Journal of Econometrics 161 (2011) 262–283

Contents lists available at ScienceDirect

Journal of Econometrics

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

Subsampling high frequency data☆

Ilze Kalnina *

Département de sciences économiques, Université de Montréal, C.P. 6128, succ. Centre-Ville, Montréal, H3C 3J7, QC, Canada

ARTICLE INFO

Article history:
Received 14 February 2010
Received in revised form
21 December 2010
Accepted 23 December 2010
Available online 28 December 2010

JEL classification:
C12
C13
C14

Keywords:
Subsampling
Market microstructure noise
High frequency data
Realised volatility

ABSTRACT

The main contribution of this paper is to propose a novel way of conducting inference for an important general class of estimators that includes many estimators of integrated volatility. A subsampling scheme is introduced that consistently estimates the asymptotic variance for an estimator, thereby facilitating inference and the construction of valid confidence intervals. The new method does not rely on the exact form of the asymptotic variance, which is useful when the latter is of complicated form. The method is applied to the volatility estimator of Aït-Sahalia et al. (2011) in the presence of autocorrelated and heteroscedastic market microstructure noise.

1. Introduction

Volatility estimation is a key component in the evaluation of financial risk. Financial econometrics continues to make progress in developing more robust and efficient estimators of volatility. But for some estimators, the asymptotic variance is hard to derive or may take a complicated form and be difficult to estimate. To tackle these problems, the current paper develops a method of inference that is automatic in the sense that it does not rely on the exact form of the asymptotic variance. In the traditional stationary time series framework, this task can be accomplished by traditional bootstrap and subsampling variance estimators. However, these are inconsistent with high frequency data, which is potentially contaminated with market microstructure noise, see Section 2.1.

A new subsampling method is developed, which enables us to conduct inference for a general class of estimators that includes many estimators of integrated volatility. The question of inference on volatility estimates is important due to volatility being unobservable. For example, one might want to test whether volatility is the same on two different days, or in two different time periods within the same day. The latter corresponds to testing for diurnal variation in the volatility. Also, a common way of testing for jumps in prices is to compare two different volatility estimates, which converge to the same quantity under the null hypothesis of no jumps, but are different asymptotically under the alternative hypothesis of jumps in prices. Then, a consistent inferential method is needed to determine whether the two volatility estimates are significantly different.

To illustrate the robustness of the new method, this paper considers the example of the inference problem for the integrated variance estimator of Aït-Sahalia et al. (2011), in the presence of market microstructure noise. As several assumptions about the market microstructure noise are relaxed, the expression for the asymptotic variance becomes more complicated, and it becomes more challenging to estimate each component of the variance separately. On the other hand, the new subsampling method delivers consistent confidence intervals that are simple to calculate.

According to the fundamental theorem of asset pricing (see Delbaen and Schachermayer, 1994), the price process should follow a semimartingale. In this model, integrated variance (sometimes called integrated volatility) is a natural measure of variability of the price path (see, e.g. Andersen et al., 2001). With moderate frequency data, say 5 or 15 min data, this can be estimated by the so-called realized variance (RV), a sum of squared returns (also referred to as realized volatility).[1] The nonparametric

☆ I would like to thank Donald Andrews, Peter Bühlmann, Valentina Corradi, Kirill Evdokimov, Silvia Gonçalves, Yuichi Kitamura, Nour Meddahi, Per Mykland, Mark Podolskij, Mathieu Rosenbaum, Myung Hwan Seo, Kevin Sheppard, and Mathias Vetter for helpful comments. I am particularly grateful to Oliver Linton and Peter Phillips for their help and encouragement. This research was supported by the ESRC and the Leverhulme foundation. I am also grateful to Cowles Foundation for financial support.

* Tel.: +1 5143432400.
E-mail address: ilze.kalnina@umontreal.ca.

[1] See Eq. (6) for the definition of realized variance.

0304-4076/$ – see front matter © 2010 Elsevier B.V. All rights reserved.
doi:10.1016/j.jeconom.2010.12.011

nature of realized variance and the simplicity of its calculation have made it popular among practitioners. It has been used for asset allocation (Fleming et al., 2003), forecasting of Value at Risk (Giot and Laurent, 2004), evaluation of volatility forecasting models (Andersen and Bollerslev, 1998), and other purposes. The Chicago Board Options Exchange (CBOE) started trading S&P 500 Three-Month realized volatility options on October 21, 2008. Over the counter, these and other derivatives written on RV have been traded for several years. These financial products allow one to bet on the direction of the volatility, or to hedge against exposure to volatility. One way of pricing these derivatives is by using the theory of quadratic variation.

Suppose the log-price X_t follows a Brownian semimartingale process,

$$dX_t = \mu_t dt + \sigma_t dW_t, \tag{1}$$

where μ, σ, and W are the drift, volatility, and Brownian Motion processes, respectively. Our interest is in estimating volatility over some interval, say one day, which we normalize to be [0, 1]. The quantity of interest is captured by integrated variance, or quadratic variation over the interval, which is defined as

$$IV = \int_0^1 \sigma_s^2 ds.$$

Realized variance is a consistent estimator of integrated variance in infill asymptotics, i.e., when the approximation is made as the time distance between adjacent observations shrinks to zero. According to this approximation, therefore, the estimation error in RV should be smaller for even higher frequency data than 5 min. Ironically, this is not the case in practice. For the highest frequencies, the data is more and more clearly affected by the bid–ask spread and other market microstructure frictions, rendering the semimartingale model inapplicable and RV inconsistent. Zhou (1996) proposed to model high frequency data as a Brownian semimartingale with an additive measurement error. This model can reconcile the main stylized facts of prices both in moderate and high frequencies. Zhang et al. (2005) were the first to propose a consistent estimator of integrated variance in this model, in the presence of i.i.d. microstructure noise, which they named the Two Scales Realized Volatility (TSRV) estimator; it is also known as the Two Time Scale estimator in the literature.[2] Consistent estimators in this framework were also proposed by Barndorff-Nielsen et al. (2008), Christensen et al. (2010, 2009) and Jacod et al. (2009). Aït-Sahalia et al. (2011) extend the TSRV estimator to the case of stationary autocorrelated microstructure noise, but do not propose an inference method. The problem with inference arises from the complicated structure of the asymptotic variance of the TSRV estimator. The method proposed in this paper can be used to conduct inference for the Two Time Scales estimator in the presence of not only autocorrelated, but also heteroscedastic measurement error. This allows the model to accommodate the stylized fact in the empirical market microstructure literature about the U-shape in observed returns and spreads.[3]

This new subsampling scheme is useful in practice when available estimators of the asymptotic variance are complicated and hence present difficulties in constructing confidence intervals. In such cases, a common procedure is to estimate the asymptotic variance as a sample variance of the bootstrap estimator. It turns out that this procedure is inconsistent for noisy high frequency data (see Section 2.1).

The subsampling method of Politis and Romano (1994) has been shown to be useful in many situations as a way of conducting inference under weak assumptions and without utilizing knowledge of limiting distributions. The basic intuition for constructing an estimator of the asymptotic variance is as follows. Imagine the standard setting of discrete time with long-span (also called increasing domain) asymptotics. Take some general estimator $\widehat{\theta}_n$ (think of i.i.d. $Y_i's$, a parameter of interest $\theta = E(Y)$, and $\widehat{\theta}_n = \frac{1}{n}\sum Y_i$). Suppose we know its asymptotic distribution

$$\tau_n(\widehat{\theta}_n - \theta) \Longrightarrow N(0, V)$$

as $n \to \infty$, where $\Longrightarrow$ denotes convergence in distribution, and τ_n is the rate of convergence when n observations are used. Suppose we would like to estimate V, in order to be able to construct confidence intervals for $\widehat{\theta}_n$. This can be done with the help of many subsamples, for which the estimator $\widehat{\theta}_n$ has the same asymptotic distribution. In particular, suppose we construct K different subsamples of $m = m(n)$ consecutive observations, starting at different values (whether they are overlapping or not is irrelevant here), where $m = m(n) \to \infty$ as $n \to \infty$ but $m/n \to 0$. Denote by $\widehat{\theta}_{n,m,l}$ the estimator $\widehat{\theta}_n$ calculated using the lth block of m observations, with n being the total number of observations. Then, the asymptotic distribution of $\tau_m(\widehat{\theta}_{n,m,l} - \theta)$ is the same, i.e.,

$$\tau_m\left(\widehat{\theta}_{n,m,l} - \theta\right) \Longrightarrow N(0, V) \tag{2}$$

for each subsample l, $l = 1, \ldots, K$. Hence, V can be estimated by the sample variance of $\tau_m\widehat{\theta}_{n,m,l}$ (with centering around $\widehat{\theta}_n$, a proxy for the true value θ). This yields the following estimator of V

$$\widehat{V} = \tau_m^2 \times \frac{1}{K}\sum_{l=1}^{K}\left(\widehat{\theta}_{n,m,l} - \widehat{\theta}_n\right)^2, \tag{3}$$

and we have

$$\widehat{V} \xrightarrow{p} V,$$

where $\xrightarrow{p}$ denotes convergence in probability. Notice that the estimator in (3) is like an average of squared $\tau_m\left(\widehat{\theta}_{n,m,l} - \theta\right)$ over all subsamples, except that $\widehat{\theta}_n$ plays the role of θ. The difference between $\widehat{\theta}_n$ and θ is negligible because $\widehat{\theta}_n$ converges faster to θ than $\widehat{\theta}_{n,m,l}$ does.

It is shown that a direct application of the above method to the high frequency framework fails. This fact is illustrated for the RV example in model (1). That is, $\widehat{\theta}_n$ is taken to be realized variance and θ its probability limit, integrated variance. The intuition behind the failure is straightforward. The problem is that $\widehat{\theta}_{n,m,l}$ and $\widehat{\theta}_n$ do not converge to the same quantity and so (2) cannot be satisfied. The underlying reason is that the spot (or infinitesimal) volatility σ_t is changing over time. The estimator calculated on a small block cannot estimate the integrated variance θ, because θ contains information about spot volatility on the whole interval.

Politis et al. (1997) show, in the long span asymptotic framework, that the traditional subsampling scheme is valid under weaker assumptions than stationarity. Instead of stationarity, they assume that the normalized $\widehat{\theta}_{n,m,l}$ is on average close to the limiting distribution of $\widehat{\theta}_n$. This allows for, e.g., considerable local heteroscedasticity. However, in an infill asymptotic framework, changes of volatility and its moments over time are not local in nature. Lahiri (1996) illustrates the problems infill asymptotics creates by proving inconsistency of some commonly-used estimators under this asymptotic scheme.

A novel subsampling scheme is proposed that can estimate the asymptotic variance of RV. Importantly, it can also be applied to the Two Time Scales estimator of Aït-Sahalia et al. (2011), in

[2] A note on terminology: many authors have called TSRV the subsampling estimator of IV. It is very different from, and should not be confused with, the subsampling method of Politis et al. (1999).

[3] See Andersen and Bollerslev (1997), Gerety and Mulherin (1994), Harris (1986), Kleidon and Werner (1996), Lockwood and Linn (1990) and McInish and Wood (1992).

the presence of autocorrelated measurement error with diurnal heteroscedasticity. There are no alternative inferential methods available in the literature for this case. Moreover, this subsampling scheme can, under some conditions, estimate the asymptotic variance of a general class of estimators, which includes many estimators of the integrated variance.

The remainder of this paper is organized as follows. Section 2 describes the usual subsampling method of Politis and Romano (1994) and proposes a new subsampling method. It also introduces an alternative scheme that is robust to jumps in volatility. Section 3 shows how inference can be conducted for the Two Time Scales estimator in the presence of autocorrelated and heteroscedastic microstructure noise. Section 4 applies the subsampling method to a general class of estimators. Section 5 investigates the numerical properties of the proposed method in a set of simulation experiments. Section 6 applies the method to high frequency stock returns. Section 7 concludes.

2. Description of resampling schemes

The aim of this section is to motivate and introduce a new subsampling scheme in a relatively simple framework. Since the proposed method does not change across models or estimators, the motivation and intuition is given for the example of realized volatility in the absence of any market microstructure noise.

We first describe the setting for the realized volatility example. Suppose that log-price X_t is the following Brownian semimartingale process

$$\mathrm{d}X_t = \mu_t \mathrm{d}t + \sigma_t \mathrm{d}W_t, \tag{4}$$

where W_t is standard Brownian motion, the stochastic process μ_t is locally bounded, and σ_t is a càdlàg spot volatility process.[4] Suppose that we have observations on X on the interval $[0, T]$, where T is fixed. Without loss of generality set $T = 1$. Assume observation times are equidistant, so that the distance between observations is $1/n$. The asymptotic scheme is infill as $n \to \infty$.

Suppose the quantity of interest is integrated variance (also called integrated volatility),

$$IV = \int_0^1 \sigma_s^2 \mathrm{d}s. \tag{5}$$

IV is a random variable depending on the realization of the volatility path $\{\sigma_t, t \in [0, 1]\}$. The usual estimator of IV is the realized variance (often called realized volatility)

$$RV_n = \sum_{i=1}^{n} \left(X_{i/n} - X_{(i-1)/n}\right)^2. \tag{6}$$

This satisfies

$$\sqrt{n}\,(RV_n - IV) \Longrightarrow MN(0, V) \tag{7}$$

$$V = 2IQ = 2\int_0^1 \sigma_s^4 \mathrm{d}s$$

where $MN(0, V)$ denotes a mixed normal distribution with random conditional variance V independent of the underlying normal distribution.[5] The convergence (7) follows from Barndorff-Nielsen and Shephard (2002) and Jacod (2008), and is stable in law, see Aldous and Eagleson (1978). Stable convergence is slightly stronger than the usual convergence in distribution. Stable asymptotics are particularly convenient because it permits division of both sides of (7) by the square root of any consistent estimator of V to obtain a standardized asymptotic distribution for conducting inference on RV_n.

In fact, for the realized variance example, inference can be conducted relatively easily. Barndorff-Nielsen and Shephard (2002) propose to estimate V as twice the realized quarticity, $\widetilde{V} = 2IQ_n$, where realized quarticity is the sum of fourth powers of returns, properly scaled,

$$IQ_n = \frac{n}{3}\sum_{i=1}^{n} \left(X_{i/n} - X_{(i-1)/n}\right)^4. \tag{8}$$

The estimator $\widetilde{V}$ is consistent for V in the sense that $\widetilde{V}/V \xrightarrow{p} 1$. This result allows the construction of consistent confidence intervals for IV. For example, a two-sided level $1-\alpha$ interval is given by $\widetilde{\mathcal{C}}_\alpha = RV_n \pm z_{\alpha/2}\widetilde{V}^{1/2}/\sqrt{n}$, where z_α is the α quantile from a standard normal distribution, and this has the property that $\Pr[IV \in \widetilde{\mathcal{C}}_\alpha] \to 1-\alpha$. Mykland and Zhang (2009) have proposed an alternative estimator of V that is more efficient than $\widetilde{V}$ under the sampling scheme (4) and can also be used to construct intervals based on the studentized limit theory.

2.1. Failure of the traditional resampling schemes

Recently, Gonçalves and Meddahi (2009) have proposed a bootstrap algorithm for RV, in the setting of no noise. They use the i.i.d. and wild bootstrap applied to studentized RV. They show that resampling the studentized RV gives confidence intervals for RV with better properties than the $2IQ_n$ estimator of asymptotic variance. Their procedure relies on an estimator of the asymptotic variance, which is not always available. A more widely used bootstrap procedure would be to estimate asymptotic variance as the sample variance of the bootstrap statistic. This procedure is simple, but only consistent for the wild bootstrap with certain external random variables. Podolskij and Ziggel (2007) show that, to first order, all methods proposed by Gonçalves and Meddahi (2009) apply in exactly the same way to the Bipower Variation estimator.

All the above bootstrap methods become inconsistent in the presence of any market microstructure noise. While the current section also keeps this simplifying assumption for expositional purpose, Section 3 shows robustness of the proposed subsampling estimator to the market microstructure noise, which enables its application to data at the highest frequencies.

We now consider the popular method of Politis and Romano (1994). This subsampling scheme fails in our setting due to variability of the volatility over time. It is however instructive to consider, as subsequently proposed methods use a similar underlying idea.

Let $\widehat{\theta}_n$ be the RV calculated on the full sample, and let $\widehat{\theta}_{n,m,l}$ be the RV calculated on the lth block of m observations,[6]

$$\widehat{\theta}_{n,m,l} = \sum_{i=m(l-1)}^{ml} \left(X_{i/n} - X_{(i-1)/n}\right)^2,$$

see Fig. 1. In the above, $0 < l \le K$, where K is the number of subsamples, $K = \lfloor n/m \rfloor$.

Assumption 5.3.1 of Politis et al. (1999) is satisfied, i.e., the sampling distribution of $\tau_n(\widehat{\theta}_n - \theta)$ converges weakly. Therefore,

[4] In other words, the sample paths of the volatility process are left continuous with right limits.

[5] In other words, the limiting p.d.f. is of the form $f(x) = \int \phi_{0,v}(x) f_V(v)\mathrm{d}v$, where f_V denotes the p.d.f. of V and $\phi_{0,v}(x) = \exp(-x^2/2v^2)/\sqrt{2\pi v}$.

[6] For simplicity, all subsampling schemes in this paper are presented with non-overlapping subsamples. However, it is inconvenient to display non-overlapping subsamples in figures, so Figs. 1–3 show maximum overlap versions of the subsampling schemes.

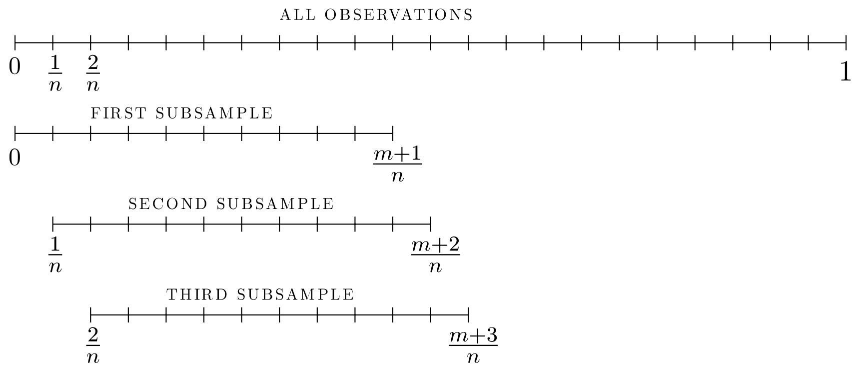

Fig. 1. The subsampling scheme of Politis and Romano (1994).

in the setting of stationary and mixing processes, V should be approximated well by

$$\widehat{V}_{PR} = m \times \frac{1}{K}\sum_{l=1}^{K}\left(\widehat{\theta}_{n,m,l} - \widehat{\theta}_n\right)^2.$$

However, in our setting, it is easy to see that $\widehat{V}_{PR}$ does not converge to V.

Proposition 1. *Let X satisfy (4) and $\widehat{\theta}_n$ be the realized variance defined in (6). Let $m \to \infty$ and $m/n \to 0$ as $n \to \infty$. Then,*

$$\widehat{V}_{PR} - m\theta^2 = o_p(m).$$

The estimator on the full sample converges to the true value, $\widehat{\theta}_n \to^p \theta$. On the other hand, the estimator on a subsample converges to zero. This is because each high frequency return is of order $n^{-1/2}$, so a sum of m squared returns is of order $m/n \to 0$. Therefore, Proposition 1 obtains that $\widehat{V}_{PR}$ is asymptotically equal to $m\theta^2$. Notice that the value θ^2 is not related to V, which is the parameter of interest.

