Essays in Volatility Estimation Based on High Frequency Data

Autor: Yucheng Sun

TESI DOCTORAL UPF / ANY 2017

DIRECTOR DE LA TESI
Prof. Christian Brownlees and Prof. Eulalia Nualart
Departament d'Economia i Empresa

This thesis is dedicated to Jianshe, Shuhua and Shujuan.

Acknowledgements

Pursuing a Ph.D degree at Universitat Pompeu Fabra is a wonderful journey in my life. I have learned much and gained mental strength from many people in this process. Without their generous support, it would have been extremely difficult for me to complete my studies.

First of all, no words can fully express my gratitude to my advisors, Professor Christian Brownlees and Professor Eulalia Nualart. When you agreed to guide me during my Ph.D studies, I was quite naive in almost all aspects of academic research. It is your years of patient instructions that helps me grow. I am grateful to every effort you have made in order to turn me into a qualified researcher. Besides regular academic supervision, you calm me down when I feel anxious, and encourage me when I am depressed. From you I can see the right attitudes and methodology of performing academic research and educating students, which I firmly believe will benefit me all my life.

I am also thankful to Professors Gabor Lugosi and Majid Al-Sadoon, who have given me valuable advice on my research. Majid's course has given me an introduction to asymptotic theory for statisticians, which is extremely important in my thesis. Gabor's course has promoted my interest and ability in mathematical analysis which also occupies much space in the thesis. I would like to thank Professors Barbara Rossi, Javier Gil-Bazo, Piotr Zwiernik, Mihalis Markakis and Filippo Ippolito as well, since your comments have also contributed to a better thesis.

I also owe many thanks to my Ph.D colleagues, especially Jianxing Wei, Haozhou Tang, Rui Guan, Tongbin Zhang, Guohao Yang, Shangyu Liu, Erqi Ge, Yiru Wang, Lang Zhou. Being together with you is not only good for my academic research, you have also improved my ability on various subjects like cooking and playing card games. I also want to thank Xuanyu Ouyang and Yu Zhang for sharing your colorful stories which are an indispensable pleasure in my life, in spite of the long distance between us. I am also grateful to Marta and Laura, as your daily work provides great convenience to economics Ph.D students at UPF.

Needless to say, I owe too much to my parents, and there have been few chances for me to repay you during these years. Hopefully I will have more time to be around and take care of you after graduation, which is the best way to express my gratitude.

I also want to tell Shujuan that an important source that motivates me to move on is the fear of failing you. When faced with difficulties and feeling depressed, I always recall the happy

time that we have together. I am willing to make whatever large efforts in order to live up with your expectation. I can never thank you enough for providing me with the strength to be a better self.

Abstract

Based on high-frequency price data, this thesis focuses on estimating the realized covariance and the integrated volatility of asset prices, and applying volatility estimation to price jump detection. The first chapter uses the LASSO procedure to regularize some estimators of high dimensional realized covariance matrices. We establish theoretical properties of the regularized estimators that show its estimation precision and the probability that they correctly reveal the network structure of the assets. The second chapter proposes a novel estimator of the integrated volatility which is the quadratic variation of the continuous part in the price process. This estimator is obtained by truncating the two-scales realized variance estimator. We show its consistency in the presence of market microstructure noise and finite or infinite activity jumps in the price process. The third chapter employs this estimator to design a test to explore the existence of price jumps with noisy price data.

Resum

Basándonos en datos de precios de alta frecuencia, esta tesis se centra en la estimación de la covarianza realizada y la volatilidad integrada de precios de activos, y la aplicación de la estimación de la volatilidad para la detección de saltos en los precios. El primer capítulo utiliza el procedimiento LASSO para regularizar algunos estimadores de matrices de covarianza realizada de alta dimensión. Establecemos propiedades teóricas de los estimadores regularizados que muestran su precisión de estimación y la probabilidad de que revelen correctamente la estructura de red de los activos. En el segundo capítulo se propone un nuevo estimador de la volatilidad integrada que es la variación cuadrática de la parte continua en el proceso de precios. Este estimador se obtiene truncando el estimador de varianza realizado en dos escalas. Demostramos su consistencia en presencia de ruido de microestructura del mercado y saltos de actividad finitos o infinitos en el proceso de precios. El tercer capítulo emplea este estimador para diseñar un test para explorar la existencia de saltos en los precios con ruido.

Preface

There is strong evidence from literatures showing that asset price volatilities and jumps in the price processes play important roles in financial economics. Accordingly, many econometric papers focus on estimating price volatilities and detecting the arrivals of price jumps. On the other hand, the availability of intra-daily high-frequency price data has contributed much to the development of financial econometrics over the last decade. Based on high-frequency price data, this thesis proposes new methodologies to measure price volatilities and test for the arrivals of price jumps. The thesis also takes the pains to deal with the market microstructure noise which is quite common in financial markets and makes observed prices different from the efficient ones.

The first and second chapters are joint papers with my advisors, Christian Brownlees and Eulalia Nualart. In the first chapter, we apply the LASSO-type shrinkage to regularize some estimators on the covariance matrix of the asset prices. The entries of the unregularized matrix estimators are volatility estimators derived from high-frequency price data, which are commonly referred to as realized volatility estimators. The regularized matrix is supposed to be an estimator on the inverse covariance matrix. The regularization can shrink some entries to zero, and zero entries of the inverse covariance matrix indicate zero partial correlations. The regularized estimator can help us recover the network structure of the assets, as the linkage between two assets is defined as non-zero partial correlation of their prices. For the regularized estimators, we establish theories concerning their estimation errors and the probability that they recover the network structure of the asset prices. Simulations and empirical applications indicate that when the partial correlation structure of the prices is sparse, the regularization can significantly improve the estimation precision.

There is a large segment of literatures trying to disentangle the quadratic variation of the continuous part in the price process, which is the integrated volatility, from the total quadratic variation when there are price jumps. This problem would be more complicated when there is market microstructure noise. The second chapter proposes a truncated volatility estimator in order to deal with this challenge. This estimator is derived by truncating the two-scales realized volatility (TSRV) estimator which has been widely used as an integrated volatility estimator robust to the noise. The truncation removes the intervals where jumps occur from the estimator, so that the convergence of the truncated estimator is not affected by jumps. Accordingly,

we show that the asymptotic distribution of the TSRV is not affected by finite activity jumps. Moreover, we prove the consistency of our estimator when there are infinite activity jumps.

In Chapter 3, we adopt the truncated volatility estimator derived in the second chapter to design a test for the arrivals of jumps. We divide the whole period under consideration into many intervals, and test for each interval, whether there is a jump on it or not. The length of each interval converges to zero as the sampling frequency increases. Therefore, besides indicating the existence of jumps on the whole period, this test can also give us inference about the time when jumps occur. The statistic for a given interval is in the form of a ratio: Its nominator is the rescaled local average return, and the denominator is a spot volatility estimator which is constructed based on the proposed estimator in Chapter 2. A jump can significantly increase the absolute value of the corresponding statistic. Thus the test is such that we believe there is a jump on the interval if the statistic is larger than some threshold in absolute value. We also derive the theory to show that the test is consistent, and perform simulations to evaluate the efficiency of our test.

Contents

List of Figures

List of Tables

Introduction

Volatility estimation based on high-frequency price data is an important topic in financial econometrics, and the estimators are typically called realized volatility estimators. In this thesis we propose novel realized volatility estimators, and apply a proposed estimator to price jump detection. The first chapter regularizes some covariance matrix estimators of asset prices, whose entries are realized volatility estimators, with the LASSO-type shrinkage. We show that the regularization can improve estimation precision when the network of the assets is sparse. The second chapter assumes the presence of jumps in the price process, and derives a truncated realized volatility estimator on the volatility of the continuous component of the price process. The third chapter adopts the proposed estimator in the second chapter in order to detect jumps. Throughout the thesis, our methodology is robust to market microstructure noise which is prevalent in financial markets. Moreover, the papers which yield the first and second chapters are respectively submitted for the Journal of Applied Econometrics and Scandinavian Journal of Statistics, and the paper which yields the third chapter is not submitted yet.

In Chapter 1 we adopt the Graphical LASSO procedure (Friedman et al. (2011)) in order to regularize the covariance matrix estimators whose entries are two-scales realized volatilities (Zhang et al. (2005)) or the realized kernel estimators (Barndorff-Nielsen et al. (2008), Varneskov (2016)). The regularized matrix directly estimates the inverse covariance matrix, and it can also recover the network structure of the assets. We derive the theory which shows estimation precision and the probability of successful recovery on the graph structure. An empirical application shows that this regularized estimator can deliver global minimum variance portfolio that has the smallest average out-of-sample variance compared to portfolios yielded by some other covariance matrix estimators.

Chapter 2 proposes a truncated realized volatility estimator by applying the truncation technique in Mancini (2008, 2009) to the two-scales realized variance (TSRV) estimator in Zhang et al. (2005). In the finite activity jump case, the truncated TSRV has the same asymptotic

distribution as the TSRV which is constructed with the no-jump data, since the truncation removes intervals that may contain jumps from the estimator. We also prove the consistency of the truncated TSRV when jumps have infinite activity.

Chapter 3 adopts the truncated TSRV in order to test the arrivals of jumps. The statistics are derived by standardizing the rescaled local average returns with the spot volatility estimator which is constructed based on the truncated TSRV. As a jump can significantly increase the value of the statistic, we show that when the threshold is set properly, the probabilities of both type I and type II will converge to zero as the number of price observations increases. Moreover, in both Chapters 2 and 3, we perform simulations to evaluate the efficiency of our methodology and illustrate its advantage compared to some other competitors.

Chapter 1

REALIZED NETWORKS

1.1 Introduction

The covariance matrix of the log–prices of financial assets is a fundamental ingredient in many applications ranging from asset allocation to risk management. For more than a decade now the econometric literature has made a number of significant leaps forward in the estimation of covariance matrices using financial high frequency data. This new generation of estimators, commonly referred to as realized covariance estimators, measure precisely the daily covariance of log–prices using intra-daily price information. The literature has proposed an extensive number of procedures that allow to estimate the covariance efficiently under general assumptions, such as the presence of market microstructure noise and asynchronous trading in the data generating process.

Despite the significant leaps forward, the estimation of large realized covariance matrices has a number of hurdles. First, as it has been put forward by Hautsch et al. (2012) and Nikolaus et al. (2013), it is hard to estimate precisely the covariance matrix when the number of assets is large. Second, in large systems it is challenging to synthesize effectively the information contained in the covariance matrix and unveil the cross–sectional dependence structure of the data. In this chapter we propose a realized covariance estimation strategy that tackles simultaneously both of these challenges. The estimation approach consists of using LASSO–type shrinkage to regularize realized covariance estimators. The LASSO procedure detects and estimates the nonzero partial correlations among the daily log–prices. The set of nonzero partial correlations can then be represented as a network. Our proposed estimation approach has different high-

lights. If the partial correlation structure of the daily log–prices is sufficiently sparse, then the regularized estimator can deliver substantial accuracy gains over its unregularized counterpart. Moreover, detecting the network of interconnections among the daily log prices is interesting in the light of the recent strand of research on networks in economics by, among others, Acemoglu et al. (2012), which shows that in highly interconnected systems the most highly interconnected entities influence the aggregate behaviour of the entire system.

In its more general version, the framework we work in makes a number of fairly common assumptions on the dynamics of the asset prices (cf Bandi and Russell, 2006; Aït-Sahalia et al., 2005; Fan et al., 2012). We assume that observed log prices are equal to the efficient log prices, which are Brownian semi–martingales, plus a noise term that is due to market microstructure frictions. Prices are observed according to the realization of a counting process driving the arrival of trades/quotes of each asset and are allowed to be asynchronous. The target estimation of interest is integrated covariance matrix of the efficient daily log–prices.

We introduce a network definition built upon the integrated covariance, which we call the integrated partial correlation network. Assets i and j are connected in the integrated partial correlation network if and only if the partial correlation between i and j implied by the integrated covariance is nonzero. As is well known, the network is entirely characterized by the inverse of the integrated covariance matrix, which we call hereafter the integrated concentration matrix. In fact, it has been known since at least Dempster (1972) that if the (i, j)–th entry of the inverse covariance matrix is zero, then variables i and j are partially uncorrelated, that is, are uncorrelated conditional on all other assets. Thus, the sparsity structure of the integrated concentration matrix determines the partial correlation network dependence structure among the daily log–prices.

We use LASSO to obtain a sparse integrated concentration matrix estimator. The procedure consists of regularizing a consistent realized covariance estimator. Several realized covariance estimators have been introduced in the literature in the presence of market microstructure effects and asynchronous trading. In this chapter we focus in particular on the Two–Scales Realized Covariance estimators (TSRC) and the Multivariate Realized Kernel (MRK) based on pairwise refresh sampling (Aït-Sahalia et al., 2005; Barndorff-Nielsen et al., 2011; Fan et al., 2012). These estimators are then regularized using the GLASSO (Friedman et al., 2011); which shrinks the off–diagonal elements of the inverse of the covariance estimators to zero. The procedure allows to detect the nonzero linkages of the integrated partial correlation network. Moreover,

the sparse integrated concentration matrix estimator can be inverted to obtain an estimator of the integrated covariance.

We study the large sample properties of the realized network estimator, and establish conditions for consistent estimation of the integrated concentration and consistent selection of the integrated partial correlation network. We develop the theory for the TSRC and MRK estimators based on pairwise refresh–sampling built upon the general asymptotic theory developed by Ravikumar et al. (2011). The MRK estimator results are obtained by developing a novel concentration inequality, while for the TSRC estimator we apply a concentration inequality derived in Fan et al. (2012). Results are established in a high–dimensional setting, that is, we allow for the total number of parameters to be larger than the amount of observations available to the extent that the proportion of nonzero parameters is small relative to the total. Other realized covariance estimators satisfying an appropriate concentration assumption lead to regularized estimators with similar properties.

A simulation study is used to investigate the finite sample properties of the procedure. Different specifications of the integrated covariance matrix of the efficient price process are used to assess the precision of the realized network estimator. The procedure is also benchmarked against a set of alternative regularization techniques proposed in the literature, including shrinkage (Ledoit and Wolf, 2004) and factor based approaches. Among others results, simulations show that when the integrated concentration matrix is indeed sparse the realized network achieves the best performance among the set of candidate regularization procedures we consider.

We apply the realized network methodology to analyse the network structure of a panel of US blue chip stocks throughout 2009 using the TSRC, MRK as well as the classic Realized Covariance (RC) estimators. More precisely, we use the realized network to regularize what we call idiosyncratic realized covariance matrix, that is the residual covariance matrix of the assets after netting out the influence of the market factor. Results show that after controlling for the market factor, assets still exhibit a significant amount of cross–sectional dependence. The estimated networks are indeed sparse, with the number of estimated links being roughly 5% of the total possible number of linkages. The distribution of the connections of the assets exhibits power law behavior, that is the number of connections is heterogeneous and the most interconnected stocks have a large number of connections relative to the total number of links. The stocks in the industrial and energy sectors show a high degree of sectoral clustering, that

is there is a large number of connections among firms in these industry groups. Technology companies and Google in particular are the most highly interconnected firms throughout the year. We investigate the usefulness of our procedure from a forecasting perspective by carrying out a Markowitz type Global Minimum Variance (GMV) portfolio prediction exercise. We run a horse race among different (regularized) covariance estimators to assess which estimator produces GMV portfolio weights that deliver the minimum out–of–sample GMV portfolio variance. Results show that the realized network significantly improves prediction accuracy irrespective of the covariance estimator used.

We build upon the literature on realized volatility and realized covariance estimation. Important contributions in this area include the work of Andersen et al. (2003), Barndorff-Nielsen and Shephard (2004a), Aït-Sahalia et al. (2005), Bandi and Russell (2006), Barndorff-Nielsen et al. (2008), Barndorff-Nielsen et al. (2011), Zhang (2011), and Fan et al. (2012). More precisely, this chapter is related to the strand of the literature concerned with the estimation and regularization of possibly large realized covariances. Important research in this field includes Wang and Zou (2010), Hautsch et al. (2012), Nikolaus et al. (2013), Tao et al. (2013), Corsi et al. (2014), Lunde et al. (2011). This chapter also relates on the network modelling literature in statistics and econometrics, which includes Meinshausen and Bühlmann (2006), Billio et al. (2012), Diebold and Yilmaz (2014), Hautsch et al. (2014a,b), Barigozzi and Brownlees (2013), Banerjee and Ghaoui (2008). Last, this chapter is related to the literature on covariance matrix regularization. Contributions in this area include the work of Ledoit and Wolf (2004), Ledoit and Wolf (2012), Fan et al. (2011) and Fan et al. (2013). Pourahmadi (2013) provides an introduction to high–dimensional covariance regularization, which includes several of the recent developments of the area.

It is important to observe that in the high-dimensional realized covariance estimation literatures such as Wang and Zou (2010), Tao et al. (2013) and Kim et al. (2016), the sparsity assumption is imposed directly to the integrated covariance matrix. While in this chapter, we assume it on the inverse. Moreover, they use the thresholding technique in order to regularize the realized covariance matrix, while in this chapter we use the GLASSO technique. We remark that the thresholding technique used in Tao et al. (2013) applied to the realized covariance estimator yields an estimation error of the same order as ours, which is smaller than the estimation error obtained in Wang and Zou (2010) and Kim et al. (2016). In addition, we believe that one advantage of employing the GLASSO technique is that we recover the network struc-

ture without requiring zero correlations between asset prices. Instead, we require zero partial correlations, which is a much weaker condition than zero correlations when the sample size is large.

The rest of this chapter is structured as follows. In Section 1.2 we introduce the base framework and the realized network estimator. The theoretical properties of the estimation procedure are analysed in Section 1.3. Section 1.4 introduces a number of important extensions to the baseline framework. Section 1.5 contains a simulation exercise to study the properties of the realized network estimator. Section 1.6 presents an application to a panel of US blue chip stocks. Concluding remarks follow in Section 1.7.

1.2 Methodology

In this section we introduce the baseline framework and estimation approach. Important extensions of the baseline methodology, including allowing for market microstructure frictions, are considered later in Section 1.4.

1.2.1 Model

Let $y(t)$ denote the n-dimensional log–price vector of n assets at time t. We assume that the dynamics of $y(t)$ are given by

$$y(t) = \int_0^t b(u)\, du + \int_0^t \mathbf{\Theta}(u)\, dB(u)\ ,\ t \in [0,1], \tag{1.1}$$

where $B(t)$ is an n-dimensional Brownian motion on a filtered probability space $(\Omega, \mathcal{F}, \mathcal{F}_t, \mathrm{P})$. The drift $b(t)$ is an n-dimensional process, and the spot covolatility process $\mathbf{\Theta}(t)$ is an $n \times n$ positive definite random matrix. The entries of $b(t)$ and $\mathbf{\Theta}(t)$ are assumed to be adapted to $\mathcal{F}_t$ and uniformly bounded on $[0,1]$, and those of $\mathbf{\Theta}(t)$ are also assumed to be càdlàg. Like most of the realized covariance literature, we do not consider jumps in the process of $y(t)$.

We are considering our process over a fixed time interval of length 1, which typically represents a day. We set $y = y(1)$. Also, we define the i–th entry of y as y_i, for $i = 1, \ldots, n$. One of

the main objects of interest in this chapter is the integrated covariance matrix

$$\mathbf{\Sigma}^{\star} = \int_0^1 \mathbf{\Sigma}(t)dt = (\sigma_{ij}^{\star}), \tag{1.2}$$

where $\mathbf{\Sigma}(t) = \mathbf{\Theta}(t)\mathbf{\Theta}(t)' = (\sigma_{ij}(t))$ is the spot covariance matrix. Notice that when B is independent of $(b, \mathbf{\Theta})$, conditional on $(b, \mathbf{\Theta})$, $\mathbf{\Sigma}^{\star}$ is the covariance matrix of y.

In this chapter we introduce a network representation of y based on the partial linear dependence structure induced by the integrated covariance matrix. We call this network the integrated partial correlation network. Recall that a network is defined as an undirected graph $\mathcal{G} = (\mathcal{V}, \mathcal{E})$, where $\mathcal{V}$ is the set of vertices $\mathcal{V} = \{1, 2, \ldots, n\}$ and $\mathcal{E}$ is the set of edges $\mathcal{E} \subset \mathcal{V} \times \mathcal{V}$. In the integrated partial correlation network the set of vertices corresponds to the set of assets, and a pair of assets is connected by an edge iff their daily returns are partially correlated given all other daily returns in the panel, that is,

$$\mathcal{E} = \{(i, j) \in \mathcal{V} \times \mathcal{V}, \rho^{ij} \neq 0, i \neq j\} \ ,$$

where ρ^{ij} is the partial correlation, a measure of linear cross-sectional conditional dependence defined as

$$\rho^{ij} = \text{Cor}(y_i, y_j | \{y_k, \quad k \neq i, j\}) \ .$$

The partial correlations can be characterized by the inverse covariance matrix (Dempster, 1972; Lauritzen, 1996). That is,

$$\rho^{ij} = \frac{-k_{ij}^{\star}}{\sqrt{k_{ii}^{\star} k_{jj}^{\star}}} \ ,$$

where $\mathbf{K}^{\star} = (\mathbf{\Sigma}^{\star})^{-1} = (\mathbf{k}_{ij}^{\star})$. Recall that if B is independent of $(b, \mathbf{\Theta})$, conditional on $(b, \mathbf{\Theta})$, $\mathbf{K}^{\star}$ is the inverse integrated covariance matrix, which we call hereafter integrated concentration matrix. Thus, the integrated partial correlation network can be equivalently defined as

$$\mathcal{E} = \{(i, j) \in \mathcal{V} \times \mathcal{V}, k_{ij}^{\star} \neq 0, i \neq j\} \ .$$

It is important to emphasize that the integrated partial correlation network definition represents particular correlation relations among the daily log–prices. Obviously enough, the absence of correlation between the log–daily prices of two assets does not necessarily imply that the spot prices are also uncorrelated.

1.2.2 Estimation

We are interested in (i) estimating the integrated covariance and concentration matrices of the daily log–prices, and (ii) detecting the nonzero entries of the integrated concentration matrix. The estimation strategy we follow consists of applying LASSO type regularization on the standard realized covariance estimators proposed in the literature.

We assume that the log-prices $y_i(t)$ of all assets $i = 1, \ldots, n$, are discretely observed at a same time grid $T = \{t_1, t_2, \ldots, t_m\}$ where $t_0 = 0 < t_1 < \cdots < t_m = 1$. We consider a generic estimator of the integrated covariance $\mathbf{\Sigma}^\star$ denoted $\overline{\mathbf{\Sigma}} = (\overline{\sigma}_{ij})$ based on the observations $y_i(t_\ell)$, $i = 1, ..., n$, $\ell = 1, ..., m$. We assume that this estimator satisfies the following concentration inequality.

ASSUMPTION 1.1. *There exist positive constants a_1, a_2 and a_3 such that for all $i, j \in \{1, ..., n\}$, $x \in [0, a_1]$, and m large,*

$$\mathrm{P}\left(\left|\overline{\sigma}_{ij} - \sigma^\star_{ij}\right| > x\right) \leq a_2 m^{\alpha_0} \exp(-a_3 (m^\beta x)^a), \tag{1.3}$$

for some positive exponents β, a and $\alpha_0 \in \{0, 1\}$.

A natural estimator of the integrated covariance of y in this setting is the so called Realized Covariance (RC) estimator. This estimator is the multivariate extension of the realized variance, whose working mechanism is that the quadratic variation of the univariate price process can be approximated by the sum of squared returns over small intervals.

Realized Covariance Estimator. The realized covariance estimator $\overline{\mathbf{\Sigma}}_{\mathsf{RC}}$ is defined as

$$\overline{\sigma}_{\mathsf{RC},ij} = \sum_{k=1}^{m} \left(y_{i\,k} - y_{i\,k-1}\right)\left(y_{j\,k} - y_{j\,k-1}\right),$$

where $y_{i\,k} = y_i(t_k)$.

Assume that the time grid satisfies that

$$\sup_{\ell \in \{1,\ldots,m\}} |t_\ell - t_{\ell-1}| \leq \frac{c}{m}, \tag{1.4}$$

for some constant $c > 0$. Then Barndorff-Nielsen and Shephard (2004c) shows that the difference between $\overline{\sigma}_{\mathsf{RC},ij}$ and $\sigma^\star_{ij}$ is asymptotically mixed normal with mean zero and variance of

order $O\left(m^{-1}\right)$. Also, it is proved in Fan et al. (2012, Lemma 3), that under the conditions of this Section the estimator satisfies Assumption 1.1 with $\alpha_0 = 0, a = 2$ and $\beta = \frac{1}{2}$.

Given an estimator of the integrated covariance $\overline{\mathbf{\Sigma}}$ satisfing Assumption 1.1, we use the Graphical LASSO procedure (GLASSO) to estimate the integrated concentration matrix $\mathbf{K}^\star$.

Realized Network Estimator. Let $\overline{\mathbf{\Sigma}}$ be an estimator of the integrated covariance, then we define the realized network estimator of the integrated concentration matrix as

$$\widehat{\mathbf{K}}_\lambda = \arg\min_{\mathbf{K}\in\mathcal{S}^n} \left\{ \text{tr}(\overline{\mathbf{\Sigma}}\mathbf{K}) - \log\det(\mathbf{K}) + \lambda \sum_{i\neq j} |k_{ij}| \right\} , \tag{1.5}$$

where $\lambda \geq 0$ is the GLASSO tuning parameter and $\mathcal{S}^n$ is the set of $n \times n$ symmetric positive definite matrices. The entries of $\widehat{\mathbf{K}}_\lambda$ are denoted by $(\widehat{k}_{\lambda\, ij})$. The corresponding realized covariance estimator based on the realized network is $\widehat{\mathbf{\Sigma}}_\lambda = \widehat{\mathbf{K}}_\lambda^{-1}$.

Observe that (1.5) defines a shrinkage type estimator. If we set $\lambda = 0$ in (1.5), we obtain the normal log-likelihood function of the covariance matrix, which is minimized by the inverse realized covariance estimator $(\overline{\mathbf{\Sigma}})^{-1}$. If λ is positive, (1.5) becomes a penalized likelihood function with penalty equal to the sum of the absolute values of the non-diagonal entries in the estimator. The important feature of the absolute value penalty is that for $\lambda > 0$, some of the entries of the realized network estimator are going to be set to zero. The highlight of this estimator is that it simultaneously estimates and selects the nonzero entries of $\mathbf{K}^\star$, thus providing an estimate of the linkages in the network. For this reason we dub the estimator realized network estimator. Banerjee and Ghaoui (2008) show that the optimization problem in (1.5) can be solved through a series of LASSO regressions, which motivates an iterative algorithm to solve (1.5) given in Friedman et al. (2007). For completeness, we provide a description of the algorithm in the appendix. The highlight of the procedure is that it is straightforward to carry out the minimization of equation (1.5) even when the number of series is large. Importantly, the algorithm is also guaranteed to provide a positive definite estimate of the concentration matrix provided that the initial value of the algorithm is a positive definite matrix. Moreover, the algorithm only requires the $\overline{\mathbf{\Sigma}}$ estimator to be positive semi–definite (provided that λ is larger than zero).

In order to apply the estimator in empirical applications we need to use a selection criterion to pick the value of the tuning parameter λ. In this chapter we resort to a BIC–type criterion

defined as

$$\text{BIC}(\lambda) \;=\; m \times \left[-\log\det \widehat{\mathbf{K}}_\lambda + \text{tr}\left(\widehat{\mathbf{K}}_\lambda \overline{\mathbf{\Sigma}}\right)\right] + \log m \times \#\{(i,j) : 1 \leq i \leq j \leq n, \hat{k}_{\lambda\, ij} \neq 0\}\ ,$$

as it is suggested in, among others, Yuan and Lin (2007).

1.3 Theory

In this section, we apply the theory of Ravikumar et al. (2011) to our particular case of an exponential concentration inequality to establish the large sample properties of the realized network estimator defined in (1.5).

In order to state the results we need to adopt the following notations. Given a matrix $\mathbf{U} = (u_{ij}) \in \mathbb{R}^{\ell \times m}$, we set $||\mathbf{U}||_\infty, ||\mathbf{U}||_1$, and $|||\mathbf{U}|||_\infty$ to denote $\max\limits_{i,j}|u_{ij}|, \sum_{i,j}|u_{ij}|$, and $\max\limits_{j}\sum_{k=1}^{m}|u_{jk}|$, where $i \in \{1,2,\ldots,\ell\}$ and $j \in \{1,2,\ldots,m\}$. If $\mathbf{A} = (a_{ij})$ is a $p \times q$ matrix and $\mathbf{B}$ is an $m \times n$ matrix, the Kronecker product of matrices $\mathbf{A}$ and $\mathbf{B}$ is the $pm \times qn$ matrix given by

$$\mathbf{A} \otimes \mathbf{B} = \begin{bmatrix} a_{11}\mathbf{B} & \cdots & a_{1q}\mathbf{B} \\ \vdots & \ddots & \vdots \\ a_{p1}\mathbf{B} & \cdots & a_{pq}\mathbf{B} \end{bmatrix}.$$

We index the pm rows of $\mathbf{A} \otimes \mathbf{B}$ by

$$R = \{(1,1),(2,1),...,(m,1),(1,2),(2,2),...,(m,2),....,(1,p),...,(m,p)\},$$

and the qn columns by

$$C = \{(1,1),(2,1),...,(n,1),(1,2),(2,2),...,(n,2),....,(1,q),...,(n,q)\}.$$

For any two subsets $\bar{R} \subset R$ and $\bar{C} \subset C$, we denote by $(\mathbf{A} \otimes \mathbf{B})_{\bar{R}\bar{C}}$ the matrix such that $(\mathbf{A} \otimes \mathbf{B})_{(i,j)(c,d)}$ is an entry of $(\mathbf{A} \otimes \mathbf{B})_{\bar{R}\bar{C}}$ iff $(i,j) \in \bar{R}$ and $(c,d) \in \bar{C}$.

ASSUMPTION 1.2. *Consider the $n^2 \times n^2$ matrix $\mathbf{\Gamma}^\star = \mathbf{\Sigma}^\star \otimes \mathbf{\Sigma}^\star$. There exists some $\alpha \in (0,1]$ such that*

$$\max_{e \in S^c} ||\mathbf{\Gamma}^\star_{eS}\left(\mathbf{\Gamma}^\star_{SS}\right)^{-1}||_1 \leq 1 - \alpha,$$

where $S = \mathcal{E} \cup \{(i,i)|i \in \mathcal{V}\}$ *and* $S^c = \left\{(i,j) \in \mathcal{V} \times \mathcal{V}, k^\star_{ij} = 0\right\}$.

Assumption 1.2 limits the amount of dependence between the non-edge terms (indexed by S^c) and the edge-based terms (indexed by S). The limit is controlled by α: the bigger the α the smaller the dependence. In other words, if we set

$$X_{(j,k)} = y_j y_k - \mathrm{E}\left(y_j y_k\right), \quad \text{for all } j,k \in \mathcal{V},$$

then conditional on $(b, \mathbf{\Theta})$, the correlation between $X_{(j,k)}$ and $X_{(\ell,m)}$ is low for any $(j,k) \in S$ and $(\ell,m) \in S^c$.

In the following, Theorem 1 shows that: (a) the rate at which the realized network estimator converges to the true value as the sample size m increases, and (b) a lower bound on the probability of correctly detecting the nonzero partial correlations (as well as their signs) as a function of the sample size m. In particular, the estimator is model selection consistent with high probability, when n is large.

THEOREM 1.1. *Assume Assumptions 1.1 and 1.2 hold, and choose* $\lambda = \frac{8}{\alpha} m^{-\beta} \left(\frac{\log(a_2 m^{\alpha_0} n^\tau)}{a_3}\right)^{\frac{1}{a}}$ *in* (1.5)*, where* $\tau > 2$ *is arbitrary.*

(a) *Assume that*

$$m > \left(\frac{2^{\alpha_0}}{a_3} \log\left(a_2 n^\tau \left(a_3^{-\frac{1}{a\beta}} c_0^{\frac{1}{\beta}}\right)^{\alpha_0}\right)\right)^{\frac{1}{a\beta}} c_0^{\frac{1}{\beta}}, \tag{1.6}$$

where

$$c_0 := \max\left(\frac{1}{a_1}, 6(1+8\alpha^{-1})^2 d \max(C_{\mathbf{\Sigma}^\star} C_{\mathbf{\Gamma}^\star}, C^3_{\mathbf{\Sigma}^\star} C^2_{\mathbf{\Gamma}^\star}), \frac{a_3^{\frac{1}{a}}}{a_2^{\beta}} \exp\left(\frac{2}{a^2\beta}\right) \mathbf{1}_{\{\alpha_0 = 1\}}, \frac{1}{\sigma_n}\right). \tag{1.7}$$

Here, $\sigma_n = \min_i \sigma^*_{ii}$*,* d *is the maximum degree of the network, that is, the maximum number of edges that include a vertex, and we have set* $C_{\mathbf{\Gamma}^\star} = |||(\mathbf{\Gamma}^\star_{SS})^{-1}|||_\infty$ *and* $C_{\mathbf{\Sigma}^\star} = |||\mathbf{\Sigma}^\star|||_\infty$.

Then,

$$\mathrm{P}\left(||\widehat{\mathbf{K}}_\lambda - \mathbf{K}^\star||_\infty \leq 2\left(1+8\alpha^{-1}\right) C_{\mathbf{\Gamma}^*} m^{-\beta} \left(\frac{\log(a_2 m^{\alpha_0} n^\tau)}{a_3}\right)^{\frac{1}{a}}\right) \geq 1 - \frac{1}{n^{\tau-2}}\,. \tag{1.8}$$

(b) *Define* $\bar{c}_0 = \max\left(c_0, \frac{2C_{\mathbf{\Gamma}^\star}\left(1+8\alpha^{-1}\right)}{k_m}\right)$*, where* k_m *is the minimum absolute value of the*

non-zero entries of $\mathbf{K}^\star$. *Assume that*

$$m > \left(\frac{2^{\alpha_0}}{a_3}\log\left(a_2 n^\tau\left(a_3^{-\frac{1}{a\beta}}\overline{c}_0^{\frac{1}{\beta}}\right)^{\alpha_0}\right)\right)^{\frac{1}{a\beta}}\overline{c}_0^{\frac{1}{\beta}}.$$

Then,

$$\mathrm{P}\left(\mathrm{sign}(\widehat{k}_{\lambda\, ij}) = \mathrm{sign}(k^\star_{ij}),\ \forall i,j \in \mathcal{V}\right) \geq 1 - \frac{1}{n^{\tau-2}}.$$

Let us now give the intuition of the proof. Assumption (3.2) implies that m is sufficiently large in order to use Assumption 1.1. The effect of the parameter σ_n in (1.7) is to make $|\overline{\sigma}_{ii} - \sigma^\star_{ii}|$ not larger than σ_n with high probability for all i, so that $\overline{\mathbf{\Sigma}}$ will have positive diagonal entries. In this case, (1.5) is an optimization problem for a convex function, and there is a unique solution. Set $\widetilde{\mathbf{K}}_\lambda = (\widetilde{k}_{\lambda\, ij})$ be the solution to (1.5) with the additional restriction that $\widetilde{k}_{\lambda\, ij} = 0$ if $k^\star_{ij} = 0$ (see (1.19)). Based on the primal-dual witness construction (see for example Ravikumar et al. (2011)), when the ℓ_∞-norms of $\widetilde{\mathbf{K}}_\lambda - \mathbf{K}^\star$ and $\overline{\mathbf{\Sigma}} - \mathbf{\Sigma}^\star$ are not larger than some quantity determined by $C_{\mathbf{\Gamma}^\star}$, $C_{\mathbf{\Sigma}^\star}$,$\alpha$, d and λ,we have $\widetilde{\mathbf{K}}_\lambda = \widehat{\mathbf{K}}_\lambda$. On the other hand, using some matrix algebra we can bound $||\widetilde{\mathbf{K}}_\lambda - \mathbf{K}^\star||_\infty$ in terms of $||\overline{\mathbf{\Sigma}} - \mathbf{\Sigma}^\star||_\infty, C_{\mathbf{\Gamma}^\star}, C_{\mathbf{\Sigma}^\star}, \alpha$ and λ. Then we define λ as in Theorem 1.1, and choose m large enough to make $||\overline{\mathbf{\Sigma}} - \mathbf{\Sigma}^\star||_\infty$ and $||\widehat{\mathbf{K}}_\lambda - \mathbf{K}^\star||_\infty$ satisfy the conditions. Therefore, with high probability we have $\widetilde{\mathbf{K}}_\lambda = \widehat{\mathbf{K}}_\lambda$, and so the same bound for $||\widetilde{\mathbf{K}}_\lambda - \mathbf{K}^\star||_\infty$ is also satisfied for $||\widehat{\mathbf{K}}_\lambda - \mathbf{K}^\star||_\infty$. In part (b) of Theorem 1.1, we need to introduce the new parameter k_n in the lower bound of m. Its effect is to guarantee that with high probability $|\widehat{k}_{\lambda\, ij} - k^\star_{ij}|$ is smaller than k_m. Then given the definition of k_m, we can see that the signs of $\widehat{k}_{\lambda\, ij}$ and $k^\star_{ij}$ are the same.

Observe that when Assumption 1.1 holds with $\alpha_0 = 0$, Theorem 1.1 is a direct application of Theorems 1 and 2 of Ravikumar et al. (2011). We give the proof of Theorem 1.1 for the case $\alpha_0 = 1$ in the appendix.

The parameter d controls the degree of sparsity of our network. It ranges from 0 (fully sparse) to n (fully interconnected). To explore its effects, let us assume that the parameters $C_{\mathbf{\Sigma}^*}, C_{\mathbf{\Gamma}^*}$, α and σ_n in Theorem 1.1 remain constant as a function (n, m, d). In this case, when $d = 0$, condition (3.2) means that m should not be smaller than $O\left((\log n)^{\frac{1}{a\beta}}\right)$. When $d = n$, condition (3.2) means that m should not be smaller than $O\left((\log n)^{\frac{1}{a\beta}} n^{\frac{1}{\beta}}\right)$, since in this case $c_0 = O(n)$. In other words, the sparser the network is, the lesser observations are required to estimate the concentration matrix accurately.

1.4 Extensions

1.4.1 Microstructure Noise and Asynchronycity

Rather than the efficient price, it is customary to assume that the econometrician observes the transaction (or midquote) price. This differs from the efficient price because trades (quotes) are affected by an array of market frictions that go under the umbrella term of market microstructure. Moreover, it is common to assume that the trades (quotes) of different assets are executed (posted) asynchronously. In this section we extend the baseline framework of Section 1.2 and introduce a number of realized covariance estimators designed to handle microstructure noise and asynchronous trading.

We assume that the log-prices of each asset i are observed asynchronously on different time grids $T_i = \{t_{i1}, \ldots, t_{i\,m_i}\}$, $i = 1, ..., n$. For each asset $i = 1, \ldots, n$ the econometrician observes the transaction (or midquote) prices $x_i(t_{i\ell})$ defined as

$$x_i(t_{i\ell}) = y_i(t_{i\ell}) + u_i(t_{i\ell}) \ ,$$

where $u_i(t_{i\ell})$ denotes the microstructure noise associated with the ℓth trade. We assume that the microstructure noise satisfies

$$u_i(t_{i\ell}) \sim N(0, \eta_i^2) \ ,$$

where η_i^2 is some positive constant. We assume that the noise process is independent of the efficient price process, and use the same notation $(\Omega, \mathcal{F}, \mathcal{F}_t, \mathrm{P})$ for the product filtered probability space of both the efficient price and the noise processes.

A standard technique used to handle asynchronous trading for realized covariance estimation is refresh time sampling, which was introduced by Martens (2004). Several variants of this technique exist, like the pairwise and groupwise refresh time approaches, used in Fan et al. (2012), Lunde et al. (2011) and Hautsch et al. (2012). In this work we use pairwise refresh time sampling, and its procedure is explained in the appendix. Pairwise refresh time sampling based covariance estimation consists of estimating each entry of the covariance separately. The i, j–entry of the matrix is computed by first synchronizing the observations of assets i and j using refresh time and then estimating the covariance between assets i and j using the synchronized data. Notice that this approach does not guarantee that the covariance estimator is

positive definite, as each covariance entry is estimated using different subsets of observations. Suppose for assets i and j the refresh sample prices are respectively $x_i^r = \{x_i(t_{i\,1}^r), \ldots, x_i(t_{i\,m}^r)\}$ and $x_j^r = \left\{x_j(t_{j\,1}^r), \ldots, x_j(t_{j\,m}^r)\right\}$. We use the shorthand notation $x_{\ell\,k}^r$, $y_{\ell\,k}^r$ and $u_{\ell,\,k}^r$ to denote $x_\ell(t_{\ell\,k}^r)$, $y_\ell(t_{\ell\,k}^r)$ and $u_\ell(t_{\ell\,k}^r)$, respectively. Also, we define M_0 as the minimum pairwise refresh sample size across all pairs of assets.

After the data have been opportunely synchronized, a number of market microstructure noise robust estimators can be applied. First, we consider the two–scales realized covariance estimator (TSRC) proposed in Zhang (2011), which is a multivariate extension of the two–scales estimator introduced by Aït-Sahalia et al. (2005).

Two-Scales Realized Covariance Estimator. The Two-Scales Realized Covariance Estimator $\overline{\mathbf{\Sigma}}_{\mathsf{TS}}$ based on pairwise refresh time is defined as

$$\overline{\sigma}_{\mathsf{TS},ij} = \frac{1}{K}\sum_{k=K+1}^{m}\left(x_{i\,k}^r - x_{i\,k-K}^r\right)\left(x_{j\,k}^r - x_{j\,k-K}^r\right) - \frac{m_K}{m_J}\frac{1}{J}\sum_{k=J+1}^{m}\left(x_{i\,k}^r - x_{i\,k-J}^r\right)\left(x_{j\,k}^r - x_{j\,k-J}^r\right)$$

where $m_K = \frac{m-K+1}{K}$ and $m_J = \frac{m-J+1}{J}$.

For this estimator, we assume the noise u_i as an i.i.d. process, and the noise processes of different assets are assumed to be independent of each other.

Zhang (2011) shows that the optimal choice of K has order $K = O(m^{\frac{2}{3}})$, and J can be taken to be a constant such as 1. The first component of this estimator is the average of K realized variances, and it converges to $\sigma_{ij}^\star$ in the absence of noise. The second component is set to correct the bias caused by the noise. Under the optimal choice of K and J, the estimation error is asymptotically mixed normal with mean 0 and variance of order $O\left(m^{-\frac{1}{3}}\right)$. If we further assume that $\frac{1}{2}m^{\frac{1}{3}} \leq m_K \leq 2m^{\frac{1}{3}}$, then, under the condition that the syncronized observation times satisfy condition,

$$\sup_{\ell\in\{1,\ldots,m\}}|\tau_\ell - \tau_{\ell-1}| \leq \frac{c}{m}, \tag{1.9}$$

Fan et al. (2012) show that this estimator satisfies Assumption 1.1 with $\beta = \frac{1}{6}$, $a = 2$ and $\alpha_0 = 0$, and thus Theorem 1.1. Observe that since M_0 is defined as the minimum pairwise refresh sample size across all pairs of assets, we should replace m with M_0 when applying Theorem 1 to TSRC.

In the empirical implementation, for each pair of assets we choose J as 1 and K as the

(rounded) average of the optimal bandwidth for the realized two-scales volatility estimator of the two assets, following the procedure detailed in Aït-Sahalia et al. (2005).

Another robust estimator is the Multivariate Generalized Flat-top Realized Kernel estimator (MGFRK), which is proposed in Varneskov (2016). Compared to the realized kernel estimator in Barndorff-Nielsen et al. (2008), the MGFRK has the advantage of dealing with dependent noise series.

