Journal of Econometrics 195 (2016) 33–50

Contents lists available at ScienceDirect

Journal of Econometrics

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

Efficient estimation of integrated volatility incorporating trading information

Yingying Li [a], Shangyu Xie [b], Xinghua Zheng [a,*]

[a] Hong Kong University of Science and Technology, Hong Kong Special Administrative Region

[b] RCAF and School of Banking and Finance, University of International Business and Economics, Beijing, China

ARTICLE INFO

Article history:
Received 21 September 2014
Received in revised form
9 May 2016
Accepted 31 May 2016
Available online 15 July 2016

JEL classification:
C14
C13
D40

Keywords:
High frequency data
Integrated volatility
Market microstructure noise
Realized volatility
Efficiency

ABSTRACT

We consider a setting where market microstructure noise is a parametric function of trading information, possibly with a remaining noise component. Assuming that the remaining noise is $O_p(1/\sqrt{n})$, allowing irregular times and jumps, we show that we can estimate the parameters at rate n, and propose a volatility estimator which enjoys $\sqrt{n}$ convergence rate. Simulation studies show that our method performs well even with model misspecification and rounding. Empirical studies demonstrate the practical relevance and advantages of our method. Furthermore, we find that a simple model can account for a high percentage of the total variation in microstructure noise.

1. Introduction

High-frequency data has attracted tremendous attention in recent years. In the vast literature of high frequency data studies, a central focus is to estimate volatilities consistently and efficiently. A major challenge arises from the presence of market microstructure noise, which is an integral part of the financial market.

A widely used assumption about market microstructure noise in the volatility estimation literature is that they are independent and identically distributed (i.i.d.) and additive to the log-price process. More specifically, over a time interval of interest $[0, T]$, one observes at times $0 = t_0 < t_1 < \cdots < t_n = T$,

$$Y_{t_k} = X_{t_k} + \varepsilon_{t_k}, \quad k = 0, 1, \ldots, n, \tag{1}$$

where X_{t_k} and ε_{t_k} denote the latent log-price and market microstructure noise at the observation time t_k respectively, and ε_{t_k}'s are i.i.d. and independent of X. Consistent estimators of the integrated volatility under this setting include the two-scales realized volatility (TSRV, Zhang et al. (2005)), multi-scale realized volatility (MSRV, Zhang (2006)), realized kernels (RK, Barndorff-Nielsen et al. (2008)), pre-averaging estimator (PAV, Jacod et al. (2009) and Podolskij and Vetter (2009)), and quasi-maximum likelihood estimator (QMLE, Xiu (2010)). The optimal rate of convergence is $n^{1/4}$ (Gloter and Jacod, 2001). MSRV, RK, PAV and QMLE are all rate-optimal.

On the other hand, studies on market microstructure noise can be traced back to the 1980s; see, Black (1986), Madhavan (2000), O'Hara (1995), Stoll (2003), and Hasbrouck (2007), among many others. An example of a simple model for microstructure noise is the "implicit measure of the effective bid–ask spread" as in Roll (1984):

$$\varepsilon_{t_k} = \alpha I_{b/s}(t_k), \tag{2}$$

where $I_{b/s}(t_k)$ denotes the trade type, indicating if the trade is buyer-initiated (+1) or seller-initiated (−1); and the coefficient α can be interpreted as one-half of the effective bid–ask spread. Roll's model was extended in Glosten and Harris (1988) by incorporating the trading volume:

$$\varepsilon_{t_k} = I_{b/s}(t_k)\big(\alpha + \beta V_{t_k}\big), \tag{3}$$

* Correspondence to: Department of Information Systems, Business Statistics and Operations Management, Hong Kong University of Science and Technology, Clear Water Bay, Kowloon, Hong Kong Special Administrative Region.

E-mail address: xhzheng@ust.hk (X. Zheng).

http://dx.doi.org/10.1016/j.jeconom.2016.05.017
0304-4076/© 2016 Elsevier B.V. All rights reserved.

where V_{t_k} denotes the trading volume at time t_k. Almgren and Chriss (2000) consider an optimal execution problem and they model the market impact as a function of trade type and trading rate. A variant in the spirit of (3) is then the following

$$\varepsilon_{t_k} = I_{b/s}(t_k)\big(\alpha + \beta V_{t_k}/\Delta t_k\big), \tag{4}$$

where $\Delta t_k := t_k - t_{k-1}$ denotes the duration between two consecutive transactions. A pioneer paper in high-frequency volatility estimation literature Aït-Sahalia et al. (2005) also models the market microstructure noise in a parametric way (but without covariates), and shows that even with model misspecification, such parametric modeling enables one to estimate the volatility at the optimal rate $n^{1/4}$ under Model (1).

Rich information is available in high-frequency data. For example, in trade data, in addition to transaction prices, trading volumes are also reported. Furthermore, quotes data are also publicly available, which contain even richer information. Individuals or institutions can also have additional trading information. This motivates us to consider taking advantage of the rich information available in the market and study the setting where the noise term in (1) can be further modeled using available trading information through a parametric function such as (2)–(4). The function can be either linear or nonlinear. We show that in this case, even with irregular observation times and jumps, the parameters in the noise model can be estimated with high precision (with convergence rate n instead of $\sqrt{n}$ as in usual parametric estimations), and consequently the latent log-prices can be estimated highly accurately. This allows us to further obtain an efficient volatility estimator, based on the estimated log-prices. We call this estimator "estimated-price realized volatility" (ERV). We show that the proposed ERV, which is based on noisy observations, provides $\sqrt{n}$ rate of convergence and the same asymptotic properties as realized volatility (RV) based on latent log-prices.

