Mattiussi, V. (2010). Non parametric Estimation of high-frequency Volatility and Correlation Dynamics. (Unpublished Doctoral thesis, City University London)

CITY UNIVERSITY LONDON
EST 1894

Original citation: Mattiussi, V. (2010). Non parametric Estimation of high-frequency Volatility and Correlation Dynamics. (Unpublished Doctoral thesis, City University London)

Permanent City Research Online URL: http://openaccess.city.ac.uk/12095/

Nonparametric Estimation of High-frequency Volatility and Correlation Dynamics

Vanessa Mattiussi

Submitted in partial fulfillment of the requirements
for the degree of Doctor of Philosophy in Economics

at

City University
Department of Economics

December 2010

Declaration

I hereby declare that this submission is my own work and that, to the best of my knowledge and belief, it contains no material previously published or written by another person nor material which has been accepted for the award of any other degree or diploma of the university or other institute of higher learning.

I also hereby grant power of discretion to the University Librarian to allow this thesis to be copied in whole or in part without further reference to me. This permission covers only single copies made for study purposes, subject to normal conditions of acknowledgment.

Acknowledgements

Over the years of my Ph.D. program I had the opportunity to collaborate with a number of superb researchers, whose contribution to the making of this thesis deserved special mention. It is a pleasure to convey my gratitude to them all in my humble acknowledgment.

I first wish to thank my supervisor Giulia Iori for making this experience productive and interesting, and for always giving me the freedom to explore, develop and incorporate my own ideas into the research projects.

My very special thanks go to Roberto Renò for his crucial contribution, his endlessly creative ideas and boundless enthusiasm, which made him a backbone of my research and so of this thesis. He provided me unflinching encouragement and support in various ways, while triggering and nourishing my intellectual maturity I know I will always benefit from.

I am much indebted to Michele Tumminello for his great intuition and passion, his valuable help, and for using his precious time to read a chapter of this thesis and provide his critical comments about it. The quality and amount of research that I was able to produce as a result of our collaboration goes far behind what I would have been able to do alone.

I wish to express my warm and sincere gratitude to the people I met during my visiting year at the University of Siena for creating a stimulating environment in which to learn and grow. In particular, I am truly grateful to Fulvio Corsi, Claudio Pacati and Simona Sanfelici.

My gratitude also goes out to the members of my oral defense committee, Angeles Carnero and Gabriel Monte-Rojas, for their time, interest and very helpful comments.

Lastly, I would like to thank my infinitely patient parents, my very supportive brother and my close friends. I love you all and always will.

Contents

Abstract

This thesis addresses the problem of quantitatively evaluating the temporal dynamics that characterized financial time series. In particular, we perform an accurate analysis of the Fourier estimator, a newly proposed nonparametric methodology to measure ex-post volatility and cross-volatilities as functions of time, when financial assets are observed at different high-frequency levels over the day. The estimator has the peculiar feature to employ the observed data in their original form, therefore exploiting all the available information in the sample. We first show how to considerably improve the numerical performance of the Fourier method making possible the analysis of large sets of data, as it is usually the case with high-frequency series. Secondly, we use Monte Carlo simulation methods to study the behavior of three driving parameters in the estimation procedure, when the effects of both irregular sampling and microstructure noise are taken into account. The estimator is showed to be particularly sensitive to one of these quantities, which is in turn used to control the contribution of the above effects. Integrated financial correlation is also analyzed within two distinct comparative studies that involve other multivariate measures. The analysis is then extended to consider the entire evolution of the underlying correlation process. Finally, we propose a new class of nonparametric spot volatility estimators, which is showed to include the Fourier method as a particular case. The full limit theory under infill asymptotics in the pure diffusive settings of the class is derived. Empirical evidence in support of our conclusions is also provided.

Introduction

In the last decade, high-frequency financial data have become increasingly available for a wide range of securities allowing for a deeper understanding of complex intraday patterns. Within a high-frequency domain the price formation is followed in real time, or tick-by-tick, resulting in a large amount of observed values and, therefore, in a virtually continuous process. This has spawned considerable interest in the use of these data to the study of financial market volatility and correlation, two fundamental parameters that play a central role in the theory and practice of asset pricing, portfolio selection, and risk management. Although several traditional models still assume volatilities and correlations to be constant, it is widely recognized among both finance academics and practitioners that, instead, they vary importantly over time, rising the need to properly analyze their distributional and dynamic properties.

The existing measurement procedures for these variables can be coarsely divided into *parametric* and *nonparametric* models. Within the first category, volatility is considered as unobservable quantity and is modelled by a fully specified functional structure, usually rather complex. Both the large ARCH-GARCH family and the stochastic volatility models belong to the category. A nonparametric approach allows instead to treat volatility as it was observable, making possible to directly analyze, model, and forecast the variable itself. See, for instance, Andersen et al. (2009) review. *Realized volatility* is the most popular and widely used nonparametric estimator. It is simply defined as the sum of squared high-frequency returns over a given time interval. The main idea is to aggregate squared intra-daily returns to approximate the daily increments of the semimartingale that drives the underlying logarithmic price process, the so-called *integrated volatility*. We refer to the works by Andersen et al. (2001) and Barndorff-Nielsen and Shephard (2002a,b) as the most influential on the topic. When prices are observed at the highest possible frequency, realized volatility is in principle a very

accurate nonparametric measure of the integrated variance. Unfortunately, the presence of microstructure noise in empirical data generates a divergence between the observed price process and the true or frictionless price process. In pure probability terms, the observed log-return process is no longer a semimartingale. This divergence could, for example, be induced by transaction price changes occurring as multiples of ticks (price discreteness) or by the existence of multiple prices for buyers and sellers (bid-ask spreads), or it may be due to liquidity or information reasons. Under these settings, it turns out that the realized volatility actually estimates the variance of the contamination noise, rather than the underlying return volatility. This is because, at least for the class of continuous semimartingale processes, the volatility is of the same order of magnitude as the time interval, while the microstructure noise has a roughly constant variance. Therefore, changes in transaction prices over very small time intervals are mainly composed of noise and carry little information about the underlying return volatility. The effect of noise components on the estimation method has been analyzed in Bandi and Russell (2006b, 2008), Aït-Sahalia et al. (2005, 2011), Barndorff-Nielsen et al. (2008, 2006), Zhang et al. (2005), Hansen and Lunde (2006) and Zhang (2006a). The concept of realized volatility can be easily extended to the multivariate case through the definition of *realized covariance*. It is though necessary to take into account an additional effect related to the non-synchronous arrival times of the traded assets: non-synchronicity relates to the so-called Epps phenomenon (Epps, 1979), which induces a downward bias in the covariance estimates as the sampling frequency increases. The problem is addressed in Bandi and Russell (2005), Zhang (2006b), Voev and Lunde (2007) and Griffin and Oomen (2010) among others. An alternative, but similar in spirit to the realized covariation estimator, is proposed in Hayashi and Kusuoka (2004), and Hayashi and Yoshida (2005, 2006). The peculiarity of their estimator is that it does not rely on any synchronization methods, instead required for the realized covariance, but rather use all the available data. More importantly, it is shown to be unbiased and consistent under the assumption that the observations are uncontaminated by noise.

The methodology here studied follows a very different approach respect to the realized estimator: it is based on harmonic analysis and exploits high-frequency data to recover the time evolution of volatility and cross-volatility processes over a fixed time window. The method was first introduced by Malliavin and Mancino (2002) in the context of Fourier series following the idea to expand the instantaneous multivariate volatility function into trigonometric polynomials with Fourier coefficients depending on the log-return process. An estimate of the integrated volatility and cross-volatility can be obtained as a particular case. In this respect, early recognition on the validity of the method, in absence of microstructure noise, can be found in Barucci and Renò (2002a,b), Kanatani (2004) and Hansen and Lunde (2005). In a comparative study, Høg and Lunde (2003) show that the their proposed wavelet estimator is virtually indistinguishable from the Fourier estimator in terms of bias and variance, al-

though the former method is computationally faster. However, as in the last cited works, they do not take into account any market effect. Nielsen and Frederiksen (2008) bridge the gab and also include in the analysis the realized volatility, as well as two alternative estimators specifically designed to account for the bias induced by microstructure noise. Their conclusion is that the Fourier estimator is superior to both the realized volatility and the wavelet estimator, and has smaller mean square error, but slightly higher bias, with respect to the bias-corrected methods. Nevertheless, this analysis is purely empirical and a precise treatment of market microstructure noise effects on the Fourier estimator is needed. In their article, Mancino and Sanfelici (2008b) derive an analytic expressions for the bias and the mean squared error of the estimator under the condition the noise component is an independent random variable respect to the Brownian motion driving the stock price dynamics. They show that, by choosing an appropriate number of Fourier coefficients, the bias of the Fourier estimator is smaller than the bias of the realized volatility in finite samples, while the mean square error converges to a positive constant, i.e. nearly consistency. Hence, the estimator can be made sufficiently robust to this type of market noise. One limitation of this study is that prices are assumed to be equidistant or homogeneous in time, therefore requiring interpolation methods to construct intraday returns, while the most appealing feature of the Fourier estimator is that it can be applied directly to market data, making use of all the available information. In addition, a more realistic microstructure noise dependence should also be analyzed as in Hansen and Lunde (2006), where the noise is time-dependent and correlated with efficient returns. In the multivariate framework, we first recall the work by Renò (2003), who exploits the methodology to investigate the determinants of the aforementioned Epps effect. Using real time series, Precup and Iori (2007) make use of signature plots to show that the Fourier method generates smoother curves with respect to interpolation based methods such as the standard Pearson coefficient and the co-volatility weighted measure proposed in Dacorogna et al. (2001). Finally, Mancino and Sanfelici (2008a) valuate the effects of both asynchronous trading and microstructure noise on the Fourier estimator and conclude that: (i) is consistent under asynchronous trading and uncontaminated prices; (ii) is asymptotically unbiased and nearly consistent in the presence of independent noise.

All these approaches, valuable as they are, focus on realized measures. We are instead interested in applying the Fourier methodology to estimate the variability of the price dynamics at a particular point in time, the so-called *instantaneous* or *spot* volatility. A first step towards this direction is the work by Renò (2008). In his stimulating paper, the author employs the Fourier estimator to reconstruct the volatility trajectory of a stochastic model. Similarly, Ogawa and Sanfelici (2008) applied the method to spot volatility estimation but also consider the case of slightly contaminated data. However, the significant impact that irregular sampling might have on the estimation procedure is not considered, since both studies employ series of evenly spaced transactions. In this respect, the first purpose of this thesis is to

place further emphasis on the estimation of univariate and multivariate volatility dynamics, while providing a deeper insight into the Fourier method. Secondly, we would like to wide and sustain the existing class of spot volatility estimators by proposing a new and efficient volatility measure. In particular, the thesis consists of four main chapters as follow.

Chapter 1 describes in details the Fourier methodology and the asymptotic properties of the estimator under univariate and multivariate settings.

Chapter 2 first shows how to apply the Fast Fourier Trasform (FFT) to enhance the calculation speed of the method. The original Fourier procedure can be quite expensive in terms of memory requirement and computational time, but we show that it can be substantially improved by means of the FFT algorithm. This is of great advantage for the analysis of large sets of data, as it is usually the case when dealing with tick-by-tick prices. In the second part of the chapter, we consider different levels of inhomogeneity and of microstructure noise to study the behavior of three driving quantities in the estimation process, namely the number of price and volatility Fourier coefficients, and the degree of resolution at which to reconstruct the volatility trajectory. We simultaneously derive their optimal values using the genetic algorithm for numerical optimization introduce by Storn and Price (1997), and known as Differential Evolutions. Monte Carlo results reveal that the number of price coefficients is a key parameter in controlling the efficiency of the estimator. In addition, we show that the estimation process is affected by the joint contribution of irregular sampling and market noise, especially when the level of these two 'ugly' facts is significant. The chapter also presents two empirical applications that, overall, provide evidence in favor of the method.

Chapter 3 deals with correlation analysis. We first evaluate the performance of the Fourier method against the popular Pearson coefficient to estimate the integrated correlation of a large set of stocks. Although very different in their design, the two estimators result to be equivalent proxies of the 'true' correlation matrix, both in the case of synchronous and asynchronous data. To measure their level of accuracy we use the Kullback-Leibler divergence, a dissimilarity measure well-known in Information Theory. The peculiar features of this measure, namely, the asymmetry and the independence of the expectation values from the original model, suit particularly well a comparative analysis between correlation matrices. In this respect, a small Kullback-Leibler distance indicates that a large amount of information about the model is retained by the sample matrix, here calculated with the Fourier and Pearson estimators respectively. In a second stage of the analysis we instead compare the Fourier method against the highly competitive correlation estimator introduced by Hayashi and Yoshida (2005, 2006). Both methods can be applied directly to series of tick-by-tick prices without any prior manipulation, but the latter is computationally simpler and, most importantly, immune to Epps effects by construction. However, differently from the Fourier approach, the Hayashi-Yoshida correlation matrix might fail to be positive de-

fined, a mandatory requirement in the application of the Kullback-Leibler measure. To solve the problem, we apply the *shrinkage* technique proposed in Ledoit and Wolf (2004b) and construct a convex combination between the Fourier and Hayashi-Yoshida sample matrices. Our results clearly show that the obtained shrinkage estimator outperforms its single components, leaving space for a better estimation procedure once these are optimally combined. The last part of the chapter deals with the estimate of time-varying and stochastic correlation. The Fourier estimator is here applied for the first time and the results are only preliminary.

Chapter 4 introduces a new class of spot volatility estimators. Members of this class are fully nonparametric measures based on *delta sequences*, i.e. sequences of functions that converges at a certain rate to a Dirac delta function. In particular, we extend the kernel estimator of Kristensen (2010) by proving that a traditional kernel function can be seen as a delta sequence. Our class is shown to be reasonably wide and also includes the trigonometric functions largely used in the Fourier estimator. The study of the asymptotic theory reveals that the estimators within the class are normally distributed when the number of observations diverges to infinity, and the maximum interval between observations (not necessarily equally spaced) shrinks to zero. Our findings are derived for diffusions under mild assumptions on the driving coefficients of the stochastic differential equation. We then allow for microstructure noise in the data and use a two-scale volatility technique, similar to the one in Zhang et al. (2005), to make our estimator robust against the noise effects. In addition, we tackle the problem of discontinuities in the return dynamics using a threshold estimator as in Mancini (2009) to filter out jumps from the observed price process. As a by-product, we contribute to the result in Malliavin and Mancino (2009) by further investigating the asymptotic behavior of the Fourier estimator. Finally, we present an empirical analysis using high-frequency transactions, where the above estimators are applied to detect intraday volatility dynamics.

The Fourier estimator

This chapter introduces the Fourier estimator under univariate and multivariate settings. Some basics concepts in stochastic analysis that underlie the main theoretical results regarding the method are first given.

1.1 Basic notions of stochastic processes

A *stochastic process* $X = \{X(t) : t \geq 0\}$ is a collection of random variables $X(t) = X(t, \omega)$ indexed by t and defined on a common probability space $(\Omega, \mathcal{F}, P)$. It can also be regarded as the function $X : \mathbb{R}_+ \times \Omega \to \mathbb{R}$ and $X(t, \omega)$ can be interpreted as the value of the process at t given the outcome ω. For a fixed $\omega \in \Omega$ the map $t \to X(t, \omega)$ is called a *trajectory of the process X*. For a specific t, the process is simply a random variable on Ω.

A stochastic process X is *càdlàg* if all its trajectories are almost surely (a.s.) right-continuous with left-hand limits everywhere. In this case, we can define two other processes: the left-limit process $X_- = \{X(t-)\}$ and the jump process $\Delta X = \{\Delta X(t)\}$, respectively given by

$$
\begin{aligned}
X(t-) &= \lim_{s \nearrow t} X(s), \qquad X(0-) = X(0) \\
\Delta X(t) &= X(t) - X(t-)
\end{aligned}
$$

If the trajectory is continuous in t, then $X(t-) = X(t)$ and $\Delta X(t) = 0$. Càdlàg functions are important in the study of stochastic processes that admit, or even require, jumps.

In what follows, we assume that the price evolution of a financial asset is modelled by a stochastic process X. In particular, we assume that the price process evolves in continuous time over the interval $[0, T]$, with T finite integer. We denote the associated natural filtration as $\{\mathcal{F}_t\}_{0 \leq t \leq T} \subseteq \mathcal{F}$, where the information set $\mathcal{F}_t$ represents the available history of the asset price movement and other relevant, possibly latent, state variables up to time t. We will refer to $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t \geq 0}, P)$ as to the *filtered probability space*.

We say that the process X is *adapted* to the filtration $\{\mathcal{F}_t\}_{0 \leq t \leq T}$ if it is $\mathcal{F}_t$-measurable, i.e. $X(t) \in \mathcal{F}_t$, $\forall t \geq 0$. The interpretation of this definition is that for every fixed t, the process value $X(t)$ is completely determined by the information $\mathcal{F}_t$ that we have access to at time t (Bjork, 1998). Therefore, an adapted process does not look into the future.

Given the above filtered probability space, we can define the *predictable σ-algebra* $\mathcal{P}$ as the smallest σ-algebra on $\mathbb{R}_+ \times \Omega$ generated by the class of all adapted, left-continuous processes. A stochastic process X is called *predictable* if it is measurable respect to $\mathcal{P}$.

A stochastic process X is a *martingale* if

1. $X(t)$ is integrable for each t;

2. X is adapted to $\{\mathcal{F}_t\}_{0 \leq t \leq T}$;

3. $X(s) = E[X(t)|\mathcal{F}_s]$, a.s. $\forall s, t$ such that $0 \leq s \leq t$.

It the equality sign is replaced by $\leq$ ($\geq$), then X is said to be a *submartingale (supermartingale)*.

A random variable τ taking values in $\mathbb{R}_+ \cup \{+\infty\}$ is a *stopping time* with respect to $\{\mathcal{F}_t\}_{0 \leq t \leq T}$ if $\{\tau \leq t\} \in \mathcal{F}_t$, $\forall t \geq 0$.

An adapted, càdlàg process X is a *local martingale* if there exists a sequence of increasing stopping times τ_n, with $\lim_{n \to \infty} \tau_n = \infty$ a.s., such that $X(t \wedge \tau_n)$ $_{\{\tau_n > 0\}}$ is a uniformly integrable martingale for each n, with $t \wedge \tau = \min(t, \tau)$.

A process X is said to have *finite* or *bounded variation* over $[0, T]$ if the sum of the absolute values of the price increments over subintervals is bounded. In formulas

$$V_X([0, T]) = \sup \sum_{i=1}^{n} |X(t_i) - X(t_{i-1})| < \infty$$

where the supremum is taken over all the partitions $0 = t_0 < t_1 < \ldots < t_n = T$ of $[0, T]$.

1.2   Semimartingales, quadratic variation and covariation

In the fundamental theory of asset pricing the arbitrage-free logarithmic price process is a
(special) *semimartingale* and, as such, admits the following unique decomposition

$$X = A + M,$$

where A is a predictable process of finite variation and M is a local martingale. The process
A (the signal) has relatively smooth sample paths, while M (the noise) is characterized by
an erratic and unpredictable behavior. The increments in A may be thought of as rewards
for investing in the risky process X. The class of semimartingales is very large and includes
all the standards models applied in financial econometrics like the Itô process, the jump and
mixed jump diffusion processes. A semimartingale X defined on $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t \geq 0}, P)$ is called
Brownian semimartingale if the process M is of the form

$$M(t) = \int_0^t a(s)ds + \int_0^t \sigma(s)dW(s),$$

where a is a predictable locally bounded drift, $\sigma(t)$ is a càdlàg volatility process, and $W(t)$
is a Brownian motion. Crucial to the economics of financial risk is the concept of *quadratic
variation* defined for any semimartingale as

$$[X, X]_t = X^2(t) - 2 \int_0^t X(s-)dX(s), \qquad 0 < t \leq T, \tag{1.1}$$

where $X(s-)$ is a càdlàg process. Equivalently, let $\{0 = \tau_0^n \leq \tau_1^n \leq \tau_2^n \leq \cdots \leq \tau_n^n = T\}$ be a
sequence of partitions of $[0, T]$ such that $\sup_{i \geq 0}(\tau_{i+1}^n - \tau_i^n) \to 0$ for $n \to \infty$, i.e. the maximum
distance between observations goes to zero in the limit. Then

$$[X, X]_t = \plim_{n \to \infty} \sum_{i \geq 1} (X(t \wedge \tau_i^n) - X(t \wedge \tau_{i-1}^n))^2, \tag{1.2}$$

where the probability limit exists for all semimartingales. Theoretical results in stochastic
processes (Karatzas and Shreve, 1988) state that almost surely

$$[X, X]_t = \int_0^t \sigma^2(s)ds, \tag{1.3}$$

where the quantity on the right-hand sise is known as *integrated volatility*. Therefore, without
any knowledge on the underlying volatility process, it is possible to obtain an accurate esti-
mate of the integrated volatility by summing increasingly finer sampled squared returns. This
feature was first applied to the context of empirical volatility measurement using high fre-
quency data by Andersen and Bollerslev (1998), Andersen et al. (2001) and, in a concurrent

and independent work, by Barndorff-Nielsen and Shephard (2001). The proposed estimator, known as *realized volatility*, take the simple form

$$RV = \sum_{i=1}^{m} r_i^2,$$

where $r_i = X_{i\Delta} - X_{(i-1)\Delta}$ is the intraday return sampled at time interval $\Delta = \frac{T}{m}$ per period, usually one day. The theory of quadratic variation shows that the realized variance converges uniformly in probability to the quadratic variation process when the sampling frequency of returns approaches infinity, i.e. for $m \to \infty$,

$$RV \xrightarrow{p} \int_0^t \sigma^2(s)ds,$$

providing a consistent estimate of the integrated volatility. This convergence property states that, in theory and under very broad assumptions, the measurement error of the realized volatility could be arbitrarily reduced by simply increasing the sampling frequency of returns. However, in practice empirical data differ in many ways from the arbitrage-free continuous time price process making this estimator strongly biased and inconsistent for small return intervals.

The extension to the multivariate case of the above results easily follow. In particular, the *quadratic covariation* of two processes $X(t)$ and $Y(t)$ can be conveniently obtained through the polarization identity

$$[X,Y]_t = \frac{1}{4} \left([X+Y, X+Y] - [X-Y, X-Y] \right), \tag{1.4}$$

although similar expressions to (1.1) and (1.2) can also be derived. In the next section we will see that, in the pure diffusive settings, the quadratic covariation effectively measure the association between the random movements of two different continuous processes.

1.3 The Fourier method

Assume that the logarithmic prices $X(t) = (X_1(t), \ldots, X_p(t))$ are Brownian semimartingales satisfying the stochastic differential equations

$$dX(t) = \Theta(t)dW(t) \tag{1.5}$$

where the *instantaneous* or *spot volatility* process Θ has elements which are all càdlàg, and W is a vector standard Brownian motion defined on a filtered probability space.

From the representation in (1.5), define the spot *covariance matrix* as

$$\Sigma(t) = \Theta(t)\Theta(t)^T$$

where $\Sigma_{ij}(t)$ will denote the cross-volatility between the ith and the jth price process, with $i,j = 1,\ldots,p$. We are interested in estimating this quantity. In the previous sections we have seen that the following identity holds almost surely

$$[X_i, X_j]_t = \int_0^t \Sigma_{ij}(s)ds, \tag{1.6}$$

where the quantity on the right-hand side is called *integrated covariance*. This result implies that

$$\Sigma_{ij}(t) = \frac{d[X_i, X_j]_t}{dt},$$

for every $1 \leq i,j \leq p$. However, high-frequency financial data are inherently non-synchronous, in the sense that usually two transaction prices are not recorded at the same time, while the above differentiation procedure requires to have observations available for both X_i and X_j at each point on the grid. The problem can be successfully overcome using the Fourier procedure introduce by Malliavin and Mancino (2002) that allows to estimate $\Sigma(t)$ through time series of data as their observed on the market, without any prior manipulation. This is possible according to the fact that the method is based on integration, rather than differentiation, of financial returns. The idea is to combine classical harmonic analysis with stochastic calculus to derive the Fourier coefficients of Σ_{ij} from the Fourier coefficients of the return process dX_i defined as

$$a_0(dX_i) = \frac{1}{2\pi}\int_0^{2\pi} dX_i \qquad\qquad a_0(\Sigma_{ij}) = \frac{1}{2\pi}\int_0^{2\pi} \Sigma_{ij}(t)dt$$

$$a_k(dX_i) = \frac{1}{\pi}\int_0^{2\pi} \cos(kt)dX_i \qquad\qquad a_k(\Sigma_{ij}) = \frac{1}{\pi}\int_0^{2\pi} \Sigma_{ij}(t)\cos(kt)dt$$

$$b_k(dX_i) = \frac{1}{\pi}\int_0^{2\pi} \sin(kt)dX_i, \qquad\qquad b_k(\Sigma_{ij}) = \frac{1}{\pi}\int_0^{2\pi} \Sigma_{ij}(t)\sin(kt)dt$$

for $k = 1,\ldots,N$. Note that the coefficients of price are not Fourier coefficients in the usual sense but proper stochastic integrals. Also note that, by rescaling the unit of time, we can easily move from the interval $[0,T]$ to $[0,2\pi]$. The volatility matrix $\Sigma(t)$ is then reconstructed point wise on the fixed window $[0,2\pi]$ by the classical Fourier inversion formula

$$\hat{\Sigma}_{N,M}^{ij}(t) = a_0(\Sigma_{ij}) + \lim_{M\to\infty}\sum_{q=1}^{M} \left[a_q(\Sigma_{ij})\cos(qt) + b_q(\Sigma_{ij})\sin(qt)\right]. \tag{1.7}$$

Due to the presence of the limit function, the above expression cannot be calculated in practice and is substituted by the partial Fourier series

$$\hat{\Sigma}^{ij}_{N,M}(t) = a_0(\Sigma_{ij}) + \sum_{q=1}^{M} \left(1 - \frac{q}{M}\right) \left[a_q(\Sigma_{ij})\cos(qt) + b_q(\Sigma_{ij})\sin(qt)\right]. \quad (1.8)$$

Note that now the coefficients $a_q(\Sigma_{ij})$ and $b_q(\Sigma_{ij})$ are weighted by the Fèjer window $w_q = \left(1 - \frac{q}{M}\right)$, which is responsible for reducing the contribution of the high modes in the recovering process. Alternatively, we can employ a modified version of the Lanczos sigma factor proposed in Malliavin and Thalmaier (2005) and defined as

$$\varphi(x) = \frac{\sin^2(x)}{x^2}, \quad \varphi(0) = 0. \quad (1.9)$$

In particular, Eq. (1.8) becomes

$$\hat{\Sigma}^{ij}_{N,M}(t) = a_0(\Sigma_{ij}) + \sum_{q=0}^{M} \varphi(\delta q)\left[a_q(\Sigma_{ij})\cos(qt) + b_q(\Sigma_{ij})\sin(qt)\right]. \quad (1.10)$$

The parameter δ can be interpreted as the time scale at which we want to reconstruct the volatility trajectory: the higher is δ, the smoother is the estimated path but at the price of a less detailed reconstruction.

1.3.1 The univariate case

For ease of exposition, we will first present the main result in the work by Malliavin and Mancino (2002) for the univariate case. In particular, we will show how to derive Fourier coefficients of the volatility from the coefficients of price. The coefficients of the cross-volatility can then be obtained by polarization as we will see shortly. Set $\Sigma_{i,i} = \sigma^2$ and recall that $L^2([0, 2\pi])$ represents the class of functions that are square integrable over $[0, 2\pi]$.

Theorem 1.1. *Consider a process satisfying (1.5) with $p = 1$. For a fixed integer $n_0 > 0$ and $1 \leq q \leq M$, the Fourier coefficients of the volatility process are given by*

$$a_0(\sigma^2) = \lim_{N\to\infty} \frac{\pi}{N+1-n_0} \sum_{k=n_0}^{N} a_k^2(dX) = \lim_{N\to\infty} \frac{\pi}{N+1-n_0} \sum_{k=n_0}^{N} b_k^2(dX) \quad (1.11)$$

$$a_q(\sigma^2) = \lim_{N\to\infty} \frac{2\pi}{N+1-n_0} \sum_{k=n_0}^{N-q} a_k(dX)a_{k+q}(dX) = \lim_{N\to\infty} \frac{2\pi}{N+1-n_0} \sum_{k=n_0}^{N-q} b_k(dX)b_{k+q}(dX) \quad (1.12)$$

$$b_q(\sigma^2) = \lim_{N\to\infty} \frac{2\pi}{N+1-n_0} \sum_{k=n_0}^{N-q} a_k(dX)b_{k+q}(dX) = -\lim_{N\to\infty} \frac{2\pi}{N+1-n_0} \sum_{k=n_0}^{N-q} b_k(dX)a_{k+q}(dX),$$

$$(1.13)$$

where all the limits are attained in probability.

The proof of the theorem as given in the original paper is not completely correct, the main problem laying in the statement

$$E\left[\left(\frac{1}{N+1-n_0}\sum_{k=1}^{N-q}G_kG_{k+q}\right)^2\right]=\frac{1}{(N+1-n_0)^2}\sum_{k_1=0}^{N-q}\sum_{k_2=0}^{N-q}E(G_{k_1}^2G_{k_2+q}^2),$$

where $G_k := a_k(dX)$, which is wrong because

$$E\left[\left(\frac{1}{N+1-n_0}\sum_{k=1}^{N-q}G_kG_{k+q}\right)^2\right]=\frac{1}{(N+1-n_0)^2}\sum_{k_1=0}^{N-q}\sum_{k_2=0}^{N-q}E(G_{k_1}G_{k_1+q}G_{k_2}G_{k_2+q}).$$

In addition, the summation indexes on both sides should start from n_0. We will now show that it is still possible to achieve the final convergence result.

Proof of Theorem 1.1. We restrict our attention to the case where $\sigma(t)$ is a deterministic function of time. The extension to the stochastic case can be obtained as in the proof of Theorem 3.1 in Malliavin and Mancino (2002).

Choose $k, h \in \mathbb{N}$ such that $k > h > 1$ and set $G_k := a_k(dX)$. Then, by the Itô isometry,

$$\begin{aligned}E[G_kG_h] &= E\left[\frac{1}{\pi}\int_0^{2\pi}\cos(kt)dX(t)\cdot\frac{1}{\pi}\int_0^{2\pi}\cos(ht)dX(t)\right]\\&= E\left[\frac{1}{\pi^2}\int_0^{2\pi}\cos(kt)\sigma(t)dW(t)\cdot\int_0^{2\pi}\cos(ht)\sigma(t)dW(t)\right]\\&= \frac{1}{\pi^2}\int_0^{2\pi}\sigma^2(t)\cos(kt)\cos(ht)dt.\end{aligned}$$

Using the identity

$$2\cos(kt)\cos(ht)=\cos[(k-h)t]+\cos[(k+h)t]$$

we obtain

$$E[G_kG_h]=\frac{1}{2\pi}[a_{|k-h|}(\sigma^2)+a_{k+h}(\sigma^2)]. \tag{1.14}$$

Moreover, given that $\int_0^{2\pi}\cos^2(kt)dt=2\pi$ and $\int_0^{2\pi}\sin^2(kt)dt=2\pi$, by Eq. (1.8) we can calculate the energy identity

$$||\sigma^2||_{L^2}^2 := \int_0^{2\pi}\sigma^2(t)dt=2\pi\sum_{k=0}^{+\infty}[a_k^2(\sigma^2)+b_k^2(\sigma^2)].$$

For ease of notation, in the following we will set $\bar{N} := N - q$.

Given an integer $n_0 > 0$ and for $q \in \mathbb{N}$ define

$$U_N^q = \frac{1}{N+1-n_0} \sum_{k=n_0}^{\bar{N}} G_k G_{k+q}.$$

By means of Eq. (1.14) we get

$$E[U_N^q] = \frac{1}{2\pi} \frac{1}{N+1-n_0} \sum_{k=n_0}^{\bar{N}} [a_q(\sigma^2) + a_{2k+q}(\sigma^2)] = \frac{1}{2\pi} a_q(\sigma^2) + R_N,$$

where

$$|R_N| = \frac{1}{2\pi} \frac{1}{N+1-n_0} \left| \sum_{k=n_0}^{\bar{N}} a_{2k+q}(\sigma^2) \right| \leq \frac{1}{\sqrt{N+1-n_0}} ||\sigma^2||_{L^2}.$$

For the last passage use the Cauchy-Schwartz inequality and the fact that

$$\sum_k a_k^2(\sigma^2) = ||\sigma^2||_{L^2}^2 + \sum_k b_k^2(\sigma^2) \leq ||\sigma^2||_{L^2}^2,$$

being $b_k^2(\sigma^2)$ a positive quantity. Therefore, $R_N \to 0$ and

$$a_q(\sigma^2) = 2\pi \lim_{N\to\infty} E[U_N^q].$$

We now want to prove that $a_q(\sigma^2) = 2\pi \lim_{N\to\infty} U_N^q$. To this purpose we first compute

$$E[U_N^q]^2 = \frac{1}{(N+1-n_0)^2} \sum_{n_0 \leq k_1, k_2 \leq \bar{N}} E[a_{k_1}(dX)a_{k_1+q}(dX)] E[a_{k_2}(dX)a_{k_2+q}(dX)].$$

Since $a_k(dX)$ is a zero-mean Gaussian random variable, we can then use the well-known formula for the product of four Gaussian variables[1] to obtain

$$E[(U_N^q)^2] = \frac{1}{(N+1-n_0)^2} \sum_{n_0 \leq k_1, k_2 \leq \bar{N}} E[a_{k_1}(dX)a_{k_1+q}(dX)a_{k_2}(dX)a_{k_2+q}(dX)]$$

$$= \frac{1}{(N+1-n_0)^2} \sum_{n_0 \leq k_1, k_2 \leq \bar{N}} \big\{ E[a_{k_1}(dX)a_{k_1+q}(dX)] E[a_{k_2}(dX)a_{k_2+q}(dX)] +$$

$$+ E[a_{k_1}(dX)a_{k_2}(dX)] E[a_{k_1+q}(dX)a_{k_2+q}(dX)] +$$

$$+ E[a_{k_1}(dX)a_{k_2+q}(dX)] E[a_{k_1+q}(dX)a_{k_2}(dX)] \big\}.$$

---

[1] Given four Gaussian random variables $x_i, i = 1, 2, 3, 4$, the expectation of the product is given by

$$E[x_1 x_2 x_3 x_4] = E[x_1 x_2] E[x_3 x_4] + E[x_1 x_3] E[x_2 x_4] + E[x_1 x_4] E[x_2 x_3] - 2E[x_1]E[x_2]E[x_3]E[x_4].$$

Using again Eq. (1.14), we can now calculate

$$E[(U_N^q - E[U_N^q])^2] = c \sum_{n_0 \leq k_1, k_2 \leq \bar{N}} \{[a_{k_1+k_2}(\sigma^2) + a_{|k_1-k_2|}(\sigma^2)][a_{k_1+k_2+2q}(\sigma^2) + a_{|k_1-k_2|}(\sigma^2)]+$$
$$+ [a_{k_1+k_2+q}(\sigma^2) + a_{|k_1-k_2-q|}(\sigma^2)][a_{k_1+k_2+q}(\sigma^2) + a_{|k_1-k_2+q|}(\sigma^2)]\},$$

where $c = \frac{1}{4\pi^2} \frac{1}{(N+1-n_0)^2}$. Finally, we use Cauchy-Schwartz to show that

$$E[(U_N^q - E[U_N^q])^2] \leq$$
$$\leq c \left[\left(\sum_{n_0 \leq k_1, k_2 \leq \bar{N}} (a_{k_1+k_2}(\sigma^2) + a_{|k_1-k_2|}(\sigma^2))^2 \sum_{n_0 \leq k_1, k_2 \leq \bar{N}} (a_{k_1+k_2+2q}(\sigma^2) + a_{|k_1-k_2|}(\sigma^2))^2 \right)^{\frac{1}{2}} + \right.$$
$$\left. + \left(\sum_{n_0 \leq k_1, k_2 \leq \bar{N}} (a_{k_1+k_2+q}(\sigma^2) + a_{|k_1-k_2-q|}(\sigma^2))^2 \sum_{n_0 \leq k_1, k_2 \leq \bar{N}} (a_{k_1+k_2+q}(\sigma^2) + a_{|k_1-k_2+q|}(\sigma^2))^2 \right)^{\frac{1}{2}} \right] \leq$$
$$\leq \frac{1}{4\pi^2} \frac{1}{(N+1-n_0)^2} ||\sigma^2||_{L^2}^2.$$

The above inequality implies convergence in L^2, and then in probability. This proves the first part of Eq. (1.12). For the second part, it is enough to repeat the proof with $G_k = b_k(dX)$ and use the trigonometric formula

$$2\sin(kt)\sin(ht) = \cos[(k-h)t] - \cos[(k+h)t].$$

To derive Eq. (1.13), we need to replace $a_k(dX), a_h(dX)$ with $a_k(dX), b_h(dX)$ and compute

$$E[a_k(dX), b_h(dX)] = \frac{1}{\pi^2} \int_0^{2\pi} \sigma^2(t)\cos(kt)\sin(ht)dt.$$

Then, by the identity

$$2\cos(kt)\sin(ht) = \sin[(k-h)t] + \sin[(k+h)t],$$

we get

$$E[a_k(dX), b_h(dX)] = \frac{1}{2\pi}[a_{|k-h|}(\sigma^2) + a_{k+h}(\sigma^2)].$$

Proceeding as above, it is than necessary to calculate the expected value of

$$V_N^q = \frac{1}{N+1-n_0} \sum_{k=n_0}^{\bar{N}} a_k(dX)b_{k+q}(dX)$$

and

$$W_N^q = \frac{1}{N+1-n_0} \sum_{k=n_0}^{\bar{N}} b_k(dX)a_{k+q}(dX).$$

Finally, Eq. (1.11) follows from

$$E[a_k^2(dX)] = E[b_k^2(dX)] = \frac{1}{2\pi}[a_0(\sigma^2) + a_{2k}(\sigma^2)].$$

This completes the proof. □

Corollary 1.2. *For $1 \leq q \leq M$, the Fourier coefficients of $\sigma^2(t)$ are calculated in $L^2([0, 2\pi])$ as*

$$a_0(\sigma^2) = \lim_{N \to \infty} \frac{\pi}{N+1-n_0} \sum_{k=n_0}^{N} \frac{1}{2}\left[a_k^2(dX) + b_k^2(dX)\right] \qquad (1.15)$$

$$a_q(\sigma^2) = \lim_{N \to \infty} \frac{\pi}{N+1-n_0} \sum_{k=n_0}^{N-q} \left[a_k(dX)a_{k+q}(dX) + b_k(dX)b_{k+q}(dX)\right] \qquad (1.16)$$

$$b_q(\sigma^2) = \lim_{N \to \infty} \frac{\pi}{N+1-n_0} \sum_{k=n_0}^{N-q} \left[a_k(dX)b_{k+q}(dX) - b_k(dX)a_{k+q}(dX)\right], \qquad (1.17)$$

Proof. The proof easily follows from Theorem 1.1. □

In Chapter 2, Section 2.4.2, we will study in detail the role played in the estimation procedure by the parameters N and M, namely, the number of price and volatility coefficients.

1.3.2 The multivariate case

The multivariate case is a straightforward extension of the previous analysis. In particular, the main result can be simply obtained by polarization as follow.

Theorem 1.3. *Consider a process satisfying (1.5). For a fixed integer $n_0 > 0$, the Fourier coefficients of the cross-volatility process are defined as*

$$a_0(\Sigma_{ij}) = \lim_{N \to \infty} \frac{\pi}{N+1-n_0} \sum_{k=n_0}^{N} \frac{1}{2}\left[a_k(dX_i)a_k(dX_j) + b_k(dX_i)b_k(dX_j)\right] \qquad (1.18)$$

$$a_q(\Sigma_{ij}) = \lim_{N \to \infty} \frac{\pi}{N+1-n_0} \sum_{k=n_0}^{N-q} \left[a_k(dX_i)a_{k+q}(dX_j) + b_k(dX_i)b_{k+q}(dX_j)\right] \qquad (1.19)$$

$$b_q(\Sigma_{ij}) = \lim_{N \to \infty} \frac{\pi}{N+1-n_0} \sum_{k=n_0}^{N-q} \left[a_k(dX_i)b_{k+q}(dX_j) - b_k(dX_i)a_{k+q}(dX_j)\right], \qquad (1.20)$$

where the limits are attained in probability.

Proof. It is enough to show how to derive coefficient $a_0(\Sigma_{ij})$. Using the polarization formula

given in Eq. (1.4) and introduced in Section 1.2

$$[X_i, X_j]_t = \frac{1}{4}\left\{[X_i + X_j, X_i + X_j] - [X_i - X_j, X_i - X_j]\right\}$$

we simply substitute $a_k^2(dX)$ in Eq. (1.15) with

$$\frac{1}{4}\left\{\left[a_k(dX_i) + a_k(dX_j)\right]^2 - \left[a_k(dX_i) - a_k(dX_j)\right]^2\right\} = a_k(dX_i)a_k(dX_j),$$

according to the fact that $a_k(dX_i \pm dX_j) = a_k(dX_i) \pm a_K(dX_j)$. $\qquad\square$

1.4 The complex form of the Fourier method

The Fourier series as given in Eq. (1.8) can be expressed in an algebraically simpler form involving complex exponentials. To relate the trigonometric and exponential functions, we make use of the well-known Euler's formula given by

$$e^{iqt} = \cos(qt) + i\sin(qt).$$

It is an immediate consequence of this formula that

$$\cos(qt) = \frac{e^{iqt} + e^{-iqt}}{2} \qquad\qquad \sin(qt) = \frac{e^{iqt} - e^{-iqt}}{2i}.$$

Substituting these expressions in (1.8), adapted to the univariate case, we obtain

$$\hat{\sigma}_{N,M}^2(t) \approx a_0 + \sum_{q=1}^{M}\left(1 - \frac{q}{M}\right)\left[\frac{a_q - ib_q}{2}e^{iqt} + \frac{a_q + ib_q}{2}e^{-iqt}\right],$$

where $a_q := a_q(\sigma^2)$ and $b_q := b_q(\sigma^2)$. If we set

$$c_0 = a_0 \qquad c_q = \frac{a_q - ib_q}{2} \qquad c_{-q} = \frac{a_q + ib_q}{2},$$

with $c_q := c_q(\sigma^2)$, then

$$\hat{\sigma}_{N,M}^2(t) \approx c_0 + \sum_{q=1}^{M}\left(1 - \frac{q}{M}\right)\left[c_q e^{iqt} + c_{-q}e^{-iqt}\right] = \sum_{|q|\leq M}\left(1 - \frac{q}{M}\right)c_q e^{iqt}.$$

This is the *complex form* of the Fourier series of $\sigma^2(t)$ and the coefficients c_q are called the *complex Fourier coefficients* of the series. It is straightforward to see that

$$c_q(\sigma^2) = \frac{1}{2\pi}\int_0^{2\pi}\sigma^2(t)e^{-iqt}dt, \qquad q = 0, \pm 1, \ldots, \pm M.$$

The last expression is analogous to the *Fourier transform* of $\sigma^2(t)$ on the real line $\mathbb{R}$. In their recent paper, Malliavin and Mancino (2009) use this complex form to define the asymptotic properties of the Fourier estimator. In particular, they prove that, in absence of microstructure noise, the estimator is consistent and also derive a weak convergence result. We will now present these properties in the univariate case for ease of comparison with the limiting theory developed in Chapter 4. The results are stated without proof.

Suppose that the process $X(t)$ is observed at instants $0 = t_0 < t_1 < \ldots < t_n = 2\pi$ over the interval $[0, T]$; the resulting tick-by-tick prices are not necessarily equally spaced in time. Moreover, define $\rho(n) := \max_{0 \leq h \leq n-1} |t_{h+1} - t_h|$ such that $\rho(n) \to 0$, as $n \to \infty$. For $|k| \leq N$, let

$$c_k(dX) = \frac{1}{2\pi} \sum_{j=0}^{n-1} e^{-ikt} \Delta X_j,$$

where $\Delta X_j = X(t_j) - X(t_{j-1})$, and consider the convolution product

$$\alpha_q(N) = \frac{2\pi}{2N+1} \sum_{|k| \leq N} c_k(dX) c_{q-k}(dX).$$

Finally, define the random function

$$\hat{\sigma}_{N,M}^2(t) = \sum_{|q| \leq M} \left(1 - \frac{q}{M}\right) \alpha_q(N) e^{iqt}. \tag{1.21}$$

Theorem 1.4. *For every $q \in \mathbb{N}$, the following convergence in probability holds*

$$\lim_{n,N \to \infty} \alpha_q(N) = c_q(\sigma^2).$$

Moreover, the estimator of instantaneous volatility $\sigma^2(t)$ is consistent uniformly in time

$$\lim_{n,N \to \infty} \sup_{t \in [0,2\pi]} |\hat{\sigma}_{N,M}^2(t) - \sigma^2(t)| = 0.$$

The limiting distribution result is based on the following set of assumptions defined in Mykland and Zhang (2006) and necessary to deal with unevenly sampled series of data

(i) $\rho(n) \to 0$ and $\Delta_n \rho(n) = O(1)$, where $\Delta_n = \frac{2\pi}{n}$

(ii) $H_n(t) = \dfrac{\sum_{t_{j+1,n} \leq t} (t_{j+1,n} - t_{j,n})^2}{\Delta_n} \to H(t)$ as $n \to \infty$

(iii) $H(t)$ is continuously differentiable.

Theorem 1.5. *Let $\hat{\sigma}_{N,M}^2(t)$ be defined as in Eq. (1.21) and assume that conditions (i)-(iii) are satisfied. Then, for $n, N, M \to \infty$ and any Lipschitz continuous function $h(t)$ of order $\alpha > \frac{1}{2}$ with compact support in $(0, 2\pi)$, and provided that $\rho(n)N^{2\alpha} \to \infty$ we have*

$$\frac{1}{\sqrt{\rho(n)}} \int_0^{2\pi} h(t)\big[\hat{\sigma}_{N,M}^2(t) - \sigma^2(t)\big] dt \xrightarrow{\mathcal{L}} \mathbf{MN}\left(0, 2\int_0^{2\pi} H'(t)h^2(t)\sigma^4(t)dt\right),$$

where the above convergence is in law.

In order extend the theorem to the multivariate case, suppose that the log-price processes X_1 and X_2 are discretized on two distinct irregular grids $0 = t_0^1 < t_1^1 < \ldots < t_{n_1}^1 = 2\pi$ and $0 = t_0^2 < t_1^2 < \ldots < t_{n_2}^2 = 2\pi$ respectively. Then, $\rho^s(n) := \max\limits_{0 \le h \le n-1} |t_{h+1}^s - t_h^s|$, $s = 1, 2$, and assume that $\rho(n) := \rho^1(n) \vee \rho^1(n) \to 0$ as $n \to \infty$. Finally, define the rescaled Dirichelet kernel as

$$D_N(t) := \frac{1}{2N+1} \sum_{|k| \le N} e^{ikt}.$$

Theorem 1.6. *Under the hypothesis of Theorem 1.5, the following result holds*

$$\frac{1}{\sqrt{\rho(n)}} \int_0^{2\pi} h(t)\big[\hat{\Sigma}_{N,M}^{12}(t) - \Sigma^{12}(t)\big] dt =$$

$$= \frac{1}{\sqrt{\rho(n)}} \sum_{i=0}^{n_1-1} \sum_{j=0}^{n_2-1} h_M(t_j^2) D_N(t_j^2 - t_i^1) \left(\int_{t_{i-1}^1 \vee t_{j-1}^2}^{t_i^1 \wedge t_j^2} \int_{t_{i-1}^1 \vee t_{j-1}^2}^{t} dX_1(s) dX_2(s) \right.$$

$$\left. + \int_{t_{i-1}^1 \vee t_{j-1}^2}^{t_i^1 \wedge t_j^2} \int_{t_{i-1}^1 \vee t_{j-1}^2}^{t} dX_2(s) dX_1(s) \right) + o_p(1),$$

where $o_p(\cdot)$ is defined for $n \to \infty$ and

$$h_M(t) = \sum_{|q| \le M} \left(1 - \frac{q}{M}\right) c_q(h) e^{iqt}$$

with $c_q(h)$ denoting the q-th Fourier coefficient of the function h.

