




## Article

# Symmetric Positive Semi-Definite Fourier Estimator of Spot Covariance Matrix with High Frequency Data

Jiro Akahori <sup>1</sup>, Reika Kambara <sup>1</sup>, Nien-Lin Liu <sup>2</sup>, Maria Elvira Mancino <sup>3,\*</sup>, Tommaso Mariotti <sup>4,\*</sup>  
and Yukie Yasuda <sup>1</sup>

<sup>1</sup> Department of Mathematical Sciences, Ritsumeikan University, Kusatsu 525-8577, Japan; akahori@se.ritsumei.ac.jp (J.A.); ra0082vs@ed.ritsumei.ac.jp (R.K.); 25v01388@gst.ritsumei.ac.jp (Y.Y.)

<sup>2</sup> Department of Business Economics, Tokyo University of Science, Tokyo 113-8654, Japan; nl-liu@rs.tus.ac.jp

<sup>3</sup> Department of Economics and Management, University of Florence, Via delle Pandette 9, 50127 Firenze, Italy

<sup>4</sup> Department of Economics, Social Studies, Applied Mathematics and Statistics, University of Turin, Corso Unione Sovietica 218/bis, 10134 Torino, Italy

\* Correspondence: mariaelvira.mancino@unifi.it (M.E.M.); tommaso.mariotti@unito.it (T.M.)

## Abstract

This paper proposes a nonparametric estimator of the spot volatility matrix with high-frequency data. Our newly proposed Positive Definite Fourier (PDF) estimator produces symmetric positive semi-definite estimates and is consistent with a suitable choice of the localizing kernel. The PDF estimator is based on a modification of the Fourier estimation method introduced by Malliavin and Mancino. The estimator has two parameters: the frequency  $N$ , which controls the biases due to the asynchronicity effect and the market microstructure noise effect; and the localization parameter  $M$  for the employed Gaussian kernel. The sensitivity of the PDF estimator to the choice of these two parameters is studied in a simulated environment. The accuracy and the ability of the estimator to produce positive semi-definite covariance matrices are evaluated by an extensive numerical analysis, against competing estimators present in the literature. The results of the simulations are confirmed under different scenarios, including the dimensionality of the problem, the asynchronicity of data, and several different specifications of the market microstructure noise. The computational time required by the estimator and the stability of estimation are also tested with empirical data.

**Keywords:** nonparametric covariance estimator; risk management; factor analysis; Fourier analysis; positive semidefiniteness



Academic Editors: Steven Haberman and Anatoliy Swishchuk

Received: 30 April 2025

Revised: 11 July 2025

Accepted: 17 September 2025

Published: 9 October 2025

**Citation:** Akahori, Jiro, Reika Kambara, Nien-Lin Liu, Maria Elvira Mancino, Tommaso Mariotti, and Yukie Yasuda. 2025. Symmetric Positive Semi-Definite Fourier Estimator of Spot Covariance Matrix with High Frequency Data. *Risks* 13: 197. <https://doi.org/10.3390/risks13100197>

**Copyright:** © 2025 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

## 1. Introduction

Empirical studies have pointed out the importance of evaluating distinct time variations in correlations between asset prices. In particular, several recent studies address the efficient estimation of covariances using high-frequency data asynchronously sampled across different assets. While the literature about the estimation of integrated covariances has been steadily growing, it remains still sparse for the spot covariances estimation. An early proposal to cope with spot covariances estimation using asynchronous high-frequency data was made in Malliavin and Mancino (2009). Differently from other estimators that pre-process data to make them synchronous (e.g., linear interpolation, piecewise-constant previous-tick interpolation or the refresh-time procedure proposed by Barndorff-Nielsen et al. (2011)), the Fourier estimator is based on an integration procedure that uses all the

available data. Using all data, with no preliminary manipulation (such as pre-averaging, see, e.g., [Aït-Sahalia and Jacod \(2014\)](#)), allows for directly feeding unevenly sampled returns, or even asynchronous data in the multivariate case. A significant property for an estimator of integrated or spot covariances requires the estimated covariance matrix to be positive semi-definite. This has important consequences in several contexts, such as the recently developed field of principal component analysis with high-frequency data ([Liu and Ngo \(2017\)](#), [Aït-Sahalia and Xiu \(2019\)](#), [Chen et al. \(2020\)](#)) or the asset allocation framework (see, e.g., [Engle and Colacito \(2006\)](#)). This property has sometimes been studied for integrated covariances estimators (see, e.g., [Barndorff-Nielsen and Shephard \(2004\)](#), [Barndorff-Nielsen et al. \(2011\)](#), [Mancino et al. \(2017\)](#), [Park et al. \(2016\)](#), [Shepard and Xiu \(2017\)](#), [Cui et al. \(2019\)](#)). However, to the best of our knowledge, the spot covariances estimator proposed in the present paper is the first to guarantee positive semi-definiteness for the estimated matrix. For example, when dealing with spot volatility, [Chen et al. \(2020\)](#) integrate the estimations before computing the eigenvalues of the covariance matrix; instead, [Bu et al. \(2023\)](#) impose positive semi-definiteness by applying suitable shrinkage techniques that effectively impose manipulation of the estimated matrix.

This work aims to propose a novel spot covariance estimator, to prove its positivity and consistency, and to analyze its finite-sample properties in a simulated environment. Our starting point is the spot Fourier estimator by [Malliavin and Mancino \(2009\)](#) (see also [Mancino et al. \(2017\)](#) for several results and applications of the estimation method). The spot Fourier estimator is defined in three steps. First, the discrete Fourier coefficients of the asset returns are computed using all the available observations. Second, it is proven that their convolution product leads to a consistent estimator of the Fourier coefficients of the volatility function. Finally, the Fourier-Fejér inversion formula provides the final spot volatility estimator. Due to a lack of symmetry in the Fejér kernel, however, the spot Fourier estimator may fail to yield positive semi-definite estimations when the asset prices are observed on asynchronous grids. To overcome this difficulty and ensure symmetric and positive semi-definite estimates, this paper introduces a modified version of the Fourier estimator, which we call the PDF estimator. Theorem 1 proves that the modified estimator fulfills the desired property. Theorem 2 gives bounds for its asymptotic error, providing sufficient conditions for the consistency of the estimator based on the rates of  $N$  and  $M$  with respect to the sampling frequency.

The new estimator relies on two parameters: the cutting frequency  $N$ , and the localizing frequency  $M$ . We address the question of how to optimally choose them to minimize the error, according to the asymptotic conditions in Theorem 2, via a simulation study. By setting  $N = c_N \rho_n^{-\alpha}$  and  $M = c_M \rho_n^{-\beta}$ , where  $\rho_n$  is the mesh of the given sampling, and  $\alpha, \beta$  are suggested by Theorem 2, a grid of possible values of the constants  $c_N$  and  $c_M$  is tested against several different model specifications for both the efficient price process and the additive microstructure component. Concerning the parameter  $c_M$ , which controls the localization Gaussian kernel, we find that it exhibits a stable optimal value across the scenarios considered, necessitating only a small downward correction in the presence of noise. A similar behavior was observed for the original Fourier estimator in [Mancino et al. \(2025\)](#). Moreover, for the four models of the efficient price, we find that the difference between close values of  $c_N$  and  $c_M$  is relatively small: therefore, making a slightly sub-optimal choice causes no significant increase in the error. Finally, we evaluate the finite-sample performance of the proposed PDF estimator by comparing its accuracy and the percentage of its positive semi-definite estimates against those obtained using the smoothed two-scale estimator by [Mykland et al. \(2019\)](#) and the local method of moments estimator by [Bibinger et al. \(2019\)](#), which are both able to manage asynchronous observations. We run the comparison focusing on the main problems that affect the estimation of variance-covariance

matrices from high-frequency data. First of all, we address the dimensionality problem, by evaluating the estimates computed when the number of assets increases. Secondly, we focus on the level of asynchronicity, considering different intensities of the Poisson processes that drive the observation frequency. Lastly, we analyze market microstructure noise, considering noise coming from rounding, i.i.d. noise, auto-correlated noise, noise correlated with the efficient price process, and heteroskedastic noise. The exercise shows that only the PDF estimator consistently produces positive semi-definite estimations in 100% of the cases (as guaranteed by the theory), while maintaining a hedge with respect to its competitors in terms of mean square error.

The robustness of all the simulation results is confirmed by modifying the simulation model behind the analysis; in particular, we consider: a Heston Stochastic Volatility model (Heston (1993)), a One Factor Volatility model and a Two Factor Volatility model (Chernov et al. (2003)), and a Rough Heston model (El Euch and Rosenbaum (2019)), and in each case we get comparable results.

Finally, in an empirical exercise that employs trade data from twelve US firms collected from January 2023 to December 2023, we verify the stability of the estimations to changes in the cutting frequency  $N$  and in the location parameter  $M$ , looking at the same time at the computational time required in such real-world cases, which turns out to be not sensible to the choice of  $M$ .

The remainder of this work is organized as follows. Section 2 introduces the positive semi-definite (PDF) Fourier estimator of spot covariance, and its positivity is proved. Section 3 studies the asymptotic error of the PDF estimator with Gaussian kernel, proving its consistency and delivering its rate of convergence both for irregular and regular sampling schemes. Section 4 presents the simulation study, including a sensitivity analysis on the parameters of the estimator in terms of integrated mean square error, and a comparison against alternative estimators present in the literature, in which accuracy and ability to produce positive semi-definite matrices are considered. Section 5 presents the study with empirical data, while Section 6 concludes.

## 2. The Positive Semi-Definite Spot Covariance Estimator

Assume that the asset price is described by a  $d$ -dimensional Itô semimartingale  $X = (X^1, \dots, X^d)$

$$X_t^j = x_0^j + \int_0^t b^j(s)ds + \sum_{k=1}^d \int_0^t \sigma_{jk}(s)dW_s^k, \quad j = 1, \dots, d$$

with  $W = (W^1, \dots, W^d)$  a  $d$ -dimensional Brownian motion on the filtered probability space  $(\Omega, (\mathcal{F}_t)_{t \in [0, T]}, P)$  and  $b_j$  and  $\sigma_{jk}$  are adapted continuous processes. The  $d \times d$  instantaneous (spot) covariance matrix  $V(t)$  has entries

$$V^{j,j'}(t) := \sum_{k=1}^d \sigma_{jk}(t)\sigma_{j'k}(t), \quad \text{for } j, j' = 1, \dots, d \quad \text{and } t \in [0, T].$$

For simplicity of notation, we assume  $T = 1$ , without loss of generality.

We assume that the prices are observed on discrete, irregular and asynchronous time grids

$$0 = t_0^j < t_1^j < \dots < t_{n_j}^j = 1 \quad \text{for } j = 1, \dots, d.$$

Let  $\rho_{n_j} = \max_{0 \leq h \leq n_j-1} |t_{h+1}^j - t_h^j|$  and  $\rho_n = \max_{j=1, \dots, d} \rho_{n_j}$ . In the following,  $\Delta(X_l^j)$  denotes the discrete return  $X_{t_l}^j - X_{t_{l-1}}^j$  for  $j = 1, \dots, d$  and  $l = 1, \dots, n_j$ .

We start introducing the classical Fourier estimation method, in order to better clarify the issue arising for the lack of symmetry and to understand the modification of this estimator we propose here.

For  $j = 1, \dots, d$ , define the discrete Fourier coefficients of the log-returns as

$$c_k(\Delta X_{n_j}^j) := \sum_{l=1}^{n_j} e^{-2\pi i k t_l^j} \Delta(X_l^j),$$

and, for any  $j, j'$ , define an unbiased estimator of the  $k$ -th Fourier coefficient of the covariance process, through the following convolution formula:<sup>1</sup>

$$c_k(V_{n,N}^{j,j'}) := \frac{1}{2N+1} \sum_{|h| \leq N} c_h(\Delta X_{n_j}^j) c_{k-h}(\Delta X_{n_{j'}}^{j'}). \quad (1)$$

Finally, for any  $t \in (0, 1)$ , the spot covariance estimator is defined as

$$\hat{V}_{n,N,M}^{j,j'}(t) := \sum_{|k| \leq M} \left(1 - \frac{|k|}{M+1}\right) e^{2\pi i k t} c_k(V_{n,N}^{j,j'}). \quad (2)$$

By means of the Dirichlet kernel  $D_N(x) := \frac{1}{2N+1} \sum_{|k| \leq N} e^{2\pi i k x}$ , it is possible to express (1) as

$$c_k(V_{n,N}^{j,j'}) = \frac{1}{2N+1} \sum_{l=1}^{n_j} \sum_{l'=1}^{n_{j'}} D_N(t_l^j - t_{l'}^{j'}) e^{-2\pi i k t_l^j} \Delta(X_l^j) \Delta(X_{l'}^{j'}).$$

Therefore, the Fourier spot volatility estimator (2) can be rewritten, using two kernels, as

$$V_{n,N,M}^{j,j'}(t) = \frac{1}{2N+1} \sum_{l=1}^{n_j} \sum_{l'=1}^{n_{j'}} F_M(t - t_{l'}^{j'}) D_N(t_l^j - t_{l'}^{j'}) \Delta(X_l^j) \Delta(X_{l'}^{j'}), \quad (3)$$

where  $F_M(x) := \sum_{|k| \leq M} \left(1 - \frac{|k|}{M+1}\right) e^{2\pi i k x}$  is the Féjér kernel.

The Fourier estimation method was introduced in [Malliavin and Mancino \(2009\)](#) to estimate instantaneous multivariate volatilities from high-frequency observations of diffusion processes in a non-parametric way and without any stationarity assumption. The authors aimed at overcoming some difficulties arising from the application of the quadratic variation formula in the commonly used realized covariation methods with financial data. Detailed results and discussion on both statistical properties as well as numerical implementation and empirical applications can be found in the book [Mancino et al. \(2017\)](#).

The asymptotic properties of the Fourier spot volatility estimator (2) have been studied in [Malliavin and Mancino \(2009\)](#) (in the absence of noise) and in [Mancino et al. \(2025\)](#) (in the presence of noise). However, while the positivity of the Fourier estimator of the integrated covariance matrix is shown and debated also in [Mancino et al. \(2017\)](#), the spot covariance estimator may fail in producing symmetric positive semi-definite estimations, being  $F_M(t - t_l^j) D_N(t_l^j - t_{l'}^{j'})$  not symmetric in  $j, j'$  and could lead to complex eigenvalues in  $\hat{V}_{\mathcal{K},S}(t)$ . In addition, simple symmetrizations such as  $(\hat{V}_{\mathcal{K},S}^{j,j'} + \hat{V}_{\mathcal{K},S}^{j',j})/2$  are not guaranteed to be positive-definite, possibly with negative eigenvalues.

In this paper, we propose the following estimator of spot covariance, which overcomes the issue of positive semidefiniteness of the Fourier estimator (3).

