Journal of Econometrics 162 (2011) 149–169

Contents lists available at ScienceDirect

Journal of Econometrics

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

Multivariate realised kernels: Consistent positive semi-definite estimators of the covariation of equity prices with noise and non-synchronous trading[☆]

Ole E. Barndorff-Nielsen[a,b], Peter Reinhard Hansen[c], Asger Lunde[d,b,*], Neil Shephard[e,f]

[a] The T.N. Thiele Centre for Mathematics in Natural Science, Department of Mathematical Sciences, University of Aarhus, Ny Munkegade, DK-8000 Aarhus C, Denmark
[b] CREATES, University of Aarhus, Denmark
[c] Department of Economics, Stanford University, Landau Economics Building, 579 Serra Mall, Stanford, CA 94305-6072, USA
[d] School of Economics and Management, Aarhus University, Bartholins Allé 10, DK-8000 Aarhus C, Denmark
[e] Oxford-Man Institute, University of Oxford, Eagle House, Walton Well Road, Oxford OX2 6ED, UK
[f] Department of Economics, University of Oxford, UK

ARTICLE INFO

Article history:
Received 14 July 2010
Received in revised form
14 July 2010
Accepted 28 July 2010
Available online 27 January 2011

Keywords:
HAC estimator
Long run variance estimator
Market frictions
Quadratic variation
Realised variance

ABSTRACT

We propose a multivariate realised kernel to estimate the ex-post covariation of log-prices. We show this new consistent estimator is guaranteed to be positive semi-definite and is robust to measurement error of certain types and can also handle non-synchronous trading. It is the first estimator which has these three properties which are all essential for empirical work in this area. We derive the large sample asymptotics of this estimator and assess its accuracy using a Monte Carlo study. We implement the estimator on some US equity data, comparing our results to previous work which has used returns measured over 5 or 10 min intervals. We show that the new estimator is substantially more precise.

1. Introduction

The last seven years has seen dramatic improvements in the way econometricians think about time-varying financial volatility, first brought about by harnessing high frequency data and then by mitigating the influence of market microstructure effects. Extending this work to the multivariate case is challenging as this needs to additionally remove the effects of non-synchronous trading while simultaneously requiring that the covariance matrix estimator be positive semi-definite. In this paper we provide the first estimator which achieves all these objectives. This will be called the multivariate realised kernel, which we will define in Eq. (1).

We study a d-dimensional log price process $X = (X^{(1)}, X^{(2)}, \ldots, X^{(d)})'$. These prices are observed irregularly and non-synchronous over the interval $[0, T]$. For simplicity of exposition we take $T = 1$ throughout the paper. These observations could be trades or quote updates. The observation times for the i-th asset will be written as $t_1^{(i)}, t_2^{(i)}, \ldots$. This means the available database of prices is $X^{(i)}(t_j^{(i)})$, for $j = 1, 2, \ldots, N^{(i)}(1)$, and $i = 1, 2, \ldots, d$. Here $N^{(i)}(t)$ counts the number of distinct data points available for the i-th asset up to time t.

X is assumed to be driven by Y, the efficient price, abstracting from market microstructure effects. The efficient price is modelled as a *Brownian semimartingale* ($Y \in \mathcal{BSM}$) defined on some filtered probability space $(\Omega, \mathcal{F}, (\mathcal{F}_t), P)$,

$$Y(t) = \int_0^t a(u)\mathrm{d}u + \int_0^t \sigma(u)\mathrm{d}W(u),$$

where a is a vector of elements which are predictable locally bounded drifts, σ is a càdlàg volatility matrix process and W is a vector of independent Brownian motions. For reviews of the econometrics of this type of process see, for example, Ghysels et al. (1996). If $Y \in \mathcal{BSM}$ then its ex-post covariation, which we will focus on for reasons explained in a moment, is

$$[Y](1) = \int_0^1 \Sigma(u)\mathrm{d}u, \quad \text{where } \Sigma = \sigma\sigma',$$

where

$$[Y](1) = \operatorname*{plim}_{n\to\infty} \sum_{j=1}^{n} \{Y(t_j) - Y(t_{j-1})\}\{Y(t_j) - Y(t_{j-1})\}',$$

☆ The second and fourth author are also affiliated with CREATES, a research center funded by the Danish National Research Foundation. The Ox language of Doornik (2006) was used to perform the calculations reported here. We thank Torben Andersen, Tim Bollerslev, Ron Gallant, Xin Huang, Oliver Linton and Kevin Sheppard for comments on a previous version of this paper.