Different orders of magnitude of $\widehat{\theta}_{n,m,l}$ and $\widehat{\theta}_n$ could be accounted for by using $\frac{m}{n}\widehat{\theta}_{n,m,l}$ instead of $\widehat{\theta}_{n,m,l}$ as in

$$\widehat{V}'_{PR} = m \times \frac{1}{K}\sum_{l=1}^{K}\left(\frac{n}{m}\widehat{\theta}_{n,m,l} - \widehat{\theta}_n\right)^2.$$

However, it still holds that $\frac{n}{m}\widehat{\theta}_{n,m,l} - \widehat{\theta}_n \overset{p}{\nrightarrow} 0$ so long as the spot volatility changes over time. This is because $\frac{m}{n}\widehat{\theta}_{n,m,l}$ estimates the spot variance $\sigma^2(\cdot)$ at some point, instead of the integrated variance θ.[7] Assuming no drift, no leverage, and sufficiently smooth volatility sample paths, one can show that the resulting estimator has a diverging bias, conditional on the volatility sample path,

$$\mathrm{E}\widehat{V}'_{PR} - V = m\left(IQ - IV^2\right) + o(1). \tag{9}$$

Therefore, the underlying reason for the failure of the subsampling method of Politis and Romano is the fact that the spot volatility changes over time. The latter effect is captured by the term $IQ - IV^2$ in Eq. (9) above, which is zero if and only if volatility is constant over the whole interval [0, 1].

An intuitive alternative would be to sample prices at some lower frequency instead of taking a sub-block of consecutive high frequency observations. In a way, sub-blocks are mimicking the long span asymptotic scheme, and the infill asymptotic scheme equivalent would be subsamples formed by lower frequency prices. Thus, for example, $\widehat{\theta}_{n,m,1}$ would be RV calculated with 5 min returns starting with the first second, $\widehat{\theta}_{n,m,2}$ would be RV calculated with 5 min returns starting with the second, and so on.

[7] For estimation of the spot variance using realized variance on a shrinking interval, see Foster and Nelson (1996), Andreou and Ghysels (2002), Mikosch and Starica (2003) and Kristensen (2010).

It then holds that $\widehat{\theta}_{n,m,l} - \widehat{\theta}_n \overset{p}{\to} 0$, $\forall l$. This, however, is not sufficient for consistency of $\widehat{V}$. The problem is that for every n, any two subsample estimators would be highly correlated. The resulting $\widehat{V}$ would be asymptotically unbiased, but inconsistent.

2.2. The new subsampling scheme

We now introduce and explain the new subsampling scheme. The current subsection describes this scheme for the RV example, and Section 3 applies it to the Two Time Scales estimator. Section 4 applies this subsampling scheme to a more general class of estimators.

In the subsampling scheme of Politis and Romano (1994), the problem was that the estimator on a subsample $\widehat{\theta}_{n,m,l}$ was centered at "the wrong quantity". In the formula

$$\widehat{V}_{PR} = m \times \frac{1}{K}\sum_{l=1}^{K}\left(\widehat{\theta}_{n,m,l} - \widehat{\theta}_n\right)^2,$$

the quantity $\widehat{\theta}_n$ plays the role of θ, but the problem is that the leading term in $\widehat{\theta}_{n,m,l}$ is integrated variance over a shrinking interval,

$$\theta_l = \int_{(l-1)m/n}^{lm/n} \sigma_u^2 \mathrm{d}u. \tag{10}$$

Thus, $\widehat{\theta}_{n,m,l}$ either converges to zero or the spot volatility depending on whether it is scaled by n/m, but in any case it cannot estimate θ, the integrated volatility over the whole interval [0, 1]. Therefore, $\widehat{\theta}_{n,m,l} - \widehat{\theta}_n$ does not converge to zero, causing $\widehat{V}_{PR}$ to explode.

Consider an alternative approach. We aim to center estimators at θ_l (as defined by Eq. (10)), in order to extract the information about the variance of $\widehat{\theta}_{n,m,l}$. The leading term of the variance of $\widehat{\theta}_{n,m,l}$ is

$$V_l = 2\int_{(l-1)m/n}^{lm/n} \sigma_u^4 \mathrm{d}u.$$

It is of course not equal to V, which we want to estimate, but we can use the fact that these add up to V over subsamples,

$$V = 2\int_0^1 \sigma_u^4 \mathrm{d}u = \sum_{l=1}^{K} V_l.$$

Given the additive structure of $\widehat{V}$, this approach can still give a consistent estimator of V, despite volatility changing over time. The only question left is, how to obtain an estimator of the centering factor θ_l. So consider using two subsamples, one with length J and one with length m, such that J is of smaller order than m. Then, both $\frac{n}{m}\widehat{\theta}_{n,m,l}$ and $\frac{n}{J}\widehat{\theta}_{n,J,l}$ estimate the spot variance, but they have different convergence rates. This in turn means one can be used to center the other. To simplify the presentation, we use the notation $\widehat{\theta}_l^{\text{long}}$ and $\widehat{\theta}_l^{\text{short}}$ instead of $\widehat{\theta}_{n,m,l}$ and $\widehat{\theta}_{n,J,l}$.

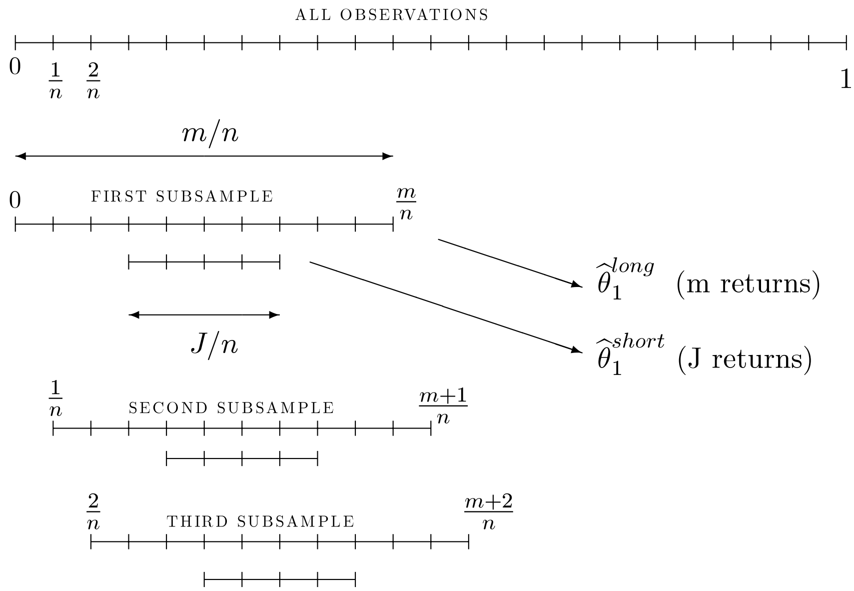

Fig. 2. The new subsampling scheme.

Since the rate of convergence of $\frac{n}{J}\widehat{\theta}^{\text{short}}$ is $\sqrt{J}$, the estimator of V becomes

$$\widehat{V}_{\text{sub}} = J \times \frac{1}{K}\sum_{l=1}^{K}\left(\frac{n}{J}\widehat{\theta}_l^{\text{short}} - \frac{n}{m}\widehat{\theta}_l^{\text{long}}\right)^2 \tag{11}$$

where $K = \lfloor n/m \rfloor$. $\widehat{\theta}_l^{\text{short}}$ and $\widehat{\theta}_l^{\text{long}}$ are realized variances calculated on the short subsample with J observations, and the long subsample with m observations. Fig. 2 provides a graphical illustration. The corresponding time intervals used are $\left[\frac{(l-1)m}{n}, \frac{(l-1)m+J}{n}\right]$ and $\left[\frac{(l-1)m}{n}, \frac{lm}{n}\right]$, so the expressions for estimators on subsamples become

$$\widehat{\theta}_l^{\text{short}} = \sum_{i=1}^{J}\left(X_{\frac{(l-1)m+i}{n}} - X_{\frac{(l-1)m+i-1}{n}}\right)^2$$

$$\widehat{\theta}_l^{\text{long}} = \sum_{i=1}^{m}\left(X_{\frac{(l-1)m+i}{n}} - X_{\frac{(l-1)m+i-1}{n}}\right)^2.$$

For an arbitrary volatility process, $nJ^{-1}\widehat{\theta}_l^{\text{short}}$ and $nm^{-1}\widehat{\theta}_l^{\text{long}}$ cannot be guaranteed to be close. For example, if the volatility process has a large jump on the interval covered by $\widehat{\theta}_l^{\text{long}}$, but not covered by $\widehat{\theta}_l^{\text{short}}$, then $nJ^{-1}\widehat{\theta}_l^{\text{short}}$ and $nm^{-1}\widehat{\theta}_l^{\text{long}}$ can differ substantially. Therefore, some kind of smoothness condition on the volatility paths is needed. Importantly, we do not require differentiable sample paths. It can be shown that a sufficient condition is to assume that volatility itself evolves like a Brownian semimartingale. This is a common way of modeling volatility in practice.

Assumption A1. The volatility process $\{\sigma_t, t \in [0, 1]\}$ is a Brownian semimartingale of the form

$$\mathrm{d}\sigma_t = \tilde{\mu}_t \mathrm{d}t + \tilde{\sigma}_t \mathrm{d}\widetilde{W}_t$$

where $\widetilde{W}_t$ is standard Brownian motion, the stochastic process $\widetilde{\mu}_t$ is locally bounded and the stochastic process $\widetilde{\sigma}_t$ is càdlàg.

Proposition 2. *Suppose (A1) holds and X satisfies (4). Let $\widehat{\theta}_n$ be the realized variance defined in (6), $m \to \infty$, $J \to \infty$, $m/n \to 0$, $J/m \to 0$, and $mJ^2/n \to 0$ as $n \to \infty$. Then,*

$$\widehat{V}_{\text{sub}} \xrightarrow{p} V.$$

The cost of not relying on the exact expression of V is that the proposed method is data intensive. J should be large enough for $\widehat{\theta}_J$ to have reasonable finite sample properties. As can be seen from the conditions above, m should be even larger, and n, the total number of observations, should be much larger than J.

Sections 3 and 4 show that Proposition 2 can be extended to more general settings than RV in a Brownian semimartingale model. This is because the subsampling method does not rely on the exact form of V, which it estimates.

2.3. An alternative subsampling scheme

The new estimator introduced in the previous section, $\widehat{V}_{\text{sub}}$, has the disadvantage that it does not allow for jumps in the volatility. The current section presents an alternative subsampling scheme that allows for such jumps.[8]

This subsampling scheme is illustrated in Fig. 3. On every block of m observations, calculate the estimator $\widehat{\theta}^n$ twice as follows. First, calculate it using all m observations, and denote it as $\widehat{\theta}_l^{\text{fast}}$. Then, calculate the estimator $\widehat{\theta}^n$ using every Qth price observation in the block of m observations, and denote it as $\widehat{\theta}_l^{\text{slow}}$.[9]

Now, $\widehat{\theta}_l^{\text{fast}}$ can be used to center the $\widehat{\theta}_l^{\text{slow}}$, because they both converge to (10), and because $\widehat{\theta}_l^{\text{fast}}$ converges to (10) faster than $\widehat{\theta}_l^{\text{slow}}$ does. The new estimator of V becomes

$$\widehat{V}'_{\text{sub}} = \frac{m}{Q} \times \frac{1}{K}\sum_{l=1}^{K}\left(\frac{n}{m}\widehat{\theta}_l^{\text{slow}} - \frac{n}{m}\widehat{\theta}_l^{\text{fast}}\right)^2$$

$$= \frac{n}{Q}\sum_{l=1}^{n/m}\left(\widehat{\theta}_l^{\text{slow}} - \widehat{\theta}_l^{\text{fast}}\right)^2$$

8 We conjecture that this alternative subsampling scheme is also robust to jumps in the price process in those special cases when these jumps do not appear in the expression of V. One example of such a case is the multipower variation when the sum of all powers is smaller than one, see Barndorff-Nielsen et al. (2005).

9 The subsampling scheme is similar in structure to the one in Lahiri et al. (1999). They similarly use two grids for subsampling to predict stochastic cumulative distribution functions in a spatial framework. However, they assume that the underlying process is stationary and their asymptotic framework is mixed infill and increasing domain.

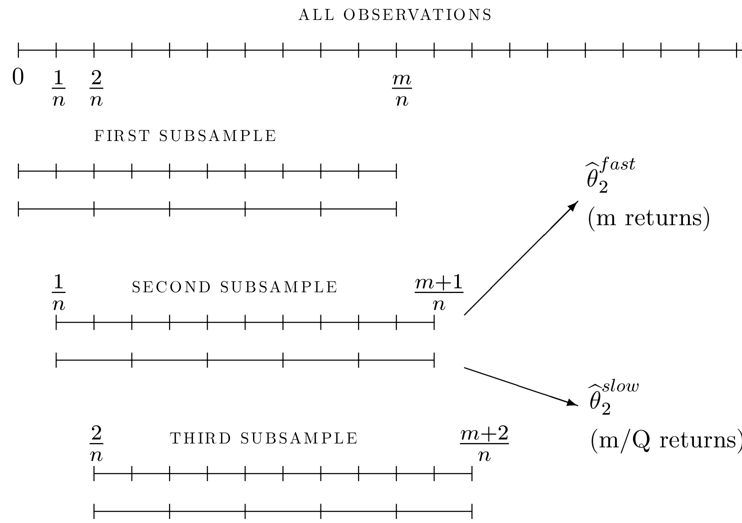

Fig. 3. An alternative subsampling scheme.

where

$$\widehat{\theta}_k^{\text{fast}} = \sum_{i=1}^{m} \left(X_{\frac{i+m(k-1)}{n}} - X_{\frac{i-1+m(k-1)}{n}} \right)^2$$

$$\widehat{\theta}_k^{\text{slow}} = \sum_{i=1}^{\lfloor m/Q \rfloor} \left(X_{\frac{iQ+m(k-1)}{n}} - X_{\frac{(i-1)Q+m(k-1)}{n}} \right)^2 .$$

The consistency result does not need Assumption A1 anymore.

Proposition 3. *Suppose X satisfies (4). Let $m \to \infty$, $Q \to \infty$, $m/n \to 0$, and $Q/m \to 0$ as $n \to \infty$. Then,*

$$\widehat{V}'_{\text{sub}} \xrightarrow{p} V. \tag{12}$$

The new estimator $\widehat{V}'_{\text{sub}}$ uses sparse data and hence cannot capture autocorrelated market microstructure noise. Since the Two Scales estimator with autocorrelated noise is the focus of this paper, $\widehat{V}'_{\text{sub}}$ is not used beyond the current section.

3. Inference for the two scales realized volatility estimator

This section shows how the new subsampling scheme can be applied to the Two Time Scales estimator of integrated variance proposed by Aït-Sahalia et al. (2011). Although only this example is discussed in detail, this subsampling scheme could also be applied to other integrated variance estimators in the presence of market microstructure noise, such as Multiscale estimator of Zhang (2006), Realized Kernels of Barndorff-Nielsen et al. (2008), and the pre-averaging estimator of Jacod et al. (2009).

Stock price data at highest frequencies is well known to be affected by market microstructure noise. For example, trades are not executed in practice at the efficient price. Typically, they are executed either at the prevailing bid or ask price. Therefore, observed transaction prices alternate between bid and ask prices (the so-called bid–ask bounce), creating negative autocorrelation in observed returns, which is a stylized fact in high frequency data. This was the motivation for Zhou (1996) to introduce an additive market microstructure noise model where the observed log-price Y is a sum of a Brownian semimartingale component X and an i.i.d. noise ϵ,

$$Y_t = X_t + \epsilon_t. \tag{13}$$

In this model, observed log-returns display negative first order autocovariance,

$$\begin{aligned}
&\text{Cov}\left(\Delta Y_{i/n}, \Delta Y_{(i-1)/n}\right) \\
&\quad = \text{Cov}\left(\Delta X_{i/n} + \epsilon_{i/n} - \epsilon_{(i-1)/n}, \Delta X_{(i-1)/n} + \epsilon_{(i-1)/n} - \epsilon_{(i-2)/n}\right) \\
&\quad = -\text{Var}\left(\epsilon_{(i-1)/n}\right).
\end{aligned} \tag{14}$$

Another stylized fact is that realized variances calculated at the highest frequencies become very large. This is in contradiction to the Brownian semimartingale model, where RV has roughly the same expectation irrespective of the frequency at which it is calculated. Also, RV should converge to *IV* when higher and higher frequencies are used. This difficulty lies behind the underlying reason for the common practice not to calculate realized variance at higher frequencies than 5 or 15 min. The problem with this approach is that it implies discarding most of the available data. There are only 72 five minute returns in a day, and only 24 fifteen minute returns in a day, while the available high frequency data is usually measured in thousands. In order to be able to use all the available data, one has to work with a model that can accommodate the above stylized facts.

Zhang et al. (2005) were the first to introduce a consistent estimator of integrated variance of the efficient price *IV* within the additive measurement error model of Zhou (1996), see Eq. (13) above. The noise ϵ_t is i.i.d., zero mean with variance $\text{Var}(\epsilon) = \omega^2$ and $\text{E}\epsilon^4 < \infty$, and independent from the latent log-price X_t. In this model, Zhang et al. (2005) propose the following consistent estimator for the integrated variance of X_t,

$$\widehat{\theta}_n = [Y, Y]^{(G_1)} - \frac{\overline{n}_{G_1}}{n} [Y, Y]^{(1)}, \tag{15}$$

where, for any parameter b,

$$[Y, Y]^{(b)} = \frac{1}{b} \sum_{i=1}^{n-b} \left(Y_{(i+b)/n} - Y_{i/n} \right)^2$$

$$\overline{n}_b = \frac{n-b+1}{b}.$$

Notice that $[Y, Y]^{(1)}$ coincides with the RV estimator, while $[Y, Y]^{(G_1)}$ consists of lower frequency returns. In particular, $[Y, Y]^{(G_1)}$ consists of returns calculated from prices that are G_1 high frequency observations apart. Thus, time distance is n^{-1} between high frequency observations and $G_1 n^{-1}$ between lower frequency observations. In empirical applications, a common choice for G_1 is such that the lower frequency returns are sampled at 5 min. Zhang et al. (2005) call the above estimator the Two Scales Realized

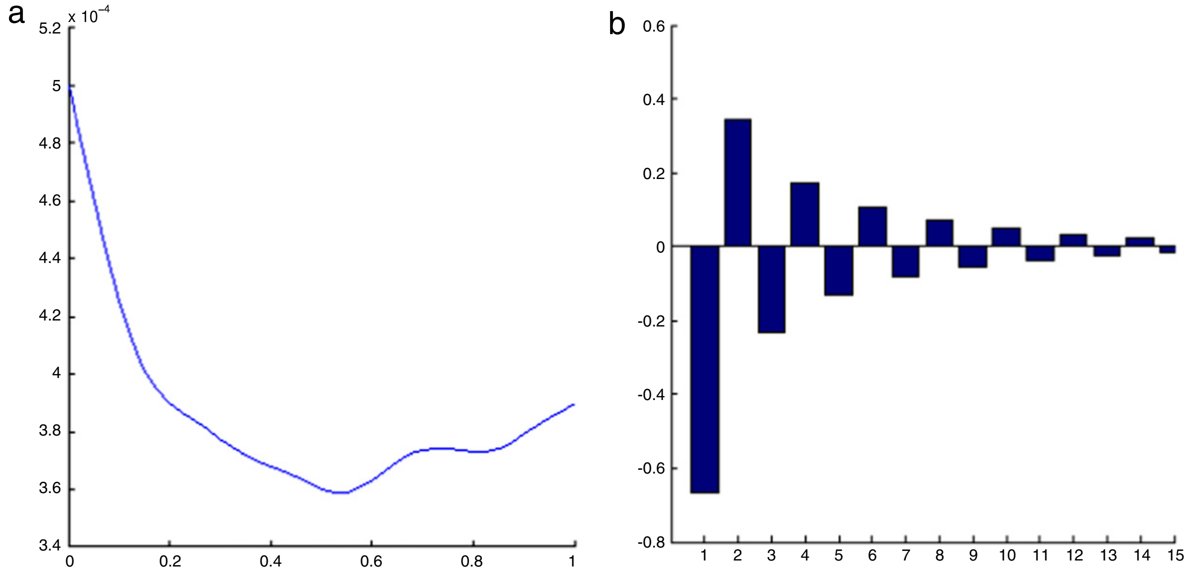

Fig. 4. Properties of returns of Microsoft (MSFT) stock. Returns are constructed from transaction prices over the whole year 2006. See Section 6 for data cleaning procedures. Panel (a) shows the estimated heteroscedasticity function $\omega(\cdot)$, averaged over all days in 2006. Panel (b) shows the autocorrelogram of returns calculated in tick time.