**Definition 1.** Let  $\mathcal{K}$  be a finite subset of  $\mathbb{Z}$ ,  $\mathcal{S} := \{S(k) \subset_{\text{finite}} \mathbb{Z}^2 : k \in \mathcal{K}, (s, s') \in S(k) \implies s + s' = k\}$ , and  $c$  be a complex function on  $\mathcal{K}$ ; we define the estimator for  $V_{j,j'}(t)$  as

$$\hat{V}_{\mathcal{K},\mathcal{S}}^{j,j'}(t) = \sum_{l=1}^{n_j} \sum_{l'=1}^{n_{j'}} \sum_{k \in \mathcal{K}} c(k) e^{2\pi i k t} \sum_{(s,s') \in \mathcal{S}} e^{-2\pi i s t_l^j} e^{2\pi i s' t_{l'}^{j'}} \Delta(X_l^j) \Delta(X_{l'}^{j'}). \quad (4)$$

**Remark 1.** If we take  $\mathcal{K} = \{0, \pm 1, \pm 2, \dots, \pm M\}$  for some positive integer  $M$  and  $\mathcal{S}(k) = \{(s, s') | s + s' = k, s = 0, \pm 1, \pm 2, \dots, \pm N\}$  for some positive integer  $N$ , and

$$c(k) = \left(1 - \frac{|k|}{M+1}\right) \frac{1}{2N+1},$$

we obtain

$$\hat{V}_{\mathcal{K},\mathcal{S}}^{j,j'}(t) = \sum_{k=-M}^M \left(1 - \frac{|k|}{M+1}\right) e^{2\pi i k t} \sum_{s=-N}^N \sum_{l=1}^{n_j} \sum_{l'=1}^{n_{j'}} e^{-2\pi i s t_l^j} e^{2\pi i (k-s) t_{l'}^{j'}} \Delta(X_l^j) \Delta(X_{l'}^{j'}). \quad (5)$$

The estimator (5) can be expressed as follows

$$\hat{V}_{\mathcal{K},\mathcal{S}}^{j,j'}(t) = \frac{1}{2N+1} \sum_{l=1}^{n_j} \sum_{l'=1}^{n_{j'}} F_M(t - t_l^j) D_N(t_l^j - t_{l'}^{j'}) \Delta(X_l^j) \Delta(X_{l'}^{j'}). \quad (6)$$

Therefore, with a suitable choice of function  $c(\cdot)$ , the estimator (4) coincides with the Fourier spot covariance estimator (3) introduced by [Malliavin and Mancino \(2009\)](#).

The main theoretical result of this work concerns the positive semi-definiteness of the proposed estimator (4) under suitable hypothesis and is stated in the following theorem.

**Theorem 1.** Let  $N$  and  $M$  be positive integers. Suppose that  $\mathcal{K} = \{0, \pm 1, \pm 2, \dots, \pm 2N\}$ ,  $c_M(k)$  is a positive semi-definite function on  $\mathcal{K}$  and

$$\mathcal{S}(k) = \begin{cases} \{(-N + k + v, N - v) : v = 0, \dots, 2N - k\} & 0 \leq k \leq 2N \\ \{(N + k - v, -N + v) : v = 0, \dots, 2N + k\} & -2N \leq k < 0. \end{cases}$$

Then,  $\hat{V}_{\mathcal{K},\mathcal{S}}(t)$  defined in (4) is symmetric and positive semi-definite.

The proof of Theorem 1 is reported in the Appendix A.

Moreover, it emerges that  $\hat{V}_{\mathcal{K},\mathcal{S}}(t)$  can be rewritten as

$$\hat{V}_{N,M}^{j,j'}(t) = \frac{1}{2N+1} \sum_{l=1}^{n_j} \sum_{l'=1}^{n_{j'}} \sum_{u=-N}^N \sum_{u'=-N}^N c_M(u - u') e^{2\pi i u(t - t_l^j)} e^{-2\pi i u'(t - t_{l'}^{j'})} \Delta(X_l^j) \Delta(X_{l'}^{j'}), \quad (7)$$

for two assets  $j$  and  $j'$  and  $t \in [0, 1]$ , where  $c_M(k)$  is still a positive semi-definite function.<sup>2</sup> We call the class of the estimators parameterized by the positive semi-definite function  $c_M$  the positive semi-definite Fourier (PDF) estimator.

By Bochner's theorem, we know that, for each positive semi-definite function  $c_M$ , there exists a bounded measure  $\mu_M$  on  $\mathbf{R}$  such that

$$c_M(x) = \int_{\mathbf{R}} e^{2\pi i y x} \mu_M(dy).$$

Therefore, we may also rewrite the PDF estimator (7) using the measure  $\mu_M$  instead of the positive semi-definite function  $c_M(k)$ , and obtain

$$\hat{V}_{N,M}^{j,j'}(t) = \frac{1}{2N+1} \sum_{l=1}^{n_j} \sum_{l'=1}^{n_{j'}} \int_{\mathbf{R}} D_N(t-t_l^j+y) D_N(t-t_{l'}^{j'}+y) \mu_M(dy) \Delta(X_l^j) \Delta(X_{l'}^{j'}). \quad (8)$$

Thus, we can also say that the PDF estimators are parameterized by a measure  $\mu_M$ .

In the next Section, we prove the consistency of the estimator (7) (equivalently, (8)).

### 3. Asymptotic Properties of the PDF Estimator with Gaussian Kernel

In this section, we consider the case where  $\mu_M$  is the Gaussian kernel, or more precisely,

$$\mu_M(dy) = \sqrt{\frac{M}{2\pi}} e^{-\frac{My^2}{2}} dy,$$

which is equivalent to

$$c_M(x) = e^{-\frac{2\pi^2 x^2}{M}}.$$

While the parameter  $N$  controls the microstructure noise effect, as it will appear in the intensive simulation study carried on in the next Section, the parameter  $M$  controls the localizing kernel and the estimation error. As we will see, it is needed  $N, M \rightarrow \infty$  as  $\rho_n \rightarrow 0$  with appropriate rates. We will call the estimator *Gaussian PDF estimator*, or GPDF for short.

In this section, we give an estimate of the error  $V^{j,j'} - \hat{V}_{N,M}^{j,j'}$  of the GPDF estimator under the following assumptions.

For simplicity, we consider  $d = 2$ . Moreover, it is not restrictive to assume that the drift  $b \equiv 0$  for the efficient price process<sup>3</sup>.

Further, assume that

(A) the volatility processes  $V^{j,j'}, j, j' = 1, 2$  satisfy

$$\|V\|_{\infty} := \max_{j,j'} (\mathbf{E}[\sup_{t \in [0, 2\pi]} |\sum_{j,j'} |V^{j,j'}(t)|^2])^{1/2} < \infty$$

and  $\sigma^j := (\sigma_1^j, \sigma_2^j), j = 1, 2$  are all twice Malliavin differentiable and

$$C_{\nabla} := \max_{j,j'} \mathbf{E}[\sup_{s,u,v \in [0, 2\pi]} |\sigma^{j'}(v) \nabla_v(\sigma^j(u) \nabla_s V^{j,j'}(u))|] < \infty,$$

where  $\nabla$  denotes the Malliavin derivative. Further, we assume that  $V^{j,j'}, j, j' = 1, 2$  are  $\kappa$ -Hölder continuous for some  $\kappa \in (0, 1)$  in the sense that

$$\sum_{k \in \mathbf{Z}} |k|^{2\kappa} \mathbf{E}[|(\mathcal{F}V^{j,j'})(k)|^2] =: C_{\kappa} < \infty, \quad (9)$$

where  $(\mathcal{F}V^{j,j'})(k)$  is the  $k$ -th Fourier coefficient of  $V^{j,j'}$ , i.e.,

$$(\mathcal{F}V^{j,j'})(k) = \int_0^1 V^{j,j'}(s) e^{-2\pi i k s} ds.$$

**Remark 2.** The Malliavin differentiability assumption for the volatility process is not restrictive, as it is satisfied by most relevant stochastic volatility models, e.g., Heston model, and also by rough volatility models, see Clement and Gloter (2011); Alòs and Ewald (2008); Inahama (2014). Thus, our result applies to all those models. As far as it concerns volatility jumps, they are not ruled out by the Malliavin differentiability but it is required here that the volatility process is Hölder-

continuous. In fact, if the volatility process  $V^{j,j'}(t)$  has a fixed time of discontinuity  $t$ , then the Fourier-Fejér inversion formula gives that (6) converges to  $(V^{j,j'}(t-) + V^{j,j'}(t))/2$  for a fixed time of discontinuity and  $V^{j,j'}(t)$  for any continuity time.

**Theorem 2.**

- (i) Under the assumption (A), for any  $j, j' = 1, 2$  the  $L^2$ -error between  $V^{j,j'}$  and the estimator  $\hat{V}_{N,M}^{j,j'}$  is estimated as

$$\begin{aligned} \mathbf{E} \left[ \int_0^1 (V^{j,j'}(t) - \hat{V}_{N,M}^{j,j'}(t))^2 dt \right] &\leq \pi^2 \|V\|_\infty^2 \rho_n^2 N^2 \sqrt{\frac{M}{2\pi}} \\ &+ (4C_\nabla + 2\|V\|_\infty^2) \left( 4\pi^2 \rho_n^2 N^2 + (2N+1)^{-1} \right) \sqrt{\frac{M}{2\pi}} \\ &+ 2C_\kappa \left( (2N)^{-2\kappa} + \left( \frac{2\pi^2}{M} \right)^\kappa \right). \end{aligned} \quad (10)$$

- (ii) In the case of synchronous and regular sampling, when  $t_k^j = k/n$  for  $k = 0, 1, \dots, n$ ,  $j = 1, 2$ , Equation (10) is improved as

$$\begin{aligned} \mathbf{E} \left[ \int_0^1 (V^{j,j'}(t) - \hat{V}_{N,M}^{j,j'}(t))^2 dt \right] &\leq \pi^2 \|V\|_\infty^2 \rho_n^2 \sqrt{\frac{M}{2\pi}} \left( \frac{M}{4\pi^2} + 1 \right) \\ &+ (4C_\nabla + 2\|V\|_\infty^2) (2N+1)^{-1} \sqrt{\frac{M}{2\pi}} \\ &+ 2C_\kappa \left( (2N)^{-2\kappa} + \left( \frac{2\pi^2}{M} \right)^\kappa \right). \end{aligned} \quad (11)$$

- (iii) Consequently, for the general sampling scheme, if<sup>4</sup>  $N \asymp \rho_n^{-\alpha}$  and  $M \asymp \rho_n^{-\beta}$ , the consistency is attained if

$$0 < \beta < \frac{4}{3}, \quad \frac{\beta}{2} < \alpha < -\frac{1}{4}\beta + 1 \quad (12)$$

and

$$\mathbf{E} \left[ \int_0^1 (V^{j,j'}(t) - \hat{V}_{N,M}^{j,j'}(t))^2 dt \right]^{1/2} = O \left( \rho_n^{\min(1-\alpha-\frac{\beta}{4}, \frac{\alpha}{2}-\frac{\beta}{4}, \frac{\kappa\beta}{2})} \right).$$

Further, the best rate is given as

$$\max_{0 < \beta < \frac{4}{3}, \frac{\beta}{2} < \alpha < -\frac{1}{4}\beta + 1} \min \left( 1 - \alpha - \frac{\beta}{4}, \frac{\alpha}{2} - \frac{\beta}{4}, \frac{\kappa\beta}{2} \right) = \frac{2\kappa}{6\kappa + 3},$$

where the maximum is attained when  $\alpha = 2/3$  and  $\beta = 4/(6\kappa + 3)$ .

- (iv) In the case of synchronous and regular sampling, when  $t_k^j = k/n$  for  $k = 0, 1, \dots, n$ ,  $j = 1, 2$ , the consistency is attained if

$$\alpha > \frac{\beta}{2}, \quad 0 < \beta < \frac{4}{3} \quad (13)$$

and

$$\mathbf{E} \left[ \int_0^1 (V^{j,j'}(t) - \hat{V}_{N,M}^{j,j'}(t))^2 dt \right]^{1/2} = O \left( n^{-\min(\frac{\alpha}{2}-\frac{\beta}{4}, 1-\frac{3}{4}\beta, \frac{\kappa\beta}{2})} \right).$$



The best rate is given as

$$\max_{\alpha > \frac{\beta}{2}, 0 < \beta < \frac{4}{3}} \min\left(\frac{\alpha}{2} - \frac{\beta}{4}, 1 - \frac{3}{4}\beta, \frac{\kappa\beta}{2}\right) = \frac{2\kappa}{2\kappa + 3},$$

where the maximum is attained when  $\beta = 4/(2\kappa + 3)$  and  $\alpha > \frac{\beta}{2}$ .

A proof of Theorem 2 will be given in Appendix B.

**Remark 3.** In Theorem 2, when  $\kappa = 1/2$ , the best rate under the general sampling scheme is  $1/6$  and under synchronous and equally spaced sampling, it is  $1/4$ .

## 4. Simulation Study

### 4.1. Simulation Settings

In this section we present an extensive numerical simulated study. The aim of this study is twofold: first in Section 4.2 we analyze the sensitivity of the estimator to the choice of parameters  $N$  and  $M$  and, with an unfeasible optimization, we find their optimal choices in different scenarios, having as a guide the theoretical results established in the previous section. Secondly, in Section 4.3 we evaluate the accuracy of the proposed GPDF estimator and its ability to produce symmetric and positive semi-definite estimations in a comparison with two alternative estimators that are present in the literature.

To give robustness to the results of our study, in the following analysis we consider many different simulation scenarios, focusing on both the two components of high-frequency financial data: the efficient price and the additive noise component given by market microstructure, so that the observed price  $\tilde{X}$  is

$$\tilde{X}_t^j = X_t^j + \eta_t^j, \quad j = 1, \dots, d, \quad (14)$$

with  $\eta$  being the noise component.

In particular, for the efficient price process we consider the following specifications:

- the Heston stochastic volatility model, by [Heston \(1993\)](#);
- the One Factor stochastic volatility model (SVF1);
- the Two Factor stochastic volatility model (SVF2), by [Chernov et al. \(2003\)](#);
- the Rough Heston model (RH), by [El Euch and Rosenbaum \(2019\)](#);

while for the additive microstructure noise we take into account the following cases:

- no noise case;
- noise coming from rounding;
- i.i.d. noise;
- autocorrelated noise;
- general noise auto-correlated and dependent of the efficient price process.

Since in the different cases where noise is present we analyze, respectively, 2, 4, 3, 4, different parameterizations, in our simulated analysis we study a total of 56 different scenarios.

For simplicity of the computations, through Sections 4.2 and 4.3, all the simulated analysis is conducted on the interval  $[0, 1]$ ; for that reason, and in the light of Section 3, we use the GPDF estimator (7) given by

$$\hat{V}_{N,M}^{j,j'}(t) = \frac{1}{2N+1} \sum_{l=1}^{n_j} \sum_{l'=1}^{n_{j'}} \sum_{u=-N}^N \sum_{u'=-N}^N e^{-\frac{2\pi^2(u-u')^2}{M}} e^{2\pi i u(t-t_l^j)} e^{-2\pi i u'(t-t_{l'}^{j'})} \Delta(X_l^j) \Delta(X_{l'}^{j'}). \quad (15)$$



where not stated otherwise, the simulations consist of  $K = 5000$  daily trajectories, considering a trading day of length 6.5 h, and are carried out on an equally spaced grid of width 2 s. To introduce asynchronicity in the data, observations are drawn from a Poisson process with an average of one observation every 10 s. Moreover, where not explicitly stated, the correlation between Brownian motions driving the efficient processes of different assets, following [Bibinger et al. \(2019\)](#), is fixed to mimic the median estimated realized correlation of the Nasdaq components, i.e.,

$$\langle W^j, W^i \rangle = 0.312, \quad j, i = 1, \dots, d \quad j \neq i.$$

In Sections 4.1.1 and 4.1.2, we define the models used for the efficient price process and the microstructure noise. As a robustness check, additional unreported simulations have been carried out under slightly different choices for the parameters of the reference models, with results analogous to those reported in Section 4.3.

