Volatility Modelling with High-Frequency Financial Data on a Continuous Time Scale

GEORGIOS SOFRONIS, B.Sc., M.Sc.

Submitted for the degree of Doctor of Statistics at Lancaster university.

May 21, 2023

Abstract

Georgios Sofronis

Volatility Modelling with High-Frequency Financial Data on a Continuous Time Scale

Financial volatility is the core of multiple sectors in finance. This work investigates different aspects of volatility using high-frequency data. High-frequency data offer a complete picture of the dynamics of the intraday patterns, contributing to a more precise inference about these patterns. However, their complex structural form yields several challenges in the analysis for the practitioners. Our research takes place in both the univariate and multivariate space, meaning that we explore the data characteristics for every asset separately and as a factor of interactions among the assets.

In terms of the analysis in the univariate space, Chapters 2 and 4 develop some volatility estimators in discrete and continuous time scales, respectively. More specifically, we develop several estimators of the intraday volatility in Chapter 2 where each estimator approximates the intraday volatility, exploiting different characteristics of the dataset. On the other hand, we consider an estimator of the daily volatility along with its theoretical framework in Chapter 4. Our simulation study shows that our estimator is superior to standard estimators of daily volatility when the variance of the noise

incorporated in the intraday observations takes values of normal size.

In the multivariate space, Chapter 3 studies whether we can decompose the daily volatility traits to some components, inferring the assets which drive these components the most. Also, we extend the relevant methodology to volatility estimates with high frequency, as those provided by the estimators in Chapter 2. Through our proposed approach, we can deduce the stocks which present the highest variability as well as the intraday periods this variability is observed more intensely.

In Chapter 5, we develop a technique for estimating the conditional dependence structure between the assets using the concept of graphical models. This chapter treats high-frequency data as functional data, allowing us to exploit their virtues to draw inferences about the assets' conditional interdependencies.

Acknowledgement

In this section, I would like to acknowledge the contribution of some people who have played a decisive role during my educational journey.

I would like to express my warmest thanks to my supervisor, Dr Juhyun Park, for the cooperation, her guidance and the countless hours we have spent discussing my research issues. Dr Park is a mentor who taught me how a statistician should think in order to confront complex research challenges.

Also, I highly appreciate the acceptance of my supervisor's request from Dr Alex Gibberd and Dr Mikko Pakkanen to be my viva's internal and external examiners, respectively.

I am deeply thankful to Lancaster University and its staff for honouring me with a full scholarship for my doctoral studies and making my investigation more painless, supplying all the comforts I have needed for my research.

I would like to thank Prof. Ingmar Nolte and Dr Sandra Nolte for their insightful comments on my research during the first year of my doctoral studies and the provision of the dataset for this study.

In addition, my special thanks to Dr Stylianos Xanthopoulos and Dr Markus Jochmann who supervised my dissertation during my undergraduate and postgraduate studies, respectively. The positive interaction with my former supervisors was a key factor in my decision to extend my studies with a PhD.

Next, I would like to thank Effrosini Daouti, Eirini Daouti, Prof. Emmanuel Koukios

and Dr Panagiotis Korovessis for their support, motivation and guidance while I was planning my next steps after the completion of my undergraduate studies.

Finally, I would like to express my gratitude to my parents, Ioannis Sofronis and Ioanna Davouti, my sister, Eirini Sofroni and my grandmother, Eirini Solomou. All their aid, backing and encouragement have always been priceless to me.

Declaration

I hereby declare that the work in this thesis titled "Volatility Modelling with High-Frequency Financial Data on a Continuous Time Scale" has been done by me. Also, this thesis has not been submitted elsewhere for the award of any other degree. The results in this study were produced using the programming language MATLAB, where all codes and algorithms either made by me or they are standard MATLAB codes. For the ACF plots in Chapter 3, the function 'autocorr' is used and requires the installation of the toolbox 'Econometrics Toolbox'. Besides, the figures in this thesis have been created by me using MATLAB, unless something else is clearly specified in the main body of this work. The word count of this thesis is 54,272 words.

Contents

List of Figures

List of Tables

Chapter 1

Introduction, Motivation and Preliminaries

1.1 Introduction

In the last decades, the rapid growth of technology has led to the exploration of new fields in sciences, where first the deep investigation of these fields was not an easy task. Nowadays, we are able to collect and handle large datasets more easily than 50 years ago. The swift development of computer science has enabled us to manipulate and study the patterns that arise in a dataset of millions or even billions of observations in a relatively painless interval of time. Finance is one of those areas that has benefited from this development. Now researchers are able to analyse financial data with higher frequencies than the common daily scale, contributing to more precise inferences about the intraday assets' behaviour. This work deals with the concept of financial volatility in the stock market with high-frequency financial data. In this study, we often remove the term 'financial' from both terms for conciseness.

High-frequency data refer to intraday data which contain every piece of information observed in the market within a day, as opposed to daily data which summarises in a

scalar-valued observation the intraday behaviour of an asset. Thus, high-frequency data enables us to investigate the dynamics in the dataset which are not evident in the daily time series, such as the stock closing prices. This fact, in turn, allows us to shed light on the data's intraday trends or even more about the factors that direct these trends. In finance, some of the most common high-frequency observations are the stock prices, stock returns and transaction times.

In literature, logarithmic returns (log-returns) are preferred over stock prices because they exhibit more handy properties in terms of the analysis procedures (Tsay, 2005). As we will see later, the log-returns are defined as the difference between two consecutive logarithms of prices. In daily time series, the last intraday logarithmic prices (closing log-prices) are used for the daily log-returns calculation. The time point at which a closing price is released does not substantially differ from day to day and it occurs in the very last minutes or seconds of the trading session. Therefore, the log-returns are considered equally spaced observations in practice since every return refers to a period of approximately 24 hours.

However, this assumption does not hold for high-frequency returns, making their analysis intricate sometimes. In particular, pointwise intraday returns can occur at different time points, which may significantly differ from intraday interval to intraday interval and from day to day. Hence, the pointwise returns of an asset are likely to refer to intraday periods with noticeable time differences depending on the asset's trading frequency. This event can generate confusion. For example, two returns of an asset with a similar value may refer to intraday periods with a distinctive length difference. We can overcome this kind of irregularity in the realisation of the dataset in three different ways.

First, the researcher can rescale the pointwise returns of irregularly spaced observation times so that they define the returns of equilength periods using Formula (1.6). As we discuss later, this rescaling procedure could produce returns of different sizes when

the frequency of trading varies greatly, distorting the intraday patterns of the return values. Second, we can define the intraday returns over lower frequencies than the actually observed frequency by subsampling the data (e.g. every five minutes) and ignoring the other intraday data in between. Lastly, we can study the intraday returns in continuous time. This work shall regard the latter two techniques.

High-frequency data show some peculiarities compared to the daily time series data. In short, this kind of data is observed for inequally spaced intraday time points, as opposed to daily time series data. Also, the noise in the high-frequency data is more significant than in the daily data (Taylor, 2005), making their research more challenging. Often, the modelling procedure of this dataset needs to be defined over lower frequencies. Estimating the optimal modelling frequency requires effort, although frequencies from five to 30 minutes are usually utilised in the literature (Andersen et al. (2001a), Zhang et al. (2005), Aït-Sahalia and Yu (2009) among others).

Besides, we may observe no trades within intraday periods, likely triggering an error for some automatic procedures. In this case, the practitioner needs to define a non-misleading datum for this period so that the analysis process will continue normally. Moreover, the dataset may contain some mistaken records which need to be removed through a data-cleaning procedure. The steps of such a procedure are overviewed in Hautsch (2011). In general, the analysis of high-frequency data is more computationally expensive than for daily time series because of the incomparably larger data size. For example, on average, we have 2,428 observed prices per stock within a day in the dataset we use in this work. Assuming 252 trading days in a year, this data size corresponds to a dataset that refers to a 10-year period in daily time series. This number substantially increases for highly traded stocks since we can observe more than 6,000 observed prices daily, which approximately corresponds to a dataset of 25 years for daily time series. Furthermore, the accessibility to high-frequency data is more restricted than for the daily time series of which the closing stock price is released to the public every day.

However, as Corsi et al. (2000) show, high-frequency data are preferred compared to daily time series in terms of volatility estimation and forecasting evaluation. This fact is also visible in Figure 1.1, where we plot the high-frequency prices of the stock Pfizer Inc. (PFE) against its daily prices for all the trading dates of January 2012. This figure illustrates how more informative high-frequency data are compared to daily data. Also, Taylor (2005) concludes that "*analysis of high-frequency data is rewarding and well worth the additional effort*" and the field is continuously expanding.

Most development in this field has focused on the problem of dimensionality in the framework of discrete time series data analysis. This thesis develops new statistical modelling approaches for high-frequency data from the perspective of continuous function estimation problem.

Chapter 1 is an introductory chapter where we briefly describe how the stock market works in section 1.2. Beyond the manipulation of high-frequency data, the investigation of financial volatility is the secondary axis that this work lies, and thus we discuss some basic concepts of financial volatility in section 1.3. Broadly speaking, considerable fluctuations are observed in the stock markets during periods of financial uncertainty. These fluctuations can cause significant economic losses to investors. Therefore, estimating or predicting the magnitude of these fluctuations is vital for protecting the investors' wealth. A measure of the size of fluctuations is financial volatility, however volatility cannot be represented by an explicit formula. Specialists define volatility according to their needs, making several convenient assumptions each time depending on the particular scope of their interest. In financial markets, the standard deviation is a widely used measure of daily volatility computed from daily returns.

Although its simple form, the standard deviation is a valid measure under the strong assumption of constant variance over the days, which has been proved wrong through time. For example, a sudden market announcement can abruptly increase the fluctuations in the market within a day. This event is not visible in the volatility estimate of

standard deviation which is given as a scalar number and refers to a long data period. Usually, standard deviation is utilised for daily or monthly time series data and similar models have been proposed for high-frequency data. In section 1.4, we introduce some preliminaries, notations and necessary definitions for high-frequency data that will be utilised in the rest of this work. Chapter 1 is completed with section 1.5 where we introduce our dataset. We also conduct an exploratory analysis on our dataset, which will permit us to compare the results to the findings in the later stages of this study.

In Chapter 2, we construct different types of volatility measures exploiting intraday information about the timing of the events or the trading frequency. After a brief discusion of the relevant context, we report two standard estimators of the daily volatility that utilise intraday returns for the estimation. These estimators are considered superior to the standard deviation in the sense that they exploit the intraday information, delivering more accurate daily volatility estimates. Moreover, the volatility estimates can considerably vary from day to day, as opposed to the standard deviation estimator where its approximation defines the volatility of a predetermined period. Therefore, the examination of the volatility patterns becomes easier through these measures.

In section 2.4 we develop five estimators of the intraday volatility, that is, the volatility observed over the intraday intervals of a day, along with their statistical framework. These estimators follow a similar logic to the aforementioned daily volatility measures and we can use them both for intraday and daily volatility approximation. Each of these five measures uses different information for the estimation. Indicatively, such information is the time space between the intraday observations or the number of observations that occurred within the intraday intervals in a day. In the last section of Chapter 2, we provide an example using the defined intraday volatility measures. In this example, the intraday and daily volatility of two stocks are estimated for three days.

In Chapter 3, we investigate the stocks' common volatility patterns as derived by the proposed volatility estimators of Chapter 2. Factor analysis is a popular method for

a multivariate dataset, and it summarises the volatility features from a high-dimension problem to a lower-dimensional scale. Through this technique, we intend to explore whether we can efficiently outline the volatility estimates to a few volatility components. The benefit of factor analysis is that it allows us to investigate the resulting components instead of the whole dataset, facilitating the analysis procedure.

This study uses the factor analysis approach presented by Lam et al. (2012). The latter work takes a time series process of multiple variables as input, assuming that this process does not violate the stationarity conditions. In financial literature, the time series process consists of daily observations; nevertheless in our work, we also test this method using the intraday volatility estimates, as given by the estimators of Chapter 2. Considering the intraday estimates as a time series process for each stock, we show that it violates the stationarity conditions. For this reason, we extend the Lam et al. (2012) approach under the intraday framework. In this framework, we can monitor which stocks highly contribute to the resulting volatility components as well as which intraday intervals drive the volatility patterns the most. Chapter 3 concludes with two empirical applications. In the first application, we apply the factor analysis to the daily volatility estimates, whilst in the second one, the factor analysis is applied to the intraday volatility estimates.

Chapter 4 develops a daily volatility measure using the local polynomial regression method. Local polynomial regression, presented in section 4.2, is a popular statistical technique that estimates the observations' mean function, enabling specialists to model the data patterns continuously. Further, this method allows us to filter out the noise admittedly contained in financial observed stock prices (Zhang et al. (2005), Hansen and Lunde (2006) among others).

Our proposed estimator, as defined in section 4.3, is similar to the measure of Kristensen (2010); however, the latter was developed assuming that the observed data mirror the true data values. In contrast, the theoretical background of our estimator consid-

ers that financial observations are blurred by noise, which has been proved in practice. Section 4.6 conducts a simulation study where the superiority of our estimator against standard volatility estimators is demonstrated for noise levels which follow normal conditions. This chapter ends with an empirical example.

Besides, this work investigates the stocks' dependence network when the data have a functional form. Functional data refer to a dataset where every observation is given in the form of functions instead of a scalar value (Yao et al., 2005; Ramsay and Silverman, 2005). On the one hand, because of the great abundance of information, functional data provide a more comprehensive picture of the population characteristics. However, they require special manipulation because of their complex structural form.

After a brief discussion of some relevant preliminary concepts, Chapter 5 develops the methodology for predicting the (conditional) dependence network in the functional domain, utilising the Gaussian graphical model. The Gaussian graphical model is a way to estimate and visualise the variables which share partial correlation. Qiao et al. (2019) introduces the theoretical and practical context concerning the estimation of the dependence structure of a set of variables in the functional domain with the Gaussian graphical model. Our approach is based on this work but differs in the methodological part. In addition, as far as we know, this is the first study which applies functional graphical models in the financial field and even more to high-frequency data. The simulation study of section 5.4 indicates that our approach works efficiently under the specified settings. Chapter 5 concludes with an empirical application in section 5.5. Appendices A-E correspond to Chapters 1-5, in respective order.

1.2 Microstructure of Financial Markets

In this section, we outline how the stock market works. When a company needs money to satisfy some goals, it can join the stock exchange to find them after completing an institutionalised procedure. By entering the stock exchange, a company makes

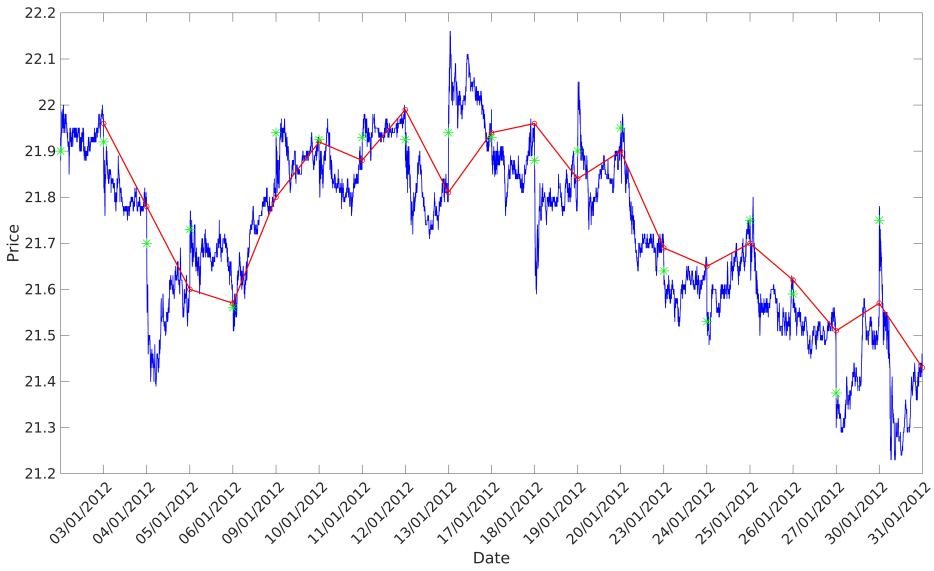

Figure 1.1: The observed high-frequency stock prices (blue line) of Pfizer Inc. (PFE) are plotted against the closing prices (red line) of the same company for the trading days of January 2012. The observed open prices are also given as green asterisks in the figure.

available to the public through the stock market a number of shares of the underlying company's stock at a predetermined price. In this way, the company asks for capital admission so that it will be able to broaden, improve and enrich its services.

On the other hand, when the public -more formally, the investors- believe that this capital admission will lead to the growth of this company, they become shareholders of the underlying company, expecting to receive several gains from this growth. Such gains can be the annual bonuses the companies often share with their shareholders from their earnings (also known as dividends) or the shareholders can profit from an increase in the stock price of the underlying company.

The price of a stock is determined by the financial markets. Financial markets rely on the concept of supply and demand. This concept shortly describes the situation where a person A is willing to sell to a person B and, simultaneously, a person B is willing to buy from a person A an asset. If the amount the potential buyer is willing to pay for this asset meets or exceeds the potential seller's requirements, then the transaction becomes

real. On the other hand, if the amount that the potential buyer is willing to pay for this asset does not reach the lowest amount the potential seller is expecting to receive by selling the asset, then either at least one has to revise his plan or the transaction is cancelled. This is a descriptive explanation of what happens in financial markets with the difference that there are more than one potential buyers, sellers, assets and shares. There are different types of assets, such as stocks, bonds, currencies, but this work with the term assets refers exclusively to stocks.

Every market has its own rules; however, the trading process follows a similar rationale across the markets. In particular, all the orders from potential buyers and sellers are collected in the market's electronic system, called the order book. In the sequel, the buyers' and the sellers' orders (also called the bid and ask prices, respectively) are sorted into preferable (for the investors on the other side of the market) order. Then, the transaction between the two counterparties takes place when the ask and the bid prices match up. This price dominates the market, consequently defining the current stock price. The reader can refer to the second chapter of Hautsch (2011) for more details about the microstructure of the financial markets.

New York Stock Exchange (NYSE) is the stock exchange that we use data from in the empirical applications of this work. NYSE is a continuous auction market where the matchup between trades is done immediately. Also, it is one of the world's most liquid markets and is active from 9:30 to 16:00 (Eastern Time Zone); six and a half hours in total. In NYSE, investors can place (but not execute) their orders before the market's open time. Upon the beginning of the trading session, these orders are offset, defining the open price of the asset. From that time on, regular orders arrive in the market as described above.

1.3 Financial Volatility

The exploration of financial volatility is of central interest in this study. Loosely speaking, volatility refers to the overall magnitude of the price change over a period. Volatility is directly linked to financial risk. As aptly Poon (2005) points out, risk refers to a situation around the possibility of an investor to lose money through a period of financial uncertainty. A reason why volatility is mainly linked to an adverse market shock compared to a positive one is the so-called leverage effect. The leverage effect refers to the widely observed event where market volatility is higher during a downward than an upward market trend. The reader can refer to Black (1976) and Christie (1982) for more information on this topic. Thus, volatility analysis plays a key role in investors' strategy to hedge or modify their portfolios in such a way to minimise risk.

Volatility is a random quantity that can be distinguished differently depending on the question one needs to address. Taylor (2005) summarises the main ways to interpret volatility. In particular, a popular way to interpret volatility is by historical volatility, which refers to the volatility approximated by observed data. Variance and standard deviation are broadly used tools for the historical volatility estimation. The variance of a random variable Y is defined by:

$$Var[Y] = \mathbb{E}[(Y - \mathbb{E}[Y])^2].$$

For a finite sample of daily stock log-returns R_d, defined by:

$$R_d = log(P_d) - log(P_{d-1}) \tag{1.1}$$

where P_d indicates the closing price of the d day, then an empirical estimator of the daily variance is given by:

$$\widehat{Var}_N(R) = \frac{1}{N-2}\sum_{d=2}^{N}(R_d - \overline{R}_d)^2 \tag{1.2}$$

where $\overline{R}_d$ denotes the mean value of the observed log-returns R_d for $d = 2, ..., N$ and N refers to the number of working days we consider in the model. The reduction by one unit in the denominator of Equation (1.2) stems from the degrees of freedom of the estimator. The degrees of freedom (i.e. the maximum number of values that can vary) are defined as the difference between the number of observations and the parameters that need to be estimated. In our case, the only parameter that needs to be approximated is the population mean $\mathbb{E}[X]$, resulting in $N - 2$ degrees of freedom.

The daily standard deviation is defined as the square root of the daily variance. Hence, an estimator of the standard deviation is given by taking the square root of the empirical variance. Moreover, the monthly and the annualised standard deviations are estimated as a proportion of the daily standard deviation estimate:

$$\begin{aligned} \widehat{\sigma}_{StDev}^{(Daily)}(R) &= \sqrt{\widehat{Var}(R)} \\ \widehat{\sigma}_{StDev}^{(Annual)}(R) &= \sqrt{252\widehat{Var}(R)} \\ \widehat{\sigma}_{StDev}^{(Monthly)}(R) &= \sqrt{\frac{252}{12}\widehat{Var}(R)} \end{aligned} \tag{1.3}$$

since there are 12 months and, in general, 252 trading days in a year.

However, as Poon (2005) states, the standard deviation is preferred from variance for the following reasons. The standard deviation is chosen regarding volatility estimation and forecasting since "*variance is much less stable*". Also, standard deviation has a natural representation of the unit. For example, the standard deviation of the returns is expressed in dollars '$', as opposed to variance, which should be expressed in '^{2}', giving rise to a bizarre representation. A drawback of using standard deviation to estimate volatility is that this is a valid estimator under the assumption of fixed variance (i.e. homoscedasticity) over the understudied period. Nevertheless, in the real world, the returns' variability may vary considerably over time, leading to an erroneous inference.

A second popular notion of volatility is conditional volatility. This term refers to the

volatility of a future stock price change, given the past information about price changes. A popular model under the framework of time-varying conditional volatility estimation is the ARCH model, developed by Engle (1982). This model considers the conditional variance as a function of a number of the preceding error terms of returns. A model that permits "*longer memory and a more flexible lag structure*" (Bollerslev, 1986) is the GARCH model. In this model, the conditional variance further considers the dependence in the preceding conditional variances to be formed. Hence, conditional volatility covers the case where the variance can change over time (i.e. heteroscedasticity), as opposed to historical volatility.

Implied volatility is based on the concept of option pricing in the framework of financial derivatives. Options are financial instruments widely used in the market, and the reader can find more about this topic in Hull (2012). Implied volatility is defined as the current volatility value that will give the option derivative's theoretical price when placed as an input in the option pricing model. It relies on the market volatility expectation for a predetermined future period, as opposed to historical volatility which exclusively depends on the observed price evolution of the past. The most famous option pricing model is the Black-Scholes model developed by Black and Scholes (1973).

Lastly, some studies are interested in understanding how volatility behaves through time. Then, the concept of stochastic volatility arises. In this area, the Itô process is a popular process for modelling the price evolution of a stock using a time-varying stochastic volatility term. The Black-Scholes model assumes constant volatility in the asset price changes throughout the option's life, which may sometimes be a strong assumption. One can overcome this situation by considering stochastic volatility models instead. Heston is such a model, developed by Heston (1993).

Overall, historical volatility is the most popular concept out of the ones mentioned above since it is also utilised for comparison purposes. For example, implied volatility is a more prevalent measure for derivative traders since it provides a forward-looking

estimation (as opposed to historical volatility). Nevertheless, experts often compare the implied volatility value to the historical one to make investment decisions on derivative markets.

This study focuses on the concept of historical volatility under the high-frequency sampling context. Realised variance is a measure that utilises intraday high-frequency data to approximate the notional daily volatility. A drawback of the conventional daily volatility estimators, like the standard deviation in Equation (1.3) is that they deliver a fixed estimate over the examined period. In contrast, realised variance gives a daily estimate over each day, providing more accurate daily volatility estimates and enabling us to model the volatility patterns of a stock. The concept of realised variance is explicitly given in the next chapter with the pertinent theoretical background.

1.4 High-Frequency Financial Data

When two counterparties agree to make a transaction for a specified asset, at a specified price, for a specified quantity, then all of these pieces of information are recorded on an intraday basis, offering high-frequency financial data. Engle (2000) refers to this kind of dataset -which contains all of the intraday trading information of an asset- as "*ultra-high-frequency data*". Usually in relevant literature, studies use intraday frequencies of a lower scale, such as frequencies of five, ten, or 30 minutes (Andersen et al. (2001a), Zhang et al. (2005), Aït-Sahalia and Yu (2009) among others). In this study, by the term 'high-frequency data' we refer to the full records of the intraday stock information where any modification on the frequency scale will be explicitly specified.

It is more convenient for specialists to analyse returns than prices (Tsay, 2005). In addition, returns are a direct measure of the price change and, thus, of volatility, demonstrating the situation prevails in the market. High-frequency observations occur several special characteristics compared to daily observations. In particular, high-frequency returns yield heavier tails than daily returns. As Tsay (2005) shows, the non-synchronous

trading, as observed in continuous markets, is responsible for the introduction of negative autocorrelation, which introduces some noise in the observed data. The noise incorporated in high-frequency data is higher than for common daily data (Taylor, 2005). Furthermore, the intraday prices react to economic news promptly, increasing stock volatility in a short time. The highest variability in stock prices is usually observed during the beginning of the trading session, followed by a downward trend by lunchtime, and the variability increases again before the end of the trading session (Andersen and Bollerslev (1997), Taylor (2005), Li et al. (2015) among others).

1.4.1 Preliminaries

At this point, it would be useful to establish the notational and theoretical framework around high-frequency data. Also, the observed values will be considered as a function of time for the rest of this work.

In general, financial markets are active for six and a half hours in a day, translated as 23,400 trading seconds. When a stock transaction occurs in the market information like the transaction price, the transaction volume, the number of transactions and the specific timestamp are monitored. A timestamp refers to the time point at which a stock transaction is monitored in the market and expressed in seconds. Transaction prices define the observed prices in a day; consequently, the observed timestamps correspond to the observation times of the observed prices on this day.

In literature, statisticians rescale the observation times to values between zero and one by dividing these time points by 23,400. In this way, they align the scale of observed ticks to the relevant theory, as explained later in this work.

Let $P(t_i)$ be a sequence of the actual intraday prices of a stock observed at time points:

- $t_i = i/n$ (i.e. equally spaced time points), or
- $0 < t_1 < ... < t_i < ... < t_n \leq T = 1$ (i.e. unequally spaced time points),

where t_i denotes the intraday time point when the i-th transaction takes place for $i = 1, ..., n$. Here, n indicates the number of observed prices in a trading day.

In literature, it is commonplace for researchers to consider the stock prices in their logarithmic form. The reason they make this modification is linked to the concept of log-returns, which is explained in the next subsection. Let $Y(t_i)$ be the observed log-price of a stock at time point t_i, given by:

$$Y(t_i) = \log\Big(P(t_i)\Big)$$

where $P(t_i)$ denotes the observed actual price of a stock at t_i. In an ideal world, $Y(t_i)$ reflects the 'true' (or effective) log-price $X(t_i)$ of this stock:

$$Y(t_i) = X(t_i). \tag{1.4}$$

Nevertheless, Equation (1.4) rarely holds because of the existence of microstructure noise in the observed log-prices. Under the existence of noise, the observed log-price at a given time point $Y(t_i)$ can be represented as a function of the true log-price and an error term:

$$Y(t_i) = X(t_i) + E(t_i)$$

where $X(t_i)$ depicts the i-th true log-price -which is a latent variable- and $E(t_i)$ shows the corresponding error term.

The noise comes from several microstructure effects that dominate the market. Corsi (2005) refers to the main sources of the microstructure noise:

- **Non-Synchronous Trading**: we compute the stock returns assuming that they refer to periods of equal length, which may be a strong assumption on high-frequency scale;
- **Asymmetric Information**: All investors do not share the same information

about companies. Investors who acquire exclusive companies' information can take advantage compared to other investors;

- **Bid-Ask Bounce**: The observed stock price commute between the same values, being stable in fact;
- **Price Discriteness**: We represent the stock price using a limited number of decimals;
- **Rounding Error**: Some noise is introduced into the stock prices by rounding the observed prices;
- **Infrequent Trading**: Prices of illiquid stocks often diverge from their true price;
- **Price Jumps**: Abrupt price changes can introduce some error in stock prices.

The reasons behind the introduction of noise are extensively studied in Madhavan (2000).

1.4.2 Returns

Volatility is associated with the concept of stock returns. A stock return refers to the rate of change in the price of a stock for a predetermined period. The intraday pointwise log-returns are defined as:

$$R(t_i) = \log\Big(P(t_i)\Big) - \log\Big(P(t_{i-1})\Big) = Y(t_i) - Y(t_{i-1}) \tag{1.5}$$

for $i = 2, ..., n$ and n denotes the number of the intraday observed prices in a day. In the rest of this work, we refer to log-returns by the general term 'returns' for conciseness.

A cursory look at the returns may overlook the fact that a transaction can happen at any moment during trading hours. In Figure 1.4.2, a visual illustration of the observed prices of a stock for two consecutive days is provided. As we observe, the sampling frequency of the transactions can change at different intervals of the day. This can

happen because of several reasons; for example, economic news arrives in the market, increasing the market transactions instantly. Therefore, comparing the intraday returns can lead us to a confusing inference since they refer to unequal intraday periods. Hence, it seems useful for statisticians to convert the corresponding returns to a comparable time scale.

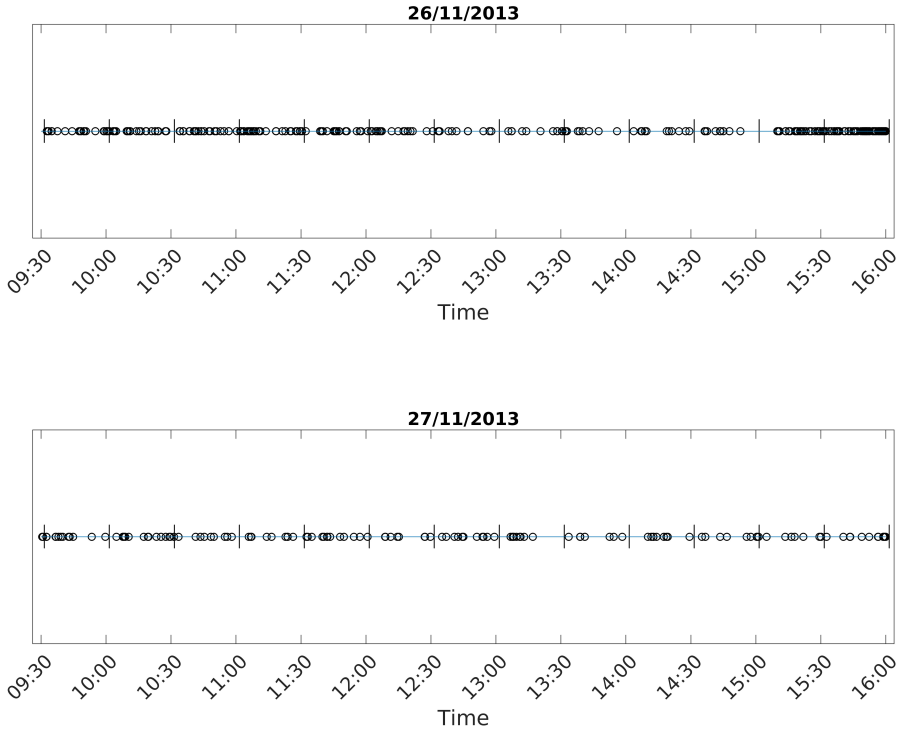

Figure 1.2: The observation times of Mastercard Inc. (MA) for two consecutive dates, 26 and 27 of November 2013.

We can distinguish high-frequency returns to the following two different cases:

- **Returns of periods of different length**: refers to the case where prices are observed at different rates within a day. The observed time points are given at $0 < t_1 < ... < t_n \leq T = 1$. In order to convert the intraday returns for periods of

different length to a comparable scale, one can compute:

$$\widetilde{R}(t_i) = \Big(Y(t_i) - Y(t_{i-1})\Big) z_{t_i} \tag{1.6}$$

where $Y(t_i)$ denotes the log-price observed at time point t_i and $z_{t_i} = \Delta\tau/\Delta t_i$. Here, $\Delta\tau = 1/n$ denotes the time space supposing equilength observation times and $\Delta t_i = t_i - t_{i-1}$ denotes the time difference between the i-th observed prices and its preceding one. The following example highlights the weakness of converting and analysing returns using the above technique. Suppose a stock has a transaction approximately every five minutes within a day. As described at the beginning of section 1.4.1, this means $\Delta t_i \approx 0.026$. If suddenly this stock realises two transactions placed one second away, then $\Delta t_i \approx 0.000043$, while $\Delta\tau$ remains constant. Placing these two hypothetical Δt_i values into Equation (1.6) would give rise to two returns of completely different size, affecting the analysis of the intraday return patterns for this stock.

- **Returns of periods of same length**: refers to the case of returns of equally spaced observations, given at $t_i = i/n$ where $i = 1, ..., n$. In this case, we have that:

$$\widetilde{R}(t_i) = R(t_i) = Y(t_i) - Y(t_{i-1}).$$

Researchers sometimes assume intraday returns of equidistant observations in their studies, however, this is an unrealistic scenario in continuous markets. An alternative way to handle the issue arising from this event is to study the returns data patterns in continuous time, which is the approach of this work.

1.5 Dataset Description and Exploratoty Data Analysis

This section introduces the dataset of this study. We conduct an exploratory analysis of the dataset, and the results are compared against later findings in this work. The dataset refers to the high-frequency stock data of 50 underlying companies for the period between 3 January 2012 and 31 December 2014; that is, 754 trading days. The dates are also presented in the format 'DD/MM/YYYY' in this work. The selected companies were considered among the most prominent stocks an investor could hold in these three years, and they are related to different sectors (see Table 1.1). The companies' names, identifier codes, corresponding sectors and related industries are listed in Table 1.2. We extracted the information of these companies from the website finance.yahoo.com (nd). All of the selected companies traded at NYSE, where, generally speaking, 2012 and 2013 are considered good years for the market, whereas 2014 was a bad one.

Sectors:
Basic Materials
Communication Services
Consumer Cyclical
Consumer Defensive
Energy
Financial Services
Healthcare
Industrials
Technology

Table 1.1: Company's sectors.

In the context of high-frequency data the term 'tick data' is broadly used in order to describe a dataset of information recorded when we observe a trade execution in the trading session. Tick data can signify different types of information, such as stock (trade data) and quote (quote data) prices. This study works with trade data which were kindly granted from the Lancaster University Management School's database by Prof. Nolte and Dr Nolte. Our trade data consist of the following intraday observations:

Asset Name	Identifier Code	Sector	Industry
3M Co.	MMM	Industrials	Conglomerates
AT&T Inc.	T	Communication Services	Telecom Services
Alcoa Inc.	AA	Basic Materials	Aluminium
American Express Co.	AXP	Financial Services	Credit Services
American International Group	AIG	Financial Services	InsuranceDiversified
Bank of America	BAC	Financial Services	Banks-Diversified
Boeing Co.	BA	Industrials	Aerospace & Defense
Caterpillar Inc.	CAT	Industrials	Farm & Heavy Construction Machinery
Chesapeake Energy Corp.	CHK	Energy	Oil & Gas E&P
Chevron Corp.	CVX	Energy	Oil & Gas Integrated
Citigroup Inc.	C	Financial Services	Banks-Diversified
Cleveland-Cliffs Inc.	CLF	Basic Materials	Steel
Coca-Cola Co.	KO	Consumer Defensive	Beverages-Non-Alcoholic
Corning Inc.	GLW	Technology	Electronic Components
Delta Air Lines Inc.	DAL	Industrials	Airlines
Devon Energy Corp.	DVN	Energy	Oil & Gas E&P
E.I. DuPont de Nemours & Co.	DD	Basic Materials	Specialty Chemicals
Eli Lilly and Co.	LLY	Healthcare	Drug Manufacturers-General
Exxon Mobil Corp.	XOM	Energy	Oil & Gas Integrated
FedEx Corp.	FDX	Industrials	Integrated Freight & Logistics
Ford Motor Co.	F	Consumer Cyclical	Auto Manufacturers
Freeport-McMoRan Inc.	FCX	Basic Materials	Copper
General Electric Co.	GE	Industrials	Specialty Industrial Machinery
General Mills Inc.	GIS	Consumer Defensive	Packaged Foods
General Motors Corp.	GM	Consumer Cyclical	Auto Manufacturers
Halliburton Co.	HAL	Energy	Oil & Gas Equipment & Services
Hewlett-Packard Co.	HPQ	Technology	Computer Hardware
Home Depot Inc.	HD	Consumer Cyclical	Home Improvement Retail
International Business Machines Corp.	IBM	Technology	Information Technology Services
JPMorgan Chase & Co.	JPM	Financial Services	Banks-Diversified
Johnson & Johnson	JNJ	Healthcare	Drug Manufacturers-General
Lockheed Martin Corp.	LMT	Industrials	Aerospace & Defense
Lowes Companies Inc.	LOW	Consumer Cyclical	Home Improvement Retail
Mastercard Inc.	MA	Financial Services	Credit Services
McDonald's Corp.	MCD	Consumer Cyclical	Restaurants
Merck & Co. Inc.	MRK	Healthcare	Drug Manufacturers-General
National Oilwell Varco Inc.	NOV	Energy	Oil & Gas Equipment & Services
Newmont Goldcorp Corp.	NEM	Basic Materials	Gold
Nike Inc.	NKE	Consumer Cyclical	Footwear & Accessories
Pfizer Inc.	PFE	Healthcare	Drug Manufacturers-General
Procter & Gamble Co.	PG	Consumer Defensive	Household & Personal Products
Schlumberger Ltd.	SLB	Energy	Oil & Gas Equipment & Services
Target Corp.	TGT	Consumer Defensive	Discount Stores
US Bancorp	USB	Financial Services	Banks-Regional
United Technologies Corp.	UTX	Industrials	Aerospace & Defense
Verizon Communications Inc.	VZ	Communication Services	Telecom Services
Visa Inc.	V	Financial Services	Credit Services
Wal-Mart Stores Inc.	WMT	Consumer Defensive	Discount Stores
Walt Disney Co.	DIS	Communication Services	Entertainment
Wells Fargo & Co.	WFC	Financial Services	Banks-Diversified

Table 1.2: The name of the companies, their ticker, their sector and the industry they are related to. Businesses information was taken from finance.yahoo.com (nd).

- the timestamp at which a transaction took place during the trading session (given in seconds),
- the number of transactions that happened at the particular time points,
- the stock price as formed by the corresponding transactions,
- the number of shares traded at the given time point (also known as transaction volume),

however, this study exclusively utilises the transaction timestamps and the corresponding stock prices in the investigation procedures, whilst both information will be transformed appropriately. In particular, the timestamps are transformed into values between zero and one in the analysis procedures. The stock prices are taken in their logarithmic form and are usually used to calculate the returns.

In the financial literature, researchers prefer to work with stock returns instead of prices. This happens because the return series presents several convenient properties (Tsay, 2005). In particular, stock returns are believed to follow a weakly stationary process (Tsay, 2005), meaning that the mean, the variance and the autocorrelation structure of the process remain constant over time without showing any periodic pattern. These properties are not the case for stock prices. So return series perform stable patterns through time, making the explorational analysis less challenging than price series.

In past times, a broadly used assumption in mathematical finance was that the daily returns, as given by Equation (1.1), are Independent and Identically Distributed (IID) where their mean value and their variance are constants (Tsay, 2005). In addition, one can view the daily returns as a sum of all the intraday returns of a similar nature (IID). Then assuming that the number of transactions is sufficiently large, by virtue of the central limit theorem, the normality assumption was considered to be reasonable for the daily returns (Fama, 1963). Kendall and Hill (1953) and Osborne (1959) are among the first who observed that the bell-shaped distribution yielded from the returns

approximates the normal distribution. Nevertheless, numerous studies have shown that the empirical distribution of the returns presents fatter tails than the normal distribution (Officer (1972), Mota (2012) among others). Especially, a leptokurtotic distribution arises, which means that the kurtosis value of this distribution is greater than three (Kendall and Hill, 1953; Franke et al., 2019; Tsay, 2005).

One can directly gain intuition about the overall stock behaviour by looking at the descriptive statistics of the stocks' daily returns. The descriptive statistics of the chosen stocks are given in Table 1.3, and a visual representation is offered in Figure A.1 in appendix A. Before proceeding with the calculation of the descriptive statistics, a pre-processing of the data for some particular stocks has been applied to treat anomalies related to stock splits.

More specifically, on 22 of January 2014 the stock price of Mastercard Inc. (MA) dropped from \$818.97 (closing price on 21/01/2014) to \$82.35 (open price on 22/01/2014). This happened due to the stock split realised in that period. Stock split refers to the situation where a shareholder receives a number of shares of an underlying stock for every share he owns, decreasing the stock price proportionally. In particular, every shareholder of MA on 22 of January 2014 received nine additional shares for every individual one, making the stock price drop ten times. Three additional companies that conducted a stock split were Coca-Cola Co. (KO) on the 13 of August 2012, Nike (NKE) on the 26 of December 2012 and National Oilwell Varco Inc. (NOV) on the 2 of June 2014. For the NOV, every shareholder received 1109 shares for every 1000 shares he owned, decreasing the stock price by about 10%, while for KO and NKE the shareholders received two shares for every share they had in their possession, making the stock price drop by 50%.

A company conducts a stock split because this technique makes the stock price more affordable to the public, attracting new investors. Ikenberry et al. (1996) show that the companies that have experienced a stock split have high returns in a three-year horizon. Although the market worth of a company remains the same after a stock split,

this technique ruins the behavioural patterns of the stock data, generating misleading statistic values. For this reason, we rescale the intraday stock prices after the stock split to previous level values by multiplying these prices by 2, 10, 1.109 and 2 for KO, MA, NOV and NKE, respectively. Now, we can securely analyse the data patterns without removing any stock from the dataset. Table 1.3 lists the descriptive statistics of the daily returns after the rescaling procedure for the stocks that have experienced stock split.

According to the data features marked above, we expect that the distribution of the daily returns of the stocks will approach the form of the normal distribution but with heavier tails. This happens for a skewness value near zero and a kurtosis value higher than three. Regarding kurtosis, the returns distribution presents heavy tails for all companies in Table 1.3, consistent with the studies cited above. According to Ruppert and Matteson (2011), the skewness of stock returns is expected to be approximately zero, which is the case for most stocks in this table. In addition, we observe a slight left-skewed performance for most stocks, given as a negative skewness values. This event is more visible for International Business Machines Corp. (IBM) and Freeport-McMoRan Inc. (FCX). As Fan and Yao (2017) notices, this comes from the leverage effect. Table 1.3 shows that most companies have location parameters (mean, median and mode) close to zero. This is consistent with the empirical results of other studies (for example, Larson (1960), Tsay (2005)). Further, we do not observe any distinct performance regarding stocks' variability, where Cleveland-Cliffs Inc. (CLF) is the stock with the highest variability.

Taylor (2005) concentrates the characteristics of the empirical distribution of the high-frequency returns from a few studies. More specifically, he reports that the empirical distribution of 10-minute returns has location parameters close to zero. Besides, he cites the work of other studies which show a kurtosis value between 16 and 38 for returns of different frequencies and a skewness value around zero based on the analysis of market indices. In Table 1.4, we list the descriptive statistics of our dataset for

	Mean	Median	Mode	Variance	St. Deviation	Max	Min	Skewness	Kurtosis
MMM	0.0009	0.0009	0	0.0001	0.0091	0.0427	-0.0439	-0.2947	5.817
T	0.0001	0.0009	0	0.0001	0.0093	0.0369	-0.0509	-0.5357	5.9243
AA	0.0007	0	0	0.0003	0.0173	0.0858	-0.0742	0.3269	4.9863
AXP	0.0009	0.0007	0	0.0001	0.0118	0.0503	-0.044	-0.0942	3.9319
AIG	0.0011	0.0015	0	0.0002	0.0154	0.0576	-0.0741	-0.2959	5.6376
BAC	0.0015	0.0012	0	0.0003	0.0183	0.0824	-0.0726	0.2395	4.7375
BA	0.0007	0.001	-0.0527	0.0002	0.0125	0.0529	-0.0527	-0.0848	4.7265
CAT	0	0	0	0.0002	0.0136	0.0571	-0.0616	-0.0974	4.9883
CHK	-0.0002	0	0	0.0006	0.0254	0.1574	-0.1567	-0.5106	9.5898
CVX	0	0.0007	-0.0554	0.0001	0.0104	0.0412	-0.0554	-0.4664	5.246
C	0.0009	0.0006	0	0.0003	0.0171	0.0616	-0.0855	-0.1217	4.681
CLF	-0.003	-0.0038	0	0.0013	0.0364	0.2025	-0.2244	-0.3547	9.3234
KO	0.0002	0.0001	0	0.0001	0.0092	0.0556	-0.0614	-0.1866	8.1313
GLW	0.0008	0.0011	0	0.0003	0.0162	0.1322	-0.1115	-0.341	16.0921
DAL	0.0024	0.0024	0	0.0005	0.0229	0.0986	-0.1228	-0.3113	5.1815
DVN	-0.0001	-0.0005	0	0.0003	0.0163	0.0951	-0.0822	0.0661	6.822
DD	0.0006	0.0006	0	0.0001	0.011	0.0511	-0.0951	-0.64	11.1409
LLY	0.0007	0.0009	-0.0426	0.0001	0.0109	0.0528	-0.0426	0.099	5.5475
XOM	0.0001	-0.0002	0	0.0001	0.0094	0.0339	-0.043	-0.3107	5.077
FDX	0.0009	0.0005	0	0.0002	0.013	0.06	-0.0711	-0.0218	5.6849
F	0.0004	0.0008	0	0.0002	0.0149	0.0749	-0.077	-0.3449	5.4047
FCX	-0.0007	-0.0003	0	0.0004	0.0193	0.0819	-0.1742	-1.1139	12.749
GE	0.0004	0.0002	0	0.0001	0.0107	0.0447	-0.0423	0.0622	4.5791
GIS	0.0004	0.001	0	0.0001	0.0084	0.0305	-0.0451	-0.7572	6.1595
GM	0.0007	-0.0003	0	0.0003	0.0173	0.0912	-0.0611	0.4722	5.6335
HAL	0.0002	0.0001	0	0.0003	0.018	0.0594	-0.1157	-0.8439	7.6444
HPQ	0.0005	0.0006	0	0.0004	0.021	0.157	-0.1397	-0.1955	15.2078
HD	0.0012	0.0009	0	0.0001	0.0111	0.0561	-0.0365	0.1651	4.9357
IBM	-0.0002	-0.0002	-0.0877	0.0001	0.0109	0.0437	-0.0877	-1.2606	13.6061
JPM	0.0008	0.001	0	0.0002	0.014	0.068	-0.0973	-0.3635	7.5392
JNJ	0.0006	0.0005	0	0.0001	0.0078	0.0255	-0.0268	-0.2039	4.3181
LMT	0.0011	0.0012	-0.0401	0.0001	0.0101	0.0372	-0.0401	-0.1483	4.2148
LOW	0.0013	0.0011	0	0.0002	0.0144	0.0615	-0.1068	-0.5033	9.0374
MA	0.0011	0.0014	-0.0512	0.0002	0.0136	0.0893	-0.0512	0.304	6.5651
MCD	-0.0001	0.0005	0	0.0001	0.0082	0.0369	-0.0459	-0.5147	6.0954
MRK	0.0005	0.0005	-0.0023	0.0001	0.0108	0.0627	-0.0446	0.2633	5.992
NOV	0	0.0006	0	0.0002	0.0154	0.0848	-0.0765	0.069	5.6405
NEM	-0.0016	-0.0015	0	0.0005	0.0223	0.0837	-0.1104	-0.113	5.0058
NKE	0.0009	0.001	-0.0119	0.0002	0.0135	0.1151	-0.0987	0.8586	18.6612
PFE	0.0005	0	0	0.0001	0.0095	0.0411	-0.0452	-0.0896	4.6844
PG	0.0004	0.0002	0	0.0001	0.0087	0.0396	-0.0679	-0.3245	9.9714
SLB	0.0003	0	0	0.0002	0.0147	0.0622	-0.0765	-0.121	4.8079
TGT	0.0005	0.0005	0	0.0001	0.011	0.0713	-0.0452	0.3832	7.9048
USB	0.0006	0.0008	0	0.0001	0.0096	0.044	-0.0504	-0.1894	4.8093
UTX	0.0006	0.0006	-0.0343	0.0001	0.0106	0.0401	-0.0343	0.0671	3.6356
VZ	0.0002	0.0005	0	0.0001	0.0098	0.0341	-0.0418	-0.1077	4.3121
V	0.0012	0.0011	-0.0729	0.0002	0.0131	0.0976	-0.0729	0.3238	8.7434
WMT	0.0005	0.0003	0	0.0001	0.0088	0.0463	-0.0479	-0.2476	7.1689
DIS	0.0012	0.0014	0	0.0001	0.0112	0.0514	-0.0617	-0.2572	4.9861
WFC	0.0009	0.0007	0	0.0001	0.011	0.0558	-0.0614	-0.1048	5.6001

Table 1.3: Descriptive statistics of the daily returns for the 50 stocks of our dataset.

10-minute returns. The return of an intraday period is defined later in Equation (2.7). In Table 1.4, we observe that the location parameters of all stocks are almost zero (up to the first three decimals), the skewness value is close to zero for most stocks and the kurtosis values do not differ substantially from those documented in Taylor (2005).

There are three shorter trading sessions in NYSE annually. Namely, these sessions are the session before U.S. Independence Day, the session after Thanksgiving Day and the session on Christmas Eve. On these days, the market opens at 09:30 and closes at 13:00 local time. Given that one of the aspects that this work investigates is the intraday volatility patterns through the days, we discard half trading days from our dataset for the intraday analysis. In this way, we avoid the effect of days of different lengths in the empirical applications of the following chapters. This approach is also adopted in Ding et al. (2022).

In terms of the estimation procedures, very rarely do not exist enough observations within the intraday intervals for the computation of the intraday returns. Then, we set as return for such an interval the return of its preceding interval in order to facilitate the automatic estimation process. Also, if there are not enough observations within the first intraday interval of a day, the return for this interval is defined as zero.

1.5.1 Correlation

The correlation coefficient (also known as Pearson's correlation coefficient) is a measure of the linear dependence between two variables. The correlation coefficient shows the degree to which two variables move together in the market and it takes values between -1 and 1. The closer the correlation coefficient to -1, the higher the indication that a pair of variables move oppositely to one another. Similarly, the closer the correlation coefficient to 1, the higher the signal that a pair of variables move identically in the market. Furthermore, when the correlation coefficient is equal to 0, this signifies that the two variables do not share linear dependence, meaning that we cannot export

	Mean	Median	Mode	Variance	St. Deviation	Max	Min	Skewness	Kurtosis
MMM	0	0	0	1E-06	0.0012	0.0213	-0.0114	0.3692	13.3085
T	0	0	0	2E-06	0.0013	0.0161	-0.0163	-0.2975	14.4566
AA	0	0	0	5E-06	0.0023	0.021	-0.0268	0.016	8.6687
AXP	0	0	0	2E-06	0.0015	0.0175	-0.0152	0.1677	10.3501
AIG	0	0	0	5E-06	0.0022	0.0218	-0.0265	0.0122	10.3936
BAC	0	0	0	5E-06	0.0023	0.0211	-0.0286	0.037	10.5492
BA	0	0	0	3E-06	0.0017	0.0217	-0.0479	-1.0819	42.9939
CAT	0	0	0	3E-06	0.0018	0.0154	-0.0222	-0.1546	10.3219
CHK	0	0	0	1E-05	0.0032	0.0318	-0.061	-0.4743	16.9599
CVX	0	0	0	2E-06	0.0014	0.0146	-0.0181	-0.0936	9.5448
C	0	0	0	5E-06	0.0022	0.0209	-0.0288	0.0862	10.5412
CLF	-0.0001	0	0	2.2E-05	0.0047	0.0701	-0.0739	0.3052	19.3905
KO	0	0	0	2E-06	0.0012	0.013	-0.0152	-0.1699	11.4563
GLW	0	0	0	4E-06	0.0021	0.0264	-0.0599	-1.2198	39.6196
DAL	0	0	0	1.2E-05	0.0034	0.052	-0.0821	-0.0347	24.5613
DVN	0	0	0	5E-06	0.0022	0.0377	-0.0189	0.2924	13.6606
DD	0	0	0	2E-06	0.0015	0.0342	-0.0149	0.6638	23.3097
LLY	0	0	0	2E-06	0.0015	0.0168	-0.0193	0.14	13.2699
XOM	0	0	0	2E-06	0.0013	0.0164	-0.0127	0.1198	8.7044
FDX	0	0	0	3E-06	0.0017	0.0238	-0.0174	0.3851	12.1
F	0	0	0	4E-06	0.0021	0.0169	-0.039	-0.5787	16.9481
FCX	0	0	0	6E-06	0.0025	0.0372	-0.0219	0.0958	10.2483
GE	0	0	0	2E-06	0.0014	0.013	-0.0205	-0.1302	9.8567
GIS	0	0	0	1E-06	0.0012	0.015	-0.0155	-0.047	13.9338
GM	0	0	0	6E-06	0.0024	0.0606	-0.0305	0.6021	27.1196
HAL	0	0	0	6E-06	0.0025	0.0516	-0.0349	0.2643	18.0685
HPQ	0	0	0	6E-06	0.0025	0.0885	-0.0322	2.004	74.006
HD	0	0	0	2E-06	0.0015	0.0151	-0.032	-0.2371	16.5212
IBM	0	0	0	2E-06	0.0013	0.0199	-0.0131	0.3636	11.7234
JPM	0	0	0	4E-06	0.0019	0.0197	-0.03	0.09	13.5463
JNJ	0	0	0	1E-06	0.0011	0.0237	-0.0241	-0.0703	24.2609
LMT	0	0	0	2E-06	0.0014	0.0167	-0.0263	-0.0578	16.5968
LOW	0	0	0	4E-06	0.0019	0.0183	-0.0222	0.1603	11.6814
MA	0	0	0	3E-06	0.0018	0.0311	-0.0355	0.3489	24.1874
MCD	0	0	0	1E-06	0.0011	0.0256	-0.0134	0.5614	22.9194
MRK	0	0	0	2E-06	0.0015	0.0195	-0.0166	0.1733	12.6049
NOV	0	0	0	5E-06	0.0022	0.019	-0.0328	-0.2087	11.3351
NEM	0	0	0	8E-06	0.0028	0.0328	-0.0265	0.1014	8.4859
NKE	0	0	0	3E-06	0.0016	0.0152	-0.0213	-0.1254	10.9261
PFE	0	0	0	2E-06	0.0015	0.0161	-0.0136	0.1296	9.2228
PG	0	0	0	1E-06	0.0012	0.019	-0.0168	0.2504	15.158
SLB	0	0	0	4E-06	0.002	0.0247	-0.0234	0.0271	10.223
TGT	0	0	0	2E-06	0.0015	0.0166	-0.0153	0.0505	11.5442
USB	0	0	0	2E-06	0.0015	0.0146	-0.0189	-0.0427	10.4204
UTX	0	0	0	2E-06	0.0015	0.0199	-0.0239	-0.0301	15.7468
VZ	0	0	0	2E-06	0.0014	0.0115	-0.0211	-0.53	12.9925
V	0	0	0	3E-06	0.0017	0.0249	-0.0288	0.1049	25.7592
WMT	0	0	0	2E-06	0.0012	0.0119	-0.0144	-0.0306	11.6003
DIS	0	0	0	2E-06	0.0015	0.0287	-0.0154	0.4436	15.5094
WFC	0	0	0	3E-06	0.0016	0.0195	-0.0199	0.1984	12.2224

Table 1.4: Descriptive statistics of the 10-minute returns for the 50 stocks of our dataset.

any inference about the movement of a variable by looking at the direction of the other variable. We are going to use the findings of this section for comparison purposes with Chapter 5 mainly, where the conditional dependence structure between the pairs of the selected stocks is explored.

1.5.2 Correlation Between the Returns of the Stocks

This subsection focuses on understanding the linear dependence structure between the stocks' returns. As we have mentioned above, volatility is related to financial risk. A way for specialists to eliminate market risk is to build a portfolio with uncorrelated stocks. This happens because stocks with a high correlation may show similar adverse behaviour during a negative market move. Therefore, it would be helpful for us to have a picture of the correlation between the stocks for this period. Due to the different number of intraday transactions observed for each stock, we use the stocks daily returns in this analysis. The explicit correlation matrix of the daily returns for the underlying stocks is given in Table A.1 in appendix A, while the heatmap in Figure 1.3 gives a visual representation of this matrix. In Figure 1.3, the red colour stands for a correlation coefficient near 1, the white colour for a coefficient near 0 and the blue colour stands for a correlation coefficient close to -1. We see that most stocks are more or less positively correlated during these three years, with few uncorrelated pairs of stocks.

Moreover, we may obtain more insights investigating the linear dependence of the stocks related to the same sector. More specifically, we anticipate that stocks in the same sector will be highly correlated. The heatmaps in Figure 1.4 depict the correlation between the daily returns of the stocks in the same sector. We observe that stocks in the same sector are positively correlated distinctly, while just few pairs of stocks have a small correlation. The exact correlation coefficients between the stocks related to the same sector are given in Tables A.2-A.10 in appendix A.

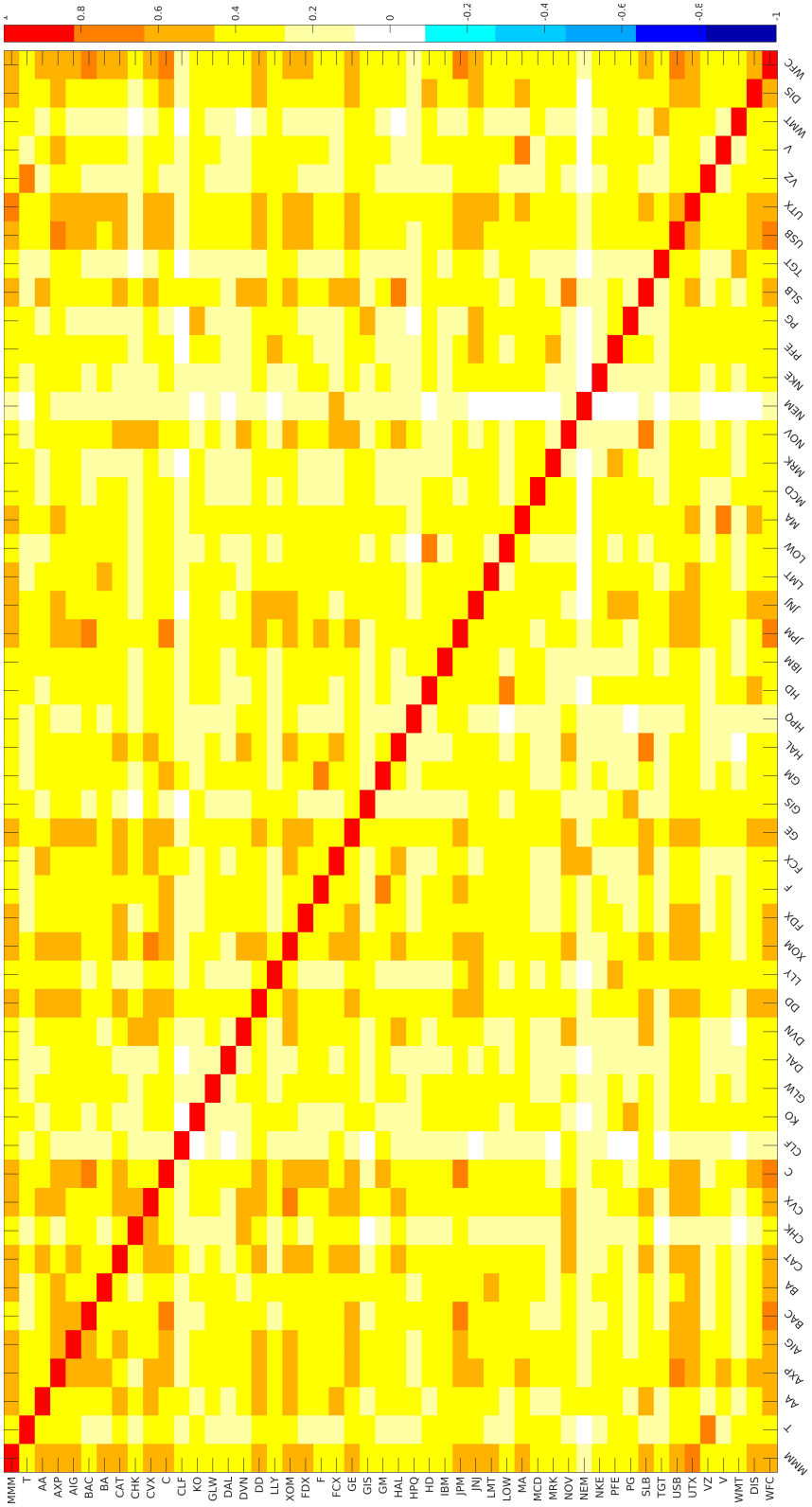

Figure 1.3: This heatmap visualises the correlation coefficient between the daily returns of the underlying stocks.

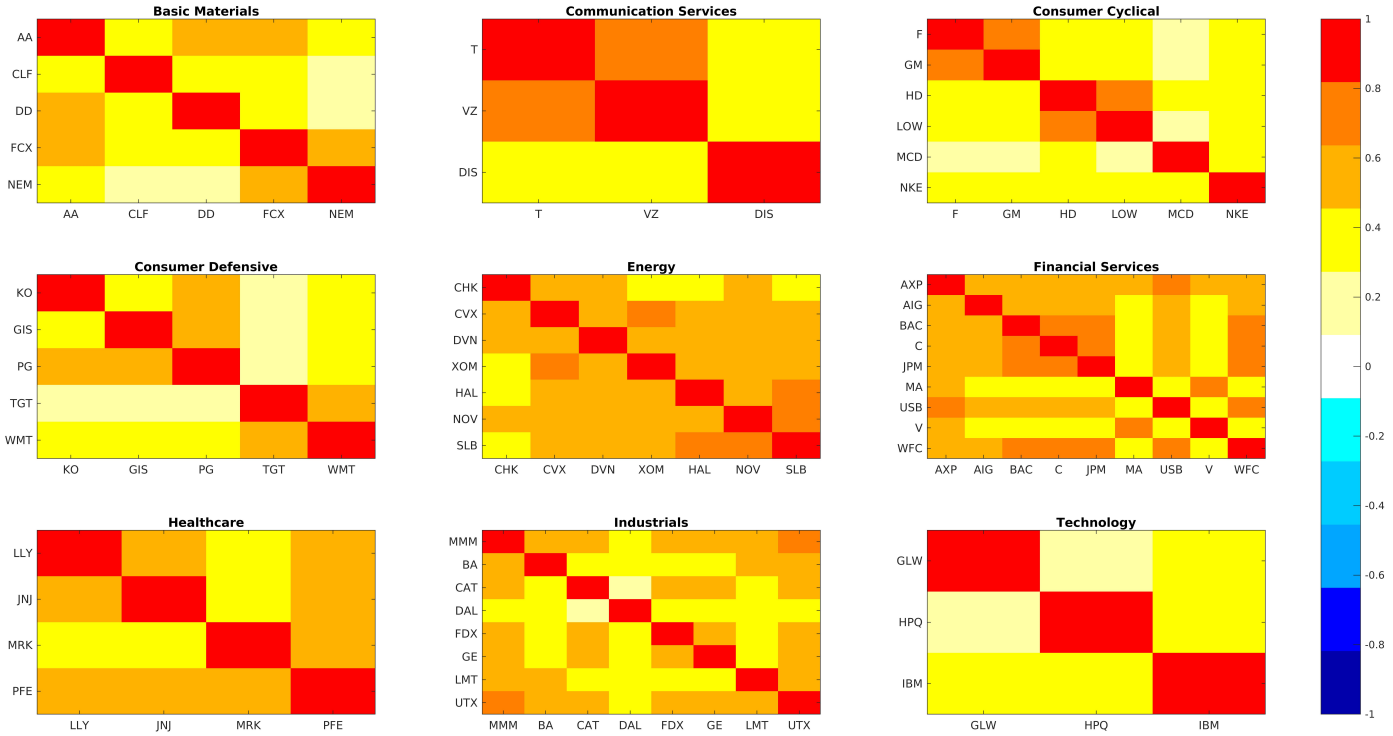

Figure 1.4: These heatmaps visualise the correlation coefficient between the daily returns of the stocks related to the same sector.

1.5.3 Correlation Between the Returns of Stocks and Sectors

The correlation coefficient between the returns of the stocks and the sectors could provide insights into how the stock patterns behave with respect to the patterns of stocks related to a specific sector. For this reason, we have created an index composed of the stocks related to a specific sector, given by:

$$sector\ index_s(d) = \sum_{j=1}^{p_s} w_{s,d}^{(j)} Y_{s,d}^{(j)}$$

where s denotes the specific sector for $s = 1, ..., 9$, $Y_{s,d}^{(j)}$ denotes the daily log-price for the j-th stock related to the sector s for the d day where $j = 1, ..., p_s$. Besides, p_s denotes the number of companies related to the sector s. The weights in the formula above are given by:

$$w_{s,d}^{(j)} = \frac{Y_{s,d}^{(j)}}{\sum_{a=1}^{p_s} Y_{s,d}^{(a)}}$$

where the denominator is the summation of the stocks' daily log-prices for the companies that compose the sector s and refer to the day d.

However, sometimes it would be safer to use the market capitalisation instead of the log-prices for the weights' calculation. The market capitalisation is calculated by:

$$MktCap = \text{total number of company's outstanding shares} \times \text{stock price.} \tag{1.7}$$

In this case, the weights are calculated by:

$$\tilde{w}_{s,d}^{(j)} = \frac{MktCap_{s,d}^{(j)}}{\sum_{a=1}^{p_s} MktCap_{s,d}^{(a)}}$$

where $MktCap_{s,d}^{(j)}$ denotes the market capitalisation of the j-th stock related to the sector s for the d day where $j = 1, ..., p_s$. The denominator indicates the summation of the market capitalisation of the companies which form the sector s and the market

capitalisation refers to the day d. Then, the composite index is defined by:

$$sector\ index_s(d) = \sum_{j=1}^{p_s} \tilde{w}_{s,d}^{(j)} Y_{s,d}^{(j)}$$

where $Y_{s,d}^{(j)}$ denotes the daily log-price for the j-th stock related to the sector s for the d day where $j = 1, ..., p_s$ and $sector\ index_s = [sector\ index_s(1), ..., sector\ index_s(754)] \in \mathbb{R}_+^{1\times 754}$ for the 754 days of our dataset.

The annual market capitalisation is given in Table A.11 in appendix A for all companies. As we see in Equation (1.7), the market capitalisation changes continually during the trading hours since the stock price incurs intraday changes, and occasionally, the number of outstanding shares of a company can change too. However, the annual market capitalisation values have been considered in the explorational procedure of this subsection. Nevertheless, we do not expect that the dynamics of the composite sector indices will be significantly affected by this event. Thus, we can receive valuable insights into the overall image of the correlation between the sectors and the stocks. In the case of sectors, the weights are calculated using the annual market capitalisation. Table 1.5 displays the correlation coefficient between the daily returns of the underlying stocks and the underlying sectors.

We expect positive correlation between most stocks and the sectors' composite index. Further, we anticipate a high correlation between a stock and the related to this stock sector. As we notice, both cases are fulfilled to a high degree, with several exceptional cases. Significantly, the composite index of the sector 'Technology' appears to be negatively correlated with almost every stock. A negative correlation coefficient appears even with two out of the three companies composing this index. The reason behind this event is that the mean value of the market capitalisation for the company International Business Machines (IBM) is \$190.24 billion for the three years. The corresponding mean values of the other two companies are \$24.23 billion for Corning Inc. (GLW) and \$51.59

billion for Hewlett-Packard Co. (HPQ). Thus, IBM ultimately gives the direction of the underlying composite index, allowing the rest of the two companies a proportionally low involvement in the performance of the index 'Technology'. This is also visible in Figure 1.4, which demonstrates a weak and moderate correlation coefficient between IBM and the other two companies. The small number of companies composing this index is an additional factor that inflates this behaviour. With a higher number of companies in this index, the proportion of involvement would have been determined more uniformly among the stocks, allowing us to make safer inferences.

1.5.4 Clustering

A popular unsupervised classification method based on the similarities among the variables of a dataset is the k-means clustering. In short, the idea behind the standard technique is to classify the variables into different groups (also known as clusters) according to the euclidean distance from some centroid values. After an iterative procedure, when the variables' classification ceases to change based on the centroid values, the resulting clusters identify the variables that share common features. Beyond the euclidean distance, alternative distances can be used depending on the analyst's needs each time. In this analysis, the daily squared returns are used and the classification of the stocks has been made with respect to their correlation coefficient. Squared returns are measures of the price changes of the stocks, so by this technique we aim to explore whether we can classify the stocks' price changes into distinct groups.

The standard algorithm in MATLAB does not deliver any strong inference about the optimal number of clusters. In particular, always the highest number of specified possible clusters is chosen. Hence, we cannot draw unequivocal conclusions about the grouping of the stocks with respect to the linear dependence of their squared returns. More elaborate methods will be developed in Chapter 5.

Id. Code	Sector	Basic Materials	Communication Services	Consumer Cyclical	Consumer Defensive	Energy	Financial Services	Healthcare	Industrials	Technology
MMM	I	0.4053	0.9245	0.9658	0.8693	0.8259	0.9266	0.9735	0.9912	-0.8978
T	CS	-0.3316	0.5308	0.1596	0.5435	0.2503	0.2827	0.3236	0.1643	-0.1845
AA	BM	0.7808	0.6018	0.6311	0.444	0.6914	0.4927	0.6554	0.7427	-0.5814
AXP	FS	0.3562	0.9242	0.9552	0.8473	0.8253	0.9355	0.9611	0.9793	-0.88
AIG	FS	0.2375	0.9443	0.958	0.8998	0.8089	0.965	0.9679	0.9486	-0.8809
BAC	FS	0.2128	0.8916	0.9714	0.8683	0.7559	0.982	0.955	0.9353	-0.911
BA	I	0.3934	0.8542	0.9476	0.7958	0.8122	0.9124	0.9218	0.9689	-0.8553
CAT	I	0.8058	0.0265	0.1773	-0.1423	0.4222	0.0415	0.1328	0.2846	-0.12
CHK	E	0.5545	0.5048	0.6323	0.4044	0.8079	0.6103	0.5944	0.7007	-0.5255
CVX	E	0.1865	0.758	0.7108	0.7159	0.8589	0.7668	0.7427	0.6961	-0.5986
C	FS	0.1188	0.8139	0.9444	0.8479	0.6987	0.9692	0.8905	0.8529	-0.836
CLF	BM	-0.0814	-0.9181	-0.8824	-0.9189	-0.6179	-0.869	-0.9228	-0.8438	0.8501
KO	CD	0.0122	0.8608	0.7023	0.8816	0.5554	0.7179	0.7741	0.6811	-0.6143
GLW	T	0.6313	0.7966	0.8664	0.6775	0.8295	0.7739	0.8624	0.9351	-0.7602
DAL	I	0.4171	0.9011	0.9616	0.82	0.8074	0.9154	0.9606	0.9874	-0.9052
DVN	E	0.8579	0.1876	0.2576	-0.0087	0.6252	0.1419	0.2401	0.3996	-0.1189
DD	BM	0.5969	0.7984	0.8755	0.7102	0.7804	0.7961	0.8561	0.9338	-0.7516
LLY	H	0.2117	0.9159	0.8609	0.8811	0.7003	0.8515	0.932	0.8483	-0.8643
XOM	E	0.4994	0.7903	0.7434	0.6814	0.9508	0.7276	0.8068	0.8284	-0.683
FDX	I	0.5242	0.8364	0.9109	0.7645	0.7786	0.8418	0.9054	0.9612	-0.8574
F	CC	0.259	0.6958	0.8956	0.6718	0.743	0.8868	0.8054	0.8273	-0.7614
FCX	BM	0.4347	-0.5219	-0.4892	-0.6233	-0.0043	-0.521	-0.504	-0.3875	0.5111
GE	I	0.2786	0.9244	0.9184	0.8898	0.8499	0.9376	0.9454	0.9418	-0.8536
GIS	CD	0.2113	0.9064	0.9503	0.8843	0.793	0.9463	0.9531	0.9203	-0.8622
GM	CC	0.2054	0.7408	0.9089	0.7344	0.7507	0.9224	0.8338	0.8476	-0.7956
HAL	E	0.5118	0.8168	0.855	0.6741	0.9425	0.8289	0.8714	0.9033	-0.79
HPQ	T	0.6513	0.5944	0.723	0.4901	0.6541	0.603	0.6768	0.7831	-0.5877
HD	CC	0.1148	0.9399	0.9479	0.9479	0.6998	0.9549	0.9595	0.8979	-0.883
IBM	T	-0.2681	-0.4279	-0.559	-0.4512	-0.2769	-0.4617	-0.5217	-0.5846	0.591
JPM	FS	0.2647	0.8575	0.97	0.8425	0.7766	0.9727	0.9346	0.9265	-0.8626
JNJ	H	0.3194	0.9465	0.9704	0.9026	0.8299	0.9469	0.9884	0.9709	-0.8944
LMT	I	0.5126	0.8697	0.9255	0.7831	0.812	0.8577	0.9311	0.9755	-0.8533
LOW	CC	0.2278	0.8724	0.982	0.8775	0.701	0.9627	0.9445	0.9321	-0.8926
MA	FS	0.3095	0.9083	0.9649	0.8623	0.7861	0.952	0.9587	0.9737	-0.9075
MCD	CC	0.2429	0.2576	0.4068	0.263	0.4485	0.3953	0.3372	0.3681	-0.2628
MRK	H	0.4077	0.9271	0.8751	0.8382	0.8271	0.8346	0.9513	0.9281	-0.8452
NOV	E	0.7948	0.3676	0.436	0.2315	0.7665	0.3505	0.4161	0.5428	-0.2709
NEM	BM	-0.2276	-0.9032	-0.9484	-0.8653	-0.712	-0.9281	-0.9462	-0.9494	0.8983
NKE	CC	0.4242	0.8283	0.9511	0.7895	0.7118	0.8828	0.9001	0.946	-0.8397
PFE	H	0.063	0.9077	0.9074	0.9125	0.7044	0.9521	0.9393	0.8763	-0.888
PG	CD	0.1692	0.8852	0.9548	0.9345	0.7165	0.9543	0.9388	0.8996	-0.8533
SLB	E	0.6361	0.7806	0.8247	0.6356	0.9455	0.7694	0.8416	0.9046	-0.7569
TGT	CD	-0.5091	0.473	0.3907	0.6905	0.0651	0.4867	0.3987	0.2368	-0.2729
USB	FS	0.3877	0.9254	0.9107	0.8499	0.8203	0.8886	0.9495	0.9607	-0.8562
UTX	I	0.3577	0.846	0.953	0.7973	0.8271	0.9403	0.9289	0.961	-0.8776
VZ	CS	-0.1126	0.9062	0.7548	0.9068	0.6229	0.822	0.8254	0.7067	-0.7036
V	FS	0.2041	0.9351	0.9636	0.9094	0.7505	0.9646	0.9731	0.9501	-0.9242
WMT	CD	-0.0598	0.873	0.7262	0.9465	0.5618	0.7758	0.8165	0.7123	-0.7036
DIS	CS	0.3427	0.97	0.9448	0.8949	0.8166	0.9186	0.9834	0.9679	-0.8959
WFC	FS	0.404	0.9316	0.9485	0.8579	0.8336	0.9118	0.9676	0.9771	-0.8665

Table 1.5: The first column of this table contains the stocks identifier code. The second column contains the initial letters of the underlying stock's sector. The rest columns exhibit the correlation coefficient between the daily returns of the stocks and the daily returns of the market capitalisation-weighed index of each sector.

Chapter 2

RV-Based Estimators of Intraday Volatility

2.1 Introduction

In the introductory chapter, the notion of historical volatility was outlined. The empirical variance and the standard deviation are two estimators of widespread usage to estimate the daily volatility. However, both measures provide estimates under the assumption of homoscedasticity in the dataset. Furthermore, these estimators utilise the daily returns, omitting a vast number of intraday observations in the estimation.

Plenty of estimators have been proposed in the literature which consider high-frequency financial data to approximate the daily notional volatility (Zhang et al. (2005), Barndorff-Nielsen et al. (2008), Podolskij et al. (2009) among others). The most common measure is the realised variance (also called realised volatility), an empirical estimator of daily volatility. This estimator evaluates the daily volatility with greater efficiency since the intraday observations aid in a more accurate volatility estimation. Realised variance is defined by summing up all the pointwise squared returns of a stock in a day. However, due to the microstructure noise contained in the observed prices and, in turn, in the re-

turns, the realised variance is proven to be a biased estimator of the true daily volatility. Thus, an alternative realised variance-based method has been developed. This estimator makes use of a lower number of intraday data, resulting in the reduction of the bias introduced by the noise aggregation. We define these two estimators in section 2.3.

It would be interesting to investigate how volatility behaves on an intraday basis. This work considers five estimators defined similarly to the realised variance, but they are used for intraday volatility estimation instead. Two of these measures exploit part of the intraday observations for the volatility estimation, whereas the rest three estimators use the total intraday observations as input.

The proposed estimators take into consideration different aspects of the intraday stock behaviour, such as the difference in time between the observed transactions and the number of observations that occurred within some specified intraday intervals. The purpose of defining these volatility estimators is the following. By these estimators, we can investigate the intraday volatility dynamics, as we intend to do in Chapter 3.

The rest of this section has the following structure. Section 2.2 establishes the relevant preliminary background, and two standard estimators of the daily volatility are discussed in section 2.3. In section 2.4, we develop five intraday volatility estimators and report their properties. An empirical example is offered in section 2.5, and this chapter concludes with a discussion in section 2.6.

2.2 Theoretical Framework

In this section, we review the theoretical background of continuous-time stochastic processes required for the definition of realised variance and the related models developed in this thesis. More details about the context of continuous-time stochastic processes can be found in Jacod and Shiryaev (2003), Øksendal (2003), Protter (2005), Capasso and Bakstein (2015) and Klebaner (2012). In particular, one can refer to Elliott and Kopp (1999), Cvitanić and Zapatero (2004), Shreve et al. (2004), Duffie (2010), Lamberton

and Lapeyre (2011), Mykland and Zhang (2012) and Aït-Sahalia and Jacod (2014) for a centralised description of this context in the field of finance. The aforementioned studies are also the sources of the standard definitions reported in this section.

Let $(\Omega, \mathcal{F}, \mathbb{P})$ be the probability space where Ω denotes the sample space, $\mathcal{F}$ denotes the σ-algebra and $\mathbb{P}$ denotes the probability measure. We assume that the true log-price evolution, $\Big(X(t)\Big)_{t\geq 0}$ is defined on the probability space $(\Omega, \mathcal{F}, \mathbb{P})$ and it is a stochastic process.

Definition 2.2.1 (Stochastic Process) *A stochastic process X on $(\Omega, \mathcal{F}, \mathbb{P})$ is a collection of real-valued random variables $\Big(X(t)\Big)_{t\geq 0}$.*

In particular, we assume that the true log-price evolution $\Big(X(t)\Big)_{t\geq 0}$ is a stochastic process which is defined on the filtered probability space $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t\geq 0}, \mathbb{P})$ where $(\mathcal{F}_t)_{t\geq 0}$ represents a family of σ-algebras called filtration.

Definition 2.2.2 (Filtration) *A filtration $(\mathcal{F}_t)_{t\geq 0}$ is an increasing family of sub-σ-algebras of $\mathcal{F}$ (i.e. $\mathcal{F}_t \subset \mathcal{F}$ and if $u < v$, then $\mathcal{F}_u \subset \mathcal{F}_v$).*

Also, we suppose that the filtration $(\mathcal{F}_t)_{t\geq 0}$ satisfies the 'usual conditions' (Liptser and Shiryayev, 1989; Elliott and Kopp, 1999; Protter, 2005), that is, $(\mathcal{F}_t)_{t\geq 0}$ is complete and $\mathcal{F}_u = \cap_{v>u}\mathcal{F}_v$ (i.e. right continuous) for all $u \geq 0$. Broadly speaking, $\mathcal{F}_t$ expresses the knowledge available to investors and researchers up to (and including) time t. In this framework, a process $X(t)$ is called adapted with respect to the filtration $(\mathcal{F}_t)_{t\geq 0}$ or just $\mathcal{F}_t$-adapted when $X(t)$ is $\mathcal{F}_t$-measurable for every t. Given the information in $\mathcal{F}_t$, this means that the value of $X(t)$ is known at time t along with its past history.

A convenient assumption for the true log-price evolution is that it follows a standard Brownian motion.

Definition 2.2.3 (Standard Brownian Motion) *A standard Brownian motion is a process $\Big(W(t)\Big)_{t\geq 0}$ defined by the following properties:*

- $W(0) = 0$ *almost surely;*
- *For any times* u *and* v *such that* $u > v$, $W(u) - W(v) \sim N(0, u - v)$*;*
- *For any times* $t_0, ..., t_n$ *such that* $0 \leq t_0 < t_1 < \cdots < t_n < \infty$*, the random variables* $W(t_0), W(t_1) - W(t_0), \cdots, W(t_n) - W(t_{n-1})$ *are independently distributed;*
- *For each* $w \in \Omega$*, the sample path* $t \mapsto W(w, t)$ *is continuous.*

However, considering that the true log-price evolution follows a standard Brownian motion is a "*too restrictive*" assumption (Duffie, 2010). In this context, a more flexible assumption is that the evolution of the true log-price follows an Itô process.

Definition 2.2.4 (Itô Process) *Let* $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t\geq 0}, \mathbb{P})$ *be a filtered probability space and* $\Big(W(t)\Big)_{t\geq 0}$ *a standard* $\mathcal{F}_t$*-Brownian motion. The process* $\Big(X(t)\Big)_{0\leq t\leq T}$ *is a real-valued Itô process if it can be written as:*

$$X(t) = X(0) + \int_0^t \mu(\varsigma)d\varsigma + \int_0^t \sigma(\varsigma)dW(\varsigma), \qquad 0 \leq t \leq T \tag{2.1}$$

where $X(0)$ *is* $\mathcal{F}_0$*-measurable,* $\Big(\mu(t)\Big)_{0\leq t\leq T}$ *and* $\Big(\sigma(t)\Big)_{0\leq t\leq T}$ *are* $\mathcal{F}_t$*-adapted processes such that* $\int_0^T |\mu(\varsigma)|d\varsigma < \infty$ *almost surely and* $\int_0^T |\sigma(\varsigma)|^2 d\varsigma < \infty$ *almost surely, respectively.*

In high-frequency literature, it is a common assumption that the true log-prices follow an Itô Process (2.1) (Zhang et al., 2005; Barndorff-Nielsen et al., 2008; Podolskij et al., 2009) and we make the same assumption in this work.

In Equation (2.1), the second term is an ordinary integral and the third term is a stochastic integral (for the definition and the properties, see Shreve et al. (2004) and Protter (2005) among others). As opposed to the ordinary integral, where standard calculus can be used to differentiate this integral, the same does not happen for the stochastic integral. This comes from the properties of the standard Brownian motion

contained as a component in this integral. More specifically, although the sample paths of the standard Brownian motion are continuous functions of time t, its paths are not differentiable anywhere (Cvitanić and Zapatero, 2004; Aït-Sahalia and Jacod, 2014).

The Itô Process (2.1) can be equivalently expressed by the following Stochastic Differential Equation (SDE):

$$dX(t) = \mu(t)dt + \sigma(t)dW(t), \qquad 0 \leq t \leq T.$$

The above SDE intuitively indicates that the log-price change of time t is a random process where $\mu(t)$ indicates the instantaneous drift, $\sigma(t)$ signifies the spot volatility of this process and $W(t)$ denotes the standard Brownian motion.

Before proceeding further, we define several necessary concepts at this point. In particular, the concepts of a càdlàg process, stopping times, martingale process, local martingale process, predictable σ-algebra and predictable process are given. Then, we would be able to introduce what a semi-martingale process is, an essential notion in the field of finance, and formalise our assumptions.

Firstly, we define the notions of càdlàg and local martingale processes, while for the latter process we first need to introduce the concepts of a martingale process and stopping times.

Definition 2.2.5 (Càdlàg Process) *A stochastic process X is said to be càdlàg if it almost surely has sample paths which are right continuous with left limits.*

Definition 2.2.6 (Stopping Time) *A random variable $\tau : \Omega \to [0, +\infty]$ is a stopping time if the event $\{\tau \leq t\} \in \mathcal{F}_t$, every t, $0 \leq t \leq +\infty$.*

The intuitive meaning of the definition of stopping times is that τ is called a stopping time when we are able to identify whether τ has occurred (or not) up to (and including) time t, given $\mathcal{F}_t$, for every t.

Definition 2.2.7 (Martingale Process) *A real-valued, adapted process* $\Big(X(t)\Big)_{t\geq 0}$ *is called a martingale with respect to the filtration* $(\mathcal{F}_t)_{t\geq 0}$ *if:*

- $\mathbb{E}[|X(t)|] < \infty$*;*
- *if* $s \leq t$*, then* $\mathbb{E}[X(t)|\mathcal{F}_s] = X(s)$*, almost surely.*

A martingale process describes a process where, loosely speaking, the best guess for a future stock price is given by the current stock price. This process is widely used to describe a fair environment where no one can predict the future value by exploiting the price history up to the current time. Using the definitions of stopping times and martingale processes, we can now define the local martingale process.

Definition 2.2.8 (Local Martingale Process) *An* $\mathcal{F}_t$*-adapted real-valued stochastic process* $\Big(X(t)\Big)_{t\geq 0}$ *is called a local martingale with respect to the given filtration* $(\mathcal{F}_t)_{t\geq 0}$ *if there exists an increasing sequence of* $\mathcal{F}_t$*-stopping times* τ_k *such that:*

- $\tau_k \to \infty$ *almost surely as* $k \to \infty$*;*
- $X(t \wedge \tau_k)$ *is an* $\mathcal{F}_t$*-martingale for all* k*.*

The concept of a local martingale process is used to show that a process "*behaves like a martingale up to suitably chosen stopping times*" (Aït-Sahalia and Jacod, 2014). If a process is a martingale, then it is a local martingale process, whereas the opposite is not necessarily true (Cherubini et al., 2011; Aït-Sahalia and Jacod, 2014).

At this point, we introduce the notion of a predictable σ-algebra.

Definition 2.2.9 (Predictable σ-Algebra) *The* σ*-algebra generated by the adapted left-continuous processes is called the predictable* σ*-algebra.*

Now, we can give the definition of a predictable process. Informally speaking, a predictable process refers to a process where the value of this process at t is known just before t. A predictable process is given by the following definition.

Definition 2.2.10 (Predictable Process) *A process is called predictable if it is measurable with respect to the predictable σ-algebra.*

Using the definitions above, we can now introduce the concept of a semi-martingale process.

Definition 2.2.11 (Semi-Martingale Process) *$X(t)$ is a semi-martingale if it can be written:*

$$X(t) = X(0) + A(t) + M(t), \qquad 0 \leq t \leq T$$

where $X(0)$ is F_0-measurable, $M(t)$ is a local martingale and $A(t)$ is a process of finite variation, i.e.,

$$\sup \sum_i |A(t_i) - A(t_{i-1})| < \infty,$$

where the supremum is over all grids $0 \leq t_0 < t_1 < \cdots < t_n = T$, and all n.

The second term of (2.1) is a process of finite variation. Let $A(t_i) = \int_0^{t_i} \mu(\varsigma) d\varsigma$, then this holds since:

$$\sum_i |A(t_i) - A(t_{i-1})| = \sum_i \Big| \int_{t_{i-1}}^{t_i} \mu(\varsigma) d\varsigma \Big| \leq \sum_i \int_{t_{i-1}}^{t_i} |\mu(\varsigma)| d\varsigma = \int_0^T |\mu(\varsigma)| d\varsigma < \infty, \quad \text{a.s.},$$

for all grids $0 \leq t_0 < t_1 < \cdots < t_n = T$. Also, the last term in (2.1) is a local martingale (see for example Shreve et al. (2004) and Aït-Sahalia and Jacod (2014)). Therefore, the Itô Process (2.1) is a semi-martingale process. Also, the Itô process is an adapted and càdlàg process (Jacod and Shiryaev, 2003; Aït-Sahalia and Jacod, 2014). These are important characteristics since they allow the pricing process to behave similar to the price evolution in continuous trading.

As Cherubini et al. (2011) state "*The decomposition of a semi-martingale is not unique, however; there is at most one such decomposition where the process A can be chosen predictably*". In the case that $A(t)$ is a predictable process of finite variation and

$M(0) = A(0) = 0$, then the semi-martingale process can be represented in a unique way, which gives rise to the concept of a special semi-martingale process (Protter, 2005).

We make the following assumptions on the true log-prices of stocks and the coefficients of the Itô process.

Assumption 2.2.1 *The true log-prices $X(t)$ follow an Itô semi-martingale process, given by (2.1), where the coefficient $\mu(t)$ is a locally bounded predictable process and the coefficient $\sigma(t)$ is a strictly positive càdlàg stochastic process.*

Assumption 2.2.2 *The coefficients $\mu(t)$ and $\sigma(t)$ are jointly independent of $W(t)$ for all t.*

Assumption 2.2.1 is commonly used in the framework of high-frequency volatility estimation (Christensen and Podolski (2005), Barndorff-Nielsen et al. (2008), Podolskij et al. (2009), Gonçalves et al. (2014) among others). It is known that Assumption 2.2.2 implies the absence of the leverage effect in the true price process $X(t)$ (Kristensen, 2010). Hansen and Lunde (2006), Kristensen (2010) and Nolte and Voev (2012) argue that this is not a crucial assumption when it comes to estimation, however it can simplify the derivation of the properties of the estimators.

The semi-martingale assumption enables us to consider a frictionless market (i.e. a market without trading costs and constraints) without arbitrage opportunities (Barndorff-Nielsen et al., 2008). Arbitrage is a term of widespread usage in finance that describes an investor's ability to make money in the market without being exposed to risk. For instance, such a situation occurs when an investor sells a company's stock at a specific price and, at the same time, buys this company's stock at a lower price in a different market. After a short time, the stock price will be the same across the markets since many investors will aim at exploiting such a misprice, offsetting the stock price to the same value across the markets. Hence, the investor has sold a stock expensively and has purchased the same stock cheaply, profiting from the difference in price across the

markets, with no risk.

