting that the stable level of correlation coefficient is different from the real one under the model. As assumed in the underlying model, the correlation coefficient is equal to 0.9 which is different with 0.8 as shown in both two figures. Nevertheless, they are quite close.

3.5.6 Bi-AR(1) with zero cross-correlation

A typical case for the Bi-AR(1) model we implemented here is to specify the underlying Bi-AR(1) with zero cross-correlation, essentially, the simulation experiment will be based on Equation 3.4.15 as described in Section 3.4.5. As introduced in Section 3.4.6, this specific model assume that two time series $\{R_1(t)\}$ and $\{R_2(t)\}$ are auto-correlated and concurrently correlated. So the parameter $\beta_{11} \neq 0$ which measures the auto-correlation of $\{R_1(t)\}$ and the parameter $\beta_{22} \neq 0$ that measures for $\{R_2(t)\}$, whereas $\sigma_{12} = \sigma_{21} \neq 0$ measure the concurrent correlation between the two time series $\{R_1(t)\}$ and $\{R_2(t)\}$. Note that σ_{12} and σ_{21} are the off-diagonal elements of covariance matrix Σ. More details refer to Section 3.4.6.

To simulate logreturns based on this modeling framework, we assume

$$\begin{cases} R_1(t_i) = -0.5R_1(t_{i-1}) + \epsilon_1(t_i), \\ R_2(t_i) = -0.6R_2(t_{i-1}) + \epsilon_2(t_i), \end{cases} \tag{3.5.4}$$

where the covariance matrix of $\epsilon_1(t_i)$ and $\epsilon_2(t_i)$ is $\Sigma = \begin{pmatrix} 0.1 & 0.1 \\ 0.1 & 0.2 \end{pmatrix}$.

In this specific modeling framework, it implies that $\boldsymbol{\beta}_0 = [0, 0]'$ and $\boldsymbol{\beta}_1 = \begin{pmatrix} -0.5 & 0 \\ -0 & -0.6 \end{pmatrix}$ (see Formular 3.4.15). So the two time series $\{R_1(t)\}$ and $\{R_2(t)\}$ are auto-correlated with coefficient -0.5 for $\{R_1(t)\}$ and -0.6 for $\{R_2(t)\}$, and zero cross-correlation with a lag. Moreover, they are concurrently correlated through the white noise $\epsilon_1(t_i)$ and $\epsilon_2(t_i)$ with covariance matrix $\Sigma = \begin{pmatrix} 0.1 & 0.1 \\ 0.1 & 0.2 \end{pmatrix}$. As before, the negative auto-correlation of logreturns can be interpreted that the movement of prices present wave shape, so it moves up and down and so forth. The cross-correlation between two logreturns can be interpreted as the lead-lag relationship, in this case, there is no lead-lag relationship between the underlying two time series $\{R_1(t)\}$ and $\{R_2(t)\}$.

Based on Equation 3.4.17, the concurrent covariance matrix $\boldsymbol{\Gamma}_0$ is $\begin{pmatrix} 0.1333 & 0.1429 \\ 0.1429 & 0.3125 \end{pmatrix}$ and it is easy to compute the correlation coefficient which is equal to $\rho = 0.699$. For the auto- and cross-correlation with lag-1, according to Equation 3.4.18, we have $\boldsymbol{\Gamma}_1 = \begin{pmatrix} -0.0667 & -0.0714 \\ -0.0857 & -0.1875 \end{pmatrix}$. $\rho_{12}(1) = -0.35$, $\rho_{21}(1) = -0.42$, $\rho_{11}(1) = -0.33$ and $\rho_{22} = -0.92$, according to Equation 3.4.21 and Equation 3.4.22. From this result, it is surprised that the time series $\{R_1(t)\}$ and $\{R_2(t)\}$ are cross-correlated with a lag, $\rho_{12}(1) = -0.35$, $\rho_{21}(1) = -0.42$, which seems contrary to the assumption of the model in that $\beta_{12} = \beta_{21} = 0$. Nevertheless, if we go through this model in deep insight, this obstacle result can be attributed to both

auto-correlation and concurrent correlation between $R_1(t_i)$ and $R_2(t_i)$. Basically, if $R_1(t_i)$ is concurrently correlated with $R_2(t_i)$, and $R_2(t_i)$ is auto-correlated with $R_2(t_{i-1})$, then $R_1(t_i)$ is also correlated with $R_2(t_{i-1})$, moreover, the sign of the cross-correlation depends on both sign of concurrent correlation and auto-correlation. If positive concurrent correlation combines with negative auto-correlation, we will get negative cross-correlation, like this modeling framework. In principle, the sign is the same as the result of product, mathematically.

On the other hand, if we comparing this model with more general previous modeling framework, we find that the only difference between them is $\beta_{12} = \beta_{21} = 0$ for this model (Bi-AR(1) by specifying zero cross-correlation) while $\beta_{12} = 0.2, \beta_{21} = -0.3$ for the previous model (Bi-AR(1)), one interesting result is that there is much stronger linear dependence for the underlying model (Bi-AR(1) by specifying zero cross-correlation) both auto- and cross-correlation with a lag, by comparing $\boldsymbol{\rho}(l) = \left(\begin{smallmatrix} -0.2 & 0.13 \\ -0.43 & -0.56 \end{smallmatrix}\right)$ for Bi-AR(1) modeling framework with $\boldsymbol{\rho}(l) = \left(\begin{smallmatrix} -0.33 & -0.35 \\ -0.42 & -0.92 \end{smallmatrix}\right)$ for Bi-AR(1) modeling framework. The following scatterplots will also prove this result.

Model based on independent DSPP and Bi-AR(1) with zero cross-correlation

We propose two methods for time generating as describe before. Firstly, we consider the time generating process based on the independent DSPP (in Section 3.5.1), the simulated times are shown on the bottom two figures in Figure 3.7. Then we pool these two sequences of times orderly as shown in Figure 3.6. This leads the way for simulating logreturn processes, which require the same length of each component of the bivariate logreturn processes.

After generate bivariate logreturn processes, data matching work is proceeded so that the number of event times for each time serie is the same as that of original time generating model based on the intensity process. Then, previous-tick interpolation should be applied given the asynchronous data, this is because the data matching work restore the data to the inherent asynchronicity.

Figure 3.18 shows some scatterplots of logreturns $\{R_1(t)\}$ and $\{R_2(t)\}$ based on different sampling scales. Note that the grid of the previous-tick interpolation is the same as sampling scale. Several results can be obtained. First of all, the plots present significant linear dependence between two time series $\{R_1(t)\}$ and $\{R_2(t)\}$. Second, the correlation increase as sampling increase then remain at quite stable level, about 0.98 as shown in Figure 3.15. It peaks at the level of 0.98 with grid 20 (left bottom panel) and remain there with sampling interval 20. The same effect is better viewed on the top figure of Figure 3.23 where the correlation coefficients are plotted as the function of the sampling scale.

Finally, the slope of linear regression equation, as shown on the left shoulder of each panel, can be different from the correlation coefficient in general. $\rho = k \cdot \sigma_{R_1}/\sigma_{R_2}$, where k is the slope of the linear regression and σ_{R_1} is the variance of logreturn $\mathbf{R}_1$, σ_{R_2} is the variance of logreturn $\mathbf{R}_2$. In some extent, the slope k product with ratio $\sigma_{R_1}/\sigma_{R_2}$ can be treated as an estimator of correlation coefficient ρ, denoted as $\hat{\rho}$. If this is the case, then $\hat{\rho} = 0.09$ when sampling interval is 0.1 (left top figure); $\hat{\rho} = 0.28$ in case sampling interval is 0.5 (right top figure); and $\hat{\rho} = 0.344$ for sampling interval 4 (left middle figure); $\hat{\rho} = 0.417$ for sampling interval 10 (right middle figure) etc. However, it is important to be clear in mind that one unique correlation coefficient is assumed in the model, that is,

$\rho = 0.6999$.

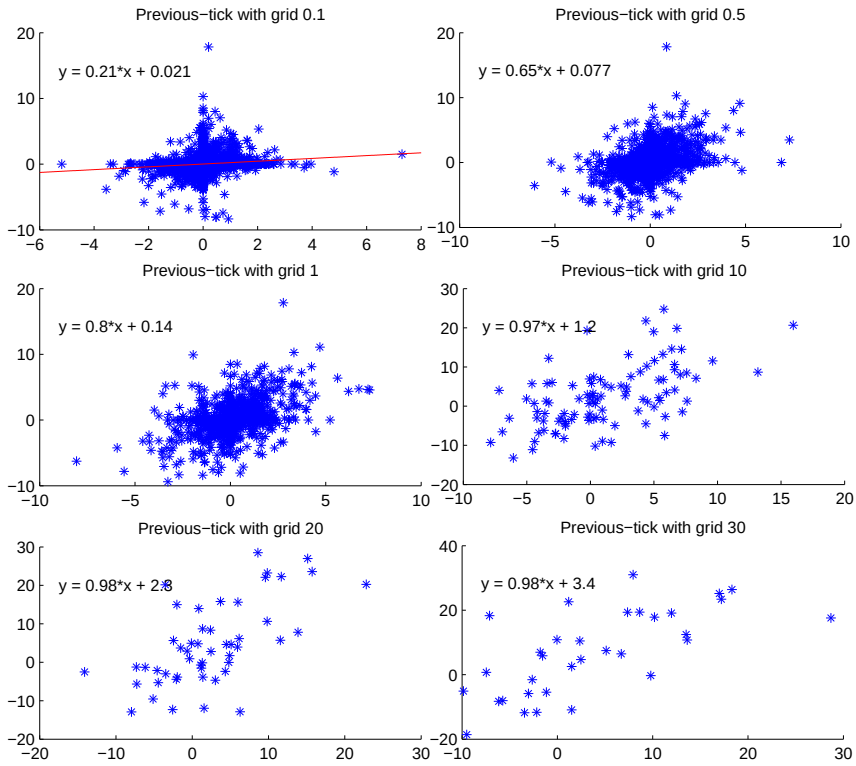

Figure 3.18: Some scatterplots of logreturns $\{R_1(t)\}$ and $\{R_2(t)\}$ based on different sampling scales. Here we apply the previous-tick interpolation. The left top figure shows the scatterplot based on the grid with 0.1 (previous-tick); the right top figure shows the scatterplot with grid equals to 0.5 (previous-tick). The left middle figure with grid 1 (previous-tick); the right middle figure with grid 10 (previous-tick). The left bottom figure is based on grid of 20 (previous-tick); the right bottom figure is based on grid of 30 (previous-tick). We also provide the linear regression equation as shown on the left shoulder of each plot. This result is based on independent Wiener processes for the logreturns generating processes and independent DSPP for time generating.

Model based on a common factor and Bi-AR(1) with zero cross-correlation

Secondly, in case that the time generating process based on the common factor model (see Section 3.5.2), graphically, the simulated times are presented on the bottom figures in Figure 3.8. Given the two sequences of the simulated times, in practice, we pool them orderly as shown in Figure 3.6. Then, we go on the simulation work for logreturn processes which are based on the pooled times. Since we generate some false logreturns (a trade-off with asynchronization), time matching work should be carried out then.

When the time matching work is done, the inherent asynchronicity for two time series is turned out again. However, the asynchronicity for this time is not so disagreeable as before, indeed, it is even closer to the real data which are asynchronous. Moreover, the number of logreturns is in line with the time generating process based on the intensity. In

order to present scatterplot, which is a way to show the correlation between two series, the previous-tick interpolation will be applied (see Figure 1.9). This is because scatterplot is dealing with bivariate vectors with the same length.

Figure 3.19 provides some scatterplots of logreturns $\{R_1(t)\}$ and $\{R_2(t)\}$ according to different sampling scales. Here the grid of previous-tick interpolation (as shown in figures) also refers to sampling scale. Similar results are obtained as the model based on independent DSPP. First, the plots present significant linear dependence between two time series $\{R_1(t)\}$ and $\{R_2(t)\}$. Second, the correlation increase as sampling increase then remain at quite stable level, about 1.3 as shown in Figure 3.16. The slope is much greater than the previous one (see Figure 3.18) when sampling interval equals to 0.2 (left top panel), 1 (right top panel), 4 (left middle panel) and 10 (right middle panel). It peaks at the level of 1.3 with grid 10 (right middle panel) and oscillates around 1.2 when sampling interval increase (greater than 10). The same effect is better viewed in Figure 3.20 where the correlation coefficients are plotted as the function of the sampling scale.

Finally, the slope of linear regression equation, as shown on the left shoulder of each panel, can be different from the correlation coefficient in general. $\rho = k \cdot \sigma_{R_1}/\sigma_{R_2}$, where k is the slope of the linear regression and σ_{R_1} is the variance of logreturn $\mathbf{R}_1$, σ_{R_2} is the variance of logreturn $\mathbf{r}_2$. In some extent, the slope k product with ratio $\sigma_{R_1}/\sigma_{R_2}$ can be treated as an estimator of correlation coefficient ρ, denoted as $\hat{\rho}$. If this is the case, then $\hat{\rho} = 0.262$ when sampling interval is 0.1 (left top figure); $\hat{\rho} = 0.43$ in case sampling interval is 0.5 (right top figure); and $\hat{\rho} = 0.473$ for sampling interval 1 (left middle figure); $\hat{\rho} = 0.56$ for sampling interval 10 (right middle figure) etc. However, it is important to be clear in mind that one unique correlation coefficient is assumed in the model, that is, $\rho = 0.6999$.

Correlation versus sampling scale

To better gauge the linear dependence between two time series, the correlation coefficient measurement takes the first place. When the standard Pearson correlation coefficient is calculated between two time series, the assumption is that this quantity is evaluated for equally spaced data. For the case of ultra-high frequency financial time series this is invalide, since they are irregularly spaced in time and asynchronous for any two time series. What matters here is the simulation progress is based on the character of the ultra-high frequency data, the generated logreturn processes thereby possess the feature of irregular spaced in time and non-synchronicity when two time series are involved. This issu is addressed in the previous part concerning on time generating context.

Given the simulated logreturns, the dependence between these two series varying with the sampling scale is shown in the scatterplots. From the shape of the scatterplots, it confirms that the involved two time series have linear relationship but not other kind of relationship, such as quadratic, cubic, non-linear etc. The limite of the scatterplot, that measuring in one fix time scale, goes against understanding the linear dependence varying with different sampling scale in a complex way. We thus calculate correlations literally as a function of sampling scale for two time series in an effort to better understand the timing and level of change that can occur.

Results of correlation coefficient calcualtion versus sampling scale are shown in Figure 3.20. Two panels display the correlation coefficient of logreturns based on the un-

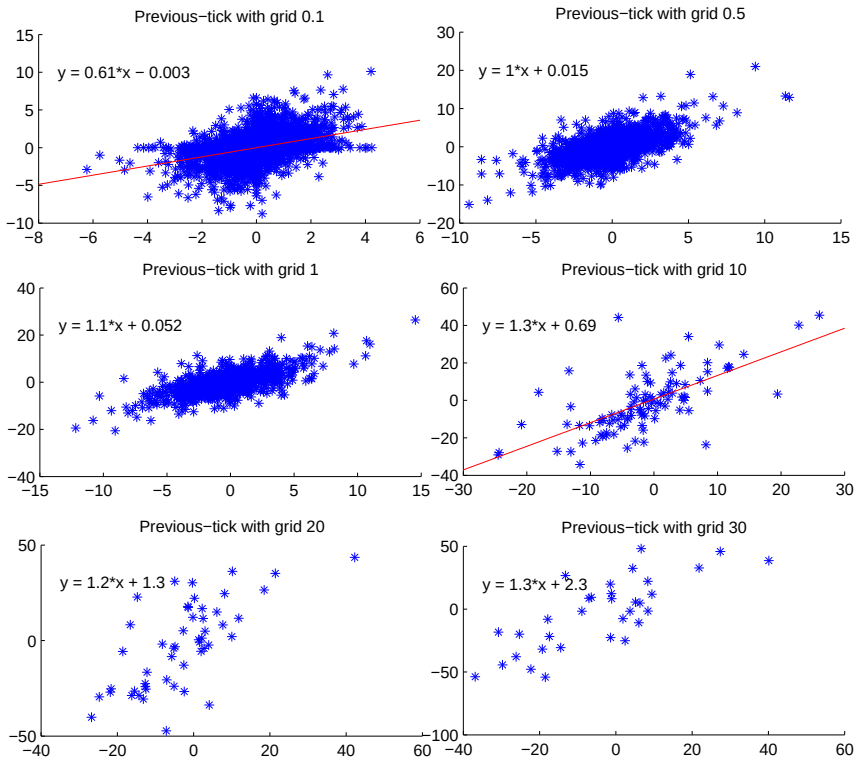

Figure 3.19: Some scatterplots of logreturns $\{R_1(t)\}$ and $\{R_2(t)\}$ based on different sampling scales. The previous-tick interpolation is applied. The left top figure shows the scatterplot based on the grid with 0.1 (previous-tick); the right top figure shows the scatterplot with grid equals to 0.5 (previous-tick). The left middle figure with grid 1 (previous-tick); the right middle figure with grid 10 (previous-tick). The left bottom figure is based on grid of 20 (previous-tick); the right bottom figure is based on grid of 30 (previous-tick). We also provide the linear regression equation as shown on the left shoulder of each plot. This result is based on the concurrently correlated Wiener process for the logreturn generating processes and the common factor intensity model for time generating.

derlying two different time generating model. The upper panel shows the result based on independent DSPP for time generating process and VAR(1) with zero cross-corrrelation for logreturn simulation. The bottom panel shows the result based on common factor model for time generating process and VAR(1) with zero cross-correlation for logreturn simulation. Note that the sampling scale refers to previous-take sampling interval.

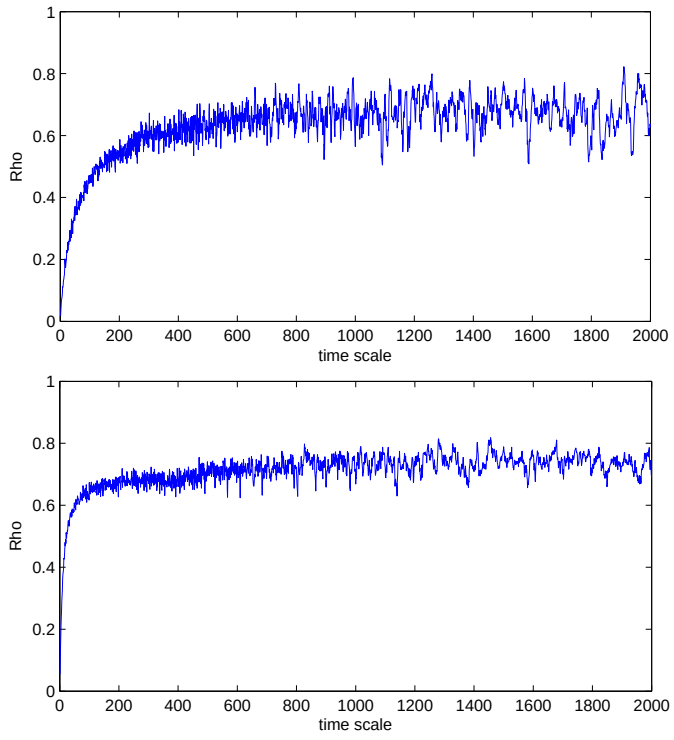

Figure 3.20: The linear correlation coefficients evaluated as a function of the sampling scale. The time generating process is based on the independent DSPP and the logreturn process is according to the VAR(1) with zero cross- correlation (top panel). The time generating process is based on the common factor model and the logreturn process is based on the VAR(1) with zero cross-correlation (bottom panel).

3.5.7 Independent Bi-AR(1)

As for mutually independent Bi-AR(1) model, basically, it is dealing with a model that two time series $\{R_1(t)\}$ and $\{R_2(t)\}$ are just autocorrelated. The simulation experiment will be carried on based on Equation 3.4.25 (in Section 3.4.7).

For the simulation, we assume

$$\begin{cases} R_1(t_i) = -0.5R_1(t_{i-1}) + \epsilon_1(t_i), \\ R_2(t_i) = -0.6R_2(t_{i-1}) + \epsilon_2(t_i), \end{cases} \tag{3.5.5}$$

where the covariance matrix of $\epsilon_1(t_i)$ and $\epsilon_2(t_i)$ is $\Sigma = \begin{pmatrix} 0.1 & 0 \\ 0 & 0.2 \end{pmatrix}$.

Under this specific modeling framework, it implies that $\boldsymbol{\beta}_0 = [0,0]'$ and $\boldsymbol{\beta}_1 = \begin{pmatrix} -0.5 & 0 \\ -0 & -0.6 \end{pmatrix}$ (see Formular 3.4.15). So the two time series $\{R_1(t)\}$ and $\{R_2(t)\}$ are just auto-correlated

with coefficient -0.5 for $\{R_1(t)\}$ and -0.6 for $\{R_2(t)\}$, and zero cross-correlation with a lag. Moreover, the only difference from the previous two modeling frameworks is that assuming the white noise $\epsilon_1(t_i)$ and $\epsilon_2(t_i)$ are not correlated with covariance matrix $\Sigma = \begin{pmatrix} 0.1 & 0 \\ 0 & 0.2 \end{pmatrix}$.

According to Equation 3.4.17, the concurrent covariance matrix $\mathbf{\Gamma}_0$ is $\begin{pmatrix} 0.1333 & 0 \\ 0 & 0.3125 \end{pmatrix}$ and it is easy to compute the correlation coefficient which is equal to $\rho = 0$. For the auto- and cross-correlation with lag-1, according to Equation 3.4.18, we have $\mathbf{\Gamma}_1 = \begin{pmatrix} -0.0667 & 0 \\ 0 & -0.1875 \end{pmatrix}$. $\rho_{12}(1) = 0$, $\rho_{21}(1) = 0$, $\rho_{11}(1) = -0.33$ and $\rho_{22} = -0.92$ according to Equation 3.4.21 and Equation 3.4.22. Hence, the two time series $\{R_1(t)\}$ and $\{R_2(t)\}$ are not cross-correlated either concurrently or with a lag. Based on this, we would expect that the following scatterplots show quite round shapes.

Model based on independent DSPP and independent Bi-AR(1)

In what follows, we will implement the underlying modeling framework by simulation in that two time generating processes are considered. Firstly, we consider the time generating process based on the independent DSPP (in Section 3.5.1), the simulated times are shown on the bottom two figures in Figure 3.7. Then we pool these two sequences of times orderly as shown in Figure 3.6. Now the remaining part is to simulate two logreturn processes, in principle, this can be obtained by data generating according to Equation 3.5.3, given the underlying pooled times.

After generate bivariate logreturn processes, as before, data matching work must be used so that the number of event times for each time serie is the same as that of original time generating model based on the intensity process. Since the intensity based time generating suffers asynchronous problem for two time series, some sort of interpolation should be applied, here we use previous-tick interpolation. Then the correlation measurement based on any sampling scale can be proceeded, though it is different from the true one.

Figure 3.21 shows some scatterplots of logreturns $\{R_1(t)\}$ and $\{R_2(t)\}$ based on different sampling scales or grids of previous-tick interpolation. All the scatterplots of Figure 3.21 show no significant linear dependence between two time series $\{R_1(t)\}$ and $\{R_2(t)\}$. The same effect is better viewed from the top figure of Figure 3.11 where the correlation coefficients are plotted as a function of the sampling scale and where the same data are used in Figure 3.9 and Figure 3.11 (top).

Model based on a common factor and independent Bi-AR(1)

Secondly, in case that the time generating process based on the common factor model (see Section 3.5.2), graphically, the simulated times are presented on the bottom figures in Figure 3.8. Given the underlying two sequences of the simulated times, in practice, we pool them orderly as shown in Figure 3.6. Accordingly, we go on the simulation work about logreturn processes which are based on the pooled times. Since we generate some false logreturns (a trade-off with asynchronization), perhaps a lot, time matching work should be carried out then.

When the time matching work is done, the inherent asynchronicity for two time series would be turned out again. However, the asynchronicity for this time is not so disagreeable as before, indeed, it is closer to the real data which are asynchronous originally and make the simulation work meaningful. Moreover, this procedure keep the number of lo-

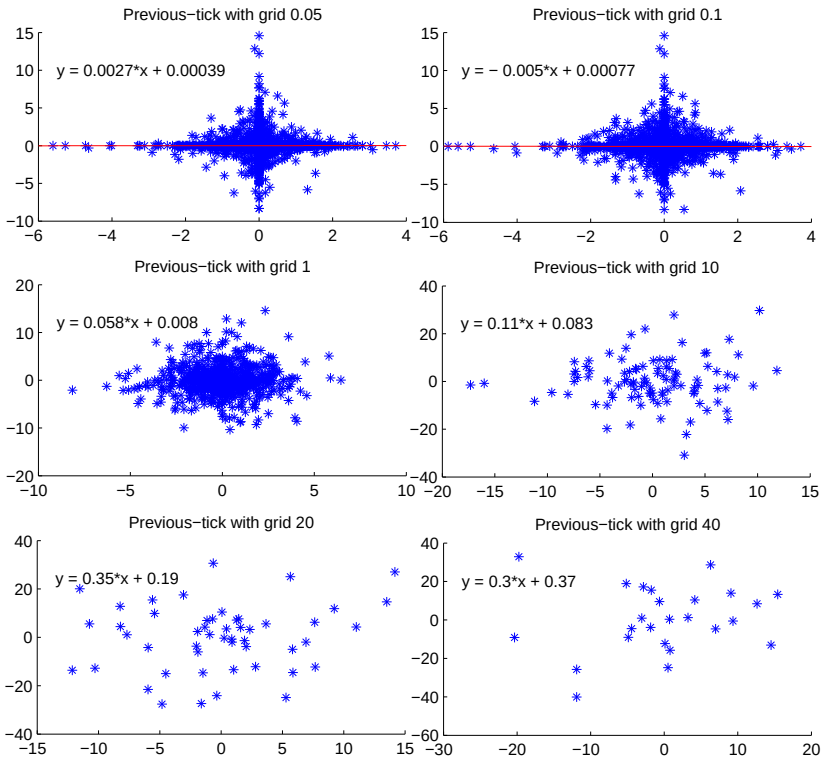

Figure 3.21: Some scatterplots of logreturns $\{R_1(t)\}$ and $\{R_2(t)\}$ based on different sampling scales. The previous-tick interpolation is applied here. The left top figure shows the scatterplot based on the grid with 0.05 (previous-tick); the right top figure shows the scatterplot with grid equals to 0.1 (previous-tick). The left middle figure with grid 1 (previous-tick); the right middle figure with grid 10 (previous-tick). The left bottom figure is based on grid of 20 (previous-tick); the right bottom figure is based on grid of 40. We also provide the linear regression equation as shown on the left shoulder of each plot. This result is based on independent Wiener processes for the logreturns generating processes and independent DSPP for time generating.

greturns in line with the time generating process based on the intensity. In order to present scatterplot, which is a way to show the correlation (linear dependence) between two series visually, the previous-tick interpolation will be applied (see Figure 1.9) such that the log-prices are equally spaced. This is because the scatterplot dealing with bivariate vectors that have the same length.

Figure 3.22 provides some scatterplots of logreturns $\{R_1(t)\}$ and $\{R_2(t)\}$ according to different sampling scales. Note that the grid of previous-tick interpolation as shown in Figure 3.22 is the same as sampling scale. As seen from Figure 3.22, there is no significant linear relationship between two time series $\{R_1(t)\}$ and $\{R_2(t)\}$. The slope of each linear regression, as shown on the left shoulder of each panel, is oscillating around zero. The same effect is better viewed in the bottom figure of Figure 3.23 where correlation coefficients are plotted as a function of the sampling scale. Note that the same data are used in Figure 3.22 and Figure 3.23 (bottom).

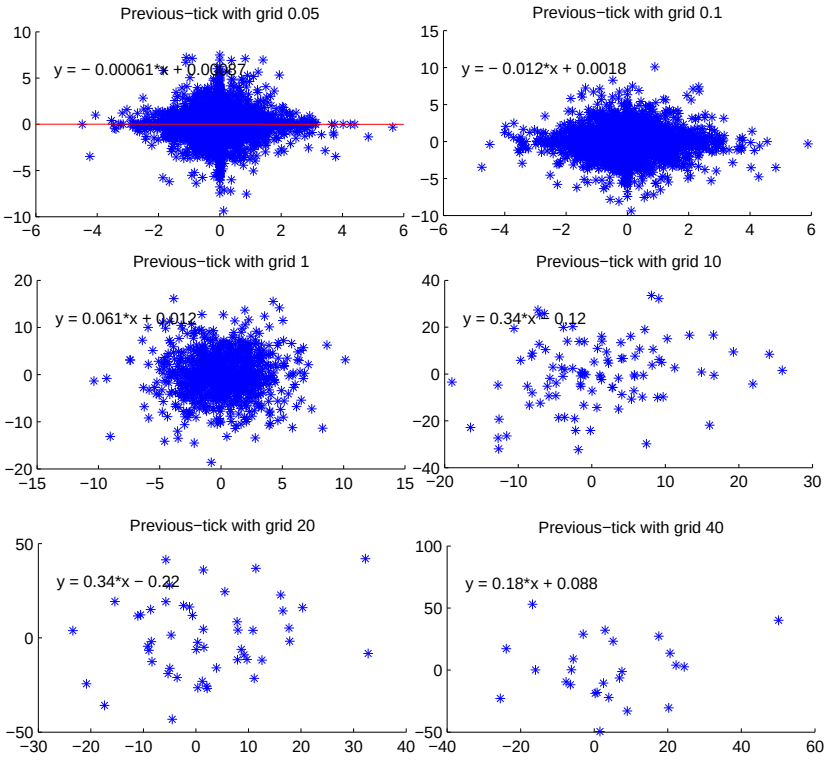

Figure 3.22: Some scatterplots of logreturns $\{R_1(t)\}$ and $\{R_2(t)\}$ based on different grids of previous-tick interpolation. The left top figure shows the scatterplot based on the grid with 0.05 (previous-tick); the right top figure shows the scatterplot with grid equals to 0.1 (previous-tick). The left middle figure with grid 1 (previous-tick); the right middle figure with grid 10; The left bottom figure is based on grid of 20 (previous-tick); the right bottom figure is based on grid of 40 (previous-tick). Here we also provide the linear regression function as shown on the left shoulder of each panel. This result is based on the concurrently correlated Wiener process for the logreturn generating processes and the common factor intensity model for time generating.

