Scandinavian Journal of Statistics Theory and Applications

Scandinavian Journal of Statistics, Vol. 43: 1078–1102, 2016
doi: 10.1111/sjos.12230

Inference for Multi-dimensional High-frequency Data with an Application to Conditional Independence Testing

MARKUS BIBINGER and PER A. MYKLAND
Institute of Mathematics, Humboldt University of Berlin and Department of Statistics, University of Chicago

ABSTRACT. We find the asymptotic distribution of the multi-dimensional multi-scale and kernel estimators for high-frequency financial data with microstructure. Sampling times are allowed to be asynchronous and endogenous. In the process, we show that the classes of multi-scale and kernel estimators for smoothing noise perturbation are asymptotically equivalent in the sense of having the same asymptotic distribution for corresponding kernel and weight functions. The theory leads to multi-dimensional stable central limit theorems and feasible versions. Hence, they allow to draw statistical inference for a broad class of multivariate models, which paves the way to tests and confidence intervals in risk measurement for arbitrary portfolios composed of high-frequently observed assets. As an application, we enhance the approach to construct a test for investigating hypotheses that correlated assets are independent conditional on a common factor.

Key words: asymptotic distribution theory, asynchronous observations, conditional independence, high-frequency data, microstructure noise, multivariate limit theorems

1. Introduction

The estimation of daily integrated volatility and covolatility has become a key topic of statistics of high-frequency data and a central building block in model calibration for financial risk analysis. Recent years have seen a tremendous increase in trading activities along with ongoing build-up of computer-based trading. The broad availability of recorded asset prices at such high frequencies magnifies the appeal of statistical methods to efficiently exploit information from the high-frequency data. This article contributes to this strand of literature by considering a continuous semimartingale

$$X_t = X_0 + \int_0^t \mu_s \, ds + \int_0^t \sigma_s \, dW_s \,, t \in \mathbb{R}^+ \,, \tag{1}$$

with drift μ, volatility σ and a standard Brownian motion W, comprising current stochastic volatility models, observed over a fixed time span $[0, T]$ on a discrete grid and by investigating asymptotics when the mesh size of the grid tends to zero. The natural estimator for the quadratic variation (integrated volatility) from equidistant observations of X at $iT/n, i = 0, \ldots, n$ is the discrete version called the realized volatility. In the one-dimensional framework, it gives a consistent estimator that weakly converges with the usual $\sqrt{n}$-rate to a mixed normal distribution where twice the integrated quarticity occurs as random asymptotic variance (cf. Jacod & Protter (1998), Zhang (2001), Barndorff-Nielsen & Shephard (2002)). Therefore, the concept of stable weak convergence by Rényi (1963) has been called into play to pave the way for statistical inference and confidence intervals. In our setting, stable convergence is equivalent to joint weak convergence with every measurable bounded random variable and thus, accompanied by a consistent estimator of the asymptotic variance, allows to conclude a feasible central limit theorem. This reasoning makes stable convergence a key element in high-frequency asymptotic statistics.

The aspiration to progress to more complex statistical models in this research area has been mainly motivated by economic issues. First of all, in a multi-dimensional framework, different assets are usually not traded and recorded at synchronous sampling times but geared to individual observation schemes. Employing simple interpolation approaches has led to the so-called Epps effect (cf. Epps (1979)) that covariance estimates get heavily biased downwards at high frequencies by the distortion from an inadequate treatment of non-synchronicity. In the absence of microstructure, the estimator by Hayashi & Yoshida (2005) remedies this flaw of naively interpolated realized covolatilities, and a feasible central limit theorem has been attained in Hayashi & Yoshida (2011). For synchronous equidistant high-frequency observations of (1), increasing sample sizes are expected to render the estimation error by discretization smaller and smaller. Contrary to the feature of the statistical model, in many situations, high-frequency financial data exhibit an exploding realized volatility when the sampling frequency is too high. This effect is ascribed to market microstructure frictions as bid-ask spreads and trading costs. A favoured way to capture this influence is to extend the classical semimartingale model, where the semimartingale acts to describe dynamics of the evolution of a latent efficient log-price, which is corrupted by an independent additive noise. Following this philosophy from Zhang *et al.* (2005), several integrated volatility estimators have been designed that smooth out noise contamination first. The optimal minimax convergence rate for this model declines to $n^{1/4}$ what is known from the mathematical groundwork provided by Gloter & Jacod (2001). This rate can be attained using the multi-scale realized volatility by Zhang (2006), pre-averaging as described in Jacod *et al.* (2009), the kernel estimator by Barndorff-Nielsen *et al.* (2008) or a quasi-maximum-likelihood approach by Xiu (2010). Although the estimators have been found in independent works and rely on various principles, it turned out that they are in a certain asymptotic sense equivalent, which is clarified in Section 3 below.

Recently, methods to deal with noise and non-synchronicity in one go have been established in the literature. In fact, to each of the aformentioned smoothing techniques (at least) one extension to non-synchronous observation schemes has been proposed. First, the multivariate realized kernels by Barndorff-Nielsen *et al.* (2011) using refresh time sampling are eligible to estimate integrated volatility matrices and guarantee for positive semi-definite estimates at the cost of a sub-optimal convergence rate. Aït-Sahalia *et al.* (2010) suggested to combine a generalized synchronization algorithm with the quasi-maximum-likelihood approach. Park *et al.* (2016) use Fourier methods on the same problem. Eventually, a feasible asymptotic distribution theory for the general non-synchronous and noisy setup has been provided by Bibinger (2012) and Christensen *et al.* (2013) for hybrid approaches built on the Hayashi–Yoshida estimator and the multi-scale and pre-average smoothing, respectively. Although these estimators combine similar ingredients, they behave quite differently, because for the approach in Bibinger (2012), interpolation takes place on the high-frequency scale after smoothing is adjusted with respect to a synchronous approximation, whereas Christensen *et al.* (2013) suggest to denoise each process first and take the Hayashi–Yoshida estimator from pre-averaged blocks, which results in interpolation with respect to a lower-frequency scale.

Presented limit theorems and asymptotic distributions of several of these estimators in the literature are univariate, that is, only the asymptotic variances of (co-)variation estimators are established. An apparent problem pertinent to applications is, however, to quantify the risk of a collection of high-frequently observed assets. When X in (1) is d-dimensional, for instance, estimating the quadratic variation of some portfolio as $w_1 X^{(1)} + w_2 X^{(2)}$ with weights w_1, w_2 is based on estimates for the integrated volatilities and the integrated covolatility. As the three estimates are correlated, we are in need of a multivariate limit theorem to deduce the asymptotic variance of the compound estimator. In this work, we establish multivariate stable limit theorems with the asymptotic variance–covariance matrix of the (generalized) multi-scale estimator and a related realized kernel estimator, along with feasible versions. Thereto, beyond techniques from statistics of high-frequency data, we exploit elements of matrix calculus. Introduce the multivariate notation by the stable central limit theorem for the realized volatility matrix from regular observations as estimator of the integrated volatility matrix $\int_0^T \Sigma_s \, ds$, $\Sigma = \sigma\sigma^\top$:

$$n^{\frac{1}{2}} \operatorname{vec}\left(\sum_{i=1}^{n}\left(X_{\frac{iT}{n}} X_{\frac{(i-1)T}{n}}\right)\left(X_{\frac{iT}{n}} X_{\frac{(i-1)T}{n}}\right)^\top \int_0^T \Sigma_s ds\right) \xrightarrow{st} MN\left(0, T\int_0^T (\Sigma_s \otimes \Sigma_s)\, \mathcal{Z}\, ds\right). \tag{2}$$

The vec-operator transforms the $(d \times d)$ matrix on the left-hand side into a d^2-dimensional vector by stacking the columns below each other:

$$\operatorname{vec}(A) = \left(A^{(11)}, A^{(21)}, \ldots, A^{(d1)}, A^{(12)}, A^{(22)}, \ldots, A^{(d2)}, \ldots, A^{(d(d-1))}, A^{(dd)}\right)^\top \in \mathbb{R}^{d^2},$$

for $A = \left(A^{(pq)}\right)_{1\le p,q\le d} \in \mathbb{R}^{d\times d}$. The mixed normal limit right-hand side comprises a $(d^2 \times d^2)$ random asymptotic variance–covariance matrix with the Kronecker square of Σ. The Kronecker product $A \otimes B \in \mathbb{R}^{d^2\times d^2}$ for $A, B \in \mathbb{R}^{d\times d}$ is defined by

$$(A \otimes B)^{(d(p-1)+q, d(p'-1)+q')} = A^{(pp')} B^{(qq')}, \quad p, q, p', q' = 1, \ldots, d.$$

The matrix $\mathcal{Z}$ describes the variance–covariance structure of the empirical covariance matrix of a standard Gaussian vector

$$\mathcal{Z} = \mathbb{C}\text{ov}(\operatorname{vec}(ZZ^\top)) \in \mathbb{R}^{d^2\times d^2} \text{ for } Z \sim N(0, I_d), \tag{3}$$

with I_d the $(d \times d)$ identity matrix. $\mathcal{Z}$ is explicit, that is, with $\delta_{p,q} = \mathbb{1}_{\{p=q\}}$:

$$\mathcal{Z}^{(d(p-1)+q, d(p'-1)+q')} = (1 + \delta_{p,q})\delta_{\{p,q\},\{p',q'\}}, \; p, q, p', q' = 1, \ldots, d,$$

by the property $\mathcal{Z}\operatorname{vec}(A) = \operatorname{vec}(A + A^\top)$ for all $A \in \mathbb{R}^{d\times d}$. The matrix $\mathcal{Z}$ is twice the so-called symmetrizer matrix from Abadir & Magnus (2005). For realized volatilities to estimate $\int_0^T \sigma_s^2 \, ds$ with σ_s one-dimensional, we recover their well-known asymptotic variance $2T\int_0^T \sigma_s^4 \, ds$. In a two-dimensional setup with volatilities $\sigma_s^{(1)}, \sigma_s^{(2)}$ and a correlation process ρ_s, we derive as limit variance of the realized covolatility $T\int_0^T (1 + \rho_s^2)\left(\sigma_s^{(1)}\sigma_s^{(2)}\right)^2 ds$. Less familiar are the limiting covariances between realized volatility and realized covolatility $2T\int_0^T \rho_s \left(\sigma_s^{(1)}\right)^3 \sigma_s^{(2)} \, ds$ and symmetrically. The form of the asymptotic variance–covariance in (2) is proved in Appendix A.

Relying on the asymptotic distribution of the considered quadratic covariation matrix estimators, we design a statistical test for investigating hypotheses, if two processes have zero covariation conditioned on a third one. We obtain an asymptotic distribution free test. This test, which we call conveniently conditional independence test, renders information about the dependence structure in multivariate portfolios and can be applied to test for zero covariation of idiosyncratic factors in typical portfolio dependence structure models, as the one by Eberlein *et al.* (2008). In particular, we may identify dependencies between single assets not carried in common macroeconomic factors that influence the whole portfolio and disentangle those from correlations induced by market influences.

The outline of the article is as follows. In Section 2, we first unify the asymptotic analysis of quadratic covariation estimation under noise by proving equivalence of methods. Then, multivariate stable limit theorems are developed. Section 3 proceeds to statistical experiments with noise and non-synchronous endogenous observation times. The conditional independence test is introduced in Section 4 and applied in an empirical study in Section 5 to high-frequency financial data. The proofs can be found in the Appendix.

2. Estimating the quadratic covariation matrix in presence of noise

Assumption 1. Consider a continuous d-dimensional Itô semimartingale (1) adapted with respect to a right-continuous and complete filtration $(\mathcal{F}_t)$ on a filtered probability space $(\Omega, \mathcal{F}, (\mathcal{F}_t), \mathbb{P})$ with adapted locally bounded drift process μ, a d-dimensional $(\mathcal{F}_t)$-Brownian motion W and adapted $(d \times d')$ càdlàg volatility process σ. Suppose that σ itself is a continuous Itô semimartingale again, given by an equation similar to (1). The processes σ and W can be dependent, allowing for leverage effect.

The d-dimensional continuous semimartingale X from (1) is discretely observed on $[0, T]$ *with additive noise:*

$$Y_j = X_{t_j} + \epsilon_j \;, j = 0, \ldots, n \;.$$

The synchronous observation times $t_j, 0 \le j \le n$, satisfy

$$\delta_n = \sup_j \left(\left(t_j - t_{j-1}\right), t_0, T - t_n \right) = \mathcal{O}\left(n^{-\frac{8}{9} - \alpha}\right) \tag{4}$$

for a constant $0 < \alpha \le 1/9$, stating that we allow for a maximum time instant tending to zero slower than with n^{-1}, but not too slow. The microstructure noise is given as a discrete-time process for which the observation errors are assumed to be i. i. d. and independent of the efficient process X. Furthermore, the errors have mean zero, and eighth moments exist.

