Contents lists available at ScienceDirect

Mathematics and Computers in Simulation

journal homepage: www.elsevier.com/locate/matcom

Original articles

The Fourier–Malliavin Volatility (FMVol) MATLAB[®] library

Simona Sanfelici^{a,*}, Giacomo Toscano^b

^a Università degli Studi di Parma, Via Kennedy, 6, Parma, 43125, Italy

^b Università degli Studi di Firenze, Via delle Pandette, 9, Firenze, 50127, Italy

ARTICLE INFO

Keywords: Non-parametric estimation Fourier analysis Volatility Quarticity Volatility of volatility Leverage MATLAB®

ABSTRACT

This paper presents the Fourier–Malliavin Volatility (FMVol) estimation library for MATLAB[®]. This library includes functions that implement Fourier–Malliavin estimators (see Malliavin and Mancino (2002, 2009)) of the volatility and co-volatility of continuous stochastic volatility processes and second-order quantities, like the quarticity (the squared volatility), the volatility of volatility and the leverage (the covariance between changes in the process and changes in its volatility). The Fourier–Malliavin method is fully non-parametric, does not require equally-spaced observations and is robust to measurement errors, or noise, without any preliminary bias correction or pre-treatment of the observations. Furthermore, in its multivariate version, it is intrinsically robust to irregular and asynchronous sampling. Although originally introduced for a specific application in financial econometrics, namely the estimation of asset volatilities, the Fourier–Malliavin method is a general method that can be applied whenever one is interested in reconstructing the latent volatility and second-order quantities of a continuous stochastic volatility process from discrete observations.

1. Introduction

The accurate estimation of the latent diffusion component of a stochastic process, which gauges the degree of unpredictability, or randomness, of its trajectories, represents a relevant issue in several fields. For instance, in high-frequency financial econometrics, the price of an asset (e.g., a stock, a bond, a currency) is typically modelled as an Itô stochastic differential equation (see, e.g., Chapter 1 in [1]) and the focus is on the estimation of the diffusion coefficient thereof, which is commonly referred to as the *volatility* process.¹ The availability of efficient estimates of asset volatility paths is indeed fundamental for many financial applications, e.g., portfolio allocation, market risk management and derivatives pricing. However, the issue of estimating the volatility of a random process is not exclusive to financial econometrics but is relevant also for applications in other fields, such as, for instance, climate change studies, computational biology, and medicine (see, e.g., [11,27,28,34], respectively).

The Fourier–Malliavin method, originally introduced in [16,17], represents an efficient and easily implementable tool to address this task in a multivariate setting. As detailed in Section 3, the method involves three steps. Firstly, one computes the Fourier coefficients of the increments of the variables of interest, e.g., asset price returns. Then, these coefficients are exploited to reconstruct those of the latent co-volatility matrix, using a convolution formula. Finally, co-volatility paths are obtained from the co-volatility coefficients via the Fourier–Fejer inversion formula.

The possibility of recovering the Fourier coefficients of the latent co-volatility matrix from those of the increments of the observed process may not come as a surprise if one recalls the notion of *Fourier transform*, an integral transform that allows the

¹ Throughout this paper, we will use the terms *volatility* and *variance* as synonyms, as is common practice in the literature on financial econometrics.

https://doi.org/10.1016/j.matcom.2024.07.003

Received 31 January 2024; Received in revised form 25 May 2024; Accepted 3 July 2024

Available online 14 July 2024

0378-4754/© 2024 The Author(s). Published by Elsevier B.V. on behalf of International Association for Mathematics and Computers in Simulation (IMACS). This is an open access article under the CC BY license (http://creativecommons.org/licenses/by/4.0/).







^{*} Corresponding author.

E-mail addresses: simona.sanfelici@unipr.it (S. Sanfelici), giacomo.toscano@unifi.it (G. Toscano).

representation of a periodic function as a linear combination of projections on a trigonometric basis, termed the *frequency-domain representation*. Specifically, the term Fourier transform refers to both the frequency-domain representation and the mathematical operation that links the latter to a function of time. The Fourier transform represents a flexible tool that has been widely used for applications in different fields, including engineering, physics and also finance. Indeed, operations performed in one domain (time or frequency) have corresponding operations in the other domain that may sometimes be easier to perform. For example, the operation of convolution in the time domain corresponds to multiplication in the frequency domain and vice versa. Therefore, one may apply the Fourier transform to a given function, perform the desired operations more easily in the frequency domain, and finally transform the result back to the time domain by Fourier inversion. For further details, see, e.g., [4,12,29].

In this paper, we introduce the Fourier–Malliavin Volatility (FMVol) estimation library for MATLAB[®]. The primary function of this library allows one to estimate the co-volatility matrix of a *d*-dimensional stochastic process from discrete observations of the latter. Further, the library also yields univariate estimates of the quarticity (i.e the squared volatility), the volatility of the volatility process and the leverage (namely, the covariance between changes in the observable variable of interest and changes in its latent volatility).² The Fourier coefficients of the volatility of volatility and the leverage are obtained through an iteration of the convolution formula (see Section 3.2). The Fourier coefficients of the quarticity, instead, are obtained through a product formula (see Section 3.3). Note that the library produces both instantaneous (spot) estimates, that is, estimates of trajectories on a discrete grid, and integrated estimates over the observation period. Other existing software implementations are provided by [5], who examine the computational cost of different implementations and the time-scale behaviour of the Fourier–Malliavin estimator.

The Fourier–Malliavin method is intrinsically robust to asynchronous and irregular sampling and to the presence of noise in the observations, due for instance to measurement errors. In financial applications, such noise is typically due to the presence of market microstructure (see Section 4.2). Moreover, its computational stability is ensured by the fact that it is based only on integration procedures. Indeed, alternative methods for the estimation of the covariance path are typically based on numerical differentiation (see Chapter 8 of [1]), which is more vulnerable to numerical instabilities. It is worth stressing that, even though the Fourier method has been introduced to address a specific issue in financial econometrics, namely the estimation of asset co-volatilities with high-frequency prices, its applicability is more general, that is, the method can be applied whenever one is interested in estimating the latent diffusion component of a (multi-dimensional) stochastic process. An application to atmospheric temperature data is discussed in Section 5.2.

The outline of the paper is as follows. Section 2 recalls the notion of Fourier transform, while Section 3 illustrates the Fourier–Malliavin estimation method. The FMVol library is introduced in Section 4 and Section 5 contains two empirical applications of the latter, with financial and weather data, respectively. Finally, Section 6 concludes.

2. The Fourier transform and Fejer's convergence theorem

Given a function f defined and integrable on [0, T], for any integer k, the kth Fourier coefficient is defined as

$$\mathcal{F}(f)(k) := \frac{1}{T} \int_0^T f(t) e^{-i\frac{2\pi}{T}kt} dt,$$
(1)

where $i = \sqrt{-1}$. When the independent variable *t* represents time, the transform variable *k* represents the frequency. Accordingly, $\mathcal{F}(f)(k)$ is also called the Fourier transform of *f* at frequency *k*. Evaluating $\mathcal{F}(f)(k)$ for all values of *k* produces the frequency-domain function. For example, if time is measured in seconds, then the frequency is measured in hertz.

By the Fejer theorem, if the function f is continuous on [0, T], the trigonometric series

$$\sum_{|k| \le N} \left(1 - \frac{|k|}{N+1} \right) \mathcal{F}(f)(k) e^{i\frac{2\pi}{T}kt}$$
(2)

converges uniformly and in mean square to f(t) on [0,T] as $N \to \infty$, see, e.g., [15].

