

## CREATES Research Paper 2008-23

# An Econometric Analysis of Modulated Realised Covariance, Regression and Correlation in Noisy Diffusion Models

Silja Kinnebrock and Mark Podolskij



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



**Aarhus School of Business**  
**University of Aarhus**

Handelshøjskolen  
Aarhus Universitet

UNIVERSITY OF  
COPENHAGEN



# An Econometric Analysis of Modulated Realised Covariance, Regression and Correlation in Noisy Diffusion Models \*

Silja Kinnebrock

*Oxford-Man Institute of Quantitative Finance*

*Oxford University*

*Oxford, OX1 4JD, UK*

*e-mail: silja.kinnebrock@oxford-man.ac.uk*

Mark Podolskij

*University of Aarhus and CREATES*

*School of Economics and Management*

*DK-8000 Aarhus C*

*e-mail: mpodolskij@creates.au.dk*

## Abstract

This paper introduces a new estimator to measure the ex-post covariation between high-frequency financial time series under market microstructure noise. We provide an asymptotic limit theory (including feasible central limit theorems) for standard methods such as regression, correlation analysis and covariance, for which we obtain the optimal rate of convergence. We demonstrate some positive semidefinite estimators of the covariation and construct a positive semidefinite estimator of the conditional covariance matrix in the central limit theorem. Furthermore, we indicate how the assumptions on the noise process can be relaxed and how our method can be applied to non-synchronous observations. We also present an empirical study of how high-frequency correlations, regressions and covariances change through time.

---

\*We thank Neil Shephard and Matthias Vetter for valuable comments. The first author gratefully acknowledges financial support from the Rhodes Trust and Oxford-Man Institute of Quantitative Finance. The second author is grateful for financial support from CREATES funded by the Danish National Research Foundation.

**Keywords:** Central Limit Theorem; Diffusion Models; Market Microstructure Noise; Non-synchronous Trading; High-Frequency Data; Semimartingale Theory; .

**AMS 2000 Subject Classification:** Primary 60F05, 62M05; secondary 60G44, 60H05.

## 1 Introduction

In recent years the availability of high frequency data changed the possibilities to accurately estimate financial volatility and correlation dramatically. The underlying idea is to use quadratic covariation as the ex-post covariation of asset prices whose increments can be estimated using a variety of instruments. A prominent estimator in the univariate case is realised variance which has been extensively discussed in Andersen et al. (2001) and Barndorff-Nielsen and Shephard (2002). But researchers faced the problem that frequent sampling would not allow them to ignore the contamination of financial data by market frictions which is known as market microstructure noise. This goes back to Zhou (1998).

The presence of market microstructure noise in high frequency financial data is well-known and complicates the estimation of both the volatility of and the covariance between asset prices. The problem is that observed asset price processes do not exhibit the efficient prices, but are contaminated by some noise. Estimators for financial volatility or covariance which work well in the absence of noise, are severely affected by the contamination of the data while sampling at high frequencies and are thus unreliable. However, the effects this has on realised volatility and realised covariance are opposite. Hansen and Lunde (2006) show that realised variance exhibits a large positive bias and a variance which diverges with the sampling frequency when using trade data. Epps (1979) documented that when estimating from empirical data, realised covariance converges to zero as the sampling frequency increases. When sampling at a very high frequency, there are more and more zero-returns, when there is non-synchronous trading, this effect dominates realised covariance and all related statistics (e.g. realised correlation).

A key to understanding the nature of market microstructure noise and a possible tool how to deal with it is that market microstructure noise induces autocorrelation in the intraday returns and this autocorrelation is the reason for the bias problem. Currently there are three main univariate approaches towards estimating the integrated variance in the presence of microstructure noise: the two-scale estimator proposed by Zhang et al. (2005) (see also the multiscale approach by Zhang (2006) which is the generalisation of the two-scale estimator), the realised kernel introduced in Barndorff-Nielsen et al. (2007) and the concept of modulated bipower variation proposed by Podolskij and Vetter (2007) which was extended in Jacod et al. (2007). The latter method is based on pre-averaging (or pre-filtering) procedure which delivers

a consistent estimator of the integrated variance.

In this paper we tackle the multivariate version of this problem. The problem of estimating correlation under non-synchronous high frequency data has been studied by Lundin et al. (1999) and Brandt and Diebold (2003). They independently proposed an estimation procedure to cope with non-synchronicity in a various model assumption. Most existing approaches, however, rely on synchronization of the original data. Furthermore, Hayashi and Yoshida (2003) introduced an estimator which is capable of dealing with non-synchronous data, but not with market microstructure noise.

The scope of this paper is to extend the theory stated in Podolskij and Vetter (2007) and Jacod et al. (2007) to the multivariate case and also address problems which occur additionally in the multivariate setting. We will call this multivariate extension Modulated Realised Covariation (MRC) in the following. We show that MRC is a simple consistent estimator of the covariation and prove the corresponding central limit theorem under weak assumptions on the involved processes. Our estimator converges at the rate  $n^{-1/4}$ , which is known to be optimal (see Jacod and Glotter (2001)). Furthermore, we construct a positive semidefinite estimator of the conditional covariance matrix in the central limit theorem to obtain feasible asymptotic results. Finally, we explain how a positive semidefinite estimator of the covariation can be obtained and how the assumptions on the noise process can be relaxed.

The MRC estimator as it is introduced in this paper, is not directly capable of dealing with non-synchronous data. However, we will use a method to clean data in a way so that empirical results are still rather good. This method has been inspired by the concept of refresh time which goes back to Barndorff-Nielsen, Hansen, Lunde and Shephard (2008).

The remainder of the paper is organised as follows. After a section on notation we present the multivariate asymptotic theory and apply it to deduce a central limit theorem for estimator for regression and correlation. Then we test our results with an extensive simulation study and finally present an empirical illustration.

## **2 Basic notations and definitions**

No-arbitrage based characterisations of securities prices (as in Delbaen and Schachermeyer (1994)) suggest they must follow Brownian semimartingales. They satisfy the fundamental law

of asset pricing and have been used extensively in order to model the evolution of asset prices in time. A Brownian semimartingale is characterised by the equation

$$Z_t = Z_0 + \int_0^t a_u du + \int_0^t \sigma_u dW_u, \quad t \in [0, 1], \quad (2.1)$$

where  $a$  is a  $d$ -dimensional predictable locally bounded drift vector,  $\sigma$  a càdlàg  $d \times d$  covolatility matrix and  $W$  is  $d$ -dimensional Brownian motion. All processes are defined on the filtered probability space  $(\Omega^0, \mathcal{F}^0, (\mathcal{F}_t^0)_{t \in [0,1]}, P^0)$ . For all continuous time stochastic processes the quadratic covariation process is defined as

$$[Z]_t = P - \lim_{n \rightarrow \infty} \sum_{j=1}^n (Z_{t_j} - Z_{t_{j-1}}) (Z_{t_j} - Z_{t_{j-1}})' \quad (2.2)$$

for any sequence of deterministic partitions  $0 = t_0 < t_1 < \dots < t_n = t$  with  $\sup_j \{t_j - t_{j-1}\} \rightarrow 0$  for  $n \rightarrow \infty$ . For Brownian semimartingales the quadratic covariation is given as

$$[Z]_t = \int_0^t \sigma_u \sigma_u' du \quad (2.3)$$

(2.3) is the object of interest as the quadratic covariation matrix gives us a good way to measure both the volatility of an asset price process as well as the covolatility of different processes. However, we do not observe the efficient price (2.1) in the market, but a process  $Z^*$  which is contaminated by market microstructure noise. More precisely, we consider the process  $Z^*$ , observed at time points  $i/n$ ,  $i = 0, 1, \dots, n$ , which is given as

$$Z_t^* = Z_t + \epsilon_t, \quad (2.4)$$

where  $(\epsilon_t)$  is an i.i.d. process that is independent of  $Z$ . Such a process can be constructed as follows. We define a second probability space  $(\Omega^1, \mathcal{F}^1, (\mathcal{F}_t^1)_{t \in [0,1]}, P^1)$ , where  $\Omega^1$  denotes  $\mathbb{R}^{[0,1]}$  and  $\mathcal{F}^1$  the product Borel- $\sigma$ -field on  $\Omega^1$ . Next, let  $Q$  be a probability measure on  $\mathbb{R}$ . 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

$(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t \in [0,1]}, P)$ , on which we define the process  $Z^*$ , is given as

$$\left. \begin{aligned} \Omega &= \Omega^0 \times \Omega^1, & \mathcal{F} &= \mathcal{F}^0 \times \mathcal{F}^1, & \mathcal{F}_t &= \bigcap_{s>t} \mathcal{F}_s^0 \times \mathcal{F}_s^1, \\ P &= P^0 \otimes P^1. \end{aligned} \right\} \quad (2.5)$$

The multivariate noise process  $\epsilon$  is assumed to satisfy the following assumption:

$$E\epsilon_t = 0, \quad E\epsilon_t \epsilon_t' = \Psi, \quad (2.6)$$

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

**Remark 1** Hansen and Lunde (2006) pointed out that the i.i.d. assumption on the noise process  $(\epsilon_t)$  and the independence between  $Z$  and  $\epsilon$  is quite unrealistic for ultra high frequencies. Jacod et al. (2007) consider a more general type of (1-dimensional) noise processes. Roughly speaking, they allow for noise processes of the form

$$\epsilon_t = \alpha_t \nu_t, \quad (2.7)$$

where  $(\alpha_t)$  is  $\mathcal{F}^0$ -measurable and càdlàg,  $E[\nu_t | \mathcal{F}^0] = 0$ ,  $E[\nu_t^2 | \mathcal{F}^0] = 1$  for any  $t \in [0, 1]$ , and, conditionally on  $\mathcal{F}^0$ ,  $\nu_t$  is independent of  $\nu_s$  for  $t \neq s$ . An interesting example of a 1-dimensional process  $Z^*$ , for which the above conditions on the noise are fulfilled, is given by ( $\gamma > 0$ )

$$Z_t^* = \gamma \left\lfloor \frac{Z_t + U_t}{\gamma} \right\rfloor,$$

where  $U_t$  is an i.i.d. process with  $U([0, \gamma])$  distribution, which is independent of  $Z$  (see Li and Mykland (2007) or Jacod et al. (2007)). In this case we have  $Z_t^* = Z_t + \epsilon_t$  with  $\epsilon_t = \alpha_t \nu_t$  and

$$\alpha_t^2 = \gamma^2 \left( \left\{ \frac{Z_t}{\gamma} \right\} - \left\{ \frac{Z_t}{\gamma} \right\}^2 \right),$$

where  $\{x\}$  denotes the rational part of  $x$ . Obviously,  $(\alpha_t)$  is a càdlàg process.

The asymptotic theory developed in this paper remains true for the multivariate noise processes of the type (2.7). However, we restrict ourselves to the models of the form (2.4) for the ease of exposition.

Next, we choose a sequence  $k_n$  of integers and a number  $\theta \in (0, \infty)$  satisfying

$$\frac{k_n}{\sqrt{n}} = \theta + o\left(n^{-1/4}\right) \quad (2.8)$$

(for example  $k_n = \lfloor \theta \sqrt{n} \rfloor$ ). 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 which are associated with  $g$ :

$$\phi_1(s) = \int_s^1 g'(u) g'(u-s) du, \quad \phi_2(s) = \int_s^1 g(u) g(u-s) du, \quad \psi_1 = \phi_1(0), \quad \psi_2 = \phi_2(0),$$

$$\Phi_{11} = \int_0^1 \phi_1^2(s) ds, \quad \Phi_{12} = \int_0^1 \phi_1(s) \phi_2(s) ds, \quad \Phi_{22} = \int_0^1 \phi_2^2(s) ds.$$

The functions  $\phi_1$  and  $\phi_2$  are assumed to be 0 outside the interval  $[0, 1]$ .

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

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

The core statistic of this paper is the multivariate extension of the estimator which was introduced in Jacod et al. (2007). We call it the Modulated Realised Covariation (MRC) estimator and it is defined as

$$MRC(Z^*)_n = \frac{1}{\theta \psi_2 \sqrt{n}} \sum_{i=0}^{n-k_n+1} \bar{Z}_i^{*n} (\bar{Z}_i^{*n})' - \frac{\psi_1}{2\theta^2 \psi_2 n} \sum_{i=1}^n \Delta_i^n Z^* (\Delta_i^n Z^*)'. \quad (2.9)$$

The last term in (2.9) is introduced to remove the bias due to the noise, but it does not play any role in the central limit theorem given below. In fact, it holds that

$$\hat{\Psi} = \frac{1}{2n} \sum_{i=1}^n \Delta_i^n Z^* (\Delta_i^n Z^*)' \xrightarrow{P} \Psi.$$

**Remark 2** The intuition behind the quantity  $\bar{Z}_i^{*n}$  can be explained as follows. Assume for a

moment that the integer  $k_n$  is an even number and define

$$\hat{Z}_i^{*n} = \frac{2}{k_n} \sum_{j=0}^{\frac{k_n}{2}-1} Z_{\frac{i+j}{n}}^*.$$

Due to this pre-averaging procedure the quantity  $\tilde{Z}_i^{*n}$  is closer to the efficient price  $Z_{\frac{i}{n}}$ . Next, we compute the realised covariation estimator based on the filtered increments

$$2(\hat{Z}_{i+\frac{k_n}{2}}^{*n} - \hat{Z}_i^{*n}).$$

(However, the latter induces a bias which is corrected by means of  $\hat{\Psi}$ ). This method was originally proposed by Podolskij and Vetter (2006) and  $\bar{Z}_i^{*n} := 2(\hat{Z}_{i+\frac{k_n}{2}}^{*n} - \hat{Z}_i^{*n})$  corresponds to the function

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

which is the most intuitive example. In this case the constants are given as

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

### 3 Asymptotic theory

In this section we study the asymptotic behaviour of  $MRC(Z)_n$ .

#### 3.1 Consistency

**Theorem 1** *Assume that  $E|\epsilon^j|^4 < \infty$  for all  $j = 1, \dots, d$ . If  $k_n$  and  $\theta$  satisfy (2.8), then the convergence in probability*

$$MRC(Z^*)_n \xrightarrow{P} \int_0^1 \sigma_s \sigma_s' ds \tag{3.1}$$

*holds.*

#### 3.2 The central limit theorem

In this section we present the central limit theorem for  $MRC(Z^*)_n$ . As in Jacod et al. (2007) we only require a moment condition on the noise process  $\epsilon$  to prove the next theorem. We use



the concept of stable convergence. Let us briefly recall the definition. A sequence of random variables  $Z^n$  converges stably in law with limit  $Z$  (throughout this paper we write  $Z^n \xrightarrow{\mathcal{D}_{st}} Z$ ), defined on an appropriate extension  $(\Omega', \mathcal{F}', P')$  of the probability space  $(\Omega, \mathcal{F}, P)$ , if and only if for any  $\mathcal{F}$ -measurable and bounded random variable  $Z$  and any bounded and continuous function  $f$  the convergence

$$\lim_{n \rightarrow \infty} E[Zf(Y^n)] = E[Zf(Y)]$$

holds. This is obviously a slightly stronger mode of convergence than weak convergence (see Renyi (1963), Aldous & Eagleson (1978) or Jacod & Shiryaev (2003) for more details on stable convergence).

**Theorem 2** *Assume that  $E|\epsilon^j|^8 < \infty$  for all  $j = 1, \dots, d$  and  $k_n$  and  $\theta$  satisfy (2.8). Then the sequence*

$$n^{1/4} \left( MRC(Z^*)_n - \int_0^1 \sigma_s \sigma'_s ds \right) \quad (3.2)$$

*converges stably in law towards a limiting variable defined on an extension of the original space which is of the form*

$$U = \sum_{j', k'=1}^d \int_0^1 \gamma_s^{j'k', j'k'} dB_s^{j'k'},$$

*where  $B$  is a standard  $d^2$ -dimensional Wiener process, independent of  $\mathcal{F}$ , and*

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

*Here  $\Lambda$ ,  $\Theta$  and  $\Upsilon$  are  $d \times d \times d \times d$  arrays with elements*

$$\begin{aligned} \Lambda_s &= \left\{ \Sigma_s^{kk'} \Sigma_s^{ll'} + \Sigma_s^{kl'} \Sigma_s^{lk'} \right\}_{k, k', l, l'=1, \dots, d}, \\ \Theta_s &= \left\{ \Sigma_s^{kk'} \Psi^{ll'} + \Sigma_s^{kl'} \Psi^{k'l} + \Sigma_s^{k'l} \Psi^{kl'} + \Sigma_s^{ll'} \Psi^{kk'} \right\}_{k, k', l, l'=1, \dots, d}, \\ \Upsilon &= \left\{ \Psi^{kk'} \Psi^{ll'} + \Psi^{kl'} \Psi^{lk'} \right\}_{k, k', l, l'=1, \dots, d}. \end{aligned} \quad (3.3)$$

An equivalent way of writing the convergence statement in Theorem 2 is the following.

$$n^{1/4} \left( MRC(Z^*)_n - \int_0^1 \sigma_s \sigma'_s ds \right) \xrightarrow{\mathcal{D}_{st}} MN(0, L)$$

where

$$L = \frac{1}{\psi_2^2} \left( \Phi_{22} \theta \int_0^1 \Lambda_s ds + \frac{\Phi_{12}}{\theta} \int_0^1 \Theta_s ds + \frac{\Phi_{11}}{\theta^3} \Upsilon \right) \quad (3.4)$$

is a conditional covariance matrix.