* Corresponding author at: School of Economics and Management, Aarhus University, Bartholins Allé 10, DK-8000 Aarhus C, Denmark.
E-mail addresses: oebn@imf.au.dk (O.E. Barndorff-Nielsen), peter.hansen@stanford.edu (P.R. Hansen), alunde@asb.dk (A. Lunde), neil.shephard@economics.ox.ac.uk (N. Shephard).

0304-4076/$ – see front matter © 2011 Elsevier B.V. All rights reserved.
doi:10.1016/j.jeconom.2010.07.009

(e.g. Protter (2004, p. 66–77) and Jacod and Shiryaev (2003, p. 51)) for any sequence of deterministic synchronized partitions $0 = t_0 < t_1 < \cdots < t_n = 1$ with $\sup_j\{t_{j+1} - t_j\} \to 0$ for $n \to \infty$. This is the quadratic variation of Y.

The contribution of this paper is to construct a consistent, positive semi-definite (psd) estimator of $[Y](1)$ from our database of asset prices. The challenges of doing this are three-fold: (i) there are market microstructure effects $U = X - Y$, (ii) the data is irregularly spaced and non-synchronous, (iii) the market microstructure effects are not statistically independent of the Y process.

Quadratic variation is crucial to the economics of financial risk. This is reviewed by, for example, Andersen et al. (2010) and Barndorff-Nielsen and Shephard (2007), who provide very extensive references. The economic importance of this line of research has recently been reinforced by the insight of Bollerslev et al. (2009) who have showed that expected stock returns seem well explained by the variance risk premium (the difference between the implied and realised variance) and this risk premium is only detectable using the power of high frequency data. See also the papers by Drechsler and Yaron (2011), Fleming et al. (2003) and de Pooter et al. (2008).

Our analysis builds upon earlier work on the effect of noise on univariate estimators of $[Y](1)$ by, amongst others, Zhou (1996), Andersen et al. (2000), Bandi and Russell (2008), Zhang et al. (2005), Hansen and Lunde (2006), Hansen et al. (2008), Kalnina and Linton (2008), Zhang (2006), Barndorff-Nielsen et al. (2008), Renault and Werker (2011), Hansen and Horel (2009), Jacod et al. (2009) and Andersen et al. (2011). The case of no noise is dealt with in the same spirit as the papers by Andersen et al. (2001), Barndorff-Nielsen and Shephard (2002, 2004), Mykland and Zhang (2006, 2009) Goncalves and Meddahi (2009) and Jacod and Protter (1998).

A distinctive feature of multivariate financial data is the phenomenon of non-synchronous trading or nontrading. These two terms are distinct. The first refers to the fact that any two assets rarely trade at the same instant. The latter to situations where one assets is trading frequently over a period while some other assets do not trade. The treatment of non-synchronous trading effects dates back to Fisher (1966). For several years researchers focused mainly on the effects that stale quotes have on daily closing prices. Campbell et al. (1997, Chapter 3) provides a survey of this literature. When increasing the sampling frequency beyond the inter-hour level several authors have demonstrated a severe bias towards zero in covariation statistics. This phenomenon is often referred to as the Epps effect. Epps (1979) found this bias for stock returns, and it has also been demonstrated to hold for foreign exchange returns, see Guillaume et al. (1997). This is confirmed in our empirical work where realised covariances computed using high frequency data, over specified fixed time periods such as 15 s, dramatically underestimate the degree of dependence between assets. Some recent econometric work on this topic includes Malliavin and Mancino (2002), Reno (2003), Martens (unpublished paper), Hayashi and Yoshida (2005), Bandi and Russell (unpublished paper), Voev and Lunde (2007), Griffin and Oomen (2011) and Large (unpublished paper). We will draw ideas from this work.