The Itô semi-martingale process can be extended to a model that includes price jumps in the process by adding an extra term. However, we assume the absence of price jumps in the true log-price evolution in this study. The reader can refer to Aït-Sahalia and Jacod (2012) who provide the mathematical framework around the price jumps in the above process along with an extensive study on high-frequency stock returns under the framework of this model.

The pathwise volatility of the Itô Process (2.1) is given by the Quadratic Variation (QV):

$$QV = \operatorname*{plim}_{n\to\infty} \sum_{i=1}^{n} \Big(X(t_i) - X(t_{i-1}) \Big)^2 = \int_0^T \sigma^2(t) dt$$

and the proof is provided in Shreve et al. (2004). Under the assumption of the absence of price jumps in the Itô process, the QV is synonymous with the Integrated Variance (IV) defined as:

$$IV = \int_0^T \sigma^2(t) dt \,. \tag{2.2}$$

Here, we set $T = 1$ to refer to one day period; hence IV expresses the daily notional volatility.

Let $0 = t_0 < t_1 < t_2 < \cdots < t_n \leq T$ be the n (irregularly spaced) observation times within the horizon $[0, T]$. A realistic assumption in high-frequency econometrics is that the observed log-prices are contaminated by noise. Under the existence of noise, the observed log-price at a given time point $Y(t_i)$ can be expressed as:

$$Y(t_i) = X(t_i) + E(t_i) \tag{2.3}$$

where $X(t_i)$ depicts the i-th intraday true log-price, given by (2.1), and $E(t_i)$ depicts the corresponding error term. In volatility modelling, $\sigma(t)$ is the coefficient of interest, however it is not observable. Hence, we should rely on the observed log-prices to derive

insights about it. We make the following modelling assumptions.

Assumption 2.2.3 *The observed log-price at a given time point t_i, $Y(t_i)$, satisfies the relation in (2.3).*

Assumption 2.2.4 *The error terms $E(t_i)$ are IID with $\mathbb{E}[E(t_i)] = 0$, $\mathbb{E}[E^2(t_i)] = \omega^2 < \infty$ and $\mathbb{E}[E^4(t_i)] < \infty$.*

Assumption 2.2.5 *The price process is independent of the errors.*

Assumptions 2.2.3 and 2.2.4 are common assumptions in this context (Podolskij et al., 2009; Gonçalves et al., 2014; Nolte and Voev, 2012). As regards Assumption 2.2.5, Barndorff-Nielsen et al. (2008) comment that "*from a market microstructure theory viewpoint is a strong assumption as one may expect E to be correlated with increments in $X(t)$. However, the empirical work of Hansen and Lunde (2006) suggests this independence assumption is not too damaging statistically when we analyze data in thickly traded stocks recorded every minute*".

2.3 Realised Variance

An empirical estimator of QV, and by extension of IV, is given by the Realised Variance (RV), defined as:

$$RV^{(all)} = \sum_{i=2}^{n} R^2(t_i). \tag{2.4}$$

where $R(t_i)$ stands for the intraday returns, given by (1.5). As we notice, the formula of $RV^{(all)}$ is nothing more than the daily empirical variance of returns in Equation (1.2) under the assumption of $\mathbb{E}[R(t)] = 0$ for all t, without averaging the squared returns. However, instead of the daily returns, the pointwise returns are utilised in this estimator.

As Andersen et al. (2003) report, under the absence of microstructure noise and price jumps we have that $RV^{(all)} \xrightarrow{p} IV$. However, considering that the Equation (2.3) holds,

then $RV^{(all)}$ is a biased and inconsistent estimator of IV. Further, we recall that the pointwise returns are given by taking the difference of consecutive intraday log-prices. This often results in a (negative) autocorrelation between the successive intraday returns (Taylor, 2005; Tsay, 2005), affecting, in turn, the unbiasedness condition of $RV^{(all)}$. Several alternative RV-based estimators have been proposed to overcome this problem without improving the properties considerably.

Before defining the typical RV estimator, we introduce relevant notations. Suppose we divide a day into J non-overlapping intervals of equal length. Then, each interval I contains a number of observed prices denoted by n_I, for $I = 1, ..., J$. Usually, the length of the intraday intervals is arbitrarily set by the specialist, whilst the number of the J intervals in a trading day is calculated by:

$$J = \frac{23,400}{q \cdot 60}, \quad \text{such that } J \in \mathbb{N}$$

where q denotes the length of the intraday intervals, given in minutes, and 23,400 refers to the total trading seconds in a trading session. For example, using 1-minute intervals we obtain $J = 390$ intraday intervals. Similarly, defining 5-minute intervals we obtain $J = 78$ intervals and so on. In the literature, frequencies of 30, 15, 10 and especially 5-minute returns are used (Andersen et al. (2001a), Zhang et al. (2005), Aït-Sahalia and Yu (2009) among others), given for $J = 13$, 26, 39 and 78 intervals, respectively.

Considering that $Y(t_{I,i})$ denotes the i-th observed log-price within the interval I, the Equation (2.3) can be re-expressed as:

$$Y(t_{I,i}) = X(t_{I,i}) + E(t_{I,i}) \tag{2.5}$$

where $X(t_{I,i})$ denotes the i-th true log-price within the interval I and $E(t_{I,i})$ stands for the corresponding error term. To avoid any misconception, it is important to mention that the subscript I refers to the particular interval for $I = 1, ..., J$ and the subscript i

refers to the i-th observed log-price within this interval.

Let $R(t_{I,i})$ be the intraday pointwise return at t_i within the interval I, given by:

$$R(t_{I,i}) = Y(t_{I,i}) - Y(t_{I,i-1}) \tag{2.6}$$

where $i = 2, ..., n_I$. Then, we define the return of the interval I by:

$$R(t_I) = Y(t_{I,n_I}) - Y(t_{I,1}) \tag{2.7}$$

where n_I denotes the number of observed prices in the interval I and thus $Y(t_{I,n_I})$ represents the last log-price in the same interval.

The standard RV estimator (Andersen et al., 2001b; Barndorff-Nielsen and Shephard, 2002) is expressed as the aggregation of the J squared interval returns:

$$RV = \sum_{I=1}^{J} R^2(t_I) \tag{2.8}$$

where $R(t_I)$ denotes the return of the interval I, given by Equation (2.7) for $I = 1, ..., J$.

In the RV estimator, the effect of noise in the IV estimation is weakened. This happens because RV considers fewer returns as inputs than $RV^{(all)}$, hence less noise is aggregated in the estimation. For this reason, RV is preferred compared to $RV^{(all)}$ in the literature. Ideally, the chosen sampling frequency should yield a high number of intraday returns, so we incorporate many intraday features in the estimation. However, we should also consider the minimum possible number of error terms (embedded in returns) in the aggregation scheme of this measure. The sampling frequency can be optimally chosen with respect to the trade-off between the estimation bias and variance. Nevertheless, specialists often define sampling frequencies of five or ten minutes. The estimators mentioned above are extensively studied -along with similar techniques- by Zhang et al. (2005).

2.4 Intraday Volatility Estimation

One of the aspects that this study investigates is the intraday volatility. By this term, we refer to the volatility within the equilength and non-overlapping intraday intervals of a day. This work develops five empirical estimators that follow a similar rationale to those in the previous section, but the new estimators are utilised for the intraday volatility estimation. Every estimator offers higher estimates for the stocks with particular properties in their observations. One can utilise the intraday estimates of these measures for further analysis. In the next chapter, we develop a methodology that exploits the patterns of the intraday estimates to infer the stocks and the intraday periods with the highest intraday variability based on the different characteristics of the dataset. By these estimators, we aim to incorporate different traits of data that signify volatility in the intraday and daily volatility estimation. These data traits are observed on the high-frequency scale and customarily ignored by conventional methods.

Following the same theoretical background as defined above, the log-price in a given interval I is expressed by (2.5) and the pointwise returns in the interval I are given by (2.6). Then, the pointwise return $R(t_{I,i})$ in an interval I is a function of the true point return and an error term V:

$$\begin{aligned} Y(t_{I,i}) - Y(t_{I,i-1}) = X(t_{I,i}) - X(t_{I,i-1}) + E(t_{I,i}) - E(t_{I,i-1}) \Leftrightarrow \\ R(t_{I,i}) = R^*(t_{I,i}) + V(t_{I,i}) \end{aligned}$$

where $R^*(t_{I,i})$ denotes the true return at time point t_i within the interval I and $V(t_{I,i})$ denotes the difference of the error terms at time points $t_{I,i}$ and $t_{I,i-1}$ within the same interval.

In this context, it is more meaningful to transform Equation (2.2) into:

$$IV_I = \int_{t_{I-1}}^{t_I} \sigma^2(\varsigma) d\varsigma \tag{2.9}$$

where I refers to the corresponding interval for $I = 1, ..., J$. Equation (2.9) describes the notional volatility that arises in the intraday interval I.

We now define two estimators of (2.9). These estimators follow similar rationale as the standard measures $RV^{(all)}$ and RV; however, they are used for intraday estimations. The first estimator is defined as:

$$RV_I^{(Intr-all)} = \sum_{i=2}^{n_I} R^2(t_{I,i})$$

where $R(t_{I,i})$ stands for the poinwise returns within the interval I, given by Equation (2.6), for $I = 1, ..., J$. The second estimator is defined by:

$$RV_I^{(Intr)} = R^2(t_I) = \Big(Y(t_{I,n_I}) - Y(t_{I,1})\Big)^2 \tag{2.10}$$

where $R(t_I)$ denotes the return of the interval I, as given by the Equation (2.7), and n_I denotes the number of observed prices occurred within I.

Let $t_I = I/J$ defines the grid points of the J intraday intervals in a day where $I = 1, \cdots, J$. Using the internal addition property of integrals we notice that:

$$\begin{aligned} IV &= \int_0^1 \sigma^2(\varsigma)d\varsigma = \int_{t_0}^{t_J} \sigma^2(\varsigma)d\varsigma \\ &= \int_{t_0}^{t_1} \sigma^2(\varsigma)d\varsigma + \int_{t_1}^{t_2} \sigma^2(\varsigma)d\varsigma + ... + \int_{t_{I-1}}^{t_I} \sigma^2(\varsigma)d\varsigma + ... + \int_{t_{J-1}}^{t_J} \sigma^2(\varsigma)d\varsigma \\ &= IV_1 + IV_2 + ... + IV_I + ... + IV_J. \end{aligned}$$

Therefore, by aggregating the J estimates of $RV^{(Intr)}$ results in the standard RV estimator in Equation (2.8). In contrast, by aggregating the estimates of the estimator $RV^{(Intr-all)}$, we do not receive the same estimation as $RV^{(all)}$ since $J-1$ fewer returns (at the boundary of the intervals) are used. However, the difference in the estimation is negligible since J is tiny compared to n. Similar to the case between RV and $RV^{(all)}$ for

noisy observations, we expect that $RV^{(Intr)}$ provides superior performance compared to $RV^{(Intr-all)}$ when the intraday estimates are aggregated for the approximation of IV. However, $RV^{(Intr-all)}$ incorporates more observations in the estimation, inevitably discarded by $RV^{(Intr)}$.

In the market, a stock presents some additional traits which can contribute to volatility other than high intraday price fluctuations. These traits are related to the stock trading behaviour, that is how frequently a trade occurs in the market. For example, suppose we observe the same price for a stock at the beginning and the end of the intraday interval I, with many different price values in between. In that case, we notice that there is some volatility, which is not captured by the Estimator (2.10) since its estimate would be zero for this interval. Another example which reveals the weakness of Estimator (2.10) to capture features that show volatility is the following. Suppose a liquid and an illiquid stock which have the same prices at the beginning and the end of the intraday interval I, the Estimator (2.10) will provide the same estimate for both stocks, even if the first one had a more intense trading behaviour in this interval.

High-frequency data have some particular characteristics which include fruitful information about the intraday volatility behaviour of an asset, such as the difference in time between the stock trades and the number of observed prices within an intraday interval. We intend to incorporate these data features in the volatility estimation, introducing the following volatility estimators for the intraday interval I:

$$stRV_I^{(Intr)} = \frac{R^2(t_I)}{\Delta t_I},$$

$$stRV_I^{(Intr-all)} = \sum_{i=2}^{n_I} \frac{R^2(t_{I,i})}{\Delta t_{I,i}},$$

$$wRV_I^{(Intr-all)} = \frac{1}{n_I - 1} \sum_{i=2}^{n_I} R^2(t_{I,i})$$

where $R(t_I)$ is given by Equation (2.7) and $R(t_{I,i})$ is given by Equation (2.6). Here, Δt_I

denotes the difference in time between the last and the first observed prices that occurred in the interval I, i.e. $\Delta t_I = t_{I,n_I} - t_{I,1}$. Similarly, $\Delta t_{I,i}$ denotes the time difference between the consecutive pointwise observed prices in I, i.e. $\Delta t_{I,i} = t_{I,i} - t_{I,i-1}$, and n_I stands for the number of observed prices within I. The intraday estimates are delivered as a $1 \times J$ vector, such that:

$$
\begin{aligned}
RV^{(Intr-all)} &= \left(RV_1^{(Intr-all)}, ..., RV_J^{(Intr-all)}\right) \in \mathbb{R}_+^J, \\
RV^{(Intr)} &= \left(RV_1^{(Intr)}, ..., RV_J^{(Intr)}\right) \in \mathbb{R}_+^J, \\
stRV^{(Intr)} &= \left(stRV_1^{(Intr)}, ..., stRV_J^{(Intr)}\right) \in \mathbb{R}_+^J, \\
stRV^{(Intr-all)} &= \left(stRV_1^{(Intr-all)}, ..., stRV_J^{(Intr-all)}\right) \in \mathbb{R}_+^J, \\
wRV^{(Intr-all)} &= \left(wRV_1^{(Intr-all)}, ..., wRV_J^{(Intr-all)}\right) \in \mathbb{R}_+^J.
\end{aligned}
$$

A list of the main symbols is given in Table 2.1 on page 70.

In the estimators proposed above, the squared pointwise and interval returns are weighted using trading frequency features of the stocks. In this way, we include additional volatility characteristics in the estimation, driven by the intensity of the trading, other than the magnitude of the price fluctuations. Such features are the duration between the observed prices and the number of observed prices within I.

The different weights we introduce in the estimation procedure contribute to higher interval estimates for stocks that exhibit particular properties within the interval I. The main differences among the above estimators are briefly explained. The estimate of $RV_I^{(Intr)}$ exclusively relies on the interval return, whilst the estimator $stRV_I^{(Intr)}$ makes further use of the time space between the first and the last observed prices in the interval I. Hence, we expect the latter to provide higher estimates for intervals with high squared returns in which the time difference between the last and the first observed prices is short. We recall that the observed time points are given in milliseconds after the rescaling procedure to values between zero and one. Thus, the lower the time difference

in the denominator, the higher the volatility estimates for $stRV_I^{(Intr)}$. Practically, this event happens more often for illiquid stocks. So, a high volatility estimation over the interval I is a function of a relatively high squared interval return and a low intraday activity for a stock through $stRV_I^{(Intr)}$.

For $stRV_I^{(Intr-all)}$, a high interval estimate depends on the high squared pointwise returns and the short time space between the sequential observed prices. So, it provides higher estimates for intensively traded stocks with considerable price changes over the given interval. On the other hand, $wRV_I^{(Intr-all)}$ offers high interval estimates for stocks with relatively high squared returns and low trading frequency because of the averaging scheme applied to every interval.

Also, the different weights applied to the returns make the estimates non-comparable across the estimators. As we explain in section 2.4.1, assuming enough observations within I ($n_I \geq 2$), we have that $wRV^{(Intr-all)} \leq RV^{(Intr-all)} \leq stRV^{(Intr-all)}$ since $1/(n_I - 1) \leq 1 \leq 1/\Delta t_{I,i}$. Similarly, $RV^{(Intr)} \leq stRV^{(Intr)}$ since $1 \leq 1/\Delta t_I$. As opposed to $RV^{(Intr-all)}$ and $RV^{(Intr)}$ where the aggregation of the interval estimates provides an approximation of IV, this does not hold for $stRV^{(Intr)}$, $stRV^{(Intr-all)}$ and $wRV^{(Intr-all)}$ due to the additional weights. Thus, the daily volatility estimation can be considered a function of $\widehat{IV}$ for the latter three cases. The benefit of using these estimators is that they allow us to compare the volatility estimates across the stocks, exploiting different characteristics of the dataset. The empirical example of section 2.5 discusses more about the distinctions of these estimators based on an analysis of a liquid and an illiquid stock.

2.4.1 Properties of the Estimators

In this section, we provide the properties of the intraday volatility estimators for the intraday interval I given R^* in terms of mean and variance. For the derivation, along with the assumptions in section 2.2, the following assumption is made.

Notation:	Denotes:
I	A specified intraday interval
J	Number of intraday intervals in a day
n_I	Number of the observed prices in the interval I
$Y(t_{I,i})$	The i-th observed log-price of stock in the interval I
$X(t_{I,i})$	The i-th true log-price of stock in the interval I
$R(t_I)$	The observed interval return for the interval I
$R^*(t_I)$	True interval return for the interval I
$R(t_{I,i})$	The i-th observed pointwise return of stock in the interval I
$R^*(t_{I,i})$	The i-th true pointwise return in the interval I
ω^2	Noise variance
Δt_I	Time difference between the last and the first observed prices in the interval I
$\Delta t_{I,i}$	Time difference between the i-th and its preceding observed price in the interval I

Table 2.1: Summary of important notations.

Assumption 2.4.1 *The observed time points are non-random but not equidistant, thus they are independent of the price process.*

Assumption 2.4.1 rules out the case of observation times dependent on the price process (i.e. endogeneity). This assumption simplifies the derivations and is often made in this context, see for example Mykland and Zhang (2009) and Nolte and Voev (2012) among others. Barndorff-Nielsen et al. (2011) assume random and non-endogenous times for the definition of their estimator. Endogeneity could affect the properties of the RV estimator, as studied by Li et al. (2014b). Mykland and Zhang (2012) review different assumptions on the observation times and their effect on the properties of RV.

More specific modelling assumptions could be made on the time process. For example, if we assume that the observation times are random variables which follow a homogeneous Poisson process, the number of observations N_I in an interval follow a Poisson random variable. Then, conditional on the number of observations within I, $N_I = n_I$, the arrival times within the interval have the same distribution as the order

statistics of Uniform random variables (Renault and Werker, 2011; Mykland and Zhang, 2012). This could give an alternative explanation for almost regular but non-equidistant time points. A general inhomogeneous assumption for the time process requires more complex justifications, which we do not pursue further but refer to Hautsch (2011) for more details.

Before stating the properties of the estimators, we make some informal comparisons of different measures of volatility. Recall that $t_i \in [0,1]$ and $n_I \approx n/J$. The proposed measures can be expressed as a weighted version of the standard realised variance estimators:

$$wRV_I^{(Intr-all)} = \sum_{i=2}^{n_I} m_i R^2(t_{I,i})\,, \quad stRV_I^{(Intr-all)} = \sum_{i=2}^{n_I} s_i R^2(t_{I,i})$$

where, provided that $n_I \geq 2$,

$$m_i = \frac{1}{n_I - 1} \leq 1 \leq \frac{1}{\Delta t_{I,i}} = s_i\,,$$

leading to:

$$wRV_I^{(Intr-all)} \leq RV_I^{(Intr-all)} \leq stRV_I^{(Intr-all)}\,. \tag{2.11}$$

The weight m_i transfers the overall volatility to volatility per transaction, while s_i transfers that to volatility per unit time, thereby revealing different dynamics over the period. For highly traded stocks, m_i decreases, s_i increases and the variation in s_i is small ($s_i \approx s_j$, $i \neq j$). Thus, if s_i or $\Delta t_{I,i}$ does not vary much, $stRV_I^{(Intr-all)}/RV_I^{(Intr-all)}$ is approximately constant for all I. This does not hold for infrequently traded stocks. In particular, the diverging shape between $stRV_I^{(Intr-all)}$ and $RV_I^{(Intr-all)}$ is a sign of irregular trading. On the contrary, $stRV^{(Intr)}/RV^{(Intr)} = 1/\Delta t_I = O(J)$ remains approximately constant in both cases.

Considering the assumptions above, we can now provide the explicit formulas of the

conditional mean and variance of the estimators, given the fixed sample size n_I. The complete derivation of the properties is given in appendix B.1.

Proposition 2.4.1 (Conditional Mean of the Intraday Volatility Estimators) *The conditional expectation of the intraday volatility estimators with respect to R^* over the interval I have the following form:*

$$\begin{aligned}
\mathbb{E}[RV_I^{(Intr)}|R^*] &= \Big(R^*(t_I)\Big)^2 + 2\omega^2, \\
\mathbb{E}[RV_I^{(Intr-all)}|R^*] &= \sum_{i=2}^{n_I}\Big(R^*(t_{I,i})\Big)^2 + 2(n_I-1)\omega^2, \\
\mathbb{E}[stRV_I^{(Intr)}|R^*] &= \frac{\Big(R^*(t_I)\Big)^2}{\Delta t_I} + \frac{2\omega^2}{\Delta t_I}, \\
\mathbb{E}[stRV_I^{(Intr-all)}|R^*] &= \sum_{i=2}^{n_I}\frac{\Big(R^*(t_{I,i})\Big)^2}{\Delta t_{I,i}} + \frac{2(n_I-1)\omega^2}{\Delta t_{I,i}}, \\
\mathbb{E}[wRV_I^{(Intr-all)}|R^*] &= \frac{1}{n_I-1}\sum_{i=2}^{n_I}\Big(R^*(t_{I,i})\Big)^2 + 2\omega^2.
\end{aligned}$$

The interpretation of the above notations is given in Table 2.1. As expected from (2.11), the conditional mean retains the same ordering and shows different factors affecting the finite sample behaviour. In relation to the assumption on the time points, had we assumed random time points, it would have been a simple matter to replace $\Delta t_{I,i}$ and Δt_I by their respective expectations. For example, under a Poisson process with a constant rate λ, the duration $\Delta t_{I,i}$ follows an independent Exponential distribution with mean $1/\lambda$ (Hautsch, 2011) but the mean of the $1/\Delta t_{I,i}$ does not exist so the measure $stRV_I^{(Intr-all)}$ can be very large for highly regular trading stocks. On the contrary, $\Delta t_I = \sum_{i=2}^{n_I}\Delta t_{I,i}$, as the sum of Exponential variables follows Gamma(n_I-1,λ) distribution with mean $\mathbb{E}[\Delta t_I] = (n_I-1)/\lambda$ and $1/\Delta t_I$ follow an Inverse Gamma distribution with mean $\mathbb{E}[1/\Delta t_I] = \lambda/(n_I-2)$ if $n_I > 2$ and variance $\lambda^2/[(n_I-2)^2(n_I-3)]$ if $n_I > 3$. As $\lambda \approx n \approx n_I J$, $stRV_I^{(Intr)}/RV_I^{(Intr)} = O(J)$ so the global behaviour of the estimators

would be similar between random and fixed time points.

For the variance, since all estimators are in the form of $\sum_i w_i R_i^2$ with $R_i = R(t_{I,i})$ and $R_i = R_i^* + V_i$, we derive a general form of the conditional variance below. The complete derivation is found in appendix B.1.3.

Proposition 2.4.2 (Conditional Variance of the Intraday Volatility Estimators) *The conditional variance of the estimator* $\widetilde{RV} = \sum_{i=1}^n w_i R_i$, *given* R^*, *has the following form:*

$$\begin{aligned} Var\Big[\widetilde{RV}|R^*\Big] = 2a_4\Big(\sum_{i=1}^{n-1} w_i(w_i + w_{i+1}) + w_n^2\Big) + 2a_2^2\Big(\sum_{i=1}^{n-1} w_i(w_i - w_{i+1}) + w_n^2\Big) \\ + 4a_3 \sum_{i=1}^{n-1} w_i w_{i+1}(R_i^* - R_{i+1}^*) + 8a_2 \left(\sum_{i=1}^{n-1} w_i R_i^*(w_i R_i^* - w_{i+1} R_{i+1}^*) + w_n^2 (R_n^*)^2\right) \end{aligned}$$

where

$$RV_I^{(Intr-all)} : w_i = 1, \quad stRV_I^{(Intr-all)} : w_i = \frac{1}{\Delta t_{I,i}}, \quad wRV_I^{(Intr-all)} : w_i = \frac{1}{n}$$

and $a_k = \mathbb{E}[E^k(t_i)]$, *for* $k = 1, \ldots, 4$.

In the proposition above, by Assumption 2.2.4 we have that $\mathbb{E}[E^2(t_I)] = a_2 = \omega^2$. Effectively, the most general form with $w_i = s_i$ is for $stRV_I^{(Intr-all)}$. The first term is the leading term in the conditional variance formula, which is proportional to the fourth moment of the error. For $stRV_I^{(Intr-all)}$, although it is not possible to simplify the formula without further assumptions on the model, we can deduce various effects of neighbouring observations in the additional terms. For example, if either $w_i = w_{i+1}$, $R_i^* = R_{i+1}^*$ or $w_i R_i^* = w_{i+1} R_{i+1}^*$, the corresponding terms will disappear.

For specific estimators, we can apply the variance formula with $n = n_I - 1$ or $n = 1$.

Case I: $RV^{(Intr-all)}$ **with** $w_i = 1$ **for all** $i = 1, \ldots, n$**:**

$$Var\Big[RV_I^{(Intr-all)}|R^*\Big]$$

$$= 2n(a_4 + a_2^2) + 2(n-1)(a_4 - a_2^2) + 8a_2 \sum_i (R_i^*)^2 - 8a_2 \sum_i R_i^* R_{i+1}^* + 4a_3 \sum_i (R_i^* - R_{i+1}^*)$$

$$= 4a_4 n - 2(a_4 - a_2^2) + 8a_2 \sum_i (R_i^*)^2 - 8a_2 \sum_i R_i^* R_{i+1}^* - 4a_3(R_n^* - R_1^*)$$

$$= 4a_4 n + O_p(1)\,;$$

Case II: $wRV^{(Intr-all)}$ **with constant weight** $w_i = 1/n$ **for all** $i = 1, \ldots, n$**:**

$$Var\Big[wRV_I^{(Intr-all)}|R^*\Big] = \frac{1}{n^2} Var[RV_I^{(Intr-all)}|R^*] = \frac{4a_4}{n} + O_p(n^{-2})\,;$$

Case III: $RV^{(Intr)}$ **and** $stRV^{(Intr)}$ **with** $n = 1$**:**

$$Var\Big[RV_I^{(Intr)}|R^*\Big] = 2a_4 + 2a_2^2 + 8a_2(R_I^*)^2\,,$$
$$Var\Big[stRV_I^{(Intr)}|R^*\Big] = \frac{1}{(\Delta t_I)^2}(2a_4 + 2a_2^2 + 8a_2(R_I^*)^2)\,.$$

In the above results, R_I^* denotes the true interval return for the intraday interval I, $a_k = \mathbb{E}[E^k(t_i)]$ denotes the k-th moment of the errors $E(t_i)$ for $k = 1, \ldots, 4$, and Δt_I denotes the time space between the first and the last prices within I. Similar derivation is found in Zhang et al. (2005).

2.5 Empirical Example

This section gives an empirical example for the estimators of section 2.4. In this example, the intraday volatility estimators are applied to two stocks for the dates 25, 26 and 27 of November 2013. It would be interesting to comprehend how these estimators behave for stocks with a different trading frequency behaviour. For this reason, we select the most and the least liquid stock of the data period for this investigation. The most liquid stock in terms of the total number of transactions is Citigroup Inc. (C), with an average of 8,387 transactions per day. For the chosen dates, 8,248, 4,679 and

4,799 transactions were monitored. On the other hand, the least active stock concerning the total number of intraday transactions is Mastercard Inc. (MA), with an average of 1,896.7 transactions per day. For the selected dates, 522, 516 and 208 transactions were observed. Indicatively, 4,346.7 intraday transactions occurred per stock, on average, for the data period. As regards the analysis settings, the estimators consider intervals of 30 minutes, resulting in 13 intraday intervals.

In the panels on the left-hand side in Figure 2.1, we present the logarithmic value of the intraday estimates of $RV^{(Intr-all)}$, $stRV^{(Intr-all)}$, $wRV^{(Intr-all)}$, $RV^{(Intr)}$, $stRV^{(Intr)}$ (from top to bottom) for the liquid stock. In the panels on the right-hand side of the same figure, the estimates of the aforementioned estimators are depicted (given in the same order) for the stock with the least frequent trading behaviour. Note that as the scale of the estimates is not comparable, the y-axis varies among the estimators and the stocks. In addition, the second intraday return is zero for $RV^{(Intr)}$ and $stRV^{(Intr)}$ on 27 November 2013, and thus its logarithmic value is not defined. The latter event is not unusual for the estimators $RV^{(Intr)}$ and $stRV^{(Intr)}$. Although there is some price variation within the specific interval, the estimated volatility is zero since the last and first log-price are the same in this interval for the liquid stock. This is one of the aspects the estimators $RV^{(Intr-all)}$, $stRV^{(Intr-all)}$, $wRV^{(Intr-all)}$ come to cover, each of them exploiting different data characteristics for the estimation.

As we notice, the estimators $stRV^{(Intr-all)}$ and $stRV^{(Intr)}$ provide the highest estimates among the estimators for both stocks because of the standardised scheme applied to the squared returns. Also, the aggregation of the squared pointwise returns is the reason for the higher estimates of $stRV^{(Intr-all)}$ compared to $stRV^{(Intr)}$.

In addition, we can see that the intraday performance of $RV^{(Intr)}$ and $RV^{(Intr-all)}$ are similar to $stRV^{(Intr)}$ and $stRV^{(Intr-all)}$, respectively, for plenty cases albeit on a lower scale. This comes from the fact that they use a similar estimation approach. However, the latter estimators additionally standardise the squared returns, providing higher

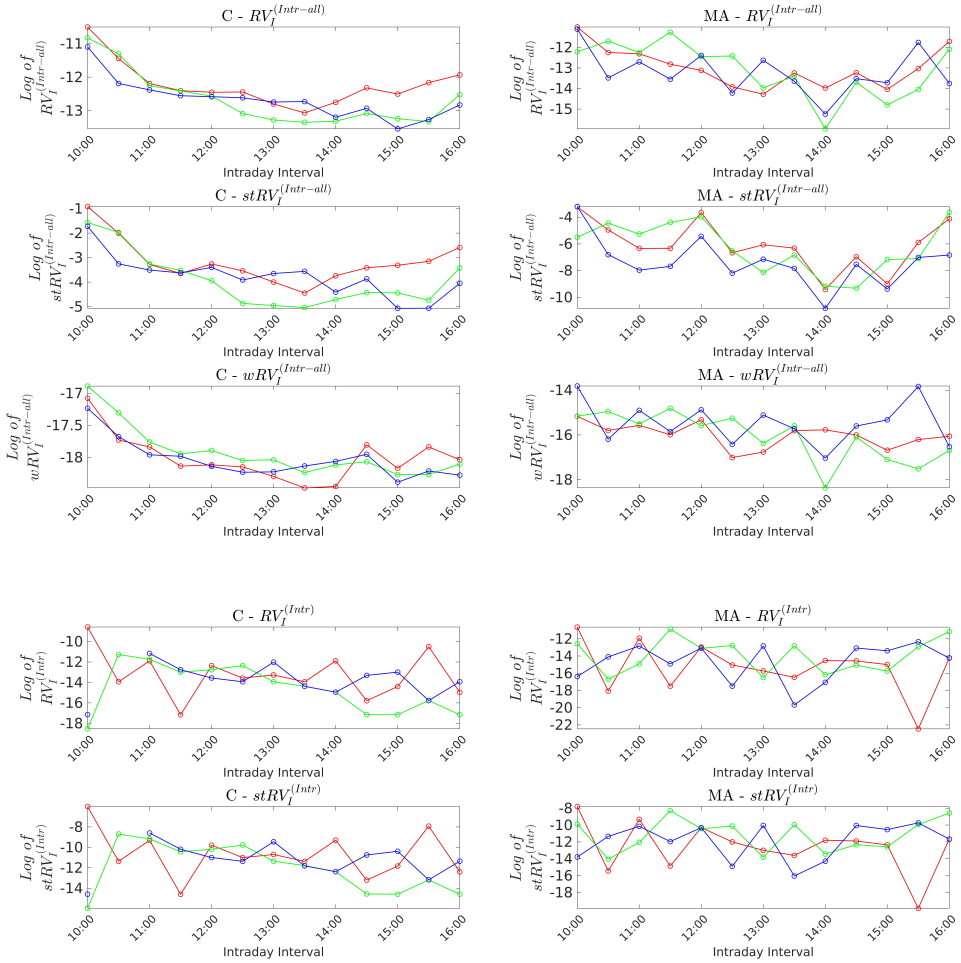

Figure 2.1: The left-hand side column presents the logarithmic value of the intraday volatility estimates of the following estimators $RV^{(Intr-all)}$, $stRV^{(Intr-all)}$, $wRV^{(Intr-all)}$, $RV^{(Intr)}$, $stRV^{(Intr)}$ (from top to bottom) for the liquid stock Citigroup Inc. (C). In the right-hand side column, the logarithmic value of the estimates of these estimators are presented (given in the same order) for the illiquid stock Mastercard Inc. (MA). The trading session has been split into 13 intraday intervals, i.e. $I = 1, \cdots, 13$. The data refer to the dates 25 (red line), 26 (green line) and 27 (blue line) of November 2013.

estimates. In Figure 2.2, we depict the frequency of the durations between consecutive transactions and Table 2.2 reports the average value of these durations for both stocks. It is straightforward that the liquid stock is traded more frequently than the illiquid one; thus, the durations are smaller and usually of compatible size. In contrast, the time difference between consecutive transactions for the illiquid stock is more likely to differ in size. Indicatively, the range of the time difference for the liquid stock (C) on 25/11/2012 is between $[4.274 \cdot 10^{-5}, 3.2 \cdot 10^{-3}]$ whereas the corresponding range for the illiquid stock (MA) on the same day is $[4.274 \cdot 10^{-5}, 4.47 \cdot 10^{-2}]$.

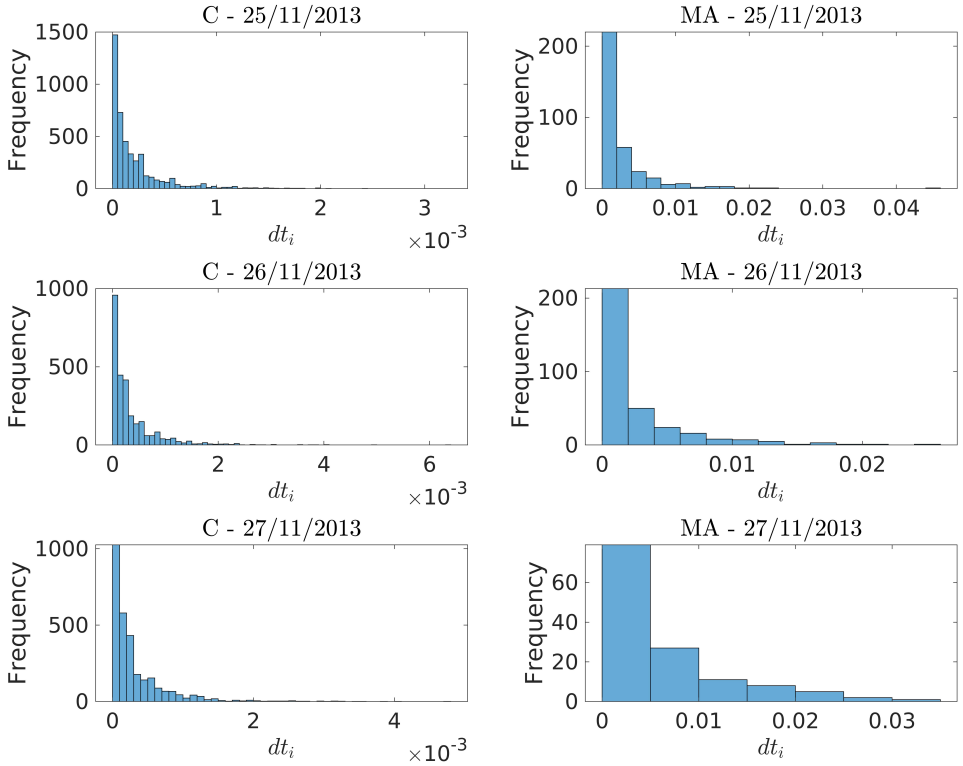

Figure 2.2: The frequency of the durations between the observed prices for the liquid (C) and the illiquid (MA) stock for the three examined days.

The facts above make the behavioural patterns of $stRV^{(Intr-all)}$ and $RV^{(Intr-all)}$ more similar for the liquid stock than the illiquid one, as we also see in Figure 2.1. This

Stock / Date	C	MA
25/11/2013	2.2172E-04	2.5664E-03
26/11/2013	3.5865E-04	2.5403E-03
27/11/2013	3.3281E-04	6.0632E-03

Table 2.2: The average duration between the observed prices for the liquid (C) and the illiquid (MA) stock for the three examined days.

is because the intense trading behaviour of the liquid stock results in comparable weights to the pointwise returns for the volatility estimation. For the illiquid stock, the duration between transactions is more likely to vary substantially across the interval. Thus, the squared pointwise returns may be divided by durations of different length, changing the behavioural patterns of the estimates between $stRV^{(Intr-all)}$ and $RV^{(Intr-all)}$. Hence, we are able to incorporate some additional traits in the volatility estimation through $stRV^{(Intr-all)}$, related to the intensity of the trading frequency, where they are not taken into consideration by the standard-based volatility measure $RV^{(Intr-all)}$.

The estimator $stRV^{(Intr)}$ utilises the squared interval returns divided by the time difference between the last and the first observations within the intervals for the estimation. Similar to the case between $RV^{(Intr-all)}$ and $stRV^{(Intr-all)}$, we anticipate that $stRV^{(Intr)}$ incorporates further information in its estimation, compared to $RV^{(Intr)}$, related to the duration between the first and the last observations within the intervals. Therefore, when this duration differs in size for an interval compared to the other intervals, this will result in a different behavioural pattern for the estimate of $stRV^{(Intr)}$ compared to $RV^{(Intr)}$. Across the intraday intervals, the range of the time difference is $[0.0750, 0.0769]$ (on 25/11/2013), $[0.0747, 0.0768]$ (on 26/11/2013) and $[0.0740, 0.0769]$ (on 27/11/2013) for the liquid stock. The ranges for the illiquid stock are $[0.0480, 0.0767]$, $[0.0423, 0.0751]$ and $[0.0262, 0.0766]$ for these dates (in respective order). The ranges do not substantially vary across days for both stocks; thus, the volatility estimates of $stRV^{(Intr)}$ follow similar patterns to $RV^{(Intr)}$. This is visible in Figure 2.1 where the

volatility estimates of $RV^{(Intr)}$ and $stRV^{(Intr)}$ are identical but of different scale. However, when the duration between the first and the last observations within an interval is of different size compared to the rest intervals, this would be illustrated in the intraday patterns of the estimates.

Even if it is not the case for all days, the estimators provide the highest volatility value for the first intraday interval. In many cases, the volatility gradually falls after the beginning of the trading session and sometimes increases again towards the end of the session. This behaviour is most common for the liquid stock. This is related to a widely observed phenomenon in the markets where the highest volatility usually occurs at the beginning of the trading sessions. Then, volatility declines by lunchtime and increases again before the end of the trading hours (Andersen and Bollerslev (1997), Taylor (2005), Li et al. (2015) among others). We will discuss again about this phenomenon in the next chapter.

After that, the defined intraday estimators can be used for the daily volatility measurement by aggregating the intraday estimates. The summation of the intraday estimates does not offer an estimate of IV for the estimators $stRV^{(Intr)}$, $stRV^{(Intr-all)}$ and $wRV^{(Intr-all)}$, as opposed to $RV^{(Intr)}$ and $RV^{(Intr-all)}$. However, we can use the aggregation of the intraday estimates as an alternative measure to distinguish liquid and illiquid stocks or other characteristics of the stocks which may contribute to volatility.

In Table 2.3, the daily estimates are given for all estimators. We recall that the summation of the intraday estimates of $RV^{(Intr)}$ results in the RV estimator. $RV^{(Intr-all)}$ and $stRV^{(Intr-all)}$ offer the highest volatility estimate for the liquid stock for all days. This is reasonable since the aggregation of the squared pointwise returns for a liquid stock usually results in higher estimates than for a stock with considerably less frequent trading performance.

The estimators $RV^{(Intr)}$ and $stRV^{(Intr)}$ also give higher estimates for the liquid stock, except for the second date in this table where the illiquid stock has the highest estimates

on that date. A high volatility estimate for both estimators is based on high squared interval returns, as opposed to $RV^{(Intr-all)}$ and $stRV^{(Intr-all)}$ where a high estimate is based on high squared pointwise returns. This fact implies that a high trading frequency does not necessarily signal a high volatility estimate for $RV^{(Intr)}$ and $stRV^{(Intr)}$.

For the liquid stock, we notice that the highest estimates are observed on 25 November 2013 for all estimators, except $wRV^{(Intr-all)}$. This can be justified by the high number of transactions that occurred on that date compared to the following two dates. For the stock with the low trading frequency, the estimators demonstrate the highest volatility estimation for the second examined date, except $wRV^{(Intr-all)}$ and $stRV^{(Intr-all)}$. The latter estimator gives the highest estimate for the first date. Both days present a similar number of intraday transactions, whereas the third day has a considerably lower trading frequency.

As regards $wRV^{(Intr-all)}$, it provides the highest estimates for the illiquid stock across the days. By this finding, one can deduce that even if the liquid stock has the most intense trading performance, the average squared return over the intervals is higher for the illiquid stock. Also, the highest daily estimate is offered for the day with the lowest trading frequency (27/11/2013). This is a sign that small but dense price variations contribute to small estimates for $wRV^{(Intr-all)}$. This event can be related to some extent to the occasion where the stock prices commute repeatedly around the same values (bid-ask bounce). Thus, $wRV^{(Intr-all)}$ discounts such a trading behaviour, producing low estimates.

2.6 Discussion

The realised variance is an ideal estimator of IV using high-frequency data without noise under a dense, regular sampling scheme. RV and $RV^{(all)}$ are two common measures of IV. In general, RV is preferred in the literature since it results in less biased estimates of IV. However, a notable drawback of RV compared to $RV^{(all)}$ is that it discards a

Estimator	25/11/2013		26/11/2013		27/11/2013	
	C	MA	C	MA	C	MA
$RV^{(all)}$	8.3065E-05	4.9877E-05	6.0915E-05	5.3542E-05	5.0822E-05	4.5673E-05
$RV/\sum_{I=1}^{13} RV_I^{(Intr)}$	2.3682E-04	3.5706E-05	3.2138E-05	4.7472E-05	3.1076E-05	1.7318E-05
$\sum_{I=1}^{13} RV_I^{(Intr-all)}$	8.2959E-05	4.7698E-05	6.0773E-05	4.9339E-05	5.0657E-05	4.1584E-05
$\sum_{I=1}^{13} stRV_I^{(Intr)}$	3.1083E-03	5.6659E-04	4.2011E-04	6.7223E-04	4.0677E-04	2.7214E-04
$\sum_{I=1}^{13} stRV_I^{(Intr-all)}$	9.0591E-01	1.0143E-01	5.2989E-01	8.3055E-02	4.4461E-01	5.0657E-02
$\sum_{I=1}^{13} wRV_I^{(Intr-all)}$	2.1065E-07	1.6426E-06	2.3527E-07	2.0203E-06	2.0054E-07	3.8750E-06

Table 2.3: The aggregated estimates of the defined estimators are given along with the estimator $RV^{(all)}$ for the liquid (C) and the illiquid stock (MA). The findings refer to the dates 25, 26 and 27 of November 2013.

considerable number of observations. Thus, the intraday data patterns are only exploited to a partial extent, disregarding the initial scope of their usage. Hence, even if RV is a more accurate estimator of the daily notional volatility, we expect that the estimates of $RV^{(all)}$ embody more information about the intraday volatility patterns.

Although useful at the aggregation level, often over a day, we do not gain insights into the volatility's intraday behaviour. We have proposed alternative measures for intraday volatility estimation with high-frequency data by exploiting variable trading characteristics observed in the market linked to the finite sampling effect within irregularly sampled data. The main components accounted for are the varying number of observations and the variation in the times between the transactions over the period of interest. Similar characteristics are often investigated as liquidity measures independently (Aït-Sahalia and Yu, 2009). Our estimators directly incorporate these features into the volatility measure. The estimation procedure is easy and quick, offering the researcher insightful information about the intraday trading behaviour of a stock against the other stocks in terms of volatility dynamics. This information was discussed in section 2.4. Briefly, a high estimate provides a picture of the overall volatility of a stock within a period ($RV_I^{(Intr)}$ and $RV_I^{(Intr-all)}$), the average squared return ($wRV_I^{(Intr-all)}$) or the intensity in the trading frequency of a stock in connection with the squared value of the returns in the understudied period ($stRV_I^{(Intr)}$ and $stRV_I^{(Intr-all)}$).

The empirical findings of section 2.5 showed that all estimators generally give the highest measures for the liquid stock, except $wRV^{(Intr-all)}$. This is due to the averaging scheme applied to the observations per interval for the latter estimator. So even if the liquid stock had a more intense trading behaviour, $wRV^{(Intr-all)}$ showed that the average squared price change was bigger for the illiquid stock.

Most estimators deliver the highest estimates for the first intraday intervals. Then, they often follow a downward trend and sometimes increase again before the end of the trading session. This intraday behaviour occurs more often for the liquid stock, and it is consistent with a well-documented phenomenon in the market which states that volatility is at its highest level at the beginning of the trading session, then gradually falls and it increases again before the end of the session.

As it is natural, the volatility estimates are incomparable because of the different weights we apply to the squared intraday returns in the estimation procedure. Indicatively, $RV^{(Intr-all)}$ and $RV^{(Intr)}$ estimate IV using squared pointwise and interval returns, respectively. The estimates of $stRV^{(Intr-all)}$ and $stRV^{(Intr)}$ are greater than those of the other estimators since the squared pointwise and interval returns, respectively, are divided by the time difference between the observed points. Under normal market circumstances, this time difference is much smaller than 1, giving rise to high estimation values. Besides, the time space between the transactions plays an essential role in the intraday patterns of the estimates of $stRV^{(Intr-all)}$ and $stRV^{(Intr)}$ compared to the patterns of $RV^{(Intr-all)}$ and $RV^{(Intr)}$, as we have explained in sections 2.4.1 and 2.5.

Chapter 3

Factor Analysis with RV-Based Estimators

3.1 Introduction

The understudy dataset consists of the intraday observations of 50 stocks over a period of 745 days. A dataset containing numerous variables is called a multivariate or multidimensional dataset since the data are distributed in a multidimensional space. A common problem that yields when specialists work with a multidimensional dataset is the curse of dimensionality. By this expression, statisticians refer to the effect where the observations are sparsely distributed in the space, making the analysis procedure complex and computationally costly.

Factor analysis is a method of broad usage for the dimensional reduction of the problem. Intuitively, through factor analysis the necessary information is incorporated into several components called factors. Moreover, the dimensions that do not contain comparably valuable information are eliminated in favour of the problem simplification. By this method, practitioners use some specified factors that absorb the essential information instead of the full multivariate dataset, facilitating their study.

In this context, the term information is attributed to the dataset variation in the multidimensional space. Hence by the factor analysis, the common variation observed in the data is summarised into a number of factors, where ideally this number will be substantially lower than the initial high dimensional problem.

Two widely used methods in factor analysis are the maximum likelihood method and the Principal Component Analysis (PCA). The rationale behind the first method is the following. Through an iterative process, the maximum likelihood estimates the factors which make the data most likely to occur. However, the number of the factors is predetermined from the beginning (Tsay, 2005) and it assumes that the variables follow a multivariate normal distribution (Howard, 2016).

On the other hand, PCA is applied to a non-negative definite matrix. In the financial literature, this matrix (usually) refers to the correlation or the sample variance-covariance matrix of the stock returns (Tsay, 2005; Alexander, 2008). The rationale behind this method is the decomposition of the above matrix into some components. These components capture the unique variations of the dataset in the multidimensional space and also these components are uncorrelated to each other. To achieve this, PCA uses the notion of eigenanalysis, as we further describe in section 3.2.1.

In financial studies, the factor analysis is often conducted to the daily returns (Tsay, 2005; Goyal et al., 2008; Făt and Dezsi, 2012). Recently, the factor analysis has been applied to historical daily volatility estimates (Askanazi and Warren, 2017; Shen et al., 2020; Ding et al., 2022). More specifically, Askanazi and Warren (2017) explore the degree to which the factor of the market volatility and the idiosyncratic volatility are connected, concluding that they are highly correlated for stock data. Shen et al. (2020) use the framework of factor modelling in order to express the data's realised covariance matrix (see Barndorff-Nielsen and Shephard (2004) for this topic) in a lower dimension. Ding et al. (2022) re-express the standard factor model in a comparable (multiplicative) form based on the empirical findings in US stocks, demonstrating superior performance

in volatility forecasting compared to standard techniques in this field. Ding et al. (2022) have also shown that applying PCA to idiosyncratic RV estimates instead of idiosyncratic returns contributes to more valuable inferences about the form of the resulting factors.

Dynamic factor analysis is a related popular dimension reduction technique in macroeconomics and finance (for example Luciani and Veredas (2015), Barigozzi and Hallin (2016), Calzolari et al. (2021) among others). A main difference from the standard factor analysis is in the structural form of the model since the variables are expressed as a function of a factor component which contains time dynamics in its form. We are not going to expand further to the concept of dynamic factor models; one can refer to Breitung and Eickmeier (2006) and Barhoumi et al. (2013) for an overview of this notion and existing literature.

In this work, we explore the performance of factor analysis utilising the RV-based estimates as data. Often a period with high stock price fluctuations is followed by a few periods of high volatility. An approach of factor analysis that exploits the information arising over different time lags for a multivariate dataset is considered by Lam et al. (2012). Instead of the data correlation or sample variance-covariance matrix, their approach forms a new matrix, using the data sample autocovariance matrix as a component over different time lags. This approach was developed under the stationarity assumption for the dataset. By this approach, we are able to profit from the cross-correlation and the autocorrelation that the RV estimates can exhibit in their series. This chapter aims to investigate whether we can efficiently summarise the common variation of the multidimensional dataset into a few factors.

A crucial part of factor analysis is the estimation of the number of factors that need to be considered in the model. A simple estimator is developed by Lam et al. (2012). This estimator takes as inputs a number of elements (the eigenvalues of the non-negative definite matrix) for the estimation. The authors recommend that the number of eigenvalues can be practically one-half of the total number of variables. We indicate through

a simulation procedure (given in appendix C.1) that this estimator does not work efficiently when all the available eigenvalues are utilised for the estimation. Furthermore, we confirm that using one-half or one-third of the most informative eigenvalues is adequate for estimating the optimal number of factors we should include in the factor model.

Beyond the simulation procedure, this chapter's novelty lies in two axes. Firstly, in Lam et al. (2012) their methodology is applied to the daily returns of 123 stocks. In our analysis, we conduct their methodology to the daily volatility estimates of the estimators presented in sections 2.3 and 2.4. Hence, our interest is in estimating the variability that daily volatility estimates exhibit over the days for the RV-based estimators which exploit different intraday information to estimate volatility. Secondly, the factor analysis procedure is explored using the intraday volatility estimates as observations. However, we indicate that these estimates exhibit seasonality in their series, violating the stationarity assumption. For this reason, we appropriately modify the approach of Lam et al. (2012) so that we can apply it to this kind of dataset under a theoretically consistent framework.

In particular, analysing the intraday volatility estimates as a time series process can lead to a false inference since these estimates exhibit seasonality over the days, infringing the theoretical framework of Lam et al. (2012). This means that their sample autocovariance matrix is a non-stationary matrix, affecting the validity of their methodology. Our approach treats the whole intraday volatility trajectory estimates over a day as a unit of observation, an instance of functional data (Ramsay and Silverman, 2005; Ferraty and Vieu, 2006; Horváth and Kokoszka, 2012). Hence, although the sample autocovariance matrix is a non-stationary matrix for a day, the intraday autocovariance matrices across the days are considered stationary (Ferraty and Vieu, 2006).

This chapter is set out as follows. In the first two subsections of section 3.2, we report the concepts of PCA and factor analysis. In subsection 3.2.3, an estimator of the optimal number of factors is defined, as proposed by Lam et al. (2012). In section 3.3,

we develop an extension of the methodology of Lam et al. (2012), valid for the (non-stationary) intraday volatility estimates. Besides, this chapter conducts two empirical studies. In the first study, the factor analysis is applied to the daily estimates of the RV-based estimators. The second empirical study is applied to the intraday estimates of these estimators. This chapter ends up with a discussion in section 3.6.

3.2 Factor Analysis

3.2.1 Principal Component Analysis

Principal Component Analysis (PCA) is a popular method utilised in factor analysis. In this subsection, we outline the general idea behind this technique. Loosely speaking, PCA converts the dataset into a lower dimensional summary measure called principal components. The components are ordered in such a way that every principal component represents the direction in which the remaining variation of the dataset exhibits the maximum variation, which means the most relevant information for the dataset.

PCA is linked with the concept of the eigendecomposition of a non-negative definite matrix. By eigendecomposition, a $p \times p$ non-negative definite matrix is decomposed into a simpler form based on p orthogonal vectors, called eigenvectors, corresponding to the ordered eigenvalues. Usually, the dataset's sample correlation or variance-covariance matrix is used for the eigendecomposition (Tsay, 2005; Alexander, 2008). Therefore, a necessary assumption is that the first two moments of the data exist. In the investigation, the volatility estimates of the measures of Chapter 2 are used as data. Thus, we suppose that the first two moments of the volatility estimates exist. This is reasonable under our setting excluding extreme jumps in price changes.

Let $Y = [Y_1, \ldots, Y_p]^T$ denotes a p-dimensional random vector with mean μ_Y and common variance-covariance matrix Σ_Y. Further, let $\widetilde{Y} = [Y_1 - \mu_Y^{(1)}, \ldots, Y_p - \mu_Y^{(p)}]^T$ be the mean-centred observations. Let v be a p-dimensional vector and write $PC = v^T\tilde{Y}$.

PCA aims to find the weight vector v in such a way that satisfies the follwoing condition:

$$\max_{v} Var[v^T \tilde{Y}] = \max_{v} v^T Var\left[\widetilde{Y}\right] v\,, \quad \text{subject to} \quad v^T v = 1\,.$$

Denote the solution by v_1. The successive components v_j are defined in a similar way with additional orthogonality constraint $v_j^T v_j = 1$ and $v_{j-1}^T v_j = 0$ for $j \geq 2$. The solutions to such optimization problem turn out to be the eigenvectors of Σ_Y corresponding to the ordered eigenvalues. The derivation is explicitly given in chapters 1 and 2 in Jolliffe (2011). The j-th eigenvalue and eigenvector satisfy:

$$\Sigma_Y v_j = \lambda_j v_j\,, \quad j = 1, \ldots, p\,,$$

where $\lambda_1 \geq \lambda_2 \geq \ldots \geq \lambda_p \geq 0$ are ordered eigenvalues. Let V be a $p \times p$ matrix where the j-th column represents the eigenvector v_j. By eigendecompoision, we have:

$$\Sigma_Y = V \text{diag}(\lambda_j)_{j=1}^p V^T = \sum_{j=1}^{p} \lambda_j v_j v_j^T\,,$$

where $v_j^T v_j = 1$ and $v_j^T v_k = 0$ for $j \neq k$. Hence, the first principle component PC_1 is given by $v_1^T \widetilde{Y}$, the second principle component PC_2 is given by $v_2^T \widetilde{Y}$ and so on. An essential property of PCA is that the first component corresponds to the eigenvector which captures the highest degree of the data (i.e. the RV-based estimates in this study) variation. Similarly, the second component corresponds to the eigenvector which captures the second-highest degree of variation of the dataset and so on. To be more specific, the following properties hold.

Theorem 3.2.1 (Moments of Principle Components - Härdle et al. (2007)) *For a given $Y \sim (\mu_Y, \Sigma_Y)$ let PC be the principle component transformation. Then:*

$$\mathbb{E}[PC_j] = 0\,, \quad Var[PC_j] = \lambda_j\,, \quad j = 1, \ldots, p\,,$$

and

$$Cov(PC_{j_1}, PC_{j_2}) = 0\,, \quad j_1 \neq j_2\,.$$

By the properties above, we notice three things:

- The lower the eigenvalue λ_j, the lower the variation of the principal component PC_j;
- In the case of a zero eigenvalue, this results in a constant principal component;
- The principle components are uncorrelated to each other.

The above features make PCA a desirable method in the context of dimensionality reduction techniques. More specifically, low eigenvalues demonstrate relatively few pieces of information towards the direction given by the corresponding eigenvectors. Hence, we can reduce the problem's dimensionality by removing the principal components associated with these eigenvalues. Besides, the principle components are uncorrelated to each other. So, the variability in a principal component's form is unique and any other principle component does not contain it. This comes from the orthogonal condition of eigenvectors.

3.2.2 Methodology of Factor Analysis

The idea behind factor analysis is to capture the common variation that arises in the dataset to several uncorrelated components, called factors. This method intends to express this common variation on a lower-dimensional scale. The decomposition of data's variability into some factors can be done by PCA. This section presents the methodology we use in factor analysis, as proposed in Lam et al. (2012).

We consider that the following formulas can describe a p-dimensional variable at the i-th time point:

$$Y_1(t_i) = \Lambda_{1,1}F_1(t_i) + \Lambda_{1,2}F_2(t_i) + ... + \Lambda_{1,k}F_k(t_i) + \epsilon_1(t_i)$$

$$Y_2(t_i) = \Lambda_{2,1}F_1(t_i) + \Lambda_{2,2}F_2(t_i) + ... + \Lambda_{2,k}F_k(t_i) + \epsilon_2(t_i)$$
$$\vdots$$
$$Y_p(t_i) = \Lambda_{p,1}F_1(t_i) + \Lambda_{p,2}F_2(t_i) + ... + \Lambda_{p,k}F_k(t_i) + \epsilon_p(t_i)$$

where $i = 1, ..., n$. Here, p denotes the number of variables and n indicates the number of observations in the dataset. The above formulas can be concisely represented pointwisely by the factor model:

$$Y(t) = \Lambda F(t) + \epsilon(t) \tag{3.1}$$

where $Y(t)$ is a $p \times 1$ vector of the observations at time t. In this section, Y indicates the centred daily volatility estimates of the estimators presented in sections 2.3. The component $F(t)$ represents the model's factors which is a latent variable of size $k \times 1$ and Λ denotes the factor loadings, given as an unknown matrix of size $p \times k$ that has to be estimated along with the factors. The term $\epsilon(t)$ represents the errors, expressed as a $p \times 1$ vector. The elements of this vector follow a zero-mean White Noise (WN) process denoted by $\epsilon(t) \sim WN(0, \Sigma_\epsilon)$.

We make the following assumptions, as reported in Lam et al. (2012).

Assumption 3.2.1 *In Model (3.1), the error terms* $\epsilon(t) \sim WN(0, \Sigma_\epsilon)$. *If* $c'F(t)$ *is WN for a constant* $c \in \mathbb{R}^p$, *then* $c'Cov\Big(F(t+m), \epsilon(t)\Big) = 0$ *for any non-zero integers* m.

Assumption 3.2.2 *In Model (3.1),* $\Lambda'\Lambda = I_k$ *where* I_k *indicates the* $k \times k$ *identity matrix.*

Assumption 3.2.1 assures that the factors are non-trivial and Assumption 3.2.2 is the identifiability condition.

The factors F capture the common variation of the p variables across the days and PCA is the method that will be utilised to decompose the common variation of the dataset into the factors. Later, we will see that the factors are linked to the eigenvectors

through a linear relationship. Thus, exploiting the association between eigenvectors and eigenvalues explained in subsection 3.2.1, we arrive at the following induction. The factor associated with the highest corresponding eigenvalue, via the eigenvector, delivers the highest magnitude of common variation. Accordingly, the factor associated with the second-highest eigenvalue includes the second-highest amount of the common variation and so on. Furthermore, the variation contained in the factors is unique for every factor because of the condition of orthogonal eigenvectors.

Thereafter, the factor loadings, given by the matrix Λ, have a valuable interpretation. More specifically, the squared value of this matrix's elements indicates the variation of a variable captured by a factor. The latent variable F is of an undetermined length k where $k \leq p$. The parameter k defines the number of factors considered in the model, and it is not determined directly by the dataset, so it needs to be estimated.

The value of k plays a crucial role in terms of the problem simplification since it specifies the number of the factors in F that will be used in the Model (3.1). Consequently, a relatively small number for k -compared to p- will result in a small number of factors. These factors follow the data patterns, embedding the higher percentage of the data variation, reducing the problem's dimensionality from p to k at the same time. So, the lower the k value, the better the problem simplification on a lower-dimensional scale. A direct and easy estimation of k is proposed by Lam et al. (2012) and it is given in subsection 3.2.3.

In addition, the following assumptions are made on the factors.

Assumption 3.2.3 *In Model (3.1), $F(t)$ is weakly stationary. Also, for any $m \geq 0$, then $Cov\Big(F(t), \epsilon(t+m)\Big) = 0$.*

Assumption 3.2.3 states that the factors are uncorrelated with any current or future error components. This relaxes a common assumption in factor analysis where the factors are assumed to be uncorrelated with the error components at any time point (Lam et al., 2012). Also, the factors are assumed to be weakly stationary. Although this may seem

a strong assumption at first sight for the daily estimates of RV (Luciani and Veredas, 2015), Ding et al. (2022) show that it is not a damaging assumption in practice. So, it is reasonable to assume that the daily RV estimates are stationary (Ding et al., 2022).

The method developed by Lam et al. (2012) is valid under the assumption of stationarity for the observations. However, this assumption is violated for the intraday volatility estimates, as we will see later, where the intraday volatility series of our estimators exhibit periodicity. In order to analyse data with this kind of characteristic, we properly extend the technique of this section for non-stationary time series in section 3.3.

Factor analysis typically utilises PCA on the correlation or the variance-covariance matrix of the data. Under a theoretically compatible framework to ours, assuming that IV follows a stationary process and the factors' structure exists, Ding et al. (2022) show that conducting PCA on the sample variance-covariance matrix of the RV estimates can consistently approximate the factors' form. Lam et al. (2012) apply PCA to the non-negative definite matrix $\widehat{M}$. This matrix is defined as the summation of the product between the data sample autocovariance matrix and its transpose matrix for all lags up to a specified lag number ℓ:

$$\widehat{M} = \sum_{\ell_0=1}^{\ell} \widehat{\Sigma}_Y(\ell_0)\widehat{\Sigma}_Y^T(\ell_0). \tag{3.2}$$

In the formula above, $\widehat{\Sigma}_Y(\ell)$ denotes the sample autocovariance matrix, given by:

$$\widehat{\Sigma}_Y(\ell_0) = \frac{1}{n-\ell_0} \sum_{i=1}^{n-\ell_0} (Y(t_{i+\ell_0}) - \overline{Y})(Y(t_i) - \overline{Y})^T$$

where $Y(t_i)$ denotes the p-dimensional data at time point t_i for $i = 1, ..., n$ and n indicates the number of observations. Also, ℓ_0 represents the number of the lags we consider in the sample autocovariance matrix each time, while $\overline{Y}$ denotes the variables mean value,

given by:

$$\overline{Y} = \frac{1}{n}\sum_{i=1}^{n} Y(t_i).$$

This is an ideal approach when the data exhibit stable autocorrelation or cross-correlation since it exploits information arising from different lags, which are inevitably ignored by other methods. As Lam et al. (2012) state, data usually have the most potent form of its auto/cross-correlation at its first time-lags, so the chosen ℓ value is suggested to be a low number. We add that for a large dataset, a small number for ℓ is preferred in the sense of a parsimonious approach; that is, for the formation of the matrix $\widehat{M}$ we need fewer matrix operations. However, Lam et al. (2011) show that the choice of ℓ does not substantially distort the validity of the number of factors we choose in the model or the quality of the matrix $\widehat{\Lambda}$. The $\widehat{\Lambda}$ matrix is of size $p \times k$ and constituted by the $\widehat{k}$ eigenvectors with the highest eigenvalue.

By the Model (3.1), the resulting factors are given by:

$$\widehat{F}(t) = \widehat{\Lambda}^T Y(t) \tag{3.3}$$

and the error terms are estimated through the matrix equation:

$$\widehat{\epsilon}(t) = (I_p - \widehat{\Lambda}\widehat{\Lambda}^T)Y(t)$$

where I_p denotes the $p \times p$ identity matrix.

3.2.3 Optimal Number of Factors

The resulting factors capture the highest proportion of the common data variability. So, the first factor captures the highest degree of the joint variation of the dataset, the second factor captures the second-highest degree of the joint variation of the dataset, and so on. A feature of the factor analysis is that the variation captured by a factor

is not included in the variation captured by the rest factors. So every factor contains the common variability observed in the dataset, and at the same time, this variability is unique for every factor.

A central question in the factor analysis is how many factors are needed to describe an adequate high degree of data variation. The magnitude of the variation captured by every factor is related to the corresponding eigenvalues. So, the higher the eigenvalue, the higher the variation captured by the related factor. Hence, the eigenvalues are significantly involved in estimating the number of factors that should be included in the model.

Numerous different approaches have been proposed for the selection of the number of factors that we should incorporate in the Model (3.1). A simple approach is presented by Kaiser (1960) who selects only the factors with eigenvalues higher than one in the model. An alternative technique is suggested by Cattell (1966) who plots the eigenvalues in descending order, choosing the factors whose eigenvalues occur visible change in value in the plot. Besides, Horn (1965) establishes the parallel analysis, another popular method in this area. The techniques above are of widespread usage in the context of factor retention methods; nevertheless, these approaches are based on a heuristic rather than a statistical framework. Under the Model (3.1), Bai and Ng (2002) have developed a criterion based on a squared-residuals minimisation problem for a different number of factors, subject to a penalty term.

However, the ratio-based estimator is preferred in this study, as proposed by Lam and Yao (2011) and Lam et al. (2012). This estimator relies on the theoretical framework of subsection 3.2.2 and appears to have desirable properties. Following their approach, we place the positive eigenvalues of the $\widehat{M}$ matrix in descending order and compute their consecutive ratios. Then, $\widehat{k}$ is chosen to equal the ratio with the lowest value:

$$\widehat{k} = \underset{1 \leq i \leq q}{\arg\min} \frac{\widehat{\lambda}_{i+1}}{\widehat{\lambda}_i}$$

where $\widehat{\lambda}_1 > ... > \widehat{\lambda}_i > ... > \widehat{\lambda}_q > 0$ and q is a constant such that $q < p$. In practice, Lam et al. (2011) and Lam et al. (2012) suggest that we can consider $q = p/2$ in the estimator. As we show in the simulation study in appendix C.1, the estimator works quite well when we exploit one-half or one-third of the eigenvalues. On the other hand, when we use all the eigenvalues, the efficiency of this estimator substantially decreases. So, by utilising one-half of the available eigenvalues, we avoid the effect of very low ratios derived from negligible eigenvalues.

If the lowest value is given for the first ratio, then we form $\widehat{\Lambda}$ using the eigenvector with the highest eigenvalue. In this case, we obtain one proposed factor by Equation (3.3). Similarly, assuming that the lowest value is given for the second ratio, we form $\widehat{\Lambda}$ using the two eigenvectors with the highest eigenvalues, obtaining two proposed factors and so forth. This intuitively means that we are looking for the most remarkable change between the positive and descending eigenvalues. The eigenvectors which correspond to eigenvalues with a lower value than $\widehat{\lambda}_k$ are ignored from the $\widehat{\Lambda}$ matrix specification and the determination of the factors. Lam et al. (2012) display in their simulation study that the higher the sample size, the better the estimation of k.

3.3 Factor Analysis for Intraday Volatility Estimates

An ACF plot visually illustrates the autocorrelation observed in a variable at different time lags. In Figure 3.1, we provide the ACF plots of the intraday volatility estimates of $RV^{(Intr)}$ for the first nine stocks in Table 1.2. The ACF plots depict the autocorrelation of the first 78 lags. This representation uses the intraday estimates as a time series process over the days. Also, the sampling frequency has been set to intraday intervals of 30 minutes, implying that a day consists of 13 intraday intervals. Therefore, the ACF plots refer to a six days period.

The ACF plots demonstrate three essential pieces of information. Initially, there is a high autocorrelation between the first and the second intraday volatility estimates on

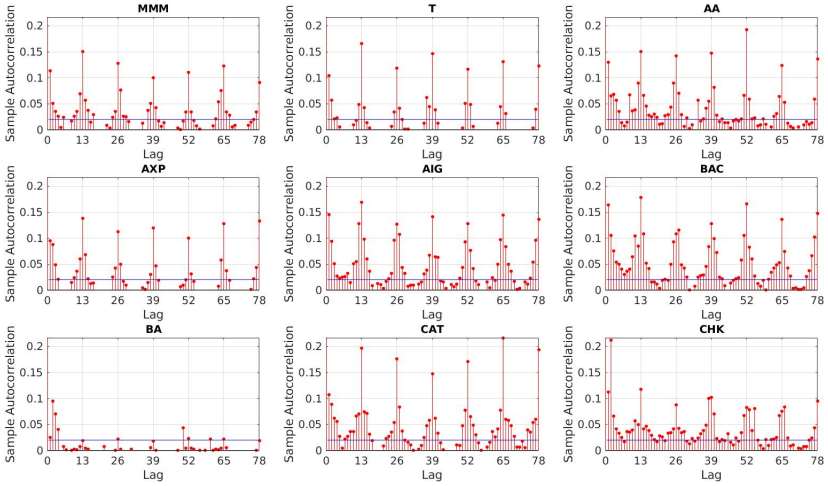

Figure 3.1: This figure depicts the autocorrelation of the volatility estimates of $RV^{(Intr)}$, for the stocks with identifier code MMM, T, AA AXP, AIG, BAC, BA, CAT, CHK, CVX, C, CLF.

day one. This autocorrelation decays over the subsequent interval estimates, and then it follows an upward trend up to the end of the trading session, that is, the 12-th time lag. Secondly, we notice that this diurnal pattern continues over the days, no matter the order of the day in the plot. Thirdly, the autocorrelation is at its highest level every 13 lags, where this lag distance refers to the first intraday interval of each day.

The above periodic patterns can be related to a well-documented event in the financial markets. More specifically, it is widely observed (Andersen and Bollerslev (1997), Taylor (2005), Li et al. (2015) among others) that the volatility is at its highest level at the beginning of the day. Then, it gradually falls until lunchtime and exhibits a slight increase by the end of the trading session, depicted as a U-shaped curve. By the ACF plots, we deduce that the high autocorrelation performed in the plots is directly related to increased volatility estimates. These diurnal features are also visible in the ACF plots of the rest intraday RV-based estimators, given in Figures C.1-C.5 in appendix C.2.

As we have mentioned, the methodology of factor analysis described in section 3.2 is valid when the estimate series follows a stationary process. However, the periodic

pattern in the autocorrelation between the estimates violates the stationarity condition. In this section, we modify the methodology of section 3.2 in order to become a reliable approach under the context of the intraday volatility estimates of the estimators proposed in section 2.4.

We consider an extended factor model for multivariate responses. The unit of time is days, as if daily time series, but the responses over a day constitute multivariate responses on the day t. Let $Y^{(I)}(t) = \left(Y_1^{(I)}(t), ..., Y_p^{(I)}(t)\right)^T$ be a p-variate random vector, representing the I-th response at time t. We assume that the factor model is of the form:

$$Y^{(I)}(t) = \Lambda^{(I)} F(t) + \epsilon^{(I)}(t), \qquad \text{for } I = 1, ..., J \tag{3.4}$$

where $Y^{(I)}(t)$ is a $p \times 1$ vector, $\Lambda^{(I)}$ is a $p \times k$ matrix, $F(t)$ is a $k \times 1$ vector and $e^{(I)}$ is the p-dimensional error term. Note that we assume that there is a common latent process over time t but allows for the effect to vary for different responses within a day. The error terms are assumed to be a WN process. Amalgamating the above notations in a matrix, we have that:

$$\begin{pmatrix} Y^{(1)}(t) \\ \vdots \\ Y^{(J)}(t) \end{pmatrix} = \begin{pmatrix} \Lambda^{(1)} \\ \vdots \\ \Lambda^{(p)} \end{pmatrix} \begin{pmatrix} F_1(t) \\ \vdots \\ F_k(t) \end{pmatrix} + \begin{pmatrix} e^{(1)}(t) \\ \vdots \\ e^{(J)}(t) \end{pmatrix}.$$

To simplify the notation, we can express the extended factor model as:

$$Y(t) = \Lambda F(t) + \epsilon(t)$$

where $Y(t)$ represents a $(Jp) \times 1$ vector of the response variable, that is our intraday RV-based estimates. Here, J denotes the number of intraday intervals and p stands for the number of variables in the dataset. The first p elements in Y represent the estimates of the first intraday interval of the p stocks. Likewise, the following p elements refer to

the estimates of the second intraday interval of the p stocks and so forth.

We claim that $Y^{(I)}(t)$ is likely to be stationary, as their behaviour is similar across days, so is $Y(t)$. However, the different responses within a day are not necessarily stationary, as observed in the empirical analysis. Nevertheless, this is not necessary as the unit of our analysis is the whole responses within a day as a multivariate vector.

Assumption 3.3.1 *In Model (3.4), $F(t)$ is a weakly stationary process across days.*

Under Assumption 3.3.1, $Y(t)$ is a stationary process. So, we define the autocovariance function at lag ℓ as:

$$\Sigma_Y(\ell) = Cov\Big(Y(t+\ell), Y(t)\Big).$$

This covariance matrix can be decomposed into:

$$\Sigma_Y(\ell) = \begin{pmatrix} \Sigma_{1,1}(\ell) & \cdots & \Sigma_{1,J}(\ell) \\ \vdots & \ddots & \vdots \\ \Sigma_{J,1}(\ell) & \cdots & \Sigma_{J,J}(\ell) \end{pmatrix} \tag{3.5}$$

where $\Sigma_{I_1,I_2}(\ell) = Cov\Big(Y^{(I_1)}(t+\ell), Y^{(I_2)}(t)\Big)$ is the covariance between the I_1-th response at time $t+\ell$ and the I_2-th response at time t, for $I_1, I_2 = 1, \ldots, J$.

Similarly to section 3.2, to recover the factors we base our analysis on the M matrix, defined as:

$$M = \sum_{\ell_0=1}^{\ell} \Sigma_Y(\ell_0)\Sigma_Y^T(\ell_0).$$

For a finite sample of data, we can construct an empirical estimator $\widehat{M}$ as:

$$\widehat{M} = \sum_{\ell_0=1}^{\ell} \widehat{\Sigma}_Y(\ell_0)\widehat{\Sigma}_Y^T(\ell_0)$$

where $\widehat{\Sigma}_Y(\ell_0)$ has the same structure as the Matrix (3.5) but with the empirical estimates of $\Sigma_{I_1,I_2}(\ell)$ as elements. In particular, for a finite sample of data $y^{(I)}(t_i)$, $i = 1, ..., n$,

$I = 1, ..., J$, an empirical covariance estimator can be constructed as:

$$\widehat{\Sigma}_{I_1,I_2}(\ell_0) = \sum_{i=1}^{n-\ell_0} \left(y^{(I_1)}(t_{i+\ell_0}) - \overline{y}^{(I_1)}(t_{i+\ell_0})\right)\left(y^{(I_2)}(t_i) - \overline{y}^{(I_2)}(t_i)\right)^T.$$

The selection of the optimal number of factors is delivered through the ratio-based estimator:

$$\widehat{k} = \underset{1\leq i\leq q}{\arg\min} \frac{\widehat{\lambda}_{i+1}}{\widehat{\lambda}_i}$$

where q represents the q highest eigenvalues for $q = (Jp)/2$.

Again, the $\widehat{\Lambda}$ matrix consists of the $\widehat{k}$ eigenvectors corresponding to the $\widehat{k}$ highest eigenvalues. It is worth noting that under this modifying context, the squared factor loadings not only indicate the stocks with the highest captured proportion of variation by the corresponding factors but additionally show the intraday interval that motivates this variation the most. Finally, the model factors and the error terms are estimated through:

$$\widehat{F}(t) = \widehat{\Lambda}^T Y(t),$$

$$\widehat{\epsilon}(t) = (I_{Jp} - \widehat{\Lambda}\widehat{\Lambda}^T)Y(t)$$

where I_{Jp} denotes the $Jp \times Jp$ identity matrix.

In summary, as we illustrate in Figure C.1, the univariate intraday volatility estimates as given by the $RV^{(Intr)}$ estimator exhibit periodicity in their autocorrelation structure. This event is also apparent for the rest of the intraday RV-based estimators in appendix C.2. This fact makes the approach reported in Lam et al. (2012) a precarious method for analysing this kind of intraday observations. For this reason, we modified this approach by transforming the univariate intraday series of a stock into multivariate series so that we obliterate the periodic pattern that arises in their autocorrelation structure. An empirical application of this technique is presented in section 3.5.

3.4 Empirical Analysis - Daily RV-Based Estimates

This study investigates whether we can summarise the common variability yields among the volatility estimates of multiple stocks to a low number of factors for the daily volatility estimators of Chapter 2. More specifically, the factor analysis is conducted on the daily estimates of $RV^{(all)}$, RV, $RV^{(Intr-all)}$, $stRV^{(Intr)}$, $stRV^{(Intr-all)}$ and $wRV^{(Intr-all)}$ for the stocks given in Table 1.2. We remind that the daily volatility estimates for $RV^{(Intr-all)}$, $stRV^{(Intr)}$, $stRV^{(Intr-all)}$ and $wRV^{(Intr-all)}$ are computed by summing up the intraday estimation values. We set 30-minute intervals, yielding 13 intervals within a day. Furthermore, the analysis considers the entire data period from 3 January 2012 to 31 December 2014. The stocks that have experienced stock split during this period are not removed from the dataset in this analysis. This is because the investigation of the daily volatility series is not affected by this event.

The daily estimates of $RV^{(all)}$, RV, $RV^{(Intr-all)}$, $stRV^{(Intr)}$ and $wRV^{(Intr-all)}$ are low numbers which may need even seven or eight decimals so that they will be depicted as non-zero estimates. In the calculation of the sample autocovariance matrix, the pairwise multiplication of these low numbers results in autocovariance approximations of even smaller sizes. This fact implies that the eigenvalue estimates of the $\widehat{M}$ matrix will be tiny numbers, requiring in some cases more than 20 decimals to be illustrated as non-zero values. These eigenvalues are considered practically zero, and for this reason, we multiply the daily estimates of $RV^{(all)}$, RV, $RV^{(Intr-all)}$, $stRV^{(Intr)}$ and $wRV^{(Intr-all)}$ by 1000. In this way, we rescale the estimates to a convenient numerical size which, in turn, results in more easily manageable eigenvalues. Furthermore, the explorational procedure is not affected by this numerical modification. On the other hand, we keep the daily estimates of $stRV^{(Intr-all)}$ unchanged since their range does not affect the interpretation of the eigenvalues of the $\widehat{M}$ matrix.

The observations' mean value can be different from zero, which also happens for the RV-based estimates. Therefore, we use the mean-centred estimates in this analysis to

align the application with the theoretical framework of PCA, as outlined in section 3.2.1. To do so, we subtract the functional mean value of the daily estimates from their value. In this case, the factor model takes the following form:

$$\widetilde{RV}(t) = RV(t) - \mu_{RV}(t) = \Lambda F(t) + \epsilon(t) \tag{3.6}$$

where $RV(t)$ denotes the RV-based estimates, $\mu_{RV}(t) = \mathbb{E}[RV(t)]$ displays the functional mean value of the estimates, and the rest components are the same as defined in section 3.2.2. We depict the centred estimates with the tilde symbol above the corresponding estimator.

Several methods have been proposed in the literature to estimate $\mathbb{E}[RV(t)]$. A simple method is the local polynomial regression method, discussed later in section 4.2. This method estimates the functional mean value of the dataset, taking the observed time points and the corresponding observed values as inputs. Then, the mean value of each point is estimated through a weighting allocation scheme applied to the neighbouring observation values. In this application, we have considered that the weighting allocation scheme is conducted in the observations placed up to four days (i.e. four estimates) away from the given observation each time. This is the case for a smoothing parameter value equal to 0.0027. The weights are normally distributed across the specified neighbourhood through a Gaussian kernel function. The Gaussian kernel function spreads the weights to observations beyond the selected area -because the bell-shaped curve's tails never touch zero- but with a much smaller impact on the estimation of the mean value. The plot of the daily volatility estimates against their estimated mean function is presented in Figure C.6 in appendix C.3 for the first three stocks of Table 1.2, for all estimators.

The first step in the model determination is the estimation of the matrix M in (3.2) for a specified number of time lags ℓ. Lam et al. (2012) comment that ℓ could be a low number since the highest amount of information in a time series process is often observed

at its first lags. In Figure 3.2, we provide an illustrative representation regarding the frequency of the estimated number of factors where the parameter ℓ takes values from 1 to 100. We notice that the choice of ℓ does not determine the number of factors the model selects. For all the estimators in this analysis, we always obtain one factor, no matter the value of ℓ. Thus, we use $\ell = 1$ for reasons of parsimony in the rest of this investigation procedure. In appendix C.3, we also provide the boxplots of the estimated eigenvalues and their ratios for $\ell = 1 - 100$ (see Figure C.7). The boxplots show how these eigenvalues and their ratios vary over the different ℓ values.

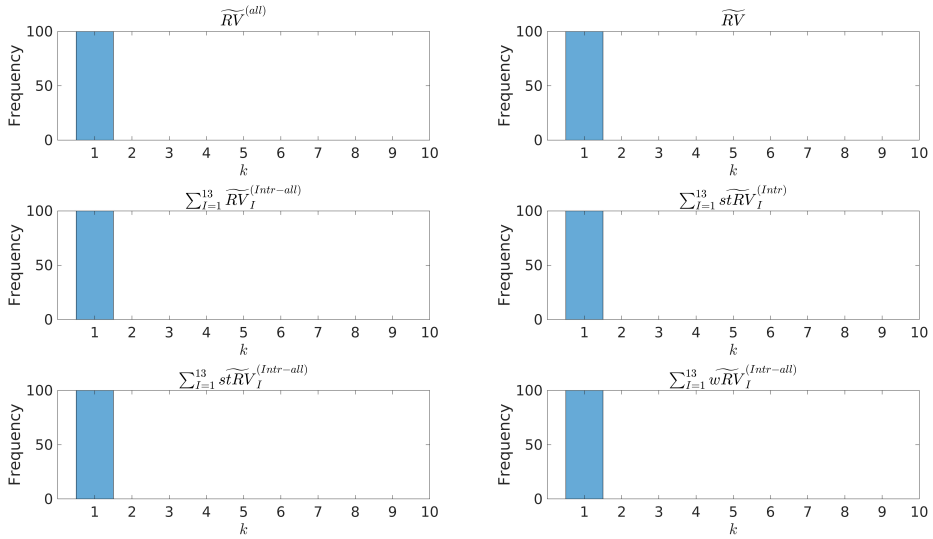

Figure 3.2: This figure presents the frequency of the proposed number of factors given by the ratio-based estimator using $\ell = 1 - 100$ for the daily RV-based estimators.

We estimate one proposed factor for all volatility estimators via the ratio-based estimator. This means that a single factor is sufficient to incorporate the variation of the daily volatility estimates over the data period. In the empirical analysis of Luciani and Veredas (2015), Askanazi and Warren (2017) and Ding et al. (2022), they also conclude that a single factor is adequate to capture the variability of their daily volatility estimates.

An eigenvalue shows the variance of the corresponding factor. Therefore, the proportion of explained variation of each factor is given by dividing the factor's eigenvalue by the sum of all the eigenvalues higher than zero. In Table 3.1, we display the amount of variation explained by the proposed factor for all estimators. The percentage of variation explained by the factor lies between 75.72% and 91.03% where $\sum_{I=1}^{13} \widetilde{stRV}_I^{(Intr-all)}$ and $\sum_{I=1}^{13} \widetilde{stRV}_I^{(Intr)}$ are the estimators with the lowest and highest percentage, respectively. As the ratio-based estimator suggests, more possible factors would have added only a small proportion of captured variation, and hence they have been left outside of the factor model determination in favour of the problem simplification.

	$\widetilde{RV}^{(all)}$	$\widetilde{RV}$	$\sum_{I=1}^{13} \widetilde{RV}_I^{(Intr-all)}$	$\sum_{I=1}^{13} \widetilde{stRV}_I^{(Intr)}$	$\sum_{I=1}^{13} \widetilde{stRV}_I^{(Intr-all)}$	$\sum_{I=1}^{13} \widetilde{wRV}_I^{(Intr-all)}$
	Factor 1	Factor 1	Factor 1	Factor 1	Factor 1	Factor 1
Variation	86.89	90.02	87.20	91.03	75.72	88.93

Table 3.1: The amount of variation explained by the proposed factor, as given for $\ell = 1$.

Thereafter, the squared elements of the $\widehat{\Lambda}$ matrix (the so-called squared factor loadings or briefly squared loadings) show the amount of each stock's variability explained in the resulting factor. In Table 3.2, we list the squared loadings for the most important stocks. The complete table is given in Table C.4 in appendix C.3. We observe that the captured variation of the first factor is mainly driven by the variation in the estimates of the stock Cleveland-Cliffs Inc. (CLF) for all estimators. More specifically, the percentage of CLF in the explained variation of the first factor always exceeds 97%, whereas the rest of the stocks share a negligible proportion. Since the first factor captures the highest degree of data variability, we expect that CLF presents a relatively high variance compared to the rest of the stocks. Indeed as we see in Tables 1.3 and 1.4 in the exploratory analysis of Chapter 1, CLF is the stock with the highest returns variance in the data period.

The volatility estimators we have defined in Chapter 2 use different aspects of data features that contribute to volatility in the estimation. So, the estimated factors are

composed of the variability of different stocks, depending on the features the estimators exploit. Our analysis shows that the stock with the second-highest explained variation by the factor differs across the estimators. For the daily estimates of $\widetilde{RV}^{(all)}$, $\widetilde{RV}^{(Intr-all)}$ and $\widetilde{stRV}^{(Intr-all)}$ is Newmont Goldcorp Corp. (NEM). For $\widetilde{RV}$, $\widetilde{stRV}^{(Intr)}$ is General Motors Corp. (GM) and for $\widetilde{wRV}^{(Intr-all)}$ is Devon Energy Corp. (DVN). In general, all of these stocks display high variation in Tables 1.3 and 1.4 in section 1.5.

Stock	$RV^{(all)}$	$\widetilde{RV}$	$\sum_{I=1}^{13} \widetilde{RV}^{(Intr-all)}$	$\sum_{I=1}^{13} \widetilde{stRV}_I^{(Intr)}$	$\sum_{I=1}^{13} \widetilde{stRV}_I^{(Intr-all)}$	$\sum_{I=1}^{13} \widetilde{wRV}_I^{(Intr-all)}$
	Factor 1	Factor 1	Factor 1	Factor 1	Factor 1	Factor 1
CLF	0.9739	0.976	0.9737	0.9775	0.9717	0.9893
DVN	0.0003	0.0005	0.0004	0.0003	0.0005	0.0033
GM	0.0003	0.005	0.0003	0.004	0	0.0001
NEM	0.0055	0.0006	0.0055	0.0005	0.0053	0.0001

Table 3.2: Squared loadings of the most distinct stocks.

In Figure 3.3, we plot the proposed factor along with centred volatility values of the two stocks which exhibit the highest involvement in the factor's explained variation. The factor is displayed as a solid black line, the most variable stock as a blue dash-dotted line and the stock with the second-highest variability captured by the factor as an orange dash-dotted line. In this figure, the proposed factor efficiently captures the most considerable variation of the most volatile stock (CLF). On the other hand, the variation of the second most variable stock does not drive the factor's variability in such a visible way. Looking at Table 3.2, this is plausible since the second most volatile stock has a tiny involvement in the explained factor's variation.