Volatility (TSRV) estimator. They derive the following asymptotic distribution of the estimator,

$$n^{1/6}\left(\widehat{\theta}_n - \theta\right) \Rightarrow \sqrt{V}Z$$

where the asymptotic (conditional) variance takes the form

$$V = \underbrace{c\frac{4}{3}\int_0^1 \sigma_u^4 \mathrm{d}u}_{\text{signal}} + \underbrace{8c^{-2}\omega^4}_{\text{noise}}, \tag{16}$$

i.e., it consists of a signal part, which is due to the efficient price, and a noise part. In the above, Z is a standard normal random variable, independent from V, and c is the constant in $G_1 = \lfloor cn^{2/3} \rfloor$. With i.i.d. noise, V can be estimated component by component. $\mathrm{Var}(\epsilon) = \omega^2$ can be estimated using the following estimator proposed by Bandi and Russell (2008),

$$\widehat{\omega^2} = \frac{RV}{2n} \xrightarrow{p} \omega^2.$$

We saw in Section 2 that in a model without noise, integrated quarticity $\int \sigma_u^4 du$ can be estimated by realized quarticity defined in (8). This becomes more difficult in the presence of noise. However, Barndorff-Nielsen et al. (2008) have proposed an estimator for $\int \sigma_u^4 \mathrm{d}u$, which is consistent in the presence of i.i.d. noise, see Section 5.

This model is for i.i.d. noise, so the noise is assumed to be homoscedastic. A well known stylized fact in the empirical market microstructure literature is that intradaily spreads (difference between bid and ask price) and intradaily stock price volatility are described typically by a U-shape (see footnote 3 for some references). In other words, prices are more volatile in mornings and afternoons than at noon; spreads are also larger in mornings and afternoons. Fig. 4(a) presents an estimate of heteroscedasticity function $\omega^2(\cdot)$ for transaction prices of Microsoft stock, averaged over all days in the year 2006. The diurnal variation is evident.

Kalnina and Linton (2008) introduce diurnal heteroscedasticity in the microstructure noise in model (13). Suppose the efficient log-price X is the same as above in (13), but the noise displays unconditional heteroscedasticity. In particular, suppose the noise ϵ_t satisfies

$$\epsilon_t = \omega(t) u_t \tag{17}$$

where $\omega(t)$ is a nonstochastic differentiable function of time t, and u_t is i.i.d. with $\mathrm{E}(u_t) = 0$, and $\mathrm{Var}(u_t) = 1$. As a result of this generalization, the asymptotic variance of $\widehat{\theta}_n$ changes to

$$V = \underbrace{c\frac{4}{3}\int_0^1 \sigma_u^4 \mathrm{d}u}_{\text{signal}} + \underbrace{8c^{-2}\int_0^1 \omega^4(u)\mathrm{d}u}_{\text{noise}}.$$

In this model, the previous estimator of the noise part of V ceases to be consistent as

$$\widehat{\omega^2} = \frac{RV}{2n} \xrightarrow{p} \int_0^1 \omega^2(u)\mathrm{d}u,$$

so, by Jensen's inequality, its square would be always strictly smaller than the target $\int \omega^4(u)\mathrm{d}u$ as long as there is any diurnal variation at all. Kalnina and Linton (2008) show that $\omega(\cdot)$ can be estimated at any fixed point τ using kernel smoothing,

$$\widetilde{\omega}^2(\tau) = \frac{1}{2}\sum_{i=1}^{n} K_h(t_{i-1} - \tau)\left(\Delta Y_{t_{i-1}}\right)^2.$$

In the above, h is a bandwidth that tends to zero asymptotically and $K_h(.) = K(./h)/h$, where $K(.)$ is a kernel function satisfying some regularity conditions. This suggests estimating the noise part of V by

$$8c^{-2}\int_0^1 \widetilde{\omega}^4(u)\mathrm{d}u.$$

As we saw earlier in (14), the i.i.d. measurement error model is consistent with negative first order autocorrelations in the observed returns. However, returns can sometimes exhibit autocorrelation beyond the first lag in practice. For example, Fig. 4(b) graphs the autocorrelogram of the returns of Microsoft stock for the whole year 2006. We see that Microsoft stock returns display strong negative autocorrelation well beyond the first lag. While the model (13) does generate a negative first autocorrelation, it implies that any further autocorrelations have to be zero. Since increments of a Brownian semimartingale are uncorrelated in time, any such autocorrelation has to be due to noise ϵ_t.[10]

[10] In a Brownian semimartingale model, the only source of autocorrelations of increments is drift, which is negligible for high frequencies.

Aït-Sahalia et al. (2011) generalize the i.i.d. measurement error model (13) in a different direction. They allow for autocorrelated stationary microstructure noise. In particular, they make the following assumption about the noise.

Assumption A2. The noise ϵ_{t_i} is independent from the efficient log-price process X_t, and it is (when viewed as a process in index i) stationary and strong mixing with the mixing coefficients decaying exponentially. Also, for some $\kappa > 0$, $\mathrm{E}\epsilon^{4+\kappa} < \infty$.

In model (13) with ϵ_t satisfying Assumption A2, Aït-Sahalia et al. (2011) propose the following consistent estimator for the integrated variance of X_t,

$$\widehat{\theta}_n = [Y, Y]^{(G_1)} - \frac{\overline{n}_{G_1}}{\overline{n}_{G_2}} [Y, Y]^{(G_2)} \tag{18}$$

where G_1 and G_2 satisfy the following assumption,

Assumption A3. The G_1 parameter of the Two Time Scales estimator $\widehat{\theta}_n$ defined by (18) satisfies $G_1 = \lfloor cn^{2/3} \rfloor$ for some constant c. The G_2 parameter is such that $\mathrm{Cov}(\epsilon_0, \epsilon_{G_2/n}) = o(n^{-1/2})$, $G_2 \to \infty$, $G_2/G_1 \to 0$.[11]

The Two Time Scales estimator defined by (18) is more general than the one in (15), which is a special case when $G_2 = 1$ and $G_1 \to \infty$ as $n \to \infty$. Aït-Sahalia et al. (2011) show that the new Two Time Scales estimator $\widehat{\theta}_n$ has the same asymptotic properties except it has a more complicated asymptotic variance,

$$V = \underbrace{c\frac{4}{3}\int_0^1 \sigma_u^4 \mathrm{d}u}_{\text{signal}} + \underbrace{8c^{-2}\mathrm{Var}\,(\epsilon)^2 + 16c^{-2}\lim_{n\to\infty}\sum_{i=1}^{n}\mathrm{Cov}\left(\epsilon_0, \epsilon_{i/n}\right)^2}_{\text{noise}}, \tag{19}$$

where c is the constant in $G_1 = \lfloor cn^{2/3} \rfloor$.

The literature does not provide any estimator of V or an alternative method for constructing confidence intervals for $\widehat{\theta}_n$. Here we can estimate the asymptotic variance of the Two Time Scales estimator $\widehat{\theta}_n$ using the subsampling scheme.

Theorem 4. *Suppose model (13) holds, and ϵ_{t_i} satisfy Assumption A2. Let $\widehat{\theta}_n$ be the Two Time Scales estimator defined by (18), with parameters G_1 and G_2 that satisfy Assumption A3. Let V be defined by (19). Let $J \to \infty$, $m \to \infty$, $J/m \to 0$, $m/n \to 0$, $G_1/J \to 0$, and $Jmn^{-5/3} \to 0$. Then,*

$$\widehat{V}_{\mathrm{sub}} \xrightarrow{p} V$$

where

$$\widehat{V}_{\mathrm{sub}} = Jn^{-2/3} \times \frac{1}{K}\sum_{l=1}^{K}\left(\frac{n}{J}\widehat{\theta}_l^{\mathrm{short}} - \frac{n}{m}\widehat{\theta}_l^{\mathrm{long}}\right)^2 \tag{20}$$

with $K = \lfloor n/m \rfloor$.

In the above, $\widehat{\theta}_l^{\mathrm{short}}$ is simply $\widehat{\theta}^n$ calculated on a smaller block of J observations inside the lth larger block of m observations, with exactly the same parameters G_1 and G_2 as $\widehat{\theta}^n$ uses. See Fig. 2 for an illustration. In particular,

$$\widehat{\theta}_l^{\mathrm{short}} = [Y, Y]_l^{(G_1)} - \frac{\bar{J}_{G_1}}{\bar{J}_{G_2}} [Y, Y]_l^{(G_2)}$$

where

$$[Y, Y]_l^{(G_i)} = \frac{1}{G_i}\sum_{i=1}^{J-G_i}\left(Y_{(l-1)m/n+(i+G_i)/n} - Y_{(l-1)m/n+i/n}\right)^2, \quad i = 1, 2$$

$$\bar{J}_{G_i} = \frac{J - G_i + 1}{G_i}, \quad i = 1, 2.$$

One obtains $\widehat{\theta}_l^{\mathrm{short}}$ by substituting J for m above. In Fig. 2, the version with maximum overlap is presented. In practice, it is much quicker to compute the no overlap version, for which Theorem 4 is formulated. While this does not alter the conclusion of Theorem 4, the maximum overlap version is slightly more efficient. In this case, $\widehat{V}_{\mathrm{sub}}$ is defined by (20) with $K = n - m + 1$.

To the author's knowledge, this is the only available method in the literature to construct confidence intervals for the Two Time Scales estimator when the noise is autocorrelated. Similarly, one can apply this method to the Multiscale estimator of Aït-Sahalia et al. (2011) when microstructure noise is autocorrelated. The advantage of using the Multiscale estimator is that it has the optimal rate of convergence $n^{1/4}$.

However, the above model of Aït-Sahalia et al. (2011) rules out any diurnal heteroscedasticity of the noise. When both autocorrelation and heteroscedasticity are taken into account, we have

Lemma 5. *Suppose the observed price satisfies $Y_{i/n} = X_{i/n} + \epsilon_{i/n}$ where the efficient log-price X_t follows a Brownian semimartingale process (4) and microstructure noise $\epsilon_{i/n}$ satisfies*

$$\epsilon_t = \omega(t)u_t$$

where $\omega(\cdot)$ is a differentiable, nonstochastic function of time, u_t satisfies Assumption A2 and $\mathrm{Var}\,(u_t) = 1$. Then, $\widehat{\theta}_n$ defined in (18) is such that

$$n^{1/6}\left(\widehat{\theta}_n - \theta\right) \Rightarrow \sqrt{V}Z$$

where

$$V = c\frac{4}{3}\int_0^1 \sigma_u^4 \mathrm{d}u + 8c^{-2}\int_0^1 \omega^4(u)\mathrm{d}u + 16c^{-2}\int_0^1 \omega^4(u)\mathrm{d}u \lim_{n\to\infty}\sum_{i=1}^{n}\mathrm{Cov}\left(\epsilon_0, \epsilon_{i/n}\right)^2.$$

In this case of autocorrelated and heteroscedastic noise, Theorem 4 easily generalizes and subsampling again delivers a consistent estimate of V. This is because both are special cases of the consistency result of the subsampling estimator in the general case, which is described in the next section. To estimate this more complicated V, exactly the same formula $\widehat{V}_{\mathrm{sub}}$ should be used as for the homoscedastic case. In this model, this is the only available method in the literature to construct confidence intervals for the Two Time Scales estimator.

Importantly, this section illustrates the robustness of the subsampling estimator of V across different sets of assumptions. Moreover, it is also easy to implement. All that is necessary is to compute $\widehat{\theta}_n$ on several sub-blocks of observations. We conjecture that the subsampling estimator $\widehat{V}_{\mathrm{sub}}$ would be consistent for V under even more general assumptions than considered above, for example, in the case when autocorrelations of the noise are changing through time, or when the efficient returns have fat tails as in Meddahi and Mykland (2010).

4. Inference for a general estimator

This section shows how to use the new subsampling scheme (as described in Sections 2.2 and 3) to conduct inference for a general class of estimators of volatility measures. A set of assumptions

[11] The restriction on $\mathrm{Cov}(\epsilon_0, \epsilon_{G_2/n})$ should be considered in the light of the fact that Assumption A2 implies that there exists a constant ϕ such that, for all i,

$$\left|\mathrm{Cov}\left(\epsilon_{i/n}, \epsilon_{(i+l)/n}\right)\right| \le \phi^l \mathrm{Var}\,(\epsilon).$$

is introduced and explained, under which subsampling delivers a consistent estimate of the asymptotic variance of an estimator $\widehat{\theta}_n$. As we shall see, there are two essential ingredients for the subsampling method to work. One is additivity over subsamples of the asymptotic variance of $\widehat{\theta}_n$. The second is that the asymptotic distribution of $\widehat{\theta}_n$ calculated on a block of observations is similar, in a sense explained below, to the asymptotic distribution of $\widehat{\theta}_n$ calculated using all available data.

We do not assume a specific process for X. It could be a pure diffusion or a diffusion contaminated with noise, as long as the regularity assumptions below are satisfied. All arguments in this section are made conditional on the volatility path $\{\sigma_u, u \in [0, 1]\}$. Suppose there is an estimator $\widehat{\theta}_n$, for which the asymptotic distribution is known to be as follows

$$\tau_n \left(\widehat{\theta}_n - \theta\right) \Rightarrow \sqrt{V} Z. \tag{21}$$

In the above, τ_n is a known rate of convergence of $\widehat{\theta}_n$. For example, $\tau_n = n^{1/2}$ for realized variance, $\tau_n = n^{1/6}$ for the Two Time Scales estimator. Z is a random variable that is known to satisfy $\mathrm{E}(Z) = 0$ and $\mathrm{Var}(Z) = 1$. A consistent estimator of V thus enables a researcher to construct consistent confidence intervals for $\widehat{\theta}_n$.

We recall the subsampling scheme introduced in Section 2.2. Divide the total number of returns into blocks of m consecutive returns. Thus, we obtain $\lfloor n/m \rfloor$ subsamples. Denote by $\widehat{\theta}_l^{\text{long}}$ the estimator $\widehat{\theta}_n$ calculated using all m returns of the lth block, $l = 1, \dots, \lfloor n/m \rfloor$. Denote by $\widehat{\theta}_l^{\text{short}}$ the estimator $\widehat{\theta}_n$ calculated using only J returns of the lth block, where $J < m$. See Fig. 2 in Section 2.2 for a graphical illustration.

In order to guarantee that $\frac{n}{J}\widehat{\theta}_l^{\text{short}}$ and $\frac{n}{m}\widehat{\theta}_l^{\text{long}}$ converge to the same quantity, despite being defined on different time intervals, we need to impose some smoothness on the volatility paths. In particular, we use the following assumption.

Assumption A4. (21) holds, where θ and V are the following functions of the volatility path $\{\sigma_u, u \in [0, 1]\}$,

$$\theta = \int_0^1 g_1 (\sigma(u))\, \mathrm{d}u$$
$$V = \int_0^1 g_2 (\sigma(u))\, \mathrm{d}u$$

where $g_1, g_2 \in C^1[0, 1]$ and σ is a Brownian semimartingale as in (4).

For example, we obtain integrated variance IV with $g_1(u) = \sigma^2(u)$ and the asymptotic variance of realized variance with $g_2(\sigma(u)) = 2\sigma^2(u)$.

The type of estimators that are likely to satisfy the assumptions of this section are those that are *approximately* additive over subsamples, i.e.,

$$\widehat{\theta}_n = \sum_{l=1}^{\lfloor n/m \rfloor} \frac{m}{J} \widehat{\theta}_l^{\text{short}} + o_p(1) \tag{22}$$

or

$$\widehat{\theta}_n = \sum_{l=1}^{\lfloor n/m \rfloor} \widehat{\theta}_l^{\text{long}} + o_p(1). \tag{23}$$

All currently available estimators of integrated variance and related quantities satisfy this additivity property. We also impose the following assumption, which ensures that estimators on subsamples are mixing.

Assumption A5. For any fixed n, the returns process $\left\{R_{i/n}^{(n)}\right\}_{i=1,\dots,n}$ with $R_{i/n}^{(n)} = X_{i/n} - X_{(i-1)/n}$ is strong mixing. Also, $\widehat{\theta}_n = \phi\left(R_{1/n}^{(n)}, R_{2/n}^{(n)}, \dots, R_1^{(n)}\right)$ where $\phi : \mathbb{R}^n \longmapsto \mathbb{R}$.

This is a rather strong assumption. For example, when X follows a semimartingale (4), this assumption rules out leverage effects. Why could leverage effects be allowed for in Proposition 2? Proposition 2 assumed that X follows a semimartingale. Therefore, after a discretization approximation, a proof could be based on the powerful martingale methods as in, e.g., Jacod and Shiryaev (2003). Here, however, market microstructure noise is allowed for, which is not a semimartingale. Therefore, without imposing more structure on the estimator (such as approximate additivity in X and the noise as in the Two Scales estimator example), this technique cannot be used.

As discussed in previous sections, $\widehat{\theta}_l^{\text{long}}$ and $\widehat{\theta}_l^{\text{short}}$ do not estimate θ, since they use only information about the volatility path on a small time interval, whereas the volatility is changing throughout the interval $[0, 1]$. Let us denote by θ_l^{long} and θ_l^{short} the respective quantities they estimate, and by V_l^{short} and V_l^{long} what can be thought of as their asymptotic variances. They can be defined as follows,

$$\begin{aligned}
\theta_l^{\text{short}} &= \int_{(l-1)m/n}^{[(l-1)m+J]/n} g_1(\sigma(u))\, \mathrm{d}u, \\
V_l^{\text{short}} &= \int_{(l-1)m/n}^{[(l-1)m+J]/n} g_2(\sigma(u))\, \mathrm{d}u \\
\theta_l^{\text{long}} &= \int_{(l-1)m/n}^{lm/n} g_1(\sigma(u))\, \mathrm{d}u, \\
V_l^{\text{long}} &= \int_{(l-1)m/n}^{lm/n} g_2(\sigma(u))\, \mathrm{d}u.
\end{aligned} \tag{24}$$

Finally, we make the following assumption,

Assumption A6. For every n, define θ_l^{short} and V_l^{short} by (24), and define a triangular array

$$\zeta_l^{(n)} = \frac{n}{J}\left[\tau_n^2 \left(\widehat{\theta}_l^{\text{short}} - \theta_l^{\text{short}}\right)^2 - V_l^{\text{short}}\right].$$

The array $\left\{\zeta_j^{(n)}\right\}$ satisfies the following conditions

(i)

$$\text{as } n \to \infty, \quad \sup_l \mathrm{E}\left(\zeta_l^{(n)}\right) \to 0.$$

(ii) $\left\{\zeta_j^{(n)}\right\}$ is L^p bounded for some $p > 1$.

We now discuss Assumption A6. Assumption A6(i) can be written equivalently as follows,

$$\text{as } n \to \infty, \quad \sup_l \mathrm{E}\left(\left(V_l^{\text{short}}\right)^{-1} \tau_n^2 \left(\widehat{\theta}_l^{\text{short}} - \theta_l^{\text{short}}\right)^2\right) \to 1,$$

as long as V_l^{short} is of order J/n. In other words, Assumption A6(i) requires that the square of the standardized statistic $\widehat{\theta}_l^{\text{short}}$ has asymptotic expectation one. On the full sample, we know from (21) that the standardized $\widehat{\theta}^n$ is asymptotically a random variable Z with $\mathrm{E}\left(Z^2\right) = 1$. Therefore, a sufficient condition for Assumption A6(i) to hold is that the asymptotic distribution of $\widehat{\theta}_l^{\text{short}}$ satisfies the same condition on a subsample. Roughly speaking, we need the estimator on a subsample, $\widehat{\theta}_l^{\text{short}}$, to behave similarly to the estimator on a full sample, $\widehat{\theta}_n$.