### 3.3 Positive semidefinite estimators

Here we present the asymptotic theory for a non-optimal choice of  $k_n$ , i.e.

$$\frac{k_n}{n^{1/2+\delta}} = \theta + o(n^{-\frac{1}{4}+\frac{\delta}{2}}) \quad (3.5)$$

for some  $0 < \delta < 1/2$ . We will see that for a non-optimal rate we obtain a positive semidefinite estimator of covariation, which consistency holds under much weaker assumptions on the noise process.

We define the estimator

$$MRC(Z^*)_n^\delta = \frac{1}{\theta \psi_2 n^{1/2+\delta}} \sum_{i=0}^{n-k_n+1} \bar{Z}_i^{*n} (\bar{Z}_i^{*n})'. \quad (3.6)$$

Our first result is a consequence of the univariate asymptotic theory developed in Vetter (2008).

**Theorem 3** *Assume that  $E|\epsilon^j|^4 < \infty$  for all  $j = 1, \dots, d$ . If  $k_n$  and  $\theta$  satisfy (3.5), then the convergence in probability*

$$MRC(Z^*)_n^\delta \xrightarrow{P} \int_0^1 \sigma_s \sigma'_s ds \quad (3.7)$$

*holds.*

Notice that the characteristics of the noise process do not appear in the limit, because the influence of the noise is negligible for the choice of  $k_n$  made in (3.5). This procedure has two advantages: (i)  $MRC(Z^*)_n^\delta$  is obviously positive semidefinite and (ii) the convergence in (3.7) holds under much weaker assumptions on the noise process. More precisely, we do not require the i.i.d. assumption on the process  $(\epsilon_t)$  (as long as  $\bar{\epsilon}_i^n$  admits asymptotic normality at the

usual rate  $k_n^{-1/2}$ ) and we can allow for any dependence structure between  $Z$  and  $\epsilon$ .

To prove the associated central limit theorem we require a further restriction on the parameter  $\delta$ . This is due to the bias caused by the noise process, which is negligible in Theorem 3, but can be large when multiplying with the convergence rate. The next result follows through the lines of the proof of Theorem 2 (see again Vetter (2008) for more details).

**Theorem 4** *Assume that  $E|\epsilon^j|^8 < \infty$  for all  $j = 1, \dots, d$  and  $k_n$  and  $\theta$  satisfy (3.5). Then we obtain the following results.*

(i) *If  $\delta > \frac{1}{10}$  we have*

$$n^{1/4-\delta/2} \left( MRC(Z^*)_n^\delta - \int_0^1 \sigma_s \sigma'_s ds \right) \xrightarrow{\mathcal{D}_{st}} MN \left( 0, \frac{\Phi_{22}\theta}{\psi_2^2} \int_0^1 \Lambda_s ds \right), \quad (3.8)$$

where  $(\Lambda_s)$  is defined in (3.3).

(ii) *If  $\delta = \frac{1}{10}$  we have*

$$n^{1/5} \left( MRC(Z^*)_n^\delta - \int_0^1 \sigma_s \sigma'_s ds \right) \xrightarrow{\mathcal{D}_{st}} MN \left( \frac{\psi_1}{\theta^2 \psi_2} \Psi, \frac{\Phi_{22}\theta}{\psi_2^2} \int_0^1 \Lambda_s ds \right). \quad (3.9)$$

Note that the asymptotic conditional covariance matrix of Theorem 4 is just the first summand of the quantity  $L$  given in (3.4). The drift which appears in (3.9) is the limit of the bias caused by the noise process multiplied with the convergence rate  $n^{-1/5}$ . Clearly, the drift can be estimated by  $\frac{\psi_1}{\theta^2 \psi_2} \hat{\Psi}$ . Consistent estimates of the conditional covariance matrix of Theorem 4 can be obtained as described in Section 4.

**Remark 3** Having a positive semidefinite estimator  $MRC(Z^*)_n^\delta$  of  $\int_0^1 \sigma_s \sigma'_s ds$  is certainly an important issue. However, the rate optimal estimator  $MRC(Z^*)_n$  (which is not necessarily positive semidefinite for finite  $n$ ) appears to be positive semidefinite for moderate sampling frequencies both in simulations and empirical applications.

### 3.4 The bivariate case

The general results are quite compact so that it helps to examine the bivariate case in order to get some further understanding. We will look at two assets 1 and 2 and with bivariate log-price

process  $Z = (X, Y)$ . We consider  $\epsilon_t$  an i.i.d. noise process with expectation 0 and covariance matrix  $\Psi$  which is of the form

$$\Psi = \begin{pmatrix} \omega_1^2 & \omega_1\omega_2c \\ \omega_1\omega_2c & \omega_2^2 \end{pmatrix}, \quad (3.10)$$

where  $c$  stands for the correlation between  $\epsilon^1$  and  $\epsilon^2$ . In that case Theorem 2 gives us the following central limit theorem.

**Corollary 5** *Under the assumptions of Theorem 2, the following limit theorem holds*

$$\begin{aligned} & n^{1/4} \begin{pmatrix} MRC(Z^*)_n^{1,1} - \int_0^1 \Sigma_u^{11} du \\ MRC(Z^*)_n^{1,2} - \int_0^1 \Sigma_u^{12} du \\ MRC(Z^*)_n^{2,2} - \int_0^1 \Sigma_u^{22} du \end{pmatrix} \\ & \xrightarrow{\mathcal{D}_{st}} MN \left[ 0, \frac{2}{\psi_2^2} \left( \Phi_{22}\theta \int_0^1 \tilde{\Lambda}_u du + \frac{\Phi_{12}}{\theta} \int_0^1 \tilde{\Theta}_u du + \frac{\Phi_{11}}{\theta^3} \tilde{\Upsilon} \right) \right] \end{aligned} \quad (3.11)$$

and

$$\begin{aligned} \tilde{\Lambda}_u &= \begin{pmatrix} 2(\Sigma_u^{11})^2 & 2\Sigma_u^{11}\Sigma_u^{12} & 2(\Sigma_u^{12})^2 \\ 2\Sigma_u^{11}\Sigma_u^{12} & \Sigma_u^{11}\Sigma_u^{22} + (\Sigma_u^{12})^2 & 2\Sigma_u^{22}\Sigma_u^{12} \\ 2(\Sigma_u^{12})^2 & 2\Sigma_u^{22}\Sigma_u^{12} & 2(\Sigma_u^{22})^2 \end{pmatrix}, \\ \tilde{\Theta}_u &= \begin{pmatrix} 4\omega_1^2\Sigma_u^{11} & \bullet & \bullet \\ 2\omega_1\omega_2c_{12}\Sigma_u^{11} + 2\omega_1^2\Sigma_u^{12} & \Sigma_u^{11}\omega_2^2 + \Sigma_u^{22}\omega_1^2 + 2\Sigma_u^{12}\omega_1\omega_2c_{12} & \bullet \\ 4\Sigma_u^{12}\omega_1\omega_2c_{12} & 2\omega_1\omega_2c_{12}\Sigma_u^{22} + 2\omega_2^2\Sigma_u^{12} & 4\omega_2^2\Sigma_u^{22} \end{pmatrix}, \\ \tilde{\Upsilon} &= \begin{pmatrix} 2\omega_1^4 & 2\omega_1^3\omega_1c_{12} & 2\omega_1^2\omega_2^2c_{12}^2 \\ 2\omega_1^3\omega_2c_{12} & \omega_1^2\omega_2^2 + \omega_1^2\omega_2^2c_{12}^2 & 2\omega_2^3\omega_1c_{12} \\ 2\omega_1^2\omega_2^2c_{12}^2 & 2\omega_2^3\omega_1c_{12} & 2\omega_2^4 \end{pmatrix}. \end{aligned}$$

In particular, we obtain the following result.

$$\frac{n^{1/4} \left( MRC(Z^*)_n^{1,2} - \int_0^1 \Sigma_{12}(u) du \right)}{\sqrt{L^{22}}} \xrightarrow{\mathcal{D}_{st}} N(0, 1), \quad (3.12)$$

where

$$\begin{aligned}
 L^{22} = & 2 \frac{\Phi_{22}\theta}{\psi_2^2} \int_0^1 \left( \Sigma_u^{11} \Sigma_u^{22} + (\Sigma_u^{12})^2 \right) du + 2 \frac{\Phi_{12}}{\theta \psi_2^2} \left( \omega_2^2 \int_0^1 \Sigma_u^{11} du \right. \\
 & \left. + \omega_1^2 \int_0^1 \Sigma_u^{22} du + 2\omega_1\omega_2 c_{12} \int_0^1 \Sigma_u^{12} du \right) + 2 \frac{\Phi_{11}}{\theta^3 \psi_2^2} (\omega_1^2 \omega_2^2 + \omega_1^2 \omega_2^2 c_{12}^2). \tag{3.13}
 \end{aligned}$$

### 3.5 The optimal choice of $\theta$

The efficiency of our estimates presented in previous sections depends on the parameter  $\theta$ . Here we fix the function  $g$  and demonstrate a particular choice of  $\theta$  which minimizes a certain norm of the conditional covariance matrix in the central limit theorem. For the sake of simplicity we only consider the bivariate case. Recall the expression for the conditional covariance matrix given in (3.11):

$$A(\theta) := \frac{2}{\psi_2^2} \left( \Phi_{22}\theta \int_0^1 \tilde{\Lambda}_u du + \frac{\Phi_{12}}{\theta} \int_0^1 \tilde{\Theta}_u du + \frac{\Phi_{11}}{\theta^3} \tilde{\Upsilon} \right).$$

For any matrix  $A$  we set  $|A|_{tr}^2 = \text{trace}(AA')$ . We choose  $\hat{\theta}$  as a solution of the following minimization problem

$$\hat{\theta} = \operatorname{argmin}_{\theta > 0} |A(\theta)|_{tr}^2.$$

The latter results in solving a polynomial equation of degree 4. In the next section we will present consistent estimates of the random quantities  $\int_0^1 \tilde{\Lambda}_u du$ ,  $\int_0^1 \tilde{\Theta}_u du$  and  $\tilde{\Upsilon}$ , so  $\hat{\theta}$  can be estimated from the data.

In a 1-dimensional case the afore-mentioned minimization problem has a much simpler solution (since we only have to solve a quadratic equation). Consider the function  $g(x) = \min(x, 1 - x)$  and assume that  $Z = \sigma W$  for some  $\sigma > 0$ . In this case we have that

$$\hat{\theta} = 4.777 \frac{\sqrt{\Psi}}{\sigma}, \quad A(\hat{\theta}) = 8.545 \sigma^3 \sqrt{\Psi}.$$

Recall that the variance of the maximum likelihood estimator for this problem is given by  $8\sigma^3 \sqrt{\Psi}$  (see Gloter and Jacod (2001)).

## 4 Estimating the asymptotic covariance matrix

Now we demonstrate how Theorem 2 can be applied in order to compute confidence intervals for the covariation. Note that in order to make Theorem 2 feasible, we need to estimate the asymptotic covariance matrix  $L$ , as it is given by (3.4), and we give an explicit estimator for  $L$  next.

**Proposition 6** *Assume that  $E | \epsilon^j |^8 < \infty$  for all  $j = 1, \dots, d$  and  $k_n$  and  $\theta$  satisfy (2.8). Then we define*

$$X_i^n = \text{vec} \left( \bar{Z}_i^{*n} \left( \bar{Z}_i^{*n} \right)' \right), \quad (4.1)$$

where the *vec* notation stacks the columns of a matrix into a vector, and

$$\begin{aligned} H_n &= \frac{1}{\sqrt{n\theta^2\psi_2^2}} \sum_{i=0}^{n-2k_n+1} X_i^n \left( X_i^n - X_{i+k_n}^n \right)' \\ &+ \sum_{j=1}^{k_n-1} X_i^n \left( X_{i+j}^n - X_{i+k_n}^n \right)' + \left( X_i^n \left( X_{i+j}^n - X_{i+k_n}^n \right)' \right)'. \end{aligned} \quad (4.2)$$

Then the convergence

$$H_n \xrightarrow{P} L.$$

holds.

(4.1) enables us to obtain a standard central limit theorem for the integrated covariance

$$\frac{n^{1/4} \left( \text{MRC} (Z_n^*)^{12} - \int_0^1 \Sigma_u^{12} du \right)}{\alpha_n} \xrightarrow{D} N(0, 1), \quad (4.3)$$

where

$$\alpha_n^2 = \frac{1}{\sqrt{n\theta^2\psi_2^2}} \sum_{i=0}^{n-k_n+1} \left( \bar{X}_i^{*n} \right)^2 \left( \bar{Y}_i^{*n} \right)^2 + 2 \sum_{j=1}^{k_n-1} \bar{X}_i^{*n} \bar{Y}_i^{*n} \left( \bar{X}_{i+j}^{*n} \bar{Y}_{i+j}^{*n} - \bar{X}_{i+k_n}^{*n} \bar{Y}_{i+k_n}^{*n} \right). \quad (4.4)$$

Unfortunately (4.1) is not necessarily positive semidefinite (it is not even necessarily positive in the univariate case). In the following we construct an estimator which is both positive semidefinite and symmetric. For simplicity let us consider the 2-dimensional process  $Z^* =$

$(X^*, Y^*)$ . Next, set

$$\chi_i^n = \left( |\bar{X}_i^{*n}|^2, \bar{X}_i^{*n} \bar{Y}_i^{*n}, |\bar{Y}_i^{*n}|^2 \right)'$$

and 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  $\theta$  and the function  $g$ ). Clearly,  $V_n(g)$  is positive semidefinite. Moreover, we have

$$V_n(g) \xrightarrow{P} a_B(g, \theta) \int_0^1 \Lambda_u du + a_M(g, \theta) \int_0^1 \Theta_u du + a_N(g, \theta) \Upsilon,$$

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 Proposition 6). We need to estimate the quantity

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

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 we obtain

$$\tilde{H}_n = C^{(1)}(g_1, g_2, g_3) V_n(g_1) + C^{(2)}(g_1, g_2, g_3) V_n(g_2) + C^{(3)}(g_1, g_2, g_3) V_n(g_3) \quad (4.5)$$

$$\xrightarrow{P} \frac{2}{\psi_2^2} \left( \Phi_{22}\theta \int_0^1 \Lambda_u du + \frac{\Phi_{12}}{\theta} \int_0^1 \Theta_u du + \frac{\Phi_{11}}{\theta^3} \Upsilon \right).$$

Moreover, the estimator is positive semidefinite, because  $V_n(g_k)$  are positive semidefinite and all  $C^{(k)}(g_1, g_2, g_3)$  ( $k = 1, 2, 3$ ) are positive.

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. Throughout this paper we are going to work with

$$g_1(x) = \sqrt{56}x^4(1-x), \quad g_2(x) = \sqrt{611}x^4(1-x), \quad g_3(x) = \sqrt{209}x^2(1-x)$$

and obtain the constants

$$\psi_1^{g_1} = 3.56, \quad \psi_2^{g_1} = 0.113, \quad \psi_1^{g_2} = 38.79, \quad \psi_2^{g_2} = 1.234, \quad \psi_1^{g_3} = 27.867, \quad \psi_2^{g_3} = 1.99.$$

Finally we obtain the following result.

$$\frac{n^{1/4} \left( MRC(Z^*)^{1,2} - \int_0^1 \Sigma_{12}(u) du \right)}{\tilde{\alpha}_n} \xrightarrow{D} N(0, 1), \quad (4.6)$$

where

$$\tilde{\alpha}_n^2 = C^{(1)}(g_1, g_2, g_3)V_n^{22}(g_1) + C^{(2)}(g_1, g_2, g_3)V_n^{22}(g_2) + C^{(3)}(g_1, g_2, g_3)V_n^{22}(g_3) \quad (4.7)$$

and

$$V_n^{22}(g) = \sum_{i=0}^{n-k_n+1} (\bar{X}_i^n)^2 (\bar{Y}_i^n)^2 - \sum_{i=0}^{n-2k_n+1} \bar{X}_i^n \bar{Y}_i^n \bar{Y}_{i+k_n}^n \bar{X}_{i+k_n}^n.$$

## 5 Asymptotic theory for regression and correlation

In this section we study the asymptotic behaviour of some statistics that are transformations of modulated realised covariation. The focus will be on modulated realised regression and correlation. We start with the regression case.



