Journal of Econometrics 208 (2019) 101–119

Contents lists available at ScienceDirect

Journal of Econometrics

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

The algebra of two scales estimation, and the S-TSRV: High frequency estimation that is robust to sampling times[☆]

Per A. Mykland[a,*], Lan Zhang[b], Dachuan Chen[b]

[a] The University of Chicago, United States
[b] University of Illinois at Chicago, United States

ARTICLE INFO

Article history:
Available online 22 October 2018

JEL classification:
C02
C13
C14
C15
C22

Keywords:
Asynchronous times
Consistency
Discrete observation
Efficiency
Endogenous times
Equivalent martingale measure
Irregular times
Itô process
Leads and lags
Leverage effect
Microstructure
Pre-averaging
Realized volatility
Robust estimation
Stable convergence
Two scales estimation

ABSTRACT

In this paper, we derive a new algebraic property of two scales estimation in high frequency data, under which the effect of sampling times is canceled to high order. This is a particular robustness property of the two scales construction. In general, irregular, asynchronous, or endogenous times can cause problems in estimators based on equidistant observation of (trade or quote) times.

The new algebraic property can be combined with pre-averaging, giving rise to the *smoothed two-scales realized volatility (S-TSRV)*. We derive a finite sample solution to controlling edge effects and for handling irregular and endogenous observation times and asynchronously observed multivariate data. In connection with this development, we use the algebraic approach to define a version of the S-TSRV which has particularly small edge effect in microstructure noise. The main result of the paper is a representation of the statistical error of the estimator in terms of simple components. As an application of this representation, the paper develops a central limit theory for multivariate volatility estimators. The approach can also handle leads and lags in the signal process.

1. Introduction

High frequency financial data is an increasingly important source of knowledge about financial markets, originally focused on the concept of *realized volatility* (RV) (Andersen and Bollerslev, 1998a,b; Barndorff-Nielsen and Shephard, 2002), and later branching out to covariance, regression, leverage effect, etc.

Microstructure noise, however, is a main barrier to inference in such data. The purpose of this paper is to develop a family of volatility estimators that have a finite sample algebraic representation in terms of classical (de-noised, but unobserved)

☆ We would like to thank the editor and referees for suggestions that substantially improved the paper. Financial support from the National Science Foundation, United States under grants DMS 14-07812 and DMS 17-13129 (Mykland), and DMS 14-07820 and DMS 17-13118 (Zhang) is also gratefully acknowledged.

* Corresponding author.
E-mail address: mykland@pascal.uchicago.edu (P.A. Mykland).

https://doi.org/10.1016/j.jeconom.2018.09.007
0304-4076/

realized volatility. This will provide small sample guidance on the properties of the estimator, and also an easy link to asymptotic results developed for the no-noise case. It will, for example, permit the handling of endogenous times in the case where there is microstructure noise. The approach extends to estimators of realized covariance.

The early development of this area was based on the assumption that log prices follow a semimartingale process X_t of the form

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

where J_t is a pure jump process and W_t is Brownian motion; μ_t and σ_t are random processes that can be dependent with W. This semimartingale model for prices is required by the no-arbitrage principle in finance theory (Delbaen and Schachermayer, 1994, 1995, 1998). The data, however, had an unexpected feedback to the theory: log prices are not semimartingales after all. This was clarified by the so-called *signature plot* (introduced by Andersen et al. (2000). See also the discussion in Mykland and Zhang (2005)).

Among several researchers, Zhang et al. (2005) investigated a model where the efficient price X_t is latent, and one actually observes

$$Y_{t_j} = X_{t_j} + \epsilon_{t_j}. \tag{2}$$

The distortion ϵ_i is called either "microstructure noise" or "measurement error", depending on one's academic field (O'Hara, 1995; Hasbrouck, 1996). The t_j can be transaction times, or quote times.

Zhang et al. (2005) proposed a *Two Scales Realized Volatility (TSRV)*, which we shall revisit in this paper. The approach consists of combining two RVs that are formed from different sampling scales. A more general approach consists in averaging several scales (Zhang, 2006). There are by now a number of different angles on the estimation of volatility under microstructure, and another method which will be central to our development is called *Pre-Averaging*: take weighted local averages of the data (log prices) before taking squares (Jacod et al., 2009a; Podolskij and Vetter, 2009a,b; Jacod et al., 2009b). There is a symmetry to these two approaches: two- and multi-scale estimation could also be described as "post-averaging".[1] The point of departure in the present paper is the following. The cited papers show results when the latent process X_t is continuous and the noise is iid The observation times are variously taken to be non-endogenous (two- and multiscale) or simply equidistant (in the case of pre-averaging). It has subsequently been quite difficult to figure out how sensitive these estimators are to such assumptions, and also to find asymptotic laws under more general conditions. Despite the years that have passed since 2005, there is still no comprehensive solution to these questions.

For example, Kalnina and Linton (2008) find that a strong diurnal pattern in the noise may severely affect estimators. On the other hand, Aït-Sahalia et al. (2011) find that two- and multiscale estimators are robust to noise that is stationary and sufficiently fast mixing. Second, in the matter of observation times, Li et al. (2014) find that endogenous times will introduce asymptotic bias in the case where there is no noise. However, this result is hard to adapt to the case with microstructure and still preserve a convergence rate that is close to efficient (Li et al., 2013). Third, the case with jumps is less well explored when there is microstructure; compare, for example, Theorem 5.4.2 (p. 162) and Theorem 16.6.1 (p. 554) in Jacod and Protter (2012).

The problem comes up with a vengeance when the process (1)–(2) is multidimensional. Here data can be observed at times which are *asynchronous*. There are available solutions to this issue in the two- and multi-scale (Zhang, 2011; Bibinger and Mykland, 2016), and Fourier (Park, 2011; Park et al., 2016; Mancino et al., 2017) approaches. For pre-averaging, however, the approach makes assumptions about the irregularity of times being benign, in the sense that observation times are a fixed transformation of an equidistant grid (Christensen et al. (2013), cf. their Assumption T1 p. 4–5, this is "mildly irregular" in the typology of Mykland and Zhang (2016, Section 2.6, pp. 248–249)). See the further discussion in Section 2 in the current paper. The cited papers also assume continuity of X_t.

In summary, volatility estimators have mostly been defined and studied to cope with one specific deviation from the continuous version of model (1), and it has turned out to be hard to study microstructure together with either jumps, and/or with various forms of irregular, asynchronous and/or endogenous observation times.

We here present a way out. This is to use a combination of the two-scales and pre-averaging constructions. We call it smoothed two-scales realized volatility (S-TSRV). We shall show in Section 3 that the S-TSRV to high order approximates an (unobserved) RV that is only based on X.

The quality of the S-TSRV is not an accidental phenomenon. In fact, it is due to an *algebraic* cancellation of terms in finite samples. We shall see this in Theorems 1–2 in Section 3. Conceptually, this is the main finding of the paper.

Meanwhile, there does not seem to be a trade-off with efficiency. We shall investigate a procedure which combines pre-averaging and two-scales estimation, as follows. First pre-average the data in small time intervals, say, 15 s, then use a two-scales estimator. The algebraic cancellation property is still valid, and thus, in particular, an efficient rate of convergence can be achieved.

[1] Three other main approaches to this estimation problem are the *Realized Kernel* approach, which uses weighted autocovariances (Barndorff-Nielsen et al., 2008), *Quasi-likelihood* (Xiu, 2010), and the *spectral* approach of Bibinger and Reiß (2014) and Altmeyer and Bibinger (2015). These will not be central to our current narrative. These approaches do, however, have similar problems to the ones discussed here concerning the use of observation times when these times are irregular and/or asynchronous. A further main approach is the *Fourier* method (Park, 2011; Park et al., 2016; Mancino et al., 2017), which is more well posed in its handling of times.

For multi-dimensional data, the proposed procedure also serves as a synchronization device. One can pre-average each series over the same 15 s, and they then become synchronous. The two scales algebraic cancellation then takes care of residual asynchronicity. Further development of the S-TSRV in the high-dimensional case, along with related literature, can be found in Chen et al. (2018).

For the purposes of asymptotics, our algebraic results serve as a *strong representation* of the S-TSRV in terms of the RV of X. Asymptotic results can therefore be obtained more easily through this path. We do not pursue this in the most general case, but show among other results that endogenous times can be tackled under microstructure noise with the efficient $n^{1/4}$ rate of convergence. This is also discussed in Item 2 in Section 4.3.

The strong representation property raises the question of whether the entire estimation error also can be given a strong representation, along the lines of Wu (2007). This would require an extension of Wu's theory to the stable convergence case.

We emphasize that our results cover the original TSRV estimator. As micro-structure declines in increasingly efficient markets, loss of efficiency of TSRV is often a non-problem, and the transparency of the estimator recommends it in many situations.

In the following, we shall see in Section 2 that observation times impact pre-averaging. Section 3 presents the main algebraic results, and also studies a modification of the two scales construction that completely eliminates edge effects due to "squared-noise". We finally show that a J- and K-scales estimator has a representation as a single K–J scale estimator in the signal process X. Section 4.1 gives asymptotic bounds for the edge effect (the error in the finite sample representation). Section 4.2 gives an asymptotic representation for the estimator in terms of average realized volatilities and noise U-statistics. This is for the case where the sparse scale $K \to \infty$, and it yields an uncluttered form with which to analyze the S-TSRV. Section 5 provides asymptotic theory for smoothed two scales realized covariances. Section 6 establishes robustness for finite K and for estimates of spot volatility. In Section 7, we show how the setup extends to other schemes for dealing with irregularity and asynchronicity. We also discuss how to implement a rolling windows approach.

2. Pre-averaging does not work on its own

We here recall the concept of pre-averaging (Section 2.1), and then analyze the variance of the return on the pre-averaged signal (Section 2.2) when observation times are irregular. We finally show (in Section 2.3) that pre-averaging by itself does not assure consistency of estimators when times are irregular.

2.1. Pre-averaging

Our general theory starts with approximating the efficient price in small neighborhoods. Specifically, we assume that observations on the form (1)–(2) are made at times $0 = t_0 < \cdots < t_i < \cdots < t_n = T$. The index n represents the total number of observations, and our arguments will be based on asymptotics as $n \to \infty$ while T is fixed. Meanwhile, neighborhoods or blocks are defined by a much less dense grid of τ_i, also spanning $[0, T]$, so that block # $i = \{\tau_{i-1} < t_j \le \tau_i\}$ (the first block, however, includes 0). We define the block size by $M_{n,i} = M_i = \#\{j : \tau_{i-1} < t_j \le \tau_i\}$, We then seek an estimate of the value of the efficient price in the time period $(\tau_{i-1}, \tau_i]$ by *pre-averaging*, which is defined as follows. Define block averages for block i, $(\tau_{i-1}, \tau_i]$:

$$\bar{Y}_i = \frac{1}{M_i} \sum_{\tau_{i-1} < t_j \le \tau_i} Y_{t_j},$$

and let $\bar{X}_i$ be defined similarly based on X. The averaging yields a reduction of the size of microstructure noise (see, for example, Mykland and Zhang (2016, Example 1, p. 244)). This is obtained if $M_{n,i} \to \infty$ with n, but sufficiently slowly that the actual time interval $(\tau_{i-1}, \tau_i]$ stays small.[2] The number of blocks will be called N_n.

2.2. The return on the pre-averaged signal

We here give a first order analysis of pre-averaging when observation times are not assumed to be equidistant or otherwise regular. At this time, assume that the times are exogenous (times are allowed to be endogenous in subsequent sections).