Assumption A6(ii) is a stronger assumption, and it illustrates the main idea of the subsampling method. Recall the basic idea of subsampling as described in the introduction of the paper. Roughly speaking, in a stationary world, the way subsampling estimates V is by constructing many random variables with V as their asymptotic variance. In our nonstationary case, continuity in time plays the role of stationarity as it ensures that the same feature in V is estimated by many subsamples. Assumption A6(ii) effectively imposes V_j^{short} to be of order J/n, i.e., that there is enough continuity in V with respect to time. Apart from this consideration, Assumption A6(ii) requires existence of moments. This is not an issue for a Brownian semimartingale model due to the local boundedness assumption on the drift and volatility functions, but becomes a constraint if X also contains other components. For example, consider a model where observations are sampled from a Brownian semimartingale with an additive noise ϵ. In this model, corresponding moments have to be assumed on ϵ for Assumption A6(ii) to hold. In the case of the Two Time Scales estimator discussed below, $L^{4+\varepsilon}$ boundedness of ϵ is needed, which is exactly what has been assumed by the authors of Two Time Scales estimator to derive its asymptotic distribution.

We have the following result.

Theorem 6. *Assume (A4), (A5), and (A6). Let $J \to \infty, m \to \infty$, $J/m \to 0$, $m/n \to 0$, and $Jm\tau_n^2 n^{-2} \to 0$. Then,*

$$\widehat{V}_{\text{sub}} \xrightarrow{p} V$$

where

$$\widehat{V}_{\text{sub}} = \frac{Jm}{n^2} \sum_{l=1}^{\lfloor n/m \rfloor} \tau_n^2 \left(\frac{n}{J} \widehat{\theta}_l^{\text{short}} - \frac{n}{m} \widehat{\theta}_l^{\text{long}} \right)^2 .$$

The assumption $Jm\tau_n^2 n^{-2} \to 0$ is determined by the smoothness of the volatility paths. Special cases show up in Proposition 2 and Theorem 4 with τ_n^2 being replaced by $\sqrt{n}$ and $n^{1/6}$, respectively. All these results assume that volatility follows a Brownian semimartingale (Assumption A1). If one assumed more smoothness, this assumption could be weakened.

Importantly, exactly the same formula is applied to all models and estimators, which satisfy the above assumptions. All that is necessary to calculate the estimator for V is to calculate the estimator $\widehat{\theta}^n$ on several subsamples, as well as to know the convergence rate τ_n. In particular, $\widehat{V}_{\text{sub}}$ simplifies to the formula for the realized variance in (11) with $\tau_n = \sqrt{n}$, and to the formula for the Two Time Scales estimator in (20) with $\tau_n = n^{1/6}$.

5. Simulation study

In this section numerical properties of the proposed estimator are studied for the example of the Two Time Scales estimator of Aït-Sahalia et al. (2011) in the case of i.i.d., autocorrelated, or heteroscedastic market microstructure noise.

The observed log-price Y_t is a sum of the efficient log-price X_t and noise u_t. The paths of the efficient log-price are simulated from the Heston (1993) model:

$$dX_t = (\alpha_1 - v_t/2)\,dt + \sigma_t dW_t$$

$$dv_t = \alpha_2 (\alpha_3 - v_t)\,dt + \alpha_4 v_t^{1/2} dB_t$$

where $v_t = \sigma_t^2$, W_t and B_t are independent Brownian motions. The parameters of the efficient log-price process X are chosen to be the same as in Zhang et al. (2005). They are $\alpha_1 = 0.05$, $\alpha_2 = 5$, $\alpha_3 = 0.04$, and $\alpha_4 = 0.5$, and they correspond to one year being a unit of time. For the sake of consistency, we keep these time units for the rest of the section. We simulate $n = 35,000$ observations over one week, i.e., five business days of 6.5 h each. This is motivated by the fact that GE stock has on average 35,000 observations per week in year 2006, see Section 6. We aim to estimate weekly integrated variance or $IV = \int_0^t \sigma_s^2 ds$ where t is one week or 1/50. The volatility path is fixed over simulations to facilitate comparisons. The volatility path used is plotted in Fig. 5. Varying the volatility path across simulations does not affect the theory nor the simulation results.

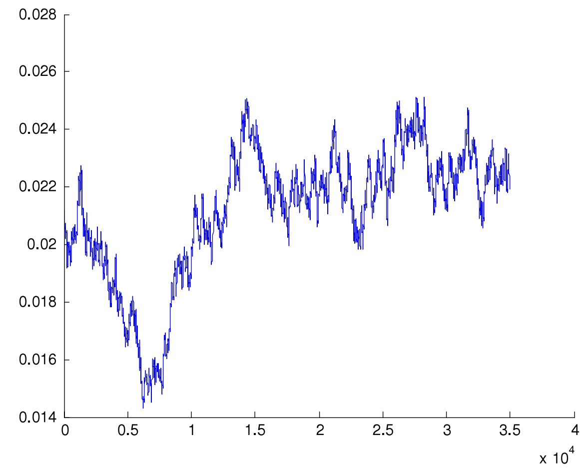

Fig. 5. Simulated volatility sample path.

Simulation of the noise u_t is described in Sections 5.1 and 5.2.

The parameters of the Two Time Scales estimator and of the subsampling procedure are chosen as follows. We set $G_1 = 100$, which in our data corresponds to 5 min lower frequency. This is a very popular choice in practice. We set $G_2 = 10$ in all autocorrelated noise simulations, and $G_2 = 1$ for heteroscedastic noise simulations. Two values of J are considered. The first is $J = 2G_1 = 200$, and the second is $J = 5G_1 = 500$. For m, three different values are considered, $m = 4J$, $10J$, and $15J$.

The literature does not propose ways of estimating the asymptotic variance of the Two Time Scales when noise is autocorrelated or diurnal. However, in the case of i.i.d. noise, there is an alternative, and this can serve as a benchmark for the simulation results. In the case of i.i.d. noise, the expression for asymptotic variance V of the Two Time Scales estimator is

$$V = c\frac{4}{3}t \int_0^t \sigma_u^4 du + 8c^{-2} [\text{Var}(u)]^2$$

and the alternative is to estimate each component of V separately. The easiest component to estimate is $[\text{Var}(u)]^2$. A popular estimator of $\text{Var}(u) = \omega^2$ is

$$\widehat{\omega^2} = \frac{RV}{2n}.$$

This has been proposed by, for example, Bandi and Russell (2006, 2008). To estimate integrated quarticity $IQ = t\int_0^t \sigma_u^4 du$ in the presence of noise is more difficult. A consistent estimator in the presence of i.i.d. noise, $\widehat{IQ}_{\text{BNHLS}}$, has been proposed by Barndorff-Nielsen et al. (2008).[12] Therefore, we can define the benchmark

12

$$\widehat{IQ}_{\text{BNHLS}}\left(\widetilde{\delta}, S\right) = \max\left[\left(\widehat{\theta}_n^*\right)^2, \frac{1}{\widetilde{n}} \sum_{j=1}^{\widetilde{n}} \widetilde{\delta}^{-2} \left(y_{j,\cdot}^2 - 2\widetilde{\omega^2}\right)\left(y_{j-2,\cdot}^2 - 2\widetilde{\omega^2}\right)\right],$$

where

$$y_{j,\cdot}^2 = \frac{1}{S} \sum_{s=0}^{S-1} \left(Y_{\widetilde{\delta}(j+\frac{s}{S})} - Y_{\widetilde{\delta}(j-1+\frac{s}{S})} \right)^2, \quad j = 1, \ldots, \widetilde{n}$$

$$\widetilde{\omega^2} = \exp\left\{\log\left(\widehat{\omega^2}\right) - \widehat{\theta}_n^*/RV\right\}$$

$$\widetilde{n} = \lfloor 1/\widetilde{\delta} \rfloor$$

and where $\widehat{\theta}_n^*$ is a consistent estimator of integrated variance IV. We take $\widehat{\theta}_n^*$ to be the TSRV estimator $\widehat{\theta}_n$. This estimator requires us to choose $\widetilde{\delta}$ and S. We use the same choice as Barndorff-Nielsen et al. (2008) do, for real and simulated data. This choice is $S = n^{1/2}$ and $\widetilde{\delta} = n^{-1/2}$. Estimator $\widetilde{\omega^2}$ corrects the small sample bias in $\widehat{\omega^2}$. With large number of observations, there is no difference between the two estimators in practice, but we keep the version of Barndorff-Nielsen et al. (2008) anyway.

estimator

$\widehat{V}_b = c\frac{4}{3}\widehat{IQ}_{\text{BNHLS}} + 8c^{-2}\left[\widehat{\omega^2}\right]^2$,

which is consistent for V when noise is i.i.d.

To simulate the market microstructure noise, we consider two cases, autocorrelated and heteroscedastic noise. Combination of the two is straightforward both in theory and practice and is therefore omitted.

5.1. Autocorrelated noise

The market microstructure noise is simulated as an MA(1) process

$$u_{i\Delta} = \epsilon_{i\Delta} + \rho\epsilon_{(i-1)\Delta}, \quad \epsilon \sim N\left(0, \frac{\omega^2}{1-\rho^2}\right),$$

so that $\text{Var}(u) = \omega^2$. In the above, $\Delta = t/n$, $t = 1/50$ is one week, and $n = 35,000$. Four different values of ρ are considered, $\rho = 0, -0.3, -0.5$, and $\rho = -0.7$.

The size of the noise, ω^2, is an important parameter. Here we build on the careful empirical study of Hansen and Lunde (2006), who investigate 30 stocks of Dow Jones Industrial Average. As in Hansen and Lunde (2006), we define noise-to-signal ratio as the ratio of the variance of the noise to the daily integrated volatility. Thus, we introduce

$$\lambda = \frac{\omega^2}{\int_0^{1/250}\sigma_u^2\mathrm{d}u}.$$

Results are simulated for three different noise-to-signal ratios, $\lambda = 0.0001, 0.001$, and 0.01, motivated by Hansen and Lunde (2006). The range of estimated λ for the data of our empirical study is between 0.001 and 0.0019, see Section 6.

Results are represented in terms of coverage probabilities of 95% two-sided, left-sided, and right-sided confidence intervals for IV. Results for noise-to-signal ratios $\lambda = 0.0001, 0.001$, and 0.01 are collected in Tables 2–4, respectively. We see that the subsampling estimator performs well in all scenarios. $\widehat{V}_b$ performs well in the scenario it is designed for, which is the uncorrelated noise case. As the correlation increases, estimated values of $\widehat{V}_b$ decrease, resulting in undercoverage. This effect is less pronounced for smaller noise cases. This is to be expected given that $\widehat{V}_b$ is consistent for V when noise is zero.

This simulation study effectively documents the well known fact that one should not calculate estimators that are not robust to autocorrelation with autocorrelated data. In practice, it can be partly remedied by using sparse data. However, this strategy does not help in general when noise is time-varying. Time-varying noise is an empirical fact that has not received much attention in the nonparametric volatility literature.

5.2. Heteroscedastic noise

We now adopt the noise model of Kalnina and Linton (2008) as in Eq. (17) where noise displays time-varying heteroscedasticity. How to find the closest equivalent of the noise-to-signal ratio for this case? We know that

$$\frac{RV}{2n} \xrightarrow{p} \frac{1}{t}\int_0^t \omega^2(u)\mathrm{d}u.$$

Hence, the most natural definition of noise-to-signal ratio is the ratio of the integrated variance of the noise to the integrated variance of the latent price,

$$\lambda = \frac{\frac{1}{t}\int_0^t \omega^2(u)\mathrm{d}u}{\int_0^t \sigma_u^2\mathrm{d}u},$$

where $t = 1/250$ keeps the horizon of integration to be one day.

This is a lucky situation where conventional estimates of λ motivated by the misspecified homoscedastic framework estimate consistently the λ in the true more general heteroscedastic framework. Thus, exactly the same values of λ are appropriate for the simulation setup, $\lambda = 0.0001, 0.001$, and 0.01. This equivalence does not hold for higher moments of noise, as discussed in Section 3, having implications on the conventional estimates of the asymptotic variance of the Two Scales estimator.

For the shape of heteroscedasticity, we take the simplest possible design motivated by "U-shape", a parabola. It is simple and easy to replicate, but is not meant to be realistic and can be improved in many directions.[13] We set

$$\omega^2(u) = a\left(\frac{u}{t} - \frac{1}{2}\right)^2, \quad u \in [0, t] \tag{25}$$

where $t = 1/50$ and a is a constant chosen to deliver values of $\lambda = 0.0001, 0.001$, or 0.01. Simple calculation shows it implies setting $a = 12\lambda\int_0^t \sigma^2(u)\mathrm{d}u$.

Results are collected in Table 5. Perhaps surprisingly, both methods seem to work well. Is this result to be expected? In this case, the asymptotic variance of the Two Scales estimator is

$$V = \underbrace{c\frac{4}{3}t\int_0^t \sigma_u^4\mathrm{d}u}_{V_{\text{signal}}} + \underbrace{8c^{-2}\frac{1}{t}\int_0^t \omega^4(u)\mathrm{d}u}_{V_{\text{noise}}}. \tag{26}$$

As discussed in Section 3, by Jensen's inequality, V_{noise} would be underestimated by $\widehat{V}_a$. In particular, V_{noise} is underestimated by a factor of

$$\frac{\frac{1}{t}\int_0^t \omega^4(u)\mathrm{d}u}{\left(\frac{1}{t}\int_0^t \omega^2(u)\mathrm{d}u\right)^2},$$

which equals 1.8 when Eq. (25) is true. Estimation of the first part depends on how $\widehat{IQ}_{\text{BNHLS}}$ behaves in the presence of heteroscedastic noise, and it turns out it has a positive bias in all our simulations. Although $\widehat{V}_b$ consists of two components that are both strongly biased, they tend to cancel out, and $\widehat{V}_b$ is at most 20% away from the true value.

We conclude this section with two remarks. First, a class of volatility estimators not used in this paper are the pre-averaging estimators recently proposed by Jacod et al. (2009). This method can estimate IV, IQ, and other volatility functionals in the presence of noise. Although not robust to autocorrelation in the noise, it is robust with respect to heteroscedasticity considered here.

Second, the Two Time Scales estimator is in fact inconsistent in the presence of heteroscedasticity of the noise of this form. This has been shown by Kalnina and Linton (2008) who propose a modification, *jittered* TSRV,[14] which restores consistency. Inconsistency arises due to a bias from the end effects. In our simulations, *jittered* TSRV reduces the bias of the TSRV estimator on average 4 times in the large noise case. However, the magnitude of this bias is too small to show very different results in terms of coverage probabilities. Therefore, we do not report the results for the *jittered* TSRV.

[13] We consider the interval of a week. Thus, a very stylized model of diurnal heteroscedasticity would be 5 parabolas instead,

$$\omega^2(u) = a\sum_{i=1}^{5}\mathbf{1}\left\{u \in \left[\frac{(i-1)t}{5}, \frac{it}{5}\right]\right\}\left(\frac{5u}{t} - i + \frac{1}{2}\right)^2, \quad u \in [0, t].$$

Moreover, a reverse J-shape is typically a better approximation. Also, a data driven method would be more realistic, but that would decrease the transparency and replicability of the simulation setup.

[14] This is not in any way related to jittering of Barndorff-Nielsen et al. (2008). They introduce a modification for Realized Kernels needed to enable estimation of confidence intervals for Realized Kernels in the presence of i.i.d. noise.

Table 1
Summary statistics.

	Num. of obs.	$\widehat{\omega}$	$\widehat{IV}_{\text{daily}}$	$\widehat{\lambda}$
MMM	810,835	0.0004	0.00096	0.0019
MSFT	2,368,013	0.0003	0.00089	0.0011
IBM	1,226,468	0.0003	0.00076	0.0013
AIG	1,054,541	0.0003	0.00070	0.0017
GE	1,835,057	0.0002	0.00057	0.0010
INTC	2,651,006	0.0004	0.00155	0.0010

However, it is important to use the jittered version if noise appears to be heteroscedastic and if avoiding bias is important. Moreover, this correction is strictly positive and in practice almost completely solves the problem that TSRV can be negative.

6. Empirical analysis

This section applies the proposed subsampling method to high frequency data from the NYSE TAQ database, and compares it to the benchmark estimator $\widehat{V}_b$, which is introduced in the previous section. The data consists of full record transaction price data of 6 stocks for year 2006. The six stocks are American International Group (AIG), General Electric (GE), International Business Machines (IBM), Intel (INTC), 3M (MMM), and Microsoft (MSFT).

We first describe the data pre-processing steps. First we obtain raw data of these six stocks for the whole year 2006, time stamped between 9:30 a.m. till 4 p.m. The first column of Table 5 in the Appendix lists the number of observations in this raw data set for each stock. Following Aït-Sahalia et al. (2011), data from all exchanges is retained and zero returns are removed. This means deleting a large part of data (see the second column of Table 5), since these flat trading periods can be quite long. Griffin and Oomen (2008) show that, in the Realized Volatility case, this adjustment of data improves precision of estimation. Jumps are also removed,[15] since the additive market microstructure noise model (13) does not allow for jumps (see the third column of Table 5). There is also an additional issue to consider, which Barndorff-Nielsen et al. (2009) denote as local trends or "gradual" jumps. These authors notice that the realized kernel, which is the estimator of integrated variance they propose, does not behave well in the presence of these "gradual" jumps. Barndorff-Nielsen et al. (2009) notice that these local trends are associated with high volumes traded, and conjecture that they are due to non-trivial liquidity effects. The authors replace them with one genuine jump, but conclude that they do not have an automatic way of detecting episodes of local trends. The subsampling method proposed in the current paper is also vulnerable to such price behavior. Our strategy to identify these gradual jumps is based on the fact that they should look like genuine jumps on a lower frequency. Therefore, we construct a time series of lower (five minute) frequency data, and set to zero those lower frequency returns that are larger than seven weekly standard deviations.

Table 1 contains some summary statistics of the resulting data set. The first column contains the number of observations used for estimation for each stock. The second column reports a measure of the noise (a square root of $RV/2n$, calculated on skip-10-ticks data for the whole year 2006). For IV estimation, we calculate the Two Time Scales estimator for each day in 2006, then average across days to obtain $\widehat{IV}_{\text{daily}}$ (G_1 is the average number of transactions in 5 min; $G_2 = 10$). The fourth column reports

$$\widehat{\lambda} = \frac{RV/2n}{\widehat{IV}_{\text{daily}}}.$$

The returns of all these stocks display large negative autocorrelation similar to GE in Fig. 4(b).

The asymptotic variance of the Two Time Scales estimator is estimated for each of the 52 weeks in year 2006. We conjecture that as long as the distance between observations is of order $1/n$, the underlying theory can be extended to the non-equidistant observations case, at least when the observation times are nonstochastic. Therefore, the estimation is done in tick time, as suggested in Barndorff-Nielsen et al. (2008) and other authors. This also applies to summary statistics.

The results are displayed in Fig. 6 in the Appendix, in terms of 95% confidence intervals for weekly integrated variance. The Two Time Scales estimate $\widehat{\theta}_n$ is in the center of both confidence intervals by construction. The subsampling confidence intervals for Two Time Scales are usually wider than confidence intervals of the benchmark method $\widehat{V}_b$. From our simulations, we conclude this might be due to negative bias of the $\widehat{V}_b$ estimator in the presence of negatively autocorrelated returns. This is because all six stocks have strongly negatively correlated returns, and we know from Section 5 that $\widehat{V}_b$ is downward biased in this case. On the other hand, the subsampling estimator is immune to autocorrelation. The figures also show a lot of variability in the estimates of V. This is mainly due to variability of the Two Time Scales estimates, with large estimates of V corresponding to large $\widehat{\theta}_n$ and vice versa. Thus, episodes of high volatility generally correspond to episodes of high volatility of volatility. Though not reported here, these also correspond to weeks with very large numbers of transactions and large volumes traded.

7. Conclusion

This paper develops an automated method for estimating the asymptotic variance of an estimator in noisy high frequency data. The method applies to an important general class of estimators, which includes many estimators of integrated variance. The new method can substantially simplify the inference question for an estimator, which has an asymptotic variance that is hard to derive or takes a complicated form. An example of such a case is the integrated variance estimator of Aït-Sahalia et al. (2011), in the presence of autocorrelated heteroscedastic market microstructure noise. There is no alternative inferential method available in the literature in this case.

