



SCHOOL OF ECONOMICS AND MANAGEMENT
FACULTY OF SOCIAL SCIENCES
AARHUS UNIVERSITY



CREATES
Center for Research in Econometric
Analysis of Time Series

CREATES Research Paper 2009-45

Pre-averaging estimators of the ex-post covariance matrix

Kim Christensen, Silja Kinnebrock and Mark Podolskij

School of Economics and Management
Aarhus University
Bartholins Allé 10, Building 1322, DK-8000 Aarhus C
Denmark

Pre-averaging estimators of the ex-post covariance matrix in noisy diffusion models with non-synchronous data

Kim Christensen

Silja Kinnebrock

Mark Podolskij*

September, 2009

Abstract

In this paper, we show how simple pre-averaging can be applied to measure the ex-post covariance of high-frequency financial time series under market microstructure noise and non-synchronous trading. A modulated realised covariance based on pre-averaged data is proposed and studied in this setting, and we provide complete large sample asymptotics for this new estimator, including feasible central limit theorems for standard methods such as covariance, regression, and correlation analysis. We discuss several versions of the modulated realised covariance, which can be designed to possess an optimal rate of convergence or to guarantee positive semi-definite covariance matrix estimates. We also derive a pre-averaged version of the Hayashi-Yoshida estimator that can be applied directly to the noisy and non-synchronous data without any prior alignment of prices. An empirical study illustrates how high-frequency covariances, regression coefficients, and correlations change through time.

Keywords: Central limit theorem; Diffusion models; High-frequency data; Market microstructure noise; Non-synchronous trading; Pre-averaging; Realised covariance.

JEL Classification: C10; C22; C80.

*Christensen is with CREATES, University of Aarhus, Denmark. E-mail: kchristensen@creates.au.dk. Kinnebrock is with Oxford-Man Institute of Quantitative Finance, Oxford University, UK. E-mail: silja.kinnebrock@merton.oxon.org. Podolskij is with ETH Zürich, Switzerland and also affiliated with CREATES, University of Aarhus, Denmark. E-mail: mark.podolskij@math.ethz.ch. Christensen and Podolskij were supported by CREATES, which is funded by the Danish National Research Foundation. Kinnebrock received financial support from the Rhodes Trust and Oxford-Man Institute of Quantitative Finance. We thank Neil Shephard and Matthias Vetter for comments on a previous draft of this paper.

1 Introduction

The theory of financial economics is often cast in multivariate settings, where the covariance structure of assets plays a key role to the solution of fundamental economic problems, such as optimal asset allocation and risk management. In recent years, a broader access to financial high-frequency data has improved our ability to accurately estimate and draw inference about financial covariation. The underlying idea is to use quadratic covariation, which we can estimate using realised covariance, as an ex-post measure, whose increments can be studied to learn about the properties of the true asset return covariation (e.g. Andersen, Bollerslev, Diebold, and Labys, 2003; Barndorff-Nielsen and Shephard, 2004).

In practice, implementing realised covariance is hampered by two empirical phenomena, namely the presence of market microstructure noise (e.g., price discreteness or bid-ask spread bounce) and non-synchronous trading. The impact of microstructure noise has received much attention in the univariate setting, where its effect on the realised variance has been well-documented. This builds on previous work in the noiseless case, including Andersen, Bollerslev, Diebold, and Labys (2001), Barndorff-Nielsen and Shephard (2002), or Mykland and Zhang (2006, 2008). A key to understanding the nature of the noise and a possible tool of how to deal with it is that microstructure noise induces autocorrelation in high-frequency returns and this leads to a bias problem (see, e.g., Zhou, 1996; Aït-Sahalia, Mykland, and Zhang, 2005; Hansen and Lunde, 2006). Currently, there are three main univariate approaches, where the damage caused by the noise is explicitly fixed: the two-scale estimator proposed by Zhang, Mykland, and Aït-Sahalia (2005) or its multi-scale version of Zhang (2006), the realised kernel introduced in Barndorff-Nielsen, Hansen, Lunde, and Shephard (2008a), which relies on autocovariance-based corrections, and finally the pre-averaging estimator of Podolskij and Vetter (2009) and Jacod, Li, Mykland, Podolskij, and Vetter (2009).

The multivariate version of this problem is, however, more complicated in that not only does the estimator need to be robust against various types of noise, it also has to cope with non-synchronous trading (see, e.g., Fisher, 1966). Asynchronicity causes high-frequency covariance estimates to be biased towards zero as the sampling frequency increases. This feature of the data was highlighted by Epps (1979) and has become known as the Epps effect. Intuitively, as the sampling frequency is increased, there are more and more zero-returns in the presence of non-synchronous trading, and this effect will dominate realised covariance and related statistics (e.g. realised correlation). Hayashi and Yoshida (2005) introduced an estimator, which is capable of dealing with non-synchronous data, but not with market microstructure noise. More recently, Zhang (2008) extended the two-scale RV to integrated covariance estimation in the simultaneous presence of noise and non-synchronicity, while in concurrent and independent work Barndorff-Nielsen, Hansen, Lunde, and Shephard (2008b) proposed a multivariate realised kernel. Additional work in this growing line of research includes Malliavin and Mancino (2002), Martens (2003), Renò (2003), Bandi and Russell (2005), Griffin and Oomen (2006), Large (2007), Voev and Lunde (2007), and Boudt, Croux, and Laurent (2008), among others.

In this paper, we propose to use a modulated realised covariance (MRC) to estimate the ex-post integrated covariance.

The econometric technique employed here for dealing with microstructure noise relies on rather simple pre-averaging of the high-frequency data, which makes the estimator both intuitive and trivial to implement. This relates to previous work in the univariate case, where pre-averaging has been suggested in Podolskij and Vetter (2009) and Jacod, Li, Mykland, Podolskij, and Vetter (2009). The current article draws ideas from these papers, but the multivariate extension is challenging, as it faces the additional complexity of non-synchronous trading and requires that the resulting estimator be positive semi-definite.

The pre-averaging approach depends on a bandwidth parameter, or window length, that grows with the sample and dictates the amount of averaging to be carried out. In turn, the choice of this tuning parameter controls the influence of microstructure noise on the MRC and, hence, also its asymptotic properties. In the optimal case, called balanced pre-averaging, this leads to an efficient $n^{-1/4}$ rate of convergence, which is known to be the fastest attainable (see, e.g., Gloter and Jacod, 2001a,b). This baseline MRC estimator, however, needs a bias-correction to be consistent for the integrated covariance. As a result, it is not guaranteed to be positive semi-definite in finite samples, though our empirical work indicates this shortcoming is not too much of a concern for more recent data. Nonetheless, as we show in the paper, it is straightforward to design a positive semi-definite estimator by increasing the pre-averaging window length slightly, which can also serve to make the MRC robust against more general noise processes.

The MRC is, in all its essence, a realised covariance computed on the back of pre-averaged high-frequency returns. As such, it depends on receiving synchronous observations as input, which clashes with the irregular spacing of real high-frequency data. We propose two distinct ways in which pre-averaging can be applied in the context of non-synchronous trading. First, we use traditional imputation schemes to map asynchronous data onto a common time grid, for example using previous-tick or refresh time, where the latter approach has been used in Barndorff-Nielsen, Hansen, Lunde, and Shephard (2008b). An MRC computed from such returns will be asymptotically robust to non-synchronous trading. Second, we extend the Hayashi and Yoshida (2005) estimator to the case of microstructure noise by using pre-averaging and show that it is consistent. This second estimator has the property that it can be implemented directly on the irregular, non-synchronous and noisy observations without any form of imputation. It therefore omits throwing away information in the sample and further avoids potential biases arising from artificially imputed returns.

An appealing feature of pre-averaging is that it is a general statistical tool that can be applied to many estimation problems. This proves useful in our setting, because as usual the mixed Gaussian central limit theorems feature an unknown conditional covariance matrix. In practice, this must be robustly estimated from data in the presence of noise and non-synchronous trading to make the distributional results feasible, such that confidence bands for elements of the integrated covariance matrix can be constructed. We outline how this can be done based on pre-averaged high-frequency data.

The paper progresses as follows. In section 2, we formulate the theoretical setup and define the MRC estimator. In section 3, we first show consistency of the MRC based on balanced pre-averaging and then derive its asymptotic

distribution. As discussed above, this estimator needs a bias-correction, so we carry on to study a modified MRC estimator, in which the degree of pre-averaging is increased. We also discuss the application of the MRC to non-synchronous data, derive a pre-averaged version of the Hayashi-Yoshida estimator, and relate the MRC to the multivariate realised kernel. In section 4, we propose a positive semi-definite estimator of the conditional covariance matrix that appears in the central limit theorem of MRC, which can be used to transform infeasible limit results into feasible ones. In section 5, the focus is shifted towards realised regression and correlation analysis. An empirical illustration is conducted in section 6, while section 7 draws conclusions and presents some ideas for future work. The appendix contains the derivations of our theoretical results.

2 Theoretical setup

We consider a log-price X defined on the probability space $(\Omega^0, \mathcal{F}^0, P^0)$ and equipped with an information filtration $(\mathcal{F}_t^0)_{t \geq 0}$. X has dimension d - the number of assets under consideration.

A standard no-arbitrage condition suggests security prices must follow semimartingales (see, e.g., Back, 1991; Delbaen and Schachermayer, 1994). These processes obey the fundamental theorem of asset pricing and, as a result, are used extensively to model the evolution of asset prices through time. In accordance with this, we model X as a semimartingale that follows the equation

$$X_t = X_0 + \int_0^t a_u du + \int_0^t \sigma_u dW_u, \quad t \geq 0, \quad (1)$$

where $a = (a_t)_{t \geq 0}$ is a d -dimensional predictable locally bounded drift vector, $\sigma = (\sigma_t)_{t \geq 0}$ an adapted càdlàg $d \times d$ covolatility matrix and $W = (W_t)_{t \geq 0}$ is d -dimensional Brownian motion.

This model is a Brownian semimartingale, or stochastic volatility model with drift, which permeates financial economics (cf., Ghysels, Harvey, and Renault, 1996, for a review). We think of this construct as governing an underlying efficient price process - the price that would prevail in the absence of market frictions, which we then subject to microstructure noise.

Of importance to our analysis is the quadratic covariation process of X , which is defined as

$$[X]_t = \text{p-lim}_{n \rightarrow \infty} \sum_{i=1}^n (X_{t_i} - X_{t_{i-1}}) (X_{t_i} - X_{t_{i-1}})' \quad (2)$$

for any sequence of deterministic partitions $0 = t_0 < t_1 < \dots < t_n = t$ with $\sup_i \{t_i - t_{i-1}\} \rightarrow 0$ for $n \rightarrow \infty$. In our setting, the quadratic covariation of X is given by

$$[X]_t = \int_0^t \Sigma_u du, \quad (3)$$

where $\Sigma = \sigma \sigma'$. The quadratic covariation is pivotal in financial economics (see, e.g., the reviews by Barndorff-Nielsen and Shephard, 2007; Andersen, Bollerslev, and Diebold, 2009), and we thus take Eq. (3) as defining the target that we are

interested in estimating. We note that for obvious reasons the matrix in Eq. (3) is also called the integrated covariance and both terms are used interchangeably.

Throughout the remainder of the paper, and without loss of generality, we restrict the clock t to evolve in the unit interval $[0, 1]$, which we think of as representing the passing of an economic event, for example a trading day.

2.1 Microstructure noise

In practice, market microstructure noise leads to a departure from the pure semimartingale model. Microstructure noise has many forms, including price discreteness and bid-ask spread bounce, which creates spurious variation in asset prices. As a result, we do not observe X from Eq. (1) in the market but a process Y , which is the efficient price distorted by noise. More precisely, we consider the process Y , observed at time points i/n , $i = 0, 1, \dots, n$, which is given as

$$Y_t = X_t + \epsilon_t, \quad (4)$$

where (ϵ_t) is an i.i.d. process with $X \perp\!\!\!\perp \epsilon$ (the symbol $\perp\!\!\!\perp$ is used to denote stochastic independence).

The noise process can be constructed as follows. We define a second probability space $(\Omega^1, \mathcal{F}^1, (\mathcal{F}_t^1)_{t \geq 0}, P^1)$, where Ω^1 denotes $\mathbb{R}^{[0,1]}$ and \mathcal{F}^1 the product Borel- σ -field on Ω^1 . Next, let Q be a probability measure on \mathbb{R} (Q is the marginal distribution of ϵ). For any $t \in [0, 1]$, $P_t^1 = Q$ and P^1 denotes the product $\otimes_{t \in [0,1]} P_t^1$. The filtered probability space $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t \geq 0}, P)$, on which we define the process Y , is given as

$$\left. \begin{aligned} \Omega &= \Omega^0 \times \Omega^1, & \mathcal{F} &= \mathcal{F}^0 \times \mathcal{F}^1, & \mathcal{F}_t &= \bigcap_{s > t} \mathcal{F}_s^0 \times \mathcal{F}_s^1, \\ P &= P^0 \otimes P^1. \end{aligned} \right\} \quad (5)$$

The multivariate noise process ϵ is assumed to satisfy:

$$\mathbb{E}(\epsilon_t) = 0, \quad \mathbb{E}(\epsilon_t \epsilon_t') = \Psi, \quad (6)$$

where Ψ is a positive definite $d \times d$ -matrix.

Remark 1 The empirical results found by Hansen and Lunde (2006) show that both the i.i.d. assumption on (ϵ_t) and the independence $X \perp\!\!\!\perp \epsilon$ can be called into question when sampling the data at very high frequencies, e.g., below the 1-minute mark (see also Diebold and Strasser, 2008). Jacod, Li, Mykland, Podolskij, and Vetter (2009) consider more general types of (1-dimensional) noise processes. Roughly speaking, they assume that the errors ϵ_t 's are, conditionally on X , centered and independent. The asymptotic theory developed in this paper still holds true for the multivariate version of such noise processes, but we restrict attention to models of the form in Eq. (4) to ease the exposition.

2.2 Pre-averaging of the data

It is intuitive that under mean zero i.i.d. microstructure noise some form of smoothing of the observed log-price Y should tend to diminish the impact of the noise. Effectively, we are going to approximate X_t , X being a continuous function of t , by an average of observations of Y in a neighborhood of t , the noise being averaged away.

Here, we describe in more detail how to conduct the pre-averaging. In particular, we consider a sequence of integers, k_n , and a number $\theta \in (0, \infty)$ such that

$$\frac{k_n}{\sqrt{n}} = \theta + o(n^{-1/4}). \quad (7)$$

An example of this would be $k_n = \lfloor \theta \sqrt{n} \rfloor$.