Finally, we define

$$\mathcal{F}(df)(k) := \frac{1}{T} \int_0^{2\pi} e^{-i\frac{2\pi}{T}kt} df(t).$$
(3)

Using integration by parts, one obtains that

$$\mathcal{F}(df)(k) = ik \frac{2\pi}{T} \mathcal{F}(f)(k) + \frac{f(T) - f(0)}{T}.$$
(4)

3. The Fourier-Malliavin estimation method

We describe the Fourier–Malliavin estimation method, originally introduced in [16,17]. The building block of the method is the *convolution formula*, which yields an estimator of the Fourier coefficients of the co-variation of two processes. Firstly, we present this formula in the case of a *d*-dimensional diffusion process. Then we address iterated covariances. Finally, we examine the *product formula*, which may be needed for estimating auxiliary quantities. An example of auxiliary quantity is provided by the quarticity. The latter appears in the asymptotic error variance of volatility estimators (see, e.g., [14]) and thus its estimation is needed for obtaining a feasible Central Limit Theorem.

² The term leverage comes from the literature on financial economics, where the (usually negative) covariance between asset price increments and volatility increments is typically described as a result of the 'leverage effect'. See, e.g., [6] for an economic explanation of this empirical stylized fact.

3.1. Convolution formula

Consider the *d*-variate stochastic differential system

$$dx^{j}(t) = b^{j}(t)dt + \sum_{i=1}^{l} \sigma_{i}^{j}(t)dW^{i}(t), \qquad x^{j}(0) = x_{0}^{j} \in \mathbb{R}, \quad j = 1, \dots, d,$$
(5)

where W^1, \ldots, W^l are *l* independent Brownian motions, while, for any *i* and any *j*, σ_i^j and b^j are random processes, adapted to the Brownian filtration.

The variance–covariance matrix $\Sigma(t)$, whose entries are equal to

$$\Sigma^{i,j}(t) = \sum_{k=1}^{l} \sigma_k^i(t) \sigma_k^j(t), \quad i, j = 1, \dots, d,$$

is also called volatility matrix.

In the case of d dependent Brownian motions, if we let the system of stochastic differential equations read as

$$dx^{j}(t) = b^{j}(t)dt + \sigma^{j}(t)dW^{j}(t), \quad x^{j}(0) = x_{0}^{j} \in \mathbb{R}, \quad j = 1, \dots, d,$$
(6)

the volatility matrix $\Sigma(t)$ is then given by

$$\Sigma^{i,j}(t) = \sum_{k=1}^{i} \sigma_k^{i}(t) \sigma_k^{j}(t) \rho^{i,j}(t), \quad i, j = 1, \dots, d,$$

where $\rho^{i,j}(t)$ is the instantaneous correlation at time t between the Brownian motions W^{i} and W^{j} .

From the mathematical point of view, the entries of $\Sigma^{i,j}(t)$ are related to the quadratic co-variation process $\langle x^i, x^j \rangle_t^3$ by the relation

$$d\langle x^i, x^j \rangle_t = \Sigma^{i,j}(t)dt$$

This remark is at the basis of the Fourier-Malliavin estimation method of covariances.

For any integer k, denote by $\mathcal{F}(dx^i)(k)$ and $\mathcal{F}(\Sigma^{i,j})(k)$ the Fourier coefficients of the increments of x^i and of the entries of the volatility matrix, respectively (see definitions (3) and (1)). The fundamental result of [16,17] is that for any integer k, it is possible to compute the Fourier coefficients $\mathcal{F}(\Sigma^{i,j})(k)$ of the latent spot covariances $\Sigma^{i,j}(t)$ by means of the Fourier coefficients $\mathcal{F}(dx^i)(k)$ and $\mathcal{F}(dx^j)(k)$. Precisely, for any i, j = 1, ..., d, it holds that

$$\frac{1}{T}\mathcal{F}(\Sigma^{i,j}) = \mathcal{F}(dx^i) * \mathcal{F}(dx^j),\tag{7}$$

where the convolution product in (7) is defined as

$$(\mathcal{F}(dx^{i}) * \mathcal{F}(dx^{j}))(k) := \lim_{N \to \infty} \frac{1}{2N+1} \sum_{|s| \le N} \mathcal{F}(dx^{i})(s) \mathcal{F}(dx^{j})(k-s),$$
(8)

for any i, j and for all integers k. The convergence of the convolution product in (8) is attained in probability.

Once the Fourier coefficients of the volatility matrix have been computed, it is possible to reconstruct the spot volatility matrix $\Sigma(t)$ using the Fourier–Fejer inversion formula given in (2). Specifically, if the volatility matrix has continuous paths, namely the functions $t \to \Sigma^{i,j}(t)$ are continuous,⁴ then the Fourier–Fejer summation gives, almost surely, that

$$\lim_{M \to \infty} \sum_{|k| < M} \left(1 - \frac{|k|}{M+1} \right) \mathcal{F}(\Sigma^{i,j})(k) e^{i\frac{2\pi}{T}kt} = \Sigma^{i,j}(t), \quad \text{for all } t \in (0,T),$$
(9)

for i, j = 1, ..., d.

When the diffusion processes are known by discrete observations, we have to define discrete analogues of the quantities introduced above. Without loss of generality, let us consider the case of two processes, observed on the discrete grids $\{0 = t_0^j < t_1^j < \cdots < t_{n_j}^j = T\}$, j = 1, 2. It is worth noting that we allow for irregularly spaced and asynchronous observation times.

For any integer k, $|k| \le 2N$, let us define the discrete Fourier transform of the increment of x^i as

$$c_k(dx_{n_i}^i) := \frac{1}{T} \sum_{l=0}^{n_i-1} e^{-i\frac{2\pi}{T}kt_l^i} \delta_l(x^j), \tag{10}$$

$$\langle X, Y \rangle_t := \lim_{\|II\| \to 0} \sum_{i=1}^n (X_{t_i} - X_{t_{i-1}}) (Y_{t_i} - Y_{t_{i-1}}),$$

where the limit, if it exists, is defined using convergence in probability.

⁴ If $\Sigma^{i,j}(t)$ has cadlag paths, then the limit in (9) gives $(\Sigma^{i,j}(t) + \Sigma^{i,j}(t^{-}))/2$, see, e.g., [15].

³ Let Π represent a partition of the interval [0,t] and $\|\Pi\|$ represent the mesh size of Π . The quadratic co-variation between the stochastic processes X and Y is defined as

S. Sanfelici and G. Toscano

where $\delta_l(x^i) := x^i(t^i_{l+1}) - x^i(t^i_l)$, $l = 0, ..., n_i - 1$, i = 1, 2. For any $|k| \le N$ and i, j = 1, 2, let us consider the discrete analogue of the convolution in (8), that is,

$$\frac{1}{2N+1} \sum_{|s| \le N} c_s(dx_{n_i}^i) c_{k-s}(dx_{n_j}^j).$$

Given the identity in (7), the previous term, when multiplied by *T*, provides an estimate of the *k*th Fourier coefficient of $\Sigma^{i,j}$. Therefore, we define

$$c_k(\Sigma_{n_i,n_j,N}^{i,j}) := \frac{T}{2N+1} \sum_{|s| \le N} c_s(dx_{n_i}^i) c_{k-s}(dx_{n_j}^j).$$
(11)

Finally, the random function of time

$$\widehat{\Sigma}_{n_{i},n_{j},N,M}^{i,j}(t) := \sum_{|k| \le M} \left(1 - \frac{|k|}{M+1} \right) c_{k} (\Sigma_{n_{i},n_{j},N}^{i,j}) e^{i\frac{2\pi}{T}kt}, \ t \in (0,T),$$
(12)

will be called the *Fourier–Malliavin estimator* of the instantaneous volatility matrix $\Sigma^{i,j}(t)$.