The variance–covariance matrix of $\epsilon_j, 0 \le j \le n$ is denoted by $\mathbf{H}$ and $\mathbb{C}\text{ov}\left(\epsilon_j \epsilon_j^\top\right) = \mathbf{H}^{\otimes}\mathcal{Z}, 0 \le j \le n$. In case that $\epsilon_j \sim N(0, \mathbf{H})$, we already know that $\mathbf{H}^{\otimes} = (\mathbf{H} \otimes \mathbf{H})$, but we allow for much more general noise. We write

$$\Delta_j Y = Y_{t_j} - Y_{t_{j-1}} \text{ and } \Delta_j^i Y = Y_{t_j} - Y_{t_{j-i}}, 1 \le j \le n, 2 \le i \le j \;, \tag{5}$$

for the increments and for increments to longer lags, respectively. An i. i. d. assumption on the noise is standard in related literature, an extension to m-dependence and mixing errors can be attained as in Aït-Sahalia *et al.* (2011). For notational convenience and to find the multivariate analogues of known one-dimensional asymptotic variances of considered estimators, we also restrict ourselves to i. i. d. noise here. Increments in this microstructure noise model

$$\Delta_j Y = \int_{t_{j-1}}^{t_j} \mu_s \, ds + \int_{t_{j-1}}^{t_j} \sigma_s \, dW_s + \epsilon_j - \epsilon_{j-1}$$

are substantially governed by the noise, because any component of the second addend is $\mathcal{O}_{\mathbb{P}}(\delta_n^{1/2})$ and the drift acts only as nuisance term of order in probability $\mathcal{O}_{\mathbb{P}}(\delta_n)$ for each component. For an accurate estimation of the quadratic covariation matrix in the presence of noise, smoothing methods are applied. We now discuss several main approaches and integrate them in a unifying theory. To this end, we show that two prominent methods are asymptotically equivalent.

The asymptotic distributions of considered estimators hinge on the random volatility process σ_s. Thus, stable weak convergence is an essential concept. Let Z_n be a sequence of $\mathcal{X}$-measurable random variables, with $\mathcal{F}_T \subseteq \mathcal{X}$. We say that Z_n *converges stably in law to* Z as $n \to \infty$ if Z is measurable with respect to an extension of $\mathcal{X}$ so that for all $A \in \mathcal{F}_T$ and for all bounded continuous g, $\mathbb{E}[I_A g(Z_n)] \to \mathbb{E}[I_A g(Z)]$ as $n \to \infty$. I_A denotes the indicator function of A, and $=1$ on A and $=0$ otherwise. Here, we have $\mathcal{X}=\mathcal{F}_T$. We refer to Jacod (1997) and Jacod & Protter (1998) for background information on stable convergence for this estimation problem. Stable central limit theorems allow for feasible limit theorems and hence confidence if the asymptotic variance–covariance matrix can be estimated consistently.

2.1. The multivariate multi-scale and kernel estimators

For the estimation of the quadratic variation, the following rate-optimal estimators with similar asymptotic behaviour have been proposed in the literature: the multi-scale approach by Zhang (2006), pre-averaging by Jacod *et al.* (2009), the kernel estimator by Barndorff-Nielsen *et al.* (2008) and a quasi-maximum-likelihood estimator by Xiu (2010). We investigate the variance–covariance structure of the multivariate multi-scale estimator explicitly, but because all these estimators have a similar structure as quadratic form of the discrete observations, analogous reasoning will apply to the other methods. In particular, we shed light on the connection to the kernel approach to profit at the same time from the considerations by Barndorff-Nielsen *et al.* (2008) pertaining parametric efficiency and the asymptotic features of different kernel functions. The multivariate multi-scale estimator

$$\widehat{[X,X]}_T^{(multi)}=\sum_{i=1}^{M_n}\frac{\alpha_i}{i}\sum_{j=i}^{n}\Delta_j^i Y\left(\Delta_j^i Y\right)^{\top} \tag{6}$$

arises as linear combination of averaged lower-frequent realized volatility matrices using frequencies $i=1,\ldots,M_n$. Estimator (6) is the multi-dimensional version of the estimator from Zhang (2006).

For discrete weights $\alpha_i, 1\le i\le M_n$, with $\sum_{i=1}^{M_n}\alpha_i=1$ and $\sum_{i=1}^{M_n}(\alpha_i/i)=0$, the expression

$$\alpha_i=\frac{i}{M_n^2}h\left(\frac{i}{M_n}\right)-\frac{i}{2M_n^3}h'\left(\frac{i}{M_n}\right)+\frac{i}{6M_n^4}(h'(1)-h'(0))-\frac{i}{24M_n^5}(h''(1)-h''(0)), \tag{7}$$

adopted from Zhang (2006), with twice continuously differentiable functions h satisfying $\int_0^1 xh(x)\,dx=1$ and $\int_0^1 h(x)\,dx=0$, gives access to a tractable class of estimators. The multi-scale frequency is chosen $M_n=c\sqrt{n}$ with a constant c, minimizing the overall mean square error to order $n^{-1/4}$. The estimator is thus rate-optimal according to the lower bounds for convergence rates by Gloter & Jacod (2001) and Bibinger (2011).

At the present day, it is commonly known that the non-parametric smoothing approaches to cope with noise contamination have a connatural structure and related asymptotic distributions. A prominent intensively studied alternative to the multi-scale approach is the (realized) kernel estimator

$$\widehat{[X,X]}_T^{(kernel)}\sum_{j=1}^{n}\Delta_j Y(\Delta_j Y)^{\top}\sum_{h=1}^{H_n}\mathfrak{K}\left(\frac{h}{H_n}\right)\left(\sum_{j=h+1}^{n}\Delta_j Y(\Delta_{j-h}Y)^{\top}\Delta_{j-h}Y(\Delta_j Y)^{\top}\right), \tag{8}$$

with a four times continuously differentiable kernel $\mathfrak{K}$ on $[0,1]$, which satisfies the conditions

$$\max\left\{\int_0^1\mathfrak{K}^2(x)\,dx,\int_0^1(\mathfrak{K}'(x))^2\,dx,\int_0^1(\mathfrak{K}''(x))^2dx\right\}<\infty, \mathfrak{K}(0)=1, \mathfrak{K}(1)=\mathfrak{K}'(0)=\mathfrak{K}'(1)=0.$$

This is the multi-dimensional version of the non-flat-top realized kernel estimator considered in Section 4.6 of Barndorff-Nielsen *et al.* (2008). In the one-dimensional setup, (8) has been motivated as linear combination of realized autocovariances of the discretely observed process. The subsequent explicit relation between kernel and multi-scale estimator enables us to embed the findings about several kernels and the construction of an asymptotically efficient one for the parametric model provided by Barndorff-Nielsen *et al.* (2008). Because the multi-scale approach exhibits good finite-sample properties in the treatment of end-effects, it can be worth to road-test resulting transferred multi-scale estimators in practice.

2.2. *Asymptotic equivalence of the multi-scale and kernel estimators*

The multi-scale and kernel estimators defined in (6) and (8) are sensitive to end-effects, which is caused by the dominating noise component whose variance–covariance matrix $\mathbf{H}$ does not depend on n. Because of end-effects, on Assumption 2, the estimators (6) and (8) with weights determined by (7) and corresponding kernels have a bias $-2\,\mathbf{H}$ and $2\,\mathbf{H}$, respectively. We here investigate a correction to each of the two types of estimator:
Correction to multi-scale: Follow Zhang (2006) by modifying the first two weights

$$\alpha_1 \mapsto \alpha_1 + 2/n, \alpha_2 \mapsto \alpha_2 - 2/n, (\alpha_i)_{3\le i\le M_n} \mapsto (\alpha_i)_{3\le i\le M_n}. \tag{9}$$

Correction to the kernel estimator:

$$\text{multiplying the realized volatility matrix in the first addend with } \frac{n-1}{n}. \tag{10}$$

This correction is different from the 'jittering' approach provided in Barndorff-Nielsen *et al.* (2008), Section 2.6. The bias-corrections do not affect the asymptotic variance–covariance structure of the estimators. We call the adjusted estimators, respectively, $\widehat{[X,X]}_T^{(multi,adj)}$ and $\widehat{[X,X]}_T^{(kernel,adj)}$. We then obtain the following direct asymptotic equivalence of the two estimators.

Theorem 2.1. *For each kernel function $\mathfrak{K}$ matching the assertions earlier, for the estimators defined in (6) and (8) with weights determined by (7) and $h = \mathfrak{K}''$, we have*

$$n^{\frac{1}{4}}\left(\widehat{[X,X]}_T^{(multi,adj)} - \widehat{[X,X]}_T^{(kernel,adj)}\right) \xrightarrow{p} 0\,, \tag{11}$$

as $n \to \infty$, $M_n = H_n = c\sqrt{n}$ with some constant c.

Remark 1. (Dependent noise.) In the case of m-dependence, it will be convenient to discard the first m frequencies and renormalize in (6). The adjusted estimator is robust.

Remark 2. (Strong representation.) The result of Theorem 2.1 is similar to other 'strong representation' results in the high-frequency literature, such as in Zhang (2011) (see key equation (39) on p. 41) and Mykland *et al.* (2012), Theorem 4. (The convergence is in probability, but is comparable with strong representation through a standard subsequence-of-subsequence argument.)

Because the motivation of the multi-scale and the kernel approach is quite different, the asymptotic equivalence in Theorem 2.1 is an intriguing result. The equivalence and its proof also reveal how refinements and results for one estimator can be transferred to the other.

2.3. Optimal choice of weights and asymptotic distribution

The standard weights employed in Zhang (2006)

$$\alpha_i = \left(\frac{12i^2}{(M_n^3 - M_n)} - \frac{6i}{(M_n^2 - 1)} - \frac{6i}{(M_n^3 - M_n)} \right) = \frac{12i^2}{M_n^3} - \frac{6i}{M_n^2}(1 + \mathcal{O}(1)) \tag{12}$$

minimize the variance by noise and lead to, as mentioned by Barndorff-Nielsen *et al.* (2008), the same asymptotic properties as for the kernel estimator (8) with a cubic kernel. However, as derived by Barndorff-Nielsen *et al.* (2008), there are kernels surpassing the cubic kernel in efficiency by shrinking the signal and cross parts of the variance while allowing for an increase in the noise variance and striving for the best balance of all three. A fourth term appearing in the asymptotic (co-)variances, see (14), induced by end-effects and noise, can be circumvented by their 'jittering' technique. Asymptotically, Tukey-Hanning kernels as listed in Table 1 combined with this 'jittering' can attain the optimal asymptotic variance in the one-dimensional parametric case known from the inverse Fisher information in Gloter & Jacod (2001). All weights (7) satisfy the relations $\sum_{i=1}^{M_n} \alpha_i = 1$ and $\sum_{i=1}^{M_n} \alpha_i / i = 0$. Classical pre-averaging is asymptotically equivalent to the Parzen kernel. This linkage has been shown by Christensen *et al.* (2010); see also the discussion in Jacod *et al.* (2009) (Remark 1, p. 2255). At this stage, we derive the multivariate stable central limit theorem along with the asymptotic variance–covariance matrix for the equidistant observations setup.

Theorem 2.2. *On the Assumptions 1 and 2 with $t_i = iT/n, 0 \le i \le n$, the multi-scale estimator (6) with $M_n = c\sqrt{n}$, and weights (7), and by the equivalence also the corresponding kernel estimator obey multivariate stable central limit theorems*

$$n^{\frac{1}{4}} \operatorname{vec}\left(\widehat{[X,X]}_T^{(multi)} - \int_0^T \Sigma_s \, ds \right) \xrightarrow{st} MN\left(0, \mathbb{ACOV}\right), \tag{13}$$

with mixed normal limit distribution and with the asymptotic variance–covariance matrix

$$\begin{aligned} \mathbb{ACOV} = {} & 4\mathfrak{D}^\alpha \, c \, T \int_0^T (\Sigma_s \otimes \Sigma_s) \mathcal{Z} \, ds + 2\mathfrak{N}_1^\alpha c^{-3} (\mathbf{H} \otimes \mathbf{H}) \mathcal{Z} \\ & + 2c^{-1}\mathfrak{M}^\alpha \int_0^T (\mathbf{H} \otimes \Sigma_s + \Sigma_s \otimes \mathbf{H}) \mathcal{Z} \, ds + 2c^{-1}\mathfrak{N}_2^\alpha \mathbf{H}^{\otimes} \mathcal{Z}, \end{aligned} \tag{14}$$

with constants $\mathfrak{D}^\alpha$, $\mathfrak{N}_1^\alpha$, $\mathfrak{N}_2^\alpha$ and $\mathfrak{M}^\alpha$ depending on the specific kernel (Table 2).

