Economic Modelling 43 (2014) 416–425

Contents lists available at ScienceDirect

Economic Modelling

journal homepage: www.elsevier.com/locate/ecmod

Covariance estimation using high-frequency data: Sensitivities of estimation methods ☆

Erik Haugom [a,b], Gudbrand Lien [a,*], Steinar Veka [a,b], Sjur Westgaard [b]

[a] Faculty of Economics and Organization Science, Lillehammer University College, Norway
[b] Department of Industrial Economics and Technology Management, Norwegian University of Science and Technology, Norway

A R T I C L E I N F O

Article history:
Accepted 14 August 2014
Available online xxxx

JEL classification:
G1
G13
C13
C58

Keywords:
Realized correlation
High-frequency data
ICE crude oil
Gasoil
Gas futures

A B S T R A C T

In this study we examine three widely used realized correlation estimators for natural gas, gasoil, and crude oil futures using data from IntercontinentalExchange (ICE). The objective is to illustrate sensitivities of estimation methods on the resulting realized correlation estimates. The empirical results show that the choice between the various correlation estimators is not at all trivial and depends strongly on the specific features and liquidity of the observed price processes. These findings suggest that great care must be taken when using high-frequency data in portfolio risk applications.

1. Introduction

Portfolio optimization performance and all kinds of risk assessment and risk management are determined by how accurately the dependency or correlation structure between asset returns is measured. Over the last decade, an increasing interest in correlation estimates obtained from high-frequency intraday data has been observed within the financial econometrics literature. Previous studies have compared the properties of these nonparametric variance and correlation measures with more commonly used parametric variance and correlation measures such as ARCH, GARCH, and RiskMetrics.

Wang et al. (2008) is the first study to examine thoroughly realized volatilities and correlations in energy markets, by analysing NYMEX crude oil and Henry-Hub natural gas futures contracts for the period 1995 to 1999. In that study they used equally-spaced five-minute returns as input in the standard realized covariance estimator, as introduced by Andersen et al. (2001). This technique assumes that the assets under consideration are observed simultaneously and without contamination (such as microstructure noise).

Unfortunately, when the sampling frequency exceeds the frequency at which a given asset has actually been traded, more zero-returns will occur. Increasing the sampling frequency will therefore decrease the probability that a simultaneous price change in the true price process for two assets (caused by the same event) will be materialized in an observable price change (through actual trading) in the same sampling interval. This will result in decreasing correlation with increasing sampling frequency, in addition to an imposed correlation between successive sampling intervals. This leads to a trade-off between the formerly mentioned effect known as the Epps effect (Epps, 1979) on the one hand and inefficient utilization of data on the other hand.

To deal with this zero returns/non-synchronous trading problem, several realized covariance estimators have been proposed in the literature. Some recent work on this topic includes: the realized covariance estimator extended with lead–lag adjustment (Zhang et al., 2005); the multivariate realized kernel covariance estimator, which is an extension of the lead–lag estimator (Barndorff-Nielsen et al., 2008, 2011); the functional regression approach to estimate and test the dependence structure between the entire distributions of dependent and independent (high-frequency) variables (Park and Qian, 2012); and the Hayashi–Yoshida

☆ This paper is a part of the project "ELCARBONRISK — Modelling and forecasting risk in electricity, carbon and related energy markets (oil, gas, coal)," financed by the Research Council of Norway, Eidsiva Energi and Tafjord Kraftproduksjon AS (Project number: 199904).

* Corresponding author at: Faculty of Economics and Organization Science, Lillehammer University College, NO-2604 Lillehammer, Norway. Tel.: +47 9248 8335.
E-mail address: gudbrand.lien@hil.no (G. Lien).

http://dx.doi.org/10.1016/j.econmod.2014.08.016
0264-9993/

covariance estimator, which computes directly from asynchronously observed prices (Hayashi and Yoshida, 2005).[1] However, it seems that it is not well understood how estimators of covariance/correlation behave when the estimation uses high-frequency tick-by-tick data (e.g., Voev and Lunde, 2007; de Pooter et al., 2008; Zhang, 2011).

In this study we empirically examine the realized correlation between three important energy commodities, using three alternative realized correlation estimators to illustrate to what extent the results from previous studies are estimator dependent. The estimators are the previous-tick estimator, the refresh-time estimator and the estimator by Hayashi and Yoshida.

The selected energy commodities are crude oil, gasoil, and natural gas traded at IntercontinentalExchange (ICE). These commodities are chosen for several reasons. First, they represent three different liquidity levels. We may therefore illustrate how liquidity affects the correlation estimates. Second, they are all fossil energy sources that to some extent are interchangeable in energy production. Gasoil is even derived from crude oil through the refining process. Natural gas on the other hand is also one of the key energy sources for the industrial sector and is used heavily in electricity generation. This creates the expectation that commodities will be fundamentally linked and there should be correlation that can be measured. Finally, the assets are all important financial assets, such that even if our results are generalizable and transferable to other asset classes, analysis of these commodities is interesting in itself.[2]

The importance of correlation timing is illustrated in, e.g., Della Corte et al. (2012). However, the estimation of this measure is not straightforward, and in this paper we illustrate several challenges in that respect. First, we illustrate the sensitivity of correlation estimates to sampling frequency using the standard realized correlation estimator. Second, we illustrate the sensitivity of correlation estimates to two highly-liquid assets (crude oil and gasoil), and to one highly-liquid asset (crude oil) and one low-liquid asset (natural gas). Third, we present correlation estimates using refresh-time sampling (instead of equally spaced previous-tick sampling) and using the Hayashi–Yoshida covariance estimator, and compare them with the estimators obtained from the traditional realized correlation measure. We also discuss the pros, cons and pitfalls of these estimation techniques.