The Fourier estimator is consistent in the in-fill asymptotics sense, that is, if $\rho^{1,2} := \max\left(\max_{1,\dots,n_1} |t_i^1 - t_{i-1}^1|, \max_{1,\dots,n_1} |t_i^2 - t_{i-1}^2|\right)$ tends to zero (or, equivalently, the sampling frequency tends to infinity), and if suitable assumptions on the growing rate of the cutting frequencies *N* and *M* are satisfied, then the estimator converges in probability to $\Sigma^{i,j}(t)$, i, j = 1, 2. The choice of suitable values for the cutting frequencies *N* and *M* is of the utmost importance to achieve accurate estimates from (potentially noise-affected) sample observations. We refer the reader to [16–19] for details on asymptotic results.

Additionally, based on (11), a consistent estimator of the *integrated volatility matrix* $\int_0^T \Sigma^{i,j}(t) dt$ is given by

$$\hat{\Sigma}_{n_i,n_j,N}^{i,j} := T \ c_0(\Sigma_{n_i,n_j,N}^{i,j}) = \frac{T^2}{2N+1} \sum_{|s| \le N} c_s(dx_{n_i}^i) c_{-s}(dx_{n_j}^j).$$

Alternatively, a Fejer kernel can be added

$$\hat{\Sigma}_{n_i,n_j,N}^{i,j} \coloneqq \frac{T^2}{N+1} \sum_{|s| \le N} \left(1 - \frac{|s|}{N+1} \right) c_s(dx_{n_i}^i) c_{-s}(dx_{n_j}^j) \tag{13}$$

to have a more efficient estimator of the integrated volatility matrix. The finite sample properties of the Fourier–Malliavin estimators of the integrated co-volatilities have been studied in [21,23].

Finally, some comments are in order. First, note that we consider inference problems that are defined on a finite time horizon [0, T] and are based on in-fill asymptotics. Such inference problems differ from those that one usually encounters in time-series analysis, where the sampling interval is fixed and T goes to infinity. Operating on a finite horizon allows us to handle samples from data-generating processes that fail to satisfy stationarity or ergodic properties, which are crucial for long-span $(T \to \infty)$ asymptotics. Moreover, our inference problems are also characterized by the fact that we can observe only a single path of the processes of interest and thus we estimate path-wise realized co-volatility trajectories.

3.2. Iterated convolutions

The knowledge of the Fourier coefficients of the latent instantaneous variances and covariances allows us to handle these processes as observable variables so that we can iterate the convolution formula and compute, for instance, the Fourier coefficients of the co-variation $B^i(t)$ between the stochastic-variance process $\Sigma^{i,i}$ and the process x^i , or those of the quadratic variation $C^i(t)$ of the stochastic-variance process $\Sigma^{i,i}$. In the following sections, we will omit the index *i* to simplify notations. We refer the reader to the bibliography for a deeper understanding of the different estimators and more specifically to [8–10,25,30–32] for asymptotic results and finite-sample properties.

Indeed, formula (7) can be generalized to estimate the Fourier coefficients of the instantaneous co-variation φ of any two square-integrable processes *X* and *Y* defined on [0,*T*], that is, we can define

$$\mathcal{F}(\varphi)(k) := \lim_{N \to \infty} \frac{T}{2N+1} \sum_{|s| \le N} \mathcal{F}(dX)(s) \mathcal{F}(dY)(k-s), \tag{14}$$

where the convergence of the convolution product is attained in probability. Then, we can reconstruct the process φ as

$$\varphi(t) = \lim_{M \to \infty} \sum_{|k| \le M} \left(1 - \frac{|k|}{M+1} \right) \mathcal{F}(\varphi)(k) e^{i\frac{2\pi}{T}kt}.$$
(15)

However, we stress the fact that the resolution at which we can reconstruct the process reduces at each iteration of the convolution formula (see also Section 4.3).

Finally, we remark that this procedure allows us to compute also integrated quantities on [0, T], that is, $\int_0^T \varphi(t) dt$ is obtained as

$$T \mathcal{F}(\varphi)(0) = \lim_{N \to \infty} \frac{T^2}{2N+1} \sum_{|s| \le N} \mathcal{F}(dX)(s) \mathcal{F}(dY)(-s).$$
(16)

3.2.1. Example 1. The leverage

Consider from now on the diffusion process

$$dx(t) = b(t)dt + \sigma(t)dW(t), \qquad x(0) = x_0 \in \mathbb{R},$$

where the volatility process is driven by the stochastic differential equation

$$d\sigma^{2}(t) = a(t)dt + \gamma(t)dZ(t), \qquad \sigma^{2}(0) = \sigma_{0}^{2} > 0.$$

The Brownian motions *W* and *Z* may be correlated, with instantaneous correlation ρ . The processes $a(\cdot)$ and $\gamma(\cdot)$ are the drift and volatility of the variance process.

We use the formula in (14) to estimate the *leverage*, i.e. the co-variation $B(t) := \rho\sigma(t)\gamma(t)$ between σ^2 and the process *x*. We assume that the diffusion process *x* is known by *n* discrete observations. The estimator of the *k*th Fourier coefficients of B(t) is defined by

$$c_k(B_{n,N,M}) := \frac{T}{2M+1} \sum_{|j| \le M} c_j(dx_n) c_{k-j}(d\sigma_{n,N}^2), \tag{17}$$

where $c_j(d\sigma_{nN}^2)$ is computed as⁵

$$c_{j}(d\sigma_{n,N}^{2}) = i j \frac{2\pi}{T} c_{j}(\sigma_{n,N}^{2}).$$
(18)

Note that the estimator (17) ultimately depends only on the Fourier coefficients $c_j(dx_n)$, which are computed from sample observations. In fact, the Fourier coefficients of the variance $c_j(\sigma_{n,N}^2)$ have been estimated in the previous step via (11) with j = i, which depends only on $c_j(dx_n)$.

The leverage process can therefore be estimated via

$$\widehat{B}_{n,N,M,L}(t) := \sum_{|k| \le L} \left(1 - \frac{|k|}{L+1} \right) c_k(B_{n,N,M}) e^{i\frac{2\pi}{T}kt}.$$
(19)

Finally, based on (16), a consistent estimator of the *integrated leverage* $\int_0^T B(t)dt$ is given by

$$\widehat{B}_{n,N,M} := T \ c_0(B_{n,N,M}) = \frac{T^2}{M+1} \sum_{|j| \le M} ij \frac{2\pi}{T} \left(1 - \frac{|j|}{M+1} \right) c_j(dx_n) c_{-j}(\sigma_{n,N}^2),$$

where the Fejer kernel has been included to reduce the variance of the estimation error.

3.2.2. Example 2. The volatility of volatility

We can use the formula in (14) also to estimate *the volatility of volatility*, i.e. the quadratic variation C(t) of the stochastic variance process $\sigma^2(t)$, given by $\gamma^2(t)$. We assume again that the diffusion processes are known by *n* discrete observations. The estimator of the *k*th Fourier coefficients of C(t) is defined by

$$c_k(C_{n,N,M}) := \frac{T}{2M+1} \sum_{|j| \le M} c_j(d\sigma_{n,N}^2) c_{k-j}(d\sigma_{n,N}^2),$$
(20)

Based on (18), the estimator in (20) depends on the Fourier coefficients of the volatility, $c_j(\sigma_{n,N}^2)$, which in turn (see (11)) depend only on the Fourier coefficients $c_j(dx_n)$.