#### 4.1.1. Efficient Price Process

##### Heston Model

The Heston stochastic volatility model by [Heston \(1993\)](#) is possibly the most widely used stochastic volatility model in the high-frequency econometric literature. It takes the form

$$\begin{cases} dX_t^j &= (\mu - (\sigma_t^j)^2/2)dt + \sigma_t^j dW_t^j \\ d(\sigma_t^j)^2 &= \gamma(\theta - (\sigma_t^j)^2)dt + \nu\sigma_t^j dZ_t^j, \end{cases}$$

with  $\langle W^j, Z^j \rangle = \lambda$  to account for the leverage effect. The parameters are set to be

$$(\mu, \gamma, \theta, \nu, \lambda) = (0.05/252, 5/252, 0.1, 0.5/252, -0.5),$$

that is the same choice made by [Zu and Boswijk \(2014\)](#); [Mancino et al. \(2017\)](#) and [Figueroa-López and Wu \(2022\)](#).

##### Factor Volatility Models

Factor volatility models have been long used in the literature; see, for example, [Huang and Tauchen \(2005\)](#). First, we consider the One-Factor Stochastic Volatility model (SV1F) of the form

$$\begin{cases} dX_t^j &= \mu dt + \sigma_t^j dW_t^j \\ \sigma_t^j &= e^{\beta_0 + \beta_1 \tau_t^j} \\ d\tau_t^j &= \alpha \tau_t^j + dZ_t^j \end{cases}$$

for  $j = 1, \dots, d$ , with  $\langle W^j, Z^j \rangle = \lambda$ , and  $\langle Z^j, Z^{j'} \rangle = 0$  for  $j \neq j'$ . The simulation is carried out using as parameters

$$(\mu, \beta_1, \alpha, \beta_0, \lambda) = (0.03, 0.125, -0.025, \beta_1/(2\alpha), -0.3),$$

which are the parameters used also in [Zu and Boswijk \(2014\)](#); [Mancino et al. \(2017\)](#); [Figueroa-López and Wu \(2022\)](#) and [Mancino et al. \(2024\)](#).

Second, we consider the Two-Factor Stochastic Volatility model (SV2F), proposed by [Chernov et al. \(2003\)](#), which is able to reproduce higher values of volatility of volatility. It takes the form

$$\begin{cases} dX_t^j &= \mu dt + s \cdot \exp[\beta_0 + \beta_1 \tau_t^{j,1} + \beta_2 \tau_t^{j,2}] dW_t^j \\ d\tau_t^{j,1} &= \alpha_1 \tau_t^{j,1} + dZ_t^{j,1} \\ d\tau_t^{j,2} &= \alpha_2 \tau_t^{j,2} + (1 + \beta_v \tau_t^{j,2}) dZ_t^{j,2} \end{cases}$$

for  $j = 1, \dots, d$ , with  $\langle W^j, Z^{j,1} \rangle = \langle W^j, Z^{j,2} \rangle = \lambda$ , and  $\langle Z^{j,i}, Z^{j',i'} \rangle = 0$  for  $j \neq j'$  and  $i \neq i'$ ,  $i, i' = 1, 2$ . For the parameters involved, our choice is to use

$$(\mu, \beta_0, \beta_1, \beta_2, \beta_v, \alpha_1, \alpha_2, \lambda) = (0.03, -1.1, 0.04, 0.3, -0.003, -0.6, 0.25).$$

### Rough Volatility

Starting with the seminal paper by [Gatheral et al. \(2018\)](#), a new strand of financial econometric literature has grown considering dynamics of the volatility process that are not driven by a standard Brownian motion but instead are driven by a fractional Brownian motion, with Hurst index  $H < 0.5$ , which corresponds to the cases where  $\kappa < 0.5$ . Theorem 2 states that the proposed PDF estimator is consistent even in the presence of rough volatility.

Rough volatility may also be modeled using the stochastic Volterra equation, as in the rough Heston model studied by [El Euch and Rosenbaum \(2019\)](#) and which we intend to use in this study.

$$\begin{cases} X_t^j &= X_0^j + \int_0^t X_s^j \sigma_s^j dW_s^j \\ (\sigma_t^j)^2 &= (\sigma_0^j)^2 + \int_0^t K(t-s) \left( (\theta - \gamma(\sigma_s^j)^2) ds + v \sigma_s^j dZ_s^j \right), \end{cases}$$

with  $\langle W^j, Z^j \rangle = \lambda$  and  $K(t) = Ct^{H-\frac{1}{2}}$  for  $H \in (0, \frac{1}{2})$  and constant  $C$ . In order to simulate the rough Heston model we apply the discrete-time Euler-type scheme studied in [Richard et al. \(2023\)](#), referring in particular to Equation (14) thereof. The parameters of the model are set to ensure that in the exercise the simulated volatility process does not exhibit negative values, and in particular they take values:

$$(\theta, \gamma, v, \lambda, H) = (0.2, 0.3, 0.2, -0.7, 0.1),$$

where the choice for the Hurst parameters is driven by the empirical evidence present in the literature; see, e.g., [Gatheral et al. \(2018\)](#).

#### 4.1.2. Market Microstructure Noise Specifications

It is a known fact (see, e.g., [Bandi and Russell \(2008\)](#)) that high-frequency data are contaminated by the so-called market microstructure noise. In particular, the observed price is the sum of the efficient price and the noise component. The origin of noise is linked to specific characteristics of the microstructure of financial markets, such as bid-ask spread, rounding, strategic trading (see, e.g., [Hasbrouck \(2007\)](#)), and several models specifications for the noise have been proposed in the literature of high-frequency financial econometrics.

#### Noise Coming from Rounding

In the presence of rounding the observed price process has the following form:

$$\tilde{X}_t^j = \log([\exp(X_t^j)/r]r),$$

where  $X$  denotes the efficient price process.

We consider two levels of rounding, corresponding to  $r = 1$  or 5 cents, which are the most used in financial markets.

Noise i.i.d.

The most widely used characterization of noise is to consider it an i.i.d. additive component, as in (14), with mean equal to zero and given constant variance

$$\eta \sim i.i.d. \quad E[\eta_t^j] = 0, \quad E[(\eta_t^j)^2] = \text{var}(\Delta X_{10sec}^j) \sigma_\eta^2, \quad \forall i, j$$

where  $X_{10sec}^j$  denotes the regularly spaced series obtained by subsampling the originally simulated series every 10 s.

Here, as can often be found in the literature, we specify a Gaussian distribution for the noise. We consider four values for the variance of the noise component:  $\sigma_\eta^2 = 1, 1.5, 2, 2.5$ .

Autocorrelated Noise

Here autocorrelation is introduced in the noise component, while keeping the additive form of Equation (14). In particular, it is modeled through an Ornstein–Uhlenbeck (OU) process defined as

$$d\eta_t^j = -\theta_\eta \eta_t^j dt + \sigma_\eta dE_t^j,$$

where  $E$  is a standard Brownian motion independent of  $W$ . Three different levels of autocorrelation are considered, using  $\theta_\eta = 0.2, 0.3, 0.4$ .  $\sigma_\eta^2$  is set to obtain a level of variance comparable to the second case in the previous scenario.

General Noise Correlated with the Efficient Price Process

In the final noise scenario, we opt for the general structure used in [Jacod et al. \(2019\)](#), which allows for both autocorrelation and dependence on the price process. For  $j \in \{1, 2\}$  and  $i = 0, 1, \dots, n_j$ , using the simplified notation  $\eta_i^j := \eta_{t_i}^j$ , the latter reads as

$$\eta_i^j = \psi_i^j \chi_i^j,$$

where  $\chi_i$  satisfies

$$\chi_i^j = Z_i^j + \sum_{l=1}^L \frac{g(1+g) \dots (l-1+g)}{l!} Z_{i-l}^j, \quad g \in (-0.5, 0.5), \quad Z_i^j \sim_{i.i.d.} \mathcal{N}(0, z), \quad (16)$$

and  $\psi_i$  is sampled from

$$d\psi^j(t) = u(h(t) - \psi^j(t))dt + v dW_t^j, \quad h(t) = 1 + w \cos\left(\frac{2\pi}{T}t\right), \quad (17)$$

with  $W^1$  and  $W^2$  being the Brownian motions driving the dynamics of the efficient prices  $p^1$  and  $p^2$ .

This last model attempts to replicate the slow-decaying autocorrelation in the noise process empirically observed by [Jacod et al. \(2017\)](#), while accounting for the possible dependence between noise and the efficient price component, as observed, for example, by [Hansen and Lunde \(2006\)](#). This formulation still includes an OU dynamics in Equation (17), but modified to also account for heteroskedasticity in the noise, reproducing in particular a U-shaped pattern for the volatility of the noise, given by the deterministic component  $h$ .

For the simulation of the noise, we use the following parameter selection:

$$(z, L, u, v) = (\text{var}(\Delta X_{10sec}^j) \cdot 2.5, 100, 10, 0.5),$$

with the four possible cases obtained coupling  $g = 0.3, 0.45$  and  $w = 0.3, 0.9$ .

In a last, unreported, exercise, we also add a rounding of 1 cent to this simulation scheme, without significant changes in the results.

#### 4.2. Selection of Parameters $N$ and $M$

In this section we want to evaluate the sensitivity of the estimator to the choice of the parameter  $N$  and  $M$  appearing in the definition of the GPDF estimator. The performance of the estimator for each couple of parameters is evaluated over the entire time interval, across  $K$  simulated independent trajectories. In all the following analyses, the path of the spot variance is reconstructed on a regular grid of width 30 min.

In the optimization study, following the results of Theorem 2, we specify the cutting frequency  $N$  and the localization parameter  $M$  in terms of  $\rho_n$ , using the fact that  $N\rho_n^\alpha \sim c_N$  and  $M\rho_n^\beta \sim c_M$ , for a suitable choice of  $\alpha$  and  $\beta$ , depending on the  $\kappa$ -Hölder continuity of the simulated volatility process<sup>5</sup>. In particular, we optimize over a grid defined by

$$c_N = 0.5, 1, 3, 5, 7, 9$$

and

$$c_M = 0.5, 1, 2, 3, 4, 5.$$

For each scenario, in a setting with  $d = 2$ , on the defined grid of values for the couple  $(c_N, c_M)$ , we look at the estimation error of variance  $\hat{V}^{1,1}$  and covariance  $\hat{V}^{1,2}$ , using in particular the integrated mean squared error

$$MISE_j = K^{-1} \sum_{k=1}^K \int_0^1 (\hat{V}^{1,j} - V^{1,j})^2 dt, \quad j = 1, 2$$

and choosing as optimal the pair that minimizes

$$0.1 \cdot MISE_1 + 0.9 \cdot MISE_2,$$

where a higher weight is given to the estimation of the covariance, being the dominant component of a generic variance–covariance matrix with  $d^2 - d$  covariance terms.<sup>6</sup>

Table 1 shows the optimal couple of  $c_N$  and  $c_M$  for each scenario. The first result we notice is that the optimal value for  $c_M$  is pretty stable across the models considered in the analysis for the efficient price process. When the data are affected by noise, the optimal  $c_N$  is reduced, coherently with known results for the other Fourier-type estimators of both integrated and spot volatility, see, e.g., Mancino et al. (2017) and Mancino et al. (2024), with a reduction that is stronger when the variance of noise is higher and in the presence of strong autocorrelation. Concerning the parameter  $c_M$ , it exhibits a more stable optimal value in the scenario considered, with only a small downward correction needed in the presence of noise. In this case,  $c_M$  manages the localizing Gaussian kernel, but a similar behavior was also observed for the original Fourier estimator in Mancino et al. (2024). Moreover, for the four models of the efficient price, the difference between close values of  $c_N$  and  $c_M$  is relatively small, meaning that making a slightly sub-optimal choice does not induce a significant increase in the error. This is shown in Table 2, where the  $MISE_2$  on the defined grid is reported for selected scenarios of under the Heston and the SVF2 model. From Table 2 it is also clear that overall, the estimator, on the grid used for this exercise, is more sensible to the choice of  $c_M$  than to the choice of  $c_N$ , with suboptimal values of  $c_M$  producing a higher increase in the estimation error w.r.t. suboptimal values of  $c_N$ . The figures for the remaining scenarios not reported in Table 2 are analogous.

**Table 1.** Optimal couple of  $c_N$ ,  $c_M$  in the considered grid across the different models for volatility and microstructure noise.

	Heston	SVF1	SVF2	RH
No noise				
	5, 1	5, 1	5, 1	5, 1
r	Noise from rounding			
0.01	5, 1	5, 1	5, 1	5, 1
0.05	5, 1	5, 1	5, 1	5, 1
$\sigma_\eta$	I.i.d. noise			
1	3, 0.5	3, 0.5	3, 0.5	3, 0.5
1.5	3, 0.5	3, 0.5	3, 0.5	3, 0.5
2	3, 0.5	3, 0.5	3, 0.5	3, 0.5
2.5	1, 0.5	1, 0.5	1, 0.5	1, 0.5
$\theta$	Auto-correlated noise			
0.2	1, 0.5	1, 0.5	1, 0.5	1, 0.5
0.3	1, 0.5	1, 0.5	1, 0.5	1, 0.5
0.4	3, 0.5	3, 0.5	3, 0.5	3, 0.5
$g, w$	General noise			
0.3, 0.3	3, 0.5	3, 0.5	3, 0.5	3, 0.5
0.3, 0.9	3, 0.5	3, 0.5	1, 0.5	1, 0.5
0.45, 0.3	1, 0.5	1, 0.5	1, 0.5	1, 0.5
0.45, 0.9	1, 0.5	1, 0.5	1, 0.5	1, 0.5