To find a compact characterization of the effect of such times, define (as in Section 2.6 of Mykland and Zhang (2016)) the random variable $I_i = I_{n,i}$ inside each block i as follows. Let t_{j_0} be the first $t_j \in (\tau_{i-1}, \tau_i]$, and set

$$I_i = \begin{cases} \dfrac{M_i - j}{M_i} \text{ with probability } \dfrac{\Delta t_{j_0+j}}{\Delta \tau_i} \\ 1 \text{ with probability } \dfrac{t_{j_0} - \tau_{i-1}}{\Delta \tau_i} \\ 0 \text{ with probability } \dfrac{\tau_i - t_{j_0+M_i-1}}{\Delta \tau_i} \end{cases} \tag{3}$$

[2] When reference to the total number n of observations is needed, we write $t_{n,j}$ instead of t_j, $\tau_{n,i}$ instead of τ_i, $M_{n,i}$ instead of $M_{n,i}$, and so on.

where $j = 1, 2, \ldots, M_{i-1}$ and $\Delta t_{j_0+j} = t_{j_0+j} - t_{j_0+j-1}$. In particular, definition (3) means that for each block i,

$$E(I_i) = \sum_{t_j \in (\tau_{i-1}, \tau_i]} \frac{M_i - j}{M_i} \frac{\Delta t_{j_0+j}}{\Delta \tau_i} + \frac{t_{j_0} - \tau_{i-1}}{\Delta \tau_i} \text{ and } E(I_i^2) = \sum_{t_j \in (\tau_{i-1}, \tau_i]} \left(\frac{M_i - j}{M_i} \right)^2 \frac{\Delta t_{j_0+j}}{\Delta \tau_i} + \frac{t_{j_0} - \tau_{i-1}}{\Delta \tau_i}.$$

Denote $V_i = \bar{X}_i - X_{\tau_{i-1}}$, and $V_i' = X_{\tau_i} - \bar{X}_i$. Decompose

$$V_i = \tilde{V}_i + E(I_i)\Delta X_{\tau_i} \text{ and } V_i' = \Delta X_{\tau_i} - V_i = -\tilde{V}_i + (1 - E(I_i))\Delta X_{\tau_i}, \tag{4}$$

where the first equation in (4) serves as definition of $\tilde{V}_i$. Obviously, $\tilde{V}_i$ is uncorrelated with ΔX_{τ_i}. We have

$$\Delta X_{\tau_i} = V_i + V_i' \tag{5}$$

while

$$\Delta \bar{X}_{i+1} = V_{i+1} + V_i'. \tag{6}$$

Assume for the rest of Section 2 that X_t is continuous. Because of (6), if we approximate σ_t^2 by $\sigma_{\tau_{i-1}}^2$ over the interval $(\tau_{i-1}, \tau_{i+1}]$, the squared-return of a pre-averaged observation becomes

$$\begin{aligned}
E[(\Delta \bar{X}_{i+1})^2 \mid \mathcal{F}_{\tau_{i-1}}] &= E[(V_i')^2 \mid \mathcal{F}_{\tau_{i-1}}] + E[(V_{i+1})^2 \mid \mathcal{F}_{\tau_{i-1}}] \\
&= (1 - E(I_i))^2 E[(\Delta X_{\tau_i})^2 \mid \mathcal{F}_{\tau_{i-1}}] + E[(\tilde{V}_i)^2 \mid \mathcal{F}_{\tau_{i-1}}] \\
&+ (E(I_{i+1}))^2 E[(\Delta X_{\tau_{i+1}})^2 \mid \mathcal{F}_{\tau_{i-1}}] + E[(\tilde{V}_{i+1})^2 \mid \mathcal{F}_{\tau_{i-1}}] \\
&\approx \sigma_{\tau_{i-1}}^2 [\Delta \tau_i ((1 - E(I_i))^2 + \mathrm{Var}(I_i)) + \Delta \tau_{i+1} (E(I_{i+1})^2 + \mathrm{Var}(I_{i+1}))] \\
&= \sigma_{\tau_{i-1}}^2 [\Delta \tau_i (E(1 - I_i)^2) + \Delta \tau_{i+1} E((I_{i+1})^2)],
\end{aligned} \tag{7}$$

where $\mathcal{F}_{\tau_{i-1}}$ is the sigma-field representing the information at time τ_{i-1}, cf. Condition 1. This is a standard approximation in this setting. One way of seeing the validity is to use the contiguity results in Mykland and Zhang (2016).

2.3. The pre-averaged RV

We shall here see that pre-averaging by itself does not estimate volatility. To clarify the implications of (7) above, define that a sequence of times $t_{n,j}$ is *regular* provided, for any sequence i_n, $n \to \infty$, $I_{i_n} = I_{n,i_n}$ converges in law to a uniform (0, 1) random variable. For regular times,

$$E(I_{n,i_n}) \approx \frac{1}{2}, \quad E(I_{n,i_n}^2) \approx \frac{1}{3}, \text{ and } E(2I_{n,i_n}^2 - 2I_{n,i_n} + 1) \approx \frac{2}{3} \tag{8}$$

in the sense of limit in probability as $n \to \infty$. Regular times include equidistant observations, and times distributed by a Poisson process, for which (8) holds exactly. (More generally, see Mykland and Zhang (2016, Section 2.6, pp. 248–249).)

The RV of the pre-averaged signal then behaves as follows as $n \to \infty$ (*ibid*, Theorem 5, p. 249):

$$\sum_i (\Delta \bar{X}_{i+1})^2 \approx \sum_i \sigma_{\tau_{i-1}}^2 \Delta \tau_i E(2I_i^2 - 2I_i + 1) \tag{9}$$

$$\text{(good news for regular times:)} \xrightarrow{p} \frac{2}{3} \int_0^T \sigma_t^2 dt \tag{10}$$

$$\text{(bad news for general times:)} \xrightarrow{p} \text{limit depends on spacings, or may not even exist} \tag{11}$$

The pre-averaged RV (the sum of squared pre-averaged returns) $\sum_i (\Delta \bar{Y}_{i+1})^2$ therefore depends on $E(I_i)$ and $E(I_i^2)$, and cannot obviously be parlayed into an estimator of the volatility of X. For example, if the ϵ_j are iid and with mean zero, and if the $M_{n,i} \equiv M_n$, one obtains $\sum_i (\Delta \bar{Y}_{i+1})^2 = \sum_i (\Delta \bar{X}_{i+1})^2 + \mathrm{Var}(\epsilon)/M_n$, which will only converge appropriately if the times are regular. In the case of equidistant times, the factor 2/3 goes back to Jacod et al. (2009a). For a discussion of the intuition, see Mykland and Zhang (2017, Corollary 1 (p. 204) and Remark 3 (p. 204–205)).

Trading times are, however, typically irregular. We illustrate this with Table 1 and Fig. 1. One therefore needs more than pre-averaging to estimate volatility.

3. Two-scales estimation to the rescue: finite sample representations

In the following, we shall see that the problem from Eq. (11) can be avoided by using a two scales construction. This relies on algebraic strong representation, and is robust to virtually all scenarios of things that can go wrong. We consider two cases: an original smoothed TSRV (S-TSRV) (Section 3.1, and in particular Fact 2), and a tapered version of the S-TSRV (Section 3.2, and in particular Theorem 1). The rest of the paper is mainly concerned with the latter (modified) estimator, but most of the results carry over to the original estimator from Section 3.1. The two estimators differ only in their edge effects,

Table 1
Irregularity of trading times. The fluctuation of $E(I_i)$, $E(I_i^2)$, and $E(2I_i^2-2I_i+1)$ for the S&P E-mini future as traded on the Chicago Mercantile Exchange, over 1620 bins of 15 s each during the day of May 1, 2007. Note in particular that for $E(2I_i^2-2I_i+1)$, the value 2/3 (from (8)) is only about the first quantile.

	Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
$E(I_i)$	0.0000	0.3621	0.4511	0.4476	0.5364	0.8685
$E(I_i^2)$	0.0000	0.2159	0.2984	0.3066	0.3865	0.7958
$E(2I_i^2-2I_i+1)$	0.5293	**0.6715**	0.7114	0.7184	0.7583	1.0000

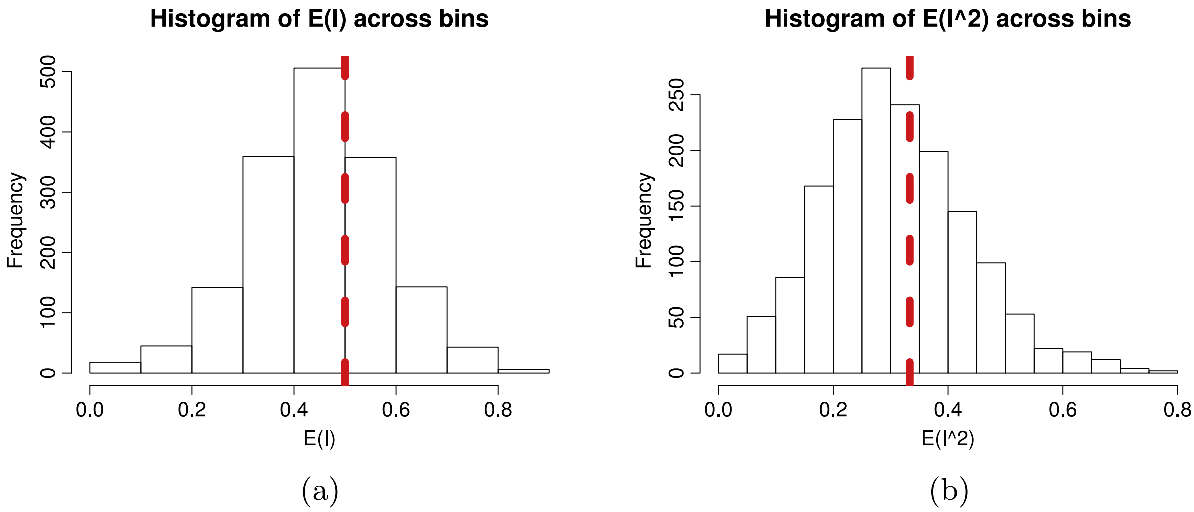

Fig. 1. Irregularity of trading times. The fluctuation of $E(I_i)$ and $E(I_i^2)$ for the S&P E-mini future as traded on the Chicago Mercantile Exchange, over 1620 bins of 15 s each during the day of May 1, 2007. The red dashed vertical lines represent the values for regular times.

and these effects are quantified for both estimators in Proposition 1 in Section 4.1. Note that by choosing block size $M_{n,i}\equiv 1$, our finding also covers the original TSRV as a special case.

We assume the following.

Condition 1. *There are n observations, of the form $Y_{t_{n,j}} = X_{t_{n,j}} + \epsilon_{n,t_{n,j}}$, where X_t is a square integrable martingale (which is right continuous and with left limits), adapted to a history of events (filtration) $(\mathcal{F}_t)$. The observation times $t_{n,j}$ and the block separation times $\tau_{n,i}$ are $(\mathcal{F}_t)$-stopping times. For each (n,j), the noise $\epsilon_{n,t_{n,j}}$ is observed at time $t_{n,j}$ (i.e., is $\mathcal{F}_{t_{n,j}}$-measurable), and* $\sup_{n,j} E\epsilon^2_{n,t_{n,j}} < \infty$*, and $E\epsilon_{n,t_{n,j}} = 0$. In the preceding, the signal X_t may not depend on n.*

The martingale condition on the signal is purely a notational convenience. In almost all circumstances, one can start with a semi-martingale X_t under an original data generating probability distribution Q, and pass to an equivalent (mutually absolutely continuous) probability P under which X_t is a martingale. The only further condition needed is then that Q and P may not depend on n.[3] The equivalent measure device is standard (not only in finance but also in econometrics), and it facilitates many inference arguments. This is because measure change commutes with stable convergence, cf. Mykland and Zhang (2009, Section 2.2). Stable convergence is defined in Footnote 4 of the current paper. In the simple setup where the noise ϵ_{n,t_j} is independent of the efficient signal X_t, the distributions Q and P are the same for the noise, and the measure change only happens for the X_t process. At the cost of further notation, our assumptions can be generalized to processes that can be localized to have the behavior in Condition 1, cf. Jacod and Protter (2012, Ch. 4.4.1, pp. 114–121) and Mykland and Zhang (2012, Ch. 2.4.4–2.4.5, pp. 156–161).

We shall also use the following concept.

Definition 1. A *martingale U-statistic* is a sum of the form $\sum_{i=m_1+m_2+1}^{N}(L_i^{(1)}-L_{i-m_1}^{(1)})(L_{i-m_1}^{(2)}-L_{i-(m_1+m_2)}^{(2)})$, where $L_i^{(1)}$ and $L_i^{(2)}$ are zero mean local square integrable martingales under P, and for positive integers m_1 and m_2.

The significance of this concept is that a martingale U-statistic is a term of mean zero which is typically of small order compared to the constituent local martingales. A term of this form usually does not affect consistency, but will often enter the asymptotically mixed normal limit term. This phenomenon will find concrete embodiment in the results in Sections 4–5.

[3] The history of events may also depend on n and take the form $(\mathcal{F}_{n,t})$. In this case, however, X_t needs to be adapted to "core" history of events $(\mathcal{F}_t)$ given by $\mathcal{F}_t = \cap_n \mathcal{F}_{n,t}$, and the (if required) likelihood ratio dQ/dP must be $\mathcal{F}_T$ measurable.

3.1. Algebraic cancellation and strong representation

We first study the standard two scales construction based on averages. Define as usual

$$[\bar{Y},\bar{Y}]^{(K)} = \frac{1}{K}\sum_{i=1}^{N-K}(\bar{Y}_{i+K}-\bar{Y}_i)^2. \tag{12}$$

Set

$$\eta_i = V_i + \bar{\epsilon}_i \text{ and } \eta_i' = V_i' - \bar{\epsilon}_i,\ \text{where } \bar{\epsilon}_i\ =\ \frac{1}{M_i}\sum_{\tau_{i-1}<t_j\le\tau_i}\epsilon_{t_j} \tag{13}$$

and where V_i and V_i' are defined in (4). Observe that

$$\bar{Y}_{i+K} - \bar{Y}_i = \eta_{i+K} + (X_{\tau_{i+K-1}} - X_{\tau_i}) + \eta_i' \tag{14}$$

Hence,

$$K[\bar{Y},\bar{Y}]^{(K)} = \sum_{i=1}^{N-K}\left[(X_{\tau_{i+K-1}}-X_{\tau_i})^2 + \eta_{i+K}^2 + (\eta_i')^2\right] + C_K \tag{15}$$

where C_K are the sum of the cross terms from dissolving the square of (14), given by (A.1) in Appendix A.1. We now get two important facts.