The volatility of volatility process can therefore be estimated by

$$\widehat{C}_{n,N,M,L}(t) := \sum_{|k| \le L} \left(1 - \frac{|k|}{L+1} \right) c_k(C_{n,N,M}) e^{i\frac{2\pi}{T}kt}.$$
(21)

Finally, from (16) a consistent estimator of the *integrated volatility of volatility* $\int_0^T C(t)dt$ is given by

$$\widehat{C}_{n,N,M} := T \ c_0(C_{n,N,M}) = \frac{T^2}{2M+1} \sum_{|j| \le M} j^2 \left(\frac{2\pi}{T}\right)^2 \left(1 - \frac{|j|}{M+1}\right) c_j(\sigma_{n,N}^2) c_{-j}(\sigma_{n,N}^2),$$

where the Fejer kernel has again been included to reduce the variance of the estimation error.

3.3. Product formula

Besides iterated covariances, the Fourier–Malliavin methodology allows us also to estimate the product of two given processes. Given two square-integrable processes X and Y, defined on [0, T], the Fourier coefficients of their product XY can be obtained by the product formula

$$\mathcal{F}(XY)(k) = \lim_{M \to \infty} \sum_{|s| \le M} \mathcal{F}(X)(s)\mathcal{F}(Y)(k-s),$$
(22)

⁵ This formula derives from Eq. (4). Note that the summand f(T) - f(0) can be neglected in the construction of the estimator (17), see [8] for further details.

see [15]. Once again, the knowledge of the Fourier coefficients of the process XY allows us to reconstruct the latter via the Fourier–Fejer inversion formula.

3.3.1. Example 3. The quarticity

The result in (22) allows estimating the path of relevant statistical quantities such as the so-called *quarticity* $\sigma^4(t)$, namely the second power of the squared diffusion coefficient of the process *x*. Additionally, to estimate the integral of the product of two given processes, one only needs to compute the 0-th Fourier coefficient of the product. For instance, the *integrated quarticity* $\int_0^T \sigma^4(t) dt$ can be estimated via

$$T\mathcal{F}(\sigma^4)(0) = T \lim_{M \to \infty} \sum_{|s| \le M} \mathcal{F}(\sigma^2)(s)\mathcal{F}(\sigma^2)(-s).$$

As for the estimators of the leverage and the volatility of volatility, to obtain spot and integrated quarticity estimates via the Fourier– Malliavin method one needs to estimate the Fourier coefficients of the volatility but, nonetheless, the knowledge of the instantaneous volatility path is not required.

See [14,24] for further details on asymptotic results and finite-sample properties of the quarticity estimator.

4. The library

The MATLAB® codes for the Fourier-Malliavin estimators are available by getting the free add-on

Flexible Statistics and Data Analysis Toolbox (FSDA)

developed by the Department of Economics and Management and the Interdepartmental Centre of Robust Statistics (Ro.S.A.) of the University of Parma, and the Joint Research Centre of the European Commission.

All the .m source codes can be found after installing FSDA from the Mathworks file exchange

https://it.mathworks.com/matlabcentral/fileexchange/72999-fsda

The associated GitHub repo containing the latest versions of the codes can be found at

https://github.com/UniprJRC/FSDA

The related MATLAB[®] documentation is also available at

http://rosa.unipr.it/FSDA/index.html

under the category Fourier-Malliavin Volatility (FMVol) estimation.

The library contains functions to estimate the integrated and instantaneous variance/covariance, leverage, volatility of volatility and quarticity from sample observations of a diffusion process, possibly affected by measurement errors.

In this section, we describe in detail some of these functions and provide also some examples of their implementation with simulated data from a parametric stochastic volatility model. For a full comprehension of the totality of the routines, we refer the reader to the related MATLAB[®] documentation and the bibliography cited therein. The documentation contains several implementation examples, with different choices of the input variables of the functions. All functions included in the library are listed in Table 1 and are freely and easily downloadable from the GitHub repo.

4.1. Simulation of input data

We assume our time series data are discrete observations from a bi-variate diffusion process. To describe our MATLAB[®] library, we generate a single trajectory of a bi-variate version of the well-known stochastic volatility model by [13]:

$$dx^{1}(t) = \left(\mu_{1} - \frac{1}{2}v^{1}(t)\right)dt + \sqrt{v^{1}(t)}dW_{t}^{1}$$
(23)

$$dx^{2}(t) = \left(\mu_{2} - \frac{1}{2}v^{2}(t)\right)dt + \sqrt{v^{2}(t)}dW_{t}^{2}$$
(24)

$$dv^{1}(t) = \theta_{1}(\alpha_{1} - v^{1}(t))dt + \gamma_{1}\sqrt{v^{1}(t)}dW_{t}^{3}$$
(25)

$$dv^{2}(t) = \theta_{2}(\alpha_{2} - v^{2}(t))dt + \gamma_{2}\sqrt{v^{2}(t)}dW_{t}^{4}$$
(26)

where W^1, W^2, W^3, W^4 are correlated Brownian motions such that $\langle dW_t^k, dW_t^r \rangle = \rho_{k,r} dt$. The simulation approach allows us to appreciate the quality of our estimates. The Appendix describes how to simulate this model by using the MATLAB[®] functions Heston1D.m and Heston2D.m, which are available in our library.

For simplicity, we assume that observations from the bivariate model are synchronous. Accordingly, for any positive integer *n*, we let $S_n := \{0 = t_0 \le \dots \le t_n = T\}$ denote the set of observation times.

We simulate also the case when observations are affected by noise (see also [21]), that is, the case when one can only observe

$$\tilde{x}^{j}(t_{i}) = x^{j}(t_{i}) + \eta^{j}(t_{i}), \quad i = 1, \dots, n, \ j = 1, 2,$$
(27)

where the noise terms η^1 and η^2 are assumed to be i.i.d. sequences of random variables with mean equal to zero and finite variance. Moreover, they are assumed to be independent of each other.

S. Sanfelici and G. Toscano

Table 1

List of the primary functions (top) and supplementary functions (bottom) related to the Fourier-Malliavin estimation method that are available in the FMVol library of the FLEXIBLE STATISTICS AND DATA ANALYSIS TOOLEOX.

Function	Description
Heston1D.m	Generates a sample trajectory from the Heston model
Heston2D.m	Generates a sample trajectory from the bivariate Heston model
FM_spot_vol.m	Estimates the spot variance of a diffusion process
FM_spot_cov.m	Estimates the spot covariance of a bi-variate diffusion process
FM_spot_volvol.m	Estimates the spot volatility of volatility of a stochastic volatility process
FM_spot_lev.m	Estimates the spot leverage of a stochastic volatility process
FM_spot_quart.m	Estimates the spot quarticity of a diffusion process
FM_int_vol.m	Estimates the integrated variance of a diffusion process
FM_int_cov.m	Estimates the integrated covariance of a bi-variate diffusion process with Fejer kernel
FM_int_volvol.m	Estimates the integrated volatility of volatility of a stochastic volatility process
FM_int_lev.m	Estimates the integrated leverage of a stochastic volatility process
FM_int_quart.m	Estimates the integrated quarticity of a diffusion process
FM_cov_matrix.m	Estimates the integrated variance-covariance matrix on any number of diffusion processes
	with Fejer kernel
OptimalCuttingFrequency.m	Computes the optimal cutting frequencies for the Fourier estimator of the integrated variance
	in the presence of noise (see [21]).
Supplementary function	Description
FE_spot_vol.m	Estimates the instantaneous variance of a diffusion process by the Fourier estimator
FE_spot_vol_FFT.m	Estimates the spot variance of a diffusion process via the Fourier estimator with the FFT algorithm
FE_int_vol.m	Estimates the spot variance from a diffusion process via the Fourier estimator with Dirichlet kernel
FE_int_vol_Fejer.m	Estimates the spot variance from a diffusion process via the Fourier estimator with Fejer kernel

4.2. Variance and covariance estimation

The function

computes the spot volatility of a diffusion process via the Fourier-Malliavin estimator in (12) by using as input a vector of observations x and the related observation times t, the time horizon T, the cutting frequencies N and M and the vector tau of estimation times. If tau is not specified, then it is set equal to 0:T/(2*M):T and is given as an output variable tau_out. When not specified, default values for the cutting frequencies are

$$N = \lfloor n/2 \rfloor, M = \lfloor (n/2)^{0.5} \rfloor$$

The default choices for these and other cutting frequencies in the FMVol library reflect the rate-efficient conditions in the absence of noise, derived in the previously cited papers.

