

# Subsampling High Frequency Data\*

Ilze Kalnina<sup>†</sup>

The London School of Economics

December 26, 2008

## Abstract

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 an automated method of inference that does not rely on the exact form of the asymptotic variance. The need for a new approach is motivated by the failure of traditional bootstrap and subsampling variance estimators with high frequency data, which is explained in the paper. 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 is applied to the integrated volatility estimator of Aït-Sahalia et al. (2006a) in the presence of autocorrelated and heteroscedastic market microstructure noise, for which there is no alternative inferential method in the literature. Monte Carlo study illustrates the finite sample properties of the proposed method.

---

\*I would like to thank Donald Andrews, Peter Bühlmann, Valentina Corradi, Kirill Evdokimov, Yuichi Kitamura, Nour Meddahi, Per Mykland, Mark Podolskij, 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.

<sup>†</sup>Department of Economics, London School of Economics, Houghton Street, London WC2A 2AE, United Kingdom. E-mail address: [i.kalnina@lse.ac.uk](mailto:i.kalnina@lse.ac.uk)

# 1 Introduction

This paper proposes the first automated method for conducting inference with high frequency data. In particular, it proposes to estimate the asymptotic variance of some estimator without relying on the exact expression of the asymptotic variance. In the traditional stationary time series framework, this task can be accomplished by bootstrap and subsampling variance estimators, but these are inconsistent with high frequency data.

A new subsampling method is developed, which enables 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 inference problem for the integrated variance estimator of Aït-Sahalia et al. (2006a), 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, Bollerslev, Diebold, and Labys, 2001). With moderate frequency data, say 5 or 15 minute data, this can be estimated by the so called realized variance (RV), a sum of squared returns. The nonparametric nature of realized variance and the simplicity of its calculation have made it popular among practitioners. It has been used for asset allocation (Fleming, Kirby, and Ostdiek, 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. Pricing of these derivatives is done according to 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  $\mu$ ,  $\sigma$ , 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

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

Realized variance (or empirical quadratic variation) is a consistent estimator of integrated variance in infill asymptotics, i.e., when the 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 minutes. 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, Mykland, and Aït-Sahalia (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 Time Scale estimator. Consistent estimators in this framework were also proposed by Barndorff-Nielsen, Hansen, Lunde, and Shephard (2008a), Christensen, Oomen, and Podolskij (2008), Christensen, Podolskij, and Vetter (2006), and Jacod, Li, Mykland, Podolskij, and Vetter (2007). Aït-Sahalia, Mykland, and Zhang (2006a) extend the Two Time Scale 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 Two Time Scale estimator. The method proposed in this paper can be used to conduct inference for the Two Time Scale estimator in 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.<sup>1</sup>

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 even in the simple case of RV, this procedure is inconsistent, as the sample variance of the bootstrap

---

<sup>1</sup>See Andersen and Bollerslev (1997), Gerety and Mulherin (1994), Harris (1986), Lockwood and Linn (1990), and McInish and Wood (1992).

estimator does not converge to the asymptotic variance of the original estimator, see Goncalvez and Meddahi (2008).

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  $\hat{\theta}_n$  (think of i.i.d.  $Y_i$ 's, a parameter of interest  $\theta = E(Y)$ , and  $\hat{\theta}_n = \frac{1}{n} \sum Y_i$ ). Suppose we know its asymptotic distribution

$$\tau_n(\hat{\theta}_n - \theta) \implies N(0, V)$$

as  $n \rightarrow \infty$ , where  $\implies$  denotes convergence in distribution, and  $\tau_n$  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  $\hat{\theta}_n$ . This can be done with the help of many subsamples, for which the estimator  $\hat{\theta}_n$  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  $m = m(n) \rightarrow \infty$  as  $n \rightarrow \infty$  but  $m/n \rightarrow 0$ . Denote by  $\hat{\theta}_{n,m,l}$  the estimator  $\hat{\theta}_n$  calculated using the  $l^{\text{th}}$  block of  $m$  observations, with  $n$  being the total number of observations. Then, the asymptotic distribution of  $\tau_m(\hat{\theta}_{n,m,l} - \theta)$  is the same, i.e.

$$\tau_m(\hat{\theta}_{n,m,l} - \theta) \implies N(0, V) \tag{2}$$

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

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

and we have

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

where  $\xrightarrow{p}$  denotes convergence in probability. Notice that the estimator in (3) is like average of squared  $\tau_m(\hat{\theta}_{n,m,l} - \theta)$  over all subsamples, except that  $\hat{\theta}_n$  plays the role of  $\theta$ . The difference between  $\hat{\theta}_n$  and  $\theta$  is negligible because  $\hat{\theta}_n$  converges faster to  $\theta$  than  $\hat{\theta}_{n,m,l}$  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,  $\hat{\theta}_n$  is taken to be realized variance and  $\theta$  its probability limit, integrated variance. The intuition behind the failure is straightforward. The problem is that  $\hat{\theta}_{n,m,l}$  and  $\hat{\theta}_n$  do not converge to the same quantity and so (2) cannot be satisfied. The underlying

reason is that the spot (or infinitesimal) volatility  $\sigma_t$  is changing over time. The estimator calculated on a small block cannot estimate the integrated variance  $\theta$ , because  $\theta$  contains information about spot volatility on the whole interval.

A novel subsampling scheme is proposed that can estimate the asymptotic variance of RV. Importantly, it can also be applied to the Two Time Scale estimator of Aït-Sahalia et al. (2006a), in 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 can be potentially useful. Section 3 shows how inference can be conducted for the Two Time Scale 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 methodology and intuition behind it is illustrated with the example of realized volatility. The first subsection explains the failure of bootstrap and subsampling (Politis and Romano, 1994) methods for the estimation of the asymptotic variance of RV. The second subsection introduces a new subsampling scheme that can estimate the asymptotic variance of RV consistently. The third subsection describes an alternative scheme that is of theoretical interest, but which will not be used beyond the RV example.

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

$$dX_t = \mu_t dt + \sigma_t dW_t, \tag{4}$$

where  $W_t$  is standard Brownian motion, the stochastic process  $\mu_t$  is locally bounded, and  $\sigma_t$  is a càdlàg spot volatility process.<sup>2</sup> 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

---

<sup>2</sup>In other words, the sample paths of the volatility process are left continuous with right limits.

observations is  $1/n$ . The asymptotic scheme is infill as  $n \rightarrow \infty$ .

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

$$IV_X = \int_0^1 \sigma_s^2 ds. \quad (5)$$

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

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

This satisfies

$$\begin{aligned} \sqrt{n}(RV_n - IV_X) &\implies MN(0, V) \\ V &= 2IQ = 2 \int_0^1 \sigma_s^4 ds \end{aligned} \quad (7)$$

where  $MN(0, V)$  denotes a mixed normal distribution with random conditional variance  $V$  independent of the underlying normal distribution.<sup>3</sup> The convergence (7) follows from Barndorff-Nielsen and Shephard (2002) and Jacod (2006), 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 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,  $\tilde{V} = 2IQ_n$ , where realized quarticity is sum of fourth powers of returns, properly scaled,

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

The estimator  $\tilde{V}$  is consistent for  $V$  in the sense that  $\tilde{V}/V \xrightarrow{p} 1$ . This result allows the construction of consistent confidence intervals for  $QV_X$ . For example, a two-sided level  $\alpha$  interval is given by  $\tilde{\mathcal{C}}_\alpha = RV_n \pm z_{\alpha/2} \tilde{V}^{1/2} / \sqrt{n}$ , where  $z_\alpha$  is the  $\alpha$  quantile from a standard normal distribution, and this has the property that  $\Pr[IV_X \in \tilde{\mathcal{C}}_\alpha] \rightarrow 1 - \alpha$ . Mykland and Zhang (2006,7) have proposed an alternative estimator of  $V$  that is more efficient than  $\tilde{V}$  under the sampling scheme (4) and can also be used to construct intervals based on the studentized limit theory.

The next subsection explains why the usual bootstrap and subsampling methods cannot be used to estimate  $V$  in this framework. Then, a new subsampling method is introduced. Section 2 concludes with description of an alternative subsampling scheme.

---

<sup>3</sup>In other words, the limiting p.d.f. is of the form  $f(x) = \int \phi_{0,v}(x) f_V(v) dv$ , where  $f_V$  denotes the p.d.f. of  $V$  and  $\phi_{0,v}(x) = \exp(-x^2/2v^2) / \sqrt{2\pi v}$

## 2.1 Failure of the Traditional Resampling Schemes

Recently, Goncalvez and Meddahi (2008) have proposed a bootstrap algorithm for RV. 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. An essential feature of their procedures is reliance on an estimator of the asymptotic variance, which is not always available. A more widely used bootstrap procedure is to estimate asymptotic variance as the sample variance of the bootstrap statistic. This procedure is simple, but inconsistent in the high frequency framework, as even in the simple case of RV, the bootstrap estimator has a different asymptotic variance than the original estimator, see Goncalvez and Meddahi (2008). This means that confidence intervals constructed using the usual bootstrap method are inconsistent.

We now consider the popular method of Politis and Romano (1994). This subsampling scheme fails in our setting, highlighting the difference that high frequency framework brings. It is however instructive to consider, as subsequent methods proposed use the same underlying idea.

Let  $\hat{\theta}_n$  be the RV calculated on the full sample, and let  $\hat{\theta}_{n,m,l}$  be the RV calculated on the  $l^{\text{th}}$  block of  $m$  observations,<sup>4</sup>

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

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

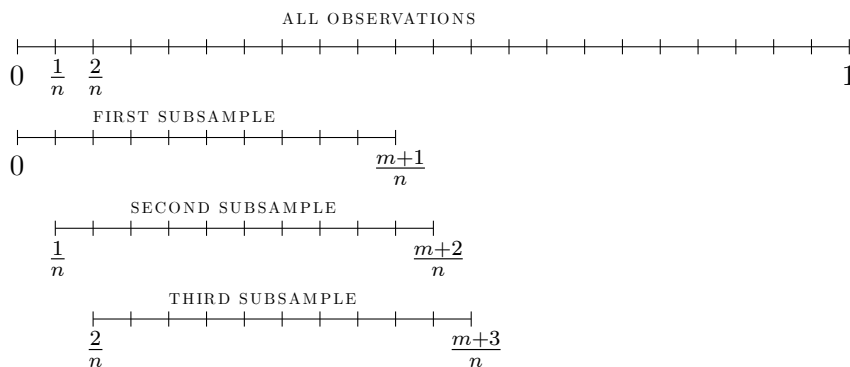


Figure 1: *The Subsampling Scheme of Politis and Romano (1994).*

<sup>4</sup>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 Figures 1,2, and 3 show maximum overlap versions of the subsampling schemes.

Assumption 5.3.1 of Politis, Romano and Wolf (1999) is satisfied, i.e., the sampling distribution of  $\tau_n(\widehat{\theta}_n - \theta)$  converges weakly. Therefore, 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  $V$ . The estimator on the full sample converges to the true value,  $\widehat{\theta}_n \rightarrow^p \theta$ . 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  $m/n \rightarrow 0$ . Thus,  $\left( \widehat{\theta}_{n,m,l} - \widehat{\theta}_n \right)^2$  converges to  $\theta$  and  $\widehat{V}_{PR}$  is asymptotically equal to  $m\theta^2$ . Notice that the value  $\theta^2$  is not related to  $V$ , which is the parameter of interest. A formal proof of the following proposition is in the appendix,

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

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

A crucial ingredient of the subsampling method of Politis and Romano is that  $\widehat{\theta}_{n,m,l}$  and  $\widehat{\theta}_n$  estimate the same quantity. A direct application of their method to high frequency framework violates this basic principle. Part of the reason is the different rates of magnitude. This could be accounted for by using  $\frac{m}{n}\widehat{\theta}_{n,m,l}$  instead of  $\widehat{\theta}_{n,m,l}$ . In this case, it still holds that  $\frac{n}{m}\widehat{\theta}_{n,m,l} - \widehat{\theta}_n \xrightarrow{p} 0$ . 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  $\theta$ .<sup>5</sup> Therefore, the underlying reason for the failure of the subsampling method of Politis and Romano is the fact that the spot variance changes over time.

## 2.2 The New Subsampling Scheme

We now introduce and motivate the new subsampling scheme. The current subsection describes this scheme for the RV example, and Section 3 applies it to the Two Time Scale 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

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

---

<sup>5</sup>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 (2005), and Kristensen (2008).

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

$$\theta_l = \int_{(l-1)m/n}^{lm/n} \sigma^2(u) du. \quad (9)$$

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  $\theta$ , 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  $\theta_l$ , 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

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

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,

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

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  $\theta_l$ . 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 notation  $\widehat{\theta}_l^{long}$  and  $\widehat{\theta}_l^{short}$  instead of  $\widehat{\theta}_{n,m,l}$  and  $\widehat{\theta}_{n,J,l}$ .