Fact 1. *Assume that $\bar{\epsilon}_{n,i}$ is a martingale difference. Then the cross terms in C_K are martingale U-statistics.* □

Fact 1 follows since $E[\eta_{i+K}|\mathcal{F}_{\tau_{i+K-1}}] = 0$, and since η_i' is $\mathcal{F}_{\tau_i}$-measurable. For consistency, it is the non-martingale terms in (15) that cause trouble. The first order effect of the two scales construction is to remove these terms, up to edge effect. We are also able to relax the conditions on the noise:

Fact 2. *Assume that $E(\bar{\epsilon}_{n,i} \mid \mathcal{F}_{\tau_{i-J}}) = 0$. For $K > J$,*

$$K[\bar{Y},\bar{Y}]^{(K)} - J[\bar{Y},\bar{Y}]^{(J)} = \sum_{i=1}^{N-K}(X_{\tau_{i+K-1}}-X_{\tau_i})^2 - \sum_{i=1}^{N-J}(X_{\tau_{i+J-1}}-X_{\tau_i})^2 + C_{K,J} + e_{K,J}, \tag{16}$$

where $C_{K,J}$ are cross terms that are martingale U-statistics, given by (A.2), and where $e_{K,J}$ is an edge effect, given by (A.3), which is normally (but not always) negligible; see Section 4.1 for a precise discussion. □

In other words, the two scales construction has removed the effect of the observation times from the main effect in (16), and this effect has been relegated to the asymptotic variance and the edge effect. The construction also allows noise averages $\bar{\epsilon}_i$ to be an m-dependent sequence, with $m \le J-1$. If instead $\bar{\epsilon}_i$ is α-mixing, related results can be established with suitable modification (an o_p term), as in Aït-Sahalia et al. (2011) and Zhang (2011).

As discussed in Appendix A.2, the edge effect can be partially offset (in expectation only, and under strong assumptions) by normalizing as follows

$$\widehat{\langle X,X\rangle}\ =\ \frac{1}{\left(1-\frac{K+J-\frac{1}{3}}{N}\right)(K-J)}\left\{K[\bar{Y},\bar{Y}]^{(K)} - J[\bar{Y},\bar{Y}]^{(J)}\right\}. \tag{17}$$

In particular, when $J = 1$ and $K = 2$, up to martingale U-statistics and edge effect,

$$\widehat{\langle X,X\rangle}\ \approx\ \text{the realized volatility } of\ the\ signal: \sum_{i=1}^{N-2}(X_{\tau_{i+1}}-X_{\tau_i})^2. \tag{18}$$

We shall in the following work with a modified estimator, but most of the results carry over to the estimator above. The two estimators differ in their edge effects, but not in the respective asymptotic behavior of their martingale U-statistics. We shall refer to both estimators as smoothed two-scales realized volatility (S-TSRV).

3.2. Getting rid of the edge effect from noise: an estimator for the very cautious

The estimator (17) retains a small amount of edge effect, as described in Appendix A. There are two components to this effect, one relating to σ_t^2, and one to $\bar{\epsilon}_i^2$. There does not seem to be a way of eliminating both these components, but one can be eliminated at the expense of the other.

We here take the view that the edge effect of noise $\bar{\epsilon}_i^2$ is the most concerning. While both types of edge effect are asymptotically negligible in most models for the data generating process, it is possible to construct scenarios where the noise effect matters asymptotically, see Kalnina and Linton (2008), and also the discussion in Section 4.1. On the other hand, while the spot volatility σ_t^2 can evolve fast, it has higher persistence, and one can thus live with some edge effect so long as it is understood and controlled.

In other words, we propose to eliminate the edge effect in $\bar{\epsilon}_i^2$, and pay the price that there may be a slightly higher such effect in the volatility. In Appendix B, we show that the following estimator is completely free of terms of type $\bar{\epsilon}_i^2$. For a pair (J, K), set

$$K\widetilde{[\bar{Y},\bar{Y}]}^{(K)} = \frac{1}{2}\sum_{i=1}^{J}(\bar{Y}_{i+K}-\bar{Y}_i)^2 + \sum_{i=J+1}^{N-b}(\bar{Y}_{i+K}-\bar{Y}_i)^2 + \frac{1}{2}\sum_{i=N-b+1}^{N-K}(\bar{Y}_{i+K}-\bar{Y}_i)^2. \tag{19}$$

where

$$b = K + J. \tag{20}$$

We define $J\widetilde{[\bar{Y},\bar{Y}]}^{(J)}$ similarly by switching J and K. (This is the same as Eq. (B.13) in case this is unclear. An alternative representation is given by (B.17).)

We obtain in Appendix B that $K\widetilde{[\bar{Y},\bar{Y}]}^{(K)} - J\widetilde{[\bar{Y},\bar{Y}]}^{(J)}$ has no edge effect in $\bar{\epsilon}^2$ terms. Also, the times do not affect the estimator to first order. Following (B.20), the exact result is:

Theorem 1 (*Algebraic Representation of Two Scales Combination*)**.** *Assume $K > J \geq 1$, and that $E(\bar{\epsilon}_{n,i} \mid \mathcal{F}_{\tau_{i-J}}) = 0$. Then*

$$\begin{aligned} K\widetilde{[\bar{Y},\bar{Y}]}^{(K)} - J\widetilde{[\bar{Y},\bar{Y}]}^{(J)} &= \left\{\frac{1}{2}\sum_{i=1}^{J}(X_{\tau_{i+K-1}}-X_{\tau_i})^2 + \sum_{i=J+1}^{N-b}(X_{\tau_{i+K-1}}-X_{\tau_i})^2 + \frac{1}{2}\sum_{i=N-b+1}^{N-K}(X_{\tau_{i+K-1}}-X_{\tau_i})^2\right\} \\ &- \left\{\frac{1}{2}\sum_{i=1}^{K}(X_{\tau_{i+J-1}}-X_{\tau_i})^2 + \sum_{i=K+1}^{N-b}(X_{\tau_{i+J-1}}-X_{\tau_i})^2 + \frac{1}{2}\sum_{i=N-b+1}^{N-J}(X_{\tau_{i+J-1}}-X_{\tau_i})^2\right\} \\ &+ \textit{cross terms} + \textit{edge term}, \end{aligned} \tag{21}$$

where the cross terms are martingale U-statistics and are given as $\tilde{C}_{K,J}$ in (B.18)–(B.19), and the edge term is given by

$$\tilde{e}_{K,J} = \left(-\sum_{i=J+1}^{K} + \sum_{i=N-K+1}^{N-J}\right)\left(\frac{1}{2}(\eta_i - \eta_i')\Delta X_{\tau_i} + \eta_i(X_{\tau_{i-1}} - X_{\tau_{i-J}})\right). \tag{22}$$

Following further derivation in Appendix B, we propose a normalization so that the final estimator is

$$\widehat{\langle X, X\rangle} = \frac{1}{(1-b/N)(K-J)}\left\{K\widetilde{[\bar{Y},\bar{Y}]}^{(K)} - J\widetilde{[\bar{Y},\bar{Y}]}^{(J)}\right\}. \tag{23}$$

From (22), we shall see in Section 4.1 that the edge effect is negligible except under highly unusual combinations of K and smoothing parameters $M_{n,i}$. This substantially limits the effect of mis-recording of the observation times on the estimator.

A version of this estimator (with $J = 1$ and no pre-averaging) was proposed in Kalnina and Linton (2008) as a remedy for non-stationary noise. They carried out an asymptotic analysis. The current results show that the estimator is robust to very general forms of noise, and also that the desirable properties hold in a small sample setting.

3.3. *The estimator is close to a $K - J$ single scale estimator in X*

One can go one step further, as shown in Appendix C:

Theorem 2. *The squared terms (the first two terms on the r.h.s.) in (21) are equal to*

$$\sum_{i=J+1}^{N-K-1}(X_{\tau_{i+K-J}}-X_{\tau_i})^2 + \frac{1}{2}(X_{\tau_K}-X_{\tau_J})^2 + \frac{1}{2}(X_{\tau_{N-J}}-X_{\tau_{N-K}})^2 + \textit{cross terms}, \tag{24}$$

where the cross terms are martingale U-statistics. The cross terms are explicitly given in Eqs. (C.23) and (C.26).

One can alternatively write the squared terms in (24) as

$$\sum_{i=J}^{N-K}(X_{\tau_{i+K-J}}-X_{\tau_i})^2 - \frac{1}{2}(X_{\tau_K}-X_{\tau_J})^2 - \frac{1}{2}(X_{\tau_{N-J}}-X_{\tau_{N-K}})^2. \tag{25}$$

The whole estimator (23) now gets the form

$$\widehat{\langle X,X\rangle} = \frac{1}{(1-b/N)(K-J)}\left\{\sum_{i=J+1}^{N-K-1}(X_{\tau_{i+K-J}}-X_{\tau_i})^2+\frac{1}{2}(X_{\tau_K}-X_{\tau_J})^2+\frac{1}{2}(X_{\tau_{N-J}}-X_{\tau_{N-K}})^2\right.$$
$$\left. + \text{ the edge term from (22) } + \text{ the cross terms from (B.19), (C.23), and (C.26).}\right\} \quad (26)$$

4. Asymptotic representation of the two scales estimator

The preceding has been concerned with exact properties. We here show how the various edge terms and martingale U-statistics vanish asymptotically.

Condition 2 (*Structure of the Efficient Price*). *Assume that $X_t = X_t^{(1)} + X_t^{(2)}$, where $X_t^{(2)}$ is a pure jump martingale with finitely many jumps, all at predictable times, and where $X_t^{(1)}$ is a square integrable martingale whose predictable quadratic variation $\langle X^{(1)}, X^{(1)}\rangle_t$ is continuously differentiable.*

This condition covers most cases that are studied, and for $X_t^{(1)}$ see Jacod and Protter (2012, Definition 2.1.1. and Eqs. (2.1.15) and (2.1.31), pp. 35–39). For a continuous process, $d\langle X^{(1)}, X^{(1)}\rangle_t/dt = \sigma_t^2$.

To put asymptotic orders in context, we note that a slight extension of Theorems 2 and 3 (p. 1401) of Zhang et al. (2005) yields that

$$\frac{1}{K-J}\sum_{i=J+1}^{N-K}(X_{\tau_{i+K-J}}-X_{\tau_i})^2-[X,X]_T = O_p((\Delta\tau_n^+(K-J))^{1/2}), \quad (27)$$

where $[X,X]_T$ is the quadratic variation of X_t. If X is continuous, then $[X,X]_T = \int_0^T \sigma_t^2 dt$. Quantities of smaller order are thus negligible.

We now look at the block and noise structure. We note that the blocks are fixed by the econometrician.

Condition 3 (*Structure of Blocks*). *We assume that for each n, there are nonrandom $\Delta\tau_n^+$ and $M_n^- \geq 1$, so that $\Delta\tau_n^+ \geq \max_i \Delta\tau_{n,i}$ and $M_n^- \leq \min_i M_{n,i}$. Also assume that $K_n\Delta\tau_n^+ \to 0$ as $n\to\infty$, and that $K > J \geq 1$.*

The nonrandomness condition is for notational convenience. If the $\Delta\tau_{n,i}$'s and $M_{n,i}$'s are nonrandom, we can set $\Delta\tau^+ = \max_i \Delta\tau_{n,i}$ and $M_n^- = \min_i M_{n,i}$. Note that under this condition,

$$\frac{T}{\Delta\tau_n^+} \leq N_n \leq \frac{n}{M_n^-}. \quad (28)$$

Also, it will normally be the case that $\Delta\tau_n^+ \propto M_n^-/n$, in which case the number of blocks $N = N_n$ is of exact order $O(n/M_n^-)$.

4.1. Negligibility of edge effects

Proposition 1. *Assume Condition 1–3, and also that $\mathrm{Var}(\bar\epsilon_{n,i}) = O(M_{n,i}^{-1})$, uniformly in (n,i). Then*

$$\frac{e_{K,J}}{K-J} = O_p\left(J^{1/2}(\Delta\tau_n^+ + (M_n^-)^{-1})\right) \text{ and}$$
$$\frac{\tilde e_{K,J}}{K-J} = O_p\left(J^{1/2}(\Delta\tau_n^+ + (M_n^-)^{-1})^{1/2}(\Delta\tau_n^+)^{1/2}\right) \quad (29)$$

The result is shown at the end of Appendix A.1. When comparing (29) to the order in (27), we note that $\tilde e_{K,J}$ is asymptotically negligible unless M_n^- and K are both chosen to be asymptotically finite, which does not yield a consistent estimator of volatility. (Both the block sizes and K are under the control of the econometrician).