Similarly, the function

C_spot=FM_spot_cov(x1,x2,t1,t2,T,'N',N,'M',M,'tau',tau)

computes the spot covariance of a bivariate diffusion process via the Fourier–Malliavin estimator in (12). When not specified, the cutting frequencies are set to

 $\mathbb{N} = [\min(n_1, n_2)/2], \ \mathbb{M} = [(\min(n_1, n_2)/2)^{0.5}].$

If the estimation time vector tau is not specified, then it is set equal to 0:T/(2*M):T and given as an output variable tau_out.

Fig. 1 shows a comparison between discrete simulated trajectories of the instantaneous variance and covariance processes of the model in (23)–(26) and estimates thereof, obtained via the Fourier–Malliavin estimator. The latter has been implemented by using as input the vectors of simulated processes and by selecting default values for N, M and tau. For the simulation, we set T=1, n=23400 and select the following model parameters (see the Appendix⁶):

parameters=[0,0;0.4,0.4;2,2;1,1], Rho=[0.5,\tmin0.5,0,0,\tmin0.5,0.5],

x0=[log(100); log(100)], V0=[0.4; 0.4].

The comparison with the true variance/covariance trajectories shows that the estimation process is very efficient.

⁶ Note that we choose the vector of initial variances V0 to be equal to the mean reversion levels [0.4, 0.4]. When this is not the case, an additional source of finite-sample bias is introduced, see [33]. However, the discussion of this aspect is outside the scope of the paper.



Fig. 1. Instantaneous variance and covariance estimation for a bi-variate Heston process.

Fig. 2 is analogous to Fig. 1, except for the fact that this time we use vectors of noise-affected observations as inputs. Noise has been simulated according to Eq. (27), by adding the i.i.d variables $\eta^j(t_i) \sim N(0, \xi_j^2)$, where ξ_j is three times the standard deviation of the increments of the process x^j . Based on the finite-sample optimal results in the presence of noise by [21,23], the cutting frequencies have been set equal to N= 465 and M= 21.⁷ The resulting equally spaced estimation grid tau includes 43 nodes. The comparison with the true variance/covariance trajectories shows that the estimation process is still very efficient.

Specifically, Fig. 1 is obtained by writing the following lines in the MATLAB® command window:

```
T=1; n=23400; parameters=[0,0;0.4,0.4;2,2;1,1];
Rho=[0.5,-0.5,0,0,-0.5,0.5];
x0=[log(100); log(100)]; V0=[0.4; 0.4]; rng(123456);
[x,V,t]=Heston2D(T,n,parameters,Rho,x0,V0);
```

```
t1=t; t2=t; C_spot=FM_spot_cov(x(:,1),x(:,2),t1,t2,T);
M=(length(C_spot)-1)/2; tau=0:T/(2*M):T;
Cov=Rho(6)*sqrt(V(:,1)).*sqrt(V(:,2));
```

```
V_spot1=FM_spot_vol(x(:,1),t1,T);
V_spot2=FM_spot_vol(x(:,2),t2,T);
```

further, Fig. 2 is obtained by typing:

```
xi1=3; xi2=3; std_r=std(diff(x(:,1))); std_r_bis=std(diff(x(:,2)));
eta=randn(n+1,1); eta_bis= randn(n+1,1);
eta2=xi1*std_r*eta; eta2_bis=xi2*std_r_bis*eta_bis;
```

⁷ The optimal cutting frequencies in the presence of noise can be computed with the commands N=OptimalCuttingFrequency(x,t) and M=floor(sqrt(N)), where x and t are the vectors of observation values and times, respectively. This function is based on the paper by [21] and selects the optimal cut-off frequency N minimizing the finite sample MSE of the estimated volatility Fourier coefficients in the presence of noise. To compute the covariance of two processes, the minimum N of the two processes should be selected.



Fig. 2. Instantaneous variance and covariance estimation for a noisy bi-variate Heston process.

x1_noise= x(:,1)+eta2; x2_noise= x(:,2)+eta2_bis;

```
N1=OptimalCuttingFrequency(x1_noise,t1); N2=OptimalCuttingFrequency(x2_noise,t2);
N=min(N1,N2); M=floor(sqrt(N)); tau=0:T/(2*M):T;
C_spot=FM_spot_cov(x1_noise, x2_noise,t1,t2,T,'N',N,'M',M,'tau',tau);
```

V_spot1=FM_spot_vol(x1_noise,t1,T,'N',N,'M',M,'tau',tau); V_spot2=FM_spot_vol(x2_noise,t2,T,'N',N,'M',M,'tau',tau);

The supplementary function

[spotvar,tau]=FE_spot_vol_FFT(x,t)

implements the instantaneous variance estimator based on the *Fast Fourier Transform* algorithm (FFT) for equally-spaced data. The advantage of using FFT is mainly computational. In fact, the FFT reduces the complexity of computing the discrete Fourier transform from $O(n^2)$, which arises if one simply uses the basic algorithm for the computation of the Fourier coefficients of the process increments, to $O(n \log n)$, where *n* is the data size. This function makes use of the MATLAB[®] built-in routines fft(x) and ifft(x), which implement discrete Fourier and inverse transforms using the algorithm by [7]. It is worth noting that the best performance of the FFT algorithm is obtained by choosing *n*, *N* and *M* to be a power of two. However, the function works even for values of *n*, *N* and *M* different from powers of two. The relative performance improvement is considerable: for instance, when the dataset dimension is n = 23400, the execution CPU time passes from approximately 20 s to approximately 4 s, with a speedup factor of 5. For further details about the FFT implementation of the Fourier–Malliavin volatility estimator, we refer the reader to [20].⁸

Table 2 displays the estimated integrated variances and covariance, compared with their actual values. The estimated values are obtained by calling

V_int1=FM_int_vol(x1,t1,T)
V_int2=FM_int_vol(x2,t2,T)
C_int=FM_int_cov(x1,x2,t1,t2,T)

⁸ We also mention the paper by [5] which extends the use of FFT to non-uniform data and provides the related Julia implementation codes.

Table 2

Estimates of integrated variances and covarian	ice and	l actual	values.
------------------------------------------------	---------	----------	---------

	Int. Var. process 1	Int. Var. process 2	Int. Covar.
Actual values	0.1700	0.1568	0.0805
Estimates	0.1723	0.1594	0.0821
Estimates (Noisy data)	0.1760	0.1741	0.0819

with default values of the cutting frequency N.⁹ The last row of the table displays the estimated values in the noisy case, obtained by setting the optimized frequencies N= 465. We note that estimates are still very good in the presence of noise, except in the case of process 2, where the estimate is slightly less efficient for N= 465.