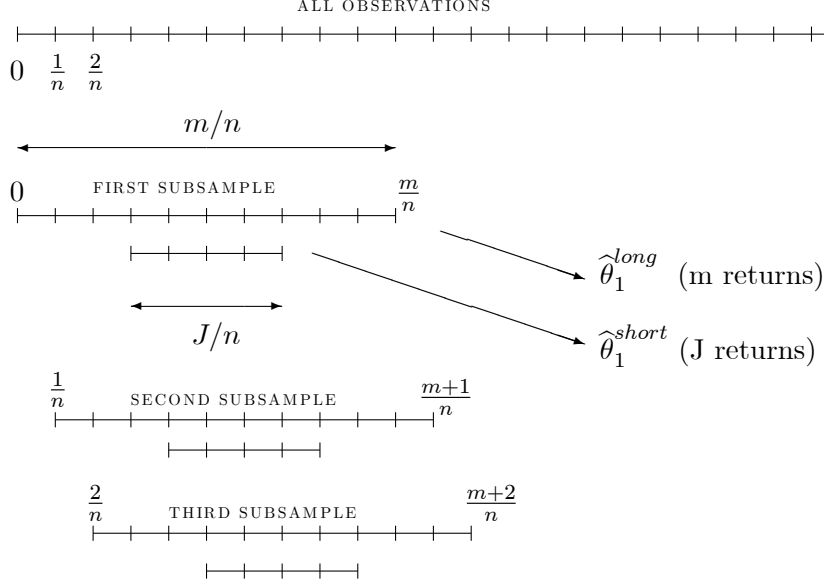


Figure 2: *The New Subsampling Scheme*

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

$$\hat{V}_{sub} = J \times \frac{1}{K} \sum_{l=1}^K \left( \frac{n}{J} \hat{\theta}_l^{short} - \frac{n}{m} \hat{\theta}_l^{long} \right)^2 \quad (10)$$

where  $K = \lfloor n/m \rfloor$ .  $\hat{\theta}_l^{short}$  and  $\hat{\theta}_l^{long}$  are realized variances calculated on the short subsample with  $J$  observations, and the long subsample with  $m$  observations. Figure 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

$$\begin{aligned} \hat{\theta}_l^{short} &= \sum_{i=1}^J \left( X_{\frac{(l-1)m+i}{n}} - X_{\frac{(l-1)m+i-1}{n}} \right)^2 \\ \hat{\theta}_l^{long} &= \sum_{i=1}^m \left( X_{\frac{(l-1)m+i}{n}} - X_{\frac{(l-1)m+i-1}{n}} \right)^2. \end{aligned}$$

For an arbitrary volatility process,  $nJ^{-1}\hat{\theta}_l^{short}$  and  $nm^{-1}\hat{\theta}_l^{long}$  cannot be guaranteed to be close. For example, if the volatility process has a large jump on the interval covered by  $\hat{\theta}_l^{long}$ , but not covered by  $\hat{\theta}_l^{short}$ , then  $nJ^{-1}\hat{\theta}_l^{short}$  and  $nm^{-1}\hat{\theta}_l^{long}$  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*

$$d\sigma_t = \tilde{\mu}_t dt + \tilde{\sigma}_t d\tilde{W}_t$$

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

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

$$\hat{V}_{sub} \xrightarrow{p} V.$$

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

This subsection presents an alternative subsampling scheme that is of theoretical interest. In general, it can be applied to cases when the asymptotic variance of an estimator on a sub-block of lower frequency observations has the same structure as the asymptotic variance of the estimator on the full sample. This scheme will not be used in further sections due to its inability to estimate the asymptotic variance of the Two Scale estimator in the presence of autocorrelated noise. We now describe it for the RV example.

Consider the following subsampling scheme. On every block of  $m$  observations, calculate the estimator  $\hat{\theta}^n$  twice as follows. First, calculate it using all  $m$  observations, denote it as  $\hat{\theta}_l^{fast}$ . Then, calculate the estimator  $\hat{\theta}^n$  using every  $Q^{th}$  price observation in the block of  $m$  observations, and denote it as  $\hat{\theta}_l^{slow}$ .

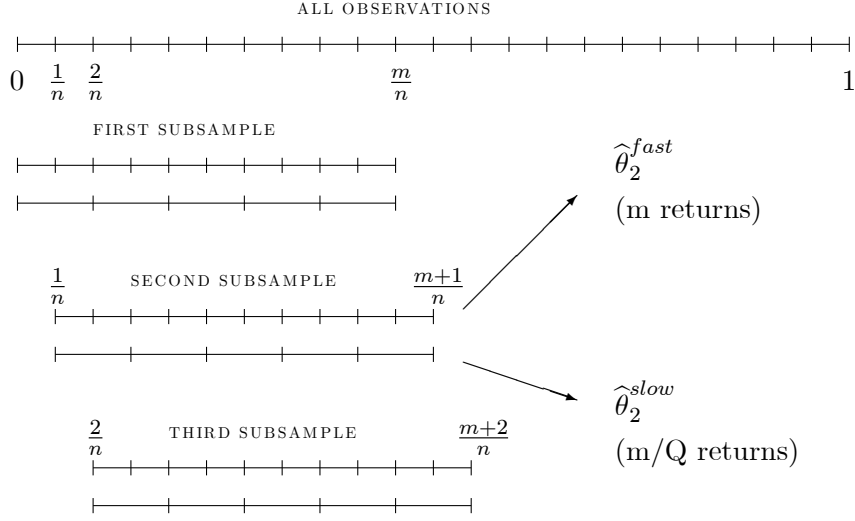


Figure 3: *An Alternative Subsampling Scheme.*

The corresponding expressions for RV calculated on these subsamples are

$$\begin{aligned}\hat{\theta}_k^{fast} &= \sum_{i=1}^m \left( X_{\frac{i+m(k-1)}{n}} - X_{\frac{i-1+m(k-1)}{n}} \right)^2 \\ \hat{\theta}_k^{slow} &= \sum_{i=1}^{\lfloor m/Q \rfloor} \left( X_{\frac{iQ+m(k-1)}{n}} - X_{\frac{(i-1)Q+m(k-1)}{n}} \right)^2.\end{aligned}$$

Now,  $\hat{\theta}_l^{fast}$  can be used to center the  $\hat{\theta}_l^{slow}$ , because they both converge to (9), and because  $\hat{\theta}_l^{fast}$  converges to (9) faster than  $\hat{\theta}_l^{slow}$  does. See Figure 3 for a graphical illustration. The estimator of  $V$  becomes

$$\hat{V}_a = \frac{m}{Q} \times \frac{1}{K} \sum_{l=1}^K \left( \frac{n}{m} \hat{\theta}_l^{slow} - \frac{n}{m} \hat{\theta}_l^{fast} \right)^2 = \frac{n}{Q} \sum_{l=1}^{n/m} \left( \hat{\theta}_l^{slow} - \hat{\theta}_l^{fast} \right)^2$$

This construction shows that lower frequency data can be used to achieve the same effect as taking a shorter block of observations. In our RV example, sparse observations still convey all the features of the model, so this subsampling scheme delivers a consistent estimator of  $V$ .

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

$$\hat{V}_a \xrightarrow{p} V. \quad (11)$$

**Remarks.** 1. *Brownian semimartingale model (4) assumes the paths of  $X$  are continuous. Instead, suppose now that  $X$  is a Brownian semimartingale with jumps. In other words, define  $X$  on some probability space*

$(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t \geq 0}, \mathbb{P})$  as the process  $dX_t = \mu_t dt + \sigma_t dW_t + dJ_t$ . The continuous part is as in (4) and  $J_t$  is some jumps process, see for example Aït-Sahalia and Jacod (2008). In that case, the asymptotic variance of RV contains jumps, and the subsampling estimator  $\widehat{V}_a$  only estimates consistently the continuous part of the  $V$ . In particular, suppose  $m \rightarrow \infty$ ,  $Q \rightarrow \infty$ ,  $m/n \rightarrow 0$ , and  $Q/m \rightarrow 0$  as  $n \rightarrow \infty$ . Then,

$$\widehat{V}_a \xrightarrow{p} 2 \int_0^1 \sigma_u^4 du + 4 \sum_{p: T_p \in [0,1]} Y_{T_p}^2 (\sqrt{\kappa_p} U_p \sigma_{T_{p-}} + \sqrt{1 - \kappa_p} U'_p \sigma_{T_p})^2$$

where  $\kappa_p, p = 1, 2, \dots$  are uniform random variables, independent from  $F$ ;  $U_p, U'_p, p = 1, 2, \dots$  are standard normal random variables independent from  $F$  and from  $\kappa_p, p = 1, 2, \dots$ ;  $T_p, p = 1, 2, \dots$  are jump times. This shows the inconsistency of  $\widehat{V}_a$  because the random variables  $U_p$  and  $U'_p$  do not appear in the expression of the asymptotic variance of RV. If  $X$  and  $\sigma$  do not jump together,  $\widehat{V}_a$  is unbiased, conditionally on  $F$  because the random variable  $(U_p + U'_p)^2$  has expectation one. This illustrates the fact that the subsampling method needs  $V$  to be continuous in time. This prevents a situation when there is some feature of  $V$  that is only represented by one subsample. See also discussion of Assumption A6(ii) in Section 4.

2. Suppose, as in Remark 1, that  $X$  is a Brownian semimartingale with jumps. In that case, integrated volatility can be estimated by Bipower Variation (see Barndorff-Nielsen and Shephard, 2007),

$$\widehat{\theta}^n = n^{\frac{r+l}{2}-1} \sum_{i=1}^n \left| X_{\frac{i}{n}} - X_{\frac{i-1}{n}} \right|^r \left| X_{\frac{i-1}{n}} - X_{\frac{i-2}{n}} \right|^l.$$

Then, the subsampling estimator of the asymptotic variance  $V$  of  $\widehat{\theta}^n$  is only consistent if  $V$  does not contain jumps. This happens if  $\max(r, l) < 1$ .

The subsampling scheme is similar in structure to the one in Lahiri, Kaiser, Cressie, and Hsu (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.

This alternative subsampling method can also be applied to noisy diffusion setting, as long as noise is independent in time. However, if noise is autocorrelated and this autocorrelation appears in the expression of  $V$ , this subsampling scheme will not be able to estimate  $V$ . This is due to sparsely sampled data not containing all the needed information about autocorrelations. Therefore,  $\widehat{\theta}_l^{slow}$  is not be able to replicate that part in  $V$ , which pertains to autocorrelations.

### 3 Inference for the Two Time Scale Estimator

This section shows how the new subsampling scheme can be applied to the Two Time Scale estimator of integrated variance proposed by Aït-Sahalia et al. (2006a). 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 estimators of Zhang et al. (2005) and Aït-Sahalia et al. (2006a), Realized Kernels of Barndorff-Nielsen et al. (2008a), and pre-averaging estimator of Jacod et al. (2007).

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  $\epsilon$ , see (13) below. In this model, observed log-returns display negative first order autocovariance,

$$\begin{aligned} & \text{Cov}(\Delta Y_{i/n}, \Delta Y_{(i-1)/n}) \\ &= \text{Cov}(\Delta X_{i/n} + \epsilon_{i/n} - \epsilon_{(i-1)/n}, \Delta X_{(i-1)/n} + \epsilon_{(i-1)/n} - \epsilon_{(i-2)/n}) \\ &= -\text{Var}(\epsilon_{(i-1)/n}). \end{aligned} \tag{12}$$

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  $IV_X$  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 minutes. 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, Mykland, and Aït-Sahalia (2005) were the first to introduce a consistent estimator of integrated variance of the efficient price  $IV_X$  within the additive measurement error model of Zhou (1996). Their model is as follows. Log-price  $X$  is a continuous Brownian semimartingale (4). Observations are contaminated by some additive measurement error, so there are discrete observations on the noisy log-price  $Y$  available where

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

The noise  $\epsilon_t$  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  $X_t$ ,

$$\hat{\theta}_n = [Y, Y]^{(G_1)} - \frac{\bar{n}_{G_1}}{n} [Y, Y]^{(1)}, \tag{14}$$

where, for any parameter  $b$ ,

$$\begin{aligned} [Y, Y]^{(b)} &= \frac{1}{b} \sum_{i=1}^{n-b} (Y_{(i+b)/n} - Y_{i/n})^2 \\ \bar{n}_b &= \frac{n-b+1}{b}. \end{aligned}$$

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 minutes. Zhang et al. (2005) call the above estimator the Two Scale Realized Variance (TSRV) estimator. They derive the following asymptotic distribution of the estimator,

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

where the asymptotic (conditional) variance takes the form

$$V = c \underbrace{\frac{4}{3} \int_0^1 \sigma_u^4 du}_{\text{signal}} + \underbrace{8c^{-2} \omega^4}_{\text{noise}}, \quad (15)$$

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.  $\text{Var}(\epsilon) = \omega^2$  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. (2008a) have proposed an estimator for  $\int \sigma_u^4 du$ , which is consistent in 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 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 2 for references). In other words, prices are more volatile in mornings and afternoons than at noon; spreads are also larger in mornings and afternoons. Figure 4(a) presents an estimate of heteroscedasticity function  $\omega^2(\cdot)$  for transaction prices of Microsoft stock, averaged over all days in year 2006. The diurnal variation is evident.

