Journal of Econometrics 193 (2016) 203–214

Contents lists available at ScienceDirect

Journal of Econometrics

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

A bias-corrected estimator of the covariation matrix of multiple security prices when both microstructure effects and sampling durations are persistent and endogenous

Shin S. Ikeda

Department of Policy Analysis, National Graduate Institute for Policy Studies, 7-22-1 Roppongi, Minato-ku, Tokyo 106-8677, Japan

ARTICLE INFO

Article history:
Received 3 April 2015
Received in revised form
30 January 2016
Accepted 24 February 2016
Available online 8 March 2016

JEL classification:
C14
C32

Keywords:
High frequency data
Covariation
Microstructure
Endogenous durations

ABSTRACT

I propose a bias-corrected non-parametric estimator of the covariation matrix of log security prices, designed as a convex combination of two realized kernels. The estimator is simple but possesses desirable statistical properties including consistency, asymptotic normality and the parametric rate of convergence in the presence of persistent, diurnally heteroskedastic and endogenous microstructure effects. It is robust to the asynchronous trading of multiple securities with persistent and endogenous refresh-time durations. I also prove the consistency of a subsampling-based estimator of the asymptotic covariance matrix of the proposed estimator. In simulations, the non-linear functions of the proposed estimator exhibit smaller bias than those based on a realized kernel, while only slightly increasing the variance. Thereby, the proposed estimator reduces the mean squared error.

1. Introduction

The covariation matrix is a multivariate and stochastic generalization of the univariate constant volatility parameter in continuous time finance models. It is central to important financial quantities, such as the no-arbitrage price of an option with stochastic volatility, optimal portfolio allocation, hedge ratios, market betas, and correlations among security returns. However, estimating the covariation matrix is difficult for several reasons. First, the integral is defined over a continuous time interval, but the available data are discrete. In any estimator, the discretization error creates an extra covariance component, which depends on the frequency of available data. Second, observed prices may exhibit a market microstructure effect, capturing imperfections and measurement errors in the market. The cumulative microstructure effect hinders the use of data recorded at very high frequency so that $n^{1/2}$-consistency is infeasible given n observations of returns. Third, the microstructure effect may persist with an unknown pattern. If the persistence of the microstructure effect is incorrectly specified, it may interfere with an otherwise consistent estimator of the covariation, which invalidates a parametric approach. Fourth, the durations of intra-daily sampling times may be neither synchronized among multiple securities, nor exogenous from the evolution of the market risk. The former may cause the under-estimation of co-movement among several markets (Epps, 1979), and the latter may influence the asymptotic bias and covariance of the estimator.

The previous studies have gradually fixed these issues. Gloter and Jacod (2001, (2.4)) indicate the convergence rate $n^{1/4}$ for the volatility estimation in a noisy environment. Aït-Sahalia, Mykland and Zhang (AMZ, 2005, Proposition 1) have already recognized the importance of a joint estimation of the volatility and the microstructure variance to achieve the rate $n^{1/4}$ for the former in their maximum likelihood framework. Zhang, Mykland and Aït-Sahalia (ZMA, 2005) incrementally improve a sparse-sampling realized variance estimator by subsampling, averaging, and two-scale bias-correcting to arrive at the first consistent non-parametric estimator with the rate $n^{1/6}$, which is improved to $n^{1/4}$ by Zhangs's (2006) multi-scale bias correction. The volatility estimation with random sampling times is an active area of recent research. For a subsampling-based approach with a bias correction, see Li, Mykland, Renault, Zhang, and Zheng (LMRZZ, 2014) in a univariate case without a microstructure effect, Li, Zhang and Zheng (LZZ, 2013) in a univariate case with a serially independent microstructure effect, and Bibinger and Mykland (2014) in a multivariate case with a serially independent microstructure effect. Aït-Sahalia, Fan and Xiu (AFX, 2010) and Shephard and Xiu (SX, 2014) extend Xiu's (2010) quasi

E-mail address: s-ikeda@grips.ac.jp.