We also choose a function g on $[0, 1]$, which is continuous, piecewise continuously differentiable with a piecewise Lipschitz derivative g' with $g(0) = g(1) = 0$ and which satisfies $\int_0^1 g^2(s) ds > 0$. Furthermore, we introduce the following functions and numbers that are associated with g :

$$\begin{aligned} \phi_1(s) &= \int_s^1 g'(u) g'(u-s) du, & \phi_2(s) &= \int_s^1 g(u) g(u-s) du, & \psi_1 &= \phi_1(0), & \psi_2 &= \phi_2(0), \\ \Phi_{11} &= \int_0^1 \phi_1^2(s) ds, & \Phi_{12} &= \int_0^1 \phi_1(s) \phi_2(s) ds, & \Phi_{22} &= \int_0^1 \phi_2^2(s) ds. \end{aligned}$$

The functions ϕ_1 and ϕ_2 are assumed to be 0 outside the interval $[0, 1]$.

Next, with any process $V = (V_t)_{t \geq 0}$ we associate the following random variables

$$\Delta_i^n V = V_{\frac{i}{n}} - V_{\frac{i-1}{n}}, \text{ for } i = 1, \dots, n \quad \bar{V}_i^n = \sum_{j=1}^{k_n-1} g\left(\frac{j}{k_n}\right) \Delta_{i+j}^n V, \text{ for } i = 0, \dots, n - k_n + 1.$$

Applying this notation to Y , it can be seen that $\Delta_i^n Y$ represents the noisy high-frequency returns, while \bar{Y}_i^n is the pre-averaged return data, using the weight function g . It follows that the stochastic order of $\bar{Y}_i^n = \bar{X}_i^n + \bar{\epsilon}_i^n$ is controlled by the constant k_n , since

$$\bar{X}_i^n = O_p\left(\sqrt{\frac{k_n}{n}}\right), \quad \bar{\epsilon}_i^n = O_p\left(\sqrt{\frac{1}{k_n}}\right). \quad (8)$$

Thus, taking $k_n = O(\sqrt{n})$ implies that the orders of the two terms in Eq. (8) are equal, so that $\bar{Y}_i^n = O_p(n^{-1/4})$. This is called balanced pre-averaging and delivers the best rate of convergence. As shown below, it is also useful to look at cases in which a higher order of k_n is chosen. This results in a suboptimal rate of convergence, but it has desirable side-effects on the finite sample properties of our estimator.

The pre-averaging window length, k_n , depends on the tuning parameter θ , which needs to be chosen by the user. We will later discuss how to sensibly make this choice.

2.3 Modulated realised covariance

The core statistic of this paper is the multivariate extension of the estimator, which was introduced in Jacod, Li, Mykland, Podolskij, and Vetter (2009). We call it the modulated realised covariance (MRC) and define it as

$$MRC[Y]_n = \frac{n}{n - k_n + 2} \frac{1}{\psi_2 k_n} \sum_{i=0}^{n-k_n+1} \bar{Y}_i^n (\bar{Y}_i^n)'. \quad (9)$$

The factor $n/(n - k_n + 2)$ is a finite sample correction for the true number of summands in $MRC[Y]_n$ relative to the sample size n . It is sometimes left out in the presentation below, but it is always included in implementations on data.

Remark 2 The sum of outer products in Eq. (9) is realised covariance based on pre-averaged data. To build some intuition for our approach, we explain the usage of \bar{Y}_i^n in more detail. Suppose k_n is an even number and write

$$\hat{Y}_i^n = \frac{2}{k_n} \sum_{j=0}^{k_n/2-1} Y_{\frac{i+j}{n}},$$

which is a simple average of Y over $k_n/2$ terms. Because of this pre-averaging, \hat{Y}_i^n will be closer to the efficient price $X_{\frac{i}{n}}$. Next, we compute the realised covariation estimator based on these filtered increments by setting

$$\bar{Y}_i^n = 2(\hat{Y}_{i+\frac{k_n}{2}}^n - \hat{Y}_i^n) = \frac{1}{k_n} \left(\sum_{j=k_n/2}^{k_n-1} Y_{\frac{i+j}{n}} - \sum_{j=0}^{k_n/2-1} Y_{\frac{i+j}{n}} \right).$$

(However, as we shall see this induces a bias, which is a function of Ψ). This method was originally proposed by Podolskij and Vetter (2009) and using the above definition of \bar{Y}_i^n corresponds to choosing the weight function

$$g(x) = \min(x, 1 - x), \quad (10)$$

which is the most intuitive example. This function g will be used for all our empirical work, in which case the asymptotic constants are given as

$$\psi_1 = 1, \quad \psi_2 = \frac{1}{12}, \quad \Phi_{11} = \frac{1}{6}, \quad \Phi_{12} = \frac{1}{96}, \quad \Phi_{22} = \frac{151}{80640}.$$

Remark 3 As noted in Jacod, Li, Mykland, Podolskij, and Vetter (2009), to avoid biases in small samples we should replace the asymptotic constants and functions $\psi_1, \psi_2, \phi_1, \phi_2, \Phi_{11}, \Phi_{12}$, and Φ_{22} by their Riemann approximations:

$$\psi_1^{k_n} = k_n \sum_{i=1}^{k_n} \left(g\left(\frac{i}{k_n}\right) - g\left(\frac{i-1}{k_n}\right) \right)^2, \quad \psi_2^{k_n} = \frac{1}{k_n} \sum_{i=1}^{k_n-1} g^2\left(\frac{i}{k_n}\right),$$

$$\phi_1^{k_n}(j) = \sum_{i=j+1}^{k_n-1} \left(g\left(\frac{i-1}{k_n}\right) - g\left(\frac{i}{k_n}\right) \right) \left(g\left(\frac{i-j-1}{k_n}\right) - g\left(\frac{i-j}{k_n}\right) \right), \quad \phi_2^{k_n}(j) = \sum_{i=j+1}^{k_n-1} g\left(\frac{i}{k_n}\right) - g\left(\frac{i-j}{k_n}\right),$$

$$\Phi_{11}^{k_n} = k_n \left(\sum_{j=0}^{k_n-1} \left(\phi_1^{k_n}(j) \right)^2 - \frac{1}{2} \left(\phi_1^{k_n}(0) \right)^2 \right), \quad \Phi_{12}^{k_n} = \frac{1}{k_n} \left(\sum_{j=0}^{k_n-1} \phi_1^{k_n}(j) \phi_2^{k_n}(j) - \frac{1}{2} \phi_1^{k_n}(0) \phi_2^{k_n}(0) \right),$$

$$\Phi_{22}^{k_n} = \frac{1}{k_n^3} \left(\sum_{j=0}^{k_n-1} \left(\phi_2^{k_n}(j) \right)^2 - \frac{1}{2} \left(\phi_2^{k_n}(0) \right)^2 \right).$$

These are the actual terms that appear in the computations of the MRC. Note that for all appropriate indices of i and j , $\psi_i^{k_n} \rightarrow \psi_i$, $\phi_i^{k_n} \rightarrow \phi_i$, $\Phi_{ij}^{k_n} \rightarrow \Phi_{ij}$ as $n \rightarrow \infty$ at smaller order than $n^{-1/4}$, so the finite sample versions can be replaced in the central limit theorems given below.

3 Asymptotic properties of MRC

3.1 Consistency

Our first result inspects the probability limit of $MRC[Y]_n$.

Theorem 1 Assume that $\mathbb{E}(|\epsilon^j|^4) < \infty$ for all $j = 1, \dots, d$ and (k_n, θ) satisfy Eq. (7). As $n \rightarrow \infty$, it holds that

$$MRC[Y]_n \xrightarrow{p} \int_0^1 \Sigma_s ds + \frac{\psi_1}{\theta^2 \psi_2} \Psi. \quad (11)$$

Proof See appendix. ■

A couple of points are worth highlighting. First, as Theorem 1 shows $MRC[Y]_n$ is consistent for $\int_0^1 \Sigma_s ds$ up to a bias correction. The bias term depends on the unknown Ψ , which must be estimated from the data.

We set

$$\hat{\Psi}_n = \frac{1}{2n} \sum_{i=1}^n \Delta_i^n Y (\Delta_i^n Y)', \quad (12)$$

which is linked to the univariate estimator proposed by Aït-Sahalia, Mykland, and Zhang (2005) and Bandi and Russell (2006, 2008). Then, we obtain the convergence $\hat{\Psi}_n \xrightarrow{p} \Psi$, such that

$$MRC[Y]_n - \frac{\psi_1^{k_n}}{\theta^2 \psi_2^{k_n}} \hat{\Psi}_n \xrightarrow{p} \int_0^1 \Sigma_s ds.$$

Hence, for the remainder of the paper, we shall incorporate the bias correction term into the definition of $MRC[Y]_n$.¹ In doing so, we are no longer ensured that $MRC[Y]_n$ is positive semi-definite in finite samples. To deal with this problem,

¹Since $\sum_{i=1}^n \Delta_i^n Y (\Delta_i^n Y)' = 2n\Psi + \int_0^1 \Sigma_s ds + o_p(n^{-1})$, where the error of this approximation has expectation zero, the bias-corrected $MRC[Y]_n$ actually estimates $\left(1 - \frac{\psi_1^{k_n}}{\theta^2 \psi_2^{k_n}} \frac{1}{2n}\right) \int_0^1 \Sigma_s ds$ and thus needs to be rescaled by $1 / \left(1 - \frac{\psi_1^{k_n}}{\theta^2 \psi_2^{k_n}} \frac{1}{2n}\right)$. We include this bias correction in our empirical work but omit it throughout the remainder of the text to simplify notation.

we propose an alternative formulation of the MRC below, which uses a larger pre-averaging window k_n to avoid this step. Second, since $\hat{\Psi}_n$ is a \sqrt{n} -estimator of Ψ , it will not affect the CLT of $MRC[Y]_n$, as the latter converges at a slower rate. Third, our initial MRC estimator is based on synchronous data, assuming all log-prices are observed at the same instant in time. We will later extend the MRC to the non-synchronous setting.

3.2 The central limit theorem

We proceed with the central limit theorem for $MRC[Y]_n$. As in Jacod, Li, Mykland, Podolskij, and Vetter (2009), we only require a moment condition on ϵ to prove this result.

The concept of stable convergence is used, which we describe here. A sequence of random variables V^n is said to converge stably in law towards V , where V is defined on an appropriate extension $(\Omega', \mathcal{F}', P')$ of the probability space (Ω, \mathcal{F}, P) , if and only if for any \mathcal{F} -measurable, bounded random variable W and any bounded, continuous function f , the convergence

$$\lim_{n \rightarrow \infty} \mathbb{E}[Wf(V^n)] = \mathbb{E}'[Wf(V)]$$

holds.

We write this as $V^n \xrightarrow{d_s} V$ and note that stable convergence is a slightly stronger mode of convergence than weak convergence, or convergence in law, which is the special case obtained by taking $Y = 1$ (see, e.g., Rényi, 1963; Aldous and Eagleson, 1978, for further details on stable convergence). Jacod and Shiryaev (2003) discuss the extension of this notion to stable convergence of processes. The key reason we require the convergence in law stably is that the conditional covariance matrix in the CLT of $MRC[Y]_n$, avar_{MRC_n} , is a function of σ and therefore random, and the usual convergence in law is insufficient to ensure joint convergence of the bivariate vector $(MRC[Y]_n, \text{avar}_{MRC_n})$, which we need to apply the delta method to the joint asymptotic distribution and to construct confidence intervals.

Theorem 2 Assume that $\mathbb{E}(|\epsilon^j|^8) < \infty$ for all $j = 1, \dots, d$ and (k_n, θ) satisfy Eq. (7). As $n \rightarrow \infty$, it holds that

$$n^{1/4} \left(MRC[Y]_n - \int_0^1 \Sigma_s ds \right) \xrightarrow{d_s} \sum_{j', k'=1}^d \int_0^1 \gamma_s^{jk, j'k'} dB_s^{j'k'}, \quad (13)$$

where B is a standard d^2 -dimensional Brownian motion defined on an extension of $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t \geq 0}, P)$ with $B \perp \mathcal{F}$,

$$\sum_{j, m=1}^d \gamma_s^{kl, jm} \gamma_s^{k'l', jm} = \frac{2}{\psi_2^2} \left(\Phi_{22} \theta \Lambda_s^{kl, k'l'} + \frac{\Phi_{12}}{\theta} \Theta_s^{kl, k'l'} + \frac{\Phi_{11}}{\theta^3} \Upsilon^{kl, k'l'} \right),$$

and where Λ , Θ and Υ are $d \times d \times d \times d$ arrays with elements

$$\begin{aligned}\Lambda_s &= \left\{ \Sigma_s^{kk'} \Sigma_s^{ll'} + \Sigma_s^{kl'} \Sigma_s^{lk'} \right\}_{k,k',l,l'=1,\dots,d}, \\ \Theta_s &= \left\{ \Sigma_s^{kk'} \Psi^{ll'} + \Sigma_s^{kl'} \Psi^{k'l} + \Sigma_s^{k'l} \Psi^{kl'} + \Sigma_s^{ll'} \Psi^{kk'} \right\}_{k,k',l,l'=1,\dots,d}, \\ \Upsilon &= \left\{ \Psi^{kk'} \Psi^{ll'} + \Psi^{kl'} \Psi^{lk'} \right\}_{k,k',l,l'=1,\dots,d}.\end{aligned}\tag{14}$$

Proof See appendix. ■

Because $B \perp\!\!\!\perp \mathcal{F}$, we can write the convergence statement in Theorem 2 as follows:

$$n^{1/4} \left(MRC[Y]_n - \int_0^1 \Sigma_s ds \right) \xrightarrow{d_s} MN(0, \text{avar}_{\text{MRC}}),$$

where

$$\text{avar}_{\text{MRC}} = \frac{2}{\psi_2^2} \left(\Phi_{22} \theta \int_0^1 \Lambda_s ds + \frac{\Phi_{12}}{\theta} \int_0^1 \Theta_s ds + \frac{\Phi_{11}}{\theta^3} \Upsilon \right)\tag{15}$$

is the conditional covariance matrix. This means that the asymptotic distribution of $MRC[Y]_n$ is mixed normal. To make use of this result to construct confidence intervals for elements of $\int_0^1 \Sigma_u du$ in practice, we need to estimate avar_{MRC} , which is addressed in section 4.²

3.3 Choosing θ in practice

The avar_{MRC} matrix in Theorem 2 depends on the parameter θ , or in other words the window size k_n . If the purpose is to estimate some one-dimensional parameters (such as realised covariance, regression or correlation) by real-valued functions of the MRC, it is in principle possible to minimize avar_{MRC} by choosing the "best" θ for a fixed function g .³ To illustrate this point, we focus on the estimation problem in the univariate case, $d = 1$. In this situation, the expressions reduce to

$$\text{avar}_{\text{MRC}} = \frac{4}{\psi_2^2} \left(\Phi_{22} \theta \int_0^1 \sigma_s^4 ds + \frac{2\Phi_{12}}{\theta} \Psi^2 \int_0^1 \sigma_s^2 ds + \frac{\Phi_{11}}{\theta^3} \Psi^4 \right),$$

where $IV = \int_0^1 \sigma_s^2 ds$ and $IQ = \int_0^1 \sigma_s^4 ds$ are called the integrated variance and integrated quarticity, respectively. Minimizing this term with respect to θ results in solving a quadratic equation. Thus, for the optimal choice of θ , say θ^* , we

²The assumption that data be equidistant is not required for the consistency to hold true. It would also apply under identical observation times (i.e., synchronous but non-equidistant data). Here, $MRC[Y]_n$ is consistent with $\Delta_i^n V$ redefined as $\Delta_i^n V = V_{t_i} - V_{t_{i-1}}$ for any process V and $k_n = \theta\sqrt{n} + o(n^{1/4})$. This is not surprising: the realised covariance also remains consistent for irregular observations (by definition). The CLT also holds, but here the variable of integration ds needs to be replaced by dH_s , where H_s is the so-called "quadratic variation of time", see, e.g., Barndorff-Nielsen, Hansen, Lunde, and Shephard (2008b).