An important issue in the applications is ensuring that the estimated covariance matrix remains positive semi-definite. [22] proved that adding a Fejer kernel in the Fourier estimator of the integrated variances and co-variances guarantees positive semi-definiteness. The function FM_cov_matrix.m computes the integrated variance-covariance matrix on any number of diffusion processes, sampled over the same time horizon [0, T], using the Fourier-Malliavin estimator (13) with Fejer kernel. Input data can be of different sizes and non-synchronous, with and without missing data. This represents an important strength of the estimator, which does not require any preliminary synchronization procedure of the data. However, we highlight that when observed data are asynchronous special care must be taken in the choice of the cutting frequency N, to contrast the so-called *Epps effect*, i.e., the bias towards zero appearing in correlations among processes as the sampling frequency increases. The discussion of this aspect is outside the scope of the paper and we refer the reader to [23].¹⁰

4.3. Leverage, volatility of volatility and quarticity estimation

The estimation of the instantaneous leverage process B(t) by the Fourier estimator in (19) can be performed by calling the function

L_spot=FM_spot_lev(x,t,T,'N',N,'M',M,'L',L,'tau',tau)

The user must provide as inputs a vector of observations x and the related observation times t, the time horizon T, the cutting frequencies N, M and L, and the vector tau of estimation times. If tau is not specified, it is set equal to 0:T/(2*L):T and given as an output variable tau_out. When not specified, default values for the cutting frequencies are

 $\mathbb{N} = \lfloor n/2 \rfloor, \ \mathbb{M} = \lfloor (n/2)^{0.5} \rfloor, \ \mathbb{L} = \lfloor (n/2)^{0.25} \rfloor.$

Similarly, the estimation of the integrated leverage is performed by calling¹¹

L_int= FM_int_lev(x,t,T,'N',N,'M',M).

Further, the estimation of the instantaneous volatility of volatility process C(t) by the Fourier estimator in (21) can be performed by calling the function

VV_spot=FM_spot_volvol(x,t,T,'N',N,'M',M,'L',L,'tau',tau).

As for the inputs, the same considerations apply as for the case of the leverage, with the exception that in the case of the volatility of volatility the default values of the cutting frequencies are^{12}

 $\mathbb{N} = \lfloor n/2 \rfloor, \ \mathbb{M} = \lfloor (n/2)^{0.4} \rfloor, \ \mathbb{L} = \lfloor (n/2)^{0.2} \rfloor.$

Similarly, the estimation of the integrated volatility of volatility is performed by calling

VV_int=FM_int_volvol(x,t,T,'N',N 'M', M).

Finally, the estimation of the instantaneous quarticity process $\sigma^4(t)$ by the Fourier estimator can be performed by calling the function

Q_spot=FM_spot_quart(x,t,T,'N',N,'M',M,'L',L,'tau',tau)

As for the inputs, the same considerations apply as for the case of the leverage. In particular, the default values of the cutting frequencies are the same as for the leverage. Similarly, the estimation of the integrated quarticity is obtained by calling

Q_int=FM_int_quart(x,t,T,'N',N,'M',M).

⁹ Note that the code for the integrated covariance is based on Eq. (13).

¹⁰ We also remark that, by construction, the Fourier–Malliavin estimator of the instantaneous volatility matrix (12) is not even symmetric. [2] propose a modified estimator that can be proved to be symmetric and positive semi-definite.

¹¹ The conditions $N = \lfloor n/2 \rfloor$ and $M = O(\sqrt{n})$ are shown to be rate-optimal for the estimation of the integrated leverage by [25].

¹² As shown in [32], the default choice of $M = O((n/2)^{0.4})$ guarantees asymptotic unbiasedness for the estimation of the Fourier coefficients of the volatility of volatility in the absence of noise but is not rate-optimal; to obtain a rate-optimal estimator one has to subtract a multiple of the estimated quarticity.



Fig. 3. Intraday BID prices of MSFT on April 15, 2024, from 9.30 (market opening) to 16.00 (market closure).

Table 3 Estimates of integra	ated quantities for MSFT on	n December 20, 2023.	
Volatility	Quarticity	Vol. of vol.	Leverage
$1.378 \cdot 10^{-4}$	$2.718 \cdot 10^{-8}$	$1.464 \cdot 10^{-5}$	$-2.966 \cdot 10^{-6}$

5. Empirical applications

In this Section, we provide two examples of applications of the FMVol library with empirical data. The first example, which employs high-frequency prices of equity stocks, represents the typical application of the Fourier–Malliavin method. The second example, which involves intraday atmospheric temperature measurements, demonstrates the flexibility of the Fourier–Malliavin method for applications outside the field of financial econometrics.

5.1. Application to financial data

The first application of the FMVol library that we propose employs the series of tick-by-tick BID prices of Microsoft (MSFT) sampled on April 15, 2024. This series has been downloaded from the database of Dukascopy Bank SA.¹³ Fig. 3 shows the plot of the series.

The sample size of the series is n = 50162 and observations are irregularly spaced. Further, observations are affected by measurement error, due to the presence of microstructure noise. In this setting, the consistent estimation of the Fourier coefficients of the volatility requires suitably reducing the cutting frequency N, that is, excluding the highest frequencies from the convolution formula in (11). Accordingly, we select the cutting frequencies by using the finite-sample optimal procedure used in the simulated exercise of Section 4.2, that is, letting x and t denote, respectively, the vector of observation values and times, we select N=OptimalCuttingFrequency(x,t), M=floor(sqrt(N)) and L=floor(sqrt(M)). Moreover, note that we measure time in days, that is, we set T= 1, and estimate spot quantities every half an hour, that is, we select tau=0:T/13:T.

The results of the application of the FMVol functions for the estimation of spot quantities are displayed in Fig. 4, while Table 3 resumes the results of the application of the functions for estimating integrated quantities. Fig. 4 shows that MSFT spot volatility estimates follow the typical U-shape pattern observed in intraday asset volatility series – due to a more intense trading activity near market open and close time – with the exception of a peak at 13.30. A similar pattern is mechanically reflected in the intraday behaviour of spot quarticity estimates. Moreover, we note that the estimated spot volatility of volatility of MSFT also follows a U-shaped intraday pattern. Finally, we observe that the spot leverage of MSFT is prevalently negative, consistently with the negative correlation between asset returns and their volatilities typically observed on financial markets.

For the specific case of spot volatility, it is possible to evaluate the estimation accuracy by means of the empirical test presented in [19]. The test is as follows. For a given h > 0, define the standardized log-return at t as

$$z_h(t) := \frac{x(t+h) - x(t)}{\sqrt{\sigma^2(t) h}}.$$

¹³ www.dukascopy.com/swiss/english/marketwatch/historical



Fig. 4. Estimated spot volatility, quarticity, leverage and volatility of volatility for MSFT on April 15, 2024.

Under the model in Eq. (6), if *h* is sufficiently small, the sequence $\{z_h(t)\}_{t \in \tau}$, $\tau = \{0, h, 2h, ..., (\lfloor T/h \rfloor - 1)h\}$, approximates a sequence of i.i.d. standard normal random variables.

Thus, letting

$$\hat{z}_h(t) := \frac{x(t+h) - x(t)}{\sqrt{\hat{\sigma}^2(t)h}}$$

denote the reconstructed standardized increments of *x*, by testing the distribution of the sample sequence $\{\hat{z}_h(t)\}_{t \in \tau}$ for normality, one can indirectly assess the accuracy of the spot volatility estimates $\{\hat{\sigma}^2(t)\}_{t \in \tau}$.

Following [3], we set h = 5 minutes and test for the standard normality of the log-price increments of MSFT standardized by the Fourier spot volatility estimates displayed in Fig. 4.¹⁴ Specifically, we use the Kolmogorov–Smirnov test. As a result, we obtain a *p*-value equal to 0.513, which suggests accepting the null hypothesis of standard normality. This finding supports the accuracy of the spot volatility estimates.

We also apply the FMVol library to estimate the covariance of MSFT BID prices with those of Google (GOOGL) on April 15, 2024.¹⁵ The sample size of GOOGL BID prices on this day results in 75773 observations. These observations are also irregularly spaced and, moreover, non-synchronous with those of MSFT.