Table 1. *Collection of important kernels and corresponding weights for the multi-scale*

kernel	$\mathfrak{K}$
cubic	$1 - 3x^2 + 2x^3$
Parzen	$(1 - 6x^2 + 6x^3)\mathbb{1}_{\{x \le 1/2\}} + 2(1-x)^3 \mathbb{1}_{\{x > 1/2\}}$
rth Tukey-Hanning	$\sin\left(\frac{\pi}{2}(1-x)^r\right)^2$
kernel	**first-order weights α_i**
cubic	$\frac{12i^2}{(M)^3} - \frac{6i}{(M)^2}$
Parzen	$\frac{i}{M^2}\left(\frac{36i}{M} - 12\right)$ for $i \le M/2$ and $\frac{i}{M^2}\left(12 - \frac{12i}{M}\right)$ for $i > M/2$
rth Tukey-Hanning	$\frac{\pi i r (1-\frac{i}{M})^{r-2}\left((r-1)\sin(\pi(1-\frac{i}{M})^r) + \pi r (\frac{i}{M}-1)^r \cos(\pi(1-\frac{i}{M})^r)\right)}{2M^2}$

Table 2. *Constants in asymptotic covariance for important kernels*

kernel	$\mathfrak{N}_1^\alpha$	$\mathfrak{D}^\alpha$	$\mathfrak{M}^\alpha$	$\mathfrak{N}_2^\alpha$
cubic	12	13/70	6/5	6/5
Parzen	24	3/4	151/560	15/40
1st Tukey-Hanning	$\pi^4/8$	$\pi^2/16$	3/8	$\pi^2/8$
16th Tukey-Hanning	14374	5.132	0.0317	10.264

A generalization for irregular sampling is covered by Theorem 3.1 in Section 3. In the special case $d = 1$, we obtain the asymptotic variance of the one-dimensional multi-scale estimator as given in Zhang (2006). The last addend involving $\mathbf{H}^{\otimes}\mathcal{Z} = \mathbb{C}\mathrm{ov}\left(\epsilon_j \epsilon_j^\top\right)$ is induced by end-effects and noise and can be circumvented by the jittering technique, see Section 2.6 of Barndorff-Nielsen *et al.* (2008). For the cross terms, note the identity

$$(\mathbf{H} \otimes \Sigma_s + \Sigma_s \otimes \mathbf{H})\,\mathcal{Z} = \mathcal{Z}\,(\mathbf{H} \otimes \Sigma_s)\,\mathcal{Z} = \mathcal{Z}\,(\Sigma_s \otimes \mathbf{H})\,\mathcal{Z}\,.$$

3. Estimating the quadratic covariation matrix under asynchronicity and noise

3.1. Asymptotic distribution of the generalized multi-scale estimator

This section focuses on the general model – comprising non-synchronous observation times and noise perturbation – and a hybrid approach founded on a combination of the estimators from Section 2 and the estimator for non-synchronous non-noisy observations by Hayashi & Yoshida (2005). First, observation times are deterministic or random and independent of Y. In Section 3.2, robustness against endogenous sampling is established.

Assumption 3.1. The process X is observed non-synchronously with additive microstructure noise:

$$Y^{(p)}_{t_j^{(p)}} = X^{(p)}_{t_j^{(p)}} + \epsilon_j^{(p)}\,,\; j = 0,\ldots,n_p,\, p = 1,\ldots,d, \text{ on } [0,T]\,.$$

The sequences of observation times are regular in the sense that $n_p/n_q \to K_{pq}$ with constants $0 < K_{pq} < \infty$. For a constant $0 < \alpha \leq 1/9$, it holds that

$$\delta_n = \sup_{(i,p)}\left(\left(t_i^{(p)} - t_{i-1}^{(p)}\right), t_0^{(p)}, T - t_{n_p}^{(p)}\right) = \mathcal{O}\left(\sup_p (n_p)^{-\frac{8}{9}-\alpha}\right)\,. \tag{15}$$

The observation errors are i. i. d. sequences, independent of the efficient processes, centred and eighth moments exist. Noise components can be mutually correlated only at synchronous observations.

We establish the asymptotic variance–covariance matrix for a generalized multi-scale method proposed in Bibinger (2011). It arises as a convenient composition of the multi-scale estimator from Section 2 and a synchronization approach inspired by the estimator suggested in Hayashi & Yoshida (2005). To handle non-synchronicity, introduce next-tick and previous-tick interpolations:

$$t_p^+(s) = \min_{i\in\{0,\ldots,n_p\}}\left(t_i^{(p)} | t_i^{(p)} \geq s\right) \text{ and } t_p^-(s) = \max_{i\in\{0,\ldots,n_p\}}\left(t_i^{(p)} | t_i^{(p)} \leq s\right)$$

for $p = 1, \ldots, d$ and $s \in [0, T]$. An important synchronous grid is given by the refresh times introduced in Barndorff-Nielsen *et al.* (2011):

$$T_0 = \max_p \left(t_p^+(0) \right), \; T_i = \max_p \left(t_p^+(T_{i-1}) \right), i = 1, \ldots, N\,.$$

For the construction of the estimator, virtually, we can think of an idealized synchronous approximation given by the $(N + 1)$ refresh times, apply subsampling and the multi-scale extension to this scheme and afterwards interpolate to the next observed values on the highest available frequency. This generalized multi-scale estimator is

$$\begin{aligned} \widehat{[X,X]}_T^{(multi)} &= \sum_{i=1}^{M_N} \frac{\alpha_i}{i} \sum_{j=i}^{N} \left(Y_{T_j}^+ - Y_{T_{j-i}}^- \right) \left(Y_{T_j}^+ - Y_{T_{j-i}}^- \right)^\top, \\ \text{with } Y_{T_j}^+ &= \left(Y_{t_p^+(T_j)}^{(p)} \right)_{1 \le p \le d}^\top, \; Y_{T_j}^- = \left(Y_{t_p^-(T_j)}^{(p)} \right)_{1 \le p \le d}^\top, j = 0, \ldots, N. \end{aligned} \tag{16}$$

Without loss of generality, suppose all next-tick interpolations for $j = N$ and previous-tick interpolations for $j = 0$ exist (simply exclude the first and last refresh time else). This estimator crucially differs from the approach by Christensen *et al.* (2013), which mimics the form of the traditional Hayashi–Yoshida estimator but bound to a low-frequency scheme of pre-averaged observations over blocks of order $\sqrt{n}$ high-frequency observations. The estimator (16) relies more on the principle of the refresh-time approximation and exhibits a simpler and for most setups much smaller variance. Contrarily to Barndorff-Nielsen *et al.* (2011), we utilize pre-tick and next-tick interpolations such that the final estimator has no bias because of non-synchronicity. For the reason of various estimators in the general model using different compositions of the methods, the article on hand cannot accomplish a unified theory that is applicable to all alternative approaches as Aït-Sahalia *et al.* (2010), Barndorff-Nielsen *et al.* (2011) and Christensen *et al.* (2013). Unlike their roots from Section 2, they are not asymptotically equivalent any more. We focus on (16) because the method is rate-optimal, and a feasible univariate central limit theorem is accessible from Bibinger (2012).

Remark 3. (Identical results for kernel estimators.) Because Eqns (4) and (15) are the same, it follows from Section 2 that our results on irregular sampling for the synchronous case, where the generalized multi-scale estimator (16) coincides with the original one (6), in the following apply identically to kernel estimators. Furthermore, all results for the estimator (16) apply to a generalized kernel estimator with refresh time sampling as in (16).

Definition 1. For observation times $t_j^{(p)}, 0 \le j \le n_p, 1 \le p \le d$, define the functional sequences

$$\mathfrak{G}_{N,r}(t) = \frac{N}{r} \sum_{T_l \le t} (T_l - T_{l-1}) \sum_{q=1}^{r \wedge l} \left(T_{l-q+1} - T_{l-q} \right), \tag{17}$$

and $S^N(t) \in \mathbb{R}^{d \times d}$ for each $t \in [0, T]$ with entries

$$\left(S^N(t) \right)^{(pq)} = \frac{1}{N} \sum_{T_l \le t} \left(\mathbb{1}_{\left\{ t_p^+(T_l) = t_q^+(T_l) \right\}} + 2\, \mathbb{1}_{\left\{ t_p^+(T_l) = t_q^+(T_{l-1}) \right\}} + \sum_{u=0}^{l} \mathbb{1}_{\left\{ t_p^+(T_l) = t_q^-(T_u) \right\}} \right). \tag{18}$$

Assumption 3.2. Assume that the sequence $\mathfrak{G}_{N,r}$ from (17) and the sequences (18) satisfy the convergences

(i) As $N \to \infty$ and $r \to \infty$ with $r = \mathcal{O}(N)$: $\mathfrak{G}_{N,r}(t) \to \mathfrak{G}(t)$ and $S^N(t) \to S(t)$, for continuous differentiable (in t) limiting functions $\mathfrak{G}$ and S on $[0, T]$.

(ii) For any null sequence (h_N), $h_N = \mathcal{O}\left(N^{-1}\right)$:

$$\frac{\mathfrak{G}_{N,r}(t+h_N)-\mathfrak{G}_{N,r}(t)}{h_N} \to \mathfrak{G}'(t)\,, \quad \frac{S^N(t+h_N)-S^N(t)}{h_N} \to S'(t)\,, \tag{19}$$

uniformly on [0,T] as $N \to \infty$.

(iii) Assume that for all $p, p', q, q' \in \{1, \ldots, d\}$, the following limits exist:

$$\begin{aligned}\chi_{qq'}^{pp'} = \lim_{N\to\infty} \frac{M_N^3}{N} \sum_{i=1}^{M_N} \frac{\alpha_i^2}{i^2} \sum_{j=i+1}^{N} \Big(& \mathbb{1}_{\{t_p^+(T_j)=t_{p'}^+(T_j)\}} \mathbb{1}_{\{t_q^-(T_{j-i})=t_{q'}^-(T_{j-i})\}} \\ & +2\,\mathbb{1}_{\{t_p^+(T_j)=t_{p'}^+(T_{j-1})\}} \mathbb{1}_{\{t_q^-(T_{j-i-1})=t_{q'}^-(T_{j-i-1})\}} \Big)\,.\end{aligned} \tag{20}$$

(iv) Assume the existence of

$$\begin{aligned}\lim_{N\to\infty} M_N^{-1} \Bigg(& \sum_{j=1}^{M_N} \left(\mathbb{1}_{\{t_p^+(T_j)=t_q^+(T_j)\}} + 2\,\mathbb{1}_{\{t_p^+(T_j)=t_q^+(T_{j-1})\}} \right) \\ & + \sum_{j=N-M_N}^{N} \mathbb{1}_{\{t_p^-(T_j)=t_q^-(T_j)\}} \Bigg)\,.\end{aligned} \tag{21}$$

The existence of the limit $\mathfrak{G}$ of $\mathfrak{G}_{N,r}$ is essential to establish an asymptotic distribution theory, because it dominates the terms that appear in the (co-)variances of the multi-scale and related estimators and contribute to the asymptotic (co-)variance, namely, the following existing limit:

$$D^{\alpha}(t) = \lim_{N\to\infty} \left(\frac{N}{M_N} \sum_{T_l \le t} (T_l - T_{l-1}) \sum_{i,k=1}^{M_N} \alpha_i \alpha_k \sum_{q=1}^{\min(l,i,k)} \left(1-\frac{q}{i}\right)\left(1-\frac{q}{k}\right)(T_{l-q+1}-T_{l-q}) \right). \tag{22}$$

In the equidistant synchronous setup $D^{\alpha}(t) = \mathfrak{D}^{\alpha} t\, T$, with the constant $\mathfrak{D}^{\alpha}$ found in Theorem 2.2.

Theorem 3.1. *On the Assumptions 1, 3.1 and 3.2, the generalized multi-scale estimator (16) with $M_N = c\sqrt{N}$ and weights (7) obeys the multivariate stable central limit theorem:*

$$N^{1/4}\left(\widehat{[X,X]}_T^{(multi)} - \int_0^T \Sigma_s\, ds \right) \stackrel{st}{\to} MN\left(0, \mathbb{ACOV}\right)\,, \tag{23}$$

with mixed normal limit distribution and with the asymptotic variance–covariance matrix

$$\begin{aligned}\mathbb{ACOV} = & 4\,c \int_0^T (D^{\alpha})'(s)(\Sigma_s \otimes \Sigma_s)\mathcal{Z}\, ds + 2\,c^{-3}\,(\mathbf{H}\otimes\mathbf{H})^*\,\mathcal{Z} \\ & + c^{-1}\,\mathfrak{M}^{\alpha} \int_0^T \left(\tilde{\mathbf{H}}_s \otimes \Sigma_s + \Sigma_s \otimes \tilde{\mathbf{H}}_s\right)\mathcal{Z}\, ds + c^{-1}\,\mathfrak{N}_2^{\alpha}\,\tilde{\mathbf{H}}^{\otimes}\mathcal{Z}\,,\end{aligned} \tag{24}$$

with (22) and the following existing limits:

$$\left(\tilde{\mathbf{H}}_s\right)^{(pq)} = \mathbf{H}^{(pq)}\left(S'(s)\right)^{(pq)} , \tag{25a}$$

$$\left((\mathbf{H}\otimes\mathbf{H})^*\right)^{(d(p-1)+q\,,\,d(p'-1)+q')} = \mathbf{H}^{(pp')}\mathbf{H}^{(qq')}\chi_{qq'}^{pp'} , \tag{25b}$$

$$\left(\tilde{\mathbf{H}}^{\otimes}\right)^{(pq)} = \left(\mathbf{H}^{\otimes}\right)^{(pq)}\left(\lim_{N\to\infty} M_N^{-1}\sum_{j=1}^{M_N}\left(\mathbb{1}_{\left\{t_p^+(T_j)=t_q^+(T_j)\right\}} + 2\mathbb{1}_{\left\{t_p^+(T_j)=t_q^+(T_{j-1})\right\}}\right) + \lim_{N\to\infty} M_N^{-1}\sum_{j=N-M_N}^{N}\mathbb{1}_{\left\{t_p^-(T_j)=t_q^-(T_j)\right\}}\right) , \tag{25c}$$

for $p, p', q, q' \in \{1, \ldots, d\}$ with S' from (19) and $\chi_{qq'}^{pp'}$ from (20).