³The optimal choice of θ will depend on the original real-valued functions of the MRC. In this sense, there is no universal optimal θ .

get $\theta^* = f(IV, IQ, \Psi)$. Then, we can proceed using an iterative procedure to find an approximation of θ^* : (i) Choose a "reasonable" value for θ and compute \hat{IV} , \hat{IQ} and $\hat{\Psi}$, (ii) then compute $\hat{\theta}^*$ by setting $\hat{\theta}^* = f(\hat{IV}, \hat{IQ}, \hat{\Psi})$ and (iii) repeat this process until the values of $\hat{\theta}^*$ converge.

In practice, it might be preferable with a simpler guidance of how to choose θ for small values of n . However, as θ comes from asymptotic statistics, it does not give any precise instruction about this. Jacod, Li, Mykland, Podolskij, and Vetter (2009) report in their simulations studies that the univariate estimator is fairly robust to the choice of k_n , and they suggest to take $\theta = 1/3$. Furthermore, Christensen, Oomen, and Podolskij (2009) use simulation-based evidence to show that conservative (higher) choices of k_n are preferred from an MSE perspective. In the empirical illustration below, we follow the rule-of-thumb by choosing $\theta = 1/3$ and see that the results look good.

3.4 Positive semi-definite estimators

In the previous section, we used an optimal pre-averaging window length to construct the MRC, which balances the impact of the noise with the estimation of the integrated covariance matrix. This choice leads to an optimal rate of convergence - $n^{-1/4}$ - but requires that we subtract an estimate of Ψ to eliminate the bias induced by noise, and the final estimator is then not positive semi-definite in general. Here, we demonstrate how positive semi-definite estimates of $\int_0^1 \Sigma_s ds$ can be formed by increasing the bandwidth parameter k_n as to kill the influence of the noise, rather than balancing it. This comes at the cost of slowing down the speed at which MRC converges to the true integrated covariation.

Now, we take:

$$\frac{k_n}{n^{1/2+\delta}} = \theta + o\left(n^{-1/4+\delta/2}\right) \quad (16)$$

for some $0 < \delta < 1/2$, and set

$$MRC[Y]_n^\delta = \frac{n}{n - k_n + 2} \frac{1}{\psi_2 k_n} \sum_{i=0}^{n-k_n+1} \bar{Y}_i^n (\bar{Y}_i^n)'. \quad (17)$$

The following result shows that $MRC[Y]_n^\delta$ is consistent without a bias correction.

Theorem 3 Assume that $\mathbb{E}(|\epsilon^j|^4) < \infty$ for all $j = 1, \dots, d$ and (k_n, θ) satisfy Eq. (16). As $n \rightarrow \infty$, it holds that

$$MRC[Y]_n^\delta \xrightarrow{p} \int_0^1 \Sigma_s ds. \quad (18)$$

Proof See appendix. ■

In Theorem 3, the properties of the noise process do not show up in the stochastic limit of Eq. (18), because the influence of the noise is negligible by the choice of order for k_n made in Eq. (16) (refer back to Eq. (8)). This has some appealing advantages. First, $MRC[Y]_n^\delta$ is positive semi-definite by construction. Second, although we state and prove this result

in the i.i.d. noise case, Theorem 3 does in fact allow for more general noise dynamics than in Theorem 1. In particular, so long as $\mathbb{E}(\epsilon_t | X) = 0$ and $\bar{\epsilon}_i^n$ admits asymptotic normality at the usual rate $k_n^{-1/2}$, the theorem will hold (so we do not require any assumptions on the dependence between X and ϵ). Of course, the rate $k_n^{-1/2}$ is achieved in the i.i.d. case, but there are more general cases where it also holds (e.g., for q -dependent and mixing processes).

To show the CLT, we require a restriction on δ . This is because the bias caused by the noise, which is negligible in Theorem 3, becomes more substantial when multiplying with the rate of convergence.

Theorem 4 Assume that $\mathbb{E}(|\epsilon^j|^8) < \infty$ for all $j = 1, \dots, d$ and (k_n, θ) satisfy Eq. (16). As $n \rightarrow \infty$, it holds that

(i) If $\delta > 0.1$

$$n^{1/4-\delta/2} \left(MRC[Y]_n^\delta - \int_0^1 \Sigma_s ds \right) \xrightarrow{d_s} MN \left(0, \frac{2\Phi_{22}\theta}{\psi_2^2} \int_0^1 \Lambda_s ds \right), \quad (19)$$

where (Λ_s) is defined in (14).

(ii) If $\delta = 0.1$

$$n^{1/5} \left(MRC[Y]_n^\delta - \int_0^1 \Sigma_s ds \right) \xrightarrow{d_s} MN \left(\frac{\psi_1}{\theta^2 \psi_2} \Psi, \frac{2\Phi_{22}\theta}{\psi_2^2} \int_0^1 \Lambda_s ds \right). \quad (20)$$

Proof See appendix. ■

Theorem 4 amounts to a classical bias-variance trade-off. It shows the expected result that using a longer pre-averaging window $k_n = O(n^{1/2+\delta})$ averages enough to make $MRC[Y]_n^\delta$ consistent without a bias correction, but it also slows down its rate of convergence. The larger is δ , the harder is this effect. Note that the asymptotic variance depends solely on the volatility process σ , while the two noise-dependent terms (a cross-term and a pure noise term) appearing in Eq. (15) are wiped out.

The optimal choice $\delta = 0.1$ results in a $n^{-1/5}$ rate of convergence, and a large sample bias of the order $n^{-1/5} \frac{\psi_1}{\theta^2 \psi_2} \Psi$. If desired, the bias can be estimated using $\hat{\Psi}_n$, but it should be small for more recent data and could be ignored.⁴

This result bears some resemblance to that of the multivariate realised kernel of Barndorff-Nielsen, Hansen, Lunde, and Shephard (2008b). A flat-top kernel estimator can be designed to converge at rate $n^{-1/4}$, but the resulting estimator may go negative. To guarantee positive semi-definiteness, the authors propose kernel corrections that are not entirely flat-top and they show this results in $n^{-1/5}$ rate of convergence and a non-zero asymptotic mean as produced here as well.

3.5 Mapping MRC into a realised kernel estimator

We can indeed make a stronger link between the multi-scale RV, the kernel approach and our pre-averaging method, when estimating quadratic variation. Here, we provide some more details about this relationship. To fix ideas, we take $d = 1$

⁴If we do bias-correct $MRC[Y]_n^\delta$, the resulting estimator is then again not ensured to be positive semi-definite.

(i.e., the univariate setting) and compare only to the kernel approach. Barndorff-Nielsen, Hansen, Lunde, and Shephard (2008a) then show how the multi-scale RV fits into the realised kernel setting.

When we explicitly include the bias correction and ignore finite sample issues, the MRC estimator is given by the formula

$$MRC[Y]_n = \frac{1}{\psi_2 k_n} \sum_{i=0}^{n-k_n+1} |\bar{Y}_i^n|^2 - \frac{\psi_1}{\theta^2 \psi_2} \hat{\Psi}_n. \quad (21)$$

Consider a flat-top kernel-based estimator with kernel weights

$$k(s) = \frac{\phi_2(s)}{\psi_2},$$

where $\phi_2(s)$ and ψ_2 are defined as in Section 2.2. We call the function k a flat-top kernel, if (i) $k(0) = 1$, (ii) $k(1) = 0$ and (iii) $k'(0) = k'(1) = 0$. We note that $k(s) = \phi_2(s)/\psi_2$ satisfies all three conditions: (i) and (ii) are trivial, while the third condition follows from the identity:

$$k'(s) = -\frac{1}{\psi_2} \int_s^1 g(x)g'(x-s)dx,$$

and integration by parts (recall that $g(0) = g(1) = 0$).

In the next step, we map the MRC statistic into a kernel-like one as follows

$$MRC[Y]_n = \sum_{i=1}^n \delta_{0i} |\Delta_i^n Y|^2 + 2 \sum_{h=1}^{k_n-2} \sum_{i=1}^{n-h} \delta_{hi} \Delta_i^n Y \Delta_{i+h}^n Y \quad (22)$$

with

$$\delta_{0i} = \begin{cases} \frac{1}{\psi_2 k_n} \sum_{j=1}^i g^2\left(\frac{j}{k_n}\right) - \frac{\psi_1}{\theta^2 \psi_2 2n} : & 1 \leq i \leq k_n - 2 \\ \frac{1}{\psi_2 k_n} \sum_{j=1}^{k_n-1} g^2\left(\frac{j}{k_n}\right) - \frac{\psi_1}{\theta^2 \psi_2 2n} : & k_n - 1 \leq i \leq n - k_n + 2 \\ \frac{1}{\psi_2 k_n} \sum_{j=1}^{n-i+1} g^2\left(\frac{k_n-j}{k_n}\right) - \frac{\psi_1}{\theta^2 \psi_2 2n} : & n - k_n + 3 \leq i \leq n \end{cases} \quad (23)$$

and

$$\delta_{hi} = \begin{cases} \frac{1}{\psi_2 k_n} \sum_{j=1}^i g\left(\frac{j}{k_n}\right) g\left(\frac{j+h}{k_n}\right) : & 1 \leq i \leq k_n - h - 2 \\ \frac{1}{\psi_2 k_n} \sum_{j=1}^{k_n-h-1} g\left(\frac{j}{k_n}\right) g\left(\frac{j+h}{k_n}\right) : & k_n - h - 1 \leq i \leq n - k_n + 2 \\ \frac{1}{\psi_2 k_n} \sum_{j=1}^{n-i+1} g\left(\frac{k_n-j}{k_n}\right) g\left(\frac{k_n-j+h}{k_n}\right) : & n - k_n + 3 \leq i \leq n \end{cases} \quad (24)$$

for $1 \leq h \leq k_n - 2$.

An example and some remarks are now in order.

Example 1 Take $g(x) = \min(x, 1-x)$ for $x \in [0, 1]$, which is our canonical weight function. Then, the corresponding kernel $k(s) = \phi_2(s)/\psi_2$ is the Parzen kernel.

Remark 4 Apart from border terms, i.e. terms close to 0 and 1, the pre-averaging estimator coincides with the kernel-based estimator using the flat-top kernel function $k(s) = \phi_2(s)/\psi_2$. Both estimators possess the optimal rate of convergence $n^{-1/4}$, but the border terms affect their asymptotic distribution. In fact, to arrive at their central limit theorem for the realised kernel, Barndorff-Nielsen, Hansen, Lunde, and Shephard (2008b) need to apply some averaging (or "jittering") to edge terms, while the pre-averaging estimator is asymptotically mixed normal "by construction".

Remark 5 Using the definition $k(s) = \phi_2(s)/\psi_2$, we learn that for every weight function g there exists a flat-top kernel k . However, the converse statement is not true in general. Thus, the class of flat-top kernels is larger than our class of pre-averaging functions, but this does not appear to be a noticeable disadvantage for practical applications (Barndorff-Nielsen, Hansen, Lunde, and Shephard, 2008b, for example, advocate using the Parzen kernel in practice, which corresponds to our g function by Example 1).

It is worthwhile to underscore that pre-averaging is a general approach, which can be used to estimate various properties of semimartingales, when they are cloaked with noise. Jacod, Li, Mykland, Podolskij, and Vetter (2009), for example, pre-average realised variance, but it has also been used in jump-robust inference in conjunction with the bipower variation statistic (e.g., Podolskij and Vetter, 2009) or the quantile-based realised variance (e.g., Christensen, Oomen, and Podolskij, 2009). This is also the reason, why we can estimate the asymptotic conditional variance in the CLT using the same type of estimator to deliver a feasible result. Other estimation methods are typically designed to estimate quadratic variation and cannot, in general, be used to solve other estimation problems.

3.6 Applying the MRC to non-synchronous data

Non-synchronous trading has long been a recognized feature of multivariate high-frequency data (see, e.g., Fisher, 1966; Lo and MacKinlay, 1990). This causes spurious cross-autocorrelation amongst asset price returns sampled at regular intervals in calendar time, as new information gets built into prices at varying intensities. It is well-known that high-frequency realised covariance estimates, using for example the previous-tick method to align returns, are biased towards zero in this setting (e.g., Epps, 1979).

Motivated by these shortcomings of realised covariance, a number of alternative procedures have been proposed in the literature. Scholes and Williams (1977) and Dimson (1979) suggested to include leads and lags of sample autocovariances of high-frequency returns into the realised covariance to correct for stale prices. This results in a bias reduction but also increases the variance of the estimator (e.g., Griffin and Oomen, 2006). Importantly, the lead-lag realised covariance is generally inconsistent in the presence of noise. More recently, Zhang (2008) extends the two-scale RV to the multivariate setting and shows that it is consistent under non-synchronous prices and noise. An analytic derivation of the impact of Epps effect on realised covariance under previous-tick prices is also given. In independent and concurrent work to ours,

Barndorff-Nielsen, Hansen, Lunde, and Shephard (2008b) propose a multivariate realised kernel. They rely on the concept of refresh time sampling to match returns in calendar time and show that stale pricing errors under refresh time do not change the asymptotic distribution of the multivariate realised kernel.

A related approach can be taken to make the MRC robust to non-synchronous observation times. In particular, under suitable regularity conditions on the sampling times, non-synchronous trading does not alter the asymptotic distribution of the MRC, when constructing synchronous time series of returns. A multivariate time series of high-frequency returns obtained in this fashion can therefore be plugged into the asymptotic theory developed above without concern. The proof of these results are highly technical, but their validity should be clear in the light of the comparison with the multivariate realised kernel given above. Therefore, we omit the detailed proofs of these results.