Correlation versus sampling scale

To better gauge the linear dependence between two time series, the correlation coefficient measurement takes the first place. When the standard Pearson correlation coefficient is calculated between two time series, the assumption is that this quantity is evaluated for equally spaced data. For the ultra-high-frequency financial time series, however, this method cannot be applied directly, since they are irregularly spaced in time and asynchronous for any two time series. What matters here is the simulation progress is based on the character of the ultra-high-frequency data, the generated logreturn processes thereby possess the feature of irregular spaced in time and non-synchronicity when two time series are involved. This issue is addressed in Section 3.3.

Given the simulated logreturns, the dependence between these two series varying with the sampling scale is shown in the scatterplots. From the shape of the scatterplots, it confirms that the involved two time series have linear relationship but not other kind of relationship, such as quadratic, cubic, non-linear etc. The limite of the scatterplot, that measuring in one fix time scale, goes against understanding the linear dependence varying with different sampling scale in a complex way. We thus calculate correlations literally as a function of sampling scale for two time series in an effort to better understand the timing and level of change that can occur.

Results of correlation coefficient calcualtion versus sampling scale are shown in Figure 3.23. Two panels display the correlation coefficient of logreturns based on the underlying two different time generating model. The upper panel shows the result based on independent DSPP for time generating process and independent VAR(1) for logreturn simulation. The bottom panel shows the result based on common factor model for time generating process and independent VAR(1) for logreturn simulation. Note that the sampling scale refers to previous-take sampling interval. Both of these two panels in Figure 3.23 show no linear relationship between two time series $\{R_1(t)\}$ and $\{R_2(t)\}$ based on data generating process with independent Bi-AR(1), neither on independent DSPP nor on common factor model for time generating work. Thus the common component of intensity does not convey any correlation into two time series.

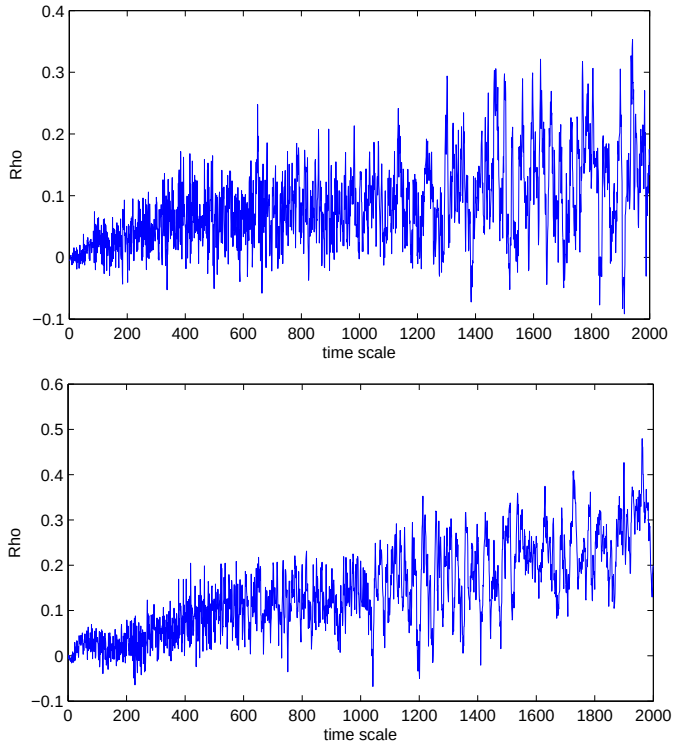

Figure 3.23: The linear correlation coefficients evaluated as a function of the sampling scale. The time generating process is based on the independent DSPP and the logreturn process is according to the independent VAR(1) (top panel). The time generating process is based on the common factor model and the logreturn process is based on the independent VAR(1) (bottom panel).

3.6 Conclusions

In this Chapter we studied the correlation implied by five modeling frameworks for logreturn processes through extensive Monte Carlo simulations. The underlying five frameworks are (1) returns generated from independent Wiener process; (2) concurrently correlated Wiener process; (3) bivariate autoregressive with first lag (Bi-AR(1)); (4) Bi-AR(1) by specializing zero cross-correlation; and (5) independent Bi-AR(1). In particular, we investigated the effect of the sampling interval on the correlation between pairs of logreturns, and compare the simulated correlations with the correlations calculated on the empirical data. The aim is to compare these five frameworks to see which one fits the empirical data better from the point of view of the implied correlations, which are among the basic characteristics that we are trying to model.

As shown in above simulation studies, the behavior of the correlation coefficient as a function of the sampling interval of logreturns has been investigated for each modeling framework. The results we found can be grouped into the following two types: first, modeling frameworks (1) (Figure 3.11) and (5) (Figure 3.23) show almost zero correlations between two time series $\{R_1(t)\}$ and $\{R_2(t)\}$ for both two time generating process (independent DSPP and common factor model). According to what we observed in empirical UHF data (through Figure 1.12 to 1.18), nonzero correlations of logreturns are dramatically attenuated when this interval decreases and enters the intrahour region. This behavior is called the Epps effect. Thus, these two modeling framworks are not qualified. Second, modeling frameworks (2) (Figure 3.14), (3) (Figure 3.17) and (4) (Figure 3.20) show nonzero correlations of logreturns and present the phenomenon of Epps effect. When compare two time generating frameworks, the independent DSPP and the common factor model, for all these three logreturn generating frameworks (2), (3) and (4), the correlation coefficients over sampling intervals based on the common factor model is closer than based on the independent DSPP to the Epps effect.

Figure 3.24 shows a comparison of Epps effect and simulated correlation coefficients based on the VAR(1) logreturn model. As we can see, the bottom right panel shows correlation coefficient as function of sampling intervals which is based on the common factor model for time generating process and VAR(1) for logreturn present a dramatic decrease in correlation as the sampling interval become small. Whilst the top right panel displays a bit less dramatic when the sampling intervals are small. When we compare the Epps effect (left panel) and two correlation coefficient figures (right panels) in Figure 3.24, we find that there exisits a common factor of intensities that jointly drives all the intensities.

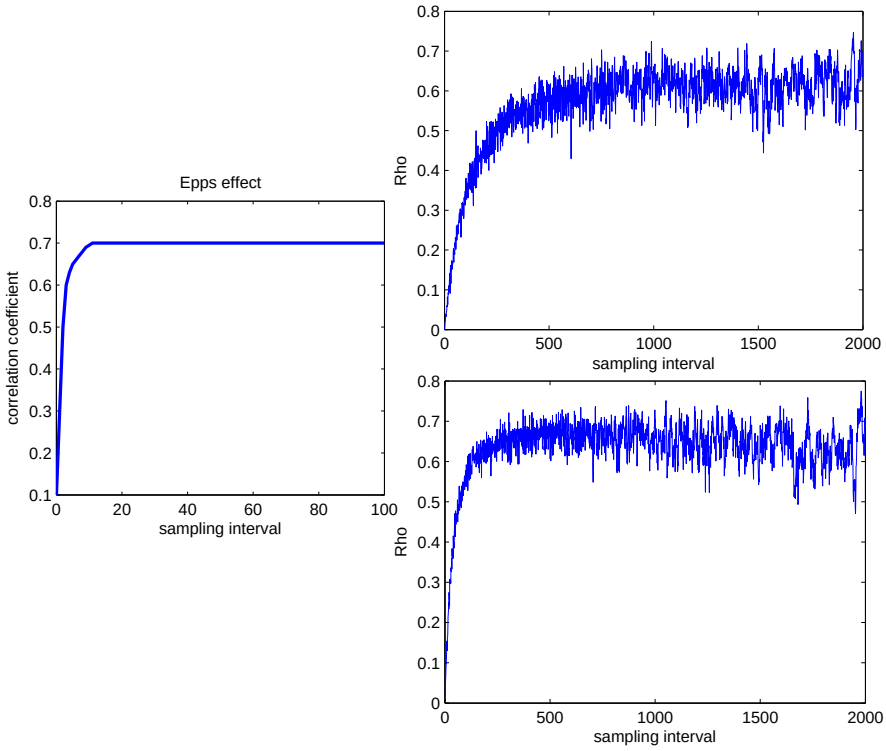

Figure 3.24: Comparison of Epps effect and simulated correlation coefficients. The Epps effect (left panel) is the same as shown in Figure 1.11. The simulated correlation coefficients (right panels) are the same as shown in Figure 3.17. The time generating process is based on the independent DSPP and the logreturn process is according to VAR(1) (top right panel). The time generating process is based on the common factor model and the logreturn process is based on VAR(1) (bottom right panel).

Chapter 4

Filtering by reversible jump Markov chain Monte Carlo methods

4.1 Introduction

As introduced in Chapter 3, our model is based on a class of marked doubly stochastic Poisson processes (DSPPs), which are characterized by the number of events in any time interval as being conditionally Poisson distributed given another positive stochastic process (called intensity). In this multivariate model, we assume that the intensities behind the DSPPs are themselves the sum of specific and common components, which are generalization of the classical shot noise process (Cox & Isham, 1980), which in turn moves up with jumps and tails off with exponential decay. For this component we assumed that the times and sizes of jumps have a probability distribution that can be expressed analytically. Univariate marked doubly stochastic Poisson processes with shot noise intensity have found applications in various fields, such as quantum electronics (Teich & Saleh, 2000), insurance (Dassios & Jang, 2003) and finance (Rydberg & Shephand, 2000; Duffie & Garleanu, 2001).

In this chapter, we concerns on filtering of the underlying (unobservable) intensities and components. Following Centanni & Minozzo (2006), in the univariate case, the filtering problem can be stated in terms of the computation of the conditonal distribution of the whole intensity, in a given time interval, given a realized trajectory, on the same interval, of the MPP $(T_i, Z_i)_{i\in\mathbb{N}}$, where $T_1, \ldots, T_N$ is a sequence of times and $Z_1, \ldots, Z_N$ are the associated marks. This is a nonlinear problem that does not admit an explicit solution and that is usually tackled by approximation or simulation techniques. Centanni & Minozzo (2006) proposed a simulation method based on reversible jump Markov chain Monte Carlo (RJMCMC) algorithm tailored to a particular class of marked DSPPs. Due to the structure of our model, we will follow and extend this approach to filter both the intensities and their components.

The RJMCMC algorithm was first introduced by Green (1995) in a Bayesian inferencial context for the simulation of the posterior distribution. The fact is that the RJMCMC algorithm plays a central role in the filtering context by supplying a simulated sample from the conditional distribution of the intensity process, over any fixed time interval, given an observed trajectory of the process $(T_i, Z_i)_{i\in\mathbb{N}}$. This sample can be viewed, in the optic of particle filters (Pitt & Shephard, 1999), as a discrete approximation with random support of

the distribution of interest. From this approximate distribution, the conditional expectation of the intensity, given the observations, at any given time point, can be obtained. Generally, this approximate distribution can be presented using some density estimation tool. This estimate of the density function can then be used for decision making, classification, and for summarizing Bayesian posteriors.

Notice that this filtering method does not require the intensity to be a Markov process and can be implemented recursively in time as data arrives continuously. Nevertheless, in the case in which the intensity is a deterministic function of the MPP $(T_i, Z_i)_{i\in\mathbb{N}}$, this intensity is indeed a Markov process.

In the literature, various filtering techniques, tailored to the particular situation at hand, have been proposed for models based on marked DSPPs. Among many others, Richardson & Green (1997), Dellaportas et al. (1998), Denison et al. (1997), Troughton & Godsill (1997), Insua & Müller (1998), Barbieri & O'Hagan (1996) and Huerta & West (1999) applied the reversible jump sampler to mixture models, variable selection, curve fitting, autoregressive models, neural networks, autoregressive moving average (ARMA) models and component structure in autoregressive (AR) models, respectively. In particular, in the financial literature, Frey & Runggaldier (2001) proposed a filtering technique based on the reference probability method for a model in which the intensity depends on the level of an unobserved Markovian state process. Rydberg & Shephard (2000) proposed a particle filtering method based on the auxiliary sampling importance resampling algorithm of Pitt & Shephard (1999) for a DSPP with unknown Markovian intensity. In this particle filtering technique, it needs to aggregate the time events in time intervals of fixed length, called bins.

Finally, it is worth stressing that Green's RJMCMC algorithm can be viewed as a development of Markov chain Monte Carlo (MCMC) methods, or so-called advanced MCMC methods. This is due to the fact that the MCMC method for Bayesian computation is usually restricted to problems where the joint distribution of all variables has a density with respect to some fixed standard underlying measure. Nevertheless, the RJMCMC algorithm dismisses the limit of the Bayesian model determination and allows transdimensional MCMC. In the sense that this algorithm allows samples from different parameter spaces with different dimensions. Before go through the RJMCMC algorithm, we survey MCMC methods and discuss an application of MCMC to a simple model.

The remainder of this chapter is organized as follows. Section 4.2 introduces MCMC algorithms and the most popular MCMC algorithms — Metropolis-Hastings algorithms — will be discussed. Section 4.3 introduces RJMCMC algorithm. In Section 4.4, we present a filtering method based on RJMCMC algorithm and its implementation on our specifed model. In addition, we also illustrate the simulation results based on the bivariate modeling framework, some drawbacks of this implementation will be discussed. Section 4.5 deals with the alterations in the implementation of RJMCMC algorithm and the simulation results will be presented.

4.2 Markov chain Monte Carlo algorithms

In a Bayesian setting, inference is based on the posterior density for the parameters of interest. In many situations to obtain this posterior we need to solve high-dimentional integrals, this may be computationaly very difficult. One way to avoid this integration

is that to resort to Markov chain Monte Carlo methods, which attempt to simulate direct draws from the posterior distribution of interest. What is important in these methods is to construct so that the Markov chain so that it sweeps around the most important regions of the sample space. In particular, they are constructed so that the samples generated from them can be considered as mimic samples from the target distribution.

A general strategy to construct MCMC algorithms has been proposed by Hastings (1970), which is developed based on the remarkable work done by Metropolis, Rosenbluth, & Teller (1953). In the literature, the most general used class of schemes for stochastic simulation using Markov chains is referring to Metropolis-Hastings algorithms, which is under a generic name of Metropolis-Hastings. Naturally, this is acknowledge of importance of the contribution from both papers by Metropolis et al.(1953) and Hastings (1970). These are considered as basic papers for the characterization of the method, although other papers including Barker (1965) and Peskun (1973) have also brought relevant contributions to the method.

One particular MCMC method is so called Gibbs sampling (Geman & Geman, 1984), which is very widely applicable to a broad class of Bayesian problems has witnessed a major increase in the application of Bayesian analysis. However, Metropolis-Hastings algorithms is extremely versatile and gives rise to the Gibbs sampler as a special case, as pointed out by Gelman (1992). In the following section, we will focus on the general MCMC method – Metropolis-Hastings algorithms, for the Gibbs sampling (Geman & Geman, 1984), since it can be viewed as a special case of the Metropolis-Hastings algorithms, we would not give a burdensome description, rather refer to an excellent review given by A. Gelfand (2000).

4.2.1 Metropolis-Hastings algorithm

Considering π as our target distribution from which samples can be drawn via Markov chains. the usual approach to Markov chain theory on a continuous state space is to start with a transition kernel $p(x, A)$, for $x \in \mathbb{R}^d$, $A \in \mathcal{B}$, where $\mathcal{B}$ is Borel σ-field on $\mathbb{R}^d$. The transition kernel $p(x, A)$ is a conditional distribution function that represents the probability of moving from x to a point in the set A. By virtue of its being a distribution function, $p(x, \mathbb{R}^d) = 1$, where it is permitted that the chain can make a transition from the point x to x, that is, remain at the same state with probability $p(x, \{x\})$ which is not necessary to be zero.

In statistics, a Markov process is said to satisfy the detailed balance condition if the transition rates between each pair of states (x, x') in the state space obey

$$\pi(x')p(x', x) = \pi(x)p(x, x'), \quad \forall x, x' \in \mathbb{R}^d, \tag{4.2.1}$$

where p is the Markov transition kernel, π is the equilibrium distribution of the chain.

Equation 4.2.1 means that the system (chain) moves from x to x' at the same rate at which it moves from x' to x when the system (chain) is in equilibrium. A Markov process that satisfies the detailed balance equation (Equation 4.2.1) is said to be a reversible Markov process with respect to invariant distribution π. Note that the detailed balance condition (Equation 4.2.1) is a sufficient condition such that π be the equilibrium (invariant) distribution of the chain.

MCMC method is concerning on the matter that the equilibrium distribution is known (perhaps up to a normalising constant) which is the target distribution $\pi(\cdot)$, but the transition kernel p is unknown. The problem then is to find an appropriate $p(x, \mathrm{d}x')$ to satisfy the detailed balance condition (Equation 4.2.1). One way suggested by Metropolis-Hastings algorithms is assuming the kernel $p(x, \mathrm{d}x')$ consists of two elements: an arbitrary proposal transition $q(x, \mathrm{d}x')$ and the probability $\alpha(x, x')$ such that

$$p_{MH}(x, \mathrm{d}x') = q(x, \mathrm{d}x') \cdot \alpha(x, x'), \quad \text{if} x \neq x', x, x' \in \mathbb{R}^d, \tag{4.2.2}$$

where $p_{MH}(x, \mathrm{d}x')$ refers to Metropolis-Hastings transition kernel, $q(x, \mathrm{d}x')$ is proposal transition, also known as candidate-generating density, $\alpha(x, x')$ is acceptance probability.

Given the definition of transition kernel (Equation 4.2.2), the detailed balance condition (Equation 4.2.1) can be rewritten as

$$\pi(x')q(x', \mathrm{d}x)\alpha(x, x') = \pi(x)q(x, \mathrm{d}x')\alpha(x', x). \tag{4.2.3}$$

One tricky made in Equation 4.2.3 is getting rid of the transition kernel $p(x, \mathrm{d}x')$ which is actually unknown and place an arbitrary proposal transition $q(x, \mathrm{d}x')$ to keep the chain move ahead. In order to garantee the reversibility of the chain, the acceptance probability $\alpha(x, x')$ plays a role to correct this problem. For this acceptance probability $\alpha(x, x')$, Hastings (1970) proposed to define it in such a way that combined with the arbitrary transition proposal q, the expression is given by

$$\alpha(x, x') = \min\left\{1, \frac{\pi(x')q(x', x)}{\pi(x)q(x, x')}\right\}. \tag{4.2.4}$$

Assume that our target distribution is $\pi(\mathrm{d}x)$ and we want to generate a draw from the transition kernel $p(x, \mathrm{d}x')$ with invariant measure π. The Metropolis-Hastings algorithms allow us to draw samples from $p(x, \mathrm{d}x')$ knowing π only up to a normalizing constant, $\pi(x) = f(x)/h$, where $x \in \mathbb{R}^d$, $f(x)$ is unnormalized density and h is the (possibly unknown) normalizing constant. Suppose that we can draw $x' \sim q(x, \mathrm{d}x')$, a proposal transition for x' with $\int q(x, \mathrm{d}x')\mathrm{d}x' = 1$. In practice, starting with an initial value x_0 satisfying $\pi(x_0) > 0$ and assuming that Markov chain is in state x_j, the j^{th} iteration $(j = 1, 2, \ldots)$ of the Metropolis-Hastings algorithm can be summarized as follows:

1. sample a candidate state x' from a proposal transition $q(x_j, \cdot)$;
2. given the candidate state x', calculate the acceptance probability $\alpha(x_j, x')$

$$\alpha(x_j, x') = \min\left\{1, \frac{\pi(x')q(x', x_j)}{\pi(x_j)q(x_j, x')}\right\};$$

3. draw $u \sim \mathrm{U}(0, 1)$ and if $u < \alpha(x_j, x')$, then $x_{j+1} = x'$, otherwise $x_{j+1} = x_j$;
4. set $j = j + 1$ and return to step 1 until the convergence is reached.

This algorithm provides a way to construct the transition kernel $p(x_j, x')$, where

$$p(x_j, x') = q(x_j, x') \cdot \alpha(x_j, x'),$$

such that the detailed balance (Equation 4.2.1) is satisfied with the target distribution $\pi(\cdot)$ as the invariant distribution. Note that the detailed balance (Equation 4.2.1) is satisfied by the transition kernel $p(\cdot,\cdot)$ but not the proposal $q(\cdot,\cdot)$. In general, $\pi(x)q(x,x') \neq \pi(x')q(x',x)$, if, for example, $\pi(x)q(x,x') > \pi(x')q(x',x)$, according to Equation 4.2.4, the acceptance probability $\alpha(x,x')$ might be very small, it is more likely the new proposed value x' would be rejected. Given the reversibility of the process, the Markov chain moves from x to x' too often and from x' to x too rarely. However, Metropolis-Hastings algorithms correct this problem with the acceptance probability $\alpha(\cdot,\cdot)$ so that reduce the number of moves from x to x', see Equation 4.2.3.

One advantage of Metropolis algorithm is keeping the transition proposal arbitrary and thus providing a flexibile tool for the construction of the algorithm. As in any MCMC method, the draws are regarded as a sample from the target ditribution π only after the chain has passed the transient period (or burn-in period) and the effect of the fixed starting value x_0 has become so small that it can be ignored. After a sufficient burn-in period, say k iterations, the chain approaches its stationary distribution $\pi(\cdot)$, that is the marginal distribution of x_j will converge to π. In fact, this convergence to the invariant distribution occurs under mild regularity conditions. The regularity conditions require the chain to be irreducible and aperiodic in the sense that, if x and x' are in the domain of $\pi(\cdot)$, irreducibility means that it is possible to move from x to x' in a finite number of iterations with nonzero probability, while aperiodicity means that the number of moves required to move from x to x' is not required to be a multiple of some integer. These conditions are usually satisfied if $q(x, \mathrm{d}x')$ has the same support of $\pi(\cdot)$.

Clearly, Metropolis-Hastings algorithm generates a Markov chain $(x_0, x_1, \ldots, x_k, \ldots)$, such that the transition probablity from x_j to x_{j+1} depends only on x_j but not on the previous values $(x_0, \ldots, x_{j-1})$. Whether the chain is irreducible and aperiodic depends on the choice of proposal distribution, so we must check these conditions diligently for any implementation. If this check confirms irreducibility and aperiodicity, then the chain generated by the Metropolis-Hastings algorithm has a unique limiting stationary distribution. This result would seem to follow the limiting theory of Markov chains. Nevertheless, irreducibility and aperiodicity remain sufficient conditions for convergence of Metropolis-Hastings chains.

To find the unique stationary distribution of an irreducile aperiodic Metropolis-Hastings chain, suppose $x' \sim \pi(x)$, and consider two points in the state space of the chain, say x_1 and x_2, for which $\pi(x_1) > 0$ and $\pi(x_2) > 0$. Without loss of generality, label these points in the manner such that $\pi(x_2)q(x_1|x_2)\pi(x_1)q(x_2|x_1)$.

It follows that that unconditional joint density of $x^{(t)} = x_1$ and $x^{(t+1)} = x_2$ can be written as $f(x_1)q(x_2|x_1)$, because if $x^{(t+1)} = x_1$ and $x' = x_2$ then $\alpha = 1$ so $x^{(t+1)} = x_2$, where $x^{(t)}$ denotes the t^{th} iteration. The unconditional joint density of $x^{(t)} = x_2$ and $x_{(t+1)} = x_1$ is

$$\pi(x_2)q(x_1|x_2)\frac{\pi(x_1)q(x_2|x_1)}{\pi(x_2)q(x_1|x_2)}. \tag{4.2.5}$$

This is ecause we need to start with $x^{(t)} = x_2$, to propose $x' = x_1$, and then to set $x^{(t+1)}$ equal to x' with probability α. Note that Equation 4.2.5 reduces to $\pi(x_1)q(x_2|x_1)$, which matches the joint density of $x^{(t)} = x_1$ and $x^{(t+1)} = x_2$. Therefore, the joint distribution of $x^{(t)}$ and $x^{(t+1)}$ is symmetric. Hence $x^{(t)}$ and $x^{(t+1)}$ have the same marginal distributions.

Thus the marginal distribution of $x^{(t+1)}$ is π, and π must be the stationary distribution of the chain.

As described above, there is a total flexibility for the choice of the proposal transition $q(\cdot,\cdot)$. However, specific features of good proposal distributions can greatly enhance the performance of the Metropolis-Hastings algorithm. A well chosen proposal distribution produces candidates values that cover the support of the stationary distribution in a reasonable number of iterations and, similarly, produces candidate values that are not accepted or rejected too often. Both of those factors are related to the spread of the proposal distribution. If the proposal distribution is too diffuse to the target distribution, the candidate values will be rejected frequently and thus the chain will require many iterations to adequately explore the space of the target distribution. On the other hand, if the proposal distribution is too focused (e.g. has too small variance), then the chain will remain in one small region of the target distribution for many iterations while other regions of the target distribution will not be adequately explored. Thus a proposal distribution whose spread is either too small or too large can produce a chain that requires many iterations to adequately explore the regions supported by the target distribution.

In what follows, we list three most common used in the literature, as the special cases of proposal transition. Note that the special case does not violate the arbitrarity of the proposal transition, yet the arbitrarity allows some special cases.

Choosing a proposal transition $q(x, dx')$

Considerable work devoted to the question of how to choose an appropriate transition kernel has been done in the past. Some special cases of probability transition $q(\cdot,\cdot)$, such as symmetric chains, random walk chains and independent chains, have been studied. Below we introduce these thress different classes of proposal transitions.

Symmetric chains. A chain is said to be symmetric if its transition kernel p is symmetric in its arguments, namely, $p(x, x') = p(x', x)$, for every pair (x, x') of states. For the Metropolis-Hastings algorithms, the notion of symmetric is applied to the proposed transition q. An example of a symmetric chain is the Metropolis version of the algorithm. If q dependents on (x, x') only through $|x - x'|$ then $q(x, x') = q(x', x)$. In this case, the acceptance probability reduces to $\min\{1, \pi(x')/\pi(x)\}$ and does not depend on q. A computational simplification is then achieved.

Random walk chains. Construct a random walk chain is proposed by Metropolis et al. (1953), in which a Markov chain is evoluted by $x^{(j)} = x^{(j-1)} + y_j$ where y_j is a random variable with distribution independence of the chain. In general, the disturbances y_j are independent and indentically distributed with density f_y. The chain has proposed moves according to $q(x, x') = f_y(x' - x)$. If f_y is symmetric around 0, the chain is symmetric and all comments above are valide here. So symmetric chain can be seen as a special case of a random chain.

In this case, the proposed value x' is then based around the previous value x of the chain. An important point still remaining is the choice of the dispersion of f_y. Large values for the variance allow moves that are very distant from previous values but at the likely cost of very small acceptance rates. On the other hand, small values for the variance only allow moves

close to the previous values but with high acceptance rates. Tierney (1994) suggested setting the variance matrix of f_y as $c \cdot V$ where c is a multiplying scalar playing the role of a tunning constant and V is some form of approximation for the posterior variance. This allows the moves along the components of x to be of the same size relative to the spread of the posterior distribution.

In general, common choices for f_y include a uniform distribution over a ball centered at the origin, a scaled normal distribution, and a scaled Student's t distribution. If the support region of π is connected and f_y is positive in a neighbourhood of 0, the resulting chain is irreducible and aperiodic (see Roberts & Casella, 1999).

Independent chains. Suppose that the proposal distribution for the Metropolis-Hastings is chosen such that $q(x, x') = g(x')$ for some fixed density g. This yields an independent chain, where each candidate value is drawn independent of the past. It may seem that the proposed value x' independent from the previous state x disagrees with the Markovian property of the chain. It is worthing remembering that q enters the detailed balance condition jointly with an acceptance probability α. For the independent chain, Equation 4.2.3 can be repressed as

$$\pi(x)g(x')\alpha(x, x') = \pi(x')g(x)\alpha(x', x), \tag{4.2.6}$$

and the acceptance probability as

$$\alpha(x, x') = \min\left\{1, \frac{\pi(x')g(x)}{\pi(x)g(x')}\right\}. \tag{4.2.7}$$

Hence, this transition from state x to x' depends on the previous state x through the acceptance probability (the Markov structure is preserved). Moreover, the resulting Markov chain is irreducible and aperiodic if $g(x) > 0$ whenever $\pi(x) > 0$.

One popular choice for g is the prior density, in this case, the acceptance probability is reduced to

$$\alpha(x, x') = \min\left\{1, \frac{l(x')}{l(x)}\right\}, \tag{4.2.8}$$

where $l(x)$ is the likelihood function. In other words, we propose from the prior and the acceptance ratio (or so-called Metropolis-Hastings ratio) $A = l(x')/l(x)$ equals to the likelihood ratio. By definition, the support of the prior covers the support of the target distribution, so the stationary distribution of this chain is the desired posterior. There are often more specialized MCMC algorithms to sample various types of posteriors in more efficient manners, but this is perhaps the simplest generic approach. To this end, it becomes a particularly popular tool for Bayesian inference. Since it allows to obtain a sample from a Markov chain whose stationary distribution is the target posterior, and this sample may be used to estimate the posterior density and many other useful quantities, such as posterior moments, tail probability etc.

Finally, it is worth pointing out that although a chain is defined by its transition kernel $p(\cdot, \cdot)$ and not by a proposal transition $q(\cdot, \cdot)$, the names used to categorize the algorithm generally refer to properties of $q(\cdot, \cdot)$ rather than $p(\cdot, \cdot)$. The previous three cases address this problem quite clear.

Filtering of a simple inhomogeneous Poisson process

To understand Metropolis-Hastings algorithms, let us consider a trival example which is a simple case of our model (as described in Chapter 3). What we are interest in is to apply Metropolis-Hastings algorithm to infer the parameters of interest. In this case, we use independent chain sampling approach for proposal distribution and the posterior distribution as the target distribution.