We have used the notation $\mathbf{MN}(0, V)$ to denote a mixed normal distribution with stochastic variance V.

Spot volatility estimation via Fourier method

In this chapter, we will discuss some practical issues related to the implementation and application of the Fourier methodology previously introduced. We recall that the procedure employs the observations in their original form and allows to recover the time evolution of a multivariate volatility process over a fixed window. We first show how to apply the Fast Fourier Transform algorithm to calculate the Fourier coefficients of price, obtaining a remarkable improvement in terms of computational efficiency. We then contribute to clarify two important aspects in the application of the method, namely, the choice of the number of price and volatility coefficients, and the choice of the scale at which to reconstruct the volatility trajectory. We show how to simultaneously estimate these quantities using the Differential Evolution optimization algorithm introduced by Storn and Price (1997). Finally, we perform a detailed analysis on the impact that different levels of inhomogeneity in the data and microstructure noise might have on the above parameters, and on the performance of the Fourier estimator as a consequence. This is a relevant step in the estimation procedure as applied to recover the temporal dynamics of a stochastic process, and it is here investigated for the first time. All the results in the chapter are derived for an univariate process, but can be easily extended to the multivariate case.

2.1 Preliminaries

In the Fourier method proposed by Malliavin and Mancino (2002), the instantaneous volatility process $\sigma^2(t)$ is expanded into trigonometric polynomials whose coefficients depend on the observed log returns. In order to apply spectral analysis to random processes, $X(t)$ is

assumed to be a periodic functions of period T on the interval $[0, T]$. A function $f(x)$ is called *periodic* if there exists a constant $\varphi > 0$ such that $f(x + \varphi) = f(x)$, for every x in the domain of definition of $f(x)$. Therefore, a periodic function repeats itself at regular intervals, and the period φ is the time between successive repetitions. The functions $\cos(t)$ and $\sin(t)$ involved into the estimation procedure are periodic with period 2π, and any linear combination of them will also be periodic, i.e. if we consider a generic function of the form

$$f(t) = \sum_{k=0}^{\infty} [a_k \cos(kt) + b_k \sin(kt)],$$

where a_n and b_n are the standard Fourier coefficients, then $f(t)$ must be a periodic function of period 2π itself, whatever values we assign to a_n and b_n. To model a process with period T rather than 2π, it is necessary to rescale the unit of time and move from the frequency domain on $[0, 2\pi]$ to the time domain on $[0, T]$. If we assume that prices are observed on the market at times $0 = t_0 \leq t_1 \leq \ldots \leq t_n = T$, this can be achieved by simply taking

$$\tau_j = \frac{2\pi(t_j - t_0)}{t_n - t_0} = \frac{2\pi t_j}{T},$$

where $\frac{2\pi}{T}$ is called the *angular frequency* and $\frac{1}{T}$ is the *fundamental frequency*.

2.2 A faster way to calculate the Fourier coefficients of price

It is well-recognized that the Fourier algorithm is not numerically efficient, both in terms of memory requirement and computational time. In this section we will show that it can be substantially improved by applying a Fast Fourier Trasform to the calculation of the coefficients of price.

2.2.1 Preliminary results

The Fourier coefficients of price introduced in Chapter 1, Section 1.3, now become

$$a_0(dX) = \frac{1}{T} \int_0^T dX(t) \tag{2.1}$$

$$a_k(dX) = \frac{2}{T} \int_0^T \cos(\theta_k t) dX(t) \tag{2.2}$$

$$b_k(dX) = \frac{2}{T} \int_0^T \sin(\theta_k t) dX(t), \qquad k = 1, \ldots, N \tag{2.3}$$

where $\theta_k = \frac{2\pi k}{T}$ are the angular frequencies.

The implementation is carried out by computing the integrals via integration by parts as follow

$$
\int_0^T \cos(\theta_k t) dX(t) = X(T) - X(0) + \theta_k \int_0^T \sin(\theta_k t) X(t) dt
$$
$$
\int_0^T \sin(\theta_k t) dX(t) = -\theta_k \int_0^T \cos(\theta_k t) X(t) dt
$$

We now take the process $X(t)$ to be constant and left continuous on the interval $[t_{j-1}, t_j]$, i.e. $X(t) = X(t_{j-1})$ for $t \in [t_{j-1}, t_j]$ and $j = 1, \ldots, n$. This is a reasonable assumption because prices do not evolve continuisly on the market, but are instead recorded at discrete times, as already stated. Then, the first integral above becomes

$$
\theta_k \int_{t_{j-1}}^{t_j} \sin(\theta_k t) X(t) dt = X(t_j) \theta_k \int_{t_{j-1}}^{t_i} \sin(\theta_k t) dt = X(t_{j-1}) \big[\cos(\theta_k t_{j-1}) - \cos(\theta_k t_j) \big],
$$

and similar arguments holds for the second one. It follows that the coefficients can be calculated approximately as

$$
a_0(dX) = \frac{X(T) - X(0)}{T} \tag{2.4}
$$
$$
a_k(dX) = \frac{X(T) - X(0)}{T/2} + \frac{2}{T} \sum_{j=1}^n X(t_{j-1}) \left[\cos(\theta_k t_{j-1}) - \cos(\theta_k t_j) \right] \tag{2.5}
$$
$$
b_k(dX) = \frac{2}{T} \sum_{j=1}^n X(t_{j-1}) \left[\sin(\theta_k t_{j-1}) - \sin(\theta_k t_j) \right]. \tag{2.6}
$$

Although very simple, the following result represents an important step in our effort to computationally improve the Fourier estimator.

Proposition 2.1. *Equations (2.5) and (2.6) can be equivalently written as*

$$
a_k(dX) = \frac{2}{T} \sum_{j=1}^{n-1} \cos(\theta_k t_j) \Delta X_j \tag{2.7}
$$
$$
b_k(dX) = \frac{2}{T} \sum_{j=1}^{n-1} \sin(\theta_k t_j) \Delta X_j, \qquad k = 1, \ldots, N \tag{2.8}
$$

where $\Delta X_j = X(t_j) - X(t_{j-1})$ is the log-price return observed at time t_j.

Proof. It is enough to show how to derive Eq. (2.8). To this end, let us consider the simple

case where $n = 4$. The extension to a general n easily follows. We have

$$
\begin{aligned}
b_k(dX) &= \frac{2}{T} \sum_{j=1}^{4} X(t_{j-1}) \big[\sin(\theta_k t_{j-1}) - \sin(\theta_k t_j) \big] \\
&= \frac{2}{T} \big[X(t_0)\sin(\theta_k t_0) - X(t_0)\sin(\theta_k t_1) + X(t_1)\sin(\theta_k t_1) - X(t_1)\sin(\theta_k t_2) + \\
&\qquad + X(t_2)\sin(\theta_k t_2) - X(t_2)\sin(\theta_k t_3) + X(t_3)\sin(\theta_k t_3) - X(t_3)\sin(\theta_k t_4) \big] \\
&= \frac{2}{T} \big[X(t_0)\sin(\theta_k t_0) + \sin(\theta_k t_1)(X(t_1) - X(t_0)) + \sin(\theta_k t_2)(X(t_2) - X(t_1)) + \\
&\qquad + \sin(\theta_k t_3)(X(t_3) - X(t_2)) - X(t_3)\sin(\theta_k t_4) \big] \\
&= \frac{2}{T} \big[X(t_0)\sin(\theta_k t_0) - X(t_3)\sin(\theta_k t_4) \big] + \frac{1}{\pi} \big[\sin(\theta_k t_1)(X(t_1) - X(t_0)) + \\
&\qquad + \sin(\theta_k t_2)(X(t_2) - X(t_1)) + \sin(\theta_k t_3)(X(t_3) - X(t_2)) \big] \\
&= \frac{2}{T} \sum_{j=1}^{3} \sin(\theta_k t_j) \Delta X_j,
\end{aligned}
$$

where the periodicity of the trigonometric functions implies

$$
\frac{2}{T} \big[X(t_0)\sin(\theta_k t_0) - X(t_3)\sin(\theta_k t_4) \big] = 0
$$

given that $t_0 = 0$ and $t_4 = T$ for a sample of length $n = 4$. $\qquad\square$

Corollary 2.2. *The following relation holds*

$$
a_k(dX) - i b_k(dX) = \frac{2}{T} \sum_{j=1}^{n-1} \Delta X_j \, e^{-i\theta_k t_j}, \qquad k = 1, \dots, N \tag{2.9}
$$

where the quantity on the left-hand side is a complex number and $i = \sqrt{-1}$.

Proof. Given Eq. (2.7) and (2.8), we easily obtain

$$
\begin{aligned}
a_k(dX) - i b_k(dX) &= \frac{2}{T} \left[\sum_{j=1}^{n-1} \cos(\theta_k t_j) \Delta X_j - i \sum_{j=1}^{n-1} \sin(\theta_k t_j) \Delta X_j \right] \\
&= \frac{2}{T} \sum_{j=1}^{n-1} \Delta X_j \big[\cos(\theta_k t_j) - i \sin(\theta_k t_j) \big] \\
&= \frac{2}{T} \sum_{j=1}^{n-1} \Delta X_j \, e^{-i\theta_k t_j},
\end{aligned}
$$

where the last equality follows from the Euler's identity $e^{-i\phi} = \cos\phi - \sin\phi$. $\qquad\square$

The summation in Eq. (2.9) is called the *Discrete Fourier Transform (DFT)*, and it maps the $n-1$ returns ΔX_j defined on the time domain into $n-1$ complex numbers defined on the frequency domain. The transform requires a discrete number of points as input data, which in our case are obtained by sampling the continuous process $X(t)$. In addition, the points must be sampled at evenly spaced intervals in time, i.e. $t_j = j\Delta$, where $\Delta = \frac{T}{N}$ is the sampling interval, and therefore

$$\frac{2\pi(t_j - t_0)}{T} = j\frac{2\pi}{N}.$$

This requirement seems to be in contrast with the peculiar property of the Fourier estimator to employ data as they are observed on the market. However, in the next section we will provide evidence that the estimator is not affected by the imputation procedure used to construct equally spaced prices, if this is carefully chosen. From the above result, it is clear that the Fourier coefficients of price are simply the real and imaginary part of the complex numbers mapped by the DFT from a finite sequence of intraday returns. We will see now how they can be calculated efficiently using Fast Fourier Trasform.

2.2.2   The Fast Fourier Transform algorithm

The Discrete Fourier Transform requires a considerable amount of time to be computed using the definition in Eq. (2.9): there are N outputs and for each output a sum of N terms is needed. The complexity of the calculation becomes more clear if we define the root of unity $w = e^{-i\frac{2\pi}{N}}$, traditionally called *twiddle factor*, and rewrite the DTF as

$$y_k = \sum_{j=1}^{N} x_j w^{jk}, \qquad k = 1, \ldots, N \tag{2.10}$$

where we have also set $y_k = a_k(dX) - ib_k(dX)$ and $x_j = \Delta X_j$ to simplify the notation. The equation can be expressed in matrix form as follows

$$\begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_N \end{bmatrix} = \begin{bmatrix} w & w^2 & \ldots & w^N \\ w^2 & w^4 & \ldots & w^{2N} \\ \vdots & \vdots & & \vdots \\ w^N & w^{2N} & \ldots & w^{N^2} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_N \end{bmatrix}.$$

Equivalently, we can use the matrix notation and write

$$\mathbf{y} = W\mathbf{x},$$

where W is a matrix of $N \times N$ complex numbers. Note that W is symmetric, i.e. $W = W^T$.

For instance, if $N = 4$

$$W = \begin{bmatrix} w & w^2 & w^3 & w^4 \\ w^2 & w^4 & w^6 & w^8 \\ w^3 & w^6 & w^9 & w^{12} \\ w^4 & w^8 & w^{12} & w^{16} \end{bmatrix} = \begin{bmatrix} w & w^2 & w^3 & 1 \\ w^2 & 1 & w^2 & 1 \\ w^3 & w^2 & w & 1 \\ 1 & 1 & 1 & 1 \end{bmatrix} = \begin{bmatrix} -i & -1 & i & 1 \\ -1 & 1 & -1 & 1 \\ i & -1 & -i & 1 \\ 1 & 1 & 1 & 1 \end{bmatrix},$$

since $w = \cos\left(\frac{2\pi}{4}\right) - i\sin\left(\frac{2\pi}{4}\right) = -i$ and $w^4 = 1$, the latter obtained using the property $i^2 = -1$. It should be now easier to see that the above matrix-vector product involves N^2 multiplications and N^2 additions. Hence, for an input sequence of N intraday returns, computing the DTF directly would require a number of arithmetic operations proportional to N^2. This number can be dramatically reduced to be proportional to $N\log_2 N$ by applying the *Fast Fourier Trasform (FFT)*. The difference between the two speed factors is extremely large: with $n = 10^6$ data points, the FFT can perform the DFT in, roughly, 30 seconds of CPU time against the 2 weeks of CPU time of the original DFT formula on a microsecond time cycle computer. The algorithm became very popular after the work of Cooley and Tukey (1965), however, it was later discovered that the German mathematician Carl Friedrich Gauss essentially proposed a very similar method back in 1805 (the paper was published posthumously in 1866). Here we will describe the simplest and most common form of the Cooley and Tukey algorithm, namely, the *Radix-2 FFT*. The method recursively breaks down the original DFT into smaller DFTs of size $\frac{N}{2}$ at each step, and makes use of intermediate calculations to reduce the overall computational time. In particular, the DFT is first rearranged into two parts: one is obtained from the even-numbered points of the original N, the other from the odd-numbered points as follow

$$
\begin{aligned}
y_k &= \sum_{j=1}^{N} x_j w^{jk} \\
&= \sum_{j=1}^{\frac{N}{2}} x_{2j} w^{2jk} + \sum_{j=1}^{\frac{N}{2}} x_{2j+1} w^{(2j+1)k} \\
&= \sum_{j=1}^{\frac{N}{2}} x_{2j} (w^2)^{jk} + w^k \sum_{j=1}^{\frac{N}{2}} x_{2j+1} (w^2)^{jk} \\
&= \sum_{j=1}^{\frac{N}{2}} x_{2j} w_{\frac{N}{2}}^{jk} + w^k \sum_{j=1}^{\frac{N}{2}} x_{2j+1} w_{\frac{N}{2}}^{jk} \\
&=: y_k^e + w^k y_k^o,
\end{aligned}
$$

where $w^2 = e^{-2i\frac{2\pi}{N}} = e^{-i\frac{2\pi}{\frac{N}{2}}} = w_{\frac{N}{2}}$.

In the last equation, y_k^e and y_k^o denote the DFTs of length $\frac{N}{2}$ obtained from the even and odd components of the original sequence x_j respectively. The sample size N is necessarily restricted to be a power of 2, but this does not represent a limit in the application of the algorithm to our context, as we will see shortly. It is also important to note that the index k varies from 0 to N, while the above summations are stopped at $\frac{N}{2}$. However, the outputs for $\frac{N}{2} \leq k < N$ from a DFT of length $\frac{N}{2}$ are identical to the outputs for $0 \leq k < \frac{N}{2} - 1$, being y_k^e and y_k^o periodic functions in k with period $\frac{N}{2}$. This is related to the periodicity of the twiddle factor given by

$$w^{k+\frac{N}{2}} = e^{-i\frac{2\pi}{N}(k+\frac{N}{2})} = e^{-i\pi} e^{-i\frac{2\pi}{N}k} = -e^{-i\frac{2\pi}{N}k} = -w^k$$

$$w_{\frac{N}{2}}^{k+\frac{N}{2}} = e^{-i\frac{2\pi}{\frac{N}{2}}(k+\frac{N}{2})} = e^{-i2\pi} e^{-i\frac{2\pi}{\frac{N}{2}}k} = e^{-i\frac{2\pi}{\frac{N}{2}}k} = w_{\frac{N}{2}}^k.$$

Proceeding as above, we obtain that

$$y_{k+\frac{N}{2}} = \sum_{j=1}^{\frac{N}{2}-1} x_{2j} w_{\frac{N}{2}}^{jk} - w^k \sum_{j=1}^{\frac{N}{2}-1} x_{2j+1} w_{\frac{N}{2}}^{jk}$$

$$=: y_k^e - w^k y_k^o.$$

Therefore, it is enough to calculate the terms y_k^e and y_k^o for $k = 1, \ldots, \frac{N}{2}$, and use them to obtain both y_k and $y_{k+\frac{N}{2}}$ simultaneously. This relation is often described by the following *butterfly* diagram
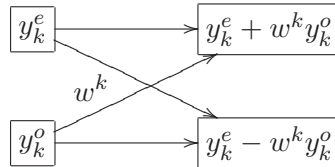
Each stage of the FFT calculation is composed of $\frac{N}{2}$ radix-2 butterflies for a total of $\frac{N}{2}\log_2 N$ diagrams per FFT. The computational effort implied in the FFT algorithm after the first cycle is here summarized:

- y_k^e requires $\frac{N}{2}$ complex multiplications and additions to be calculated, and so does y_k^o. For all $k = 1, \ldots, \frac{N}{2}$ terms, this results in $2(\frac{N}{2})^2$ arithmetic operations;

- $\frac{N}{2}$ multiplications are necessary to calculate $w^k y_k^o$ for all $k = 1, \ldots, \frac{N}{2}$;

- $N = \frac{N}{2} + \frac{N}{2}$ additions are necessary to calculate $y_k^e + w^k y_k^o$ and $y_k^e - w^k y_k^o$.

Hence, the total number of arithmetic steps is approximately $\frac{N^2}{2}$, half the number required by the direct DFT formula. After having reduced the problem to the evaluation of y_k^e and y_k^o,

the procedure can be used recursively on y_k^e and y_k^o themselves to get

$$y_k^e = y_k^{ee} + (w^2)^k y_k^{eo}, \qquad y_{k+\frac{n}{4}}^e = y_k^{ee} - (w^2)^k y_k^{eo}$$

and

$$y_k^o = y_k^{oe} + (w^2)^k y_k^{oo}, \qquad y_{k+\frac{n}{4}}^e = y_k^{oe} - (w^2)^k y_k^{oo}$$

This process of halving the order of the DFTs continues until the data are subdivided into $\log_2 N$ transforms of length 1, where the Fourier transform of a number is the number itself, in the sense that $y_k^{oeooeoe\cdots eeo} = x_q$ for some $q > 1$.

2.2.3 The Zero-padding effect

Although the Cooley-Tukey FFT procedure only works with sets of data equally distant in time and of length 2^m, where m is a positive integer, it is still possible to apply the algorithm to our series of prices. The solution is equivalent to the *zero-padding* technique of adding a number of zero valued data to increase the length of the data sequence and match the transform size supported by the algorithm. This practice will generally decrease the overall system efficiency because a portion of the available processing power is wasted on zero valued data. However, in our case the gain in terms of calculation time is still impressive. In addition, by looking into the literature, it is clear as fast algorithms for transform size rather than a power of two appear to be more complex and computational expensive[1].

To illustrate the zero-padding approach, we first need to artificially create a series of evenly spaced prices starting from the row, and irregularly spaced, transactions. If the latter are given by $t_0 \leq t_1 \ldots \leq t_n$, then the price at time $\bar{t} \in [t_{j-1}, t_j]$ can be obtained using either the *previous tick* or the *linear interpolation* method, respectively given by

$$X(\bar{t}) := X(t_{j-1})$$

$$X(\bar{t}) := X(t_{j-1}) + \frac{\bar{t} - t_{j-1}}{t_j - t_{j-1}}[X(t_j) - X(t_{j-1})].$$

Both methods are discussed in Dacorogna et al. (2001). However, the linear interpolation method is showed to induce a downward bias in the Fourier estimation of the integrated volatility, which intensifies as the sampling frequency increases (Barucci and Renò, 2002a). We will therefore avoid using it. There is also another reason to prefer the previous tick method: setting the price at time $\bar{t}$ to be equal to the last available observation is equivalent to our initial assumption that the price process $X(t)$ is piecewise constant on the interval $[t_{j-1}, t_j]$, see Section 2.2.1. Hence, if we look at the alternative expressions for the coefficients

---

[1]Our choice in favor of the Radix-2 Cooley-Tukey technique is also motived by practical reasons. All the applications in this thesis have been implemented using the C programming language, and the source code for this specific algorithm is publicly available on the Net.

of price given in Eqs. (2.7) and (2.8), it should be clear that any two interpolated prices between two consecutive (not equally distant) market transactions, a zero return is added into the formulas, but without effecting the overall result. Let us consider a simple example. Suppose that our sample is made of $X(1)$, $X(3)$, $X(4)$ and $X(8)$ observations, where $t = 1, 3, 4, 8$ are seconds. After applying the previous tick interpolation scheme to obtain a price every second, and calculating the returns ΔX_j, we get

$$X(1) \quad X(1) \quad X(3) \quad X(4) \quad X(4) \quad X(4) \quad X(4) \quad X(8)$$
$$ 0 \quad \Delta X_3 \quad \Delta X_4 \quad 0 \quad 0 \quad 0 \quad \Delta X_8$$

The number of non-zero returns before and after interpolating the data is exactly the same. Therefore, to obtain a equally space sequence of length 2^m, it is enough to first interpolate the observed prices and then keep adding the last transaction until the desired transform size is reached. Zero-padding the input sequence of returns has no consequences on the DFT calculation: it does not change the real and imaginary part of the output spectrum, but only increases the density of the DFT points in the frequency domain by interpolating the original, unpadded data. To gain a better understanding of the effect, let us consider the simple case of extending the length of the sample by filling in zeros at the end of the last available observation. Define a new time series of returns that consists of the original series followed by $N^* - N$ zeros

$$\mathbf{x} = [x_1, x_2, \ldots, x_N, 0, 0, \ldots, 0].$$

The new DFT is then given by

$$y_k = \sum_{j=1}^{N^*} x_j e^{i \frac{2\pi}{N^*} jk} = \sum_{j=1}^{N} x_j e^{i \frac{2\pi}{N^*} jk}, \qquad k = 1, \ldots, N^*.$$

Respect to the previous formula, two details should be noted: the index k goes up to N^* and $\frac{2\pi}{N^*} < \frac{2\pi}{N}$, as $N^* > N$. The last quantity is called the *frequency bin* of the DFT, and refers to the spacing between two adjacent points in the frequency domain. It turns out that, by zero padding, the output frequency spectrum now contains more data points in the same frequency range. However, these additional points are interpolated from the unpadded values and, therefore, no extra information is carried in respect to that already contained in the original series. In other words, zero padding simply provides a very accurate spectral representation of the original N-sample sequence (see the example in Appendix A.1).

2.2.4 Application

In all our application the input values x_j are real numbers. The theoretical results presented so far are still valid but the real to complex FFT performs about twice as fast as the complex

to complex FFT, when they are both applied on the same finite sequence of data. This is possible because the transform, when calculated with real values, satisfies the symmetry

$$y_k^* = y_{N-k},$$

where y_k^* denotes the complex conjugate[2] of y_k. The property exploits the periodicity of the twiddle factor and can be easily proved. The main idea of the real to complex FFT is the following: the input real sequence is treated as a complex array of half the length, and then the output is cleverly rearranged to reconstruct the desired y_k values. In this way, it is not necessary to change the implementation of the algorithm to welcome real instead of complex numbers. The idea can be expressed in more details as follows. Take the real sequence $x_1, x_2, \ldots, x_N$ of length $N = 2^m$ and treat it as a sequence of complex values $x_{2j} + i x_{2j+1}$ of length $\frac{N}{2}$, that is, $\frac{N}{2}$ elements of the input vector represent the real part of the these values, and the remaining $\frac{N}{2}$ elements the imaginary part. Then, the linearity of the DFT implies that

$$z_k = \sum_{j=1}^{\frac{N}{2}} (x_{2j} + i x_{2j+1}) w_{\frac{N}{2}}^{jk}$$

$$= \sum_{j=1}^{\frac{N}{2}} x_{2j} w_{\frac{N}{2}}^{jk} + i \sum_{j=1}^{\frac{N}{2}} x_{2j+1} w_{\frac{N}{2}}^{jk}, \qquad k = 1, \ldots, \frac{N}{2}.$$

To work these results into the transform of the original data set x_j, compare the last equation with that derived in Section 2.2.2 and given by

$$y_k = \sum_{j=1}^{\frac{N}{2}} x_{2j} w_{\frac{N}{2}}^{jk} + w^k \sum_{j=1}^{\frac{N}{2}} x_{2j+1} w_{\frac{N}{2}}^{jk}, \qquad k = 1, \ldots, N.$$

The similarity leads to the following efficient algebraic solution for the y_k in terms of the z_k values

$$y_k = \frac{1}{2} \left(z_k + z_{\frac{N}{2}-k}^* \right) + \frac{1}{2i} \left(z_k - z_{\frac{N}{2}-k}^* \right) w^k, \qquad k = 1, \ldots, N$$

The result is derived in Appendix A.2. It follows that there is no need to save the entire frequency spectrum because $y_{\frac{N}{2}+1}, y_{\frac{N}{2}+2}, \ldots, y_N$ can be obtained by taking the complex conjugate of the previously computed y_k's. Hence, in the implementation process it is only necessary to store $\frac{N}{2}$ of the output (complex) numbers . Going back to the identity between the Fourier coefficients of price and the DFT in Eq. (2.9), it means that the FFT output

---

[2]For any complex number $a + ib$, the complex conjugate is given by $a - ib$. It has the same real part but imaginary part of opposite sign.

will consist in $\frac{N}{2}$ coefficients $a_k(X)$ and an equal number of coefficients $b_k(X)$. As we will see shortly, this number correspond to the Nyquist frequency that we expect to choose when the data are equally spaced in time.

We conclude the section showing that the coefficients $a_k(X)$ and $b_k(X)$ can be effectively calculated by means of the FFT algorithm. To this purpose, we generate a series of prices using the diffusion model described in Chapter 1 with a sampling interval of 1 second for a total of $n = 7200$ seconds. In order to mirror the market behavior and generate an irregularly traded series of prices, we randomly delete 20% of the transactions from the sample; the previous tick imputation scheme is then employed to interpolate the missing data and extend the length of the sequence to $n = N + 1 = 8193$, with $N = 2^{13}$. Finally, we calculate the Fourier coefficients using both the standard procedure and the one based on the FFT algorithm, we will hereafter refer to as the *improved procedure*. In particular,

- we apply the standard procedure to the extended series of prices (scenario A);

- we apply the improved procedure to the extended series of prices (scenario B).

The output for the first 10 coefficients is summarized in the following table.

k	$a_k(dX)$ A	$a_k(dX)$ B	$b_k(dX)$ A	$b_k(dX)$ B
1	-4.178395	-4.178395	0.000000	0.000000
2	1.014212	1.014211	-3.347712	-3.347711
3	-2.386532	-2.386531	-4.540582	-4.540581
4	1.491427	1.491421	-2.898089	-2.898089
5	-7.557976	-7.557971	-3.706359	-3.706359
6	12.506399	12.506399	3.512145	3.512146
7	-0.803598	-0.803598	-3.857507	-3.857507
8	-2.326363	-2.326363	1.533958	1.533958
9	-0.424159	-0.424159	-6.790314	-6.790314
10	14.788901	14.788900	0.901543	0.901544

Table 2.1: Fourier coefficients of price calculated with the standard procedure (column A) as opposed to the FFT algorithm (column B).

Although the size of the initial sequence is too small to fully appreciate the advantages of using the fast transform in terms of computational time, it is apparent as the results corroborate the theoretical analysis developed in the previous sections. The differences in the values are mainly due to rounding errors linked to the floating point type variables employed in the implementation process.

2.3   The optimal choice of the Fourier parameters

In applying the Fourier method to study *ex post* volatility dynamics, one is left with the problem of choosing the number of Fourier coefficients of price N and the number of volatility coefficients M to include in the estimation process, and to set the level of resolution δ. Recall that the latter two quantities appear in the Fourier-Fèjer inversion formula

$$\hat{\sigma}_{N,M}^2(t) = a_0(\sigma^2) + \sum_{q=1}^{M} \varphi(\delta q) \left[a_q(\sigma^2)\cos(qt) + b_q(\sigma^2)\sin(qt) \right], \qquad (2.11)$$

where $\varphi(\delta q) = \frac{\sin^2(\delta q)}{(\delta q)^2}$, and with the volatility coefficients $a_q(\sigma^2)$ and $b_q(\sigma^2)$ as derived in Theorem 1.1, Chapter 1, for the case $i = j$. If the process is observed at regular intervals, it is possible to define an upper limit for N. In particular, suppose that n observations are available; then N should not be set above the Nyquist critical level (see Priestley, 1979), in our case given by

$$f_{Nyq} = \frac{1}{2\Delta} \qquad (2.12)$$

where $\Delta = \frac{T}{n}$ is the sampling interval, i.e. the time interval between consecutive samples. This is the fastest frequency that can be represented when data are recorded every Δ units of time. If N was set to a larger value, than the spectral density of the original process that lies outside the frequency range $(-f_{Nyq}, f_{Nyq})$ would be *aliased*, i.e. falsely translated, into the range itself resulting in a corrupted reconstruction of the volatility trajectory. Before proceeding with the analysis, let us consider a simple example to illustrate the problem of aliasing from another prospective. Figure 2.1 shows that a set of five data points (the dots) recorded every second can be delivered by sampling two sinusoids at different frequencies. In particular, it is clear that the cosine wave with the higher frequency (black line), which exceeds the Nyquist level, can be easily mistaken for the one with a lower frequency (gray line), since they are observationally indistinguishable. Also note that, by increasing the frequency even further, we would certainly find another sinusoid able to match the sample due to the periodicity of the trigonometric functions. Therefore, under the assumption of equally spaced data, we need to use enough Fourier coefficients of price to preserve the information on the original process, but we cannot exceed the maximum limit given in Eq. (2.12) either, otherwise the observed data will not adequately represent the price process itself, and so the underlying volatility dynamics.

After a continuous process has been evenly sampled, the aliasing effect cannot be removed. An anti-aliasing filter could be applied to the process to cut the high frequency out *before sampling* but, needless to say, the solution is not suitable to our context. The good news is that usually prices are not observed on the market at regular intervals, and this is enough to guarantee absence of spurious frequencies in the spectrum. On the other end, a natural
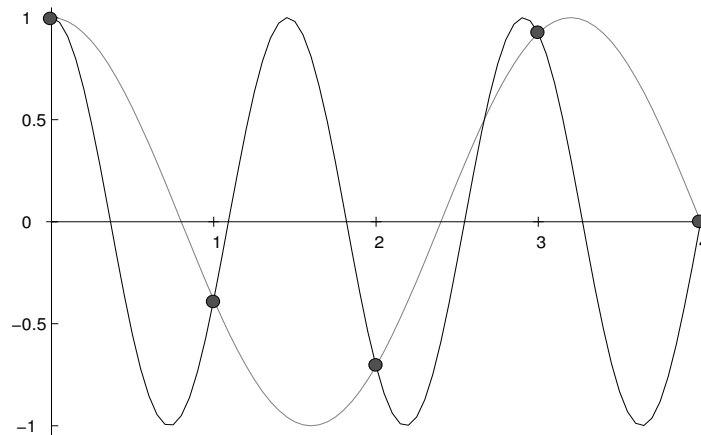
Figure 2.1: Aliasing effect: the sinusoid with a frequency larger than the Nyquist (black line) and that one obtained with a lower frequency (gray line) are observationally indistinguishable. The dots represent the observed and equally spaced data points.

limit for the maximum frequency is no longer available. Intuitively, we can only expect N to also exceed the Nyquist level when some points are located much closer than the average sampling rate, as it might be the case with irregularly spaced data. It is therefore important to know the maximum frequency content of the original process.

To deal with the problem of choosing an adeguate value for N but also for the frequency M and the smoothing parameter δ, we propose to employ a stochastic optimization algorithm to simultaneously estimate the three quantities. The algorithm is presented in the next section.

2.3.1 Joint optimization with Differential Evolution

The Differential Evolution (DE) algorithm belongs to the class of genetic or evolutionary methods that use mechanisms inspired by biological evolution, such as reproduction, mutation and natural selection, to solve difficult optimization problems. First introduced by Storn and Price (1997), DE is a population based method, in the sense that several solutions are considered at the same time to find the global minimum of a multidimensional, nonlinear and multimodal object function. Unlike other stochastic techniques that usually perturb existing results in accordance with a random quantity to locate new points on the search domain, DE uses weighted differences between decision space vectors to mutate elements of a current population and then generates new offspring's.

In the Differential Evolution algorithm, individuals are represented by d-dimensional vectors of candidates $\mathbf{x}_i$ with $i = 0, 1, \ldots, P-1$, where d is the number of parameters and P is the population size. For each solution $\mathbf{x}_i$ of the current population, a new offspring solution $\mathbf{o}_i$ is generated. The new solution $\mathbf{u}_i$ consists of a crossed over combination between $\mathbf{x}_i$ and

$\mathbf{o}_i$ where each element has a probability of π to come from $\mathbf{o}_i$ and $\mathbf{x}_i$ otherwise. The new candidate solution $\mathbf{u}_i$ therefore contains (combination of) values up to four current solutions: $\mathbf{x}_{p_1}, \mathbf{x}_{p_2}, \mathbf{x}_{p_3}$ and $\mathbf{x}_i$. If one element of $\mathbf{x}$ has the same value in all of these parenting solutions, then so will the offspring; if they disagree, the offspring will either inherit $\mathbf{x}_i$'s values, with a low probability of $(1 - \pi)$, or will be given a new value which is in the neighborhood of the corresponding value in $\mathbf{x}_{p_1}$. Finally, if $\mathbf{u}_i$ has lower objective value than $\mathbf{x}_i$, the offspring replaces $\mathbf{x}_i$, otherwise solution $\mathbf{x}_i$ survives.

The fundamental steps of the algorithm are here summarized.

1. INITIALIZATION: select an upper and lower bound for each parameter d and collect these values into two d-dimensional vectors $\mathbf{b}_l$ and $\mathbf{b}_u$. Then, each element of the candidate parameters $\mathbf{x}_i$ is initialized by extracting a random value within this range as follows

$$x_{i,j} = u_j \cdot (b_{j,u} - b_{j,l}) + b_{j,l} \qquad j = 1, \ldots, d$$

   where $u_j \sim U(0,1)$ is a uniformly distributed random number.

2. MUTATION: recombine the population to obtain a vector of P trial values given by

$$\mathbf{o}_i = \mathbf{x}_{p_1} + F \cdot (\mathbf{x}_{p_2} - \mathbf{x}_{p_3})$$

   with $i \neq p_1 \neq p_2 \neq p_3 \in [0, P-1]$ randomly sampled integers and $F > 0$ is a real constant factor that determines the rate at which the populations evolves. This difference based mutation operator is the distinctive feature of the DE algorithm.

3. CROSSOVER: construct a trial vector through a crossover operation which combines components from the current vector and from the mutant vector, according to a control parameter $\pi \in (0,1)$

$$\mathbf{v}_i = \begin{cases} \mathbf{o}_i & \text{if } U(0,1) < \pi \\ \mathbf{x}_i & \text{otherwise.} \end{cases}$$

4. SELECTION: replace $\mathbf{x}_i$ with $\mathbf{v}_i$ if $f(\mathbf{v}_i) < f(\mathbf{x}_i)$, for a given objective function $f(x)$.

Once the population is initialized, the process of mutation, crossover and selection is repeated through a finite number of generations G until the optimum is located on the parameter space, or another termination criteria is satisfied, e.g. the maximum number of iterations $G_{\max}$ is reached. There are three critical parameters in the heuristics: the population size P, the crossover probability π and the scaling factor F. As a general rule, it is advisable to set P to 10 times the number of the decision variables, while there is no upper limit for F, although effective values are seldom grater than 1.

Depending on the programming language in used, the implementation process of the the DE algorithm can be quite demanding as several technical details must be carefully considered. Moreover, to adapt the algorithm to the Fourier context an additional and substantial effort is required. To partially assist our work, we have referred to the book by Price et al. (2005), which offers an extended description of the method with a comprehensive overview of the aforementioned details (see Chapter 2 therein), including useful pseudo-code and sample applications. See also Maringer and Meyer (2008) for an accurate comparison of different evolutionary algorithms and the most recent developments in the field.

2.4 Numerical analysis

We will start with a simple example of spot volatility estimation in order to understand how the Fourier method works. A detailed analysis of the estimation process will be developed in the second part of the section. Consider the short interest rate model introduced by Chan et al. (1992). This is a broad class of processes that includes the mean reverting version of the Ornstein-Uhlenbeck process proposed by Vasicek (1977) and the one-factor general equilibrium model developed in Cox et al. (1985). It is defined as the solution of the following stochastic differential equation

$$d\,r(t) = \beta(\alpha - r(t))dt + \eta r^{\gamma}(t)dW(t). \tag{2.13}$$

To simulate the process, we have used the parameters estimated in Jiang (1998) on the 3-month Treasury bill rates via an indirect inference approach and given by $\hat{\alpha} = 0.079$, $\hat{\beta} = 0.093$, $\hat{\gamma} = 1.474$, and $\hat{\eta} = 0.794$. Figure 2.2 illustrates the temporal behavior of the diffusion coefficient $\sigma^2(t) = \hat{\eta}^2 r^{2\hat{\gamma}}(t)$ on $5,000$ daily prices obtained by sampling the process $r(t)$. The estimate obtained with the Fourier methodology is showed to well follow the dynamics of the simulated path, leading to a good pointwise reconstruction of the volatility trajectory. In particular, it is important to point out that the latter was obtained directly from the generated interest rates values and not from the simulated diffusion process, here used only for comparison.

2.4.1 Simulation design

From the previous simple application, it should be clear that the Fourier approach can be a successful procedure to estimate spot volatility. However, its performance crucially depends on the choice of the frequencies N and M, as well as of the level of resolution δ in Eq. (2.11). In our opinion, the existing literature on the method is not very clear about the actual role played by these quantities: it mainly follows the convention to set M equal to N and to adjust δ to filter out the high frequency noise modes. We would like to see whether this
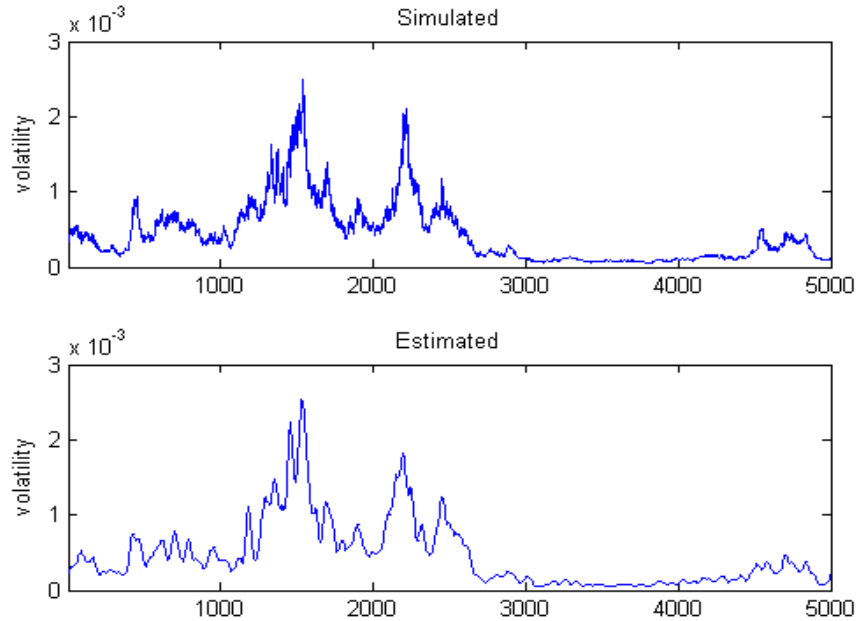
Figure 2.2: Top panel: squared diffusion coefficient $\hat{\eta}^2 r^{2\hat{\gamma}}(t)$ as simulated by model (2.13). Bottom panel: estimated trajectory via Fourier method (daily data).

could be a reasonable solution, or it is possible to define a less intuitive and more structured approach instead. To this purpose, we apply the estimation methodology to a model where both the log-price and the volatility are allowed to follow a continuous time diffusion process. Note that in Eq. (2.13) the randomness of the volatility component is instead generated by the state variable itself with the parameter γ measuring the degree of dependence of the variance from the interest rate level. The new settings refer to the popular Heston model (Heston, 1993)

$$
\begin{aligned}
dX(t) &= \sigma(t)dW_1 \\
d\sigma^2(t) &= \kappa(\theta - \sigma^2(t))dt + \eta\sigma(t)dW_2,
\end{aligned}
\tag{2.14}
$$

where the driving Brownian motions W_1 and W_2 are correlated, i.e.

$$
dW_1 \cdot dW_2 = \rho\, dt.
$$

The parameter κ represents the speed of reversion of the variance $\sigma^2(t)$ to its long-run mean θ, while η is the volatility of volatility term. The correlation parameter is typically found to be negative implying that an increase in volatility usually comes after a fall in prices, which is sometimes referred to as *leverage effect*. A negative ρ also implies that the conditional, on initial stock price $X(t)$ and volatility $\sigma(t)$, returns distribution is skewed to the left.

In order to simulate data from the model, we divide the unit interval $[0, 1]$ into m segments of equal length $\Delta = \frac{1}{m}$ and then apply the Euler-Maruyama discretization scheme to the Heston process (2.14), see Kloeden and Platen (1999),

$$X(t_i) = X(t_{i-1}) + \sigma(t_{i-1})\epsilon_1\sqrt{\Delta}$$
$$\sigma^2(t_i) = \sigma^2(t_{i-1}) + \kappa(\theta - \sigma^2(t_{i-1}))\Delta + \eta\sigma(t_{i-1})\left(\rho\epsilon_1 + \sqrt{1 - \rho^2}\epsilon_2\right)\sqrt{\Delta},$$

where ϵ_1 and ϵ_2 are i.i.d standard normal variables and $t_i = i\Delta$ for $i = 1, \ldots, m$. In the last equation we also make use of the Cholesky decomposition to correlate the variables.

We use the Differential Evolution method previously introduced to find an optimal value for the frequencies N and M, and the appropriate level of resolution δ. The objective function to be evaluated at each iteration of the algorithm is given by the root mean square error (RMSE) defined as

$$\text{RMSE} = \left(\frac{1}{\overline{n}}\sum_{k=1}^{\overline{n}}\left[\sigma^2(t_k) - \hat{\sigma}^2_{N,M}(t_k)\right]^2\right)^{\frac{1}{2}},$$

where $\hat{\sigma}^2_{N,M}(t_k)$ is the Fourier volatility and $\sigma^2(t_k)$ is the simulated volatility, both estimated at point t_k. The difference in the above equation is taken point by point over the two trajectories for a total of $\overline{n}$ points. We can also choose to set $\overline{n} \equiv m$. It follows that the optimal tern $(\hat{N}, \hat{M}, \hat{\delta})$ is the one able to deliver the minimum RMSE. Overall, the performance of the estimator is evaluated by looking at the average of the minimum values obtained after L replications of the optimization process over L different trajectories.

To introduce more empirical realism in our Monte Carlo study, we consider also the case where data are contaminated by market microstructure effects. As described in Hasbrouck (1993, 1996), a general way to model the impact of various sources of microstructure effects is to decompose the observed price into the sum of two unobservable components: a martingale component representing the informationally efficient price process, and a stationary pricing error component expressing the discrepancy between the efficient price and the observed one. According to Hasbrouck, the contaminated intraday log-price observed at time t_i is defined as

$$Y(t_i) = X(t_i) + \eta(t_i), \tag{2.15}$$

where $\eta(t_i)$ is the market noise component. We can now look at $X(t_i)$ as the price in equilibrium, that is, the price that would be observed in absence of microstructure effects. The divergence between the observed and the true prices can be induced by many sources, such as the random arrivals of buy and sell orders on the market that make the price bounce back and forth between buyers and sellers, the so-called bid-ask spread (Roll, 1984), or price discreteness (Harris, 1990), and properties of the trading mechanism (Black, 1976). We assume that the η's are i.i.d Gaussian variables with parameters $(0, \xi^2)$, and that are inde-

pendent from X. This market microstructure model, that we will call *independent noise*, is usually justified by bid-ask bounce effects. In Hansen and Lunde (2006), it is considered an inaccurate description of the noise component when data are sampled at ultra-high frequencies, that is, less than 1 minute as in our simulation study, and a noise correlated with the price process would be a more suitable choice. However, Griffin and Oomen (2008) show that such dependence can be triggered by the sampling scheme adopted for the analysis, i.e. calendar time against tick time, even when the actual noise process is independent (see also Bandi and Russell, 2006a). In addition, several noteworthy studies evaluate the effects of independent noise, including Corsi et al. (2001), Curci and Corsi (2003), Zhang et al. (2005), Zhang (2006a) and Bandi and Russell (2008). We will therefore limit our attention to this case.

If we denote the intraday return observed at time t_i as $\Delta Y_i = Y(t_i) - Y(t_{i-1})$, than we can easily calculate

$$\text{Var}(\Delta Y_i) = \int_{t_{i-1}}^{t_i} \sigma^2(s)ds + 2\xi^2 \quad \text{and} \quad \text{Cov}(\Delta Y_i, \Delta Y_{i-1}) = -\xi^2,$$

where the process $Y(t)$ clearly exhibits spurious volatility and negative serial correlation. Hence, according to this model, the observed variance is equal to the true integrated variance plus an additional term coming from the noise component. This last term is likely to induced a bias in the volatility estimates[3]. As long as the length of the return interval is sufficiently long, i.e. one day or one week in physical time, the contribution of the microstructure noise is negligible and so is the bias of the volatility estimation. However, when high-frequency data are used the contribution of the additional component increases and the size of the bias is no longer negligible. In the next section, we will see that is possible to contrast this effect by a suitable choice of the frequency N.

In order to simulate the volatility model (2.14), we take $m = 32,768$, corresponding to 2^{15}, and normalize 1 second to be $\frac{1}{m}$ so that $[0,1]$ is thought to span over roughly 9 hours of trading. In this way we can easily apply the Fast Fourier Transform to the calculation of the price coefficients. For the parameters of the Heston model, we consider the estimation results in Bakshi et al. (1997) and set $\kappa = 1.15$, $\theta = 0.04$, $\eta = 0.39$, $\rho = -0.64$ and $\sigma^2(0) = 0.04$. To mirror the inherently non homogeneous nature of data, we assume that the intervals between two contiguous trades, the so called durations, are exponentially distributed with mean $\lambda > 0$[4]. By setting $\lambda = 1, 2, 5, 8, 10, 15, 20, 30$ seconds, we generate different levels

---

[3]Mancino and Sanfelici (2008b) compute analytically the bias of the Fourier estimator for the case of integrated volatility and under the condition of regularly sampled data.