3.6.1 A pre-averaged Hayashi-Yoshida estimator

Hayashi and Yoshida (2005, 2008) develop an alternative procedure for covariance measurement in the noise-free case, which is based on the original non-synchronous data (see, e.g., de Jong and Nijman, 1997; Martens, 2003; Palandri, 2006; Corsi and Audrino, 2007, for related work). This estimator has the profound advantage that it does not throw away information that is typically lost using a synchronization procedure. Here, we show how this estimator can also be made robust to noise by using pre-averaging.

Given the vector of log-prices $Y = (Y^1, \dots, Y^d)'$, which is defined by the noisy diffusion model in Eq. (4), we now assume that the component processes (Y^k) are observed at non-random time points $t_i^{(k)}$, for $i = 1, \dots, n_k$, with $(t_i^{(k)})$ being a partition of the interval $[0, 1]$ and $k = 1, \dots, d$. In addition, we need some regularity conditions on the sampling such that all time schemes are comparable in the following way: $\max |t_i^{(k)} - t_{i-1}^{(k)}| \rightarrow 0$ as $n_k \rightarrow \infty$, for $k = 1, \dots, d$ and

$$\max_{1 \leq i \leq n_k} \# \left\{ t_j^{(k)} \mid t_j^{(k)} \in [t_i^{(l)}, t_{i+1}^{(l)}] \right\} \leq K, \quad (25)$$

for $1 \leq k, l \leq d$ and some $K > 0$, where K is independent of n_k , for $k = 1, \dots, d$. The latter condition says that data from one process do not cluster in any single interval of the others. Finally, the following condition is needed

$$\frac{\max |t_i^{(k)} - t_{i-1}^{(k)}|}{\min |t_i^{(k)} - t_{i-1}^{(k)}|} \leq c, \quad (26)$$

where $c > 0$ is a constant independent of n_k for all $1 \leq k \leq d$. This condition implies that $n_k \max |t_i^{(k)} - t_{i-1}^{(k)}| \leq c$, which could be assumed instead.

A consistent estimator of the integrated covariance $\int_0^1 \sum_s^{kl} ds$ between assets Y^k and Y^l can then be constructed as follows. We set $n = \sum_{k=1}^d n_k$ and define the statistic

$$HY[Y]_n^{(k,l)} = \frac{1}{(\psi_{HY} k_n)^2} \sum_{i=0}^{n_k - k_n + 1} \sum_{j=0}^{n_l - k_n + 1} Y_i^k Y_j^l \mathbb{I}_{\{(t_i^{(k)}, t_{i+k_n}^{(k)}) \cap (t_j^{(l)}, t_{j+k_n}^{(l)}) \neq \emptyset\}}, \quad (27)$$

where $\psi_{HY} = \int_0^1 g(x)dx$ and $\mathbb{I}_{\{\bullet\}}$ is the indicator function discarding pre-averaged returns that do not overlap in time. Hence, $HY[Y]_n^{(k,l)}$ is a pre-averaged version of the Hayashi and Yoshida (2005) estimator. Note that under the previous conditions, n , n_k and n_l are of the same order and that n controls the pre-averaging window k_n .

Theorem 5 Assume that $\mathbb{E}(|\epsilon^j|^4) < \infty$ for all $j = 1, \dots, d$, (k_n, θ) satisfy Eq. (7) and $g(x) > 0$ for $x \in (0, 1)$. As $n \rightarrow \infty$, it holds that

$$HY[Y]_n^{(k,l)} \xrightarrow{P} \int_0^1 \Sigma_s^{kl} ds, \quad (28)$$

for $1 \leq k, l \leq d$.

Proof See appendix. ■

Interestingly, Theorem 5 shows the somewhat surprising result that there is no asymptotic noise-induced bias in $HY[Y]_n^{(k,l)}$, not even when the spacings $(t_i^{(k)})$ and $(t_j^{(l)})$ are identical. This can be seen as follows. First, under the i.i.d. structure on the noise only products of the form $\epsilon_{t_i^{(k)}}^k \epsilon_{t_j^{(l)}}^l$ with $t_i^{(k)} = t_j^{(l)} = t$ contribute to potential bias. We consider that set of points and assume that $k_n \leq i \leq n_1 - k_n$ and $k_n \leq j \leq n_2 - k_n$ (this is innocent, for the summands which do not fulfill this are negligible). Then, an inspection of $HY[Y]_n^{(k,l)}$ shows that all products $\epsilon_t^k \epsilon_t^l$ appear with the factor

$$\left(\sum_{j=0}^{k_n-1} g\left(\frac{j+1}{k_n}\right) - g\left(\frac{j}{k_n}\right) \right)^2$$

in front. But $\sum_{j=0}^{k_n-1} g\left(\frac{j+1}{k_n}\right) - g\left(\frac{j}{k_n}\right) = 0$, because $g(0) = g(1) = 0$, so these terms drop out of the summation.

While $HY[Y]_n^{(k,l)}$ has the advantage that it is free of prior alignment of log-prices and hence does not throw away information in the sample, it does suffer from being a pairwise estimator, which means that once we assemble all the single variance/covariance estimates into a full covariance matrix, it is not guaranteed to be positive semi-definite. Still, there are some problems in financial economics, in which it is only the accuracy of the estimator that matters and positive semi-definiteness is less important, for example in asset allocation and risk management under gross exposure constraints (e.g. Fan, Zhang, and Yu, 2009). Moreover, our empirical results suggest that $(HY[Y]_n^{(k,l)})_{1 \leq k, l \leq d}$ rarely fails to be positive semi-definite for recent data.

Proposition 1 Assume that $\mathbb{E}(|\epsilon^j|^4) < \infty$ for all $j = 1, \dots, d$. If (k_n, θ) satisfy Eq. (7) and $g(x) > 0$ for $x \in (0, 1)$, then

$$\text{var}(HY[Y]_n^{(k,l)}) = O(n^{-1/2}),$$

that is, $HY[Y]_n^{(k,l)}$ has the optimal rate of convergence.

Proposition 1 shows that the rate of convergence associated with $HY[Y]_n^{(k,l)}$ is $n^{-1/4}$. However, we dispense with the actual proof of the proposition and CLT, as the resulting formulas are quite involved (it would go along the lines of the proofs given in the appendix). In the empirical section, we gauge the properties of this estimator on actual data and find that the pre-averaged version performs very well.

4 A positive semi-definite estimator of the asymptotic covariance matrix

In order to make Theorem 2 and 4 feasible, we need to estimate the asymptotic covariance matrix avar_{MRC} , as it is given by Eq. (15), and we give an explicit estimator for avar_{MRC} next. More precisely, we present a positive semi-definite estimator of the asymptotic covariance matrix of the vectorized statistic in Eq. (13).

We set

$$\chi_i^n = \text{vec}\left(\bar{Y}_i^n (\bar{Y}_i^n)'\right), \quad (29)$$

where vec denotes the well-known vectorisation operator that stacks the columns of a matrix below one another. Define the statistic

$$V_n(g) = \sum_{i=0}^{n-k_n+1} \chi_i^n (\chi_i^n)' - \frac{1}{2} \sum_{i=0}^{n-2k_n+1} \left(\chi_i^n (\chi_{i+k_n}^n)' + \chi_{i+k_n}^n (\chi_i^n)' \right)$$

(Notice that the above estimator depends on the parameter θ and the function g). Clearly, $V_n(g)$ is positive semi-definite. Moreover, we have for any $1 \leq k, k', l, l' \leq d$

$$V_n^{(k-1)d+k', (l-1)d+l'}(g) \xrightarrow{p} a_B(g, \theta) \int_0^1 \Lambda_u^{kk', ll'} du + a_M(g, \theta) \int_0^1 \Theta_u^{kk', ll'} du + a_N(g, \theta) \Upsilon^{kk', ll'},$$

where $a_B(g, \theta) = \theta^2 \psi_2^2$, $a_M(g, \theta) = \psi_1 \psi_2$ and $a_N(g, \theta) = \frac{\psi_1^2}{\theta^2}$ (this convergence is obtained by similar arguments as presented in the proof of Theorem 1; see Kinnebrock and Podolskij (2008) for more details). We need to estimate the quantity

$$\text{avar}_{\text{MRC}} = \frac{2}{\psi_2^2} \left(\Phi_{22} \theta \int_0^1 \Lambda_u du + \frac{\Phi_{12}}{\theta} \int_0^1 \Theta_u du + \frac{\Phi_{11}}{\theta^3} \Upsilon \right),$$

where all the constants refer to some given function g_0 . Suppose that $g_0 = \min(x, 1-x)$. Now let us consider three different functions g_1, g_2 and g_3 such that the matrix

$$A(g_1, g_2, g_3) = \begin{pmatrix} a_B(g_1, \theta) & a_M(g_1, \theta) & a_N(g_1, \theta) \\ a_B(g_2, \theta) & a_M(g_2, \theta) & a_N(g_2, \theta) \\ a_B(g_3, \theta) & a_M(g_3, \theta) & a_N(g_3, \theta) \end{pmatrix}$$

is invertible and all components of the vector

$$C(g_1, g_2, g_3) = \left(\frac{2\Phi_{22}\theta}{\psi_2^2}, \frac{2\Phi_{12}}{\psi_2^2\theta}, \frac{2\Phi_{11}}{\psi_2^2\theta^3} \right) A^{-1}(g_1, g_2, g_3)$$

are positive. Finally, consider the estimators $V_n(g_k)$ associated with the functions g_k ($k = 1, 2, 3$). Then the statistic

$$\widehat{\text{avar}}_{\text{MRC},n} = C^{(1)}(g_1, g_2, g_3)V_n(g_1) + C^{(2)}(g_1, g_2, g_3)V_n(g_2) + C^{(3)}(g_1, g_2, g_3)V_n(g_3) \quad (30)$$

is positive semi-definite, because $V_n(g_k)$ are positive semi-definite and all $C^{(k)}(g_1, g_2, g_3)$ ($k = 1, 2, 3$) are positive. Moreover, it holds that

$$\widehat{\text{avar}}_{\text{MRC},n}^{(k-1)d+k', (l-1)d+l'} \xrightarrow{p} \text{avar}_{\text{MRC}}^{kk', ll'},$$

for any $1 \leq k, k', l, l' \leq d$.

It is important to choose g_1, g_2 and g_3 such that $C^{(1)}(g_1, g_2, g_3)$, $C^{(2)}(g_1, g_2, g_3)$ and $C^{(3)}(g_1, g_2, g_3)$ are positive. There are various examples of functions that fulfill this requirement. By searching in the class of functions for which $g(x) = x^a(1-x)^b$ with $a, b > 0$, we found one such set of functions:

$$g_1(x) = x(1-x)^6, \quad g_2(x) = x(1-x)^2, \quad g_3(x) = x^4(1-x)^5.$$

5 Asymptotic theory for covariance, regression and correlation

The results in section 3 can be applied in order to compute confidence intervals for some functionals of $\int_0^1 \Sigma_u du$ that are important in practice, such as covariance, regression and correlation. For the i th and j th asset, these quantities are given by

$$\int_0^1 \Sigma_u^{ij} du, \quad \beta^{(ji)} = \frac{\int_0^1 \Sigma_u^{ij} du}{\int_0^1 \Sigma_u^{ii} du}, \quad \rho^{(ji)} = \frac{\int_0^1 \Sigma_u^{ij} du}{\sqrt{\int_0^1 \Sigma_u^{ii} du \int_0^1 \Sigma_u^{jj} du}}. \quad (31)$$

Theorem 1, 3 or 5 can be invoked to provide consistent estimates of $\int_0^1 \Sigma_u^{ij} du$, $\beta^{(ji)}$ and $\rho^{(ji)}$, e.g. for $\text{MRC}[Y]_n$

$$\begin{aligned} \text{MRC}[Y]_n^{i,j} &\xrightarrow{p} \int_0^1 \Sigma_u^{ij} du, \quad \hat{\beta}_n^{(ji)} = \frac{\text{MRC}[Y]_n^{i,j}}{\text{MRC}[Y]_n^{i,i}} \xrightarrow{p} \beta^{(ji)}, \\ \hat{\rho}_n^{(ji)} &= \frac{\text{MRC}[Y]_n^{i,j}}{\sqrt{\text{MRC}[Y]_n^{i,i} \text{MRC}[Y]_n^{j,j}}} \xrightarrow{p} \rho^{(ji)} \end{aligned} \quad (32)$$

for any $1 \leq i, j \leq d$. The estimators in (32) are called modulated realised covariance, regression and correlation, respectively. In the next theorem we present the associated feasible central limit theorems, which follow from Theorem 2 and the delta method for stable convergence.

Theorem 6 Assume that $\mathbb{E}(|\epsilon^j|^8) < \infty$ for all $j = 1, \dots, d$ and (k_n, θ) satisfy Eq. (7). As $n \rightarrow \infty$, it holds that

$$\frac{n^{1/4} \left(MRC[Y]_n^{i,j} - \int_0^1 \Sigma_u^{ij} du \right)}{\widehat{avar}_{MRC,n}^{(i-1)d+j, (i-1)d+j}} \xrightarrow{d} N(0, 1), \quad (33)$$

$$\frac{n^{1/4} (\hat{\beta}_n^{(ji)} - \beta^{(ji)})}{\sqrt{\left(MRC[Y]_n^{i,i} \right)^{-2} g_n^{(ji)}}} \xrightarrow{d} N(0, 1), \quad (34)$$

$$\frac{n^{1/4} (\hat{\rho}_n^{(ji)} - \rho^{(ji)})}{\sqrt{\left(MRC[Y]_n^{i,i} MRC[Y]_n^{j,j} \right)^{-1} h_n^{(ji)}}} \xrightarrow{d} N(0, 1), \quad (35)$$

where $\widehat{avar}_{MRC,n}$ is given in (30) and $g_n^{(ji)}$ and $h_n^{(ji)}$ are defined by

$$g_n^{(ji)} = \left(1, -\hat{\beta}_n^{(ji)} \right) \Gamma_n \left(1, -\hat{\beta}_n^{(ji)} \right)', \quad h_n^{(ji)} = \left(-\frac{1}{2} \hat{\beta}_n^{(ji)}, 1, -\frac{1}{2} \hat{\beta}_n^{(ij)} \right) \bar{\Gamma}_n \left(-\frac{1}{2} \hat{\beta}_n^{(ji)}, 1, -\frac{1}{2} \hat{\beta}_n^{(ij)} \right)'$$

with

$$\Gamma_n = \begin{pmatrix} \widehat{avar}_{MRC,n}^{(i-1)d+j, (i-1)d+j} & \widehat{avar}_{MRC,n}^{(i-1)d+j, (i-1)d+i} \\ \bullet & \widehat{avar}_{MRC,n}^{(i-1)d+i, (i-1)d+i} \end{pmatrix},$$

$$\bar{\Gamma}_n = \begin{pmatrix} \widehat{avar}_{MRC,n}^{(i-1)d+i, (i-1)d+i} & \widehat{avar}_{MRC,n}^{(i-1)d+j, (i-1)d+i} & \widehat{avar}_{MRC,n}^{(i-1)d+i, (j-1)d+j} \\ \bullet & \widehat{avar}_{MRC,n}^{(i-1)d+j, (i-1)d+j} & \widehat{avar}_{MRC,n}^{(i-1)d+j, (j-1)d+j} \\ \bullet & \bullet & \widehat{avar}_{MRC,n}^{(j-1)d+j, (j-1)d+j} \end{pmatrix}.$$

All the required terms are easy to compute, so it is rather simple to implement the estimators.

