Electronic Journal of Statistics
Vol. 7 (2013) 2032–2053
ISSN: 1935-7524
DOI: 10.1214/13-EJS834

Stationary bootstrapping realized volatility under market microstructure noise

Eunju Hwang and Dong Wan Shin*

Institute of Mathematical Sciences and Department of Statistics
Ewha University
e-mail: ehwang@ewha.ac.kr; shindw@ewha.ac.kr

Abstract: Large-sample validity is proved for stationary bootstrapping of a bias-corrected realized volatility under market microstructure noise, which enables us to construct a bootstrap confidence interval of integrated volatility. A finite-sample simulation shows that the stationary bootstrapping confidence interval outperforms existing ones which are constructed ignoring market microstructure noise or using asymptotic normality for the bias-corrected realized volatility.

MSC 2010 subject classifications: Primary 62G09, 60G10; secondary 62E20.
Keywords and phrases: Confidence interval, market microstructure noise, stationary bootstrap.

Received January 2013.

Contents

1 Introduction . . . 2032
2 Preliminary setup . . . 2034
  2.1 Existing theories . . . 2034
  2.2 Stationary bootstrap procedure . . . 2036
3 Asymptotic validity and confidence interval . . . 2037
4 Monte Carlo study . . . 2039
5 Conclusion . . . 2040
6 Proofs . . . 2043
Appendix A . . . 2050
Acknowledgements . . . 2051
References . . . 2051

1. Introduction

In a recent decade, volatility estimations and inferences have attracted much attention in the econometrical and statistical literature. Realized volatility based

*Corresponding author. Mailing Address: Dept. of Statistics, Ewha University, Seoul, Korea. Tel: +82-2-3277-2614; Fax: +82-2-3277-3606; *E-mail address*: shindw@ewha.ac.kr (D.W. Shin)

on high frequency data, as an estimator of the integrated volatility, is very popular and has been actively researched in both academia and financial markets. The realized volatility is defined as the sum of squared intraday returns, and is consistent in the ideal situation of no market micro structure noise under some general condition, (see e.g. [4, 6, 17, 26] etc.). In particular, [4] and [6] made significant contributions in theory of realized volatility by establishing asymptotic theorems including central limit theorems of bipower variation in financial econometrics.

In practice, the prices recorded at the high frequency are contaminated by market microstructure noise. The presence of market microstructure noise in high frequency financial data complicates volatility estimation and causes some statistically serious problems such as bias problem and inconsistency, (see [3] and [29]). Recently, many econometricians and statisticians have studied the effects of market microstructure noise and developed asymptotic validity theories for the realized volatility under market microstructure noise. [13] studied empirical properties of market microstructure noise and analyzed its implications for the realized volatility. In particular, they established asymptotic normality for the bias-corrected estimator of [30]. The estimator of [13] and [30] incorporates the first-order autocovariance, which amounts to a bias correction that works in the same way that robust covariance estimators achieve their consistency. [3] provided a general treatment of the effect of market microstructure noise on realized volatility estimates, and specifically considered both asymptotic and finite sample effects of noise. [2] analyzed the impact of time series dependence in market microstructure noise on the properties of estimators of the integrated volatility, based on data sampled at frequencies high enough for that noise to be a dominant consideration. [24] provided theoretical reviews and comparisons of high-frequency based volatility estimators and the impact of different types of noise.

Some bootstrapping methods are developed for realized volatility by [9] and [11]. In the absence of market microstructure noise, [11] proposed bootstrap methods for a class of nonlinear transformation of realized volatility including the raw version and its logarithmic transformation, by means of i.i.d. bootstrap and wild bootstrap, and [9] studied bootstrap methods for statistics that are functions of multivariate high frequency returns such as realized regression coefficients, covariances and correlations.

This paper considers a bootstrap approach for realized volatility under market microstructure noise by adopting the stationary bootstrap of [25]. The stationary bootstrap, which is an extension of the moving block bootstrap or circular block bootstrap by allowing the block length to be a geometric random variable, has been widely used as a powerful resampling technique for approximating the sampling distribution of non-parametric estimators. Recent applications of stationary bootstrapping are found in [14, 15, 20], regarding nonparametric estimation and in [23, 27, 28] regarding nonstationary time series analysis. Some new properties of the stationary bootstrap are established by [10, 12, 16, 21].

We verify the first order asymptotic validity of the stationary bootstrap method for the bias-corrected realized volatility as a consistent estimator of the integrated volatility. This allows us to construct a stationary bootstrap confidence interval (CI) of integrated volatility, which improves other existing CIs by [4] based on asymptotic normality of uncorrected realized volatility, by [11] based on i.i.d. bootstrapping of uncorrected realized volatility, and by [13] based on asymptotic normal theory of bias-corrected realized volatility.

For cases of market microstructure noises, the CIs of [4] and [11] are not valid while the proposed CI and the CI of [13] are valid. The proposed bootstrap CI has an advantage over the CI based on asymptotic normality as [13] in that, unlike the latter, the former does not require estimates of market microstructure noise variance to which coverage probability of the latter is very sensitive.

A Monte-Carlo experiment shows that the proposed stationary bootstrapping CI has better coverage performance than the existing ones.

The remaining of the paper is organized as follows. In Section 2 we describe the setup and the existing theory that will be used in this work, and review the stationary bootstrap procedure. Main theoretical results and construction of a bootstrap confidence interval are presented in Section 3, a Monte-Carlo result is given in Section 4, and a conclusion is provided in Section 5 while technical results and proofs are found in Section 6.

2. Preliminary setup

2.1. Existing theories

We assume that the latent log-price process $\{\tilde{S}(t) : t \geq 0\}$ follows a continuous time process $d\tilde{S}(t) = \mu(t)dt + \sigma(t)dW(t)$ where $\mu(t)$ is the drift process, $\sigma(t)$ is a volatility process, and $W(t)$ is the standard Brownian motion. Let $S(t)$ denote the observable log-price process, and thus the noise process is given by

$$u(t) = S(t) - \tilde{S}(t).$$

We are interested in estimating the integrated volatility (IV) over a fixed time interval $[0, 1]$ defined by

$$IV = \int_0^1 \sigma^2(s)\, ds.$$

In order to define an estimator of IV, we partition the interval $[0, 1]$ into n subintervals; $0 = t_0 < t_1 < \cdots < t_n = 1$, and define the *intraday returns* by

$$\tilde{r}_i = \tilde{S}(t_i) - \tilde{S}(t_{i-1}), \quad i = 1, 2, \ldots, n.$$

The increments in S and u are defined similarly by

$$r_i = S(t_i) - S(t_{i-1}), \quad e_i = u(t_i) - u(t_{i-1}), \qquad i = 1, 2, \ldots, n.$$

A consistent estimator of the integrated volatility is the realized volatility (RV) defined by the sum of squared high frequency returns

$$\widetilde{RV}_{(n)} = \sum_{i=1}^{n} \tilde{r}_i^2.$$

Since $\tilde{S}$ is latent, $\widetilde{RV}_{(n)}$ is not a feasible estimator. The realized volatility of process S defined by

$$RV_{(n)} = \sum_{i=1}^{n} r_i^2$$

is observable but suffers from the bias problem and inconsistency. To overcome these problems, [30] proposed the bias-corrected estimator incorporated with the empirical first-order autocovariance given by

$$RV_{(n)}^{AC_1} = \sum_{i=1}^{n} r_i^2 + \sum_{i=2}^{n} r_{i-1} r_i + \sum_{i=1}^{n-1} r_i r_{i+1}. \tag{2.1}$$

[13] and [31] reconsidered the estimator $RV_{(n)}^{AC_1}$ of [30]. In particular, [13] analyzed a special case of their work under simplistic assumptions of the independent noise, and discussed when its theoretical results provide reasonable approximations. According to the empirical analysis of [13] the noise may be ignored when intraday returns are sampled at relatively low frequencies, such as 20-minute sampling. When intraday returns are sampled every 15 ticks or so, assumption of independent noise seems to be reasonable.

The case with independent market microstructure noise has been dealt with by several authors including [1, 3, 7, 29]. In a general setting of the noise, [19] studied a theoretical comparison between IV and RV by characterizing the noise term and by quantifying the importance of the noise. The case with dependent market microstructure noise has been discussed by [2], where the noise process is assumed to be strong mixing. [13] and [22] considered a finite time-dependent noise process to analyze bias-corrected RVs.

The works mentioned above are about the case of exogenous noise process, that is, the noise $u(t)$ and the process $\tilde{S}(t)$ are independent, whereas [5] and [18] dealt with the case of endogenous noise. In particular, [5] have studied realized kernels which is robust to endogenous noise.

To derive our main goal of this paper, we need the following assumptions, which are based on the special case of [13].

Assumption 1. The efficient price process satisfies $d\tilde{S}(t) = \sigma(t)dW(t)$, where $W(t)$ is a standard Brownian motion, $\sigma(t)$ is a time-varying (random) function that is independent of W, and $\sigma^2(t)$ is Lipschitz (almost surely).

Assumption 2. The noise process u is independent of the process $\tilde{S}$, and $u(s)$ and $u(t)$ are independent of each other for $s \neq t$. Let $\omega^2 = E[u(t)]^2 < \infty$ for all t.

In [13], we find several good arguments for analyzing the properties of the RV and related quantities under the independent noise, in spite of dismissing this form of noise as an accurate description of the noise in the data. The independent noise assumption makes the analysis tractable and provides valuable insight into the issues related to market microstructure noise.