3.5 Empirical Analysis - Intraday RV-Based Estimates

In this empirical application, we implement the factor analysis in the intraday RV-based estimates of $RV^{(Intr)}$, $RV^{(Intr-all)}$, $stRV^{(Intr)}$, $stRV^{(Intr-all)}$ and $wRV^{(Intr-all)}$, as defined in section 2.4. The intraday volatility estimates stem from the 50 stocks listed in Table 1.2, and they refer to the period between 3 January 2012 and 31 December 2014. We consider 30-minute intervals daily. Also, in order to avoid the effect of tiny eigenval-

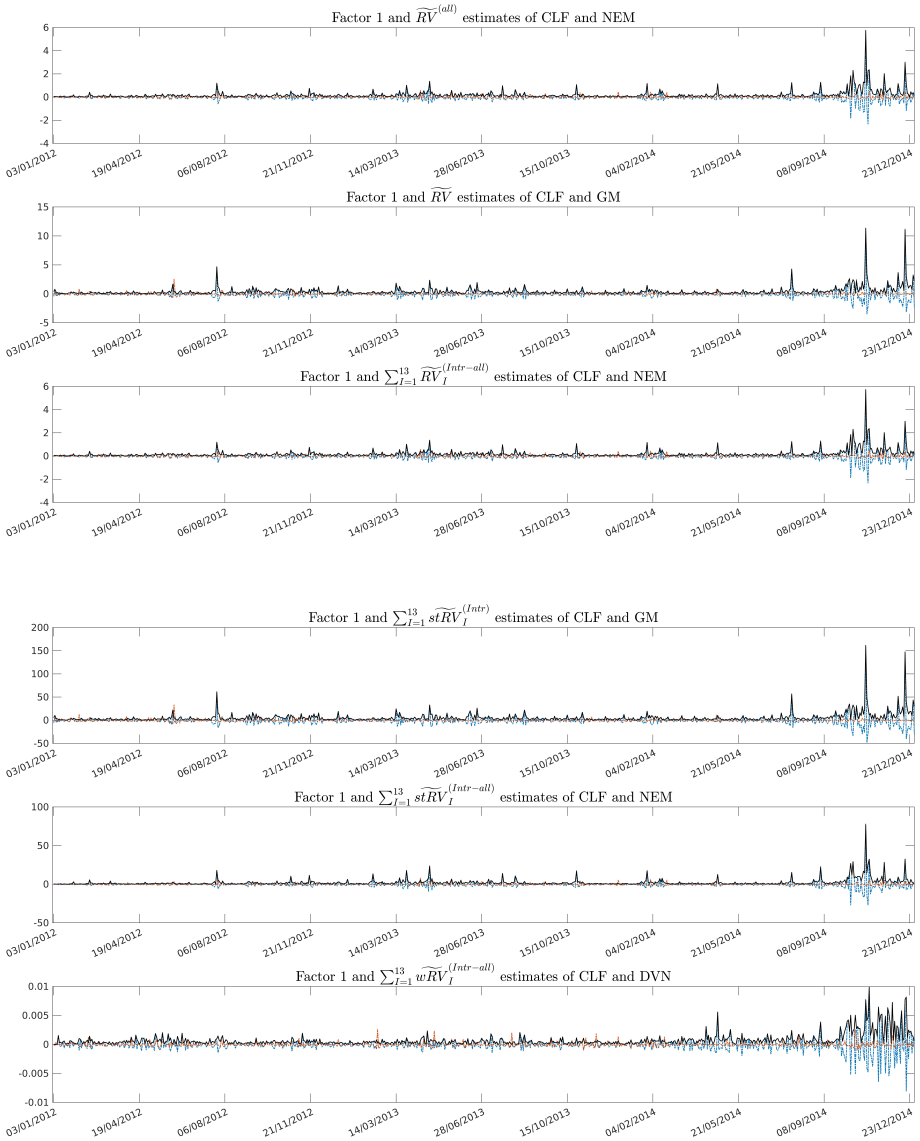

Figure 3.3: This figure presents the proposed factor (solid black line) with the centred volatility estimates of the two stocks that direct this factor the most (blue and orange dash-dotted lines).

ues, we multiply the estimates of $RV^{(Intr)}$, $RV^{(Intr-all)}$, $stRV^{(Intr)}$ and $wRV^{(Intr-all)}$ by 1000.

Similar to the investigation in section 3.4, we analyse the centred values in this application. Thus, we consider the Model (3.6) and the centred estimates are given by subtracting the functional mean values from the original volatility estimates of every intraday interval across days for every stock. The functional mean $\mu_{RV}(t)$ is estimated through the local polynomial regression. The Gaussian kernel function is considered for the weights allocation around the four neighbouring values of the understudy value each time, given for a smoothing parameter value equal to 0.0027. In Figures C.8-C.12 in appendix C.4, we illustrate the interval estimates across the days against their mean functions for the first two stocks of Table 1.2. The centred estimates are depicted with the tilde symbol above the corresponding estimator in the rest of this section.

Figure 3.4 indicates the frequency of the estimated k for lags $\ell = 1 - 100$. Again, a single factor is enough to explain the variation in the centred estimates of the estimators that define the returns over lower frequencies (i.e. $\widetilde{RV}^{(Intr)}$ and $\widetilde{stRV}^{(Intr)}$). On the other hand, the situation seems more ambiguous for the estimators that use all the pointwise returns in a day (i.e. $\widetilde{RV}^{(Intr-all)}$, $\widetilde{stRV}^{(Intr-all)}$ and $\widetilde{wRV}^{(Intr-all)}$). For the estimator $\widetilde{RV}^{(Intr-all)}$, we observe two factors for $\ell = 12 - 15$, three factors for $\ell = 16 - 41$, whilst one factor is enough to explain the data variation for the remaining lag values. In the case of $\widetilde{stRV}^{(Intr-all)}$, usually the ratio-based estimator suggests four factors. Besides, we receive one factor for $\ell = 1 - 12$ and $\ell = 38 - 41$, two factors for $\ell = 15, 16$ and three factors for $\ell = 17 - 37$. As regards the estimator $\widetilde{wRV}^{(Intr-all)}$, we find one factor for the most ℓ values, whereas two factors arise for the lags $\ell = 16 - 39$. The boxplots which depict the range of the estimated eigenvalues and their ratios for the different ℓ values are given in Figure C.13 in the appendix section C.4.

For the estimation of the M matrix, we select $\ell = 1$ for $\widetilde{RV}^{(Intr)}$, $\widetilde{RV}^{(Intr-all)}$, $\widetilde{stRV}^{(Intr)}$ and $\widetilde{wRV}^{(Intr-all)}$. This time lag provides the k with the highest frequency

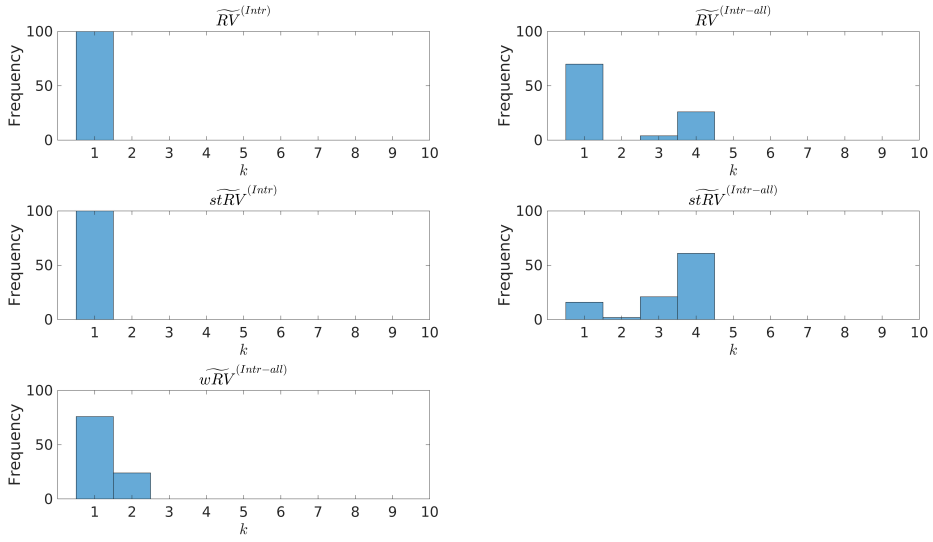

Figure 3.4: This figure presents the frequency of the proposed number of factors given by the ratio-based estimator using $\ell = 1 - 100$ for the intraday RV-based estimators.

for the aforementioned estimators and, at the same time, results in the highest explained variation compared to the lag values that produce more factors in the model. In the case of $\widetilde{stRV}^{(Intr-all)}$, we obtain four factors more frequently across the different lag values. Nonetheless, four factors also yield the highest total explained variation; we define $\ell = 1$, which summarises the data variation in a single factor. We make this choice since we receive three fewer factors by the ratio-based estimator and the cost in the total explained variation is merely about 3%.

Presumably, the more estimated factors in this analysis across the different lags compared to the analysis of the daily estimates are because of the following two reasons. Firstly, when we extend the scale on an intraday basis there are more pieces of information about the stock's variability, not directly visible on the daily basis. Thus, more factors are required to explain this variability sufficiently. Secondly, more factors are evident for the intraday estimators which exploit the full high-frequency observations. Hence, when all the intraday data contribute to volatility estimation, the factor model

often needs more than one factors to capture the volatility patterns effectively.

In Table 3.3, we report the amount of variation expressed by the estimated factors. The variation captured by the first factor lies between 48.94% and 63.71% for the estimators $\widetilde{RV}^{(Intr)}$, $\widetilde{RV}^{(Intr-all)}$, $\widetilde{stRV}^{(Intr)}$ and $\widetilde{stRV}^{(Intr-all)}$, whereas this percentage drops to 35.27% for $\widetilde{wRV}^{(Intr-all)}$. In this table, we additionally report the total variation as would have been determined for the second most beneficial k value for comparison purposes. The total variation is computed as the sum of variation of the proposed factors.

	$\widetilde{RV}^{(Intr)}$	$\widetilde{RV}^{(Intr-all)}$	$\widetilde{stRV}^{(Intr)}$	$\widetilde{stRV}^{(Intr-all)}$	$\widetilde{wRV}^{(Intr-all)}$
	Factor 1	Factor 1	Factor 1	Factor 1	Factor 1
Variation	48.94	63.71	49.95	51.49	35.27
Total explained variation by considering the second most desirable k in the model					
ℓ/k:	-	$\ell = 16$ / $k = 4$	-	$\ell = 13$ / $k = 4$	$\ell = 16$ / $k = 2$
Total Variation	-	59.57	-	54.56	33.76

Table 3.3: The first row in the matrix on the top indicates the amount of variation explained by each factor. The matrix on the bottom reports the total explained variation for the second most desirable case of k, as given for the lowest ℓ value.

Under the framework of daily estimates, the squared loadings indicate the proportion that a stock contributes to a factor's variation. In the case of the intraday estimates, the squared loadings provide the proportion of variability in the intraday interval of a stock as captured by the factor. This investigation offers insights into which intraday periods drive the factor's variation the most. In Table 3.4, we list the factors' squared loadings for the most distinct stocks related to every estimator. CLF remains the stock with the highest proportion of captured variability by the first factor across the estimators. Also, the overall proportion of CLF is higher for the estimators that use interval returns in the volatility estimation. The proportion of the stock with the second-highest involvement in the captured factors's variation differs across the estimators due to the different characteristics the estimators exploit to measure volatility. However, this proportion is considerably lower than that of CLF. Indicatively, the second most variable stock across

the intraday intervals, as captured by the factors, is Chesapeake Energy Corp. (CHK) for $\widetilde{RV}^{(Intr)}$ and $\widetilde{stRV}^{(Intr)}$, Walt Disney Co. (DIS) for $\widetilde{RV}^{(Intr-all)}$, Johnson & Johnson (JNJ) for $\widetilde{stRV}^{(Intr-all)}$ and Mastercard Inc. (MA) for $\widetilde{wRV}^{(Intr-all)}$.

Furthermore, the variation contained in the factors is mainly driven by the first intraday intervals of the most important stocks. An increased captured variability is often visible in the squared loadings for the intraday intervals towards the end of the trading sessions, usually for the volatility measures which use all intraday observations. We can link these findings with the diurnal volatility pattern, described in section 3.3, where the intraday volatility is represented as a U-shaped curve. Our outcome can also be related to the empirical results of Chapter 2 (see Figure 2.1). In that empirical example, we generally observe similar patterns for the stocks' intraday volatility estimates, as found in this analysis. Nevertheless, only a small sample was considered in that application, discouraging us from making generalisations. The full tables of the squared loadings are given in appendix C.4 for all estimators (see Tables C.5-C.9).

In Figure 3.5, the first factor of $\widetilde{RV}^{(Intr)}$, $\widetilde{RV}^{(Intr-all)}$, $\widetilde{stRV}^{(Intr)}$, $\widetilde{stRV}^{(Intr-all)}$ and $\widetilde{wRV}^{(Intr-all)}$ is plotted (from top to bottom) against the intraday centred volatility estimates of the two most volatile stocks across the intraday intervals. The factor is given by the solid black line and the two stocks which drive the variability in this factor by the blue and the orange dash-dotted line. We notice that the factors adequately capture the variability that arises in the intraday intervals of these stocks. CLF is the stock that drives the variability of the proposed factor for all volatility estimators. The second most variable stock does not motivate the variability of the factors so clearly because of the considerably lower involvement of this stock in the factor's variation compared to CLF.

I. I. / Stock	1	2	3	4	5	6	7	8	9	10	11	12	13
	Factor 1 - $\widetilde{RV}^{(Intr)}$												
CHK	0.0095	0.0005	0.0012	0.0008	0.0007	0	0	0	0.0005	0	0.0045	0.0002	0.0004
CLF	0.7554	0.0911	0.0428	0.0001	0.004	0.0063	0.0023	0.0017	0.0001	0.0013	0.0004	0.0003	0.0004
DAL	0.001	0	0.0005	0	0.0002	0	0	0	0.002	0.0001	0.0053	0.0005	0.0006
FCX	0.0069	0.0012	0.0025	0.0011	0.0002	0.0001	0.0008	0	0	0	0	0	0
MA	0.0041	0.0001	0.0008	0.0002	0	0	0.0002	0.0001	0	0	0.0001	0	0
NEM	0.0001	0.0014	0.0004	0	0	0	0.0001	0	0	0	0	0	0
SLB	0.0013	0.0003	0	0	0.0001	0	0.0003	0	0	0.0002	0	0	0
	Factor 1 - $\widetilde{RV}^{(Intr-all)}$												
CHK	0.0001	0	0	0	0.0001	0.0001	0.0002	0	0.0001	0	0	0.0002	0
CLF	0.4738	0.0968	0.031	0.0199	0.0003	0.0073	0.0023	0.0103	0.001	0.0003	0.0027	0.3089	0.005
GLW	0	0	0.0003	0.0001	0	0	0.0003	0	0	0	0	0	0
DAL	0.0002	0	0	0.0002	0	0	0.0001	0	0.0001	0.0002	0	0.0032	0.0001
JNJ	0.0095	0.0014	0.0004	0.0001	0	0	0	0	0	0	0	0.0011	0
MA	0.0002	0	0	0.0004	0	0	0.0002	0.0001	0	0	0.0001	0	0
DIS	0.0017	0.0002	0	0	0	0.0004	0	0.0008	0.0006	0.0001	0	0.0002	0.0005
	Factor 1 - $\widetilde{stRV}^{(Intr)}$												
AA	0.0001	0.0005	0.0032	0.0001	0	0	0.0001	0	0	0	0	0	0
CHK	0.011	0.0004	0.0015	0.0009	0.0008	0	0	0	0.0007	0	0.0053	0.0002	0.0006
CLF	0.7429	0.0795	0.0647	0.0004	0.004	0.0063	0.0016	0.001	0.0002	0.001	0.0004	0.0004	0.0004
DAL	0.0014	0	0.0004	0	0.0001	0	0	0	0.0018	0.0001	0.0045	0.0004	0.0005
FCX	0.0074	0.0011	0.0031	0.0011	0.0002	0.0001	0.0007	0	0	0	0	0	0
MA	0.0045	0.0001	0.0009	0.0002	0	0	0.0002	0.0001	0	0	0.0001	0	0
NEM	0.0001	0.0014	0.0004	0.0001	0	0	0.0001	0	0	0	0	0	0
SLB	0.0011	0.0003	0	0	0.0001	0	0.0003	0	0	0.0002	0	0	0
	Factor 1 - $\widetilde{stRV}^{(Intr-all)}$												
CLF	0.4984	0.0493	0.0212	0.0139	0.0001	0.0003	0.0003	0.0008	0	0.0019	0.0024	0.3475	0.0035
DAL	0.0001	0	0	0.0003	0	0	0.0001	0.0002	0	0.0002	0.0001	0.0056	0.0002
XOM	0.0016	0.0001	0	0	0	0.0001	0	0.0002	0.0002	0	0	0.0002	0.0007
JNJ	0.0222	0.0017	0.0007	0.0001	0	0	0	0	0	0	0	0.0027	0
MA	0.0002	0	0	0.0006	0	0	0.0002	0	0	0	0.0002	0	0
SLB	0.001	0	0	0	0	0	0	0	0	0	0	0.0001	0.0001
DIS	0.002	0.0002	0	0.0001	0	0.0002	0	0.0005	0.0006	0.0001	0	0.0007	0.0019
	Factor 1 - $\widetilde{wRV}^{(Intr-all)}$												
AXP	0.0024	0.0004	0.0001	0	0.0001	0	0	0.0001	0.0002	0	0	0	0.0001
CVX	0	0	0	0	0	0	0	0	0	0	0	0	0
CLF	0.3594	0.3041	0.056	0.0411	0.0396	0.0261	0.0065	0.0002	0.0003	0.0021	0.0002	0.1185	0.0021
DAL	0	0	0	0	0	0.0007	0.0003	0	0.0003	0.0002	0.0001	0.0004	0.0001
DVN	0.0007	0.0001	0	0	0.0002	0	0	0.0001	0.0007	0	0.0001	0.0001	0.0001
LMT	0.0014	0.0001	0	0.0001	0.0005	0	0	0.0002	0	0.0005	0	0	0.0001
MA	0.0014	0.001	0.0008	0.0003	0.0012	0.0005	0.0002	0.0005	0.0003	0	0.0002	0	0
NEM	0	0.0001	0.0001	0	0	0.0001	0	0	0	0.0001	0	0.0006	0
V	0.0001	0.0002	0.0001	0	0.0001	0.0001	0.0005	0.0005	0	0	0	0	0

Table 3.4: This table illustrates the squared loadings of the most distinct stocks for all estimators. The rows refer to the particular stock and the columns refer to the specific intraday interval.

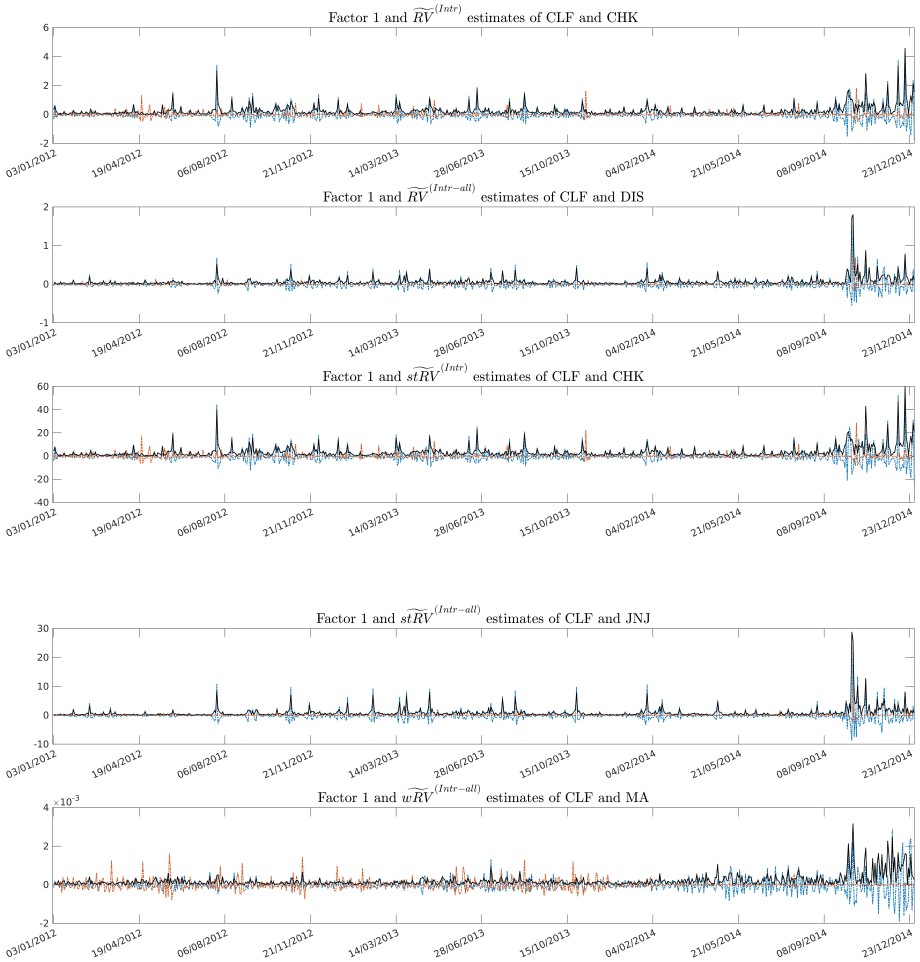

Figure 3.5: In the panels of this figure, we plot the resulting factor (solid black line) along with the centred interval estimates of the stocks with the highest variability (blue and orange dash-dotted lines), as given by the corresponding factor. The headlines include the estimator and the names of the two stocks with the most variable performance captured by the factor.

3.6 Discussion

In this chapter, we investigate the factor analysis approach of Lam et al. (2012) utilising the daily RV-based estimates of the estimators developed in Chapter 2, and we extend this method for the intraday volatility estimates of these estimators. More specifically, we show that considering the intraday volatility estimates as a time series sequence violates the usual stationarity assumption, a critical assumption for the factor analysis approach of Lam et al. (2012). We observe that the non-stationarity is driven by the periodic behaviour across days. For this reason, we extend their methodology to a vector-valued process to properly treat the non-stationarity inherited in the intraday volatility estimates of our dataset.

An essential outcome of the empirical applications is that we can effectively summarise the variability of the daily and intraday volatility estimates in only a few factors. The analysis of the daily volatility estimates reveals that one factor is enough to capture the variability for the estimates of the RV-based measures. This finding is consistent with Luciani and Veredas (2015), Askanazi and Warren (2017) and Ding et al. (2022) who estimate one factor for their daily volatility estimates.

As regards the investigation of the intraday volatility estimates, the empirical results suggest that one factor is enough to explain the variability of the estimates for the volatility estimators which utilise a part of the intraday observations (i.e. for $RV_I^{(Intr)}$ and $stRV_I^{(Intr)}$). In contrast, the estimators which use all the pointwise returns (i.e. $RV_I^{(Intr-all)}$, $stRV_I^{(Intr-all)}$ and $wRV_I^{(Intr-all)}$) need from one to four factors, depending on the lag we choose to form the autocovariance-based matrix. This can be interpreted in the following way. The intraday volatility estimates contain more information about the stock intraday patterns; thus, more factors are often needed to capture their variability effectively. This event is perceptible for the volatility estimators which make use of all intraday data.

All empirical findings in this chapter suggest that CLF exhibit the most variable

performance, which fact is also evident in Tables 1.4 and 1.5 in the exploratory analysis of Chapter 1. The rest stocks share a significantly smaller proportion in the factors' explained variation. This explained variation depends on the volatility estimator we use and the data features that the estimators exploit for the volatility estimation. Also, analysing the intraday volatility estimates permits us to investigate some intraday patterns, not visible in the daily estimates. This is the reason why the stocks with the second-highest captured variability by the factors differ between the daily and the intraday estimates.

An additional advantage of conducting the factor analysis on the intraday compared to the daily estimates is that we can further infer which part of the day motivates the variability explained by the factors the most. The empirical investigation of the intraday volatility estimates shows that most factors' variability is largely driven by the variability observed within the first 30-minute periods over the days. Besides, a high variability is also noticed in some intraday intervals towards the end of the trading session, mainly for the estimators which utilise all the intraday observations. This can be linked with the well-documented diurnal volatility pattern, presented as a U-shaped curve. This finding has also been demonstrated in the empirical example in Chapter 2 for several cases.

Chapter 4

Integrated Variance Estimation with Local Polynomial Regression

4.1 Introduction

Accurate volatility estimation is of vital importance in areas like financial risk analysis and financial risk management since it plays a key role in understanding the causes of market fluctuations or predicting future fluctuations. In this way, experts would be able to protect their portfolios from expected fluctuations that are possible to happen in the future. This chapter develops an estimator of IV, using all of the pointwise intraday returns in the estimation.

The RV-based estimators of section 2.3 are practical tools that allow the investigator to quickly and easily approximate IV. However, several technical and problematic issues can occur when one models the pointwise intraday returns through these estimators. In particular, the aggregation of all the intraday squared returns for the measurement of daily IV can cause high bias in the estimation. For this reason, practitioners prefer to aggregate interval returns to reduce the estimation bias. This technique has been proven superior since less noise is aggregated into the estimate. However, it leaves a part of the

data untapped, resulting in a loss of information in favour of bias elimination. Also, this approach could create computational issues. For example, suppose no transaction can occur within an intraday interval. In that case, the calculation of the interval return is unfeasible, calling the investigator to define a non-misleading return for this interval. In addition, pointwise returns refer to periods of different lengths because of the spatially random design of the observed price realisations in a day. So, an obscure interpretation arises when the investigator wants to compare the returns of these intraday periods.

One can get over the points above by analysing the intraday patterns in continuous time. A popular nonparametric method that permits us to investigate the data patterns in continuous time and, at the same time, the observations are softened towards the local mean value is the Local Polynomial Regression (LPR). Specialists utilise this method to estimate the data mean function based on a weighting allocation scheme. In this chapter, we intend to mitigate the size of the noise in the pointwise squared returns of a stock by using their mean function instead. Then, the smoothed squared returns are used to estimate IV.

In particular, we show that aggregating the elements of the mean function of the squared returns will result in better IV approximations when the noise variance takes values of normal daily size. A similar technique has been proposed by Kristensen (2010) for the instantaneous volatility (also called spot volatility) approximation. The author applies LPR to the pointwise squared returns and then states that aggregating the pointwise estimates can provide a measure of IV. However, his approach relies on the assumption of the absence of noise in the observed values. Also, as far as we know, it has yet to be examined for its accuracy in estimating IV. On the contrary, our estimator assumes that the observed log-prices are blurred by noise and we apply a bias correction, as driven from the properties of the estimator.

Moreover, we exploit the LPR characteristics to overcome the issues that can arise, as quoted above. More specifically, every intraday (smoothed) return is involved in

the IV estimation; thus, the daily estimate absorbs the complete intraday information. Consequently, there is no need for the specification of intraday intervals, and further, the computational problems disappear. In addition, this method enables us to approximate the mean function at any time we wish through interpolation. Hence, we can determine the pointwise observations so they refer to periods of identical length. This technique is mainly used in the methodology of the next chapter.

This chapter proceeds as follows. In section 4.2, we define the LPR method and the proposed IV estimator is presented along with its asymptotic statistical properties in section 4.3. In section 4.4, three benchmark estimators of IV are overviewed. In section 4.6, we examine the performance of the LPR-based estimator against the standard IV estimators through a simulation study. In section 4.7, we highlight a few empirical applications of our estimator to two stocks. Finally, this chapter concludes with a discussion in section 4.8.

4.2 Local Polynomial Regression

Regression analysis is a common way to investigate the dependence structure between the explanatory and the response variables. In particular, when the data exhibit a linear relationship, linear regression is ideal for manipulating such a situation. However, linear regression cannot express this association when the data follow non-linear patterns. This happens because the linear line used to describe this relationship fails to capture the random non-linear trends, generating a significant bias. A famous but not efficient way to fix this feature is by using the notion of polynomial regression. This approach adds more parameters to the model, allowing more flexibility in the estimator. However, this way does have some drawbacks. Fan and Gijbels (1996) mention several of them, such as the degree of the polynomial is hard to be manipulated on a continuous time scale. Also, the polynomial functions have the constraint that all derivatives exist, making the modelling procedure complex enough. Further, some extreme data points can affect the

efficiency of the estimation.

A different way to handle such a situation is by applying a linear regression at various points of the dataset. This gives rise to the concept of local linear regression around a specified neighbourhood. This approach can also be extended in more forms than the linear one by using polynomials of a higher degree. In this case, the idea of the local polynomial regression arises. Fan and Gijbels (1996) is an extensive study which discuss the context around this method.

The central topic around the concept of local polynomial regression is the estimation of the mean function $\mathbb{E}[Y|X = x]$ for n pairs of observations $(x_1, y_1), ..., (x_n, y_n)$. The underlying statistical model can be expressed as follows:

$$Y = m(x) + \sigma(x)\epsilon \tag{4.1}$$

where the function $m(x) = \mathbb{E}[Y|X = x]$ denotes the unknown mean function. Also, $\sigma(x) = \sqrt{Var[Y|X = x]}$ denotes the conditional volatility of Y given X and the errors ϵ have the properties $\mathbb{E}[\epsilon] = 0$, $Var[\epsilon] = 1$ and $X \perp\!\!\!\perp \epsilon$. Furthermore, $m(x)$ and $\sigma(x)$ are fixed functions. The $m(x)$ can be approximated with polynomials around a given point x_0 with the Taylor series:

$$\begin{aligned} m(x) &\approx m(x_0) + m'(x_0)(x - x_0) + \frac{m''(x_0)}{2!}(x - x_0)^2 + ... + \frac{m^{(D)}(x_0)}{D!}(x - x_0)^D \\ &= \beta_0 + \beta_1(x - x_0) + \beta_2(x - x_0)^2 + ... + \beta_D(x - x_0)^D \end{aligned} \tag{4.2}$$

where $\beta_0 = m(x_0)$, $\beta_1 = m'(x_0)$, $\beta_2 = \frac{m''(x_0)}{2!}$,..., $\beta_D = \frac{m^{(D)}(x_0)}{D!}$. Also, x stands for the observations and D denotes the degree of the polynomial in the Taylor series. In the formula above, we assume that all of the derivatives exist. Now, given a finite sample of data (x_i, y_i) for $i = 1, ...n$, our interest is centred on how we will minimise the corresponding error term by utilising the proper amount of contribution of x to the understudy point x_0. This describes a weighted least of squares minimisation problem,

expressed as:

$$\sum_{i=1}^{n} \left(Y_i - \beta_0 - \beta_1(x_i - x_0) + ... + \beta_D(x_i - x_0)^D \right)^2 K_h(x_i - x_0)$$

where the function $K_h(x_i - x_0)$ is called kernel function, and it indicates a non-negative function which is associated with the weights allocation around x_0 for the estimation of the regression function. More details on kernel functions are found in subsection 4.2.1. Then, the minimisation of the empirical error at x_0 is achieved with:

$$\widehat{\beta_{q-1}} = e_q^T (X_{x_0}^T W_{x_0} X_{x_0})^{-1} X_{x_0}^T W_{x_0} Y \tag{4.3}$$

where e_q^T is a $1 \times (D+1)$ vector of zeros except for the position q which is one. So, we obtain the point approximation at x_0 for $q = 1$. For $q = 2$, we obtain an approximation of the first derivative of the function at x_0. Similarly, we obtain a measure of the second derivative at this point for $q = 3$ and so on. This results in the following coefficients for the point x_0:

$$\beta_{x_0} = (\beta_0, \beta_1, ..., \beta_D)^T.$$

The remaining terms of Equation (4.3) are given by:

$$X_{x_0} = \begin{pmatrix} 1 & (x_1 - x_0) & (x_1 - x_0)^2 & \dots & (x_1 - x_0)^D \\ 1 & (x_2 - x_0) & (x_2 - x_0)^2 & \dots & (x_2 - x_0)^D \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & (x_n - x_0) & (x_n - x_0)^2 & \dots & (x_n - x_0)^D \end{pmatrix}, \tag{4.4}$$

$$W_{x_0} = \begin{pmatrix} K_h(x_1 - x_0) & 0 & \dots & 0 \\ 0 & K_h(x_2 - x_0) & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & K_h(x_n - x_0) \end{pmatrix}, \tag{4.5}$$

$$Y = (Y_1, Y_2, ..., Y_n)^T.$$

As we can see in Equation (4.2), the higher the degree of the polynomial, the higher the derivative order used in this equation, and thus, the more precise the point approximation at x_0. Although, this statement is not completely correct. This is because there are some interesting properties about the selection of the polynomial degree in conjunction with the balance between the estimation bias and variance asymptotically.

In Theorem 3.1, Fan and Gijbels (1996) report that considering polynomials of even degree will result in the increase of the estimation bias asymptotically, compared to polynomials with a degree higher by one unit. On the other hand, both polynomial degrees share the same size of conditional variance. Consequently, polynomials of degree $D = 1$ are preferred compared to polynomials of degree $D = 0$, polynomials of degree $D = 3$ are preferred compared to degree $D = 2$ and so on. This happens because of an extra factor in the asymptotic conditional bias for polynomials of odd degree compared to polynomials of even degree (Fan and Gijbels, 1996). So, one can claim that an increase in the polynomial degree results in a drop of the estimation bias; however, as they conclude, it sometimes seems wiser for experts to choose polynomials of an odd than an even degree for the estimation of the regression function. Besides, they propose the linear fit (given for $D = 1$) as a sufficient technique for the point approximation for most cases.

Some indicative cases of local polynomial estimators for a specified point x_0, given the parameter h, are the following:

- For $D = 0$, we end up with the Nadaraya-Watson estimator:

$$\widehat{\beta}_0 = \frac{\sum_{i=1}^{n} w_{ii} Y_i}{\sum_{i=1}^{n} w_{ii}}.$$

- For $D = 1$, we end up with the local linear estimator:

$$\widehat{\beta}_0 = \sum_{i=1}^{n} \frac{(s_2 - s_1 d_i) w_{ii} Y_i}{s_0 s_2 - s_1^2},$$

$$\widehat{\beta}_1 = \sum_{i=1}^{n} \frac{(-s_1 + s_0 d_i) w_{ii} Y_i}{s_0 s_2 - s_1^2}.$$

In the equations above, we have that $w_{ii} = K_h(x_i - x_0)$, $s_q = \sum_{i=1}^{n} d_i^q w_{ii}$ and $d_i^q = (x_i - x_0)^q$, for $q = 0, 1, 2$. In the case of $q = 1$, the power in d_i^q is omitted. By Equation (4.2), we also see that the derivatives of the local polynomial regression at x_0 are given by $m(x_0) = 0!\beta_0$, $m'(x_0) = 1!\beta_1$ and so forth for polynomials of higher degree. The derivations of the above estimators are provided along with the derivations of the local quadratic (given for $D = 2$) and local cubic (given for $D = 3$) estimators in appendix D.1.

4.2.1 Kernel Function

The $K_h(x_i - x_0)$ function is called kernel function and it is a non-negative function which is defined by:

$$K_h(x_i - x_0) = \frac{K(\frac{x_i - x_0}{h})}{h}$$

where h is a positive constant whose definition is given in the subsection 4.2.2. The function K can take different forms according to how we wish to allocate the weights around a specified local neighbourhood. For example, the weights are equally distributed across the observations in this neighbourhood for the uniform kernel function. For the Gaussian kernel, the weights are allocated according to the distance of each observation to x_0. So, the smaller the distance of a particular observation to x_0, the higher its contribution. In Table 4.1, we provide the form of the most popular kernel functions and Figure 4.1 depicts how the weights are distributed around the given point x_0 for these kernel functions.

Kernel Function:	Function:
Uniform	$K(u) = \frac{1}{2} I_{(\lvert u \rvert \leq 1)}$
Gaussian	$K(u) = \frac{1}{\sqrt{2\pi}} e^{-u^2/2}$
Epanechnikov	$K(u) = \frac{3}{4}(1-u^2) I_{(\lvert u \rvert \leq 1)}$
Biweight	$K(u) = \frac{15}{16}(1-u^2)^2 I_{(\lvert u \rvert \leq 1)}$
Triweight	$K(u) = \frac{35}{32}(1-u^2)^3 I_{(\lvert u \rvert \leq 1)}$

Table 4.1: Common kernel functions.

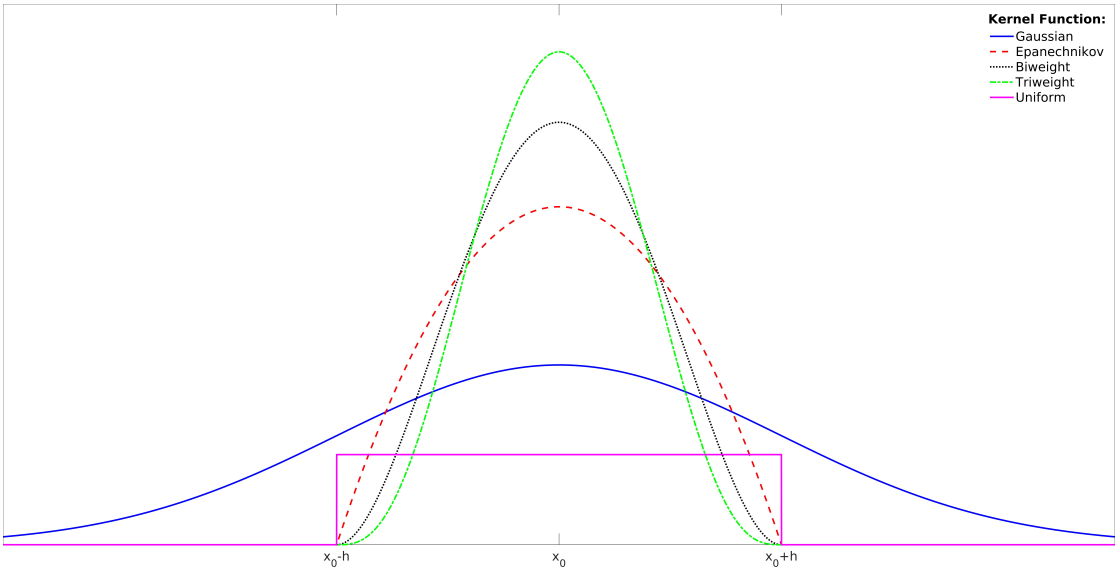

Figure 4.1: This figure illustrates how the weights are allocated around a given point x_0 for the Gaussian (solid blue line), Epanechnikov (dashed red line), biweight (dotted black line), triweight (dash-dotted green line) and the uniform kernel function (solid magenta line).

This study considers kernel functions with the following properties. They are non-negative continuous functions, and hence Probability Density Functions (PDFs) where $\int_{-\infty}^{+\infty} K(u)du = 1$. Furthermore, they are symmetric functions $K(u) = K(-u)\ \forall u \in \mathbb{R}$ and their two first moments satisfy:

$$\int_{-\infty}^{+\infty} uK(u)du = 0 \quad \text{and} \quad \int_{-\infty}^{+\infty} u^2 K(u)du \neq 0.$$

The kernel functions in Table 4.1 are called second-order kernels since the first non-zero moment is the second moment.

Higher-order kernels emerge when the first non-zero moment is observed for higher moments. The main trait of higher-order kernels is that they permit negative weights to the most distant observations from x_0 in the specified neighbourhood. In this way, a bias correction is achieved where sometimes this correction can be remarkable. However, Marron (1994) shows that a substantial improvement may depend on the kind of curve one wants to estimate and there is a cost in the interpretability (the higher-order kernel ceases being a PDF) because of the negative weights. In addition, the same author states that in the context of local polynomial regression, we can eliminate bias by considering a higher degree polynomial for the local approximation. By the term kernel functions, we exclusively refer to second-order kernel functions in this study and the form of the kernel function is specified in the empirical applications.

Nevertheless, the selection of the kernel function does not play a crucial role in the quality of the estimation. On the contrary, determining the optimal range for the local neighbourhood is often essential for the quality of the analysis. This range is given by the smoothing parameter h (also called bandwidth), and its selection ultimately controls the number of observations included in the understudy neighbourhood $x_0 \pm h$. So, for a relatively small bandwidth value, the neighbourhood will consist of a small number of observations, delivering a significant variance in the estimation. On the other hand, a relatively high bandwidth value will decrease the variance, producing, inevitably, considerable bias. Hence, we have to select the proper bandwidth to offset the balance between variance and bias optimally. This topic is discussed in the following subsection.

4.2.2 Bandwidth Selection

The estimation of an ideal bandwidth is of great importance. The bandwidth h essentially controls the neighbourhood range in which the weights of a kernel function

are allocated around a specific point x_0. A relatively high bandwidth sets a wide neighbourhood around x_0, including many observations in this neighbourhood. This means that many data points will contribute to the point approximation at x_0. In this case, we observe an oversmoothed estimation, portrayed as a curve with high bias but low variance. Similarly, a relatively low bandwidth considers a short neighbourhood around x_0, embedding an insufficient number of observations. In this case, we notice an undersmoothed estimation, given as an estimation with low bias but great variance. It is straightforward that as the sample size increases ($n \to \infty$), the local neighbourhood range should decrease ($h \to 0$), so it contains an analogous number of observations in this local window.

In Figure 4.2, we plot the intraday prices (blue curve) of Cleveland-Cliffs Inc. (CLF) for the date 03 January 2012, along with the local linear fit for three different bandwidth values. In particular, a low bandwidth value (red curve), a mid-range bandwidth (green curve) and a high bandwidth value (magenta curve) are considered in the local estimator, presenting from undersmoothed to oversmoothed estimation. In this figure, we notice that the bandwidth value can play a central role in the estimation quality.

Choosing a suitable bandwidth is a demanding task. An efficient way to choose the optimal bandwidth value is by minimising the Mean Squared Error (MSE), expressed as a function of a bias and a variance term:

$$\begin{aligned} MSE(x_0; h) &= \mathbb{E}\left[\left(\widehat{m}^{(k)}(x_0; h) - m^{(k)}(x_0)\right)^2 \Big| x_1, ..., x_n\right] \\ &= Bias^2(x_0; h, x_1, ..., x_n) + Var[x_0; h, x_1, ..., x_n] \end{aligned}$$

where $\widehat{m}^{(k)}(x_0; h)$ indicates the estimated k-th derivative of the regression function at x_0 given the bandwidth h and x_i denotes the i-th data point. The above derivation of MSE is given in appendix D.2.

MSE provides a picture of the estimator's performance at a local point x_0 for a

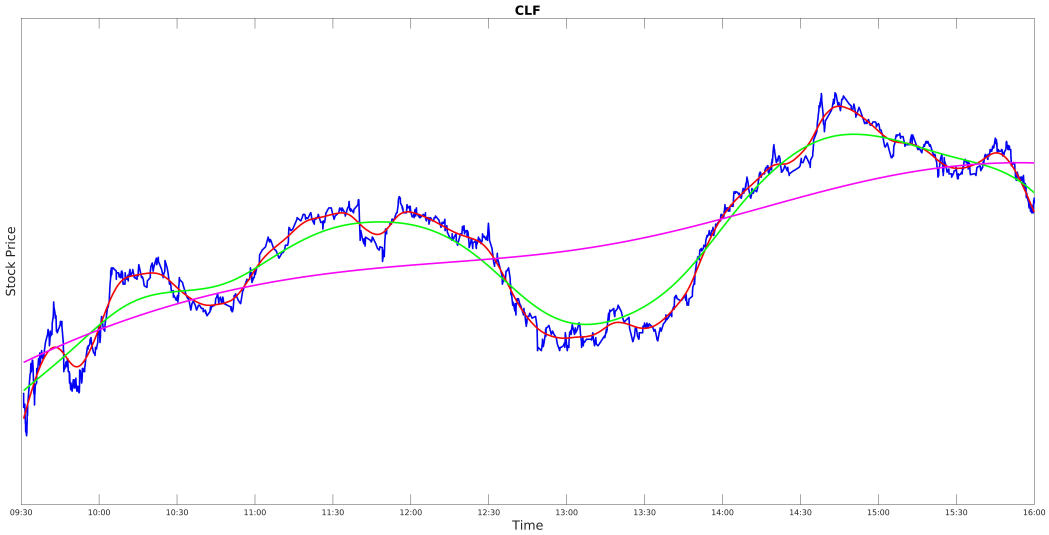

Figure 4.2: This figure depicts the intraday prices (blue curve) of Cleveland-Cliffs Inc. (CLF) over the date 03 January 2012, along with the local linear fit using a low bandwidth value (red curve), a mid-range bandwidth value (green curve) and a high bandwidth value (magenta curve).

specified bandwidth h. In the context of global optimal bandwidth selection, the Mean Integrated Squared Error (MISE) is a more appropriate measure to deal with this problem globally, intending to utilise the properties of MSE for the whole data curve. MISE also considers an additional non-negative weight component, referring to the observations' density. Consequently, the bandwidth selection for MISE is given by minimising the following expression for different h values:

$$MISE(h) = \int MSE(x; h)w(x)dx.$$

Here, $w(x)$ stands for the non-negative weight component, which is used to erase the boundary errors. The boundary errors refer to the misestimation that often occurs at the beginning and the end of the regression function because of the low number of data that contribute to the estimation in these areas. So, one can edit out this component when he refers to the middle part of the data curve.

A method that approximates the above expression is Cross-Validation (CV) (Clark,

1977). In a nutshell, the procedure of this method goes as follows. We remove an observation or a part of observation values (test data) from the dataset. Then, using the remaining data (train data), we attempt to estimate the test data for a grid of bandwidth values through LPR. The optimal bandwidth is the value h that minimises the prediction error. We can express CV as:

$$CV(h) = \frac{1}{n}\sum_{i=1}^{n}(\widehat{m}_{(-i)}(x_i;h) - y_i)^2.$$

where y_i stands for the i-th observed value for $i = 1, ..., n$. Moreover, $\widehat{m}_{(-i)}(x_i;h)$ denotes the estimated regression function where the i-th observation or part of observations have been removed from the dataset. In appendix D.3, we show that for the regression function ($k = 0$), we obtain:

$$\begin{aligned}\mathbb{E}[CV(h)] &= \mathbb{E}\left[\frac{1}{n}\sum_{i=1}^{n}\Big(\widehat{m}_{(-i)}(x_i;h) - m(x_i)\Big)^2\right] + \frac{1}{n}\sum_{i=1}^{n}\sigma^2(x_i) \\ &\approx \int_0^1 \mathbb{E}\left[\Big(\widehat{m}(x;h) - m(x)\Big)^2\right]dx + \widetilde{\sigma^2}\end{aligned}$$

which is nothing more than an estimator of the MISE in discrete time with an additional constant term $\widetilde{\sigma^2}$. The term $\widetilde{\sigma^2}$ of the above quantity does not considerably affect the quality of the estimation (Bowman and Azzalini, 1997) since it is not a function of h. Therefore, the minimisation of MISE can be considered equivalent to the minimisation of CV.

In the applications of this study, the k-fold CV is conducted for choosing the optimal h. The difference between the standard CV and the k-fold CV is that the latter method partition the observations into a number of groups of approximately equal size, accelerating the estimation procedure. One group is considered as test dataset and the rest are considered training datasets. Then, we use the training datasets to estimate the regression function using the time points of the test dataset in the Matrices (4.4) and (4.5).

The optimal h is chosen to be this value which allows the lowest possible discrepancy between the test dataset and the regression function as derived by the training dataset across the groups.

The under examination bandwidths are given as some grid points of the interval $[h_{min}, h_{max}]$. In practice, the bandwidth h_{min} can be set as that value that takes at least five observations in the specified neighbourhood for each point. On the other hand, h_{max} can take a value that includes a large number of observations within the local neighbourhood. Intuitively, we can set a value for h_{max} so that the point approximation will take place using roughly one-third of the total observations on each side in the local neighbourhood. The steps of the CV are summarised in ALGORITHM 1.

In the context of linear fitting models, the Generalised Cross-Validation (GCV) is an alternative estimator of MISE. GCV is proposed by Craven and Wahba (1978) and it is defined by:

$$GCV(h) = \frac{1}{n}\sum_{i=1}^{n}\left(\frac{y_i - \widehat{m}(x_i;h)}{1 - trace(S)/n}\right)^2$$

where S indicates the global smoothing matrix for a given bandwidth h. The S matrix is related to the concept of the local smoothing matrix. In particular, the local smoothing matrix at x_0 for a bandwidth h, $S(x_0;h)$, is derived from:

$$\widehat{m}(x_0;h) = X_{x_0}\widehat{\beta}_{x_0} = X_{x_0}(X_{x_0}^T W_{x_0} X_{x_0})^{-1} X_{x_0}^T W_{x_0} Y = S(x_0;h)Y$$

where $S(x_0;h) = X_{x_0}(X_{x_0}^T W_{x_0} X_{x_0})^{-1} X_{x_0}^T W_{x_0}$. For example, evaluating at x_1 gives $\widehat{m}(x_1;h) = S(x_1;h)Y$. The matrix $S(x_1;h)$ is of size $n \times n$ and the first row of this matrix is called effective kernel. Likewise, using x_2 as evaluation point, then the effective kernel is the second row of the local smoothing matrix $S(x_2;h)$. In like manner, the i-th row of the matrix $S(x_i;h)$ is the effective kernel for the evaluation point x_i. Collecting the n effective kernels in a new matrix $S(x;h)$, then this matrix is called global smoothing matrix. The definition and the derivation of the local and the global

Algorithm 1 k-fold Cross-Validation for Bandwidth Estimation in LPR

- Specify $[h_{min}, h_{max}]$;
- Divide the dataset into G groups;

for h_k, where $k = 1, ..., K$ **do**:

for group g, where $g = 1, ..., G$, **do**:

- Remove the group g from X and Y (test data) and keep the rest (train data);
- Specify the evaluation points of the regression function, as given by the test data of X;
- Estimate the regression function using the bandwidth h_k and the train data at the particular evaluation points;
- Calculate the Prediction Error:

$$PE_g(h_k) = mean\left(\left(Y^{(test)} - \widehat{m}^{(-g)}\right)^2\right)$$

where $Y^{(test)}$ denotes the test dataset of Y and $\widehat{m}^{(-g)}$ denotes the regression function where the group g has been removed;

end for

- Calculate the mean error for the given h_k by:

$$CV(h_k) = \frac{1}{G}\sum_{g=1}^{G} PE_g(h_k);$$

end for

- The optimal tuning parameter is chosen by:

$$\widehat{h_{CV}} = \arg\min_{h_k>0} CV(h_k).$$

smoothing matrices are given in appendix D.1 for the local smoothers with polynomial degree $D = 0, \ldots, 3$.

Thus, we end up that the optimal bandwidth h for the CV and GCV score functions is given by:

$$\widehat{h_{CV}} = \arg\min_{h_k>0} CV(h_k),$$

$$\widehat{h_{GCV}} = \arg\min_{h_k>0} GCV(h_k).$$

where $k = 1, ..., K$.

Simonoff (2012) points out that both the CV and the GCV estimators often choose bandwidths that lead to undersmoothed estimations, whilst Hastie et al. (2009) remark that this event is more frequent for CV. Nevertheless, choosing the proper bandwidth is a demanding task. Although one follows the methods mentioned above, the estimation can still perform poorly. Seasonality or autocorrelation between the data series can lead to bandwidth misestimation (Opsomer et al., 2001; De Brabanter et al., 2018). Alternative approaches for optimal bandwidth selection have been proposed in the literature. Some studies develop the theoretical framework for a locally adaptive bandwidth selection. These studies treat the bandwidth as a variable, allowing the bandwidth to change across the data points.

Müller and Stadtmüller (1987), Staniswalis (1989), Fan and Gijbels (1996), Lepski et al. (1997) are some popular studies which propose a methodology for the estimation of local bandwidths. The reader can refer to Köhler et al. (2014) who provide an extensive overview and comparison of the most popular bandwidth selection practices.

4.3 LPR-Based Estimator of IV

In this section, we develop an empirical estimator of IV, along with its relevant statistical framework. In the proposed approach, we utilise the mean function of the pointwise returns (in their standard and squared form) for the estimation. LPR, as defined in section 4.2, is the method used to approximate the mean functions; hence we refer to this measure as LPR-based estimator of IV.

LPR is a method which estimates the observations' mean value in continuous time based on a weighting allocation scheme around a specified local neighbourhood. Several daily volatility estimators (Andersen et al., 2001a; Podolskij et al., 2009) assume equidistant observed time points in order to be defined. Although, this is not a damaging assumption in terms of the estimation procedure it requires pre-processing to ensure that every intraday interval contains sufficient (non-zero) number of observations and

the volatility is defined at the predefined time points. The continuous trading is more realistically represented by non-equidistant observed time points and such restriction is not necessary under the continuous function estimation framework. In the LPR-based estimator, we can specify the range of the neighbourhood in order to contain enough observations for the volatility estimation. On the other hand, boundary errors can affect the precision of the LPR-based estimator.

We consider the same theoretical framework as described in Chapter 2, including all the assumptions made there. Recall that:

$$X(t_i) - X(t_{i-1}) = \int_{t_{i-1}}^{t_i} \mu(\varsigma)d\varsigma + \int_{t_{i-1}}^{t_i} \sigma(\varsigma)dW(\varsigma).$$

In section 2.2, μ and σ are supposed to be random processes. In addition, we assume that they are smooth functions in this chapter.

Assumption 4.3.1 *The coefficients $\mu(t)$ and $\sigma(t)$ are continuously differentiable and the derivatives of the sample paths are uniformly Lipschitz continuous: there exists a constants $L > 0$ such that a.s.*

$$|\mu'(s) - \mu'(t)| \leq L|s-t|\,, \quad |\sigma'(s) - \sigma'(t)| \leq L|s-t| \qquad \textit{for all } s,t \in \mathcal{I} = [0,1]\,.$$

In addition, there exists a constant C such that

$$\sup_{t\in\mathcal{I}} |\mu^{(\nu)}(t)| < C\,, \quad \sup_{t\in\mathcal{I}} |\mu^{(\nu)}(t)| < C\,, \ \textit{for } \nu = 0,1\,.$$

Similar smoothness conditions are considered in Müller et al. (2006), Fan and Wang (2008), Kristensen (2010) and Müller et al. (2011). It is possible to make a weaker assumption with (Hölder) continuity of order $0 < \alpha < 1$ or consider some specific classes of the models for σ to refine the order of smoothness but for simplicity we stay with this assumption. From now on, our derivations are understood as being conditional on

μ and σ. Effectively, they are treated as if they are fixed functions.

The i-th observed log-price $Y(t_i)$ of a stock at time t_i is defined by:

$$Y(t_i) = X(t_i) + E(t_i)$$

where $X(t_i)$ denotes the i-th intraday true log-price, which is unknown and follows a stochastic process given by Equation (2.1). Furthermore, $E(t_i)$ displays the corresponding error term. Consequently, the i-th intraday return is defined by:

$$R(t_i) = R^*(t_i) + V(t_i)$$

where the return $R(t_i) = Y(t_i) - Y(t_{i-1})$ and $R^*(t_i)$ stands for the corresponding true return, given by:

$$R^*(t_i) = \int_{t_{i-1}}^{t_i} \mu(\varsigma)d\varsigma + \int_{t_{i-1}}^{t_i} \sigma(\varsigma)dW(\varsigma)$$

with $\mathbb{E}[R^*(t_i)] = \int_{t_{i-1}}^{t_i} \mu(\varsigma)d\varsigma$ and $Var[R^*(t_i)] = \int_{t_{i-1}}^{t_i} \sigma^2(\varsigma)d\varsigma$. Besides, the term $V(t_i)$ indicates the difference between the error terms $E(t_i)$ and $E(t_{i-1})$. By taking the expectation of the squared returns and using Assumption 2.2.4, we have that:

$$\mathbb{E}[R^2(t_i)] = \mathbb{E}\left[\left(R^*(t_i)\right)^2\right] + \mathbb{E}[V^2(t_i)] = Var[R^*(t_i)] + (\mathbb{E}[R^*(t_i)])^2 + 2\,\mathbb{E}[E^2(t_i)].$$

Observe that $\mathbb{E}[R^*(t_i)] = \mathbb{E}[R(t_i)]$. Thus, we conclude with the following local volatility relation:

$$\int_{t_{i-1}}^{t_i} \sigma^2(\varsigma)d\varsigma = \mathbb{E}[R^2(t_i)] - (\mathbb{E}[R(t_i)])^2 - 2\,\mathbb{E}[E^2(t_i)]. \tag{4.6}$$

An approximation of the local volatility can be obtained by estimating the right-hand side terms of Equation (4.6). An estimator of the term $\mathbb{E}[R^2(t_i)]$ is given by the mean function of the squared returns. Similarly, an estimation of the term $\mathbb{E}[R(t_i)]$ is provided

by estimating the mean function of the observed returns. For the estimation of the mean functions $\mathbb{E}[R^2(t_i)]$ and $\mathbb{E}[R(t_i)]$, we apply LPR to the squared and the standard returns, respectively. The term $\mathbb{E}[E^2(t_i)]$ defines the noise variance ω^2. For the estimation of ω^2, several common estimators are outlined in section 4.3.1. Hence, a natural estimator of the local volatility by direct application of moment matching may be defined as:

$$\int_{t_{i-1}}^{t_i} \sigma^2(\varsigma)d\varsigma \approx \widehat{m_{R^2}}(t_i) - \left(\widehat{m_R}(t_i)\right)^2 - 2\widehat{\omega^2} \tag{4.7}$$

where $\widehat{m_{R^2}}(t_i)$ denotes the estimated mean function of the pointwise squared returns $R^2(t_i)$, $\widehat{m_R}(t_i)$ indicates the estimated mean function of the pointwise returns $R(t_i)$ and $\widehat{\omega^2}$ is the noise variance estimate. This defines our proposed estimator. The approximation error will be made more precisely by studying the asymptotic properties of these estimators later in section 4.3.2. The impact of the error correction will depend on the magnitude of the error ω^2 and this will be investigated in the simulation study as well. The daily IV is defined by:

$$IV = \int_0^1 \sigma^2(\varsigma)d\varsigma$$

where the limits of integration denote one day period. Exploiting the fact that we can re-write the above integral as the summation of integrals using the interior points of the interval $[0, 1]$ as limits, we end up that a measure of IV is provided by aggregating the local volatility estimates of Equation (4.7):

$$\widehat{IV} = \sum_{i=2}^{n} \left(\widehat{m_{R^2}}(t_i) - \left(\widehat{m_R}(t_i)\right)^2 - 2\widehat{\omega^2} \right).$$

In Chapter 2, we have defined that t_i represents the time point of the i-th observed price in the trading session. Assuming n observed prices within a day, this impies that we obtain $n-1$ pointwise returns. So, the functions $\widehat{m_{R^2}}(t_i)$ and $\widehat{m_R}(t_i)$ would be more meaningful to start from the time point t_2 in the mathematical representation above.

4.3.1 Estimation of Noise Variance

For the estimation of the noise variance ω^2, Zhang et al. (2005) propose the following estimator:

$$\widehat{\omega^2} = \frac{RV^{(all)}}{2(n-1)} \xrightarrow{p} \omega^2, \qquad \text{as} \quad n \to \infty \tag{4.8}$$

where the derivation of this estimator comes from the theoretical properties of $RV^{(all)}$, given the increments of the true price process. Besides, Zhang et al. (2005) on page 1398 further prove that:

$$n^{1/2}(\widehat{\omega}^2 - \omega^2) \to N(0, \mathbb{E}[E^4])\,, \tag{4.9}$$

where

$$\widehat{\mathbb{E}[E^4])} = \frac{\sum_{i=1}^{n}(R_{t_i})^4}{2(n-1)} - 3\Big(\widehat{\omega^2}\Big)^2.$$

Two alternative estimators of ω^2 suggested by Hansen and Lunde (2006). The first estimator has the following form:

$$\widehat{\omega^2} = \frac{RV^{(all)} - RV^{(30-\text{min})}}{2\Big((n-1) - 13\Big)} \tag{4.10}$$

where $RV^{(30-\text{min})}$ denotes the RV estimator using 30-minute returns and the term '13' in the denominator of (4.10) comes from the 13 equilength and non-overlapping intraday intervals which yield setting 30-minute returns. The second estimator is a generalisation of the Estimator (4.10), given by:

$$\widehat{\omega^2} = \frac{RV^{(all)} - \widehat{IV}}{2(n-1)} \tag{4.11}$$

where $\widehat{IV}$ depicts any alternative measure of IV. Hansen and Lunde (2006) show that even if all of the aforementioned estimators are identical as regards the probability limit for $n \to \infty$, the Estimator (4.8) can be biased for finite n.

4.3.2 Asymptotic Properties

Recall that the estimator of the local volatility is defined as $\widehat{m}(t_i) = \widehat{m_{R^2}}(t_i) - \left(\widehat{m_R}(t_i)\right)^2 - 2\widehat{\omega^2}$, which are constructed from two regression models

$$R_i = m_R(t_i) + \phi_i\,, \quad R_i^2 = m_{R^2}(t_i) + \eta_i\,,$$

where the errors ϕ_i and η_i are assumed to have mean zero and finite variance. We consider that the design points of the regression are fixed but satisfy the following standard assumption (Mykland and Zhang, 2009; Nolte and Voev, 2012).

Assumption 4.3.2 *For the observed time points $0 \leq t_1 < \cdots < t_n \leq 1$, we have that:*

$$\max_{2\leq i\leq n} |t_i - t_{i-1}| = O\Big(1/n\Big)$$

as $n \to \infty$.

Denote by h_1 and h_2 the bandwidths applied to the returns and the squared returns by the LPR, respectively. Using the properties of the local polynomial smoothers, we derive the asymptotic properties of the estimators below.

Lemma 4.3.1 *(Properties of the Local Linear Estimator of $m(x)$ - Fan and Gijbels (1996)) Suppose that $(X_i, Y_i), i = 1, \ldots, n$ are IID copies of (X, Y) and $Y = m(x)+\sigma(x)\epsilon$ where $m(x) = \mathbb{E}[Y|X = x]$, $\sigma^2(x) = Var[Y|X = x]$ and ϵ has mean 0 and variance 1. Assume that $X \sim f(x) > 0$ and that $f(\cdot)$, $m''(\cdot)$ and $\sigma^2(\cdot)$ are continuous in a neighborhood of x. Further, assume that $h \to 0$ and $nh \to \infty$. Then, for the point approximation with the local linear smoother the asymptotic conditional bias is given by:*

$$aBias\Big(\widehat{m}(x)\Big) = \frac{h^2}{2} k_2 m''(x) + o_p(h^2)\,,$$

and the asymptotic conditional variance is given by:

$$aVar\Big[\widehat{m}(x)\Big] = \frac{\nu_0 \sigma^2(x)}{nhf(x)} + o_p\Big(\frac{1}{nh}\Big)\,,$$

where K denotes the kernel function, h stands for the bandwidth, $\nu_0 = \int K^2(u)du$ and $k_2 = \int u^2 K(u)du$ denotes the second-order moment of the kernel function K.

The same result holds for stationary correlated errors as long as the correlation decays fast enough (Simonoff, 2012; De Brabanter et al., 2018). This is the case in our setting as the correlation of the errors is zero beyond finite lags, as shown in appendix B.1.1.

In order to apply the results of Lemma 4.3.1 to our setting, we need to verify that

$$m''_R(t) = O_p(1)\,, \quad m''_{R^2}(t) = O_p(1)\,,$$

and

$$\sigma^2_R(t) = O_p(1)\,, \quad \sigma^2_{R^2}(t) = O_p(1)\,.$$

Assumption 4.3.1 is sufficient to guarantees this. Below we derive the bias and variance for our combined estimator, conditional on μ and σ, and we make the following assumptions on the bandwidths.

Assumption 4.3.3 *As $n \to \infty$, $h_1 \to 0$ and $h_2 \to 0$ and $nh_1 \to \infty$ and $nh_2 \to \infty$.*

Assumption 4.3.4 *Assume that $h_1 = O(n^{-1/5})$ and h_1 and h_2 are of the same order.*

Assumption 4.3.3 is a common assumption in this context (Fan and Gijbels, 1996; Opsomer et al., 2001; De Brabanter et al., 2018). In addition, the bandwidths h_1 and h_2 are assumed by Assumption 4.3.4 to be of the same order and h_1 is of order $O(n^{-1/5})$, which is optimal in the sense of asymptotic MSE (Fan and Gijbels, 1996). The latter assumption is not essential but can simplify the derivations as we show below.

To consider at arbitrary time point t, we assume that the gap between transaction times, $t_i - t_{i-1}$, is independent of t_i and approximately uniform, $t_i - t_{i-1} = \delta_i \approx \delta$. For simplicity, let $\delta_i = \delta$ and we write $R^*(t) = X(t) - X(t-\delta) \sim N\Big(r(t), v(t)\Big)$ where:

$$r(t) = \int_{t-\delta}^{t} \mu(\varsigma) d\varsigma, \quad v(t) = \int_{t-\delta}^{t} \sigma^2(\varsigma) d\varsigma\,,$$

and $m(t) = v(t)$. For correspondence to the Equations in appendix D.4, $v(t_i) = s_i^2$.

Asymptotic Bias

Moment properties of returns are summarized in the appendix section D.4. From (D.1) and (D.2), the mean of the returns and that of the squared returns are:

$$m_R(t) = \int_{t-\delta}^{t} \mu(\varsigma) d\varsigma = r(t), \quad m_{R^2}(t) = \int_{t-\delta}^{t} \sigma^2(\varsigma) d\varsigma + \Big(\int_{t-\delta}^{t} \mu(\varsigma) d\varsigma \Big)^2 + 2\omega^2.$$

The bias of $\widehat{m}$ is given by:

$$\begin{aligned} \mathbb{E}[\widehat{m}(t) - m(t)] &= \mathbb{E}\left[\widehat{m_{R^2}}(t) - \Big(\widehat{m_R}(t)\Big)^2 - 2\widehat{\omega^2} - \Big(m_{R^2}(t) - \Big(m_R(t)\Big)^2 - 2\omega^2\Big)\right] \\ &= \mathbb{E}\left[\widehat{m_{R^2}}(t) - m_{R^2}(t)\right] - \mathbb{E}\left[\Big(\widehat{m_R}(t)\Big)^2 - \Big(m_R(t)\Big)^2\right] - 2\,\mathbb{E}\left[\widehat{\omega^2} - \omega^2\right] \end{aligned} \tag{4.12}$$

where the noise does not contribute to smoothing bias, see (4.9). To express the second term, we observe that:

$$\begin{aligned} \mathbb{E}\left[\Big(\widehat{m_R}(t)\Big)^2\right] &= \mathbb{E}\left[\Big(\widehat{m_R}(t) - m_R(t) + m_R(t)\Big)^2\right] \\ &= \mathbb{E}\left[\Big(\widehat{m_R}(t) - m_R(t)\Big)^2 + m_R^2(t) + 2\Big(\widehat{m_R}(t) - m_R(t)\Big) m_R(t)\right] \end{aligned}$$

so the second term is given by:

$$\mathbb{E}\left[\Big(\widehat{m_R}(t)\Big)^2 - m_R^2(t)\right] = \mathbb{E}\left[\Big(\widehat{m_R}(t) - m_R(t)\Big)^2\right] + 2m_R(t)\,\mathbb{E}\left[\Big(\widehat{m_R}(t) - m_R(t)\Big)\right]$$

$$= Var\Big[\widehat{m_R}(t) - m_R(t)\Big] + \Big(\mathbb{E}[\widehat{m_R}(t) - m_R(t)]\Big)^2 + 2m_R(t)\,\mathbb{E}[\widehat{m_R}(t) - m_R(t)].$$

Plugging the quantity above into (4.12) gives:

$$\begin{aligned}\mathbb{E}[\widehat{m}(t) - m(t)] &= \mathbb{E}[\widehat{m_{R^2}}(t) - m_{R^2}(t)] - Var[\widehat{m_R}(t) - m_R(t)] \\ &\quad - \Big(\mathbb{E}[\widehat{m_R}(t) - m_R(t)]\Big)^2 - 2m_R(t)\,\mathbb{E}[\widehat{m_R}(t) - m_R(t)] - 2\,\mathbb{E}\left[\widehat{\omega^2} - \omega^2\right].\end{aligned} \tag{4.13}$$

From Lemma 4.3.1, the leading terms in asymptotic bias are:

$$aBias\Big(\widehat{m_{R^2}}(t)\Big) = \frac{h_2^2}{2}k_2 m''_{R^2}(t)\,, \quad aBias\Big(\widehat{m_R}(t)\Big) = \frac{h_1^2}{2}k_2 m''_R(t)\,,$$

and the leading term in asymptotic variance of $\widehat{m_R}(t)$ is:

$$aVar\Big[\widehat{m_R}(t)\Big] = \frac{\nu_0\sigma_R^2(t)}{(n-1)h_1 f_T(t)} \tag{4.14}$$

where $\sigma_R^2(t) = Var[R(t)] = m(t) + 2\omega^2$ (see (D.3) in appendix D.4) and $f_T(t) = 1$. Thus, the asymptotic bias of $\widehat{m}$ in (4.13) can now be expressed as:

$$\begin{aligned}&aBias\Big(\widehat{m}(t)\Big) = aBias\Big(\widehat{m_{R^2}}(t)\Big) - 2m_R(t)aBias\Big(\widehat{m_R}(t)\Big) - \Big(aBias(\widehat{m_R}(t))\Big)^2 - aVar\Big[\widehat{m_R}(t)\Big] \\ &= \frac{h_2^2}{2}k_2 m''_{R^2}(t) - h_1^2 k_2 m_R(t) m''_R(t) - \frac{h_1^4}{4}k_2^2\Big(m''_R(t)\Big)^2 - \frac{\nu_0\sigma_R^2(t)}{(n-1)h_1} \\ &= \frac{h_2^2}{2}k_2\Big(m''(t) + 2\Big(m'_R(t)\Big)^2 + 2m_R(t)m''_R(t)\Big) - h_1^2 k_2 m_R(t) m''_R(t) - \frac{h_1^4}{4}k_2^2\Big(m''_R(t)\Big)^2 - \frac{\nu_0\sigma_R^2(t)}{(n-1)h_1} \\ &= \frac{h_2^2}{2}k_2\left[\Big((\sigma^2)'(t) - (\sigma^2)'(t-\delta)\Big) + 2\Big(\mu(t) - \mu(t-\delta)\Big)^2 + 2\Big(\int_{t-\delta}^{t}\mu(\varsigma)d\varsigma\Big)\Big(\mu'(t) - \mu'(t-\delta)\Big)\right] \\ &\quad - h_1^2 k_2\Big(\int_{t-\delta}^{t}\mu(\varsigma)d\varsigma\Big)\Big(\mu'(t) - \mu'(t-\delta)\Big) - \frac{h_1^4 k_2^2}{4}\Big(\mu'(t) - \mu'(t-\delta)\Big)^2 - \frac{\nu_0\int_{t-\delta}^{t}\sigma^2(\varsigma)d\varsigma + 2\nu_0\omega^2}{(n-1)h_1}.\end{aligned}$$

In (4.13), the second and the third term express the MSE of $\widehat{m_R}(t)$. So, the expression can be further simplified if h_1 is chosen optimally, that is, to minimize the asymptotic MSE:

$$aMSE\Big(\widehat{m_R}(t)\Big) = O_p(h_1^4 + (nh_1)^{-1})\,,$$

leading to $h_1 = O(n^{-1/5})$ (see Assumption 4.3.4), in which case $aBias^2 = O_p(n^{-4/5})$ is negligible compared to $aBias = O_p(n^{-2/5})$. Thus, the asymptotic bias of $\widehat{m}$ is simplified to:

$$
\begin{aligned}
&aBias\Big(\widehat{m}(t)\Big) = aBias\Big(\widehat{m_{R^2}}(t)\Big) - 2m_R(t)aBias\Big(\widehat{m_R}(t)\Big) + o_p(1) \\
&= \frac{h_2^2}{2}k_2\Bigg[\Big((\sigma^2)'(t) - (\sigma^2)'(t-\delta)\Big) + 2\Big(\mu(t) - \mu(t-\delta)\Big)^2 + 2\Big(\int_{t-\delta}^{t}\mu(\varsigma)d\varsigma\Big)\Big(\mu'(t) - \mu'(t-\delta)\Big)\Bigg] \\
&\quad - h_1^2k_2\Big(\int_{t-\delta}^{t}\mu(\varsigma)d\varsigma\Big)\Big(\mu'(t) - \mu'(t-\delta)\Big) = O_p(h_2^2\delta + h_1^2\delta^2) = O_p(h_2^2\delta)
\end{aligned}
$$

where the constants involving μ and σ and their derivatives are controlled by the smoothness conditions for σ and μ in Assumption 4.3.1. As the gap between transaction times $\delta = O(n^{-1})$ (see Assumption 4.3.2), if the bandwidth h_2 is chosen to have the same order as h_1 (see Assumption 4.3.4), the bias is dominated by the estimation of the mean of the squared returns. This is consistent with the common assumption that the effect of μ is negligible compared to that of σ (Fan and Wang, 2008; Müller et al., 2011). Consequently, the asymptotic bias of $\widehat{IV}$ is:

$$
\begin{aligned}
&\mathbb{E}\Big[\widehat{IV} - IV\Big] = \mathbb{E}\Big[\sum_{i=2}^{n}\widehat{m}(t_i) - IV\Big] \\
&\approx \frac{k_2h_2^2}{2}\sum_{i=2}^{n}\Bigg[\Big((\sigma^2)'(t_i) - (\sigma^2)'(t_{i-1})\Big) + 2\Big(\mu(t_i) - \mu(t_{i-1})\Big)^2 + 2\Big(\int_{t_{i-1}}^{t_i}\mu(\varsigma)d\varsigma\Big)\Big(\mu'(t_i) - \mu'(t_{i-1})\Big)\Bigg] \\
&\quad - h_1^2k_2\sum_{i=2}^{n}\Big(\mu'(t_i) - \mu'(t_{i-1})\Big)\int_{t_{i-1}}^{t_i}\mu(\varsigma)\,d\varsigma \\
&= O_p(h_2^2n\delta + h_2^2n\delta^2 + n\delta^2h_1^2) = O_p(h_2^2)\,.
\end{aligned}
$$

As discussed above, the leading error is driven by the estimation of the variance term from squared returns.

Asymptotic Variance

To complement the calculation, we can also consider the variance by observing that:

$$\begin{aligned} Var[\widehat{m}] &= Var[\widehat{m_{R^2}} - (\widehat{m_R})^2 - 2\widehat{\omega^2}] \\ &= O\Big(Var[\widehat{m_{R^2}}] + Var[(\widehat{m_R})^2] + 4Var\Big[\widehat{\omega^2}\Big]\Big)\,. \end{aligned} \tag{4.15}$$

Thus, it is sufficient to consider the leading term of the variance of each term. More specifically, the asymptotic variance of $\widehat{m_{R^2}}(t)$ is given by:

$$aVar\Big[\widehat{m_{R^2}}(t)\Big] = \frac{\nu_0 \sigma^2_{R^2}(t)}{nh_2 f_T(t)}$$

where $\sigma^2_{R^2}(t)$ from (D.4) can be expressed as:

$$\begin{aligned} \sigma^2_{R^2}(t) = Var\Big[R^2(t)\Big] &= 2v^2(t) + 4m^2_R(t)v(t) + 8a_2\Big(m^2_R(t) + v(t)\Big) + 2(a_4 + a_2^2) \\ &= 2v^2(t) + 4v(t)(m^2_R(t) + 2a_2) + 8a_2 m^2_R(t) + 2(a_4 + a_2^2) \\ &= 2\left(\int_{t-\delta}^{t} \sigma^2(\varsigma)\, d\varsigma\right)^2 + 4\left(\int_{t-\delta}^{t} \sigma^2(\varsigma)\, d\varsigma\right)\left\{\left(\int_{t-\delta}^{t} \mu(\varsigma)\, d\varsigma\right)^2 + 2a_2\right\} \\ &\quad + 8a_2\left(\int_{t-\delta}^{t} \mu(\varsigma)\, d\varsigma\right)^2 + 2(a_4 + a_2^2) \\ &= 2(a_4 + a_2^2) + 8a_2\left(\int_{t-\delta}^{t} \sigma^2(\varsigma)\, d\varsigma\right) + O_p(\delta^2)\,. \end{aligned}$$

where $a_k = \mathbb{E}[E^k(t_i)]$. Thus,

$$aVar\Big[\widehat{m_{R^2}}(t)\Big] = \frac{2\nu_0}{nh_2}\left[4a_2\int_{t-\delta}^{t} \sigma^2(\varsigma)\, d\varsigma + (a_4 + a_2^2)\right] + O_p\left(\frac{\delta^2}{nh_2}\right)\,. \tag{4.16}$$

The constant term $a_4 + a_2^2$ is due to the error component. If the errors are assumed to be Gaussian, then $a_4 = 3a_2^2$ with $a_2 = \omega^2$ so that the last term is simply $4\omega^4$. Here we assume that the noise variance is constant. In other circumstances, it is possible to consider the case where the noise variance is a function of the sample size (Aït-

Sahalia and Yu, 2009), and then the interplay between the noise variance (ω^2) and the discretization (δ) could be looked into further.

For the second term $Var\Big[\Big(\widehat{m_R}(t)\Big)^2\Big]$ in (4.15), we use delta method, that is, a first-order Taylor approximation of $f(x) = x^2$, together with (4.14) to obtain:

$$\begin{aligned} Var\Big[\Big(\widehat{m_R}(t)\Big)^2\Big] &\approx Var\Big[\,\mathbb{E}[\widehat{m_R}(t)] + 2\,\mathbb{E}[\widehat{m_R}(t)](\widehat{m}_R(t) - \mathbb{E}[\widehat{m_R}(t)])\Big] \\ &= 4\Big(\,\mathbb{E}[\widehat{m_R}(t)]\Big)^2 Var[\widehat{m_R}(t)] = 4\Big[m_R(t) + Bias\Big(\widehat{m}_R(t)\Big)\Big]^2 Var[\widehat{m_R}(t)]\,. \end{aligned}$$

It follows that:

$$\begin{aligned} &aVar\Big[\Big(\widehat{m_R}(t)\Big)^2\Big] \approx 4Var[\widehat{m_R}(t)]\Big[m_R(t) + aBias\Big(\widehat{m}_R(t)\Big)\Big]^2 \\ &= \frac{4\nu_0\left(\int_{t-\delta}^t \sigma^2(s)\,ds + 2a_2\right)}{nh_1}\left[\int_{t-\delta}^t \mu(s)\,ds + \frac{h_1^2\kappa_2\{\mu'(t) - \mu'(t-\delta)\}}{2}\right]^2 \\ &\le \frac{8\nu_0\left(\int_{t-\delta}^t \sigma^2(s)\,ds + 2a_2\right)}{nh_1}\left[\left(\int_{t-\delta}^t \mu(s)\,ds\right)^2 + \frac{h_1^4\kappa_2^2}{4}\left\{\mu'(t) - \mu'(t-\delta)\right\}^2\right] \\ &= \frac{8\nu_0}{nh_1}\left(\int_{t-\delta}^t \sigma^2(s)\,ds + 2a_2\right)O_p(\delta^2 + h_1^4\delta^2) \\ &= O_p\left(\frac{\delta^2 + \delta^3 + h_1^4\delta^2}{nh_1}\right) = O_p\Big(\frac{\delta^2}{nh_1}\Big). \end{aligned} \tag{4.17}$$

From (4.9), $\widehat{\omega}^2$ is a $\sqrt{n}$-consistent estimator so $aVar\Big[\widehat{\omega}^2\Big] = O(n^{-1})$. Combining (4.16) and (4.17) gives:

$$aVar\Big[\widehat{m}(t)\Big] = \frac{2\nu_0}{nh_2}\left[4a_2\int_{t-\delta}^t \sigma^2(\varsigma)\,d\varsigma + (a_4 + a_2^2)\right] + O_p\left(\frac{\delta^2}{nh_2}\right) + O_p\left(\frac{\delta^2}{nh_1}\right) + O_p\left(\frac{1}{n}\right).$$

Similar to bias, the variance is also dominated by the estimation of the mean of the squared returns. Consequently, the asymptotic variance of $\widehat{IV}$ is dominated by:

$$Var[\widehat{IV}] = Var\Big[\sum_{i=2}^n \widehat{m}(t_i)\Big] \approx \frac{8\nu_0 a_2\int_0^1 \sigma^2(s)\,ds + 2\nu_0 n(a_4 + a_2^2)}{nh_2} + O_p(1)\,.$$

where the latter follows from $\delta = O(n^{-1})$ (see Assumption 4.3.2). Although it is odd to have increasing variance, this is natural because the estimator $\widehat{IV}$ is defined as the sum of the individual estimator. Perhaps our approximation is rather crude as the effect of the noise variance is not cancelled out but the asymptotic analysis suggests that h_2 should be relatively large to attenuate the effect of noise variance. Here, we assume that the noise variance is constant independent of the sample size. In practice, the noise variance and noise-to-signal ratio tend to be lower for highly traded stocks (Aït-Sahalia and Yu, 2009). Then, one could consider the case where the noise variance decreases as a function of n with a certain order.

4.4 Standard IV Estimators

In this section, we report three standard IV measures under the presence of microstructure noise. These estimators are utilised in the simulation study of this chapter for comparison purposes. More specifically, the two-scales RV estimator, the realised kernel estimator and the pre-averaging approach are briefly explained where the first estimator converges to the rate $n^{-1/6}$, whereas the latter two measures achieve the optimal rate of convergence $n^{-1/4}$. For the presented estimators, we assume the same theoretical framework described in section 2.2.

4.4.1 Two-Scales RV

The Two-Scales RV estimator (TSRV) was proposed by Zhang et al. (2005). Assuming that we observe a new transaction at time points $0 < t_1 < ... < t_i < ... < t_n \leq T$, where $i = 1, ..., n$, then the TSRV model is defined by:

$$TSRV = RV^{(avg)} - \frac{\overline{g}}{n-1} RV^{(all)}$$

and

$$RV^{(avg)} = \frac{1}{\ell}\sum_{i=1}^{\ell}\sum_{k=1}^{g_i}\Big(Y(t_{i+k\ell}) - Y(t_{i+(k-1)\ell})\Big)^2$$

where $RV^{(all)}$ represents the Estimator (2.4) and $Y(t_{i+k\ell})$ and $Y(t_{i+(k-1)\ell})$ show the log-prices that are placed ℓ lags away. Further, g_i indicates the highest number such that $i + k\ell < n$ using the i-th point as starting point for $i = 1, ..., \ell$. Also, n denotes the number of the observed prices in a day and:

$$\overline{g} = \frac{1}{\ell}\sum_{i=1}^{\ell} g_i.$$

TSRV is a simple measure of IV and it has been proven superior to the standard RV-based estimators because of the bias correction realised within its application. Zhang et al. (2005) show that TSRV is an unbiased and consistent estimator, reaching a convergence rate equal to $n^{-1/6}$.

4.4.2 Pre-Averaging Approach

The Pre-Averaging (PA) estimator was introduced by Podolskij et al. (2009) and a generalised version of this approach is provided by Jacod et al. (2009). For this technique, we consider that the observed prices occur at equidistant time points $t_i = i/n$ where $i = 1, ..., n$. Then, the PA estimator is defined by:

$$PA(Y) = \sum_{I=1}^{J}\left(\frac{1}{n_I - \ell}\sum_{i=1}^{n_I-\ell}\Big(Y(t_{I,i+\ell}) - Y(t_{I,i})\Big)\right)^2$$

where n_I denotes the number of observed prices within the intraday interval I and J refers to the number of non-overlapping and equilength intraday intervals we set in a day such that $I = 1, ..., J$. Besides, $Y(t_{I,i})$ represents the i-th log-price within the intraday interval I and $Y(t_{I,i+\ell})$ depicts the log-price within this interval, observed ℓ lags later than the log-price $Y(t_{I,i})$.

Also, Podolskij et al. (2009) suggests that J and ℓ can be computed optimally by:

$$J = n/(c_2\ell), \qquad \ell = c_1\sqrt{n}$$

where n denotes the number of the observed prices within the day and c_1, c_2 display two constant values. Assuming that the volatility process σ does not vary through time, then c_1 and c_2 are given by:

$$c_1 = \sqrt{\frac{18}{(c_2-1)(4-c_2)}}\frac{\omega}{\sigma},$$
$$c_2 = 1.6.$$

In their simulation study, Podolskij et al. (2009) mark that we can use $\widehat{IV}$ as a measure of σ in the above formula, however, they specify $c_1 = 1$ and $c_2 = 1.6$ in their simulation study because of the intricacy in the relevant computations.

According to Theorem 1 in Podolskij et al. (2009), the following expression holds:

$$PA_n(Y) \xrightarrow{p} PA(Y) = \frac{1}{c_1c_2}\int_0^1 (\nu_1\sigma_u^2 + \nu_2\omega^2)du$$

where $\nu_1 = \frac{c_1\max(3c_2-4+(2-c_2)^3,0)}{3(c_2-1)^2}$, $\nu_2 = \frac{2\min(c_2-1),1)}{c_1(c_2-1)^2}$. Therefore, we have that:

$$\frac{c_1c_2PA(Y) - \nu_2\omega^2}{\nu_1} \xrightarrow{p} \int_0^1 \sigma_u^2 du$$

Hence, they end up with the Modulated Realised Variance (MRV) estimator of IV:

$$MRV(Y) = \frac{c_1c_2PA(Y) - \nu_2\omega^2}{\nu_1}$$

which is a superior estimator compared to PA. We can estimate the noise variance ω^2 through the estimators in subsection 4.3.1.

The advantages of this technique are the following. Using a modified form of the PA estimator, we can obtain alternative measures of volatility. Also, for the IV estimation MRV achieves the optimal rate of convergence $n^{-1/4}$.

Although the magnitude of the noise is mitigated through this approach, the pre-averaging scheme may also induce some noise (Jacod et al., 2009). In addition, this estimator is defined for regularly spaced data. In the real world, the investigator may confront computational obstacles using automatic procedures. For example, when the understudy stock occurs with low trading behaviour and the time points have a random design. This happens because we define the interval returns based on a constant number of time lags, which is the same for each interval. Therefore, pre-averaging may be a challenging task for intraday intervals with different trading frequencies.

4.4.3 Realised-Kernel Estimator

The Realised-Kernel (RK) estimator was developed by Barndorff-Nielsen et al. (2008), who provided the theoretical framework around this estimator. A more practical view of this estimator is delivered by Barndorff-Nielsen et al. (2009), who describe how we can use it in the real world.

We consider that the observed prices occur at random time points $0 < t_1 < ... < t_i < ... < t_n \leq T$ where $i = 1, ..., n$. In this context, the RK estimator is defined as:

$$RK(Y) = \sum_{h=-H}^{H} K\Big(\frac{h}{H+1}\Big)\gamma_h \tag{4.18}$$

where K represents a kernel function with bandwidth h. The component γ_h is given by:

$$\gamma_h = \sum_{i=|h|+1}^{n} R(t_i)R(t_{i+h})$$

where $R(t_i)$ denotes the i-th return and $R(t_{i+h})$ indicates the return placed h posi-

tions apart from the return i. Barndorff-Nielsen et al. (2009) suggest the Parzen kernel function because it always provides non-negative estimates, given by:

$$K(u) = \begin{cases} 0, & \text{if } \ u > 1 \\ 2(1-u)^3, & \text{if } \ 1/2 \leq u \leq 1. \\ 6(u^3 - u^2) + 1, & \text{if } \ 0 \leq u \leq 1/2 \end{cases}$$

In the Estimator (4.18), the initial optimal bandwidth value H^* can be estimated by minimising the bias-variance trade-off asymptotically. Barndorff-Nielsen et al. (2009) report that this value is given for:

$$H^* = c\, \xi^{4/5} n^{3/5}$$

where c refers to a constant that depends on the kernel function. Indicatevely, for the Parzen kernel $c = 3.5134$. Moreover, n shows the sample size and an estimator of ξ is given by:

$$\widehat{\xi} = \frac{\widehat{\omega^2}}{\widehat{IV}}.$$

In the above formula, we can estimate IV by (2.8), whilst Barndorff-Nielsen et al. (2009) use the following estimator for ω^2:

$$\widehat{\omega^2} = \frac{1}{\ell} \sum_{i=1}^{\ell} \left(\frac{1}{2z_i} \sum_{k=1}^{g_i} \left(Y(t_{i+k\ell}) - Y(t_{i+(k-1)\ell}) \right)^2 \right)$$

where ℓ refers to a specified lag, g_i indicates the highest number such that $i+k\ell < n$ using the i-th point as starting point for $i = 1, ..., \ell$. Also, $Y(t_{i+k\ell}) - Y(t_{i+(k-1)\ell})$ indicates the differences of log-prices placed ℓ lags away for $k = 1, ..., g_i$, using the i-th log-price as the starting point. Besides, z_i denotes the number of non-zero elements for all the differences starting from the i-th log-price each time. Barndorff-Nielsen et al. (2009) recommend

a value for ℓ that places the log-prices $Y(t_{i+k\ell})$ and $Y(t_{i+(k-1)\ell})$ two minutes away, on average.

Some features of this estimator follow. RK is a consistent estimator of IV and achieves the optimal convergence rate $n^{-1/4}$. Also, it is defined for irregularly spaced data, making its application more convenient than for the PA approach.

4.5 Spot Volatility

The pointwise estimates of our proposed approach can be utilised to approximate the spot volatilities. As opposed to volatility -a diachronic notion that measures the daily, monthly or annual fluctuations- the investigation of the spot volatility patterns has developed over the last three decades. Spot volatility refers to the volatility observed at a specific time over a day.

Foster and Nelson (1996) first focused on exploring spot volatility, proposing an estimator given as the summation of the squared returns within an intraday interval based on a weighting scheme. Later, Andreou and Ghysels (2002) modified this model by subtracting the drift from the returns before squaring this quantity. On the other hand, Fan and Wang (2008) proposed a kernel-based estimator for the spot volatility approximation. Kristensen (2010) applies the local constant smoother to the intraday squared returns for the estimation of spot volatilities under the assumption that the observed returns mirror the true returns (absence of noise).

Other popular studies on this topic were published by Zu and Boswijk (2014) who define an estimator that utilises as component the TSRV model, as given in the subsection 4.4.1, over intraday periods for the point approximation of σ^2 through time. Ogawa and Sanfelici (2011) measure the spot volatility by applying a two-step procedure to filter out the microstructure noise. In particular, they average the observed prices and calculate the squared returns based on these averaged prices. Then, they implement a regularisation scheme on the resulting squared returns for the spot volatility estimation.

4.6 Simulation

This section applies a simulation procedure where the accuracy of the LPR-based estimator is tested against the common IV estimators, outlined in section 4.4. We perform the simulation procedure 100 times, and the error between the estimated and the true IV is collected for each iteration:

$$error(iter) = \widehat{IV}_{iter} - IV_{iter}$$

where $\widehat{IV}_{iter} - IV_{iter}$ denote the difference between the estimated and true IV for a particular iteration. Once we have collected the 100 errors from the estimators, the mean, the MSE and the variance are computed. The simulation model rarely offers outliers and for this reason the median, the median squared error and the squared value of the median absolute deviation of the errors are also considered.

The 'Stochastic Volatility - Factor One' model (or briefly SVF1) simulates the IV for every iteration. The acronym SVF1 was given by Huang and Tauchen (2005), and it is a special case of a broad family of stochastic processes, introduced by Chernov et al. (2003). This is a popular simulation model in the field of stochastic volatility modelling with high-frequency data (Barndorff-Nielsen et al. (2008), Podolskij et al. (2009) among others). The SVF1 model has the following form:

$$dX(t) = \mu(t)dt + \sigma(t)dW(t), \tag{4.19}$$

$$d\tau(t) = \alpha\tau(t)dt + dB(t), \tag{4.20}$$

where

$$\sigma(t) = exp\Big(\beta_0 + \beta_1\tau(t)\Big), \tag{4.21}$$

$$Corr\Big(dW(t), dB(t)\Big) = \rho.$$

In the formulas above, $W(t)$ and $B(t)$ indicate two standard Brownian motions that are correlated with a correlation equal to ρ. This can be constructed from $B(t) = \rho W(t) + \sqrt{1-\rho^2}Z(t)$, where $Z(t)$ is a Brownian motion independent of $W(t)$. In the literature, this correlation coefficient is referred as 'leverage parameter' (Barndorff-Nielsen et al., 2008) or 'leverage correlation' (Huang and Tauchen, 2005) and enables the existence of leverage effect (Nolte and Voev, 2012). Although we assume the absence of leverage effect for the derivation of our estimator (see Assumption 2.2.2), we assess its performance under the presence of this effect. The component μ refers to the drift of the log-price change $dX(t)$ and $\sigma(t)$ indicates the spot volatility through time, specified by Equation (4.21). In addition, $\tau(t)$ refers to a component that determines the volatility process through time.