## 5.1 Modulated realised regression

Regression plays a central role both in theoretical and empirical financial economics. In this section we use our distribution theory for modulated realised covariation to derive a theory for univariate regression. We consider a bivariate log-price process  $Z^* = (X^*, Y^*)$ . We regress asset 2 with log-price process  $Y$  on asset 1 with log-price process  $X$  to obtain the modulated realised regression

$$\hat{\beta}_n^{(21)} = \frac{MRC(Z^*)_n^{1,2}}{MRC(Z^*)_n^{1,1}}. \quad (5.1)$$

Modulated realised regression involves just elements of the matrix  $MRC(Z^*)$  and so we can use the asymptotic theory of the previous chapter to derive its asymptotic distribution. The probability limit for the regression case follows from Theorem 1. In particular, if  $Z \in BSM$ , then

$$\hat{\beta}_n^{(21)} \xrightarrow{P} \frac{[X, Y]}{[X]} = \beta^{(21)}, \quad (5.2)$$

where  $[X, Y]$  is the  $(1, 2)$ -th element of  $[Z]$ . As  $Z$  is a Brownian semimartingale, (5.2) can be reduced to the much simpler form

$$\beta^{(21)} = \frac{\int_0^1 \Sigma_u^{12} du}{\int_0^1 \Sigma_u^{11} du}. \quad (5.3)$$

The asymptotic distribution of  $\hat{\beta} - \beta$  is derived using the  $\Delta$ -method.

**Theorem 7** *Under the conditions of Theorem 2, as  $n \rightarrow \infty$*

$$\frac{n^{1/4} \left( \hat{\beta}_n^{(21)} - \beta^{(21)} \right)}{\sqrt{\left( \int_0^1 \Sigma_u^{11} du \right)^{-2} g^{21}}} \xrightarrow{D} N(0, 1) \quad (5.4)$$

where  $g^{21} = d^{(21)'} \Gamma^{(21)} d^{(21)}$ . Here  $d^{(21)} = (1, -\beta^{(21)})'$  and

$$\hat{\Gamma}^{(21)} = \left\{ \frac{\Phi_{22}\theta}{\psi_2^2} \int_0^1 \begin{pmatrix} \Sigma_u^{11} \Sigma_u^{22} + (\Sigma_u^{12})^2 & \bullet \\ 2\Sigma_u^{11} \Sigma_u^{12} & 2(\Sigma_u^{11})^2 \end{pmatrix} du \right. \\ \left. + \frac{\Phi_{12}}{\theta\psi_2^2} \int_0^1 \begin{pmatrix} \Sigma_u^{11} \omega_2^2 + \Sigma_u^{22} \omega_1^2 + \Sigma_u^{12} \omega_1 \omega_2 c_{12} & \bullet \\ 2\omega_1 \omega_2 c_{12} \Sigma_u^{11} + 2\omega_1^2 \Sigma_u^{12} & 4\omega_1^2 \Sigma_u^{11} \end{pmatrix} du + \frac{\Phi_{11}}{\theta^3 \psi_2^2} \begin{pmatrix} \omega_1^2 \omega_2^2 + \omega_1^2 \omega_2^2 c_{12}^2 & \bullet \\ 2\omega_1^3 \omega_2 c_{12} & 2\omega_1^4 \end{pmatrix} \right\}.$$

In practice it is obviously essential to replace  $d^{(21)}$  and  $\Gamma^{(21)}$  by consistent estimators in order to make the regression theory above feasible. In the previous section we suggested two estimators for the asymptotic covariance matrix.

**Proposition 8** *Under the conditions given in Theorem 2, as  $n \rightarrow \infty$*

$$\frac{n^{1/4}(\hat{\beta}_n^{(21)} - \beta^{(21)})}{\sqrt{\left(MRC(Z^*)_n^{1,1}\right)^{-2} \hat{g}_n^{21}}} \xrightarrow{D} N(0, 1) \quad (5.5)$$

and

$$\frac{n^{1/4}(\hat{\beta}_n^{(21)} - \beta^{(21)})}{\sqrt{\left(MRC(Z^*)_n^{1,1}\right)^{-2} \tilde{g}_n^{21}}} \xrightarrow{D} N(0, 1). \quad (5.6)$$

Here  $\hat{g}_n^{21} = \hat{d}_n^{(21)'} \hat{\Gamma}_n^{(21)} \hat{d}_n^{(21)}$ ,  $\tilde{g}_n^{21} = \hat{d}_n^{(21)'} \tilde{\Gamma}_n^{(21)} \hat{d}_n^{(21)}$ ,

$$\hat{\Gamma}_n^{21} = \frac{1}{\sqrt{n}\theta^2\psi_2^2} \sum_{i=0}^{n-k_n+1} \begin{pmatrix} (\bar{X}_i^{*n})^2 (\bar{Y}_i^{*n})^2 & \bullet \\ (\bar{X}_i^{*n})^3 \bar{Y}_i^{*n} & (\bar{X}_i^{*n})^4 \end{pmatrix} \quad (5.7)$$

$$+ \sum_{j=1}^{k_n-1} \begin{pmatrix} \bar{X}_i^{*n} \bar{Y}_i^{*n} \left( \bar{X}_{i+j}^{*n} \bar{Y}_{i+j}^{*n} - \bar{X}_{i+k_n}^{*n} \bar{Y}_{i+k_n}^{*n} \right) & \bullet \\ (\bar{X}_i^{*n})^2 \left( \bar{X}_{i+j}^{*n} \bar{Y}_{i+j}^{*n} - \bar{X}_{i+k_n}^{*n} \bar{Y}_{i+k_n}^{*n} \right) & (\bar{X}_i^{*n})^2 \left( (\bar{X}_{i+j}^{*n})^2 - (\bar{X}_{i+k_n}^{*n})^2 \right) \end{pmatrix}$$

$$+ \sum_{j=1}^{k_n-1} \begin{pmatrix} \bar{X}_i^{*n} \bar{Y}_i^{*n} \left( \bar{X}_{i+j}^{*n} \bar{Y}_{i+j}^{*n} - \bar{X}_{i+k_n}^{*n} \bar{Y}_{i+k_n}^{*n} \right) & \bullet \\ \bar{X}_i^{*n} \bar{Y}_i^{*n} \left( (\bar{X}_{i+j}^{*n})^2 - (\bar{X}_{i+k_n}^{*n})^2 \right) & (\bar{X}_i^{*n})^2 \left( (\bar{X}_{i+j}^{*n})^2 - (\bar{X}_{i+k_n}^{*n})^2 \right) \end{pmatrix},$$

$$\tilde{\Gamma}_n^{21} = \begin{pmatrix} \tilde{H}_n^{22} & \bullet \\ \tilde{H}_n^{12} & \tilde{H}_n^{11} \end{pmatrix} \quad (5.8)$$

and  $H_n$  and  $\tilde{H}_n$  have been introduced in (4.1) and (4.5).

Obviously, all the required terms are straightforward to compute, so it is rather easy to implement the estimators.

## 5.2 Modulated realised correlation

We can apply the same strategy in order to derive the asymptotic distribution of the modulated realised correlation coefficient. Generally, the correlation between two stochastic processes  $X$

and  $Y$  is defined as

$$\rho^{(21)} = \frac{[X, Y]}{\sqrt{[X][Y]}}.$$

In the case of  $Z = (X, Y)$  being a Brownian semimartingale this reduces to the much simpler form

$$\rho^{(21)} = \frac{\int_0^1 \Sigma_u^{12} du}{\sqrt{\int_0^1 \Sigma_u^{11} du \cdot \int_0^1 \Sigma_u^{22} du}}. \quad (5.9)$$

The modulated realised correlation coefficient is defined as

$$\hat{\rho}_n^{(21)} = \frac{MRC(Z^*)_n^{1,2}}{\sqrt{MRC(Z^*)_n^{1,1} MRC(Z^*)_n^{2,2}}}.$$

From Theorem 1 it naturally follows that it converges to (5.9). Again its asymptotic distribution is derived from Theorem 2 using the  $\Delta$ -method.

**Theorem 9** *Under the conditions in Theorem 2, as  $n \rightarrow \infty$*

$$\frac{n^{1/4} (\hat{\rho}_n^{(21)} - \rho^{(21)})}{\sqrt{\left(\int_0^1 \Sigma_u^{11} du \int_0^1 \Sigma_u^{22} du\right)^{-1} g^{21}}} \xrightarrow{D} N(0, 1) \quad (5.10)$$

Here  $g^{21} = d^{(21)'} \Gamma^{(21)} d^{(21)}$ , where  $\Gamma^{(21)} = L$  and  $d^{(21)} = (-\frac{1}{2}\beta^{(21)}, 1, -\frac{1}{2}\beta^{(12)})$ . Furthermore,  $\beta^{(21)}$  is given by (5.3).

The infeasible results can be used in practice by applying the following result.

**Proposition 10** *Under conditions given in Theorem 2, as  $n \rightarrow \infty$*

$$\frac{n^{1/4} (\hat{\rho}_n^{(21)} - \rho^{(21)})}{\sqrt{\left(MRC(Z^*)_n^{1,1} MRC(Z^*)_n^{2,2}\right)^{-1} \hat{g}_n^{(21)}}} \xrightarrow{D} N(0, 1) \quad (5.11)$$

and

$$\frac{n^{1/4} (\hat{\rho}_n^{(21)} - \rho^{(21)})}{\sqrt{\left(MRC(Z^*)_n^{1,1} MRC(Z^*)_n^{2,2}\right)^{-1} \tilde{g}^{(21)}}} \xrightarrow{D} N(0, 1). \quad (5.12)$$

Here  $\hat{g}_n^{21} = \hat{d}_n^{(21)'} \hat{\Gamma}_n^{(21)} \hat{d}_n^{(21)}$  and  $\tilde{g}_n^{21} = \hat{d}_n^{(21)'} \tilde{\Gamma}_n^{(21)} \hat{d}_n^{(21)}$ , where  $\hat{d}_n^{(21)} = \left(-\frac{1}{2}\hat{\beta}_n^{(21)}, 1, -\frac{1}{2}\hat{\beta}_n^{(12)}\right)$ , and  $\hat{\Gamma}_n^{(21)}$  and  $\tilde{\Gamma}_n^{(21)}$  are defined as in (5.7) and (5.8).

## 6 Simulation Study

In this section we study the finite sampling behaviour of the modulated realised covariation estimator by carrying out some Monte Carlo experiments. We will focus on the following questions.

- How good is modulated realised volatility, covariance, regression and correlation as an estimator of integrated volatility, covariance, actual regression and correlation?
- How close to standard normality is the finite sample performance of both the unfeasible test statistics (3.12), (5.4) and (5.10) and the feasible test statistics ((4.3), resp. (4.6), (5.5), resp. (5.6) and (5.11), resp. (5.12) for different values of  $n$ ?

### 6.1 Simulation Design

Throughout this section we work with a bivariate stochastic volatility model which is very similar to the model that Barndorff-Nielsen & Shephard (2004) use. Their theory holds under the assumption that the volatility process and the Brownian motion which drives the diffusion process are uncorrelated. However, empirical studies have revealed that asset returns and volatility tend to be negatively correlated (which is called leverage effect). The results of this paper have been proved allowing for leverage. In order to reflect this in the simulation study, we use the following slightly modified model:

$$dZ_t = \sigma_t dW_t, \quad \Sigma_t = \sigma_t \sigma_t'$$

where

$$\Sigma_t = \begin{pmatrix} \Sigma_t^{11} & \Sigma_t^{12} \\ \Sigma_t^{12} & \Sigma_t^{22} \end{pmatrix} = \begin{pmatrix} (\sigma_t^X)^2 & \sigma_t^X \sigma_t^Y \rho_t \\ \sigma_t^X \sigma_t^Y \rho_t & (\sigma_t^Y)^2 \end{pmatrix}.$$

This model goes back to Barndorff-Nielsen and Shephard (2002) who used realised variances to fit the variance of the DM/Dollar rate from 1986 to 1996 by the sum of two uncorrelated processes

$$(\sigma_t^X)^2 = (\sigma_t^{X(1)})^2 + (\sigma_t^{X(2)})^2.$$

In particular, the processes  $(\sigma_t^{X(1)})^2$  and  $(\sigma_t^{X(2)})^2$  satisfy the following SDE.

$$\left(\sigma_t^{X(s)}\right)^2 = -\lambda_s \left\{ \left(\sigma_t^{X(s)}\right)^2 - \xi_s \right\} dt + \omega_s \left(\sigma_t^{X(s)}\right)^2 dB_{\lambda_s t}^s, \quad \xi_s \geq \omega_s^2/2, \quad (6.1)$$

where  $B$  is a bivariate vector of standard Brownian motion. In order to incorporate leverage into the model, we use the following construction:

$$\text{corr}(B_t^s, W_t^s) = -0.7, \quad s = 1, 2.$$

The process (6.1) has a gamma marginal distribution

$$d\left(\sigma_t^{X(s)}\right)^2 \sim Ga\left(2\omega_s^{-2}\xi_s, 2\omega_s^{-2}\right) = Ga\left(v_s, a_s\right), \quad v_s \geq 1,$$

with a mean of  $v_s/a_s$  and a variance of  $v_s/a_s^2$ . The parameters  $\omega_2$ ,  $\lambda_s$  and  $\xi_s$  were calibrated by Barndorff-Nielsen and Shephard (2002b) as follows. They estimated

$$E\left[\left(\sigma_t^{X(s)}\right)^2\right] = p_s \cdot 0.509, \quad \text{Var}\left[\left(\sigma_t^{X(s)}\right)^2\right] = p_s \cdot 0.461, \quad s = 1, 2,$$

$$p_1 = 0.218, \quad p_2 = 0.782, \quad \lambda_1 = 0.0429, \quad \lambda_2 = 3.74.$$

The model for  $(\sigma_t^Y)^2$  takes the form

$$d\left(\sigma_t^Y\right)^2 = -0.35 \left\{ \left(\sigma_t^Y\right)^2 - 0.636 \right\} dt + 0.236 \left(\sigma_t^Y\right)^2 dB_t^3.$$

Finally, the model for  $\rho_t$  is specified as  $\rho_t = \frac{\exp(2x_t)-1}{\exp(2x_t)+1}$ , where  $x_t$  follows the GARCH diffusion

$$dx_t = -0.03 \{x_t - 0.64\} dt + 0.118x_t dB_t^4.$$

Again, we assume the observed price process  $Z^*$  to be a decomposition of the efficient price process  $Z$  and a noise process  $\epsilon$ . We assume the noise process to be i.i.d. normally distributed with mean 0 and covariance matrix

$$\Psi = \begin{pmatrix} 0.0005 & 0.0001 \\ 0.0001 & 0.0005 \end{pmatrix}.$$

To produce an impression we have drawn Figures 1(a) and 1(b). They report results based on 1024 observations per day which on an average trading day from 9.30 am to 4.00 pm corresponds to one observation every 22.8 seconds. We are simulating up to 200 trading days. Figure 1(a) shows the first 10 days of the sample, plotting the bivariate 30 minute return data. The  $x$ -axis represents days in this picture. Figure 1(b) shows the daily returns drawn against the trading days.