[13] showed unbiasedness and asymptotic normality of $RV_{(n)}^{AC_1}$ of [30] in (2.1), which are given in Proposition 2.1 below under Assumptions 1 and 2 of no drift and independent noise.

Proposition 2.1. *([13]) Under Assumptions 1 and 2, we have*

(a) $E(RV_{(n)}^{AC_1}) = IV$,

(b)

$$Var\left(RV_{(n)}^{AC_1}\right) = 8\omega^4 n + 8\omega^2 \sum_{i=1}^{n} \sigma_{i,n}^2 - 6\omega^4 + 6\sum_{i=1}^{n} \sigma_{i,n}^4 + O(n^{-2})$$

where $\sigma_{i,n}^2 = \int_{t_{i-1}}^{t_i} \sigma^2(s)ds$, $i = 1, \ldots, n$, *and*

(c)

$$\frac{RV_{(n)}^{AC_1} - IV}{\sqrt{8\omega^4 n}} \xrightarrow{\text{d}} N(0,1) \quad as \quad n \to \infty.$$

An important result of Proposition 2.1 is that $RV_{(n)}^{AC_1}$ is unbiased for the IV at any sampling frequency. Also, a remarkable result of Proposition 2.1 is that the bias-corrected estimator $RV_{(n)}^{AC_1}$ has a smaller asymptotic variance as $n \to \infty$ than the unadjusted estimator $RV_{(n)}$, (see Lemma 2 of [13]). [13] noted that the asymptotic results in Proposition 2.1 are more useful than those of the unadjusted estimator $RV_{(n)}$, for example, the results of Proposition 2.1 are used by [3] and [29] to estimate the variance of the noise, ω^2. Based on the asymptotic results in Proposition 2.1, our main results will be established via stationary bootstrapping.

In other cases with dependent and endogenous noises as in [3] and [13], bootstrapping estimators need to be proposed, but it is out of our scope in this work and remains as a further study.

2.2. Stationary bootstrap procedure

The stationary bootstrapping of $\{r_1, \ldots, r_n\}$ is described. First we define a new time series $\{r_{ni} : i \geq 1\}$ by a periodic extension of the observed data set as follows. For each $i \geq 1$, define $r_{ni} := r_j$ where j is such that $i = qn+j$ for some q. The sequence $\{r_{ni} : i \geq 1\}$ is obtained by wrapping the data $r_1, \ldots, r_n$ around a circle, and relabelling them as $r_{n1}, r_{n2}, \ldots$. Next, for a positive integer ℓ, define the blocks $B(i,\ell), i \geq 1$ as $B(i,\ell) = \{r_{ni}, \ldots, r_{n(i+\ell-1)}\}$ consisting of ℓ observations starting from r_{ni}. Bootstrap observations under the stationary bootstrap method are obtained by selecting a random number of blocks from collection $\{B(i,\ell) : i \geq 1, \ell \geq 1\}$. To do this, we generate random variables $I_1, \ldots, I_n$ and $L_1, \ldots, L_n$ as follows: (i) $I_1, \ldots, I_n$ are i.i.d. discrete uniform on $\{1, \ldots, n\}$: $P(I_1 = i) = \frac{1}{n}$, for $i = 1, \ldots, n$, (ii) $L_1, \ldots, L_n$ are i.i.d. random variables having the geometric distribution with a parameter $p \in (0,1)$: $P(L_1 = \ell) = p(1-p)^{\ell-1}$ for $\ell = 1, 2, \ldots$, where $p = p(n)$ depends on the sample size n and (iii) the collections $\{I_1, \ldots, I_n\}$ and $\{L_1, \ldots, L_n\}$ are independent.

For notational simplicity, we suppress dependence of the variables $I_1, \ldots, I_n$, $L_1, \ldots, L_n$ and of the parameter p on n. We assume that $p = p(n)$ goes to 0 as $n \to \infty$. Under the stationary bootstrap the block length variables $L_1, \ldots, L_n$ are random and the expected block length EL_1 is p^{-1}, which tends to ∞ as $n \to \infty$. Now, a pseudo-time series $r_1^*, \ldots, r_n^*$ is generated in the following way. Let $\tau = \inf\{k \geq 1 : L_1 + \cdots + L_k \geq n\}$. Then select τ blocks $B(I_1, L_1), \ldots, B(I_\tau, L_\tau)$. Note that there are $L_1 + \cdots + L_\tau$ elements in the resampled blocks $B(I_1, L_1), \ldots, B(I_\tau, L_\tau)$. Arranging these elements in a series and deleting the last $L_1 + \cdots + L_\tau - n$ elements, we get the bootstrap observations $r_1^*, \ldots, r_n^*$. Conditional on $\{r_1, \ldots, r_n\}$, the process $\{r_t^*, t = 1, 2, \ldots\}$ is stationary. In the following, P^*, E^*, and Var^* denote the conditional probability, expectation, and variance, respectively, given $r_1, \ldots, r_n$.

3. Asymptotic validity and confidence interval

In this section we propose the stationary bootstrap realized volatility and establish its asymptotic validity. The asymptotic theory enables us to construct a bootstrap confidence interval. In addition to Assumptions 1 and 2 above, we need weak dependence conditions on the data $\{r_1, r_2, \ldots, r_n\}$ as follows.

Assumption 3. Let $\{r_1, r_2, \ldots, r_n\}$ satisfy $E|r_i|^{4+2\delta} < \infty$ for some $\delta > 0$, and $\frac{1}{n}\sum_{i=1}^{n} iCov(g_1(r_1, r_2),\ g_2(r_{i+1}, r_{i+2})) \to 0$ as $n \to \infty$, where functions $g_1,\ g_2 : R^2 \to R$ are either $g(x, y) = x^2$ or $g(x, y) = xy$.

Now we define the stationary bootstrap version of $RV_{(n)}^{AC_1}$ by

$$RV_{(n)}^{AC_1*} = \sum_{i=1}^{n} r_i^{*2} + \sum_{i=2}^{n} r_{i-1}^* r_i^* + \sum_{i=1}^{n-1} r_i^* r_{i+1}^*$$

and we state main asymptotics results for the stationary bootstrap realized volatility.

Theorem 3.1. *We suppose Assumptions 1, 2 and 3 above. If parameter p of geometric distribution of random block length in the stationary bootstrap procedure is chosen so that $np^2 \to \infty$, then*

(a)

$$\frac{1}{n}\left|Var^*\left(RV_{(n)}^{AC_1*}\right) - Var\left(RV_{(n)}^{AC_1}\right)\right| \xrightarrow{\text{p}} 0,$$

(b)

$$\sup_{x\in R}\left|P^*\left(\frac{RV_{(n)}^{AC_1*} - E^*(RV_{(n)}^{AC_1*})}{\sqrt{8\omega^4 n}} \leq x\right) - P\left(\frac{RV_{(n)}^{AC_1} - IV}{\sqrt{8\omega^4 n}} \leq x\right)\right| \xrightarrow{\text{p}} 0$$

as $n \to \infty$.

Theorem 3.2. *We suppose Assumptions 1, 2 and 3 above. If parameter p of geometric distribution of random block length in the stationary bootstrap procedure satisfies $p = c \cdot n^{-\frac{\delta-2\rho}{2(2+\delta)}}$ for some $c > 0$, $0 < \rho < \delta/2$ and $0 < \delta < 1$, then*

(a)

$$\frac{1}{n}\left|Var^*\left(RV_{(n)}^{AC_1*}\right) - Var\left(RV_{(n)}^{AC_1}\right)\right| \xrightarrow{\text{a.s}} 0,$$

(b)

$$\sup_{x\in R}\left|P^*\left(\frac{RV_{(n)}^{AC_1*} - E^*(RV_{(n)}^{AC_1*})}{\sqrt{8\omega^4 n}} \le x\right) - P\left(\frac{RV_{(n)}^{AC_1} - IV}{\sqrt{8\omega^4 n}} \le x\right)\right| \xrightarrow{\text{a.s}} 0$$

as $n \to \infty$.

The above Theorem 3.1 enables us to construct a bootstrap confidence interval of the integrated volatility. According to Theorem 3.1, we have

$$\begin{aligned} &P^*\left[RV_{(n)}^{AC_1*} \le x\right] \\ &\quad = P^*\left[\left(RV_{(n)}^{AC_1*} - E^*(RV_{(n)}^{AC_1*})\right)\Big/\sqrt{8\omega^4 n} \le \left(x - E^*(RV_{(n)}^{AC_1*})\right)\Big/\sqrt{8\omega^4 n}\right] \\ &\quad \cong P\left[\left(RV_{(n)}^{AC_1} - IV\right)\Big/\sqrt{8\omega^4 n} \le \left(x - E^*(RV_{(n)}^{AC_1*})\right)\Big/\sqrt{8\omega^4 n}\right] \\ &\quad = P\left[RV_{(n)}^{AC_1} - \left(x - E^*(RV_{(n)}^{AC_1*})\right) \le IV\right]. \end{aligned}$$

Therefore, observing that $0.95 = P^*[\,q^*_{0.025} \le RV_{(n)}^{AC_1*} \le q^*_{0.975}\,] \cong$

$$P\left[RV_{(n)}^{AC_1} - \left(q^*_{0.975} - E^*(RV_{(n)}^{AC_1*})\right) \le IV \le RV_{(n)}^{AC_1} - \left(q^*_{0.025} - E^*(RV_{(n)}^{AC_1*})\right)\right],$$

we construct, for example, a 95% bootstrap confidence interval for IV:

$$\left[\,RV_{(n)}^{AC_1} + \overline{RV}_{(n)}^{AC_1*} - q^*_{0.975},\ \ RV_{(n)}^{AC_1} + \overline{RV}_{(n)}^{AC_1*} - q^*_{0.025}\,\right] \tag{3.1}$$

where $q^*_{0.025}$ and $q^*_{0.975}$ are the 2.5% and 97.5% quantiles, respectively, of the B bootstrap replications $RV_{(n)}^{AC_1*}(b)$, $b = 1, \ldots, B$, and

$$\overline{RV}_{(n)}^{AC_1*} = B^{-1}\sum_{b=1}^{B} RV_{(n)}^{AC_1*}(b).$$

One of the main advantages of the proposed stationary bootstrap method is that it does not require estimation of the noise variance, ω^2, as seen in the bootstrap confidence interval in (3.1) above. On the other hand, the confidence interval of [13] based on Proposition 2.1(c) requires estimation of ω^2. This fact gives the former confidence interval stabler finite sample coverage probability than the latter confidence interval as investigated in the following section.