**Table 2.**  $MISE_2$  in estimating covariance over the considered grid, for selected scenarios.

Heston—No noise						
$c_N/c_M$	0.5	1	2	3	4	5
0.5	$3.068 \cdot 10^{-4}$	$4.370 \cdot 10^{-4}$	$6.193 \cdot 10^{-4}$	$7.539 \cdot 10^{-4}$	$8.646 \cdot 10^{-4}$	$9.510 \cdot 10^{-4}$
1	$1.539 \cdot 10^{-4}$	$2.124 \cdot 10^{-4}$	$2.971 \cdot 10^{-4}$	$3.619 \cdot 10^{-4}$	$4.164 \cdot 10^{-4}$	$4.331 \cdot 10^{-4}$
3	$5.172 \cdot 10^{-5}$	$7.101 \cdot 10^{-5}$	$1.002 \cdot 10^{-4}$	$1.232 \cdot 10^{-4}$	$1.427 \cdot 10^{-4}$	$1.553 \cdot 10^{-4}$
5	$3.985 \cdot 10^{-5}$	$5.238 \cdot 10^{-5}$	$7.081 \cdot 10^{-5}$	$8.488 \cdot 10^{-5}$	$9.660 \cdot 10^{-5}$	$1.023 \cdot 10^{-4}$
7	$4.657 \cdot 10^{-5}$	$5.541 \cdot 10^{-5}$	$6.846 \cdot 10^{-5}$	$7.847 \cdot 10^{-5}$	$8.687 \cdot 10^{-5}$	$9.636 \cdot 10^{-5}$
9	$5.732 \cdot 10^{-5}$	$6.524 \cdot 10^{-5}$	$6.989 \cdot 10^{-5}$	$8.874 \cdot 10^{-5}$	$9.113 \cdot 10^{-5}$	$9.995 \cdot 10^{-5}$
Heston—I.i.d. noise $\sigma_\eta = 2.5$						
$c_N/c_M$	0.5	1	2	3	4	5
0.5	$3.215 \cdot 10^{-4}$	$4.562 \cdot 10^{-4}$	$6.456 \cdot 10^{-4}$	$7.856 \cdot 10^{-4}$	$9.006 \cdot 10^{-4}$	$9.228 \cdot 10^{-4}$
1	$1.046 \cdot 10^{-4}$	$1.509 \cdot 10^{-4}$	$2.173 \cdot 10^{-4}$	$2.676 \cdot 10^{-4}$	$3.096 \cdot 10^{-4}$	$3.261 \cdot 10^{-4}$
3	$1.543 \cdot 10^{-4}$	$2.125 \cdot 10^{-4}$	$2.968 \cdot 10^{-4}$	$3.613 \cdot 10^{-4}$	$4.159 \cdot 10^{-4}$	$4.365 \cdot 10^{-4}$
5	$1.768 \cdot 10^{-4}$	$2.408 \cdot 10^{-4}$	$3.349 \cdot 10^{-4}$	$4.073 \cdot 10^{-4}$	$4.683 \cdot 10^{-4}$	$4.883 \cdot 10^{-4}$
7	$2.506 \cdot 10^{-4}$	$3.286 \cdot 10^{-4}$	$4.447 \cdot 10^{-4}$	$5.347 \cdot 10^{-4}$	$6.115 \cdot 10^{-4}$	$6.510 \cdot 10^{-4}$
9	$2.732 \cdot 10^{-4}$	$3.682 \cdot 10^{-4}$	$4.770 \cdot 10^{-4}$	$5.618 \cdot 10^{-4}$	$6.663 \cdot 10^{-4}$	$6.798 \cdot 10^{-4}$
SVF2—No noise						
$c_N/c_M$	0.5	1	2	3	4	5
0.5	$8.197 \cdot 10^{-4}$	$1.234 \cdot 10^{-3}$	$1.855 \cdot 10^{-3}$	$2.328 \cdot 10^{-3}$	$2.723 \cdot 10^{-3}$	$2.982 \cdot 10^{-4}$
1	$4.804 \cdot 10^{-4}$	$6.877 \cdot 10^{-4}$	$9.867 \cdot 10^{-4}$	$1.212 \cdot 10^{-3}$	$1.403 \cdot 10^{-3}$	$1.633 \cdot 10^{-4}$
3	$2.392 \cdot 10^{-4}$	$3.066 \cdot 10^{-4}$	$4.066 \cdot 10^{-4}$	$4.826 \cdot 10^{-4}$	$5.462 \cdot 10^{-4}$	$5.701 \cdot 10^{-4}$
5	$1.860 \cdot 10^{-4}$	$2.161 \cdot 10^{-4}$	$2.639 \cdot 10^{-4}$	$3.020 \cdot 10^{-4}$	$3.354 \cdot 10^{-4}$	$3.411 \cdot 10^{-4}$

Table 2. Cont.

7	$2.068 \cdot 10^{-4}$	$2.183 \cdot 10^{-4}$	$2.449 \cdot 10^{-4}$	$2.699 \cdot 10^{-4}$	$2.931 \cdot 10^{-4}$	$3.028 \cdot 10^{-4}$
9	$2.236 \cdot 10^{-4}$	$2.421 \cdot 10^{-4}$	$2.563 \cdot 10^{-4}$	$2.784 \cdot 10^{-4}$	$2.988 \cdot 10^{-4}$	$3.295 \cdot 10^{-4}$
SVF2—I.i.d. noise $\sigma_\eta = 2.5$						
$c_N/c_M$	0.5	1	2	3	4	5
0.5	$8.616 \cdot 10^{-4}$	$1.311 \cdot 10^{-3}$	$1.974 \cdot 10^{-3}$	$2.473 \cdot 10^{-3}$	$2.884 \cdot 10^{-3}$	$3.115 \cdot 10^{-3}$
1	$4.529 \cdot 10^{-4}$	$6.001 \cdot 10^{-4}$	$8.176 \cdot 10^{-4}$	$9.825 \cdot 10^{-4}$	$1.117 \cdot 10^{-3}$	$1.228 \cdot 10^{-3}$
3	$5.701 \cdot 10^{-4}$	$8.084 \cdot 10^{-4}$	$1.144 \cdot 10^{-3}$	$1.394 \cdot 10^{-3}$	$1.604 \cdot 10^{-3}$	$1.773 \cdot 10^{-3}$
5	$6.232 \cdot 10^{-4}$	$8.579 \cdot 10^{-4}$	$1.163 \cdot 10^{-3}$	$1.377 \cdot 10^{-3}$	$1.553 \cdot 10^{-3}$	$1.640 \cdot 10^{-3}$
7	$6.173 \cdot 10^{-4}$	$8.607 \cdot 10^{-4}$	$1.254 \cdot 10^{-3}$	$1.574 \cdot 10^{-3}$	$1.849 \cdot 10^{-3}$	$1.930 \cdot 10^{-3}$
9	$6.454 \cdot 10^{-4}$	$8.773 \cdot 10^{-4}$	$1.296 \cdot 10^{-3}$	$1.602 \cdot 10^{-3}$	$1.890 \cdot 10^{-3}$	$1.981 \cdot 10^{-3}$

#### 4.3. Performance Comparison

After having analyzed the sensibility of our estimator to the choice of  $N$  and  $M$ , in this section we replicate the extensive simulation study adopted in Section 4.2 to evaluate the accuracy of the proposed PDF estimator in comparison with the one of two competing estimators that are present in the literature. Focusing on the estimators consistent in the presence of asynchronous observations contaminated by microstructure noise, we consider the following.

- The Gaussian positive definite Fourier estimator proposed in this work (GPDF);
- the smoothed two-scale spot estimator, by [Mykland et al. \(2019\)](#) (STS);
- the local method of moments spot estimators, by [Bibinger et al. \(2019\)](#) (LMM).

The kernel-based estimator proposed by [Bu et al. \(2023\)](#), while it may be extended to manage irregular and asynchronous observations, relies on specifically tuned shrinkage techniques to impose positive semi-definiteness of the estimation and is therefore not considered in this analysis.

The analysis is carried out considering a maximum of  $d = 40$  assets, and the performances of each estimator are evaluated according to the mean integrated square error (MISE) and its relative counterpart (RMISE), defined as

$$MISE = (Kd^2)^{-1} \sum_{k=1}^K \int_0^1 \sum_{j,i=1}^d (\hat{V}_k^{ij}(t) - V_k^{ij}(t))^2 dt,$$

$$RMISE = (Kd^2)^{-1} \sum_{k=1}^K \int_0^1 \sum_{j,i=1}^d (\hat{V}_k^{ij}(t) - V_k^{ij}(t))^2 / V_k^{ij}(t)^2 dt.$$

Unreported results show that using a different loss function, in particular the Frobenius norm, the Euclidean norm, or the  $l - 1$  norm of the difference between the estimated and the real spot volatility matrix, does not affect the rankings that emerge from Tables 3–8. Most importantly, in this analysis we pay particular attention to the percentage of symmetric positive semi-definite (spsd) variance–covariance matrix that each estimator is able to produce in the different scenarios.

For the LMM estimator, we use the parameters that were found to be optimal in the numerical analysis by [Bibinger et al. \(2019\)](#), while for the STS estimator we choose the parameters in a neighborhood of the one used by [Chen et al. \(2020\)](#), minimizing the mean square error obtained on auxiliary simulations. In the following, when not explicitly stated otherwise, the results are meant to be achieved under the Heston specification for the efficient price process.

#### 4.3.1. Absence of Noise

In Sections 4.3.1 and 4.3.2, we report the results of the comparison, in terms of MISE and percentage of spsd estimates produced by the three competing estimators, when the efficient price process follows the Heston model. In this setting, the results in terms of RMISE are analogous. We begin considering noise-free data and we focus on two major features of high-frequency covariance matrix estimation: the dimensionality of the matrix and the asynchronicity of observations.

Table 3 shows the results for increasing values of  $d$ . In this simulated exercise, the modified Fourier estimator performs the best in terms of MSE, for any dimension of the volatility matrix. The effectiveness of the STS estimator to produce positive semi-definite estimations seems to decrease as the number of assets increases, in particular with  $d > 20$ , while the other two estimators both produce 100% of spsd matrices, with a slight decrease for the LMM estimator observed for  $d = 40$ . For the important role that it plays in estimating variance–covariance matrices and for the influence that it has on the positivity of the estimation, dimensionality will always be taken into consideration in the remaining analysis. For simplicity of exposition, the dimensions considered in the following are limited to  $d = 5, 10, 15, 20$ .

**Table 3.** Accuracy (MISE) and % of spsd matrix produced by each estimator, when the dimension  $d$  of  $V$  increases.

Estimator	MISE	% SPSD	MISE	% SPSD
d = 2			d = 20	
GPDF	$6.773 \cdot 10^{-5}$	100%	$5.395 \cdot 10^{-5}$	100%
LMM	$2.563 \cdot 10^{-4}$	100%	$1.133 \cdot 10^{-4}$	100%
STS	$2.240 \cdot 10^{-4}$	100%	$2.156 \cdot 10^{-4}$	88.60%
d = 5			d = 25	
GPDF	$5.670 \cdot 10^{-5}$	100%	$5.258 \cdot 10^{-5}$	100%
LMM	$1.495 \cdot 10^{-4}$	100%	$9.659 \cdot 10^{-5}$	100%
STS	$2.201 \cdot 10^{-4}$	100%	$2.138 \cdot 10^{-4}$	66.22%
d = 10			d = 30	
GPDF	$5.316 \cdot 10^{-5}$	100%	$5.142 \cdot 10^{-5}$	100%
LMM	$1.215 \cdot 10^{-4}$	100%	$9.604 \cdot 10^{-5}$	100%
STS	$2.022 \cdot 10^{-4}$	100%	$2.006 \cdot 10^{-4}$	9.87%
d = 15			d = 40	
GPDF	$5.299 \cdot 10^{-5}$	100%	$5.223 \cdot 10^{-5}$	100%
LMM	$9.977 \cdot 10^{-4}$	100%	$9.501 \cdot 10^{-5}$	99.54%
STS	$2.019 \cdot 10^{-4}$	95.43%	$1.994 \cdot 10^{-4}$	3.02%

In the no-noise setting, we also address the issue of different levels of asynchronicity in the data. To do so, we examine the changes in the performance of the three estimators as the average time between two consecutive observations increases. In particular, we extract the observations from the simulated trajectories according to homogeneous Poisson processes that produce on average one observation every 15, 20, and 30 s. Table 4 shows that, still maintaining an edge in terms of MISE with respect to the competitors, the proposed estimator is the only one that is able to produce spsd estimations in 100% of the cases, while both the STS and the LMM estimator can fail with increased frequency when the asynchronicity increases, even though the impact of this kind of changes seems to be quite small in terms of percentage of positive estimation obtained. It also seems that the accuracy



of all the estimators decreases with higher values of  $\bar{\Delta}t$ ; this is of course in line with the fact that the consistency of these estimators is an asymptotic property.

**Table 4.** Accuracy and % of spsd matrix produced by each estimator, when the average time between consecutive observations  $\bar{\Delta}t$  changes.

Estimator	MISE	% SPSPD	MISE	% SPSPD	MISE	% SPSPD
d = 5, $\bar{\Delta}t = 15$			d = 5, $\bar{\Delta}t = 20$		d = 5, $\bar{\Delta}t = 30$	
GPDF	$6.314 \cdot 10^{-5}$	100%	$7.798 \cdot 10^{-5}$	100%	$9.723 \cdot 10^{-5}$	100%
LMM	$1.610 \cdot 10^{-4}$	100%	$1.885 \cdot 10^{-4}$	100%	$2.503 \cdot 10^{-4}$	99.54%
STS	$2.426 \cdot 10^{-4}$	100%	$2.603 \cdot 10^{-4}$	100%	$2.686 \cdot 10^{-4}$	100%
d = 10, $\bar{\Delta}t = 15$			d = 10, $\bar{\Delta}t = 20$		d = 10, $\bar{\Delta}t = 30$	
GPDF	$6.397 \cdot 10^{-5}$	100%	$7.339 \cdot 10^{-5}$	100%	$9.188 \cdot 10^{-5}$	100%
LMM	$1.956 \cdot 10^{-4}$	100%	$1.882 \cdot 10^{-4}$	100%	$1.975 \cdot 10^{-4}$	98.58%
STS	$2.303 \cdot 10^{-4}$	100%	$2.344 \cdot 10^{-4}$	100%	$2.395 \cdot 10^{-4}$	100%
d = 15, $\bar{\Delta}t = 15$			d = 15, $\bar{\Delta}t = 20$		d = 15, $\bar{\Delta}t = 30$	
GPDF	$6.322 \cdot 10^{-5}$	100%	$7.451 \cdot 10^{-5}$	100%	$9.169 \cdot 10^{-5}$	100%
LMM	$1.717 \cdot 10^{-4}$	100%	$1.445 \cdot 10^{-4}$	98.05%	$1.918 \cdot 10^{-4}$	97.68%
STS	$2.285 \cdot 10^{-4}$	99.90%	$2.310 \cdot 10^{-4}$	99.15%	$2.383 \cdot 10^{-4}$	98.54 %
d = 20, $\bar{\Delta}t = 15$			d = 20, $\bar{\Delta}t = 20$		d = 20, $\bar{\Delta}t = 30$	
GPDF	$6.325 \cdot 10^{-5}$	100%	$7.258 \cdot 10^{-5}$	100%	$9.224 \cdot 10^{-5}$	100%
LMM	$1.651 \cdot 10^{-4}$	99.83%	$1.322 \cdot 10^{-4}$	90.18%	$1.782 \cdot 10^{-4}$	96.67%
STS	$2.213 \cdot 10^{-4}$	91.45%	$2.289 \cdot 10^{-5}$	88.01%	$2.330 \cdot 10^{-4}$	82.66 %

#### 4.3.2. Data Contaminated by Microstructure Noise

In this Section we run our comparison considering the noise specification described in Section 4.1.2. It is useful to note that the LMM estimator entails an explicit noise correction, and the STS estimator relies on pre-averaging of the observed data on a synchronous and equally spaced grid. For the proposed GPDF estimator instead, in line with the original Fourier estimator of spot volatility, there is no need to manipulate the data or correct the estimator to manage the presence of noise, but it is sufficient to cut the frequency  $N$ , as shown in Section 4.2.

Table 5 shows that the presence of rounding does not appear to significantly affect the accuracy of the estimators and the positive semi-definiteness of the estimations. This effect may be due to the scheme adopted to simulate irregularly sampled data, that implies a sub-sampling with respect to the rounded simulated series, reducing the intensity of this source of noise, whose only impact is to slightly decrease the percentage of spsd estimation for the STS estimator.

