ARTICLE IN PRESS

Journal of Econometrics ▮ (▮▮▮▮) ▮▮▮–▮▮▮

Contents lists available at ScienceDirect

Journal of Econometrics

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

Estimation of integrated quadratic covariation with endogenous sampling times[☆]

Yoann Potiron[a,*], Per A. Mykland[b,1]

[a] Faculty of Business and Commerce, Keio University, 2-15-45 Mita, Minato-ku, Tokyo, 108-8345, Japan

[b] Department of Statistics, The University of Chicago, 5734 S. University Avenue Chicago, IL 60637, United States

ARTICLE INFO

Article history:
Received 30 January 2016
Received in revised form
21 June 2016
Accepted 27 October 2016
Available online xxxx

JEL classification:
C01
C02
C13
C14
C22
C32
C58

Keywords:
Asymptotic bias
Asynchronous times
Endogenous model
Hayashi–Yoshida estimator
High-frequency data
Quadratic covariation
Time endogeneity

ABSTRACT

When estimating high-frequency covariance (quadratic covariation) of two arbitrary assets observed asynchronously, simple assumptions, such as independence, are usually imposed on the relationship between the prices process and the observation times. In this paper, we introduce a general endogenous two-dimensional nonparametric model. Because an observation is generated whenever an auxiliary process called *observation time process* hits one of the two boundary processes, it is called the *hitting boundary process with time process* (HBT) model. We establish a central limit theorem for the Hayashi–Yoshida (HY) estimator under HBT in the case where the price process and the observation price process follow a continuous Itô process. We obtain an asymptotic bias. We provide an estimator of the latter as well as a bias-corrected HY estimator of the high-frequency covariance. In addition, we give a consistent estimator of the associated standard error.

1. Introduction

Covariation between two assets is a crucial quantity in finance. Fundamental examples include optimal asset allocation and risk management. In the past few years, using the increasing amount of high-frequency data available, many papers have been published about estimating this covariance. Suppose that the latent log-price of two arbitrary assets $X_t = (X_t^{(1)}, X_t^{(2)})$ follows a continuous Itô process

$$dX_t^{(1)} := \mu_t^{(1)} dt + \sigma_t^{(1)} dW_t^{(1)}, \tag{1}$$

$$dX_t^{(2)} := \mu_t^{(2)} dt + \sigma_t^{(2)} dW_t^{(2)}, \tag{2}$$

where $\mu_t^{(1)}, \mu_t^{(2)}, \sigma_t^{(1)}, \sigma_t^{(2)}$ are random processes, and $W_t^{(1)}$ and $W_t^{(2)}$ are standard Brownian motions, with (random) high-frequency correlation $d\langle W^{(1)}, W^{(2)}\rangle_t = \rho_t dt$. Econometrics usually seeks to infer the *integrated covariation*

$$\langle X^{(1)}, X^{(2)}\rangle_t = \int_0^t \rho_u \sigma_u^{(1)} \sigma_u^{(2)} du.$$

Earlier results were focused on estimating the integrated variance of a single asset, starting from the probabilistic point of view (Genon-Catalot and Jacod, 1993; Jacod, 1994). Barndorff-Nielsen and Shephard (2001, 2002) introduced the problem in econometrics. Adapted to two dimensions, if each process is observed simultaneously at (possibly random) times $\tau_{0,n} := 0, \tau_{1,n}, \ldots, \tau_{N_n,n}$ the

☆ We would like to thank Simon Clinet, Mathieu Rosenbaum, Steven Lalley, Jianqing Fan (the Editor), an anonymous Associate Editor, and one anonymous referee for helpful discussions and advice. Financial support from the National Science Foundation under grant DMS 14-07812 is greatly acknowledged.

