Article

Not peer-reviewed version

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

Jiro Akahori , Reika Kambara , Nien-Lin Liu , Maria Elvira Mancino * , Tommaso Mariotti * , Yukie Yasuda

Posted Date: 9 May 2025

doi: 10.20944/preprints202505.0726.v1

Keywords: Nonparametric covariance estimator; risk management, factor analysis; Fourier analysis

Article

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

Jirô Akahori [1], Reika Kambara [1], Nien-Lin Liu [2], Maria Elvira Mancino [3,*] Tommaso Mariotti [4,*] and Yukie Yasuda [1]

[1] Ritsumeikan University, Department of Mathematical Sciences, Kusatsu, Japan; akahori@se.ritsumei.ac.jp (J.K.)
[2] Tokyo University of Science, Department of Business Economics, Tokyo, Japan
[3] University of Florence, Dept. of Economics and Management, Via delle Pandette 9, 50127 Firenze, Italy
[4] University of Turin, Dept. of Economics, Social Studies, Applied Mathematics and Statistics, Corso Unione Sovietica 218/bis, 10134 Torino, Italy
* Correspondence: mariaelvira.mancino@unifi.it (M.E.M.); tommaso.mariotti@unito.it (T.M.)

Abstract: In this paper we propose a nonparametric estimator of the spot volatility matrix with high-frequency data. The newly proposed Positive Definite Fourier (PDF) estimator is proved to produce symmetric positive semi-definite estimates and to be consistent with a suitable choice of the localizing kernel. The PDF estimator relies on a modification of the Fourier estimation method introduced by [22]. The estimator relies on two parameters: the frequency N, which is responsible for controlling both the biases due to the asynchronicity effect and the market microstructure noise effect, and the localization parameter M of the employed Gaussian kernel. The sensitivity of the estimator to the choice of the two parameters is studied in a simulated environment. The accuracy and the ability of the estimator to produce positive semi-definite covariance matrices is evaluated with an extensive numerical study, in comparison with the competing estimators present in the literature. The results of the simulation study are confirmed under many scenarios, that consider the dimensionality of the problem, the asynchronicity of data and the presence of several specifications of market microstructure noise.

Keywords: nonparametric covariance estimator; risk management; factor analysis; fourier analysis

---

1. Introduction

Empirical studies have pointed out the importance of considering distinct time variations in correlations between asset prices. Then, in the last years, several studies have addressed the issue of efficiently estimating covariances using high frequency data asynchronously sampled across different assets. While the literature is becoming rich as it concerns the estimation of integrated covariances, it is still sparse for the spot covariances estimation. An early proposal to cope with spot covariances estimation with asynchronous high frequency data, has been given in Malliavin and Mancino [22]. In contrast with the other estimators which rely on a pre-processing of data in order to make them synchronous, such as linear interpolation, piecewise constant (previous-tick) interpolation or the refresh-time procedure proposed by Barndorff-Nielsen et al. [5], the Fourier estimator uses all the available data, being based on an integration procedure. The possibility of using all data, avoiding any preliminary manipulation of them (such as pre-averaging, see, e.g., Aït-Sahalia and Jacod [1]), translates into the direct use of unevenly sampled returns and even asynchronous data in the multivariate case.

A substantial property for an estimator of integrated or spot covariances relies in the positive semi-definiteness of the estimated covariance matrix. This property has important consequences in several contexts, such as the recently developed field of principal component analysis with high-frequency data (Liu and Ngo [21], Aït-Sahalia and Xiu [2], Chen et al. [8]) or the asset allocation framework (see, e.g., Engle and Colacito [12]). While this point has been addressed by some authors for the integrated covariances estimators (see, e.g., Barndorff-Nielsen et al. [5], Mancino and Sanfelici [28] Park et al. [30], Cui et al. [10]), at the best of our knowledge the estimator of spot covariances proposed in the present paper is the first to guarantee positive semi-definiteness of the estimation itself, a problem that so far has not been addressed in the literature. For example, when dealing with spot volatility, Chen et al. [8] integrate the estimations before computing the eigenvalues of the covariance matrix, while in Bu et al. [7] positive semi-definiteness is imposed applying suitable shrinkage techniques to the estimation, thus introducing a manipulation of the estimated matrix.

The aim of this work is to propose a novel spot covariance estimator, prove its positivity and consistency, and analyze its finite-sample properties in a simulated environment. Our starting point is the spot Fourier estimator by Malliavin and Mancino [23]. This estimator, however, due to lack of symmetry in the Fejér kernel, may fail to provide positive semi-definite estimations when the asset prices are observed on asynchronous grids. To guarantee that the estimations are symmetric and positive semi-definite, in this paper, we introduce a modified version of the Fourier estimator, which we call the PDF estimator. In Theorem 1 we prove that it indeed fulfills the desired property, while Theorem 2 gives bounds for the asymptotic error, providing conditions on the rates of N and M with respect to the sampling frequency to ensure the consistency of the estimator.

The proposed estimator relies on two parameters: the cutting frequency N, and the localizing frequency M. The question of how to optimally choose them in order to minimize the error is assessed, 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 ρ_n is the mesh of the given sampling, and α, β 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. We find that, concerning the parameter c_M, which controls the localization Gaussian kernel, it exhibits a more stable optimal value in the scenarios considered, with only a small downward correction needed in the presence of noise. A similar behavior was also observed for the original Fourier estimator in [25]. 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.

Moreover, to evaluate the finite-sample performance of the proposed PDF estimator, we compare its accuracy and the percentage of positive semi-definite estimates that it is able to produce with the ones obtained employing the smoothed two-scale estimator by Mykland et al. [29] and the local method of moments estimator by Bibinger et al. [6], which are both able to manage asynchronous observations. Developing this comparison we focus on the main problems that may affect the estimation of variance-covariance matrices using high frequency data. First of all, we address the problem of dimensionality, evaluating the produced estimations when the number of assets increase; secondly, we focus on the level of asynchronicity, considering different intensities of the Poisson processes that drives the observation frequency; lastly, we analyze the presence of 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. It is shown that, in this exercise, the PDF estimator is the only one to consistently produce positive semi-definite estimations in 100% of the cases, as guaranteed by the theory, while maintaining a hedge with respect to the competitors in terms of mean square error.

The robustness of all the simulation results are confirmed changing the simulation model behind the analysis; in particular, we consider: an Heston Stochastic Volatility model (Heston [17]), a One Factor Volatility model and a Two Factor Volatility model (Chernov et al. [9]), and a Rough Heston model (El Euch and Rosenbaum [11]), getting in each case comparable results.

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

2. The Positive Semi-Definite Spot Covariance Estimator

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

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

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

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

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

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

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

In this setting, we propose the following estimator of spot covariance The estimator was introduced in the earlier version of the present paper Akahori et al. [3].

