

Inference for Multi-Dimensional High-Frequency Data: Equivalence of Methods, Central Limit Theorems, and an Application to Conditional Independence Testing

Markus Bibinger*
Per A. Mykland**



* Humboldt-Universität zu Berlin, Germany

** University of Chicago, U.S.A.

This research was supported by the Deutsche Forschungsgemeinschaft through the SFB 649 "Economic Risk".

<http://sfb649.wiwi.hu-berlin.de>
ISSN 1860-5664

SFB 649, Humboldt-Universität zu Berlin
Spandauer Straße 1, D-10178 Berlin



Inference for Multi-Dimensional High-Frequency Data: Equivalence of Methods, Central Limit Theorems, and an Application to Conditional Independence Testing

Markus Bibinger & Per A. Mykland

Humboldt-Universität zu 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. The central limit theorem is shown to have a feasible version. 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. We also include the analysis for the Hayashi-Yoshida estimator in absence of microstructure.

The theory leads to multi-dimensional stable central limit theorems for respective estimators and hence allows to draw statistical inference for a broad class of multivariate models and linear functions of the recorded components. This 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 cover more complex functions and in order 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

JEL Classification: C14, C32, C58, G10

1 Introduction

The estimation of daily integrated variance and covariance¹ has become a central building block in model calibration for financial risk analysis. Recent years have seen a tremendous increase in trading activities along with ongoing buildup of computer-based trading. The availability of recorded asset prices at such high frequencies magnifies the appeal of asset price models grounded on continuous-time stochastic processes which are a cornerstone of financial modeling since the seminal works by Black & Scholes (1973) and Heston (1993). Increasing observation frequencies makes it possible to consider efficient estimation from underlying statistical experiments. Rising demand for an advanced theoretical foundation thus gave birth to the field of statistics for high-frequency data, going back to the path-breaking work of Andersen & Bollerslev (1998), Andersen *et al.* (2001, 2003), and Barndorff-Nielsen & Shephard (2001, 2002).

This article contributes to this strand of literature by considering a continuous-time stochastic process, i. e. a continuous semimartingale X comprising current stochastic volatility models, observed on a fixed time span $[0, T]$ at $(n + 1)$ points of a discrete grid and by investigating asymptotics when the mesh size of the grid tends to zero. A natural estimator for the integrated variance of a process is the discrete version called realized variance or realized volatility. For a continuous semimartingale, this estimator is consistent, and it weakly converges with usual \sqrt{n} -rate to a mixed normal distribution where twice the integrated quarticity occurs as random asymptotic variance (Barndorff-Nielsen & Shephard (2002), Jacod & Protter (1998), Zhang (2001)). 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

¹More accurately known as integrated volatilities and covolatilities, but we here stick to the more heavily used terminology.

convergence is equivalent to joint weak convergence with every measurable bounded random variable² 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, and it completes the asymptotic distribution theory for a univariate setup.³ Yet, an apparent problem pertinent to applications is to quantify the risk of a collection of high-frequently observed assets. Suppose we wish to estimate the quadratic variation of a sum $X_1 + X_2$, both processes X_1 and X_2 documented as high-frequency data and modeled by continuous semimartingales. The quadratic variation of the sum is the sum of integrated variances and twice the integrated covariance. For the latter, the derivation of a feasible central limit theorem is evident as for its one-dimensional counterpart. However, when estimating $[X_1 + X_2]$ with the sum of these estimates, we do not obtain a feasible asymptotic distribution theory for this combined estimator for free. This is due to the fact, that the single estimates are correlated. To deduce the asymptotic variance of the compound estimator, we are in need of a multivariate limit theorem involving the asymptotic covariance matrix of the estimators. The first part of this article is devoted to that task and provides asymptotic covariances of covariance matrix estimators within prominent specific models for high-frequency data.

The aspiration to progress to more complex statistical models in this research area, has again 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 covariances and a feasible central limit theorem has been attained in Hayashi & Yoshida (2011).

For one-dimensional high-frequency data, increasing sample sizes are expected to render the estimation error by discretization smaller and smaller – which is clearly the case if we assume an underlain continuous semimartingale. Contrary to the feature of the statistical model, in many situations high-frequency financial data exhibit an exploding realized variance when the sampling frequency is too high.⁴ This effect is ascribed to market microstructure frictions as bid-ask spreads and trading costs. A favored 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 variance estimators have been designed which 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 variance 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). Though the estimators have been found in independent works and rely on various principles, it turned out that they are actually quite similar and in a certain asymptotic sense equivalent which is clarified in Section 3 below.

The approaches to cope with microstructure noise analogously carry over to the synchronous multi-dimensional setting. Recently, methods to deal with noise and non-synchronicity in one go have been established in the literature. In fact, to each of the abovementioned smoothing techniques one extension to non-synchronous observation schemes has been proposed. First, the multivariate realised kernels by Barndorff-Nielsen *et al.* (2011) using refresh time sampling are eligible to estimate integrated covariance 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. Eventually, a feasible asymptotic distribution theory for the general non-synchronous and noisy setup has been provided by Bibinger (2012) and Christensen *et al.* (2011) 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, since for the approach in Bibinger (2012) interpolation takes place on the high-frequency scale after smoothing is adjusted with

²For a discussion of the general case, see p. 270 of Jacod & Protter (1998).

³See Section 2 for definition and further discussion.

⁴This is usually seen with the help of a so-called signature plot, see Andersen *et al.* (2000) and also the discussion in Chapter 2.5.2 of Mykland & Zhang (2012).

respect to a synchronous approximation whereas Christensen *et al.* (2011) 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. Park & Linton (2012) use Fourier methods on the same problem.

Remarkably, when interpolations in the fashion of the Hayashi-Yoshida estimator are performed on the high-frequency scale their impact on the asymptotic discretization variance of the hybrid generalized multi-scale estimator vanishes asymptotically at the slower convergence rate in the presence of noise.⁵ This reveals that non-synchronicity becomes less important in the latent observation model with noise dilution. In all four models, discretely observed continuous semimartingales with or without noise, synchronously or non-synchronously, we develop the asymptotic covariance structure of the respective estimation methods. We choose the realized covariance matrix, the multi-scale, the Hayashi-Yoshida and the generalized multi-scale estimator to establish the multivariate limit theorems. While the asymptotic covariances for the synchronous settings are found following a similar strategy as for the asymptotic variances, the most intricate challenge arises for non-synchronous sampling schemes. The asymptotic variance of the Hayashi-Yoshida estimator can be illuminated as in Bibinger (2011a) by a synchronous approximation and interpolations and hinges on an interplay of the two different sampling grids. For covariances, we consider two Hayashi-Yoshida estimates with generally four different sampling schemes. Nevertheless, utilizing an illustration with refresh times of pairs and quadruplets will reveal the nature of the asymptotic covariance. For the generalized multi-scale estimator in the most general setup we benefit again from the fact that interpolation effects fall out asymptotically. Yet, the effects by superposition with noise require a tedious notation to cover all possible sampling designs. In typical situations, as a completely asynchronous sampling design, only the signal parts will contribute to asymptotic covariances, and the multivariate distribution simplifies leading to a tractable general approach.

Relying on the asymptotic distribution of the considered integrated covariance matrix estimators, we strive to design a statistical test for investigating hypotheses, if two processes have zero covariation conditioned on a third one. We end up with a feasible stable central limit theorem for the test statistic involving products of estimators and thus 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. We start in Section 2 by discussing asymptotic covariance matrices for the realized covariances in a simple equidistant discretely observed Itô process setup. We proceed to statistical experiments with noise in Section 3, non-synchronous sampling in Section 4 and both at the same time in Section 5. In Section 6, the results are gathered to conclude feasible multivariate central limit theorems. The conditional independence test is introduced in Section 7 and after an empirical study of asymptotic covariances applied in Section 8 to high-frequency financial data. The proofs can be found in Appendix A.

2 The simple case: Asymptotic covariance matrix of realized covariances

Assumption 1. Consider a continuous p -dimensional Itô semimartingale (Itô process)

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

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 p -dimensional Brownian motion W and adapted $p \times p'$ càdlàg volatility process σ . Suppose that σ itself is an Itô process again, given by an equation similar to (1). The processes σ and W can be dependent, allowing for leverage effect.

⁵ Similar findings were made for the two-scales estimator in Zhang (2011), and for local likelihood in Bibinger *et al.* (2012).

Since we have in mind financial applications in which X is assumed to represent an hypothetical underlying log-price process, we will call X efficient (log-price) process in this context. For the ease of exposition we restrict ourselves to $p = 4$, which suffices to reveal the general asymptotic covariance form. The target of inference is the integrated covariance matrix $\int_0^T \Sigma_s ds$, over a certain fixed time span $[0, T]$, where $\Sigma = \sigma\sigma^\top$, with σ from (1). We denote in the following

$$\Sigma_s = \begin{pmatrix} \sigma_s^{(11)} & \sigma_s^{(12)} & \sigma_s^{(13)} & \sigma_s^{(14)} \\ \sigma_s^{(12)} & \sigma_s^{(22)} & \sigma_s^{(23)} & \sigma_s^{(24)} \\ \sigma_s^{(13)} & \sigma_s^{(23)} & \sigma_s^{(33)} & \sigma_s^{(34)} \\ \sigma_s^{(14)} & \sigma_s^{(24)} & \sigma_s^{(34)} & \sigma_s^{(44)} \end{pmatrix} \quad (2)$$

in the four-dimensional setting.

Assumption 2. *The Itô process X from Assumption 1 with $p = 4$ is discretely observed at equidistant observation times $iT/n, i \in \{0, \dots, n\}$.⁶*

We write

$$\Delta_i X = X_{\frac{iT}{n}} - X_{\frac{(i-1)T}{n}}, 1 \leq i \leq n, \text{ and } \Delta_i^j X = X_{\frac{iT}{n}} - X_{\frac{(i-j)T}{n}}, 1 \leq i \leq n, 2 \leq j \leq i, \quad (3)$$

for the increments and for increments to longer lags, respectively, and analogously for the single components.

The standard estimators for the integrated covariance matrix based on discrete observations considered in this section, but also the estimators designed to deal with noise and non-synchronicity below, will have variances hinging on the random volatility process σ . This is the main motivation why for high-frequency asymptotics in the strand of literature on integrated variance estimation, stable weak convergence is inherent as an essential concept.⁷ Stable central limit theorems allow for feasible limit theorems if the asymptotic variance can be estimated consistently and thus for statistical inference and confidence bands.⁸ The stability of weak convergence with respect to \mathcal{F} is established for all estimators considered throughout this article. In the sequel, we use the notation $\mathbb{V}\text{ar}_\Sigma(\cdot), \text{Cov}_\Sigma(\cdot)$ for random (co-)variances dependent on Σ . The stochastic limits multiplied with the convergence rate will be denoted **AVAR** and **ACOV** for asymptotic variance and covariance, respectively.

Proposition 2.1. *On Assumptions 1 and 2, for $p = 4$, the asymptotic covariance between realized covariances yields:*

$$\mathbb{A}\text{COV} \left(\sum_{i=1}^n \Delta_i X^{(k)} \Delta_i X^{(l)}, \sum_{i=1}^n \Delta_i X^{(r)} \Delta_i X^{(q)} \right) = T \int_0^T (\sigma_s^{(kr)} \sigma_s^{(lq)} + \sigma_s^{(kq)} \sigma_s^{(lr)}) ds, \quad (4)$$

for all $k, l, r, q \in \{1, 2, 3, 4\}$. In particular, we have

$$\mathbb{A}\text{COV} \left(\sum_{i=1}^n \Delta_i X^{(1)} \Delta_i X^{(2)}, \sum_{i=1}^n \Delta_i X^{(3)} \Delta_i X^{(4)} \right) = T \int_0^T (\sigma_s^{(13)} \sigma_s^{(24)} + \sigma_s^{(14)} \sigma_s^{(23)}) ds.$$

⁶The discussion in this section extends to data that are synchronous but mildly irregular, cf. Mykland & Zhang (2012), Chapter 2.7.1.

⁷Let 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 $n \rightarrow \infty$ 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 , $E I_A g(Z_n) \rightarrow E I_A g(Z)$ as $n \rightarrow \infty$. I_A denotes the indicator function of A , and $= 1$ if A and $= 0$ otherwise. In the case of no microstructure, $\mathcal{X} = \mathcal{F}_T$. If there is microstructure, \mathcal{X} is formed as the smallest sigma-field containing \mathcal{F}_T and also making the microstructure measurable. We refer to Jacod (1997), Jacod & Protter (1998), and Bibinger (2011a) for background information on stable convergence for this estimation problem.

⁸Stable convergence also permits the suppression of drift through measure change, see Section 2.2 of Mykland & Zhang (2009), which draws on Rootzén (1980). The device is similar to the passage to risk neutral measures in finance, going back to Ross (1976), Harrison & Kreps (1979), and Harrison & Pliska (1981). This mode of convergence also permits the localization of processes such as volatility, so they can be assumed bounded, see Chapter 2.4.5 of Mykland & Zhang (2012).

A generalization for non-equidistant sampling is covered by Proposition 4.1 in Section 4. From now on we express the general asymptotic covariances using indices 1, 2, 3, 4 as in the second formula above, and obtain special cases by inserting ‘(1) = (2)’ etc. Proposition 2.1 includes the well-known results that

$$n \mathbb{V}\text{ar}_\Sigma \left(\sum_{i=1}^n (\Delta_i X)^2 \right) \xrightarrow{p} 2T \int_0^T \sigma_s^4 ds$$

for the realized variance in a one-dimensional setup, where the asymptotic variance hinges on the so-called integrated quarticity, and

$$n \mathbb{V}\text{ar}_\Sigma \left(\sum_{i=1}^n \Delta_i X^{(1)} \Delta_i X^{(2)} \right) \xrightarrow{p} T \int_0^T (1 + \rho_s^2) (\sigma_s^{(1)} \sigma_s^{(2)})^2 ds$$

in a bivariate model with spot correlation process ρ_s . Already in the two-dimensional model we additionally obtain asymptotic covariances between realized variances and the realized covariance

$$n \mathbb{C}\text{ov}_\Sigma \left(\sum_{i=1}^n \Delta_i X^{(1)} \Delta_i X^{(2)}, \sum_{i=1}^n (\Delta_i X^{(1)})^2 \right) \xrightarrow{p} 2T \int_0^T \rho_s (\sigma_s^{(1)})^3 \sigma_s^{(2)} ds.$$

The key steps for proving (4) are the approximation

$$\Delta_i X \approx \sigma_{\frac{(i-1)T}{n}} \left(W_{\frac{iT}{n}} - W_{\frac{(i-1)T}{n}} \right), \quad (5)$$

more precisely given in the Appendix A, and the formula

$$\mathbb{C}\text{ov} \left(Z^{(i)} Z^{(l)}, Z^{(m)} Z^{(u)} \right) = \Sigma_{im} \Sigma_{lu} + \Sigma_{iu} \Sigma_{lm}. \quad (6)$$

for a multivariate normal $Z \sim \mathbf{N}(0, \Sigma)$ with covariance matrix (Σ_{ij}) . The right-hand side of (5) is conditionally on $\mathcal{F}_{(i-1)T/n}$ centered Gaussian and this finding will be helpful, since by the martingale structure of realized (co-)variances and estimation errors in the upcoming sections below, the asymptotic covariances are given as limit of the sequence of conditional covariances. Hence it will be possible to apply (6) which is a special case of the general formula for moments from a multivariate normal by Isserlis (1918).

3 Inference for observations with microstructure noise

Assumption* 2. *The process X is observed synchronously on $[0, T]$ with additive microstructure noise:*

$$Y_i = X_{t_i} + \epsilon_i, i = 0, \dots, n.$$

The $t_i, 0 \leq i \leq n$, are the observation times and we assume that there is a constant $0 < \alpha \leq 1/9$, such that

$$\delta_n = \sup_i ((t_i - t_{i-1}), t_0, T - t_n) = \mathcal{O} \left(n^{-8/9-\alpha} \right), \quad (7)$$

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. Furthermore, the errors have mean zero, and fourth moments exist.

Exact orders in (7) and below in (18) and (25) arise from upper bounds of remainder terms after applying Hölder inequality. We keep to the notation

$$\Delta_i X = X_{t_i} - X_{t_{i-1}} \text{ and } \Delta_i^j X = X_{t_i} - X_{t_{i-j}}, 1 \leq i \leq n, 2 \leq j \leq i.$$

Since notation varies between papers, note the correspondence to the other main form:

$$\Delta_i X \text{ is the same as } \Delta X_{t_i}.$$

The covariance matrix of the vectors $\epsilon_j, 0 \leq j \leq n$, is denoted \mathbf{H} and for $p = 4$ we set

$$\mathbf{H} = \begin{pmatrix} \eta_1^2 & \eta_{12} & \eta_{13} & \eta_{14} \\ \eta_{12} & \eta_2^2 & \eta_{23} & \eta_{24} \\ \eta_{13} & \eta_{23} & \eta_3^2 & \eta_{34} \\ \eta_{14} & \eta_{24} & \eta_{34} & \eta_4^2 \end{pmatrix}. \quad (8)$$

Note that 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 of asymptotic variances, we restrict ourselves to i. i. d. noise in this section – in the general asynchronous framework below asymptotic covariances of generalized multi-scale estimates are not affected by the noise. Increments in such a 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, since the second addend is $\mathcal{O}_p(\delta_n^{1/2})$ and the drift acts only as nuisance term of order in probability $\mathcal{O}_p(\delta_n)$. For an accurate estimation of the integrated covariance 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.

