Economics Letters 150 (2017) 59–62

Contents lists available at ScienceDirect

Economics Letters

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

On estimating market microstructure noise variance

Yingjie Dong [a,*], Yiu-Kuen Tse [b]

[a] Business School, University of International Business and Economics, Beijing, China
[b] School of Economics, Singapore Management University, Singapore

HIGHLIGHTS

- Propose modified estimates of market microstructure noise (MMN) variance based on Hansen and Lunde (2006).
- Report Monte Carlo results of comparison of different estimates of MMN variance.
- Report better performance of modified estimates at different frequencies.
- Report empirical estimates of MMN variance for some liquid NYSE stocks in 2010–2013.

ARTICLE INFO

Article history:
Received 21 September 2016
Received in revised form
5 November 2016
Accepted 9 November 2016
Available online 16 November 2016

JEL classification:
C410
G120

Keywords:
Microstructure noise
Noise-to-signal ratio
Realized variance
High-frequency data

ABSTRACT

We study the market microstructure noise-variance estimation of high-frequency stock prices. Based on the Hansen and Lunde (2006) approach, we propose estimates using subsampling method at different time scales. We conduct a Monte Carlo study to compare our method against others in the literature. Our results show that our proposed estimates have lower (absolute) mean error and root mean-squared error, and their performance is quite stable at different time scales.

© 2016 Elsevier B.V. All rights reserved.

1. Introduction

Market microstructure noise is a widely studied topic in the ultra high-frequency financial econometrics literature. Hansen and Lunde (2006) investigate the empirical properties of microstructure noise of the Dow Jones Industrial Average (DJIA) stocks. They find the noise-to-signal ratio (NSR) to be slightly below 0.1%.[1] Zhou (1996) and Zhang et al. (2005), among others, study the effect of the NSR on the optimal sampling frequencies of high-frequency data for the estimation of realized volatility. Awartani et al. (2009) examine the impact of microstructure noise on realized volatilities computed at different sampling frequencies. These studies highlight the importance of research on microstructure noise estimation.

Hansen and Lunde (2006) propose to estimate the noise variance using the realized variance, with trades sampled tick-by-tick. There are, however, drawbacks of these estimators, as their validity depends on the assumption of serially independent noise. Empirically we find this assumption to be violated when trades are selected at ultra-high frequencies (such as tick-by-tick), which may result from the behavior of economic agents (Diebold and Strasser, 2013). On the other hand, we find that the noise-variance estimates using trades sampled at low frequencies have substantially large estimation errors. Thus, careful choice of sampling method is important in balancing between inducing serially independent noise and avoiding large estimation errors.

In this paper, we re-examine the Hansen and Lunde (2006) approach and propose the use of subsampling to improve the noise-variance estimate. First, we advocate avoiding the use of tick-by-tick data, to circumvent the existence of serially dependent

* Correspondence to: Business School, University of International Business and Economics, 10 Huixin Dongjie, Beijing, 100029, China.
E-mail address: 02541@uibe.edu.cn (Y. Dong).

[1] NSR is defined as the ratio of microstructure-noise variance to daily stock return realized variance. The Hansen and Lunde (2006) results are based on data in 2000.

http://dx.doi.org/10.1016/j.econlet.2016.11.009
0165-1765/© 2016 Elsevier B.V. All rights reserved.

noise at ultra-high frequency. Second, we use the realized variance at different time scales, without relying on returns at low sampling frequencies, such as 30 min, as commonly used in the literature. Our Monte Carlo study shows that the root mean-squared error (RMSE) of our proposed method is lower than other methods currently available in the literature.

The rest of the paper is organized as follows. Section 2 outlines our modified noise-variance estimates and their asymptotic distributions. We report the results of our Monte Carlo study in Section 3 on comparing different noise-variance estimates. Section 4 draws conclusions. Supplementary materials are provided in the Online Appendix A.

2. Microstructure noise-variance estimates

Let $\{X_t\}$ denote a latent efficient log-price process in continuous time and $\{Y_t\}$ denote the observed log-price process. The noise process is $\{\varepsilon_t\}$, with $\varepsilon_t = Y_t - X_t$. We define the full grid containing all sampled time points (in sec) by $\mathcal{G} = \{t_0, t_1, t_2, \ldots, t_n\}$.

We follow Hansen and Lunde (2006), among others, and postulate the following two assumptions for the efficient log-price process $\{X_t\}$ and the noise process $\{\varepsilon_t\}$.

Assumption 1. The efficient log-price $\{X_t\}$ follows the Itô process $dX_t = \mu_t dt + \sigma_t dW_t$.[2]

Assumption 2. The noise process $\{\varepsilon_t\}$ is covariance stationary with mean 0, and auto-covariance function $\pi(s) \equiv E\left(\varepsilon_{t_i}\varepsilon_{t_{i-s}}\right)$ for any $t_i, t_{i-s} \in \mathcal{G}$ and $s \in \mathbb{Z}$.