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

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

Remark 1. *If we take $\mathcal{K} = \{0, \pm1, \pm2, ..., \pm M\}$ for some positive integer M and $\mathcal{S}(k) = \{(s,s')|s + s' = k, s = 0, \pm1, \pm2, ..., \pm N\}$ for some positive integer N, and*

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

we obtain:

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

The estimator (2) can be expressed, using the Dirichlet and the Fejér kernels $D_N(x) = \sum_{k=-N}^{N} e^{2\pi ikx}$ and $F_M(x) = \sum_{k=-M}^{M}\left(1 - \frac{|k|}{M+1}\right)e^{2\pi ikx}$, as follows

$$\hat{V}_{\mathcal{K},\mathcal{S}}^{j,j'}(t) = \frac{1}{2N+1}\sum_{l=1}^{n_j}\sum_{l'=1}^{n_{j'}} F_M(t - t_l^j)D_N(t_l^j - t_{l'}^{j'})\Delta(X_l^j)\Delta(X_{l'}^{j'}). \tag{3}$$

Therefore, with a suitable choice of function $c(\cdot)$, the estimator (1) coincides with the Fourier spot covariance estimator introduced by Malliavin and Mancino [23]. The asymptotic properties have been studied in Malliavin and Mancino [23] (in the absence of noise) and in Mancino et al. [25] (in the presence of noise). However, while *the positivity of the Fourier estimator of the integrated covariance matrix is proved in Mancino and Sanfelici [28], the spot covariance estimator may fail in producing symmetric positive semi-definite estimations, being $F_M(t-t^j_l)D_N(t^j_l-t^{j'}_{l'})$ not symmetric in j, j', leading to complex eigenvalues in $\hat{V}_{\mathcal{K},\mathcal{S}}(t)$. In addition, simple symmetrizations such as $(\hat{V}^{j,j'}+\hat{V}^{j',j})/2$ are still not positive-definite, possibly with negative eigenvalues.*

The main theoretical result of this work concerns the positive semi-definiteness of the proposed estimator and is stated in the following theorem.

Theorem 1. *Let N and M be positive integers. Suppose that $\mathcal{K}=\{0,\pm1,\pm2,...,\pm2N\}$, $c_M(k)$ is a positive semi-definite function on $\mathcal{K}$ and*

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

Then, $\hat{V}_{\mathcal{K},\mathcal{S}}(t)$ defined in (1) 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:

$$\hat{V}^{j,j'}_{N,M}(t)=\frac{1}{2N+1}\sum_{l=1}^{n_j}\sum_{l'=1}^{n_{j'}}\sum_{u=-N}^{N}\sum_{u'=-N}^{N}c_M(u-u')e^{2\pi iu(t-t^j_l)}e^{-2\pi iu'(t-t^{j'}_{l'})}\Delta(X^j_l)\Delta(X^{j'}_{l'}), \tag{4}$$

for two asset j and j' and $t\in[0,1]$, where $c_M(k)$ is still a positive semi-definite function Here the notation $\hat{V}^{j,j'}_{N,M}(t)$ highlights the dependence on the two parameters N, M.. 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 μ_M on $\mathbf{R}$ such that

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

Therefore, we may also rewrite the PDF estimator (4) using the measure μ_M instead of the positive semi-definite function $c_M(k)$, and obtain

$$\hat{V}^{j,j'}_{N,M}(t)=\frac{1}{2N+1}\sum_{l=1}^{n_j}\sum_{l'=1}^{n_{j'}}\int_{\mathbf{R}}D_N(t-t^j_l+y)D_N(t-t^{j'}_{l'}+y)\mu_M(dy)\Delta(X^j_l)\Delta(X^{j'}_{l'}). \tag{5}$$

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

In the next Section, we prove the consistency of the estimator (4) (equivalently, (5)).

3. Asymptotic Properties of the PDF Estimator with Gaussian Kernel

In this section, we consider the case where μ_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

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

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 $N, M \to \infty$ as $\rho_n \to 0$ 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 $V^{j,j'} - \hat{V}^{j,j'}_{N,M}$ 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 The fact that the drift does not contribute to the asymptotics can be proved analogously as in Malliavin and Mancino [23].

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 $\sigma^j := (\sigma^j_1, \sigma^j_2)$, $j = 1, 2$ 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 ∇ denotes the Malliavin derivative. Further, we assume that $V^{j,j'}, j, j' = 1, 2$ are κ-*Hölder continuous* for some $\kappa \in (0, 1)$ in the sense that

$$\sum_{k\in\mathbf{Z}} |k|^{2\kappa} \mathbf{E}[|(\mathcal{F}V^{j,j'})(k)|^2] =: C_\kappa < \infty, \tag{6}$$

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

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

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

$$\begin{aligned} &\mathbf{E}\left[\int_0^1 (V^{j,j'}(t) - \hat{V}^{j,j'}_{N,M}(t))^2 dt\right] \le \pi^2 \|V\|^2_\infty \rho_n^2 N^2 \sqrt{\frac{M}{2\pi}} \\ &+ (4C_\nabla + 2\|V\|^2_\infty)\left(4\pi^2\rho_n^2 N^2 + (2N+1)^{-1}\right)\sqrt{\frac{M}{2\pi}} \\ &+ 2C_\kappa\left((2N)^{-2\kappa} + \left(\frac{2\pi^2}{M}\right)^\kappa\right). \end{aligned} \tag{7}$$

(ii) In the case of synchronous and regular sampling, when $t^j_k = k/n$ for $k = 0, 1, \cdots, n$, $j = 1, 2$, eq. (7) is improved as

$$\begin{aligned} &\mathbf{E}\left[\int_0^1 (V^{j,j'}(t) - \hat{V}^{j,j'}_{N,M}(t))^2 dt\right] \le \pi^2 \|V\|^2_\infty \rho_n^2 \sqrt{\frac{M}{2\pi}}\left(\frac{M}{4\pi^2} + 1\right) \\ &+ (4C_\nabla + 2\|V\|^2_\infty)(2N+1)^{-1}\sqrt{\frac{M}{2\pi}} \\ &+ 2C_\kappa\left((2N)^{-2\kappa} + \left(\frac{2\pi^2}{M}\right)^\kappa\right). \end{aligned} \tag{8}$$

(iii) Consequently, for the general sampling scheme, if Here $a_n \asymp b_n$ means both $\limsup_{n\to\infty} a_n/b_n$ and $\limsup_{n\to\infty} b_n/a_n$ are finite. $N \asymp \rho_n^{-\alpha}$ and $M \asymp \rho_n^{-\beta}$, the consistency is attained if

$$0 < \beta < \frac{4}{3}, \qquad \frac{\beta}{2} < \alpha < -\frac{1}{4}\beta + 1 \tag{9}$$

and

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

Further, the best rate is given as

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

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 $k = 0,1,\cdots,n$, $j = 1,2$, the consistency is attained if

$$\alpha > \frac{\beta}{2}, \qquad 0 < \beta < \frac{4}{3} \tag{10}$$

and

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

The best rate is given as

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

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 2. *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:

$$\tilde{X}_t^j = X_t^j + \eta_t^j, \qquad j = 1, ..., d, \tag{11}$$

with η being the noise component.

