IRTG 1792 Discussion Paper 2020-025

Non-Parametric Estimation of Spot Covariance Matrix with High-Frequency Data

Konul Mustafayeva [*]
Weining Wang [*2]

[*] King's College London, UK
[*2] University of York, UK

This research was supported by the Deutsche Forschungsgesellschaft through the International Research Training Group 1792 "High Dimensional Nonstationary Time Series".

http://irtg1792.hu-berlin.de
ISSN 2568-5619

Non-Parametric Estimation of Spot Covariance Matrix with High-Frequency Data

Konul Mustafayeva[1] and Weining Wang [2]

[1]Department of Mathematics, King's College London, London, UK
[2]Department of Economics, University of York, York, UK

Correspondence to: Konul Mustafayeva, Department of Mathematics, King's College London, Strand, London, WC2R 2LS.
E-mail: konul.mustafayeva@kcl.ac.uk
Phone: +44(0)7707833139

Abstract

Estimating spot covariance is an important issue to study, especially with the increasing availability of high-frequency financial data. We study the estimation of spot covariance using a kernel method for high-frequency data. In particular, we consider first the kernel weighted version of realized covariance estimator for the price process governed by a continuous multivariate semimartingale. Next, we extend it to the threshold kernel estimator of the spot covariances when the underlying price process is a discontinuous multivariate semimartingale with finite activity jumps. We derive the asymptotic distribution of the estimators for both fixed and shrinking bandwidth. The estimator in a setting with jumps has the same rate of convergence as the estimator for diffusion processes without jumps. A simulation study examines the finite sample properties of the estimators. In addition, we study an application of the estimator in the context of covariance forecasting. We discover that the forecasting model with our estimator outperforms a benchmark model in the literature.

Keywords: high-frequency data; kernel estimation; jump; forecasting covariance matrix
MOS subject classification: 62F12, 62G05, 60J75.

1 Introduction

The broad availability of high-frequency intra-day data of asset returns has given rise to a considerable collection of works dedicated to estimating integrated and spot (co)variances. While integrated (co)variance is an important quantity in risk management and optimal hedging, in recent years the usage of spot volatility estimators is also increasing in financial applications. For example, spot covariance estimates have been shown to be beneficial, with respect to the integrated covariance, in estimating infinitesimal cross-moments Bandi and Renò (2016) and in co-jump tests Bibinger and Winkelmann (2015), estimating parametric multivariate stochastic volatility models Kanaya and Kristensen (2016). Moreover, understanding covariance dynamics is crucial for effective portfolio choice and derivative pricing with stochastic volatility where the initial volatility value, in addition to the initial value of the underlying, is needed to price the option.

We consider the nonparametric estimation of spot covariance with high-frequency financial data. Our study is at the intersection of two fields of literature. The first strand of literature is on estimating integrated covariance matrices over a fixed period. This topic has been studied extensively in high-frequency econometrics. For example, the highly celebrated paper by Barndorff-Nielsen and Shephard (2004a) makes important contributions to the use of realized covariance to estimate integrated covariance matrix in a setup without market microstructure noise. The quasi-maximum likelihood estimator in Aït-Sahalia et al. (2010), the multivariate pre-averaging estimator in Christensen et al. (2013), the two-scale estimator in Zhang (2011) are robust to microstructure noise.

The second strand focuses on spot volatility estimation. Several approaches of estimating spot volatility were proposed. Foster et al. (1996) were the first to introduce the spot volatility estimator: rolling and sampling filters. Later, kernel-type estimators were introduced in

Fan and Wang (2008) and Kristensen (2010). These estimators of spot variance neglect the microstructure noise and jumps. The examples of spot variance estimators accounting for microstructure noise include Zu and Boswijk (2014), Bos et al. (2012), Mykland and Zhang (2008). Yu et al. (2014) extended kernel spot volatility estimator of Kristensen (2010) to the case when the underlying price process has jumps.

The estimation of spot covariance matrix is, however, an area that has been studied less. For a multi-dimensional continuous semimartingale log-asset price process, Bibinger et al. (2017) proposed an estimator for spot covariance which is constructed based on a local average of block-wise parametric spectral covariance estimates. Aiming to fill this gap in the literature we make the following contribution to the spot covariance estimation for both continuous and discontinuous semimartingales.

For a setup without jumps, we establish asymptotic properties of the kernel covariance estimator, which was mentioned in Kristensen (2010) as an extension to the multivariate case and was left for the future research. The estimator is a kernel-weighted version of the standard integrated covariance estimator, which depends on a kernel function and choice of bandwidth. It can be regarded as a kernel regression in the time domain. The bandwidth choice allows us to focus on the covariance behavior at specific points in time, and give different weights to the covariance matrix over the window used. As the bandwidth shrinks to zero, the spot covariance can be extracted. We prove asymptotic normality of the estimator for both fixed and shrinking bandwidth. The proofs are component-wise. We first derive the mean and the variance of the estimators using the main results of Barndorff-Nielsen and Shephard (2004a) and the lemmas from Kristensen (2010) and then prove the asymptotic normality via the Cramér-Wold device.

Second, we consider the case when the underlying price process is a discontinuous semimartingale with finite activity jumps. We extend the kernel estimator to the threshold kernel covariance estimator and derive the asymptotic distribution of this estimator for a fixed bandwidth. The estimator is an extension to the multivariate case of the threshold kernel volatility estimator proposed by Yu et al. (2014). In the proof of this theorem we combine our results from the first theorem, techniques from Yu et al. (2014) and employ Cramér-Wold device.

The rate of convergence of both kernel and threshold kernel covariance estimators is $n^{-1/2}$. The local method of moments estimator of the spot covariances of Bibinger et al. (2017) attains slower optimal rate of convergence ($n^{-1/8}$). However, it should be noted this is due to the fact that Bibinger et al. (2017) consider the setting with market microstructure noise, whereas we consider a jump case. The kernel and threshold kernel covariance estimators are fairly easy to implement.

We examine with simulated data the finite sample properties of the estimators with different kernel functions using the integrated mean square error and the integrated bias performance measurements. We find that the estimators with one sided kernel performs the best. Furthermore, we study an application of the kernel estimator in the context of covariance forecasting. Considerable efforts has been put into covariance forecasting (see

e.g. Alexander (2018), Andersen et al. (2013)). Multivariate GARCH models are a standard tool used in modelling and forecasting covariances. However, more recent studies propose models based on high-frequency data and options implied data. In a comprehensive empirical study by Symitsi et al. (2018) several approaches to the covariance forecasting are compared based on statistical and economic criteria. It is concluded that models based on high-frequency data offer a clear advantage in terms of statistical accuracy. In particular, a Vector Heterogeneous Autoregressive (VHAR) model achieves the best performance amongst the competing models. The VHAR model is a linear combination of past daily, weekly and monthly realized covariance estimators of Barndorff-Nielsen and Shephard (2004a).

Motivated by this we use the VHAR model to forecast covariance, however instead of the realized we use the kernel covariance estimator. We further show that with the VHAR model the kernel covariance estimator outperforms the benchmark realized covariance estimator in all three measures of accuracy: the Euclidean loss function, the Frobenius distance and the multivariate quasi-likelihood loss function.

The paper is structured as follows. In Section 2.1 we review theoretical setup of the problem and the kernel covariance estimator which was proposed in Kristensen (2010) and left for the future research. In Section 2.2 we study the asymptotic properties of the estimator for a fixed and small (tending to zero) bandwidth. In Section 3 we introduce the setup with jumps, propose the estimator for jump case and derive its asymptotic distribution. In Section 4 we conduct Monte Carlo simulations and investigate the finite sample properties of both estimators. In Section 5 we present an application of the estimator in the context of covariance forecasting. Finally, in Section 6 we summarise our findings.

2 Kernel Covariance Estimation

2.1 Theoretical Setup and the Kernel Covariance Estimator

In this Section we start by considering a multidimensional continuous semimatingale, describe the theoretical setup and review the kernel covariance estimator in Kristensen (2010). Our aim is to accurately estimate the spot covariance matrix of a d-dimensional log-price process $(X(t))_{t\geq 0} = (X_1(t), X_2(t), ..., X_d(t))_{t\geq 0}$. We assume that, for $t \in [0, T]$, $X(t)$ follows a continuous semimartingale

$$\begin{pmatrix} dX_1(t) \\ dX_2(t) \\ \vdots \\ dX_d(t) \end{pmatrix} = \begin{pmatrix} \mu_1(t)dt \\ \mu_2(t)dt \\ \vdots \\ \mu_d(t)dt \end{pmatrix} + \begin{pmatrix} \theta_{11}(t) & \cdots & \theta_{1d}(t) \\ \theta_{21}(t) & \cdots & \theta_{2d}(t) \\ \vdots & \ddots & \vdots \\ \theta_{d1}(t) & \cdots & \theta_{dd}(t) \end{pmatrix} \begin{pmatrix} dW_1(t) \\ dW_2(t) \\ \vdots \\ dW_d(t) \end{pmatrix} \tag{1}$$

defined on a filtered probability space $(\Omega, \mathcal{F}, (\mathcal{F})_{t\geq 0}, P)$, with an initial condition $X(0) \in \mathbb{R}^d$, the drift vector $\mu(t)$, the d-dimensional standard Brownian motion $W(t)$ and the $(d \times d)$-dimensional instantaneous volatility matrix $\theta(t)$ which has elements that are all càdlàg. The latter yields the $(d \times d)$-dimensional *spot covariance matrix* $\Sigma(t) = \theta(t)\theta(t)^\top$, which is our

object of interest. We also denote the *integrated covariance matrix* by $\Sigma^*(t) = \int_0^t \Sigma(s)ds$.

We consider the finite and fixed time horizon $[0, T]$ with $n + 1$ high-frequency discrete observations $X_k(t_0), X_k(t_1), ..., X_k(t_{n-1}), X_k(t_n)$ of the realization of k-th asset, with $k = 1, 2, ..., d$. For an arbitrary partition $0 = t_0 < t_1 < ... < t_n = T$ of the interval $[0, T]$ we require that $\max_{i=1,...,n} |t_i - t_{i-1}|$ approaches zero under the asymptotic limit. For simplicity, we consider the case of equally spaced and synchronous observation times. We denote $\delta = T/n$, so that $t_i = i\delta$ for $i = 1, 2, \cdots, n$.

A kernel is a non-negative integrable function K satisfying the following condition: $\int_{\mathbb{R}} K(u)du = 1$. The kernel weighted measure of the integrated covariance, which is an extension of the measure of the integrated variance introduced in Kristensen (2010), is of the following form

$$KCV(t) = \int_0^T K_h(s - t)\Sigma(s)ds, \tag{2}$$

where the function $K_h(z)$ is given by $K\left(\frac{z}{h}\right)/h$, satisfies $\int_{\mathbb{R}} K(z)dz = 1$, and $h > 0$ is the fixed bandwidth. $KCV(t)$ delivers a kernel weighted quadratic covariation.