Multivariate Generalized Flat-Top Realized Kernel Estimator. The MGFRK $\overline{\Sigma}_{\mathsf{K}}$ based on pairwise refresh time is defined as

$$\overline{\sigma}_{\mathsf{K}\,ij} = \gamma_0\left(x_i^r, x_j^r\right) + \frac{1}{2}\sum_{h=1}^{H} k\left(\frac{h-1}{H}\right)\left(\gamma_h(x_i^r, x_j^r) + \gamma_{-h}(x_i^r, x_j^r) + \gamma_h(x_j^r, x_i^r) + \gamma_{-h}(x_j^r, x_i^r)\right), \tag{1.10}$$

where for each $h \in \{-H, -H+1, \ldots, -1, 0, 1, \ldots, H-1, H\}$,

$$\gamma_h(x_i^r, x_j^r) = \sum_{p=1}^{m-1}(x_{i\,p}^r - x_{i\,p-1}^r)(x_{j\,p-h}^r - x_{j\,p-h-1}^r).$$

The kernel function $k : [0,1] \to \mathbb{R}$ satisfies the following conditions: (1) $k(x) = 1$ for $x \in \left[0, \frac{M}{m}\right]$; (2) $k(x)$ is twice differentiable with bounded derivatives on $\left(\frac{M}{m}, 1\right)$; (3) $k(1) = k'\left(\frac{M}{m}\right) = k'(1) = 0$. In order to find a suitable kernel function, it is obvious that we can assume it has the polynomial form $k(x) = d_p x^p + d_{p-1}x^{p-1} + \cdots + d_1 x + d_0$ when $x \in \left(\frac{M}{m}, 1\right)$, for a fixed integer $p \geq 3$, and work out the values of the coefficients.

For this estimator we assume the same hypothesis on the noise as in Tao et al. (2013). That is, we assume $(u_{1,k}^r, \ldots, u_{n,k}^r)$ as multivariate M-dependent time series for generic k, where M is a fixed integer. Therefore, for generic $\{p_1, \ldots, p_n\}$ and $\{q_1, \ldots, q_n\}$, $\{u_{1,p_1}^r, \ldots, u_{n,p_n}^r\}$ are jointly independent of $\{u_{1,q_1}^r, \ldots, u_{n,q_n}^r\}$ when $|p-q| > M$ for all $p \in \{p_1, \ldots, p_n\}$ and $q \in \{q_1, \ldots, q_n\}$.

For a choice of H of $O(m^{\frac{1}{2}})$, Varneskov (2016) shows that the estimation error is asymptotically mixed normal with mean 0 and variance of order $O\left(m^{-\frac{1}{2}}\right)$. In the appendix, we show that this estimator follows the following concentration inequality.

THEOREM 1.2. *If the synchronized observation times satisfy condition (1.4), then there exist*

positive constants a_1, a_2 *and* a_3 *such that for all* $i, j \in \{1, ..., n\}$, $x \in [0, a_1]$, *and* M_0 *large,*

$$\mathrm{P}\left(\left|\overline{\sigma}_{\mathsf{K}\,ij} - \sigma^{\star}_{ij}\right| > x\right) \leq a_2 M_0 \exp(-a_3 M_0^{1/4} x).$$

Therefore, the MGFRK satisfies Assumption 1.1 with $\beta = \frac{1}{4}$, $a = 1$ and $\alpha_0 = 1$. Hence, we can apply Theorem 1.1, and we obtain that the estimation error of $\widehat{\mathbf{K}}_{\lambda}$ converges to zero at rate $M_0^{-\frac{1}{4}}\sqrt{\log n}$ (assuming that all other parameters including $\alpha, d, C_{\mathbf{\Gamma}^*}$ and C_{Σ^*} are constants). Thus, in this case, $M_0^{\frac{1}{2}}$ is required to be large compared to $\log n$ to make the error small in probability. Notice that this result is analogous to the one obtained in Tao et al. (2013), where the same convergence rate $M_0^{-\frac{1}{4}}\sqrt{\log n}$ is obtained for a multi-scales realized covariance estimator. Moreover, according to Tao et al. (2013), this rate is optimal on estimating the integrated covariance matrix with noisy high-frequency data.

1.4.2 Factor Structure

Classical asset pricing theory models like the CAPM or APT imply that the unexpected rate of return of risky assets can be expressed as a linear function of few common factors and an idiosyncratic component. Factors induce a fully interconnected partial correlation network structure. In this case, it is natural to carry out network analysis on the partial correlation structure of the assets after netting out the influence of common sources of variation. In this section we propose a modification of our network definition for such systems. Also, we propose a modified covariance estimation strategy analogous to the one put forward in Fan et al. (2008) and Fan et al. (2011) which is based on the particular structure of the system.

We augment the y process with additional k components representing factors. The dynamics of the augmented system are assumed to be the same as the one described in (2.1). Moreover, the factors are assumed to be observed, as it is commonly done in the empirical finance literature and also as in Fan et al. (2008). The integrated covariance of the augmented system can then be partitioned as an $(n + k) \times (n + k)$ matrix

$$\mathbf{\Sigma}^{\star} = \begin{bmatrix} \mathbf{\Sigma}^{\star}_{AA} & \mathbf{\Sigma}^{\star}_{FA} \\ \mathbf{\Sigma}^{\star}_{AF} & \mathbf{\Sigma}^{\star}_{FF} \end{bmatrix}, \tag{1.11}$$

where A and F denote, respectively, the blocks of assets and factors.

The covariance of the assets can be expressed as the sum of the systematic and idiosyncratic components, that is

$$\mathbf{\Sigma}^{\star}_{AA} = \mathbf{B}\mathbf{\Sigma}^{\star}_{FF}\mathbf{B}' + \mathbf{\Sigma}^{\star}_{I},$$

where

$$\mathbf{B} = \mathbf{\Sigma}^{\star}_{AF}\left[\mathbf{\Sigma}^{\star}_{FF}\right]^{-1} \text{ and } \mathbf{\Sigma}^{\star}_{I} = \mathbf{\Sigma}^{\star}_{AA} - \mathbf{\Sigma}^{\star}_{AF}\left[\mathbf{\Sigma}^{\star}_{FF}\right]^{-1}\mathbf{\Sigma}^{\star}_{FA}.$$

If the factors are pervasive ($\mathbf{B}$ is not sparse), then the concentration matrix of the assets cannot be sparse. In these cases, rather than proposing a network definition on the basis of the partial correlations of the system, we propose a network definition based on the idiosyncratic partial correlations, that is the partial correlations implied by the idiosyncratic covariance matrix $\mathbf{\Sigma}^{\star}_{I}$. Precisely, we define the idiosyncratic integrated partial correlation network as the network whose set of edges is given by

$$\mathcal{E}_I = \left\{(i,j) \in \mathcal{V}\times\mathcal{V}, k^{\star}_{I\,ij} \neq 0, i \neq j\right\} \ ,$$

where $k^{\star}_{I\,ij}$ is the i,j–entry of the matrix $\mathbf{K}^{*}_{I} = (\mathbf{\Sigma}^{*}_{I})^{-1}$.

Let $\overline{\mathbf{\Sigma}}$ be an appropriate estimator of the integrated covariance of the augmented system and consider partitioning the estimated covariance matrix analogously to equation (1.11)

$$\overline{\mathbf{\Sigma}} = \begin{bmatrix} \overline{\mathbf{\Sigma}}_{AA} & \overline{\mathbf{\Sigma}}_{FA} \\ \overline{\mathbf{\Sigma}}_{AF} & \overline{\mathbf{\Sigma}}_{FF} \end{bmatrix} .$$

Then, a natural estimator of the idiosyncratic realized covariance estimator $\overline{\mathbf{\Sigma}}_{I} = (\overline{\sigma}_{I\,ij})$ is

$$\overline{\mathbf{\Sigma}}_{I} = (\overline{\sigma}_{I\,ij}) = \overline{\mathbf{\Sigma}}_{AA} - \overline{\mathbf{\Sigma}}_{FA}\left[\overline{\mathbf{\Sigma}}_{FF}\right]^{-1}\overline{\mathbf{\Sigma}}_{AF} \ . \tag{1.12}$$

The following corollary establishes the concentration inequality of the estimator $\overline{\mathbf{\Sigma}}_{I}$ using the one for $\overline{\mathbf{\Sigma}}$.

COROLLARY 1.1. *If Assumption 1.1 holds, then there exist positive constants b_1, b_2 and b_3 such that for all $i,j \in \{1,...,n\}$, $x \in [0, b_1]$, and M_0 large,*

$$\mathrm{P}\left(\left|\overline{\sigma}_{I\,ij} - \sigma^{*}_{I\,ij}\right| > x\right) \leq b_2 M_0^{\alpha_0}\exp(-b_3(M_0^{\beta}x)^{a}),$$

where β, a and α_0 are the constants from Assumption 1.1.

The realized network estimator can thus be applied to regularize the idiosyncratic realized covariance matrix and estimate the idiosyncratic partial correlation network. Moreover, the covariance matrix of the assets can be estimated as

$$\widehat{\mathbf{\Sigma}}_{AA} = \overline{\mathbf{B}}\,\overline{\mathbf{\Sigma}}_{FF}\,\overline{\mathbf{B}}' + \widehat{\mathbf{\Sigma}}_{I\lambda},$$

where $\widehat{\mathbf{\Sigma}}_{I\lambda}$ denotes the realized covariance estimator implied by the realized network. Notice, that this estimation strategy is analogous to the one proposed in Fan et al. (2011).

1.5 Simulation Study

In this section we carry out a simulation study to assess the finite sample properties of the realized network estimator. The simulation exercise consists of simulating one day of high frequency data and to apply the realized network estimator to estimate the integrated covariance and the integrated concentration matrices. Different specifications of the covariance matrix of the efficient price process are used to assess the usefulness of the realized network estimator depending on the underlying cross-sectional dependence structure of the data. The realized network estimator is also benchmarked against a set of alternative regularization procedures proposed in the literature.

In our simulation study, the efficient price $y(t)$ is an n–dimensional zero drift diffusion with n equal to 50. As in Hautsch et al. (2012), the efficient price follows a simple diffusion with constant covariance $\mathbf{\Sigma}$,

$$y(t) = \int_0^t \mathbf{\Theta}\, dB(u).$$

where $\mathbf{\Theta}$ is the Cholesky factorization of $\mathbf{\Sigma}$. The econometrician observes the microstructure contaminated version of the efficient price, and we assume that the noise component u_{ik} is an independent zero mean normal random variable with variance equal to $(0.05)^2$ for all stocks. Finally, prices of each stock are observed asynchronously, according to the realization Poisson process with a constant intensity calibrated to have 5 trades per minute on average. In our numerical implementation, a trading day is 6.5 hours and the simulation of the continuous time process is carried out using the Euler scheme with a discretization step of 5 seconds.

Three different specifications of the covariance matrix $\mathbf{\Sigma}$ are adopted. In the first simulation design (Design 1), we pick a specification for $\mathbf{\Sigma}$ which induces a sparse partial correlation

structure among the assets in the panel. In particular, we choose $\mathbf{\Sigma}$ as a function of a realization of an Erdös–Renyi random graph. The Erdös–Renyi random graph $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ is an undirected graph defined over a fixed set of vertices $\mathcal{V} = \{1, \ldots, n\}$ and a random set of edges where the existence of an edge between any pair of vertices is determined by an independent Bernoulli trial with probability p. We generate $\mathbf{\Sigma}$ by first simulating an Erdös–Renyi random graph $\mathcal{G}$ and then setting $\mathbf{\Sigma}$ equal to

$$\mathbf{\Sigma} = \left[\mathbf{I}_n + \mathbf{D} - \mathbf{A}\right]^{-1},$$

where $\mathbf{I}_n$ is the n dimensional identity matrix, and $\mathbf{D}$ and $\mathbf{A}$ are respectively the degree matrix and the adjancy matrix of the random graph $\mathcal{G}$. The model for $\mathbf{\Sigma}$ is such that the underlying random graph structure determines the sparsity structure of the integrated concentration matrix. Also, note that $\mathbf{\Sigma}$ is symmetric positive definite by construction. In the simulation we set p equal to $2/n$. In this scenario (i.e. when np is greater than one) the Erdös–Renyi random graph will almost surely have a giant component, that is a connected component containing a constant fraction of the entire set of network vertices. Thus, the highlight of the model is that the generated concentration matrix is sparse while the corresponding covariance matrix is not.

In the second simulation design (Design 2), we pick a specification for $\mathbf{\Sigma}$ based on a factor model. We set $\mathbf{\Sigma}$ as

$$\mathbf{\Sigma} = \mathbf{B}\,\mathbf{I}_k\,\mathbf{B}' + \mathbf{I}_n,$$

where $\mathbf{B}$ is an $n \times k$ matrix whose entries are iid normal Gaussian draws with mean zero and unit variance. In the simulation we set the number of factors k to 2. Notice that in this scenario it is challenging for the realized network estimator in that the inverse covariance matrix implied by the model is not sparse.

Last, in the third simulation design (Design 3), we set the $\mathbf{\Sigma}$ matrix as

$$[\mathbf{\Sigma}]_{ij} = \begin{cases} \rho + \rho^{|i-j|} & \text{if } i \neq j \\ 1 & \text{otherwise} \end{cases},$$

and set ρ equal to 0.25. Notice that also in this scenario the inverse covariance matrix is not sparse and the covariance matrix does not have a factor representation.

Different approaches are used to estimate the integrated covariance and integrated concentration matrices. First, we estimate the integrated covariance using the 5–minute frequency RC,

the pairwise–refreshed time TSRC and the pairwise–refreshed time MRK. The bandwidth parameters of the TSRC and MRK are computed using the plug–in rules previously described. It is important to stress that the TSRC and MRK estimators are not guaranteed to be positive definite. When the estimators are indefinite we apply eigenvalue cleaning (as described in Hautsch et al. (2012)), which is introduced in the appendix, to obtain a positive definite estimator. For each realized covariance estimator, we consider a number of different regularization procedures. First, we consider the shrinkage estimator proposed by Ledoit and Wolf (2004). Let $\overline{\mathbf{\Sigma}}$ denote the unregularized realized covariance (computed using either the RC, TSRC or MRK estimators and without applying eigenvalue cleaning). The shrinkage estimator is defined as

$$\widehat{\mathbf{\Sigma}}_{LW} = \alpha_1 \mathbf{I}_n + \alpha_2 \overline{\mathbf{\Sigma}}, \tag{1.13}$$

where α_1 and α_2 are two tuning parameters chosen to minimize the risk of the estimator that we set following Ledoit and Wolf (2004) and Nikolaus et al. (2013). The second regularized estimator is based on a factor approximation of the covariance matrix. It is defined as

$$\widehat{\mathbf{\Sigma}}_{F} = \sum_{i=1}^{k} \hat{\xi}_i \hat{e}_i \hat{e}_i' + \widehat{\mathbf{R}}_k, \tag{1.14}$$

where $\hat{\xi}_i$ and $\hat{e}_i$ denote the eigenvalues (in increasing order) and corresponding eigenvectors obtained from the spectral decomposition of the unregularized realized estimator $\overline{\mathbf{\Sigma}}$, and $\widehat{\mathbf{R}}_k$ is $\text{diag}(\overline{\mathbf{\Sigma}} - \sum_{i=1}^{k} \hat{\xi}_i \hat{e}_i \hat{e}_i')$. Notice that the shrinkage and factor regularization procedures are also not guaranteed to provide a positive definite estimator when applied to the TSRC and MRK estimators. In these cases we apply eigenvalue cleaning whenever the resulting estimator is not positive definite (which however happens rarely in the simulation designs we consider). Last, we use the realized network estimator defined in Section 1.2 using the BIC criterion to determine the optimal amount of shrinkage to apply. Notice that one of the inputs of the BIC criterion is the number of observations used to compute the estimator. When using pairwise–refresh sampling however, this number is different for each entry of the covariance matrix. Similarly to Hautsch et al. (2012) we opt for a conservative choice of this quantity, and we set it to the minimum number of refresh time observations across all pairs. The realized network estimator is guaranteed to be positive definite as long as the pre–estimator $\overline{\mathbf{\Sigma}}$ is positive semi-definite. If the unregularized estimator is indefinite, we add to it the identity matrix times the absolute value of its smallest eigenvalue to obtain a positive semi-definite estimator.

Different metrics are used to evaluate the performance of the estimators. A classic loss function used for the evaluation of covariance matrix estimators is the Kullback–Liebler loss proposed by Stein (Stein, 1956; Pourahmadi, 2013), which is defined as

$$\mathsf{KL}\left(\widehat{\mathbf{\Sigma}}\right) = \text{tr}(\widehat{\mathbf{\Sigma}}\mathbf{K}^*) - \log|\widehat{\mathbf{\Sigma}}\mathbf{K}^*| - n\,.$$

Following (Hautsch et al., 2012), we also consider a Root Means Square Error (RMSE) type loss based on the scaled Frobenius norm of the covariance matrix, which is defined as

$$\mathsf{RMSE}\left(\widehat{\mathbf{\Sigma}}\right) = \sqrt{\frac{1}{n}\sum_{i,j=1}^{n}(\widehat{\sigma}_{ij} - \sigma_{ij}^*)^2}\,.$$

We perform 10'000 Monte Carlo replications of the simulation exercise for each simulation design and report summary statistics on the performance of the estimators in table 1.1. The table reports the average of the KL and RMSE losses of the estimators in the three simulation designs. Results convey that using a regularization technique whose shrinking target is closer to the true underlying structure of the data produces the best results and large improvements over the unregularized realized covariance matrix estimator. In particular, it is easy to see that when the partial correlation structure of the data is sparse (Design 1) the realized network estimator is the best performing regularization technique. Analogously, the factor based regularization works best in Design 2 and shrinkage regularization works in Design 3. We note that using some form of regularization is always advantageous with respect to the unregularized estimator even when the assumptions of the DGP do not match exactly the structure of the shrinkage. Thus in general our results show that it is alwasy convinien to apply shrinkages as perfromance is always better and some times much better. The results convey that the gains by using the realized network estimator when the partial correlation structure is sparse can be substantial.

1.6 Empirical Application

We use the realized network estimator to analyse the dependence structure of a panel of US blue chip stocks from the NYSE throughout 2009. We then engage in a Markowitz style Global Minimum Variance portfolio prediction exercise to highlight the advantages of the methodology for forecasting.

1.6.1 Data and Estimation

We consider a sample of 100 liquid US blue chip stocks that have been part of the S&P100 index for most of the 2000's. We also include in the panel the SPY ETF, the ETF tracking the S&P 500 index. We work with tick–by–tick transaction prices obtained from the NYSE–TAQ database. Before proceeding with the econometric analysis, the data are filtered using standard techniques described in Brownlees and Gallo (2006) and Barndorff-Nielsen et al. (2009). The full list of tickers, company names and industry groups is reported in table 1.2.

We estimate the integrated covariance of the assets throughout 2009. More precisely, for each of the 52 weeks of 2009, we use the data on the last weekday available of each week to construct the realized covariance estimators. On each of these days, we only consider the tickers that have at least 1000 trades. Exploratory analysis (non reported in this chapter) confirms the well documented evidence of a CAPM–type factor structure in the panel. To this extent, our realized covariance estimation strategy consists of first decomposing the realized covariance in systematic and idiosyncratic covariance components and then regularizing the idiosyncratic part with the realized network. More precisely, we compute the realized covariance of the assets in the panel together with the SPY ticker (the proxy of the market), and then obtain the systematic and idiosyncratic components of the realized covariance of the assets on the basis of formula (1.12). Finally, we apply the realized network regularization procedure to the idiosyncratic realized covariance. On each week of 2009, we estimate the realized network using three (idiosyncratic) realized covariance estimators: MRK, TSRC as well as the classic RC using data sampled at a 1 minute frequency.

1.6.2 Realized Network Estimates

In this section we present the realized network estimation results. We first provide details for one specific date only that roughly corresponds to the mid of the sample (June 26, 2009), and then report summaries for all estimated networks in 2009. In the interest of space we report the TSRC estimator results only. The MRK and RC provide similar evidence.

We begin by showing in figure 1.1 the heatmap of the idiosyncratic correlation matrix associated with the idiosyncratic realized covariance estimator on June 26. Notice that the heatmap is constructed by sorting stocks by industry group and then by alphabetical order. The picture clearly shows that after netting out the influence of the market factor, a fair amount of cross–

sectional dependence is still present across stocks. Inspection of the heatmap reveals that the majority of estimated correlation coefficients are positive. The correlation matrix exhibits a block diagonal structure hinting that correlation is stronger among firms in the same industry. On this date, the intra-industry group correlation is particularly strong for energy companies.

We estimate the realized network using the GLASSO and use the BIC to choose the optimal amount of shrinkage. To give insights on the sensitivity of the estimated network to the shrinkage parameter λ, the top panel of figure 1.2 reports the so called trace, which is the graph of the estimated partial correlations as a function of the shrinkage parameter. The bottom panel of the same figure also shows the value of the BIC criterion as a function of the shrinkage parameter. The plot highlights how the sparsity of the estimated network varies substantially over the range of lambda values considered and that the majority of estimated partial correlations are positive irrespective of the value of shrinkage imposed on the estimator. The estimated network corresponding to the optimal λ has 188 edges, which correspond to approximately the 5% of the total number of possible edges in the network on this date. The number of companies that have interconnections are 66 (roughly 2/3 of the total) and are all connected in a unique giant component.

It is useful to provide details on the amount of variability explained by the systematic and idiosyncratic components of the covariance matrix of the panel. To this extent, we introduce the systematic coefficient of determination, defined as

$$R^2_{F\,i} = \frac{\overline{\mathbf{B}}_i' \overline{\mathbf{\Sigma}}_{FF} \overline{\mathbf{B}}_i}{\overline{\mathbf{B}}_i' \overline{\mathbf{\Sigma}}_{FF} \overline{\mathbf{B}}_i + \hat{\sigma}_{I\,ii}},$$

which measures the amount of variability of asset i explained by the market factor. We also introduce the idiosyncratic coefficient of determination, defined as

$$R^2_{I\,i} = \frac{\hat{\sigma}_{I\,ii} - 1/\hat{k}_{I\,ii}}{\hat{\sigma}_{I\,ii}},$$

which measures the amount of variability of asset i explained by the remaining assets conditional on the market factor. On June 26, the average of the systematic R^2_F is equal to 22.8% while the average of the idiosyncratic R^2_I (for those assets with at least one neighbour) is 7.3%. Overall the systematic component is the most relevant one in terms of explained variability, however, the idiosyncratic component captures a non negligible portion of variability as well.

Figure 1.3 displays the idiosyncratic partial correlation network. A number of comments on the empirical characteristics of the network are in order. First, on this date, Google (GOOG) emerges as a particularly highly interconnected firm, with linkages spreading to several other industry groups. The estimated network also shows some degree of industry clustering, that is linkages are more frequent among firms in the same industry group. In order to get better insights on the industry linkages in table 1.3 we report the total number of links across industry groups. The table shows that firms in the industrials, energy, technology and financials groups are particularly interconnected among each other. On the other hand, consumer discretionary, cosumer staples and healthcare have few intra-industry linkages. In figure 1.4 we report the degree distribution of the estimated network and the distribution of the nonzero partial correlations. As far as the degree distribution is concerned, the network exhibits the typical features of Power Law networks, that is the number of connections is heterogeneous and the most interconnected stocks have a large number of connections relative to the total number of links. The histogram of the partial correlation shows that the majority of the partial correlations are positive and that positive partial correlations are on average larger than the negative ones.

Last, we are interested in determining which companies are more interconnected and central in the network. We measure the degree of interconnectedness of a firm using different approaches: (i) the degree of a company in the network (that is, the number of links); (ii) the sum of nonzero square partial correlations of a company; and (iii) the centrality index of the page–rank algorithm. The page–rank algorithm is a famous network based centrality index used by web search engines to rank web pages. It turns out that the indices provide substantially close rankings. The rank correlation among the different measures are all above 0.9. We report the top ten most central companies in table 1.4 according to page rank. The page rank algorithm shows that Google is indeed the most central stock on this date.

We report a number of summary statistics for the sequence of networks estimated in 2009. First, in figure 1.5 we report the proportion of number of links in the network throughout the year. The picture shows that sparsity is stable a value slightly below 5% of the total possible number of linkages. The plots show the sparsity rate vis–à–vis the VIX volatility index to show that no particular time series pattern emerges in the network sparsity and that in particular the sparsity is unrelated to the level of volatility of the market.

Figure 1.6 shows the total number of links of each industry group divided by the total number of possible edges. The plot omits the series for materials, telecom and utilities due to their

small size. Technology, energy, financial and industrials are the most interconnected sectors also throughout 2009, and the level and cross sectional rankings are fairly stable across the sample. In order to give more insights on the degree of concentration within each group, in figure 1.7 we report the concentration of links in each industry group measured using the Herfindahl index. Again, materials, telecom and utilities are omitted from the graph. Once again, no particular time series pattern emerges from the plot and cross sectional concentration rankings are quite stable. The picture shows that the most interlinked sectors have quite different concentration characteristics. Technology is one of the most highly concentrated sector. Detailed inspection of the results reveals that this is driven by the fact that in 2009 Google is essentially the most interconnected ticker in the sample. On the other hand, industrials have the smaller average concentration, in that the number of links is quite uniformly distributed across firms and no specific "hub" emerges among these tickers.

Overall results convey that after conditioning on a one factor structure that data still has a fair amount of cross–sectional dependence and that networks provide a useful device to synthesize such information. The main empirical features of the network are stable throughout 2009. Firms in the energy and industrials sectors are strongly interconnected. Technology companies and Google in particular are the most highly interconnected firms throughout the year.

1.6.3 Predictive Analysis

In order to assess the ability of the regularized network methodology to provide more precise estimates of the integrated covariance we carry out an asset allocation prediction exercise (Hautsch et al., 2012; Nikolaus et al., 2013; Engle and Colacito, 2006). The forecasting exercise is designed as follows. For each week of 2009, we construct the Markowitz Global Minimum Variance (GMV) portfolio weights using the formula

$$\widehat{\mathbf{w}} = \frac{\widehat{\mathbf{\Sigma}}^{-1}\mathbf{1}}{\mathbf{1}'\widehat{\mathbf{\Sigma}}^{-1}\mathbf{1}} , \tag{1.15}$$

where $\mathbf{1}$ is n–dimensional vector of ones, and $\widehat{\mathbf{\Sigma}}$ is some estimator of the integrated covariance over day t. The resulting GMV portfolio weights are used to construct a portfolio of the assets which is held for the following week. At the end of the week, we compute the daily sample variance of the portfolio return for that period. We repeat the exercise for all the weeks of 2009. The performance of the covariance estimators is evaluated by assessing which estimator

delivers the smallest average out–of–sample GMV portfolio variance.

The set of estimators we consider are based on the systematic/idiosyncratic decomposition of the covariance matrix

$$\widehat{\mathbf{\Sigma}} = \overline{\mathbf{B}}\,\overline{\mathbf{\Sigma}}_{FF}\,\overline{\mathbf{B}}' + \widehat{\mathbf{\Sigma}}_I, \tag{1.16}$$

and differ on the choice of the estimator of the idiosyncratic realized covariance matrix $\widehat{\mathbf{\Sigma}}_I$. The set of candidate idiosyncratic realized covariance estimators contains: (i) unregularized covariance estimator; (ii) constrained covariance estimator, obtained by setting all the off–diagonal elements of the unregularized covariance estimator to zero; (iii) shrinkage covariance estimator of Ledoit and Wolf (2004) (see equation (1.13)); (iv) factor regularized covariance estimator (see equation (1.14)) based on three factors; (v) block-factor regularized estimator, obtained by applying factor regularization of equation (1.14) based on one factor to each industry block and setting the rest of the covariance matrix to zero; and (vi) realized network estimator. The exercise is carried out using the MRK, TSRC and RC estimators.

We report summary results on the forecasting exercise in table 1.5. The table shows the average annualized volatility of the GMV portfolios. The three different covariance estimators deliver analogous inference. The constrained estimator that ignores cross–sectional dependence in the idiosyncratic realized covariance matrix typically performs worst than the baseline unconstrained realized covariance estimators. Interestingly, the factor and block factor regularization schemes do not produce large gains out–of-sample in comparison to the benchmark. We interpret this as the consequence that after controlling for the market factor there is only weak evidence of the presence of additional factors. The block-factor regularization might be not particularly successful because while some sectors exhibit strong dependence (Industrials) a large number of stocks in other sectors do not (Consumer Discretionary, Consumer Staples and Healthcare). The shrinkage and realized network regularization schemes provide the best out–of–sample results, and the realized network estimator in particular achives the lower out–of–sample variance. Also, it is interesting to point out that among the two market friction robust estimators, the MRK delivers lower out of sample losses than the TSRC. Last we note that the difference in the forecasting performance across the different realized volatility estimators is substantially smaller after carrying out network regularization.

1.7 Conclusions

In this chapter we propose a regularization procedure for realized covariance estimators. The regularization consists of shrinking the off–diagonal elements of the inverse realized covariance matrix to zero using a LASSO–type penalty. Since estimating a sparse inverse realized covariance matrix is equivalent to detecting the partial correlation network structure of the daily log-prices, we call our procedure realized network. The technique is specifically designed for the Two–Scales Realized Covariance (TSRC) and the Multivariate Realized Kernel (MRK) estimators based on refresh time sampling, which are state–of–the–art consistent covariance estimators that allow for market microstructure effects and asynchronous trading. We establish the large sample properties of the procedure estimator and show that the realized network consistently estimates the inverse integrated covariance matrix and consistently detects the nonzero partial correlations of the network. An empirical exercise is used to highlight the usefulness of the procedure and an out—of–sample GMV portfolio asset allocation exercise is carried out to compare our procedure against a number of benchmarks. Results convey that realized network enhances the prediction properties of classic realized covariance estimators and performs well relative to a set of alternative regularization procedures.

1.8 Appendix

1.8.1 Figures and Tables

Table 1.1: SIMULATION STUDY

		No regular.	Shrinkage	Factor	Network
		Design 1			
RC	KL	65.89	55.13	48.04	46.06
	RMSE	55.77	44.98	57.15	40.43
TSRC	KL	51.95	48.72	30.32	28.32
	RMSE	29.97	29.56	29.13	19.35
MRK	KL	53.44	48.64	36.95	30.93
	RMSE	32.50	29.56	29.38	16.94
		Design 2			
RC	KL	73.54	52.07	27.85	46.65
	RMSE	68.65	54.59	51.98	45.55
TSRC	KL	52.30	10.99	3.94	36.67
	RMSE	30.09	20.84	19.02	20.11
MRK	KL	60.67	46.32	6.19	38.45
	RMSE	31.03	22.33	23.43	26.69
		Design 3			
RC	KL	36.29	15.53	17.61	31.16
	RMSE	39.46	19.28	37.69	37.28
TSRC	KL	40.26	4.64	3.35	9.36
	RMSE	15.95	10.65	12.01	12.09
MRK	KL	28.34	4.59	5.43	7.10
	RMSE	19.38	10.52	19.49	18.38

The table reports the KL and RMSE average losses of the unregularized RC, TSRC, and MRK estimators (No regular) as well as their regularized versions (Shrinkage, Factor, Network) in the three simulation designs of Section 1.5.

Table 1.2: Tickers, Company Names and Sectors

Ticker Symbol	Company Name	GICS Sector	Ticker Symbol	Company Name	GICS Sector
AMZN	Amazon.com	Consumer Discretionary	ABT	Abbott Laboratories	Health Care
CMCSA	Comcast	Consumer Discretionary	AMGN	Amgen	Health Care
DIS	Walt Disney	Consumer Discretionary	BAX	Baxter International	Health Care
F	Ford Motor	Consumer Discretionary	BMY	Bristol-Myers Squibb	Health Care
FOXA	Twenty-First Century Fox	Consumer Discretionary	GILD	Gilead Sciences	Health Care
GM	General Motors	Consumer Discretionary	JNJ	Johnson & Johnson	Health Care
HD	Home Depot	Consumer Discretionary	LLY	Lilly & Co.	Health Care
LOW	Lowes	Consumer Discretionary	MDT	Medtronic	Health Care
MCD	McDonalds	Consumer Discretionary	MRK	Merck	Health Care
NKE	NIKE	Consumer Discretionary	PFE	Pfizer	Health Care
SBUX	Starbucks	Consumer Discretionary	UNH	United Health Group	Health Care
TGT	Target	Consumer Discretionary	BA	Boeing Company	Industrials
TWX	Time Warner Inc.	Consumer Discretionary	CAT	Caterpillar	Industrials
CL	Colgate-Palmolive	Consumer Staples	EMR	Emerson Electric	Industrials
COST	Costco Co.	Consumer Staples	FDX	FedEx	Industrials
CVS	CVS Caremark	Consumer Staples	GD	General Dynamics	Industrials
KO	The Coca Cola Company	Consumer Staples	GE	General Electric	Industrials
MDLZ	Mondelez International	Consumer Staples	HON	Honeywell Intl	Industrials
MO	Altria Group Inc	Consumer Staples	LMT	Lockheed Martin	Industrials
PEP	PepsiCo Inc.	Consumer Staples	MMM	3M Company	Industrials
PG	Procter & Gamble	Consumer Staples	NSC	Norfolk Southern	Industrials
PM	Philip Morris	Consumer Staples	RTN	Raytheon Co.	Industrials
WAG	Walgreen	Consumer Staples	UNP	Union Pacific	Industrials
WMT	Wal-Mart Stores	Consumer Staples	UPS	United Parcel Service	Industrials
APA	Apache	Energy	UTX	United Technologies	Industrials
APC	Anadarko Petroleum	Energy	AAPL	Apple	Information Technology
COP	ConocoPhillips	Energy	ACN	Accenture	Information Technology
CVX	Chevron	Energy	CSCO	Cisco Systems	Information Technology
DVN	Devon Energy	Energy	EBAY	eBay	Information Technology
HAL	Halliburton Co.	Energy	EMC	EMC	Information Technology
NOV	National Oilwell Varco	Energy	FB	Facebook	Information Technology
OXY	Occidental Petroleum	Energy	GOOG	Google Inc.	Information Technology
SLB	Schlumberger Ltd.	Energy	HPQ	Hewlett-Packard	Information Technology
XOM	Exxon Mobil	Energy	IBM	IBM	Information Technology
AIG	AIG	Financials	INTC	Intel	Information Technology
ALL	Allstate	Financials	MA	Mastercard	Information Technology
AXP	American Express Co	Financials	MSFT	Microsoft	Information Technology
BAC	Bank of America	Financials	ORCL	Oracle	Information Technology
BK	Bank of New York	Financials	QCOM	QUALCOMM Inc.	Information Technology
BRK.B	Berkshire Hathaway	Financials	TXN	Texas Instruments	Information Technology
C	Citigroup Inc.	Financials	V	Visa Inc.	Information Technology
COF	Capital One Financial	Financials	DD	Du Pont	Materials
GS	Goldman Sachs Group	Financials	DOW	Dow Chemical	Materials
JPM	JPMorgan Chase & Co.	Financials	FCX	Freeport-McMoran	Materials
MET	MetLife Inc.	Financials	MON	Monsanto Co.	Materials
MS	Morgan Stanley	Financials	T	AT&T Inc	Telecommunications
SPG	Simon Property Group	Financials	VZ	Verizon Communications	Telecommunications
USB	U.S. Bancorp	Financials	AEP	American Electric Power	Utilities
WFC	Wells Fargo	Financials	EXC	Exelon	Utilities
ABBV	AbbVie	Health Care	SO	Southern Co.	Utilities

The table reports the list of company tickers, company names and industry sectors.

Table 1.3: LINKS ON 2009-06-26

	Disc	Stap	Ener	Fin	Heal	Ind	Tech	Mat	Tel	Util
Disc	1	2				4	8	1		
Stap	2	1	1		1	1	6	1		
Ener		1	35		2	2	12	6		
Fin				11		2	14	3		
Heal		1	2		3		7	3		
Ind	4	1	2	2		12	20	4		
Tech	8	6	12	14	7	20	14	6		2
Mat	1	1	6	3	3	4	6	3		
Tel										
Util							2			

The table reports the number of estimated links among industry groups on June 26, 2009.

Table 1.4: CENTRALITY ON 2009-06-26

Rank	Ticker	Sector
1	GOOG	Information Technology
2	MA	Information Technology
3	SLB	Energy
4	FCX	Materials
5	APA	Energy
6	COP	Energy
7	OXY	Energy
8	APC	Energy
9	DVN	Energy
10	CVX	Energy

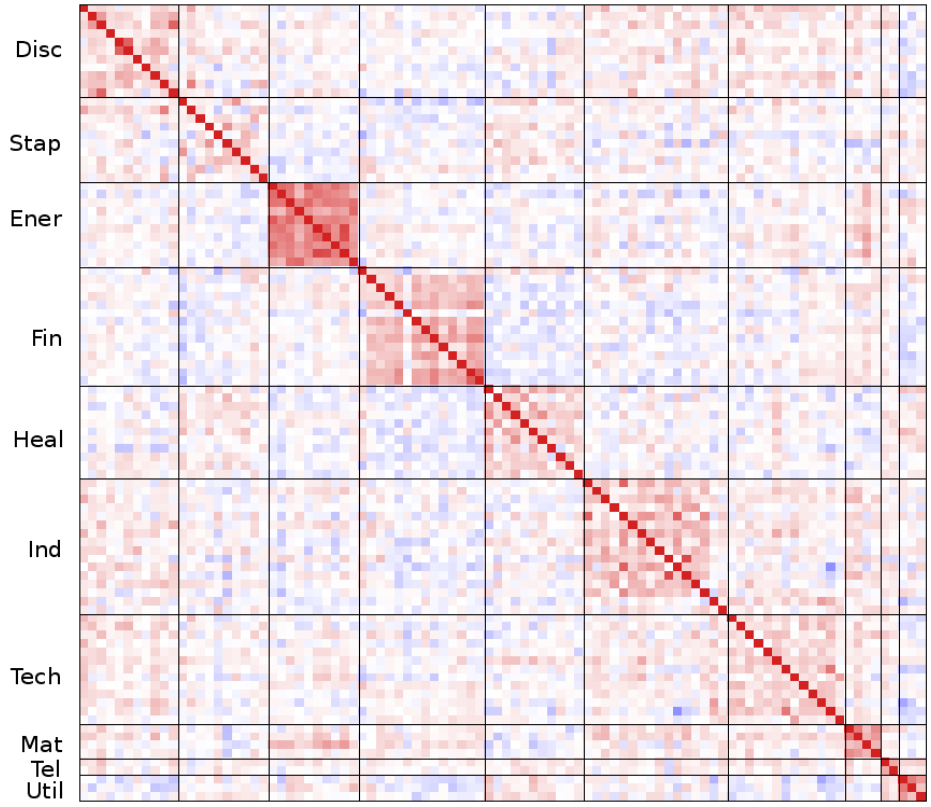
The table reports the top tickers by eigenvector centrality on June 26, 2009.

Table 1.5: GMV FORECASTING

	No regular	Diagonal	Network	Shrinkange	Factor	Block-Factor
RC	39.10	40.53	26.16	31.86	31.38	32.68
TSRC	37.22	41.41	26.22	29.38	30.60	31.58
MRK	32.83	35.51	24.52	27.81	28.49	28.02

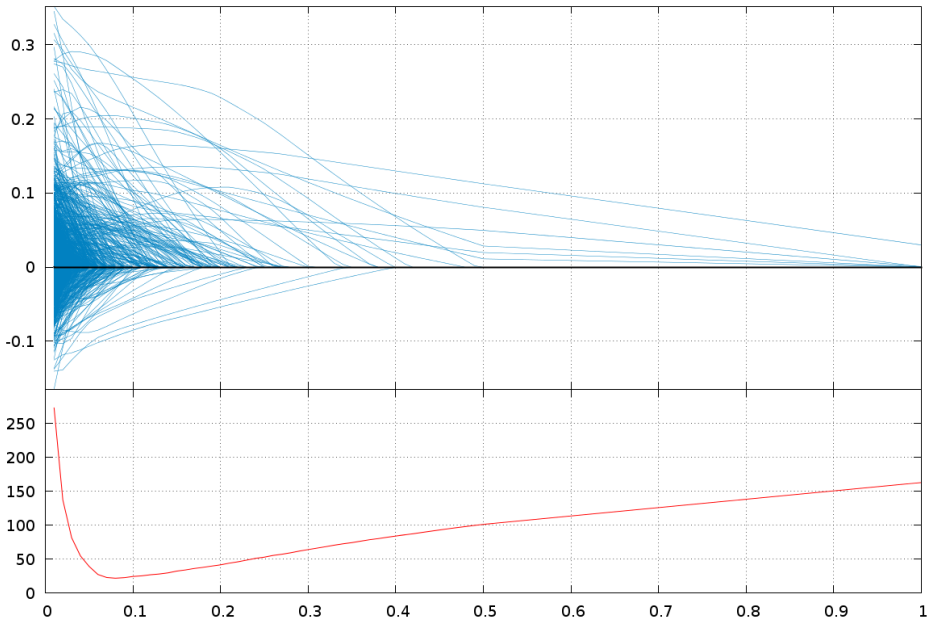
The table reports the results of the GMV forecasting comparison exercise. The table reports the annualized out of sample volatilities of the GMV portfolios constructed for the unregularized RC, TSRC and MRK estimators (No regular.) as well as their regularized versions (Diagonal, Network, Shrinkage, Factor, Block–Factor).

Figure 1.1: IDIOSYNCRATIC CORRELATION HEATMAP ON 2009-06-26



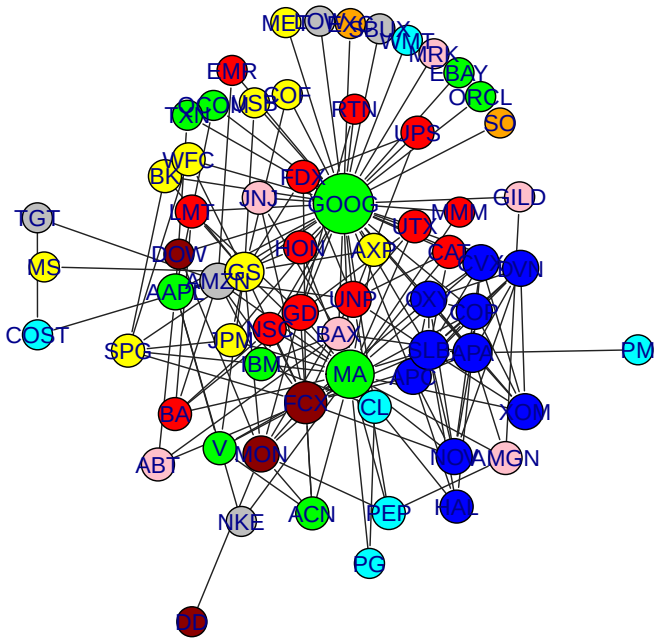
The figure shows the heatmap of the idiosyncratic realized correlation matrix on June 26, 2009 estimated using the TSRC estimator. Darker colors indicate higher correlations in absolute value.

Figure 1.2: TRACE ON 2009-06-26



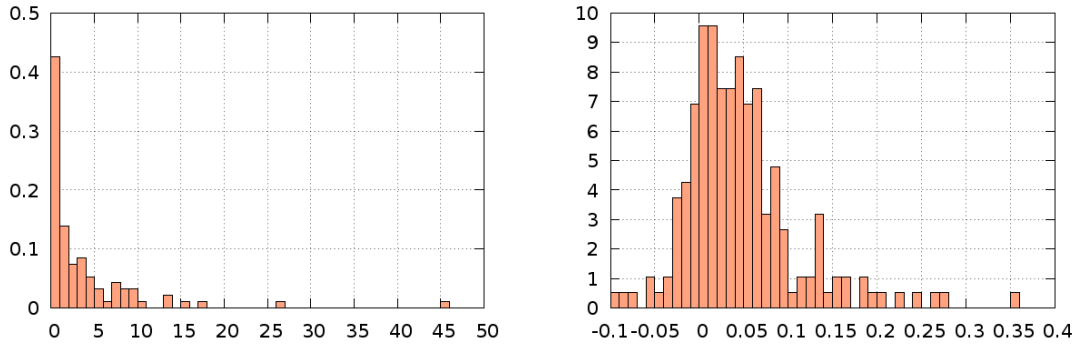
The figure shows the trace and BIC of the realized network estimator on June 26, 2009.