In a synchronous setting $(\mathbf{H}\otimes\mathbf{H})^* = \mathfrak{N}_1^{\alpha}\,(\mathbf{H}\otimes\mathbf{H})$ ($\mathfrak{N}_1^{\alpha} = 12$ for the cubic kernel), $\tilde{\mathbf{H}}_s = 2\mathbf{H}$ and $\tilde{\mathbf{H}}^{\otimes} = 2\mathbf{H}^{\otimes}$, and then (24) coincides with (14) except for the influence of irregular sampling. In particular, the asymptotic variance–covariance matrix of the multi-scale estimator for synchronous but non-equidistant sampling coincides with (14), but in the discretization part the derivative of (22), analogously defined for the one observation scheme replaces the constant $\mathfrak{D}^{\alpha}T$.

Interestingly, in most situations, non-diagonal entries of $S(t)$ equal zero as well as $\chi_{qq'}^{pp'}$ whenever $p \neq p'$ or $q \neq q'$, such that the noise part of covariances vanishes. We obtain the following important result for the completely non-synchronous case.

Corollary 3.3. *In the case that no synchronous observations take place, $t_i^{(p)} \neq t_j^{(q)}$ for all i, j and $p \neq q$ (or the amount of synchronous observations tends to zero as $N \to \infty$), (23) holds and (25a), (25b) and (25c) simplify, with $\delta_{p,q} = \mathbb{1}_{\{p=q\}}$, to*

$$\left(\tilde{\mathbf{H}}_s\right)^{(pq)} = 2\,\mathbf{H}^{(pq)}\delta_{p,q} \tag{26a}$$

$$\left((\mathbf{H}\otimes\mathbf{H})^*\right)^{(d(p-1)+q\,,\,d(p'-1)+q')} = \delta_{p,p'}\delta_{q,q'}\mathbf{H}^{(pp)}\mathbf{H}^{(qq)}\mathfrak{N}_1^{\alpha} \tag{26b}$$

$$\left(\tilde{\mathbf{H}}^{\otimes}\right)^{(pq)} = 2\left(\mathbf{H}^{\otimes}\right)^{(pq)}\delta_{p,q} . \tag{26c}$$

Remark 4. Our major focus is not on the theoretical limits $\mathfrak{G}$ and of other sequences, because in the general case, they are specified only as limits. We do not need these values, however, for feasible inference. Convergence of (17) is the natural assumption to derive a central limit theorem for irregularly spaced (non-equidistant) observations already in the one-dimensional framework. It emulates the asymptotic quadratic variation of time for realized volatility to an asymptotic long-run variation of time emerging in the variance for subsampling and the other smoothing approaches. Not directly, the limit of (17) will appear in the asymptotic variance, but some limiting function additionally involves specific weights (the kernel). If we think of random sampling independent of Y, the structure of (17) will be particularly simple for i. i. d. time instants. Virtually, only the expectation will matter, and we can apply the standard law of large numbers. Assuming (19) is less restrictive than the assertion in Zhang (2006). Remarkably, for the popular model of homogenous Poisson sampling independent of Y with expected time instants T/n, the asymptotic variance of the integrated volatility estimator is the same as for equidistant observations. This emanates from the i. i. d. nature of time instants and the vanishing influence of the first addend $2T/(nr)$ in (17) as $r \to \infty$. The finite sample correction factor of the term with $\mathfrak{G}_{N,r}$ in (19) for this Poisson setup is thus $(r+1)/r$.

Remark 5. (Pairwise refresh times) Instead of subsampling geared to the refresh time scheme in (16), we can as well use pairwise refresh times to estimate each entry of the integrated volatility matrix, that is, to estimate $\int_0^T \Sigma_s^{(pq)}\,ds$, we work with refresh times build from $(t_i^{(p)})_{1\le p\le n_p}, (t_i^{(q)})_{1\le q\le n_q}$. Especially in case of very different liquidities, the pairwise estimation can be more efficient in finite samples. The variance–covariance structure for a pairwise generalized multi-scale estimator is slightly more cumbersome – but of the same nature as (24).

At first glance, the simple appearance of the variance–covariance of generalized multi-scale estimates in the typical setup where all observations are non-synchronous is intriguing. It hinges only on the discretization error as if we had synchronous observations at the refresh times $T_i, i = 0, \ldots, N$. The noise falls out of the asymptotic covariances on the assumption that observation errors at different observation times are independent.
This constitutes another nice property of the generalized multi-scale method that a multivariate limit theorem (23) is available and covariances are pretty simple. Here, we benefit from the construction of (16), where interpolation effects and hence the discretization error due to non-synchronicity is asymptotically negligible. This is in line with the result of Bibinger *et al.* (2014) that in this general model with microstructure noise and non-synchronicity, the noise prevails such that the discretization variance–covariance is asymptotically not affected by non-synchronicity.
With a consistent estimator $\widehat{\mathbb{ACOV}} = \widehat{\mathbb{ACOV}\mathcal{Z}}$ of (24), which is provided in the Supporting Information, we derive the feasible multivariate central limit theorem

$$N^{1/4}\widehat{\mathbb{ACOV}}^{-1/2}\left(\widehat{[X,X]}_T^{(multi)} - \int_0^T \Sigma_s\,ds\right) \xrightarrow{st} N(0,\mathcal{Z})\,. \tag{27}$$

3.2. Robustness of multi-scale estimators under endogenous sampling

One crucial limitation of the observation model with Assumption 3.1 is that observation times are supposed to be exogenous and not dependent on the process Y. This appears unrealistic when observations come at random trading times. A prominent contribution in which volatility estimation in presence of endogenous random observation times has been considered is Fukasawa (2010), other works dedicated to endogenous sampling include Li *et al.* (2013, 2014). Especially, the limit theorem for realized volatility by Fukasawa (2010) has attained a lot of attention as the limit law is, in general, different to the case of exogenous sampling. This pointed out that endogeneities can lead to completely new surprising effects and complicate estimators' asymptotic properties. We provide a concise review of the main findings of Fukasawa (2010) in the Supporting Information. Here, we shall reveal that under mild regularity conditions, similar effects of endogeneity do not arise for multi-scale estimators. Thus, the estimation approach is robust against endogeneity of observation times. This finding is in line with recent works by Koike (2014) and Koike (2016) proving that asymptotic properties of pre-average estimators are not affected by endogenous sampling. The different impact of endogenous sampling on the multi-scale estimator compared with realized volatility is due to the smoothing. Realized volatility is the sum of squared increments such that fourth powers of increments trigger its variance. For the multi-scale approach instead, the variance–covariance induced by squared increments is asymptotically negligible and instead cross products of increments over disjoint time segments trigger the (co-)variances of the discretization error (see (43) in the proofs). While the higher moments of increments driving the asymptotics of realized volatility are quite sensitive to endogenities, similar effects do not occur for multi-scale estimators, and generalized Itô isometry implies (co-)variances of the same type as under exogenous observation times.

Assumption 3.4. We have observations at random times $t_i^{(p)}, i = 0, \ldots, n_p, p = 1, \ldots, d$ with $0 < \mathbb{E}[n_p]/\mathbb{E}[n_q] < \infty$. We introduce a sequence of sub-filtrations $(\mathcal{F}_t^N)$ of the augmented $(\mathcal{F}_t)$ such that $t_i^{(p)}, i = 0, \ldots, n_p, p = 1, \ldots, d$ are sequences of $(\mathcal{F}_t^N)$-stopping times and X is adapted to $(\mathcal{F}_t^N)$. For a constant $0 < \alpha \le 1/9$, it holds that

$$\delta_n = \sup_{(i,p)} \left(\left(t_i^{(p)} - t_{i-1}^{(p)} \right), t_0^{(p)}, T - t_{n_p}^{(p)} \right) = \mathcal{O}_{\mathbb{P}} \left(\sup_p (n_p)^{-\frac{8}{9} - \alpha} \right) . \tag{28}$$

Assume as $N \to \infty$ and for $r \to \infty, r = \mathcal{O}(N)$, that

$$\sqrt{\frac{N}{r}} \sum_{l=1}^{N} \left(\int_{T_{l-1}}^{T_l} \mu_s \, ds \right) \sum_{q=1}^{r \wedge l} \left(\int_{T_{l-q-1}}^{T_{l-q}} \sigma_s \, dW_s \right)^{\top} \xrightarrow{p} 0 \,,$$

$$\frac{N}{r} \sum_{l=1}^{N} \left(\int_{T_{l-1}}^{T_l} \Sigma_s \, ds \right) \otimes \left(\sum_{q=1}^{r \wedge l} \left(\int_{T_{l-q-1}}^{T_{l-q}} \Sigma_s \, ds - \sum_{q'=1}^{r \wedge l} \int_{T_{l-q-1}}^{T_{l-q}} \sigma_s \, dW_s \left(\int_{T_{l-q'-1}}^{T_{l-q'}} \sigma_s \, dW_s \right)^{\top} \right) \right) \xrightarrow{p} 0.$$

Furthermore, assume stochastic convergence of the sequences in Assumption 3.2 (i). When the indicator functions in Assumption 3.2 are replaced by $\mathbb{P}\left(t_p^+(T_j) = t_{p'}^+(T_j) | \mathcal{F}_{T_j}^N \right)$ and analogously for the other sets, assume convergence of the respective series. We use the same notation for the limit objects as earlier.

Additional conditions of Assumption 3.4 set mild constraints on the random sampling times and the drift required to obtain a stable central limit theorem of the same type as in the exogenous case. These conditions are ensured, for example, by some constraint on long-range dependence of sampling times, see Assumption [A4](v) of Koike (2016), and by some regularity of the drift as in Assumption [A1] of Koike (2016).

Corollary 3.5. *On the Assumptions 1 and 3.4, as well as Assumption 3.1 on the noise, the generalized multi-scale estimator (16) with deterministic* M_N, $M_N N^{-1/2} \to c$ *for some constant* c, *and weights (7) obeys the multivariate stable central limit theorem (23) with asymptotic variance–covariance matrix (24).*

4. An application to conditional independence testing

This section is devoted to the design of a statistical test in order to investigate if the correlation of two assets is only induced by a factor to which both are correlated. For portfolio modelling and management, information about such relations can provide valuable information. Conclusions that significant integrated covolatilities between high-frequency assets are fully explained by their dependence on a joint factor or another asset, respectively, facilitate dimension reduction of covariance matrix estimation, which is particularly important when considering multivariate limit theorems with variance–covariance matrices of dimension (35).