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

- the Heston stochastic volatility model, by Heston [17];
- the One Factor stochastic volatility model (SVF1);
- the Two Factor stochastic volatility model (SVF2), by Chernov et al. [9];
- the Rough Heston model (RH), by El Euch and Rosenbaum [11];

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;
- autocorrelated noise dependent of the 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 - 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 (4) given by:

$$\hat{V}_{N,M}^{j,j'}(t) = \frac{1}{2N+1} \sum_{l=1}^{n_j} \sum_{l'=1}^{n_{j'}} \sum_{u=-N}^{N} \sum_{u'=-N}^{N} e^{-\frac{2\pi^2(u-u')^2}{M}} e^{2\pi i u (t-t_l^j)} e^{-2\pi i u'(t-t_{l'}^{j'})} \Delta(X_l^j)\Delta(X_{l'}^{j'}). \tag{12}$$

Where not stated otherwise, the simulations consist of $K = 500$ daily trajectories, considering a trading day of length 6.5 hours, and are carried out on an equally spaced grid of width 2 seconds. To introduce asynchronicity in the data, observations are drawn from a Poisson process with an average of one observation every 10 seconds. Moreover, where not explicitly stated, the correlation between Brownian motions driving the efficient processes of different assets, following Bibinger et al. [6], is fixed to mimic the median estimated realized correlation of the Nasdaq components, i.e.:

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

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 [17] 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 [33], Mancino and Recchioni [26] and Figueroa-López and Wu [13].

Factor volatility models

Factor volatility models have been long used in the literature; see, for example, Huang and Tauchen [18]. 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 $j = 1, ..., d$, with $\langle W^j, Z^j \rangle = \lambda$, and $\left\langle Z^j, Z^{j'} \right\rangle = 0$ 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 [33], Mancino and Recchioni [26], Figueroa-López and Wu [13] and Mancino et al. [24].

Second, we consider the Two-Factor Stochastic Volatility model (SV2F), proposed by Chernov et al. [9], that is able to reproduce higher values of volatility of volatility. It takes the form:

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

for $j = 1, ..., d$, with $\langle W^j, Z^{j,1} \rangle = \langle W^j, Z^{j,2} \rangle = \lambda$, and $\left\langle Z^{j,i}, Z^{j',i'} \right\rangle = 0$ 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. [14] 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 [11] and which we intend to use in this study.

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

with $\langle W^j, Z^j \rangle = \lambda$ and $K(t) = Ct^{H-\frac{1}{2}}$ for $H \in \left(0, \frac{1}{2}\right)$ and constant C. In order to simulate the rough Heston model we apply the discrete-time Euler-type scheme studied in Richard et al. [31], referring in particular to equation (11) thereof. The parameters of the model are set to ensure that in the exercise the simulated volatility process does not exhibits negative values, and in particular they take values:

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

where the choice for the Hurst parameters is driven by the empirical evidence present in the literature; see, e.g., Gatheral et al. [14].

4.1.2. Market Microstructure Noise Specifications

It is a known fact (see, e.g. Bandi and Russell [4]) 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 [16]), 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 (11), with mean equal to zero and given constant variance

$$\eta \sim i.i.d. \quad E[\eta] = 0, \quad E[(\eta^j)^2] = \mathrm{var}(\Delta X^j_{10sec})\sigma^2_\eta,$$

where X^j_{10sec} denotes the regularly spaced series obtained by subsampling the originally simulated series every 10 seconds.

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: $\sigma^2_\eta = 1, 1.5, 2, 2.5$.

Autocorrelated Noise

Here autocorrelation is introduced in the noise component, while keeping the additive form of eq. (11). 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 $\theta_\eta = 0.2, 0.3, 0.4$. σ^2_η 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. [20], which allows for both autocorrelation and dependence on the price process. For $j \in \{1, 2\}$ and $i = 0, 1, \ldots, n_j$, using the simplified notation $\eta_i^j := \eta^j_{t_i^j}$, the latter reads as

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

where χ_i satisfies

$$\chi_i^j = Z_i^j + \sum_{l=1}^{L} \frac{g(1+g)...(l-1+g)}{l!} Z^j_{i-l}, \quad g \in (-0.5, 0.5), \quad Z_i^j \sim_{i.i.d.} \mathcal{N}(0, z), \tag{13}$$

Preprints.org (www.preprints.org) | NOT PEER-REVIEWED | Posted: Posted: 9 May 2025 doi:doi:10.20944/preprints202505.0726.v1

and ψ_i is sampled from

$$d\psi^j(t) = u(h(t) - \psi^j(t))dt + v dW_t^j, \quad h(t) = 1 + w\cos\left(\frac{2\pi}{T}t\right), \tag{14}$$

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 [19], while accounting for the possible dependence between noise and the efficient price component, as observed, for example, by [15]. This formulation still includes an OU dynamics in eq. (14), 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:

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

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 minutes.

In the optimization study, following the results of Theorem 2, we specify the cutting frequency N and the localization parameter M in terms of ρ_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 α and β, depending on the κ-*Hölder* continuity of the simulated volatility process 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.. 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:

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

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. 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.

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. [27] and [24], 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 [24]. 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 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
σ_η	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
θ	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

Preprints.org (www.preprints.org) | NOT PEER-REVIEWED | Posted: Posted: 9 May 2025 doi:doi:10.20944/preprints202505.0726.v1