3.1 Asymptotic Equivalence of the Multi-Scale and Kernel Estimators

For the estimation of integrated variance the following rate-optimal estimators with similar asymptotic behavior have been proposed in the literature: a multi-scale approach by Zhang (2006), pre-averaging noisy returns first as in Jacod *et al.* (2009), the kernel approach by Barndorff-Nielsen *et al.* (2008) and a Quasi-Maximum-Likelihood-Estimator by Xiu (2010). We investigate the covariance structure of the multi-scale estimator explicitly, but since 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 multi-scale estimator

$$[\widehat{X^{(1)}, X^{(2)}}]_T^{(multi)} = \sum_{i=1}^{M_n^{(12)}} \frac{\alpha_i}{i} \sum_{j=i}^n \Delta_j^i Y^{(1)} \Delta_j^i Y^{(2)}, \quad (9)$$

and analogous for other components, arises as linear combination of subsampling estimators that are averaged lower-frequent realized covariances using frequencies $i = 1, \dots, M_n^{(12)}$. For discrete weights $\alpha_i, 1 \leq i \leq M_n$, with $\sum_{i=1}^{M_n} \alpha_i = 1$ and $\sum_{i=1}^{M_n} (\alpha_i/i) = 0$, the expression

$$\alpha_i = \frac{i}{M_n^2} h\left(\frac{i}{M_n}\right) - \frac{i}{2M_n^3} h'\left(\frac{i}{M_n}\right) + \frac{i}{6M_n^4} (h'(1) - h'(0)) - \frac{i}{24M_n^5} (h''(1) - h''(0)), \quad (10)$$

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 frequencies are chosen $M_n^{(kl)} = c_{kl} \sqrt{n}$ with constants $c_{kl}, (k, l) \in \{1, 2, 3, 4\}^2$, 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 (2011b).

At the present day, it is commonly known that the nonparametric smoothing approaches to cope with noise contamination have a connatural structure and related asymptotic distribution. A prominent intensively

studied alternative to the multi-scale approach is the kernel estimator by Barndorff-Nielsen *et al.* (2008)

$$\begin{aligned} [\widehat{X^{(1)}, X^{(2)}}]_T^{(kernel)} &= \sum_{j=1}^n \Delta_j Y^{(1)} \Delta_j Y^{(2)} \\ &+ \sum_{h=1}^{H_n} \mathfrak{K}\left(\frac{h}{H_n}\right) \left(\sum_{j=h+1}^n \Delta_j Y^{(1)} \Delta_{j-h} Y^{(2)} + \Delta_{j-h} Y^{(1)} \Delta_j Y^{(2)} \right), \end{aligned} \quad (11)$$

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

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

In the one-dimensional setup (11) has the shape of a 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 scalar model provided by Barndorff-Nielsen *et al.* (2008). Since 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.

Theorem 3.1. *For each kernel function \mathfrak{K} matching the assertions above, for the estimator defined in (11) and the multi-scale estimator (9) with weights determined by (10) and $h = \mathfrak{K}'$, we have*

$$n^{1/4} \left([\widehat{X^{(1)}, X^{(2)}}]_T^{(multi)} - [\widehat{X^{(1)}, X^{(2)}}]_T^{(kernel)} + 4\eta_{12} \right) \xrightarrow{p} 0, \quad (12)$$

as $n \rightarrow \infty$, $M_n = H_n \rightarrow \infty$. The term $4\eta_{12}$ is due to the different impact of end-effects in (11) and (9) and, since the variance-covariance structure carries over to adjusted unbiased versions of the estimators, is not crucial for the relation of the asymptotic (co)variances.

3.2 Asymptotic Equivalence of Adjusted Estimators

The multi-scale and kernel estimators defined in (9) and (11) are sensitive to end-effects which is caused by the dominating noise component (which does not depend on n). Due to end-effects, on Assumption *2, the estimators (9) and (11) with weights determined by (10) and corresponding kernels have a bias $-2\eta_{12}$ and $2\eta_{12}$, 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

$$\alpha_1 \mapsto \alpha_1 + 2/n, \alpha_2 \mapsto \alpha_2 - 2/n, (\alpha_i)_{3 \leq i \leq M_n} \mapsto (\alpha_i)_{3 \leq i \leq M_n}. \quad (13)$$

Correction to the Kernel estimator:

$$\text{multiplying the realized covariance in the first addend with } \frac{n-1}{n}. \quad (14)$$

This correction is different from the ‘jittering’ approach provided in Barndorff-Nielsen *et al.* (2008).⁹

We call the adjusted estimators, respectively,

$$[\widehat{X^{(1)}, X^{(2)}}]_T^{(multi, adj)} \text{ and } [\widehat{X^{(1)}, X^{(2)}}]_T^{(kernel, adj)}.$$

With these adjustments, we obtain the following direct equivalence of the two estimators.

Theorem 3.2. *Under the assumptions of Theorem 3.1, for each kernel function \mathfrak{K} , we have*

$$n^{1/4} \left([\widehat{X^{(1)}, X^{(2)}}]_T^{(multi, adj)} - [\widehat{X^{(1)}, X^{(2)}}]_T^{(kernel, adj)} \right) \xrightarrow{p} 0, \quad (15)$$

as $n \rightarrow \infty$, $M_n = c_{multi}\sqrt{n}$ and $H_n = c_{kernel}\sqrt{n}$.

⁹Section 2.6 p. 1487-88 of Barndorff-Nielsen *et al.* (2008).

kernel	\mathfrak{K}
cubic	$1 - 3x^2 + 2x^3$
Parzen	$(1 - 6x^2 + 6x^3)\mathbb{1}_{\{x \leq 1/2\}} + 2(1 - x)^3\mathbb{1}_{\{x > 1/2\}}$
r th 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 $i \leq M/2$ and $\frac{i}{M^2} \left(12 - \frac{12i}{M} \right)$ for $i > M/2$
r th Tukey-Hanning	$\frac{\pi i r (1 - \frac{i}{M})^{r-2} ((r-1) \sin(\pi(1 - \frac{i}{M})^r) + \pi r (\frac{i}{M} - 1)^r \cos(\pi(1 - \frac{i}{M})^r))}{2M^2}$

Table 1: Collection of important kernels and corresponding weights for the multi-scale (first order term).

The extension from $H_n = M_n$ in Theorem 3.1 to asymptotically of the same (optimal) order follows directly, by inserting the minimum in the transformations in the proof, and by elementary bounds for the remainder.

Remark 1. (*Dependent noise.*) In the case of m -dependence it will be convenient to discard the first m frequencies and renormalize in (9). The adjusted estimator is robust.

Remark 2. (*Strong representation.*) The results in Theorems 3.1-3.2 are 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 to strong representation through a standard subsequence-of-subsequence argument.)

3.3 Optimal choice of weights, and Asymptotic distribution

The standard weights employed in Zhang (2006) and Bibinger (2011b):

$$\begin{aligned} \alpha_i &= \left(\frac{12i^2}{((M_n^{(12)})^3 - M_n^{(12)})} - \frac{6i}{((M_n^{(12)})^2 - 1)} - \frac{6i}{((M_n^{(12)})^3 - M_n^{(12)})} \right) \\ &= \frac{12i^2}{(M_n^{(12)})^3} - \frac{6i}{(M_n^{(12)})^2} (1 + o(1)) \end{aligned} \quad (16)$$

minimize the asymptotic noise variance and lead to, as mentioned by Barndorff-Nielsen et al. (2008), the same asymptotic properties as for the kernel estimator (11) 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-)variance, see (17) below, 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’ attain the optimal asymptotic variance in the scalar case known from the inverse Fisher information in Gloter & Jacod (2001). All weights (10) 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 asymptotic covariance structure for the typically considered equidistant observations setup and we will extend this to irregular sampling below in our general non-synchronous model.

kernel	\mathfrak{N}_1^α	\mathfrak{D}^α	\mathfrak{M}^α	\mathfrak{N}_2^α
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

Table 2: Constants in asymptotic covariance for important kernels.

Proposition 3.1. *On the Assumptions 1 and * 2 with $t_i = iT/n, 0 \leq i \leq n$, the asymptotic covariance of the multi-scale estimates (9) with $M_n^{(12)} = c_{12} \sqrt{n}$, $M_n^{(34)} = c_{34} \sqrt{n}$, $c = \min(c_{12}, c_{34})$, and weights (10), and by the equivalence also of the corresponding kernel estimates, is*

$$\begin{aligned} \mathbb{ACOV} \left([X^{(1)}, X^{(2)}]_T^{(multi)}, [X^{(3)}, X^{(4)}]_T^{(multi)} \right) &= 2\mathfrak{D}^\alpha c T \int_0^T (\sigma_s^{(13)} \sigma_s^{(24)} + \sigma_s^{(14)} \sigma_s^{(23)}) ds \\ &+ 2\mathfrak{N}_1^\alpha c^{-3} (\eta_{13}\eta_{24} + \eta_{14}\eta_{23}) + c^{-1}\mathfrak{M}^\alpha \int_0^T (\eta_{13}\sigma_s^{(24)} + \eta_{24}\sigma_s^{(13)} + \eta_{14}\sigma_s^{(23)} + \eta_{23}\sigma_s^{(14)}) ds \\ &+ c^{-1}\mathfrak{N}_2^\alpha (\eta_{13}\eta_{24} + \eta_{14}\eta_{23}), \end{aligned} \quad (17)$$

with constants \mathfrak{D}^α , \mathfrak{N}_1^α , \mathfrak{N}_2^α and \mathfrak{M}^α depending on the specific kernel, see Table 2.¹⁰

A generalization for non-equidistant sampling is covered by Proposition 5.1 in Section 5. In this case, the first addend of (17) (signal term) hinges on a function (29), while the other terms are analogous. Inserting ‘(1) = (2) = (3) = (4)’ in the general formula (17), we obtain the asymptotic variance of the one-dimensional multi-scale estimator as given in Zhang (2006). Also, for (1) = (3) and (2) = (4), we have the asymptotic variance of the integrated covariance multi-scale estimator as given for the special case in which \mathbf{H} is diagonal in Bibinger (2012).

4 Inference for non-synchronous observations in the Absence of Microstructure noise

This section is devoted to the estimation problem under sampling with different observation schemes in each component. Denote $n = n_1 + n_2 + n_3 + n_4$, the total number of observations. For deterministic real sequences we introduce the notation $a_n \sim b_n$ to express shortly that $a_n = \mathcal{O}(b_n)$ and $b_n = \mathcal{O}(a_n)$. For sequences of random variables we analogously use $a_n \sim^p b_n$ if $a_n = \mathcal{O}_p(b_n)$ and $b_n = \mathcal{O}_p(a_n)$.

Assumption 2.** *Components of the process X are observed discretely at times $t_i^{(l)}, 0 \leq i \leq n_l, l = 1, 2, 3, 4$, which follow sequences of observation schemes for which $n_1 \sim n_2 \sim n_3 \sim n_4$, and with a constant $0 < \alpha \leq 1/3$ it holds that*

$$\delta_n = \sup_{(i,l)} \left(\left(t_i^{(l)} - t_{i-1}^{(l)} \right), t_0^{(l)}, T - t_{n_l}^{(l)} \right) = \mathcal{O} \left(n^{-2/3-\alpha} \right). \quad (18)$$

Similarly as before, we denote componentwise

$$\Delta_i X^{(l)} = X_{t_i^{(l)}}^{(l)} - X_{t_{i-1}^{(l)}}^{(l)} \text{ and } \Delta_i^j X^{(l)} = X_{t_i^{(l)}}^{(l)} - X_{t_{i-j}^{(l)}}^{(l)}, 1 \leq i \leq n_l, 2 \leq j \leq i, l \in \{1, 2, 3, 4\}.$$

¹⁰General expressions for the constants \mathfrak{D}^α , \mathfrak{N}_1^α , \mathfrak{N}_2^α and \mathfrak{M}^α are the same as in Zhang (2006) and Bibinger (2012), since they apply equally in the special cases discussed there. In this paper, we do not focus so much on the theoretical expressions for constants, as our focus is in the feasible CLT.

Under non-synchronous sampling the equitable estimator for the integrated covariance is the following generalized realized covariance by Hayashi & Yoshida (2005):

$$[\widehat{X^{(1)}, X^{(2)}}]_T^{(HY)} = \sum_{i=1}^{n_1} \sum_{j=1}^{n_2} \Delta_i X^{(1)} \Delta_j X^{(2)} \mathbb{1}_{[\min(t_i^{(1)}, t_j^{(2)}) > \max(t_{i-1}^{(1)}, t_{j-1}^{(2)})]} , \quad (19)$$

where the sum comprises all products of increments with overlapping observation time instants. This estimator is asymptotically unbiased, i.e. unbiased and UMVU in the absence of a drift term, and can be deduced as Maximum-Likelihood estimator in a model with deterministic function Σ_t as illustrated in Mykland (2012). If supposed that we have sequences of sampling schemes for which some characteristic features, specified in detail below, have a limit describing an asymptotic behavior of asynchronicity, a stable central limit theorem with optimal convergence rate \sqrt{n} has been established and there are also feasible versions (cf. Hayashi & Yoshida (2011) and Bibinger (2011a)). The asymptotic variance is in general larger than for (4) in the synchronous case and hinges on functions capturing the superposition of the two sampling times designs. For this reason, the analysis of the covariance structure in a multi-dimensional setup gets more involved, since for a covariance we confront a superposition of four different sampling schemes instead of two. For a more illustrative description, we use the illuminative rewriting of the Hayashi-Yoshida estimator from Bibinger (2011a) with a synchronous approximation and an uncorrelated addend due to the lack of synchronicity.

For this purpose we introduce the notion of next- and previous-tick interpolations:

$$t_l^+(s) = \min_{i \in \{0, \dots, n_l\}} (t_i^{(l)} | t_i^{(l)} \geq s) \text{ and } t_l^-(s) = \max_{i \in \{0, \dots, n_l\}} (t_i^{(l)} | t_i^{(l)} \leq s)$$

for $l \in \{1, 2, 3, 4\}$ and $s \in [0, T]$. One way to rewrite (19) using telescoping sums is:

$$\begin{aligned} [\widehat{X^{(1)}, X^{(2)}}]_T^{(HY)} &= \sum_{i=1}^{n_1} \Delta_i X^{(1)} \left(X_{t_2^+(t_i^{(1)})}^{(2)} - X_{t_2^-(t_{i-1}^{(1)})}^{(2)} \right) \\ &= \sum_{j=1}^{n_2} \Delta_j X^{(2)} \left(X_{t_1^+(t_j^{(2)})}^{(1)} - X_{t_1^-(t_{j-1}^{(2)})}^{(1)} \right). \end{aligned}$$

For the generalization of the idea of closest synchronous approximations define

$$T_0^{12} = \max(t_1^+(0), t_2^+(0)), T_i^{12} = T_{i-1}^{12} + \max(t_1^+(T_{i-1}^{12}), t_2^+(T_{i-1}^{12})), i = 1, \dots, N_{12},$$

$$T_0^{34} = \max(t_3^+(0), t_4^+(0)), T_i^{34} = T_{i-1}^{34} + \max(t_3^+(T_{i-1}^{34}), t_4^+(T_{i-1}^{34})), i = 1, \dots, N_{34}.$$

The times T_i^{12}, T_i^{34} , are the refresh times from Barndorff-Nielsen *et al.* (2011) built for each pair of processes and thus we will refer to these times, which coincide with the ones defined in a slightly different manner in Bibinger (2011a), as refresh times in the following. The notion of next- and previous-ticks will be applied analogously as above to refresh times:

$$T_{12}^+(s) = \min_{i \in \{0, \dots, N_{12}\}} (T_i^{12} | T_i^{12} \geq s) \text{ and } T_{34}^+(s) = \min_{i \in \{0, \dots, N_{34}\}} (T_i^{34} | T_i^{34} \geq s),$$

and $T_{12}^-(s), T_{34}^-(s)$ in the same fashion for $s \in [0, T]$. Writing $X_{T_{12}^+}^{(l),+}$ for $X_{t_l^+(T_{12})}^{(l)}$, $l = 1, 2$, and $X_{T_{12}^+}^{(l),-}$ for $X_{t_l^-(T_{12})}^{(l)}$, $l = 1, 2$, the Hayashi-Yoshida estimator (19) can be illustrated

$$[\widehat{X^{(1)}, X^{(2)}}]_T^{(HY)} = \sum_{i=1}^{N_{12}} \left(X_{T_i^{12}}^{(1),+} - X_{T_{i-1}^{12}}^{(1),-} \right) \left(X_{T_i^{12}}^{(2),+} - X_{T_{i-1}^{12}}^{(2),-} \right), \quad (20)$$

where $N_{12} \leq \min(n_1, n_2)$ is the number of refresh times T_i^{12} . This illustration is particularly useful

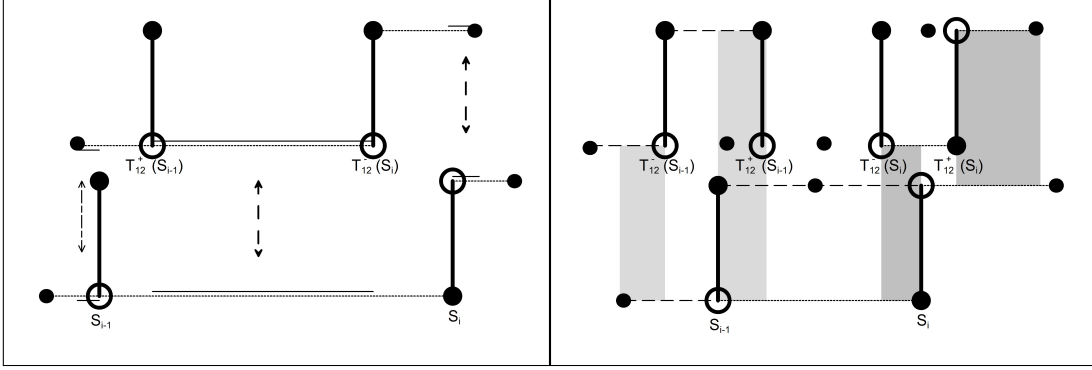


Figure 1: Interaction of next-tick and previous-tick interpolation terms affecting the covariance.

to decompose the Hayashi-Yoshida estimator in two uncorrelated addends - a fictional non-observable synchronous realized covariance and the error due to the lack of synchronicity:

$$[\widehat{X^{(1)}, X^{(2)}}]_T^{(HY)} = D_T^{12} + A_T^{12} \quad (21a)$$

with

$$D_T^{12} = \sum_{j=1}^{N_{12}} \left(X_{T_j^{12}}^{(1)} - X_{T_{j-1}^{12}}^{(1)} \right) \left(X_{T_j^{12}}^{(2)} - X_{T_{j-1}^{12}}^{(2)} \right), \quad (21b)$$

$$\begin{aligned} A_T^{12} = & \sum_{j=1}^{N_{12}} \left(X_{T_j^{12}}^{(1),+} - X_{T_j^{12}}^{(1)} \right) \left(X_{T_j^{12}}^{(2)} - X_{T_{j-1}^{12}}^{(2),-} \right) + \left(X_{T_j^{12}}^{(2),+} - X_{T_j^{12}}^{(2)} \right) \left(X_{T_j^{12}}^{(1)} - X_{T_{j-1}^{12}}^{(1),-} \right) \\ & + \left(X_{T_{j-1}^{12}}^{(1)} - X_{T_{j-1}^{12}}^{(1),-} \right) \left(X_{T_j^{12}}^{(2)} - X_{T_{j-1}^{12}}^{(2)} \right) + \left(X_{T_{j-1}^{12}}^{(2)} - X_{T_{j-1}^{12}}^{(2),-} \right) \left(X_{T_j^{12}}^{(1)} - X_{T_{j-1}^{12}}^{(1)} \right). \end{aligned} \quad (21c)$$

Each increment in (20) is the sum of three addends, the increment over the refresh time instant and a previous- and next-tick interpolated increment, where in each product only one of the next-tick and previous-tick increments is non-zero.