4. Monte Carlo study

This section compares finite sample empirical coverage probability of the proposed stationary bootstrapping confidence interval with those of existing ones in the presence of both conditional heteroscedasticity and market microstructure noise.

For the comparison, we consider the same data generating process studied by [11] given by $d\tilde{S}(t) = \mu dt + \sigma(t)(\rho_1 dW_1(t) + \rho_2 dW_2(t) + \sqrt{1-\rho_1^2-\rho_2^2}dW_3(t))$ with the following two volatility processes: GARCH(1, 1) diffusion, $d\sigma^2(t) = 0.035(0.636 - \sigma^2(t))dt + 0.144\sigma^2(t)dW_1(t)$; two-factor affine diffusion, $\sigma(t) = \exp(-1.2 + 0.04\sigma_1^2(t) + 1.5\sigma_2^2(t))$, $d\sigma_1^2(t) = -0.00137\sigma_1^2(t)dt + dW_1(t)$, $d\sigma_2^2(t) = -1.386\sigma_2^2(t)dt + (1 + 0.25\sigma_2^2(t))dW_2(t)$, where $W_1(t), W_2(t), W_3(t)$ are independent standard Brownian motions. The observed prices are $p(t_i) = \tilde{p}(t_i) + u_i, i = 1, \ldots, n$, where $t_i = i/n$ and u_i are iid N$(0,\, \omega^2)$ noise independent of the latent price process $\tilde{S}(t)$. The observed return $r_i = \tilde{r}_i + e_i$ is subject to noise $e_i = u_i - u_{i-1}$.

For n, three values 10, 100, 1000 are considered, which correspond to sampling periods of 0.6 hour, 3.6 minute, 21.6 second, respectively, in a trading day. For (μ, ρ_1, ρ_2), two cases are considered: $\{(0,\ 0,\ 0),\ (0.0314,\ -0.576,\ 0)\}$ for GARCH(1, 1) model and $\{(0, 0, 0),\ (0.030, -0.3, -0.3)\}$ for two-factor diffusion model. The case with $(\mu, \rho_1, \rho_2) = 0$ corresponds to a model with no drift and no leverage and the other case correspond to a model with drift and leverage. For the noise variance ω^2, four cases $\omega^2 = 0, 0.01, 0.02, 0.03$ are considered. The case with $\omega^2 = 0$ correspond to a case with no noise. Magnitude of noise increases as n or i_u increase because the noise-to-signal ratio is given by $\lambda = \sum_{i=1}^{n} var(e_i)/E(IV) = 2n\omega^2/E(IV)$.

The normal errors are simulated using standard normal errors generated by RNNOA, a FORTRAN subroutine in IMSL. The volatility process is initiated with $\sigma^2(0) = 0.636$ for the GARCH(1, 1) diffusion and $\sigma_1^2(0) = \sigma_2^2(0) = 0.1$ for the two-factor affine diffusion. The integrated volatility $IV = \int_0^1 \sigma^2(s)ds$ is approximated by the noise-free realized volatility $\widetilde{RV}_{(n_0)} = \sum_{i=1}^{n_0} \tilde{r}_i^2$ computed with large $n_0 = 100,000$.

Four confidence intervals of nominal coverage probability 95% are compared. The first one, denoted by $NA1$, is based on the normal approximation theory of [4] and [6] for $RV_{(n)}$: $RV_{(n)} \pm 1.96\sqrt{2nRQ_{(n)}}$, where $RQ_{(n)} = (n/3)\sum_{i=1}^{n} r_i^4$. The second one, denoted by IB, is based on i.i.d. bootstrapping of [11]: $RV_{(n)} \pm t^*_{0.95}\sqrt{2nRQ_{(n)}}$, $t^*_{0.95}$ is the 95% percentile of the B bootstrap values of $\sqrt{n}(RV^*_{(n)} - RV_{(n)})/\sqrt{Q^*_{(n)}}$, where $Q^*_{(n)} = n\sum_{i=1}^{n} r_i^{*4} - (RV^*_{(n)})^2$. Next one, denoted by $NA2$, is based on the normal approximation theory of [13] for $RV^{AC_1}_{(n)}$: $RV^{AC_1}_{(n)} \pm 1.96\sqrt{8n\hat{\omega}^4}$, where $\hat{\omega}^2 = \sum_{i=1}^{n} r_i^2/(2n)$. The last one, denoted by SB, is based on stationary bootstrapping of $RV^{AC_1}_{(n)}$: $[\, RV^{AC_1}_{(n)} + \overline{RV}^{AC_1*}_{(n)} - q^*_{0.975},\ RV^{AC_1}_{(n)} + \overline{RV}^{AC_1*}_{(n)} - q^*_{0.025}\,]$ as given by (3.1) in Section 3. The first two intervals $NA1$ and IB are valid for cases without market micro structure noise, $\omega^2 = 0$. The last two intervals $NA2$ and SB are valid for both cases with and without market micro structure noise, $\omega^2 \geq 0$.

The bootstrap confidence intervals by IB and SB are computed using $B = 1,000$ bootstrap replications. For SB, we use the block-length parameter $p = 0.2i_p(n/100.)^{-1/3}$, $i_p = 1, 2$. SB with $i_p = 1, 2$ are denoted by SB_1, SB_2, respectively. The noise-to-signal ratio, $\lambda = 2n\omega^2/E(IV)$ is also computed by approximating $E(IV)$ by the average of the 10, 000 values of IV.

Table 1 reports empirical coverage probabilities computed using 10, 000 independent replications. The table shows us that the proposed confidence interval SB has the best overall coverage performance. We observe that SB tends to have uniformly better coverage than $NA2$. For both noise-free cases and noise cases, SB based on stationary bootstrapping of $RV_{(n)}^{AC_1}$ provides more stable confidence intervals than $NA2$ based on normal limit theory of $RV_{(n)}^{AC_1}$.

The poor performance of NA2 for small n or for small λ dues to poor estimation of ω^2 as observed by [13] that small noise causes the estimator $\hat{\omega}^2 = \sum_{i=1}^{n} r_i^2/2n$ to have a large bias unless n is very large. The proposed bootstrap interval SB, requiring no estimation of ω^2, has stabler coverage than NA2 for all n and λ considered here.

We also observe that SB based on $RV_{(n)}^{AC_1}$ is better than $NA1$ and IB based on $RV_{(n)}$. Ignoring noise, $NA1$ and IB have serious problems for large n and $\omega^2 \neq 0$. The confidence intervals $NA1$ and IB fail to work having empirical coverage values below 10% for non-ignorable noise cases of $n = 1000$ and $\omega^2 = 0.01, 0.02, 0.03$. On the other hand, the proposed confidence interval SB has stable coverage probability.

Table 2 reports average lengths of the confidence intervals. For the GARCH(1, 1) diffusion, the average lengths of the 5 confindence intervals are not much different. For the two-factor affine diffusion, the average lengths are very different: IB has the largest length and NA2 has the smallest length; the stabler coverage probability of SB than NA2 is obtained by the cost of enlarged average length. The intervals $NA2$ and SB with coverage probabilities closer to the nominal coverage 95% tend to have larger average lengths than those with smaller coverage probabilities. The intervals $NA1$ and IB show similar aspects for the cases of $n = 10$ or $\lambda = 0$ in which bias of $RV_{(n)}$ is negligible. For the other cases of (n, λ) in which bias of $RV_{(n)}$ is not negligible, even though $NA1$ and IB have average lengths not much smaller than those of ($NA2$, SB), owning to the bias, $NA1$ and IB have seriously small coverage probabilities close to 0.