6 Empirical Illustration

To illustrate some empirical features of modulated realised covariance, regression and correlation, we retrieved high-frequency data for a five-dimensional vector of assets from Wharton Research Data Services (WRDS).⁵ We picked four equities at random from the S&P 500 constituents list as of July 1, 2009. We then added a 5th element, namely the S&P 500 Depository Receipt (ticker symbol SPY), which is an exchange-traded fund that tracks the large-cap segment of the U.S. stock market. As such, it can be viewed as generating market-wide index returns. The four remaining stocks are the following (with ticker symbol and industry classification in parenthesis): Bristol-Myers Squibb (BMY, health care), Lockheed Martin (LMT, industrials), Oracle (ORCL, information technology) and Sara Lee (SLE, consumer staples),

⁵An extensive simulation study can be found in Kinnebrock and Podolskij (2008).

Table 1: Descriptive statistics and number of data before and after filtering

Stock	BMY	LMT	ORCL	SLE	SPY
Exchange	N	N	Q	N	P
<i>Panel A: Transaction data</i>					
Raw trades	1085420	811607	8153413	533517	7685215
Corrected/Abnormal/Zeros	111	61	13806	85	2584
Time aggregation	231826	139181	6599286	78039	6000898
#Trades	853483	672365	1540321	455393	1681733
Intensity	3400	2679	6137	1814	6700
Noise ratio, γ	0.363	0.336	0.484	0.656	0.202
<i>Panel B: Quotation data</i>					
Raw quotes	5402607	3245315	23411495	3208830	17536447
Negative/Wide/Zeros	643	3917	2623	604	256
Time aggregation	2547054	1075914	20050224	979299	12851464
#Quotes	2854910	2165484	3358648	2228927	4684727
Intensity	11374	8627	13381	8880	18664
Avg. spread (in cents)	1.273	2.389	1.017	1.215	1.575
Noise ratio, γ	0.205	0.219	0.203	0.310	0.109

Note. This table reports some descriptive statistics and liquidity measures for the selection of stocks included in our empirical application. We show the exchange from which data are extracted. The exchange code is: N = NYSE, Q = NASDAQ and P = Pacific. Raw trades/quotes is the total number of data available from these exchanges during the trading session, while # trades/quotes is the total sample remaining after filtering the data. Intensity is the average number of data pr. day, while the noise ratio is defined in Oomen (2006).

thus representing a broad category of industries. We use both trades and quotes data for the sample period that covers the whole of 2006, which results in 251 trading days.

Table 1 reports some descriptive statistics for our universe of stocks and sample period. As can be seen, these equities display varying degrees of liquidity with ORCL and SPY being the most liquid, while LMT and SLE are the least liquid. Also reported in the table is the univariate noise ratio statistic, γ , which is a noise-to-signal measure that describes the level of microstructure noise to integrated variance (see, e.g., Oomen, 2006, for further details on the noise ratio). Generally speaking, there is a tendency for more frequently traded companies to contain less microstructure noise, the notable exception being the transaction data for ORCL.

6.1 Filtering procedures

As a preliminary step, we subduced the sample to data cleaning. Pre-cleaning high-frequency data is necessary, because the raw data has many invalid observations (e.g., data with misplaced decimal points, or trades that are reported out-of-sequence). Our filter is roughly identical to that used by Barndorff-Nielsen, Hansen, Lunde, and Shephard (2008b) with some minor differences. Here, we briefly describe the filtering rules we employ.

Trades and quotes: The following rules are applied to both trades and quotes data. a) We keep data from a single exchange: (Pacific for SPY and primary exchange for the 4 remaining equities, see Table 1), b) we delete data with time stamps outside the regular exchange opening hours from 9:30am to 4:00pm, c) we delete rows with a transaction price, bid or ask quote of zero, and d) we aggregate data with identical time stamp using volume-weighted average prices (using total transaction volume or quoted bid and ask volume, respectively).

Trades only: We delete entries with a correction indicator $\neq 0$ or with abnormal sales condition.

Quotes only: We delete quotes with negative spreads and rows where the quoted spread exceeds 10 times the median spread for that day.

Table 1 also reports how many observations that are lost by passing these filters through the data. It should be noted that the "Trades Only" and "Quotes Only" filters generally tend to reduce the sample by only a very small fraction.

6.2 Implementation of MRC estimators

We implement both the $MRC[Y]_n$ and $MRC[Y]_n^\delta$ estimators of section 2 and 3.4, respectively. Recall that $MRC[Y]_n$ converges at rate $n^{-1/4}$, it needs to be corrected for bias, and as a result is not guaranteed to be positive semi-definite. We base $MRC[Y]_n^\delta$ on $\delta = 0.1$, among many plausible choices, which results in a $n^{-1/5}$ rate of convergence, a small finite sample bias that we omit correcting and, hence, positive semi-definiteness by construction. Two sampling schemes are used, calendar time and refresh time sampling, which yields a total of four combinations. We use pre-averaging windows found as $k_n = \lfloor \theta n^{\delta'} \rfloor$, where $\theta = 1/3$ and $\delta' = (0.5, 0.6)$ for our two choices. The selection of θ follows the guide from Jacod, Li, Mykland, Podolskij, and Vetter (2009). We set $n = 390$ for the calendar time-based estimators, which is 1-minute sampling, while n is determined automatically by the data for the refresh time sampling scheme.

6.3 High-frequency covariance analysis

Here, we inspect the outcome of applying the estimators introduced above, after which we look at transforms of the covariance matrix to study regression and correlation estimates. As a comparison, we also compute the standard realised covariance from 15-min, 5-min, and 15-sec previous-tick data.

In Table 2, we report the sample covariance matrix estimates averaged across the 251 days. This table is constructed in the usual way, displaying the results based on transaction prices in the upper diagonal (including the main diagonal), while the strict lower diagonal elements are the corresponding results based on quotation data. Consistent with prior literature, we see from the table that the standard realised covariance is suffering from Epps effect, when sampling runs quickly. All estimated covariance terms lie in the positive region, but for Cov^{15s} they are heavily compressed towards zero. This is less of a concern for $Cov^{avg(5m,15m)}$, which should tend to capture the average level of the covariance structure well, while not being seriously influenced by microstructure frictions and Epps effect. Turning next to the estimators proposed in this paper, we note that the time series average of both MRC versions and the noise-robust HY estimator are in line with that produced by $Cov^{avg(5m,15m)}$, showing that they appear free of any systematic bias. The Hayashi and Yoshida (2005) estimator produces a strong downwards bias in the covariance estimates, when it is applied directly to noisy and irregular high-frequency data. This reaffirms previous empirical work (see, e.g., Barndorff-Nielsen, Hansen, Lunde, and Shephard, 2008b; Griffin and Oomen, 2006; Voev and Lunde, 2007), so these results are not surprising or novel. The pre-averaged version $HY[Y]_n^{(k,l)}$, however, does a much better job and tends to agree with the average level of other noise-robust estimators.

Finally, we turn to the issue of positive semi-definiteness. As noted above, $MRC[Y]_n$ and $HY[Y]_n^{(k,l)}$ are not guaranteed to possess this property. Nonetheless, they do not fail to be positive semi-definite on a single instance across our sample period. This is true for both trade and quote data, and both combinations of the bias-corrected MRC estimator. Thus, while theoretically a concern, this problem does not appear to occur frequently in practice, although the conclusion might change for other data sets.

6.4 Analysing realised beta

We now focus on estimating $\beta^{(ji)}$ by $\hat{\beta}_n^{(ji)} = MRC[Y]_n^{i,j} / MRC[Y]_n^{i,i}$, where we take $i = SPY$ and form regressions by using $j = BMY, LMT, ORCL, SLE$. This type of regression, where individual equity covariances with the market are regressed onto a market-wide realised variance measure, is important in financial economics, for example within conditional CAPM, since only systematic risk should be rewarded with expected excess returns.

In Figure 1, we plot the MRC-based betas from transaction prices and refresh time sampling. The corresponding plots from the other estimators proposed in this paper are qualitatively similar. As in Barndorff-Nielsen, Hansen, Lunde, and Shephard (2008b), we smooth the daily beta estimates by passing them through an ARMA(1,1) filter. The figure shows that beta is time-varying and predictable, and that it tends to fluctuate around its mean level. Evidently, the estimated processes exhibit substantial memory with autoregressive roots at 0.90 or higher (see also Andersen, Bollerslev, Diebold, and Wu, 2006, who study the persistence of quarterly realised beta estimated from daily asset returns).

Table 2: Average of high-frequency covariance matrix estimates

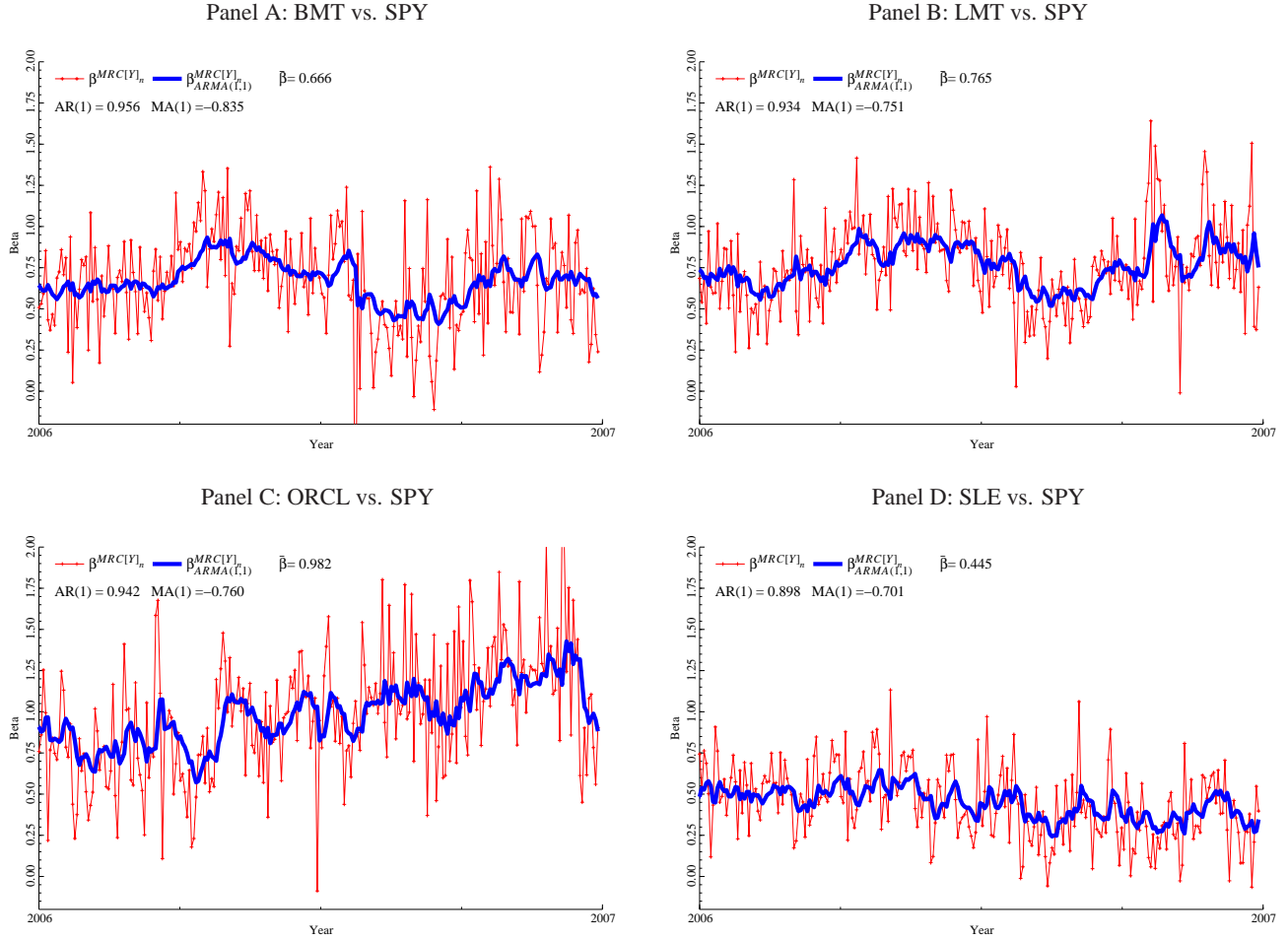
$MRC[Y]_{n=CT(390)}$						$MRC[Y]_{n=CT(390)}^{\delta=0.1}$				
	BMY	LMT	ORCL	SLE	SPY	BMY	LMT	ORCL	SLE	SPY
BMY	1.465	0.203	0.259	0.154	0.233	1.396	0.195	0.259	0.152	0.231
LMT	0.201	0.975	0.232	0.149	0.247	0.194	0.905	0.222	0.158	0.235
ORCL	0.255	0.228	1.804	0.143	0.316	0.256	0.218	1.718	0.151	0.316
SLE	0.144	0.149	0.136	0.955	0.162	0.144	0.156	0.143	0.911	0.167
SPY	0.228	0.249	0.307	0.154	0.317	0.228	0.237	0.309	0.162	0.310
$MRC[Y]_{n=RT}$						$MRC[Y]_{n=RT}^{\delta=0.1}$				
	BMY	LMT	ORCL	SLE	SPY	BMY	LMT	ORCL	SLE	SPY
BMY	1.394	0.197	0.226	0.134	0.220	1.397	0.198	0.249	0.141	0.227
LMT	0.193	0.955	0.224	0.134	0.242	0.201	0.924	0.222	0.147	0.241
ORCL	0.196	0.198	1.756	0.128	0.297	0.228	0.220	1.726	0.139	0.310
SLE	0.117	0.114	0.105	0.878	0.147	0.131	0.134	0.122	0.898	0.157
SPY	0.200	0.237	0.243	0.120	0.310	0.216	0.249	0.285	0.138	0.311
HY						$HY[Y]_n^{(k,l)}$				
	BMY	LMT	ORCL	SLE	SPY	BMY	LMT	ORCL	SLE	SPY
BMY	1.160	0.093	0.097	0.103	0.111	1.417	0.192	0.255	0.144	0.227
LMT	0.049	0.761	0.075	0.072	0.097	0.202	0.902	0.223	0.147	0.238
ORCL	0.032	0.036	1.922	0.077	0.112	0.240	0.227	1.903	0.135	0.309
SLE	0.029	0.026	0.013	1.053	0.095	0.134	0.137	0.127	0.900	0.159
SPY	0.041	0.051	0.040	0.023	0.253	0.219	0.250	0.292	0.143	0.309
Cov^{15s}						$Cov^{avg(5m,15m)}$				
	BMY	LMT	ORCL	SLE	SPY	BMY	LMT	ORCL	SLE	SPY
BMY	1.823	0.088	0.118	0.091	0.120	1.499	0.192	0.288	0.149	0.221
LMT	0.095	0.854	0.084	0.059	0.094	0.191	0.953	0.209	0.152	0.221
ORCL	0.095	0.086	4.043	0.069	0.162	0.261	0.206	1.936	0.153	0.306
SLE	0.056	0.049	0.043	1.947	0.072	0.146	0.148	0.142	1.008	0.161
SPY	0.111	0.118	0.117	0.057	0.293	0.219	0.233	0.290	0.152	0.303

Note. This table reports average covariance matrix estimates. In all subpanels, the numbers in the upper diagonal (including diagonal elements) are based on transaction prices, while the lower diagonal is based on mid-quote data. $Cov^{avg(5m,15m)}$ is a simple time series average of the realised covariance computed from 5- and 15-minute returns.