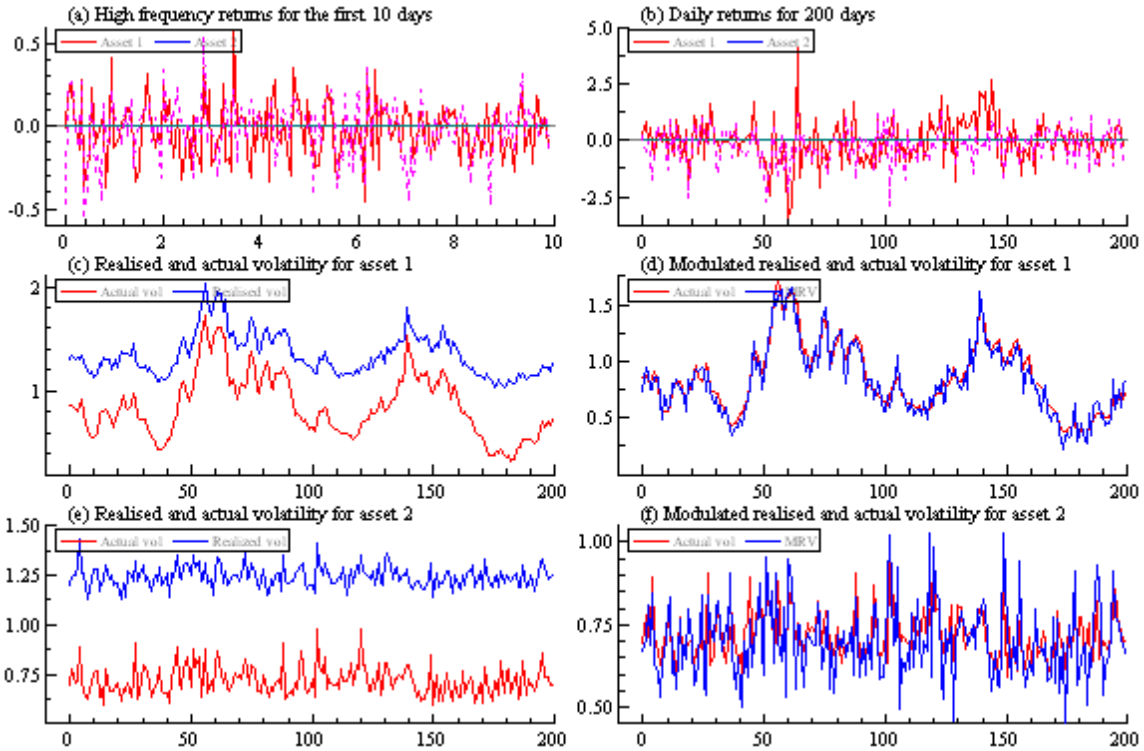


Figure 1: Simulated bivariate stochastic volatility model using  $n = 1024$ : (a) 30 minute returns  $\Delta_i^n X$ ,  $\Delta_i^n Y$  for 10 trading days; (b) daily returns  $\sum_{i=1}^n \Delta_i^n X$  and  $\sum_{i=1}^n \Delta_i^n Y$  for 200 trading days; (c) realised volatility and actual volatility for asset 1; (d)  $\sqrt{MRC(Z^*)_n^{1,1}}$  and actual volatility for asset 1; (e) realised volatility and actual volatility for asset 1; (f)  $\sqrt{MRC(Z^*)_n^{2,2}}$  and actual volatility for asset 2;

## 6.2 Simulation Results

Jacod et al. (2007) report that the univariate estimator is fairly robust to the choice of  $k_n$  and as  $\theta$  comes from asymptotic statistics, it does not give any precise instruction about the choice of  $k_n$  for small values of  $n$ . They suggest to choose  $\theta = 1/3$  for simulations.

Jacod et al. (2007) show the importance of a correction of the univariate version of modulated realised covariation, when dealing with small sampling sizes. Therefore they suggest the replacement of the parameters  $\psi_i$  and  $\Phi_{ij}$  by their finite sample analogues. Set  $g_i^n = g(i/k_n)$ .

$$\psi_1^{k_n} = k_n \sum_{i=1}^{k_n} (g_{i+1}^n - g_i^n)^2, \quad \psi_2^{k_n} = k_n^{-1} \sum_{i=1}^{k_n-1} (g_i^n)^2,$$

$$\phi_1^{k_n}(j) = \sum_{i=j+1}^{k_n-1} (g_{i-1}^n - g_i^n) (g_{i-j-1}^n - g_{i-j}^n), \quad \phi_2^{k_n}(j) = \sum_{i=j+1}^{k_n-1} g_i^n g_{i-j}^n,$$

$$\begin{aligned} \Phi_{11}^{k_n} &= k_n \left( \sum_{j=0}^{k_n-1} \left( \phi_1^{k_n}(j) \right)^2 - \frac{1}{2} \left( \phi_1^{k_n}(0) \right)^2 \right), & \Phi_{12}^{k_n} &= \frac{1}{k_n} \left( \sum_{j=0}^{k_n-1} \phi_1^{k_n}(j) \phi_2^{k_n}(j) - \frac{1}{2} \phi_1^{k_n}(0) \phi_2^{k_n}(0) \right) \\ \Phi_{22}^{k_n} &= \frac{1}{k_n^3} \left( \sum_{j=0}^{k_n-1} \left( \phi_2^{k_n}(j) \right)^2 - \frac{1}{2} \left( \phi_2^{k_n}(0) \right)^2 \right) \end{aligned}$$

The parameters stated above are the ones which are used in the proof, but each of them converges at a smaller order than  $n^{-1/4}$ . We will use these finite sample expressions when applying  $H_n$  as an estimator of the asymptotic covariance matrix. Unfortunately, computations have shown that it is not possible to do so, when applying  $\tilde{H}_n$  as otherwise  $C^{(1)}(g_1, g_2, g_3)$ ,  $C^{(2)}(g_1, g_2, g_3)$  or  $C^{(3)}(g_1, g_2, g_3)$  in (4.5) can get negative which is not feasible. So we will use the original constants for  $\tilde{H}_n$ .

Figure 1(c) shows the realised volatility for asset 1  $\sqrt{\sum_{i=1}^n (\Delta_i^n X)^2}$  together with its corresponding actual volatility  $\sqrt{\int_0^1 (\sigma_s^X)^2 ds}$ . The corresponding results for asset 2 are given in Figure 1(e). This time series looks very jagged, reflecting the fast mean-reverting component in this process. We can see the substantial positive bias that realised variance exhibits due to market microstructure noise effects. Figure 1(d) shows  $\sqrt{MRC(Z^*)_{n,1}^{1,1}}$  plotted against the integrated volatility and Figure 1(f) the same for asset 2. Under the realistic assumption of sampling every 22.8 seconds modulated realised covariation performs significantly better than realised variance. It does not exhibit any bias, it also captures the time series of actual volatility rather well, but it is obviously still noisy. However, assuming  $n = 4096$ , which corresponds to trading every 4.8 seconds on an 8.5 hour trading day, (see Figures 12 (c) -(f) in the Appendix) we obtain very good results. We see that for this sampling frequency modulated realised variation captures the original time series very well and there hardly any large differences between

the actual and the estimated volatility.

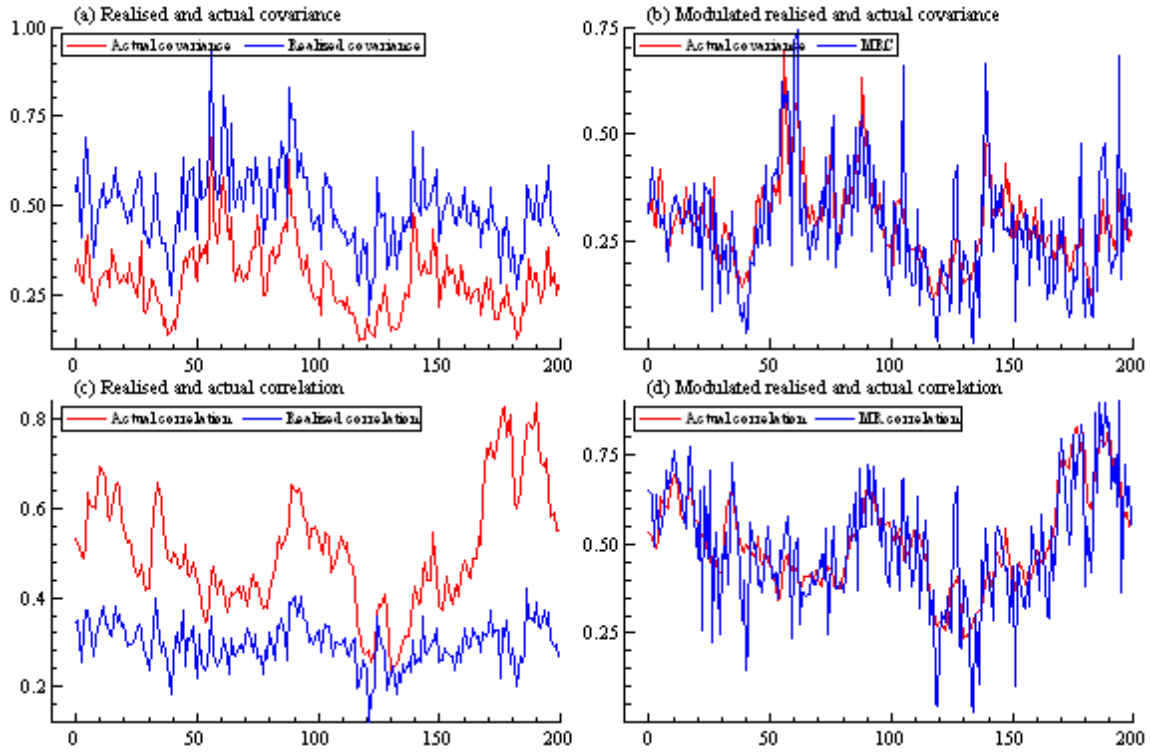


Figure 2: Simulation of measures of dependence, using  $n = 1024$ : (a) realised covariance and actual covariance; (b)  $MRC(Z^*)_{n}^{1,2}$  and actual covariance; (c) realised and actual correlation; (d) MRC based and actual correlation.

Table 1: Pearson correlation coefficients for regressions of realised volatility and modulated realised volatility on integrated volatility

$n$	RV vs IV	MRV vs IV
256	0.109794	0.63512
4096	0.000111	0.895634
65536	$2.05486 \exp(-6)$	0.998934

In order to gain a better understanding of how an increasing sampling frequency improves the estimator, we look at the Pearson coefficient regression of modulated realised volatility against integrated volatility. For  $n = 256$  it is 0.63 and thus substantially above the Pearson coefficient for the relationship between realised volatility and integrated volatility which is 0.11. For a sampling frequency corresponding to  $n = 4096$ , realised volatility is not very useful anymore, whereas regressing modulated realised volatility on integrated volatility yields a Pearson coefficient of 0.9. Finally under the rather unrealistic assumption of  $n = 65536$ , the



Pearson coefficient is almost 1.

Figure 2(a) gives the corresponding results for covariance and correlation between the two sets of asset returns. Figure 2(a) depicts  $\sum_{i=1}^n \Delta_i^n X \cdot \Delta_i^n Y$  and  $\int_0^1 \sigma_s^X \sigma_s^Y \rho_s ds$  the realised and actual covariances for each trading day. Figure 3(b) shows  $MRC(Z^*)_{n,2}^{1,2}$  and  $\int_0^1 \sigma_s^X \sigma_s^Y \rho_s ds$  for each trading day. Just like realised variance, realised covariance exhibits a substantial positive bias which increases with the sampling frequency. However, compared to realised volatility the bias is significantly smaller which is due to the fact that the bias is of order  $2\omega^2 n$  for realised volatility and  $\omega_1 \omega_2 c_{12} n$  for realised covariance. So the bias of realised covariance depends on the assumption on the correlation of the noise. In this work we assume it to be equal to 0.2, but there has been strong indication in the literature that it might be equal to 0. In Figure 2(b) ( $n = 1024$ ) modulated realised covariation is still a very noisy estimator. However, with  $n = 4096$  we obtain a rather good result in Figure 13 in the Appendix. In both Figures 2(c)

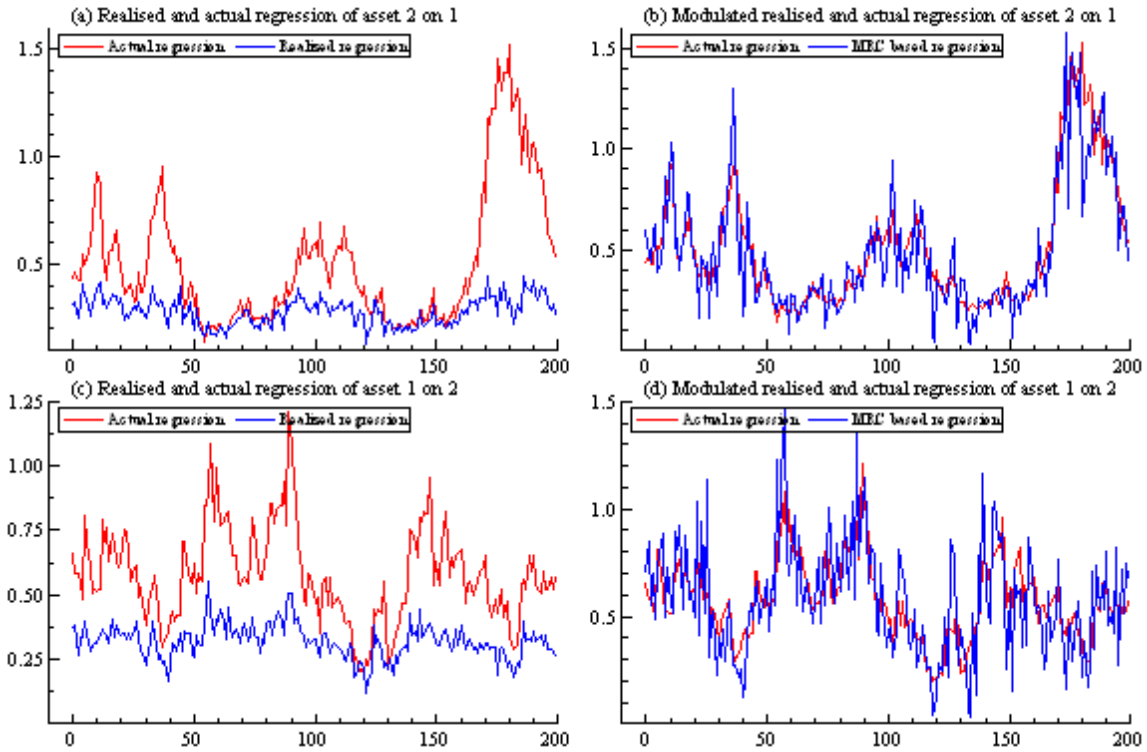


Figure 3: Simulations of measures of dependence, using  $n = 1024$ ; (a) realised and actual regressions of returns on asset 2 on asset 1; (b) MRC based and actual regressions of returns on asset 2 on asset 1; (c) same as (a) but asset 1 on 2; (d) same as (b) but asset 1 on 2;

and 13(c), resp. 2(d) and 13(d) we depict the realised correlation

$$\frac{\sum_{i=1}^n \Delta_i^n X \Delta_i^n Y}{\sqrt{\sum_{i=1}^n (\Delta_i^n X)^2 \cdot \sum_{i=1}^n (\Delta_i^n Y)^2}}, \quad \frac{MRC(Z^*)_n^{1,2}}{\sqrt{MRC(Z^*)_n^{1,1} \cdot MRC(Z^*)_n^{2,2}}}$$

and the actual correlation

$$\frac{\int_0^1 \sigma_s^X \sigma_s^Y \rho_s ds}{\sqrt{\int_0^1 (\sigma_s^X)^2 ds \cdot \int_0^1 (\sigma_s^Y)^2 ds}}$$

amongst two time series. The bias induced by market microstructure noise effects is negative in that case which reflects the fact that realised variance exhibits a larger positive bias than realised covariance. Again modulated realised correlation plotted in Figures 2(d) and 13(d) manages to overcome this issue, however, it is still rather noisy for  $n = 1024$ .

Table 2: Pearson correlation coefficients for regressions of realised covariance and modulated realised covariance on integrated covariance

$n$	RC vs IC	MRC vs IC
256	0.409953	0.611243
4096	0.858672	0.676787
65536	0.489819	0.99425

A similar picture appears in Figures 3 and 13 where realised, modulated realised and actual regression coefficient are plotted for  $n = 1024$  and 4096 for 200 trading days, i.e. we look at the statistics

$$\frac{\sum_{i=1}^n \Delta_i^n X \Delta_i^n Y}{\sum_{i=1}^n (\Delta_i^n X)^2}, \quad \frac{MRC(Z^*)_n^{1,2}}{MRC(Z^*)_n^{1,1}}, \quad \frac{\int_0^1 \sigma_s^X \sigma_s^Y \rho_s ds}{\int_0^t (\sigma_s^X)^2 ds}.$$

and the corresponding regressions for asset 1 on asset 2. Again, the realised statistics turn out to be negatively biased and rather noisy. However, also modulated realised regression exhibits a rather noisy behaviour at a sampling frequency corresponding to  $n = 1024$ .

### 6.3 Assessing the Performance of the Feasible Asymptotic Theory

In this section we will examine how close to standard normality the finite sample performance of the standardised test statistics of modulated realised covariance, regression and correlation is. We will consider both the unfeasible and the feasible theory and compare the two estimators

for the asymptotic covariance matrix,  $H_n$  and  $\tilde{H}_n$ , that we introduced in section 4. As  $H_n$  is not positive semidefinite, some realisations of the simulation happen to result in a negative variance. This obviously does not happen when using  $\tilde{H}_n$  as it is positive definite. On the other hand,  $\tilde{H}_n$  tends to underestimate the asymptotic variance for small values of  $n$ , e.g.  $n = 256$ .