[4]The Exponential distribution is also employed in the work by Barucci and Renò (2002a) where the authors claim that the empirical distribution of the time intervals for the one-year DM-US\$ exchange rate time series, used in their application, can be approximated with an Exponential curve with $\lambda = 14$ seconds. We have repeated a similar analysis on a different data set given by the 3-month futures contracts for the S&P index, the JPY-USD and AUD-USD exchange rates respectively. Despite the different level of liquidity of the three contracts, the Exponential appears to be a good approximation of the empirical distribution also
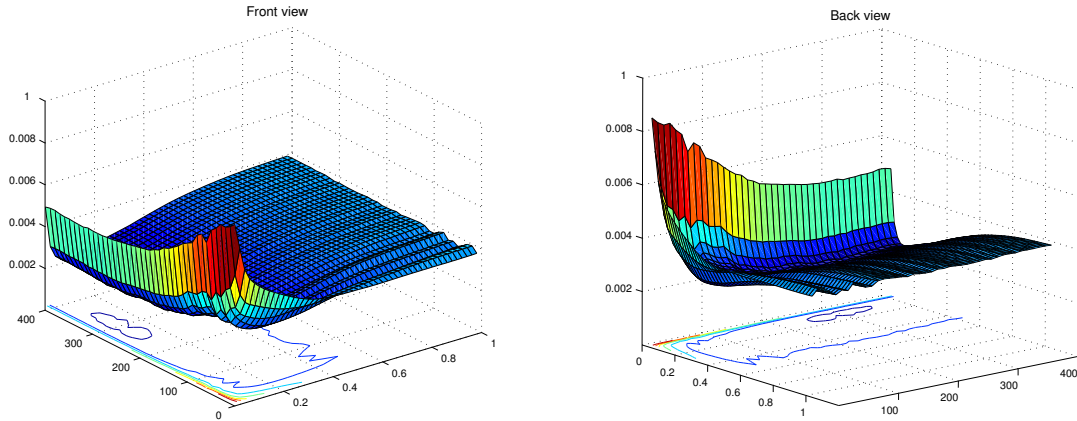
Figure 2.3: Front and back view of the RMSE surface for a series of $n = 16,384$ evenly sampled prices with $N = \frac{n}{2}$ and a selected grid of values for the frequency $M \in [0, 400]$ and the smoothing parameter $\delta \in [0.01, 1]$.

of trading activity, and obtain different sets $\{t_i, X(t_i), i = 1, \ldots, n\}$ of unevenly prices as a result. In particular, if $\lambda = 1$ it means that we are taking all the second-by-second data points in the simulated series, i.e. $n \equiv m$. The series are then extended back to the original length m using the previous tick interpolation as describe in Section 2.2.3. The implementation of the Differential Evolution algorithm requires to define an upper and lower bound for the three parameters under study. We decide to take $N, M \in [5, \frac{n}{2}]$ and $\delta \in [0.01, 1]$, where $\frac{n}{2}$ corresponds to the Nyquist level in the case of data sampled every second. To complete the set up of the optimization problem, we first explore the behavior of the RMSE object function. Figure 2.3 shows a two-side view of the RMSE surface obtained for an evenly sequence of points over a fixed grid of values for the frequency M and scale factor δ, with $N = f_{Nyq}$. The area of the local minimum is visible by the contour plot and appear to be quite smooth with a considerable large range of possible candidates. Therefore, we decide to set the population size to $P = 35$, i.e. 5 elements more than the suggested limit of 10 times the number of parameters, and to iterate the DE algorithm $G_{\max} = 50$ times for each trajectory, for a total of $L = 500$ trajectories. The trajectories are reconstructed point wise over the interval $[0, 1]$ with a time step of 10 seconds for a total of $\overline{n} = 3,276$ points. As a termination criteria, we simply choose to reach the maximum number of generations $G_{\max}$. This is a convenient strategy given the dynamics of the RMSE function, but it is slightly more expensive in terms of computational time. Finally, preliminary tests confirm that $\pi = 0.5$ and $F = 0.7$ are reasonable choices for the cross probability and the scaling factor respectively.

in our case. However, we believe that the Autoregressive Conditional Duration (ACD) model proposed by Engle and Russell (1998) would be a better choice but it is here left as a future development.

λ	$f_{Nyq}(\lambda)$	$\hat{N}$	$\hat{M}$	$\hat{\delta}$	RMSE	$\hat{N}/f_{Nyq}$
1	16384	16359 [16060, 16658]	498 [362, 434]	0.038 [0.036, 0.040]	0.00377	0.9985
2	8192	6963 [6772, 7154]	323 [292, 353]	0.066 [0.062, 0.071]	0.00426	0.8500
5	3277	2594 [2484, 2704]	255 [227, 282]	0.073 [0.068, 0.078]	0.00534	0.7916
8	2048	1598 [1530, 1666]	214 [187, 241]	0.091 [0.085, 0.097]	0.00612	0.7802
10	1638	1273 [1213, 1332]	209 [183, 234]	0.109 [0.102, 0.116]	0.00691	0.7769
15	1092	792 [747, 837]	203 [178, 230]	0.121 [0.112, 0.130]	0.00781	0.7250
20	819	601 [564, 638]	190 [166, 214]	0.152 [0.142, 0.162]	0.00874	0.7331
30	546	403 [383, 423]	179 [156, 202]	0.187 [0.175, 0.199]	0.00952	0.7387

Table 2.2: Summary results for the optimization procedure via Differential Evolution with uncontaminated prices. The 95% confidence intervals for the Fourier parameter N, M and δ are shown in brackets.

2.4.2 Simulation results

We first consider the case of uncontaminated prices. The results are presented in Table 2.2. In general, we note that the optimal values for the two frequencies N and M are inversely related to the size of the time interval between consecutive prices, while an opposite behavior characterized the resolution parameter δ. Let us first analyze the frequency N. The second column in the table corresponds to the Nyquist levels that we would obtain if we generate regularly spaced series of prices with a uniform sampling interval λ, and are calculated as $f_{Nyq}(\lambda) = \frac{32,768}{2 \cdot \lambda}$. These numbers are clearly higher than the maximum frequencies $\hat{N}$ selected by the optimization method, an aspect that can be explained by taking a closer look at the effect of the induced inhomogeneity in the data. With uneven sampling, clusters of adjacent prices are opposed to observations that may fall well apart from each other, and while in the former case a sampling frequency around the Nyquist critical level would be appropriate, the distant points would instead require a much lower number of weaves harmonics to be detected, i.e. a lower frequency. Although the following might be related to the arrival times dynamics imposed by the Exponential generator, when we plot the sets of values for $f_{Nyq}(\lambda)$ and $\hat{N}$, other features become apparent, see Figure 2.4. We first note that the curves get closer and seem to converge for durations on average larger than 15 seconds.

Table 2.2 also shows the frequency ratios defined as $\hat{N}/f_{Nyq}$ and, indeed, the values decrease consistently for $\lambda \leq 15$ to revert the tendency and increase at a much lower rate afterwards. In addition, if we take the associated optimal frequencies $\hat{N}$ and compute the corresponding time step as $\hat{\lambda} = \frac{32,768}{2 \cdot \hat{N}}$, we obtain
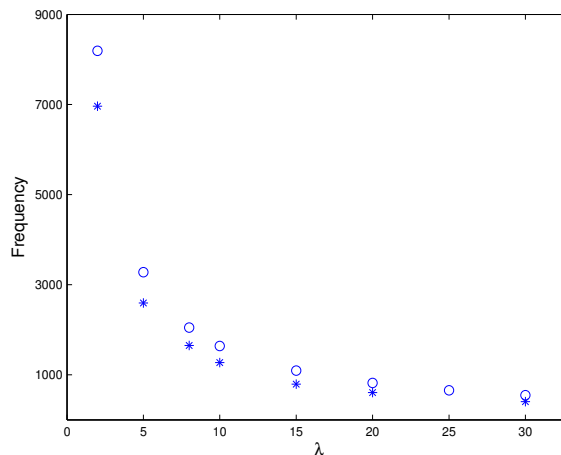
Figure 2.4: Theoretical Nyquist levels $f_{Nyq}(\lambda)$ (circles) and estimated values $\hat{N}$ (stars) behavior.

λ	1.00	2.00	5.00	8.00	10.00	15.00	20.00	30.00
$\hat{\lambda}$	1.00	2.35	6.32	10.25	12.87	20.69	27.28	40.61

We can then conclude that for sampling rates below or, equivalently, for time steps above a certain level, the Fourier method becomes more vulnerable to the absence of information in the sample at the point that it cannot distinguish between the homogeneous and inhomogeneous scenario due to the significant contribution of periods with no trading activity in both cases.

As far as the frequency M is concerned, from the results in Table 2.2 we first note that its estimated values are always well under those obtained for N, which is in contrast with the intuition in Renò (2008), then adopted in Ogawa and Sanfelici (2008), to set M to the maximum possible value, i.e. $M = N - 1$, and then randomly adjust δ to refine the volatility reconstruction. Note that taking $M = N - 1$ volatility coefficients would also increase the computational effort necessary to complete the point-by-point estimation process. In addition, the $a_q(\sigma^2)$ and $b_q(\sigma^2)$ terms are measured with a precision that deteriorates with q, since the calculation exploits $N - q$ coefficients of price (see Chapter 1, Section 1.3.1). Therefore, we can expect the optimal values for M to be smaller than those for N, in terms of minimum RMSE. More in general, having to work with samples of finite size, the performance of the Fourier estimator will always be affected by two error sources: one related to the approximation of the $a_k(dX)$ and $b_k(dX)$ coefficients, as derived at the beginning of the chapter, and the other due to the presence of the limits $N, M \to \infty$ in the price and volatility coefficient formulas. However, our optimization study reveals that both N and M can not be set arbitrarily high in the attempt to minimize the error but, instead, they have to be properly chosen to account for the different degrees of information provided by the data. It is at this stage that the contribution of the modified Lanczos sigma factor $\varphi(\delta q)$ in Eq. 2.11 becomes apparent. In the applications, the most convenient truncated Fourier series would be the one
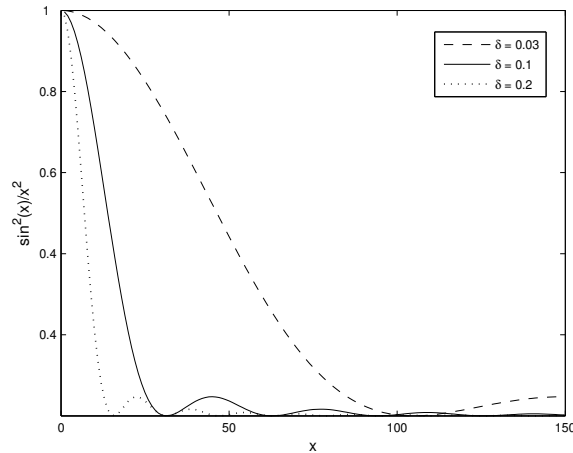
Figure 2.5: The modified Lanczos sigma factor $\varphi(x)$ at different values of the smoothing parameter δ.

with coefficients that decrease rapidly. In this case, the first few terms of the series would suffice to give a faithful representation of the true function, since the sum of all the remaining terms would be small. Hence, the faster the coefficients decrease, the fewer terms of the series are needed to guarantee an appropriate degree of accuracy. This is the idea behind the Lanczos factor: it is a smoothing procedure that attenuates the oscillations associated with the higher modes in the trigonometric approximant. Its behavior is graphically represented in Figure 2.5 where we can see that the function barely oscillates after a certain level of the frequency M. Moreover, this cut-off level becomes smaller as the value of δ is increased. In this way, it is possible to shape the partial Fourier series more gradually rather than truncating it abruptly. We can also easily infer that the optimal values for the driving parameter δ and frequency M are tightly connected and that the former affects the level of the latter. Consistently to our intuition, Table 2.2 shows that increasing values of the parameter δ have been selected by the optimization algorithm: the more the data are sparsely sampled, the higher is the degree of smoothing necessary to compensate for the consequent lack of precision.

The jointly effect of the three parameters on the estimated trajectories is illustrated in Figure 2.6 where the quality of the reconstruction is clearly altered by durations on average larger than 15 seconds. Also note as progressively higher values of δ results into a smoother but biased trajectory. However, the obtained reconstructions are still able to follow the leading dynamics of the original volatility process in Eq. (2.14) with a good representation of both the abrupt changes and the more regular sections in the volatility path. We emphasize that the time evolution of the volatility process in the figures was recovered by applying the Fourier methodology directly to the simulated prices $X(t_i)$ without making any assumption on the volatility process itself. Since the algorithm perform badly at the boundaries due to
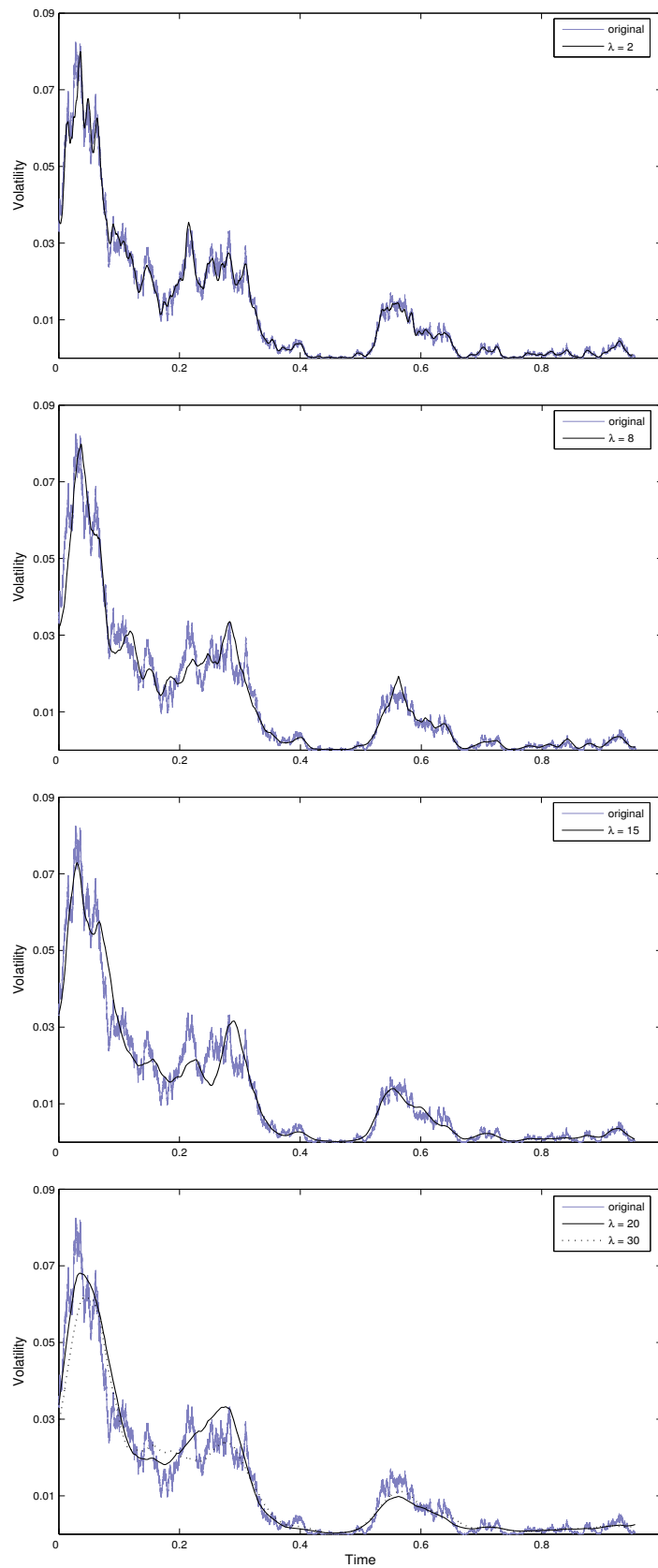
Figure 2.6: Fourier spot volatility estimation with unevenly sampled data.

	Low noise				High noise			
λ	$\hat{N}$	$\hat{M}$	$\hat{\delta}$	RMSE	$\hat{N}$	$\hat{M}$	$\hat{\delta}$	RMSE
1	1223 [1201, 1245]	358 [324, 392]	0.055 [0.052, 0.058]	0.00493	216 [199, 233]	139 [115, 163]	0.142 [0.137, 0.147]	0.01105
2	945 [919, 971]	231 [204, 257]	0.072 [0.067, 0.077]	0.00686	113 [104, 122]	90 [68, 112]	0.153 [0.146, 0.160]	0.01221
5	768 [736, 801]	181 [157, 205]	0.091 [0.085, 0.097]	0.00728	81 [73, 87]	70 [50, 90]	0.162 [0.153, 0.171]	0.01356
8	628 [602, 655]	178 [155, 201]	0.107 [0.100, 0.114]	0.00768	66 [61, 71]	69 [50, 88]	0.180 [0.171, 0.189]	0.01407
10	577 [549, 604]	172 [149, 194]	0.123 [0.115, 0.131]	0.00809	61 [57, 65]	67 [48, 86]	0.191 [0.182, 0.200]	0.01638
15	505 [476, 534]	161 [140, 183]	0.151 [0.143, 0.162]	0.00886	53 [49, 57]	63 [45, 81]	0.225 [0.213, 0.237]	0.01707
20	420 [394, 446]	153 [132, 173]	0.172 [0.161, 0.183]	0.00957	44 [41, 47]	59 [42, 76]	0.251 [0.238, 0.264]	0.01974
30	379 [360, 398]	142 [122, 162]	0.203 [0.190, 0.216]	0.01201	40 [37, 43]	55 [38, 72]	0.283 [0.268, 0.298]	0.02257

Table 2.3: Summary results of the optimization procedure via Differential Evolution with prices affected by Gaussian microstructure noise with standard deviation $\xi = 0.00005$ (Low noise) and $\xi = 0.0119$ (High noise) respectively. The 95% confidence intervals for the Fourier parameter N, M and δ are shown in brackets.

the well-known diffraction phenomena of Fourier series (see Malliavin and Thalmaier, 2005), we have excluded 2% of the outocomes from the final representation.

The effect of market noise

We will now analyze the impact of the market microstructure noise on the optimum values for the three parameters. To this purpose, we first set the standard deviation of the independent noise to $\xi = 0.00005$ (Low noise), and then consider a more realistic level given by $\xi = 0.0119$ (High noise). The latter is derived in the unpublished appendix of Bandi and Russell (2008) using IBM equity returns; values of similar magnitude are also obtained in Hansen and Lunde (2006). We report the results of the optimization procedure in Table 2.3. It is immediate to note that the number of price coefficients N is now much smaller than in the case of uncontaminated prices, especially when the microstructure noise corresponds to the value calibrated with market data. This aspect was already uncovered in Barucci and Renò (2002a), and then employed in the work by Mancino and Sanfelici (2008b). Both papers deal with the estimate of the integrated volatility using the Fourier method. Therefore, our findings corroborate those in Barrucci and Renò and confirmed that they remain valid also when the estimation process is extended to the point wise reconstruction of the entire volatility trajectory.

The value of N remains small to undo the microstructure effects and guarantee a reliable computation of the Fourier coefficients also at high frequencies. This becomes clear when we plot the first volatility coefficient $a_0(\sigma^2)$, given by the sum of the squared price coefficients $a_k(dX)$ and $b_k(dX)$ (see Eq. (1.15) in Chapter 1), while letting the number of its components
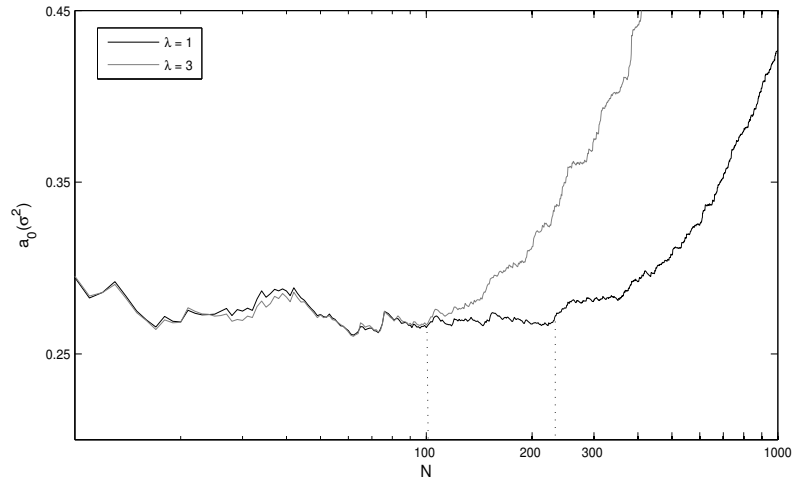
Figure 2.7: First volatility coefficient $a_0(\sigma^2)$ as a function of the frequency N. The x-axis is represented in logarithmic scale.

N varying over the interval $[0, \frac{n}{2}]$. The result is illustrated in Figure 2.7. The plotted curve can be interpreted as the mean of the power spectrum of the price process $X(t)$ over a fixed range of frequencies. It is similar in spirit to a *signature plot*, where the average realized volatility (see Chapter 1) is displayed as a function of the sampling interval[5]. Indeed, by integrating Eq. (2.11), we can easily obtained that

$$\int_0^T \sigma^2(t)dt = a_0(\sigma^2).$$

If $dX(t)$ was normally distributed, as in a model like (2.14), the spectrum would be flat and would converge to $a_0(\sigma^2)$. Instead, the plot clearly shows that the curve sharply increases starting from a certain value of N as a consequence of the negative serial correlation induced by the noise on the returns. Therefore, it is in principle possible to make the Fourier estimator invariant to market noise by choosing a suitable number of coefficients[6] to include in the estimation such that $N << \frac{n}{2}$. From this point of view N becomes the moderate sampling frequency that satisfies the trade-off between the bias induced by the noise and the precision of the estimate. For instance, the optimal value $N = 1223$ in Table 2.3 corresponds to a sampling rate of $\frac{32,768}{2 \cdot N \cdot 60} = 1.23$ minutes, which means that only the long-run dynamics contained in the data are retained in the volatility computation, while those related to a shorter time horizon are excluded to eliminate, or at least reduce, the effect of microstructure noise. In their paper on the Fourier-based estimation of integrated volatility, Mancino and Sanfelici

---

[5]Signature plots first appeared in an unpublished thesis by Fang (1996) but were named and made popular by Andersen et al. (2000).

[6]Whilst exploring the behavior of the power spectrum to choose N may represent just a 'visual' solution to the problem, we believe it remains a reasonable method, easily implementable and particularly welcome in practice.

(2008b) claim that the estimator does not need any bias correction in order to be robust to some kind of market noise if N is appropriately chosen. They also emphasize that this peculiarity of the method relies on the use of all the available data by integration and on the convolution product is based on, which let to incorporate not only the squared increments but also the autocovariances of all orders along the time window. The existing literature in the field includes important examples of biased-corrected integrated volatility estimators that utilize autocovariances to compensate for the presence of microstructure, see Zhou (1996), Hansen and Lunde (2006) and Barndorff-Nielsen et al. (2008). However, this seems to be a valid conclusion only when prices are slightly contaminated by noise, as in the work by Mancino and Sanfelici (2008b). We will provide further insight on this point in Chapter 4. Moving back to the point estimation problem, we can see from Figure 2.8 that the correction terms embedded into the Fourier algorithm allows for a good reconstruction of the volatility trajectories when they are combined with a smaller number of coefficients as showed in Table 2.3 (Low noise). The RMSE is now just above that in the case of uncontaminated prices, and so is the smoothing parameter δ. With an increased level of microstructure noise, it is apparent from Figure 2.9 that the estimator becomes less accurate in recovering the volatility path, even when the data are sparsely sampled with a moderate average duration of $\lambda \leq 5$. The results in Table 2.3 (High noise) clearly highlight the effort required in the estimation process: N is optimally set to a very low value in order to account for the distortion induced by the noise, while the delivered RMSE is almost as twice as the error measured in the previous case. In the attempt to explain such behavior, we first note how the quality of the reconstruction seems to be, in general, highly influenced by the level of the parameter N. As already stated in our first experiment, the Fourier coefficients of price convey the information on the underlying volatility process contained in the data into the estimated points along the trajectory. Moreover, the error related to the limit $N \to \infty$ in the calculation formulas is further increased by the smaller N. Therefore, the recovered process would follow the original path dynamics with different degrees of accuracy, depending on how we choose N. Due to the microstructure effects, we are forced to limit the contribution of the price coefficients and the result is clearly a less effective estimation procedure. As we have seen above, this is equivalent to sampling the data at a frequency that is lower than the actual frequency rate at which the data itself arrives on the market, therefore, foregoing details about the original process. In addition, we let the distance between consecutive prices to increase through λ, but sampling over a coarser grid can only imply a larger discretization error. From the right-hand side values in Table 2.3 we also observe that the frequency M initially remains lower than N and then inverts its tendency, in particular, starting from $\lambda = 8$. We assume that the estimator is attempting to compensate for the lack of precision triggered by N by adding additional volatility coefficients, although setting $M = N$ does not produce any substantial change in the final reconstruction.
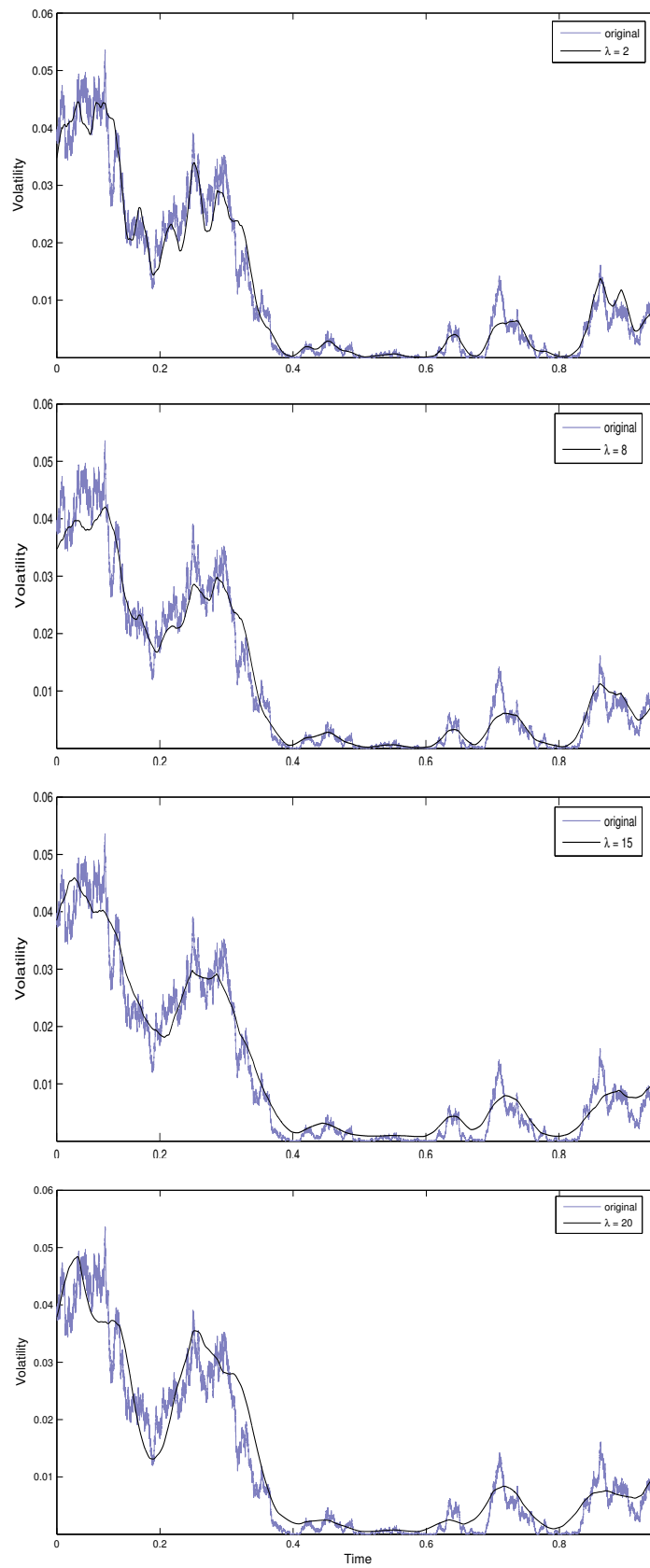
Figure 2.8: Fourier spot volatility estimation with unevenly sampled data and low microstructure noise.
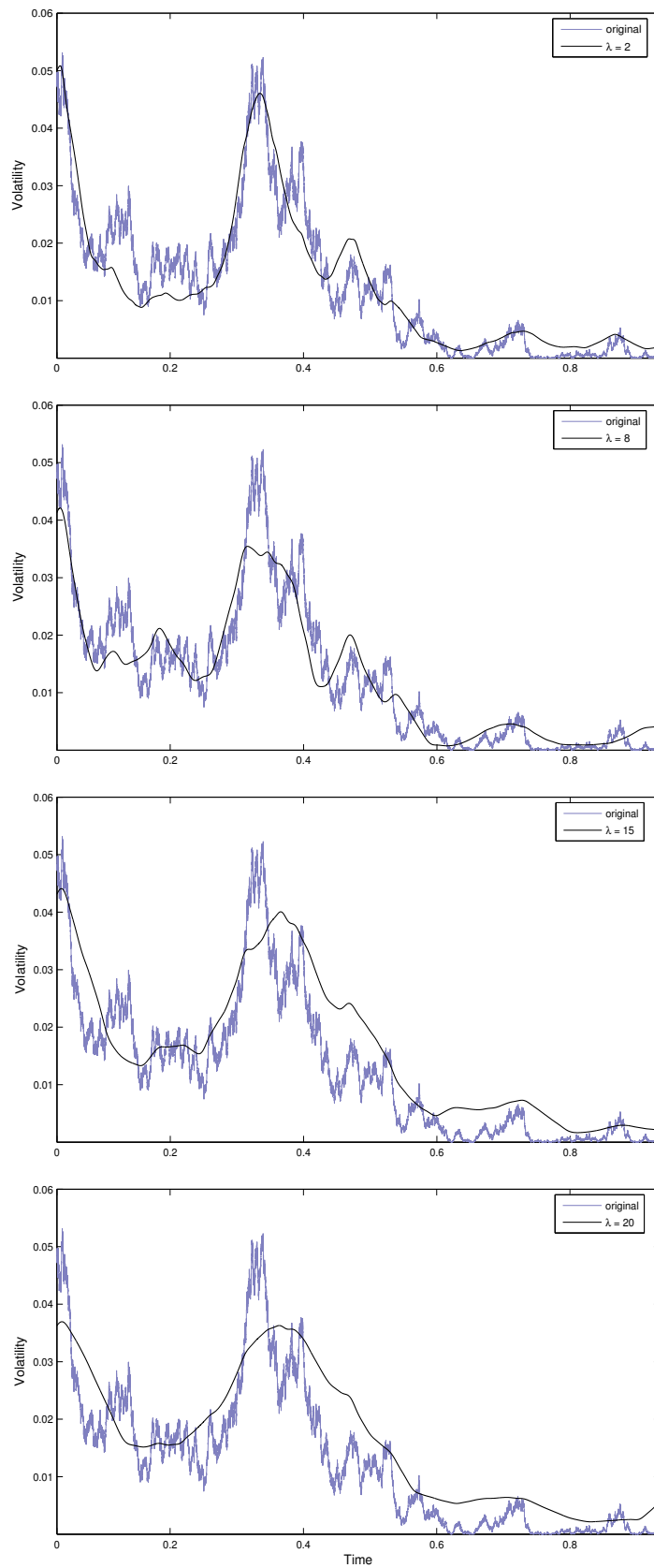
Figure 2.9: Fourier spot volatility estimation with unevenly sampled data and high microstructure noise.

2.5 Empirical application

The purpose of this study is to use market data to verify the goodness of the Fourier volatility estimates within a more realistic context. Obviously, in empirical analysis the true volatility is not observable, hence no direct evaluation criteria of the quality of the volatility estimators exist. However, general indirect methods can be employed. We start by making the usual assumption that the log-price $X(t)$ follows a standard continuous time stochastic volatility diffusion process without drift, i.e. $dX(t) = \sigma(t)dW(t)$. It follows that the univariate series of price returns can be naturally decomposed as $r(t) = \sigma(t)z(t)$ for each $t \in [0, T]$, where $z(t)$ are i.i.d. standard Normal variables. By rearranging this decomposition, we obtain the σ-standardized return

$$z(t) = \frac{r(t)}{\sigma(t)}. \tag{2.16}$$

If the Fourier volatility estimator was an adequate measure for $\sigma(t)$, then under the above assumption the distribution and dependence structure of this quantity should be Gaussian. Using a set of high-frequency transactions for the S&P 500 stock index futures, we now show that the actual empirical distribution is in fact consistent with this theoretical prediction, indicating that the Fourier estimator is a reliable spot volatility measure.

Our dataset is composed of 251 trading days from 4th January 1999 to 30th December 1999. We restrict our attention to contracts closer to maturity and we take prices between $8:30$ a.m. to $3:15$ p.m. every day for a total of $704,406$ observations. The resulting tick-by-tick series is very liquid with transactions occurring on average every 7 seconds, and an average of $3,054$ prices every day. By means of signature plots, we first set the number of price coefficients to $N = 150$, and then we choose $M = 100$ and $\delta = 0.15$ according to our simulation study. We obtain a volatility trajectory estimated at each observed tick time in order to calculate the above standardized returns. To verify our results, we first plot the returns in a histogram as showed in Figure 2.10 (left panel), and note that the characteristic symmetric bell-shape pattern is clearly detected. We add to the plot a standard normal distribution curve as a further evidence obtaining a good fit. The diagnostic statistics in Table 2.5 also confirm our first impression: following the Jarque-Bera test the distribution is Gaussian, with only slightly fatter tails. This feature is also well represented in the Q-Q plot in Figure 2.10 (right panel), which indeed deviates from the linear trend at the margins. The presence of jumps in the observed price returns is likely to result in fat-tailed distributions. It is well recognized in literature that volatility jumps and market price jumps occur in most cases at the same time and exhibit negative dependence, i.e. a large negative price movement is usually followed by a positive jump in volatility (see, for instance, the recent paper by Todorov and Tauchen, 2010). Although the Fourier estimator does not allow to separate the diffusion part from the jump component in the volatility estimation process, our intuition suggests that at each point of

Mean	0.033
Median	0.019
Maximum	3.671
Minimum	-3.173
Std. Dev.	1.122
Skewness	0.016
Kurtosis	3.306
Jarque-Bera	1.788
JB p-value	0.409

Table 2.4: Descriptive statistics for the σ-standardized returns
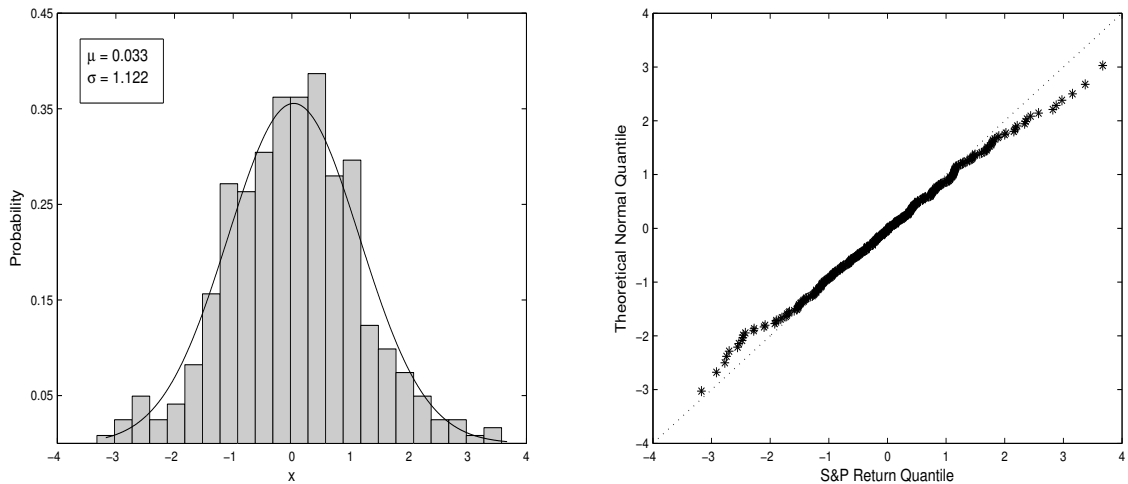
Figure 2.10: Left panel: histogram of the σ-standardized returns in Eq. 2.16 and added standard normal distribution. Right panel: relative qq-plot. The spot volatility is estimated via Fourier method as applied to a series of high-frequency transactions for the S&P 500 stock index futures.

discontinuity it calculates the average of the variance values corresponding to the left and right-hand sides of the point. This would mean that it does not overestimate the volatility at that point. Therefore, we can conclude that the Fourier method is able to deliver reliable estimates of the spot volatility process, and the excess kurtosis can be mainly explained by the presence of large price variations in the returns employed in Eq. (2.16)[7].

---

[7]In Chapter 4 we will use the same set of market data in a different empirical application and show that this is indeed the case.

2.6   An analysis of the East Asian Crisis period

Comforted by the results in the previous section, we now extend the empirical application to include a larger set of high-frequency data. In particular, we consider time series for two currency futures, the Australian dollar and the Japanese yen (both in terms of the US dollar), and for the S&P 500 index future, all observed tick-by-tick from April to December 1997. This time window includes the first phase of the East Asian Crisis, which slowly started in July of that year. The currency futures were chosen because of the geographical proximity of these countries to the center of the crisis and the index future because of the role of the underlying index as leading indicator of the US stock market performance. The objective is to test the Fourier estimator amid a period of financial turmoil, where large price movements triggered by a market crash are usually followed by a sudden increase in the volatility, as previously stated. We expect the estimator to be able to clearly link changes in market volatility to well defined periods.

The futures on the S&P 500 is the most liquid contract in our dataset with $567, 465$ tick prices, occurring on average every 8 seconds, followed by the JPY-USD and the AUD-USD series with $206, 360$ and $19, 027$ available quotes. The average durations for the last two contracts are equal to 28 seconds and almost 4 minutes respectively. The Australian market is therefore characterized by the smaller turnover of transactions. A closer look to the AUD-USD prices also reveals that long periods of no trading activity alternate with hours of liquid trading. This suggests to break down the series in order to gain a better control of the Fourier settings. Indeed, if a small value for the leading frequency N would accommodate the most active intervals to account for the microstructure noise, a higher value would be more appropriate for the other case, when the noise should instead be negligible[8]. After a careful inspections of the sample, we apply the estimator independently to the following periods: April-May, June, July-August, September, October and November-December, using Tables 2.2 and 2.3 as benchmarks to set the three parameters. For the other contracts, we resort to signature plots to roughly estimate N and then look at Table 2.3 to lead the choice of the remaining quantities. The results are the sets $\{N, M, \delta\} = \{1200, 1100, 0.18\}$ for the S&P 500, and $\{300, 400, 0.25\}$ for the JPY-USD currency futures, considering the more realist case of high market noise.

We first represent the times series in Figure 2.11 together with their relative logarithmic returns. The effect of the Hong Kong stock market crash, also known as "mini-crash", caused by the Asian crisis on October 27 is evident and translates into a clear price drop for both the AUD-USD and the S&P 500 index futures. On the other side, by looking at the graphs for the JPY-USD contract, it appears that the Asian events did not have a remarkable

---

[8]Bandi and Russell (2008) show that market data recorded at a frequency above 2 minutes are only slightly affected by microstructure noise.
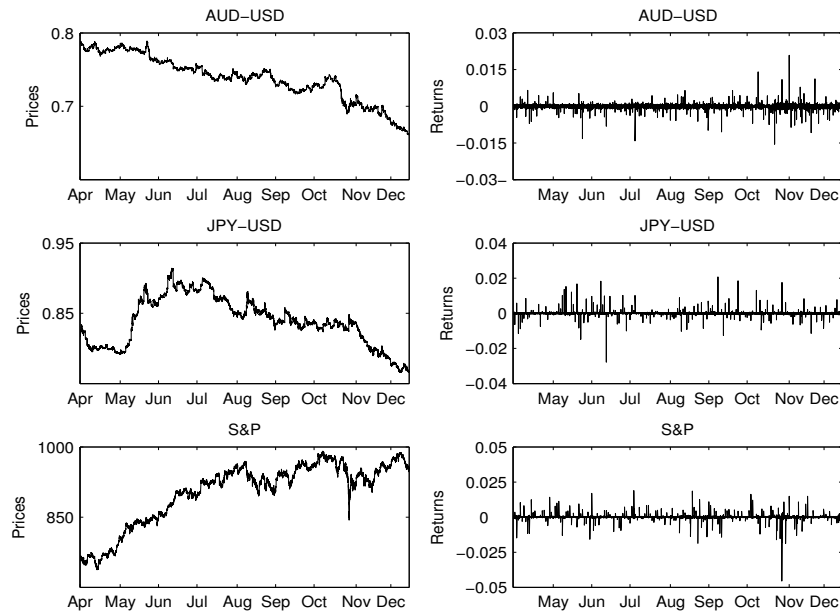
Figure 2.11: Left panel: series of tic-by-tick prices for Australian dollar and the Japanese yen currency futures and the S&P 500 index future. Right panel: relative logarithmic returns.

effect on the Japanese market. In Figure 2.12 we plot the volatility estimates for the three contracts in hand over the period under consideration. We observe that the volatilities of the AUD-USD and the S&P 500 futures follow each other rather closely between July and November 1997 and in particular around the mini-crash of October 27. A peak on May 21, following Thailand announced (on May 15) of wide-ranging capital controls, is detected on the AUD-USD futures but not on the other two contracts. The JPY-USD futures also presents a spike on October 27, but not a persistently high volatility after the event. The volatility of the JPY-USD is highest during the end of May and mid June 1997, before the crisis properly started, and cluster of high volatility are also detected around June 10. A second period of increased volatility is observed from August to September 1997. The Japanese economy was only marginally affected by the 1997 turmoil (Dungey et al., 2003, 2004). Milton (1999) argued that the severe period of recession and stagnation Japan was going through actually predated and transcended the Asian crisis. Ellis and Lewis (2001), by analyzing daily market-close data for stock prices, bond futures prices and exchange rates, found that developments in the US market generally had a much greater influence on price movements and volatility than cross-market shocks originating in the Asian crisis economies. They also provide evidence that stock markets reacted to the developments in Asia after the United States did, instead of responding directly to the news itself. Volatility of both the Australian dollar and the New Zealand dollar exchange rates against the US dollar increased remarkably during the Asian crisis, building towards the end of the period, and remained high into the world crisis
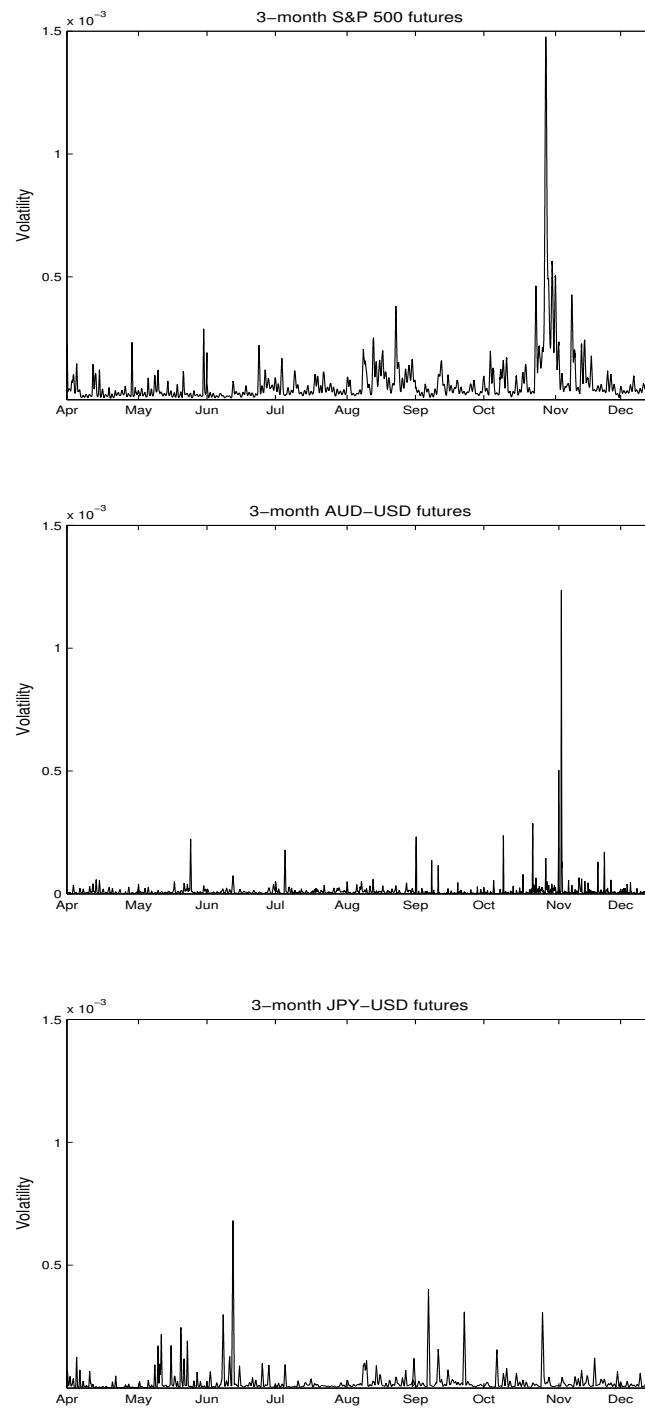
Figure 2.12: Fourier spot volatility estimation of the S&P 500 index futures (top panel), the AUD-USD (middle panel) and JPY-USD currency (bottom panel) currency futures over the period Apr-Dec 1997.

period. Our results seem to strengthen the analysis of Ellis and Lewis and show that the volatility of the AUD-USD futures follows closely the volatility of the S&P 500 index futures. This conclusion is well supported by the clear, large spike in the volatility trajectory of the Australian futures which occurs soon after a similar level of volatility is detected for the futures on the index.

2.7   Summary

We have thoroughly studied the Fourier estimator proposed by Malliavin and Mancino (2002) and provided strong evidence that the method is able to recover the temporal dynamics of a latent volatility process, also when the input series of data are not sampled at regular intervals, as it is usually the case with high-frequency transactions. We have showed in details how to use the Fast Fourier Transform in the implementation process, obtaining a significant improvement of the numerical performance of the estimator. An important part of this chapter is devoted to the analysis of three crucial quantities in the estimation process, namely, the number of price and volatility coefficients, and the resolution parameters, the first being the most influential. The problem of choosing an optimal value for these quantities is tackled through the Differential Evolution algorithm. The principal limitation of this approach is that it cannot be applied to real market data, unless an analytical formula for the mean square error of the Fourier estimator is previously obtained. An alternative objective function would work equally well. Nonetheless, the optimization procedure allowed us to highlight the separate contribution of the three variables under reasonable market conditions. It was of particular interest realize that the quality of the reconstructed trajectories strongly depends upon the number of added Fourier price coefficients, which also can be seen as the dominating frequency in the volatility spectrum. The role of this frequency is also determinant in the case of contaminated prices as it helps to control the bias induced by the microstructure noise. However, when a higher level of noise is considered, a further drop in the frequency level also generates a larger approximation error (related to the coefficients formula), penalizing the accuracy of the estimator. The effect is further exacerbated when the degree of inhomogeneity in the data is also increased. Empirical evidence is provided by using an indirect criterion to asses the quality of the Fourier volatility estimates. The method is showed capable to measure the adequacy of the commonly used continuous time model for the observed asset prices. We have then applied the estimator to market futures time series, observed at different trading frequencies and spanning the period over the East Asian crisis. Our results are coherent with the exiting literature. We have found that the Australian economy was not impacted by the Asian events directly but its reaction was rather driven by the developments in the US market. Also, we have observed that the Asian turmoil did not have a noticeable effect on the Japanese market.

CHAPTER 3

Correlation analysis

According to Markowitz (1959), the asset selection process should focus on identifying stocks that eliminate firm-specific risks while isolating the impact of market-risk on the overall return of the portfolio. Such analyses are important to avoid naive diversification and dealing with idiosyncratic risk. Thus, a well-diversified portfolio is not simply a collection of good dissimilar assets, but a collection of assets that minimize the total risk of the portfolio itself. In terms of risk reduction, market co-movements between stocks become crucial. It is common practice to measure such movements by means of the popular Pearson correlation coefficient. However, two major effects may invalidate the reliability of this estimator: the statistical uncertainty triggered by the finite length of the sample; the asynchronous nature of the time series of prices under analysis. It is therefore important to compare different correlation measures. In this chapter, we examine the performance of the Pearson coefficient against the Fourier estimator, here applied to a large set of stocks for the first time. In particular, we evaluate the performance of the two methods by means of the Kullback-Leibler divergence, a dissimilarity measure between two completely determined probability distributions. The analysis is then extended to include the correlation estimator proposed by Hayashi and Yoshida (2005), which is showed to be highly competitive in absence of microstructure noise. However, differently from the other methods, this procedure does not guarantee the positive definiteness of the output correlation matrix, therefore invalidating the use of the Kullback-Leibler distance. We handle this problem by means of the shrinkage technique developed in Ledoit and Wolf (2004b). Finally, in a pioneering study, we apply the Fourier measure to recover the time-varying and stochastic dynamics of a correlation process, and also unveil further aspects of the estimation procedure.