5. Conclusion

Asymptotic normality is established for stationary bootstrapping of a bias-corrected realized volatility in the presence of market microstructure noise. Applying stationary bootstrapping to the bias-corrected realized volatility, we construct the stationary bootstrapping confidence interval of the integrated volatil-

TABLE 1. *Coverage probabilities (%) of confidence intervals*

			no drift & no leverage						drift & leverage					
model	n	ω^2	λ	$NA1$	IB	$NA2$	SB_1	SB_2	λ	$NA1$	IB	$NA2$	SB_1	SB_2
1	10	0.00	.000	85.2	92.9	74.4	83.0	86.8	.000	85.1	92.8	73.6	82.7	86.4
1	10	0.01	.003	85.3	93.1	74.4	83.2	86.9	.003	85.2	92.8	73.6	82.5	86.3
1	10	0.02	.013	85.8	93.4	74.4	83.3	86.9	.013	85.6	93.1	74.0	83.2	86.6
1	10	0.03	.028	86.4	93.6	75.2	83.9	87.3	.028	85.7	93.3	74.8	83.6	87.1
1	100	0.00	.000	93.6	94.3	74.1	92.3	93.1	.000	94.0	94.6	73.1	91.9	92.9
1	100	0.01	.031	95.3	95.5	74.5	92.3	93.6	.031	94.9	95.7	74.1	91.9	93.1
1	100	0.02	.126	91.9	93.2	77.4	93.1	94.3	.126	92.1	92.9	77.0	92.6	94.1
1	100	0.03	.283	69.3	74.2	81.1	94.1	95.3	.283	69.6	74.4	80.4	93.7	94.9
1	1000	0.00	.000	94.9	94.1	74.7	94.4	94.5	.000	94.9	94.3	74.6	94.8	94.9
1	1000	0.01	.315	.0	.0	81.8	95.3	95.7	.315	.0	.0	81.8	95.1	95.5
1	1000	0.02	1.259	.0	.0	90.6	95.8	96.8	1.259	.0	.0	90.3	95.7	96.5
1	1000	0.03	2.832	.0	.0	93.5	96.2	97.1	2.832	.0	.0	93.6	96.4	97.2
2	10	0.00	.000	77.9	89.7	65.5	75.9	79.2	.000	78.0	89.4	64.9	76.1	79.4
2	10	0.01	.000	79.0	89.9	66.0	76.5	79.8	.000	79.0	89.8	66.0	76.7	79.8
2	10	0.02	.001	82.0	91.1	67.8	78.1	81.6	.001	81.6	91.3	67.8	78.1	81.8
2	10	0.03	.003	84.4	92.7	70.5	79.9	83.6	.003	84.1	92.6	70.4	79.9	83.8
2	100	0.00	.000	90.8	93.0	59.4	88.7	89.1	.000	90.6	92.9	59.5	88.9	89.2
2	100	0.01	.004	83.7	88.3	66.3	90.3	91.1	.004	83.9	88.2	65.6	89.9	90.6
2	100	0.02	.014	44.9	51.9	75.5	92.7	93.6	.014	45.0	51.8	74.8	92.2	93.1
2	100	0.03	.032	23.7	29.1	81.6	94.4	95.2	.032	23.8	29.0	80.8	94.1	95.0
2	1000	0.00	.000	94.3	94.1	57.8	96.9	95.8	.000	94.2	93.8	58.2	96.9	95.8
2	1000	0.01	.036	6.2	6.5	80.9	97.1	97.0	.035	6.1	6.4	81.1	97.0	96.7
2	1000	0.02	.143	1.6	1.8	89.6	96.8	97.1	.141	1.6	1.7	89.3	96.7	96.9
2	1000	0.03	.321	.8	.9	92.3	96.4	96.9	.317	.8	.9	92.0	96.5	97.2

Note: Model 1 is GARCH(1, 1) diffusion and model 2 is two-factor affine diffusion. Nominal coverage = 95%; number of replications = 10, 000; number of bootstrap repetitions B = 1, 000.

Table 2. *Average lengths of confidence intervals*

			no drift & no leverage						drift & leverage					
model	n	ω^2	λ	$NA1$	IB	$NA2$	SB_1	SB_2	λ	$NA1$	IB	$NA2$	SB_1	SB_2
1	10	0.00	.000	1.02	2.62	1.13	1.73	1.97	.000	1.02	2.60	1.13	1.73	1.97
1	10	0.01	.003	1.02	2.63	1.13	1.73	1.98	.003	1.02	2.61	1.13	1.73	1.97
1	10	0.02	.013	1.03	2.64	1.14	1.74	2.00	.013	1.03	2.63	1.14	1.74	1.99
1	10	0.03	.028	1.05	2.67	1.16	1.76	2.02	.028	1.04	2.66	1.16	1.76	2.02
1	100	0.00	.000	.35	.39	.35	.59	.61	.000	.35	.39	.35	.60	.61
1	100	0.01	.031	.36	.40	.37	.60	.62	.031	.36	.40	.37	.61	.62
1	100	0.02	.126	.39	.43	.40	.63	.66	.126	.39	.43	.40	.63	.66
1	100	0.03	.283	.45	.49	.45	.68	.72	.283	.45	.49	.45	.69	.73
1	1000	0.00	.000	.11	.11	.11	.19	.19	.000	.11	.11	.11	.19	.19
1	1000	0.01	.315	.15	.15	.15	.22	.22	.315	.15	.15	.15	.22	.22
1	1000	0.02	1.259	.25	.25	.25	.31	.32	1.259	.25	.25	.25	.31	.32
1	1000	0.03	2.832	.43	.43	.43	.48	.50	2.832	.43	.43	.43	.48	.50
2	10	0.00	.000	9.28	1134	5.68	12.93	14.95	.000	8.35	1067	5.30	12.03	13.37
2	10	0.01	.000	9.28	1174	5.68	12.45	14.98	.000	8.35	1072	5.30	11.69	13.70
2	10	0.02	.001	9.29	1128	5.69	12.43	14.97	.001	8.36	1117	5.31	11.33	13.44
2	10	0.03	.003	9.30	1137	5.71	12.41	15.00	.003	8.37	1182	5.33	11.49	13.28
2	100	0.00	.000	9.10	103.41	2.69	13.23	12.90	.000	9.19	142.71	2.62	11.06	10.90
2	100	0.01	.004	9.11	63.29	2.70	13.21	12.85	.004	9.20	63.90	2.64	10.75	10.69
2	100	0.02	.014	9.14	61.34	2.74	13.73	12.77	.014	9.22	68.11	2.67	11.01	10.98
2	100	0.03	.032	9.18	59.62	2.79	13.17	13.35	.032	9.27	57.88	2.72	11.26	11.21
2	1000	0.00	.000	5.48	12.15	.92	9.35	8.84	.000	5.62	17.15	.93	8.76	7.75
2	1000	0.01	.036	5.51	13.08	.96	9.55	8.66	.035	5.64	16.70	.97	8.45	7.97
2	1000	0.02	.143	5.61	13.37	1.06	9.52	9.10	.141	5.74	18.44	1.07	8.41	7.94
2	1000	0.03	.321	5.77	13.73	1.24	9.98	9.03	.317	5.91	18.60	1.25	8.78	8.00

Note: Model 1 is GARCH(1, 1) diffusion and model 2 is two-factor affine diffusion. Nominal coverage = 95%; number of replications = 10, 000; number of bootstrap repetitions B = 1, 000.

ity. A Monte-Carlo experiment reveals advantages of the proposed stationary bootstrapping confidence interval over existing methods based on asymptotic normality established by [4] and [13] or those based on i.i.d. bootstrapping of uncorrected realized volatility by [11].

6. Proofs

In the proofs, $X_n \xrightarrow{\text{p}} X$, $X_n \xrightarrow{\text{a.s}} X$ and $X_n \xrightarrow{\text{d}} X$ mean that X_n converges to X in probability, almost surely, and in distribution, respectively. Also, $X_n \xrightarrow{\text{p}^*} X$, $X_n \xrightarrow{\text{a.s}^*} X$ and $X_n \xrightarrow{\text{d}^*} X$ mean that X_n converges to X in probability, almost surely, in distribution, respectively, conditionally given $\{r_1, \dots, r_n\}$, respectively. Proof of Theorem 3.1(b) is given first because it is the main part. Proof of Theorem 3.1(a) is given after proofs of Lemmas 6.1, 6.2 below.

Proof of Theorem 3.1(b). Given Proposition 2.1(c), it suffices to show that

$$\frac{1}{\sqrt{n}}\left(RV_{(n)}^{AC_1*} - E^*(RV_{(n)}^{AC_1*})\right) \xrightarrow{\text{d}^*} N(0, 8\omega^4). \tag{6.1}$$

Let

$$Q_{n,\tau}^* = \sum_{i=1}^{s_\tau} r_i^{*2} + 2\sum_{i=1}^{s_\tau - 1} r_i^* r_{i+1}^*, \quad U_{n,\tau} = \sum_{j=1}^{\tau}\left(S_{I_j,L_j} + 2\cdot T_{I_j,L_j}\right),$$

where $s_\tau = L_1 + \cdots + L_\tau$, $S_{i,\ell} = \sum_{j=i}^{i+\ell-1} r_{nj}^2$, $T_{i,\ell} = \sum_{j=i}^{i+\ell-1} r_{nj}r_{n(j+1)}$. Then, thanks to Lemma 6.1, $RV_{(n)}^{AC_1*}$ is approximated by the average of τ conditionally i.i.d. random variables $S_{I_j,L_j} + 2\cdot T_{I_j,L_j}, j = 1, \dots, \tau$ whose conditional expectation is approximately $E^*(RV_{(n)}^{AC_1*}) \cong RV_{(n)}^{AC_1}$ by Lemma 6.2(b) below and conditional variance is approximately $8\omega^4$ by Theorem 3.1(b). Then the classical central limit theorem gives the result Lemma 6.2(a) and hence (6.1). More formally, in order to verify (6.1), we will show that $\frac{1}{\sqrt{n}}|Q_{n,\tau}^* - U_{n,\tau}| \xrightarrow{\text{p}^*} 0$, $\frac{1}{\sqrt{n}}|Q_{n,\tau}^* - RV_{(n)}^{AC_1*}| \xrightarrow{\text{p}^*} 0$ in Lemma 6.1 below and that

$$\frac{1}{\sqrt{n}}\left(U_{n,\tau} - RV_{(n)}^{AC_1}\right) \xrightarrow{\text{d}^*} N(0, 8\omega^4), \quad \frac{1}{\sqrt{n}}\left(RV_{(n)}^{AC_1} - E^*(RV_{(n)}^{AC_1*})\right) \xrightarrow{\text{p}} 0$$

in Lemma 6.2 below. □

Lemma 6.1. *Under the same assumptions as in Theorem 3.1, we have*

$$(a)\ \ \frac{1}{\sqrt{n}}|Q_{n,\tau}^* - U_{n,\tau}| \xrightarrow{\text{p}^*} 0, \quad (b)\ \ \frac{1}{\sqrt{n}}\left|Q_{n,\tau}^* - RV_{(n)}^{AC_1*}\right| \xrightarrow{\text{p}^*} 0.$$