The paper is organized as follows. In Section 2 we briefly introduce the concept of realized covariance and correlation, then the various realized covariance and correlation estimators are described. This is followed by data description in Section 3 and presentation of the results in Section 4, and finally in Section 5 conclusions are drawn and implications discussed.

2. Alternative estimation methods

Assume that the *true* price process, realizations of which are the (log of the) trading prices that we occasionally observe, follows a diffusion process of the following kind:

$$dp(t) = \mu(t)dt + \Omega(t)dw(t) \tag{1}$$

where the drift, $\mu(t)$, is an N-dimensional vector process, the instantaneous volatility, $\Omega(t)$, is an $(N \times N)$ matrix such that $\sum(t) = \Omega(t)\Omega'(t)$ is the covariance matrix process of the continuous sample path component and $w(t)$ is a vector of N independent Brownian motions.

In the following we will consider two assets, denoted as asset a, and asset b, such that $z \to 0$, so (1) could be written as:

$$\begin{bmatrix} p_t^{(a)} - p_{t-z}^{(a)} \\ p_t^{(b)} - p_{t-z}^{(b)} \end{bmatrix} = \begin{bmatrix} \mu_t^{(a)} \\ \mu_t^{(b)} \end{bmatrix} \begin{bmatrix} t-z \\ t-z \end{bmatrix} + \begin{bmatrix} \Omega_t^{(aa)} & \Omega_t^{(ab)} \\ \Omega_t^{(ba)} & \Omega_t^{(bb)} \end{bmatrix} \begin{bmatrix} w_t^{(a)} - w_{t-z}^{(a)} \\ w_t^{(b)} - w_{t-z}^{(b)} \end{bmatrix}.$$

Assuming that the returns do not allow for arbitrage and have a finite instantaneous mean, the multivariate log-price process in (1) belongs to the class of semi-martingales. Because the price process in (1) is a semi-martingale it has a well-defined quadratic variance/covariance process.[3] Then, the quadratic covariance, *QCov* in (2) below is the theoretical covariance of the price process in (1).

$$QCov = \int \sum(t)dt \tag{2}$$

Given access to tick-by-tick high-frequency data, a number of possibilities for approximating the quadratic covariance have been proposed. In the following we will present the three selected alternatives.

Allowing for a slight misuse of notation we denote $r_{t,i}$ as the i-th intraday return of day t, defined as an $(N \times 1)$ vector. In our two dimensional example, $r_{t,i} = \{r_{t,i}^{(a)}, r_{t,i}^{(b)}\}$ and $i = \{1, 2, \ldots, M\}$ where M depends on the sampling frequency. For example, collecting trading prices each hour between 09:00 and 15:00 gives $M = 6$. A vector of time stamps for the actual trading times for asset j is denoted as $q^{(j)}$.

2.1. Previous-tick realized correlation estimator (RCorr)

Using tick-by-tick high-frequency data, the quadratic covariance in (2) may be approximated directly by the realized covariance measure, (RCov_t):

$$\mathrm{RCov}_t = \sum_{i=1}^{M} r_{t,i} r'_{t,i}. \tag{3}$$

When $M \to \infty$, (3) converges to the theoretical quadratic covariance measure in (2) (Andersen et al., 2001, 2003).

In our example with assets a and b:

$$\mathrm{RCov}_t^{(ab)} = \begin{bmatrix} \mathrm{RCov}_t^{(aa)} & \mathrm{RCov}_t^{(ab)} \\ \mathrm{RCov}_t^{(ba)} & \mathrm{RCov}_t^{(bb)} \end{bmatrix}.$$

Realized correlation is then simply calculated as:

$$\mathrm{RCorr}_t^{(ab)} = \frac{\mathrm{RCov}_t^{(ab)}}{\sqrt{\left(\mathrm{RCov}_t^{(aa)} \times \mathrm{RCov}_t^{(bb)}\right)}}. \tag{4}$$

Andersen et al. (2001) and Barndorff-Nielsen and Shephard (2004) proposed that from the tick-by-tick high frequency data we can construct M synchronized returns using the previous-tick method. At each M sampling points the most recently observed log price for each asset, j, is recorded:

$$p_{t,i}^{(j)} = p_{Q_j(i)}^{(j)} \quad \text{where } Q_j(i) = \sup_q \left\{ \left\{ q^{(j)} | q^{(j)} \le t_0 + i\Delta \right\} \right\}. \tag{5}$$

[1] Griffin and Oomen (2011) analysed the properties of the standard realized covariance estimator (Andersen et al., 2001), the realized covariance plus lead–lag adjustment (Zhang et al., 2005) and the Hayashi–Yoshida covariance estimator (Hayashi and Yoshida, 2005) in cases with non-synchronous trading and microstructure noise using simulation techniques.

[2] Crude oil is the single most traded commodity in the world, and gasoil is the most significant refined petroleum product globally. The ICE gasoil futures are a global benchmark for all heating oil. The natural gas market is also one of the largest and most established energy markets (ICE Futures Europe, 2013).

[3] Semi-martingales allow for a decomposition of the price process into a predictable finite variation process and a local martingale including an infinite variation component. More specific details about the theory of quadratic variation and semi-martingale assumptions are described in, for example, Andersen et al. (2001) and Andersen et al. (2003).