3.1 The Pearson coefficient

The Pearson coefficient measures the tendency of two random processes to move together. Differently from the Fourier method, it can only be applied to synchronous data and, therefore, it is usually necessary to first synchronize the tick prices on a common grid using the interpolation methods presented in Section 2.2.3. Suppose now that $X(t_i)$ and $Y(t_i)$ are the observed prices of two financial assets $X(t)$ and $Y(t)$. Let $\Delta X_i = X(t_i) - X(t_{i-1})$ and $\Delta Y_i = Y(t_i) - Y(t_{i-1})$ the logarithmic returns sampled at regular intervals $\frac{1}{n}$, where n is the sample size. Then the Pearson correlation coefficient is defined as

$$\rho_{ij}^{\mathrm{P}} = \frac{\sum_{i=1}^{n} \left(\Delta X_i - \sum_{i=1}^{n} \Delta X_i\right)\left(\Delta Y_i - \sum_{i=1}^{n} \Delta Y_i\right)}{\sqrt{\sum_{i=1}^{n} \left(\Delta X_i - \sum_{i=1}^{n} \Delta X_i\right)^2 \sum_{i=1}^{n} \left(\Delta Y_i - \sum_{i=1}^{n} \Delta Y_i\right)^2}} \tag{3.1}$$

The values of the coefficient range in the interval $[-1, 1]$. Preliminary to the analysis developed in this chapter, we recall that the Pearson correlation matrix relative to a system of p assets is always positive definite, i.e. has no negative eigenvalues.

3.2 Modelling integrated correlation via Fourier method

We first recall that the Fourier method is based on the assumption that the logarithmic prices $X(t) = (X_1(t), \ldots, X_p(t))$ are Brownian semimartingales satisfying the stochastic differential equations

$$dX_i(t) = \sigma_i(t)dW_i(t), \qquad i = 1, \ldots, p$$

where σ_i is the instantaneous volatility of asset i. We assume to work with correlated Brownian motions, i.e. $d[W_i, W_j]_t = \rho_{ij}\, dt$, for $i, j = 1, \ldots, p$. It follows that the elements of the volatility matrix $\Sigma(t)$ are given by $\Sigma_{ij}(t) = \rho_{ij}\sigma_i(t)\sigma_j(t)$. This is a time-dependent covariance matrix evolving according to the dynamics of the volatilities and the constant correlation among asset returns. In Section 3.7 we will allow the correlation to be stochastic. We have already seen that a consistent estimator of $\Sigma_{ij}(t)$ is given by

$$\hat{\Sigma}_{N,M}^{ij}(t) \approx a_0(\Sigma_{ij}) + \sum_{q=1}^{M} \varphi(\delta q)\left[a_q(\Sigma_{ij})\cos(qt) + b_q(\Sigma_{ij})\sin(qt)\right] \qquad t \in [0, 2\pi]. \tag{3.2}$$

Using this inversion formula, it is easy to show that an estimate of the integrated covariance is given by

$$\int_0^{2\pi} \hat{\Sigma}_{N,M}^{ij}(t)dt = 2\pi a_0(\Sigma_{ij}),$$

where

$$a_0(\Sigma_{ij}) \approx \frac{\pi}{N + 1 - n_0} \sum_{k=n_0}^{N} \frac{1}{2}\left[a_k(dX_i)a_k(dX_j) + b_k(dX_i)b_k(dX_j)\right], \tag{3.3}$$

which implies that

$$\rho_{ij}^{\mathrm{F}} = \frac{a_0(\Sigma_{ij})}{\sqrt{a_0(\Sigma_{ii})a_0(\Sigma_{jj})}}. \tag{3.4}$$

Therefore, the calculation of the integrated correlation is only affected by the choice of the convenient frequency N at which to stop the expansion (3.3). We will see in Section 3.5 that the role played by this frequency is again determinant in driving the accuracy of the estimates. This would be particularly evident when we extend our analysis to consider the case of non-synchronous data.

3.3 The Kullback Leibler divergence

In probability and in information theory the Kullback-Leibler divergence (Kullback and Leibler, 1951), also known as relative entropy, is a dissimilarity measure between two probability distributions, the *true* distribution Q_1 and a *target* distribution Q_2. It is defined as

$$K(Q_1, Q_2) = E_{Q_1}\left[\ln\left(\frac{Q_1}{Q_2}\right)\right],$$

where $E_{Q_1}[\cdot]$ indicates the expected value with respect to the distribution Q_1. The relative entropy has the property to be always non negative and equals zero when the two distributions are equivalent. However, it is clearly not symmetric and does not satisfy the triangle property either; therefore, it cannot be considered as a proper distance, hence the term 'divergence' to characterized it (in common practice the two terms are nonetheless used interchangeably). The smaller the entropy, the more similar are the distributions of the two variables. This motivate the use of the divergence to measure, for instance, the error one commits in considering two random variables X and Y as uncorrelated variables by evaluating $K[q(X, Y), q(X)q(Y)]$, where $q(X, Y)$ is the joint probability density function while $q(X)$ and $q(Y)$ are the corresponding marginal distributions.

In what follow, we will mainly consider the Kullback-Leibler divergence between multivariate Gaussian distributions. We will refer to the probability density function (pdf) of such distribution as $\eta(\Psi, X)$, where Ψ is the correlation matrix of the system X. Given two different multivariate Gaussian distributions $\eta(\Psi_1, X)$ and $\eta(\Psi_2, X)$, Tumminello et al. (2007b) show that

$$\begin{aligned}
K[\eta(\Psi_1, X), \eta(\Psi_2, X)] &= E_{\eta(\Psi_1, X)}\left[\ln\left(\frac{\eta(\Psi_1, X)}{\eta(\Psi_2, X)}\right)\right] \\
&= \int \eta(\Psi_1, X)\ln\left(\frac{\eta(\Psi_1, X)}{\eta(\Psi_2, X)}\right) dX \\
&= \frac{1}{2}\left[\ln\left(\frac{|\Psi_2|}{|\Psi_1|}\right) + \mathrm{tr}(\Psi_2^{-1}\Psi_1) - p\right],
\end{aligned} \tag{3.5}$$

where $|\cdot|$ indicates the determinant of Ψ.

From the last equation it is apparent as the Kullback-Leibler divergence is an explicit function of the correlation matrices Ψ_1 and Ψ_2, and we can then simplify the notation by referring to the entropy as to $K(\Psi_1, \Psi_2)$. Note that, differently from a typical distance measure such as the Frobenius distance, the divergence (3.5) can be calculated only if Ψ_1 and Ψ_2 are positive definite.

In the next section we will largely use the expectations of the Kullback-Liebler divergence as an absolute term of comparison in our correlation analysis. We present here the theoretical results as derived in Tumminello et al. (2007b) by exploiting the theory of Wishart matrices

$$E[K(\Psi, R_1)] = \frac{1}{2}\left[p\ln\left(\frac{2}{n}\right) + \sum_{j=n-p+1}^{n} \frac{\Gamma'(\frac{j}{2})}{\Gamma(\frac{j}{2})} + \frac{p(p+1)}{n-p-1}\right] \tag{3.6}$$

$$E[K(R_1, \Psi)] = \frac{1}{2}\left[p\ln\left(\frac{n}{2}\right) + \sum_{j=n-p+1}^{n} \frac{\Gamma'(\frac{j}{2})}{\Gamma(\frac{j}{2})}\right] \tag{3.7}$$

$$E[K(R_1, R_2)] = \frac{1}{2}\frac{p(p+1)}{n-p-1}, \tag{3.8}$$

where R_1 and R_2 are two sample correlation matrix obtained from two *independent* realizations of the system, and Γ' is the derivative of the Gamma function Γ. With asynchronous data, the effective length of the series differs from stock to stock and it is necessary to modify the expected values of the Kullback-Leibler distance accordingly. In particular, for two time series of length n_1 and n_2 the last equation becomes

$$E[K(R_1, R_2)] = \frac{1}{2}\left[p\ln\left(\frac{n_1}{n_2}\right) + \sum_{j=n_2-p+1}^{n_2} \varphi\left(\frac{j}{2}\right) - \sum_{j=n_1-p+1}^{n_1} \varphi\left(\frac{j}{2}\right) + \frac{p(p+1)}{n_2-p-1}\right], \quad (3.9)$$

where $\varphi(z) = \Gamma'(z)/\Gamma(z)$ is the first-order Polygamma function. Note that the formula in Eq. (3.5) is independent of T and there is no need to modify it accordingly.

The asymmetry of the distance turns out to be very useful when it is necessary to distinguish between quantities assumed to be 'true', such as the correlation matrix in a real system, and quantities that are estimated empirically. It is also apparent that all the expectation values are independent of the specific model because Ψ does not appear in any of the above formulas. This peculiar aspect has two important implications: (i) the expected value of the Kullback-Leibler divergence can be calculated also in the common case where the underlying system model is unknown; (ii) the relative entropy is a valid approach to measure the statistical uncertainty of the correlation matrix due to the finite length of the data sample.

3.4   Simulation design

The comparative analysis presented in the next section is based on a large number of simulated high-frequency series of prices. To introduce more empirical realism in the data, we first apply a clustering technique to extract useful information from a set of 100 highly capitalized equity stocks traded on the New York Stock Exchange, following the intuition that financial assets within a specific market sector tend to move broadly together. This allows to identify the hierarchical structure of an initial correlation matrix, that is then used to establish interdependencies among the underlying-driving processes of the simulated assets[1]. The above structure is described in terms of a factor model. The next section formalize the whole idea in details.

3.4.1   A hierarchically nested factor model

Consistently with the assumption that certain economic factors play the role of determinants in affecting the price returns of a broad portfolio of traded stocks (Ross, 1976), we can look at financial markets as complex systems organized in a nested hierarchical structure: the elements of the system can be partitioned in clusters which in turn can be partitioned in subclusters and so on, up to a certain level (Mantegna, 1999). The shared information within each cluster can be evaluated in terms of a similarity measure, usually given by the correlation between elements within the cluster. Therefore, a correlation matrix R of a multivariate complex system can be used to extract information about the hierarchical organization of the system itself. As a result of the clustering procedure, a *hierarchical tree*, or *dendrogram*, is associated with the matrix providing a graphic representation of the hierarchies. An illustrative example is given in Figure 3.1. A dendrogram is a rooted tree where a special node, the root, is singled out respect to the remaining leaves and internal nodes of the tree. In our example, this node is labeled by α_1. An ordered set of internal nodes called *genealogy* $G(i)$ $(G(\alpha_h))$ is connecting the leaf i (internal node α_h) to the root α_1. For instance, by looking at the graph, we note that the genealogy of leaf 8 is $G(8) = \{\alpha_4, \alpha_3, \alpha_1\}$, and that of node α_8 is $G(\alpha_8) = \{\alpha_8, \alpha_5, \alpha_4, \alpha_3, \alpha_1\}$, where the internal node α_8 is included in $G(\alpha_8)$. Depending on the clustering algorithm in use, the output dendrogram can be slightly different as different aspects of the sample correlation matrix are highlighted by the method. A large number of hierarchical clustering procedures can be found in the literature. Here we will use the Average Linkage Cluster Analysis where clusters that exhibits high level of similarity within the cluster are merged together by taking the average of the correlation coefficients

---

[1]It is interesting to observe that the relevance of economic sectors and subsectors in the correlation structure is shown to depend on the sampling time horizon of the stock returns as well. Specifically, the system turns out to be more hierarchically structured at daily time horizon confirming that the market needs a finite amount of time to assess the correct degree of cross correlation between pairs of stocks.
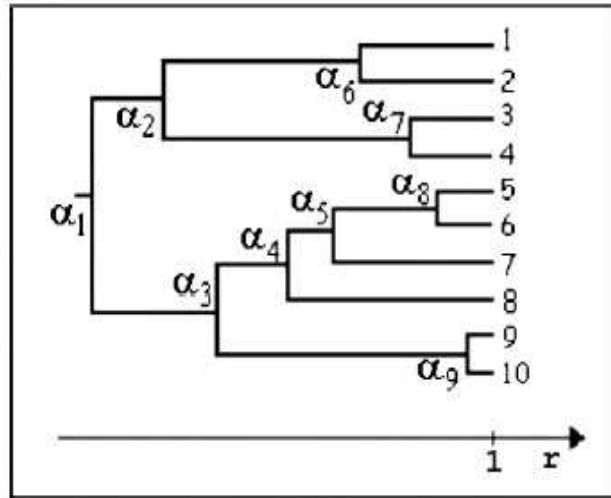
Figure 3.1: Example of dendrogram related to a system of $p = 10$ elements (leaves in the tree), where $\{\alpha_1, \ldots, \alpha_9\}$ represent the $p - 1 = 9$ internal nodes.

between all pairs of elements in the clusters (Anderberg, 1973). In general, starting from a empirical correlation matrix R, the clustering algorithm associates the correlation coefficient r_{α_i} of R with each internal node α_i. The whole information contained in the hierarchical tree is then stored in a filtered similarity matrix $\bar{R}$ of elements $\bar{r}_{ij} = r_{\alpha_k}$, where α_k is the first internal node in which leaves i and j merge together. For instance, from Figure 3.1 we have $\bar{r}_{37} = r_{\alpha_1}$ and $\bar{r}_{57} = r_{\alpha_5}$. Note that, by construction, in $\bar{R}$ there are at most $p - 1$ distinct coefficients against the $p(p-1)/2$ of the original matrix. Finally, we use the notation $b = g(a)$ to indicate that an internal node b is the *parent* of the node a if b immediately precedes a on the path from the root to a, e.g. $\alpha_5 = g(\alpha_8)$ in our graphical example.

Tumminello et al. (2007a) show that it is possible to describe the above structure in terms of a factor model[2] as long as this is constructed in such a way that its correlation matrix coincides with the similarity matrix filtered by the clustering procedure as applied to R. The proposed model is called Hierarchically Nested Factor Model (HNFM) and is defined as

$$x_i(t) = \sum_{\alpha_h \in G(i)} \gamma_{\alpha_h} f^{\alpha_h}(t) + \psi_i \epsilon_i(t), \qquad i = 1, \ldots, p \tag{3.10}$$

with $\psi_i = \sqrt{1 - \sum_{\alpha_h \in G(i)} \gamma_{\alpha_h}^2}$. The model postulates that the variable $x_i(t)$ is linearly dependent upon a few unobservable random variables $f^{\alpha_h}(t)$ called *factors*, and an additional source of variation $\epsilon_i(t)$. In particular, the h-th factor $f^{\alpha_h}(t)$ and $\epsilon_i(t)$ are i.i.d. variables with zero mean and unit variance. The bijective relation between the factor model and the

---

[2]A factor model is motivated by the following argument: suppose that some variables within a particular group are highly correlated among themselves, but have relatively small correlations with variables in a different group. Than it is possible that each group of variables represents a single underlying factor that is responsible for the observed correlations.

filtered matrix $\bar{R}$ is established by setting

$$\gamma_{\alpha_1} = \sqrt{r_{\alpha_1}}$$
$$\gamma_{\alpha_h} = \sqrt{r_{\alpha_h} - r_{g(\alpha_h)}}, \qquad h = 2, \ldots, p - 1 \qquad (3.11)$$

where we assume that $r_{\alpha_1} \geq 0$[3]. To prove that $\bar{R}$ is the correlation matrix of the factor model (3.10), we need to show that the cross correlation $E(x_i, x_j)$ is given by $\bar{r}_{ij} = r_{\alpha_k}$, where (i, j) is a generic pair of elements merging together at the node α_k and corresponding to the correlation level r_{α_k}. We first note that $E(x_i, x_j)$ depends only on the factors f^{α_h} which are common to x_i and x_j, and since there is a factor for each internal node in the dendrogram, it is necessary to identify the internal nodes belonging to both the genealogies $G(i)$ and $G(j)$. To this purpose, we see that $G(i) \cap G(j) = G(\alpha_k)$ holds true. The relation can be easily verified with the example in Figure 3.1 where we have that $G(8) = \{\alpha_4, \alpha_3, \alpha_1\}$ and $G(9) = \{\alpha_9, \alpha_3, \alpha_1\}$, and therefore, $G(8) \cap G(9) = \{\alpha_3, \alpha_1\} = G(\alpha_3)$. Finally, by Eqs. (3.10) and (3.11) we have that the cross-correlation between variable x_i and x_j is

$$E(x_i, x_j) = \sum_{\alpha_h \in G(\alpha_h)} \gamma_{\alpha_h}^2$$
$$= r_{\alpha_k}$$
$$= \bar{r}_{ij}.$$

For instance, $E(x_8, x_9) = \gamma_{\alpha_3}^2 + \gamma_{\alpha_1}^2 = r_{\alpha_3} - r_{g(\alpha_1)} + r_{g(\alpha_1)} = r_{g(\alpha_3)} = \bar{r}_{89}$. Hence, the HNFM is an opportunely designed factor model able to take into account the hierarchical properties of the investigated system. Moreover, when a factor model with correlation matrix $\bar{R}$ exists, it implies that $\bar{R}$ is always positive definite if $r_{\alpha_1} \geq 0$.

The model as given in Eq. (3.10) is based on $p - 1$ factors coming from a dendrogram of m elements. However, it is possible to remove the factors that are not statistically 'reliable' respect to a predefined standard threshold. In Tumminello et al. (2007a), this is achieved through a method based on a non-parametric bootstrap technique (Efron, 1979). The method allows to associate a bootstrap value to each internal node of a hierarchical tree; due to the one-to-one relation between the nodes in the hierarchical tree and the factors in the HNFM, the bootstrap value associated with a certain node is also associated with the corresponding factor in the HNFM. The bootstrap value corresponds to the fraction of bootstrap replicas preserving the internal node in the dendrogram, i.e. given an internal node α_h in the original dendrogram, a bootstrap replica is preserving the node if and only if there exists a node $\hat{\alpha}_h$ in a replica dendrogram connected to the same leaves of α_h. For example, we say that the node α_2 in Figure 3.1 is preserved if and only if a node in some replica dendrogram belongs to the

---

[3]Should this value be negative, it is possible to define a suitable linear transformation of the similarity measure that does not modify the final structure of the dendrogram.

genealogy of all the leaves 1,2,3 and 4. The authors than proposed to remove the nodes, and therefore the corresponding factors, with bootstrap value smaller than a suitable threshold, whose value is inferred from the data.

3.5 Comparing Fourier and Pearson estimators

Our comparison study between the Pearson coefficient and the Fourier estimator is based on 100 sets of $p = 100$ series of stock prices of length n generated from HNFM. In particular, the log prices satisfy the diffusion process

$$dX_i(t) = \sigma_i dW_i(t)$$

where

$$dW_i(t) = \sum_{k=1}^{Q} \gamma_{ik} f_k(t) + \psi_i \epsilon_i(t), \qquad i = 1, \ldots, p$$

with Q indicating the number of significative factors and $\psi_i = \sqrt{1 - \sum_{k=1}^{Q} \gamma_{ik}^2}$. In our framework the factors $f_k(t)$ and the noise component $\epsilon_i(t)$ are independent and identically distributed Gaussian random variables with zero mean and unite variance. To correlate the Brownian motions, we employ the correlation matrix obtained in Tumminello et al. (2007a) using the aforementioned bootstrapping technique as applied to a set of daily price returns of 100 highly capitalized equity stocks traded on the New York Stock Exchange. The resulting number of factors is equal to $Q = 23$.

3.5.1 Synchronous data

We start considering the simple case of synchronous data with prices simulated every 1 second for a total of $n = 32,768$ transactions. The length of the series is set to $n = 2^{15}$ in order to easily apply the Fast Fourier Trasform (FFT) technique as shown in the previous chapter. The significant gain in terms of computational time is especially welcome in this context given the size of the dataset under analysis. We then estimate the correlation matrices following both the Pearson and the Fourier approaches, and compare the results through the Kullback-Leibler distance $K(\Psi, R)$, where Ψ is the true correlation matrix of HNFM and R is the estimated one. By closely looking at Eq (3.3), we note that the Fourier procedure can provide a correlation point, or a correlation matrix if we consider p assets, for each value of frequency N. In particular, the upper bound of this range of values corresponds to the fastest frequency representable using data sampled at 1 second, namely the Nyquist frequency introduced in Chapter 2 and here given by $N = \frac{n}{2}$. By means of the relation $N = \frac{n}{2 \cdot \tau}$, it is then possible to associate at each frequency the corresponding sampling rate τ in the time domain. In particular, we restrict our attention to a selected number of rates ranging
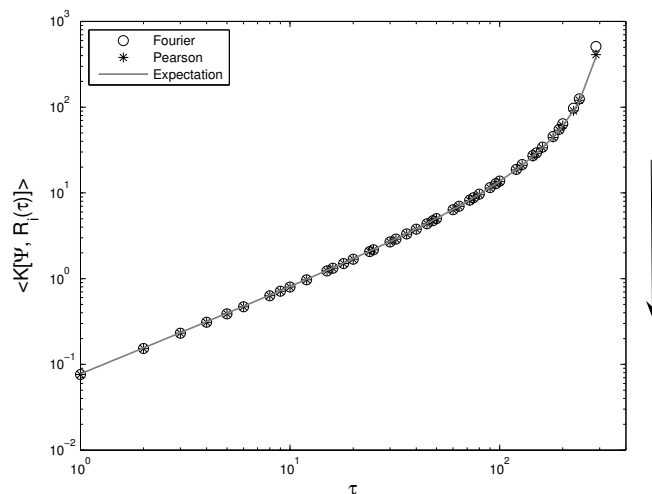
Figure 3.2: Average Kullback-Leibler distance between the true correlation matrix of the model Ψ and the Fourier (Pearson) correlation matrix R estimated from $n = 100$ synchronously sampled time series of length $T = 32,768$. The average is taken over 100 independent datasets. The straight line represents the theoretical expected value given in Eq. (3.6). The increasing level of accuracy follows the arrow direction.

from $\tau = 1$ to $\tau = 288$ seconds, for a total of 41 correlation matrices for each stock. The Pearson coefficients are calculated by taking prices every τ seconds with no need to previously interpolate them since the data are already synchronous. We first employ the Kullback-Lieber distance to measure the precision of the two methods. In Figure 3.2 the Pearson and Fourier estimators are showed to be in complete agreement with respect to the theoretical expected value of the distance given in Eq. (3.6). Deviations from this value would mean that the sample matrix R is not consistent with the true matrix Ψ. It is also clear that the most accurate estimate is obtained at $\tau = 1$ second or, equivalently, when N equals the Nyquist frequency $\frac{n}{2}$. Being the time series synchronous and sampled every second, Fourier delivers the best result in correspondence of the maximum attainable frequency, therefore when all the available information is carried into the estimated points. A larger value (not displayed in the graph) would trigger aliasing effects, whilst a smaller frequency, i.e. a larger τ, would amplify the approximation error due to the limit in the coefficient formulas, reducing the accuracy of the outcomes as a consequence. This is clearly showed in the graph. Similarly, the accuracy of the Pearson coefficient decreases as the data in the sample become more sparsely, implying a higher discretization error. Overall, we observe that the relative entropy between the correlation matrix of the model and its Fourier or Pearson estimation at different sampling intervals is an increasing function of the interval itself: as expected, increasing the sample rate reduces the statistical uncertainty and clearly improve the resulting correlation estimates in both cases. Nonetheless, it is of great interest to realize that both methods are on average equally accurate, despite being based on a very different theoretical background.
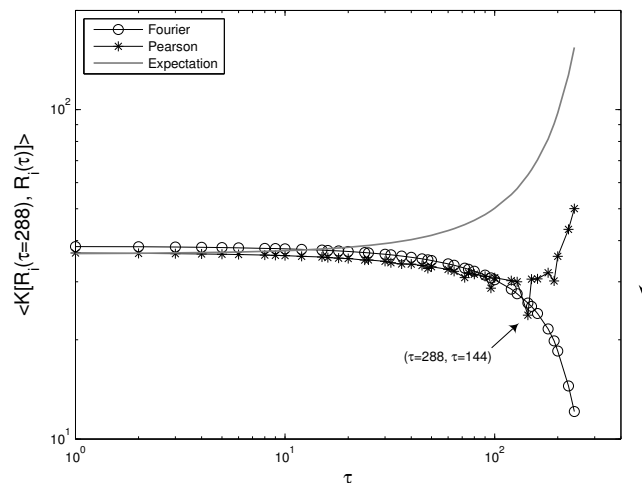
Figure 3.3: Average Kullback-Leibler distance between the Fourier (Pearson) correlation matrix at $\tau = 288$ and the correlation matrices relative to the remaining time intervals. Each correlation matrix is estimated from the same dataset of $n = 100$ synchronously sampled time series of length $T = 32,768$ for a total of 100 independent datasets. The theoretical expected value in Eq. (3.9) is also reported. The increasing level of smoothness follows the arrow direction.

Our conclusions seem to be in contrast with previous studies where the Fourier estimator is claimed to perform better than the Pearson estimator. We refer in particular to the work by Precup and Iori (2007). The authors employ real high-frequency transactions and signature plots[4] to compare the Fourier method against the Pearson coefficient and the co-volatility adjusted correlation measure proposed in Dacorogna et al. (2001). The correlation spectrum associated to the Fourier estimator and found by the authors is undoubtedly smoother than that for the other two methods, which shows instead a more erratic behavior. This has been interpreted as an index of higher precision in favor of the Fourier approach, whilst we will show here that it is simply due to a peculiar feature of the estimator.

We start by considering correlation estimates within the same sample of $p = 100$ stocks. In particular, we calculate the distance between the correlation matrix obtained at $\tau = 288$, the largest available sampling interval (smaller values of τ can be equally chosen), and each single correlation matrix calculated at the remaining values of τ ranging from 1 to 240. We then repeat the procedure for all the 100 realizations of the system. Figure 3.3 plots the average distance $\langle K[R_i(\tau = 288), R_i(\tau)]\rangle$. It is apparent as the entropy trajectory related to the Fourier estimator behaves more regularly than that one relative to the Pearson coefficient, especially at low frequencies. However, it does *not* mean that the Fourier measure outperforms the Pearson coefficient, but rather that the errors implied in the estimation procedure are more correlated at different τ's than are in the Pearson case. Let us make a simple example to clarify this point. Suppose that the true correlation between asset i and j is $\psi_{ij} = 0.3$.

---

[4]In the case of the Pearson and Fourier estimator, the signature plots are obtained by representing, respectively, the correlation coefficients (3.1) and (3.4) as a function of τ.
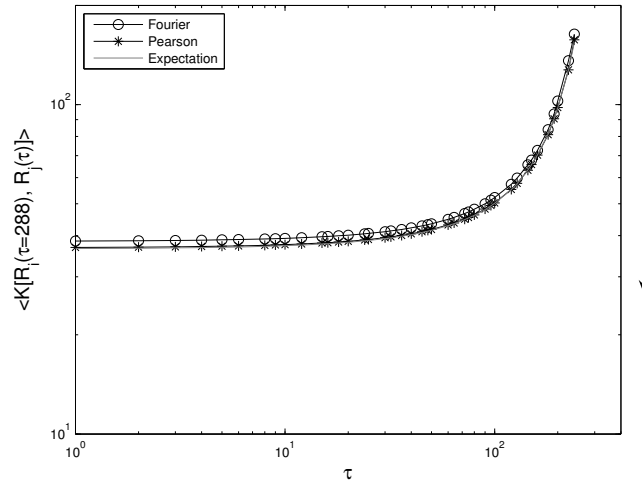
Figure 3.4: Average Kullback-Leibler distance between the Fourier (Pearson) correlation matrix at $\tau = 288$ and the correlation matrices relative to the remaining time intervals. The matrices are estimated from 100 independent datasets of $n = 100$ synchronously sampled time series of length $T = 32,768$ and the average is taken across sets. The theoretical expected value in Eq. (3.9) is also reported. The increasing level of smoothness follows the arrow direction.

Should Fourier estimates a value of $\rho_{ij}^{\mathrm{F}} = 0.32$ at $\tau = 288$, then also that at $\tau = 240$ would be likely larger than the true one. For the Pearson coefficient this is still possible but the probability of the event is smaller. In particular, given the way the coefficient is defined, it can be shown that this mainly happens for values of τ which are exact divisors of $\tau = 288$, e.g. the point $\langle K[R_i(\tau = 288), R_i(\tau = 144)] \rangle$ indicated by an arrow in Figure 3.3. Therefore, the curves in the graph are only representative of the possible connections between the estimation error relative to a certain sampling frequency and that associated to a subsequent frequency. It follows that the Fourier estimator is characterized by stronger autocorrelated errors than the Pearson coefficient. The interpretation of this distinctive behavior is twofold. From one side, it means that we can count on a certain degree of flexibility in choosing the value of the frequency N; from the other we should be aware that choosing a value for N well below the optimal, and usually unknown, level would carry an estimation error definitely higher than that generated by the opposite choice. But Figure 3.3 reveals another interesting aspect. When we add the theoretical expected value of the Kullback-Lieber divergence as calculated between *independent* realizations of the system (gray straight line), and given in Eq. (3.9), we first note that the Pearson curve gets much closer for large time horizons than Fourier's does, indeed because the autocorrelation effect is less pronounced. Secondly, we observe that increasing the sampling frequency by reducing the time scale, i.e. for small τ's, both the curves converge to the analytical value indicating that the estimation error is vanishing. This result corroborates our expectations given that both measures provide consistent estimates of the true correlation matrix Ψ.

In support of our findings, we repeat the above experiment by considering correlations between independent realizations, i.e. $K[R_i(\tau = 288), R_j(\tau)]$. The results are plotted in Figure 3.4. It is clear as both the Fourier and the Pearson methodology are now able to provide correlation matrices that, on average, match almost perfectly the theoretical expected value without showing any autocorrelation effect as in the previous case. At large frequencies, i.e. small tau's, the information carried by the data reaches the maximum level and both estimators deliver a correlation structure very close to Ψ, the predictable feature also visible in the previous figure. Note that, in both cases, the Pearson coefficient slightly outperforms the Fourier estimator for arrival times shorter than 10 seconds where the curve perfectly overlaps the theoretical value.

3.5.2 Asynchronous data

To get a deeper insight into the analysis, we induce asynchronicity in the above generated series by extracting the time intervals between consecutive transactions from an Exponential distribution with mean equal to $\lambda = 25$ seconds[5]. We aim to study the estimators behavior in the presence of the non-trading effect triggered by asynchronicity (Lo and MacKinlay, 1990). In order to apply the Pearson coefficient, the series are first interpolated over a common grid using the previous tick method. We set the grid points to be regularly spaced with a time step of 1 second. It is important to emphasize that this way of sampling prices does not eliminate the above effect but simply ensures that returns across assets are measured over matching intervals. In addition, we have seen in the previous chapter that the Fourier estimator can be equivalently applied to interpolated series of data, allowing to calculate the Fourier coefficients of price using the Fast Fourier Transform algorithm.

In Figure 3.5, above panel, a first comparison of the Fourier and Pearson estimators is given. Respect to the synchronous case, the Kullback-Leibler curves now clearly display a minimum point indicating the existence of an 'optimal' sampling rate at which both estimators maximally approach the real correlation matrix Ψ. This is due to the contribution of two effects: (i) the discretization error that generates a loss of accuracy for large time horizons; (ii) the aforementioned non-synchronous trading effect. The latter is described as a dramatic drop in the absolute value of the covariance and correlation measures when the sampling frequency is increased. The phenomenon was first reported by Epps (1979) and since then has been studied by several other authors including Dacorogna and Lundin (1999), Renò (2003), Martens (2004) and Voev and Lunde (2007). In the case of the Pearson coefficient, the Epps effect is a consequence of assuming that two time series are sampled simultaneously when instead are non-synchronous. The previous tick interpolation method used to this purpose artificially sets equal prices over periods of no trading activity that generate null returns and

---

[5]Although the distribution of the intervals is common to all the 100 stocks, we still obtain sets of asynchronous series of different length due to the random nature of the generating process.

zero cross-product values as a consequence, resulting in a correlation estimate biased downwards. Oversampling increases the contribution of these values and, therefore, exacerbates the negative bias even further. Looking back at the graph, we can see that the deterioration of the Kullback-Leibler distance for small τ's is clearly depicted: the divergence between the model and the estimator is widened and the accuracy is indeed reduced.

As far as the Fourier estimator is concerned, we first note that the null returns should not be account for the bias clearly visible in the figure. This becomes apparent when we rewrite the covariance term as follow

$$
\begin{aligned}
2\pi a_0(\Sigma_{12}) &\approx \frac{2\pi^2}{N+1} \sum_{k=1}^{N} \Big[a_k(dX_1)a_k(dX_2) + b_k(dX_1)b_k(dX_2) \Big] \\
&\approx \frac{2}{N+1} \sum_{k=1}^{N} \Bigg[\sum_{i=1}^{n} \cos(kt_i^1)\Delta X_i^1 \cdot \sum_{j=1}^{n} \cos(kt_j^2)\Delta X_j^2 + \\
&\qquad\qquad\qquad\qquad + \sum_{i=1}^{n} \sin(kt_i^1)\Delta X_i^1 \cdot \sum_{j=1}^{n} \sin(kt_j^2)\Delta X_j^2 \Bigg] \\
&\approx \frac{2}{N+1} \sum_{k=1}^{N} \Bigg[\sum_{i=1}^{n}\sum_{j=1}^{n} \Delta X_i^1 \Delta X_j^2 \Big(\cos(kt_i^1)\cos(kt_j^2) + \sin(kt_i^1)\sin(kt_j^2) \Big) \Bigg] \\
&\approx \frac{2}{N+1} \sum_{k=1}^{N} \Bigg[\sum_{i=1}^{n}\sum_{j=1}^{n} \Delta X_i^1 \Delta X_j^2 \cos(k(t_i^1 - t_j^2)) \Bigg] \\
&\approx \frac{2}{N+1} \sum_{i=1}^{n}\sum_{j=1}^{n} \Delta X_i^1 \Delta X_j^2 \sum_{k=1}^{N} \cos[k(t_i^1 - t_j^2)] \\
&= \frac{2}{N+1} \sum_{i=1}^{n_1}\sum_{j=1}^{n_2} \Delta X_i^1 \Delta X_j^2 \sum_{k=1}^{N} \cos[k(t_i^1 - t_j^2)],
\end{aligned}
\tag{3.12}
$$

where we have used the trigonometric identity $\cos(\alpha - \beta) = \cos(\alpha)\cos(\beta) + \sin(\alpha)\sin(\beta)$ and Eqs (2.7) and (2.8) in Section 2.2. In addition we have

$$
\omega_N := \sum_{k=1}^{N} \cos(kx) = \begin{cases} N & \text{if } x = 0 \\ \frac{\sin\frac{(N+1)x}{2}\cos\frac{Nx}{2}}{\sin\frac{x}{2}} & \text{otherwise.} \end{cases}
\tag{3.13}
$$

In Eq. (3.12), n_1 and n_2, where $n_1, n_2 < n = 32,768$, represent the effective length of the series after taking into account the zero returns. Hence, for every $N \geq 1$, their contribution is 'absorbed' through the calculation of the price coefficients and the covariance formula turns out to be identical to the one we would obtain by applying the Fourier method directly to the original, and not interpolated, data. The same reasoning holds true for the variance estimation.

The cross-product term $\sum_{i=1}^{n_1}\sum_{j=1}^{n_2}\Delta X_i^1 \Delta X_j^2$ also suggest that the price returns of two assets not only correlate contemporaneously, but the returns of one asset also correlate with *all* possible lead-lag returns of the other asset. In principle, the presence of leads and lags in the covariance measure should reduce the effect of non synchronicity, as thoroughly demonstrated in the work by Griffin and Oomen (2010) using a modified first-order autocovariance corrected estimator. The Fourier methods includes higher order cross terms, but these seem to be redundant and to generate unnecessary lead-lag adjustments. However, from Figure 3.5, it is clear that if N is properly chosen, for instance by means of covariance signature plots as previously suggested, than it is possible to minimize the non-trading effect. In particular, the covariance level is driven by N through the quantity ω_N, which weights the contribution of the cross increments to the above summation. In their interesting paper, Griffin and Oomen (2010) also show that the size of the negative bias due to non synchronicity is mainly determinate by the slowest asset, when prices are not contaminated by microstructure noise as in our case. This is consistent with the work by Dacorogna and Lundin (1999) where the authors find that correlation is inversely related to trading activity: the more an asset is traded, the less evident is the Epps effect. Renò (2003) employs the Fourier estimator to investigate the effect, and also provides further evidence on the above inverse relationship using both simulated and real time series. Moreover, he shows that the impact of non-synchronicity can be considerably reduced by using only synchronous transactions, i.e. tick prices that occur at the same time without any prior interpolation, although this method sacrifices a large portion of the available data as expected.

Finally, we complete our study calculating the Kullback-Leibler distance between the sample correlation matrices R obtained from both dependent and independent realizations of the system. The results are plotted in Figure 3.5, middle and below panels respectively. It is important to note that the two scenarios differs from those found in the synchronous case only for the discrepancy due to the Epps effect, visible for short time horizons. These fundings corroborate our preliminary conclusion that the Fourier and Pearson estimators are equally accurate also when more realistic market features are taken into account. The combination of asynchronicity and discretization error implies a careful choice of the sampling interval for both estimators, however, the Pearson coefficient seems to control the first effect on average slightly better than the Fourier estimator.
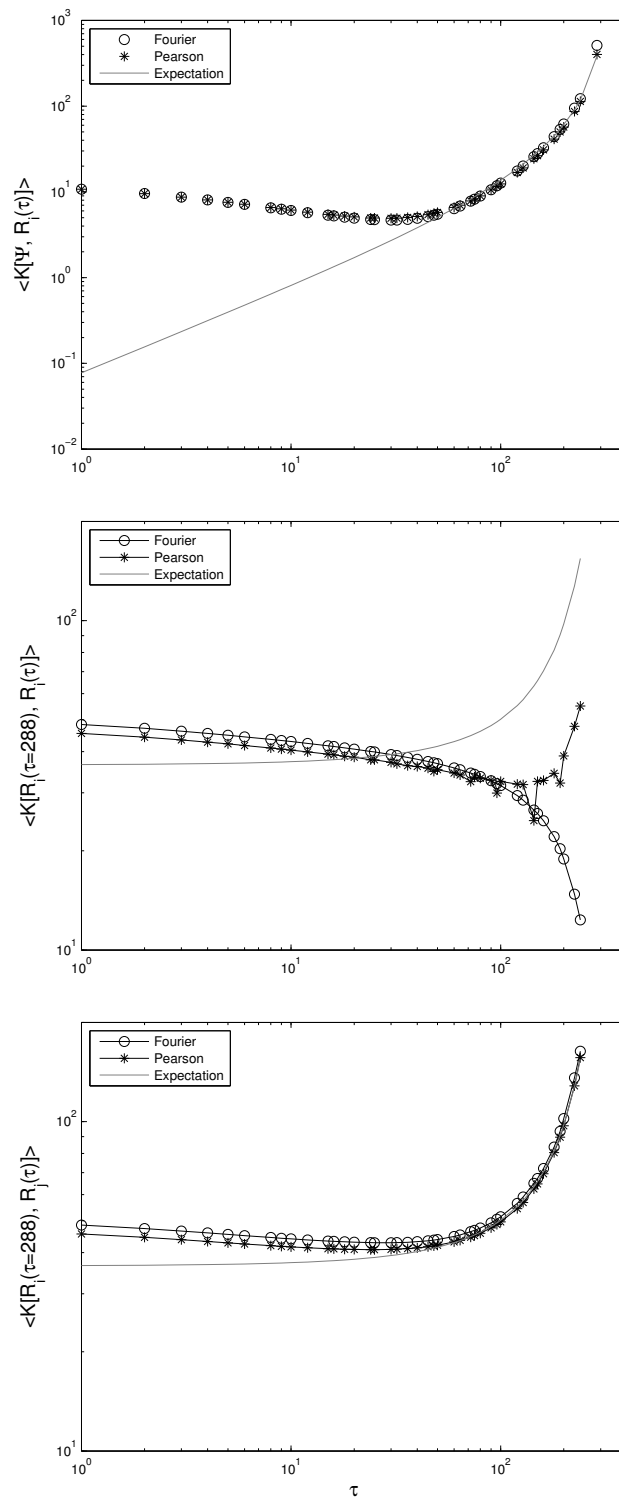
This space intentionally left blank

Figure 3.5: As in Figure 3.2 (above panel), in Figure 3.3 (middle panel) and in Figure 3.4 (below panel) but with asynchronously sampled data. The average interval time between consecutive transactions is generated from an Exponential distribution with parameter $\lambda = 25$ seconds.

3.6 The Hayashi-Yoshida covariance estimator

In a recent paper Hayashi and Yoshida (2005) propose a new covariance procedure to deal with the problem of non-synchronous trading. The resulting estimator is free of bias and can be applied directly to the observed data, with no need to first interpolate them on a regular grid. It is also showed to be consistent as the observation time intensity increases to infinity. In order to define the estimator, suppose that X and Y are two continuous semimartingales. Let $\Pi^X = \{t_i : i = 1, \ldots, n_1\}$ and $\Pi^Y = \{t_j : j = 1, \ldots, n_2\}$ denote the sets of random times at which transaction are recorded with n_1 and n_2 being the number of quotes for each asset. The results are derived under the assumption that both assets trade at $t_0 = 0$ and $t_i = t_j = 1$, where $t = 1$ is the end of the trading day[6]. Let $\Delta X_i = X(t_i) - X(t_{i-1})$ and $\Delta Y_j = Y(t_j) - Y(t_{j-1})$ be the tick-by-tick returns of X and Y respectively. The cumulative covariance estimator is then defined as

$$HY = \sum_{i=1}^{n_1} \sum_{j=1}^{n_2} \mathrm{I}\left\{(t_{i-1}, t_i] \cap (t_{j-1}, t_j] \neq \emptyset\right\} \Delta X_i \Delta Y_j, \qquad (3.14)$$

where $\mathrm{I}\{\cdot\}$ denotes the usual indicator function. If we compare this expression with Eq. (3.12), we note that the Fourier and HY estimators share a similar structure but assign different weights to the cross returns. Surprisingly, the presence of the indicator function seems to highly improve the performance of the estimator. Indeed, differently from the Fourier estimator, the product of the returns contributes to the sum as long as the corresponding intervals overlap: the tick return of the base asset, say X, is multiplied by all the possible tick returns of Y, but the cross products are limited to the interval starting before or at t_{i-1} up to or beyond t_i. Because of this particular feature, the estimator does not suffer from the Epps effect. This should also better clarify our comment regarding the Fourier estimator and the distortion generated by an overflow of leads and lags components, as discussed in the previous section. Following the intuition of Hayashi and Yoshida, we note that if ω_N in Eq. (3.13) was instead defined as

$$\omega_N := \sum_{k=1}^{N} \cos[k(t_i - t_j)] = \begin{cases} N & \text{if } (t_{i-1}, t_i] \cap (t_{j-1}, t_j] \neq \emptyset \\ \dfrac{\sin\left(\frac{(N+1)(t_i - t_j)}{2}\right)\cos\left(\frac{N(t_i - t_j)}{2}\right)}{\sin\left(\frac{(t_i - t_j)}{2}\right)} & \text{otherwise,} \end{cases}$$

the Fourier estimator would become a generalized version of HY, where the contribution of the non overlapping terms would be controlled by a correct choice of the frequency N. This is an interesting results that it would be worth exploring in a future work.

---

[6]Without this assumption, left and right censoring will induce an additional bias, which would be negligible for actively traded assets. By censoring we mean that usually assets start trading some time after the start of the trading day (left censoring) and stop trading before the end of the trading day (right censoring)

Using Eq. (3.14), a non synchronous correlation estimator is given by

$$\rho_{ij}^{\mathrm{HY}} = \frac{\sum_{i,j} \mathrm{I}\left\{(t_{i-1}, t_i] \cap (t_{j-1}, t_j] \neq \emptyset\right\} \Delta X_i \Delta Y_j}{\sqrt{\sum_i \Delta X_i}\sqrt{\sum_j \Delta Y_j}},$$

where the quantities at the denominator represent the realized volatilities calculated using raw data. Besides being consistent, Hayashi and Yoshida (2006) establish the asymptotic normality of this estimator when the correlation ρ between Brownian motions is constant.

3.6.1   Comparing Fourier and Hayashi-Yoshida estimators

The two estimators have been already analyzed by Hoshikawa et al. (2008) and Mancino and Sanfelici (2008a) in two separate comparative studies that also involve other cross-volatility measures. In the first paper the authors show that the Hayashi-Yoshida estimator outperforms the Fourier method in terms of minimum MSE and bias, but the results are restricted to a very small range of possible values for the frequency N. In addition, their study does not contemplate the presence of microstructure noise. From this prospective, the work by Mancino and Sanfelici (2008a) is more complete since it takes into account different levels of both independent and correlated noise, and N is let to vary widely, as we also did in our numerical analysis. Both the Fourier and Hayashi-Yoshida estimators are implemented using the same settings as in Voev and Lunde (2007) leading to similar conclusions as far as HY is concerned: the estimator is unbiased and quite robust to microstructure effects under independent noise, as long as this is set to a moderate level (see also Griffin and Oomen, 2010). It also performs better than the Fourier estimator both in terms of MSE and bias. However, when the noise is allowed to be correlated with the price process and serially correlated with the noise itself, although for a limited amount of time as in Bandi and Russell (2005), HY turns out to be both biased and inconsistent. Voev and Lunde (2007) propose a simply but effective way to bias correct the estimator, while they suggest to employ the subsampling technique introduced in Zhang et al. (2005) to also achieve consistency. In this respect, Mancino and Sanfelici (2008a) prove that the Fourier estimator is asymptotically unbiased in the presence of independent and correlated noise, but a 'nearly' consistency, i.e. the MSE converges to a constant value as the number of observations increases, is derived only in the case of i.i.d noise. Nonetheless, Monte Carlo simulations provide evidence that these properties are maintained also when the noise process is of more general type, once the frequency N is properly chosen. Indeed, the Fourier method delivers the best results among all the other estimators, which include HY, the realized covariance and the realized covariance with lead-lag adjustment. On the other side, the more efficient versions of the Hayashi-Yoshida estimator proposed by Voev and Lunde (2007) is showed to outperform a larger range of covariance measurements when the effects of different scenarios are investi-

gated.

It would be therefore interesting to complete the Mancino and Sanfelici study with a comparison between the Fourier method and these alternative estimators under similar conditions.

The purpose of our comparative analysis is different. Firstly, we are interested in modelling co-movements of a very large number of stocks; secondly, the estimation methods are not examined directly but by means of the Kullback-Leibler distance as previously done. In this respect, we note that the Kullback-Leibler formula, as derived in Eq. (3.5), requires the computation of the inverse of a matrix and of the logarithm of its determinant. It follows that the correlation matrices under analysis must be positive definite and, therefore, nonsingular. Unfortunately, the nonsingularity of the Hayashi-Yoshida correlation matrix is not always guaranteed: differently from the Fourier estimator, the full correlation matrix obtained by assembling all the pairwise estimates ρ_{ij}^{HY} might not be positive definite. One way to tackle this problem is to apply the method proposed in Rebonato and Jackel (1999), which generates a positive semidefinite correlation matrix that closely approximates the target symmetric, but singular matrix. However, an early stage of the procedure consists in setting to zero all the negative eigenvalues of the ill-conditioned matrix, and this would violate the requirement of positive determinant compulsory to the calculation of the Kullback-Leibler distance[7]. It is then necessary to employ a methodology that fulfills such condition, but is also able to guarantee the positive definiteness of the output correlation matrix. A possible solution is the *shrinkage* technique proposed in Ledoit and Wolf (2004b). Following this idea, we compute a convex combination between the sample correlation matrices R_F and R_{HY} obtained using the Fourier and Hayashi-Yoshida estimators respectively, and defined as

$$R_S = \eta R_F + (1 - \eta)R_{HY}, \tag{3.15}$$

where $\eta \in [0, 1]$ is called the *shrinkage constant*, a measure of the weight that is given to each estimator. In the above expression, R_F represents the 'structured' estimator R_{HY} is shrunk to, also known as *shrinkage target*. The purpose is to reduce the singularity of R_{HY} by safely pulling its small eigenvalues away from zero. This can be easily seen by taking the spectral decomposition of the two matrices.