For the construction of the two correlated standard Brownian motions, the next steps are followed:

1. Generate one two-dimensional correlated sequence Q of size n such that:

$$Q \sim N\left([0,0], \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix}\right)$$

   where ρ is chosen to be -0.03.

2. Then, the two (correlated) standard Brownian motions in Equations (4.19) and (4.20) obtained on finite grids by:

$$Q_{1n} = W_n = [0, w(t_1), w(t_1) + w(t_2), ..., \sum_{i=1}^{k} w(t_i), ..., \sum_{i=1}^{n-1} w(t_{n-1})]\sqrt{dt},$$
$$Q_{2n} = B_n = [0, b(t_1), b(t_1) + b(t_2), ..., \sum_{i=1}^{k} b(t_i), ..., \sum_{i=1}^{n-1} b(t_{n-1})]\sqrt{dt}.$$

For the construction of the SVF1 model, we follow the next steps:

1. Generate n equidistant time points t_i within the interval $[0, 1]$ and the sample size is taken to be $n = [256,\ 1024,\ 4096,\ 9216]$.

2. Generate a process of n normally distributed random variables for τ_t, such that:

$$\tau \sim N\Big(0, -1/(2\alpha)\Big)$$

where α is chosen equal to -0.025 in this model.

3. We randomly choose a value from $N\Big(0, -1/(2\alpha)\Big)$ as the initial value for τ_1.

4. Then, $\sigma(t_1)$ can be computed by (4.21).

5. The initial value of the true log-price X_0 is chosen from $N\Big(3, \sigma^2(t_1)\Big)$.

6. The two SDE models have the following form:

$$dX(t) = \mu(t)dt + \sigma(t)dW(t),$$

$$d\tau(t) = \alpha\tau(t)dt + dB(t).$$

The formulas above can be expressed as:

$$\begin{aligned} X(t_n) - X(t_0) = \int_{t_0}^{t_n} \mu(t)dt + \int_{t_0}^{t_n} \sigma(t)dW(t) \Leftrightarrow \\ X(t_n) = X(t_0) + \int_{t_0}^{t_n} \mu(t)dt + \int_{t_0}^{t_n} \sigma(t)dW(t). \end{aligned}$$

These quantities can be estimated by the Euler-Maruyama approximation method:

$$\begin{aligned} X(t_i) &= X(t_{i-1}) + \mu(t_{i-1})dt + \sigma(t_{i-1})(W(t_i) - W(t_{i-1})) \\ \tau(t_i) &= \tau(t_{i-1}) + \alpha\tau(t_{i-1})dt + B(t_i) - B(t_{i-1}) \\ \sigma(t_i) &= exp\Big(\beta_0 + \beta_1\tau(t_i)\Big) \end{aligned} \tag{4.22}$$

where $W(t) - W(t - \Delta t) = \Delta W(t) \sim N(0, \Delta t)$. Also, μ is set equal to 0.03, $\beta_1 = 0.125$ and $\beta_0 = b_1^2/(2\alpha) = -0.3125$, so that $\mathbb{E}[\sigma^2(t)] = 1$.

7. The true IV is computed by:

$$IV = \sum_{t=1}^{n} \sigma^2(t)\Delta t.$$

8. After that, the noisy log-prices that will be used as inputs in the estimators are given by:

$$Y(t_i) = X(t_i) + E(t_i).$$

The pointwise true log-prices $X(t_i)$ are simulated by Equation (4.22), and the error terms $E(t_i)$ are generated from a normal distribution, such that $E \sim N(0, \omega^2)$.

4.6.1 Simulation - Results

In this subsection, we examine the performance of the LPR-based model against some standard estimators of IV through a simulation procedure based on 100 iterations. The simulation design is the same as determined at the beginning of this section. The difference $\widehat{IV} - IV$ is collected for every iteration, and relevant statistics of this difference are computed.

In this application, the estimators RK, MRV, TSRV and RV are tested against our LPR-based estimator. Our estimator is considered in the following two forms:

$$LPR^{(1)} = \sum_{i=2}^{n} \left(\widehat{m_{R^2}}(t_i) - \left(\widehat{m_R}(t_i) \right)^2 \right),$$
$$LPR^{(2)} = \sum_{i=2}^{n} \left(\widehat{m_{R^2}}(t_i) - \left(\widehat{m_R}(t_i) \right)^2 \right) - 2(n-1)\widehat{\omega^2}$$

where $\widehat{m_{R^2}}(t_i)$ denotes the mean curve of the pointwise squared returns, $\widehat{m_R}(t_i)$ displays the mean curve of the pointwise returns and ω^2 refers to the noise variance. The es-

timator $LPR^{(2)}$ expresses the standard estimator, as defined in section 4.3, whereas in $LPR^{(1)}$ the term $2(n-1)\widehat{\omega^2}$ is ignored. This modification aims to infer whether the third term of our proposed estimator is essential for the IV estimation.

We use the Gaussian kernel in both alternatives for the weights allocation, and the CV method is used to estimate the optimal bandwidth value. In terms of RK, the Tukey-Hanning$_2$ kernel is used, given by:

$$K(u) = \sin^2\left(\frac{\pi}{2}(1-u)^2\right).$$

Barndorff-Nielsen et al. (2008) show the superiority of this kernel compared to other kernel functions, and this kernel is also utilised in the simulation studies of Barndorff-Nielsen et al. (2008) and Podolskij et al. (2009). For RV and TSRV, we specify intraday returns based on 10-minute intervals. In the model $LPR^{(2)}$, the noise variance is approximated by (4.10).

Hansen and Lunde (2006) explore the annual mean noise variance for 30 stocks traded in the Dow Jones Industrial Average for the years 2000-2004 using the estimators presented in subsection (4.3.1). They found that the annual absolute mean value of $\widehat{\omega^2}$ lies between 0.00001 and 0.02891. In general, values around 0.01 are not observed too often in their findings, whereas values around 0.0001-0.001 are regarded as normal. The simulation study of Barndorff-Nielsen et al. (2008) and Podolskij et al. (2009) examine their estimator for noise variance values $\omega^2 = 0.01, 0.001, 0.0001$ and $\omega^2 = 0.01, 0.001$, respectively.

Table 4.2 reports the performance of the estimators based on these noise variance values. More specifically, the mean, MSE and variance are computed along with the median, Median Squared Error (MedSE) and the Squared value of the Median Absolute Deviation (SMAD). The latter three statistics are also considered because of the very few extreme values the simulation model produces. As sample size we choose $n =$

[256, 1024, 4096, 9216] where the estimators of Barndorff-Nielsen et al. (2008) and Podolskij et al. (2009) are also evaluated for similar sample sizes. Indicatively, we notice 2,428 intraday prices, on average, through the data period in our dataset.

In Table 4.2, RK outperforms the rest estimators for all the pairs of ω^2 and n, followed by TSRV. The performance of MRV only exceeds TSRV for the occasions where $n = 9216$. In terms of the LPR-based methods, they present poor results because much noise is aggregated in the estimation. Among the two LPR-based measures we have specified in the simulation study, $LPR^{(2)}$ is the estimator which performs better due to the additional term compared to $LPR^{(1)}$. In some cases, it can also be considered a slightly better measure than RV, especially when $\omega^2 = 0.01$. However, it fails to afford accurate estimates compared to the estimators RK, MRV and TSRV for these ω^2 values.

Nevertheless, noise variance values of size 0.0001-0.01 can be regarded as extremely high on the daily scale. In the findings of Hansen and Lunde (2006), the noise variance values converted on the daily scale correspond to values between $3.9E^{-8}$ and $9.87E^{-5}$. Table 4.3 summarises the findings of the simulation study where the noise variance takes values $\omega^2 = 1E^{-5}$, $\omega^2 = 1E^{-6}$ and $\omega^2 = 1E^{-7}$. For these values, $LPR^{(1)}$ exhibits the best performance across the estimators, followed by RK. More specifically, RK offers better MedSE values only for the cases where the sample size is 4096 and 9216 for $\omega^2 = 1E^{-5}$, that is, for the case of high sampling frequency when the noise variance is large. This is a natural result due to the noise aggregation in $LPR^{(1)}$. Overall, we conclude that $LPR^{(1)}$ outperform the rest estimators for common daily ω^2 values, especially for $\omega^2 = 1E^{-6}$ and $\omega^2 = 1E^{-7}$. On the other hand, $LPR^{(2)}$ has the worst performance among the measures due to the noise variance subtraction in the third term of this model.

Estimator	n	$\omega^2 = 0.01$			$\omega^2 = 0.001$			$\omega^2 = 0.0001$		
		Mean	MSE	Variance	Mean	MSE	Variance	Mean	MSE	Variance
		Median	MedSE	SMAD	Median	MedSE	SMAD	Median	MedSE	SMAD
RK	256	-0.09607	1.4583	1.4637	0.0072	0.45828	0.46285	-0.0229	1.3461	1.3592
		-0.04006	0.06319	0.06537	0.01796	0.05012	0.04271	-0.07668	0.05424	0.0481
	1024	-0.02673	0.5561	0.561	0.04951	0.16716	0.16638	-0.13041	0.34246	0.32875
		-0.00922	0.04312	0.0431	0.02096	0.02472	0.02212	-0.07019	0.01523	0.01534
	4096	0.01991	0.20133	0.20296	0.02831	0.06636	0.06622	-0.09177	0.12687	0.11964
		0.00935	0.0136	0.01587	0.00383	0.00787	0.00787	-0.02813	0.00542	0.00486
	9216	-0.01883	0.11366	0.11445	0.0292	0.05673	0.05644	-0.05439	0.06561	0.06328
		-0.00535	0.00853	0.00955	-0.00637	0.00457	0.00421	-0.01387	0.00392	0.0049
MRV	256	-0.30498	4.7295	4.6833	0.26961	2.9957	2.9525	0.11897	10.5163	10.6083
		-0.15156	0.33715	0.25064	-0.03647	0.25541	0.24161	-0.21123	0.3906	0.42099
	1024	-0.12687	3.8849	3.9079	0.16238	2.3185	2.3153	-0.05233	2.1913	2.2107
		-0.10318	0.12235	0.12721	0.01846	0.13631	0.12376	-0.08905	0.17595	0.1428
	4096	0.10263	0.93545	0.93426	0.22401	0.74044	0.69723	0.05165	1.4466	1.4585
		-0.02264	0.09443	0.09545	0.05945	0.07931	0.08256	-0.01171	0.0828	0.07746
	9216	-0.0651	0.96246	0.9679	-0.02093	0.4266	0.43047	-0.0946	0.51878	0.51499
		-0.00947	0.04331	0.04004	0.017	0.03016	0.03291	-0.05556	0.04489	0.05006
TSRV	256	-0.70079	3.0186	2.553	-0.43466	1.0414	0.86108	-0.45116	1.587	1.3974
		-0.24367	0.17238	0.11225	-0.19355	0.12292	0.06788	-0.25136	0.12801	0.08826
	1024	-0.33351	1.2499	1.1501	-0.15224	0.54848	0.53061	-0.17322	1.2913	1.274
		-0.13664	0.08403	0.05927	-0.08199	0.05034	0.05377	-0.12935	0.07116	0.06853
	4096	-0.22897	1.047	1.0046	-0.06669	0.47013	0.47039	-0.08622	1.3003	1.3059
		-0.08375	0.07661	0.06728	-0.03981	0.04863	0.03989	-0.09813	0.06503	0.05748
	9216	-0.22259	1.0309	0.99122	-0.05514	0.46419	0.46581	-0.07402	1.2981	1.3057
		-0.08043	0.08862	0.07974	-0.03835	0.04201	0.03937	-0.08828	0.06518	0.05459
RV	256	-0.00803	4.5191	4.5647	-0.39651	1.7863	1.6455	-0.33778	1.8127	1.7157
		0.52347	0.50794	0.18059	-0.13129	0.07805	0.08894	-0.20002	0.0922	0.06645
	1024	0.50772	2.8464	2.6148	-0.0475	0.9692	0.97671	-0.00832	1.7655	1.7833
		0.72662	0.67085	0.19838	-3e-05	0.14005	0.14006	-0.03973	0.08225	0.0715
	4096	0.67445	2.8442	2.4135	0.11778	1.0013	0.99737	0.11415	2.4679	2.4796
		0.74973	0.71413	0.18858	0.09185	0.13924	0.1109	-0.0059	0.0976	0.09694
	9216	0.6214	2.4452	2.0798	0.17557	0.98549	0.9643	0.11609	2.1733	2.1816
		0.681	0.67219	0.09781	0.07975	0.17295	0.1213	-0.00119	0.10395	0.10392
$LPR^{(1)}$	256	5.1513	27.3067	0.77852	0.5223	0.45366	0.18269	-0.09657	0.361	0.35523
		5.0541	25.5441	0.19597	0.49485	0.25309	0.01556	0.0269	0.00739	0.00632
	1024	20.4288	418.8159	1.4957	2.0794	4.4038	0.08071	0.16948	0.08729	0.05915
		20.3641	414.6978	0.73548	2.0731	4.2976	0.01721	0.20036	0.04523	0.00281
	4096	91.5496	8387.4111	6.1418	9.1758	84.3413	0.14671	0.91675	0.85524	0.01496
		91.5041	8373.0087	2.5176	9.1927	84.505	0.03125	0.92214	0.85034	0.00203
	9216	184.7359	34139.9017	12.6601	18.3724	337.713	0.16822	1.8375	3.3841	0.00777
		185.2634	34322.5219	6.0815	18.385	338.01	0.05745	1.8392	3.3827	0.00202
$LPR^{(2)}$	256	-0.46372	5.6449	5.4847	-0.02866	2.1276	2.1483	-0.01034	7.446	7.5211
		-0.19278	0.33081	0.30497	-0.07832	0.14772	0.13178	-0.14177	0.1613	0.14506
	1024	-0.36693	4.3962	4.3046	0.18523	4.2197	4.2277	0.09907	8.9608	9.0414
		-0.27913	0.36	0.41941	0.00912	0.13946	0.13964	-0.09352	0.14531	0.12807
	4096	-0.30094	5.8976	5.8657	0.20453	4.1755	4.1754	0.17582	9.0161	9.076
		-0.30611	0.58225	0.40806	-0.01786	0.16246	0.15891	-0.07415	0.1421	0.14047
	9216	-0.35637	5.2916	5.2168	0.21246	4.2206	4.2176	0.18849	9.2651	9.3228
		-0.34992	0.98566	0.65487	0.04564	0.15094	0.14574	-0.04389	0.13154	0.13301

Table 4.2: Simulation results based on 100 iteration for all the combinations $n = [256,\ 1024,\ 4096,\ 9216]$ and $\omega^2 = [0.01,\ 0.001,\ 0.0001]$.

Estimator	n	$\omega^2 = 0.00001$			$\omega^2 = 0.000001$			$\omega^2 = 0.0000001$		
		Mean	MSE	Variance	Mean	MSE	Variance	Mean	MSE	Variance
		Median	MedSE	SMAD	Median	MedSE	SMAD	Median	MedSE	SMAD
RK	256	-0.21081	1.7896	1.7628	0.23207	1.8363	1.8005	-0.04613	0.79815	0.80406
		-0.01964	0.0485	0.05167	-0.00301	0.03662	0.03662	-0.02211	0.03155	0.02948
	1024	-0.13374	0.78666	0.77654	0.15193	0.43992	0.42105	-0.02685	0.24314	0.24486
		-0.0084	0.01869	0.01952	0.00254	0.01132	0.01186	-0.00804	0.0145	0.0128
	4096	-0.04201	0.18799	0.18811	0.13432	0.34855	0.33384	0.0306	0.12507	0.12539
		0.00046	0.00528	0.00526	0.00712	0.00303	0.00353	-0.00702	0.01097	0.01032
	9216	-0.01042	0.10928	0.11027	0.1149	0.29112	0.28072	0.02348	0.07094	0.0711
		-0.00059	0.00385	0.00378	0.00028	0.00275	0.00278	-0.00494	0.00693	0.00613
MRV	256	0.04274	3.5835	3.6178	-0.06374	5.0231	5.0698	-0.43376	4.3941	4.2485
		-0.09733	0.25092	0.17018	-0.02788	0.2661	0.27169	-0.13329	0.24295	0.26648
	1024	-0.29997	4.5364	4.4913	0.39121	10.0716	10.0187	0.2936	5.5296	5.4984
		-0.0404	0.10991	0.11862	-0.05599	0.14277	0.14273	-0.02179	0.14105	0.12742
	4096	-0.1471	1.7195	1.715	0.41519	2.3835	2.2334	-0.09913	1.3136	1.3169
		-0.02387	0.03275	0.03583	0.03625	0.04806	0.05311	-0.02023	0.04622	0.04458
	9216	-0.10286	1.6158	1.6214	0.23349	1.7627	1.7254	-0.07899	0.61691	0.61684
		-0.02261	0.03016	0.036	0.00622	0.0305	0.0293	-0.025	0.02901	0.02551
TSRV	256	-0.77088	4.3063	3.7496	-0.29024	1.4288	1.3582	-0.55486	1.5728	1.2777
		-0.21877	0.12757	0.10265	-0.20414	0.06554	0.05197	-0.24601	0.083	0.07319
	1024	-0.41339	2.5697	2.423	0.01245	1.3089	1.322	-0.24258	0.8972	0.84682
		-0.09813	0.07751	0.06233	-0.07464	0.03321	0.03373	-0.08959	0.03229	0.03229
	4096	-0.29548	2.0189	1.9511	0.10607	1.3787	1.3812	-0.13808	0.76376	0.75222
		-0.06396	0.05917	0.04949	-0.04417	0.02524	0.0275	-0.04469	0.03894	0.03334
	9216	-0.27848	1.9563	1.8977	0.11921	1.3865	1.3861	-0.12264	0.75249	0.7449
		-0.05644	0.05634	0.04893	-0.03656	0.02511	0.0268	-0.03728	0.03981	0.03459
RV	256	-0.81056	4.9281	4.3143	-0.28792	0.87311	0.79819	-0.75585	2.2433	1.6889
		-0.24826	0.1252	0.09799	-0.20334	0.06505	0.05426	-0.23532	0.10169	0.05337
	1024	-0.36173	2.4943	2.3873	0.16625	2.3731	2.3691	-0.25518	0.89814	0.84144
		-0.08635	0.10521	0.1046	-0.00653	0.04365	0.04127	-0.13211	0.05409	0.04634
	4096	-0.25364	2.0856	2.0417	0.2477	2.2338	2.1943	-0.13096	0.94865	0.94091
		-0.02762	0.07411	0.07969	-0.01497	0.04346	0.04193	-0.07191	0.05865	0.04587
	9216	-0.24193	2.0361	1.9976	0.25688	2.2561	2.2122	-0.12621	0.93428	0.92763
		-0.02209	0.06112	0.07253	0.00834	0.04654	0.04606	-0.06916	0.06105	0.04577
$LPR^{(1)}$	256	-0.10966	0.22138	0.21147	0.07411	0.24629	0.24323	-0.01145	0.14195	0.14325
		-0.03287	0.0158	0.01227	-0.00875	0.00616	0.00487	-0.00692	0.00715	0.00685
	1024	0.08122	0.09263	0.0869	0.05741	0.1249	0.12284	-0.02893	0.04559	0.04521
		0.02318	0.00314	0.00313	-0.00048	0.00216	0.0022	-0.00841	0.00323	0.00367
	4096	0.09672	0.0234	0.01419	0.01049	0.0149	0.01494	-0.00488	0.01803	0.01818
		0.08847	0.00821	0.00125	0.01313	0.00055	0.00043	0.00037	0.00052	0.00052
	9216	0.19365	0.05086	0.01349	0.01318	0.00562	0.00551	-0.00596	0.01773	0.01788
		0.18705	0.0366	0.00056	0.01625	0.00045	0.00016	0.00167	0.00023	0.00024
$LPR^{(2)}$	256	-0.14367	2.7056	2.7121	-0.12314	1.9141	1.9181	-0.53386	3.3052	3.0507
		-0.07605	0.17483	0.16119	-0.05539	0.13788	0.11545	-0.24411	0.3439	0.21138
	1024	-0.05885	2.9209	2.9469	0.11333	3.7835	3.8088	-0.32285	3.3644	3.2931
		-0.0444	0.17184	0.18732	-0.03812	0.17576	0.17184	-0.19607	0.16727	0.17357
	4096	0.12443	2.7556	2.7678	0.14454	4.8124	4.8399	-0.26671	3.4519	3.4149
		0.01637	0.15474	0.15621	-0.00738	0.18864	0.18676	-0.16257	0.17815	0.1803
	9216	0.13725	3.2534	3.2673	0.13382	4.7811	4.8113	-0.26752	3.3813	3.3432
		0.01918	0.18856	0.18083	-0.02011	0.19734	0.19443	-0.13148	0.17036	0.19052

Table 4.3: Simulation results based on 100 iteration for all the combinations $n = [256,\ 1024,\ 4096,\ 9216]$ and $\omega^2 = [0.00001,\ 0.000001,\ 0.0000001]$.

4.7 Empirical Application

This section estimates the daily IV of Cleveland-Cliffs Inc. (CLF) and Wells Fargo & Co. (WFC) over the years 2012-2014. As we have noticed in Table 1.3 and in the empirical applications of Chapter 3, CLF is the most variable stock. Thus, we expect that it will exhibit high IV estimates, whilst WFC is a liquid stock generally, with 2,860 observed prices per day, on average.

This application affords a comparative view of the behaviour of our proposed LPR-based estimator against the estimators RK, TSRV and RV. As we will see later in this analysis, the daily noise variances of the two stocks need more than six decimals to be presented as non-zero values. Following the results of the simulation study, our estimator is considered in the form:

$$LPR = \sum_{i=1}^{n-1} \left(\widehat{m_{R^2}}(t_i) - \left(\widehat{m_R}(t_i) \right)^2 \right).$$

We choose the Gaussian kernel function to allocate the weights and the optimal bandwidth is chosen via CV. In addition, we consider 10-minute intervals daily for RV and TSRV, whilst the Parzen kernel function is used in the RK estimator.

Figure 4.3 illustrates the logarithmic value of the daily IV estimates of the aforementioned measures over the trading days. Due to the difference in the size of the IV estimates for the two stocks, we choose to plot the estimates on different scales so that one can have a clear view of the estimators' behaviour. Overall, the IV estimates are higher for CLF than for WFC. This fact is not surprising because of the high variability of this stock, reported in Tables 1.3, 1.4 and the empirical applications of Chapter 3. Beyond that, RV (cyan line) often presents the highest estimates for both CLF and WFC, while the estimates of TSRV (green line) are usually lower than the other models' estimates. On the other hand, the LPR-based and the RK models (given by blue and red line, respectively) appear to have more comparable performance. As we also see

in the simulation analysis with noise variances of normal daily size, the MSE and the median squared error values of LPR and RK are much smaller than the corresponding values of RV and TSRV.

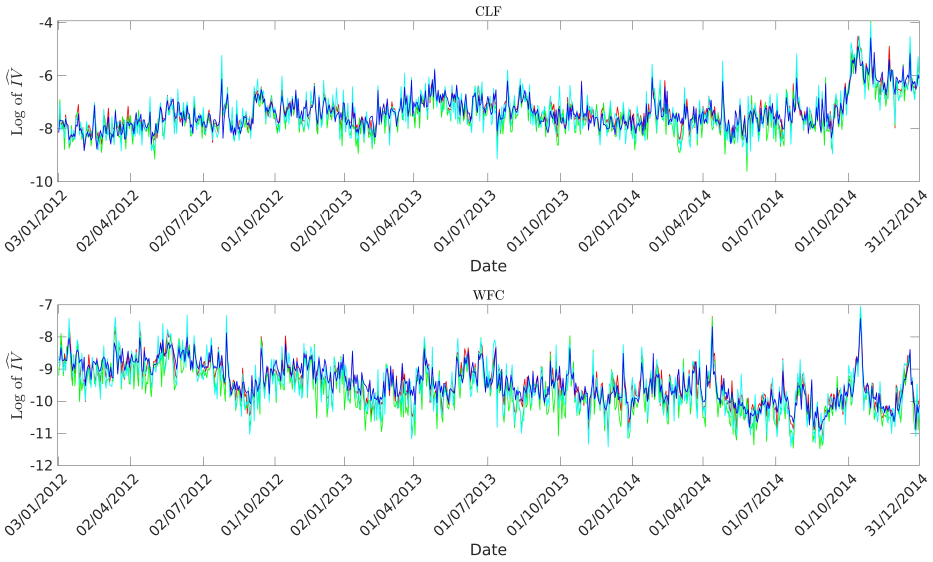

Figure 4.3: This figure depicts the logarithmic value of the daily IV estimates of the LPR-based (blue line), RK (red line), TSRV (green line) and RV (cyan line) estimator for the stocks CLF and WFC over the trading days of the years 2012-2014.

Furthermore, it would be interesting to estimate the magnitude of the noise variance for these assets. Figure 4.4 displays the daily noise variance estimates of the two stocks over the working days of the data period. Again, we plot the estimates on a different scale for the two stocks in favour of a clearer view. In this figure, the Estimators (4.10) and (4.11) are used and they are depicted by cyan and blue colour, respectively. In the latter estimator, our LPR-based model is used to estimate IV. In Figure 4.4, the Estimator (4.11) offers less variable estimates of ω^2 than the Estimator (4.10), mainly for CLF. To avoid intricate statements, we use the terms 'higher' and 'lower' to refer to the absolute value of the estimates in Figure 4.4. This is because negative values for noise variance are likely to occur due to the subtraction that appears in the nominator of the Estimators (4.10) and (4.11). This is also the case in the findings

of Hansen and Lunde (2006).

Concerning WFC, the estimates of (4.10) are higher for the years 2012 and 2013, whereas the estimates of (4.11) considerably increase within 2014, exceeding the corresponding values of (4.10) for the second half of this year. For the Estimator (4.10), the range of $\widehat{\omega^2}$ values is between -2.2E-06 and 7.4E-07 for CLF, while these values lie between -9.5E-08 and 2.2E-08 for WFC. As regards the Estimator (4.11), the corresponding values are between -1.1E-07 and 6.9E-08 for CLF and between -3.5E-09 and 3.2E-09 for WFC. Therefore, we can say that the noise variance can take higher values for CLF compared to WFC.

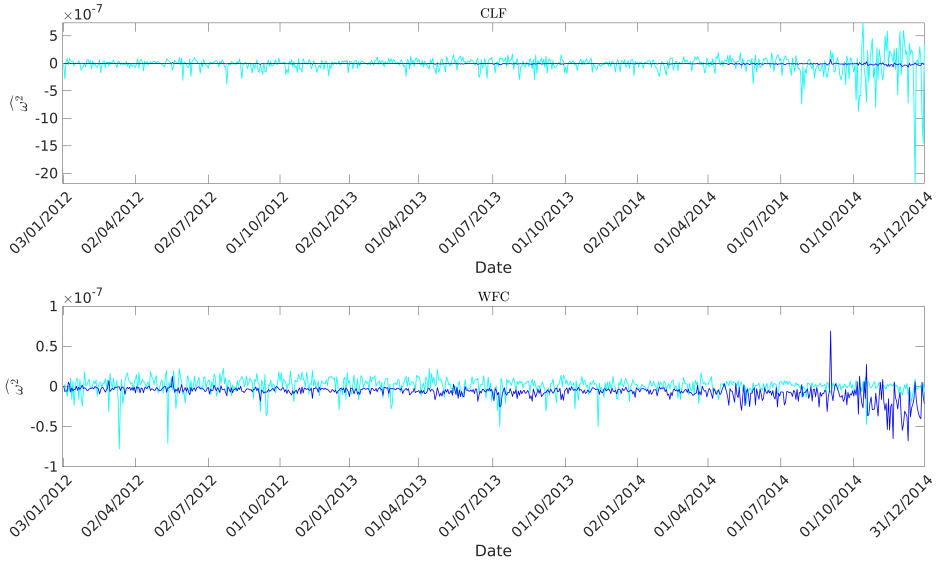

Figure 4.4: This figure depicts the daily noise variance estimates of CLF and WFC using the Estimators (4.10) and (4.11) with cyan and blue line, respectively, over the trading days of the years 2012-2014.

Besides, we can get insights about the volatility locally using the local elements of our proposed estimator. Figure 4.5 illustrates the local estimates of our LPR-based estimator for the two stocks over the first five trading days of 2012. We choose to portray the estimates on a different scale for the two stocks since they are of different magnitudes. We observe that the highest estimates are given at the beginning of the trading hours

for both assets, while CLF has greater estimates at this part of the day. Then, the estimates of both stocks follow a smoother behaviour in later hours with several distinct exceptions, mainly for CLF.

CLF is generally a more volatile stock than WFC on these five days. For CLF, the most volatile time point was observed at the beginning of the trading session of the second day (red line), whereas the most volatile time point was noticed at the same time point but on the fifth day for WFC (black line). During the last part of the trading sessions, we do not see any particular volatile performance for the two stocks on these five days. Note that each day has a different number of observations, resulting in curves of different lengths along the x-axis.

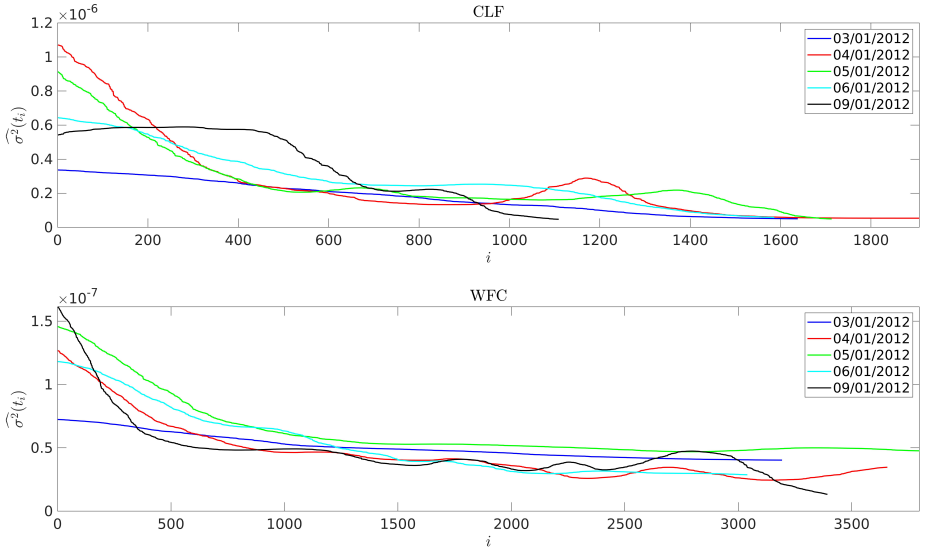

Figure 4.5: This figure depicts the local estimates of the LPR-based estimator over the first five trading days of 2012.

4.8 Discussion

We have proposed a new nonparametric estimator of IV based on local polynomial regression for high-frequency volatility estimation under the irregular design points with the presence of noise. Although similar ideas of nonparametric estimators have been around long time, some of them were studied under either the regular design points or without noise setting. Our estimator can be easily implemented without the need of pre-processing. The LPR-based method can also be utilised for spot volatility estimation. The simulation study showed that as the noise variance is of normal daily size, our proposed approach (discarding the error-correction term) can be considered superior compared to the benchmark methods based on the MSE and median squared error. Especially, the advantage of our approach to estimate IV for intermediate and low ω^2 daily values is evident, no matter the sample size. For these values, our estimator's MSE and the median squared error values are always from four to 51 times less than the corresponding statistic values of the estimator with the second-best performance.

In the empirical analysis, we observed that the LPR-based estimator performs similarly to RK regarding the IV estimation across the days. In many cases, RV provides more extreme values and TSRV remarkably lower estimates than the other measures. In general, our estimator offers higher estimates for the most variable stock of our dataset (as found in the exploratory analysis of Chapter 1 and the empirical investigations in Chapter 3). This is also the case for the other IV measures. In addition, our estimator can be used as a component for the estimation of the noise variance via the Estimator (4.11). Our investigation also used the Estimator (4.10). The estimator which incorporates the LPR-based measure to approximate the noise variance usually offers lower values than the Estimator (4.10), generally. Applying our approach to multiple days on two stocks, we notice that the highest local estimates occur at the beginning of the days, followed by a smooth decrease before keeping a constant trend with a few exceptions.

Although our estimator shows comparable performance, theoretical analysis suggests that the effect of constant noise variance could be a problematic factor asymptotically. One simple approach to remove this effect in our estimator would be to use smoothed log-prices to calculate the returns. On the other hand, it has been noted that the effect of noise tends to decrease as the sample size increases (Aït-Sahalia and Yu, 2009), which suggests an alternative modelling framework where the noise variance is a function of the sample size.

Chapter 5

Conditional Dependence Structure Estimation with Functional Graphical Models

5.1 Introduction

Dependencies in the data patterns of stocks concern plenty of financial sectors, such as financial risk management, financial risk analysis and portfolio analysis. Experts, being aware of these dependencies, are able to construct or modify a portfolio so that they will minimise the market risk. Market risk refers to the danger of heavy financial losses for a portfolio when steep fluctuations occur in the market. During these fluctuations, it is highly probable that dependent stocks will behave similarly, magnifying the portfolio losses when a sharp downward trend is observed. For this reason, specialists are used to composing portfolios with independent stocks so that they will offset the possible losses during an unexpected unfavourable event.

Thus, a correct interpretation of the dependence structure of a set of variables is crucial for risk management; for example, for protecting investors' wealth during periods

of uncertainty. Graphical models are useful tools in this field since they permit us to estimate and visualise the conditional dependence structure between stocks. Given a set of variables, the conditional dependence between two variables expresses the relationship between these variables when we control the rest of them.

Our particular interest is in the estimation of the stocks' dependence networks using high-frequency financial data. A necessary step is the estimation of the variance-covariance matrix. High-frequency financial data are irregularly spaced noisy observations and their number differs from stock to stock and across the days. Hence, we cannot approximate the conventional variance-covariance matrix with standard techniques. However, we can study this kind of data as functional data defined on a continuous domain (Ramsay and Silverman, 2005; Ferraty and Vieu, 2006; Horváth and Kokoszka, 2012).

In a functional dataset, every sample point of a variable is given in the form of a function, as opposed to a scalar or a vector value. Because of the plurality of information in their form and the fact that their realisations can follow an arbitrary process, we cannot treat them as usual time series data. This fact makes a functional dataset more challenging to handle. Nevertheless, the vast amount of information contributes to a more accurate approximation of the population features. In general, a conventional multivariate dataset is of fixed finite dimension, it is defined for discrete time points and does not include any shared information. On the other hand, functional data are "*intrinsically infinite dimensional*" (Wang et al., 2016) and are defined at any point in the continuous space. Also, they incorporate several special characteristics, such as common behaviour and common trends over the continuous domain, which differentiate this dataset from a conventional multivariate one. So, this kind of data includes tremendous information and thus interesting data patterns can be investigated, not directly observable in a multivariate dataset (Wang et al., 2016).

Smoothing methods and interpolation are valuable tools in this research area (Ullah

and Finch, 2013). We can exploit the smoothing methods to mitigate the noise size in the intraday observations and express the variance-covariance matrix in continuous time. That is, the variance-covariance matrix can be defined for irregularly spaced observations as opposed to the conventional matrix, which is defined for regularly spaced time points that do not differ across the variables. Besides, highly dependent variables might be a problem for determining the conventional variance-covariance matrix since singularities are likely to occur and hence its inverse matrix could be undefined. On the other hand, functional data as continuous smooth functions are highly dependent but we can approximate well its inverse function without removing any points. A conventional variance-covariance matrix is defined for finite points of observation. Even if the variance-covariance matrix under the functional data context is also determined for fixed time points, we can specify the points we want to evaluate the variables' dependence on the continuous time scale, independently of the observation time points. All of these considerations make the functional graphical analysis preferable for the high-frequency time series data to the multivariate graphical analysis based on a conventional variance-covariance matrix.

This chapter assumes that the data come from a multivariate Gaussian stochastic process, which means that any sample point from the data follows a multivariate normal distribution. Under this normality setting, the correlation coefficient sufficiently describes the dependence relationship of a set of variables (Embrechts et al., 2002; Meucci, 2005; Sweeting, 2017). In this case, the graphical model summarises the variables' conditional correlation, also called partial correlation.

Undoubtedly, one can claim that the normality assumption is a constraining conjecture; for example, Solea and Li (2022). Li and Solea (2018) report the three consequences we accept by the normality assumption. Namely, statisticians often consider the normal distribution because it is convenient for the investigation; however, it may be different in reality. Secondly, this assumption implies that the variables are exclusively connected

through a linear relationship. Thirdly, this association does not incorporate any additional (independent) component, which brings any linear dependence down. Beyond that, Rachev et al. (2009) mark several reasons why correlation is not an adequate measure of dependence. The above assertions may seem strong at first sight since the model does not capture non-linear dependencies, providing a delusive image of the variables' relationship. However, the assumption that the dataset follows a Gaussian process is convenient, and Gómez et al. (2020) cite several studies demonstrating that this is not damaging in practice under the framework of functional graphical models.

Considering the above context, Qiao et al. (2019) developed a methodology for the prediction of the variables' dependence structure through the graphical models. In their approach, the variance-covariance matrix is estimated using the most critical coefficients in their analysis, as suggested by the functional PCA. Furthermore, they develop an algorithm for the estimation of the dependence network of the coefficients. As regards our analysis, it follows the same theoretical environment as Qiao et al. (2019). However, the methodology for the estimation of the variance-covariance matrix differs. More specifically, we use the complete dataset to estimate this matrix by taking the cross-product of each function across the variables, and then we smooth it through a bivariate LPR. This technique is considered by Yao et al. (2005). Its advantages are that we exploit the complete data information for the estimation and the variance-covariance matrix can be defined for irregularly spaced time points. Once the variance-covariance matrix has been approximated, we adopt the algorithm of Qiao et al. (2019) on this matrix for the conditional dependence structure estimation. Alternative works in this area are discussed later in this chapter.

The main contribution of this chapter can be summarised as follows:

1. High-frequency data present diurnal patterns over the days. We exploit their special characteristics by studying high-frequency data as functional data. Thus, the curves consist of high-frequency data instead of common daily or monthly time

series data.

2. Our scope is the estimation of the stocks' conditional dependence through the functional graphical LASSO approach. Most of the existing methods eliminate the dimension of the problem using the best coefficients in the basis function representation (conditional dependence of the coefficient vectors). We develop a new methodology for estimating conditional dependence through functions.

3. We introduce a practical CV-based method for the estimation of the sparsity of the dependence network. This method approximates the optimal level of sparsity, that is, which variables appear to be conditionally independent, by examining the similar patterns of the variance-covariance matrix over different periods.

4. To the best of our knowledge, this is the first work that develops a comprehensive technique for estimating the dependence structure through graphical models and using functional data in the field of finance.

This chapter is organised as follows. Section 5.2 discusses the concept of graphical models, while the idea of graphical models is extended in the context of functional data in section 5.3. In the same section, we also analyse our proposed methodology for the estimation of the conditional dependence structure for a set of variables. In section 5.4, we conduct a simulation study for the proposed approach. In section 5.5, an empirical analysis is applied to stock data. This work comes to an end with a discussion in section 5.6.

5.2 Graphical Models

A probabilistic graphical model (or briefly graphical model) expresses the conditional dependence structure of a set of variables. This structure is given in the form of networks where each variable, represented as a node, is connected through edges to the variables

it shares dependence with.

The graphical model is denoted by:

$$\mathcal{G} = (\mathcal{V}, \mathcal{E})$$

where $\mathcal{V}$ denotes the graph's vertices (i.e. the variables) and $\mathcal{E}$ denotes the set of edges, such that $\mathcal{E} \subset \mathcal{V} \times \mathcal{V}$. The set of edges visualises the conditional dependence observed in each pair of variables.

There are two types of graphical models, the directed and the undirected graphical models. The directed graphical models depict the conditional dependence structure among the variables as well as the variables from which the dependence is driven. However, we focus on the undirected graphical models in this work. An undirected graphical model illustrates the dependence network of a number of variables, and the dependent variables are linked with edges.

A useful term in the context of graphical models is the 'clique'. A clique defines the sets of nodes where they are connected to each other through edges. In addition, a 'maximal clique' refers to a clique that cannot be extended by including more vertices in it. To highlight the differences, we portray a graphical model $\mathcal{G}$ in Figure 5.1. In this figure, we include six variables $\mathcal{V}(\mathcal{G}) = (1, 2, 3, 4, 5, 6)$ -given as nodes- in graph $\mathcal{G}$ and their conditional dependence structure is given by the edges $\mathcal{E}(\mathcal{G}) = (< 1, 2 >, < 1, 3 >, < 1, 4 >, < 2, 3 >, < 2, 4 >, < 3, 4 >, < 4, 6 >, < 5, 6 >)$. The first four variables are fully connected in this graph, showing that these variables share (conditional) dependence. On the other hand, variable 6 shares conditional dependence with variables 4 and 5; however, variable 5 is conditionally independent of the rest variables. The subgraphs $\mathcal{E}(\mathcal{G}_1) = (< 2, 3 >, < 2, 4 >, < 3, 4 >)$, $\mathcal{E}(\mathcal{G}_2) = (< 1, 3 >, < 1, 4 >, < 3, 4 >)$, $\mathcal{E}(\mathcal{G}_3) = (< 1, 2 >, < 1, 4 >, < 2, 4 >)$ and $\mathcal{E}(\mathcal{G}_4) = (< 1, 2 >, < 1, 3 >, < 2, 3 >)$ form four cliques since the variables of each subgraph are adjacent to each other. However, these cliques

cannot be considered maximal cliques since we can summarise the dependence networks of these four subgraphs in the broader subgraph $\mathcal{E}(\mathcal{G}_5) = (< 1,2 >, < 1,3 >, < 1,4 >, < 2,3 >, < 2,4 >, < 3,4 >)$. The subgraph $\mathcal{E}(\mathcal{G}_5)$ forms a maximal clique because it is a clique itself and cannot be extended by incorporating more variables in it. Moreover, variable 6 only shares dependence with variable 4 from the variables in subgraph $\mathcal{G}_5$ and thus is left outside the maximal clique determination. For the same reasons, variables 4, 6 and 5, 6 form two cliques which, at the same time, are maximal cliques because we cannot extend these cliques by including more variables into them. Therefore, these two maximal cliques are given by the edges $\mathcal{E}(\mathcal{G}_6) = (< 4,6 >)$ and $\mathcal{E}(\mathcal{G}_7) = (< 5,6 >)$.

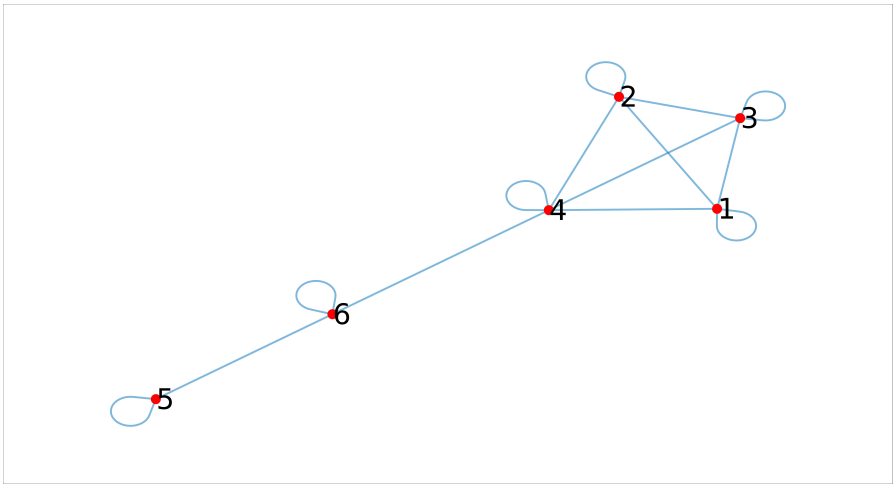

Figure 5.1: An illustrative example of a graphical model with six variables.

By understanding the variables' conditional dependence structure, we can express their joint distribution utilising their conditional distributions (Gómez et al., 2020). In this way, we reduce the problem's dimensionality, facilitating our analysis. In the case of undirected graphical models, the joint distribution satisfies the following:

$$p(X_1, X_2, ..., X_p) = \frac{1}{Z} \prod_{c \in \mathcal{C}(\mathcal{G})} \psi_c(X_c) \tag{5.1}$$

where c stands for the cliques, $\mathcal{C}(\mathcal{G})$ denotes the set of all the maximal cliques in the

graph $\mathcal{G} = (\mathcal{V}, \mathcal{E})$, ψ_c represents a non-negative function related to the cliques and Z is called partition function and is given by:

$$Z = \sum_{X_1, X_2, \ldots, X_p} \prod_{c \in \mathcal{C}(\mathcal{G})} \psi_c(X_c).$$

Hence, the graphical model gives the means to decompose and visualise the conditional dependencies arising in the joint distribution of p variables $(X_1, X_2, ..., X_p)$ so we can utilise it for the determination of the joint probability in Equation (5.1). Indicatively, the joint probability in Figure 5.1 is given by:

$$p(X_1, X_2, X_3, X_4, X_5, X_6) = \frac{1}{Z}\psi_{1,2,3,4}(x_1, x_2, x_3, x_4)\psi_{4,6}(x_4, x_6)\psi_{5,6}(x_5, x_6)$$

and the variables' conditional independence structure in this figure is:

$$\begin{array}{lll} 1 \perp\!\!\!\perp 5|2,3,4,6 & 2 \perp\!\!\!\perp 5|1,3,4,6 & 3 \perp\!\!\!\perp 5|1,2,4,6 \\ 1 \perp\!\!\!\perp 6|2,3,4,5 & 2 \perp\!\!\!\perp 6|1,3,4,5 & 3 \perp\!\!\!\perp 6|1,2,4,6 \\ & 4 \perp\!\!\!\perp 5|1,2,3,6. & \end{array}$$

The question now is centred on how we can estimate the conditional dependence network of the p variables. This can be done using the graphical LASSO approach outlined in the following subsection. In this work, we assume that the observations stem from a joint Gaussian distribution, and thus we focus on the linear conditional dependence, also known as partial correlation.

5.2.1 Graphical LASSO Approach

The Graphical LASSO (GLASSO) approach was developed by Friedman et al. (2008), and it makes use of the concept of LASSO (Tibshirani, 1996). In a nutshell, assuming the linear regression model the scope of the LASSO (the initials of the letters Least Absolute

Shrinkage and Selection Operator) approach is to eliminate the explanatory variables which have a relatively small impact on the response variable using a penalty term in favour of a problem simplification. LASSO is a popular technique in sparsity modelling. In statistics, the latter term refers to the procedure where we identify and discard the parameters of low importance in the model to facilitate the model interpretation. One can refer to Hastie et al. (2015) for more details about LASSO and its generalisations.

Relative to LASSO, the target of GLASSO is to shrink small conditional dependencies between the variables to zero, utilising a penalty term. Supposing that the variables come from a Gaussian distribution, then it is known that the variance-covariance matrix Σ expresses the dependence between a set of variables, where:

$$\Sigma_{j_1 j_2} = 0 \Leftrightarrow X_{j_1} \perp\!\!\!\perp X_{j_2}$$

for two variables j_1 and j_2. Our purpose via GLASSO is the estimation of the precision matrix Θ. The Θ matrix is defined as $\Theta = \Sigma^{-1}$ and gives the partial correlation between the variables. This means that when:

$$\Theta_{j_1 j_2} = 0 \Leftrightarrow X_{j_1} \perp\!\!\!\perp X_{j_2} | X_{-j_1 j_2}.$$

By this method, we are able to visualise the (linear) conditional dependence structure between the variables, where the conditionally uncorrelated variables correspond to zero

elements in the Θ matrix. For instance, in Figure 5.1 the Θ matrix would be of the form:

$$\Theta = \begin{pmatrix} NZ & NZ & NZ & NZ & 0 & 0 \\ NZ & NZ & NZ & NZ & 0 & 0 \\ NZ & NZ & NZ & NZ & 0 & 0 \\ NZ & NZ & NZ & NZ & 0 & NZ \\ 0 & 0 & 0 & 0 & NZ & NZ \\ 0 & 0 & 0 & NZ & NZ & NZ \end{pmatrix}$$

where the acronym NZ shows the non-zero elements of the Θ matrix.

Suppose a p-dimensional vector X, such that $X \sim N(0, \Sigma)$, then the corresponding multivariate Gaussian PDF is given by:

$$f(x) = \frac{e^{\{-\frac{1}{2}x^T\Sigma^{-1}x\}}}{(2\pi)^{p/2}\det(\Sigma)^{1/2}} \propto e^{\{-\frac{1}{2}x^T\Theta x\}}\det(\Theta)^{1/2} \tag{5.2}$$

where $\Sigma = \mathbb{E}[XX^T]$ is the non-negative definite variance-covariance matrix, Θ denotes the precision matrix such that $\Theta = \Sigma^{-1}$, assuming that Σ^{-1} exists, and the notation 'det' refers to the determinant of the particular matrix.

The maximum likelihood is a popular method for estimating the optimal Θ. To do so, we have to calculate the log-likelihood function. Assuming n IID samples x_i, the log-likelihood function is given by:

$$\begin{aligned} \mathbb{L}(\Theta) &= \frac{1}{n}\sum_{i=1}^{n}\log f(x_i) = \frac{1}{2}\log\Big(\det(\Theta)\Big) - \frac{1}{2n}\sum_{i=1}^{n}x_i^T\Theta x_i \\ &\propto \log\Big(\det(\Theta)\Big) - \mathrm{tr}(S\Theta) \end{aligned}$$

where $S = \frac{1}{n}\sum_{i=1}^{n}x_i x_i^T$ is the sample variance-covariance matrix and the notation 'tr' refers to the trace of the particular matrix. The derivation of the above quantity is given in appendix E.2. Then, the optimal Θ matrix is given by maximising the penalised log-

likelihood function based on a ℓ_1-norm (see appendix E.1) penalty:

$$\widehat{\Theta} = \arg\max_{\Theta \succ 0} \left(\log\Big(\det(\Theta)\Big) - \mathrm{tr}(S\Theta) - \lambda \sum_{j=1}^{p} ||\Theta_j||_1 \right) \tag{5.3}$$

where Θ_j is the j-th column-vector of the Θ matrix and λ stands for the non-negative tuning parameter which determines the magnitude of the penalty term. The basic estimation methods for the tuning parameter are outlined in section 5.2.2. In appendix E.2, we show that Equation (5.3) describes a convex optimisation problem and the optimal Θ parameter is given by solving the gradient equation:

$$W - S - \lambda\Gamma = 0_{p\times p} \tag{5.4}$$

where $W = \Theta^{-1}$, assuming that Θ^{-1} is defined. In terms of Γ, we explain in appendix E.2 that it satisfies:

$$\Gamma_{ij} : \begin{cases} = 1, & \text{if } \Theta_{ij} > 0 \\ \in [-1, 1], & \text{if } \Theta_{ij} = 0 \\ = -1, & \text{if } \Theta_{ij} < 0 \end{cases}$$

where the subscripts i and j specify the elements of the Γ and Θ matrices.

The matrices W, S and Γ are of size $p \times p$ and they have the following block-matrix structure:

$$W = \begin{bmatrix} W_{11} & w_{12} \\ w_{21}^T & W_{22} \end{bmatrix}, S = \begin{bmatrix} S_{11} & s_{12} \\ s_{21}^T & S_{22} \end{bmatrix}, \Gamma = \begin{bmatrix} \Gamma_{11} & \gamma_{12} \\ \gamma_{21}^T & \Gamma_{22} \end{bmatrix}$$

where $W_{11},\ S_{11},\ \Gamma_{11}$ are matrices of size $(p-1)\times(p-1)$, $w_{12},\ s_{12},\ \gamma_{12}$ are vectors of size $(p-1)\times 1$ and $W_{22},\ S_{22},\ \Gamma_{22}$ are scalars. Equation (5.4) can be blockwisely optimised using standard mathematical operations on the blockwise inversion of a matrix. This is the logic behind a popular technique developed by Friedman et al. (2008) for the

optimisation of Equation (5.4). Briefly, the GLASSO algorithm of Friedman et al. (2008) optimises Equation (5.4) with respect to each row/column-vector of W (and hence of Θ, since $\Theta^{-1} = W$) through an iterative procedure until convergence. In Figure 5.2, we graphically represent how this algorithm operates to optimise the j-th row/column-vector of the matrix W for $j = 1, ..., 4$. Also, the exact algorithm's steps are summarised in ALGORITHM 5 in appendix E.2.1, given the tuning parameter λ. The efficiency of this algorithm depends on the assumption that $p < n$. A crucial step in constructing an accurate dependence network is the estimation of the optimal tuning parameter value. This topic is overviewed in the following section.

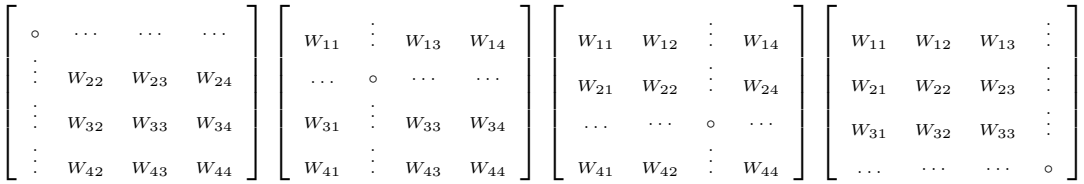

Figure 5.2: This figure visualises the iterative optimisation process applied to the row/-column vector of the W matrix until convergence. This example considers four variables.

5.2.2 Tuning Parameter Estimation in Graphical LASSO

Ultimately, the tuning parameter λ determines the size of the penalty. For a relatively high value of λ, the essential conditional dependencies are considered as independent cases. On the other hand, the model considers more than the necessary conditional dependencies for a relatively low value of λ. Hence, we require a tuning parameter that reveals merely the critical dependencies.

The tuning parameter in GLASSO can be estimated using several methods and CV is one of them. The scope of CV is the estimation of that λ_k, for $k = 1, \ldots, K$, which

minimises the prediction error over the groups g:

$$PE_g(\lambda_k) = \text{tr}\Big(S^{(test)}\widehat{\Theta}(\lambda_k)\Big) - \log\left(\det\left(\widehat{\Theta}(\lambda_k)\right)\right)$$

based on the train and the test data where $g = 1, \ldots, G$. The equation of the prediction error above is derived from the Formula (5.3).

Alternative ways to estimate the tuning parameter are through the AIC and BIC score functions:

$$AIC(\lambda_k) = -2\mathbb{L}\Big(\widehat{\Theta}(\lambda_k)\Big) + 2df(\lambda_k),$$
$$BIC(\lambda_k) = -2\mathbb{L}\Big(\widehat{\Theta}(\lambda_k)\Big) + \log(n)df(\lambda_k)$$

where $\mathbb{L}\Big(\widehat{\Theta}(\lambda_k)\Big)$ indicates the log-likelihood function of Θ evaluated at the estimates, given λ_k, $df(\lambda_k)$ refers to the number of non-zero elements in the upper diagonal of the $\widehat{\Theta}(\lambda_k)$ matrix and n refers to the number of observations. In the framework of GLASSO, the CV and BIC methods are studied by Lafit et al. (2019) along with a great deal of alternative approaches. Hastie et al. (2015) uses $\widehat{\lambda} = 2\sqrt{\frac{\log(p)}{n}}$ where p indicates the number of variables and n stands for the number of observations.

5.3 Functional Graphical Models

This section deals with the estimation of the variables' dependence structure when the dataset is of functional nature (Ramsay and Silverman, 2005; Ferraty and Vieu, 2006; Horváth and Kokoszka, 2012). The latter term refers to a set of variables where each observation is given as a curve instead of scalar values. In particular, we assume that the dataset is given by the p-dimensional variable Y_d for the day $d = 1, \ldots, D$, where $Y_d = (Y_{1,d}, ..., Y_{j,d}, ..., Y_{p,d})$. Further, we assume that each variable $Y_{j,d}$ is expressed in the form of curves. In this work, we assume that the dataset follows a noisy Gaussian

stochastic process.

Similar to the non-functional case, the variance-covariance matrix is directly related to the precision matrix, which defines the variables' conditional dependence structure. Qiao et al. (2019) develop a theoretical framework and methodology for estimating the precision matrix in the functional domain via the graphical models. Their work will be a guideline regarding the development of our approach. The methodology for defining the variance-covariance matrix is a distinctive feature between the two studies.

In their work, they reduce the dimension of the dataset by conducting a PCA under the functional regime. For the approximation of the variance-covariance matrix, they use the most critical curve coefficients instead of the data themselves. After that, they develop an algorithm for the prediction of the dependence network by solving the GLASSO optimisation function in (5.3) under relevant changes in the functional environment. Consequently, we can infer the conditional independence between the coefficient's vectors, not the underlying functions.

On the other hand, we estimate the variance-covariance matrix using all the covariance functions across the variables. We conduct a bivariate LPR to smooth the above matrix, allowing the matrix to keep its continuous characteristics. Then, we approximate the precision matrix using the algorithm suggested by Qiao et al. (2019). Our methodology exploits the full dataset to estimate the variables' conditional dependence structure as opposed to the methodology of Qiao et al. (2019) which only utilises the most essential coefficients of the functions. Thus, we expect that more information about the interdependencies will be incorporated into the estimation in our methodology, resulting in safer conclusions. Under the theoretical context of Qiao et al. (2019), Zapata et al. (2022) define the functional covariance function by considering a new structural form of the eigendecomposition.

Other relevant studies in the context of inference about the functional variables' dependence networks and under Gaussian settings are outlined below. Gómez et al.

(2020) deal with the concept of directed graphical models. They study two case scenarios when the structure of the model is known and when it is not. In the second case, they propose an algorithm which targets to minimise a loss function which is composed of a group-LASSO penalty (see Yuan and Lin (2006) for this technique) and a ℓ_2-norm (see appendix E.1) penalty in the estimation procedure. Applying both penalties results in better dependence estimates when multiple variables drive the patterns of a variable.

By the normality assumption, the predicted network corresponds to the partial correlation between the variables. Li and Solea (2018) relax this assumption to the non-linear and heteroscedastic case by considering the concept of additive conditional independence, developed by Li et al. (2014a). Later, Solea and Li (2022) estimate the most important functions' coefficients and they model the chosen coefficients through the Gaussian copula model (see Joe (2014) for this concept) to define the variables' dependence network, which technique also goes beyond the simple correlation structure. On the other hand, the studies of Zhu et al. (2016) and Lunagmez et al. (2021) deal with the concept of graphical models under Bayesian statistical assumptions.

5.3.1 Notation for Functional Variables

Standard functional data analysis considers a random curve as an element of L^2 space (Bosq, 2000; Horváth and Kokoszka, 2012). For a functional variable $X = \{X(t), t \in \mathcal{I} = [0,1]\}$ with $\mathbb{E}\,\|X\|^2 = \mathbb{E}[\int_0^1 X^2(t)\,dt] < \infty$, we write the mean and covariance functions as:

$$\mu(t) = \mathbb{E}[X(t)]\,, \quad c(s,t) = Cov\Big(X(s), X(t)\Big)\,,$$

and $Var[X(t)] = c(t,t)$ for $t \in \mathcal{I}$. We assume that the mean and covariance functions are continuous. If we evaluate at finite grid points $\tau_1, \ldots, \tau_J$, we write the corresponding

covariance matrix as:

$$C = \begin{pmatrix} c(\tau_1, \tau_1) & \dots & c(\tau_1, \tau_J) \\ \vdots & \vdots & \vdots \\ c(\tau_J, \tau_1) & \dots & c(\tau_J, \tau_J) \end{pmatrix}$$

and simply refer to functional covariance matrix.

Similarly, for a multivariate functional data $X = (X_1, \dots, X_p) = \Big\{ \Big(X_1(t), \dots, X_p(t) \Big), t \in [0,1] \Big\}$ with $\mathbb{E}[\int_0^1 X_j^2(t)\, dt] < \infty, j = 1, \dots, p$, the mean for each variable and the (cross-)covariance between a pair of variables are defined as:

$$\mu_j(t) = \mathbb{E}[X_j(t)]\,, \quad c_{j_1,j_2}(s,t) = Cov\Big(X_{j_1}(s), X_{j_2}(t) \Big)\,,$$

where $j_1 \neq j_2$, so the covariance function for the j-th variable is $Cov\Big(X_j(s), X_j(t) \Big) = c_{j,j}(s,t)$. The corresponding functional variance-covariance matrix with evaluation at finite grid points at $\tau_1, \dots, \tau_J$ has a block-matrix structure:

$$\mathbb{C} = \begin{pmatrix} C_{(1,1)} & \dots & C_{(1,p)} \\ \vdots & \ddots & \vdots \\ C_{(p,1)} & \dots & C_{(p,p)} \end{pmatrix}$$

where $C_{(j_1,j_2)}$, $j_1, j_2 = 1, \dots, p$, is the $J \times J$ cross-covariance function evaluated at finite grid points.

For D realizations of X denoted by $X_1, \dots, X_D$, $X_d = (X_{1,d}, \dots, X_{p,d})$ represents a p-dimensional functional variable. An additive noisy functional data can be expressed as:

$$Y_{j,d} = X_{j,d} + E_{j,d}, \quad j = 1, \dots, p;\ d = 1, \dots, D$$

where $E_{j,d}$ has mean zero and finite variance. In addition, if the data are only available at finite incidents, we can express as $Y_{j,d}(t_i)$ for $i = 1, \dots, n_{j,d}$.

5.3.2 Estimation of Functional Covariance Function

For the estimation of the conditional dependence structure, our first step is the estimation of the functional variance-covariance matrix since the approximation of the dependence network is based on the analysis of this matrix. For this reason, an accurate estimation would be essential. However, due to the specific traits of a dataset of functional nature, standard methods for time series data would result in mistaken matrix estimates. To approximate the variance-covariance matrix using functional data with irregularly sampled observations, we adopt the methodology proposed by Yao et al. (2005). This methodology was developed for one-dimensional functional data but can be easily extended to multivariate functional data as we consider here.

We assume that the i-th intraday observation of the j variable on the d day, $Y_{j,d}(t_i)$, can be expressed as a function of its corresponding true value $X_{j,d}(t_i)$ and an error term $E_{j,d}(t_i)$:

$$Y_{j,d}(t_i) = X_{j,d}(t_i) + E_{j,d}(t_i) \tag{5.5}$$

for $j = 1, ..., p$, $d = 1, ..., D$ where p denotes the number of variables, D indicates the number of days and $i = 1, ..., n_{jd}$. The number of observations, n, can vary over the variable j and the day d. We further consider that the error terms are IID, with mean zero and finite variance ω^2. In addition, we assume that the error processes are independent of the latent process X.

The covariance function for the variables j_1 and j_2 over the day d is given for:

$$Cov\Big(Y_{j_1,d}(t_{i_1}), Y_{j_2,d}(t_{i_2})\Big) = \mathbb{E}\left[\Big(Y_{j_1,d}(t_{i_1}) - \mu_{j_1,d}(t_{i_1})\Big)\Big(Y_{j_2,d}(t_{i_2}) - \mu_{j_2,d}(t_{i_2})\Big)\right] \tag{5.6}$$

where $\mu_{j_1,d}(t_{i_1}) = \mathbb{E}[Y_{j_1,d}(t_{i_1})]$ and $\mu_{j_2,d}(t_{i_2}) = \mathbb{E}[Y_{j_2,d}(t_{i_2})]$, for $j_1, j_2 = 1, \ldots, p$, $d =$

$1, \ldots, D$, $i_1 = 1, \ldots, n_{j_1,d}$ and $i_2 = 1, \ldots, n_{j_2,d}$. It is shown in appendix E.3 that:

$$Cov\Big(Y_{j_1,d}(t_{i_1}), Y_{j_2,d}(t_{i_2})\Big) = \begin{cases} Var[X_{j_1,d}(t_{i_1})] + \omega^2, & \text{if } j_1 = j_2 \text{ and } i_1 = i_2 \\ Cov\Big(X_{j_1,d}(t_{i_1}), X_{j_2,d}(t_{i_2})\Big), & \text{otherwise} \end{cases}.$$

For the development of our methodology, we make the following assumption.

Assumption 5.3.1 *The covariance function of $X_{j,d}$ does not depend on the days d but on the stocks.*

Assumption 5.3.1 considers a covariance function that does not vary much over the days (not time-dependent across days) but depends on the stocks. Therefore, we can pool all data and estimate the common covariance function across the days. If we are interested in time-dynamics we could apply the same methodology pooling the data over a shorter period of time and update the estimate regularly.

Regarding the covariance function estimation, the component $\mu_{j,d}(t_i)$ of Equation (5.6) can be estimated through the LPR, as discussed in section 4.2. Hence, an estimator of the covariance function over the day d can be constructed using the pseudo-data of the cross-product of the residuals:

$$V_{j_1,j_2,d}(t_{i_1}, t_{i_2}) = \Big(Y_{j_1,d}(t_{i_1}) - \widehat{\mu_{j_1,d}}(t_{i_1})\Big)\Big(Y_{j_2,d}(t_{i_2}) - \widehat{\mu_{j_2,d}}(t_{i_2})\Big). \tag{5.7}$$

This is justified because:

$$Cov\Big(X_{j_1,d}(t_{i_1}), X_{j_2,d}(t_{i_2})\Big) \approx \begin{cases} \mathbb{E}[V_{j_1,j_2,d}(t_{i_1}, t_{i_2})] + \omega^2, & \text{if } j_1 = j_2 \text{ and } i_1 = i_2 \\ \mathbb{E}[V_{j_1,j_2,d}(t_{i_1}, t_{i_2})], & \text{otherwise}. \end{cases}$$

The LPR technique can be utilised on the elements in (5.7) for the approximation of $\mathbb{E}[V_{j_1,j_2,d}]$. As opposed to Chapter 4 where we discuss the notion of LPR for a single variable, we now need to define the LPR in the bivariate space. The detailed methodology

of the bivariate local linear smoother is summarised later in section 5.3.3.

In particular, on the basis of Assumption 5.3.1, we follow the next steps in order to estimate the functional covariance matrix. For two variables j_1, j_2 we pool the pseudo-data of (5.7) for all D days into:

$$\widetilde{V}_{j_1,j_2} = \Big[V_{j_1,j_2,1} \ \cdots \ V_{j_1,j_2,d} \ \cdots \ V_{j_1,j_2,D}\Big]$$

where $\widetilde{V}_{j_1,j_2}$ is a vector such that $\widetilde{V}_{j_1,j_2} \in \mathbb{R}^N$ where $N = \sum_{d=1}^{D} n_{j_1,d}\ n_{j_2,d}$. This constitutes the response variable Y in (5.11) as a function of two coordinates (t_{i_1}, t_{i_2}). Then, we smooth $\widetilde{V}_{j_1,j_2}$ using the bivariate LPR for some pairs of evaluation points in the two-dimensional space.

In Chapter 4, the point approximation took place utilising as evaluation points the observation times. However, this is not a limitation and we can specify the time points we want to approximate our regression function. Considering the spatially random time points of our financial dataset, we choose to evaluate the above function at J specified points via the bivariate LPR method in favour of the analysis facilitation. Based on our high-frequency framework, we can use as evaluation points the grid points of J intraday intervals. This corresponds to the points from $\tau_1 = 1/J$ to $\tau_J = 1$ with step $1/J$ for every pair of variables. Under the new notational specification, the subscript i in the observation times t_i is replaced by r for the evaluation times τ_r such that $r = 1, \ldots, J$.

When $j_1 \neq j_2$, an estimate of the functional covariance function for these two variables at the evaluation grid points is given by:

$$\widehat{C}_{(j_1,j_2)} = \begin{pmatrix} \widehat{c_{j_1,j_2}}(\tau_1,\tau_1) & \widehat{c_{j_1,j_2}}(\tau_1,\tau_2) & \cdots & \widehat{c_{j_1,j_2}}(\tau_1,\tau_J) \\ \widehat{c_{j_1,j_2}}(\tau_2,\tau_1) & \widehat{c_{j_1,j_2}}(\tau_2,\tau_2) & \cdots & \widehat{c_{j_1,j_2}}(\tau_2,\tau_J) \\ \vdots & \vdots & \ddots & \vdots \\ \widehat{c_{j_1,j_2}}(\tau_J,\tau_1) & \widehat{c_{j_1,j_2}}(\tau_J,\tau_2) & \cdots & \widehat{c_{j_1,j_2}}(\tau_J,\tau_J) \end{pmatrix} \tag{5.8}$$

where $\widehat{c_{j_1,j_2}}(\tau_{r_1},\tau_{r_2})$ denotes the smoothed cross-product at time points (τ_{r_1},τ_{r_2}) for $r_1, r_2 = 1, ..., J$. Hence, $\widehat{C}_{(j_1,j_2)}$ is a matrix such that $\widehat{C}_{(j_1,j_2)} \in \mathbb{R}^{J\times J}$.

Putting all the estimates of the functional covariance functions in the same matrix offers the estimate of the functional variance-covariance matrix as follows:

$$\widehat{\mathbb{C}} = \begin{pmatrix} \widetilde{C}_{(1,1)} & \widehat{C}_{(1,2)} & \cdots & \widehat{C}_{(1,p)} \\ \widehat{C}_{(2,1)} & \widetilde{C}_{(2,2)} & \cdots & \widehat{C}_{(p,2)} \\ \vdots & \vdots & \ddots & \vdots \\ \widehat{C}_{(p,1)} & \widehat{C}_{(p,2)} & \cdots & \widetilde{C}_{(p,p)} \end{pmatrix} \tag{5.9}$$

where $\widehat{\mathbb{C}} \in \mathbb{R}^{Jp\times Jp}$. However, the diagonal elements of the above matrix include additional error. The way to eliminate this error is explained below.

Noise Elimination in the Functional Variance-Covariance Matrix

In the case where $j_1 = j_2$, the diagonal elements (variances) of the Matrix (5.8) contaminated by measurement error, as shown in appendix E.3. This noise is equal to the errors' variance and this subsection describes the steps we follow to remove this noise, as suggested by Yao et al. (2005).

Once we have estimated the Matrix (5.8) for $j_1 = j_2 = j$, we collect the diagonal elements of this matrix in a new vector B:

$$B = [\widehat{c_{j,j}}(\tau_1,\tau_1) \quad \widehat{c_{j,j}}(\tau_2,\tau_2) \quad \cdots \quad \widehat{c_{j,j}}(\tau_J,\tau_J)]$$

where $B \in \mathbb{R}^J_+$ and we replace the elements of the main diagonal in the estimated

functional covariance matrix by zero, obtaining:

$$\widetilde{C}_{(j,j)} = \begin{pmatrix} 0 & \widehat{c_{j,j}}(\tau_1, \tau_2) & \cdots & \widehat{c_{j,j}}(\tau_1, \tau_J) \\ \widehat{c_{j,j}}(\tau_2, \tau_1) & 0 & \cdots & \widehat{c_{j,j}}(\tau_2, \tau_J) \\ \vdots & \vdots & \ddots & \vdots \\ \widehat{c_{j,j}}(\tau_J, \tau_1) & \widehat{c_{j,j}}(\tau_J, \tau_2) & \cdots & 0 \end{pmatrix}.$$

We collect the cross-product of the residuals in Equation (5.7) which refer to the same observation times ($t_{i_1} = t_{i_2} = t_i$) over the D days for the variable j in the following vector:

$$Q = [V_{j,j,1}(t_i, t_i) \ \cdots \ V_{j,j,d}(t_i, t_i) \ \cdots \ V_{j,j,D}(t_i, t_i)]$$

where $V_{j,j,d}(t_i, t_i) = \Big(Y_{j,d}(t_i) - \widehat{\mu_{j,d}}(t_i)\Big)^2$, for $d = 1, ..., D$ and $i = 1, ..., n_{j,d}$. Then, we apply a (univariate) LPR on the Q vector using as evaluation points the same points we smoothed the Matrix (5.8); i.e. $\tau_1 = 1/J$ to $\tau_J = 1$ with step $1/J$. We illustrate the smoothed values of Q by $\widetilde{Q}$.

An estimator of ω^2 can be expressed by:

$$\widehat{\omega^2} = \int_t \Big(\widetilde{Q}(t) - B(t)\Big) dt$$

and ω^2 can be approximated by:

$$\widehat{\omega^2} = \frac{1}{(3J/4 - J/4)} \sum_{r=\lfloor J/4 \rfloor}^{\lfloor 3J/4 \rfloor} \Big(\widetilde{Q}(\tau_r) - B(\tau_r)\Big).$$

where we set $\widehat{\omega^2} = 0$, if $\widehat{\omega^2} < 0$. For the estimation of ω^2, we ignore the elements placed at the beginning and the end of the vectors $\widetilde{Q}(\tau)$ and $B(\tau)$ in order to avoid the boundary effects. Generally, when the weight allocation takes place for the point approximation close to the boundary, the weighting scheme is applied to insufficient and spatially random -in time- observations, affecting the accuracy of the estimation at these

points. For example, when the understudy point is the first observation, the weighting scheme will consider only the later observation for the point approximation, reducing the length of the local neighbourhood by one-half.

Consequently, when $j_1 = j_2 = j$ an estimate of the functional covariance matrix is given by:

$$\widehat{C}_{(j,j)} = \begin{pmatrix} \widetilde{Q}(\tau_1) - \widehat{\omega^2} & \widehat{c_{j,j}}(\tau_1, \tau_2) & \cdots & \widehat{c_{j,j}}(\tau_1, \tau_J) \\ \widehat{c_{j,j}}(\tau_2, \tau_1) & \widetilde{Q}(\tau_2) - \widehat{\omega^2} & \cdots & \widehat{c_{j,j}}(\tau_2, \tau_J) \\ \vdots & \vdots & \ddots & \vdots \\ \widehat{c_{j,j}}(\tau_J, \tau_1) & \widehat{c_{j,j}}(\tau_J, \tau_2) & \cdots & \widetilde{Q}(\tau_J) - \widehat{\omega^2} \end{pmatrix}. \tag{5.10}$$

Hence, the complete estimate of the functional variance-covariance matrix is given by replacing every diagonal element in (5.9) by (5.10), for each $j = 1, \ldots, p$.

5.3.3 Bivariate LPR

This subsection discusses the smoothing approach through LPR when two variables are considered in the local linear regression. Similar to the univariate LPR, the purpose of this technique is the estimation of the conditional mean function $m(x) = \mathbb{E}[Y|X = x]$ of the response variable Y:

$$Y_i = m(\underline{X}_i) + \sigma(\underline{X}_i)\epsilon_i, \qquad i = 1, ..., N\,, \tag{5.11}$$

where $\underline{X}_i = (X_{1,i}, X_{2,i})$ and $m(\underline{X})$ is now a function of two variables $\underline{x} = (x_1, x_2)$, $\sigma(\underline{X})$ is the conditional volatility of Y, given the covariates $X = x$, and ϵ indicates the error terms.

For given points $\underline{x}$, the estimation of the optimal coefficients is provided by minimising the following:

$$\sum_{i=1}^{N} \Big(Y_i - \beta_0 - \beta_1(X_{1,i} - x_1) - \beta_2(X_{2,i} - x_2) \Big)^2 K_{h_1}(X_{1,i} - x_1) \cdot K_{h_2}(X_{2,i} - x_2) \tag{5.12}$$

with respect to coefficient $\beta_{\underline{x}} = (\beta_0, \beta_1, \beta_2)^T$ where:

$$\beta_0 = m(x), \qquad \beta_1 = \frac{\partial m(x)}{\partial x_1}, \qquad \beta_2 = \frac{\partial m(x)}{\partial x_2}.$$

In Formula (5.12), K_h indicates the kernel function and h stands for the bandwidth, such that $h > 0$. The estimation of the coefficients at given points $\underline{x}$ is provided by:

$$\widehat{\beta}_{\underline{x}} = (X_{\underline{x}}^T W_{\underline{x}} X_{\underline{x}})^{-1} X_{\underline{x}}^T W_{\underline{x}} Y,$$

where $X_{\underline{x}} = [1|X_1 - x_1|X_2 - x_2]$ is the $N \times 3$ design matrix, W is the $N \times N$ matrix of kernel weights and Y is the response vector.

When applying to the problem of estimating the functional covariance matrix, the estimation of the coefficients is given at the evaluation points $\tau_0 = (\tau_{r_1}, \tau_{r_2})$. In this case, the matrix X_{τ_0} contains a vector of ones in the first column. The second column of this matrix contains the distance of the observed time points of the variable j_1 over the D days from the evaluation time point τ_1. Similarly, the third column contains the distance of the observed time points of the variable j_2 over the D days from the evaluation time point τ_2. We need to place the observed time points so that all possible combinations of the time points and the evaluation points for the two variables have been considered in the last two columns.

On the other hand, the Y vector contains the pseudo-data of the cross-product of the residuals (see Formula (5.7)), taken in respective order to correspond to the elements of the matrix X_{τ_0}. Figure 5.3.3 shows the form of the X_{τ_0} matrix and the Y vector. In this figure, the general notation $t_i^{(j,d)}$ refers to the i-th time point of the variable j for the day d, such that $i = 1, \ldots, n_{j,d}$, $j = 1, \ldots, p$ and $d = 1, \ldots, D$. In notation t_i, the superscript (j, d) is added compared to the standard notation in Equation (5.5). This happens for illustrative purposes, so we can discriminate the variable and the day the time points refer to. Besides, τ_r denotes the points we want to evaluate our function for

$r = 1, \ldots, J$.

As regards the kernel weights matrix W_{τ_0}, it is provided by:

$$W_{\tau_0} = \begin{pmatrix} w_{j_1,1}^{(h_1,r_1)}(t_1) \cdot w_{j_2,1}^{(h_2,r_2)}(t_1) & & & & 0 \\ & \ddots & & & \\ & & w_{j_1,d}^{(h_1,r_1)}(t_{i_1}) \cdot w_{j_2,d}^{(h_2,r_2)}(t_{i_2}) & & \\ & & & \ddots & \\ 0 & & & & w_{j_1,D}^{(h_1,r_1)}(t_{n_{j_1,d}}) \cdot w_{j_2,D}^{(h_2,r_2)}(t_{n_{j_2,d}}) \end{pmatrix}$$

where each weight component can be represented in the general form $w_{j,d}^{(h,r)}(t_i) = K_h(t_i^{(j,d)} - \tau_r)$, for a bandwidth h. The notations h_1, i_1 and r_1 show that the these elements refer to the variable j_1 in W_{τ_0}. Analogous notation holds when the elements refer to the variable j_2. In W_{τ_0}, we have considered that the two variables' weight components are independent for simplicity.

By mathematical matrix calculations (similar to those given in appendix D.1), the estimation of the coefficients at $\tau_0 = (\tau_{r_1}, \tau_{r_2})$ in the linear smoother are given for:

$$\begin{aligned}
\widehat{\beta}_0 &= \frac{(s_3 s_5 - s_4^2)u_0 + (s_2 s_4 - s_1 s_5)u_1 + (s_1 s_4 - s_2 s_3)u_2}{s_0 s_3 s_5 + 2(s_1 s_2 s_4) - s_0 s_4^2 - s_1^2 s_5 - s_2^2 * s_3}, \\
\widehat{\beta}_1 &= \frac{(s_2 s_4 - s_1 s_5)u_0 + (s_0 s_5 - s_2^2)u_1 + (s_1 s_2 - s_0 s_4)u_2}{s_0 s_3 s_5 + 2(s_1 s_2 s_4) - s_0 s_4^2 - s_1^2 s_5 - s_2^2 * s_3}, \\
\widehat{\beta}_2 &= \frac{(s_1 s_4 - s_2 s_3)u_0 + (s_1 s_2 - s_0 s_4)u_1 + (s_0 s_3 - s_1^2)u_2}{s_0 s_3 s_5 + 2(s_1 s_2 s_4) - s_0 s_4^2 - s_1^2 s_5 - s_2^2 * s_3}
\end{aligned}$$

where

$$\begin{aligned}
s_0 &= \sum_{d=1}^{D} \sum_{i_1=1}^{n_{j_1,d}} \sum_{i_2=1}^{n_{j_2,d}} w_{j_1,d}^{(h_1,r_1)}(t_{i_1}) \cdot w_{j_2,1}^{(h_2,r_2)}(t_{i_2}), \\
s_1 &= \sum_{d=1}^{D} \sum_{i_1=1}^{n_{j_1,d}} \sum_{i_2=1}^{n_{j_2,d}} w_{j_1,d}^{(h_1,r_1)}(t_{i_1}) \cdot w_{j_2,1}^{(h_2,r_2)}(t_{i_2}) \cdot (t_{i_1}^{(j_1,d)} - \tau_{r_1}), \\
s_2 &= \sum_{d=1}^{D} \sum_{i_1=1}^{n_{j_1,d}} \sum_{i_2=1}^{n_{j_2,d}} w_{j_1,d}^{(h_1,r_1)}(t_{i_1}) \cdot w_{j_2,1}^{(h_2,r_2)}(t_{i_2}) \cdot (t_{i_2}^{(j_1,d)} - \tau_{r_2}),
\end{aligned}$$

$$X_{\tau_0} = \begin{pmatrix}
1 & t_1^{(j_1,1)} - \tau_{r_1} & t_1^{(j_2,1)} - \tau_{r_2} \\
1 & t_1^{(j_1,1)} - \tau_{r_1} & t_2^{(j_2,1)} - \tau_{r_2} \\
\vdots & \vdots & \vdots \\
1 & t_1^{(j_1,1)} - \tau_{r_1} & t_{n_{j_2,1}}^{(j_2,1)} - \tau_{r_2} \\
1 & t_2^{(j_1,1)} - \tau_{r_1} & t_1^{(j_2,1)} - \tau_{r_2} \\
1 & t_2^{(j_1,1)} - \tau_{r_1} & t_2^{(j_2,1)} - \tau_{r_2} \\
\vdots & \vdots & \vdots \\
1 & t_2^{(j_1,1)} - \tau_{r_1} & t_{n_{j_2,1}}^{(j_2,1)} - \tau_{r_2} \\
\vdots & \vdots & \vdots \\
1 & t_{n_{j_1,1}}^{(j_1,1)} - \tau_{r_1} & t_1^{(j_2,1)} - \tau_{r_2} \\
1 & t_{n_{j_1,1}}^{(j_1,1)} - \tau_{r_1} & t_2^{(j_2,1)} - \tau_{r_2} \\
\vdots & \vdots & \vdots \\
1 & t_{n_{j_1,1}}^{(j_1,1)} - \tau_{r_1} & t_{n_{j_2,1}}^{(j_2,1)} - \tau_{r_2} \\
1 & t_1^{(j_1,2)} - \tau_{r_1} & t_1^{(j_2,2)} - \tau_{r_2} \\
1 & t_1^{(j_1,2)} - \tau_{r_1} & t_2^{(j_2,2)} - \tau_{r_2} \\
\vdots & \vdots & \vdots \\
1 & t_2^{(j_1,2)} - \tau_{r_1} & t_{n_{j_2,2}}^{(j_2,2)} - \tau_{r_2} \\
1 & t_2^{(j_1,2)} - \tau_{r_1} & t_1^{(j_2,2)} - \tau_{r_2} \\
1 & t_2^{(j_1,2)} - \tau_{r_1} & t_2^{(j_2,2)} - \tau_{r_2} \\
\vdots & \vdots & \vdots \\
1 & t_2^{(j_1,2)} - \tau_{r_1} & t_{n_{j_2,2}}^{(j_2,2)} - \tau_{r_2} \\
\vdots & \vdots & \vdots \\
1 & t_{n_{j_1,2}}^{(j_1,2)} - \tau_{r_1} & t_1^{(j_2,2)} - \tau_{r_2} \\
1 & t_{n_{j_1,2}}^{(j_1,2)} - \tau_{r_1} & t_2^{(j_2,2)} - \tau_{r_2} \\
\vdots & \vdots & \vdots \\
1 & t_{n_{j_1,2}}^{(j_1,2)} - \tau_{r_1} & t_{n_{j_2,2}}^{(j_2,2)} - \tau_{r_2} \\
\vdots & \vdots & \vdots \\
1 & t_{n_{j_1,D}}^{(j_1,D)} - \tau_{r_1} & t_1^{(j_2,D)} - \tau_{r_2} \\
1 & t_{n_{j_1,D}}^{(j_1,D)} - \tau_{r_1} & t_2^{(j_2,D)} - \tau_{r_2} \\
\vdots & \vdots & \vdots \\
1 & t_{n_{j_1,D}}^{(j_1,D)} - \tau_{r_1} & t_{n_{j_2,D}}^{(j_2,D)} - \tau_{r_2}
\end{pmatrix} \qquad Y = \begin{pmatrix}
V_{j_1,j_2,1}(t_1,t_1) \\
V_{j_1,j_2,1}(t_1,t_2) \\
\vdots \\
V_{j_1,j_2,1}(t_1,t_{n_{j_2,1}}) \\
V_{j_1,j_2,1}(t_2,t_1) \\
V_{j_1,j_2,1}(t_2,t_2) \\
\vdots \\
V_{j_1,j_2,1}(t_2,t_{n_{j_2,1}}) \\
\vdots \\
V_{j_1,j_2,1}(t_{n_{j_1,1}},t_1) \\
V_{j_1,j_2,1}(t_{n_{j_1,1}},t_2) \\
\vdots \\
V_{j_1,j_2,1}(t_{n_{j_1,1}},t_{n_{j_2,1}}) \\
V_{j_1,j_2,2}(t_1,t_1) \\
V_{j_1,j_2,2}(t_1,t_2) \\
\vdots \\
V_{j_1,j_2,2}(t_1,t_{n_{j_2,2}}) \\
V_{j_1,j_2,2}(t_2,t_1) \\
V_{j_1,j_2,2}(t_2,t_2) \\
\vdots \\
V_{j_1,j_2,2}(t_2,t_{n_{j_2,2}}) \\
\vdots \\
V_{j_1,j_2,2}(t_{n_{j_1,2}},t_1) \\
V_{j_1,j_2,2}(t_{n_{j_1,2}},t_2) \\
\vdots \\
V_{j_1,j_2,2}(t_{n_{j_1,2}},t_{n_{j_2,2}}) \\
\vdots \\
V_{j_1,j_2,D}(t_{n_{j_1,D}},t_1) \\
V_{j_1,j_2,D}(t_{n_{j_1,D}},t_2) \\
\vdots \\
V_{j_1,j_2,D}(t_{n_{j_1,D}},t_{n_{j_2,D}})
\end{pmatrix}$$

Figure 5.3: A visual representation of the matrix X_{τ_0} and the vector Y in the bivariate local linear regression.

$$s_3 = \sum_{d=1}^{D} \sum_{i_1=1}^{n_{j_1,d}} \sum_{i_2=1}^{n_{j_2,d}} w_{j_1,d}^{(h_1,r_1)}(t_{i_1}) \cdot w_{j_2,1}^{(h_2,r_2)}(t_{i_2}) \cdot (t_{i_1}^{(j_1,d)} - \tau_{r_1})^2,$$

$$s_4 = \sum_{d=1}^{D} \sum_{i_1=1}^{n_{j_1,d}} \sum_{i_2=1}^{n_{j_2,d}} w_{j_1,d}^{(h_1,r_1)}(t_{i_1}) \cdot w_{j_2,1}^{(h_2,r_2)}(t_{i_2}) \cdot (t_{i_1}^{(j_1,d)} - \tau_{r_1}) \cdot (t_{i_2}^{(j_1,d)} - \tau_{r_2}),$$

$$s_5 = \sum_{d=1}^{D} \sum_{i_1=1}^{n_{j_1,d}} \sum_{i_2=1}^{n_{j_2,d}} w_{j_1,d}^{(h_1,r_1)}(t_{i_1}) \cdot w_{j_2,1}^{(h_2,r_2)}(t_{i_2}) \cdot (t_{i_2}^{(j_1,d)} - \tau_{r_2})^2,$$

$$u_0 = \sum_{d=1}^{D} \sum_{i_1=1}^{n_{j_1,d}} \sum_{i_2=1}^{n_{j_2,d}} w_{j_1,d}^{(h_1,r_1)}(t_{i_1}) \cdot w_{j_2,1}^{(h_2,r_2)}(t_{i_2}) \cdot V_{j_1,j_2,d}(t_{i_1}, t_{i_2}),$$

$$u_1 = \sum_{d=1}^{D} \sum_{i_1=1}^{n_{j_1,d}} \sum_{i_2=1}^{n_{j_2,d}} w_{j_1,d}^{(h_1,r_1)}(t_{i_1}) \cdot w_{j_2,1}^{(h_2,r_2)}(t_{i_2}) \cdot (t_{i_1}^{(j_1,d)} - \tau_{r_1}) \cdot V_{j_1,j_2,d}(t_{i_1}, t_{i_2}),$$

$$u_2 = \sum_{d=1}^{D} \sum_{i_1=1}^{n_{j_1,d}} \sum_{i_2=1}^{n_{j_2,d}} w_{j_1,d}^{(h_1,r_1)}(t_{i_1}) \cdot w_{j_2,1}^{(h_2,r_2)}(t_{i_2}) \cdot (t_{i_2}^{(j_1,d)} - \tau_{r_2}) \cdot V_{j_1,j_2,d}(t_{i_1}, t_{i_2})$$

where $r_1, r_2 = 1, \ldots, J$, for given variables j_1, j_2 and bandwidths h_1, h_2. However, only the β_0 estimation concerns this study.

Bandwidth Selection in Bivariate LPR

The CV approach is conducted to estimate the optimal bandwidth in the bivariate case of LPR. Our scope is to find the bandwidth value which minimises the average prediction error based on a grid of bandwidth values. This method follows the same logic as that presented in the univariate case of section 4.2.2, and ALGORITHM 2 provides the full steps.

5.3.4 Estimation of the Inverse of the Variance-Covariance Matrix

For functional variables, the inverse of the covariance function is not well defined. We explain how this is estimated based on functional principal component analysis, the basis of which is Mercer's Lemma (Bosq, 2000; Horváth and Kokoszka, 2012).