Proof. Denoting $I_{\tau+1} = 0$, we first observe that

$$\begin{aligned}
&\frac{1}{\sqrt{n}}|Q^*_{n,\tau} - U_{n,\tau}| \\
&\quad = \frac{1}{\sqrt{n}}\left|\sum_{i=1}^{s_\tau} r_i^{*2} + 2\sum_{i=1}^{s_\tau-1} r_i^* r_{i+1}^* - \sum_{j=1}^{\tau}\left(\sum_{i=I_j}^{I_j+L_j-1} r_{ni}^2 + 2\sum_{i=I_j}^{I_j+L_j-1} r_{ni}r_{n(i+1)}\right)\right| \\
&\quad = \frac{2}{\sqrt{n}}\left|\sum_{i=1}^{s_\tau-1} r_i^* r_{i+1}^* - \sum_{j=1}^{\tau}\left(\sum_{i=I_j}^{I_j+L_j-1} r_{ni}r_{n(i+1)}\right)\right| \\
&\quad = \frac{2}{\sqrt{n}}\left|\sum_{j=1}^{\tau}\left(\sum_{i=I_j}^{I_j+L_j-2} r_{ni}r_{n(i+1)} + r_{n(I_j+L_j-1)}r_{nI_{j+1}}\right)\right. \\
&\qquad\qquad \left. - \sum_{j=1}^{\tau}\left(\sum_{i=I_j}^{I_j+L_j-1} r_{ni}r_{n(i+1)}\right)\right| \\
&\quad \le \frac{2}{\sqrt{n}}\sum_{j=1}^{\tau}\left|r_{n(I_j+L_j-1)}(r_{nI_{j+1}} - r_{n(I_j+L_j)})\right|.
\end{aligned}$$

Let $Y_j = r_{n(I_j+L_j-1)}(r_{nI_{j+1}} - r_{n(I_j+L_j)})$. Then $\{Y_j : j = 1, 2, \dots\}$ is a sequence of i.i.d. random variables since $\{(I_j, L_j) : j = 1, 2, \dots\}$ are i.i.d. Note that $\tau = np + O_p(\sqrt{np})$ by [25]. For any sequence m with $m/(np) \to 1$, we will show that $\frac{2}{\sqrt{n}}\sum_{j=1}^{m}|Y_j| \xrightarrow{\mathrm{p}^*} 0$. For any $\epsilon > 0$ and $\delta > 0$,

$$P^*\left(\frac{2}{\sqrt{n}}\sum_{j=1}^{m}|Y_j| > \epsilon\right) \le \frac{1}{n^{1+\delta/2}\epsilon^{2+\delta}} E^*\left|\sum_{j=1}^{m} Y_j\right|^{2+\delta} = C\left(\frac{m}{n}\right)^{1+\delta/2}$$

observing $E^*|Y_j|^{2+\delta} < \infty$ a.s. The last term above tends to zero as $n \to \infty$, and thus (a) holds.

Secondly, we observe

$$\frac{1}{\sqrt{n}}\left(Q^*_{n,\tau} - RV^{AC_1*}_{(n)}\right) = \frac{1}{\sqrt{n}}\sum_{i=n+1}^{s_\tau}(r_i^{*2} + 2r_{i-1}^* r_i^*).$$

Let $\eta_1 = n - s_{\tau-1}$ where $s_{\tau-1} = L_1 + \cdots + L_{\tau-1}$, and let $\eta = L_\tau - \eta_1$. Note that, from the memoryless property of the geometric distribution of the block length of the stationary bootstrap procedure, the random variable η has a geometric distribution with mean $1/p$, conditional on $(\eta_1, s_{\tau-1})$. Hence, $\frac{1}{\sqrt{n}}\sum_{i=n+1}^{s_\tau}(r_i^{*2} + 2r_{i-1}^* r_i^*)$ is equal in distribution to $\frac{1}{\sqrt{n}}[S_{I,\eta} + 2\cdot T_{I-1,\eta}]$ where I is uniform on $\{1, \dots, n\}$. Thus, we will show that

$$\frac{1}{\sqrt{n}}\sum_{j=I}^{I+\eta-1}(r_{nj}^2 + 2r_{n(j-1)}r_{nj}) \xrightarrow{\mathrm{p}} 0. \tag{6.2}$$

For any $\epsilon > 0$,

$$P\left(\frac{1}{\sqrt{n}}\sum_{j=I}^{I+\eta-1}(r_{nj}^2 + 2r_{n(j-1)}r_{nj}) > \epsilon\right)$$
$$= \frac{1}{n}\sum_{i=1}^{n}\sum_{\ell=1}^{\infty} p(1-p)^{\ell-1} P\left(\frac{1}{\sqrt{n}}\sum_{j=i}^{i+\ell-1}(r_{nj}^2 + 2r_{n(j-1)}r_{nj}) > \epsilon\right).$$

We note that, for $\delta > 0$,

$$P\left(\frac{1}{\sqrt{n}}\sum_{j=i}^{i+\ell-1}(r_{nj}^2 + 2r_{n(j-1)}r_{nj}) > \epsilon\right) \le \frac{1}{n^{1+\delta/2}\epsilon^{2+\delta}} E|S_{i,\ell} + 2T_{i-1,\ell}|^{2+\delta}.$$

In Appendix (i) it is shown that $E|S_{i,\ell} + 2T_{i-1,\ell}|^{2+\delta} \le C\ell^{1+\delta/2}$ by applying Minkowski's inequality. Thus,

$$P\left(\frac{1}{\sqrt{n}}\sum_{j=I}^{I+\eta-1}(r_{nj}^2 + 2r_{n(j-1)}r_{nj}) > \epsilon\right) \le C\sum_{\ell=1}^{\infty} p(1-p)^{\ell-1}\frac{\ell^{1+\delta/2}}{n^{1+\delta/2}} = \frac{C}{(np)^{1+\delta/2}}$$

since $\sum_{\ell=1}^{\infty}(1-p)^{\ell-1}\ell^a = O(1/p^{a+1})$ for $a \ge 1$. The last term tends to zero since $np \to \infty$. Thus the convergence in (conditional) probability in (6.2) holds and so does the second result (b). □

Lemma 6.2. *Under the same assumptions as in Theorem 3.1, we have*

$$(a) \quad \frac{1}{\sqrt{n}}\left(U_{n,\tau} - RV_{(n)}^{AC_1}\right) \xrightarrow{\mathrm{d}^*} N(0, 8\omega^4),$$
$$(b) \quad \frac{1}{\sqrt{n}}\left(RV_{(n)}^{AC_1} - E^*(RV_{(n)}^{AC_1 *})\right) \xrightarrow{\mathrm{p}} 0.$$

Proof. In order to show the first convergence (a), it suffices to show that, for any sequence m with $m/(np) \to 1$,

$$\frac{1}{\sqrt{n}}\left[U_{n,m} - \frac{m}{np}RV_{(n)}^{AC_1}\right] \xrightarrow{\mathrm{d}^*} N(0, 8\omega^4), \tag{6.3}$$

since $\tau = np + O_p(\sqrt{np})$ by [25].

For $1 \le j \le m$, let $Z_{n,j} = \sqrt{m/n}(S_{I_j,L_j} + 2\cdot T_{I_j,L_j})$. Note that $\bar{Z}_m = \frac{1}{m}\sum_{j=1}^{m} Z_{n,j}$ is the average of i.i.d. variables since $\{I_j\}$ and $\{L_j\}$ are i.i.d. Also, observing

$$E^*[S_{I_j,L_j}|L_j = \ell] = \frac{1}{n}\sum_{i=1}^{n}\sum_{j=i}^{i+\ell-1} r_{nj}^2 = \frac{\ell}{n}\sum_{i=1}^{n} r_i^2,$$

$$E^*[T_{I_j,L_j}|L_j = \ell] = \frac{1}{n}\sum_{i=1}^{n}\sum_{j=i}^{i+\ell-1} r_{nj}r_{n(j+1)} = \frac{\ell}{n}\sum_{i=1}^{n} r_i r_{i+1} \quad \text{with} \quad r_{n+1} = r_1,$$

we have $E^*[S_{I_j,L_j} + 2 \cdot T_{I_j,L_j}|L_j] = \frac{1}{n}L_j(RV_{(n)}^{AC_1} + 2r_1r_n)$ and thus $E^*[\bar{Z}_m] = \frac{\sqrt{m}}{pn\sqrt{n}}[RV_{(n)}^{AC_1} + 2r_1r_n]$. Note that $\frac{\sqrt{m}}{pn\sqrt{n}}r_1r_n \xrightarrow{\text{p}} 0$ as $n \to \infty$, and the left term of (6.3) is equal to $\sqrt{m}[\bar{Z}_m - E^*\bar{Z}_m] + o_p(1)$.