Now consider a simple factor model for the intensity. Basically, the intensities $\tilde{\lambda}^{01}$ and $\tilde{\lambda}^{02}$ are assumed as a linear combination of the components $\tilde{\lambda}^0$, $\tilde{\lambda}^1$, $\tilde{\lambda}^2$, which is given by

$$\begin{cases} \tilde{\lambda}^{01} = \tilde{\lambda}^1 + a_1\tilde{\lambda}^0, \\ \tilde{\lambda}^{02} = \tilde{\lambda}^2 + a_2\tilde{\lambda}^0, \end{cases} \tag{4.2.9}$$

where $\tilde{\lambda}^1 = \nu_1 x_1$, $\tilde{\lambda}^2 = \nu_2 x_2$, $\tilde{\lambda}^0 = \nu_0 x_0$, $x_i \sim \Gamma(\alpha_i, \beta_i)$, and ν_i, α_i, β_i are parameters, $i = 0, 1, 2$. Note that $\tilde{\lambda}^k$ are specific components, $k = 1, 2$, $\tilde{\lambda}^0$ is the common component. The coefficient a_k are scalar parameters which drive the common component $\tilde{\lambda}^0$ impact on the specific component $\tilde{\lambda}^k$ contemporaneously.

In order to mimic the financial data, we also assume that the return processes follow independent Wiener process and the associated transaction times $\{t_j^k\}_{j\in\mathbb{N}}$, are Poisson distributed with underlying intensity $\tilde{\lambda}^{0k}$, $k = 1, 2$, on the time interval $(0, T]$. Thus, the return processes become

$$\begin{cases} R^1_{t_j^1} = \mu_1 + \xi^1_{t^1_{j-1}}, \\ R^2_{t_j^2} = \mu_2 + \xi^2_{t^2_{j-1}}, \end{cases} \tag{4.2.10}$$

where $\xi^k_{t_j} \sim \mathbf{N}\Big(0, (t_j^k - t_{j-1}^k)\Big)$, and $\mu_k = 0$, $k = 1, 2$.

In this specific bivariate model, the inter-arrival times $t_j^k - t_{j-1}^k$ between two trades are independent and identically exponentially distributed with mean $\tilde{\lambda}^{0k}$, $k = 1, 2$. Note that inter-arrival time depends on $\tilde{\lambda}^{0k}$, and so in turn, on the single components $\tilde{\lambda}^i$, where $i = 0, 1, 2$. The variance for each return process $\{R_j^k\}_{j\in\{1,2,\ldots,n^k\}}$ is $t_j^k - t_{j-1}^k$, where N^k denotes the number of transactions for kth return. This implies that the return process R_t^k would be influented by the intensity $\tilde{\lambda}^{0k}$. The higher intensity $\tilde{\lambda}^{0k}$, the more transactions might be observed and the less variance of the returns may be found. On the other hand, the lower intensity $\tilde{\lambda}^{0k}$, the less transactions might be observed and the more variance of the returns may be found. Hence, the speed of the transaction is incorporated into the intensity.

Practically, we observe returns and the assosciating transaction times $(R_j^k, t_j^k)_{j\in\{1,\ldots,n^k\}}$, $k = 1, 2$, but we do not observe the underlying intensities $\tilde{\lambda}^{0k}$, and their components, $\tilde{\lambda}^0$, $\tilde{\lambda}^1$, $\tilde{\lambda}^2$, either. The objective is then to predict the unobservable intensities, not only $\tilde{\lambda}^{01}$ and $\tilde{\lambda}^{02}$ but also components, $\tilde{\lambda}^0$, $\tilde{\lambda}^1$, $\tilde{\lambda}^2$, given the observation $(R_j^k, t_j^k)_{j\in\{1,\ldots,n^k\}}$, $k = 1, 2$. Notice that under this simple factor model of the intensity (Formular 4.2.9), it is easy to attain $\tilde{\lambda}^{01}$, $\tilde{\lambda}^{02}$ if the components $\tilde{\lambda}^0$, $\tilde{\lambda}^1$, $\tilde{\lambda}^2$, are known. Therefore, our objective is reduced to predict the components $\tilde{\lambda}^0$, $\tilde{\lambda}^1$, $\tilde{\lambda}^2$.

Accordingly, we have three unobservable intensities (components) which we want to predict on the basis of two sequences of transaction times and returns. A graphic example is illustrated in Figure 4.1 (setting parameter $\alpha_1 = 2$, $\beta_1 = 5$, $\alpha_2 = 3$, $\beta_2 = 1$, $\alpha_0 = 5$, $\beta_0 = 0.5$, $\nu_1 = 2$, $\nu_2 = 1$, $\nu_0 = 1$, $a_1 = 1$, $a_2 = 0.5$. The upper left shows the constant trajectory

of $\tilde{\lambda}^1$ and the upper right gives constant trajectory of $\tilde{\lambda}^2$. The middle one is the trajectory of common component $\tilde{\lambda}^0$. The bottom two figures present the return processes R_t^1 (left) and R_t^2 (right), where each dot can be interpreted as one transaction which contains its occurrence time and the associated return. Though we know all the parameters, we do not know the realized values x_i of the Gamma distributed variables $\mathbf{X}_i$, $i = 0, 1, 2$). To this end, the number of transaction times N^k on the time interval $(0, T]$ follows inhomogeneous Poisson process, where $k = 1, 2$.

From Figure 4.1, the bottom two return processes show the frequency of the transactions based on the number of dots. As we can see, the first return (Return1) is much more liquid than the second one (Return2), almost 3 times more. This implies that the intensity $\tilde{\lambda}^{01}$ is greater than $\tilde{\lambda}^{02}$, in fact, it is true if we check the realized value of components (see the upper and the middle figures in the Figure 4.1) with $a_1 = 1 > a_2 = 0.5$, according to Equation 4.2.9, we obtain the realized value of the underlying intensity $\tilde{\lambda}^{01} = 15.91 > \tilde{\lambda}^{02} = 4.435$. Thus, the frequency of the transactions contain the information of the unobserved intensities. To predict the unobserved quantities, $\tilde{\lambda}^0$, $\tilde{\lambda}^1$, $\tilde{\lambda}^2$, given the observations, $(R_j^k, t_j^k)_{j\in\{1,\dots,n^k\}}$, $k = 1, 2$, we use independent Metropolis-Hastings algorithms and consider posterior density as the target distribution.

Before applying the Metropolis-Hastings algorithms to our simple model, one interesting issue may be addressed here with the following question: is this Metropolis-Hastings algorithm able to distinguish three components $\tilde{\lambda}^0$, $\tilde{\lambda}^1$, $\tilde{\lambda}^2$, given the two sequencies of observation $(R_j^k, t_j^k)_{j\in\{1,\dots,n^k\}}$, $k = 1, 2$? As a matter of fact, from the factor model of the intensities (Formular 4.2.9), the intensities, $\tilde{\lambda}^{01}$, $\tilde{\lambda}^{02}$, incorporate the full information of the transaction times but not the components, $\tilde{\lambda}^0$, $\tilde{\lambda}^1$, $\tilde{\lambda}^2$. Therefore, given the transaction times, it is not difficult to infer the intensity $\tilde{\lambda}^{01}$ and $\tilde{\lambda}^{02}$ by Metropolis-Hastings algorithms treating the posterior distribution as the target distribution. However, it is quite tricky to predict each component $\tilde{\lambda}^0$, $\tilde{\lambda}^1$, $\tilde{\lambda}^2$, since the transaction times do not contain any information about the components. If this is the case, we may probabily meet the problem of overprediction for one component whilst underprediction for the other.

To verify this plausible issue, we run simulation based on Metropolis-Hastings algorithm. Assuming that we do not have knowledge about intenities, neither components $\tilde{\lambda}^0$, $\tilde{\lambda}^1$, $\tilde{\lambda}^2$, nor intensity as a whole $\tilde{\lambda}^{01}$, $\tilde{\lambda}^{02}$, but we know they are constant trajectories. We have full knowledge of the data $(r_j^k, t_j^k)_{j\in\{1,\dots,n^k\}}$, $k = 1, 2$, as shown on the bottom two figures of Figure 4.1, and also know the parameters $\alpha_1 = 2$, $\beta_1 = 5$, $\alpha_2 = 3$, $\beta_2 = 1$, $\alpha_0 = 5$, $\beta_0 = 0.5$, $\nu_1 = 2$, $\nu_2 = 1$, $\nu_0 = 1$, $a_1 = 1$, $a_2 = 0.5$. Our task is to predict the components of the intensity, given the data. Graphically, we want to know the values shown on the top and the middle of Figure 4.1, given the data as shown at the bottom of Figure 4.1.

Given the arbitrary propery of the proposal transition $q(\cdot, \cdot)$, we choose an arbitrary exponential distribution as our candidate density, where the marginal distributions are independent and identical exponential distribution with mean $\lambda_0 = 1.5$, $\lambda_1 = 2$, $\lambda_2 = 2.5$, respectively. In this case, we would apply independent chain sampling approach, so $q(\mathbf{Y}, \mathbf{Y}^*) = g(\mathbf{Y}^*)$, where $\mathbf{Y} = (y_0, y_1, y_2)$ and $\mathbf{Y}^* = (y_0^*, y_1^*, y_2^*)$. Generate y_i^* from independent and identical exponential distribution with mean λ_i, $i = 0, 1, 2$. Then calculation the acceptance probability

$$\alpha(\mathbf{Y}, \mathbf{Y}^*) = \min\left\{1, \frac{\pi(\mathbf{Y}^*)q(\mathbf{Y}^*, \mathbf{Y})}{\pi(\mathbf{Y})q(\mathbf{Y}, \mathbf{Y}^*)}\right\} = \min\left\{1, \frac{\pi(\mathbf{Y}^*) \cdot g(\mathbf{Y})}{\pi(\mathbf{Y}) \cdot g(\mathbf{Y}^*)}\right\},$$

where $q(\mathbf{Y}, \mathbf{Y}^*) = g(\mathbf{Y}^*) = (\lambda_0 e^{-\lambda_0 y_0^*}) \cdot (\lambda_1 e^{-\lambda_1 y_1^*}) \cdot (\lambda_2 e^{-\lambda_2 y_2^*})$. Note that $q(\cdot)$ is not symmetric, therefore, we must apply Metropolis-Hastings sampling.

Considering the posterior distribution as the target distribution which is given by

$$\pi(\mathbf{Y}|\mathbf{t}) = \frac{f(\mathbf{t}|\mathbf{Y}) \cdot f(\mathbf{Y})}{\int_\Theta f(\mathbf{t}|\mathbf{Y}') \cdot f(\mathbf{Y}')\mathrm{d}\mathbf{Y}'}, \tag{4.2.11}$$

where $\mathbf{t} = \{t_j^k\}_{j\in\{1,2,\ldots,n^k\}}$, N^k denotes the number of transaction times for kth asset, $k = 1, 2$.

The target distribution π enters the algorithm through the acceptance ratio A in the form of the ratio $\pi(\mathbf{Y}^*)/\pi(\mathbf{Y})$ as in the resampling methods, so the complete knowledge of π is not required, in particular, the normalising constant $\int_\Theta f(\mathbf{t}|\mathbf{Y}) \cdot f(\mathbf{Y})\mathrm{d}\mathbf{Y}$ is not needed. When π is a posterior density, even though its functional form is always known, the value of the propotionality constant is usually unknown or difficult to calculate. So, the algorithm is particularly useful for applications to Bayesian inference.

The acceptance probability can be rewritten as

$$\alpha = \min\left\{1, \frac{f(\mathbf{t}|\mathbf{Y}^*)f(\mathbf{Y}^*) \cdot g(\mathbf{Y})}{f(\mathbf{t}|\mathbf{Y})f(\mathbf{Y}) \cdot g(\mathbf{Y}^*)}\right\},$$

where $f(\mathbf{t}|\mathbf{Y}^*) = \tilde{\lambda}^{*01} e^{-\tilde{\lambda}^{*01}T} \cdot \tilde{\lambda}^{*02} e^{-\tilde{\lambda}^{*02}T}$, $\tilde{\lambda}^{*01} = \nu_1 y_1^* + a_1 \nu_0 y_0^*$ and $\tilde{\lambda}^{*02} = \nu_2 y_2^* + a_2 \nu_0 y_0^*$. $f(\mathbf{Y}^*) = \mathrm{C} \cdot y_0^{*(\alpha_0-1)} \cdot y_1^{*(\alpha_1-1)} \cdot y_2^{*(\alpha_2-1)}$ and $f(\mathbf{Y}) = \mathrm{C} \cdot y_0^{(\alpha_0-1)} \cdot y_1^{(\alpha_1-1)} \cdot y_2^{(\alpha_2-1)}$, where C is a constant and is equivalent to $\beta_0^{\alpha_0}/\Gamma(\alpha_0) \cdot \beta_1^{\alpha_1}/\Gamma(\alpha_1) \cdot \beta_2^{\alpha_2}/\Gamma(\alpha_2)$.

Simulation results are presented on Figure 4.2, inference for x_0, x_1, x_2 can be based on the sample obtained from running the chain. After 100,000 iterations on Metropolis-Hastings algorithm and discarding the first 50,000 iterations (as burn-in period), the posterior means are approximately given by 3.8, 6, 2.5 respectively. This is very close to the true one, $x_0 = 3.85$, $x_1 = 6.03$, $x_2 = 2.51$. However, the samples from Metropolis-Hastings algorithm is quite disperse, take x_0 for instance, range from 0.6 to 8.5, but the arithmetic mean is around the true value 3.8. Note that we do not expected to have a single point with high frequency but rather to have some diversity such that exploring the sampling space.

Recall the issue addressed before, concerning on whether Metropolis-Hastings algorithm is able to distinct components of intensity, given the knowledge of the transactions. From the simulation results, we find this algorithm perform well on the question of interest. This may probably devote to samples provided by Metropolis-Hastings algorithm are succeed to explore the region of the target distribution (posterior distribution in the example), where the chain does not remain in one small region of the target distribution but also other regions of the target distribution. After sufficient long iterations, the chain converge to the target distribution which is concentrate in the area of interest.

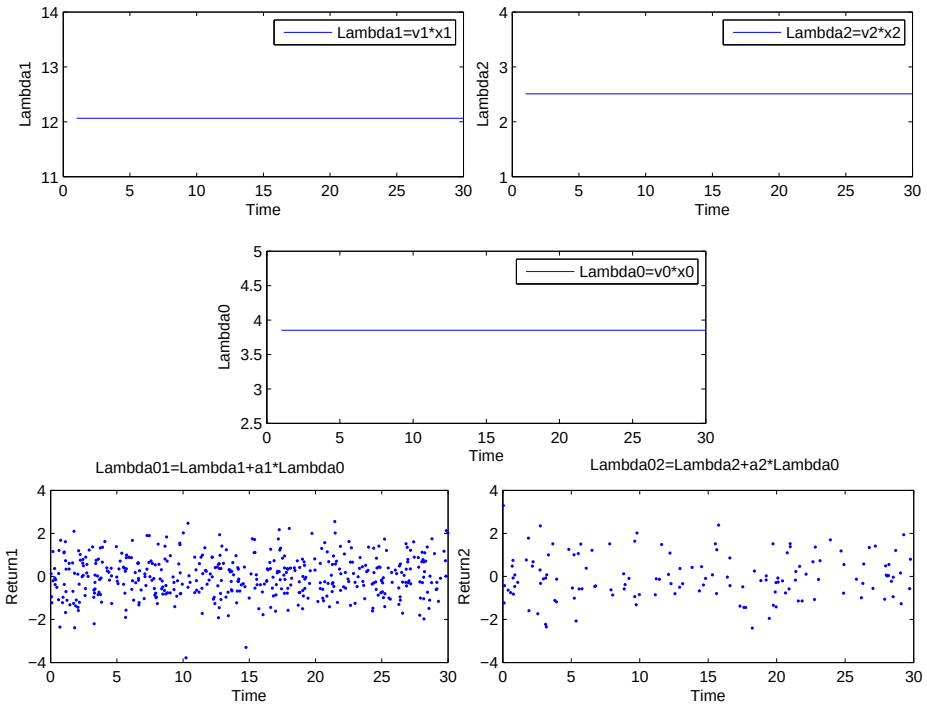

Figure 4.1: The true (simulated) specific component of the intensity, $\lambda_1 = \nu_1 x_1$, with $x_1 \sim \Gamma(\alpha_1, \beta_1)$ (upper left). The true (simulated) specific component of the intenity, $\lambda_2 = \nu_2 x_2$, with $x_2 \sim \Gamma(\alpha_2, \beta_2)$ (upper right). The true (simulated) common component of intensity, $\lambda_0 = \nu_0 x_0$, with $x_0 \sim \Gamma(\alpha_0, \beta_0)$ (on the middle). The return process under the combined intensity λ^{01} (bottom left). The return process under the combined intensity λ^{02} (bottom right). Set parameters as: $\alpha_1 = 2$, $\beta_1 = 5$, $\alpha_2 = 3$, $\beta_2 = 1$, $\alpha_0 = 5$, $\beta_0 = 0.5$, $\nu_1 = 2$, $\nu_2 = 1$, $\nu_0 = 1$, $a_1 = 1$, $a_2 = 0.5$ (Equation 4.2.9 and 4.2.10).

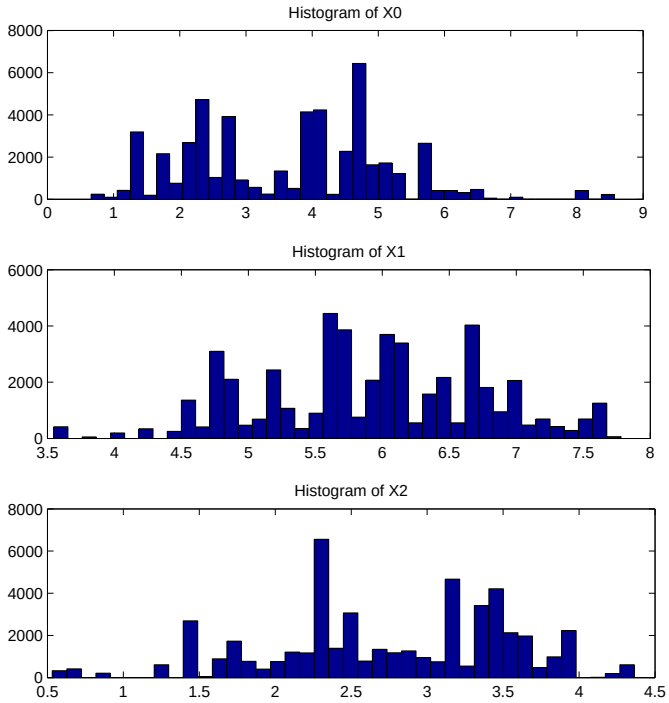

Figure 4.2: Run of 100,000 iterations of Metropolis-Hastings algorithm with 50,000 burn-in period. Histogram of x_0 (top). Histogram of x_1 (middle). Histogram of x_2 (bottom). The true value: $x_0 = 3.85$, $x_1 = 6.03$, $x_2 = 2.51$.

4.3 The reversible jump MCMC algorithm

In the previous section, we considered MCMC methods for simulating x' from a Markov chain with a stationary distribution $\pi(\cdot)$. The methods described in Section 4.2 required that the dimensionality of x' (i.e. of its state space) does not change with different iterations. It may be of interest to develop a chain that allows for changes in the dimension of the parameter space from one iteration to the next. As described in the introduction, Green's reversible jump Markov chain Monte Carlo method permits transdimentional Markov chain Monte Carlo simulation. We explain this simulation method in the context of Bayesian selection.

Now consider constructing a Markov chain to explore a space of candidate models, each of which might be used to fit observed data $\mathbf{y}$. Let $\{\mathcal{M}_j\}_{j\in\{1,2,\ldots,J\}}$ be a family of M models under consideration. A parameter vector $\boldsymbol{\theta}_j$ denotes the parameters in j^{th} model. Different models may have different numbers of parameters, so we let n_j denote the number of parameters in the j^{th} model. A more formal treatment is given to the problem of choosing between models. This is done by indexing all models under consideration by j and treating this index j as another parameter to be treated jointly with all other model parameters, in the Bayesian paradigm, we can define random variables $\mathbf{C} = (j, \boldsymbol{\theta}_j)$ as a index for that model. We may assign prior distributions to these parameters, then seek to simulate from their posterior distribution using a MCMC method for which the t^{th} random draw is $\mathbf{C}^{(t)} = (j^{(t)}, \boldsymbol{\theta}^{(t)}_{j^{(t)}})$, where $\boldsymbol{\theta}^{(t)}_{j^{(t)}}$ denotes the parameters drawn for the model indexed by $j^{(t)}$, has dimension $n_{j^{(t)}}$ that can vary with t. For simplicity, we avoid the treatment of nonparametetric model averaging techniques.

The goal of RJMCMC is to generate samples with joint posterior density $f(j, \boldsymbol{\theta}_j|\mathbf{y})$. This posterior arises from Bayes' theorem via

$$f(j, \boldsymbol{\theta}_j|\mathbf{y}) \propto f(\mathbf{y}|j, \boldsymbol{\theta}_j) f(\boldsymbol{\theta}_j|j) f(j), \tag{4.3.1}$$

where $f(\mathbf{y}|m, \boldsymbol{\theta}_j)$ denotes the density of the observed data under the j^{th} model and its parameters, $f(\boldsymbol{\theta}_j|j)$ denotes the prior density for the parameters in the j^{th} model, and $f(j)$ denotes the prior density of the j^{th} model satisfying $\sum_{j=1}^{J} f(j) = 1$.

RJMCMC enables to construct an appropriate Markovian process for $\mathbf{C}$ that jumps between models with parameter spaces of different dimensions. Like simpler MCMC methods, RJMCMC proceeds with the generation of a proposed step from the current state $\mathbf{C}^{(t)}$ to $\mathbf{C}^{*}$, and then a decision whether to accept the proposal or to keep the previous. The stationary distribution for our chain is based on the specification of a transition kernel $p(x, dx')$ satisfying the detailed balance equation

$$\int_A \int_B \pi(\mathrm{d}\mathbf{C}) p(\mathbf{C}, \mathrm{d}\mathbf{C}') = \int_B \int_A \pi(\mathrm{d}\mathbf{C}') p(\mathbf{C}', \mathrm{d}\mathbf{C}), \tag{4.3.2}$$

for all appropriate A, B where the move is allowed. Chains that satisfy this detailed balance condition are termed reversible, because the direction of time does not matter in the dynamics of the chain. After long run of the simulation, this chain can obtain dependent, approximate sample from $\pi(\mathrm{d}\mathbf{C})$.

The transition will be constructed in two stages: a proposal transition and an acceptance probability, correcting the proposal to ensure the detailed balance condition. The

main difference here is that many models are simultaneouly being considered, and therefore many qualitatively different moves are involved. Green (1995) explore this idea by imposing detailed balance for all possible moves between models. Thus, detailed balance would be attained globally.

Consider for each move r between models, an arbitrary transition kernel $q_r(\mathbf{C}, \mathrm{d}\mathbf{C}')$ and a yet to be specified acceptance probability $\alpha_r(\mathbf{C}, \mathbf{C}')$. Green (1995) admits that the chain may not move at every iteration such that $\sum_r q_r(x, \boldsymbol{\ell}) \leq 1$, where $\boldsymbol{\ell}$ is the combined parameter space $\boldsymbol{\ell} = \cup_{j \in \Im}$ and $\Im = 1, 2, \ldots, J$. Jump moves r are proposed according to probability $q_r(\mathbf{C}, \boldsymbol{\ell})$ and there is also a probability $1 - \sum_r q_r(\mathbf{C}, \boldsymbol{\ell})$, no change to the present state is proposed. Naturally, it is possible to have $\sum_r q_r(\mathbf{C}, \boldsymbol{\ell}) = 1$ and in this case, a move will always be proposed.

The transition kernel can be written as

$$P(\mathbf{C}, B) = \sum_r \int_B q_r(\mathbf{C}, \mathrm{d}\mathbf{C}')\alpha_r(\mathbf{C}, \mathbf{C}') + s(\mathbf{C})I(\mathbf{C} \in B), \tag{4.3.3}$$

where B is Borel sets in $\boldsymbol{\ell}$, $I(\cdot)$ denotes the indicator function, and

$$\begin{aligned} s(\mathbf{C}) &= 1 - \sum_r \int_{\boldsymbol{\ell}} q_r(\mathbf{C}, \mathrm{d}\mathbf{C}')\alpha_r(\mathbf{C}, \mathbf{C}') \\ &= \sum_r \int_{\boldsymbol{\ell}} q_r(\mathbf{C}, \mathrm{d}\mathbf{C}')[1 - \alpha_r(\mathbf{C}, \mathbf{C}')] + 1 - \sum_r q_r(\mathbf{C}, \boldsymbol{\ell}), \end{aligned} \tag{4.3.4}$$

is the probability of remaining at previous state $\mathbf{C}$.

To satisfy the detailed balance requirement (Equation 4.3.2), the equilibrium probability of moving from A to B equal that from B to A, for all Borel sets A, B in $\boldsymbol{\ell}$, the key assumption made by Green (1995) is that $\pi(\mathrm{d}\mathbf{C})q_r(\mathbf{C}, \mathrm{d}\mathbf{C}')$ has a finite density $f_r(\mathbf{C}, \mathbf{C}')$ with respect to a symmetric measure ϵ_r on $\boldsymbol{\ell} \times \boldsymbol{\ell}$. In general, A, $B \subset \boldsymbol{\ell}$, set

$$\epsilon(A \times B) = \{(A \cap \ell_j) \times (B \cap \ell_k)\} + \{(A \cap \ell_k) \times (B \cap \ell_j)\},$$

where $j, k \in \Im$. This is symmetric. Then, the appropriate acceptance probability for the proposed transition from $\mathbf{C}$ to $\mathbf{C}'$ by moving type r,

$$\alpha_r = \min\left\{1, \frac{f_r(\mathbf{C}', \mathbf{C})}{f_r(\mathbf{C}, \mathbf{C}')}\right\}. \tag{4.3.5}$$

Less formally,

$$\alpha_r = \min\left\{1, \frac{\pi(\mathrm{d}\mathbf{C}')q_r(\mathbf{C}', \mathrm{d}\mathbf{C})}{\pi(\mathrm{d}\mathbf{C})q_r(\mathbf{C}, \mathrm{d}\mathbf{C}')}\right\}. \tag{4.3.6}$$

The key to the RJMCMC algorithm is the introduction of auxiliary random variables at times t and $t + 1$ with dimensions chosen so that the augmented variables (that is, $\mathbf{C}$ and the auxiliary variables) at time t ans $t + 1$ have equal dimensions. we can then construct a Markov transition for the augmented variable at time t that maintains dimensionality. This dimension-matching stragegy enables the time-reversibility condition to be met by using a suitable acceptance probability, thereby ensuring that the Markov chain converges to the joint posterior for $\mathbf{C}$.

To understand dimension matching, it is simplest to begin by considering how one might propose parameters $\boldsymbol{\theta}_2$ corresponding to a proposed move (r) from a model $\mathcal{M}_1$ with n_1 parameters to a model $\mathcal{M}_2$ with n_2 parameters when $n_2 > n_1$. A simple approach is to generate $\boldsymbol{\theta}_2$ from an invertible deterministic function of both $\boldsymbol{\theta}_1$ and an independent random component $\mathbf{U}_1$. We can write $\boldsymbol{\theta}_2 = \delta_{1,2}(\boldsymbol{\theta}_1, \mathbf{U}_1)$. Proposing parameters for reverse move can be carried out via the inverse transformation, $(\boldsymbol{\theta}_1, \mathbf{U}_1) = \delta_{1,2}^{-1}(\boldsymbol{\theta}_2) = \delta_{2,1}(\boldsymbol{\theta}_2)$. Note that $\delta_{2,1}$ is an entirely deterministic function of a given $\boldsymbol{\theta}_2$.

If we generalize this idea to generate an augmented candidate parameter vector ($\boldsymbol{\theta}'_j$ and auxiliary variables $\mathbf{U}'$), given a proposed move to j' from the current model, $j^{(t)}$. We can apply an invertible deterministic function $\delta_{t,'}$ to $\boldsymbol{\theta}^{(t)}$ and some auxiliary random variables $\mathbf{U}$ to generate

$$(\boldsymbol{\theta}'_{j'}, \mathbf{U}') = \delta_{t,'}(\boldsymbol{\theta}^{(t)}, \mathbf{U}),$$

where $\mathbf{U}$ is generated from proposal density. The auxiliary variables $\mathbf{U}'$ and $\mathbf{U}$ are used so that $\delta_{t,'}$ maintains dimensionality during the Markov chain transition at time t.

In general, the reversible jump Markov chain Monte Carlo algorithm can be summarized as follows:

1. Initialize the iteration counter $t = 1$ and set an arbitrary initial value $(\mathbf{C}^{(0)}, j^{(0)})$;
2. Choose one move type (r) with probability p_r;
3. Given j', using the current state $\mathbf{C}^{(t)} = (j^{(t)}, \boldsymbol{\theta}^{(t)}_{j^{(t)}})$, generate an augmenting variable $\mathbf{U}|(j^{(t)}, \boldsymbol{\theta}^{(t)}_{j^{(t)}}, j')$ from an arbitrary proposal density q_r, three choices of proposal transition is summarized in the previous section. Let

$$(\boldsymbol{\theta}'_{j'}, \mathbf{U}') = \delta_{t,'}(\boldsymbol{\theta}^{(t)}_{j^{(t)}}, \mathbf{U}),$$

   where $\delta_{t,'}$ is an invertible mapping from $(\boldsymbol{\theta}^{(t)}_{j^{(t)}}, \mathbf{U})$ to $(\boldsymbol{\theta}'_{j'}, \mathbf{U}')$ and the auxiliary varibles have dimensions satisfying $n_{j^{(t)}} + n_{\mathbf{U}} = n_{j'} + n_{\mathbf{U}'}$;
4. Calculate the acceptance probability α_r given by

$$\alpha_r = \frac{f(j', \boldsymbol{\theta}'_{j'}|\mathbf{y})}{f(j^{(t)}, \boldsymbol{\theta}_{j^{(t)}}|\mathbf{Y})} \cdot \frac{g(j^{(t)}|j')}{g(j'|j^{(t)})} \cdot \frac{q_r(\mathbf{U}'|j', \boldsymbol{\theta}'_{j'}, j^{(t)})}{q_r(\mathbf{U}|j^{(t)}, \boldsymbol{\theta}^{(t)}_{j^{(t)}}, j')} \cdot \left|\mathbf{J}(t)\right|, \tag{4.3.7}$$

   where

$$\mathbf{J}(t) = \left.\frac{\mathrm{d}\delta_{t,'}(\boldsymbol{\theta}, \mathbf{u})}{\mathrm{d}(\boldsymbol{\theta}, \mathbf{u})}\right|_{(\boldsymbol{\theta}, \mathbf{u}) = (\boldsymbol{\theta}^{(t)}_{j^{(t)}}, \mathbf{U})}. \tag{4.3.8}$$

   If the move is accepted, $\mathbf{C}^{(t+1)} = (j', \boldsymbol{\theta}'_{j'})$; otherwise, $\mathbf{C}^{(t+1)} = \mathbf{C}^{(t)}$.
5. Set $t = t + 1$ and return to step 2 until the convergence is reached.