An estimator of the integrated covariance in equation (2) is the kernel smoothed sample average of the increments:

$$\widehat{KCV}(t) = \sum_{i=1}^{n} K_h(t_{i-1} - t)\Delta X(t_{i-1})\Delta X^\top(t_{i-1}), \tag{3}$$

where $\Delta X(t_{i-1}) = X(t_i) - X(t_{i-1})$ is the d-dimensional vector (d is fixed) of the increments of the process X over time interval $[t_{i-1}, t_i]$. As introduced above, for a fixed $h > 0$, $KCV(t)$ gives a weighted measure of the integrated covariance. However, as $h \to 0$, the instantaneous covariance can be recovered at any point of continuity t of $t \mapsto \Sigma(t)$:

$$\Sigma(t) = \operatorname*{plim}_{h\to 0} KCV(t), \tag{4}$$

where plim denotes probability limit. To emphasize that we are working with an estimator of the instantaneous covariance at time t, we shall denote:

$$\widehat{\Sigma}(t) = \sum_{i=1}^{n} K_h(t_{i-1} - t)\Delta X(t_{i-1})\Delta X(t_{i-1})^\top. \tag{5}$$

Note that, $\sum_{i=1}^{n} K_h(t_{i-1} - t)\Delta X(t_{i-1})\Delta X(t_{i-1})^\top$ can be regarded as the Nadaraya-Watson estimator (see Nadaraya (1964), Watson (1964)). An overview of these types of kernel can be found in Silverman (1986). In the univariate case, i.e. when $d = 1$, we recover the spot variance estimator from Kristensen (2010).

2.2 Asymptotic Properties of the Kernel Covariance Estimator

In this Section we state the necessary assumptions and present a couple of results: a theorem that derives the asymptotic distribution of the kernel covariance estimator for the fixed

bandwidth, and a theorem that proves asymptotic normality of the kernel covariance estimator for a tending to zero bandwidth. Throughout our work we shall consider the following set of assumptions:

Assumption 1. The processes μ and Σ are jointly independent of W.

Assumption 1 greatly facilitates the proof by allowing us to make all arguments conditional on μ and Σ. Under Assumption 1, the volatility process being independent of W, the model falls into the case without leverage effects. However, this assumption does not appear to be strictly necessary as demonstrated in Kanaya and Kristensen (2016). When examining the performance of the estimator with simulated data, we relax this assumption.

It is convenient to have short expressions for terms that converge in probability to zero or are uniformly tight. The notation $o_P(1)$ is used for a sequence of random variables that converges to zero in probability. The expression $O_p(1)$ denotes a sequence that is bounded in probability (see Van der Vaart, (1998), Aït-Sahalia and Jacod (2014)). More generally, for random variables u_n and $v_n > 0$ we have

$$u_n = o_p(v_n) \quad \text{if} \quad \frac{u_n}{v_n} \to 0 \quad \text{in probability,} \tag{6}$$

$$u_n = O_p(v_n) \quad \text{if} \quad \frac{u_n}{v_n} \quad \text{is bounded in probability,} \quad \text{i.e. for all } \epsilon > 0 \ \exists \text{ an } M \text{ for which } \sup_n P\left(\left|\frac{u_n}{v_n}\right| > M\right) < \epsilon. \tag{7}$$

Assumption 2. For any sequences $(i-1)\delta \leq s_i \leq t_i \leq i\delta$, with $i = 1, \cdots, n$ and every $k = 1, \cdots, d$, as $\delta \to 0$ we have

$$\delta \sum_{i=1}^{n} |\mu_k^2(s_i) - \mu_k^2(t_i)| = o_P(1), \qquad \delta \sum_{i=1}^{n} |\Omega(s_i) - \Omega(t_i)| = o_P(1), \tag{8}$$

where $\Omega(t) =: \{\Sigma_{kk'}(t)\Sigma_{ll'}(t) + \Sigma_{kl'}(t)\Sigma_{lk'}(t)\}_{k,k',l,l'=1,\cdots,d}$.

Assumptions 2 imposes a restriction on the local behavior of the mean and covariance processes. It allows for the deterministic patterns, jumps, and nonstationarity, and is automatically satisfied when the mean and volatility processes have continuous trajectories. In particular, standard diffusion models such as Heston (1993), Hull and White (1987) satisfy this assumption.

Assumption 3. For every $k = 1, \cdots, d$ and $i = 1, \cdots, n$ the quantities

$$\delta^{-1} \int_{t_{i-1}}^{t_i} \Sigma_{kk}(s) ds \tag{9}$$

are bounded away from 0 and infinity uniformly in δ.

Equation (9) in Assumption 3 essentially means that, on any bounded interval, $\Sigma_{kk}(t)$ itself is bounded away from zero and infinity. This is the case, for example for Cox-Ingersoll-Ross (CIR) process in Cox et al. (1985) and Ornstein-Uhlenbeck (OU) process in Uhlenbeck

and Ornstein (1930). The above mentioned assumptions are sufficient to derive the asymptotic distribution of $\widehat{KCV}(t)$, however in order to get the asymptotics of $\widehat{\Sigma}(t)$, when $h \to 0$, the general smoothness condition needs to be imposed on the covariance process.

Assumption 4. The space $C^{m,\gamma}[0,T]$ for some $m \geq 0$ and $0 < \gamma < 1$ consists of functions $f : [0,T] \mapsto \mathbb{R}$ that are m times differentiable with the m-th derivative $f^{(m)}(t)$, satisfying

$$|f^{(m)}(t+\delta) - f^{(m)}(t)| \leq \mathcal{L}(t,|\delta|)|\delta|^{\gamma} + o_P(|\delta|^{\gamma}), \quad \delta \to 0, \quad (a.s.), \tag{10}$$

where $\mathcal{L}(t,\delta)$ is Lipschitz coefficient, a slowly varying function at zero and $t \mapsto \mathcal{L}(t,0)$ is continuous. The mapping $t \mapsto \Sigma_{k,l}(t)$ for $k,l = (1,...,d)$ lies in $C^{m,\gamma}[0,T]$ for some $m \geq 0$ and $\gamma \geq 0$.

As stated in Yu et al. (2014) this condition is satisfied by commonly used diffusion processes. For any model driven by Brownian motion Assumption 5 holds with $m = 0$ and $\gamma < 0.5$ (see e.g. Revuz and Yor (1998)).

We also impose requirements on the kernel function:

Assumption 5. The kernel $K : \mathbb{R} \mapsto \mathbb{R}$

(a) satisfies $\int_{\mathbb{R}} K(x)dx = 1$ and continuously differentiable, i.e. $K \in C^{1,0}$, such that

$$\bar{K}_z :=:= \sup_{0 \leq u \leq T} |K^{(z)}(u)| < \infty, \qquad z = 0, 1.$$

(b) satisfies the condition that there exists some constants Λ, L and $\Gamma_i < \infty$ such that $|K^{(i)}(u)| \leq \Lambda$, and for some $v > 1$, $|K^{(i)}(u)| \leq \Gamma_i |u|^{-v}$ for $|u| \geq L$, $i = 0, 1$.

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

The assumptions above are satisfied by most standard kernels for $r \leq 2$. When $r > 2$, K is called a higher-order kernel. If $m > 2$ as well, the higher-order kernels can be used to reduce the bias in the estimation of more than twice differentiable functions. Although, as mentioned in Kristensen (2010), since $m = 0$ is a usual case, Cline and Hart (1991) demonstrated that higher-order kernels can potentially reduce bias even when the object of interest is non-smooth and has jumps.

We denote by $\xrightarrow{\mathcal{L}}$ *convergence in law* for random variables. If $X_1, X_2, ...$ is a sequence of random variables with cumulative distribution functions (cdf's) $\{F_i\}_{i \geq 1}$ of X_i, we say that $\{X_n\}_n$ converges in law to the random variable X with cdf $F(x)$, written $X_n \xrightarrow{\mathcal{L}} X$, if $F_n(x) \to F(x)$ for every x at which $F(x)$ is continuous.

Now we derive the asymptotics of the kernel covariance estimator for a fixed bandwidth.

Theorem 1. *If Assumptions 1-5 hold, we have that for fixed h and any $t \in [0,T]$*

$$\sqrt{\delta^{-1}} \left\{ vec\left(\widehat{KCV}(t)\right) - vec\left(\int_0^T K_h(s-t)\Sigma(s)ds\right) \right\} \xrightarrow{\mathcal{L}} N\left(0, \int_0^T K_h^2(s-t)\Omega(s)ds\right) \tag{11}$$

(convergence in law for a fixed t). Here $\Omega(t) = \Sigma(t) \otimes \Sigma(t)$ *is a* $d^2 \times d^2$ *array with elements*

$$\Omega(t) =: \{\Sigma_{kk'}(t)\Sigma_{ll'}(t) + \Sigma_{kl'}(t)\Sigma_{lk'}(t)\}_{k,k',l,l'=1,\cdots,d}. \tag{12}$$

Proof. See Appendix A. □

This theorem is an intermediate step in the derivation of the asymptotic distribution of the estimator for a shrinking bandwidth. The Theorem 1 is necessary for the proof of the asymptotic normality of the spot kernel covariance estimator in (5).

Theorem 2. *If Assumptions 1-5 hold with* $r \geq m+\gamma$*, then as* $nh \to \infty$ *and* $nh^{2(m+\gamma)+1} \to 0$ *for any* $t \in (0,T)$ *we have*

$$\sqrt{\delta^{-1}h}\left\{vec\left(\widehat{\Sigma}(t)\right) - vec\left(\Sigma(t)\right)\right\} \xrightarrow{\mathcal{L}} N\left(0, \Omega(t)\int_{\mathbb{R}} K^2(z)dz\right) \tag{13}$$

(convergence in law for a fixed t). Here $\Omega(t)$ *is a* $d^2 \times d^2$ *array with elements*

$$\Omega(t) =: \{\Sigma_{kk'}(t)\Sigma_{ll'}(t) + \Sigma_{kl'}(t)\Sigma_{lk'}(t)\}_{k,k',l,l'=1,\cdots,d}. \tag{14}$$

Proof. See Appendix B. □

For a given rate of smoothness, the highest rate of convergence is $n^{-(m+\gamma)/(2(m+\gamma)+1)}$ when the bandwidth is chosen as $h = O(-1/(2(m+\gamma)+1))$. In particular, when $m = 0$ and $\gamma = 1/2$ we obtain the convergence rate $n^{-1/4}$. Bibinger et al. (2017) propose spot covariance estimator which is constructed based on local averages of block-wise parametric spectral covariance estimates. This is an extension of the local method of moments (LMM) in Bibinger and Reiss (2014). Since Bibinger et al. (2017) consider a setting with market microstructure noise, their estimator attains the optimal rate of convergence ($n^{-1/8}$) which is slower compared to the convergence rate of the kernel covariance estimator ($n^{-1/4}$). The kernel estimator in equation (5) is fairly easy to implement.