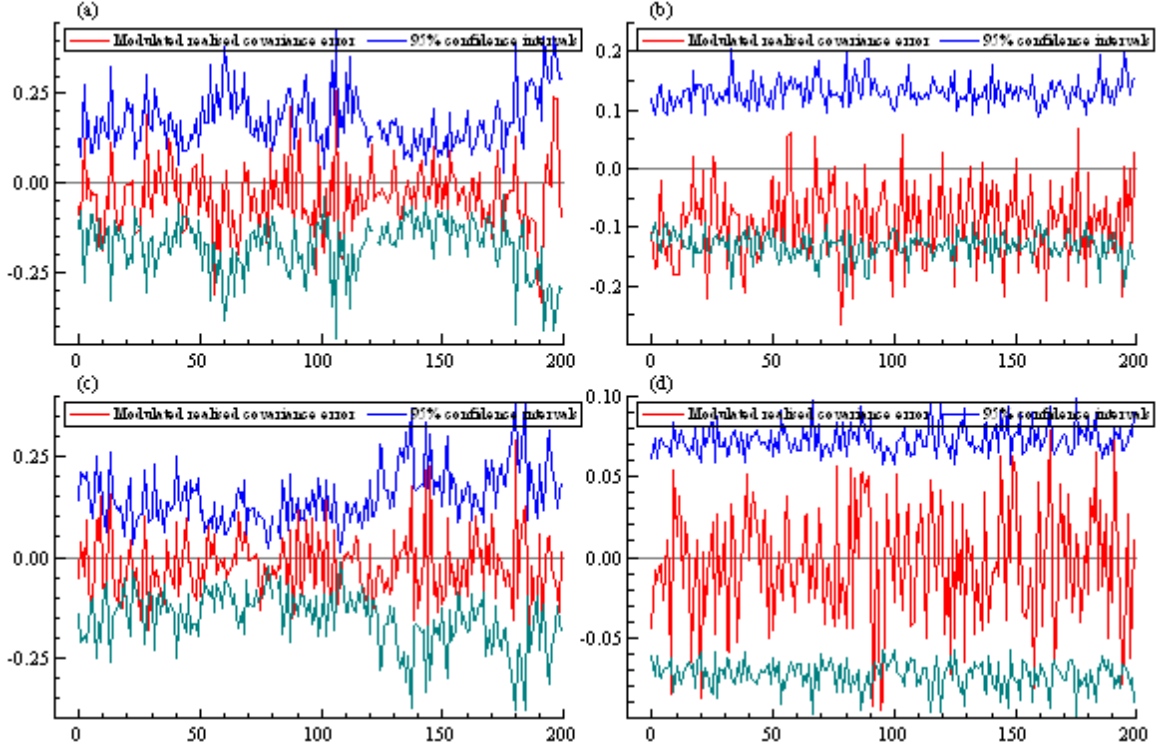


Figure 4: Modulated realised covariance. The modulated realised covariance errors and their 95% confidence intervals w.r.t.  $\alpha_n$  resp.  $\tilde{\alpha}_n$ ; (a)  $n = 256$ ,  $\alpha_n$ . (b)  $n = 256$ ,  $\tilde{\alpha}_n$ . (c)  $n = 4096$ ,  $\alpha_n$ . (d)  $n = 4096$ ,  $\tilde{\alpha}_n$ .

The asymptotic theory for modulated realised covariance tells us that the normalized estimation error

$$\frac{n^{1/4} \left( MRC(Z_n^*) - \int_0^1 \Sigma_u^{12} du \right)}{\alpha_n} \xrightarrow{D} N(0, 1) , \quad (6.2)$$

and

$$\frac{n^{1/4} \left( MRC(Z_n^*) - \int_0^1 \Sigma_u^{12} du \right)}{\tilde{\alpha}_n} \xrightarrow{D} N(0, 1) , \quad (6.3)$$

where  $\alpha_n$  and  $\tilde{\alpha}_n$  are given by (4.4) and (4.7). We want to find out how close to normality this ratio is for small and moderate values of  $n$ . Figure 4 plots the modulated realised covariance errors,  $MRC(Z_n^*) - \int_0^1 \Sigma_u^{12} du$ , for each trading day. Hereby we use the Monte Carlo design

discussed in the previous section. The figure also gives the 95% confidence intervals for the errors generated using the feasible limit theory (6.2) and (6.3). We can observe a slightly negative bias for  $n = 256$ . With an increasing value of  $n$  we can see a decrease in the spread of the error and in the confidence intervals for both  $\alpha_n$  and  $\tilde{\alpha}_n$ . Furthermore, we observe that the analysis based on  $\tilde{\alpha}_n$  produces significantly smaller confidence intervals than using  $\alpha_n$ , but is rather unreliable for small values of  $n$ . For example, for  $n = 256$  the standard error exceeds the lower 95% quantile rather often.

We back up the coverage of the limit theory by giving the t-statistics in Table 15 in the Appendix. We repeat the above analysis, but using 20,000 days and focus on the distribution of (6.2) and (6.3) and (3.12).

We observe that the confidence intervals are captured very well already for small values of  $n$ , particularly in the unfeasible case. In the feasible case even for  $n = 65536$  the mean of the standard error is  $-0.11$  when using  $\tilde{H}_n$  as an estimator for the asymptotic covariance matrix. A similar effect has been observed in Jacod, Li, Mykland, Podolskij and Vetter (2007). They observe a non-vanishing standardised asymptotic mean when using the various constants instead of their Taylor expansions. On the other hand, we see an effect of  $H_n$ , the other estimator for the asymptotic covariance matrix, not being positive semidefinite. As the variance can get negative, we do not obtain a value of either mean or standard deviation of the standard error for  $n = 256$ .

The limit theory for the normalized estimation error for modulated realised regression of the returns of asset 2 on asset 1 is

$$\frac{n^{1/4}(\hat{\beta}^{(lk)} - \beta^{(lk)})}{\sqrt{(MRC(Z^*))^{-2} \hat{g}^{lk}}} \xrightarrow{D} N(0, 1) \quad (6.4)$$

$$\frac{n^{1/4}(\hat{\beta}^{(lk)} - \beta^{(lk)})}{\sqrt{(MRC(Z^*))^{-2} \tilde{g}^{lk}}} \xrightarrow{D} N(0, 1). \quad (6.5)$$

Figure 5 shows the modulated realised regression errors,  $\hat{\beta}^{(21)} - \beta^{(21)}$ , plotted for each trading day, together with 95% asymptotic confidence intervals based on the asymptotic limit theory (6.4), resp. (6.5). Again we observe that the confidence intervals relying on (6.5) are smaller than the ones based on (6.4). Furthermore, they are more reliable for small values of  $n$  than

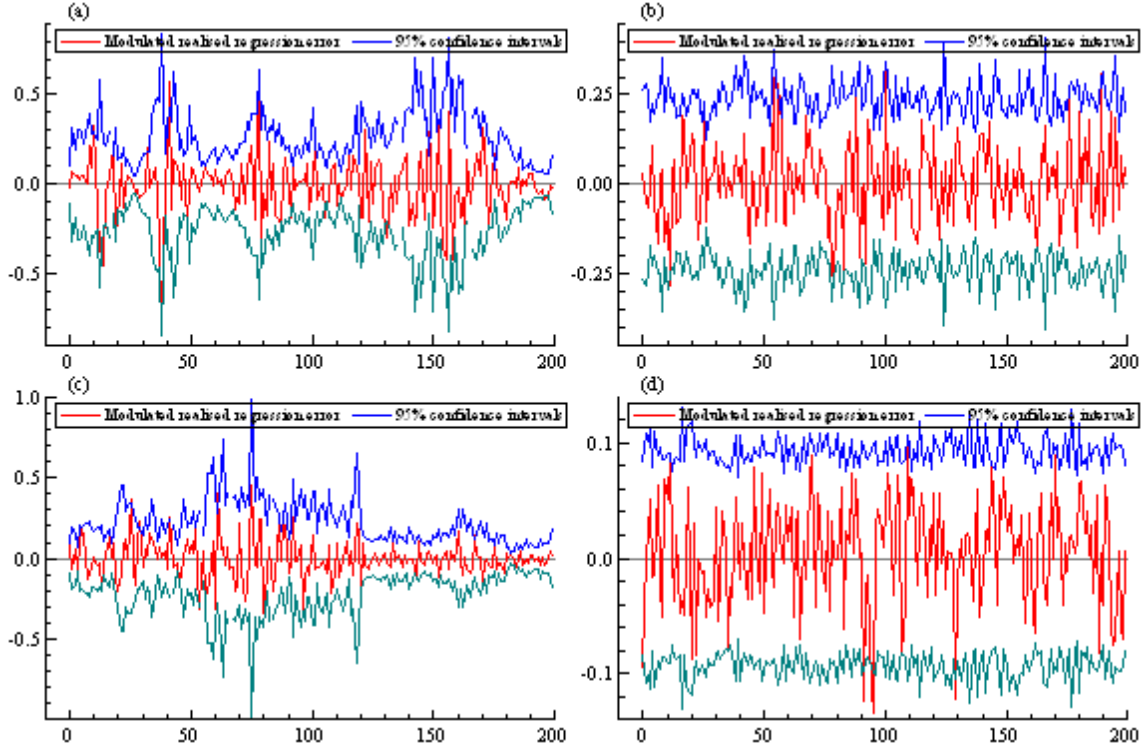


Figure 5: Modulated realised regression. The modulated realised regression errors and their 95% confidence intervals based on (6.4) resp. (6.5). (a)  $n = 256$ , (6.4). (b)  $n = 256$ , (6.5). (c)  $n = 4096$ , (6.4), (d)  $n = 4096$ , (6.5)

in the case of modulated realised covariance.

Again we show the results of the asymptotic analysis of the  $t$ -statistics in the Appendix for both (5.6) and the unfeasible theory given by (5.4).

For modulated realised regression, the confidence intervals converge even faster than in the covariance case, just like mean and variance of the standard error. However, for small values of  $n$  we can obtain negative variance when using  $H_n$  to make the theory feasible. The feasible limit theory for correlation of the returns of asset 1 and asset 2 is

$$\frac{n^{1/4}\hat{\rho}^{(21)} - \rho^{(21)}}{\sqrt{\left(MRC(Z^*)_n^{1,1} MRC(Z^*)_n^{2,2}\right)^{-1} \hat{g}^{(21)}}} \xrightarrow{D} N(0, 1) \quad (6.6)$$

and

$$\frac{n^{1/4}\hat{\rho}^{(21)} - \rho^{(21)}}{\sqrt{\left(MRC(Z^*)_n^{1,1} MRC(Z^*)_n^{2,2}\right)^{-1} \tilde{g}^{(21)}}} \xrightarrow{D} N(0, 1). \quad (6.7)$$

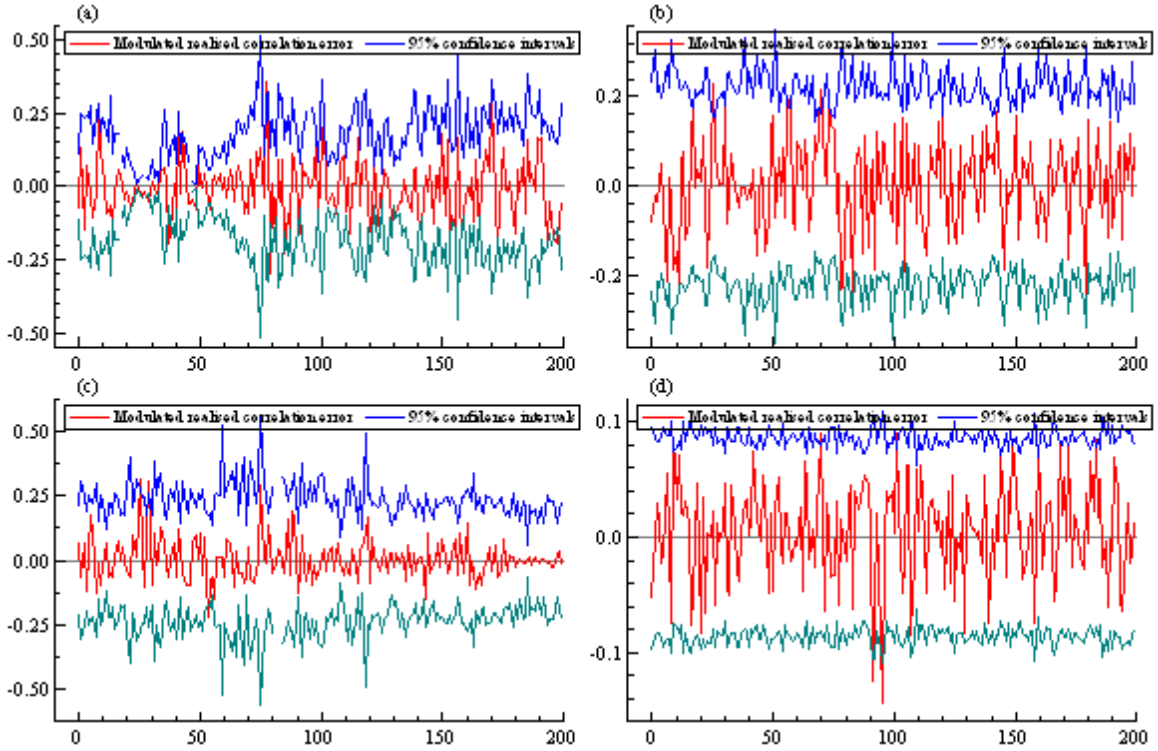


Figure 6: Modulated realised correlation. The modulated realised correlation errors and their 95% confidence intervals based on (6.6) resp. (6.7). (a)  $n = 256$ , (6.6). (b)  $n = 256$ , (6.7). (c)  $n = 4096$ , (6.6), (d)  $n = 4096$ , (6.7)

The graphs in Figure 6 show the correlation errors  $\hat{\rho}^{(21)} - \rho^{(21)}$ .

The scaling  $\left( MRC(Z^*)_{n,1,1}^{1,1} MRC(Z^*)_{n,2,2}^{2,2} \right)^{-1}$  adjusts the denominator in (6.6) and (6.7) to make it invariant as we scale either of the asset returns within each time period. This suggests it should be less sensitive to changes in the level of volatility in either of the assets.

## 7 Empirical Illustration

Let us finally apply the theory to some real data. To illustrate some of the empirical features of modulated realised volatility, covariance, regression and correlation and in particular their precision as estimators of the actual quantities, we perform an empirical study which focuses on high frequency equity quote data. For comparison reasons we consider two bivariate data series of Lehman Brothers and Merrill Lynch intraday TAQ data resp. IBM and UTX intraday TAQ data, available at WRDS. The time series starts on 1st March 2005 and covers the next 50 trading days. Before analysing the data we have cleaned the data. Following the methods

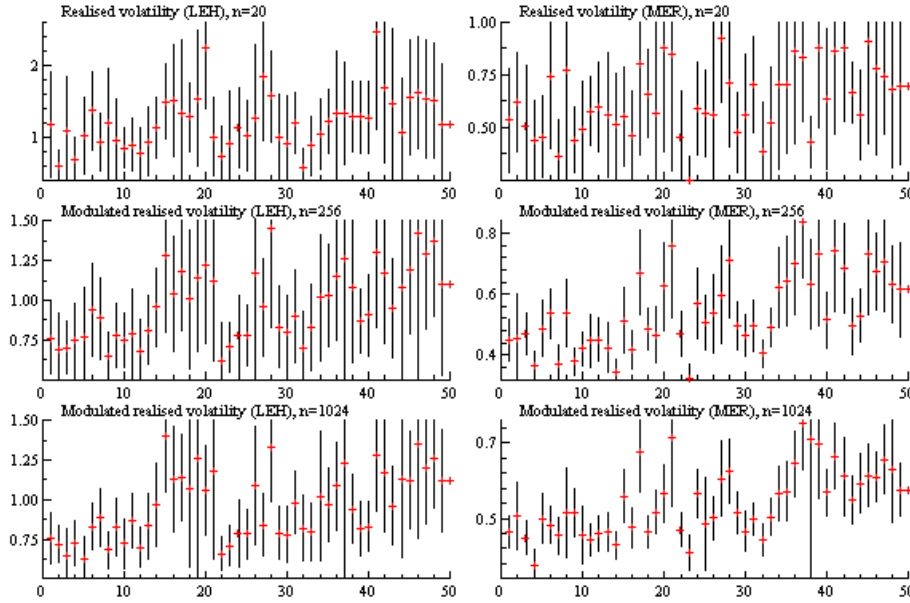


Figure 7: Lehman Brothers (first column) and Merrill Lynch (second column) equity data. Data are 1st March 2005 onwards for 50 active trading days. (a) Realised volatility for Lehman based on 20 intra-day observations. This is marked with a cross while the bars denote 95% confidence intervals. (b) Modulated realised volatility for Lehman based on 256 intra-day observations. (c) Modulated realised for Lehman volatility based on 1024 intra-day observations. (d) Realised volatility for Merrill Lynch based on 20 intra-day observations. This is marked with a cross while the bars denote 95% confidence intervals. (e) Modulated realised volatility for Merrill Lynch based on 256 intra-day observations. (f) Modulated realised volatility for Merrill Lynch based on 1024 intra-day observations. All the confidence intervals of modulated realised covariation are based on  $\tilde{H}_n$ .

used by Hansen and Lunde (2006) we concentrate on quote data from one stock exchange only. Here we have chosen NYSE. We only consider quotes, where both the bidsize and the asksize are greater than 0 and which are quoted in a normal trading environment (quote condition = 12 in the TAQ database). We concentrate on data from 9.30 pm to 4pm and only consider offer prices. In order to construct a time series of the required sampling frequency, we use a "bivariate previous-tick method" which is inspired by work by Barndorff-Nielsen et al. (2008). In the following we will describe this method in some detail.

