

# Efficient Covariance Estimation for Asynchronous Noisy High-Frequency Data

Markus Bibinger <sup>1</sup>

Interdisciplinary Center for Scientific Computing  
of the Ruprecht-Karls-University of Heidelberg

and

Department of Mathematics  
Humboldt University of Berlin

December 2008

## Abstract

We focus on estimating the integrated covariance of log-price processes in the presence of market microstructure noise. We construct an efficient unbiased estimator for the quadratic covariation of two Itô processes in the case where high-frequency asynchronous discrete returns under market microstructure noise are observed. This estimator is based on synchronization and multi-scale methods and attains the optimal rate of convergence. A Monte Carlo study analyzes the finite sample size characteristics of our estimator.

**Key words:** Quadratic covariation estimator, Asynchronous observations, Market microstructure noise, Subsampling, Multi-scale estimator, Optimal rate

**MSC classes:** 62F12, 62G05

---

<sup>1</sup>E-mail: bibinger@math.hu-berlin.de

# 1 Introduction

Estimating the quadratic covariation, also called integrated covariance, of asset returns is a central theme in finance. With the availability of high-frequency intraday returns the estimation of daily integrated covariances using high-frequency observations became an issue of great interest. The problems occurring in covariance estimation using high-frequency data are mainly the lack of synchronicity and market microstructure noise.

In this article we propose a new estimator for the integrated covariance  $\langle \tilde{X}, \tilde{Y} \rangle_T$  of two log-price processes over a fixed time horizon  $[0, T]$  (usually one trading day) when we observe high-frequent noisy asynchronous data. The problem of asynchronous data without noise was solved by Hayashi and Yoshida (2005) and there are as well estimators developed in recent literature that solve the problem of noisy but synchronous data (see e. g. Barndorff-Nielsen and Shephard (2004)).

We work within the model where the efficient asset processes  $\tilde{X}$  and  $\tilde{Y}$  (without noise) are assumed to be Itô processes

$$\begin{aligned} d\tilde{X}_t &= \mu_t^X dt + \sigma_t^X dB_t^X, \\ d\tilde{Y}_t &= \mu_t^Y dt + \sigma_t^Y dB_t^Y, \quad t \in [0, T] \end{aligned}$$

with Brownian motions  $B^X$  and  $B^Y$  which are correlated with  $\text{corr}(B_t^X, B_t^Y) = \rho_t$  and continuous, bounded and adapted stochastic processes  $\mu_t^X, \mu_t^Y, \sigma_t^X, \sigma_t^Y$ . It is well known that for synchronous observations without noise the realized covariance  $\sum_{t_i+1 \leq T} (X_{t_{i+1}} - X_{t_i})(Y_{t_{i+1}} - Y_{t_i})$  is a consistent estimator for  $\langle \tilde{X}, \tilde{Y} \rangle_T = \int_0^T \rho_t \sigma_t^X \sigma_t^Y dt$  if  $\sup_i (t_{i+1} - t_i) \rightarrow 0$ .

In the case of high-frequency data, e. g. tick-by-tick data, the observations of the asset processes are usually not simultaneous and estimation methods are hence based on synchronization of the data by interpolation (e. g. linear or previous-tick interpolation) before calculating a realized covariance estimator.

We use the construction of a synchronized time grid for two processes following the method proposed by Palandri (2006). The realized covariance calculated with the synchronized observations corresponds to the Hayashi-Yoshida estimator given by the sum of all products of increments with overlapping time intervals. This estimator is proved to be a consistent estimator in the absence of noise (Hayashi and Yoshida (2005)) but becomes inconsistent when market microstructure effects are relevant. The behaviour of this estimator and the realized covariance depending on the sample frequencies are studied in Hansen et al. (2007) and Griffin and Oomen (2006).

Market microstructure effects and estimators for the integrated volatility under its influence were studied intensively in recent literature. Estimators with an optimal rate of convergence  $N^{1/4}$ , where  $N$  denotes the number of observations, are presented by Zhang (2006a) and Barndorff-Nielsen and Shephard (2006). Merging those techniques for asynchronous data and the subsampling approach to high-frequency observations contaminated by market microstructure noise as presented in Zhang et al. (2005) a consistent estimator for asynchronous noisy data can be achieved. By construction of an adequate synchronized time-scale for two asset processes and subsampling an estimator with  $N^{1/6}$ -rate of convergence can be obtained, where  $N$  denotes the number of synchronized observations in this context, which is less than or equal to the minimum of observations of  $X$  and  $Y$ . This result is presented in Palandri (2006). We use the same methods to rearrange the observations in a synchronized grid and show how an extension of subsampling to a multi-scale approach can afford an estimator with a more efficient rate of convergence  $N^{1/4}$ , which we prove to be the best attainable rate. For this purpose we prove local asymptotic normality with rate  $N^{-1/4}$  for a simplified model and obtain bounds for the asymptotic Fisher information. With the minimax theorem we conclude that  $N^{1/4}$  is

a lower bound for the rate of convergence even in the synchronous equidistant case. Our estimator hence upgrades the estimator proposed in Palandri (2006) to a rate-optimal consistent estimator for integrated covariances and leads to a suitable implementation of integrated variance and covariance estimation following the multi-scale approach invented by Zhang (2006a) and our extension for the covariance case. Barndorff-Nielsen et al. (2008) present a multivariate realized kernel estimator that, furthermore, guarantees to be positive semi-definite, which is aside from non-synchronicity and noise a third important issue in multivariate considerations. Their estimator has a  $N^{1/5}$ -rate of convergence. In Section 2 we introduce the model and our basic notation. We present a brief outline and the two main results of this article concerning the asymptotics of our multi-scale estimator for the quadratic covariation and local asymptotic normality.

## 2 Model and Main Results

We want to obtain a consistent estimator for the covariation  $\langle \tilde{X}, \tilde{Y} \rangle_T$  of two Itô processes  $\tilde{X}_t$  and  $\tilde{Y}_t$  over a fixed time interval  $[0, T]$ , e. g. one trading day, when we observe discrete asynchronous returns contaminated by market microstructure noise. The observations of  $X$  will be denoted by

$$X_{t_0}, \dots, X_{t_n}, \quad 0 \leq t_0 < t_1 < \dots < t_n \leq T, \text{ with increments } \Delta X_{t_i} = X_{t_i} - X_{t_{i-1}},$$

and the observations of another log-price process  $Y$  by

$$Y_{\tau_0}, \dots, Y_{\tau_m}, \quad 0 \leq \tau_0 < \tau_1 < \dots < \tau_m \leq T, \text{ with increments } \Delta Y_{\tau_j} = Y_{\tau_j} - Y_{\tau_{j-1}}.$$

Synchronous data would mean that  $m = n$  and  $t_i = \tau_i$  for all  $i \in \{0, \dots, n\}$ . We consider the general case where the number of observations may differ and the sets of observation times  $\mathcal{O}^X = \{t_0, \dots, t_n\}$  and  $\mathcal{O}^Y = \{\tau_0, \dots, \tau_m\}$  also contain points  $t_i \notin \mathcal{O}^Y$  and  $\tau_j \notin \mathcal{O}^X$ . Usually the considered time interval  $[0, T]$  starts with the first observation which means  $t_0 = 0$  or  $\tau_0 = 0$ . We work within the model imposed by the following two assumptions:

**Assumption 1.** *The observed log-price processes are described by the sums of efficient Itô processes and independent (discrete-time) noise processes  $\epsilon_{t_i}^X$  and  $\epsilon_{\tau_j}^Y$ :*

$$\begin{aligned} X_{t_i} &= \tilde{X}_{t_i} + \epsilon_{t_i}^X, \quad i \in \{0, \dots, n\} \\ Y_{\tau_j} &= \tilde{Y}_{\tau_j} + \epsilon_{\tau_j}^Y, \quad j \in \{0, \dots, m\}. \end{aligned}$$

On a filtered probability space  $(\Omega, \mathcal{F}, (\mathcal{F}_t))$  the efficient processes are defined by

$$\begin{aligned} d\tilde{X}_t &= \mu_t^X dt + \sigma_t^X dB_t^X, \\ d\tilde{Y}_\tau &= \mu_\tau^Y d\tau + \sigma_\tau^Y dB_\tau^Y, \end{aligned}$$

where  $B^X$  and  $B^Y$  are two  $\mathcal{F}_t$ -adapted correlated standard Brownian motions with correlation  $\rho_t$ . The drift  $\mu_t$  and spot volatility  $\sigma_t$  for both are  $\mathcal{F}_t$ -adapted, continuous and bounded stochastic processes.

**Assumption 2.** *The errors  $\epsilon_{t_i}^X, i \in \{0, \dots, n\}$  and  $\epsilon_{\tau_j}^Y, j \in \{0, \dots, m\}$  due to market microstructure noise are assumed to be i.i.d. processes and independent to each other and the efficient processes. We also assume*

$$\mathbb{E}\epsilon^X := \mathbb{E}\epsilon_{t_i}^X = \mathbb{E}\epsilon^Y := \mathbb{E}\epsilon_{\tau_j}^Y = 0 \quad \forall i, j$$

and  $\mathbb{E}(\epsilon^X)^4, \mathbb{E}(\epsilon^Y)^4 < \infty$  where  $\mathbb{E}(\epsilon^X)^k := \mathbb{E}[(\epsilon_{t_i}^X)^k]$  and  $\mathbb{E}(\epsilon^Y)^k := \mathbb{E}[(\epsilon_{\tau_j}^Y)^k]$  analogously.

The variances of the noise processes will be denoted by

$$\eta_X^2 := \mathbb{E} \left[ (\epsilon_{t_i}^X)^2 \right] \quad \text{and} \quad \eta_Y^2 := \mathbb{E} \left[ (\epsilon_{\tau_j}^Y)^2 \right] .$$

We want to obtain a consistent rate-optimal estimator for the integrated covariance of the efficient processes  $\langle \tilde{X}, \tilde{Y} \rangle_T = \int_0^T \rho_t \sigma_t^X \sigma_t^Y dt$  under asymptotics where  $\sup_i \Delta t_i \rightarrow 0$  and  $\sup_j \Delta \tau_j \rightarrow 0$ . Assumptions concerning the observation times are imposed later on after constructing a synchronized joint grid. Evidently, we need a form of regularization criterion to ensure the number of observations for both processes and the length of (usually not equidistant) time intervals  $\Delta t_i = (t_i - t_{i-1})$  and  $\Delta \tau_j = (\tau_j - \tau_{j-1})$  being of the same order. Recall that for our analysis we regard the conditional law given the observation times. A precise analysis for the case of random trading times (e. g. event times of counting processes) require some additional concepts that are not the focus of this article although we will use Poisson processes to generate the observation times in Section 7. The latter could be interesting when tick-by-tick data are considered, where trading times can be modeled as Poisson arrivals. We refer to Zhang (2006b) for an analysis of this special case.

At this point we present an outline and an outlook on the the two main results of this article. The article is organized as follows: in Section 3 we describe the method of synchronization and show that the Hayashi-Yoshida estimator becomes inconsistent in the presence of market microstructure noise. In Section 4 we present the subsampling estimator, which is similar to the one proposed by Palandri (2006), but using a different representation, which will be useful for our further analysis. In Section 5 we develop our new estimator using a multi-scale approach that improves the rate of convergence to  $N^{1/4}$  resulting in a more efficient estimation method for the quadratic covariation  $\langle \tilde{X}, \tilde{Y} \rangle_T$ . This is one of our main results and can be summarized in the following Theorem 1:

**Theorem 1.** *Let Assumptions 1, 2 and 3, that will be stated in Section 3, be satisfied.*

*We will prove that the multi-scale estimator  $\widehat{\langle \tilde{X}, \tilde{Y} \rangle}_T^{(mult)}$ , which will be constructed in Sections 3-5, is unbiased and has the following asymptotic property:*

$$\left( \mathbb{E} \left[ \left( \widehat{\langle \tilde{X}, \tilde{Y} \rangle}_T^{mult} - \langle \tilde{X}, \tilde{Y} \rangle_T \right)^2 \right] \right)^{1/2} = \mathcal{O} \left( N^{-1/4} \right) .$$

*N denotes the number of synchronized observations emanating from the synchronization method that will be presented in the next section.*

In Section 5, after the construction of our multi-scale estimator, Proposition 9 gives a more detailed version of Theorem 1.

In Section 6 we consider a simplified parametric model with constant parameters and equidistantly and synchronously observed data contaminated with market microstructure noise. The key result is the following Theorem 2:

**Theorem 2.** *For a constant correlation coefficient  $\rho$  and synchronously, equidistantly observed Brownian Motions with i. i. d. Gaussian noise local asymptotic normality (LAN) holds with  $N^{-1/4}$ -rate. We conclude that  $N^{1/4}$  is a lower bound for the rate of convergence for any sequence of estimators and hence our multi-scale estimator is rate-optimal and asymptotically efficient.*

In Proposition 11 we also give bounds for the asymptotic Fisher information. Simulation results follow in Section 7, where we compare the two proposed estimators and see that

the multi-scale estimator also performs better in the case of finite sample sizes. We observe that if the influence of market microstructure effects is incisive the multi-scale approach provides a much more efficient estimation compared to the estimator based on subsampling. For the simulations we generated the observation times as arrivals of Poisson processes.

### 3 Dealing with asynchronicity: Synchronizing data and covariance estimation for observations without noise

Hayashi and Yoshida (2005) proved the consistency of their estimator

$$\widehat{\langle \tilde{X}, \tilde{Y} \rangle}_T^{(HY)} = \sum_{i=1}^n \sum_{j=1}^m \Delta \tilde{X}_{t_i} \Delta \tilde{Y}_{\tau_j} \mathbb{1}_{[\min(t_i, \tau_j) > \max(t_{i-1}, \tau_{j-1})]},$$

where the product terms include all increments over overlapping time intervals, for the case where we observe the efficient price processes without noise.