Remark 1 (The bivariate case)**.** *It is helpful to focus on the bivariate case in order to gain further understanding. We will look at the results for the assets k and l, whose log-prices will be written as* X_k *and* X_l *respectively. Then the high-frequency returns at time* t_i *is*

$$\Delta X_k(t_i) = X_k(t_i) - X_k(t_{i-1}) \quad \textit{and} \quad \Delta X_l(t_i) = X_l(t_i) - X_l(t_{i-1}) \quad \textit{for } i = 1, \cdots, n.$$

In order to avoid the symmetric replication in the covariation matrix we employ a half-vectorization, or alternatively, a vech transformation. The half-vectorization of a symmetric matrix is obtained by vectorizing only the lower triangular part of the matrix (see Kollo and Rosen (2005), Lütkeohl (1996)). In this case, for a fixed T, Theorem 1 tells us that the joint asymptotic distribution for identifying elements of realized covariation of two assets X_k *and*

X_l becomes

$$\sqrt{\delta^{-1}}\begin{pmatrix} \sum_{i=1}^{n} K_h(t_{i-1}-t)\Delta X_k^2(t_i) - \int_0^T K_h(s-t)\Sigma_{kk}(s)ds \\ \sum_{i=1}^{n} K_h(t_{i-1}-t)\Delta X_k(t_i)\Delta X_l(t_i) - \int_0^T K_h(s-t)\Sigma_{kl}(s)ds \\ \sum_{i=1}^{n} K_h(t_{i-1}-t)\Delta X_l^2(t_i) - \int_0^T K_h(s-t)\Sigma_{ll}(s)ds \end{pmatrix} \xrightarrow{\mathcal{L}}$$
$$N\left[0, \int_0^T K_h^2(s-t)\begin{pmatrix} 2\Sigma_{kk}^2(s) & 2\Sigma_{kk}(s)\Sigma_{kl}(s) & 2\Sigma_{kl}^2(s)) \\ 2\Sigma_{kk}(s)\Sigma_{kl}(s) & \Sigma_{kk}(s)\Sigma_{ll}(s)+\Sigma_{kl}^2(s) & 2\Sigma_{ll}(s)\Sigma_{kl}(s) \\ 2\Sigma_{kl}^2(s) & 2\Sigma_{ll}(s)\Sigma_{kl}(s) & 2\Sigma_{ll}^2(s)) \end{pmatrix} ds\right].$$

3 Extension to the case with jumps

In this Section we assume that the price process is governed by a discontinuous semimartingale with finite activity jumps. We propose a threshold kernel spot covariance estimator. Theorem 3 derives the asymptotic distribution of the threshold kernel covariance estimator for a fixed bandwidth.

Consider a filtered probability space $(\Omega, (\mathcal{F})_{t\in[0,T]}, \mathcal{F}, P)$. Let the d-dimensional log-price $X(t) = (X_1(t), X_2(t), ..., X_d(t))$ be defined on the this space and satisfy the following stochastic differential equation:

$$\begin{pmatrix} dX_1(t) \\ dX_2(t) \\ \vdots \\ dX_d(t) \end{pmatrix} = \begin{pmatrix} \mu_1(t)dt \\ \mu_2(t)dt \\ \vdots \\ \mu_d(t)dt \end{pmatrix} + \begin{pmatrix} \theta_{11}(t) & \cdots & \theta_{1d}(t) \\ \theta_{21}(t) & \cdots & \theta_{2d}(t) \\ \vdots & \ddots & \vdots \\ \theta_{d1}(t) & \cdots & \theta_{dd}(t) \end{pmatrix} \begin{pmatrix} dW_1(t) \\ dW_2(t) \\ \vdots \\ dW_d(t) \end{pmatrix} + \begin{pmatrix} dJ_1(t) \\ dJ_2(t) \\ \vdots \\ dJ_d(t) \end{pmatrix} \tag{15}$$

where $\mu(t)$ is the drift vector, $\theta(t)$ is the instantaneous volatility matrix, $W(t)$ is the d-dimensional Brownian motion and $J(t)$ is a compound Poisson process, which can be written as $J(t) = \sum_{i=1}^{N(t)}(Z_1(t_i), ..., Z_d(t_i)) = \left(\sum_{i=1}^{N(t)} Z_1(t_i), ..., \sum_{i=1}^{N(t)} Z_d(t_i)\right)$. Here $(N(t))_{t\geq 0}$ is a homogeneous Poisson process with constant intensity $\lambda > 0$ and $(Z_k)_{k\in\mathbb{N}}$ is a sequence of i.i.d. random variables with values in $\mathbb{R}^d$, which denotes the jump size at the jump location t_i. We assume $Z_k(t_i)$ for $k = 1, 2, ...d$ are independent of N_t. Denote the $(d \times d)$-dimensional *spot covariance matrix* by $\Sigma(t) = \theta(t)\theta(t)^\top$.

Suppose that on a finite and fixed time horizon $[0, T]$, we have $n + 1$ high-frequency discrete observations $X_k(t_0), X_k(t_1), ..., X_k(t_{n-1}), X_k(t_n)$ of the realization of k-th asset, with $k = 1, 2, ..., d$. Here, $0 = t_0 < t_1 < ... < t_n = T$ is an arbitrary partition of the interval $[0, T]$. Although the observations are not necessarily equidistant, we require that $\max_{i=1,...,n} |t_i - t_{i-1}|$ approaches zero under the asymptotic limit. We consider the case of equally spaced and synchronous observation times, though this assumption can easily be lifted. Denote $\delta = T/n$, so that $t_i = i\delta$ for $i = 1, 2, \cdots, n$.

The quantity of interest is the spot covariance matrix $\Sigma(t)$. The threshold kernel covariance estimator is an extension of threshold kernel volatility estimator of Yu et. al. (2014).

We denote the threshold kernel covariance estimator by $\widehat{TCV}$, and defined it as

$$\widehat{TCV}(t) = \sum_{i=1}^{n} K_h(t_{i-1} - t)\Delta X(t_{i-1})\Delta X^{\top}(t_{i-1})\mathbb{1}_{\{\|\Delta X_{t_{i-1}}\| \leq r(\delta)\}}, \tag{16}$$

where $\mathbb{1}(\cdot)$ is the indicator function and $\Delta X(t_{i-1}) = X(t_i) - X(t_{i-1})$ is the d-dimensional vector of increments of process X over time interval $[t_{i-1}, t_i]$. The function $K_h(x)$ is given by $K(x/h)/h$, where h is bandwidth and the kernel function $K(x)$ satisfies $\int_{\mathbb{R}} K(x)dx = 1$. The threshold function $r(\delta)$ is a deterministic function of the step length δ. As the bandwidth $h \to 0$ we recover the spot covariance. The threshold function $r(\delta)$ has to vanish more slowly than the modulus of the continuity of the Brownian motion in order to have the convergence in probability. Thus we have the following additional assumption.

Assumption 6. *$r(\delta)$ is a deterministic function of the step length δ such that $\lim_{\delta \to 0} r(\delta) = 0$ and $\lim_{\delta \to 0} \frac{\delta \log \frac{1}{\delta}}{r(\delta)} = 0$.*

We now can derive the asymptotics of the threshold kernel covariance estimator.

Theorem 3. *If Assumptions 1-6 hold, we have that for fixed h and any $t \in [0, T]$*

$$\sqrt{\delta^{-1}} \left\{ vec\left(\widehat{TCV}(t)\right) - vec\left(\int_0^T K_h(s-t)\Sigma(s)ds\right)\right\} \xrightarrow{\mathcal{L}} N\left(0, \int_0^T K_h^2(s-t)\Omega(s)ds\right) \tag{17}$$

(convergence in law for a fixed t). where $\Omega(t)$ is a $d^2 \times d^2$ array with elements

$$\Omega(t) =: \{\Sigma_{kk'}(t)\Sigma_{ll'}(t) + \Sigma_{kl'}(t)\Sigma_{lk'}(t)\}_{k,k',l,l'=1,\cdots,d}. \tag{18}$$

Proof. See Appendix C. □

In Theorem 3 we derived asymptotic distribution for the estimator for a fixed bandwidth h of the kernel. The similar results as in Theorem 3 was achieved for univariate case in Yu et al. (2014).

4 Examining the performance of estimators with simulated data

In this Section we examine the performance of the kernel and threshold kernel covariance estimators. In particular, we investigate the finite-sample performances of the estimators relative to the time distance between observations. Throughout we work with bivariate stochastic volatility model. First, we examine the kernel covariance estimator in a setup without jumps and assume that asset prices, $Y(t) = (Y_1(t), Y_2(t))$, follow Heston model:

$$dY(t) = \mu Y(t)dt + \theta(t)Y(t)dW(t), \qquad \Sigma(t) = \theta(t)\theta'(t), \tag{19}$$

Table 1: Interior performance of the KCV estimator

	Gaussian kernel		**One-sided kernel***		Beta kernel	
Data Frequency	IMSE	ISB	**IMSE**	**ISB**	IMSE	ISB
5 seconds	0.14	0.37	**0.11**	**0.21**	0.13	0.28
20 seconds	0.73	0.63	**0.43**	**0.49**	0.66	0.46
1 minute	0.80	0.74	**0.59**	**0.71**	0.76	0.69
5 minutes	1.85	1.97	**1.17**	**1.24**	2.03	1.43
10 minutes	3.88	4.21	**2.16**	**2.14**	2.85	3.16

Note: Integrated mean squared error ($\times 10^{-5}$) and integrated squared bias ($\times 10^{-5}$).

where

$$\Sigma(t) = \begin{pmatrix} \Sigma_{11}(t) & \Sigma_{12}(t) \\ \Sigma_{12}(t) & \Sigma_{22}(t) \end{pmatrix} = \begin{pmatrix} \sigma_1^2(t) & \sigma_{1,2}(t) \\ \sigma_{1,2}(t) & \sigma_2^2(t) \end{pmatrix} \tag{20}$$

with the covariance $\sigma_{1,2}(t) = \sigma_1(t)\sigma_2(t)\rho$, the drift vector $\mu(t) = (\mu_1(t), \mu_2(t))$ and a standard two dimensional Brownian motion $W(t) = (W_1(t), W_2(t))$ such that $d\langle W_1, W_2\rangle_t = \rho dt$. The variance processes, $\sigma_i(t)$ for $i = 1, 2$, follow the CIR model Cox et al. (1985):

$$d\sigma_i^2(t) = \kappa_i(\theta_i - \sigma_i^2(t))dt + \eta_i\sigma_i(t)dZ_i(t). \tag{21}$$

The data generating parameters are chosen to match the estimated parameter values in Barndorff-Nielsen and Shephard (2002). Note that although in Assumption 1 we assume no leverage for theoretical results, in the simulations we include the leverage to demonstrate that the estimation works in realistic model.

In our simulation we set $T = 2$ (48 hours). We consider frequencies $\Delta^{-1} = 12 \times 60 \times 24, 2 \times 60 \times 24, 60 \times 24, 12 \times 24, 6 \times 24$ corresponding to sampling every 5 seconds, 20 seconds, 1 minute, 5 minutes and 10 minutes. In order to simulate the data using model (4) we employ the Euler discretization scheme from Kloeden and Platen (1999). We simulate one trajectory of each $\{\sigma_i^2(t)\}$ for $i = 1, 2$ and keep them fixed. Then we run 500 Monte Carlo repetitions for prices of two assets $\{Y_1(t), Y_2(t)\}$. In each repetition we compute $\hat{\Sigma}_{kl}(t)$ for $i = 1, 2$ based on sampling frequencies.