The covariance is hence given by

$$\begin{aligned} \mathbb{E} [D_T^{12} D_T^{34} + A_T^{12} D_T^{34} + A_T^{34} D_T^{12} + A_T^{12} A_T^{34}] - \mathbb{E} [D_T^{12}] \mathbb{E} [D_T^{34}] \\ = \mathbb{E} [D_T^{12} D_T^{34}] + \mathbb{E} [A_T^{12} A_T^{34}] - \mathbb{E} [D_T^{12}] \mathbb{E} [D_T^{34}] + o(1). \end{aligned}$$

The products $D_T^{12} A_T^{34}$ and $D_T^{34} A_T^{12}$ tend to zero in probability, since the expectation for the Brownian parts equals zero by one factor in each addend of an interpolated increment over an interpolation time instant disjoint to the instants of the other factors. The product $D_T^{12} D_T^{34}$ comprises the refresh time sampling for the pairs of processes. It can be decomposed passing over to refresh times of refresh times (which are the refresh times of all four processes in the definition of Barndorff-Nielsen *et al.* (2011)) in the way: $D_T^{12} D_T^{34} = D_T^{12,34} + A_T^{12,34}$. The idea is the same as for the usual Hayashi-Yoshida estimator (20) as we have two non-synchronous sampling designs by the sequences $T_i^{12}, 0 \leq i \leq N_{12}, T_l^{34}, 0 \leq l \leq N_{34}$. For this purpose define

$$S_0 = \max(T_0^{12}, T_0^{34}), S_i = S_{i-1} + \max(T_{12}^+(S_{i-1}), T_{34}^+(S_{i-1})), i = 1, \dots, N.$$

Here, different as for the variance of one Hayashi-Yoshida estimator the term $A_T^{12,34}$, comprising interpolations from $T_i^{12}, 0 \leq i \leq N_{12}$, and $T_l^{34}, 0 \leq l \leq N_{34}$, to the times $S_k, 0 \leq k \leq N$, will not contribute to the asymptotic covariance. The term has in general a non-zero expectation depending on $\sigma^{(12)}\sigma^{(34)}$,

but is compensated by terms in $\mathbb{E} [D_T^{12}] \mathbb{E} [D_T^{34}]$. Therefore, two terms will constitute the asymptotic covariance of two Hayashi-Yoshida estimators. A simple covariance term as in the presence of synchronous observations at the refresh times $S_k, 0 \leq k \leq N$, and the product $A_T^{12} A_T^{34}$. For the latter all products of terms of the above given type including next- and previous ticks which overlap in time can contribute to its expectation. To quantify this for general observation schemes we consider an auxiliary structure founded on the partition of $[0, T]$ in the refresh time increments $(S_k - S_{k-1}), k = 1, \dots, N$. For each increment $(S_k - S_{k-1}), k = 1, \dots, N$, we consider the refresh times T_i^{12}, T_j^{34} , lain next outside the interval and within the interval, i. e. $T_{12}^+(S_k), T_{12}^-(S_k), T_{34}^+(S_k), T_{34}^-(S_k)$, where at least one of the pairs equals S_k and the same for S_{k-1} . This is visualized in Figure 1 in which observations at pairwise refresh times are illustrated by dots and “missing” observations by circles. On the left-hand side arrows and bars and on the right-hand side colored rectangles mark the intervals where interpolations for pairs intersect and thus contribute to the covariance. Although we have no further interest what observations take place between the considered refresh times, in particular more pairwise refresh times of one pair can lie in this interval, interpolation terms of those will not overlap with one of the other pair. We only need to focus on the depicted borders of the intervals and the correlation between adjacent steps. As visualized on the left side of Figure 1, next-tick interpolation terms at S_i and the last precedent refresh time of the other pair have a non-empty intersection and contribute to the covariance. It is also possible - as highlighted on the right side of Figure 1 - that the next-tick interpolation term at the subsequent refresh time of the other pair and at S_i overlap and have a non-zero correlation. The same findings apply to previous-tick interpolation terms. Together with the covariance between next- and previous-tick steps at the same S_i of adjacent terms, which can be quantified similarly as for the variance in Bibinger (2011a), we have captured all possible terms triggering the covariance by non-synchronous sampling.

Definition 1. Define the deterministic sequences

$$G^N(t) = \frac{N}{T} \sum_{S_i \leq t} (S_i - S_{i-1})^2, \quad (22a)$$

$$\begin{aligned} F_{24}^{13,N}(t) = \frac{N}{T} \sum_{S_{i+1} \leq t} & (\min(t_1^+(S_i), t_3^+(S_i)) - S_i)^+ (S_i - \max(t_2^-(S_{i-1}), t_4^-(S_{i-1}))) \\ & + (S_i - \min(t_1^-(S_{i-1}), t_3^-(S_{i-1})))^+ (S_i - S_{i-1}) \\ & + (\min(t_2^+(S_i), t_4^+(S_i)) - S_i)^+ (S_i - \max(t_1^-(S_{i-1}), t_3^-(S_{i-1}))) \\ & + (S_i - \min(t_2^-(S_{i-1}), t_4^-(S_{i-1})))^+ (S_i - S_{i-1}), \end{aligned} \quad (22b)$$

$$\begin{aligned} H_{24}^{13,N}(t) = \frac{N}{T} \sum_{S_{i+1} \leq t} & (\min(t_1^+(S_i), t_3^+(S_i)) - S_i)^+ (S_{i+1} - \max(t_2^-(S_{i+1}), t_4^-(S_{i+1})))^+ \\ & + (\min(t_2^+(S_i), t_4^+(S_i)) - S_i)^+ (S_{i+1} - \max(t_1^-(S_{i+1}), t_3^-(S_{i+1})))^+, \end{aligned} \quad (22c)$$

$$I_{24}^{13,N}(t) = \frac{N}{T} \sum_{S_{i+1} \leq t} (S_i - \max(T_{12}^-(S_i), T_{34}^-(S_i)))^+ \quad (22d)$$

$$\begin{aligned} & \times (\min(\max(t_1^+(S_i), t_3^+(S_i), \max(t_1^+(T_{12}^+(S_i)), t_3^+(T_{34}^+(S_i)))) - \max(T_{12}^+(S_i), T_{34}^+(S_i)))^+ \\ & + (\min(T_{12}^-(S_{i-1}), T_{34}^-(S_{i-1})) - \max(t_1^-(S_{i-1}), t_3^-(S_{i-1})))^+ (\min(T_{12}^+(S_{i-1}), T_{34}^+(S_{i-1})) - S_{i-1})^+ \\ & + (S_i - \max(T_{12}^-(S_i), T_{34}^-(S_i)))^+ \\ & \times (\min(\max(t_2^+(S_i), t_4^+(S_i), \max(t_2^+(T_{12}^+(S_i)), t_4^+(T_{34}^+(S_i)))) - \max(T_{12}^+(S_i), T_{34}^+(S_i)))^+ \\ & + (\min(T_{12}^-(S_{i-1}), T_{34}^-(S_{i-1})) - \max(t_2^-(S_{i-1}), t_4^-(S_{i-1})))^+ (\min(T_{12}^+(S_{i-1}), T_{34}^+(S_{i-1})) - S_{i-1})^+ \end{aligned}$$

where $(\cdot)^+$ denotes the positive part, and analogous sequences $F_{23}^{14,N}, H_{23}^{14,N}, I_{23}^{14,N}$ of monotone increasing functions on $[0, T]$. These are called quadratic covariations of times in the sequel.

Assumption 3. Assume that for the sequences (22a), (22b), (22c) and (22d) from Definition 1 the following convergence assertions hold:

- (i) $G^N(t) \rightarrow G(t)$, $F_{24}^{13,N}(t) \rightarrow F_{24}^{13}$, $F_{23}^{14,N}(t) \rightarrow F_{23}^{14}$, $H_{24}^{13,N}(t) \rightarrow H_{24}^{13}(t)$, $H_{23}^{14,N}(t) \rightarrow H_{23}^{14}(t)$, $I_{24}^{13,N}(t) \rightarrow I_{24}^{13}(t)$, $I_{23}^{14,N}(t) \rightarrow I_{23}^{14}(t)$ as $N \rightarrow \infty$, where the limits are continuously differentiable functions on $[0, T]$.

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

$$\frac{G^N(t + h_N) - G^N(t)}{h_N} \rightarrow G'(t), \quad (23a)$$

$$\frac{F_{24}^{13,N}(t + h_N) - F_{24}^{13,N}(t)}{h_N} \rightarrow F_{24}^{13'}(t), \quad (23b)$$

$$\frac{H_{24}^{13,N}(t + h_N) - H_{24}^{13,N}(t)}{h_N} \rightarrow H_{24}^{13'}(t), \quad (23c)$$

$$\frac{I_{24}^{13,N}(t + h_N) - I_{24}^{13,N}(t)}{h_N} \rightarrow I_{24}^{13'}(t), \quad (23d)$$

uniformly on $[0, T]$ as $N \rightarrow \infty$ and analogously for $H_{23}^{14,N}$, $F_{23}^{14,N}$, $I_{23}^{14,N}$. The limiting functions are called asymptotic quadratic covariations of times in the sequel.

Assumption 3, which generalizes the notion of a quadratic variation of time for the one-dimensional case by Zhang *et al.* (2005), directly postulates that sequences of time instants appearing in the sequences of conditional covariances after applying Itô isometry and (5) converge to some limit. This is a rather weak condition and necessary for the existence of asymptotic covariances.

Proposition 4.1. On the Assumptions 1, **2 and 3 the asymptotic covariance of two Hayashi-Yoshida estimators is

$$\begin{aligned} \mathbb{A}\text{COV} \left(\left[\widehat{X^{(1)}}, \widehat{X^{(2)}} \right]_T^{(HY)}, \left[\widehat{X^{(3)}}, \widehat{X^{(4)}} \right]_T^{(HY)} \right) &= T \int_0^T G'(s) (\sigma_s^{(13)} \sigma_s^{(24)} + \sigma_s^{(14)} \sigma_s^{(23)}) ds \\ &+ T \int_0^T (F_{24}^{13'} + H_{24}^{13'} + I_{24}^{13'}) \sigma_s^{(13)} \sigma_s^{(24)} ds + T \int_0^T (F_{23}^{14'} + H_{23}^{14'} + I_{23}^{14'}) \sigma_s^{(23)} \sigma_s^{(14)} ds. \end{aligned} \quad (24)$$

For ‘(1) = (3)’ and ‘(2) = (4)’, we find the asymptotic variance of the Hayashi-Yoshida estimator as illustrated in Bibinger (2011a). In this case $I_{24}^{13'}$, $I_{23}^{14'}$ and $H_{23}^{14'}$ are zero.

5 The general case: Asymptotic covariance matrix of the generalized multi-scale estimates under Asynchronicity and Microstructure

This section focuses on the general model – comprising non-synchronous sampling and noise perturbation – and an hybrid approach founded on a combination of the estimators from Sections 3 and 4.

Assumption* 2.** The process X is observed non-synchronously with additive microstructure noise:

$$Y_{t_j^{(l)}}^{(l)} = X_{t_j^{(l)}}^{(l)} + \epsilon_j^{(l)}, j = 0, \dots, n_l, l \in \{1, 2, 3, 4\}.$$

The sequences of observation schemes are regular in the sense that $n_1 \sim n_2 \sim n_3 \sim n_4$ and with a constant $0 < \alpha \leq 1/9$ it holds that

$$\delta_n = \sup_{(i,l)} \left(\left(t_i^{(l)} - t_{i-1}^{(l)} \right), t_0^{(l)}, T - t_{n_l}^{(l)} \right) = \mathcal{O} \left(n^{-s/9-\alpha} \right). \quad (25)$$

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

In the following we establish the asymptotic covariance matrix for the generalized multi-scale method from Bibinger (2011b) and Bibinger (2012). It arises as a convenient composition of the multi-scale realized (co-)variance by Zhang (2006) from Section 3 and a synchronization approach inspired by the estimator by Hayashi & Yoshida (2005). Virtually we can think of an idealized synchronous approximation by 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. Reviving the notation from Section 4, the generalized multi-scale estimator can be illustrated:

$$[\widehat{X^{(1)}, X^{(2)}}]_T^{(multi)} = \sum_{i=1}^{M_n^{(12)}} \frac{\alpha_i}{i} \sum_{j=i}^{N_{12}} \left(X_{T_j^{12}}^{(1),+} - X_{T_{j-i}^{12}}^{(1),-} \right) \left(X_{T_j^{12}}^{(2),+} - X_{T_{j-i}^{12}}^{(2),-} \right). \quad (26)$$

This estimator crucially differs from the approach by Christensen *et al.* (2011), which has 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 (26) relies more on the principle of the refresh-time approximation – but contrary to Barndorff-Nielsen *et al.* (2011) – we use pre- and next-tick interpolations such that the final estimator has no bias due to non-synchronicity. For this reason the article on hand can not accomplish a unified theory that is applicable to alternative approaches by Barndorff-Nielsen *et al.* (2011), Aït-Sahalia *et al.* (2010) and Christensen *et al.* (2011), since unlike their roots from Section 3 they are not connatural any more. We consider (26) because the method attains a much smaller discretization variance in comparison to the one by Christensen *et al.* (2011), is rate-optimal and a feasible central limit theorem is accessible from Bibinger (2012).

Remark 3. (*Identical results for kernel estimators.*) Since equations (7) and (25) are the same, it follows from Section 3 that our results on irregular sampling for the synchronous case, where the generalized multi-scale estimator (26) coincides with the original one (9), in the following apply identically to kernel estimators. Furthermore, all results for the estimator (26) apply to a generalized kernel estimator with pairwise refresh time sampling as in (26).

Definition 2. For given sampling schemes $t_j^{(l)}, 0 \leq j \leq n_l, 1 \leq l \leq p = 4$ and $r < n_l$, define the functional sequences

$$\mathfrak{G}_{n,r}^{(l)}(t) := \frac{n_l}{rT} \sum_{t_j^{(l)} \leq t} (t_j^{(l)} - t_{j-1}^{(l)}) \sum_{q=0}^{r \wedge j} (t_{j-q}^{(l)} - t_{j-q-1}^{(l)}), \quad (27)$$

for each component and analogously for refresh times $T_j^{kl}, 0 \leq j \leq N_{kl}, (k, l) \in \{1, 2, 3, 4\}^2$, introduced in Section 3:

$$\mathfrak{G}_{N_{kl},r}^{(l,k)}(t) := \frac{N_{kl}}{rT} \sum_{T_j^{kl} \leq t} (T_j^{kl} - T_{j-1}^{kl}) \sum_{q=0}^{r \wedge j} (T_{j-q}^{kl} - T_{j-q-1}^{kl}). \quad (28)$$

Denote $\mathfrak{G}_{N,r}(t)$ as the function build in the same fashion from the refresh times of all four observed components $S_j, 0 \leq j \leq N$.

The existence of a limit \mathfrak{G} of the sequence in Definition 2 is essential to establish an asymptotic distribution theory, since 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:

$$D^\alpha(t) := \lim_{N \rightarrow \infty} \left(\frac{N}{M_N T} \sum_{S_r \leq t} \Delta S_r \sum_{i,k=1}^{M_N} \alpha_i \alpha_k \sum_{q=0}^{r \wedge i \wedge k} \left(1 - \frac{q}{i}\right) \left(1 - \frac{q}{k}\right) \Delta S_{r-q} \right). \quad (29)$$

In the equidistant setup $D^\alpha(t) = \mathfrak{D}^\alpha t$ with the constant \mathfrak{D}^α found in Proposition 3.1. We will call the limit \mathfrak{G} in case of existence local asymptotic sampling autocorrelation (LASA). If we focus on the special case ‘(1)=(2)=(3)=(4)’, convergence of (27) is assumed for the one component.

The preceding definition suffices to quantify the influence of non-equidistant synchronous schemes on the asymptotic properties of the multi-scale estimator, but to give a very general asymptotic covariance structure of the generalized multi-scale estimator in a transparent form, we can not avoid to introduce some tedious notation in the following. Readers interested mainly in the usual completely non-synchronous setup, where the asymptotic covariances of estimates involving different components only hinge on the signal parts, may proceed with Corollary 5.2.