http://dx.doi.org/10.1016/j.jeconom.2016.02.016
0304-4076/

maximum likelihood framework in which researchers deliberately mis-specify the stochastic volatility matrix as constant. Their estimators achieve consistency (AFX, Theorem 3) and the joint asymptotic mixed normality (SX, Assumption 6 and Equation (9)) given independence of sampling times from the efficient price and serial independence of microstructure effects. Bibinger et al. (2014, p.1319 and Theorem 4.4) provide a multivariate extension of Reiss's (2011) semi-parametric efficiency bound for their local method-of-moment estimator in the spectral domain. The benefit of a spectrum-based approach is natural, because the realized kernel approach (Barndorff-Nielsen et al., 2008; Ikeda, 2015) and the pre-averaging approach (Jacod, 2012; Christensen et al., 2010) resemble the kernel-based and tapering-based spectrum estimation (Priestley, 1981, Section 6.2 and Brillinger, 2001, Section 5.2). Koike (2014, 2016) apply Hayashi and Yoshida's (2005) estimator of the covariation to pre-averaged data given endogenous sampling times and a serially independent microstructure effect. Jacod and Mykland (2015, Theorem 4.1 and Remark 4.4) show that their univariate pre-averaging method with an adaptive local pre-averaging window can achieve the semi-parametric efficiency bound up to some constant scale. Note that all of these recent studies rely on independence of sampling times from the efficient price process and/or on serial independence of microstructure effects. A desirable estimator of the covariation should simultaneously alleviate all of these complications.

Barndorff-Nielsen, Hansen, Lunde and Shephard (BNHLS, 2011b) develop the first non-parametric covariation matrix estimator in the presence of a persistent microstructure effect. This estimator is called a realized kernel because it constitutes a kernel-weighted sum of sample autocovariation matrices of log price vectors. Realized kernels deal with a persistent microstructure effect by increasing the number of autocovariation matrices in a data-dependent manner. Unfortunately, the limiting distribution of this estimator is asymptotically biased by the long-run covariance of the microstructure effect. Consequently, the estimator converges at a rate slower than $n^{1/4}$. Moreover, non-linear functions of a finite-sample estimate become strongly biased. To simultaneously solve the above problems, this paper proposes a new non-parametric estimator of the covariation matrix with a built-in asymptotic bias correction. In constructing this estimator, I reformulate the problem as a *joint* estimation of the covariation of the efficient log price and the long-run covariance of the microstructure effect. By jointly estimating these two variations within the realized kernel framework, I create a new estimator of the covariation; a particular convex combination of two different realized kernels, which I therefore name the "Two-Scale Realized Kernel" estimator. The convex combination defining the estimator automatically eliminates the asymptotic biases of the two realized kernels in the proposed estimator. The built-in asymptotic bias correction also centers the limiting distribution of the estimator at the true quantity. This result is intuitive because the long-run covariance is jointly estimated with the covariation; hence, the former is no longer a nuisance parameter for the latter. The method also produces a new estimator of the long-run covariance of a persistent microstructure effect.

Because my estimator reduces the asymptotic bias, it converges more rapidly than that of the realized kernel because the bandwidth parameter can be designed to minimize the asymptotic covariance. Particularly, my estimator achieves an $n^{1/4}$-consistency in the presence of all of the aforementioned complications in high-frequency financial data, thereby solving the issues raised by LMRZZ (2014, Remarks 5 and 6). Moreover, this rate is robust to persistence, heteroskedasticity, and endogeneity in both microstructure effects and refresh-time durations imposed on multiple securities. I also establish a subsampling-based estimator of the asymptotic covariance of the proposed estimator and its consistency, facilitating an asymptotically valid inference. In simulation studies, the non-linear functions of my estimator reduce the biases in finite samples while only slightly increasing the variances. Consequently, the estimator tends to yield a smaller mean squared error (MSE) than the realized kernel. My estimator also appears robust to serially- and cross-sectionally dependent microstructure effects and asynchronous sampling durations. Because the subsampling, kernel and pre-averaging approaches share some asymptotic equivalence (BNHLS, 2008; Jacod et al., 2009; Bibinger and Mykland, 2014), this paper will be useful in extending these recent studies for a more general microstructure effect.