Table 2. $MISE_2$ Error 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	$3.068\cdot10^{-4}$	$4.370\cdot10^{-4}$	$6.193\cdot10^{-4}$	$7.539\cdot10^{-4}$	$8.646\cdot10^{-4}$	$9.510\cdot10^{-4}$
1	$1.539\cdot10^{-4}$	$2.124\cdot10^{-4}$	$2.971\cdot10^{-4}$	$3.619\cdot10^{-4}$	$4.164\cdot10^{-4}$	$4.331\cdot10^{-4}$
3	$5.172\cdot10^{-5}$	$7.101\cdot10^{-5}$	$1.002\cdot10^{-4}$	$1.232\cdot10^{-4}$	$1.427\cdot10^{-4}$	$1.553\cdot10^{-4}$
5	$3.985\cdot10^{-5}$	$5.238\cdot10^{-5}$	$7.081\cdot10^{-5}$	$8.488\cdot10^{-5}$	$9.660\cdot10^{-5}$	$1.023\cdot10^{-4}$
7	$4.657\cdot10^{-5}$	$5.541\cdot10^{-5}$	$6.846\cdot10^{-5}$	$7.847\cdot10^{-5}$	$8.687\cdot10^{-5}$	$9.636\cdot10^{-5}$
9	$5.732\cdot10^{-5}$	$6.524\cdot10^{-5}$	$6.989\cdot10^{-5}$	$8.874\cdot10^{-5}$	$9.113\cdot10^{-5}$	$9.995\cdot10^{-5}$
			Heston - I.i.d. noise $\sigma_\eta = 2.5$			
c_N/c_M	0.5	1	2	3	4	5
0.5	$3.215\cdot10^{-4}$	$4.562\cdot10^{-4}$	$6.456\cdot10^{-4}$	$7.856\cdot10^{-4}$	$9.006\cdot10^{-4}$	$9.228\cdot10^{-4}$
1	$1.046\cdot10^{-4}$	$1.509\cdot10^{-4}$	$2.173\cdot10^{-4}$	$2.676\cdot10^{-4}$	$3.096\cdot10^{-4}$	$3.261\cdot10^{-4}$
3	$1.543\cdot10^{-4}$	$2.125\cdot10^{-4}$	$2.968\cdot10^{-4}$	$3.613\cdot10^{-4}$	$4.159\cdot10^{-4}$	$4.365\cdot10^{-4}$
5	$1.768\cdot10^{-4}$	$2.408\cdot10^{-4}$	$3.349\cdot10^{-4}$	$4.073\cdot10^{-4}$	$4.683\cdot10^{-4}$	$4.883\cdot10^{-4}$
7	$2.506\cdot10^{-4}$	$3.286\cdot10^{-4}$	$4.447\cdot10^{-4}$	$5.347\cdot10^{-4}$	$6.115\cdot10^{-4}$	$6.510\cdot10^{-4}$
9	$2.732\cdot10^{-4}$	$3.682\cdot10^{-4}$	$4.770\cdot10^{-4}$	$5.618\cdot10^{-4}$	$6.663\cdot10^{-4}$	$6.798\cdot10^{-4}$
			SVF2 - No noise			
c_N/c_M	0.5	1	2	3	4	5
0.5	$8.197\cdot10^{-4}$	$1.234\cdot10^{-3}$	$1.855\cdot10^{-3}$	$2.328\cdot10^{-3}$	$2.723\cdot10^{-3}$	$2.982\cdot10^{-4}$
1	$4.804\cdot10^{-4}$	$6.877\cdot10^{-4}$	$9.867\cdot10^{-4}$	$1.212\cdot10^{-3}$	$1.403\cdot10^{-3}$	$1.633\cdot10^{-4}$
3	$2.392\cdot10^{-4}$	$3.066\cdot10^{-4}$	$4.066\cdot10^{-4}$	$4.826\cdot10^{-4}$	$5.462\cdot10^{-4}$	$5.701\cdot10^{-4}$
5	$1.860\cdot10^{-4}$	$2.161\cdot10^{-4}$	$2.639\cdot10^{-4}$	$3.020\cdot10^{-4}$	$3.354\cdot10^{-4}$	$3.411\cdot10^{-4}$
7	$2.068\cdot10^{-4}$	$2.183\cdot10^{-4}$	$2.449\cdot10^{-4}$	$2.699\cdot10^{-4}$	$2.931\cdot10^{-4}$	$3.028\cdot10^{-4}$
9	$2.236\cdot10^{-4}$	$2.421\cdot10^{-4}$	$2.563\cdot10^{-4}$	$2.784\cdot10^{-4}$	$2.988\cdot10^{-4}$	$3.295\cdot10^{-4}$
			SVF2 - I.i.d. noise $\sigma_\eta = 2.5$			
c_N/c_M	0.5	1	2	3	4	5
0.5	$8.616\cdot10^{-4}$	$1.311\cdot10^{-3}$	$1.974\cdot10^{-3}$	$2.473\cdot10^{-3}$	$2.884\cdot10^{-3}$	$3.115\cdot10^{-3}$
1	$4.529\cdot10^{-4}$	$6.001\cdot10^{-4}$	$8.176\cdot10^{-4}$	$9.825\cdot10^{-4}$	$1.117\cdot10^{-3}$	$1.228\cdot10^{-3}$
3	$5.701\cdot10^{-4}$	$8.084\cdot10^{-4}$	$1.144\cdot10^{-3}$	$1.394\cdot10^{-3}$	$1.604\cdot10^{-3}$	$1.773\cdot10^{-4}$
5	$6.232\cdot10^{-4}$	$8.579\cdot10^{-4}$	$1.163\cdot10^{-3}$	$1.377\cdot10^{-3}$	$1.553\cdot10^{-3}$	$1.640\cdot10^{-3}$
7	$6.173\cdot10^{-4}$	$8.607\cdot10^{-4}$	$1.254\cdot10^{-3}$	$1.574\cdot10^{-3}$	$1.849\cdot10^{-3}$	$1.930\cdot10^{-3}$
9	$6.454\cdot10^{-4}$	$8.773\cdot10^{-4}$	$1.296\cdot10^{-3}$	$1.602\cdot10^{-3}$	$1.890\cdot10^{-3}$	$1.981\cdot10^{-3}$

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. [29] (STS);
- the local method of moments spot estimators, by Bibinger et al. [6] (LMM).

The kernel-based estimator proposed by Bu et al. [7], 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. [6], while for the STS estimator we choose the parameters in a neighborhood of the one used by Chen et al. [8], minimizing the mean square error obtained on auxiliary simulations. In the following, when not explicitly stated otherwise, the results are meant to be achieve 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 derease 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	$6.641 \cdot 10^{-5}$	100%	$5.302 \cdot 10^{-5}$	100%
LMM	$2.522 \cdot 10^{-4}$	100%	$1.023 \cdot 10^{-4}$	100%
STS	$2.310 \cdot 10^{-4}$	100%	$2.034 \cdot 10^{-4}$	89.98%
	d=5		d=25	
GPDF	$5.778 \cdot 10^{-5}$	100%	$5.252 \cdot 10^{-5}$	100%
LMM	$1.531 \cdot 10^{-4}$	100%	$9.796 \cdot 10^{-5}$	100%
STS	$2.155 \cdot 10^{-4}$	100%	$2.032 \cdot 10^{-4}$	64.45%
	d=10		d=30	
GPDF	$5.435 \cdot 10^{-5}$	100%	$5.193 \cdot 10^{-5}$	100%
LMM	$1.191 \cdot 10^{-4}$	100%	$9.539 \cdot 10^{-5}$	100%
STS	$2.079 \cdot 10^{-4}$	100%	$2.033 \cdot 10^{-4}$	9.94%
	d=15		d=40	
GPDF	$5.353 \cdot 10^{-5}$	100%	$5.194 \cdot 10^{-5}$	100%
LMM	$1.078 \cdot 10^{-4}$	100%	$9.528 \cdot 10^{-5}$	99.54%