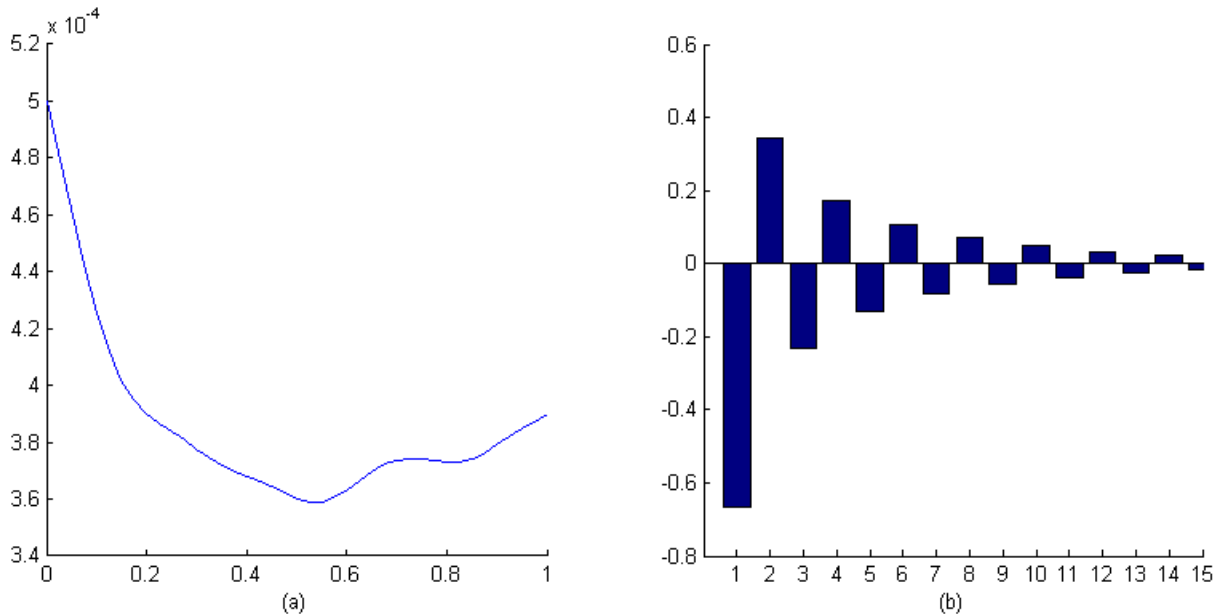


Figure 4: *Properties of returns of Microsoft (MSFT) stock. Returns are constructed from transaction prices over the whole year 2006. See Section 6 for data cleaning procedures.*

*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.*

Kalnina and Linton (2006) 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  $\epsilon_t$  satisfies

$$\epsilon_t = \omega(t) u_t \quad (16)$$

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

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

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

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

so, by Jensen's inequality, its square would be always strictly smaller than the target  $\int \omega^4(u) du$  as long as there is any diurnal variation at all. Kalnina and Linton (2006) show that  $\omega(\cdot)$  can be estimated at any fixed point  $\tau$  using kernel smoothing,

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

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

$$8c^{-2} \int_0^1 \tilde{\omega}^4(u) du.$$

As we saw earlier in (12), 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, Figure 4(b) graphs the autocorrelogram of the returns of the 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  $\epsilon_t$ .<sup>6</sup>

Aït-Sahalia et al. (2006a) 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  $\epsilon_t$  is independent from the efficient log-price  $X_t$ . Also, when viewed as a process in index  $i$ ,  $\epsilon_{t_i}$  is stationary and strong mixing with the mixing coefficients decaying exponentially*

In model (13) with  $\epsilon_t$  satisfying Assumption A2, Aït-Sahalia et al. (2006a) propose the following consistent estimator for the integrated variance of  $X_t$ ,

$$\hat{\theta}_n = [Y, Y]^{(G_1)} - \frac{\bar{n}_{G_1}}{\bar{n}_{G_2}} [Y, Y]^{(G_2)} \quad (17)$$

where  $G_1$  and  $G_2$  satisfy the following assumption,

**Assumption A3.** *The  $G_1$  parameter of the Two Time Scale estimator  $\hat{\theta}_n$  defined by (17) satisfies  $G_1 = \lfloor cn^{2/3} \rfloor$  for some constant  $c$ .  $G_2$  parameter is such that  $\text{Cov}(\epsilon_0, \epsilon_{G_2/n}) = o(n^{-1/2})$ ,  $G_2 \rightarrow \infty$ ,  $G_2/G_1 \rightarrow 0$ .<sup>7</sup>*

<sup>6</sup>In a Brownian semimartingale model, the only source of autocorrelations of increments is drift, which is negligible for high frequencies.

<sup>7</sup>The restriction on  $\text{Cov}(\epsilon_0, \epsilon_{G_2/n})$  should be considered in the light of the fact that Assumption A2 implies that there exists a constant  $\phi$  such that, for all  $i$ ,

$$|\text{Cov}(\epsilon_{i/n}, \epsilon_{(i+1)/n})| \leq \phi^l \text{Var}(\epsilon).$$

The Two Time Scale realized variance estimator defined by (17) is a more general than the one in (14), which is a special case when  $G_2 = 1$  and  $G_1 \rightarrow \infty$  as  $n \rightarrow \infty$ . Ait-Sahalia et al. (2006a) show that the new TSRV estimator  $\widehat{\theta}_n$  has the same asymptotic properties except it has a more complicated asymptotic variance,

$$V = \underbrace{c \frac{4}{3} \int_0^1 \sigma_u^4 du}_{\text{signal}} + \underbrace{8c^{-2} \text{Var}(\epsilon)^2 + 16c^{-2} \lim_{n \rightarrow \infty} \sum_{i=1}^n \text{Cov}(\epsilon_0, \epsilon_{i/n})^2}_{\text{noise}}, \quad (18)$$

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

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

**Theorem 4.** *Suppose model (13) holds, and  $\epsilon_{t_i}$  satisfy Assumption A2. Let  $\widehat{\theta}_n$  be the TSRV estimator defined by (17), with parameters  $G_1$  and  $G_2$  that satisfy Assumption A3. Let  $V$  be defined by (18). Assume  $\{\sigma\}$  and  $\{\mu\}$  are independent of  $\{W\}$ . Let  $J \rightarrow \infty$ ,  $m \rightarrow \infty$ ,  $J/m \rightarrow 0$ ,  $m/n \rightarrow 0$ ,  $G_1/J \rightarrow 0$  and  $Jmn^{-5/3} \rightarrow 0$  as  $n \rightarrow \infty$ . Then,*

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

where

$$\widehat{V}_{sub} = Jn^{-2/3} \times \frac{1}{K} \sum_{l=1}^K \left( \frac{n}{J} \widehat{\theta}_l^{short} - \frac{n}{m} \widehat{\theta}_l^{long} \right)^2 \quad (19)$$

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

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

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

where

$$\begin{aligned} [Y, Y]_l^{(G_i)} &= \frac{1}{G_i} \sum_{i=1}^{J-G_i} (Y_{(l-1)m/n+(i+G_i)/n} - Y_{(l-1)m/n+i/n})^2, \quad i = 1, 2 \\ \bar{J}_{G_i} &= \frac{J - G_i + 1}{G_i}, \quad i = 1, 2. \end{aligned}$$

One obtains  $\widehat{\theta}_l^{short}$  by substituting  $J$  for  $m$  above. In Figure 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,  $\widehat{V}_{sub}$  is defined by (19) 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 Scale estimator when the noise is autocorrelated. Similarly, one can apply this method to Multiscale Estimator of Aït-Sahalia et al. (2006a) when microstructure noise is autocorrelated. The advantage of using Multiscale estimator is that it has the optimal rate of convergence  $n^{1/4}$ .

However, the above model of Aït-Sahalia et al. (2006a) rules out any diurnal heteroscedasticity of the noise. When both autocorrelation and heteroscedasticity is 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  $\text{Var}(u_t) = 1$ . Then,  $\widehat{\theta}_n$  defined in (17) satisfies

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

where

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

In this case of autocorrelated and heteroscedastic noise, Theorem 4 easily generalizes and subsampling again delivers 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  $\widehat{V}_{sub}$  should be used as for homoscedastic case. In this model, this is the only available method in the literature to construct confidence intervals for the Two Time Scale estimator.

Importantly, this section illustrates 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. It seems that the subsampling estimator  $\widehat{V}_{sub}$  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.

## 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 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 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

$$\tau_n \left( \widehat{\theta}_n - \theta \right) \Rightarrow \sqrt{V} Z. \quad (20)$$

In the above,  $\tau_n$  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 TSRV estimator.  $Z$  is a random variable that is known to satisfy  $E(Z) = 0$  and  $\text{Var}(Z) = 1$ . 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  $\widehat{\theta}_l^{long}$  the estimator  $\widehat{\theta}_n$  calculated using all  $m$  returns of the  $l^{th}$  block,  $l = 1, \dots, \lfloor n/m \rfloor$ . Denote by  $\widehat{\theta}_l^{short}$  the estimator  $\widehat{\theta}_n$  calculated using only  $J$  returns of the  $l^{th}$  block, where  $J < m$ . See Figure 2 in Section 2.2 for a graphical illustration.

In order to guarantee that  $\frac{n}{J} \widehat{\theta}_l^{short}$  and  $\frac{n}{m} \widehat{\theta}_l^{long}$  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.** (20) holds, where  $\theta$  and  $V$  are the following functions of the volatility path  $\{\sigma_u, u \in [0, 1]\}$ ,

$$\begin{aligned} \theta &= \int_0^1 g_1(\sigma(u)) du \\ V &= \int_0^1 g_2(\sigma(u)) du \end{aligned}$$

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

For example, we obtain integrated variance  $IV_X$  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 assumptions of this section are those that are *approximately* additive over subsamples, i.e.,

$$\widehat{\theta}_n = \sum_{l=1}^{\lfloor n/m \rfloor} \frac{m}{J} \widehat{\theta}_l^{short} + o_p(1) \quad (21)$$

or

$$\widehat{\theta}_n = \sum_{l=1}^{\lfloor n/m \rfloor} \widehat{\theta}_l^{long} + o_p(1). \quad (22)$$

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  $\{R_{i/n}^{(n)}\}_{i=1,\dots,n}$  with  $R_{i/n}^{(n)} = X_{i/n} - X_{(i-1)/n}$  is strong mixing. Also,  $\widehat{\theta}_n = \phi(R_{1/n}^{(n)}, R_{2/n}^{(n)}, \dots, R_{1/n}^{(n)})$  where  $\phi: \mathbb{R}^n \mapsto \mathbb{R}$ .

For example, suppose  $X$  is a Brownian semimartingale as in (4). Then, Assumption A5 is satisfied if we assume  $\{\mu\}, \{\sigma\} \perp \{W\}$  and consider all arguments conditional on  $\{\mu_t, \sigma_t, t \in [0, 1]\}$ . This means that the conclusion about consistency of  $\widehat{V}_{sub}$  of Theorem 6 below holds, conditionally on  $\{\mu_t, \sigma_t, t \in [0, 1]\}$ . Hence, the same conclusion also holds unconditionally.

As discussed in previous sections,  $\widehat{\theta}_l^{long}$  and  $\widehat{\theta}_l^{short}$  do not estimate  $\theta$ , 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  $\theta_l^{long}$  and  $\theta_l^{short}$  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,

$$\begin{aligned} \theta_l^{short} &= \int_{(l-1)m/n}^{[(l-1)m+J]/n} g_1(\sigma(u)) du, & V_l^{short} &= \int_{(l-1)m/n}^{[(l-1)m+J]/n} g_2(\sigma(u)) du \\ \theta_l^{long} &= \int_{(l-1)m/n}^{lm/n} g_1(\sigma(u)) du, & V_l^{short} &= \int_{(l-1)m/n}^{[(l-1)m+J]/n} g_2(\sigma(u)) du. \end{aligned} \quad (23)$$

Finally, we make the following assumption,

**Assumption A6.** For every  $n$ , define  $\theta_l^{short}$  and  $V_l^{short}$  by (23), and define a triangular array

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

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

(i)

$$\text{as } n \rightarrow \infty, \sup_l \mathbb{E} \left( \zeta_l^{(n)} \right) \rightarrow 0,$$

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

We now discuss Assumption (A6). The appendix contains verification of Assumption (A6) for the TSRV estimator, as this is how Theorem 4 is proved.