2. Basic setup and motivation

2.1. Model and assumptions

The setup is based on the constructs of BNHLS (2011b), Jacod (2012), and Ikeda (2015).[1] All integrals are defined element-wise under appropriate integrability and measurability conditions. The first difference operator with respect to any discrete index is denoted by Δ, and the Frobenius norm of any matrix A is given by $\|A\| := tr(A'A)$. $d \in \mathbb{N}$ is the number of risky securities, I_d is the $d \times d$ identity matrix, 0_d and O_d are zero matrices of respective sizes $d \times 1$ and $d \times d$, and $A|_B$ is the variable A under the condition B. Other notations are the Kronecker product $\otimes$, the indicator function $1_{\{\cdot\}}$, $x^{\pm} := \max\{\pm x, 0\}$ for $x \in \mathbb{R}$, the interval $[0, 1]$ of the trading period under consideration, the continuous-time indices $s, t \in [0, 1]$, the covariation matrix $\langle X, Y\rangle_t \in \mathbb{R}^{d\times d}$ of two processes $X_s, Y_s \in \mathbb{R}^d$ for $s \in [0, t] \subset [0, 1]$ (Protter, 2005, p. 66), and $\langle X\rangle_t := \langle X, X\rangle_t$. A process (X_t) is called uniformly L_q-bounded ($q \in \mathbb{N}$) if $\sup_t E[\|X_t\|^q] < \infty$. A collection of non-random multiple-indexed matrices $\Omega_{a_1,\ldots,a_k}$ is called q-summable ($q \in \mathbb{N}$) if $\sum_{a_i\in\mathbb{Z}}(1 + |a_i|^q)\|\Omega_{a_1,\ldots,a_k}\| < \infty$ for any $i = 1, \ldots, k$.

Assumption 1. (A-1) $(\mathcal{A}, \mathcal{F}, \mathcal{P})$ is a complete probability space with two independent filtrations $\mathbb{G} = (\mathcal{G}_t)_{t\in[0,1]}$ and $\mathbb{H} = (\mathcal{H}_t)_{t\in[0,1]}$ satisfying the usual conditions, $\mathcal{G} = \mathcal{G}_1$, and $E_t[\cdot] := E[\cdot|\mathcal{G}_t]$.

(A-2) Trading times of multiple securities are synchronized at refresh times $\{t_i\}_{i=1,\ldots,n+2m-1}$ for $n, m \in \mathbb{N}$ given $m = o(n)$ (BNHLS, 2011b), and $\Delta t_i = D_{n,i}/n$ for $i \geq 2$.
- (a) $D_{n,i}$ is $\mathcal{G}_{t_i}$-measurable, uniformly positive and uniformly L_4-bounded.
- (b) $E_{t_{\lfloor ns\rfloor-1}}[D^r_{n,\lfloor ns\rfloor}] \xrightarrow{p} \mathcal{X}_{r,s}$ as $n \to \infty$ for any $r \in (0, 2]$ and any $s \in (0, 1]$.
- (c) $\mathcal{X}_{r,s}$ is left continuous at any $s \in (0, 1]$, $0 < \mathcal{X}_{r,s} < \infty$ and $\langle \mathcal{X}_r\rangle_1 < \infty$ uniformly.

(A-3) $p(t_i)$ is the $d \times 1$ vector of log security prices in the refresh-time sampling such that

$$p(t_i) = p^*(t_i) + \epsilon(t_i) \tag{1}$$

in which $\epsilon(t_i) = u(t_i) + v(t_i)$, $u(t_i) = \xi_{t_i}\bar{u}_{t_i}$, and $v(t_i) = n^{1/2}\sum_{\gamma\in\mathbb{Z}} \psi_{\gamma,t_{i-\gamma-1}}\Delta W_{t_{i-\gamma}}$.
- (a) ξ_t is $\mathbb{G}$-adapted, positive definite, uniformly bounded, and Lipschitz continuous.
- (b) $\bar{u}_{t_i}$ is $(\mathcal{H}_{t_i})$-adapted, zero-mean, α-mixing of size $-\varpi/(\varpi - 4)$, uniformly L_ϖ-bounded, and fourth-order stationary with ϖ-summable jth order cumulant matrices $\bar{\Omega}_{a_1,\ldots,a_{j-1}}$ ($j = 2, 3, 4$; $a_1, \ldots, a_{j-1} \in \mathbb{Z}$) for some $\varpi > 4$.

[1] Special thanks to Jean Jacod for kindly providing an earlier draft of Jacod (2012).