Lemma 5.3.1 *(Mercer's lemma) Let c be a covariance function continuous over $[0,1]^2$.*

Algorithm 2 k-fold Cross-Validation for Bandwidth Estimation in Bivariate LPR

- Specify $[h_{min}, h_{max}]$;
- Divide our dataset into G groups;

for h_k, where $k = 1, ..., K$ **do**:

for group g, where $g = 1, ..., G$, **do**:

- Remove group g from X_1, X_2, Y (test data) and keep the rest (train data);
- Specify as evaluation points the observation times $t_1^{(test)}$ and $t_2^{(test)}$, of the test data of X_1 and X_2, respectively ;
- Estimate the regression function using the understudy h_k, the train data and the specified evaluation points ;
- Calculate the Prediction Error:

$$PE_g(\lambda_k) = mean\left(\left(Y^{(test)} - \widehat{m}^{(-g)}\right)^2\right);$$

end for

- Calculate the mean error for the given h_k by:

$$CV(h_k) = \frac{1}{G}\sum_{g=1}^{G} PE_g(h_k);$$

end for

- The optimal bandwidth is chosen through:

$$\widehat{h_{CV}} = \arg\min_{h_k>0} CV(h_k).$$

Then there exists a sequence $(\psi_\ell)_{\ell=1}^{\infty}$ *of continuous functions and a decreasing sequence* $(\lambda_\ell)_{\ell=1}^{\infty}$ *of positive numbers such that:*

$$\int_0^1 c(s,t)\psi_\ell(s)\,ds = \lambda_\ell\psi_\ell(t)\,, \quad t \in [0,1], \quad \ell \geq 1, \tag{5.13}$$

and

$$\int_0^1 \psi_\ell(s)\psi'_\ell(s)\,ds = \delta_{\ell,\ell'} \quad \ell, \ell' \geq 1\,.$$

Moreover,

$$c(s,t) = \sum_{\ell=1}^{\infty} \lambda_\ell\psi_\ell(s)\psi_\ell(t)\,, \quad s,t \in [0,1]\,,$$

where the series converges uniformly on $[0,1]^2$*; hence:*

$$\sum_{\ell=1}^{\infty} \lambda_\ell = \int_0^1 c(t,t)\, dt < \infty\,.$$

The pair $(\lambda_\ell, \psi_\ell)$ are the eigenvalues and the eigenfunctions of the covariance operator $\mathcal{C}$, defined by:

$$\mathcal{C}(x) = \mathbb{E}[\langle X - \mu, x\rangle (X - \mu)]\,, \quad x \in L^2.$$

By Mercer's lemma, one obtains the decomposition:

$$\mathcal{C}(x) = \sum_{\ell=1}^{\infty} \lambda_\ell \langle x, \psi_\ell\rangle \psi_\ell\,, \quad x \in L^2\,.$$

The eigendecomposition of c implies that the inverse operator $\mathcal{C}^{-1}$ which satisfies $\mathcal{C}^{-1}\mathcal{C}(x) = x$ may be defined as:

$$\mathcal{C}^{-1}(y) = \sum_{\ell=1}^{\infty} \lambda_\ell^{-1} \langle y, \psi_\ell\rangle \psi_\ell\,, \quad y \in L^2\,.$$

However, the inverse does not exist for all $y \in L^2$ and the sum converges only if:

$$\|\mathcal{C}^{-1}(y)\|^2 = \sum_{\ell=1}^{\infty} \lambda_\ell^{-2} \langle y, \psi_\ell\rangle^2 < \infty\,.$$

Instead, one can define a pseudo-inverse. Suppose that $\lambda_1 \geq \lambda_2 \geq \ldots \geq \lambda_L > 0$ and define:

$$\mathcal{C}_L^{-1}(y) = \sum_{\ell=1}^{L} \lambda_\ell^{-1} \langle y, \psi_\ell\rangle \psi_\ell\,.$$

The operator $\mathcal{C}_L^{-1}$ is defined on the whole of L^2. The sample counterpart of this pseudo-inverse is defined by:

$$\widehat{\mathcal{C}_L^{+}}(y) = \sum_{\ell=1}^{L} \widehat{\lambda}_\ell^{-1} \langle y, \widehat{\psi}_\ell\rangle \widehat{\psi}_\ell \tag{5.14}$$

where $\widehat{\lambda}_\ell, \hat{\psi}_\ell$ are the estimated eigenvalues and eigenfunctions of the empirical covariance operator $\widehat{\mathcal{C}}$. The same principle applies to both scalar and vector-valued functional variables.

In order to compute the eigenvalues and eigenfunctions of $\mathcal{C}$, we use the corresponding empirical covariance function $\widehat{c}$. Section 5.3.2 explains how to obtain the empirical covariance function $\widehat{c}$ and the corresponding matrix $\widehat{C}$ evaluated at finite grids $t_1, \ldots, t_J \in [0,1]$ with $\Delta = t_i - t_{i-1}$. In order to find the eigenvalues and eigenfunctions, we need to solve the Integral Equation (5.13) with $\widehat{c}$. Let $\boldsymbol{\psi} = \Big(\psi(t_1), \ldots, \psi(t_J)\Big)^\top$ be the vector of the evaluation of the function at the finite grids. The integral equation can be approximated by:

$$\widehat{C}\boldsymbol{\psi}_\ell \Delta = \lambda \boldsymbol{\psi}_\ell \,, \quad \boldsymbol{\psi}_\ell^\top \boldsymbol{\psi}_\ell' \Delta = \delta_{\ell,\ell'} \,.$$

Define $\widetilde{\boldsymbol{\psi}}_\ell = \Delta^{1/2}\boldsymbol{\psi}_\ell$. It follows that:

$$\widehat{C}\widetilde{\boldsymbol{\psi}}_\ell = \widetilde{\lambda}_\ell \widetilde{\boldsymbol{\psi}}_\ell \,, \quad \widetilde{\boldsymbol{\psi}}_\ell^\top \widetilde{\boldsymbol{\psi}}_\ell' = \delta_{\ell,\ell'}$$

where $\widetilde{\lambda}_\ell = \lambda_\ell / \Delta$. Thus, it is sufficient to compute the eigenvalues and eigenvectors $(\widetilde{\lambda}_\ell, \widetilde{\boldsymbol{\psi}_\ell})$ of the matrix $\widehat{C}$. Then, we can obtain:

$$\widehat{\boldsymbol{\psi}}_\ell = \Delta^{-1/2}\widetilde{\boldsymbol{\psi}}_\ell \,, \quad \widehat{\lambda}_\ell = \Delta \widetilde{\lambda}_\ell \,, \quad \ell \geq 1 \,.$$

Let $\widehat{\lambda}_1 \geq \widehat{\lambda}_2 \geq \ldots \geq \widehat{\lambda}_L > 0$ be the positive eigenvalues. From (5.14), the corresponding element of the inverse matrix $\widehat{C}^{-1}$ is estimated by:

$$\widehat{c}_L^+(t_i, t_j) = \sum_{\ell=1}^{L} \widehat{\lambda}_\ell^{-1} \widehat{\psi}_\ell(t_i) \widehat{\psi}_\ell(t_j) \,.$$

The determinant of $\widehat{C}$ is computed by:

$$\det(\widehat{C}) = \prod_{\ell=1}^{L} \widehat{\lambda}_\ell \, .$$

5.3.5 Functional GLASSO Algorithm

This section analyses the steps for estimating the precision matrix Θ under the functional regime. This matrix demonstrates the conditional dependence form of the variables. The algorithm we use in this study and its derivation were developed by Qiao et al. (2019) who used the acronym FGLASSO, which stands for the initials of Functional Graphical LASSO. The derivation of this algorithm follows a similar rationale as the algorithm of Friedman et al. (2008) in the case of functional data. The reader can find the complete derivation in appendix E.4 of this work. In a nutshell, similar to the graphical LASSO approach, presented in subsection 5.2.1, the purpose of this technique is the maximisation of the following penalised log-likelihood with respect to Θ:

$$\widehat{\Theta} = \arg\max_{\Theta} \left(\log\Big(\det(\Theta)\Big) - \operatorname{tr}(S\Theta) - \lambda \sum_{j \neq l} ||\Theta_{jl}||_F \right) \tag{5.15}$$

where Θ is the precision matrix, S stands for the estimated functional variance-covariance matrix, given by (5.9), and $||\cdot||_F$ denotes the Frobenius norm. Our approach exploits the full data information since all data contribute to the estimation of S and Θ, as opposed to the method of Qiao et al. (2019) who merely utilise the most significant coefficients of the data curves. Therefore, we expect our estimated functional variance-covariance matrix to result in better inference regarding the variables' conditional dependence network. This is because more information about the interdependencies among the variables is considered in the estimation.

The precision matrix Θ and the estimate of the functional variance-covariance matrix

S have the following form:

$$\Theta = \begin{bmatrix} \Theta_{jj} & \theta_j^T \\ \theta_j & \Theta_{-j} \end{bmatrix}, S = \begin{bmatrix} S_{jj} & s_j^T \\ s_j & S_{-j} \end{bmatrix}.$$

A visual representation of the Θ matrix is given in Figure (5.4) while the matrix S has an identical structure. In this representation, J denotes the number of evaluation points we have considered in the bivariate LPR (the grid points of J intraday intervals) and p denotes the number of variables.

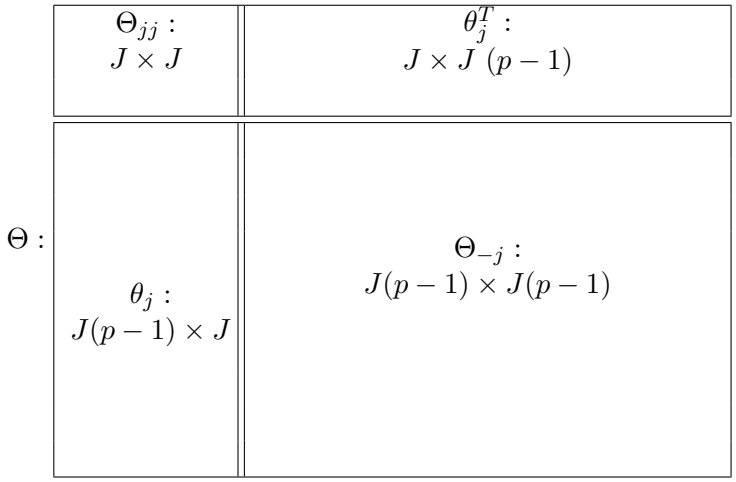

Figure 5.4: An illustration of the structure of the precision matrix Θ.

The solution of (5.15) over the fixed j row/column block-vector is given by:

$$\widehat{\Theta}_{jj} = S_{jj}^{-1} + \widehat{\theta}_j^T \Theta_{-j}^{-1} \widehat{\theta}_j \,,$$

for such $\widehat{\theta}_j$:

$$\widehat{\theta}_j = \arg\min_{\theta_j} \left(\mathrm{tr}(S_{jj} \theta_j^T \Theta_{-j}^{-1} \theta_j) + 2\mathrm{tr}(s_j^T \theta_j) + 2\lambda \sum_{l=1}^{p-1} ||\theta_{jl}||_F \right) \tag{5.16}$$

where s_j and θ_j denote the j-th column block-vector of the matrices S and Θ, respectively. Both vectors are composed of $(p-1)$ matrices of size $J \times J$ each. Also, $||\theta_{jl}||_F$ denotes the Frobenius norm (see appendix E.1) of the l-th matrix in θ_j. Appendix E.4.3 reports the derivation of (5.16). The zero block-matrices in $\widehat{\Theta}$ show the conditional independence cases for the corresponding variables.

The detailed steps of the FGLASSO algorithm are presented in ALGORITHM 3, and standard calculations on blockwise inversion of a matrix are utilised for the derivation. In ALGORITHM 3, the $\widehat{\Sigma}$ matrix has the same form as given in Figure (5.4). In addition, the notation $\{\Theta_{-j}^{-1}\}_{kk} \bigotimes S_{jj}$ indicates the Kronecker product (see appendix E.1) between the inverse of the k-th matrix on the main diagonal of Θ_{-j} and the matrix S_{jj}. Also, r_{jk} denotes the block-residual, which is a $J \times J$ matrix, and λ denotes the tuning parameter, such that $\lambda > 0$. Besides, the acronym *vec* demonstrates that the particular matrix has been transformed into a vector.

The algorithm only requires the estimated variance-covariance matrix S as an input to begin with, which is obtained by the Matrix (5.9) in our case. For updating steps, we need to be able to compute the inverse of the part of the matrix. One of the main difficulties with direct approximation of functional variables is that the dimension of the matrix is very large, so the inverse matrices $\widehat{\Sigma}_{jj}^{-1}$ and Θ_{-j}^{-1} are not well defined. To overcome this issue, we have used the approximation based on the eigendecomposition explained in section 5.3.4 to estimate $\widehat{\Sigma}_{jj}^{-1}$, then $\widehat{\Theta}_{-j}^{-1}$ is given by Equation (5.17).

5.3.6 Tuning Parameter Estimation in Functional Graphical LASSO

Optimal tuning parameter estimation is of great importance since it controls which dependencies will be considered negligible by the graphical model. For a relatively high λ, more than the necessary non-zero conditional dependencies will be set equal to zero. On the other hand, the model will not shrink the minor conditional dependencies to zero for a relatively low λ. Hence, the optimal λ has to be of that size that keeps the meaningful

Algorithm 3 FGLASSO Algorithm

• Initialise $\widehat{\Theta} = I_{Jp}$ and $\widehat{\Sigma} = I_{Jp}$;
for $iter = 1, \ldots$, until convergence **do**:
for $j = 1, ..., p$ **do**:
• Estimation of Θ_{-j}^{-1} through:

$$\widehat{\Theta}_{-j}^{-1} = \widehat{\Sigma}_{-j} - \widehat{\sigma}_j \widehat{\Sigma}_{jj}^{-1} \widehat{\sigma}_j^T; \tag{5.17}$$

• Initialise $\widehat{\theta}_j$, e.g.:

$$\widehat{\theta}_j = 0_{J(p-1)\times J} + 1e - 10;$$

for $k = 1, ..., p-1$ **do**:
• Estimate block-residual matrix r_{jk}:

$$\widehat{r}_{jk} = \sum_{l \neq k} \{\Theta_{-j}^{-1}\}_{lk}^T \theta_{jl} S_{jj} + s_{jk};$$

if $||\widehat{r}_{jk}||_F \leq \lambda$ **then**

$$\widehat{\theta}_{jk} = 0_{J\times J};$$

else
$\widehat{\theta}_{jk}$ is given for that θ_{jk} which solves:

$$\left(\{\Theta_{-j}^{-1}\}_{kk} \bigotimes S_{jj} \right) vec(\theta_{jk}) + vec(r_{jk}) + \lambda \frac{vec(\theta_{jk})}{||\theta_{jk}||_F} = 0;$$

end if
end for
• Reformulate $\widehat{\Theta}$:

$$\widehat{\Theta} = \begin{pmatrix} \widehat{\Theta}_{jj} & \widehat{\theta}_j^T \\ \widehat{\theta}_j & \widehat{\Theta}_{-j} \end{pmatrix};$$

• Reformulate $\widehat{\Sigma}$ applying:

$$\begin{aligned} \widehat{\Sigma}_{jj} &= S_{jj} \\ \widehat{\sigma}_j &= -U_j S_{jj} \\ \widehat{\Sigma}_{-j} &= \widehat{\Theta}_{-j}^{-1} + U_j S_{jj} U_j^T \end{aligned}$$

where $U_j = \widehat{\Theta}_{-j}^{-1} \widehat{\theta}_j$;
end for
if mean$(|\widehat{\Theta}^{(iter)} - \widehat{\Theta}^{(iter-1)}|) < 1E^{-3}$ **then**
Return Θ;
end if
where $\widehat{\Theta}^{(iter)}$ is the $\widehat{\Theta}$ matrix as results by the corresponding iteration and $\widehat{\Theta}^{(0)}$ refers to the initial $\widehat{\Theta}$ matrix.
end for

conditional dependencies discarding, at the same time, the trivial dependencies and simplifying the variables' dependence structure.

Qiao et al. (2019) and Gómez et al. (2020) state that the practitioner can apply alternative methods for the λ estimation, such as CV, GCV, AIC or BIC. In this work, we have conducted a k-fold CV based on the specific characteristics of our dataset. In particular, our technique takes as input the estimate of the functional variance-covariance matrix over different neighbouring periods, and it computes the prediction error based on the following formula:

$$PE_m(\lambda) = \text{tr}\Big(\widehat{\mathbb{C}}_m \cdot \widehat{\Theta}(\lambda)\Big) - \log\left(\det\left(\widehat{\Theta}(\lambda)\right)\right) \tag{5.18}$$

where $\widehat{\mathbb{C}}_m$ indicates the estimate of the functional variance-covariance matrix for the period m, such that $m = 1, ..., M$. For example, m can refer to weeks or months. Besides, Θ denotes the precision matrix for given λ. The formula of the prediction error is the log-likelihood component of the maximisation problem in (5.15), where it has been transformed into a minimisation problem by taking the opposite quantity. To select λ, we set an interval of grid points $[\lambda_{min}, \lambda_{max}]$ where λ_{min} is a value lower than the smallest Frobenius Norm of the submatrices in $\widehat{\mathbb{C}}$ for the period of our interest. Similarly, λ_{max} is chosen to be a high value so that it will shrink all conditional dependencies to zero in the Θ matrix. ALGORITHM 4 reports the complete steps. By this method, we aim to estimate the optimal λ which minimises the prediction error, exploiting similar conditional interdependencies among the variables across neighbouring periods.

5.4 Simulation Study

Our proposed methodology procedure is examined via a simulation study in this section. This study simulates an intraday observation every 15 minutes (arising 26 intraday intervals) for five and ten variables over six months. Also, we consider 20

Algorithm 4 k-fold Cross-Validation for GLASSO

- Specify $[\lambda_{min}, \lambda_{max}]$;

for λ_k, where $k = 1, ..., K$ **do**:

- Estimate Θ for the desired period through the FGLASSO algorithm;

for period m, where $m = 1, ..., M$, **do**:

- Calculate the Prediction Error with $\widehat{\Theta}$ and the matrix $\widehat{\mathbb{C}}$ for different periods:

$$PE_m(\lambda_k) = \mathrm{tr}\Big(\widehat{\mathbb{C}}_m \cdot \widehat{\Theta}(\lambda_k)\Big) - \log\left(\det\left(\widehat{\Theta}(\lambda_k)\right)\right);$$

end for

- Calculate the cross-validation score for the given λ_k by:

$$CV(\lambda_k) = \sum_{m=1}^{M} PE_m(\lambda_k);$$

end for

- The optimal tuning parameter is chosen by:

$$\widehat{\lambda_{CV}} = \arg\min_{\lambda_k \geq 0} CV(\lambda_k).$$

working days in a month period, and each day is simulated to be independent of the rest days across the sample period and the variables. The properties of the model are given in appendix E.5. Furthermore, the Epanechnikov kernel function is used in both the univariate and bivariate local linear smoother.

The reported technique is computationally expensive, so we examine our model for five and ten variables. This computational cost mainly stems from the selection procedure of the optimal bandwidth in the bivariate LPR, the estimation of the proper tuning parameter in the model and the optimisation process by iteration.

This study uses the simulation model considered in Yao et al. (2005), where several minor changes have been applied. More specifically, the true observation of the j-th variable on the d day is simulated through the formula:

$$X_{j,d}(t_i) = \mu_{j,d}(t_i) + \xi_{j,d}^{(1)}\phi_{j,d}^{(1)}(t_i) + \xi_{j,d}^{(2)}\phi_{j,d}^{(2)}(t_i)$$

where $j = 1, ..., p$ (for $p = 5$ and $p = 10$) and $d = 1, ..., 120$. The mean function $\mu_{j,d}(t_i)$ is formed by:

$$\mu_{j,d}(t_i) = t_i + \sin(t_i)$$

and the eigenfunctions are expressed as:

$$\phi_{j,d}^{(1)}(t_i) = \frac{-\cos(0.1 \cdot \pi \cdot t_i)}{\sqrt{5}},$$
$$\phi_{j,d}^{(2)}(t_i) = \frac{\sin(0.1 \cdot \pi \cdot t_i)}{\sqrt{5}}$$

where $1/26 \leq t_i \leq 6$ with step $1/26$ at a time.

For simplicity, we have considered that variable 1 is dependent on variables 2, 5 (using five variables) and variables 2, 5, 8, 9, 10 (using ten variables). The rest variables are independent of each other. The concept of the adjacency matrix is helpful in the context of graphical models. When a pair of variables are conditionally dependent, the corresponding matrix element is given as 1; otherwise, it is given as 0. Figure 5.5 shows a graphical representation of the specified dependence structure, and Figure 5.6 depicts the adjacency matrices for this structure. When $j = 1, 3, 4$ (for five variables) and $j = 1, 3, 4, 6, 7$ (for ten variables), the components $\xi_{j,d}^{(1)}$, $\xi_{j,d}^{(2)}$ for the variable j, are simulated through a univariate random normal distribution with zero mean and variance equal to 4 and 0.6, respectively. In the case where $j = 2, 5$ (for five variables) and $j = 2, 5, 8, 9, 10$ (for ten variables), $\xi_{j,d}^{(1)}$ and $\xi_{j,d}^{(2)}$ are specified through:

$$\xi_{1,d}^{(1)} = \rho\xi_{j,d}^{(1)} + \epsilon_{j,d}^{(1)},$$
$$\xi_{1,d}^{(2)} = \rho\xi_{j,d}^{(2)} + \epsilon_{j,d}^{(2)}$$

where ρ introduces the correlation between the processes. The correlation coefficient is set equal to $\rho = 0.6$ in this work. Moreover, $\epsilon_{j,d}^{(1)}$, $\epsilon_{j,d}^{(2)}$ incorporate some error in the process. These errors are simulated through a random normal distribution with zero

mean and variance $1 - \rho^2$.

Finally, we produce the noisy intraday observations through:

$$Y_{j,d}(t_i) = X_{j,d}(t_i) + E_{j,d}(t_i)$$

where the error terms follow a random normal distribution with zero mean and the variance equals 0.25.

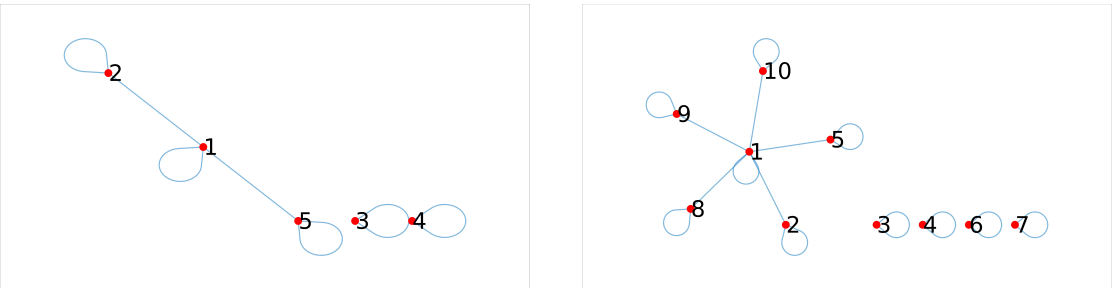

Figure 5.5: The variables dependence structure using five (left panel) and ten (right panel) variables.

Two measures that evaluate the proposed technique's precision are the True Positive Rate (TPR) and the False Positive Rate (FPR). Ultimately, TPR refers to the probability that the estimated adjacency matrix successfully captures the conditional dependence between two variables. On the other hand, FPR indicates the probability that two independent variables are estimated as conditionally dependent. These two rates are computed by:

$$TPR = \frac{TP}{TP + FN}, \qquad FPR = \frac{FP}{TN + FP}$$

where:

TP : conditional dependence cases estimated as conditional dependencies,

FN : conditional dependence cases estimated as conditional independencies,

$$Adj M_5 = \begin{pmatrix} 1 & 1 & 0 & 0 & 1 \\ 1 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 1 & 0 & 0 & 0 & 1 \end{pmatrix}, Adj M_{10} = \begin{pmatrix} 1 & 1 & 0 & 1 & 1 & 0 & 0 & 1 & 1 & 1 \\ 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{pmatrix}$$

Figure 5.6: The true adjacency matrices for five (left panel) and ten (right panel) variables.

FP : conditional independence cases estimated as conditional dependencies,

TN : conditional independence cases estimated as conditional independencies.

Both rates take values within the interval $[0, 1]$. For TPR, a value close to one shows that the estimated adjacency matrix captures the conditional dependencies between the variables with high accuracy, whereas a value near zero displays that the estimated adjacency reports the conditional dependencies as independent cases. As regards the FPR, a value close to zero indicates that the estimated adjacency matrix correctly captures the conditional independencies between the variables with high accuracy. In contrast, a value close to one shows that the estimated adjacency predicts the conditional independencies between the variables as dependencies. For both rates, a value close to 0.5 is a sign of random estimates for the estimated adjacency matrix. Therefore, a rate close to one is desirable for TPR. On the other hand, a rate near zero is required for FPR in terms of an accurate conditional dependence structure estimation.

These preferred values highly depend on the tuning parameter. However, a meagre λ value would result in TPR and FPR rates (almost) equal to one, and an extremely high λ value would imply that the values of TPR and FPR are (nearly) zero. An ideal

λ would bring about an adjacency matrix with a TPR rate equal to one and FPR equal to zero. In practice, this is often a somewhat unrealistic scenario, mainly because of the noise incorporated in the observations (yielding a noisy approximation of the functional variance-covariance matrix) and the selection of a suboptimal tuning parameter through the estimation methods.

Table 5.1 summarises the mean value of TPR and FPR over 100 iterations for five and ten variables. This table demonstrates that the proposed approach successively captures the conditional dependence between the variables since TPR is higher than 90% in both cases. On the other hand, FPR takes values 7.75% and 6.06% for five and ten variables, respectively. These results summarise that the proposed model, under these settings, is able to capture the conditional dependencies between the variables for a rate higher than 90%. Besides, it fails to capture the corresponding independencies by less than 8%, on average.

Rate / p	TPR	FPR
5	0.92	0.0775
10	0.905	0.0616

Table 5.1: The mean values of TPR and FPR over 100 iterations.

5.5 Empirical Analysis

This section investigates the conditional dependence structure of the 50 stocks of Table 1.2 for January 2012. In this analysis, we work both with the log-prices and the returns of the stocks. Because of the high computational cost of this method for a considerable number of observations, we have decided to specify an intraday frequency of 15 minutes in the procedure, resulting in 26 intraday intervals for each day.

For the estimation of the functional variance-covariance matrix, the evaluation points are set to be 26 equispaced points between 0 and 1. The Gaussian kernel function

is selected for the weights allocation through the local linear smoother for both the univariate and the bivariate cases. In the latter case, we consider the same bandwidth in the kernel functions K_{h_1} and K_{h_2}, i.e. $h_1 = h_2$. The CV method is used for the estimation of the optimal bandwidths (see sections 4.2.2, 5.3.3).

The CV method is also utilised to estimate the optimal tuning parameter in the functional GLASSO approach, as presented in section 5.3.6. Applying direct multivariate method can produce a negative value for the determinant of the $\widehat{\Theta}$ matrix, making its logarithmic value undefined for some tuning parameter values. For this reason, we apply an eigenanalysis to this matrix and define the determinant as the product of the positive eigenvalues of this matrix, as explained in section 5.3.4. The CV score is computed by aggregating the absolute value of the prediction errors for every tuning parameter. For the calculation of the prediction errors in (5.18), the $\widehat{\mathbb{C}}$ Matrix in (5.9) is specified for all the months from January to June of 2012. None of the stocks realised a stock split during the chosen months; hence our dataset does not need any pre-processing for the empirical investigation. In this analysis, we refer to the particular stocks by their identifier code for brevity and one can find their full names and related sectors in Table 1.2. Appendix E.6 contains supplementary results.

Log-Prices

In Figure 5.7, we present the stocks' conditional dependence structure for eight different tuning parameter values. As expected, the model provides more independent stock pairs as we increase the tuning parameter. In this figure, the missing stocks refer to those which are conditionally independent of the other stocks. These stocks are given in Figure E.2 in appendix E.6.1 for every tuning parameter case.

For the tuning parameters in Figure 5.7, we have applied a CV to infer which value offers the optimal conditional dependence structure. The CV method estimates $\lambda = 0.0826$ as the optimal tuning parameter, given by the bottom right panel. For this

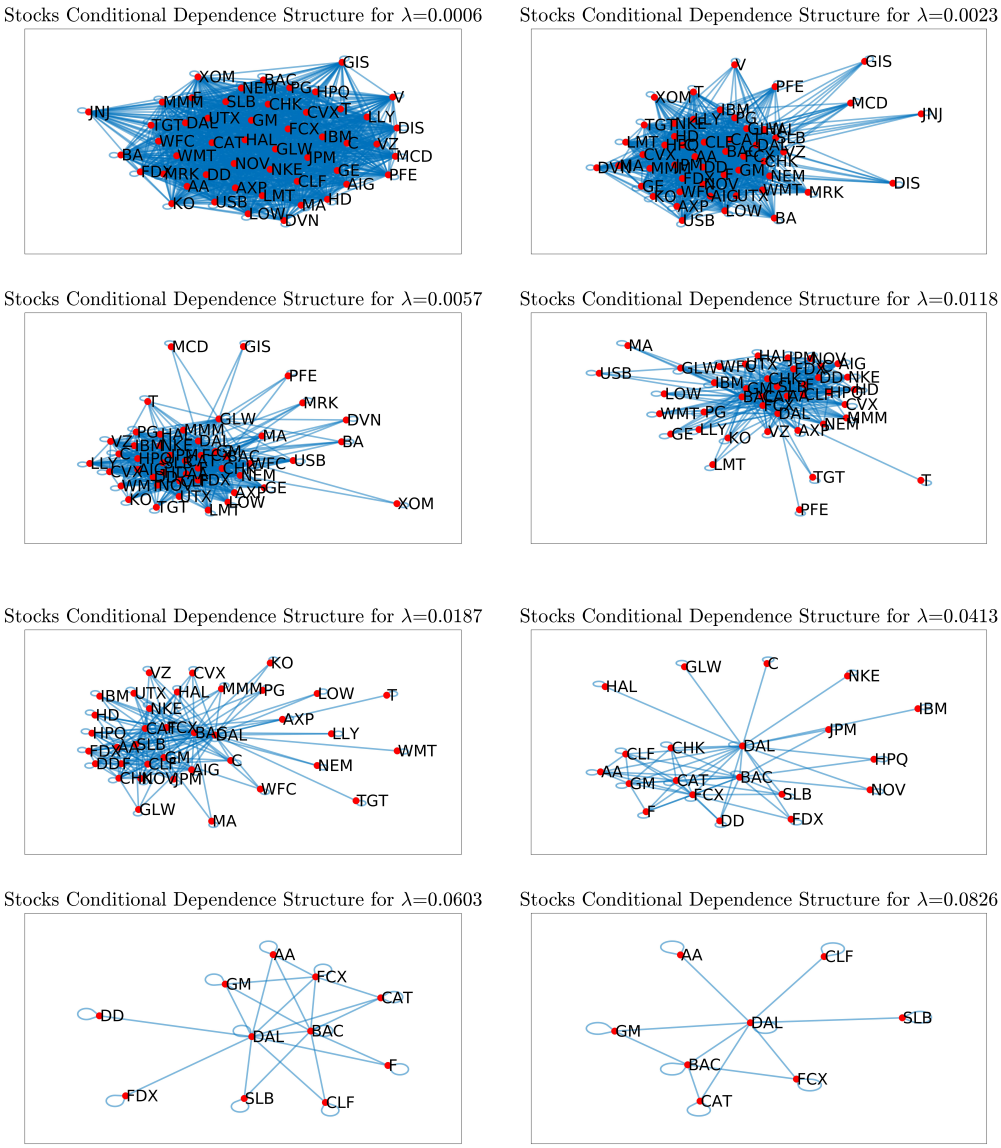

Figure 5.7: The conditional dependence structure of the log-prices of the stocks for eight different tuning parameter values.

value, we obtain six maximal cliques and they are defined for the following subgraphs:

$$\mathcal{E}(\mathcal{G}_1) = (< DAL, AA >),$$

$$\mathcal{E}(\mathcal{G}_2) = (< DAL, CLF >),$$

$$\mathcal{E}(\mathcal{G}_3) = (< DAL, SLB >),$$

$$\mathcal{E}(\mathcal{G}_4) = (< DAL, BAC >, < DAL, FCX >, < BAC, FCX >),$$

$$\mathcal{E}(\mathcal{G}_5) = (< DAL, BAC >, < DAL, CAT >, < BAC, CAT >),$$

$$\mathcal{E}(\mathcal{G}_6) = (< DAL, BAC >, < DAL, GM >, < BAC, GM >).$$

Except for the stocks DAL and CAT, which are related to the sector 'Industrials', all the other conditionally dependent stocks are related to different sectors. This finding is surprising since, generally, it is anticipated that stocks of the same sector would be highly correlated. Further, taking into consideration Figure 1.3 and Table A.1, which provide insights about the correlation coefficients of the above stocks, we notice that their coefficients are relatively weak.

The stocks BAC and DAL, which show the most dependencies in the estimated network, have the lowest prices in the examined period among the stocks. On the other hand, CAT and CLF are among the stocks with the highest prices. The fact that the two groups of stocks are highly connected suggest that the stocks with low price have the tendency to follow the patterns of stocks with higher prices. As the stock returns tend to be of similar size, we also investigate the conditional interdependencies with respect to returns.

Returns

In Figure 5.8, we visualise the conditional dependence structure of the stock returns, again for eight different tuning parameter values. Besides, the stocks which do not share conditional dependence with the other stocks are given in Figure E.3 in appendix E.6.2 for the chosen tuning parameter values. As also happens for the log-prices, the dependence network becomes more sparse as we increase the tuning parameter. Besides, the cliques given by the highest tuning parameter are also formed for the lower tuning parameter values.

The findings show more reasonable performance compared to the results with the log-prices. Indicatively, the resulting networks for the two highest tuning parameters mainly comprise stocks related to the same sector. For instance, the maximal cliques of the network with a tuning parameter equal to 0.4125 are given with the following subgraphs:

$$\mathcal{E}(\mathcal{G}_1) = (< SLB, HAL >),$$
$$\mathcal{E}(\mathcal{G}_2) = (< HAL, NOV >),$$
$$\mathcal{E}(\mathcal{G}_3) = (< BAC, C >, < BAC, WFC >, < C, WFC >),$$
$$\mathcal{E}(\mathcal{G}_4) = (< WFC, JPM >),$$
$$\mathcal{E}(\mathcal{G}_5) = (< AA, FCX >)$$

where the subgraphs $\mathcal{G}_1$ and $\mathcal{G}_2$ are related to the sector 'Energy', the following two subgraphs are related to the sector 'Financial Sevices' and the fifth subgraph to the sector 'Basic Materials'. In Table A.2, we notice that the correlation coefficient between AA and FCX is about 52%, while the correlation coefficients between the underlying stocks in subgraphs $\mathcal{G}_1 - \mathcal{G}_4$ exceeds 60%, as we observe in Tables A.6 and A.7 in appendix A.

CV estimates that the optimal conditional dependence structure is given for the graph with tuning parameter equal to 0.1602 (second row, left panel). This structure is more clearly depicted in Figure 5.9. The dependence structure offers a more complex structure compared to the proposed structure for the log-prices; however, this seems more natural given the interdependencies that occur in the market. Some indicative maximal cliques in this graph are given by:

$$\mathcal{E}(\mathcal{G}_1) = (< MMM, CLF >, < MMM, NOV >, < CLF, NOV >),$$
$$\mathcal{E}(\mathcal{G}_2) = (< AA, F >, < AA, HAL >, < AA, NOV >, < AA, SLB >,$$

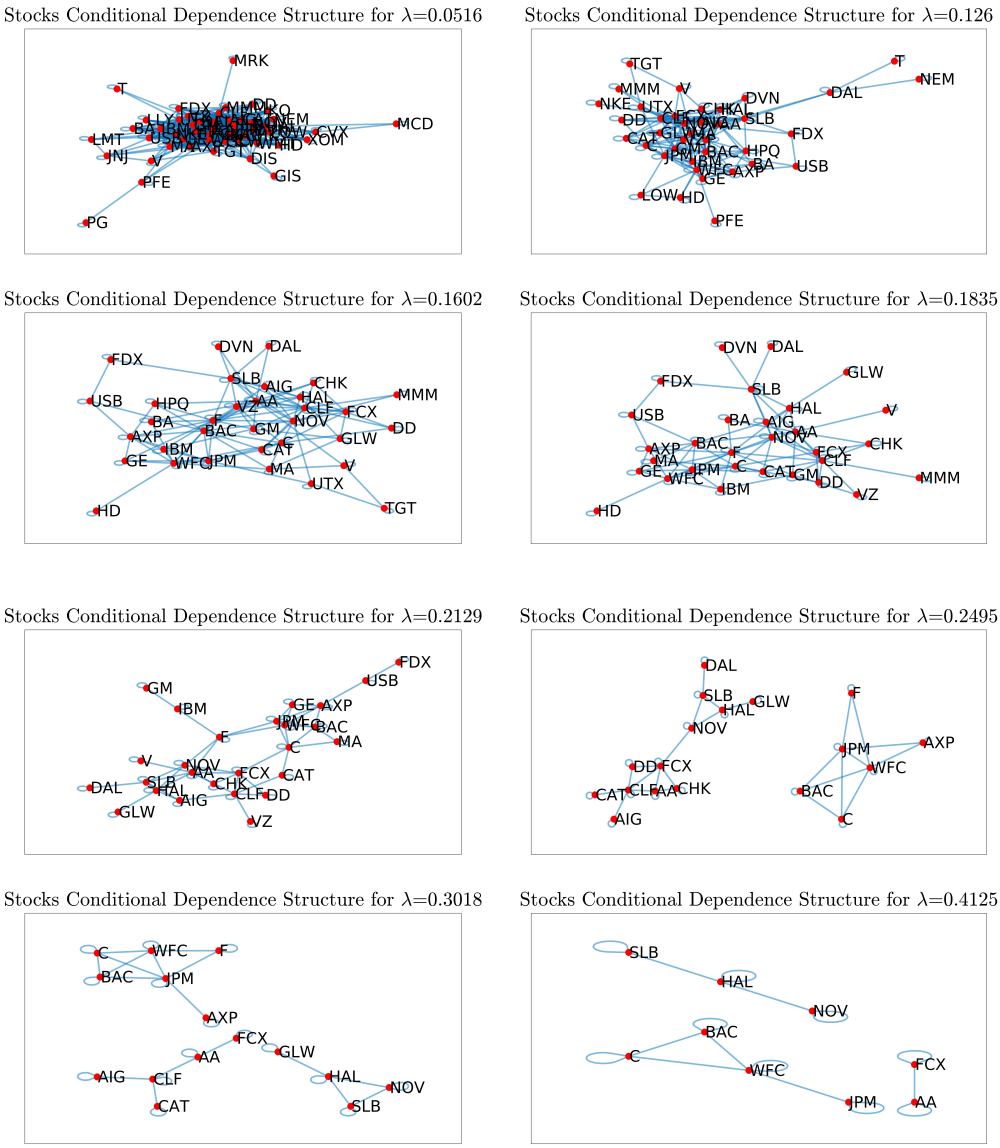

Figure 5.8: The conditional dependence structure of the returns of the stocks for eight different tuning parameter values.

$$< F, HAL >, < F, NOV >, < F, SLB >, < HAL, NOV >,$$

$$< HAL, SLB >, < NOV, SLB >),$$

$$\mathcal{E}(\mathcal{G}_3) = (< AXP, F >, < AXP, IBM >, < AXP, JPM >, < AXP, WFC >,$$

$$< F, IBM >, < F, JPM >, < F, WFC >, < IBM, JPM >,$$
$$< IBM, WFC >, < JPM, WFC >),$$
$$\mathcal{E}(\mathcal{G}_4) = (< AXP, GE >, < AXP, JPM >, < AXP, WFC >, < GE, JPM >,$$
$$< GE, WFC >, < JPM, WFC >),$$
$$\mathcal{E}(\mathcal{G}_5) = (< AIG, HAL >, < AIG, NOV >, < AIG, SLB >, < HAL, NOV >,$$
$$< HAL, SLB >, < NOV, SLB >),$$
$$\mathcal{E}(\mathcal{G}_6) = (< BAC, C >, < BAC, MA >, < BAC, WFC >, < C, MA >,$$
$$< C, WFC >, < MA, WFC >),$$
$$\mathcal{E}(\mathcal{G}_7) = (< AXP, F >, < AXP, GE >, < AXP, JPM >, < AXP, WFC >$$
$$< F, GE >, < F, JPM >, < F, WFC >, < GE, JPM >,$$
$$< GE, WFC >, < JPM, WFC >),$$
$$\mathcal{E}(\mathcal{G}_8) = (< AA, CLF >, < AA, F >, < AA, SLB >, < AA, VZ >$$
$$< CLF, F >, < CLF, SLB >, < CLF, VZ >, < F, SLB >,$$
$$< F, VZ >, < SLB, VZ >).$$

Also, a visual representation of the aforementioned maximal cliques is given in Figure E.4 in the appendix section E.6.2.

5.6 Discussion

Modelling high-frequency data in the standard time series framework poses many challenges. The dimension of the data is large, they are irregularly sampled with noise and can exhibit complex trends and complex dependence structure. On the other hand, due to the regularity of human activities and common structure of the system, it is possible to identify a certain degree of commonality. In fact, high-frequency financial data often exhibit particular diurnal patterns. Exploiting such information can be beneficial

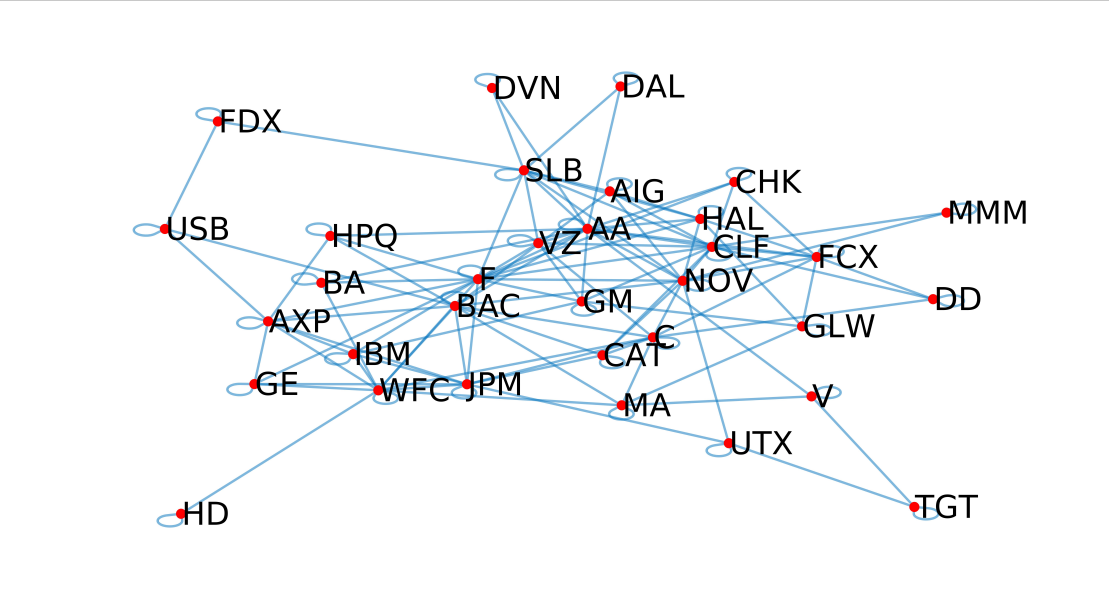

Figure 5.9: The conditional dependence structure of the stock returns for the estimated optimal tuning parameter.

in making inferences about the data co-movements and risk management.

We have proposed a functional data approach to investigate patterns in high-frequency data. Functional data approach allows us to easily treat large-dimensional and irregularly sampled data with noise and makes it possible to extract the commonality and to characterise variability in the continuous domain. In particular, this chapter aimed to estimate the conditional dependence structure through the graphical LASSO approach using high-frequency data as continuous functions. As far as we know, this is the first study which develops a methodology utilising the graphical LASSO approach, considering high-frequency observations as functional data in finance.

Under this framework, a necessary step is the estimation of the covariance structure of functional data, which expresses the co-movements between the stocks, and its inverse, which defines the stocks' conditional dependence network. Estimation of these quantities in functional data context is challenging, as the problem is posed in the infinite-dimensional setting. Previous studies apply basis expansion methods to es-

timate the curves' coefficients and then apply multivariate methods to the truncated coefficients (Qiao et al., 2019). In our extension, we estimate the corresponding quantities in functional covariance function and its inverse using the full dataset, with a proper functional approximation. A similar idea appears in recent work by Zapata et al. (2022) who develops an alternative theoretical framework to propose a new algorithm in the functional context.

A basic assumption we make for developing our proposed methodology is that the covariance function is time-independent across days but varies according to the pairs of stocks. Although the pointwise data (the intraday observations) within a day can be non-stationary, they remain similar across the days. Further, we utilise the functional variance-covariance matrix to express the variability of the dataset in this chapter as opposed to the Chapters 2 and 4 of this thesis where we estimate the volatility of a stock by the proposed volatility estimators, without considering the dependence between different time periods within a day.

Besides, we have also proposed a CV-based method for the estimation of the optimal tuning parameter in this chapter, assuming that the functional variance-covariance matrix shares similar behaviour to the corresponding matrices of neighbouring periods. The simulation study showed that the methodology works satisfactorily; however, its computational cost increases when we consider more variables in the model. Exploring alternative algorithms to improve efficiency is left for future work.

In the empirical analysis of this chapter, we attempted to estimate the conditional dependence structure between the stocks of our dataset. The investigation was conducted on both the log-prices and the returns of the stocks in our dataset. Between the two, the returns presented a more plausible dependence structure.

For the log-prices of the stocks, we mainly found strong conditional dependencies between stocks of different sectors. Possibly, this is due to the fact that the price patterns of stocks with low prices are affected by the patterns of stocks with noticeably higher

prices through the covariance functions. In this case, implementing our approach to the stocks' correlation matrix could reduce the effect of price differences among the stocks.

On the other hand, the dependence network seems to have a more sensible formation for returns. As we increase the tuning parameter, the dependencies between the stocks of the same sector become more apparent. The correlation coefficient is also strong for these cases, as found in the exploratory analysis in section 1.5.

Bibliography

Aït-Sahalia, Y. and Jacod, J. (2012). Analyzing the spectrum of asset returns: Jump and volatility components in high frequency data. *Journal of Economic Literature*, 50(4):1007–50.

Aït-Sahalia, Y. and Jacod, J. (2014). *High-frequency financial econometrics*. Princeton University Press.

Aït-Sahalia, Y. and Yu, J. (2009). High frequency market microstructure noise estimates and liquidity measures. *The Annals of Applied Statistics*, 3(1):422 – 457.

Alexander, C. (2008). *Quantitative methods in finance*. Wiley.

Andersen, T. G. and Bollerslev, T. (1997). Intraday periodicity and volatility persistence in financial markets. *Journal of empirical finance*, 4(2-3):115–158.

Andersen, T. G., Bollerslev, T., Diebold, F. X., and Ebens, H. (2001a). The distribution of realized stock return volatility. *Journal of financial economics*, 61(1):43–76.

Andersen, T. G., Bollerslev, T., Diebold, F. X., and Labys, P. (2001b). The distribution of realized exchange rate volatility. *Journal of the American statistical association*, 96(453):42–55.

Andersen, T. G., Bollerslev, T., Diebold, F. X., and Labys, P. (2003). Modeling and forecasting realized volatility. *Econometrica*, 71(2):579–625.

Andreou, E. and Ghysels, E. (2002). Rolling-sample volatility estimators: some new theoretical, simulation and empirical results. *Journal of Business & Economic Statistics*, 20(3):363–376.

Askanazi, R. and Warren, J. (2017). Factor analysis for volatility. *Available at SSRN 2986619*.

Bai, J. and Ng, S. (2002). Determining the number of factors in approximate factor models. *Econometrica*, 70(1):191–221.

Barhoumi, K., Darné, O., and Ferrara, L. (2013). Dynamic factor models: A review of the literature. *Available at SSRN 2291459*.

Barigozzi, M. and Hallin, M. (2016). Generalized dynamic factor models and volatilities: recovering the market volatility shocks.

Barndorff-Nielsen, O. E., Hansen, P. R., Lunde, A., and Shephard, N. (2008). Designing realized kernels to measure the ex post variation of equity prices in the presence of noise. *Econometrica*, 76(6):1481–1536.

Barndorff-Nielsen, O. E., Hansen, P. R., Lunde, A., and Shephard, N. (2009). Realized kernels in practice: trades and quotes.

Barndorff-Nielsen, O. E., Hansen, P. R., Lunde, A., and Shephard, N. (2011). Multivariate realised kernels: consistent positive semi-definite estimators of the covariation of equity prices with noise and non-synchronous trading. *Journal of Econometrics*, 162(2):149–169.

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

Barndorff-Nielsen, O. E. and Shephard, N. (2004). Econometric analysis of realized covariation: High frequency based covariance, regression, and correlation in financial economics. *Econometrica*, 72(3):885–925.

Bertsekas, D., Nedic, A., and Ozdaglar, A. (2003). *Convex Analysis and Optimization*. Athena Scientific optimization and computation series. Athena Scientific.

Black, F. (1976). Studies of stock market volatility changes. *1976 Proceedings of the American statistical association bisiness and economic statistics section*.

Black, F. and Scholes, M. (1973). The pricing of options and corporate liabilities. *Journal of political economy*, 81(3):637–654.

Bollerslev, T. (1986). Generalized autoregressive conditional heteroskedasticity. *Journal of econometrics*, 31(3):307–327.

Bosq, D. (2000). *Linear processes in function spaces: theory and applications*, volume 149. Springer Science & Business Media.

Bowman, A. W. and Azzalini, A. (1997). *Applied smoothing techniques for data analysis: the kernel approach with S-Plus illustrations*, volume 18. OUP Oxford.

Boyd, S., Boyd, S. P., and Vandenberghe, L. (2004). *Convex optimization*. Cambridge university press.

Breitung, J. and Eickmeier, S. (2006). Dynamic factor models. *Allgemeines Statistisches Archiv*, 90:27–42.

Calzolari, G., Halbleib, R., and Zagidullina, A. (2021). A latent factor model for forecasting realized variances. *Journal of Financial Econometrics*, 19(5):860–909.

Capasso, V. and Bakstein, D. (2015). *Introduction to Continuous-Time Stochastic Processes*. Springer.

Cattell, R. B. (1966). The scree test for the number of factors. *Multivariate behavioral research*, 1(2):245–276.

Chernov, M., Gallant, A. R., Ghysels, E., and Tauchen, G. (2003). Alternative models for stock price dynamics. *Journal of Econometrics*, 116(1-2):225–257.

Cherubini, U., Mulinacci, S., Gobbi, F., and Romagnoli, S. (2011). *Dynamic copula methods in finance*. John Wiley & Sons.

Christensen, K. and Podolski, M. (2005). Asymptotic theory for range-based estimation of integrated variance of a continuous semi-martingale. Technical report, Technical Report.

Christie, A. A. (1982). The stochastic behavior of common stock variances: Value, leverage and interest rate effects. *Journal of financial Economics*, 10(4):407–432.

Clark, R. (1977). Non-parametric estimation of a smooth regression function. *Journal of the Royal Statistical Society: Series B (Methodological)*, 39(1):107–113.

companiesmarketcap.com (n.d.). Companies ranked by market cap. https://companiesmarketcap.com/.

Corsi, F. (2005). *Measuring and modelling realized volatility*. PhD thesis, Università della Svizzera italiana.

Corsi, F., Dacorogna, M., Müller, U., and Zumbach, G. (2000). High frequency data do improve volatility and risk estimation. *Olsen & Associates Discussion Paper, Zurich, Switzerland*.

Craven, P. and Wahba, G. (1978). Smoothing noisy data with spline functions. *Numerische mathematik*, 31(4):377–403.

Cvitanić, J. and Zapatero, F. (2004). *Introduction to the economics and mathematics of financial markets*. MIT press.

De Brabanter, K., Cao, F., Gijbels, I., and Opsomer, J. (2018). Local polynomial regression with correlated errors in random design and unknown correlation structure. *Biometrika*, 105(3):681–690.

Ding, Y., Engle, R. F., Li, Y., and Zheng, X. (2022). Factor modeling for volatility. *SSRN Electronic Journal*.

Duffie, D. (2010). *Dynamic asset pricing theory*. Princeton University Press.

Elliott, R. J. and Kopp, P. E. (1999). *Mathematics of financial markets*, volume 2. Springer.

Embrechts, P., McNeil, A., and Straumann, D. (2002). Correlation and dependence in risk management: properties and pitfalls. *Risk management: value at risk and beyond*, 1:176–223.

Engle, R. F. (1982). Autoregressive conditional heteroscedasticity with estimates of the variance of united kingdom inflation. *Econometrica: Journal of the Econometric Society*, pages 987–1007.

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

Fama, E. F. (1963). Mandelbrot and the stable paretian hypothesis. *The journal of business*, 36(4):420–429.

Fan, J. and Gijbels, I. (1996). *Local polynomial modelling and its applications: monographs on statistics and applied probability 66*, volume 66. CRC Press.

Fan, J. and Wang, Y. (2008). Spot volatility estimation for high-frequency data. *Statistics and its Interface*, 1(2):279–288.

Fan, J. and Yao, Q. (2017). *The elements of financial econometrics*. Cambridge University Press.

Făt, C. M. and Dezsi, E. (2012). A factor analysis approach of international portfolio diversification: does it pay off? *Procedia Economics and Finance*, 3:648–653.

Ferraty, F. and Vieu, P. (2006). *Nonparametric functional data analysis: theory and practice*, volume 76. Springer.

finance.yahoo.com (n.d.). Yahoo finance - stock market live, quotes, business & finance news. https://finance.yahoo.com/.

Foster, D. P. and Nelson, D. B. (1996). Continuous record asymptotics for rolling sample variance estimators. Technical Report 1.

Franke, J., Härdle, W. K., and Hafner, C. M. (2019). *Statistics of financial markets*, volume 2. Springer.

Friedman, J., Hastie, T., and Tibshirani, R. (2008). Sparse inverse covariance estimation with the graphical lasso. *Biostatistics*, 9(3):432–441.

Gómez, A. M. E., Paynabar, K., and Pacella, P. (2020). Functional directed graphical models and applications in root-cause analysis and diagnosis. *Journal of Quality Technology*.

Gonçalves, S., Hounyo, U., and Meddahi, N. (2014). Bootstrap inference for pre-averaged realized volatility based on nonoverlapping returns. *Journal of Financial Econometrics*, 12(4):679–707.

Goyal, A., Pérignon, C., and Villa, C. (2008). How common are common return factors across the nyse and nasdaq? *Journal of Financial Economics*, 90(3):252–271.

Hansen, P. R. and Lunde, A. (2006). Realized variance and market microstructure noise. *Journal of Business & Economic Statistics*, 24(2):127–161.

Härdle, W., Simar, L., et al. (2007). *Applied multivariate statistical analysis*, volume 22007. Springer.

Hastie, T., Tibshirani, R., Friedman, J. H., and Friedman, J. H. (2009). *The elements of statistical learning: data mining, inference, and prediction*, volume 2. Springer.

Hastie, T., Tibshirani, R., and Wainwright, M. (2015). *Statistical learning with sparsity: the lasso and generalizations*. CRC press.

Hautsch, N. (2011). *Econometrics of financial high-frequency data*. Springer Science & Business Media.

Heston, S. L. (1993). A closed-form solution for options with stochastic volatility with applications to bond and currency options. *The review of financial studies*, 6(2):327–343.

Horn, J. L. (1965). A rationale and test for the number of factors in factor analysis. *Psychometrika*, 30(2):179–185.

Horváth, L. and Kokoszka, P. (2012). *Inference for functional data with applications*, volume 200. Springer Science & Business Media.

Howard, M. C. (2016). A review of exploratory factor analysis decisions and overview of current practices: What we are doing and how can we improve? *International Journal of Human-Computer Interaction*, 32(1):51–62.

Huang, X. and Tauchen, G. (2005). The relative contribution of jumps to total price variance. *Journal of financial econometrics*, 3(4):456–499.

Hull, J. (2012). *Options, Futures, and Other Derivatives*. Prentice Hall.

Ikenberry, D. L., Rankine, G., and Stice, E. K. (1996). What do stock splits really signal? *Journal of Financial and Quantitative analysis*, 31(3):357–375.

Jacod, J., Li, Y., Mykland, P. A., Podolskij, M., and Vetter, M. (2009). Microstructure noise in the continuous case: the pre-averaging approach. *Stochastic processes and their applications*, 119(7):2249–2276.

Jacod, J. and Shiryaev, A. (2003). *Limit theorems for stochastic processes*, volume 288. Springer Science & Business Media.

Joe, H. (2014). *Dependence modeling with copulas*. CRC press.

Jolliffe, I. (2011). Principal component analysis. In *International encyclopedia of statistical science*, pages 1094–1096. Springer.

Kaiser, H. F. (1960). The application of electronic computers to factor analysis. *Educational and psychological measurement*, 20(1):141–151.

Kendall, M. G. and Hill, A. B. (1953). The analysis of economic time-series-part i: Prices. *Journal of the Royal Statistical Society. Series A (General)*, 116(1):11–34.

Klebaner, F. C. (2012). *Introduction to stochastic calculus with applications*. World Scientific Publishing Company.

Köhler, M., Schindler, A., and Sperlich, S. (2014). A review and comparison of bandwidth selection methods for kernel regression. *International Statistical Review*, 82(2):243–274.

Kristensen, D. (2010). Nonparametric filtering of the realized spot volatility: A kernel-based approach. *Econometric Theory*, 26(1):60–93.

Lafit, G., Tuerlinckx, F., Myin-Germeys, I., and Ceulemans, E. (2019). A partial correlation screening approach for controlling the false positive rate in sparse gaussian graphical models. *Scientific reports*, 9(1):1–24.

Lam, C. and Yao, Q. (2011). Factor modelling for high-dimensional time series: A dimension-reduction approach. Technical report, Technical Report.

Lam, C., Yao, Q., and Bathia, N. (2011). Estimation of latent factors for high-dimensional time series. *Biometrika*, 98(4):901–918.

Lam, C., Yao, Q., et al. (2012). Factor modeling for high-dimensional time series: inference for the number of factors. *The Annals of Statistics*, 40(2):694–726.

Lamberton, D. and Lapeyre, B. (2011). *Introduction to stochastic calculus applied to finance*. CRC press.

Larson, A. B. (1960). Measurement of a random process in futures prices. In *Proceedings of the Annual Meeting (Western Farm Economics Association)*, volume 33, pages 101–112. JSTOR.

Lepski, O. V., Mammen, E., Spokoiny, V. G., et al. (1997). Optimal spatial adaptation to inhomogeneous smoothness: an approach based on kernel estimates with variable bandwidth selectors. *The Annals of Statistics*, 25(3):929–947.

Li, B., Chun, H., and Zhao, H. (2014a). On an additive semigraphoid model for statistical networks with application to pathway analysis. *Journal of the American Statistical Association*, 109(507):1188–1204.

Li, B. and Solea, E. (2018). A nonparametric graphical model for functional data with application to brain networks based on fmri. *Journal of the American Statistical Association*, 113(524):1637–1655.

Li, Y., Mykland, P. A., Renault, E., Zhang, L., and Zheng, X. (2014b). Realized volatility when sampling times are possibly endogenous. *Econometric theory*, 30(3):580–605.

Li, Y., Nolte, I., and Nolte-Lechner, S. (2015). High-frequency volatility estimation and the relative importance of market microstructure variables: An autoregressive conditional intensity approach. *SSRN Electronic Journal. doi*, 10.

Liptser, R. S. and Shiryayev, A. N. (1989). *Basic Concepts and the Review of Results of The General Theory of Stochastic Processes*, pages 1–84. Springer Netherlands, Dordrecht.

Lu, T.-T. and Shiou, S.-H. (2002). Inverses of 2× 2 block matrices. *Computers & Mathematics with Applications*, 43(1-2):119–129.

Luciani, M. and Veredas, D. (2015). Estimating and forecasting large panels of volatilities with approximate dynamic factor models. *Journal of Forecasting*, 34(3):163–176.

Lunagmez, S., Olhede, S. C., and Wolfe, P. J. (2021). Modeling network populations via graph distances. *Journal of the American Statistical Association*, 116(536):2023–2040.

Madhavan, A. (2000). Market microstructure: A survey. *Journal of financial markets*, 3(3):205–258.

Marron, J. (1994). Visual understanding of higher-order kernels. *Journal of Computational and Graphical Statistics*, 3(4):447–458.

Meucci, A. (2005). *Risk and asset allocation*, volume 1. Springer.

Mota, P. (2012). Normality assumption for the log-return of the stock prices. *Discussiones Mathematicae Probability and Statistics*, 32(1-2):47–58.

Müller, H.-G., Sen, R., and Stadtmüller, U. (2011). Functional data analysis for volatility. *Journal of Econometrics*, 165(2):233–245.

Müller, H.-G. and Stadtmüller, U. (1987). Variable bandwidth kernel estimators of regression curves. *The Annals of Statistics*, pages 182–201.

Müller, H.-G., Stadtmüller, U., and Yao, F. (2006). Functional variance processes. *Journal of the American Statistical Association*, 101(475):1007–1018.

Mykland, P. A. and Zhang, L. (2009). Inference for continuous semimartingales observed at high frequency. *Econometrica*, 77(5):1403–1445.

Mykland, P. A. and Zhang, L. (2012). The econometrics of high frequency data. *Statistical methods for stochastic differential equations*, 124:109.

Nolte, I. and Voev, V. (2012). Least squares inference on integrated volatility and the relationship between efficient prices and noise. *Journal of Business & Economic Statistics*, 30(1):94–108.

Officer, R. R. (1972). The distribution of stock returns. *Journal of the american statistical association*, 67(340):807–812.

Ogawa, S. and Sanfelici, S. (2011). An improved two-step regularization scheme for spot volatility estimation. *Economic Notes*, 40(3):107–134.

Øksendal, B. (2003). Stochastic differential equations. In *Stochastic differential equations*, pages 65–84. Springer.

Opsomer, J., Wang, Y., and Yang, Y. (2001). Nonparametric regression with correlated errors. *Statistical Science*, pages 134–153.

Osborne, M. F. (1959). Brownian motion in the stock market. *Operations research*, 7(2):145–173.

Podolskij, M., Vetter, M., et al. (2009). Estimation of volatility functionals in the simultaneous presence of microstructure noise and jumps. *Bernoulli*, 15(3):634–658.

Poon, S.-H. (2005). *A practical guide to forecasting financial market volatility*. John Wiley & Sons.

Protter, P. E. (2005). Stochastic differential equations. In *Stochastic integration and differential equations*, pages 249–361. Springer.

Qiao, X., Guo, S., and James, G. M. (2019). Functional graphical models. *Journal of the American Statistical Association*, 114(525):211–222.

Rachev, S., Sun, W., and Stein, M. (2009). Copula concepts in financial markets. *Portfolio Institutionell (Ed.)*, 4:12–15.

Ramsay, J. O. and Silverman, B. W. (2005). *Functional Data Analysis*. Springer.

Renault, E. and Werker, B. J. (2011). Causality effects in return volatility measures with random times. *Journal of Econometrics*, 160(1):272–279.

Ross, S. (2010). *A first course in probability*. Pearson.

Ruppert, D. and Matteson, D. S. (2011). *Statistics and data analysis for financial engineering*, volume 13. Springer.

Shen, K., Yao, J., and Li, W. K. (2020). Forecasting high-dimensional realized volatility matrices using a factor model. *Quantitative Finance*, 20(11):1879–1887.

Shreve, S. E. et al. (2004). *Stochastic calculus for finance II: Continuous-time models*, volume 11. Springer.

Simonoff, J. S. (2012). *Smoothing methods in statistics*. Springer Science & Business Media.

Solea, E. and Li, B. (2022). Copula gaussian graphical models for functional data. *Journal of the American Statistical Association*, 117(538):781–793.

Staniswalis, J. G. (1989). Local bandwidth selection for kernel estimates. *Journal of the American Statistical Association*, 84(405):284–288.

Sweeting, P. (2017). *Financial enterprise risk management*. Cambridge University Press.

Taylor, S. J. (2005). *Asset Price Dynamics, Volatility, and Prediction*. Princeton University Press, stu - student edition edition.

Tibshirani, R. (1996). Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society: Series B (Methodological)*, 58(1):267–288.

Tsay, R. S. (2005). *Analysis of financial time series*, volume 543. John Wiley & Sons.

Ullah, S. and Finch, C. F. (2013). Applications of functional data analysis: A systematic review. *BMC medical research methodology*, 13:1–12.

Wang, J.-L., Chiou, J.-M., and Müller, H.-G. (2016). Functional data analysis. *Annual Review of Statistics and its application*, 3:257–295.

Yao, F., Müller, H.-G., and Wang, J.-L. (2005). Functional data analysis for sparse longitudinal data. *Journal of the American statistical association*, 100(470):577–590.

Yuan, M. and Lin, Y. (2006). Model selection and estimation in regression with grouped variables. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 68(1):49–67.

Zapata, J., Oh, S.-Y., and Petersen, A. (2022). Partial separability and functional graphical models for multivariate gaussian processes. *Biometrika*, 109(3):665–681.

Zhang, L., Mykland, P. A., and Aït-Sahalia, Y. (2005). A tale of two time scales: Determining integrated volatility with noisy high-frequency data. *Journal of the American Statistical Association*, 100(472):1394–1411.

Zhu, H., Strawn, N., and Dunson, D. B. (2016). Bayesian graphical models for multivariate functional data.

Zu, Y. and Boswijk, H. P. (2014). Estimating spot volatility with high-frequency financial data. *Journal of Econometrics*, 181(2):117–135.

Appendix A

Appendix for Chapter 1

his appendix provides the figures and the tables of the exploratory analysis of section 1.5 in Chapter 1. In particular, we report the boxplots of the descriptive statistics across the stocks (see Figure A.1), the correlation coefficients between the stocks (see Table A.1), the correlation coefficients between the stocks related to the same sector (see Tables A.2-A.10) and the companies' annual market capitalisation for the data period (see Table A.11).

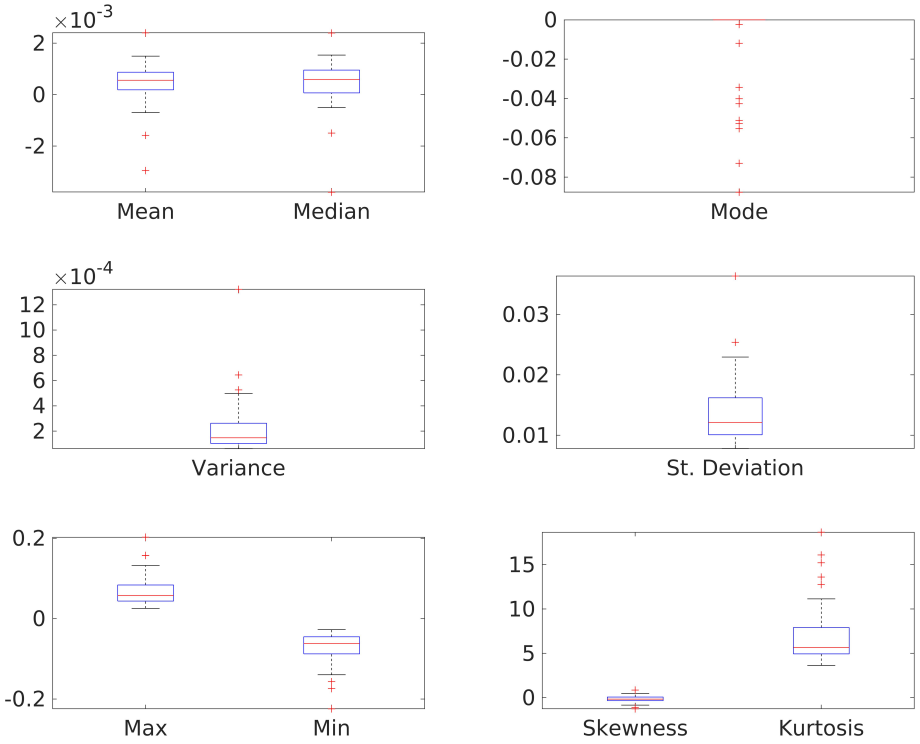

Figure A.1: The boxplots depict the descriptive statistics of the daily returns across the stocks.

	MMM	T	AA	AXP	AIG	BAC	BA	CAT	CHK	CVX	C	CLF	KO	GLW	DAL	DVN	DD	LLY	XOM	FDX	F	FCX	GE	GIS	GM	HAL	HPQ	HD	IBM	JPM	JNJ	LMT	LOW	MA	MCD	MRK	NOV	NEM	NKE	PFE	PG	SLB	TGT	USB	UTX	VZ	V	WMT	DIS	WFC
MMM	1	0.41	0.49	0.58	0.52	0.45	0.51	0.54	0.29	0.53	0.52	0.27	0.4	0.44	0.33	0.41	0.6	0.43	0.59	0.52	0.39	0.41	0.63	0.37	0.43	0.45	0.33	0.45	0.42	0.5	0.53	0.54	0.37	0.48	0.37	0.37	0.45	0.13	0.34	0.45	0.38	0.51	0.31	0.57	0.66	0.36	0.44	0.28	0.54	0.56
T	0.41	1	0.29	0.37	0.31	0.26	0.25	0.3	0.16	0.33	0.31	0.13	0.37	0.2	0.18	0.24	0.34	0.31	0.39	0.27	0.23	0.24	0.4	0.42	0.27	0.25	0.2	0.28	0.28	0.32	0.41	0.25	0.22	0.31	0.29	0.31	0.22	0.08	0.21	0.37	0.38	0.27	0.26	0.38	0.36	0.72	0.25	0.29	0.36	0.38
AA	0.49	0.29	1	0.44	0.39	0.45	0.37	0.52	0.37	0.49	0.44	0.33	0.22	0.37	0.25	0.44	0.47	0.3	0.48	0.38	0.39	0.52	0.45	0.22	0.37	0.45	0.32	0.26	0.31	0.4	0.28	0.36	0.26	0.39	0.26	0.22	0.45	0.31	0.27	0.29	0.16	0.49	0.18	0.43	0.43	0.24	0.33	0.15	0.37	0.48
AXP	0.58	0.37	0.44	1	0.54	0.53	0.46	0.41	0.26	0.46	0.54	0.14	0.4	0.41	0.34	0.38	0.5	0.4	0.48	0.45	0.44	0.32	0.5	0.34	0.41	0.34	0.26	0.45	0.29	0.55	0.49	0.45	0.38	0.55	0.34	0.33	0.39	0.12	0.4	0.41	0.33	0.43	0.35	0.64	0.55	0.38	0.55	0.31	0.52	0.62
AIG	0.52	0.31	0.39	0.54	1	0.53	0.42	0.47	0.32	0.44	0.58	0.22	0.33	0.38	0.33	0.39	0.49	0.31	0.46	0.41	0.45	0.37	0.47	0.32	0.41	0.38	0.27	0.41	0.35	0.52	0.38	0.41	0.33	0.44	0.33	0.3	0.41	0.15	0.31	0.39	0.26	0.45	0.29	0.51	0.51	0.27	0.42	0.23	0.45	0.53
BAC	0.45	0.26	0.45	0.53	0.53	1	0.35	0.45	0.32	0.4	0.78	0.25	0.26	0.37	0.33	0.36	0.44	0.25	0.4	0.43	0.45	0.37	0.47	0.22	0.44	0.34	0.28	0.34	0.29	0.71	0.29	0.32	0.36	0.36	0.24	0.2	0.43	0.16	0.24	0.28	0.18	0.41	0.22	0.6	0.46	0.22	0.33	0.21	0.43	0.64
BA	0.51	0.25	0.37	0.46	0.42	0.35	1	0.37	0.24	0.36	0.42	0.19	0.31	0.42	0.29	0.27	0.43	0.32	0.4	0.42	0.38	0.36	0.43	0.29	0.4	0.31	0.26	0.33	0.29	0.37	0.38	0.55	0.28	0.43	0.29	0.27	0.35	0.1	0.29	0.31	0.23	0.36	0.23	0.42	0.58	0.25	0.38	0.22	0.42	0.46
CAT	0.54	0.3	0.52	0.41	0.47	0.45	0.37	1	0.31	0.49	0.5	0.39	0.26	0.3	0.25	0.45	0.49	0.24	0.53	0.46	0.43	0.52	0.55	0.24	0.45	0.5	0.32	0.33	0.33	0.44	0.32	0.37	0.28	0.37	0.32	0.26	0.54	0.22	0.27	0.28	0.21	0.54	0.23	0.46	0.59	0.23	0.31	0.16	0.39	0.46
CHK	0.29	0.16	0.37	0.26	0.32	0.32	0.24	0.31	1	0.46	0.34	0.22	0.14	0.23	0.13	0.6	0.28	0.15	0.4	0.24	0.29	0.37	0.3	0.08	0.26	0.44	0.25	0.21	0.16	0.35	0.13	0.19	0.16	0.25	0.16	0.11	0.47	0.2	0.15	0.11	0.09	0.45	0.06	0.26	0.25	0.12	0.21	0.02	0.23	0.32
CVX	0.53	0.33	0.49	0.46	0.44	0.4	0.36	0.49	0.46	1	0.44	0.23	0.38	0.38	0.15	0.61	0.51	0.36	0.77	0.31	0.38	0.46	0.54	0.31	0.35	0.56	0.34	0.34	0.38	0.44	0.45	0.36	0.29	0.42	0.33	0.33	0.55	0.26	0.25	0.38	0.34	0.63	0.25	0.49	0.5	0.29	0.37	0.2	0.42	0.51
C	0.52	0.31	0.44	0.54	0.58	0.78	0.42	0.5	0.34	0.44	1	0.24	0.31	0.38	0.36	0.38	0.51	0.32	0.46	0.49	0.49	0.37	0.55	0.26	0.49	0.39	0.32	0.38	0.37	0.77	0.41	0.4	0.34	0.43	0.29	0.27	0.44	0.16	0.28	0.36	0.26	0.44	0.3	0.64	0.51	0.26	0.38	0.27	0.46	0.66
CLF	0.27	0.13	0.33	0.14	0.22	0.25	0.19	0.39	0.22	0.23	0.24	1	0.08	0.22	0.08	0.25	0.29	0.09	0.25	0.23	0.22	0.41	0.24	0.07	0.27	0.25	0.2	0.1	0.17	0.23	0.08	0.18	0.1	0.17	0.14	0.05	0.31	0.26	0.1	0.09	0.01	0.29	0.01	0.2	0.25	0.1	0.14	-0.01	0.15	0.2
KO	0.4	0.37	0.22	0.4	0.33	0.26	0.31	0.26	0.14	0.38	0.31	0.08	1	0.2	0.19	0.25	0.36	0.29	0.39	0.26	0.24	0.21	0.38	0.43	0.25	0.21	0.19	0.31	0.3	0.31	0.43	0.33	0.26	0.34	0.39	0.29	0.23	0.04	0.24	0.37	0.46	0.3	0.26	0.34	0.39	0.35	0.3	0.33	0.42	0.37
GIW	0.44	0.2	0.37	0.41	0.38	0.37	0.42	0.3	0.23	0.38	0.38	0.22	0.2	1	0.24	0.35	0.39	0.26	0.38	0.34	0.36	0.33	0.41	0.25	0.39	0.38	0.26	0.29	0.31	0.35	0.33	0.37	0.29	0.33	0.25	0.21	0.35	0.09	0.21	0.28	0.18	0.38	0.19	0.41	0.44	0.21	0.31	0.15	0.36	0.43
DAL	0.33	0.18	0.25	0.34	0.33	0.33	0.29	0.25	0.13	0.15	0.36	0.08	0.19	0.24	1	0.16	0.31	0.22	0.18	0.39	0.34	0.17	0.32	0.22	0.33	0.16	0.19	0.24	0.16	0.32	0.25	0.29	0.19	0.29	0.2	0.22	0.16	0	0.23	0.22	0.15	0.14	0.19	0.33	0.3	0.18	0.25	0.19	0.35	0.34
DVN	0.41	0.24	0.44	0.38	0.39	0.36	0.27	0.45	0.6	0.61	0.38	0.25	0.25	0.35	0.16	1	0.4	0.23	0.55	0.32	0.33	0.45	0.44	0.18	0.32	0.57	0.27	0.2	0.28	0.36	0.28	0.25	0.19	0.34	0.27	0.19	0.57	0.25	0.2	0.24	0.19	0.62	0.13	0.36	0.36	0.19	0.26	0.07	0.33	0.38
DD	0.6	0.34	0.47	0.5	0.49	0.44	0.43	0.49	0.28	0.51	0.51	0.29	0.36	0.39	0.31	0.4	1	0.38	0.52	0.43	0.44	0.44	0.52	0.28	0.42	0.42	0.31	0.38	0.39	0.48	0.46	0.41	0.34	0.41	0.31	0.33	0.4	0.15	0.29	0.36	0.28	0.48	0.25	0.51	0.53	0.31	0.39	0.24	0.47	0.53
LLY	0.43	0.31	0.3	0.4	0.31	0.25	0.32	0.24	0.15	0.36	0.32	0.09	0.29	0.26	0.22	0.23	0.38	1	0.36	0.27	0.22	0.22	0.33	0.3	0.21	0.24	0.17	0.25	0.2	0.32	0.48	0.33	0.24	0.35	0.26	0.44	0.24	0.06	0.24	0.49	0.36	0.28	0.27	0.39	0.4	0.31	0.31	0.31	0.34	0.38
XOM	0.59	0.39	0.48	0.48	0.46	0.4	0.4	0.53	0.4	0.77	0.46	0.25	0.39	0.38	0.18	0.55	0.52	0.36	1	0.39	0.38	0.46	0.56	0.36	0.42	0.54	0.3	0.4	0.39	0.46	0.47	0.42	0.36	0.43	0.36	0.33	0.53	0.25	0.27	0.39	0.36	0.61	0.28	0.5	0.5	0.33	0.36	0.24	0.44	0.52
FDX	0.52	0.27	0.38	0.45	0.41	0.43	0.42	0.46	0.24	0.31	0.49	0.23	0.26	0.34	0.39	0.32	0.43	0.27	0.39	1	0.39	0.34	0.49	0.26	0.4	0.34	0.25	0.34	0.29	0.44	0.38	0.42	0.32	0.41	0.26	0.26	0.35	0.09	0.33	0.31	0.22	0.38	0.22	0.46	0.49	0.23	0.33	0.25	0.43	0.49
F	0.39	0.23	0.39	0.44	0.45	0.45	0.38	0.43	0.29	0.38	0.49	0.22	0.24	0.36	0.34	0.33	0.44	0.22	0.38	0.39	1	0.36	0.42	0.24	0.66	0.32	0.24	0.39	0.31	0.46	0.28	0.34	0.34	0.38	0.25	0.19	0.34	0.14	0.34	0.24	0.18	0.36	0.25	0.45	0.42	0.22	0.34	0.23	0.43	0.45
FCX	0.41	0.24	0.52	0.32	0.37	0.37	0.36	0.52	0.37	0.46	0.37	0.41	0.21	0.33	0.17	0.45	0.44	0.22	0.46	0.34	0.36	1	0.41	0.18	0.39	0.47	0.25	0.26	0.29	0.34	0.25	0.28	0.28	0.3	0.24	0.17	0.49	0.47	0.17	0.19	0.16	0.5	0.17	0.33	0.41	0.18	0.26	0.13	0.31	0.36
GE	0.63	0.4	0.45	0.5	0.47	0.47	0.43	0.55	0.3	0.54	0.55	0.24	0.38	0.41	0.32	0.44	0.52	0.33	0.56	0.49	0.42	0.41	1	0.34	0.45	0.44	0.29	0.42	0.43	0.52	0.45	0.42	0.34	0.43	0.38	0.32	0.46	0.17	0.3	0.44	0.34	0.55	0.29	0.55	0.53	0.37	0.39	0.28	0.48	0.55
GIS	0.37	0.42	0.22	0.34	0.32	0.22	0.29	0.24	0.08	0.31	0.26	0.07	0.43	0.25	0.22	0.18	0.28	0.3	0.36	0.26	0.24	0.18	0.34	1	0.24	0.22	0.2	0.25	0.23	0.26	0.44	0.32	0.26	0.35	0.33	0.31	0.19	0.09	0.23	0.37	0.46	0.24	0.25	0.33	0.33	0.37	0.3	0.33	0.41	0.3
GM	0.43	0.27	0.37	0.41	0.41	0.44	0.4	0.45	0.26	0.35	0.49	0.27	0.25	0.39	0.33	0.32	0.42	0.21	0.42	0.4	0.66	0.39	0.45	0.24	1	0.32	0.25	0.33	0.3	0.44	0.29	0.35	0.3	0.37	0.26	0.21	0.39	0.16	0.28	0.26	0.18	0.34	0.21	0.43	0.45	0.23	0.32	0.21	0.38	0.44
HAL	0.45	0.25	0.45	0.34	0.38	0.34	0.31	0.5	0.44	0.56	0.39	0.25	0.21	0.38	0.16	0.57	0.42	0.24	0.54	0.34	0.32	0.47	0.44	0.22	0.32	1	0.24	0.23	0.26	0.37	0.31	0.32	0.21	0.32	0.25	0.21	0.63	0.21	0.24	0.24	0.16	0.74	0.16	0.34	0.39	0.21	0.27	0.07	0.36	0.38
HPQ	0.33	0.2	0.32	0.26	0.27	0.28	0.26	0.32	0.25	0.34	0.32	0.2	0.19	0.26	0.19	0.27	0.31	0.17	0.3	0.25	0.24	0.25	0.29	0.2	0.25	0.24	1	0.14	0.31	0.29	0.2	0.25	0.08	0.21	0.18	0.18	0.28	0.13	0.15	0.2	0.07	0.26	0.13	0.27	0.27	0.17	0.22	0.12	0.22	0.27
HD	0.45	0.28	0.26	0.45	0.41	0.34	0.33	0.33	0.21	0.34	0.38	0.1	0.31	0.29	0.24	0.2	0.38	0.25	0.4	0.34	0.39	0.26	0.42	0.25	0.33	0.23	0.14	1	0.28	0.36	0.38	0.32	0.65	0.39	0.29	0.27	0.27	0.04	0.31	0.33	0.29	0.29	0.35	0.42	0.42	0.26	0.35	0.29	0.46	0.41
IBM	0.42	0.28	0.31	0.29	0.35	0.29	0.29	0.33	0.16	0.38	0.37	0.17	0.3	0.31	0.16	0.28	0.39	0.2	0.39	0.29	0.31	0.29	0.43	0.23	0.3	0.26	0.31	0.28	1	0.34	0.34	0.38	0.24	0.33	0.34	0.21	0.25	0.1	0.21	0.25	0.21	0.34	0.17	0.36	0.4	0.22	0.29	0.19	0.31	0.36
JPM	0.5	0.32	0.4	0.55	0.52	0.71	0.37	0.44	0.35	0.44	0.77	0.23	0.31	0.35	0.32	0.36	0.48	0.32	0.46	0.44	0.46	0.34	0.52	0.26	0.44	0.37	0.29	0.36	0.34	1	0.39	0.39	0.35	0.39	0.27	0.28	0.37	0.1	0.31	0.37	0.25	0.42	0.28	0.64	0.48	0.28	0.35	0.23	0.45	0.66
JNJ	0.53	0.41	0.28	0.49	0.38	0.29	0.38	0.32	0.13	0.45	0.41	0.08	0.43	0.33	0.25	0.28	0.46	0.48	0.47	0.38	0.28	0.25	0.45	0.44	0.29	0.31	0.2	0.38	0.34	0.39	1	0.44	0.29	0.44	0.38	0.45	0.25	0.05	0.29	0.53	0.46	0.36	0.32	0.49	0.48	0.4	0.41	0.36	0.46	0.51
LMT	0.54	0.25	0.36	0.45	0.41	0.32	0.55	0.37	0.19	0.36	0.4	0.18	0.33	0.37	0.29	0.25	0.41	0.33	0.42	0.42	0.34	0.28	0.42	0.32	0.35	0.32	0.25	0.32	0.38	0.39	0.44	1	0.25	0.4	0.32	0.34	0.3	0.06	0.33	0.36	0.28	0.33	0.24	0.45	0.59	0.29	0.37	0.25	0.45	0.44
LOW	0.37	0.22	0.26	0.38	0.33	0.36	0.28	0.28	0.16	0.29	0.34	0.1	0.26	0.29	0.19	0.19	0.34	0.24	0.36	0.32	0.34	0.28	0.34	0.26	0.3	0.21	0.08	0.65	0.24	0.35	0.29	0.25	1	0.32	0.21	0.22	0.2	0.04	0.28	0.22	0.28	0.25	0.31	0.34	0.38	0.17	0.29	0.25	0.38	0.35
MA	0.48	0.31	0.39	0.55	0.44	0.56	0.43	0.37	0.25	0.42	0.43	0.17	0.34	0.33	0.29	0.34	0.41	0.35	0.43	0.41	0.38	0.3	0.43	0.35	0.37	0.32	0.21	0.39	0.33	0.39	0.44	0.4	0.32	1	0.35	0.28	0.35	0.07	0.4	0.38	0.29	0.38	0.32	0.44	0.46	0.27	0.72	0.25	0.47	0.45
MCD	0.37	0.29	0.26	0.34	0.33	0.24	0.29	0.32	0.16	0.33	0.29	0.14	0.39	0.25	0.2	0.27	0.31	0.26	0.36	0.26	0.25	0.24	0.38	0.33	0.26	0.25	0.18	0.29	0.34	0.27	0.38	0.32	0.21	0.35	1	0.29	0.28	0.05	0.33	0.28	0.31	0.3	0.24	0.33	0.33	0.25	0.27	0.28	0.35	0.34
MRK	0.37	0.31	0.22	0.33	0.3	0.2	0.27	0.26	0.11	0.33	0.27	0.05	0.29	0.21	0.22	0.19	0.33	0.44	0.33	0.26	0.19	0.17	0.32	0.31	0.21	0.21	0.18	0.27	0.21	0.28	0.45	0.34	0.22	0.28	0.29	1	0.18	0.03	0.2	0.49	0.35	0.23	0.22	0.32	0.31	0.3	0.28	0.25	0.33	0.32
NOV	0.45	0.22	0.45	0.39	0.41	0.43	0.35	0.54	0.47	0.55	0.44	0.31	0.23	0.35	0.16	0.57	0.4	0.24	0.53	0.35	0.34	0.49	0.46	0.19	0.39	0.63	0.28	0.27	0.25	0.37	0.25	0.3	0.2	0.35	0.28	0.18	1	0.23	0.24	0.19	0.13	0.69	0.15	0.39	0.44	0.17	0.3	0.09	0.35	0.42
NEM	0.13	0.08	0.31	0.12	0.15	0.16	0.1	0.22	0.2	0.26	0.16	0.26	0.04	0.09	0	0.25	0.15	0.06	0.25	0.09	0.14	0.47	0.17	0.09	0.16	0.21	0.13	0.04	0.1	0.1	0.05	0.06	0.04	0.07	0.05	0.03	0.23	1	-0.01	0.06	0.02	0.26	0.02	0.1	0.1	0.06	0.05	0.03	0.03	0.11
NKE	0.34	0.21	0.27	0.4	0.31	0.24	0.29	0.27	0.15	0.25	0.28	0.1	0.24	0.21	0.23	0.2	0.29	0.24	0.27	0.33	0.34	0.17	0.3	0.23	0.28	0.24	0.15	0.31	0.21	0.31	0.29	0.33	0.28	0.4	0.33	0.2	0.24	-0.01	1	0.21	0.18	0.26	0.26	0.34	0.35	0.2	0.33	0.22	0.38	0.34
PFE	0.45	0.37	0.29	0.41	0.39	0.28	0.31	0.28	0.11	0.38	0.36	0.09	0.37	0.28	0.22	0.24	0.36	0.49	0.39	0.31	0.24	0.19	0.44	0.37	0.26	0.24	0.2	0.33	0.25	0.37	0.53	0.36	0.22	0.38	0.28	0.49	0.19	0.06	0.21	1	0.41	0.3	0.25	0.42	0.39	0.35	0.38	0.29	0.43	0.41
PG	0.38	0.38	0.16	0.33	0.26	0.18	0.23	0.21	0.09	0.34	0.26	0.01	0.46	0.18	0.15	0.19	0.28	0.36	0.36	0.22	0.18	0.16	0.34	0.46	0.18	0.16	0.07	0.29	0.21	0.25	0.46	0.28	0.28	0.29	0.31	0.35	0.13	0.02	0.18	0.41	1	0.21	0.26	0.31	0.33	0.32	0.28	0.37	0.37	0.3
SLB	0.51	0.27	0.49	0.43	0.45	0.41	0.36	0.54	0.45	0.63	0.44	0.29	0.3	0.38	0.14	0.62	0.48	0.28	0.61	0.38	0.36	0.5	0.55	0.24	0.34	0.74	0.26	0.29	0.34	0.42	0.36	0.33	0.25	0.38	0.3	0.23	0.69	0.26	0.26	0.3	0.21	1	0.19	0.41	0.47	0.22	0.34	0.12	0.39	0.46
TGT	0.31	0.26	0.18	0.35	0.29	0.22	0.23	0.23	0.06	0.25	0.3	0.01	0.26	0.19	0.19	0.13	0.25	0.27	0.32	0.22	0.25	0.17	0.29	0.25	0.21	0.16	0.13	0.35	0.17	0.28	0.32	0.24	0.31	0.32	0.24	0.22	0.15	0.02	0.26	0.25	0.26	0.19	1	0.31	0.32	0.24	0.26	0.46	0.28	0.34
USB	0.57	0.38	0.43	0.64	0.51	0.6	0.42	0.46	0.26	0.49	0.64	0.2	0.34	0.41	0.33	0.36	0.51	0.39	0.5	0.46	0.45	0.33	0.55	0.33	0.43	0.34	0.27	0.42	0.36	0.64	0.49	0.45	0.34	0.44	0.33	0.32	0.39	0.1	0.34	0.42	0.31	0.41	0.31	1	0.53	0.37	0.42	0.3	0.52	0.78
UTX	0.66	0.36	0.43	0.55	0.51	0.46	0.58	0.59	0.25	0.5	0.51	0.25	0.39	0.44	0.3	0.36	0.53	0.4	0.5	0.49	0.42	0.41	0.53	0.33	0.45	0.39	0.27	0.42	0.4	0.48	0.48	0.59	0.38	0.46	0.33	0.31	0.44	0.1	0.35	0.39	0.33	0.47	0.32	0.53	1	0.31	0.43	0.28	0.53	0.56
VZ	0.36	0.72	0.24	0.38	0.27	0.22	0.25	0.23	0.12	0.29	0.26	0.1	0.35	0.21	0.18	0.19	0.31	0.31	0.33	0.23	0.22	0.18	0.37	0.37	0.23	0.21	0.17	0.26	0.22	0.28	0.4	0.29	0.17	0.27	0.25	0.3	0.17	0.06	0.2	0.35	0.32	0.22	0.24	0.37	0.31	1	0.25	0.29	0.33	0.34
V	0.44	0.25	0.33	0.55	0.42	0.33	0.38	0.31	0.21	0.37	0.38	0.14	0.3	0.31	0.25	0.26	0.39	0.31	0.36	0.33	0.34	0.26	0.39	0.3	0.32	0.27	0.22	0.35	0.29	0.35	0.41	0.37	0.29	0.72	0.27	0.28	0.3	0.05	0.33	0.38	0.28	0.34	0.26	0.42	0.43	0.25	1	0.26	0.42	0.41
WMT	0.28	0.29	0.15	0.31	0.23	0.21	0.22	0.16	0.02	0.2	0.27	-0.01	0.33	0.15	0.19	0.07	0.24	0.31	0.24	0.25	0.23	0.13	0.28	0.33	0.21	0.07	0.22	0.29	0.19	0.23	0.36	0.25	0.25	0.25	0.28	0.25	0.09	0.03	0.22	0.29	0.37	0.12	0.46	0.3	0.28	0.29	0.26	1	0.35	0.29
DIS	0.54	0.36	0.37	0.52	0.45	0.43	0.42	0.39	0.23	0.42	0.46	0.15	0.42	0.36	0.35	0.33	0.47	0.34	0.44	0.43	0.43	0.31	0.48	0.41	0.38	0.36	0.22	0.46	0.31	0.45	0.46	0.45	0.38	0.47	0.35	0.33	0.35	0.03	0.38	0.43	0.37	0.39	0.28	0.52	0.53	0.33	0.42	0.35	1	0.53
WFC	0.56	0.38	0.48	0.62	0.53	0.64	0.46	0.46	0.32	0.51	0.66	0.2	0.37	0.43	0.34	0.38	0.53	0.38	0.52	0.49	0.45	0.36	0.55	0.3	0.44	0.38	0.27	0.41	0.36	0.66	0.51	0.44	0.35	0.45	0.34	0.32	0.42	0.11	0.34	0.41	0.3	0.46	0.34	0.78	0.56	0.34	0.41	0.29	0.53	1

Table A.1: The correlation matrix of the daily returns for the underlying stocks.