Our estimator, the *multivariate realised kernel*, differs from the univariate realised kernel estimator by Barndorff-Nielsen et al. (2008) in important ways. The latter converges a rate $n^{1/4}$ but critically relies on the assumption that the noise is a white noise process, and Barndorff-Nielsen et al. (2008) stress that their estimator cannot be applied to tick-by-tick data. In order not to be in obvious violation of the iid assumption, Barndorff-Nielsen et al. (2008) apply their estimator to prices that are (on average) sampled every minute or so. Here, in the present paper, we allow for a general form of noise that is consistent with the empirical features of tick-by-tick data. For this reason we adopt a larger bandwidth that has the implication that our multivariate realised kernel estimator converges at rate $n^{1/5}$. Although this rate is slower than $n^{1/4}$ it is, from a practical viewpoint, important to acknowledge that there are only 390 one-minute returns in a typical trading day, while many shares trade several thousand times, and $390^{1/4} < 2000^{1/5}$. So the rates of convergence will not (alone) tell us which estimators will be most accurate in practice — even for the univariate estimation problem. In addition to being robust to noise with a general form of dependence, the $n^{1/5}$ convergence rate enables us to construct an estimator that is guaranteed to psd, which is not the case for the estimator by Barndorff-Nielsen et al. (2008). Moreover, our analysis of irregularly spaced and non-synchronous observations causes the asymptotic distribution of our estimator to be quite different from that in Barndorff-Nielsen et al. (2008). We discuss the differences between these estimators in greater details in Section 6.1.

The structure of the paper is as follows. In Section 2 we synchronise the timing of the multivariate data using what we call Refresh Time. This allows us to refine high frequency returns and in turn the multivariate realised kernel. Further we make precise the assumptions we make use of in our theorems to study the behaviour of our statistics. In Section 3 we give a detailed discussion of the asymptotic distribution of realised kernels in the univariate case. The analysis is then extended to the multivariate case. Section 4 contains a summary of a simulation experiment designed to investigate the finite sample properties of our estimator. Section 5 contains some results from implementing our estimators on some US stock price data taken from the TAQ database. We analyse up to 30 dimensional covariance matrices, and demonstrate efficiency gains that are around 20-fold compared to using daily data. This is followed by a section on extensions and further remarks, while the main part of the paper is finished by a conclusion. This is followed by an Appendix which contains the proofs of various theorems given in the paper, and an Appendix with results related to Refresh Time sampling. More details of our empirical results and simulation experiments are given in a web appendix which can be found at http://mit.econ.au.dk/vip_htm/alunde/BNHLS/BNHLS.htm.

2. Defining the multivariate realised kernel

2.1. Synchronising data: refresh time

Non-synchronous trading delivers fresh (trade or quote) prices at irregularly spaced times which differ across stocks. Dealing with non-synchronous trading has been an active area of research in financial econometrics in recent years, e.g. Hayashi and Yoshida (2005), Voev and Lunde (2007) and Large (unpublished paper). Stale prices are a key feature of estimating covariances in financial econometrics as recognised at least since Epps (1979), for they induce cross-autocorrelation amongst asset price returns.

Write the number of observations in the i-th asset made up to time t as the counting process $N^{(i)}(t)$, and the times at which trades are made as $t_1^{(i)}, t_2^{(i)}, \ldots$. We now define *refresh time* which will be key to the construction of multivariate realised kernels. This time scale was used in a cointegration study of price discovery by Harris et al. (1995), and Martens (unpublished paper) has used the same idea in the context of realised covariances.

Definition 1. Refresh Time for $t \in [0, 1]$. We define the first refresh time as $\tau_1 = \max(t_1^{(1)}, \ldots, t_1^{(d)})$, and then subsequent refresh times as

$$\tau_{j+1} = \max(t^{(1)}_{N^{(1)}_{\tau_j}+1}, \ldots, t^{(d)}_{N^{(d)}_{\tau_j}+1}).$$