Three different estimators of instantaneous covariance: Gaussian kernel estimator, one-sided kernel estimator and beta kernel estimator are implemented. For all three estimators cross-validation was used to select the bandwidth (see Kristensen (2010)). We used the following integrated squared error (ISE) as the goodness-of-fit criterion:

$$\text{ISE}(h) = \int_{t_l}^{t_u} \left\| \Sigma(s) - \widehat{\Sigma}(s) \right\|_F^2 ds, \qquad \text{for } \ 0 \le t_l < t_u \le T, \tag{22}$$

where $\|\cdot\|_F$ denotes the Frobenius norm, $\Sigma(s)$ and $\widehat{\Sigma}(s)$ are the true and the estimated spot covariances. Two performance measurements are used to evaluate the finite-sample

Table 2: Interior performance of the TKCV estimator

	Gaussian kernel		**One-sided kernel***		Beta kernel	
Data Frequency	IMSE	ISB	**IMSE**	**ISB**	IMSE	ISB
5 seconds	1.76	1.38	**1.25**	**1.22**	2.34	1.75
20 seconds	2.24	1.13	**1.87**	**1.34**	2.13	2.03
1 minute	3.76	1.45	**2.31**	**1.67**	3.54	2.43
5 minutes	9.35	1.67	**7.31**	**1.35**	3.52	6.67
10 minutes	5.53	1.25	**3.65**	**7.38**	1.83	4.39

Note: Integrated mean squared error ($\times 10^{-5}$) and integrated squared bias ($\times 10^{-5}$).

properties of the estimators: the integrated mean squared error and the integrated bias

$$\text{IMSE} = E\left[\int_{t_l}^{t_u} \left\|\Sigma_{kl}(s) - \widehat{\Sigma}_{kl}(s)\right\|_F^2 ds\right], \quad \text{ISB} = \int_{t_l}^{t_u} \left(\left\|E[\Sigma_{kl}(s) - \widehat{\Sigma}_{kl}(s)]\right\|_F^2 ds\right), \quad (23)$$

where $0 \leq t_l < t_u \leq T$. The results for the performance of the estimator of the covariance, $\widehat{\Sigma}_{12}(t)$, are reported in Table 1. Figure 1 displays QQ plot for observed standardized error terms of Kernel Covariance Estimator using minute-by-minute data.

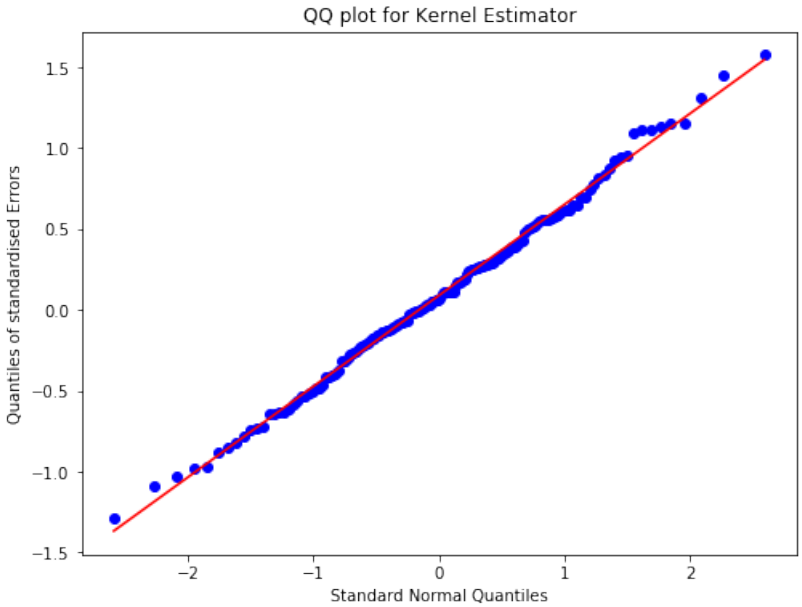

Figure 1: Normal residuals based on QQ plot for observed standardized error terms of Kernel Covariance Estimator using minute-by-minute data

Next, we examine the finite sample performance of the threshold covariance estimator. Though several models combining jumps and stochastic volatility appeared in the literature, we use the model from Bates (1996), one of the most popular examples of the class, an independent jump component is added to the Heston stochastic volatility model:

$$dX(t) = \mu dt + \theta(t)dW(t) + dJ(t), \qquad \Sigma(t) = \theta(t)\theta'(t), \quad (24)$$

with

$$\Sigma(t) = \begin{pmatrix} \Sigma_{11}(t) & \Sigma_{12}(t) \\ \Sigma_{12}(t) & \Sigma_{22}(t) \end{pmatrix} = \begin{pmatrix} \sigma_1^2(t) & \sigma_{1,2}(t) \\ \sigma_{1,2}(t) & \sigma_2^2(t) \end{pmatrix}, \tag{25}$$

where $\sigma_{1,2}(t) = \sigma_1(t)\sigma_2(t)\rho(t)$, $\mu = (\mu_1, \mu_2)$ is the drift vector, $J(t) = \sum_{i=1}^{N(t)}(N_1(t_i), N_2(t_i))$ is a two dimensional compound Poisson jump process and $W(t) = (W_1(t), W_2(t))$ is a two dimensional Brownian motion such that $d\langle W_1, W_2\rangle_t = \rho dt$. The variance processes, $\sigma_i(t)$ for $i = 1, 2$, follow the CIR model:

$$d\sigma_i^2(t) = \kappa_i(\theta_i - \sigma_i^2(t))dt + \eta_i\sigma_i^2(t)dZ_i(t). \tag{26}$$

As in simulations for Heston model without jumps we set $T = 2$ (48 hours) and consider sampling frequencies 5 seconds, 30 seconds, 1 minute. We employ Euler discretization scheme from Kloeden and Platen (1999) for the simulation. We simulate one trajectory of each $\{\sigma_i^2(t)\}$ for $i = 1, 2$ and keep them fixed. Then we run 500 repetitions of $(X_1(t), X_2(t))$. For each simulated path of the bivariate log asset price we compute $\widehat{TKCV}$ based on sampling frequencies. We use the threshold function $r(\delta) = \delta^\alpha$. We set $\alpha = 0.49$ as it yields the best estimates of spot variance for all sample frequencies considered in simulations (see Yu et al. (2014)). We use two IMSE and ISB performance measurements in equation (23)

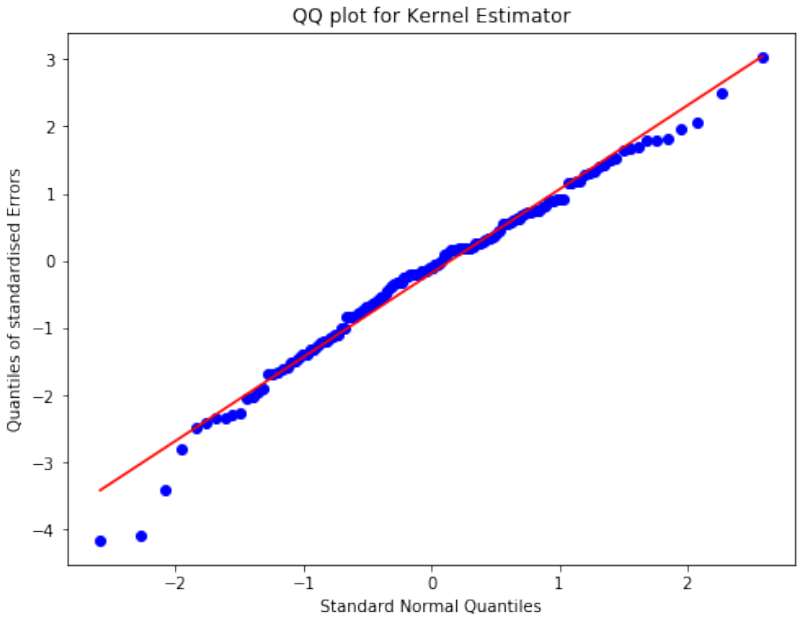

Figure 2: Normal residuals based on QQ plot for observed standardized error terms of Threshold Kernel Covariance Estimator using minute-by-minute data

for three different estimators: Gaussian, beta and one-sided kernel estimator. The results for the performance of the $\widehat{TCV}$ estimator are reported in Table 2. Figure 2 displays QQ plot for observed standardized error terms of Threshold Kernel Covariance Estimator using minute-by-minute data.

5 Application to covariance forecasting

Forecasting covariance has an important economic value in the context of asset pricing and portfolio allocation. Multivariate GARCH model is a standard tool of modelling and forecasting covariances. However, the more recent approaches advocate the use of high-frequency data.

Symitsi et al. (2018) undertake a comprehensive empirical comparison of two generic families of covariance forecasting models: multivariate GARCH models that employ daily data and models that use high-frequency and options data. The authors conclude that models based on high-frequency data offer both a clear advantage in terms of statistical accuracy and yield more theoretically consistent predictions leading to superior out-of-sample portfolio performance. In particular, a Vector Heterogeneous Autoregressive Model (VHAR) achieves the best performance out of the models under consideration. Motivated by this, we use the VHAR model to forecast the integrated covariance, however, when implementing for a finite sample, we use the kernel covariance estimator (3) in Section 2.1 instead of the realized covariance estimator of Barndorff-Nielsen and Shephard (2004a).

Heterogeneous Autoregressive model (HAR), see Corsi (2009), was proposed as a simple way to approximate the long-memory behaviour of volatility. Vector HAR, implemented in Chiriac (2011), is a multivariate extension of HAR. In the VHAR the realized covariance is expressed as a linear combination of past daily, weekly and monthly realized covariances:

$$RC_{t+1} = \alpha + \beta_d RC_t + \beta_w RC_{t-5:t} + \beta_m RC_{t-22:t} + \epsilon_{t+1}, \tag{27}$$

where RC_t is obtained from Cholesky decomposition of realized covariance matrix. If H_t is a matrix of realized covariances, its Cholesky decomposition gives $H_t = C_t C_t'$ and then $RC_t = vech(C_t)$. In order to allow direct comparison among quantities defined over various time horizons, these multiperiod factors are normalized sums of the daily realized factors, i.e.

$$RC_{t-k:t} = \frac{1}{k} \sum_{i=0}^{k-1} RC_{t-i} \tag{28}$$

is the past k day values of RC, α is a constant term and $\beta_d, \beta_w, \beta_m$ are, respectively, the parameters of daily, weekly and monthly components of the model. The covariance forecasts, H_t, are obtained by the reverse transformations of the RC_t's. Modelling the Cholesky factors rather than covariances directly is done in order to avoid unnecessary restrictions that ensure positive definiteness.

We simulate the log-prices of two assets and their volatilises using model (4) in Section 4. Since we use simulated data, we have the true integrated covariance matrix and we propose to forecast the true covariance matrix using two measures of integrated covariance: standard in the literature realized covariance estimator of Barndorff-Nielsen and Shephard (2002) and newly proposed kernel estimator of the covariance in equation (3). Thus we have two models for forecasting integrated covariance. First model is VHAR model where we use the realized

covariance as a measure of integrated covariance:

$$IC_{t+1} = \alpha + \beta_d RC_t + \beta_w RC_{t-5:t} + \beta_m RC_{t-22:t} + \epsilon_{t+1}, \tag{29}$$

where IC is the half-vectorized Cholesky decomposition of the integrated covariance matrix.

In light of this it is natural to define the VHAR-KCV model, in which we borrow the VHAR model above to predict the integrated covariance matrix, however we use kernel covariance estimator:

$$IC_{t+1} = \alpha + \beta_d \widehat{KCV}_t + \beta_w \widehat{KCV}_{t-5:t} + \beta_m \widehat{KCV}_{t-22:t} + \epsilon_{t+1}, \tag{30}$$

where $\widehat{KCV}$ is the half-vectorized Cholesky decomposition of the kernel covariance estimator in (3). We benchmark the VHAR-KCV against the VHAR.