Beyond this practical implication, our test reveals the dependence structure useful, for example, for default contagion as well as for many other economic applications. For instance, we can think of two observed asset processes X_1 and X_2 listed within one index Z being conditionally on Z independent. To put it the other way round, pairs that are not conditionally independent exhibit significant covariance that carries information about the direct mutual influence. We understand independence here in terms of orthogonal quadratic covariation processes and test for zero integrated covolatility – so the term 'independence' is used here for a simple illustrative phrasing. X_1 and X_2 are orthogonally decomposed in the sum of Z and a process independent of Z. The constants ρ^{X_1}, ρ^{X_2} quantify the degree of dependence on Z.

$$X_1 = \rho^{X_1} Z + Z^{\perp}\,,\ X_2 = \rho^{X_2} Z + Z^{\dagger}\ \text{ with } [Z, Z^{\perp}] \equiv 0\,,\ [Z, Z^{\dagger}] \equiv 0\,. \tag{29}$$

With $[X_1, X_2] \equiv 0$ for two semimartingales X_1, X_2, we express that $[X_1, X_2]_s = 0$ for all $s \in [0, T]$. For the conditional independence hypothesis, we set

$$\mathbb{H}_0 : [Z^{\perp}, Z^{\dagger}]_T = 0\,. \tag{30}$$

Essentially, we do not distinguish between pairs for which the orthogonal parts are uncorrelated on the whole line and pairs for which this correlation process integrates to zero. Our focus is on a resulting zero quadratic covariation over $[0, T]$.
A suitable test statistic to decide whether we reject $\mathbb{H}_0$ or not is

$$\mathfrak{T}(X_1, X_2, Z) = [X_1, Z]_T [X_2, Z]_T - [X_1, X_2]_T [Z, Z]_T\,, \tag{31}$$

which is zero under $\mathbb{H}_0$.
In our high-frequency framework, we can estimate the single integrated (co-)volatilities via the approaches considered in the preceding sections. The vital point is to deduce the asymptotic distribution of the estimated version

$$\hat{\mathfrak{T}}_n = \widehat{[X_1, Z]}_T^{(multi)} \widehat{[X_2, Z]}_T^{(multi)} - \widehat{[X_1, X_2]}_T^{(multi)} \widehat{[Z, Z]}_T^{(multi)}\,, \tag{32}$$

with one of the estimators (6) or (16). This test statistic is more complex to analyse than linear combinations, because we face products of our estimators. Therefore, the asymptotic law of (31) is not directly obtained from Theorem 2.2 or Theorem 3.1, respectively. In lieu of determining the distribution of the test statistic, we apply the Δ-method for stable convergence.

$$\begin{aligned}\mathfrak{T} - \hat{\mathfrak{T}}_n = {} & [X_2, Z]_T \left([X_1, Z]_T - \widehat{[X_1, Z]}_T^{(multi)}\right) + [X_1, Z]_T \left([X_2, Z]_T - \widehat{[X_2, Z]}_T^{(multi)}\right) \\ & - [X_1, X_2]_T \left([Z, Z]_T - \widehat{[Z, Z]}_T^{(multi)}\right) - [Z, Z]_T \left([X_1, X_2]_T - \widehat{[X_1, X_2]}_T^{(multi)}\right) \\ & + \mathcal{O}_{\mathbb{P}}\left(n^{-\frac{1}{2}}\right)\,.\end{aligned} \tag{33}$$

The asymptotic variance of the test statistic is random as a linear combination of the quadratic (co-)variations and entries of the asymptotic variance–covariance matrix. Denote by $\mathbf{AVAR}(U)$ and $\mathbb{ACOV}(U, V)$ in the sequel asymptotic variances and covariances of one-dimensional random variables U, V. An elementary calculation yields

$$\begin{aligned}\mathbf{AVAR}(\hat{\mathfrak{T}}_n) &= [X_2,Z]_T^2\,\mathbf{AVAR}\left(\widehat{[X_1,Z]}_T^{(multi)}\right) + [X_1,Z]_T^2\,\mathbf{AVAR}\left(\widehat{[X_2,Z]}_T^{(multi)}\right)\\ &+ [X_1,X_2]_T^2\,\mathbf{AVAR}\left(\widehat{[Z,Z]}_T^{(multi)}\right) + [Z,Z]_T^2\,\mathbf{AVAR}\left(\widehat{[X_1,X_2]}_T^{(multi)}\right)\\ &+ 2\,[Z,Z]_T[X_1,X_2]_T\,\mathbb{ACOV}\left(\widehat{[X_1,X_2]}_T^{(multi)},\widehat{[Z,Z]}_T^{(multi)}\right)\\ &+ 2\,[X_1,Z]_T[X_2,Z]_T\,\mathbb{ACOV}\left(\widehat{[X_1,Z]}_T^{(multi)},\widehat{[X_2,Z]}_T^{(multi)}\right)\\ &- 2\,[X_1,Z]_T[Z,Z]_T\,\mathbb{ACOV}\left(\widehat{[X_1,X_2]}_T^{(multi)},\widehat{[X_2,Z]}_T^{(multi)}\right)\\ &- 2\,[X_2,Z]_T[Z,Z]_T\,\mathbb{ACOV}\left(\widehat{[X_1,X_2]}_T^{(multi)},\widehat{[X_1,Z]}_T^{(multi)}\right)\\ &- 2\,[X_1,X_2]_T[X_1,Z]_T\,\mathbb{ACOV}\left(\widehat{[X_2,Z]}_T^{(multi)},\widehat{[Z,Z]}_T^{(multi)}\right)\\ &- 2\,[X_1,X_2]_T[X_2,Z]_T\,\mathbb{ACOV}\left(\widehat{[X_1,Z]}_T^{(multi)},\widehat{[Z,Z]}_T^{(multi)}\right).\end{aligned}$$

Inserting consistent estimators for the asymptotic (co-)variances, we obtain with our multivariate stable central limit theorem that

$$n^{\frac{1}{4}}\left(\widehat{\mathbf{AVAR}\left(\hat{\mathfrak{T}}_n\right)}\right)^{-1/2}\hat{\mathfrak{T}}_n \overset{st}{\to} N(0,1)\,, \tag{34}$$

or with scaling $N^{1/4}$ for non-synchronous observations, under $\mathbb{H}_0$ what gives an asymptotic distribution free test.

The role of Z in the model can be also some macro variable that is either known or can be estimated with faster rate of convergence, which simplifies the terms earlier. For regularly observed high-frequency data without noise, the same kind of test can be constructed using the realized volatility matrix, and the faster rate $n^{1/2}$ is attained.

5. An empirical example

We survey our methods in an application study on NASDAQ intra-day trading data, reconstructed from first-level order book data from August 2015. We consider a sample portfolio with five assets, namely, Apple (AAPL), Microsoft (MSFT), Oracle (ORCL), Exxon Mobil Corporation (XOM) and Pfizer (PFE). Traded prices are recorded at non-synchronous times, and market microstructure noise is clearly indicated such that we suppose the model from Assumption 3.1. We quantify the integrated volatility matrix over the whole month (where we discard overnight returns) and for the first trading day, 3 August 2015, respectively, using generalized multi-scale estimates (16) with weights (12) and pairwise refresh times. Precise selection of tuning parameters for the univariate asymptotic variance estimation is outlined in Algorithm 2 of Bibinger (2012). Pre-analysing diagonal entries according to this algorithm, in view of robustness of the method against moderate changes of M_N, we set for simplicity $M_N = 0.2\sqrt{N}$ permanently for the analysis here. In order to infer covariances between covariation estimates, we use for each entry the number of refresh times of all involved assets as N and the histogram multi-scale approach with tuning parameters given dependent on N and M_N in Algorithm 2 of Bibinger (2012) and in the Supporting Information. The complete variance–covariance matrix of the estimates is quantified. For a d-dimensional portfolio, the number of free entries of this symmetric variance–covariance matrix is given by

$$\frac{1}{2}\frac{d(d+1)}{2}\left(\frac{d(d+1)}{2}+1\right)=d+3\binom{d}{4}+3\cdot 2\binom{d}{3}+4\binom{d}{2}. \tag{35}$$

In Table 3, we list the estimates for the integrated volatility matrices ± estimated standard deviations. The estimated variance–covariance matrices, not rescaled with the rates, of these estimates are listed in Table 4. One key insight is that involving covariances of estimates is indispensable when facing questions for multivariate portfolio management. The estimated quadratic variation of a sum of all five assets is $92.04 \cdot 10^{-3}$ for August 2015 and $90.16 \cdot 10^{-5}$ for 3 August 2015. The risk of the estimated volatilities for these portfolios, $4.88 \cdot 10^{-6}$ and $12.72 \cdot 10^{-10}$, is mainly induced by covariances (3.97/8.96), whereas the trace of the variance–covariance matrix, that is, the sum of estimated variances, is much smaller. If one would mistakenly act as if the estimators were uncorrelated, this leads to a tremendous underestimate of uncertainty.

We perform the test from Section 4 to investigate three hypotheses: if MSFT and ORCL have a zero covariation conditional on XOM; ORCL and XOM conditional on MSFT and MSFT and XOM conditional on ORCL. We obtain the following p-values as test results

$$p = 1.27 \cdot 10^{-37}; 2.26 \cdot 10^{-114}; 0.02 \text{ (August 2015)},$$

$$p = 4.24 \cdot 10^{-6};\ 0.87\ ;\ 2.62 \cdot 10^{-4} \text{ (3 August 2015)}.$$

In conclusion, this empirical evidence suggests that on 3 August 2015 MSFT and ORCL as well as MSFT and XOM have some dependence not explained by conditioning on the third asset (and further tests show that also not by conditioning on all others). On the contrary, we can not reject a zero covariation for XOM and ORCL conditional on MSFT. For August 2015, the hypothesis that XOM and MSFT have conditionally on ORCL zero covariation is the only one with non-vanishing, but still small, p-value. This gives a heuristic that the portfolio dependence structure is not completely persistent. Although there are some limitations where the additive noise model does not perfectly fit the stylized facts of the considered high-frequency data as discreteness of returns and zero returns, the approaches developed in this research area and advancements of this article provide reliable tools to quantify risk measures from high-frequency asset prices and to determine confidence intervals for the estimates.

Table 3. *Estimates for the integrated volatility matrix* ($\cdot$ $\mathbf{10^3}$) *2015/08 (top) and* ($\cdot$ $\mathbf{10^5}$) *2015/08/03 (bottom)*

$\widehat{[X,X]}_T^{(multi)}$	AAPL	MSFT	ORCL	XOM	PFE
AAPL	9.46 ± 0.01	4.14 ± 0.02	3.68 ± 0.03	2.76 ± 0.01	4.51 ± 0.03
MSFT		6.38 ± 0.03	3.59 ± 0.05	2.55 ± 0.02	4.21 ± 0.07
ORCL			4.39 ± 0.06	2.90 ± 0.03	0.33 ± 0.10
XOM				4.74 ± 0.01	1.68 ± 0.04
PFE					6.37 ± 0.09

$\widehat{[X,X]}_T^{(multi)}$	AAPL	MSFT	ORCL	XOM	PFE
AAPL	20.58 ± 0.80	4.22 ± 0.51	1.79 ± 0.41	2.03 ± 0.65	1.64 ± 0.39
MSFT		10.09 ± 0.47	2.00 ± 0.34	2.29 ± 0.60	2.06 ± 0.34
ORCL			4.83 ± 0.26	0.53 ± 0.45	0.98 ± 0.26
XOM				13.25 ± 0.83	0.85 ± 0.44
PFE					4.63 ± 0.27

Table 4. *Estimated covariance matrix of estimates from Table 3 (* $\cdot\, \mathbf{10^8}$ *), 2015/08 (top) and (* $\cdot\, \mathbf{10^{12}}$ *) 2015/08/03 (below), A=AAPL, M=MSFT, O=ORCL, X=XOM, P=PFE*

	[A,A]	[A,M]	[A,O]	[A,X]	[A,P]	[M,M]	[M,O]	[M,X]	[M,P]	[O,O]	[O,X]	[O,P]	[X,X]	[X,P]	[P,P]
[A,A]	6.42	2.91	2.03	1.65	1.31	1.41	0.98	0.83	0.65	0.71	0.59	0.48	0.51	0.40	0.41
[A,M]		5.98	2.71	2.03	1.83	2.77	1.28	0.98	0.88	0.88	0.68	0.60	0.57	0.51	0.45
[A,O]			2.53	1.66	0.95	1.28	1.18	0.79	0.48	0.85	0.56	0.35	0.47	0.29	0.29
[A,X]				3.36	0.18	0.98	0.79	1.53	0.13	0.56	1.05	0.13	0.85	0.11	0.18
[A,P]					10.51	0.88	0.48	0.13	4.71	0.35	0.13	0.29	0.11	2.40	1.59
[M,M]						5.97	2.61	1.94	1.77	1.19	0.90	0.82	0.69	0.62	0.61
[M,O]							2.35	1.57	0.85	1.09	0.73	0.43	0.56	0.34	0.33
[M,X]								3.21	0.05	0.73	1.45	0.91	1.09	0.09	0.11
[M,P]									11.33	0.43	0.09	4.61	0.09	3.33	3.12
[O,O]										1.06	0.69	0.43	0.47	0.27	0.22
[O,X]											1.37	0.15	0.89	0.09	0.13
[O,P]												3.81	0.09	2.66	1.17
[X,X]													1.87	0.10	0.19
[X,P]														5.76	0.20
[P,P]															25.19
[A,A]	63.76	9.73	3.26	3.26	3.34	2.57	1.00	1.08	1.04	0.59	0.42	0.40	0.81	0.51	0.62
[A,M]		26.24	6.31	10.00	6.48	6.22	1.55	1.52	1.66	0.83	0.94	0.76	0.93	0.89	0.66
[A,O]			16.60	1.17	4.69	1.56	3.49	0.33	0.88	1.71	0.18	0.48	0.16	0.37	0.24
[A,X]				42.82	2.46	1.51	0.32	7.76	0.55	0.18	3.81	0.31	4.10	0.29	0.20
[A,P]					15.46	1.67	0.88	0.55	3.76	0.48	0.31	1.70	0.28	1.95	1.64
[M,M]						21.84	5.45	8.80	5.80	1.54	2.39	1.50	6.35	2.44	1.68
[M,O]							11.72	0.89	3.56	3.21	0.31	1.01	0.37	1.92	1.01
[M,X]								35.45	1.88	0.30	9.01	0.51	19.08	1.07	0.53
[M,P]									12.03	1.00	0.53	3.12	1.09	4.58	3.32
[O,O]										6.91	0.61	2.18	0.19	0.16	0.75
[O,X]											20.31	1.16	1.07	0.17	0.34
[O,P]												6.52	0.17	0.57	1.94
[X,X]													69.25	3.26	0.34
[X,P]														19.23	1.18
[P,P]															7.16

Acknowledgements

The authors would like to thank Johannes Schmidt-Hieber, Till Sabel, Ruey Tsay, Dan Christina Wang and Dacheng Xiu for inspiring discussions. We also thank an anonymous referee for valuable comments on an earlier version. Markus Bibinger was supported by a fellowship of the German Academic Exchange Service (DAAD), the CRC 649 'Economic Risk' at Berlin, supported by the Deutsche Forschungsgemeinschaft (DFG), and from the Stevanovich Center for Financial Mathematics at the University of Chicago. Per Mykland was supported under the National Science Foundation grant SES 11-24526 and DMS 14-07812. We gratefully acknowledge these financial supports.
The data used in Section 5 has been provided by LOBSTER academic data powered by NASDAQ OMX.