In the case of high-frequency data contaminated by market microstructure noise, however, the estimator becomes inconsistent and explodes (tends to infinity) for  $\min(n, m) \rightarrow \infty$ . We will prove this in Proposition 1, but we focus first on an alternative useful method to deal with the asynchronicity of the data. This method was presented in Palandri (2006) (which he calls pseudo-aggregation).

For this purpose we construct a set with one or more than one observation times as elements which we call the joint grid. Those elements will be denoted by  $\mathcal{H}^i$  and  $\mathcal{G}^i, i \in \{0, \dots, N\}$ , and  $(N + 1)$  is the number of sets contained in the joint grid. The resulting synchronous realized covariance estimator for the integrated covariance will coincide with the one of Hayashi and Yoshida (2005) but this approach will be useful for our analysis of noise terms. In particular, this construction will enable us to deal with the noise contamination by applying subsampling techniques in Section 4. Based on this construction we get a joint grid with  $(N + 1)$  sets  $\mathcal{H}^i$  and  $\mathcal{G}^i$  where  $N < \min(n, m)$ . The last fact indicates heuristically that the efficiency of such techniques of covariance estimation depends on the number of observations available for the lower frequent process. In the ideal case both observation frequencies do not differ too much, i. e.  $n$  and  $m$  are of the same order and both assets show similar liquidity. In Assumption 3 we will give a more precise statement on the conditions imposed on the observation times needed for our analysis.

The method of constructing a joint grid for the observations of both processes is described by the following iterative algorithm:

first step:

- for  $t_0 < \tau_0$  and  $\tilde{w}_0 := \min(w \in \{1, \dots, n\} | t_{w-1} < \tau_0 \leq t_w)$ :

$$\mathcal{H}^0 = \{t_0, \dots, t_{\tilde{w}_0}\} \text{ and } \mathcal{G}^0 = \{\tau_0\}$$

$$q_1 := \begin{cases} \tilde{w}_0 + 1 & \text{if } \tau_0 = t_{\tilde{w}_0} \\ \tilde{w}_0 & \text{if } \tau_0 < t_{\tilde{w}_0} \end{cases} \text{ and } r_1 := 1$$

- for  $t_0 = \tau_0$ :

$$\mathcal{H}^0 = \{t_0\} \text{ and } \mathcal{G}^0 = \{\tau_0\}$$

$$q_1 := 1 \text{ and } r_1 := 1$$

- for  $t_0 > \tau_0$  and  $\tilde{l}_0 := \min(l \in \{1, \dots, m\} | \tau_{l-1} < t_0 \leq \tau_l)$ :

$$\mathcal{H}^0 = \{t_0\} \text{ and } \mathcal{G}^0 = \{\tau_0, \dots, \tau_{\tilde{l}_0}\}$$

$$q_1 := 1 \text{ and } r_1 := \begin{cases} \tilde{l}_0 + 1 & \text{if } t_0 = \tau_{\tilde{l}_0} \\ \tilde{l}_0 & \text{if } t_0 < \tau_{\tilde{l}_0} \end{cases}$$

i-th step (given  $\mathcal{H}^{i-1}$  and  $\mathcal{G}^{i-1}$ ):

- for  $t_{q_i} < \tau_{r_i}$  and  $\tilde{w}_i := \min(w \in \{q_i + 1, \dots, n\} | t_{w-1} < \tau_{r_i} \leq t_w)$ :

$$\mathcal{H}^i = \{t_{q_i}, \dots, t_{\tilde{w}_i}\} \text{ and } \mathcal{G}^i = \{\tau_{r_i}\}$$

$$q_i \dashrightarrow \begin{cases} q_{i+1} = \tilde{w}_i + 1 & \text{if } \tau_{r_i} = t_{\tilde{w}_i} \\ q_{i+1} = \tilde{w}_i & \text{if } \tau_{r_i} < t_{\tilde{w}_i} \end{cases} \text{ and } r_i \dashrightarrow r_{i+1} = r_i + 1$$

- for  $t_{q_i} = \tau_{r_i}$ :

$$\mathcal{H}^i = \{t_{q_i}\} \text{ and } \mathcal{G}^i = \{\tau_{r_i}\}$$

$$q_i \dashrightarrow q_{i+1} = q_i + 1 \text{ and } r_i \dashrightarrow r_{i+1} = r_i + 1$$

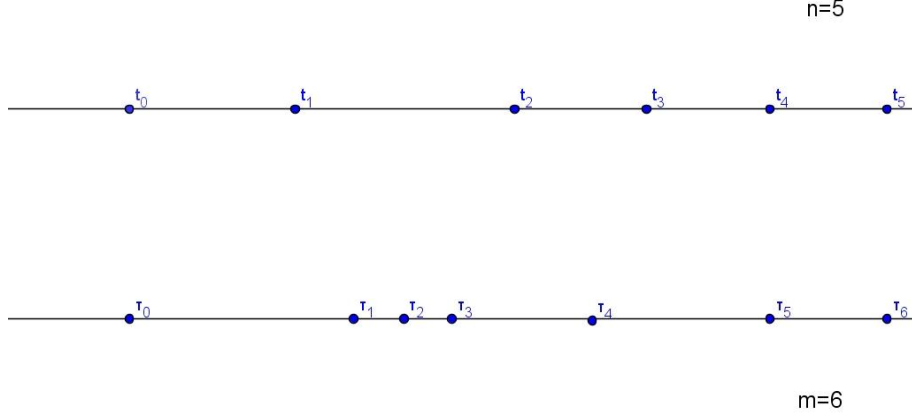
- for  $t_{q_i} > \tau_{r_i}$  and  $\tilde{l}_i := \min(l \in \{r_i + 1, \dots, m\} | \tau_{l-1} < t_{q_i} \leq \tau_l)$ :

$$\mathcal{H}^i = \{t_{q_i}\} \text{ and } \mathcal{G}^i = \{\tau_{r_i}, \dots, \tau_{\tilde{l}_i}\}$$

$$q_i \dashrightarrow q_{i+1} = q_i + 1 \text{ and } r_i \dashrightarrow \begin{cases} r_{i+1} = \tilde{l}_i + 1 & \text{if } t_{q_i} = \tau_{\tilde{l}_i} \\ r_{i+1} = \tilde{l}_i & \text{if } t_{q_i} < \tau_{\tilde{l}_i} \end{cases}$$

Let us give an example for the construction of the joint grid. Of course the example is just for illustration and the number of observations  $m = 6$  and  $N = n = 5$  is restricted and much smaller than in practice. The example emphasizes some important issues appearing when observations are asynchronous.

## Example



In the example illustrated above we have  $\mathcal{H}^0 = \{t_0\}$ ,  $\mathcal{G}^0 = \{\tau_0\}$ ,  $\mathcal{H}^1 = \{t_1, t_2\}$ ,  $\mathcal{G}^1 = \{\tau_1\}$ ,  $\mathcal{H}^2 = \{t_2\}$ ,  $\mathcal{G}^2 = \{\tau_2, \tau_3, \tau_4\}$ ,  $\mathcal{H}^3 = \{t_3\}$ ,  $\mathcal{G}^3 = \{\tau_4, \tau_5\}$ ,  $\mathcal{H}^4 = \{t_4\}$ ,  $\mathcal{G}^4 = \{\tau_5\}$ ,  $\mathcal{H}^5 = \{t_5\}$ ,  $\mathcal{G}^5 = \{\tau_6\}$ . The example shows the important fact that the sets  $\mathcal{H}^i$  and  $\mathcal{G}^i$  are in general not disjoint and the maxima of consecutive sets can be the same time points. The minimum of a successive set can as well equal the maximum of the previous one. For further examples see Palandri (2006).

Next we pass over from the original observations to the sums of observed **increments** of the noisy log-prices over sets  $\mathcal{H}^i$  and  $\mathcal{G}^i$ , respectively:

$$X^{\mathcal{H}^i} := \sum_{t_j \in \mathcal{H}^i} \Delta X_{t_j}, \quad Y^{\mathcal{G}^i} := \sum_{\tau_j \in \mathcal{G}^i} \Delta Y_{\tau_j}, \quad i \in \{0, \dots, N\}.$$

We observe that the realized covariance of the synchronized observations

$$\text{SRC} := \sum_{i=0}^N X^{\mathcal{H}^i} Y^{\mathcal{G}^i} = \sum_{i=1}^n \sum_{j=1}^m \Delta \tilde{X}_{t_i} \Delta \tilde{Y}_{\tau_j} \mathbb{1}_{[\min(t_i, \tau_j) > \max(t_{i-1}, \tau_{j-1})]}$$

is the well-known Hayashi-Yoshida estimator. We use a different representation of this estimator compared to Palandri (2006) using telescoping sums. If we define

$$\begin{aligned} \mu_i &:= \max(k | t_k \in \mathcal{H}^i), & \tilde{\mu}_i &:= \max(k | \tau_k \in \mathcal{G}^i) \text{ and} \\ \nu_i &:= \min(k | t_k \in \mathcal{H}^i), & \tilde{\nu}_i &:= \min(k | \tau_k \in \mathcal{G}^i), \quad i \in \{0, \dots, N\} \end{aligned}$$

and for the purpose of a simpler notation

$$\begin{aligned} X_{g_i} &:= X_{t_{\mu_i}}, & Y_{\gamma_i} &:= Y_{\tau_{\tilde{\mu}_i}}, \quad i \in \{0, \dots, N\} \text{ and} \\ X_{l_i} &:= X_{t_{\nu_{i-1}}}, & Y_{\lambda_i} &:= Y_{\tau_{\tilde{\nu}_{i-1}}}, \quad i \in \{1, \dots, N\} \end{aligned}$$

with  $l_0 := t_0$ ,  $\lambda_0 := \tau_0$  we can write  $X^{\mathcal{H}^i}$  and  $Y^{\mathcal{G}^i}$  as telescoping sums  $X^{\mathcal{H}^i} = (X_{g_i} - X_{l_i})$ ,  $Y^{\mathcal{G}^i} = (Y_{\gamma_i} - Y_{\lambda_i})$  which leads to

$$\widehat{\langle \tilde{X}, \tilde{Y} \rangle}_T^{(HY)} = \sum_{i=0}^N (X_{g_i} - X_{l_i}) (Y_{\gamma_i} - Y_{\lambda_i}).$$

In this notation  $g_i$  denotes the greatest and  $l_i$  the last observation time before the least element of the set  $\mathcal{H}^i$  and analogously  $\gamma_i$  and  $\lambda_i$  of  $\mathcal{G}^i$ .

We are interested in asymptotics when  $n, m \rightarrow \infty$  being of the same order and the time lags between returns tending to zero and hence impose the following assumption on the observation design:

**Assumption 3.** *For the time lags between the observations*

$$\delta_N^X := \sup_{i \in \{1, \dots, n\}} (t_i - t_{i-1}) = \mathcal{O}\left(\frac{1}{N}\right),$$

$$\delta_N^Y := \sup_{j \in \{1, \dots, m\}} (\tau_j - \tau_{j-1}) = \mathcal{O}\left(\frac{1}{N}\right)$$

holds.

By this assumption we exclude data where the lengths of the observed time intervals vary vigorously or the number of observations  $n$  and  $m$  are of different order. The assumption seems not too restrictive for most applications. Similar conditions are often imposed for asymptotic analysis, see e. g. Zhang (2006b). Further results are deduced under asymptotics for  $n, m \rightarrow \infty$  and Assumption 3. In the following we will use the notation  $\mathbb{E}_{\tilde{X}, \tilde{Y}}[\cdot] := \mathbb{E}[\cdot | \tilde{X}, \tilde{Y}]$  for the conditional expectation given the paths of both efficient processes.

**Proposition 1.** *Under Assumptions 1, 2 and 3 for the synchronized realized covariance estimator*

$$\mathbb{E}_{\tilde{X}, \tilde{Y}} \left[ \widehat{\langle \tilde{X}, \tilde{Y} \rangle}_T^{(HY)} \right] = \langle \tilde{X}, \tilde{Y} \rangle_T, \quad \text{Var}_{\tilde{X}, \tilde{Y}} \left( \widehat{\langle \tilde{X}, \tilde{Y} \rangle}_T^{(HY)} \right) = \mathcal{O}_p(N)$$

holds.

*Proof.* Unbiasedness follows directly from Assumption 2 and the conditional variance can be simplified to:

$$\begin{aligned} \text{Var}_{\tilde{X}, \tilde{Y}} \left( \widehat{\langle \tilde{X}, \tilde{Y} \rangle}_T^{(HY)} \right) &= \mathbb{E} \left[ \sum_{i=0}^N (\epsilon_{g_i}^X - \epsilon_{l_i}^X) (\epsilon_{\gamma_i}^Y - \epsilon_{\lambda_i}^Y) \right]^2 + \mathcal{O}_p(1) = \\ &= \mathbb{E} \left[ \sum_{i=0}^N (\epsilon_{g_i}^X - \epsilon_{l_i}^X)^2 (\epsilon_{\gamma_i}^Y - \epsilon_{\lambda_i}^Y)^2 \right] + \mathcal{O}_p(1) = 4N\eta_X^2\eta_Y^2 + \mathcal{O}_p(1) = \mathcal{O}_p(N). \end{aligned}$$

The variances of  $\sum_i (\epsilon_{g_i}^X - \epsilon_{l_i}^X) (\epsilon_{\gamma_i}^Y - \epsilon_{\lambda_i}^Y)$  and the second sum including increments of  $\tilde{X}$  and  $\epsilon^Y$  lead to the term of order 1 in probability. The mixed terms in the remaining second moment have an expectation equal to zero although consecutive sets  $\mathcal{H}^i$  and  $\mathcal{H}^{i+1}$  (or  $\mathcal{G}^i$  and  $\mathcal{G}^{i+1}$ ) are not generally disjoint. Nevertheless, our synchronization method was defined such that if the intersection of  $\mathcal{H}^i$  and  $\mathcal{H}^{i+1}$  is non-empty,  $\mathcal{G}^i \cap \mathcal{G}^{i+1} = \emptyset$  holds. Assumption 2 yields that each summand has expectation  $2\eta_X^2 \cdot 2\eta_Y^2$ .  $\square$

We conclude that in the presence of market microstructure frictions the Hayashi-Yoshida estimator cannot yield a consistent estimation of the integrated covariance of the underlying efficient processes.