In line with Symitsi et al. (2018) we evaluate forecasting ability of the the VHAR-KCV model (30) based on three multivariate loss functions and compare its performance to the performance of the benchmark VHAR model (29). We use the Euclidean loss function, $\mathcal{L}_E$, which is equally-weighted elements of the forecast error matrix; the Frobenius distance, $\mathcal{L}_F$, which is the extension of the mean squared error to the multivariate space and the multivariate quasi-likelihood loss function, $\mathcal{L}_Q$, which is scale invariant:

$$\mathcal{L}_E = vech(\Sigma_t - H_t)' vech(\Sigma_t - H_t), \tag{31}$$

$$\mathcal{L}_F = Tr[(\Sigma_t - H_t)'(\Sigma_t - H_t)], \tag{32}$$

$$\mathcal{L}_Q = \log|H_t| + Tr(H_t^{-1}\Sigma_t). \tag{33}$$

Here Tr denotes the trace of square matrix, Σ_t denotes the integrated covariance matrix at time t and H_t is time t matrix of conditional covariance forcasts.

Results are reported in Table 3. Based on these results the VHAR-KCV model outperforms the VHAR at all forecasting horizons. In the large study by Symitsi et al. (2018) out of twelve models under consideration the VHAR model was shown to be the best model for forecasting covariance matrix. Thus, the VHAR-KCV is already significant improvement. This improvement maybe due to the fact that the VHAR-KCV model with kernel covariance estimator simply puts higher weight to the more recent data, whereas the VHAR with realized covariance estimator puts equal weight to all data points.

6 Concluding Remarks

In this paper we developed estimators of spot covariances for two types of the underlying price process: continuous and discontinuous semimartingales. We showed the asymptotic normality of the estimators. An important result is that we are able to attain the convergence rate of for a bandwidth tending to zero, which is $n^{-1/4}$. The convergence rate of spot covariance matrix estimator for continuous martingales in a setup with microstructure noise proposed by Bibinger et al. (2017) is, in turn, $n^{-1/8}$. In financially realistic scenarios, we

Table 3: The table reports the out of sample forecast loses for the 1-, 5-, 22-day horizons, respectively. The model with the lowes out-of-sample loss is market with asterisk (*).

	1-day horizon		1-week horizon		2-week horizon	
	$VHAR$	$VHAR-KCV^*$	$VHAR$	$VHAR-KCV^*$	$VHAR$	$VHAR-KCV^*$
α	0.3243	0.3213	0.4987	0.4896	0.4124	0.4126
β_d	0.6904	0.6064	0.2443	0.2032	0.2295	0.2175
β_w	0.6909	0.6028	0.1765	0.1483	0.2257	0.1591
β_m	0.8922	0.8374	0.9007	0.7289	0.5219	0.4328
$\mathcal{L}_E$	0.1267	0.0529	0.1831	0.0772	0.2412	0.1841
$\mathcal{L}_F$	0.1387	0.0546	0.1796	0.0797	0.2981	0.1902
$\mathcal{L}_Q$	-10.143	-14.0537	-9.893	-13.2624	-7.8503	-11.5561

conducted Monte Carlo experiments to study the finite sample properties of our estimators. In addition, we investigated one of the possible applications of the estimator, the forecasting of covariance matrix. We concluded that our estimator performs better in the context of forecasting than the benchmark realized covariance estimator of Barndorff-Nielsen and Shephard (2004a). One of the possible extensions of the estimators is to consider a market-microstructure noise.

References

Aït-Sahalia, Y., and Jacod, J. 2014. *High-Frequency Financial Econometrics.* Princeton: Princeton University Press.

Aït-Sahalia Y., Fan J. and Xiu D. 2010. High-frequency estimates with noisy and asynchronous financial data. *Journal of the American Statistical Association* **105**: 1504-1516.

Alj, A, Azrak, R., and Mélard, G. 2010. An alternative central limit theorem for martingale difference arrays. *SemanticScholar.*

Alexander C. 2008. *Market Risk Analysis (vol. 2): Practical Financial Econometrics.* Chichester: John Wiley & Sons.

Andersen T.G., Bollerslev T., Christoffersen P.F., and Diebold F.X. 2013. Financial risk measurement for financial risk management. *Handbook of the Economics of Finance* **53**: 1127-1220.

Bandi, F.M. and Renò, R. 2016. Price and volatility co-jumps. *Journal of Financial Economics* **119**: 1007-146.

Barndorff-Nielsen O.E. and Shephard N. 2004a. Econometric analysis of realised covariation: high frequency based covariance, regression and correlation in financial economics. *Econometrica* **72**: 885–925.

Barndorff-Nielsen, O.E. and Shephard N. 2002. Econometric analysis of realized volatility and its use in estimating stochastic volatility models. *Journal of the Royal Statistical Society* **64**: 253-280.

Bates D. 1996. Jumps and stochastic volatility: the exchange rate processes implicit in Deutschemark options. *The Review of Financial Studies* **9**: 69-107.

Bibinger M., Hautsch N., Malec P. and Reiss M. 2017. Estimating the spot covariation of asset prices — statistical theory and empirical evidence. *Journal of Business and Economic Statistics*. **37:** 1504-1516.

Bibinger M. and Reiss M. 2014. Spectral estimation of covolatility from noisy observations using local weights. *Scandinavian Journal of Statistics* **6**: 23-50.

Bibinger M., and Winkelmann L. 2015. Econometrics of co-jumps in high-frequency data with noise. *Journal of Econometrics* **184**: 361-378.

Bos C.S., Janus P. and Koopman S.J. 2012. Spot variance path estimation and its application to high-frequency jump testing. *Journal of Financial Econometrics* **10**: 354-389.

Chiriac R. and Voev V. 2011. Modelling and forecasting multivariate realized volatility. *Journal of Applied Econometrics* **26**: 922-947.

Christensen K., Podolskij M. and Vetter M. 2013. On covariation estimation for multivariate continuous Itô semimartingales with noise in non-synchronous observation schemes. *Journal of Multivariate Analysis* **120**: 59-84.

Cline D.B.H. and Hart J.D. 1991. Kernel estimation of densities with discontinuities or discontinuous derivatives. *Statistics* **22**: 69-84.

Corsi F. 2009. A simple approximate long-memory model of realized volatility. *Journal of Financial Econometrics* **7**: 174-196.

Cox J., Ingersoll J. and Ross S. 1985. A theory of the term structure of interest rates. *Econometrica* **53**: 385-407.

Fan J. and Wang Y. 2008. Spot volatility estimation for high-frequency data. *Statistics and Its Interface* **1**: 279-288.

Foster D.P. and Nelson D.B. 1996. Continuous record asymptotics for rolling sample variance estimators. *Econometrica* **64**: 139-174.

Heston S.L. 1993. A closed-form solution for options with stochastic volatility with applications to bond and currency options. *Review of Financial Studies* **6**: 327-343.

Hull J. and White A. 1987. The pricing of options on assets with stochastic volatility. *Journal of Finance* **42**: 281-300.

Kanaya S. and Kristensen D. 2016. Estimation of stochastic volatility models by nonparametric filtering. *Econometric Theory* **32**: 861-916.

Karatzas I. and Shreve S.E. 1999. *Brownian motion and stochastic calculus*. New York: Springer.

Kloeden P. and Platen E. 1999. *Numerical solutions of stochastic differential equations.* Berlin: Springer-Verlag.

Kollo T. and Rosen D. 2005. *Advanced multivariate statistics with matrices.* Dordrecht: Springer.

Kristensen D. 2010. Nonparametric filtering of the realized spot volatility: a kernel-based approach. *Econometric Theory* **26**: 60–93.

Lütkeohl, H. 1996. *Handbook of matrices.* Chichester: John Wiley & Sons Ltd.

Mancini, C. 2009. Non-parametric threshold estimation for models with stochastic diffusion coefficient and jumps. *Scandinavian Journal of Statistics* **36**: 270–296.

Mykland P.A. and Zhang L. 2008. Inference for volatility-type objects and implications for hedging. *Statistics and Its Interface* **1**: 255-278

Nadaraya, E. A. 1964. On estimating regression. *Theory of Probability and Its Applications* **9**: 141–142

Revuz D. and Yor M. 1998. *Continuous martingales and Brownian motion.* Berlin: Springer-Verlag.

Silverman B.W. 1986. *Density estimation for statistics and data analysis.* New York: Chapman and Halls.

Symitsi E., Symeonidis L., Kourtis A and Markellos R. 2018. Covariance forecasting in equity markets. *Journal of Banking and Finance* **96**: 153-168.

Uhlenbeck G.E. and Ornstein L.S. 1930. On the theory of Brownian Motion. *Phys.Rev* **36**: 823-41.

Van der Vaart, A. W. 1998. *Asymptotic Statistics.* Cambridge: Cambridge University Press.

Watson, G.S. 1964. Smooth regression analysis. *Sankhya: The Indian Journal of Statistics, Series A.* **26**: 359–372.

Yu C., Fang Y., Li Z. and Zhao X. 2014. Non-parametric estimation of high-frequency spot volatility for Brownian semimartingale with jumps. *Journal of Time Series Analysis* **35**: 572-591.

Zhang L. 2011. Estimating covariation: Epps effect and microstructure noise. *Journal of Econometrics* **160**: 33-47.

Zu Y.and Boswijk H.P. 2014. Estimating spot volatility with high-frequency financial data. *Journal of Econometrics* **181**: 117-135.

A Proof of Theorem 1

Before presenting the proof, let us first rewrite the Lemma 6 in Kristensen (2010) in terms of the components of covariance matrix.

Lemma 1. *Under Assumption 2 and Assumption 5(a,b), we have for every* $k, l = 1, \cdots, d$

$$(i) \quad \sum_{i=1}^{n} K_h(t_{i-1} - t) \int_{t_{i-1}}^{t_i} \Sigma_{kl}(s)ds = \int_0^T K_h(s-t)\Sigma_{kl}(s)ds + o_P(\delta)\bar{K}_1,$$

$$(ii) \quad \delta^{-1} \sum K_h^2(t_{i-1} - t) \left(\int_{t_{i-1}}^{t_i} \Sigma_{kl}(s)ds \right)^2 = \int_0^T K_h^2(s-t)\Sigma_{kl}^2(s)ds + o_P(1) \times \bar{K}_0 + O_P(\delta) \times \bar{K}_1$$

uniformly over $t \in [0, T]$, *as* $\delta \to 0$.

Proof. See Kristensen (2010). □

The proof of Theorem 1 is component-wise and consists of several steps. First step is to derive the means and covariances of the variates

$$\widehat{KCV}_{kl}(t) = \sum_{i=1}^{n} K_h(t_{i-1} - t)\Delta X_k(t_{i-1})\Delta X_l(t_{i-1}) \tag{34}$$

$$= \sum_{i=1}^{n} K_h(t_{i-1} - t) \left(X_k(t_i) - X_k(t_{i-1})\right) \left(X_l(t_i) - X_l(t_{i-1})\right) \tag{35}$$

with $k, l = 1, 2, \cdots, d$. Next, the Theorem 1 is proved for the case, where the mean processes μ_k $(k = 1, \cdots, d)$ are identically 0. Finally, the latter restriction is lifted. The proof relies on results and techniques employed in Barndorff-Nielsen and Shephard (2004a) and Kristensen (2010).