Figure 1.3: IDIOSYNCRATIC PARTIAL CORRELATION NETWORK ON 2009-06-26



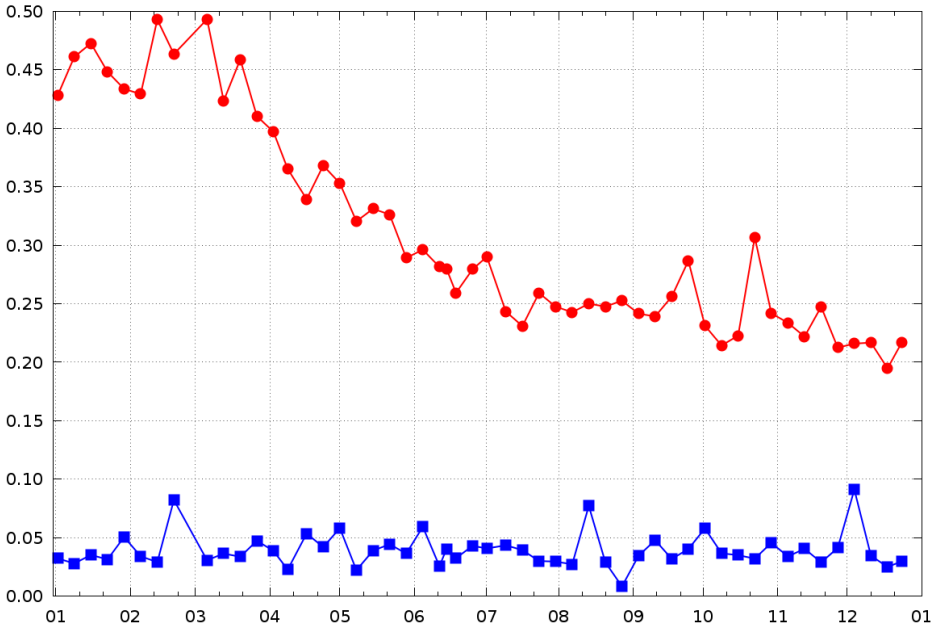
The figure shows the optimal realized network estimated on June 26, 2009.

Figure 1.4: DEGREE AND PARTIAL CORRELATION DISTRIBUTION ON 2009-06-26



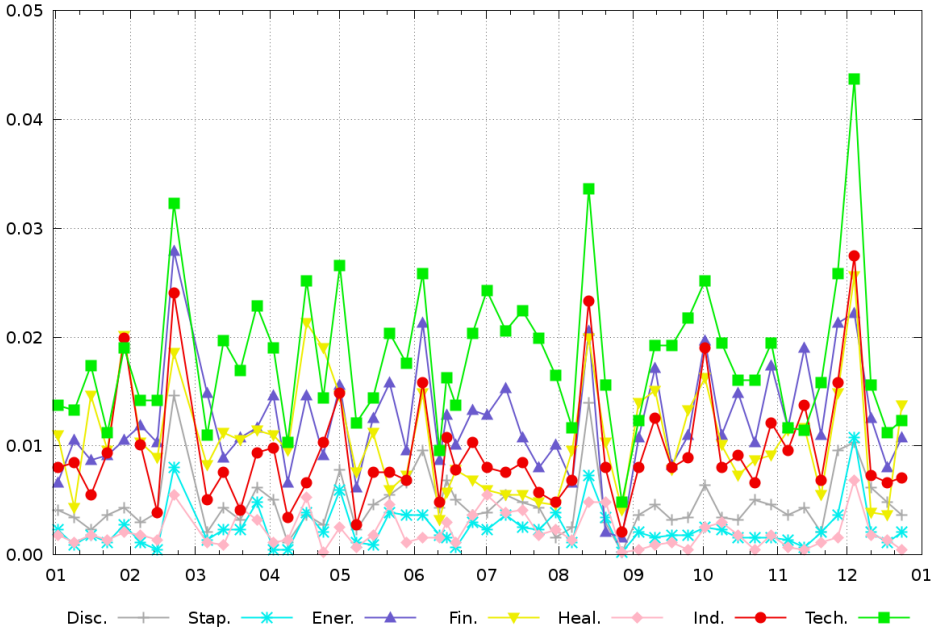
The figure shows the degree distribution and the distribution of partial correlations on June 26, 2009.

Figure 1.5: SPARSITY VS. VOLATILITY



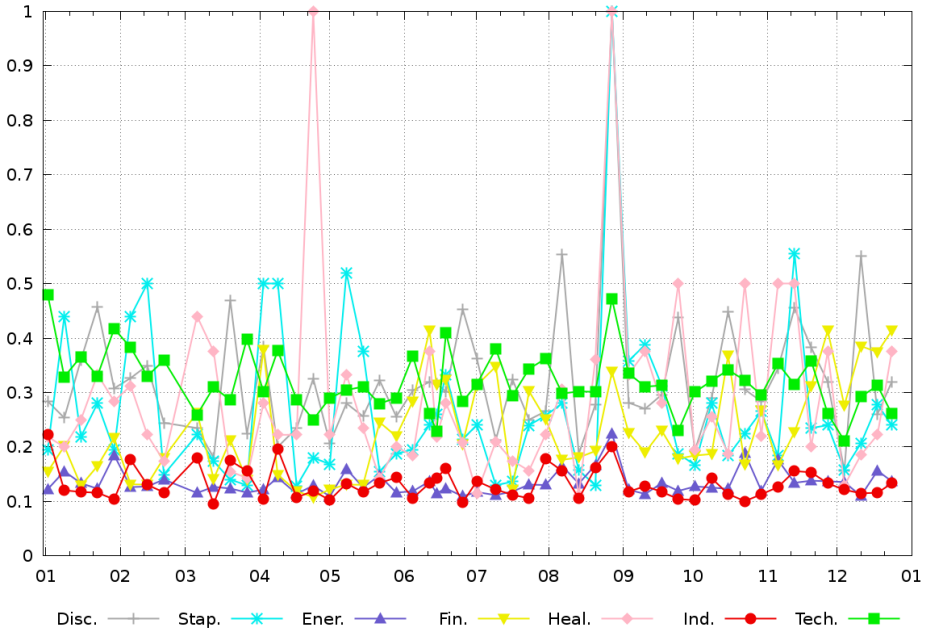
The figure shows the sparsity of the estimated network (square) vis–à–vis the level of volatility measured by the VIX (circle) for each week of 2009.

Figure 1.6: SECTORIAL LINKS



The figure shows the number of linkages of the different industry groups over the total number of possible linkages for each week of 2009.

Figure 1.7: SECTORIAL CONCENTRATION



The figure shows the link concentration (measured using the Herfindahl index) of the different industry for each week of 2009.

1.8.2 Algorithm

The GLASSO Algorithm

Banerjee and Ghaoui (2008) show that the optimization problem in (1.5) can be solved through a series of LASSO regressions. Define

$$\overline{\mathbf{\Sigma}} = \begin{bmatrix} \overline{\mathbf{\Sigma}}_{11} & \overline{\Sigma}_{12} \\ \overline{\Sigma}'_{12} & \overline{\Sigma}_{22} \end{bmatrix} \quad \text{and} \quad \widehat{\mathbf{K}}_{\lambda}^{-1} = \mathbf{W}_{\lambda} = \begin{bmatrix} \mathbf{W}_{\lambda\,11} & w_{\lambda\,12} \\ w'_{\lambda\,12} & w_{\lambda\,22} \end{bmatrix}, \tag{1.17}$$

where $\overline{\mathbf{\Sigma}}_{11}$ and $\mathbf{W}_{\lambda\,11}$ are $(n-1)\times(n-1)$ matrices, and $\overline{\Sigma}_{12}$ and $w_{\lambda\,12}$ are $(n-1)\times 1$ vectors. Suppose that that the values of $\mathbf{W}_{\lambda\,11}$ and $w_{\lambda\,22}$ are known and that we are concerned with finding the $w_{\lambda\,12}$ which minimizes (1.5). Banerjee and Ghaoui (2008) consider the following minimization problem

$$\min_{\beta}\left\{\frac{1}{2}\left(\mathbf{W}_{\lambda\,11}^{\frac{1}{2}}\beta - b\right)'\left(\mathbf{W}_{\lambda\,11}^{\frac{1}{2}}\beta - b\right) + \lambda||\beta||_1\right\}, \tag{1.18}$$

where $b = \mathbf{W}_{\lambda\,11}^{-\frac{1}{2}}\overline{\Sigma}_{12}$ and $||\beta||_1$ is the sum of absolute values of entries in β. The authors then show that (1.18) is the the dual problem of (1.5) in the sense that if β solves (1.18), then $w_{\lambda\,12} = \mathbf{W}_{\lambda\,11}\beta$ solves (1.5). Note that the optimization problem in (1.18) can be cast as a LASSO regression estimation problem.

These results motivate the following iterative algorithm to solve (1.5) (Friedman et al., 2007). Let $\widehat{\mathbf{W}}$ denote an initial estimate of $\mathbf{W}_{\lambda}$, for example $\widehat{\mathbf{W}} = \overline{\mathbf{\Sigma}} + \lambda\mathbf{I}_n$, where $\mathbf{I}_n$ is the identity matrix of order n. In each iteration of the algorithm, for each asset i we permute the rows and columns of $\widehat{\mathbf{W}}$ so that the n–the column/row of the matrix contains the covariances with respect to i. We then partition $\widehat{\mathbf{W}}$ as in (1.17) and update the values of $\widehat{w}_{12}$ and $\widehat{w}'_{12}$ by solving equation (1.18). The algorithm is iterated until convergence. Finally, we obtain the $\widehat{\mathbf{K}}_{\lambda}$ estimator by taking the inverse of $\mathbf{W}_{\lambda}$.

Pairwise Refresh Time Sampling

Consider two assets i and j observed at times $T_i = \{t_{i1}, \ldots, t_{i\,m_i}\}$ and $T_j = \{t_{j1}, \ldots, t_{j\,m_j}\}$ on $[0,1]$. The set of pairwise refresh time points is $T_{ij} = \left\{\tau_0, \tau_1, \ldots, \tau_{m-1}, \tau_m, \tau_{m+1}\right\}$, where $0 = \tau_0 < \tau_1 \ldots < \tau_{m-1} < \tau_m \leq \tau_{m+1} = 1$ and m is the amount of pairwise refresh time

observations for stocks i and j on $[0,1]$. For $0 < k \leq m$, τ_k is determined by

$$\tau_k = \max\left\{\min\left\{t \in T_i : t > \tau_{k-1}\right\}, \min\left\{t \in T_j : t > \tau_{k-1}\right\}\right\}.$$

The timestamps for assets i and j that correspond to the refresh time τ_k are respectively $t^r_{i,k} = \max\left\{t \in T_i : t \leq \tau_k\right\}$ and $t^r_{j,k} = \max\left\{t \in T_j : t \leq \tau_k\right\}$.

Eigenvalue Cleaning

The eigenvalue cleaning technique we use here is the same as that used in Hautsch et al. (2012), and it is introduced by Laloux et al. (1999). When the correlations between asset returns are zeros, according to the Marchenko-Pastur law, the maximum eigenvalues of the estimated correlation matrix of asset returns is $\lambda_{\max} = \sigma^2\left(1 + \frac{n}{m} + 2\sqrt{\frac{n}{m}}\right)$, as $m \to \infty$, where $\sigma = 1$ at this step. If the largest eigenvalue of the estimated correlation matrix is larger than $\lambda_{\max}$, we believe this is a market signal of price correlations. Then we reset the value of σ^2 as $\sigma^2 = 1 - \frac{\widehat{\lambda}_1}{n}$, and update the value of $\lambda_{\max}$ accordingly in order to detect smaller signals. In the same way we compare $\lambda_{\max}$ with the other estimated eigenvalues and update $\lambda_{\max}$ based on the comparison results. The eigenvalues which are larger than the updated $\lambda_{\max}$ are supposed to reflect market signals, and those smaller than $\lambda_{\max}$ are believed to be driven by the measurement noise on the correlation matrix. Thus we update the eigenvalues in the following way:

$$\widetilde{\lambda}_i = \begin{cases} \widehat{\lambda}_i & \text{if } \widehat{\lambda}_i > \lambda_{\max} \\ \delta & \text{otherwise,} \end{cases}$$

where

$$\delta = \frac{trace(\widehat{R}_+) - \sum_{(\widehat{\lambda}_i > \lambda_{\max})} \widehat{\lambda}_i}{n - (\text{ No. of } \widehat{\lambda}_i > \lambda_{\max})}$$

and $\widehat{R}_+$ is the positive semi-definite projection of the estimated correlation matrix. δ is defined in order to preserve the trace of the correlation matrix, and the regularized correlation matrix can be finally obtained by $\widetilde{R} = \widehat{Q}\widetilde{\Lambda}\widehat{Q}'$, where $\widehat{Q}$ is the eigenvector matrix of the original estimated correlation matrix, and $\widetilde{\Lambda} = diag(\widetilde{\lambda}_i)$.

Page-Rank Algorithm

Suppose there are n assets represented as $A_1, A_2, \ldots, A_n$, and we denote the page-rank index of the ith asset by PR_i. Then the indices of these assets can be obtained by solving the following equations:

$$\begin{cases} PR_1 = \left(\frac{PR_2\alpha_{1\,2}}{L_2} + \frac{PR_3\alpha_{1\,3}}{L_3} + \cdots + \frac{PR_n\alpha_{1\,n}}{L_n}\right) q + 1 - q \\ PR_2 = \left(\frac{PR_1\alpha_{1\,2}}{L_1} + \frac{PR_3\alpha_{2\,3}}{L_3} + \cdots + \frac{PR_n\alpha_{2\,n}}{L_n}\right) q + 1 - q \\ \cdots \\ PR_n = \left(\frac{PR_1\alpha_{1\,n}}{L_1} + \frac{PR_2\alpha_{2\,3}}{L_n} + \cdots + \frac{PR_{n-1\,n}\alpha_{n-1\,n}}{L_{n-1}}\right) q + 1 - q \end{cases}$$

Here $\alpha_{i\,j} = 1$ if there is a link between assets i and j, and $\alpha_{i\,j} = 0$ if there is no link. L_i is the number of links of asset i. q is a constant and its effect is to make the solutions nonzero.

1.8.3 Proofs

Following the same notation as in Ravikumar et al. (2011), we set

$$\widetilde{\mathbf{K}} = \arg \min_{\mathbf{K} \in \mathcal{S}^n, \mathbf{K}_{\mathcal{S}^c} = 0} \left\{ \text{tr}(\overline{\mathbf{\Sigma}}\mathbf{K}) - \log\det(\mathbf{K}) + \lambda \sum_{i \neq j} |k_{ij}| \right\}, \tag{1.19}$$

$\mathbf{W} = (w_{ij}) = \overline{\mathbf{\Sigma}} - \mathbf{\Sigma}^\star, \triangle = \widetilde{\mathbf{K}} - \mathbf{K}^\star, \boldsymbol{R}(\triangle) = \widetilde{\mathbf{K}} - \mathbf{K}^* + \mathbf{K}^\star \triangle \mathbf{K}^\star$.

We need the following lemma in order to prove Theorem 1 for $\alpha_0 = 1$.

LEMMA 1.1. *Assume that* $m \geq m_0 := \left(-\frac{2}{a_3 x^a} \log\left(\frac{a_3^{\frac{1}{a\beta}}}{a_2} x^{\frac{1}{\beta}} n^{-\tau}\right)\right)^{\frac{1}{a\beta}}$*, for some* $\tau > 2$*, where* a_i, a *and* β *are the constants in Assumption 1.1. Then for all* $x \in \left[0, \min\left(a_1, \frac{a_2^\beta}{a_3^{\frac{1}{a}}} \exp\left(-\frac{2}{a^2\beta}\right)\right)\right]$,

$$\mathrm{P}\left(||\mathbf{W}||_\infty \geq x\right) \leq \frac{1}{n^{\tau-2}}.$$

Proof. According to Assumption 1.1, when $x \in [0, a_1]$, for generic i, j we have

$$\mathrm{P}\left(|w_{ij}| \geq x\right) \leq a_2 m \exp\left(-a_3 \left(m^\beta x\right)^a\right) = P_1 P_2, \tag{1.20}$$

where $P_1 = \frac{a_2}{a_3^{\frac{1}{a\beta}}} x^{-\frac{1}{\beta}} \exp\left(-\frac{a_3\left(m^\beta x\right)^a}{2}\right)$, $P_2 = \left(a_3 m^{a\beta} x^a\right)^{\frac{1}{a\beta}} \exp\left(-\frac{a_3\left(m^\beta x\right)^a}{2}\right)$. P_1 is decreasing in m. Thus, when $m \geq m_0$, $P_1 \leq n^{-\tau}$. On the other hand, $a_3 m^{a\beta} x^a \geq -2\log\left(\frac{a_3^{\frac{1}{a\beta}}}{a_2} x^{\frac{1}{\beta}} n^{-\tau}\right)$.

Thus, when $x \leq \frac{a_2^{\beta}}{a_3^{\frac{1}{a}}} \exp\left(-\frac{2}{a^2\beta}\right)$, $a_3 m^{a\beta} x^a \geq -2\log\left(\frac{a_3^{\frac{1}{a\beta}}}{a_2} x^{\frac{1}{\beta}}\right) \geq \frac{4}{a^2\beta^2}$.

Consider the function $f(y) = \left(y \exp\left(-\frac{a\beta y}{2}\right)\right)^{\frac{1}{a\beta}}$. It is easy to see that when $y \geq \frac{4}{a^2\beta^2}$, $f(y) \leq 1$. As $P_2 = f\left(a_3 m^{a\beta} x^a\right)$, and $a_3 m^{a\beta} x^a \geq \frac{4}{a^2\beta^2}$, we get that $\mathrm{P}_2 \leq 1$. Since $P_1 \leq n^{-\tau}$, we obtain that

$$\mathrm{P}\left(|w_{ij}| \geq x\right) \leq n^{-\tau}.$$

Using this inequality over all n^2 entries of $\mathbf{W}$, we conclude that

$$\mathrm{P}\left(||W||_\infty \geq x\right) = \mathrm{P}\left(\max_{i,j} |w_{ij}| \geq x\right) \leq n^{2-\tau}.$$

□

Proof of Theorem 1.1. First we prove (a). Set $h(x) = a_2 m \exp\left(-a_3\left(m^{\beta}x\right)^a\right)$ for $x > 0$, and $a_0 := \min\left(a_1, \frac{a_2^{\beta}}{a_3^{\frac{1}{a}}} \exp\left(-\frac{2}{a^2\beta}\right)\right)$. In the proof of Lemma 1, we have shown that when $m \geq m_0$, which is true according to (3.2) and (1.7), $h(a_0) \leq n^{-\tau}$. On the other hand, $h\left(\frac{\alpha}{8}\lambda\right) = n^{-\tau}$. Because the function h is decreasing, we get that $\frac{\alpha}{8}\lambda \leq a_0$. Therefore, by Lemma 1.1, we conclude that

$$\mathrm{P}\left(||\mathbf{W}||_\infty \leq \frac{\alpha}{8}\lambda\right) \geq 1 - \frac{1}{n^{\tau-2}}. \tag{1.21}$$

By the same argument, if $\sigma_m \leq a_0$, as $m \geq \left(-\frac{2}{a_3\sigma_m^a}\log\left(\frac{a_3^{\frac{1}{a\beta}}}{a_2}\sigma_m^{\frac{1}{\beta}} n^{-\tau}\right)\right)^{\frac{1}{a\beta}}$, we have $h(\sigma_m) \leq n^{-\tau}$, so $\frac{\alpha\lambda}{8} \leq \sigma_m$. When $\sigma_m > a_0$, we still have $\frac{\alpha\lambda}{8} \leq \sigma_m$, since $\frac{\alpha\lambda}{8} \leq a_0$. Therefore,

$$\mathrm{P}\left(||\mathbf{W}||_\infty \leq \sigma_m\right) \geq \mathrm{P}\left(||\mathbf{W}||_\infty \leq \frac{\alpha}{8}\lambda\right) \geq 1 - \frac{1}{n^{\tau-2}}, \tag{1.22}$$

and when the event $A := \left\{||\mathbf{W}||_\infty \leq \frac{\alpha}{8}\lambda = m^{-\beta}\left(\frac{\log(a_2 m n^{\tau})}{a_3}\right)^{\frac{1}{a}}\right\}$ holds, $\overline{\mathbf{\Sigma}}$ has positive diagonal entries, and by Lemma 3 in Ravikumar et al. (2011), $\widehat{\mathbf{K}}$ is unique. Proceeding as before, we can obtain that $\left(6\left(1+8\alpha^{-1}\right)^2 d \max\left(C_{\mathbf{\Sigma}^\star} C_{\mathbf{\Gamma}^\star}, C_{\mathbf{\Sigma}^\star}^3 C_{\mathbf{\Gamma}^\star}^2\right\}\right)^{-1} \geq \frac{\alpha\lambda}{8}$. Thus, we get that, if A holds, then

$$2C_{\mathbf{\Gamma}^\star}(||\mathbf{W}||_\infty + \lambda) \leq 2C_{\mathbf{\Gamma}^\star}\left(1+8\alpha^{-1}\right) m^{-\beta}\left(\frac{\log(a_2 m n^{\tau})}{a_3}\right)^{\frac{1}{a}} \leq \left(3d \max\left(C_{\mathbf{\Sigma}^\star}, C_{\mathbf{\Sigma}^\star}^3 C_{\mathbf{\Gamma}^\star}\right\}\right)^{-1}.$$

Therefore, by Lemma 6 in Ravikumar et al. (2011), it holds that

$$||\triangle||_\infty \leq 2C_{\mathbf{\Gamma}^\star}\left(1+8\alpha^{-1}\right)m^{-\beta}\left(\frac{\log(a_2mn^\tau)}{a_3}\right)^{\frac{1}{a}}.$$

Now, appealing to Lemma 5 in Ravikumar et al. (2011), we obtain that

$$\begin{aligned}||\boldsymbol{R}(\triangle)||_\infty &\leq \frac{3}{2}d||\triangle||_\infty^2 C_{\mathbf{\Sigma}^\star}^3 \leq 6C_{\mathbf{\Sigma}^\star}^3 C_{\mathbf{\Gamma}^\star}^2 d\left(1+8\alpha^{-1}\right)^2\left(m^{-\beta}\left(\frac{\log(a_2mn^\tau)}{a_3}\right)^{\frac{1}{a}}\right)^2 \\ &= 6C_{\mathbf{\Sigma}^\star}^3 C_{\mathbf{\Gamma}^\star}^2 d\left(1+8\alpha^{-1}\right)^2 m^{-\beta}\left(\frac{\log(a_2mn^\tau)}{a_3}\right)^{\frac{1}{a}}\frac{\alpha\lambda}{8} \leq \frac{\alpha\lambda}{8}.\end{aligned}$$

By Lemma 4 in Ravikumar et al. (2011), we conclude that $\widetilde{\mathbf{K}} = \widehat{\mathbf{K}}$. Thus, $||\widehat{\mathbf{K}} - \mathbf{K}^\star||_\infty$ satisfies the same bound as $||\triangle||_\infty$.

Now let us prove (b). Using the same argument as before, it is easy to see hat

$$\frac{k_m}{2C_{\mathbf{\Gamma}^\star}\left(1+8\alpha^{-1}\right)} \geq \frac{\alpha\lambda}{8} = m^{-\beta}\left(\frac{\log(a_2mn^\tau)}{a_3}\right)^{\frac{1}{a}}.$$

Then we have

$$k_m \geq 2C_{\mathbf{\Gamma}^\star}\left(1+8\alpha^{-1}\right)m^{-\beta}\left(\frac{\log(a_2mn^\tau)}{a_3}\right)^{\frac{1}{a}} \geq 2C_{\mathbf{\Gamma}^\star}(||\mathbf{W}||_\infty + \lambda).$$

Finally appealing to Lemma 7 in Ravikumar et al. (2011), the proof is completed. □

Proof of Corollary 3.1. Notice that $\mathbf{\Sigma}^*_{FF} = \sigma^*_{n+1,n+1}$ and $\overline{\mathbf{\Sigma}}_{FF} = \overline{\sigma}_{n+1,n+1}$. Then for all $i, j \in \mathcal{V}$,

$$\sigma^*_{I\,ij} = \sigma^*_{ij} - \frac{\sigma^*_{n+1,i}\sigma^*_{j,n+1}}{\sigma^*_{n+1,n+1}} \quad \text{and} \quad \overline{\sigma}_{I\,ij} = \overline{\sigma}_{ij} - \frac{\overline{\sigma}_{n+1,i}\overline{\sigma}_{j,n+1}}{\overline{\sigma}_{n+1,n+1}}.$$

Therefore,

$$\mathrm{P}(\left|\overline{\sigma}_{I\,ij} - \sigma^*_{I\,ij}\right| > x) \leq \mathrm{P}\left(\left|\sigma^*_{ij} - \overline{\sigma}_{ij}\right| \geq \frac{x}{2}\right) + \mathrm{P}\left(\left|\frac{\overline{\sigma}_{n+1,i}\overline{\sigma}_{j,n+1}}{\overline{\sigma}_{n+1,n+1}} - \frac{\sigma^*_{n+1,i}\sigma^*_{j,n+1}}{\sigma^*_{n+1,n+1}}\right| \geq \frac{x}{2}\right).$$

To the first term, we can apply the concentration inequality (3.22). In order to bound the second

term, set $b_1 = \overline{\sigma}_{n+1,n+1} - \sigma^*_{n+1,n+1}$, $b_2 = \overline{\sigma}_{n+1,i} - \sigma^*_{n+1,i}$, and $b_3 = \overline{\sigma}_{j,n+1} - \sigma^*_{j,n+1}$. Then,

$$\begin{aligned}&\left|\frac{\overline{\sigma}_{n+1,i}\overline{\sigma}_{j,n+1}}{\overline{\sigma}_{n+1,n+1}} - \frac{\sigma^*_{n+1,i}\sigma^*_{j,n+1}}{\sigma^*_{n+1,n+1}}\right| \\ &= \left|\frac{b_3\sigma^*_{n+1,n+1}\sigma^*_{n+1,i} + b_2\sigma^*_{n+1,n+1}\sigma^*_{j,n+1} + b_2b_3\sigma^*_{n+1,n+1} - b_1\sigma^*_{n+1,i}\sigma^*_{j,n+1}}{\sigma^*_{n+1,n+1}(\sigma^*_{n+1,n+1} + b_1)}\right|.\end{aligned} \tag{1.23}$$

Without loss of generality, we assume that $\sigma^*_{j,n+1}$ and $\sigma^*_{n+1,i}$ are non-zero. Otherwise, the proof follows similarly. Then,

$$\begin{aligned}\mathrm{P}\left(\left|\frac{\overline{\sigma}_{n+1,i}\overline{\sigma}_{j,n+1}}{\overline{\sigma}_{n+1,n+1}} - \frac{\sigma^*_{n+1,i}\sigma^*_{j,n+1}}{\sigma^*_{n+1,n+1}}\right| \geq 8x\right) &\leq \mathrm{P}\left(|b_1| \geq \min\left(\frac{\sigma^*_{n+1,n+1}}{2}, \frac{\sigma^{*2}_{n+1,n+1}x}{|\sigma^*_{n+1,i}\sigma^*_{j,n+1}|}\right)\right) \\ &+ \mathrm{P}\left(|b_2| \geq \min\left(\frac{\sigma^*_{n+1,n+1}x}{|\sigma^*_{j,n+1}|}, 1\right)\right) + \mathrm{P}\left(|b_3| \geq \frac{\sigma^*_{n+1,n+1}x}{|\sigma^*_{n+1,i}|}\right).\end{aligned}$$

Observe that in the last inequality we have used the fact that if $|b_1| < \frac{\sigma^*_{n+1,n+1}}{2}$, then we can lower bound the denominator in (1.23) by $\frac{\sigma^{*2}_{n+1,n+1}}{2}$. Then appealing again to the concentration inequality (3.22), we conclude the desired result. □

Proof of Theorem 1.2. We start proving the concentration inequality for a generic asset i. To simplify the exposition, we assume that the drift is zero in (2.1), otherwise it is easy to see that the same result follows. By definition (1.10), the realized kernel estimator of σ^*_{ii} equals

$$\overline{\sigma}_{\mathsf{K}\,ii} = \gamma_0(y,y) + A + B + C, \tag{1.24}$$

where

$$\begin{aligned}A &= \sum_{h=1}^{H} k\left(\frac{h-1}{H}\right)(\gamma_h(y,y) + \gamma_{-h}(y,y)), \\ B &= 2\gamma_0(y,u) + \sum_{h=1}^{H} k\left(\frac{h-1}{H}\right)(\gamma_h(y,u) + \gamma_h(u,y) + \gamma_{-h}(y,u) + \gamma_{-h}(u,y)), \\ C &= \gamma_0(u,u) + \sum_{h=1}^{H} k\left(\frac{h-1}{H}\right)(\gamma_h(u,u) + \gamma_{-h}(u,u)),\end{aligned}$$

and for each $h \in \{-H, \ldots, H\}$, $\gamma_h(y,u) = \sum_{j=1}^{m-1}(y_j - y_{j-1})(u_{j-h} - u_{j-h-1})$.

By Lemma 3 in Fan et al. (2012), there exist constants $c_1, c_2 > 0$ such that for large m and

$x \in [0, c_1]$,

$$\mathrm{P}\left(|\gamma_0(y,y) - \sigma_{ii}^*| \geq x\right) \leq 2\exp(-c_2 m x^2). \tag{1.25}$$

Therefore, by (3.9) and (3.11), it suffices to bound the terms A, B and C in probability.

Bound in probability of the term A in (3.9). We decompose A as

$$\begin{aligned} A &= \sum_{j=1}^{m-1} \int_{t_{j-1}}^{t_j} \theta(u) dB(u) \sum_{h=1}^{H} k\left(\frac{h-1}{H}\right) \left(\int_{t_{j-h-1}}^{t_{j-h}} \theta(u) dB(u) + \int_{t_{j+h-1}}^{t_{j+h}} \theta(u) dB(u) \right) \\ &= A_1 + A_2. \end{aligned}$$

We start bounding A_1. For any $x > 0$ and $\alpha > 0$, we have

$$\begin{aligned} \mathrm{P}\left(|A_1| \geq \frac{x}{2}\right) \leq \mathrm{P}\left(\left| \sum_{j=1}^{m-1} \int_{t_{j-1}}^{t_j} \theta(u) dB(u) D_j \mathbf{1}_{\left\{ \sup_{j \in \{1,\ldots,m-1\}} |D_j| \leq m^{-\frac{1}{4}+\alpha} \right\}} \right| \geq \frac{x}{4} \right) \\ + \mathrm{P}\left(\sup_{j \in \{1,\ldots,m-1\}} |D_j| > m^{-\frac{1}{4}+\alpha} \right), \end{aligned}$$

where

$$D_j = \sum_{h=1}^{H} k\left(\frac{h-1}{H}\right) \int_{t_{j-h-1}}^{t_{j-h}} \theta(u) dB(u).$$

Since k and θ are bounded, it holds that

$$\sum_{h=1}^{H} k^2\left(\frac{h-1}{H}\right) \int_{t_{j-h-1}}^{t_{j-h}} |\theta(u)|^2 du \leq cH \frac{C_\triangle}{m} = c m^{-\frac{1}{2}}.$$

Therefore, appealing to the exponential martingale inequality, we get that

$$\mathrm{P}\left(\sup_{j \in \{1,\ldots,m-1\}} |D_j| > m^{-\frac{1}{4}+\alpha} \right) \leq 2m \exp\left(-c m^{2\alpha}\right).$$

On the other hand, since

$$\sum_{j=1}^{m-1} \int_{t_{j-1}}^{t_j} |\theta(u)|^2 du D_j^2 \mathbf{1}_{\left\{ \sup_{j \in \{1,\ldots,m-1\}} |D_j| \leq m^{-\frac{1}{4}+\alpha} \right\}} \leq c m^{2\alpha - \frac{1}{2}},$$

again using the exponential martingale inequality, we obtain that for all $x > 0$,

$$\mathrm{P}\left(\left|\sum_{j=1}^{m-1}\int_{t_{j-1}}^{t_j}\theta(u)dB(u)D_j\mathbf{1}_{\left\{\sup_{j\in\{1,\ldots,m-1\}}|D_j|\leq m^{-\frac{1}{4}+\alpha}\right\}}\right|\geq\frac{x}{4}\right)\leq 2\exp\left(-cm^{\frac{1}{2}-2\alpha}x^2\right).$$

Therefore, choosing $\alpha > 0$ such that $m^{2\alpha} = m^{\frac{1}{4}}x$, we conclude that

$$\mathrm{P}\left(|A_1|\geq\frac{x}{2}\right)\leq 2m\exp\left(-cm^{\frac{1}{4}}x\right). \tag{1.26}$$

We next bound A_2. We write $A_2 = A_{21} + A_{22} + A_{23}$, where

$$\begin{aligned}
A_{21} &= \sum_{i=1}^{H-1}\int_{t_i}^{t_{i+1}}\theta(u)dB(u)\sum_{j=0}^{i-1}k\left(\frac{i-j}{H}\right)\int_{t_j}^{t_{j+1}}\theta(u)dB(u)\\
A_{22} &= \sum_{i=H}^{m-1}\int_{t_i}^{t_{i+1}}\theta(u)dB(u)\sum_{j=i-H}^{i-1}k\left(\frac{i-j}{H}\right)\int_{t_j}^{t_{j+1}}\theta(u)dB(u)\\
A_{23} &= \sum_{i=m}^{m+H-1}\int_{t_i}^{t_{i+1}}\theta(u)dB(u)\sum_{j=i-H}^{m-2}k\left(\frac{i-j}{H}\right)\int_{t_j}^{t_{j+1}}\theta(u)dB(u).
\end{aligned}$$

We start bounding A_{22}. For a generic i, we have that

$$\sum_{j=i-H}^{i-1}\int_{t_j}^{t_{j+1}}k^2\left(\frac{i-j}{H}\right)|\theta(u)|^2du\leq cm^{-\frac{1}{2}}.$$

Thus, according to the exponential martingale inequality, for any $\alpha > 0$, we obtain that

$$\mathrm{P}\left(\sup_{i\in\{H,\ldots,m-1\}}|F_i| > m^{\alpha-\frac{1}{4}}\right)\leq 2m\exp\left(-cm^{2\alpha}\right),$$

where $F_i = \sum_{j=i-H}^{i-1}k\left(\frac{i-j}{H}\right)\int_{t_j}^{t_{j+1}}\theta(u)dB(u)$. Therefore, using again the exponential martingale inequality, we conclude that for any $x > 0$ and $\alpha > 0$,

$$\begin{aligned}
&\mathrm{P}\left(|A_{22}|\geq\frac{x}{6}\right)\leq\mathrm{P}\left(\left|\sum_{i=H}^{m-1}\int_{t_i}^{t_{i+1}}\theta(u)F_i\mathbf{1}_{\left\{\sup_{i\in\{H,\ldots,m-1\}}|F_i|\leq m^{\alpha-\frac{1}{4}}\right\}}dB(u)\right|\geq\frac{x}{12}\right)\\
&+\mathrm{P}\left(\sup_{i\in\{H,\ldots,m-1\}}|F_i| > m^{\alpha-\frac{1}{4}}\right)\leq 2\exp\left(-cm^{\frac{1}{2}-2\alpha}x^2\right)+2m\exp\left(-cm^{2\alpha}\right),
\end{aligned}$$

since

$$\sum_{i=H}^{m-1}\int_{t_i}^{t_{i+1}}|\theta(u)|^2F_i^2\mathbf{1}_{\left\{\sup_{i\in\{H,\dots,m-1\}}|F_i|\leq m^{\alpha-\frac{1}{4}}\right\}}du\leq cm^{2\alpha-\frac{1}{2}}.$$

Thus, choosing α as in (1.26), we conclude that $A_{2,2}$ also satisfies (1.26). Similarly, so do $A_{2,1}$, $A_{2,3}$, and A.

Bound in probability of the term B in (3.9). We decompose B as $B=B_1+B_2$, where

$$B_1=\gamma_0(y,u)+\sum_{h=1}^{H}k\left(\frac{h-1}{H}\right)(\gamma_h(y,u)+\gamma_{-h}(y,u))$$
$$B_2=\gamma_0(u,y)+\sum_{h=1}^{H}k\left(\frac{h-1}{H}\right)(\gamma_h(u,y)+\gamma_{-h}(u,y)).$$

We start bounding B_1. We write

$$B_1=\sum_{j=1}^{m-1}\int_{t_{j-1}}^{t_j}\theta(u)dB(u)G_j,$$

where

$$G_j=\sum_{h=1}^{H}k\left(\frac{h-1}{H}\right)(u_{j-h}-u_{j-h-1}+u_{j+h}-u_{j+h-1})+u_j-u_{j-1}. \tag{1.27}$$

Observe that since $k(0)=1,k(1)=0$, we have

$$G_j\sim N\left(0,2\eta^2\sum_{h=1}^{H}\left(k\left(\frac{h}{H}\right)-k\left(\frac{h-1}{H}\right)\right)^2\right).$$

Moreover, there exists $\beta_h\in\left[\frac{h-1}{H},\frac{h}{H}\right]$ such that

$$\sum_{h=1}^{H}\left(k\left(\frac{h}{H}\right)-k\left(\frac{h-1}{H}\right)\right)^2=\frac{1}{H}\sum_{h=1}^{H}\frac{1}{H}k'(\beta_h)^2\leq cm^{-\frac{1}{2}}.$$

Therefore, for all $\alpha>0$,

$$\mathrm{P}\left(\sup_{j\in\{1,\dots,m-1\}}|G_j|>m^{\alpha-\frac{1}{4}}\right)>2m\exp\left(-cm^{2\alpha}\right).$$

Consequently, by the exponential martingale inequality, for any $x > 0$,

$$\mathrm{P}\left(|B_1| \geq x\right) \leq \mathrm{P}\left(\left|\sum_{j=1}^{m-1} \int_{t_{j-1}}^{t_j} \theta(u) dB(u) G_j \mathbf{1}_{\left\{\sup_{j\in\{1,\ldots,m-1\}} |G_j| \leq m^{\alpha-\frac{1}{4}}\right\}}\right| \geq x\right)$$
$$+ \mathrm{P}\left(\sup_{j\in\{1,\ldots,m-1\}} |G_j| > m^{\alpha-\frac{1}{4}}\right) \leq 2\exp\left(-cm^{\frac{1}{2}-2\alpha}x^2\right) + 2m\exp\left(-cm^{2\alpha}\right),$$

since

$$\sum_{i=H}^{m-1} \int_{t_i}^{t_{i+1}} |\theta(u)|^2 G_j^2 \mathbf{1}_{\left\{\sup_{j\in\{1,\ldots,m-1\}} |G_j| \leq m^{\alpha-\frac{1}{4}}\right\}} du \leq cm^{2\alpha-\frac{1}{2}}.$$

Thus, choosing α as in (1.26), we conclude that B_1 also satisfies (1.26).

We next bound B_2. We write

$$B_2 = \sum_{j=1}^{m-1} (u_j - u_{j-1}) S_j = -S_1 u_0 + \sum_{j=2}^{m-1} (S_{j-1} - S_j) u_{j-1} + S_{m-1} u_{m-1},$$

where

$$S_j = \sum_{h=1}^{H} k\left(\frac{h-1}{H}\right) \left(\int_{t_{j-h-1}}^{t_{j-h}} \theta(u) dB(u) + \int_{t_{j+h-1}}^{t_{j+h}} \theta(u) dB(u)\right) + \int_{t_{j-1}}^{t_j} \theta(u) dB(u).$$

Now, for all $x > 0$ and $\alpha > 0$, we write

$$\mathrm{P}\left(|B_2| \geq x\right) \leq \mathrm{P}\left(\left|B_2 \mathbf{1}_{\left\{E \leq m^{2\alpha-\frac{1}{2}}\right\}}\right| \geq \frac{x}{2}\right) + \mathrm{P}\left(E > m^{2\alpha-\frac{1}{2}}\right),$$

where $E = S_1^2 + \sum_{j=2}^{m-1} \left(S_{j-1} - S_j\right)^2 + S_{m-1}^2$.

Since the random variables $u_0, u_1, ..., u_m$ are iid centered Gaussian with variance η^2,

$$\mathrm{P}\left(\left|B_2 \mathbf{1}_{\left\{E \leq m^{2\alpha-\frac{1}{2}}\right\}}\right| \geq \frac{x}{2}\right) \leq 2\exp\left(cm^{\frac{1}{2}-2\alpha}x^2\right).$$

In order to bound the second term, we observe that

$$E = S_1(2S_1 - S_2) + \sum_{j=2}^{m-2} S_j(2S_j - S_{j-1} - S_{j+1}) + S_{m-1}(2S_{m-1} - S_{m-2}).$$

Thus,

$$\mathrm{P}\left(E > 7m^{2\alpha-\frac{1}{2}}\right) \leq \mathrm{P}\left(\sup_{j\in\{2,\dots,m-2\}} |2S_j - S_{j-1} - S_{j+1}| > m^{\alpha-\frac{5}{4}}\right) + \mathrm{P}\left(\sup_{j\in\{1,\dots,m-1\}} |S_j| > m^{\alpha-\frac{1}{4}}\right).$$

Straightforward computations show that both terms are bounded by $cm\exp\left(-cm^{2\alpha}\right)$. Therefore, choosing α as in (1.26), we conclude that B_2, and thus B also satisfies (1.26).

Bound in probability of the term C in (3.9). We write

$$C = -u_0 e_1 + \sum_{j=1}^{m-2}(e_j - e_{j+1})u_j + e_{m-1}u_{m-1},$$

where $e_j = \sum_{h=2}^{H} k\left(\frac{h-1}{H}\right)(u_{j-h} - u_{j-h-1} + u_{j+h} - u_{j+h-1}) + u_{j+1} - u_{j-2}$.

For all $x > 0$ and $\alpha > 0$, we write

$$\mathrm{P}\left(|C| \geq x\right) \leq \mathrm{P}\left(\left|C\mathbf{1}_{\left\{F\leq m^{2\alpha-\frac{1}{2}}\right\}}\right| \geq \frac{x}{2}\right) + \mathrm{P}\left(F > m^{2\alpha-\frac{1}{2}}\right),$$

where $F = e_1^2 + \sum_{j=2}^{m-1}(e_j - e_{j-1})^2 + e_{m-1}^2$.

Again, straightforward computations show that

$$\mathrm{P}\left(F > 3m^{2\alpha-\frac{1}{2}}\right) \leq \mathrm{P}\left(\sup_{j\in\{2,\dots,m-1\}}(e_j - e_{j-1})^2 > m^{2\alpha-\frac{3}{2}}\right) + \mathrm{P}\left(e_1^2 > m^{2\alpha-\frac{1}{2}}\right) + \mathrm{P}\left(e_{m-1}^2 > m^{2\alpha-\frac{1}{2}}\right) \leq cm\exp\left(-cm^{2\alpha}\right).$$

Hence, the rest of the proof follows as for B_2, which shows that C also satisfies (1.26).