Definition 3. Depending on the sequences of sampling schemes, define the following sequences of functions:

$$S_{13}^N(t) = \frac{1}{N} \sum_{t_j^{(1)} \leq t} \sum_{t_k^{(3)} \leq t} \mathbb{1}_{\{t_j^{(1)} = t_k^{(3)}\}}, \quad (30a)$$

and in the same way $S_{14}^N(t)$, $S_{23}^N(t)$ and $S_{24}^N(t)$. Define in the case of existence for given $M_N^{(12)}$, $M_N^{(34)}$ with $M_N = \min(M_N^{(12)}, M_N^{(34)})$:

$$\begin{aligned} \mathfrak{S}_{13}^{24} = \lim_{N \rightarrow \infty} N^{-1} M_N^{-1} \sum_{j=0}^{N_{12}} \sum_{k=0}^{N_{34}} \sum_{r=1}^{j \wedge M_N^{(12)}} \sum_{q=1}^{k \wedge M_N^{(34)}} & \left(\mathbb{1}_{\{t_1^+(T_j^{12}) = t_3^+(T_k^{34}), t_2^-(T_{j-r}^{12}) = t_4^-(T_{j-q}^{34})\}} \right. \\ & \left. + \mathbb{1}_{\{t_2^+(T_j^{12}) = t_4^+(T_k^{34}), t_1^-(T_{j-r}^{12}) = t_3^-(T_{j-q}^{34})\}} \right), \end{aligned} \quad (31a)$$

$$\begin{aligned} \mathfrak{S}_{14}^{23} = \lim_{N \rightarrow \infty} N^{-1} M_N^{-1} \sum_{j=0}^{N_{12}} \sum_{k=0}^{N_{34}} \sum_{r=1}^{j \wedge M_N^{(12)}} \sum_{q=1}^{k \wedge M_N^{(34)}} & \left(\mathbb{1}_{\{t_1^+(T_j^{12}) = t_4^+(T_k^{34}), t_2^-(T_{j-r}^{12}) = t_3^-(T_{j-q}^{34})\}} \right. \\ & \left. + \mathbb{1}_{\{t_2^+(T_j^{12}) = t_3^+(T_k^{34}), t_1^-(T_{j-r}^{12}) = t_4^-(T_{j-q}^{34})\}} \right), \end{aligned} \quad (31b)$$

$$\begin{aligned} \tilde{\mathfrak{S}}_{13}^{24} = \lim_{N \rightarrow \infty} M_N^{-1} \sum_{j=0}^{M_N^{(12)}-1} \sum_{k=0}^{M_N^{(34)}-1} & \left(\mathbb{1}_{\{t_1^+(T_j^{12}) = t_3^+(T_k^{34}), t_2^+(T_j^{12}) = t_4^+(T_k^{34})\}} \right. \\ & \left. + \mathbb{1}_{\{t_1^-(T_{N_{12}-j}^{12}) = t_3^-(T_{N_{34}-k}^{34}), t_2^-(T_{N_{12}-j}^{12}) = t_4^-(T_{N_{34}-k}^{34})\}} \right), \end{aligned} \quad (31c)$$

$$\begin{aligned} \tilde{\mathfrak{S}}_{14}^{23} = \lim_{N \rightarrow \infty} M_N^{-1} \sum_{j=0}^{M_N^{(12)}-1} \sum_{k=0}^{M_N^{(34)}-1} & \left(\mathbb{1}_{\{t_1^+(T_j^{12}) = t_4^+(T_k^{34}), t_2^+(T_j^{12}) = t_3^+(T_k^{34})\}} \right. \\ & \left. + \mathbb{1}_{\{t_1^-(T_{N_{12}-j}^{12}) = t_4^-(T_{N_{34}-k}^{34}), t_2^-(T_{N_{12}-j}^{12}) = t_3^-(T_{N_{34}-k}^{34})\}} \right). \end{aligned} \quad (31d)$$

Assumption* 3. Assume that for the sequences (27)/(28) from Definition 2, and the sequences from Definition 3, the following convergence assumptions hold:

- (i) As $N \rightarrow \infty$ and $r \rightarrow \infty$ with $r = \mathcal{O}(N)$: $\mathfrak{S}_{N,r}(t) \rightarrow \mathfrak{S}(t)$, for a continuous differentiable limiting function \mathfrak{S} on $[0, T]$.
- (ii) For any null sequence (h_N) , $h_N = \mathcal{O}(N^{-1})$:

$$\frac{\mathfrak{S}_{N,r}(t + h_N) - \mathfrak{S}_{N,r}(t)}{h_N} \rightarrow \mathfrak{S}'(t) \quad (32)$$

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

- (iii) $S_{13}^N(t) \rightarrow S_{13}$, $S_{14}^N(t) \rightarrow S_{14}(t)$, $S_{23}^N(t) \rightarrow S_{23}(t)$, $S_{24}^N(t) \rightarrow S_{24}(t)$ as $N \rightarrow \infty$, where the limits are continuously differentiable functions on $[0, T]$.

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

$$\frac{S_{13}^N(t + h_N) - S_{13}^N(t)}{h_N} \rightarrow S'_{13}(t), \quad \frac{S_{14}^N(t + h_N) - S_{14}^N(t)}{h_N} \rightarrow S'_{14}(t), \quad (33a)$$

$$\frac{S_{23}^N(t + h_N) - S_{23}^N(t)}{h_N} \rightarrow S'_{23}(t), \quad \frac{S_{24}^N(t + h_N) - S_{24}^N(t)}{h_N} \rightarrow S'_{24}(t), \quad (33b)$$

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

Proposition 5.1. *On the Assumptions 1, **2 and *3, the asymptotic covariance of generalized multi-scale estimates (26) with $M_n^{(12)} = c_{12}\sqrt{N_{12}} = \tilde{c}\sqrt{N}$, $M_n^{(34)} = c_{34}\sqrt{N_{34}} = c^*\sqrt{N}$, $c = \min(\tilde{c}, c^*)$, and weights (10) is given by*

$$\begin{aligned} \mathbb{ACOV} \left(\left[\widehat{X^{(1)}}, \widehat{X^{(2)}} \right]_T^{(multi)}, \left[\widehat{X^{(3)}}, \widehat{X^{(4)}} \right]_T^{(multi)} \right) &= 2cT \int_0^T (D^\alpha)'(s) (\sigma_s^{(13)} \sigma_s^{(24)} + \sigma_s^{(14)} \sigma_s^{(23)}) ds \\ &+ c^{-3} (\mathfrak{S}_{13}^{24} \mathfrak{C}_{13,\alpha}^{24} \eta_{13} \eta_{24} + \mathfrak{S}_{14}^{23} \mathfrak{C}_{14,\alpha}^{23} \eta_{14} \eta_{23}) + c^{-1} \left(\tilde{\mathfrak{S}}_{13}^{24} \tilde{\mathfrak{C}}_{13,\alpha}^{24} \eta_{13} \eta_{24} + \tilde{\mathfrak{S}}_{14}^{23} \tilde{\mathfrak{C}}_{14,\alpha}^{23} \eta_{14} \eta_{23} \right) \\ &+ c^{-1} \mathfrak{M}^\alpha \int_0^T \left(\eta_{13} S'_{13} \sigma_s^{(24)} + \eta_{24} S'_{24} \sigma_s^{(13)} + \eta_{14} S'_{14} \sigma_s^{(23)} + \eta_{23} S'_{23} \sigma_s^{(14)} \right) ds, \end{aligned} \quad (34)$$

with the existing differentiable limiting function (29) hinging on (28). All constants $\mathfrak{S}\mathfrak{C}_{,\alpha}$, $\tilde{\mathfrak{S}}\tilde{\mathfrak{C}}_{,\alpha}$ depend on the asymptotic proportion of sampling times where pairs of two components are recorded synchronously and the selected weights.

In a synchronous setting clearly $\mathfrak{S}_{13}^{24} = \mathfrak{S}_{14}^{23} = \tilde{\mathfrak{S}}_{13}^{24} = \tilde{\mathfrak{S}}_{14}^{23} = 1$, $\mathfrak{C}_{13}^{24} = \mathfrak{C}_{14}^{23} = 2\mathfrak{M}_1^\alpha$ (=24 for the cubic kernel), $\tilde{\mathfrak{C}}_{13}^{24} = \tilde{\mathfrak{C}}_{14}^{23} = \mathfrak{M}_2^\alpha$ (=6/5 for the cubic kernel) and $S'_{13} = S'_{14} = S'_{23} = S'_{24} = \mathbb{1}_{[0,T]}$, and (34) coincides with (17) except the influence of irregular sampling. In particular, the asymptotic variance of the multi-scale estimator for synchronous non-equidistant sampling yields

$$\begin{aligned} \mathbf{AVAR} \left(\left[\widehat{X^{(1)}}, \widehat{X^{(2)}} \right]_T^{(multi)} \right) &= 2cT \int_0^T (D^\alpha)'(s) (\sigma_s^{(11)} \sigma_s^{(22)} + (\sigma_s^{(12)})^2) ds \\ &+ 2\mathfrak{M}_1^\alpha c^{-3} (\eta_1^2 \eta_2^2 + \eta_{12}^2) + c^{-1} \mathfrak{M}^\alpha \int_0^T \left(\eta_1^2 \sigma_s^{(22)} + \eta_2^2 \sigma_s^{(11)} + 2\eta_{12} \sigma_s^{(12)} \right) ds \\ &+ c^{-1} \mathfrak{M}_2^\alpha (\eta_1^2 \eta_2^2 + \eta_{12}^2). \end{aligned} \quad (35)$$

Interestingly, in most situations if $\mathfrak{S}_{13}^{24} = \mathfrak{S}_{14}^{23} = 0$, the noise part will vanish. Furthermore, we obtain the following important result for the completely non-synchronous case:

Corollary 5.2. *In the case that no synchronous observations take place: $t_i^{(l)} \neq t_j^{(k)}$ for all $l \neq k$ and $(i, j) \in \{0, \dots, n_l\} \times \{0, \dots, n_k\}$, or the amount of synchronous observations tends to zero as $n \rightarrow \infty$, using the same notation as in Proposition 5.1 and on the same (remaining) assumptions, we have:*

$$\mathbb{ACOV} \left(\left[\widehat{X^{(1)}}, \widehat{X^{(2)}} \right]_T^{(multi)}, \left[\widehat{X^{(3)}}, \widehat{X^{(4)}} \right]_T^{(multi)} \right) = 2cT \int_0^T (D^\alpha)'(s) (\sigma_s^{(13)} \sigma_s^{(24)} + \sigma_s^{(14)} \sigma_s^{(23)}) ds, \quad (36)$$

with the existing derivative of the limiting function (29).

Here, formula (36) makes sense only for different components and we may not insert '(1)=(3)' or likewise as above. However, the relation is meaningful for the asymptotic covariance between integrated variance estimators.

Remark 4. *Our major focus is not on the theoretical limits \mathfrak{G} and of other sequences, since in the general case they are specified only as limits. We do not need these values, however, for inference, as we shall see in the next section on feasible inference.*

Note that convergence of (27) 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 variance to an asymptotic local autocorrelation of sampling time instants which constitutes the counterpart to the sum of squared time instants emerging in the variance for subsampling and the other smoothing approaches. The only difference is that not directly the limit of (27) will appear in the asymptotic variance, but some limiting function additionally involving specific weights (the kernel). If we think of random sampling independent of Y , the structure of (27) will be particularly simple for i. i. d. time instants. Virtually, only the expectation will matter and we can apply the law of large numbers. Assuming (32) is less restrictive than the assertion in Zhang (2006), i. e. sampling needs not to be close to an equidistant scheme in the sense that asymptotic quadratic variation of time converges to T at T . Remarkably, for the popular model of homogenous Poisson sampling independent of X with expected time instants T/n , the asymptotic variance (35) 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 (27) as $r \rightarrow \infty$. The finite sample correction factor in (32) for this Poisson setup is thus $(r + 1)/r$.

At first glance the simple appearance of the covariance between generalized multi-scale estimates in the typical general setup where all observations are non-synchronous and in the presence of microstructure noise is intriguing. The covariance hinges only on the discretization error as if we had synchronous observations at the refresh times $S_i, i = 0, \dots, N$. The noise falls out of the asymptotic covariance on the assumption that observation errors at different observation times are independent.

This feature constitutes another nice property of the generalized multi-scale method that not only the asymptotic variance of the estimator has a simple form, where the impact of non-synchronicity falls out in the signal part, but the asymptotic covariances are even simpler and particularly are not influenced by the superposition of the four underlying different sampling schemes. Here, we benefit by the fact that for the construction of (26) a smoothing method to reduce noise contamination is utilized which at the same time smoothes out interpolation effects and hence the error due to non-synchronicity.

6 A feasible multivariate stable central limit theorem

In the sequel, we conclude a feasible stable central limit theorem for linear combinations of estimated entries of the integrated covariance matrix with estimators of the types considered above. A remaining step towards a feasible asymptotic distribution theory allowing to draw statistical inference, is to provide consistent estimators for the asymptotic covariances from Propositions 2.1, 3.1, 4.1 and 5.1. In the vein of Bibinger (2012), we construct consistent asymptotic covariance estimators in the general non-synchronous framework, following a histogram-type approach. For the simple case in the absence of noise and non-synchronicity, we give a simple-structured estimator resembling the prominent bipower variation.

Proposition 6.1. *In the setting of Section 2, the estimator*

$$\begin{aligned} \widehat{\mathbb{ACOV}} & \left(\sum_{i=1}^n \Delta_i X^{(1)} \Delta_i X^{(2)}, \sum_{i=1}^n \Delta_i X^{(3)} \Delta_i X^{(4)} \right) \\ &= \frac{n}{T} \sum_{i=1}^{n-1} \Delta_i X^{(1)} \Delta_{i+1} X^{(2)} \Delta_i X^{(3)} \Delta_{i+1} X^{(4)} + \Delta_{i+1} X^{(1)} \Delta_i X^{(2)} \Delta_i X^{(3)} \Delta_{i+1} X^{(4)}, \end{aligned} \quad (37)$$

is a consistent estimator of the general asymptotic covariance according to (4).

On the assumptions imposed in Proposition 5.1, the estimator

$$\begin{aligned}
& \widehat{\text{ACOV}} \left(\left[\widehat{X^{(1)}}, \widehat{X^{(2)}} \right]_T^{(multi)}, \left[\widehat{X^{(3)}}, \widehat{X^{(4)}} \right]_T^{(multi)} \right) \\
&= 2cT \sum_{j=1}^{K_N} \frac{\Delta[\widehat{X^{(1)}}, \widehat{X^{(3)}}]^n \Delta[\widehat{X^{(2)}}, \widehat{X^{(4)}}]^n + \Delta[\widehat{X^{(1)}}, \widehat{X^{(4)}}]^n \Delta[\widehat{X^{(2)}}, \widehat{X^{(3)}}]^n}{(\Delta D_j^N)^2} \frac{D_N^\alpha(T)}{K_N} \\
&+ c^{-3} \left(\hat{\mathfrak{S}}_{13}^{24} \hat{\mathfrak{C}}_{13,\alpha}^{24} \hat{\eta}_{13} \hat{\eta}_{24} + \hat{\mathfrak{S}}_{14}^{23} \hat{\mathfrak{C}}_{14,\alpha}^{23} \hat{\eta}_{14} \hat{\eta}_{23} \right) + c^{-1} \left(\hat{\mathfrak{S}}_{13}^{24} \hat{\mathfrak{C}}_{13,\alpha}^{24} \hat{\eta}_{13} \hat{\eta}_{24} + \hat{\mathfrak{S}}_{14}^{23} \hat{\mathfrak{C}}_{14,\alpha}^{23} \hat{\eta}_{14} \hat{\eta}_{23} \right) \quad (38) \\
&+ c^{-1} \mathfrak{M}^\alpha \left(\sum_{j=1}^{K_N} \frac{\Delta[\widehat{X^{(2)}}, \widehat{X^{(4)}}]^n}{\Delta_j S_{13}^N} \frac{S_{13}^N(T)}{K_N} \hat{\eta}_{13} + \sum_{j=1}^{K_N} \frac{\Delta[\widehat{X^{(1)}}, \widehat{X^{(3)}}]^n}{\Delta_j S_{24}^N} \frac{S_{24}^N(T)}{K_N} \hat{\eta}_{24} \right. \\
&\left. + \sum_{j=1}^{K_N} \frac{\Delta[\widehat{X^{(1)}}, \widehat{X^{(4)}}]^n}{\Delta_j S_{23}^N} \frac{S_{23}^N(T)}{K_N} \hat{\eta}_{23} + \sum_{j=1}^{K_N} \frac{\Delta[\widehat{X^{(2)}}, \widehat{X^{(3)}}]^n}{\Delta_j S_{14}^N} \frac{S_{14}^N(T)}{K_N} \hat{\eta}_{14} \right),
\end{aligned}$$

gives a consistent histogram-wise estimation of the general asymptotic covariance from (34). Here we use

$$D_j^N := \inf \{t \in [0, T] | D^\alpha(t) \geq j D^\alpha(T)/K_N\}, 0 \leq j \leq K_N, \Delta D_j^N = D_j^N - D_{j-1}^N,$$

$$(S_{lk}^N)_j := \inf \{t \in [0, T] | S_{lk}(t) \geq j S_{lk}(T)/K_N\}, 0 \leq j \leq K_N, l = 1, 2, k = 3, 4,$$

$$\Delta_j S_{lk}^N = (S_{lk}^N)_j - (S_{lk}^N)_{j-1}, 1 \leq j \leq K_N,$$

$$\hat{\eta}_{kl} = -(N S_{kl}(T))^{-1} \sum_{i=1}^{N S_{kl}(T)} \Delta_i X^{(k)} \Delta_{i+1} X^{(l)}, k = 1, 2, l = 3, 4,$$

and empirical realizations for all functions based on the sampling design from Definition 3. The number of bins K_N satisfies $K_N N^{-1/3} \rightarrow 0$ as $K_N \rightarrow \infty$. The binwise evaluated estimators in (38) are multi-scale estimators on each bin with multi-scale frequencies $M_N(j), 1 \leq j \leq K_N$. One possible choice is $K_N = cN^{1/5}$ and $M_N(j) = c^{5/4} N^{3/5}$.