Notation For the purpose of simplifying the proof we will use index (or equivalently, tensor) notation instead of vector or matrix notation in a similar way to Barndorff-Nielsen and Shephard (2004a). We rewrite the d stochastic processes X_k, $(k = 1, \cdots, d)$ in equation (1) in index notation as

$$X_k(t) = \int_0^t \mu_k(s)ds + \int_0^t \theta_k^a(u)dW_a(s) \tag{36}$$

with initial condition $X_k(0) = 0$. Here

$$\Theta(t) = \{\theta_k^a(t)\}_{k,a=1,2,\cdots,d}.$$

In index notation Einstein summation convention is used, which means if an index variable appears twice in a single expression then it implies summation over that index. Thus (36)

is understood to mean

$$X_k(t) = \int_0^t \mu_k(s)ds + \sum_{a=1}^{d} \int_0^t \theta_k^a(u)dW_a(s) \tag{37}$$

We apply summation convention to indices $a, b, c, d,$ but not to indices $k, l, k', l',$ unless otherwise specified. Furthermore, we write

$$\theta_{kl}^{ab} = \theta_k^a \theta_l^b, \tag{38}$$

with similar notation for other index combination. In (38) no superscripts or subscripts are repeated and so no summation operator is generated. Combining the Einstein summation convention and the notional rule for θ_{kl}^{ab}, the (k, l)th element of the spot covariance matrix of model (1) is

$$\Sigma_{kl}(t) = \theta_{kl}^{aa} = \sum_{a=1}^{d} \theta_k^a(t)\theta_l^a(t), \tag{39}$$

Mean and variances Throughout the rest of this proof we reason conditionally on the paths of θ_t and μ_t: with $\mathcal{H}$ denoting the σ-field generated by all variables (θ_t, μ_t), $t \geq 0$. We have $\mathcal{H} = \bigcup_{i=1}^{\infty} \mathcal{H}_{t_i}$. We condition on $\mathcal{H}$, which is independent of $\mathcal{F}$ by assumption 1. Therefore the Brownian motion, W, is independent of $\mathcal{H}$. We start by computing the conditional expectation of $\widehat{KCV}_{kl}(t)$ in equation (35):

$$\begin{aligned}
\mathrm{E}\left[\widehat{KCV}_{kl}(t)|\mathcal{H}\right] &= \mathrm{E}\left[\sum_{i=1}^{n} K_h(t_{i-1} - t)\left(X_k(t_i) - X_k(t_{i-1})\right)\left(X_l(t_i) - X_l(t_{i-1})\right)|\mathcal{H}\right] \\
&= \sum_{i=1}^{n} K_h(t_{i-1} - t)\mathrm{E}\left[\left(X_k(t_i) - X_k(t_{i-1})\right)\left(X_l(t_i) - X_l(t_{i-1})\right)|\mathcal{H}_{t_{i-1}}\right] \\
&= \sum_{i=1}^{n} K_h(t_{i-1} - t)\int_{t_{i-1}}^{t_i} \theta_{kl}^{aa}(s)ds,
\end{aligned} \tag{40}$$

where the final equation is due to the results of Barndorff-Nielsen and Shephard (2004a) on p.918 :

$$\mathrm{E}\left[\left(X_k(t_i) - X_k(t_{i-1})\right)\left(X_l(t_i) - X_l(t_{i-1})\right)|\mathcal{H}_{t_{i-1}}\right] = \int_{t_{i-1}}^{t_i} \theta_{kl}^{aa}(s)ds. \tag{41}$$

Next, we apply Lemma 1 and have

$$\sum_{i=1}^{n} K_h(t_{i-1} - t)\int_{t_{i-1}}^{t_i} \theta_{kl}^{aa}(s)ds = \int_0^T K_h(s - t)\theta_{kl}^{aa}(s)ds + o(\sqrt{\delta}). \tag{42}$$

Thus

$$\mathrm{E}\left[\widehat{KCV}_{kl}(t)|\mathcal{H}\right] = \int_0^T K_h(s - t)\theta_{kl}^{aa}(s)ds. \tag{43}$$

In order to compute covariance of (35) we use the following results from Barndorff-Nielsen and Shephard (2004a)

$$\mathrm{Cov}\left\{\left[\Delta X_k(t_{i-1})\Delta X_l(t_{i-1}\right],\left[\Delta X_{k'}(t_{i-1})\Delta X_{l'}(t_{i-1}\right]|\mathcal{H}_{t_{i-1}}\right\} = \tag{44}$$

$$\int_{t_{i-1}}^{t_i}\theta^{aa}_{kk'}(s)\theta^{cc}_{ll'}(s)ds + \int_{t_{i-1}}^{t_i}\theta^{aa}_{kl'}(s)\theta^{cc}_{lk'}(s)ds. \tag{45}$$

Now, using the definition of covariance and equations (43), (42) and (44) we have

$$\begin{aligned}
&\mathrm{Cov}\left\{\widehat{KCV}_{kl}(t),\widehat{KCV}_{k'l'}(t)|\mathcal{H}\right\}\\
&=\mathrm{Cov}\left\{\left[\sum_{i=1}^{n}K_h(t_{i-1}-t)\Delta X_k(t_{i-1})\Delta X_l(t_{i-1})\right],\left[\sum_{i=1}^{n}K_h(t_{i-1}-t)\Delta X_{k'}(t_{i-1})\Delta X_{l'}(t_{i-1})\right]\Bigg|\mathcal{H}\right\}\\
&=\mathrm{E}\Bigg\{\left[\sum_{i=1}^{n}K_h(t_{i-1}-t)\Delta X_k(t_{i-1})\Delta X_l(t_{i-1})-\int_0^T K_h(s-t)\theta^{aa}_{kl}(s)ds\right]\\
&\times\left[\sum_{i=1}^{n}K_h(t_{i-1}-t)\Delta X_{k'}(t_{i-1})\Delta X_{l'}(t_{i-1})-\int_0^T K_h(s-t)\theta^{cc}_{k'l'}(s)ds\right]\Bigg|\mathcal{H}\Bigg\}\\
&=\sum_{i=1}^{n}K_h^2(t_{i-1}-t)\Bigg\{\int_{t_{i-1}}^{t_i}\theta^{aa}_{kk'}(s)\theta^{cc}_{ll'}(s)ds+\int_{t_{i-1}}^{t_i}\theta^{aa}_{kl'}(s)\theta^{cc}_{lk'}(s)ds\Bigg\}.
\end{aligned}$$

By applying Lemma 1 and invoking Riemann integration, we get as $\delta\to 0$

$$\begin{aligned}
&\delta^{-1}\sum_{i=1}^{n}K_h^2(t_{i-1}-t)\left\{\int_{t_{i-1}}^{t_i}\theta^{aa}_{kk'}(s)ds\int_{t_{i-1}}^{t_i}\theta^{cc}_{ll'}(s)ds+\int_{t_{i-1}}^{t_i}\theta^{aa}_{kl'}(s)\theta^{cc}_{lk'}(s)ds\right\}\\
&\qquad\qquad\to\int_0^T K_h^2(s-t)\Omega_{kl,k'l'}(s)ds,
\end{aligned}$$

where $\Omega(t)$ is given in (12).

Asymptotic normality To prove the results of Theorem 1 in the case where the mean processes μ_k are identically 0, we apply Cramer-Wold device, i.e. it suffices to show that for any real constants a^{kl} we have, as $\delta\to 0$

$$\begin{aligned}
&\frac{1}{\sqrt{\delta}}\left(\sum_{i=1}^{n}a^{kl}\left[K_h(t_{i-1}-t)\Delta X_k(t_{i-1})\Delta X_l(t_{i-1})-\int_{t_{i-1}}^{t_i}K_h(s-t)\theta^{aa}_{kl}(s)ds\right]\right)\\
&\qquad\qquad\xrightarrow{\mathcal{L}}\mathcal{N}\left(0,a^{kl}a^{k'l'}(\int_0^T K_h^2(s-t)\Omega_{kl,k'l'}(s)ds)\right).
\end{aligned}\tag{46}$$

Here we apply Einstein summation convention also to the indices k, l. By the above calculations,

$$\mathrm{Var}\left\{\sum_{i=1}^{n} a^{kl}\left(K_h(t_{i-1}-t)\Delta X_k(t_{i-1})\Delta X_l(t_{i-1}) - \int_{t_{i-1}}^{t_i} K_h(s-t)\theta_{kl}^{aa}(s)ds\right)\Big|\mathcal{H}\right\}$$
$$\to a^{kl}a^{k'l'}\sum_{i=1}^{n}\int_{t_{i-1}}^{t_i} K_h^2(s-t)\Omega_{kl,k'l'}(s)ds. \quad (47)$$

We apply central limit theorem for martingale difference array from Alj et al. (2010). Martingale array $\{s_{ni}, \mathcal{F}_{ni}, 1 \le i \le k_n\}$ is a zero-mean, square-integrable martingales for each $n \ge 1$. Martingale difference is $y_{ni} = s_{ni} - s_{n,i-1}$. Let $y_n = y_{n1} + ... + y_{nk_n}$. Notice that $E[y_{ni}|\mathcal{F}_{n,i-1}] = 0$ and conditional variance $Var[y_n] = \sum_{j=1}^{k_n} E[y_{nj}^2|\mathcal{F}_{n,j-1}]$.

Theorem 4. *Suppose that*

(a)

$$\sum_{j=1}^{k_n} E[|y_{nj}|^{2+\epsilon}] \to 0 \quad as \quad n \to \infty \quad (48)$$

for some $\epsilon > 0$.

(b)

$$\sum_{j=1}^{k_n} E[y_{nj}^2|\mathcal{F}_{n,j-1}] \to 1 \quad as \quad n \to \infty. \quad (49)$$

Then $y_n \xrightarrow{\mathcal{L}} N(0,1)$.

Proof. See Alj et al. (2010). □

Now we state the Corollary 3 from Barndorff-Nielsen and Shephard (2004a) below.

Corollary 1. *Suppose that* $\mathrm{E}[y_{ni}] = 0$ *for all* n *and* i *and there exists a non-negative number* v *that* $Var[y_n] \to v$ *for* $n \to \infty$. *Then*

$$y_n \xrightarrow{\mathcal{L}} \mathcal{N}(0, v) \quad (50)$$

if and only if (48) *is satisfied.*

Proof. See Barndorff-Nielsen and Shephard (2004a). □

Thus, we only need to show that condition (48) holds. Let

$$y_{ni} = \frac{1}{\sqrt{\delta}} a^{kl}\{K_h(t_{i-1}-t)\Delta X_k(t_{i-1})\Delta X_l(t_{i-1}) - \int_{t_{i-1}}^{t_i} K_h(s-t)\theta_{kl}^{aa}(s)ds\}. \quad (51)$$