## 4 Dealing with noise contamination: the subsample estimator

In this section we show that a subsample approach as presented in Zhang et al. (2005) leads to a consistent integrated covariance estimator with  $N^{1/6}$ -rate of convergence. The sets  $\mathcal{H}^i$  and  $\mathcal{G}^i$  are grouped in  $K_N$  subsamples and for each subsample the (lower-frequency) realized covariances are calculated. So the synchronized observations are arranged to  $K_N$  subsets of observations and for each subset we can calculate a realized covariance for which the error due to noise is smaller than for the highest-frequency realized covariance because of Proposition 1. Averaging the realized covariations calculated with lower frequencies leads to the resulting estimator:

$$\widehat{\langle \tilde{X}, \tilde{Y} \rangle}_T^{sub} = \frac{1}{K_N} \sum_{i=K_N}^N (X_{g_i} - X_{l_{i-K_N}}) (Y_{\gamma_i} - Y_{\lambda_{i-K_N}}). \quad (1)$$

Recall that we use the synchronized data and the joint grid that we presented in Section 3 to calculate the subsampling estimator.  $K$  depends on  $N$ , but we drop the index in the following. The sum over  $K$  subsamples and the sum over all increments in each subsample can be simplified to the form of the estimator stated above if we assume a regular allocation to subsamples ( $\mathcal{H}^0, \mathcal{H}^K, \dots$  in the first subsample,  $\mathcal{H}^1, \mathcal{H}^{K+1}, \dots$  in the second, etc.). We do not consider boundary effects (we focus on asymptotics  $K = o(N)$  and  $K, n \rightarrow \infty$ ) and hence leave out the weights imposed by Palandri (2006). His (realized) covariance estimator is defined as an average of weighted sums of products over overlapping increments also divided in subsamples. If we regard the unions

$$A_{v,w} = \bigcup_{i=(w-1)K+v}^{wK+(v-1)} \mathcal{H}^i \quad \text{and} \quad B_{v,w} = \bigcup_{i=(w-1)K+v}^{wK+(v-1)} \mathcal{G}^i$$

of  $K$  sets  $\mathcal{H}^i$  and  $\mathcal{G}^i$ , respectively, and  $X_{v,w} = \sum_{t_j \in A_{v,w}} \Delta X_{t_j}$  and  $Y_{v,w} = \sum_{\tau_j \in B_{v,w}} \Delta Y_{\tau_j}$ , respectively, his ‘Consistent Realized Covariance’ estimator

$$\mathbf{CRC} = \frac{1}{K} \sum_{v=0}^{K-1} w_v \sum_{w=1}^{N/K-1} X_{v,w} Y_{v,w}$$

corresponds to our proposed estimator (except the weights  $w_v$ ). Assuming that the noise processes across assets are independent, this estimator is unbiased and as we will prove in the following has for optimal choice  $K = \mathcal{O}(N^{2/3})$  an asymptotic variance of order  $N^{-1/3}$ . The unbiasedness is the reason why we do not need a bias-correction term in contrast to the realized variance case (two time-scales estimator presented by Zhang et al. (2005)).

Palandri (2006) has proved this result for his similar estimator, but it is reasonable for our further analysis concerning the multi-scale estimator in chapter 5 to give a short calculation of the asymptotic variance of the subsample estimator in our illustration. We are only interested in the order of the asymptotic variance and we impose mild assumptions on the grid and time intervals that are inherent in the method of subsampling and the data. In particular, we assume a regular allocation to subsamples as stated above and Assumption 3 that ensures (together with Assumption 1)  $\Delta \tilde{X}_{t_i} = \mathcal{O}_p(\sqrt{1/N})$ ,  $\Delta \tilde{Y}_{\tau_j} = \mathcal{O}_p(\sqrt{1/N})$ .

The total variance can be written as

$$\begin{aligned} \text{Var} \left( \widehat{\langle \tilde{X}, \tilde{Y} \rangle}_T^{sub} \right) &= \frac{1}{K^2} \sum_{i=K}^N \sum_{j=K}^N \text{Cov} \left( (X_{g_i} - X_{l_{i-K}}) (Y_{\gamma_i} - Y_{\lambda_{i-K}}) , \right. \\ &\quad \left. (X_{g_j} - X_{l_{j-K}}) (Y_{\gamma_j} - Y_{\lambda_{j-K}}) \right) \\ &= \frac{1}{K^2} \sum_{i=K}^N \sum_{j=K}^N \left( \text{Cov} (\mathfrak{e}_i, \mathfrak{e}_j) + \text{Cov} (\mathfrak{m}_i, \mathfrak{m}_j) + \text{Cov} (\mathfrak{v}_i, \mathfrak{v}_j) + \text{Cov} (\mathfrak{n}_i, \mathfrak{n}_j) \right) \quad (2) \end{aligned}$$

with the four uncorrelated terms:

$$\begin{aligned} \mathfrak{e}_i &= (\tilde{X}_{g_i} - \tilde{X}_{l_{i-K}}) (\tilde{Y}_{\gamma_i} - \tilde{Y}_{\lambda_{i-K}}) , \\ \mathfrak{m}_i &= (\tilde{X}_{g_i} - \tilde{X}_{l_{i-K}}) (\epsilon_{\gamma_i}^Y - \epsilon_{\lambda_{i-K}}^Y) , \\ \mathfrak{v}_i &= (\epsilon_{g_i}^X - \epsilon_{l_{i-K}}^X) (\tilde{Y}_{\gamma_i} - \tilde{Y}_{\lambda_{i-K}}) , \\ \mathfrak{n}_i &= (\epsilon_{g_i}^X - \epsilon_{l_{i-K}}^X) (\epsilon_{\gamma_i}^Y - \epsilon_{\lambda_{i-K}}^Y) . \end{aligned}$$

We will consider the four summands consecutively.

We start our analysis of the asymptotic orders of the different summands in the total variance focusing on the sum of covariances  $\text{Cov} (\mathfrak{e}_i, \mathfrak{e}_j)$ . This is the variance due to discretization and would be the total variance of a subsampling estimator calculated with observations of the efficient processes without noise. First we deduce the order of increments for the efficient processes from Assumptions 1 and 3.

**Lemma 2.** *If Assumptions 1 and 3 hold, we obtain the following asymptotic orders for the efficient processes without microstructure noise:*

$$\mathbb{E} \left[ (\tilde{X}_{g_i} - \tilde{X}_{l_{i-K}})^2 \right] = \mathcal{O} (K/N) , \quad (3a)$$

$$\mathbb{E} \left[ (\tilde{Y}_{\gamma_i} - \tilde{Y}_{\lambda_{i-K}})^2 \right] = \mathcal{O} (K/N) , \quad (3b)$$

$$\mathbb{E} \left[ (\tilde{X}_{g_i} - \tilde{X}_{l_{i-K}}) (\tilde{Y}_{\gamma_i} - \tilde{Y}_{\lambda_{i-K}}) \right] = \mathcal{O} (K/N) , \quad (3c)$$

$$\mathbb{E} \left[ (\tilde{X}_{g_i} - \tilde{X}_{l_{i-K}})^2 (\tilde{Y}_{\gamma_i} - \tilde{Y}_{\lambda_{i-K}})^2 \right] = \mathcal{O} (K^2/N^2) . \quad (3d)$$

*Proof.* By Assumption 1 the drifts of the efficient processes are bounded and thus

$$\mathbb{E} \left[ (\tilde{X}_{g_i} - \tilde{X}_{l_{i-K}})^2 \right] = \mathbb{E} \left[ \left( \int_{l_{i-K}}^{g_i} \sigma_t^X dB_t^X \right)^2 \right] + \mathcal{O} \left( \frac{K}{N} \right)$$

holds because the squared drift term is of order  $K^2/N^2$  and the mixed term is of order  $(K/N)^{3/2}$ . Therefore, to prove (3a) it suffices to apply Itô isometry and the mean value theorem using again Assumption 1 for the spot volatilities:

$$\begin{aligned} \mathbb{E} \left[ \left( \int_{l_{i-K}}^{g_i} \sigma_t^X dB_t^X \right)^2 \right] &= \mathbb{E} \left[ \int_{l_{i-K}}^{g_i} (\sigma_t^X)^2 dt \right] \\ &= \sigma_{l_{i-K}, g_i}^{MVT} (g_i - l_{i-K}) = \mathcal{O} \left( \frac{K}{n} \right) . \end{aligned}$$

The asymptotic orders of the time increments are given by Assumption 3. The constant  $\sigma_{l_{i-K}, g_i}^{MVT}$  occurring by application of the mean value theorem is finite because of Assumption 1. The proof of (3b) follows analogously.

Using (3a) and (3b) we obtain (3c) by the Cauchy-Schwarz inequality:

$$\begin{aligned} \mathbb{E} \left[ (\tilde{X}_{g_i} - \tilde{X}_{l_{i-K}}) (\tilde{Y}_{\gamma_i} - \tilde{Y}_{\lambda_{i-K}}) \right] &\leq \sqrt{\mathbb{E} \left[ (\tilde{X}_{g_i} - \tilde{X}_{l_{i-K}})^2 \right]} \cdot \sqrt{\mathbb{E} \left[ (\tilde{Y}_{\gamma_i} - \tilde{Y}_{\lambda_{i-K}})^2 \right]} \\ &= \mathcal{O} \left( \frac{K}{n} \right). \end{aligned}$$

The fourth moments of the increments can be bounded by the squared quadratic covariation using the Burkholder-Davis-Gundy inequality, so it is adequate to prove (3d) again using the Cauchy-Schwarz inequality:

$$\begin{aligned} \mathbb{E} \left[ (\tilde{X}_{g_i} - \tilde{X}_{l_{i-K}})^2 (\tilde{Y}_{\gamma_i} - \tilde{Y}_{\lambda_{i-K}})^2 \right] &\leq \sqrt{\mathbb{E} \left[ (\tilde{X}_{g_i} - \tilde{X}_{l_{i-K}})^4 \right]} \cdot \sqrt{\mathbb{E} \left[ (\tilde{Y}_{\gamma_i} - \tilde{Y}_{\lambda_{i-K}})^4 \right]} \\ &\leq \sqrt{c \mathbb{E} \left( \langle \tilde{X}, \tilde{X} \rangle_{g_i} - \langle \tilde{X}, \tilde{X} \rangle_{l_{i-K}} \right)^2} \cdot \sqrt{c^* \mathbb{E} \left( \langle \tilde{Y}, \tilde{Y} \rangle_{\gamma_i} - \langle \tilde{Y}, \tilde{Y} \rangle_{\lambda_{i-K}} \right)^2} = \mathcal{O} \left( \frac{K^2}{N^2} \right) \end{aligned}$$

with constants  $c$  and  $c^*$  from the application of the Burkholder-Davis-Gundy inequality. The order  $K^2/N^2$  then easily follows e. g. using again the mean value theorem as above.  $\square$

**Corollary 3.** *We obtain for the variance due to discretization, when there is no microstructure noise present, that we denote by  $\mathbb{V}\text{ar}_{\eta_X=\eta_Y=0}(\cdot)$ :*

$$\mathbb{V}\text{ar}_{\eta_X=\eta_Y=0} \left( \widehat{\langle \tilde{X}, \tilde{Y} \rangle}_T^{sub} \right) = \mathcal{O} \left( \frac{K}{N} \right).$$

*Proof.*

$$\begin{aligned} \mathbb{V}\text{ar}_{\eta_X=\eta_Y=0} \left( \widehat{\langle \tilde{X}, \tilde{Y} \rangle}_T^{sub} \right) &= \frac{1}{K^2} \sum_{i=K}^N \sum_{j=K}^N \text{Cov}(\mathbb{e}_i, \mathbb{e}_j) = \\ &= \frac{2}{K^2} \sum_{i=K}^N \sum_{j=K}^{i-1} \text{Cov} \left( (\tilde{X}_{g_i} - \tilde{X}_{l_{i-K}}) (\tilde{Y}_{\gamma_i} - \tilde{Y}_{\lambda_{i-K}}), (\tilde{X}_{g_j} - \tilde{X}_{l_{j-K}}) (\tilde{Y}_{\gamma_j} - \tilde{Y}_{\lambda_{j-K}}) \right) \\ &\quad + \frac{1}{K^2} \sum_{i=K}^N \text{Var}(\mathbb{e}_i) \\ &= \frac{2}{K^2} \sum_{i=K}^N \sum_{j=K}^{i-1} \mathbb{1}_{\{|i-j| \leq K\}} \text{Cov} \left( (\tilde{X}_{g_i} - \tilde{X}_{g_j} + \tilde{X}_{g_j} - \tilde{X}_{l_{i-K}}) (\tilde{Y}_{\gamma_i} - \tilde{Y}_{\gamma_j} + \tilde{Y}_{\gamma_j} - \tilde{Y}_{\lambda_{i-K}}), \right. \\ &\quad \left. (\tilde{X}_{g_j} - \tilde{X}_{l_{i-K}} + \tilde{X}_{l_{i-K}} - \tilde{X}_{l_{j-K}}) (\tilde{Y}_{\gamma_j} - \tilde{Y}_{\lambda_{i-K}} + \tilde{Y}_{\lambda_{i-K}} - \tilde{Y}_{\lambda_{j-K}}) \right) + \frac{1}{K^2} \sum_{i=K}^N \text{Var}(\mathbb{e}_i) \\ &= \frac{2}{K^2} \sum_{i=K}^N \sum_{j=K}^{i-1} \mathbb{1}_{\{|i-j| \leq K\}} \text{Var} \left( (\tilde{X}_{g_j} - \tilde{X}_{l_{i-K}}) (\tilde{Y}_{\gamma_j} - \tilde{Y}_{\lambda_{i-K}}) \right) + \frac{1}{K^2} \sum_{i=K}^N \text{Var}(\mathbb{e}_i) \\ &= \mathcal{O} \left( \frac{1}{K^2} N K \frac{K^2}{N^2} \right) = \mathcal{O} \left( \frac{K}{N} \right). \end{aligned}$$