Meanwhile, $e_{K,J}$ is asymptotically negligible if $(\Delta\tau_n^+)^{-1}(M_n^-)^{-2} = o((K-J)/J)$. This condition can fail to be satisfied, which is consistent with the results of Kalnina and Linton (2008). Observe, however, that $e_{K,J}$ will be negligible with only moderate pre-averaging. When $\Delta\tau_n^+ \propto M_n^-/n$, $e_{K,J}$ is asymptotically negligible if $M_n^- = O(n^{1/3})$ when $K\to\infty$ with n, and if $n^{-1/3}M_n^- \to \infty$ with n when K is asymptotically finite. Also, this is a worst case scenario, and the edge effect will disappear if the noise is homoscedastic and independent of the latent process X_t.

The condition $\mathrm{Var}(\bar\epsilon_{n,i}) = O(M_{n,i}^{-1})$ is a CLT style requirement which is assured under martingale, Markov, or mixing assumptions (Hall and Heyde, 1980; Nummelin, 1984).

4.2. Asymptotic representation

We here work with the assumption that $K-J\to\infty$, to be able to give a sense of what are main terms. This clarifies that to first order, there is no effect of the observation times, either on the quantity being estimated, *or on the asymptotic variance*.

The development also clarifies what are the main terms to handle to characterize the estimation error of the S-TSRV estimator: the two terms on the first line in (30). We resume this discussion in Section 4.3.

The implications of requiring $K-J\to\infty$ are discussed at the end of Section 5.2. Robustness to finite K is considered in Section 6. Also, J may or may not go to infinity with N.

To state the asymptotic representation, we also need more assumptions on the noise:

Condition 4 (*Structure of the Noise*)**.** *Assume that $E\epsilon_{n,t_{n,j}}=0$ and $\sup_{n,j}E\epsilon^2_{n,t_{n,j}}<\infty$ (as in Condition 1). Also assume that $E(\bar\epsilon_{n,i}\mid\mathcal{F}_{\tau_{i-J}})=0$ (as in Fact 2 and Theorem 1), and that $E\sup_i E(\bar\epsilon^2_{n,i}\mid\mathcal{F}_{\tau_{i-J}})=o_p(\Delta\tau_n^+(K-J)^{1/2})$.*

An important feature of Condition 4 is that the noise and the process may be dependent on each other. As an extreme example, one can have the noise reflect short term leads and lags. For instance, the condition permits $\epsilon_{t_{n,j}}=X_{t_{n,j}}-X_{t_{n,j-p}}+$ m-dependent independent noise, where p and m may vary from dimension to dimension (for the case where X is multi-dimensional).

In order to capture the dependent noise possibility without too much notation, we have not sought to optimize Condition 4 in other respects. If one assumes that the ϵ_{n,t_j} are independent of X and the times t_j and τ_i, then Condition 4 may be replaced by the assumption that the ϵ_{n,t_j} are exponentially α-mixing, with uniformly bounded fourth moment. See, e.g., McLeish (1975) and Hall and Heyde (1980), cf. also Aït-Sahalia et al. (2011) and Zhang (2011) in connection with noise in high frequency observations. Extra conditions on the X process may also be required, cf. Jacod and Protter (2012, Chapter 2.1.5, pp. 39–44).

Theorem 3 (*Asymptotic Representation of the Two Scales Estimator*)**.** *Assume Condition 1–4. Also suppose that $K-J$, M_n^-, and N_n all tend to infinity with n, and that $N_n\Delta\tau_n^+=O(1)$. Then*

$$\widehat{\langle X,X\rangle}=\frac{1}{(1-b/N)(K-J)}\left\{\sum_{i=J+1}^{N-K}(X_{\tau_{i+K-J}}-X_{\tau_i})^2+\textit{ noise U-statistics}\right\}$$
$$+\,o_p((\Delta\tau^+(K-J))^{1/2})+O_p\left((\Delta\tau^+)^{1/2}/(K^{1/2}+J^{1/2})\times\textit{ noise U-statistics}\right)\tag{30}$$

where

$$\textit{noise U-statistics}=2\left(\frac12\sum_{i=1}^{b-K}+\sum_{i=b-K+1}^{N-b}+\frac12\sum_{i=N-b+1}^{N-K}\right)\bar\epsilon_{i+K}\bar\epsilon_i$$
$$-\,2\left(\frac12\sum_{i=1}^{b-J}+\sum_{i=b-J+1}^{N-b}+\frac12\sum_{i=N-b+1}^{N-J}\right)\bar\epsilon_{i+J}\bar\epsilon_i.\tag{31}$$

We have here seen that the estimator mainly depends on the two terms on the first line in (30): the average subsampled (at the τ_i's) realized variance (RV) $(K-J)^{-1}\sum_{i=J+1}^{N-K}(X_{\tau_{i+K-J}}-X_{\tau_i})^2$, and the noise U-statistic. The latter is easily handled with the help of Condition 4.

4.3. The final piece: discretization error

From Theorem 3, it remains to analyze the average subsampled RV, *i.e.*, the discretization error

$$\frac{1}{(1-b/N)(K-J)}\sum_{i=J+1}^{N-K}(X_{\tau_{i+K-J}}-X_{\tau_i})^2-\langle X,X\rangle.\tag{32}$$

It is not the purpose of this paper to provide a most general Central Limit Theorem (CLT) for this piece, since at the time of writing it is difficult to foresee all contingencies. In a certain sense, the representation in Theorem 3 *is* the general result, from which CLTs for $\widehat{\langle X,X\rangle}$ can be harvested for a range of situations. Theorem 3 clarifies that controlling (32) is the only remaining piece in the CLT problem, and that this does not involve the microstructure noise.

The analysis of (32) can be done with a variety of existing approaches.

1. Under continuity of X and exogenous τ_i, asymptotics follow from Theorems 2 and 3 (p. 1401) of Zhang et al. (2005). As proof of concept, we have generalized this theory to covariance matrices in Sections 5.2 and 6, including the spot case in Section 6.2.

2. If the τ_i are endogenous, one can use the approach from Li et al. (2014). For the purposes of this paper, it is the τ_is that would then be endogenous. This would not occur if $\Delta\tau_i$ is constant in clock time, but may be the case is the block size $M_{n,i}$ is constant.
3. When there are jumps, one can go to Chapter 16.6 of Jacod and Protter (2012). If there are jumps and endogeneity of times, one may be able to combine the two latter directions of study.

The clear separation of noise and signal also points to situations where the central limit theory for the signal does not yet exist (as far as we know). For example, if jump times are infinitely many, and endogenous, the X process may not follow the Itô semimartingale form which underlies many of the CLTs that currently exist (see, *e.g.*, Jacod and Protter (2012, Assumption (H), p. 126).) This is because stopping times due to jumps in Itô semimartingales are *totally inaccessible* (Jacod and Shiryaev, 2003, Definition 2.20, p. 29), which may collide with endogeneity.

In view of the spot volatility discussion in Section 6.2, the results in this section also have bearing on, for example, the estimation of functionals of volatility (Jacod and Rosenbaum, 2013, 2015).

5. Application: estimation of realized covariances, and a central limit theorem

5.1. Definitions and asymptotic representation

As an application of the above, we consider the multivariate case. We here assume that the τ_is are synchronous, which is plausible in many cases, since they are under the control of the econometrician.

By much the same analysis as above, we obtain the estimator

$$\widehat{\langle X^{(r)}, X^{(s)}\rangle} = \frac{1}{(1-b/N)(K-J)}\left\{K\widetilde{[\bar{Y}^{(r)}, \bar{Y}^{(s)}]}^{(K)} - J\widetilde{[\bar{Y}^{(r)}, \bar{Y}^{(s)}]}^{(J)}\right\}. \tag{33}$$

Here, $X^{(r)}$ is component # r in vector X. The rest of the nomenclature extends similarly. The I_i variable defined in Section 2 also becomes a vector, and the probability distribution is then defined in obvious fashion on a hyper-cube of the same dimension as X.

The finite representation results from the scalar case continue to hold, but the discussion would be tedious. For clarity, however, we provide a multivariate extension of Theorem 3. The proof is similar.

Theorem 4 (*Asymptotic Representation of the Two Scales Estimator*)**.** *Under the same conditions as Theorem 3, but extended to the multivariate case (X_t is a vector, etc.), the following is valid.*

$$\widehat{\langle X^{(r)}, X^{(s)}\rangle} = \frac{1}{(1-b/N)(K-J)}\left\{\sum_{i=J+1}^{N-K}(X^{(r)}_{\tau_{i+K-J}} - X^{(r)}_{\tau_i})(X^{(s)}_{\tau_{i+K-J}} - X^{(s)}_{\tau_i}) + \text{ noise U-statistics}\right\}$$
$$+\ o_p((\Delta\tau^+(K-J))^{1/2}) + O_p\left((\Delta\tau^+)^{1/2}/(K^{1/2}+J^{1/2}) \times \text{ noise U-statistics}\right) \tag{34}$$

where

$$\text{noise U-statistics} = \left(\frac{1}{2}\sum_{i=1}^{b-K} + \sum_{i=b-K+1}^{N-b} + \frac{1}{2}\sum_{i=N-b+1}^{N-K}\right)\bar{\epsilon}^{(r)}_{i+K}\bar{\epsilon}^{(s)}_i[2]$$
$$- \left(\frac{1}{2}\sum_{i=1}^{b-J} + \sum_{i=b-J+1}^{N-b} + \frac{1}{2}\sum_{i=N-b+1}^{N-J}\right)\bar{\epsilon}^{(r)}_{i+J}\bar{\epsilon}^{(s)}_i[2]. \tag{35}$$

The notation $\bar{\epsilon}^{(r)}_{i+K}\bar{\epsilon}^{(s)}_i[2]$ means $\bar{\epsilon}^{(r)}_{i+K}\bar{\epsilon}^{(s)}_i + \bar{\epsilon}^{(s)}_{i+K}\bar{\epsilon}^{(r)}_i$, see, for example McCullagh (1987).

5.2. Central limit theory

To provide a simple application, we here state and show a CLT in the multivariate case under (by now) classical conditions. In particular, the ϵ are independent of X and the times t_j and τ_i, and the latter are exogenous.

The two main terms in (34) are then asymptotically independent. For the signal term, the univariate central limit theorem reduces to Theorem 3 (p. 1401) of Zhang et al. (2005). In the multivariate case, we obtain as follows. First the signal term, which works the same way as in the earlier paper.

Theorem 5 (*CLT for Signal Term*)**.** *Assume the conditions of Theorem 4. Also assume that X_t is a continuous martingale, with locally bounded spot volatility σ_t. Assume that the τ's are exogenous. Define the matrix*

$$D^{(r,s)} = \frac{1}{(1-b/N)(K-J)}\sum_{i=J+1}^{N-K}(X^{(r)}_{\tau_{i+K-J}} - X^{(r)}_{\tau_i})(X^{(s)}_{\tau_{i+K-J}} - X^{(s)}_{\tau_i}) - \langle X^{(r)}, X^{(s)}\rangle_T \tag{36}$$

If the τ's are equidistant, define $G(t) = 4t/3$, and otherwise define it using Eq. (44) (p. 1401) and (A.33)-(A.34) (p. 1411) of Zhang et al. (2005). Then $(N/(K-J))^{1/2}D^{(r,s)}$ converges stably in law[4] *to a normal distribution with mean zero and covariance tensor*

$$ACOV(D^{(r_1,s_1)}, D^{(r_2,s_2)}) = \frac{1}{4}T\int_0^T \langle X^{(r_1)}, X^{(r_2)}\rangle_t' \langle X^{(s_1)}, X^{(s_2)}\rangle_t' dG(t)[2][2], \tag{37}$$

where the "[2][2]" means summation over four terms where r_1 and r_2 can change place with s_1 and s_2. In other words, $a^{(r_1,r_2)}a^{(s_1,s_2)}[2][2] = a^{(r_1,r_2)}a^{(s_1,s_2)} + a^{(s_1,r_2)}a^{(r_1,s_2)} + a^{(r_1,s_2)}a^{(s_1,r_2)} + a^{(s_1,s_2)}a^{(r_1,r_2)}$.

The noise term in (34) is more complex since we have allowed for the possibility of pre-averaging. There is an averaging across $M_{n,i}^{(r)}$ observations of interval #i. This alters variance calculations and convergence rates. Normality as such is straightforward given the short run dependence of the ϵ's.