Assumption A6 (i) can be written equivalently as follows,

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

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  $\widehat{\theta}_l^{short}$  has asymptotic expectation one. On the full sample, we know from (20) that standardized  $\widehat{\theta}^n$  is asymptotically a random variable  $Z$  with  $\mathbb{E}(Z^2) = 1$ . Therefore, a sufficient condition for Assumption A6 (i) to hold is that the asymptotic distribution of  $\widehat{\theta}_l^{short}$  satisfies the same condition on a subsample. Roughly speaking, we need the estimator on a subsample,  $\widehat{\theta}_l^{short}$ , 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  $\epsilon$ . In this model, corresponding moments have to be assumed on  $\epsilon$  for assumption A6 (ii) to hold. In the case of the TSRV estimator discussed below,  $L^{4+\varepsilon}$  boundedness of  $\epsilon$  is needed, which is exactly what has been assumed by the authors of TSRV estimator to derive its asymptotic distribution.

We have the following result.

**Theorem 6.** *Assume (A4), (A5), and (A6). Then,*

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

where

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

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  $\tau_n$ . In particular,  $\widehat{V}_{sub}$  simplifies to formula for the realized variance in (10) with  $\tau_n = \sqrt{n}$ , and to the formula for the Two Time Scale estimator in (19) with  $\tau_n = n^{1/6}$ .

## 5 Simulation Study

In this section numerical properties of the proposed estimator are studied for the example of TSRV estimator of Ait-Sahalia et al. (2006a) in the case of i.i.d. or autocorrelated microstructure noise.

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

$$\begin{aligned} dX_t &= (\alpha_1 - v_t/2) dt + \sigma_t dW_t \\ dv_t &= \alpha_2 (\alpha_3 - v_t) dt + \alpha_4 v_t^{1/2} dB_t \end{aligned}$$

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  $\alpha_4 = 0.5$  (unit of time is one year). We simulate 35,000 observations over one week, i.e., five business days of 6.5 hours each. This is motivated by the fact that GE stock has on average 35,000 observations per week in year 2006, see Section 6.

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

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

Four different values of  $\rho$  are considered,  $\rho = 0$ ,  $-0.3$ ,  $-0.5$ , and  $\rho = -0.7$ .

Size of the noise is an important parameter. Denote  $\text{Var}(u) = \omega^2$ . Denote the noise-to-signal ratio by

$$\xi^2 = \frac{\omega^2}{\sqrt{\int_0^1 \sigma_u^4 du}}.$$

Results are simulated for two different noise-to-signal ratios, which are suggested by Barndorff-Nielsen et al. (2008a) and are  $\xi^2 = 0.001$  and  $0.0001$ . These are motivated by the careful empirical study of Hansen and Lunde (2006), who investigate 30 stocks of Dow Jones Industrial Average. The volatility path is fixed over simulations to facilitate comparisons. The volatility path used is plotted in Figure 5. Varying volatility path across simulations does not affect the theory nor simulation results.

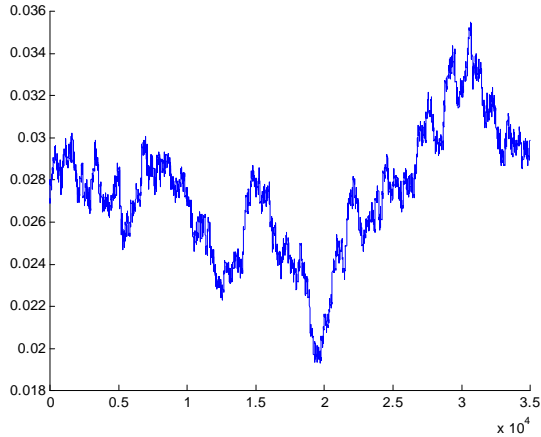


Figure 5: *Simulated volatility sample path.*

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

The literature does not propose ways of estimating asymptotic variance of TSRV when noise is auto-correlated or diurnal. However, in the case of i.i.d. noise, there is an alternative, and this will serve as a benchmark for the simulation results. In the case of i.i.d noise, the expression for asymptotic variance  $V$  of TSRV estimator is

$$V = c \frac{4}{3} \int_0^1 \sigma_u^4 du + 8c^{-2} [\text{Var}(u)]^2$$

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  $IQ$  (the first term in  $V$ ) in the presence of microstructure noise is more difficult. A consistent estimator in the presence of i.i.d. noise has been proposed by Barndorff-Nielsen et al. (2008a)

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

where

$$\begin{aligned} y_{j,\cdot}^2 &= \frac{1}{S} \sum_{s=0}^{S-1} \left( Y_{\tilde{\delta}(j+\frac{s}{S})} - Y_{\tilde{\delta}(j-1+\frac{s}{S})} \right)^2, \quad j = 1, \dots, \tilde{n} \\ \widetilde{\omega}^2 &= \exp \left\{ \log \left( \widehat{\omega}^2 \right) - \widehat{\theta}_n^*/RV \right\} \\ \tilde{n} &= \left\lfloor 1/\tilde{\delta} \right\rfloor \end{aligned}$$

and where  $\widehat{\theta}_n^*$  is a consistent estimator of integrated variance  $IV_X$ . We take  $\widehat{\theta}_n^*$  to be the Two Time Scale estimator  $\widehat{\theta}_n$ . This estimator requires to choose  $\tilde{\delta}$  and  $S$ . We use the same choice as Barndorff-Nielsen et al. (2008a) do, for real and simulated data. This choice is  $S = n^{1/2}$  and  $\tilde{\delta} = n^{-1/2}$ . Estimator  $\widetilde{\omega}^2$  corrects 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. (2008a) anyway. Thus, the following alternative estimator of  $V$  is constructed

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

This estimator is consistent for  $V$  in the presence of i.i.d. noise.

When noise is autocorrelated, the estimator  $\widehat{IQ}_{BNHLS}$  is inconsistent, but the sign of the bias depends on the exact parameters of the model. We now give heuristic explanation about what behavior can be expected of  $\widehat{IQ}_{BNHLS}$  in the presence of autocorrelated noise. To simplify the exposition, notice that  $y_{j,\cdot}^2$  can be thought of as simply one low frequency return squared. This is because  $y_{j,\cdot}^2$  is an average over returns that are very highly correlated given large overlaps in time they have. Also, use  $\widetilde{\omega}^2$  instead of  $\widehat{\omega}^2$ . Consider the building block of  $\widehat{IQ}_{BNHLS}$ , for  $j = 1$ ,

$$y_{1,\cdot}^2 - 2\widetilde{\omega}^2 \approx \left( Y_{\frac{1}{n} + \frac{1}{\sqrt{n}}} - Y_{\frac{1}{n}} \right)^2 - \frac{1}{n} \sum_{i=1}^n \left( Y_{\frac{i}{n}} - Y_{\frac{i-1}{n}} \right)^2.$$

Suppose  $\epsilon \perp X$  and that low frequency noise is uncorrelated. Then,

$$\mathbb{E} \left( y_{1,\cdot}^2 - 2\widetilde{\omega}^2 \right) \approx \mathbb{E} \left( X_{\frac{1}{n} + \frac{1}{\sqrt{n}}} - X_{\frac{1}{n}} \right)^2 - \frac{1}{n} \sum_{i=1}^n \mathbb{E} \left( Y_{\frac{i}{n}} - Y_{\frac{i-1}{n}} \right)^2 + 2\mathbb{E} \left( \epsilon_{\frac{i}{n}} \epsilon_{\frac{i-1}{n}} \right).$$

The middle term is of smallest order and so can be ignored. Notice that last term is negative in practice, so  $\mathbb{E} \left( y_{1,\cdot}^2 - 2\widetilde{\omega}^2 \right)$  becomes smaller. If noise is not too large,  $y_{1,\cdot}^2 - 2\widetilde{\omega}^2$  is biased towards zero. Thus, the final estimator of  $IQ$  has a negative bias.

Results are represented in terms of coverage probabilities of 95% two-sided, left-sided, and right-sided confidence intervals for  $IV_X$ . Table 1 contains the larger noise-to-signal case  $\xi^2 = 0.001$ , and Table 2 contains results for the smaller noise-to-signal case  $\xi^2 = 0.0001$ . We see that the subsampling estimator performs well in all scenarios.  $\widehat{V}_a$  performs well in the scenario it is designed for, which is the uncorrelated noise

case. As the correlation increases, estimated values of  $\widehat{V}_a$  decrease, resulting in undercoverage. For the  $\xi^2 = 0.001$  scenario, two sided coverage probabilities of  $\widehat{V}_a$  decrease from 0.93 to 0.78 as autocorrelation becomes stronger. For the smaller noise scenario with  $\xi^2 = 0.0001$ , two sided coverage probabilities decrease from 0.93 to 0.88. This improvement as noise becomes smaller is to be expected, given that  $\widehat{V}_a$  is consistent for  $V$  when noise is zero.

## 6 Empirical Analysis

This section applies the proposed subsampling method to tick data of AIG, GE, IBM, INTC, MMM, and MSFT stocks obtained from the NYSE TAQ database, and compares it to the estimator  $\widehat{V}_a$ , which is introduced in the previous section.

We use the whole year of 2006 of transaction prices for AIG, GE, IBM, INTC, MMM, and MSFT stocks obtained from the NYSE TAQ database. Zero returns are removed, as in Aït-Sahalia et al. (2006a). Griffin and Oomen (2008) show that, in Realized Volatility case, this adjustment of data improves precision of estimation. Jumps are also removed,<sup>8</sup> since the additive market microstructure noise model (13) does not allow for jumps. There is also an additional issue to consider, which Barndorff-Nielsen et al. (2008b) 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. Such episodes occur rarely, but are nonetheless important. Barndorff-Nielsen et al. (2008b) 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 also is 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 (one minute) frequency data, and remove those lower frequency returns that are larger than seven weekly standard deviations. Such "gradual" jumps occur rarely, and most weeks do not contain any.

In the resulting data set, the average number of trades per week is 20,341 for AIG, 35,361 for GE, 23,657 for IBM, 51,092 for INTC, 15,642 for MMM, and 45,646 for MSFT. The returns of all these stocks display large negative autocorrelation similar to GE in Figure 4(b).

The asymptotic variance of the Two Time Scale estimator is estimated for each of the 52 weeks in year 2006. As long as the distance between observations is of order  $1/n$ , the underlying theory can be extended

---

<sup>8</sup>Jumps are identified as deviations of the log-returns that are larger than five standard deviations on a moving window. 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 (2007).

to the non equidistant observations case. Therefore, the estimation is done in tick time, as suggested in Barndorff-Nielsen et al. (2008a) and other authors.

The results are displayed in Figures 6 - 11 in the appendix, in terms of 95% confidence intervals for integrated variance. Confidence intervals with bars correspond to subsampling method and confidence intervals with lines correspond to the alternative method  $\widehat{V}_a$ . The TSRV estimate  $\widehat{\theta}_n$  is in the center of both confidence intervals by construction. The subsampling confidence intervals for TSRV are usually wider than confidence intervals of the alternative method  $\widehat{V}_a$ . From our simulations, we conclude this might be due to negative bias of the  $\widehat{V}_a$  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  $\widehat{V}_a$  is downward biased in this case. On the other hand, 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 TSRV 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 the first automated method for estimating the asymptotic variance of an estimator in high frequency data. The method applies to an important general class of estimators, which include 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 case is the integrated variance estimator of Aït-Sahalia et al. (2006a), 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 Scale 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.