Now let $Z^*_{n,j} = Z_{n,j} - E^*Z_{n,j}$, and then

$$\sqrt{m}[\bar{Z}_m - E^*\bar{Z}_m] = \frac{1}{\sqrt{m}}\sum_{j=1}^{m} Z^*_{n,j}.$$

$\{Z^*_{n,j} : 1 \leq j \leq m\}$ are i.i.d. variables with mean zero under P^*. Thus we obtain, (for $\iota = \sqrt{-1}$),

$$\begin{aligned} E^*\left[e^{\iota(t/\sqrt{m})\sum_{j=1}^{m} Z^*_{n,j}}\right] &= \left(E^*\left[e^{\iota(t/\sqrt{m})Z^*_{n,1}}\right]\right)^m \\ &= \left[1 + \frac{\iota t}{\sqrt{m}}E^*Z^*_{n,1} - \frac{t^2}{2m}(1+o(1))E^*(Z^*_{n,1})^2\right]^m. \end{aligned}$$

By Proposition 2.1(b) and Theorem 3.1(a),

$$\frac{1}{n}Var^*\left(RV_{(n)}^{AC_1*}\right) = Var^*(Z^*_{n,1}) \xrightarrow{\text{p}} 8\omega^4,$$

and thus the above term tends to $e^{-4t^2\omega^4}$ in probability, and the desired result (a) is obtained. Now, by this result and by Lemma 6.1, the second convergence (b) is obvious. $\square$

Proof of Theorem 3.1(a). We write $RV_{(n)}^{AC_1*} = \sum_{i=1}^n X_i^*$ where $X_i^* = r_i^{*2} + 2r_i^*r_{i+1}^*$ for $i = 1, \dots, n-1$, and $X_n^* = r_n^{*2}$, and observe

$$\begin{aligned} \frac{1}{n}Var^*\left(RV_{(n)}^{AC_1*}\right) &= \frac{1}{n}Var^*\left(\sum_{i=1}^{n-1}X_i^* + X_n^*\right) \\ &= \frac{1}{n}\left[Var^*\left(\sum_{i=1}^{n-1}X_i^*\right) + Var^*(X_n^*) + 2Cov^*\left(\sum_{i=1}^{n-1}X_i^*, X_n^*\right)\right] \\ &= \frac{n-1}{n}\left[Var^*(X_1^*) + 2\sum_{i=1}^{n-2}\left(1 - \frac{i}{n-1}\right)Cov^*(X_1^*, X_{1+i}^*)\right] \\ &\quad + \frac{1}{n}Var^*(X_n^*) + \frac{2}{n}\sum_{i=1}^{n-1}Cov^*(X_i^*, X_n^*). \end{aligned}$$

Similarly we can express $\frac{1}{n}Var(RV_{(n)}^{AC_1})$ in terms of $X_i := r_i^2 + 2r_ir_{i+1}$ for $i = 1, \dots, n-1$ and $X_n = r_n^2$.

We will show that $\frac{1}{n}Var^*(RV_{(n)}^{AC_1*})$ has the same limiting as that of $\frac{1}{n}Var(RV_{(n)}^{AC_1})$, as $n \to \infty$, which can be expressed as $c(0) + 2\lim_{n\to\infty}\sum_{i=1}^{n-2}c(i)$ where $c(i) = Cov(r_1^2 + 2r_1r_2, r_{1+i}^2 + 2r_{1+i}r_{2+i})$.

First, we show that

$$\left|\sum_{i=1}^{n-2}\left(1-\frac{i}{n-1}\right)Cov^*(X_1^*, X_{1+i}^*) - \sum_{i=1}^{n-2} c(i)\right| \xrightarrow{\text{p}} 0$$

as $n \to \infty$. Write $Cov^*(X_1^*, X_{1+i}^*) =$

$$\begin{aligned}
&Cov^*(r_1^{*2}, r_{1+i}^{*2}) + 2Cov^*(r_1^{*2}, r_{1+i}^* r_{2+i}^*) + 2Cov^*(r_1^* r_2^*, r_{1+i}^{*2}) \\
&\quad + 4Cov^*(r_1^* r_2^*, r_{1+i}^* r_{2+i}^*) \\
&\quad =: c_1^*(i) + 2c_2^*(i) + 2c_3^*(i) + 4c_4^*(i)
\end{aligned}$$

and $c(i) = c_1(i) + 2c_2(i) + 2c_3(i) + 4c_4(i)$ where $c_1(i) = Cov(r_1^2, r_{1+i}^2)$, $c_2(i) = Cov(r_1^2, r_{1+i}r_{2+i})$, $c_3(i) = Cov(r_1 r_2, r_{1+i}^2)$ and $c_4(i) = Cov(r_1 r_2, r_{1+i} r_{2+i})$. It suffices to show that, for $k = 1, 2, 3, 4$,

$$\sum_{i=1}^{n-2}\left[\left(1-\frac{i}{n-1}\right)c_k^*(i) - c_k(i)\right] \xrightarrow{\text{p}} 0. \tag{6.4}$$

For $k = 1$, we observe $Cov^*(r_1^{*2}, r_{1+i}^{*2}) = E^*[r_1^{*2} r_{1+i}^{*2}] - (E^* r_i^{*2})^2$, $E^* r_i^{*2} = \frac{1}{n}\sum_{j=1}^n r_j^2$ and

$$\begin{aligned}
E^*[r_1^{*2} r_{1+i}^{*2}] &= E[r_1^{*2} r_{1+i}^{*2} | L_1 > i] P(L_1 > i) + E[r_1^{*2} r_{1+i}^{*2} | L_1 \leq i] P(L_1 \leq i) \\
&= \frac{1}{n}\sum_{j=1}^n r_{nj}^2 r_{n(j+i)}^2 (1-p)^i + \left(\frac{1}{n}\sum_{j=1}^n r_j^2\right)^2 [1-(1-p)^i]
\end{aligned}$$

and thus

$$Cov^*(r_1^{*2}, r_{1+i}^{*2}) = \frac{(1-p)^i}{n}\sum_{j=1}^n \left[(r_{nj}^2 - A_1)(r_{n(j+i)}^2 - A_1)\right]$$

where $A_1 = \frac{1}{n}\sum_{j=1}^n r_j^2$. Let $\hat{d}_n(i) = \frac{1}{n}\sum_{j=1}^{n-i}[(r_j^2 - A_1)(r_{j+i}^2 - A_1)]$, and then

$$Cov^*(r_1^{*2}, r_{1+i}^{*2}) = (1-p)^i [\hat{d}_n(i) + \hat{d}_n(n-i)].$$

Thus we have

$$\sum_{i=1}^{n-2}\left(1-\frac{i}{n-1}\right)Cov^*(r_1^{*2}, r_{1+i}^{*2}) = \sum_{i=1}^{n-2} b_n(i)\hat{d}_n(i)$$

where $b_n(i) = (1-\frac{i}{n-1})(1-p)^i + \frac{i}{n-1}(1-p)^{n-1-i}$.

In order to show the convergence in probability for $k = 1$ in (6.4), we observe

$$\begin{aligned}
&\sum_{i=1}^{n-2}\left[\left(1-\frac{i}{n-1}\right)c_1^*(i)-c_1(i)\right]\\
&\quad=\sum_{i=1}^{n-2}\left[b_n(i)\hat{d}_n(i)-c_1(i)\right]\\
&\quad=\sum_{i=1}^{n-2}\left[\frac{b_n(i)}{n}\sum_{j=1}^{n-i}\left[(r_j^2-A_1)(r_{j+i}^2-A_1)\right]-c_1(i)\right]\\
&\quad\le\sum_{i=1}^{n-2}\left(\frac{b_n(i)}{n}\sum_{j=1}^{n-i}\left[(r_j^2-A_1)(r_{j+i}^2-A_1)-c_1(i)\right]\right)\\
&\qquad+\sum_{i=1}^{n-2}\left(\frac{b_n(i)}{n}(n-i)c_1(i)-c_1(i)\right).
\end{aligned}\tag{6.5}$$

Since $b_n(i) - 1 \to 0$ for all i, the second term of (6.5) has the same limit as $\frac{1}{n}\sum_{i=1}^{n-2} ic_1(i)$, which tends to zero by Assumption 3.

For the convergence in probability of the first term in (6.5), note that $\sum_{i=1}^{n-2} b_n(i) \le 2/p$ (see [25]), and we show that

$$\frac{1}{np}\sum_{j=1}^{n-i}\left[(r_j^2-A_1)(r_{j+i}^2-A_1)-c_1(i)\right]\xrightarrow{\text{p}}0.$$

For any $\epsilon > 0$ and $\delta > 0$

$$P\left(\frac{1}{np}\sum_{j=1}^{n-i}\left[(r_j^2-A_1)(r_{j+i}^2-A_1)-c_1(i)\right]>\epsilon\right)\le\frac{1}{(np\epsilon)^{2+\delta}}E\left|\sum_{j=1}^{n-i}Y_{j,i}\right|^{2+\delta}\tag{6.6}$$

where $Y_{j,i} = (r_j^2 - A_1)(r_{j+i}^2 - A_1) - c_1(i)$. By the fact that $E|\sum_{j=1}^{n-i} Y_{j,i}|^{2+\delta} \le Cn^{1+\delta/2}$, given in Appendix (ii), the left term of (6.6) is less than or equal to $c/(n^{1+\delta/2}p^{2+\delta})$, which tends to zero if $np^2 \to \infty$. Hence the convergence in probability of the first term in (6.5) holds, and so does the convergence in probability for $k = 1$ in (6.4). The convergence in probability for $k = 2$ in (6.4) can be shown in the same way.