Note that the last term in Equation 4.3.7 is the obsolute value of the determinant of the Jacobian matrix arising fromt the change of variables from $(\boldsymbol{\theta}^{(t)}_{j^{(t)}}, \mathbf{U})$ to $(\boldsymbol{\theta}'_{j'}, \mathbf{U}')$. If $n_{j^{(t)}} = n_{j'}$, then $|\mathbf{J}(t)| = 1$ and it reduces to simple standard Metropolis-Hastings algorithm. Note that it is implicitly assumed that the transformation $\delta_{t,'}$ is differentiable.

Finally, it will be helpful to summarize the acceptance probability (Equation 4.3.7) as the following form:

$$\alpha_r = \min\left\{1, \text{likelihood ratio} \times \text{prior ratio} \times \text{proposal ratio} \times \text{Jacobian}\right\}. \tag{4.3.9}$$

4.4 Filtering of the intensities with the RJMCMC algorithm

In the previous section, a simulation method based on Markov chain called RJMCMC algorithm is introduced. The aim of this section is to present the filtering of several intensities based on the RJMCMC algorithm. Specially, a simple model — bivariate factor model — will be studied and the performance of the algorithm will also be discussed.

Before going through some details of application of RJMCMC algorithm, let us recap the model specification and the problem of interest. Up to this section, we have introduced a class of marked double stochastic Poisson processes, in which we consider the intensity process is a deterministic function of another marked point process, more detail is referred to Chapter 3.

4.4.1 A bivariate factor model

Consider the multivariate dynamic intensity model described in Chapter 3. In this specific model, each intensity is assumed to be composed of 2 components: one is a common component which allows to capture the comovements of the underlying point process, the other is an individual specific components, and both are driven by their own characteristics. This model combines the idea of latent factor models with the concept of dynamic intensity processes.

In this section, let us consider the simplest bivariate case, $K = 2$. Recall the intensity process:

$$\begin{cases} \tilde{\lambda}_t^{01} = \lambda_t^{(1)} + a_1\lambda_t^{(0)}, \\ \tilde{\lambda}_t^{02} = \lambda_t^{(2)} + a_2\lambda_t^{(0)}, \end{cases} \tag{4.4.1}$$

where $\lambda_t^{(1)}$ and $\lambda_t^{(2)}$ are two specific components, $\lambda_t^{(0)}$ is a common component, a_1 and a_2 (scalar parameters) are coefficients responsible for the impact of the common component $\lambda_t^{(0)}$ on the individual components $\lambda_t^{(1)}$ and $\lambda_t^{(2)}$ contemporaneously.

As described in Chapter 3, the intensity function completely characterizes the evolution of the point process , in particular, the intensity is viewed as a deterministic function of another MPP. Typically, for our factor model, each component $\lambda_t^{(i)}$ is a deterministic function of another marked point process, denoted by $\Phi'^{(i)} = \left(\tau_j^{(i)}, X_j^{(i)}\right)_{j\in\mathbf{N}}$, $i = 0, 1, 2$. As introduced in Section 3.2, each component $\lambda_t^{(i)}$ is assumed to be a shot noise process

(Cox & Isham, 1980, 1986). An explicit expression is given by

$$\lambda_t^{(i)} = \lambda_0^{(i)} e^{(-\kappa_i t)} + \sum_{j=0}^{N_t^{'(i)}} X_j^{(i)} \cdot e^{(-\kappa_i (t - \tau_j^{(i)}))}, \quad t \geq 0, \tag{4.4.2}$$

where:

$\lambda_0^{(i)}$ the initial value of the process $\lambda_t^{(i)}$ is Gamma distributed with shape parameter ν_i/κ_i and rate parameter γ_i;

$X_j^{(i)}$ the jump size has independent and identical exponential distribution with mean $1/\gamma_i$;

$\tau_j^{(i)}$ the time at which jump happens, which is assumed as Poissson process with constant intensity ν_i.

In the above model formulation, all of the elements of Equation 4.4.1, $\lambda_t^{(i)}$, $i = 0, 1, 2$, are assumed as mutually independent shot noise process (Cox & Isham 1980, 1986). The independence of each component can be guaranteed by the independence of associated marked point process $\Phi^{'(i)}$, so $\boldsymbol{\tau}^{(i)}$ are independent from $\boldsymbol{\tau}^{(j)}$ and $\mathbf{X}^{(j)}$, $i \neq j, i, j = 0, 1, 2$, where the vector $\boldsymbol{\tau}^{(i)}$ refers to all of the arrival times of i^{th} asset and the vector $\mathbf{X}^{(j)}$ is the corresponding jump size.

As for the intensity $\tilde{\lambda}_t^{0k}$ itself, the linear combination of two components, is still following shot noise process (Cox & Isham 1980,1986). The jump times are actually the combination of two sequences of jump times, $(\tau_1^{(0)}, \tau_2^{(0)}, \ldots, \tau_{N_t^{'(0)}}^{(0)})$ and $(\tau_1^{(k)}, \tau_2^{(k)}, \ldots, \tau_{N_t^{'(k)}}^{(k)})$, by order. It is worth to mention that the length of jump times of the intensity is $N_t^{'(0)} + N_t^{'(k)}$. The associated jump sizes are a bit complicated, when the common component is involved, the jump size should multiplicate the coefficient a_k. If j^{th} item is driven by the common component, for instance, the corresponding jump size is $a_k \cdot X_l^{(0)}$. Note that j and l are not required to be the same, in general, they are different.

In this modeling framework, a number of models can be specified under different hypothesis on the frequency and magnitude of these news. For instance, the inter-arrival times $\tau_j^{(i)} - \tau_{j-1}^{(i)}$ and the jump sizes $X_j^{(i)}$ can be serially correlated to account for reactions to the arrival of a piece of news; Barndorff-Nielsen & Shephard (2001) proposed the intensity λ_t has the dependence with marks. As described in Chapter 3, we specify our model with the following assumption:

1. the initial value $\lambda_0^{(i)}$ of the intensity process $(\lambda_t^{(i)})$ have Gamma distribution with scale parameter ν_i/k_i and rate parameter γ_i (that is, $\mathrm{E}(\lambda_0^{(i)}) = \nu_i/\kappa_i\gamma_i$);
2. the number of news arrivals $N^{'(i)}$ are Poisson processes with constant intensity ν_i;
3. the jump size $X_j^{(i)}$ have independent and identical exponential distribution with mean $1/\gamma_i$;
4. the inter-arrival time of news $\tau_j^{(i)} - \tau_{j-1}^{(i)}$ have independent and identical exponential distribution with mean $1/\nu_i$, and also independent of $\{X_j^{(i)}\}_{j \in \mathbf{N}}$.

Figure 4.3 illustrates a simulated realization of the common factor model, including the realization of the intensities, the realization of their components, and the realization of transaction times. Setting parameters: $\nu_0 = 0.1$, $\kappa_0 = 0.1$, $\gamma_0 = 0.2$, $\nu_1 = 0.5$, $\kappa_1 = 0.5$, $\gamma_1 = 1$, $\nu_2 = 1$, $\kappa_2 = 0.4$, $\gamma_2 = 1$, $a_1 = 1$ and $a_2 = 1$. Given three intensity components, $\lambda_t^{(0)}$ (upper panel), $\lambda_t^{(1)}$ (middle left panel) and $\lambda_t^{(2)}$ (middle right panel), according to Equation 4.4.1), the intensities $\tilde{\lambda}_t^{01}$ and $\tilde{\lambda}_t^{02}$ can be obtained. Transaction times $\{t_i^{(1)}\}_{i\in\mathbf{N}}$ (bottom left panel) and $\{t_i^{(2)}\}_{i\in\mathbf{N}}$ (bottom right panel) are simulated based on the intensities $\tilde{\lambda}_t^{01}$ and $\tilde{\lambda}_t^{02}$. As we can see, there is a cluster in transaction times during time 60 to 80 for both the two assets, if we turn back to the individual intensities $\lambda_t^{(1)}$ (middle left panel) and $\lambda_t^{(2)}$ (middle right panel), getting quite low intensities in that duration (even lower than their neighbour), however, we find the common intensity $\lambda_t^{(0)}$ have extremely high value during time 60 to 80. Such cluster could therefore impute to the common intensity.

In this setting, it is interesting to get the filtering of intensities, not only the 'combined' intensity $\tilde{\lambda}_t^{ok}$ but also the components of the underlying intensity, $\lambda_t^{(k)}$ and $\lambda_t^{(0)}$. Under the Bayesian context, it is to approximate the conditional distribution of intensity $\lambda_t^{(i)}$ from time 0 to T, given the observed trajecory of MPPs $\Phi^{'(i)}$ from 0 to T, denoted as $\Phi_{[0,T]}^{(i)}$, $i = 0, 1, 2$. Note that we set T as fixed time instant. This will be progressed in the following section.

It will be helpful to summarize some notations before to go through the filtering procedure. Let t denote physical time and let $\left\{t_j^k\right\}_{j\in\{1,2,\ldots,N^k\}}$, $k = 1, 2$, denote 2 sequences of transaction times with respect to 2 different types of events. Abbreviate the random vector $(\lambda_0^{(i)}, \tau_1^{(i)}, \ldots, \tau_{N_t^{'(i)}}^{(i)}, X_1^{(i)}, \ldots, X_{N_t^{'(i)}}^{(i)})$ by $\boldsymbol{\beta}(N_t^{'(i)})$, $i = 0, 1, 2$. In fact, the intensity $\lambda_t^{(i)}$ is completely determined by $\boldsymbol{\beta}(N_t^{'(i)})$ and 'combined' intensity $\tilde{\lambda}_t^{0k}$ is essentially the linear combination function of $(N_t^{'(0)})$ and $\boldsymbol{\beta}(N_t^{'(k)})$. The dimension of $\boldsymbol{\beta}(N_t^{'(i)})$ is then $2N_t^{'(i)}+1$. Note that the term 'combined' intensity is used in this context in order to distinguish the intensities from their components.

We observe the transaction times, $\left\{t_j^k\right\}_{j\in\{1,2,\ldots,N^k\}}$, but the underlying intensity is unknown, so it is difficult to determine the number of jumps for the underlying intensity. High frequency of the transaction can be driven by large jump size or by another new jump, on the other hand, low frequency of the transaction might be caused by new jump but with small jump size or essentially no jumps. $N_t^{'(i)}$ between 0 to T is therefore not fixed. However, the standard Markov chain Monte Carlo methods are restricted to problems where the joint distribution of all variables has a density with respect to some fixed standard measure. They have therefore not been available for application to Bayesian model, where the dimensionality of the parameter vector is not fixed. Nevertheless, Green (1995) proposed a framework for construction of reversible Markov chain samplers that jump between parameter subspaces with different dimensions which is introduced in the previous section. It is now extensively applied in model determination problems.

Concerning on the problem we have arised, it is convenient to apply Green's method to construct a reversible Markov chain such that the conditional distribution $\Pr(n^{(i)}, \boldsymbol{\beta}^{(i)}|\Phi_{[0,T]}^{(i)})$ as the target distribution, where $n^{(i)} = N_t^{'(i)}$, $\boldsymbol{\beta}^{(i)} = \boldsymbol{\beta}^{(i)}(N_t^{'(i)})$, the space of the pairs $(N_t^{'(i)}, \boldsymbol{\beta}(N_t^{'(i)}))$ is given by $C = U_{n=1}^{\infty} C_n$, where $C_n = \{n\} \times \mathbb{R}_1^{2n+1}$. In the following section, the implementation of the simplest bivariate factor model will be in progress.

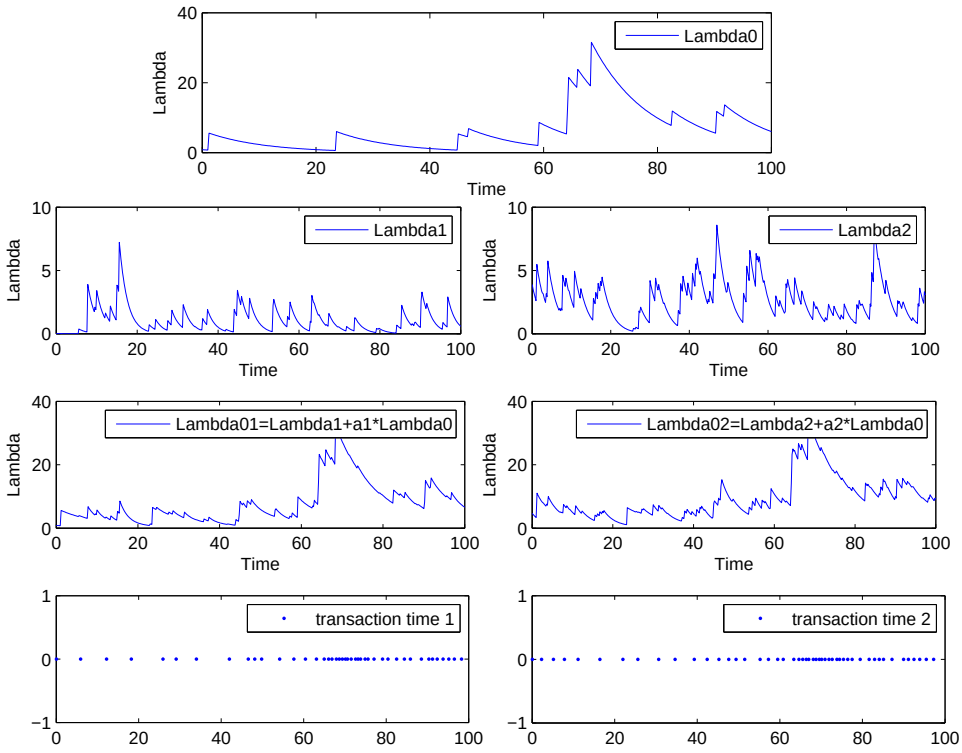

Figure 4.3: Simulated common factor intensity model and the resulting transaction times with parameters $\nu_0 = 0.1$, $\kappa_0 = 0.1$, $\gamma_0 = 0.2$, $\nu_1 = 0.5$, $\kappa_1 = 0.5$, $\gamma_1 = 1$, $\nu_2 = 1$, $\kappa_2 = 0.4$, $\gamma_2 = 1$, $a_1 = 1$ and $a_2 = 1$. The common component λ_0 (top panel). Individual specific component 1 λ_1 (upper middle left panel). Individual specific component 2 λ_2 (upper middle right panel). Intensity $\tilde{\lambda}^{01}$ (lower middle left panel). Intensity $\tilde{\lambda}^{02}$ (lower middle right panel). Transaction times 1 (left bottom panel). Transaction times 2 (right bottom panel). (Equation 4.4.1).

4.4.2 Implementation of the RJMCMC algorithm

In this section, we apply a set of transition moves proposed by Centanni & Minozzo (2006) on our bivariate intensity model described in the previous Chapter (Section 3.5.2 in Chapter 3). To assess the performance of this RJMCMC algorithm, some simulation experiments are required.

As for the definition of the transition moves, given the structure of the basic intensity model (Equation 3.2.3), there are totally three random variables involving with RJMCMC algorithm, $\lambda_0^{(i)}, X_j^{(i)}, \tau_j^{(i)}, N_t^{'(i)}$, $j = 1, 2, \ldots, N_t^{'(i)}$, $i = 0, 1, 2$. As suggested by Centanni & Minozzo (2006), the following five types of move can adequately explore the sampling space:

s) change the starting value $\lambda_0^{(i)}$;

h) change the height of a randomly chosen jump $X_j^{(i)}$;

p) change the position in time of a randomly chosen jump $\tau_j^{(i)}$;

n) change the number of jumps $N_t^{'(i)}$, this will relate to the dimension changing problem. Specifically, we propose two types of move concerning on the change of dimension:

b) generate a new jump at a randomly chosen time in $(0, t]$, this can be referred to 'birth' such that $N_t^{'(i)} = N_t^{'(i)} + 1$;

d) delete a randomly chosen jump, this can be referred to 'death' such that $N_t^{'(i)} = N_t^{'(i)} - 1$;

It is worth noting that all of these five types of move is suggested by aims to explore the sampling space and also to ease calculation. For the reversibility condition, these five types of move are satisfied. For example, type (s), generate a new starting value $\lambda_0^{*(i)}$ from Gamma distribution, and switch back, just generate another starting value from the same distribution, imagining as the previous one $\lambda_0^{(i)}$. The same for type (h) and type(p) that are involving with the same type, but a bit different for type (b) and type (d). Specially, the reverse transition for type (b) (type (d)) is type (d) (type (b)), removing (adding) one randomly chosen point $(\tau_j^{(i)}, X_j^{(i)})$. Since this pair 'birth-death' type involve with the dimension changing problem, the standard MCMC algorithm can not be applied. It is therefore turned to RJMCMC algorithm proposed by Green (1995) which is directed against the differing dimensionality issue, more details refer to Section 4.4.

Recall from our common factor model that three components of intensity are involved in the RJMCMC algorithms and each component is suggested to move around the sample space by engaging in the above five types of move. The simplest way to distribute the probability for each type of move and for each component is given by Figure 4.4. Especially, each component has the same importance (with probability $1/3$), but we do give more weight to type of moves, such as (s), (h), (p), with probability of 0.3, while less for type (b) and (d), with probablility of 0.05. This is just to avoid samples jumping between subspaces with differing dimensions too frequently.

In the following, we will go through more detail for implemetation of the RJMCMC algorithms. Specially, to identify each elements of acceptance probablity α (Equation 4.3.9)

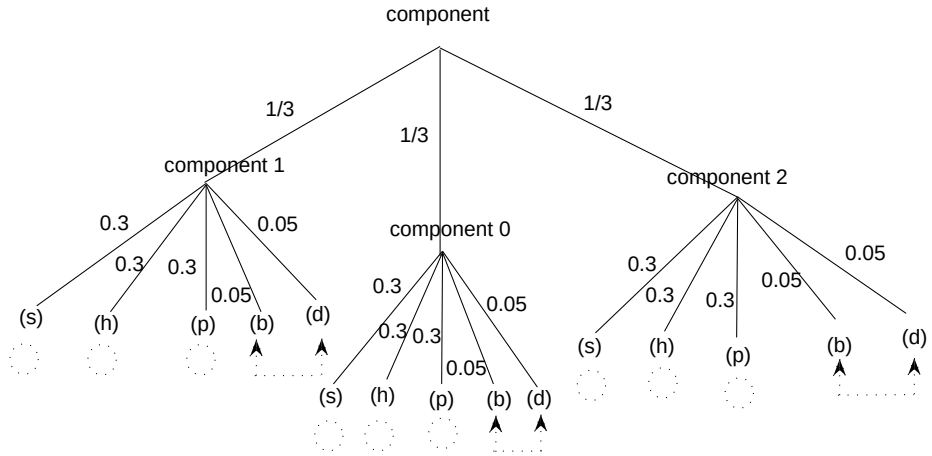

Figure 4.4: This graph illustrates the probability for each type of move. Three components (component (1), component (0) and component (2)) have the same probability distribution for each type of move: 0.3 for type (s), type (h) and type (p); 0.05 for type (b) and type (d). Note that (s), (h), (p), (b) and (d) stand for starting value, height, position, birth and death respectively.

for each type of move. For each updating of RJMCMC algorithm, we choose one of these five types of move with probability $p(r_i|n_i)$, where r_i indicates the type of move, (s_i), (h_i), (p_i), (b_i), (d_i), where the index i denotes the component (i), $i = 0, 1, 2$, such that $\sum_i \sum_{r_i} p(r_i|n_i) = 1$. Naturally, if the number of jumps n_i is equal to 0, the only type of move available for a proposal are the change of the starting value, type (s_i), and the 'birth' of a jump, type(b_i), that is, $p(r_i|n_i = 0) = 0$, $r_i = (h_i), (p_i), (d_i)$.

In practice, calculating the acceptance probability $\alpha(x, x')$ is a problem of crucial importance to the MCMC algorithms. Of course, it is important to the RJMCMC as well. For the RJMCMC algorithm, the decomposition of α is given by Equation 4.3.9.

For the proposal distribution, one popular choice is the prior density. In case of jumping between subspaces with the same dimension, the acceptance probability α (Equation 4.3.9) is reduced to

$$\alpha = \min \Big\{ 1, \text{likelihood ratio} \Big\}.$$

For instance, type (s_i), (h_i), (p_i). However, if the transition involving different dimensions, the acceptance ratio remains as the Formular 4.3.9, such as type (b_i) and type (d_i).

Now let us go through Equation 4.3.9 item by item. First, the likelihood ratio, which can be fully characterized by likelihood function. Under bivariate common factor model,

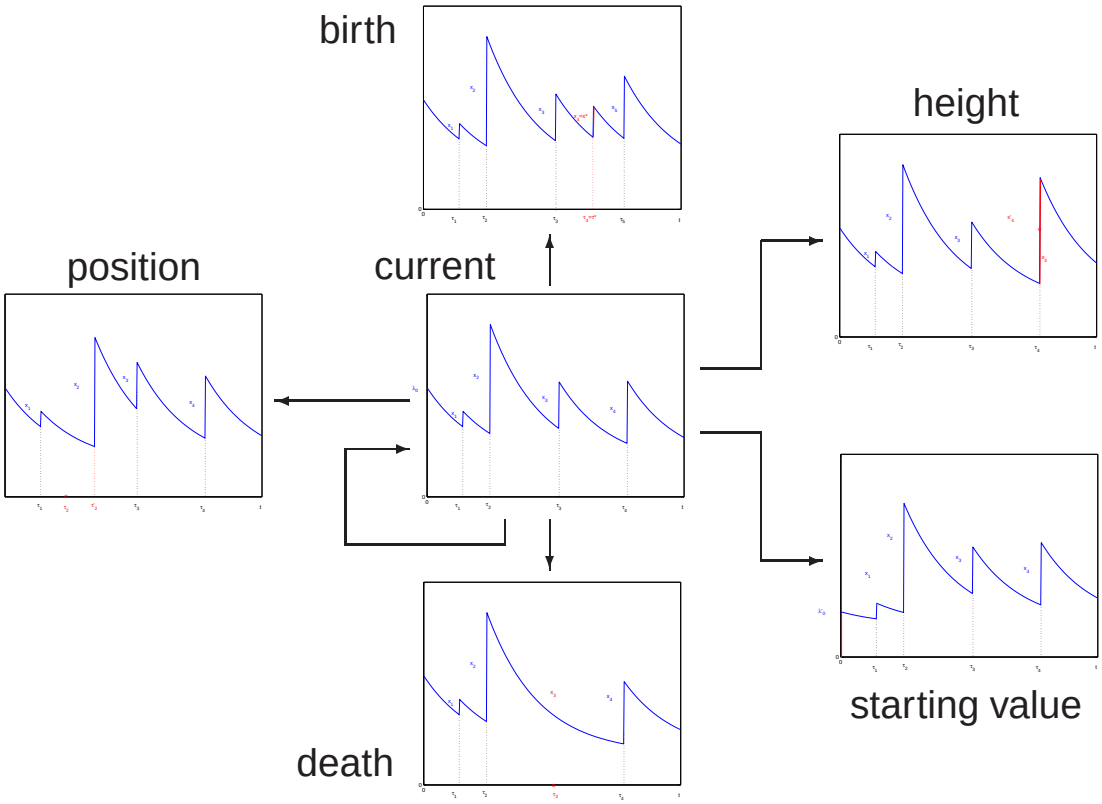

Figure 4.5: This figure illustrates five type of moves proposed for the RJMCMC algorithm. The red line indicates the new proposed value.

the likelihood function can be expressed as

$$
\begin{aligned}
&f\big((\mathbf{t}^{(1)},\mathbf{t}^{(2)})|\mathbf{N}'_{t,0}\,,\boldsymbol{\tau}^{(0)},\boldsymbol{\tau}^{(1)},\boldsymbol{\tau}^{(2)},\mathbf{X}^{(0)},\mathbf{X}^{(1)},\mathbf{X}^{(2)}\big)\\
&=\prod_{k=1}^{2} f(\mathbf{t}^{(k)}|N_t^{'(0)},N_t^{'(k)},\lambda_0^{(0)},\lambda_0^{(k)},\boldsymbol{\tau}^{(0)},\boldsymbol{\tau}^{(k)},\mathbf{X}^{(0)},\mathbf{X}^{(k)})\\
&=\prod_{k=1}^{2}\Bigg(\big[\prod_{l=1}^{N_t^{(k)}} f(t_l^{(k)}|t_{l-1}^{(k)})\big]\cdot \mathrm{P}_r\big(t^{(k)}_{N_t^{(k)}+1}>t|t^{(k)}_{N_t^{(k)}}\big)\Bigg)\\
&=\prod_{k=1}^{2}\Bigg(\big[\prod_{l=1}^{N_t^{(k)}} \tilde{\lambda}^{0k}_{t_l^{(k)}}\big]\cdot\exp\big\{-\int_0^t \tilde{\lambda}^{0k}_{t_s^{(k)}}\mathrm{d}s\big\}\Bigg),
\end{aligned}
$$

where $\tilde{\lambda}_t^{0k}$ denotes kth intensity at time t (see Equation 4.4.1).

For the prior function, it is given by

$$
\begin{aligned}
&\Pr(\mathbf{N}'_t,\boldsymbol{\Lambda}_0,(\boldsymbol{\tau}^{(0)},\boldsymbol{\tau}^{(1)},\boldsymbol{\tau}^{(2)}),(\mathbf{X}^{(0)},\mathbf{X}^{(1)},\mathbf{X}^{(2)}))\\
&=\prod_{i=0}^{2}\big(\Pr(N_t^{'(i)},\lambda_0^{(i)},\boldsymbol{\tau}^{(i)},\mathbf{X}^{(i)})\big)\\
&=\prod_{i=0}^{2}\Bigg(\Pr(\mathbf{X}^{(k)}|\boldsymbol{\tau}^{(k)},N_t^{'(k)})\cdot\Pr(\boldsymbol{\tau}^{(k)}|N_t^{'(k)})\cdot\Pr(N_t^{'(k)})\cdot f(\lambda_0^{(k)})\Bigg).
\end{aligned}
$$

And the posterior distribution function can be expressed as

$$
\begin{aligned}
&\Pr(\mathbf{N}'_t,\boldsymbol{\Lambda}_0,\boldsymbol{\tau}^{(0)},\boldsymbol{\tau}^{(1)},\boldsymbol{\tau}^{(2)},\mathbf{X}^{(0)},\mathbf{X}^{(1)},\mathbf{X}^{(2)}|\mathbf{t}^{(1)},\mathbf{R}^{(1)},\mathbf{t}^{(2)},\mathbf{R}^{(2)}) &(4.4.3)\\
&=\Pr(\mathbf{N}'_t,\boldsymbol{\Lambda}_0,\boldsymbol{\tau}^{(0)},\boldsymbol{\tau}^{(1)},\boldsymbol{\tau}^{(2)},\mathbf{X}^{(0)},\mathbf{X}^{(1)},\mathbf{X}^{(2)}|\mathbf{t}^{(1)},\mathbf{t}^{(2)}), &(4.4.4)
\end{aligned}
$$

where $\mathbf{N}'_t=[N_t^{'(0)},N_t^{'(1)},N_t^{'(2)}]$, $\boldsymbol{\Lambda}_0=[\lambda_0^{(0)},\lambda_0^{(1)},\lambda_0^{(2)}]$, $\boldsymbol{\tau}^{(0)}=[\tau_1^{(0)},\tau_2^{(0)},\ldots,\tau^{(0)}_{N_t^{'(0)}}]$, $\boldsymbol{\tau}^{(k)}=[\tau_1^{(k)},\tau_2^{(k)},\ldots,\tau^{(k)}_{N_t^{'(k)}}]$, $\mathbf{X}^{(k)}=[X_1^{(k)},X_2^{(k)},\ldots,X^{(k)}_{N_t^{'(k)}}]$, $\mathbf{t}^{(k)}=[t_1^{(k)},t_2^{(k)},\ldots,t_{(k)N_t^{(k)}}]$, $\mathbf{R}^{(k)}=\big[R_1^{(k)},R_2^{(k)},\ldots,R_{N_t^{(k)}}]$, where $N_t^{(k)}$ refers the transaction time of kth asset, $k=1,2$.

Equation 4.4.3 implies that the only observations from the data are transaction times $\mathbf{t}_j^{(k)}$ and logreturns $\mathbf{R}_j^{(k)}$, $j=1,2,\ldots,N_t^k$, $k=1,2$. The pass from Equation 4.4.3 to Equation 4.4.4, we assume logreturns are independent from intensity process $\tilde{\lambda}_t^{0k}$ which relys on $\mathbf{N}'_t$, $\boldsymbol{\Lambda}_0$, $\boldsymbol{\tau}^{(0)}$, $\boldsymbol{\tau}^{(k)}$, $\mathbf{X}^{(0)}$, $\mathbf{X}^{(k)}$, given the transaction time $\mathbf{t}^{(k)}$.