In this calculation we also used characteristics of our synchronization method. The increments  $(\tilde{X}_{g_i} - \tilde{X}_{l_{i-K}})$  and  $(\tilde{X}_{g_j} - \tilde{X}_{l_{j-K}})$  are non-overlapping and hence uncorrelated for  $|i - j| > K$ . Taking the construction procedure of the joint grid into account, increments  $(\tilde{X}_{g_i} - \tilde{X}_{l_{i-K}})$  and  $(\tilde{Y}_{\gamma_j} - \tilde{Y}_{\lambda_{j-K}})$  are uncorrelated as well for  $|i - j| > K$ .  $\square$

So, the asymptotic order of the discretization variance is

$$\frac{1}{K^2} \sum_{i=K}^N \sum_{j=K}^N \text{Cov}(\mathbb{e}_i, \mathbb{e}_j) = \mathcal{O}\left(\frac{K}{N}\right). \quad (4)$$

For the mixed summands in (2) we obtain under Assumption 1, 2 and 3

$$\begin{aligned} \frac{1}{K^2} \sum_{i=K}^N \sum_{j=K}^N \text{Cov}(\mathbb{m}_i, \mathbb{m}_j) &= \frac{1}{K^2} \sum_{i=K}^N \text{Var}(\mathbb{m}_i) + \frac{2}{K^2} \sum_{i=K}^{N-1} \text{Cov}(\mathbb{m}_i, \mathbb{m}_{i+1}) \\ &\leq \frac{3}{K^2} \sum_{i=1}^N \text{Var}(\mathbb{m}_i) = C_{\mathbf{m}} \frac{\eta_Y^2}{K} = \mathcal{O}\left(\frac{\eta_Y^2}{K}\right) \end{aligned} \quad (5)$$

with a constant  $C_{\mathbf{m}}$  and analogously with a constant  $C_{\nu}$

$$\frac{1}{K^2} \sum_{i=K}^N \sum_{j=K}^N \text{Cov}(\mathbb{v}_i, \mathbb{v}_j) \leq C_{\nu} \frac{\eta_X^2}{K} = \mathcal{O}\left(\frac{\eta_X^2}{K}\right). \quad (6)$$

The fourth (noise) term in (2) is of order  $N/K^2$  because

$$\begin{aligned} \frac{1}{K^2} \sum_{i=K}^N \sum_{j=K}^N \text{Cov}(\mathbb{n}_i, \mathbb{n}_j) &= \frac{1}{K^2} \sum_{i=K}^N \text{Var}(\mathbb{n}_i) + \frac{2}{K^2} \sum_{i=K}^{N-1} \text{Cov}(\mathbb{n}_i, \mathbb{n}_{i+1}) \\ &\leq \frac{3}{K^2} \sum_{i=1}^N \text{Var}(\mathbb{n}_i) = C_{\mathbf{n}} \frac{\eta_X^2 \eta_Y^2 N}{K^2} = \mathcal{O}\left(\frac{\eta_X^2 \eta_Y^2 N}{K^2}\right) \end{aligned} \quad (7)$$

with a constant  $C_{\mathbf{n}}$  holds. We used Assumption 2 that the noise is i. i. d., but recall that  $g_i = g_{i+1}$  and  $l_i = g_{i-1}$  is possible. The reason why we do not include moments of the distribution of the noise processes in the constants is that these distributions may depend on the number of observations  $N$  although we disclaimed on further indices.

For a choice  $K = \mathcal{O}(N^{2/3})$  the first and fourth term are of the same order  $N^{-1/3}$  and thus we see that the  $\widehat{\langle \tilde{X}, \tilde{Y} \rangle}_T^{sub}$ -estimator is consistent with rate  $N^{1/6}$ .

We summarize the properties of the subsample estimator in the following proposition.

**Proposition 4.** *If we choose  $K_N = \mathcal{O}(N^{2/3})$  the subsample estimator*

$$\widehat{\langle \tilde{X}, \tilde{Y} \rangle}_T^{sub} = \frac{1}{K_N} \sum_{i=K_N}^N (X_{g_i} - X_{l_{i-K_N}}) (Y_{\gamma_i} - Y_{\lambda_{i-K_N}})$$

*is a consistent unbiased estimator with asymptotic variance of order  $N^{-1/3}$ :*

$$\mathbb{E} \left[ \widehat{\langle \tilde{X}, \tilde{Y} \rangle}_T^{sub} - \langle \tilde{X}, \tilde{Y} \rangle_T \right] = 0, \quad (8a)$$

$$\text{Var} \left( \widehat{\langle \tilde{X}, \tilde{Y} \rangle}_T^{sub} \right) = \mathcal{O} \left( N^{-1/3} \right) . \quad (8b)$$

*Bias-variance decomposition yields that*

$$\mathbb{E} \left( \left[ \left( \widehat{\langle \tilde{X}, \tilde{Y} \rangle}_T^{sub} - \langle \tilde{X}, \tilde{Y} \rangle_T \right)^2 \right] \right)^{1/2} = \mathcal{O} \left( N^{-1/6} \right) . \quad (8c)$$

In the next section we will show that a multi-scale approach can improve this  $N^{1/6}$  rate of convergence to  $N^{1/4}$ .

## 5 Upgrading the subsample estimator using a multi-scale approach

In this section we show that using different lower frequencies for subsampling instead of one singular fixed  $K_N$  and calculating a weighted mean of the different subsample estimators leads to a more efficient consistent estimator with a better rate of convergence  $N^{1/4}$ . We calculate  $M_N$  subsample estimators using the regular sequence  $i = 1, 2, 3, \dots, M_N - 1, M_N$  instead of the fixed  $K_N$  in Section 4. We remark that there is no advantage using a more general sequence of subsample frequencies. We focus on the variance of our multi-scale estimator for the integrated covariance due to the noise terms first, which is the conditional variance given the paths of both efficient processes, and calculate noise-optimal weights that minimize this variance due to market microstructure frictions. In the following we skip the index  $N$  for  $M$ . The general multi-scale estimator

$$\widehat{\langle \tilde{X}, \tilde{Y} \rangle}_T^{mult} = \sum_{i=1}^M \alpha_i \widehat{\langle \tilde{X}, \tilde{Y} \rangle}_T^{sub, i} = \sum_{i=1}^M \alpha_i \frac{1}{i} \sum_{j=i}^N (X_{g_j} - X_{l_{j-i}}) (Y_{\gamma_j} - Y_{\lambda_{j-i}})$$

with weights  $\alpha_i$  will give a consistent estimator by choosing the weights optimally.

To determine noise-optimal weights, we will impose side conditions on the weights that simplify the minimization problem for the variance due to noise. After that, we will prove that the variance due to mixed terms is asymptotically negligible and we will calculate the discretization variance. As for the subsample estimator there is a trade-off between variance due to noise and variance due to discretization. Choosing  $M$  optimally in the way that the mean square error is minimized, we will see that the total variance is of order  $N^{-1/2}$ . An important fact is that the weights as well as the order of  $M$  and the rate of convergence of the multi-scale estimator are in line with the variance case presented by Zhang (2006a). We impose the condition

$$\sum_{i=1}^M \alpha_i = 1 , \quad (9a)$$

that ensures unbiasedness of the resulting estimator, and the auxiliary condition

$$\sum_{i=1}^M \frac{\alpha_i}{i} = 0 , \quad (9b)$$

that will guarantee that the ‘leading’ term in the variance equals zero, on the weights that gives

$$\begin{aligned}
& \sum_{i=1}^M \frac{\alpha_i}{i} \sum_{j=i}^N (\epsilon_{g_j}^X - \epsilon_{l_{j-i}}^X) (\epsilon_{\gamma_j}^Y - \epsilon_{\lambda_{j-i}}^Y) \\
&= \sum_{i=1}^M \frac{\alpha_i}{i} \left( \sum_{j=0}^N (\epsilon_{g_j}^X \epsilon_{\gamma_j}^Y + \epsilon_{l_j}^X \epsilon_{\lambda_j}^Y) + R_N - \sum_{j=i}^N (\epsilon_{g_j}^X \epsilon_{\lambda_{j-i}}^Y + \epsilon_{l_{j-i}}^X \epsilon_{\gamma_j}^Y) \right) \\
&\stackrel{(9b)}{=} \sum_{i=1}^M \frac{\alpha_i}{i} \left( R_N - \sum_{j=i}^N (\epsilon_{g_j}^X \epsilon_{\lambda_{j-i}}^Y + \epsilon_{l_{j-i}}^X \epsilon_{\gamma_j}^Y) \right)
\end{aligned}$$

with the remainder term

$$R_N = - \sum_{j=0}^{i-1} \epsilon_{g_j}^X \epsilon_{\gamma_j}^Y - \sum_{j=N-i+1}^N \epsilon_{l_j}^X \epsilon_{\lambda_j}^Y$$

that is asymptotically negligible in the sum because of Assumption 2 and  $i \leq M = o(N)$ . Hence we only have to focus on the residual term for the analysis of the variance due to noise. Define

$$U_i := - \sum_{j=i}^N (\epsilon_{g_j}^X \epsilon_{\lambda_{j-i}}^Y + \epsilon_{l_{j-i}}^X \epsilon_{\gamma_j}^Y) .$$

**Lemma 5.** *The variance of  $U_i$  satisfies for all  $i \in \{1, \dots, M\}$  the following asymptotic inequality:*

$$\text{Var}(U_i) \leq 6N \eta_X^2 \eta_Y^2 . \quad (10)$$

*Proof.* The summands have variances  $2\eta_X^2 \eta_Y^2$ . Because of Assumption 2 only covariances

$$\text{Cov} \left( (\epsilon_{g_j}^X \epsilon_{\lambda_{j-i}}^Y + \epsilon_{l_{j-i}}^X \epsilon_{\gamma_j}^Y) , (\epsilon_{g_k}^X \epsilon_{\lambda_{k-i}}^Y + \epsilon_{l_{k-i}}^X \epsilon_{\gamma_k}^Y) \right)$$

with  $k = j \pm 1$  can be non-zero. Those covariances are smaller or equal than  $2\eta_X^2 \eta_Y^2$  and we obtain the inequality by separating the variance of  $U_i$  in the sum over all variances and covariances.  $\square$

Under our assumptions the random variables  $U_i$  are not necessarily uncorrelated. For different subsample frequencies  $i$  and  $k \in \{i-1, \dots, i+1\}$  the covariances

$$\text{Cov} \left( \sum_{j=i}^N \epsilon_{g_j}^X \epsilon_{\lambda_{j-i}}^Y + \epsilon_{l_{j-i}}^X \epsilon_{\gamma_j}^Y , \sum_{r=k}^N \epsilon_{g_r}^X \epsilon_{\lambda_{r-k}}^Y + \epsilon_{l_{r-k}}^X \epsilon_{\gamma_r}^Y \right)$$

can be non-zero. In fact, only addends with  $r \in \{j-1, j, j+1\}$  and  $(j-i) = (r-k)$  **could** have non-zero covariance (if the considered maximum **and** minimum are equal) and hence correlation effects will be very small but in any case a mathematical analysis gives an upper bound and the exact asymptotic order using an inequality similar to those in the last section.

**Lemma 6.** *For the variance of the general multi-scale estimator due to noise the asymptotic inequality*

$$\text{Var} \left( \sum_{i=1}^M \frac{\alpha_i}{i} \sum_{j=i}^N (\epsilon_{g_j}^X - \epsilon_{l_{j-i}}^X) (\epsilon_{\gamma_j}^Y - \epsilon_{\lambda_{j-i}}^Y) \right) \leq \sum_{i=1}^M \frac{\alpha_i^2}{i^2} 18 N \eta_X^2 \eta_Y^2 \quad (11)$$

*holds.*

*Proof.* Applying the Cauchy-Schwarz inequality to the covariance terms considered above and using inequality (10) we conclude that

$$\begin{aligned} \mathbb{V}\text{ar} \left( \sum_{i=1}^M \frac{\alpha_i}{i} \sum_{j=i}^N (\epsilon_{g_j}^X - \epsilon_{l_{j-i}}^X) (\epsilon_{\gamma_j}^Y - \epsilon_{\lambda_{j-i}}^Y) \right) &\leq \sum_{i=1}^M \frac{\alpha_i^2}{i^2} 3 \mathbb{V}\text{ar} (U_i) \\ &\leq \sum_{i=1}^M \frac{\alpha_i^2}{i^2} 18 \eta_X^2 \eta_Y^2 (N + o(N)) \end{aligned}$$

which gives the result of Lemma 6.  $\square$

Minimization with side conditions yields for an arbitrary constant  $c \in \mathbb{R}$ :

$$\begin{aligned} \frac{\partial}{\partial \alpha_i} \left( c \sum_j \frac{\alpha_j^2}{j^2} + \lambda_1 \left( \sum_j \alpha_j - 1 \right) + \lambda_2 \left( \sum_j \frac{\alpha_j}{j} \right) \right) &= 0 \\ \Leftrightarrow 2c \frac{\alpha_i}{i^2} + \lambda_1 + \frac{\lambda_2}{i} &= 0 \\ \Leftrightarrow \alpha_i &= -\frac{1}{2c} (i^2 \lambda_1 + \lambda_2 i) . \end{aligned}$$

Since  $1 = \sum \alpha_i = -\frac{1}{2c} (\lambda_1 \sum i^2 + \lambda_2 \sum i)$  and  $0 = \sum \frac{\alpha_i}{i} = -\frac{1}{2c} (\lambda_1 \sum i + \lambda_2 M)$  we get the result

$$\lambda_1 = \frac{-24c}{M^3 - M} , \quad \lambda_2 = \frac{12c}{(M-1)M} , \quad \alpha_i = \frac{12i^2}{M^3 - M} - \frac{6i}{(M-1)M} .$$

The noise-optimal weights are the same as for the MSRV-estimator invented by Zhang (2006a) for efficient high-frequency realized variance estimation from noisy observations. This is a positive aspect of using these methods because the weights are calculated once and serve for variance as well as covariance estimation.