To evaluate the efficacy of the method, we employ the 100 data sets of $n = 100$ asynchronous time series used in Section 3.5.2 and calculate the Kullback-Leibler divergence between the true correlation matrix Ψ and the shrinkage estimator R_S, letting the shrinkage constant η vary in $[0, 1]$. Note that R_F is estimated by setting the frequency N to the best value obtained with the accuracy test performed in the same section. The obtained results are shown in Figure 3.6. Only the values of the shrinkage constant for which R_S is positive definite are plotted, given that it would not be possible to calculate the relative entropy otherwise. By looking at the figure, we can see that the curve reaches a minimum point

---

[7]We recall the reader that the determinant of a square matrix is equal to the product of its eigenvalues.
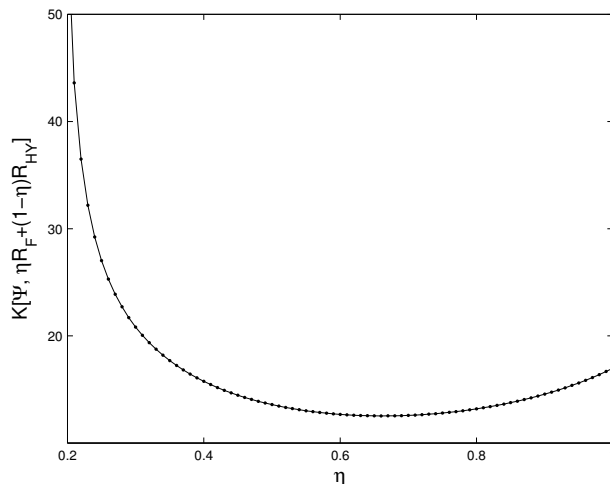
Figure 3.6: Kullback-Leibler divergence between the true correlation matrix Ψ and the shrinkage estimator $\eta R_F + (1-\eta)R_{HY}$ as a function of the shrinkage constant η. Only the values of the shrinkage constant that lead to the positive definiteness of R_S are considered.

at $\eta = 0.66$, therefore in favor of the Fourier estimator. We believe this is related to the presence of positive but almost zero eigenvalues in correspondence of small η's that clearly jeopardize the performance of the measure. Indeed, the resulting correlation matrix R_S is still invertible but numerically instable, which means that inverting the matrix amplifies the estimation error dramatically, as it is evident from the plot. As η is increased, the critical eigenvalues are pulled away from their initial minimal level and the positive contribution of the well-structured Fourier estimator slowly prevails over R_{HY}. It is important to note that, at the end, only an optimally weighted combination of the two methodologies leads to the best result, indicating that there is still space for improvement.

It would be now interesting to find a way to optimally select the shrinkage constant. At the present time, we do not have a valid solution to the problem but we believe we can rely on the estimation procedures developed in Ledoit and Wolf (2003, 2004a,b) to find one. In general, their method consists in minimizing the quadratic distance between the true and the shrinked correlation matrices, which leads to the following optimization problem

$$\min_{\eta} E\big[||\Psi - R_S||^2\big]$$

$$\text{s.t.} \quad R_S = \eta R_F + (1 - \eta)R_{HY},$$

where $||\Psi - R_S|| = \sqrt{\sum_{i=1}^{n} \sum_{j=1}^{n} (\rho_{ij} - r_{ij})^2}$ is the Frobenius distance between the two matrices $\Psi = [\psi_{ij}]_{n \times n}$ and $R_S = [r_{ij}]_{n \times n}$. It is important to emphasize that the method does not involve the inverse of any matrix and, therefore, it will not break down in case of singularity issues. Unfortunately, the optimal shrinkage constant $\hat{\eta}$ turns out to depend on the

true but unobservable correlation matrix Ψ. Ledoit and Wolf (2004b) find a way to consistently estimate this quantity obtaining a shrinkage intensity of the form $\hat{\eta} = \frac{k}{n}$, where k is a constant and n is the number of observations. In our case, we expect the intensity to be a function of the different length of the times series in use and, therefore, of the different level of asynchronicity in the data. To be more precise, the shrinkage intensity should depend, among other things, on the estimation error contained in the sample estimator: when the number of observations is high, the shrinkage intensity will tend to be small and vice versa. Our analysis employs tick-by-tick data that makes the values of n vary across assets with the consequence that pairwise correlations obtained with less observations should be shrinked more than those computed from highly liquid assets.

A naive solution to our optimal problem, would be to use $\hat{\eta}$ as derived in Ledoit and Wolf, but instead of applying the same constant to all the elements of the correlation matrix, we could construct a matrix of weights Γ with elements $\Gamma_{ij} = 1/\min\{n_i, n_j\}$, where n_i is the number of ticks for asset i. The correlation shrinkage estimator would then take the form

$$\hat{R}_S = \hat{\eta}\,\Gamma \circ R_F + (\mathrm{I} - \hat{\eta}\,\Gamma) \circ R_{HY},$$

where I is a $n \times n$ matrix of ones, while the symbol $\circ$ indicates the element-by-element Hadamard, or Schur, product. For the time being, we will postpone any test on the validity of this estimator to future work, together with a deeper analysis of the overall optimal problem.

3.7 Modelling stochastic correlation via Fourier method

Although correlation is traditionally modelled as a constant variable, it is well known in practice that is instead highly time-varying and can be even more unstable than volatility, as clearly displayed by the example in Figure 3.7. This feature is also highlighted in several empirical studies. For instance, Longin and Solnik (1995) use a GARCH model to investigate the behavior of monthly international equity returns and conclude that the resulting correlation change dynamically. They also provide evidence that correlations increase during highly volatile periods. Ball and Torous (2000) use instead filtering methods to show that the estimated correlation structure is changing over time. However, most of the existing financial econometrics research in the field is mainly based on complex multivariate GARCH models, the last cited study being an exception, which suffer from increasing parameter dimensionality and can be often estimated only after imposing severe restrictions. A breakthrough solution is represented by the Dynamic Conditional Correlation (DCC) estimator developed by Engle (2002) and characterized by the same flexibility of an univariate GARCH model. Moreover, the number of parameters implied by the model is independent of the number of series to be correlated and very large correlation matrices can be estimated. Recently,
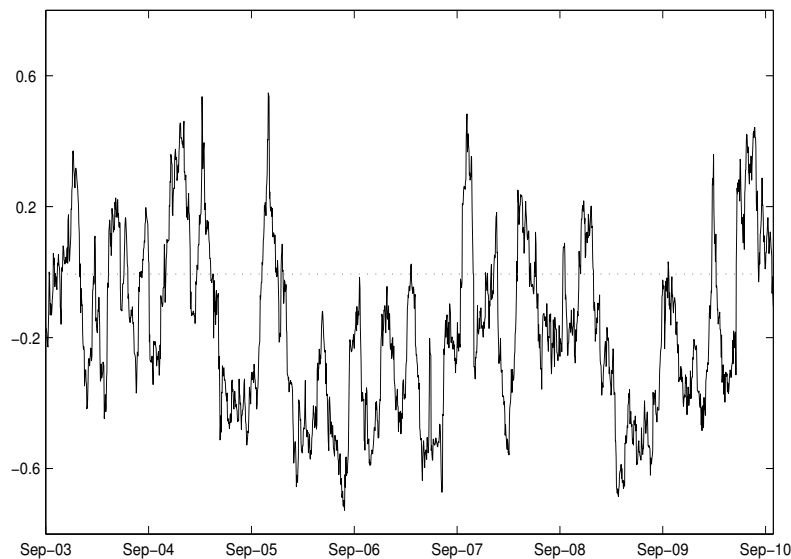
Figure 3.7: Daily correlation between the Dow Jones and the EUR/USD exchange rate estimated through the Pearson coefficient given in Eq. (3.1) using 10-minute returns.

Engle and Rangel (2008) introduce the Factor-Spline-GARCH model to generalize the DCC estimator. The model is able to captures the correlation dynamic patterns at high and low frequencies, and also allows to incorporate features of the underlying asset pricing framework and/or of the economic fundamentals. However, their approach is based on a simple on-factor CAPM model for the stock returns and becomes over-complicated for more complex models. Once we concentrate back on nonparametric modelling, we become aware that none of the existing methods concerning spot volatility estimation has been successfully extended to the multivariate case, at least to the best of our knowledge. We will now take some tentative steps in this direction by employing the Fourier method to recover the stochastic behavior of a simulated correlation path. In practice, simple methods such as rolling historical correlations and the exponential smoothing used by RiskMatrics are widely used to detect the time-varying identity of correlations.

3.7.1   Fully stochastically correlated Brownian motions

Consider two independent standard Brownian motions $W_1(t)$ and $W_2(t)$ with respect to some filtration $\mathcal{F}$. For instance, we could take the natural filtration generated by $W = (W_1, W_2)$. Assume that

$$d\rho(t) = a\left(t, \rho(t)\right) dt + b\left(t, \rho(t)\right) d\tilde{W}(t)$$

is a measurable stochastic process taking values in $[-1, 1]$, where $\tilde{W}$ is a one-dimensional Brownian motion independent of both W_1 and W_2, and a and b are two suitable random

functions. Then set

$$V(t) = W_1(t)$$

$$Z(t) = \int_0^t \rho(s)dW_1(s) + \int_0^t \sqrt{1 - \rho^2(s)}dW_2(s), \quad Z(0) = 0,$$

for every $t \in [0, T]$. Following the definition of quadratic variation, we have

$$[Z, Z]_t = \int_0^t \rho^2(s)d[W_1, W_1]_s + \int_0^t \left(1 - \rho^2(s)\right) d[W_2, W_2]_s + 2 \int_0^t \rho(s)\sqrt{1 - \rho^2(s)}d[W_1, W_2]_s$$

$$= \int_0^t \rho^2(s)ds + \int_0^t \left(1 - \rho^2(s)\right) ds$$

$$= t,$$

since the assumed independence implies that $[W_1, W_2]_t = 0$ and $d[W_i, W_i]_t = dt$, for $i = 1, 2$. Therefore, by the Lévy characterization theorem (see Musiela and Rutkowski, 1998), the process Z is also a standard one-dimensional Brownian motion with respect to $\mathcal{F}$. Moreover, the cross-variation between V and Z satisfies $d[V, Z]_t = \rho(t)dt$ given that, proceeding as above,

$$[V, Z]_t = \int_0^t \rho(s)d[W_1, W_1]_s + \int_0^t \sqrt{1 - \rho^2(s)}d[W_1, W_2]_s = \int_0^t \rho(s)ds.$$

Note that the result agrees with the case of constant correlation where the quadratic covariation would be $d[V, Z]_t = \rho\, dt$, as assumed in Section 3.2.

According to the definitions in Section 3.2, the elements of the instantaneous volatility matrix $\Sigma(t)$ are now given by $\Sigma_{ij}(t) = \rho_{ij}(t)\sigma_i(t)\sigma_j(t)$. From this, we get the instantaneous correlation matrix $R(t)$ by setting

$$R(t) = D(t)\Sigma(t)D(t),$$

where $D(t)$ is a diagonal matrix with entries $d_{ii} = \frac{1}{\sigma_i(t)}$. It follows that a Fourier estimate of the stochastic pairwise correlation $\rho_{i,j}(t)$ can be obtained as

$$\hat{\rho}_{i,j}(t) = \frac{\hat{\Sigma}_{N,M}^{ij}(t)}{\sqrt{\hat{\Sigma}_{N,M}^{ii}(t)\hat{\Sigma}_{N,M}^{jj}(t)}}, \tag{3.16}$$

where $\hat{\Sigma}_{N,M}^{ij}(t)$ is given in Eq. (3.2). It is important to note that, differently from the integrated case, it is now necessary to select three parameters, i.e. N, M and δ, for both the numerator and the denominator. We will have the chance to return on this point in Section 3.7.3 below.

3.7.2   The model

The practical example in the introductory section shows that correlation is not deterministic. This suggests that the risk implied by correlation movements is not only significant, affecting in particular optimal portfolio selection problems, but it can be also very different with respect to the risk implied by stochastic volatility, which practitioners have indeed found to be smaller. Although our conclusions should be based on further experiments, we can already infer from Figure 3.7 that: (i) correlation seems to vary stochastically around a mean; (ii) it reverts to that mean over time. We also note that the obtained trajectory never reaches the boundaries, at least in this particular case. According to Skintzi and Refens (2005) another important feature of asset return correlation is persistence: periods of high (low) correlation are likely to be followed by periods of high (low) correlation again. By introducing the implied correlation index for the DJIA, they provide evidence that the autocorrelations remains significant up to 40 days suggesting a long-run dependence in the correlation index.

To model correlation, we follow the idea in Driessen et al. (2005) and van Emmerich (2006) and choose ρ to be a mean-reverting Jacobi process defined by the following stochastic differential equation

$$dp(t) = \kappa \left(\theta - \rho(t) \right) dt + \alpha \sqrt{\left(\rho(t) - a \right) \left(b - \rho(t) \right)} d\tilde{W}(t), \quad \rho(0) = r \in [a, b], \qquad (3.17)$$

where $a < b$ are the boundaries of the process, $\kappa \geq 0$ is the rate of mean reversion, $\alpha > 0$ is the diffusion parameter and $\theta \in (a, b)$ represent the level of mean reversion. Note that the modelling of stochastic correlation through this process still admits basic autocorrelation functions of the form $e^{-\kappa \Delta t}$ (Bibby et al., 2005), which implies that the process has persistence properties, especially if the mean reversion parameter κ is low. It is now important to derive the parameters conditions that ensure the boundaries of (3.17) to be *unattainable*. Concretely speaking, the boundaries a and b are considered to be unattainable if the process does not reach them in finite time with positive probability. We want the boundaries of stochastic correlation to be unattainable in order to avoid situations where the correlation process sticks to either of its boundaries and in order to avoid degenerate correlation structures. For the specific case of our interest where $[a, b] = [-1, 1]$, van Emmerich (2006) shows that the lower boundary is attainable if $\frac{\kappa}{\alpha^2}(\theta + 1) < 1$, while the upper boundary is attainable if $\frac{\kappa}{\alpha^2}(\theta - 1) < 1$. Therefore, the boundaries of the considered Jacobi process are unattainable if and only if

$$\kappa \geq \frac{\alpha^2}{1 \pm \theta}.$$

As an alternative, we will also consider the ad hoc correlation model

$$\rho(t) = \frac{e^{2x(t)} - 1}{e^{2x(t)} + 1} \qquad (3.18)$$

introduced by Barndorff-Nielsen and Shephard (2004), where $x(t)$ follows the GARCH diffusion

$$dx(t) = \kappa(\theta - \rho(t))dt + \alpha\, x(t)dW(t), \qquad (3.19)$$

which is again characterized by a mean reverting property. This model is also employed in Voev and Lunde (2007).

3.7.3 Preliminary results

The aim of our study is to show whether the Fourier method is able to provide an estimate of the cross-volatility dynamics as claimed in Malliavin and Mancino (2002, 2009), bridging theory and practice. In particular, we will take an exploratory approach to uncover some of the practical aspects that need to be addressed in the estimation procedure. It follows that, for the time being, the results will be based on a single realization of the generating process, leaving a deeper investigation of the method to future research.

With Section 3.7.1 in mind, we start considering a stochastic correlation model of the form

$$dX_1(t) = \sigma_1 dW_1(t)$$
$$dX_2(t) = \sigma_2 dW_2(t)$$

with

$$d[W_1, W_2]_t = \rho(t)dt,$$

and where

$$d\rho(t) = \kappa\left(\theta - \rho(t)\right)dt + \alpha\sqrt{1 - \rho^2(t)}d\tilde{W}(t) \qquad (3.20)$$

is the pairwise correlation between stock X_1 and X_2 defined as the Jacobi process previously introduced. We arbitrarily set $\sigma_1 = 0.2$ and $\sigma_2 = 0.4$, while the parameters of the correlation process results from a last-square fitting of (3.20) to the historical data in Figure 3.7 under the constrain of unattainability, and are given by $\theta = -0.2$, $\kappa = 9.4$ and $\alpha = 1$. Although the underlying volatilities are chosen to be constant, we will still employ the Fourier method to get an estimate of these quantities[8]. We use the Euler-Maruyama scheme to discretized the model, and then simulate $n = 2^{13} = 8192$ tick-by-tick prices for both assets with return intervals of one second. We start by applying the Fourier estimator to the original, synchronous series. The result is showed in Figure 3.8. As expected, when prices are only contemporaneously correlated, i.e. the non-synchronous trading effect is not present, the method is able to provide a consistent estimate of the trajectory.

---

[8]A simpler but less realistic solution consists in setting the denominator of the correlation coefficient equal to $\sigma_1\sigma_2$ rather to $\sqrt{\hat{\Sigma}_{N,M}^{11}(t)\hat{\Sigma}_{N,M}^{22}(t)}$, the latter being the Fourier estimate of the former quantity.
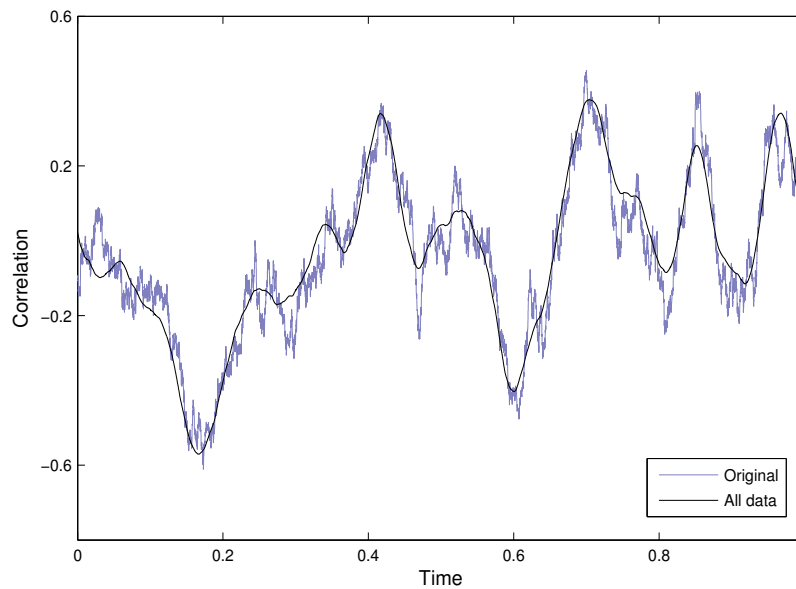
Figure 3.8: Fourier spot correlation estimate of two stochastically correlated assets using synchronous tick-by-tick prices.

To simulate asynchronous data, we proceed as in Chapter 2 and generate the inter-trade times, the so-called durations, from two independent Exponential distributions with means λ_1 and λ_2 respectively. In particular, we compare two different scenarios by first setting $\lambda_1 = 2$ and $\lambda_2 = 2$, and then $\lambda_1 = 3$ and $\lambda_2 = 6$. Given that the durations are randomly sampled, the first set of parameters implies that the two assets are trading at the same speed but *not* necessarily at the same time, while with the second combination we obtain that X_1 is trading twice as fast as X_2 leading to a higher level of asynchronicity. The previous tick imputation scheme is then employed to interpolate the missing data and to extend the sequences back to the original size n in order to employ the FFT algorithm in the calculation process. To limit the bias due to the Epps effect, we choose the frequency N or, equivalently, the number of Fourier coefficients of price, by looking at two distinct signature plots, one for each scenario, where the integrated covariances are plotted against a range of possible values for N. The plots stabilize around $N = 900$ and $N = 500$ respectively. The remaining parameters M and δ are instead selected by simply comparing the MSE of a few hand trials. Figure 3.9 shows the estimated trajectories. Although both curves are able to mirror the main features of the original correlation process, it is apparent as the quality of the reconstruction is rather affected by the level of asynchronicity. In Chapter 2 we have seen that increasing the number of Fourier coefficients would increase the precision of the final estimate. However, this gain in precision is balanced out by the bias due to the non-synchronous trading effect. If from one side the trade-off between the two implies a bias reduction, from the other we obtain an accumulation of measurement error since the limits in the Fourier formulas are approximated
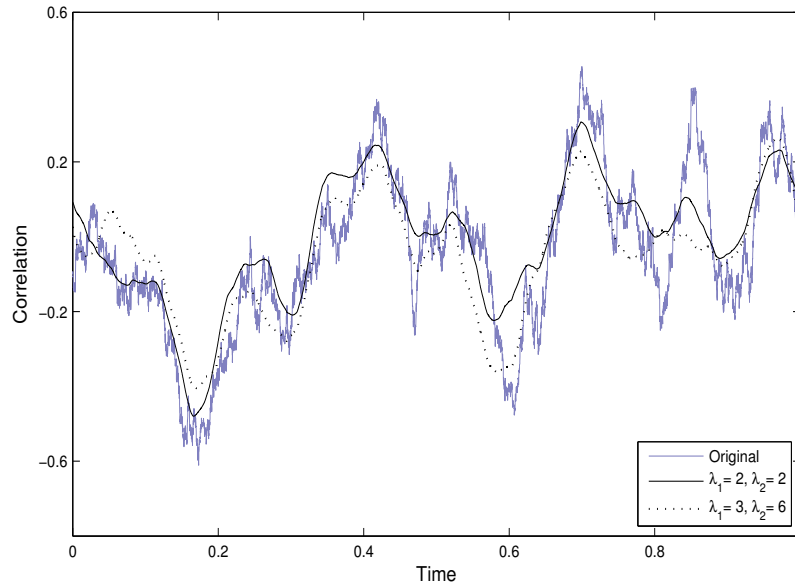
Figure 3.9: Fourier spot correlation estimate of two stochastically correlated assets through model (3.17) using non-synchronous tick-by-tick prices.

with a lower number of terms. Moreover, the latter source of error lies both at the numerator and the denominator of the correlation coefficient, leading to a weaker reconstruction respect to the one-dimensional case, even when both assets are highly traded.

Let us now consider the more realist case where both the volatilities of the underlying assets follow a stochastic process themselves. In particular, we simulate two series of tick prices using the bivariate model

$$dX_k(t) = \sigma_k(t)dW_k(t)$$
$$d\sigma_k^2(t) = \lambda_k \left(\omega_1 - \sigma_k^2(t) \right) dt + \sqrt{2\lambda_k\theta_k}\sigma_k^2(t)dW_h(t),$$

for $k = 1, 2$ and $h = 3, 4$, and where the correlation is captured through $d[W_1, W_2]_t = \rho(t)dt$. This is the bivariate continuous time limit of the well-known GARCH(1,1) model as developed in Nelson (1990), with parameter estimated in Andersen and Bollerslev (1998) using daily return times series of the DEM-USD and JPY-USD exchange rates and given by

$$\theta_1 = 0.035 \quad \theta_2 = 0.054$$
$$\omega_1 = 0.636 \quad \omega_2 = 0.476$$
$$\lambda_1 = 0.296 \quad \lambda_2 = 0.480.$$

It is worth noting that these estimates were obtained independently for each of the above time series and are here used only for illustrative purposes. The same set of parameters is employed in the Monte Carlo studies by Andreou and Ghysels (2002), Renò (2003) and
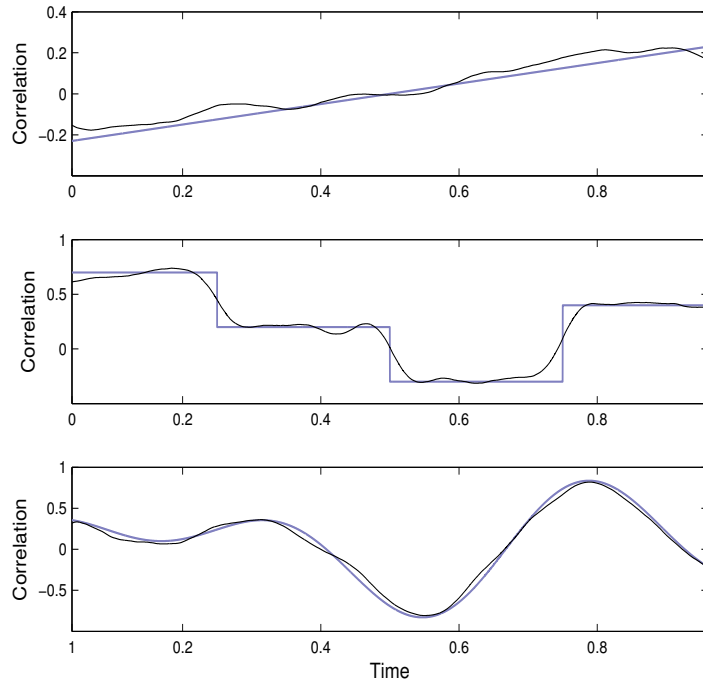
Figure 3.10: Fourier spot correlation estimate of two correlated assets through the time-varying functions in Eqs. (3.21)-(3.23) using synchronous tick-by-tick prices.

Andersen et al. (2004). We first define the correlation process $\rho(t)$ to be a simple function of time. In particular, for $t \in [0,1]$, we consider

- Linear function

$$\rho(t) = -0.2 + \frac{t}{2n} \tag{3.21}$$

- Step function

$$\rho(t) = \begin{cases} 0.7 & t < \frac{n}{4} \\ 0.2 & \frac{n}{4} \leq t < \frac{n}{2} \\ -0.3 & \frac{n}{2} \leq t < \frac{3}{4}n \\ 0.4 & \frac{3}{4}n \leq t < n \end{cases} \tag{3.22}$$

- Trigonometric function

$$\rho(t) = 0.5\big[\sin(3\bar{t}) + 0.8\cos(5\bar{t} + 0.5)\big], \qquad \bar{t} = \frac{2\pi}{n}t. \tag{3.23}$$

Figure 3.10 shows that the estimator is able to deliver a good reconstruction of the simulated trajectories in accordance with our expectations, given that the series of tick prices are synchronous. However, when we analyze the case where the correlations is a stochastic process itself, we obtain a different result, see Figure 3.11. In particular, $\rho(t)$ is the
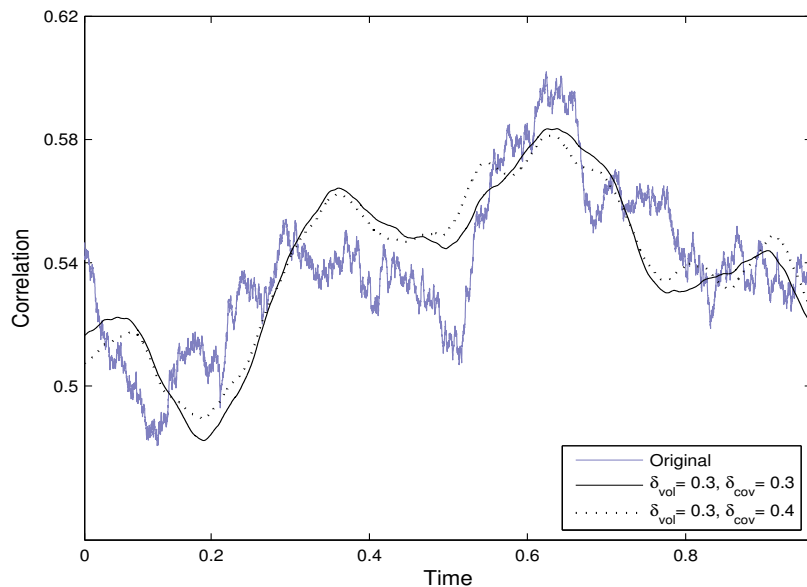
Figure 3.11: Fourier spot correlation estimate of two stochastically correlated assets through model (3.18)-(3.19) using synchronous tick-by-tick prices.

process defined in Eqs. (3.18)-(3.19) with $\kappa = 0.03$, $\delta = 0.64$ and $\alpha = 0.118$, as in Barndorff-Nielsen and Shephard (2004). Although the Fourier method is applied to the same data set, clearly the quality of the estimates is not as good as before. We first point out that the frequency N was set equal to the Nyquist level, given by $\frac{n}{2}$. This is because with synchronous prices there is no need to adjust N to account for the Epps effect. Less trivial would be the case with non-synchronous data as the computation of the cross-volatility $\hat{\Sigma}_{N,M}^{ij}(t)$ would require a lower frequency N respect to the volatility components to contrast the downward bias induced by the effect. This would result into two distinct values for N. As far as the frequency M is concerned, we remind the reader that in general M should be adjusted according to the degree of resolution δ, beside being $M << N$ (see Chapter 2, Section 2.4.2). Therefore, when the input data are synchronous, we can only improve the quality of the Fourier estimates by tuning the resolution of the covariance at the numerator and that of the variances at the denominator of the correlation coefficient, respectively defined by δ_{cov} and δ_{vol}. In Figure 3.11 we demonstrate that selecting $\delta_{cov} \neq \delta_{vol}$ can indeed result in a slightly better reconstruction as the right combination of parameters can reduce the overall estimation error. However, this is not enough: respect to the case illustrated in Figure 3.9, it is now apparent as the performance of the Fourier method is clearly compromised by the additional source of randomness related to the stochastic nature of the volatility processes[9]. Our intuition is supported by the analysis developed in this section: when the volatility is fixed to a constant level or the correlation is a simple time-varying function, the

---

[9]This can be true only in small sample analysis as the estimator is asymptotically consistent.

method can indeed reproduce the main features of the correlation dynamics, also when data are unevenly sampled. Finally, we observe that if we induced asynchronicity in the data, the non-synchronous trading effect would undermine the accuracy of the estimates even further. In addition, we would be left with the problem of choosing not only $(\delta_{cov}, \delta_{vol})$, but also (N_{cov}, N_{vol}) as previously anticipated, and (M_{cov}, M_{vol}). A simplified solution would consist in taking the same set of parameters for both the covariance and the volatilities but at the price of a potentially higher estimation error. Given the unsatisfactory results obtained with synchronous prices, we will not provide evidence on the asynchronous case.

3.8 Summary

We have first performed a comparative study on correlation estimation involving the popular Pearson coefficient and the Fourier method. The two estimators are built on a very different theoretical ground and also differ on the way they make use of the raw input data: the Pearson coefficient requires the construction of a regular grid depending on the frequency of tick arrivals, while the Fourier estimator directly and efficiently employs the observations in their original form. Nonetheless, the two approaches turned out to be very similar, in the sense that they equally convey the information about the true correlation of the system from the data under analysis. The performance is evaluated in terms of the Kullback-Leibler distance and in the presence of synchronous and asynchronous trading. In particular, when the time interval between transactions is not constant, we observed the existence of a optimal sampling frequency at which returns should be computed in order to deliver the best approximation of the true matrix.

In a second study, the Fourier estimator is compared against the Hayashi-Yoshida method, which is also designed to directly handle asynchronous data. The validity of the method is widely recognized in the econometric literature. As above, the analysis exploits the Kullback-Leibler divergence and is based on a large set of stock, but Hayashi-Yoshida matrix fails to satisfy the requirement of positive definiteness. We suggested to use the shrinkage approach to address the problem, and also showed that the best estimator in terms of minimum distance can be obtain by properly combining the Fourier and Hayashi-Yoshida estimators.

Finally, we have performed an exploratory study to evaluate the behavior of the Fourier method as applied to the estimate of a both time-varying and stochastic correlation processes. Both in the case where the volatilities of the price processes are kept constant, but the correlation is stochastic, and the case with both stochastic volatilities and correlation, the method is showed to deliver good estimates of the underlying correlation dynamics. However, when the two sources of randomness are combined together, the result is clearly unsatisfactory, despite the data are synchronous and finely sampled on a regular grid.

CHAPTER 4

Spot Volatility Estimation Using Delta Sequences

In Chapter 2 we have showed the potentials but also the shortcomings of the Fourier method as applied to spot volatility estimation. Motivated by our findings, we here introduce a new class of nonparametric estimators based on delta sequences and conceived to include many of the existing measures as special cases. The full limit theory under infill asymptotics and finite time-horizon is first derived in the pure diffusive settings. We then extend our class to include jumps or microstructure noise effects in the observed price process. As a development of our results, we study the distribution theory of the Fourier estimator and further improve the computational aspects of the algorithm. Empirical evidence from the stock index futures market is also provided.

4.1 Introduction

One way of estimating instantaneous volatility consists in assuming that the volatility process is a deterministic function of the observable state variable, and nonparametric techniques can be applied both in absence (Florens-Zmirou, 1993; Bandi and Phillips, 2003; Renò, 2008) and in presence of jumps (Bandi and Nguyen, 2003; Johannes, 2004; Mancini and Renò, 2009). Fully nonparametric methods where volatility is instead a (random) function of time include the idea of rolling sample volatility estimators in Foster and Nelson (1996), see also Andreou and Ghysels (2004), the spot volatility estimation approach of Bandi and Renò (2008), and the kernel based method of Fan and Wang (2008), Mykland and Zhang (2008), and Kristensen (2010). Alternatives are given by the Fourier estimator previously introduced, and the wavelet-based estimators in Genon-Catalot et al.

(1992) and Hoffmann (1999, 2002). In addition, Ogawa and Sanfelici (2008) propose a two-step regularization scheme designed to filter out microstructure noise, and Boudt et al. (2008) apply integrated measures to estimate the periodic component in intraday volatility. Finally, when volatility becomes the (unobservable) state variable in a bivariate process, nonparametric estimation of the functionals driving the volatility dynamics have been studied in Renò (2008) using the Fourier method, in Comte et al. (2010) using mean-square methods, and in Bandi and Renò (2008) using infinitesimal moments.

The purpose of our study is to define a large class of nonparametric estimators in order to generalize the aforementioned methods. We will follow the intuition that a spot volatility estimator can be written as the convolution of squared price returns with a function, known as *delta sequence*, which is asymptotically equivalent to the distribution of a Dirac delta function, i.e. it concentrates all the probability mass around one point in the limit. For applications of delta sequences in statistics we refer to Watson and Leadbetter (1964), and Walter and Blum (1979). In particular, Watson and Leadbetter provide sufficient conditions for a sequence of functions to be a delta sequence.

4.2 Spot Volatility Estimation in the Basic Setting

In what follows, we will consider a univariate logarithmic price process $X(t)$ defined on a filtered probability space $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{0 \leq t \leq T}, \mathcal{P})$ as specified in Chapter 1. Our results are based on the set of assumptions outlined below.

Assumption 1. *The logarithmic price $X(t)$ is the solution of the following stochastic differential equation*

$$dX(t) = \mu(t)dt + \sigma(t)dW(t), \tag{4.1}$$

where the initial condition X_0 is measurable with respect to $\mathcal{F}_0$, and $W(t)$ is a standard Brownian motion defined on the filtered probability space; the drift $\mu(t)$ and the diffusion term $\sigma(t)$ are adapted and uniformly bounded processes on $[0, T] \times \Omega$. Given a fixed point $\bar{t} \in [0, T]$, there exists a constant $C > 0$ and $0 < \gamma \leq 1$ such that

$$|\sigma^2(\bar{t}) - \sigma^2(s)| \leq C|\bar{t} - s|^\gamma \tag{4.2}$$

almost surely in a neighborhood of $\bar{t}$.

The class of processes for $\sigma(t)$ we wish to estimate point wise is larger than the class of differentiable functions, and it includes the important case where $\sigma(t)$ is generated itself by a Brownian motion as in a stochastic volatility model, see Revuz and Yor (2001) or, more recently, Kristensen (2010). Condition (4.2) imposes restrictions on the local behavior of the volatility process, allowing for deterministic patterns, nonstationarity, and is automatically

satisfied when $\sigma^2(t)$ has continuous trajectories. Moreover, it also allows for finite, but not infinite, jump activity in the volatility process itself (see Bandi and Nguyen, 2003, and the methods used therein). In this respect, the estimator we are going to propose below is robust to jumps in volatility, and so is the Fourier estimator in Section 4.4.

Denote the log-price return by $\Delta X_i = X_{t_i} - X_{t_{i-1}}$. Our proposed estimator takes the form of a discrete convolution

$$\widehat{\sigma}_{n,f}^2(\bar{t}) = \sum_{i=1}^{n} f_n(\bar{t} - t_{i-1}) \left(\Delta X_i\right)^2, \tag{4.3}$$

where $f_n(\cdot)$ is a sequence of real functions whose properties are specified below.

Assumption 2.

i) *Let D be an open subset of $\mathbb{R}$ such that $0 \in \overline{D}$. For each integer n, the real function $f_n(x) : \mathbb{R} \to \mathbb{R}$ is zero outside D, belongs to $\mathcal{L}^4(\mathbb{R})$, and is continuous and differentiable in D with $\sup_{x \in D} |f_n'(x)| \le C f_n(0)$ for a suitable constant $C > 0$ which does not depend on n.*

ii) *Define by $\mathcal{C}_t^+$ the class of stochastic processes defined on $[0, T]$ and almost surely non-negative, bounded and continuous in a neighborhood of $\bar{t} \in [0, T]$. For all $\varphi \in \mathcal{C}_t^+$, we assume that*

$$\int_0^T f_n(\bar{t} - s)\varphi(s)ds = \varphi(\bar{t}) + R_n(\bar{t}), \tag{4.4}$$

$$\frac{1}{f_n(0)} \int_0^T f_n^2(\bar{t} - s)\varphi(s)ds = c_f \varphi(\bar{t}) + o_p(1), \tag{4.5}$$

$$\frac{1}{f_n^2(0)} \int_0^T f_n^4(\bar{t} - s)\varphi(s)ds = O_p(f_n(0)), \tag{4.6}$$

where the constant $c_f \in]0, 1]$, and $R_n(\bar{t}) = o_p(1)$. The asymptotics $o_p(\cdot)$ and $O_p(\cdot)$ are defined for $n \to \infty$ (see Appendix B.1).

We first ask mild smoothness conditions on $f_n(x)$ in a neighborhood of zero to be applied in the asymptotic results derived below. Note that $f_n(x)$ is assumed to be zero outside the open subset D in order to include the indicator function, which leads to the case of realized volatility estimators. Eq. (4.4) can be reformulated with the short-hand notation

$$f_n(x) \xrightarrow[n \to \infty]{p} \delta(x), \tag{4.7}$$

where $\delta(x)$ is known as the *Dirac delta function*. It follows that $f_n(x)$ is a *delta sequence*[1]. However, we cannot use all the possible sequences since we have to impose some regularity

---

[1]See Appendix B.2 and B.3 for a formal definition of Dirac function and delta sequence.

conditions to make the estimation of spot volatility possible. In Eq. (4.5) we choose to normalize $f_n^2(x)$ by $f_n(0)$, but $f_n(0)$ can be replaced by any sequence a_n able to deliver the same result, such as $a_n = \int f_n^2(x)dx$ or $a_n = \sup_{x \in D} f_n(x)$. Eq. (4.5) also implies that $g_n(x) = f_n^2(x)/(c_f f_n(0))$ fulfills Eq. (4.4) and thus is a delta sequence as well. Eq. (4.6) is a sufficient condition for the central limit Theorem 4.3 below and is automatically met if $g_n(x)$ satisfies Eq. (4.5).

Some relevant examples of sequences $f_n(x)$ satisfying Assumption 2 are the followings. Other examples can be found in Walter and Blum (1979).

EXAMPLE 1: Trigonometric functions

The *Dirichlet kernel* is defined as

$$f_n(x) = D_{N_n}(x), \tag{4.8}$$

where

$$D_{N_n}(x) := \sum_{|h| \leq N_n} e^{ihx} = \frac{\sin\left[\left(N_n + \frac{1}{2}\right)x\right]}{\sin\frac{x}{2}}, \tag{4.9}$$

and N_n is a sequence such that $N_n/n \to 0$. In this case, $f_n(0) = 2N_n + 1$ and $c_f = 1$. If the Dirichlet kernel can be negative at some points, a positive trigonometric example is given by $f_n(x) = F_{N_n}(x)$ where $F_{N_n}(x)$ is the *Fejér kernel*

$$F_{N_n}(x) := \sum_{|s| \leq N_n} \left(1 - \frac{|s|}{N_n + 1}\right) e^{isx} = \frac{1}{N_n + 1} \left(\frac{\sin\frac{N_n+1}{2}x}{\sin\frac{x}{2}}\right)^2, \tag{4.10}$$

with $f_n(0) = N_n + 1$ and $c_f = \frac{2}{3}$. Conditions (4.4)-(4.6) can be verified using the Dirichlet and Fejér kernel's properties in the proof of Proposition 4.11.

EXAMPLE 2: Kernel functions

Consider a non-negative function $K(x)$ such that

(i) $\int_{\mathbb{R}} K(x)dx = 1$; $|x|\left|K^{(i)}(x)\right| \to 0$ as $|x| \to \infty$, $i = 0, 1$; there exists $\phi, \varphi < \infty$ such that $\left|K^i(x)\right| \leq \phi$ and, for some $v > 1$, $\left|K^i(x)\right| \leq \phi \|x\|^{-v}$ for $\|x\| \geq \varphi$, $i = 0, 1$.

(ii) $\int_{\mathbb{R}} x^i K(x)dx = 0$, $i = 1, \ldots, r-1$, and $\int_{\mathbb{R}} |x|^r |K(x)| \, dx < \infty$ for some $r \geq 0$.

These regularity conditions are satisfied by most standard kernels for $r \leq 2$. See also Appendix C for a detailed introduction to kernel functions. Then define

$$f_n(x) = \frac{1}{h_n} K\left(\frac{x}{h_n}\right), \tag{4.11}$$

where the bandwidth $h_n \to 0$. Since $f_n(0) = \frac{1}{h_n} K(0)$, we can interpret $f_n(0)$ as the inverse

of the bandwidth. Moreover, we have

$$\int_0^T \frac{1}{h_n} K\left(\frac{\bar{t}-s}{h_n}\right) \sigma^2(s) ds = \sum_{i=1}^n \frac{1}{h_n} K\left(\frac{\bar{t}-t_{i-1}}{h_n}\right) \sigma^2(t_{i-1}) \Delta_n^i + o_{a.s.}\left(\frac{\overline{\Delta}_n}{h_n}\right)$$

$$= \sigma^2(\bar{t}) + o_p\left(\frac{\overline{\Delta}_n}{h_n}\right),$$

as shown in Kristensen (2010), with $c_f = K_2/K(0)$. In particular, for the Epanechnikov kernel $c_f = 4/5$, while $c_f = 1/\sqrt{2}$ in the case of the most commonly used Gaussian kernel. In a similar way, it is possible to verify conditions (4.5) and (4.6). Note that both the Dirichlet and Fejér kernels previously introduced cannot be casted in the form (4.11).

$\underline{\text{Example}}$ 3: Estimation at boundaries
When estimating $\sigma^2(\bar{t})$ at the boundary $\bar{t} = 0$, a suitable choice is a delta sequence $f_n(x)$ of the form

$$f_n(x) = \begin{cases} 0 & \text{if } x < 0 \text{ or } x > n^{-\frac{1}{2}} \\ n^{\frac{1}{2}} & \text{if } 0 \leq x \leq n^{-\frac{1}{2}} \end{cases}$$

with $f_n(0) = n^{\frac{1}{2}}$ and $c_f = 1$.

From Assumption 2, we can easily deduce the rate of convergence of $R_n(\bar{t})$ when $\varphi(t) = \sigma^2(t)$. This is an important first result because $R_n(\bar{t})$ estimates the asymptotic bias term. Hereafter, proofs will be postponed to Appendix E.

Proposition 4.1. *Let Assumptions 1, 2 and 3 hold. Then $R_n(\bar{t}) = O_p(f_n(0)^{-\gamma})$, if $\varphi(t) = \sigma^2(t)$.*

Note that, by taking $\varphi(s) = I_{\{\bar{t}-s \in D\}}$, Eq. (4.4) implies

$$\int_D f_n(x) dx \xrightarrow[n \to \infty]{} 1,$$

and Eq. (4.5) implies that the value of the constant c_f can be recovered by

$$\frac{1}{f_n(0)} \int_D f_n^2(x) dx \xrightarrow[n \to \infty]{} c_f.$$

In order to work with irregular sampling, we adapt to our settings the concept of quadratic variation of time defined in Mykland and Zhang (2006).

Assumption 3.

i) *Suppose that the process X_t is observed at instants $0 = t_0 < t_1 < \ldots < t_n = T$, not necessarily equally spaced, and with T fixed. Then set $\Delta_n^i = t_i - t_{i-1}$ and $\overline{\Delta}_n = \frac{T}{n}$*

with $\max_{i=1,\dots,n}\{\Delta_n^i\} = O(\overline{\Delta}_n)$. The quadratic variation of time is defined as $H(t) = \lim_{n\to\infty} H_n(t)$, where

$$H_n(t) = \frac{1}{\overline{\Delta}_n} \sum_{t_i \le t} \left(\Delta_n^i\right)^2. \tag{4.12}$$

Assuming that the above limit exists, we require $H(t)$ to be a continuously differentiable function.

ii) Uniformly in $[0, T]$, we have

$$\lim_{n\to\infty} f_n(0) \left(H_n(t) - H_n\left(t - \frac{1}{f_n(0)} \right) \right) = H'(t). \tag{4.13}$$

The last condition is necessary to interchange limits and differentiability when using the quadratic variation of time. In the restrictive case of equally spaced observations, $\Delta_n^i = \overline{\Delta}_n$ and $H'(t) = 1$. When observations are more (less) concentrated around t, we have $H'(t) < 1$ ($H'(t) > 1$).

With the next theorem we derive the asymptotic distribution of the proposed volatility estimator (4.3). Preliminary to this result is the definition of stable convergence, Jacod (1997) or, more recently, Podolskij and Vetter (2010).

Definition 4.2. *A sequence of random variables G_n converges stably in law with limit G, hereafter $G_n \xrightarrow{\mathcal{L}_{st}} G$, defined on an appropriate extension $(\tilde{\Omega}, \tilde{\mathcal{F}}, \tilde{\mathcal{P}})$ of a probability space $(\Omega, \mathcal{F}, \mathcal{P})$, if and only if for any $\mathcal{F}$-measurable and bounded random variable H, and any bounded and continuous function g we have that*

$$\lim_{n\to\infty} E\left[Hg(G_n) \right] = \tilde{E}\left[Hg(G) \right].$$

Stable convergence is slightly stronger than the mere convergence in law, and it is here necessary to account for the case where $\sigma^2(t)$ is a stochastic process. In the following, we will use the $\mathbf{MN}(0, V)$ to denote a mixed normal distribution with stochastic variance V.

Theorem 4.3. *Let Assumptions 1, 2 and 3 hold. If $n, f_n(0) \to \infty$ in such a way that $f_n(0)\overline{\Delta}_n \to 0$ and $R_n = o_p(1)$, then we have*

$$\widehat{\sigma}_{n,f}^2(\overline{t}) \xrightarrow{p} \sigma^2(\overline{t}).$$

If furthermore $R_n(\overline{t}) = o_p\left(\sqrt{f_n(0)\overline{\Delta}_n} \right)$, then

$$\frac{1}{\sqrt{f_n(0)\overline{\Delta}_n}} \left[\widehat{\sigma}_{n,f}^2(\overline{t}) - \sigma^2(\overline{t}) \right] \xrightarrow{\mathcal{L}_{st}} \mathbf{MN}\left(0, 2c_f H'(\overline{t})\sigma^4(\overline{t}) \right).$$

A similar result is found in Kristensen (2010) when $f_n(x)$ is a kernel function. The central limit theorem condition $R_n(\bar{t}) = o_p\left(\sqrt{f_n(0)\overline{\Delta}_n}\right)$ can be rewritten as $f_n(0)^{1+2\gamma}\overline{\Delta}_n \to \infty$ by Proposition 4.1, and conforms to the requirement that $f_n(0)\overline{\Delta}_n \to 0$, for every $\gamma \in]0, 1]$.

Remark 4.4. *(Choice of the optimal f_n)*
The choice of the optimal f_n relies on the usual bias and variance trade-off considerations (see, Appendix C). In the proof of Theorem 4.3, we have seen that the bias of $\widehat{\sigma}^2_{n,f}(\bar{t})$ is proportional to $R_n(\bar{t})$, the latter being $O_p(f_n(0)^{-\gamma})$ by Proposition 4.1. It follows that the mean-square error optimal $f_n(0)$ is proportional to $(\overline{\Delta}_n)^{-\frac{1}{1+2\gamma}}$. Note that γ is unknown in practice. By imposing further differentiability to f_n (similar to the higher order conditions typically imposed on kernels), we find that, asymptotically, the mean integrated square error (MISE) optimal choice is the Epanechnikov kernel, i.e. a function f_n such that

$$f_n(x) = f_n(0)\left[1 - \left(\frac{4}{3}f_n(0)x\right)^2\right]^+,$$

where $x^+ = \max(x, 0)$. The form of the optimal kernel when estimating at boundaries is derived in Zhang et al. (1999).

Remark 4.5. *(Small sample correction)*
In small samples, it is advisable to use the estimator

$$\widehat{\widehat{\sigma}}^2_{n,f}(\bar{t}) = \frac{\sum_{i=1}^n f_n(\bar{t} - t_{i-1})(\Delta X_i)^2}{\sum_{i=1}^n f_n(\bar{t} - t_{i-1})\Delta^i_n}, \tag{4.14}$$

for which the asymptotic results in Theorem 4.3 are still valid given that

$$\sum_{i=1}^n f_n(\bar{t} - t_{i-1})\Delta^i_n \longrightarrow 1, \quad \text{as } n \to \infty.$$

Remark 4.6. *(Equivalent estimators)*
All the estimators of the form
$$\widehat{\sigma}^2_{n,f}(\bar{t}) + \widehat{s}_n$$
have the same asymptotics of $\widehat{\sigma}^2_{n,f}(\bar{t})$ if $\widehat{s}_n = o_p\left(\sqrt{f_n(0)\overline{\Delta}_n}\right)$ and $cov(s_n, \widehat{\sigma}^2_{n,f}(\bar{t})) \to 0$. Therefore, the class of our estimators can be defined modulo the difference of $o_p\left(\sqrt{f_n(0)\overline{\Delta}_n}\right)$ uncorrelated terms.