Remark 5. The feasible CLT remains valid when relaxing Assumptions of Proposition 5.1 on existence of the limit \mathfrak{G} in Assumption *3, since every subsequence of (29) has an in probability converging subsequence, see the discussion at the end of p. 1411 in Zhang et al. (2005) for analogous reasoning and more details.

The estimator (38) simplifies in many cases, i.e. the completely non-synchronous setup, to the first addend. An estimator for the asymptotic covariance (17) of multi-scale estimators in the synchronous case is inherent in (38) when inserting appropriate constants. One estimator for the non-noisy but non-synchronous model and the asymptotic covariance given in (24) may be constructed in a similar fashion as (38) with histogram-type estimators based on equispaced grids with respect to the quadratic covariations of times introduced in Definition 1. Since the principle is clear from the above given estimator, we forego to explicitly state that estimator. In the synchronous setup or for $l = k$, the noise (co-)variance can be estimated $\sqrt{n_l}$ consistently by the realized (co-)variance or using adjacent increments as stated above. For '(1) = (2) = (3) = (4)', our estimator (37) becomes $2n \sum_i (\Delta_i X)^2 (\Delta_{i+1} X)^2$ and differs from the standard estimator $(2n/3) \sum_i (\Delta_i X)^4$ proposed in Barndorff-Nielsen & Shephard (2002) which is preferable because its variance $42/3 n^{-1}$ is slightly smaller than $48 n^{-1}$ of (37) for this case. The constant c in Proposition 6.1 is fixed from the minimum constant for the multi-scale estimates here. For an algorithm how to select the tuning parameter adaptively involving pilot estimates to calibrate the whole estimation procedure first, we refer to Bibinger (2012) for the generalized multi-scale and Barndorff-Nielsen et al. (2008) for the univariate kernel estimator.

For a quadratic symmetric $(p \times p)$ matrix $A \in \text{sym}(p) \subset \mathbb{R}^{p \times p}$, we denote the mapping to the vector of its $p(p+1)/2$ free entries

$$\begin{aligned} \text{SVEC}(A) &= ((A_{rq})_{1 \leq r \leq p, 1 \leq q \leq r})^\top \\ &= (A_{11}, A_{12}, \dots, A_{1p}, A_{22}, A_{23}, \dots, A_{2p}, \dots, A_{p-1p}, A_{pp})^\top. \end{aligned}$$

Theorem 6.1. Denote $[\widehat{X^{(k)}}, \widehat{X^{(l)}}]^n, 1 \leq k \leq p, 1 \leq l \leq p$, one of the integrated covariance matrix estimators from Sections 2–5 and consider the vector $\text{SVEC} \left(\left([\widehat{X^{(k)}}, \widehat{X^{(l)}}]^n \right)_{1 \leq k \leq p, 1 \leq l \leq p} \right)$ of estimated entries. The estimators fulfill a multivariate stable central limit theorem with rate r_n :

$$r_n \left(\text{SVEC} \left(\left([\widehat{X^{(k)}}, \widehat{X^{(l)}}]^n - [X^{(k)}, X^{(l)}] \right)_{1 \leq k \leq p, 1 \leq l \leq p} \right) \right) \xrightarrow{\text{stably}} \mathbf{N}(0, \text{COV}), \quad (39)$$

with a symmetric $p(p+1)/2 \times p(p+1)/2$ -dimensional asymptotic covariance matrix COV determined by Propositions 2.1, 3.1, 4.1 and 5.1, respectively. The rate r_n equals \sqrt{n} , $n = \sum_{l=1}^p n_l$, in the non-noisy experiment and $n^{1/4}$ under microstructure noise. For linear combinations

$$Z := \left[\sum_k c_k \widehat{X^{(k)}} \right]^n = \sum_{k,l} c_k c_l [\widehat{X^{(k)}}, \widehat{X^{(l)}}]^n$$

with the consistent estimators

$$\widehat{\text{AVAR}}(Z) = \sum_{k, \bar{k}, l, \bar{l}} c_k c_{\bar{k}} c_l c_{\bar{l}} \widehat{\text{ACOV}} \left([\widehat{X^{(\bar{k})}}, \widehat{X^{(\bar{l})}}]^n, [\widehat{X^{(k)}}, \widehat{X^{(l)}}]^n \right),$$

if the latter is strictly positive, a feasible central limit theorem holds:

$$r_n \left(Z / \sqrt{\widehat{\text{AVAR}}(Z)} \right) \xrightarrow{\text{weakly}} \mathbf{N}(0, 1). \quad (40)$$

Note that we rescale the entries of the asymptotic covariance matrix COV from Propositions 2.1, 3.1, 4.1 and 5.1, respectively, and its estimates with factors $r_N/r_n \left((N/n)^{(1/2)} \right)$ without and $(N/n)^{(1/4)}$ with noise) to obtain (39) and (40), where N denotes the number of refresh times of the involved components. The multivariate central limit theorem (39) is directly derived from the multi-dimensional version of Theorem 2–1 by Jacod (1997) which provides the basis for stable limit theorems in this research area. By virtue of the asymptotic covariance structure deduced in this work, it suffices to check the conditions that covariations of the componentwise estimation errors with the underlying reference Brownian motion W driving our efficient process tend to zero, as well as the covariations with all \mathcal{F}_t -adapted bounded martingales orthogonal to W . The proof by establishing elementary bounds for the variances of these terms is along the same lines as the univariate analysis, see Bibinger (2012) and Christensen *et al.* (2011), among others. Suppose without loss of generality for notational convenience that we have

$$(X_n, Y_n) \xrightarrow{\text{stably}} (X, Y) \stackrel{\mathcal{L}}{=} \mathbf{N} \left(0, \begin{pmatrix} V_X & C_{XY} \\ C_{XY} & V_Y \end{pmatrix} \right).$$

By continuous mapping we conclude $X_n + Y_n \rightarrow X + Y$ stably. Assume we have at hand consistent estimators $V_X^n \xrightarrow{p} V_X, V_Y^n \xrightarrow{p} V_Y, C_{XY}^n \xrightarrow{p} C_{XY}$, and set $V^n = V_X^n + V_Y^n + 2C_{XY}^n$ which is assumed to be strictly positive. As V^n is a bounded \mathcal{F} -measurable random variable, stable convergence implies $(X_n + Y_n, V^n) \xrightarrow{\text{weakly}} (X + Y, V_X + V_Y + 2C_{XY})$ and hence also $(X_n + Y_n)/\sqrt{V^n} \xrightarrow{\text{weakly}} \mathbf{N}(0, 1)$.

7 An application for 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 third one to which both are correlated. For multi-dimensional portfolio modeling and

management, information about such relations can provide valuable information and access to a new angle on the covariance structure. Conclusions in the way that a significant integrated covariance between two high-frequency assets might be fully explained by their dependence on a joint factor or another asset, respectively, will be of revealing impact on the used models. One interesting case we may think of is that two observed asset processes X_1 and X_2 are listed within one index recorded as high-frequency process Z . We then ask if X_1 and X_2 are conditionally on Z independent. To put it the other way round, pairs which are not conditionally independent given some other asset exhibit some crucial covariance that carries information about the direct mutual influence between the two assets. We understand independence here in terms of orthogonal quadratic covariation processes and will further restrict to specific hypotheses as below where we test for zero integrated covariances – so the term ‘independence’ is misused here for a simple illustrative phrasing. The role of Z in the model can be also some macro variable that is known or can be estimated with faster rate. We set up a statistical experiment in which 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 .

$$X_1 = \rho^{X_1} Z + Z^\perp, \quad X_2 = \rho^{X_2} Z + Z^\dagger \quad \text{with } [Z, Z^\perp] \equiv 0, [Z, Z^\dagger] \equiv 0. \quad (41)$$

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

$$\mathbb{H}_0 : [Z^\perp, Z^\dagger]_T = 0. \quad (42)$$

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 covariance over $[0, T]$.

A suitable test statistic to decide whether we reject \mathbb{H}_0 or not is

$$\mathfrak{T}(X_1, X_2, Z) = [X_1, Z]_T [X_2, Z]_T - [X_1, X_2]_T [Z]_T, \quad (43)$$

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

$$\hat{\mathfrak{T}}_n = [\widehat{X_1}, \widehat{Z}]_T^n [\widehat{X_2}, \widehat{Z}]_T^n - [\widehat{X_1}, \widehat{X_2}]_T^n [\widehat{Z}]_T^n, \quad (44)$$

where $[\cdot]_T^n$ stands for one of the aforementioned estimators satisfying (39) whichever model fits the data best. This test statistic, though based on the simple function $g(x, y, u, v) = xy - uv$, is more complex to analyze than linear combinations, since we face products of our estimators. In lieu of determining the distribution of the test statistic directly, we find help in the asymptotic covariance structure provided in this paper and the Δ -method for stable convergence. Here, the methodology is similar to the prominent propagation of error concept from experimental science. For each quadratic covariation, the estimation error gets small for large n and hence we can profit when we Taylor the underlying function g . Indeed, this will give us the leading term of the variance of $\hat{\mathfrak{T}}_n$:

$$\begin{aligned} \mathfrak{T} - \hat{\mathfrak{T}}_n &= [X_2, Z]_T \left([X_1, Z]_T - [\widehat{X_1}, \widehat{Z}]_T^n \right) + [X_1, Z]_T \left([X_2, Z]_T - [\widehat{X_2}, \widehat{Z}]_T^n \right) \\ &\quad - [X_1, X_2]_T \left([Z]_T - [\widehat{Z}]_T^n \right) - [Z]_T \left([X_1, X_2]_T - [\widehat{X_1}, \widehat{X_2}]_T^n \right) + \mathcal{O}_p(r_n^{-4}). \end{aligned} \quad (45)$$

The asymptotic variance of the test statistic is random as we are familiar with from the usual mixed normal limits in this field. It is a linear combination of the unknown covariations we estimate and the asymptotic covariances which are the topic of this article and can be estimated consistently as found in Proposition

6.1. An elementary calculation yields

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

Inserting consistent estimators for the asymptotic covariances above, we finally obtain with (39) that

$$\mathbb{P}_0 \left(r_n \frac{\hat{\mathfrak{T}}_n}{\widehat{\mathbf{AVAR}}(\hat{\mathfrak{T}}_n)} \right) \xrightarrow{weakly} \mathbf{N}(0, 1), \quad (46)$$

where \mathbb{P}_0 denotes the distribution under \mathbb{H}_0 and an asymptotic distribution free test for \mathbb{H}_0 .

The conditional independence test also provides a tool for investigating dependencies within vast dimensional portfolios in order to choose small blocks, i. e. subportfolios for which covariances of the estimates are quantified explicitly.

8 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 2010. We consider a sample portfolio with 5 assets, namely Apple (AAPL), Microsoft (MSFT), Oracle (ORCL), Exxon Mobil Corporation (XOM) and Pfizer (PFE). We quantify the complete asymptotic covariance matrix of generalized multi-scale estimates with weights (16) for the integrated covariance matrix over the whole month (where we discard over-night returns) and for the first trading day, 2010/08/02. The data provides a good training data set to analyze estimation and test procedures.

For a p -dimensional portfolio, the number of free entries of the symmetric covariance matrix is given by

$$\frac{1}{2} \frac{p(p+1)}{2} \left(\frac{p(p+1)}{2} + 1 \right) = p + 3 \binom{p}{4} + 3 \cdot 2 \binom{p}{3} + 4 \binom{p}{2}. \quad (47)$$

The left-hand illustration is derived by the reasoning that we estimate $p(p+1)/2$ different entries of the symmetric integrated covariance matrix which leads to a $(p(p+1)/2)^2$ -dimensional covariance matrix which is symmetric again. Since the dimension of the asymptotic covariance matrix increases proportional to p^4 , an evaluation of all covariances is tractable only for small values of p (our estimates of risk in a p -dimensional portfolio have dimension proportional to p^2 and the risk of the estimates proportional to p^4). In Table 3, we list the estimates for integrated covariances \pm estimated standard deviation. With the estimators (38) and the numbers of refresh times of incorporated components, we quantify covariances of the generalized multi-scale estimates which are listed in Table 4. The bottom line is that involving asymptotic covariances is indispensable when facing questions for multivariate portfolio management. The estimated quadratic variation of a sum of all five assets is $(41.57 \pm 0.26) \cdot 10^{-3}$ for 2010/08 and $(129.22 \pm 6.88) \cdot 10^{-5}$ for 2010/08/02. The estimated asymptotic variances, $6.93 \cdot 10^{-8}$ and $47.42 \cdot 10^{-10}$, are mainly induced by estimated covariance terms (6.34/42.91), whereas the trace of the matrix, i. e. the sum of estimates for asymptotic variances, is smaller. If one would mistakenly act as if the estimators were uncorrelated, this leads to a tremendous underestimate of the estimation risk. Note that in principle, e. g. for negatively correlated assets, the risk could sometimes also decrease by adding covariances. For [MSFT+ORCL], which have the highest correlation in our portfolio, the ratios of the estimated sum of variances to the estimated sum of all covariances are 0.14/0.45 and 0.85/3.30, respectively, and for the least correlated pairs, [AAPL+XOM] and [PFE+XOM], respectively, we have 0.09/0.20 and 0.84/1.59.

We perform the test from Section 7, to investigate three hypotheses: if MSFT and ORCL have a zero co-variation conditional on PFE; ORCL and PFE conditional on MSFT and MSFT and PFE conditional on

$[\widehat{X_1}, \widehat{X_2}]_T^{multi}$	AAPL	MSFT	ORCL	XOM	PFE
AAPL	2.82 ± 0.01	1.23 ± 0.02	1.66 ± 0.03	0.63 ± 0.01	0.99 ± 0.03
MSFT		2.87 ± 0.03	2.05 ± 0.05	0.94 ± 0.02	1.25 ± 0.07
ORCL			3.81 ± 0.06	1.31 ± 0.03	1.63 ± 0.10
XOM				2.36 ± 0.01	1.25 ± 0.04
PFE					3.83 ± 0.09

$[\widehat{X_1}, \widehat{X_2}]_T^{multi}$	AAPL	MSFT	ORCL	XOM	PFE
AAPL	8.16 ± 0.10	3.42 ± 0.20	3.74 ± 0.20	2.28 ± 0.11	2.47 ± 0.26
MSFT		12.43 ± 0.38	7.43 ± 0.37	2.95 ± 0.20	4.69 ± 0.47
ORCL			11.90 ± 0.39	3.73 ± 0.22	2.90 ± 0.49
XOM				6.42 ± 0.12	1.75 ± 0.29
PFE					19.59 ± 0.70

Table 3: Estimates for integrated covariances ($\cdot 10^3$) 2010/08 (top) and ($\cdot 10^5$) 2010/08/02 (bottom).

ORCL. From the higher correlation of MSFT and ORCL, we may expect that the first will probably not apply. We obtain the following three test results (two-sided test) with p -values p and statistics Q which are asymptotically standard normal under H_0 :

$$Q = -8.10; -2.47; -1.85 \quad p = 0.00; 0.014; 0.064 \quad (2010/08),$$

$$Q = -10.00; -0.05; -1.24 \quad p = 0.00; 0.96; 0.21 \quad (2010/08/02).$$

Tests if MSFT and ORCL have zero covariation conditional on the sum of all assets yield p -values 0.13 for 2010/08 and 0.23 for 2010/08/02. Tests for ORCL and PFE conditional on the sum of the three other assets yield p -values 0.19 and 0.99. In conclusion, this empirical evidence suggests that MSFT and ORCL have some dependence not explained by a common macro factor influencing all NASDAQ assets. Contrary, we can not reject this for several other combinations. Some differences between 2010/08 and 2010/08/02 give an heuristic that the portfolio dependence structure is not completely persistent. Though 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 in this article provide reliable tools to quantify risk measures and determine confidence intervals from high-frequently documented asset prices.

	[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]	4.01	4.53	5.47	1.55	5.26	6.97	6.75	4.58	6.92	9.33	5.15	8.28	3.02	3.46	10.89
[A,M]		4.91	4.63	1.89	2.72	7.48	7.74	3.33	2.72	8.74	2.88	9.30	1.22	2.88	12.23
[A,O]			7.21	1.85	4.50	5.85	2.89	2.91	7.26	11.62	4.70	9.61	2.03	2.80	3.85
[A,X]				2.83	2.85	6.15	5.98	3.44	6.71	8.03	1.94	7.49	3.97	4.59	2.96
[A,P]					7.03	4.30	5.63	2.72	8.96	6.71	0.63	11.34	1.37	4.71	13.72
[M,M]						13.31	12.29	5.81	8.18	8.06	3.33	6.72	1.21	2.90	5.33
[M,O]							12.36	5.86	10.18	15.29	5.71	11.07	1.35	3.61	6.32
[M,X]								5.24	6.95	6.03	5.48	6.15	2.72	4.51	6.54
[M,P]									13.86	6.82	4.02	14.42	1.82	6.25	12.47
[O,O]										21.04	7.85	12.93	1.46	3.98	7.50
[O,X]											6.67	9.08	2.97	4.74	7.98
[O,P]												17.28	1.72	7.63	15.67
[X,X]													4.90	4.92	13.11
[X,P]														8.20	12.23
[P,P]															29.98

	[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]	1.55	2.01	1.75	0.77	0.65	1.52	1.50	0.41	0.62	1.91	0.73	1.01	0.27	0.31	2.05
[A,M]		2.64	2.32	0.92	1.78	5.67	4.32	2.46	4.27	3.12	1.79	2.89	0.56	1.08	3.05
[A,O]			2.35	1.12	1.22	2.96	4.05	1.42	2.35	5.21	2.47	3.54	0.70	0.77	2.63
[A,X]				0.86	0.68	1.84	1.85	1.38	0.95	2.18	1.53	1.05	1.04	0.90	2.14
[A,P]					3.29	2.52	1.80	0.91	5.57	1.81	0.61	5.64	0.28	3.29	7.99
[M,M]						13.78	9.94	4.79	11.05	5.72	2.73	4.59	0.68	1.74	4.66
[M,O]							7.27	4.1	5.86	9.65	4.33	6.79	0.84	1.64	3.03
[M,X]								2.3	2.68	3.15	2.46	2.05	1.35	2.35	2.53
[M,P]									9.21	2.60	0.90	9.49	0.34	4.56	15.32
[O,O]										12.21	5.63	5.51	0.92	1.14	2.84
[O,X]											2.63	1.78	1.72	3.17	1.43
[O,P]												9.88	0.33	5.18	7.03
[X,X]													1.62	1.08	2.2
[X,P]														3.84	5.42
[P,P]															34.16

Table 4: Estimated asymptotic covariance matrix ($\cdot 10^8$), 2010/08 (top) and ($\cdot 10^{10}$) 2010/08/02 (below), A=AAPL, M=MSFT, O=ORCL, X=XOM, P=PFE.