**Table 5.** Accuracy and % of spsd matrix produced by each estimator, when a rounding of 1 or 5 cents is present.

Estimator	MISE	% SPSPD	MISE	% SPSPD
d = 5, r = 0.01			d = 5, r = 0.05	
GPDF	$5.583 \cdot 10^{-5}$	100%	$5.587 \cdot 10^{-5}$	100%
LMM	$1.638 \cdot 10^{-4}$	100%	$1.641 \cdot 10^{-4}$	100%
STS	$2.188 \cdot 10^{-4}$	100%	$2.188 \cdot 10^{-4}$	100%
d = 10, r = 0.01			d = 10, r = 0.05	
GPDF	$5.376 \cdot 10^{-5}$	100%	$5.377 \cdot 10^{-5}$	100%
LMM	$1.094 \cdot 10^{-4}$	100%	$1.095 \cdot 10^{-4}$	100%

Table 5. Cont.

Estimator	MISE	% SPSP	MISE	% SPSP
STS	$2.022 \cdot 10^{-4}$	100%	$2.023 \cdot 10^{-4}$	100%
$d = 15, r = 0.01$		$d = 15, r = 0.05$		
GPDP	$5.284 \cdot 10^{-5}$	100%	$5.284 \cdot 10^{-5}$	100%
LMM	$1.111 \cdot 10^{-4}$	100%	$1.113 \cdot 10^{-4}$	100%
STS	$1.998 \cdot 10^{-4}$	99.97%	$1.997 \cdot 10^{-4}$	99.85%
$d = 20, r = 0.01$		$d = 20, r = 0.05$		
GPDP	$5.203 \cdot 10^{-5}$	100%	$5.204 \cdot 10^{-5}$	100%
LMM	$1.011 \cdot 10^{-4}$	100%	$1.011 \cdot 10^{-4}$	100%
STS	$1.902 \cdot 10^{-4}$	97.66%	$1.902 \cdot 10^{-4}$	97.03%

Table 6 shows that i.i.d. noise, especially with high noise variance  $\sigma_\eta^2$ , is able to negatively affect the ability of the STS and, marginally, of the LMM estimators to produce spsd estimations, with a stronger impact as the dimension of the estimated matrix grows. Also, the accuracy of all the estimators deteriorates with higher noise, with the GPDP confirmed as the top performer also in this scenario, but with reduced differences in accuracy among the three estimators. Since i.i.d. noise, from a market microstructure perspective, is usually linked to the presence bid-ask spread as modeled, e.g., as in Roll (1984), and being the bid-ask spread usually related to the liquidity of an asset, the ability of managing this kind of noise may be regarded as the ability to estimate correctly covariance also for illiquid assets.

Table 6. Accuracy and % of spsd matrix produced by each estimator, when the data is contaminated by i.i.d. noise.

Estimator	MISE	% SPSP	MISE	% SPSP	MISE	% SPSP	MISE	% SPSP
$d = 5, \sigma_\eta = 1$		$d = 5, \sigma_\eta = 1.5$		$d = 5, \sigma_\eta = 2$		$d = 5, \sigma_\eta = 2.5$		
GPDP	$8.229 \cdot 10^{-5}$	100%	$8.242 \cdot 10^{-5}$	100%	$1.350 \cdot 10^{-4}$	100%	$2.054 \cdot 10^{-4}$	100%
LMM	$1.485 \cdot 10^{-4}$	100%	$1.588 \cdot 10^{-4}$	100%	$1.727 \cdot 10^{-4}$	100%	$1.991 \cdot 10^{-4}$	100%
STS	$2.478 \cdot 10^{-4}$	100%	$2.567 \cdot 10^{-4}$	100%	$2.910 \cdot 10^{-4}$	100%	$3.361 \cdot 10^{-4}$	97.25%
$d = 10, \sigma_\eta = 1$		$d = 10, \sigma_\eta = 1.5$		$d = 10, \sigma_\eta = 2$		$d = 10, \sigma_\eta = 2.5$		
GPDP	$6.922 \cdot 10^{-5}$	100%	$7.901 \cdot 10^{-5}$	100%	$1.503 \cdot 10^{-4}$	100%	$1.889 \cdot 10^{-4}$	100%
LMM	$1.320 \cdot 10^{-4}$	100%	$1.191 \cdot 10^{-4}$	100%	$1.583 \cdot 10^{-4}$	100%	$2.198 \cdot 10^{-4}$	99.25%
STS	$2.236 \cdot 10^{-4}$	98.85%	$2.369 \cdot 10^{-4}$	98.25%	$2.610 \cdot 10^{-4}$	86.47%	$3.028 \cdot 10^{-4}$	55.67%
$d = 15, \sigma_\eta = 1$		$d = 15, \sigma_\eta = 1.5$		$d = 15, \sigma_\eta = 2$		$d = 15, \sigma_\eta = 2.5$		
GPDP	$6.873 \cdot 10^{-5}$	100%	$7.588 \cdot 10^{-5}$	100%	$1.474 \cdot 10^{-4}$	100%	$1.801 \cdot 10^{-4}$	100%
LMM	$1.225 \cdot 10^{-4}$	100%	$1.348 \cdot 10^{-4}$	100%	$1.920 \cdot 10^{-4}$	100%	$1.955 \cdot 10^{-4}$	98.42%
STS	$2.042 \cdot 10^{-4}$	97.43%	$2.187 \cdot 10^{-4}$	76.01%	$2.738 \cdot 10^{-4}$	28.10%	$3.008 \cdot 10^{-4}$	20.13%
$d = 20, \sigma_\eta = 1$		$d = 20, \sigma_\eta = 1.5$		$d = 20, \sigma_\eta = 2$		$d = 20, \sigma_\eta = 2.5$		
GPDP	$6.468 \cdot 10^{-5}$	100%	$7.622 \cdot 10^{-5}$	100%	$1.497 \cdot 10^{-4}$	100%	$1.628 \cdot 10^{-4}$	100%
LMM	$1.008 \cdot 10^{-4}$	100%	$1.581 \cdot 10^{-4}$	97.83%	$1.621 \cdot 10^{-4}$	96.26%	$2.044 \cdot 10^{-4}$	92.10%
STS	$2.250 \cdot 10^{-4}$	65.27%	$2.390 \cdot 10^{-4}$	15.17%	$2.503 \cdot 10^{-4}$	6.27%	$2.817 \cdot 10^{-4}$	0.0%

Table 7 shows that autocorrelated noise is again able to significantly effect the ability of the STS estimators to produce spsd matrices, in particular with low values of  $\theta_\eta$ , while the LMM estimator seems to be only slightly affected. Low values of  $\theta_\eta$ , i.e., higher values

of autocorrelation in the noise process, reduce the accuracy of the three competitors, while maintaining their ranking substantially unchanged.

**Table 7.** Accuracy and % of spsd matrix produced by each estimator, when the data is contaminated by autocorrelated noise.

Estimator	MISE	% SPSD	MISE	% SPSD	MISE	% SPSD
	d = 5, $\theta_\eta = 0.2$		d = 5, $\theta_\eta = 0.3$		d = 5, $\theta_\eta = 0.4$	
GPDF	$2.093 \cdot 10^{-4}$	100%	$1.956 \cdot 10^{-4}$	100%	$1.743 \cdot 10^{-4}$	100%
LMM	$2.187 \cdot 10^{-4}$	100%	$2.632 \cdot 10^{-4}$	100%	$2.693 \cdot 10^{-4}$	100%
STS	$4.134 \cdot 10^{-4}$	90.47%	$3.026 \cdot 10^{-4}$	98.98%	$2.945 \cdot 10^{-4}$	97.23%
	d = 10, $\theta_\eta = 0.2$		d = 10, $\theta_\eta = 0.3$		d = 10, $\theta_\eta = 0.4$	
GPDF	$1.887 \cdot 10^{-4}$	100%	$1.810 \cdot 10^{-4}$	100%	$1.752 \cdot 10^{-4}$	100%
LMM	$2.789 \cdot 10^{-4}$	100%	$2.863 \cdot 10^{-4}$	98.94%	$1.953 \cdot 10^{-4}$	100%
STS	$3.934 \cdot 10^{-4}$	22.71%	$3.007 \cdot 10^{-4}$	65.58%	$2.666 \cdot 10^{-4}$	85.43%
	d = 15, $\theta_\eta = 0.2$		d = 15, $\theta_\eta = 0.3$		d = 15, $\theta_\eta = 0.4$	
GPDF	$1.755 \cdot 10^{-4}$	100%	$1.721 \cdot 10^{-4}$	100%	$1.599 \cdot 10^{-4}$	100%
LMM	$2.662 \cdot 10^{-4}$	100%	$2.659 \cdot 10^{-4}$	93.44%	$1.781 \cdot 10^{-4}$	95.90%
STS	$3.728 \cdot 10^{-4}$	2.01%	$3.001 \cdot 10^{-4}$	6.25%	$2.789 \cdot 10^{-4}$	20.99%
	d = 20, $\theta_\eta = 0.2$		d = 20, $\theta_\eta = 0.3$		d = 20, $\theta_\eta = 0.4$	
GPDF	$1.743 \cdot 10^{-4}$	100%	$1.657 \cdot 10^{-4}$	100%	$1.580 \cdot 10^{-4}$	100%
LMM	$2.514 \cdot 10^{-4}$	93.87%	$2.501 \cdot 10^{-4}$	93.48%	$1.751 \cdot 10^{-4}$	93.60%
STS	$3.899 \cdot 10^{-4}$	1.02%	$2.921 \cdot 10^{-4}$	2.03%	$2.700 \cdot 10^{-4}$	2.58%

Table 8 shows the results for the last specification of noise that we consider, that is, with general noise allowing for auto-covariance, correlation with the efficient price and time-varying noise variance. Our results show that, in this setting, both the LMM and the STS estimators may have difficulties in reaching satisfactory percentages of spsd estimations, depending on the intensity of the microstructure component. We confirm once again the ability of the GPDF estimator to produce variance–covariance matrix with the desired property, and with relatively low estimation error. at the same time, we still find that increasing the dimensionality of the estimation exercise hinders the ability of traditional estimators to produce spsd matrices.

**Table 8.** Accuracy and % of spsd matrix produced by each estimator, when the data is contaminated by the general noise process.

Estimator	MISE	% SPSD	MISE	% SPSD	MISE	% SPSD	MISE	% SPSD
	d = 5, g = 0.3, w = 0.3		d = 5, g = 0.3, w = 0.9		d = 5, g = 0.45, w = 0.3		d = 5, g = 0.45, w = 0.9	
GPDF	$2.543 \cdot 10^{-4}$	100%	$2.521 \cdot 10^{-4}$	100%	$2.845 \cdot 10^{-4}$	100%	$2.981 \cdot 10^{-4}$	100%
LMM	$2.701 \cdot 10^{-4}$	100%	$2.695 \cdot 10^{-4}$	99.52%	$3.103 \cdot 10^{-4}$	100%	$3.499 \cdot 10^{-4}$	99.40%
STS	$2.919 \cdot 10^{-4}$	99.87%	$3.184 \cdot 10^{-4}$	99.00%	$4.396 \cdot 10^{-4}$	99.83%	$4.716 \cdot 10^{-4}$	98.51%
	d = 10, g = 0.3, w = 0.3		d = 10, g = 0.3, w = 0.9		d = 10, g = 0.45, w = 0.3		d = 5, g = 0.45, w = 0.9	
GPDF	$1.762 \cdot 10^{-4}$	100%	$2.123 \cdot 10^{-4}$	100%	$2.458 \cdot 10^{-4}$	100%	$2.702 \cdot 10^{-4}$	100%
LMM	$2.198 \cdot 10^{-4}$	99.76%	$2.707 \cdot 10^{-4}$	97.17%	$2.660 \cdot 10^{-4}$	97.89%	$3.543 \cdot 10^{-4}$	91.06%
STS	$2.702 \cdot 10^{-4}$	99.26%	$3.005 \cdot 10^{-4}$	79.58%	$3.722 \cdot 10^{-4}$	90.36%	$4.601 \cdot 10^{-4}$	77.10%
	d = 15, g = 0.3, w = 0.3		d = 15, g = 0.3, w = 0.9		d = 15, g = 0.45, w = 0.3		d = 5, g = 0.45, w = 0.9	
GPDF	$1.508 \cdot 10^{-4}$	100%	$1.702 \cdot 10^{-4}$	100%	$2.344 \cdot 10^{-4}$	100%	$2.523 \cdot 10^{-4}$	100%
LMM	$1.891 \cdot 10^{-4}$	98.04%	$3.085 \cdot 10^{-4}$	87.25%	$2.622 \cdot 10^{-4}$	90.93%	$3.582 \cdot 10^{-4}$	81.48%

Table 8. Cont.

Estimator	MISE	% SPSPD	MISE	% SPSPD	MISE	% SPSPD	MISE	% SPSPD
STS	$2.671 \cdot 10^{-4}$	50.90%	$2.901 \cdot 10^{-4}$	47.12%	$3.600 \cdot 10^{-4}$	39.94%	$4.543 \cdot 10^{-4}$	40.23%
	$d = 20, g = 0.3, w = 0.3$		$d = 20, g = 0.3, w = 0.9$		$d = 20, g = 0.45, w = 0.3$		$d = 5, g = 0.45, w = 0.9$	
GPDF	$1.305 \cdot 10^{-4}$	100%	$1.587 \cdot 10^{-4}$	100%	$2.317 \cdot 10^{-4}$	100%	$2.458 \cdot 10^{-4}$	100%
LMM	$2.674 \cdot 10^{-4}$	96.15%	$2.991 \cdot 10^{-4}$	81.34%	$2.956 \cdot 10^{-4}$	85.17%	$3.478 \cdot 10^{-4}$	78.63%
STS	$2.667 \cdot 10^{-4}$	8.21%	$2.901 \cdot 10^{-4}$	21.94%	$3.478 \cdot 10^{-4}$	3.77%	$4.293 \cdot 10^{-4}$	16.42%

#### 4.4. Alternative Volatility Models

In the previous Sections the comparison results have been obtained in the case when the simulated efficient price process is an Heston model. Even though the error produced by the three estimators may be different changing the simulation model behind our analysis, and in particular the differences in MISE between the GPDF and the LMM estimators are reduced when using the SVF2 or the Rough Heston model, the results are substantially confirmed: the positive semi-definite Fourier estimator remain the only one able to consistently produce positive semi-definite estimations, and is the best performer in terms of MSE in almost any scenario. Table 9 shows the percentage of psd estimations obtained under the alternative volatility models, in absence of microstructure noise, together with the ranking of the estimators in terms of MISE. We can see that, in this exercise, it seems that, moving to the SV1F, to the SV2F or Rough Heston, does not influence significantly the ability of the estimators of producing positive matrices. Also the ranking of the estimators is essentially unaffected. More extensive results about the alternative models, showing the percentage of spsd estimations in the cases with rounding, i.i.d. and general noise, are reported in Appendix C. It is worth noting, as a final consideration, that the results in terms of RMISE are analogous, and always see the GPDF estimator having a competitive edge.

**Table 9.** The % of psd matrix produced by each estimator, when the efficient price process is produced by alternative models.