A question that is yet to be addressed rigorously is a data-driven bandwidth choice. Several choices for the Two Time Scales estimator are suggested in the Monte Carlo section.

A very promising extension that will be considered in a future paper is inference for a multivariate parameter. Subsampling naturally produces positive semi-definite estimated variance–covariance matrices, which can be very important for applications. For estimators like Realized Volatility, all the results extend readily to the multivariate case. The real challenge, however, arises due to the additional complications, which are not present in the univariate case. These concern the fact that different stocks do not trade at the same time or so-called asynchronous trading. Also, uncertainty about the observation times becomes much more important in the multivariate context.

[15] Jumps are identified as deviations of the log-returns that are larger than five standard deviations on a moving window of 500 observations. This is motivated by the thresholding technique of filtering out jumps, first proposed by Cecilia Mancini in a series of papers (e.g., Mancini, 2004), see also Aït-Sahalia and Jacod (2009), Eq. (21). Returns containing an identified jump are deleted.

Appendix A. Proofs

Since $\{\sigma_t\}, \{\widetilde{\sigma}_t\}, \{\mu_t\}$ and $\{\widetilde{\mu}_t\}$ are locally bounded, it can be assumed, without loss of generality, that they are uniformly bounded by C_σ (see Barndorff-Nielsen et al. (2006), Section 3). We use C to denote a generic constant that is different from line to line.

A.1. Proof of Proposition 1

By Cauchy–Schwarz and Burkholder–Davis–Gundy inequality (Revuz and Yor, 2005, p. 160),

$$\begin{aligned}
\mathrm{E}\widehat{\theta}_{n,m,l} &= \sum_{i=m(l-1)}^{ml} \mathrm{E}\left(X_{i/n} - X_{(i-1)/n}\right)^2 \\
&\le C \sum_{i=m(l-1)}^{ml} \int_{(i-1)/n}^{i/n} \sigma_u^4 du \le CC_\sigma \frac{m}{n},
\end{aligned}$$

$$\begin{aligned}
&\mathrm{Var}\widehat{\theta}_{n,m,l} \\
&= \sum_{i'=m(l-1)}^{ml} \sum_{i=m(l-1)}^{ml} \mathrm{Cov}\left[\left(X_{i/n} - X_{(i-1)/n}\right)^2, \left(X_{i'/n} - X_{(i'-1)/n}\right)^2\right] \\
&\le \sum_{i'=m(l-1)}^{ml} \sum_{i=m(l-1)}^{ml} \mathrm{E}\left[\left(X_{i/n} - X_{(i-1)/n}\right)^2 \left(X_{i'/n} - X_{(i'-1)/n}\right)^2\right] \\
&\le \sum_{i'=m(l-1)}^{ml} \sum_{i=m(l-1)}^{ml} \mathrm{E}\left[\left(X_{i/n} - X_{(i-1)/n}\right)^4\right]^{1/2} \\
&\quad \times \mathrm{E}\left[\left(X_{i'/n} - X_{(i'-1)/n}\right)^4\right]^{1/2} \\
&\le C \sum_{i'=m(l-1)}^{ml} \sum_{i=m(l-1)}^{ml} \mathrm{E}\left[\left(\int_{(i-1)/n}^{i/n} \sigma_u^4 du\right)^2\right]^{1/2} \\
&\quad \times \mathrm{E}\left[\left(\int_{(i'-1)/n}^{i'/n} \sigma_u^4 du\right)^2\right]^{1/2} \\
&\le CC_\sigma m^2 n^{-2}
\end{aligned}$$

for some constant C. Hence,

$$\widehat{\theta}_{n,m,l} = O_p\left(\frac{m}{n}\right)$$

and

$$\begin{aligned}
\widehat{V}_{PR} &= m \times \frac{1}{K} \sum_{l=1}^{K} \left(\widehat{\theta}_{n,m,l} - \widehat{\theta}_n\right)^2 \\
&= m\widehat{\theta}_n^2 - 2\widehat{\theta}_n m \times \frac{1}{K} \sum_{l=1}^{K} \widehat{\theta}_{n,m,l} + m \times \frac{1}{K} \sum_{l=1}^{K} \left(\widehat{\theta}_{n,m,l}\right)^2 \\
&= m\widehat{\theta}_n^2 - 2\frac{m}{K}\widehat{\theta}_n^2 + \frac{m}{K} \sum_{l=1}^{K} \left(\widehat{\theta}_{n,m,l}\right)^2 \\
&= m\widehat{\theta}_n^2 + o_p(m).
\end{aligned}$$

The result now follows by consistency of $\widehat{\theta}_n$ for θ. □

A.2. Proof of Proposition 2

Before proceeding to the main proof, we state two useful inequalities that hold when X and its volatility are Brownian semimartingales. First, for any $q > 0$

$$\mathrm{E}\left(|\sigma_{t+s} - \sigma_t|^q \,|\mathcal{F}_t\right) \le Cs^{q/2}. \tag{27}$$

This holds because

$$\begin{aligned}
\mathrm{E}\left(|\sigma_{t+s} - \sigma_t|^q \,|\mathcal{F}_t\right) &= \mathrm{E}\left(\left|\int_t^{s+t} \widetilde{\mu}_u du + \int_t^{s+t} \widetilde{\sigma}_u d\widetilde{W}_u\right|^q |\mathcal{F}_t\right) \\
&\le \mathrm{E}\left(\left|\int_t^{s+t} \widetilde{\mu}_u du\right|^q |\mathcal{F}_t\right) \\
&\quad + \mathrm{E}\left(\left|\int_t^{s+t} \widetilde{\sigma}_u d\widetilde{W}_u\right|^q |\mathcal{F}_t\right) \\
&\le Cs^q + C\mathrm{E}\left(\left|\int_t^{s+t} \widetilde{\sigma}_u^2 du\right|^{q/2} |\mathcal{F}_t\right) \\
&\le Cs^{q/2}
\end{aligned}$$

where the Davis–Burkholder–Gundy inequality (Revuz and Yor, 2005, p. 160) is used to obtain the second transition.

The second inequality is as follows, see Jacod (2007). For for all $q > 1$,

$$\mathrm{E}\left[|\mathcal{X}_{k.i}|^q \,\Big|\mathcal{F}_{\frac{(k-1)m+i-1}{n}}\right] \le C\left(\frac{1}{n}\right)^{1\wedge q/2} \tag{28}$$

where

$$\begin{aligned}
\mathcal{X}_{k.i} &= \sqrt{n}\left[\sigma_{\frac{m(k-1)}{n}} \Delta W_{\frac{(k-1)m+i}{n}} - \Delta X_{\frac{(k-1)m+i}{n}}\right] \\
&= \sqrt{n} \int_{[(k-1)m+i-1]/n}^{[(k-1)m+i]/n} \left(\mu_u du + \left(\sigma_u - \sigma_{\frac{m(k-1)}{n}}\right) dW_u\right).
\end{aligned}$$

Introduce the following notation,

$$\widehat{V}_{\text{sub}}^{\text{DISCR}} = \frac{1}{K} \sum_{k=1}^{K} 2\sigma_{\frac{k-1}{K}}^4$$

$$\mathcal{E}\left(\widehat{V}\right)^{\text{DISCR}} = \frac{J}{K} \sum_{k=1}^{K} \mathrm{E}\left[\gamma_k^{\text{DISCR}} |\mathcal{F}_{\frac{k-1}{K}}\right]$$

$$\mathcal{E}\left(\widehat{V}\right) = \frac{J}{K} \sum_{k=1}^{K} \mathrm{E}\left[\gamma_k |\mathcal{F}_{\frac{k-1}{K}}\right]$$

$$\widehat{\alpha}_k^{\text{short}} = \frac{n}{J} \sum_{i=1}^{J} \sigma^2_{\frac{m(k-1)}{n}} \left(W_{\frac{(k-1)m+i}{n}} - W_{\frac{(k-1)m+i-1}{n}}\right)^2$$

$$\gamma_k = \left(\widehat{\theta}_k^{\text{short}} - \widehat{\theta}_k^{\text{long}}\right)^2$$

$$\widehat{\alpha}_k^{\text{long}} = \frac{n}{m} \sum_{i=1}^{m} \sigma^2_{\frac{m(k-1)}{n}} \left(W_{\frac{(k-1)m+i}{n}} - W_{\frac{(k-1)m+i-1}{n}}\right)^2$$

$$\gamma_k^{\text{DISCR}} = \left(\widehat{\alpha}_k^{\text{short}} - \widehat{\alpha}_k^{\text{long}}\right)^2.$$

We want to show

$$\begin{aligned}
\widehat{V}_{\text{sub}} &= J \times \frac{1}{K} \sum_{l=1}^{K} \left(\frac{n}{J}\widehat{\theta}_l^{\text{short}} - \frac{n}{m}\widehat{\theta}_l^{\text{long}}\right)^2 \xrightarrow{p} V \\
&= 2\int_0^1 \sigma_u^4 du.
\end{aligned}$$

First, by Riemann integrability of σ,

$$V^{\text{DISCR}} \xrightarrow{p} V = 2\int_0^1 \sigma_u^4 du.$$

To prove Proposition 2, proceed in three steps. Prove $\widehat{V}_{\text{sub}} - \mathcal{E}\left(\widehat{V}\right) \xrightarrow{p} 0$, then $\mathcal{E}\left(\widehat{V}\right)^{\text{DISCR}} - \mathcal{E}\left(\widehat{V}\right) \xrightarrow{p} 0$, and finally $\mathcal{E}\left(\widehat{V}\right)^{\text{DISCR}} - V^{\text{DISCR}} \xrightarrow{p} 0$.

The first step is to show

$$\widehat{V}_{\text{sub}} - \mathcal{E}\left(\widehat{V}\right) = \frac{J}{K}\sum_{k=1}^{K}\left(\gamma_k - \mathrm{E}\left[\gamma_k|\mathcal{F}_{\frac{k-1}{K}}\right]\right) \xrightarrow{p} 0.$$

By Lenglart's inequality (see e.g. Podolskij, 2006), it is sufficient to show that

$$\sum_{k=1}^{K}\mathrm{E}\left[\left|\frac{J}{K}\gamma_k\right|^2\left|\mathcal{F}_{\frac{k-1}{K}}\right.\right] \xrightarrow{p} 0.$$

We have

$$\mathrm{E}\left[\left|\frac{J}{K}\gamma_k\right|^2\left|\mathcal{F}_{\frac{k-1}{K}}\right.\right] = \frac{J^2}{K^2}\mathrm{E}\left[\left\{\frac{n}{J}\sum_{i=1}^{J}\left(X_{\frac{(k-1)m+i}{n}} - X_{\frac{(k-1)m+i-1}{n}}\right)^2 - \frac{n}{m}\sum_{i=1}^{m}\left(X_{\frac{(k-1)m+i}{n}} - X_{\frac{(k-1)m+i-1}{n}}\right)^2\right\}^4\left|\mathcal{F}_{\frac{k-1}{K}}\right.\right]$$

$$\leq C\frac{J^2}{K^2} = \frac{J^2m^2}{n^2}$$

for some constant C not depending on k, by repeated use of the Cauchy–Schwarz inequality and

$$\mathrm{E}\left[\left|X_{\frac{(k-1)m+i}{n}} - X_{\frac{(k-1)m+i-1}{n}}\right|^q\left|\mathcal{F}_{\frac{k-1}{K}}\right.\right] \leq C_q\left(\frac{1}{n}\right)^{q/2}$$

for all $q > 0$, $i = 1, \ldots, m$, and C_q some constant depending on q only. Hence,

$$\sum_{k=1}^{K}\mathrm{E}\left[\left|\frac{J}{K}\gamma_k\right|^2\left|\mathcal{F}_{\frac{k-1}{K}}\right.\right] \leq C\frac{J^2}{K} = C\frac{mJ^2}{n}.$$

The first step is thus proved, provided $mJ^2n^{-1} \to 0$.

The second step is to show

$$\mathcal{E}\left(\widehat{V}\right)^{\text{DISCR}} - \mathcal{E}\left(\widehat{V}\right) = \frac{J}{K}\sum_{k=1}^{K}\mathrm{E}\left[\gamma_k^{\text{DISCR}} - \gamma_k\left|\mathcal{F}_{\frac{k-1}{K}}\right.\right] \xrightarrow{p} 0.$$

We have

$$\begin{aligned}
&\mathrm{E}\left[\left|\gamma_k^{\text{DISCR}} - \gamma_k\right|\left|\mathcal{F}_{\frac{k-1}{K}}\right.\right] \\
&= \mathrm{E}\left[\left|\widehat{\alpha}_k^{\text{long}} - \widehat{\alpha}_k^{\text{short}} + \widehat{\theta}_k^{\text{long}} - \widehat{\theta}_k^{\text{short}}\right| \times \left|\left\{\widehat{\alpha}_k^{\text{long}} - \widehat{\theta}_k^{\text{long}}\right\} - \left\{\widehat{\alpha}_k^{\text{short}} - \widehat{\theta}_k^{\text{short}}\right\}\right|\left|\mathcal{F}_{\frac{k-1}{K}}\right.\right] \\
&= \mathrm{E}_k\left[\left|\widehat{\alpha}_k^{\text{long}} - \widehat{\alpha}_k^{\text{short}} + \widehat{\theta}_k^{\text{long}} - \widehat{\theta}_k^{\text{short}}\right| \right.\\
&\quad \times \left|\frac{n}{m}\sum_{i=1}^{m}\left[\sigma^2_{\frac{m(k-1)}{n}}\left(\Delta W_{\frac{(k-1)m+i}{n}}\right)^2 - \left(\Delta X_{\frac{(k-1)m+i}{n}}\right)^2\right]\right.\\
&\quad \left.\left. - \frac{n}{J}\sum_{i=1}^{J}\left[\sigma^2_{\frac{m(k-1)}{n}}\left(\Delta W_{\frac{(k-1)m+i}{n}}\right)^2 - \left(\Delta X_{\frac{(k-1)m+i}{n}}\right)^2\right]\right|\right] \\
&\leq \sqrt{\mathrm{E}_kA^2}\sqrt{\mathrm{E}_kB^2}.
\end{aligned}$$

Let $c_i = n/m - n/J$ for $i = 1, \ldots, J$. $c_i = n/m$ for $i = J + 1, \ldots, m$. The second part is the square root of

$$\begin{aligned}
\mathrm{E}_kB^2 &= \mathrm{E}_k\left[\frac{n}{m}\sum_{i=1}^{m}\left(\sigma^2_{\frac{m(k-1)}{n}}\left(\Delta W_{\frac{(k-1)m+i}{n}}\right)^2 - \left(\Delta X_{\frac{(k-1)m+i}{n}}\right)^2\right) - \frac{n}{J}\sum_{i=1}^{J}\left(\sigma^2_{\frac{m(k-1)}{n}}\left(\Delta W_{\frac{(k-1)m+i}{n}}\right)^2 - \left(\Delta X_{\frac{(k-1)m+i}{n}}\right)^2\right)\right]^2 \\
&= \mathrm{E}_k\left[\left\{\sum_{i=1}^{m}c_i\left[\sigma^2_{\frac{m(k-1)}{n}}\left(\Delta W_{\frac{(k-1)m+i}{n}}\right)^2 - \left(\Delta X_{\frac{(k-1)m+i}{n}}\right)^2\right]\right\}^2\right] \\
&= \sum_{i=1}^{m}c_i^2\mathrm{E}_k\left[\sigma^2_{\frac{m(k-1)}{n}}\left(\Delta W_{\frac{(k-1)m+i}{n}}\right)^2 - \left(\Delta X_{\frac{(k-1)m+i}{n}}\right)^2\right]^2 \\
&= \sum_{i=1}^{m}\sum_{i'=1}^{m}c_ic_{i'}\mathrm{E}_k\left[\sigma^2_{\frac{m(k-1)}{n}}\left(\Delta W_{\frac{(k-1)m+i}{n}}\right)^2 - \left(\Delta X_{\frac{(k-1)m+i}{n}}\right)^2\right] \\
&\quad \times \left[\sigma^2_{\frac{m(k-1)}{n}}\left(\Delta W_{\frac{(k-1)m+i'}{n}}\right)^2 - \left(\Delta X_{\frac{(k-1)m+i'}{n}}\right)^2\right] \\
&\leq Cn^{-5/2}\sum_{i=1}^{m}c_i^2 + Cn^{-3}\sum_{i=1}^{m}\sum_{i'=1}^{m}|c_i|\,|c_{i'}| \\
&\leq Cn^{-5/2}\frac{n^2}{J} + Cn^{-3}n^2 \\
&= Cn^{-1/2}J^{-1} + Cn^{-1}
\end{aligned}$$

because

$$\begin{aligned}
&\mathrm{E}_k\left[\sigma^2_{\frac{m(k-1)}{n}}\left(\Delta W_{\frac{(k-1)m+i}{n}}\right)^2 - \left(\Delta X_{\frac{(k-1)m+i}{n}}\right)^2\right]^2 \\
&= \mathrm{E}_k\left[\sigma_{\frac{m(k-1)}{n}}\Delta W_{\frac{(k-1)m+i}{n}} - \Delta X_{\frac{(k-1)m+i}{n}}\right]^2 \times \left[\sigma_{\frac{m(k-1)}{n}}\Delta W_{\frac{(k-1)m+i}{n}} + \Delta X_{\frac{(k-1)m+i}{n}}\right]^2 \\
&\leq \sqrt{\mathrm{E}_k\left[\sigma_{\frac{m(k-1)}{n}}\Delta W_{\frac{(k-1)m+i}{n}} - \Delta X_{\frac{(k-1)m+i}{n}}\right]^4} \times \sqrt{\mathrm{E}_k\left[\sigma_{\frac{m(k-1)}{n}}\Delta W_{\frac{(k-1)m+i}{n}} + \Delta X_{\frac{(k-1)m+i}{n}}\right]^4} \\
&\leq C\sqrt{\frac{1}{n^3}}\sqrt{\frac{1}{n^2}} = Cn^{-5/2}
\end{aligned}$$

and, for $i < i'$,

$$\begin{aligned}
&\left|\mathrm{E}_k\left[\sigma^2_{\frac{m(k-1)}{n}}\left(\Delta W_{\frac{(k-1)m+i}{n}}\right)^2 - \left(\Delta X_{\frac{(k-1)m+i}{n}}\right)^2\right] \times \left[\sigma^2_{\frac{m(k-1)}{n}}\left(\Delta W_{\frac{(k-1)m+i'}{n}}\right)^2 - \left(\Delta X_{\frac{(k-1)m+i'}{n}}\right)^2\right]\right| \\
&\leq \mathrm{E}_k\left(\left|\sigma^2_{\frac{m(k-1)}{n}}\left(\Delta W_{\frac{(k-1)m+i}{n}}\right)^2 - \left(\Delta X_{\frac{(k-1)m+i}{n}}\right)^2\right| \right.\\
&\quad \left.\times \mathrm{E}\left[\left|\sigma^2_{\frac{m(k-1)}{n}}\left(\Delta W_{\frac{(k-1)m+i'}{n}}\right)^2 - \left(\Delta X_{\frac{(k-1)m+i'}{n}}\right)^2\right|\left|\mathcal{F}_{\frac{(k-1)m+i}{n}}\right.\right]\right) \\
&\leq Cn^{-3/2}\mathrm{E}_k\left[\left|\sigma^2_{\frac{m(k-1)}{n}}\left(\Delta W_{\frac{(k-1)m+i}{n}}\right)^2 - \left(\Delta X_{\frac{(k-1)m+i}{n}}\right)^2\right|\right] \\
&\leq Cn^{-3}.
\end{aligned}$$

The first part is the square root of

$$\begin{aligned}
\mathrm{E}_kA^2 &= \mathrm{E}_k\left(\widehat{\alpha}_k^{\text{long}} - \widehat{\alpha}_k^{\text{short}} + \widehat{\theta}_k^{\text{long}} - \widehat{\theta}_k^{\text{short}}\right)^2 \\
&= \mathrm{E}_k\left[\left\{\sum_{i=1}^{m}c_i\left[\sigma^2_{\frac{m(k-1)}{n}}\left(\Delta W_{\frac{(k-1)m+i}{n}}\right)^2 + \left(\Delta X_{\frac{(k-1)m+i}{n}}\right)^2\right]\right\}^2\right] \\
&\leq C.
\end{aligned}$$

Combining both A and B terms, we obtain

$$\mathrm{E}\left[\left|\gamma_k^{\text{DISCR}} - \gamma_k\right|\left|\mathcal{F}_{\frac{k-1}{K}}\right.\right] \leq Cn^{-1/4}J^{-1/2} + Cn^{-1/2},$$

from which the second step

$$\left|\mathcal{E}\left(\widehat{V}\right)^{\mathrm{DISCR}}-\mathcal{E}\left(\widehat{V}\right)\right| \le \frac{J}{K}\sum_{k=1}^{K}\mathrm{E}\left[\left|\gamma_k^{\mathrm{DISCR}}-\gamma_k\right|\left|\mathcal{F}_{\frac{k-1}{K}}\right.\right] \le CJn^{-1/4}J^{-1/2}+CJn^{-1/2}\xrightarrow{p}0$$

follows, provided $J^2/n \to 0$, which is implied by $mJ^2n^{-1} \to 0$.