Acknowledgements

The authors would like to thank Dacheng Xiu, Ruey Tsay and Dan Christina Wang for inspiring discussions. We also thank Johannes Schmidt-Hieber and Till Sabel for a discussion on the relation of estimators via quadratic forms for noise smoothing who have worked on the transformations between several of these estimators.

Markus Bibinger was supported by a fellowship within the Postdoc-Programme of the German Academic Exchange Service (DAAD) and the CRC 649 ‘Economic Risk’ at Berlin, supported by the Deutsche Forschungsgemeinschaft (DFG), and also from the Stevanovich Center for Financial Mathematics at the University of Chicago. Per Mykland was supported under National Science Foundation grant SES 11-24526. We gratefully acknowledge this financial support.

The NASDAQ trading data used in Section 8 has been reconstructed from the order book with LOBSTER provided by the high-frequency econometrics team at the Chair of Econometrics, School of Business and Economics, Humboldt-Universität zu Berlin, Germany.

References

- Aït-Sahalia, Y., J. Fan, & D. Xiu (2010). High-frequency estimates with noisy and asynchronous financial data. *Journal of the American Statistical Association* 105(492), 1504–1516.
- Aït-Sahalia, Y., L. Zhang, & P. A. Mykland (2011). Ultra high frequency volatility estimation with dependent microstructure noise. *Journal of Econometrics*, 160, 160–165.
- Andersen, T. G. & T. Bollerslev (1998). Answering the skeptics: Yes, standard volatility models do provide accurate forecasts. *International Economic Review* 39, 885–905.
- Andersen, T. G., T. Bollerslev, F. X. Diebold, & P. Labys (2000). Great realizations. *Risk* 13, 105–108.
- Andersen, T. G., T. Bollerslev, F. X. Diebold, & P. Labys (2001). The distribution of realized exchange rate volatility. *Journal of the American Statistical Association* 96, 42–55.
- Andersen, T. G., T. Bollerslev, F. X. Diebold, & P. Labys (2003). Modeling and forecasting realized volatility. *Econometrica* 71, 579–625.
- Barndorff-Nielsen, O. E., P. R. Hansen, A. Lunde, & N. Shephard (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., P. R. Hansen, A. Lunde, & N. Shephard (2011). Multivariate realised kernels: consistent positive semi-definite estimators of the covariation of equity prices with noise and non-synchronous trading. *Journal of Econometrics* 162(2), 149–169.
- Barndorff-Nielsen, O. E. & N. Shephard (2001). Non-Gaussian Ornstein-Uhlenbeck-based models and some of their uses in financial economics. *Journal of the Royal Statistical Society, B* 63, 167–241.
- Barndorff-Nielsen, O. E. & N. Shephard (2002). Econometric analysis of realized volatility and its use in estimating stochastic volatility models. *Journal of the Royal Statistical Society* 64(2), 253–280.
- Bibinger, M. (2011a). Asymptotics of asynchronicity. *Technical Report, Humboldt-Universität zu Berlin*, URL=<http://sfb649.wiwi.hu-berlin.de/papers/pdf/SFB649DP2011--033.pdf>.
- Bibinger, M. (2011b). Efficient covariance estimation for asynchronous noisy high-frequency data. *Scandinavian Journal of Statistics* 38, 23–45.
- Bibinger, M. (2012). An estimator for the quadratic covariation of asynchronously observed itô processes with noise: Asymptotic distribution theory. *Stochastic Processes and their Applications* 122, 2411–2453.

- Bibinger, M., M. Reiß, N. Hautsch, & P. Malec (2012). Asymptotically efficient covolatility matrix estimation via local likelihoods. *preprint, HU Berlin*.
- Black, F. S. & M. S. Scholes (1973). The pricing of options and corporate liabilities. *Journal of Political Economy* 81, 637–654.
- Christensen, K., S. Kinnebrock, & M. Podolskij (2010). Pre-averaging estimators of the ex-post covariance matrix in noisy diffusion models with non-synchronous data. *Journal of Econometrics*, 159(1), 116–133.
- Christensen, K., M. Podolskij, & M. Vetter (2011). On covariation estimation for multivariate continuous itô semimartingales with noise in non-synchronous observation schemes. *CREATES research paper 2011-53*.
- Eberlein, E., R. Frey, & E. A. von Hammerstein (2008). Advanced credit portfolio modeling and cdo pricing. In *Mathematics – Key Technology for the Future*, edited by W. Jäger and H.-J. Krebs, Springer, 253–280.
- Epps, T. W. (1979). Comovements in stock prices in the very short run. *Journal of the American Statistical Association* 74, 291–298.
- Gloter, A. & J. Jacod (2001). Diffusions with measurement errors 1 and 2. *ESAIM Probability and Statistics* 5, 225–242.
- Harrison, M. & D. Kreps (1979). Martingales and arbitrage in multiperiod securities markets. *Journal of Economic Theory* 20, 381–408.
- Harrison, M. & S. Pliska (1981). Martingales and stochastic integrals in the theory of continuous trading. *Stochastic Processes and Their Applications* 11, 215–260.
- Hayashi, T. & N. Yoshida (2005). On covariance estimation of non-synchronously observed diffusion processes. *Bernoulli* 11(2), 359–379.
- Hayashi, T. & N. Yoshida (2011). Nonsynchronous covariation process and limit theorems. *Stochastic Processes and their Applications* 121, 2416–2454.
- 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.
- Isserlis, L. (1918). On a formula for the product-moment coefficient of any order of a normal frequency distribution in any number of variables. *Biometrika* 12, 134–139.
- Jacod, J. (1997). On continuous conditional gaussian martingales and stable convergence in law. *Séminaire de Probabilités, Strasbourg, tome 31*, 232–246.
- Jacod, J. (2012). Statistics and high frequency data. *Proceedings of the 7th Séminaire Européen de Statistique, La Manga, 2007: Statistical methods for stochastic differential equations*, edited by M. Kessler, A. Lindner and M. Sørensen.
- Jacod, J., Y. Li, P. A. Mykland, M. Podolskij, & M. Vetter (2009). Microstructure noise in the continuous case: the pre-averaging approach. *Stochastic Processes and their Applications* 119, 2803–2831.
- Jacod, J. & P. Protter (1998). Asymptotic error distributions for the euler method for stochastic differential equations. *The Annals of Probability* 26, 267–307.
- Mykland, P. & L. Zhang (2012). The econometrics of high frequency data. *Proceedings of the 7th Séminaire Européen de Statistique, La Manga, 2007: Statistical methods for stochastic differential equations*, edited by M. Kessler, A. Lindner and M. Sørensen.
- Mykland, P. A. (2012). A gaussian calculus for inference from high frequency data. *Annals of Finance* 8, 235–258.

- Mykland, P. A., N. Shephard, & K. Sheppard (2012). Efficient and feasible inference for the components of financial variation using blocked multipower variation. Technical Report, University of Oxford.
- Mykland, P. A. & L. Zhang (2009). Inference for continuous semimartingales observed at high frequency. *Econometrica* 77, 1403–1455.
- Park, S. & O. Linton (2012). Estimating the quadratic covariation matrix for an asynchronously observed continuous time signal masked by additive noise. *FMG Discussion Papers* 703.
- Rényi, A. (1963). On stable sequences of events. *Sankhya: The Indian Journal of Statistics, Series A* 25(3), 293–302.
- Rootzén, H. (1980). Limit distributions for the error in approximations of stochastic integrals. *Annals of Probability* 8, 241–251.
- Ross, S. M. (1976). The arbitrage theory of capital asset pricing. *Journal of Economic Theory* 13, 341–360.
- Xiu, D. (2010). Quasi-maximum likelihood estimation of volatility with high frequency data. *Journal of Econometrics* 159, 235–250.
- Zhang, L. (2001). *From Martingales to ANOVA: Implied and Realized Volatility*. Ph. D. thesis, The University of Chicago, Department of Statistics.
- 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. *Journal of Econometrics* 160, 33–47.
- Zhang, L., P. A. Mykland, & Y. Aït-Sahalia (2005). A tale of two time scales: Determining integrated volatility with noisy high-frequency data. *Journal of the American Statistical Association* 100(472), 1394–1411.

APPENDIX: PROOFS

A Proofs

A.1 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 $\Delta_i W = W_{iT/n} - W_{(i-1)T/n}$, $i = 1, \dots, n$, for the Brownian motion W driving the SDE with solution X in (1) and $\Delta_i \sigma = \sigma_{iT/n} - \sigma_{(i-1)T/n}$. Consider some norm $\|\cdot\|$, e. g. the euclidean norm, on \mathbb{R}^p . Suppose Assumption 1 holds. By several applications of the Burkholder-Davis-Gundy and Hölder inequality one can obtain the following estimates:

$$\mathbb{E} [\|\Delta_i X\|^2 + \|\Delta_i W\|^2 | \mathcal{F}_{(i-1)/n}] \leq Cn^{-1}, \quad \mathbb{E} [\|\Delta_i \sigma\|^2 | \mathcal{F}_{(i-1)/n}] \leq Cn^{-1}, \quad (\text{A.1a})$$

$$\mathbb{E} [\|\Delta_i X - \Delta_i W\|^2 | \mathcal{F}_{(i-1)/n}] \leq Cn^{-2}, \quad (\text{A.1b})$$

for equidistant observation schemes. For general synchronous sampling (A.1a) and (A.1b) remain valid when replacing n by δ_n^{-1} with $\delta_n = \sup_i (t_i - t_{i-1})$. The estimates (A.1a) and (A.1b) are proven in Jacod (2012), among others. They are used repeatedly in the analysis below. We will as well make use of the direct extensions for increments with respect to subsampled time lags and the componentwise versions of (A.1a) and (A.1b).

Recall (6), which is a special case of the result by Isserlis (1918) and used implicitly throughout the proofs below. For notational convenience we write again $a_n \sim^p b_n$ if $a_n = \mathcal{O}_p(b_n)$ and $b_n = \mathcal{O}_p(a_n)$.

A.2 Proof of Proposition 2.1

Applying (6) and (A.1b), by the Itô isometry we deduce for the random conditional covariance

$$\begin{aligned} \text{Cov}_\Sigma \left(\sum_{i=1}^n \Delta_i X^{(1)} \Delta_i X^{(2)}, \sum_{i=1}^n \Delta_i X^{(3)} \Delta_i X^{(4)} \right) &\sim^p \sum_{i=1}^n \text{Cov}_{i-1} \left(\Delta_i X^{(1)} \Delta_i X^{(2)}, \Delta_i X^{(3)} \Delta_i X^{(4)} \right) \\ &\sim^p \sum_{i=1}^n \left(\sigma_{\frac{(i-1)T}{n}}^{(13)} \sigma_{\frac{(i-1)T}{n}}^{(24)} + \sigma_{\frac{(i-1)T}{n}}^{(14)} \sigma_{\frac{(i-1)T}{n}}^{(23)} \right) \frac{T^2}{n^2}, \end{aligned}$$

where we write $\text{Cov}_{i-1}(\cdot)$ for $\text{Cov}(\cdot | \mathcal{F}_{(i-1)T/n})$ and $\text{Cov}_\Sigma(\cdot)$ for the random covariance dependent on Σ . By convergence of the Riemann sum to the integral, this yields

$$n \text{Cov}_\Sigma \left(\sum_{i=1}^n \Delta_i X^{(1)} \Delta_i X^{(2)}, \sum_{i=1}^n \Delta_i X^{(3)} \Delta_i X^{(4)} \right) \xrightarrow{p} T \int_0^T \left(\sigma_s^{(13)} \sigma_s^{(24)} + \sigma_s^{(14)} \sigma_s^{(23)} \right) ds.$$

□

A.3 Proof of Theorem 3.1

For the proof that

$$\left(\widehat{[X^{(1)}, X^{(2)}]}_T^{(multi)} - \widehat{[X^{(1)}, X^{(2)}]}_T^{(kernel)} \right) = -4\eta_{12} + \mathcal{O}_p(n^{-1/4})$$

if $\mathfrak{R}'' = h$ in (10), it suffices to focus on the first-order term of the weights. Transforming (9) yields

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

where we suppress additional component superscripts on M_n . The term R_n induced by end-effects

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

has an expectation by noise:

$$\eta_{12} \sum_{i=1}^{M_n} \alpha_i \left(\sum_{j=1}^{i-1} \frac{i-j}{i} - \sum_{j=2}^{i-1} \frac{i-j}{i} + \sum_{j=n-i+1}^{n-1} \frac{i-n+j}{i} - \sum_{j=n-i+1}^{n-2} \frac{i-n+j}{i} \right) = 4\eta_{12}.$$

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

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

by partial integration under the restrictions made on \mathfrak{K} . This yields the form (11) of the transformed kernel estimator and our claim. That the integral approximation does not harm the above equality up to the $\mathcal{O}(n^{-1/4})$ -term, can be seen by the estimate

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

with generic constant C , $i \geq l$, for the Lipschitz function

$$f(x) = \mathfrak{R}''(x) \left(x - \frac{l}{M_n} \right)$$

on the compact support $[0, 1]$, where Lipschitz continuity is ensured by the preconditioned continuous differentiability.

A.4 Proof of Proposition 3.1

Decompose the multi-scale estimator (9) for $[X^{(k)}, X^{(l)}]_T, (k, l) \in \{1, 2, 3, 4\}^2$

$$\begin{aligned} \sum_{i=1}^{M_n^{(kl)}} \frac{\alpha_i}{i} \sum_{j=i}^n \Delta_j^i Y^{(l)} \Delta_j^i Y^{(k)} &= \sum_{i=1}^{M_n^{(kl)}} \frac{\alpha_i}{i} \sum_{j=i}^n \Delta_j^i X^{(l)} \Delta_j^i X^{(k)} + \sum_{i=1}^{M_n^{(kl)}} \frac{\alpha_i}{i} \sum_{j=i}^n \Delta_j^i \epsilon^{(l)} \Delta_j^i \epsilon^{(k)} \\ &\quad + \sum_{i=1}^{M_n^{(kl)}} \frac{\alpha_i}{i} \sum_{j=i}^n \Delta_j^i X^{(l)} \Delta_j^i \epsilon^{(k)} + \sum_{i=1}^{M_n^{(kl)}} \frac{\alpha_i}{i} \sum_{j=i}^n \Delta_j^i \epsilon^{(l)} \Delta_j^i X^{(k)}, \end{aligned}$$

with $\Delta_j^i \epsilon^{(l)}$ analogously defined as for $X^{(l)}$, in a signal part, a noise part and cross terms which are uncorrelated for each component. We analyze the covariances which stem from covariances of the signal, noise parts and cross terms consecutively. The elements of the proof of a stable central limit theorem is similar as in Zhang (2006) and Bibinger (2012), respectively, and hence we restrict ourselves to the evaluation of the general covariance including possibly different frequencies and weights for equidistant sampling.