Estimator	SVF1		SVF2		Rough H.	
	MISE	% SPSPD	MISE	% SPSPD	MISE	% SPSPD
$d = 2$						
GPDF	$2.433 \cdot 10^{-5}$	100%	$2.441 \cdot 10^{-3}$	100%	$4.575 \cdot 10^{-3}$	100%
LMM	$6.799 \cdot 10^{-5}$	100%	$4.663 \cdot 10^{-3}$	100%	$6.020 \cdot 10^{-3}$	100%
STS	$9.345 \cdot 10^{-5}$	100%	$6.217 \cdot 10^{-3}$	100%	$6.995 \cdot 10^{-3}$	100%
$d = 5$						
GPDF	$1.805 \cdot 10^{-5}$	100%	$1.496 \cdot 10^{-3}$	100%	$2.421 \cdot 10^{-3}$	100%
LMM	$4.022 \cdot 10^{-5}$	100%	$2.036 \cdot 10^{-3}$	100%	$3.640 \cdot 10^{-3}$	100%
STS	$7.002 \cdot 10^{-5}$	100%	$3.391 \cdot 10^{-3}$	100%	$4.473 \cdot 10^{-3}$	100%
$d = 10$						
GPDF	$1.756 \cdot 10^{-5}$	100%	$6.802 \cdot 10^{-4}$	100%	$1.588 \cdot 10^{-3}$	100%
LMM	$3.670 \cdot 10^{-5}$	100%	$1.215 \cdot 10^{-3}$	100%	$2.036 \cdot 10^{-3}$	99.98%
STS	$7.287 \cdot 10^{-5}$	100%	$1.902 \cdot 10^{-3}$	100%	$3.654 \cdot 10^{-3}$	99.95%
$d = 15$						
GPDF	$1.699 \cdot 10^{-5}$	100%	$5.185 \cdot 10^{-4}$	100%	$1.403 \cdot 10^{-3}$	100%
LMM	$3.476 \cdot 10^{-5}$	100%	$9.004 \cdot 10^{-4}$	100%	$1.890 \cdot 10^{-3}$	99.64%
STS	$6.995 \cdot 10^{-5}$	100%	$1.634 \cdot 10^{-3}$	99.92%	$3.338 \cdot 10^{-3}$	99.74%

Table 9. Cont.

Estimator	SVF1		SVF2		Rough H.	
	MISE	% SPSP	MISE	% SPSP	MISE	% SPSP
d = 20						
GPDF	$1.781 \cdot 10^{-5}$	100%	$4.561 \cdot 10^{-4}$	100%	$1.241 \cdot 10^{-3}$	100%
LMM	$3.286 \cdot 10^{-5}$	100%	$8.639 \cdot 10^{-4}$	99.76%	$1.703 \cdot 10^{-3}$	99.55%
STS	$6.994 \cdot 10^{-5}$	99.56%	$1.464 \cdot 10^{-3}$	96.85%	$3.263 \cdot 10^{-3}$	98.23%

## 5. Empirical Study

In this section, we aim at using historical market data to comment, in a realistic case of use, on the usability of the proposed GPDF estimator in real world applications, complementing the analysis of the theoretical properties and finite sample performance that was carried out in the previous sections. In particular, we focus on two topics at same time: first, the stability of the estimation in the region of parameters identified as optimal in Section 4.2, since a clear identification of the appropriate scenario for noise might be challenging; second the speed of execution of the computation, in relation to the choice of  $N$  and  $M$ .

The exercise is carried out on a selection of 12 firms traded on the NYSE, coming from a variety of business sectors. In particular we consider: Accenture (ACN), American Tower corporation (AMT), The Boeing Company (BA), ConocoPhillips (COP), General Electric Company (GE), The Goldman Sachs Group (GS), Eli Lilly and Company (LLY), McDonald's Corporation (MCD), NextEra Energy (NEE), Netflix (NFLX), Nike (NKE) and Walmart (WMT). The stocks, although all of them are from large companies, may exhibit different liquidity profiles, thus creating a mixture of different specification of noise that might be better captured in the most general form defined in Section 4.1.2. The selected stocks are recorded at the trade frequency for the time horizon of one year, from January 2023 to December 2023. Similar to the approach used in the previous sections, the volatility matrix is estimated every 30 min for each of the 248 full trading days of the sample.

In Section 4.2, the optimal choice for  $c_M$  in the presence of noise is consistently 0.5, while the value of  $c_N$  that minimize the MISE appears to lay in the range between 1 and 3, thus we perform the estimation exercise focusing on  $c_N \in \{1, 2, 3\}$  and  $c_M \in \{0.5, 1\}$ , where  $c_M=1$  is added to evaluate the impact of moving to a different  $c_M$  on the computational time and the impact of shifting to more sub-optimal parameterizations on the estimation itself. Table 10 shows the results of this analysis. In particular, for each couple of selected  $(c_N, c_M)$ , it reports the computational time in seconds<sup>7</sup> (computed as the average time needed to estimate the trajectory of the volatility matrix in one day on the defined time grid) and the average relative distance between the estimation and the baseline case of  $c_N = 3$  and  $c_M = 0.5$ , when the estimation is computed in the middle of the trading day (i.e., in  $t = 0.5$  for  $t \in [0, 1]$ ) and the relative distance is defined as the ratio between the Euclidean norm of the difference between the two estimations and the Euclidean norm of the baseline case:

$$d_{emp} = \sum_{k=1}^K \frac{\|\hat{V}^{0.5,k} - \hat{V}_{base}^{0.5,k}\|}{\|\hat{V}_{base}^{0.5,k}\|},$$

with  $K$  being the number of days in the sample.

Overall, the distance in estimation seems to be relatively close and, in most cases, it is smaller or comparable to the relative difference observed in the simulation exercise between the MISE of the GPDF and the ones of the alternative competitors. This confirms that, in the region identified in the simulation study (and even for worsen values of  $c_M$ ), the estimations appear to be sufficiently stable even when applied to empirical data.

**Table 10.** Average computational time (in seconds) for estimation of volatility paths for different couple of  $(c_N, c_M)$  and relative distance to baseline case  $(c_N = 3, c_M = 0.5)$ .

$(c_N, c_M)$	(1, 0.5)	(2, 0.5)	(3, 0.5)	(1, 1)	(2, 1)	(3, 1)
avg. comp. time	1.0552	2.113	3.1014	1.0666	2.3161	3.6283
$d_{emp}$	0.2674	0.1298	0	0.3966	0.3404	0.3124

## 6. Conclusions

In the present paper, a modified version of the classical Fourier estimator for spot covariance by [Malliavin and Mancino \(2009\)](#) has been proposed to overcome the difficulty of obtaining symmetric and positive semi-definite estimation of the spot variance-covariance matrix. We showed that the proposed estimator is positive semi-definite and consistent with a suitable choice of the tuning parameters  $N, M$ . To the best of our knowledge, this is the first non-parametric estimator of the spot covariance that guarantees the positiveness. Based on the theoretical results obtained, a numerical study has been carried out to evaluate the optimal choice of the parameters in a variety of settings. The optimal couple seems to be quite stable, and, as usual for the class of the Fourier estimators, in the presence of asynchronicity and noisy data, the parameter  $N$  should be reduced with respect to the optimal no-noise case, which is the Nyquist frequency. Moreover, a thorough simulation study has been carried out to evaluate the accuracy of the estimator and its actual ability in producing psd estimations. Comparing the results with the ones of two alternative estimators present in the literature, the STS and the LMM estimators, we found out that the proposed PDF estimator usually outperforms the competitors in terms of mean square error and is the only one that, in this study, was able to always produce psd estimations. The simulation analysis was focused on many challenging aspects of high-frequency covariance estimations, such as the dimensionality of the problem, the degree of asynchronicity between assets and the presence of multiple specifications of market microstructure noise. The robustness of our results are confirmed using alternative data generating processes. The empirical exercise demonstrates the applicability of the method using real-world data. Therefore, we believe that the PDF estimator of spot covariance may be more competitive in terms of empirical applications due to its properties.

**Author Contributions:** Methodology, J.A., M.E.M., T.M. and Y.Y.; Investigation, R.K.; Data curation, T.M.; Writing—original draft, J.A., N.-L.L. and M.E.M. All authors have read and agreed to the published version of the manuscript.

**Funding:** Maria Elvira Mancino was partially supported from the European Union Next-GenerationEU—National Recovery and Resilience Plan (NRRP)—MISSION 4 COMPONENT 2, INVESTIMENT 1.1 grant number P2022AX5TH.

**Data Availability Statement:** The original contributions presented in this study are included in the article. Further inquiries can be directed to the corresponding author(s).

**Conflicts of Interest:** The authors declare no conflict of interest.

## Appendix A. Proof of Theorem 1

Let  $a_j$  for  $j = 1, 2, 3$  be arbitrary functions on  $\mathbb{Z}$ , from the definitions of  $\mathcal{K}$  and  $\mathcal{S}(k)$  we notice that:

$$\begin{aligned}
 & \sum_{k \in \mathcal{K}} \sum_{(s, s') \in \mathcal{S}(k)} a_1(k) a_2(s) a_3(s') \\
 &= \sum_{k=0}^{2N} \sum_{v=0}^{2N-k} a_1(k) a_2(-N+k+v) a_3(N-v) + \sum_{k=-2N}^{-1} \sum_{v=0}^{2N+k} a_1(k) a_2(N+k-v) a_3(-N+v) =: A + B.
 \end{aligned}$$

For the first term we have

$$\begin{aligned}
 A &= \sum_{k=0}^{2N} \sum_{u=k-N}^N a_1(k) a_2(k-u) a_3(u) \\
 &= \sum_{u=-N}^N \sum_{k=0}^{u+N} a_1(k) a_2(k-u) a_3(u) \\
 &= \sum_{u=-N}^N \sum_{u'=-N}^N a_1(u+u') a_2(u') a_3(u),
 \end{aligned}$$

where we set  $u = N - v$  in the first line, changed the order of the summations in the second line, and put  $u' = k - u$ . Similarly, using the convention that  $\sum_{u=0}^{-1} = 0$ , for the second term we have

$$\begin{aligned}
 B &= \sum_{k=-2N}^{-1} \sum_{u=-N}^{N+k} a_1(k) a_2(k-u) a_3(u) \\
 &= \sum_{u=-N}^N \sum_{k=u-N}^{-1} a_1(k) a_2(k-u) a_3(u) \\
 &= \sum_{u=-N}^N \sum_{u'=-N}^{-u-1} a_1(u+u') a_2(u') a_3(u).
 \end{aligned}$$

Thus we see that

$$\sum_{k \in \mathcal{K}} \sum_{(s,s') \in \mathcal{S}(k)} a_1(k) a_2(s) a_3(s') = \sum_{u=-N}^N \sum_{u'=-N}^N a_1(u+u') a_2(u') a_3(u).$$

When  $a_1(k) = c(k) e^{2\pi i k t}$ ,  $a_2(s) = e^{-2\pi i s t^j l'}$  and  $a_3(s') = e^{-2\pi i s' t^j l}$ , using the change of variable  $u \rightarrow -u'$ , we obtain

$$\hat{V}_N^{j,j'}(\tau) = \sum_{l=1}^{n_j} \sum_{l'=1}^{n_{j'}} \sum_{u=-N}^N \sum_{u'=-N}^N c(u-u') e^{2\pi i u(\tau-t_l^j)} e^{-2\pi i u'(\tau-t_{l'}^{j'})} \Delta(X_l^j) \Delta(X_{l'}^{j'}).$$

Then, for  $x \in \mathbb{C}^d$

$$\begin{aligned}
 &\sum_{j,j'} \hat{V}_N^{j,j'}(\tau) x_j \overline{x_{j'}} \\
 &= \sum_{u=-N}^N \sum_{u'=-N}^N \left( \sum_{j=1}^d x_j \sum_{l=1}^{n_j} e^{2\pi i u(\tau-t_l^j)} \Delta X_l^j \right) \left( \sum_{j'=1}^d x_{j'} \sum_{l'=1}^{n_{j'}} e^{-2\pi i u'(\tau-t_{l'}^{j'})} \Delta X_{l'}^{j'} \right) \\
 &= \sum_{u=-N}^N \sum_{u'=-N}^N f(u) \overline{f(u')} \geq 0.
 \end{aligned}$$

with  $f(u) =: \sum_{j=1}^d x_j \sum_{l=1}^{n_j} e^{2\pi i u(\tau-t_l^j)} \Delta X_l^j$ . The proof is complete.  $\square$

## Appendix B. Proof of Theorem 2

For simplicity let  $j = 1, j' = 2$ . By introducing the notation

$$\varphi_n^j(s) = \sum_{k=0}^{n_j-1} t_k^j 1_{[t_k^j, t_{k+1}^j)}(s), \quad s \in [0, 1),$$



we can rewrite  $\hat{V}_{N,M}^{1,2}$

$$\begin{aligned}\hat{V}_{N,M}^{1,2}(t) &= \frac{1}{2N+1} \int_{\mathbf{R}} \int_0^1 D_N(t - \varphi_n^1(s) + y) dX_s^1 \int_0^1 D_N(t - \varphi_n^2(u) + y) dX_u^2 \mu_M(dy) \\ &= \frac{1}{2N+1} \int_{\mathbf{R}} \int_0^1 D_N(t - \varphi_n^1(s) + y) D_N(t - \varphi_n^2(s) + y) V^{1,2}(s) ds \mu_M(dy) \\ &\quad + \frac{1}{2N+1} \int_{\mathbf{R}} \int_0^1 \int_0^s D_N(t - \varphi_n^1(u) + y) D_N(t - \varphi_n^2(s) + y) dX_u^1 dX_s^2 \mu_M(dy) \\ &\quad + \frac{1}{2N+1} \int_{\mathbf{R}} \int_0^1 \int_0^s D_N(t - \varphi_n^2(u) + y) D_N(t - \varphi_n^1(s) + y) dX_u^2 dX_s^1 \mu_M(dy).\end{aligned}$$

We put

$$I(t) := \frac{1}{2N+1} \int_{\mathbf{R}} \int_0^1 \left( D_N(t - \varphi_n^1(s) + y) D_N(t - \varphi_n^2(s) + y) - D_N(t - s + y)^2 \right) V^{1,2}(s) ds \mu_M(dy)$$

$$\begin{aligned}II(t) &:= \frac{1}{2N+1} \int_{\mathbf{R}} \int_0^1 \int_0^s D_N(t - \varphi_n^1(u) + y) D_N(t - \varphi_n^2(s) + y) dX_u^1 dX_s^2 \mu_M(dy) \\ &\quad + \frac{1}{2N+1} \int_{\mathbf{R}} \int_0^1 \int_0^s D_N(t - \varphi_n^1(s) + y) D_N(t - \varphi_n^2(u) + y) dX_u^2 dX_s^1 \mu_M(dy)\end{aligned}$$

and

$$\begin{aligned}III(t) &:= \frac{1}{2N+1} \int_0^1 \int_{\mathbf{R}} D_N(t - s + y)^2 \mu_M(dy) V^{1,2}(s) ds - V^{1,2}(t) \\ &= \int_{\mathbf{R}} \int_0^1 F_{2N}(t - s + y) (V^{1,2}(s) - V^{1,2}(t)) ds \mu_M(dy),\end{aligned}$$

where  $F_{2N}$  is the Fejér kernel defined in Remark 1. Then,

$$V^{1,2}(t) - \hat{V}_N^{1,2}(t) = I(t) + II(t) + III(t).$$

The following  $L^2$ -estimates of  $I(t)$ ,  $II(t)$ , and  $III(t)$  are true for any  $\mu_M$  so far as  $0 < c_M(k) < 1$ , which is true for the Gaussian case.