* Corresponding author.
E-mail addresses: potiron@fbc.keio.ac.jp (Y. Potiron), mykland@pascal.uchicago.edu (P.A. Mykland).
1 Fax: +1 773 702 9810.

http://dx.doi.org/10.1016/j.jeconom.2016.10.004
0304-4076/

Please cite this article in press as: Potiron, Y., Mykland, P.A., Estimation of integrated quadratic covariation with endogenous sampling times. Journal of Econometrics (2016), http://dx.doi.org/10.1016/j.jeconom.2016.10.004

realized covariation $\left[X^{(1)}, X^{(2)}\right]_t$ is defined as the sum of cross log returns

$$\left[X^{(1)}, X^{(2)}\right]_t = \sum_{\tau_{i,n} \le t} \Delta X^{(1)}_{\tau_{i,n}} \Delta X^{(2)}_{\tau_{i,n}}, \tag{3}$$

where for any positive integer i, $\Delta X^{(k)}_{\tau_{i,n}} = X^{(k)}_{\tau_{i,n}} - X^{(k)}_{\tau_{i-1,n}}$ corresponds to the increment of the kth process between the last two sampling times. As the observation intervals $\Delta\tau_{i,n}$ get closer (and the number of observations N_n goes to infinity), $\left[X^{(1)}, X^{(2)}\right]_t \xrightarrow{\mathbb{P}} \langle X^{(1)}, X^{(2)}\rangle_t$ (see e.g. Theorem I.4.47 in Jacod and Shiryaev (2003)). Furthermore, when the observation times $\tau_{i,n}$ are independent of the prices process X_t, its estimation error follows a mixed normal distribution (Jacod and Protter, 1998; Zhang, 2001; Mykland and Zhang, 2006). This gives us insight on how to estimate the integrated covariation. However, in practice, these two assumptions are usually not satisfied. The observation times of the two assets are rarely *synchronous* and there is *endogeneity* in the price sampling times.

The first issue has been studied for a long time. The lack of synchronicity often creates undesirable effects in inference. If we sample at very high frequencies, we observe the Epps effect (Epps, 1979), i.e. the correlation estimates are drastically decreased compared to an estimate with sparse observations. Hayashi and Yoshida (2005) introduced the so-called *Hayashi–Yoshida estimator* (HY)

$$\widehat{\langle X^{(1)}, X^{(2)}\rangle}^{HY}_t = \sum_{\tau^{(1)}_{i,n}, \tau^{(2)}_{j,n} < t} \Delta X^{(1)}_{\tau^{(1)}_{i,n}} \Delta X^{(2)}_{\tau^{(2)}_{j,n}} \mathbf{1}_{\left\{[\tau^{(1)}_{i-1,n}, \tau^{(1)}_{i,n}) \cap [\tau^{(2)}_{j-1,n}, \tau^{(2)}_{j,n}) \neq \emptyset\right\}}, \tag{4}$$

where $\tau^{(k)}_{i,n}$ are the observation times of the kth asset. Note that if the observations of both processes occur simultaneously, (3) and (4) are equal. The consistency of this estimator was achieved in Hayashi and Yoshida (2005) and Hayashi and Kusuoka (2008). The corresponding central limit theorems were investigated in Hayashi and Yoshida (2008, 2011) under strong predictability of observation times, which is a more restrictive assumption than only assuming they are stopping times but still allows some dependence between prices and observation times. Recently, Koike (2014, 2015) extended the pre-averaged Hayashi–Yoshida estimator first under predictability of observation times, and then under a more general endogenous setting of stopping times. Other examples of high-frequency covariance estimators can be found in Zhang (2011), Barndorff-Nielsen et al. (2011), Aït-Sahalia et al. (2010), Christensen et al. (2010, 2013).

In a general one-dimensional endogenous model, the asymptotic behavior of the realized volatility (3) has been investigated in the case of sampling times given by hitting times on a grid (Fukasawa, 2010a; Robert and Rosenbaum, 2011, 2012; Fukasawa and Rosenbaum, 2012). Due to the regularity of those three models (see the discussion in the latter paper), they do not obtain any bias in the limit distribution of the normalized error. Also, the case of strongly predictable stopping times is treated in Hayashi and Yoshida (2011). Finally, two general results (Fukasawa, 2010b; Li et al., 2014) showed that we can identify and estimate the asymptotic bias.