Supporting information

Additional information for this article is available online including:

S1 Review on Realized Volatility under Endogenous Sampling.
S2 The Feasible Multivariate Limit Theorem.
S3 Supplementary information for the proofs.

References

Abadir, K. M. & Magnus, J. R. (2005). *Matrix algebra*, Econometric Exercises, vol. 1, Cambridge University Press, Cambridge.

Aït-Sahalia, Y., Fan, J. & Xiu, D. (2010). High-frequency estimates with noisy and asynchronous financial data. *J. Amer. Stat. Assoc.* **105**, (492), 1504–1516.

Aït-Sahalia, Y., Zhang, L. & Mykland, P. A. (2011). Ultra high frequency volatility estimation with dependent microstructure noise. *J. Econometrics,* **160**, 160–165.

Barndorff-Nielsen, O. E., Hansen, P. R., Lunde, A. & Shephard, N. (2008). Designing realised kernels to measure the ex-post variation of equity prices in the presence of noise. *Econometrica* **76**, (6), 1481–1536.

Barndorff-Nielsen, O. E., Hansen, P. R., Lunde, A. & Shephard, N. (2011). Multivariate realised kernels: Consistent positive semi-definite estimators of the covariation of equity prices with noise and non-synchronous trading. *J. Econometrics* **162**, (2), 149–169.

Barndorff-Nielsen, O. E. & Shephard, N. (2002). Econometric analysis of realized volatility and its use in estimating stochastic volatility models. *J. R. Stat. Soc.* **64**, (2), 253–280.

Bibinger, M. (2011). Efficient covariance estimation for asynchronous noisy high-frequency data. *Scand. J. Stat.* **38**, 23–45.

Bibinger, M. (2012). An estimator for the quadratic covariation of asynchronously observed Itô processes with noise: Asymptotic distribution theory. *Stoch. Process. Appl.* **122**, 2411–2453.

Bibinger, M., Hautsch, N., Malec, P. & Reiß, M. (2014). Estimating the quadratic covariation matrix from noisy observations: Local method of moments and efficiency. *Ann. Stat.* **42**, (4), 1312–1346.

Christensen, K., Kinnebrock, S. & Podolskij, M. (2010). Pre-averaging estimators of the ex-post covariance matrix in noisy diffusion models with non-synchronous data. *J. Econometrics,* **159**, (1), 116–133.

Christensen, K., Podolskij, M. & Vetter, M. (2013). On covariation estimation for multivariate continuous itô semimartingales with noise in non-synchronous observation schemes. *J. Multivariate Anal.* **120**, 59–84.

Eberlein, E., Frey, R. & Von, H. (2008). Advanced credit portfolio modeling and cdo pricing. In *Mathematics – key technology for the future* (eds W. Jäger & H.-J. Krebs), 253–280. Springer, Berlin.

Epps, T. W. (1979). Comovements in stock prices in the very short run. *J. Amer. Stat. Assoc.* **74**, 291–298.

Fukasawa, M. (2010). Realized volatility with stochastic sampling. *Stochastic. Process. Appl.* **120**, 209–233.

Gloter, A. & Jacod, J. (2001). Diffusions with measurement errors 1 and 2. *ESAIM, Probab. Stat.* **5**, 225–242.

Hayashi, T. & Yoshida, N. (2005). On covariance estimation of non-synchronously observed diffusion processes. *Bernoulli* **11**, (2), 359–379.

Hayashi, T. & Yoshida, N. (2011). Nonsynchronous covariation process and limit theorems. *Stochastic. Process. Appl.* **121**, 2416–2454.

Jacod, J. (1997). On continuous conditional gaussian martingales and stable convergence in law. *Séminaire de Probabilitiés, Strasbourg, Tome* **31**, 232–246.
Jacod, J. (2012). Statistics and high frequency data. In *Proceedings of the 7th Séminaire Européen de Statistique, La Manga, 2007: Statistical Methods for Stochastic Differential Equations* (eds M. Kessler, A. Lindner & M. Sørensen), 191–310. CRC Press, Boca Raton, FL.
Jacod, J., Li, Y., Mykland, P. A., Podolskij, M. & Vetter, M. (2009). Microstructure noise in the continous case: The pre-averaging approach. *Stochastic. Process. Appl.* **119**, 2803–2831.
Jacod, J. & Protter, P. (1998). Asymptotic error distributions for the euler method for stochastic differential equations. *Ann. Probab.* **26**, 267–307.
Koike, Y. (2014). Limit theorems for the pre-averaged Hayashi–Yoshida estimator with random sampling. *Stochastic. Process. Appl.* **124**, (8), 2699–2753.
Koike, Y. (2016). Time endogeneity and an optimal weight function in pre-averaging covariance estimation. *Stat. Inference Stoch. Process*. DOI: 10.1007/s11203-016-9135-3, to appear in print.
Li, Y., Mykland, P. A., Renault, E., Zhang, L. & Zheng, X. (2014). Realized volatility when sampling times are possibly endogenous. *Economet. Theory* **30**, 580–605.
Li, Y., Zhang, Z. & Zheng, X. (2013). Volatility inference in the presence of both endogenous time and microstructure noise. *Stochastic Process.' Appl.* **123**, (7), 2696–2727 (English).
Mykland, P. A., Shephard, N. & Sheppard, K. (2012). Efficient and feasible inference for the components of financial variation using blocked multipower variation, University of Oxford. Available at: http://papers.ssrn.com/sol3/papers.cfm?abstract_id=2008690 [Accessed on 1 December 2015].
Park, S., Hong, S. Y. & Linton, O. (2016). Estimating the quadratic covariation matrix for asynchronously observed high frequency stock returns corrupted by additive measurement error. *J. Econometrics* **191**, (2), 325–347.
Rényi, A. (1963). On stable sequences of events. *Sankhya, Ser. A* **25**, (3), 293–302.
Xiu, D. (2010). Quasi-maximum likelihood estimation of volatility with high frequency data. *J. Econometrics* **159**, 235–250.
Zhang, L. (2001). *From martingales to ANOVA: Implied and realized volatility. Ph.D. Thesis*, Department of Statistics, University of Chicago.
Zhang, L. (2006). Efficient estimation of stochastic volatility using noisy observations: A multi-scale approach. *Bernoulli* **12**, (6), 1019–1043.
Zhang, L. (2011). Estimating covariation: Epps effect and microstructure noise. *J. Econometrics* **160**, 33–47.
Zhang, L., Mykland, P. A. & Aït-Sahalia, Y. (2005). A tale of two time scales: Determining integrated volatility with noisy high-frequency data. *J. Amer. Stat. Assoc.* **100**, (472), 1394–1411.

Received November 2014, in final form January 2016

Markus Bibinger, Institute of Mathematics, Humboldt University of Berlin, Unter den Linden 6, 10099 Berlin, Germany.
Email: bibinger@math.hu-berlin.de

Appendix A: Preliminaries

The local boundedness condition in Assumption 1 can be strengthened to uniform boundedness on $[0, T]$ by a localization procedure carried out in Jacod (2012), Lemma 6. 6 of Section 6 .3. Let C be a generic constant and denote $\Delta_i W = W_{t_i} - W_{t_{i-1}}, i = 1, \ldots, n$, for the Brownian motion W driving the SDE with solution X in (1) and $\Delta_i \sigma = \sigma_{t_i} - \sigma_{t_{i-1}}$. Consider some norm $\| \cdot \|$, for example, the euclidean norm, on $\mathbb{R}^d$. Suppose Assumption 1 holds. By several applications of the Burkholder–Davis–Gundy and Hölder inequality, one can obtain the following estimates:

$$\mathbb{E}\left[\|\Delta_i X\|^2 + \|\Delta_i W\|^2 \middle| \mathcal{F}_{t_{i-1}}\right] \le C n^{-1}, \quad \mathbb{E}\left[\|\Delta_i \sigma\|^2 \middle| \mathcal{F}_{t_{i-1}}\right] \le C n^{-1}, \tag{36a}$$

$$\mathbb{E}\left[\|\Delta_i X - \sigma_{t_{i-1}} \Delta_i W\|^2 \middle| \mathcal{F}_{t_{i-1}}\right] \le C n^{-2}, \tag{36b}$$

for equidistant observation schemes $t_i = iT/n$. For general synchronous sampling, (36a) and (36b) remain valid when replacing n by δ_n^{-1} with $\delta_n = \sup_i (t_i - t_{i-1})$. The estimates (36a) and (36b) are proven in Jacod (2012), among others. They are used repeatedly in the analysis below. We write $a_n \asymp^p b_n$ if $a_n = \mathcal{O}_{\mathbb{P}}(b_n)$ and $b_n = \mathcal{O}_{\mathbb{P}}(a_n)$ and express analogously $a_n \asymp b_n$ for $a_n = \mathcal{O}(b_n)$ and $b_n = \mathcal{O}(a_n)$.

A summary including the elements of matrix algebra, which are heavily used throughout the proofs, can be found in Sections 10.1, 10.2 and 11.2 of Abadir & Magnus (2005). Let us calculate next the asymptotic variance–covariance matrix of the realized volatility matrix in (2), which serves as well as preparation for the proofs in the succeeding sections. Denote by $Z_i \in \mathbb{R}^d, i = 0, \ldots, n$, independent standard normally distributed random vectors. We apply the rule $\operatorname{vec}(ABC) = (C^\top \otimes A)\operatorname{vec}(B)$ for matrices A, B, C frequently in the following texts. The multivariate stable central limit theorems are proved based on Theorem 3–1 of Jacod (1997). The limiting variance–covariance matrix in (2) is random and obtained as the stochastic limit of the sum of conditional variance–covariance matrices. We find that

$$\begin{aligned}
&\sum_{i=1}^{n} \mathbb{C}\mathrm{ov}\left(\operatorname{vec}\left(\Delta_i X(\Delta_i X)^\top\right) \Big| \mathcal{F}_{\frac{(i-1)T}{n}}\right) \\
&\asymp^p \sum_{i=1}^{n} \mathbb{C}\mathrm{ov}\left(\operatorname{vec}\left(\sqrt{\frac{T}{n}}\Sigma^{1/2}_{\frac{(i-1)T}{n}}\left(Z_i(Z_i)^\top\right)\sqrt{\frac{T}{n}}\Sigma^{1/2}_{\frac{(i-1)T}{n}}\right) \Big| \mathcal{F}_{\frac{(i-1)T}{n}}\right) \\
&= \frac{T^2}{n^2}\sum_{i=1}^{n} \mathbb{C}\mathrm{ov}\left(\left(\Sigma^{1/2}_{\frac{(i-1)T}{n}} \otimes \Sigma^{1/2}_{\frac{(i-1)T}{n}}\right)\operatorname{vec}\left(Z_i Z_i^\top\right) \Big| \mathcal{F}_{\frac{(i-1)T}{n}}\right) \\
&= \frac{T^2}{n^2}\sum_{i=1}^{n}\left(\Sigma^{1/2}_{\frac{(i-1)T}{n}} \otimes \Sigma^{1/2}_{\frac{(i-1)T}{n}}\right)\mathcal{Z}\left(\Sigma^{1/2}_{\frac{(i-1)T}{n}} \otimes \Sigma^{1/2}_{\frac{(i-1)T}{n}}\right) \\
&= \frac{T}{n}\sum_{i=1}^{n}\left(\Sigma_{\frac{(i-1)T}{n}} \otimes \Sigma_{\frac{(i-1)T}{n}}\right)\frac{T}{n}\mathcal{Z} \asymp^p \frac{T}{n}\int_0^T (\Sigma_s \otimes \Sigma_s)\,\mathcal{Z}\,ds\,.
\end{aligned}$$

All other ingredients required to conclude (2) by Theorem 3–1 of Jacod (1997) are readily established here and we skip the details. We use analogous transforms for computing terms of the form

$$\mathbb{C}\mathrm{ov}\left(\operatorname{vec}\left(\Delta_j X \otimes (\Delta_l X)^\top\right)\right) = \mathbb{C}\mathrm{ov}\left(\operatorname{vec}\left(\Delta_j X(\Delta_l X)^\top\right)\right)$$

frequently in the succeeding texts without repeating each step.