**Lemma A1.** *We have*

$$\mathbf{E} \int_0^1 (I(t))^2 dt \leq \pi^2 \|V\|_{\infty}^2 \rho_n^2 N^2 \sum_{|k| \leq 2N} |c_M(k)|^2, \quad (\text{A1})$$

and, in the synchronous and regular case, when  $\varphi_n^1 \equiv \varphi_n^2$ ,

$$\mathbf{E} \int_0^1 (I(t))^2 dt \leq \pi^2 \|V\|_{\infty}^2 \rho_n^2 \sum_{|k| \leq 2N} |c_M(k)|^2 k^2. \quad (\text{A2})$$

**Proof.** Since

$$\begin{aligned}
 \int_0^1 (I(t))^2 dt &= \frac{1}{(2N+1)^2} \int_{[0,1]^3} dt ds du \int_{\mathbf{R}} \mu_M(dy) \int_{\mathbf{R}} \mu_M(dy') V^{1,2}(s) V^{1,2}(u) \\
 &\times (D_N(t - \varphi_n^1(s) + y) D_N(t - \varphi_n^2(s) + y) - D_N(t - s + y)^2) \\
 &\times (D_N(t - \varphi_n^1(u) + y') D_N(t - \varphi_n^2(u) + y') - D_N(t - u + y')^2) \\
 &= \frac{1}{(2N+1)^2} \sum_{-N \leq k_1, k_2, k_3, k_4 \leq N} \int_{\mathbf{R}} e^{2\pi i(k_1+k_2)y} \mu_M(dy) \int_{\mathbf{R}} e^{2\pi i(k_3+k_4)y'} \mu_M(dy') \\
 &\times \int_0^1 e^{2\pi i(k_1+k_2+k_3+k_4)t} dt \int_0^1 (e^{-2\pi i k_1 \varphi_n^1(s) - 2\pi i k_2 \varphi_n^2(s)} - e^{-2\pi i(k_1+k_2)s}) V^{1,2}(s) ds \\
 &\times \int_0^1 (e^{-2\pi i k_3 \varphi_n^1(u) - 2\pi i k_4 \varphi_n^2(u)} - e^{-2\pi i(k_3+k_4)u}) V^{1,2}(u) du,
 \end{aligned}$$

we obtain (A1) once we establish

$$\mathbf{E} \int_0^1 (e^{-2\pi i k_1 \varphi_n^1(s) - 2\pi i k_2 \varphi_n^2(s)} - e^{-2\pi i(k_1+k_2)s}) V^{1,2}(s) ds \leq \pi \|V\|_{\infty} (|k_1| + |k_2|) \rho_n. \quad (\text{A3})$$

To prove Equation (A3), we first observe that

$$\begin{aligned}
 &| \int_0^1 (e^{-2\pi i k_1 \varphi_n^1(s) - 2\pi i k_2 \varphi_n^2(s)} - e^{-2\pi i(k_1+k_2)s}) V^{1,2}(s) ds | \\
 &\leq \sup_{t \in [0,1]} |V^{1,2}(t)| \int_0^1 |1 - e^{2\pi i(k_1(s - \varphi_n^1(s)) + k_2(s - \varphi_n^2(s)))}| ds,
 \end{aligned}$$

and

$$\begin{aligned}
 &\int_0^1 |1 - e^{2\pi i(k_1(s - \varphi_n^1(s)) + k_2(s - \varphi_n^2(s)))}| ds \\
 &\leq 2\pi \int_0^1 |k_1(s - \varphi_n^1(s)) + k_2(s - \varphi_n^2(s))| ds \\
 &\leq 2\pi |k_1| \int_0^1 |s - \varphi_n^1(s)| ds + 2\pi |k_2| \int_0^1 |s - \varphi_n^2(s)| ds.
 \end{aligned}$$

Then, (A3) holds since

$$\begin{aligned}
 \int_0^1 |s - \varphi_n^j(s)| ds &= \sum_{k=0}^{n_j-1} \int_{t_k^j}^{t_{k+1}^j} (s - t_k^j) ds = \frac{1}{2} \sum_{k=0}^{n_j-1} (t_{k+1}^j - t_k^j)^2 \\
 &\leq \frac{\rho(n_j)}{2} \sum_{k=0}^{n_j-1} (t_{k+1}^j - t_k^j) \leq \frac{\rho_n}{2}
 \end{aligned} \quad (\text{A4})$$

for  $j = 1, 2$ .

For (A2), we just need

$$\begin{aligned}
 &\int_0^1 |1 - e^{2\pi i(k_1(s - \varphi_n^1(s)) + k_2(s - \varphi_n^2(s)))}| ds \\
 &\leq 2\pi |k_1 + k_2| \int_0^1 |s - \varphi_n^1(s)| ds.
 \end{aligned}$$

□

**Lemma A2.** For the general case, it holds

$$\mathbf{E} \int_0^1 (II(t))^2 dt \leq (4C_{\nabla} + 2\|V\|_{\infty}^2) \left( 4\pi^2 \rho_n^2 N^2 + (2N+1)^{-1} \right) \sum_{|k| \leq 2N} c_M(k)^2, \quad (\text{A5})$$

and, when  $t_k^j = k/n$  for  $k = 0, 1, \dots, n, j = 1, 2$ ,

$$\int_0^1 \mathbf{E}(II(t))^2 dt \leq \frac{4C_{\nabla} + 2\|V\|_{\infty}^2}{2N+1} \sum_{|k| \leq N} c_M(k)^2. \quad (\text{A6})$$

**Proof.** We first show that

$$\mathbf{E}[(II(t))^2] \leq \frac{4C_{\nabla} + 2\|V\|_{\infty}^2}{(2N+1)^2} \int_{[0,1]^2} (G(s,u))^2 ds du, \quad (\text{A7})$$

where

$$G(s,u) \equiv G^{1,2}(s,u) := \int_{\mathbf{R}} \mu_M(dy) D_N(t - \varphi_n^1(s) + y) D_N(t - \varphi_n^2(u) + y).$$

Let

$$A^{j,j'} := \int_{\mathbf{R}} \int_0^1 \int_0^s D_N(t - \varphi_n^j(u) + y) D_N(t - \varphi_n^{j'}(s) + y) dX_u^j dX_s^{j'} \mu_M(dy)$$

for  $j, j' \in \{1, 2\}$ , so that

$$|II(t)|^2 = \frac{1}{(2N+1)^2} |A^{1,2} + A^{2,1}|^2 \leq \frac{2}{(2N+1)^2} (|A^{1,2}|^2 + |A^{2,1}|^2).$$

Then, we have

$$\begin{aligned} & \mathbf{E}[|A^{j,j'}|^2] \\ &= \int_{\mathbf{R}^2} \mu_M^{\otimes 2}(dy dy') \int_0^1 (D_N(t - \varphi_n^j(s) + y) D_N(t - \varphi_n^{j'}(s) + y')) \\ & \quad \times \mathbf{E} \left[ V^{j,j}(s) \int_0^s D_N(t - \varphi_n^{j'}(u) + y) dX_u^{j'} \int_0^s D_N(t - \varphi_n^{j'}(u) + y') dX_u^{j'} \right] ds \\ &= \int_0^1 \int_0^s (G^{j,j'}(s,u))^2 \mathbf{E}[V^{j,j}(s) V^{j',j'}(u)] du ds \\ &+ \int_{\mathbf{R}^2} \mu_M^{\otimes 2}(dy dy') \left( \int_0^1 (D_N(t - \varphi_n^j(s) + y) D_N(t - \varphi_n^{j'}(s) + y')) \right. \\ & \quad \times \mathbf{E} \left[ V^{j,j}(s) \int_0^s \int_0^u \left( D_N(t - \varphi_n^{j'}(u) + y) D_N(t - \varphi_n^{j'}(v) + y') \right. \right. \\ & \quad \left. \left. + D_N(t - \varphi_n^{j'}(u) + y') D_N(t - \varphi_n^{j'}(v) + y) \right) dX_v^{j'} dX_u^{j'} \right] ds \Big). \end{aligned}$$

By the Malliavin integration by parts formula,

$$\begin{aligned} & \mathbf{E}[V^{j,j}(s) \int_0^s \int_0^u D_N(t - \varphi_n^{j'}(u) + y) D_N(t - \varphi_n^{j'}(v) + y') dX_v^{j'} dX_u^{j'}] \\ &= \int_0^s D_N(t - \varphi_n^{j'}(u) + y) \mathbf{E}[\sigma^{j'}(u) \nabla_s V^{j,j}(u) \int_0^u D_N(t - \varphi_n^{j'}(v) + y') dX_v^{j'}] du \\ &= \int_0^s D_N(t - \varphi_n^{j'}(u) + y) \int_0^u \mathbf{E}[\sigma^{j'}(v) \nabla_v (\sigma^{j'}(u) \nabla_s V^{j,j}(u))] D_N(t - \varphi_n^{j'}(v) + y') dv du. \end{aligned}$$

Then applying Malliavin integration by parts formula again, we see that

$$\begin{aligned}
 & \int_0^1 ds \int_{\mathbf{R}^2} \mu_M^{\otimes 2}(dydy') (D_N(t - \varphi_n^j(s) + y) D_N(t - \varphi_n^j(s) + y')) \\
 & \mathbf{E}[V^{jj}(s) \int_0^s \int_0^u (D_N(t - \varphi_n^{j'}(u) + y) D_N(t - \varphi_n^{j'}(v) + y') \\
 & \quad + D_N(t - \varphi_n^{j'}(u) + y') D_N(t - \varphi_n^{j'}(v) + y)) dX_v^{j'} dX_u^{j'}] \\
 & = 2 \int_0^1 \int_0^s \int_0^u G^{jj'}(s, u) G^{jj'}(s, v) \mathbf{E}[\sigma^{j'}(v) \nabla_v(\sigma^{j'}(u) \nabla_s V^{jj}(u))] dv du ds \\
 & \leq 2C_{\nabla} \int_0^1 \int_0^s \int_0^u |G^{jj'}(s, u) G^{jj'}(s, v)| dv du ds \\
 & \leq C_{\nabla} \int_{[0,1]^2} (G^{jj'}(s, u))^2 du ds.
 \end{aligned}$$

Thus we obtain (A7).

We proceed to prove (A5) and (A6). Observe that

$$\begin{aligned}
 & \int_{[0,2\pi]^3} (G(s, u))^2 du ds dt \\
 & = \int_{[0,2\pi]^2} ds du \int_{\mathbf{R}^2} \mu_M^{\otimes 2}(dydy') \\
 & \quad \sum_{-N \leq k_1, k_2, k_3, k_4 \leq N} e^{2\pi i(k_1+k_2+k_3+k_4)t} e^{-2\pi i(k_1+k_2)\varphi_n^1(s)} e^{-2\pi i(k_3+k_4)\varphi_n^2(u)} e^{2\pi i(k_1+k_3)y+2\pi i(k_2+k_4)y'} \\
 & = \sum_{\substack{-N \leq k_1, k_2, k_3, k_4 \leq N \\ k_1+k_2+k_3+k_4=0}} c_M(k_1+k_3) c_M(k_2+k_4) \int_0^1 e^{-2\pi i(k_1+k_2)(\varphi_n^1(s)-s)} ds \int_0^1 e^{-2\pi i(k_3+k_4)(\varphi_n^2(s)-s)} ds.
 \end{aligned} \tag{A8}$$

When  $t_k \equiv k/n$ , we have

$$\int_0^1 e^{2\pi i k \varphi(s)} ds = \frac{1}{n} \sum_{l=0}^{n-1} e^{\frac{2\pi i l k}{n}} = 1_{\{k=0\}},$$

hence we obtain (A6). For the general case, we have

$$\begin{aligned}
 \int_0^{2\pi} e^{2\pi i k \varphi(s)} ds & = \int_0^1 (e^{2\pi i k \varphi(s)} - e^{2\pi i k s}) ds + \int_0^1 e^{2\pi i k s} ds \\
 & = \int_0^1 (e^{2\pi i k \varphi(s)} - e^{2\pi i k s}) ds + 1_{\{k=0\}}.
 \end{aligned}$$

By (A4), we obtain (A5).  $\square$

**Lemma A3.**

$$\mathbb{E} \int_0^1 (III(t))^2 dt \leq 2C_{\kappa} \left( (2N)^{-2\kappa} + \sup_{0 < |k| \leq 2N} \left( \frac{1 - c_M(k)}{|k|^2} \right)^{\kappa} \right). \tag{A9}$$

**Proof.** We first note that

$$\begin{aligned}
 & \frac{1}{2\pi} \int_0^{2\pi} (III(t))^2 dt \\
 & = \sum_{|k| > 2N} |(\mathcal{F}V)(k)|^2 + \sum_{-2N \leq k \leq 2N} \left( 1 - \left( 1 - \frac{|k|}{2N+1} \right) c_M(k) \right)^2 |(\mathcal{F}V)(k)|^2,
 \end{aligned}$$

and

$$\sum_{|k|>2N} |(\mathcal{F}V)(k)|^2 \leq (2N)^{-\kappa} \sum_{|k|>2N} |k|^{2\kappa} |(\mathcal{F}V)(k)|^2. \quad (\text{A10})$$

On the other hand, since  $0 < c_M(k) < 1$ ,

$$0 < 1 - \left(1 - \frac{|k|}{2N+1}\right) c_M(k) < 1,$$

we have

$$\begin{aligned} \left(1 - \left(1 - \frac{|k|}{2N+1}\right) c_M(k)\right)^2 &\leq 2(1 - c_M(k))^2 + 2\left(\frac{|k|}{2N+1} c_M(k)\right)^2 \\ &\leq 2(1 - c_M(k))^\kappa + 2\left(\frac{|k| c_M(k)}{2N+1}\right)^{2\kappa}, \end{aligned}$$

and therefore we have

$$\begin{aligned} &\sum_{-2N \leq k \leq 2N} \left(1 - \left(1 - \frac{|k|}{2N+1}\right) c_M(k)\right)^2 |(\mathcal{F}V)(k)|^2 \\ &\leq 2 \left( \sup_{0 < |k| \leq 2N} \left(\frac{1 - c_M(k)}{|k|^2}\right)^\kappa + (2N+1)^{-2\kappa} \right) \sum_{-2N \leq k \leq 2N} |k|^{2\kappa} |(\mathcal{F}V)(k)|^2. \end{aligned} \quad (\text{A11})$$

Combining (A10) and (A11), we get (A9) by the assumption (9).  $\square$

Now we are ready to prove Theorem 2

**Proof of Theorem 2.** First we prove (i) and (ii). We now set

$$c_M(k) = e^{-\frac{2\pi^2 k^2}{M}}.$$

Then,

$$\sum_{|k| \leq 2N} |c_M(k)|^2 \leq 2 \sum_{k=1}^{2N} \int_{k-1}^k e^{-\frac{2\pi^2 x^2}{M}} dx = \int_{\mathbf{R}} e^{-\frac{2\pi^2 x^2}{M}} dx = \sqrt{\frac{M}{2\pi}}$$

and

$$\begin{aligned} \sum_{|k| \leq 2N} |c_M(k)|^2 k^2 &\leq 2 \sum_{l=0}^{2N} \int_l^{l+1} (x+1)^2 e^{-\frac{2\pi^2 x^2}{M}} dx \\ &= \int_{\mathbf{R}} (x+1)^2 e^{-\frac{2\pi^2 x^2}{M}} dx = \sqrt{\frac{M}{2\pi}} \left( \frac{M}{4\pi^2} + 1 \right). \end{aligned}$$

Further, we have

$$\frac{1 - c_M(k)}{|k|^2} = \frac{1 - e^{-\frac{2\pi^2 k^2}{M}}}{|k|^2} \leq \frac{2\pi^2}{M}$$

so that

$$\sup_{0 < |k| \leq 2N} \left( \frac{1 - c_M(k)}{|k|^2} \right)^\kappa \leq \left( \frac{2\pi^2}{M} \right)^\kappa.$$

We prove now (iii) and (iv). If  $N \asymp \rho_n^{-\alpha}$  and  $M \asymp \rho_n^{-\beta}$ , we have

$$\rho_n^2 N^2 \sqrt{\frac{M}{2\pi}} \asymp \rho_n^{2-2\alpha-\frac{\beta}{2}},$$

$$(2N+1)^{-1} \sqrt{\frac{M}{2\pi}} \asymp \rho_n^{\alpha-\frac{\beta}{2}},$$

$$2C_\kappa(2N)^{-2\kappa} \asymp \rho_n^{2\kappa\alpha}. \quad 2C_\kappa\left(\frac{2\pi^2}{M}\right)^\kappa \asymp \rho_n^{\kappa\beta},$$

and

$$\rho_n^2 \sqrt{\frac{M}{2\pi}} \left( \frac{M}{4} + \pi^2 \right) \asymp n^{2-\frac{3\beta}{2}}.$$

Finally, in order to attain the consistency of the proposed estimator under the general sampling scheme, we need to assume

$$2 - 2\alpha - \frac{\beta}{2} > 0, \alpha > \frac{\beta}{2}, \alpha > 0, \beta > 0,$$

which is equivalent to the condition (12). In such a case  $\kappa\beta < 2\kappa\alpha$ .