For the estimation of the covariance between the two stocks we make the following choices of the cutting frequencies. Firstly, we compute the finite-sample optimal values of N for each stock by means of N1=OptimalCuttingFrequency(x1,t1) and N2=OptimalCuttingFrequency(x2,t2), where x1 and x2 are the vectors of log-price values of MSFT and GOOGL, while t1 and t2 are the vectors of the corresponding observation times. Then we select N for the covariance as N=min(N1,N2). Finally, we choose M=floor(sqrt(N)).

As for T and tau, we maintain the same choice as in the case of the univariate application of the FMVol library with MSFT BID prices. The estimated spot covariance path is shown in Fig. 5. The value of the estimated daily integrated covariance is $8.248 \cdot 10^{-5}$. We note that the spot covariance between the two assets is always positive on the day under study. The positive sign of the spot and integrated covariance can be explained by the fact that both stocks belong to the same industrial sector. Furthermore, we note that the spot covariance is higher near market open and close times.

5.2. Application to weather data

The second application of the FMVol library that we illustrate considers a time series of atmospheric temperature measurements. Specifically, it involves sample intraday atmospheric temperature measurements recorded by a weather station near Sydney, Australia, with geographic coordinates given by -34.0, 151.1. This time series, which we downloaded from the Copernicus database,¹⁶ is sampled at the hourly frequency and covers the period from January 1, 2022, to December 31, 2022, for a total of 8760 hourly observations. Fig. 6 plots the measurement series, expressed in Celsius degrees.

¹⁴ As price observations are irregularly spaced, we perform downsampling at the 5-minute frequency by considering the observations whose timestamp is the closest to the points composing the 5-minute grid. Further, note that the last volatility estimate obtained at 16.00, that is, at market close, is excluded from the test because it is not possible to compute the forward-looking 5-minute log-price increment.

¹⁵ GOOGL BID prices have also been downloaded from the Dukascopy database.

¹⁶ www.copernicus.eu



Fig. 6. Hourly atmospheric temperature measurements corresponding to the geographic coordinates -34.0, 151.1 for the period January 1, 2022–December 31, 2022.

As detailed in [26], climate time series can be assumed to be a stochastic process composed of a trend (drift) and an i.i.d. noise (diffusion), plus possible outliers (jumps). We note that the effect of the drift on the convolution in is negligible (see lemma 2.2 by [17]) and thus, at high frequencies, the presence of a drift component does not impact the finite-sample efficiency of the Fourier–Malliavin method in reconstructing the coefficients of the volatility. However, the time series employed in this example are sampled at the (not so-high) hourly frequency. Hence, we prefer to take a conservative approach and de-trend the series, that is, isolate the diffusion component, in order to avoid a possible finite-sample bias in the estimates of the Fourier coefficients of the volatility. Note that weather time series can incorporate multiple seasonal trends, with different periods. In fact, in our specific instance, the presence of at least two trends can be recognized: a short-span trend related to the repetition of the intraday pattern of temperatures and a long-span trend related to the alternation of summer, autumn, winter and spring throughout the year.

Specifically, we apply the following procedure to remove trends. Firstly, to eliminate the presence of seasonal trends with multiweek periods, we split the 1-year atmospheric temperature sample into 73 sub-samples of the same size, corresponding to 5 days. Note that this splitting maintains a sufficiently large number of observations in each sub-sample for the asymptotic theory to be effective, namely 120. Further, we use the MATLAB[®] function *trenddecomp* to find and remove short-span intraday trends from sub-samples. As a result, we obtain a time series of detrended residuals, which represent the diffusion component of the atmospheric temperature series. We use such de-trended residuals as input to the FMVol functions.

For the application of the FMVol functions, we set T = 5, n = 120 and tau = 0 : 1/6 : T-1/6, and use the default values of the cutting frequencies. Note that the choice of tau is such that estimates are obtained every 4 hours. The resulting sequence of



Fig. 7. Estimated atmospheric temperature spot volatility, quarticity, leverage and volatility of volatility for the geographic coordinates -34.0, 151.1 over 2022.

 Table 4

 Estimates of integrated quantities for the atmospheric temperature series corresponding to the geographic coordinates -34.0, 151.1 during 2022.

Volatility	Quarticity	Vol. of vol.	Leverage
1060.552	11 150.248	18816.550	-89.279



Fig. 8. Estimated spot covariance of the atmospheric temperature recorded at coordinates -34.0, 151.1 and -33.7, 150.8 during 2022.

spot estimates is displayed in Fig. 7, while Table 4 contains the aggregate values of integrated quantities, obtained as the sum of sub-sample estimates. Fig. 7 shows that the spot volatility, quarticity and volatility of volatility show clusters, displaying larger spikes between February and May and smaller spikes near the end of the year. As for the spot leverage,¹⁷ Fig. 7 shows a value close to zero on average, with significant negative and positive spikes occurring approximately in correspondence with spot volatility spikes.

We also estimate the spot and integrated covariance with de-trended residuals of the temperature time series recorded from the weather station with coordinates -33.7, 150.8, which has also been downloaded from the Copernicus database. Since the two series are equally spaced and synchronous, for the estimation of the integrated and spot covariance we maintain the same values of T, N, M, L and tau as in the univariate case. The resulting spot estimates are displayed in Fig. 8. The aggregate integrated covariance

¹⁷ We refer to the quadratic covariation between atmospheric temperature and its volatility as *leverage* to underline that it has been estimated via the homonymous function of the FMVol library. However, we are aware that this is an abuse of terminology when input time series are not related to asset prices.

amounts to 536.253. We note that the spot covariance tends to be most often positive, consistently with the fact that the two locations are quite close in terms of geographic coordinates, with occasional negative spikes.

6. Conclusions

This paper presents the Fourier–Malliavin Volatility (FMVol) estimation library for MATLAB[®], which includes functions that implement the Fourier–Malliavin estimators of the spot and integrated volatility, co-volatility, quarticity, volatility of volatility and leverage. By introducing a ready-to-use library, this paper aims to facilitate the implementation of Fourier–Malliavin estimators even for users who are not familiar with Fourier analysis, in view of the flexibility of the method and the related wide scope of potential applications. Although originally introduced for the estimation of asset volatilities, the Fourier–Malliavin method is indeed a general non-parametric method that can be applied whenever one is interested in reconstructing the latent volatility (and/or second-order quantities) of a stochastic volatility process from discrete observations thereof. We remark that the method, which relies on infill asymptotics, does not impose any stationarity and ergodicity assumptions. Furthermore, we stress that the method does not require equally-spaced observations and is robust to measurement errors, or noise, without any preliminary bias correction or data pre-treatment. Moreover, in its multivariate version, it is intrinsically robust to irregular and asynchronous sampling.

Funding

S. Sanfelici is a member of the INdAM-GNCS research group, Italy. G. Toscano gratefully acknowledges financial support from the Institut Louis Bachelier 2022 grant 'Risk management in times of unprecedented geo-political volatility: a machine learning approach'.

CRediT authorship contribution statement

Simona Sanfelici: Conceptualization, Formal analysis, Investigation, Project administration, Software, Supervision, Writing – original draft, Writing – review & editing, Methodology, Resources, Validation, Visualization. Giacomo Toscano: Conceptualization, Data curation, Formal analysis, Funding acquisition, Investigation, Methodology, Resources, Software, Supervision, Validation, Visualization, Writing – original draft, Writing – review & editing.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

The authors are thankful to Marco Riani and Aldo Corbellini for their support in developing the FMVol library.

Appendix. Bivariate data simulation

This Appendix describes the functions Heston2D.m and Heston1D.m that we used to simulate data samples in Section 4. See [13] for more details on the parametric model employed.