## References

- [1] AÏT-SAHALIA, Y. AND J. JACOD (2008). Estimating the Degree of Activity of Jumps in High Frequency Data. *Annals of Statistics*, forthcoming.
- [2] AÏT-SAHALIA, Y., P. MYKLAND, AND L. ZHANG (2005). How Often to Sample a Continuous-Time Process in the Presence of Market Microstructure Noise. *Review of Financial Studies*, 18, 351-416.
- [3] AÏT-SAHALIA, Y., P. MYKLAND, AND L. ZHANG (2006a). Ultra High Frequency Volatility Estimation with Dependent Microstructure Noise. Working Paper.
- [4] AÏT-SAHALIA, Y., P. MYKLAND, AND L. ZHANG (2006b). Comments on ‘Realized Variance and Market Microstructure Noise,’ by P. Hansen and A. Lunde, *Journal of Business and Economic Statistics*.
- [5] AÏT-SAHALIA, Y., ZHANG, L. , AND P. MYKLAND (2005). Edgeworth Expansions for Realized Volatility and Related Estimators. Working paper, Princeton University.
- [6] ANDERSEN, T. G. AND T. BOLLERSLEV (1997). Intraday periodicity and volatility persistence in financial markets. *Journal of Empirical Finance* 5, 115 - 158.
- [7] ANDERSEN, T. G., AND T. BOLLERSLEV (1998). Answering the Skeptics: Yes, Standard Volatility Models Do Provide Accurate Forecasts. *International Economic Review*. Vol 39, No. 4, 885 - 905.
- [8] ANDERSEN, T. G., T. BOLLERSLEV, F. X. DIEBOLD, AND P. LABYS (2001). The distribution of exchange rate volatility. *Journal of the American Statistical Association* 96, 42 - 55. Correction published in 2003, volume 98, page 501.
- [9] ANDREOU, E. AND E. GHYSELS (2002). Rolling-Sample Volatility Estimators: Some New Theoretical, Simulations and Empirical Results. *Journal of Business and Economic Statistics* 20, 363-375
- [10] ANDREWS, D. W. K. (1998). Laws of Large Numbers for Dependent Non-Identically Distributed Random Variables. *Econometric Theory*, Vol. 4, No. 3., 458 - 467
- [11] AWARTANI, B., CORRADI, V., AND W. DISTASO (2004). Testing and Modelling Market Microstructure Effects with and Application to the Dow Jones Industrial Average. Working paper.
- [12] BANDI, F. M. AND J. R. RUSSELL (2006). Separating microstructure noise from volatility. *Journal of Financial Economics* 79, 655-692.
- [13] BANDI, F. M. AND J. R. RUSSELL (2008). Microstructure noise, realized variance, and optimal sampling. *Review of Economic Studies* 75, 339–369.

- [14] BARNDORFF-NIELSEN, O. E., S. E. GRAVERSEN, J. JACOD, AND N. SHEPHARD (2006). Limit theorems for bipower variation in financial econometrics. *Econometric Theory*, 22, 677-719.
- [15] BARNDORFF-NIELSEN, O. E., HANSEN, P. R., LUNDE, A., AND N. SHEPHARD (2006). Subsampling Realized Kernels. *Journal of Econometrics*, forthcoming.
- [16] BARNDORFF-NIELSEN, O. E., HANSEN, P. R., LUNDE, A., AND N. SHEPHARD (2008a). Designing Realised Kernels to Measure the Ex-post Variation of Equity Prices in the Presence of Noise. *Econometrica*, Vol. 76, No. 6, 1481-1536
- [17] BARNDORFF-NIELSEN, O. E., HANSEN, P. R., LUNDE, A., AND N. SHEPHARD (2008b). Realised Kernels in Practice: Trades and Quotes. Working Paper.
- [18] BARNDORFF-NIELSEN, O. E. AND SHEPHARD, N. (2002). Econometric analysis of realised volatility and its use in estimating stochastic volatility models. *Journal of the Royal Statistical Society B* 64, 253–280.
- [19] BARNDORFF-NIELSEN, O. E. AND SHEPHARD, N. (2004). Power and Bipower Variation with Stochastic volatility and Jumps. *Journal of Financial Econometrics*, Vol. 2, No. 1, 1-37
- [20] BARNDORFF-NIELSEN, O. E. AND SHEPHARD, N. (2007). Variation, jumps, market frictions and high frequency data in financial econometrics. *Advances in Economics and Econometrics. Theory and Applications*, Ninth World Congress, (edited by R. Blundell, P. Torsten, and W. K. Newey), *Econometric Society Monographs*, Cambridge University Press.
- [21] CHRISTENSEN, K., R. C. A. OOMEN, AND M. PODOLSKIJ (2008). Realised Quantile-Based Estimation of the Integrated Variance. Available at SSRN: <http://ssrn.com/abstract=1085553>
- [22] CHRISTENSEN, K., M. PODOLSKIJ, AND M. VETTER (2006). Bias-Correcting the Realized Range-Based Variance in the Presence of Market Microstructure Noise. Available at SSRN: <http://ssrn.com/abstract=935419>
- [23] DELBAEN, F. AND W. SCHACHERMAYER (1994). A general version of the fundamental theorem of asset pricing. *Math. Ann.*, 300(3): 463-520.
- [24] FOSTER, D. P. AND D. B. NELSON (1996). Continuous Record Asymptotics for Rolling Sample Variance Estimators. *Econometrica* 64, 139-174.
- [25] GERETY, M. S., AND H. MULHERIN (1994). Price formation on stock exchanges: The evolution of trading within the day, *Review of Financial Studies* 6, 23-56.

- [26] GONÇALVES, S. AND N. MEDDAHI (2008). Bootstrapping Realized Volatility. Forthcoming in *Econometrica*.
- [27] GRIFFIN, J. AND R. OOMEN (2008). Sampling Returns for Realized Variance Calculations: Tick Time or Transaction Time? *Econometric Reviews*, Vol. 27, 230-253.
- [28] HANSEN, P. R., AND A. LUNDE (2006). Realized variance and market microstructure noise (with comments and rejoinder). *Journal of Business and Economic Statistics* 24, 127–218.
- [29] HARRIS, L. (1986). A transaction data study of weekly and intradaily patterns in stock returns, *Journal of Financial Economics*, Vol. 16, Issue 1, 99-117
- [30] 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.
- [31] JACOD, J. (2006). Asymptotic properties of realized power variations and related functionals of semimartingales. Unpublished paper
- [32] JACOD, J. (2007). Statistics and high frequency data. Lecture notes. Séminaire Européen de Statistique 2007.
- [33] JACOD, J., Y. LI, P. A. MYKLAND, M. PODOLSKIJ, AND M. VETTER (2007). Microstructure Noise in the Continuous Case: The Pre-Averaging Approach. Working Paper.
- [34] JACOD, J. AND A. N. SHIRYAEV (2003). *Limit Theorems for Stochastic Processes*. Springer.
- [35] KALNINA, I. AND O. B. LINTON (2006). Estimating Quadratic Variation Consistently in the Presence of Correlated Measurement Error. Forthcoming in *Journal of Econometrics*.
- [36] KARATZAS, I. AND S. E. SHREVE (2005). *Brownian Motion and Stochastic Calculus*, New York: Springer-Verlag.
- [37] KRISTENSEN, D. (2008). Nonparametric Filtering of the Realised Spot Volatility: A Kernel-Based Approach. Forthcoming in *Econometric Theory*.
- [38] LAHIRI, S. N., M. S. KAISER, N. CRESSIE, AND N. HSU (1999). Prediction of Spatial Cumulative Distribution Functions Using Subsampling. *Journal of the American Statistical Association*, 94, 86–97
- [39] LOCKWOOD, L. J., AND S. C. LINN (1990). An examination of stock market return volatility during overnight and intraday periods, 1964–1989. *Journal of Finance* 45, 591-601.

- [40] MANCINI, C. (2004). Estimation of the characteristics of the jumps of a general Poisson-diffusion model, *Scandinavian Actuarial Journal*, 2004, 1: 42-52
- [41] MCINISH, T. H., AND R. A. WOOD (1992). An analysis of intraday patterns in bid/ask spreads for NYSE stocks. *Journal of Finance* 47, 753–764.
- [42] MIKOSCH, T. AND C. STARICA (2003). Stock Market Risk-Return Inference. An Unconditional, Non-Parametric Approach. Available at SSRN: <http://ssrn.com/abstract=882820>
- [43] MYKLAND, P. AND L. ZHANG (2007). Locally parametric inference in high frequency data. University of Chicago.
- [44] PODOLSKIJ, M. (2006). New Theory on Estimation of Integrated Volatility with Applications. PhD Thesis, Bochum University
- [45] PODOLSKIJ, M. AND M. VETTER (2006). Estimation of Volatility Functionals in the Simultaneous Presence of Microstructure Noise and Jumps. Working Paper.
- [46] PODOLSKIJ, M. AND D. ZIGGEL (2007). Bootstrapping Bipower Variation. Technical Report. Ruhr-University of Bochum.
- [47] POLITIS, D. N. AND J. P. ROMANO (1994), “Large sample confidence regions based on subsamples under minimal assumptions.” *Annals of Statistics* 22, 2031-2050.
- [48] POLITIS, D. N., J. P. ROMANO AND M. WOLF (1999). *Subsampling*, Springer-Verlag, New York.
- [49] REVUZ, D., AND M. YOR (2005). *Continuous Martingales and Brownian Motion*. Springer-Verlag, New York.
- [50] ZHANG, L. (2006). Efficient estimation of stochastic volatility using noisy observations: a multi-scale approach. *Bernoulli*, 12 (6), 1019-1043.
- [51] ZHANG, L., P. MYKLAND, AND Y. AÏT-SAHALIA (2005). A tale of two time scales: determining integrated volatility with noisy high-frequency data. *Journal of the American Statistical Association*, 100, 1394–1411
- [52] ZHOU, B. (1996). High-frequency data and volatility in foreign-exchange rates. *Journal of Business and Economic Statistics* 14, 45–52.

## A Proofs

Since  $\{\sigma_t\}$ ,  $\{\tilde{\sigma}_t\}$ ,  $\{\mu_t\}$  and  $\{\tilde{\mu}_t\}$  are locally bounded, it can be assumed, without loss of generality, that they are uniformly bounded by  $C_\sigma$  (see Barndorff-Nielsen, Graversen, Jacod, and Shephard (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),

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

for some constant  $C$ . Hence,

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

and

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

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

## 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  $q > 0$

$$\mathbb{E}(|\sigma_{t+s} - \sigma_t|^q | \mathcal{F}_t) \leq Cs^{q/2}. \quad (24)$$

This holds because

$$\begin{aligned} \mathbb{E}(|\sigma_{t+s} - \sigma_t|^q | \mathcal{F}_t) &= \mathbb{E} \left( \left| \int_t^{s+t} \tilde{\mu}_u du + \int_t^{s+t} \tilde{\sigma}_u d\tilde{W}_u \right|^q \middle| \mathcal{F}_t \right) \\ &\leq \mathbb{E} \left( \left| \int_t^{s+t} \tilde{\mu}_u du \right|^q \middle| \mathcal{F}_t \right) + \mathbb{E} \left( \left| \int_t^{s+t} \tilde{\sigma}_u d\tilde{W}_u \right|^q \middle| \mathcal{F}_t \right) \\ &\leq Cs^q + CE \left( \left| \int_t^{s+t} \tilde{\sigma}_u^2 du \right|^{q/2} \middle| \mathcal{F}_t \right) \\ &\leq Cs^{q/2} \end{aligned}$$

where 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  $q > 1$ ,

$$\mathbb{E} \left[ |\mathcal{X}_{k,i}|^q \middle| \mathcal{F}_{\frac{(k-1)m+i-1}{n}} \right] \leq C \left( \frac{1}{n} \right)^{1 \wedge q/2} \quad (25)$$

where

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

Introduce the following notation,

$$\begin{aligned} \widehat{V}_{sub}^{DISCR} &= \frac{1}{K} \sum_{k=1}^K 2\sigma_{\frac{k-1}{K}}^4 & \mathcal{E} \left( \widehat{V} \right)^{DISCR} &= \frac{J}{K} \sum_{k=1}^K \mathbb{E} \left[ \gamma_k^{DISCR} \middle| \mathcal{F}_{\frac{k-1}{K}} \right] \\ \mathcal{E} \left( \widehat{V} \right) &= \frac{J}{K} \sum_{k=1}^K \mathbb{E} \left[ \gamma_k \middle| \mathcal{F}_{\frac{k-1}{K}} \right] & \widehat{\alpha}_k^{short} &= \frac{n}{J} \sum_{i=1}^J \sigma_{\frac{m(k-1)}{n}}^2 \left( W_{\frac{(k-1)m+i}{n}} - W_{\frac{(k-1)m+i-1}{n}} \right)^2 \\ \gamma_k &= \left( \widehat{\theta}_k^{short} - \widehat{\theta}_k^{long} \right)^2 & \widehat{\alpha}_k^{long} &= \frac{n}{m} \sum_{i=1}^m \sigma_{\frac{m(k-1)}{n}}^2 \left( W_{\frac{(k-1)m+i}{n}} - W_{\frac{(k-1)m+i-1}{n}} \right)^2 \\ \gamma_k^{DISCR} &= \left( \widehat{\alpha}_k^{short} - \widehat{\alpha}_k^{long} \right)^2 \end{aligned}$$

We want to show

$$\widehat{V}_{sub} = \frac{1}{K} \sum_{k=1}^K \left( \widehat{\theta}_k^{slow} - \widehat{\theta}_k^{fast} \right)^2 \xrightarrow{p} V = 2 \int_0^1 \sigma_u^4 du.$$