Conclusion. The bounds obtained for A, B, and C, together with (3.9) and (3.11), show that there exist constant $c_1, c_2, c_3, c_4 > 0$ such that for large m and $x \in [0, c_1]$,

$$\mathrm{P}\left(|\overline{\sigma}_{\mathsf{K}\,ii} - \sigma_{ii}^*| \geq x\right) \leq 2\exp(-c_2 m x^2) + c_3 m\exp\left(-c_4 m^{\frac{1}{4}}x\right).$$

Now, when $m^{-\frac{3}{4}} < x \leq c_1$, we have $x^2 m \geq x m^{\frac{1}{4}}$, and thus, the second term wins. On the other

hand, when $0 \leq x \leq m^{-\frac{3}{4}}$, we have that

$$m \exp\left(-c_4 x m^{\frac{1}{4}}\right) \geq m \exp\left(-c_4 m^{-\frac{1}{2}}\right) > 1,$$

and thus the same bound holds. This concludes the proof of the theorem for the diagonal terms.

We are now going to prove the concentration inequality for two generic assets i and j. By (2.1), the integrated covariance of the log-returns of assets i and j on $[0,1]$ is given by

$$\begin{aligned}\int_0^1 \Theta_i(u)\Theta_j(u)'du = \frac{1}{4}\left(\int_0^1 \left(\Theta_i(u)+\Theta_j(u)\right)\left(\Theta_i(u)+\Theta_j(u)\right)' du\right) \\ - \frac{1}{4}\left(\int_0^1 \left(\Theta_i(u)-\Theta_j(u)\right)\left(\Theta_i(u)-\Theta_j(u)\right)' du\right),\end{aligned} \tag{1.28}$$

where $\Theta_k(u)$, $k \in \{1, ..., n\}$ denote the rows of the matrix $\Theta(u)$. Therefore, in order to estimate the integrated covariance, it suffices to estimate the two terms in (1.28).

The MRK estimator of $s^*_{ij} = \int_0^1 \left(\Theta_i(u)+\Theta_j(u)\right)\left(\Theta_i(u)+\Theta_j(u)\right)' du$ is given by

$$s_{\mathsf{K}\,ij} = \gamma_0(x_i^r + x_j^r) + \sum_{h=1}^{H} k\left(\frac{h-1}{H}\right)(\gamma_h(x_i^r + x_j^r) + \gamma_{-h}(x_i^r + x_j^r)),$$

where $\gamma_h(x_i^r + x_j^r) = \gamma_h(x_i^r + x_j^r, x_i^r + x_j^r)$, for all $h \in \{-H, \ldots, H\}$.

Observe that if there is no asynchronicity among the observations, then we can derive the concentration inequality for $s_{\mathsf{K}\,ij}$ as we did for the case of a single asset. Therefore, it suffices to split the estimator as $s_{\mathsf{K}\,ij} = \overline{s}_{\mathsf{K}\,ij} + \sum_{q=1}^{10} F_q$, where $\overline{s}_{\mathsf{K}\,ij}$ is the analogous of $s_{\mathsf{K}\,ij}$ but replacing $x_i^r + x_j^r$ by $\overline{x}_i + \overline{x}_j$, where $\overline{x}_{\ell k} := y_\ell(\tau_k) + u^r_{\ell k}$, for $\ell = i, j$ and $k \in \{1, ..., m\}$. The terms F_q, are those that contain the points $\triangle y_{\ell\, k} := y_\ell(\tau_k) - y^r_{\ell k}$, that is,

$$\begin{aligned}
F_1 &= \gamma_0(\triangle y_i, \overline{x}_i + \overline{x}_j) + \sum_{h=1}^{H} k\left(\frac{h-1}{H}\right)(\gamma_h(\triangle y_i, \overline{x}_i + \overline{x}_j) + \gamma_{-h}(\triangle y_i, \overline{x}_i + \overline{x}_j)),\\
F_3 &= \gamma_0(\overline{x}_i + \overline{x}_j, \triangle y_i) + \sum_{h=1}^{H} k\left(\frac{h-1}{H}\right)(\gamma_h(\overline{x}_1 + \overline{x}_2, \triangle y_i) + \gamma_{-h}(\overline{x}_i + \overline{x}_j, \triangle y_i)),\\
F_5 &= \gamma_0(\triangle y_i, \triangle y_j) + \sum_{h=1}^{H} k\left(\frac{h-1}{H}\right)(\gamma_h(\triangle y_i, \triangle y_j) + \gamma_{-h}(\triangle y_i, \triangle y_j)),\\
F_7 &= \gamma_0(\triangle y_i, u_i^r + u_j^r) + \sum_{h=1}^{H} k\left(\frac{h-1}{H}\right)(\gamma_h(\triangle y_i, u_i^r) + \gamma_{-h}(\triangle y_i, u_i^r + u_j^r)),
\end{aligned}$$

$$F_9 = \gamma_0(u_i^r + u_j^r, \triangle y_i) + \sum_{h=1}^{H} k\left(\frac{h-1}{H}\right)(\gamma_h(u_i^r + u_j^r, \triangle y_i) + \gamma_{-h}(u_i^r + u_j^r, \triangle y_i)).$$

F_2, F_4, F_6, F_8 and F_{10} are equal to F_1, F_3, F_5, F_7 and F_9, respectively, by replacing $\triangle y_i$ by $\triangle y_j$ and viceversa.

The proof of the concentration inequality for a single asset i shows that $|\overline{s}_{\mathsf{K}\,ij} - \sigma_{ij}^*|$ satisfies (1.26). Starightfoward computations show that the same bound is satisfied for all $|F_q|$. Finally, proceeding similarly, we obtain the same estimate for the MRK estimator of the second term in (1.28), which concludes the desired proof. □

Chapter 2

A TRUNCATED TWO–SCALES REALIZED VOLATILITY ESTIMATOR

2.1 Introduction

The volatility of asset prices is a fundamental ingredient for asset pricing, risk management and portfolio allocation. Over the last decade, the financial econometrics literature has developed a new generation of estimators of the daily volatility of asset prices based on intra–daily data typically referred to as realized volatility estimators. The classic realized volatility estimator of Andersen et al. (2003) for example is defined as the sum of the squares of high–frequency intra–daily returns. Under appropriate assumptions, this estimator provides a consistent estimate of the quadratic variation of asset prices when prices follow a continuous stochastic model and are directly observed (see e.g. Barndorff-Nielsen and Shephard (2002) and Andersen et al. (2003)).

It is well acknowledged in the literature that asset prices exhibit discontinuities in their sample paths and that are also contaminated by market microstructure noise (see e.g. Barndorff-Nielsen and Shephard (2006) and Hansen and Lunde (2012)). The presence of discontinuities has motivated modelling prices as a combination of a continuous and a jump process. However, allowing for a jump component makes it more challenging to estimate the quadratic variation of the continuous part, which is typically the object of interest from an economic perspective. The presence of market microstructure noise also poses challenges to the estimation of the quadratic variation. In fact, in the presence of noise standard realized volatility estimators are inconsistent as the sampling frequency of the data increases.

These two important stylized facts of asset prices have motivated the development of a number of estimators which are consistent in the presence of jumps, noise or both. An estimator that is robust to price jumps is the truncated realized volatility estimator introduced by Mancini (2008, 2009), which deals with both finite and infinite activity jumps. Moreover, different realized power and multipower variation estimators that are also robust to price jumps have been introduced, see e.g. Barndorff-Nielsen and Shephard (2004b) and Corsi et al. (2010) for the case of finite activity jumps, and Barndorff-Nielsen et al. (2006), Woerner (2006), Jacod (2008), and Jacod and Todorov (2014) for infinite activity jumps. Consistent estimators in the presence of noise are the two-scales realized volatility (Zhang et al. (2005)) and the realized kernels (Barndorff-Nielsen et al. (2008), Barndorff-Nielsen et al. (2011)). Finally, contributions that propose estimators that are consistent in the presence of both finite activity jumps and noise include, among others, Podolskij and Vetter (2009), Fan and Wang (2007), Barunik and Vacha (2015) and Christensen et al. (2010).

This chapter contributes to this latter strand of the literature by developing a novel realized volatility estimator that is consistent in the presence of both finite or infinity activity jumps and noise. We do so by combining a truncation technique in the spirit of Mancini (2008, 2009) to deal with the jumps, together with the idea of local average of intra-daily returns developed in Zhang et al. (2005) to deal with the market microstructure noise. The standard truncation technique introduced by Mancini (2008) consists of excluding the intra-daily returns larger than a threshold (in absolute value) from the estimation of the quadratic variation, as these are likely to contain a realization of a jump. However, this jump truncation strategy fails if the efficient price process is contaminated by market microstructure noise. We overcome this hurdle by introducing a truncation technique based on a local average of intra-daily returns. We show that such local average smooths away the effect of the noise and retains the property of being large when a jump is realized. We then estimate the quadratic variation using the two–scales realized volatility estimator after truncating the intervals in which the local average is larger than a threshold. We call our estimator the truncated two–scales realized volatility estimator (TTSRV). We show that this estimator is consistent in the presence of finite or infinite activity jumps. In the case where jumps have finite activity, we derive its asymptotic distribution. A simulation study is used to assess the finite sample properties of the TTSRV estimator. The study shows that when the price process is affected by both jumps and noise, our proposed estimator delivers a significant improvement over other commonly employed realized volatility estimators: the

truncated realized volatility (Mancini, 2008, 2009), the two-scales realized volatility (Zhang et al., 2005), the bipower variation (Barndorff-Nielsen and Shephard, 2004b), and the modulated bipower variation (Podolskij and Vetter, 2009).

This chapter is primarily related to the contributions of Podolskij and Vetter (2009), Fan and Wang (2007), Barunik and Vacha (2015) and Christensen et al. (2010). The estimator proposed in Podolskij and Vetter (2009) is a modified version of the modulated bipower variation (MBV) obtained using an estimator of the variance of the market microstructure noise. The resulting estimator is shown to be consistent in the presence of finite activity jumps and noise, but no asymptotic distribution or convergence rate is established. The simulation study of this chapter shows that our estimator is significantly more efficient than this new MBV. The estimators proposed in Fan and Wang (2007) and Barunik and Vacha (2015) are based on wavelet techniques. They first use wavelets to detect the locations and sizes of price jumps and then remove those jumps from the price series. The authors then apply noise robust realized volatility estimators to the jump adjusted data. The asymptotic distribution of the wavelet-based two-scales estimator is derived in Fan and Wang (2007). Our truncated two-scales estimator is similar in spirit to this approach since both estimators are based on detecting the jumps first and then applying a noise robust estimator to the jump adjusted data. In fact, the TTSRV has the same asymptotic distribution of the estimator proposed in Fan and Wang (2007). Our estimator however is easier to compute than the wavelet-based one, and we also show that it is robust to both infinite activity jumps and noise. Moreover, we are able to derive a technique to estimate the asymptotic variance of the estimation error, which is absent in Fan and Wang (2007). The estimator proposed in Christensen et al. (2010) is based on intra–daily quantile ranges. It achieves the optimal convergence rate of a realized volatility estimator in the presence of noise and finite activity jumps. However, its efficiency relies on specific assumptions on the dynamics of the spot volatility that are quite restrictive. In fact, the other estimators cited in this section as well as the estimator proposed in this chapter do not rely on such assumptions. It is important to emphasize that the analysis of all of these estimators has been developed under finite activity jumps only, while in this chapter we establish the properties of our estimator under both finite and infinite activity jumps. This chapter is also related to the contributions of Jacod and Protter (2012) and Aït-Sahalia and Jacod (2014) who propose truncated estimators on volatilty functionals, including the truncated bipower and multipower estimators. These estimators however do not consistently estimate the integrated volatility in the presence of noise. Besides volatil-

ity estimation, truncation techniques are widely used, for instance, to explore the relationship between jumps and spot volatility (Jacod and Todorov, 2010) and to estimate the covariation between asset returns and changes in volatility (Aït-Sahalia et al., 2016). Another strand of the literature relevant to this chapter is the one that concerns testing for the presence of price jumps and cojumps (with or without noise), as e.g. in Jacod and Todorov (2009), Jacod et al. (2010), Aït-Sahalia et al. (2012), Lee and Mykland (2012) and Li et al. (2016). These papers have inspired the jump detection indicator based on local averages used in this chapter.

The outline of this chapter is as follows. Section 2 introduces basic notation and the definition of the TTSRV estimator. In Section 3 we establish the asymptotic properties of the estimator when price jumps have finite activity. In Section 4 we derive the consistency of the estimator when jumps are of infinite activity. Section 5 contains the result of the simulation study.

2.2 Methodology

We denote by $(y_t, t \in [0,1])$ the efficient log–price process of an asset, where 0 typically represents the opening of the trading day and 1 the closing. The process starts at an initial value $y_0 \in \mathbb{R}$ and its dynamics are given by

$$dy_t = a_t dt + \sigma_t dB_t + dJ_t, \qquad t \in]0,1], \tag{2.1}$$

where B is a standard Brownian motion and J is a pure jump Lévy process, both defined on a filtered probability space $(\Omega, (\mathcal{F}_t)_{t\geq 0}, \mathcal{F}, \mathrm{P})$. We assume that the coefficients a and σ are progressively measurable processes, which ensures that $(y_t, t \in [0,1])$ is adapted and càdlàg (see e.g. Ikeda and Watanabe, 1981).

We assume that the efficient log–price is contaminated by market microstructure noise. That is, rather than the efficient price y_t the econometrician observes at discrete times its contaminated counterpart x_t. Specifically, we assume that the observed price x_t is measured at equally spaced timestamps $0 = t_0 < t_1 < \cdots < t_m = 1$ where $h = t_i - t_{i-1} = \frac{1}{m}$ and is generated as

$$x_{t_i} = y_{t_i} + u_{t_i}, \quad i = 1, ..., m,$$

where u_{t_i} denotes the microstructure noise associated to the i–th trade. We assume that u_{t_i} is a

discrete i.i.d. process, independent of the efficient price process and such that $u_{t_i} \sim N(0, \eta^2)$ where η is a positive constant. In order to simplify the exposition, we denote by x_i, y_i and u_i the processes x_{t_i}, y_{t_i} and u_{t_i}.

The theory developed in this chapter differs depending on whether the pure jump Lévy process J has finite activity (FA), that is, it jumps a.s. a finite number of times on each finite time interval, or it has infinite activity (IA). It is well-known (see e.g. Sato (1999)), that we can always decompose J as $J_t = J_{1,t} + J_{2,t}$, where

$$J_{1,t} = \int_0^t \int_{|z|>1} z\, \mu(dz, ds), \quad \text{and} \quad J_{2,t} = \int_0^t \int_{|z|\leq 1} z\, \left(\mu(dz, ds) - \nu(dz)ds\right), \tag{2.2}$$

where μ is the Poisson random measure associated to the jumps of J, $\mu(dz, ds) - \nu(dz)ds$ is the compensated measure and ν is the Lévy measure. Observe that J_2 is a square integrable martingale with IA, and for each t, $\text{var}(J_{2,t}) = t \int_{|x|\leq 1} x^2 \nu(dx) < \infty$. On the other hand, J_1 is a compound Poisson process with FA, so we can write $J_{1,t} = \sum_{i=1}^{N_t} Y_i$, where N is a Poisson process with constant intensity $\lambda > 0$, and Y_i are i.i.d. random variables independent of N. Recall that N_t counts the number of jumps occurred in the interval $[0, t]$, and the Y_i's are the different jump sizes. We set $\triangle N_t = N_t - N_{t^-}$ and $N_i = N_{t_i}$, for $i = 1, ..., m$. Observe that we can also consider a slightly more general jump process $J_t = J_{1,t} + J_{2,t}$, where J_2 is an IA Lévy pure jump process and J_1 is a general FA jump process, that is, $J_{1,t} = \sum_{i=1}^{N_t} Y_i$, where N is a non-explosive counting process with not necessarily constant intensity, and the random variables Y_i are not necessarily i.i.d., nor independent of N.

The aim of this chapter is to provide a consistent estimator of the integrated volatility

$$\text{IV} = \int_0^1 \sigma_t^2 dt.$$

In case of no microstructure noise nor price jumps, IV is the quadratic variation of the price process, and it is well-known that the realized volatility $\sum_{i=1}^m (y_i - y_{i-1})^2$ is a consistent estimator of IV (see e.g. Barndorff-Nielsen and Shephard (2002)). Several strategies have been put forward in the literature to estimate IV in case of jumps. The approach introduced by Mancini (2008, 2009) consists of excluding from the realized volatility the intervals $(t_{i-1}, t_i]$ where jumps are likely to have occurred. In order to identify such intervals, Mancini proposes a truncation method that consists of comparing the value of the squared return over each interval with a given threshold. If this value is larger than the threshold, then it is likely that the interval

contains a jump. However, in the presence of microstructure noise this method does not consistently detect jumps since large returns can be observed over short intervals because of the noise.

In order to obtain a consistent estimator of IV in the presence of jumps and noise we propose the following estimation strategy. We first introduce a jump signaling device that is able to detect the location of jumps in the presence of microstructure noise. We then use the jump signaling device to truncate a noise robust realized volatility estimator.

In order to detect the presence of a jump in a given time interval $(t_{i-1}, t_i]$, we consider the following measure

$$\beta_i = \frac{1}{K_1} \sum_{j=i}^{i+K_1-1} (x_j - x_{j-K_1}), \quad \text{for} \quad i = 1, \ldots, m,$$

where $K_1 = K_1(m)$ satisfies $\lim_{m\to\infty} \frac{K_1}{m} = 0$ and $\lim_{m\to\infty} K_1 = \infty$. The β_i measure is a local average of overlapping returns. In the interval $(t_{i-K_1}, t_{i+K_1-1}]$ we construct K_1 overlapping intervals of the form of $(t_j, t_{j+K_1}]$. The β_i measure is the average of the returns on these K_1 intervals. A jump realized on $(t_{i-1}, t_i]$ will affect all the returns and will make $|\beta_i|$ large. On the other hand, if no jump is realized on $(t_{i-1}, t_i]$ the local averaging smooths away the effect of the noise and $|\beta_i|$ will be small. In fact, Theorem 2.1 below quantifies these facts in terms of a threshold $r(h)$ as m is large. Notice the intervals $(t_{i-K_1}, t_{i+K_1-1}]$ are not likely to contain jumps for large m since its length $K_1 h$ is assumed to be small for large m.

A classic estimator of IV that is consistent in the presence of microstructure noise is the two-scale realized volatility (TSRV) estimator, which is defined as (see Zhang et al. (2005))

$$\widehat{\sigma}^2 = \frac{1}{K} \sum_{j=K}^{m} (x_j - x_{j-K})^2 - \frac{m-K+1}{mK} \sum_{j=1}^{m} (x_j - x_{j-1})^2,$$

where $K = cm^{2/3}$ and c is a positive constant. When there is no microstructure noise nor jumps, the first term converges to IV in probability. In the presence of microstructure noise, the second term corrects the bias and $\widehat{\sigma}^2$ becomes a consistent estimator of IV. However, this estimator is not consistent in the presence of jumps. Our strategy consists of truncating the TSRV estimator using the β_i measure to obtain a consistent estimator in the presence of microstructure noise as well as jumps. However, the TSRV estimator is difficult to truncate due to the different scale of

both sums. Therefore, we work with the following modified version of the TSRV estimator

$$\widehat{\sigma}^2_{\mathsf{TS}} = \frac{1}{K}\sum_{j=K}^{m}(x_j - x_{j-K})^2 - \frac{1}{K}\sum_{j=K}^{m}(x_j - x_{j-1})^2.$$

We observe that $\widehat{\sigma}^2_{\mathsf{TS}}$ is also a consistent estimator of IV in the presence of microstructure noise since the second term is of the same order as the second term in the TSRV. Moreover, it is easy to show that both estimators $\widehat{\sigma}^2_{\mathsf{TS}}$ and $\widehat{\sigma}^2$ have the same asymptotic distribution derived in Zhang et al. (2005) for $\widehat{\sigma}^2$.

We finally introduce the truncated two-scales realized volatility estimator (TTSRV)

$$\widehat{\sigma}^2_{\mathsf{TTS}} = \frac{1}{K}\sum_{j=K}^{m}(x_j - x_{j-K})^2\mathbf{1}_{E_j} - \frac{1}{K}\sum_{j=K}^{m}(x_j - x_{j-1})^2\mathbf{1}_{E_j}, \tag{2.3}$$

where $E_j = \{|\beta_i| \leq r(h), \text{ for all } i = j-K+1, \ldots, j\}$, where $r(h)$ is the threshold introduced in Theorem 2.1. In Section 3 we study the asymptotic properties of this estimator when $J_2 \equiv 0$ so that J has FA, while in Section 4 we allow J_2 to be non-zero.

2.3 Theory: Finite Activity Jumps

2.3.1 Jump Detection

In this section, we consider the case where J has FA. That is, $J = J_1$ and J_1 is a compound Poisson process with FA or a more general FA jump process as above. In this case, the quadratic variation of y on $[0,1]$ is given by

$$[y]_1 = \text{IV} + \sum_{i=1}^{N_1} Y_i^2.$$

The following theorem shows that the absolute values of β_i can be used to identify the intervals where no jumps occured.

THEOREM 2.1. *Suppose that*

(1) *For all* $t \in [0,1]$, $\mathrm{P}\left(\triangle N_t \neq 0, Y_{N_t} = 0\right) = 0$.

(2) $\limsup_{h\to 0} \frac{\sup_{i\in\{1,\ldots,m+K_1-1\}}\left|\int_{t_{i-K_1}}^{t_i} a_s\, ds\right|}{\sqrt{K_1 h \log\frac{1}{h}}} \leq C(\omega) < \infty$ *a.s.*

(3) $\limsup_{h\to 0} \frac{\sup_{i\in\{1,\ldots,m+K_1-1\}} \left|\int_{t_{i-K_1}}^{t_i} \sigma_s^2 ds\right|}{K_1 h} \leq M(\omega) < \infty$ *a.s.*

(4) $r(h)$ *is a deterministic function such that*

$$\lim_{h\to 0} r(h) = 0, \ \lim_{h\to 0} \frac{\sqrt{\frac{\log \frac{1}{h}}{K_1}}}{r(h)} = 0, \text{ and } \lim_{h\to 0} \frac{\sqrt{K_1 h \log \frac{1}{h}}}{r(h)} = 0.$$

Then, for P*-almost all* ω*, there exists* $\overline{h}(\omega) > 0$ *such that for all* $h \leq \overline{h}(\omega)$ *we have for all* $i = 1, \ldots, m$*,*

$$\mathbf{1}_{\left\{N_{i+K_1-1}-N_{i-K_1}=0\right\}}(\omega) \leq \mathbf{1}_{\{|\beta_i|\leq r(h)\}}(\omega) \text{ and } \mathbf{1}_{\{|\beta_i|\leq r(h)\}}(\omega) \leq \mathbf{1}_{\{N_i-N_{i-1}=0\}}(\omega).$$

REMARK 2.1. (i) *Assumption* (1) *says that the sizes of the price jumps are a.s. non-zero. Note that any FA Lévy process satisfies this condition, since* $\nu(\{0\}) = 0$*.*

(ii) *Assumptions* (2) *and* (3) *are satisfied by common assumptions on* a *and* σ *such as being a.s. bounded on* $[0, 1]$*. In particular, they are satisfied as soon as* a *and* σ *have càdlàg paths.*

(iii) *Assumption* (4) *indicates how to choose the threshold. If e.g.* $K_1 = m^{\alpha_1}$ *with* $0 < \alpha_1 < 1$*, then we can choose* $r(h) = h^{\alpha_2}$ *with* $0 < \alpha_2 < \min\left(\frac{\alpha_1}{2}, \frac{1-\alpha_1}{2}\right)$*.*

(iv) *As in Mancini (2009) and Zhang et al. (2005), the results of this chapter can be extended to not necessarily equally spaced observations, but for the sake of conciseness, we leave it to the interested reader.*

Theorem 2.1 quantifies the intuition of the β_i measure given after its definition in terms of a threshold. The first inequality implies that when there are no jumps on the interval $(t_{i-K_1}, t_{i+K_1-1}]$ (which recall it is likely to occur when m is large), $|\beta_i|$ will be smaller than the threshold for large m. This is because condition (4) implies that the absolute value of the local average of the returns over that interval when there are no jumps goes faster to zero than the threshold. On the other hand, the second inequality implies that if there is a jump on the interval $(t_{i-1}, t_i]$, $|\beta_i|$ will be larger than the threshold for large m. Therefore, in this case, the intervals $(t_{j-K}, t_j]$ that contain $(t_{i-1}, t_i]$ will be removed from the estimator to eliminate the impact of the jump.

2.3.2 Consistency and Asymptotic Mixed Normality

When there are no price jumps, the next result shows that $\widehat{\sigma}^2$ and $\widehat{\sigma}^2_{\mathsf{TS}}$ have the same asymptotic distribution (derived in Zhang et al. (2005, Theorem 4) for $\widehat{\sigma}^2$).

PROPOSITION 2.1. *Consider the framework of Section 2 with $J = 0$, and assume the drift coefficient a and the diffusion coefficient σ are a.s. continuous on $[0, 1]$, and σ is a.s. bounded away from 0. Then, as $m \to \infty$,*

$$m^{1/6}\left(\widehat{\sigma}^2_{\mathsf{TS}} - \int_0^1 \sigma_t^2 dt\right) \xrightarrow{\mathcal{L}} \left(8c^{-2}\eta^4 + \frac{4}{3}c\int_0^1 \sigma_t^4 dt\right)^{1/2} N(0,1),$$

where the convergence is stable in law, and c is the constant such that $K = cm^{2/3}$.

As in Zhang et al. (2005), stable convergence means that the left-side converges to the right-side jointly with the x process, and the $N(0,1)$ random variable is independent of x.

We now turn to the analysis of the TTSRV estimator defined in (2.3). We first show that it is a consistent estimator of IV.

THEOREM 2.2. *Consider the assumptions of Theorem 2.1, and those of Proposition 2.1 on a and σ. Assume also that $\lim_{m\to\infty} \frac{K_1 \log m}{m^{1/3}} = 0$. Then as $m \to \infty$,*

$$\widehat{\sigma}^2_{\mathsf{TTS}} \xrightarrow{\mathrm{P}} \int_0^1 \sigma_t^2 dt.$$

The proof of this Theorem is divided into two steps. First, since by Theorem 2.1 price jumps can be detected as m is large and are removed from the computation of the TTSRV estimator, we first show that the difference between the TTSRV computed when there is a jump component in the price and when there is not is a.s. zero when m is large (see the proof of (2.11) below). Second, we show that when there are no price jumps, the difference between the TTSRV and $\widehat{\sigma}^2_{\mathsf{TS}}$ is $O_{\mathrm{P}}(K_1 m^{-1/3}\log m)$ (see the proof of (2.12) below). Thus, the assumption that $\lim_{m\to\infty} \frac{K_1 \log m}{m^{1/3}} = 0$ gives the desired result. We can e.g. choose $K_1 = m^{\alpha_1}$ and $r(h) = h^{\alpha_2}$ with $0 < \alpha_1 < \frac{1}{3}$ and $0 < \alpha_2 < \frac{\alpha_1}{2}$.

Using Proposition 2.1 and the same steps of the proof of Theorem 2.2 we obtain that asymptotic distribution of the TTSRV, which is the same as in Proposition 2.1.

THEOREM 2.3. *Consider the assumptions of Theorem 2.1, and those of Proposition 2.1 on a*

and σ. Assume also that $\lim_{m\to\infty} \frac{K_1 \log m}{m^{1/6}} = 0$*. Then as* $m \to \infty$*,*

$$m^{1/6}\left(\widehat{\sigma}^2_{\mathsf{TTS}} - \int_0^1 \sigma_t^2 dt\right) \xrightarrow{\mathcal{L}} \left(8c^{-2}\eta^4 + \frac{4}{3}c\int_0^1 \sigma_t^4 dt\right)^{1/2} N(0,1),$$

where the convergence is stable in law.

We can e.g. choose $K_1 = m^{\alpha_1}$ and $r(h) = h^{\alpha_2}$ with $0 < \alpha_1 < \frac{1}{6}$ and $0 < \alpha_2 < \frac{\alpha_1}{2}$.

2.3.3 Estimating the Asymptotic Variance

In this section we follow similar steps as in Zhang et al. (2005) in order to estimate the asymptotic variance of the TTSRV estimator, which by Theorem 2.3 equals $8c^{-2}\eta^4 + \frac{4}{3}c\int_0^1 \sigma_s^4 ds$. More specifically, we apply our truncation technique to the estimator proposed in Zhang et al. (2005).

First, we begin by noting that η^2 can be estimated by $\widehat{\eta}^2 = \frac{\sum_{i=1}^m (x_i - x_{i-1})^2}{2m}$. In our setting it can be easily checked that $\widehat{\eta}^2$ is still a consistent estimator of η^2. The reason is that since we have finite price jumps on $[0,1]$, they will only affect finite terms in $\sum_{i=1}^m (x_i - x_{i-1})^2$, so the jump effect will vanish as $m \to \infty$.

Next, following the notation in Section 6 of Zhang et al. (2005), we divide $[0,1]$ into segments $(T_n, T_{n+1}]$, where $T_n = \frac{nM}{m}$, for $n = 1, \ldots, m/M$. The value of M will be specified later and for simplicity we assume m/M is an integer. We define the TTSRV estimator for the period $[0, T_n]$ as

$$< \widehat{X,X} >^K_{T_n} = \frac{1}{K}\sum_{j=K}^{Mn} (x_j - x_{j-K})^2 \mathbf{1}_{E_j} - \frac{1}{K}\sum_{j=K}^{Mn} (x_j - x_{j-1})^2 \mathbf{1}_{E_j}.$$

Consider the following truncated version of an estimator introduced in Zhang et al. (2005)

$$\widehat{s}_0^2 = m^{1/3}\sum_{n=1}^{m/M}\left(< \widehat{X,X} >^{K_2}_{T_n} - < \widehat{X,X} >^{K_2}_{T_{n-1}} - \left(< \widehat{X,X} >^{K_3}_{T_n} - < \widehat{X,X} >^{K_3}_{T_{n-1}}\right)\right)^2,$$

where $K_2 = c_2 m^{2/3}$, $K_3 = c_3 m^{2/3}$, and c_2, c_3 are positive constants. Zhang et al. (2005) show that the untruncated version of $\widehat{s}_0^2$ converges in probability when there are no price jumps as $m \to \infty$ to

$$\mathrm{V} = \frac{4}{3}\left(c_2^{1/2} - c_3^{1/2}\right)^2 \int_0^1 \sigma_s^4 ds + 8\eta^4(c_2^{-2} + c_3^{-2} - c_2^{-1}c_3^{-1}).$$

In the presence of jumps, we obtain the analogous result in our setting

THEOREM 2.4. *Assume the hypotheses of Theorem 2.3, and that*

$$\lim_{m\to\infty}\frac{K_1(\log m)^2}{m^{1/6}}=0,\quad \lim_{m\to\infty}\frac{M}{m^{2/3}}=\infty,\quad \textit{and}\quad \lim_{m\to\infty}\frac{M}{m^{5/6}\log m}=0.$$

Then as $m\to\infty$, $\widehat{s}_0^2\xrightarrow{\mathrm{P}}\mathrm{V}$.

Finally, in order to estimate the asymptotic variance of the TTSRV estimator it suffices to combine the estimators $\widehat{\eta}^2$ and $\widehat{s}_0^2$ using the formula provided in Zhang et al. (2005). Notice that we can choose $K_1=m^{\alpha_1}$ and $M=m^{\alpha_2}$ with $0<\alpha_1<\frac{1}{6}$ and $\frac{2}{3}<\alpha_2<\frac{5}{6}$.

2.4 Theory: Infinite Activity Jumps

In this section, we allow J_2 to be a non-zero IA Lévy pure jump process, and J_1 is assumed to be a general FA jump process with counting process N. As $J=J_1+J_2$, the quadratic variation of the process y up time 1 becomes

$$[y]_1=\mathrm{IV}+\sum_{t\leq 1}\left(\Delta J_{1,t}\right)^2+\sum_{t\leq 1}\left(\Delta J_{2,t}\right)^2. \tag{2.4}$$

The following theorem shows that in this IA jumps setting, the TTSRV estimator still consistently estimates IV.

THEOREM 2.5. *Assume the hypotheses of Theorem 2.3, and that*

$$\lim_{m\to\infty}\frac{K_1^3\log m}{r^2(h)m^{1/3}}=0. \tag{2.5}$$

Assume also that $\mathrm{P}\left(N_i-N_{i-1}>0\right)=O\left(\frac{1}{m}\right)$ *and that* J_2 *is independent of* N. *Then as* $m\to\infty$, $\widehat{\sigma}^2_{\mathsf{TTS}}\xrightarrow{\mathrm{P}}\int_0^1\sigma_t^2dt$.

For example, we can set $K_1=m^{\alpha_1}$ and $r(h)=h^{\alpha_2}$, with $0<3\alpha_1+2\alpha_2<\frac{1}{3}$ and $0<\alpha_2<\frac{\alpha_1}{2}$. The structure of the proof of this theorem is similar to that of Mancini (2009, Theorem 4). On one hand, we use the measure β_i and the threshold $r(h)$ in order to cut off the jumps from J_1. On the other hand, we truncate the jumps in J_2 with absolute value larger than $\sqrt{\delta+16r^2(h)}$, where $\delta>0$ is arbitrary, and show that the information loss caused by the truncation is negligible.

It is more challenging to establish a central limit theorem result in the infinite activity case, and we leave this problem open for future research. We point out that in Cont and Mancini (2011) a central limit theorem for the truncated realized volatility estimator in the case of infinite activity jumps and no noise is established. However, their result relies on an argument which we are not able to use when prices are contaminated by microstructure noise. In particular, their asymptotic result relies on choosing a threshold low enough such that the error caused by the infinite jumps is negligible. In our framework, because of the noise component, the threshold has to be sufficiently large.

2.5 Simulation Study

In this section we perform a simulation study to assess the performance of the TTSRV estimator. The simulation exercise consists of simulating one day of high frequency data and then applying the TTSRV estimator to estimate the integrated volatility. We consider two different specifications of the prices process, the first one has finite active jumps while in the second the jump activity is infinite. The TTSRV estimator is also benchmarked against a set of alternative estimators proposed in the literature.

We simulate the observed price x_t according to

$$x_t = \int_0^t \sigma_s dB_s + J_t + u_t.$$

The spot volatility σ_s follows a CIR process

$$d\sigma_s = \kappa(\upsilon - \sigma_s) + \tau\sqrt{\sigma_s}dW_s,$$

where $\upsilon = 9$ is the long run mean of the process, $\tau = 2.74$ is its volatility, $\kappa = 0.1$ is the mean reversion parameter, and W is a standard Brownian motion independent of B. The noise u_t is a discrete i.i.d. $N(0, \eta^2)$, where $\eta > 0$. Two different specifications for the jump process J are used. In first case, labelled as model 1, J is a compound Poisson process with a constant intensity $\lambda = 2$. The jump sizes are i.i.d. $N(0, \xi^2)$, where $\xi > 0$. In the second case, labelled as model 2, J is a variance gamma (VG) process, which is a pure jump process with infinite

activity and finite variation. The process is defined as

$$J_s = d_1 G_s + d_2 \overline{W}_{G_s},$$

where G_s is a Gamma random variable with shape parameter s/b and scale parameter $b > 0$, and $\overline{W}$ is a standard Brownian motion independent of B, W and G. We fix the values of d_1 and d_2 to respectively -0.8 and 0.8. We assume that a trading day is eight hours long and the observed price x_t is measured each second (that is, $m = 28,800$). The simulation is carried out using the Euler simulation scheme. Throughout this section we compute the TTSRV for K_1 set to 4 (which is approximately $m^{1/7}$) and K set to 30. For each model and parameter setting the simulation is replicated 1000 times.

Figures 2.2 and 2.3 show the plot of the MSE of the TTSRV estimator as a function of $r(h)$ in, respectively, model 1 and model 2 for different magnitudes of η (0.05, 0.10, 0.15). For model 1, ξ is fixed to 2 while for model 2, b is fixed to 2. The figures show that in both cases the MSE is a decreasing function of $r(h)$ when $r(h)$ is small. This is because when the threshold is too small, many intervals that do not contain jumps are truncated, and this increases the variance of the TTSRV estimator. On the other hand, the MSE is an increasing function of $r(h)$ when $r(h)$ is large. This is because when the threshold is too large, the intervals that contain price jumps are not truncated, and this increases the bias of the the TTSRV estimator.

Figures 2.4 and 2.5 show the plot of the MSE of the TTSRV estimator as a function of $r(h)$ in, respectively, model 1 and model 2 for different magnitudes of the jump component. For model 1, ξ is set to 1, 2 or 3 while for model 2, b is set to 1, 2 and 3. In both sets of simulations η is fixed to 0.1. The MSE has the same convex shape documented in Figures 2.2 and 2.3.

Next we investigate the finite sample distribution of the TTSRV estimator. We do this under model 1 only and for different values of ξ. We define the standardized estimation error as

$$z = \frac{m^{1/6}\left(\widehat{\sigma}^2_{\mathsf{TTS}} - \int_0^1 \sigma_s^2 ds\right)}{\left(8c^{-2}\eta^4 + \frac{4}{3}c\int_0^1 \sigma_s^4 ds\right)^{1/2}}.$$

Theorem 2.3 implies that if the sample size is sufficiently large, z should be approximately normally distributed. We compute z keeping the value of the threshold $r(h)$ fixed at 0.75 and $\eta = 0.1$. As mentionned in Mancini (2009, Remark in p.278) the optimality of the threshold varies for each model and one usually chooses the one that performs better the simulations.

Figure 2.6 shows the histogram and the normal qqplot of z while Table 2.1 reports summary statistics. The TTSRV has a bias that increases with the standard deviation of the jumps' size. We note that, comparing with the results of the simulation study of Mancini (2009), the bias is roughly of the same order of the one of the TRV estimator when no noise is present (see also Mancini, 2008). Overall, the approximation provided by the asymptotic theory is adequate.

Last, we compare the efficiency of the TTSRV with other estimators of the integrated volatility: The truncated realized volatility (TRV), the bipower variation (BPV), the modulated bipower variation (MBV) and the TSRV. In this exercise the threshold $r(h)$ of the TTSRV estimator is fixed to 0.75.

The TRV proposed by Mancini (2008, 2009) is a truncatated version of the classic realized volatility estimator. It is defined as

$$\widehat{\sigma}^2_{\mathsf{TRV}} = \sum_{i=2}^{m} (x_i - x_{i-1})^2 \mathbf{1}_{\{|x_i - x_{i-1}| \leq \overline{r}(h)\}}.$$

As we have previously pointed out this estimator is not consistent in the presence of noise. Moreover, the truncation devices used by this estimator does not truncate jumps with high probability in this setting. To minimize the impact of the noise, we use the optimal sampling scheme proposed by Zhang et al. (2005), which shows that the optimal amount of equidistant observations for constructing $\widehat{\sigma}^2_{\mathsf{RV}}$ is given by

$$\overline{m} = \left(\frac{1}{4\eta^4} \int_0^1 \sigma_s^4 ds \right)^{\mathbf{1}/3}.$$

Note however that this is only optimal in case jumps are absent. Also, as mentionned before it is not clear from the literature how to set $\overline{r}(h)$. Here, we set it to 0.7 as this approximately minimizes the MSE of the estimator in the scenarios considered in this study.

Another important jump robust estimator proposed in the literature is the bipower variation (BPV), introduced by Barndorff-Nielsen and Shephard (2004b). It is defined as

$$\widehat{\sigma}^2_{\mathsf{BPV}} = \frac{\pi}{2} \sum_{j=3}^{m} \left(x_j - x_{j-1} \right) \left(x_{j-1} - x_{j-2} \right).$$

The performance of $\widehat{\sigma}^2_{\mathsf{BPV}}$ is also compromised in the presence of noise. Again, in this case reducing the sampling frequency can relieve the problem, and we use the same low frequency

sampling scheme used for computing $\widehat{\sigma}^2_{\mathsf{TRV}}$.

We also consider the MBV proposed by Podolskij and Vetter (2009). The estimator is defined as

$$\widehat{\sigma}^2_{\mathsf{MBV}} = \frac{\frac{e_1 e_2}{\mu_1^2}\widetilde{\sigma}^2 - v_2\widehat{\eta}^2}{v_1},$$

where

$$\widetilde{\sigma}^2 = \sum_{n=1}^{M} \left|\overline{X}_n^{(\overline{K})}\overline{X}_{n+1}^{(\overline{K})}\right|, \qquad \widehat{\eta}^2 = \frac{1}{2m}\sum_{i=1}^{m}(x_i - x_{i-1})^2,$$
$$\overline{X}_n^{(\overline{K})} = \frac{1}{m/M - \overline{K} + 1}\sum_{i=\frac{(n-1)m}{M}}^{\frac{mn}{M}-\overline{K}}(x_{i+\overline{K}} - x_i),$$

$e_1 > 0$, $e_2 > 1$, $\mu_1 = \sqrt{\frac{2}{\pi}}$, $M = \frac{n}{e_2\overline{K}} = \frac{n^{1/2}}{e_1 e_2}$, $\overline{K} = e_1 m^{1/2}$, $v_1 = \frac{e_1(3e_2 - 4 + \max((2-e_2)^3, 0))}{3(e_2-1)^2}$ and $v_2 = \frac{2\min(e_2-1,1)}{e_1(e_2-1)^2}$ (which is an estimator of η^2). As pointed out by Podolskij and Vetter (2009), the computation of the optimal values of e_1 and e_2 involves solving polynomial equations of degree higher than two. Following Podolskij and Vetter (2009), we set $e_1 = 0.8$ and $e_2 = 2.3$.

For reference purposes, we also consider the TSRV estimator. Note that when jumps are present this estimator converges in probability to (see Zhang et al., 2005)

$$\int_0^1 \sigma_t^2 dt + \sum_{0\leq t\leq 1}(\Delta J_t)^2.$$

Figures 2.7 and 2.8 show the plot of the MSE of the estimators in models 1 and 2 as function of η. In model 1, ξ is 0.7 and in model 2, b is also 0.7. Tables 2.2 and 2.3 report the MSE of the estimators for selected values of η. The pictures show that the TTSRV estimator dominates the competing estimators almost uniformly over the range of η considered. It only performs worse than the BPV and slightly worse than the TRV estimators when the magnitude of the noise is negligible. The MSE of the TRV and BPV estimators increases steadily as η increases as these estimators are not consistent in the presence of noise. The MSE of the TSRV estimator does not change as the variance of the noise increases, however it is much larger as it is not consistent in the presence of jumps. In this setting, the MBV estimator does not perform particularly well (especially in model 1) despite being robust to both noise and jumps.

Figures 2.9 and 2.10 show the plot of the MSE of the estimators under model 1 as function of ξ and in model 2 as a function of b. In both models η is fixed to 0.05. Tables 2.4 and 2.5

report the MSE of the estimators for selected values of, respectively, ξ and b. Again, the TTSRV achieves the best performance overall. The TSRV performs well in this setting only when the standard deviation of the jump size or the scale of the VG process are so small that the impact of the jump component in negligible. The TRV, MBV and BPV estimators roughly have the same performance and are not affected by the variability of the jump component.