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 seconds. Table 4 shows that, still maintaining an edge in terms of MISE with respect to the competitors, the PDF 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	% SPSD	MISE	% SPSD	MISE	% SPSD
	d=5, $\bar{\Delta}t = 15$		d=5, $\bar{\Delta}t = 20$		d=5, $\bar{\Delta}t = 30$	
PDF	$6.283 \cdot 10^{-5}$	100%	$7.886 \cdot 10^{-5}$	100%	$9.671 \cdot 10^{-5}$	100%
LMM	$1.648 \cdot 10^{-4}$	100%	$1.810 \cdot 10^{-4}$	100%	$2.503 \cdot 10^{-4}$	99.83%
STS	$2.492 \cdot 10^{-4}$	100%	$2.536 \cdot 10^{-4}$	100%	$2.646 \cdot 10^{-4}$	100%
	d=10, $\bar{\Delta}t = 15$		d=10, $\bar{\Delta}t = 20$		d=10, $\bar{\Delta}t = 30$	
PDF	$6.379 \cdot 10^{-5}$	100%	$7.492 \cdot 10^{-5}$	100%	$9.120 \cdot 10^{-5}$	100%
LMM	$1.991 \cdot 10^{-4}$	100%	$1.805 \cdot 10^{-4}$	100%	$1.936 \cdot 10^{-4}$	99.66%
STS	$2.255 \cdot 10^{-4}$	100%	$2.327 \cdot 10^{-4}$	100%	$2.401 \cdot 10^{-4}$	100%
	d=15, $\bar{\Delta}t = 15$		d=15, $\bar{\Delta}t = 20$		d=15, $\bar{\Delta}t = 30$	
PDF	$6.345 \cdot 10^{-5}$	100%	$7.389 \cdot 10^{-5}$	100%	$9.203 \cdot 10^{-5}$	100%
LMM	$1.679 \cdot 10^{-4}$	100%	$1.485 \cdot 10^{-4}$	98.99%	$1.906 \cdot 10^{-4}$	98.52%
STS	$2.201 \cdot 10^{-4}$	99.80%	$2.265 \cdot 10^{-4}$	99.80%	$2.341 \cdot 10^{-4}$	98.79 %
	d=20, $\bar{\Delta}t = 15$		d=20, $\bar{\Delta}t = 20$		d=20, $\bar{\Delta}t = 30$	
PDF	$6.325 \cdot 10^{-5}$	100%	$7.337 \cdot 10^{-5}$	100%	$9.196 \cdot 10^{-5}$	100%
LMM	$1.660 \cdot 10^{-4}$	99.78%	$1.306 \cdot 10^{-4}$	90.02%	$1.877 \cdot 10^{-4}$	96.25%
STS	$2.187 \cdot 10^{-4}$	92.59%	$2.241 \cdot 10^{-5}$	87.16%	$2.311 \cdot 10^{-4}$	82.14 %

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	% SPSD	MISE	% SPSD
	d=5, r=0.01		d=5, r=0.05	
GPDF	$5.678{\cdot}10^{-5}$	100%	$5.679{\cdot}10^{-5}$	100%
LMM	$1.540{\cdot}10^{-4}$	100%	$1.541{\cdot}10^{-4}$	100%
STS	$2.208{\cdot}10^{-4}$	100%	$2.208{\cdot}10^{-4}$	100%
$\cdot 10^{-5}$	d=10, r=0.01		d=10, r=0.05	
GPDF	$5.342{\cdot}10^{-5}$	100%	$5.344{\cdot}10^{-5}$	100%
LMM	$1.183{\cdot}10^{-4}$	100%	$1.183{\cdot}10^{-4}$	100%
STS	$2.074{\cdot}10^{-4}$	100%	$2.074{\cdot}10^{-4}$	100%
	d=15, r=0.01		d=15, r=0.05	
GPDF	$5.308{\cdot}10^{-5}$	100%	$5.308{\cdot}10^{-5}$	100%
LMM	$1.078{\cdot}10^{-4}$	100%	$1.072{\cdot}10^{-4}$	100%
STS	$2.034{\cdot}10^{-4}$	99.97%	$2.034{\cdot}10^{-4}$	99.97%
	d=20, r=0.01		d=20, r=0.05	
GPDF	$5.244{\cdot}10^{-5}$	100%	$5.245{\cdot}10^{-5}$	100%
LMM	$1.009{\cdot}10^{-4}$	100%	$1.009{\cdot}10^{-4}$	100%
STS	$2.022{\cdot}10^{-4}$	97.35%	$2.022{\cdot}10^{-4}$	97.32%