We have that $\Delta X_k(t_i) \stackrel{\mathcal{L}}{=} \sqrt{\int_{t_{i-1}}^{t_i} \theta_{kk}^{aa}(s)ds} U_{ki}$, where U_{ki} is a standard normal random variable. Thus we have

$$\begin{aligned} y_{ni} &\stackrel{\mathcal{L}}{=} \frac{1}{\sqrt{\delta}} a^{kl} \Bigg\{ K_h(t_{i-1}-t) \sqrt{\int_{t_{i-1}}^{t_i} \theta_{kk}^{aa}(s)ds} \sqrt{\int_{t_{i-1}}^{t_i} \theta_{ll}^{cc}(s)ds} U_{ki} U_{li} \\ &\qquad - \int_{t_{i-1}}^{t_i} K_h(s-t)\theta_{kl}^{aa}(s)ds \Bigg\} \\ &\stackrel{\mathcal{L}}{=} \frac{1}{\sqrt{\delta}} a^{kl} \Bigg\{ K_h(t_{i-1}-t) \sqrt{\int_{t_{i-1}}^{t_i} \theta_{kk}^{aa}(s)ds} \sqrt{\int_{t_{i-1}}^{t_i} \theta_{ll}^{cc}(s)ds} U_{ki} U_{li} \\ &\qquad - K_h(t_{i-1}-t) \int_{t_{i-1}}^{t_i} \theta_{kl}^{aa}(s)ds \Bigg\} \\ &\stackrel{\mathcal{L}}{=} \frac{1}{\sqrt{\delta}} \Bigg\{ a^{kl} K_h(t_{i-1}-t) \left(\sqrt{\int_{t_{i-1}}^{t_i} \theta_{kk}^{aa}(s)ds} \sqrt{\int_{t_{i-1}}^{t_i} \theta_{ll}^{cc}(s)ds} (U_{ki}U_{li} - \rho_{kl}) \right) \Bigg\} \\ &\stackrel{\mathcal{L}}{=} \sqrt{\delta} a^{kl} \Bigg\{ K_h(t_{i-1}-t) \sqrt{\widehat{\Gamma}_{ki}\widehat{\Gamma}_{li}} \left(U_{ki}U_{li} - \rho_{kl} \right) \Bigg\} \end{aligned} \tag{52}$$

where

$$\hat{\Gamma}_{ki} = \frac{1}{\delta} \int_{t_{i-1}}^{t_i} \theta_{kk}^{aa}(s)ds \tag{53}$$

and

$$\rho_{kl} = \frac{\int_{t_{i-1}}^{t_i} \theta_{kl}^{aa}(s)ds}{\sqrt{\int_{t_{i-1}}^{t_i} \theta_{kk}^{cc}(s)ds} \sqrt{\int_{t_{i-1}}^{t_i} \theta_{ll}^{dd}(s)ds}} \tag{54}$$

is the correlation coefficient between U_k and U_l. By our Assumption on the process Σ, as δ varies the quantities $\hat{\Gamma}$ are bounded away from 0 and infinity, uniformly in k and j. This implies that

$$\mathrm{E}[|a^{kl} K_h(t_{i-1}-t) \sqrt{\widehat{\Gamma}_{ki}\widehat{\Gamma}_{li}} (U_{ki}U_{li} - \rho_{kl}))|^{2+\epsilon} | \mathcal{H}] \tag{55}$$

is uniformly bounded above, and hence, by (51), we have

$$\sum_{i=1}^{n} \mathrm{E}[|y_{ni}|^{2+\epsilon}] \to 0 \tag{56}$$

as to be shown. Next, we show that the effect of a nonzero drift term is negligible.

$$\begin{aligned} \widehat{KCV}_{kl}(t) - \widehat{KCV}_{kl}^{*}(t) &= \sum_{i=1}^{n} K_h(t_{i-1}-t) \left(\int_{t_{i-1}}^{t_i} \mu_k(s)ds \right) \left(\int_{t_{i-1}}^{t_i} \mu_l(s)ds \right) \\ &\quad + \sum_{i=1}^{n} K_h(t_{i-1}-t) \left(\int_{t_{i-1}}^{t_i} \mu_k(s)ds \right) \left(\int_{t_{i-1}}^{t_i} \theta_l(s)dW(s) \right) \\ &\quad + \sum_{i=1}^{n} K_h(t_{i-1}-t) \left(\int_{t_{i-1}}^{t_i} \mu_l(s)ds \right) \left(\int_{t_{i-1}}^{t_i} \theta_k(s)dW(s) \right) \end{aligned} \tag{57}$$

By Lemma 1 the first term in equation (57) is

$$\sum_{i=1}^{n} K_h(t_{i-1}-t)\left(\int_{t_{i-1}}^{t_i}\mu_k(s)ds\right)\left(\int_{t_{i-1}}^{t_i}\mu_l(s)ds\right) = \delta\int_0^T K_h(s-t)\mu_k(s)\mu_l(s)ds + o(\delta).$$

The second term is

$$\begin{aligned}&\sum_{i=1}^{n} K_h(t_{i-1}-t)\left(\int_{t_{i-1}}^{t_i}\mu_k(s)ds\right)\left(\int_{t_{i-1}}^{t_i}\theta_l(s)dW(s)\right)\\ &\sim \mathcal{N}\left(0, \sum_{i=1}^{n} K_h^2(t_{i-1}-t)(\int_{t_{i-1}}^{t_i}\mu_k(s)ds)^2\int_{t_{i-1}}^{t_i}\theta_{ll}^{cc}(s)ds\right)\end{aligned}$$

and, similarly, the third term

$$\begin{aligned}&\sum_{i=1}^{n} K_h(t_{i-1}-t)\left(\int_{t_{i-1}}^{t_i}\mu_l(s)ds\right)\left(\int_{t_{i-1}}^{t_i}\theta_k(s)dW(s)\right)\\ &\sim \mathcal{N}\left(0, \sum_{i=1}^{n} K_h^2(t_{i-1}-t)\left(\int_{t_{i-1}}^{t_i}\mu_l(s)ds\right)^2\int_{t_{i-1}}^{t_i}\theta_{kk}^{aa}(s)ds\right).\end{aligned}$$

where

$$\begin{aligned}&\sum_{i=1}^{n} K_h^2(t_{i-1}-t)\left(\int_{t_{i-1}}^{t_i}\mu_k(s)ds\right)^2\int_{t_{i-1}}^{t_i}\theta_{ll}^{cc}(s)ds\\ &\le \delta\sup_s\theta_{ll}^{cc}(s)\times\sum_{i=1}^{n} K_h^2(t_{i-1}-t)\left(\int_{t_{i-1}}^{t_i}\mu_k(s)ds\right)^2\\ &= \delta^2\sup_s\theta_{ll}^{cc}(s)\times\left(\int_0^T K_h^2(s-t)\mu_k^2(s)ds + o(1)\right)\end{aligned}$$

and

$$\begin{aligned}&\sum_{i=1}^{n} K_h^2(t_{i-1}-t)\left(\int_{t_{i-1}}^{t_i}\mu_l(s)ds\right)^2\int_{t_{i-1}}^{t_i}\theta_{kk}^{aa}(s)ds\\ &\le \delta\sup_s\theta_{kk}^{aa}(s)\times\sum_{i=1}^{n} K_h^2(t_{i-1}-t)\left(\int_{t_{i-1}}^{t_i}\mu_l(s)ds\right)^2\\ &= \delta^2\sup_s\theta_{kk}^{aa}(s)\times\left(\int_0^T K_h^2(s-t)\mu_l^2(s)ds + o(1)\right).\end{aligned}$$

B Proof of Theorem 2

Before presenting the proof, we state the Lemma 7 in Kristensen (2010) in terms of the components of covariance matrix.

Lemma 2. *Under Assumption 4 and Assumption 5, uniformly over* $t \in [a, T-a]$, *as*

$\delta, h, a/h \to 0$ *we have:*

$$(i) \quad \sum_{i=1}^{n} K_h(t_{i-1}-t)\int_{t_{i-1}}^{t_i}\Sigma_{kl}(s)ds = \Sigma_{kl}(t) + h^{m+\gamma}\mathcal{L}(t,0)\int_{\mathbb{R}} K(z)z^{m+\gamma}dz + O_P(\frac{\delta}{h}) + o_P(h^{m+\gamma}),$$

$$(ii) \quad \sum K_h^2(t_{i-1}-t)\left(\int_{t_{i-1}}^{t_i}\Sigma_{kl}(s)ds\right)^2 = \frac{\delta}{h}\Sigma_{kl}^2(t)\int_{\mathbb{R}} K^2(z)dz + O_P(\frac{\delta^{1+\gamma}}{h}) + O_P(\frac{\delta^2}{h^2}).$$

Proof. See Kristensen (2010). □

The convergence results in the proof of Theorem 1 still hold when $h \to 0$. Now, we consider shrinking bandwidth, $h \to 0$, and we derive means and covariances of varieties

$$\hat{\Sigma}_{kl}(t) = \sum_{i=1}^{n} K_h(t_{i-1}-t)\Delta X_k(t_{i-1})\Delta X_l(t_{i-1}) \tag{58}$$

$$= \sum_{i=1}^{n} K_h(t_{i-1}-t)\left(X_k(t_i)-X_k(t_{i-1})\right)\left(X_l(t_i)-X_l(t_{i-1})\right). \tag{59}$$

Following the proof of Theorem 1 and applying Lemma 2 instead of Lemma 1 we obtain:

$$\sum_{i=1}^{n} K_h(t_{i-1}-t)\int_{t_i-1}^{t_i}\theta_{kl}^{aa}(s)ds = \theta_{kl}^{aa} + h^{m+\gamma}\mathcal{L}(t,0)\int_{\mathbb{R}} K(z)z^{m+\gamma}dz + O_P\left(\frac{\delta}{h}\right) + O_P(h^{m+\gamma})$$

where $\mathcal{L}(t,0)$ denotes "Lipschitz coefficient" of $\theta_{kl}(s)$. Thus we have:

$$\mathrm{E}\left[\hat{\Sigma}_{kl}(t)|\mathcal{H}\right] = \Sigma_{kl}(t). \tag{60}$$

For deriving the covariance of the components in (59) we use the following result from proof of Theorem 1:

$$\mathrm{Cov}\left\{\hat{\Sigma}_{kl}(t), \hat{\Sigma}_{k'l'}(t)|\mathcal{H}\right\} =$$
$$\sum_{i=1}^{n} K_h^2(t_{i-1}-t)\left\{\int_{t_{i-1}}^{t_i}\theta_{kk'}^{aa}(s)\theta_{ll'}^{cc}(s)ds + \int_{t_{i-1}}^{t_i}\theta_{kl'}^{aa}(s)\theta_{lk'}^{cc}(s)ds\right\}.$$

Now we using Lemma 2 and invoking Riemann integration for $h \to 0$ we obtain:

$$\delta^{-1}h\sum_{i=1}^{n} K_h^2(t_{i-1}-t)\left\{\int_{t_{i-1}}^{t_i}\theta_{kk'}^{aa}(s)ds\int_{t_{i-1}}^{t_i}\theta_{ll'}^{cc}(s)ds + \int_{t_{i-1}}^{t_i}\theta_{kl'}^{aa}(s)ds\int_{t_{i-1}}^{t_i}\theta_{lk'}^{cc}(s)ds\right\} \to \Omega_{kl,k'l'}(t)\int_{\mathbb{R}} K^2(z)dz.$$

One can easily show the asymptotic normality by following p.21-24 in the proof of Theorem 1 and applying Cramer-Wold device, i.e. to show that for any real constants a^{kl} we have, as

$\delta \to 0$ and $h \to 0$:

$$\sqrt{\delta^{-1}h}(a^{kl}(\widehat{\Sigma}_{kl}(t) - \Sigma_{kl}(t)) \xrightarrow{\mathcal{L}} \mathcal{N}\left(0, a^{kl}a^{k'l'}\Omega_{kl,k'l'}(t)\int_{\mathbb{R}} K^2(z)dz\right). \tag{61}$$

C Proof of Theorem 3

Here we follow the notation in Section A with $\Sigma_{kl}(t)$ denoting the (k,l)-th element of spot covariance matrix at time t (see equation (39)). We first derive the asymptotic distribution of elements $(\widehat{TCV}_{kl})$ for $k, l = 1, ...d$ by following Ye et al. (2014) and then using Cramér-Wold theorem prove multivariate convergence in distribution using univariate results.