7 Concluding remarks

In this paper, we outline how simple pre-averaging of financial tick data can be used in the estimation of the ex-post integrated covariance matrix, possibly in the simultaneous presence of market microstructure noise and non-synchronous

Figure 1: MRC-based beta.



Note. $\beta^{MRC[Y]_n}$ is the time series of daily $MRC[Y]_n$ -based beta estimates, using transaction prices and refresh time sampling for various asset combinations (in subpanels). $\beta_{ARMA(1,1)}^{MRC[Y]_n}$ are fitted values from an ARMA(1,1) filter, with estimated autoregressive parameter AR(1) and moving average parameter MA(1). The sample mean MRC beta is reported as $\bar{\beta}$.

trading. We introduce and rigorously study the properties of a modulated realised covariance (MRC) and carry on to show its consistency and asymptotic mixed normality under mild conditions on the dynamics of the price process. This new estimator bears close resemblance to the standard realised covariance, being a sum of outer products of high-frequency returns, but it relies instead on pre-averaging to reduce the impact of microstructure noise. The MRC can be configured to possess an optimal rate of convergence or to guarantee positive semi-definite covariance matrix estimates. In the presence of non-synchronous trading, we show how to modify the MRC by using an imputation scheme (for example previous-tick rule or refresh time sampling) to align high-frequency returns. An MRC constructed on the back of such returns will again be consistent for the integrated covariance. Another novelty in this paper is a pre-averaging extension of the Hayashi and

Yoshida (2005) estimator that can be implemented directly on the raw non-synchronous *and* noisy observations, without any prior alignment of returns. We demonstrate with an empirical illustration that all these new estimators can be applied to real high-frequency data and that they perform well.

Proofs

In the following, we assume that the processes a and σ are bounded. This is without loss of generality and can be justified by a standard localization procedure (see, e.g., Barndorff-Nielsen, Graversen, Jacod, Podolskij, and Shephard (2006)). Moreover, we denote constants by C , or C_p if they depend on an additional parameter p . The main parts of the proofs are based upon Podolskij and Vetter (2009) and Jacod, Li, Mykland, Podolskij, and Vetter (2009).

Proof of Theorem 1: Due to the triangular equality $[V, W] = \frac{1}{4}([V + W, V + W] - [V - W, V - W])$, it suffices to prove the univariate case $d = 1$ (i.e. all processes are 1-dimensional). We use the decomposition

$$MRC[Y]_n = \frac{1}{k_n} \sum_{l=0}^{k_n-1} MRC[Y]_n^l - \frac{\psi_1}{\theta^2 \psi_2} \hat{\Psi}_n, \quad (36)$$

with

$$MRC[Y]_n^l = \frac{1}{\theta \psi_2 \sqrt{n}} \sum_{j=0}^{[n/k_n]-1} |\bar{Y}_{l+jk_n}^n|^2.$$

Notice that, for any $l = 0, \dots, k_n - 1$, the summands in the definition of $MRC[Y]_n^l$ are asymptotically uncorrelated. This type of estimators have been discussed in Podolskij and Vetter (2009) and we can deduce by the methods presented therein (see the proof of Theorem 1) that

$$MRC[Y]_n^l \xrightarrow{p} \int_0^1 \sigma_s^2 ds + \frac{\psi_1}{\psi_2 \theta^2} \Psi,$$

where the convergence holds uniformly in l (due to the boundedness of the processes a and σ). On the other hand we have that

$$\hat{\Psi}_n = \frac{1}{2n} \sum_{i=1}^n |\Delta_i^n Y|^2 \xrightarrow{p} \Psi.$$

This implies the convergence

$$MRC[Y]_n \xrightarrow{p} \int_0^1 \sigma_s^2 ds,$$

which completes the proof. □

Proof of Theorem 2: Here we apply the "big blocks & small blocks"-technique used in Jacod, Li, Mykland, Podolskij, and Vetter (2009). The role of the small blocks (which will be asymptotically negligible) is to ensure the asymptotic independence of the big blocks. More precisely, we choose an integer p , set

$$a_i(p) = i(p+1)k_n \quad \text{and} \quad b_i(p) = i(p+1)k_n + pk_n,$$

and let $A_i(p)$ denote the set of integers l satisfying $a_i(p) \leq l < b_i(p)$ and $B_i(p)$ the integers satisfying $b_i(p) \leq l < a_{i+1}(p)$. We further define $j_n(p)$ to be the largest integer j such that $b_j(p) \leq n$ holds, which gives the identity

$$j_n(p) = \left\lfloor \frac{n}{k_n(p+1)} \right\rfloor - 1. \quad (37)$$

Moreover, we use the notation $i_n(p) = (j_n(p) + 1)(p + 1)k_n$.

Next, we introduce the random variable

$$\bar{Y}_{i,m}^n = \sum_{j=1}^{k_n-1} g\left(\frac{j}{k_n}\right) (\sigma_{\frac{m}{n}} \Delta_{i+j}^n W + \Delta_{i+j}^n \epsilon), \quad (38)$$

which can be interpreted as an approximation of some \bar{Y}_j^n . Moreover, we set

$$\Upsilon_{j,m}^n = \bar{Y}_{i,m}^n (\bar{Y}_{i,m}^n)' - \mathbb{E} \left[\bar{Y}_{i,m}^n (\bar{Y}_{i,m}^n)' | \mathcal{F}_{\frac{m}{n}} \right], \quad (39)$$

and define

$$\tilde{Y}_j^n = \begin{cases} \Upsilon_{j,a_i(p)}^n, & j \in A_i(p) \\ \Upsilon_{j,b_i(p)}^n, & j \in B_i(p) \\ \Upsilon_{j,i_n(p)}^n, & j \geq i_n(p) \end{cases}$$

as well as

$$\zeta(p, 1)_j^n = \sum_{l=a_j(p)}^{b_j(p)-1} \tilde{Y}_l^n, \quad \zeta(p, 2)_j^n = \sum_{l=b_j(p)}^{a_{j+1}(p)-1} \tilde{Y}_l^n.$$

Notice that $\zeta(p, 1)_j^n$ contains pk_n summands ("big block") whereas $\zeta(p, 2)_j^n$ contains k_n summands ("small block"). Finally, we set

$$M(p)^n = n^{-\frac{1}{2}} \sum_{j=0}^{j_n(p)} \zeta(p, 1)_j^n, \quad N(p)^n = n^{-\frac{1}{2}} \sum_{j=0}^{j_n(p)} \zeta(p, 2)_j^n, \quad C(p)^n = n^{-\frac{1}{2}} \sum_{j=i_n(p)}^n \tilde{Y}_j^n$$

and note that

$$\mathbb{E} \left[\zeta(p, 1)_j^n | \mathcal{F}_{\frac{a_j(p)}{n}} \right] = 0 = \mathbb{E} \left[\zeta(p, 2)_j^n | \mathcal{F}_{\frac{b_j(p)}{n}} \right] \quad (40)$$

by construction.

Now, by the same approximations as presented in Jacod, Li, Mykland, Podolskij, and Vetter (2009) (see the identity (5.14), Lemma 5.5 and Lemma 5.6 therein) we obtain that

$$n^{1/4} \left(MRC[Y]_n - \int_0^1 \Sigma_s ds \right) = \frac{n^{\frac{1}{4}}}{\theta \psi_2} (M(p)^n + N(p)^n + C(p)^n) + R(p)_n \quad (41)$$

where the last three summands satisfy the convergence

$$\lim_{p \rightarrow \infty} \limsup_{n \rightarrow \infty} P(|n^{\frac{1}{4}} N(p)^n| + |n^{\frac{1}{4}} C(p)^n| + |R(p)^n| > \delta) = 0 \quad (42)$$

for any $\delta > 0$. Notice that the term $R(p)_n$ stands for the approximation error in Eq. (38).

In the next lemma we show the stable convergence $\frac{n^{\frac{1}{4}}}{\theta \psi_2} M(p)^n \xrightarrow{d_s} U(p)$ (for any fixed p). On the other hand, we will see that, as $p \rightarrow \infty$, $U(p) \xrightarrow{P} U$, where U is the limiting variable defined in Theorem 2. By combining this with Eqs. (41-42) we obtain the assertion of Theorem 2.

Lemma 1 *If the assumptions of Theorem 2 are satisfied we obtain (for any fixed p)*

$$\frac{n^{\frac{1}{4}}}{\theta\psi_2} M(p)^n \xrightarrow{d_s} U(p) = \sum_{j',k'=1}^d \int_0^1 \gamma_s^{jk,j'k'}(p) dB_s^{j'k'},$$

and

$$\begin{aligned} \sum_{j,m=1}^d \gamma_s^{kl,jm}(p) \gamma_s^{k'l',jm}(p) &= A_s^{kl,k'l'} = \frac{2}{\psi_2^2} \left(\frac{\theta p}{p+1} \Lambda_s^{kl,k'l'} \int_0^1 \left(1 - \frac{u}{p}\right) \phi_2^2(u) du \right. \\ &\quad \left. + \frac{p}{\theta(p+1)} \Theta_s^{kl,k'l'} \int_0^1 \left(1 - \frac{u}{p}\right) \phi_1(u) \phi_2(u) du + \frac{p}{\theta^3(p+1)} \Upsilon^{kl,k'l'} \int_0^1 \left(1 - \frac{u}{p}\right) \phi_1^2(u) du \right), \end{aligned}$$

where the processes Λ_s , Θ_s and Υ are given in Theorem 2.

Notice that $\sum_{j,m=1}^d \gamma_s^{kl,jm}(p) \gamma_s^{k'l',jm}(p) \xrightarrow{p} \sum_{j,m=1}^d \gamma_s^{kl,jm} \gamma_s^{k'l',jm}$ ($1 \leq k, k', l, l' \leq d$), where γ_s is defined in Theorem 2. From this we deduce the convergence $U(p) \xrightarrow{p} U$.

Proof of Lemma 1: Due to Theorem IX 7.28 in Jacod and Shiryaev (2003) the following conditions need to be shown (for all $1 \leq k, k', l, l' \leq d$)

$$\frac{n^{-1/2}}{\theta^2 \psi_2^2} \sum_{j=0}^{j_n(p)} \mathbb{E} \left[\zeta(p, 1)_j^{n,kl} \zeta(p, 1)_j^{n,k'l'} | \mathcal{F}_{\frac{a_j(p)}{n}} \right] \xrightarrow{p} \int_0^1 A_u^{kl,k'l'} du, \quad (43)$$

$$n^{-1} \sum_{j=0}^{j_n(p)} \mathbb{E} \left[||\zeta(p, 1)_j^n||^4 | \mathcal{F}_{\frac{a_j(p)}{n}} \right] \xrightarrow{p} 0, \quad (44)$$

$$n^{-1/4} \sum_{j=0}^{j_n(p)} \mathbb{E} \left[\zeta(p, 1)_j^{n,kl} \Delta W(p)_j^{n,k'} | \mathcal{F}_{\frac{a_j(p)}{n}} \right] \xrightarrow{p} 0, \quad (45)$$

$$n^{-1/4} \sum_{j=0}^{j_n(p)} \mathbb{E} \left[\zeta(p, 1)_j^{n,kl} \Delta N(p)_j^n | \mathcal{F}_{\frac{a_j(p)}{n}} \right] \xrightarrow{p} 0, \quad (46)$$

where $\Delta V(p)_j^n = V_{n/b_j(p)} - V_{n/a_j(p)}$ for any process V and Eq. (46) holding for any 1-dimensional bounded martingale N being orthogonal to W . For proving Eqs. (44) and (46), it is no restriction to assume that $d = 1$. Then these conditions are already shown in Jacod, Li, Mykland, Podolskij, and Vetter (2009) (Lemma 5.7). On the other hand, the functional $\zeta(p, 1)_j^n$ is even in W . Since W and ϵ are independent, we readily deduce that

$$\mathbb{E} \left[\zeta(p, 1)_j^{n,kl} \Delta W(p)_j^{n,k'} | \mathcal{F}_{\frac{a_j(p)}{n}} \right] = 0,$$

which implies the condition in Eq. (45). Hence, we are left to proving Eq. (43).