Table 6 shows that i.i.d. noise, especially with high noise variance σ_η^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 GPDF 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 [32], 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	% SPSD	MISE	% SPSD	MISE	% SPSD	MISE	% SPSD
	d=5, $\sigma_\eta = 1$		d=5, $\sigma_\eta = 1.5$		d=5, $\sigma_\eta = 2$		d=5, $\sigma_\eta = 2.5$	
GPDF	$8.017{\cdot}10^{-5}$	100%	$8.026{\cdot}10^{-5}$	100%	$1.453{\cdot}10^{-4}$	100%	$1.894{\cdot}10^{-4}$	100%
LMM	$1.278{\cdot}10^{-4}$	100%	$1.378{\cdot}10^{-4}$	100%	$1.697{\cdot}10^{-4}$	100%	$2.109{\cdot}10^{-4}$	100%
STS	$2.360{\cdot}10^{-4}$	100%	$2.573{\cdot}10^{-4}$	100%	$2.930{\cdot}10^{-4}$	100%	$3.448{\cdot}10^{-4}$	98.45%
	d=10, $\sigma_\eta = 1$		d=10, $\sigma_\eta = 1.5$		d=10, $\sigma_\eta = 2$		d=10, $\sigma_\eta = 2.5$	
GPDF	$7.072{\cdot}10^{-5}$	100%	$7.653{\cdot}10^{-5}$	100%	$1.425{\cdot}10^{-4}$	100%	$1.835{\cdot}10^{-4}$	100%
LMM	$1.204{\cdot}10^{-4}$	100%	$1.392{\cdot}10^{-4}$	100%	$1.684{\cdot}10^{-4}$	100%	$2.033{\cdot}10^{-4}$	99.94%
STS	$2.198\ {\cdot}10^{-4}$	99.83%	$2.384{\cdot}10^{-4}$	99.08%	$2.697{\cdot}10^{-4}$	88.39%	$3.189{\cdot}10^{-4}$	54.33%
	d=15, $\sigma_\eta = 1$		d=15, $\sigma_\eta = 1.5$		d=15, $\sigma_\eta = 2$		d=15, $\sigma_\eta = 2.5$	
GPDF	$6.763{\cdot}10^{-5}$	100%	$7.575{\cdot}10^{-5}$	100%	$1.390{\cdot}10^{-4}$	100%	$1.789{\cdot}10^{-4}$	100%
LMM	$1.198{\cdot}10^{-4}$	100%	$1.405{\cdot}10^{-4}$	100%	$1.687{\cdot}10^{-4}$	100%	$2.005{\cdot}10^{-4}$	99.05%
STS	$2.155{\cdot}10^{-4}$	97.98%	$2.335{\cdot}10^{-4}$	77.01%	$2.647{\cdot}10^{-4}$	25.93%	$3.177{\cdot}10^{-4}$	21.80%
	d=20, $\sigma_\eta = 1$		d=20, $\sigma_\eta = 1.5$		d=20, $\sigma_\eta = 2$		d=20, $\sigma_\eta = 2.5$	
GPDF	$6.653{\cdot}10^{-5}$	100%	$7.554{\cdot}10^{-5}$	100%	$1.392{\cdot}10^{-4}$	100%	$1.783{\cdot}10^{-4}$	100%
LMM	$1.180{\cdot}10^{-4}$	100%	$1.384{\cdot}10^{-4}$	99.88%	$1.662{\cdot}10^{-4}$	98.80%	$1.986{\cdot}10^{-4}$	94.60%
STS	$2.163{\cdot}10^{-4}$	67.46%	$2.325{\cdot}10^{-4}$	14.18%	$2.622{\cdot}10^{-4}$	8.33%	$3.184{\cdot}10^{-4}$	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 θ_η, while the LMM estimator seems to be only slightly affected. Low values of θ_η, 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, $\theta_\eta = 0.4$	
PDF	$2.114 \cdot 10^{-4}$	100%	$1.924 \cdot 10^{-4}$	100%	$1.711 \cdot 10^{-4}$	100%
LMM	$2.093 \cdot 10^{-3}$	100%	$2.957 \cdot 10^{-4}$	100%	$2.704 \cdot 10^{-4}$	100%
STS	$4.319 \cdot 10^{-4}$	91.99%	$3.139 \cdot 10^{-4}$	98.91%	$2.984 \cdot 10^{-4}$	99.86%
	d=10, $\theta_\eta = 0.2$		d=10, $\theta_\eta = 0.3$		d=10, $\theta_\eta = 0.4$	
PDF	$1.937 \cdot 10^{-4}$	100%	$1.790 \cdot 10^{-4}$	100%	$1.625 \cdot 10^{-4}$	100%
LMM	$2.727 \cdot 10^{-4}$	100%	$2.892 \cdot 10^{-4}$	99.77%	$1.972 \cdot 10^{-4}$	100%
STS	$4.012 \cdot 10^{-4}$	21.66%	$3.099 \cdot 10^{-4}$	65.18%	$2.758 \cdot 10^{-4}$	86.84%
	d=15, $\theta_\eta = 0.2$		d=15, $\theta_\eta = 0.3$		d=15, $\theta_\eta = 0.4$	
PDF	$1.878 \cdot 10^{-4}$	100%	$1.750 \cdot 10^{-4}$	100%	$1.596 \cdot 10^{-4}$	100%
LMM	$2.703 \cdot 10^{-4}$	100%	$2.819 \cdot 10^{-4}$	95.98%	$1.703 \cdot 10^{-4}$	96.80%
STS	$3.900 \cdot 10^{-4}$	0.09%	$3.017 \cdot 10^{-4}$	5.16%	$2.697 \cdot 10^{-4}$	22.69%
	d=20, $\theta_\eta = 0.2$		d=20, $\theta_\eta = 0.3$		d=20, $\theta_\eta = 0.4$	
PDF	$1.869 \cdot 10^{-4}$	100%	$1.746 \cdot 10^{-4}$	100%	$1.595 \cdot 10^{-4}$	100%
LMM	$2.712 \cdot 10^{-4}$	94.90%	$2.439 \cdot 10^{-4}$	92.98%	$1.705 \cdot 10^{-4}$	95.48%
STS	$3.867 \cdot 10^{-4}$	0.0%	$2.987 \cdot 10^{-4}$	0.0%	$2.671 \cdot 10^{-4}$	1.98%

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 PDF 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, $g = 0.45, w = 0.9$	
PDF	$2.667 \cdot 10^{-4}$	100%	$2.587 \cdot 10^{-4}$	100%	$2.823 \cdot 10^{-4}$	100%	$3.009 \cdot 10^{-4}$	100%
LMM	$2.799 \cdot 10^{-4}$	100%	$2.788 \cdot 10^{-4}$	99.54%	$3.014 \cdot 10^{-4}$	100%	$3.693 \cdot 10^{-4}$	99.14%
STS	$2.965 \cdot 10^{-4}$	99.93%	$3.266 \cdot 10^{-4}$	99.04%	$4.387 \cdot 10^{-4}$	99.80%	$4.616 \cdot 10^{-4}$	99.05%
	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, $g = 0.45, w = 0.9$	
PDF	$1.754 \cdot 10^{-4}$	100%	$2.099 \cdot 10^{-4}$	100%	$2.449 \cdot 10^{-4}$	100%	$2.643 \cdot 10^{-4}$	100%
LMM	$2.145 \cdot 10^{-4}$	99.89%	$2.678 \cdot 10^{-4}$	96.57%	$2.534 \cdot 10^{-4}$	98.50%	$3.687 \cdot 10^{-4}$	90.17%
STS	$2.730 \cdot 10^{-4}$	99.93%	$2.995 \cdot 10^{-4}$	79.46%	$3.795 \cdot 10^{-4}$	91.10%	$4.685 \cdot 10^{-4}$	76.46%
	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, $g = 0.45, w = 0.9$	
PDF	$1.452 \cdot 10^{-4}$	100%	$1.714 \cdot 10^{-4}$	100%	$2.362 \cdot 10^{-4}$	100%	$2.499 \cdot 10^{-4}$	100%
LMM	$1.882 \cdot 10^{-4}$	99.37%	$3.265 \cdot 10^{-4}$	86.47%	$2.526 \cdot 10^{-4}$	92.31%	$3.522 \cdot 10^{-4}$	79,70%
STS	$2.656 \cdot 10^{-4}$	52.98%	$2.915 \cdot 10^{-4}$	46.86%	$3.616 \cdot 10^{-4}$	39.72%	$4.420 \cdot 10^{-4}$	40.54%
	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, $g = 0.45, w = 0.9$	
PDF	$1.313 \cdot 10^{-4}$	100%	$1.523 \cdot 10^{-4}$	100%	$2.263 \cdot 10^{-4}$	100%	$2.417 \cdot 10^{-4}$	100%
LMM	$2.752 \cdot 10^{-4}$	98.14%	$2.890 \cdot 10^{-4}$	80.25%	$3.024 \cdot 10^{-4}$	94.03%	$3.373 \cdot 10^{-4}$	80.46%
STS	$2.643 \cdot 10^{-4}$	7.53%	$2.891 \cdot 10^{-4}$	22.55%	$3.555 \cdot 10^{-4}$	3.47%	$4.298 \cdot 10^{-4}$	16.93%

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 PDF and the LMM estimators are reduced when using the SVF2 or the Rough Heston model, the results are substantially confirmed: the PDF 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 or RMISE are analogous, and always see the PDF estimator having a competitive edge.