We can write the signal part of the addends of $[X^{(1)}, X^{(2)}]_T$ in the way

$$\begin{aligned} \frac{1}{i_{12}} \sum_{j=i_{12}}^n \Delta_j^{i_{12}} X^{(1)} \Delta_j^{i_{12}} X^{(2)} &= \sum_{j=1}^n \Delta_j X^{(1)} \Delta_j X^{(2)} + \sum_{l=1}^n \Delta_l X^{(1)} \sum_{j=1}^{i_{12} \wedge l} \left(1 - \frac{j}{i_{12}}\right) \Delta_{l-j} X^{(2)} \\ &\quad + \sum_{l=1}^n \Delta_l X^{(2)} \sum_{j=1}^{i_{12} \wedge l} \left(1 - \frac{j}{i_{12}}\right) \Delta_{l-j} X^{(1)}. \end{aligned}$$

The first addend is the usual realized covariance and will contribute only asymptotically negligible covariance terms in the noisy setting with slower convergence rate. We derive that

$$\text{Cov}_\Sigma \left(\frac{1}{i_{12}} \sum_{j=i_{12}}^n \Delta_j^{i_{12}} X^{(1)} \Delta_j^{i_{12}} X^{(2)}, \frac{1}{i_{34}} \sum_{j=i_{34}}^n \Delta_j^{i_{34}} X^{(3)} \Delta_j^{i_{34}} X^{(4)} \right) \sim^p \Gamma_{13}^n + \Gamma_{14}^n + \Gamma_{23}^n + \Gamma_{24}^n,$$

with conditional covariance terms

$$\begin{aligned} \Gamma_{13}^n &= \sum_{l=1}^n \mathbb{E} \left[\Delta_l X^{(1)} \Delta_l X^{(3)} \right] \sum_{j=1}^{\min(i_{12}, i_{34}, l)} \left(1 - \frac{j}{i_{12}}\right) \left(1 - \frac{j}{i_{34}}\right) \Delta_{l-j} X^{(2)} \Delta_{l-j} X^{(4)} \\ &\sim^p \sum_{l=1}^n \sigma_{\frac{(l-1)T}{n}}^{(13)} \frac{1}{n} \sum_{j=1}^{\min(i_{12}, i_{34}, l)} \left(1 - \frac{j}{i_{12}}\right) \left(1 - \frac{j}{i_{34}}\right) \sigma_{\frac{(l-1)T}{n}}^{(24)} \frac{1}{n}, \end{aligned} \quad (\text{A.2})$$

and analogously for the other addends. The smoothness of Σ ensured by Assumption 1 giving the bound in (A.1a) suffices that we can approximately take $\sigma_{\frac{(l-1)T}{n}}^{(24)}$ in all addends of the second sum. Hence, for the equidistant case the analysis boils down to simply evaluating the deterministic inner sum:

$$\sum_{j=1}^m \left(1 - \frac{j}{i_{12}}\right) \left(1 - \frac{j}{i_{34}}\right) = \frac{m}{2} - \frac{m^2}{6M} - \frac{1}{8} + \frac{1}{12M} \sim \frac{m}{6} \left(3 - \frac{m}{M}\right),$$

where $m = \min(i_{12}, i_{34})$ and $M = \max(i_{12}, i_{34})$. For $M_n = \min(M_n^{(12)}, M_n^{(34)})$ denote the limit of the following series including weights according to (10)

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

This leads to the result

$$\frac{n}{M_n} \sum_{i_{12}, i_{34}=1}^{M_n} \alpha_{i_{12}} \alpha_{i_{34}} \Gamma_{13}^n \xrightarrow{p} \mathfrak{D}^\alpha T \int_0^T \sigma_s^{(13)} \sigma_s^{(24)} ds.$$

Similar results hold for $\Gamma_{14}^n, \Gamma_{23}^n, \Gamma_{24}^n$. Thereby we conclude the signal term of (17). \mathfrak{D}^α is a constant showing up in the asymptotic discretization variance depending on the weights, where for the standard weights (16) or cubic kernel $\mathfrak{D}^\alpha = 13/70$. Eventually, we also obtain the general form of the asymptotic discretization variance of the one-dimensional estimator.

The addend induced by market microstructure noise in the multi-scale estimator (9) is

$$\begin{aligned} \sum_{i=1}^{M_n^{(12)}} \frac{\alpha_i}{i} \sum_{j=i}^n \left(\epsilon_j^{(1)} - \epsilon_{j-i}^{(1)} \right) \left(\epsilon_j^{(2)} - \epsilon_{j-i}^{(2)} \right) &= \sum_{i=1}^{M_n^{(12)}} \frac{\alpha_i}{i} \left(2 \sum_{j=1}^n \epsilon_j^{(1)} \epsilon_j^{(2)} \right. \\ &\quad \left. - \sum_{j=i}^n \left(\epsilon_j^{(1)} \epsilon_{j-i}^{(2)} + \epsilon_j^{(2)} \epsilon_{j-i}^{(1)} \right) - \sum_{j=n-i+1}^n \epsilon_j^{(1)} \epsilon_j^{(2)} - \sum_{j=0}^{i-1} \epsilon_j^{(1)} \epsilon_j^{(2)} \right). \end{aligned}$$

The last two sums lead for the non-adjusted multi-scale estimator (9) to the negative bias by noise and end-effects. We have focused on the bias and an adjusted corrected estimator in Section 3 and will concentrate on the adjusted non-biased version in the sequel. The first inner sum on the right-hand side does not depend on i and vanishes since $\sum_{i=1}^{M_n^{(12)}} \alpha_i/i = 0$. The covariances of the remaining uncorrelated addends contribute to the total covariance due to noise perturbation. Denote the constant limits

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

and rewriting

$$\sum_{i=1}^{M_n^{(12)}} \frac{\alpha_i}{i} \sum_{j=i}^n \epsilon_j^{(1)} \epsilon_{j-i}^{(2)} = \sum_{j=i}^n \sum_{i=1}^{M_n^{(12)} \wedge j} \frac{\alpha_i}{i} \epsilon_j^{(1)} \epsilon_{j-i}^{(2)}$$

and also

$$\sum_{i=1}^{M_n^{(12)}} \frac{\alpha_i}{i} \left(\sum_{j=0}^{i-1} \epsilon_j^{(1)} \epsilon_j^{(2)} + \sum_{j=n-i+1}^n \epsilon_j^{(1)} \epsilon_j^{(2)} \right) = \sum_{j=0}^{M_n^{(12)}-1} \left(\epsilon_j^{(1)} \epsilon_j^{(2)} + \epsilon_{n-j}^{(1)} \epsilon_{n-j}^{(2)} \right) \sum_{i=j+1}^{M_n^{(12)}} \frac{\alpha_i}{i},$$

we obtain with $M_n = \min(M_n^{(12)}, M_n^{(34)})$ that

$$\sqrt{\frac{M_n^3}{n}} \text{Cov} \left(\sum_{i=1}^{M_n^{(12)}} \frac{\alpha_i}{i} \sum_{j=i}^n \epsilon_j^{(1)} \epsilon_{j-i}^{(2)}, \sum_{i=1}^{M_n^{(34)}} \frac{\alpha_i}{i} \sum_{j=i}^n \epsilon_j^{(3)} \epsilon_{j-i}^{(4)} \right) \longrightarrow \mathfrak{N}_1^\alpha \eta_{13} \eta_{24}, \quad (\text{A.3})$$

and that

$$\begin{aligned} \sqrt{M_n} \text{Cov} \left(\sum_{i=1}^{M_n^{(12)}} \frac{\alpha_i}{i} \left(\sum_{j=n-i+1}^n \epsilon_j^{(1)} \epsilon_j^{(2)} + \sum_{j=0}^{i-1} \epsilon_j^{(1)} \epsilon_j^{(2)} \right), \sum_{i=1}^{M_n^{(34)}} \frac{\alpha_i}{i} \left(\sum_{j=n-i+1}^n \epsilon_j^{(3)} \epsilon_j^{(4)} + \sum_{j=0}^{i-1} \epsilon_j^{(3)} \epsilon_j^{(4)} \right) \right) \\ \longrightarrow \mathfrak{N}_2^\alpha (\eta_{13} \eta_{24} + \eta_{14} \eta_{23}). \end{aligned} \quad (\text{A.4})$$

The analogous implications for all addends contributing to the total covariance by noise lead to the noise parts in (17). For (16), we have $\mathfrak{N}_2^\alpha = 6/5$ and $\mathfrak{N}_1^\alpha = 12$, which gives the minimum of the variance due to noise (see Zhang (2006)). Note that for the weights from (10), the respective frequencies $M_n^{(kl)}, 1 \leq k \leq 4, 1 \leq l \leq 4$, are inserted, but we leave out further indices for a better readability.

Finally, consider the cross terms of (9). They can be decomposed in addends of the form

$$\sum_{i=1}^{M_n^{(kl)}} \frac{\alpha_i}{i} \sum_{j=0}^n \zeta_{i,j}^{(k)} \epsilon_j^{(l)} \quad (\text{A.5})$$

for components (k, l) , where

$$\zeta_{i,j}^{(k)} = \begin{cases} -\Delta_{i-j}^{(k)} X^{(k)} & , 0 \leq j \leq (i-1) \\ \Delta_j^{(k)} X^{(k)} - \Delta_{j+i}^{(k)} X^{(k)} & , i \leq j \leq (n-i) \\ \Delta_j^{(k)} X^{(k)} & , n-i+1 \leq j \leq n \end{cases}.$$

The asymptotic covariance coming from cross terms, by the martingale structure of the above term can be deduced by the limit of the sum over j covariances conditionally on $\mathcal{F}_T \vee \sigma(\epsilon_l, l \leq j-1)$. We thus have

$$\mathbb{Cov}_\Sigma \left(\sum_{i=1}^{M_n^{(12)}} \frac{\alpha_i}{i} \sum_{j=0}^n \zeta_{i,j}^{(1)} \epsilon_j^{(2)}, \sum_{i=1}^{M_n^{(34)}} \frac{\alpha_i}{i} \sum_{j=0}^n \zeta_{i,j}^{(3)} \epsilon_j^{(4)} \right) = \sum_{j=0}^n \mathbb{E} \left[\epsilon_j^{(1)} \epsilon_j^{(3)} \right] \sum_{i=1}^{M_n^{(12)}} \frac{\alpha_i}{i} \zeta_{i,j}^{(2)} \sum_{i=1}^{M_n^{(34)}} \frac{\alpha_i}{i} \zeta_{i,j}^{(4)}.$$

Now, similarly as in Zhang (2006) and Bibinger (2012), if we assume without loss of generality $1 \leq i_{12} \leq i_{34} \leq M_n = \min(M_n^{(12)}, M_n^{(34)})$, it holds that

$$\begin{aligned} \zeta_{i_{12},j}^{(k)} \zeta_{i_{34},j}^{(l)} &= \left(\sum_{r=j-i_{12}}^{j-1} \Delta_r X^{(k)} - \sum_{r=j}^{i_{12}+j-1} \Delta_r X^{(k)} \right) \left(\sum_{r=j-i_{34}}^{j-1} \Delta_r X^{(l)} - \sum_{r=j}^{i_{34}+j-1} \Delta_r X^{(l)} \right) \\ &= \zeta_{i_{12}}^{(k)} \zeta_{i_{12}}^{(l)} + \zeta_{i_{34}}^{(k)} \left(\sum_{r=j-i_{34}}^{j-i_{12}-1} \Delta_r X^{(l)} - \sum_{r=j+i_{12}}^{i_{34}+j-1} \Delta_r X^{(l)} \right) \quad k=1,2; l=3,4. \end{aligned}$$

The conditional covariances from these addends yield for $u, k, l, v \in \{1, 2, 3, 4\}$

$$\begin{aligned} M_n \mathbb{Cov}_\Sigma \left(\sum_{i=1}^{M_n^{(12)}} \frac{\alpha_i}{i} \sum_{j=0}^n \zeta_{i,j}^{(k)} \epsilon_j^{(u)}, \sum_{i=1}^{M_n^{(34)}} \frac{\alpha_i}{i} \sum_{j=0}^n \zeta_{i,j}^{(l)} \epsilon_j^{(v)} \right) \\ \sim^p 2M_n \eta_{uv} \sum_{i=1}^{M_n} \sum_{r=1}^{M_n} \frac{\alpha_i \alpha_r}{i r} (i \wedge r) \left(\frac{1}{(i \wedge r)} \sum_{j=(i \wedge r)}^n \Delta_j^{(i \wedge r)} X^{(k)} \Delta_j^{(i \wedge r)} X^{(l)} \right) \\ \xrightarrow{p} \mathfrak{M}^\alpha \eta_{uv} \left[X^{(k)}, X^{(l)} \right]_T. \end{aligned}$$

For the specific weights (16), the constant takes the value $\mathfrak{M}^\alpha = 6/5$. This leads to the total covariance by cross terms completing the proof and we conclude Proposition 3.1. \square

A.5 Proof of Proposition 4.1

Recall the illustration (20) of the Hayashi-Yoshida estimator with refresh times and interpolations. We first consider the term $D_T^{12} D_T^{34}$ from the product of the two synchronous-type approximations. Only overlapping increments contribute to the asymptotic covariance:

$$\begin{aligned} \mathbb{E} [D_T^{12} D_T^{34}] &= \sum_{i,j} \mathbb{1}_{\{\max(T_{i-1}^{12}, T_{j-1}^{34}) < \min(T_i^{12}, T_j^{34})\}} \\ &\quad \times \mathbb{E} \left[\left(X_{T_i^{12}}^{(1)} - X_{T_{i-1}^{12}}^{(1)} \right) \left(X_{T_i^{12}}^{(2)} - X_{T_{i-1}^{12}}^{(2)} \right) \left(X_{T_j^{34}}^{(3)} - X_{T_{j-1}^{34}}^{(3)} \right) \left(X_{T_j^{34}}^{(4)} - X_{T_{j-1}^{34}}^{(4)} \right) \right] \end{aligned}$$

and thus

$$\mathbb{Cov}(D_T^{12}, D_T^{34}) = \sum_{i=1}^N \mathbb{Cov} \left(\left(X_{S_i}^{(1)} - X_{S_{i-1}}^{(1)} \right) \left(X_{S_i}^{(2)} - X_{S_{i-1}}^{(2)} \right), \left(X_{S_i}^{(3)} - X_{S_{i-1}}^{(3)} \right) \left(X_{S_i}^{(4)} - X_{S_{i-1}}^{(4)} \right) \right).$$

Only on intervals where increments of all four processes coincide, the expectation of the product with four factors does not equal the product of expectations. For that reason the refresh times $S_i, i = 0, \dots, N$, of all

processes give the crucial synchronous approximation here. The other remaining addend is the covariance of the terms due to non-synchronicity for each pair $A_T^{12} A_T^{34}$. We have that

$$\begin{aligned} \mathbb{E} [A_T^{12} A_T^{34}] &= \sum_{i,j} \mathbb{1}_{\{\max(t_1^-(T_{i-1}^{12}), t_2^-(T_{i-1}^{12}), t_3^-(T_{j-1}^{34}), t_4^-(T_{j-1}^{34})) < \min(t_1^+(T_i^{12}), t_2^+(T_i^{12}), t_3^+(T_j^{34}), t_4^+(T_j^{34}))\}} \\ &\times \text{Cov} \left(\left(X_{T_i^{12}}^{(1),+} - X_{T_i^{12}}^{(1)} + X_{T_{i-1}^{12}}^{(1)} - X_{T_{i-1}^{12}}^{(1),-} \right) \left(X_{T_i^{12}}^{(2),+} - X_{T_i^{12}}^{(2)} + X_{T_{i-1}^{12}}^{(2)} - X_{T_{i-1}^{12}}^{(2),-} \right), \right. \\ &\left. \left(X_{T_j^{34}}^{(3),+} - X_{T_j^{34}}^{(3)} + X_{T_{j-1}^{34}}^{(3)} - X_{T_{j-1}^{34}}^{(3),-} \right) \left(X_{T_j^{34}}^{(4),+} - X_{T_j^{34}}^{(4)} + X_{T_{j-1}^{34}}^{(4)} - X_{T_{j-1}^{34}}^{(4),-} \right) \right) \end{aligned}$$

As explained in Section 3, we can arrange the coinciding interpolation terms using the structuring generated by the S_i s. Consider now the sequence of covariances conditionally on $\mathcal{F}_{S_{i-1}}$, $i = 1, \dots, N$, and its sum. For the first term with the synchronous approximation this is

$$\begin{aligned} &\sum_{i=1}^N \text{Cov} \left(\left(X_{S_i}^{(1)} - X_{S_{i-1}}^{(1)} \right) \left(X_{S_i}^{(2)} - X_{S_{i-1}}^{(2)} \right), \left(X_{S_i}^{(3)} - X_{S_{i-1}}^{(3)} \right) \left(X_{S_i}^{(4)} - X_{S_{i-1}}^{(4)} \right) \middle| \mathcal{F}_{S_{i-1}} \right) \\ &\sim^p \sum_{i=1}^N \left(\sigma_{S_{i-1}}^{(13)} \sigma_{S_{i-1}}^{(24)} + \sigma_{S_{i-1}}^{(14)} \sigma_{S_{i-1}}^{(23)} \right) (S_i - S_{i-1})^2 \\ &= \sum_{i=1}^N \left(\sigma_{S_{i-1}}^{(13)} \sigma_{S_{i-1}}^{(24)} + \sigma_{S_{i-1}}^{(14)} \sigma_{S_{i-1}}^{(23)} \right) \frac{(G(S_i) - G(S_{i-1}))}{S_i - S_{i-1}} (S_i - S_{i-1}) \end{aligned}$$

and analogously for $\text{Cov}(A_T^{12}, A_T^{34})$ with all interpolations. The difference quotients will converge on Assumption 3 and the Riemann sums to the corresponding integrals. With all functions from Definition 1 and Assumption 3, we obtain

$$\begin{aligned} \mathbb{A}\text{COV} &= p - \lim_{N \rightarrow \infty} \left(\sum_{i=1}^N \left(\sigma_{S_{i-1}}^{(13)} \sigma_{S_{i-1}}^{(24)} + \sigma_{S_{i-1}}^{(14)} \sigma_{S_{i-1}}^{(23)} \right) \frac{(G(S_i) - G(S_{i-1}))}{S_i - S_{i-1}} (S_i - S_{i-1}) \right. \\ &+ \sum_{i=1}^N \sigma_{S_{i-1}}^{(13)} \sigma_{S_{i-1}}^{(24)} \frac{(F_{24}^{13}(S_i) - F_{24}^{13}(S_{i-1})) + (H_{24}^{13}(S_i) - H_{24}^{13}(S_{i-1})) + (I_{24}^{13}(S_i) - I_{24}^{13}(S_{i-1}))}{S_i - S_{i-1}} (S_i - S_{i-1}) \\ &\left. + \sum_{i=1}^N \sigma_{S_{i-1}}^{(14)} \sigma_{S_{i-1}}^{(23)} \frac{(F_{14}^{23}(S_i) - F_{14}^{23}(S_{i-1})) + (H_{14}^{23}(S_i) - H_{14}^{23}(S_{i-1})) + (I_{14}^{23}(S_i) - I_{14}^{23}(S_{i-1}))}{S_i - S_{i-1}} (S_i - S_{i-1}) \right) \end{aligned}$$

and this limit in probability of the above sum equals (24). \square

A.6 Proof of Proposition 5.1

Consider four possibly different observation schemes, but also allowing for subsets of synchronous sampling times. We use the notation from Section 3 for interpolations and refresh times. For the error due to noise (A.3) becomes

$$\begin{aligned} &\frac{M_N^3}{N} \text{Cov} \left(\sum_{j=0}^{N_{12}} \sum_{i=1}^{M_N^{(12)} \wedge (j+1)} \frac{\alpha_i}{i} \left(\epsilon_{t_1^+(T_j^{12})}^{(1)} \epsilon_{t_2^-(T_{j-i}^{12})}^{(2)} + \epsilon_{t_2^+(T_j^{12})}^{(2)} \epsilon_{t_1^-(T_{j-i}^{12})}^{(1)} \right) \right. \\ &\quad \left. , \sum_{j=0}^{N_{34}} \sum_{i=1}^{M_N^{(34)} \wedge (j+1)} \frac{\alpha_i}{i} \left(\epsilon_{t_3^+(T_j^{34})}^{(3)} \epsilon_{t_4^-(T_{j-i}^{34})}^{(4)} + \epsilon_{t_4^+(T_j^{34})}^{(4)} \epsilon_{t_3^-(T_{j-i}^{34})}^{(3)} \right) \right) \\ &\longrightarrow (\eta_{13} \eta_{24} \mathfrak{S}_{13}^{24} \mathfrak{C}_{13}^{24} + \eta_{14} \eta_{23} \mathfrak{S}_{14}^{23} \mathfrak{C}_{14}^{23}) \end{aligned}$$

on the assumption that the limits in Definition 3 exist with $\mathfrak{C}_{13}^{24} = \lim_{N \rightarrow \infty} M_N^3 \sum_{i,k=1}^{M_N} \frac{\alpha_i \alpha_k}{ik}$ over all (i, k) for which the corresponding indicator functions in the Definition 3 equal one. It holds that $\max(\mathfrak{C}_{13}^{24}, \mathfrak{C}_{14}^{23})$