Given the number of subgrids K, we define the subgrids $\mathcal{G}_K^k$, for $k = 1, \ldots, K$, as

$$\mathcal{G}_K^k = \{t_{k-1},\ t_{k-1+K},\ t_{k-1+2K}, \ldots,\ t_{k-1+(m_K^k-1)K}\}, \tag{1}$$

where $m_K^k = \max\{z \in \mathbb{N} : t_{k-1+(z-1)K} \le t_n\}$. Andersen et al. (2003) propose to estimate the integrated volatility IV by the realized variance of subgrid $\mathcal{G}_K^k$, defined as

$$RV_K^k = \sum_{t_i \in \mathcal{G}_K^k} (Y_{t_i} - Y_{t_{i-K}})^2.$$

Empirically we only observe $\{Y_t\}$, while the presence of the microstructure noise $\{\varepsilon_t\}$ influences the stock return volatility estimates. Under Assumption 1 and assumption of iid noise, the bias of RV_K^k as an estimate of IV is

$$\mathrm{E}[RV_K^k - \mathrm{IV}] = 2m_K^k \pi(0). \tag{2}$$

To alleviate the noise effect, Andersen et al. (2003) and Barndorff-Nielsen and Shephard (2002) suggest to sample transactions at low sampling frequency, such as 5 min or 20 min. A drawback of using only one subgrid $\mathcal{G}_K^k$, for $k \neq 1$, is that we lose information. To make use of all available observations in $\mathcal{G}$, Zhang et al. (2005) propose the subsampling estimate given by

$$\overline{RV}_K = \frac{1}{K}\sum_{k=1}^{K} RV_K^k. \tag{3}$$

Based on Assumptions 1 and 2, we have

$$\mathrm{E}\left[\overline{RV}_K - \mathrm{IV}\right] = 2\overline{m}_K\left[\pi(0) - \pi(K)\right] + 2\rho_K, \tag{4}$$

where $\rho_K = E\left(\sum_{k=1}^{K}\sum_{t_i \in \mathcal{G}_K^k} X_{t_i}\varepsilon_{t_i}\right)/K$ and $\overline{m}_K = \sum_{k=1}^{K} m_K^k/K$. By Cauchy–Schwarz inequality, $[\pi(0) - \pi(K)]$ is always non-negative. Thus, $E[\overline{RV}_K - \mathrm{IV}]$ is negative only when $\rho_K < -\overline{m}_K[\pi(0) - \pi(K)]$. Hence, $\overline{RV}_K$ underestimates IV only when the cross-correlation between the noise and latent returns is negative. Empirically, we calculate the volatility signature plots using NYSE transaction data from January 2010 to April 2013 of the top 40 market-capitalization stocks. For more than half of stocks the mean $\overline{RV}_K$ is smaller than the mean Realized Kernel (RK) estimate of Barndorff-Nielsen et al. (2008) at sampling frequency from tick-by-tick to 30 s. This result provides direct empirical evidence of the negative cross-correlation between the microstructure noise and the latent transaction returns at ultra-high frequency.[3]

Hansen and Lunde (2006) propose to estimate the microstructure noise variance $\pi(0)$ by

$$\widehat{\pi} = \frac{RV_K^k - \mathrm{IV}}{2m_K^k}, \tag{5}$$

where the integrated volatility IV is unknown and an unbiased estimate is needed. Hansen and Lunde (2006) use the first-order bias-corrected realized variance of Zhou (1996) to calculate IV. In this paper, we adopt the noise-robust RK estimate for IV.

The derivation of $\widehat{\pi}$ is based on Eq. (2) and its validity depends on the independent noise assumption. However, we find empirically that the noises are not serially independent at ultra-high frequencies. Thus, to use $\widehat{\pi}$, we may need to calculate RV_K^k using sparsely selected trades, instead of using the tick-by-tick data.[4] This may introduce two problems. First, when the noise variance is estimated by RV_K^k, with $K \neq 1$, we use transactions in one subgrid only, resulting in loss of information. Second, as shown in our Monte Carlo study, the finite sample performance of the noise-variance estimates deteriorates as the sampling frequency decreases.

To address the above issues, we propose to utilize all available transactions using subsampling method. First, we modify $\widehat{\pi}$ by replacing RV_K^k using $\overline{RV}_K$ to obtain

$$\overline{\pi} = \frac{\overline{RV}_K - \mathrm{IV}}{2\overline{m}_K}. \tag{6}$$

We denote this estimate by M1. If IV is known, the following proposition states the properties of M1.

Proposition 1. *Suppose the efficient log-price $\{X_t\}$ is an Itô process satisfying Assumption 1 and IV is known. If the noise process $\{\varepsilon_t\}$ is iid with $E(\varepsilon_t) = 0$, and $\{\varepsilon_t\}$ and $\{X_t\}$ are independent, with $E(\varepsilon_t^4) < \infty$, we have*

$$\sqrt{n}\left(\overline{\pi} - E(\varepsilon_t^2) \mid X\right) \xrightarrow{\mathcal{L}} \mathbf{N}\left(0, E(\varepsilon_t^4)\right). \tag{7}$$

The proof of Proposition 1 is in the Online Appendix A.

In implementation, we need to use an unbiased IV estimate for M1, and the estimation error introduced will distort the estimated noise variance. In this paper, we propose to estimate IV by a subsampling realized volatility estimate. Thus, we consider the following modified noise-variance estimate

$$\widetilde{\pi} = \frac{\overline{RV}_{K_1} - \overline{RV}_{K_2}}{2(\overline{m}_{K_1} - \overline{m}_{K_2})}, \tag{8}$$

where $K_1 < K_2$. We denote this estimate by M2. The following proposition states the properties of M2.

---

[2] Under Assumption 1, the integrated volatility of the price process over $[t_0, t_n]$ is $\mathrm{IV} = \int_{t_0}^{t_n} \sigma_s^2 ds$. According to standard arguments, the mean drift of the price process has negligible impact on the analysis of high-frequency data. Thus, without loss of generality we assume $\mu_t = 0$.

[3] We provide a representative volatility signature plot (stock JPM) in the Online Appendix A.

[4] Note that empirically both Hansen and Lunde (2006) and Zhang et al. (2005) use tick-by-tick transactions to compute RV_K^k in Eq. (5).

Download English Version:

https://daneshyari.com/en/article/5057874

Download Persian Version:

https://daneshyari.com/article/5057874

Daneshyari.com