Now we show the convergence in probability for $k = 3$ in (6.4). For $k = 3$, we observe $Cov^*(r_1^*r_2^*, r_{1+i}^{*2}) = E^*[r_1^*r_2^*r_{1+i}^{*2}] - (E^*r_1^*r_2^*)(E^*r_{1+i}^{*2})$, $E^*r_{1+i}^{*2} = \frac{1}{n}\sum_{j=1}^n r_j^2 = E(r_1^2) + o_p(1)$ and $E^*[r_1^*r_2^*] = E[r_1^*r_2^*|L_1 > 1]P(L_1 > 1) + E[r_1^{*2}r_2^{*2}|L_1 \le 1]P(L_1 \le 1) = \frac{1}{n}\sum_{j=1}^n r_{nj}r_{n(j+1)}(1-p) + (\frac{1}{n}\sum_{j=1}^n r_j)^2 p = E(r_1r_2) + o_p(1)$. Now we observe $E^*[r_1^*r_2^*r_{1+i}^{*2}]$. If $i = 1$, then we have

$E^*[r_1^* r_2^* r_{1+i}^{*2}] = E[r_1 r_2^3] + o_p(1)$. If $i \geq 2$, then

$$\begin{aligned} E^*[r_1^* r_2^* r_{1+i}^{*2}] &= E[r_1^* r_2^* r_{1+i}^{*2} | L_1 > i] P(L_1 > i) + E[r_1^* r_2^* r_{1+i}^{*2} | L_1 \leq i] P(L_1 \leq i) \\ &= \frac{1}{n} \sum_{j=1}^{n} r_{nj} r_{n(j+1)} r_{n(j+i)}^2 (1-p)^i \\ &\quad + \left(\frac{1}{n} \sum_{j=1}^{n} r_{nj} r_{n(j+1)} \right) \left(\frac{1}{n} \sum_{j=1}^{n} r_j^2 \right) [1 - (1-p)^i] \end{aligned}$$

and thus

$$c_3^*(i) = Cov^*(r_1^* r_2^*, r_{1+i}^{*2}) = \frac{(1-p)^i}{n} \sum_{j=1}^{n} \left[(r_{nj} r_{n(j+1)} - A_2)(r_{n(j+i)}^2 - A_1) \right] + o_p(1)$$

where $A_2 = \frac{1}{n} \sum_{j=1}^{n} r_{nj} r_{n(j+1)}$. For the left term in (6.4) with $k = 3$, we have

$$\begin{aligned} &\left(1 - \frac{1}{n-1}\right) c_3^*(1) - c_3(1) + \sum_{i=2}^{n-2} \left[\left(1 - \frac{i}{n-1}\right) c_3^*(i) - c_3(i) \right] \\ &= \sum_{i=2}^{n-2} \left[(1-p)^i \left(1 - \frac{i}{n-1}\right) \frac{1}{n} \sum_{j=1}^{n} \left[(r_{nj} r_{n(j+1)} - A_2)(r_{n(j+i)}^2 - A_1) - c_3(i) \right] \right] \qquad (6.7) \\ &\quad + \sum_{i=2}^{n-2} \left[(1-p)^i \left(1 - \frac{i}{n-1}\right) c_3(i) - c_3(i) \right] + o_p(1). \qquad (6.8) \end{aligned}$$

By inequality $\sum_{i=2}^{n-2} (1-p)^i - \frac{1}{n-1} \sum_{i=2}^{n-2} i(1-p)^i \leq 2/p$, and by the fact that $E|\sum_{j=1}^{n-1} W_{j,i}|^{2+\delta} \leq Cn^{1+\delta/2}$, given in Appendix (iii), where $W_{j,i} = (r_{nj} r_{n(j+1)} - A_2)(r_{n(j+i)}^2 - A_1) - c_3(i)$, the term in (6.7) converges to zero in probability like the argument in (6.6) if $np^2 \to \infty$. The first term in (6.8) is less than or equal to $\frac{1}{n-1} \sum_{i=2}^{n-2} i c_3(i)$, which tends to zero by Assumption 3. Therefore, the convergence in probability for $k = 3$ in (6.4) holds. Similarly, for $k = 4$ in (6.4) the convergence in probability can be shown.

Secondly, we show that

$$\frac{1}{n} \sum_{i=1}^{n-1} Cov^*(X_i^*, X_n^*) \xrightarrow{\text{p}} 0.$$

We have $Cov^*(X_i^*, X_n^*) = Cov^*(r_i^{*2}, r_n^{*2}) + 2Cov^*(r_i^* r_{i+1}^*, r_n^{*2}) = c_1^*(n-i) + 2c_3^*(n-i)$ and thus $\frac{1}{n} \sum_{i=1}^{n-1} Cov^*(X_i^*, X_n^*) = \frac{1}{n} \sum_{i=1}^{n-1} [c_1^*(n-i) + 2c_3^*(n-i)] = \frac{1}{n} \sum_{i=1}^{n-1} [c_1^*(i) + 2c_3^*(i)] \xrightarrow{\text{p}} 0$ by the similar arguments above and Assumption 3.

Finally, we show that

$$\frac{n-1}{n} Var^*(X_1^*) + \frac{1}{n} Var^*(X_n^*) - c(0) \xrightarrow{\text{p}} 0.$$

We observe $Var^*(X_1^*) = Var^*(r_i^{*2}) + 4Var^*(r_1^* r_2^*) + 4Cov^*(r_1^{*2}, r_1^* r_2^*)$, $Var^*(r_i^{*2}) = \frac{1}{n}\sum_{i=1}^n r_i^4 - (\frac{1}{n}\sum_{i=1}^n r_i^2)^2 = Var(r_1^2) + o_p(1)$, $E^*(r_1^* r_2^*) = E^*[r_1^* r_2^* | L_1 > 1] P(L_1 > 1) + E^*[r_1^* r_2^* | L_1 = 1] P(L_1 = 1) = \frac{1}{n}\sum_{i=1}^n r_{ni} r_{n(i+1)}(1-p) + (\frac{1}{n}\sum_{i=1}^n r_i)^2 p = E[r_1 r_2] + o_p(1)$, similarly $E^*(r_1^{*2} r_2^{*2}) = E[r_1^2 r_2^2] + o_p(1)$, and thus $Var^*(r_1^* r_2^*) = Var(r_1 r_2) + o_p(1)$. In the same way we have $Cov^*(r_1^{*2}, r_1^* r_2^*) = Cov(r_1^2, r_1 r_2) + o_p(1)$, and hence $Var^*(X_1^*) = Var(X_1) + o_p(1)$, noting that $c(0) = Var(X_1)$, and $\frac{1}{n} Var^*(X_n^*) = \frac{1}{n} Var^*(r^{*4}) \xrightarrow{\text{p}} 0$, we obtain the desired result. This finishes the proof. $\square$

Proof of Theorem 3.2. If $p = c \cdot n^{-\frac{\delta - 2\rho}{2(2+\delta)}}$ for some $c > 0$, $0 < \rho < \delta/2$ and $0 < \delta < 1$, then all convergences in probability in Lemmas 6.1–6.2 and Theorem 3.1 and their proofs are replaced by almost sure convergences by Borel-Cantelli Lemma and by the following:

$$\sum_{n=1}^{\infty} \left(\frac{1}{np^2}\right)^{1+\delta/2} = c \sum_{n=1}^{\infty} \frac{1}{n^{1+\rho}} < \infty.$$

See [16] for the almost sure convergence of the stationary bootstrap. $\square$

Appendix A

(i) $E|S_{i,\ell} + 2T_{i-1,\ell}|^{2+\delta} \le C\ell^{1+\delta/2}$ for some constant C depending only on δ. Here and below, C is a generic constant.

(ii) $E|\sum_{j=1}^{n-i} Y_{j,i}|^{2+\delta} \le Cn^{1+\delta/2}$ where $Y_{j,i}$ is either $(r_j^2 - A_1)(r_{j+i}^2 - A_1) - c_1(i)$ or $(r_j^2 - A_1)(r_{j+i} r_{j+i+1} - A_2) - c_2(i)$.

(iii) $E|\sum_{j=1}^{n-1} W_{j,i}|^{2+\delta} \le Cn^{1+\delta/2}$ where $W_{j,i}$ is either $(r_{nj} r_{n(j+1)} - A_2)(r_{n(j+i)}^2 - A_1) - c_3(i)$ or $(r_{nj} r_{n(j+1)} - A_2)(r_{n(j+i)} r_{n(j+i+1)} - A_2) - c_4(i)$.