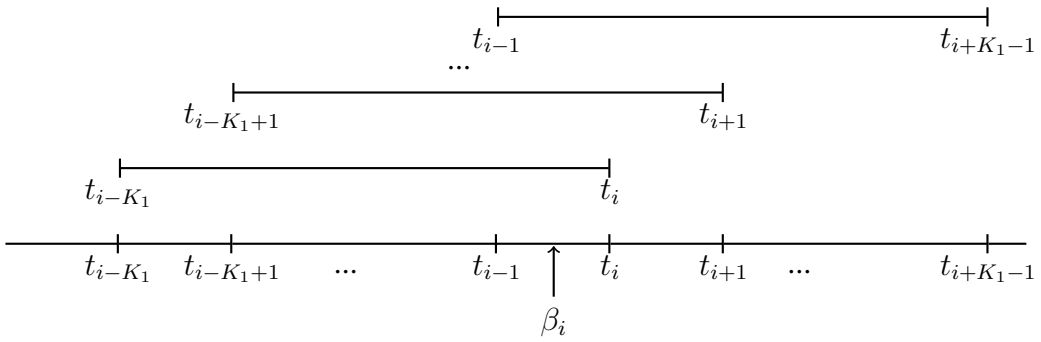
2.6 Conclusions

In this chapter we introduce a novel estimator of the integrated volatility of asset prices that is consistent in the presence of both finite or infinite activity price jumps and market microstructure noise. We first introduce a jump detection indicator which consistently detects jumps in the presence of noise. We then use our jump detection methodology to construct a truncated version of the two–scales realized volatility estimator that we call truncated two–scales realized volatility (TTSRV) estimator. In the finite activity jumps case, we show that this estimator is consistent as well as asymptotically normal. Because the intervals that contain price jumps will be a.s. truncated from the estimator and the information loss is negligible, the estimator has the same asymptotic properties as the standard two–scales realized volatility estimator in the case of no jumps. Moreover, we introduce an estimator of the asymptotic variance. In the infinite activity jumps case, we show that the TTSRV estimator is consistent. A simulation study is used to compare our proposed approach to other important realized volatility estimators proposed in the literature. The study shows that the TTSRV estimator performs favorably relative to its competitors when the price process contains jumps and is contaminated by noise.

2.7 Appendix

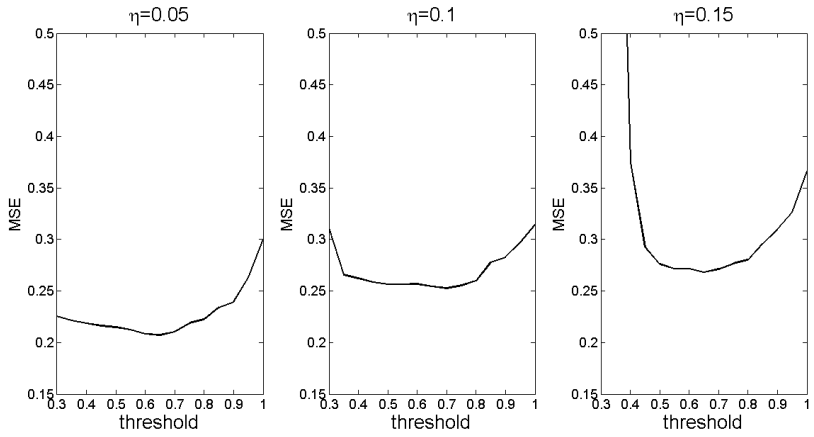
2.7.1 Figures and Tables

Figure 2.1: COMPUTING THE LOCAL AVERAGE RETURN



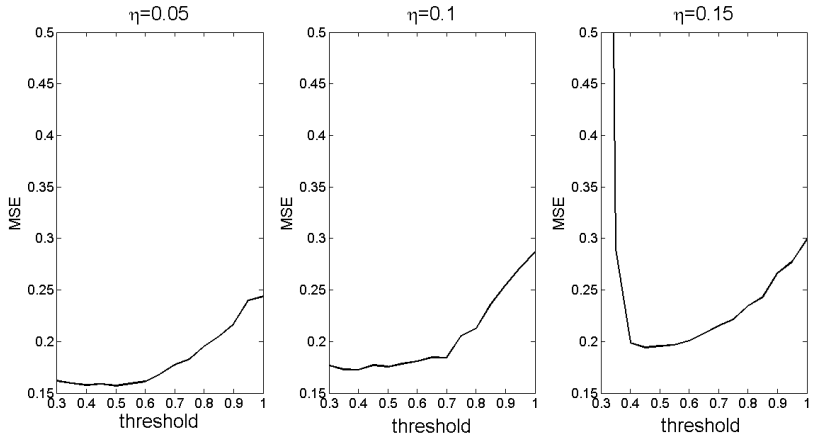
The figure provides a schematic representation of the computation of the β_i measure.

Figure 2.2: MSE CURVES OF THE TTSRV ESTIMATOR (I)



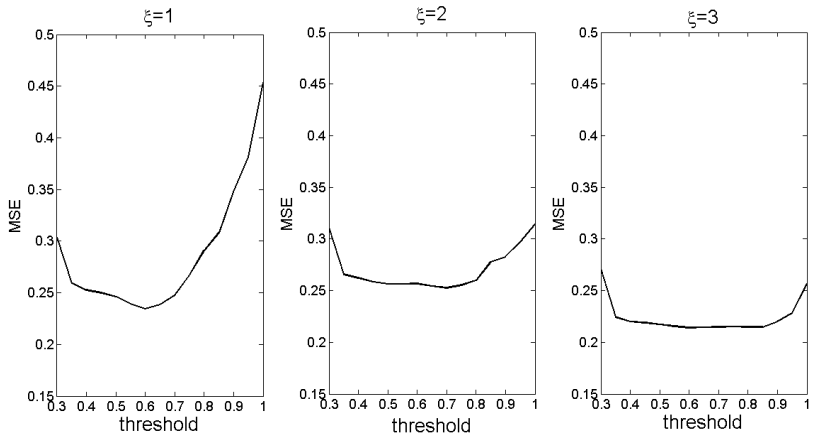
The figure shows the plot of the MSE of the TTSRV estimator in model 1 as a function of the threshold $r(h)$ for different values of the market microstructure noise standard deviation η ($\eta = 0.05, 0.10, 0.15$).

Figure 2.3: MSE CURVES OF THE TTSRV ESTIMATOR (II)



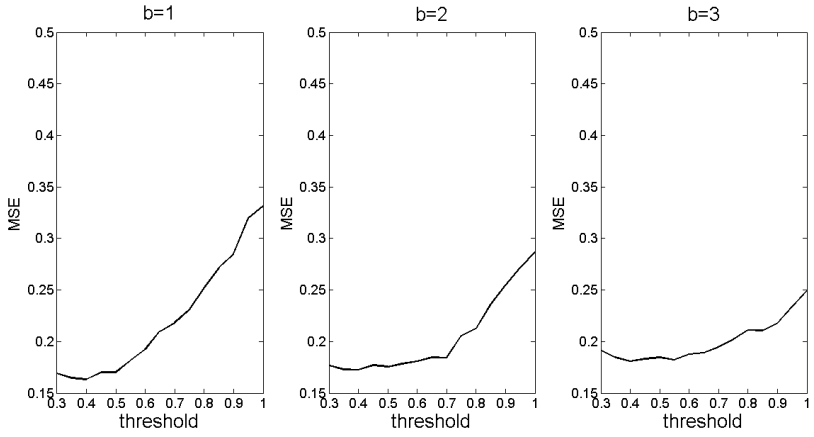
The figure shows the plot of the MSE of the TTSRV estimator in model 2 as a function of the threshold $r(h)$ for different values of the market microstructure noise standard deviation η ($\eta = 0.05, 0.10, 0.15$).

Figure 2.4: MSE CURVES OF THE TTSRV ESTIMATOR (III)



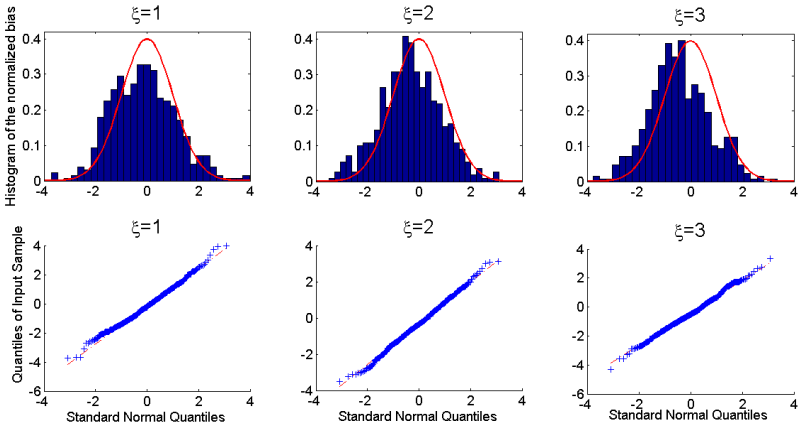
The figure shows the plot of the MSE of the TTSRV estimator in model 1 as a function of the threshold $r(h)$ for different values of the jump size standard deviation ξ ($\xi = 1, 2, 3$).

Figure 2.5: MSE CURVES OF THE TTSRV ESTIMATOR (IV)



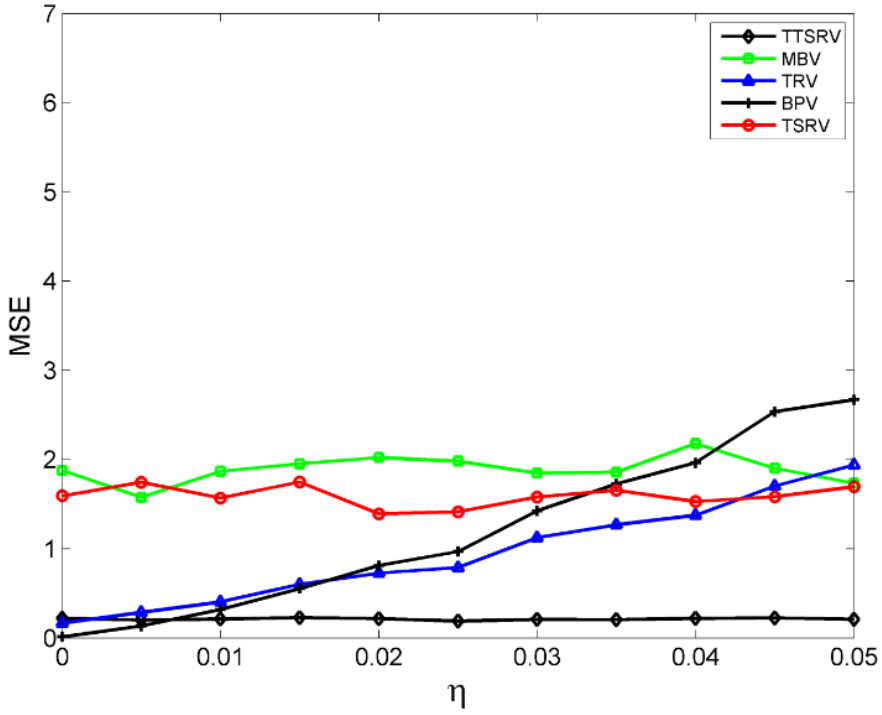
The figure shows the plot of the MSE of the TTSRV estimator in model 2 as a function of the threshold $r(h)$ for different values of VG process scale parameter b ($b = 1, 2, 3$).

Figure 2.6: ASYMPTOTIC NORMALITY



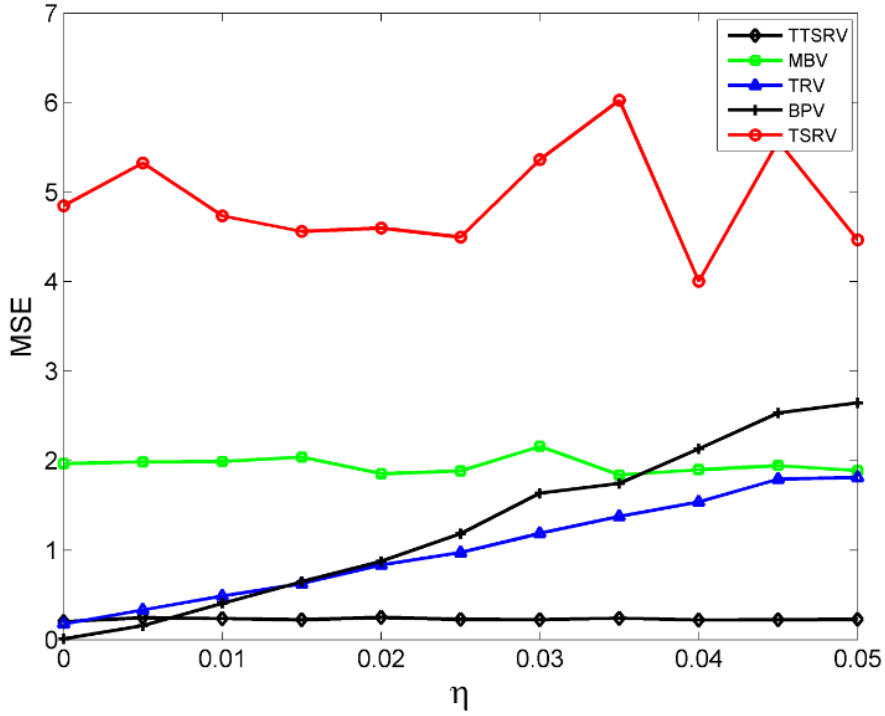
The figure shows the histograms and qqplots of the standardized estimation errors of the TTSRV estimator in model 1 for different values of the jump size standard deviation ξ ($\xi = 1, 2, 3$).

Figure 2.7: MSE CURVES OF THE DIFFERENT ESTIMATORS (I)



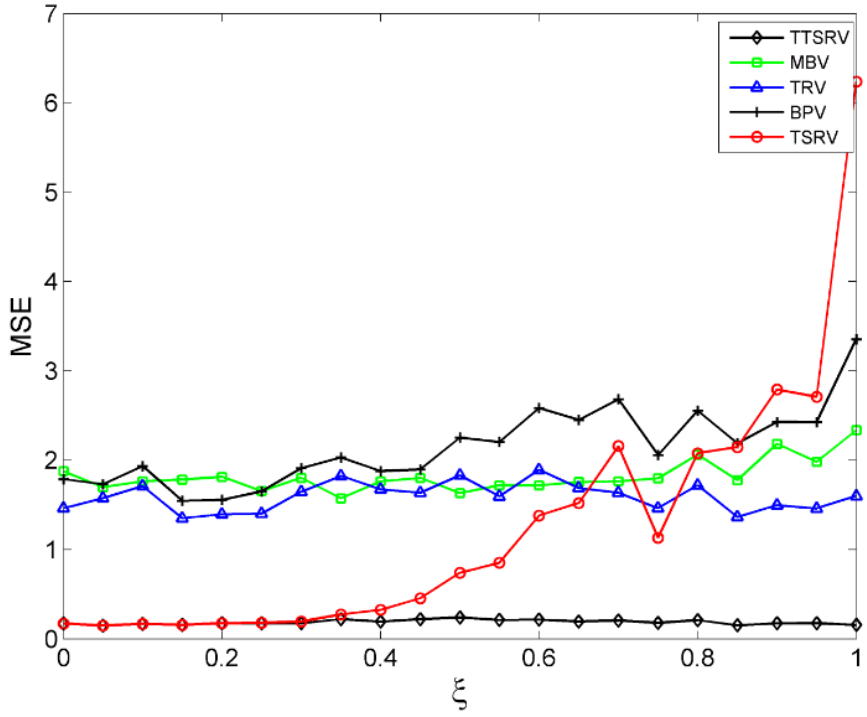
The figure shows the plot of the MSE of the TTSRV, TSRV, TRV, MBV and BPV estimators in model 1 as a function of the market microstructure noise standard deviation η.

Figure 2.8: MSE CURVES OF THE DIFFERENT ESTIMATORS (II)



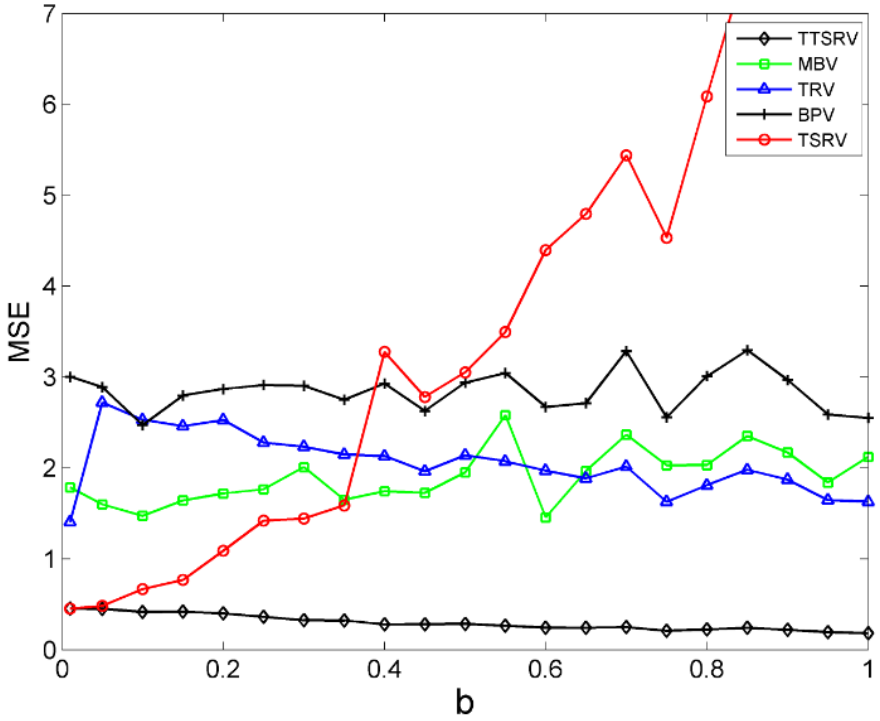
The figure shows the plot of MSE of the TTSRV, TSRV, TRV, MBV and BPV estimators in model 2 as a function of the market microstructure noise standard deviation η.

Figure 2.9: MSE CURVES OF THE DIFFERENT ESTIMATORS (III)



The figure shows the plot of MSE of the TTSRV, TSRV, TRV, MBV and BPV estimators in model 1 as a function of the jump size standard deviation ξ.

Figure 2.10: MSE CURVES OF THE DIFFERENT ESTIMATORS (IV)



The figure shows the plot of MSE of the TTSRV, TSRV, TRV, MBV and BPV estimators in model 1 as a function of VG process scale parameter b.

Table 2.1: STATISTICS OF THE STANDARDIZED ESTIMATION ERRORS

	$\xi = 1$	$\xi = 2$	$\xi = 3$
Mean	-0.148	-0.324	-0.456
Std Dev	1.247	1.177	1.170
Skewness	0.243	0.036	0.114
Kurtosis	3.126	2.988	3.079
JB test	0.065	0.113	0.546

The table reports the mean, standard deviation, skewness, kurtosis and p–value of the Jarque–Bera normality test of the standardized estimation errors of the TTSRV estimator in model 1 for different values of the jump size standard deviation ξ.

Table 2.2: COMPARISON OF THE MSE (I)

	$\eta = 0$	$\eta = 0.01$	$\eta = 0.02$	$\eta = 0.03$	$\eta = 0.04$	$\eta = 0.05$
TTSRV	0.220	0.212	0.218	0.207	0.219	0.211
TSRV	1.594	1.570	1.392	1.580	1.531	1.695
TRV	0.167	0.404	0.727	1.125	1.375	1.940
MBV	1.878	1.868	2.023	1.848	2.180	1.735
BPV	0.012	0.319	0.812	1.425	1.964	2.670

The table reports the MSE of the TTSRV, TSRV, TRV, MBV and BPV estimators in model 1 for different values of the market microstructure noise standard deviation η.

Table 2.3: COMPARISON OF THE MSE (II)

	$\eta = 0$	$\eta = 0.01$	$\eta = 0.02$	$\eta = 0.03$	$\eta = 0.04$	$\eta = 0.05$
TTSRV	0.203	0.237	0.250	0.226	0.221	0.227
TSRV	4.850	4.733	4.598	5.362	4.003	4.469
TRV	0.177	0.489	0.836	1.187	1.539	1.814
MBV	1.966	1.991	1.854	2.159	1.899	1.889
BPV	0.010	0.408	0.877	1.638	2.133	2.646

The table reports the MSE of the TTSRV, TSRV, TRV, MBV and BPV estimators in model 2 for different values of the market microstructure noise standard deviation η.

Table 2.4: COMPARISON OF THE MSE (III)

	$\xi = 0$	$\xi = 0.2$	$\xi = 0.4$	$\xi = 0.6$	$\xi = 0.8$	$\xi = 1$
TTSRV	0.173	0.175	0.192	0.215	0.209	0.155
TSRV	0.167	0.179	0.261	0.943	1.493	4.780
TRV	1.462	1.393	1.673	1.894	1.719	1.600
MBV	1.876	1.815	1.763	1.718	2.063	2.335
BPV	1.788	1.556	1.877	2.583	2.554	3.355

The table reports the MSE of the TTSRV, TSRV, TRV, MBV and BPV estimators in model 1 for different values of the jump size standard deviation ξ.

Table 2.5: COMPARISON OF THE MSE (IV)

	$b = 0.01$	$b = 0.2$	$b = 0.4$	$b = 0.6$	$b = 0.8$	$b = 1$
TTSRV	0.454	0.401	0.280	0.244	0.223	0.182
TSRV	0.454	1.089	3.278	4.397	6.085	9.025
TRV	1.405	2.528	2.131	1.970	1.811	1.629
MBV	1.786	1.721	1.744	1.456	2.033	2.121
BPV	3.004	2.868	2.929	2.673	3.009	2.553

The table reports the MSE of the TTSRV, TSRV, TRV, MBV and BPV estimators in model 2 for different values of the VG process scale parameter b

2.7.2 Technical Appendix

Proof of Theorem 2.1. We start proving the first inequality of Theorem 2.1. For each ω, set $J_{0,h} = \{i \in \{1, ..., m\} : N_{i+K_1-1} - N_{i-K_1} = 0\}$. It suffices to prove that a.s. for small h

$$\frac{\sup_{i\in J_{0,h}} |\beta_i|}{r(h)} \leq 1. \tag{2.6}$$

To evaluate this sup, using the definition of β_i, we have that a.s.

$$\begin{aligned}
\sup_{i\in J_{0,h}} |\beta_i| \leq & \sup_{i\in J_{0,h}} \left| \frac{1}{K_1} \sum_{j=i}^{i+K_1-1} \int_{t_{j-K_1}}^{t_j} a_s ds \right| + \sup_{i\in J_{0,h}} \left| \frac{1}{K_1} \sum_{j=i}^{i+K_1-1} \int_{t_{j-K_1}}^{t_j} \sigma_s dB_s \right| \\
& + \sup_{i\in J_{0,h}} \left| \frac{1}{K_1} \sum_{j=i}^{i+K_1-1} (u_j - u_{j-K_1}) \right| \\
\leq & \sup_{i\in\{1,...,m+K_1-1\}} \left| \int_{t_{i-K_1}}^{t_i} a_s ds \right| + \sup_{i\in\{1,...,m+K_1-1\}} \left| \int_{t_{i-K_1}}^{t_i} \sigma_s dB_s \right| \\
& + \sup_{i\in\{1,...,m\}} \left| \frac{1}{K_1} \sum_{j=i}^{i+K_1-1} (u_j - u_{j-K_1}) \right|.
\end{aligned}$$

We evaluate the three terms separately. We first write

$$\frac{\sup_{i\in\{1,...,m+K_1-1\}} \left| \int_{t_{i-K_1}}^{t_i} a_s ds \right|}{r(h)} = \frac{\sup_{i\in\{1,...,m+K_1\}} \left| \int_{t_{i-K_1}}^{t_i} a_s ds \right|}{\sqrt{K_1 h \log \frac{1}{h}}} \times \frac{\sqrt{K_1 h \log \frac{1}{h}}}{r(h)}.$$

Then, by assumptions (2) and (4), a.s. for small h the left hand side is bounded by 1.

For the second term, we write

$$\begin{aligned}
\frac{\sup_{i\in\{1,...,m+K_1-1\}} \left| \int_{t_{i-K_1}}^{t_i} \sigma_s dB_s \right|}{r(h)} \leq & \sup_{i\in\{1,...,m+K_1-1\}} \frac{\left| \int_{t_{i-K_1}}^{t_i} \sigma_s dB_s \right|}{\sqrt{2 \int_{t_{i-K_1}}^{t_i} \sigma_s^2 ds \log \frac{1}{\int_{t_{i-K_1}}^{t_i} \sigma_s^2 ds}}} \times \\
\sup_{i\in\{1,...,m+K_1-1\}} & \frac{\sqrt{2 \int_{t_{i-K_1}}^{t_i} \sigma_s^2 ds \log \frac{1}{\int_{t_{i-K_1}}^{t_i} \sigma_s^2 ds}}}{\sqrt{2K_1 hM \log \frac{1}{K_1 hM}}} \times \frac{\sqrt{2K_1 hM \log \frac{1}{K_1 hM}}}{\sqrt{2K_1 h \log \frac{1}{K_1 h}}} \times \frac{\sqrt{2K_1 h \log \frac{1}{K_1 h}}}{r(h)},
\end{aligned} \tag{2.7}$$

where M is the random constant of assumption (3).

Since $\int_{t_{i-K_1}}^{t_i} \sigma_s dB_s$ is a time changed Brownian motion, by Karatzas and Shreve (1999,

p.114 theorem 9.25), we have that a.s.

$$\limsup_{h\to 0} \sup_{i\in\{1,\ldots,m+K_1-1\}} \frac{\left|\int_{t_{i-K_1}}^{t_i} \sigma_s dB_s\right|}{\sqrt{2\int_{t_{i-K_1}}^{t_i} \sigma_s^2 ds \log \frac{1}{\int_{t_{i-K_1}}^{t_i} \sigma_s^2 ds}}} \leq 1.$$

Moreover, by assumption (3) and the monotonicity of the function $f(x) = x \log \frac{1}{x}$, we get that a.s.

$$\limsup_{h\to 0} \sup_{i\in\{1,\ldots,m+K_1-1\}} \frac{\sqrt{2\int_{t_{i-K_1}}^{t_i} \sigma_s^2 ds \log \frac{1}{\int_{t_{i-K_1}}^{t_i} \sigma_s^2 ds}}}{\sqrt{2K_1 hM \log \frac{1}{K_1 hM}}} \leq 1.$$

Finally, since $K_1 h \to 0$ as $h \to 0$, and $M(\omega) < \infty$, we have that a.s.

$$\lim_{h\to 0} \frac{\sqrt{2K_1 hM \log \frac{1}{K_1 hM}}}{\sqrt{2K_1 h \log \frac{1}{K_1 h}}} = \sqrt{M}.$$

Therefore, using assumption (4), we conclude that a.s. for small h the left hand side of (2.7) is bounded by 1.

For the last term, we write

$$\frac{\sup_{i\in\{1,\ldots,m\}} \left|\frac{1}{K_1}\sum_{j=i}^{i+K_1-1}(u_j - u_{j-K_1})\right|}{r(h)} = \frac{\sup_{i\in\{1,\ldots,m\}} \left|\frac{1}{K_1}\sum_{j=i}^{i+K_1-1}(u_j - u_{j-K_1})\right|}{\sqrt{\frac{\log\frac{1}{h}}{K_1}}} \times \frac{\sqrt{\frac{\log\frac{1}{h}}{K_1}}}{r(h)}. \quad (2.8)$$

Now, since $\frac{1}{K_1}\sum_{j=i}^{i+K_1-1}(u_j - u_{j-K_1}) \stackrel{\mathcal{L}}{=} N\left(0, \frac{2\eta^2}{K_1}\right)$, we have that for all $x > 0$,

$$\mathrm{P}\left(\sup_{i\in\{1,\ldots,m\}} \left|\frac{1}{K_1}\sum_{j=i}^{i+K_1-1}(u_j - u_{j-K_1})\right| \geq x\right) \leq 2m \exp\left(-\frac{x^2 K_1}{2\eta^2}\right).$$

Therefore, for all $\epsilon > 0$,

$$\sum_{m\geq 1} \mathrm{P}\left(\frac{\sup_{i\in\{1,\ldots,m\}} \left|\frac{1}{K_1}\sum_{j=i}^{i+K_1-1}(u_j - u_{j-K_1})\right|}{\sqrt{\frac{\log m}{K_1}}} \geq \sqrt{(2+\epsilon)2\eta^2}\right) \leq 2\sum_{m\geq 1} m^{-(1+\epsilon)} < \infty.$$

Thus, using Borell-Cantelli Lemma and letting $\epsilon \downarrow 0$, we obtain that a.s.

$$\limsup_{h\to 0} \frac{\sup_{i\in\{1,\dots,m\}} \left| \frac{1}{K_1} \sum_{j=i}^{i+K_1-1} (u_j - u_{j-K_1}) \right|}{\sqrt{\frac{\log \frac{1}{h}}{K_1}}} \leq 2\eta,$$

which together with assumption (4) implies that a.s. for h small the left hand side of (2.8) is bounded by 1. This concludes the proof of (2.6).

We next show the second inequality of Theorem 2.1. For each ω, set $J_{1,h} = \{i \in \{1, ..., m\} : N_i - N_{i-1} > 0\}$. It suffices to prove that a.s. for small h

$$\frac{\inf_{i\in J_{1,h}} |\beta_i|}{r(h)} > 1. \tag{2.9}$$

In order to evaluate this infimum, observe that for all $i \in J_{1,h}$, $|\beta_i| = |A_i + B_i|$, where

$$A_i = \frac{1}{K_1} \sum_{j=i}^{i+K_1-1} \left(\int_{t_{j-K_1}}^{t_j} a_s ds + \int_{t_{j-K_1}}^{t_j} \sigma_s dB_s + (u_j - u_{j-K_1}) \right) \qquad \text{and}$$

$$B_i = \frac{1}{K_1} \sum_{j=i}^{i+K_1-1} \sum_{\ell=1}^{N_{t_j} - N_{t_{j-K_1}}} Y_\ell.$$

In the first part of the theorem we have shown that a.s. the term $\frac{A_i}{r(h)}$ tends to 0 as $h \to 0$ uniformly with respect to i. On the other hand, we observe that

$$t_{i+K_1-1} - t_{i-K_1} \leq 2K_1 h \to 0 \quad \text{as} \quad h \to 0.$$

Therefore, a.s. for h small we have that for each i, $N_{i+K_1-1} - N_{i-K_1} \leq 1$. Thus, a.s.

$$\lim_{h\to 0} \frac{\inf_{i\in J_{1,h}} |\beta_i|}{r(h)} \geq \lim_{h\to 0} \frac{\inf_{i\in J_{1,h}} |B_i|}{r(h)} \geq \lim_{h\to 0} \frac{\inf_{i\in N_1} |Y_i|}{r(h)} = \infty,$$

where in the last equality we have used assumptions (1) and (4). This proves (2.9). □

Proof of Proposition 2.1. For simplicity, we assume that the dirft is zero. Since we are assuming that $J = 0$, we have that

$$\widehat{\sigma}^2_{\mathsf{TS}} - \int_0^1 \sigma_s^2 ds = A + B + C,$$

where

$$A = \left(\frac{1}{K} \sum_{i=K}^{m} \left(\int_{t_{i-K}}^{t_i} \sigma_s dB_s \right)^2 - \int_0^1 \sigma_s^2 ds \right) - \frac{1}{K} \sum_{i=K}^{m} \left(\int_{t_{i-1}}^{t_i} \sigma_s dB_s \right)^2,$$

$$B = \frac{2}{K} \sum_{i=K}^{m} \left(\int_{t_{i-K}}^{t_i} \sigma_s dB_s \right) (u_i - u_{i-K}) - \frac{2}{K} \sum_{i=K}^{m} \left(\int_{t_{i-1}}^{t_i} \sigma_s dB_s \right) (u_i - u_{i-1}),$$

$$C = \frac{1}{K} \sum_{i=0}^{K-2} u_i^2 - \frac{1}{K} \sum_{i=m-K+1}^{m-1} u_i^2 - \frac{2}{K} \sum_{i=K}^{m} u_{i-K} u_i + \frac{2}{K} \sum_{i=K}^{m} u_{i-1} u_i.$$

According to Zhang et al. (2005, Theorem 3), we have that as $m \to \infty$

$$m^{1/6} \left(\frac{1}{K} \sum_{i=K}^{m} \left(\int_{t_{i-K}}^{t_i} \sigma_s dB_s \right)^2 - \int_0^1 \sigma_s^2 ds \right) \xrightarrow{\mathcal{L}} \sqrt{\frac{4c}{3} \int_0^1 \sigma_s^4 \, ds} \, \mathrm{N}(0,1), \tag{2.10}$$

where the convergence is stable in law.

On the other hand, since $\sum_{i=K}^{m} \left(\int_{t_{i-1}}^{t_i} \sigma_s dB_s \right)^2 = O_{\mathrm{P}}(1)$ and $K = cm^{2/3}$, we get that as $m \to \infty$,

$$m^{1/6} \frac{1}{K} \sum_{i=K}^{m} \left(\int_{t_{i-1}}^{t_i} \sigma_s dB_s \right)^2 \xrightarrow{\mathrm{P}} 0.$$

Moreover, equations (A.8) and (A.13) in Zhang et al. (2005) imply that as $m \to \infty$,

$$m^{1/6} \frac{2}{K} \sum_{i=K}^{m} \left(\int_{t_{i-1}}^{t_i} \sigma_s dB_s \right) (u_i - u_{i-1}) \xrightarrow{\mathrm{P}} 0,$$

and

$$m^{1/6} \frac{2}{K} \sum_{i=K}^{m} \left(\int_{t_{i-K}}^{t_i} \sigma_s dB_s \right) (u_i - u_{i-K}) \xrightarrow{\mathrm{P}} 0.$$

Finally, straightforward computations as in Zhang et al. (2005) show that as $m \to \infty$,

$$m^{1/6} C \xrightarrow{\mathcal{L}} 2\sqrt{2} c^{-1} \eta^2 \mathrm{N}(0,1).$$

Since A and C are independent, this concludes the desired result. □

Proof of Theorem 2.2. For simplicity we assume that the dirft is zero. Set

$$\overline{x}_t = x_t - J_{1,t} = \int_0^t \sigma_s dB_s + u_t,$$

and consider the corresponding estimators

$$\overline{\sigma}^2_{\mathsf{TS}} = \frac{1}{K}\sum_{j=K}^{m}(\overline{x}_j - \overline{x}_{j-K})^2 - \frac{1}{K}\sum_{j=K}^{m}(\overline{x}_j - \overline{x}_{j-1})^2,$$

and

$$\overline{\sigma}^2_{\mathsf{TTS}} = \frac{1}{K}\sum_{j=K}^{m}(\overline{x}_j - \overline{x}_{j-K})^2\mathbf{1}_{E_j} - \frac{1}{K}\sum_{j=K}^{m}(\overline{x}_j - \overline{x}_{j-1})^2\mathbf{1}_{E_j},$$

where recall that $E_j = \{|\beta_i| \leq r(h), \quad \text{for all } i = j-K+1, \ldots, j\}$ and

$$\beta_j = \frac{1}{K_1}\sum_{i=j}^{j+K_1-1}(x_i - x_{i-K_1}).$$

Since as $m \to \infty$, $\overline{\sigma}^2_{\mathsf{TS}}$ converges in probability to IV, it suffices to show that as $m \to \infty$,

$$\overline{\sigma}^2_{\mathsf{TTS}} - \widehat{\sigma}^2_{\mathsf{TTS}} \xrightarrow{\mathrm{P}} 0, \tag{2.11}$$

and

$$\overline{\sigma}^2_{\mathsf{TS}} - \overline{\sigma}^2_{\mathsf{TTS}} \xrightarrow{\mathrm{P}} 0. \tag{2.12}$$

Proof of (2.11). By the second inequality of Theorem 2.1, we have that a.s. as $m \to \infty$, for all $j \in \{K, ..., m\}$,

$$\mathbf{1}_{E_j} = \mathbf{1}_{\{|\beta_{j-K+1}| \leq r(h)\}}\mathbf{1}_{\{|\beta_{j-K+2}| \leq r(h)\}} \cdots \mathbf{1}_{\{|\beta_j| \leq r(h)\}} \leq \mathbf{1}_{\{N_j - N_{j-K} = 0\}}.$$

Thus, a.s. as $m \to \infty$, for all $j \in \{K, ..., m\}$,

$$\left((\overline{x}_j - \overline{x}_{j-K})^2 - (x_j - x_{j-K})^2\right)\mathbf{1}_{E_j} = 0.$$

Similarly, a.s. as $m \to \infty$

$$\left((\overline{x}_j - \overline{x}_{j-1})^2 - (x_j - x_{j-1})^2\right)\mathbf{1}_{E_j} = 0.$$

Therefore, a.s. as $m \to \infty$, $\overline{\sigma}^2_{\mathsf{TTS}} - \widehat{\sigma}^2_{\mathsf{TTS}} = 0$, which proves (2.11).

Proof of (2.12). We have

$$\overline{\sigma}^2_{\mathsf{TS}} - \overline{\sigma}^2_{\mathsf{TTS}} = \frac{1}{K}\sum_{j=K}^{m}\left(\left(\overline{x}_j - \overline{x}_{j-K}\right)^2 - \left(\overline{x}_j - \overline{x}_{j-1}\right)^2\right)\mathbf{1}_{E_j^c},$$

where

$$\begin{aligned}\mathbf{1}_{E_j^c} &= \mathbf{1}_{\left\{\sup\limits_{i\in\{j-K+1,\ldots,j\}}|\beta_i|>r(h)\right\}} \\ &= \mathbf{1}_{|\beta_{j-K+1}|>r(h)} + \sum_{i=j-K+2}^{j}\left(\mathbf{1}_{|\beta_i|>r(h)}\prod_{p=j-K+1}^{i-1}\mathbf{1}_{|\beta_p|\leq r(h)}\right).\end{aligned}$$

Therefore,

$$|\overline{\sigma}^2_{\mathsf{TS}} - \overline{\sigma}^2_{\mathsf{TTS}}| \leq \sup_{\substack{i\in\{K,\ldots,m\}\\ \ell\in\{0,\ldots,K-1\}}}\left|\frac{1}{K}\sum_{j=i}^{\min(i+\ell,m)}\left(\left(\overline{x}_j - \overline{x}_{j-K}\right)^2 - \left(\overline{x}_j - \overline{x}_{j-1}\right)^2\right)\right|\sum_{i=1}^{m}\mathbf{1}_{|\beta_i|>r(h)}.$$

By the first inequality of Theorem 2.1, a.s. as $m\to\infty$,

$$\sum_{i=1}^{m}\mathbf{1}_{|\beta_i|>r(h)} \leq \sum_{i=1}^{m}\mathbf{1}_{\{N_{i+K_1-1}-N_{i-K_1}>0\}} \leq (2K_1-1)\left(N_{m+K_1-1}-N_{1-K_1}\right),$$

and $N_{m+K_1-1}-N_{1-K_1}$ is a.s. finite.

Without loss of generality, we assume that $i\in\{K,...,m-K+1\}$. Otherwise, the proof follows along the same lines. In this case, for all $i\in\{K,\ldots,m-K+1\}$ and $\ell\in\{0,\ldots,K-1\}$, $\min(i+\ell,m)=i+\ell$. Moreover, we have that

$$\frac{1}{K}\sum_{j=i}^{i+\ell}\left(\left(\overline{x}_j - \overline{x}_{j-K}\right)^2 - \left(\overline{x}_j - \overline{x}_{j-1}\right)^2\right) = D_1 + D_2 + D_3,$$

where

$$\begin{aligned}D_1 &= \frac{1}{K}\sum_{j=i}^{i+\ell}\left(\left(\int_{t_{j-K}}^{t_j}\sigma_s dB_s\right)^2 - \left(\int_{t_{j-1}}^{t_j}\sigma_s dB_s\right)^2\right), \\ D_2 &= \frac{2}{K}\sum_{j=i}^{i+\ell}\left(\left(\int_{t_{j-K}}^{t_j}\sigma_s dB_s\right)(u_j-u_{j-K}) - \left(\int_{t_{j-1}}^{t_j}\sigma_s dB_s\right)(u_j-u_{j-1})\right), \\ D_3 &= \frac{1}{K}\sum_{j=i}^{i+\ell}\left((u_j-u_{j-K})^2 - (u_j-u_{j-1})^2\right).\end{aligned}$$

We now claim that for all $k = 1, 2, 3$, as $m \to \infty$,

$$K_1 \sup_{\substack{i\in\{K,\ldots,m-K+1\}\\ \ell\in\{0,\ldots,K-1\}}} |D_k| \xrightarrow{\mathrm{P}} 0.$$

We start with D_1. Since $\ell < K$, $|D_1|$ is not larger than $\sup_{j\in\{K,\ldots,m\}} \left(\int_{t_{j-K}}^{t_j} \sigma_s dB_s\right)^2$ which is $O_{\mathrm{P}}\left(m^{-1/3}\log m\right)$. Because u is independent of σ and B, and $\frac{4}{K^2}\sum_{j=i}^{i+\ell}\left(\int_{t_{j-K}}^{t_j} \sigma_s dB_s\right)^2$ is $O_{\mathrm{P}}\left(m^{-1}\right)$, we get that $\sup|D_2|$ is $O_{\mathrm{P}}\left(m^{-1/2}\log m\right)$. Finally, since u_i is a $N(0,\eta^2)$ i.i.d sequence, $\sup|D_3|$ is $O_{\mathrm{P}}\left(m^{-1/3}\log m\right)$. Therefore, since $K_1 m^{-1/3}\log m \to 0$ as $m \to \infty$, (2.12) holds true. □

Proof of Theorem 2.3. Using the same notation as in the proof of Theorem 2.2, we have that Proposition 2.1 applies to $\overline{\sigma}^2_{\mathsf{TS}}$. Thus, it suffices to show that as $m \to \infty$,

$$m^{1/6}(\overline{\sigma}^2_{\mathsf{TTS}} - \widehat{\sigma}^2_{\mathsf{TTS}}) \xrightarrow{\mathrm{P}} 0, \tag{2.13}$$

and

$$m^{1/6}(\overline{\sigma}^2_{\mathsf{TS}} - \overline{\sigma}^2_{\mathsf{TTS}}) \xrightarrow{\mathrm{P}} 0. \tag{2.14}$$

(2.13) follows the fact that a.s. $\overline{\sigma}^2_{\mathsf{TTS}} - \widehat{\sigma}^2_{\mathsf{TTS}} = 0$ as $m \to \infty$, as shown in the proof of (2.11). On the other hand, in the proof of (2.12), we have shown that $\overline{\sigma}^2_{\mathsf{TS}} - \overline{\sigma}^2_{\mathsf{TTS}} = O_{\mathrm{P}}\left(K_1 m^{-1/3}\log m\right)$, which implies (2.14) since $K_1 m^{-1/6}\log m \to 0$ as $m \to \infty$. □

Proof of Theorem 3.3. As in the proof of Theorem 2.2, we denote by $\overline{< X, X >}^K_{T_n}$ and $\overline{s}^2_0$ the estimators $\widehat{< X, X >}^K_{T_n}$ and $\hat{s}^2_0$, where x is replaced with $\overline{x}$ (no jumps), and by $\widetilde{< X, X >}^K_{T_n}$ and $\tilde{s}^2_0$ where x is replaced with $\overline{x}$ and there is no truncation.

Following similarly as in the proof of (2.11), we can easily check that a.s. as $m \to \infty$, $\widehat{s}^2_0 = \overline{s}^2_0$. Moreover, by Zhang et al. (2005) $\widetilde{s}^2_0 \xrightarrow{\mathrm{P}} \mathrm{V}$ as $m \to \infty$. Thus, it suffices to show that $\overline{s}^2_0 - \widetilde{s}^2_0 \xrightarrow{\mathrm{P}} 0$ as $m \to \infty$.