In order to avoid zero-returns which becomes a problem when it comes to the estimation of covariation, in a first step we extract a time series where the only requirement is that consecutive prices differ in absolute values. In a second step we extract time series sampled under a certain sampling frequency. This works as follows: We compare the two univariate non-zero-return time series w.r.t. their length and declare the smaller time series to be the leading time series. We

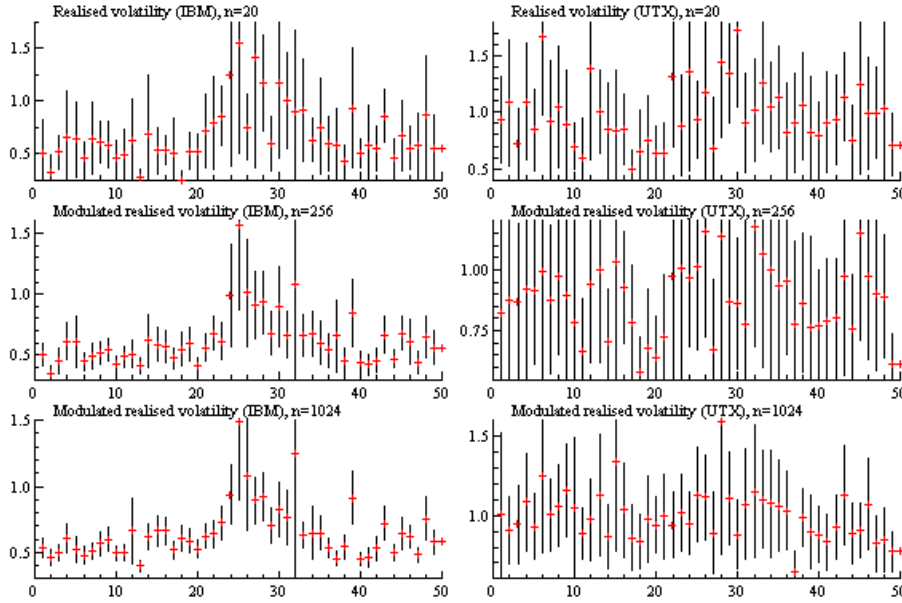


Figure 8: IBM (first column) and UTX (second column) equity data. Data are 1st March 2005 onwards for 50 active trading days. (a) Realised volatility for IBM based on 20 intra-day observations. This is marked with a cross while the bars denote 95% confidence intervals. (b) Modulated realised volatility for IBM based on 256 intra-day observations. (c) Modulated realised for IBM volatility based on 1024 intra-day observations. (d) Realised volatility for UTX based on 20 intra-day observations. This is marked with a cross while the bars denote 95% confidence intervals. (e) Modulated realised volatility for UTX based on 256 intra-day observations. (f) Modulated realised volatility for UTX based on 1024 intra-day observations. All the confidence intervals of modulated realised covariation are based on  $\tilde{H}_n$ .

take the first data point of the leading time series and the closest next data point of the second time series and record them as the first datapoint of the synchronised bivariate time series. Next we pick the next data point of the leading time series under consideration of the desired sampling frequency and repeat the procedure. Obviously, this works the better, the more data points remain after the first step of extraction. After having cleaned the data, we obtain 6 bivariate time series consisting of 50 business days with 20 observations of data sampled every 20 minutes, 256 observations of data sampled every 1.7 minutes and 1024 observations for data sampled every 26 seconds.

We have computed the realised volatility for a sampling frequency of 20 minutes and the modulated realised volatility for higher sampling frequencies and the respective 95% confidence intervals for the Lehman and Merrill Lynch (Figure 7) and IBM and UTX (Figure 8) equity data. The confidence intervals are based on the asymptotic covariance estimator  $\tilde{H}_n$  which was introduced in (4.5). We see in the summary statistics in Tables 3 and 4 that modulated



realised volatility is very close to the 6-month-open-to-close realised volatility. Note that the tables obtain information about the standard error being estimated both by  $H_n$  and  $\tilde{H}_n$ .

estimator	average	standard deviation			bias
		( $H_n$ for MRV)	( $\tilde{H}_n$ for MRV)	empirical	
MRV (256, LEH)	0.972	0.164	0.188	0.222	-0.022
MRV (1024, LEH)	0.956	0.112	0.140	0.212	-0.038
RV (20min, LEH)	0.925	0.505		0.632	-0.069
RV (OtoC, LEH)	0.994				
MRV (256, MER)	0.544	0.09	0.06	0.123	-0.024
MRV (1024, MER)	0.543	0.06	0.041	0.09	-0.025
RV (20min, MER)	0.553	0.273		0.348	-0.015
RV (OtoC, MER)	0.568				

Table 3: LEH and MER equity data January to June 2005. Summary statistics for modulated realised volatility for  $n = 256$  and  $n = 1024$ . The same statistics are computed for the corresponding realised volatility ( $n = 20$ ) and the open-to-close realised volatility.

estimator	average	standard deviation			bias
		( $H_n$ for MRV)	( $\tilde{H}_n$ for MRV)	empirical	
MRV (256, IBM)	0.624	0.175	0.188	0.235	-0.027
MRV (1024, IBM)	0.653	0.149	0.151	0.219	0.002
RV (20min, IBM)	0.724	0.200		0.285	0.073
RV (OtoC, IBM)	0.651				
MRV (256, UTX)	0.882	0.162	0.179	0.143	-0.081
MRV (1024, UTX)	0.998	0.099	0.141	0.153	0.036
RV (20min, UTX)	1.069	0.302		0.313	0.107
RV (OtoC, UTX)	0.962				

Table 4: IBM and UTX equity data January to June 2005. Summary statistics for modulated realised volatility for  $n = 256$  and  $n = 1024$ . The same statistics are computed for the corresponding realised volatility ( $n = 20$ ) and the open-to-close realised volatility.)

Figure 9 displays the equivalent results for realised and modulated realised covariance and correlation for Lehman Brothers and Merrill Lynch (Figure 9), resp. IBM and UTX (Figure 10). Modulated realised covariance does not behave as nicely as modulated realised volatility which is due to the Epps effect. Our estimator relies on synchronous trading which is obviously rather unrealistic in practice. Epps (1979) documented that when estimating from empirical data, realised covariance converges to 0 as the sampling frequency increases. When sampling at a very high frequency, there are more and more zero-returns, when there is non-synchronous trading, this effect dominates both realised covariance and - to a smaller extent though - modulated realised covariance. Average results over the duration of 6 months can be seen in Tables 5 and 6. All high-frequency estimators exhibit a negative bias compared to open-to-close realised covariance. However, we observe that the IBM/UTX dataset exhibits a better performance

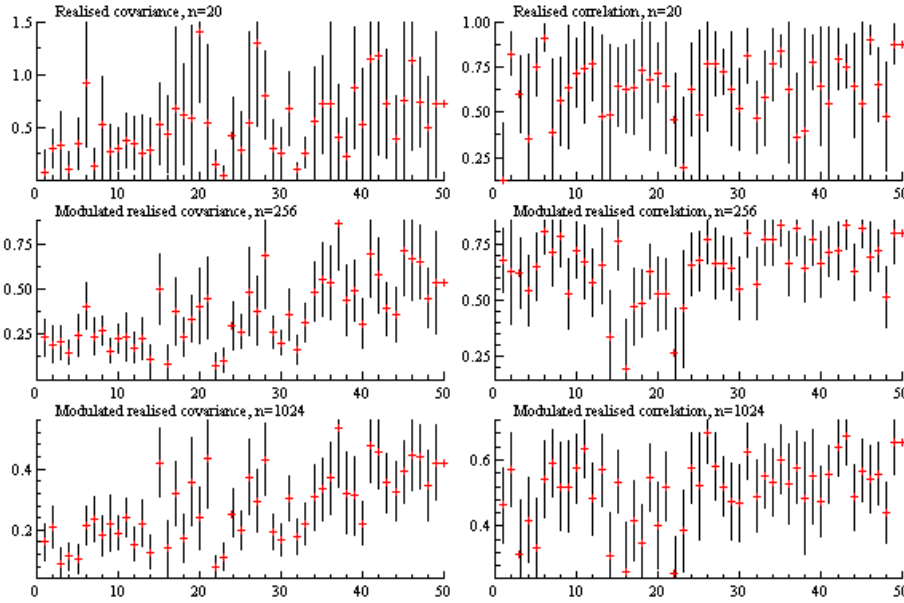


Figure 9: Dependence measures between Merrill Lynch and Lehman equity data. Data are 1st March 2005 onwards for 50 active trading days. (a) Realised covariance based on 20 intra-day observations. This is marked with a cross while the bars denote 95% confidence intervals. (b) Modulated realised covariance based on 256 intra-day observations. (c) Modulated realised covariance based on 1024 intra-day observations. (d) Modulated realised correlation based on 4096 intra-day observations. (e) Realised correlation based on 20 intra-day observations. (f) Modulated realised correlation based on 256 intra-day observations. (g) Modulated realised correlation based on 1024 intra-day observations. All the confidence intervals of modulated realised covariance are based on  $\tilde{H}_n$ .

when it comes to non-synchronous trading effects than the LEH/MER dataset. The average 6-month realised covariance of LEH/MER is 0.422, whereas modulated realised covariance estimates 0.399 ( $n = 256$ ), resp. 0.304 ( $n = 1024$ ). However, the estimates for the integrated covariance of the 6-month IBM/UTX time series appears rather stable (Table 6). Note that the estimator which was proposed by Hayashi and Yoshida (2005) exhibits a pronounced negative bias. This is probably due to the fact that it is by construction only capable of dealing with non-synchronicity, but not necessarily with market microstructure noise. Barndorff-Nielsen et al. (2008) have shown similar results. Note that the estimator for the standard deviation which is based on  $H_n$  generally estimates the standard error slightly bigger than  $\tilde{H}_n$ . Furthermore, the empirical standard error is always bigger than the estimated one.

Modulated realised correlation, as displayed in Figures 9 and 10 and Tables 7 and 8, obviously inherits the behaviour of the covariance to converge to 0 as the sampling frequency increases and both datasets show a pronounced negative bias of both realised and modulated

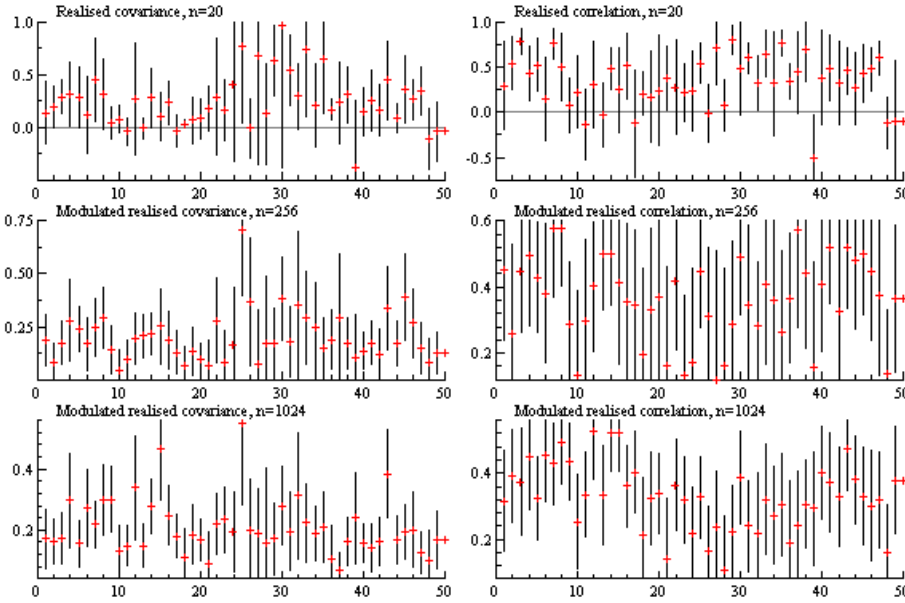


Figure 10: Dependence measures between IBM and UTX equity data. Data are 1st March 2005 onwards for 50 active trading days. (a) Realised covariance based on 20 intra-day observations. This is marked with a cross while the bars denote 95% confidence intervals. (b) Modulated realised covariance based on 256 intra-day observations. (c) Modulated realised covariance based on 1024 intra-day observations. (d) Realised correlation based on 20 intra-day observations. (e) Modulated realised correlation based on 256 intra-day observations. (f) Modulated realised correlation based on 1024 intra-day observations. All the confidence intervals of modulated realised covariance are based on  $\hat{H}_n$ .

realised correlation as compared to open-to-close correlation. In the LEH/MER dataset and for  $n = 256$  the effect of non-synchronous trading becomes particularly obvious with a negative bias of 0.219. But also for the IBM/UTX dataset modulated realised correlation has a surprisingly pronounced bias even though the results for both realised volatility and realised covariance in this dataset are rather good. The reason for that is that modulated realised volatility has a small positive bias whereas modulated realised covariance has a small negative bias which - as combined in modulated realised correlation - results in a rather big bias.

In Figure 11 we display modulated realised regression of MER on LEH equity data, resp. UTX on IBM equity data. For the LEH/MER dataset realised regression provides very similar results to modulated realised regression ( $n = 256$ ), but exhibits a significant downwards bias for  $n = 1024$ . The results for IBM/UTX data are more stable. Throughout the analysis the confidence intervals for modulated realised covariance based estimators are smaller than the related confidence intervals for the realised statistics.

estimator	average	standard deviation			bias
		$(H_n \text{ for MRC})$	$(\tilde{H}_n \text{ for MRC})$	empirical	
MRC (256, LEH/MER)	0.399	0.099	0.054	0.185	-0.054
MRC (1024, LEH/MER)	0.304	0.055	0.054	0.121	-0.149
RC (20min, LEH/MER)	0.422	0.472		0.701	-0.031
RC (OtoC, LEH/MER)	0.453				
Hayashi/Yoshida	0.108				0.064
					-0.349

Table 5: LEH and MER equity data January to June 2005. Summary statistics for modulated realised covariance for  $n = 256$  and  $n = 1024$ , realised covariance ( $n = 20$ ), open-to-close realised covariance and Hayashi-Yoshida estimator. The first column identifies the estimator, and the second gives the average value, followed by the estimated standard deviation. For modulated realised statistics the standard deviation is estimated using either  $H_n$  or  $\tilde{H}_n$ . The fourth column presents the empirical standard deviation. The last column presents the bias.

estimator	average	standard deviation			bias
		$(H_n \text{ for MRC})$	$(\tilde{H}_n \text{ for MRC})$	empirical	
MRC (256, IBM/UTX)	0.212	0.091	0.082	0.125	-0.065
MRC (1024, IBM/UTX)	0.229	0.062	0.059	0.089	-0.048
RC (20min, IBM/UTX)	0.252	0.222		0.228	-0.018
RC (OtoC, IBM/UTX)	0.277				
Hayashi/Yoshida	0.149				0.107
					-0.121

Table 6: IBM and UTX equity data January to June 2005. Summary statistics for modulated realised covariance for  $n = 256$  and  $n = 1024$ , realised covariance ( $n = 20$ ), open-to-close realised covariance and Hayashi-Yoshida estimator. The first column identifies the estimator, and the second gives the average value, followed by the estimated standard deviation. For modulated realised statistics the standard deviation is estimated using either  $H_n$  or  $\tilde{H}_n$ . The fourth column presents the empirical standard deviation. The last column presents the bias.

Let us finally understand why the results for the IBM/UTX dataset are substantially better than for the LEH/MER time series. Table 11 gives us the 6-month average of the number of data points of a daily time series. The first line exhibits the number of quotes available when considering all stock exchanges, the second line concentrates on NYSE only and the third line shows the number of data points we obtain if we clean the data in the way described at the beginning of the section. We see that the IBM/UTX dataset is significantly larger than the LEH/MER dataset. Thus the chance of constructing a truly synchronous bivariate time series of non-zero returns is much higher for the larger dataset which leads to better estimation results.