First, notice the identity

$$\bar{V}_i^n = \sum_{j=1}^{k_n} g\left(\frac{j}{k_n}\right) \Delta_{i+j}^n V = - \sum_{j=0}^{k_n-1} \left(g\left(\frac{j+1}{k_n}\right) - g\left(\frac{j}{k_n}\right) \right) V_{\frac{i+j}{n}}.$$

The second equality is useful for the computation of the moments of $\bar{\epsilon}_l^n$. By the smoothness assumption on the function g and the above identity we obtain the approximations ($1 \leq k, l \leq d$)

$$\mathbb{E}[\bar{W}_j^{n,k} \bar{W}_{j'}^{n,l}] = \delta_{kl} \frac{k_n}{n} \psi_2\left(\frac{|j-j'|}{k_n}\right) + O(n^{-1}), \quad \mathbb{E}[\bar{\epsilon}_j^{n,k} \bar{\epsilon}_{j'}^{n,l}] = \frac{\Psi^{kl}}{k_n} \psi_1\left(\frac{|j-j'|}{k_n}\right) + O(n^{-1}) \quad (47)$$

for $|j-j'| < k_n$, whereas the above expectations vanish when $|j-j'| \geq k_n$ (here δ_{kl} denotes the Kronecker symbol). Next, we introduce the decomposition

$$\zeta(p, 1)_j^n = v(p, 1)_j^n + v(p, 2)_j^n + v(p, 3)_j^n,$$

where the terms $v(p, 1)_j^n$, $v(p, 2)_j^n$ and $v(p, 3)_j^n$ are given by

$$\begin{aligned} v(p, 1)_j^n &= \sum_{l=a_j(p)}^{b_j(p)-1} \sigma_{\frac{a_j(p)}{n}} \bar{W}_l^n \left(\sigma_{\frac{a_j(p)}{n}} \bar{W}_l^n \right)' - \mathbb{E} \left[\sigma_{\frac{a_j(p)}{n}} \bar{W}_l^n \left(\sigma_{\frac{a_j(p)}{n}} \bar{W}_l^n \right)' \middle| \mathcal{F}_{\frac{a_j(p)}{n}} \right], \\ v(p, 2)_j^n &= \sum_{l=a_j(p)}^{b_j(p)-1} \bar{\epsilon}_l^n \left(\bar{\epsilon}_l^n \right)' - \mathbb{E} \left[\bar{\epsilon}_l^n \left(\bar{\epsilon}_l^n \right)' \right], \\ v(p, 3)_j^n &= \sum_{l=a_j(p)}^{b_j(p)-1} \sigma_{\frac{a_j(p)}{n}} \bar{W}_l^n \left(\bar{\epsilon}_l^n \right)' + \bar{\epsilon}_l^n \left(\sigma_{\frac{a_j(p)}{n}} \bar{W}_l^n \right)'. \end{aligned}$$

By a straightforward calculation (and Eq. (47)) we obtain for all $1 \leq k, l, k'l' \leq d$

$$\begin{aligned} \mathbb{E}[v(p, 1)_j^{n,kl} v(p, 1)_j^{n,k'l'} | \mathcal{F}_{\frac{a_j(p)}{n}}] &= \frac{2pk_n^4}{n^2} \Lambda_{\frac{a_j(p)}{n}}^{kl,k'l'} \int_0^1 \left(1 - \frac{u}{p}\right) \phi_2^2(u) du + o_p(1), \\ \mathbb{E}[v(p, 2)_j^{n,kl} v(p, 2)_j^{n,k'l'} | \mathcal{F}_{\frac{a_j(p)}{n}}] &= 2p \Upsilon^{kl,k'l'} \int_0^1 \left(1 - \frac{u}{p}\right) \phi_1^2(u) du + o_p(1), \\ \mathbb{E}[v(p, 3)_j^{n,kl} v(p, 3)_j^{n,k'l'} | \mathcal{F}_{\frac{a_j(p)}{n}}] &= \frac{2pk_n^2}{n} \Theta_{\frac{a_j(p)}{n}}^{kl,k'l'} \int_0^1 \left(1 - \frac{u}{p}\right) \phi_1(u) \phi_2(u) du + o_p(1), \end{aligned}$$

where the approximation holds uniformly in j . Now recall that $j_n(p) = \left\lfloor \frac{n}{k_n(p+1)} \right\rfloor - 1$. Consequently, by Riemann integrability we deduce that

$$\frac{n^{-\frac{1}{2}}}{\theta^2 \psi_2^2} \sum_{j=0}^{j_n(p)} \mathbb{E}[\zeta(p, 1)_j^{n,kl} \zeta(p, 1)_j^{n,k'l'} | \mathcal{F}_{\frac{a_j(p)}{n}}] \xrightarrow{p} \int_0^1 A_u^{kl,k'l'} du,$$

which completes the proof of Lemma 1. □

Proof of Theorem 3 and 4: Recall that $\frac{k_n}{n^{1/2+\delta}} = \theta + o(n^{-\frac{1}{4}+\frac{\delta}{2}})$. A straightforward calculation shows that

$$\mathbb{E} \left[MRC[\epsilon]_n^\delta \right] = \frac{\psi_1}{\theta^2 \psi_2 n^{2\delta}} \Psi + o(n^{-\frac{1}{4}+\frac{\delta}{2}}).$$

By similar methods as presented in the proof of Theorem 1 we deduce that

$$MRC[Y]_n^\delta - \left(\int_0^1 \Sigma_s ds + \frac{\psi_1}{\theta^2 \psi_2 n^{2\delta}} \Psi \right) \xrightarrow{p} 0.$$

Hence, we obtain the convergence in probability

$$MRC[Y]_n^\delta \xrightarrow{P} \int_0^1 \Sigma_s ds,$$

which implies the assertion of Theorem 3. Following the same scheme as demonstrated in the proof of Theorem 2 we get

$$n^{1/4-\delta/2} \left(MRC[Y]_n^\delta - \left(\int_0^1 \Sigma_s ds + \frac{\psi_1}{\theta^2 \psi_2 n^{2\delta}} \Psi \right) \right) \xrightarrow{d_3} MN \left(0, \frac{2\Phi_{22}\theta}{\psi_2^2} \int_0^1 \Lambda_s ds \right),$$

since $\frac{\psi_1}{\theta^2 \psi_2 n^{2\delta}} \Psi$ is an appropriate centering for the noise term in this case. Now

$$\frac{\psi_1 n^{1/4-\delta/2}}{\theta^2 \psi_2 n^{2\delta}} \Psi \xrightarrow{P} 0$$

for $\delta > 1/10$, whereas

$$\frac{\psi_1 n^{1/4-\delta/2}}{\theta^2 \psi_2 n^{2\delta}} \Psi = \frac{\psi_1}{\theta^2 \psi_2} \Psi$$

for $\delta = 1/10$. Hence, Theorem 4 follows. \square

Proof of Theorem 5: First, we start with the decomposition

$$\begin{aligned} HY[Y]_n^{(k,l)} &= \frac{1}{(\psi_{HY} k_n)^2} \left(\sum_{i,j} \bar{X}_i^{k_n} \bar{X}_j^{l_n} \mathbb{I}_{\{(t_{i-1}^{(k)}, t_i^{(k)}) \cap (t_{j-1}^{(l)}, t_j^{(l)}) \neq \emptyset\}} \right. \\ &\quad + \sum_{i,j} (\bar{X}_i^{k_n} \bar{\epsilon}_j^{l_n} + \bar{\epsilon}_i^{k_n} \bar{X}_j^{l_n}) \mathbb{I}_{\{(t_{i-1}^{(k)}, t_i^{(k)}) \cap (t_{j-1}^{(l)}, t_j^{(l)}) \neq \emptyset\}} \\ &\quad \left. + \sum_{i,j} \bar{\epsilon}_i^{k_n} \bar{\epsilon}_j^{l_n} \mathbb{I}_{\{(t_{i-1}^{(k)}, t_i^{(k)}) \cap (t_{j-1}^{(l)}, t_j^{(l)}) \neq \emptyset\}} \right) =: HY[Y]_n^1 + HY[Y]_n^2 + HY[Y]_n^3. \end{aligned}$$

As X and ϵ are independent, it follows that $\mathbb{E}(HY[Y]_n^2) = 0$, and a simple computation shows that

$$\text{var}(HY[Y]_n^2) = 0.$$

Thus, $HY[Y]_n^2 \xrightarrow{P} 0$. Next, we consider the term $HY[Y]_n^3$. This expression can be further decomposed as

$$HY[Y]_n^3 = \frac{1}{(\psi_{HY} k_n)^2} \left(\sum_{t_i \in J_{k,l}} a_i^n(k, l) \epsilon_{t_i}^k \epsilon_{t_i}^l + \sum_{t_i \in J_{k,l}^c} b_i^n(k, l) \epsilon_{t_i}^k \epsilon_{t_i}^l + \sum_{|i-j| < k_n} c_{ij}^n(k, l) \epsilon_{t_i}^k \epsilon_{t_j}^l 1_{\{t_i^{(k)} \neq t_j^{(l)}\}} \right),$$

for certain numbers $a_i^n(k, l)$, $b_i^n(k, l)$ and $c_{ij}^n(k, l)$. Here $J_{k,l}$ denotes the set of common points of $(t_i^{(k)})_{k_n \leq i \leq n_k - k_n}$ and $(t_i^{(l)})_{k_n \leq i \leq n_l - k_n}$, and $J_{k,l}^c$ denotes the set of all common points of $(t_i^{(k)})_{1 \leq i \leq n_k}$ and $(t_i^{(l)})_{1 \leq i \leq n_l}$ excluded $J_{k,l}$. Since $\bar{\epsilon}_i^{k_n} = -\sum_{j=1}^{k_n} \left(g\left(\frac{j}{k_n}\right) - g\left(\frac{j-1}{k_n}\right) \right) \epsilon_{t_{i+j}}^k$ and g is piecewise differentiable it holds that

$$|b_i^n(k, l)| + |c_{ij}^n(k, l)| \leq C.$$

A straightforward computation shows that

$$a_i^n(k, l) = \left(\sum_{j=1}^{k_n} g\left(\frac{j}{k_n}\right) - g\left(\frac{j-1}{k_n}\right) \right)^2 = (g(1) - g(0))^2 = 0,$$

because $g(0) = g(1) = 0$. Thus, the first summand in the decomposition of $HY[Y]_n^3$ disappears, which is absolutely crucial for the proof. On the other hand, $\#J_{k,l}^c \leq Ck_n$ which means that

$$\frac{1}{(\psi_{HY}k_n)^2} \sum_{t_i \in J_{k,l}^c} b_i^n(k, l) \epsilon_{t_i}^k \epsilon_{t_i}^l \xrightarrow{p} 0.$$

Finally, note that the summands $\epsilon_{t_i}^k \epsilon_{t_j}^l \mathbb{I}_{\{t_i^{(k)} \neq t_j^{(l)}\}}$ have expectation 0 and are mutually uncorrelated. This implies that

$$\frac{1}{(\psi_{HY}k_n)^2} \sum_{|i-j| < k_n} c_{ij}^n(k, l) \epsilon_{t_i}^k \epsilon_{t_j}^l \mathbb{I}_{\{t_i^{(k)} \neq t_j^{(l)}\}} \xrightarrow{p} 0.$$

Hence,

$$HY[Y]_n^3 \xrightarrow{p} 0.$$

Now we consider the term $HY[Y]_n^1$. We decompose

$$\begin{aligned} HY[Y]_n^1 &= \frac{1}{(\psi_{HY}k_n)^2} \left(\sum_{(i,j) \in I_{k,l}} \bar{a}_{ij}^n(k, l) \Delta_i^{n_k} X^k \Delta_j^{n_l} X^l \mathbb{I}_{\{(t_{i-1}^{(k)}, t_i^{(k)}) \cap (t_{j-1}^{(l)}, t_j^{(l)}) \neq \emptyset\}} \right. \\ &\quad + \sum_{(i,j) \in I_{k,l}^c} \bar{b}_{ij}^n(k, l) \Delta_i^{n_k} X^k \Delta_j^{n_l} X^l \mathbb{I}_{\{(t_{i-1}^{(k)}, t_i^{(k)}) \cap (t_{j-1}^{(l)}, t_j^{(l)}) \neq \emptyset\}} \\ &\quad \left. + \sum_{|i-j| < k_n} \bar{c}_{ij}^n(k, l) \Delta_i^{n_k} X^k \Delta_j^{n_l} X^l \mathbb{I}_{\{(t_{i-1}^{(k)}, t_i^{(k)}) \cap (t_{j-1}^{(l)}, t_j^{(l)}) = \emptyset\}} \right), \end{aligned} \quad (48)$$

for some constants $\bar{a}_{ij}^n(k, l)$, $\bar{b}_{ij}^n(k, l)$, $\bar{c}_{ij}^n(k, l)$, $I_{k,l} = \{(i, j) : k_n \leq i \leq n_k - k_n, k_n \leq j \leq n_l - k_n\}$ and $I_{k,l}^c = \{(i, j) : 1 \leq i \leq n_k, 1 \leq j \leq n_l\} - I_{k,l}$. Notice that all "border terms" are collected in the second summand whereas all terms with empty intersection of the intervals are in the third summand (in fact, we will see that both are negligible).

Notice that

$$|a_{ij}^n(k, l)| + |b_{ij}^n(k, l)| + |c_{ij}^n(k, l)| \leq Cn.$$

Furthermore, we have

$$\mathbb{E} [|\Delta_i^{n_k} X^k \Delta_j^{n_l} X^l|] \leq C \sqrt{(t_i^{(k)} - t_{i-1}^{(k)})(t_j^{(l)} - t_{j-1}^{(l)})}$$

and $\#(I_{k,l}^c \cap \{(i, j) : (t_{i-1}^{(k)}, t_i^{(k)}) \cap (t_{j-1}^{(l)}, t_j^{(l)}) \neq \emptyset\}) \leq Ck_n$ by Eq. (25). This implies by Eq. (26) that

$$\frac{1}{(\psi_{HY}k_n)^2} \sum_{(i,j) \in I_{k,l}^c} \bar{b}_{ij}^n(k, l) \Delta_i^{n_k} X^k \Delta_j^{n_l} X^l \mathbb{I}_{\{(t_{i-1}^{(k)}, t_i^{(k)}) \cap (t_{j-1}^{(l)}, t_j^{(l)}) \neq \emptyset\}} = O_p(n^{-1/2}),$$

and thus the second summand in Eq. (48) is negligible.