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

	SVF1		SVF2		Rough H.	
Estimator	MISE	% SPSD	MISE	% SPSD	MISE	% SPSD
			d=2			
PDF	$2.401 \cdot 10^{-5}$	100%	$2.405 \cdot 10^{-3}$	100%	$4.667 \cdot 10^{-3}$	100%
LMM	$6.659 \cdot 10^{-5}$	100%	$4.644 \cdot 10^{-3}$	100%	$6.493 \cdot 10^{-3}$	100%
STS	$9.919 \cdot 10^{-5}$	100%	$6.733 \cdot 10^{-3}$	100%	$8.021 \cdot 10^{-3}$	100%
			d=5			
PDF	$1.971 \cdot 10^{-5}$	100%	$1.415 \cdot 10^{-3}$	100%	$2.433 \cdot 10^{-3}$	100%
LMM	$4.203 \cdot 10^{-5}$	100%	$2.440 \cdot 10^{-5}$	100%	$3.300 \cdot 10^{-3}$	100%
STS	$7.798 \cdot 10^{-5}$	100%	$3.410 \cdot 10^{-3}$	100%	$4.962 \cdot 10^{-3}$	100%
			d=10			
PDF	$1.841 \cdot 10^{-5}$	100%	$6.743 \cdot 10^{-4}$	100%	$1.639 \cdot 10^{-3}$	100%
LMM	$3.96 \cdot 10^{-5}$	100%	$1.171 \cdot 10^{-3}$	100%	$2.120 \cdot 10^{-3}$	100%
STS	$7.242 \cdot 10^{-5}$	100%	$1.916 \cdot 10^{-3}$	100%	$3.731 \cdot 10^{-3}$	100%
			d=15			
PDF	$1.784 \cdot 10^{-5}$	100%	$5.261 \cdot 10^{-4}$	100%	$1.387 \cdot 10^{-3}$	100%
LMM	$3.541 \cdot 10^{-5}$	100%	$9.254 \cdot 10^{-4}$	100%	$1.844 \cdot 10^{-3}$	100%
STS	$7.067 \cdot 10^{-5}$	100%	$1.601 \cdot 10^{-3}$	99.95%	$3.379 \cdot 10^{-3}$	99.86%
			d=20			
PDF	$1.757 \cdot 10^{-5}$	100%	$4.506 \cdot 10^{-4}$	100%	$1.253 \cdot 10^{-3}$	100%
LMM	$3.347 \cdot 10^{-5}$	100%	$8.660 \cdot 10^{-4}$	99.84%	$1.661 \cdot 10^{-3}$	99.50%
STS	$6.960 \cdot 10^{-5}$	99.54%	$1.441 \cdot 10^{-3}$	96.75%	$3.368 \cdot 10^{-3}$	96.06%

5. Conclusions

In the present paper a modified version of the classical Fourier estimator for spot covariance by Malliavin and Mancino [23] 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. We believe that the PDF estimator of spot covariance has a competitive advantage in terms of empirical applications due to its properties.

Conflicts 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:

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

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 $a_1(k) = c(k)e^{2\pi ikt}$, $a_2(s) = e^{-2\pi ist^{j'} l'}$ and $a_3(s') = e^{-2\pi ist^j_l}$, using the change of variable $u- \to -u'$, we obtain:

$$\hat{V}_N^{j,j'}(\tau) = \sum_{l=1}^{n_j} \sum_{l'=1}^{n_{j'}} \sum_{u=-N}^{N} \sum_{u'=-N}^{N} c(u-u')e^{2\pi iu(\tau-t^j_l)}e^{-2\pi iu'(\tau-t^{j'}_{l'})}\Delta(X^j_l)\Delta(X^{j'}_{l'}).$$

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

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

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. □

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

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

We put

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

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

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 μ_M so far as $0 < c_M(k) < 1$, which is true for the Gaussian case.

Lemma A1. *We have*

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

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

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

Proof. Since

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

we obtain (A1) once we establish

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

To prove eq. (A3), we first observe that

$$\begin{aligned}
&\left|\int_0^1 (e^{-2\pi i k_1 \varphi_n^1(s) - 2\pi i k_2 \varphi_n^2(s)} - e^{-2\pi i(k_1+k_2)s}) V^{1,2}(s) ds\right| \\
&\leq \sup_{t\in[0,1]} |V^{1,2}(t)| \int_0^1 |1 - e^{2\pi i(k_1(s-\varphi_n^1(s)) + ik_2(s-\varphi_n^2(s))}| ds,
\end{aligned}$$

and

$$\begin{aligned}
&\int_0^1 |1 - e^{2\pi i(k_1(s-\varphi_n^1(s)) + ik_2(s-\varphi_n^2(s))}| ds \\
&\leq 2\pi \int_0^1 |k_1(s-\varphi_n^1(s)) + k_2(s-\varphi_n^2(s))| ds \\
&\leq 2\pi |k_1| \int_0^1 |s-\varphi_n^1(s)| ds + 2\pi |k_2| \int_0^1 |s-\varphi_n^2(s)| ds.
\end{aligned}$$

Then, (A3) holds since

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

Preprints.org (www.preprints.org) | NOT PEER-REVIEWED | Posted: Posted: 9 May 2025 doi:doi:10.20944/preprints202505.0726.v1

for $j = 1, 2$.

For (A2), we just need

$$\int_0^1 |1 - e^{2\pi i (k_1(s-\varphi_n^1(s)) + ik_2(s-\varphi_n^2(s))}| ds$$
$$\leq 2\pi |k_1 + k_2| \int_0^1 |s - \varphi_n^1(s)| ds.$$

□

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

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

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

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

Proof. We first show that

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

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

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

By the Malliavin integration by parts formula,

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

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

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

Thus we obtain (A7).

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

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

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

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

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

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

By (A4), we obtain (A5). □

Lemma A3.

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

Proof. We first note that

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

and

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

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

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

we have

$$\left(1-(1-\frac{|k|}{2N+1})c_M(k)\right)^2 \le 2(1-c_M(k))^2 + 2(\frac{|k|}{2N+1}c_M(k))^2$$
$$\le 2(1-c_M(k))^\kappa + 2\left(\frac{|k|c_M(k)}{2N+1}\right)^{2\kappa},$$

and therefore we have

$$\sum_{-2N\le k\le 2N}\left(1-(1-\frac{|k|}{2N+1})c_M(k)\right)^2|(\mathcal{F}V)(k)|^2$$
$$\le 2\left(\sup_{0<|k|\le 2N}\left(\frac{1-c_M(k)}{|k|^2}\right)^\kappa + (2N+1)^{-2\kappa}\right)\sum_{-2N\le k\le 2N}|k|^{2\alpha}|(\mathcal{F}V)(k)|^2. \tag{A11}$$

Combining (A10) and (A11), we get (A9) by the assumption (6). □

Now we are ready to prove Theorem 2

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

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

Then,

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

and

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

Further, we have

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

so that

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

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

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

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

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

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 (9). In such a case $\kappa\beta < 2\kappa\alpha$.