Following similarly as in the proof of (2.12), we have that

$$\begin{aligned}
&\left| < \overline{X, X} >^K_{T_n} - < \overline{X, X} >^K_{T_{n-1}} - (< \widetilde{X, X} >^K_{T_n} - < \widetilde{X, X} >^K_{T_{n-1}} \right| \\
&\leq \sup_{\substack{i\in\{M(n-1)+1,\ldots,Mn\}\\ \ell\in\{0,\ldots,K\}}} \left| \frac{1}{K}\sum_{j=i}^{i+\ell} \overline{X}^K_j \right| \sum_{i=M(n-1)+1}^{Mn} \mathbf{1}_{|\beta_i|>r(h)},
\end{aligned}$$

where $\overline{X}_j^K = (\overline{x}_j - \overline{x}_{j-K})^2 - (\overline{x}_j - \overline{x}_{j-1})^2$, and a.s. as $m \to \infty$

$$\sum_{i=M(n-1)+1}^{Mn} \mathbf{1}_{|\beta_i|>r(h)} \leq 2K_1(N_{Mn+K_1-1} - N_{M(n-1)+1-K_1}).$$

Therefore,

$$|\overline{s}_0^2 - \widetilde{s}_0^2| \leq m^{1/3} \sum_{k=1}^{7} A_k,$$

where

$$A_1 = \sup_{\substack{i\in\{K_2,\ldots,m\}\\ \ell\in\{0,\ldots,K_2\}}} \left| \frac{1}{K_2} \sum_{j=i}^{i+\ell} \overline{X}_j^{K_2} \right|^2 4K_1^2 \sum_{n=1}^{m/M} (N_{Mn+K_1-1} - N_{M(n-1)+1-K_1})^2$$

$$A_2 = \sup_{\substack{i\in\{K_3,\ldots,m\}\\ \ell\in\{0,\ldots,K_3\}}} \left| \frac{1}{K_3} \sum_{j=i}^{i+\ell} \overline{X}_j^{K_3} \right|^2 4K_1^2 \sum_{n=1}^{m/M} (N_{Mn+K_1-1} - N_{M(n-1)+1-K_1})^2$$

$$A_3 = 2 \sup_{n\in\{1,\ldots,m/M\}} \left| < \widetilde{X,X} >_{T_n}^{K_2} - < \widetilde{X,X} >_{T_{n-1}}^{K_2} \right| \sup_{\substack{i\in\{K_2,\ldots,m\}\\ \ell\in\{0,\ldots,K_2\}}} \left| \frac{1}{K_2} \sum_{j=i}^{i+\ell} \overline{X}_j^{K_2} \right|$$
$$\times 2K_1 \sum_{j=1}^{m/M} (N_{Mj+K_1-1} - N_{M(j-1)+1-K_1})$$

$$A_4 = 2 \sup_{n\in\{1,\ldots,m/M\}} \left| < \widetilde{X,X} >_{T_n}^{K_2} - < \widetilde{X,X} >_{T_{n-1}}^{K_2} \right| \sup_{\substack{i\in\{K_3,\ldots,m\}\\ \ell\in\{0,\ldots,K_3\}}} \left| \frac{1}{K_3} \sum_{j=i}^{i+\ell} \overline{X}_j^{K_3} \right|$$
$$\times 2K_1 \sum_{j=1}^{m/M} (N_{Mj+K_1-1} - N_{M(j-1)+1-K_1})$$

$$A_5 = 2 \sup_{n\in\{1,\ldots,m/M\}} \left| < \widetilde{X,X} >_{T_n}^{K_3} - < \widetilde{X,X} >_{T_{n-1}}^{K_3} \right| \sup_{\substack{i\in\{K_2,\ldots,m\}\\ \ell\in\{0,\ldots,K_2\}}} \left| \frac{1}{K_2} \sum_{j=i}^{i+\ell} \overline{X}_j^{K_2} \right|$$
$$\times 2K_1 \sum_{j=1}^{m/M} (N_{Mj+K_1-1} - N_{M(j-1)+1-K_1})$$

$$A_6 = 2 \sup_{n\in\{1,\ldots,m/M\}} \left| < \widetilde{X,X} >_{T_n}^{K_3} - < \widetilde{X,X} >_{T_{n-1}}^{K_3} \right| \sup_{\substack{i\in\{K_3,\ldots,m\}\\ \ell\in\{0,\ldots,K_3\}}} \left| \frac{1}{K_3} \sum_{j=i}^{i+\ell} \overline{X}_j^{K_3} \right|$$
$$\times 2K_1 \sum_{j=1}^{m/M} (N_{Mj+K_1-1} - N_{M(j-1)+1-K_1})$$

$$A_7 = 2 \sup_{i\in\{K_2,\ldots,m\},\ell\in\{0,\ldots,K_2\}} \left| \frac{1}{K_2} \sum_{j=i}^{i+\ell} \overline{X}_j^{K_2} \right| \sup_{\substack{i\in\{K_3,\ldots,m\}\\ \ell\in\{0,\ldots,K_3\}}} \left| \frac{1}{K_3} \sum_{j=i}^{i+\ell} \overline{X}_j^{K_3} \right|$$

$$\times 4K_1^2 \sum_{n=1}^{m/M} (N_{Mn+K_1-1} - N_{M(n-1)+1-K_1})^2.$$

Also similarly as in the proof of (2.12), we have that

$$\sup_{\substack{i\in\{K_2,\ldots,m\}\\ \ell\in\{0,\ldots,K_2\}}} \left| \frac{1}{K_2} \sum_{j=i}^{i+\ell} \overline{X}_j^{K_2} \right|^2 = O_{\mathrm{P}}\left(m^{-2/3}(\log m)^2\right).$$

As m is large, there will be at most 1 jump in $(t_{M(n-1)+1-K_1}, t_{Mn+K_1-1}]$, which is contained in at most 2 intervals of this form. Thus,

$$\begin{aligned} \sum_{n=1}^{m/M} (N_{Mn+K_1-1} - N_{M(n-1)+1-K_1})^2 &= \sum_{n=1}^{m/M} (N_{Mn+K_1-1} - N_{M(n-1)+1-K_1}) \\ &\leq 2(N_{m+K_1-1} - N_{1-K_1}). \end{aligned}$$

which is a.s. finite. Finally, since $K_1^2 = o\left(\frac{m^{1/3}}{(\log m)^2}\right)$, we conclude that $m^{1/3}A_1 \xrightarrow{\mathrm{P}} 0$ as $m \to \infty$. A similar argument holds for A_2.

In order to treat the other terms, observe that by Fan et al. (2012, Theorem 1), for all $x \in [0, a_1]$ and m large,

$$\mathrm{P}\left(\sup_{n\in\{1,\ldots,\frac{m}{M}\}} \left| < \widetilde{X,X} >_{T_n}^{K_2} - < \widetilde{X,X} >_{T_{n-1}}^{K_2} - \int_{T_{n-1}}^{T_n} \sigma_s^2 ds \right| \geq x \right) \leq a_2 m \exp\left(-a_3 x^2 m^{1/3}\right).$$

Since $[T_{n-1}, T_n] = O\left(m^{-1/6}\log m\right)$, we conclude that

$$\sup_{n\in\{1,\ldots,m/M\}} \left| < \widetilde{X,X} >_{T_n}^{K_2} - < \widetilde{X,X} >_{T_{n-1}}^{K_2} \right| = O_{\mathrm{P}}(m^{-1/6}\log m).$$

Finally, since $K_1 = o\left(\frac{m^{1/6}}{(\log m)^2}\right)$, we obtain that $m^{1/3}A_3 \xrightarrow{\mathrm{P}} 0$ as $m \to \infty$. A similar argument holds for the rest of the terms, which concludes the desired result. □

We need the following preliminary lemma for the proof of Theorem 3.4.

LEMMA 2.1. *Under the assumptions of Theorem 3.4, there exists a subsequence m_k such that a.s. for any $\delta > 0$, as $k \to \infty$,*

$$\sup_{i\in\{1,\ldots,m_k\}} \sup_{t\in]t_{i-1},t_i]} (\triangle J_{2,t})^2 \mathbf{1}_{A_{i,k}} \leq \delta + 16r^2(h_k),$$

where $A_{i,k} := \left\{ \left| \frac{1}{K_1} \sum_{j=i}^{i+K_1-1} (J_{2,j} - J_{2,j-K_1}) \right| \leq 2r(h_k) \right\}$ *and* $h_k = \frac{1}{m_k}$.

Proof. Set $R^m(t) = \sum_{i=1}^{m} \left(J_{2,t\wedge t_i} - J_{2,t\wedge t_{i-1})} \right)^2$. Then according to Metivier (1982, Theorem 25.1), there exists a subsequence m_k for which a.s. as $k \to \infty$,

$$R^{m_k}(t) \to [J_2]_t \quad \text{uniformly in } \ t \in [0,1],$$

where $[J_2]_t = \sum_{s\leq t} (\triangle J_{2,s})^2$ (see Cont and Tankov (2004)). Thus, defining $f_k(t) = R^{m_k}(t) - [J_2]_t$, we have a.s. as $k \to \infty$

$$\begin{aligned} \sup_{i\in\{1,\ldots,m_k\}} \left| (J_{2,i} - J_{2,i-1})^2 - \sum_{s\in]t_{i-1},t_i]} (\triangle J_{2,s})^2 \right| &= \sup_{i\in\{1,\ldots,m_k\}} |f_k(t_i) - f_k(t_{i-1})| \\ &\leq 2 \sup_{t\in[0,1]} |f_k(t)| \to 0. \end{aligned}$$

Therefore, given arbitrary $\delta > 0$, as $k \to \infty$, we have

$$\sup_{i\in\{1,\ldots,m_k\}} \left| (J_{2,i} - J_{2,i-1})^2 - \sum_{s\in]t_{i-1},t_i]} (\triangle J_{2,s})^2 \right| < \delta.$$

Therefore, we are left to show that a.s. as $k \to \infty$,

$$\sup_{i\in\{1,\ldots,m_k\}} (J_{2,i} - J_{2,i-1})^2 \mathbf{1}_{A_{i,k}} \leq 16 r^2(h_k). \tag{2.15}$$

We write,

$$\begin{aligned} &|J_{2,i} - J_{2,i-1}| \mathbf{1}_{A_{i,k}} \left(\mathbf{1}_{\{|J_{2,i}-J_{2,i-1}|\leq 4r(h_k)\}} + \mathbf{1}_{\{|J_{2,i}-J_{2,i-1}|>4r(h_k)\}} \right) \\ &\leq |J_{2,i} - J_{2,i-1}| \left(\mathbf{1}_{\{|J_{2,i}-J_{2,i-1}|\leq 4r(h_k)\}} + \mathbf{1}_{\{|J_{2,i}-J_{2,i-1}|>4r(h)\}} \left(\mathbf{1}_{B_{i,k}^c} + \mathbf{1}_{B_{i,k}} \mathbf{1}_{A_{i,k}} \right) \right), \end{aligned}$$

where

$$B_{i,k} = \left\{ \sup_{\substack{\ell\in\{i-K_1+1,\ldots,i+K_1-1\} \\ \ell\neq i}} |J_{2,\ell} - J_{2,\ell-1}| \leq \frac{r(h_k)}{K_1} \right\}.$$

Observe that for any $j \in \{i, ..., i + K_1 - 1\}$ it holds that

$$|J_{2,j} - J_{2,j-K_1}| \geq |J_{2,i} - J_{2,i-1}| - K_1 \sup_{\substack{\ell\in\{i-K_1+1,\ldots,i+K_1-1\} \\ \ell\neq i}} |J_{2,\ell} - J_{2,\ell-1}|.$$

Thus, on the event $\{|J_{2,i} - J_{2,i-1}| > 4r(h_k)\} \cap B_{i,k}$ we have that

$$|J_{2,j} - J_{2,j-K_1}| \geq 4r(h_k) - \frac{r(h_k)}{K_1} \cdot K_1 = 3r(h_k).$$

On the other hand, on that same event, it holds that

$$|J_{2,i} - J_{2,i-1}| > \Big| \sum_{\substack{\ell \in \{j-K_1+1,\dots,j\} \\ \ell \neq i}} (J_{2,\ell} - J_{2,\ell-1}) \Big|,$$

which implies that all the increments $J_{2,j} - J_{2,j-K_1}$ have the same sign as $J_{2,i} - J_{2,i-1}$.

Thus, we conclude that on the event $\{|J_{2,i} - J_{2,i-1}| > 4r(h_k)\} \cap B_{i,k}$,

$$\left| \frac{1}{K_1} \sum_{j=i}^{i+K_1-1} (J_{2,j} - J_{2,j-K_1}) \right| = \frac{1}{K_1} \sum_{j=i}^{i+K_1-1} |J_{2,j} - J_{2,j-K_1}| \geq 3r(h_k).$$

This implies that $\mathbf{1}_{\{|J_{2,i}-J_{2,i-1}|>4r(h)\}} \mathbf{1}_{B_{i,k}} \mathbf{1}_{A_{i,k}} = 0$.

On the other hand, since J_2 has independent and stationary increments with zero expectation and variance ch, using Chebyshev's inequality, we get that

$$\mathrm{P}\left(\sup_{i \in \{1,\dots,m_k\}} \mathbf{1}_{\{|J_{2,i}-J_{2,i-1}|>4r(h)\}} \mathbf{1}_{B_{i,k}^c} = 1 \right) \leq C \frac{K_1^3 h_k}{r^4(h_k)} = C \left(\frac{K_1^3 \log m}{r^2(h) m^{1/3}} \right)^2 \frac{1}{K_1^3 m^{1/3} \log^2 m},$$

which tends to zero as $k \to \infty$ by hypothesis (2.5). Therefore, a.s. as $k \to \infty$,

$$\sup_{i \in \{1,\dots,m_k\}} |J_{2,i} - J_{2,i-1}| \, \mathbf{1}_{A_{i,k}} \leq \sup_{i \in \{1,\dots,m_k\}} |J_{2,i} - J_{2,i-1}| \, \mathbf{1}_{\{|J_{2,i}-J_{2,i-1}| \leq 4r(h_k)\}},$$

which shows (2.15). □

COROLLARY 2.1. *Under the assumptions of Theorem 3.4, as* $m \to \infty$

$$Z_m := \frac{1}{K} \sum_{j=K}^{m} (J_{2,j} - J_{2,j-K})^2 \mathbf{1}_{R_{j,K}} \xrightarrow{\mathrm{P}} 0.$$

where $R_{j,K} = \left\{ \sup_{i \in \{j-K+1,\dots,j\}} \left| \frac{1}{K_1} \sum_{j=i}^{i+K_1-1} (J_{2,j} - J_{2,j-K_1}) \right| \leq 2r(h) \right\}$.

Proof. By Lemma 2.1, there exists a subsequence m_k such that a.s. for any $\delta > 0$, as $k \to \infty$,

$$\sup_{t \in]t_{j-K_k}, t_j]} |\triangle J_{2,t}| \, \mathbf{1}_{R_{j,K_k}} \leq \sqrt{\delta + 16r^2(h_k)},$$

where $K_k = cm_k^{2/3}$. Therefore, a.s. for any δ and large k,

$$Z_{m_k} \leq \frac{1}{K_k} \sum_{j=K_k}^{m_k} (J_{2,j} - J_{2,j-K_k})^2 \mathbf{1}_{\left\{ \sup_{t \in]t_{j-K_k}, t_j]} |\triangle J_{2,t}| \leq \sqrt{\delta + 16r^2(h_k)} \right\}}.$$

Define the process Y by

$$Y_t = \int_0^t \int_{|x| \leq \sqrt{\delta + 16r^2(h_k)}} x \, (\mu(dx, ds) - \nu(dx)ds) - t \int_{\sqrt{\delta + 16r^2(h_k)} \leq |x| \leq 1} x\nu(dx).$$

Then, we have that a.s. for any δ and large k,

$$Z_{m_k} \leq \frac{1}{K_k} \sum_{j=K_k}^{m_k} (Y_j - Y_{j-K_k})^2 .$$

On the other hand, observe that

$$\frac{1}{K_k} \sum_{j=K_k}^{m_k} (Y_j - Y_{j-K_k})^2 \leq \sup_{i \in \{0, \ldots, K_k - 1\}} \sum_{j=0}^{c_{i\, m_k}} \left(Y_{i+jK_k} - Y_{(i+(j-1)K_k) \vee 0} \right)^2 ,$$

where $c_{i\, m_k}$ is the largest positive integer not larger than $\frac{m_k - i}{K_k}$.

For any $0 \leq t \leq 1$, define $d_{t\, m_k}$ as the largest positive integer not larger than $\frac{t m_k}{K_k}$, $\Pi_{t\, m_k} = \{0, t - d_{t\, m_k} \frac{K_k}{m_k}, t - (d_{t\, m_k} - 1)\frac{K_k}{m_k}, \ldots, t - \frac{K_k}{m_k}, t\}$, and

$$L^{m_k}(t) = \sum_{s_i \in \Pi_{t\, m_k}} \left(Y_{s_i} - Y_{(s_i - \frac{K_k}{m_k}) \vee 0} \right)^2 .$$

Observe that the partition $\Pi_{t\, m_k}$ is different from Metivier (1982, Theorem 25.1), but since $s_i - (s_i - \frac{K_k}{m_k}) \vee 0 \to 0$ uniformly in $s_i \in [0, 1]$ as $k \to \infty$, we have the same conclusion as that theorem, that is, there exists a subsequence m_{k_ℓ} such that a.s. as $\ell \to \infty$

$$\sup_{t \in [0,1]} (L^{m_{k_\ell}}(t) - [Y]_t) \to 0.$$

This implies that a.s. as $\ell \to \infty$

$$\sup_{i\in\{0,\dots,K_{k_\ell}-1\}} \left| \sum_{j=0}^{c_i m_{k_\ell}} \left(Y_{i+jK_{k_\ell}} - Y_{(i+(j-1)K_{k_\ell})\vee 0} \right)^2 - [Y]_{t_{i+c_i m_{k_\ell}} K_{k_\ell}} \right| \to 0.$$

Since δ is arbitrary and $r(h_{k_\ell}) \to 0$ as $\ell \to \infty$, we get that a.s. as $\delta \to 0$ and $\ell \to \infty$,

$$[Y]_{t_{i+c_i m_{k_\ell}}} \leq \int_0^1 \int_{|x|\leq\sqrt{\delta+16r^2(h_{k_\ell})}} x^2 \mu(ds,dx) \to 0.$$

Therefore, a.s. as $\ell \to \infty$ and $\delta \to 0$,

$$\sup_{i\in\{0,\dots,K_{k_\ell}-1\}} \sum_{j=0}^{c_i m_{k_\ell}} \left(Y_{i+jK_{k_\ell}} - Y_{(i+(j-1)K_{k_\ell})\vee 0} \right)^2 \to 0.$$

Thus, we have shown that there exists a subsequence m_{k_ℓ} such that as $\ell \to \infty$, $Z_{m_{k_\ell}}$ tends to zero in probability. So we can extract a subsequence for which it tends to zero a.s. Reapeating our reasoning, from each subsequence $Z_{m_{k_\ell}}$ we can extract a subsequence tending to zero a.s. Thus, Z_{m_k} tends to zero in probability as $k \to \infty$. Repeating again the same reasoning, we conclude that Z_m tends to zero in probability as $m \to \infty$. □

Proof of Theorem 3.4. For simplicity we assume the drift is zero. Set $\overline{x}_t = x_t - J_{2,t}$, $\overline{\beta}_j = \frac{1}{K_1}\sum_{i=j}^{j+K_1-1}(\overline{x}_i - \overline{x}_{i-K_1})$, $\widehat{x}_t = x_t - J_{1,t}$, $\widehat{\beta}_j = \frac{1}{K_1}\sum_{i=j}^{j+K_1-1}(\widehat{x}_i - \widehat{x}_{i-K_1})$, $\widetilde{x}_t = \overline{x}_t - J_{1,t}$, and $\widetilde{\beta}_j = \frac{1}{K_1}\sum_{i=j}^{j+K_1-1}(\widetilde{x}_i - \widetilde{x}_{i-K_1})$.

Then,

$$\left| \widehat{\sigma}^2_{\mathsf{TTS}} - \int_0^1 \sigma_t^2 dt \right| \leq Y_1 + Y_2 + Y_3 + Y_4,$$

where

$$\begin{aligned}
Y_1 =& \frac{1}{K}\sum_{j=K}^{m} \left((\overline{x}_j - \overline{x}_{j-K})^2 - (\overline{x}_j - \overline{x}_{j-1})^2 \right) \mathbf{1}_{A_j} - \int_0^1 \sigma_t^2 dt \Big| \\
Y_2 =& \left| \frac{1}{K}\sum_{j=K}^{m} \left((\overline{x}_j - \overline{x}_{j-K})^2 - (\overline{x}_j - \overline{x}_{j-1})^2 \right) \left(\mathbf{1}_{E_j} - \mathbf{1}_{A_j} \right) \right| \\
Y_3 =& \frac{2}{K}\sum_{j=K}^{m} (J_{2,j} - J_{2,j-K})^2 \mathbf{1}_{E_j} + \frac{2}{K}\sum_{j=K}^{m} (J_{2,j} - J_{2,j-1})^2 \mathbf{1}_{E_j} \\
Y_4 =& \left| \frac{2}{K}\sum_{j=K}^{m} (\overline{x}_j - \overline{x}_{j-K})(J_{2,j} - J_{2,j-K}) \mathbf{1}_{E_j} \right| + \left| \frac{2}{K}\sum_{j=K}^{m} (\overline{x}_j - \overline{x}_{j-1})(J_{2,j} - J_{2,j-1}) \mathbf{1}_{E_j} \right|,
\end{aligned}$$

$$A_j = \left\{ \sup_{i\in\{j-K+1,\ldots,j\}} |\overline{\beta}_i| \leq 2r(h) \right\} \text{ and } E_j = \{|\beta_i| \leq r(h), i = j-K+1,\ldots,j\}.$$

According to Theorem 2.3, $Y_1 \xrightarrow{\mathrm{P}} 0$ as $m \to \infty$. We are left to proof that the other terms also converge to zero in probability as $m \to \infty$.

Proof that $Y_2 \xrightarrow{\mathrm{P}} 0$. Observe that

$$\mathbf{1}_{E_j} - \mathbf{1}_{A_j} = \mathbf{1}_{E_j \cap A_j^c} - \mathbf{1}_{A_j \cap E_j^c}.$$

Now, using the fact that

$$|\beta_\ell| \geq |\overline{\beta}_\ell| - \frac{1}{K_1} \left| \sum_{i=\ell}^{\ell+K_1-1} (J_{2,i} - J_{2,i-K_1}) \right|,$$

on the event $E_j \cap A_j^c$, we get that

$$\sup_{\ell\in\{j-K+1,\ldots,j+K_1-1\}} |J_{2,\ell} - J_{2,\ell-K_1}| \geq r(h),$$

which implies the event

$$C_j = \left\{ \sup_{\ell\in\{j-K-K_1+1,\ldots,j+K_1-2\}} |J_{2,\ell+1} - J_{2,\ell}| \geq \frac{r(h)}{K_1} \right\}.$$

Therefore, we have shown that

$$E_j \cap A_j^c \subset C_j \cap A_j^c. \tag{2.16}$$

By the first inequality of Theorem 2.1, we have a.s. as $m \to \infty$ for all $j \in \{K,\ldots,m\}$

$$A_j^c \subset D_j = \{N_{j+K_1-1} - N_{j-K+1-K_1} > 0\}. \tag{2.17}$$

Next, using Chebyshev's inequality, the fact that J_2 is independent of N and that $K > K_1$ for large m, we get that

$$\mathrm{P}\left(\sup_{j\in\{K,\ldots,m\}} \mathbf{1}_{C_j\cap D_j} = 1 \right) \leq m^{-1/3} \frac{K_1^2}{r^2(h)} = \frac{K_1^3 \log m}{r^2(h) m^{1/3}} \frac{1}{K_1 \log m}, \tag{2.18}$$

which tends to zero as $m \to \infty$ by hypothesis (2.5).

Consequently, we conclude that as $m \to \infty$

$$\left| \frac{1}{K} \sum_{j=K}^{m} \left((\overline{x}_j - \overline{x}_{j-K})^2 - (\overline{x}_j - \overline{x}_{j-1})^2 \right) \mathbf{1}_{E_j \cap A_j^c} \right| \xrightarrow{\mathrm{P}} 0.$$

Thus, we are left to show that as $m \to \infty$

$$\left| \frac{1}{K} \sum_{j=K}^{m} \left((\overline{x}_j - \overline{x}_{j-K})^2 - (\overline{x}_j - \overline{x}_{j-1})^2 \right) \mathbf{1}_{A_j \cap E_j^c} \right| \xrightarrow{\mathrm{P}} 0. \tag{2.19}$$

By the second inequality of Theorem 2.1, a.s. as $m \to \infty$ for all $j \in \{K, \ldots, m\}$,

$$A_j \subset \{N_j - N_{j-K} = 0\}, \tag{2.20}$$

which implies that a.s. as $m \to \infty$

$$\begin{aligned} &\frac{1}{K} \sum_{j=K}^{m} \left((\overline{x}_j - \overline{x}_{j-K})^2 - (\overline{x}_j - \overline{x}_{j-1})^2 \right) \mathbf{1}_{A_j \cap E_j^c} \\ &= \frac{1}{K} \sum_{j=K}^{m} \left((\widetilde{x}_j - \widetilde{x}_{j-K})^2 - (\widetilde{x}_j - \widetilde{x}_{j-1})^2 \right) \mathbf{1}_{A_j \cap E_j^c} \left(\mathbf{1}_{D_j} + \mathbf{1}_{D_j^c} \right). \end{aligned}$$

Similar computations as in the proof of Theorem 2.2 show that

$$\sup_{j \in \{K, \ldots, m\}} \frac{1}{K} \left| (\widetilde{x}_j - \widetilde{x}_{j-K})^2 - (\widetilde{x}_j - \widetilde{x}_{j-1})^2 \right| = O_{\mathrm{P}} \left(m^{-2/3} \log m \right).$$

On the other hand, using (2.20), we obtain that

$$\begin{aligned} \sum_{j=K}^{m} \mathbf{1}_{A_j \cap E_j^c \cap D_j} &\leq \sum_{j=K}^{m} \mathbf{1}_{\{N_j - N_{j-K} = 0\}} \mathbf{1}_{D_j} \\ &\leq \sum_{j=K}^{m} \mathbf{1}_{\{N_{j+K_1-1} - N_j > 0\}} + \sum_{j=K}^{m} \mathbf{1}_{\{N_{j-K} - N_{j-K-K_1+1} = 0\}} \\ &\leq K_1 (N_{m+K_1} - N_K) + K_1 (N_{m-K} - N_{1-K_1}). \end{aligned}$$

Therefore,

$$\begin{aligned}
&\frac{1}{K}\sum_{j=K}^{m}\left((\widetilde{x}_j-\widetilde{x}_{j-K})^2-(\widetilde{x}_j-\widetilde{x}_{j-1})^2\right)\mathbf{1}_{A_j\cap E_j^c\cap D_j}\\
&\leq \sup_{j\in\{K,\ldots,m\}}\frac{1}{K}\left|(\widetilde{x}_j-\widetilde{x}_{j-K})^2-(\widetilde{x}_j-\widetilde{x}_{j-1})^2\right|\sum_{j=K}^{m}\mathbf{1}_{A_j\cap E_j^c\cap D_j}\\
&= O_{\mathrm{P}}\left(K_1 m^{-2/3}\log m\right)=O_{\mathrm{P}}\left(\frac{K_1^3\log m}{r^2(h)m^{1/3}}\frac{r^2(h)}{K_1^2m^{1/3}}\right),
\end{aligned}$$

which tends to 0 as $m\to\infty$ by hypothesis (2.5).

Thus, we are left to show that as $m\to\infty$

$$\frac{1}{K}\sum_{j=K}^{m}\left((\widetilde{x}_j-\widetilde{x}_{j-K})^2-(\widetilde{x}_j-\widetilde{x}_{j-1})^2\right)\mathbf{1}_{A_j\cap E_j^c\cap D_j^c}\xrightarrow{\mathrm{P}}0. \tag{2.21}$$

By (2.17), we have that a.s. as $m\to\infty$ for all $j\in\{K,...,m\}$,

$$A_j\cap E_j^c\cap D_j^c=E_j^c\cap D_j^c.$$

We next show that a.s. as $m\to\infty$ for all $j\in\{K,...,m\}$,

$$E_j^c\cap D_j^c=F_j:=\left\{\sup_{i\in\{j-K+1,\ldots,j\}}|\widehat{\beta}_i|>r(h)\right\}. \tag{2.22}$$

Observe that

$$F_j=(E_j^c\cap D_j^c)\cup(F_j\cap D_j).$$

Moreover,

$$\begin{aligned}
\sup_{j\in\{K,\ldots,m\}}(F_j\cap D_j)\leq&\left\{\sup_{j\in\{1,\ldots,m\}}|\widetilde{\beta}_j|>\frac{r(h)}{2}\right\}\cup\\
&\sup_{j\in\{K,\ldots,m\}}\left(\left\{\sup_{i\in\{j-K-K_1+1,\ldots,j+K_1-2\}}|J_{2,i+1}-J_{2,i}|\geq\frac{r(h)}{2K_1}\right\}\cap D_j\right).
\end{aligned}$$

Therefore, from the second inequality of Theorem 2.1 and (2.18), we obtain (2.22).

Following similarly as in the proof of (2.12), we have that

$$\left|\frac{1}{K}\sum_{j=K}^{m}\left((\widetilde{x}_j-\widetilde{x}_{j-K})^2-(\widetilde{x}_j-\widetilde{x}_{j-1})^2\right)\mathbf{1}_{F_j}\right|$$
$$\leq \sup_{\substack{i\in\{K,\dots,m\}\\ \ell\in\{0,\dots,K\}}}\left|\frac{1}{K}\sum_{j=i}^{\min(i+\ell,m)}\left((\widetilde{x}_j-\widetilde{x}_{j-K})^2-(\widetilde{x}_j-\widetilde{x}_{j-1})^2\right)\right|\sum_{i=K}^{m}\mathbf{1}_{\left\{|\widehat{\beta}_i|>r(h)\right\}},$$

and as $m\to\infty$,

$$\sup_{\substack{i\in\{K,\dots,m\}\\ \ell\in\{0,\dots,K\}}}\left|\frac{1}{K}\sum_{j=i}^{\min(i+\ell,m)}\left((\widetilde{x}_j-\widetilde{x}_{j-K})^2-(\widetilde{x}_j-\widetilde{x}_{j-1})^2\right)\right|=O_{\mathrm{P}}\left(m^{-1/3}\log m\right).$$

Moreover, by the definition of $\widehat{\beta}_i$, we have a.s. as $m\to\infty$ that

$$\begin{aligned}\sum_{i=K}^{m}\mathbf{1}_{\left\{|\widehat{\beta}_i|>r(h)\right\}}&\leq\sum_{i=K}^{m}\mathbf{1}_{\left\{\frac{1}{K_1}\left|\sum_{\ell=i}^{i+K_1-1}\left(J_{2,\ell}-J_{2,\ell-K_1}\right)\right|>\frac{r(h)}{2}\right\}}\\&\leq\sum_{i=K}^{m}\mathbf{1}_{\left\{\sup_{\ell\in\{i-K_1,\dots,i+K_1-2\}}|J_{2,\ell+1}-J_{2,\ell}|>\frac{r(h)}{2K_1}\right\}}\\&\leq 2K_1\sum_{i=K-K_1}^{m+K_1-2}\mathbf{1}_{\left\{|J_{2,i+1}-J_{2,i}|>\frac{r(h)}{2K_1}\right\}}.\end{aligned}\tag{2.23}$$

Next, using Chebyshev's inequality, we get that

$$\mathrm{E}\left(K_1m^{-1/3}\log m\sum_{i=K-K_1}^{m+K_1-2}\mathbf{1}_{\left\{|J_{2,i+1}-J_{2,i}|>\frac{r(h)}{2K_1}\right\}}\right)\leq C\frac{K_1^3\log m}{m^{1/3}r^2(h)},$$

which tends to zero as $m\to\infty$ by hypothesis (2.5). This shows (2.21) and $Y_2\xrightarrow{\mathrm{P}}0$.

Proof of $Y_3\xrightarrow{\mathrm{P}}0$. We only treat the first term since the second can be treated similarly. Consider the event $R_{j,K}$ of Corollary 3.2. We write

$$(E_j\cap R_{j,K}^c)\subset(E_j\cap R_{j,K}^c\cap D_j^c)\cup(R_{j,K}^c\cap D_j).$$

Hence, by the definition of β_i and $\widetilde{\beta}_i$,

$$\sup_{j\in\{K,\ldots,m\}}(E_j\cap R_{j,K}^c)\leq \sup_{i\in\{1,\ldots,m\}}\{|\widetilde{\beta}_i|>r(h)\}\cup \sup_{j\in\{K,\ldots,m\}}\left(\left\{\sup_{\ell\in\{j-K-K_1+1,\ldots,j+K_1-2\}}|J_{2,\ell+1}-J_{2,\ell}|\geq\frac{r(h)}{2K_1}\right\}\cap D_j\right).$$

Therefore, by the first inequality of Theorem 2.1 and (2.18), we obtain that as $m\to\infty$,

$$\mathrm{P}\left(\sup_{j\in\{K,\ldots,m\}}\mathbf{1}_{E_j\cap R_{j,K}^c}=1\right)\longrightarrow 0. \tag{2.24}$$

Thus, we conclude that a.s. as $m\to\infty$

$$\frac{1}{K}\sum_{j=K}^{m}(J_{2,j}-J_{2,j-K})^2\mathbf{1}_{E_j\cap(R_{j,K}^c\cup R_{j,K})}\leq\frac{1}{K}\sum_{j=K}^{m}(J_{2,j}-J_{2,j-K})^2\mathbf{1}_{R_{j,K}},$$

which converges to zero in probability as $m\to\infty$ according to Corollary 3.2.

Proof of $Y_4\xrightarrow{\mathrm{P}}0$. We only treat the first term since the second can be treated similarly. By Cauchy-Schwarz inequality,

$$\left|\frac{1}{K}\sum_{j=K}^{m}\left(\int_{t_{j-K}}^{t_j}\sigma_u\,dB_u\right)(J_{2,j}-J_{2,j-K})\mathbf{1}_{E_j}\right| \leq\sqrt{\frac{1}{K}\sum_{j=K}^{m}\left(\int_{t_{j-K}}^{t_j}\sigma_u dB_u\right)^2}\sqrt{\frac{1}{K}\sum_{j=K}^{m}(J_{2,j}-J_{2,j-K})^2\mathbf{1}_{E_j}}\xrightarrow{\mathrm{P}}0,$$

as $m\to\infty$, since the first term converges in probability to $\sqrt{\int_0^1\sigma_u^2du}$ according to (3.12), and the second term converges to zero in probability as shown in the proof of $Y_3\xrightarrow{\mathrm{P}}0$.

By (2.24), we have a.s. as $m\to\infty$,

$$\left|\frac{1}{K}\sum_{j=K}^{m}(u_j-u_{j-K})(J_{2,j}-J_{2,j-K})\mathbf{1}_{E_j}\right|=\left|\frac{1}{K}\sum_{j=K}^{m}(u_j-u_{j-K})(J_{2,j}-J_{2,j-K})\mathbf{1}_{E_j\cap R_{j,K}}\right| \leq|I_1|+|I_2|,$$

where

$$I_1 = \frac{1}{K}\sum_{j=K}^{m}(u_j - u_{j-K})(J_{2,j} - J_{2,j-K})\mathbf{1}_{R_{j,K}}$$
$$I_2 = \frac{1}{K}\sum_{j=K}^{m}(u_j - u_{j-K})(J_{2,j} - J_{2,j-K})\mathbf{1}_{E_j^c \cap R_{j,K}}.$$

Now, for each $j \in \{K, \ldots, m\}$, set $d_{j\,K}$ as the largest positive integer not larger than $\frac{j}{K}$, and we separate the sum in I_1 between odd and even terms. Because the u_i are i.i.d. $N(0, \eta^2)$ independent of J_2, the random variable

$$\frac{\frac{1}{K}\sum_{odd\, d_{j\,K}}(u_j - u_{j-K})(J_{2,j} - J_{2,j-K})\mathbf{1}_{R_{j,K}}}{\sqrt{\frac{1}{K^2}\sum_{odd\, d_{j\,K}}(J_{2,j} - J_{2,j-K})^2\mathbf{1}_{R_{j,K}}}}$$

is $N(0, \eta^2)$. By Corollary 3.2, as $m \to \infty$,

$$\frac{1}{K^2}\sum_{odd\, d_{j\,K}}(J_{2,j} - J_{2,j-K})^2\mathbf{1}_{R_{j,K}} \xrightarrow{\mathrm{P}} 0.$$

Therefore, as $m \to \infty$,

$$\frac{1}{K}\sum_{odd\, d_{j\,K}}(u_j - u_{j-K})(J_{2,j} - J_{2,j-K})\mathbf{1}_{R_{j,K}} \xrightarrow{\mathrm{P}} 0.$$

A similar arugment follows for the even terms. Thus, we conclude that as $m \to \infty$, I_1 tends to zero in probability. We next treat I_2. Following similarly as in the proof of (2.12), we have that

$$|I_2| \leq \sup_{\substack{i\in\{K,\ldots,m\}\\ \ell\in\{0,\ldots,K\}}}\left|\frac{1}{K}\sum_{j=i}^{\min(i+\ell,m)}(u_j - u_{j-K})(J_{2,j} - J_{2,j-K})\mathbf{1}_{R_{j,K}}\right|\sum_{i=1}^{m}\mathbf{1}_{\{|\beta_i|>r(h)\}}.$$

Because the u_i are i.i.d. $N(0, \eta^2)$ independent of J_2, and using Corollary 3.2, we have that as $m \to \infty$,

$$\sup_{\substack{i\in\{K,\ldots,m\}\\ \ell\in\{0,\ldots,K\}}}\left|\frac{1}{K}\sum_{j=i}^{\min(i+\ell,m)}(u_j - u_{j-K})(J_{2,j} - J_{2,j-K})\mathbf{1}_{R_{j,K}}\right| = O_{\mathrm{P}}\left(\sqrt{\frac{\log m}{K}}\right).$$

On the other hand, using (2.23), we get that

$$\sum_{i=1}^{m} \mathbf{1}_{\{|\beta_i|>r(h)\}} \leq \sum_{i=1}^{m} \left(\mathbf{1}_{\left\{N_{i+K_1-1}-N_{i-K_1}>0\right\}} + \mathbf{1}_{\left\{\left|\widehat{\beta}_i\right|>r(h)\right\}} \right)$$
$$\leq 2K_1(N_{m+K_1-1} - N_{1-K_1}) + 2K_1 \sum_{i=K-K_1}^{m+K_1-2} \mathbf{1}_{\left\{|J_{2,i+1}-J_{2,i}|>\frac{r(h)}{2K_1}\right\}}.$$

Next, using Chebyshev's inequality, we have that

$$\mathrm{E}\left(K_1\sqrt{\frac{\log m}{K}} \sum_{i=K-K_1}^{m+K_1-2} \mathbf{1}_{\left\{|J_{2,i+1}-J_{2,i}|>\frac{r(h)}{2K_1}\right\}} \right) \leq C\sqrt{\frac{\log m}{K}}\frac{K_1^3}{r^2(h)} = \tilde{C}\frac{K_1^3 \log m}{m^{1/3}r^2(h)}\frac{1}{\sqrt{\log m}},$$

which tends to zero as $m \to \infty$ by hypothesis (2.5). Since $K_1\sqrt{\frac{\log m}{K}}$ also converges to zero as $m \to \infty$ by hypothesis (2.5), we conclude that I_2 tends to 0 in probability as $m \to \infty$. Thus, we have shown that as $m \to \infty$,

$$\frac{1}{K}\sum_{j=K}^{m}(u_j - u_{j-K})(J_{2,j} - J_{2,j-K})\mathbf{1}_{E_j} \xrightarrow{\mathrm{P}} 0.$$

We are left to show that as $m \to \infty$

$$\frac{1}{K}\sum_{j=K}^{m}(J_{1,j} - J_{1,j-K})(J_{2,j} - J_{2,j-K})\mathbf{1}_{E_j} \xrightarrow{\mathrm{P}} 0.$$

This follows easily since on the event $E_j \cap A_j^c$ it tends to zero by (2.16), (2.17) and (2.18), and on $E_j \cap A_j$ all the terms of the sum are zero by (2.20). Hence, we conclude that $Y_4 \xrightarrow{\mathrm{P}} 0$. $\square$

Chapter 3

DETECTING PRICE JUMPS IN THE PRESENCE OF MARKET MICROSTRUCTURE NOISE

3.1 Introduction

Discontinuities in the path of asset prices are typically modeled with jumps, and jumps are the focus of a large segment of the literature. For example, Bakshi et al. (1997) show how jumps affect option values. Moreover, Tauchen and Zhou (2011) study the influence of the jump risk factor on the variation of the credit default swap spreads. Accordingly, several contributions in the literature focus on detecting jumps based on high-frequency data, and some examples include the works by Andersen et al. (2007), Aït-Sahalia and Jacod (2009), Corsi et al. (2010), Lee and Mykland (2008), Aït-Sahalia and Jacod (2011) and Li et al. (2016).

In this chapter, we propose a novel non-parametric jump detection technique. We assume the presence of market microstructure noise in the observed high-frequency prices. In order to explore the jump arrival time, we divide the whole period under consideration into small intervals. Then for each interval we derive a statistic to detect whether there is a jump on it. The statistic is defined in terms of the TTSRV estimator proposed in Chapter 2. We compare the absolute value of the statistic with a threshold. The choice of the threshold depends on the significance level and the asymptotic distribution of the statistic when no jump is present in the interval. If the threshold is small, we reject the null hypothesis and conclude that there is a jump

in the interval. A generic jump is likely to be detected, since it significantly increases with the magnitude of the statistic. Once we have identified an interval as containing a jump, the jump size can be consistently estimated by averaging the overlapping local returns that include the interval. The test proposed by Lee and Mykland (2008) has the same goal as ours, since we both aim to detect jump arrival times. Our advantage is that this method is vulnerable to the noise which is common in financial markets according to the work by Hansen and Lunde (2012).

The simulations indicate that when the sampling frequency increases, the test becomes more efficient, and when the sampling frequency is high enough, the test reaches satisfactory performance. This supports our theory that the test is consistent. Moreover, the simulations show that jumps with larger size are easier to detect, which is in line with the intuition. The simulations also demonstrate the superiority of our test compared to the tests by Lee and Mykland (2008) and Jiang and Oomen (2008) when assuming a moderate size of the variance of the noise.

In the literature there are essentially two approaches for testing jumps: parametric and non-parametric. The parametric approach typically assumes a specific model to characterize jumps. Such models can then be estimated on the basis of observed data. For example, Chernov et al. (2003), and Pan (2002) adopt the Efficient Method of Moments (EMM) and the Generalized Method of Moments (GMM) to measure the jump-related parameters, respectively. Other parametric approaches related to jump models can be found, among others, in Bakshi et al. (1997), Bates (2000), and Piazzesi (2005) . Since jump occurence in financial markets is quite irregular, parametric approaches are likely to commit misspecification errors, and results can be unreliable. Accordingly, the results obtained can be counterintuitive and insignificant. The non-parametric jump tests typically propose statistics to detect the existence of realized jumps, since the jumps can significantly affect the behavior of the statistics. For example, Andersen et al. (2007) consider the difference between the bipower variation (BPV) and the realized volatility (RV) as the statistic to test jumps. When there are no jumps, both the BPV and the RV are consistent estimators of the quadratic variation of the asset price process. Jumps can increase the difference between the BPV and the RV, since the BPV is immune to jumps and the RV is not. Jiang and Oomen (2008) explore jumps also by comparing the RV to another volatility measure, since their difference is sensitive to jumps. Some other non-parametric jump detection methods can be found, among others, in Aït-Sahalia and Jacod (2009), Aït-Sahalia et al. (2012), Corsi et al. (2010) and Lee and Mykland (2008).

However, most of the approaches concerning non-parametric jump tests are not able to pin

down the precise jump arrival time or size. That is, they can only test whether there are jumps on a given period, without further information on their number, arrival time or size. Another obstacle of jump detection is market microstructure noise. It makes the observed price different from the efficient one, and it turns out that most non-parametric jump tests are invalid in this sense. A notable exception appears in Aït-Sahalia et al. (2012), where a statistic robust to noise is proposed. Jiang and Oomen (2008) also propose a noise robust statistic, but they require the spot volatility to be constant, which is hardly true in reality. In this chapter, we introduce a methodology that allows to detect jump times and sizes in the presence of noise.