First, by Riemann integrability of  $\sigma$ ,

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

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

The first step is to show

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

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

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

We have

$$\begin{aligned} & \mathbb{E} \left[ \left| \frac{J}{K} \gamma_k \right|^2 \middle| \mathcal{F}_{\frac{k-1}{K}} \right] \\ &= \frac{J^2}{K^2} \mathbb{E} \left[ \left\{ \frac{n}{J} \sum_{i=1}^J \left( X_{\frac{(k-1)m+i}{n}} - X_{\frac{(k-1)m+i-1}{n}} \right) \right)^2 - \frac{n}{m} \sum_{i=1}^m \left( X_{\frac{(k-1)m+i}{n}} - X_{\frac{(k-1)m+i-1}{n}} \right) \right\}^4 \middle| \mathcal{F}_{\frac{k-1}{K}} \right] \\ &\leq C \frac{J^2}{K^2} = \frac{J^2 m^2}{n^2} \end{aligned}$$

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

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

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

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

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

Second step is to show

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

We have

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

Let  $c_i = n/m - n/J$  for  $i = 1, \dots, J$ .  $c_i = n/m$  for  $i = J+1, \dots, m$ . Second part is square root of

$$\begin{aligned}
& \mathbb{E}_k B^2 \\
&= \mathbb{E}^k \left[ \frac{n}{m} \sum_{i=1}^m \left( \sigma_{\frac{m(k-1)}{n}}^2 \left( \Delta W_{\frac{(k-1)m+i}{n}} \right)^2 - \left( \Delta X_{\frac{(k-1)m+i}{n}} \right)^2 \right) - \frac{n}{J} \sum_{i=1}^J \left( \sigma_{\frac{m(k-1)}{n}}^2 \left( \Delta W_{\frac{(k-1)m+i}{n}} \right)^2 - \left( \Delta X_{\frac{(k-1)m+i}{n}} \right)^2 \right) \right]^2 \\
&= \mathbb{E}^k \left[ \left\{ \sum_{i=1}^m c_i \left[ \sigma_{\frac{m(k-1)}{n}}^2 \left( \Delta W_{\frac{(k-1)m+i}{n}} \right)^2 - \left( \Delta X_{\frac{(k-1)m+i}{n}} \right)^2 \right] \right\}^2 \right] \\
&= \sum_{i=1}^m c_i^2 \mathbb{E}^k \left[ \sigma_{\frac{m(k-1)}{n}}^2 \left( \Delta W_{\frac{(k-1)m+i}{n}} \right)^2 - \left( \Delta X_{\frac{(k-1)m+i}{n}} \right)^2 \right]^2 \\
&= \sum_{i=1}^m \sum_{i'=1}^m c_i c_{i'} \mathbb{E}^k \left[ \sigma_{\frac{m(k-1)}{n}}^2 \left( \Delta W_{\frac{(k-1)m+i}{n}} \right)^2 - \left( \Delta X_{\frac{(k-1)m+i}{n}} \right)^2 \right] \left[ \sigma_{\frac{m(k-1)}{n}}^2 \left( \Delta W_{\frac{(k-1)m+i'}{n}} \right)^2 - \left( \Delta X_{\frac{(k-1)m+i'}{n}} \right)^2 \right] \\
&\leq C n^{-5/2} \sum_{i=1}^m c_i^2 + C n^{-3} \sum_{i=1}^m \sum_{i'=1}^m |c_i| |c_{i'}| \\
&\leq C n^{-5/2} \frac{n^2}{J} + C n^{-3} n^2 \\
&= C n^{-1/2} J^{-1} + C n^{-1}
\end{aligned}$$

because

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

and, for  $i < i'$ ,

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

First part is square root of

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

Combining both  $A$  and  $B$  terms, we obtain

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

from which second step

$$\left| \mathcal{E} \left( \widehat{V} \right)^{DISCR} - \mathcal{E} \left( \widehat{V} \right) \right| \leq \frac{J}{K} \sum_{k=1}^K \mathbf{E} \left[ |\gamma_k^{DISCR} - \gamma_k| \middle| \mathcal{F}_{\frac{k-1}{K}} \right] \leq C J n^{-1/4} J^{-1/2} + C J n^{-1/2} \xrightarrow{p} 0$$

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

Now we prove the third step.

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

Thus,

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

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

### 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.

$$\begin{aligned}
V^{DISCR} &= \frac{m}{n} \sum_{k=1}^K 2\sigma_{\frac{k-1}{K}}^4 & \mathcal{E}(\widehat{V})^{DISCR} &= \frac{m}{n} \sum_{k=1}^K \left[ \mathbb{E} \gamma_k^{DISCR} \middle| \mathcal{F}_{\frac{k-1}{K}} \right] \\
\mathcal{E}(\widehat{V}) &= \frac{m}{n} \sum_{k=1}^K \mathbb{E} \left[ \gamma_k \middle| \mathcal{F}_{\frac{k-1}{K}} \right] & \widehat{\alpha}_k^{slow} &= \sigma_{\frac{m(k-1)}{n}}^2 \left( \Delta_{\frac{m}{n}} W_{\frac{mk}{n}} \right)^2 \\
\gamma_k &= \left( \widehat{\theta}_k^{slow} - \widehat{\theta}_k^{fast} \right)^2 & \widehat{\alpha}_k^{fast} &= \sigma_{\frac{m(k-1)}{n}}^2 \sum_{i=1}^m \left( \Delta_{\frac{1}{n}} W_{\frac{i+m(k-1)}{n}} \right)^2 \\
\gamma_k^{DISCR} &= \left( \widehat{\alpha}_k^{slow} - \widehat{\alpha}_k^{fast} \right)^2
\end{aligned}$$

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

$$\widehat{V}_a \xrightarrow{p} V = 2 \int_0^1 \sigma_u^4 du$$

where

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

First, by Riemann integrability,

$$V^{DISCR} \xrightarrow{p} V = 2 \int_0^1 \sigma_u^4 du. \quad (26)$$

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

The first step is to show

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

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

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

Notice that, by Burkholder-Davis-Gundy inequality, Cauchy-Schwarz inequality, and uniform boundedness of  $\sigma$ ,

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

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

$$\begin{aligned}
\mathbb{E}_{k-1}^n \left[ \left( \widehat{\theta}_k^{fast} \right)^4 \right] & \leq \frac{C}{K^4}, \quad \mathbb{E}_{k-1}^n \left[ \left( \widehat{\theta}_k^{fast} \right)^3 \right] \leq \frac{C}{K^4}, \quad \mathbb{E}_{k-1}^n \left[ \left( \widehat{\theta}_k^{fast} \right)^2 \right] \leq \frac{C}{K^2}, \\
\mathbb{E}_{k-1}^n \left[ \left( \widehat{\theta}_k^{slow} \right)^4 \right] & \leq \frac{C}{K^4}, \quad \mathbb{E}_{k-1}^n \left[ \left( \widehat{\theta}_k^{slow} \right)^3 \right] \leq \frac{C}{K^3}, \quad \mathbb{E}_{k-1}^n \left[ \left( \widehat{\theta}_k^{slow} \right)^2 \right] \leq \frac{C}{K^2}.
\end{aligned} \tag{27}$$

From here,

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

and

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

The second step is to show

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

It is sufficient to show

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

Write

$$\begin{aligned}
K \sum_{k=1}^K \mathbb{E} \left[ |\gamma_k^{DISCR} - \gamma_k| \right] & = K \sum_{k=1}^K \mathbb{E} \left[ \left| \widehat{\alpha}_k^{fast} - \widehat{\alpha}_k^{slow} + \widehat{\theta}_k^{fast} - \widehat{\theta}_k^{slow} \right| \left| \left\{ \widehat{\alpha}_k^{fast} - \widehat{\theta}_k^{fast} \right\} - \left\{ \widehat{\alpha}_k^{slow} - \widehat{\theta}_k^{slow} \right\} \right| \right] \\
& \equiv A + B
\end{aligned}$$

As to the first term, we have

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

In above, to obtain second inequality, we used (27). To obtain the fourth inequality, we used

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

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 (2001). For every  $i$  and  $k$ , there exists a constant  $c_{i,k}$  s.t.

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

and

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

Notice that

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

by right-continuity and boundedness of  $\sigma$ . Then,

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

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

The final step is to show

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

We have

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

Therefore,

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

The result follows immediately.  $\square$

#### A.4 Proof of Theorem 4

To simplify the proof, assume  $\mu \equiv 0$ . Below arguments can easily be extended to nonzero drift. Recall that Theorem 4 assumes  $\{\sigma\} \perp \{W\}$ . Therefore, all arguments are done conditional on volatility. This Theorem is proved by validating the conditions of Theorem 6 for the Two Time Scale estimator of Ait-Sahalia et al. (2006a).

We first introduce some additional notation. For some arguments  $w$  and  $q$ , define

$$\bar{q}_w = \frac{q - w + 1}{w}.$$

This is the average numbers of observations in a (Infill Price type) subsample if total number of observations is  $q$  and there are  $w$  subsamples.

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

$$\begin{aligned} V_l^{signal} &= \frac{4}{3}c \int_{(l-1)m/n}^{[(l-1)m+J]/n} \sigma_u^4 du \\ V_l^{noise} &= 8c^{-2} \frac{J}{n} \text{Var}(\epsilon)^2 + 16 \frac{J}{n} c^{-2} \lim_{n \rightarrow \infty} \sum_{i=1}^n \text{Cov}(\epsilon_0, \epsilon_{i/n})^2. \end{aligned}$$

We now prove Assumption A6(i) is satisfied. For that, we decompose  $\zeta_l^{(n)}$  into signal and noise components. For the signal part, we show that arguments of Aït-Sahalia et al. (2006a) carry over to the subsample. For the noise part, their arguments apply directly.

The decomposition is

$$\begin{aligned} & \mathbb{E} \left( \frac{n}{J} \left( n^{1/3} \left( \hat{\theta}_l^{short} - \theta_l^{short} \right)^2 - V_l^{short} \right) \right) \\ &= \frac{n}{J} \left( n^{1/3} \mathbb{E} \left[ \left( [Y, Y]_l^{(G_1)} - \frac{\bar{J}_{G_1}}{\bar{J}_{G_2}} [Y, Y]_l^{(G_2)} - \theta_l^{short} \right)^2 \right] - V_l^{short} \right) \\ &= \frac{n}{J} \left( n^{1/3} \mathbb{E} \left[ \left( [X, X]_l^{(G_1)} - \theta_l^{short} + [\epsilon, \epsilon]_l^{(G_1)} - \frac{\bar{J}_{G_1}}{\bar{J}_{G_2}} [\epsilon, \epsilon]_l^{(G_2)} \right)^2 \right] - V_l^{short} \right) + R_1 \\ &= \frac{n^{4/3}}{J} \mathbb{E} \left( [X, X]_l^{(G_1)} - \theta_l^{short} \right)^2 + \frac{n^{4/3}}{J} \mathbb{E} \left( [\epsilon, \epsilon]_l^{(G_1)} - \frac{\bar{J}_{G_1}}{\bar{J}_{G_2}} [\epsilon, \epsilon]_l^{(G_2)} \right)^2 - \frac{n}{J} V_l^{short} + R_1 + R_2. \end{aligned}$$

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

$$\frac{n^{4/3}}{J} \mathbb{E} \left( [X, X]_l^{(G_1)} - \theta_l^{short} \right)^2 - \frac{n}{J} V_l^{signal} = o(1). \quad (28)$$

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

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

where

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

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

$$\mathbb{E}(R_3) \leq C \frac{G_1}{n}, \quad \mathbb{E}(R_3^2) \leq C \frac{G_1^2}{n^2}$$

By (29) and (30), we have  $E(S_l^2) \leq Jn^{-4/3}$  and  $n^{1/3}E\left([X, X]_l^{(1)} - \theta_l^{short}\right)^2 \leq Cn^{-5/3}$ . Therefore, to prove (28), it is sufficient to show

$$\frac{n^{4/3}}{J}E\left([X, X]_l^{(1)} + S_l - \theta_l^{short}\right)^2 - \frac{n}{J}V_l^{signal} = o(1).$$

We have

$$\begin{aligned} & E(S_l^2) \tag{29} \\ &= E\left[\left(2\sum_{i=1}^{J-1}(\Delta X_{[(l-1)m+i]/n})\sum_{j=1}^{G_1 \wedge i}\left(1 - \frac{j}{G_1}\right)(\Delta X_{[(l-1)m+i-j]/n})\right)^2\right] \\ &= 4\sum_{i=1}^{J-1}\int_{[(l-1)m+i-1]/n}^{[(l-1)m+i]/n}\sigma_u^2 du\sum_{j=1}^{G_1 \wedge i}\left(1 - \frac{j}{G_1}\right)^2\int_{((l-1)m+i-j-1)/n}^{((l-1)m+i-j)/n}\sigma_u^2 du \\ &= 4\sum_{i=G_1}^{J-1}\int_{[(l-1)m+i-1]/n}^{[(l-1)m+i]/n}\sigma_u^2 du\sum_{j=1}^{G_1}\left(1 - \frac{j}{G_1}\right)^2\int_{((l-1)m+i-j-1)/n}^{((l-1)m+i-j)/n}\sigma_u^2 du + o\left(\frac{J}{n^{4/3}}\right) \\ &= 4\sum_{i=G_1}^{J-1}\frac{\sigma_{[(l-1)m+i]/n}^2}{n}\sum_{j=1}^{G_1}\left(1 - \frac{j}{G_1}\right)^2\frac{\sigma_{[(l-1)m+i]/n}^2}{n} + R_4 + R_5 + o\left(\frac{J}{n^{4/3}}\right) \\ &= \frac{4}{3}\frac{G_1}{n^2}\sum_{i=G_1}^{J-1}\sigma_{[(l-1)m+i]/n}^4 + R_4 + R_5 + o\left(\frac{J}{n^{4/3}}\right) \\ &= \frac{4}{3}\frac{c}{n^{1/3}}\int_0^{J/n}\sigma_u^4 du + R_4 + R_5 + o\left(\frac{J}{n^{4/3}}\right) \end{aligned}$$

where we use  $G_1 = cn^{2/3}$ .