Appendix B: Proofs of Section 2

Proof of Theorem 2.1. For the proof that $\widehat{[X,X]}_T^{(multi)} - \widehat{[X,X]}_T^{(kernel)} = -4\mathbf{H} + \mathcal{O}_{\mathbb{P}}\left(n^{-\frac{1}{4}}\right)$ if $\mathfrak{K}'' = h$ in (7), it suffices to focus on the first-order term of the weights. Transforming (6) yields

$$\begin{aligned}
\sum_{i=1}^{M_n}\frac{\alpha_i}{i}\sum_{j=i}^{n}\Delta_j^i Y(\Delta_j^i Y)^\top &= \sum_{i=1}^{M_n}\alpha_i\left(\sum_{j=2}^{n}\sum_{l=1}^{i\wedge(j-1)}\left(1-\frac{l}{i}\right)\left(\Delta_j Y(\Delta_{j-l}Y)^\top\right.\right. \\
&\qquad \left.\left. + \Delta_{j-l}Y(\Delta_j Y)^\top\right)\right) + \sum_{j=1}^{n}\Delta_j Y \Delta_j Y^\top - R_n \\
&= \sum_{l=1}^{M_n}\sum_{j=l+1}^{n}\sum_{i=l}^{M_n}\alpha_i\left(1-\frac{l}{i}\right)\left(\Delta_j Y(\Delta_{j-l}Y)^\top + \Delta_{j-l}Y(\Delta_j Y)^\top\right) \\
&\quad + \sum_{j=1}^{n}\Delta_j Y(\Delta_j Y)^\top - R_n\,.
\end{aligned}$$

The term R_n induced by end-effects

$$\sum_{i=1}^{M_n}\alpha_i\left(\sum_{j=1}^{i-1}\left(\frac{i-j}{i}\Delta_j Y(\Delta_j Y)^\top+\sum_{l=1}^{(j-1)\wedge 1}\frac{i-j}{i}\left(\Delta_j Y(\Delta_{j-l}Y)^\top+\Delta_{j-l}Y(\Delta_j Y)^\top\right)\right)\right.$$
$$\left.+\sum_{j=n-i+2}^{n}\left(\frac{i-n+j-1}{i}\left(\Delta_j Y(\Delta_j Y)^\top+\sum_{l=1}^{i\wedge(n-j)}\left(\Delta_j Y(\Delta_{j-l}Y)^\top+\Delta_{j-l}Y(\Delta_j Y)^\top\right)\right)\right)\right)$$

has an expectation by noise:

$$2\,\mathbf{H}\sum_{i=1}^{M_n}\alpha_i\left(\sum_{j=1}^{i-1}\frac{i-j}{i}-\sum_{j=2}^{i-1}\frac{i-j}{i}+\sum_{j=n-i+1}^{n-1}\frac{i-n+j}{i}-\sum_{j=n-i+1}^{n-2}\frac{i-n+j}{i}\right)=4\,\mathbf{H}.$$

The variance–covariance matrix of this term is asymptotically negligible, what can be shown with standard bounds. For the main term above, we can detach the inner sum and find that

$$\sum_{i=l}^{M_n}\alpha_i\left(1-\frac{l}{i}\right)=\sum_{i=l}^{M_n}\frac{i}{M_n^2}\frac{(i-l)}{i}\mathfrak{K}''\left(\frac{i}{M_n}\right)+\mathcal{O}\left(n^{-\frac{1}{4}}\right)$$
$$=\int_{l/M_n}^{1}\mathfrak{K}''(x)\left(x-\frac{l}{M_n}\right)dx+\mathcal{O}\left(n^{-\frac{1}{4}}\right)=\mathfrak{K}\left(\frac{l}{M_n}\right)+\mathcal{O}\left(n^{-\frac{1}{4}}\right),$$

by partial integration under the restrictions made on $\mathfrak{K}$. This yields the form (8) of the transformed kernel estimator and our claim. That the integral approximation does not harm the aforementioned equality up to the $\mathcal{O}\left(n^{-1/4}\right)$-term can be seen by the estimate

$$\int_{i/M_n}^{(i+1)/M_n}\left|f(x)-f\left(\frac{i}{M_n}\right)\right|dx\le\int_{i/M_n}^{(i+1)/M_n}C\left|x-\frac{i}{M_n}\right|dx\le C\,M_n^{-2}$$

with generic constant C, $i\ge l$, for the Lipschitz function $f(x)=\mathfrak{K}''(x)\left(x-\frac{l}{M_n}\right)$ on the compact support $[0,1]$, where Lipschitz continuity is ensured by the preconditioned continuous differentiability. □

Proof of Theorem 2.2. Decompose the multi-scale estimator (6)

$$\sum_{i=1}^{M_n}\frac{\alpha_i}{i}\sum_{j=i}^{n}\Delta_j^i Y(\Delta_j^i Y)^\top=\sum_{i=1}^{M_n}\frac{\alpha_i}{i}\sum_{j=i}^{n}\Delta_j^i X(\Delta_j^i X)^\top+\sum_{i=1}^{M_n}\frac{\alpha_i}{i}\sum_{j=i}^{n}\Delta_j^i\epsilon(\Delta_j^i\epsilon)^\top$$
$$+\sum_{i=1}^{M_n}\frac{\alpha_i}{i}\sum_{j=i}^{n}\Delta_j^i X(\Delta_j^i\epsilon)^\top+\sum_{i=1}^{M_n}\frac{\alpha_i}{i}\sum_{j=i}^{n}\Delta_j^i\epsilon(\Delta_j^i X)^\top,$$

with $\Delta_i^j\epsilon=\epsilon_j-\epsilon_{j-i}$, in a signal part, a noise part and cross terms, which are uncorrelated. We analyse the variance–covariance matrices consecutively. Write the signal term

$$\frac{1}{i}\sum_{j=i}^{n}\Delta_j^i X(\Delta_j^i X)^\top = \sum_{j=1}^{n}\Delta_j X(\Delta_j X)^\top + \sum_{l=1}^{n}\Delta_l X\sum_{j=1}^{i\wedge l}\left(1-\frac{j}{i}\right)(\Delta_{l-j}X)^\top + \sum_{l=1}^{n}\sum_{j=1}^{i\wedge l}\left(1-\frac{j}{i}\right)\Delta_{l-j}X(\Delta_l X)^\top . \tag{37}$$

The first addend is the realized volatility matrix, converging with rate $n^{1/2}$ to the integrated volatility matrix, and thus contributing only an asymptotically negligible error term. Because of

$$\text{vec}\left(\Delta_l X(\Delta_{l-j}X)^\top + \Delta_{l-j}X(\Delta_l X)^\top\right) = \mathcal{Z}\text{vec}\left(\Delta_l X(\Delta_{l-j}X)^\top\right),$$

it is enough to consider one addend. Using

$$\begin{aligned}\mathbb{C}\text{ov}\left(\text{vec}\left(\Delta_l X(\Delta_{l-j}X)^\top\right)\middle|\mathcal{F}_{\frac{(l-1)T}{n}}\right) &= \mathbb{C}\text{ov}\left(\Delta_{l-j}X\otimes\Delta_l X\middle|\mathcal{F}_{\frac{(l-1)T}{n}}\right)\\ &= \mathbb{E}\Big[(\Delta_{l-j}X\otimes\Delta_l X)\left((\Delta_{l-j}X)^\top\otimes(\Delta_l X)^\top\right)\Big|\\ &\quad\times\mathcal{F}_{\frac{(l-1)T}{n}}\Big] \asymp^p \Delta_{l-j}X(\Delta_{l-j}X)^\top\otimes\Sigma_{\frac{(l-1)T}{n}}\frac{T}{n},\end{aligned}$$

$$\mathbb{E}\left[\Delta_{l-j}X(\Delta_{l-j}X)^\top\otimes\Sigma_{\frac{(l-1)T}{n}}\frac{T}{n}\right]\asymp^p T^2n^{-2}\left(\Sigma_{\frac{(l-j-1)T}{n}}\otimes\Sigma_{\frac{(l-1)T}{n}}\right),$$

and $\mathcal{Z}^2 = 2\mathcal{Z}$, we derive that

$$\begin{aligned}&\sum_{l=1}^{n}\mathbb{C}\text{ov}\left(\sum_{j=1}^{i\wedge l}\left(1-\frac{j}{i}\right)\mathcal{Z}\text{vec}\left(\Delta_l X\left(\Delta_{l-j}X\right)^\top\right)\middle|\mathcal{F}_{\frac{(l-1)T}{n}}\right)\\ &\quad\asymp^p\sum_{l=1}^{n}T^2n^{-2}\left(\Sigma_{\frac{(l-1)T}{n}}\otimes\Sigma_{\frac{(l-1)T}{n}}\right)2\mathcal{Z}\sum_{j=1}^{i\wedge l}\left(1-\frac{j}{i}\right)^2.\end{aligned}$$

The bounds in (36a), (36b) suffice that the approximation errors above are asymptotically negligible. Verifying all other conditions of Theorem 3–1 of Jacod (1997), which readily follow along the same lines as in the proof of Proposition A.3 of Bibinger (2012), we obtain first stable central limit theorems for discretization errors of subsampling estimators with fixed subsampling frequencies. The covariances between them are determined with

$$\begin{aligned}&\sum_{l=1}^{n}\mathbb{E}\left[\sum_{j=1}^{i\wedge l}\left(1-\frac{j}{i}\right)\mathcal{Z}\text{vec}\left(\Delta_l X\left(\Delta_{l-j}X\right)^\top\right)\left(\sum_{j=1}^{i'\wedge l}\left(1-\frac{j}{i'}\right)\mathcal{Z}\text{vec}\left(\Delta_l X\left(\Delta_{l-j}X\right)^\top\right)\right)^\top\right|\\ &\quad\times\mathcal{F}_{\frac{(l-1)T}{n}}\Bigg]\asymp^p\sum_{l=1}^{n}T^2n^{-2}\left(\Sigma_{\frac{(l-1)T}{n}}\otimes\Sigma_{\frac{(l-1)T}{n}}\right)2\mathcal{Z}\sum_{j=1}^{\min(i,i',l)}\left(1-\frac{j}{i}\right)\left(1-\frac{j}{i'}\right).\end{aligned}$$

Hence, we are left to evaluate the deterministic sum:

$$\sum_{j=1}^{m}\left(1-\frac{j}{i}\right)\left(1-\frac{j}{i'}\right) = \frac{m}{2}-\frac{m^2}{6M}-\frac{1}{8}+\frac{1}{12M}\asymp\frac{m}{6}\left(3-\frac{m}{M}\right),$$

where $m = \min(i, i')$ and $M = \max(i, i')$. Including the weights according to (7), we set

$$\mathfrak{D}^{\alpha} = \lim_{n\to\infty} M_n^{-1} \sum_{k=1}^{M_n} \sum_{l=1}^{k} \frac{l}{6M_n}\left(3 - \frac{l}{k}\right)\alpha_k \alpha_l \,.$$

With the covariances for different subsample frequencies above we obtain a stable central limit theorem for vectors spanning over finite sets of different frequencies. The Cramér-Wold device implies central limit theorems for linear combinations of the components. The final stable limit theorem

$$n^{\frac{1}{4}} \text{vec}\left(\sum_{i=1}^{n} \frac{\alpha_i}{i} \sum_{j=i}^{n} \Delta_j^i X \left(\Delta_j^i X\right)^{\top} - \int_0^T \Sigma_s \, ds\right) \xrightarrow{st} MN\left(0, 4\mathfrak{D}^{\alpha} cT \int_0^T (\Sigma_s \otimes \Sigma_s)\, \mathcal{Z}\, ds\right) \tag{38}$$

is concluded by extending this to infinitely many subsample frequencies adopting the analogous step from Zhang (2006) for the univariate multi-scale estimator. Thereby we conclude the signal term of (14). $\mathfrak{D}^{\alpha}$ is a constant showing up in the asymptotic discretization variance depending on the weights, where for the standard weights (12) or cubic kernel $\mathfrak{D}^{\alpha} = 13/70$.
Next, consider the noise term

$$\sum_{i=1}^{M_n} \frac{\alpha_i}{i} \sum_{j=i}^{n} \Delta_j^i \epsilon (\Delta_j^i \epsilon)^{\top} = \sum_{i=1}^{M_n} \frac{\alpha_i}{i}\left(2\sum_{j=1}^{n} \epsilon_j \epsilon_j^{\top} - \sum_{j=i}^{n}\left(\epsilon_j \epsilon_{j-i}^{\top} + \epsilon_{j-i}\epsilon_j^{\top}\right) - \sum_{j=n-i+1}^{n} \epsilon_j \epsilon_j^{\top} - \sum_{j=0}^{i-1} \epsilon_j \epsilon_j^{\top}\right).$$