4.3   Estimation in presence of jumps and microstructure noise

In this section we show that, with proper adjustments, the estimator $\widehat{\sigma}^2_{n,f}(\bar{t})$ can be employed to the analysis of a more general data generating process where prices are affected by microstructure noise or can display a finite number of jumps, two important aspects that play a relevant role in the study of financial time series.

4.3.1   Robustness to microstructure effects

We assume that the logarithmic prices $X(t_i)$ are observed at discrete times $t_0, \ldots, t_n$ and are subject to an observation error due to microstructure noise.

Assumption 4. *Let*

$$X(t_i) = Y(t_i) + \varepsilon_i, \tag{4.15}$$

where $Y(t_i)$ is the unobservable efficient price satisfying Assumption 1, and ε_i denotes the noise component. The noise process $\{\varepsilon_i\}_{0 \le i \le n}$ is i.i.d. and independent of Y with $\mathrm{E}[\epsilon_i] = 0$, $V_\varepsilon = \mathrm{E}[\epsilon_i^2] < +\infty$ and $\kappa_\varepsilon = \mathrm{E}[\epsilon_i^4] < +\infty$.

Lemma 4.7. *Let Assumption 2,3 and 4 hold. If $R_n(\bar{t}) = o_p(\overline{\Delta}_n f_n(0))$ and $f_n(0)\overline{\Delta}_n \to 0$ as $n \to \infty$, then*

$$\frac{1}{\sqrt{f_n(0)\overline{\Delta}_n}} \left(\frac{1}{2}\overline{\Delta}_n H'(\bar{t})\widehat{\sigma}^2_{n,f}(\bar{t}) - V_\varepsilon \right) \longrightarrow \mathbf{N}\left(0, \frac{1}{2}c_f H'(\bar{t})\left(\kappa_\varepsilon + V_\varepsilon^2 \right) \right), \tag{4.16}$$

where the above convergence is in distribution.

It is immediate to see that the market microstructure-induced bias is given by

$$\mathrm{E}[\widehat{\sigma}^2_{n,f}(\bar{t}) - \sigma^2(\bar{t})] = \frac{2V_\varepsilon}{\Delta_n H'(\bar{t})} + o\left(\frac{1}{\Delta_n} \right), \tag{4.17}$$

which diverges at rate n. However, when appropriately corrected by a factor $\frac{1}{2}\overline{\Delta}_n H'(\bar{t})$, a consistent estimate of the noise variance can be obtained and is of the form

$$\widehat{V}_\varepsilon = \frac{1}{2}\overline{\Delta}_n H'(\bar{t})\widehat{\sigma}^2_{n,f}(\bar{t}).$$

To derive a consistent and normally distributed estimator of the spot variance, we follow the two-scale approach in Zhang et al. (2005) and propose an estimator with overlapping prices at the lower frequencies. The idea is to remove the market microstructure noise by subtracting volatility estimated at two different frequencies, low and high. Another possible method is to smooth the observed time series via pre-averaging as in Jacod et al. (2007), or to use autocovariances and a flat-top kernel as in Barndorff-Nielsen et al. (2008).

Define an integer $\bar{n} < n$. Our small sample bias-corrected estimator is defined as

$$\widehat{\sigma}_{n,\bar{n}}^{2,TS}(\bar{t}) = \overline{\sigma}_{n,\bar{n}}^2(\bar{t}) - \frac{1}{\bar{n}}\widehat{\sigma}_{n,f}^2(\bar{t}), \tag{4.18}$$

where

$$\overline{\sigma}_{n,\bar{n}}^2(\bar{t}) = \frac{1}{\bar{n}}\sum_{i=1}^{n-\bar{n}+1} f_n(\bar{t} - t_{i-1})\left(X_{t_{i+\bar{n}-1}} - X_{t_{i-1}}\right)^2 \tag{4.19}$$

is the low-frequency estimator obtained by sampling sparsely over subgrids of observations. The following Theorem shows that $\widehat{\sigma}_{n,\bar{n}}^{2,TS}(\bar{t})$ is a consistent and normally distributed estimator in the presence of microstructure noise.

Theorem 4.8. *Let Assumptions 2,3 and 4 hold. If $n, f_n(0), \bar{n} \to \infty$ in such a way that $\bar{n}f_n(0)\overline{\Delta}_n \to 0$ and $R_n(\bar{t}) = o_p(1)$, then we have*

$$\widehat{\sigma}_{n,\bar{n}}^{2,TS}(\bar{t}) \xrightarrow{p} \sigma^2(\bar{t}).$$

Furthermore, if $R_n(\bar{t}) = o_p\left(\sqrt{f_n(0)\overline{\Delta}_n\bar{n}}\right)$ and $\bar{n} = c(\overline{\Delta}_n)^{-\frac{2}{3}}$ with $c \in \mathbb{R}$, then

$$\frac{1}{\sqrt{f_n(0)(\overline{\Delta}_n)^{\frac{1}{3}}}}\left[\widehat{\sigma}_{n,\bar{n}}^{2,TS}(\bar{t}) - \sigma^2(\bar{t})\right] \xrightarrow{\mathcal{L}_{st}} \mathbf{MN}\left(0, 2c_f\left[\frac{V_\varepsilon^2}{H'(\bar{t})} + cH'(\bar{t})\sigma^4(\bar{t})\right]\right).$$

Remark 4.9. *(Choice of the optimal f_n or $\bar{n}$ under microstructure noise)*
The rate $\bar{n} \sim c(\overline{\Delta}_n)^{-\frac{2}{3}}$ in Theorem 4.8 corresponds to the optimal choice for minimizing the asymptotic variance, as for the case of integrated volatility in Zhang et al. (2005). However, other options for $\bar{n}$ are possible and still lead to the asymptotic normality but at a different rate and with a different asymptotic variance. In particular, from the proof of Theorem 4.8, it can be inferred that Proposition 2.1 still holds, and that the leading term in the bias calculation is again $f_n(0)^{-\gamma}$. Moreover, following the same reasoning as in Remark 4.4, if $\bar{n} \sim (\overline{\Delta}_n)^{-\frac{2}{3}}$, then the MISE-optimal $f_n(0)$ turns out to be proportional to $(\overline{\Delta}_n)^{-\frac{1}{3+6\gamma}}$. Instead, if $f_n(0)$ is given, the MISE-optimal $\bar{n}$ for subsampling is found when $f_n(0)^{-2\gamma} \sim \frac{f_n(0)}{\bar{n}^2\overline{\Delta}_n} + \bar{n}f_n(0)\overline{\Delta}_n$, that is, when $\bar{n} \sim \sqrt{\frac{f_n(0)^{1+2\gamma}}{\overline{\Delta}_n}}$ if $\overline{\Delta}_n f_n(0)^{1+2\gamma} \to 0$, and when $\bar{n} \sim \overline{\Delta}_n^{-\frac{2}{3}}$ otherwise.

4.3.2 Robustness to jumps

We now consider the case where a finite number of Poisson jumps is added to the stochastic process driving the state variable dynamics.

Assumption 5. *The adapted process $X(t)$ defined on $[0, T]$ satisfies*

$$X(t) = Y(t) + J(t) \tag{4.20}$$

with $dJ(t) = c_J(t)dN(t)$, where $Y(t)$ fulfills Assumption 1, $J(t)$ is a doubly stochastic Poisson process, and $N(t)$ is a non-explosive Poisson counting process whose intensity is an adapted stochastic process $\lambda(t)$. The size of the jumps occurring at times $\tau_1, \ldots, \tau_{N(t)}$ is measured by adapted, i.i.d. random variables $c_J(\tau_j)$ such that $\mathcal{P}(\{c_J(\tau_j) = 0\}) = 0 \ \forall t \in [0, T]$.

Following the approach in Mancini (2009), we define our estimator to be

$$\widehat{\phi}^2_{n,f}(\bar{t}) = \sum_{i=1}^n f_n(\bar{t} - t_{i-1})(\Delta X_i)^2 I_{\{(\Delta X_i)^2 \leq \vartheta_n\}}, \tag{4.21}$$

where $I_{\{\cdot\}}$ denotes the indicator function and ϑ_n is a suitable sequence. The aim of the threshold ϑ_n is to disentangle the discontinuous variation induced by the Poisson jumps from the continuous variation induced by the Brownian motion. Asymptotically, this happens when ϑ_n converges to zero slower than the modulus of continuity of the Brownian motion, as specified in the next Theorem. Note that ϑ_n can also be either a function of time or a stochastic process (see Mancini and Renò, 2009). Alternative options to (4.21) are the flat kernel estimator in Aït-Sahalia and Jacod (2009), or the locally averaged bipower variation proposed by Veraart (2010). Both approaches admit infinite jump activity in the data.

Theorem 4.10. *Let Assumptions 2,3 and 5 hold. If $n, f_n(0) \to \infty$ and $\vartheta_n \to 0$ in such a way that $f_n(0)\overline{\Delta}_n \to 0$, $\vartheta_n\left(\overline{\Delta}_n \log \frac{1}{\overline{\Delta}_n}\right) \longrightarrow \infty$ and $R_n(\bar{t}) = o_p(1)$, then we have*

$$\widehat{\phi}^2_{n,f}(\bar{t}) \xrightarrow{p} \sigma^2(\bar{t}).$$

Furthermore, if $R_n(\bar{t}) = o_p\left(\sqrt{f_n(0)\overline{\Delta}_n}\right)$, then

$$\frac{1}{\sqrt{f_n(0)\overline{\Delta}_n}}\left[\widehat{\phi}^2_{n,f}(\bar{t}) - \sigma^2(\bar{t})\right] \xrightarrow{\mathcal{L}_{st}} \mathbf{MN}\left(0, 2c_f H'(\bar{t})\sigma^4(\bar{t})\right).$$

This results also indicates how to choose the threshold ϑ_n. Given that the absolute value of the increments of the Brownian motion tends to zero a.s. at the same speed as the deterministic function $\sqrt{2\overline{\Delta}_n \ln\frac{1}{\overline{\Delta}_n}}$, when we find that the squared increment $(\Delta X_i)^2$ is larger than $\vartheta_n > 2\overline{\Delta}_n \ln\frac{1}{\overline{\Delta}_n}$ for small $\overline{\Delta}_n$, than it is likely that a jump has occurred.

4.4 Relation to the Fourier estimator

In this section, we show that the Fourier estimator can be a member of our class under certain conditions. This is related to the role played by the Fejér and Dirichlet kernels in the estimation procedure, although this cannot be clearly inferred from the original formula. Recall that the two kernels were defined in Example 2 of Section 4.2. Due to the large amount

of calculations required, we will consider the estimator in its complex form (see Chapter 1), which implies that the volatility estimate is now given by

$$\widehat{\sigma}_{n,N,M}^{2,F}(\bar{t}) = \sum_{|k|\leq M} \left(1 - \frac{|k|}{M}\right) \alpha_k(N) e^{ik\tau}, \tag{4.22}$$

where

$$\alpha_k(N) := \frac{T}{2N+1} \sum_{|s|\leq N} c_s(dX) c_{k-s}(dX), \tag{4.23}$$

and

$$c_s(dX) := \frac{1}{T} \sum_{j=1}^{n} e^{-is\tau_{j-1}} \Delta X_j. \tag{4.24}$$

The last quantity is the discrete Fourier transform of dX_t. Here $\tau = 2\pi\bar{t}/T$ and $\tau_j = 2\pi t_j/T$ are rescaled times. Malliavin and Mancino (2009) prove that

$$\widehat{\sigma}_{n,N,M}^{2,F}(\bar{t}) \xrightarrow{p} \sigma^2(\bar{t}),$$

when $n, N, M \to \infty$. They also provide a weak convergence result for the whole stochastic process $\sigma^2(t)$ but the asymptotic distribution properties of $\widehat{\sigma}_{n,N,M}^{2,F}(\bar{t})$ are, to our knowledge, still unknown. In order to apply the Fourier estimator, it is necessary to set the number of coefficients of the price N, and the number of volatility coefficients M. Note that both parameters are functions of n. We have seen in Chapter 2 that a typical value for the first parameter is $N = n/2$, also known as Nyquist frequency (Priestley, 1979), while the second should be chosen in a way that $\frac{M}{N} \to 0$.

The next proposition states the central limit theorem for $\widehat{\sigma}_{n,N,M}^{2,F}(\bar{t})$ in the pure diffusive settings, i.e. in absence of microstrucure noise, also specializing the weak convergence result of Theorem 1.5, Chapter 1, to the case of spot volatility.

Proposition 4.11. *The Fourier estimator given in (4.22) is such that*

$$\widehat{\sigma}_{n,N,M}^{2,F}(\bar{t}) = \widehat{\sigma}_{n,f}^2(\bar{t}) + \phi_{n,f,g}(\bar{t}) \tag{4.25}$$

where

$$\phi_{n,f,g}(\bar{t}) = \frac{2}{g_n(0)} \sum_{i=1}^{n} \sum_{j=1}^{i-1} f_n(\bar{t} - t_{j-1}) g_n(t_{j-1} - t_{i-1}) \Delta X_i \Delta X_j, \tag{4.26}$$

with

$$f_n(x) = \frac{1}{T} F_{M-1}\left(\frac{2\pi x}{T}\right) \qquad g_n(x) = \frac{1}{T} D_N\left(\frac{2\pi x}{T}\right).$$

Moreover, let Assumption 1,2 and 3 hold and N and M be sequences depending on n such that, when $n, N, M \to \infty$, we have $\frac{M}{n} \to 0$. If $\frac{M}{N} \to 0$ and $\frac{N}{n} \to 0$, then

$$\sqrt{\frac{N}{M}} \left[\widehat{\sigma}_{n,N,M}^{2,F}(\bar{t}) - \sigma^2(\bar{t}) \right] \xrightarrow{\mathcal{L}_{st}} \mathbf{MN}\left(0, \frac{4}{3}\sigma^4(\bar{t}) \right).$$

Instead, if $\frac{N}{n} \to c \in]0, \frac{1}{2}]$ then

$$\sqrt{\frac{n}{M}} \left[\widehat{\sigma}_{n,N,M}^{2,F}(\bar{t}) - \sigma^2(\bar{t}) \right] \xrightarrow{\mathcal{L}_{st}} \mathbf{MN}\left(0, \frac{4}{3}H'(\bar{t})\sigma^4(\bar{t}) + \frac{8}{3}\left(\frac{1}{2} - cH'(\bar{t}) \right)\sigma^4(\bar{t}) \right).$$

Therefore, the Fourier estimator does not belong directly to our class but it can be rearranged into the sum of two independent terms: the newly proposed volatility estimator $\widehat{\sigma}_{n,f}^2(\bar{t})$, with $f_n(\cdot)$ equal to a rescaled version of the Fejér kernel, and a cross-product term. Let us analyze the role played by this term in details.

Depending on the ratio $\frac{N}{n}$, we can distinguish two main scenarios: if $\frac{N}{n} \to 0$, then $\phi_{n,f,g}(\bar{t}) = O_p\left(\frac{M}{N}\right)$, dominating the variance and leading to a slower speed of convergence; instead, if $\frac{N}{n} \to c \in]0, \frac{1}{2}]$, then $\phi_{n,f,g}(\bar{t}) = o_p\left(\frac{M}{n}\right)$, implying a faster rate of convergence. In the latter case, the asymptotic variance reaches its minimum at $cH'(\bar{t}) = \frac{1}{2}$, which occurs when, for instance, prices are observed at regular time intervals, i.e. $H'(\bar{t}) = 1$, and $N = \frac{n}{2}$. This is in agreement with our findings in Chapter 2, where the best volatility reconstruction, in terms of minimum MSE, is indeed obtained when all the evenly sampled prices are used. We have also noted that $N > \frac{n}{2}$ would lead to aliasing effects, however, we can now assert more precisely that a larger N is actually allowed if $H'(\bar{t}) < 1$, that is, if observations are unevenly spaced and more concentrated around $\bar{t}$ than they are on average. To state it concisely, when $\frac{N}{n} \to \frac{1}{2H'(\bar{t})}$ the Fourier estimator falls within the class of estimators (4.3), with $f_n(\cdot)$ given by the Fejér kernel (4.10), plus a zero-mean term whose variance becomes negligible in the limit (see Remark 4.6). Under these settings, it is then convenient to remove the cross-product term as it leads to the same asymptotic distribution but the number of computational steps is dramatically reduced from $O(n^2)$ to $O(n)$.

Finally, we perform a simple numerical test and plot the outcomes in Figure 4.1, when $\sigma^2(t)$ is both a deterministic function of time and a stochastic process, the latter obtained using the Heston model (see Chapter 2). All the evenly sampled data are used for a total of $n = 1000$ tick-by-tick prices with a time interval of 1 second. It is apparent as the trajectory estimated with the Fourier method without cross products (circles), and the one obtained setting $N = \frac{n}{2}$ (black line) perfectly overlap, while a lower N clearly generates higher variance.

It is important to note that the conclusions from Proposition 4.11 are obtained under pure diffusive settings and might not be necessarily true when microstructure effects are considered. Following the intuition in Barndorff-Nielsen et al. (2008) for the case of integrated volatility,
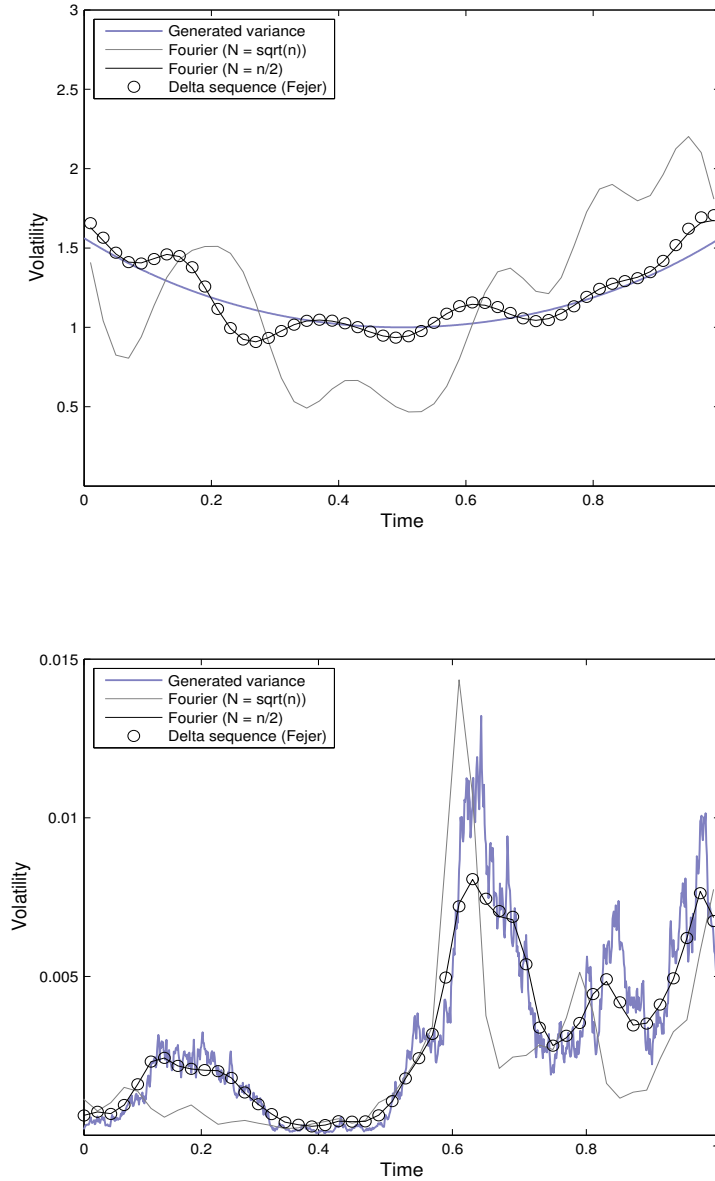
Figure 4.1: Fourier estimate of the variance of a single generated path over a sample of $n = 1000$ tick prices when $\sigma^2(t)$ is both deterministic (above panel) and stochastic (below panel). The parameters are set to $M = 8$, $N = n/2$ (black line) and $N = \sqrt{n}$ (gray line). The estimate (4.3) with the Fejér delta sequence is also displayed (circles).

then adopted in Mancino and Sanfelici (2008b), the product of disjoint increments in $\phi_{n,f,g}(\bar{t})$ could effectively act as control variates against such effects and, therefore, it might not be convenient to remove the term. It follows that our numerical study in Chapter 2 remains valid and provides useful guidelines to the finite sample behavior of the estimator.

Figure 4.2: Intraday spot volatility for the S&P500 stock index futures over one year of data calculated using the two scale estimator (4.18). Days with relevant jump activity are previously removed from the sample. The inset shows the average estimate of the microstructure noise variance V_ϵ.

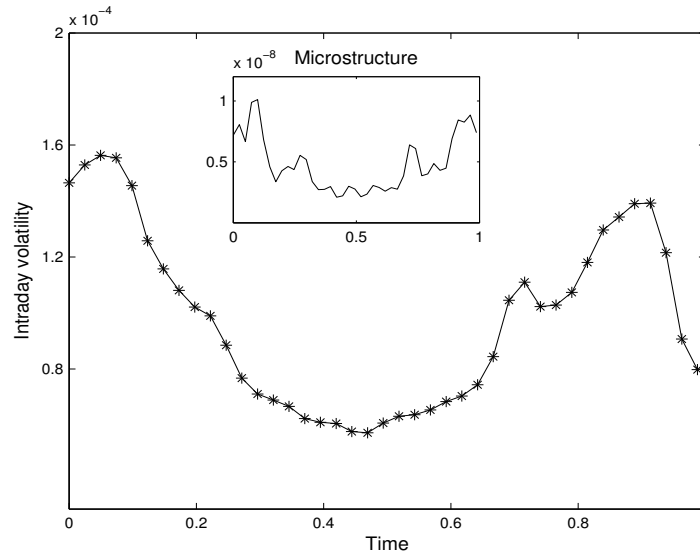## 4.5 Empirical application

In this final Section, we apply the proposed estimators (4.18) and (4.21) to the set of market data used in Chapter 2 and consisting in high-frequency transactions of the S&P 500 stock index futures. For both estimators we use the Epanechnikov kernel with $h = 15$ minutes, the latter estimated using the normal reference bandwidth selector in Eq. (C.13).

To calculate the low frequency estimator $\overline{\sigma}^2_{n,\bar{n}}(\bar{t})$ on the right-hand side of (4.18), we interpolate the data on a grid of 5 seconds using the previous-tick scheme, i.e. $X(t) = X(t_j)$ for $t \in [t_j, t_{j+1})$, and then apply a subsamplying technique similar to the one described in Zhang et al. (2005) with $\overline{n} = 12$, which corresponds to one-minute return. In order to avoid the effect of jump dynamics in the observed data, we first remove from the sample all the days characterized by significant price changes using the procedure described below. Figure 4.2 plots the estimated intraday spot volatility averaged across days and calculated in daily time units. The well known $U-$shape is clearly detected, as it was already observed in previous studies, see for instance Andersen and Bollerslev (1997). The estimate of the microstructure variance V_ε is also provided.

To show that our threshold estimator $\widehat{\phi}^2_{n,f}(\bar{t})$ is robust to price jumps, we compare it with the original spot volatility estimator (4.3) using a dataset created by removing all days with relevant jump activity. The resulting intraday volatility curves then should be almost identical. To identify the jumps, we employ the C-Tz statistics in Corsi et al. (2009). The test is similar in spirit to the z statistics of Barndorff-Nielsen and Shephard (2006) but is based

on an alternative e newly introduced estimator of realized volatility in the presence of jumps: the threshold multipower variation (see Appendix D). After setting the daily significance level of a jump to 99%, a total of 28 days are detected and then excluded from the sample. It is important to point out that we have applied the C-Tz test to series of 5 minutes returns to avoid microstructure effects. We estimate θ_n through the iterating procedure developed in Corsi et al. (2009). The result is a time-dependent threshold. The above panel in Figure 4.3 shows that the volatility curves obtained with the two aforementioned estimators match almost everywhere, meaning that that $\widehat{\phi}_{n,f}^2(\bar{t})$ is not affected by large price movements and is instead able to provide robust estimates of the intraday volatility dynamics. To highlight this feature, we apply the same estimation procedure to a sample made of the 28 days initially removed; the result in plotted in Figure 4.3, below panel. As expected, now the two curves behave quite differently, especially around the market opening time.

4.6 Summary

In this chapter, we have enlarged the class of spot volatility estimators using the peculiar property of localizing functions to converge in probability to a Dirac delta function. As a consequence, only their differentiability conditions around zero become relevant, together with the speed of convergence to the delta function. Under mild hypotheses on the data generating process, a complete asymptotic theory for the estimators within the class is provided, and suitable changes to assess the effect of microstructure noise or price discontinuities are proposed. We then applied the resulting estimators to a dataset of high-frequency stock index futures and successfully recovered the traditional U-shaped intraday volatility pattern. As a special case, we found the asymptotic distribution of the Fourier spot volatility estimator. In particular, we carefully examined the estimation methodology and showed that it can be significantly improved by removing a cross-product term. However, it remains an inefficient member of our class because the Fejér kernel is based on is not the optimal kernel.

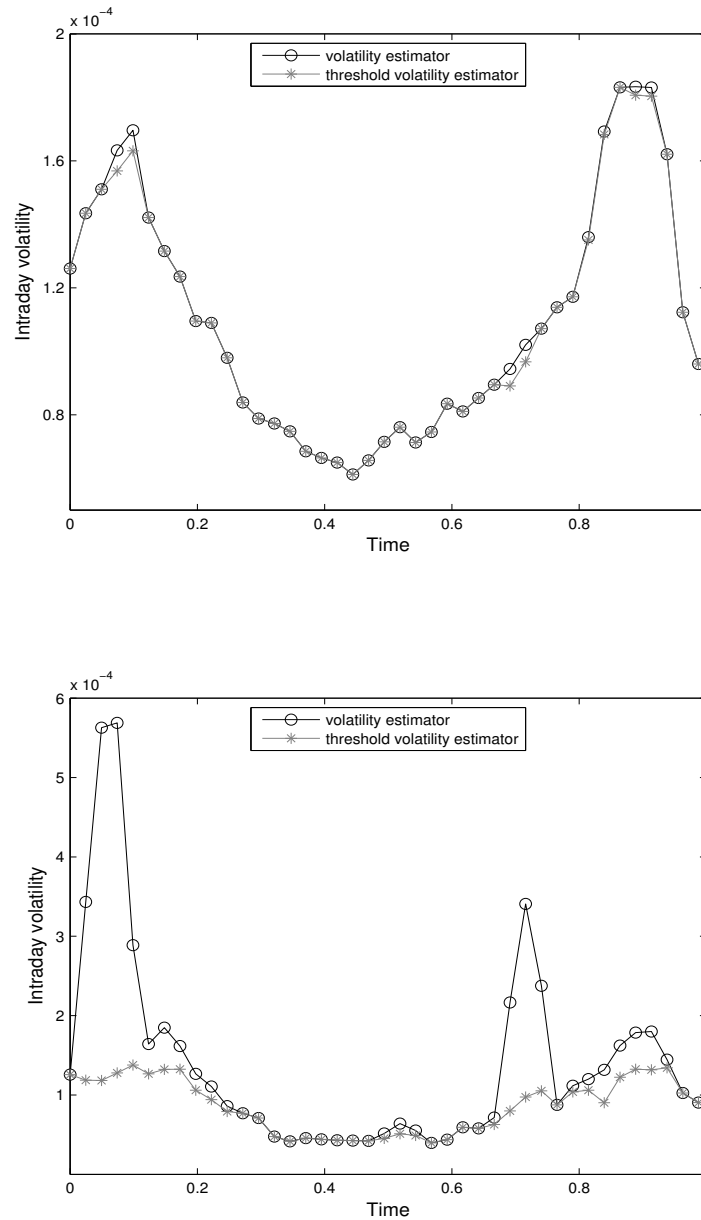Figure 4.3: Intraday spot volatility for the S&P500 stock index futures over one year of data calculated using the original volatility estimator (4.3) and the threshold estimator (4.21) respectively. Above panel: original dataset without relevant jump activity measured on daily basis. Below panel: sample made of 28 days characterized by large price movements. The significance level of jump detection is set to 99%.

Conclusions and directions for future research

When the Fourier estimator was first proposed by Malliavin and Mancino back in 2002, it was welcome as a very promising methodology to estimate volatility and correlation dynamics. In particular, the advantage of using the observed data directly into the estimation process without any prior manipulation, that it was showed to induce a bias in the estimate, has always symbolized one of its most attractive features. Although it was mainly studied by an inner circle of people, rather than being widely recognized as a nonparametric estimator, numerical and empirical evidence have highly supported its application to different financial problems. Further and more recent applications has then focused on the peculiarity of the estimator to deal with market noise effects by simply choosing a convenient number of Fourier coefficients. However, all these studies, valuable as they are, employ the procedure as integrated measure, therefore only partially exploring the potential, but also the possible shortcomings of the method. This thesis bridges the gab by applying the Fourier method to locally estimate the univariate or bivariate process driving the underlying price behavior.

Our research has emphasized new and interesting aspects of the methodology. In particular, with the numerical study on volatility estimation we provide novel and useful guidelines for setting the fundamental Fourier parameters to an appropriate level. Analytically derived optimal values would be ideal but they would also require a substantial calculation effort due to the synthesis of harmonic analysis and stochastic calculus the estimator is built on. From the study it is also clear that the estimator requires finely sampled series of data to deliver reliable volatility reconstructions, indicating that it should be mainly apply to the analysis of liquid stocks. In this way, we would obtain a better trade-off between the amount of information contained in the sample and that effectively carried into the final estimate

by a proper choice of the leading frequency N. Our empirical applications also support this conclusion. Unfortunately, when we moved to investigate the multivariate case the obtained results failed to fulfill our expectations, despite the chosen settings were in favor of the estimator. In addition, we observed that the more realistic context of asynchronous series would force to select distinct Fourier parameters for the covariance and the variance components of the correlation coefficient, leaving space to a number of possible combinations that would not be easy to handle in practice. Although our analysis is based only on simulated data, and therefore cannot completely reflect market realism, we believe it is accurate enough to let us draw the preliminary conclusion that the Fourier method should not be yet applied to multivariate spot volatility estimation using real data, unless a deeper inspection on the compounded role of the frequencies and resolution parameter is first conducted. When instead the procedure is used to estimate integrated correlation, we obtained a considerable improvement in the quality of the outcomes: the parameter space shrinks to one frequency for each asset, and so does the estimation error. On the other side, we found that the Fourier method and the plain Pearson coefficient on average perform similarly, also suggesting that the advantage of using the observed market prices directly, as mentioned above, should no longer be interpreted as a distinctive feature of the estimator. This becomes also evident when we introduce the FFT algorithm to enhance the computational time involved in the estimation process.

The above comments on the application of the method, together with the significant effort required by the implementation itself, has motivated our interest in searching alternative approaches. The new class of spot volatility estimators developed in this thesis, is characterized by a much simpler structure, and it is flexible enough to be successfully adjusted against the impact of microstructure noise and jumps. In this respect, we recall that the Fourier estimator can be made robust to the presence of market noise but it cannot be informative about the contribution of jumps in the observed price process because it simultaneously measures both the variance of the instantaneous return and the quadratic variation of the jump component. Being able to distinguish between jumps and continuous price movements can be relevant in terms of risk management and asset allocation purposes. We also proved that the estimator is a member of our class combined with an additional term, which is showed to have disruptive effects on the final results under certain conditions.

To conclude, our research has shed new light on the Fourier estimator but has also revealed some important limitations of the method. At the time we started this work, the estimator was one of the very few existing nonparametric measures able to recover ex-post market volatility and, potentially, correlation dynamics. In the meantime, we have witnessed a growing econometric literature in the field, especially around kernel-based methods, with our newly introduced class of spot volatility estimators being a very recent example. However, there is still space for improvement in both directions.

In our study involving the Kullback-Leibler distance, we have found an optimal frequency, respectively time scale, at which price returns should be evaluated in order to obtain the best estimate of the true correlation matrix. It would be now useful to derive this value analytically, also taking into account both the length of tick-by-tick series and possibly different levels of asynchronicity in the data. This would allow us to safely choose the correct number of Fourier price coefficient to insert in the estimation procedure. Another possible extension would consist in the analysis of microstructure noise effects, from which the natural step forward would be to analyze real market data. The Kullback-Leibler distance would easily accommodate the latter case since the theoretical expectation formulas are independent of the true, and unknown, correlation matrix.

The work on the shrinkage estimator was only preliminary but still very encouraging. Being able to find and analytical expression for the shrinkage constant would let us define a new correlation estimator that would be better, in terms of Kullback-Leibler distance, than both the Fourier and Pearson measures, once its components are optimally combined.

Regarding our new class of estimators, there are two possible developments along the line. First, we would like to widen our theoretical framework to include the combined effect of microstructure noise and jumps. To this purpose, the techniques developed in Jacod et al. (2007) and in Podolskij and Vetter (2009) would represent a good reference source to start from. An extension to the multivariate case would also be ideal to bridge the gap with the existing literature. This would also imply the need to adjust the estimator against the Epps effect.

Finally, the role played by the cross-product term in the equivalent expression we have derived for the Fourier estimator should be better investigate in the presence of microstructure noise. Although it is believed to positively contribute to lower the bias generated by the noise, a precise analytical study is needed to support our intuition. A good understanding of this peculiar feature would give us the confidence to further extend the application of the method. For instance, it would be interesting to estimate the variance of the volatility process, the so-called *volatility of volatility*, which is pivotal to the pricing process of many volatility-based derivatives. The Fourier procedure would allow for a direct approach to this issue since it can be easily iterated, as a result of basic trigonometric properties combined to stochastic integration.

Appendices

FFT and Zero-padding

A.1   Zero-padding: an example

In order to illustrate the effect of zero-padding in the FFT calculation, we generate a sequence of $n = 100$ stock prices from a simple Geometric Brownian Motion (GBM) model using the formula

$$X(t) = X(t-1)e^{(\mu - \frac{\sigma^2}{2})t + \sigma W(t)}, \qquad t = 1, \dots, 99$$

setting the initial price and the parameters to $x(0) = 100$, $\mu = 0.05$ and $\sigma = 0.2$ respectively. We follow the steps describe in the chapter to first create uneven transactions, and then interpolate and extend the data by zero-padding to reach three different lengths: $N = 128, 256, 512$. Figure A.1 shows the resulting log returns.

We finally apply the Fast Fourier Trasform to the above series and obtain the results plotted in Figure A.2. First note how the series has been decomposed into a linear combination of periodic waves with different frequencies. It is then apparent as the two-sided spectrum does not change shape but only gets visually denser as the sample size increases. This is because adding zeros in the time domain leads to interpolation in the frequency domain with new spectrum lines added between the original $\frac{N}{2}$ harmonic frequencies, although the information about the frequency content of the input data remains the same.
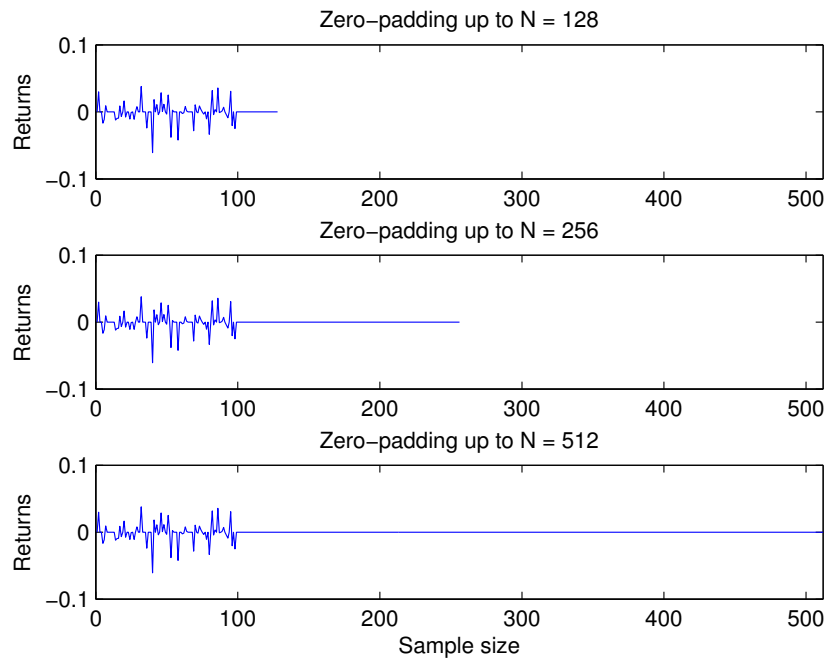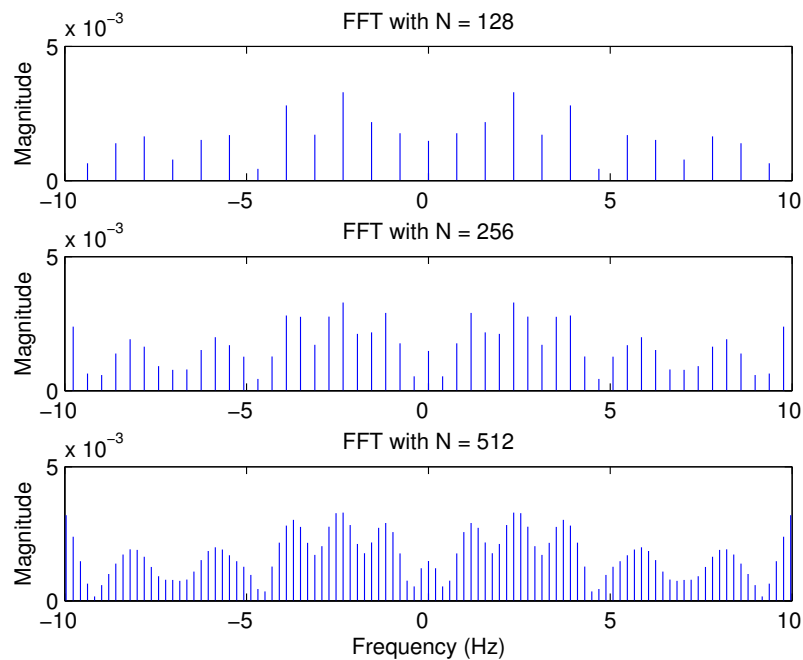
Figure A.1: Series of price returns extended by zero-padding

Figure A.2: Effect of zero-padding on the output spectrum. Only the frequency interval [-10,10] is plotted to get a clearer view.

A.2   The real valued FFT

Assume $x_1, x_2, \ldots, x_N$ is a real sequence of length N and arrange it as it was a sequence of $\frac{N}{2}$ complex numbers by taking $x_{2j} + ix_{2j+1}$. For the linearity of the DFT, we get

$$
\begin{aligned}
z_k &= \sum_{j=1}^{\frac{N}{2}} (x_{2j} + ix_{2j+1}) w_{\frac{N}{2}}^{jk} \\
&= \sum_{j=1}^{\frac{N}{2}} x_{2j} w_{\frac{N}{2}}^{jk} + i \sum_{j=1}^{\frac{N}{2}} x_{2j+1} w_{\frac{N}{2}}^{jk} \\
&=: f_k + ig_k \qquad\qquad k = 1, \ldots, \frac{N}{2}.
\end{aligned}
\tag{A.1}
$$

Then, note that

$$
f_{\frac{N}{2}-k}^* = \sum_{j=1}^{\frac{N}{2}} x_{2j}^* \left[w_{\frac{N}{2}}^{j(\frac{N}{2}-k)} \right]^* = \sum_{j=1}^{\frac{N}{2}} x_{2j} w_{\frac{N}{2}}^{jk} = f_k
$$

because x_{2j} is real, i.e. the complex conjugate of a real number is the number itself, and

$$
\left[w_{\frac{N}{2}}^{j(\frac{N}{2}-k)} \right]^* = e^{i\frac{2\pi}{N}(\frac{N}{2}-k)} = e^{i2\pi} e^{-i\frac{2\pi}{N}k} = e^{-i\frac{2\pi}{N}k} = w_{\frac{N}{2}}^{k},
$$

implied by the the periodicity of the twiddle factor. Since x_{2j+1} are also real, $g_{\frac{N}{2}-k}^* = g_k$. The complex conjugate of z_k can be now expressed in terms of f_k and g_k as shown below

$$
z_{\frac{N}{2}-k}^* = f_{\frac{N}{2}-k}^* - ig_{\frac{N}{2}-k}^* = f_k - ig_k.
\tag{A.2}
$$

Combining (A.1) and (A.2), we immediately obtain

$$
f_k = \frac{1}{2}\left(z_k + z_{\frac{N}{2}-k}^* \right) \qquad g_k = \frac{1}{2i}\left(z_k - z_{\frac{N}{2}-k}^* \right)
$$

Finally, we use these quantities to recover the original y_k values

$$
\begin{aligned}
y_k &= \sum_{j=1}^{\frac{N}{2}} x_{2j} w_{\frac{N}{2}}^{jk} + w^k \sum_{j=1}^{\frac{N}{2}} x_{2j+1} w_{\frac{N}{2}}^{jk} \\
&=: f_k + w^k g_k, \qquad\qquad k = 1, \ldots, N
\end{aligned}
$$

by direct substitution as follow

$$
y_k = \frac{1}{2}\left(z_k + z_{\frac{N}{2}-k}^* \right) + \frac{1}{2i}\left(z_k - z_{\frac{N}{2}-k}^* \right) w^k.
$$

Mathematical definitions

B.1   Order of probability notation: big-O_p and little-o_p

Given two sets of random variables X_n and Y_n, where the latter can also be a non-random sequence, the notation $X_n = o_p(Y_n)$ means that $\frac{X_n}{Y_n} \overset{p}{\to} 0$, as n approaches an appropriate limit. It can be equivalently written as $X_n = Y_n o_p(1)$. It follows that the statement $X_n = o_p(1)$ implies $X_n \overset{p}{\to} 0$. The notation $X_n = O_p(Y_n)$ instead means that the set of values $\frac{X_n}{Y_n}$ is bounded in probability.

B.2   The Delta function

The delta function was first introduced by Dirac (1930) as a technical device in the mathematical formulation of quantum mechanics. It is defined as

$$\delta(t) = \begin{cases} 0 & \text{if } t \neq 0 \\ \infty & \text{if } t = 0 \end{cases}$$

with

$$\int_{-\infty}^{\infty} \delta(t)dt = 1.$$

Dirac referred to $\delta(t)$ as an improper function because is not a function in the usual mathematical sense: if a function is zero everywhere except at a single point, its integral is necessarily zero and not one. However, it can be regarded as the limit of a sequence of ordinary functions

and becomes meaningful when it is integrated against other functions. For this reason the Dirac's delta is rather known as *generalized* function, whose properties can be very different from the properties of an ordinary function. For instance, a generalized function is always differentiable and has derivatives of all orders.

By looking at the definition of delta function given above, it should be clear that only the behavior of the function in the neighborhood of $t = 0$ really matters. Alternatively, it can be defined as

$$\delta_\epsilon(t) = 0 \quad \text{if } |t| > \epsilon, \qquad \int_{-\epsilon}^{\epsilon} \delta_\epsilon(t)dt = 1$$

where $\epsilon > 0$ and $\delta_\epsilon(t) \to \infty$ within a small-sized interval of $t = 0$. This clearly represents the delta function as an infinitely high pick in a neighborhood of $t = 0$ with the area under the curve equal to 1.

B.2.1 Basic properties

If $f(t)$ is an arbitrary function which is bounded, integrable and continuous at $t = 0$, then

$$\int_{-\infty}^{\infty} \delta(t)f(t)dt = f(0).$$

Since the delta function takes only a large value at $t = 0$ and is zero otherwise, the result of multiplying $\delta(t)$ by a function $f(t)$ is simply the value of function itself at zero. The integration is needed to deal with the infinite height and infinitesimal width of the unit point mass. Therefore, we could informally write $\delta(t)f(0)$ instead of $\delta(t)f(t)$ and pull $f(0)$ outside the integral getting the final result. This important relation is called the *sifting property* and it can be easily generalized to the case where $t = a$, for any constant a, as follow

$$\int_{-\infty}^{\infty} \delta(t - a)f(t)ds = f(a).$$

The other properties are:

- $\delta(-t) = \delta(t) \quad \text{or} \quad \delta(a - t) = \delta(t - a)$

- $t\delta(t) = 0 \quad \text{or} \quad (t - a)\delta(t - a) = 0$

- $\delta(at) = |a|^{-1}\delta(t), \quad (a \neq 0).$

B.3 The Delta Sequence

By a delta sequence we mean any sequence of n continuous and differentiable functions $\varphi_n(t)$ defined on an open interval $J \in \mathbb{R}$ with the following properties:

(i) $\int_{-\infty}^{\infty} \varphi_n(t)dt = 1$ for all $n \in \mathbb{N}$

(ii) There is a sequence of positive numbers $\alpha_n \to 0$ such that

$$\varphi_n(t) = 0 \quad \text{for} \quad |t| \geq \alpha_n, \ n \in \mathbb{N}$$

(iii)

$$\lim_{n \to \infty} \int_{-\infty}^{\infty} \varphi_n(t - a)f(t)dt = f(a)$$

Therefore, in the limit the sequence becomes a delta function. Generally, a delta sequence can be recognized by observing that the functions are increasingly sharply picked around the point $t = a$.

Kernel functions overview

Suppose we are given a realization $X_1, \cdots, X_n$ of an unknown process with marginal distribution $f(x)$, and we wish to estimate f at the point x_0. Since the density is the derivative of the cdf $F(x_0)$, we have

$$
\begin{aligned}
f(x_0) &= \lim_{h \to 0} \frac{F(x_0 + h) - F(x_0 - h)}{2h} \\
&= \lim_{h \to 0} \frac{\Pr(x_0 - h < x < x_0 + h)}{2h}
\end{aligned}
$$

Given our sample of size n, it is possible to use the estimator

$$
\hat{f}(x_0) = \frac{1}{n} \sum_{i=1}^{n} \frac{1}{h} \, \mathrm{I} \left(x_0 - \frac{h}{2} < X_i < x_0 + \frac{h}{2} \right) \tag{C.1}
$$

where $\mathrm{I}(\cdot)$ denotes the usual indicator function. The estimator $\hat{f}(x_0)$ is an histogram centered at x_0 with bin width $2h$. Since the intervals are not overlapping and the number of observations in each interval is counted and then divided by the sample size N, we necessarily obtain a estimator that satisfies the based requirement for a density: it sums to one. Although the histogram represents the easiest and most popular density estimator, it retains some undesirable properties that can be summarized as follow:

- The resulting density estimate is a step function, even if the underlying density is continuous, with jumps at the boundaries of the bins. It is not differentiable at the jumps and has zero derivative elsewhere.

- Sample size affects the relative frequencies and the visual appearance of the histogram.

- The shape of the histogram crucially depends on the bandwidth, i.e. on the bin width. There are several objective ways to select a bin width Silverman (1986). However, even in the ideal case where we somehow know in advance that the true distribution is, for instance, bell-shaped, there are additional proposed rules for the appropriate number of bins according to the sample size and, in some cases, also to other factors Scott (1992).