Inserting the noise-optimal weights

$$\alpha_{i,opt} = \frac{12i^2}{M^3} - \frac{6i}{M^2} (1 + o(1)) \quad (12)$$

in the noise-variance term above yields the result stated in the following proposition.

**Proposition 7.** *If we insert the noise-optimal weights  $\alpha_{i,opt}$  from (12) in the general multi-scale estimator and assume that the variances  $\eta_X^2, \eta_Y^2$  of the noise distributions are of order 1, the asymptotic variance due to noise satisfies*

$$\mathbb{V}\text{ar} \left( \sum_{i=1}^M \frac{\alpha_i}{i} \sum_{j=i}^N (\epsilon_{g_j}^X - \epsilon_{\lambda_{j-i}}^Y) (\epsilon_{l_{j-i}}^X - \epsilon_{\gamma_j}^Y) \right) = o \left( \frac{N}{M^3} \right) . \quad (13)$$

*Proof.* We calculate the minimal noise-variance using the optimal weights (12). For the occurring sums it suffices to use the asymptotic formula  $\sum_{i=1}^l i^k = \frac{l^{k+1}}{k+1} + o(l^{k+1})$  and Lemma 5 and 6 to deduce the asymptotic order of the variance:

$$\begin{aligned} \sum_{i=1}^M \frac{(\alpha_{i,opt})^2}{i^2} \eta_X^2 \eta_Y^2 N &= \sum_{i=1}^M \left( \frac{144i^2}{M^6} - \frac{144i}{M^5} + \frac{36}{M^4} \right) \eta_X^2 \eta_Y^2 N \\ &= \frac{12N}{M^3} \eta_X^2 \eta_Y^2 + o \left( \frac{N}{M^3} \right) . \end{aligned}$$

$\square$

We have shown that our resulting multi-scale estimator with noise-optimal weights has a variance due to noise contamination of asymptotic order  $N/M^3$ . Next we focus on the other terms occurring in the total variance.

Under the stated assumptions the multi-scale estimator for integrated covariance is unbiased and the variance induced by the mixed summands is asymptotically negligible. Unbiasedness holds by Condition (9a) on the weights and we focus on the variance of the mixed summands now.

The analysis of the terms with the  $\mathbb{M}_i$ s and  $\mathbb{D}_i$ s (see (2) for definition) is analogous and we only mention the analysis of the first term. Inserting the noise-optimal weights (12) the variance equals

$$\begin{aligned} & \text{Var} \left( \sum_{i=1}^M \frac{\alpha_{i,opt}}{i} \sum_{j=i}^N (\tilde{X}_{g_j} - \tilde{X}_{l_{j-i}}) (\epsilon_{\gamma_j}^Y - \epsilon_{\lambda_{j-i}}^Y) \right) \\ &= 2 \sum_{i=1}^M \sum_{k=1}^i \left( \frac{144ik}{M^6} + \frac{36}{M^4} - \frac{72i}{M^5} - \frac{72k}{M^5} \right) \\ & \quad \times \sum_{j=i}^N \sum_{r=k}^N \text{Cov} \left( (\tilde{X}_{g_j} - \tilde{X}_{l_{j-i}}) (\epsilon_{\gamma_j}^Y - \epsilon_{\lambda_{j-i}}^Y), (\tilde{X}_{g_r} - \tilde{X}_{l_{r-k}}) (\epsilon_{\gamma_r}^Y - \epsilon_{\lambda_{r-k}}^Y) \right). \end{aligned}$$

The covariances can only be non-zero if the time intervals of the increments of the efficient process  $\tilde{X}$  are overlapping and because of Assumption 2 if  $r \in \{j-1, j, j+1\}$  or  $(r-k) = (j-i)$  holds. Therefore, we conclude that a constant  $C^*$  exists such that

$$2 \sum_{i=1}^M \sum_{k=1}^i \left( \frac{144ik}{M^6} + \frac{36}{M^4} - \frac{72i}{M^5} - \frac{72k}{M^5} \right) N \frac{i}{N} 2 C^* \eta_Y^2$$

is an upper bound for the variance and we obtain the asymptotic order  $1/M$  for the mixed terms. We have deduced that the mixed terms are negligible in the asymptotic total variance and hence we focus next on the terms containing the  $\mathbb{C}_i$ s (see (2)) and the variance due to discretization.

Thus the variance term of interest is

$$\begin{aligned} & \text{Var} \left( \sum_{i=1}^M \frac{\alpha_i}{i} \sum_{j=i}^N (\tilde{X}_{g_j} - \tilde{X}_{l_{j-i}}) (\tilde{Y}_{\gamma_j} - \tilde{Y}_{\lambda_{j-i}}) \right) = \\ & 2 \sum_{i=1}^M \sum_{k=1}^i \frac{\alpha_i \alpha_k}{ik} \text{Cov} \left( \sum_{j=i}^N (\tilde{X}_{g_j} - \tilde{X}_{l_{j-i}}) (\tilde{Y}_{\gamma_j} - \tilde{Y}_{\lambda_{j-i}}), \right. \\ & \quad \left. \sum_{r=k}^N (\tilde{X}_{g_r} - \tilde{X}_{l_{r-k}}) (\tilde{Y}_{\gamma_r} - \tilde{Y}_{\lambda_{r-k}}) \right). \end{aligned}$$

Considering next the single covariance terms in the discretization-variance using

$$\mathbb{E} \left[ (\tilde{X}_{t_i} - \tilde{X}_{t_{i-l}}) (\tilde{X}_{t_j} - \tilde{X}_{t_{j-k}}) \right] = \mathbb{E} \left( \tilde{X}_{t_{\min(i,j)}} - \tilde{X}_{t_{\max(i-l, j-k)}} \right)^2 \mathbb{1}_{\{\min(i,j) > \max(i-l, j-k)\}}$$

for arbitrary  $i, j$  and  $l, k$  leads to

$$\begin{aligned} & \text{Var} \left( \sum_{i=1}^M \frac{\alpha_i}{i} \sum_{j=i}^N (\tilde{X}_{g_j} - \tilde{X}_{l_{j-i}}) (\tilde{Y}_{\gamma_j} - \tilde{Y}_{\lambda_{j-i}}) \right) \\ & \leq 2 \sum_{k=1}^M \sum_{l=1}^k \left( \frac{144lk}{M^6} + \frac{36}{M^4} - \frac{72l}{M^5} - \frac{72k}{M^5} \right) C^* \cdot \frac{l^2 k}{N} \leq C^{**} \frac{M}{N} \end{aligned}$$



with constants  $C^*$  and  $C^{**}$ . The inequality is deduced in the usual way analyzing which increments are overlapping and hence correlated and using the asymptotic orders of the increments known by Assumptions 1 and 3.

**Proposition 8.** *For the variance of the noise-optimal multi-scale estimator due to discretization the following asymptotic inequality holds:*

$$\mathbb{V}\text{ar} \left( \sum_{i=1}^M \frac{\alpha_i}{i} \sum_{j=i}^N (\tilde{X}_{g_j} - \tilde{X}_{l_{j-i}}) (\tilde{Y}_{\gamma_j} - \tilde{Y}_{\lambda_{j-i}}) \right) = \mathcal{O} \left( \frac{M}{N} \right). \quad (14)$$

The discretization variance is of order  $M/N$  and we have to choose

$$M = \mathcal{O}(\sqrt{N})$$

to reduce the total variance to order  $1/M$  or rather  $1/\sqrt{N}$ . There is a trade-off between the variance terms due to microstructure noise and discretization and the total variance is minimized by a choice of  $M$  that induces both being of the same asymptotic order. Calculating  $M = \mathcal{O}(\sqrt{N})$  different subsample estimators and calculating the weighted sum with noise-optimal weights results in obtaining an estimator with asymptotic total variance of order  $1/\sqrt{N}$  upgrading the rate of convergence to  $N^{1/4}$  compared with  $N^{1/6}$  for the simple one scale estimator presented in Section 3. Although the new estimator requires more computing time, the new estimator gains a higher efficiency in covariance estimation in the case of high-frequency noisy observations.

We present the results derived in this section again in the following proposition which implies Theorem 1.

**Proposition 9.** *If we choose  $M_N = \mathcal{O}(\sqrt{N})$  and calculate the noise-optimal multi-scale estimator for the integrated covariance*

$$\widehat{\langle \tilde{X}, \tilde{Y} \rangle}_T^{mult} = \sum_{i=1}^{M_N} \left( \frac{12i}{M_N^3} - \frac{6}{M_N^2} \right) \sum_{j=i}^N (X_{g_j} - X_{l_{j-i}}) (Y_{\gamma_j} - Y_{\lambda_{j-i}}), \quad (15)$$

*we obtain a consistent unbiased estimator with asymptotic variance of order  $M^{-1} = N^{-1/2}$ :*

$$\mathbb{E} \left[ \widehat{\langle \tilde{X}, \tilde{Y} \rangle}_T^{mult} - \langle \tilde{X}, \tilde{Y} \rangle_T \right] = 0 \quad (16a)$$

$$\mathbb{V}\text{ar} \left( \widehat{\langle \tilde{X}, \tilde{Y} \rangle}_T^{mult} \right) = \mathcal{O} \left( \frac{1}{M} \right) = \mathcal{O} \left( \sqrt{\frac{1}{N}} \right). \quad (16b)$$

*The bias-variance decomposition yields that*

$$\left( \mathbb{E} \left[ \left( \widehat{\langle \tilde{X}, \tilde{Y} \rangle}_T^{mult} - \langle \tilde{X}, \tilde{Y} \rangle_T \right)^2 \right] \right)^{1/2} = \mathcal{O} \left( N^{-1/4} \right). \quad (16c)$$

*We will prove the rate of convergence  $N^{1/4}$  to be optimal in the following section.*

**Remark 10.** *We suppose that an extension of Proposition 9 for non-i. i. d. noise is possible under a milder assumption of exponentially decreasing mixing coefficients such that the equations (16a)-(16c) still hold. This extension for the one-dimensional case has been developed in Ait-Sahalia et al. (2005).*

## 6 A lower bound for the rate of convergence

In the following we show the LAN (local asymptotic normality) property for a constant correlation coefficient  $\rho = \text{corr}(B^X, B^Y)$  of the two Brownian motions of  $\tilde{X}$  and  $\tilde{Y}$  with rate  $N^{-1/4}$  within the following simplified model and conclude the rate-optimality of our estimator defined in the last chapter.

We have the observations:

$$\begin{aligned} X_{t_i} &= \int_0^{t_i} dB_t^X + \epsilon_{t_i}^X \\ Y_{t_i} &= \int_0^{t_i} dB_t^Y + \epsilon_{t_i}^Y \quad i = 0, \dots, N. \end{aligned}$$

We restrict ourselves to synchronous observations and equidistant time intervals  $\Delta t_i = \Delta t = 1/N$ . Furthermore we assume the discrete noise processes to be independent of the efficient processes and independent to each other (as in Assumption 2 before). We strengthen the i.i.d. assumption for the noise to an i.i.d. -Gaussian assumption:

$$\epsilon_{t_i}^X \stackrel{iid}{\sim} \mathcal{N}(0, \eta_X^2), \quad \epsilon_{t_i}^Y \stackrel{iid}{\sim} \mathcal{N}(0, \eta_Y^2), \quad i = 0, \dots, N.$$

We want to estimate the parameter  $\rho$  from observed increments  $(\Delta X_{t_1}, \dots, \Delta X_{t_N}, \Delta Y_{t_1}, \dots, \Delta Y_{t_N})$  taking values in a measurable space  $(\Omega_{2N}, \mathcal{F}_{2N})$  with law  $\mathbb{P}_\rho^{2N}$ . Local asymptotic normality with rate  $N^{-1/4}$  means that for a real sequence  $h_N \rightarrow h$  the sequence of log-likelihoods converges in law to a limit of the following form:

$$\log \left( \frac{d\mathbb{P}^{2N}_{\rho + N^{-1/4} h_N}}{d\mathbb{P}_\rho^{2N}} \right) \xrightarrow{\mathbb{P}_\rho^{2N}} hZ\sqrt{I(\rho)} - \frac{h^2 I(\rho)}{2}$$

with  $Z \sim \mathcal{N}(0, 1)$  and  $I(\rho)$  denoting the Fisher information. Then the limit distribution of a sequence of estimators  $\hat{\rho}_{2N}$  is under regularity conditions the convolution of a Gaussian distribution and a noise factor. The maximum risk of any estimator is bounded below by the Gaussian risk and the minimax theorem gives the result on how well the parameter can be estimated asymptotically. See e. g. van der Vaart (1998) and van der Vaart and Wellner (1996) for further information on LAN and optimal convergence rates.