The remainder terms  $R_4$  and  $R_5$  have expressions as below, and they are of smaller order than  $Jn^{-4/3}$  by (24),

$$\begin{aligned} & \frac{n^{4/3}}{J}R_4 \\ &= \frac{n^{4/3}}{J}4\sum_{i=G_1}^{J-1}\frac{\sigma_{[(l-1)m+i]/n}^2}{n}\sum_{j=1}^{G_1}\left(1 - \frac{j}{G_1}\right)^2\int_{((l-1)m+i-j-1)/n}^{((l-1)m+i-j)/n}(\sigma_u^2 - \sigma_{i/n}^2) du \\ &\leq C\frac{n^{4/3}}{J}4\sum_{i=G_1}^{J-1}\frac{\sigma_{[(l-1)m+i]/n}^2}{n}\sum_{j=1}^{G_1}\left(1 - \frac{j}{G_1}\right)^2\frac{1}{n}\sqrt{\frac{G_1}{n}} \leq C\frac{n^{4/3}}{J}\frac{1}{n^2}\sqrt{\frac{G_1}{n}}JG_1 \rightarrow 0, \end{aligned}$$

$$\begin{aligned} & \frac{n^{4/3}}{J}R_5 \\ &= \frac{n^{4/3}}{J}\frac{n^{4/3}}{J}4\sum_{i=G_1}^{J-1}\int_{[(l-1)m+i-1]/n}^{[(l-1)m+i]/n}(\sigma_u^2 - \sigma_{i/n}^2) du\sum_{j=1}^{G_1}\left(1 - \frac{j}{G_1}\right)^2\int_{((l-1)m+i-j-1)/n}^{((l-1)m+i-j)/n}\sigma_u^2 du \\ &\leq C\frac{n^{4/3}}{J}4\sum_{i=G_1}^{J-1}\frac{1}{n^{3/2}}\sum_{j=1}^{G_1}\left(1 - \frac{j}{G_1}\right)^2\frac{1}{n} \leq C\frac{n^{4/3}}{J}\frac{1}{n^{3/2}}\frac{1}{n}JG_1 \rightarrow 0. \end{aligned}$$

Therefore,

$$\frac{n^{4/3}}{J} \mathbb{E} (S_l^2) - \frac{n}{J} V_l^{signal} = o(1).$$

Finally,

$$\frac{n^{4/3}}{J} \mathbb{E} \left( [X, X]_l^{(1)} - \theta_l^{short} \right)^2 = \frac{n^{4/3}}{J} \text{Var} \left( [X, X]_l^{(1)} \right) \leq C \frac{n^{4/3}}{J} \frac{J}{n^2} = o(1). \quad (30)$$

This concludes the proof of (28). Next, we turn to the noise part and prove

$$n^{1/3} \mathbb{E} \left( [\epsilon, \epsilon]_l^{(G_1)} - \frac{\bar{J}_{G_1}}{\bar{J}_{G_2}} [\epsilon, \epsilon]_l^{(G_2)} \right)^2 - V_l^{short} = o\left(\frac{J}{n}\right). \quad (31)$$

In this case, Proposition 1 of Aït-Sahalia et al. (2006a) can be applied directly, with  $J$  instead of  $n$  (this is the number of observations used above) to obtain

$$\begin{aligned} \frac{G_1}{\sqrt{J}} \left( [\epsilon, \epsilon]_l^{(G_1)} - \frac{\bar{J}_{G_1}}{\bar{J}_{G_2}} [\epsilon, \epsilon]_l^{(G_2)} \right) &\Rightarrow N \left( 0, 8\text{Var}(\epsilon)^2 + 16 \lim_{n \rightarrow \infty} \sum_{i=1}^n \text{Cov}(\epsilon_0, \epsilon_{i/n})^2 \right) \\ \frac{G_1^2}{J} \mathbb{E} \left( [\epsilon, \epsilon]_l^{(G_1)} - \frac{\bar{J}_{G_1}}{\bar{J}_{G_2}} [\epsilon, \epsilon]_l^{(G_2)} \right)^2 &\rightarrow 8\text{Var}(\epsilon)^2 + 16 \lim_{n \rightarrow \infty} \sum_{i=1}^n \text{Cov}(\epsilon_0, \epsilon_{i/n})^2 \\ \frac{n^{4/3}}{J} \mathbb{E} \left( [\epsilon, \epsilon]_l^{(G_1)} - \frac{\bar{J}_{G_1}}{\bar{J}_{G_2}} [\epsilon, \epsilon]_l^{(G_2)} \right)^2 - \frac{n}{J} V^{noise} &\rightarrow 0, \end{aligned}$$

which immediately gives (31).

The final step is to show  $R_1 + R_2 = o(1)$ . We have

$$R_1 = \frac{n}{J} n^{1/3} \mathbb{E} \left[ \left( \frac{\bar{J}_{G_1}}{\bar{J}_{G_2}} [X, X]_l^{(G_2)} \right)^2 \right] + \frac{n}{J} n^{1/3} \mathbb{E} \left[ \left( 2[X, \epsilon]_l^{(G_1)} \right)^2 \right] + \left( \frac{n}{J} \right)^{1/3} \mathbb{E} \left[ \left( \frac{\bar{J}_{G_1}}{\bar{J}_{G_2}} [X, \epsilon]_l^{(G_2)} \right)^2 \right] + R'_1$$

where, for  $i = 1, 2$ ,

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

First term is  $o(1)$  because  $[X, X]_l^{(G_2)} \xrightarrow{p} \theta_l^{short}$  by substituting  $G_2$  for  $G_1$  in (28). Second and third terms are of smaller order than  $m/n$  by proof of Lemma 1 of Aït-Sahalia et al. (2006a), which implies, for  $i = 1, 2$ ,

$$\mathbb{E} \left( \left( [X, \epsilon]_l^{(G_i)} \right)^2 \mid X \right) \leq C \frac{1}{G_i^2} [X, X]_l^{(G_i)}.$$

The final terms  $R'$  and  $R_2$  contain cross terms that are negligible by Cauchy-Schwarz inequality. This concludes verification of the assumption A6(i).

Assumption A6(ii) can be verified using straightforward calculations, by using the mixing property of the noise  $\epsilon$ , as well as  $L^{4+\delta}$  boundedness of  $\epsilon$  for some  $\delta$ , and finally

$$\mathbb{E} (|X_t - X_{t-s}|^q) \leq C_q s^{q/2},$$

for some constant  $C_q$  depending on  $q$  only, for all  $q > 0$  and all  $s$ . Above follows from Burkholder-Davis-Gundy inequality.

Assumption A5 is immediate due to  $\{\mu\}, \{\sigma\} \perp \{W\}$  assumption. This implies that conditional on volatility and drift path, returns are independent over non-overlapping intervals. Hence, strong mixing of  $\zeta_l^{(n)}$  follows from strong mixing of the  $\epsilon$ , which in turn holds by assumption. This concludes the proof of Theorem 4.  $\square$

## A.5 Proof of Lemma 5

Most of the proof of the asymptotic distribution of TSRV estimator of Aït-Sahalia et al. (2006a) 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. (2006a). Given that  $G_1/G_2 \rightarrow 0$  and

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

due to differentiability of  $\omega$ , the desired result follows.  $\square$

## A.6 Proof of Theorem 6

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

$$G^{(n)} = \frac{mJ}{n^2} \sum_{l=1}^{n/m} \tau_n^2 \left( \frac{n}{J} \hat{\theta}_l^{short} - \frac{n}{J} \theta_l^{short} \right)^2 \xrightarrow{p} V \quad (32)$$

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

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

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

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

By A4, we have

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

and so (32) follows.

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

$$\begin{aligned}
\widehat{V} - G^{(n)} &= \frac{Jm\tau_n^2}{n^2} \sum_{l=1}^K \left( \frac{n}{J} \widehat{\theta}_l^{short} - \frac{n}{m} \widehat{\theta}_l^{long} \right)^2 - \frac{Jm\tau_n^2}{n^2} \sum_{l=1}^{n/m} \tau_n^2 \left( \frac{n}{J} \widehat{\theta}_l^{short} - \frac{n}{J} \theta_l^{short} \right)^2 \\
&= \frac{Jm\tau_n^2}{n^2} \sum_{l=1}^K \left( \frac{n}{m} \widehat{\theta}_l^{long} - \frac{n}{m} \theta_l^{long} \right)^2 + \frac{Jm\tau_n^2}{n^2} \sum_{l=1}^K \left( \frac{n}{m} \theta_l^{long} - \frac{n}{J} \theta_l^{short} \right)^2 \\
&\quad + 2 \frac{Jm\tau_n^2}{n^2} \sum_{l=1}^K \left( \frac{n}{m} \widehat{\theta}_l^{long} - \frac{n}{m} \theta_l^{long} \right) \left( \frac{n}{m} \theta_l^{long} - \frac{n}{J} \theta_l^{short} \right) \\
&\quad - 2 \frac{Jm\tau_n^2}{n^2} \sum_{l=1}^{n/m} \tau_n^2 \left( \frac{n}{J} \widehat{\theta}_l^{short} - \frac{n}{J} \theta_l^{short} \right) \left( \frac{n}{m} \widehat{\theta}_l^{long} - \frac{n}{J} \theta_l^{short} \right).
\end{aligned} \tag{33}$$

We have the following decomposition,

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

These terms are small enough due to A4 and (24) as follows,

$$\begin{aligned}
& \mathbb{E} \left| \frac{Jm\tau_n^2}{n^2} \sum_{l=1}^K \left( \frac{n}{m} \int_{(l-1)m/n}^{lm/n} (f(u) - f((l-1)m/n)) du \right)^2 \right| \\
& \leq \frac{Jm\tau_n^2}{n^2} \sum_{l=1}^K \mathbb{E} \left( \frac{n}{m} \int_{(l-1)m/n}^{lm/n} (f(u) - f((l-1)m/n)) du \right)^2 \\
& = \frac{Jm\tau_n^2}{n^2} \sum_{l=1}^K \mathbb{E} (f(s_l) - f((l-1)m/n))^2 \\
& \leq C \frac{Jm\tau_n^2}{n^2} \sum_{l=1}^K \mathbb{E} (\sigma(s_l) - \sigma((l-1)m/n))^2 \\
& \leq C \frac{Jm\tau_n^2}{n^2} \sum_{l=1}^K \frac{m}{n} = C \frac{Jm\tau_n^2}{n^2} \rightarrow 0
\end{aligned} \tag{34}$$

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

$$\int_{(l-1)m/n}^t (f(u) - f((l-1)m/n)) du \tag{35}$$

in time.

Next, we show

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

By substituting  $m$  for  $J$  in

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

we obtain

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

and so by multiplying left hand side by  $J/m$ , (35) follows since  $J/m \rightarrow 0$ .