Let $\widehat{TCV}_{kl}$ denote the (k,l)-th component of the estimator and X^* denote the diffusion part of X. So, we have

$$\begin{aligned}
& \sqrt{n}\frac{\widehat{TCV}_{kl}(t) - \int_0^T K_h(s-t)\Sigma_{kl}(s)ds}{\sqrt{\int_0^T K_h^2(s-t)\Omega_{kl,k'l'}(s)ds}} \\
= \; & \sqrt{n}\frac{\sum_{i=1}^n K_h(t_{i-1}-t)\Delta_{i-1}X^*_{kl}\Delta_{i-1}X^*_{k'l'}\mathbb{1}_{\{\Delta_{i-1}N=0\}} - \int_0^T K_h(s-t)\Sigma_{kl}(s)ds}{\sqrt{\int_0^T K_h^2(s-t)\Omega_{kl,k'l'}(s)ds}} \\
= \; & \sqrt{n}\frac{\sum_{i=1}^n K_h(t_{i-1}-t)\Delta_{i-1}X^*_{kl}\Delta_{i-1}X^*_{k'l'} - \int_0^T K_h(s-t)\Sigma_{kl}(s)ds}{\sqrt{\int_0^T K_h^2(s-t)\Omega_{kl,k'l'}(s)ds}} \\
- \; & \sqrt{n}\frac{\sum_{i=1}^n K_h(t_{i-1}-t)\Delta_{i-1}X^*_{kl}\Delta_{i-1}X^*_{k'l'}\mathbb{1}_{\{\Delta_{i-1}N\neq 0\}}}{\sqrt{\int_0^T K_h^2(s-t)\Omega_{kl,k'l'}(s)ds}},
\end{aligned} \tag{62}$$

where the the second line is due to results of Mancini (2009) (p.273, Theorem 1). The first term in equation (62) for the fixed h, as $\delta \to 0$ is

$$\sqrt{n}\frac{\sum_{i=1}^n K_h(t_{i-1}-t)\Delta_{i-1}X^*_{kl}\Delta_{i-1}X^*_{k'l'} - \int_0^T K_h(s-t)\Sigma_{kl}(s)ds}{\sqrt{\int_0^T K_h^2(s-t)\Omega_{kl,k'l'}(s)ds}} \xrightarrow{\mathcal{L}} N(0,1). \tag{63}$$

Now, the Assumption 5 states that the kernel K is bounded $|K_h(t_{i-1}-t_i)| \leq \Lambda/h$ for some constant Λ. The number of jumps occurring over the interval $[0,T]$ is finite. Then the second term in equation (62)

$$\begin{aligned}
\sqrt{n}\frac{\left|\sum_{i=1}^n K_h(t_{i-1}-t)\Delta_{i-1}X^*_{kl}\Delta_{i-1}X^*_{k'l'}\mathbb{1}_{\{\Delta_{i-1}N\neq 0\}}\right|}{\sqrt{\int_0^T K_h^2(s-t)\Omega_{kl,k'l'}(s)ds}} & \leq \\
\sqrt{n}\frac{\frac{\Lambda}{h}\sum_{i=1}^n \Delta_{i-1}X^*_{kl}\Delta_{i-1}X^*_{k'l'}\mathbb{1}_{\{\Delta_{i-1}N\neq 0\}}}{\sqrt{\int_0^T K_h^2(s-t)\Omega_{kl,k'l'}(s)ds}} & \leq \\
\sqrt{n}\frac{N_T \times \frac{\Lambda}{h} \times \sup \int_{t_{i-1}}^{t_i}\theta_{kl}(s)dW_s \int_{t_{i-1}}^{t_i}\theta_{k'l'}(s)dB_s}{\sqrt{\int_0^T K_h^2(s-t)\Omega_{kl,k'l'}(s)ds}}.
\end{aligned} \tag{64}$$

Here the integral $\int_0^t \theta_{kl}(s)dW_s$, $\int_0^t \theta_{k'l'}(s)dB_s$ are time changed Brownian motions (see Revuz and Yor (1998), ch.5, thm.9) and by the Levy law of the modulus of continuity of Brownian motion's path Karatzas and Shreve (1999), for small δ we have

$$\sup_{i\in 1,...,n} \frac{\left|\int_{t_{i-1}}^{t_i} \theta_{kl}(s)dW_s\right|}{\sqrt{2\delta \log \frac{1}{\delta}}} \leq \sqrt{M}, \qquad \sup_{i\in 1,...,n} \frac{\left|\int_{t_{i-1}}^{t_i} \theta_{k'l'}(s)dB_s\right|}{\sqrt{2\delta \log \delta^{-1}}} \leq \sqrt{L}, \tag{65}$$

where M, L are a non-negative constants. Therefore the last term in equation (62) is

$$\begin{aligned} &\sqrt{n}\frac{\left|\sum_{i=1}^n K_h(t_{i-1}-t)\Delta_{i-1}X^*_{kl}\Delta_{i-1}X^*_{k'l'}\mathbb{1}_{\{\Delta_{i-1}N\neq 0\}}\right|}{\sqrt{\int_0^T K_h^2(s-t)\Omega_{kl,k'l'}(s)ds}} \leq \\ &\sqrt{n}\frac{N_T \times \frac{\Lambda}{h} \times \sqrt{M} \times \sqrt{L} \times 2\delta \log \delta^{-1}}{\sqrt{\int_0^T K_h^2(s-t)\Omega_{kl,k'l'}(s)ds}} \xrightarrow{P} 0. \end{aligned} \tag{66}$$

To prove multivariate convergence, given that we have asymptotic distribution of the elements of the covariance matrix, we employ Cramér-Wold device:

Lemma 3. *For any real $a \in \mathbb{R}^{d\times d}$, as $\delta \to 0$*

$$\begin{aligned} &\frac{1}{\sqrt{\delta}}\left(\sum_{i=1}^n a^T\left[vec\left(K_h(t_{i-1}-t)\Delta X(t_{i-1})\Delta X^\top(t_{i-1})\mathbb{1}_{\{\|\Delta X_{t_{i-1}}\|\leq r(\delta)\}}\right)\right.\right. \\ &\left.\left. -vec\left(\int_{t_{i-1}}^{t_i} K_h(s-t)\Sigma(t)ds\right)\right]\right) \xrightarrow{\mathcal{L}} N\left(0, a^T\int_0^T K_h^2(s-t)\Omega(s)dsa\right). \end{aligned}$$

Proof. This follows from univariate case, since or $k, l = 1, ..., d$ the variates $\frac{1}{\sqrt{\delta}}\left(\sum_{i=1}^n a^{kl}\left[K_h(t_{i-1}-t)\Delta X(t_{i-1})\Delta X^\top(t_{i-1})\mathbb{1}_{\{\|\Delta X_{t_{i-1}}\|\leq r(\delta)\}} - \int_{t_{i-1}}^{t_i} K_h(s-t)\Sigma_{kl}(s)ds\right]\right)$ are independent and identically distributed with variance $a^{kl}a^{k'l'}\int_0^T K_h^2(s-t)\Omega_{kl,k'l'}(s)ds$ and mean 0. □

IRTG 1792 Discussion Paper Series 2020

For a complete list of Discussion Papers published, please visit http://irtg1792.hu-berlin.de.

001 "Estimation and Determinants of Chinese Banks' Total Factor Efficiency: A New Vision Based on Unbalanced Development of Chinese Banks and Their Overall Risk" by Shiyi Chen, Wolfgang K. Härdle, Li Wang, January 2020.

002 "Service Data Analytics and Business Intelligence" by Desheng Dang Wu, Wolfgang Karl Härdle, January 2020.

003 "Structured climate financing: valuation of CDOs on inhomogeneous asset pools" by Natalie Packham, February 2020.

004 "Factorisable Multitask Quantile Regression" by Shih-Kang Chao, Wolfgang K. Härdle, Ming Yuan, February 2020.

005 "Targeting Cutsomers Under Response-Dependent Costs" by Johannes Haupt, Stefan Lessmann, March 2020.

006 "Forex exchange rate forecasting using deep recurrent neural networks" by Alexander Jakob Dautel, Wolfgang Karl Härdle, Stefan Lessmann, Hsin-Vonn Seow, March 2020.

007 "Deep Learning application for fraud detection in financial statements" by Patricia Craja, Alisa Kim, Stefan Lessmann, May 2020.

008 "Simultaneous Inference of the Partially Linear Model with a Multivariate Unknown Function" by Kun Ho Kim, Shih-Kang Chao, Wolfgang K. Härdle, May 2020.

009 "CRIX an Index for cryptocurrencies" by Simon Trimborn, Wolfgang Karl Härdle, May 2020.

010 "Kernel Estimation: the Equivalent Spline Smoothing Method" by Wolfgang K. Härdle, Michael Nussbaum, May 2020.

011 "The Effect of Control Measures on COVID-19 Transmission and Work Resumption: International Evidence" by Lina Meng, Yinggang Zhou, Ruige Zhang, Zhen Ye, Senmao Xia, Giovanni Cerulli, Carter Casady, Wolfgang K. Härdle, May 2020.

012 "On Cointegration and Cryptocurrency Dynamics" by Georg Keilbar, Yanfen Zhang, May 2020.

013 "A Machine Learning Based Regulatory Risk Index for Cryptocurrencies" by Xinwen Ni, Wolfgang Karl Härdle, Taojun Xie, August 2020.

014 "Cross-Fitting and Averaging for Machine Learning Estimation of Heterogeneous Treatment Effects" by Daniel Jacob, August 2020.

015 "Tail-risk protection: Machine Learning meets modern Econometrics" by Bruno Spilak, Wolfgang Karl Härdle, October 2020.

016 "A data-driven P-spline smoother and the P-Spline-GARCH models" by Yuanhua Feng, Wolfgang Karl Härdle, October 2020.

017 "Using generalized estimating equations to estimate nonlinear models with spatial data" by Cuicui Lu, Weining Wang, Jeffrey M. Wooldridge, October 2020.

018 "A supreme test for periodic explosive GARCH" by Stefan Richter, Weining Wang, Wei Biao Wu, October 2020.

019 "Inference of breakpoints in high-dimensional time series" by Likai Chen, Weining Wang, Wei Biao Wu, October 2020.

IRTG 1792, Spandauer Strasse 1, D-10178 Berlin
http://irtg1792.hu-berlin.de

This research was supported by the Deutsche Forschungsgemeinschaft through the IRTG 1792.

IRTG 1792 Discussion Paper Series 2020

For a complete list of Discussion Papers published, please visit http://irtg1792.hu-berlin.de.

020 "Long- and Short-Run Components of Factor Betas: Implications for Stock Pricing" by Hossein Asgharian, Charlotte Christiansen, Ai Jun Hou, Weining Wang, October 2020.

021 "Improved Estimation of Dynamic Models of Conditional Means and Variances" by Weining Wang, Jeffrey M. Wooldridge, Mengshan Xu, October 2020.

022 "Tail Event Driven Factor Augmented Dynamic Model" by Weining Wang, Lining Yu, Bingling Wang, October 2020.

023 "The common and speci
fic components of infl ation expectation across European countries" by Shi Chen, Wolfgang Karl Härdle, Weining Wang, October 2020.

024 "Dynamic Spatial Network Quantile Autoregression" by Xiu Xu, Weining Wang, Yongcheol Shin, October 2020.

025 "Non-Parametric Estimation of Spot Covariance Matrix with High-Frequency Data" by Konul Mustafayeva, Weining Wang, October 2020.

IRTG 1792, Spandauer Strasse 1, D-10178 Berlin
http://irtg1792.hu-berlin.de

This research was supported by the Deutsche Forschungsgemeinschaft through the IRTG 1792.