Given the complexity of market microstructure noise, we further consider the setting where market microstructure noise admits an extra noise component. Under the assumption that the extra noise component is $O_p(1/\sqrt{n})$, we propose another volatility estimator ERV_{ext} which still enjoys $\sqrt{n}$ rate of convergence. Numerically, we demonstrate that ERV_{ext} (and E-QMLE, another estimator that we propose without establishing its asymptotic properties) performs well even in the situations where there are rounding errors and model misspecification on the parametric model. More importantly, extensive empirical studies demonstrate the relevance of our method and the advantages of our estimator. An interesting additional empirical finding is that, for various stocks examined, a simple model for market microstructure noise, which incorporates only trade type and trading rate, can account for around 70%–80% of the total variation in noise. Our analysis also provides a useful framework for studying the market microstructure.

An independent and concurrent research, Chaker (2013), shares the same spirit as this paper. There are however quite a few major differences. In our paper, the models for market microstructure noise are allowed to be nonlinear[1]; the observation times are allowed to be irregularly spaced, in fact the observation times can even be endogenous as what is considered in Li et al. (2014, 2013); and jumps are allowed in the latent price process. Furthermore, our small additional noise assumption leads to rather different estimators and asymptotic properties. Some earlier works along this line include Hansen and Lunde (2006) and Engle and Sun (2007). Hansen and Lunde (2006) consider in the Section 6 of their paper how to estimate the efficient prices from bid and ask quotes and transaction prices, based on a vector autoregressive model. Engle and Sun (2007) use GARCH model for the efficient price process and a two-component ARMA model for the noise.

The rest of this paper is organized as follows. Section 2 presents our proposed ERV estimator and its extensions, together with their asymptotic properties. Sections 3 and 4 are devoted to simulation studies and empirical studies, respectively. Section 5 concludes and discusses related issues. Proofs are given in the Appendix.

2. Estimated-price realized volatility

2.1. When noise can be completely modeled

We assume that the latent log-price process has the following representation:

$$dX_t = \mu_t\,dt + \sigma_t\,dW_t + dJ_t, \quad t \in [0, T], \tag{5}$$

where W_t is a Brownian motion, μ_t and σ_t are adapted locally bounded random processes, and J_t is a pure jump process, all defined on a common filtered probability space $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t\geq 0}, P)$. The quantity of interest is the quadratic variation (QV) $\int_0^T \sigma_t^2 dt + \sum_{t\leq T}(\Delta J_t)^2$ with $\Delta J_t := J_t - J_{t-}$, or more often, the continuous part of QV, commonly referred to as the integrated volatility IV $:= \int_0^T \sigma_t^2 dt$. Without loss of generality, we set $T = 1$.

Following Li et al. (2014) we shall allow the observation times to be endogenous and adopt some of the notation therein. Denote the observation times at stage n by

$$0 = t_{n,0} < t_{n,1} < \cdots < t_{n,N} \leq 1. \tag{6}$$

Here n is a latent number that characterizes the observation frequency, and $N = N(n)$, which may be random, stands for the actual number of observations before time 1. See Section 3 of Li et al. (2014) for various examples in this regard. In the exogenous case when observation times are either deterministic or random but independent of the price process, without loss of generality, we can and will take $n = N$. More generally, we will establish a feasible asymptotic theory in terms of N under the assumption that N/n has a (possibly random) probability limit F. Let us mention that in the endogenous setting, while in general n may not be uniquely defined[2], the feasible asymptotic theory will be independent of n (see also Remark 1 in Li et al. (2014) or the discussion following Assumption (O) in Jacod et al. (2014)). For notational ease, when there is no confusion we shall write $t_{n,1}$, $t_{n,k}$ as t_1, t_k etc.

In this subsection we consider the setting where the market microstructure noise can be completely modeled by trading information, through a parametric function g:

$$Y_{t_k} = X_{t_k} + g(\mathbf{Z}_{t_k}; \boldsymbol{\theta}_0), \tag{7}$$

where Y_{t_k} is the observed log-prices at time t_k, $\mathbf{Z}_{t_k}$ is the information set which can include, but not limited to, trade type, trading volume and bid–ask bounds; and $\boldsymbol{\theta}_0$ is a (finite-dimensional) parameter. The aforementioned Models (2)–(4) are all examples of g. We shall also consider some other forms of g in the numerical studies in Section 3. In our theoretical analysis, we allow the

[1] Nonlinear models are relevant in practice. For example, Keim and Madhavan (1996) show that the price impact of block trades is a concave function of order size.

[2] To see this, similar to Examples 4–6 in Li et al. (2014), define the observation times $t_{n,i}$ to be successive hitting times: $t_{n,i+1} := \inf\{t > t_{n,i} : |X_t - X_{t_{n,i}}| \geq Z_{i+1}/\sqrt{n}\}$, where Z_i's are random variables which may or may not be i.i.d. Such a definition suggests that n is a natural characterization of the observation frequency. However, if another person takes $\widetilde{Z}_{i+1} = \sqrt{2}Z_{i+1}$, then $t_{n,i+1}$ can be equivalently defined as $\widetilde{t}_{2n,i+1} := \inf\{t > \widetilde{t}_{2n,i} : |X_t - X_{\widetilde{t}_{2n,i}}| \geq \widetilde{Z}_{i+1}/\sqrt{2n}\}$. The latter definition suggests $2n$ as another characterization of the observation frequency.