## 8 Conclusions

In this paper we have introduced the modulated realised covariation estimator as an estimator of ex-post covariation of high-frequency asset prices under market microstructure noise. The novelty of this paper is as follows: we suggest a very simple way of estimating the covariation of

estimator	average	standard deviation			bias
		$(H_n \text{ for MRC})$	$(\tilde{H}_n \text{ for MRC})$	empirical	
MRCorr (256, LEH/MER)	0.755	0.081	0.096	0.138	-0.029
MRCorr (1024, LEH/MER)	0.585	0.057	0.053	0.101	-0.140
RCorr (20min, LEH/MER)	0.824	0.252		0.448	0.098
RCorr (OtoC, LEH/MER)	0.726				

Table 7: IBM and UTX equity data January to June 2005. Summary statistics for modulated realised correlation for  $n = 256$  and  $n = 1024$ , realised correlation ( $n = 20$ ) and open-to-close realised correlation. The first column identifies the estimator, and the second gives the average value, followed by the estimated standard deviation. For modulated realised statistics the standard deviation is estimated using either  $H_n$  or  $\tilde{H}_n$ . The fourth column presents the empirical standard deviation. The last column presents the bias.

estimator	average	standard deviation			bias
		$(H_n \text{ for MRC})$	$(\tilde{H}_n \text{ for MRC})$	empirical	
MRCorr (256, IBM/UTX)	0.362	0.125	0.117	0.183	-0.08
MRCorr (1024, IBM/UTX)	0.329	0.08	0.07	0.143	-0.113
RCorr (20min, IBM/UTX)	0.339	0.284		0.435	-0.103
RCorr (OtoC, IBM/UTX)	0.442				

Table 8: LEH and MER equity data January to June 2005. Summary statistics for modulated realised correlation for  $n = 256$  and  $n = 1024$ , realised correlation ( $n = 20$ ) and open-to-close realised correlation. The first column identifies the estimator, and the second gives the average value, followed by the estimated standard deviation. For modulated realised statistics the standard deviation is estimated using either  $H_n$  or  $\tilde{H}_n$ . The fourth column presents the empirical standard deviation. The last column presents the bias.

price processes and provide a feasible asymptotic limit theory for modulated realised covariance, correlation and regression. Moreover, we construct a positive semidefinite estimator of the conditional covariance matrix of the limiting variable, some positive semidefinite estimators of the covolatility and indicate how the assumptions on the noise process can be relaxed. Finally, we present a way of arranging the data such that the estimator is capable of dealing with non-synchronous trading and we obtain very accurate empirical results for dependence measures between assets. Thus we have tackled the most important issues of the multivariate problem.

It remains an open problem to extend the theory to a setting where non-synchronous trading cannot only be dealt with by a reasonable way of cleaning the data, but where the estimator itself is capable of dealing with it. One approach would be to combine the ideas of modulated realised covariation with the estimator of Hayashi and Yoshida (2003).

## 9 Appendix: Proof

In the following we assume without loss of generality that the processes  $a$  and  $\sigma$  are bounded. This can be justified by a standard localization procedure (see e.g. Barndorff-Nielsen et al.

estimator	average	standard deviation			bias
		( $H_n$ for MRC)	( $\tilde{H}_n$ for MRC)	empirical	
MRRegr (256, LEH/MER)	1.350	0.212	0.184	0.281	0.08
MRRegr (1024, LEH/MER)	1.032	0.136	0.131	0.243	-0.238
RRegr (20min, LEH/MER)	1.379	0.561		1.22	0.109
RRegr (OtoC, LEH/MER)	1.270				

Table 9: LEH and MER equity data January to June 2005. Summary statistics for modulated realised regression for  $n = 256$  and  $n = 1024$ , realised regression ( $n = 20$ ) and open-to-close realised regression. The first column identifies the estimator, and the second gives the average value, followed by the estimated standard deviation. For modulated realised statistics the standard deviation is estimated using either  $H_n$  or  $\tilde{H}_n$ . The fourth column presents the empirical standard deviation. The last column presents the bias.

estimator	average	standard deviation			bias
		( $H_n$ for MRC)	( $\tilde{H}_n$ for MRC)	empirical	
MRRegr (256, IBM/UTX)	0.249	0.125	0.095	0.272	-0.038
MRRegr (1024, IBM/UTX)	0.210	0.077	0.053	0.221	-0.078
RRegr (20min, IBM/UTX)	0.2469	0.153		0.378	-0.041
RRegr (OtoC, IBM/UTX)	0.288				

Table 10: IBM and UTX equity data January to June 2005. Summary statistics for modulated realised regression for  $n = 256$  and  $n = 1024$ , realised regression ( $n = 20$ ) and open-to-close realised regression. The first column identifies the estimator, and the second gives the average value, followed by the estimated standard deviation. For modulated realised statistics the standard deviation is estimated using either  $H_n$  or  $\tilde{H}_n$ . The fourth column presents the empirical standard deviation. The last column presents the bias.

(2006)). Moreover, we denote all constants by  $C$  or  $C_p$  if they depend on an additional parameter  $p$ . The main parts of the proofs are based upon the methods presented in Podolskij and Vetter (2006) and Jacod et al. (2007).

*Proof of Theorem 1:* Due to the triangular equality  $[X, Y] = \frac{1}{4}([X + Y, X + Y] - [X - Y, X - Y])$  it suffices to consider the case  $d = 1$  (i.e. all processes are 1-dimensional). Next, we use the decomposition

$$MRC(Z^*)_n = \frac{1}{k_n} \sum_{l=0}^{k_n-1} MRC(Z^*)_n^l - \frac{\psi_1}{2\theta^2\psi_2 n} \sum_{i=1}^n |\Delta_i^n Z^*|^2, \quad (9.1)$$

with

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

Notice that, for any  $l = 0, \dots, k_n - 1$ , the summands in the definition of  $MRC(Z^*)_n^l$  are asymptotically uncorrelated. This type of estimators have been discussed in Podolskij and Vetter (2006) and we can deduce by the methods presented therein (see the proof of Theorem

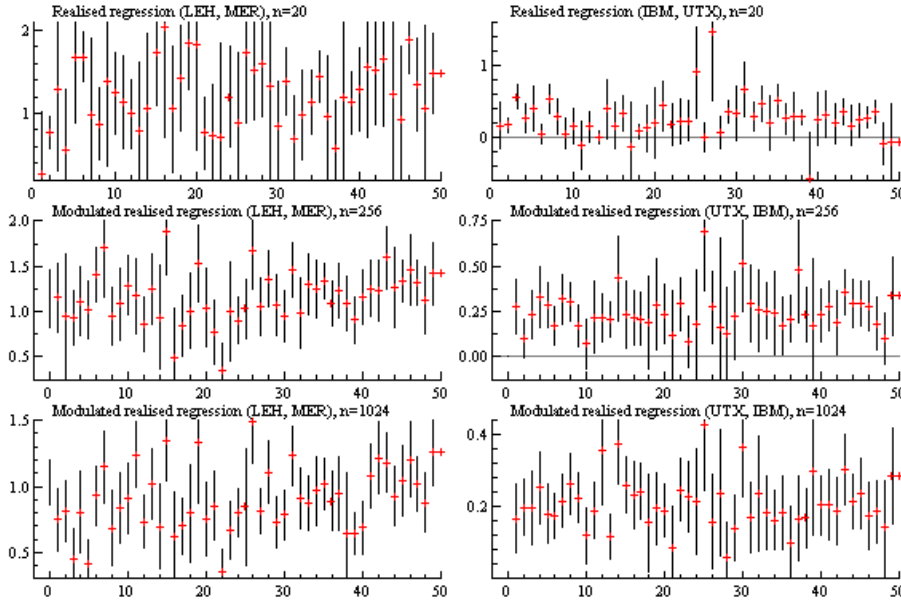


Figure 11: Regression of Merrill Lynch on Lehman equity data, resp. UTX on IBM equity data. Data are 1st March 2005 onwards for 50 active trading days. (a) Realised regression based on 20 intra-day observations (MER LEH). This is marked with a cross while the bars denote 95% confidence intervals. (b) Modulated realised regression based on 256 intra-day observations (MER LEH). (c) Modulated realised regression based on 1024 intra-day observations (MER LEH). (d) Realised regression based on 20 intra-day observations (UTX IBM). (e) Modulated realised regression based on 256 intra-day observations (UTX IBM). (f) Modulated realised regression based on 1024 intra-day observations (UTX IBM). All the confidence intervals of modulated realised covariation are based on  $\tilde{H}_n$ .

1) that

$$MRC(Z^*)_n^l \xrightarrow{P} \int_0^1 \sigma_s^2 ds + \frac{\psi_1}{\psi_2 \theta^2} \Psi ,$$

where the convergence holds uniformly in  $l$  (due to the boundedness of the processes  $a$  and  $\sigma$ ).

On the other hand we have that

$$\hat{\Psi} = \frac{1}{2n} \sum_{i=1}^n |\Delta_i^n Z^*|^2 \xrightarrow{P} \Psi .$$

This implies the convergence

$$MRC(Z^*)_n \xrightarrow{P} \int_0^1 \sigma_s^2 ds ,$$

which completes the proof.  $\square$

*Proof of Theorem 2:* Here we apply the "big blocks & small blocks"-technique used in Ja-

Firm	LEH	MER	IBM	UTX
all data	36322	45461	51533	39781
NYSE data	12497	14983	17036	17036
cleaned data	2625	2623	3376	3187

Table 11: Average number of data points in daily time series over a 6-month period (January 1, 2005 to June 30, 2005) for various firms. First line: all data available from WRDS; second line: NYSE data available from WRDS; third line: data is cleaned in such a way that if two consecutive prices have the same value, the second price is not included in the cleaned dataset.

cod et al. (2007). 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

$$j_n(p) = \left\lfloor \frac{n}{k_n(p+1)} \right\rfloor - 1. \tag{9.2}$$

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

Next, we introduce the random variable

$$\bar{Z}_{i,m}^{*n} = \sum_{j=1}^{k_n-1} g\left(\frac{j}{k_n}\right) (\sigma_{\frac{m}{n}} \Delta_{i+j}^n W + \Delta_{i+j}^n \epsilon) , \tag{9.3}$$

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

$$\Upsilon_{j,m}^n = \bar{Z}_{j,m}^{*n} (\bar{Z}_{j,m}^{*n})' - E[\bar{Z}_{j,m}^{*n} (\bar{Z}_{j,m}^{*n})' | \mathcal{F}_{\frac{m}{n}}] , \tag{9.4}$$

and define

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



as well as

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

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

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

and note that

$$E[\zeta(p, 1)_j^n | \mathcal{F}_{\frac{a_j(p)}{n}}] = 0 = E[\zeta(p, 2)_j^n | \mathcal{F}_{\frac{b_j(p)}{n}}] \quad (9.5)$$

by construction.

Now, by the same approximations as presented in Jacod et al. (2007) (see the identity (5.14), Lemma 5.5 and Lemma 5.6 therein) we obtain that

$$n^{1/4} \left( MRC(Z^*)_n - \int_0^1 \sigma_s \sigma'_s ds \right) = \frac{n^{1/4}}{\theta\psi_2} (M(p)^n + N(p)^n + C(p)^n) + R(p)_n \quad (9.6)$$

where the last three summands satisfy the convergence

$$\lim_{p \rightarrow \infty} \limsup_{n \rightarrow \infty} P(\|n^{1/4} N(p)^n\| + \|n^{1/4} C(p)^n\| + \|R(p)_n\| > \delta) = 0 \quad (9.7)$$

for any  $\delta > 0$ . Notice that the term  $R(p)_n$  stands for the error made in the approximation (9.3).

In the next lemma we show the stable convergence  $\frac{n^{1/4}}{\theta\psi_2} M(p)^n \xrightarrow{\mathcal{D}_{st}} U(p)$  (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 (9.6) and (9.7) we obtain the assertion of Theorem 2.

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

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

and

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

where the processes  $\Lambda_s$ ,  $\Theta_s$  and  $\Upsilon$  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  $\gamma_s$  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 have to be shown (for all  $1 \leq k, k', l, l' \leq d$ )

$$\frac{n^{-\frac{1}{2}}}{\theta^2 \psi_2^2} \sum_{j=0}^{j_n(p)} E[\zeta(p, 1)_j^{n,kl} \zeta(p, 1)_j^{n,k'l'} | \mathcal{F}_{\frac{a_j(p)}{n}}] \xrightarrow{P} \int_0^1 A_u^{kl,k'l'} du \quad (9.8)$$

$$n^{-1} \sum_{j=0}^{j_n(p)} E[|\zeta(p, 1)_j^n|^4 | \mathcal{F}_{\frac{a_j(p)}{n}}] \xrightarrow{P} 0 \quad (9.9)$$

$$n^{-\frac{1}{4}} \sum_{j=0}^{j_n(p)} E[\zeta(p, 1)_j^{n,kl} \Delta W(p)_j^{n,k'} | \mathcal{F}_{\frac{a_j(p)}{n}}] \xrightarrow{P} 0 \quad (9.10)$$

$$n^{-\frac{1}{4}} \sum_{j=0}^{j_n(p)} E[\zeta(p, 1)_j^{n,kl} \Delta N(p)_j^n | \mathcal{F}_{\frac{a_j(p)}{n}}] \xrightarrow{P} 0 \quad (9.11)$$

where  $\Delta V(p)_j^n = V_{n/b_j(p)} - V_{n/a_j(p)}$  for any process  $V$  and (9.11) holding for any 1-dimensional bounded martingale  $N$  being orthogonal to  $W$ . For proving (9.9) and (9.11) it is no restriction to assume that  $d = 1$ . Then these conditions are already shown in Jacod et al. (2007) (Lemma 5.7). On the other hand, the functional  $\zeta(p, 1)_j^n$  is even in  $W$ . Since  $W$  and  $\epsilon$  are independent, we readily deduce that

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

which implies condition (9.10). Hence, we are left to proving (9.8).

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  $\bar{\epsilon}_i^n$ . By the smoothness assumption on the function  $g$  and the above identity we obtain the approximations ( $1 \leq k, l \leq d$ )

$$E[\bar{W}_j^{n,k} \bar{W}_{j'}^{n,l}] = \delta_{kl} \frac{k_n}{n} \psi_2\left(\frac{|j-j'|}{k_n}\right) + O(n^{-1}), \quad E[\bar{\epsilon}_j^{n,k} \bar{\epsilon}_{j'}^{n,l}] = \frac{\Psi^{kl}}{k_n} \psi_1\left(\frac{|j-j'|}{k_n}\right) + O(n^{-1}) \quad (9.12)$$

for  $|j-j'| < k_n$ , whereas the above expectations vanish when  $|j-j'| \geq k_n$  (here  $\delta_{kl}$  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

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

By a straightforward calculation (and (9.12)) we obtain for all  $1 \leq k, l, k'l' \leq d$

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

where the approximation holds uniformly in  $j$ . Now recall that  $j_n(p) = \left\lfloor \frac{n}{k_n(p+1)} \right\rfloor - 1$ . Conse-

quently, by Riemann integrability we deduce that

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

which completes the proof of Lemma 1. □

*Proof of Proposition 6:* Due to the triangular equality  $[X, Y] = \frac{1}{4}([X+Y, X+Y] - [X-Y, X-Y])$  it suffices to consider the case  $d = 1$  (i.e.  $X_i^n = |\bar{Z}_i^{*n}|^2$ ). Now we replace the quantity  $\bar{Z}_i^{*n}$  in the definition of the estimator  $H_n$  by its approximation  $\bar{Z}_{i,i}^{*n}$  given in (9.3). As in Podolskij and Vetter (2006) we deduce that

$$\begin{aligned} H_n &= \frac{1}{\sqrt{n\theta^2\psi_2^2}} \sum_{i=0}^{n-2k_n+1} |\bar{Z}_{i,i}^{*n}|^2 \left( |\bar{Z}_{i,i}^{*n}|^2 - |\bar{Z}_{i+k_n,i}^{*n}|^2 \right) + 2 \sum_{j=1}^{k_n-1} |\bar{Z}_{i,i}^{*n}|^2 \left( |\bar{Z}_{i+j,i}^{*n}|^2 - |\bar{Z}_{i+k_n,i}^{*n}|^2 \right) + o_p(1) \\ &= \sum_{i=0}^{n-2k_n+1} \xi_i^n + o_p(1). \end{aligned}$$

A straightforward calculation using (9.12) shows that

$$\sum_{i=0}^{n-2k_n+1} E[\xi_i^n | \mathcal{F}_i^n] \xrightarrow{P} \frac{1}{\psi_2^2} \left( 2\Phi_{22}\theta \int_0^1 \sigma_s^4 ds + \frac{4\Phi_{12}\Psi}{\theta} \int_0^1 \sigma_s^2 ds + \frac{2\Phi_{11}\Psi^2}{\theta^3} \right).$$

On the other hand we have that

$$E[\xi_i^n \xi_j^n] \leq Cn^{-1} , \quad |i - j| \leq 2k_n ,$$

whereas the above expression is 0 for  $|i - j| > 2k_n$ . This implies the convergence

$$\sum_{i=0}^{n-2k_n+1} \xi_i^n \xrightarrow{P} \frac{1}{\psi_2^2} \left( 2\Phi_{22}\theta \int_0^1 \sigma_s^4 ds + \frac{4\Phi_{12}\Psi}{\theta} \int_0^1 \sigma_s^2 ds + \frac{2\Phi_{11}\Psi^2}{\theta^3} \right) ,$$

which completes the proof of Proposition 6. □

## References

- [1] Andersen, T.G., Bollerslev, T., Diebold, F.X. and Ebens, H. (2001). The distribution of realized stock return volatility. *Journal of Financial Economics* 61, 63-76.