The function Heston2D.m implements the Euler scheme to integrate the stochastic differential Eq. (23)–(26). The function reads as

[x,V] = Heston2D(T,n,parameters,Rho,x0,V0),

where:

- T (e.g., T=1) denotes the length of the integration interval [0, T];
- n (e.g., n=10000) is the number of sub-intervals used for the discretization of the equations;
- parameters is 4×2 matrix whose *j*th column contains the values for the parameters μ_j , α_j , θ_j and γ_j , j = 1, 2, in (23)–(26);
- Rho is a column vector of length equal to 6 that contains the correlation coefficients of the Brownian motions, namely $\rho_{1,2}$, $\rho_{1,3}$, $\rho_{1,4}$, $\rho_{2,3}$, $\rho_{2,4}$, $\rho_{3,4}$;
- x0 and V0 are vectors that contain the initial values for x^1, x^2 and their related variances, respectively.

The function provides as outputs the $(n + 1) \times 2$ matrices x and V containing the n + 1 observations for x^1 and x^2 and their related instantaneous variance observations, respectively.

The function

[x,V]=Heston1D(T,n,parameters,rho,x0,V0)

similarly generates a univariate diffusion process and its instantaneous variance from the Heston model.

We refer the reader to the related MATLAB $^{(\! 8\!)}$ documentation and help files available in the FSDA Toolbox for more details.

References

- [1] Y. Aït-Sahalia, J. Jacod, High-frequency financial econometrics, Princeton University Press, 2014.
- [2] J. Akahori, N.L. Liu, Mancino, T. M. E, Mariotti, Y. Yasuda, The Fourier estimation method with positive semi-definite estimators, Working Paper, 2023.
- [3] T.G. Andersen, T. Bollerslev, F.X. Diebold, P. Labys, Modeling and forecasting realized volatility, Econometrica 45 (1) (2001) 137–185.
- [4] P. Bloomfield, Fourier analysis of time series. An introduction., Wiley Series in Probability and Statistics, 2000.
- [5] P. Chang, E. Pienaar, T. Gebbie, Malliavin-Mancino estimators implemented with non-uniform fast Fourier transforms, SIAM J. Sci. Comput. 42 (6) (2020).
- [6] A.A. Christie, The stochastic behavior of common stock variances: value, leverage and interest rate effects, J. Financ. Econ. 10 (4) (1982) 407-432.
- [7] J.W. Cooley, J.W. Tukey, An algorithm for the machine computation of the complex Fourier series, Math. Comp. 19 (1965) 297-301.
- [8] I.V. Curato, Estimation of the stochastic leverage effect using the Fourier transform method, Stochastic Process. Appl. 129 (9) (2019) 3207–3238.
 [9] I.V. Curato, S. Sanfelici, Measuring the leverage effect in a high-frequency framework, in: G. Gregoriou (Ed.), The handbook of high-frequency trading, Elsevier, 2015, pp. 425–446.
- [10] I.V. Curato, S. Sanfelici, Stochastic leverage effect in high-frequency data: a Fourier based analysis, Econometr, Stat. 23 (2022) 53-82.
- [11] E.L. Grogan, J.A. Morris, P.R. Norris, D.J. France, A. Ozdas, R. Stiles, P. Harris, B. Dawant, T. Speroff, Reduced heart rate volatility: an early predictor of death in trauma patients, Ann. Surg. 240 (3) (2004) 547–556.
- [12] E.J. Hannan, Multiple time series, John Wiley and Sons, 1970.
- [13] S. Heston, A closed-form solution for options with stochastic volatility with applications to bond and currency options, Rev. Financ. Stud. 6 (1993) 327-343.
- [14] G. Livieri, M.E. Mancino, S. Marmi, Asymptotic results for the Fourier estimator of the integrated quarticity, Decis. Econ. Finance 42 (2) (2019) 471–502.
 [15] P. Malliavin, Integration and probability, Springer-Verlag, 1995.
- [16] P. Malliavin, M.E. Mancino, Fourier series method for measurement of multivariate volatilities, Finance Stoch. 4 (2002) 49-61.
- [17] P. Malliavin, M.E. Mancino, A Fourier transform method for nonparametric estimation of multivariate volatility, Ann. Statist. 37 (4) (2009) 1983–2010.
- [18] M.E. Mancino, T. Mariotti, G. Toscano, Asymptotic normality and finite-sample robustness of the Fourier spot volatility estimator in the presence of microstructure noise, Working Paper, 2022.
- [19] M.E. Mancino, M.C. Recchioni, Fourier spot volatility estimator: asymptotic normality and efficiency with liquid and illiquid high-frequency data, PLoS One 10 (9) (2015) e0139041.
- [20] M.E. Mancino, M.C. Recchioni, S. Sanfelici, Fourier-Malliavin volatility estimation: theory and practice, Springer International Publishing, 2017.
- [21] M.E. Mancino, S. Sanfelici, Robustness of Fourier estimator of integrated volatility in the presence of microstructure noise, Comput. Statist. Data Anal. 52 (6) (2008) 2966–2989.
- [22] M.E. Mancino, S. Sanfelici, Covariance estimation and dynamic asset-allocation under microstructure effects via Fourier methodology, in: G.N. Gregoriou, R. Pascalau (Eds.), Financial econometrics modeling: market microstructure, factor models and financial risk measures, Palgrave Macmillan UK, London, 2011, pp. 3–32.
- [23] M.E. Mancino, S. Sanfelici, Estimating covariance via Fourier method in the presence of asynchronous trading and microstructure noise, J. Financ. Econom. 9 (2) (2011) 367–408.
- [24] M.E. Mancino, S. Sanfelici, Estimation of quarticity with high frequency data, Quant. Finance 12 (4) (2012) 607-622.
- [25] M.E. Mancino, G. Toscano, Rate-efficient asymptotic normality of the Fourier estimator of the leverage process, Stat. Interface 15 (1) (2022) 73-89.
- [26] M. Mudelsee, Climate time series analysis: classical statistical and bootstrap methods, Springer International Publishing, 2014.
- [27] C. Mumenthaler, O. Renaud, R. Gava, T. Brosch, The impact of local temperature volatility on attention to climate change: evidence from Spanish tweets, Global Environ. Change 69 (2021) 102286.
- [28] G. Papantonopoulos, K. Takahashi, T. Bountis, B.G. Loos, Mathematical modeling suggests that periodontitis behaves as a non-linear chaotic dynamical process, J. Periodontol. 84 (10) (2013) e29-e39.
- [29] M.B. Priestley, Spectral analysis and time series, Academic Press, 1983.
- [30] S. Sanfelici, I.V. Curato, M.E. Mancino, High-frequency volatility of volatility estimation free from spot volatility estimates, Quant. Finance 15 (8) (2015) 1331–1345.
- [31] G. Toscano, The price-leverage covariation as a measure of the response of the leverage effect to price and volatility changes, Appl. Stoch. Models Bus. Ind. 38 (3) (2022) 497-511.
- [32] G. Toscano, G. Livieri, M.E. Mancino, S. Marmi, Volatility of volatility estimation: central limit theorems for the Fourier transform estimator and empirical study of the daily time series stylized facts, J. Financ. Econom. 22 (1) (2024) 252–296.
- [33] G. Toscano, M.C. Recchioni, Bias-optimal vol-of-vol estimation: the role of window overlapping, Decis. Econ. Finance 45 (1) (2021) 137-185.
- [34] J.S. Yap, J. Fan, R. Wu, Nonparametric modeling of longitudinal covariance structure in functional mapping of quantitative trait loci, Biometrics 65 (4) (2009) 1068–1077.