Now we prove the third step.

$$\mathrm{E}\left[\gamma_k^{\mathrm{DISCR}}\left|\mathcal{F}_{\frac{k-1}{K}}\right.\right] = \sigma^4_{\frac{m(k-1)}{n}}\mathrm{E}\left[\left\{\frac{n}{J}\sum_{i=1}^{J}\left(W_{\frac{(k-1)m+i}{n}}-W_{\frac{(k-1)m+i-1}{n}}\right)^2 - \frac{n}{m}\sum_{i=1}^{m}\left(W_{\frac{(k-1)m+i}{n}}-W_{\frac{(k-1)m+i-1}{n}}\right)^2\right\}^2\right] = \sigma^4_{\frac{m(k-1)}{n}}\frac{2}{J}-\sigma^4_{\frac{m(k-1)}{n}}\frac{2}{m}.$$

Thus,

$$\mathcal{E}\left(\widehat{V}\right)^{\mathrm{DISCR}} = \frac{J}{K}\sum_{k=1}^{K}\mathrm{E}\left[\gamma_k^{\mathrm{DISCR}}\left|\mathcal{F}_{\frac{k-1}{K}}\right.\right] = \frac{J}{K}\sum_{k=1}^{K}\sigma^4_{\frac{m(k-1)}{n}}\frac{2}{J}-\frac{J}{K}\sum_{k=1}^{K}\sigma^4_{\frac{m(k-1)}{n}}\frac{2}{m} = \widehat{V}^{\mathrm{DISCR}}_{\mathrm{sub}}-O_p\left(\frac{J}{m}\right).$$

This proves consistency of the subsampling method for RV, provided $mJ^2n^{-1} \to 0$ and σ satisfies A1. □

A.3. Proof of Proposition 3

Proposition 3 is proved for the special case $Q = m$. The general Q case follows by the same steps, but the notation is more involved. Denote $K = \lfloor n/m \rfloor$ and $\Delta_\delta X_t = X_t - X_{t-\delta}$.

Introduce the same notation as in Proposition 2.

$$V^{\mathrm{DISCR}} = \frac{m}{n}\sum_{k=1}^{K}2\sigma^4_{\frac{k-1}{K}}$$

$$\mathcal{E}\left(\widehat{V}\right)^{\mathrm{DISCR}} = \frac{m}{n}\sum_{k=1}^{K}\left[\mathrm{E}\gamma_k^{\mathrm{DISCR}}\left|\mathcal{F}_{\frac{k-1}{K}}\right.\right]$$

$$\mathcal{E}\left(\widehat{V}\right) = \frac{m}{n}\sum_{k=1}^{K}\mathrm{E}\left[\gamma_k\left|\mathcal{F}_{\frac{k-1}{K}}\right.\right] \qquad \widehat{\alpha}_k^{\mathrm{slow}} = \sigma^2_{\frac{m(k-1)}{n}}\left(\Delta_{\frac{m}{n}}W_{\frac{mk}{n}}\right)^2$$

$$\gamma_k = \left(\widehat{\theta}_k^{\mathrm{slow}}-\widehat{\theta}_k^{\mathrm{fast}}\right)^2 \qquad \widehat{\alpha}_k^{\mathrm{fast}} = \sigma^2_{\frac{m(k-1)}{n}}\sum_{i=1}^{m}\left(\Delta_{\frac{1}{n}}W_{\frac{i+m(k-1)}{n}}\right)^2$$

$$\gamma_k^{\mathrm{DISCR}} = \left(\widehat{\alpha}_k^{\mathrm{slow}}-\widehat{\alpha}_k^{\mathrm{fast}}\right)^2.$$

Also, denote $\mathrm{E}\left[\gamma_k\left|\mathcal{F}_{\frac{k-1}{K}}\right.\right]$ by $\mathrm{E}^n_{k-1}[\gamma_k]$. We want to show

$$\widehat{V}'_{\mathrm{sub}}\xrightarrow{p}V = 2\int_0^1\sigma_u^4\mathrm{d}u$$

where

$$\widehat{V}'_{\mathrm{sub}} = \frac{n}{m}\sum_{k=1}^{K}\left(\widehat{\theta}_k^{\mathrm{slow}}-\widehat{\theta}_k^{\mathrm{fast}}\right)^2 = \frac{n}{m}\sum_{k=1}^{K}\left\{\left(\Delta_{\frac{m}{n}}X_{\frac{mk}{n}}\right)^2-\sum_{i=1}^{m}\left(\Delta_{\frac{1}{n}}X_{\frac{i+m(k-1)}{n}}\right)^2\right\}^2.$$

First, by Riemann integrability,

$$V^{\mathrm{DISCR}}\xrightarrow{p}V = 2\int_0^1\sigma_u^4\mathrm{d}u. \tag{29}$$

To prove Proposition 3, use the following three steps. Prove $\widehat{V}'_{\mathrm{sub}} - \mathcal{E}\left(\widehat{V}\right)\xrightarrow{p}0$, then $\mathcal{E}\left(\widehat{V}\right)^{\mathrm{DISCR}} - \mathcal{E}\left(\widehat{V}\right)\xrightarrow{p}0$, and finally $\mathcal{E}\left(\widehat{V}\right)^{\mathrm{DISCR}} - V^{\mathrm{DISCR}}\xrightarrow{p}0$.

The first step is to show

$$\widehat{V}'_{\mathrm{sub}}-\mathcal{E}\left(\widehat{V}\right) = K\sum_{k=1}^{K}\left(\gamma_k-\mathrm{E}\left[\gamma_k\left|\mathcal{F}_{\frac{k-1}{K}}\right.\right]\right)\xrightarrow{p}0.$$

By Lenglart's inequality (see e.g. Podolskij, 2006), it is sufficient to show that

$$\sum_{k=1}^{K}\mathrm{E}\left[|K\gamma_k|^2\left|\mathcal{F}_{\frac{k-1}{K}}\right.\right]\xrightarrow{p}0.$$

Notice that, by the Burkholder–Davis–Gundy inequality, Cauchy–Schwarz inequality, and uniform boundedness of σ,

$$\mathrm{E}_{k-1}\left[\left(\widehat{\theta}_k^{\mathrm{fast}}\right)^4\right] \le \sum_{i'''=1}^{m}\sum_{i''=1}^{m}\sum_{i'=1}^{m}\sum_{i=1}^{m}\sqrt[4]{\mathrm{E}_{k-1}\left[\left(\Delta_{\frac{1}{n}}X_{\frac{i+m(k-1)}{n}}\right)^8\right]} \times\sqrt[4]{\mathrm{E}_{k-1}\left[\left(\Delta_{\frac{1}{n}}X_{\frac{i'+m(k-1)}{n}}\right)^8\right]} \times\sqrt[4]{\mathrm{E}_{k-1}\left[\left(\Delta_{\frac{1}{n}}X_{\frac{i''+m(k-1)}{n}}\right)^8\right]} \times\sqrt[4]{\mathrm{E}_{k-1}\left[\left(\Delta_{\frac{1}{n}}X_{\frac{i'''+m(k-1)}{n}}\right)^8\right]} \le C\frac{m^4}{n^4} = C\frac{1}{K^4}$$

for some constant C, which does not depend on any of the above parameters. Hence, and by similarity,

$$\begin{aligned}&\mathrm{E}_{k-1}\left[\left(\widehat{\theta}_k^{\mathrm{fast}}\right)^4\right]\le\frac{C}{K^4}, \qquad \mathrm{E}_{k-1}\left[\left(\widehat{\theta}_k^{\mathrm{fast}}\right)^3\right]\le\frac{C}{K^4},\\ &\mathrm{E}_{k-1}\left[\left(\widehat{\theta}_k^{\mathrm{fast}}\right)^2\right]\le\frac{C}{K^2},\\ &\mathrm{E}_{k-1}\left[\left(\widehat{\theta}_k^{\mathrm{slow}}\right)^4\right]\le\frac{C}{K^4}, \qquad \mathrm{E}_{k-1}\left[\left(\widehat{\theta}_k^{\mathrm{slow}}\right)^3\right]\le\frac{C}{K^3},\\ &\mathrm{E}_{k-1}\left[\left(\widehat{\theta}_k^{\mathrm{slow}}\right)^2\right]\le\frac{C}{K^2}.\end{aligned} \tag{30}$$

From here,

$$\mathrm{E}_{k-1}\left[\gamma_k^2\right] = \mathrm{E}_{k-1}\left[\left(\widehat{\theta}_k^{\mathrm{fast}}-\widehat{\theta}_k^{\mathrm{slow}}\right)^4\right]\le C\frac{1}{K^4}$$

and

$$\sum_{k=1}^{K}\mathrm{E}\left[|K\gamma_k|^2\left|\mathcal{F}_{\frac{k-1}{K}}\right.\right]\le C\frac{1}{K} = o(1).$$

The second step is to show

$$\mathcal{E}\left(\widehat{V}\right)^{\mathrm{DISCR}}-\mathcal{E}\left(\widehat{V}\right) = K\sum_{k=1}^{K}\mathrm{E}\left[\gamma_k^{\mathrm{DISCR}}-\gamma_k\left|\mathcal{F}_{\frac{k-1}{K}}\right.\right]\xrightarrow{p}0.$$

It is sufficient to show

$$K\sum_{k=1}^{K}\mathrm{E}\left[\left|\gamma_k^{\mathrm{DISCR}}-\gamma_k\right|\right]\to 0.$$

Write

$$K\sum_{k=1}^{K}\mathrm{E}\left[\left|\gamma_k^{\mathrm{DISCR}}-\gamma_k\right|\right]=K\sum_{k=1}^{K}\mathrm{E}\left[\left|\widehat{\alpha}_k^{\mathrm{fast}}-\widehat{\alpha}_k^{\mathrm{slow}}+\widehat{\theta}_k^{\mathrm{fast}}-\widehat{\theta}_k^{\mathrm{slow}}\right|\right.$$
$$\left.\times\left|\left\{\widehat{\alpha}_k^{\mathrm{fast}}-\widehat{\theta}_k^{\mathrm{fast}}\right\}-\left\{\widehat{\alpha}_k^{\mathrm{slow}}-\widehat{\theta}_k^{\mathrm{slow}}\right\}\right|\right]$$
$$\equiv A+B.$$

As to the first term, we have

$$\begin{aligned}
A &= K\sum_{k=1}^{K}\mathrm{E}\left[\left|\widehat{\alpha}_k^{\mathrm{fast}}-\widehat{\alpha}_k^{\mathrm{slow}}+\widehat{\theta}_k^{\mathrm{fast}}-\widehat{\theta}_k^{\mathrm{slow}}\right|\left|\widehat{\alpha}_k^{\mathrm{fast}}-\widehat{\theta}_k^{\mathrm{fast}}\right|\right]\\
&\le K\sum_{k=1}^{K}\left\{\mathrm{E}\left[\widehat{\alpha}_k^{\mathrm{fast}}-\widehat{\alpha}_k^{\mathrm{slow}}+\widehat{\theta}_k^{\mathrm{fast}}-\widehat{\theta}_k^{\mathrm{slow}}\right]^2\right\}^{1/2}\\
&\quad\times\left\{\mathrm{E}\left[\widehat{\alpha}_k^{\mathrm{fast}}-\widehat{\theta}_k^{\mathrm{fast}}\right]^2\right\}^{1/2}\\
&\le C\sum_{k=1}^{K}\left\{\mathrm{E}\left[\widehat{\alpha}_k^{\mathrm{fast}}-\widehat{\theta}_k^{\mathrm{fast}}\right]^2\right\}^{1/2}\\
&= C\sum_{k=1}^{K}\left\{\mathrm{E}\left[\sum_{i=1}^{m}\left\{\sigma_{\frac{m(k-1)}{n}}\Delta_{\frac{1}{n}}W_{\frac{i+m(k-1)}{n}}-\Delta_{\frac{1}{n}}X_{\frac{i+m(k-1)}{n}}\right\}\right.\right.\\
&\quad\left.\left.\times\left\{\sigma_{\frac{m(k-1)}{n}}\Delta_{\frac{1}{n}}W_{\frac{i+m(k-1)}{n}}+\Delta_{\frac{1}{n}}X_{\frac{i+m(k-1)}{n}}\right\}\right]^2\right\}^{1/2}\\
&\le C\sum_{k=1}^{K}\left\{\sum_{i'=1}^{m}\sum_{i=1}^{m}\sqrt[4]{\mathrm{E}\left[\left\{\sigma_{\frac{m(k-1)}{n}}\Delta_{\frac{1}{n}}W_{\frac{i+m(k-1)}{n}}-\Delta_{\frac{1}{n}}X_{\frac{i+m(k-1)}{n}}\right\}^4\right]}\right.\\
&\quad\times\sqrt[4]{\mathrm{E}\left[\left\{\sigma_{\frac{m(k-1)}{n}}\Delta_{\frac{1}{n}}W_{\frac{i+m(k-1)}{n}}+\Delta_{\frac{1}{n}}X_{\frac{i+m(k-1)}{n}}\right\}^4\right]}\times\\
&\quad\sqrt[4]{\mathrm{E}\left[\left\{\sigma_{\frac{m(k-1)}{n}}\Delta_{\frac{1}{n}}W_{\frac{i'+m(k-1)}{n}}-\Delta_{\frac{1}{n}}X_{\frac{i'+m(k-1)}{n}}\right\}^4\right]}\\
&\quad\left.\times\sqrt[4]{\mathrm{E}\left[\left\{\sigma_{\frac{m(k-1)}{n}}\Delta_{\frac{1}{n}}W_{\frac{i'+m(k-1)}{n}}+\Delta_{\frac{1}{n}}X_{\frac{i'+m(k-1)}{n}}\right\}^4\right]}\right\}^{1/2}\\
&\le\frac{C}{\sqrt{n}}\sum_{k=1}^{K}\sum_{i=1}^{m}\sqrt[4]{\mathrm{E}\left[\left\{\sigma_{\frac{m(k-1)}{n}}\Delta_{\frac{1}{n}}W_{\frac{i+m(k-1)}{n}}-\Delta_{\frac{1}{n}}X_{\frac{i+m(k-1)}{n}}\right\}^4\right]}\\
&\le\frac{C}{\sqrt{n}}\sum_{k=1}^{K}\sum_{i=1}^{m}\left\{\mathrm{E}\left[\int_{\frac{i-1+m(k-1)}{n}}^{\frac{i+m(k-1)}{n}}\left(\sigma_{\frac{m(k-1)}{n}}-\sigma_u\right)^2\mathrm{d}u\right]^2\right\}^{1/4}\\
&=\frac{C}{\sqrt{n}}\sum_{k=1}^{K}\sum_{i=1}^{m}\left\{\mathrm{E}\left[\int_{\frac{i-1+m(k-1)}{n}}^{\frac{i+m(k-1)}{n}}\left(\sigma_{\frac{\lfloor Ku\rfloor}{K}}-\sigma_u\right)^2\mathrm{d}u\right]^2\right\}^{1/4}.
\end{aligned}$$

In the above, to obtain the second inequality, we used (30). To obtain the fourth inequality, we used

$$\mathrm{E}\left[\left\{\sigma_{\frac{m(k-1)}{n}}\Delta_{\frac{1}{n}}W_{\frac{i+m(k-1)}{n}}+\Delta_{\frac{1}{n}}X_{\frac{i+m(k-1)}{n}}\right\}^4\right]$$
$$\le C\mathrm{E}\left[\left\{\int_{\frac{i-1+m(k-1)}{n}}^{\frac{i+m(k-1)}{n}}\left(\sigma_{\frac{m(k-1)}{n}}+\sigma_u\right)^2\mathrm{d}u\right\}^2\right]\le\frac{C}{n^2},$$

which follows by Burkholder–Davis–Gundy inequality. To proceed with term A, we use the arguments along the lines of the proof of Lemma 1 of Barndorff-Nielsen and Shephard (2002). For every i and k, there exists a constant $c_{i,k}$ s.t.

$$\inf_{\frac{i-1+m(k-1)}{n}\le u\le\frac{i+m(k-1)}{n}}\left(\sigma_{\frac{\lfloor Ku\rfloor}{K}}-\sigma_u\right)^2$$
$$\le c_{i,k}\le\sup_{\frac{i-1+m(k-1)}{n}\le u\le\frac{i+m(k-1)}{n}}\left(\sigma_{\frac{\lfloor Ku\rfloor}{K}}-\sigma_u\right)^2$$

and

$$\int_{\frac{i-1+m(k-1)}{n}}^{\frac{i+m(k-1)}{n}}\left(\sigma_{\frac{\lfloor Ku\rfloor}{K}}-\sigma_u\right)^2\mathrm{d}u=c_{i,k}\frac{1}{n}.$$

Notice that

$$\sup_{i,k}c_{i,k}\to 0$$

by right-continuity and boundedness of σ. Then,

$$\begin{aligned}
A&\le\frac{C}{\sqrt{n}}\sum_{k=1}^{K}\sum_{i=1}^{m}\left\{\mathrm{E}\left[\int_{\frac{i-1+m(k-1)}{n}}^{\frac{i+m(k-1)}{n}}\left(\sigma_{\frac{\lfloor Ku\rfloor}{K}}-\sigma_u\right)^2\mathrm{d}u\right]^2\right\}^{1/4}\\
&=\frac{C}{\sqrt{n}}\sum_{k=1}^{K}\sum_{i=1}^{m}\left\{\mathrm{E}\left[c_{i,k}\frac{1}{n}\right]^2\right\}^{1/4}=C\sum_{k=1}^{K}\sum_{i=1}^{m}\sqrt[4]{\mathrm{E}c_{i,k}^2}\frac{1}{n}\to 0
\end{aligned}$$

by Monotone Convergence Theorem. $B\to 0$ is proved using exactly the same steps. This proves the second step.

The final step is to show

$$\mathcal{E}\left(\widehat{V}\right)^{\mathrm{DISCR}}-V^{\mathrm{DISCR}}\xrightarrow{p}0.$$

We have

$$\begin{aligned}
&\mathrm{E}\left[\gamma_k^{\mathrm{DISCR}}\middle|\mathcal{F}_{\frac{k-1}{K}}\right]\\
&=\sigma_{\frac{m(k-1)}{n}}^4\mathrm{E}\left[\left\{\left(\Delta_{\frac{m}{n}}W_{\frac{mk}{n}}\right)^2-\sum_{i=1}^{m}\left(\Delta_{\frac{1}{n}}W_{\frac{i+m(k-1)}{n}}\right)^2\right\}^2\right]\\
&=\frac{2}{K^2}\sigma_{\frac{m(k-1)}{n}}^4+o_p\left(\frac{1}{K^2}\right).
\end{aligned}$$

Therefore,

$$\begin{aligned}
\mathcal{E}\left(\widehat{V}\right)^{\mathrm{DISCR}}&=K\sum_{k=1}^{K}\mathrm{E}\left[\gamma_k^{\mathrm{DISCR}}\middle|\mathcal{F}_{\frac{k-1}{K}}\right]\\
&=K\sum_{k=1}^{K}\left(\frac{2}{K^2}\sigma_{\frac{m(k-1)}{n}}^4+o_p\left(\frac{1}{K^2}\right)\right)\\
&=\sum_{k=1}^{K}\frac{2}{K^1}\sigma_{\frac{m(k-1)}{n}}^4+o_p(1)=V^{\mathrm{DISCR}}+o_p(1).
\end{aligned}$$

The result follows immediately. □

A.4. Proof of Theorem 4

It is convenient to decompose V_l^{short} into the signal and noise parts, $V_l^{\mathrm{short}}=V_l^{\mathrm{signal}}+V_l^{\mathrm{noise}}$ where

$$V_l^{\mathrm{signal}}=\frac{4}{3}c\int_{(l-1)m/n}^{[(l-1)m+J]/n}\sigma_u^4\mathrm{d}u$$

$$V_l^{\mathrm{noise}}=8c^{-2}\frac{J}{n}\mathrm{Var}\left(\epsilon\right)^2+16\frac{J}{n}c^{-2}\lim_{n\to\infty}\sum_{i=1}^{n}\mathrm{Cov}\left(\epsilon_0,\epsilon_{i/n}\right)^2.$$

We first state the following lemma (see Appendix A.4.1 for proof).