The rest of this chapter is organized as follows: In Section 3.2 we define the basic setting, the statistic and discuss its asymptotic distribution. In Section 3.3 we obtain the limiting probabilities of testing errors. In Section 3.4 we perform Monte Carlo simulations to evaluate the efficiency of our test and compare it to the tests proposed by Lee and Mykland (2008) and Jiang and Oomen (2008). Section 3.5 concludes.

3.2 Framework and Asymptotic Theory

We use y_t to denote the value of the efficient log–price of some asset at time t, where $t \in [0,1]$. The process y_t starts at $y_0 \in \mathbb{R}$, and is defined on a filtered probability space $(\Omega, (\mathcal{F}_t)_{t\geq 0}, \mathcal{F}, \mathrm{P})$. Its dynamics are characterized by

$$dy_t = a_t dt + \sigma_t dB_t + dJ_t, \qquad t \in]0,1]. \tag{3.1}$$

Here B_t is a standard Brownian motion. The drift a_t and the spot volatility σ_t are progressively measurable processes on $[0,1]$. J_t is a finite activity jump process independent of (σ, B), of the form $J_t = \sum_{i=1}^{N_t} Y_i$. N_t is a non-explosive counting process, and Y_i are i.i.d. random variables and independent of N. Thus the solution $(y_t, t \in [0,1])$ to equation (3.1) is unique in the strong sense, adapted and càdlàg (see for example Ikeda and Watanabe (1981)).

In addition, we assume that due to the presence of microstructure noise in the market, the efficient price y_t is not observable. That is, the observed transaction price denoted as x_t is different from y_t. To this extent, we assume that the timestamps when we observe the transaction

prices are $0 = t_0 < t_1 < \cdots < t_m = 1$ where $t_i = \frac{i}{m}$, and x_{t_i} and y_{t_i} are related by

$$x_{t_i} = y_{t_i} + u_{t_i}, \quad i = 1, ..., m,$$

where u_{t_i} is the noise component for the ith trade. u_{t_i} is assumed to be a discrete i.i.d process, independent of the process y_t and such that $u_{t_i} \sim N(0, \eta^2)$, where η is a positive constant. We use x_i, y_i and u_i to denote respectively the processes x_{t_i}, y_{t_i} and u_{t_i} to simplify the notation.

Moreover, we assume the following conditions on the dynamics of a_t and σ_t.

ASSUMPTION 3.1. *σ_t is positive, and there exists constants $a > 0, \sigma^+ > \sigma^- > 0$ such that $|a_t| < a$, $\sigma^- < \sigma_t < \sigma^+$ for all $t \in (0, 1]$.*

ASSUMPTION 3.2. *For any constants $c > 0, \epsilon > 0, 0 < \alpha < 1$,*

$$\sup_i \sup_{t_i \le u \le t_i + cm^{\alpha-1}} |\sigma_u - \sigma_{t_i}| = O_{\mathrm{P}}(m^{\frac{1}{2}(\alpha-1)+\epsilon}), \tag{3.2}$$

as $m \to \infty$.

Intuitively, Assumption 3.2 implies that the spot volatility σ_t does not change much over a short time interval. It is reasonable in view of Lemma 2 in Mykland and Zhang (2006), which shows that (3.2) holds when σ_t is a general Itô process. For example, in Podolskij and Vetter (2009) and Christensen et al. (2010), they assume that the dynamics of σ_t are:

$$d\sigma_t = b_t dt + c_t dW_t + d_t dV_t, \tag{3.3}$$

where b_t, c_t, d_t are bounded and adapted processes, and W_t, V_t are independent Brownian motions.

3.2.1 The statistic

In this subsection we define the statistic that will be used in order to test jumps. As mentioned, we divide the whole period into many small intervals, and detect whether there is a jump on each interval.Specifically, as shown in Figure 3.1, the period $(0, 1]$ is splitted into $\frac{m}{2K_1}$ intervals, and the ith interval is $\left(\frac{2(i-1)K_1}{m}, \frac{2iK_1}{m}\right]$, where $K_1 = m^{\alpha_1}$ and $\alpha_1 \in (0, 1)$. The value of α_1 will be choosen later, and for simplicity, we assume $\frac{m}{2K_1}$ is an integer.

In order to construct the statistic for the interval $\left(\frac{2(i-1)K_1}{m}, \frac{2iK_1}{m}\right]$, we first define the rescaled sum of returns that will smooth away the effect of the noise by:

$$A_i = M \sum_{j=K_1}^{2K_1} \left(x_{j+2(i-1)K_1} - x_{j-K_1+2(i-1)K_1}\right), \tag{3.4}$$

where $M = \sqrt{\frac{3m}{K_1(K_1+1)(2K_1+1)}}$. As shown in Figure 3.2, A_i is obtained by rescaling the sum of the returns over the intervals that lie inside $\left(\frac{2(i-1)K_1}{m}, \frac{2iK_1}{m}\right]$. Each of these intervals has the form $\left(\frac{2(i-1)K_1+j-1}{m}, \frac{(2i-1)K_1+j-1}{m}\right]$, for $j \in \{K_1, K_1+1, \ldots, 2K_1\}$. For large m, since the contribution to A_i by the drift a_t and the noise u_j is negligible, A_i is dominated by the dynamics of (σ_t, B_t). When there is no jump on $\left(\frac{2(i-1)K_1}{m}, \frac{2iK_1}{m}\right]$, it can be checked that

$$A_i \approx \sigma_{\frac{2(i-1)K_1}{m}} \cdot Z, \tag{3.5}$$

where Z is a standard normal random variable.

In order to get rid of the spot volatility $\sigma_{\frac{2(i-1)K_1}{m}}$ in the asymptotic distribution of A_i, we need to construct an estimator of $\sigma_{\frac{2(i-1)K_1}{m}}$. We will use the TTSRV introduced in Chapter 2 to define an estimator of $\int_{\frac{2(i-1)K_1-K_2}{m}}^{\frac{2(i-1)K_1}{m}} \sigma_t^2 dt$. Here $K_2 = cm^{\alpha_2}$ and $\alpha_2 \in (0,1)$ whose value will be specified later. Because of Assumption 1, $\int_{\frac{2(i-1)K_1-K_2}{m}}^{\frac{2(i-1)K_1}{m}} \sigma_t^2 dt \approx \frac{K_2}{m}\sigma^2_{\frac{2(i-1)K_1}{m}}$. Then we can estimate $\sigma_{\frac{2(i-1)K_1}{m}}$ by taking the square root of the rescaled TTSRV. That is,

$$\widehat{\sigma}_{\frac{2(i-1)K_1}{m}} = \sqrt{\frac{m}{K_2}\widehat{\sigma}^2_{\mathsf{TTS}\,i}}. \tag{3.6}$$

Observe the TTSRV in the interval $\left[\frac{2(i-1)K_1-K_2}{m}, \frac{2(i-1)K_1}{m}\right]$, we have

$$\widehat{\sigma}^2_{\mathsf{TTS}\,i} = \frac{1}{K_3} \sum_{j=2(i-1)K_1-K_2+K_3}^{2(i-1)K_1} (x_j - x_{j-K_3})^2 \mathbf{1}_{E_j} - \frac{1}{K_3} \sum_{j=2(i-1)K_1-K_2+K_3}^{2(i-1)K_1} (x_j - x_{j-1})^2 \mathbf{1}_{E_j}, \tag{3.7}$$

where $E_j = \{|\beta_i(K_4)| \leq r(m), \quad \text{for all } i = j-K_3+1, \ldots, j\}$. Here $K_3 = m^{\alpha_3}, K_4 = m^{\alpha_4}$ and $r(m) = m^{-\alpha_5}$, and $\alpha_3, \alpha_4, \alpha_5 \in (0,1)$.

Then the statistic S_i can be obtained by standardizing A_i with $\widehat{\sigma}_{\frac{2(i-1)K_1}{m}}$, that is,

$$S_i = \frac{A_i}{\widehat{\sigma}_{\frac{2(i-1)K_1}{m}}}. \tag{3.8}$$

The following lemma shows how to choose the values of $\alpha_2, \alpha_3, \alpha_4, \alpha_5$ in order to make the denominator a consistent estimator of $\sigma_{\frac{2(i-1)K_1}{m}}$.

LEMMA 3.1. *Assume* $0 < \alpha_3 < \alpha_2 < 1$, $\alpha_5 < \frac{\alpha_4}{2}$ *and* $0 < \alpha_4 < \frac{1}{4}$. *Then for any* $\epsilon > 0$, *we have*

$$\sup_{i \in \{1,\ldots,\frac{m}{2K_1}\}} \left| \frac{m}{K_2} \widehat{\sigma}^2_{\mathsf{TTS}\,i} - \sigma^2_{t_{2(i-1)K_1}} \right| = O_{\mathrm{P}}\left(m^{\widetilde{\alpha}+\epsilon}\right), \tag{3.9}$$

as $m \to \infty$, *where* $\widetilde{\alpha} = \max(\frac{1}{2}(\alpha_3 - \alpha_2), 1 - \frac{1}{2}\alpha_2 - \alpha_3, \frac{1}{2}(\alpha_2 - 1), \frac{2}{3} + \alpha_4 - \alpha_2)$.

The proof of Lemma 3.1 is similar to the proof of the consistency of the TTSRV in Chapter 2. First we show that since the returns affected by jumps are removed from $\widehat{\sigma}^2_{\mathsf{TTS}\,i}$, the difference between $\widehat{\sigma}^2_{\mathsf{TTS}\,i}$ constructed with the jump-involved price data and the TSRV based on the no-jump data is caused only by the information loss due to the truncations. Then we compute the magnitude of the information loss, and the difference between the TSRV and $\int_{t_{2(i-1)K_1-K_2}}^{t_{2(i-1)K_1}} \sigma_t^2 dt$.

Given (3.9), in order to minimize the difference between $\frac{m}{K_2}\widehat{\sigma}^2_{\mathsf{TTS}\,i}$ and $\sigma^2_{t_{2(i-1)K_1}}$, it is natural to minimize $\widetilde{\alpha}$ over the values of $\alpha_2, \alpha_3, \alpha_4$. It is easy to check that the minimum of $\widetilde{\alpha}$ is $-\frac{1}{12}$, which is achieved when $\alpha_2 = \frac{5}{6}, \alpha_3 = \frac{2}{3}$, $0 < \alpha_4 < \frac{1}{12}$ and $0 < \alpha_5 < \frac{\alpha_4}{2}$. In the rest of the chapter we will assume this set of values for $\alpha_2, \alpha_3, \alpha_4, \alpha_5$ unless stated otherwise.

3.2.2 Under the null hypothesis

In this subsection we compute the asymptotic distribution of S_i under the null hypothesis that there is no jump on $\left(\frac{2(i-1)K_1}{m}, \frac{2iK_1}{m}\right]$. Since for large m, A_i is roughly $\sigma_{\frac{2(i-1)K_1}{m}}$ times a standard normal random variable, and $\widehat{\sigma}_{\frac{2(i-1)K_1}{m}}$ is approximately $\sigma_{\frac{2(i-1)K_1}{m}}$, S_i is asymptotically standard normal, as shown in the following theorem where

$$B_m = \left\{ i \in \left\{1, \ldots, \frac{m}{2K_1}\right\}, \text{ such that } N_{t_{2iK_1}} - N_{t_{2(i-1)K_1}} = 0 \right\}. \tag{3.10}$$

THEOREM 3.1. *For any $\epsilon > 0$, as $m \to \infty$,*

$$\sup_{i \in B_m} |S_i - \widehat{S}_i| \;=\; O_{\mathrm{P}}\left(m^{\beta+\epsilon}\right), \tag{3.11}$$

where $\beta = \max\left(\frac{1}{2} - \alpha_1, \frac{1}{2}\alpha_1 - \frac{1}{2}, -\frac{1}{12}\right)$, and

$$\widehat{S}_i \;=\; M \sum_{j=0}^{K_1} \left(B_{t_{(2i-1)K_1+j}} - B_{t_{2(i-1)K_1+j}}\right). \tag{3.12}$$

Moreover, $\widehat{S}_i$ is a standard normal random variable.

Let $\alpha_1 \in \left(\frac{1}{2}, 1\right)$ as a constant on $\left(\frac{1}{2}, 1\right)$. Then Theorem 3.1 asserts that the difference between S_i and $\widehat{S}_i$ is $o_{\mathrm{P}}(1)$, which means that asymptotically, S_i is also standard normal. In addition, the $\widehat{S}_i$'s are pairwise independent. Thus asymptotically, so are the S_i's.

3.2.3 Under the alternative hypothesis

In this subsection we study the property of S_i under the alternative hypothesis that there is a jump on $(t_{2(i-1)K_1}, t_{2iK_1}]$. Then the following Theorem 3.2 can help us check whether a jump can significantly affect the behavior of S_i, which is important for jump detection.

THEOREM 3.2. *Suppose there is a jump at time τ and i is the integer such that $\tau \in (t_{2(i-1)K_1}, t_{2iK_1}]$. Then for any $\epsilon > 0$ we have*

$$S_i \;=\; \widehat{S}_i + M\frac{D_m(\tau)}{\sigma_{t_{2(i-1)K_1}}} Y(\tau) + o_{\mathrm{P}}\left(m^{\beta+\epsilon} + MD_m(\tau)\right), \tag{3.13}$$

where $\widehat{S}_i$ is defined in Theorem 3.1, $Y(\tau)$ is the jump size at time τ, and

$$D_m(\tau) = \min\left(\left[\left(\tau - \frac{2(i-1)K_1}{m}\right)m\right] + 1, K_1, \left[\left(\frac{2iK_1}{m} - \tau\right)m\right] + 1\right),$$

where $[x]$ means the largest integer not larger than x.

$D_m(\tau)$ shows how the jump time τ affects the contribution by the jump to S_i. In order to know the magnitude of $D_m(\tau)$ when τ is random, we assume the following condition for N_t:

$$\mathrm{P}\left(N_{t_i} - N_{t_{i-1}} > 0\right) = O\left(\frac{1}{m}\right), \tag{3.14}$$

for all $i \in \{1, \ldots, m\}$. Notice that this condition holds for the Poisson process, and it is also adopted by Mancini (2009). Then we have the following corollary:

COROLLARY 3.1. *For any* $\epsilon > 0$*,*

$$\frac{D_m(\tau)}{K_1^{1-\epsilon}} \xrightarrow{\mathrm{P}} \infty, \tag{3.15}$$

as $m \to \infty$*.*

Corollary 3.1 implies that $D_m(\tau)$ is approximately of the same order as K_1. The intuition is that when K_1 is large, $D_m(\tau)$ will also be large unless τ is close to $t_{2(i-1)K_1}$ or t_{2iK_1}, but the probability that τ is near $t_{2(i-1)K_1}$ or t_{2iK_1} is small. Given the second component on the right-side of (3.13), the contribution to S_i by the jump is approximately of order $m^{\frac{1}{2}-\frac{1}{2}\alpha_1}$. As $\alpha_1 \in \left(\frac{1}{2}, 1\right)$, the jump is likely to make the value of $|S_i|$ large for large m.

3.2.4 Selection of rejection region

From Theorems 3.1 and 3.2, when there is a jump on $(t_{2(i-1)K_1}, t_{2iK_1}]$, $|S_i|$ is likely to be larger than $|S_j|$ for a generic j such that there is no jump on $(t_{2(i-1)K_1}, t_{2iK_1}]$. From the next lemma, we can see that $|S_i|$ is also likely to be larger than the maximum of $|S_j|$ across different values of j such that there is no jump on $(t_{2(i-1)K_1}, t_{2iK_1}]$.

LEMMA 3.2. *As* $m \to \infty$*,*

$$\frac{\max\limits_{i \in B_m} |S_i| - C_m}{L_m} \to X, \tag{3.16}$$

where $\mathrm{P}(X \leq x) = \exp\left(-e^{-x}\right)$*,*

$$C_m = \left(2 \log \frac{m}{2K_1}\right)^{\frac{1}{2}} - \frac{\log \pi + \log\left(\log \frac{m}{2K_1}\right)}{2\left(2 \log \frac{m}{2K_1}\right)^{\frac{1}{2}}}$$

and $L_m = \frac{1}{\left(2 \log \frac{m}{2K_1}\right)^{\frac{1}{2}}}$.

The principle of our test is that if $|S_i|$ is too large compared with the asymptotic distribution of $\max\limits_{i \in B_m} |S_i|$, we reject the null hypothesis that there is no jump on $(t_{2(i-1)K_1}, t_{2iK_1}]$. Thus the value of the threshold for the test is determined by the significance level and the asymptotic distribution of $\max\limits_{i \in B_m} |S_i|$. For example, if we define the significance level as 5%, because the 95% quantile of the random variable X is 2.97, as $\exp\left(-e^{-2.97}\right) = 0.95$, we believe there is a jump on $(t_{2(i-1)K_1}, t_{2iK_1}]$ if $\frac{|S_i| - c_m}{L_m} > 2.97$.

3.3 Misclassifications

We consider four types of misclassifications. The first is such that a jump occurs in interval $(t_{2(i-1)K_1}, t_{2iK_1}]$, but the test does not detect it. We call this a $failure\ to\ detect\ actual\ jump\ in\ (t_{2(i-1)K_1}, t_{2iK_1}]$ (FTD_i). The second is such that the test indicates there is a jump in $(t_{2(i-1)K_1}, t_{2iK_1}]$, but actually there is not. We call this a $spurious\ detection\ of\ jump\ in$ $(t_{2(i-1)K_1}, t_{2iK_1}]$ (SD_i). If we commit an FTD_i for any $i \in \left\{1, \ldots, \frac{m}{2K_1}\right\}$, a $global\ failure\ to$ $detect\ actual\ jump$ $(GFTD)$ happens. In other words, we commit a $GFTD$ unless we detect all the jumps in $(0, 1]$. If we commit an SD_i for any $i \in \left\{1, \ldots, \frac{m}{2K_1}\right\}$, a $global\ spurious$ $detection\ of\ jump$ (GSD) happens. Thus we commit a GSD unless we correctly identify all the intervals $(t_{2(i-1)K_1}, t_{2iK_1}]$ where $i \in \overline{B}_m$ as not containing a jump. We define $\overline{J}_i$ as the event that there is a jump in $(t_{2(i-1)K_1}, t_{2iK_1}]$, and E_i as the event that based on the value of $|S_i|$, we declare there is a jump in $(t_{2(i-1)K_1}, t_{2iK_1}]$. Then the four types of misclassifications can be expressed as:

$$failure\ to\ detect\ actual\ jump\ in\ (t_{2(i-1)K_1}, t_{2iK_1}](local\ property)(FTD_i) = J_i \bigcap E_i^c,$$

$$spurious\ detection\ of\ jump\ in\ (t_{2(i-1)K_1}, t_{2iK_1}](local\ property)(SD_i) = J_i^c \bigcap E_i,$$

$$failure\ to\ detect\ actual\ jumps(global\ property)(GFTD) = \bigcup_{i=1}^{\frac{m}{2K_1}} \left(J_i \bigcap E_i^c\right),$$

$$spurious\ detection\ of\ jumps(global\ property)(GSD) = \bigcup_{i=1}^{\frac{m}{2K_1}} \left(J_i^c \bigcap E_i,\right).$$

Define α_m as the significance level of our test, γ_m as the $(100 - 100\alpha_m)th$ quantile of the distribution of X in Lemma 3.2, and $F(y)$ as the cumulative distribution function of the absolute value of jump size. Then Theorem 3.3 shows the limiting probability of $GFTD$ as $m \to \infty$.

THEOREM 3.3. *Assume* $1-\alpha_m$ *bounded away from zero. Let* $\{\tau_1, \tau_2, \ldots, \tau_{N_1}\}$ *be the time points when jumps occur on* $(0, 1]$. *Then*

$$\mathrm{P}(GFTD|\{\tau_1, \tau_2, \ldots, \tau_{N_1}\}) \to 1 - \prod_{i=1}^{N_1} (1 - F(G_{i\,m})) \tag{3.17}$$

as $m \to \infty$*, where*

$$G_{i\,m} = \frac{1}{M}\frac{\sigma_{t_{2(i-1)K_1}}}{D_{i\,m}}(\gamma_m L_m + C_m),$$

$$D_{i\,m} = \min\left(\left[\left(\tau_i - \frac{2(b_{i\,m}-1)K_1}{m}\right)m\right]+1, K_1, \left[\left(\frac{2b_{i\,m}K_1}{m} - \tau_i\right)m\right]+1\right),$$

and $b_{i\,m} = \left[\frac{m\tau_i}{2K_1}\right] + 1$.

Now we assume for each i, Y_i is a.s. nonzero. Then the following corollary shows how to set the value of γ_m such that the probability of $GFTD$ converges to zero as $m \to \infty$.

COROLLARY 3.2. *Set* γ_m *such that there exists* $\epsilon > 0$ *such that*

$$\frac{\gamma_m}{\sqrt{\log(m)}} m^{-\frac{1}{2}+\frac{1}{2}\alpha_1+\epsilon} \to 0 \tag{3.18}$$

as $m \to \infty$*. Then we have*

$$\mathrm{P}(GFTD) \to 0, \tag{3.19}$$

as $m \to \infty$.

The event that $GFTD$ does not occur means that for the $S_i's$ corresponding to the intervals that contain a jump, their absolute values are all larger than the threshold $\gamma_m L_m + C_m$. On the right-side of (3.13), the main contribution by the jump is the second term, and $\widehat{S}_i$ would be negligible compared to the threshold for large m. Thus the probability that $|S_i|$ is larger than the threshold approximately equals to the probability that the second term is larger than the threshold $L_m\gamma_m + C_m$.

From (3.17) we can see that when $G_{i\,m} \to 0$ as $m \to \infty$, the probability of $GFTD$ converges to 0. Same as $D_m(\tau)$ in Theorem 3.2, $D_{i\,m}$ shows how the position of the jump affects its contribution to $|S_i|$. From Corollary 3.1 we can see that the order of $D_{i\,m}$ is approximately the same as K_1. Thus (3.19) can be obtained by setting the value of γ_m as, for example, any constant or m^ϵ for any $0 < \epsilon < \frac{1}{2} - \frac{1}{2}\alpha_1$.

By definition GSD is the same as the event that $\max_{i\in B_m}|S_i|$ is larger than the threshold $L_m\gamma_m + C_m$. Then from the asymptotic distribution of $\max_{i\in B_m}|S_i|$, we can obtain the limiting probability of GSD as follows.

THEOREM 3.4. *As* $m \to \infty$,

$$\mathrm{P}(GSD) \to 1 - \exp\left(-e^{-\gamma_m}\right) = \alpha_m. \tag{3.20}$$

Theorems 3.3 and 3.4 suggest that if $\lim_{m\to\infty} \gamma_m = \infty$, and there exists $\epsilon > 0$ such that $\lim_{m\to\infty} \frac{\gamma_m}{\sqrt{\log(m)}} m^{-\frac{1}{2}+\frac{1}{2}\alpha_1+\epsilon} = 0$, then both probabilities of $GFTD$ and GSD converge to zero as $m \to \infty$. For example, we can set $\gamma_m = m^\alpha$ for any $0 < \alpha < \frac{1}{2} - \frac{1}{2}\alpha_1$.

If we find there is a jump on $(t_{2(i-1)K_1}, t_{2iK_1}]$, the jump size can be estimated by taking the average of the returns over the intervals that contain $(t_{2(i-1)K_1}, t_{2iK_1}]$. For example, there are $2K_1 + 1$ intervals in the form of $(t_j, t_{j+4K_1}]$ that contain $(t_{2(i-1)K_1}, t_{2iK_1}]$, which are $(t_{2iK_1-4K_1}, t_{2iK_1}], \ldots, (t_{2(i-1)K_1}, t_{2iK_1+2K_1}]$ (see Figure 3.3). Then we can consistently estimate the jump size by taking the average of the $2K_1+1$ returns over these intervals. It can be checked that the contribution to the average return from the diffusion process is $O_{\mathrm{P}}\left(m^{\frac{1}{2}(\alpha_1-1)}\right)$, and that from the noise is $O_{\mathrm{P}}\left(m^{-\frac{1}{2}\alpha_1}\right)$. Thus for large m, the effect from the noise and the diffusion will vanish, and the value of the average return will be approximately equal to the jump size.

3.4 Monte Carlo Simulation

In this section we perform Monte Carlo simulations to explore the efficiency of our jump test, and we compare it with the tests by Lee and Mykland (2008) and Jiang and Oomen (2008). We assume there are 8 hours in a trading day. Thus for example, if the sampling frequency is 1 second, we have $\triangle t = t_i - t_{i-1} = \frac{1}{3600\times 8}$. We assume that there is one price jump in the entire day, and its occurence time is uniformly distributed on $[0, 1]$. Moreover, we set the drift term $a_t = 0$ and the standard deviation of the noise as 0.1, unless stated otherwise. We also set $K_1 = m^{\frac{7}{12}}, K_2 = m^{\frac{5}{6}}, K_3 = 0.25m^{\frac{2}{3}}, K_4 = 2m^{\frac{1}{12}}$, which is in line with our theory to make the test consistent. We use the Euler scheme to approximate the continuous-time process y_t which is assumed to satisfy (3.1), with the discretization being 0.0625 seconds. We assume two conditions for the spot volatility σ_t. In the first setting, σ_t is a constant whose value is 0.6. In the second condition, σ_t follows the CIR process:

$$d\sigma_t = \kappa(\upsilon - \sigma_t)dt + \theta\sqrt{\sigma_t}dW_t, \tag{3.21}$$

where the mean reversion parameter $\kappa = 0.1$, the long run mean of the process $\upsilon = 0.6$, and the volatility $\theta = 0.3$. The sampling frequency we consider ranges from 0.125 second to 1 second, and for each sampling frequency, we simulate 1000 different series of price observations. Table 3.1 shows the probability of spurious detection for a generic interval $(t_{2(i-1)K_1}, t_{2iK_1}]$. The significance level is set as $\alpha = 5\%$. From Table 3.1 we can see that more frequent observations can reduce the probability of spurious detection.

Table 3.2 shows the probability of correctly detecting a jump on $(t_{2(i-1)K_1}, t_{2iK_1}]$ with respect to different jump sizes and sampling frequencies, with the significance level $\alpha = 5\%$. We set the jump size relative to the spot volatility σ_t, since in the stochastic volatility condition, υ is the long-run mean of σ_t. From Table 3.2 we can see that it would be easier to detect large jumps than small ones, and sampling more frequently improves the chance of successful detection.

We define the event global misclassification (GM) such that it happens if and only if either GSD or $GFTD$ occurs. Table 3.3 shows the probability of GM under the condition where $\sigma_t = 0.6$. Like the results of Table 3.2, from Table 3.3 we can also see that more frequent sampling makes the test more efficient, and larger jumps are easier to detect, which reduces the probability of GM by decreasing the likelihood of $GFTD$.

We also compare the probabilities of $GFTD$, GSD and GM for our test and the tests by Lee and Mykland (2008) (LM test) and Jiang and Oomen (2008) (JO test). Now let us introduce LM and JO tests. The statistic proposed by Lee and Mykland (2008) to detect whether there is a jump on $(t_{i-1}, t_i]$ is defined as follows:

$$L_i = \frac{x_i - x_{i-1}}{\widehat{\sigma}_i}, \tag{3.22}$$

where

$$\widehat{\sigma}_i^2 = \frac{1}{K_5 - 2} \sum_{j=i-K+2}^{i-1} (x_j - x_{j-1})(x_{j-1} - x_{j-2}). \tag{3.23}$$

Lee and Mykland (2008) set $K_5 = m^\delta$ and δ is a constant on $(0, \frac{1}{2})$. They do not show how to choose the optimal value of δ. We have tried many values on δ, and we will show the results when $\delta = \frac{1}{2}$, because the results for other values of δ are not better. Lee and Mykland (2008) justify the test in the absence of the noise. In this case, under the null hypothesis that there is

no jump on $(t_{i-1}, t_i]$,

$$\sqrt{m}(x_i - x_{i-1}) \approx \sigma_{t_i} U_i, \tag{3.24}$$

where U_i is a standard normal random variable generated by the Brownian motion on $(t_{i-1}, t_i]$. Moreover, in the noiseless setting, $\sqrt{m}\widehat{\sigma}_i$ is approximately equal to the local volatility σ_{t_i}. Thus without the impact of the noise and jump, L_i follows asymptotically a standard normal distribution and they are independent of each other. Still in the noiseless setting, a jump on $(t_{i-1}, t_i]$ can significantly increase the value of $|L_i|$. Thus the principle that Lee and Mykland (2008) adopt to determine the threshold based on the significance level is similar to ours. However, neither the nominator nor the denominator on the right-side of (3.22) are immune to the noise, so the efficiency of LM test decreases with the presence of noise.

The statistic proposed by Jiang and Oomen (2008) to detect jumps on $(0, 1]$ is as follows:

$$S_m = \frac{SwV_m - RV_m}{\sqrt{\widehat{\Omega}}}, \tag{3.25}$$

where $SwV_m = 2\sum_{i=1}^{m} \left(e^{x_i - x_{i-1}} - 1 - (x_i - x_{i-1})\right)$, $RV_m = \sum_{i=1}^{m} (x_i - x_{i-1})^2$, and $\widehat{\Omega}$ is some approximation of

$$\Omega = 4m\eta^6 + 12\eta^4 \int_0^1 \sigma_t^2 dt + 8\eta^2 \frac{1}{m} \int_0^1 \sigma_t^4 dt + \frac{5}{3}\frac{1}{m^2} \int_0^1 \sigma_t^6 dt.$$

We use the procedure described by Jiang and Oomen (2008) to construct $\widehat{\Omega}$. Under the null hypothesis that there is no jump on $(0, 1]$, S_m has approximately zero expectation and variance 1. In order to obtain acceptable results from the test, as shown in Jiang and Oomen (2008), we need to impose several conditions. For example, σ_t is a constant, $\eta^2 \ll \sigma_t^2$ and $m\sigma_t^4$ is small. Here recall that η is the standard deviation of the noise. The first two conditions are basically true in our simulations, since we define $\sigma_t = 0.6$ and $\eta^2 = 0.01$ or 0.0144, much smaller than $\sigma_t^2 = 0.36$. The last condition implies that m cannot be too large, so it may not be proper to sample frequently for JO test with noisy data, and we will see that the simulation results also illustrate it. Thus we sample much less frequenly in performing JO test than the other two tests. Moreover, unlike our and LM tests, for a fixed period $(0, T]$, JO test can only indicate whether there is a jump on that period, without further indication on the jump arrival time, like locating the jump arrival within a small interval. Thus in order to make the results comparable, we use

the JO test to detect jump respectively on $\left(0, \frac{1}{2}\right]$ and $\left(\frac{1}{2}, 1\right]$ in one simulation.

For all the tests, the significance level is set as 5%, and the jump size is subject to normal distribution with mean 0 and standard deviation 1.2. Table 3.4 reports the results. We can see that across different levels of sampling frequency and η, our test is less likely to commit misspecifications than LM and JO test, since the probability of GM is uniformly lower for our test. Further investigation reveals that this is because the probability of $GFTD$ of our test is always smaller than LM and JO tests. It is obvious that relative to our and LM tests, we use much less frequent price observations to construct the JO test. This is because for JD test, when m is large, the absolute value of the statistic will be significantly increased by the noise, so it is likely to generate a misleading result when there is no jump on the interval. For example, when the sampling frequency is 10 seconds, over 90 percent of the JO tests will commit a spurious detection. Unreported results suggest that for JO test, the optimal sampling frequency is around 120 seconds. In addition, when $\eta = 0.12$, more frequent sampling does not reduce the likelihood of misspecification for LM test. This implies that the consistency of LM test, which is established in the noiseless setting, is not immune to the noise. However, sampling more frequently can always improve the efficiency of our test.

Table 3.5 displays the values of the mean squared errors (MSE) on estimating the jump size by considering the local average return as the estimator. The procedure to construct the estimator is the same as we have explained in the last paragraph of Section 3. That is, for any jump on $(0, 1]$, it is contained in an interval in the form of $(t_{2(i-1)K_1}, t_{2iK_1}]$ for some $i \in \left\{1, \ldots, \frac{m}{2K_1}\right\}$. Then we take the average of the returns over the intervals $(t_{(2i-4)K_1}, t_{2iK_1}]$, $(t_{(2i-4)K_1+1}, t_{2iK_1+1}]$, $\ldots$, $(t_{(2i-2)K_1}, t_{(2i+2)K_1}]$ as the estimator on the jump size. We consider 4 combinations in terms of the spot volatility σ_t and jump size Y_i. σ_t can be either a constant 0.6 or follows the stochastic process as specified in (3.21). Y_i can be either a constant 1.2 or subject to the distribution N(0,1.44). From Table 3.5 we can see that sampling more frequently increases the precision of estimating the jump size.

3.5 Conclusions

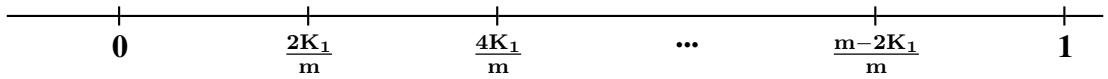
In this chapter we propose a novel test on detecting jumps in the process of the efficient asset price. This test is based on high-frequency price data that are affected by market microstructure noise. We perform multiple tests over the whole period, and each single test checks whether

there is a jump in a small interval. The statistic is the rescaled local average return standardized by an estimator of spot volatility. The volatility estimator is derived from the truncated two-scales realized variance (TTSRV), because the TTSRV consistently estimates the integrated volatility in the presence of noise and jumps. Since a jump can greatly imcrease the absolute value of the statistic, we will reject the null hypothesis that there is no jump in the interval, if the corresponding statistic is large in its absolute value. We show that when the threshold is defined properly, the probability of testing errors will converge to zero as the number of price observations increases. Simulations also illustrate the consistency of our test, and show that our procedure performs well. Simulations also indicate that our test performs well compared to some other prevalent jump tests which are not immune to the noise.

3.6 Appendix

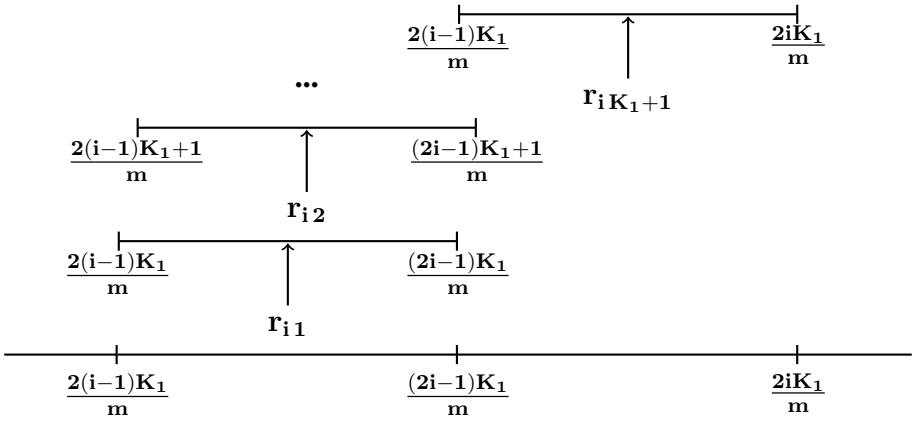
3.6.1 Figures and Tables

Figure 3.1: DIVISION OF THE WHOLE PERIOD



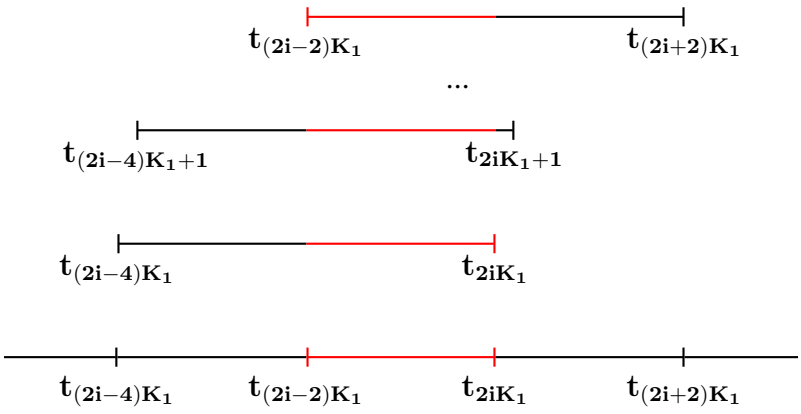
The figure shows how we divide the whole horizon $(0, 1]$ into $\frac{m}{2K_1}$ intervals. Then test can be performed on each interval to detect jumps.

Figure 3.2: CONSTRUCTION OF THE NOMINATOR



The figure provides a schematic representation of how we compute the nominator of S_i

Figure 3.3: THE OVERLAPPING INTERVALS



The figure shows there are $2K_1+1$ intervals in the form of $(t_j, t_{j+4K_1}]$ that contain the interval $(t_{2i-2}K_1, t_{2iK_1}]$ in red color.

Table 3.1: PROBABILITY OF SPURIOUS DETECTION

$freq$	$\sigma_t = 0.6$	(SE)	SV	(SE)
1-second	3.89×10^{-3}	(3.54×10^{-4})	4.44×10^{-3}	(3.89×10^{-4})
0.5-second	3.13×10^{-3}	(2.71×10^{-4})	3.08×10^{-3}	(2.36×10^{-4})
0.25-second	2.19×10^{-3}	(1.99×10^{-4})	1.56×10^{-3}	(1.65×10^{-4})
0.125-second	1.53×10^{-3}	(1.36×10^{-4})	1.41×10^{-3}	(1.26×10^{-4})

The table shows means and standard errors (in parentheses) of probability of spurious detection $\mathrm{P}(SD_i)$. The significance level α is 5%. SV means the spot volatility follows the process of (3.21), and $freq$ denotes the sampling frequency.

Table 3.2: PROBABILITY OF DETECTING ACTUAL JUMP

Jump Size	$2\sigma_t$	σ_t	$0.5\sigma_t$
		constant volatility σ_t at 0.6	
$freq$=1-second	0.78	0.67	0.34
	(1.32×10^{-2})	(1.49×10^{-2})	(1.51×10^{-2})
$freq$=0.5-second	0.85	0.72	0.48
	(1.13×10^{-2})	(1.43×10^{-2})	(1.59×10^{-2})
$freq$=0.25-second	0.89	0.75	0.55
	(9.94×10^{-3})	(1.38×10^{-2})	(1.58×10^{-2})
$freq$=0.125-second	0.94	0.77	0.64
	(7.55×10^{-3})	(1.34×10^{-2})	(1.53×10^{-2})
Jump Size	2υ	υ	0.5υ
		stochastic volatility	
$freq$=1-second	0.81	0.64	0.19
	(1.25×10^{-2})	(1.53×10^{-2})	(1.25×10^{-2})
$freq$=0.5-second	0.79	0.66	0.30
	(1.29×10^{-2})	(1.51×10^{-2})	(1.46×10^{-2})
$freq$=0.25-second	0.82	0.70	0.36
	(1.22×10^{-2})	(1.46×10^{-2})	(1.53×10^{-2})
$freq$=0.125-second	0.86	0.71	0.55
	(1.10×10^{-2})	(1.44×10^{-2})	(1.58×10^{-2})

This table shows means and standard errors (in parentheses) of probability of detecting actual jump $(1 - \mathrm{P}(FTD_i))$. The significance level α is 5%. Stochastic volatility means the spot volatility follows the process of (3.21), and $freq$ denotes the sampling frequency.

Table 3.3: PROBABILITY OF GLOBAL MISCLASSIFICATION

Jump size	$2\sigma_t$	σ_t	$0.5\sigma_t$
$freq$=1-second	0.32	0.40	0.70
	(1.48×10^{-2})	(1.56×10^{-2})	(1.46×10^{-2})
$freq$=0.5-second	0.26	0.33	0.61
	(1.39×10^{-2})	(1.49×10^{-2})	(1.55×10^{-2})
$freq$=0.25-second	0.20	0.34	0.50
	(1.27×10^{-2})	(1.51×10^{-2})	(1.59×10^{-2})
$freq$=0.125-second	0.17	0.30	0.43
	(1.19×10^{-2})	(1.46×10^{-2})	(1.57×10^{-2})

This table shows means and standard errors (in parentheses) of probability of global misclassifications by either global spurious detection of jumps (GSD) or global failure to detect actual jumps($GFTD$). The significance level α is 5% and $\sigma_t = 0.6$. $freq$ denotes the sampling frequency.

Table 3.4: COMPARISON OF THE MISCLASSIFICATION PROBABILITIES

η=0.1			Our test			
$freq$	P(GM)	(SE)	P($GFTD$)	(SE)	P(GSD)	(SE)
0.5-second	0.38	(1.54×10^{-2})	0.30	(1.46×10^{-2})	0.15	(1.13×10^{-2})
0.25-second	0.36	(1.53×10^{-2})	0.28	(1.43×10^{-2})	0.15	(1.13×10^{-2})
0.125-second	0.27	(1.41×10^{-2})	0.20	(1.27×10^{-2})	0.10	(9.53×10^{-3})
η=0.1			LM test			
$freq$	P(GM)	(SE)	P($GFTD$)	(SE)	P(GSD)	(SE)
0.5-second	0.52	(1.59×10^{-2})	0.52	(1.59×10^{-2})	0.013	(3.16×10^{-3})
0.25-second	0.40	(1.56×10^{-2})	0.39	(1.55×10^{-2})	0.028	(5.42×10^{-3})
0.125-second	0.39	(1.55×10^{-2})	0.38	(1.54×10^{-2})	0.011	(3.16×10^{-3})
η=0.1			JO test			
$freq$	P(GM)	(SE)	P($GFTD$)	(SE)	P(GSD)	(SE)
120-second	0.41	(1.56×10^{-2})	0.34	(1.50×10^{-2})	0.11	(9.89×10^{-3})
60-second	0.43	(1.57×10^{-2})	0.35	(1.51×10^{-2})	0.17	(1.19×10^{-2})
10-second	0.95	(6.89×10^{-3})	0.15	(1.13×10^{-2})	0.93	(8.07×10^{-3})
η=0.12			Our test			
$freq$	P(GM)	(SE)	P($GFTD$)	(SE)	P(GSD)	(SE)
0.5-second	0.49	(1.59×10^{-2})	0.37	(1.53×10^{-2})	0.14	(1.10×10^{-2})
0.25-second	0.45	(1.58×10^{-2})	0.29	(1.44×10^{-2})	0.19	(1.25×10^{-2})
0.125-second	0.41	(1.56×10^{-2})	0.28	(1.43×10^{-2})	0.17	(1.19×10^{-2})
η=0.12			LM test			
$freq$	P(GM)	(SE)	P($GFTD$)	(SE)	P(GSD)	(SE)
0.5-second	0.53	(1.59×10^{-2})	0.51	(1.59×10^{-2})	0.032	(5.42×10^{-3})
0.25-second	0.52	(1.59×10^{-2})	0.52	(1.59×10^{-2})	0.014	(3.16×10^{-3})
0.125-second	0.58	(1.57×10^{-2})	0.56	(1.58×10^{-2})	0.023	(4.45×10^{-3})
η=0.12			JO test			
$freq$	P(GM)	(SE)	P($GFTD$)	(SE)	P(GSD)	(SE)
120-second	0.51	(1.58×10^{-2})	0.42	(1.56×10^{-2})	0.16	(1.16×10^{-2})
60-second	0.54	(1.58×10^{-2})	0.42	(1.56×10^{-2})	0.25	(1.37×10^{-2})
10-second	0.99	(3.15×10^{-3})	0.19	(1.24×10^{-2})	0.98	(4.43×10^{-3})

This table shows means and standard errors of probabilities of $GFTD$, GSD and GM for our test and the tests by Lee and Mykland (2008) (LM test) and Jiang and Oomen (2008) (JO test). The significance level α is 5% and $\sigma_t = 0.6$. The volatility of the jump size is set as 1.2. $freq$ denotes the sampling frequency, and η is the standard deviation of the noise.