For the last term of Equation 4.3.9 (Jacobian), it indeed concerns on dimension matching. Of course, if a proposed sample come from the same sample space as the previous state of the chain, dimensionality is matched by nature, then the determinant of Jacobian is simply one. In our case, moves like (s), (h) and (p) meet such situation, and the RJMCMC reduced to simply MCMC method in which the Jacobian is one. However, for moves (b) and (d), they are involving dimensionality changing problem and the Jacobian may different from one. Let us consider a birth move (b), as proposed above for our model. It is also helpful to see Figure 4.5 together, which illustrates a graphic example of each type of move. Naturally, the inverse move of birth move is death move (d). When birth a new

jump, the dimension varys from $2N_t^{(i)}+1$ to $2N_t^{(i)}+3$, the difference can be accounted by two continuous variables, the new position $\tau^{*(i)}$ and its new jump size $X^{*(i)}$. After relabel the position for the new proposal, $(\tau_{j+1}^{'(i)}, \tau_{j+2}^{'(i)}, \ldots, \tau_{N_t^{'(i)}}^{'(i)}, \tau_{N_t^{'(i)}}^{'(i)}) = (\tau^{*(i)}, \tau_j^{(i)}, \ldots, \tau_{N_t^{(i)}}^{(i)})$ and $(X_{j+1}^{'(i)}, X_{j+2}^{'(i)}, \ldots, X_{N_t^{'(i)}}^{'(i)}) = (X^{*(i)}, X_{j+1}^{(i)}, \ldots, X_{N_t^{(i)}}^{(i)})$, where $N_t^{'(i)} = N_t^{(i)}+1$, $j = 0, 1, \ldots, N_t^{(i)}$. The initial value $\lambda_0^{(i)}$ keep the same as the existing one. Therefore, the first derivative of the new proposed vector respect to the old ones becomes

$$\frac{\partial f_{b,n\to n'(x,\mathbf{u})}}{\partial(x,\mathbf{u})} = \frac{\partial(\boldsymbol{\tau}_{(j+1)^-}^{'(i)}, \mathbf{X}_{(j+1)^-}^{(i)}, \lambda_0^{'(i)}, \tau_{j+1}^{'(i)}, X_{j+1}^{'(i)})}{\partial(\boldsymbol{\tau}^{(i)}, \mathbf{X}^{(i)}, \lambda_0^{(i)}, \tau^{*(i)}, X^{*(i)})} = \text{diag}(1, \ldots, 1)$$

, where

$$\boldsymbol{\tau}_{(j+1)^-}^{'(i)} = [\tau_1^{'(i)}, \tau_2^{'(i)}, \ldots, \tau_j^{'(i)}, \tau_{j+2}^{'(i)}, \ldots, \tau_{N_t^{'(i)}}^{'(i)}]$$

and

$$\mathbf{X}_{(j+1)^-}^{(i)} = [X_1^{'(i)}, X_2^{'(i)}, \ldots, X_j^{'(i)}, X_{j+2}^{'(i)}, \ldots, X_{N_t^{'(i)}}^{'(i)}].$$

Thus the determinant of Jacobian is one, because the determinant of a diagonal matrix with elements one is simply one. Without loss of generality, the determinant of Jacobian for its inverse move (d) is also one.

To proceed the RJMCMC algorithm, given all the knowledge provided above, we will get through component by component and type by type in more details. In our specified model, there are essentially three components involved. So, firstly, we choose one component, with probability $1/3$, then proceed the type of move with some probability according to the probability tree (see Figure 4.4). If a move of type$(h^{(i)})$ is chosen, a change to a height is attempted by choosing one of $X_1^{(i)}, X_2^{(i)}, \ldots, X_{N_t^{'(i)}}^{(i)}$ randomly, say $X_j^{(i)}$, then proposing a new height $X_j^{*(i)}$ from an exponential distribution with mean $\frac{1}{\gamma^{(i)}}$. For the acceptance probability of this type of move α_r, we should consider two situations differently, one is the common component $\lambda_t^{(0)}$, the other is the individual specific components $\lambda_t^{(k)}$. Because the common component affects the specific components contemporaneously, while the specific one just do its own. For the individual specific component $\lambda_t^{(i)}$, $i = 1, 2$, the acceptance probability is given by

$$\alpha_{ir} = \min\left\{1, \frac{\prod_{m=1}^{N_t^{(k)}} \check{\tilde{\lambda}}_{t_m}^{0k}}{\prod_{m=1}^{N_t^{(k)}} \tilde{\lambda}_{t_m}^{0k}} \exp\Big\{-\frac{1}{\kappa_i}(X_j^{*(k)} - X_j^{(k)})(1 - e^{-\kappa_i(t-\tau_j^{(i)})})\Big\}\times \right.$$
$$\left. \times \frac{e^{-\gamma_i X_j^{*(i)}}}{e^{-\gamma_i X_j^{(i)}}} \times \frac{e^{-\gamma_i X_j^{(i)}}}{e^{-\gamma_i X_j^{*(i)}}} \times 1\right\}.$$

Instead, for the common component $\lambda_t^{(0)}$, the acceptance probability is given by

$$\alpha_{0r} = \min\left\{1, \prod_{l=1}^{2}\left[\frac{\prod_{m=1}^{N_t^{(0)}} \check{\tilde{\lambda}}_{t_m}^{0l}}{\prod_{m=1}^{N_t^{(0)}} \tilde{\lambda}_{t_m}^{0l}} \exp\Big\{-\frac{a_l}{\kappa_0}(X_j^{*(0)} - X_j^{(0)})(1 - e^{-\kappa_0(t-\tau_j^{(0)})})\Big\}\right]\times\right.$$

$$\times \frac{e^{-\gamma_0 X_j^{*(0)}}}{e^{-\gamma_0 X_j^{(0)}}} \times \frac{e^{-\gamma_0 X_j^{(0)}}}{e^{-\gamma_0 X_j^{*(0)}}} \times 1 \Bigg\}.$$

For a move of changing position, type $(p^{(i)})$, one of the $\tau_1^{(i)}, \tau_2^{(i)}, \ldots, \tau_{N_t^{(i)}}^{(i)}$ is drawn at random, obtaining $\tau_j^{(i)}$. The proposed replacement value is $\tau_j^{*(i)}$ which is uniformly distributed on the interval $[\tau_{j-1}^{(i)}, \tau_{j+1}^{(i)}]$. The same here, we should consider two different situations, for the individual specific component $\lambda_t^{(i)}$, $i = 1, 2$, the acceptance probability is found to be

$$\alpha_{ir} = \min \Bigg\{ 1, \frac{\prod_{m=1}^{N_t^{(i)}} \tilde{\tilde{\lambda}}_{t_m}^{0i}}{\prod_{m=1}^{N_t^{(i)}} \tilde{\lambda}_{t_m}^{0i}} \exp\Big\{ \frac{X_j^{(i)}}{\kappa_i} (e^{-\kappa_i (t - \tau_j^{*(i)})} - e^{-\kappa_i (t - \tau_j^{(i)})}) \Big\} \times$$
$$\times 1 \times \frac{\frac{1}{T}}{\frac{1}{T}} \times 1 \Bigg\}.$$

And for the common component $\lambda_t^{(0)}$, the acceptance probability turns out as

$$\alpha_{0r} = \min \Bigg\{ 1, \prod_{l=1}^{2} \Bigg[\frac{\prod_{m=1}^{N_t^{(l)}} \tilde{\tilde{\lambda}}_{t_m}^{0l}}{\prod_{m=1}^{N_t^{(l)}} \tilde{\lambda}_{t_m}^{0l}} \exp\Big\{ \frac{X_j^{(0)}}{\kappa_0} (e^{-\kappa_0 (t - \tau_j^{*(0)})} - e^{-\kappa_0 (t - \tau_j^{(0)})}) \Big\} \Bigg] \times$$
$$\times 1 \times \frac{\frac{1}{T}}{\frac{1}{T}} \times 1 \Bigg\}.$$

Similarly, for changing the initial value, type $(s^{(i)})$, we generate $\lambda_0^{*(i)}$ from Gamma distribution with mean $\nu^{(i)}/\kappa_i \gamma^{(i)}$ and variance $\nu^{(i)}/\kappa_i (\gamma^{(i)})^2$, the acceptance probability for the individual specific component $\lambda_t^{(i)}$, $i = 1, 2$, is given by

$$\alpha_{ir} = \min \Bigg\{ 1, \frac{\prod_{m=1}^{N_t^{(i)}} \tilde{\tilde{\lambda}}_{t_m}^{0i}}{\prod_{m=1}^{N_t^{(i)}} \tilde{\lambda}_{t_m}^{0i}} \exp\Big\{ -\frac{1}{\kappa_i} (\lambda_j^{*(i)} - \lambda_j^{(i)})(1 - e^{-\kappa_i t}) \Big\} \times$$
$$\times \left(\frac{\lambda_j^{*(i)}}{\lambda_j^{(i)}} \right)^{\nu_i/\kappa_i - 1} e^{-\gamma_i (\lambda_j^{*(i)} - \lambda_j^{(i)})} \times \left(\frac{\lambda_j^{(i)}}{\lambda_j^{*(i)}} \right)^{\nu_i/\kappa_i - 1} e^{-\gamma_i (\lambda_j^{(i)} - \lambda_j^{*(i)})} \times 1 \Bigg\}.$$

While for the common component $\lambda_t^{(0)}$, the acceptance probability is given by

$$\alpha_{0r} = \min \Bigg\{ 1, \prod_{l=1}^{2} \Bigg[\frac{\prod_{m=1}^{N_t^{(l)}} \tilde{\tilde{\lambda}}_{t_m}^{0l}}{\prod_{m=1}^{N_t^{(l)}} \tilde{\lambda}_{t_m}^{0l}} \exp\Big\{ -\frac{a_l}{\kappa_0} (\lambda_j^{*(0)} - \lambda_j^{(0)})(1 - e^{-\kappa_l t}) \Big\} \Bigg] \times$$
$$\times \left(\frac{\lambda_j^{*(0)}}{\lambda_j^{(0)}} \right)^{\nu_0/\kappa_0 - 1} e^{-\gamma_0 (\lambda_j^{*(0)} - \lambda_j^{(0)})} \times \left(\frac{\lambda_j^{(0)}}{\lambda_j^{*(0)}} \right)^{\nu_0/\kappa_0 - 1} e^{-\gamma_0 (\lambda_j^{(0)} - \lambda_j^{*(0)})} \times 1 \Bigg\}.$$

In case of changing the number of jumps, it leads to much more complicated situation. Let's first see the birth type $(b^{(i)})$, which attempts to add a new intensity jump for component $\lambda_t^{(i)}$, $i = 1, 2$. We need to generate a new position $\tau^{*(i)}$ uniformly distributed on the

interval (0,T] and a new jump size $X^{*(i)}$ from an exponential distribution with mean $1/\gamma^{(i)}$. This new position $\tau^{*(i)}$ must belong to an existing interval, say $(\tau_j^{(i)}, \tau_{j+1}^{(i)})$. If accepted, the number of intensity jumps is updated as $N_t^{(i)}+1$, and $\tau_{j+1}^{'(i)}$ will be set to $\tau^{*(i)}$, the following sequence after j, $\tau_{j+1}^{(i)}, \tau_{j+2}^{(i)}, \ldots, \tau_{N_t^{(i)}}^{(i)}$ will be relabeld as $\tau_{j+2}^{'(i)}, \tau_{j+3}^{'(i)}, \ldots, \tau_{N_t^{(i)}+1}^{'(i)}$.

The acceptance probability for this proposal has to be calculated to achieve detailed balance with its inverse move, death move $(d^{(i)})$. Dimension matching is achieved by reversing birth move, so that the jump position $\tau^{*(i)}$ and its associated jump size $X^{*(i)}$ are removed. Of course, the number of intensity jumps turns out to be $N_t^{(i)}$. Hence, the acceptance probability α_r for type $(b^{(i)})$, again we should consider two different cases, if individual specific component $\lambda_t^{(i)}$, $i=1,2$, is selected, α_{ir} is found to be

$$\alpha_{ir} = \min\left\{1, \frac{\prod_{m=1}^{N_t^{(i)}} \tilde{\lambda}_{t_m}^{0i}}{\prod_{m=1}^{N_t^{(i)}} \lambda_{t_m}^{0i}} \exp\Big\{-\frac{X^{*(i)}}{\kappa_i}(1-e^{-\kappa_i(t-\tau^{*(i)})})\Big\}\times \right.$$
$$\left. \times e^{-\gamma_i X^{*(i)}}\gamma_i\nu_i \times \frac{p(d^{(i)}|N_t^{'(i)}+1)}{p(b^{(i)}|N_t^{'(i)})}\frac{\frac{1}{N_t^{'(i)}+1}}{\frac{1}{T}\gamma_i e^{-\gamma_i X^{*(i)}}} \times 1\right\}.$$

Instead, for the common component $\lambda_t^{(0)}$, the acceptance probability becomes

$$\alpha_{0r} = \min\left\{1, \prod_{l=1}^{2}\left[\frac{\prod_{m=1}^{N_t^{(l)}} \tilde{\tilde{\lambda}}_{t_m}^{0l}}{\prod_{m=1}^{N_t^{(l)}} \tilde{\lambda}_{t_m}^{0l}} \exp\Big\{-\frac{a_l X^{*(l)}}{\kappa_0}(1-e^{-\kappa_0(t-\tau^{*(0)})})\Big\}\right]\times \right.$$
$$\left. \times e^{-\gamma_0 X^{*(0)}}\gamma_0\nu_0 \times \frac{p(d^{(0)}|N_t^{'(0)}+1)}{p(b^{(0)}|N_t^{'(0)})}\frac{\frac{1}{N_t^{'(0)}+1}}{\frac{1}{T}\gamma_0 e^{-\gamma_0 X^{*(0)}}} \times 1\right\}.$$

Similarly, the death move $(d^{(i)})$, according to the detailed balance condition, choose one $\tau_j^{(i)}$ at random, remove it and also its associated value $X_j^{(i)}$, then number of intensity jumps turns to $N_t^{(i)}-1$, if the death move is accepted. The acceptance probability for individual specific component $\lambda_t^{(i)}$, $i=1,2$, is give by

$$\alpha_{ir} = \min\left\{1, \frac{\prod_{m=1}^{N_t^{(i)}} \tilde{\tilde{\lambda}}_{t_m}^{0i}}{\prod_{m=1}^{N_t^{(i)}} \tilde{\lambda}_{t_m}^{0i}} \exp\Big\{\frac{X^{*(i)}}{\kappa_i}(1-e^{-\kappa_i(t-\tau^{*(i)})})\Big\} \times \frac{e^{\gamma_i X^{*(i)}}}{\gamma_i\nu_i}\times \right.$$
$$\left. \times \frac{p(b^{(i)}|N_t^{'(i)}-1)}{p(d^{(i)}|N_t^{'(i)})}\frac{\frac{1}{T}\gamma_i e^{-\gamma_i X^{*(i)}}}{\frac{1}{N_t^{'(i)}+1}} \times 1\right\}.$$

And for the common component $\lambda_t^{(0)}$, the acceptance probability is given by

$$\alpha_{0r} = \min\left\{1, \prod_{l=1}^{2}\left[\frac{\prod_{m=1}^{N_t^{(l)}} \tilde{\tilde{\lambda}}_{t_m}^{0l}}{\prod_{m=1}^{N_t^{(l)}} \tilde{\lambda}_{t_m}^{0l}} \exp\Big\{\frac{a_l X^{*(0)}}{\kappa_0}(1-e^{-\kappa_0(t-\tau^{*(0)})})\Big\}\right] \times \frac{e^{\gamma_0 X^{*(0)}}}{\gamma_0\nu_0}\times \right.$$
$$\left. \times \frac{p(b^{(0)}|N_t^{'(0)}-1)}{p(d^{(0)}|N_t^{'(0)})}\frac{\frac{1}{T}\gamma_0 e^{-\gamma_0 X^{*(0)}}}{\frac{1}{N_t^{'(0)}+1}} \times 1\right\}.$$

4.4.3 Simulation results and discussions

The goal of a RJMCMC analysis is to estimate features of the target distribution π. The reliablility of such estimates depends on the extent to which sample averages computed using realizations of the chain correspond to their expectation under the limiting stationary distribution of the chain. All of the RJMCMC (even the simple MCMC) methods described above have the correct limiting stationary distribution. In practice, however, it is necessary to determine when the chain has run sufficiently long so that it is reasonable to believe that the output adequately represents the target distribution and can be used reliably for estimation. Unfortunately, RJMCMC (and simple MCMC) methods can sometimes be quite slow to converge, requiring extremely long runs, especially for high dimension. Or sometimes the chain can not fully explores the support of the target distribution.

The performance of the filtering based on our proposed RJMCMC algorithm is evaluated for different simulation studies. In the following, we will show some simulation results of filtering for the bivariate common factor intensity model. In this model, three components (intensities) are the target of the filtering (prediction), one is the common intensity $\lambda_t^{(0)}$, the other are two individual specific intensities, $\lambda_t^{(1)}$ and $\lambda_t^{(2)}$, representing a multi-demensional Markov chain. How well this proposed RJMCMC algorithm works, several question can be made. First, has the chain run long encough? Second, has the chain traversed all portions of the region of support of the target distribution? Third, are the sampled values approximate draws from the target distribution? The first question concerns on convergence of the chain and the second and third questions can be evaluated by comparison of the filtering and of the true intensities, not only the intensity as a whole but also the intensity components.

Key considerations in the diagnosis of convergence are the *burn-in* period and run length. For any implementation, the iterations will not have exactly the correct marginal distribution, and the dependence on the starting value from which the chain was started may remain strong. To reduce the severity of this problem, the first n values from the chain are typically discarded as a *burn-in period*.

The determination of an appropriate burn-in period and run length is an active area of research. Suppose that the variable of interest is X, and its value at the t^{th} iteration of the chain is $x^{(t)}$. Thus, the n values $x^{(0)}, \ldots, x^{(n-1)}$ of the chain are discarded as burn-in and the rest values $x^{(n)}, \ldots, x^{(N)}$ are retained, where N is the length of chain. The most commonly used estimator is based on an empirical average. Discard the burn-in, then compute

$$\bar{x} = \frac{1}{N-n+1} \sum_{t=n}^{N} x^{(t)}, \tag{4.4.5}$$

which can be viewed as estimator of $\mathrm{E}(X)$.

Many times we may ask what values should be used for the number of iterations for burn-in and the length of the chain after burn-in. But most authors are reluctant to recommend generic values because appropriate choices are highly dependent on the problem at hand and the rate and efficient with which the chain explores the region supported by the target distribution.

The implementation of RJMCMC on our common factor model, as described above, we first need to define r type of move. Figure 4.4 shows the probability distribution for

each type of move. Here we make no claim of optimaliy for the particular choice of which component, but we do claim to set less weight to birth and death type of move for each component avoiding jumping subspaces with differing dimensionality frequently. The dotted line indicates the reversed move, in particular, the dotted circle means the same move, such as (s), (h) and (p), while the dotted arrow between (b) and (d) means the reversed move for (b) is (d), and vice versa.

Figure 4.6 and Figure 4.7 show the simulation results under the specified bivariate intensity model, in which we set time horizon $T = 100$ and parameters are fixed as $\nu_0 = 1.5$, $k_0 = 0.4$, $\gamma_0 = 0.25$; $\nu_1 = 2$, $k_1 = 0,2$, $\gamma_1 = 0.5$; $\nu_2 = 1.8$, $k_2 = 0.6$, $\gamma_2 = 0.1$; $a_1 = 1.6$, $a_2 = 1.8$. We ran the RJMCMC algorithm with total 100,000 iterations (including 50,000 burn-in period) on the simulated trajectory of transaction times from 0 to T. To obtain the filtering expectation of λ_t^{01} and λ_t^{02}, we take arithmetic means of $\tilde{\lambda}_{[0,T]}^{01(i)}(t)$ and $\tilde{\lambda}_{[0,T]}^{02(i)}(t)$, where $\tilde{\lambda}_{[0,T]}^{01(i)}(t)$ denotes the subset of samples in i^{th} iteration, $i > 50,000$, $t = 0, 0.2, 0.4, \ldots, T$. That is, $(1/50,000)\sum_{i=50,001}^{100,000} \tilde{\lambda}_{[0,T]}^{01(i)}(t)$.

Figure 4.6 illustrates the filtering (prediction) and the true intensity process from time 0 to $T = 100$ based on the combination of one common component $(\lambda^{(0)})$ and of one individual component $(\lambda^{(1)})/(\lambda^{(2)})$. The blue solid line denotes the true (simulated) intensity and the red dash line denotes the filtering result based on the RJMCMC algorithm which we proposed in previous section. If compare the filtered intensity (blue line) with the true one (red dash line), it produces satisfactorily mixing chains where the proposed RJMCMC algorithm succeeds to provide samples to explore the *combined* parameter space. In the sense, the posterior distribution, $\Pr(N_t^{'(0)} + N_t^{'(1)}, N_t^{'(0)} + N_t^{'(2)}, \lambda_0^{(1)} + a_1\lambda_0^{(0)}, \lambda_0^{(2)} + a_2\lambda_0^{(0)}, \tilde{\boldsymbol{\tau}}^{01}, \tilde{\boldsymbol{\tau}}^{02}, \tilde{\mathbf{X}}^{01}, \tilde{\mathbf{X}}^{02} | \mathbf{T}^{(1)}, \mathbf{T}^{(2)})$, where $\tilde{\boldsymbol{\tau}}^{01} = \{\tau_j^k\}_{j\in\{1,2,\ldots,N'^{(k)}\}}$, $k = 0, 1$, denote two sequences of arrival times orderly with respect to common news (intensity) (0) and specific news (intensity) (1), and $\tilde{\mathbf{X}}^{01} = \{aX_j^k\}_{j\in\{1,2,\ldots,N_t^{'(k)}\}}$, $a = a_1$ for $k = 0$, and $a = 0$ for $k = 1$, the order for $\tilde{\mathbf{X}}^{01}$ is the same as $\tilde{\boldsymbol{\tau}}^{01}$, regards as target distribution from which the samples were drawn via RJMCMC algorithm. The performance of filtering combined intensity, $\tilde{\lambda}_t^{01}$ and $\tilde{\lambda}_t^{02}$, demonstrates that it fully reflects the information of transaction times $\{T_j\}_{j\in\{1,2,\ldots,N\}}$.

However, for filtering of intensity components, $\lambda_t^{(0)}$, $\lambda_t^{(1)}$, $\lambda_t^{(2)}$, see Figure 4.7, the solid line denotes the true (simulated) intensity and the line denotes the filtering result based on the RJMCMC algorithm which we proposed in previous section. The upper figure shows the common component $\lambda_t^{(0)}$, the left bottom figure shows the individual component $\lambda_t^{(1)}$ and the right bottom one shows the individual component $\lambda_t^{(2)}$. The resulting filtering figures show the chain does poor job of exploring the region of posterior support. This chain has clearly not converged to their stationary distribution. Since drifts are still visable. Particularly, overprediction for the common component $\lambda_t^{(0)}$, while underprediction for the individual components $\lambda_t^{(1)}$ and $\lambda_t^{(2)}$. This simulation result presents a plausible distinction problem, as first discussed in the simple example of Metropolis-Hastings algorithm (see Section 4.2.1). To distinct the components of intensity, $\lambda_t^{(0)}$, $\lambda_t^{(1)}$, $\lambda_t^{(2)}$, when we just observe the transaction times which actually reflected in the intensity as a whole $\tilde{\lambda}_t^{01}$ and $\tilde{\lambda}_t^{02}$, seems a tricky problem that the mean value can be obtained by a combination of over and under values (in rough).

If we get deep insight into the routine of filtering, we would find this distinction prob-

lem could be relieved by modifying one assumption we made in the previous section. Recall the structure of our model, in which the same cluster in transaction times should be due to the common intensity, see Figure 4.3. Furthermore, three sequencies jump times $\{\tau_j^k\}_{j\in\{0,1,\ldots,N'^{(k)}\}}$, $k = 0, 1, 2$ are distinctive except the initial value that they all start at time 0, in other words, the news arrival times of different assets never coincide at the same time with probability 1 (this is theoretical tenable and practical realizable) except the starting value, this helps to understand that the (efficient) RJMCMC algorithm should extract such information and make the filtering works, apart the initial value.

Therefore, the real distinction problem then reduced to the initial value and the simplest way to tackle this problem is to set the initial value as fixed. Accordingly, we remove the assumption that the initial value λ_0 are Gamma distributed, treating as a scale parameter instead. So we would expect that overprediction and underprediction of the components accounted for distinction problem should be removed.

We ran RJMCMC algorithm again with the same data, but set the initial values as fixed. The result is presented in Figure 4.8, by comparison, it looks the filtering result is improved with respect to Figure 4.6 and Figure 4.7 (note that the initial value was set as fixed in Figure 4.8). Especially, in Figure 4.6, as we can see, the filtering intensity $\tilde{\lambda}^{01}$(dotted line) shift to the upside of the true intensity (solid line), whereas the second time simulation (set the initial as fixed) does better job as shown in Figure 4.8 (bottom left). Besides, filtering of intensity $\tilde{\lambda}^{02}$ is improved as well, it is perfectly fit the ture intensity in Figure 4.8 (bottom right) although it almost works well in Figure 4.6. On the other hand, for filtering three components, the second simulation is improved in the sense there is no such clear overprediction for one component while underprediction for the other as shown in Figure 4.7. It turns out rather random combination of overprediction and underprediction between components. For instance, time between 0 and 10, overprediction for common component and underprediction for two individual components, while opposit situation can be find during 75 to 80.

The question then turns to the original problem that the proposed RJMCMC algorithm, in particular, the proposed type of moves, could not efficiently explore the support of the target distribution. Although this RJMCMC algorithm outperforms for filtering the intensity as a whole $\tilde{\lambda}_t^{01}$ and $\tilde{\lambda}_t^{02}$ (see Figure 4.6 and Figure 4.8 (bottom two panels)), it is quite poor for filtering the components $\lambda_t^{(0)}$, $\lambda_t^{(1)}$, and $\lambda_t^{(2)}$. This suggests the chain does visit the important ‘combined‘ region, but inefficiently visit the specific subspaces in which the samples are drawn.

On the other hand, the chain might become stuck in one or more modes of the target distribution. In this case, all convergence diagnostics may indicate that the chain has converged, though the chain does not fully represent the target distribution. However, there is substantial overlap between the goals of diagnosing convergence to the stationary distribution and investigating the mixing properties of the chain. No diagnostic is fail-safe. For these reasons, we combine the discussion of mixing and convergence in the following section.

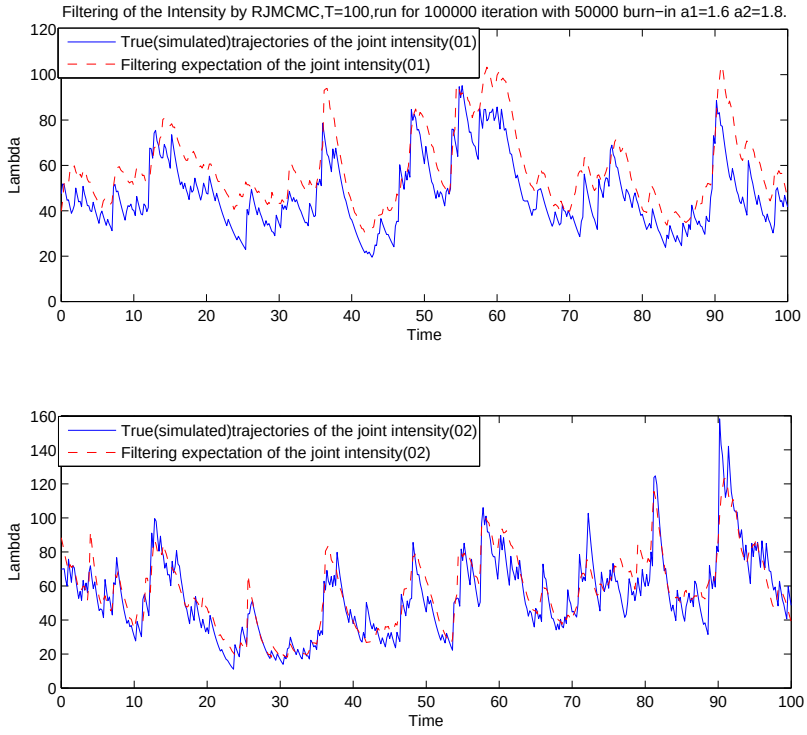

Figure 4.6: Bivariate intensty process. Combined intensity λ_t^{01} (upper figure). Combined intensity λ_t^{02} (bottom figure). The solid line denotes the true intensity and the dash line represents the filtering intensity. Results based on 100,000 updates with 50,000 burn-in period; $\nu_0 = 1.5$, $k_0 = 0.4$, $\gamma_0 = 0.25$; $\nu_1 = 2$, $k_1 = 0,2$, $\gamma_1 = 0.5$; $\nu_2 = 1.8$, $k_2 = 0.6$, $\gamma_2 = 0.1$; $a_1 = 1.6$, $a_2 = 1.8$ (Equation 4.4.1).

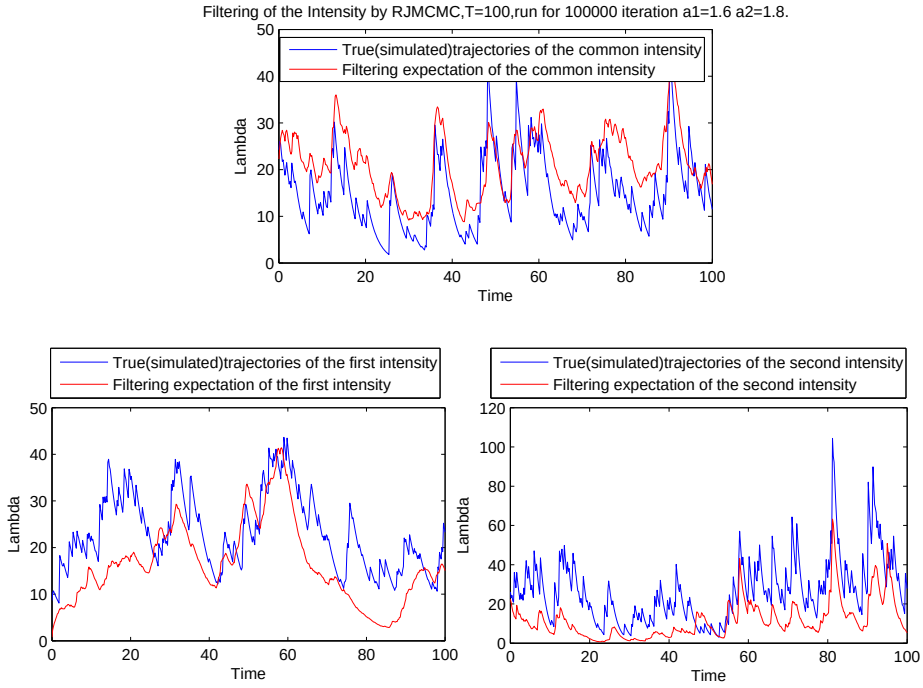

Figure 4.7: Bivariate intensity process. Common component of intensity $\lambda_t^{(0)}$ (upper figure). Individual component $\lambda_t^{(1)}$ (bottom left figure). Individual component $\lambda_t^{(2)}$ (bottom right figure). The solid line denotes the true intensity and the line represents the filtering intensity. Results based on 100,000 updates with 50,000 burn-in period; $\nu_0 = 1.5$, $k_0 = 0.4$, $\gamma_0 = 0.25$; $\nu_1 = 2$, $k_1 = 0,2$, $\gamma_1 = 0.5$; $\nu_2 = 1.8$, $k_2 = 0.6$, $\gamma_2 = 0.1$; $a_1 = 1.6$, $a_2 = 1.8$ (Equation 4.4.1).