We summarize the results of this section in the following Proposition 11:

**Proposition 11.** *In the simple model of two synchronously equidistantly observed standard Brownian motions  $\tilde{X}$  and  $\tilde{Y}$  with constant correlation  $\rho$  and an observation noise described by i.i.d. Gaussian errors with standard deviations  $\eta_X$  and  $\eta_Y$  the LAN property with  $N^{-1/4}$ -rate holds, where  $N$  denotes the number of observations in the interval  $[0, 1]$ . Assuming without loss of generality  $\eta_X \geq \eta_Y$ , we obtain the following lower and upper bound for the asymptotic Fisher information:*

$$\frac{1}{8\eta_X} \left( \frac{1}{(1+\rho)^{3/2}} + \frac{1}{(1-\rho)^{3/2}} \right) \leq I(\rho) \leq \frac{\sqrt{2}}{8} \frac{1}{\sqrt{\eta_X^2 + \eta_Y^2}} \left( \frac{1}{(1+\rho)^{3/2}} + \frac{1}{(1-\rho)^{3/2}} \right). \quad (17)$$

Particularly assuming the variance of both noise processes to be equal ( $\eta_X = \eta_Y = \eta$ ) we can calculate the exact asymptotic Fisher information. It is given by

$$I(\rho) = \frac{1}{8\eta} \left( \frac{1}{(1+\rho)^{3/2}} + \frac{1}{(1-\rho)^{3/2}} \right). \quad (18)$$

Proposition 11 implies Theorem 2 and gives, furthermore, bounds for the asymptotic Fisher information.

**Remark 12.** *We prove the LAN property with rate  $N^{-1/4}$  in this simplified model and thus the optimality of our multi-scale estimator. It has the optimal rate of convergence even in the synchronous equidistant case. The asymptotic Fisher information is enclosed between the ‘natural’ lower and an intuitive upper bound. We state that the Fisher information (18) has the following asymptotic behaviour:*

$$I(\rho) \rightarrow \infty \text{ for } \rho \rightarrow \pm 1 \text{ and } I(\rho) \rightarrow 0 \text{ for } \eta \rightarrow \infty .$$

*Proof.* First we will prove the LAN property for the simpler case of equal noise variances  $\eta_X = \eta_Y = \eta$  and calculate the asymptotic Fisher information (18). We want to derive the distribution of the increments

$$\begin{aligned} \Delta X_{t_i} &= \int_{t_{i-1}}^{t_i} dB_t^X + \epsilon_{t_i}^X - \epsilon_{t_{i-1}}^X , \\ \text{and } \Delta Y_{t_i} &= \int_{t_{i-1}}^{t_i} dB_t^Y + \epsilon_{t_i}^Y - \epsilon_{t_{i-1}}^Y . \end{aligned}$$

The constant correlation parameter is denoted by  $\theta$  in the following. There exists a Brownian motion  $B$  independent of  $\tilde{X}$  such that the following equation holds:

$$\Delta Y_{t_i} = \int_{t_{i-1}}^{t_i} \theta dB_t^X + \sqrt{1-\theta^2} \int_{t_{i-1}}^{t_i} dB_t + \epsilon_{t_i}^Y - \epsilon_{t_{i-1}}^Y .$$

Taking this into account we can easily calculate the covariations of the increments:

$$\mathbb{C}\text{ov}(\Delta X_{t_i}, \Delta X_{t_j}) = \mathbb{C}\text{ov}(\Delta Y_{t_i}, \Delta Y_{t_j}) = \begin{cases} \Delta t + 2\eta^2 & \text{if } i = j \\ -\eta^2 & \text{if } |i - j| = 1 \\ 0 & \text{if } |i - j| > 1 \end{cases} ,$$

$$\mathbb{C}\text{ov}(\Delta X_{t_i}, \Delta Y_{t_j}) = \begin{cases} \theta \Delta t & \text{if } i = j \\ 0 & \text{if } i \neq j \end{cases} .$$

The random vector  $(\Delta X_{t_1}, \dots, \Delta X_{t_N}, \Delta Y_{t_1}, \dots, \Delta Y_{t_N})^t$  has a  $2N \times 2N$  dimensional covariance matrix

$$\Sigma_\theta = \begin{pmatrix} A_N & D_N \\ D_N & A_N \end{pmatrix}$$

with the  $N \times N$  diagonal matrix

$$D_N = \begin{pmatrix} \theta \Delta t & 0 & \dots & 0 \\ 0 & \ddots & & \vdots \\ \vdots & & \ddots & 0 \\ 0 & \dots & 0 & \theta \Delta t \end{pmatrix}$$

and the  $N \times N$  tridiagonal 1-Toeplitz matrix

$$A_N = \begin{pmatrix} \Delta t + 2\eta^2 & -\eta^2 & 0 & \dots & 0 \\ -\eta^2 & \ddots & \ddots & & \vdots \\ 0 & \ddots & \ddots & \ddots & 0 \\ \vdots & & \ddots & \ddots & -\eta^2 \\ 0 & \dots & 0 & -\eta^2 & \Delta t + 2\eta^2 \end{pmatrix}.$$

This special structure of the covariance matrix makes it possible to explicitly compute the eigenvalues of  $\Sigma_\theta$ . Here the fact that we assumed the variances of both noise processes to be equal plays an important role.

We write the  $N$ -dimensional identity matrix as  $\mathbb{1}_N$ . Then the characteristic polynomial of  $\Sigma_\theta$  can be written as

$$\det(\Sigma_\theta - \lambda \mathbb{1}_{2N}) = (\det(A_N - \lambda \mathbb{1}_N))^2 - (\theta \Delta t)^{2N}.$$

Using a Laplace-expansion, the characteristic polynomials of  $A_N$  can be computed by a recursion:

$$\begin{aligned} \det(A_N - \lambda \mathbb{1}_N) &= (\Delta t + 2\eta^2 - \lambda) \det(A_{N-1} - \lambda \mathbb{1}_{N-1}) + (\eta^2)^2 \det(A_{N-2} - \lambda \mathbb{1}_{N-2}) \\ &= \sum_{k=0}^{\lfloor \frac{N}{2} \rfloor} (-1)^k \binom{N-k}{k} (\Delta t + 2\eta^2 - \lambda)^{N-2k} (\eta^2)^{2k}. \end{aligned}$$

The eigenvalues of  $A_N$  are  $\lambda_{i,N} = \Delta t + 2\eta^2 \left(1 - \cos \frac{i\pi}{N+1}\right)$ ,  $i = 1, \dots, N$ , and because of the simple structure of  $\Sigma_\theta$  we can deduce the  $2N$  eigenvalues of the covariance matrix directly:

$$\lambda_{i,N}^+(\theta) = \Delta t(1 + \theta) + 2\eta^2 \left(1 - \cos \frac{i\pi}{N+1}\right), \quad i = 1, \dots, N, \quad (19a)$$

$$\lambda_{i,N}^-(\theta) = \Delta t(1 - \theta) + 2\eta^2 \left(1 - \cos \frac{i\pi}{N+1}\right), \quad i = 1, \dots, N. \quad (19b)$$

With the notation

$$\lambda_{j,2N}(\theta) = \begin{cases} \lambda_{i,N}^+ & \text{if } j = 2i - 1, \quad i = 1, \dots, N, \\ \lambda_{i,N}^- & \text{if } j = 2i, \quad i = 1, \dots, N \end{cases}, \quad (19c)$$

we can write the  $2N \times 2N$  diagonal matrix of the eigenvalues as  $\Lambda_\theta^{2N}$  with  $(\Lambda_\theta^{2N})_{jj} = \lambda_{j,2N}(\theta)$ .  $\Sigma_\theta$  can be diagonalized by an  $2N \times 2N$  orthogonal matrix  $P^{2N}$  which is independent of  $\theta$ . The random vector  $P^{2N} \cdot (\Delta X_{t_1}, \dots, \Delta X_{t_N}, \Delta Y_{t_1}, \dots, \Delta Y_{t_N})^t$  is centered Gaussian with covariance matrix  $\Lambda_\theta^{2N}$ . We define the  $2N$ -dimensional random vector  $T^{2N}$  by

$$\begin{aligned} (T^{2N})_j &:= \frac{1}{\sqrt{\lambda_{j,2N}(\rho)}} \left( P^{2N} \cdot (\Delta X_{t_1}, \dots, \Delta X_{t_N}, \Delta Y_{t_1}, \dots, \Delta Y_{t_N})^t \right)_j \\ &\sim \mathcal{N} \left( 0, \frac{\lambda_{j,2N}(\theta)}{\lambda_{j,2N}(\rho)} \right). \end{aligned}$$

To prove the LAN property we have to examine the log-likelihood

$$\begin{aligned}
& \log \left( \frac{d\mathbb{P}^{2N}}{\rho + N^{-\frac{1}{4}}h_N} \right) \\
&= \log \left[ \left( \frac{\prod_{j=1}^{2N} \lambda_{j,2N}(\rho + N^{-\frac{1}{4}}h_N)}{\prod_{j=1}^{2N} \lambda_{j,2N}(\rho)} \right)^{-\frac{1}{2}} \right] - \frac{1}{2} \sum_{j=1}^{2N} (T^{2N})_j^2 \left( \frac{\lambda_{j,2N}(\rho)}{\lambda_{j,2N}(\rho + N^{-\frac{1}{4}}h_N)} - 1 \right) \\
&= -\frac{1}{2} \sum_{j=1}^{2N} \left( \log(1 + \gamma_j^{2N}) - (T^{2N})_j^2 \frac{\gamma_j^{2N}}{\gamma_j^{2N} + 1} \right)
\end{aligned}$$

where

$$\gamma_j^{2N} := \frac{\lambda_{j,2N}(\rho + N^{-1/4}h_N)}{\lambda_{j,2N}(\rho)} - 1 = \frac{\Delta t \cdot N^{-1/4}h_N}{\lambda_{j,2N}(\rho)}.$$

The proof is now analogous to the one dimensional case (see Gloter and Jacod (2001)) and using Theorem VIII-3.32 in Jacod and Shiryaev (2003) it remains to show that

$$\sup_{1 \leq j \leq 2N} |\gamma_j^{2N}| \rightarrow 0 \quad \text{and} \quad \sum_{j=1}^{2N} (\gamma_j^{2N})^2 \rightarrow 2h^2 I(\rho). \quad (20)$$

The first condition is obviously fulfilled. To prove the second one we write the sum of the squares as a Riemann sum and use an inequality including the corresponding integral:

$$\begin{aligned}
\sum_{j=1}^{2N} (\gamma_j^{2N})^2 &= \sum_{j=1}^N \frac{N^{-1/2}h_N^2}{\left(1 + \rho + \frac{2\eta^2}{\Delta t} \left(1 - \cos \frac{j\pi}{N+1}\right)\right)^2} + \sum_{j=1}^N \frac{N^{-1/2}h_N^2}{\left(1 - \rho + \frac{2\eta^2}{\Delta t} \left(1 - \cos \frac{j\pi}{N+1}\right)\right)^2} \\
&= \frac{N^{1/2}h_N^2 (\Delta t)^2}{(\eta^2)^2 \pi} \underbrace{\frac{\pi}{N} \sum_{j=1}^N \frac{1}{\left(2 \left(1 - \cos \frac{j\pi}{N+1}\right) + \frac{\Delta t(1+\rho)}{\eta^2}\right)^2}}_{=S_N} \\
&\quad + \frac{N^{1/2}h_N^2 (\Delta t)^2}{(\eta^2)^2 \pi} \underbrace{\frac{\pi}{N} \sum_{j=1}^N \frac{1}{\left(2 \left(1 - \cos \frac{j\pi}{N+1}\right) + \frac{\Delta t(1-\rho)}{\eta^2}\right)^2}}_{=\tilde{S}_N}.
\end{aligned}$$

For the integral

$$J = \int_0^\pi \frac{1}{\left(2(1 - \cos z) + \frac{\Delta t(1+\rho)}{\eta^2}\right)^2} dz$$

and accordingly

$$\tilde{J} = \int_0^\pi \frac{1}{\left(2(1 - \cos z) + \frac{\Delta t(1-\rho)}{\eta^2}\right)^2} dz,$$

the following inequalities with the lower and upper Darboux sums hold:

$$\frac{\pi}{N} \sum_{j=1}^N \frac{1}{\left(2 \left(1 - \cos \frac{(j+1)\pi}{N+1}\right) + \frac{\Delta t(1+\rho)}{\eta^2}\right)^2} \leq J \leq \frac{\pi}{N} \sum_{j=1}^N \frac{1}{\left(2 \left(1 - \cos \frac{j\pi}{N+1}\right) + \frac{\Delta t(1+\rho)}{\eta^2}\right)^2}$$

and accordingly

$$\frac{\pi}{N} \sum_{j=1}^N \frac{1}{\left(2 \left(1 - \cos \frac{(j+1)\pi}{N+1}\right) + \frac{\Delta t(1-\rho)}{\eta^2}\right)^2} \leq \tilde{J} \leq \frac{\pi}{N} \sum_{j=1}^N \frac{1}{\left(2 \left(1 - \cos \frac{j\pi}{N+1}\right) + \frac{\Delta t(1-\rho)}{\eta^2}\right)^2}.$$

Thus the following inequalities hold for the Riemann sums  $S_N$  and  $\tilde{S}_N$ , respectively:

$$J \leq S_N \leq J + \frac{\pi}{N} \frac{1}{N \left(4 + \frac{\Delta t(1+\rho)}{\eta^2}\right)^2} - \frac{\pi}{N} \frac{1}{\left(2 \left(1 - \cos \frac{\pi}{N+1}\right) + \frac{\Delta t+(1+\rho)}{\eta^2}\right)^2}$$

and

$$\tilde{J} \leq \tilde{S}_N \leq \tilde{J} + \frac{\pi}{N} \frac{1}{N \left(4 + \frac{\Delta t(1-\rho)}{\eta^2}\right)^2} - \frac{\pi}{N} \frac{1}{\left(2 \left(1 - \cos \frac{\pi}{N+1}\right) + \frac{\Delta t+(1-\rho)}{\eta^2}\right)^2}.$$