Proof of (i). We will use the Marcinkiewicz-Zygmund inequality and the Minkowsky inequality for the bound of $E|S_{i,\ell} + 2T_{i-1,\ell}|^{2+\delta}$. First, if $i + \ell - 1 \le n$, then

$$\begin{aligned} E|S_{i,\ell}|^{2+\delta} &= E|\sum_{j=i}^{i+\ell-1} r_{nj}^2|^{2+\delta} \\ &\le CE\left(\sum_{j=i}^{i+\ell-1} r_{nj}^4\right)^{1+\delta/2} \le C\left[\sum_{j=i}^{i+\ell-1}\left(E|r_{nj}^2|^{2+\delta}\right)^{2/(2+\delta)}\right]^{1+\delta/2} \\ &= C\ell^{1+\delta/2} \end{aligned}$$

since $E|r_{nj}^2|^{2+\delta} < \infty$. Also we have $E|T_{i-1,\ell}|^{2+\delta} = C\ell^{1+\delta/2}$. By the Minkowsky inequality again, we have

$$\begin{aligned} &E|S_{i,\ell} + 2T_{i-1,\ell}|^{2+\delta} \\ &\quad \le \left[\left(E|S_{i,\ell}|^{2+\delta}\right)^{1/(2+\delta)} + C\left(E|T_{i-1,\ell}|^{2+\delta}\right)^{1/(2+\delta)}\right]^{2+\delta} \le C\ell^{1+\delta/2}. \end{aligned}$$

Secondly, if $i+\ell-1>n$ and $\ell\leq n$, then write $S_{i,\ell}=(r_{ni}^2+\cdots+r_{nn}^2)+(r_{n1}^2+\cdots+r_{n(i+\ell-1)}^2)$ and $T_{i-1,\ell}=(r_{n(i-1)}r_{ni}+\cdots+r_{n(n-1)}r_{nn})+(r_{nn}r_{n1}+\cdots+r_{n(i+\ell-2)}r_{n(i+\ell-1)})$, and similarly we have $E|S_{i,\ell}+2T_{i-1,\ell}|^{2+\delta}\leq C\ell^{1+\delta/2}$. Finally if $i+\ell-1\geq n$ and $\ell>n$, then write $\ell=nq+\tilde{\ell}$ for some $q\geq 1$ and $1\leq\tilde{\ell}\leq n$, and

$$S_{i,\ell}=\left[\sum_{j=i}^{n}r_{nj}^2+q\sum_{j=1}^{n}r_{nj}^2+\sum_{j=1}^{i+\tilde{\ell}-1}r_{nj}^2\right]=nqA_1+S_{i,\tilde{\ell}}=(\ell-\tilde{\ell})A_1+S_{i,\tilde{\ell}}.$$

Hence, by Minkowsky inequality again, we have $E|S_{i,\ell}|^{2+\delta}=$

$$\begin{aligned}E|(\ell-\tilde{\ell})A_1+S_{i,\tilde{\ell}}|^{2+\delta}&\leq\left[\left(E|(\ell-\tilde{\ell})A_1|^{2+\delta}\right)^{1/(2+\delta)}+\left(E|S_{i,\tilde{\ell}}|^{2+\delta}\right)^{1/(2+\delta)}\right]^{2+\delta}\\&\leq C(\ell-\tilde{\ell})^{1+\delta/2}+C\tilde{\ell}^{1+\delta/2}\leq C\ell^{1+\delta/2}.\end{aligned}$$

For $T_{i-1,\ell}$, we have the same argument to get $E|T_{i-1,\ell}|^{2+\delta}=C\ell^{1+\delta/2}$, and by the Minkowsky inequality again, we obtain the desired result in Appendix (i). $\square$

Proofs of (ii), (iii). For fixed i in $\{1,\ldots,n-1\}$, note that $\{Y_{j,i}\,:\,j=1,2,\ldots\}$ and $\{W_{j,i}\,:\,j=1,2,\ldots\}$ are i-dependent sequences, which belong to a class of ψ-weakly dependent sequences, (cf. [8]). Observing $E|Y_{j,i}|^{2+\delta}<\infty$ and $E|W_{j,i}|^{2+\delta}<\infty$, by Lemma 4.1 of [16], the desired results for the bounds are obtained. $\square$

Acknowledgements

The authors are very grateful for the comments of an anonymous referee which improve the paper considerably. This work was supported by the National Research Foundation of Korea (NRF-2012-2046157) through the National Research Foundation of Korea (NRF) funded by the Ministry of Education Science and Technology.

References

[1] Aït-Sahalia, Y., Mykland, P.A., and Zhang, L. (2005). How often to sample a continuous-time process in the presence of market microstructure noise. *Review of Financial Studies* **18** 351–416.

[2] Aït-Sahalia, Y., Mykland, P.A., and Zhang, L. (2011). Ultra high frequency volatility estimation with dependent microstructure noise. *Journal of Econometrics* **160** 160–175. MR2745875

[3] Bandi, F.M. and Russel, J.R. (2008). Microstructure noise, realized variance, and optimal sampling. *Review of Economic Studies* **75** 339–369. MR2398721

[4] Barndorff-Nielsen, O.E., Graversen, S.E., Jacod, J., and Shephard, N. (2006). Limit theorems for bipower variation in financial econometrics. *Econometric Theory* **22** 677–719. MR2283032

[5] Barndorff-Nielsen, O.E., Hansen, P.H., Lunde, A., and Shephard, N. (2008). Designing realised kernels to measure the ex-post variation of equity prices in the presence of noise. *Econometrica* **76** 1481-1536. MR2468558

[6] Barndorff-Nielsen, O.E. and Shephard, N. (2002). Econometric analysis of realised volatility and its use in estimating stochastic volatility models. *Journal of the Royal Statistical Society, Ser. B* **64** 253–280. MR1904704

[7] Corsi, F., Zumbach, G., Müller, U., and Dacorogna, M. (2001). Consistent high-precision volatility from high-frequency data. *Economic Notes* **30** 183–204.

[8] Doukhan, P. and Louhichi, S. (1999). A new weak dependence condition and applications to moment inequalities. *Stochastic Processes and their Applications* **84** 313–342. MR1719345

[9] Dovonon, P., Goncalves, S., and Meddahi, N. (2013). Bootstrapping realized multivariate volatility measures. *Journal of Econometrics* **172** 49–65. MR2997129

[10] Goncalves, S. and de Jong, R. (2003). Consistency of the stationary bootstrap under weak moment conditions. *Economics Letters* **81** 273–278. MR2017286

[11] Goncalves, S. and Meddahi, N. (2009). Bootstrapping realized volatility. *Econometrica* **77** 283–306. MR2477851

[12] Goncalves, S. and White, H. (2002). The bootstrap of the mean for dependent heterogeneous arrays. *Econometric Theory* **18** 1367–1384. MR1945417

[13] Hansen, P.R. and Lunde, A. (2006). Realized variance and market microstructure noise. *Journal of Business and Economic Statistics* **24** 127–161. MR2234447

[14] Hwang, E. and Shin, D.W. (2011). Stationary bootstrapping for nonparametric estimator of nonlinear autoregressive model. *Journal of Time Series Analysis* **32** 292–303. MR2808256

[15] Hwang, E. and Shin, D.W. (2012a). Stationary bootstrap for kernel density estimators under ψ-weak dependence. *Computational Statistics and Data Analysis* **56** 1581–1593. MR2892362

[16] Hwang, E. and Shin, D.W. (2012b). Strong consistency of the stationary bootstrap under ψ-weak dependence. *Statistics and Probability Letters* **82** 488–495. MR2887463

[17] Jacod, J. and Protter, P. (1998). Asymptotic error distributions for the Euler method for stochastic differential equations. *Annals of Probability* **26** 267–307. MR1617049

[18] Kalnina, I. and Linton, O. (2006). Estimating quadratic variation consistently in the presence of correlated measurement error. Unpublished Paper, Dept. of Economics, London School of Economics.

[19] Meddahi, N. (2002). A theoretical comparison between integrated and realized volatilities. *Journal of Applied Econometrics* **17** 479–508.

[20] Mudelsee, M. (2003). Estimating Pearson's correlation coefficient with bootstrap confidence interval from serially dependent time series. *Mathematical Geology* **35** 651–665.

[21] Nordman, D.J. (2009). A note on the stationary bootstrap's variance. *Annals of Statistics* **37** 359–370. MR2488355

[22] Oya, K. (2011). Bias-corrected realized variance under dependent microstructure noise. *Mathematics and Computers in Simulation* **81** 1290–1298. MR2781660

[23] Parker, C. Paparoditis, E., and Politis, D.N. (2006). Unit root testing via the stationary bootstrap. *Journal of Econometrics* **133** 601–638. MR2252911

[24] Pigorsch, C., Pigorsch, U., and Popov, I. (2012). *Volatility estimation based on highfrequency data*. In J.C. Duan, J.E. Gentle, and W.K. Härdle (Eds.), Handbook of Computational Finance. Springer. pp. 335–369. MR2908478

[25] Politis, D.N. and Romano, J.P. (1994). The stationary bootstrap. *Journal of the American Statistical Association* **89** 1303–1313 MR1310224

[26] Protter, P. (2005). *Stochastic integration and differential equations*. Second edition. Springer-Verlag, Berlin. MR2273672

[27] Shin, D.W. and Hwang, E. (2013). Stationary bootstrapping for cointegrating regressions. *Statistics and Probability Letters* **83** 474–480. MR3006979

[28] Swensen, A.R. (2003). Bootstrapping unit root tests for integrated processes. *Journal of Time Series Analysis* **24** 99–126. MR1960385

[29] Zhang, L., Mykland, P.A., and Ait-Sahalia, Y. (2005). A tale of two time scales: Determining integrated volatility with noisy high-frequency data. *Journal of the American Statistical Association* **100** 1394–1411. MR2236450

[30] Zhou, B. (1996). High-frequency data and volatility in foreign-exchange rates. *Journal of Business and Economic Statistics* **14** 45–52.

[31] Zumbach, G., Corsi, F., and Trapletti, A. (2002). Efficient estimation of volatility using high frequency data, Manuscript, Olsen & Associates, Research Institute for Applied Economics, Zurich, Switzerland.