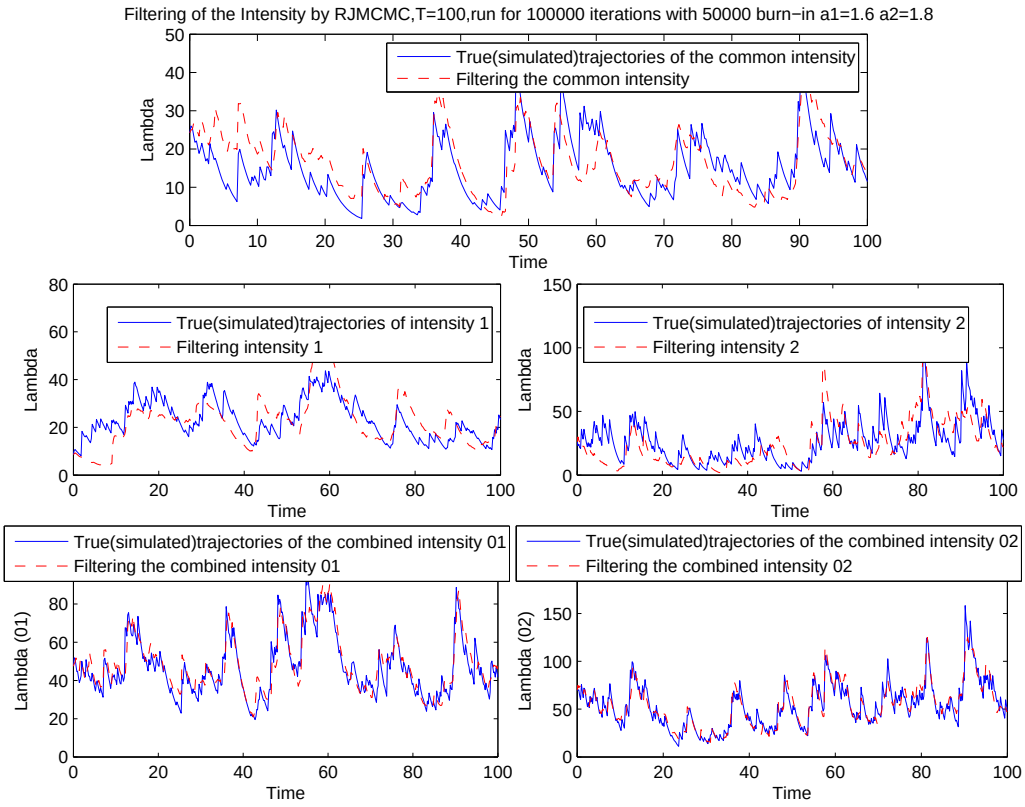

Figure 4.8: Bivariate intensity process. Common component of intensity $\lambda_t^{(0)}$ (upper figure). Individual component $\lambda_t^{(1)}$ (middle left figure). Individual component $\lambda_t^{(2)}$ (middle right figure). Combined intensity λ_t^{01} (bottom left figure). Combined intensity λ_t^{02} (bottom right figure). The solid line denotes the true intensity and the dash line represents the filtering intensity. The blue solid line denotes the true intensity and the red line represents the filtering intensity. Results based on 100,000 updates with 50,000 burn-in period; $\nu_0 = 1.5$, $k_0 = 0.4$, $\gamma_0 = 0.25$; $\nu_1 = 2$, $k_1 = 0,2$, $\gamma_1 = 0.5$; $\nu_2 = 1.8$, $k_2 = 0.6$, $\gamma_2 = 0.1$; $a_1 = 1.6$, $a_2 = 1.8$ (Equation 4.4.1).

4.5 A second RJMCMC filtering algorithm

The message from the above simulation results is that the proposed RJMCMC algorithm is not very efficiently to explore the sample space. This may due to the fact that the proposed type of moves or transition kernel provide samples, at least some of them, are far away from the target region.

For proposing a new type of move, we must concern on some combined moves as suggested by the above simulation results. However, Roberts and Papaspiliopoulos (2004) propose a move from $\Phi^{'}$ to $\Phi^{'} - \{(\tau_j, X_j)\} \cup \{(\tau^*, X^*)\}$ achieved by local change of the intensity process. Basically, they generate τ^* uniformly in $[\tau_{j-1}, \tau_{j+1}]$ with $\tau_0 = 0$ and $\tau_{n+1} = T$, and define an invertible transformation in which the new proposed height X^* is a function of its location τ^* (more detail see Roberts & Papaspiliopoulos (2004)). This move provides a local change and evaluation of only a small part of the likehood function, which may be a good start to reconsider our proposed type of move concerning on local displacement.

In case of our common factor model, we can consider a local change for three components of a mixture, adopting the idea of Roberts & Papaspiliopoulos (2004). Since the previous RJMCMC (in Section 4.4) can adequately explore the combined sample space in the sense that the RJMCMC algorithm does good job in filtering of intensity as a whole $\tilde{\lambda}_{0k}$ but poor in distinction of intensity components. Some local changes among them (components) serves as constraint to let the chain visit some other region.

Practically, we change the locations $\{\tau_j^{*(i)}\}_{j \in \mathrm{N}}$, $i = 0, 1, 2$, with a mixture, restricted by a neighbourhood condition $D = |\tau^{(1)} - \tau^{(2)}| < 2\beta$. Adapting to our model, we define a *bdd* move (abbreviation for birth-death-death) that combines two nearby individual components and a *dbb* move (abbreviation for death-birth-birth) that splits a common component into two nearby ones. The *bdd* move considers the neighbourhood condition $D = |\tau^{(1)} - \tau^{(2)}| < 2\beta$ of two individual components such that to select a single value $\tau^{*(0)}$ for common component, whose new location is

$$u^{*(0)} = \tau^{*(0)} \sim \mathrm{U}(\tau^*_{\max} - \beta, \tau^*_{\min} + \beta), \tag{4.5.1}$$

where β is a simulation parameter, $\tau^*_{\min} = \max\{\tau^{(1)}, \tau^{(2)}\}$, $\tau^*_{\min} = \min\{\tau^{(1)}, \tau^{(2)}\}$. Note that to ensure reversibility, we only perform the *bdd* move if $|\tau^{(1)} - \tau^{(2)}| < 2\beta$ is satisfied. Meanwhile, we also generate a new value $X^{*(0)}$ from exponential distribution with mean $1/\gamma_0$ for jump height.

The corresponding *dbb* move that gurantees reversibility, considers breaking a randomly choosen $\tau_j^{(0)}$ from common component into two new generated value $\tau^{*(1)}$ and $\tau^{*(2)}$ for two individual component respectively:

$$\begin{cases} \tau^{*(1)} = \tau_j^{(0)} + u^{*(1)}\beta, \\ \tau^{*(2)} = \tau_j^{(0)} + u^{*(2)}\beta, \end{cases} \tag{4.5.2}$$

where $u^{*(1)}, u^{*(2)} \sim \mathrm{U}(-1, 1)$. Meanwhile, generate two new values $X^{*(1)}$ and $X^{*(2)}$ from exponential distribution with mean $1/\gamma_1$ and $1/\gamma_2$ respectively. It worth noting, during the proceeding of simulation, that $\tau^{*(1)}$ and $\tau^{*(2)}$ could be out of the interval $(0, \mathrm{T}]$ because u^* has uniformly distribution, it ranges from -1 to 1.

A graphical example is presented in Figure 4.9. This figure shows merge and split move. S_1 refers to $\tau^{*(1)}$, S_2 refers to $\tau^{*(2)}$ and m refers to $\tau^{*(0)}$. A merge move that combines two nearby components and a split move that breaks a component into two nearby ones.

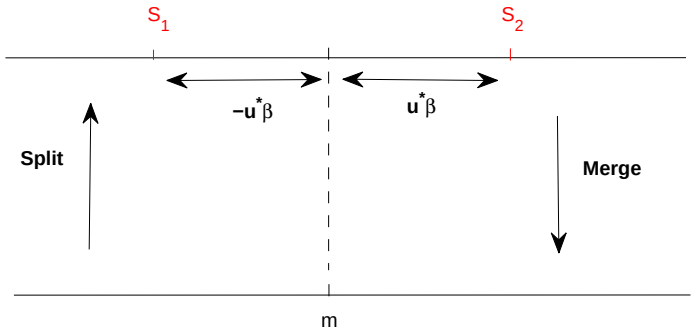

Figure 4.9: A graphic example for merge and split move. A merge move that combines two nearby components and a split move that breaks a component into two nearby ones.

For dimension matching, in this case, consider a bijection between $(u^{*(1)}, X^{*(1)}, u^{*(2)}, X^{*(2)}, \tau^{(0)}, X^{(0)})$ and $(\tau^{(1)}, X^{(1)}, \tau^{(2)}, X^{(2)}, u^{*(0)}, X^{*(0)})$ with dimension of six. In deriving an expression for the acceptance probability of type *dbb*, it is helpful to write down item by item given the Formular 4.3.9. The likelihood ratio is given by

$$\prod_{l=1}^{2}\left[\frac{\prod_{m=1}^{N_t^{(l)}}\tilde{\lambda}_{T_m}^{0l}}{\prod_{m=1}^{N_t^{(l)}}\lambda_{T_m}^{0l}}\exp\left\{\frac{X^{*(l)}}{k_l}(e^{-k_l(t-\tau^{*(l)})}-1)+\frac{a_l X^{(0)}}{k_0}(1-e^{-k_0(t-\tau^{(0)})})\right\}\right]. \quad (4.5.3)$$

The prior ratio is

$$\frac{e^{\gamma_0 X^{(0)}}}{\gamma_0\nu_0}\frac{\gamma_1\nu_1}{e^{\gamma_1 X^{(1)}}}\frac{\gamma_2\nu_2}{e^{\gamma_2 X^{(2)}}}.$$

The proposal ratio is

$$\frac{p(bdd)}{p(dbb)}\frac{(1/n_c)(\gamma_0 e^{-\gamma_0 X^{(0)}})(1/(\tau^*_{\max}-\tau^*_{\min}+2\beta))}{(1/N_t'^{(0)})(1/2)(\gamma_1 e^{-\gamma_1 X^{*(1)}})(1/2)(\gamma_2 e^{-\gamma_2 X^{(*2)}})},$$

where n_c denotes the number of couples that satisfy the neighbourhood condition. Notice that if nc $= 0$, that is, there do not exist two nearby points for its reverse move, *bdd*, then the proposal ratio becomes

$$\frac{p(bdd)}{p(dbb)}\frac{(\gamma_0 e^{-\gamma_0 X^{(0)}})(1/(\tau^*_{\max}-\tau^*_{\min}+2\beta))}{(1/N_t'^{(0)})(1/2)(\gamma_1 e^{-\gamma_1 X^{*(2)}})(1/2)(\gamma_2 e^{-\gamma_2 X^{(2)}})}.$$

And the Jacobian for move *dbb*, in the present context, it is

$$J_{dbb} = \left| \frac{\partial(u^{*(1)}, X^{*(1)}, u^{*(2)}, X^{*(2)}, \tau^{(0)}, X^{(0)})}{\partial(\tau^{(1)}, X^{(1)}, \tau^{(2)}, X^{(2)}, u^{*(0)}, X^{*(0)})} \right| = \beta^2.$$

Similarly, for the move *bdd*, the likelihood ratio is

$$\prod_{l=1}^{2} \left[\frac{\prod_{m=1}^{N_t^{(l)}} \tilde{\lambda}_{T_m}^{0l}}{\prod_{m=1}^{N_t^{(l)}} \lambda_{T_m}^{0l}} \exp\left\{ \frac{a_1 X^{*(0)}}{k_0}(e^{-k_0(t-\tau^{*(0)})} - 1) + \frac{X^{(1)}}{k_1}(1 - e^{-k_1(t-\tau^{(1)})}) \right\} \right].$$

The proposal ratio is

$$\frac{p(dbb)}{p(bdd)} \frac{1(/N_t^{'(0)})(1/2)(\gamma_1 e^{-\gamma_1 X^{(1)}})(1/2)(\gamma_2 e^{-\gamma_2 X^{(2)}})}{(1/n_c)(\gamma_0 e^{-\gamma_0 X^{*(0)}})(1/(\tau^*_{\max} - \tau^*_{\min} + 2\beta))},$$

where n_c denotes the number of couples that satisfy the neighbourhood condition. Notice that for move *bdd* if and only if when $n_c > 0$ is satisfied. This implies that the corresponding reverse move *dbb* satisfying the neighbourhood condition. Obviously, it is no sense to make death move for component 1 and 2 when they are far away. Generally, few points are survived in this case. Finally, the Jacobian for move *bdd* is just the inverse of J_{dbb}, then

$$J_{bdd} = \frac{1}{\beta^2}.$$

For the other moves $(h^{(i)})$, $(p^{(i)})$, $(b^{(i)})$ and $(d^{(i)})$, after posing the neighbourhood condition, the probability of each move $p(r_i)$ should be restricted somehow to ease the calculation. Again, we also fix the starting value in this algorithm, so type $(s^{(i)})$ is no need to be considered. The key point is worth to stress here is when one move involved in position change, it is possible that one move satisfies the neighbourhood condition, but after the jump (into another subspace), the neighbourhood condition might be unsatisfied. On the other hand, before the jump, one does not satisfy the neighbourhood condition, but after the proposed change, the condition may be satisfied. Hence, the acceptance ratio for type $(p^{(i)})$, $(b^{(i)})$ and $(d^{(i)})$ should be reconsidered, while the acceptance ratio for type $(h^{(i)})$ remains the same as in Section 4.4.

Before calculating the acceptance probability α_r, it is helpfull to go through the probability for each type of move and the associated reverse move after introducing the neighbourhood condition. The probability tree is illustrated in Figure 4.10. For each iteration of the RJMCMC algorithm, we should first check if the neighbourhood condition is satisfied or not. This is because the neighbourhood condition leads to two different results for acceptance ratio A_r as shown in the first two branches (enclosing 'Yes' and 'No') in the probability tree. Specially, for type $(p^{(k)})$, $(b^{(k)})$ and $(d^{(k)})$ which are involving with the position changing. Notice that the neighbourhood condition is the restriction for individual component (k), $k = 1, 2$, so the acceptance ratio for common component (0) remains the same as in Section 4.4.

In Figure 4.10, $|\tau_1(i) - \tau_2(j)| < 2\beta$ is the neighbourhood condition, where $i = 1, 2, \ldots, N^{'(1)}$, $j = 1, 2, \ldots, N^{'(2)}$, and β is a simulation system parameter. 'Yes' or 'No'

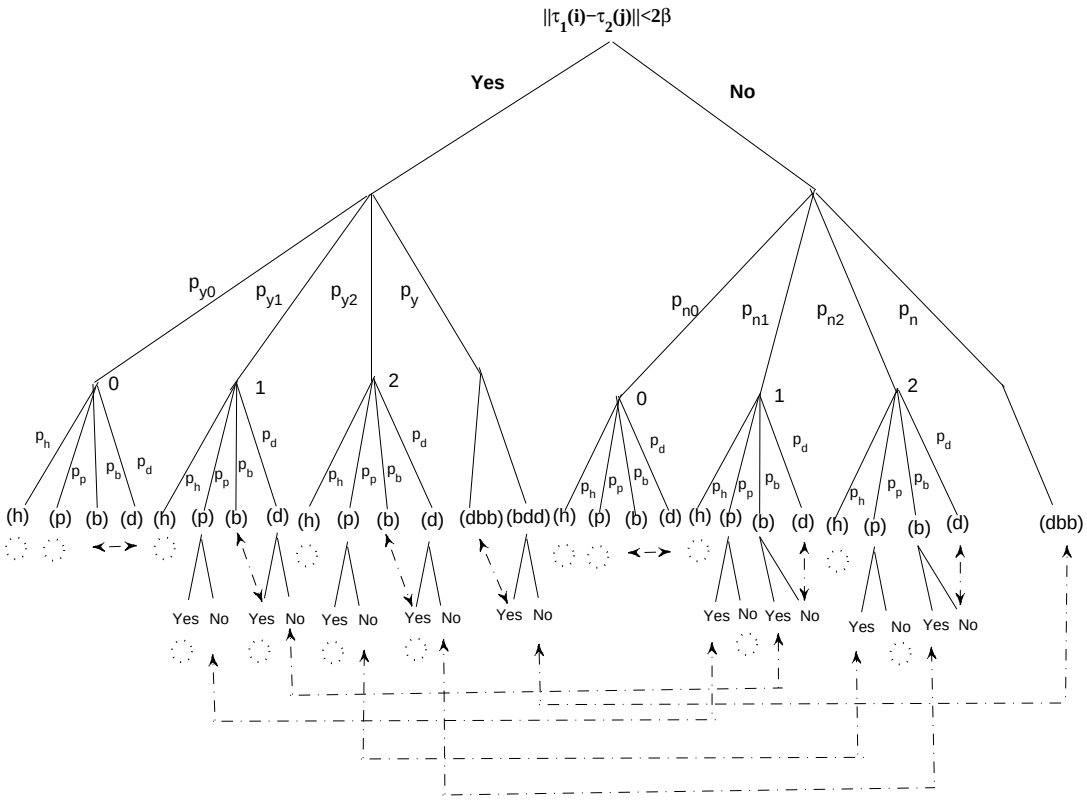

Figure 4.10: Probability tree with the neighbourhood condition. This graph illustrates the probability for each type of move and the associated reverse move (shown with arrow). The neighbourhood condition $\|\tau_1(i) - \tau_2(j)\| < 2\beta$, where $i = 1, \ldots, N'^{(1)}$, $j = 1, \ldots, N'^{(2)}$, β is a simulation system parameter. 'Yes' or 'No' here indicates the neighbourhood condition is satisfied or not. The dotted circle indicates the reverse move involved in the same type of move; The dashed arrow indicates the reverse move involved in different type of move. (h), (p), (b), (d) denote change of height, change of position, birth and death respectively. (dbb) denotes death move for the common component (0) and birth move for two individual components (1) and (2). (bdd) denotes birth move for the common component (0) and death move for two individual components (1) and (2).

here indicates the neighbourhood condition is satisfied or not. The dotted circle indicates the reverse move involved in the same type of move; The dashed arrow indicates the reverse move involved in different type of move. (h), (p), (b), (d) denote change height, position, birth and death respectively. (dbb) denotes death move for the common component (0) and birth move for two individual components (1) and (2). (bdd) denotes birth move for the common component (0) and death move for two individual components (1) and (2).

As we can see, Figure 4.10 is much more complicated than Figure 4.4, although the basic idea is the same that each move should garantee its reversed move by the nature of RJMCMC. The neighbourhood condition functions whenever the location is involved. There is no place for type of move (s) since the location is fixed, the rest of moves, (h), (p), (b), (d), (bdd) and (dbb), instead, should check diligently such condition. For instance, a move type (p) for component (1) satisfying neighbourhood condition (yes), if a new proposed position does not satisfy this neighbourhood condition, its reversed move then goes to the left branch of the probability tree where move type (p) for component (1) unsatisfying neighbourhood condition (no). Notice that no existing a type of move (bdd) in case the neighbourhood condition is not satisfied, because all the locations are far away (beyond the neighbourhood condition).

Now let us go through the acceptance ratio calculation type by type, for $(p^{(k)})$, $(b^{(k)})$ and $(d^{(k)})$. In case that the neighbourhood condition is satisfied, that is, there at least exists one pair $(\tau^{*(1)}, \tau^{*(2)})$ satisfy $D = |\tau^{*(2)} - \tau^{*(1)}| \leq 2\beta$. If move type $(p^{(k)})$ is chosen, choose a j from $\{1, 2, \ldots, N'^{(k)}\}$ at random and propose a new position $\tau^{*(k)}$ uniformly distributed from $(\tau_{j-1}^{(k)}, \tau_{j+1}^{(k)})$, if $\tau_{j+1}^{(k)} > T$ then set $\tau_{j+1}^{(k)} = T$. Now calculate the acceptance ratio A_r. According to Equation 4.3.9, we will consider each element separately, likelihood ratio, prior ratio, proposal ratio and Jacobian. Note that likelihood ratio, prior ratio and Jacobian will not be influenced by introducing the neighbourhood condition except the proposal ratio. In the following we will identify these ratios one by one. Now turn to the likelihood ratio for the move $(p^{(k)})$, it is given by

$$l_{p_k} = \frac{\prod_{m=1}^{N_t^{(k)}} \tilde{\tilde{\lambda}}_{T_m}^{0k}}{\prod_{m=1}^{N_t^{(k)}} \tilde{\lambda}_{T_m}^{0k}} \exp\left\{ \frac{X_j^{(k)}}{\kappa_k} (e^{-\kappa_k (t - \tau_j^{*(k)})} - e^{-\kappa_k (t - \tau_j^{(k)})}) \right\},$$

where l_{p_k} denotes the likelihood ratio for position move $(p^{(k)})$.

The prior ratio and Jacobian are equivalent to 1. For the proposal ratio, consider two results after the proposed move. One is that this new proposal may still hold the neighbourhood condition, if this is the case, the proposal ratio is given by

$$\frac{p_p \cdot p_{y_k}}{p_p \cdot p_{y_k}} \frac{1/N'^{(k)} \cdot 1/T}{1/N'^{(k)} \cdot 1/T},$$

otherwise, this new proposed positon destroy the neighborhood condition and move far away from the constraint, then the proposal ratio becomes

$$\frac{p_p \cdot p_{n_k}}{p_p \cdot p_{y_k}} \frac{1/N'^{(k)} \cdot 1/T}{1/N'^{(k)} \cdot 1/T}.$$

In case that the neighbourhood condition is not satisfied, so there is not exist one pair $(\tau^{*(1)}, \tau^{*(2)})$ satisfy $D = |\tau^{*(2)} - \tau^{*(1)}| \leq 2\beta$. For move type $(p^{(k)})$, a new position is

proposed, and calculate the acceptance ratio A_r. Considering two situations: one is that this new proposal move to the subspace that the condition is satisfied, in this case, the proposal ratio is given by

$$\frac{p_p \cdot p_{y_k}}{p_p \cdot p_{n_k}} \frac{1/N'^{(k)} \cdot 1/T}{1/N'^{(k)} \cdot 1/T};$$

the other is that the proposed value dose not improve the neighborhood relation and there are all of points in $\boldsymbol{\tau}^{(1)}$ and in $\boldsymbol{\tau}^{(2)}$ are still far from each other in terms of the specified distance 2β, So the proposal ratio becomes

$$\frac{p_p \cdot p_{n_k}}{p_p \cdot p_{n_k}} \frac{1/N'^{(k)} \cdot 1/T}{1/N'^{(k)} \cdot 1/T}.$$

Now consider the pair 'birth' and 'death' move. If the neighbourhood condition is satisfied and move type $(b^{(k)})$ is chosen, so give a birth for asset k in which we generate a new jump time $\tau^{*(k)}$ and also the assosciating jump size $X^{*(k)}$, $k = 1, 2$. The same as shown in Section 4.4, the new proposed $\tau^{*(k)}$ is drawing from uniform distribution $\mathbf{U}(0, T)$ and the corresponding jump size $X^{*(k)}$ is from exponential distribution with mean $1/\gamma_k$. This new position $\tau^{*(k)}$ must belong to an existing interval, say $(\tau_j^{(k)}, \tau_{j+1}^{(k)})$. If accepted, the number of intensity jumps is updated as $N_t^{(k)} + 1$, and $\tau_{j+1}^{'(k)}$ will be set to $\tau^{*(k)}$, the following sequence after j, $\tau_{j+1}^{(k)}, \tau_{j+2}^{(k)}, \ldots, \tau_{N_t^{(k)}}^{(k)}$ will be relabeld as $\tau_{j+2}^{'(k)}, \tau_{j+3}^{'(k)}, \ldots, \tau_{N_t^{(k)}+1}^{'(k)}$. For the calculation of acceptance ratio A_r, according to Equation 4.3.9, let us first consider the likelihood ratio

$$l_{b_k} = \frac{\prod_{m=1}^{N_t^{(k)}} \tilde{\tilde{\lambda}}_{T_m}^{0k}}{\prod_{m=1}^{N_t^{(k)}} \tilde{\lambda}_{T_m}^{0k}} \exp\left\{ -(X^{*(k)}/\kappa_k)(1 - e^{-\kappa_k(t - \tau^{*(k)})}) \right\},$$

where l_{b_k} denotes the likelihood ratio for birth move $(b^{(k)})$. The prior ratio is given by

$$\eta_{b_k} = e^{-\gamma_i X^{*(k)}} \gamma_k \nu_k,$$

where η_{b_k} denotes the prior ratio for birth move $(b^{(k)})$. And the determinant of Jacobian is 1, this has already been explained in Section 4.4 for birth move.

For the proposal ratio, we need to clear the resluting states of the points and the reverse move. Obviously, birth a new point under the satisfication of neighborhood condition, no matter this proposed value is accept or not, such kind of neighborhood condition is still held. Therefore, the only candidate for the reverse move is death move $(d^{(k)})$ which is satisfied with the neighborhood condition. After clearify this matter, the proposal ratio is given by

$$\frac{p_d \cdot p_{y_k}}{p_b \cdot p_{y_k}} \frac{1/(N'^{(k)} + 1)}{\gamma_k e^{-\gamma_k X^{*(k)}} \cdot 1/T}.$$

On the other hand, if the neighbourhood condition is not satisfied and move type $(b^{(k)})$ is chosen, the only difference from the previous case is the new jump may cause the neighbourhood condition to be held. Here we should consider two situations which will alter

the acceptance ratio. Firstly, if such new jump $(\tau^{*(k)}, X^{*(k)})$ do not improve the neighbourhood relation, then for its reverse move $(d^{(k)})$, which must be inconsistent with the neighbourhood condition. Then the proposal ratio becomes

$$\frac{p_d \cdot p_{n_k}}{p_b \cdot p_{n_k}} \frac{1/(N'^{(k)}+1)}{\gamma_k e^{-\gamma_k X^{*(k)}} \cdot 1/T}.$$

Secondly, if such new jump $(\tau^{*(k)}, X^{*(k)})$ result in the neighbourhood condition satisfied, the reverse move $(d^{(k)})$ must be reconsidered. Speically, we should consider the one that the neighbourhood condition is held before the death move, but such death move destroy the only one pair that satisfy the neighbourhood condition. The proposal ratio is

$$\frac{p_d \cdot p_{y_k}}{p_b \cdot p_{n_k}} \frac{1/(N'^{(k)}+1)}{\gamma_k e^{-\gamma_k X^{*(k)}} \cdot 1/T}.$$

Similarly, if the neighbourhood condition is satified and death move $(d^{(k)})$ is chosen, the same technique as described in Section 4.4, generate a random number from subset $\{1, 2, \ldots, N_t^{'(k)}\}$, where $N_t^{'(k)}$ is the number of jumps for asset k on the time interval $[0, t]$, say j, then delete the jump time $\tau_j^{(k)}$ and corresponding jump size $X_j^{(k)}$. Again, for calculating the acceptance ratio, according to Equation 4.3.9, we first see the likelihood ratio which is given by

$$l_{d_k} = \frac{\prod_{m=1}^{N_t^{(k)}} \tilde{\tilde{\lambda}}_{T_m}^{0i}}{\prod_{m=1}^{N_t^{(i)}} \tilde{\lambda}_{T_m}^{0k}} \exp\left\{\frac{X^{*(k)}}{\kappa_k}(1 - e^{-\kappa_k(t-\tau^{*(k)})})\right\},$$

where l_{d_k} denotes the likelihood ratio for death move $(d^{(k)})$. The prior ratio is

$$\eta_{d_k} = \frac{e^{\gamma_k X^{*(k)}}}{\gamma_k \nu_k},$$

where η_{d_k} denotes the prior ratio for death move $(d^{(k)})$. And the Jacobian is equal to 1.