The integrals can be computed explicitly:

$$\frac{N^{-3/2} h_N^2}{(\eta^2)^2 \pi} (J + \tilde{J}) = \frac{N^{-3/2} h_N^2}{(\eta^2)^2 \pi} \left[ \frac{\pi \left(2 + \frac{\Delta t(1+\rho)}{\eta^2}\right)}{\left(\frac{\Delta t(1+\rho)}{\eta^2} \left(\frac{\Delta t(1+\rho)}{\eta^2} + 4\right)\right)^{3/2}} + \frac{\pi \left(2 + \frac{\Delta t(1-\rho)}{\eta^2}\right)}{\left(\frac{\Delta t(1-\rho)}{\eta^2} \left(\frac{\Delta t(1-\rho)}{\eta^2} + 4\right)\right)^{3/2}} \right].$$

Since  $h_N \rightarrow h$ , we can deduce from the preceding inequalities for both summands the convergence

$$\sum_{j=1}^{2N} (\gamma_j^{2N})^2 \rightarrow \frac{h^2}{4\eta} \left( \frac{1}{(1+\rho)^{3/2}} + \frac{1}{(1-\rho)^{3/2}} \right) = 2h^2 I(\rho) \quad (21)$$

with the Fisher information

$$I(\rho) = \frac{1}{8\eta} \left( \frac{1}{(1+\rho)^{3/2}} + \frac{1}{(1-\rho)^{3/2}} \right). \quad (22)$$

We continue the proof with the generalization for different noise variances. If the noise variances are not equal  $\eta_X^2 \neq \eta_Y^2$ , the covariance matrix can be written as

$$\Sigma_\theta = \begin{pmatrix} A_N & D_N \\ D_N & B_N \end{pmatrix}$$

with the same diagonal matrix  $D_N$  as before and two tridiagonal 1-Toeplitz matrices  $A_N$  and  $B_N$  with the same structure as before where  $A_N$  has the entries  $\Delta t + 2\eta_X^2$  on the main diagonal and correspondingly,  $B_N$  the entries  $\Delta t + 2\eta_Y^2$ . The eigenvalues of  $A_N$  and  $B_N$  have been deduced before and are denoted by  $\lambda_X^{(i,N)}$  and  $\lambda_Y^{(i,N)}$  here, which emphasizes the dependence on  $\eta_X$  and  $\eta_Y$ , respectively. Because of the special structure of  $A_N$  and  $B_N$ , that are in particular symmetric and commutative, they share the same eigenvectors  $v_i, i = 1, \dots, N$ . We can calculate the  $2N$  eigenvalues of  $\Sigma_\theta$ , denoted by  $\xi_+^{(i)}, \xi_-^{(i)}, i = 1, \dots, N$ , using the approach

$$\Sigma_\theta = \begin{pmatrix} A_N & D_N \\ D_N & B_N \end{pmatrix} \cdot \begin{pmatrix} \alpha v_i \\ \beta v_i \end{pmatrix} = \xi \begin{pmatrix} \alpha v_i \\ \beta v_i \end{pmatrix}$$

for the eigenvectors with constants  $\alpha$  and  $\beta$ . This equation implies that

$$\alpha \lambda_X^{(i,N)} + \Delta t \theta \beta = \alpha \xi,$$

$$\alpha \Delta t \theta + \beta \lambda_Y^{(i,N)} = \beta \xi,$$

and by solving this system of equations we obtain the  $2N$  eigenvalues

$$\xi_+^{(i)} = \frac{\lambda_X^{(i)} + \lambda_Y^{(i)}}{2} + \sqrt{\left(\frac{\lambda_X^{(i)} - \lambda_Y^{(i)}}{2}\right)^2 + \theta^2 (\Delta t)^2},$$

$$\xi_-^{(i)} = \frac{\lambda_X^{(i)} + \lambda_Y^{(i)}}{2} - \sqrt{\left(\frac{\lambda_X^{(i)} - \lambda_Y^{(i)}}{2}\right)^2 + \theta^2 (\Delta t)^2}.$$

We have dropped the index  $N$  of the eigenvalues here.

**Lemma 13.** *If we assume  $\eta_X > \eta_Y$ , the following inequalities hold:*

$$\frac{\lambda_X^{(i)} + \lambda_Y^{(i)}}{2} + \theta \Delta t < \xi_+^{(i)} < \lambda_X^{(i)} + \theta \Delta t, \quad (23a)$$

$$\lambda_X^{(i)} - \theta \Delta t < \xi_-^{(i)} < \frac{\lambda_X^{(i)} + \lambda_Y^{(i)}}{2} - \theta \Delta t. \quad (23b)$$

*Proof.* If  $\eta_X > \eta_Y$  for the eigenvalues  $\lambda_X^{(i)} > \lambda_Y^{(i)}$  holds for all  $i \in \{1, \dots, N\}$ . Thus

$$\xi_+^{(i)} < \frac{\lambda_X^{(i)} + \lambda_Y^{(i)}}{2} + \sqrt{\left(\frac{\lambda_X^{(i)} - \lambda_Y^{(i)}}{2}\right)^2 + (\lambda_X^{(i)} - \lambda_Y^{(i)}) \theta \Delta t + \theta^2 (\Delta t)^2} = \lambda_X^{(i)} + \theta \Delta t$$

holds and analogously the lower bound for  $\xi_-^{(i)}$  is obtained by adding the mixed term to the expression under the square root. The other bounds are obvious.  $\square$

In the following we define

$$\gamma_+^{(i)} = \frac{\xi_+^{(i)} (\rho + N^{-1/4} h_N)}{\xi_+^{(i)} (\rho)} - 1 > 0 \quad \text{and}$$

$$\gamma_-^{(i)} = \frac{\xi_-^{(i)} (\rho + N^{-1/4} h_N)}{\xi_-^{(i)} (\rho)} - 1 < 0$$

in analogy to the case of equal noise variances. We use the preceding lemma to obtain bounds for these coefficients and show the LAN property with the same rate  $N^{-1/4}$  as above, including bounds for the Fisher information.

**Proposition 14.** *If  $\eta_X > \eta_Y$  the following inequalities hold:*

$$\frac{N^{-1/4} h_N \Delta t + \frac{\lambda_Y^{(i)} - \lambda_X^{(i)}}{2}}{\lambda_X^{(i)} + \rho \Delta t} < \gamma_+^{(i)} < \frac{N^{-1/4} h_N \Delta t}{\frac{\lambda_X^{(i)} + \lambda_Y^{(i)}}{2} + \rho \Delta t} \quad (24a)$$

and

$$\frac{-N^{-\frac{1}{4}}h_N\Delta t}{\frac{\lambda_X^{(i)}+\lambda_Y^{(i)}}{2}-\rho\Delta t} < \gamma_-^{(i)} < \frac{-N^{-\frac{1}{4}}h_N\Delta t + \frac{\lambda_Y^{(i)}-\lambda_X^{(i)}}{2}}{\lambda_X^{(i)}-\rho\Delta t}. \quad (24b)$$

*Proof.* Using the inequality (23a) in the preceding Lemma 13 we obtain the lower bound for  $\gamma_+^{(i)}$ . From

$$\begin{aligned} \gamma_+^{(i)} &= \frac{\frac{\lambda_X^{(i)}+\lambda_Y^{(i)}}{2} + \sqrt{\left(\frac{\lambda_X^{(i)}-\lambda_Y^{(i)}}{2}\right)^2 + \left(\rho + N^{-\frac{1}{4}}h_N\right)^2 (\Delta t)^2}}{\xi_+^{(i)}(\rho)} - 1 \\ &< \frac{\frac{\lambda_X^{(i)}+\lambda_Y^{(i)}}{2} + \sqrt{\left(\frac{\lambda_X^{(i)}-\lambda_Y^{(i)}}{2}\right)^2 + \rho^2 (\Delta t)^2 + N^{-\frac{1}{2}}h_N^2 (\Delta t)^2 + 2N^{-\frac{1}{4}}h_N\Delta t \sqrt{\left(\frac{\lambda_X^{(i)}-\lambda_Y^{(i)}}{2}\right)^2 + \rho^2 (\Delta t)^2}}}{\xi_+^{(i)}(\rho)} - 1 \\ &= \frac{\frac{\lambda_X^{(i)}+\lambda_Y^{(i)}}{2} + \sqrt{\left(\frac{\lambda_X^{(i)}-\lambda_Y^{(i)}}{2}\right)^2 + (\rho\Delta t)^2 + N^{-\frac{1}{4}}h_N\Delta t}}{\xi_+^{(i)}(\rho)} - 1 \\ &= \frac{N^{-\frac{1}{4}}h_N\Delta t}{\xi_+^{(i)}(\rho)} < \frac{N^{-\frac{1}{4}}h_N\Delta t}{\frac{\lambda_X^{(i)}+\lambda_Y^{(i)}}{2} + \rho\Delta t} \end{aligned}$$

we can deduce the upper bound using again the right-hand side of inequality (23a) in the last inequality. The bounds for  $\gamma_-^{(i)}$  follow analogously.  $\square$

Now we are able to prove the LAN property in the same way as for the case of equal noise variances using the preceding inequalities. Because of Proposition 14, the inequalities

$$\sum_{i=1}^N \left(\gamma_+^{(i)}\right)^2 + \sum_{i=1}^N \left(\gamma_-^{(i)}\right)^2 < \sum_{i=1}^N \left( \frac{N^{-\frac{1}{2}}h_N^2 (\Delta t)^2}{\left(\frac{\lambda_X^{(i)}+\lambda_Y^{(i)}}{2} + \rho\Delta t\right)^2} + \frac{N^{-\frac{1}{2}}h_n^2 (\Delta t)^2}{\left(\frac{\lambda_X^{(i)}+\lambda_Y^{(i)}}{2} - \rho\Delta t\right)^2} \right)$$

and

$$\begin{aligned} \sum_{i=1}^N \left(\gamma_+^{(i)}\right)^2 + \sum_{i=1}^N \left(\gamma_-^{(i)}\right)^2 &> \sum_{i=1}^N \left( \frac{N^{-\frac{1}{2}}h_N^2 (\Delta t)^2 + \left(\frac{\lambda_Y^{(i)}-\lambda_X^{(i)}}{2}\right)^2}{\left(\lambda_X^{(i)} + \rho\Delta t\right)^2} + \frac{N^{-\frac{1}{2}}h_N^2 (\Delta t)^2 + \left(\frac{\lambda_Y^{(i)}-\lambda_X^{(i)}}{2}\right)^2}{\left(\lambda_X^{(i)} - \rho\Delta t\right)^2} \right) \\ &> \sum_{i=1}^N \left( \frac{N^{-\frac{1}{2}}h_N^2 (\Delta t)^2}{\left(\lambda_X^{(i)} + \rho\Delta t\right)^2} + \frac{N^{-\frac{1}{2}}h_N^2 (\Delta t)^2}{\left(\lambda_X^{(i)} - \rho\Delta t\right)^2} \right) \end{aligned}$$

hold. In the lower bound the mixed terms drop out.

Using those inequalities, the proof reduces to the method used before for the equal noise variance case where we found that (Riemann) sums of this type can be approximated by integrals. We just have to do this calculation twice for the upper and the lower bound changing only the constants in the denominator of the integrated function and obtain the convergence to  $2h^2\overline{I}(\rho)$  and  $2h^2\underline{I}(\rho)$ , respectively, with the lower and upper bound for  $I(\rho)$  stated in (17).  $\square$



**Remark 15.** *Although the inequalities appearing in the proof for the case of different noise variances are strict, the asymptotic results do not yield the strict inequalities in (17) for the lower and upper bound for the asymptotic Fisher information. We suppose that the strict inequalities also hold and a numerical approximation for the Riemann sums using different special values indicated this too.*

## 7 Simulation Results

In this section we compare the simulation results for the subsampling and the multi-scale estimator. A detailed comparison of the finite sample size performance of the subsampling and the Hayashi-Yoshida estimator is given in Palandri (2006). We have shown that the multi-scale estimator is asymptotically more efficient which means that the rate of convergence is  $N^{1/4}$  compared to  $N^{1/6}$  for the subsampling estimator. The following simulations investigate the behaviour of both estimators for finite sample sizes.

To generate asynchronous observation times for the processes  $X$  and  $Y$  we take  $t_i, i = 1, \dots, n$  and  $\tau_j, j = 1, \dots, m$  as arrival times of two independent Poisson processes such that  $\Delta t_i \sim \text{Exp}(\vartheta_X)$  and  $\Delta \tau_j \sim \text{Exp}(\vartheta_Y)$ . In our notation this means  $\mathbb{E}[\Delta t_i] = \vartheta_X$  and  $\mathbb{E}[\Delta \tau_j] = \vartheta_Y$ . We set  $T = 1$  and the means of the time increments between observations equally to  $\vartheta_X = \vartheta_Y = 1/30000$ . The expected number of observations for both processes is about the number of seconds during one trading day and thus a typically high-frequency observation scheme. The sets of observations  $\mathcal{O}^X$  and  $\mathcal{O}^Y$  almost surely have no intersection points, but all time increments are of order  $1/N$  in probability (and thus all assumptions imposed in the sections before are guaranteed).

**Remark 16.** *In this special case, where the number of observations  $n$  and  $m$  follow independent Poisson distributions with parameters  $1/\vartheta_X$  and  $1/\vartheta_Y$ , we can prove that our synchronization method creates  $N$  synchronized observations with  $\mathbb{E}N = 1/\vartheta$  where*

$$\vartheta = \vartheta_X + \vartheta_Y - \frac{\vartheta_X \vartheta_Y}{\vartheta_X + \vartheta_Y}.$$

For  $\vartheta_X = \vartheta_Y (= 1/30000)$  we obtain  $\vartheta = (3/2)\vartheta_X$  and  $\mathbb{E}N = (2/3)\vartheta_X^{-1} (= 20000)$ .