$\leq \mathfrak{N}_1^\alpha$ by the Cauchy-Schwarz inequality and that each observation point can only equal one of the other scheme. The generalization of (A.4) incorporating general observation schemes is

$$\begin{aligned} & \sqrt{M_n} \text{Cov} \left(\sum_{j=0}^{M_N-1} \sum_{i=j+1}^{M_N} \frac{\alpha_i}{i} \epsilon_{t_1^+(T_j^{12})}^{(1)} \epsilon_{t_2^+(T_j^{12})}^{(2)} + \epsilon_{t_1^-(T_{N_{12}-j}^{12})}^{(1)} \epsilon_{t_2^-(T_{N_{12}-j}^{12})}^{(2)}, \right. \\ & \quad \left. \sum_{j=0}^{M_N-1} \sum_{i=j+1}^{M_N} \frac{\alpha_i}{i} \epsilon_{t_3^+(T_j^{34})}^{(1)} \epsilon_{t_4^+(T_j^{34})}^{(2)} + \epsilon_{t_3^-(T_{N_{34}-j}^{34})}^{(1)} \epsilon_{t_4^-(T_{N_{34}-j}^{34})}^{(2)} \right) \\ & \longrightarrow (\tilde{\mathfrak{S}}_{13}^{24} \tilde{\mathfrak{C}}_{13}^{24} \eta_{13} \eta_{24} + \tilde{\mathfrak{S}}_{14}^{23} \tilde{\mathfrak{C}}_{14}^{23} \eta_{14} \eta_{23}). \end{aligned}$$

Again we have $\max(\tilde{\mathfrak{C}}_{13}^{24}, \tilde{\mathfrak{C}}_{14}^{23}) \leq \mathfrak{N}_2^\alpha$, i. e. constants smaller or equal the ones in the synchronous setting. For the analysis of cross terms we can build on the findings from above and the rewriting (A.5). Directly ignoring interpolations we set

$$\tilde{\zeta}_{i,j}^{(k)} = \begin{cases} -\Delta_{i-j}^i X^{(k)} & , 0 \leq j \leq (i-1) \\ \Delta_j^i X^{(k)} - \Delta_{j+i}^i X^{(k)} & , i \leq j \leq (n-i) \\ \Delta_j^i X^{(k)} & , n-i+1 \leq j \leq n \end{cases}.$$

with $\Delta_j^i X^{(k)} = (X_{T_j^{12}}^{(k)} - X_{T_{j-i}^{12}}^{(k)})$ for $k = 1, 2$, and $\Delta_j^i X^{(k)} = (X_{T_j^{34}}^{(k)} - X_{T_{j-i}^{34}}^{(k)})$ for $k = 3, 4$, here. The conditional covariance of the cross terms ignoring asymptotically negligible interpolation steps yields

$$\sum_{j=0}^{n_1} \sum_{k=0}^{n_3} \mathbb{E} \left[\epsilon_{t_j^{(1)}}^{(1)} \epsilon_{t_k^{(3)}}^{(3)} \right] \mathbb{1}_{\{t_j^{(1)}=t_k^{(3)}\}} \sum_{i=1}^{M_N^{(12)}} \frac{\alpha_i}{i} \tilde{\zeta}_{i,j}^{(2)} \sum_{i=1}^{M_N^{(34)}} \frac{\alpha_i}{i} \tilde{\zeta}_{i,k}^{(4)}$$

and

$$\begin{aligned} & \tilde{\zeta}_{i_{12},j}^{(2)} \tilde{\zeta}_{i_{34},k}^{(4)} = \left(\sum_{r=j-i_{12}}^{j-1} \Delta_r X^{(2)} - \sum_{r=j}^{i_{12}+j-1} \Delta_r X^{(2)} \right) \left(\sum_{r=k-i_{34}}^{k-1} \Delta_r X^{(4)} - \sum_{r=k}^{i_{34}+k-1} \Delta_r X^{(4)} \right) \\ & \sim^p \left(\sum_{r=\max(j-i_{12}, k-i_{34})}^{(j \wedge k)-1} \Delta_r X^{(2)} - \sum_{r=(j \vee k)}^{\min(i_{12}+j-1, i_{34}+k-1)} \Delta_r X^{(2)} \right) \left(\sum_{r=\max(j-i_{12}, k-i_{34})}^{(j \wedge k)-1} \Delta_r X^{(4)} - \sum_{r=(j \vee k)}^{\min(i_{12}+j-1, i_{34}+k-1)} \Delta_r X^{(4)} \right). \end{aligned}$$

Compared to the synchronous case this product includes refresh time instants and in general $j \neq k$, but if $t_j^{(1)} = t_k^{(3)}$ holds, locally $\tilde{\zeta}_{i_{12},j}^{(2)} \tilde{\zeta}_{i_{34},k}^{(4)}$ estimates $\int_{\Delta} \sigma_s^2 ds$, only that here the interval Δ is determined by (j, k) and (i_{12}, i_{34}) . Taking the sum for the multi-scale estimates, we end up with the same approximation as above, but including the limiting function of S'_{13} to describe the proportion of synchronous times

$$\begin{aligned} & \text{Cov}_{\Sigma} \left(\sum_{i=1}^{M_N^{(12)}} \frac{\alpha_i}{i} \sum_{j=0}^{n_3} \tilde{\zeta}_{i,j}^{(2)} \epsilon_j^{(1)}, \sum_{i=1}^{M_N^{(34)}} \frac{\alpha_i}{i} \sum_{j=0}^{n_4} \tilde{\zeta}_{i,j}^{(4)} \epsilon_j^{(3)} \right) \\ & \sim^p 2M_N \sum_{i=1}^{M_N} \sum_{r=1}^{M_N} \frac{\alpha_i \alpha_r}{ir} (i \wedge r) \left(\frac{1}{(i \wedge r)} \sum_{j,k=(i \wedge r)}^n \Delta_j^{(i \wedge r)} X^{(2)} \Delta_k^{(i \wedge r)} X^{(4)} \eta_{13} \mathbb{1}_{\{t_j^{(1)}=t_k^{(3)}\}} \right) \\ & \xrightarrow{p} \frac{12}{5} \eta_{13} \int_0^T S'_{13}(t) \sigma_t^2 dt. \end{aligned}$$

The other addends are treated analogously.

It remains to show (36) for the discretization term, in most situations the only non-vanishing term in the asymptotic covariance. Proposition 5.1 and Corollary 5.2 state that the asymptotic covariance equals the

one of the experiment in which we observe all processes synchronously at times $S_i, 0 \leq i \leq N$. As a first step of the proof it can be shown that the discretization error induced by interpolations to refresh times $T_i^{12}, 0 \leq i \leq N_{12}$, and $T_i^{34}, 0 \leq i \leq N_{34}$, for each pair is asymptotically negligible. This is proved analogously as Proposition A.10 in Bibinger (2012) and we omit it here. The second step would be to show that the covariances between an approximation error where refresh times of one pair are interpolated to the refresh times $S_i, i = 0, \dots, N$, of all four processes and the other multi-scale estimate tends to zero in probability. As discussed in Section 3 these terms have non-zero expectation, but do not contribute to the covariance which already gives the result. Alternatively, the conclusion can be thoroughly comprehended by adapting (A.2) to the two refresh time schemes after the first approximation step. All addends with overlapping refresh time instants contribute to the sum and by the Itô isometry and approximation (A.1b), the single addends are of the form

$$(\min(T_j^{12}, T_k^{34}) - \max(T_{j-1}^{12}, T_{k-1}^{34})) \sigma_{\max(T_{j-1}^{12}, T_{k-1}^{34})}^{13}.$$

Indeed, the interpolation terms fall out of the sum and in consequence the proof boils down to the synchronous setup after passing to the refresh times proxy. Yet, we still need to develop the severally interesting asymptotic distribution theory for irregular non-equidistant synchronous sampling. Compared to Zhang (2006) and Bibinger (2012), we forego the assertion that the observation scheme is close to equidistant such that the approximation in (A.2) is still valid with an asymptotically negligible approximation error. Generally, with the limiting function

$$D^\alpha(t) := \lim_{N \rightarrow \infty} \left(\frac{N}{M_N T} \sum_{S_r \leq t} \Delta S_r \sum_{i,k=1}^{M_N} \alpha_i \alpha_k \sum_{q=0}^{r \wedge i \wedge k} \left(1 - \frac{q}{i}\right) \left(1 - \frac{q}{k}\right) \Delta S_{r-q} \right),$$

which exists under Assumption 3 (i) and if the difference quotients also converge uniformly under Assumption 3 (ii), we obtain in place of (A.2):

$$\begin{aligned} \frac{N}{M_N} \sum_{i,k=1}^{M_N} \alpha_i \alpha_k \tilde{\Gamma}_{13}^N &\sim^p \frac{N}{M_N} \sum_{l=1}^{N \Delta S_l} \sum_{i_{12}, i_{34}=1}^{M_N} \alpha_{i_{12}} \alpha_{i_{34}} \sigma_{S_{l-1}}^{(13)} \sigma_{S_{l-1}}^{(24)} \sum_{q=0}^{l \wedge i_{12} \wedge i_{34}} \left(1 - \frac{q}{i_{12}}\right) \left(1 - \frac{q}{i_{34}}\right) \Delta S_{l-q} \\ &\xrightarrow{p} T \int_0^T \sigma_s^{(13)} \sigma_s^{(24)} (D^\alpha)'(s) ds. \end{aligned}$$

The existence of

$$\frac{N}{(i \wedge k) T} \sum_{S_r \leq t} \Delta S_r \sum_{q=0}^{r \wedge i \wedge k} \left(1 - \frac{q}{i}\right) \left(1 - \frac{q}{k}\right) \Delta S_{r-q}$$

is ensured on Assumption 3 by dominated convergence and the convergence of $M_N^{-1} \sum \alpha_i \alpha_k (i \wedge k)$ on the conditions for the weights (10). For the weights (16) and the cubic kernel $M_N^{-1} \sum \alpha_i \alpha_k (i \wedge k) \sim 38/70$. The above finding is valid in the same fashion in the one-dimensional setting where

$$D^\alpha(t) := \lim_{n \rightarrow \infty} \left(\frac{n}{M_n T} \sum_{t_r \leq t} \Delta t_r \sum_{i,k=1}^{M_n} \alpha_i \alpha_k \sum_{q=0}^{r \wedge i \wedge k} \left(1 - \frac{q}{i}\right) \left(1 - \frac{q}{k}\right) \Delta t_{r-q} \right).$$

The explicit formula for the weighted local sampling autocorrelation is useful to analyze specific examples:

$$\begin{aligned} \text{wLSA}_{i,k}^r &= n \left((\Delta t_r)^2 + \left(1 - \frac{1}{i}\right) \left(1 - \frac{1}{k}\right) \Delta t_r \Delta t_{r-1} + \left(1 - \frac{2}{i}\right) \left(1 - \frac{2}{k}\right) \Delta t_r \Delta t_{r-2} + \dots \right. \\ &\quad \left. + \frac{1}{i \wedge k} \left(1 - \frac{i \wedge k}{i \vee k}\right) \Delta t_r \Delta t_{r-i \wedge k} \right). \end{aligned} \quad (\text{A.6})$$

Since the first addend is negligible as $n \rightarrow \infty$, (A.6) has a simple nature for i.i.d. random sampling independent of Y as for the equidistant setup. Analogous reasoning for the other addends contributing to the discretization variance leads to (36) which completes the proof of Proposition 5.1. \square

A.7 Proof of Proposition 6.1

Consider the estimator $\widehat{\text{ACOV}}$ from (37) for the asymptotic covariance (4) of integrated covariances. On Assumption 1 we have

$$\begin{aligned}
\mathbb{E} [\widehat{\text{ACOV}}] &\sim \frac{n}{T} \sum_{i=1}^{n-1} \mathbb{E} \left[\int_{\frac{(i-1)T}{n}}^{\frac{iT}{n}} \sigma_s^{(1)} dW_s^{(1)} \int_{\frac{(i-1)T}{n}}^{\frac{iT}{n}} \sigma_s^{(3)} dW_s^{(3)} + \int_{\frac{iT}{n}}^{\frac{(i+1)T}{n}} \sigma_s^{(2)} dW_s^{(2)} \int_{\frac{iT}{n}}^{\frac{(i+1)T}{n}} \sigma_s^{(4)} dW_s^{(4)} \right] \\
&\quad + \frac{n}{T} \sum_{i=1}^{n-1} \mathbb{E} \left[\int_{\frac{(i-1)T}{n}}^{\frac{iT}{n}} \sigma_s^{(2)} dW_s^{(2)} \int_{\frac{(i-1)T}{n}}^{\frac{iT}{n}} \sigma_s^{(3)} dW_s^{(3)} + \int_{\frac{iT}{n}}^{\frac{(i+1)T}{n}} \sigma_s^{(1)} dW_s^{(1)} \int_{\frac{iT}{n}}^{\frac{(i+1)T}{n}} \sigma_s^{(4)} dW_s^{(4)} \right] \\
&\sim \frac{n}{T} \sum_{i=1}^{n-1} \int_{\frac{(i-1)T}{n}}^{\frac{iT}{n}} \sigma_s^{(13)} ds \int_{\frac{iT}{n}}^{\frac{(i+1)T}{n}} \sigma_s^{(24)} ds + \int_{\frac{(i-1)T}{n}}^{\frac{iT}{n}} \sigma_s^{(23)} ds \int_{\frac{iT}{n}}^{\frac{(i+1)T}{n}} \sigma_s^{(14)} ds \\
&\sim \frac{n}{T} \sum_{i=1}^{n-1} \left(\sigma_{\frac{(i-1)T}{n}}^{(13)} \sigma_{\frac{iT}{n}}^{(24)} + \sigma_{\frac{(i-1)T}{n}}^{(23)} \sigma_{\frac{iT}{n}}^{(14)} \right) \frac{T^2}{n^2} \\
&\sim \sum_{i=1}^{n-1} \left(\sigma_{\frac{(i-1)T}{n}}^{(23)} \sigma_{\frac{(i-1)T}{n}}^{(14)} \sigma_{\frac{(i-1)T}{n}}^{(13)} \sigma_{\frac{(i-1)T}{n}}^{(24)} \right) \frac{T}{n} \rightarrow T \int_0^T \left(\sigma_s^{(13)} \sigma_s^{(24)} + \sigma_s^{(14)} \sigma_s^{(23)} \right) ds
\end{aligned}$$

with (A.1a), (A.1b) and Itô's isometry. The variance is of order n^{-1} , what can be seen e. g. by bounding the second moment by Burkholder-Davis-Gundy inequality, which implies consistency.

A histogram-wise approach as (38) proposed for the general setup is typical for these kind of methods and consistency can be proved similar as for the asymptotic variance estimator in Proposition 5.1 of Bibinger (2012).

SFB 649 Discussion Paper Series 2013

For a complete list of Discussion Papers published by the SFB 649, please visit <http://sfb649.wiwi.hu-berlin.de>.

- 001 "Functional Data Analysis of Generalized Quantile Regressions" by Mengmeng Guo, Lhan Zhou, Jianhua Z. Huang and Wolfgang Karl Härdle, January 2013.
- 002 "Statistical properties and stability of ratings in a subset of US firms" by Alexander B. Matthies, January 2013.
- 003 "Empirical Research on Corporate Credit-Ratings: A Literature Review" by Alexander B. Matthies, January 2013.
- 004 "Preference for Randomization: Empirical and Experimental Evidence" by Nadja Dwenger, Dorothea Kübler and Georg Weizsäcker, January 2013.
- 005 "Pricing Rainfall Derivatives at the CME" by Brenda López Cabrera, Martin Odening and Matthias Ritter, January 2013.
- 006 "Inference for Multi-Dimensional High-Frequency Data: Equivalence of Methods, Central Limit Theorems, and an Application to Conditional Independence Testing" by Markus Bibinger and Per A. Mykland, January 2013.

SFB 649, Spandauer Straße 1, D-10178 Berlin
<http://sfb649.wiwi.hu-berlin.de>

This research was supported by the Deutsche
Forschungsgemeinschaft through the SFB 649 "Economic Risk".