- Histograms vary according to the locations of the bin boundaries. This is clearly illustrated in Figure C.1 where the two histograms (panel (a) and (b)) have the same bin width and the same number of bins but different location of the left and right boundaries of the bins. We situation does not improve when we increase the number of the bins (panel (c) and (d)).
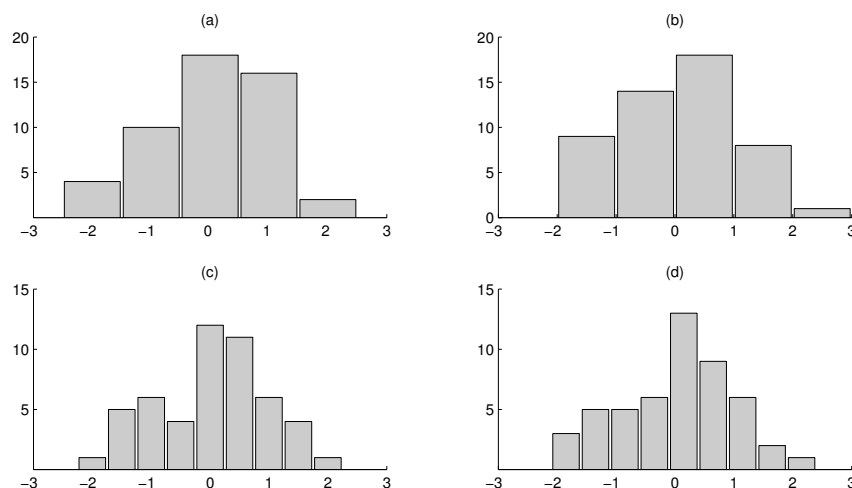
Figure C.1: Effect of locating different bin boundaries on visual appearance of an histogram given a random set of data. Reducing the bin width from 1 (panel (a) and (b)) to 0.5 (panel (c) and (d)) does not make any improvement.

The kernel density estimator introduced by Rosenblatt (1956), generalizes the histogram estimate (C.1) by using an alternative weighting function as follow

$$\hat{f}_h(x) = \frac{1}{n} \sum_{i=1}^{n} \frac{1}{h} K\left(\frac{x - X_i}{h}\right), \tag{C.2}$$

where $K(\cdot)$ is a *kernel function* and h is a *bandwidth* parameter representing the window size. Figure C.2 shows how, by introducing a kernel function, the probability mass $\frac{1}{n}$ at each observed point (the red cross) is smoothly redistributed to its vicinity. The bandwidth h is the standard deviation of the Gaussian density functions plotted under the final density estimate.

Figure C.2: Kernel density estimation spreads the probability mass at each point, depicted by a cross, to its vicinity and adds the redistributed masses together to obtain the final estimate.

Albeit rather crude, the histogram estimator can be casted in the form (C.2) by considering the simple case of a Uniform kernel function with $K(u) = \frac{1}{2}\mathrm{I}(|u| \leq 1)$. Then,

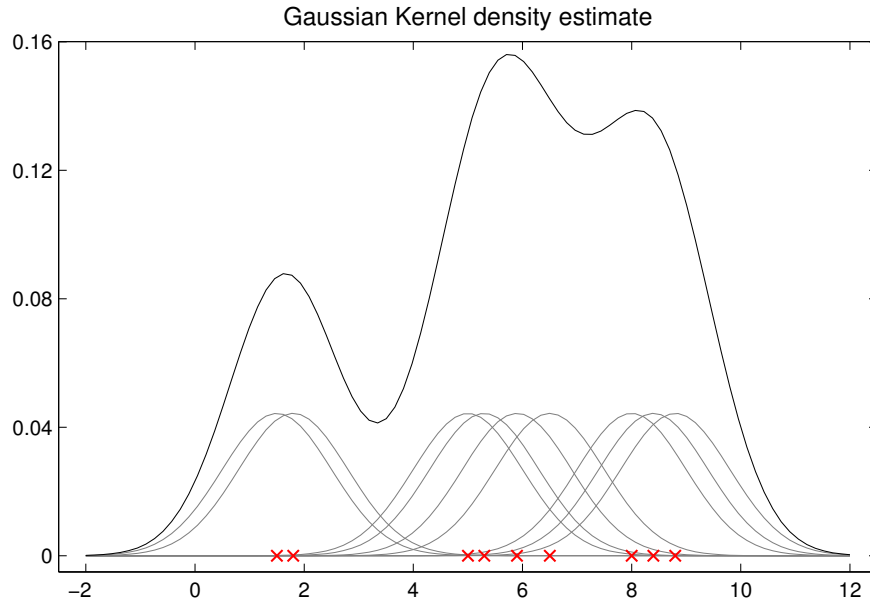$$\hat{f}_h(x) = \frac{1}{nh} \sum_{i=1}^{n} \frac{1}{2} \mathrm{I}\left(\left|\frac{x - X_i}{h}\right| \leq 1\right), \tag{C.3}$$

which is another way to rearrange Eq. (C.1). Needless to say, also this estimator is limited by the several shortcomings listed above. In particular, it is clear how the Uniform kernel function assigns equal weight $\frac{1}{2}$ to each observation X_i no matter its vicinity to x (the point at which we want estimate the pdf). In addition, points farther away from x get zero weight due to the definition of indicator function. A more suitable choice would be a continuous function able to weight the contribution coming from data points close to x more than that coming from distant observations. For example, consider

$$\hat{f}_h(x) = \frac{1}{2nh} \sum_{i=1}^{n} \frac{3}{2} \left[1 - \left(\frac{x - X_i}{h}\right)^2\right] \mathrm{I}\left(\left|\frac{x - X_i}{h}\right| \leq 1\right) \tag{C.4}$$

$$= \frac{1}{n} \sum_{i=1}^{n} \frac{1}{h} K\left(\frac{x - X_i}{h}\right), \tag{C.5}$$

where $K(\cdot)$ is the shorthand notation for a different weighting function, the well-known *Epanechnikov* kernel

$$K(u) = \frac{3}{4}(1 - u^2)\mathrm{I}(|u| \leq 1).$$

Other commonly used kernels are listed in Table C. These are continuous, non-negative, symmetric functions and share the following additional properties

(i) $\int_{\mathbb{R}} K(u)du = 1$ and $\int_{\mathbb{R}} uk(u)du = 0$

(ii) $\int_{\mathbb{R}} |k(u)|du < \infty$ and $\int_{\mathbb{R}} u^2 k(u)du < \infty$

(iii) (a) $|u|K(u) \to 0$ as $|u| \to \infty$
(b) $K(u) = 0$ if $|u| \le u_0$ for some u_0

The last property refers to the effective support of the kernel function. Therefore, even with the same bandwidth h, different kernels use different amounts of information provided by the local data points around x (see Figure C.3 for an illustration).

Kernel	$K(u)$
Triangle	$(1 - |u|)\mathrm{I}(|u| \le 1)$
Biweight	$\frac{15}{16}(1 - u^2)^2 \mathrm{I}(|u| \le 1)$
Triweight	$\frac{35}{32}(1 - u^2)^3 \mathrm{I}(|u| \le 1)$
Cosine	$\frac{\pi}{4}\cos\left(\frac{\pi}{2}u\right)\mathrm{I}(|u| \le 1)$
Gaussian	$\frac{1}{\sqrt{2\pi}}\exp\left(-\frac{1}{2}u^2\right)$

Table C.1: Kernel functions.

In using a kernel estimator, the choice of the bandwidth plays a crucial role respect to the choice of the kernel function. Once the bandwidth is optimally chosen, the form of kernel does not affect the performance of the estimator, as long as the kernel functions are both symmetric and unimodal. This is evident from Figure C.4 where the kernel density for one day of data of the S&P 500 stock index is obtained using different bandwidths. If the optimal h is calculated by Eq. (C.11) below, the resulting curve (black solid line) appears to be quite similar both in the case of Gaussian and Epanechnikov kernel. With a smaller bandwidth, less local data points are available to reduce the variance of the estimate and the curve may exhibit unrealistic multimodalities and picks (gray solid line). On the contrary, when a larger bandwidth is applied, the curve appears to be oversmoothed (dashed line), hiding the fine structure of the underlying distribution and leading to an highly biased estimate.

C.1 Bandwidth selection

It is possible to select the appropriate bandwidth by simply minimizing the mean square error (MSE) respect to h. Following the usual bias and variance decomposition, we have

$$
\begin{aligned}
\mathrm{MSE}(x) &= E[\hat{f}_h(x) - f(x)]^2 & \text{(C.6)} \\
&= \mathrm{Bias}[\hat{f}_h(x)]^2 + \mathrm{Var}[\hat{f}_h(x)]. & \text{(C.7)}
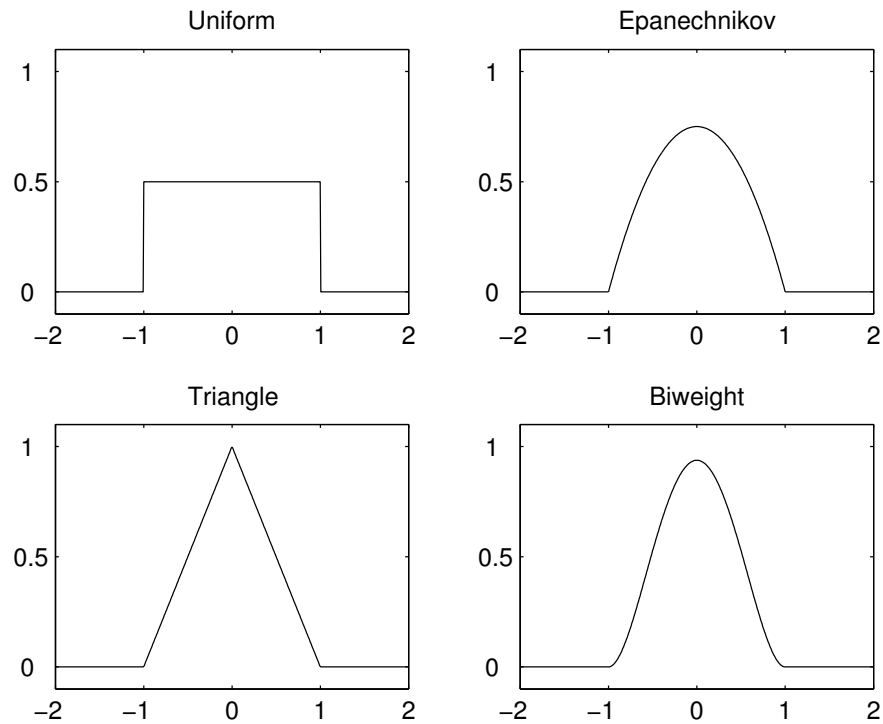\end{aligned}
$$
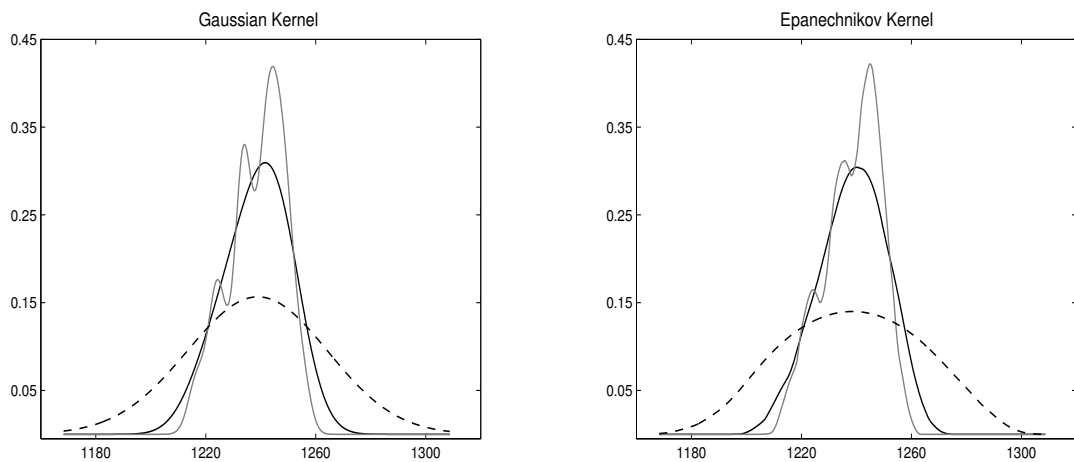
Figure C.3: Commonly used kernels.

Figure C.4: Estimated distributions for the S&P 500 stock index (1 day of data) using a Gaussian (left) and an Epanechnikov (right) kernel with a small (gray solid line), large (dashed line) and an optimally chosen (solid line) bandwidth.

The bias is simply given by

$$
\begin{aligned}
\mathrm{Bias}[\hat{f}_h(x)] &= E[\hat{f}_h(x)] - f(x) \\
&= \frac{1}{n}\sum_{i=1}^{n} E[K_h(x - X_i)] - f(x) \\
&= \int_{\mathbb{R}} \frac{1}{h}K\left(\frac{x-u}{h}\right) f(u)du - f(x)
\end{aligned}
$$

where we have used the short-hand notation $K_h(u) = \frac{1}{h}K(\frac{u}{h})$. With the change of variable $s = \frac{u-x}{h}$, the symmetry of the kernel function, i.e. $K(-s) = K(s)$ and the second order Taylor expansion of $f(u)$ around x, it can be shown that

$$
\mathrm{Bias}[\hat{f}_h(x)] = \frac{h^2}{2}f''(x)\mu_2(K) + o(h^2), \quad \text{as } h \to 0, \tag{C.8}
$$

provided that f has a continuous second derivative. Here we denote $\mu_2(K) = \int_{\mathbb{R}} u^2 K(u)du$ the variance of K. Since the bias is proportional to h^2, it can be reduced by choosing a small bandwidth or, equivalently, a kernel function with a small variance. It also depends on $f''(x)$, the curvature of the density at x and, therefore, it reaches its highest at the peaks of the distribution. Similarly for the variance we have

$$
\begin{aligned}
\mathrm{Var}[\hat{f}_h(x)] &= \mathrm{Var}\left[\frac{1}{n}\sum_{i=1}^{n} K_h(x - X_i)\right] \\
&= \frac{1}{n^2}\sum_{i=1}^{n} \mathrm{Var}[K_h(x - X_i)] \\
&= \frac{1}{n}\mathrm{Var}[K_h(x - X)] \\
&= \frac{1}{n}\{E[K_h^2(x - X)] - E[K_h(x - X)]^2\}.
\end{aligned}
$$

Using

$$
\begin{aligned}
E[K_h(x - X)]^2 &= E[\hat{f}_h(x)]^2 = O(1) \\
\frac{1}{n}E[K_h^2(x - X)] &= \frac{1}{nh^2}\int_{\mathbb{R}} K^2\left(\frac{x-u}{h}\right) f(u)du,
\end{aligned}
$$

a change of variable and Taylor expansion arguments as before, we obtain

$$
\mathrm{Var}[\hat{f}_h(x)] = \frac{1}{nh}\|K\|_2^2 f(x) + o\left(\frac{1}{nh}\right), \quad \text{as } nh \to \infty \tag{C.9}
$$

where $\|g\|_2^2 = \int_{\mathbb{R}} g^2(u)du$ is the L^2-norm (for a formal proof see Fan and Yao (2003), Theorem 5.1). It follows that, given the sample size n, we have to choose a fairly large h in order

to make the variance small. We also obtain the familiar result that the variance decreases proportionately with the size of the sample as, for a fixed value of h, increasing n will decrease the factor nh^{-1}. Finally, substituting (C.8) and (C.9) into (C.6) gives

$$\text{MSE}(x) = \left[\frac{h^2}{2} f''(x)\mu_2(K)\right]^2 + \frac{1}{nh} \|K\|_2^2 f(x) + o\left(h^2 + \frac{1}{nh}\right). \tag{C.10}$$

If we look at this formula, it is clear that we face the common trade-off between variance and bias: we surely want to keep them small but increasing h will lower the variance and rise the bias, while decreasing h will lead to the opposite effect. The optimal choice of h, obtained by minimizing $\text{MSE}(x)$, would then represents a compromise between under and oversmoothing. To this purpose, we will refer to the mean integrated square error (MISE) instead of MSE, the former having the advantage of being a global rather than a local measure of the estimation accuracy, i.e. we are interested in tailoring the bandwidth for all points x in the domain of $f(x)$ and not for a specific point x. The MISE is defined as follow

$$\begin{aligned}
\text{MISE} &= \int_{\mathbb{R}} \text{MSE}(x) \\
&\approx \frac{1}{nh} \|K\|_2^2 \int_{\mathbb{R}} f(x)dx + \frac{h^2}{4}\mu_2^2(K) \int_{\mathbb{R}} [f''(x)]^2 dx \\
&\approx \frac{1}{nh} \|K\|_2^2 + \frac{h^2}{4} \|f''\|_2^2 \mu_2^2(K).
\end{aligned}$$

Minimizing the asymptotic MISE with respect to h gives the *optimal bandwidth*

$$h_{\text{opt}} = \left(\frac{\|K\|_2^2}{n \|f''\|_2^2 \mu_2^2(K)}\right)^{\frac{1}{5}}. \tag{C.11}$$

The resulting minimum MISE is

$$\text{MISE}(h_{\text{opt}}) = \frac{5}{4} \left(\|K\|_2^2\right)^{\frac{4}{5}} \left(\mu_2(K) \|f''\|_2\right)^{\frac{2}{5}} n^{-\frac{4}{5}}. \tag{C.12}$$

Note that, since $h_{\text{opt}} = O(n^{-\frac{1}{5}})$, the consistency conditions are well satisfied, namely $h_{\text{opt}} \to 0$ and $nh_{\text{opt}} \to \infty$ as $n \to \infty$. The quantity nh is sometimes called the "the effective sample size" and the requirement that $nh \to \infty$ as $n \to \infty$ simply means that, as we get more information, we can average over a narrower region ($h \to 0$) but the amount of local information must increase at the same time. When both bias and variance goes to zero the kernel density estimator $\hat{f}_h(x)$ is consistent, i.e. converges in mean square to f. In particular, if $h = cn^{-\frac{1}{5}}$ for some $c > 0$ as above, we can achieve a consistency rate of $n^{-\frac{2}{5}}$ by taking the square root of the minimum MISE in Eq. (C.12).

Having obtained the appropriate bandwidth conditional on the choice of a particular form for $K(u)$, it can be shown that the optimal kernel function, in terms of minimal MISE, is

the Epanechnikov kernel (C.4). A derivation of this result can be found in Pagan and Ullah (1999) or in Fan and Yao (2003), Theorem 5.5. However, several studies indicate that the difference between the value of $\mathrm{MISE}(h_{\mathrm{opt}})$ attained by most kernels and the Epanechnikov kernel is actually very small (see Silverman (1986), Table 3.1 and Fan and Yao (2003), Table 5.1), resulting in a range of estimators having similar relative efficiency levels. Therefore, it is possible to select the kernel function by simply taking into account computational aspects and its simplicity, the Gaussian kernel being a popular choice.

Although h_{opt} is theoretically desirable, it depends on the unknown underlying density through the functional $\|f''\|_2^2$ and can not be applied in practice. A frequently used method to estimate this quantity consists in assuming that the unknown density f belongs to the Gaussian family of distributions with mean μ and variance σ^2. Then we have

$$\|f''\|_2^2 = \frac{1}{\sigma^5}\frac{3}{8\sqrt{\pi}} \approx \frac{0.212}{\sigma^5},$$

where σ can be replaced by the sample standard deviation s. To apply (C.11) we also need to choose a kernel function. For instance, if the take either the Gaussian kernel or the Epanechnikov kernel, we obtain the following bandwidths

$$\hat{h} = \begin{cases} 1.06n^{-\frac{1}{5}}s & \text{Gaussian kernel} \\ 2.34n^{-\frac{1}{5}}s & \text{Epanechnikov kernel} \end{cases} \tag{C.13}$$

However, the careful reader might object that assuming the normality of f is just the opposite of the philosophy of nonparametric density estimation. This is true but what we achieve by working under this assumption is an explicit, applicable formula for bandwidth selection. In practice, we do not know whether X is normally distributed. If it is, then $\hat{h}$ will give the optimal bandwidth. If not, $\hat{h}$ will give a bandwidth not to far from the optimum value if the data are nearly Gaussian. For this reason we refer to (C.13) as a simple rule of thumb or, more formally, as the *the normal reference bandwidth selector* (Silverman, 1986). The method is often a reasonable choice in many applications and, in particular, provides good results when the distributions are unimodal, fairly symmetric and the tails are not too fat, as it seems to be the case for the series in Figure C.4. This is also the method we have apply in a empirical analysis of this chapter.

The Silverman's rule of thumb belongs to the class of the *plug-in* methods for bandwidth selection. More refined versions of these methods, that also include the case of multimodal densities, can be found in Woodroofe (1970), Scott et al. (1977), Park and Marron (1990) and Sheather and Jones (1991). Another important technique for selecting the bandwidth is the *cross-validation* procedure. This data-driven approach is known to fulfill certain optimality properties and has been intensively investigated by Rudemo (1982), Bowman (1984), Scott and Terrell (1987), Hall et al. (1992) and, more recently, Hart and Yi (1998).

The C-Tz jump detection test

The test statistics proposed by Corsi et al. (2009) and used in the empirical analysis to verify the presence of significative jumps is given by

$$
\text{C-Tz} = \overline{\Delta}_n^{\frac{1}{2}} \frac{(\text{RV}(X) - \text{C-TBPV}(X)) \cdot \text{RV}(X)^{-1}}{\sqrt{\left(\frac{\pi^2}{4} + \pi + 5\right) \max\left\{1, \frac{\text{C-TTriPV}(X)}{(\text{C-TBPV}(X))^2}\right\}}}.
$$

Under the null hypothesis of no jumps in the series, the authors show that $\text{C-Tz} \to N(0, 1)$ stably in law as $\overline{\Delta}_n \to 0$. The different components of the test are defined as follow:

- Realized volatility

$$
\text{RV}(X) = \sum_{i=1}^{n} (\Delta X_i)^2
$$

- Corrected threshold multipower variation

$$
\text{C-TMPV}(X)^{(\gamma_1, \ldots, \gamma_M)} = \overline{\Delta}_n^{1 - \frac{1}{2}(\gamma_1 + \cdots + \gamma_M)} \sum_{i=M}^{[T/\overline{\Delta}_n]} \prod_{k=1}^{M} Z_{\gamma_k}(\Delta X_{j-k+1}, \theta_{j-k+1})
$$

for $\gamma_1, \ldots, \gamma_M > 0$, with

$$
Z_\gamma(x, y) = \begin{cases} |x|^\gamma & \text{if } x^2 \leq y \\ \frac{1}{2\Phi(-c_\theta)} \sqrt{\pi} \left(\frac{2}{c_\theta^2} y\right)^{\frac{\gamma}{2}} \Gamma\left(\frac{\gamma+1}{2}, \frac{c_\theta^2}{2}\right) & \text{if } x^2 > y \end{cases}
$$

and where $\Phi(x)$ is the standard normal cumulative function

- Corrected threshold bipower variation

$$\text{C-TBPV}(X) = \mu_1^{-2}\text{C-TMPV}(X)^{(1,1)}$$

- Corrected threshold tripower variation

$$\text{C-TTriPV}(X) = \mu_{\frac{4}{3}}^{-3}\text{C-TMPV}(X)^{(\frac{4}{3},\frac{4}{3},\frac{4}{3})}$$

The corrected versions of the above estimators overcome the finite sample weaknesses of the multipower variation introduced by Barndorff-Nielsen and Shephard (2004, 2006) to estimate realized variance in the presence of jumps. The estimator is defined as

$$\text{MPV}(X)^{(\gamma_1,\dots,\gamma_M)} = \overline{\Delta}_n^{1-\frac{1}{2}(\gamma_1+\dots+\gamma_M)} \sum_{i=M}^{[T/\overline{\Delta}_n]} \prod_{k=1}^{M} |\Delta X_{i-k+1}|^{\gamma^k}.$$

To give an idea of the problems arising with the use of this technique in practice, suppose that ΔX_i contains a jump. By definition, the estimator would multiply two adjacent returns, ΔX_{i-1} and and ΔX_{i+1}. Asymptotically, both these returns will vanish and the multipower variation will converge to the integrated continuous volatility. But for finite $\overline{\Delta}_n$ these returns will not vanish, causing a positive bias which will be larger as ΔX_i increases. The problem cannot be accommodated by simply shrinking the observation interval, since market microstructure effects would jeopardize the estimation in an unpredictable way. This consideration suggests that the bias of multipower variation will be extremely large in case of consecutive jumps. In the above estimators, jumps larger than θ_i are minimized using the function $Z_\gamma(x,y)$, thus correcting for the bias. The threshold is a positive random function given by

$$\theta_t = c_\theta^2 \hat{V}_t,$$

where c_θ^2 is a scale-free constant and $\hat{V}_t$ is an auxiliary estimator of $\sigma^2(t)$. In the empirical application we use $c_\theta = 3$. This parameter can be used to change the threshold and test the robustness of proposed estimators with respect to the choice of the threshold. The local variance is estimated by iterating in Q the following nonparametric filter of length $2L+1$ (Fan and Yao, 2003), adapted to the presence of jumps in Corsi et al. (2009) and defined as

$$\hat{V}_t^Q = \frac{\displaystyle\sum_{i=-L,\,i\neq-1,0,1}^{L} K\left(\frac{i}{L}\right)(\Delta X_{t+1})^2 I_{\{(\Delta X_{t+1})^2 \leq c_V^2 \hat{V}_{t+1}^{Q-1}\}}}{\displaystyle\sum_{i=-L,\,i\neq-1,0,1}^{L} K\left(\frac{i}{L}\right) I_{\{(\Delta X_{t+1})^2 \leq c_V^2 \hat{V}_{t+1}^{Q-1}\}}}, \qquad Q = 1, 2, \cdots$$

with $\hat{V}^0 = \infty$ as starting values, which corresponds to using all the available observations in the first step, and $C_V = 3$.

At each iteration, large returns are eliminated by the condition $(\Delta X_{t+1})^2 \leq c_V^2 \hat{V}_t^{Q-1}$. The iterations stop when no further large returns are removed. With high frequency data, this typically happens with $Q = 2, 3$ iterations. The bandwidth parameter L controls the number of adjacent returns included in the estimation of the local variance around point t. In our application, we set $L = 25$. We also choose to employ the Gaussian kernel $K(x) = \frac{1}{\sqrt{2\pi}} e^{\frac{x^2}{2}}$.

Proof of Proposition 4.1. By the mean value theorem, there exists a number $\xi_{n,x}$ such that $f_n(x) = f_n(0) + f'_n(\xi_{n,x})x$ for all $x \in [-1/2f_n(0), 1/2f_n(0)]$. Using the bound on the first derivative in Assumption 2, we also have $|f_n(x) - f_n(0)| \leq Cf_n(0)|x|$. Therefore, by setting the test process in Eq. (4.4) to $\varphi(s) = I_{\{|\bar{t}-s|\leq 1/(2f_n(0))\}}$, we obtain

$$\left| \int_{-\frac{1}{2f_n(0)}}^{\frac{1}{2f_n(0)}} f_n(x)dx \ - \ 1 \right| \leq \frac{1}{4}\frac{C}{f_n(0)} \longrightarrow 0.$$

Instead, with $\varphi(s) = \sigma^2(s)I_{\{|\bar{t}-s|\leq 1/(2f_n(0))\}}$ and the local Hölder property (4.2),

$$|R_n(\bar{t})| = \left| \int_0^T f_n(\bar{t}-s)\sigma^2(s)I_{\{|\bar{t}-s|\leq 1/(2f_n(0))\}}ds - \sigma^2(\bar{t}) \right| \leq C_{\bar{t}}\int_{-\frac{1}{2f_n(0)}}^{\frac{1}{2f_n(0)}} f_n(x)|x|^\gamma dx. \quad \text{(E.1)}$$

Again, by the mean value theorem and the bound on the first derivative we can conclude that

$$|R_n(\bar{t})| \leq \breve{C}f_n(0)^{-\gamma},$$

where $\breve{C}$ is a suitable constant. $\qquad\square$

Lemma E.1. *Consider a bounded process $A(t)$ defined in $[0,T]$ and satifying the same properties of $\sigma^2(t)$. Under Assumptions 2 and 3 we have*

$$\frac{1}{\overline{\Delta}_n}\sum_{i=1}^n f_n(\bar{t}-t_{i-1})A(t_{i-1})(\Delta_n^i)^2 \xrightarrow[n\to\infty]{p} H'(\bar{t})A(\bar{t}) \qquad \text{(E.2)}$$

and

$$\frac{1}{\overline{\Delta}_n f_n(0)}\sum_{i=1}^n f_n^2(\bar{t}-t_{i-1})A(t_{i-1})(\Delta_n^i)^2 \xrightarrow[n\to\infty]{p} c_f H'(\bar{t})A(\bar{t}). \qquad \text{(E.3)}$$

Proof. It is enough to show Eq. (E.2), since $f_n^2/(c_f f_n(0))$ is also a delta sequence. By Assumption 2,

$$H'(\bar{t})A(\bar{t}) = \int_0^T f_n(\bar{t} - s)H'(s)A(s)ds + o_p(1) = \sum_{i=1}^n f_n(\bar{t} - t_{i-1})H'(t_i)A(t_i)\Delta_n^i + o_p(1).$$

Moreover, we can write

$$\frac{1}{\overline{\Delta}_n}\sum_{i=1}^n f_n(\bar{t} - t_{i-1})A(t_{i-1})(\Delta_n^i)^2 - \sum_{i=1}^n f_n(\bar{t} - t_{i-1})H'(t_i)A(t_i)\Delta_n^i =$$

$$= \sum_{i=1}^n f_n(\bar{t} - t_{i-1})A(t_i)\Delta_n^i\left(H'(t_i) - \frac{\Delta_n^i}{\overline{\Delta}_n}\right).$$

If ϵ_n is a sequence such that $\epsilon_n f_n(0) \to 0$, then Eq. (4.13) implies

$$\left|H'(t_i) - \frac{H_n(t_i) - H_n(t_i - \epsilon_n)}{\epsilon_n}\right| \longrightarrow 0,$$

uniformly in $1 \leq i \leq n$. Now, set $\epsilon_n = (1 - \epsilon')\Delta_n^i$, where ϵ' is an arbitrarely small positive number. Since $\Delta_n^i f_n(0) = O(\overline{\Delta}_n f_n(0)) \to 0$, we get

$$\left|H'(t_i) - \frac{\Delta_n^i}{\overline{\Delta}_n(1 - \epsilon')}\right| \longrightarrow 0,$$

uniformly in $1 \leq i \leq n$, and since $\sum_{i=1}^n f_n(\bar{t} - t_{i-1})A(t_i)\Delta_n^i \xrightarrow{p} A(\bar{t})$ by the consistency part of Theorem 4.3, we finally obtain Eq. (E.2). $\qquad\square$

Proof of Theorem 4.3. It is not restrictive to set $\mu(t) = 0$, see Kristensen (2010) and Lee and Mykland (2008). If $\varphi(s) = \sigma^2(s)I_{\{|\bar{t}-s|>z\}}$ for a given number $z > 0$, then Assumption 2 implies

$$\int_{|\bar{t}-s|>z} f_n(\bar{t} - s)\sigma^2(s)ds \xrightarrow{p} 0.$$

Therefore, only the convergence in a neighborhood of $\bar{t}$ matters and we can assume that property (4.2) holds in $[0, T]$. We start by proving convergence in law. Consider

$$\int_0^T f_n(\bar{t} - s)\sigma^2(s)ds - \sum_{i=1}^n f_n(\bar{t} - t_{i-1})\int_{t_{i-1}}^{t_i}\sigma^2(s)ds =$$

$$= \sum_{i=1}^n \int_{t_{i-1}}^{t_i}[f_n(\bar{t} - s) - f_n(\bar{t} - t_{i-1})]\sigma^2(s)ds.$$

By the mean value theorem, there exists a point $\xi_i \in [t_{i-1}, s]$ such that

$$[f_n(\bar{t} - s) - f_n(\bar{t} - t_{i-1})] \leq f_n'(\xi_i)|s - t_{i-1}|,$$

and using the properties of f_n specified in Assumption 2, we have

$$\int_0^T f_n(\bar{t} - s)\sigma^2(s)ds - \sum_{i=1}^n f_n(\bar{t} - t_{i-1}) \int_{t_{i-1}}^{t_i} \sigma^2(s)ds =$$

$$= \sum_{i=1}^n \int_{t_{i-1}}^{t_i} f_n'(\xi_i)\sigma^2(s)|s - t_{i-1}|ds$$

$$\leq C f_n(0) \sum_{i=1}^n \frac{1}{2}(\Delta_n^i)^2 = O_{a.s.}(f_n(0)\overline{\Delta}_n).$$

The last term follows from the fact that $\sigma^2(t)$ is a bounded process, i.e.

$$\int_{t_{i-1}}^{t_i} \sigma^2(s)|s - t_{i-1}|ds \leq \left(\sup_{t \in [0,T]} \sigma^2(t)\right) \int_{t_{i-1}}^{t_i} |s - t_{i-1}|ds = \frac{1}{2}\left(\sup_{t \in [0,T]} \sigma^2(t)\right)(\Delta_n^i)^2,$$

and from the definition of the quadratic variation of time in Eq. (4.12). Finally, using Eq. (4.4),

$$\frac{1}{\sqrt{f_n(0)\overline{\Delta}_n}}\left[\sum_{i=1}^n f_n(\bar{t} - t_{i-1})\Delta X_i^2 - \sigma^2(\bar{t})\right] =$$

$$= \frac{1}{\sqrt{f_n(0)\overline{\Delta}_n}}\left[\sum_{i=1}^n f_n(\bar{t} - t_{i-1})\Delta X_i^2 - \int_0^T f_n(\bar{t} - s)\sigma^2(s)ds + R_n(\bar{t})\right]$$

$$= \frac{1}{\sqrt{f_n(0)\overline{\Delta}_n}}\left[\sum_{i=1}^n f_n(\bar{t} - t_{i-1})\left(\Delta X_i^2 - \int_{t_{i-1}}^{t_i} \sigma^2(s)ds\right) + O_{a.s.}(f_n(0)\overline{\Delta}_n) + R_n(\bar{t})\right]$$

$$= \sum_{i=1}^n \frac{f_n(\bar{t} - t_{i-1})}{\sqrt{f_n(0)\overline{\Delta}_n}}\left[\left(\int_{t_{i-1}}^{t_i} \sigma(s)dW(s)\right)^2 - \int_{t_{i-1}}^{t_i} \sigma^2(s)ds\right]$$

$$+ O_{a.s.}\left(\sqrt{f_n(0)\overline{\Delta}_n}\right) + \frac{R_n(\bar{t})}{\sqrt{f_n(0)\overline{\Delta}_n}} \tag{E.4}$$

$$:= \sum_{i=1}^n U_i + o_p(1),$$

since $R_n(\bar{t}) = o_p\left(\sqrt{f_n(0)\overline{\Delta}_n}\right)$. To derive a central limit theorem stable in law for $\sum_{i=1}^n U_i$, we refer to Theorem IX.7.28 in Jacod and Shiryaev (2003). According to this result, the following are sufficient conditions

$$(i) \quad \sum_{i=1}^n \mathrm{E}_{i-1}[U_i] \xrightarrow{p} 0 \qquad\qquad (iii) \quad \sum_{i=1}^n \mathrm{E}_{i-1}[U_i^4] \xrightarrow{p} 0$$

$$(ii) \quad \sum_{i=1}^n \mathrm{E}_{i-1}[U_i^2] \xrightarrow{p} V_t \qquad\qquad (iv) \quad \sum_{i=1}^n \mathrm{E}_{i-1}[U_i \Delta H_i] \xrightarrow{p} 0,$$

where $\mathrm{E}_{i-1}[\cdot]$ denotes the conditional expectation $\mathrm{E}[\cdot|\mathcal{F}_{t_{i-1}}]$.

In (iv), we have that $H = W$ or $H = B$, and B is any bounded martingale orthogonal (in the martingale sense) to W. It is immediate to prove the first condition using Itô isometry

$$
\begin{aligned}
\sum_{i=1}^{n} \mathrm{E}_{i-1}[U_i] &= \sum_{i=1}^{n} \frac{f_n(\bar{t} - t_{i-1})}{\sqrt{f_n(0)\overline{\Delta}_n}} \mathrm{E}_{i-1}\left[\left(\int_{t_{i-1}}^{t_i} \sigma(s)dW(s)\right)^2 - \int_{t_{i-1}}^{t_i} \sigma^2(s)ds\right] \\
&= \sum_{i=1}^{n} \frac{f_n(\bar{t} - t_{i-1})}{\sqrt{f_n(0)\overline{\Delta}_n}} \left[\int_{t_{i-1}}^{t_i} \mathrm{E}_{i-1}[\sigma^2(s)]ds - \int_{t_{i-1}}^{t_i} \mathrm{E}_{i-1}[\sigma^2(s)]ds\right] \\
&= 0.
\end{aligned}
$$

For condition (ii), consider

$$
\begin{aligned}
\sum_{i=1}^{n} \mathrm{E}_{i-1}\left[U_i^2\right] &= \sum_{i=1}^{n} \frac{f_n^2(\bar{t} - t_{i-1})}{f_n(0)\overline{\Delta}_n} \mathrm{E}_{i-1}\left[\left\{\left(\int_{t_{i-1}}^{t_i} \sigma(s)dW(s)\right)^2 - \int_{t_{i-1}}^{t_i} \sigma^2(s)ds\right\}^2\right] \\
&= \sum_{i=1}^{n} \frac{f_n^2(\bar{t} - t_{i-1})}{f_n(0)\overline{\Delta}_n} \left\{\mathrm{E}_{i-1}\left[\left(\int_{t_{i-1}}^{t_i} \sigma(s)dW(s)\right)^4\right] + \mathrm{E}_{i-1}\left[\left(\int_{t_{i-1}}^{t_i} \sigma^2(s)ds\right)^2\right] - \right. \\
&\quad \left. - 2\mathrm{E}_{i-1}\left[\left(\int_{t_{i-1}}^{t_i} \sigma(s)dW(s)\right)^2 \left(\int_{t_{i-1}}^{t_i} \sigma^2(s)ds\right)\right]\right\}.
\end{aligned}
$$

Now, write

$$
\begin{aligned}
\mathrm{E}_{i-1}\left[\left(\int_{t_{i-1}}^{t_i} \sigma(s)dW(s)\right)^4\right] &= \mathrm{E}_{i-1}\left[\left(\int_{t_{i-1}}^{t_i} [\sigma(t_{i-1}) + \sigma(s) - \sigma(t_{i-1})]\, dW(s)\right)^4\right] \\
&= 3\sigma^4(t_{i-1})(\Delta_n^i)^2 + \mathrm{E}_{i-1}\left[\left(\int_{t_{i-1}}^{t_i} [\sigma(s) - \sigma(t_{i-1})]\, dW(s)\right)^4\right] + \\
&\quad + 6\sigma^2(t_{i-1})\Delta_n^i \mathrm{E}_{i-1}\left[\left(\int_{t_{i-1}}^{t_i} [\sigma(s) - \sigma(t_{i-1})]\, dW(s)\right)^2\right].
\end{aligned}
$$

Using the Burkholder-Davis-Gundy inequality and the local Hölder property (4.2), for suitable positive constants C_1, C_2 and C_3, we have

$$
\begin{aligned}
\mathrm{E}_{i-1}\left[\left(\int_{t_{i-1}}^{t_i} [\sigma(s) - \sigma(t_{i-1})]\, dW(s)\right)^4\right] &\leq C_1 \mathrm{E}_{i-1}\left[\left(\int_{t_{i-1}}^{t_i} [\sigma(s) - \sigma(t_{i-1})]^2\, ds\right)^2\right] \\
&\leq C_2(\Delta_n^i)^2(\Delta_n^i)^{4\gamma} = o_p(\Delta_n^i)^2,
\end{aligned}
$$

and

$$\sigma^2(t_{i-1})\Delta_n^i \mathrm{E}_{i-1}\left[\left(\int_{t_{i-1}}^{t_i}\left[\sigma(s)-\sigma(t_{i-1})\right]dW(s)\right)^2\right] =$$

$$= \sigma^2(t_{i-1})\Delta_n^i \mathrm{E}_{i-1}\left[\left(\int_{t_{i-1}}^{t_i}\left[\sigma(s)-\sigma(t_{i-1})\right]^2 ds\right)\right]$$

$$\leq C_3\sigma^2(t_{i-1})\left(\Delta_n^i\right)^{2+2\gamma} = o_p(\Delta_n^i)^2.$$

Similarly,

$$\mathrm{E}_{i-1}\left[\left(\int_{t_{i-1}}^{t_i}\sigma(s)dW(s)\right)^2\left(\int_{t_{i-1}}^{t_i}\sigma^2(s)ds\right)\right] = \sigma^4(t_{i-1})(\Delta_n^i)^2 + o_p(\Delta_n^i)^2.$$

From the above results, together with Assumption (4.5) and Lemma E.1, we get

$$\sum_{i=1}^n \mathrm{E}_{i-1}\left[U_i^2\right] = 2\sum_{i=1}^n \frac{f_n^2(\bar{t}-t_{i-1})}{f_n(0)\overline{\Delta}_n}\left[\sigma^4(t_{i-1})(\Delta_n^i)^2 + o_p(\Delta_n^i)^2\right]$$

$$\xrightarrow{p} 2c_f\sigma^4(\bar{t})H'(\bar{t}).$$

Proceeding as earlier, condition (iii) is also verified since

$$\sum_{i=1}^n \mathrm{E}_{i-1}\left[U_i^4\right] = \frac{O_p(\overline{\Delta}_n)}{f_n^2(0)}\int_0^T f_n^4(\bar{t}-s)\sigma^8(s)ds + o_{a.s.}(\overline{\Delta}_n)^{4+\gamma},$$

which is $O_p(f_n(0)\overline{\Delta}_n)$ given Eq. (4.6).

Finally, we consider condition (iv) starting from the case $H = B$. Using the boundedness of B, which implies $|\Delta B_i| \leq C$, we obtain

$$\sum_{i=1}^n \mathrm{E}_{i-1}\left[U_i\Delta B_i\right] \leq C\sum_{i=1}^n \frac{f_n(\bar{t}-t_{i-1})}{\sqrt{f_n(0)\overline{\Delta}_n}}\mathrm{E}_{i-1}\left[\left(\int_{t_{i-1}}^{t_i}\sigma(s)dW(s)\right)^2 - \int_{t_{i-1}}^{t_i}\sigma^2(s)ds\right]$$

$$= O_p(1).$$

When instead $H = W$,

$$\sum_{i=1}^n \mathrm{E}_{i-1}\left[U_i\Delta W_i\right] \leq$$

$$\leq \sum_{i=1}^n \frac{f_n(\bar{t}-t_{i-1})}{\sqrt{f_n(0)\overline{\Delta}_n}}\sqrt{\mathrm{E}_{i-1}\left[\left(\int_{t_{i-1}}^{t_i}\sigma(s)dW(s)\right)^2 - \int_{t_{i-1}}^{t_i}\sigma^2(s)ds\right]^2}\sqrt{\mathrm{E}_{i-1}\left[(\Delta W_i^2)\right]}$$

$$= O_p(\overline{\Delta}_n)\cdot O_p(1).$$

This completes the proof of stable convergence of $\widehat{\sigma}_{n,f}^2(\bar{t})$.

For the convergence in probability the condition $R_n(\bar{t}) = o_p\left(\sqrt{f_n(0)\overline{\Delta}_n}\right)$ is not required. Indeed, to prove the convergency result, it is enough to multiply both sides of Eq. (E.4) by $\sqrt{f_n(0)\overline{\Delta}_n}$ and apply the law of large numbers to the sum of martingale differences, see e.g. Lemma 4.1 in Jacod (2007). $\qquad\square$

Proof of Lemma 4.7. The proof is based on that of Theorem 4.3. In what follows, we compute conditional expectations with respect to a new augmented filtration $\mathcal{F}_t^\varepsilon$ obtained by including the observed noise $(\varepsilon_i)_{t_i \leq t}$ for each $t \in [0, T]$. Set $\mu(t) = 0$ and write

$$\widehat{\sigma}_{n,f}^2(\bar{t}) = A_n + B_n + C_n,$$

where

$$A_n = \sum_{i=1}^{n} f_n(\bar{t} - t_{i-1})(\Delta Y_i)^2$$

$$B_n = 2\sum_{i=1}^{n} f_n(\bar{t} - t_{i-1})\Delta Y_i(\varepsilon_i - \varepsilon_{i-1})$$

$$C_n = \sum_{i=1}^{n} f_n(\bar{t} - t_{i-1})(\varepsilon_i - \varepsilon_{i-1})^2.$$

Note that A_n and C_n are independent and $\text{Cov}(A_n, B_n) = \text{Cov}(B_n, C_n) = 0$. For the first term A_n, we can simply apply Theorem 4.3. Next, define

$$U_{B,i} := 2f_n(\bar{t} - t_{i-1})\Delta Y_i(\varepsilon_i - \varepsilon_{i-1})/\sqrt{f_n(0)}.$$

We have

$$\sum_{i=1}^{n} \text{E}_{i-1}\left[U_{B,i}\right] = 0$$

$$\sum_{i=1}^{n} \text{E}_{i-1}\left[U_{B,i}^2\right] = 8\sum_{i=1}^{n} \frac{f_n^2(\bar{t} - t_{i-1})}{f_n(0)} V_\varepsilon \text{E}_{i-1}\left[(\Delta Y_i)^2\right] \xrightarrow{p} 8V_\varepsilon c_f \sigma^2(\bar{t})$$

$$\sum_{i=1}^{n} \text{E}_{i-1}\left[U_{B,i}^4\right] = O_p(\overline{\Delta}_n),$$

the latter equality attained using Eq. (4.6). Then

$$\sqrt{f_n(0)}B_n \longrightarrow \mathbf{MN}\left(0, 8V_\varepsilon c_f \sigma^2(\bar{t})\right).$$

Now, set $z_n = \frac{1}{2}\overline{\Delta}_n H'(\bar{t})$. We have that $z_n A_n \xrightarrow{p} 0$ with rate $\overline{\Delta}_n^{\frac{3}{2}} f_n(0)^{\frac{1}{2}}$ by Theorem 4.3, and $z_n B_n \xrightarrow{p} 0$ with rate $\overline{\Delta}_n f_n(0)^{-\frac{1}{2}}$ for the above result. Consider

$$U_{C,i} = \sqrt{\frac{1}{\overline{\Delta}_n f_n(0)}}\left(z_n f_n(\bar{t} - t_{i-1})(\varepsilon_i - \varepsilon_{i-1})^2 - V_\epsilon \Delta_n^i f_n(\bar{t} - t_{i-1})\right).$$

As in the proof of Theorem 4.3, we have that $z_n C_n - V_\varepsilon = \sum_{i=1}^{n} U_{C,i} + o_p(1)$ under the assumption $R_n(\bar{t}) = o_p(f_n(0)\overline{\Delta}_n)$ and

$$\sum_{i=1}^{n} \mathrm{E}_{i-1}\left[U_{C,i}\right] = 0$$

$$\sum_{i=1}^{n} \mathrm{E}_{i-1}\left[U_{C,i}^2\right] = \sum_{i=1}^{n} \frac{f_n^2(\bar{t} - t_{i-1})}{f_n(0)} \left(\frac{2\kappa_\varepsilon + 6V_\varepsilon^2}{4} - V_\varepsilon^2\right) H'(t_{i-1})\Delta_n^i \xrightarrow{p} \frac{1}{2}\left(\kappa_\varepsilon + V_\varepsilon^2\right) c_f H'(\bar{t})$$

$$\sum_{i=1}^{n} \mathrm{E}_{i-1}\left[U_{C,i}^4\right] = o_p(f_n(0)\overline{\Delta}_n),$$

which implies Eq. (4.16). $\qquad\qquad\square$

Proof of Theorem 4.8. As in the previous result, we set $\mu(t) = 0$ and compute conditional expectations with respect to the new augmented filtration $\mathcal{F}_t^\varepsilon$. Write

$$\widehat{\sigma}_{n,\bar{n}}^{2,TS}(\bar{t}) := S_n + R_{1,n} + R_{2,n} + R_{3,n} \tag{E.5}$$

where

$$S_n = \frac{1}{\bar{n}} \sum_{i=1}^{n-\bar{n}+1} f_n(\bar{t} - t_{i-1}) \left(Y_{t_{i+\bar{n}-1}} - Y_{t_{i-1}}\right)^2$$

$$R_{1,n} = \frac{1}{\bar{n}} \sum_{i=1}^{n-\bar{n}+1} f_n(\bar{t} - t_{i-1}) \left[\left(\varepsilon_{t_{i+\bar{n}-1}} - \varepsilon_{t_{i-1}}\right)^2 - \left(\varepsilon_{t_i} - \varepsilon_{t_{i-1}}\right)^2\right]$$

$$R_{2,n} = \frac{2}{\bar{n}} \sum_{i=1}^{n-\bar{n}+1} f_n(\bar{t} - t_{i-1}) \left[\left(\varepsilon_{t_{i+\bar{n}-1}} - \varepsilon_{t_{i-1}}\right)\left(Y_{t_{i+\bar{n}-1}} - Y_{t_{i-1}}\right) - \left(\varepsilon_{t_i} - \varepsilon_{t_{i-1}}\right)\Delta Y_i\right]$$

$$R_{3,n} = -\frac{1}{\bar{n}} \sum_{i=1}^{n-\bar{n}+1} f_n(\bar{t} - t_{i-1})(\Delta Y_i)^2.$$

By Theorem 4.3, $R_{3,n} = O_p\left(\frac{\sqrt{f_n(0)\overline{\Delta}_n}}{\bar{n}}\right) \to 0$. Now note that

$$S_n = \frac{1}{\bar{n}} \sum_{i=1}^{n-\bar{n}+1} f_n(\bar{t} - t_{i-1}) \left(\sum_{j=0}^{\bar{n}-1} \Delta Y_{i+j}\right)^2 = \underbrace{\frac{1}{\bar{n}} \sum_{i=1}^{n} (\Delta Y_i)^2 \left(\sum_{j=1}^{\bar{n} \wedge i} f_n(\bar{t} - t_{i-j})\right)}_{A_n}$$

$$+ \underbrace{\frac{2}{\bar{n}} \sum_{i=1}^{n} \Delta Y_i \left(\sum_{j=1}^{\bar{n}-1 \wedge i} f_n(\bar{t} - t_{i-j-1})\left(Y_{t_{i-1}} - Y_{t_{i-j-1}}\right)\right)}_{B_n}.$$

Using the mean value theorem and $f_n'(x) = O(f_n(0))$,

$$\sum_{j=1}^{\overline{n} \wedge i} f_n(\overline{t} - t_{i-j}) = (\overline{n} \wedge i) f_n(\overline{t} - t_{i-1}) + O\left(\overline{n}^2 f_n(0) \overline{\Delta}_n\right),$$

and together with Theorem 4.3,

$$A_n = \sigma^2(\overline{t}) + O_p\left(\overline{n} f_n(0) \overline{\Delta}_n\right).$$

End effects are $o_p(\overline{n} f_n(0) \overline{\Delta}_n)$ and can be systematically neglected. For the term $B_n := \sum_{i=1}^n \widetilde{U}_{n,i}^B$ we have $\sum_{i=1}^n \mathrm{E}_{i-1}\left[\widetilde{U}_{n,i}^B\right] = 0$ and

$$\sum_{i=1}^n \mathrm{E}_{i-1}\left[\left(\widetilde{U}_{n,i}^B\right)^2\right] =$$

$$= \frac{4}{\overline{n}^2} \sum_{i=1}^n \sigma^2(t_{i-1}) \Delta_n^i \left(\sum_{j=1}^{\overline{n}-1 \wedge i} f_n^2(\overline{t} - t_{i-j-1}) \left(Y_{t_{i-1}} - Y_{t_{i-j-1}}\right)^2\right) + o_p(1)$$

$$= \frac{4}{\overline{n}^2} \sum_{i=1}^n \sigma^4(t_{i-1}) \Delta_n^i \sum_{j=1}^{\overline{n}-1 \wedge i} f_n^2(\overline{t} - t_{i-j-1}) \left(t_{i-1} - t_{i-j-1}\right) + o_p(1)$$

$$= 4\overline{n} \sum_{i=1}^n \frac{1}{2} \sigma^4(t_{i-1}) H'(t_{i-1}) \overline{\Delta}_n f_n^2(\overline{t} - t_{i-1}) \Delta_n^i + o_p(\overline{n} f_n(0) \overline{\Delta}_n),$$

where we used the Hölder property, the mean value theorem and that

$$\sum_{j=1}^{\overline{n}-1} (t_{i-1} - t_{t_{i-j-1}}) = H'(t_{i-1}) \overline{\Delta}_n \frac{(\overline{n}-1)(\overline{n}-2)}{2} + o(1)$$

as implied by lemma E.1 for sufficiently large n. Similarly, we can show that

$$\sum_{i=1}^n \mathrm{E}_{i-1}\left[\left(\widetilde{U}_{n,i}^B\right)^4\right] = o_p(\overline{n} f_n(0) \overline{\Delta}_n),$$

which implies

$$\frac{1}{\sqrt{\overline{n} f_n(0) \overline{\Delta}_n}} B_n \xrightarrow{d} \mathbf{MN}\left(0, 2c_f \sigma^4(\overline{t}) H'(\overline{t})\right).$$

Therefore, $S_n = \sigma^2(\overline{t}) + O_p\left(\sqrt{f_n(0) \overline{\Delta}_n \overline{n}}\right)$. Proceeding as earlier, it is possible to show that

$$\sum_{i=1}^n \mathrm{E}_{i-1}\left[\left(\widetilde{U}_{n,i}^B + \widetilde{U}_{n,i}^A\right)^4\right] = o_p(\overline{n} f_n(0) \overline{\Delta}_n),$$

where $A_n := \sum_{i=1}^n \widetilde{U}_{n,i}^A$.