For our simulations we use constant parameters  $\sigma_X = \sigma_Y = 1$  and  $\rho \in [-1, 1]$  and neglect drift terms. The increments of the efficient processes are then given by

$$\Delta \tilde{X}_{t_i} = \int_{t_{i-1}}^{t_i} dB_t^X$$

and

$$\Delta \tilde{Y}_{\tau_j} = \int_{\tau_{j-1}}^{\tau_j} dB_t^Y = \int_{\tau_{j-1}}^{\tau_j} \rho dB_t^X + \sqrt{1 - \rho^2} \int_{\tau_{j-1}}^{\tau_j} dB_t,$$

where  $B_t$  is a standard Brownian motion independent of  $\tilde{X}$ . Therefore, we simulate values of  $\tilde{X}$  for all observation times in  $\mathcal{O}^X \cup \mathcal{O}^Y$  and simulate the observations of  $\tilde{Y}$  using the equation above. For the discrete noise processes we assume  $\epsilon_{t_i}^X \sim \mathcal{N}(0, \eta_X^2)$  and  $\epsilon_{\tau_j}^Y \sim \mathcal{N}(0, \eta_Y^2)$ .

To calculate the subsampling and the multi-scale estimators we first have to determine the number of subsamples  $K_N$  and the number of frequencies  $M_N$ , respectively. We know that a choice  $K_N = c_{sub}N^{2/3}$  and  $M_N = c_{multi}N^{1/2}$ , respectively, with constants  $c_{sub}$  and  $c_{multi}$ , respectively, minimizes the resulting mean square errors of the estimators. For our simulations we can calculate the optimal

noise level $\eta_X^2 = \eta_Y^2$	$K_N$	$M_N$
$(1/\sqrt{10})$	646	216
0.1	300	122
$(1/\sqrt{10}) \cdot 0.1$	139	68
0.01	65	38
$(1/\sqrt{10}) \cdot 0.01$	30	22
0.001	14	12
$(1/\sqrt{10}) \cdot 0.001$	6	7
0.0001	3	4

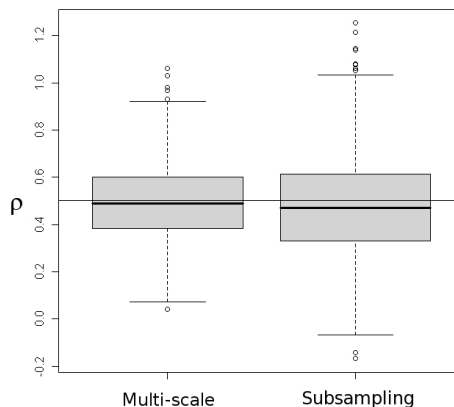


Table 1: Calculated values for  $K_N$  and  $M_N$  for different noise levels  $\eta_X^2 = \eta_Y^2$  and parameters  $\sigma^X = \sigma^Y = 1$ .

Figure 2: Boxplot for Multi-scale and subsampling estimator for  $\rho = 0.5$  when  $\eta_X^2 = \eta_Y^2 = \eta^2 = \sqrt{0.1}$ .

constants because we know all the parameters. Inserting  $K_N = c_{sub}N^{2/3}$  and  $M_N = c_{multi}N^{1/2}$ , respectively, in the asymptotic variances of the estimators, minimization yields

$$c_{sub,opt} = \sqrt[3]{3\eta_X^2\eta_Y^2}, \quad c_{multi,opt} = \sqrt[4]{\frac{36 \cdot 35}{52}\eta_x^2\eta_y^2}.$$

If one uses the estimators for data of two asset prices the parameters are unknown and one can calculate the optimal constants above in the same way, but they will depend on the variances of the noise processes and the quartilities of both underlying Itô processes. Then one can estimate the constants by using adequate estimators for these parameters as proposed in Zhang (2006a) for example.

Figure 2 shows a boxplot for 1000 Monte Carlo iterations for large noise variances  $\eta_X^2 = \eta_Y^2 = \eta^2 = \sqrt{0.1}$  that exemplifies a higher efficiency of our proposed multi-scale estimator compared to the subsampling estimator at least when microstructure noise effects are large.

Next we present a comparison of the resulting root mean square errors (RMSE) for different noise levels. The results are illustrated in Figure 3. The RMSEs are calculated for each noise level based on 1000 Monte Carlo iterations. Our simulations show that for very noisy data (noise level  $\eta_X^2 = \eta_Y^2 = \eta^2 \geq 0.01$ ) and  $\mathbb{E}N = 30000$  expected observations for both processes, the multi-scale estimator has a significant smaller root mean square error compared to the subsampling estimator. The ratio of both RMSEs is increasing when the noise level decreases in the range  $0.1 \geq \eta^2$ . For small noise levels and same (expected) sample sizes the multi-scale estimator also has a smaller RMSE but the ratio of the RMSEs gets close to 1 and fluctuates for different (small) noise levels.

Our simulations thus confirm that for not negligible market microstructure frictions our proposed multi-scale estimator for the quadratic covariation of two Itô processes performs better than the subsampling (and of course the HY-estimator) not only asymptotically but also in the case of typical sample sizes (for high-frequent intraday stock data).

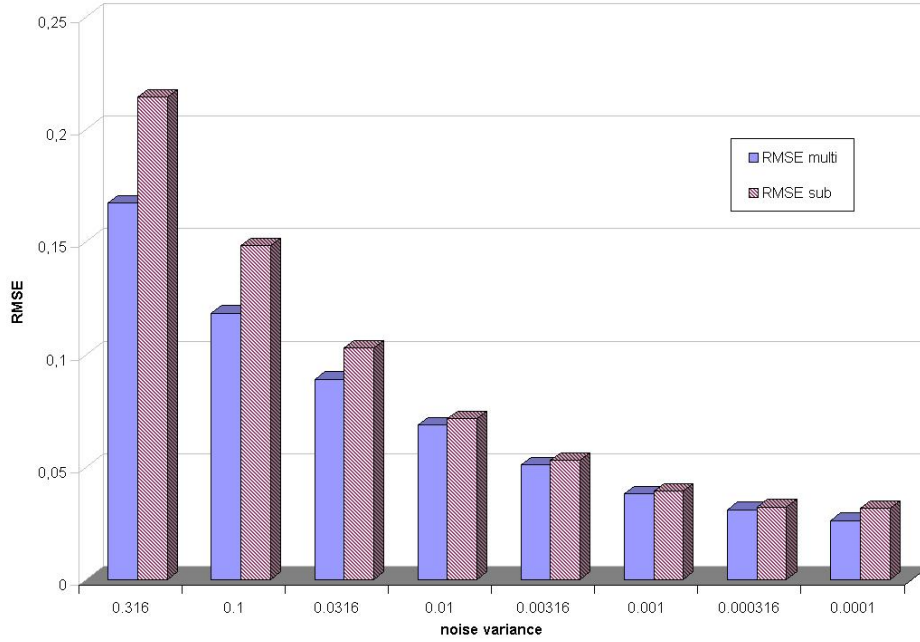


Figure 3: Root mean square errors of subsampling and multi-scale estimator for different noise levels  $\eta_X^2 = \eta_Y^2 = \eta^2$  for  $\rho = 0.5$ .

We have chosen the ranges for the noise variances such that for the illustrated values the noise variances decrease with a factor  $1/\sqrt{10}$ . Because of the factor  $\eta_X^2 \eta_Y^2$  in the variances of the estimators due to noise the root mean square error (disregarding the discretization error) should decrease linearly. This can be seen regarding the values in Figure 3 for large noise variances when the error due to noise dominates the error due to discretization whereas the influence of the discretization error is stronger for small noise levels.

In Figure 4 the root mean square errors of both estimators are diagrammed for different constant parameter values of the correlation  $\rho = k/10, k = 0, \dots, 10$  when  $\eta = 0.01$  based on 200 Monte Carlo iterations for each value. For all eleven parameter values the multi-scale estimator has a smaller root mean square error although the differences are not very large for this (small) noise level. We can announce the increasing root mean square errors when  $\rho$  increases with the dependence of the discretization error on  $\rho$ . Although we did not state precisely a formula for the asymptotic variance due to discretization it is natural that the variance analyzed in Section 5 grows for higher values of  $\rho$ . For this noise level the discretization error is influential enough to cause the different root mean square errors illustrated in Figure 4.

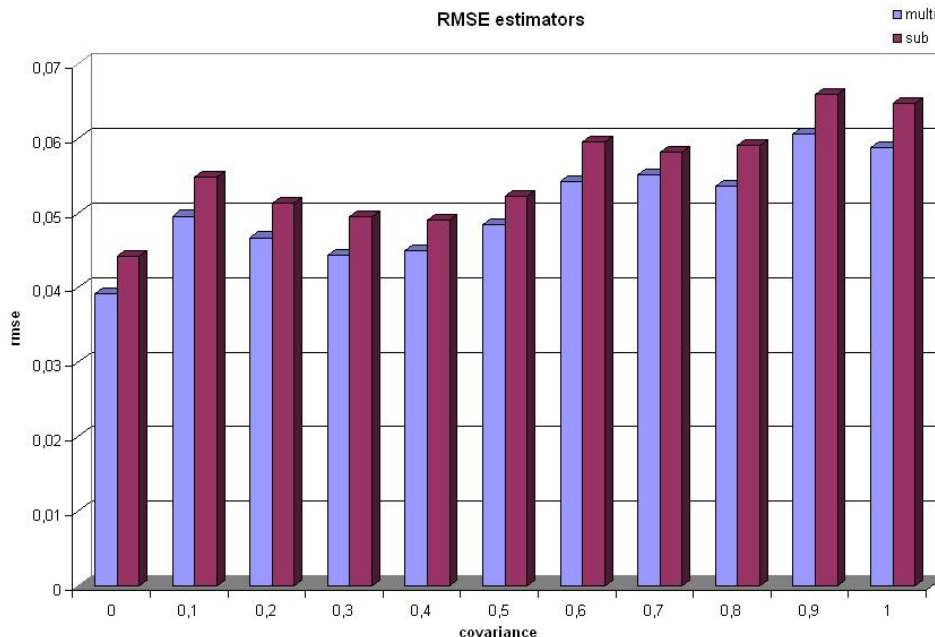


Figure 4: Root mean square errors of subsampling and multi-scale estimator for a constant noise level  $\eta_X^2 = \eta_Y^2 = \eta^2 = 0.01$  and different correlations.

## 8 Conclusion

We have presented and compared three estimators for the quadratic covariation of two Itô processes. If we have discrete asynchronous observations without market microstructure noise, the Hayashi-Yoshida estimator is a consistent estimator solving the problem of asynchronicity. However, when microstructure frictions are relevant the estimator is not consistent any more, as we stated in Proposition 1. If we deal with noisy asynchronous data, we have to synchronize the observations first. We used the method presented by Palandri (2006) in Section 3 to rearrange the observations in an adequate way. We have shown in Section 4 that a subsampling approach yields a consistent estimator with  $N^{1/6}$ -rate of convergence, where  $N$  denotes the number of synchronized observations. In Section 5 we introduced our multi-scale estimator which gains a higher efficiency and has a  $N^{1/4}$ -rate of convergence which is stated in Theorem 1. This rate is optimal what we have proved in a simplified model even for the synchronous case by giving a lower bound for the rate of convergence using the LAN property with rate  $N^{-1/4}$  in Section 6. Proposition 11 comprises this result and the asymptotic Fisher information. Simulations show that the multi-scale estimator performs better compared to the subsampling estimator if the noise level is high enough and the sample size is not too small.

## References

Aït-Sahalia, Y., L. Zhang, and P. A. Mykland (2005). Ultra high frequency volatility estimation with dependent microstructure noise. *technical report, Princeton University*.

- Bandi, F. M. and J. R. Russel (2005). Realized covariation, realized beta, and microstructure noise. *working paper, University of Chicago*.
- Barndorff-Nielsen, O. E., P. R. Hansen, A. Lunde, and N. Shephard (2008). 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 (2004). Econometric analysis of realised covariation: High frequency covariance, regression and correlation in financial economics. *Econometrica* (3), 885–925.
- Barndorff-Nielsen, O. E. and N. Shephard (2006). Power and bipower variation with stochastic volatility and jumps. *Journal of Financial Econometrics* 4, 1–30.
- Gloter, A. and J. Jacod (2001). Diffusions with measurement errors 1 and 2. *ESAIM: Probability and Statistics*, 225–242.
- Griffin, J. E. and R. C. Oomen (2006). Covariance measurement in the presence of non-synchronous trading and market microstructure noise. *working paper, University of Warwick*.
- Hansen, P. R., A. Lunde, and V. Voev (2007). Integrated covariance estimation using high-frequency data in the presence of noise. *Journal of Financial Econometrics* (1), 68–104.
- Hayashi, T. and N. Yoshida (2005). On covariance estimation of non-synchronously observed diffusion processes. *Bernoulli* 11(2), 359–379.
- Jacod, J., Y. Li, P. A. Mykland, M. Podolskij, and M. Vetter (2007). Microstructure noise in the continuous case: the pre-averaging approach. *technical report, University of Chicago*.
- Jacod, J. and A. N. Shiryaev (2003). *Limit Theorems for Stochastic Processes*. New York: Springer-Verlag.
- Palandri, A. (2006). Consistent realized covariance for asynchronous observations contaminated by market microstructure noise. *technical report, University of Copenhagen*.
- van der Vaart, A. W. (1998). *Asymptotic Statistics*. Cambridge University Press.
- van der Vaart, A. W. and J. A. Wellner (1996). *Weak Convergence and Empirical Processes*. New York: Springer.
- Zhang, L. (2006a). Efficient estimation of stochastic volatility using noisy observations: A multi-scale approach. *Bernoulli* 12(6), 1019–1043.
- Zhang, L. (2006b). Estimating covariation: Epps effect, microstructure noise. *technical report, Princeton University*.
- Zhang, L., P. A. Mykland, and Y. Ait-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.