- [2] Barndorff-Nielsen, O.E., Graversen, S.E., Jacod, J., Podolskij, M., Shephard, N. (2008). *A central limit theorem for realised power and bipower variations of continuous semi-martingales.* "From Stochastic Analysis to Mathematical Finance, Festschrift for Albert Shiryaev", Springer.
- [3] Barndorff-Nielsen, O.E., Hansen, P.R., Lunde, A., Shephard, N. (2006). Designing realised kernels to measure the ex-post variation of equity prices in the presence of noise. *Econometrica*, Forthcoming.
- [4] Barndorff-Nielsen, O.E., Shephard, N. (2002). Econometric analysis of realised volatility and its use in estimating stochastic volatility models. *J.Roy. Statist. Soc. Ser. B* 64, 253-280.
- [5] Barndorff-Nielsen, O.E., Shephard, N. (2004). Econometric analysis of realised covariation: high frequency covariance, regression and correlation in financial economics. *Econometrica* 72, 885-925.
- [6] Barndorff-Nielsen, O.E., Hansen, P., Lunde, A. and Shephard, N. (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 Oxford.*
- [7] Brandt, M.W. and Diebold, F.X. (2006). A no-arbitrage approach to range-based estimation of return covariances and correlation. *J. of Business* 79, 61-74.
- [8] Dacorogna, M., Lundin, M., Mueller, U.A. (1999). Correlation of High-Frequency Financial Time Series. *Financial Markets Tick by Tick*, Wiley & Son.
- [9] Epps, T.W. (1979). Comovements in stock prices in the very short run. *Journal of the American Statistical Society* 74, 291-296.
- [10] Gloter, A., Jacod, J. Diffusion with measurement errors II. Optimal estimators. *ESAIM Probability and Statistics* 5, 243-260.
- [11] Hansen, P., Lunde, A. (2006). Realized variance and market microstructure noise (with discussion). *Journal of Business and Economic Statistics* 24, 127-218.

- [12] Hayashi, T., Yoshida, N. (2005). On covariance estimation of non-synchronously observed diffusion processes. *Bernoulli* 11, 359-379.
- [13] Jacod, J., Li, Y. Mykland, P.A., Podolskij, M. and Vetter, M. (2007). Microstructure noise in the continuous case: the pre-averaging approach. *Working paper. Department of Statistics, University of Chicago.*
- [14] Jacod, J., Shiryaev, A.N. (2003). *Limit theorems for stochastic processes.* Springer, Berlin.
- [15] Podolskij, M., Vetter, M. (2006). Estimation of volatility functionals in the simultaneous presence of microstructure noise and jumps. *Working paper.*
- [16] Vetter, M. (2008). Estimation methods in noisy diffusion models (2008). *PhD thesis, Ruhr-University of Bochum.*
- [17] Zhang, L. (2006). Efficient estimation of stochastic volatility using noisy observations: a multiscale approach. *Bernoulli* 12(6), 1019-1043.
- [18] Zhang, L., Mykland, P.A., Ait-Sahalia, Y. (2005). A tale of two time scales: Determining integrated volatility with noisy high-frequency data. *J. Amer. Statist. Assoc.* 472, 1394-1411.

## **10 Appendix: Figures and Tables**

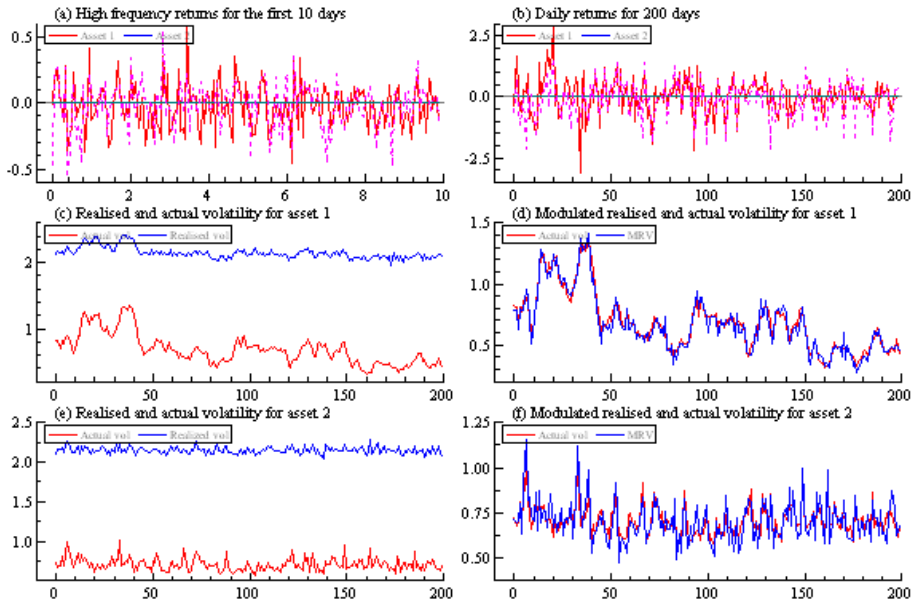


Figure 12: Simulated bivariate stochastic volatility model using  $n = 4096$ : (a) 30 minute returns  $\Delta_i^n X$ ,  $\Delta_i^n Y$  for 10 trading days; (b) daily returns  $\sum_{i=1}^n \Delta_i^n X$  and  $\sum_{i=1}^n \Delta_i^n Y$  for 200 trading days; (c) realized volatility and actual volatility for asset 1; (d)  $\sqrt{MRC(Z^*)_n^{1,1}}$  and actual volatility for asset 1; (e) realized volatility and actual volatility for asset 1; (f)  $\sqrt{MRC(Z^*)_n^{2,2}}$  and actual volatility for asset 2;

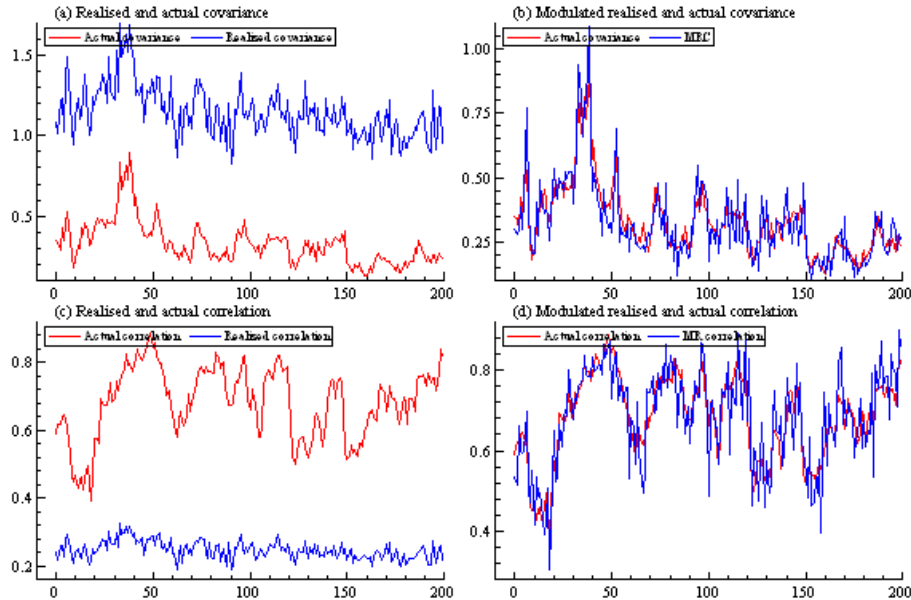


Figure 13: Simulation of measures of dependence, using  $n = 4096$ : (a) realized covariance and actual covariance; (b)  $MRC(Z^*)_n^{1,2}$  and actual covariance; (c) realized and actual correlation; (d) MRC based and actual correlation.

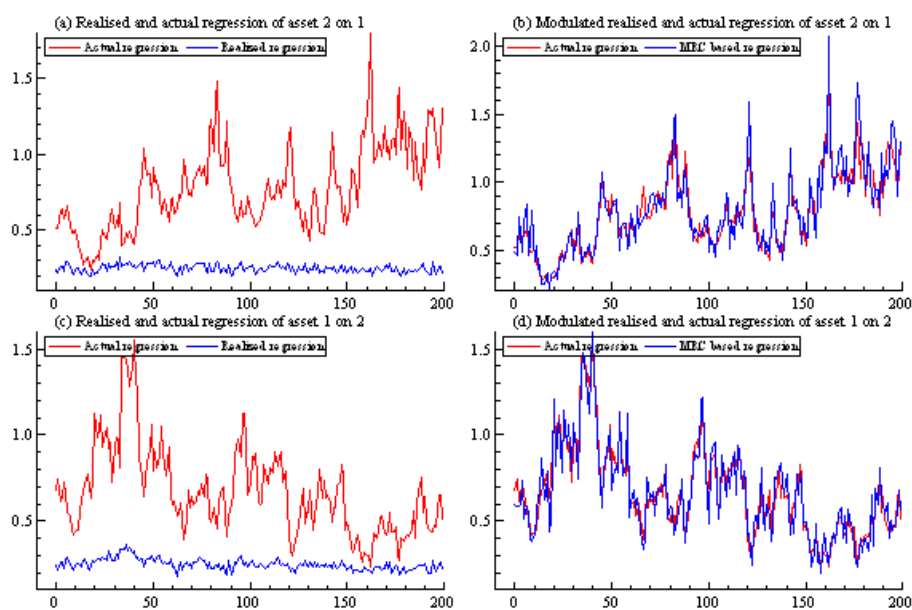


Figure 14: Simulations of measures of dependence, using  $n = 4096$ ; (a) realized and actual regressions of returns on asset 2 on asset 1; (b) MRC based and actual regressions of returns on asset 2 on asset 1; (c) same as (a) but asset 1 on 2; (d) same as (b) but asset 1 on 2;



test statistic	n	coverage						mean	variance
		0.5%	2.5%)	5%	95%	97.5%)	99.5%		
(3.7)	256	0.58	5.09	11.82	99.76	99.93	99.99	-0.33	0.6
	1024	0.39	3.14	6.91	98.15	99.13	99.88	-0.17	0.83
	4096	0.40	2.70	5.90	96.60	98.25	99.80	-0.08	0.92
	16384	0.57	2.63	5.34	95.77	97.50	99.53	-0.03	0.98
	65536	0.55	2.59	5.24	95.51	97.51	99.50	-0.01	1.02
(4.3)	256	9.68	19.75	27.13	99.90	99.98	99.99	NV	NV
	1024	3.48	8.89	13.47	98.75	99.50	99.96	-0.47	1.15
	4096	1.70	5.50	9.55	97.40	99.20	99.90	-0.19	1.06
	16384	1.27	3.63	7.34	96.63	98.40	99.83	-0.10	1.03
	65536	1.03	3.28	6.88	96.48	97.99	99.80	-0.07	1.02
(4.6)	256	16.21	32.12	42.10	99.95	99.98	100.00	-1.48	1.25
	1024	3.19	9.27	14.74	99.05	99.61	99.97	-0.57	1.05
	4096	1.94	5.85	9.91	97.64	98.98	99.98	-0.29	1.08
	16384	1.01	4.10	7.32	96.50	98.52	99.79	-0.14	1.03
	65536	0.93	3.85	6.83	96.39	98.32	99.74	-0.13	1.01
(5.4)	256	0.51	2.19	4.20	95.90	97.82	99.47	0.02	0.84
	1024	0.45	2.16	4.59	97.30	98.76	99.78	0.03	0.84
	4096	0.40	2.45	3.95	96.05	98.35	99.70	-0.01	0.89
	16384	0.40	2.33	4.70	96.67	98.53	99.87	-0.06	0.88
	65536	0.49	2.53	4.78	96.01	98.14	99.60	-0.02	0.95
(5.5)	256	1.09	3.35	5.99	95.31	97.25	99.11	NV	NV
	1024	0.65	3.07	6.08	94.95	97.05	99.03	NV	NV
	4096	0.80	3.40	5.25	94.45	97.10	99.50	-0.01	1.07
	16384	0.77	3.20	6.17	95.73	98.07	99.63	-0.07	1.03
	65536	0.67	32.75	6.04	95.47	97.84	99.65	-0.04	1.02
(5.6)	256	0.28	1.56	3.37	95.65	98.95	100	0.08	0.93
	1024	0.36	1.84	3.86	99.1	98.63	99.78	0.03	0.84
	4096	0.41	2.25	4.45	96.4	97.93	99.64	0.03	0.94
	16384	0.44	2.31	4.45	95.52	97.92	99.61	-0.003	0.92
	65536	0.47	2.44	4.62	95.37	97.65	99.58	-0.001	0.96
(5.10)	256	1.30	3.71	6.01	97.62	99.00	99.66	-0.19	1.22
	1024	0.95	3.45	6.10	98.08	99.13	99.71	-0.11	1.08
	4096	0.80	2.60	4.60	96.70	98.70	99.60	-0.09	1.10
	16384	0.80	3.20	5.67	96.90	98.57	99.57	-0.06	1.04
	65536	0.57	2.59	5.29	95.44	98.00	99.54	-0.01	1.01
(5.11)	256	0.62	2.17	4.23	93.47	95.91	98.31	-0.19	1.22
	1024	0.65	3.07	6.08	94.95	97.05	99.03	-0.11	1.08
	4096	0.35	1.95	4.25	94.10	96.50	99.25	0.03	1.05
	16384	0.50	2.90	5.77	94.80	97.13	99.37	-0.07	1.02
	65536	0.49	2.86	5.78	95.12	98.3	99.59	-0.02	1.01
(5.12)	256	0.04	0.60	1.67	92.27	95.38	99.71	0.23	0.93
	1024	0.10	0.92	2.54	95.13	97.56	99.34	0.07	0.84
	4096	0.17	1.29	3.18	94.17	96.99	99.38	0.09	0.94
	16384	0.24	1.66	3.67	95.08	98.53	99.72	0.03	0.92
	65536	0.32	1.97	3.94	95.02	98.01	99.56	0.02	0.97

Table 12: We simulate 20000 replications from the stochastic volatility model described in section 5. We compute the coverage properties, mean and standard deviation of 6 test statistics.

# Research Papers 2008



- 2008-10 Peter Christoffersen, Kris Jacobs, Christian Dorion and Yintian Wang: Volatility Components, Affine Restrictions and Non-Normal Innovations
- 2008-11 Peter Christoffersen, Kris Jacobs, Chayawat Ornthanalai and Yintian Wang: Option Valuation with Long-run and Short-run Volatility Components
- 2008-12 Tom Engsted and Stig V. Møller: An iterated GMM procedure for estimating the Campbell-Cochrane habit formation model, with an application to Danish stock and bond returns
- 2008-13 Lars Stentoft: Option Pricing using Realized Volatility
- 2008-14 Jie Zhu: Pricing Volatility of Stock Returns with Volatile and Persistent Components
- 2008-15 Jie Zhu: Testing for Expected Return and Market Price of Risk in Chinese A-B Share Market: A Geometric Brownian Motion and Multivariate GARCH Model Approach
- 2008-16: Jie Zhu: FIEGARCH-M and International Crises: A Cross-Country Analysis
- 2008-17: Almut E. D. Veraart: Inference for the jump part of quadratic variation of Itô semimartingales
- 2008-18: Michael Sørensen: Parametric inference for discretely sampled stochastic differential equations
- 2008-19: Anne Péguin-Feissolle, Birgit Strikholm and Timo Teräsvirta: Testing the Granger noncausality hypothesis in stationary nonlinear models of unknown functional form
- 2008-20: Stefan Holst Bache, Christian M. Dahl and Johannes Tang Kristensen: Determinants of Birthweight Outcomes: Quantile Regressions Based on Panel Data
- 2008-21: Ole E. Barndorff-Nielsen, José Manuel Corcuera, Mark Podolskij and Jeannette H.C. Woerner: Bipower variation for Gaussian processes with stationary increments
- 2008-22: Mark Podolskij and Daniel Ziggel: A Range-Based Test for the Parametric Form of the Volatility in Diffusion Models
- 2008-23: Silja Kinnebrock and Mark Podolskij: An Econometric Analysis of Modulated Realised Covariance, Regression and Correlation in Noisy Diffusion Models