Lemma 7. *Suppose the assumptions of Theorem 4 hold. Then,*

$$\frac{m}{J}\sum_{l=1}^{K}\left[n^{1/3}\left(\widehat{\theta}_l^{\text{short}}-\theta_l^{\text{short}}\right)^2-V_l^{\text{short}}\right]\xrightarrow{p}0.$$

We conclude from Eq. (27) that

$$V-\sum_{l=1}^{K}\frac{m}{J}V_l^{\text{short}}=o_p(1).$$

Therefore, to prove Theorem 4 it is sufficient to prove the negligibility of

$$\begin{aligned}\widehat{V}_{\text{sub}}-\sum_{l=1}^{K}\frac{m}{J}V_l^{\text{short}}&=\frac{m}{J}\sum_{l=1}^{K}\left[n^{1/3}\left(\widehat{\theta}_l^{\text{short}}-\frac{J}{m}\widehat{\theta}_l^{\text{long}}\right)^2-V_l^{\text{short}}\right]\\&=\frac{m}{J}\sum_{l=1}^{K}\Bigg[\left(n^{1/3}\left(\widehat{\theta}_l^{\text{short}}-\theta_l^{\text{short}}\right)^2-V_l^{\text{short}}\right)\\&\quad+n^{1/3}\left(\theta_l^{\overset{\text{long}}{\cancel{\text{short}}}}-\frac{J}{m}\widehat{\theta}_l^{\text{long}}\right)^2\\&\quad+n^{1/3}\left(\theta_l^{\text{short}}-\frac{J}{m}\theta_l^{\text{long}}\right)^2\Bigg]+R,\end{aligned}$$

where R contains cross terms that are $o_p(1)$ if the three main three terms are $o_p(1)$. The first of these three terms is negligible by Lemma 7. The second term is also negligible by Lemma 7 by taking m instead of J,

$$\begin{aligned}n^{1/3}\frac{m}{J}\sum_{l=1}^{K}\left(\frac{J}{m}\widehat{\theta}_l^{\text{long}}-\frac{J}{m}\theta_l^{\text{long}}\right)^2&=\frac{J}{m}n^{1/3}\sum_{l=1}^{K}\left(\widehat{\theta}_l^{\text{long}}-\theta_l^{\text{long}}\right)^2\\&=\frac{J}{m}\left(V+o_p(1)\right)=o_p(1).\end{aligned}$$

The third term is $o_p(1)$ by assumption $Jmn^{-5/3}\to 0$ and

$$\mathrm{E}\left|\theta_l^{\text{short}}-\frac{J}{m}\theta_l^{\text{long}}\right|^2\le C\frac{J^2m}{n^3},$$

which follows from Eq. (27). This concludes the proof of Theorem 4. □

A.4.1. Proof Lemma 7

We now prove Lemma 7 stated in Appendix A.4. We have the following decomposition

$$\begin{aligned}&\frac{m}{J}\sum_{l=1}^{K}\left[n^{1/3}\left(\widehat{\theta}_l^{\text{short}}-\theta_l^{\text{short}}\right)^2-V_l^{\text{short}}\right]\\&=\frac{m}{J}\sum_{l=1}^{K}\left[n^{1/3}\left([Y,Y]_l^{(G_1)}-\frac{\bar{J}_{G_1}}{\bar{J}_{G_2}}[Y,Y]_l^{(G_2)}-\theta_l^{\text{short}}\right)^2-V_l^{\text{short}}\right]\\&=\frac{m}{J}\sum_{l=1}^{K}\Bigg[n^{1/3}\Bigg([X,X]_l^{(G_1)}-\theta_l^{\text{short}}+[\epsilon,\epsilon]_l^{(G_1)}\\&\quad-\frac{\bar{J}_{G_1}}{\bar{J}_{G_2}}[\epsilon,\epsilon]_l^{(G_2)}\Bigg)^2-V_l^{\text{short}}\Bigg]+R_1\\&=\frac{m}{J}\sum_{l=1}^{K}\left[n^{1/3}\left([X,X]_l^{(G_1)}-\theta_l^{\text{short}}\right)^2-V_l^{\text{signal}}\right]\\&\quad+\frac{m}{J}\sum_{l=1}^{K}\left[n^{1/3}\left([\epsilon,\epsilon]_l^{(G_1)}-\frac{\bar{J}_{G_1}}{\bar{J}_{G_2}}[\epsilon,\epsilon]_l^{(G_2)}\right)^2-V_l^{\text{noise}}\right]\\&\quad+R_1+R_2.\end{aligned}$$

In the first step, we show negligibility of the signal part, i.e.,

$$\frac{m}{J}\sum_{l=1}^{K}\left[n^{1/3}\left([X,X]_l^{(G_1)}-\theta_l^{\text{short}}\right)^2-V_l^{\text{signal}}\right]=o_p(1). \tag{31}$$

For this, we adapt the arguments of Zhang et al. (2005) to the subsample. We have

$$[X,X]_l^{(G_1)}=[X,X]_l^{(1)}+S_l+R_3 \tag{32}$$

where

$$S_l=2\sum_{i=1}^{J-1}\left(\Delta X_{(l-1)m/n+i/n}\right)\sum_{j=1}^{G_1\wedge i}\left(1-\frac{j}{G_1}\right)\left(\Delta X_{(l-1)m/n+(i-j)/n}\right)$$

where $\Delta X_{i/n}=X_{i/n}-X_{(i-1)/n}$. R_3 arises due to the end effects, see Zhang et al. (2005), p.1410., and it satisfies $R_3=O_p\left(G_1n^{-1}\right)$.

The second term in (32) satisfies

$$\begin{aligned}S_l^2&=\left(2\sum_{i=1}^{J-1}\left(\Delta X_{[(l-1)m+i]/n}\right)\sum_{j=1}^{G_1\wedge i}\left(1-\frac{j}{G_1}\right)\left(\Delta X_{[(l-1)m+i-j]/n}\right)\right)^2\\&=4\sum_{i=1}^{J-1}\int_{[(l-1)m+i-1]/n}^{[(l-1)m+i]/n}\sigma_u^2\mathrm{d}u\Bigg(\sum_{j=1}^{G_1\wedge i}\left(1-\frac{j}{G_1}\right)\\&\quad\times\left(\Delta X_{[(l-1)m+i-j]/n}\right)\Bigg)^2+o_p\left(\frac{J}{n^{4/3}}\right)\\&=(I)+(II)+o_p\left(\frac{J}{n^{4/3}}\right)\end{aligned}$$

where

$$\begin{aligned}(I)&=4\sum_{i=1}^{J-1}\int_{[(l-1)m+i-1]/n}^{[(l-1)m+i]/n}\sigma_u^2\mathrm{d}u\sum_{j=1}^{G_1\wedge i}\left(1-\frac{j}{G_1}\right)^2\\&\quad\times\left(\Delta X_{[(l-1)m+i-j]/n}\right)^2\\&=4\sum_{i=1}^{J-1}\sigma^4_{[(l-1)m+i-1]/n}\frac{1}{n^2}\sum_{j=1}^{G_1\wedge i}\left(1-\frac{j}{G_1}\right)^2+o_p\left(\frac{J}{n^{4/3}}\right)\\&=\frac{4}{3}\frac{G_1}{n^2}\sum_{i=1}^{J-1}\sigma^4_{[(l-1)m+i-1]/n}+o_p\left(\frac{J}{n^{4/3}}\right)\end{aligned}$$

and

$$\begin{aligned}(II)&=8\sum_{i=1}^{J-1}\int_{[(l-1)m+i-1]/n}^{[(l-1)m+i]/n}\sigma_u^2\mathrm{d}u\sum_{k>r\ge 0}^{i-1}\left(\Delta X_{[(l-1)m+k]/n}\right)\\&\quad\times\left(\Delta X_{[(l-1)m+r]/n}\right)\left(1-\frac{i-k}{G_1}\right)^+\left(1-\frac{i-r}{G_1}\right)^+\\&=o_p\left(\frac{J}{n^{4/3}}\right).\end{aligned}$$

The last equality follows from Zhang et al. (2005), p.1410 and the fact that conditions $G_1=cn^{2/3}$ and $J>G_1$ imply $G_1/n<J/n^{4/3}$. Therefore,

$$\begin{aligned}S_l^2&=\frac{4}{3}\frac{G_1}{n}\int_{(l-1)m/n}^{[(l-1)m+J]/n}\sigma_u^4\mathrm{d}u+o_p\left(\frac{J}{n^{4/3}}\right)\\&=\frac{1}{n^{1/3}}V_l^{\text{signal}}+o_p\left(\frac{J}{n^{4/3}}\right).\end{aligned}$$

The final piece in (32) to deal with is to show

$$n^{1/3}\frac{m}{J}\sum_{l=1}^{K}\left([X,X]_l^{(1)}-\theta_l^{\text{short}}\right)^2=o_p(1),$$

which follows by following (a simpler version of) the steps of the proof of Proposition 2. Eq. (31) follows.

Next, we turn to the noise part and prove

$$\frac{m}{J}\sum_{l=1}^{K}\left[n^{1/3}\left([\epsilon,\epsilon]_l^{(G_1)}-\frac{\bar{J}_{G_1}}{\bar{J}_{G_2}}[\epsilon,\epsilon]_l^{(G_2)}\right)^2-V_l^{\text{noise}}\right]\xrightarrow{p}0. \quad (33)$$

Given that noise is a discrete time process, Proposition 1 of Aït-Sahalia et al. (2011) can be applied directly, with J instead of n (this is the number of observations used above) to obtain, for each l,

$$\frac{G_1}{\sqrt{J}}\left([\epsilon,\epsilon]_l^{(G_1)}-\frac{\bar{J}_{G_1}}{\bar{J}_{G_2}}[\epsilon,\epsilon]_l^{(G_2)}\right)$$
$$\Rightarrow N\left(0, 8\mathrm{Var}(\epsilon)^2+16\lim_{n\to\infty}\sum_{i=1}^{n}\mathrm{Cov}\left(\epsilon_0,\epsilon_{i/n}\right)^2\right).$$

Since noise is mixing over subsamples, we can apply the law of large numbers to obtain

$$\frac{1}{K}\sum_{l=1}^{K}\frac{G_1^2}{J}\left([\epsilon,\epsilon]_l^{(G_1)}-\frac{\bar{J}_{G_1}}{\bar{J}_{G_2}}[\epsilon,\epsilon]_l^{(G_2)}\right)^2$$
$$\xrightarrow{p}8\mathrm{Var}(\epsilon)^2+16\lim_{n\to\infty}\sum_{i=1}^{n}\mathrm{Cov}\left(\epsilon_0,\epsilon_{i/n}\right)^2=\frac{n}{J}c^2V_l^{\text{noise}},$$

which is equivalent to Eq. (33) given that $K=n/m$ and $G_1=cn^{2/3}$.

The final step to prove Lemma 7 is to show $R_1+R_2=o_p(1)$.

We have

$$\begin{aligned}R_1 &= \frac{m}{J}\sum_{l=1}^{K}n^{1/3}\left[\left(\frac{\bar{J}_{G_1}}{\bar{J}_{G_2}}[X,X]_l^{(G_2)}\right)^2\right]\\ &\quad+\frac{m}{J}\sum_{l=1}^{K}n^{1/3}\left[\left(2[X,\epsilon]_l^{(G_1)}\right)^2\right]\\ &\quad+\frac{m}{J}\sum_{l=1}^{K}n^{1/3}\left[\left(\frac{\bar{J}_{G_1}}{\bar{J}_{G_2}}2[X,\epsilon]_l^{(G_2)}\right)^2\right]+R_1'\end{aligned}$$

where, for $i=1,2$,

$$[X,\epsilon]_l^{(G_i)}=\frac{1}{G_1}\sum_{i=1}^{n-G_1}\left(X_{(i+G_1)/n}-X_{i/n}\right)\left(\epsilon_{(i+G_1)/n}-\epsilon_{i/n}\right).$$

The first term in R_1 is $o_p(1)$ because $[X,X]_l^{(G_2)}=O_p\left(Jn^{-1}\right)$ by substituting G_2 for G_1 in (31). The second and third terms are of $o_p(1)$ by proof of Lemma 1 of Aït-Sahalia et al. (2011), which implies, for $i=1,2$,

$$\mathrm{E}\left(\left([X,\epsilon]_l^{(G_i)}\right)^2|X\right)\le C\frac{1}{G_i^2}[X,X]_l^{(G_i)}.$$

The final terms R_1' and R_2 contain cross terms that are negligible by Cauchy–Schwarz inequality. □

A.5. Proof of Lemma 5

Most of the proof of the asymptotic distribution of the TSRV estimator of Aït-Sahalia et al. (2011) remains valid under the assumptions of Lemma 5. The noise component of the asymptotic distribution arises from the asymptotic distribution of

$$-2\frac{1}{\sqrt{n}}\sum_{i=0}^{n-G_1}\epsilon_{i/n}\epsilon_{(i+G_1)/n}+2\frac{1}{\sqrt{n}}\sum_{i=0}^{n-G_2}\epsilon_{i/n}\epsilon_{(i+G_2)/n},$$

see page 26 of Aït-Sahalia et al. (2011). Given that $G_1/G_2\to0$ and

$$\omega\left(\frac{i+G_1}{n}\right)-\omega\left(\frac{i}{n}\right)\le C\frac{G_1}{n}$$

due to differentiability of ω, the desired result follows. □

A.6. Proof of Theorem 6

Assume n is divisible by m by simplicity. As a first step, we prove

$$G^{(n)}=\frac{mJ}{n^2}\sum_{l=1}^{n/m}\tau_n^2\left(\frac{n}{J}\widehat{\theta}_l^{\text{short}}-\frac{n}{J}\theta_l^{\text{short}}\right)^2\xrightarrow{p}V. \quad (34)$$

For any two subsamples l and l' s.t. $l\neq l'$, $\zeta_l^{(n)}$ has no common returns with $\zeta_{l'}^{(n)}$. Therefore, $\zeta_l^{(n)}$ is strong mixing because $R^{(n)}$ is. Moreover, if we define

$$\psi_i^{(n)}=\zeta_l^{(n)}-\mathrm{E}\left(\zeta_l^{(n)}\right),$$

it is also strong mixing. Therefore, under A6, $\psi_i^{(n)}$ is a uniformly integrable L^1-mixingale as defined in Andrews (1998), to which we can apply Theorem 2 of Andrews (1998) to obtain

$$\frac{m}{n}\sum_{l=1}^{n/m}\psi_i^{(n)}=\frac{m}{n}\sum_{l=1}^{n/m}\left[\zeta_l^{(n)}-\mathrm{E}\left(\zeta_l^{(n)}\right)\right]\xrightarrow{p}0.$$

By A4, we have

$$\begin{aligned}\frac{m}{n}\sum_{l=1}^{n/m}\zeta_l^{(n)} &= \frac{m}{n}\sum_{l=1}^{n/m}\frac{n}{J}\left[\tau_n^2\left(\widehat{\theta}_l^{\text{short}}-\theta_l^{\text{short}}\right)^2-V_l^{\text{short}}\right]\xrightarrow{p}0\\ &= \frac{m}{n}\sum_{l=1}^{n/m}\frac{n}{J}\tau_n^2\left(\widehat{\theta}_l^{\text{short}}-\theta_l^{\text{short}}\right)^2-\frac{m}{n}\sum_{l=1}^{n/m}\frac{n}{J}V_l^{\text{short}}\xrightarrow{p}0\\ &= \frac{m}{n}\sum_{l=1}^{n/m}\frac{n}{J}\tau_n^2\left(\widehat{\theta}_l^{\text{short}}-\theta_l^{\text{short}}\right)^2\xrightarrow{p}V\end{aligned}$$

and so (34) follows.

In a second step, we prove that $G^{(n)}-\widehat{V}_{\text{sub}}\xrightarrow{p}0$.

$$\begin{aligned}\widehat{V}_{\text{sub}}-G^{(n)} &= \frac{Jm\tau_n^2}{n^2}\sum_{l=1}^{K}\left(\frac{n}{J}\widehat{\theta}_l^{\text{short}}-\frac{n}{m}\widehat{\theta}_l^{\text{long}}\right)^2\\ &\quad-\frac{Jm\tau_n^2}{n^2}\sum_{l=1}^{n/m}\tau_n^2\left(\frac{n}{J}\widehat{\theta}_l^{\text{short}}-\frac{n}{J}\theta_l^{\text{short}}\right)^2\\ &= \frac{Jm\tau_n^2}{n^2}\sum_{l=1}^{K}\left(\frac{n}{m}\widehat{\theta}_l^{\text{long}}-\frac{n}{m}\theta_l^{\text{long}}\right)^2\\ &\quad+\frac{Jm\tau_n^2}{n^2}\sum_{l=1}^{K}\left(\frac{n}{m}\theta_l^{\text{long}}-\frac{n}{J}\theta_l^{\text{short}}\right)^2\\ &\quad+2\frac{Jm\tau_n^2}{n^2}\sum_{l=1}^{K}\left(\frac{n}{m}\widehat{\theta}_l^{\text{long}}-\frac{n}{m}\theta_l^{\text{long}}\right)\\ &\quad\times\left(\frac{n}{m}\theta_l^{\text{long}}-\frac{n}{J}\theta_l^{\text{short}}\right)\\ &\quad-2\frac{Jm\tau_n^2}{n^2}\sum_{l=1}^{n/m}\tau_n^2\left(\frac{n}{J}\widehat{\theta}_l^{\text{short}}-\frac{n}{J}\theta_l^{\text{short}}\right)\\ &\quad\times\left(\frac{n}{m}\widehat{\theta}_l^{\text{long}}-\frac{n}{J}\theta_l^{\text{short}}\right).\end{aligned} \quad (35)$$

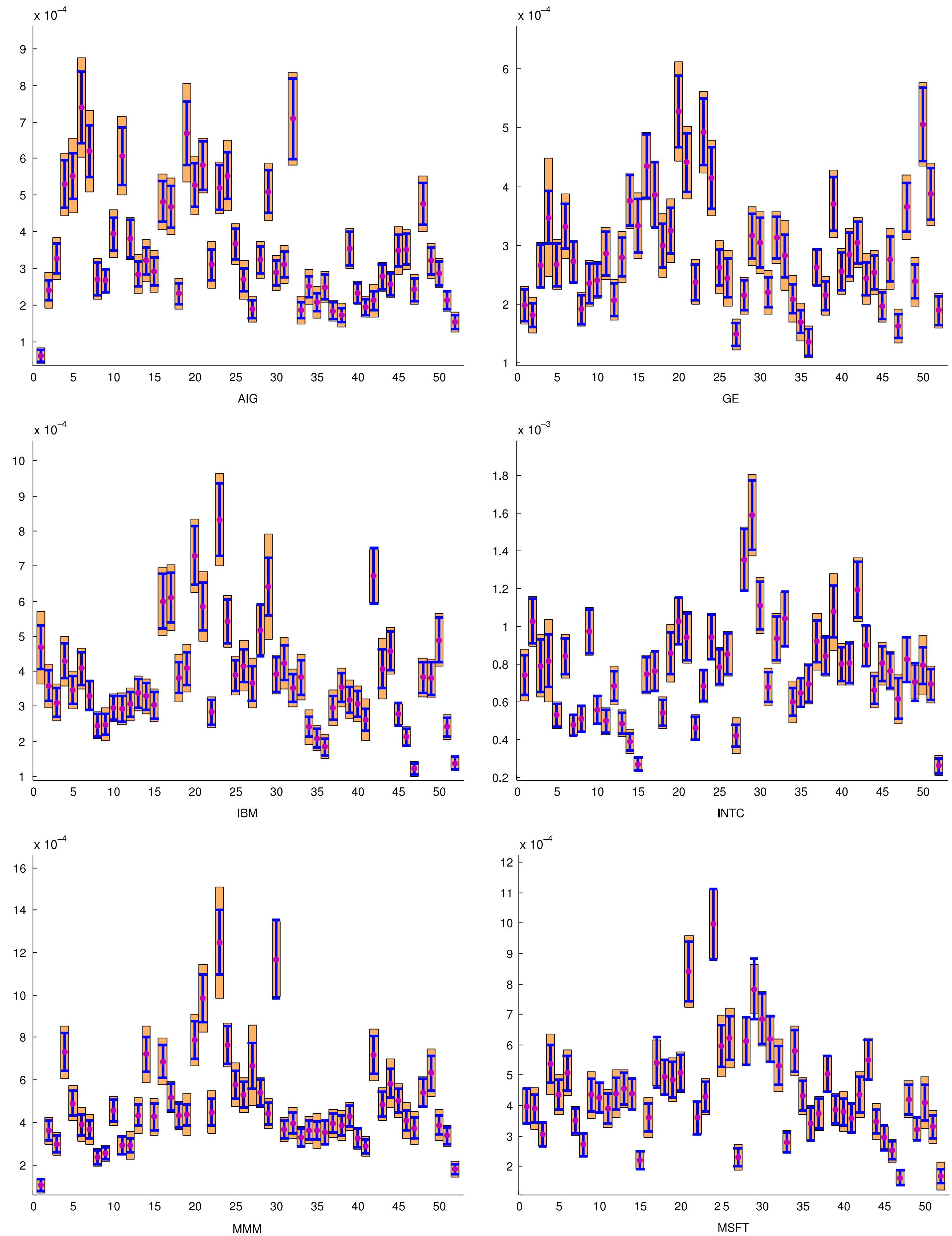

Fig. 6. 95% Confidence Intervals (CI's) for weekly IV, for each of 52 weeks in 2006, calculated using the subsampling method (CI's with bars) or $\widehat{V}_b$ (CI's with lines). TSRV is the middle of all CI's by construction.