For the proposal ratio, we need to consider the reverse move. Since death move might destroy the neighbourhood condition, two different situations will be included. If this death move make the neighbourhood condition unsatisfied, then for its reverse move, 'birh' move, should give a birh such that the neighourhood condition is satisfied. It is worth to mention the situation in which the neighbourhood condition is not satisfied before this birth move. The proposal ratio is then

$$\frac{p_b \cdot p_{n_k}}{p_d \cdot p_{y_k}} \frac{\gamma_k e^{-\gamma_k X_j^{(k)}} \cdot 1/T}{1/(N'^{(k)}+1)},$$

On the other hand, if this death move dose not destroy the neighbourhood condition, that is, the chosen index j or precisely $\tau_j^{(k)}$, which is not one member or is not the only one member (in the sense that there are many) that satisfy the neighbourhood condition. The reverse move then give a birth under the situation that the neighbourhood condition is held (before the move). The proposal ratio becomes

$$\frac{p_b \cdot p_{y_k}}{p_d \cdot p_{y_k}} \frac{\gamma_k e^{-\gamma_k X_j^{(k)}} \cdot 1/T}{1/(N'^{(k)}+1)}.$$

In the case that the neighbourhood condition is unsatisfied and death move $(d^{(k)})$ is selected, the same as before, one candidate pairs $(\tau_j^{(k)}, X_j^{(k)})$ is going to delete. It is obvious that the death move is not able to improve the neighbourhood condition, so its reverse move has to give a birth under the case that the neighbourhood condition is not held and such birth does not make the neighbourhood condition to be satisfied. The proprosal ratio is therefore

$$\frac{p_b \cdot p_{n_k}}{p_d \cdot p_{n_k}} \frac{\gamma_k e^{-\gamma_k X_j^{(k)}} \cdot 1/T}{1/(N'^{(k)}+1)}.$$

4.5.1 Simulation results and discussions

Run on the second RJMCMC algorithm with total 100,000 iterations (including 50,000 burn-in period) on the simulated trajectory of transaction times from 0 to T, Figure 4.11 and Figure 4.12 show the simulation results. To obtain the filtering expectation of λ_t^{01} and λ_t^{02}, we take arithmetic means of $\tilde{\lambda}_{[0,T]}^{01(i)}(t)$ and of $\tilde{\lambda}_{[0,T]}^{02(i)}(t)$, $(1/50,000)\sum_{i=50,001}^{100,000} \tilde{\lambda}_{[0,T]}^{01(i)}(t)$ where $\tilde{\lambda}_{[0,T]}^{01(i)}(t)$ denotes the subset of samples in i^{th} iteration, $i > 50,000$, $t = 0, 0.2, 0.4, \ldots, T$. It is worth noting that, as showed above, the probability for each type of move plays a special role in the RJMCMC algorithm. Especially, in this second RJMCMC algorithm, we would like to give more weight to the joint moves (dbb and bdd) (Note that the probability for moves are quite flexible and subjective.). The reason for biased joint moves is quite clear that we hope the neighbourhood condition would constrain the jump between different sampling space. It is helpful to go through Figure 4.10. In our case, we set $p_{y0} = 1/6$, $p_{y1} = 1/6$, $p_{y2} = 1/6$, $p_y = 1/2$, $p_{n0} = 5/12$, $p_{n1} = 1/6$, $p_{n2} = 1/6$, $p_n = 1/4$, $p_h = 0.3$, $p_p = 0.3$, $p_b = 0.2$, and $p_d = 0.2$.

Figure 4.11 shows the filtering (prediction) and the true intensity process from time 0 to $T = 100$ based on the combination of one common component $(\lambda^{(0)})$ and of one individual component $(\lambda^{(1)})/(\lambda^{(2)})$. The blue solid line denotes the true (simulated) intensity and the red dash line denotes the filtering value based on the second RJMCMC algorithm. By comparison, the filtering intensity (blue line) are close to the true one (red dash line). In other words, it produces quite satisfactorily mixing chains where the second RJMCMC algorithm somehow succeeds to provide samples to explore the *combined* parameter space.

In Figure 4.12, it gives the filtering of components, $\lambda_t^{(0)}$, $\lambda_t^{(1)}$, and $\lambda_t^{(2)}$, where the solid line denotes the true (simulated) intensity and the line denotes the filtering result based on the second RJMCMC algorithm. The upper figure shows the common component $\lambda_t^{(0)}$, the left bottom figure shows the individual component $\lambda_t^{(1)}$ and the right bottom one shows the individual component $\lambda_t^{(2)}$. By comparing the filtering and the true intensity, we find that the resulting filtering is improved by the second RJMCMC algorithm, though the drift still exist, the common component $\lambda_t^{(0)}$ (the upper panel) is almost overpredicted while two individual components are underprected, which is the same problem as in Section 4.4 but with much more mild. Generally speaking, this chain proposed by the second RJMCMC

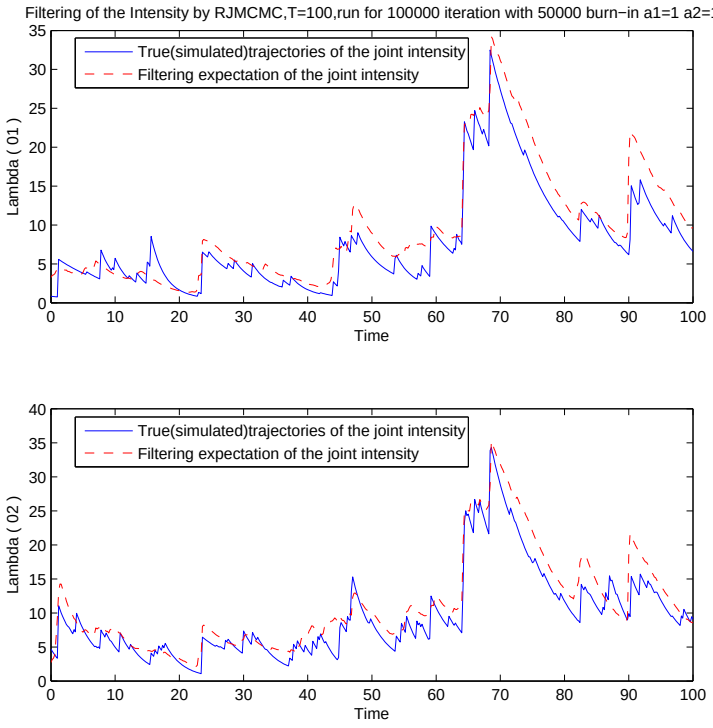

Figure 4.11: Bivariate intensty process. Combined intensity λ_t^{01} (upper figure). Combined intensity λ_t^{02} (bottom figure). The solid line denotes the true intensity and the dash line represents the filtering intensity. Results based on 100,000 updates with 50,000 burn-in period; $\nu_0 = 0.1$, $k_0 = 0.1$, $\gamma_0 = 0.2$, $\nu_1 = 0.5$, $k_1 = 0.5$, $\gamma_1 = 1$, $\nu_2 = 1$, $k_2 = 0.4$, $\gamma_2 = 1$, $a_1 = 1$, $a_2 = 1$, $\beta = 1$ (Equation 4.4.1).

algorithm does better job of exploring the region of target distribution than the previous RJMCMC (see Section 4.4).

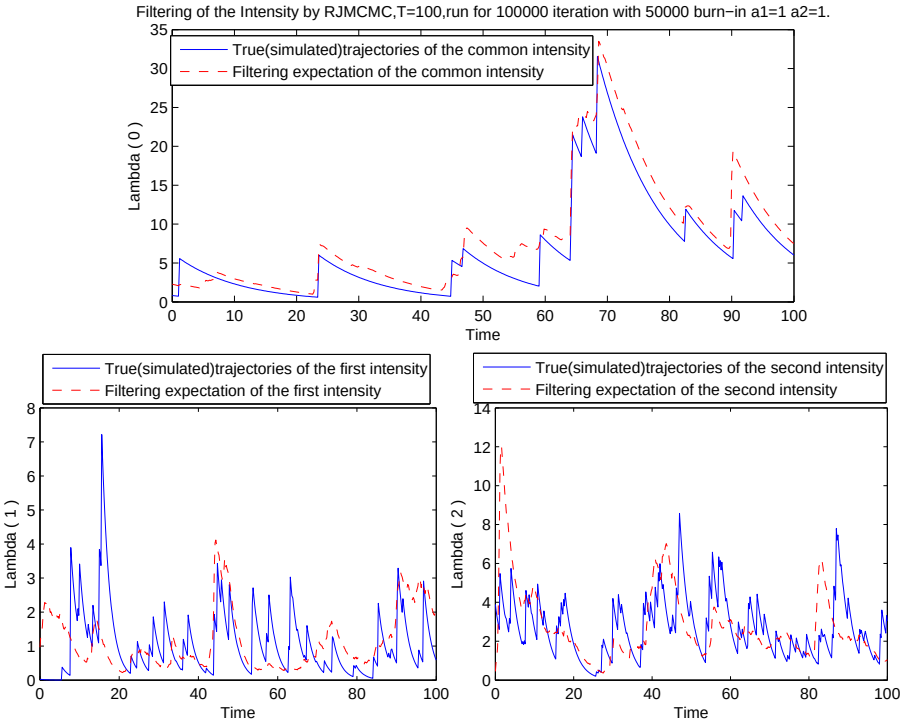

Figure 4.12: Bivariate intensity process. Common component of intensity $\lambda_t^{(0)}$ (upper figure). Individual component $\lambda_t^{(1)}$ (bottom left figure). Individual component $\lambda_t^{(2)}$ (bottom right figure). The solid line denotes the true intensity and the dotted line represents the filtering intensity. Results based on 100,000 updates with 50,000 burn-in period; $\nu_0 = 0.1$, $k_0 = 0.1$, $\gamma_0 = 0.2$, $\nu_1 = 0.5$, $k_1 = 0.5$, $\gamma_1 = 1$, $\nu_2 = 1$, $k_2 = 0.4$, $\gamma_2 = 1$, $a_1 = 1$, $a_2 = 1$, $\beta = 1$ (Equation 4.4.1).

Thus the second RJMCMC algorithm does better job than the first one, although the divergence of chain is still existing. Some diagnosis may be useful in this case. Here we will focus on the convergence and dependence of the starting value. So we run the second RJMCMC algorithm with 200,000 iterations and discarding 100,000 as burn-in period. The results is shown in Figure 4.11 and Figure 4.12. Note the second run is based on the same parameters $T = 100$, $\nu_0 = 0.1$, $k_0 = 0.1$, $\gamma_0 = 0.2$, $\nu_1 = 0.5$, $k_1 = 0.5$, $\gamma_1 = 1$, $\nu_2 = 1$, $k_2 = 0.4$, $\gamma_2 = 1$, $a_1 = 1$, $a_2 = 1$. Comparing the first run (see Figure 4.11 and Figure 4.12)) with the second run (see Figure 4.13 and Figure 4.14)), clearly, we can find the second run is better than the first one. The common component, for instance, the filtering of intensity (red dashed line) is almost the same as the ture intensity (blue solid line). On the other hand, the two individual intensity components (the bottom two panels), they are improved as well. At time 0 to 10, though the drift is not removed by doubled run of the chain, it was reduced a bit, taking the $\lambda^{(1)}$ for instance, at time 2, the predicted value is greater than 2 for the first run (100,000 iterations) while less than one for the second run

(200,000 iterations). Similarly, $\lambda^{(2)}$ is 12 for the first run with respect to 8 for the second run. Note that these two runs are with the same starting value for the chain.

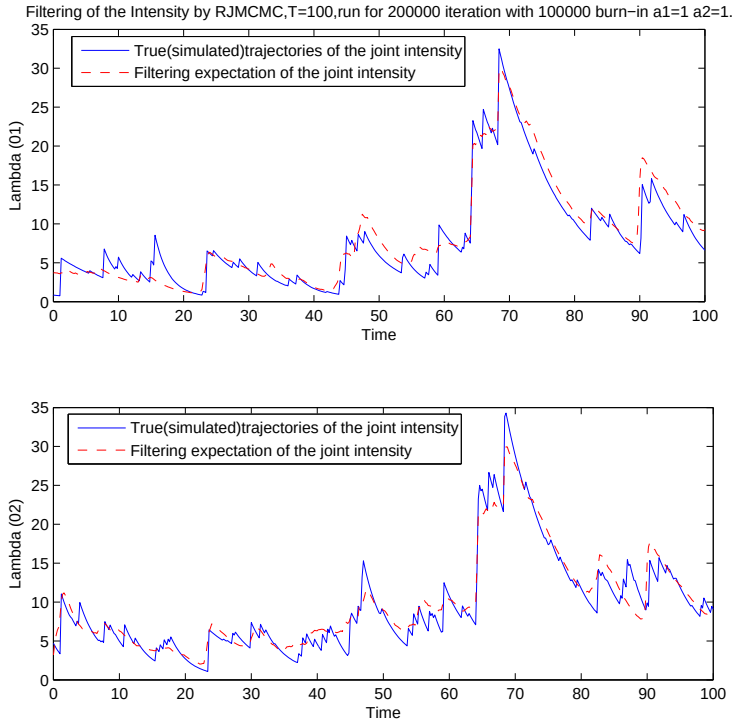

Figure 4.13: Bivariate intensty process. Combined intensity λ_t^{01} (upper figure). Combined intensity λ_t^{02} (bottom figure). The blue solid line denotes the true intensity and the red dash line represents the filtering intensity. Results based on 200,000 updates with 100,000 burn-in period; $\nu_0 = 0.1$, $k_0 = 0.1$, $\gamma_0 = 0.2$, $\nu_1 = 0.5$, $k_1 = 0.5$, $\gamma_1 = 1$, $\nu_2 = 1$, $k_2 = 0.4$, $\gamma_2 = 1$, $a_1 = 1$, $a_2 = 1$, $\beta = 1$ (Equation 4.4.1).

By running 200,000 iterations, the results is somehow leading us to believe that the chain has run sufficiently long. So it is reasonable to believe that the output adequately represents the target distribution. But, extremely speaking, we find the predicted value at some time points is different from the true one. This is indeed one of the most difficult problems to diagnose whether or not the chain has become stuck in one or more modes of the target distribution. In this case, all convergence diagnostics may indicate that the chain has converged, though the chain does not fully represent the target distribution. As suggested by many researchers and practioners, it is also important to consider a multiple chains from diverse starting values and then compare the within- and between-chain behavior.

The motivation for trying multiple runs is the hope that all interesting features (e.g. modes) of the target distribution will be explored by at least one chain, and that the failure of individual chains to find such features or to wash out the influence of their starting values can be detected, in which chains must be lengthened or the problem reparameterized to encourage better mixing. In this point, we also ran decades times of the second RJMCMC based on the same data as shown in Figure 4.11 and 4.12 with 100,000 iterations (50,000

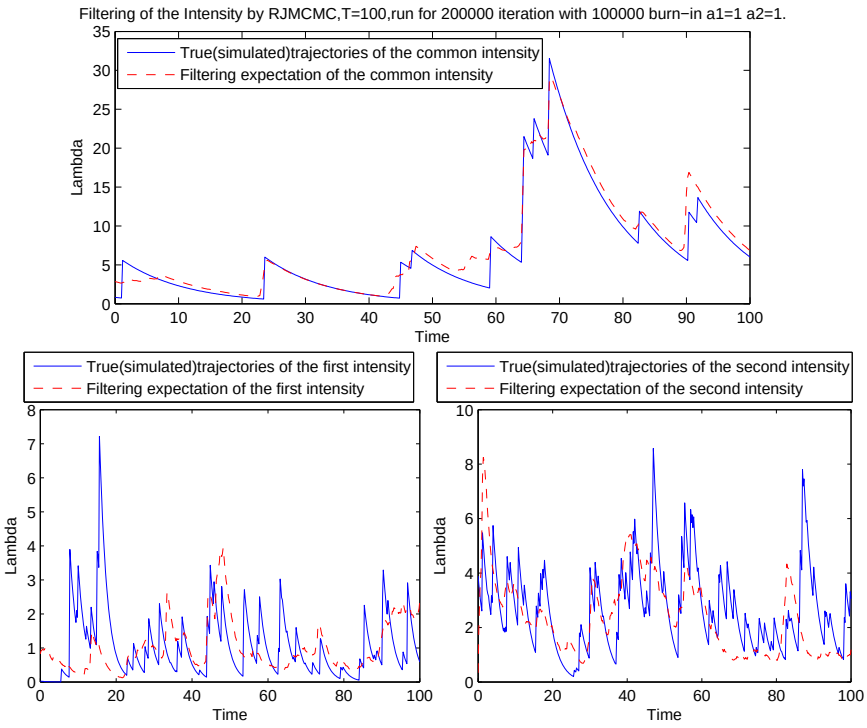

Figure 4.14: Bivariate intensity process. Common component of intensity $\lambda_t^{(0)}$ (upper figure). Individual component $\lambda_t^{(1)}$ (bottom left figure). Individual component $\lambda_t^{(2)}$ (bottom right figure). The blue solid line denotes the true intensity and the red line represents the filtering intensity. Results based on 200,000 updates with 100,000 burn-in period; $\nu_0 = 0.1$, $k_0 = 0.1$, $\gamma_0 = 0.2$, $\nu_1 = 0.5$, $k_1 = 0.5$, $\gamma_1 = 1$, $\nu_2 = 1$, $k_2 = 0.4$, $\gamma_2 = 1$, $a_1 = 1$, $a_2 = 1$, $\beta = 1$ (Equation 4.4.1).

burn-in period) by varying starting values, but the results were quite consistent with the previous one (as shown in Figure 4.11 and 4.12). Here we do not give out these results since they are very similar as Figure 4.11 and 4.12.

Last but not the least, it is important to report the time-consuming for the computing in the simulation study, but most of authors avoid addressing this problem. To some extent, the time-consuming concerns on the efficiency of the simulation, so it is useful information for the readers. Of course, there is no comparison among different simulation environment such as model specification, computing algorithm, and programming software etc. In this thesis, we use Matlab 7.4.0, a high-level computer lannguages for mathematics and statistics, to run the simulations and study statistical properties of the simulation result. After running of enormous simulations of the RJMCMC algorithm for our model, two results are obtained: first, the second RJMCMC algorithm needs more time than the first RJMCMC algorithm for running the same number of iterations on the same data, 2 times in general. The reason is quite intuitive that each iteration of the second RJMCMC algorithm needs to check the neighbourhood condition, while the first RJMCMC saves this checking procedure. Second, the dependence of the model specification, that is, parameters. This is a tricky reason for explaination of time-consuming. Figure 4.15 and Figure 4.16 show the filtering results based on different parameters by running of the second RJMCMC algorithm (100,000 iterations with 50,000 burn-in period). It takes 15572 seconds for Figure 4.15 with parameters $\nu_0 = 0.15$, $k_0 = 0.1$, $\gamma_0 = 0.2$, $\nu_1 = 0.1$, $k_1 = 0.1$, $\gamma_1 = 0.1$, $\nu_2 = 0.1$, $k_2 = 0.1$, $\gamma_2 = 0.15$, $a_1 = 1$, $a_2 = 1$, $\beta = 0.002$, whereas 41863 seconds for Figure 4.16 with parameters $\nu_0 = 0.15$, $k_0 = 0.1$, $\gamma_0 = 0.2$, $\nu_1 = 0.1$, $k_1 = 0.1$, $\gamma_1 = 0.1$, $\nu_2 = 0.1$, $k_2 = 0.15$, $\gamma_2 = 0.15$, $a_1 = 1$, $a_2 = 1$, $\beta = 0.002$. This difference (twice time-consuming) is mainly due to the intensities, if we go insight into the RJMCMC algorithm, each iteration involves with computing likelihood ratio, as shown in Equation 4.5.3, in which the number of transactions $N_t^{(k)}$ ($k = 0, 1, 2$) plays essential role in time consuming. And this $N_t^{(k)}$ ($k = 0, 1, 2$) depends on the underlying intensities. In our presented two examples, Figure 4.15 and Figure 4.16, we find the intensity 01 and the intensity 02 in Figure 4.16 are generally higher than those in Figure 4.15.

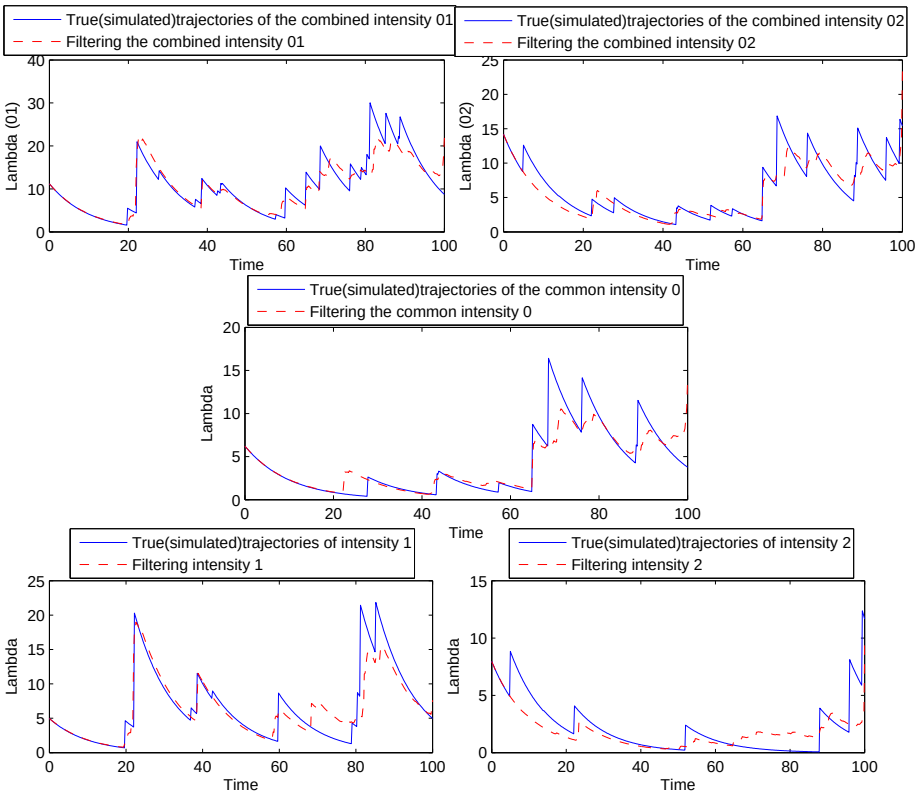

Figure 4.15: Bivariate intensity process. Combined intensity ($\lambda_t^{(01)}$) (upper left panel). Combined intensity ($\lambda_t^{(02)}$) (upper right panel). Common component of intensity ($\lambda_t^{(0)}$) (middle panel). Individual component ($\lambda_t^{(1)}$) (bottom left panel). Individual component ($\lambda_t^{(2)}$) (bottom right panel). Results based on 100,000 updates with 50,000 burn-in period; $\nu_0 = 0.15$, $k_0 = 0.1$, $\gamma_0 = 0.2$, $\nu_1 = 0.1$, $k_1 = 0.1$, $\gamma_1 = 0.1$, $\nu_2 = 0.1$, $k_2 = 0.1$, $\gamma_2 = 0.15$, $a_1 = 1$, $a_2 = 1$, $\beta = 0.002$ (Equation 4.4.1).

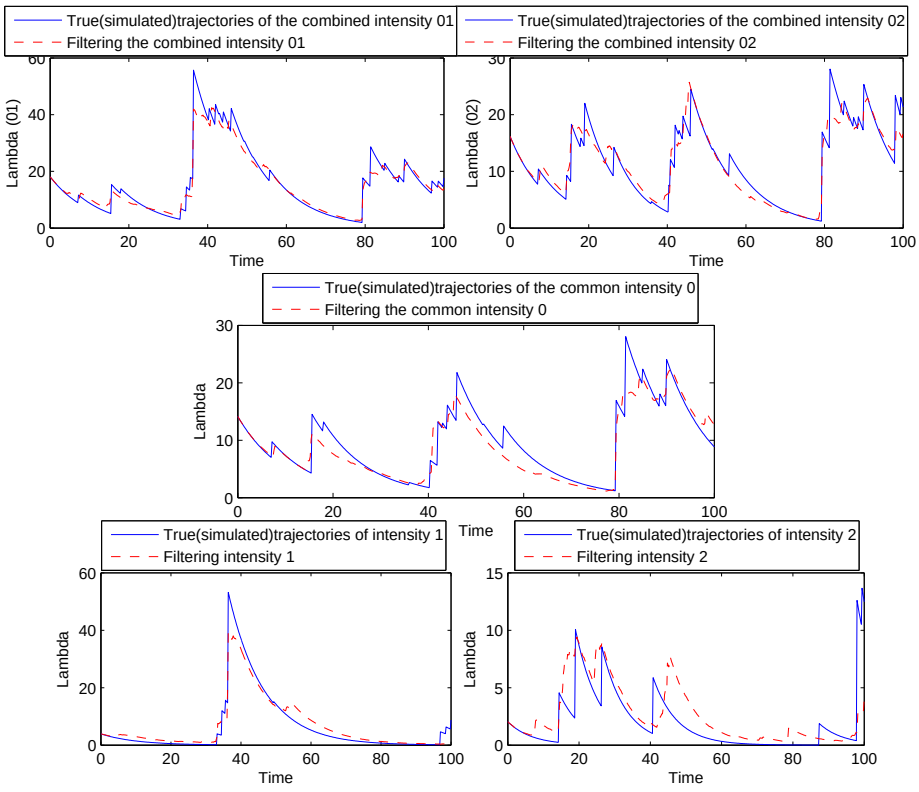

Figure 4.16: Bivariate intensity process. Combined intensity ($\lambda_t^{(01)}$) (upper left panel). Combined intensity ($\lambda_t^{(02)}$) (upper right panel). Common component of intensity ($\lambda_t^{(0)}$) (middle panel). Individual component ($\lambda_t^{(1)}$) (bottom left panel). Individual component ($\lambda_t^{(2)}$) (bottom right panel). Results based on 100,000 updates with 50,000 burn-in period; $\nu_0 = 0.15$, $k_0 = 0.1$, $\gamma_0 = 0.2$, $\nu_1 = 0.1$, $k_1 = 0.1$, $\gamma_1 = 0.1$, $\nu_2 = 0.1$, $k_2 = 0.15$, $\gamma_2 = 0.15$, $a_1 = 1$, $a_2 = 1$, $\beta = 0.002$ (Equation 4.4.1).

4.5.2 Tuning of the filtering algorithm

The essential difference between the first and the second RJMCMC algorithm is the neighbourhood condition, in which the endogenous simulaiton parameter β plays an important role. As we know, β serves as a constraint binding two components of intensity. It perhaps quite reasonable to expect large value of β would make no improvement of the second RJMCMC with respect to the first RJMCMC. On the other hand, a smaller β is likely to make the algorithm work better (for instance, $\beta = 1$ as shown in Figure 4.11- Figure 4.14). Of course, there is no precise value for β. It very depends on the algorithm itself and the model as well. In our case, the model with parameters, $\nu_0 = 0.1$, $k_0 = 0.1$, $\gamma_0 = 0.2$, $\nu_1 = 0.5$, $k_1 = 0.5$, $\gamma_1 = 1$, $\nu_2 = 1$, $k_2 = 0.4$, $\gamma_2 = 1$, $a_1 = 1$, $a_2 = 1$ (as shown in Figure 4.11- Figure 4.14), we show by simulation that the second RJMCMC works properly when β ranges from 0.001 to 0.01. It means when $0.001 \leq \beta \leq 0.01$, the filtering results are quite similar. But if $\beta > 0.01$, the neighbourhood condition does not improve the algorithm very much.

Getting insight into the performance of the second RJMCMC algorithm by varying value of β, we ran three independent simulation experiments with the same simulated trajectory $\mathbf{T}_0^T$ of event times. Figure 4.17 shows the simulation results with different values of β, $\beta = 0.5$ (red dashed line), $\beta = 0.01$ (blue dotted line), $\beta = 0.005$ (magenta dash-dot line). Note that β controls the distance of two components either for merge move or for split move. For merge move, two components are selected randomly satisfying the neighbourhood condition (controlling by β), then according to Equation 4.5.1, birth a new location for common component. On the other hand, for split move, birth two components within neighbourhood range (controlling by β) according to Equation 4.5.2. More details see Section 4.5.

Roughly, from Figure 4.17, as we expected, the smaller nearby region the better results obtainted. Here we specially focus on three components, the filtering results are improved gradually as shrinking the neighbourhood distance which is controlled by β. The most apparent improvement is individual component one as shown on the bottom left panel. If we review the combined intensities as shown on the upper panels, it is easy to find that the value of β does not affect the filtering result. This is quite obvious because the neighbourhood condition functions on components but not on combined intensities. Second, the first RJMCMC algorithm did good work for combined intensities even without the merge-split moves concerning on neighbourhood condition. Finally, again, β is very depending on the model as we declared previously.

Precisely, the mean square error (MSE) can be a natural measurement to evaluate the performance of the algorithm. The MSE of the proceeding intensities $\lambda_t^{(01)}$, $\lambda_t^{(02)}$ and components $\lambda_t^{(0)}$, $\lambda_t^{(1)}$ and $\lambda_t^{(2)}$ are computed as

$$\text{MSE}^{(0k)}(t) = \sqrt{(\hat{\text{E}}(\lambda_t^{(0k)}|_0^T) - \lambda_t^{(0k)})^2} \tag{4.5.4}$$

$$\text{MSE}^{(i)}(t) = \sqrt{(\hat{\text{E}}(\lambda_t^{(i)}|_0^T) - \lambda_t^{(i)})^2} \tag{4.5.5}$$

where $\hat{\text{E}}(\lambda_t^{(0k)}|_0^T)$ and $\hat{\text{E}}(\lambda_t^{(i)}|_0^T)$ are the simulated filtering expectation of intensities and components obtained with 200,000 iterations (50,000 burn-in period) respectively.

Through Figure 4.18 to Figure 4.20, we show the mean square erro of components of underlying intensity, $\hat{\mathrm{E}}(\lambda_t^{(i)}|_0^T)$, according to Equation 4.5.5, $i = 0, 1, 2$. The three plots are corresponding to different β, $\beta = 0.5$ (a), $\beta = 0.01$ (b) and $\beta = 0.005$ (c). Interestingly, the highest value of MSE, in any case of β, corresponds to highest peaks of the intensities. It is consistent with the simulation results as shown in Centanni & Minozzo (2006). Figure 4.18 depicts the MSE of common component based on these three different β, as we can see, it is slightly improved by shrinking β. Figure 4.19 and Figure 4.20 display the MSE of two individual components according to three different β, $\hat{\mathrm{E}}(\lambda_t^{(1)}|_0^T)$ and $\hat{\mathrm{E}}(\lambda_t^{(2)}|_0^T)$, respectively. Similarly, the neighbourhood condition improve the filtering algorithm by shrinking the region, that is, the value of β.