	AA	CLF	DD	FCX	NEM
AA	1	0.3347	0.4668	0.5164	0.3120
CLF	0.3347	1	0.2857	0.4127	0.2573
DD	0.4668	0.2857	1	0.4443	0.1496
FCX	0.5164	0.4127	0.4443	1	0.4716
NEM	0.3120	0.2573	0.1496	0.4716	1

Table A.2: The correlation matrix of the daily returns for the stocks of the sector 'Basic Materials'.

	T	VZ	DIS
T	1	0.7194	0.3586
VZ	0.7194	1	0.3346
DIS	0.3586	0.3346	1

Table A.3: The correlation matrix of the daily returns for the stocks of the sector 'Communication Servises'.

	F	GM	HD	LOW	MCD	NKE
F	1	0.6637	0.3922	0.3374	0.2473	0.3407
GM	0.6637	1	0.3332	0.3036	0.2634	0.2832
HD	0.3922	0.3332	1	0.6519	0.2859	0.3125
LOW	0.3374	0.3036	0.6519	1	0.2105	0.2793
MCD	0.2473	0.2634	0.2859	0.2105	1	0.3259
NKE	0.3407	0.2832	0.3125	0.2793	0.3259	1

Table A.4: The correlation matrix of the daily returns for the stocks of the sector 'Consumer Cuclical'.

	KO	GIS	PG	TGT	WMT
KO	1	0.4283	0.4569	0.2558	0.3297
GIS	0.4283	1	0.4647	0.2501	0.3292
PG	0.4569	0.4647	1	0.2562	0.3739
TGT	0.2558	0.2501	0.2562	1	0.4555
WMT	0.3297	0.3292	0.3739	0.4555	1

Table A.5: The correlation matrix of the daily returns for the stocks of the sector 'Consumer Defensive'.

	CHK	CVX	DVN	XOM	HAL	NOV	SLB
CHK	1	0.4649	0.6025	0.4031	0.4422	0.4723	0.4507
CVX	0.4649	1	0.6136	0.7670	0.5575	0.5539	0.6307
DVN	0.6025	0.6136	1	0.5483	0.5699	0.5740	0.6225
XOM	0.4031	0.7670	0.5483	1	0.5407	0.5308	0.6116
HAL	0.4422	0.5575	0.5699	0.5407	1	0.6304	0.7375
NOV	0.4723	0.5539	0.5740	0.5308	0.6304	1	0.6895
SLB	0.4507	0.6307	0.6225	0.6116	0.7375	0.6895	1

Table A.6: The correlation matrix of the daily returns for the stocks of the sector 'Energy'.

	AXP	AIG	BAC	C	JPM	MA	USB	V	WFC
AXP	1	0.5373	0.5255	0.5432	0.5479	0.5513	0.6415	0.5488	0.6242
AIG	0.5373	1	0.5280	0.5806	0.5217	0.4406	0.5118	0.4195	0.5343
BAC	0.5255	0.5280	1	0.7771	0.7103	0.3591	0.6011	0.3316	0.6432
C	0.5432	0.5806	0.7771	1	0.7693	0.4348	0.6356	0.3802	0.6608
JPM	0.5479	0.5217	0.7103	0.7693	1	0.3939	0.6362	0.3525	0.6613
MA	0.5513	0.4406	0.3591	0.4348	0.3939	1	0.4431	0.7191	0.4456
USB	0.6415	0.5118	0.6011	0.6356	0.6362	0.4431	1	0.4186	0.7809
V	0.5488	0.4195	0.3316	0.3802	0.3525	0.7191	0.4186	1	0.4076
WFC	0.6242	0.5343	0.6432	0.6608	0.6613	0.4456	0.7809	0.4076	1

Table A.7: The correlation matrix of the daily returns for the stocks of the sector 'Financial Services'.

	LLY	JNJ	MRK	PFE
LLY	1	0.4804	0.4359	0.4882
JNJ	0.4804	1	0.4524	0.5252
MRK	0.4359	0.4524	1	0.4899
PFE	0.4882	0.5252	0.4899	1

Table A.8: The correlation matrix of the daily returns for the stocks of the sector 'Healthcare'.

	MMM	BA	CAT	DAL	FDX	GE	LMT	UTX
MMM	1	0.5136	0.5415	0.3257	0.5160	0.6284	0.5359	0.6598
BA	0.5136	1	0.3692	0.2931	0.4217	0.4305	0.5470	0.5778
CAT	0.5415	0.3692	1	0.2471	0.4640	0.5510	0.3714	0.5856
DAL	0.3257	0.2931	0.2471	1	0.3874	0.3238	0.2948	0.3026
FDX	0.5160	0.4217	0.4640	0.3874	1	0.4888	0.4175	0.4910
GE	0.6284	0.4305	0.5510	0.3238	0.4888	1	0.4151	0.5319
LMT	0.5359	0.5470	0.3714	0.2948	0.4175	0.4151	1	0.5927
UTX	0.6598	0.5778	0.5856	0.3026	0.4910	0.5319	0.5927	1

Table A.9: The correlation matrix of the daily returns for the stocks of the sector 'Industrials'.

	GLW	HPQ	IBM
GLW	1	0.2562	0.3126
HPQ	0.2562	1	0.3095
IBM	0.3126	0.3095	1

Table A.10: The correlation matrix of the daily returns for the stocks of the sector 'Technology'.

Stock	2012	2013	2014
MMM	63.79	93.02	104.36
T	188.14	183.75	174.22
AA	8.79	8.3	8.6
AXP	63.51	96.53	95.17
AIG	52.11	74.74	77.06
BAC	125.13	164.91	188.14
BA	56.94	102	91.86
CAT	58.69	57.92	55.48
CHK	9.69	16.35	14.95
CVX	210.51	239.02	210.85
C	119.82	157.85	163.62
CLF	5.49	4.01	1.09
KO	162	181.84	184.33
GLW	18.55	24.93	29.21
DAL	10.1	23.38	40.59
DVN	21.12	25.11	25.03
DD	38.9	53.51	52.79
LLY	56.54	56.95	76.62
XOM	389.64	438.7	388.38
FDX	28.84	44.88	49.2
F	51.1	61.47	62.13
FCX	32.45	39.17	24.27
GE	218.41	282	254.14
GIS	26.13	31.17	32.19
GM	39.39	61.3	55.85
HAL	32.22	43.08	33.35
HPQ	27.76	53.4	73.6
HD	92.47	115.95	138.33
IBM	214.03	197.77	158.91
JPM	167.25	219.65	232.47
JNJ	194.77	258.34	291.04
LMT	29.62	47.71	60.85
LOW	39.94	51.82	66.93
MA	60.54	133.86	99.3
MCD	88.44	96.09	90.22
MRK	123.91	146.52	161.17
NOV	29.18	34.07	27.45
NEM	22.83	11.34	9.42
NKE	46.21	69.83	83.07
PFE	182.47	196	195.96
PG	185.45	220.73	245.98
SLB	92.04	117.8	108.92
TGT	38.5	39.99	48.35
USB	59.7	73.71	80.27
UTX	75.35	104.31	104.57
VZ	123.69	140.63	194.36
V	99.79	140.37	161.09
WMT	228.24	254.62	276.8
DIS	89.62	129.88	160.12
WFC	180	238.67	283.43

Table A.11: The companies' annual market capitalisation for the years 2012-2014. Market capitalisation data were taken from companiesmarketcap.com (nd) and they are expressed in $ billions.

Appendix B

Appendix for Chapter 2

This appendix provides the derivation of the properties of the intraday RV-based estimators, presented in Chapter 2.

B.1 Properties of the Intraday RV-Based Volatility Estimators

In this section, we derive the properties of the intraday RV-based estimators in Propositions 2.4.1 and 2.4.2. Recall that assuming that the intraday observed log-prices can be expressed by (2.3):

$$Y_i = Y(t_i) = X(t_i) + E(t_i)\,, \quad X(s) - X(t) = \int_t^s \mu(\varsigma)d\varsigma + \int_t^s \sigma^2(\varsigma)dW(\varsigma)\,,$$

the intraday pointwise returns for two consecutive time points are given by:

$$R_i = R(t_i) = Y_i - Y_{i-1} = [X(t_i) - X(t_{i-1})] + [E(t_i) - E(t_{i-1})] \equiv R_i^* + V_i$$

where $R_i^* = R^*(t_i)$ and $V_i = V(t_i)$. We write $\mathbb{E}[E^k(t_i)] = a_k, k = 1, \ldots, 4$ with $a_2 = \omega^2$. The proofs are done in steps and we start with some preliminary calculations involving

V_i and R_i.

B.1.1 Moments of V_i

For the first four moments of $V_i = E(t_i) - E(t_{i-1})$, recall that:

$$
\begin{aligned}
(a+b)^2 &= a^2 + 2ab + b^2, \\
(a+b)^3 &= a^3 + 3a^2b + 3ab^2 + b^3, \\
(a+b)^4 &= a^4 + 4a^3b + 6a^2b^2 + 4ab^3 + b^4.
\end{aligned}
$$

Applying to V_i and using the Assumptions 2.2.4 and 2.2.5, we obtain that:

$$
\begin{aligned}
\mathbb{E}[V_i] &= 0, \\
\mathbb{E}[V_i^2] &= \mathbb{E}[E^2(t_i) - 2E(t_i)E(t_{i-1}) + E^2(t_{i-1})] = 2a_2, \\
\mathbb{E}[V_i^3] &= \mathbb{E}[E^3(t_i) - 3E^2(t_i)E(t_{i-1}) + 3E(t_i)E^2(t_{i-1}) - E^3(t_{i-1})] = a_3 - a_3 = 0, \\
\mathbb{E}[V_i^4] &= \mathbb{E}[E^4(t_i) - 4E^3(t_i)E(t_{i-1}) + 6E^2(t_i)E^2(t_{i-1}) - 4E(t_i)E^3(t_{i-1}) + E^4(t_{i-1})] \\
&= 2a_4 + 6a_2^2 = 2(a_4 + 3a_2^2)
\end{aligned}
$$

where $\mathbb{E}[E^k(t_i)] = a_k$, for $k = 1, \ldots, 4$.

For covariance, first observe that:

$$
Cov(V_i, V_j) = 0\,, \quad \text{if } |i-j| > 1\,, \tag{B.1}
$$

and thus:

$$
\begin{aligned}
Cov(V_i, V_{i+1}) &= Cov\Big(E(t_i) - E(t_{i-1}), E(t_{i+1}) - E(t_i)\Big) = -Var\Big(E(t_i)\Big) = -a_2, \\
Cov(V_i, V_{i+1}^2) &= Cov\Big(E(t_i) - E(t_{i-1}), E^2(t_{i+1}) - 2E(t_{i+1})E(t_i) + E^2(t_i)\Big) \\
&= -2Cov\Big(E(t_i), E(t_{i+1})E(t_i) + Cov(E(t_i), E^2(t_i)\Big) = a_3,
\end{aligned}
$$

$$
\begin{aligned}
Cov(V_i^2, V_{i+1}) &= Cov\Big(E^2(t_i) - 2E(t_i)E(t_{i-1}) + E^2(t_{i-1}), E(t_{i+1}) - E(t_i)\Big) \\
&= -Cov\Big(E^2(t_i), E(t_i)) + 2Cov(E(t_i)E(t_{i-1}), E(t_i)\Big) = -a_3, \\
Cov(V_i^2, V_{i+1}^2) &= Cov\Big(E^2(t_i) - 2E(t_i)E(t_{i-1}) + E^2(t_{i-1}), E^2(t_{i+1}) - 2E(t_{i+1})E(t_i) + E^2(t_i)\Big) \\
&= -2Cov\Big(E^2(t_i), E(t_{i+1})E(t_i)\Big) + Cov\Big(E^2(t_i), E^2(t_i)\Big) \\
&\quad + 4Cov\Big(E(t_i)E(t_{i-1}), E(t_{i+1})E(t_i)\Big) - 2Cov\Big(E(t_i)E(t_{i-1}), E^2(t_i)\Big) \\
&= a_4 - a_2^2 \,.
\end{aligned}
$$

B.1.2 Conditional Moments of R_i

Since $R_i = R_i^* + V_i$, from the results from section B.1.1, it follows that:

$$
\begin{aligned}
&\mathbb{E}[R_i|R^*] = R_i^*, \\
&\mathbb{E}[R_i^2|R^*] = (R_i^*)^2 + 2\,\mathbb{E}[V_i^2] = (R_i^*)^2 + 2a_2, \\
&\mathbb{E}[R_i^3|R^*] = (R_i^*)^3 + 3R_i^*\,\mathbb{E}[V_i^2] + \mathbb{E}[V_i^3] = (R_i^*)^3 + 6a_2R_i^*, \\
&\mathbb{E}[R_i^4|R^*] = (R_i^*)^4 + 6(R_i^*)^2\,\mathbb{E}[V_i^2] + 4R_i^*\,\mathbb{E}[V_i^3] + \mathbb{E}[V_i^4] = (R_i^*)^4 + 12a_2(R_i^*)^2 + 2(a_4 + 3a_2^2)
\end{aligned}
$$

where $\mathbb{E}[E^k(t_i)] = a_k$, for $k = 1, \ldots, 4$. Similarly, the conditional variance and covariance of R_i^2 are given by:

$$
\begin{aligned}
Var[R_i^2|R^*] &= \mathbb{E}[R_i^4|R^*] - (\mathbb{E}[R_i^2|R^*])^2 = 8a_2(R_i^*)^2 + 2(a_4 + a_2^2), \\
Cov(R_i^2, R_j^2|R^*) &= Cov\Big((R_i^*)^2 + 2R_i^*V_i + V_i^2, (R_j^*)^2 + 2R_j^*V_j + V_j^2\Big) \\
&= 4R_i^*R_j^*Cov(V_i, V_j) + 2R_i^*Cov(V_i, V_j^2) + 2R_j^*Cov(V_i^2, V_j) + Cov(V_i^2, V_j^2)\,.
\end{aligned}
$$

Because of (B.1), the non-zero conditional covariance of R_i^2 occurs only in one lag:

$$
\begin{aligned}
Cov(R_i^2, R_{i+1}^2|R^*) &= 4R_i^*R_{i+1}^*Cov(V_i, V_{i+1}) + 2R_i^*Cov(V_i, V_{i+1}^2) + 2R_{i+1}^*Cov(V_i^2, V_{i+1}) \\
&\quad + Cov(V_i^2, V_{i+1}^2)
\end{aligned}
$$

$$= -4a_2 R_i^* R_{i+1}^* + 2a_3(R_i^* - R_{i+1}^*) + a_4 - a_2^2,$$

$$\begin{aligned} Cov(R_i^2, R_{i-1}^2|R^*) &= 4R_i^* R_{i-1}^* Cov(V_i, V_{i-1}) + 2R_i^* Cov(V_i, V_{i-1}^2) + 2R_{i-1}^* Cov(V_i^2, V_{i-1}) \\ &\quad + Cov(V_i^2, V_{i-1}^2) \\ &= -4a_2 R_i^* R_{i-1}^* + 2a_3(R_{i-1}^* - R_i^*) + a_4 - a_2^2 \,. \end{aligned}$$

Using the properties above, we derive the conditional mean and variance of the intraday volatility estimators in section B.1.3.

B.1.3 Conditional Mean and Variance of the RV-based Estimators

As the estimators are in the form of a weighted average of returns, the estimators can be expressed as:

$$U_Y = \sum_{i=1}^{n} w_i R_i^2$$

where

$$RV_I^{(Intr-all)} : w_i = 1, \quad stRV_I^{(Intr-all)} : w_i = \frac{1}{\Delta(t_{I,i})}, \quad wRV_I^{(Intr-all)} : w_i = \frac{1}{n_I - 1}.$$

The estimators $RV_I^{(Intr-all)}$ and $stRV_I^{(Intr-all)}$ with $n = 1$ and relevant straightforward notational modifications define $RV^{(Intr)}$ and $stRV^{(Intr)}$, respectively.

Conditional on R^*, we wish to compute the mean and the variance of U_Y:

$$\mathbb{E}[U_Y|R^*] = \sum_i w_i \, \mathbb{E}[R_i^2|R^*] \,, \quad Var[U_Y|R^*] = \sum_i \sum_j w_i w_j Cov(R_i^2, R_j^2|R^*) \,.$$

For the mean, we have that:

$$\mathbb{E}[U_Y|R^*] = \sum_i w_i \, \mathbb{E}[R_i^2|R^*] = \sum_i w_i \Big((R_i^*)^2 + 2a_2 \Big) = \sum_i w_i (R_i^*)^2 + 2a_2 \sum_i w_i \,.$$

where a_2 denotes the noise variance ω^2. With an appropriate choice of w_i, the results in

Proposition 2.4.1 follow.

For the variance, first observe that:

$$\begin{aligned} Var[U_Y|R^*] &= \sum_i \sum_j w_i w_j Cov(R_i^2, R_j^2|R^*) \\ &= \sum_i w_i \Big(w_i Var[R_i^2] + w_{i+1} Cov(R_i^2, R_{i+1}^2) + w_{i-1} Cov(R_i^2, R_{i-1}^2) \Big) . \end{aligned}$$

Using the derivations given in section B.1.2 gives:

$$\begin{aligned} &Var[U_Y|R^*] = \sum_i w_i^2 Var[R_i^2|R^*] + \sum_i w_i w_{i+1} Cov(R_i^2, R_{i+1}^2) + \sum_i w_i w_{i-1} Cov(R_i^2, R_{i-1}^2) \\ &= \sum_i w_i^2 \Big(8a_2 (R_i^*)^2 + 2(a_4 + a_2^2) \Big) + \sum_{i=1}^{n-1} w_i w_{i+1} \Big(-4a_2 R_i^* R_{i+1}^* + 2a_3 (R_i^* - R_{i+1}^*) + a_4 - a_2^2 \Big) \\ &\quad + \sum_{i=2}^{n} w_i w_{i-1} \Big(-4a_2 R_i^* R_{i-1}^* + 2a_3 (R_{i-1}^* - R_i^*) + a_4 - a_2^2 \Big) \\ &= 2(a_4 + a_2^2) \sum_{i=1}^{n} w_i^2 + 2(a_4 - a_2^2) \sum_{i=1}^{n-1} w_i w_{i+1} + 8a_2 \sum_i w_i^2 (R_i^*)^2 - 8a_2 \sum_i w_i w_{i+1} R_i^* R_{i+1}^* \\ &\quad + 4a_3 \sum_i w_i w_{i+1} (R_i^* - R_{i+1}^*) \end{aligned}$$

where $\mathbb{E}[E^k(t_i)] = a_k$, for $k = 1, \ldots, 4$ and by convention $a_2 = \omega^2$. With an appropriate choice of w_i, the results in Proposition 2.4.2 follow.

Appendix C

Appendix for Chapter 3

This appendix provides the analysis results of Chapter 3. More specifically, appendix C.1 offers the simulation study which shows that the ratio-based estimator is an accurate estimator of the optimal number of factors in the factor model when we discard a number of eigenvalues in the estimation procedure. Also, the ACF plots, as mentioned in section 3.3, are presented in C.2. Finally, the figures and the tables of the analysis procedures of sections 3.4 and 3.5 are reported in C.3 and C.4, respectively.

C.1 Simulation Study

In subsection 3.2.3, we outlined several factor retention methods, concluding that the ratio-based estimator is the estimator that best suits the theoretical framework of subsection 3.2.2. The ratio-based estimator determines the optimal number of factors that need to be contained in the factor model, utilising the eigenvalues of the Matrix (3.2) as components. As Lam et al. (2011) and Lam et al. (2012) report, instead of considering all the available -p- eigenvalues, they choose only the $p/2$ highest eigenvalues in this estimator. Consequently, the maximum number of factors that are likely to be incorporated in the factor model decreases from $(p-1)$ to $(p-1)/2$. This assumption may seem strong at first sight since the significance of numerous factors is not considered in the

ratio-based estimator; however, as we highlight in this section's simulation procedure it is a necessary condition in terms of the estimator's efficiency.

In the simulation study of Lam et al. (2012), the ratio-based estimator is evaluated under different settings, illustrating its success in selecting the true number of optimal factors when the sample size or the number of variables increase. In these works, the estimator is defined for part of the total number of eigenvalues. In this simulation study, we compare the efficiency of this estimator for three different cases. In the first case, we consider all the eigenvalues as inputs in the estimator, in the second case, we consider the $p/2$ highest eigenvalues and lastly, we consider the $p/3$ highest eigenvalues.

The simulation model is the same as Lam et al. (2012), and the estimator is examined under the same simulation settings as considered by these authors. In particular the simulation model is tested for all the possible combinations of a sample size $n = 50, 100, 200, 400, 800, 1600, 3200$ and number of variables $p = 0.2n, 0.5n, 0.8n, 1.2n$.

The factor model has the following form:

$$y = \Lambda F + \epsilon$$

where y indicates the $p \times n$ matrix of the simulated dataset. In the factor model, we choose three factors. So, Λ displays a $p \times 3$ matrix of the model's coefficients, F is a $3 \times n$ matrix representing the elements of the three factors. Besides, ϵ is a $p \times n$ matrix which contains the model's error terms. These error terms are uncorrelated random numbers generated through a multivariate normal distribution with zero mean and variance equal to one. The three factors are considered to follow an autoregressive AR(1) process, as given by the following equations:

$$F_1(t_i) = 0.6F_1(t_{i-1}) + e_1(t_i),$$
$$F_2(t_i) = -0.5F_2(t_{i-1}) + e_2(t_i),$$

p \ n	50	100	200	400	800	1600	3200
0.2n	0.03	0.16	0.505	0.735	0.92	0.965	0.98
0.5n	0.075	0.315	0.66	0.845	0.885	0.97	0.975
0.8n	0.065	0.39	0.695	0.86	0.925	0.935	0.965
1.2n	0	0	0	0	0	0	0
0.2n	0.22	0.57	0.885	0.995	1	1	1
0.5n	0.29	0.715	0.94	1	1	1	1
0.8n	0.4	0.77	0.98	1	1	1	1
1.2n	0.46	0.765	0.97	1	1	1	1
0.2n	0.23	0.57	0.885	0.995	1	1	1
0.5n	0.29	0.715	0.94	1	1	1	1
0.8n	0.4	0.77	0.98	1	1	1	1
1.2n	0.46	0.765	0.97	1	1	1	1

Table C.1: The three matrices separated by a black line show the ratio-based estimator's success rate for three different occasions based on 200 iterations. In the matrix on the top, the estimator exploits all of the eigenvalues for the estimation. In the middle matrix, the $p/2$ highest eigenvalues are used by the estimator, whereas in the matrix on the bottom, the $\lfloor p/3 \rfloor$ highest eigenvalues have been taken into consideration by the estimator. This table refers to the case where we have considered three strong factors in the model.

$$F_3(t_i) = 0.3F_3(t_{i-1}) + e_3(t_i)$$

where e_j's are considered to follow a multivariate normal distribution with zero mean, variance one, for $i = 1, ..., n$, and they are uncorrelated to each other and across time.

In terms of the factor coefficients Λ, they are set under two different conditions. Firstly, they are generated from a continuously uniform distribution taken from the interval [-1,1]. On this occasion, they signify strong factors. Secondly, the uniformly distributed random numbers of the first occasion are divided by $p^{0.25}$. On this occasion, the factors' strength decreases.

Tables C.1 and C.3 list the ratio of the successful estimates for 200 iterations as given by the simulation procedure. A black line separates both tables into three different matrices. These three matrices depict the estimator's accuracy when all the eigenvalues, one-half of the eigenvalues and one-third of the eigenvalues (from top to bottom, respec-

p \ n	50	100	200	400	800	1600	3200
0.2n	0.165	0.680	0.940	0.995	1	1	1
0.5n	0.410	0.800	0.980	1	1	1	1
0.8n	0.560	0.815	0.990	1	1	1	1
1.2n	0.590	0.820	0.990	1	1	1	1
0.2n	0.075	0.155	0.270	0.570	0.980	1	1
0.5n	0.090	0.285	0.285	0.820	0.960	1	1
0.8n	0.060	0.180	0.490	0.745	0.970	1	1
1.2n	0.090	0.180	0.310	0.760	0.915	1	1

Table C.2: The simulation results as given in Lam et al. (2012) under the same simulation settings for an aliquot of eigenvalues. The matrix on the top provides the simulation results considering three strong factors in the model, whereas the matrix on the bottom reports the simulation results for three weak factors in the model.

p \ n	50	100	200	400	800	1600	3200
0.2n	0	0.005	0.005	0.015	0.14	0.285	0.425
0.5n	0	0	0.025	0.035	0.135	0.3	0.44
0.8n	0	0	0	0.02	0.135	0.265	0.38
1.2n	0	0	0	0	0	0	0
0.2n	0.105	0.1	0.205	0.44	0.78	0.98	1
0.5n	0.075	0.09	0.33	0.585	0.905	1	1
0.8n	0.04	0.14	0.41	0.62	0.94	0.995	1
1.2n	0.105	0.15	0.335	0.65	0.925	1	1
0.2n	0.185	0.1	0.205	0.44	0.78	0.98	1
0.5n	0.085	0.09	0.33	0.585	0.905	1	1
0.8n	0.04	0.14	0.41	0.62	0.94	0.995	1
1.2n	0.105	0.15	0.335	0.65	0.925	1	1

Table C.3: The three matrices separated by a black line show the ratio-based estimator's success rate for three different occasions based on 200 iterations. In the matrix on the top, the estimator exploits all of the eigenvalues for the estimation. In the middle matrix, the $p/2$ highest eigenvalues are used by the estimator, whereas in the matrix on the bottom, the $\lfloor p/3 \rfloor$ highest eigenvalues have been taken into consideration by the estimator. This table refers to the case where we have considered three weak factors in the model.

tively) are utilised in the ratio-based estimator. Furthermore, Table C.1 refers to the case where we have considered strong factors in the model, whereas Table C.3 refers to the case of weak factors. The simulation results of Lam et al. (2012) are also presented in Table C.2 for three strong factors (matrix on the top) and three weak factors (matrix on the bottom).

As we observe in Table C.1, the estimator presents a poor performance when it makes use of all the eigenvalues. Especially the ratio-based estimator delivers completely inaccurate estimates when the number of variables exceeds the data sample size. On the other hand, when part of the eigenvalues is used in the estimation procedure, the estimator's efficiency increases intensely. As we detect, when one-half (middle matrix) or one-third (matrix on the bottom) of the highest eigenvalues are utilised, we arrive at identical results. Thus, we conclude that the effect of very low ratios when negligible eigenvalues are used in the estimation procedure affects the validity of the ratio-based estimator.

Table C.3 illustrates the simulation results where three weak factors have been incorporated into the simulation model. The ratio-based estimator reveals similar behaviour to the results in Table C.1 as regards the evolution of the precision across the sample size and the number of variables; however, the estimator's accuracy declines conspicuously. In addition, the failure of the estimator to correctly estimate the optimal number of factors when all of the eigenvalues are exploited is apparent in this table too.

C.2 ACF Plots

This appendix section presents the ACF plots of the intraday volatility estimates, taken as a time series sequence. These plots depict the autocorrelation of the volatility estimates, as derived by the intraday volatility models of Chapter 2, for the stocks with identifier code MMM, T, AA AXP, AIG, BAC, BA, CAT, CHK (nine first companies in Table 1.2). The intraday estimates are based on 30-minute intraday intervals and cover

a period of six days (i.e. 78 time lags).

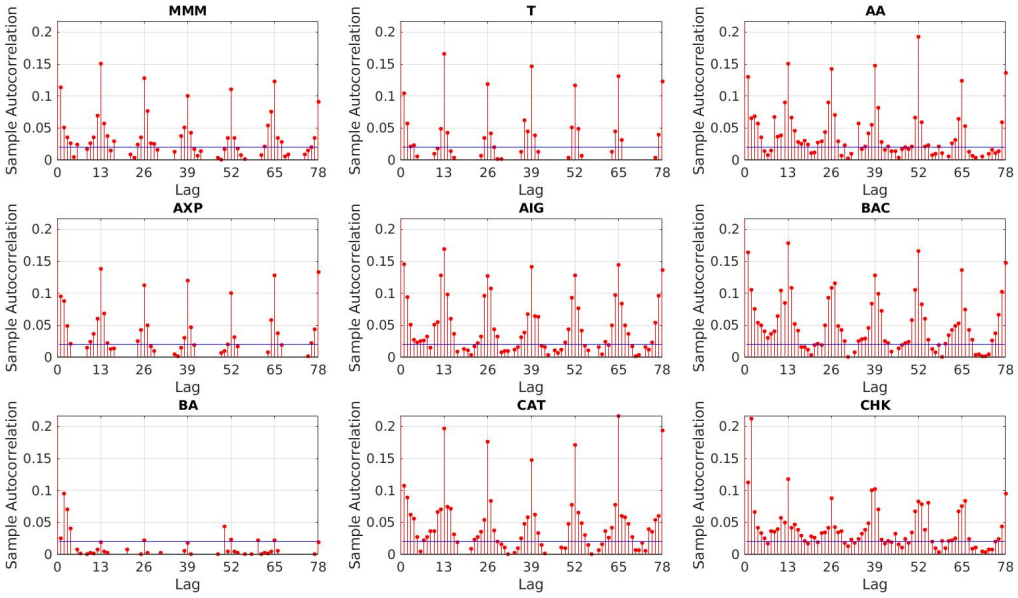

Figure C.1: ACF plots of $RV^{(Intr)}$ for the first nine stocks of Table 1.2.

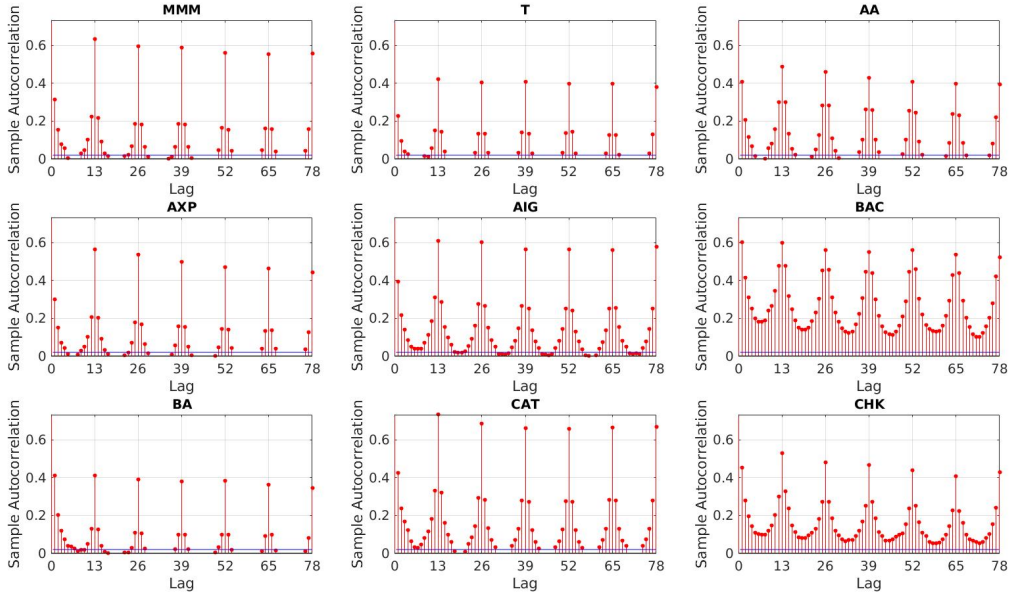

Figure C.2: ACF plots of $RV^{(Intr-all)}$ for the first nine stocks of Table 1.2.

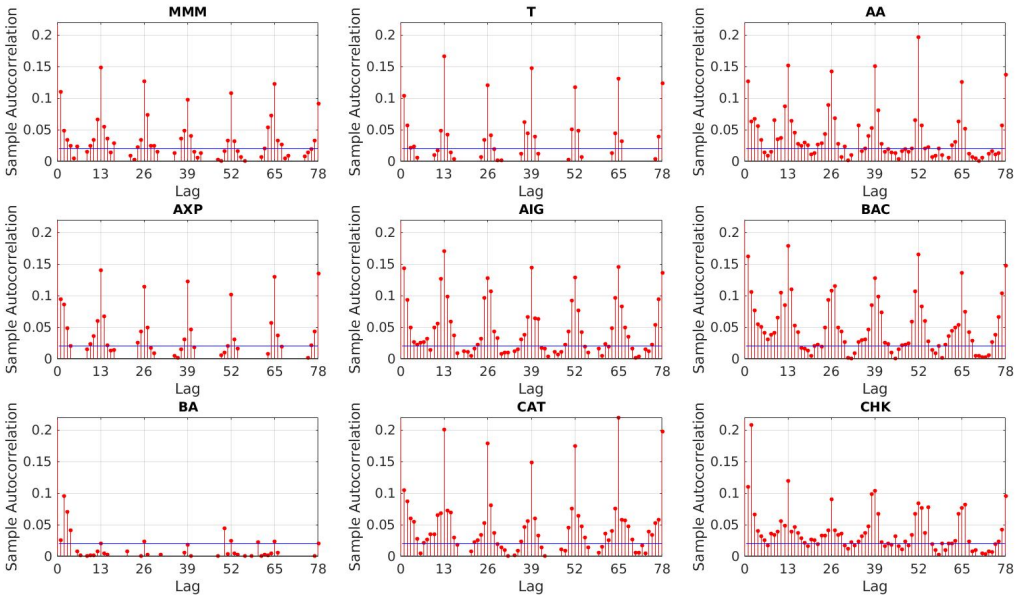

Figure C.3: ACF plots of $stRV^{(Intr)}$ for the first nine stocks of Table 1.2.

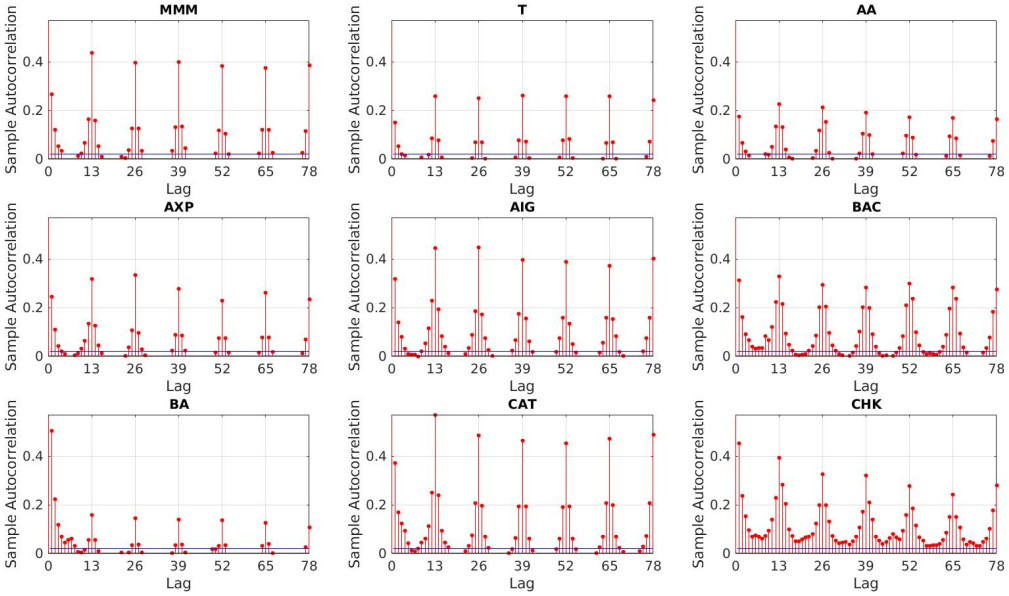

Figure C.4: ACF plots of $stRV^{(Intr-all)}$ for the first nine stocks of Table 1.2.

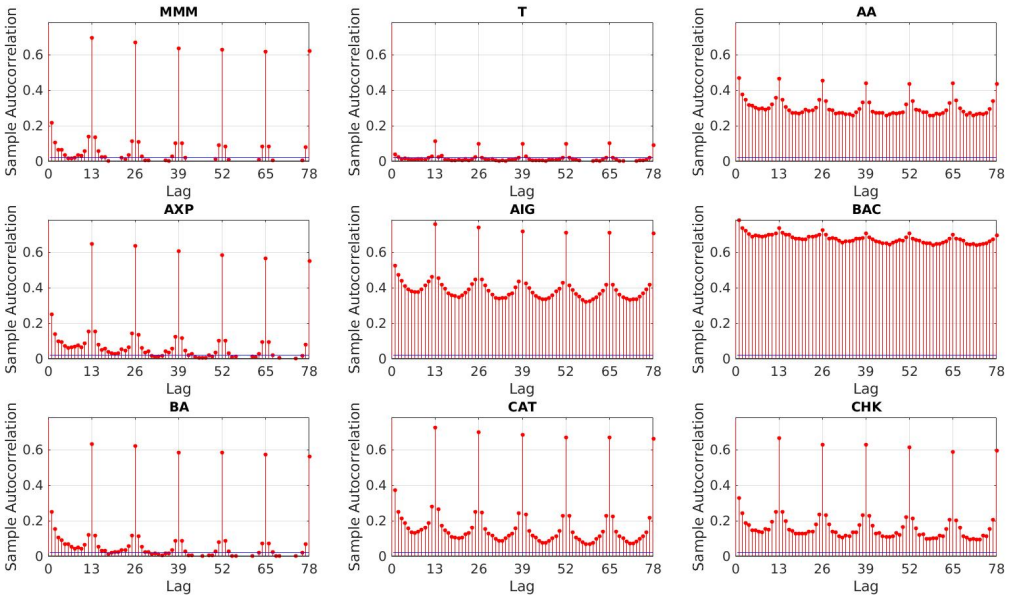

Figure C.5: ACF plots of $wRV^{(Intr-all)}$ for the first nine stocks of Table 1.2.

C.3 Factor Analysis Findings - Daily RV-Based Estimates

Stock	$\widetilde{RV}^{(all)}$	$\widetilde{RV}$	$\sum_{I=1}^{13} \widetilde{RV}_I^{(Intr-all)}$	$\sum_{I=1}^{13} \widetilde{stRV}_I^{(Intr)}$	$\sum_{I=1}^{13} \widetilde{stRV}_I^{(Intr-all)}$	$\sum_{I=1}^{13} \widetilde{wRV}_I^{(Intr-all)}$
	Factor 1	Factor 1	Factor 1	Factor 1	Factor 1	Factor 1
MMM	0	0	0	0	0	0
T	0	0.0001	0	0.0001	0	0
AA	0.0032	0.0009	0.0033	0.001	0.0023	0
AXP	0	0	0	0	0	0.0002
AIG	0.0001	0.0003	0.0001	0.0003	0.0001	0
BAC	0.0001	0	0.0001	0	0	0
BA	0.0005	0	0.0004	0	0.0002	0.0001
CAT	0.0004	0.0011	0.0004	0.001	0.0009	0.0001
CHK	0	0	0	0.0001	0.0025	0.0002
CVX	0.0003	0	0.0003	0	0	0
C	0	0	0	0	0	0
CLF	0.9739	0.976	0.9737	0.9775	0.9717	0.9893
KO	0	0	0	0	0	0
GLW	0.0019	0.0005	0.002	0.0006	0.0018	0
DAL	0.0032	0.0025	0.0033	0.0017	0.0027	0
DVN	0.0003	0.0005	0.0004	0.0003	0.0005	0.0033
DD	0	0.0001	0	0.0001	0	0
LLY	0.0005	0.0006	0.0005	0.0006	0.0005	0.0004
XOM	0.0001	0.0003	0.0001	0.0002	0.0001	0
FDX	0	0.0001	0	0	0	0.0002
F	0	0.0001	0	0.0001	0.0002	0
FCX	0.0006	0.0009	0.0005	0.0011	0.0002	0.0005
GE	0.0001	0.0001	0.0001	0.0001	0.0003	0
GIS	0	0	0	0	0	0
GM	0.0003	0.005	0.0003	0.004	0	0.0001
HAL	0	0.0007	0	0.001	0.0005	0.0002
HPQ	0.0005	0.0001	0.0005	0.0001	0.0004	0
HD	0.0001	0.0002	0.0001	0.0002	0.0002	0
IBM	0.0001	0	0.0001	0	0	0
JPM	0.0001	0.0004	0.0001	0.0003	0.0001	0
JNJ	0.0017	0.0031	0.0017	0.0031	0.0033	0
LMT	0.0003	0	0.0003	0	0.0001	0.0006
LOW	0.0003	0	0.0003	0	0.0002	0
MA	0.001	0.0022	0.0009	0.0023	0.0009	0.0022
MCD	0.0001	0.0003	0.0001	0.0003	0.0001	0.0001
MRK	0.0002	0.0006	0.0002	0.0006	0.0004	0
NOV	0.0008	0.0009	0.0008	0.0009	0.0002	0.0002
NEM	0.0055	0.0006	0.0055	0.0005	0.0053	0.0001
NKE	0	0.0003	0	0.0002	0	0.0001
PFE	0	0.0002	0	0.0002	0	0
PG	0	0.0001	0	0.0001	0.0001	0
SLB	0.0004	0.0008	0.0004	0.0006	0.0005	0.0001
TGT	0	0.0001	0	0.0001	0	0.0006
USB	0	0	0	0	0	0
UTX	0.0002	0	0.0002	0	0.0001	0.0007
VZ	0.0001	0	0.0001	0	0.0002	0
V	0.0023	0.0003	0.0023	0.0003	0.003	0
WMT	0	0	0	0	0.0001	0
DIS	0.0005	0	0.0005	0	0.0001	0
WFC	0.0002	0	0.0001	0	0	0

Table C.4: Squared factor loadings.

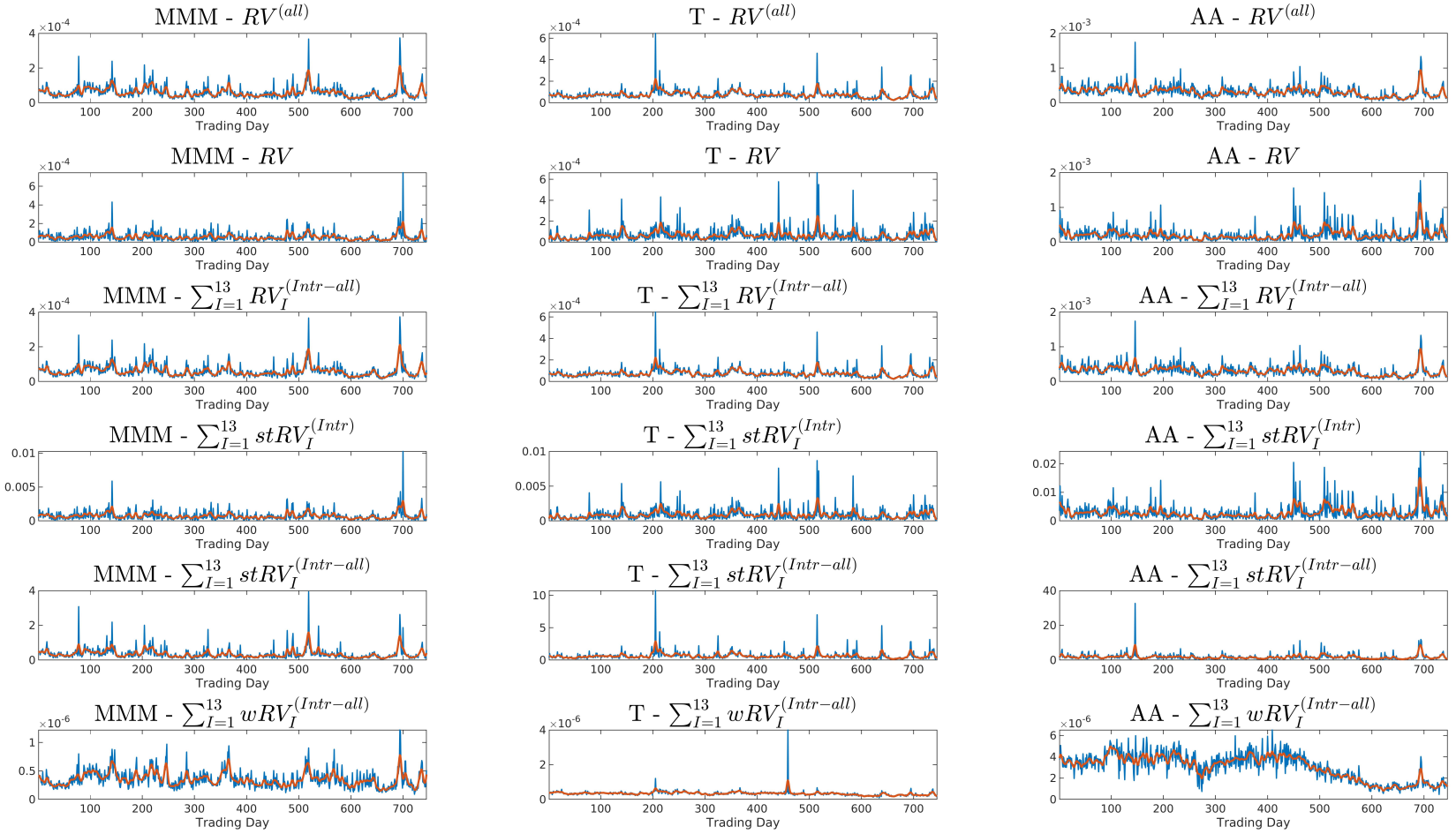

Figure C.6: This figure depicts the daily volatility estimates against the mean function of the estimates for the first three stocks of Table 1.2.

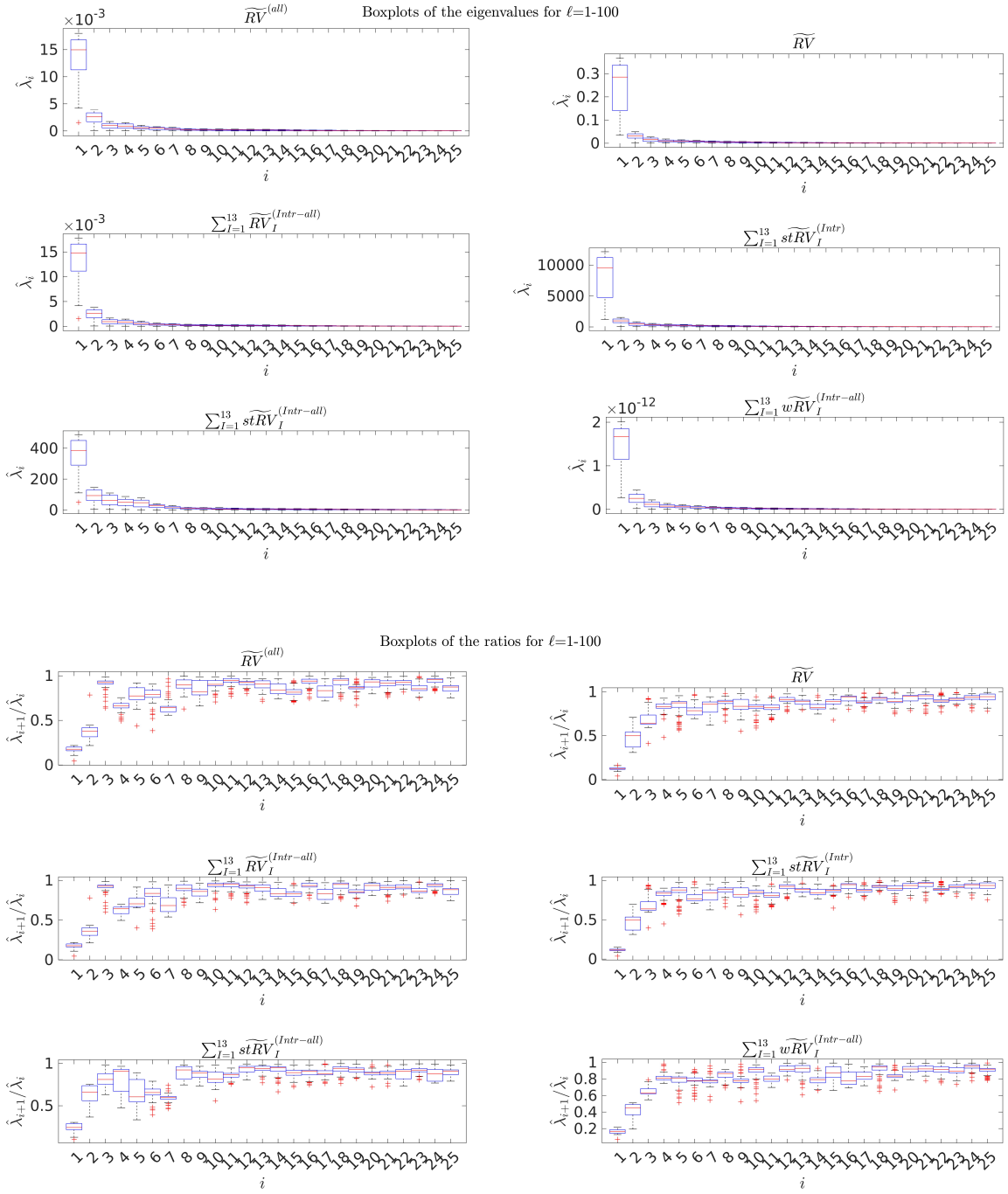

Figure C.7: These boxplots show the range of the estimated eigenvalues (first three row-panels) and the range of their ratios (last three row-panels) for the daily RV-based volatility estimates for $\ell = 1 - 100$.

C.4 Factor Analysis Findings - Intraday RV-Based Estimates

Stock \ I. I.	1	2	3	4	5	6	7	8	9	10	11	12	13
MMM	0	0.0001	0	0	0	0	0	0	0	0	0	0	0
T	0	0	0	0	0	0	0	0	0	0	0	0	0
AA	0.0001	0.0004	0.0024	0	0	0	0.0001	0	0	0	0	0	0
AXP	0.0002	0	0	0	0	0	0	0	0	0	0	0	0
AIG	0.0009	0	0	0	0	0	0	0	0	0	0	0	0
BAC	0	0.0001	0	0	0.0001	0	0	0	0	0	0.0001	0	0
BA	0.0005	0	0	0	0	0	0	0	0	0	0	0	0
CAT	0.0032	0.0001	0.0001	0.0001	0	0	0	0	0	0	0.0001	0	0
CHK	0.0095	0.0005	0.0012	0.0008	0.0007	0	0	0	0.0005	0	0.0045	0.0002	0.0004
CVX	0.0002	0	0	0.0001	0	0	0	0	0	0	0	0	0
C	0.0004	0	0	0	0	0	0	0	0	0	0	0	0
CLF	0.7554	0.0911	0.0428	0.0001	0.004	0.0063	0.0023	0.0017	0.0001	0.0013	0.0004	0.0003	0.0004
KO	0.0001	0.0001	0	0	0	0	0.0002	0	0	0	0	0	0
GLW	0.0005	0	0.0008	0.0003	0.0001	0	0	0	0	0	0	0.0001	0
DAL	0.001	0	0.0005	0	0.0002	0	0	0	0.002	0.0001	0.0053	0.0005	0.0006
DVN	0.0002	0.0001	0.0001	0.0003	0	0	0	0	0	0	0	0	0
DD	0	0	0.0002	0	0	0	0	0	0	0	0	0	0
LLY	0.0002	0.0001	0.0002	0.0002	0	0	0	0	0	0	0	0	0
XOM	0.0002	0	0	0	0	0	0	0	0	0	0	0	0
FDX	0	0.0004	0	0.0002	0	0	0	0	0	0	0	0	0
F	0.0017	0.0002	0.0002	0	0.0001	0	0	0	0	0	0.0001	0	0
FCX	0.0069	0.0012	0.0025	0.0011	0.0002	0.0001	0.0008	0	0	0	0	0	0
GE	0	0	0	0	0	0	0	0	0	0	0	0	0
GIS	0.0001	0	0	0	0	0	0	0	0	0	0	0	0
GM	0.0006	0	0	0	0	0	0	0.0001	0	0.0001	0	0	0
HAL	0.001	0	0	0	0	0	0.0004	0	0	0	0.0001	0	0
HPQ	0.0058	0	0.0004	0.0003	0.0006	0	0.0008	0.0001	0.0001	0	0.0019	0.0002	0.0001
HD	0.0005	0	0	0	0	0	0	0	0	0	0	0	0
IBM	0	0	0	0	0	0	0	0	0	0	0	0	0
JPM	0.0004	0	0	0	0	0	0	0	0	0	0	0.0001	0
JNJ	0.0015	0	0.0003	0.0003	0.0003	0	0.0002	0	0.0001	0.0001	0	0.0003	0
LMT	0.0004	0	0	0	0	0	0	0	0	0	0	0	0
LOW	0.0001	0	0	0	0	0	0	0	0	0	0	0	0
MA	0.0041	0.0001	0.0008	0.0002	0	0	0.0002	0.0001	0	0	0.0001	0	0
MCD	0	0	0.0001	0	0	0	0	0	0	0	0	0	0
MRK	0	0	0.0006	0.0002	0	0	0	0	0	0	0	0	0
NOV	0.0017	0	0.0002	0	0	0	0.0001	0	0	0	0	0	0
NEM	0.0001	0.0014	0.0004	0	0	0	0.0001	0	0	0	0	0	0
NKE	0.0008	0.0002	0	0.0002	0	0	0	0	0	0	0	0	0
PFE	0.0002	0	0	0	0	0	0.0002	0	0	0	0	0	0
PG	0.0001	0	0	0	0	0	0	0	0	0	0	0	0
SLB	0.0013	0.0003	0	0	0.0001	0	0.0003	0	0	0.0002	0	0	0
TGT	0	0	0.0001	0	0	0	0	0	0	0	0	0	0
USB	0	0	0	0	0	0	0	0	0	0	0.0001	0	0
UTX	0	0	0	0	0	0	0	0	0	0	0	0	0
VZ	0.0001	0	0	0	0	0	0	0	0	0	0	0	0
V	0.0004	0	0	0	0	0	0	0	0	0	0	0	0
WMT	0	0	0	0	0	0	0	0	0	0	0	0	0
DIS	0	0	0	0	0	0	0	0	0	0	0	0	0
WFC	0.0001	0	0.0001	0	0	0	0	0	0	0	0.0003	0	0

Table C.5: This table illustrates the squared loadings of factor 1 for the intraday volatility estimates of $\widetilde{RV}^{(Intr)}$. The rows refer to the particular stock and the columns refer to the specific intraday interval.

Stock \ I. I.	1	2	3	4	5	6	7	8	9	10	11	12	13
MMM	0.0001	0	0	0	0	0	0	0	0	0	0	0	0
T	0.0001	0	0	0	0	0	0	0	0	0	0	0	0
AA	0	0	0	0	0	0	0	0	0	0	0	0.0001	0
AXP	0.0007	0.0001	0	0	0	0	0	0.0002	0.0001	0	0	0	0.0001
AIG	0	0	0	0	0	0	0	0	0	0	0	0	0
BAC	0	0	0	0	0	0	0	0	0	0	0	0	0
BA	0.0001	0	0	0	0	0	0	0	0	0	0	0	0
CAT	0.0001	0	0	0	0	0	0	0	0	0	0	0	0
CHK	0.0001	0	0	0	0.0001	0.0001	0.0002	0	0.0001	0	0	0.0002	0
CVX	0.0001	0	0	0	0	0	0	0	0	0	0	0	0
C	0	0	0	0	0	0	0	0	0	0	0	0	0
CLF	0.4738	0.0968	0.031	0.0199	0.0003	0.0073	0.0023	0.0103	0.001	0.0003	0.0027	0.3089	0.005
KO	0	0	0	0	0	0	0	0	0	0	0	0	0
GLW	0	0	0.0003	0.0001	0	0	0.0003	0	0	0	0	0	0
DAL	0.0002	0	0	0.0002	0	0	0.0001	0	0.0001	0.0002	0	0.0032	0.0001
DVN	0	0	0	0.0002	0	0	0	0	0.0001	0	0	0	0
DD	0	0	0	0.0001	0	0	0	0	0	0	0	0	0
LLY	0.0003	0.0001	0	0	0	0	0	0	0	0	0	0	0
XOM	0.0016	0.0002	0	0	0	0.0002	0	0.0004	0.0003	0	0	0	0.0002
FDX	0.0001	0	0	0	0	0	0	0	0	0	0	0	0
F	0.0001	0	0	0	0	0	0	0	0	0	0	0	0
FCX	0.0013	0.0001	0.0001	0	0	0	0	0	0	0	0	0.0001	0
GE	0	0	0	0	0	0	0	0	0	0	0	0	0
GIS	0	0	0	0	0	0	0	0	0	0	0	0	0
GM	0.0002	0	0	0	0	0	0	0	0	0	0	0	0
HAL	0.0007	0.0003	0	0	0	0	0	0	0	0	0	0.0001	0
HPQ	0	0	0	0.0002	0	0	0	0	0.0001	0.0001	0	0	0
HD	0.0001	0	0	0	0	0	0	0	0	0	0	0	0
IBM	0	0	0	0	0	0	0	0	0	0	0	0	0
JPM	0.0008	0.0001	0	0	0	0	0	0	0	0	0	0.0001	0
JNJ	0.0095	0.0014	0.0004	0.0001	0	0	0	0	0	0	0	0.0011	0
LMT	0.0002	0	0	0	0	0	0	0	0	0	0	0	0
LOW	0	0	0	0	0	0	0	0	0	0	0	0	0
MA	0.0002	0	0	0.0004	0	0	0.0002	0.0001	0	0	0.0001	0	0
MCD	0	0	0	0	0	0	0	0	0	0	0	0	0
MRK	0	0.0001	0	0.0001	0	0	0	0	0	0	0	0.0001	0
NOV	0.0001	0	0	0.0002	0	0	0	0	0	0	0	0.0001	0
NEM	0.0001	0	0	0.0004	0	0	0	0	0	0	0	0.0001	0
NKE	0.0001	0	0	0	0	0	0	0	0	0	0	0	0
PFE	0	0	0	0	0	0	0	0	0	0	0	0	0
PG	0.0001	0	0	0	0	0	0	0	0	0	0	0	0
SLB	0.0009	0.0001	0	0	0	0	0	0.0001	0	0	0	0	0
TGT	0	0	0	0	0	0	0	0	0	0	0	0	0
USB	0	0	0	0	0	0	0	0	0	0	0	0	0
UTX	0.0001	0	0	0	0	0	0	0	0	0	0	0	0
VZ	0	0	0	0	0	0	0	0	0	0	0	0	0
V	0.0001	0	0	0.0003	0	0	0.0001	0.0001	0	0	0	0	0
WMT	0.0001	0	0	0	0	0	0	0	0	0	0	0	0
DIS	0.0017	0.0002	0	0	0	0.0004	0	0.0008	0.0006	0.0001	0	0.0002	0.0005
WFC	0	0	0	0	0	0	0	0	0	0	0	0	0

Table C.6: This table illustrates the squared loadings of factor 1 for the intraday volatility estimates of $\widetilde{RV}^{(Intr-all)}$. The rows refer to the particular stock and the columns refer to the specific intraday interval.

Stock \ I. I.	1	2	3	4	5	6	7	8	9	10	11	12	13
MMM	0	0.0001	0	0	0	0	0	0	0	0	0	0	0
T	0	0	0	0	0	0	0	0	0	0	0	0	0
AA	0.0001	0.0005	0.0032	0.0001	0	0	0.0001	0	0	0	0	0	0
AXP	0.0001	0	0	0	0	0	0	0	0	0	0	0	0
AIG	0.0009	0	0	0	0	0	0	0	0	0	0	0	0
BAC	0	0.0001	0	0	0.0001	0	0	0	0	0	0.0001	0	0
BA	0.0004	0	0	0	0	0	0	0	0	0	0	0	0
CAT	0.0031	0.0001	0.0001	0	0	0	0	0	0	0	0.0001	0	0
CHK	0.011	0.0004	0.0015	0.0009	0.0008	0	0	0	0.0007	0	0.0053	0.0002	0.0006
CVX	0.0002	0	0	0.0001	0	0	0	0	0	0	0	0	0
C	0.0004	0	0	0	0	0	0	0	0	0	0	0	0
CLF	0.7429	0.0795	0.0647	0.0004	0.004	0.0063	0.0016	0.001	0.0002	0.001	0.0004	0.0004	0.0004
KO	0.0001	0.0001	0	0	0	0	0.0002	0	0	0	0	0	0
GLW	0.0007	0	0.001	0.0002	0.0001	0	0	0	0	0	0	0.0001	0
DAL	0.0014	0	0.0004	0	0.0001	0	0	0	0.0018	0.0001	0.0045	0.0004	0.0005
DVN	0.0002	0.0001	0.0001	0.0003	0	0	0	0	0	0	0	0	0
DD	0	0	0.0003	0	0	0	0	0	0	0	0	0	0
LLY	0.0002	0.0001	0.0002	0.0002	0	0	0	0	0	0	0	0	0
XOM	0.0002	0	0	0	0	0	0	0	0	0	0	0	0
FDX	0	0.0004	0	0.0002	0	0	0	0	0	0	0	0	0
F	0.0017	0.0002	0.0002	0	0.0001	0	0	0	0	0	0	0	0
FCX	0.0074	0.0011	0.0031	0.0011	0.0002	0.0001	0.0007	0	0	0	0	0	0
GE	0	0	0	0	0	0	0	0	0	0	0	0	0
GIS	0.0001	0	0	0	0	0	0	0	0	0	0	0	0
GM	0.0006	0	0	0	0	0	0	0.0001	0	0.0001	0	0	0
HAL	0.001	0	0	0	0	0	0.0004	0	0	0	0.0001	0	0
HPQ	0.0054	0	0.0004	0.0003	0.0005	0	0.0007	0	0.0001	0	0.0017	0.0002	0.0001
HD	0.0005	0	0	0	0	0	0	0	0	0	0	0	0
IBM	0	0	0	0	0	0	0	0	0	0	0	0	0
JPM	0.0004	0	0	0	0	0	0	0	0	0	0	0.0001	0
JNJ	0.0015	0	0.0003	0.0003	0.0003	0	0.0002	0	0.0001	0.0001	0	0.0003	0
LMT	0.0004	0	0	0	0	0	0	0	0	0	0	0	0
LOW	0.0001	0	0	0	0	0	0	0	0	0	0	0	0
MA	0.0045	0.0001	0.0009	0.0002	0	0	0.0002	0.0001	0	0	0.0001	0	0
MCD	0	0	0.0001	0	0	0	0	0	0	0	0	0	0
MRK	0	0	0.0006	0.0002	0	0	0	0	0	0	0	0	0
NOV	0.0018	0	0.0002	0	0	0	0	0	0	0	0	0	0
NEM	0.0001	0.0014	0.0004	0.0001	0	0	0.0001	0	0	0	0	0	0
NKE	0.0007	0.0002	0	0.0002	0	0	0	0	0	0	0	0	0
PFE	0.0003	0	0	0	0	0	0.0002	0	0	0	0	0	0
PG	0.0001	0	0	0	0	0	0	0	0	0	0	0	0
SLB	0.0011	0.0003	0	0	0.0001	0	0.0003	0	0	0.0002	0	0	0
TGT	0	0	0.0002	0	0	0	0	0	0	0	0	0	0
USB	0	0	0	0	0	0	0	0	0	0	0.0001	0	0
UTX	0	0	0	0	0	0	0	0	0	0	0	0	0
VZ	0.0001	0	0	0	0	0	0	0	0	0	0	0	0
V	0.0005	0	0	0	0	0	0	0	0	0	0	0	0
WMT	0	0	0	0	0	0	0	0	0	0	0	0	0
DIS	0	0	0	0	0	0	0	0	0	0	0	0	0
WFC	0.0001	0	0.0001	0	0	0	0	0	0	0	0.0003	0	0

Table C.7: This table illustrates the squared loadings of factor 1 for the intraday volatility estimates of $\widetilde{stRV}^{(Intr)}$. The rows refer to the particular stock and the columns refer to the specific intraday interval.

Stock \ I. I.	1	2	3	4	5	6	7	8	9	10	11	12	13
MMM	0.0001	0	0	0	0	0	0	0	0	0	0	0	0
T	0.0001	0	0	0	0	0	0	0	0	0	0	0	0
AA	0	0	0	0.0001	0	0	0	0	0	0	0	0	0
AXP	0.0003	0	0	0	0	0	0	0	0.0001	0	0	0.0001	0.0002
AIG	0	0	0	0	0	0	0	0	0	0	0	0	0
BAC	0.0001	0	0	0	0	0	0	0	0	0	0	0	0
BA	0.0004	0	0	0	0	0	0	0	0	0	0	0	0
CAT	0.0001	0	0	0	0	0	0	0	0	0	0	0	0
CHK	0	0	0	0.0001	0	0	0.0001	0	0	0	0	0.0005	0
CVX	0.0001	0	0	0	0	0	0	0	0	0	0	0	0
C	0	0	0	0	0	0	0	0	0	0	0	0.0001	0
CLF	0.4984	0.0493	0.0212	0.0139	0.0001	0.0003	0.0003	0.0008	0	0.0019	0.0024	0.3475	0.0035
KO	0	0	0	0	0	0	0	0	0	0	0	0	0
GLW	0	0	0.0002	0.0001	0	0	0.0002	0	0	0	0	0	0
DAL	0.0001	0	0	0.0003	0	0	0.0001	0.0002	0	0.0002	0.0001	0.0056	0.0002
DVN	0	0	0	0.0001	0	0	0	0	0	0	0	0	0
DD	0	0	0	0	0	0	0	0	0	0	0	0	0
LLY	0.0003	0	0	0	0	0	0	0	0	0	0	0	0
XOM	0.0016	0.0001	0	0	0	0.0001	0	0.0002	0.0002	0	0	0.0002	0.0007
FDX	0	0	0	0	0	0	0	0	0	0	0	0	0
F	0.0003	0.0001	0	0.0001	0	0	0	0	0	0	0	0	0
FCX	0.0011	0.0001	0	0	0	0	0	0	0	0	0	0.0001	0
GE	0	0	0	0	0	0	0	0	0	0	0	0	0
GIS	0	0	0	0	0	0	0	0	0	0	0	0	0
GM	0.0005	0	0	0	0	0	0	0	0	0	0	0	0
HAL	0.0009	0.0001	0	0	0	0	0	0	0	0	0	0.0003	0
HPQ	0.0001	0	0	0.0004	0	0	0	0	0	0	0	0	0
HD	0	0	0	0	0	0	0	0	0	0	0	0	0
IBM	0.0001	0	0	0	0	0	0	0	0	0	0	0	0
JPM	0.0014	0.0001	0	0	0	0	0	0	0	0	0	0.0002	0
JNJ	0.0222	0.0017	0.0007	0.0001	0	0	0	0	0	0	0	0.0027	0
LMT	0.0001	0	0	0	0	0	0	0	0	0	0	0	0
LOW	0	0	0	0	0	0	0	0	0	0	0	0	0
MA	0.0002	0	0	0.0006	0	0	0.0002	0	0	0	0.0002	0	0
MCD	0	0	0	0	0	0	0	0	0	0	0	0	0
MRK	0	0	0	0	0	0	0	0	0	0	0	0.0001	0
NOV	0	0	0	0	0	0	0	0	0	0	0	0	0
NEM	0.0002	0	0	0.0004	0	0	0.0001	0	0	0	0	0	0
NKE	0.0001	0	0	0	0	0	0	0	0	0	0	0	0
PFE	0	0	0	0	0	0	0	0	0	0	0	0	0
PG	0.0001	0	0	0	0	0	0	0	0	0	0	0	0
SLB	0.001	0	0	0	0	0	0	0	0	0	0	0.0001	0.0001
TGT	0	0	0	0	0	0	0	0	0	0	0	0	0
USB	0	0	0	0	0	0	0	0	0	0	0	0	0
UTX	0.0001	0	0	0	0	0	0	0	0	0	0	0	0
VZ	0	0	0	0	0	0	0	0	0	0	0	0	0
V	0.0002	0	0	0.0005	0	0	0.0001	0	0	0	0	0	0
WMT	0	0	0	0	0	0	0	0	0	0	0	0	0.0001
DIS	0.002	0.0002	0	0.0001	0	0.0002	0	0.0005	0.0006	0.0001	0	0.0007	0.0019
WFC	0	0	0	0	0	0	0	0	0	0	0	0	0.0001

Table C.8: This table illustrates the squared loadings of factor 1 for the intraday volatility estimates of $\widetilde{stRV}^{(Intr-all)}$. The rows refer to the particular stock and the columns refer to the specific intraday interval.

Stock \ I. I.	1	2	3	4	5	6	7	8	9	10	11	12	13
MMM	0.0003	0	0	0	0	0	0	0	0	0	0	0	0
T	0	0	0	0	0	0	0	0	0	0	0	0	0
AA	0.0003	0.0001	0	0	0	0.0001	0.0004	0	0	0	0	0	0
AXP	0.0024	0.0004	0.0001	0	0.0001	0	0	0.0001	0.0002	0	0	0	0.0001
AIG	0	0	0	0	0	0	0	0	0	0	0	0	0
BAC	0.0001	0	0	0	0	0	0	0	0	0	0	0	0
BA	0.0003	0	0	0	0.0001	0	0.0001	0	0	0	0	0	0
CAT	0	0	0	0	0	0	0	0	0	0	0	0	0
CHK	0.0001	0	0	0	0	0.0007	0.0007	0	0.0002	0.0001	0	0	0
CVX	0	0	0	0	0	0	0	0	0	0	0	0	0
C	0	0	0	0	0	0	0	0	0	0	0	0	0
CLF	0.3594	0.3041	0.056	0.0411	0.0396	0.0261	0.0065	0.0002	0.0003	0.0021	0.0002	0.1185	0.0021
KO	0	0	0	0	0	0	0	0	0	0	0	0	0
GLW	0.0001	0.0003	0.0005	0	0	0.0002	0.0002	0	0	0	0	0.0001	0
DAL	0	0	0	0	0	0.0007	0.0003	0	0.0003	0.0002	0.0001	0.0004	0.0001
DVN	0.0007	0.0001	0	0	0.0002	0	0	0.0001	0.0007	0	0.0001	0.0001	0.0001
DD	0.0006	0.0001	0.0001	0.0001	0	0.0001	0	0	0	0	0	0	0
LLY	0.0002	0.0003	0	0	0.0001	0.0002	0	0	0	0.0001	0.0001	0.0001	0
XOM	0.0005	0.0001	0	0	0	0	0	0.0001	0.0001	0	0	0	0.0001
FDX	0	0.0003	0	0	0	0.0001	0.0006	0	0.0001	0	0	0.0001	0
F	0.0001	0	0	0.0002	0	0	0.0001	0	0	0	0	0	0
FCX	0.0006	0	0.0001	0	0	0.0001	0.0001	0	0	0	0	0	0
GE	0	0	0	0	0	0	0	0	0	0	0	0	0
GIS	0.0005	0.0001	0	0	0	0.0001	0	0	0.0001	0	0	0	0
GM	0	0.0001	0	0	0	0	0	0	0	0	0	0	0
HAL	0.0006	0.0004	0	0	0.0001	0	0	0	0	0.0001	0.0001	0	0
HPQ	0.0002	0	0	0	0.0001	0.0001	0	0	0.0001	0.0001	0	0	0
HD	0	0	0	0	0	0	0	0	0	0	0	0	0
IBM	0	0	0	0	0	0	0	0	0	0	0	0	0
JPM	0	0	0	0	0	0	0	0	0	0	0	0	0
JNJ	0.0006	0.0001	0	0	0	0	0	0	0	0	0	0.0001	0
LMT	0.0014	0.0001	0	0.0001	0.0005	0	0	0.0002	0	0.0005	0	0	0.0001
LOW	0.0001	0.0001	0	0	0	0.0001	0	0	0	0	0	0	0
MA	0.0014	0.001	0.0008	0.0003	0.0012	0.0005	0.0002	0.0005	0.0003	0	0.0002	0	0
MCD	0	0	0	0	0	0	0	0	0	0	0	0	0
MRK	0.0001	0.0001	0	0	0	0	0	0	0	0	0	0	0
NOV	0	0	0	0	0	0.0004	0	0.0001	0	0	0	0	0
NEM	0	0.0001	0.0001	0	0	0.0001	0	0	0	0.0001	0	0.0006	0
NKE	0.0004	0.0002	0	0	0	0	0	0	0	0	0	0	0
PFE	0	0	0	0	0	0	0	0	0	0	0	0	0
PG	0.0002	0	0	0	0	0	0	0	0	0	0	0	0
SLB	0.0001	0	0.0001	0	0	0.0002	0	0	0	0	0	0.0001	0
TGT	0.0007	0.0001	0	0.0001	0.0001	0.0001	0	0	0	0	0	0	0
USB	0.0001	0	0	0	0	0	0	0	0	0	0	0	0
UTX	0.0014	0.0005	0.0001	0	0.0002	0.0001	0.0001	0	0	0	0	0	0
VZ	0	0	0	0	0	0	0	0	0	0	0	0	0
V	0.0001	0.0002	0.0001	0	0.0001	0.0001	0.0005	0.0005	0	0	0	0	0
WMT	0	0	0	0	0	0	0	0	0	0	0	0	0
DIS	0.0004	0.0001	0	0	0	0	0	0.0001	0.0001	0	0	0	0.0001
WFC	0	0	0	0	0	0	0	0	0	0	0	0	0

Table C.9: This table illustrates the squared loadings of factor 1 for the intraday volatility estimates of $\widetilde{wRV}^{(Intr-all)}$. The rows refer to the particular stock and the columns refer to the specific intraday interval.

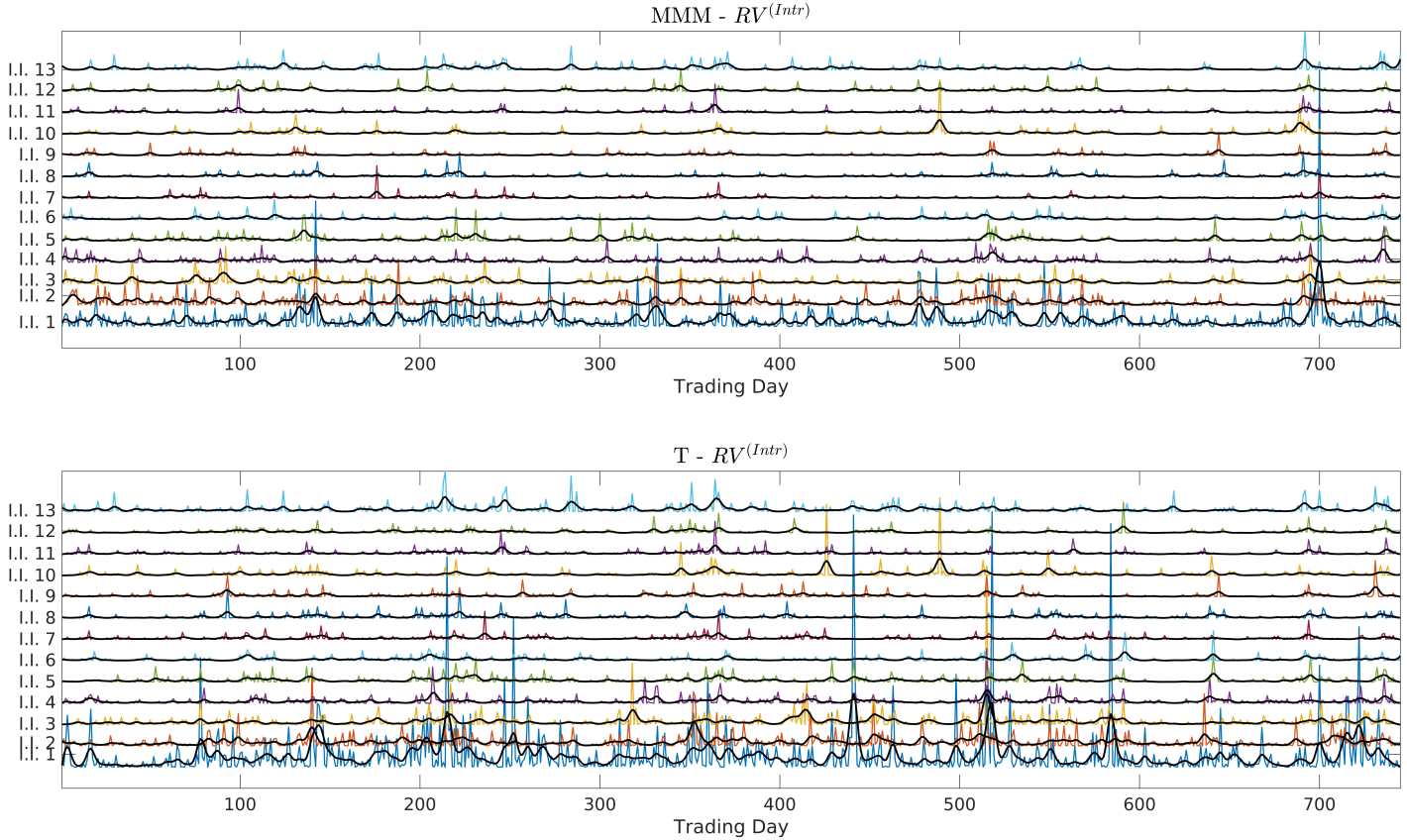

Figure C.8: This figure plots the volatility estimates of $RV^{(Intr)}$ against the mean functions of the stocks 3M Co. (MMM) and AT&T Inc. (T) for each intraday interval over the days.

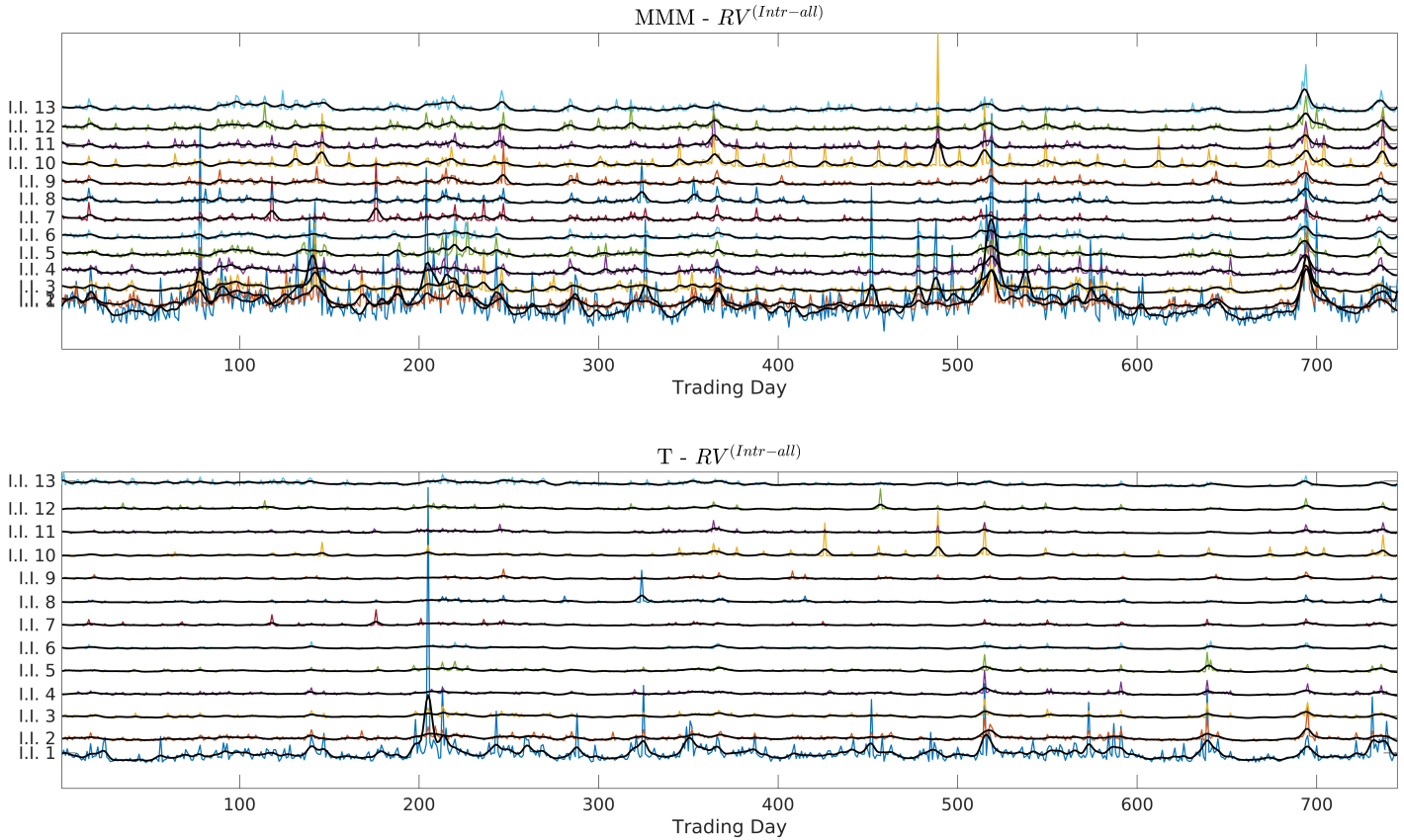

Figure C.9: This figure plots the volatility estimates of $RV^{(Intr-all)}$ against the mean functions of the stocks 3M Co. (MMM) and AT&T Inc. (T) for each intraday interval over the days.

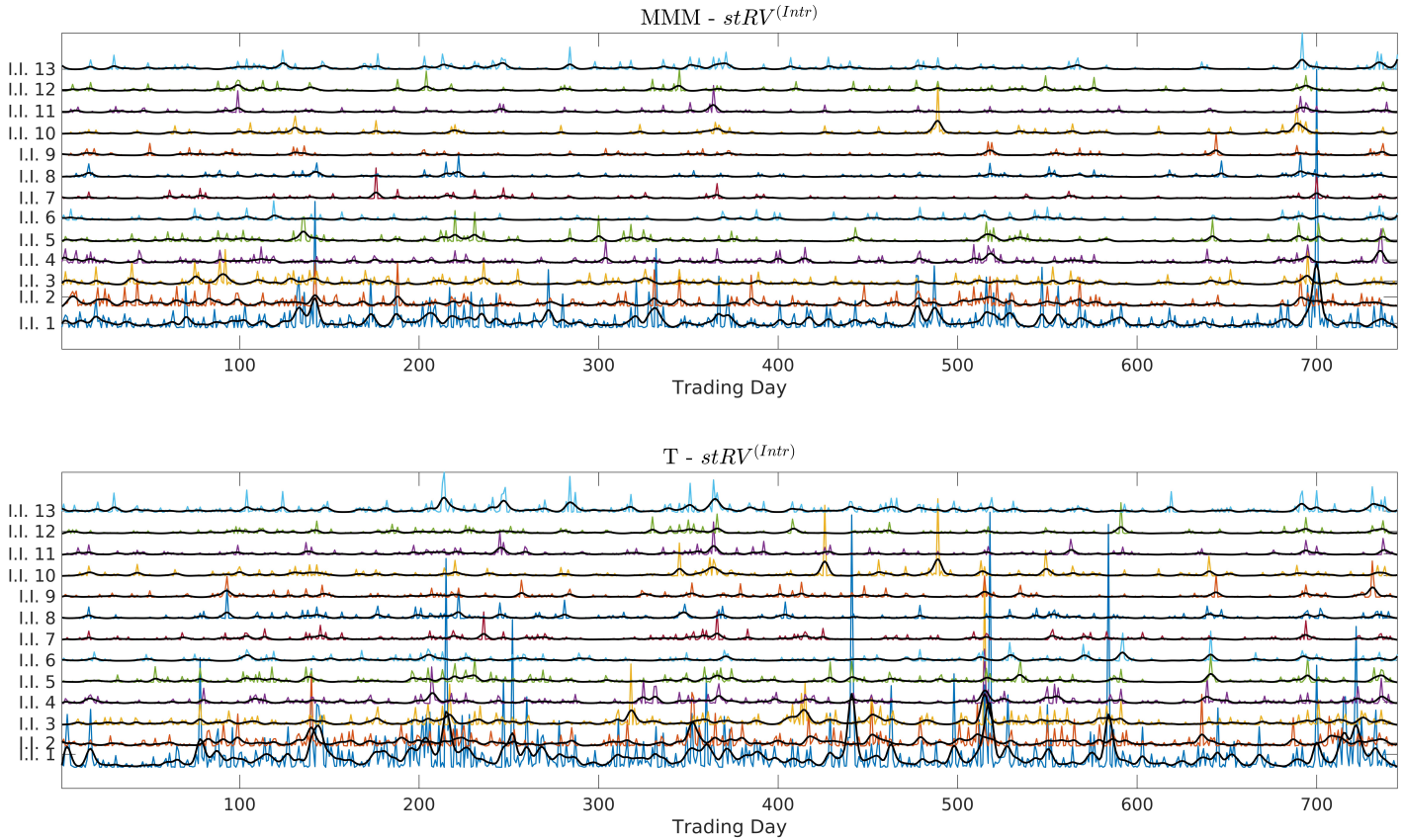

Figure C.10: This figure plots the volatility estimates of $stRV^{(Intr)}$ against the mean functions of the stocks 3M Co. (MMM) and AT&T Inc. (T) for each intraday interval over the days.

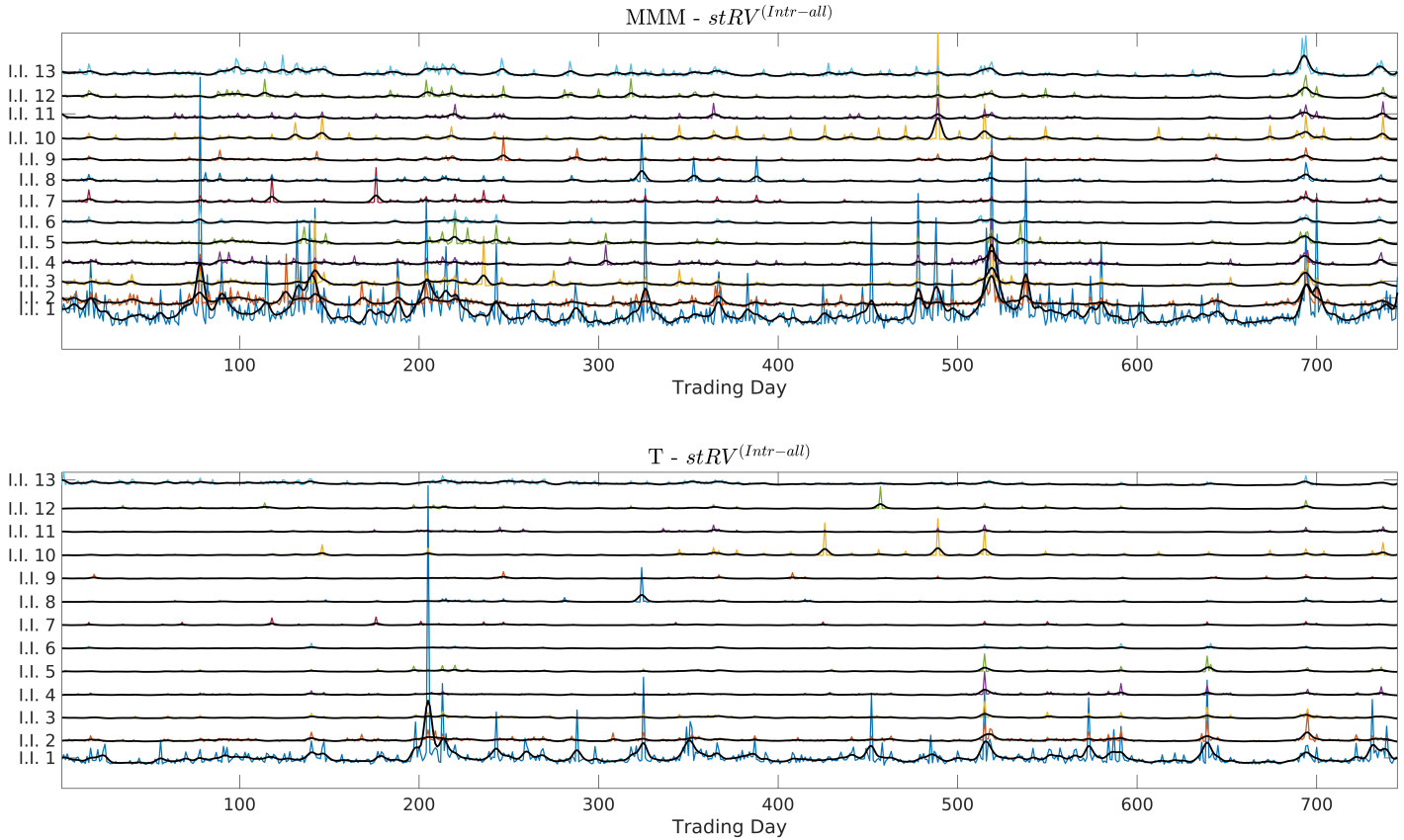

Figure C.11: This figure plots the volatility estimates of $stRV^{(Intr-all)}$ against the mean functions of the stocks 3M Co. (MMM) and AT&T Inc. (T) for each intraday interval over the days.