Denote

$$C_K^{(r,s)} = \frac{1}{(1-b/N)(K-J)}\left(\frac{1}{2}\sum_{i=1}^{b-K} + \sum_{i=b-K+1}^{N-b} + \frac{1}{2}\sum_{i=N-b+1}^{N-K}\right)\bar{\epsilon}_{i+K}^{(r)}\bar{\epsilon}_i^{(s)}[2]$$

$$C_J^{(r,s)} = \frac{1}{(1-b/N)(K-J)}\left(\frac{1}{2}\sum_{i=1}^{b-J} + \sum_{i=b-J+1}^{N-b} + \frac{1}{2}\sum_{i=N-b+1}^{N-J}\right)\bar{\epsilon}_{i+J}^{(r)}\bar{\epsilon}_i^{(s)}[2] \tag{38}$$

and assume that $J_nM_n^-$ is large enough to make $\bar{\epsilon}_i^{(s)}$ and $\bar{\epsilon}_{i+J}^{(r)}$ independent under Condition 4. In this case (which is simplified), $\mathrm{Cov}(C_K^{(r_1,s_1)}, C_J^{(r_2,s_2)}) = 0$, and

$$\begin{aligned}
&\mathrm{Cov}(C_K^{(r_1,s_1)}, C_K^{(r_2,s_2)})\\
&= \left(\frac{1}{(1-b/N)(K-J)}\right)^2\left(\frac{1}{4}\sum_{i=1}^{b-K} + \sum_{i=b-K+1}^{N-b} + \frac{1}{4}\sum_{i=N-b+1}^{N-K}\right)\mathrm{Cov}(\bar{\epsilon}_{i+K}^{(r_1)}\bar{\epsilon}_i^{(s_1)}, \bar{\epsilon}_{i+K}^{(r_2)}\bar{\epsilon}_i^{(s_2)})[2][2]\\
&= \left(\frac{1}{(1-b/N)(K-J)}\right)^2\left(\frac{1}{4}\sum_{i=1}^{b-K} + \sum_{i=b-K+1}^{N-b} + \frac{1}{4}\sum_{i=N-b+1}^{N-K}\right)\mathrm{Cov}(\bar{\epsilon}_{i+K}^{(r_1)}, \bar{\epsilon}_{i+K}^{(r_2)})\mathrm{Cov}(\bar{\epsilon}_i^{(s_1)}, \bar{\epsilon}_i^{(s_2)})[2][2],
\end{aligned} \tag{39}$$

where "[2][2]" has the same meaning as in Theorem 5.

To get a handle on rates, let us suppose that $M_{n,i}^{(r)}$ only depends on n, $= M_n$. Then $\mathrm{Cov}(\bar{\epsilon}_i^{(s_1)}, \bar{\epsilon}_i^{(s_2)}) = O(M_n^{-1})$. If we suppose further that there is stationarity enough to assure $\mathrm{Cov}(\bar{\epsilon}_i^{(s_1)}, \bar{\epsilon}_i^{(s_2)}) = M_n^{-1}c^{(s_1,s_2)}$, then

$$\begin{aligned}
\mathrm{Cov}(C_K^{(r_1,s_1)}, C_K^{(r_2,s_2)}) &= \left(\frac{1}{(1-b/N)(K-J)}\right)^2\left(N-K-\frac{3}{2}J\right)M_n^{-2}c^{(r_1,r_2)}c^{(s_1,s_2)}[2][2]\\
&\sim \frac{N}{(K-J)^2M_n^2}c^{(r_1,r_2)}c^{(s_1,s_2)}[2][2].
\end{aligned} \tag{40}$$

Similar expressions hold for $\mathrm{Cov}(C_J^{(r_1,s_1)}, C_J^{(r_2,s_2)})$.

In other words, the noise term has order $O_p\left(\frac{N^{1/2}}{(K-J)M_n}\right)$. Meanwhile, the signal term has order $O_p\left((N/(K-J))^{-1/2}\right)$. Equality between these two orders is achieved by setting

$$K - J = O_p\left((N/M_n)^{2/3}\right). \tag{41}$$

The order of convergence of the estimator thus becomes $O_p\left(N^{-1/6}M_n^{-1/3}\right) = O_p\left(n^{-1/6}M_n^{-1/6}\right)$, since $N = n/M_n$ in this case. We can thus get arbitrarily close to the optimal $O_p(n^{-1/4})$ rate. We can also achieve this rate, by setting $M_n = O(n^{1/2})$, but asymptotic expressions become more complicated as, in this case, $K - J$ stays finite. Our finite sample calculations, however, remain valid also for this case. We shall see this in the next section.

6. Asymptotic representation and normality: robustness to finite *K* and shrinking *T*

6.1. Estimation of integrated volatility

As we have seen at the end of the previous section, it can be optimal to choose K to be finite. Also, over recent years, as trading becomes more frequent and markets become more liquid, the size of the microstructure noise appears to be

[4] Let Z_n be a sequence of $\mathcal{F}_T$-measurable random variables. We say that Z_n *converges stably in law to* Z as $n \to \infty$ if Z is measurable with respect to an extension of $\mathcal{F}_T$ so that for all $A \in \mathcal{F}_T$ and for all bounded continuous g, $EI_Ag(Z_n) \to EI_Ag(Z)$ as $n \to \infty$. I_A denotes the indicator function of A, and $= 1$ if A and $= 0$ otherwise. The same definition applies to triangular arrays. In the context of inference, $Z_n = n^{1/2}(\hat{\theta}_n - \theta)$, for example, and $Z = N(b, a^2)$. For further discussion of stable convergence, and for the relationship to measure change, see Section 2.2 of Mykland and Zhang (2009), which draws on Rootzén (1980).

declining. This also points to choosing K to be finite in the two scales estimator. In addition, when estimating instantaneous volatility, or if time periods are small, one may wish to consider the possibility that $T \to 0$ with increasing data.

The expressions for variance terms in the finite K case are quite complicated. Standard errors are conjectured to be most easily set using observed asymptotic variances (Mykland and Zhang, 2017).

First, however, a result to provide the representation and limit properties in this case. Note that (27) remains valid, but with the modification

$$\frac{1}{K-J}\sum_{i=J+1}^{N-K}(X_{\tau_{i+K-J}}-X_{\tau_i})^2-\int_0^T\sigma_t^2dt=O_p\left(T\left(\frac{K-J}{N}\right)^{1/2}\right) \tag{42}$$

It is straightforward, by the same methods as above, to see that the other signal terms in $\widehat{\langle X^{(r)},X^{(s)}\rangle}$ are of at most this order. Meanwhile, the main noise U-statistic (normalized by $K-J$) is of order $O_p(N^{1/2}/\bar{M}_n(K-J))$, where $\bar{M}_n$ is the average value of $M_{n,i}$. Observe that by definition, $\bar{M}_nN = n$ (the total number of observations in $[0,T]$). Equating these two orders,[5] we obtain $K-J=O_p\left((N/T\bar{M}_n)^{2/3}\right)$ in generalization of (41), or

$$K-J=O\left((n/TM_n^2)^{2/3}\right)=O\left((\overline{\Delta t}M_n^2)^{-2/3}\right), \tag{43}$$

where the average sampling frequency $\overline{\Delta t}=T/n$ is used to represent the number of observations. This provides more transparency since n will normally depend on T, which we here do not assume to be fixed.

Theorem 6 (*Asymptotic Representation of the Two Scales Estimator*)**.** *Assume the conditions of Theorem 5, except that now K may be finite. Also, $T=O(1)$ or smaller. Assume that $\max_i M_{n,i}=O_p(\bar{M}_n)$, as well as (43). Finally, assume that the $\bar{\epsilon}_i^{(r)}$ process is stationary. Then*

$$\widehat{\langle X^{(r)},X^{(s)}\rangle}=\frac{1}{(1-b/N)(K-J)}\left\{\sum_{i=J+1}^{N-K}(X^{(r)}_{\tau_{i+K-J}}-X^{(r)}_{\tau_i})(X^{(s)}_{\tau_{i+K-J}}-X^{(s)}_{\tau_i})+\text{ cross terms}\right\}$$
$$+\,o_p(((K-J)\Delta\tau^+)^{1/2}) \tag{44}$$

where the cross terms are martingale U-statistics of order $O_p(c_n)$, where $c_n = T^{1/2}\overline{\Delta t}^{1/6}M_n^{-1/6}$. Also the cross terms are asymptotically stably normal when normalized by c_n^{-1}. In particular,

$$c_n^{-1}\left(\widehat{\langle X^{(r)},X^{(s)}\rangle}-\langle X^{(r)},X^{(s)}\rangle\right) \tag{45}$$

converges stably to a normal distribution, with variance that is random but consistently estimable ($\mathcal{F}_T$-measurable).

In analogy with the T fixed case, one can choose M_n to be of order up to $O((\overline{\Delta t})^{-1/2})$ (but if T is shrinking, no longer order $O(n^{1/2})$). The best convergence rate c_n is obtained by choosing M_n to be of exact order $O((\overline{\Delta t})^{-1/2})$, which corresponds to the case of J and K finite, yielding

$$c_n=T^{1/2}\overline{\Delta t}^{1/4}. \tag{46}$$

The proof of Theorem 6 is much the same as for the earlier asymptotic theorems, except that the $O_p(\Delta\tau^+)$ terms must be scrutinized individually for asymptotic normality, and we have invoked stationarity to complete the conditions for normality of the noise term. The condition on $\max_i M_{n,i}$ assures asymptotic negligibility. One can, obviously, go with weaker but more elaborate conditions.

6.2. Estimation of spot volatility

The spot volatility matrix is defined by $(\sigma_t^{r,s})^2=d\langle X^{(r)},X^{(s)}\rangle_t/dt$. The standard estimate of spot volatility is now to use the normalized integrated volatility divided by T, in our current case $\widehat{(\sigma_{\mathcal{T}}^{r,s})^2}=\widehat{\langle X^{(r)},X^{(s)}\rangle}/T$. We have there adopted the convention that $\widehat{\langle X^{(r)},X^{(s)}\rangle}$ is computed over observations in $[\mathcal{T}-T,\mathcal{T}]$. Similarly, we define $\langle X^{(r)},X^{(s)}\rangle=\int_{\mathcal{T}-T}^{\mathcal{T}}(\sigma_t^{r,s})^2dt$.

The error in this estimator is

$$\widehat{(\sigma_{\mathcal{T}}^{r,s})^2}-(\sigma_{\mathcal{T}}^{r,s})^2=\underbrace{\frac{1}{T}\left(\widehat{\langle X^{(r)},X^{(s)}\rangle}-\langle X^{(r)},X^{(s)}\rangle\right)}_{\text{error in estimator of integrated volatility}}+\underbrace{\frac{1}{T}\left(\int_{\mathcal{T}-T}^{\mathcal{T}}(\sigma_t^{r,s})^2dt-T(\sigma_{\mathcal{T}}^{r,s})^2\right)}_{\text{smoothing error}}$$
$$=O_p(c_nT^{-1})+O_p(T^{1/2}). \tag{47}$$

[5] We emphasize that this is not so much a requirement on the tuning parameters $K-J$ as a condition to assure that both the signal and noise terms in the error are represented in the asymptotic distribution. Any other order will lead to an asymptotics that underrepresents the actual variability in the estimator.

The first O_p term in (47) comes from (45) above, while the second O_p term is explained at the end of this section. Once again the error is asymptotically stably normal when suitably normalized.

With c_n from (46), and equating the two error terms in (47), one obtains

$$c_n = \overline{\Delta t}^{1/8},$$

which is the conjectured best rate in this estimation problem. See Mykland and Zhang (2008) end of Section 6, p. 263.

We finally return to the second O_p term in (47). This term comes from

$$\int_{\mathcal{T}-T}^{\mathcal{T}} \sigma_t^2 dt - T\sigma_{\mathcal{T}}^2 = \int_{\mathcal{T}-T}^{\mathcal{T}} (\mathcal{T}-t) d\sigma_t^2 = O_p(T^{3/2}). \tag{48}$$

The first transition in (48) is due to integration by parts (in Itô's formula): $d(\sigma_t^2 - \sigma_{\mathcal{T}}^2)(\mathcal{T}-t) = -(\sigma_t^2 - \sigma_{\mathcal{T}}^2)dt + (\mathcal{T}-t)d\sigma_t^2$, and the second transition comes from assuming that σ_t^2 is an Itô-semimartingale.[6] Observe in particular that the second term in (47) is a variance term, not a bias.