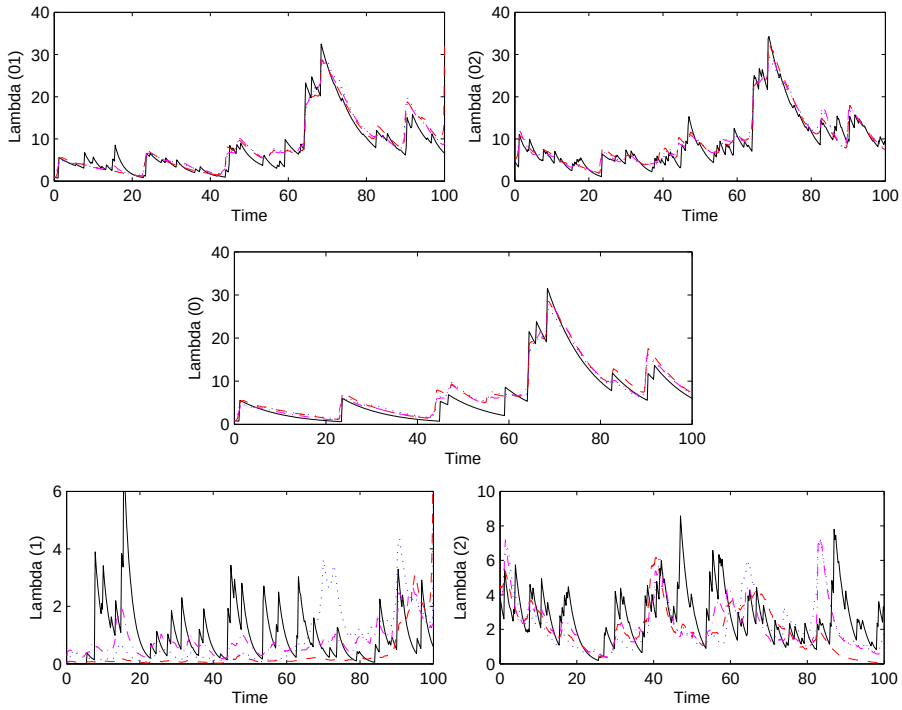

Figure 4.17: Filtering results of different β (parameter of neighbourhood condition) based on bivariate intensity process. Combined intensity ($\lambda_t^{(01)}$) (upper left panel). Combined intensity ($\lambda_t^{(02)}$) (upper right panel). Common component of intensity ($\lambda_t^{(0)}$) (middle panel). Individual component ($\lambda_t^{(1)}$) (bottom left panel). Individual component ($\lambda_t^{(2)}$) (bottom right panel). The black solid line denotes the true intensity; the red dash line represents the filtering intensity with $\beta = 0.5$; the blue dotted line is the filtering intensity with $\beta = 0.01$ and the magnenta dash-dot line is with $\beta = 0.005$. Results based on 200,000 updates with 50,000 burn-in period; $\nu_0 = 0.1$, $k_0 = 0.1$, $\gamma_0 = 0.2$, $\nu_1 = 0.5$, $k_1 = 0.5$, $\gamma_1 = 1$, $\nu_2 = 1$, $k_2 = 0.4$, $\gamma_2 = 1$, $a_1 = 1$, $a_2 = 1$ (Equation 4.4.1).

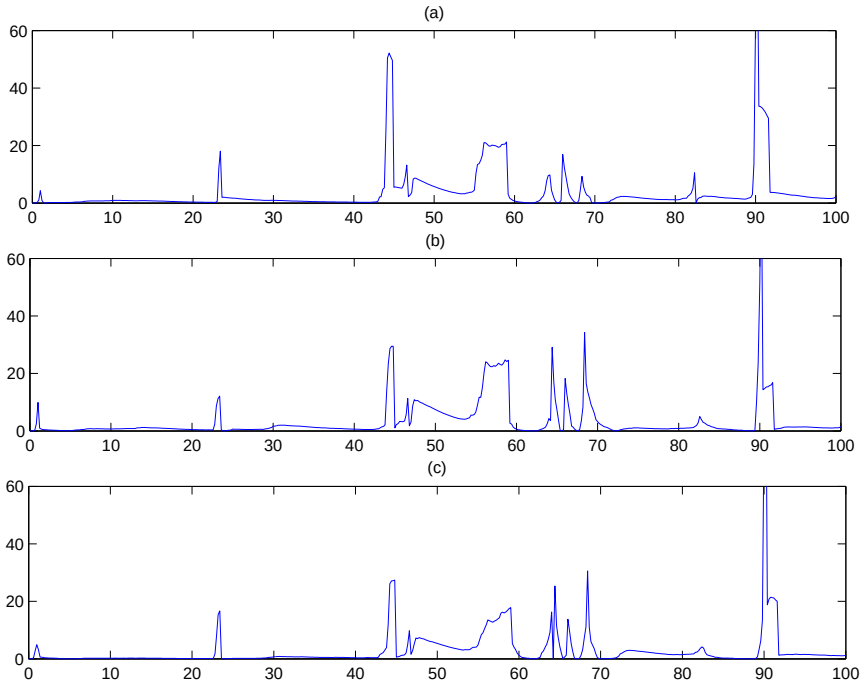

Figure 4.18: Mean Square Error of $\hat{\mathrm{E}}(\lambda_t^{(0)}|_0^T)$ (Equation 4.5.5). The three plots are corresponding to different β, $\beta = 0.5$ (a), $\beta = 0.01$ (b) and $\beta = 0.005$ (c). $t = 0, 0.2, 0.4, \ldots, 100$, $k = 1, 2$ and $i = 0, 1, 2$. Results based on 200,000 updates with 50,000 burn-in period; $\nu_0 = 0.1$, $k_0 = 0.1$, $\gamma_0 = 0.2$, $\nu_1 = 0.5$, $k_1 = 0.5$, $\gamma_1 = 1$, $\nu_2 = 1$, $k_2 = 0.4$, $\gamma_2 = 1$, $a_1 = 1$, $a_2 = 1$ (Equation 4.4.1).

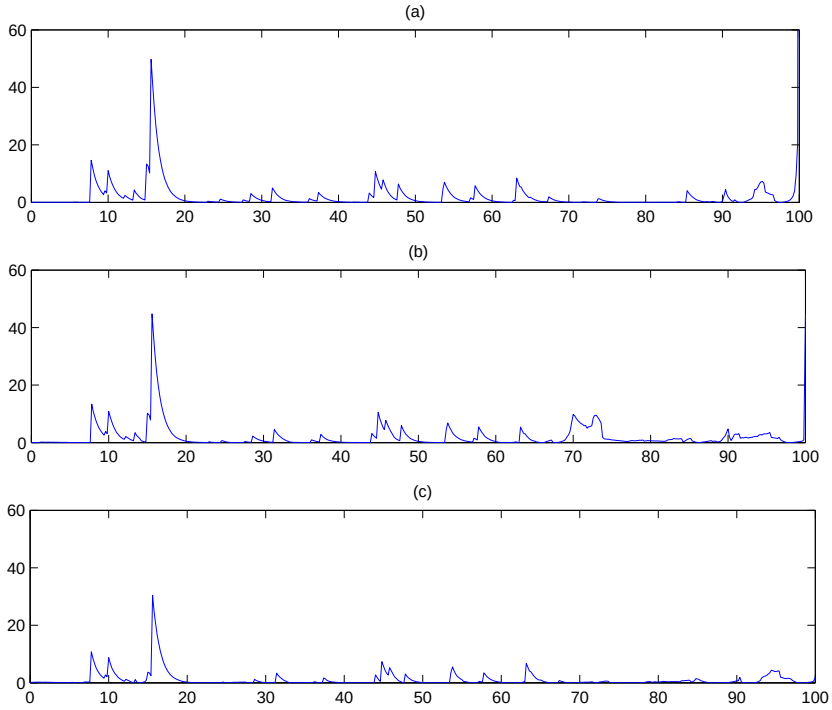

Figure 4.19: Mean Square Error of $\hat{\mathrm{E}}(\lambda_t^{(1)}|_0^T)$ (Equation 4.5.5). The three plots are corresponding to different β, $\beta = 0.5$ (a), $\beta = 0.01$ (b) and $\beta = 0.005$ (c). $t = 0, 0.2, 0.4, \ldots, 100$, $k = 1, 2$ and $i = 0, 1, 2$. Results based on 200,000 updates with 50,000 burn-in period; $\nu_0 = 0.1$, $k_0 = 0.1$, $\gamma_0 = 0.2$, $\nu_1 = 0.5$, $k_1 = 0.5$, $\gamma_1 = 1$, $\nu_2 = 1$, $k_2 = 0.4$, $\gamma_2 = 1$, $a_1 = 1$, $a_2 = 1$ (Equation 4.4.1).

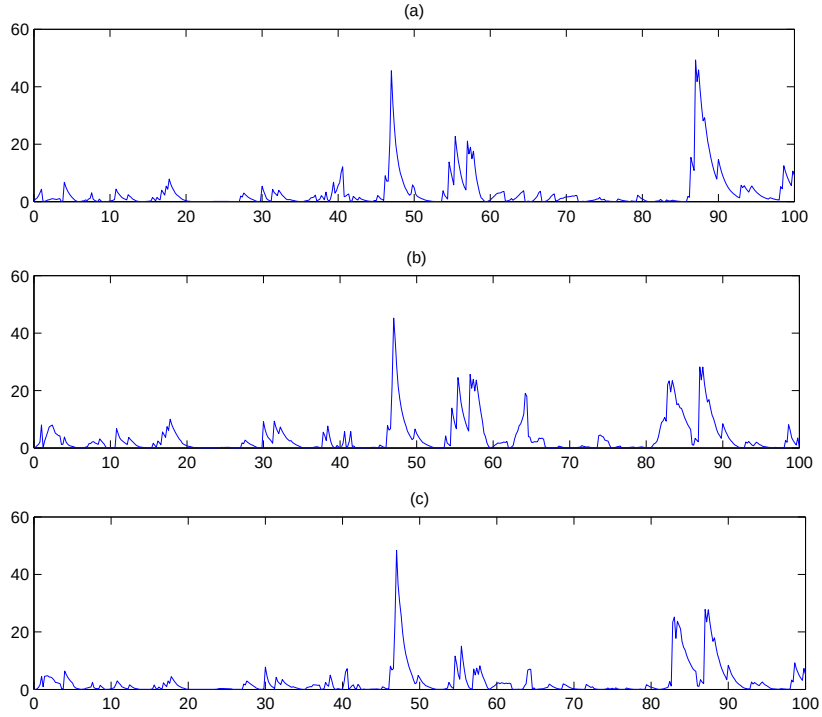

Figure 4.20: Mean Square Error of $\hat{\mathrm{E}}(\lambda_t^{(2)}|_0^T)$ (Equation 4.5.5). The three plots are corresponding to different β, $\beta = 0.5$ (a), $\beta = 0.01$ (b) and $\beta = 0.005$ (c). $t = 0, 0.2, 0.4, \ldots, 100$, $k = 1, 2$ and $i = 0, 1, 2$. Results based on 200,000 updates with 50,000 burn-in period; $\nu_0 = 0.1$, $k_0 = 0.1$, $\gamma_0 = 0.2$, $\nu_1 = 0.5$, $k_1 = 0.5$, $\gamma_1 = 1$, $\nu_2 = 1$, $k_2 = 0.4$, $\gamma_2 = 1$, $a_1 = 1$, $a_2 = 1$ (Equation 4.4.1).

4.6 Conclusions

In this thesis, we propose a new type of multivariate model based on marked point processes, in particular on marked doubly stochastic Poisson processes. The main idea is to specify the intensity function of each univariate DSPP as a combination of two components, one individual specific component and one common component, both unobservable and belonging themselves to the class of marked doubly stochastic Poisson processes. Hence, our specification extends the univariate model proposed by Centanni & Minozzo (2006) by allowing for a common component jointly driving the individual intensities. Moreover, our model could be compared with the stochastic conditional intensity model (SCI) proposed by Bauwens & Hautsch (2006), which is based on an observation-driven component (observable) and a dynamic latent component (unobservable). The latent component in the SCI model is specified as a log-linear model based on a Gaussian autoregressive process of the first order, while the two unobservable components in our model follow a shot noise process. Although the SCI model is different from our model, they share the presence of a common factor driving all the intensities of the univariate DSPPs,that is, modeling the joint dynamic of the multivariate system.

Since the intensity processes are not observable, their conditional and unconditional moments cannot typically be computed analytically, and we must resort to some simulation methods. In particular, we develop suitable reversible jump Markov chain Monte Carlo algorithms to simulate the conditional distribution of the intensities given the data. We construct two different RJMCMC algorithms for filtering the intensities of the model (and their components) and show by an extensive simulation study that the performance of the first algorithm, although able to filter the intensities, is not as good as the second in the filtering of the components.

Bibliography

E. Abbé (1879). Über Blutkörper-Zahlung. Jena Z. Med. Naturwiss, 13 (New Series 6), 98–105.

T. G. Andersen and T. Bolleslev (1997). Intraday periodicity and volatility persistence in financial markets. Journal of Empirical Finance, 4, 115–158.

G. Ballocchi, M. M. Dacorogna, R. Gencay and B. Piccinato (1999a). Intraday statistical properties of Eurofutures. Ferivatives Quarterly, 6, 28–44.

G. Ballocchi, M. M. Dacorogna, C. M. Hopman, U. A. Müller, and R. B. Olsen (1999b). The intraday multivariate structure of the Eurofutures markets. Journal of Empirical Fiance, 6, 479–513.

O. E. Barndorff-Nielsen and N. Shephard (2004). Econometic Analysis of Realized Covariation: High Frequency based Covariance, Regression, and Correlation in Financial Eonomics. Econometrica, 72 (3), 885–925.

O. E. Barndorff-Nielsen, P. R. Hansen, A. Lunde and N. Shepard (2008). Designing realized kernels to measure the ex-post variation of equity prices in presence of noise. Econometrica, 76 (6), 1481–1536.

O. E. Barndorff-Nielsen, W. S. Kendall, and M. N. M. van Lieshout (1998). Stochastic Geometry: Likelihood and Computation. Chapman and Hall.

M. S. Bartlett (1963). The spectral analysis of point processes. J. Roy. Statist. Soc. Ser. B, 25, 264–296.

M. S. Bartlett (1964). The spectral analysis of two-dimensional point processes. Biometrika, 51, 299–311.

E. Barucci, R. Renò (2002). On measuring volatility and the GARCH forecasting performance. Journal of International Financial Markets, Institutions and Money, 12, 182–200.

L. Bauwens and P. Giot (2000). The logarithmic ACD model: An application to the bid/ask quote process of two NYSE stocks. Annales d'Economie et de Statistique, 60, 117–149.

L. Bauwens and P. Giot (2001). Econometric Modelling of Stock Market Intraday Activity. Kluwer Academic Publishers, Boston, Dordrechet, London.

L. Bauwens and P. Giot (2003). Asymmetric ACD models: Introducing price information in ACD models with a two state transition model. Empirical Economics, 28, 1–23.

L. Bauwens and N. Hautsch (2006). Stochastic conditional intensity processes. J. Financial Econometrics, 4, 450–493.

L. Bauwens and N. Hautsch (2008). Modelling fiancial high frequency data using point processes. (forthcoming). T. A. Andersen, R. A. Davis, J. P. Kreiss, T. Mikosch. Handbook of Financial Times Series Econometrics.

L. Bauwens and D. Veredas (2004). The stochastic conditional duration model: a latent factor model for the analysis of financial durations. Journal of Econometrics, 119 (2), 381–421.

J. Bertoin (1996). Lévy Processes. Cambridge University Press.

J. Besag (2000). Markov chain Monte Carlo for statistical inference. Working paper. University of Washington, Center for Statistics and the Social Sciences, Seattle.

A. Beskos, O. Papaspiliopolos and G. Roberts (2007). Monte Carlo maximum likelihood estimation for discretely observed diffusion processes. Annals of Statistics.

B. Biais, T. Foucault and P. Hillion (1997). Microstructure des Marchés Financiers: Institutions, Modèles et Tests Empiriques. Paris: Presses Universitaires de France.

R. Bloomfield and M. O'Hara (1999). Market transparency: who wins and who loses? Review of Financial Studies, 12, 5–35.

T. Bollerslev (1986). Generalized autoregressive conditional heteroskedasticity. Journal of Econometrics, 31, 307–327.

T. Bollerslev and I. Domowitz (1993). Trading patterns and prices in the interbank foreign exchange market. Journal of Finance, 48, 1421–1443.

C. G. Bowsher (2007). Modelling security markets in continous time: Intensity based, multivariate point process models. Journal of Econometrics, 141, 876–912.

P. Brémaud (1981). Point Processes and Queues: Martingale Dynamics. Springer.

W. Breymann, A. Dias, P. Embrechts (2003). Dependence Structure for Multivariate High-Frequency Data in Finance. Quantitative Finance, 3 (1), 1–16.

S. Centanni and M. Minozzo (2006a). Estimation and filtering by reversible jump MCMC for a doubly stochastic Poisson model for ultra-high-frequency financial data. Statistical Modelling, 97–118.

S. Centanni and M. Minozzo (2006b). A Monte Carlo approach to filtering for a class of marked doubly stochastic Poisson processes. Journal of the American Statistical Association, 1582–1597.

K. Chan (1992). A further analysis of the lead-lag relationships between the cash market and the stock index futures market. Review of Financial Studies, 5, 123–152.

K. Chan (1993). Imperfect information and cross-autocorrelation among stock prices. Journal of Finance, 48 (4), 1211–1230.

F. Corsi (2006). Realized Correlation Tick-by-Tick. Working paper, University of Lugano.

D. R. Cox (1955). Some statistical methods connected with series of events (with discussion). J. Roy. Statist. Soc. Ser. B, 17, 129–164.

D. R. Cox and P. A. W. Lewis (1966). The Statistical Analysis of Series of Events. Methuen, London.

H. Cramér and M. R. Leadbetter (1967). Stationary and Related Stochastic Processes. Wiley.

N. A. C. Cressie (1991). Statistics for Spatial Data. (Rev. ed. 1993.). Wiley.

M. M. Dacorogna, R. Gencay, U. Müller, R. B. Olsen and O. V. Pictet (2008). An Introduction to High-Frequency Finance. Academic Press.

D. J. Daley (1974). Various concepts of orderliness for point processes. (in Harding and Kendall (1974), 148–161). Wiley.

D. J. Daley and D. Vere-Jones (1972). A summary of the theory of point processes. (in Lewis (1972), 299–383). Wiley.

D. J. Daley and D. Vere-Jones (2003). An Introduction to the Theory of Point Processes, volume I. Springer.

D. A. Dawson, K. Fleischmann, and C. Mueller (2000). Finite time extinction of superprocesses with catalysts. Ann. Probab., 28, 603–642.

F. de Jong, Th. Nijman (1997). Journal of Empirical Finance, 4, 259–277

A. Dias, P. Embrechts (2004). Dynamic copula models for multivariate high-frequency data in finance. Manuscript, ETH Zurich.

P. J. Diggle(1983). Statistical Analysis of Spatial Point Patterns. Academic Press.

A. Dufour and R. F. Engle (2000). The ACD model: Predictability of the time between consecutive trades. (working paper). ISMA Centre, University of Reading.

D. Easley and M. O'Hara (1987). Price, trade size, and information in securities market. Journal of Financial Economics, 19, 69–90.

D. Easley and M. O'Hara (1992). Time and process of security price adjustment. Journal of Finance, 47, 576–605.

P. Embrechts, C. Klüppelberg, and T. Mikosch (1997). Modelling Extremal Events. Springer.

R. F. Engel (1982). Autoregressive Conditional Heteroscedasticity with estimates of the variance of United Kingdom inflation. Econometrica, 50, 987–1006.

R. F. Engel (2000). The econometrics of ultra-high-frequency data. Econometrica, 1–22.

R. F. Engel and J. R. Russel (1998). Autoregressive conditional duration: a new model for irregularly spaced transactions data. Econometrica, 66, 1127–1161.

T. Epps (1979). Co-movements in stock prices in the very short-run. Journal of the American Statistical Association, 74, 291–298.

A. K. Erlang (1909). The theory of probabilities and telephone conversations.(Reprinted in E. Brockmeyer, H. L. Halstrom and A. Jensen (1948), The Life and Works of A. K. Erlang, Copenhagen Telephone Company, Copenhagen, 131-137). Nyt. Tidsskr. Mat. B, 20, 33–41.

P. Fearnhead, O. Papaspiliopoulos and G. O. Roberts (2008). Particle filters for partially observed diffusions. J. R. Statist. Soc. B, 70 (4), 755–777.

W. Feller (1966). An Introduction to Probability Theory and its Applications, Vol. II. Wiley.

D. Feng, G. Jiang and P. Song (2004). Stochastic conditional duration models with leverage effect for financial transaction data. Journal of Financial Econometrics, 2, 390–421.

M. Fernandes and J. Gramming (2006). A family of autoregressive conditional duration models. Journal of Econometrics, 130, 1–23.

R. Frey (2000). Risk-minimization with incomplete information in a model for high-frequency data. Mathematical Finance 10, 215–25.

R. Frey and W. J. Runggaldier (2001). A nonlinear filtering approach to volatility estimation with a view towards high frequency data. International Journal in Theoretical and Applied Finance 4, 271–300.

E. Ghysels and J. Jasiak (1995). Stochastic volatility and time deformation: An application of trading volume and leverage effects. In Proceedings of the HFDF-I Conference, Zurich, Switzerland, March 29-31, 1995, 1, 1–14.

W. R. Gilks, S. Richardson and D. J. Spiegelhalter (1996). Markov Chain Monte Carlo in Practice. Chapman and Hall.

L. Glosten and P. Milgrom (1985). Bid, ask, and the transaction prices in a specialist market with heterogeneously informed traders. Journal of Financial Economics, 13, 71–100.

C. A. E. Goodhart (1989). News and the foreign exchange market. In Proceedings of the Manchester Statistical Society, 1–79.

C. A. E. Goodhart and L. Figliuoli (1991). Every minute counts in financial markets. Journal of International Money and Finance, 10, 23–52.

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

D. M. Guillaume, M. M. Dacorogna, R. D. Davé, U. A. Müller, R. B. Olsen, and O. V. Pictet (1997). From the bird's eye to the microscope: a survey of new stylized facts of the intraday foreign exchange markets. Finance and Statistics, 1, 95–129.

N. Hautsh (2008). Capturing common components in high-frequency financial time series: a multivariate stochastic multiplicative error model. Journal of Economic Dynamics and control, 32, 3978–4015.

A. G. Hawkes (1971a). Spectra of some self-exciting and mutually exciting point processes. Biometrika, 58, 83–90.

A. G. Hawkes (1971b). Point spectra of some mutually exciting point processes. J. Roy. Statist. Soc. Ser. B, 33, 438–443.

A. G. Hawkes (1972). Spectra of some mutually exciting point processes with associated variables. (in Lewis (1972), 261–271). Wiley.

A. G. Hawkes and L. Adamopoulos (1973). Cluster models for earthquakes regional comparisons. Bull. Int. Statist. Inst., 45 (3), 454–461.

A. G. Hawkes and D. Oakes (1974). A cluster representation of a self-exciting process. J. Appl. Probab., 11, 493–503.

A. Heinen and E. Rengifo (2003). Multivariate autoregressive modelling of time series count data using copulas. (discussion Paper 2203–2225), CORE, Université Catholique de Louvain.

F. Heiss (2008). Squential Numerical Integration In Nonlinear State Space Models For Microeconometric Panel Data. Journal Of Applied Eoconometrics, 23, 373–389.

M. Jacobsen (1982). Statistical Analysis of Counting Processes. Lecture Notes in Statistics 12. Springer.

A. F. Karr (1986). Point Processes and Their Statistical Inference. (2nd ed. 1991.). Marcel Dekker, New York.

J. F. C. Kingman (1993). Poisson Processes. Clarendon Press, Oxford.

Y. A. Kutoyants (1980). Estimation of Parameters of Stochastic Processes (in Russian). Armenian Academy of Science, Erevan.

Y. A. Kutoyants (1984). Parameter Estimation for Stochastic Processes. (Translated by B. L. S. Prakasa Rao and Revised by Kutoyants (1980).) Heldermann, Berlin.

Y. A. Kutoyants (1998). Statistical Inference for Spatial Poisson Processes, Lecture Notes in Statistics 134. Springer.

G. Last and A. Brandt (1995). Marked Point Processes on the Real Line. Springer.

M. R. Leadbetter (1972). On basic results of point process theory. Proc.Sixth Berkeley Symp. Math. Statist. Probab., 3, 449–462.

M. R. Leadbetter, G. Lindgren, and H. Rootzen (1983). Extremes and Related Properties of Random Sequences and Processes. Springer.

P. A. W. Lewis (1964a). A branching Poisson process model for the analysis of computer failure patterns (with discussion). J. Roy. Statist. Soc. Ser. B, 26, 398–456.

P. A. W. Lewis (1964b). The implications of a failure model for the use and maintenance of computers. J. Appl. Probab., 1, 347–368.

P. A. W. Lewis (1970). Remarks on the theory, computation and application of the spectral analysis of series of events. J. Sound Vib., 12 (3), 353–375.

P. A. W. Lewis (1972). Stochastic Point Processes. Wiley.

P. A. W. Lewis and G. S. Shedler (1976). Simulation of nonhomogeneous Poisson processes with log linear rate function. Biometrika, 63, 501–506.

R. Liesenfeld, J. F. Richard (2008). Improving MCMC, using efficient importance sampling. Computational Statistics and Data Analysis, 53, 272–288.

R. S. Liptser and A. N. Shiryayev (1974). Statistics of Random Processes (in Russian).(Translation (1977, 1978).). Nauka, Moscow.

R. S. Liptser and A. N. Shiryayev (1977). Statistics of Random Processes, I: General Theory. Springer.

R. S. Liptser and A. N. Shiryayev (1978). Statistics of Random Processes, II: Applications. Springer.

A. Lo and A. C. MacKinlay (1990). An econometric analysis of nonsynchronous trading. Journal of Econometrics, 45, 181–211.

A. Low, J. Muthuswamy and S. Sarkar (1996). Time variation in the correlation structure of exchange rates: high frequency analyses. In Proceedings of the Third International Conference on Forecasting Financial Markets, London, 1, 1–24.

M. Lundin, M. M. Dacorogna (1999). Correlation of high-frequency financial time series. In Financial Markets Tick by Tick, (P. Lequeux Editor), Wiley.

J. F. Lyon and R. Thoma (1881). Ueber die Methode der Blutkörperzählung. Virchows. Arch. Path. Anat. Physiol., 84, 131–154.

A. Madhavan (2000). Market microstructure: a survey. Journal of Financial Markets, 3, 205–258.

S. Manganelli (2005). Duration, volume and volatility impact of trades. Journal of Financial Markets, 8, 377–399.

J. A. McFadden (1956). The axis-crossing intervals of random functions. I. Trans. Inst. Radio Engnrs, IT-2, 146–150.

J. A. McFadden (1958). The axis-crossing intervals of random functions. II. Trans. Inst. Radio Engnrs, IT-4, 14–24.

M. Meitz and T. Teräsvirta (2006). Evaluating models of autoregressive conditional duration. Journal of Business and Economic Statistics, 24, 104–124.

M. O'Hara (1995). Market Microstructure Theory. Basil Blackwell, Oxford.

Y. Ogata (1988). Statistical models for earthquake occurrence and residual analysis for point processes. Journal of the American Statistical Association, 83, 9–27.

C. Palm (1943). Intensitätsschwankungen im Fernsprechverkehr. Ericsson Technics, 44.

M. K. Pitt and N. Shephard (1999). Filtering via Simulation: Auxiliary Particle Filters. Journal of the American Statistical Association, 94, 590–599.

J. Prigent, O. Renault and O. Scaillet (2001). An Autoregressive Conditional Binomial Option Pricing Model. Selected papers from the First World Congress of the Bachelier Finance Society. Geman and Madan and Pliska and Vorst, Springer.

R. Renò (2003). A closer look at the Epps effect. International Journal of Theoretical and Applied Finance, 6 (1), 87–102.

S. I. Resnick (1987). Extreme Values, Regular Variation, and Point Processes. Springer.

S. O. Rice (1944). Mathematical analysis of random noise. (Reprinted in N. Wax (1954). Selected Papers on Noise and Stochastic Processes. Dover, New York, 133-294.). Bell Syst. Tech. J., 23, 282-332 and 24, 46–156.

S. Richardson and P. Green (1997). On Bayesian analysis of mixtures with unknown number of components. J. Roy. Stat. Soc., B, 59, 731–758.

B. D. Ripley (1981). Spatial Statistics. Wiley.

C. P. Roberts and G. Casella (1999). Monte Carlo Statistical Methods. Springer.

G. Roberts, O. Papaspilopoulos and P. Dellaportas (2004). Bayesian inference for non-Gaussian Ornstein-Uhlenbeck stochastic volatility processes. J. Roy. Stat. Soc., B, 369–393.

J. R. Russell (1999). Econometric modeling of multivariate irregulary-spaced high-frequency data. (discussion paper). University of Chicago.

J. R. Russell and R. F. Engle (2005). A discrete-state continuous-time model of financial transactions prices and times: The autoregressive conditional duration model. Journal of Business and Economic Statistics, 23, 166–180.

T. H. Rydberg and N. Shephard (2000). A modelling framework for the prices and times of trades made on the New York stock exchange. In Nonlinear and Nonstationary Signal Processing. W. J. Fitzgerald, R. L. Smith, A. T. Walden and P. C. Young, editors. Cambridge University Press, 217–26.

H. Seidel (1876). Über die Probabilitäten solcher Ereignisse welche nur seiten vorkommen, obgleich sie unbeschränkt oft möglich sind. Sitzungsber. Math. Phys. Cl. Akad.Wiss. München, 6, 44–50.

J. Serra (1982). Image Analysis and Mathematical Morphology. Academic Press, London.

O. Simonsen (2006). An empirical model for durations in stocks. Annals of Finance.

L. Spierfijk, T. E. Nijman and A. H. O. Van Soest (2002). The price impact of trades in illiquid stocks in periods of high and low market activity. (discussion paper). Center for Economic Research, Tilburg University.

D. L. Snyder (1972). Filtering and detection for doubly stochastic Poisson processes. IEEE Trans. Inf. Theory, IT-18, 97–102.

D. L. Snyder (1975). Random Point Processes. Wiley.

D. L. Snyder and M. I. Miller (1991). Random Point Processes in Time and Space.(= 2nd ed. of Snyder (1975).). Wiley.

Student (1907). On the error of counting with a haemacytometer. Biometrika, 5, 351–360.

S. J. Taylor (2008). Modelling Financial Time Series. Wiley.

D. Vere-Jones (1970). Stochastic models for earthquake occurrence. J. Roy. Stat. Soc., B, 32, 1–62.

D. Vere-Jones and T. Ozaki (1982). Some example of statistical inference applied to earthquake data. Annals of the Institute of Statistical Mathematics, 34, 189–207.

A. Zebedee (2001). A closer look at co-movements among stock returns. (working paper), San Diego State University.

M. Y. Zhang, J. Russell and R. S. Tsay (2001). A nonlinear autoregressive conditional duration model with application to financial transaction data. Journal of Econometrics, 104, 179–207.