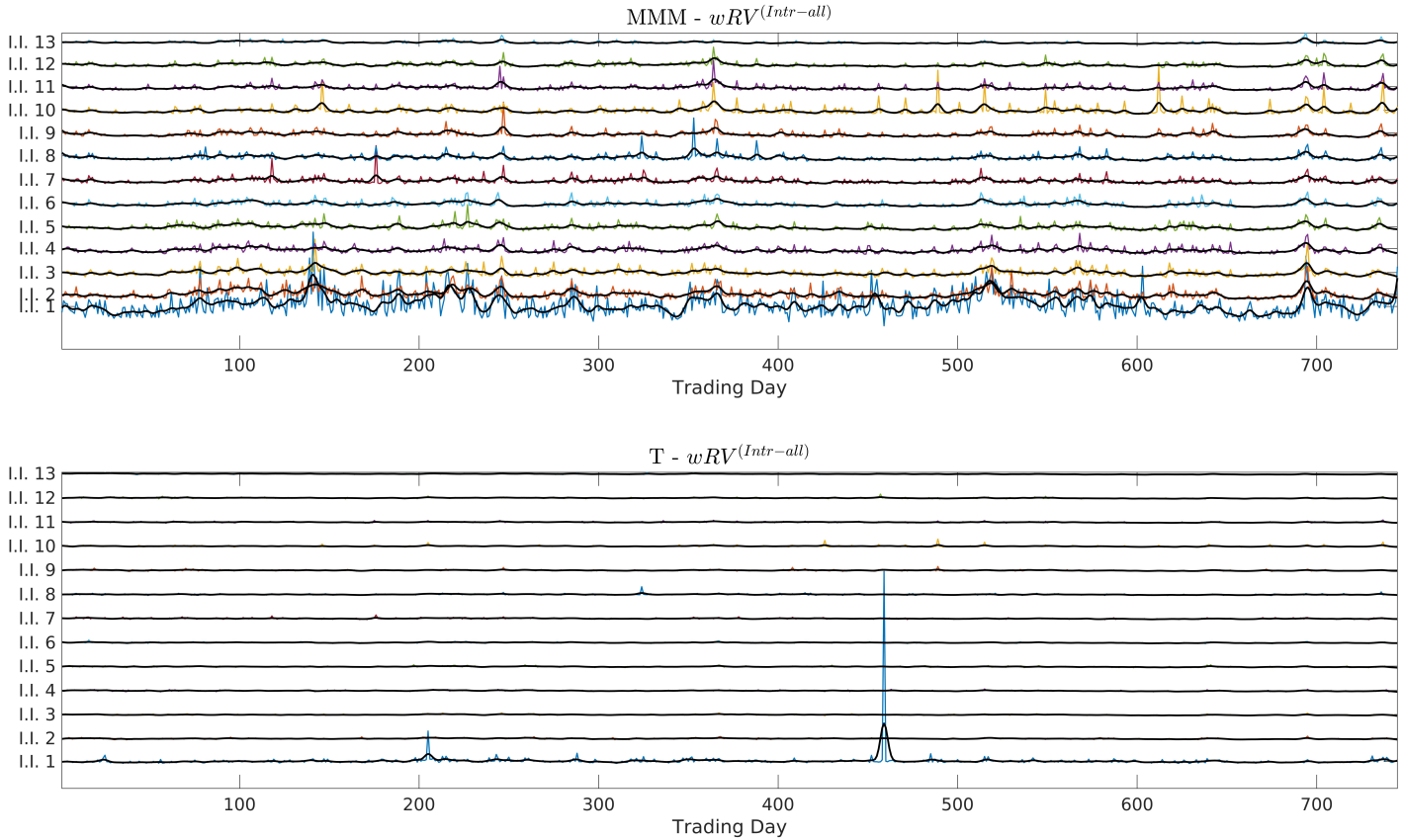

Figure C.12: This figure plots the volatility estimates of $wRV^{(Intr-all)}$ against the mean functions of the stocks 3M Co. (MMM) and AT&T Inc. (T) for each intraday interval over the days.

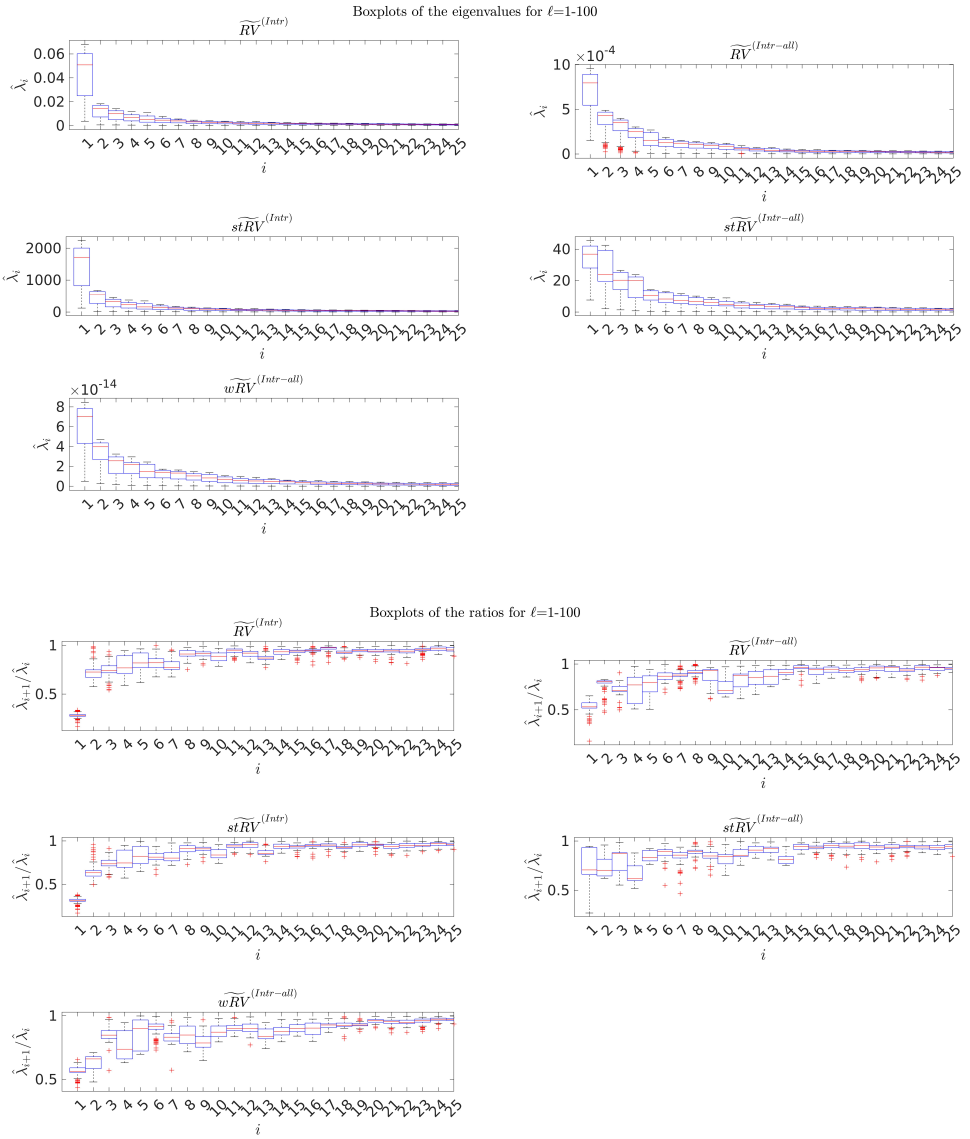

Figure C.13: These boxplots show the range of the estimated eigenvalues (first three row-panels) and the range of their ratios (last three row-panels) for the intraday RV-based volatility estimates for $\ell = 1 - 100$.

Appendix D

Appendix for Chapter 4

This appendix refers to Chapter 4. In particular, we derive the local smoothers up to the third polynomial degree in D.1. We provide the derivations of MSE and CV in D.2 and D.3, respectively. Besides, the derivation of specific properties that used in the derivation of our proposed LPR-based estimator are given in D.4.

D.1 Derivation of the Local Polynomial Estimators

This appendix section gives the derivation of the Nadaraya-Watson (appendix D.1.1), the local linear (appendix D.1.2), the local quadratic (appendix D.1.3) and the local cubic (appendix D.1.4) estimators. Unless explicitly stated otherwise, all quantities in this appendix section refer to a particular point x_0, and thus further clarification is omitted. So, for instance, by s_0 we mean $s_0(x_0)$, by w_{11} we mean $w_{11}(x_0)$ and so on.

D.1.1 Nadaraya-Watson Estimator

This appendix section provides the derivation of the local constant estimator (i.e. for the case $D = 0$). By Equation (4.3), the coefficients of the regression function at point x_0, are given by:

$$\widehat{\beta}_{x_0} = (X_{x_0}^T W_{x_0} X_{x_0})^{-1} X_{x_0}^T W_{x_0} Y$$

which can be separated into two terms, as follows.

For $X_{x_0}^T W_{x_0} Y$:

$$\begin{pmatrix} 1 & 1 & \dots & 1 \end{pmatrix} \times \begin{pmatrix} K_h(x_1 - x_0) & 0 & \dots & 0 \\ 0 & K_h(x_2 - x_0) & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & K_h(x_n - x_0) \end{pmatrix} \times \begin{pmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_n \end{pmatrix} = \sum_{i=1}^{n} w_{ii} Y_i$$

where $w_{ii} = K_h(x_i - x_0)$.

For $X_{x_0}^T W_{x_0} X_{x_0}$:

$$\begin{pmatrix} 1 & 1 & \dots & 1 \end{pmatrix} \times \begin{pmatrix} w_{11} & 0 & \dots & 0 \\ 0 & w_{22} & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & w_{nn} \end{pmatrix} \times \begin{pmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{pmatrix} = \sum_{i=1}^{n} w_{ii} = s_0.$$

Therefore, the inverse form of $X_{x_0}^T W_{x_0} X_{x_0}$ is given by:

$$(X_{x_0}^T W_{x_0} X_{x_0})^{-1} = \frac{1}{s_0}.$$

So for the understudy point x_0, the Nadaraya-Watson estimator is given by:

$$\widehat{\beta}_0 = (X_{x_0}^T W_{x_0} X_{x_0})^{-1} X_{x_0}^T W_{x_0} Y = \frac{\sum_{i=1}^{n} w_{ii} Y_i}{\sum_{i=1}^{n} w_{ii}}.$$

Global Smoothing Matrix for the Nadaraya-Watson Estimator

The local smoothing matrix, at a particular point x_0, has the following form:

$$S(x_0; h) = X_{x_0} (X_{x_0}^T W_{x_0} X_{x_0})^{-1} X_{x_0}^T W_{x_0}.$$

Therefore, by the derivations above, we conclude that the local smoothing matrix $S(x_0; h)$ can also be written as:

$$\begin{pmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{pmatrix} \times \left(\frac{1}{s_0}\right) \times \begin{pmatrix} w_{11} & w_{22} & \dots & w_{nn} \end{pmatrix} = \begin{pmatrix} \frac{w_{11}}{s_0} & \frac{w_{22}}{s_0} & \dots & \frac{w_{nn}}{s_0} \\ \frac{w_{11}}{s_0} & \frac{w_{22}}{s_0} & \dots & \frac{w_{nn}}{s_0} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{w_{11}}{s_0} & \frac{w_{22}}{s_0} & \dots & \frac{w_{nn}}{s_0} \end{pmatrix}.$$

For n random specified points, we conclude with the following global smoothing matrix:

$$\begin{pmatrix} \frac{w_{11}(x_1)}{s_0(x_1)} & \frac{w_{22}(x_1)}{s_0(x_1)} & \dots & \frac{w_{nn}(x_1)}{s_0(x_1)} \\ \frac{w_{11}(x_2)}{s_0(x_2)} & \frac{w_{22}(x_2)}{s_0(x_2)} & \dots & \frac{w_{nn}(x_2)}{s_0(x_2)} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{w_{11}(x_n)}{s_0(x_n)} & \frac{w_{22}(x_n)}{s_0(x_n)} & \dots & \frac{w_{nn}(x_n)}{s_0(x_n)} \end{pmatrix}.$$

The argument x_i in the parenthesis indicates that the corresponding value was obtained for the i-th (specified) point where we choose to evaluate the observations at. In the above derivation, the number of specified points which has been selected is equal to the number of observations. However, this is not a limitation as regards the number of points we can choose. So, a different number (higher or lower) of specified points can be selected, depending on how frequently one wants to approximate the regression function.

D.1.2 Local Linear Estimator

This appendix section provides the derivation of the local linear estimator (i.e. for the case $D = 1$). By Equation (4.3), the coefficients of the regression function at point x_0, are given by:

$$\widehat{\beta}_{x_0} = (X_{x_0}^T W_{x_0} X_{x_0})^{-1} X_{x_0}^T W_{x_0} Y$$

which can be separated into two terms, as follows.

For $X_{x_0}^T W_{x_0} Y$:

$$\begin{pmatrix} 1 & 1 & \dots & 1 \\ (x_1 - x_0) & (x_2 - x_0) & \dots & (x_n - x_0) \end{pmatrix} \times \begin{pmatrix} K_h(x_1 - x_0) & 0 & \dots & 0 \\ 0 & K_h(x_2 - x_0) & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & K_h(x_n - x_0) \end{pmatrix} \times \begin{pmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_n \end{pmatrix}$$

$$= \begin{pmatrix} \sum_{i=1}^n w_{ii} Y_i \\ \sum_{i=1}^n d_i w_{ii} Y_i \end{pmatrix}$$

where $w_{ii} = K_h(x_i - x_0)$ and $d_i^k = (x_i - x_0)^k$.

For $X_{x_0}^T W_{x_0} X_{x_0}$:

$$\begin{pmatrix} 1 & 1 & \dots & 1 \\ d_1 & d_2 & \dots & d_n \end{pmatrix} \times \begin{pmatrix} w_{11} & 0 & \dots & 0 \\ 0 & w_{22} & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & w_{nn} \end{pmatrix} \times \begin{pmatrix} 1 & d_1 \\ 1 & d_2 \\ \vdots & \vdots \\ 1 & d_n \end{pmatrix}$$

$$= \begin{pmatrix} \sum_{i=1}^n w_{ii} & \sum_{i=1}^n w_{ii} d_i \\ \sum_{i=1}^n w_{ii} d_i & \sum_{i=1}^n w_{ii} d_i^2 \end{pmatrix} = \begin{pmatrix} s_0 & s_1 \\ s_1 & s_2 \end{pmatrix}.$$

Hence, the form of $(X_{x_0}^T W_{x_0} X_{x_0})^{-1} X_{x_0}^T W_{x_0} Y$ can be expressed as:

$$(X_{x_0}^T W_{x_0} X_{x_0})^{-1} X_{x_0}^T W_{x_0} Y = \frac{1}{s_0 s_2 - s_1^2} \times \begin{pmatrix} s_2 & -s_1 \\ -s_1 & s_0 \end{pmatrix} \times \begin{pmatrix} \sum_{i=1}^n w_{ii} Y_i \\ \sum_{i=1}^n d_i w_{ii} Y_i \end{pmatrix}$$

$$= \frac{1}{s_0 s_2 - s_1^2} \times \begin{pmatrix} \sum_{i=1}^n (s_2 - s_1 d_i) w_{ii} Y_i \\ \\ \sum_{i=1}^n (-s_1 + s_0 d_i) w_{ii} Y_i \end{pmatrix}.$$

For the understudy point x_0, we end up with the following coefficient estimators:

$$\widehat{\beta}_0 = \sum_{i=1}^{n} \frac{(s_2 - s_1 d_i) w_{ii} Y_i}{s_0 s_2 - s_1^2},$$
$$\widehat{\beta}_1 = \sum_{i=1}^{n} \frac{(-s_1 + s_0 d_i) w_{ii} Y_i}{s_0 s_2 - s_1^2}.$$

Global Smoothing Matrix for the Local Linear Estimator

The local smoothing matrix, at a particular point x_0, has the following form:

$$S(x_0; h) = X_{x_0}(X_{x_0}^T W_{x_0} X_{x_0})^{-1} X_{x_0}^T W_{x_0}.$$

Hence, by the derivations above, we conclude that the local smoothing matrix $S(x_0; h)$ can also be written as:

$$\begin{pmatrix} 1 & d_1 \\ 1 & d_2 \\ \vdots & \vdots \\ 1 & d_n \end{pmatrix} \times \begin{pmatrix} \frac{s_2}{s_0 s_2 - s_1^2} & \frac{-s_1}{s_0 s_2 - s_1^2} \\ \frac{-s_1}{s_0 s_2 - s_1^2} & \frac{s_0}{s_0 s_2 - s_1^2} \end{pmatrix} \times \begin{pmatrix} w_{11} & w_{22} & \dots & w_{nn} \\ d_1 w_1 & d_2 w_{22} & \dots & d_n w_{nn} \end{pmatrix}$$

$$= \begin{pmatrix} 1 & d_1 \\ 1 & d_2 \\ \vdots & \vdots \\ 1 & d_n \end{pmatrix} \times \begin{pmatrix} \frac{(s_2 - s_1 d_1) w_{11}}{s_0 s_2 - s_1^2} & \frac{(s_2 - s_1 d_2) w_{22}}{s_0 s_2 - s_1^2} & \dots & \frac{(s_2 - s_1 d_n) w_{nn}}{s_0 s_2 - s_1^2} \\ \frac{(-s_1 + s_0 d_1) w_{11}}{s_0 s_2 - s_1^2} & \frac{(-s_1 + s_0 d_2) w_{22}}{s_0 s_2 - s_1^2} & \dots & \frac{(-s_1 + s_0 d_n) w_{nn}}{s_0 s_2 - s_1^2} \end{pmatrix}$$

$$= \begin{pmatrix} \frac{[(s_2-s_1d_1)+(-s_1+s_0d_1)d_1]w_{11}}{s_0s_2-s_1^2} & \frac{[(s_2-s_1d_2)+(-s_1+s_0d_2)d_1]w_{22}}{s_0s_2-s_1^2} & \cdots & \frac{[(s_2-s_1d_n)+(-s_1+s_0d_n)d_1]w_{nn}}{s_0s_2-s_1^2} \\ \frac{[(s_2-s_1d_1)+(-s_1+s_0d_1)d_2]w_{11}}{s_0s_2-s_1^2} & \frac{[(s_2-s_1d_2)+(-s_1+s_0d_2)d_2]w_{22}}{s_0s_2-s_1^2} & \cdots & \frac{[(s_2-s_1d_n)+(-s_1+s_0d_n)d_2]w_{nn}}{s_0s_2-s_1^2} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{[(s_2-s_1d_1)+(-s_1+s_0d_1)d_n]w_{11}}{s_0s_2-s_1^2} & \frac{[(s_2-s_1d_2)+(-s_1+s_0d_2)d_n]w_{22}}{s_0s_2-s_1^2} & \cdots & \frac{[(s_2-s_1d_n)+(-s_1+s_0d_n)d_n]w_{nn}}{s_0s_2-s_1^2} \end{pmatrix}.$$

All the above quantities refer to a particular point x_0; thus, further clarification is omitted. So, for instance, by s_0 we mean $s_0(x_0)$, by w_{11} we mean $w_{11}(x_0)$ and so on.

For n random specified points, the global smoothing matrix is given by:

$$\begin{pmatrix} \frac{[\alpha_1(x_1)+\beta_1(x_1)d_1(x_1)]w_{11}(x_1)}{\xi(x_1)} & \frac{[\alpha_2(x_1)+\beta_2(x_1)d_1(x_1)]w_{22}(x_1)}{\xi(x_1)} & \cdots & \frac{[\alpha_n(x_1)+\beta_n(x_1)d_1(x_1)]w_{nn}(x_1)}{\xi(x_1)} \\ \frac{[\alpha_1(x_2)+\beta_1(x_2)d_2(x_2)]w_{11}(x_2)}{\xi(x_2)} & \frac{[\alpha_2(x_2)+\beta_2(x_2)d_2(x_2)]w_{22}(x_2)}{\xi(x_2)} & \cdots & \frac{[\alpha_n(x_2)+\beta_n(x_2)d_2(x_2)]w_{nn}(x_2)}{\xi(x_2)} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{[\alpha_1(x_n)+\beta_1(x_n)d_n(x_n)]w_{11}(x_n)}{\xi(x_n)} & \frac{[\alpha_2(x_n)+\beta_2(x_n)d_n(x_n)]w_{22}(x_n)}{\xi(x_n)} & \cdots & \frac{[\alpha_n(x_n)+\beta_n(x_n)d_n(x_n)]w_{nn}(x_n)}{\xi(x_n)} \end{pmatrix}$$

where $\xi(x_i) = s_0(x_i)s_2(x_i) - s_1^2(x_i)$, $\alpha_j(x_i) = s_2(x_i) - s_1(x_i)d_j(x_i)$ and $\beta_j(x_i) = -s_1(x_i) + s_0(x_i)d_j(x_i)$ for $1 \leq i, j \leq n$. Additionally, the argument x_i denotes the i-th specified point where we choose to evaluate the observations at. In the above derivation, the number of specified observations which has been selected is equal to the number of the observations. However, this is not a limitation as regards the number of points we can choose. So a different number (higher or lower) of specified points can be selected, depending on how frequently one wants to approximate the regression function. Then, the global smoothing matrix will be transformed analogously.

Furthermore, in the case we choose to evaluate our regression function at the same points as those given from the observations, then the global smoothing matrix is further simplified as follows:

$$\begin{pmatrix} \frac{s_2(x_1)w_{11}(x_1)}{\xi(x_1)} & \frac{[\alpha_2(x_1)+\beta_2(x_1)d_1(x_1)]w_{22}(x_1)}{\xi(x_1)} & \dots & \frac{[\alpha_n(x_1)+\beta_n(x_1)d_1(x_1)]w_{nn}(x_1)}{\xi(x_1)} \\ \frac{[\alpha_1(x_2)+\beta_1(x_2)d_2(x_2)]w_{11}(x_2)}{\xi(x_2)} & \frac{s_2(x_2)w_{22}(x_2)}{\xi(x_2)} & \dots & \frac{[\alpha_n(x_2)+\beta_n(x_2)d_2(x_2)]w_{nn}(x_2)}{\xi(x_2)} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{[\alpha_1(x_n)+\beta_1(x_n)d_n(x_n)]w_{11}(x_n)}{\xi(x_n)} & \frac{[\alpha_2(x_n)+\beta_2(x_n)d_n(x_n)]w_{22}(x_n)}{\xi(x_n)} & \dots & \frac{s_2(x_n)w_n(x_n)}{\xi(x_n)} \end{pmatrix}.$$

D.1.3 Local Quadratic Estimator

This appendix section provides the derivation of the local quadratic estimator (i.e. for the case $D = 2$). By Equation (4.3), the coefficients of the regression function at point x_0, are given by:

$$\widehat{\beta}_{x_0} = (X_{x_0}^T W_{x_0} X_{x_0})^{-1} X_{x_0}^T W_{x_0} Y$$

which can be separated into two terms, as follows.

For $X_{x_0}^T W_{x_0} Y$:

$$\begin{pmatrix} 1 & 1 & \dots & 1 \\ (x_1 - x_0) & (x_2 - x_0) & \dots & (x_n - x_0) \\ (x_1 - x_0)^2 & (x_2 - x_0)^2 & \dots & (x_n - x_0)^2 \end{pmatrix} \times \begin{pmatrix} K_h(x_1 - x_0) & 0 & \dots & 0 \\ 0 & K_h(x_2 - x_0) & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & K_h(x_n - x_0) \end{pmatrix} \times \begin{pmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_n \end{pmatrix}$$

$$= \begin{pmatrix} \sum_{i=1}^{n} w_{ii} Y_i \\ \sum_{i=1}^{n} d_i w_{ii} Y_i \\ \sum_{i=1}^{n} d_i^2 w_{ii} Y_i \end{pmatrix}$$

where $w_{ii} = K_h(x_i - x_0)$ and $d_i^k = (x_i - x_0)^k$.

For $X_{x_0}^T W_{x_0} X_{x_0}$:

$$\begin{pmatrix} 1 & 1 & \dots & 1 \\ d_1 & d_2 & \dots & d_n \\ d_1^2 & d_2^2 & \dots & d_n^2 \end{pmatrix} \times \begin{pmatrix} w_{11} & 0 & \dots & 0 \\ 0 & w_{22} & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & w_{nn} \end{pmatrix} \times \begin{pmatrix} 1 & d_1 & d_1^2 \\ 1 & d_2 & d_2^2 \\ \vdots & \vdots & \vdots \\ 1 & d_n & d_n^2 \end{pmatrix}$$

$$= \begin{pmatrix} \sum_{i=1}^{n} w_{ii} & \sum_{i=1}^{n} w_{ii} d_i & \sum_{i=1}^{n} w_{ii} d_i^2 \\ \sum_{i=1}^{n} w_{ii} d_i & \sum_{i=1}^{n} w_{ii} d_i^2 & \sum_{i=1}^{n} w_{ii} d_i^3 \\ \sum_{i=1}^{n} w_{ii} d_i^2 & \sum_{i=1}^{n} w_{ii} d_i^3 & \sum_{i=1}^{n} w_{ii} d_i^4 \end{pmatrix} = \begin{pmatrix} s_0 & s_1 & s_2 \\ s_1 & s_2 & s_3 \\ s_2 & s_3 & s_4 \end{pmatrix}.$$

As regards the calculation of $(X_{x_0}^T W_{x_0} X_{x_0})^{-1}$:

- The determinant of $A = \begin{pmatrix} s_0 & s_1 & s_2 \\ s_1 & s_2 & s_3 \\ s_2 & s_3 & s_4 \end{pmatrix}$ is :

$$\begin{aligned} det(A) &= s_0 s_2 s_4 + s_1 s_2 s_3 + s_1 s_2 s_3 - s_2^3 - s_0 s_3^2 - s_1^2 s_4 \\ &= s_0(s_2 s_4 - s_3^2) + s_2(2 s_1 s_3 - s_2^2) - s_1^2 s_4. \end{aligned}$$

- Therefore, the matrix of co-factors of A is:

$$\begin{pmatrix} s_2 s_4 - s_3^2 & -(s_1 s_4 - s_2 s_3) & s_1 s_3 - s_2^2 \\ -(s_1 s_4 - s_2 s_3) & s_0 s_4 - s_2^2 & -(s_0 s_3 - s_1 s_2) \\ s_1 s_3 - s_2^2 & -(s_0 s_3 - s_1 s_2) & s_0 s_2 - s_1^2 \end{pmatrix}.$$

- The transpose matrix remains the same as the matrix of co-factors:

$$\begin{pmatrix} s_2s_4-s_3^2 & s_2s_3-s_1s_4 & s_1s_3-s_2^2 \\ s_2s_3-s_1s_4 & s_0s_4-s_2^2 & s_1s_2-s_0s_3 \\ s_1s_3-s_2^2 & s_1s_2-s_0s_3 & s_0s_2-s_1^2 \end{pmatrix}.$$

Hence, the inverse form of $X_{x_0}^T W_{x_0} X_{x_0}$ is given by:

$$(X_{x_0}^T W_{x_0} X_{x_0})^{-1} = \frac{1}{s_0(s_2s_4-s_3^2)+s_2(2s_1s_3-s_2^2)-s_1^2s_4} \times \begin{pmatrix} s_2s_4-s_3^2 & s_2s_3-s_1s_4 & s_1s_3-s_2^2 \\ s_2s_3-s_1s_4 & s_0s_4-s_2^2 & s_1s_2-s_0s_3 \\ s_1s_3-s_2^2 & s_1s_2-s_0s_3 & s_0s_2-s_1^2 \end{pmatrix}.$$

So, the form of $(X_{x_0}^T W_{x_0} X_{x_0})^{-1} X_{x_0}^T W_{x_0} Y$ can be expressed as:
$(X_{x_0}^T W_{x_0} X_{x_0})^{-1} X_{x_0}^T W_{x_0} Y =$

$$= \frac{1}{s_0(s_2s_4-s_3^2)+s_2(2s_1s_3-s_2^2)-s_1^2s_4} \times \begin{pmatrix} s_2s_4-s_3^2 & s_2s_3-s_1s_4 & s_1s_3-s_2^2 \\ s_2s_3-s_1s_4 & s_0s_4-s_2^2 & s_1s_2-s_0s_3 \\ s_1s_3-s_2^2 & s_1s_2-s_0s_3 & s_0s_2-s_1^2 \end{pmatrix} \times \begin{pmatrix} \sum_{i=1}^n w_{ii}Y_i \\ \sum_{i=1}^n d_i w_{ii}Y_i \\ \sum_{i=1}^n d_i^2 w_{ii}Y_i \end{pmatrix}$$

which gives: $(X_{x_0}^T W_{x_0} X_{x_0})^{-1} X_{x_0}^T W_{x_0} Y = \frac{1}{s_0(s_2s_4-s_3^2)+s_2(2s_1s_3-s_2^2)-s_1^2s_4} \times$

$$\times \begin{pmatrix} (s_2s_4-s_3^2)\sum_{i=1}^n w_{ii}Y_i + (s_2s_3-s_1s_4)\sum_{i=1}^n d_i w_{ii}Y_i + (s_1s_3-s_2^2)\sum_{i=1}^n d_i^2 w_{ii}Y_i \\ (s_2s_3-s_1s_4)\sum_{i=1}^n w_{ii}Y_i + (s_0s_4-s_2^2)\sum_{i=1}^n d_i w_{ii}Y_i + (s_1s_2-s_0s_3)\sum_{i=1}^n d_i^2 w_{ii}Y_i \\ (s_1s_3-s_2^2)\sum_{i=1}^n w_{ii}Y_i + (s_1s_2-s_0s_3)\sum_{i=1}^n d_i w_{ii}Y_i + (s_0s_2-s_1^2)\sum_{i=1}^n d_i^2 w_{ii}Y_i \end{pmatrix}.$$

For the understudy point x_0, we end up with the following coefficient estimators:

$$\begin{aligned} \widehat{\beta}_0 &= \sum_{i=1}^n \frac{(s_2s_4-s_3^2+(s_2s_3-s_1s_4)d_i+(s_1s_3-s_2^2)d_i^2)w_{ii}Y_i}{s_0(s_2s_4-s_3^2)+s_2(2s_1s_3-s_2^2)-s_1^2s_4}, \\ \widehat{\beta}_1 &= \sum_{i=1}^n \frac{(s_2s_3-s_1s_4+(s_0s_4-s_2^2)d_i+(s_1s_2-s_0s_3)d_i^2)w_{ii}Y_i}{s_0(s_2s_4-s_3^2)+s_2(2s_1s_3-s_2^2)-s_1^2s_4}, \\ \widehat{\beta}_2 &= \sum_{i=1}^n \frac{(s_1s_3-s_2^2+(s_1s_2-s_0s_3)d_i+(s_0s_2-s_1^2)d_i^2)w_{ii}Y_i}{s_0(s_2s_4-s_3^2)+s_2(2s_1s_3-s_2^2)-s_1^2s_4}. \end{aligned}$$

Global Smoothing Matrix for the Local Quadratic Estimator

The local smoothing matrix, at a particular point x_0, has the following form:

$$S(x_0; h) = X_{x_0}(X_{x_0}^T W_{x_0} X_{x_0})^{-1} X_{x_0}^T W_{x_0}.$$

Hence, by the derivations above, we conclude that the local smoothing matrix $S(x_0; h)$ can also be written as:

$$\begin{pmatrix} 1 & d_1 & d_1^2 \\ 1 & d_2 & d_2^2 \\ \vdots & \vdots & \vdots \\ 1 & d_n & d_n^2 \end{pmatrix} \times \begin{pmatrix} \frac{s_2 s_4 - s_3^2}{\xi} & \frac{s_2 s_3 - s_1 s_4}{\xi} & \frac{s_1 s_3 - s_2^2}{\xi} \\ \frac{s_2 s_3 - s_1 s_4}{\xi} & \frac{s_0 s_4 - s_2^2}{\xi} & \frac{s_1 s_2 - s_0 s_3}{\xi} \\ \frac{s_1 s_3 - s_2^2}{\xi} & \frac{s_1 s_2 - s_0 s_3}{\xi} & \frac{s_0 s_2 - s_1^2}{\xi} \end{pmatrix} \times \begin{pmatrix} w_{11} & w_{22} & \dots & w_{nn} \\ d_1 w_1 & d_2 w_{22} & \dots & d_n w_{nn} \\ d_1^2 w_{11} & d_2^2 w_{22} & \dots & d_n^2 w_{nn} \end{pmatrix} =$$

$$= \begin{pmatrix} \frac{[(\alpha+\beta d_1+\gamma d_1^2)+(\beta+\delta d_1+\epsilon d_1^2)d_1+(\gamma+\epsilon d_1+\zeta d_1^2)d_1^2]w_{11}}{\xi} & \frac{[(\alpha+\beta d_2+\gamma d_2^2)+(\beta+\delta d_2+\epsilon d_2^2)d_1+(\gamma+\epsilon d_2+\zeta d_2^2)d_1^2]w_{22}}{\xi} & \cdots & \frac{[(\alpha+\beta d_n+\gamma d_n^2)+(\beta+\delta d_n+\epsilon d_n^2)d_1+(\gamma+\epsilon d_n+\zeta d_n^2)d_1^2]w_{nn}}{\xi} \\ \frac{[(\alpha+\beta d_1+\gamma d_1^2)+(\beta+\delta d_1+\epsilon d_1^2)d_2+(\gamma+\epsilon d_1+\zeta d_1^2)d_2^2]w_{11}}{\xi} & \frac{[(\alpha+\beta d_2+\gamma d_2^2)+(\beta+\delta d_2+\epsilon d_2^2)d_2+(\gamma+\epsilon d_2+\zeta d_2^2)d_2^2]w_{22}}{\xi} & \cdots & \frac{[(\alpha+\beta d_n+\gamma d_n^2)+(\beta+\delta d_n+\epsilon d_n^2)d_2+(\gamma+\epsilon d_n+\zeta d_n^2)d_2^2]w_{nn}}{\xi} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{[(\alpha+\beta d_1+\gamma d_1^2)+(\beta+\delta d_1+\epsilon d_1^2)d_n+(\gamma+\epsilon d_1+\zeta d_1^2)d_n^2]w_{11}}{\xi} & \frac{[(\alpha+\beta d_2+\gamma d_2^2)+(\beta+\delta d_2+\epsilon d_2^2)d_n+(\gamma+\epsilon d_2+\zeta d_2^2)d_n^2]w_{22}}{\xi} & \cdots & \frac{[(\alpha+\beta d_n+\gamma d_n^2)+(\beta+\delta d_n+\epsilon d_n^2)d_n+(\gamma+\epsilon d_n+\zeta d_n^2)d_n^2]w_{nn}}{\xi} \end{pmatrix}$$

where $\xi = s_0(s_2s_4 - s_3^2) + s_2(2s_1s_3 - s_2^2) - s_1^2s_4$, $\alpha = s_2s_4 - s_3^2$, $\beta = s_2s_3 - s_1s_4$, $\gamma = s_1s_3 - s_2^2$, $\delta = s_0s_4 - s_2^2$, $\epsilon = s_1s_2 - s_0s_3$, $\zeta = s_0s_2 - s_1^2$. For n random specified points, we conclude with the following global smoothing matrix:

$$\begin{pmatrix} \frac{[\mu_1(x_1)+\nu_1(x_1)d_1(x_1)+\sigma_1(x_1)d_1^2(x_1)]w_{11}(x_1)}{\xi(x_1)} & \frac{[\mu_2(x_1)+\nu_2(x_1)d_1(x_1)+\sigma_2(x_1)d_1^2(x_1)]w_{22}(x_1)}{\xi(x_1)} & \cdots & \frac{[\mu_n(x_1)+\nu_n(x_1)d_1(x_1)+\sigma_n(x_1)d_1^2(x_1)]w_{nn}(x_1)}{\xi(x_1)} \\ \frac{[\mu_1(x_2)+\nu_1(x_2)d_2(x_2)+\sigma_1(x_2)d_2^2(x_2)]w_{11}(x_2)}{\xi(x_2)} & \frac{[\mu_2(x_2)+\nu_2(x_2)d_2(x_2)+\sigma_2(x_2)d_2^2(x_2)]w_{22}(x_2)}{\xi(x_2)} & \cdots & \frac{[\mu_n(x_2)+\nu_n(x_2)d_2(x_2)+\sigma_n(x_2)d_2^2(x_2)]w_{nn}(x_2)}{\xi(x_2)} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{[\mu_1(x_n)+\nu_1(x_n)d_n(x_n)+\sigma_1(x_n)d_n^2(x_n)]w_{nn}(x_n)}{\xi(x_n)} & \frac{[\mu_2(x_n)+\nu_2(x_n)d_n(x_n)+\sigma_2(x_n)d_n^2(x_n)]w_{22}(x_n)}{\xi(x_n)} & \cdots & \frac{[\mu_n(x_n)+\nu_n(x_n)d_n(x_n)+\sigma_n(x_n)d_n^2(x_n)]w_{nn}(x_n)}{\xi(x_n)} \end{pmatrix}$$

where the parameters mentioned above are defined as follow:

- $\mu_j(x_i) = \alpha(x_i) + \beta(x_i)d_j(x_i) + \gamma(x_i)d_j^2(x_i)$,
- $\nu_j(x_i) = \beta(x_i) + \delta(x_i)d_j(x_i) + \epsilon(x_i)d_j^2(x_i)$,
- $\sigma_j(x_i) = \gamma(x_i) + \epsilon(x_i)d_j(x_i) + \zeta(x_i)d_j^2(x_i)$,
- $\xi(x_i) = s_0(x_i)(s_2(x_i)s_4(x_i) - s_3^2(x_i)) + s_2(x_i)(2s_1(x_i)s_3(x_i) - s_2^2(x_i)) - s_1^2(x_i)s_4(x_i)$.

Additionally, the argument x_i denotes the i-th specified point. In the above derivation, the number of specified observations which has been selected is equal to the number of the observations. However, this is not a limitation as regards the number of points we can choose. So a different number (higher or lower) of specified points can be selected, depending on how frequently one wants to approximate the regression function. Then, the global smoothing matrix will be transformed analogously.

Furthermore, in the case we choose to evaluate our regression function at the same points as those given from the observations, then the global smoothing matrix is further simplified as follows:

$$\begin{pmatrix} \frac{\alpha(x_1)w_{11}(x_1)}{\xi(x_1)} & \frac{(\alpha(x_1)+\beta(x_1)d_2(x_1)+\gamma(x_1)d_2^2(x_1))w_{22}(x_1)}{\xi(x_1)} & \cdots & \frac{(\alpha(x_1)+\beta(x_1)d_n(x_1)+\gamma(x_1)d_n^2(x_1))w_{nn}(x_1)}{\xi(x_1)} \\ \frac{(\alpha(x_2)+\beta(x_2)d_1(x_2)+\gamma(x_2)d_1^2(x_2))w_{11}(x_2)}{\xi(x_2)} & \frac{\alpha(x_2)w_{22}(x_2)}{\xi(x_2)} & \cdots & \frac{(\alpha(x_2)+\beta(x_2)d_n(x_2)+\gamma(x_2)d_n^2(x_2))w_{nn}(x_2)}{\xi(x_2)} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{(\alpha(x_n)+\beta(x_n)d_1(x_n)+\gamma(x_n)d_1^2(x_n))w_{11}(x_n)}{\xi(x_n)} & \frac{(\alpha(x_n)+\beta(x_n)d_2(x_n)+\gamma(x_n)d_2^2(x_n))w_{22}(x_n)}{\xi(x_n)} & \cdots & \frac{\alpha(x_n)w_{nn}(x_n)}{\xi(x_n)} \end{pmatrix}.$$

D.1.4 Local Cubic Estimator

This appendix section provides the derivation of the local cubic estimator (i.e. for the case $D = 3$). By Equation (4.3), the coefficients of the regression function at point x_0, are given by:

$$\widehat{\beta}_{x_0} = (X_{x_0}^T W_{x_0} X_{x_0})^{-1} X_{x_0}^T W_{x_0} Y$$

which can be separated into two terms, as follows.

For $X_{x_0}^T W_{x_0} Y$:

$$\begin{pmatrix} 1 & 1 & \dots & 1 \\ (x_1 - x_0) & (x_2 - x_0) & \dots & (x_n - x_0) \\ (x_1 - x_0)^2 & (x_2 - x_0)^2 & \dots & (x_n - x_0)^2 \\ (x_1 - x_0)^3 & (x_2 - x_0)^3 & \dots & (x_n - x_0)^3 \end{pmatrix} \times \begin{pmatrix} K_h(x_1 - x_0) & 0 & \dots & 0 \\ 0 & K_h(x_2 - x_0) & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & K_h(x_n - x_0) \end{pmatrix} \times \begin{pmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_n \end{pmatrix}$$

$$= \begin{pmatrix} w_{11} & w_{22} & \dots & w_{nn} \\ d_1 w_{11} & d_2 w_{22} & \dots & d_n w_{nn} \\ d_1^2 w_{11} & d_2^2 w_{22} & \dots & d_n^2 w_{nn} \\ d_1^3 w_{11} & d_2^3 w_{22} & \dots & d_n^3 w_{nn} \end{pmatrix} \times \begin{pmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_n \end{pmatrix} = \begin{pmatrix} \sum_{i=1}^n w_{ii} Y_i \\ \sum_{i=1}^n d_i w_{ii} Y_i \\ \sum_{i=1}^n d_i^2 w_{ii} Y_i \\ \sum_{i=1}^n d_i^3 w_{ii} Y_i \end{pmatrix}$$

where $w_{ii} = K_h(x_i - x_0)$ and $d_i^q = (x_i - x_0)^q$ for $q = 1, ..., 3$.

For $X_{x_0}^T W_{x_0} X_{x_0}$:

$$\begin{pmatrix} 1 & 1 & \dots & 1 \\ d_1 & d_2 & \dots & d_n \\ d_1^2 & d_2^2 & \dots & d_n^2 \\ d_1^3 & d_2^3 & \dots & d_n^3 \end{pmatrix} \times \begin{pmatrix} w_{11} & 0 & \dots & 0 \\ 0 & w_{22} & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & w_{nn} \end{pmatrix} \times \begin{pmatrix} 1 & d_1 & d_1^2 & d_1^3 \\ 1 & d_2 & d_2^2 & d_2^3 \\ \vdots & \vdots & \vdots & \vdots \\ 1 & d_n & d_n^2 & d_n^3 \end{pmatrix}$$

$$= \begin{pmatrix} \sum_{i=1}^n w_{ii} & \sum_{i=1}^n w_{ii}d_i & \sum_{i=1}^n w_{ii}d_i^2 & \sum_{i=1}^n w_{ii}d_i^3 \\ \sum_{i=1}^n w_{ii}d_i & \sum_{i=1}^n w_{ii}d_i^2 & \sum_{i=1}^n w_{ii}d_i^3 & \sum_{i=1}^n w_{ii}d_i^4 \\ \sum_{i=1}^n w_{ii}d_i^2 & \sum_{i=1}^n w_{ii}d_i^3 & \sum_{i=1}^n w_{ii}d_i^4 & \sum_{i=1}^n w_{ii}d_i^5 \\ \sum_{i=1}^n w_{ii}d_i^3 & \sum_{i=1}^n w_{ii}d_i^4 & \sum_{i=1}^n w_{ii}d_i^5 & \sum_{i=1}^n w_{ii}d_i^6 \end{pmatrix} = \begin{pmatrix} s_0 & s_1 & s_2 & s_3 \\ s_1 & s_2 & s_3 & s_4 \\ s_2 & s_3 & s_4 & s_5 \\ s_3 & s_4 & s_5 & s_6 \end{pmatrix}.$$

As regards the calculation of $(X_{x_0}^T W_{x_0} X_{x_0})^{-1}$:

- We can see that the determinant of $A = \begin{pmatrix} s_0 & s_1 & s_2 & s_3 \\ s_1 & s_2 & s_3 & s_4 \\ s_2 & s_3 & s_4 & s_5 \\ s_3 & s_4 & s_5 & s_6 \end{pmatrix}$ is :

$$\begin{aligned} det(A) =& s_0s_2s_4s_6 + 2s_0s_3s_4s_5 + 2s_1s_3s_4^2 + s_1^2s_5^2 + 2s_1s_2s_3s_6 + 2s_2^2s_3s_5 + s_2^2s_4^2 + s_3^4 - \\ & - (s_0s_4^3 + s_0s_2s_5^2 + s_0s_3^2s_6 + s_1^2s_4s_6 + 2s_1s_3^2s_5 + 2s_1s_2s_4s_5 + s_2^3s_6 + 3s_2s_3^2s_4). \end{aligned}$$

- The elements $(row, column)$ of the matrix of co-factors have the following form:

$$(1,1): \begin{vmatrix} s_2 & s_3 & s_4 \\ s_3 & s_4 & s_5 \\ s_4 & s_5 & s_6 \end{vmatrix} = s_2s_4s_6 + 2s_3s_4s_5 - s_4^3 - s_2s_5^2 - s_3^2s_6,$$

$$(1,2): - \begin{vmatrix} s_1 & s_3 & s_4 \\ s_2 & s_4 & s_5 \\ s_3 & s_5 & s_6 \end{vmatrix} = s_1s_5^2 + s_3s_4^2 + s_2s_3s_6 - s_1s_4s_6 - s_2s_4s_5 - s_3^2s_5,$$

$$(1,3): \begin{vmatrix} s_1 & s_2 & s_4 \\ s_2 & s_3 & s_5 \\ s_3 & s_4 & s_6 \end{vmatrix} = s_1s_3s_6 + s_2s_3s_5 + s_2s_4^2 - s_3^2s_4 - s_1s_4s_5 - s_2^2s_6,$$

$$(1,4):-\begin{vmatrix} s_1 & s_2 & s_3 \\ s_2 & s_3 & s_4 \\ s_3 & s_4 & s_5 \end{vmatrix} = s_3^3 + s_1 s_4^2 + s_2^2 s_5 - s_1 s_3 s_5 - 2 s_2 s_3 s_4,$$

$$(2,1):-\begin{vmatrix} s_1 & s_2 & s_3 \\ s_3 & s_4 & s_5 \\ s_4 & s_5 & s_6 \end{vmatrix} = s_1 s_5^2 + s_3 s_4^2 + s_2 s_3 s_6 - s_1 s_4 s_6 - s_2 s_4 s_5 - s_3^2 s_5,$$

$$(2,2):\begin{vmatrix} s_0 & s_2 & s_3 \\ s_2 & s_4 & s_5 \\ s_3 & s_5 & s_6 \end{vmatrix} = s_0 s_4 s_6 + 2 s_2 s_3 s_5 - s_3^2 s_4 - s_0 s_5^2 - s_2^2 s_6,$$

$$(2,3):-\begin{vmatrix} s_0 & s_1 & s_3 \\ s_2 & s_3 & s_5 \\ s_3 & s_4 & s_6 \end{vmatrix} = s_3^3 + s_0 s_4 s_5 + s_1 s_2 s_6 - s_0 s_3 s_6 - s_1 s_3 s_5 - s_2 s_3 s_4,$$

$$(2,4):-\begin{vmatrix} s_0 & s_1 & s_2 \\ s_2 & s_3 & s_4 \\ s_3 & s_4 & s_5 \end{vmatrix} = s_0 s_3 s_5 + s_1 s_3 s_4 + s_2^2 s_4 - s_2 s_3^2 - s_0 s_4^2 - s_1 s_2 s_5,$$

$$(3,1):\begin{vmatrix} s_1 & s_2 & s_3 \\ s_2 & s_3 & s_4 \\ s_4 & s_5 & s_6 \end{vmatrix} = s_1 s_3 s_6 + s_2 s_3 s_5 + s_2 s_4^2 - s_3^2 s_4 - s_1 s_4 s_5 - s_2^2 s_6,$$

$$(3,2):-\begin{vmatrix} s_0 & s_2 & s_3 \\ s_1 & s_3 & s_4 \\ s_3 & s_5 & s_6 \end{vmatrix} = s_3^3 + s_0 s_4 s_5 + s_1 s_2 s_6 - s_0 s_3 s_6 - s_1 s_3 s_5 - s_2 s_3 s_4,$$

$$(3,3):\begin{vmatrix} s_0 & s_1 & s_3 \\ s_1 & s_2 & s_4 \\ s_3 & s_4 & s_6 \end{vmatrix} = s_0 s_2 s_6 + 2 s_1 s_3 s_4 - s_2 s_3^2 - s_0 s_4^2 - s_1^2 s_6,$$

$$(3,4): -\begin{vmatrix} s_0 & s_1 & s_2 \\ s_1 & s_2 & s_3 \\ s_3 & s_4 & s_5 \end{vmatrix} = s_2^2 s_3 + s_0 s_3 s_4 + s_1^2 s_5 - s_0 s_2 s_5 - s_1 s_3^2 - s_1 s_2 s_4,$$

$$(4,1): -\begin{vmatrix} s_1 & s_2 & s_3 \\ s_2 & s_3 & s_4 \\ s_3 & s_4 & s_5 \end{vmatrix} = s_3^3 + s_1 s_4^2 + s_2^2 s_5 - s_1 s_3 s_5 - 2 s_2 s_3 s_4,$$

$$(4,2): \begin{vmatrix} s_0 & s_2 & s_3 \\ s_1 & s_3 & s_4 \\ s_2 & s_4 & s_5 \end{vmatrix} = s_0 s_3 s_5 + s_1 s_3 s_4 + s_2^2 s_4 - s_2 s_3^2 - s_0 s_4^2 - s_1 s_2 s_5,$$

$$(4,3): -\begin{vmatrix} s_0 & s_1 & s_3 \\ s_1 & s_2 & s_4 \\ s_2 & s_3 & s_5 \end{vmatrix} = s_2^2 s_3 + s_0 s_3 s_4 + s_1^2 s_5 - s_0 s_2 s_5 - s_1 s_3^2 - s_1 s_2 s_4,$$

$$(4,4): \begin{vmatrix} s_0 & s_1 & s_3 \\ s_1 & s_2 & s_4 \\ s_2 & s_3 & s_5 \end{vmatrix} = s_0 s_2 s_4 + 2 s_1 s_2 s_3 - s_2^3 - s_0 s_3^2 - s_1^2 s_4.$$

Now, we set the following components:

$$\alpha = s_2 s_4 s_6 + 2 s_3 s_4 s_5 - s_4^3 - s_2 s_5^2 - s_3^2 s_6,$$

$$\beta = s_0 s_4 s_6 + 2 s_2 s_3 s_5 - s_3^2 s_4 - s_0 s_5^2 - s_2^2 s_6,$$

$$\gamma = s_0 s_2 s_6 + 2 s_1 s_3 s_4 - s_2 s_3^2 - s_0 s_4^2 - s_1^2 s_6,$$

$$\delta = s_0 s_2 s_4 + 2 s_1 s_2 s_3 - s_2^3 - s_0 s_3^2 - s_1^2 s_4,$$

$$\epsilon = s_1 s_5^2 + s_3 s_4^2 + s_2 s_3 s_6 - s_1 s_4 s_6 - s_2 s_4 s_5 - s_3^2 s_5,$$

$$\zeta = s_1 s_3 s_6 + s_2 s_3 s_5 + s_2 s_4^2 - s_3^2 s_4 - s_1 s_4 s_5 - s_2^2 s_6,$$

$$\eta = s_3^3 + s_1 s_4^2 + s_2^2 s_5 - s_1 s_3 s_5 - 2 s_2 s_3 s_4,$$

$$\theta = s_3^3 + s_0 s_4 s_5 + s_1 s_2 s_6 - s_0 s_3 s_6 - s_1 s_3 s_5 - s_2 s_3 s_4,$$

$$\kappa = s_0 s_3 s_5 + s_1 s_3 s_4 + s_2^2 s_4 - s_2 s_3^2 - s_0 s_4^2 - s_1 s_2 s_5,$$

$$\lambda = s_2^2 s_3 + s_0 s_3 s_4 + s_1^2 s_5 - s_0 s_2 s_5 - s_1 s_3^2 - s_1 s_2 s_4,$$

$$\xi = s_0 s_2 s_4 s_6 + 2 s_0 s_3 s_4 s_5 + 2 s_1 s_3 s_4^2 + s_1^2 s_5^2 + 2 s_1 s_2 s_3 s_6 + 2 s_2^2 s_3 s_5 + s_2^2 s_4^2 + s_3^4$$
$$- (s_0 s_4^3 + s_0 s_2 s_5^2 + s_0 s_3^2 s_6 + s_1^2 s_4 s_6 + 2 s_1 s_3^2 s_5 + 2 s_1 s_2 s_4 s_5 + s_2^3 s_6 + 3 s_2 s_3^2 s_4).$$

Then, we obtain the following form for the symmetric matrix $(X_{x_0}^T W_{x_0} X_{x_0})^{-1}$:

$$\frac{1}{\xi} \times \begin{pmatrix} \alpha & \epsilon & \zeta & \eta \\ \epsilon & \beta & \theta & \kappa \\ \zeta & \theta & \gamma & \lambda \\ \eta & \kappa & \lambda & \delta \end{pmatrix}.$$

Thus, $(X^T W X)^{-1} X^T W Y$ can be expressed as:

$$\frac{1}{\xi} \times \begin{pmatrix} \alpha & \epsilon & \zeta & \eta \\ \epsilon & \beta & \theta & \kappa \\ \zeta & \theta & \gamma & \lambda \\ \eta & \kappa & \lambda & \delta \end{pmatrix} \times \begin{pmatrix} \sum_{i=1}^n w_{ii} Y_i \\ \sum_{i=1}^n d_i w_{ii} Y_i \\ \sum_{i=1}^n d_i^2 w_{ii} Y_i \\ \sum_{i=1}^n d_i^3 w_{ii} Y_i \end{pmatrix} =$$

$$= \frac{1}{\xi} \times \begin{pmatrix} \alpha \sum_{i=1}^n w_{ii} Y_i + \epsilon \sum_{i=1}^n d_i w_{ii} Y_i + \zeta \sum_{i=1}^n d_i^2 w_{ii} Y_i + \eta \sum_{i=1}^n d_i^3 w_{ii} Y_i \\ \epsilon \sum_{i=1}^n w_{ii} Y_i + \beta \sum_{i=1}^n d_i w_{ii} Y_i + \theta \sum_{i=1}^n d_i^2 w_{ii} Y_i + \kappa \sum_{i=1}^n d_i^3 w_{ii} Y_i \\ \zeta \sum_{i=1}^n w_{ii} Y_i + \theta \sum_{i=1}^n d_i w_{ii} Y_i + \gamma \sum_{i=1}^n d_i^2 w_{ii} Y_i + \lambda \sum_{i=1}^n d_i^3 w_{ii} Y_i \\ \eta \sum_{i=1}^n w_{ii} Y_i + \kappa \sum_{i=1}^n d_i w_{ii} Y_i + \lambda \sum_{i=1}^n d_i^2 w_{ii} Y_i + \delta \sum_{i=1}^n d_i^3 w_{ii} Y_i \end{pmatrix}.$$

So for the understudy point x_0, we end up with the following coefficient estimators:

$$\widehat{\beta}_1 = \sum_{i=1}^n \frac{(\alpha + \epsilon d_i + \zeta d_i^2 + \eta d_i^3) w_{ii} Y_i}{\xi},$$

$$\widehat{\beta}_2 = \sum_{i=1}^{n} \frac{(\epsilon + \beta d_i + \theta d_i^2 + \kappa d_i^3) w_{ii} Y_i}{\xi},$$
$$\widehat{\beta}_3 = \sum_{i=1}^{n} \frac{(\zeta + \theta d_i + \gamma d_i^2 + \lambda d_i^3) w_{ii} Y_i}{\xi},$$
$$\widehat{\beta}_4 = \sum_{i=1}^{n} \frac{(\eta + \kappa d_i + \lambda d_i^2 + \delta d_i^3) w_{ii} Y_i}{\xi}.$$

Global Smoothing Matrix for the Local Cubic Estimator

The local smoothing matrix, at a particular point x_0, has the following form:

$$S(x_0; h) = X_{x_0}(X_{x_0}^T W_{x_0} X_{x_0})^{-1} X_{x_0}^T W_{x_0}.$$

Hence, by the derivations above, we conclude that this local smoothing matrix can also be written as:

$$\begin{pmatrix} 1 & d_1 & d_1^2 & d_1^3 \\ 1 & d_2 & d_2^2 & d_2^3 \\ \vdots & \vdots & \vdots & \vdots \\ 1 & d_n & d_n^2 & d_n^3 \end{pmatrix} \times \frac{1}{\xi} \begin{pmatrix} \alpha & \epsilon & \zeta & \eta \\ \epsilon & \beta & \theta & \kappa \\ \zeta & \theta & \gamma & \lambda \\ \eta & \kappa & \lambda & \delta \end{pmatrix} \times \begin{pmatrix} w_{11} & w_{22} & \dots & w_{nn} \\ d_1 w_1 & d_2 w_{22} & \dots & d_n w_{nn} \\ d_1^2 w_{11} & d_2^2 w_{22} & \dots & d_n^2 w_{nn} \\ d_1^3 w_{11} & d_2^3 w_{22} & \dots & d_n^3 w_{nn} \end{pmatrix} =$$

$$= \begin{pmatrix} 1 & d_1 & d_1^2 & d_1^3 \\ 1 & d_2 & d_2^2 & d_2^3 \\ \vdots & \vdots & \vdots & \vdots \\ 1 & d_n & d_n^2 & d_n^3 \end{pmatrix} \times \frac{1}{\xi} \begin{pmatrix} \alpha w_{11} + \epsilon d_1 w_{11} + \zeta d_1^2 w_{11} + \eta d_1^3 w_{11} & \alpha w_{22} + \epsilon d_2 w_{22} + \zeta d_2^2 w_{22} + \eta d_2^3 w_{22} & \dots & \alpha w_{nn} + \epsilon d_n w_{nn} + \zeta d_n^2 w_{nn} + \eta d_n^3 w_{nn} \\ \epsilon w_{11} + \beta d_1 w_{11} + \theta d_1^2 w_{11} + \kappa d_1^3 w_{11} & \epsilon w_{22} + \beta d_2 w_{22} + \theta d_2^2 w_{22} + \kappa d_2^3 w_{22} & \dots & \epsilon w_{nn} + \beta d_n w_{nn} + \theta d_n^2 w_{nn} + \kappa d_n^3 w_{nn} \\ \zeta w_{11} + \theta d_1 w_{11} + \gamma d_1^2 w_{11} + \lambda d_1^3 w_{11} & \zeta w_{22} + \theta d_2 w_{22} + \gamma d_2^2 w_{22} + \lambda d_2^3 w_{22} & \dots & \zeta w_{nn} + \theta d_n w_{nn} + \gamma d_n^2 w_{nn} + \lambda d_n^3 w_{nn} \\ \eta w_{11} + \kappa d_1 w_{11} + \lambda d_1^2 w_{11} + \delta d_1^3 w_{11} & \eta w_{22} + \kappa d_2 w_{22} + \lambda d_2^2 w_{22} + \delta d_2^3 w_{22} & \dots & \eta w_{nn} + \kappa d_n w_{nn} + \lambda d_n^2 w_{nn} + \delta d_n^3 w_{nn} \end{pmatrix}$$

$$= \begin{pmatrix} 1 & d_1 & d_1^2 & d_1^3 \\ 1 & d_2 & d_2^2 & d_2^3 \\ \vdots & \vdots & \vdots & \vdots \\ 1 & d_n & d_n^2 & d_n^3 \end{pmatrix} \times \frac{1}{\xi} \begin{pmatrix} (\alpha + \epsilon d_1 + \zeta d_1^2 + \eta d_1^3) w_{11} & (\alpha + \epsilon d_2 + \zeta d_2^2 + \eta d_2^3) w_{22} & \dots & (\alpha + \epsilon d_n + \zeta d_n^2 + \eta d_n^3) w_{nn} \\ (\epsilon + \beta d_1 + \theta d_1^2 + \kappa d_1^3) w_{11} & (\epsilon + \beta d_2 + \theta d_2^2 + \kappa d_2^3) w_{22} & \dots & (\epsilon + \beta d_n + \theta d_n^2 + \kappa d_n^3) w_{nn} \\ (\zeta + \theta d_1 + \gamma d_1^2 + \lambda d_1^3) w_{11} & (\zeta + \theta d_2 + \gamma d_2^2 + \lambda d_2^3) w_{22} & \dots & (\zeta + \theta d_n + \gamma d_n^2 + \lambda d_n^3) w_{nn} \\ (\eta + \kappa d_1 + \lambda d_1^2 + \delta d_1^3) w_{11} & (\eta + \kappa d_2 + \lambda d_2^2 + \delta d_2^3) w_{22} & \dots & (\eta + \kappa d_n + \lambda d_n^2 + \delta d_n^3) w_{nn} \end{pmatrix}$$

$$= \begin{pmatrix} 1 & d_1 & d_1^2 & d_1^3 \\ 1 & d_2 & d_2^2 & d_2^3 \\ \vdots & \vdots & \vdots & \vdots \\ 1 & d_n & d_n^2 & d_n^3 \end{pmatrix} \times \frac{1}{\xi} \begin{pmatrix} (\alpha + \epsilon d_1 + \zeta d_1^2 + \eta d_1^3) w_{11} & (\alpha + \epsilon d_2 + \zeta d_2^2 + \eta d_2^3) w_{22} & \dots & (\alpha + \epsilon d_n + \zeta d_n^2 + \eta d_n^3) w_{nn} \\ (\epsilon + \beta d_1 + \theta d_1^2 + \kappa d_1^3) w_{11} & (\epsilon + \beta d_2 + \theta d_2^2 + \kappa d_2^3) w_{22} & \dots & (\epsilon + \beta d_n + \theta d_n^2 + \kappa d_n^3) w_{nn} \\ (\zeta + \theta d_1 + \gamma d_1^2 + \lambda d_1^3) w_{11} & (\zeta + \theta d_2 + \gamma d_2^2 + \lambda d_2^3) w_{22} & \dots & (\zeta + \theta d_n + \gamma d_n^2 + \lambda d_n^3) w_{nn} \\ (\eta + \kappa d_1 + \lambda d_1^2 + \delta d_1^3) w_{11} & (\eta + \kappa d_2 + \lambda d_2^2 + \delta d_2^3) w_{22} & \dots & (\eta + \kappa d_n + \lambda d_n^2 + \delta d_n^3) w_{nn} \end{pmatrix}$$

$$= \frac{1}{\xi} \begin{pmatrix} (\mu_1 + \nu_1 d_1 + \sigma_1 d_1^2 + \tau_1 d_1^3) w_{11} & (\mu_2 + \nu_2 d_1 + \sigma_2 d_1^2 + \tau_2 d_1^3) w_{22} & \dots & (\mu_n + \nu_n d_1 + \sigma_n d_1^2 + \tau_n d_1^3) w_{nn} \\ (\mu_1 + \nu_1 d_2 + \sigma_1 d_2^2 + \tau_1 d_2^3) w_{11} & (\mu_2 + \nu_2 d_2 + \sigma_2 d_2^2 + \tau_2 d_2^3) w_{22} & \dots & (\mu_n + \nu_n d_2 + \sigma_n d_2^2 + \tau_n d_2^3) w_{nn} \\ \vdots & \vdots & \ddots & \vdots \\ (\mu_1 + \nu_1 d_n + \sigma_1 d_n^2 + \tau_1 d_n^3) w_{11} & (\mu_2 + \nu_2 d_n + \sigma_2 d_n^2 + \tau_2 d_n^3) w_{22} & \dots & (\mu_n + \nu_n d_n + \sigma_n d_n^2 + \tau_n d_n^3) w_{nn} \end{pmatrix}$$

where $\mu_j = \alpha + \epsilon d_j + \zeta d_j^2 + \eta d_j^3$, $\nu_j = \epsilon + \beta d_j + \theta d_j^2 + \kappa d_j^3$, $\sigma_j = \zeta + \theta d_j + \gamma d_j^2 + \lambda d_j^3$, $\tau_j = \eta + \kappa d_j + \lambda d_j^2 + \delta d_i^3$.

For n random specified points, we conclude with the following global smoothing matrix:

$$\begin{pmatrix} \frac{(\mu_1(x_1)+\nu_1(x_1)d_1(x_1)+\sigma_1(x_1)d_1^2(x_1)+\tau_1(x_1)d_1^3(x_1))w_{11}(x_1)}{\xi(x_1)} & \frac{(\mu_2(x_1)+\nu_2(x_1)d_1(x_1)+\sigma_2(x_1)d_1^2(x_1)+\tau_2(x_1)d_1^3(x_1))w_{22}(x_1)}{\xi(x_1)} & \cdots & \frac{(\mu_n(x_1)+\nu_n(x_1)d_1(x_1)+\sigma_n(x_1)d_1^2(x_1)+\tau_n(x_1)d_1^3(x_1))w_{nn}(x_1)}{\xi(x_1)} \\ \frac{(\mu_1(x_2)+\nu_1(x_2)d_2(x_2)+\sigma_1(x_2)d_2^2(x_2)+\tau_1(x_2)d_2^3(x_2))w_{11}(x_2)}{\xi(x_2)} & \frac{(\mu_2(x_2)+\nu_2(x_2)d_2(x_2)+\sigma_2(x_2)d_2^2(x_2)+\tau_2(x_2)d_2^3(x_2))w_{22}(x_2)}{\xi(x_2)} & \cdots & \frac{(\mu_n(x_2)+\nu_n(x_2)d_2(x_2)+\sigma_n(x_2)d_2^2(x_2)+\tau_n(x_2)d_2^3(x_2))w_{nn}(x_2)}{\xi(x_2)} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{(\mu_1(x_n)+\nu_1(x_n)d_n(x_n)+\sigma_1(x_n)d_n^2(x_n)+\tau_1(x_n)d_n^3(x_n))w_{11}(x_n)}{\xi(x_n)} & \frac{(\mu_2(x_n)+\nu_2(x_n)d_n(x_n)+\sigma_2(x_n)d_n^2(x_n)+\tau_2(x_n)d_n^3(x_n))w_{22}(x_n)}{\xi(x_n)} & \cdots & \frac{(\mu_n(x_n)+\nu_n(x_n)d_n(x_n)+\sigma_n(x_n)d_n^2(x_n)+\tau_n(x_n)d_n^3(x_n))w_{nn}(x_n)}{\xi(x_n)} \end{pmatrix}$$

where the parameters mentioned above are defined as follow:

- $\mu_j(x_i) = \alpha(x_i) + \epsilon(x_i)d_j(x_i) + \zeta(x_i)d_j^2(x_i) + \eta(x_i)d_j^3(x_i)$,
- $\nu_j(x_i) = \epsilon(x_i) + \beta(x_i)d_j(x_i) + \theta(x_i)d_j^2(x_i) + \kappa(x_i)d_j^3(x_i)$,
- $\sigma_j(x_i) = \zeta(x_i) + \theta(x_i)d_j(x_i) + \gamma(x_i)d_j^2(x_i) + \lambda(x_i)d_j^3(x_i)$,
- $\tau_j(x_i) = \eta(x_i) + \kappa(x_i)d_j(x_i) + \lambda(x_i)d_j^2(x_i) + \delta(x_i)d_i^3(x_i)$,
- $\xi(x_i) = s_0(x_i)s_2(x_i)s_4(x_i)s_6(x_i)+2s_0(x_i)s_3(x_i)s_4(x_i)s_5(x_i)+2s_1(x_i)s_3(x_i)s_4^2(x_i)+s_1^2(x_i)s_5^2(x_i)+2s_1(x_i)s_2(x_i)s_3(x_i)s_6(x_i)+2s_2^2(x_i)s_3(x_i)s_5(x_i)+s_2^2(x_i)s_4^2(x_i)+s_3^4(x_i)-(s_0(x_i)s_4^3(x_i)+s_0(x_i)s_2(x_i)s_5^2(x_i)+s_0(x_i)s_3^2(x_i)s_6(x_i)+s_1^2(x_i)s_4(x_i)s_6(x_i)+2s_1(x_i)s_3^2(x_i)s_5(x_i)+2s_1(x_i)s_2(x_i)s_4(x_i)s_5(x_i)+s_2^3(x_i)s_6(x_i)+3s_2(x_i)s_3^2(x_i)s_4(x_i))$.

The argument x_i denotes the values obtained for the corresponding factors for the i-th specified point. In the above derivation, the number of specified observations which has been taken is equal to the number of the observations. However, this is not a limitation as regards the number of points we can choose. So a different number (higher or lower) of specified

points can be selected, depending on how frequently one wants to approximate the regression function. Then, the global smoothing matrix will be transformed analogously.

Furthermore, in the case we choose to evaluate our regression function at the same points as those given from the observations, then the global smoothing matrix is further simplified as follows:

$$\begin{pmatrix} \frac{\alpha(x_1)w_{11}(x_1)}{\xi(x_1)} & \frac{\mu_2(x_1)w_{22}(x_1)}{\xi(x_1)} & \cdots & \frac{\mu_n(x_1)w_{nn}(x_1)}{\xi(x_1)} \\ \frac{\mu_1(x_2)w_{11}(x_2)}{\xi(x_2)} & \frac{\alpha(x_2)w_{22}(x_2)}{\xi(x_2)} & \cdots & \frac{\mu_n(x_2)w_{nn}(x_2)}{\xi(x_2)} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\mu_1(x_n)w_{11}(x_n)}{\xi(x_n)} & \frac{\mu_2(x_n)w_{22}(x_n)}{\xi(x_n)} & \cdots & \frac{\alpha(x_n)w_{nn}(x_n)}{\xi(x_n)} \end{pmatrix}.$$

D.2 Derivation of the Mean Squared Error

By definition, the bias and variance at a point x_0 can be expressed as:

$$\begin{aligned} Bias(x_0) &= \mathbb{E}[\widehat{m}^{(k)}(x_0;h)] - m^{(k)}(x_0), \\ Var[x_0] &= \mathbb{E}\left[\left(\widehat{m}^{(k)}(x_0;h) - \mathbb{E}[\widehat{m}^{(k)}(x_0;h)]\right)^2\right] \end{aligned}$$

where $\widehat{m}^{(k)}$ denotes the estimation of the k^{th}-derivative of the m function. Then, the MSE at a point x_0, given the bandwidth h, is defined as:

$$\begin{aligned} MSE(x_0;h) &= \mathbb{E}\left[\left(\widehat{m}^{(k)}(x_0;h) - m^{(k)}(x_0)\right)^2\right] \\ &= \mathbb{E}\left[\left(\widehat{m}^{(k)}(x_0;h) - \mathbb{E}[\widehat{m}^{(k)}(x_0;h)] + \mathbb{E}[\widehat{m}^{(k)}(x_0;h)] - m^{(k)}(x_0)\right)^2\right] \\ &= \mathbb{E}\left[\left(\widehat{m}^{(k)}(x_0;h) - \mathbb{E}[\widehat{m}^{(k)}(x_0;h)]\right)^2\right] + \mathbb{E}\left[\left(\mathbb{E}[\widehat{m}^{(k)}(x_0;h)] - m^{(k)}(x_0)\right)^2\right] \\ &\quad + 2\,\mathbb{E}\left[\left(\widehat{m}^{(k)}(x_0;h) - \mathbb{E}[\widehat{m}^{(k)}(x_0;h)]\right)\left(\mathbb{E}[\widehat{m}^{(k)}(x_0;h)] - m^{(k)}(x_0)\right)\right] \\ &= Var[x_0] + Bias^2(x_0). \end{aligned}$$

D.3 Derivation of the Expectation of Cross-Validation

The expectation of the cross-validation can be written as follows:

$$\begin{aligned} \mathbb{E}[CV(h)] &= \mathbb{E}\left[\frac{1}{n}\sum_{i=1}^{n}(\widehat{m}_{(-i)}(x_i;h) - y_i)^2\right] \\ &= \mathbb{E}\left[\frac{1}{n}\sum_{i=1}^{n}\left(\widehat{m}_{(-i)}(x_i;h) - (m(x_i) + \sigma(x_i)\epsilon_i)\right)^2\right] \quad \left(\text{by Equation (4.1)}\right) \\ &= \frac{1}{n}\sum_{i=1}^{n}\Bigg(\mathbb{E}\left[\left(\widehat{m}_{(-i)}(x_i;h) - m(x_i)\right)^2\right] - 2\,\mathbb{E}\left[\left(\widehat{m}_{(-i)}(x_i;h) - m(x_i)\right)\sigma(x_i)\epsilon_i\right] + \\ &\qquad + 2\,\mathbb{E}[m(x_i)\sigma(x_i)\epsilon_i] + \mathbb{E}[\sigma^2(x)\epsilon_i^2]\Bigg) \\ &= \frac{1}{n}\sum_{i=1}^{n}\left(\mathbb{E}\left[\left(\widehat{m}_{(-i)}(x_i;h) - m(x_i)\right)^2\right] + \sigma^2(x_i)\right) \quad (\text{since } \mathbb{E}[\epsilon_i] = 0,\ Var[\epsilon_i] = 1) \end{aligned}$$

$$\approx \int_0^1 \mathbb{E}\left[\left(\widehat{m}(x;h) - m(x)\right)^2\right] dx + \int_0^1 \sigma^2(x)\, dx$$
$$= MISE + \widetilde{\sigma^2}$$

where $\widetilde{\sigma^2}$ denotes the integrated variance, so it is a constant.

D.4 Moments of Returns

For the moments of normal random variables, we use the following lemma.

Lemma D.4.1 *(Moment Generating Function for a Normal Random Variable - Ross (2010)) The moment generating function for a normal random variable $X \sim N(\mu, \sigma^2)$ is given by:*

$$M(t) = \mathbb{E}\left[e^{tX}\right] = \exp\left\{t\mu + \frac{1}{2}\sigma^2 t^2\right\}.$$

So $\mathbb{E}[X^k] = M^{(k)}(0)$, for $k = 1, 2, \ldots$. In particular, the first four moments are given by:

$$\begin{aligned}
M'(t) &= M(t)(\mu + \sigma^2 t)\,, \qquad M'(0) = \mu \\
M''(t) &= M'(t)(\mu + \sigma^2 t) + \sigma^2 M(t)\,, \quad M''(0) = \mu^2 + \sigma^2 \\
M'''(t) &= M''(t)(\mu + \sigma^2 t) + 2\sigma^2 M'(t)\,, \quad M'''(0) = \mu^3 + 3\mu\sigma^2 \\
M^{(4)}(t) &= M'''(t)(\mu + \sigma^2 t) + 3\sigma^2 M''(t)\,, \quad M^{(4)}(0) = \mu^4 + 6\mu^2\sigma^2 + 3\sigma^4.
\end{aligned}$$

We recall that:

$$R^*(t_i) = \int_{t_{i-1}}^{t_i} \mu(\varsigma) d\varsigma + \int_{t_{i-1}}^{t_i} \sigma(\varsigma) dW(\varsigma) \sim N\Big(\int_{t_{i-1}}^{t_i} \mu(\varsigma) d\varsigma, \int_{t_{i-1}}^{t_i} \sigma^2(\varsigma) d\varsigma\Big) \triangleq N(r_i, s_i^2) \tag{D.1}$$

where $r_i = \mathbb{E}[R^*(t_i)]$ and $s_i^2 = Var[R^*(t_i)]$. So by Lemma D.4.1, we have that:

$$\mathbb{E}\left[\left(R^*(t_i)\right)^2\right] = r_i^2 + s_i^2, \quad \mathbb{E}\left[\left(R^*(t_i)\right)^3\right] = r_i^3 + 3r_i s_i^2, \quad \mathbb{E}\left[\left(R^*(t_i)\right)^4\right] = r_i^4 + 6r_i^2 s_i^2 + 3s_i^4.$$

Therefore, the variance of the true squared returns is given by:

$$Var\Big[\Big(R^*(t_i)\Big)^2\Big] = \mathbb{E}\Big[\Big(R^*(t_i)\Big)^4\Big] - \Big(\Big(R^*(t_i)\Big)^2\Big)^2 = 4r_i^2 s_i^2 + 2s_i^4.$$

For the errors $E(t)$, we have shown in the appendix B.1.1 that:

$$\mathbb{E}[V(t_i)] = \mathbb{E}[V^3(t_i)] = 0, \quad \mathbb{E}[V^2(t_i)] = 2\omega^2, \quad \mathbb{E}[V^4(t_i)] = 2(a_4 + 3a_2^2)$$

where $\mathbb{E}[E^k(t_i)] = a_k$ and, by convention, $\mathbb{E}[E^2(t_i)] = a_2 = \omega^2$. Putting them together, we can derive an expression for the variances of $R(t_i)$ and $R^2(t_i)$. More specifically, we have that:

$$\begin{aligned}
\mathbb{E}\Big[R(t_i)\Big] &= \mathbb{E}\Big[R^*(t_i)\Big] = r_i, \\
\mathbb{E}\Big[R^2(t_i)\Big] &= \mathbb{E}\Big[\mathbb{E}\Big[R^2(t_i)\Big|R^*(t_i)\Big]\Big] = \mathbb{E}\Big[R^{*2}(t_i)\Big] + 2a_2 = r_i^2 + s_i^2 + 2a_2, \\
\mathbb{E}\Big[R^4(t_i)\Big] &= \mathbb{E}\Big[\mathbb{E}\Big[R^4(t_i)\Big|R^*(t_i)\Big]\Big] = \mathbb{E}\Big[R^{*4}(t_i)\Big] + 6\,\mathbb{E}\Big[R^{*2}(t_i)\Big]\,\mathbb{E}\Big[V^2(t_i)\Big] + \mathbb{E}\Big[V^4(t_i)\Big] \\
&= r_i^4 + 6r_i^2 s_i^2 + 3s_i^4 + 12(r_i^2 + s_i^2)a_2 + 2a_4 + 6a_2^2.
\end{aligned} \tag{D.2}$$

So, the variances of $R^2(t_i)$ and $R^2(t_i)$ can be expressed with the following forms:

$$Var\Big[R(t_i)\Big] = \mathbb{E}\Big[R^2(t_i)\Big] - \Big(\mathbb{E}\Big[R(t_i)\Big]\Big)^2 = s_i^2 + a_2, \tag{D.3}$$

$$Var\Big[R^2(t_i)\Big] = \mathbb{E}\Big[R^4(t_i)\Big] - \Big(\mathbb{E}\Big[R^2(t_i)\Big]\Big)^2 = 2s_i^4 + 4r_i^2 s_i^2 + 8(r_i^2 + s_i^2)a_2 + 2(a_4 + a_2^2). \tag{D.4}$$

Appendix E

Appendix for Chapter 5

E.1 Description of Relevant Concepts

This section expresses mathematically some relevant concepts that we mention in Chapter 5.

Frobenius Norm

Assuming a $m \times n$ matrix A:

$$A = \begin{bmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} & \dots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m2} & \dots & a_{mn} \end{bmatrix},$$

then the Frobenius norm of A, $||A||_F$, is calculated by:

$$||A||_F = \sqrt{\sum_{row=1}^{m} \sum_{col=1}^{n} a_{row\,col}^2}.$$

Kronecker Product

Assuming two matrices A, B of size $m \times n$ and $p \times q$, respectively:

$$A = \begin{pmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} & \dots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m2} & \dots & a_{mn} \end{pmatrix}, B = \begin{pmatrix} b_{11} & b_{12} & \dots & b_{1q} \\ b_{21} & b_{22} & \dots & b_{2q} \\ \vdots & \vdots & \ddots & \vdots \\ b_{p1} & b_{p2} & \dots & b_{pq} \end{pmatrix},$$

then the Kronecker product of A and B, $A \bigotimes B$, is given by:

$$A \bigotimes B = \begin{pmatrix} a_{11}B & a_{12}B & \dots & a_{1n}B \\ a_{21}B & a_{22}B & \dots & a_{2n}B \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1}B & a_{m2}B & \dots & a_{mn}B \end{pmatrix}$$

$$= \begin{pmatrix} a_{11}b_{11} & \dots & a_{11}b_{1q} & a_{12}b_{11} & \dots & a_{12}b_{1q} & \dots & a_{1n}b_{11} & \dots & a_{1n}b_{1q} \\ a_{11}b_{21} & \dots & a_{11}b_{2q} & a_{12}b_{21} & \dots & a_{12}b_{2q} & \dots & a_{1n}b_{21} & \dots & a_{1n}b_{2q} \\ \vdots & & \vdots & \vdots & & \vdots & & \vdots & & \vdots \\ a_{11}b_{p1} & \dots & a_{11}b_{pq} & a_{12}b_{p1} & \dots & a_{12}b_{pq} & \dots & a_{1n}b_{p1} & \dots & a_{1n}b_{pq} \\ a_{21}b_{11} & \dots & a_{21}b_{1q} & a_{22}b_{11} & \dots & a_{22}b_{1q} & \dots & a_{2n}b_{11} & \dots & a_{2n}b_{1q} \\ a_{21}b_{21} & \dots & a_{21}b_{2q} & a_{22}b_{21} & \dots & a_{22}b_{2q} & \dots & a_{2n}b_{21} & \dots & a_{2n}b_{2q} \\ \vdots & & \vdots & \vdots & & \vdots & & \vdots & & \vdots \\ a_{21}b_{p1} & \dots & a_{21}b_{pq} & a_{22}b_{p1} & \dots & a_{22}b_{pq} & \dots & a_{2n}b_{p1} & \dots & a_{2n}b_{pq} \\ \vdots & & \vdots & \vdots & & \vdots & & \vdots & & \vdots \\ a_{m1}b_{11} & \dots & a_{m1}b_{1q} & a_{m2}b_{11} & \dots & a_{m2}b_{1q} & \dots & a_{mn}b_{11} & \dots & a_{mn}b_{1q} \\ a_{m1}b_{21} & \dots & a_{m1}b_{2q} & a_{m2}b_{21} & \dots & a_{m2}b_{2q} & \dots & a_{mn}b_{21} & \dots & a_{mn}b_{2q} \\ \vdots & & \vdots & \vdots & & \vdots & & \vdots & & \vdots \\ a_{m1}b_{p1} & \dots & a_{m1}b_{pq} & a_{m2}b_{p1} & \dots & a_{m2}b_{pq} & \dots & a_{mn}b_{p1} & \dots & a_{mn}b_{pq} \end{pmatrix}$$

where $A \bigotimes B \in \mathbb{R}^{mp \times nq}$.

ℓ_p-Norm :

The ℓ_p-norm of a variable X, where $X = (x_1, ..., x_n)$, is defined by:

$$||X||_p = \Big(\sum_{i=1}^{n} |x_i|^p \Big)^{1/p}.$$

E.2 Derivation of the GLASSO Algorithm

Considering a p-dimensional vector X, such that $X \sim N(0, \Sigma)$, then the corresponding multivariate Gaussian PDF is given by:

$$\begin{aligned} f(x) &= \frac{1}{(2\pi)^{p/2} \det(\Sigma)^{1/2}} e^{\{-\frac{1}{2} x^T \Sigma^{-1} x\}} \\ &= \frac{1}{(2\pi)^{p/2}} e^{\{-\frac{1}{2} x^T \Theta x\}} \det(\Theta)^{1/2} \qquad \Big(\det(A^{-1}) = \tfrac{1}{\det(A)} \Big) \\ &\propto e^{\{-\frac{1}{2} x^T \Theta x\}} \det(\Theta)^{1/2} \end{aligned}$$

where $\Sigma = \mathbb{E}[XX^T]$ is the (non-negative definite) variance-covariance matrix and $\Theta = \Sigma^{-1}$, assuming that Σ^{-1} exists. By taking the maximum log-likelihood function of the Equation (5.2) for a set of n IID samples x_i, we obtain the following quantity to maximise:

$$\begin{aligned} \mathbb{L}(\Theta) &= \frac{1}{n} \sum_{i=1}^{n} \log f(x_i) \\ &= \frac{1}{2} \log \det(\Theta) - \frac{1}{2n} \sum_{i=1}^{n} x_i^T \Theta x_i \qquad \Big(\text{by (5.2)} \Big) \\ &= \frac{1}{2} \Bigg(\log \Big(\det(\Theta) \Big) - \text{tr}(S\Theta) \Bigg) \qquad \Big(\text{see proof (E.5)} \Big) \\ &\propto \log \Big(\det(\Theta) \Big) - \text{tr}(S\Theta) \end{aligned}$$

where $S = \frac{1}{n}\sum_{i=1}^{n} x_i x_i^T$ is the sample variance-covariance matrix. Considering a ℓ_1-norm penalty, the optimal Θ matrix is given by maximising the following quantity:

$$\widehat{\Theta} = \arg\max_{\Theta \succ 0} \left(\log\Big(\det(\Theta)\Big) - \text{tr}(S\Theta) - \lambda \sum_{j=1}^{p} ||\Theta_j||_1 \right) \tag{E.1}$$

where Θ_j is the j-th column-vector (or row-vector since Θ is symmetric) of the Θ matrix and λ denotes the tuning parameter, such that $\lambda \geq 0$.

Equation (E.1) describes a convex optimisation problem, and so the optimal parameter is given by solving the following gradient equation with respect to Θ:

$$\begin{aligned} &\Theta^{-1} - \frac{1}{n}\sum_{i=1}^{n} x_i x_i^T - \lambda\Gamma = 0_{p\times p} \Rightarrow \\ &\Theta^{-1} - S - \lambda\Gamma = 0_{p\times p} \Rightarrow \\ &W - S - \lambda\Gamma = 0_{p\times p} \end{aligned} \tag{E.2}$$

where $W = \Theta^{-1}$, assuming that Θ^{-1} is defined. In the following, zero vectors and matrices are depicted with zero bold in this appendix.

As regards the first term of Equation (E.2), as Boyd et al. (2004) shows (p. 641) and Friedman et al. (2008) note, $\Big(\log\Big(\det(\Theta)\Big)\Big)' = \Theta^{-1}$. In terms of the second term of Equation (E.2), assuming that:

$$\Theta = \begin{bmatrix} \theta_{1,1} & \theta_{1,2} & \dots & \theta_{1,n} \\ \theta_{2,1} & \theta_{2,2} & \dots & \theta_{2,n} \\ \vdots & \vdots & \ddots & \vdots \\ \theta_{n,1} & \theta_{n,2} & \dots & \theta_{n,n} \end{bmatrix}, X = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix},$$

then for $\frac{1}{n}X^T\Theta X$ we have that:

$$\begin{aligned}
&\frac{1}{n}\begin{pmatrix} x_1 & x_2 & \dots & x_n \end{pmatrix}\begin{pmatrix} \theta_{1,1} & \theta_{1,2} & ... & \theta_{1,n} \\ \theta_{2,1} & \theta_{2,2} & ... & \theta_{2,n} \\ \vdots & \vdots & \ddots & \vdots \\ \theta_{n,1} & \theta_{n,2} & ... & \theta_{n,n} \end{pmatrix}\begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix} \\
&= \frac{1}{n}\begin{pmatrix} \sum_{i=1}^n x_i\theta_{i,1} & \sum_{i=1}^n x_i\theta_{i,2} & ... & \sum_{i=1}^n x_i\theta_{i,n} \end{pmatrix}\begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix} \\
&= \frac{1}{n}\left[\left(\sum_{i=1}^n x_i\theta_{i,1}\right)x_1 + \left(\sum_{i=1}^n x_i\theta_{i,2}\right)x_2 + \cdots + \left(\sum_{i=1}^n x_i\theta_{i,n}\right)x_n\right] \\
&= \frac{1}{n}\Big(\sum_{i=1}^n x_i\theta_{i,1}x_1 + \sum_{i=1}^n x_i\theta_{i,2}x_2 + \cdots + \sum_{i=1}^n x_i\theta_{i,n}x_n\Big).
\end{aligned}$$

By mathematical properties, we have that:

$$\frac{\partial f}{\partial \Theta} = \begin{pmatrix} \frac{\partial f}{\partial \theta_{1,1}} & \frac{\partial f}{\partial \theta_{1,2}} & \cdots & \frac{\partial f}{\partial \theta_{1,n}} \\ \frac{\partial f}{\partial \theta_{2,1}} & \frac{\partial f}{\partial \theta_{2,2}} & \cdots & \frac{\partial f}{\partial \theta_{2,n}} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial f}{\partial \theta_{n,1}} & \frac{\partial f}{\partial \theta_{n,2}} & \cdots & \frac{\partial f}{\partial \theta_{n,n}} \end{pmatrix}. \tag{E.3}$$

Hence, we get from Matrix (E.3):

$$\frac{\partial f}{\partial \Theta} = \frac{1}{n}\begin{pmatrix} x_1^2 & x_1x_2 & ... & x_1x_n \\ x_2x_1 & x_2^2 & ... & x_2x_n \\ \vdots & \vdots & \ddots & \vdots \\ x_nx_1 & x_nx_2 & ... & x_n^2 \end{pmatrix} \tag{E.4}$$

where $f(x_1, x_2, \dots, x_n) = \frac{1}{n}\Big(\sum_{i=1}^n x_i\theta_{i,1}x_1 + \sum_{i=1}^n x_i\theta_{i,2}x_2 + \cdots + \sum_{i=1}^n x_i\theta_{i,n}x_n\Big)$. Fur-

ther, we notice that (E.4) can be equivalently written as $\frac{1}{n}XX^T$ which defines the S matrix.