When the sampling is synchronous and regularly spaced, the necessary condition for consistency clearly becomes (13).  $\square$

## Appendix C. Additional Results of Comparison for Alternative Models

**Table A1.** The % of psd matrix produced by each estimator, when the efficient price process is produced by alternative models, in presence of i.i.d. noise.

Estimator	SV1F	SV2F	RH	SV1F	SV2F	RH	SV1F	SV2F	RH	SV1F	SV2F	RH
d = 5, $\sigma_\eta = 1$				d = 5, $\sigma_\eta = 1.5$			d = 5, $\sigma_\eta = 2$			d = 5, $\sigma_\eta = 2.5$		
PDF	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%
LMM	100%	100%	100%	100%	100%	100%	100%	99.92%	99.53%	97.12%	94.18%	95.06%
STS	100%	100%	100%	100%	100%	100%	100%	99.76%	99.62%	98.53%	96.64%	98.01%
d = 10, $\sigma_\eta = 1$				d = 10, $\sigma_\eta = 1.5$			d = 10, $\sigma_\eta = 2$			d = 10, $\sigma_\eta = 2.5$		
PDF	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%
LMM	100%	100%	100%	99.84%	100%	100%	99.58%	99.13%	95.33%	90.27%	86.01%	86.83%
STS	99.68%	99.93%	99.95%	98.82%	98.71%	98.55%	87.86%	89.35%	88.16%	55.23%	49.79%	59.43%
d = 15, $\sigma_\eta = 1$				d = 15, $\sigma_\eta = 1.5$			d = 15, $\sigma_\eta = 2$			d = 15, $\sigma_\eta = 2.5$		
PDF	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%
LMM	100%	100%	99.98%	99.97%	99.74%	99.63%	99.15%	98.02%	98.31%	78.15%	10.71%	76.37%
STS	96.69%	99.35%	96.93%	76.07%	72.45%	75.91%	34.70%	29.01%	28.72%	6.95%	7.44%	8.26%
d = 20, $\sigma_\eta = 1$				d = 20, $\sigma_\eta = 1.5$			d = 20, $\sigma_\eta = 2$			d = 20, $\sigma_\eta = 2.5$		
PDF	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%
LMM	100%	100%	99.99%	99.94%	98.38%	99.07%	97.21%	95.80%	95.99%	66.13%	62.74%	64.50%
STS	64.78%	67.24%	65.87%	12.89%	12.45%	16.93%	4.33%	2.47%	0.98%	0.0%	0.0%	0.0%

**Table A2.** The % of psd matrix produced by each estimator, when the efficient price process is produced by alternative models, in presence of general noise.

Estimator	SV1F	SV2F	RH	SV1F	SV2F	RH	SV1F	SV2F	RH	SV1F	SV2F	RH
d = 5, g = 0.3, w = 0.3			d = 5, g = 0.3, w = 0.9			d = 5, g = 0.45, w = 0.3			d = 5, g = 0.45, w = 0.9			
PDF	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%
LMM	100%	99.98%	99.84%	98.43%	98.21%	98.02%	93.21%	90.42%	90.16%	99.86%	98.41%	97.62%
STS	99.01%	99.73%	99.61%	98.23%	98.11%	99.04%	99.16%	99.73%	99.88%	98.90%	98.75%	99.00%
d = 10, g = 0.3, w = 0.3			d = 10, g = 0.3, w = 0.9			d = 10, g = 0.45, w = 0.3			d = 10, g = 0.45, w = 0.9			
PDF	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%
LMM	100%	99.56%	99.58%	95.86%	95.92%	96.02%	79.20%	80.11%	81.56%	91.25%	89.37%	90.84%
STS	92.80%	93.04%	93.52%	79.23%	73.21%	80.01%	90.93%	89.15%	89.37%	76.41%	74.86%	77.75%
d = 15, g = 0.3, w = 0.3			d = 15, g = 0.3, w = 0.9			d = 15, g = 0.45, w = 0.3			d = 15, g = 0.45, w = 0.9			
PDF	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%
LMM	100%	98.12%	99.03%	85.48%	88.23%	90.03%	68.21%	67.39%	68.98%	80.16%	77.56%	81.43%
STS	51.67%	45.47%	50.93%	45.88%	43.92%	45.78%	39.66%	35.31%	37.40%	79.65%	38.04%	39.61%
d = 20, g = 0.3, w = 0.3			d = 20, g = 0.3, w = 0.9			d = 20, g = 0.45, w = 0.3			d = 20, g = 0.45, w = 0.9			
PDF	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%
LMM	99.98%	96.15%	98.78%	84.38%	81.51%	85.00%	53.11%	52.05%	55.15%	75.82%	74.83%	72.68%
STS	8.99%	7.15%	8.01%	21.67%	21.42%	12.89%	4.57%	3.66%	3.01%	75.43%	14.38%	15.02%

## Notes

- Hereafter, we accept abuse of notation by denoting the estimator  $\hat{V}_{n_j, n_j', N, M}^{ij'}(t)$  as  $\hat{V}_{n, N, M}^{ij'}(t)$ .
- Here the notation  $\hat{V}_{N, M}^{ij'}(t)$  highlights the dependence on the two parameters  $N, M$ .
- The fact that the drift does not contribute to the asymptotics can be proved analogously as in [Malliavin and Mancino \(2009\)](#).
- Here  $a_n \asymp b_n$  means both  $\limsup_{n \rightarrow \infty} a_n / b_n$  and  $\limsup_{n \rightarrow \infty} b_n / a_n$  are finite.
- Since all the simulations are conducted under irregularly-spaced and asynchronous observations, we follow point (iii) of Theorem 2. Moreover, for the the Heston, the SVF1 and the SVF2 models the Hölder parameter is  $\kappa = \frac{1}{2}$ , while the Rough Heston model it depends on the chosen Hurst exponent.
- Note that, according with Theorem 1, the semi-definite positiveness of the proposed estimator is granted when the optimal cutting frequency  $N$  is the same for each spot volatility-covariance entries estimates.
- Computations are performed using a machine with 2.30 GHz clock speed Intel i7-11800H, 16 GB RAM on Arch Linux kernel 6.14.10-arch1-1 with MATLAB R2024a update 6.

## References

- Aït-Sahalia, Yacine, and Dacheng Xiu. 2019. Principal component analysis of high-frequency data. *Journal of the American Statistical Association* 114: 287–303.
- Aït-Sahalia, Yacine, and Jean Jacod. 2014. *High-Frequency Financial Econometrics*. Princeton: Princeton University Press.
- Alòs, Elisa, and Christian-Oliver Ewald. 2008. Malliavin differentiability of the heston volatility and applications to option pricing. *Advances in Applied Probability* 40: 144–62. [\[CrossRef\]](#)
- Bandi, Federico M., and Jeffrey R. Russell. 2008. Microstructure noise, realized variance, and optimal sampling. *The Review of Economic Studies* 75: 339–69. [\[CrossRef\]](#)
- Barndorff-Nielsen, Ole E., and Neil Shephard. 2004. Econometric analysis of realized covariation: High frequency-based covariance, regression, and correlation in financial economics. *Econometrica* 72: 885–925. [\[CrossRef\]](#)
- Barndorff-Nielsen, Ole E., Peter Reinhard Hansen, Asger Lunde, and Neil Shephard. 2011. Multivariate realised kernels: Consistent positive semi-definite estimators of the covariation of equity prices with noise and non-synchronous trading. *Journal of Econometrics* 162: 149–69. [\[CrossRef\]](#)
- Bibinger, Markus, Nikolaus Hautsch, Peter Malec, and Markus Reiss. 2019. Estimating the spot covariation of asset prices—Statistical theory and empirical evidence. *Journal of Business & Economic Statistics* 37: 419–35.
- Bu, Ruijun, Degui Li, Oliver Linton, and Hanchao Wang. 2023. Nonparametric estimation of large spot volatility matrices for high-frequency financial data. *arXiv arXiv:2307.01348*. [\[CrossRef\]](#)



- Chen, Dachuan, Per A. Mykland, and Lan Zhang. 2020. The five trolls under the bridge: Principal component analysis with asynchronous and noisy high frequency data. *Journal of the American Statistical Association* 115: 1960–77. [\[CrossRef\]](#)
- Chernov, Mikhail, A. Ronald Gallant, Eric Ghysels, and George Tauchen. 2003. Alternative models for stock price dynamics. *Journal of Econometrics* 116: 225–57. [\[CrossRef\]](#)
- Clement, Emmanuelle, and Arnaud Gloter. 2011. Limit theorems in the Fourier transform method for the estimation of multivariate volatility. *Stochastic Processes and Their Applications* 121: 1097–124. [\[CrossRef\]](#)
- Cui, Liyuan, Yongmiao Hong, Yingxing Li, and Junhui Wang. 2024. Regularized High-Dimensional Positive Definite Covariance Estimator with High-Frequency Data. *Management Science* 70: 7242–7264. [\[CrossRef\]](#)
- El Euch, Omar, and Mathieu Rosenbaum. 2019. The characteristic function of rough Heston models. *Mathematical Finance* 29: 3–38. [\[CrossRef\]](#)
- Engle, Robert, and Riccardo Colacito. 2006. Testing and valuing dynamic correlations for asset allocation. *Journal of Business & Economic Statistics* 24: 238–53. [\[CrossRef\]](#)
- Figuerola-López, José E., and Bei Wu. 2022. Kernel estimation of spot volatility with microstructure noise using pre-averaging. *Econometric Theory* 40: 558–607. [\[CrossRef\]](#)
- Gatheral, Jim, Thibault Jaisson, and Mathieu Rosenbaum. 2018. Volatility is rough. *Quantitative Finance* 18: 933–49. [\[CrossRef\]](#)
- Hansen, Peter R., and Asger Lunde. 2006. Realized variance and market microstructure noise. *Journal of Business & Economic Statistics* 24: 127–161. [\[CrossRef\]](#)
- Hasbrouck, Joel. 2007. *Empirical Market Microstructure: The Institutions, Economics, and Econometrics of Securities Trading*. Oxford: Oxford University Press.
- Heston, Steven L. 1993. A closed-form solution for options with stochastic volatility with applications to bond and currency options. *The Review of Financial Studies* 6: 327–43. [\[CrossRef\]](#)
- Huang, Xin, and George Tauchen. 2005. The relative contribution of jumps to total price variance. *Journal of Financial Econometrics* 3: 456–99. [\[CrossRef\]](#)
- Inahama, Yuzuru. 2014. Malliavin differentiability of solutions of rough differential equations. *Journal of Functional Analysis* 267: 1566–84. [\[CrossRef\]](#)
- Jacod, Jean, Yingying Li, and Xinghua Zheng. 2017. Statistical properties of microstructure noise. *Econometrica* 85: 1133–74. [\[CrossRef\]](#)
- Jacod, Jean, Yingying Li, and Xinghua Zheng. 2019. Estimating the integrated volatility with tick observations. *Journal of Econometrics* 208: 80–100. [\[CrossRef\]](#)
- Liu, Nien-Lin, and Hoang-Long Ngo. 2017. Approximation of eigenvalues of spot cross volatility matrix with a view toward principal component analysis. *Japan Journal of Industrial and Applied Mathematics* 34: 747–61. [\[CrossRef\]](#)
- Malliavin, Paul, and Maria Elvira Mancino. 2009. A Fourier transform method for nonparametric estimation of multivariate volatility. *The Annals of Statistics* 37: 1983–2010. [\[CrossRef\]](#)
- Mancino, Maria Elvira, Maria Cristina Recchioni, and Simona Sanfelici. 2017. *Fourier-Malliavin Volatility Estimation: Theory and Practice*. Berlin: Springer International Publishing.
- Mancino, Maria Elvira, Tommaso Mariotti, and Giacomo Toscano. 2024. Asymptotic normality and finite-sample robustness of the Fourier spot volatility estimator in the presence of microstructure noise. *Journal of Business & Economic Statistics* 1–23. [\[CrossRef\]](#)
- Mancino, Maria Elvira, Tommaso Mariotti, and Giacomo Toscano. 2025. Spot beta estimation with asynchronous noisy prices. *Quantitative Finance* 25: 733–55. [\[CrossRef\]](#)
- Mykland, Per A., Lan Zhang, and Dachuan Chen. 2019. The algebra of two scales estimation, and the s-tsrv: High frequency estimation that is robust to sampling times. *Journal of Econometrics* 208: 101–19. [\[CrossRef\]](#)
- Park, Sujin, Seok Young Hong, and Oliver Linton. 2016. Estimating the quadratic covariation matrix for asynchronously observed high frequency stock returns corrupted by additive measurement error. *Journal of Econometrics* 191: 325–47. [\[CrossRef\]](#)
- Richard, Alexandre, Xiaolu Tan, and Fan Yang. 2023. On the discrete-time simulation of the rough heston model. *SIAM Journal on Financial Mathematics* 14: 223–49. [\[CrossRef\]](#)
- Roll, Richard. 1984. A simple implicit measure of the effective bid-ask spread in an efficient market. *The Journal of Finance* 39: 1127–39. [\[CrossRef\]](#)
- Shepard, Neil, and Dacheng Xiu. 2017. Econometric analysis of multivariate realised qml: Estimation of the covariation of equity prices under asynchronous trading. *Journal of Econometrics* 201: 19–42. [\[CrossRef\]](#)
- Zu, Yang and H., Peter Boswijk. 2014. Estimating spot volatility with high-frequency financial data. *Journal of Econometrics* 181: 117–35. [\[CrossRef\]](#)

**Disclaimer/Publisher’s Note:** The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.