The last two sums lead for the non-adjusted multi-scale estimator (6) to the negative bias by noise and end-effects. The first inner sum on the right-hand side earlier does not depend on i, and the term vanishes because $\sum_{i=1}^{M_n} \alpha_i / i = 0$. The variance–covariance matrices of the remaining uncorrelated addends contribute to the total variance–covariance matrix because of noise perturbation. As the noise variance–covariance matrix $\mathbf{H}$ is fixed, we may work conditional on X and consider covariances directly instead of conditional covariances as for the discretization part. Denote the constant limits

$$\mathfrak{N}_2^{\alpha} = \lim_{n\to\infty} M_n \sum_{j=1}^{M_n - 1}\left(\sum_{i=j+1}^{M_n} \frac{\alpha_i}{i}\right)^2 \text{ and } \mathfrak{N}_1^{\alpha} = \lim_{n\to\infty} M_n^3 \sum_{i=1}^{M_n} \frac{\alpha_i^2}{i^2}\,.$$

Rewriting $$\sum_{i=1}^{M_n} \frac{\alpha_i}{i} \sum_{j=i}^{n} \epsilon_j \epsilon_{j-i}^{\top} = \sum_{j=1}^{n} \sum_{i=1}^{M_n \wedge j} \frac{\alpha_i}{i} \epsilon_j \epsilon_{j-i}^{\top}\,,$$

$$\sum_{i=1}^{M_n} \frac{\alpha_i}{i}\left(\sum_{j=0}^{i-1} \epsilon_j \epsilon_j^{\top} + \sum_{j=n-i+1}^{n} \epsilon_j \epsilon_j^{\top}\right) = \sum_{j=0}^{M_n - 1}\left(\epsilon_j \epsilon_j^{\top} + \epsilon_{n-j}\epsilon_{n-j}^{\top}\right) \sum_{i=j+1}^{M_n} \frac{\alpha_i}{i}\,,$$

we obtain that $$\frac{M_n^3}{n} \mathbb{C}\text{ov}\left(\text{vec}\left(\sum_{i=1}^{M_n} \frac{\alpha_i}{i} \sum_{j=i}^{n} \epsilon_j \epsilon_{j-i}^{\top}\right)\right) \to \mathfrak{N}_1^{\alpha}\left(\mathbf{H} \otimes \mathbf{H}\right)\,, \tag{39}$$

$$M_n \mathbb{C}\mathrm{ov}\left(\mathrm{vec}\left(\sum_{i=1}^{M_n}\frac{\alpha_i}{i}\left(\sum_{j=n-i+1}^{n}\epsilon_j\epsilon_j^\top+\sum_{j=0}^{i-1}\epsilon_j\epsilon_j^\top\right)\right)\right)\to 2\mathfrak{N}_2^\alpha\,\mathbf{H}^{\otimes}\mathcal{Z}\,. \tag{40}$$

Using once again that $\mathrm{vec}(A+A^\top)=\mathcal{Z}\mathrm{vec}(A)$ for $A\in\mathbb{R}^{d\times d}$, we conclude the noise parts in (14). For the specific weights (12) corresponding to the cubic kernel, we have $\mathfrak{N}_2^\alpha=6/5$ and $\mathfrak{N}_1^\alpha=12$, which gives the minimum of the variance due to noise, cf. Zhang (2006).

Finally, consider the cross terms. They can be decomposed in addends of the form

$$\sum_{i=1}^{M_n}\frac{\alpha_i}{i}\sum_{j=0}^{n}\left(\zeta_{i,j}\epsilon_j^\top+\epsilon_j\zeta_{i,j}^\top\right)\ \text{, where }\ \zeta_{i,j}=\begin{cases}-\Delta_{i-j}^iX & ,0\le j\le(i-1)\\ \Delta_j^iX-\Delta_{j+i}^iX & ,i\le j\le(n-i)\\ \Delta_j^iX & ,n-i+1\le j\le n\end{cases}.$$

In order to derive the asymptotic variance–covariance matrix, observe that

$$\mathbb{C}\mathrm{ov}\left(\mathrm{vec}\left(\zeta_{i,j}\epsilon_j^\top\right)\right)=\mathbb{C}\mathrm{ov}\left(\epsilon_j\otimes\zeta_{i,j}\right)=\mathbb{E}\left[\left(\epsilon_j\otimes\zeta_{i,j}\right)\left(\epsilon_j^\top\otimes\zeta_{i,j}^\top\right)\right]=\mathbf{H}\otimes\mathbb{E}\left[\zeta_{i,j}\zeta_{i,j}^\top\right].$$

Now, if we assume without loss of generality $1\le i\le i'\le M_n$, it holds for $M_n\le j\le(n-M_n)$ that

$$\begin{aligned}\zeta_{i,j}\zeta_{i',j}^\top&=\left(\sum_{r=j-i}^{j-1}\Delta_rX-\sum_{r=j}^{i+j-1}\Delta_rX\right)\left(\sum_{r=j-i'}^{j-1}(\Delta_rX)^\top-\sum_{r=j}^{i'+j-1}(\Delta_rX)^\top\right)\\&=\zeta_{i,j}\zeta_{i,j}^\top+\zeta_{i,j}\left(\sum_{r=j-i'}^{j-i-1}(\Delta_rX)^\top-\sum_{r=j+i}^{i'+j-1}(\Delta_rX)^\top\right).\end{aligned}$$

We obtain for the sum of conditional variance–covariance matrices the following convergence:

$$\begin{aligned}&M_n\sum_{j=0}^{n}\mathbb{C}\mathrm{ov}\left(\mathrm{vec}\left(\sum_{i=1}^{M_n}\frac{\alpha_i}{i}\left(\zeta_{i,j}\epsilon_j^\top+\epsilon_j\zeta_{i,j}^\top\right)\right)\Big|\mathcal{F}_{\frac{(j-1)T}{n}}\right)\\&=M_n\sum_{j=0}^{n}\mathbb{C}\mathrm{ov}\left(\sum_{i=1}^{M_n}\frac{\alpha_i}{i}\mathcal{Z}\mathrm{vec}\left(\zeta_{i,j}\epsilon_j^\top\right)\Big|\mathcal{F}_{\frac{(j-1)T}{n}}\right)\\&\asymp^p 2M_n\mathcal{Z}\left(\mathbf{H}\otimes\sum_{i=1}^{M_n}\sum_{r=1}^{M_n}\frac{\alpha_i\alpha_r}{ir}(i\wedge r)\left(\frac{1}{(i\wedge r)}\sum_{j=(i\wedge r)}^{n}\mathbb{E}\left[\Delta_j^{(i\wedge r)}X(\Delta_j^{(i\wedge r)}X)^\top\Big|\mathcal{F}_{\frac{(j-1)T}{n}}\right]\right)\right)\mathcal{Z}\\&\stackrel{p}{\to}2\mathfrak{M}^\alpha\mathcal{Z}\left(\mathbf{H}\otimes[X,X]_T\right)\mathcal{Z}\,,\end{aligned} \tag{41}$$

with the constant of the limit depending on the weights (7): $\mathfrak{M}^\alpha=\lim_{M_n\to\infty}\sum_{i,k=1}^{M_n}\frac{\alpha_i\alpha_k}{ik}(i\wedge k)$.

For the specific weights (12), the constant takes the value $\mathfrak{M}^\alpha=6/5$. For the asymptotic variance–covariance matrix of the cross terms, we can use that

$$\mathcal{Z}(A\otimes B)\mathcal{Z}=(A\otimes B+B\otimes A)\,\mathcal{Z}=\mathcal{Z}(B\otimes A)\mathcal{Z}\,. \tag{42}$$

We restrict ourselves to the evaluation of the general multivariate variance–covariance structure and derive (14) by (38), (39), (40) and (41). The remaining elements of the proof of the stable central limit theorem are close to Zhang (2006) and Bibinger (2012) and omitted. □

Appendix C: Proof of Theorem 3.1 and Corollary 3.5

We only highlight the impact of irregular and endogenous sampling on the discretization term, a more detailed complete proof is provided in the Supporting Information. A decomposition analogous to (37) gives

$$\frac{N}{M_N}\sum_{j=1}^{N}\mathbb{C}\text{ov}\left(\text{vec}\left(\sum_{i=1}^{M_N\wedge j}\frac{\alpha_i}{i}\left(X_{T_j}^{+}-X_{T_{j-i}}^{-}\right)\left(X_{T_j}^{+}-X_{T_{j-i}}^{-}\right)^{\top}\right)\middle|\mathcal{F}_{T_{j-1}}\right)$$
$$\asymp^p \frac{N}{M_N}\sum_{l=1}^{N}\mathbb{C}\text{ov}\left(\sum_{i=1}^{M_N\wedge l}\alpha_i\,\mathcal{Z}\text{vec}\left((X_{T_l}-X_{T_{l-1}})\sum_{j=1}^{i\wedge l}\left(1-\frac{j}{i}\right)(X_{T_{l-j}}-X_{T_{l-j-1}})^{\top}\right)\middle| \times\mathcal{F}_{T_{l-1}}\right)$$

where conditioning on $\left(\mathcal{F}_{T_{l-1}}\right)_l$ is each time replaced by $\left(\mathcal{F}^N_{T_{l-1}}\right)_l$ defined within Assumption 3.4 for the endogenous case. On the regularity conditions of Assumption 3.4, we derive that

$$\mathbb{C}\text{ov}\left(\text{vec}\left((X_{T_l}-X_{T_{l-1}})(X_{T_{l-j}}-X_{T_{l-j-1}})^{\top}\right)\right)\asymp\left[\int_{T_{l-j-1}}^{T_{l-j}}\Sigma_s\,ds\right]\otimes\left[\int_{T_{l-1}}^{T_l}\Sigma_s\,ds\right], \tag{43}$$

$$\mathbb{E}\left[\int_{T_{l-j-1}}^{T_{l-j}}\Sigma_s\,ds\otimes\int_{T_{l-1}}^{T_l}\Sigma_s\,ds\middle|\mathcal{F}_{T_{l-1}}\right]\asymp^p\left(\Sigma_{T_{l-j-1}}\otimes\Sigma_{T_{l-1}}\right)(T_l-T_{l-1})(T_{l-j}-T_{l-j-1}),$$

by generalized Itô-isometry. The following computation is analogous for exogenous and endogenous observation times:

$$\frac{N}{M_N}\sum_{j=1}^{N}\mathbb{C}\text{ov}\left(\text{vec}\left(\sum_{i=1}^{M_N\wedge j}\frac{\alpha_i}{i}\left(X_{T_j}^{+}-X_{T_{j-i}}^{-}\right)\left(X_{T_j}^{+}-X_{T_{j-i}}^{-}\right)^{\top}\right)\middle|\mathcal{F}_{T_{j-1}}\right)$$
$$\asymp^p\frac{N}{M_N}\sum_{l=1}^{N}\mathbb{C}\text{ov}\left(\mathcal{Z}\sum_{i=1}^{M_N\wedge l}\frac{\alpha_i}{i}\text{vec}\left((X_{T_l}-X_{T_{l-1}})\sum_{q=1}^{i\wedge l}\left(1-\frac{q}{i}\right)(X_{T_{l-q+1}}-X_{T_{l-q}})^{\top}\right)\middle|\mathcal{F}_{T_{l-1}}\right)$$
$$\asymp^p\frac{2N}{M_N}\sum_{l=1}^{N}\sum_{i,k=1}^{M_N\wedge l}\alpha_i\alpha_k(T_l-T_{l-1})\mathcal{Z}\sum_{q=1}^{\min(l,i,k)}\left(\left(1-\frac{q}{i}\right)\left(1-\frac{q}{k}\right)(X_{T_{l-q+1}}-X_{T_{l-q}})(X_{T_{l-q+1}}-X_{T_{l-q}})^{\top}\otimes\Sigma_{T_{l-1}}\right)\mathcal{Z}$$
$$\asymp^p\frac{2N}{M_N}\sum_{i,k=1}^{M_N}\alpha_i\alpha_k\sum_{l=1}^{N}(T_l-T_{l-1})\mathcal{Z}\left(\Sigma_{T_{l-1}}\otimes\Sigma_{T_{l-1}}\right)\mathcal{Z}\sum_{q=1}^{\min(l,i,k)}\left(1-\frac{q}{i}\right)\left(1-\frac{q}{k}\right).$$

such that Assumption 3.2 (resp. Assumption 3.4) ensures the convergence in probability to $4\int_0^T(D^\alpha)'(s)\,(\Sigma_s\otimes\Sigma_s)\,\mathcal{Z}\,ds$. □