7. Extensions

7.1. A formulation which extends to other problems

The formulation in this paper is in terms of pre-averaging. However, the same treatment extends to cases where observations are of the form (with $\hat{Y}$ replacing $\bar{Y}$)

$$\hat{Y}_{i+K} - \hat{Y}_i = V_{i+K} + (X_{\tau_{i+K-1}} - X_{\tau_i}) + V_i' + \hat{\epsilon}_{i+K} - \hat{\epsilon}_i, \tag{49}$$

where the $\hat{\epsilon}_i$ (the microstructure noise) are at most $J-1$-dependent, while V_i and V_i' are functions of the X process that essentially take place in interval # i, from τ_{i-1} to τ_i. Specifically, for the above arguments to work, we need that V_i and V_i' be $\mathcal{F}_{\tau_i}$-measurable, and also that $E(V_i \mid \mathcal{F}_{\tau_{i-1}}) = E(V_i' \mid \mathcal{F}_{\tau_{i-1}}) = 0$. For example, if one has asynchronous observation, and if the τ_i are a grid as in Zhang (2011), then the main condition for this to work is that there needs to be an observation of each process in each interval. The τ_i's can be picked as refresh times. If $\hat{Y}_i$ is the first observation of a process in interval # i, then the condition $E(V_i \mid \mathcal{F}_{\tau_{i-1}}) = E(V_i' \mid \mathcal{F}_{\tau_{i-1}}) = 0$ is satisfied.

7.2. Rolling windows

Another extension of the above would be to consider rolling windows $(\tau_{i-1}, \tau_i]$. Unless observation times are equidistant, there is not a unique natural way of defining such windows. One option is to require the number of observations M in each window to be constant, another is to let the time length $\Delta\tau$ be constant. Other, more complex schemes, may be needed in the case of multivariate observations.

To the extent that rolling windows represent an averaging (over starting points of each interval) of several estimators that have non-overlapping windows, our results (except in Sections 5.2 and 6.2) remain valid. The negligibility of edge effects are unaffected, and the representations in Theorem 3–4 are equally unaffected. Here, our algebraic results again clarify the crux of the problem: to find a natural way to roll the window so as to average the representation terms in (30).

8. Conclusion

We have discussed the finite sample and then asymptotic representation of the smoothed two-scales realized volatility (S-TSRV). We have shown that the estimator has small edge effect, and we have also derived a representation theorem for the estimator which can be used as a tool to characterize its behavior in a wide range of situations that can combine irregular, asynchronous, and endogenous times, and jumps.

Appendix A. Edge effect and calibration of the original TSRV: derivations for Sections 3.1 and 4.1

A.1. Edge effect

The explicit form of the cross terms in (15) is given by

$$C_K = 2\sum_{i=1}^{N-K} \eta_{i+K}(X_{\tau_{i+K-1}} - X_{\tau_i}) + C_K^* \text{ where } C_K^* = 2\sum_{i=1}^{N-K} \left((X_{\tau_{i+K-1}} - X_{\tau_i})\eta_i' + \eta_{i+K}\eta_i'\right). \tag{A.1}$$

[6] *Op. cit.*, Section 4. This reference also essentially provides the argument why the two terms in (47) are asymptotically (conditionally) independent and why the second term is Gaussian. If σ_t^2 is not an Itô-semimartingale, then the integration by parts still typically goes through, but the order or the term may change.

This is a sum of three martingale U-statistics under the conditions of Fact 1. Furthermore,

$$C_K - C_J = 2\left(\sum_{i=K+1}^{N} \eta_i(X_{\tau_{i-1}} - X_{\tau_{i-K}}) - \sum_{i=J+1}^{N} \eta_i(X_{\tau_{i-1}} - X_{\tau_{i-J}})\right) + C_K^* - C_J^* = C_{K,J} + R_{K,J}$$

$$\text{where } C_{K,J} = 2\sum_{i=K+1}^{N} \eta_i(X_{\tau_{i-J}} - X_{\tau_{i-K}}) \text{ and } R_{K,J} = -2\sum_{i=J+1}^{K} \eta_i(X_{\tau_{i-1}} - X_{\tau_{i-J}}). \tag{A.2}$$

$C_{K,J}$ is a sum of martingale U-statistics under the conditions of Fact 2.

In view of (13), the edge effect $e_{K,J}$ is given by

$$\begin{aligned} e_{K,J} &= \sum_{i=1}^{N-K}\left[(\eta_{i+K})^2 + (\eta_i')^2\right] - \sum_{i=1}^{N-J}\left[(\eta_{i+J})^2 + (\eta_i')^2\right] + R_{K,J} \\ &= \{\sum_{i=K+1}^{N} \eta_i^2 - \sum_{i=J+1}^{N} \eta_i^2\} + \{\sum_{i=1}^{N-K}(\eta_i')^2 - \sum_{i=1}^{N-J}(\eta_i')^2\} + R_{K,J} \\ &= -\sum_{i=J+1}^{K} \eta_i^2 - \sum_{i=N-K+1}^{N-J} (\eta_i')^2 + R_{K,J}. \end{aligned} \tag{A.3}$$

Proof of Proposition 1. Define the 2-norm by $\|R\|_2 = (E(R^2))^{1/2}$.

$e_{K,J}$ case. Under Condition 2, if $X_t^{(1)}$ replaces X_t in the definition of V_i and V_i', $E\eta_i^2$ and $E(\eta_i')^2$ are both bounded by $2(E(\Delta X_{\tau_{n,i}}^{(1)})^2 + \mathrm{Var}(\bar{\epsilon}_i)) = 2(E\Delta\langle X^{(1)}, X^{(1)}\rangle_{\tau_{n,i}} + \mathrm{Var}(\bar{\epsilon}_i))$. Since $\mathrm{Var}(\bar{\epsilon}_{n,i}) = O(M_{n,i}^{-1})$, the bound becomes $O(\max_i \Delta\tau_{n,i} + \max_i M_{n,i}^{-1})$. With the same redefinition of X_t, $\|R_{K,J}\|_2 \le 2\sum_{J+1}^{K} \|\eta_i\|_2 \|X_{\tau_{i-1}} - X_{\tau_{i-J}}\|_2 = O((K-J)(\max_i \Delta\tau_{n,i} + \max_i M_{n,i}^{-1}))^{1/2}((J-1)\max_i \Delta\tau_{n,i})^{1/2}$. Hence, again with the same redefinition of X_t,

$$e_{K,J} = O_p\left((K-J)J^{1/2}(\max_i \Delta\tau_{n,i} + \max_i M_{n,i}^{-1})\right) \tag{A.4}$$

We have here also assumed that $d\langle X^{(1)}, X^{(1)}\rangle_t/dt \le c$, where c is nonrandom. The nonrandom constant c is subsequently removed via localization, cf. the references on localization just before Definition 1. Since the event that $X_t^{(2)}$ has no jump on $(\tau_{n,J}, \tau_{n,K}]$ and $(\tau_{n,N-K}, \tau_{n,N-J}]$ has probability tending to one, the bound (A.4) also applies to the original $e_{K,J}$ from (A.3).

For compactness, we also show the $\tilde{e}_{K,J}$ case here. Refer to (22) for the form of $\tilde{e}_{K,J}$. As in the $e_{K,J}$ case, we can replace X_t by $X_t^{(1)}$. Since $\|\eta_i\|_2$ and $\|\eta_i'\|_2$ are both uniformly bounded by $O(\Delta\tau_n^+ + (M_n^-)^{-1})^{1/2}$, we obtain the second formula in (29).

A.2. Calibration of this estimator

To obtain a calibration constant, assume for simplicity that X_t is continuous and that σ is constant and that the $\Delta\tau$ are completely regular, *i.e.*, $\Delta\tau = T/N$. The first term on the right hand side of (16) has expectation $(N-K)(K-1)\Delta\tau\sigma^2$. The combined first two terms thus has expectation $[(N-K)(K-1) - (N-J)(J-1)]\Delta\tau\sigma^2 = (K-J)(N-K-J+1)\Delta\tau\sigma^2$. We then have to handle the edge effect. If we assume that the times have the same (possibly irregular) distribution inside each interval, and if we take $(\bar{\epsilon}_i)^2$ to be iid (and independent of the X process), we obtain

$$\begin{aligned} Ee_{K,J} &= -(K-J)[E(V_1^2 + (V_1')^2) + 2\mathrm{Var}(\bar{\epsilon}_1)] \\ &= -(K-J)[\Delta\tau\sigma^2(1 - 2E(I_1(1-I_1))) + 2\mathrm{Var}(\bar{\epsilon}_1)] \end{aligned} \tag{A.5}$$

If we further assume that the t_j are equidistant, and that the ϵ_j are iid, we get from (8)

$$Ee_{K,J} \approx -(K-J)[\frac{2}{3}\Delta\tau\sigma^2 + 2M^{-1}\mathrm{Var}(\epsilon)]. \tag{A.6}$$

We thus obtain

$$E[K[\bar{Y}, \bar{Y}]^{(K)} - J[\bar{Y}, \bar{Y}]^{(J)}] = (K-J)\left(1 - \frac{K+J-\frac{1}{3}}{N}\right)\sigma^2 T - 2(K-J)M^{-1}\mathrm{Var}(\epsilon) \tag{A.7}$$

If we assume that M is large, and/or that the signal-to-noise ratio $\sigma^2/\mathrm{Var}(\epsilon)$ is also large, one approach is to ignore the term due to the ϵ_j. Since in the σ^2 constant case, the integrated volatility has value $T\sigma^2$, we thus get the proposed estimator in (17).

Appendix B. Estimation without $\bar{\epsilon}_i^2$: derivations for Section 3.2

B.1. Derivation of linear combination

Write from (15) that

$$K[\bar{Y},\bar{Y}]^{(K)} = \sum_{i=1}^{N-K}(X_{\tau_{i+K-1}} - X_{\tau_i})^2 + \sum_{i=K+1}^{N}[(V_i+\bar{\epsilon}_i)^2] + \sum_{i=1}^{N-K}[(V_i'+\bar{\epsilon}_i)^2] + C_K. \tag{B.8}$$

We then wish for a modified J $(< K)$ scale measure. Consider first the measure on part of the time line, from a to b

$$\begin{aligned} J[\bar{Y},\bar{Y}]^{(J,a,b)} &= \sum_{i=a}^{b-J}(\bar{Y}_{i+J} - \bar{Y}_i)^2 \\ &= \sum_{i=a}^{b-J}(X_{\tau_{i+J-1}} - X_{\tau_i})^2 + \sum_{i=J+a}^{b}[(V_i+\bar{\epsilon}_i)^2] + \sum_{i=a}^{b-J}[(V_i'+\bar{\epsilon}_i)^2] + C_{J,a,b}. \end{aligned} \tag{B.9}$$

The contribution of the $\bar{\epsilon}^2$ terms is

$$\begin{aligned} \mathcal{E}(J,a,b) &= \sum_{i=a}^{a+J-1}(\bar{\epsilon}_i)^2 + \sum_{i=b-J+1}^{b}(\bar{\epsilon}_i)^2 + 2\sum_{i=a+J}^{b-J}(\bar{\epsilon}_i)^2 \text{ if } b-a \geq 2J \\ &= \sum_{i=a}^{b-J}(\bar{\epsilon}_i)^2 + \sum_{i=J+a}^{b}(\bar{\epsilon}_i)^2 \text{ otherwise.} \end{aligned} \tag{B.10}$$

Now assume that $b-a = K+J-1$, and note that $2K > b-a \geq 2J$ since $K > J$. We obtain

$$\mathcal{E}(K,a,b) - \mathcal{E}(J,a,b) = -2\sum_{i=a+J}^{b-J}(\bar{\epsilon}_i)^2 \tag{B.11}$$

Hence components due to $\bar{\epsilon}^2$ in the edge effect in (A.3) can be written

$$\frac{1}{2}\{\mathcal{E}(K,a,b) - \mathcal{E}(J,a,b)|_{a=1}^{b=K+J}\} + \frac{1}{2}\{\mathcal{E}(K,a,b) - \mathcal{E}(J,a,b)|_{a=N-(K+J)+1}^{b=N}\} \tag{B.12}$$

We can thus create a new two scales estimator with no $\bar{\epsilon}^2$ in the edge effect or anywhere else, by modifying the measurements of volatility as follows. With $b = K+J$, set

$$\begin{aligned} K\widetilde{[\bar{Y},\bar{Y}]}^{(K)} &= K[\bar{Y},\bar{Y}]^{(K)} - \frac{1}{2}K[\bar{Y},\bar{Y}]^{(K,1,b)} - \frac{1}{2}K[\bar{Y},\bar{Y}]^{(K,N-b+1,N)} \\ &= \frac{1}{2}\sum_{i=1}^{b-K}(\bar{Y}_{i+K} - \bar{Y}_i)^2 + \sum_{i=b-K+1}^{N-b}(\bar{Y}_{i+K} - \bar{Y}_i)^2 + \frac{1}{2}\sum_{i=N-b+1}^{N-K}(\bar{Y}_{i+K} - \bar{Y}_i)^2. \end{aligned} \tag{B.13}$$

This leads to the same definition as (19). One obtains that $K\widetilde{[\bar{Y},\bar{Y}]}^{(K)} - J\widetilde{[\bar{Y},\bar{Y}]}^{(J)}$ has no edge effect. For the $\bar{\epsilon}^2$ terms, this is how the combination was derived, but clearly the V and V' terms vanish similarly.