The remaining cross-terms in (33) are negligible by above results and Cauchy-Schwarz inequality. This concludes the proof of Theorem 6.  $\square$

## B Tables and Figures

Table 1. Coverage probabilities of 95% confidence interval of  $IV_X$ ,  $\xi^2 = 0.001$

		$J = 200$			$J = 500$			$\widehat{V}_a$
	m	800	2000	3000	2000	5000	7500	
$\rho = 0$	<i>Two-sided</i>	0.96	0.97	0.98	0.92	0.95	0.95	0.93
	<i>Left-sided</i>	0.94	0.95	0.95	0.91	0.92	0.92	0.91
	<i>Right-sided</i>	0.97	0.98	0.98	0.96	0.97	0.97	0.96
$\rho = -0.3$	<i>Two-sided</i>	0.97	0.97	0.98	0.93	0.95	0.95	0.91
	<i>Left-sided</i>	0.96	0.97	0.97	0.93	0.95	0.95	0.91
	<i>Right-sided</i>	0.97	0.98	0.98	0.95	0.97	0.96	0.92
$\rho = -0.5$	<i>Two-sided</i>	0.97	0.98	0.98	0.93	0.95	0.96	0.84
	<i>Left-sided</i>	0.95	0.96	0.96	0.93	0.94	0.94	0.87
	<i>Right-sided</i>	0.97	0.98	0.98	0.95	0.97	0.97	0.91
$\rho = -0.7$	<i>Two-sided</i>	0.97	0.98	0.99	0.95	0.95	0.96	0.78
	<i>Left-sided</i>	0.95	0.96	0.96	0.92	0.93	0.94	0.81
	<i>Right-sided</i>	0.98	0.99	0.99	0.97	0.98	0.98	0.89

Table 2. Coverage probabilities of 95% confidence interval of  $IV_X$ ,  $\xi^2 = 0.0001$

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

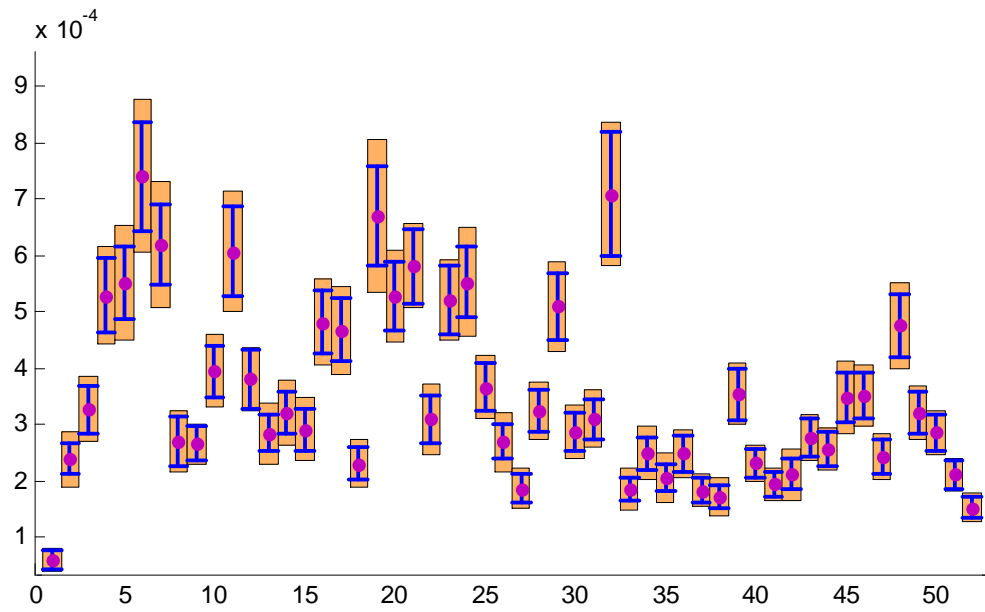


Figure 6: *AIG stock*

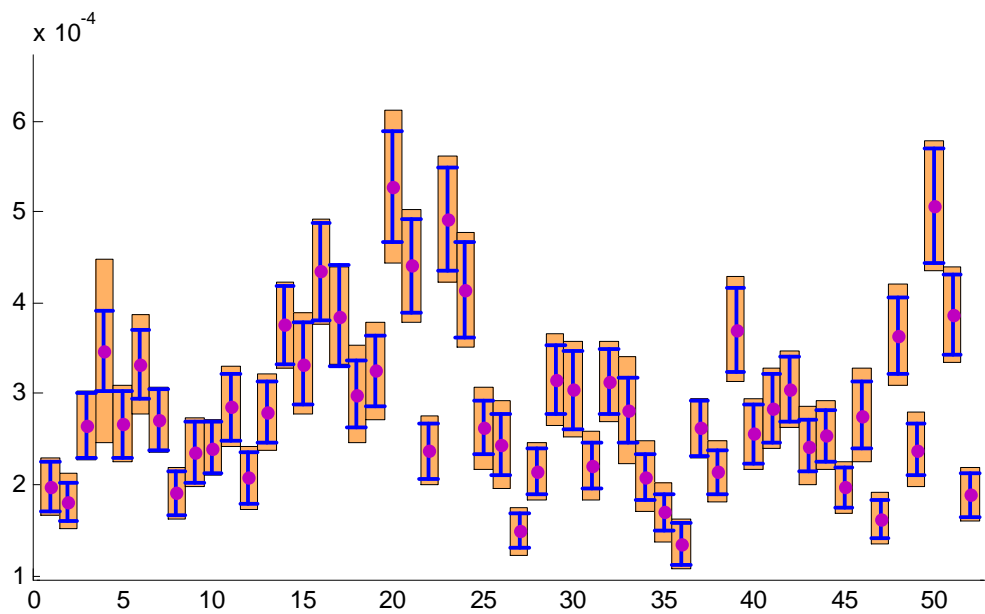


Figure 7: *GE stock*

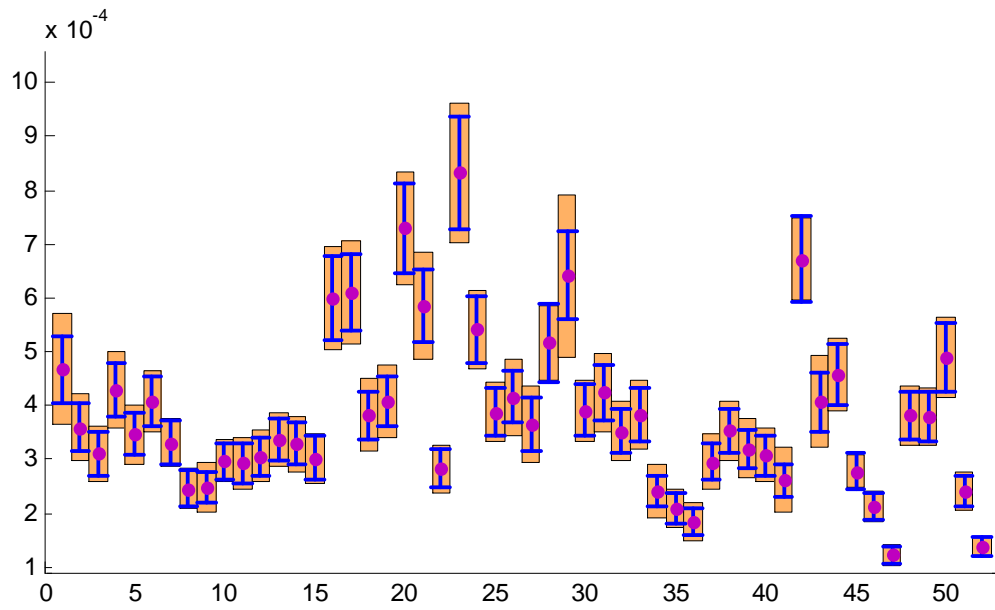


Figure 8: *IBM stock*

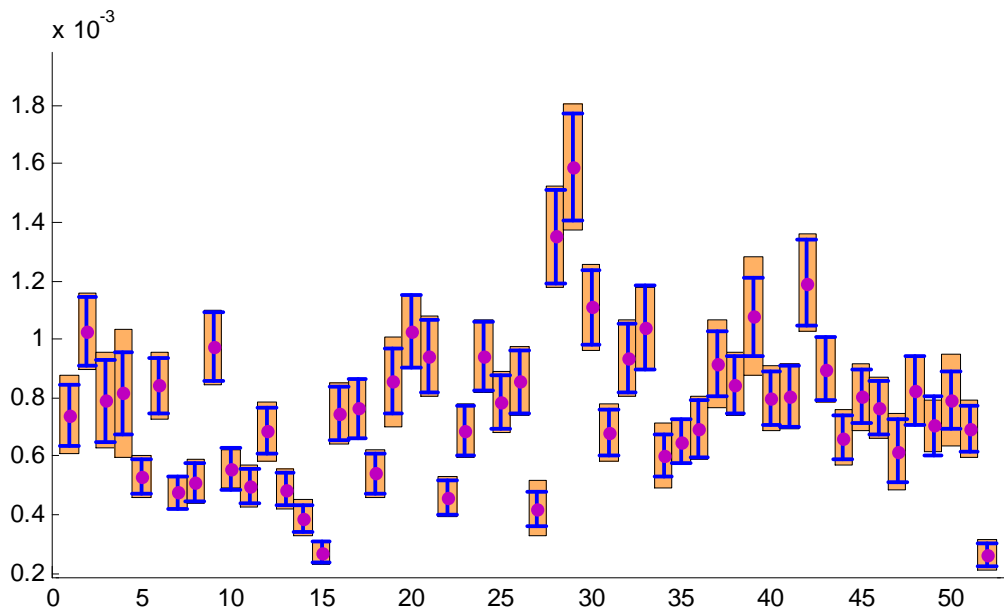


Figure 9: *INTC stock*

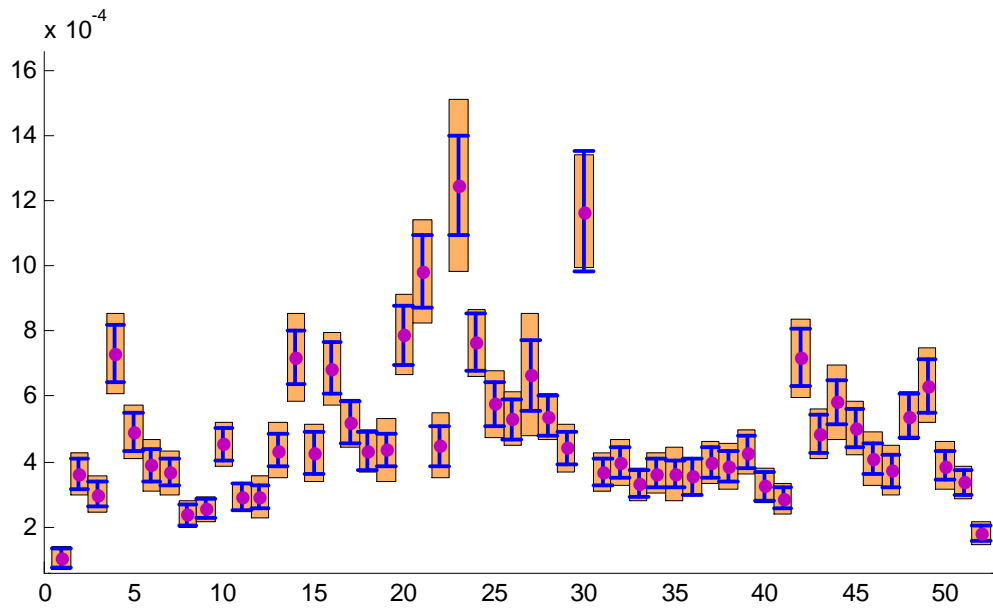


Figure 10: *MMM stock*

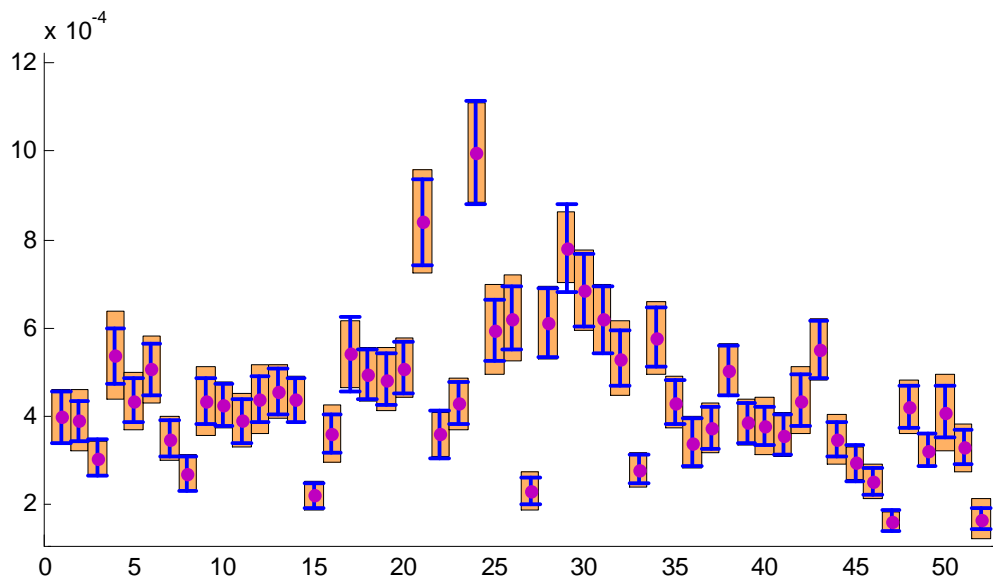


Figure 11: *MSFT stock*