The primary goal of this paper is to bias-correct the HY. Note that estimating the bias is more challenging than in the volatility case because observations are asynchronous. In particular, the estimator will involve a quantity that can be considered as the *tricity* of Li et al. (2014), but with a more intricate definition because of the asynchronicity in sampling times. This new definition can be seen as an analogy with the generalization of the RV estimator (3) by the HY estimator (4).

Another very important issue to address is the estimation of the asymptotic standard deviation. First, because the model is more general than in the no-endogeneity work, the theoretical asymptotic variance will be different. Consequently, a new variance estimator, which takes into proper account the endogeneity, will be given.

The authors want to take no position on the joint distribution of the log-return and the next observation time that corresponds to an asset price change because they know that their unknown relationship is most likely contributing to the bias and the variance of the high-frequency covariance's estimate when we (wrongly) assume full independence between the price process and observation times. For this purpose, they introduce the *hitting boundary process with time process* (HBT) model.

Finally, techniques developed in the proofs are innovative in the sense that they reduce the normalized error of the Hayashi–Yoshida estimator to a discrete process, which is locally a uniformly ergodic homogeneous Markov chain. Thus, the problem can be solved locally, and because we assume that the volatility of assets is continuous, the error of approximation between the local Markov structure and the real structure of the normalized error vanishes asymptotically. This technique is not problem-specific, and it can very much be applied to other estimators dealing with temporal data.

The paper is organized as follows. We introduce the HBT model in Section 2. Examples covered by this model are given in Section 3. The main theorem of this work, the limit distribution of the normalized error is given in Section 4. Estimators of the asymptotic bias and variance are provided in Section 5. We carry out numerical simulations in Section 6 to corroborate the theory. Proofs are developed in the Appendix.

2. Definition of the HBT model

We first introduce the model in 1-dimension. We assume that for any positive integer i, τ_{i+1} is the next arrival time (after τ_i) that corresponds to an actual change of price. In particular, several trades can occur at the same price Z_{τ_i} between τ_i and τ_{i+1}, but no trade can occur with a price different than Z_{τ_i} before τ_{i+1}. We also assume that X_t is the efficient (log) price of the security of interest. In addition, we assume that the observations are noisy and that we observe $Z_{\tau_i} := X_{\tau_i} + \epsilon_{\tau_i}$ where the microstructure noise ϵ_{τ_i} can be expressed as a known function of the observed prices $Z_0, \dots, Z_{\tau_i}$. As an example, Robert and Rosenbaum (2012) showed in (2.3) in p. 5 that the model with uncertainty zones can be written with that noise structure if we assume that we know the friction parameter η. Finally, we define $\alpha > 0$ as the tick size, and we assume that the observed price Z_{τ_i} lays on the tick grid, i.e. there exist positive integers m_i such that $Z_{\tau_i} := m_i\alpha$.

Empirically, no economical model based on rational behaviors of agents on the stock markets, that shed light on the relationship between the efficient return ΔX_{τ_i} and time before the next price change $\Delta\tau_i = \tau_i - \tau_{i-1}$, has won unanimous support. When arrival times are independent of the asset price, it follows directly from the continuous Itô-assumption that the dependence structure is such that the return ΔX_{τ_i} is a function of $\Delta\tau_i$. The longer we wait, the bigger the variance of the return is expected to be. In this paper, we take the opposite point of view by building a model in which τ_i is defined as a function of the efficient price path. For that purpose, we define the *observation time process* $X^{(t)}_t$ that will drive the observation times. We also define the *down process* $d_t(s)$ and the *up process* $u_t(s)$. Note that for any $t \ge 0$, we assume that d_t and u_t are functions on $\mathbb{R}^+$. We also assume that the down process takes only negative values and that the up process takes only positive values. A new observation time will be generated whenever one of those two processes is hit by the increment of the observation time process. Then, the increment of the observation time process will start again from 0, and the next observation time

Please cite this article in press as: Potiron, Y., Mykland, P.A., Estimation of integrated quadratic covariation with endogenous sampling times. Journal of Econometrics (2016), http://dx.doi.org/10.1016/j.jeconom.2016.10.004