Table 2
Coverage probabilities of 95% confidence interval of IV, $\lambda = 0.0001$.

		$J = 200$			$J = 500$			$\widehat{V}_a$
	m	800	2000	3000	2000	5000	7500	
$\rho = 0$	*Two-sided*	0.97	0.98	0.98	0.93	0.95	0.95	0.92
	Left-sided	0.95	0.96	0.96	0.92	0.94	0.94	0.91
	Right-sided	0.98	0.98	0.99	0.96	0.97	0.97	0.95
$\rho = -0.3$	*Two-sided*	0.97	0.98	0.98	0.93	0.96	0.96	0.92
	Left-sided	0.95	0.97	0.97	0.92	0.94	0.94	0.91
	Right-sided	0.98	0.98	0.98	0.96	0.97	0.97	0.96
$\rho = -0.5$	*Two-sided*	0.97	0.98	0.98	0.93	0.95	0.95	0.92
	Left-sided	0.95	0.97	0.97	0.92	0.94	0.94	0.92
	Right-sided	0.97	0.98	0.98	0.95	0.96	0.96	0.94
$\rho = -0.7$	*Two-sided*	0.97	0.98	0.98	0.94	0.96	0.96	0.91
	Left-sided	0.95	0.97	0.97	0.92	0.94	0.94	0.9
	Right-sided	0.98	0.98	0.98	0.96	0.97	0.97	0.96

Table 3
Coverage probabilities of 95% confidence interval of IV, $\lambda = 0.001$.

		$J = 200$			$J = 500$			$\widehat{V}_a$
	m	800	2000	3000	2000	5000	7500	
$\rho = 0$	*Two-sided*	0.97	0.98	0.98	0.93	0.95	0.95	0.92
	Left-sided	0.95	0.96	0.96	0.92	0.93	0.94	0.91
	Right-sided	0.98	0.99	0.99	0.96	0.97	0.97	0.95
$\rho = -0.3$	*Two-sided*	0.97	0.97	0.98	0.93	0.96	0.96	0.91
	Left-sided	0.95	0.96	0.97	0.91	0.94	0.94	0.9
	Right-sided	0.98	0.98	0.98	0.96	0.97	0.98	0.96
$\rho = -0.5$	*Two-sided*	0.97	0.97	0.98	0.94	0.96	0.96	0.9
	Left-sided	0.95	0.96	0.97	0.91	0.94	0.94	0.89
	Right-sided	0.98	0.98	0.98	0.96	0.97	0.98	0.95
$\rho = -0.7$	*Two-sided*	0.98	0.98	0.98	0.94	0.96	0.96	0.88
	Left-sided	0.96	0.97	0.97	0.93	0.94	0.95	0.9
	Right-sided	0.97	0.98	0.98	0.95	0.96	0.96	0.91

We have the following decomposition,

$$
\begin{aligned}
&\left(\frac{n}{J}\theta_l^{\text{short}} - \frac{n}{m}\theta_l^{\text{long}}\right)^2 \\
&= \left(\frac{n}{J}\int_{(l-1)m/n}^{[(l-1)m+J]/n} g(u)\mathrm{d}u - \frac{n}{m}\int_{(l-1)m/n}^{lm/n} g(u)\mathrm{d}u\right)^2 \\
&\leq \left(\frac{n}{J}\int_{(l-1)m/n}^{[(l-1)m+J]/n} (g(u) - g((l-1)m/n))\,\mathrm{d}u\right)^2 \\
&\quad + \left(\frac{n}{m}\int_{(l-1)m/n}^{lm/n} (g(u) - g((l-1)m/n))\,\mathrm{d}u\right)^2 \\
&\quad + 2\left|\frac{n}{J}\int_{(l-1)m/n}^{[(l-1)m+J]/n} (g(u) - g((l-1)m/n))\,\mathrm{d}u\right| \\
&\quad \times \left|\frac{n}{m}\int_{(l-1)m/n}^{lm/n} (g(u) - g((l-1)m/n))\,\mathrm{d}u\right|.
\end{aligned}
$$

These terms are small enough due to A4 and (27) as follows,

$$
\begin{aligned}
&\mathrm{E}\left|\frac{Jm\tau_n^2}{n^2}\sum_{l=1}^{K}\left(\frac{n}{m}\int_{(l-1)m/n}^{lm/n} (f(u) - f((l-1)m/n))\,\mathrm{d}u\right)^2\right| \\
&\leq \frac{Jm\tau_n^2}{n^2}\sum_{l=1}^{K}\mathrm{E}\left(\frac{n}{m}\int_{(l-1)m/n}^{lm/n} (f(u) - f((l-1)m/n))\,\mathrm{d}u\right)^2 \\
&= \frac{Jm\tau_n^2}{n^2}\sum_{l=1}^{K}\mathrm{E}\left(f(s_l) - f((l-1)m/n)\right)^2 \\
&\leq C\frac{Jm\tau_n^2}{n^2}\sum_{l=1}^{K}\mathrm{E}\left(\sigma(s_l) - \sigma((l-1)m/n)\right)^2 \\
&\leq C\frac{Jm\tau_n^2}{n^2}\sum_{l=1}^{K}\frac{m}{n} = C\frac{Jm\tau_n^2}{n^2} \to 0 \qquad (36)
\end{aligned}
$$

by assumption. In the above, the first equality follows by the mean value theorem, which applies by differentiability of

$$\int_{(l-1)m/n}^{t} (f(u) - f((l-1)m/n))\,\mathrm{d}u \qquad (37)$$

in time.

Next, we show

$$\frac{Jm\tau_n^2}{n^2}\sum_{l=1}^{K}\left(\frac{n}{m}\widehat{\theta}_l^{\text{long}} - \frac{n}{m}\theta_l^{\text{long}}\right)^2 \xrightarrow{p} 0.$$

By substituting m for J in

$$G^{(n)} = \frac{mJ}{n^2}\sum_{l=1}^{n/m}\tau_n^2\left(\frac{n}{J}\widehat{\theta}_l^{\text{short}} - \frac{n}{J}\theta_l^{\text{short}}\right)^2 \xrightarrow{p} V,$$

we obtain

$$\frac{m^2}{n^2}\sum_{l=1}^{n/m}\tau_n^2\left(\frac{n}{m}\widehat{\theta}_l^{\text{long}} - \frac{n}{m}\theta_l^{\text{long}}\right)^2 \xrightarrow{p} V,$$

and so by multiplying the left hand side by J/m, (37) follows since $J/m \to 0$.

The remaining cross-terms in (35) are negligible by the above results and Cauchy–Schwarz inequality. This concludes the proof of Theorem 6. □

Appendix B. Tables and figures

See Tables 2–6.

Table 4
Coverage probabilities of 95% confidence interval of IV_X, $\lambda = 0.01$.

		$J = 200$			$J = 500$			$\widehat{V}_a$
	m	800	2000	3000	2000	5000	7500	
$\rho = 0$	*Two-sided*	0.97	0.98	0.98	0.94	0.96	0.96	0.92
	Left-sided	0.95	0.96	0.97	0.92	0.94	0.94	0.9
	Right-sided	0.98	0.98	0.98	0.97	0.98	0.98	0.96
$\rho = -0.3$	*Two-sided*	0.97	0.98	0.98	0.93	0.96	0.96	0.82
	Left-sided	0.96	0.97	0.97	0.94	0.95	0.95	0.85
	Right-sided	0.97	0.98	0.98	0.94	0.96	0.96	0.88
$\rho = -0.5$	*Two-sided*	0.98	0.98	0.98	0.94	0.96	0.96	0.7
	Left-sided	0.95	0.96	0.96	0.93	0.94	0.94	0.8
	Right-sided	0.97	0.98	0.98	0.96	0.97	0.97	0.84
$\rho = -0.7$	*Two-sided*	0.96	0.97	0.98	0.94	0.96	0.95	0.77
	Left-sided	0.94	0.95	0.95	0.92	0.94	0.94	0.83
	Right-sided	0.97	0.98	0.98	0.96	0.97	0.97	0.84

Table 5
Heteroscedastic noise. Coverage probabilities of 95% confidence interval of IV.

		$J = 200$			$J = 500$			$\widehat{V}_a$
	m	800	2000	3000	2000	5000	7500	
$\lambda = 0.0001$	*Two-sided*	0.96	0.98	0.98	0.93	0.95	0.96	0.94
	Left-sided	0.95	0.96	0.96	0.91	0.93	0.93	0.92
	Right-sided	0.98	0.99	0.99	0.97	0.97	0.98	0.97
$\lambda = 0.001$	*Two-sided*	0.97	0.98	0.98	0.93	0.95	0.96	0.94
	Left-sided	0.95	0.96	0.96	0.92	0.93	0.94	0.93
	Right-sided	0.97	0.99	0.99	0.95	0.96	0.97	0.96
$\lambda = 0.01$	*Two-sided*	0.96	0.97	0.98	0.93	0.94	0.95	0.97
	Left-sided	0.93	0.94	0.95	0.90	0.91	0.91	0.94
	Right-sided	0.99	0.99	0.99	0.98	0.98	0.98	0.98

Table 6
Summary of data manipulations.

	Raw data	Step 1: flat trading	Step 2: jumps	Step 3: gradual jumps
MMM	1,797,107	983,705 (54.74%)	2567 (0.14%)	5,963 (0.33%)
MSFT	18,738,034	16,364,458 (87.33%)	5563 (0.03%)	18,795 (0.10%)
IBM	2,786,649	1,556,475 (55.85%)	3706 (0.13%)	7,525 (0.27%)
AIG	2,807,065	1,749,345 (62.32%)	3179 (0.11%)	10,433 (0.37%)
GE	7,288,596	5,449,832 (74.77%)	3707 (0.05%)	12,991 (0.18%)
INTC	21,155,095	18,498,295 (87.44%)	5794 (0.03%)	21,119 (0.10%)

References

Aït-Sahalia, Y., Jacod, J., 2009. Testing for jumps in a discretely observed process. Annals of Statistics 37, 184–222.

Aït-Sahalia, Y., Mykland, P., Zhang, L., 2011. Ultra high frequency volatility estimation with dependent microstructure noise. Journal of Econometrics 160, 190–203.

Aldous, D.G., Eagleson, G.K., 1978. On mixing and stability of limit theorems. The Annals of Probability 6, 325–331.

Andersen, T.G., Bollerslev, T., 1997. Intraday periodicity and volatility persistence in financial markets. Journal of Empirical Finance 5, 115–158.

Andersen, T.G., Bollerslev, T., 1998. Answering the skeptics: yes, standard volatility models do provide accurate forecasts. International Economic Review 39 (4), 885–905.

Andersen, T.G., Bollerslev, T., Diebold, F.X., Labys, P., 2001. The distribution of exchange rate volatility. Journal of the American Statistical Association 96, 42–55. Correction published in 2003, volume 98, page 501.

Andreou, E., Ghysels, E., 2002. Rolling-sample volatility estimators: some new theoretical, simulations and empirical results. Journal of Business and Economic Statistics 20, 363–375.

Andrews, D.W.K., 1998. Laws of large numbers for dependent non-identically distributed random variables. Econometric Theory 4, 458–467.

Bandi, F.M., Russell, J.R., 2006. Separating microstructure noise from volatility. Journal of Financial Economics 79, 655–692.

Bandi, F.M., Russell, J.R., 2008. Microstructure noise, realized variance, and optimal sampling. Review of Economic Studies 75, 339–369.

Barndorff-Nielsen, O.E., Shephard, N., Winkel, M., 2005. Limit theorems for multipower variation in the presence of jumps. Stochastic Processes and Applications 116, 796–806.

Barndorff-Nielsen, O.E., Graversen, S.E., Jacod, J., Shephard, N., 2006. Limit theorems for bipower variation in financial econometrics. Econometric Theory 22, 677–719.

Barndorff-Nielsen, O.E., Hansen, P.R., Lunde, A., Shephard, N., 2008. Designing realised kernels to measure the ex-post variation of equity prices in the presence of noise. Econometrica 76 (6), 1481–1536.

Barndorff-Nielsen, O.E., Hansen, P.R., Lunde, A., Shephard, N., 2009. Realised kernels in practice: trades and quotes. Econometrics Journal 12, C1–C32.

Barndorff-Nielsen, O.E., Shephard, N., 2002. Econometric analysis of realised volatility and its use in estimating stochastic volatility models. Journal of the Royal Statistical Society. Series B 64, 253–280.

Christensen, K., Oomen, R.C.A., Podolskij, M., 2010. Realised quantile-based estimation of the integrated variance. Journal of Econometrics 159, 74–98.

Christensen, K., Podolskij, M., Vetter, M., 2009. Bias-correcting the realized range-based variance in the presence of market microstructure noise. Finance and Stochastics 13, 239–268.

Delbaen, F., Schachermayer, W., 1994. A general version of the fundamental theorem of asset pricing. Mathematische Annalen 300 (3), 463–520.

Foster, D.P., Nelson, D.B., 1996. Continuous record asymptotics for rolling sample variance estimators. Econometrica 64, 139–174.

Fleming, J., Kirby, C., Ostdiek, B., 2003. The economic value of volatility timing using "realized" volatility. Journal of Financial Economics 67, 473–509.

Gerety, M.S., Mulherin, H., 1994. Price formation on stock exchanges: the evolution of trading within the day. Review of Financial Studies 6, 23–56.

Giot, P., Laurent, S., 2004. Modelling daily value-at-risk using realized volatility and ARCH type models. Journal of Empirical Finance 11, 379–398.

Gonçalves, S., Meddahi, N., 2009. Bootstrapping realized volatility. Econometrica 77, 283–306.

Griffin, J., Oomen, R., 2008. Sampling returns for realized variance calculations: tick time or transaction time? Econometric Reviews 27, 230–253.

Hansen, P.R., Lunde, A., 2006. Realized variance and market microstructure noise (with comments and rejoinder). Journal of Business and Economic Statistics 24, 127–218.

Harris, L., 1986. A transaction data study of weekly and intradaily patterns in stock returns. Journal of Financial Economics 16 (1), 99–117.

Heston, S., 1993. A closed-form solution for options with stochastic volatility with applications to bonds and currency options. Review of Financial Studies 6, 327–343.
Jacod, J., 2007. Statistics and high frequency data. Lecture notes. Séminaire Européen de Statistique 2007.
Jacod, J., 2008. Asymptotic properties of realized power variations and related functionals of semimartingales. Stochastic Processes and their Applications 118, 517–559.
Jacod, J., Li, Y., Mykland, P.A., Podolskij, M., Vetter, M., 2009. Microstructure noise in the continuous case: the pre-averaging approach. Stochastic Processes and their Applications 119, 2249–2276.
Jacod, J., Shiryaev, A.N., 2003. Limit Theorems for Stochastic Processes. Springer.
Kalnina, I., Linton, O.B., 2008. Estimating quadratic variation consistently in the presence of correlated measurement error. Journal of Econometrics 147, 47–59.
Kleidon, A., Werner, I., 1996. UK and US trading of British cross-listed stocks: an intraday analysis of market integration. Review of Financial Studies 9, 619–644.
Kristensen, D., 2010. Nonparametric filtering of the realised spot volatility: a kernel-based approach. Econometric Theory 26, 60–93.
Lahiri, S.N., 1996. On inconsistency of estimators based on spatial data under infill asymptotics. Sankya: The Indial Journal of Statistics, Series A 58, 403–417.
Lahiri, S.N., Kaiser, M.S., Cressie, N., Hsu, N., 1999. Prediction of spatial cumulative distribution functions using subsampling. Journal of the American Statistical Association 94, 86–97.
Lockwood, L.J., Linn, S.C., 1990. An examination of stock market return volatility during overnight and intraday periods, 1964–1989. Journal of Finance 45, 591–601.
Mancini, C., 2004. Estimation of the characteristics of the jumps of a general Poisson-diffusion model. Scandinavian Actuarial Journal 2004 (1), 42–52.
McInish, T.H., Wood, R.A., 1992. An analysis of intraday patterns in bid/ask spreads for NYSE stocks. Journal of Finance 47, 753–764.
Meddahi, N., Mykland, P., 2010. Fat tails or many small jumps? The near-diffusion paradigm. Work in Progress.
Mikosch, T., Starica, C., 2003. Stock market risk-return inference. An unconditional, non-parametric approach. Available at SSRN: http://ssrn.com/abstract=882820.
Mykland, P., Zhang, L., 2009. Inference for continuous semimartingales observed at high frequency: a general approach. Econometrica 77, 1403–1445.
Podolskij, M., 2006. New theory on estimation of integrated volatility with applications. Ph.D. Thesis. Bochum University.
Podolskij, M., Ziggel, D., 2007. Boostrapping bipower variation. Technical Report. Ruhr-University of Bochum.
Politis, D.N., Romano, J.P., 1994. Large sample confidence regions based on subsamples under minimal assumptions. Annals of Statistics 22, 2031–2050.
Politis, D.N., Romano, J.P., Wolf, M., 1997. Subsampling for heteroscedastic time series. Journal of Econometrics 81, 281–317.
Politis, D.N., Romano, J.P., Wolf, M., 1999. Subsampling. Springer-Verlag, New York.
Revuz, D., Yor, M., 2005. Continuous Martingales and Brownian Motion. Springer-Verlag, New York.
Zhang, L., 2006. Efficient estimation of stochastic volatility using noisy observations: a multi-scale approach. Bernoulli 12 (6), 1019–1043.
Zhang, L., Mykland, P., Aït-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.
Zhou, B., 1996. High-frequency data and volatility in foreign-exchange rates. Journal of Business and Economic Statistics 14, 45–52.