When the sampling is synchronous and regularly spaced, the necessary condition for consistency clearly becomes (10). □

Appendix C Additional results of comparison for alternative models

Table A1. % 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%	99.85%	99.68%	97.28%	93.98%	94.98%
STS	100%	100%	100%	100%	100%	100%	99.77%	99.83%	99.79%	98.28%	97.89%	98.12%
	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%	99.96%	100%	100%	99.70%	99.25%	95.26	90.03%	85.48%	87.11%
STS	99.97%	100%	99.97%	98.97%	98.68%	98.87%	88.19%	89.16%	88.26%	54.56%	50.29%	57.80%
	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%	100%	99.96%	99.80%	99.80%	99.25%	97.85%	98.02%	79.28%	2.71%	77.98%
STS	97.75%	99.90%	97.22%	75.13%	72.61%	75.86%	35.07%	29.54%	28.14%	2.25%	2.71%	3.90%
	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%	100%	99.96%	98.27%	98.82%	97.95%	95.23%	95.88%	65.91%	63.84%	65.95%
STS	64.85%	66.79%	65.05%	13.62%	12.20%	15.91%	2.97%	0.0%	0.40%	0.0%	0.0%	0.0%

 doi:doi:10.20944/preprints202505.0726.v1

Table A2. % 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%	99.96%	99.96%	99.25%	98.95%	99.18%	92.79%	91.65%	91.31%	99.2%	98.41%	97.80%
STS	98.53%	99.80%	99.90%	98.58%	98.37%	99.14%	99.83%	99.70%	99.86%	98.87%	98.77%	99.31%
	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%	99.73%	99.77%	95.78%	95.74%	95.61%	79.97%	80.03%	81.27%	90.01%	88.37%	86.54%
STS	93.91%	92.06%	94.41%	78.54%	75.96%	80.29%	91.03%	89.29%	99.20%	76.69%	75.17%	78.27%
	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%	98.02%	99.25%	86.87%	89.58%	91.22%	67.04%	66.54%	68.96%	80.01%	78.96%	78.60%
STS	50.53%	46.99%	49.86%	45.57%	44.38%	44.64%	38.59%	36.51%	38.00%	82.77%	37.86%	37.87%
	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	99.82%	97.09%	97.58%	85.26%	80.64%	83.44%	52.92%	51.88%	55.24%	75.45%	75.18%	73.78%
STS	7.17%	6.12%	6.22%	22.85%	21.69%	11.21%	2.81%	2.61%	3.08%	78.81%	16.30%	16.07%

References

1. Aït-Sahalia, Yacine and Jean Jacod. 2014. *High-frequency financial econometrics*. Princeton University Press.
2. Aït-Sahalia, Yacine and Dacheng Xiu. 2019. Principal component analysis of high-frequency data. *Journal of the American Statistical Association 114*(525), 287–303.
3. Akahori, Jirô, Nien-Lin Liu, Maria Elvira Mancino, and Yukie Yasuda. 2014. The fourier estimation method with positive semi-definite estimators. *arXiv:1410.0112 [q-fin.ST]*.
4. Bandi, Federico M and Jeffrey R Russell. 2008. Microstructure noise, realized variance, and optimal sampling. *The Review of Economic Studies 75*(2), 339–369.
5. 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*(2), 149–169.
6. 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*(3), 419–435.
7. Bu, Ruijun, Degui Li, Oliver Linton, and Hanchao Wang. 2025. Nonparametric estimation of large spot volatility matrices for high-frequency financial data. *Econometric Theory*, 1–38.
8. 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*(532), 1960–1977.
9. Chernov, Mikhail, A Ronald Gallant, Eric Ghysels, and George Tauchen. 2003. Alternative models for stock price dynamics. *Journal of Econometrics 116*(1-2), 225–257.
10. Cui, Liyuan, Yongmiao Hong, Yingxing Li, and Junhui Wang. 2019. Large-dimensional positive definite covariance estimation for high frequency data via low-rank and sparse matrix decomposition. *Available at SSRN 3414910*.
11. El Euch, Omar and Mathieu Rosenbaum. 2019. The characteristic function of rough heston models. *Mathematical Finance 29*(1), 3–38.
12. Engle, Robert and Riccardo Colacito. 2006. Testing and valuing dynamic correlations for asset allocation. *Journal of Business & Economic Statistics 24*(2), 238–253.
13. Figueroa-López, José E and Bei Wu. 2022. Kernel estimation of spot volatility with microstructure noise using pre-averaging. *Econometric Theory*, 1–50.
14. Gatheral, Jim, Thibault Jaisson, and Mathieu Rosenbaum. 2018. Volatility is rough. *Quantitative finance 18*(6), 933–949.
15. Hansen, Peter R and Asger Lunde. 2006. Realized variance and market microstructure noise. *Journal of Business & Economic Statistics 24*(2), 127–161.
16. Hasbrouck, Joel. 2007. *Empirical market microstructure: the institutions, economics, and econometrics of securities trading*. Oxford University Press.
17. 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*(2), 327–343.

18. Huang, Xin and George Tauchen. 2005. The relative contribution of jumps to total price variance. *Journal of financial econometrics 3*(4), 456–499.
19. Jacod, Jean, Yingying Li, and Xinghua Zheng. 2017. Statistical properties of microstructure noise. *Econometrica 85*(4), 1133–1174.
20. Jacod, Jean, Yingying Li, and Xinghua Zheng. 2019. Estimating the integrated volatility with tick observations. *Journal of Econometrics 208*(1), 80–100.
21. 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*(3), 747–761.
22. Malliavin, Paul and Maria Elvira Mancino. 2002. Fourier series method for measurement of multivariate volatilities. *Finance and Stochastics 6*(1), 49–61.
23. Malliavin, Paul and Maria Elvira Mancino. 2009. A fourier transform method for nonparametric estimation of multivariate volatility. *The Annals of Statistics 37*(4), 1983–2010.
24. 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.
25. Mancino, Maria Elvira, Tommaso Mariotti, and Giacomo Toscano. 2025. Spot beta estimation with asynchronous noisy prices. *Quantitative Finance, forthcoming*.
26. Mancino, Maria Elvira and Maria Cristina Recchioni. 2015. Fourier spot volatility estimator: Asymptotic normality and efficiency with liquid and illiquid high-frequency data. *PloS one 10*(9), e0139041.
27. Mancino, Maria Elvira, Maria Cristina Recchioni, and Simona Sanfelici. 2017. *Fourier-Malliavin volatility estimation: Theory and practice*. SpringerBriefs.
28. Mancino, Maria Elvira and Simona Sanfelici. 2011. Estimating covariance via fourier method in the presence of asynchronous trading and microstructure noise. *Journal of Financial Econometrics 9*(2), 367–408.
29. 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*(1), 101–119.
30. 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*(2), 325–347.
31. Richard, Alexandre, Xiaolu Tan, and Fan Yang. 2023. On the discrete-time simulation of the rough heston model. *SIAM Journal on Financial Mathematics 14*(1).
32. Roll, Richard. 1984. A simple implicit measure of the effective bid-ask spread in an efficient market. *The Journal of finance 39*(4), 1127–1139.
33. Zu, Yang and H Peter Boswijk. 2014. Estimating spot volatility with high-frequency financial data. *Journal of Econometrics 181*(2), 117–135.