Table 3.5: MSE OF ESTIMATING THE JUMP SIZE

$freq$	$Y = 1.2, \sigma_t = 0.6$	$Y = 1.2$, SV	$Y \sim N(0, 1.2), \sigma_t = 0.6$	$Y \sim N(0, 1.2)$, SV
1-second	1.46×10^{-2}	2.53×10^{-2}	1.68×10^{-2}	2.40×10^{-2}
0.5-second	1.34×10^{-2}	2.09×10^{-2}	1.42×10^{-2}	1.72×10^{-2}
0.25-second	8.16×10^{-3}	1.94×10^{-2}	9.40×10^{-3}	1.32×10^{-2}
0.125-second	7.38×10^{-3}	1.11×10^{-2}	7.16×10^{-3}	1.00×10^{-2}

The table shows the MSE values of estimating the jump size for different conditions on the spot volatility σ_t and the jump size Y. SV means σ_t follows the process of (3.21), and $freq$ denotes the sampling frequency.

3.6.2 Proofs

Proof of Lemma 3.1. For simplicity we assume $a_t = 0$. We define

$$\overline{x}_i = \int_0^{t_i} \sigma_s dB_s + u_{t_i},$$

$$\overline{\sigma}^2_{\mathsf{TTS}\,i} = \frac{1}{K_3} \sum_{j=2(i-1)K_1-K_2+K_3}^{2(i-1)K_1} (\overline{x}_j - \overline{x}_{j-K_3})^2 \mathbf{1}_{E_j} - \frac{1}{K_3} \sum_{j=2(i-1)K_1-K_2+K_3}^{2(i-1)K_1} (\overline{x}_j - \overline{x}_{j-1})^2 \mathbf{1}_{E_j},$$

and

$$\overline{\sigma}^2_{\mathsf{TS}\,i} = \frac{1}{K_3} \sum_{j=2(i-1)K_1-K_2+K_3}^{2(i-1)K_1} (\overline{x}_j - \overline{x}_{j-K_3})^2 - \frac{1}{K_3} \sum_{j=2(i-1)K_1-K_2+K_3}^{2(i-1)K_1} (\overline{x}_j - \overline{x}_{j-1})^2.$$

Following similar steps as those justifying equations (2.11) and (2.12), it can be checked that a.s.,

$$\widehat{\sigma}^2_{\mathsf{TTS}\,i} = \overline{\sigma}^2_{\mathsf{TTS}\,i}, \tag{3.26}$$

for all $i \in \left\{1, \ldots, \frac{m}{2K_1}\right\}$, as $m \to \infty$, and for any $\epsilon > 0$,

$$\sup_{i \in \left\{1, \ldots, \frac{m}{2K_1}\right\}} \left|\overline{\sigma}^2_{\mathsf{TTS}\,i} - \overline{\sigma}^2_{\mathsf{TS}\,i}\right| = O_{\mathrm{P}}\left(m^{-\frac{1}{3}+\alpha_4+\epsilon}\right). \tag{3.27}$$

Therefore,

$$\sup_{i \in \left\{1, \ldots, \frac{m}{2K_1}\right\}} \left|\widehat{\sigma}^2_{\mathsf{TTS}\,i} - \overline{\sigma}^2_{\mathsf{TS}\,i}\right| = O_{\mathrm{P}}\left(m^{-\frac{1}{3}+\alpha_4+\epsilon}\right). \tag{3.28}$$

$\overline{\sigma}^2_{\mathsf{TS}\,i}$ can be seen as a two-scales realized volatility estimator on $[t_{2(i-1)K_1-K_2}, t_{2(i-1)K_1}]$. Based on the proof of Theorem 1 in Fan et al. (2012), we have that there exists positive constants c and c' such that for all $x \in [0, c]$ and large m,

$$\mathrm{P}\left(\left|\overline{\sigma}^2_{\mathsf{TS}\,i} - \int_{t_{2(i-1)K_1-K_2}}^{t_{2(i-1)K_1}} \sigma_s^2 ds\right| \geq x\right) \leq c' \exp\left(-cx^2 m^{\min(2-\alpha_2-\alpha_3, 2\alpha_3-\alpha_2)}\right), \tag{3.29}$$

from which we can see that for any $\epsilon > 0$ we have

$$\sup_{i \in \left\{1,\ldots,\frac{m}{2K_1}+1\right\}} \left|\overline{\sigma}^2_{\mathsf{TS}\,i} - \int_{t_{2(i-1)K_1-K_2}}^{t_{2(i-1)K_1}} \sigma_s^2 ds\right| = O_{\mathrm{P}}\left(m^{\max\left(\frac{1}{2}(\alpha_2+\alpha_3)-1, \frac{1}{2}\alpha_2-\alpha_3\right)+\epsilon}\right). \tag{3.30}$$

From Assumption 3.2 for all $i \in \left\{1, \ldots, \frac{m}{2K_1}\right\}$ we have

$$\begin{aligned}
\int_{t_{2(i-1)K_1-K_2}}^{t_{2(i-1)K_1}} \sigma_s^2 ds &= \sigma^2_{t_{2(i-1)K_1}}(t_{2(i-1)K_1} - t_{2(i-1)K_1-K_2}) + \int_{t_{2(i-1)K_1-K_2}}^{t_{2(i-1)K_1}} (\sigma_s^2 - \sigma_{t^2_{2(i-1)K_1}}) ds \\
&= \frac{K_2}{m}\sigma^2_{t_{2(i-1)K_1}} + O_{\mathrm{P}}\left(m^{\frac{3}{2}(\alpha_2-1)+\epsilon}\right),
\end{aligned} \tag{3.31}$$

for $\epsilon > 0$. Therefore,

$$\sup_{i \in \left\{1,\ldots,\frac{m}{2K_1}\right\}} \left|\frac{m}{K_2}\int_{t_{2(i-1)K_1-K_2}}^{t_{2(i-1)K_1}} \sigma_s^2 ds - \sigma^2_{t_{2(i-1)K_1}}\right| = O_{\mathrm{P}}\left(m^{\frac{1}{2}(\alpha_2-1)+\epsilon}\right). \tag{3.32}$$

From (3.27) (3.30) (3.32), we get that

$$\sup_{i \in \left\{1,\ldots,\frac{m}{2K_1}\right\}} \left|\frac{m}{K_2}\widehat{\sigma}^2_{\mathsf{TS}\,i} - \sigma^2_{t_{2(i-1)K_1}}\right| = O_{\mathrm{P}}\left(m^{\max\left(\frac{2}{3}+\alpha_4-\alpha_2, \frac{1}{2}\alpha_3-\frac{1}{2}\alpha_2, -\frac{1}{2}\alpha_2-\alpha_3+1, \frac{1}{2}(\alpha_1-1)\right)+\epsilon}\right) \tag{3.33}$$

for $\epsilon > 0$. □

Proof of Theorem 3.1. For simplicity we assume $a_t = 0$. Define

$$A_{i1} = \sum_{j=0}^{K_1} \int_{t_{2(i-1)K_1+j}}^{t_{(2i-1)K_1+j}} \sigma_t dB_t, \tag{3.34}$$

and

$$A_{i2} = \sum_{j=0}^{K_1} (u_{(2i-1)K_1+j} - u_{2(i-1)K_1+j}). \tag{3.35}$$

Then when $i \in B_m$,

$$A_i = \sqrt{\frac{3m}{K_1(K_1+1)(2K_1+1)}}(A_{i1} + A_{i2}).$$

For A_{i1}, we have

$$A_{i1} = \sum_{j=2(i-1)K_1}^{(2i-1)K_1} \int_{t_j}^{t_{j+K_1}} (\sigma_t - \sigma_{t_{2(i-1)K_1}}) dB_t + \sigma_{t_{2(i-1)K_1}} \sum_{j=2(i-1)K_1}^{(2i-1)K_1} (B_{t_{j+K_1}} - B_{t_j}).$$

By Assumption 3.2, across $i \in \left\{1, \ldots, \frac{m}{2K_1}\right\}$, for any $\epsilon > 0$ we have

$$\sup_i \sup_{t \in [t_{2(i-1)K_1}, t_{2iK_1}]} (\sigma_t - \sigma_{t_{2(i-1)K_1}}) = O_{\mathrm{P}}\left(m^{\frac{1}{2}(\alpha_1 - 1)+\epsilon}\right).$$

Then it can be checked that

$$\sup_i \sum_{j=2(i-1)K_1}^{(2i-1)K_1} \int_{t_j}^{t_{j+K_1}} (\sigma_t - \sigma_{t_{2(i-1)K_1}}) dB_t = O_{\mathrm{P}}\left(m^{2\alpha_1 - 1+\epsilon}\right), \tag{3.36}$$

for $\epsilon > 0$. For A_{i2}, we have $\sup_i |A_{i2}| = O_{\mathrm{P}}\left(m^{\frac{1}{2}\alpha_1+\epsilon}\right)$. Therefore,

$$\sup_i \left| M \left(A_{i1} + A_{i2} - \sigma_{t_{2(i-1)K_1}} \sum_{j=2(i-1)K_1}^{(2i-1)K_1} (B_{t_{j+K_1}} - B_{t_j}) \right) \right|$$

is $O_{\mathrm{P}}\left(m^{\max(1-2\alpha_1, \frac{1}{2}-\alpha_1)+\epsilon}\right)$, as $M = \sqrt{\frac{3m}{K_1(K_1+1)(2K_1+1)}}$. Then we have

$$\sup_i \left| S_i \sqrt{\frac{m}{K_2} \widehat{\sigma}^2_{\mathsf{TTS}\, i}} - \sigma_{t_{2(i-1)K_1}} \widehat{S}_i \right| = O_{\mathrm{P}}\left(m^{\max(1-2\alpha_1, \frac{1}{2}-\alpha_1)+\epsilon}\right), \tag{3.37}$$

for any $\epsilon > 0$. Since we have set $\alpha_2 = \frac{5}{6}, \alpha_3 = \frac{2}{3}$, based on Lemma 3.1, it can be obtained that

$$\sup_i \left| \sqrt{\frac{m}{K_2} \widehat{\sigma}^2_{\mathsf{TTS}\, i}} - \sigma_{t_{2(i-1)K_1}} \right| = O_{\mathrm{P}}\left(m^{-\frac{1}{12}+\epsilon}\right), \tag{3.38}$$

for any $\epsilon > 0$. Then we have

$$\sup_i \left| \left(\sigma_{t_{2(i-1)K_1}} + \epsilon_i \right) S_i - \sigma_{t_{2(i-1)K_1}} \widehat{S}_i \right| = O_{\mathrm{P}} \left(m^{\max(1-2\alpha_1, \frac{1}{2}-\alpha_1)+\epsilon} \right), \tag{3.39}$$

where $\sup_i |\epsilon_i|$ is $O_{\mathrm{P}} \left(m^{-\frac{1}{12}+\epsilon} \right)$. As we will prove later, $\widehat{S}_i$ is subject to the standard normal distribution. Then $\sup_i |\widehat{S}_i|$ is $O_{\mathrm{P}}(\log m)$. Thus given that $\sigma_{t_{2(i-1)K_1}} > \sigma^-$, according to (3.39), we can see that $\sup_i |S_i|$ is $O_{\mathrm{P}} \left(\log m + m^{\max(1-2\alpha_1, \frac{1}{2}-\alpha_1)+\epsilon} \right)$, because

$$\sup_i |S_i| \leq \sup_i \left| \frac{\sigma_{t_{2(i-1)K_1}}}{\sigma_{t_{2(i-1)K_1}} + \epsilon_i} \widehat{S}_i \right| + O_{\mathrm{P}} \left(m^{\max\left(\frac{1}{2}-\alpha_1, \frac{1}{2}\alpha_1 - \frac{1}{2},\right)+\epsilon} \right). \tag{3.40}$$

Also from (3.39), we have that

$$\sup_i \left| \sigma_{t_{2(i-1)K_1}} S_i - \sigma_{t_{2(i-1)K_1}} \widehat{S}_i \right| \leq O_{\mathrm{P}} \left(m^{\max(1-2\alpha_1, \frac{1}{2}-\alpha_1)+\epsilon} \right) + \sup_i |\epsilon_i S_i|. \tag{3.41}$$

As $\sup_i |\epsilon_i S_i|$ is $O_{\mathrm{P}} \left(m^{\max\left(\frac{1}{2}-\alpha_1, \frac{1}{2}\alpha_1 - \frac{1}{2}, -\frac{1}{12}\right)+\epsilon} \right) = O_{\mathrm{P}} \left(m^{\beta+\epsilon} \right)$ and $\sigma_{t_{2(i-1)K_1}} > \sigma^-$, proof for (3.11) is completed. Now we are going to show that $\widehat{S}_i = M \sum_{j=2(i-1)K_1}^{(2i-1)K_1} (B_{t_{j+K_1}} - B_{t_j})$ is subject to the standard normal distribution.

$$\begin{aligned} \sum_{j=2(i-1)K_1}^{(2i-1)K_1} (B_{t_{j+K_1}} - B_{t_j}) &= \sum_{j=0}^{K_1-1} (K_1 - j)(B_{t_{(2i-1)K_1-j}} - B_{t_{(2i-1)K_1-j-1}}) \\ &+ \sum_{j=0}^{K_1-1} (K_1 - j)(B_{t_{(2i-1)K_1+j+1}} - B_{t_{(2i-1)K_1+j}}). \end{aligned} \tag{3.42}$$

As $1^2 + 2^2 + \cdots + K_1^2 = \frac{K_1(K_1+1)(2K_1+1)}{6}$,

$$\sum_{j=2(i-1)K_1}^{(2i-1)K_1} (B_{t_{j+K_1}} - B_{t_j}) \sim N \left(0, \frac{K_1(K_1+1)(2K_1+1)}{3m} \right). \tag{3.43}$$

As $\frac{K_1(K_1+1)(2K_1+1)}{3m} = \frac{1}{M^2}$, $\widehat{S}_i \sim N(0,1)$.

□

Proof of Theorem 3.2. For simplicity we assume $a_t = 0$. Recall that we have defined $\overline{x}_i = \int_0^{t_i} \sigma_s dB_s + u_{t_i}$. It can be checked that

$$S_i = \frac{M \sum_{j=0}^{K_1} \left(\overline{x}_{(2i-1)K_1+j} - \overline{x}_{(2i-2)K_1+j} \right)}{\sqrt{\frac{m}{K_2} \widehat{\sigma}^2_{\mathsf{TTS}\, i}}} + \frac{M D_m(\tau) Y_\tau}{\sqrt{\frac{m}{K_2} \widehat{\sigma}^2_{\mathsf{TTS}\, i}}}. \tag{3.44}$$

From the proof of Theorem 3.1, we can see that

$$\frac{M \sum_{j=0}^{K_1} \left(\overline{x}_{(2i-1)K_1+j} - \overline{x}_{(2i-2)K_1+j} \right)}{\sqrt{\frac{m}{K_2} \widehat{\sigma}^2_{\mathsf{TTS}\, i}}} = \widehat{S}_i + o_{\mathrm{P}}\left(m^{\max\left(\frac{1}{2}-\alpha_1, \frac{1}{2}\alpha_1 - \frac{1}{2}, -\frac{1}{12}\right)+\epsilon} \right), \tag{3.45}$$

where $\widehat{S}_i$ is standard normal. From Lemma 3.1, we have

$$\sqrt{\frac{m}{K_2} \widehat{\sigma}^2_{\mathsf{TTS}\, i}} = \sigma_{t_{2(i-1)K_1}} + o_{\mathrm{P}}(1). \tag{3.46}$$

Therefore,

$$\frac{M D_m(\tau) Y_\tau}{\sqrt{\frac{m}{K_2} \widehat{\sigma}^2_{\mathsf{TTS}\, i}}} = \frac{M D_m(\tau) Y_\tau}{\sigma_{t_{2(i-1)K_1}}} + o_{\mathrm{P}}\left(\frac{M D_m(\tau)}{\sigma_{t_{2(i-1)K_1}}} \right). \tag{3.47}$$

Now the proof is completed given (3.44) (3.45) (3.47). □

Proof of Corollary 3.1. When $D_m(\tau) \leq K_1^{1-\epsilon}$, we have

$$\tau \in \bigcup_{j=0}^{\frac{m}{2K_1}-1} \left(t_{2K_1 j}, t_{2K_1 j + K_1^{1-\epsilon}} \right) \text{ or } \tau \in \bigcup_{j=0}^{\frac{m}{2K_1}-1} \left(t_{2K_1(j+1)-K_1^{1-\epsilon}}, t_{2K_1(j+1)} \right).$$

Therefore,

$$\begin{aligned}
& \mathrm{P}\left(D_m(\tau) \leq K_1^{1-\epsilon} \right) \\
\leq & \sum_{j=0}^{\frac{m}{2K_1}-1} \mathrm{P}\left(\tau \in (t_{2K_1 j}, t_{2K_1 j + K_1^{1-\epsilon}}) \right) + \sum_{j=0}^{\frac{m}{2K_1}-1} \mathrm{P}\left(\tau \in (t_{2K_1(j+1)-K_1^{1-\epsilon}}, t_{2K_1(j+1)}) \right) \\
\leq & \sum_{j=0}^{\frac{m}{2K_1}-1} \mathrm{P}\left(N_{t_{2K_1 j + K_1^{1-\epsilon}}} - N_{t_{2K_1 j}} > 0 \right) + \sum_{j=0}^{\frac{m}{2K_1}-1} \mathrm{P}\left(N_{t_{2K_1(j+1)}} - N_{t_{2K_1(j+1)-K_1^{1-\epsilon}}} > 0 \right) \\
= & O\left(\frac{m}{K_1} \cdot \frac{K_1^{1-\epsilon}}{m} \right) = O\left(K_1^{-\epsilon} \right),
\end{aligned} \tag{3.48}$$

where the second to the last equation is obtained from (3.14). Thus for any $\epsilon \in (0,1)$ it is a.s. that $D_m(\tau) > K_1^{1-\epsilon}$ for large m and any $\epsilon > 0$, so we have

$$\frac{D_m(\tau)}{K_1^{1-\epsilon}} > \frac{K_1^{1-\frac{\epsilon}{2}}}{K_1^{1-\epsilon}} = K_1^{\frac{\epsilon}{2}}, \tag{3.49}$$

for large m. Thus the proof is completed given that $K_1^{\frac{\epsilon}{2}} \to \infty$ as $m \to \infty$. □

Proof of Lemma 3.2. From Theorem 3.1 we can see that as $m \to \infty$, the limiting distribution of $\max_{i \in \overline{B}_m} |S_i|$ is the same as the distribution of the maximum absolute value of $\frac{m}{2K_1} - N_1$ standard normal variables. Lemma 1 in Lee and Mykland (2008) shows the distribution of the maximum absolute value of $m - N_1$ standard normal variables, as $m \to \infty$. Thus given that $\frac{m}{2K_1} \to \infty$ as $m \to \infty$, Lemma 1 in Lee and Mykland (2008) directly leads to the results of Lemma 3.2. □

Proof of Theorem 3.3. For simplicity we assume $a_t = 0$. For a generic i, consider the ith jump which happens at time τ_i with size Y_i. Without loss of generality, we assume $\tau_i \notin \{t_{2K_1}, t_{4K_1}, \ldots, t_{m-2K_1}, t_m\}$. Then the ith jump happens on the interval $(t_{2(b_{i\,m}-1)K_1}, t_{2b_{i\,m}K_1}]$ where $b_{i\,m} = \left[\frac{m\tau_i}{2K_1}\right] + 1$. Now we study the probability that $|S_{b_{i\,m}}|$ is larger than the threshold $L_m\gamma_m + C_m$, conditional on the value of τ_i. Similarly as (3.44) we have

$$S_{b_{i\,m}} = \frac{M\sum_{j=0}^{K_1}\left(\overline{x}_{(2b_{i\,m}-1)K_1+j} - \overline{x}_{(2b_{i\,m}-2)K_1+j}\right)}{\sqrt{\frac{m}{K_2}\widehat{\sigma}^2_{\mathsf{TTS}\,b_{i\,m}}}} + \frac{MD_{i\,m}Y_i}{\sqrt{\frac{m}{K_2}\widehat{\sigma}^2_{\mathsf{TTS}\,b_{i\,m}}}}. \tag{3.50}$$

Define

$$S^1_{b_{i\,m}} = \frac{M\sum_{j=0}^{K_1}\left(\overline{x}_{(2b_{i\,m}-1)K_1+j} - \overline{x}_{(2b_{i\,m}-2)K_1+j}\right)}{\sqrt{\frac{m}{K_2}\widehat{\sigma}^2_{\mathsf{TTS}\,b_{i\,m}}}},$$

$$S^2_{b_{i\,m}} = \frac{MD_{i\,m}Y_i}{\sigma_{t_{2(b_{i\,m}-1)K_1}}},$$

and

$$S^3_{b_{i\,m}} = \frac{MD_{i\,m}Y_i}{\sqrt{\frac{m}{K_2}\widehat{\sigma}^2_{\mathsf{TTS}\,b_{i\,m}}}} - S^2_{b_{i\,m}}.$$

From the proof of Theorem 3.1, we can see that $S^1_{b_{i\,m}}$ is asymptotically standard normal as $m \to \infty$. Since $L_m\gamma_m + C_m \overset{m\to\infty}{\to} \infty$, we have as $m \to \infty$,

$$\lim_{m\to\infty} \frac{S^1_{b_{i\,m}}}{L_m\gamma_m + C_m} \xrightarrow{\mathrm{P}} 0. \tag{3.51}$$

Given (3.46), $S^3_{b_i\,m}$ is $o_{\rm P}\left(S^2_{b_i\,m}\right)$. As

$$\mathrm{P}\left(|S_{b_i\,m}| > L_m\gamma_m + C_m\right) = \mathrm{P}\left(|S^1_{b_i\,m} + S^2_{b_i\,m} + S^3_{b_i\,m}| > L_m\gamma_m + C_m\right), \tag{3.52}$$

for any positive constants c_1, c_2 we have

$$\begin{aligned}
& \mathrm{P}\left(|S^1_{b_i\,m} + S^2_{b_i\,m} + S^3_{b_i\,m}| > L_m\gamma_m + C_m\right) \\
\geq\ & \mathrm{P}\left(|S^1_{b_i\,m}| < c_1(L_m\gamma_m + C_m), (1-c_2)|S^2_{b_i\,m}| > (1+c_1)(L_m\gamma_m + C_m), |S^3_{b_i\,m}| < c_2|S^2_{b_i\,m}|\right) \\
\geq\ & 1 - \mathrm{P}\left(|S^1_{b_i\,m}| \geq c_1(L_m\gamma_m + C_m)\right) - \mathrm{P}\left((1-c_2)|S^2_{b_i\,m}| \leq (1+c_1)(L_m\gamma_m + C_m)\right) \\
-\ & \mathrm{P}\left(|S^3_{b_i\,m}| \geq c_2|S^2_{b_i\,m}|\right).
\end{aligned} \tag{3.53}$$

Since $S^1_{b_i\,m}$ is $o_{\rm P}\left(L_m\gamma_m + C_m\right)$ and $S^3_{b_i\,m}$ is $o_{\rm P}\left(S^2_{b_i\,m}\right)$,

$$\mathrm{P}\left(|S^1_{b_i\,m}| \geq c_1(L_m\gamma_m + C_m)\right) \to 0 \text{ and } \mathrm{P}\left(|S^3_{b_i\,m}| \geq c_2|S^2_{b_i\,m}|\right) \to 0, \tag{3.54}$$

as $m \to \infty$. Since c_1, c_2 can be any positive constants, when $c_1 \to 0, c_2 \to 0$, the right-side of the second inequality of (3.53) converges to $1 - \mathrm{P}\left(|S^2_{b_i\,m}| \leq L_m\gamma_m + C_m\right) = \mathrm{P}\left(|S^2_{b_i\,m}| > L_m\gamma_m + C_m\right)$, as $m \to \infty$. In addition,

$$\begin{aligned}
& \mathrm{P}\left(|S^1_{b_i\,m} + S^2_{b_i\,m} + S^3_{b_i\,m}| > L_m\gamma_m + C_m\right) \\
\leq\ & \mathrm{P}\left((1+c_2)|S^2_{b_i\,m}| > (1-c_1)(L_m\gamma_m + C_m)\right) + \mathrm{P}\left(|S^3_{b_i\,m}| \geq c_2|S^2_{b_i\,m}|\right) + \\
& \mathrm{P}\left(|S^1_{b_i\,m}| > c_1(L_m\gamma_m + C_m)\right).
\end{aligned} \tag{3.55}$$

Following a similar analysis, it can be checked that when $c_1 \to 0, c_2 \to 0$, the right-side of (3.55) also converges to $\mathrm{P}\left(|S^2_{b_i\,m}| > L_m\gamma_m + C_m\right)$ as $m \to \infty$. Therefore, we have that $\mathrm{P}\left(|S^1_{b_i\,m} + S^2_{b_i\,m} + S^3_{b_i\,m}| > L_m\gamma_m + C_m\right)$ converges to $\mathrm{P}\left(|S^2_{b_i\,m}| > L_m\gamma_m + C_m\right)$ as $m \to \infty$, and

$$\begin{aligned}
& \mathrm{P}\left(|S^2_{b_i\,m}| > L_m\gamma_m + C_m\right) \\
=\ & \mathrm{P}\left(|Y_i| > \frac{\sigma_{t_{2(b_i\,m-1)K_1}}}{MD_{i\,m}}(L_m\gamma_m + C_m)\right) \\
=\ & 1 - F(G_{i\,m}).
\end{aligned} \tag{3.56}$$

Similarly, conditional on the occurence times of all the jumps on $(0,1]$ we have

$$\begin{aligned} & \mathrm{P}\left(|S_{b_1\,m}| > L_m\gamma_m + C_m, |S_{b_2\,m}| > L_m\gamma_m + C_m, \ldots, |S_{b_{N_1}\,m}| > L_m\gamma_m + C_m\right) \\ \to\ & \mathrm{P}\left(|S^2_{b_1\,m}| > L_m\gamma_m + C_m, |S^2_{b_2\,m}| > L_m\gamma_m + C_m, \ldots, |S^2_{b_{N_1}\,m}| > L_m\gamma_m + C_m\right) \\ =\ & \prod_{i=1}^{N_1} \mathrm{P}\left(|S^2_{b_i\,m}| > L_m\gamma_m + C_m\right) = \prod_{i=1}^{N_1}(1 - F(G_{i\,m})), \end{aligned} \tag{3.57}$$

as $m \to \infty$. Therefore, by the definition of $GFTD$,

$$\begin{aligned} & \mathrm{P}\left(GFTD|\{\tau_1, \tau_2, \ldots, \tau_{N_1}\}\right) \\ =\ & 1 - \mathrm{P}\left(|S_{b_1\,m}| > L_m\gamma_m + C_m, |S_{b_2\,m}| > L_m\gamma_m + C_m, \ldots, |S_{b_{N_1}\,m}| > L_m\gamma_m + C_m\right) \\ \to\ & 1 - \prod_{i=1}^{N_1}(1 - F(G_{i\,m})), \end{aligned} \tag{3.58}$$

as $m \to \infty$. □

Proof of Corollary 3.2. As Y_i is a.s. nonzero, we have $F(0) = 0$. Thus given (3.58), (3.19) is proved if we show that $G_{i\,m} \xrightarrow{\mathrm{P}} 0$ as $m \to \infty$. We have

$$G_{i\,m} = \sqrt{\frac{K_1(K_1+1)(2K_1+1)}{3m}}\frac{\sigma_{t_{2(i-1)K_1}}}{D_{i\,m}}\gamma_m L_m + \sqrt{\frac{K_1(K_1+1)(2K_1+1)}{3m}}\frac{\sigma_{t_{2(i-1)K_1}}}{D_{i\,m}}C_m. \tag{3.59}$$

For the first part on the right-side of (3.59), recall that there exists $\epsilon > 0$ such that $\frac{\gamma_m}{\sqrt{\log(m)}}m^{-\frac{1}{2}+\frac{1}{2}\alpha_1+\epsilon} \to 0$, as $m \to \infty$. We have

$$\begin{aligned} & \sqrt{\frac{K_1(K_1+1)(2K_1+1)}{3m}}\frac{\sigma_{t_{2(i-1)K_1}}}{D_{i\,m}}\gamma_m L_m \\ =\ & \sigma_{t_{2(i-1)K_1}}\frac{\gamma_m m^{-\frac{1}{2}+\frac{1}{2}\alpha_1+\epsilon}}{\sqrt{\log(m)}}(L_m\sqrt{\log(m)})\frac{\sqrt{\frac{K_1(K_1+1)(2K_1+1)}{3m}}}{m^{\frac{3}{2}\alpha_1-\frac{1}{2}}}\frac{m^{\alpha_1-\epsilon}}{D_{i\,m}} \xrightarrow{\mathrm{P}} 0, \end{aligned} \tag{3.60}$$

as $m \to \infty$, given that $L_m\sqrt{\log(m)}$, $\frac{\sqrt{\frac{K_1(K_1+1)(2K_1+1)}{3m}}}{m^{\frac{3}{2}\alpha_1-\frac{1}{2}}}$ are $O(1)$ and based on Corollary 3.1, $\frac{m^{\alpha_1-\epsilon}}{D_{i\,m}}$ is $o_{\mathrm{P}}(1)$. It can be also checked that the second part on the right-side of (3.59) converges to zero in probability as $m \to \infty$. Thus $G_{i\,m} \xrightarrow{\mathrm{P}} 0$ as $m \to \infty$. □

Proof of Theorem 3.4. Recall the definition of X in Lemma 3.2. We have

$$\begin{aligned} \mathrm{P}\left(GSD\right) &= 1 - \mathrm{P}\left(\frac{\max_{i \in B_m} |S_i| - C_m}{L_m} \leq \gamma_m\right) \\ &\to 1 - \mathrm{P}(X \leq \gamma_m) \\ &= 1 - \exp\left(-e^{-\gamma_m}\right) = \alpha_m. \end{aligned} \tag{3.61}$$

□

Bibliography

Acemoglu, D., Carvalho, V., Ozdaglar, A., and Tahbaz-Salehi, A. (2012). The Network Origins of Aggregate Fluctuations. *Econometrica*, 80:1977–2016.

Aït-Sahalia, Y., Fan, J., Laeven, R. J., Wang, C. D., and Yang, X. (2016). Estimation of the Continuous and Discontinuous Leverage Effects. *Journal of the American Statistical Association, in press*.

Aït-Sahalia, Y. and Jacod, J. (2009). Testing for Jumps in a Discretely Observed Process. *The Annals of Statistics*, 37(1):184–222.

Aït-Sahalia, Y. and Jacod, J. (2011). Testing whether Jumps Have Finite or Infinite Activity. *The Annals of Statistics*, 39(3):1689–1719.

Aït-Sahalia, Y. and Jacod, J. (2014). High-Frequency Financial Econometrics. *Princeton, N.J.: Princeton University Press, first edition*.

Aït-Sahalia, Y., Jacod, J., and Li, J. (2012). Testing for Jumps in Noisy High-Frequency Data. *Journal of Econometrics*, 168(2):207–222.

Aït-Sahalia, Y., Mykland, P. A., and Zhang, L. (2005). How often to sample a continuous-time process in the presence of market microstructure noise. *Review of Financial Studies*, 18(2):351–416.

Andersen, T. G., Bollerslev, T., and Diebold, F. X. (2007). Roughing it up: Including jump components in the measurement, modeling, and forecasting of return volatility. *The Review of Economics and Statistics*, 89(4):701–720.

Andersen, T. G., Bollerslev, T., Diebold, F. X., and Labys, P. (2003). Modeling and Forecasting Realized Volatility. *Econometrica*, 71:579–625.

Bakshi, G., Cao, C., and Chen, Z. (1997). Empirical Performance of Alternative Option Pricing Models. *The Journal of Finance*, 52(5):2003–2049.

Bandi, F. M. and Russell, J. R. (2006). Separating microstructure noise from volatility. *Journal of Financial Economics*, 79:655–692.

Banerjee, O. and Ghaoui, L. E. (2008). Model Selection Through Sparse Maximum Likelihood Estimation for Multivariate Gaussian or Binary Data. *Journal of Machine Learning Research*, 9:485–416.

Barigozzi, M. and Brownlees, C. (2013). NETS: Network Estimation for Time Series. Technical report, Barcelona GSE.

Barndorff-Nielsen, O. and Shephard, N. (2004a). Econometric analysis of realized covariation: High frequency based covariance, regression, and correlation in financial economics. *Econometrica*, 72(3):885–925.

Barndorff-Nielsen, O. and Shephard, N. (2004b). Power and Bipower Variation with Stochastic Volatility and Jumps. *Journal of Financial Econometrics*, 2(1):1–48.

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*, 76:1481–1536.

Barndorff-Nielsen, O. E., Hansen, P. R., Lunde, A., and Shephard, N. (2009). Realised kernels in practice: Trades and quotes. *Econometrics Journal*, 12:1–32.

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, Ser. B*, 64:253–280.

Barndorff-Nielsen, O. E. and Shephard, N. (2004c). Econometric analysis of realized covariation: High frequency based covariance, regression, and correlation in financial economics. *Econometrica*, 72(3):885–925.

Barndorff-Nielsen, O. E. and Shephard, N. (2006). Econometrics of Testing for Jumps in Financial Economics Using Bipower Variation. *Journal of Financial Econometrics*, 4(1):1–30.

Barndorff-Nielsen, O. E., Shephard, N., and Winkel, M. (2006). Limit Theorems for Multipower Variation in the Presence of Jumps. *Stochastic Processes and their Applications*, 116(5):796–806.

Barunik, J. and Vacha, L. (2015). Realized Wavelet-Based Estimation of Integrated Variance and Jumps in the Presence of Noise. *Quantitative Finance*, 15(8):1347–1364.

Bates, D. S. (2000). Post-'87 Crash Fears in the S& P 500 Futures Option Market. *Journal of Econometrics*, 94(1-2):181–238.

Billio, M., Getmansky, M., Lo, A., and Pellizzon, L. (2012). Econometric Measures of Connectedness and Systemic Risk in the Finance and Insurance Sectors. *Journal of Financial Economics*, 104:535–559.

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

Chernov, M., Gallant, R., Ghysels, E., and Tauchen, G. (2003). Alternative Models for Stock Price Dynamics. *Journal of Econometrics*, 116(1-2):225–257.

Christensen, K., Oomen, R., and Podolskij, M. (2010). Realised Quantile-Based Estimation of the Integrated Variance. *Journal of Econometrics*, 159(1):74–98.

Cont, R. and Mancini, C. (2011). Nonparametric tests for pathwise properties of semimartingales. *Bernoulli*, 17(2):781–813.

Cont, R. and Tankov, P. (2004). *Financial Modelling with Jump Processes*. Chapman & Hall/CRC, Bocaraton.

Corsi, F., Peluso, S., and Audrino, F. (2014). Missing in asynchronicity: a kalman-em approach for multivariate realized covariance estimation. *Journal of Applied Econometrics*.

Corsi, F., Pirino, D., and Reno, R. (2010). Threshold Bipower Variation and the Impact of Jumps on Volatility Forecasting. *Journal of Econometrics*, 159(2):276–288.

Dempster, A. P. (1972). Covariance selection. *Biometrics*, 28:157–175.

Diebold, F. and Yilmaz, K. (2014). On the Network Topology of Variance Decompositions: Measuring the Connectedness of Financial Firms. *Journal of Econometrics*, 182:119–134.

Engle, R. and Colacito, R. (2006). Testing and valuing dynamic correlations for asset allocation. *Journal of Business and Economic Statistics*, 24:238–253.

Fan, J., Fan, Y., and Lv, J. (2008). High Dimensional Covariance Matrix Estimation using a Factor Model. *Journal of Econometrics*, 147:186–197.

Fan, J., Li, Y., and Yu, K. (2012). Vast volatility matrix estimation using high-frequency data for portfolio selection. *Journal of the American Statistical Association*, 107(497):412–428.

Fan, J., Liao, Y., and Mincheva, M. (2011). High Dimensional Covariance Matrix Estimation in Approximate Factor Models. *The Annals of Statistics*, 39:3320–3356.

Fan, J., Liao, Y., and Mincheva, M. (2013). Large Covariance Estimation by Thresholding Principal Orthogonal Complements. *Journal of the Royal Statistical Society, Series B*.

Fan, J. and Wang, Y. (2007). Multi-Scale Jump and Volatility Analysis for High-Frequency Financial Data. *Journal of the American Statistical Association*, 102(480):1349–1362.

Friedman, J., Hastie, T., and Tibshirani, R. (2011). Sparse inverse covariance estimation with the graphical lasso. *Biostatistics*, 9(3):432–441.

Friedman, J., Hastle, T., Hofling, H., and Tibshirani, R. (2007). Pathwise Coordinate Optimization. *The Annals of Applied Statistics*, 1:302–332.

Hansen, P. R. and Lunde, A. (2012). Realized Variance and Market Microstructure Noise. *Journal of Business and Economic Statistics*, 24(2):127–161.

Hautsch, N., Kyj, L. M., and Oomen, R. C. A. (2012). A blocking and regularization approach to high dimensional realized covariance estimation. *Journal of Applied Econometrics*, 27(4):625–645.

Hautsch, N., Schaumburg, J., and Schienle, M. (2014a). Financial network systemic risk contributions. *Review of Finance*.

Hautsch, N., Schaumburg, J., and Schienle, M. (2014b). Forecasting Systemic Impact in Financial Networks. *International Jounal of Forecasting*, 30(3):781–794.

Ikeda, N. and Watanabe, S. (1981). *Stochastic Differential Equations and Diffusion Processes*. North Holland, Amsterdam.

Jacod, J. (2008). Asymptotic Properties of Realized Power Variations and Related Functionals of Semimartingales. *Stochastic Processes and their Applications*, 118(4):517–559.

Jacod, J., Podolskij, M., and Vetter, M. (2010). Limit Theorems for Moving Averages of Discretized Processes plus Noise. *The Annals of Statistics*, 38(3):1478–1545.

Jacod, J. and Protter, P. (2012). Discretization of Processes. *New York: Springer-Verlag, first edition*.

Jacod, J. and Todorov, V. (2009). Testing for Common Arrivals of Jumps for Discretely Observed Multidimensional Processes. *The Annals of Statistics*, 37(4):1792–1838.

Jacod, J. and Todorov, V. (2010). Do Price and Volatility Jump Together? *The Annals of Applied Probability*, 20(4):1425–1469.

Jacod, J. and Todorov, V. (2014). Efficient Estimation of Integrated Volatility in Presence of Infinite Variation Jumps. *The Annals of Statistics*, 42(3):1029–1069.

Jiang, G. J. and Oomen, R. C. (2008). Testing for Jumps when Asset Prices are Observed with Noise-a "Swap Variance" Approach. *Journal of Econometrics*, 144(2):352–370.

Karatzas, I. and Shreve, S. (1999). *Brownian Motion and Stochastic Calculus*. New York: Springer.

Kim, D., Wang, Y., and Zou, J. (2016). Asymptotic Theory for Large Volatility Matrix Estimation based on High-frequency Financial Data. *Stochastic Processes and Their Applications*, forthcoming.

Laloux, L., Cizeau, P., Bouchaud, J.-P., and Potters, M. (1999). Noise Dressing of Financial Correlation Matrices. *Phisical Review Letters*, 83(7):1467–1470.

Lauritzen, S. L. (1996). *Graphical Models*. Oxford University Press, USA.

Ledoit, O. and Wolf, M. (2004). A Well-Conditioned Estimator for Large–Dimensional Covariance Matrices. *Journal of Multivariate Analysis*, 88:365–411.

Ledoit, O. and Wolf, M. (2012). Nonlinear Shrinkage Estimation of Large–Dimensional Covariance Matrices. *Annals of Statistics*, 40:1024–1060.

Lee, S. S. and Mykland, P. A. (2008). Jumps in Financial Markets: A New Nonparametric Test and Jump Dynamics. *Review of Financial Studies*, 21(6):2535–2563.

Lee, S. S. and Mykland, P. A. (2012). Jumps in Equilibrium Prices and Market Microstructure Noise. *Journal of Econometrics*, 168(2):396–406.

Li, J., Todorov, V., Tauchen, G., and Lin, H. (2016). Rank tests at Jump Events. *preprint*.

Lunde, A., Shephard, N., and Sheppard, K. (2011). Econometric analysis of vast covariance matrices using composite realized kernels. Technical report.

Mancini, C. (2008). Large deviation principle for an estimator of the diffusion coefficient in a jump-diffusion process. *Statistics and Probability Letters*, 78(7):869–879.

Mancini, C. (2009). Non-parametric Threshold Estimation for Models with Stochastic Diffusion Coefficient and Jumps. *Scandinavian Journal of Statistics*, 36(2):270–296.

Martens, M. (2004). Estimating Unbiased and Precise Realized Covariances. *Econometric Institute, Eramus University Rotterdam*.

Meinshausen, N. and Bühlmann, P. (2006). High Dimensional Graphs and Variable Selection with the Lasso. *The Annals of Statistics*, 34:1436–1462.

Metivier, M. (1982). *Semimartingales: A Course on Stochastic Processes*. De Gruyter, New York.

Mykland, P. A. and Zhang, L. (2006). ANOVA for Diffusions and Ito Processes. *The Annals of Statistics*, 34(4):1931–1963.

Nikolaus, H., Lada, M., K., and Peter, M. (2013). Do High-Frequency Data Improve High-Dimesional Portfolio Allocations? *Journal of Applied Econometric, forthcoming*.

Pan, J. (2002). The Jump-Risk Premia Implicit in Options: Evidence from an Integrated Time-Series Study. *Journal of Financial Economics*, 63(1):3–50.

Piazzesi, M. (2005). Bond Yields and the Federal Reserve. *Journal of Political Economy*, 113(2):311–344.

Podolskij, M. and Vetter, M. (2009). Estimation of Volatility Functionals in the Simultaneous Presence of Microstructure Noise and Jumps. *Bernoulli*, 15(3):634–658.

Pourahmadi, M. (2013). *High-dimensional covariance estimation*. John Wiley&Sons, Inc., Hoboken, New Jersey.

Ravikumar, P., Wainwright, M. J., Raskutti, G., and Yu, B. (2011). High-dimensional covariance estimation by minimizing ℓ_1-penalized log-determinant divergence. *Electronic Journal of Statistics*, 5:935–980.

Sato, K. (1999). *Lévy Processes and Infinitely Divisible Distributions*. Cambridge University Press, Cambridge.

Stein, C. (1956). Inadmissibility of the usual estimator for the mean of a multivariate normal distribution. In *Proceedings of the 3rd Berkeley Symposium on Mathematical Statistics and Probability*, pages 157–164. University of California Press.

Tao, M., Wang, Y., and Zhou, H. H. (2013). Optimal Sparse Volatility Matrix Estimation for High-dimensional Itô Processes with Measurement Errors. *The Annals of Statistics*, 41:1816–1864.

Tauchen, G. and Zhou, H. (2011). Realized jumps on financial markets and predicting credit spreads. *Journal of Econometrics*, 160(1):102–118.

Varneskov, R. T. (2016). Flat-Top Realized Kernel Estimation of Quadratic Covariation with Nonsynchronous and Noisy Asset Prices. *Journal of Business and Economic Statistics*, 34(1):1–22.

Wang, Y. and Zou, J. (2010). Vast Volatility Matrix Estimation for High-frequency Financial Data. *The Annals of Statistics*, 38:943–978.

Woerner, W. (2006). Power and Multipower Variation: Inference for High Frequency Data. *In Stochastic Finance, A.N. Shiryaev, M. do Rosário Grosshino, P. Oliviera, M. Esquivel, eds., 343-354, Springer*, pages 343–354.

Yuan, M. and Lin, Y. (2007). Model Selection and Estimation in the Gaussian Graphical Model. *Biometrika*, 94(1):19–35.

Zhang, L. (2011). Estimating covariation: Epps effect, microstructure noise. *Journal of Econometrics*, 160(1):33–47.

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.