B.2. Total edge effect for pure η and η' terms

The components due to η_i^2 and $(\eta_i')^2$ in $K\widetilde{[\bar{Y},\bar{Y}]}^{(K)}$ are given by

$$\begin{aligned} &= \left(\frac{1}{2}\sum_{i=1}^{b-K} + \sum_{i=b-K+1}^{N-b} + \frac{1}{2}\sum_{i=N-b+1}^{N-K}\right)(\eta_{i+K}^2 + (\eta_i')^2) \\ &= \left(\frac{1}{2}\sum_{i=K+1}^{b} + \sum_{i=b+1}^{N-J} + \frac{1}{2}\sum_{i=N-J+1}^{N}\right)\eta_i^2 + \left(\frac{1}{2}\sum_{i=1}^{J} + \sum_{i=J+1}^{N-b} + \frac{1}{2}\sum_{i=N-b+1}^{N-K}\right)(\eta_i')^2 \end{aligned} \tag{B.14}$$

The similar expression for $J\widetilde{[\bar{Y},\bar{Y}]}^{(J)}$ is

$$\left(\frac{1}{2}\sum_{i=J+1}^{b} + \sum_{i=b+1}^{N-K} + \frac{1}{2}\sum_{i=N-K+1}^{N}\right)\eta_i^2 + \left(\frac{1}{2}\sum_{i=1}^{K} + \sum_{i=K+1}^{N-b} + \frac{1}{2}\sum_{i=N-b+1}^{N-J}\right)(\eta_i')^2. \tag{B.15}$$

Subtracting component by component (in the same order) gives

$$
\begin{aligned}
&\text{Components due to } \eta_i^2 \text{ and } (\eta_i')^2 \text{ in } K\widetilde{[\bar{Y},\bar{Y}]}^{(K)} - J\widetilde{[\bar{Y},\bar{Y}]}^{(J)} \\
&= \left\{ \frac{1}{2}\left(\sum_{i=K+1}^{b} - \sum_{i=J+1}^{b}\right) + \left(\sum_{i=b+1}^{N-J} - \sum_{i=b+1}^{N-K}\right) + \frac{1}{2}\left(\sum_{i=N-J+1}^{N} - \sum_{i=N-K+1}^{N}\right)\right\}\eta_i^2 \\
&+ \left\{ \frac{1}{2}\left(\sum_{i=1}^{J} - \sum_{i=1}^{K}\right) + \left(\sum_{i=J+1}^{N-b} - \sum_{i=K+1}^{N-b}\right) + \frac{1}{2}\left(\sum_{i=N-b+1}^{N-K} - \sum_{i=N-b+1}^{N-J}\right)\right\}(\eta_i')^2 \\
&= \left(-\frac{1}{2}\sum_{i=J+1}^{K} + \frac{1}{2}\sum_{i=N-K+1}^{N-J}\right)\eta_i^2 + \left(\frac{1}{2}\sum_{i=J+1}^{K} - \frac{1}{2}\sum_{i=N-K+1}^{N-J}\right)(\eta_i')^2 \\
&= \frac{1}{2}\left(-\sum_{i=J+1}^{K} + \sum_{i=N-K+1}^{N-J}\right)(\eta_i^2 - (\eta_i')^2) = \frac{1}{2}\left(-\sum_{i=J+1}^{K} + \sum_{i=N-K+1}^{N-J}\right)(\eta_i - \eta_i')\Delta X_{\tau_i},
\end{aligned}
\tag{B.16}
$$

since $\eta_i + \eta_i' = V_i + V_i' = \Delta X_{\tau_i}$.

B.3. The martingale U-statistics and the complete edge effect

An alternative representation of (19) is given by

$$
K\widetilde{[\bar{Y},\bar{Y}]}^{(K)} = \frac{1}{2}\left(\sum_{i=1}^{N-b} + \sum_{i=J+1}^{N-K}\right)(\bar{Y}_{i+K} - \bar{Y}_i)^2 \tag{B.17}
$$

In analogy with (A.1), cross terms are given by

$$
\begin{aligned}
\tilde{C}_K &= \left(\sum_{i=1}^{N-b} + \sum_{i=J+1}^{N-K}\right)\eta_{i+K}(X_{\tau_{i+K-1}} - X_{\tau_i}) + \tilde{C}_K^* = \left(\sum_{i=K+1}^{N-J} + \sum_{i=b+1}^{N}\right)\eta_i(X_{\tau_{i-1}} - X_{\tau_{i-K}}) + \tilde{C}_K^* \\
\text{where } \tilde{C}_K^* &= \left(\sum_{i=1}^{N-b} + \sum_{i=J+1}^{N-K}\right)\left((X_{\tau_{i+K-1}} - X_{\tau_i})\eta_i' + \eta_{i+K}\eta_i'\right).
\end{aligned}
\tag{B.18}
$$

In parallel with (A.2), we obtain

$$
\begin{aligned}
\tilde{C}_K - \tilde{C}_J &= \left(\sum_{i=K+1}^{N-J} + \sum_{i=b+1}^{N}\right)\eta_i(X_{\tau_{i-1}} - X_{\tau_{i-K}}) - \left(\sum_{i=J+1}^{N-K} + \sum_{i=b+1}^{N}\right)\eta_i(X_{\tau_{i-1}} - X_{\tau_{i-J}}) + \tilde{C}_K^* - \tilde{C}_J^* \\
&= \tilde{C}_{K,J} + \tilde{R}_{K,J} \text{ where} \\
\tilde{C}_{K,J} &= \left(\sum_{i=K+1}^{N-J} + \sum_{i=b+1}^{N}\right)\eta_i(X_{\tau_{i-J}} - X_{\tau_{i-K}}) + \tilde{C}_K^* - \tilde{C}_J^* \text{ and} \\
\tilde{R}_{K,J} &= \left(-\sum_{i=J+1}^{K} + \sum_{i=N-K+1}^{N-J}\right)\eta_i(X_{\tau_{i-1}} - X_{\tau_{i-J}}).
\end{aligned}
\tag{B.19}
$$

By (B.16), the complete edge effect $\tilde{e}_{K,J}$ is given by

$$
\tilde{e}_{K,J} = \frac{1}{2}\left(-\sum_{i=J+1}^{K} + \sum_{i=N-K+1}^{N-J}\right)(\eta_i - \eta_i')\Delta X_{\tau_i} + \tilde{R}_{K,J}. \tag{B.20}
$$

B.4. Calibration of the modified estimator

To get an calibration constant, assume for simplicity that σ is constant that the $\Delta\tau$ are regular, *i.e.*, $\Delta\tau = T/N$. Also, for an all-purpose constant, assume that $E(I_i) = \frac{1}{2}$ for the edge values $i \in [J+1, K] \cup [N-K+1, N-J]$. The first term on the right hand side of (21) has expectation $[(b-K)+(N-b-(b-K))](K-1)\Delta\tau\sigma^2=(N-b)(K-1)\Delta\tau\sigma^2$. The whole expression thus has expectation $(N-b)(K-J)\Delta\tau\sigma^2=(1-b/N)(K-J)T\sigma^2$. Since in the σ^2 constant case, the integrated volatility has value $T\sigma^2$, we thus get the proposed estimator in (23).

Appendix C. Proof of Theorem 2 in Section 3.3

Set

$$K_1=\frac{1}{2}\sum_{i=1}^{N-b}(X_{\tau_{i+K-1}}-X_{\tau_i})^2, K_2=\frac{1}{2}\sum_{i=J+1}^{N-K}(X_{\tau_{i+K-1}}-X_{\tau_i})^2, \text{ and}$$
$$J_1=\frac{1}{2}\sum_{i=1}^{N-b}(X_{\tau_{i+J-1}}-X_{\tau_i})^2, J_2=\frac{1}{2}\sum_{i=K+1}^{N-J}(X_{\tau_{i+J-1}}-X_{\tau_i})^2. \tag{C.21}$$

We shall use that the squared terms in (21) equal $K_1+K_2-J_1-J_2$. Observe that

$$\begin{aligned}&(X_{\tau_{i+K-1}}-X_{\tau_i})^2-(X_{\tau_{i+J-1}}-X_{\tau_i})^2\\&=(X_{\tau_{i+K-1}}-X_{\tau_{i+J-1}})^2+2(X_{\tau_{i+K-1}}-X_{\tau_{i+J-1}})(X_{\tau_{i+J-1}}-X_{\tau_i}).\end{aligned} \tag{C.22}$$

(The second term is obviously a cross term.) Hence,

$$\begin{aligned}K_1-J_1&=\frac{1}{2}\sum_{i=1}^{N-b}\left\{(X_{\tau_{i+K-1}}-X_{\tau_{i+J-1}})^2+2(X_{\tau_{i+K-1}}-X_{\tau_{i+J-1}})(X_{\tau_{i+J-1}}-X_{\tau_i})\right\}\\&=\frac{1}{2}\sum_{i=J}^{N-K-1}(X_{\tau_{i+K-J}}-X_{\tau_i})^2+\sum_{i=K}^{N-J-1}(X_{\tau_i}-X_{\tau_{i-(K-J)}})(X_{\tau_{i-(K-J)}}-X_{\tau_{i-K+1}})\end{aligned} \tag{C.23}$$

Now re-index the sums in K_2 and J_2 to get

$$K_2=\frac{1}{2}\sum_{i=b}^{N-1}(X_{\tau_i}-X_{\tau_{i-K+1}})^2, \text{ and } J_2=\frac{1}{2}\sum_{i=b}^{N-1}(X_{\tau_i}-X_{\tau_{i-J+1}})^2. \tag{C.24}$$

By symmetry, one can proceed as in (C.22), on the opposite edge:

$$(X_{\tau_i}-X_{\tau_{i-K+1}})^2-(X_{\tau_i}-X_{\tau_{i-J+1}})^2=(X_{\tau_{i-K+1}}-X_{\tau_{i-J+1}})^2+2(X_{\tau_i}-X_{\tau_{i-J+1}})(X_{\tau_{i-J+1}}-X_{\tau_{i-K+1}}). \tag{C.25}$$

This yields

$$\begin{aligned}K_2-J_2&=\frac{1}{2}\sum_{i=b}^{N-1}\left\{(X_{\tau_{i-K+1}}-X_{\tau_{i-J+1}})^2+2(X_{\tau_i}-X_{\tau_{i-J+1}})(X_{\tau_{i-J+1}}-X_{\tau_{i-K+1}})\right\}\\&=\frac{1}{2}\sum_{i=J+1}^{N-K}(X_{\tau_{i+K-J}}-X_{\tau_i})^2+\sum_{i=b}^{N-1}(X_{\tau_i}-X_{\tau_{i-J+1}})(X_{\tau_{i-J+1}}-X_{\tau_{i-K+1}}).\end{aligned} \tag{C.26}$$

Combining (C.23) and (C.26), we thus obtain Theorem 2.

Appendix D. Proof of Theorem 3 in Section 4

In the following, observe that by localization (see references on localization just before Definition 1), one can take $d\langle X^{(1)},X^{(1)}\rangle_t/dt\le\nu^+$ where ν^+ is a nonrandom constant and $\langle X^{(1)},X^{(1)}\rangle_t$ is the predictable quadratic variation from Condition 2.

We give the maximum order of neglected terms.

1. **The** $(1-b/N)$ **factor** can be replaced by 1, for the same reasons as at the beginning of Appendix E.
2. **The edge term in** $\widehat{\langle X,X\rangle}$ **is covered by** Proposition 1, except that we replace M_n^- by $o_p(\Delta\tau^+(K-J)^{1/2})$. This is valid since M_n^- only entered the proof of Proposition 1 as a symbol.
3. **The two "1/2" terms in** Eq. (24) are of order $O_p(\Delta\tau^+/(K-J))$, and hence negligible.
4. **The cross terms explicitly given in Eqs.** (C.23) **and** (C.26) **are handled as follows.** For terms containing only $X^{(1)}$, the quadratic variations are of order $O_p(J(K-J)^2\Delta\tau^+)$ and $O_p(J^2(K-J)\Delta\tau^+)$ before normalization, and are thus negligible after normalization. Terms containing $X^{(1)}$ and $X^{(2)}$ are directly of same or smaller order than those containing only $X^{(1)}$, since there are only finitely many such terms. Asymptotically, no term contains only $X^{(2)}$.

In Item 4 above, and in the remaining items below, we handle (sums of) martingale U-statistics. Inference is made from the predictable quadratic variation to the original martingale with the help of Lenglart domination (Jacod and Shiryaev, 2003, Lemma 3.30, p. 35).