Since $\text{Cov}(A_n, B_n) = 0$, we also have

$$\frac{1}{\sqrt{\overline{n} f_n(0) \overline{\Delta}_n}} \left(S_n - \sigma^2(\overline{t}) \right) \longrightarrow \mathbf{MN} \left(0, 2c_f \sigma^4(\overline{t}) H'(\overline{t}) \right).$$

Note that subsampling does not improve the asymptotic variance because of the localization procedure induced by the Dirac delta sequence. For integrated volatility instead, subsampling provides a factor $\frac{4}{3}$ instead of 2.

We now evaluate the remaining terms using the independence between $Y(t)$ and ε_t and the following identities, which hold true for every $i, j > 1$ under our assumptions:

$$\begin{aligned}
\mathrm{E}\left[(\varepsilon_{i+j} - \varepsilon_i)^2 \right] &= 2V_\varepsilon \\
\mathrm{E}\left[(\varepsilon_{i+j} - \varepsilon_i)(\varepsilon_{i+1} - \varepsilon_i) \right] &= V_\varepsilon \\
\mathrm{E}\left[(\varepsilon_{i+j} - \varepsilon_i)^4 \right] &= 2\kappa_\varepsilon + 4V_\varepsilon^2 \\
\mathrm{E}\left[(\varepsilon_{i+j} - \varepsilon_i)^2 (\varepsilon_{i+1} - \varepsilon_i)^2 \right] &= \kappa_\varepsilon - V_\varepsilon^2 \\
\mathrm{E}\left[\left(Y_{t_{i+j}} - Y_{t_i} \right) \left(Y_{t_{i+1}} - Y_{t_i} \right) \right] &= \mathrm{E}[\sigma^2(t_i)]\Delta_n^{i+1} + O_{a.s.}\left(\overline{\Delta}_n^{1+\gamma} \right) \\
\mathrm{E}\left[\left(Y_{t_{i+j}} - Y_{t_i} \right)^2 \left(Y_{t_{i+1}} - Y_{t_i} \right)^2 \right] &= \mathrm{E}[\sigma^2(t_i)]\mathrm{E}[\sigma^2(t_{i+1})]\Delta_n^{i+2}\Delta_n^{i+1}(j-1) + \\
&\quad + 3\mathrm{E}[\sigma^2(t_{i-1})] \left(\Delta_n^{i+1} \right)^2 + O_{a.s}(\overline{\Delta}_n^{1+\gamma}).
\end{aligned}$$

Write $R_{1,n} := \sum_{i=1}^n U_{n,i+\overline{n}-1}^{R_1}$. Then, for sufficiently large $\overline{n}$,

$$\begin{aligned}
\sum_{i=1}^n \mathrm{E}_{i+\overline{n}-2}\left[U_{n,i+\overline{n}-1}^{R_1} \right] &= \\
&= \frac{1}{\overline{n}} \sum_{i=1}^{n-\overline{n}+1} f_n(\overline{t} - t_{i-1}) \mathrm{E}_{i+\overline{n}-2}\left[\varepsilon_{t_{i+\overline{n}-1}}^2 - 2\varepsilon_{t_{i+\overline{n}-1}} \varepsilon_{t_{i-1}} - \varepsilon_{t_i}^2 + 2\varepsilon_{t_{i-1}} \varepsilon_{t_i} \right] \\
&= \frac{1}{\overline{n}} \sum_{i=1}^{n-\overline{n}+1} f_n(\overline{t} - t_{i-1}) \left[V_\varepsilon - \varepsilon_{t_i}^2 + 2\varepsilon_{t_{i-1}} \varepsilon_{t_i} \right] \xrightarrow{p} 0,
\end{aligned}$$

where we have applied the law of large numbers. Similarly,

$$\begin{aligned}
\sum_{i=1}^n \mathrm{E}_{i+\overline{n}-2}\left[\left(U_{n,i+\overline{n}-1}^{R_1} \right)^2 \right] &= \\
&= \frac{1}{\overline{n}^2} \sum_{i=1}^{n-\overline{n}+1} f_n^2(\overline{t} - t_{i-1}) \left[\kappa_\varepsilon + 4V_\varepsilon \varepsilon_{t_{i-1}}^2 + \varepsilon_{t_i}^4 + 4\varepsilon_{t_i}^2 \varepsilon_{t_{i-1}}^2 - 4\mathrm{E}[\varepsilon^3]\varepsilon_{t_{i-1}} + \right. \\
&\qquad \left. -2V_\varepsilon \varepsilon_{t_i}^2 + 4V_\varepsilon \varepsilon_{t_i} \varepsilon_{t_{i-1}} - 4\varepsilon_{t_i}^3 \varepsilon_{t_{i-1}} \right] \\
&\xrightarrow{p} c_f \frac{f_n(0)}{\overline{n}^2 \overline{\Delta}_n H'(\overline{t})} \left(2\kappa_\varepsilon + 6V_\varepsilon^2 \right).
\end{aligned}$$

Proceeding in the same way, we can also show that $\sum_{i=1}^n \mathrm{E}_{i+\overline{n}-2}\left[\left(U_{n,i+\overline{n}-1}^{R_1} \right)^4 \right] = o_p(1)$.

This implies the asymptotic mixed normality of $R_{1,n}$ with the variance above. Now, set $R_{2,n} = \sum_{i=1}^{n} U_{n,i+\overline{n}-1}^{R_2}$. We have

$$\sum_{i=1}^{n} \mathrm{E}_{i+\overline{n}-2} = \left[U_{n,i+\overline{n}-1}^{R_2} \right] = \frac{1}{\overline{n}} \sum_{i=1}^{n-\overline{n}+1} f_n(\overline{t} - t_{i-1}) \left[-\varepsilon_{t_{i-1}} \left(Y_{t_{i+\overline{n}-2}} - Y_{t_{i-1}} \right) \right] = o_p(1)$$

and

$$\sum_{i=1}^{n} \mathrm{E}_{i+\overline{n}-2} \left[\left(U_{n,i+\overline{n}-1}^{R_2} \right)^2 \right] =$$

$$= \frac{4}{\overline{n}^2} \sum_{i=1}^{n-\overline{n}+1} f_n^2(\overline{t} - t_{i-1}) \begin{bmatrix} (V_\varepsilon + \varepsilon_{t_{i-1}}^2) [\underbrace{\sigma^2(t_{i+\overline{n}-2})\Delta_n^{i+\overline{n}-1}}_{O_p\left(\frac{\sqrt{f_n(0)}}{\overline{n}^2} \right)} + \underbrace{(Y_{t_{i+\overline{n}-2}} - Y_{t_{i-1}})^2}_{O_p\left(\frac{\sqrt{f_n(0)}}{\overline{n}} \right)} \\ + [2\Delta Y_{i+\overline{n}-1}(Y_{t_{i+\overline{n}-2}} - Y_{t_{i-1}})] \\ + \underbrace{\left(\varepsilon_{t_i} - \varepsilon_{t_{i-1}} \right)^2 (\Delta Y_i)^2}_{O_p\left(\frac{\sqrt{f_n(0)}}{\overline{n}^2} \right)} \\ + 2\varepsilon_{t_{i-1}} \left(\varepsilon_{t_i} - \varepsilon_{t_{i-1}} \right) (Y_{t_{i+\overline{n}-2}} - Y_{t_{i-1}}) \end{bmatrix}$$

$$= o_p\left(\frac{\sqrt{f_n(0)}}{\overline{n}} \right),$$

where the last asymptotic order refers to the expression within brackets. Analogously, it can be shown that $\sum_{i=1}^{n} \mathrm{E}_{i+\overline{n}-2} \left[\left(U_{n,i+\overline{n}-1}^{R_2} \right)^4 \right] = o_p(1)$. Moving to cross-terms, we have

$$\sum_{i=1}^{n} \mathrm{E}_{i+\overline{n}-2} \left[U_{n,i+\overline{n}-1}^{R_2} U_{n,i+\overline{n}-1}^{R_1} \right] = o_p\left(\frac{f_n(0)}{\overline{n}^2 \overline{\Delta}_n} \right)$$

and the same applies between $R_{2,n}$ and $R_{3,n}$ while $R_{1,n}$ and $R_{3,n}$ are independent. Therefore, the leading order terms are $(S_n - \sigma^2(\overline{t}))$ and $R_{1,n} = O_p\left(\sqrt{\frac{f_n(0)}{\overline{\Delta}_n(\overline{n})^2}} \right)$. When $\overline{n} = c\overline{\Delta}_n^{-\frac{2}{3}}$, the rates of convergence of S_n and $R_{1,n}$ are the same and the final result follows. Stable convergence in law can be derived as in the proof of Theorem 4.3. $\square$

Proof of Theorem 4.10. Denote by $X = Y + J$ where Y is a continuous semimartingale. By virtue of Theorem 1 in Mancini (2009), for n large enough, we can write

$$\widehat{\phi}_{n,f}^2(\overline{t}) = \sum_{i=1}^{n} f_n(\overline{t} - t_{i-1})(\Delta Y_i)^2 - \sum_{i=1}^{n} f_n(\overline{t} - t_{i-1})(\Delta Y_i)^2 I_{\{\Delta N_i \neq 0\}}.$$

Theorem 4.3 can be applied to the first term, while the second term is $O_p(N_T \overline{\Delta}_n f_n(0))$, or equivalently, $o_p\left(\sqrt{\overline{\Delta}_n f_n(0)} \right)$, where N_T is the Poisson counting process and is vanishing in the limit. $\square$

Lemma E.2. *Let Assumptions 1,2,3 hold. With $\phi_{n,f,g}$ defined in (4.26), let f_n, g_n be two sequences satisfying Assumption 2 with g_n symmetric around 0 and such that $\frac{g_n(0)}{f_n(0)} \to \infty$. If $\overline{\Delta}_n g_n(0) \to 0$, then*

$$\sqrt{\frac{g_n(0)}{f_n(0)}} \phi_{n,f,g}(\bar{t}) \xrightarrow{\mathcal{L}_{st,}} \mathbf{MN}(0, 2\, c_f c_g \sigma^4(\bar{t})). \tag{E.6}$$

Proof. As above, we set $\mu(t) = 0$ and define

$$\sum_{i=1}^n U_i := \sqrt{\frac{g_n(0)}{f_n(0)}} \sum_i \frac{2}{g_n(0)} \sum_{j<i} f_n(\bar{t} - t_{j-1}) g_n(t_{j-1} - t_{i-1}) \Delta X_i \Delta X_j,$$

with the U_i terms measurable with respect to $\mathcal{F}_{t_i}$. Condition (i) in the proof of Theorem 4.3 is clearly satisfied as $\mathrm{E}_{i-1}[\Delta X_i \Delta X_j] = \Delta X_j \mathrm{E}_{i-1}[\Delta X_i] = 0$. To valuate condition (ii), we use previous arguments to find

$$\sum_{i=1}^n \mathrm{E}_{i-1}\left[U_i^2\right] = \frac{4}{f_n(0)g_n(0)} \sum_i \mathrm{E}_{i-1}\left[\left(\sum_{j<i} f_n(\bar{t} - t_{j-1}) g_n(t_{j-1} - t_{i-1}) \Delta X_i \Delta X_j\right)^2\right]$$

$$= \frac{4}{f_n(0)g_n(0)} \sum_i \sum_{j<i} f_n^2(\bar{t} - t_{j-1}) g_n^2(t_{j-1} - t_{i-1}) \mathrm{E}_{i-1}[(\Delta X_i)^2](\Delta X_j)^2$$

$$= \frac{4}{f_n(0)g_n(0)} \sum_i \sum_{j<i} f_n^2(\bar{t} - t_{j-1}) g_n^2(t_{j-1} - t_{i-1}) \left(\sigma^2(t_{i-1})\sigma^2(t_{j-1}) \Delta_n^i \Delta_n^j + \right.$$

$$\left. + o_{a.s.}[(\Delta_n^i)^{1+\gamma}(\Delta_n^j)^{1+\gamma}]\right)$$

$$= \frac{4}{f_n(0)g_n(0)} \int_0^T f_n^2(\bar{t} - u)\sigma^2(u) \left(\int_0^{u - \overline{\Delta}_n H'(s)} g_n^2(u - s)\sigma^2(s) ds\right) du +$$

$$+ o_{a.s.}[(\overline{\Delta}_n)^{1+\gamma}(\overline{\Delta}_n)^{1+\gamma}],$$

where we have applied the law of large numbers. We now study the convergence of the term

$$\mathcal{G}_n := \frac{1}{g_n(0)} \int_0^{u - \overline{\Delta}_n H'(s)} g_n^2(u - s)\sigma^2(s) ds.$$

As in the proof of Proposition 4.1, we have that

$$\mathcal{G}_n' := \frac{1}{g_n(0)} \int_{-\overline{\Delta}_n H'(s)}^{\overline{\Delta}_n H'(s)} g_n^2(u - s)\sigma^2(s) ds = O_p(g_n(0)\overline{\Delta}_n),$$

and $\mathcal{G}_n$ has the same limit of $\frac{1 - \mathcal{G}_n'}{2}$ due to the symmetry of $g_n(\cdot)$. Therefore, if $\overline{\Delta}_n g_n(0) \to 0$, $\mathcal{G}_n$ converges to $\frac{1}{2}c_g \sigma^2(u)$ by Eq. (4.5) and we get

$$\sum_{i=1}^n \mathrm{E}_{i-1}\left[U_i^2\right] \xrightarrow{p} 2\, c_f c_g \sigma^4(\bar{t}).$$

As far as the fourth moment is concerned, we note that

$$\sum_{i=1}^{n} \mathrm{E}_{i-1}\left[U_i^4\right] = \frac{16}{f_n^2(0)g_n^2(0)} \sum_i \mathrm{E}_{i-1}\left[\left(\sum_{j<i} f_n(\bar{t}-t_{i-1})g_n(t_{j-1}-t_{i-1})\Delta X_i \Delta X_j\right)^4\right]$$

$$= \frac{16}{f_n^2(0)g_n^2(0)}(A_n + B_n),$$

where

$$A_n := \sum_{i=2}^{n}\sum_{j=1}^{i-1} f_n^4(\bar{t}-t_{i-1})g_n^4(t_{j-1}-t_{i-1})\mathrm{E}_{i-1}\left[\Delta X_i^4\right](\Delta X_j^4)$$

$$= 9\sum_{i=2}^{n}\sum_{j=1}^{i-1} f_n^4(\bar{t}-t_{i-1})g_n^4(t_{j-1}-t_{i-1})\left(\sigma^4(t_{i-1})\sigma^4(t_{j-1})(\Delta_n^i)^2(\Delta_n^j)^2 + \right.$$

$$\left. + o_{a.s.}[(\Delta_n^i)^{2+\gamma}(\Delta_n^j)^{2+\gamma}]\right)$$

and

$$B_n := 6\sum_{i=2}^{n}\sum_{j=1}^{i-2}\sum_{k=j+1}^{i-1} f_n^4(\bar{t}-t_{i-1})g_n^4(t_{j-1}-t_{i-1})\mathrm{E}_{i-1}\left[\Delta X_i^4\right](\Delta X_j^2)(\Delta X_k^2)$$

$$= 18\sum_{i=2}^{n}\sum_{j=1}^{i-2}\sum_{k=j+1}^{i-1} f_n^4(\bar{t}-t_{i-1})g_n^2(t_{j-1}-t_{i-1})g_n^2(t_{k-1}-t_{i-1})\cdot$$

$$\cdot \left(\sigma^4(t_{i-1})\sigma^2(t_{j-1})\sigma^2(t_{k-1})(\Delta_n^i)^2\Delta_n^j\Delta_n^k + o_{a.s.}[(\Delta_n^i)^{2+\gamma}(\Delta_n^j)^{1+\gamma}(\Delta_n^k)^{1+\gamma}]\right).$$

Moving to the integral form, we find that $\sum_{i=1}^{n} \mathrm{E}_{i-1}(U_i^4) \xrightarrow{p} 0$. Stable convergence in law follows as in the proof of Theorem 4.3. $\qquad\square$

Proof of Proposition 4.11. From Eq. (4.22)-(4.24) we have

$$\widehat{\sigma}_{n,N,M}^{2,F}(\bar{t}) = \lim_{M\to\infty} \sum_{|k|\leq M}\left(1 - \frac{|k|}{M}\right)\alpha_k(N)e^{ik\tau}$$

$$= \lim_{M\to\infty} \sum_{|k|\leq M}\left(1 - \frac{|k|}{M}\right)\left(\frac{T}{2N+1}\sum_{|s|\leq N} c_s(dX)c_{k-s}(dX)\right)e^{ik\tau}$$

$$= \lim_{M\to\infty} \frac{1}{T}\sum_{|k|\leq M}\left(1 - \frac{|k|}{M}\right)\left(\frac{1}{2N+1}\sum_{|s|\leq N}\sum_{j'=1}^{n}\sum_{j=1}^{n} e^{-is\tau_{j'-1}}e^{-i(k-s)\tau_{j-1}}\Delta X_j\Delta X_{j'}\right)e^{ik\tau}.$$

Now, note that

$$\frac{1}{T}\sum_{|k|\leq M}\left(1 - \frac{|k|}{M}\right)\left(\frac{1}{2N+1}\sum_{|s|\leq N}\sum_{j'=1}^{n}\sum_{j=1}^{n} e^{-is\tau_{j'-1}}e^{-i(k-s)\tau_{j-1}}\Delta X_{j'}\Delta X_j\right)e^{ik\tau}$$

can be written as

$$\frac{1}{T} \sum_{|k| \leq M} \left(1 - \frac{|k|}{M}\right) \left[\frac{1}{2N+1} \sum_{|s| \leq N} \left(\sum_{j'=1}^{n} e^{-ik\tau_{j'-1}}(\Delta X_{j'})^2 + \right.\right.$$

$$\left.\left. + 2 \sum_{j'=2}^{n} \sum_{j=1}^{j'-1} e^{-is\tau_{j'-1}} e^{-i(k-s)\tau_{j-1}} \Delta X_j' \Delta X_j \right)\right] e^{ik\tau}$$

$$= \frac{1}{T} \sum_{j'=1}^{n} \sum_{|k| \leq M} \left(1 - \frac{|k|}{M}\right) e^{ik(\tau-\tau_{j'-1})}(\Delta X_{j'})^2 +$$

$$+ \frac{2}{T(2N+1)} \sum_{j'=2}^{n} \sum_{j=1}^{j'-1} \sum_{|k| \leq M} \left(1 - \frac{|k|}{M}\right) e^{ik(\tau-\tau_{j-1})} \sum_{|s| \leq N} e^{is(\tau_{j-1}-\tau_{j'-1})} \Delta X_{j'} \Delta X_j.$$

By the definitions of Dirichlet and Fejér kernel as given in equations (4.9) and (4.10), we then get

$$\widehat{\sigma}_{n,N,M}^{2,F}(\bar{t}) = \sigma_{n,F}^2(\bar{t}) + \phi_{n,F,D}(\bar{t}),$$

where

$$\sigma_{n,F}^2(\bar{t}) := \frac{1}{T} \sum_{j=1}^{n} F_{M-1}(\tau - \tau_{j-1})(\Delta X_j)^2$$

$$\phi_{n,F,D}(\bar{t}) := \frac{1}{T} \frac{2}{D_N(0)} \sum_{j'=2}^{n} \sum_{j=1}^{j'-1} F_{M-1}(\tau - \tau_{j-1}) D_N(\tau_{j-1} - \tau_{j'-1}) \Delta X_{j'} \Delta X_j.$$

Since $D_N(0) = 2N + 1$, then Lemma E.2 applies if $\frac{N}{n} \to 0$. To calculate the values of c_g and c_f, it is enough to employ the properties we have derived for the Dirichlet and Fejér kernels, namely,

$(i) \quad \dfrac{1}{2\pi} \displaystyle\int_0^{2\pi} F_{M-1}(x)dx = 1 \qquad\qquad (ii) \quad \dfrac{1}{2\pi} \displaystyle\int_0^{2\pi} D_N(x)dx = 1$

$(iii) \quad D_N^2(x) = (2N+1)F_{2N}(x) \qquad (iv) \quad \dfrac{1}{F_{M-1}(0)} \displaystyle\int_0^{2\pi} F_{M-1}^2(x)dx = \dfrac{4\pi}{3},$

where $F_{M-1}(0) = M$ and $D_N(0) = 2N+1$, as above. We first perform the change of variables $s = \frac{T}{2\pi}x$, then $c_g = 1$ follows from identity (iii) and the relation

$$\frac{1}{T} \frac{1}{D_N(0)} \int_0^T D_N^2\left(\frac{2\pi}{T}s\right) ds = c_g,$$

while the value $c_f = \frac{2}{3}$ can be easily obtained using property (iv). When $\frac{N}{n} \to c \in]0, \frac{1}{2}]$, we have to evaluate

$$\frac{1}{D_N(0)} \int_{-\infty}^{-\Delta_n^{j'-1}} D_N^2(x)dx$$

(see the proof of Lemma E.2). To this purpose, we use again identity (iii) and the result

$$\int_{-\varepsilon}^{\varepsilon} F_{2N}(x)dx = 2N\varepsilon,$$

for $\varepsilon > 0$. Using Lemma E.1, we then obtain

$$\text{Var}\left[\phi_{n,F,D}(\bar{t})\right] \longrightarrow 4c_f F_{M-1}(0)\overline{\Delta}_n\left(\frac{1}{2} - cH'(\bar{t})\right).$$

Theorem 4.3 and the equality $\text{Cov}(\phi_{n,F,D}(\bar{t}), \sigma_{n,F}^2) = 0$ conclude the proof. $\square$

Aït-Sahalia, Y. and J. Jacod, 2009: Testing for jumps in a discretely observed process. *Annals of Statistics*, **37**, 184–222.

Aït-Sahalia, Y., P. Mykland, and L. Zhang, 2005: How often to sample a continuous-time process in the presence of microstructure noise. *Review of Financial Studies*, **18**, 351–416.

Aït-Sahalia, Y., P. Mykland, and L. Zhang, 2011: Ultra high frequency volatility estimation with dependent microstructure noise. Forthcoming in the Journal of Econometrics.

Anderberg, M. R., 1973: *Cluster Analyses for Applications*. Academic Press, New York.

Andersen, T. and T. Bollerslev, 1997: Intraday periodicity and volatility persistence in financial markets. *Journal of Empirical Finance*, **4**, 115–158.

Andersen, T. and T. Bollerslev, 1998: Answering the skeptics: Yes, standard volatility models do provide accurate forecasts. *International Economic Review*, **39**, 885–905.

Andersen, T., T. Bollerslev, F. Diebold, and P. Labys, 2000: Great realizations. *Risk*, **13**((3)), 105–108.

Andersen, T., T. Bollerslev, F. Diebold, and P. Labys, 2001: The distribution of realized exchange rate volatility. *Journal of the American Statistical Association*, **96**, 42–55.

Andersen, T., T. Bollerslev, and F. X. Diebold, 2009: Parametric and nonparametric volatility measurement. In *Handbook of Financial Econometrics*, Hansen, L. P. and Ait-Sahalia, Y., editors. North-Holland.

Andersen, T., T. Bollerslev, and N. Meddahi, 2004: Analytic evaluation of volatility forecasts. *International Economic Review*, **45**(4), 1079–1110.

Andreou, E. and E. Ghysels, 2002: Rolling-sampling volatility estimators: some new theo-

retical, simulation and empirical results. *Journal of Business and Economic Statistics*, **20**, 363–376.

Andreou, E. and E. Ghysels, 2004: The impact of sampling frequency and volatility estimators on change-point tests. *Journal of Financial Econometrics*, **2**, 290–318.

Bakshi, G., C. Cao, and Z. Chen, 1997: Empirical performance of alternative option pricing models. *Journal of Finance*, **52**, 2003–2049.

Ball, C. and W. Torous, 2000: Stochastic correlation across international stock markets. *Journal of Empirical Finance*, **7**, 373–388.

Bandi, F. and T. Nguyen, 2003: On the functional estimation of jump-diffusion models. *Journal of Econometrics*, **116**, 293–328.

Bandi, F. and P. Phillips, 2003: Fully nonparametric estimation of scalar diffusion models. *Econometrica*, **71**(1), 241–283.

Bandi, F. and R. Renò, 2008: Nonparametric stochastic volatility. Working Paper.

Bandi, F. and J. Russell, 2005: Realized covariation, realized beta ad microstructure noise. Working paper.

Bandi, F. and J. Russell, 2006a: Comment on "realized variance and microstructure noise" by peter hansen and asger lunde. *Journal of Business and Economic Statistics*, **24**, 167–173.

Bandi, F. and J. Russell, 2006b: Separating microstructure noise from volatility. *Journal of Financial Economics*, **79**(3), 655–92.

Bandi, F. and J. Russell, 2008: Microstructure noise, realized volatility and optimal sampling. *Review of Economic Studies*, **75**, 339–369.

Barndorff-Nielsen, O., P. Hansen, A. Lunde, and N. Shephard, 2006: Subsampling realised kernels. Forthcoming in the Journal of Econometrics.

Barndorff-Nielsen, O., P. Hansen, A. Lunde, and N. Shephard, 2008: Designing realised kernels to measure the ex-post variation of equity prices in the presence of noise. *Econometrica*, **76**(6), 1481–1536.

Barndorff-Nielsen, O. E. and N. Shephard, 2001: Non-gaussian ornstein-uhlenbeck based models and some of their uses in financial economics (with discussion). *Journal of the Royal Statistical Society. Series B*, **63**, 167–241.

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

Barndorff-Nielsen, O. E. and N. Shephard, 2002b: Estimating quadratic variation using realized variance. *Journal of Applied Econometrics*, **17**, 457–478.

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

Barndorff-Nielsen, O. E. and N. Shephard, 2006: Econometrics of testing for jumps in financial economics using bipower variation. *Journal of Financial Econometrics*, **4**, 1–30.

Barucci, E. and R. Renò, 2002a: On measuring volatility and the GARCH forecasting performance. *Journal of International Financial Markets, Institutions and Money*, **12**, 183–200.

Barucci, E. and R. Renò, 2002b: On measuring volatility of diffusion processes with high frequency data. *Economics Letters*, **74**, 371–378.

Bibby, B., M. Skovgaard, and M. Sø rensen, 2005: Diffusion-type models with given marginal distribution and autocorrelated functions. *Bernoulli*, **11**(2), 191–220.

Bjork, T., 1998: *Arbitrage theory in continuous time.* Wiley.

Black, F., 1976: Noise. *Journal of Finance*, **41**, 529–543.

Boudt, K., C. Croux, and S. Laurent, 2008: Robust estimation of intraweek periodicity in volatility and jump detection. Working Paper.

Bowman, A., 1984: An alternative method of cross-validation for the smoothing of density estimates. *Biometrika*, **71**, 353–360.

Chan, K., A. Karolyi, F. Longstaff, and A. Sanders, 1992: An empirical comparison of alternative models of the short-term interest rate. *Journal of Finance*, **47**(3), 1209–1227.

Comte, F., V. Genon-Catalot, and Y. Rozenholc, 2010: Nonparametric estimation for a stochastic volatility model. *Finance and Stochastics*, **14**, 49–80.

Cooley, J. W. and J. W. Tukey, 1965: An algorithm for the machine calculation of complex fourier series. *Mathematics of Computation*, **19**, 297–301.

Corsi, F., D. Pirino, and R. Renò, 2009: Threshold bipower variation and the impact of jumps on volatility forecasting. Working Paper.

Corsi, F., G. Zumbach, U. Muller, and M. Dacorogna, 2001: Consistent high-precision volatility from high-frequency data. *Economic Notes*, **30**(2), 183–204.

Cox, J., J. Ingersoll, and S. Ross, 1985: A theory of the term structure of interest rates.

Econometrica, **53**, 385–406.

Curci, G. and F. Corsi, 2003: A discrete sine transform approach for realized volatility measurement. Manuscript, University of Pisa.

Dacorogna, M., R. Gencay, U. Muller, R. Olsen, and O. Pictet, 2001: *An introduction to high-frequency finance.* Academic Press, London.

Dacorogna, M. and M. Lundin, 1999: Correlation of high-frequency financial time series. In *Financial Markets Tick by Tick*, Lequeux, P., editor. Wiley & Sons.

Dirac, P. M., 1930: *The Principles of Quantum Mechanics.* Clarendon Press.

Driessen, J., P. Maenhout, and G. Vilkov, 2005: Option implied correlations and the price of correlation risk. EFA 2005, Moscow Meetings.

Dungey, M., R. A. Fry, and V. L. Martin, 2003: Equity transmission mechanisms from asia to australia: interdependence or contagion? *Australian Journal of Management*, **28**(2), 157–182.

Dungey, M., R. A. Fry, and V. L. Martin, 2004: Currency market contagion in the asia-pacific region. *Australian Economic Papers*, **43**(4), 379–395.

Efron, B., 1979: Bootstrap methods: another look at the jacknife. *Annals of Statistics*, **7**, 1–26.

Ellis, L. and E. Lewis, 2001: The response of financial markets in australia and new zealand to news about the asian crisis. RBA Research Discussion Papers RDP2001-03, Reserve Bank of Australia.

Engle, R., 2002: Dynamic conditional correlation - a simple class of garch models. *Journal of Business and Economic Statistics*, **20**, 339–350.

Engle, R. and J. G. Rangel, 2008: The factor-spline-garch model for high and low frequency correlations. Working paper.

Engle, R. and J. Russell, 1998: Autoregressive conditional duration: a new model for irregularly spaced transaction data. *Econometrica*, **66**(5), 1127–1162.

Epps, T., 1979: Comovements in stock prices in the very short run. *Journal of the American Statistical Association*, **74**, 291–298.

Fan, J. and Y. Wang, 2008: Spot volatility estimation for high-frequency data. *Statistics and Its Interface*, **1**, 279–288.

Fan, J. and Q. Yao, 2003: *Nonlinear time series.* Springer.

Fang, Y., 1996: Volatility modeling and estimation of high-frequency data with gaussian noise. Phd thesis, MIT, Sloan School of Management.

Florens-Zmirou, D., 1993: On estimating the diffusion coefficient from discrete observations. *Journal of Applied Probability*, **30**, 790–804.

Foster, D. and D. Nelson, 1996: Continuous record asymptotics for rolling sample variance estimators. *Econometrica: Journal of the Econometric Society*, **64**(1), 139–174.

Genon-Catalot, V., C. Laredo, and D. Picard, 1992: Non-paramteric estimation of the diffusion coefficient by wavelets methods. *Scandinavian Journal of Statistics*, **19**, 317–335.

Griffin, J. E. and R. Oomen, 2008: Sampling returns for realized variance calculations: tick time or transaction time? *Econometric Reviews*, **27**, 230–253.

Griffin, J. E. and R. C. A. Oomen, 2010: Covariance measurement in the presence of non-synchronous trading and market microstructure noise. Forthcoming in the Journal of Econometrics.

Hall, P., J. Marron, and B. U. Park, 1992: Smoothed cross-validation. *Probability Theory and Related Fields*, **92**, 1–20.

Hansen, P. and A. Lunde, 2005: A forecast comparison of volatility models: does anything beat a GARCH(1,1)? *Journal of Applied Econometrics*, **20**(7), 873–890.

Hansen, P. R. and A. Lunde, 2006: Realized variance and market microstructure noise. *Journal of Business and Economic Statistics*, **24**, 127–161.

Harris, L., 1990: Estimation of stock price variances and serial covariances frome discrete observations. *Journal of Financial and Quantitative Analysis*, **25**, 291–306.

Hart, J. D. and S. Yi, 1998: One-sided cross-validation. *Journal of the American Statistical Association*, **93**, 620–631.

Hasbrouck, J., 1993: Assessing the quality of a security market: a new approach to transaction-cost measurement. *Review of Financial Studies*, **6**(1), 191–212.

Hasbrouck, J., 1996: Modelling market microstructure time series. In *Statistical Methods in Fianance*, Maddala, G. S. and R., R. C., editors, volume 14 of *Handbook of Statistics*. North Holland, Amsterdam, 647-692.

Hayashi, T. and S. Kusuoka, 2004: Nonsynchronous covariation measurement for continuous semimartingales. Preprint Series of the Graduate School of Mathematical Sciences, The University of Tokyo.

Hayashi, T. and N. Yoshida, 2005: On covariance estimation of non-synchronously observed

diffusion proce sses. *Bernoulli*, **11**, 359–379.

Hayashi, T. and N. Yoshida, 2006: Estimating correlations with nonsynchronous observations in continuous diffusion models. Working paper.

Heston, S., 1993: A closed-form solution for options with stochastic volatility with applications to bond and currency options. *Review of financial studies*, **6**, 327–343.

Hoffmann, M., 1999: L_p estimation of the diffusion coefficient. *Bernoulli*, **5**(3), 447–481.

Hoffmann, M., 2002: Rate of convergence for parametric estimation in a stochastic volatility model. *Stochastic Processes and Their Applications*, **97**, 147–170.

Høg, E. and A. Lunde, 2003: Wavelet estimation of integrated volatility. Unbublished, Aarhus School of Business.

Hoshikawa, T., K. Nagai, T. Kanatani, and Y. Nishiyama, 2008: Nonparametric estimation methods of integrated multivariate volatilities. *Econometric Reviews*, **27**(1-3), 112–138.

Jacod, J., 1997: On continuous conditional gaussian martingales and stable convergence in law. In *Lecture Notes in Math.*, Seminaire de Probabilits. Springer, Berlin, 232-246.

Jacod, J., 2007: Statistics and high-frequency data. Lecture notes of SEMSTAT course in La Manga. To appear.

Jacod, J., Y. Li, P. Mykland, M. Podolskij, and M. Vetter, 2007: Microstructure noise in the continuous case: the pre-averaging approach. *Stochastic Processes and Their Applications*. Forthcoming.

Jacod, J. and A. N. Shiryaev, 2003: *Limit Theorems for Stochastic Processes*. Springer.

Jiang, G., 1998: Nonparametric modeling of U.S. interest rate term structure dynamics and implications on the prices of derivative securities. *Journal of Financial and Quantitative Analysis*, **33**(4), 465–497.

Johannes, M., 2004: The statistical and economic role of jumps in continuous-time interest rate models. *Journal of Finance*, **59**, 227–260.

Kanatani, T., 2004: Integrated volatility measuring from unevenly sampled observations. *Economics Bulletin*, **3**(36), 1–8.

Karatzas, I. and E. Shreve, 1988: *Brownian Motion and Stochastic Calculus*. Springer-Verlag.

Kloeden, P. and E. Platen, 1999: *Numerical Solutions of Stochastic Differential Equations*. Springer-Verlag, 2nd edition.

Kristensen, D., 2010: Nonparametric filtering of the realised spot volatility: a kernel-based approach. *Econometric Theory*, **26**, 69–93. Forthcoming.

Kullback, S. and R. A. Leibler, 1951: On information and sufficiency. *The Annals of Mathematical Statistics*, **22**, 79–86.

Ledoit, O. and Wolf, 2003: Improved estimation of the covariance matrix of stock returns with an application to portfolio selection. *Journal of Empirical Finance*, **10**(5), 603–621.

Ledoit, O. and M. Wolf, 2004a: Honey, i shrunk the sample covariance matrix. *Journal of Portfolio Management*, **30**(4), 110–119.

Ledoit, O. and M. Wolf, 2004b: A well-conditioned estimator for large-dimensional covariance matrices. *Journal of Multivariate Analysis*, **88**(2), 365–411.

Lee, S. and P. Mykland, 2008: Jumps in financial markets: A new nonparametric test and jump dynamics. *Review of Financial studies*, **21**(6), 2535.

Lo, A. W. and A. C. MacKinlay, 1990: An econometric analysis of nonsynchronous trading. *Journal of Econometrics*, **45**, 181–212.

Longin, F. and B. Solnik, 1995: Is the correlation in international equity returns constant: 1960-1990. *Journal of International Money and Finance*, **14**, 3–26.

Malliavin, P. and M. Mancino, 2002: Fourier series method for measurement of multivariate volatilities. *Finance & Stochastics*, **6**(1), 49–61.

Malliavin, P. and M. Mancino, 2009: A Fourier transform method for nonparametric estimation of volatility. *Annals of Statistics*, **37**(4), 1983–2010.

Malliavin, P. and A. Thalmaier, 2005: *Stochastic Calculus of Variations in Mathematical Finance*. Springer Finance.

Mancini, C., 2009: Non-parametric threshold estimation for models with stochastic diffusion coefficient and jumps. *Scandinavian Journal of Statistics*, **36**(2), 270–296.

Mancini, C. and R. Renò, 2009: Threshold estimation of Markov models with jumps and interest rate modeling. *Journal of Econometrics*. Forthcoming.

Mancino, M. and S. Sanfelici, 2008a: Estimating covariance via fourier method in the presence of asynchronous trading and microstructure noise. SSRN Electronic Library.

Mancino, M. and S. Sanfelici, 2008b: Robustness of Fourier estimator of integrated volatility in the presence of microstructure noise. *Computational Statistics and Data Analysis*, **52**(6), 2966–2989.

Mantegna, R., 1999: Hierarchical structure in financial markets. *European Physics Journal B*, **11**, 193–197.

Maringer, D. G. and M. Meyer, 2008: Smooth transition autoregressive models: new approaches to the model selection problem. *Studies in Nonlinear Dynamics & Econometrics*, **12**(1), Article 5.

Markowitz, H. M., 1959: *Portfolio Selection: Efficient Diversification of Investments*. Monograph 16. Cowles Foundation, Yale University.

Martens, M., 2004: Estimating unbiased and precise realized covariances. SSRN Electronic Library.

Milton, F., 1999: How asia fell. In *Hoover Digest No 2*, Hoover Institution, Stanford University.

Musiela, M. and M. Rutkowski, 1998: *Martingale Methods in Financial Modelling*. Springer-Verlag.

Mykland, P. and L. Zhang, 2006: ANOVA for diffusions and Ito processes. *Annals of Statistics*, **34**(4), 1931.

Mykland, P. and L. Zhang, 2008: Inference for volatility-type objects and implications for hedging. *Statistics and Its Interface*, **1**, 255–278.

Nelson, D., 1990: ARCH models as diffusion approximations. *Journal of Econometrics*, **45**, 7–38.

Nielsen, M. O. and P. H. Frederiksen, 2008: Finite sample accuracy and choice of sampling frequency in integrated volatility estimation. *Journal of Empirical Finance*, **15**, 265–286. Forthcoming.

Ogawa, S. and S. Sanfelici, 2008: An improved two-step regularization scheme for spot volatility estimation. Working paper.

Pagan, A. and A. Ullah, 1999: *Nonparametric Econometrics*. Cambdrige University Press.

Park, B. and J. Marron, 1990: Comparison of data driven bandwidth selectors. *Journal of the American Statistical Association*, **85**, 66–72.

Podolskij, M. and M. Vetter, 2009: Estimation of volatility functionals in the simultaneous presence of microstructure noise and jumps. *Bernoulli*, **15**(3), 634–668.

Podolskij, M. and M. Vetter, 2010: Understanding limit theorems for semimartingales: a short survey. *Statistica Nederlandica*, **64**(3), 329–351.

Precup, O. and G. Iori, 2007: Cross-correlation measures in the high-frequency domain. *European Journal of Finance*, **13**(4), 319–331.

Price, K., R. Storn, and J. Lampinen, 2005: *Differential Evolution: a practical approach to global optimization.* Springer.

Priestley, M., 1979: *Spectral Time Series Analysis.* Wiley.

Rebonato, R. and P. Jackel, 1999: The most general methodology for creating a valid correlation matrix for risk management and option pricing purposes. *The Journal of Risk*, **2**(2), 17–27.

Renò, R., 2003: A closer look at the Epps effect. *International Journal of Theoretical and Applied Finance*, **6**(1), 87–102.

Renò, R., 2008: Nonparametric estimation of the diffusion coefficient of stochastic volatility models. *Econometric Theory*, **24**(5), 1174–1206.

Revuz, D. and M. Yor, 2001: *Continuous Martingales and Brownian Motion.* Springer.

Roll, R., 1984: A simple measure of the implicit bid-ask spread in an efficient market. *Journal of Finance*, **39**, 1127–1139.

Ross, S. A., 1976: The arbitrage theory of capital asset pricing. *Journal of Economic Theory*, **13**, 341–360.

Rudemo, M., 1982: Empirical choice of histograms and kernel density estimators. *Scandinavian Journal of Statistics*, **9**, 65–78.

Scott, D., 1992: *Multivariate Density Estimation.* John Wiley & Sons.

Scott, D., R. Tapia, and J. Thompson, 1977: Kernel density estimation revisited. *Nonlinear Analysis, Theory, Methods and Applications*, **1**, 339–372.

Scott, D. and G. Terrell, 1987: Biased and unbiased cross-validation in density estimation. *Journal of American Statistical Association*, **82**, 1131–1146.

Sheather, S. and M. Jones, 1991: A reliable data-based bandwidth selection method for kernel density estimation. *Journal of the Royal Statistical Society, Ser. B*, **53**, 683–690.

Silverman, B. W., 1986: *Density Estimation for Statistics and Data Analysis.* Chapman & Hall/CRC.

Skintzi, V. and A. Refens, 2005: Implied correlation index: a new measure of diversification. *Journal of Future Markets*, **25**, 171–195.

Storn, R. and K. Price, 1997: Differential evolution: a simple and efficient heuristic for global optimization over continuous spaces. *Journal of Global Optimization*, **11**(4), 341–359.

Todorov, V. and G. Tauchen, 2010: Volatility jumps. Forthcoming in the Journal of Business and Economic Statistics.

Tumminello, M., F. Lillo, and R. N. Mantegna, 2007a: Hierarchical nested factor model from multivariate data. *Europhysics Letters*, **78**, 30006.

Tumminello, M., F. Lillo, and R. N. Mantegna, 2007b: Kullback-leibler distance as a measure of the information filtered from multivariate data. *Physical Review E*, **76**, 031123.

van Emmerich, C., 2006: Modelling correlation as a stochastic process. *Preprint BUW-AMNA 06/03*. Bergische Universität Wuppertal.

Vasicek, O., 1977: An equilibrium characterization of the term structure. *Journal of Financial Economics*, **5**, 177–188.

Veraart, A., 2010: Inference for the jump part of quadratic variation of itô semimartingales. *Econometric Theory*, **26**(2), 331–368.

Voev, V. and A. Lunde, 2007: Integrated covariance estimation using high-frequency data in the presence of noise. *Journal of Financial Econometrics*, **5**, 68–104.

Walter, G. and J. Blum, 1979: Probability density estimation using delta sequences. *The Annals of Statistics*, **6**(2), 328–340.

Watson, G. and M. Leadbetter, 1964: Hazard analysis II. *Sankhya, Ser. A*, **26**, 101–116.

Woodroofe, M., 1970: On choosing a delta-sequence. *Annals fo Mathematical Statistics*, **41**, 1665–1671.

Zhang, L., 2006a: Efficient estimation of stochastic volatility using noisy observations : A multi-scale approach. *Bernoulli*, **12**(6), 1019–1043.

Zhang, L., 2006b: Estimating covariation: Epps effect, microstructure noise. Working Paper.

Zhang, L., P. A. Mykland, and Y. Aït-Sahalia, 2005: A tale of two time scales: Determining integrated volatility with noisy high-frequency data. *Journal of the American Statistical Association*, **100**, 1394–1411.

Zhang, S., R. Karunamuni, and M. Jones, 1999: An improved estimator of the density function at the boundary. *Journal of the American Statistical Association*, **94**, 1231–1241.

Zhou, B., 1996: High-frequency data and volatility in foreign-exchange rates. *Journal of Business and Economics Statistics*, **14**(1), 45–52.