As regards the derivative of the third term of Equation (E.2), by mathematical properties the derivative of the function $f(x) = |x|$ is given by:

$$\frac{df(x)}{dx} = \frac{d(|x|)}{dx} : \begin{cases} = 1, & \text{if } x > 0 \\ \text{not defined}, & \text{if } x = 0 \\ = -1, & \text{if } x < 0 \end{cases} \cdot$$

Since the derivative of x is not defined at $x = 0$, we use the notion of subgradients (see Bertsekas et al. (2003) among others), which generalises the idea of gradients for non-differentiable points of a convex function. If we assume a convex function f, then the subgradient g at the point x for all y is defined by:

$$f(y) \geq f(x) + g^T(y - x).$$

At the points where the function f is non-differentiable (i.e. at point $x = 0$ in our case), then the subgradient may not be unique in its domain. In this case, the set of all planes that go through the point $x = 0$ are given by the subdifferential $\partial f(x)$ at this point, obtaining:

$$\frac{\partial f(x)}{\partial x} = \frac{\partial |x|}{\partial x} : \begin{cases} = 1, & \text{if } x > 0 \\ \in [-1, 1], & \text{if } x = 0 \\ = -1, & \text{if } x < 0 \end{cases} \cdot$$

When $x = 0$, we have $\frac{\partial(|x|)}{\partial x} \in [-1, 1]$. This happens because of the following reason. Since the point $x = 0$ is not differentiable, the approximation at this point is given by the subgradients that go through this problematic point. All the possible lines (subgradients) that go through this point need a slope between $[-1, 1]$. A visual representation of

the subgradients (dashed red lines) at a non-differentiable point $(0, 0)$ for the function $f(x) = |x|$ (blue line) is given in Figure E.1.

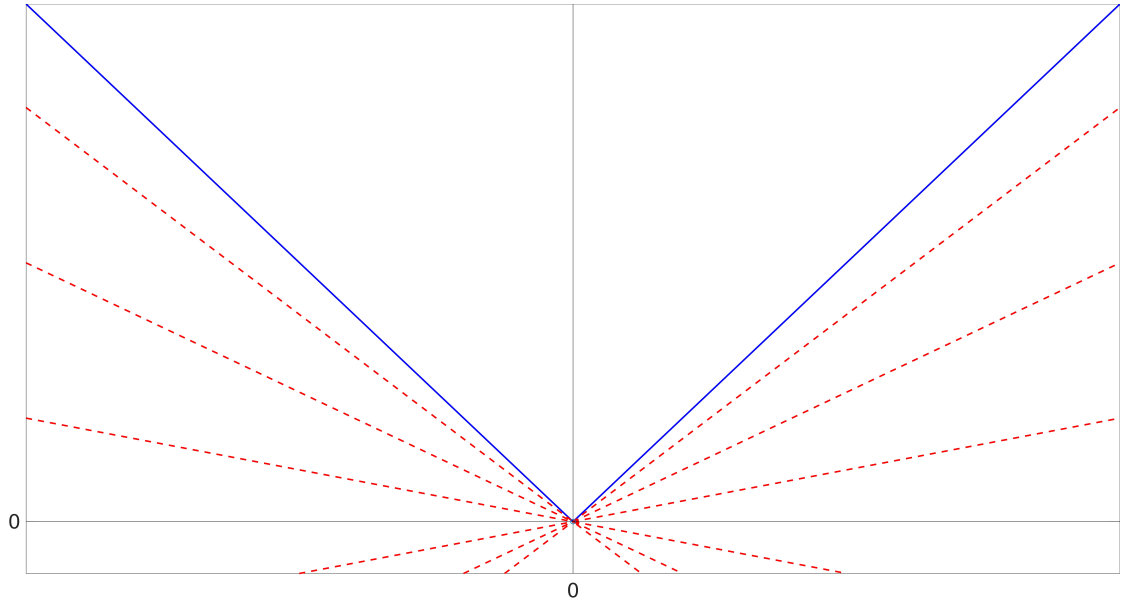

Figure E.1: The function $f(x) = |x|$ (blue line) is plotted along with several possible subgradients (dashed red lines) at point $(0, 0)$.

Similarly, in our case we obtain that:

$$\Gamma_{ij} : \begin{cases} = 1, & \text{if } \Theta_{ij} > 0 \\ \in [-1, 1], & \text{if } \Theta_{ij} = 0 \\ = -1, & \text{if } \Theta_{ij} < 0 \end{cases}$$

where i and j indicate the particular element in the matrices Γ and Θ.

E.2.1 Steps of the GLASSO Algorithm

Exploiting the derivation in appendix E.2 and using standard mathematical operations on blockwise invesion of a matrix, similar to those given in (E.8) in appendix E.4.2, we can derive the full steps of the coordinate descent algorithm. The steps of the

GLASSO algorithm are reported below.

Algorithm 5 GLASSO Algorithm

• Initialise $W = S + \lambda I_p$;
$\rightarrow$ S: the sample variance-covariance matrix;
$\rightarrow$ λ: the tuning parameter;
$\rightarrow$ I_p: the identity matrix.
for $iter = 1, 2, ...,$ until convergence **do**:
for $j = 1, ..., p$ **do**:
• Partition of W, S and Γ into:

$$W = \begin{bmatrix} W_{11} & w_{12} \\ w_{21}^T & W_{22} \end{bmatrix}, S = \begin{bmatrix} S_{11} & s_{12} \\ s_{21}^T & S_{22} \end{bmatrix}, \Gamma = \begin{bmatrix} \Gamma_{11} & \gamma_{12} \\ \gamma_{21}^T & \Gamma_{22} \end{bmatrix}$$

with respect to the j-th row/column vector;
• Solve:

$$W_{11}\beta - s_{12} + \lambda\gamma_{12} = \mathbf{0}$$

where $\gamma_{12} \in \text{sign}(\beta)$;
• Update w_{12} as follows:

$$w_{12} = W_{11}\beta;$$

end for
• Set $\widehat{\theta}_{12} = -\widehat{\theta}_{22}\beta$, where:

$$\widehat{\theta}_{22} = \frac{1}{w_{22} - w_{12}^T\beta};$$

if $\text{mean}(|\widehat{W}^{(iter)} - \widehat{W}^{(iter-1)}|) < 1E^{-6}$ **then**
Return Θ;
end if
where $\widehat{W}^{(iter)}$ is the $\widehat{W}$ matrix as results by the corresponding iteration and $\widehat{W}^{(0)}$ refers to the initial W matrix.
end for

E.3 Properties of the Functional Variance-Covariance Matrix

We assume that an observed value with noise can be represented by:

$$Y_{j,d}(t_i) = X_{j,d}(t_i) + E_{j,d}(t_i)$$

where $Y_{j,d}(t_i)$ denotes the i-th observed value of the stock j for the day d such that $i = 1, \ldots, n$, $j = 1, \ldots, p$ and $d = 1, \ldots, D$. Also, $X_{j,d}(t_i)$ is the corresponding true value and $E_{j,d}(t_i)$ indicates the corresponding error term. We consider that the error terms $E_{j,d}(t_i)$ are IID, with $\mathbb{E}[E_{j,d}(t_i)] = 0$, $\mathbb{E}[E^2_{j,d}(t_i)] = \omega^2 < \infty$ and $X(t) \perp\!\!\!\perp E(t)$. The covariance function of the noisy observations $Y_{j,d}(t_k)$ and $Y_{j,d}(t_l)$ can be written as:

$$
\begin{aligned}
&Cov\Big(Y_{j,d}(t_k), Y_{j,d}(t_l)\Big) = \mathbb{E}\left[\Big(Y_{j,d}(t_k) - \mathbb{E}[Y_{j,d}(t_k)]\Big)\Big(Y_{j,d}(t_l) - \mathbb{E}[Y_{j,d}(t_l)]\Big)\right] \\
&= \mathbb{E}\left[Y_{j,d}(t_k)Y_{j,d}(t_l) - Y_{j,d}(t_k)\,\mathbb{E}[Y_{j,d}(t_l)] - Y_{j,d}(t_l)\,\mathbb{E}[Y_{j,d}(t_k)] + \mathbb{E}[Y_{j,d}(t_k)]\,\mathbb{E}[Y_{j,d}(t_l)]\right] \\
&= \mathbb{E}\Big[X_{j,d}(t_k)X_{j,d}(t_l) + X_{j,d}(t_k)E_{j,d}(t_l) + X_{j,d}(t_l)E_{j,d}(t_k) + E_{j,d}(t_k)E_{j,d}(t_l) \\
&\quad - \Big(X_{j,d}(t_k) + E_{j,d}(t_k)\Big)\Big(\mathbb{E}[X_{j,d}(t_l)] + \mathbb{E}[E_{j,d}(t_l)]\Big) \\
&\quad - \Big(X_{j,d}(t_l) + E_{j,d}(t_l)\Big)\Big(\mathbb{E}[X_{j,d}(t_k)] + \mathbb{E}[E_{j,d}(t_k)]\Big) \\
&\quad + \Big(\mathbb{E}[X_{j,d}(t_k)] + \mathbb{E}[E_{j,d}(t_k)]\Big)\Big(\mathbb{E}[X_{j,d}(t_l)] + \mathbb{E}[E_{j,d}(t_l)]\Big)\Big] \\
&= \mathbb{E}\Big[X_{j,d}(t_k)X_{j,d}(t_l) - X_{j,d}(t_k)\,\mathbb{E}[X_{j,d}(t_l)] - X_{j,d}(t_l)\,\mathbb{E}[X_{j,d}(t_k)] \\
&\quad + \mathbb{E}[X_{j,d}(t_k)]\,\mathbb{E}[X_{j,d}(t_l)]\Big] + \mathbb{E}[E_{j,d}(t_k)E_{j,d}(t_l)] \\
&= \mathbb{E}\left[\Big(X_{j,d}(t_k) - \mathbb{E}[X_{j,d}(t_k)]\Big)\Big(X_{j,d}(t_l) - \mathbb{E}[X_{j,d}(t_l)]\Big)\right] + \mathbb{E}[E_{j,d}(t_k)E_{j,d}(t_l)] \\
&= Cov[X_{j,d}(t_k), X_{j,d}(t_l)] + \mathbb{E}[E_{j,d}(t_k)E_{j,d}(t_l)]
\end{aligned}
$$

where $k, l = 1, ..., n$.

In the case where $k = l$:

$$
\begin{aligned}
&Cov\Big(Y_{j,d}(t_k), Y_{j,d}(t_k)\Big) = \\
&= \mathbb{E}\left[\Big(X_{j,d}(t_k) - \mathbb{E}[X_{j,d}(t_k)]\Big)\Big(X_{j,d}(t_k) - \mathbb{E}[X_{j,d}(t_k)]\Big)\right] + \mathbb{E}[E_{j,d}(t_k)E_{j,d}(t_k)] \\
&= \mathbb{E}\Big[(X_{j,d}(t_k) - \mathbb{E}[X_{j,d}(t_k)])^2\Big] + \mathbb{E}[E^2_{j,d}(t_k)] \\
&= Var[X_{j,d}(t_k)] + \omega^2
\end{aligned}
$$

where $\mathbb{E}[E^2_{j,d}(t_i)] = \omega^2$ the noise variance.

In the case where $k \neq l$:

$$
\begin{aligned}
&Cov\Big(Y_{j,d}(t_k), Y_{j,d}(t_l)\Big) = \\
&= \mathbb{E}\left[\Big(X_{j,d}(t_k) - \mathbb{E}[X_{j,d}(t_k)]\Big)\Big(X_{j,d}(t_l) - \mathbb{E}[X_{j,d}(t_l)]\Big)\right] + \mathbb{E}[E_{j,d}(t_k)]\,\mathbb{E}[E_{j,d}(t_l)] \\
&= \mathbb{E}\left[\Big(X_{j,d}(t_k) - \mathbb{E}[X_{j,d}(t_k)]\Big)\Big(X_{j,d}(t_l) - \mathbb{E}[X_{j,d}(t_l)]\Big)\right] \qquad \Big(\text{by } X(t) \perp\!\!\!\perp E(t)\ \Big) \\
&= Cov\Big(X_{j,d}(t_k), X_{j,d}(t_k)\Big).
\end{aligned}
$$

E.4 Derivation of the FGLASSO Algorithm

This appendix section provides some preliminary calculations required for the derivation of the functional GLASSO approach. In this derivation, we assume that Σ denotes the variance-covariance matrix of a multivariate variable X, $\Theta = \Sigma^{-1}$ under the assumption that the Σ^{-1} matrix is defined, and S denotes the sample variance-covariance matrix of X. Also, the zero vectors and matrices are presented with zero bold.

E.4.1 Preliminary Calculations

- We show that $\frac{1}{n}\sum_{i=1}^{n} x_i^T \Theta x_i = \text{tr}(S\Theta)$.

Proof:

$$
\begin{aligned}
\frac{1}{n}\sum_{i=1}^{n} x_i^T \Theta x_i &= \frac{1}{n}(x_1^T\Theta x_1 + x_2^T\Theta x_2 + ... + x_n^T\Theta x_n) \qquad \text{(where a scalar)} \\
&= \text{tr}(\frac{1}{n}(x_1^T\Theta x_1 + x_2^T\Theta x_2 + ... + x_n^T\Theta x_n)) \qquad \Big(\text{by c= tr(c)}\Big) \\
&= \frac{1}{n}\Big(\text{tr}(x_1^T\Theta x_1) + \text{tr}(x_2^T\Theta x_2) + ... + \text{tr}(x_n^T\Theta x_n)\Big) \\
&= \frac{1}{n}\sum_{i=1}^{n}\text{tr}(x_i^T\Theta x_i)) \qquad \Big(\text{by tr(A+B)= tr(A)+ tr(B)}\Big)
\end{aligned}
$$

$$= \frac{1}{n}\sum_{i=1}^{n} \text{tr}(x_i x_i^T \Theta)) \qquad \Big(\text{by tr(AB)= tr(BA)}\Big)$$

$$= \text{tr}\Big(\frac{1}{n}\sum_{i=1}^{n} x_i x_i^T \Theta\Big) \qquad \Big(\text{by tr(A)+ tr(B)= tr(A+B)}\Big)$$

$$= \text{tr}(S\Theta). \tag{E.5}$$

□

$$\bullet \text{tr}(S\Theta) = \text{tr}(S_{jj}\Theta_{jj}) + 2\text{tr}(s_j^T \theta_j)$$

Proof: By induction:

Assuming three variables ($p = 3$), we have:

$$\Theta = \begin{bmatrix} \Theta_{11} & \theta_{12} & \theta_{13} \\ \theta_{21} & \Theta_{22} & \theta_{23} \\ \theta_{31} & \theta_{32} & \Theta_{33} \end{bmatrix}, S = \begin{bmatrix} S_{11} & s_{12} & s_{13} \\ s_{21} & S_{22} & s_{23} \\ s_{31} & s_{32} & S_{33} \end{bmatrix}$$

where S_{jj}, Θ_{jj} denote the diagonal submatrices of S and Θ, respectively, for $j = 1, 2, 3$ and s_{ij}, θ_{ij} indicate the off-diagonal submatrices $\forall i < j$. So:

$$S\Theta = \begin{bmatrix} S_{11} & s_{12} & s_{13} \\ s_{21} & S_{22} & s_{23} \\ s_{31} & s_{32} & S_{33} \end{bmatrix} \begin{bmatrix} \Theta_{11} & \theta_{12} & \theta_{13} \\ \theta_{21} & \Theta_{22} & \theta_{23} \\ \theta_{31} & \theta_{32} & \Theta_{33} \end{bmatrix}$$

$$= \begin{bmatrix} S_{11}\Theta_{11} + s_{12}\theta_{12}^T + s_{13}\theta_{13}^T & S_{11}\theta_{12} + s_{12}\Theta_{22} + s_{13}\theta_{23}^T & S_{11}\theta_{13} + s_{12}\theta_{23} + s_{13}\Theta_{33} \\ s_{12}^T\Theta_{11} + S_{22}\theta_{12}^T + s_{23}\theta_{13}^T & s_{12}^T\theta_{12} + S_{22}\Theta_{22} + s_{23}\theta_{23}^T & s_{12}^T\theta_{13} + S_{22}\theta_{23} + s_{23}\Theta_{33}. \\ s_{13}^T\Theta_{11} + s_{23}^T\theta_{12}^T + S_{33}\theta_{13}^T & s_{13}^T\theta_{12} + s_{23}^T\Theta_{22} + S_{33}\theta_{23}^T & s_{13}^T\theta_{13} + s_{23}^T\theta_{23} + S_{33}\Theta_{33} \end{bmatrix}.$$

Since the elements of the matrices S and Θ are submatrices of size $n \times n$, then the trace

of the product $S\Theta$, it is straightforward that it is given by:

$$\begin{aligned}
\text{tr}(S\Theta) &= \text{tr}\Big((S_{11}\Theta_{11} + s_{12}\theta_{12}^T + s_{13}\theta_{13}^T) + (s_{12}^T\theta_{12} + S_{22}\Theta_{22} + s_{23}\theta_{23}^T) + (s_{13}^T\theta_{13} + s_{23}^T\theta_{23} + S_{33}\Theta_{33})\Big) \\
&= \text{tr}\Big((S_{11}\Theta_{11} + S_{22}\Theta_{22} + S_{33}\Theta_{33}) + (s_{12}\theta_{12}^T + s_{13}\theta_{13}^T + s_{23}\theta_{23}^T) + (s_{12}^T\theta_{12} + s_{13}^T\theta_{13} + s_{23}^T\theta_{23})\Big) \\
&= \text{tr}(S_{11}\Theta_{11} + S_{22}\Theta_{22} + S_{33}\Theta_{33}) + \text{tr}(s_{12}\theta_{12}^T + s_{13}\theta_{13}^T + s_{23}\theta_{23}^T) + \text{tr}(s_{12}^T\theta_{12} + s_{13}^T\theta_{13} + s_{23}^T\theta_{23}) \\
&\quad \Big(\text{by the trace property } \text{tr}(A+B) = \text{tr}(A) + \text{tr}(B)\Big) \\
&= \text{tr}(S_{11}\Theta_{11} + S_{22}\Theta_{22} + S_{33}\Theta_{33}) + 2\text{tr}(s_{12}^T\theta_{12} + s_{13}^T\theta_{13} + s_{23}^T\theta_{23}) \\
&\quad \Big(\text{by the trace property } \text{tr}(A^TB) = \text{tr}(AB^T)\Big) \\
&= \text{tr}\Big(\sum_{j=1}^{3} S_{jj}\Theta_{jj}\Big) + 2\text{tr}\Big(\sum_{i<j}^{3} s_{ij}^T\theta_{ij}\Big).
\end{aligned}$$

We assume that it also holds for $(p-1)$ variables:

$$\begin{aligned}
\text{tr}(S\Theta) &= \text{tr}(S_{11}\Theta_{11} + ... + S_{(p-1)(p-1)}\Theta_{(p-1)(p-1)}) + 2\text{tr}(s_{12}^T\theta_{12} + ... + s_{(p-2)(p-1)}^T\theta_{(p-2)(p-1)}) \\
&= \text{tr}\Big(\sum_{j=1}^{(p-1)} S_{jj}\Theta_{jj}\Big) + 2\text{tr}\Big(\sum_{i<j}^{p-1} s_{ij}^T\theta_{ij}\Big).
\end{aligned}$$

Now, we show that it also holds for p variables:

$$\begin{aligned}
\text{tr}(S\Theta) = \text{tr}\Big(& (S_{11}\Theta_{11} + s_{12}\theta_{12}^T + s_{13}\theta_{13}^T + ... + s_{1(p-1)}\theta_{1(p-1)}^T + s_{1p}\theta_{1p}^T) \\
& + (s_{12}^T\theta_{12} + S_{22}\Theta_{22} + s_{23}\theta_{23}^T + s_{24}\theta_{24}^T + ... + s_{2(p-1)}\theta_{2(p-1)}^T + s_{2p}\theta_{2p}^T) \\
& + (s_{13}^T\theta_{13} + s_{23}^T\theta_{23} + S_{33}\Theta_{33} + s_{34}\theta_{34}^T + ... + s_{3(p-1)}\theta_{3(p-1)}^T + s_p\theta_{3p}^T) \\
& + (s_{14}^T\theta_{14} + s_{24}^T\theta_{24} + s_{34}^T\theta_{34} + S_{44}\Theta_{44} + s_{45}\theta_{45}^T + s_{46}\theta_{46}^T + ... + s_{4(p-1)}\theta_{4(p-1)}^T + s_{4p}\theta_{4p}^T) \\
& \qquad\qquad \vdots \\
& + (s_{1p}^T\theta_{1p} + s_{2p}^T\theta_{2p} + s_{3p}^T\theta_{3p} + ... + s_{(p-1)p}\theta_{(p-1)p}^T + S_{pp}\Theta_{pp})\Big).
\end{aligned}$$

Separating the elements of the products of the submatrices for the p-th variable to a different parenthesis factor, we have:

$$\begin{aligned}\mathrm{tr}(S\Theta) = \mathrm{tr}\Bigg(& (S_{11}\Theta_{11} + ... + s_{1(p-1)}\theta_{1(p-1)}^T) + (s_{12}^T\theta_{12} + ... + s_{2(p-1)}\theta_{2(p-1)}^T) \\ & + (s_{13}^T\theta_{13} + ... + s_{3(p-1)}\theta_{3(p-1)}^T) + (s_{14}^T\theta_{14} + ... + s_{4(p-1)}\theta_{4(p-1)}^T) \\ & \vdots \\ & + (s_{1(p-1)}^T\theta_{1(p-1)} + ... + s_{(p-1)(p-1)}\theta_{(p-1)(p-1)}^T) + (s_{1p}^T\theta_{1p} + ... + s_{pp}\theta_{pp}^T) \\ & + (s_{1p}^T\theta_{1p} + s_{2p}^T\theta_{2p} + + ... + s_{(p-1)p}\theta_{(p-1)p}^T\Bigg).\end{aligned}$$

Assuming that the under examination formula holds for $(p-1)$ variables, then:

$$\begin{aligned}\mathrm{tr}(S\Theta) &= \mathrm{tr}\Big(\sum_{j=1}^{p-1} S_{jj}\Theta_{jj}\Big) + \mathrm{tr}\Big(\sum_{i<j}^{p-1} s_{ij}^T\theta_{ij}\Big) + \mathrm{tr}\Big(\sum_{i<j}^{p-1} s_{ij}\theta_{ij}^T\Big) + \mathrm{tr}(S_{pp}\Theta_{pp}) \\ &\quad + \mathrm{tr}(s_{1j}^T\theta_{1j} + ... + s_{(p-1)p}^T\theta_{(p-1)p}) + \mathrm{tr}(s_{1j}\theta_{1j}^T + ... + s_{(p-1)p}\theta_{(p-1)p}^T) \\ &= \mathrm{tr}\Big(\sum_{j=1}^{p} S_{jj}\Theta_{jj}\Big) + \mathrm{tr}\Big(\sum_{i<j}^{p} s_{ij}^T\theta_{ij}\Big) + \mathrm{tr}\Big(\sum_{i<j}^{p} s_{ij}\theta_{ij}^T\Big) \\ &\quad \Big(\text{by the trace property } \mathrm{tr}(A+B) = \mathrm{tr}(A) + \mathrm{tr}(B)\Big) \\ &= \mathrm{tr}\Big(\sum_{j=1}^{p} S_{jj}\Theta_{jj}\Big) + 2\mathrm{tr}\Big(\sum_{i<j}^{p} s_{ij}^T\theta_{ij}\Big) \\ &\quad \Big(\text{by the trace property } \mathrm{tr}(A^TB) = \mathrm{tr}(AB^T)\Big) \\ &= \mathrm{tr}(S_{jj}\Theta_{jj}) + \mathrm{tr}\Big(\sum_{l\neq j}^{p} S_{ll}\Theta_{ll}\Big) + \mathrm{tr}\Big(\sum_{l\neq j}^{p} s_{jl}^T\theta_{jl} + s_{lj}^T\theta_{lj}\Big) + \mathrm{tr}\Big(\sum_{i\neq j<l}^{p} s_{il}^T\theta_{il} + s_{li}^T\theta_{li}\Big)\end{aligned}$$

In all the cases above, for a fixed row/column block-vector, we have that:

$$\mathrm{tr}(S\Theta) = \mathrm{tr}(S_{jj}\Theta_{jj}) + 2\mathrm{tr}(s_j^T\theta_j),$$

so:

$$\frac{\partial\Big(\mathrm{tr}(S\Theta)\Big)}{\partial\widetilde{\Theta_j}} = \frac{\partial}{\partial\widetilde{\Theta_j}}\Big(\mathrm{tr}(S_{jj}\Theta_{jj}) + 2\mathrm{tr}(s_j^T\theta_j)\Big)$$

where $\widetilde{\Theta_j} = [\Theta_{jj}\ \theta_j]$.

□

E.4.2 Derivation of the FGLASSO Approach

We consider the multivariate Gaussian model with zero mean and variance Σ:

$$\begin{aligned} f(x) &= \frac{1}{(2\pi)^{p/2}\det(\Sigma)^{1/2}} e^{\{-\frac{1}{2}x^T\Sigma^{-1}x\}} \\ &= \frac{1}{(2\pi)^{p/2}} e^{\{-\frac{1}{2}x^T\Theta x\}}\det(\Theta)^{1/2} \quad \Big(\text{by } \det(A^{-1}) = \tfrac{1}{\det(A)}\Big) \\ &\propto e^{\{-\frac{1}{2}x^T\Theta x\}}\det(\Theta)^{1/2} \end{aligned}$$

where X indicates a p-dimensional vector, such that $X \sim N(0,\Sigma)$, $\Sigma = \mathbb{E}[XX^T]$ is the (non-negative definite) variance-covariance matrix and $\Theta = \Sigma^{-1}$, assuming that Σ^{-1} is defined.

Our purpose is the maximisation of f. To do so, we take the maximum log-likelihood of the above quantity with respect to Θ for n IID samples:

$$\begin{aligned} \mathbb{L}(\Theta) &= \frac{1}{n}\log\Big(f(x_i)\Big) \\ &= \frac{1}{n}\sum_{i=1}^{n}\left(\log(e^{-\frac{1}{2}x_i^T\Theta x_i}) + \log\Big(\det(\Theta)^{1/2}\Big)\right) \\ &= \frac{1}{n}\sum_{i=1}^{n}\log(e^{-\frac{1}{2}x_i^T\Theta x_i}) + \frac{1}{n}n\log\Big(\det(\Theta)^{1/2}\Big) \\ &= -\frac{1}{n}\sum_{i=1}^{n}(\frac{1}{2}x_i^T\Theta x_i)\log(e) + \frac{1}{2}\log\Big(\det(\Theta)\Big) \\ &= \frac{\sum_{i=1}^{n}x_i^T\Theta x_i}{2n} + \frac{1}{2}\log\Big(\det(\Theta)\Big) \qquad \text{(by E.5)} \\ &\propto \log\Big(\det(\Theta)\Big) - \mathrm{tr}(S\Theta) \end{aligned}$$

where $S = \frac{1}{n}\sum_{i=1}^{n} x_i x_i^T$ is the sample variance-covariance matrix.

Assume that $\underline{x}^T = (x_1^1 \cdots x_1^n | x_2^1 \cdots x_2^n | \cdots | x_p^1 \cdots x_p^n)$, such that $\underline{x} \in \mathbb{R}^{np}$ where p denotes the number of variables and n indicates the number of observations. By adding a norm-penalty in the above quantity, then we are looking for that $\widehat{\Theta}$ which maximises:

$$\widehat{\Theta} = \arg\max_{\Theta} \left(\log\Big(\det(\Theta)\Big) - \text{tr}(S\Theta) - \lambda \sum_{j \neq l} ||\theta_{jl}||_F \right) \tag{E.6}$$

where $||\cdot||_F$ denotes the Frobenius norm, λ denotes the tuning parameter, such that $\lambda \geq 0$, and $\widehat{\Theta} \in \mathbb{R}^{np \times np}$. This represents a convex optimisation problem, and the maximisation of the above quantity is given through the gradient equation:

- $\dfrac{\partial\Big[\log\Big(\det(\Theta)\Big)\Big]}{\partial\Theta} = \Theta^{-1}$, (by Boyd et al. (2004), section A.4.1)
- $\dfrac{\partial \text{tr}(S\Theta)}{\partial\Theta} = \dfrac{\partial \text{tr}(\frac{1}{n}\sum_{i=1}^{n} x_i x_i^T \Theta)}{\partial\Theta} = \dfrac{1}{n}\sum_{i=1}^{n} x_i x_i^T = S,$
- $\dfrac{\partial(\lambda \sum_{j \neq l} ||\theta_{jl}||_F)}{\partial\Theta} = \dfrac{\lambda\partial(\sum_{j \neq l} ||\theta_{jl}||_F)}{\partial\Theta} = \lambda\Gamma$

where $\Gamma = \frac{\partial(\sum_{j \neq l} ||\theta_{jl}||_F)}{\partial\Theta}$. Amalgamating the above quantities in the same equation and setting this equation equal to zero, we have that:

$$\Theta^{-1} - S - \lambda\Gamma = \mathbf{0}. \tag{E.7}$$

Now, we define the following block-matrices:

$$\Theta = \begin{bmatrix} \Theta_{jj} & \theta_j^T \\ \theta_j & \Theta_{-j} \end{bmatrix}, S = \begin{bmatrix} S_{jj} & s_j^T \\ s_j & S_{-j} \end{bmatrix}, \Gamma = \begin{bmatrix} \Gamma_{jj} & \gamma_j^T \\ \gamma_j & \Gamma_{-j} \end{bmatrix}$$

where indicatively, the Θ matrix is of the following structure:

$$\Theta: \begin{array}{|c||c|}\hline \Theta_{jj}: & \theta_j^T: \\ n \times n & n \times n\ (p-1) \\ \hline\hline \theta_j: & \Theta_{-j}: \\ n(p-1) \times n & n(p-1) \times n(p-1) \\ \hline \end{array}$$

and the block-elements of the matrices S and Γ have the same form and dimension as the Θ block-matrix. In addition, we define the W block-matrix, such that $W = \Theta^{-1}$. By standard calculation on the inverse of a block-matrix (see Lu and Shiou (2002) among others), we have that:

$$\begin{aligned} \Theta^{-1} &= W \\ &= \begin{pmatrix} W_{jj} & w_j^T \\ w_j & W_{-j} \end{pmatrix} = \begin{pmatrix} (\Theta_{jj} - \theta_j^T \Theta_{-j}^{-1} \theta_j)^{-1} & -W_{jj} \theta_j^T \Theta_{-j}^{-1} \\ -\Theta_{-j}^{-1} \theta_j W_{jj} & \Theta_{-j}^{-1} + \Theta_{-j}^{-1} \theta_j W_{jj} \theta_j^T \Theta_{-j}^{-1} \end{pmatrix} \end{aligned} \tag{E.8}$$

assuming that Θ_{-j}^{-1} and $(\Theta_{jj} - \theta_j^T \Theta_{-j}^{-1} \theta_j)^{-1}$ exist.

Using standard mathematical matrix calculations in Equation (E.7), for a fixed row/-

column block-vector j, we have that:

$$\Theta^{-1} - S - \lambda\Gamma = \mathbf{0} \Leftrightarrow W - S - \lambda\Gamma = \mathbf{0}$$

where in matrix form, it is written as:

$$\begin{bmatrix} W_{jj} & w_j^T \\ w_j & W_{-j} \end{bmatrix} - \begin{bmatrix} S_{jj} & s_j^T \\ s_j & S_{-j} \end{bmatrix} - \lambda \begin{bmatrix} \Gamma_{jj} & \gamma_j^T \\ \gamma_j & \Gamma_{-j} \end{bmatrix} = \begin{bmatrix} \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{bmatrix}$$

obtaining the following four equations:

$$W_{jj} - S_{jj} - \lambda\Gamma_{jj} = \mathbf{0}, \tag{E.9}$$

$$w_j^T - s_j^T - \lambda\gamma_j^T = \mathbf{0},$$

$$w_j - s_j - \lambda\gamma_j = \mathbf{0},$$

$$W_{-j} - S_{-j} - \lambda\Gamma_{-j} = \mathbf{0}. \tag{E.10}$$

In the Equations (E.9) and (E.10), the diagonal elements of Γ_{jj} and Γ_{-j} are zero. Thus, solving the Equation (E.9) with respect to Θ_{jj}, we have that:

$$\begin{aligned} W_{jj} - S_{jj} - \lambda\Gamma_{jj} = \mathbf{0} \Leftrightarrow \\ (\Theta_{jj} - \theta_j^T\Theta_{-j}^{-1}\theta_j)^{-1} - S_{jj} - \mathbf{0} = \mathbf{0} \Leftrightarrow \\ \Theta_{jj} - \theta_j^T\Theta_{-j}^{-1}\theta_j = S_{jj}^{-1} \Leftrightarrow \\ \Theta_{jj} = S_{jj}^{-1} + \theta_j^T\Theta_{-j}^{-1}\theta_j, \end{aligned}$$

assuming that S_{jj}^{-1} exists. Also, Qiao et al. (2019) show that the minimisation of:

$$\widehat{\theta}_j = \arg\min_{\theta_j} \left(\text{tr}(S_{jj}\theta_j^T\Theta_{-j}^{-1}\theta_j) + 2\text{tr}(s_j^T\theta_j) + 2\lambda \sum_{l=1}^{p-1} ||\theta_{jl}||_F \right) \tag{E.11}$$

can estimate θ_j, where θ_{jl} indicates the l-th matrix of the block-vector θ_j and the complete derivation is given in the appendix section E.4.3.

The quantity in (E.11), for a fixed row/column block-vector j, is equivalent to the minimisation with respect to each one of the submatrices θ_{j1}, θ_{j2},...,$\theta_{j(p-1)}$ of the block-vector θ_j:

$$\arg\min_{\theta_{j1},\ldots,\theta_{j(p-1)}}\left(\text{tr}\Big(\sum_{l=1}^{p-1}\sum_{k=1}^{p-1}S_{jj}\theta_{jl}^T\{\Theta_{-j}^{-1}\}_{lk}\theta_{jk}+2\sum_{k=1}^{p-1}s_{jk}^T\theta_{jk}\Big)+2\lambda\sum_{k=1}^{p-1}||\theta_{jk}||_F\right)\Rightarrow$$

$$\arg\min_{\theta_{j1},\ldots,\theta_{j(p-1)}}\left(\text{tr}\Big(S_{jj}\theta_{jk}^T\{\Theta_{-j}^{-1}\}_{kk}\theta_{jk}+\sum_{l\neq k}^{p-1}S_{jj}\theta_{jl}^T\{\Theta_{-j}^{-1}\}_{lk}\theta_{jk}+2\sum_{k=1}^{p-1}s_{jk}^T\theta_{jk}\Big)+2\lambda\sum_{k=1}^{p-1}||\theta_{jk}||_F\right)\Rightarrow$$

$$\arg\min_{\theta_{j1},\ldots,\theta_{j(p-1)}}\left(\text{tr}(S_{jj}\theta_{jk}^T\{\Theta_{-j}^{-1}\}_{kk}\theta_{jk})+\text{tr}\Big(\sum_{l\neq k}^{p-1}S_{jj}\theta_{jl}^T\{\Theta_{-j}^{-1}\}_{lk}\theta_{jk}\Big)+2\text{tr}\Big(\sum_{k=1}^{p-1}s_{jk}^T\theta_{jk}\Big)+2\lambda\sum_{k=1}^{p-1}||\theta_{jk}||_F\right)$$

In order to minimise θ_{jk}, we take the partial derivative of the above formula and set this formula equal to zero with respect to θ_{jk}:

$$\frac{\partial\left(\text{tr}(S_{jj}\theta_{jk}^T\{\Theta_{-j}^{-1}\}_{kk}\theta_{jk})+\text{tr}(\sum_{l\neq k}^{p-1}S_{jj}\theta_{jl}^T\{\Theta_{-j}^{-1}\}_{lk}\theta_{jk})+2\text{tr}(\sum_{k=1}^{p-1}s_{jk}^T\theta_{jk})+2\lambda\sum_{k=1}^{p-1}||\theta_{jk}||_F\right)}{\partial\theta_{jk}}=\mathbf{0}$$

$$\frac{\partial\text{tr}(S_{jj}\theta_{jk}^T\{\Theta_{-j}^{-1}\}_{kk}\theta_{jk})}{\partial\theta_{jk}}+\frac{\partial\text{tr}(\sum_{l\neq k}^{p-1}S_{jj}\theta_{jl}^T\{\Theta_{-j}^{-1}\}_{lk}\theta_{jk})}{\partial\theta_{jk}}+2\frac{\partial\text{tr}(\sum_{k=1}^{p-1}s_{jk}^T\theta_{jk})}{\partial\theta_{jk}}$$
$$+2\lambda\frac{\partial\sum_{k=1}^{p-1}||\theta_{jk}||_F}{\partial\theta_{jk}}=\mathbf{0}.$$

Considering each term separately, we have:

- $\dfrac{\partial\text{tr}(S_{jj}\theta_{jk}^T\{\Theta_{-j}^{-1}\}_{kk}\theta_{jk})}{\partial\theta_{jk}}=\{\Theta_{-j}^{-1}\}_{kk}\theta_{jk}S_{jj}+\{\Theta_{-j}^{-1}\}_{kk}^T\theta_{jk}S_{jj}^T.$

This derivation is proven by Qiao et al. (2019) where, on page 16 in supplementary materials, they show that:

$$\frac{\partial\text{tr}(AX^T)BX}{\partial X}=BXA+B^TXA^T.$$

Also,

$$\bullet \ \frac{\partial \text{tr}(\sum_{l\neq k}^{p} S_{jj}\theta_{jl}^{T}\{\Theta_{-j}^{-1}\}_{lk}\theta_{jk})}{\partial \theta_{jk}} = \sum_{l\neq k}^{p-1}\left(\{\Theta_{-j}^{-1}\}_{lk}^{T}\theta_{jl}S_{jj}^{T} + \{\Theta_{-j}^{-1}\}_{kl}^{T}\theta_{jl}S_{jj}\right)$$

since:

$$\frac{\partial \text{tr}(AX^{T}B)}{\partial X} = BA \ \text{ and } \ \frac{\partial \text{tr}(XA)}{\partial X} = \frac{\partial \text{tr}(AX)}{\partial X} = A^{T}.$$

Furthermore,

$$\bullet \ \frac{\partial \text{tr}(\sum_{k=1}^{p} s_{jk}^{T}\theta_{jk})}{\partial \theta_{jk}} = s_{jk} \qquad \left(\text{by } \tfrac{\partial \text{tr}(A^{T}X)}{\partial X} = A\right)$$

and

$$\bullet \ \frac{\partial(\sum_{k=1}^{p-1}||\theta_{jk}||_{F})}{\partial \theta_{jk}} = \frac{\partial(\sqrt{\theta_{jk}^{(1,1)\ 2} + \theta_{jk}^{(1,2)\ 2} + ... + \theta_{jk}^{(n,n)\ 2}})}{\partial \theta_{jk}} = \frac{\theta_{jk}}{||\theta_{jk}||_{F}} = \psi_{jk}$$

where ψ_{jk} defines the quantity $\frac{\theta_{jk}}{||\theta_{jk}||_{F}}$ and $\theta_{jk}^{(i,j)}$ represents the element at i-th row, j-th column of the $n \times n$ matrix:

$$\begin{pmatrix} \theta_{jk}^{(1,1)} & \theta_{jk}^{(1,2)} & \cdots & \theta_{jk}^{(1,n)} \\ \theta_{jk}^{(2,1)} & \theta_{jk}^{(2,2)} & \cdots & \theta_{jk}^{(2,n)} \\ \vdots & \vdots & \ddots & \vdots \\ \theta_{jk}^{(n,1)} & \theta_{jk}^{(n,2)} & \cdots & \theta_{jk}^{(n,n)} \end{pmatrix}.$$

For example, for the first element of the matrix above, we have that:

$$\frac{\partial||\theta_{jk}||_{F}}{\partial \theta_{jk}^{(1,1)}} = \frac{2\theta_{jk}^{(1,1)}}{2||\theta_{jk}||_{F}} = \frac{\theta_{jk}^{(1,1)}}{||\theta_{jk}||_{F}}.$$

So, for each element of the matrix, we have that:

$$\frac{\partial||\theta_{jk}||_F}{\partial\theta_{jk}} = \begin{pmatrix} \frac{\partial||\theta_{jk}||_F}{\theta_{jk}^{(1,1)}} & \frac{\partial||\theta_{jk}||_F}{\theta_{jk}^{(1,2)}} & \dots & \frac{\partial||\theta_{jk}||_F}{\theta_{jk}^{(1,n)}} \\ \frac{\partial||\theta_{jk}||_F}{\theta_{jk}^{(2,1)}} & \frac{\partial||\theta_{jk}||_F}{\theta_{jk}^{(2,2)}} & \dots & \frac{\partial||\theta_{jk}||_F}{\theta_{jk}^{(2,n)}} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial||\theta_{jk}||_F}{\theta_{jk}^{(n,1)}} & \frac{\partial||\theta_{jk}||_F}{\theta_{jk}^{(n,2)}} & \dots & \frac{\partial||\theta_{jk}||_F}{\theta_{jk}^{(n,n)}} \end{pmatrix} = \begin{pmatrix} \frac{\theta_{jk}^{(1,1)}}{||\theta_{jk}||_F} & \frac{\theta_{jk}^{(1,2)}}{||\theta_{jk}||_F} & \dots & \frac{\theta_{jk}^{(1,n)}}{||\theta_{jk}||_F} \\ \frac{\theta_{jk}^{(2,1)}}{||\theta_{jk}||_F} & \frac{\theta_{jk}^{(2,2)}}{||\theta_{jk}||_F} & \dots & \frac{\theta_{jk}^{(2,n)}}{||\theta_{jk}||_F} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\theta_{jk}^{(n,1)}}{||\theta_{jk}||_F} & \frac{\theta_{jk}^{(n,2)}}{||\theta_{jk}||_F} & \dots & \frac{\theta_{jk}^{(n,n)}}{||\theta_{jk}||_F} \end{pmatrix}.$$

Amalgamating the partial derivatives in the same equation, we obtain that:

$$\{\Theta_{-j}^{-1}\}_{kk}\theta_{jk}S_{jj}+\{\Theta_{-j}^{-1}\}_{kk}^T\theta_{jk}S_{jj}^T+\sum_{l\neq k}^{p-1}\left(\{\Theta_{-j}^{-1}\}_{lk}^T\theta_{jl}S_{jj}^T+\{\Theta_{-j}^{-1}\}_{kl}^T\theta_{jl}S_{jj}\right)+2s_{jk}+2\lambda\psi_{jk}=\mathbf{0}.$$

Since both S and Θ are symmetric matrices we have that $S_{jj} = S_{jj}^T$ and, under the property of symmetric matrices $A = A^T \Leftrightarrow A^{-1} = (A^T)^{-1} \Leftrightarrow A^{-1} = (A^{-1})^T$, we also have that $\{\Theta_{-j}^{-1}\}_{lk}^T = \{\Theta_{-j}^{-1}\}_{kl}$. Therefore, we can simplify the above equation as follows:

$$2\{\Theta_{-j}^{-1}\}_{kk}\theta_{jk}S_{jj} + 2\sum_{l\neq k}^{p-1}\{\Theta_{-j}^{-1}\}_{lk}^T\theta_{jl}S_{jj} + 2s_{jk} + 2\lambda\psi_{jk} = \mathbf{0} \Leftrightarrow$$
$$\{\Theta_{-j}^{-1}\}_{kk}\theta_{jk}S_{jj} + \sum_{l\neq k}^{p-1}\{\Theta_{-j}^{-1}\}_{lk}^T\theta_{jl}S_{jj} + s_{jk} + \lambda\psi_{jk} = \mathbf{0}.$$

If we define the block-residuals by:

$$r_{jk} = \sum_{l\neq k}\{\Theta_{-j}^{-1}\}_{lk}^T\theta_{jl}S_{jj} + s_{jk},$$

then the equation above can be expressed as:

$$\{\Theta_{-j}^{-1}\}_{kk}\theta_{jk}S_{jj} + r_{jk} + \lambda\psi_{jk} = \mathbf{0}$$

where

$$\psi_{jk} : \begin{cases} \frac{\theta_{jk}}{||\theta_{jk}||_F}, & \text{if } \theta_{jk} \neq \mathbf{0} \\ \psi_{jk} \in \mathbb{R}^{n \times n} \text{ where } ||\psi_{jk}||_F \leq 1, & \text{if } \theta_{jk} = \mathbf{0} \end{cases}.$$

Concluding to the solution of θ_{jk}. If $\theta_{jk} = \mathbf{0}$, then $||r_{jk}||_F = \lambda||\psi_{jk}||_F \leq \lambda$. Otherwise, it is given by the following minimisation, with respect to θ_{jk}:

$$\arg\min_{\theta_{jk}} \left(\{\Theta_{-j}^{-1}\}_{kk} \theta_{jk} S_{jj} + \mathbf{r}_{jk} + \lambda \frac{\theta_{jk}}{||\theta_{jk}||_F} \right).$$

E.4.3 Derivation of the Coefficients Minimisation Problem

As we have seen in Equation (E.6) in the appendix section E.4.2, we are looking for that Θ which maximises:

$$\widehat{\Theta} = \arg\max_{\Theta} \left(\log\Big(\det(\Theta)\Big) - \text{tr}(S\Theta) - \lambda \sum_{j \neq l} ||\theta_{jl}||_F \right)$$

where θ_{jl} indicates the l-th submatrix in the j-th block-vector θ_j. Following the matrix notation on page 294 and using the Schur complement property, assuming that Θ_{-j} is invertible, we have that:

$$\det(\Theta) = \begin{pmatrix} \Theta_{jj} & \theta_j^T \\ \theta_j & \Theta_{-j} \end{pmatrix} = \log\Big(\det(\Theta_{-j})\Big) \log\Big(\det(\Theta_{jj} - \theta_j^T \Theta_{-j}^{-1} \theta_j)\Big),$$

then the first term of Equation (E.6) can be written as:

$$\begin{aligned} \log\Big(\det(\Theta)\Big) &= \log\Big(\det(\Theta_{-j}) \det(\Theta_{jj} - \theta_j^T \Theta_{-j}^{-1} \theta_j)\Big) \\ &= \log\Big(\det(\Theta_{-j})\Big) + \log\Big(\det(\Theta_{jj} - \theta_j^T \Theta_{-j}^{-1} \theta_j)\Big). \end{aligned}$$

If Θ_{jj} is constant, then the determinant of a constant number is constant, ending up with the following simplification:

$$\log\Big(\det(\Theta)\Big) = \log\Big(\det(\Theta_{-j})\Big) + \log\Big(\det(\Theta_{jj} - \theta_j^T\Theta_{-j}^{-1}\theta_j)\Big)$$

however, we are not interested in this case.

For the maximisation of the j-th row/column block-vector in Θ, it is straightforward that Θ_{-j} remains unchanged since the under examination block-vector is not contained. Therefore, $\log\Big(\det(\Theta_{-j})\Big)$ can be considered as a constant term and we can ignore it in the optimisation procedure. As regards the second term, we have shown in appendix E.4.1 that:

$$\frac{\partial \text{tr}(S\Theta)}{\partial\widetilde{\Theta_j}} = \frac{\partial}{\partial\widetilde{\Theta_j}}\Big(\text{tr}(S_{jj}\Theta_{jj}) + 2\text{tr}(s_j^T\theta_j)\Big)$$

where $\widetilde{\Theta_j} = [\Theta_{jj}\ \theta_j]$. So, we can rewrite the Equation (E.6), for fixed Θ_{-j}, as:

$$\widehat{\Theta} = \arg\max_{\Theta}\left(\log\Big(\det(\Theta_{jj} - \theta_j^T\Theta_{-j}^{-1}\theta_j)\Big) - \text{tr}(S_{jj}\Theta_{jj}) - 2\text{tr}(s_j^T\theta_j) - \lambda\sum_{j\neq l}||\theta_{jl}||_F\right). \tag{E.12}$$

Using the partial derivative of each term for the maximisation of the quantity above with respect to Θ_{jj}, we obtain that:

$$\begin{aligned}\bullet\ \frac{\partial\log\Big(\det(\Theta_{jj} - \theta_j^T\Theta_{-j}^{-1}\theta_j)\Big)}{\partial\Theta_{jj}} &= (\Theta_{jj} - \theta_j^T\Theta_{-j}^{-1}\theta_j)^{-1}(\Theta_{jj} - \theta_j^T\Theta_{-j}^{-1}\theta_j)' \\ &= (\Theta_{jj} - \theta_j^T\Theta_{-j}^{-1}\theta_j)^{-1}(I_n - \mathbf{0}) \\ &= (\Theta_{jj} - \theta_j^T\Theta_{-j}^{-1}\theta_j)^{-1}\end{aligned}$$

where this arises by the derivation of Boyd et al. (2004) in section A.4.1 who shows that

$\bigtriangledown \log\Big(\det(X)\Big) = X^{-1}$. Also,

$$\bullet\ \frac{\partial S_{jj}\Theta_{jj}^{-1}}{\partial \Theta_{jj}} = S_{jj}, \quad \bullet\ \frac{\partial s_j^T\theta_j}{\partial \Theta_{jj}} = \mathbf{0}, \quad \bullet\ \frac{\partial(\lambda\sum_{j\neq l}||\theta_{jl}||_F)}{\partial \Theta_{jj}} = \mathbf{0},$$

concluding to:

$$\begin{aligned}(\Theta_{jj} - \theta_j^T\Theta_{-j}^{-1}\theta_j)^{-1} - S_{jj} = \mathbf{0} \Rightarrow \\ (\Theta_{jj} - \theta_j^T\Theta_{-j}^{-1}\theta_j)^{-1} = S_{jj} \Rightarrow \\ \Theta_{jj} - \theta_j^T\Theta_{-j}^{-1}\theta_j = S_{jj}^{-1} \Rightarrow \\ \Theta_{jj} = S_{jj}^{-1} + \theta_j^T\Theta_{-j}^{-1}\theta_j, \qquad \text{(E.13)}\end{aligned}$$

assuming that S_{jj}^{-1} is defined.

Formula (E.12) is maximised for that Θ_{jj} given by (E.13). However, θ_j is unknown, so we have to find the profile likelihood. Considering the Equations (E.12) and (E.13), we obtain the maximum log-likelihood of θ_j:

$$\begin{aligned}\widehat{\Theta} &= \arg\max_{\theta_j}\Bigg(\log\Big(\det(S_{jj}^{-1} + \theta_j^T\Theta_{-j}^{-1}\theta_j - \theta_j^T\Theta_{-j}^{-1}\theta_j)\Big) - \mathrm{tr}\Big(S_{jj}(S_{jj}^{-1} + \theta_j^T\Theta_{-j}^{-1}\theta_j)\Big) \\ &\quad - 2\mathrm{tr}(s_j^T\theta_j) - 2\lambda\sum_{l=1}^{p-1}||\theta_{jl}||_F\Bigg) \\ &= \arg\max_{\theta_j}\Bigg(\log\Big(\det(S_{jj}^{-1})\Big) - \mathrm{tr}(I) - \mathrm{tr}(S_{jj}\theta_j^T\Theta_{-j}^{-1}\theta_j) - 2\mathrm{tr}(s_j^T\theta_j) - 2\lambda\sum_{l=1}^{p-1}||\theta_{jl}||_F\Bigg)\end{aligned}$$

(where the first two terms above remain fixed over j for θ_j)

$$\begin{aligned}&= \arg\max_{\theta_j}\Bigg(-\mathrm{tr}(S_{jj}\theta_j^T\Theta_{-j}^{-1}\theta_j) - 2\mathrm{tr}(s_j^T\theta_j) - 2\lambda\sum_{l=1}^{p-1}||\theta_{jl}||_F\Bigg) \\ &= \arg\min_{\theta_j}\Bigg(\mathrm{tr}(S_{jj}\theta_j^T\Theta_{-j}^{-1}\theta_j) + 2\mathrm{tr}(s_j^T\theta_j) + 2\lambda\sum_{l=1}^{p-1}||\theta_{jl}||_F\Bigg)\end{aligned}$$

ending up with the negative profile likelihood. Then this procedure is conducted for

every j row/column block-vector until convergence. Note that the term $\lambda \sum_{j \neq l} ||\theta_{jl}||_F$ can be equivalently written as $2\lambda \sum_{l=1}^{p-1} ||\theta_{jl}||_F$. This happens because the submatrices θ_{jl} and θ_{jl}^T are two components of the Θ matrix, which give the same outcome when we calculate their Frobenius norm.

E.5 Simulation Model

Assuming that the i-th true observation of a stock j_1, on the day d, is given by:

$$X_{j_1,d}(t_i) = \mu_{j_1,d}(t_i) + \xi_{j_1,d}^{(1)} \phi_{j_1,d}^{(1)}(t_i) + \xi_{j_1,d}^{(2)} \phi_{j_1,d}^{(2)}(t_i)$$

where $\mu_{j_1,d}(t_i)$ refers to the variable's mean function, $\phi_{j_1,d}(1)$, $\phi_{j_1,d}(2)$ indicate two eigenfunctions and $\xi_{j_1,d}^{(1)}$, $\xi_{j_1,d}^{(2)}$ indicate the functional principal component scores. According to the theoretical framework of functional PCA, the variance of $\xi^{(1)}$ and $\xi^{(2)}$ are given by the corresponding eigenvalues λ_1 and λ_2, respectively. Also, we assume that $\xi^{(1)}$ and $\xi^{(2)}$ stem from a normal distribution with zero mean. Hence, we have that:

$$\xi_{j_1,d}^{(1)} \sim N(0, \lambda_1) \quad \text{and} \quad \xi_{j_1,d}^{(2)} \sim N(0, \lambda_2). \tag{E.14}$$

Taking the expectation and the variance of $X_{j_1,d}(t_i)$, we obtain:

$$\begin{aligned}
\mathbb{E}[X_{j_1,d}(t_i)] &= \mathbb{E}[\mu_{j_1,d}(t_i) + \xi_{j_1,d}^{(1)} \phi_{j_1,d}^{(1)}(t_i) + \xi_{j_1,d}^{(2)} \phi_{j_1,d}^{(2)}(t_i)] \\
&= \mu_{j_1,d}(t_i) + \phi_{j_1,d}^{(1)}(t_i)\, \mathbb{E}[\xi_{j_1,d}^{(1)}] + \phi_{j_1,d}^{(2)}(t_i)\, \mathbb{E}[\xi_{j_1,d}^{(2)}] \\
&\quad \text{(since } \mu_{j_1,d}(t_i), \phi_{j_1,d}^{(1)}(t_i), \phi_{j_1,d}^{(2)}(t_i) \text{ constants)} \\
&= \mu_{j_1,d}(t_i), \qquad \Big(\text{by (E.14)}\Big) \\
Var[X_{j_1,d}(t_i)] &= Var[\mu_{j_1,d}(t_i) + \xi_{j_1,d}^{(1)} \phi_{j_1,d}^{(1)}(t_i) + \xi_{j_1,d}^{(2)} \phi_{j_1,d}^{(2)}(t_i)] \\
&= Var[\xi_{j_1,d}^{(1)} \phi_{j_1,d}^{(1)}(t_i) + \xi_{j_1,d}^{(2)} \phi_{j_1,d}^{(2)}(t_i)] \\
&\quad \text{(since } \mu_{j_1,d}(t_i) \text{ constant)}
\end{aligned}$$

$$
\begin{aligned}
&= Var[\xi^{(1)}_{j_1,d}\phi^{(1)}_{j_1,d}(t_i)] + Var[\xi^{(2)}_{j_1,d}\phi^{(2)}_{j_1,d}(t_i)] \\
&\quad + 2\,\mathbb{E}[(\xi^{(1)}_{j_1,d}\phi^{(1)}_{j_1,d}(t_i) - \mathbb{E}[\xi^{(1)}_{j_1,d}\phi^{(1)}_{j_1,d}(t_i)])(\xi^{(2)}_{j_1,d}\phi^{(2)}_{j_1,d}(t_i) - \mathbb{E}[\xi^{(2)}_{j_1,d}\phi^{(2)}_{j_1,d}(t_i)])] \\
&\quad (\text{since } Var[X+Y] = Var[X] + Var[Y] + 2\,\mathbb{E}[(X - \mathbb{E}[X])(Y - \mathbb{E}[Y])]) \\
&= \phi^{(1)2}_{j_1,d}(t_i)Var[\xi^{(1)}_{j_1,d}] + \phi^{(2)2}_{j_1,d}(t_i)Var[\xi^{(2)}_{j_1,d}] + 2\,\mathbb{E}[\xi^{(1)}_{j_1,d}\phi^{(1)}_{j_1,d}(t_i)\xi^{(2)}_{j_1,d}\phi^{(2)}_{j_1,d}(t_i)] \\
&\quad \Big(\text{by (E.14)}\Big) \\
&= \phi^{(1)2}_{j_1,d}(t_i)Var[\xi^{(1)}_{j_1,d}] + \phi^{(2)2}_{j_1,d}(t_i)Var[\xi^{(2)}_{j_1,d}] + 2\phi^{(1)}_{j_1,d}(t_i)\phi^{(2)}_{j_1,d}(t_i)\,\mathbb{E}[\xi^{(1)}_{j_1,d}]\,\mathbb{E}[\xi^{(2)}_{j_1,d}] \\
&\quad (\text{since } \xi^{(1)}_{j_1,d} \perp\!\!\!\perp \xi^{(2)}_{j_1,d}) \\
&= \lambda_1\phi^{(1)2}_{j_1,d}(t_i) + \lambda_2\phi^{(2)2}_{j_1,d}(t_i). \qquad \Big(\text{by (E.14)}\Big)
\end{aligned}
$$

Hence, $X_{j_1,d}(t_i) \sim N\Big(\mu_{j_1,d}(t_i), \lambda_1\phi^{(1)2}_{j_1,d}(t_i) + \lambda_2\phi^{(2)2}_{j_1,d}(t_i)\Big)$.

Now, we derive the properties of the simulation model considering a second variable $X_{j_2,d}(t_i)$, correlated to the variable $X_{j_1,d}(t_i)$ in appendix E.5.1. Also, we derive the properties of the model when the variable $X_{j_2,d}(t_i)$ is uncorrelated to the variable $X_{j_1,d}(t_i)$ in appendix E.5.2.

E.5.1 Correlated Variables

Assuming a second variable $X_{j_2,d}(t_i)$:

$$
X_{j_2,d}(t_i) = \mu_{j_2,d}(t_i) + \xi^{(1)}_{j_2,d}\phi^{(1)}_{j_2,d}(t_i) + \xi^{(2)}_{j_2,d}\phi^{(2)}_{j_2,d}(t_i)
$$

correlated to variable $X_{j_1,d}(t_i)$ with correlation coefficient ρ, then:

$$
\begin{aligned}
\xi^{(1)}_{j_2,d} &= \rho\xi^{(1)}_{j_1,d} + \epsilon^{(1)}_{j_2,d}, \\
\xi^{(2)}_{j_2,d} &= \rho\xi^{(2)}_{j_1,d} + \epsilon^{(2)}_{j_2,d}
\end{aligned}
$$

where $\epsilon^{(1)}_{j_2,d}$, $\epsilon^{(2)}_{j_2,d}$ indicate the error terms of the correlation structure. By the theory of correlated random variables, we have that:

$$\epsilon^{(1)}_{j_2,d} \sim N(0, 1-\rho^2) \quad \text{and} \quad \epsilon^{(2)}_{j_2,d} \sim N(0, 1-\rho^2). \tag{E.15}$$

The component $\xi^{(1)}_{j_2,d}$ has the following distribution properties:

$$\begin{aligned}
\mathbb{E}[\xi^{(1)}_{j_2,d}] &= \mathbb{E}[\rho\xi^{(1)}_{j_1,d} + \epsilon^{(1)}_{j_2,d}] \\
&= \rho\,\mathbb{E}[\xi^{(1)}_{j_1,d}] \qquad \Big(\text{by (E.15)}\Big) \\
&= 0, \qquad \Big(\text{by (E.14)}\Big) \\
Var[\xi^{(1)}_{j_2,d}] &= Var[\rho\xi^{(1)}_{j_1,d} + \epsilon^{(1)}_{j_2,d}] \\
&= Var[\rho\xi^{(1)}_{j_1,d}] + Var[\epsilon^{(1)}_{j_2,d}] + 2\,\mathbb{E}[(\rho\xi^{(1)}_{j_1,d} - EX[\rho\xi^{(1)}_{j_1,d}])(\epsilon^{(1)}_{j_2,d} - \mathbb{E}[\epsilon^{(1)}_{j_2,d}])] \\
&= \rho^2 Var[\xi^{(1)}_{j_1,d}] + Var[\epsilon^{(1)}_{j_2,d}] + 2\rho\,\mathbb{E}[\xi^{(1)}_{j_1,d}]\,\mathbb{E}[\epsilon^{(1)}_{j_2,d}] \\
&\quad \text{(by (E.14), (E.15) and } \xi^{(1)}_{j_1,d} \perp\!\!\!\perp \epsilon^{(1)}_{j_2,d}) \\
&= \rho^2\lambda_1 + (1-\rho^2). \qquad \Big(\text{by (E.14) and (E.15)}\Big)
\end{aligned}$$

Thus, we have that:

$$\xi^{(1)}_{j_2,d} \sim N\Big(0, \rho^2\lambda_1 + (1-\rho^2)\Big) \quad \text{and, similarly,} \quad \xi^{(2)}_{j_2,d} \sim N\Big(0, \rho^2\lambda_2 + (1-\rho^2)\Big). \tag{E.16}$$

So, the expectation and the variance of $X_{j_2,d}(t_i)$ can be expressed as:

$$\begin{aligned}
\mathbb{E}[X_{j_2,d}(t_i)] &= \mathbb{E}[\mu_{j_2,d}(t_i) + \xi^{(1)}_{j_2,d}\phi^{(1)}_{j_2,d}(t_i) + \xi^{(2)}_{j_2,d}\phi^{(2)}_{j_2,d}(t_i)] \\
&= \mu_{j_2,d}(t_i) + \phi^{(1)}_{j_2,d}(t_i)\,\mathbb{E}[\xi^{(1)}_{j_2,d}] + \phi^{(2)}_{j_2,d}(t_i)\,\mathbb{E}[\xi^{(2)}_{j_2,d}] \\
&\quad (\text{since } \mu_{j_2,d}(t_i), \phi^{(1)}_{j_2,d}(t_i), \phi^{(2)}_{j_2,d}(t_i) \text{ constants}) \\
&= \mu_{j_2,d}(t_i), \qquad \Big(\text{by (E.16)}\Big) \\
Var[X_{j_2,d}(t_i)] &= Var[\mu_{j_2,d}(t_i) + \xi^{(1)}_{j_2,d}\phi^{(1)}_{j_2,d}(t_i) + \xi^{(2)}_{j_2,d}\phi^{(2)}_{j_2,d}(t_i)]
\end{aligned}$$

$$
\begin{aligned}
&= Var[\xi^{(1)}_{j_2,d}\phi^{(1)}_{j_2,d}(t_i)] + Var[\xi^{(2)}_{j_2,d}\phi^{(2)}_{j_2,d}(t_i)] \\
&\quad + 2\,\mathbb{E}[(\xi^{(1)}_{j_2,d}\phi^{(1)}_{j_2,d}(t_i) - \mathbb{E}[\xi^{(1)}_{j_2,d}\phi^{(1)}_{j_2,d}(t_i)])(\xi^{(2)}_{j_2,d}\phi^{(2)}_{j_2,d}(t_i) - \mathbb{E}[\xi^{(2)}_{j_2,d}\phi^{(2)}_{j_2,d}(t_i)])] \\
&= \phi^{(1)2}_{j_2,d}(t_i)Var[\xi^{(1)}_{j_2,d}] + \phi^{(2)2}_{j_2,d}(t_i)Var[\xi^{(2)}_{j_2,d}] + 2\phi^{(1)}_{j_2,d}(t_i)\phi^{(2)}_{j_2,d}(t_i)\,\mathbb{E}[\xi^{(1)}_{j_2,d}]\,\mathbb{E}[\xi^{(2)}_{j_2,d}] \\
&\quad (\text{since } \xi^{(1)}_{j_2,d} \perp\!\!\!\perp \xi^{(2)}_{j_2,d}) \\
&= \phi^{(1)2}_{j_2,d}(t_i)\Big(\rho^2\lambda_1 + (1-\rho^2)\Big) + \phi^{(2)2}_{j_2,d}(t_i)\Big(\rho^2\lambda_2 + (1-\rho^2)\Big) \\
&= \rho^2\Big(\lambda_1\phi^{(1)2}_{j_2,d}(t_i) - \phi^{(1)2}_{j_2,d}(t_i) + \lambda_2\phi^{(2)2}_{j_2,d}(t_i) - \phi^{(2)2}_{j_2,d}(t_i)\Big) + \phi^{(1)2}_{j_2,d}(t_i) + \phi^{(2)2}_{j_2,d}(t_i).
\end{aligned}
$$

Concluding to the following properties for $X_{j_2,d}(t_i)$:

$$
X_{j_2,d}(t_i) \sim N\Big(\mu_{j_2,d}(t_i), \rho^2\Big(\lambda_1\phi^{(1)2}_{j_2,d}(t_i) - \phi^{(1)2}_{j_2,d}(t_i) + \lambda_2\phi^{(2)2}_{j_2,d}(t_i) - \phi^{(2)2}_{j_2,d}(t_i)\Big) + \phi^{(1)2}_{j_2,d}(t_i) + \phi^{(2)2}_{j_2,d}(t_i)\Big).
$$

The correlation between $X_{j_1,d}(t_k)$ and $X_{j_2,d}(t_l)$ is:

$$
Corr\Big(X_{j_1,d}(t_k), X_{j_2,d}(t_l)\Big) = \frac{Cov\Big(X_{j_1,d}(t_k), X_{j_2,d}(t_l)\Big)}{\sqrt{Var[X_{j_1,d}(t_k)]Var[X_{j_2,d}(t_l)]}}.
$$

The covariance can be expressed as:

$$
\begin{aligned}
&Cov\Big(X_{j_1,d}(t_k), X_{j_2,d}(t_l)\Big) = \\
&= \mathbb{E}[(X_{j_1,d}(t_k) - \mathbb{E}[X_{j_1,d}(t_k)])(X_{j_2,d}(t_l) - \mathbb{E}[X_{j_2,d}(t_l)])] \\
&= \mathbb{E}[(X_{j_1,d}(t_k) - \mu_{j_1,d}(t_k))(X_{j_2,d}(t_l) - \mu_{j_2,d}(t_l))] \\
&= \mathbb{E}[X_{j_1,d}(t_k)X_{j_2,d}(t_l)] - \mu_{j_2,d}(t_l)\,\mathbb{E}[X_{j_1,d}(t_k)] - \mu_{j_1,d}(t_k)\,\mathbb{E}[X_{j_2,d}(t_l)] + \mu_{j_1,d}(t_k)\mu_{j_2,d}(t_l) \\
&\quad (\text{since } \mu_{j_1,d}(t_k),\ \mu_{j_2,d}(t_l) \text{ constants}) \\
&= \mathbb{E}[X_{j_1,d}(t_k)X_{j_2,d}(t_l)] - \mu_{j_2,d}(t_l)\mu_{j_1,d}(t_k) \\
&= \mathbb{E}\left[\Big(\mu_{j_1,d}(t_k) + \xi^{(1)}_{j_1,d}\phi^{(1)}_{j_1,d}(t_k) + \xi^{(2)}_{j_1,d}\phi^{(2)}_{j_1,d}(t_k)\Big)\Big(\mu_{j_2,d}(t_l) + \xi^{(1)}_{j_2,d}\phi^{(1)}_{j_2,d}(t_l) + \xi^{(2)}_{j_2,d}\phi^{(2)}_{j_2,d}(t_l)\Big)\right] \\
&\qquad - \mu_{j_2,d}(t_l)\mu_{j_1,d}(t_k) \\
&= \phi^{(1)}_{j_1,d}(t_k)\phi^{(1)}_{j_2,d}(t_l)\,\mathbb{E}[\xi^{(1)}_{j_1,d}\xi^{(1)}_{j_2,d}] + \phi^{(1)}_{j_1,d}(t_k)\phi^{(2)}_{j_2,d}(t_l)\,\mathbb{E}[\xi^{(1)}_{j_1,d}\xi^{(2)}_{j_2,d}]
\end{aligned}
$$

$$
\begin{aligned}
&+ \phi^{(2)}_{j_1,d}(t_k)\phi^{(1)}_{j_2,d}(t_l)\,\mathbb{E}[\xi^{(2)}_{j_1,d}\xi^{(1)}_{j_2,d}] + \phi^{(2)}_{j_1,d}(t_k)\phi^{(2)}_{j_2,d}(t_l)\,\mathbb{E}[\xi^{(2)}_{j_1,d}\xi^{(2)}_{j_2,d}] \\
&\qquad \Big(\text{by (E.14) and (E.16)}\Big) \\
&= \phi^{(1)}_{j_1,d}(t_k)\phi^{(1)}_{j_2,d}(t_l)\,\mathbb{E}\left[\xi^{(1)}_{j_1,d}\Big(\rho\xi^{(1)}_{j_1,d} + \epsilon^{(1)}_{j_2,d}\Big)\right] + \phi^{(1)}_{j_1,d}(t_k)\phi^{(2)}_{j_2,d}(t_l)\,\mathbb{E}\left[\xi^{(1)}_{j_1,d}\Big(\rho\xi^{(2)}_{j_1,d} + \epsilon^{(2)}_{j_2,d}\Big)\right] \\
&\quad + \phi^{(2)}_{j_1,d}(t_k)\phi^{(1)}_{j_2,d}(t_l)\,\mathbb{E}\left[\xi^{(2)}_{j_1,d}\Big(\rho\xi^{(1)}_{j_1,d} + \epsilon^{(1)}_{j_2,d}\Big)\right] + \phi^{(2)}_{j_1,d}(t_k)\phi^{(2)}_{j_2,d}(t_l)\,\mathbb{E}\left[\xi^{(2)}_{j_1,d}\Big(\rho\xi^{(2)}_{j_1,d} + \epsilon^{(2)}_{j_2,d}\Big)\right] \\
&= \phi^{(1)}_{j_1,d}(t_k)\phi^{(1)}_{j_2,d}(t_l)\Big(\mathbb{E}[\rho\xi^{(1)2}_{j_1,d}] + \mathbb{E}[\xi^{(1)}_{j_1,d}\epsilon^{(1)}_{j_2,d}]\Big) + \phi^{(1)}_{j_1,d}(t_k)\phi^{(2)}_{j_2,d}(t_l)\Big(\mathbb{E}[\rho\xi^{(2)}_{j_1,d}\xi^{(1)}_{j_1,d}] + \mathbb{E}[\xi^{(1)}_{j_1,d}\epsilon^{(2)}_{j_2,d}]\Big) \\
&\quad + \phi^{(2)}_{j_1,d}(t_k)\phi^{(1)}_{j_2,d}(t_l)\Big(\mathbb{E}[\rho\xi^{(1)}_{j_1,d}\xi^{(2)}_{j_1,d}] + \mathbb{E}[\xi^{(2)}_{j_1,d}\epsilon^{(1)}_{j_2,d}]\Big) + \phi^{(2)}_{j_1,d}(t_k)\phi^{(2)}_{j_2,d}(t_l)\Big(\mathbb{E}[\rho\xi^{(2)2}_{j_1,d}] + \mathbb{E}[\xi^{(2)}_{j_1,d}\epsilon^{(2)}_{j_2,d}]\Big) \\
&\qquad \text{- At this point we assume } \xi \perp\!\!\!\perp \epsilon \text{ -} \\
&= \phi^{(1)}_{j_1,d}(t_k)\phi^{(1)}_{j_2,d}(t_l)\Big(\mathbb{E}[\rho\xi^{(1)2}_{j_1,d}] + \mathbb{E}[\xi^{(1)}_{j_1,d}]\,\mathbb{E}[\epsilon^{(1)}_{j_2,d}]\Big) + \phi^{(1)}_{j_1,d}(t_k)\phi^{(2)}_{j_2,d}(t_l)\Big(\mathbb{E}[\rho\xi^{(2)}_{j_1,d}\xi^{(1)}_{j_1,d}] + \mathbb{E}[\xi^{(1)}_{j_1,d}]\,\mathbb{E}[\epsilon^{(2)}_{j_2,d}]\Big) \\
&\quad + \phi^{(2)}_{j_1,d}(t_k)\phi^{(1)}_{j_2,d}(t_l)\Big(\mathbb{E}[\rho\xi^{(1)}_{j_1,d}\xi^{(2)}_{j_1,d}] + \mathbb{E}[\xi^{(2)}_{j_1,d}]\,\mathbb{E}[\epsilon^{(1)}_{j_2,d}]\Big) + \phi^{(2)}_{j_1,d}(t_k)\phi^{(2)}_{j_2,d}(t_l)\Big(\mathbb{E}[\rho\xi^{(2)2}_{j_1,d}] + \mathbb{E}[\xi^{(2)}_{j_1,d}]\,\mathbb{E}[\epsilon^{(2)}_{j_2,d}]\Big) \\
&= \rho\phi^{(1)}_{j_1,d}(t_k)\phi^{(1)}_{j_2,d}(t_l)\,\mathbb{E}[\xi^{(1)2}_{j_1,d}] + \rho\phi^{(1)}_{j_1,d}(t_k)\phi^{(2)}_{j_2,d}(t_l)\,\mathbb{E}[\xi^{(2)}_{j_1,d}\xi^{(1)}_{j_1,d}] + \rho\phi^{(2)}_{j_1,d}(t_k)\phi^{(1)}_{j_2,d}(t_l)\,\mathbb{E}[\xi^{(1)}_{j_1,d}\xi^{(2)}_{j_1,d}] \\
&\quad + \rho\phi^{(2)}_{j_1,d}(t_k)\phi^{(2)}_{j_2,d}(t_l)\,\mathbb{E}[\xi^{(2)2}_{j_1,d}] \qquad \Big(\text{by (E.15)}\Big) \\
&\qquad \text{- At this point we assume } \xi^{(1)}_{j_1,d} \perp\!\!\!\perp \xi^{(2)}_{j_1,d} \text{ -} \\
&= \rho\phi^{(1)}_{j_1,d}(t_k)\phi^{(1)}_{j_2,d}(t_l)\,\mathbb{E}[\xi^{(1)2}_{j_1,d}] + \rho\phi^{(1)}_{j_1,d}(t_k)\phi^{(2)}_{j_2,d}(t_l)\,\mathbb{E}[\xi^{(2)}_{j_1,d}]\,\mathbb{E}[\xi^{(1)}_{j_1,d}] \\
&\quad + \rho\phi^{(2)}_{j_1,d}(t_k)\phi^{(1)}_{j_2,d}(t_l)\,\mathbb{E}[\xi^{(1)}_{j_1,d}]\,\mathbb{E}[\xi^{(2)}_{j_1,d}] + \rho\phi^{(2)}_{j_1,d}(t_k)\phi^{(2)}_{j_2,d}(t_l)\,\mathbb{E}[\xi^{(2)2}_{j_1,d}] \qquad \Big(\text{by (E.15)}\Big) \\
&= \rho\Big(\phi^{(1)}_{j_1,d}(t_k)\phi^{(1)}_{j_2,d}(t_l)\,\mathbb{E}[\xi^{(1)2}_{j_1,d}] + \phi^{(2)}_{j_1,d}(t_k)\phi^{(2)}_{j_2,d}(t_l)\,\mathbb{E}[\xi^{(2)2}_{j_1,d}]\Big) \qquad \Big(\text{by (E.14)}\Big) \\
&= \rho\Big(\phi^{(1)}_{j_1,d}(t_k)\phi^{(1)}_{j_2,d}(t_l)Var[\xi^{(1)}_{j_1,d}] + \phi^{(2)}_{j_1,d}(t_k)\phi^{(2)}_{j_2,d}(t_l)Var[\xi^{(2)}_{j_1,d}]\Big)
\end{aligned}
$$

where the properties $Var[\xi^{(1)}_{j_1,d}] = \mathbb{E}[\xi^{(1)2}_{j_1,d}] - (\mathbb{E}[\xi^{(1)}_{j_1,d}])^2$ and $(\mathbb{E}[\xi^{(1)}_{j_1,d}])^2 = 0$ are utilised in the last equation. The same holds for $\xi^{(1)}_{j_2,d}$ too. Using the properties in (E.14), we conclude that:

$$
Cov\Big(X_{j_1,d}(t_k), X_{j_2,d}(t_l)\Big) = \rho\Big(\lambda_1\phi^{(1)}_{j_1,d}(t_k)\phi^{(1)}_{j_2,d}(t_l) + \lambda_2\phi^{(2)}_{j_1,d}(t_k)\phi^{(2)}_{j_2,d}(t_l)\Big).
$$

Furthermore, we have derived above that:

$$
X_{j_1,d}(t_i) \sim N\Big((\mu_{j_1,d}(t_i), \lambda_1\phi^{(1)2}_{j_1,d}(t_i) + \lambda_2\phi^{(2)2}_{j_1,d}(t_i)\Big),
$$

$$X_{j_2,d}(t_i) \sim N\Big(\mu_{j_2,d}(t_i), \rho^2\Big(\lambda_1\phi^{(1)2}_{j_2,d}(t_i) - \phi^{(1)2}_{j_2,d}(t_i) + \lambda_2\phi^{(2)2}_{j_2,d}(t_i) - \phi^{(2)2}_{j_2,d}(t_i)\Big) + \phi^{(1)2}_{j_2,d}(t_i) + \phi^{(2)2}_{j_2,d}(t_i)\Big),$$

so we have all components for the correlation coefficient:

$$\begin{aligned}
Cov\Big(X_{j_1,d}(t_k), X_{j_2,d}(t_l)\Big) &= \rho\Big(\lambda_1\phi^{(1)}_{j_1,d}(t_k)\phi^{(1)}_{j_2,d}(t_l) + \lambda_2\phi^{(2)}_{j_1,d}(t_k)\phi^{(2)}_{j_2,d}(t_l)\Big),\\
Var[X_{j_1,d}(t_k)] &= \lambda_1\phi^{(1)2}_{j_1,d}(t_k) + \lambda_2\phi^{(2)2}_{j_1,d}(t_k),\\
Var[X_{j_2,d}(t_l)] &= \rho^2\Big(\lambda_1\phi^{(1)2}_{j_2,d}(t_l) - \phi^{(1)2}_{j_2,d}(t_l) + \lambda_2\phi^{(2)2}_{j_2,d}(t_l) - \phi^{(2)2}_{j_2,d}(t_l)\Big) + \phi^{(1)2}_{j_2,d}(t_l) + \phi^{(2)2}_{j_2,d}(t_l).
\end{aligned}$$

E.5.2 Uncorrelated Variables

Now, we assume a second variable $X_{j_2,d}(t_i)$ with form:

$$X_{j_2,d}(t_i) = \mu_{j_2,d}(t_i) + \xi^{(1)}_{j_2,d}\phi^{(1)}_{j_2,d}(t_i) + \xi^{(2)}_{j_2,d}\phi^{(2)}_{j_2,d}(t_i)$$

uncorrelated with the variable $X_{j_1,d}(t_i)$. Following the same steps as for the derivation of the properties of variable $X_{j_1,d}(t_i)$, then:

$$\xi^{(1)}_{j_2,d} \sim N(0,\lambda_1), \quad \xi^{(2)}_{j_2,d} \sim N(0,\lambda_2) \tag{E.17}$$

and $X_{j_2,d}(t_i) \sim N\Big(\mu_{j_2,d}(t_i), \lambda_1\phi^{(1)2}_{j_2,d}(t_i) + \lambda_2\phi^{(2)2}_{j_2,d}(t_i)\Big)$. Then, the covariance between $X_{j_1,d}(t_k)$ and $X_{j_2,d}(t_l)$ has the following form:

$$\begin{aligned}
&Cov\Big(X_{j_1,d}(t_k), X_{j_2,d}(t_l)\Big) =\\
&= \mathbb{E}[(X_{j_1,d}(t_k) - \mathbb{E}[X_{j_1,d}(t_k)])(X_{j_2,d}(t_l) - \mathbb{E}[X_{j_2,d}(t_l)])]\\
&= \mathbb{E}[(X_{j_1,d}(t_k) - \mu_{j_1,d}(t_k))(X_{j_2,d}(t_l) - \mu_{j_2,d}(t_l))]\\
&= \mathbb{E}[X_{j_1,d}(t_k)X_{j_2,d}(t_l)] - \mu_{j_2,d}(t_l)\mu_{j_1,d}(t_k)\\
&= \mathbb{E}\left[\Big(\mu_{j_1,d}(t_k) + \xi^{(1)}_{j_1,d}\phi^{(1)}_{j_1,d}(t_k) + \xi^{(2)}_{j_1,d}\phi^{(2)}_{j_1,d}(t_k)\Big)\Big(\mu_{j_2,d}(t_l) + \xi^{(1)}_{j_2,d}\phi^{(1)}_{j_2,d}(t_l) + \xi^{(2)}_{j_2,d}\phi^{(2)}_{j_2,d}(t_l)\Big)\right]
\end{aligned}$$

$$
\begin{aligned}
&\quad - \mu_{j_2,d}(t_l)\mu_{j_1,d}(t_k) \\
&= \phi^{(1)}_{j_1,d}(t_k)\phi^{(1)}_{j_2,d}(t_l)\,\mathbb{E}[\xi^{(1)}_{j_1,d}\xi^{(1)}_{j_2,d}] + \phi^{(1)}_{j_1,d}(t_k)\phi^{(2)}_{j_2,d}(t_l)\,\mathbb{E}[\xi^{(1)}_{j_1,d}\xi^{(2)}_{j_2,d}] + \phi^{(2)}_{j_1,d}(t_k)\phi^{(1)}_{j_2,d}(t_l)\,\mathbb{E}[\xi^{(2)}_{j_1,d}\xi^{(1)}_{j_2,d}] \\
&\quad + \phi^{(2)}_{j_1,d}(t_k)\phi^{(2)}_{j_2,d}(t_l)\,\mathbb{E}[\xi^{(2)}_{j_1,d}\xi^{(2)}_{j_2,d}] \qquad \Big(\text{by (E.14) and (E.16)}\Big) \\
&\text{- At this point we assume } \xi^{(1)}_{j_1,d} \perp\!\!\!\perp \xi^{(2)}_{j_1,d},\ \xi^{(1)}_{j_1,d} \perp\!\!\!\perp \xi^{(1)}_{j_2,d},\ \xi^{(1)}_{j_1,d} \perp\!\!\!\perp \xi^{(2)}_{j_2,d},\ \xi^{(2)}_{j_2,d} \perp\!\!\!\perp \xi^{(2)}_{j_2,d} \text{ -} \\
&= \phi^{(1)}_{j_1,d}(t_k)\phi^{(1)}_{j_2,d}(t_l)\,\mathbb{E}[\xi^{(1)}_{j_1,d}]\,\mathbb{E}[\xi^{(1)}_{j_2,d}] + \phi^{(1)}_{j_1,d}(t_k)\phi^{(2)}_{j_2,d}(t_l)\,\mathbb{E}[\xi^{(1)}_{j_1,d}]\,\mathbb{E}[\xi^{(2)}_{j_2,d}] \\
&\quad + \phi^{(2)}_{j_1,d}(t_k)\phi^{(1)}_{j_2,d}(t_l)\,\mathbb{E}[\xi^{(2)}_{j_1,d}]\,\mathbb{E}[\xi^{(1)}_{j_2,d}] + \phi^{(2)}_{j_1,d}(t_k)\phi^{(2)}_{j_2,d}(t_l)\,\mathbb{E}[\xi^{(2)}_{j_1,d}]\,\mathbb{E}[\xi^{(2)}_{j_2,d}] \\
&= 0. \qquad \Big(\text{by (E.14) and (E.17)}\Big)
\end{aligned}
$$

In addition, $X_{j_1,d}(t_i)$ and $X_{j_2,d}(t_i)$ have the following distribution properties:

$$
\begin{aligned}
X_{j_1,d}(t_i) &\sim N\Big(\mu_{j_1,d}(t_i), \lambda_1\phi^{(1)2}_{j_1,d}(t_i) + \lambda_2\phi^{(2)2}_{j_1,d}(t_i)\Big), \\
X_{j_2,d}(t_i) &\sim N\Big(\mu_{j_2,d}(t_i), \lambda_1\phi^{(1)2}_{j_2,d}(t_i) + \lambda_2\phi^{(2)2}_{j_2,d}(t_i)\Big),
\end{aligned}
$$

so we end up with a correlation coefficient equal to zero where:

$$
\begin{aligned}
Cov\Big(X_{j_1,d}(t_k), X_{j_2,d}(t_l)\Big) &= 0, \\
Var[X_{j_1,d}(t_k)] &= \lambda_1\phi^{(1)2}_{j_1,d}(t_k) + \lambda_2\phi^{(2)2}_{j_1,d}(t_k), \\
Var[X_{j_2,d}(t_l)] &= \lambda_1\phi^{(1)2}_{j_2,d}(t_l) + \lambda_2\phi^{(2)2}_{j_2,d}(t_l).
\end{aligned}
$$

E.6 Empirical Analysis Results

This section provides the findings not reported in the empirical analysis of section 5.5. In particular, the fully conditionally independent stocks for eight different tuning parameter values are plotted using log-prices (see appendix E.6.1) and returns (see appendix E.6.2) as data. In addition, for the case of returns (in appendix E.6.2) some indicative maximal cliques are given, as formed for the estimated optimal tuning parameter.

E.6.1 Log-Prices

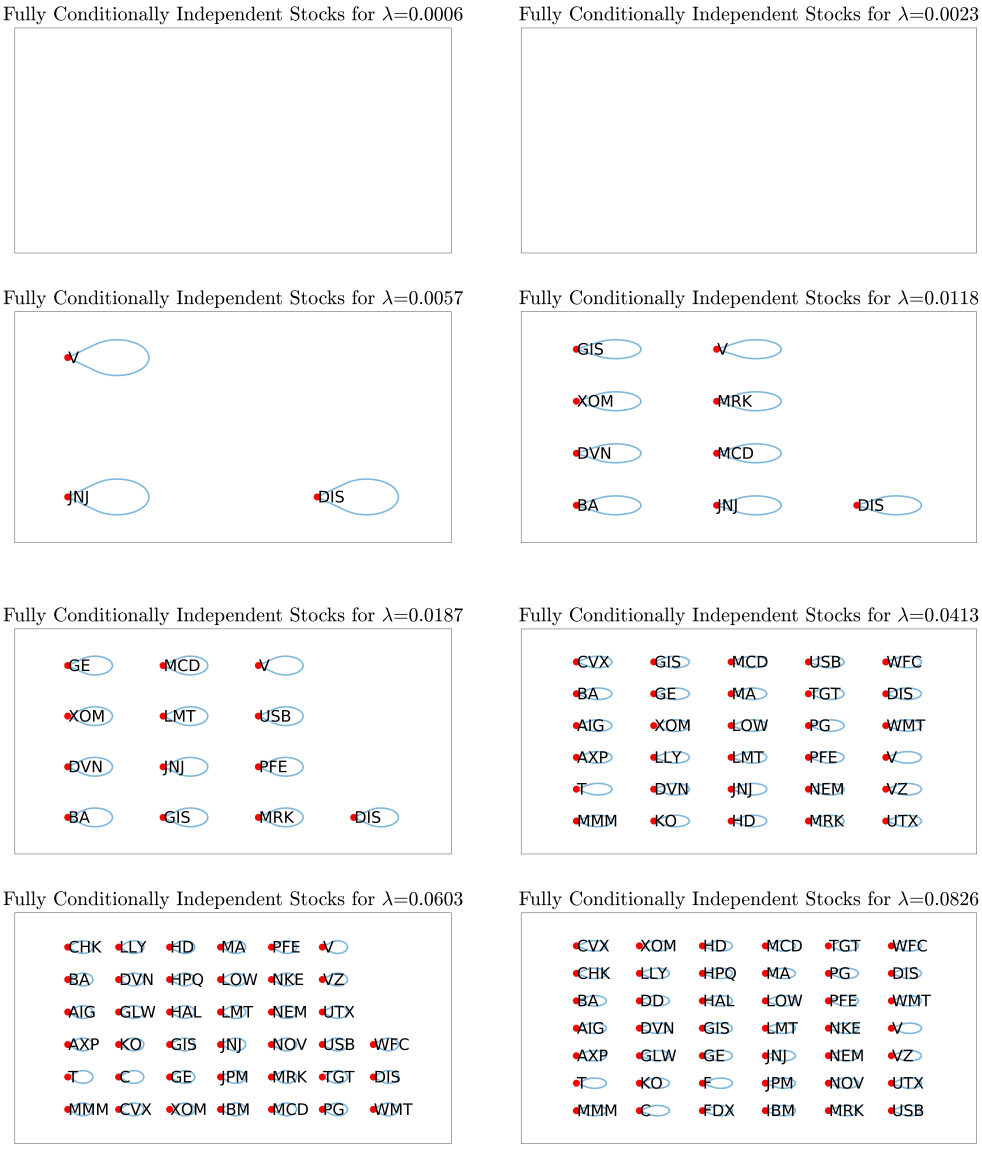

Figure E.2: This figure shows the fully conditionally independent stocks using log-prices for eight different tuning parameter values.

E.6.2 Returns

Figure E.3: This figure shows the fully conditionally independent stocks using returns for eight different tuning parameter values.

Figure E.4: This figure shows some indicative cases of maximal cliques as presented in section 5.5.

Abbreviations - Acronyms

AIC : Akaike Information Criterion

BIC : Bayesian Information Criterion

CV : Cross-Validation

FGLASSO : Functional Graphical Least Absolute Shrinkage and Selection Operator

GLASSO : Graphical Least Absolute Shrinkage and Selection Operator

FPR : False Positive Rate

IID : Independent and Identically Distributed

IV : Integrated Variance

LASSO : Least Absolute Shrinkage and Selection Operator

LPR : Local Polynomial Regression

MISE : Mean Integrated Squared Error

MSE : Mean Squared Error

MRV : Modulated Realised Variance

NYSE : New York Stock Exchange

OLS : Ordinary Least Squares

PA : Pre-Averaging Approach

PCA : Principal Component Analysis

PDF : Probability Density Function

QV : Quadratic Variation

RK : Realised Kernel Estimator

RSS : Residual Sum of Squares

RV : Realised Variance

SDE : Stochastic Differential Equation

SVF1 : Stochastic Volatility - Factor One Model

TPR : True Positive Rate

TSRV : Two Scale RV

WN : White Noise