We are now left to deal with the terms from (B.19). Set $V_i^{(2)}$ as the contribution to η_i from $X^{(2)}$. Also set $\eta_{i,1} = V_i^{(1)} + \bar{\epsilon}_i$ and $\eta'_{i,1} = V_i^{(1)\prime} - \bar{\epsilon}_i$, where $V_i^{(1)}$ is the contribution to η_i from $X^{(1)}$, and similarly for $V_i^{(1)\prime}$. Observe that $E(\eta_i^2 \mid \mathcal{F}_{i-J}) \leq 2\nu^+\Delta\tau^+ + 2E(\bar{\epsilon}_i^2 \mid \mathcal{F}_{i-J})$, whence by Condition 4,

$$E \sup_i E(\eta_{i,1}^2 \mid \mathcal{F}_{i-J}) = o(\Delta\tau_n^+(K-J)^{1/2}) \tag{D.27}$$

We now continue with the individual terms

5. **The cross term of the form** $\sum_i \eta_i(X_{\tau_{i-J}} - X_{\tau_{i-K}})$. Eventually (for any given outcome ω), $\sum_i V_i^{(2)}(X_{\tau_{i-J}} - X_{\tau_{i-K}}) = \sum_i V_i^{(2)}(X^{(1)}_{\tau_{i-J}} - X^{(1)}_{\tau_{i-K}}) = O_p(\Delta\tau^+(K-J))$. The predictable quadratic variation of the modified $\sum_i \eta_{i,1}(X_{\tau_{i-J}} - X_{\tau_{i-K}})$ is bounded by $\sum_i E(\eta_i^2 \mid \mathcal{F}_{i-J})(X_{\tau_{i-J}} - X_{\tau_{i-K}})^2 \leq o_p(\Delta\tau^+(K-J)^{3/2})$ by (D.27), and hence this term is also negligible.
6. **The cross terms** C_K^* and C_J^* **of the form** $\sum_i(X_{\tau_{i+K-1}} - X_{\tau_i})\eta'_i$. As in the previous Item 5, we can replace η_i with $\eta_{i,1}$ without loss of generality. Again decompose into two terms. On the one hand $\sum_i(X^{(2)}_{\tau_{i+K-1}} - X^{(2)}_{\tau_i})\eta'_{i,1} = o_p(\Delta\tau^+(K-J)^{1/2})$ by (D.27) since there are only finitely many terms in the sum. On the other hand, the main term in $C_K^* - C_J^*$ is of the form $\sum_i(X^{(1)}_{\tau_{i+K-1}} - X^{(1)}_{\tau_{i+J-1}})\eta'_{i,1}$, and has an observed (optional) quadratic variation which is Lenglart-dominated by $\nu^+(\Delta\tau^+)^2(K-J)^{3/2}$, and hence this term is also negligible. The remaining edge related terms disappear similarly.
7. **The cross term of the form** $\sum_i \eta_{i+K}\eta'_i$ **and** $\sum_i \eta_{i+J}\eta'_i$. As in the previous Item 5, we can replace η_i with $\eta_{i,1}$ without loss of generality. By (D.27), these terms have predictable quadratic variation bounded by $o_p(N(\Delta\tau^+)^2(K-J)) = o_p(\Delta\tau^+(K-J))$, which is also negligible.

This completes the proof.

Appendix E. Proof of Theorem 5 in Section 5.2

The factor $(1-b/N)^{-1}$ can be replaced by 1, since $(N/(K-J))^{1/2}\left((1-b/N)^{-1} - 1\right) = O((N(K-J))^{-1/2}) = o(1)$. We are thus interested in the limit of $(N/(K-J))^{1/2}\tilde{D}^{(r,s)}$, where

$$\tilde{D}^{(r,s)} = \sum_{i=J+1}^{N-K}(X^{(r)}_{\tau_{i+K-J}} - X^{(r)}_{\tau_i})(X^{(s)}_{\tau_{i+K-J}} - X^{(s)}_{\tau_i}) - \langle X^{(r)}, X^{(s)}\rangle_T = M^{(r,s)}_{N,T}[2]$$

where $M^{(r,s)}_{N,t}$ is the martingale (cf. Condition 1) with end point (value at T) given by $\sum_{i=J+1}^{N-K}\int_{\tau_i}^{\tau_{i+K-J}}(X^{(r)}_t - X^{(r)}_{\tau_i})dX^{(s)}_{\tau_i}$ and where $a^{r,s}[2] = a^{r,s} + a^{s,r}$.

Compare to the proof of Theorem 2–3 (p. 1410–1411) in Zhang et al. (2005). Replace t_i by τ_i. Since we are in a multidimensional situation, replace D_T by $\tilde{D}^{(r,s)}_T = \tilde{D}^{(r,s)}$. The martingale representation of $M^{(r,s)}_{N,T}[2]$ is similar to (*ibid.*, (A.25)). Following the same development,

$$\langle \tilde{D}^{(r_1,s_1)}, \tilde{D}^{(r_2,s_2)}_N\rangle_T = \langle M^{(r_1,s_1)}_N[2], M^{(r_2,s_2)}_N[2]\rangle_T = \langle M^{(r_1,s_1)}_N, M^{(r_2,s_2)}_N\rangle_T[2][2]$$

which, as in the earlier paper, is asymptotically the same as

$$(K-J)\overline{\Delta\tau_n}\frac{1}{4}\int_0^T \langle X^{(r_1)}, X^{(r_2)}\rangle'_t\langle X^{(s_1)}, X^{(s_2)}\rangle'_t dG(t)[2][2]$$

The result follows similarly to the theorems in Zhang et al. (2005).

References

Aït-Sahalia, Y., Mykland, P.A., Zhang, L., 2011. Ultra high frequency volatility estimation with dependent microstructure noise. J. Econometrics 160, 160–175.

Altmeyer, R., Bibinger, M., 2015. Functional stable limit theorems for quasi-efficient spectral covolatility estimators. Stochastic Process. Appl. 125, 4556–4600.

Andersen, T.G., Bollerslev, T., 1998a. Answering the skeptics: Yes, Standard volatility models do provide accurate forecasts. Internat. Econom. Rev. 39, 885–905.

Andersen, T.G., Bollerslev, T., 1998b. Deutsche Mark-Dollar volatility: Intraday activity patterns, macroeconomic announcements, and longer run dependencies. J. Finance 53, 219–265.

Andersen, T.G., Bollerslev, T., Diebold, F.X., Labys, P., 2000. Great realizations. Risk 13, 105–108.

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

Barndorff-Nielsen, O.E., Shephard, N., 2002. Econometric analysis of realized volatility and its use in estimating stochastic volatility models. J. R. Stat. Soc. Ser. B Stat. Methodol. 64, 253–280.

Bibinger, M., Mykland, P.A., 2016. Inference for multi-dimensional high-frequency data with an application to conditional independence testing. Scand. J. Stat. 43, 1078–1102.

Bibinger, M., Reiß, M., 2014. Spectral estimation of covolatility from noisy observations using local weights. Scand. J. Stat. 41 (1), 23–50.

Chen, D., Mykland, P.A., Zhang, L., 2018. The five trolls under the bridge: principal component analysis with asynchronous and noisy high frequency data. Available at SSRN https://ssrn.com/abstract=3118039.
Christensen, K., Podolskij, M., Vetter, M., 2013. On covariation estimation for multivariate continuous Itô semimartingales with noise in non-synchronous observation schemes. J. Multivariate Anal. 120, 59–84.
Delbaen, F., Schachermayer, W., 1994. A General Version of the Fundamental Theorem of Asset Pricing. Math. Ann. 300, 463–520.
Delbaen, F., Schachermayer, W., 1995. The existence of absolutely continuous local martingale measures. Ann. Appl. Probab. 5, 926–945.
Delbaen, F., Schachermayer, W., 1998. The Fundamental Theorem of Asset Pricing For Unbounded Stochastic Processes. Math. Ann. 312, 215–250.
Hall, P., Heyde, C.C., 1980. Martingale Limit Theory and Its Application. Academic Press, Boston.
Hasbrouck, J., 1996. Modeling Market Microstructure Time Series. In: Rao, C.R., Maddala, G.S. (Eds.), Handbook of Statistics, vol. 14. North-Holland, Amsterdam, pp. 647–692.
Jacod, J., Li, Y., Mykland, P.A., Podolskij, M., Vetter, M., 2009a. Microstructure noise in the continuous case: the pre-averaging approach. Stochastic Process. Appl. 119, 2249–2276.
Jacod, J., Podolskij, M., Vetter, M., 2009b. Limit theorems for moving averages of discretized processes plus noise. Ann. Statist. 38, 1478–1545.
Jacod, J., Protter, P., 2012. Discretization of Processes, first ed. Springer-Verlag, New York.
Jacod, J., Rosenbaum, M., 2013. Quarticity and other Functionals of Volatility: Efficient Estimation. Ann. Statist. 41, 1462–1484.
Jacod, J., Rosenbaum, M., 2015. Estimation of volatility functionals: The case of a $\sqrt{n}$ window. In: Friz, P., Gatheral, J., Gulisashvili, A., Jacquier, A., Teichmann, J. (Eds.), Large Deviations and Asymptotic Methods in Finance. In: the Springer Proceedings in Mathematics and Statistics, vol. 110, Springer-Verlag, Berlin, pp. 559–590.
Jacod, J., Shiryaev, A.N., 2003. Limit Theorems for Stochastic Processes, second ed. Springer-Verlag, New York.
Kalnina, I., Linton, O., 2008. Estimating quadratic variation in the presence of endogenous and diurnal microstructure noise. J. Econometrics 147, 47–59.
Li, Y., Mykland, P.A., Renault, E., Zhang, L., Zheng, X., 2014. Realized volatility when sampling times are possibly endogenous. Econom. Theory 30, 580–605.
Li, Y., Zhang, Z., Zheng, X., 2013. Volatility inference in the presence of both endogenous time and microstructure noise. Stochastic Process. Appl. 123, 2696–2727.
Mancino, M.E., Recchioni, M.C., Sanfelici, S., 2017. Fourier-Malliavin Volatility Estimation: Theory and Practice. SpringerBriefs in Quantitative Finance, Cham, Switzerland.
McCullagh, P., 1987. Tensor Methods in Statistics. Chapman and Hall, London, U.K..
McLeish, D.L., 1975. A maximal inequality and dependent strong laws. Ann. Probab. 3 (5), 829–839.
Mykland, P.A., Zhang, L., 2005. Discussion of paper A selective overview of nonparametric metods in financial econometrics by J. Fan. Statist. Sci. 20, 347–350.
Mykland, P.A., Zhang, L., 2008. Inference for volatility-type objects and implications for hedging. Stat. Interface 1, 255–278.
Mykland, P.A., Zhang, L., 2009. Inference for continuous semimartingales observed at high frequency. Econometrica 77, 1403–1455.
Mykland, P.A., Zhang, L., 2012. The econometrics of high frequency data. In: Kessler, M., Lindner, A., Sørensen, M. (Eds.), Statistical Methods for Stochastic Differential Equations. Chapman and Hall/CRC Press, New York, pp. 109–190.
Mykland, P.A., Zhang, L., 2016. Between data cleaning and inference: Pre-averaging and other robust estimators of the efficient price. J. Econometrics 194, 242–262.
Mykland, P.A., Zhang, L., 2017. Assessment of uncertainty in high frequency data: the observed asymptotic variance. Econometrica 85, 197–231.
Nummelin, E., 1984. General Irreducible Markov Chains and Non-Negative Operators. Cambridge University Press, Cambridge, UK.
O'Hara, M., 1995. Market Microstructure Theory. Blackwell Publishers, Cambridge, MA.
Park, S., 2011. Consistent Estimator of Ex-Post Covariation of Discretely Observed Diffusion Processes and Its Application To High Frequency Financial Time Series. (Ph.D. thesis), London School of Economics, Department of Statistics.
Park, S., Hong, S.Y., Linton, O., 2016. Estimating the quadratic covariation matrix asynchronously observed high frequency stock returns corrupted by additive measurement error. J. Econometrics 191, 325–347.
Podolskij, M., Vetter, M., 2009a. Bipower-type estimation in a noisy diffusion setting. Stochastic Process. Appl. 119, 2803–2831.
Podolskij, M., Vetter, M., 2009b. Estimation of volatility functionals in the simultaneous presence of microstructure noise and jumps. Bernoulli 15, 634–658.
Rootzén, H., 1980. Limit distributions for the error in approximations of stochastic integrals. Ann. Probab. 8, 241–251.
Wu, W.B., 2007. Strong invariance principles for dependent random variables. Ann. Probab. 35 (6), 2294–2320.
Xiu, D., 2010. Quasi-maximum likelihood estimation of volatility with high frequency data. J. Econometrics 159, 235–250.
Zhang, L., 2006. Efficient estimation of stochastic volatility using noisy observations: a multi-scale approach. Bernoulli 12, 1019–1043.
Zhang, L., 2011. Estimating covariation: Epps effect and microstructure noise. J. Econometrics 160, 33–47.
Zhang, L., Mykland, P.A., Aït-Sahalia, Y., 2005. A tale of two time scales: determining integrated volatility with noisy high-frequency data. J. Amer. Statist. Assoc. 100, 1394–1411.