Now, recall that $\Delta_i^{n_k} X^k$ can be replaced with $(\sigma_s \Delta_i^{n_k} W)^k$ for any s with $t_i^{(k)} - s = O(k_n/n)$ ($1 \leq k \leq d$) without changing the first-order asymptotics (see, e.g., Podolskij and Vetter, 2009). Notice also that the terms $\Delta_i^{n_k} W^k \Delta_j^{n_l} W^l \mathbb{I}_{\{(t_{i-1}^{(k)}, t_i^{(k)}) \cap (t_{j-1}^{(l)}, t_j^{(l)}) = \emptyset\}}$ are mutually uncorrelated. Hence,

$$\frac{1}{(\psi_{HY}k_n)^2} \sum_{|i-j| < k_n} \bar{c}_{ij}^n(k, l) \Delta_i^{n_k} X^k \Delta_j^{n_l} X^l \mathbb{I}_{\{(t_{i-1}^{(k)}, t_i^{(k)}) \cap (t_{j-1}^{(l)}, t_j^{(l)}) = \emptyset\}} = o_p(1).$$

Finally, consider the first summand in Eq. (48). As above, we approximate $\Delta_i^{n_k} X^k \Delta_j^{n_l} X^l$ by

$$X_{i,j}^n(k, l) = (\sigma_{t_{i-1}^{(k)} \wedge t_{j-1}^{(l)}} \Delta_i^{n_k} W)^k (\sigma_{t_{i-1}^{(k)} \wedge t_{j-1}^{(l)}} \Delta_i^{n_l} W)^l$$

whenever $(t_{i-1}^{(k)}, t_i^{(k)}) \cap (t_{j-1}^{(l)}, t_j^{(l)}) \neq \emptyset$, and set

$$\overline{HY}[X]_n^{(k,l)} = \frac{1}{(\psi_{HY} k_n)^2} \sum_{(i,j) \in I_{k,l}} \bar{a}_{ij}^n(k, l) X_{i,j}^n(k, l) \mathbb{I}_{\{(t_{i-1}^{(k)}, t_i^{(k)}) \cap (t_{j-1}^{(l)}, t_j^{(l)}) \neq \emptyset\}}.$$

Note that the quantities $HY[Y]_n^{(k,l)}$ and $\overline{HY}[X]_n^{(k,l)}$ have the same first order asymptotics. Then, it can be shown that

$$\bar{a}_{ij}^n(k, l) = \left(\sum_{h=1}^{k_n-1} g\left(\frac{h}{k_n}\right) \right)^2 = k_n^2 \left(\int_0^1 g(x) dx \right)^2 + o(k_n^2).$$

Next, note that

$$\mathbb{E} \left[\Delta_i^{n_k} W^k \Delta_i^{n_l} W^l \mathbb{I}_{\{(t_{i-1}^{(k)}, t_i^{(k)}) \cap (t_{j-1}^{(l)}, t_j^{(l)}) \neq \emptyset\}} | \mathcal{F}_{t_{i-1}^{(k)} \wedge t_{j-1}^{(l)}} \right] = (t_i^{(k)} \wedge t_j^{(l)}) - (t_{i-1}^{(k)} \vee t_{j-1}^{(l)}). \quad (49)$$

Eq. (49) and Riemann integrability then delivers the convergence

$$\frac{1}{(\psi_{HY} k_n)^2} \sum_{(i,j) \in I_{k,l}} \bar{a}_{ij}^n(k, l) \mathbb{E} \left[X_{i,j}^n(k, l) \mathbb{I}_{\{(t_{i-1}^{(k)}, t_i^{(k)}) \cap (t_{j-1}^{(l)}, t_j^{(l)}) \neq \emptyset\}} | \mathcal{F}_{t_{i-1}^{(k)} \wedge t_{j-1}^{(l)}} \right] \xrightarrow{P} \int_0^1 \Sigma_s^{kl} ds.$$

By usual martingale arguments we have that

$$\overline{HY}[X]_n^{(k,l)} - \frac{1}{(\psi_{HY} k_n)^2} \sum_{(i,j) \in I_{k,l}} \bar{a}_{ij}^n(k, l) \mathbb{E} \left[X_{i,j}^n(k, l) \mathbb{I}_{\{(t_{i-1}^{(k)}, t_i^{(k)}) \cap (t_{j-1}^{(l)}, t_j^{(l)}) \neq \emptyset\}} | \mathcal{F}_{t_{i-1}^{(k)} \wedge t_{j-1}^{(l)}} \right] \xrightarrow{P} 0.$$

On the other hand, $HY[Y]_n^{(k,l)} - \overline{HY}[X]_n^{(k,l)} \xrightarrow{P} 0$. Thus, collecting terms produces

$$HY[Y]_n^{(k,l)} \xrightarrow{P} \int_0^1 \Sigma_s^{kl} ds,$$

which completes the proof.

References

- Ait-Sahalia, Y., P. A. Mykland, and L. Zhang, 2005, “How often to sample a continuous-time process in the presence of market microstructure noise,” *Review of Financial Studies*, 18(2), 351–416.
- Aldous, D. J., and G. K. Eagleson, 1978, “On mixing and stability of limit theorems,” *Annals of Probability*, 6(2), 325–331.
- Andersen, T. G., T. Bollerslev, and F. X. Diebold, 2009, “Parametric and nonparametric volatility measurement,” in *Handbook of Financial Econometrics*, ed. by L. P. Hansen, and Y. Ait-Sahalia. North-Holland, Forthcoming.
- Andersen, T. G., T. Bollerslev, F. X. Diebold, and P. Labys, 2001, “The distribution of realized exchange rate volatility,” *Journal of the American Statistical Association*, 96(453), 42–55.
- , 2003, “Modeling and forecasting realized volatility,” *Econometrica*, 71(2), 579–625.

- Andersen, T. G., T. Bollerslev, F. X. Diebold, and J. Wu, 2006, "Realized beta: Persistence and predictability," in *Advances in Econometrics: Econometric Analysis of Economic and Financial Time Series in Honor of R. F. Engle and C. W. J. Granger*, ed. by T. Fomby, and D. Terrell. pp. 1–40, Volume B.
- Back, K., 1991, "Asset prices for general processes," *Journal of Mathematical Economics*, 20(4), 371–395.
- Bandi, F. M., and J. R. Russell, 2005, "Realized covariation, realized beta and microstructure noise," Working paper, Graduate School of Business, University of Chicago.
- , 2006, "Separating microstructure noise from volatility," *Journal of Financial Economics*, 79(3), 655–692.
- , 2008, "Microstructure noise, realized variance, and optimal sampling," *Review of Economic Studies*, 75(2), 339–369.
- Barndorff-Nielsen, O. E., S. E. Graversen, J. Jacod, M. Podolskij, and N. Shephard, 2006, "A central limit theorem for realized power and bipower variations of continuous semimartingales," in *From Stochastic Calculus to Mathematical Finance: The Shiryaev Festschrift*, ed. by Y. Kabanov, R. Lipster, and J. Stoyanov. Springer-Verlag, pp. 33–68.
- Barndorff-Nielsen, O. E., P. R. Hansen, A. Lunde, and N. Shephard, 2008a, "Designing realised kernels to measure the ex-post variation of equity prices in the presence of noise," *Econometrica*, 76(6), 1481–1536.
- , 2008b, "Multivariate realised kernels: Consistent positive semi-definite estimators of the covariation of equity prices with noise and non-synchronous trading," Working paper, University of Aarhus.
- Barndorff-Nielsen, O. E., and N. Shephard, 2002, "Econometric analysis of realized volatility and its use in estimating stochastic volatility models," *Journal of the Royal Statistical Society: Series B*, 64(2), 253–280.
- , 2004, "Econometric analysis of realized covariation: High frequency based covariance, regression, and correlation in financial economics," *Econometrica*, 72(3), 885–925.
- , 2007, "Variation, jumps, market frictions and high frequency data in financial econometrics," in *Advances in Economics and Econometrics: Theory and Applications, Ninth World Congress*, ed. by R. Blundell, P. Torsten, and W. K. Newey. Cambridge University Press, Forthcoming.
- Boudt, K., C. Croux, and S. Laurent, 2008, "Outlyingness weighted quadratic covariation," Working paper, K. U. Leuven.
- Christensen, K., R. C. A. Oomen, and M. Podolskij, 2009, "Realised quantile-based estimation of the integrated variance," Working paper, available at <http://ssrn.com/abstract=1085553>.
- Corsi, F., and F. Audrino, 2007, "Realised correlation tick-by-tick," Working paper, available at <http://ssrn.com/abstract=957997>.
- de Jong, F., and T. Nijman, 1997, "High frequency analysis of lead-lag relationships between financial markets," *Journal of Empirical Finance*, 4(2-3), 259–277.
- Delbaen, F., and W. Schachermayer, 1994, "A general version of the fundamental theorem of asset pricing," *Mathematische Annalen*, 300, 463–520.
- Diebold, F. X., and G. H. Strasser, 2008, "On the correlation structure of microstructure noise in theory and practice," Working paper, University of Pennsylvania.
- Dimson, E., 1979, "Risk measurement when shares are subject to infrequent trading," *Journal of Financial Economics*, 7(2), 197–226.
- Epps, T. W., 1979, "Comovements in stock prices in the very short run," *Journal of the American Statistical Association*, 74(366), 291–298.

- Fan, J., J. Zhang, and K. Yu, 2009, "Asset allocation and risk assessment with gross exposure constraints for vast portfolios," Working paper, Princeton University.
- Fisher, L., 1966, "Some new stock-market indexes," *Journal of Business*, 39(1-2), 191–225.
- Ghysels, E., A. C. Harvey, and E. Renault, 1996, "Stochastic volatility," in *Handbook of Statistics: Volume 14*, ed. by G. S. Maddala, and C. R. Rao. North-Holland, pp. 119–191.
- Gloter, A., and J. Jacod, 2001a, "Diffusions with measurement errors. I - local asymptotic normality," *ESAIM: Probability and Statistics*, 5, 225–242.
- , 2001b, "Diffusions with measurement errors. II - measurement errors," *ESAIM: Probability and Statistics*, 5, 243–260.
- Griffin, J. E., and R. C. A. Oomen, 2006, "Covariance measurement in the presence of non-synchronous trading and market microstructure noise," Working paper, University of Warwick.
- Hansen, P. R., and A. Lunde, 2006, "Realized variance and market microstructure noise," *Journal of Business and Economic Statistics*, 24(2), 127–161.
- Hayashi, T., and N. Yoshida, 2005, "On covariance estimation of non-synchronously observed diffusion processes," *Bernoulli*, 11(2), 359–379.
- , 2008, "Asymptotic normality of a covariance estimator for nonsynchronously observed diffusion processes," *Annals of the Institute of Statistical Mathematics*, 60(2), 367–406.
- Jacod, J., Y. Li, P. A. Mykland, M. Podolskij, and M. Vetter, 2009, "Microstructure noise in the continuous case: The pre-averaging approach," *Stochastic Processes and their Applications*, 119(7), 2249–2276.
- Jacod, J., and A. N. Shiryaev, 2003, *Limit Theorems for Stochastic Processes*. Springer-Verlag, 2 edn.
- Kinnebrock, S., and M. Podolskij, 2008, "An econometric analysis of modulated realised covariance, regression and correlation in noisy diffusion models," Working paper, CREATES, University of Aarhus.
- Large, J., 2007, "Accounting for the Epps effect: Realized covariation, cointegration and common factors," Working paper, Oxford-Man Institute, University of Oxford.
- Lo, A., and C. MacKinlay, 1990, "An econometric analysis of nonsynchronous trading," *Journal of Econometrics*, 45(1-2), 181–211.
- Malliavin, P., and M. E. Mancino, 2002, "Fourier series method for measurement of multivariate volatilities," *Finance and Stochastics*, 6(1), 49–61.
- Martens, M., 2003, "Estimating unbiased and precise realized covariances," Working paper, Erasmus School of Economics, Rotterdam.
- Mykland, P. A., and L. Zhang, 2006, "ANOVA for diffusions and Itô processes," *Annals of Statistics*, 34(4), 1931–1963.
- , 2008, "Inference for continuous semimartingales observed at high frequency: A general approach," Working paper, University of Chicago.
- Oomen, R. C. A., 2006, "Properties of realized variance under alternative sampling schemes," *Journal of Business and Economic Statistics*, 24(2), 219–237.
- Palandri, A., 2006, "Consistent realized covariance for asynchronous observations contaminated by market microstructure noise," Working paper, University of Copenhagen.

- Podolskij, M., and M. Vetter, 2009, “Estimation of volatility functionals in the simultaneous presence of microstructure noise and jumps,” *Bernoulli*, 15(3), 634–658.
- Renò, R., 2003, “A closer look at the Epps effect,” *International Journal of Theoretical and Applied Finance*, 6(1), 87–102.
- Rényi, A., 1963, “On stable sequences of events,” *Sankhya: The Indian Journal of Statistics; Series A*, 25(3), 293–302.
- Scholes, M., and J. Williams, 1977, “Estimating betas from nonsynchronous data,” *Journal of Financial Economics*, 5(3), 309–327.
- Voev, V., and A. Lunde, 2007, “Integrated covariance estimation using high-frequency data in the presence of noise,” *Journal of Financial Econometrics*, 5(1), 68–104.
- Zhang, L., 2006, “Efficient estimation of stochastic volatility using noisy observations: A multi-scale approach,” *Bernoulli*, 12(6), 1019–1043.
- , 2008, “Estimating covariation: Epps effect, microstructure noise,” Working paper, University of Illinois at Chicago.
- Zhang, L., P. A. Mykland, and 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.
- Zhou, B., 1996, “High-frequency data and volatility in foreign-exchange rates,” *Journal of Business and Economic Statistics*, 14(1), 45–52.

- 2009-32: Alessandro Palandri: The Effects of Interest Rate Movements on Assets' Conditional Second Moments
- 2009-33: Peter Christoffersen, Redouane Elkamhi, Bruno Feunou and Kris Jacobs: Option Valuation with Conditional Heteroskedasticity and Non-Normality
- 2009-34: Peter Christoffersen, Steven Heston and Kris Jacobs: The Shape and Term Structure of the Index Option Smirk: Why Multifactor Stochastic Volatility Models Work so Well
- 2009-35: Peter Christoffersen, Jeremy Berkowitz and Denis Pelletier: Evaluating Value-at-Risk Models with Desk-Level Data
- 2009-36: Tom Engsted and Thomas Q. Pedersen: The dividend-price ratio does predict dividend growth: International evidence
- 2009-37: Michael Jansson and Morten Ørregaard Nielsen: Nearly Efficient Likelihood Ratio Tests of the Unit Root Hypothesis
- 2009-38: Frank S. Nielsen: Local Whittle estimation of multivariate fractionally integrated processes
- 2009-39: Borus Jungbacker, Siem Jan Koopman and Michel van der Wel: Dynamic Factor Models with Smooth Loadings for Analyzing the Term Structure of Interest Rates
- 2009-40: Niels Haldrup, Antonio Montañés and Andreu Sansó: Detection of additive outliers in seasonal time series
- 2009-41: Dennis Kristensen: Pseudo-Maximum Likelihood Estimation in Two Classes of Semiparametric Diffusion Models
- 2009-42: Ole Eiler Barndorff-Nielsen and Robert Stelzer: The multivariate supOU stochastic volatility model
- 2009-43: Lasse Bork, Hans Dewachter and Romain Houssa: Identification of Macroeconomic Factors in Large Panels
- 2009-44: Dennis Kristensen: Semiparametric Modelling and Estimation: A Selective Overview
- 2009-45: Kim Christensen, Silja Kinnebrock and Mark Podolskij: Pre-averaging estimators of the ex-post covariance matrix