Journal of Econometrics 222 (2021) 411–428

Contents lists available at ScienceDirect

Journal of Econometrics

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

The Observed Asymptotic Variance: Hard edges, and a regression approach[☆]

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

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

ARTICLE INFO

Article history:
Available online 11 August 2020

Keywords:
Asynchronous times
Consistency
Discrete observation
Edge effect
Irregular times
Leverage effect
Microstructure
Observed information
Realized volatility
Robust estimation
Semimartingale
Standard error
Two scales estimation
Volatility of volatility

ABSTRACT

High frequency financial data has become an essential component of the digital economy, yielding an increasing number of estimators. However, it is hard to reliably assess the uncertainty of such estimators. The Observed Asymptotic Variance (observed AVAR) is a non-parametric estimator for (squared) standard error in high frequency data. The device is related to observed information in likelihood theory, but in this case it is non-parametric and uses the high-frequency data structure. An earlier paper has developed the estimator in the case where edge effects are small to moderate. In practical data, it is often more realistic to assume that edge effects can be large, and this is the problem that we tackle in the current paper. We here find a regression approach to observed AVAR which is highly robust to large edges. This approach covers most high frequency estimators.

1. Introduction: Hard edges

The Observed Asymptotic Variance is a non-parametric estimator for squared standard error in high frequency data (Mykland and Zhang, 2017a). The earlier paper develops the estimator in the case where edge effects are small to moderate. In practical data, it is often safer to assume that edge effects can be large, and this is the problem that we seek to tackle in this paper.

We consider integrated parameters and their estimators[1] over time intervals $(S, T] \subset [0, \mathcal{T}]$:

$$\Theta_{(S,T]} = \int_S^T \theta_t dt \text{ and } \hat{\Theta}_{(S,T]} = \text{ a consistent estimator of } \Theta_{(S,T]}, \tag{1}$$

☆ The authors would like to thank the co-editor and the referees for helpful suggestions and advice. Financial support from the National Science Foundation, United States of America under grants DMS 17-13118 and DMS-2015530 (Zhang) and DMS 17-13129 and DMS-2015544 (Mykland) is gratefully acknowledged.

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

1 All estimators are implicitly or explicitly indexed by the number of observations n. Consistency, convergence in law, etc, refers to behavior as $n \to \infty$.

https://doi.org/10.1016/j.jeconom.2020.07.008

where θ_t is a semi-martingale representing spot volatility, skewness, a regression coefficient, or other. See Mykland and Zhang (2017a) for examples (Section 7) and general principles for how to find estimators of this type, and for the removal of jumps (Sections 5–6).

The typical statistical situation is as follows: there is a semimartingale $M_{n,t}$ and *edge effects* $e_{n,S}$ and $\tilde{e}_{n,T}$, so that,

$$\hat{\Theta}^{(n)}_{(S,T]} - \Theta_{(S,T]} = \underbrace{M_{n,T} - M_{n,S}}_{\text{semimartingale}} + \underbrace{\tilde{e}_{n,T} - e_{n,S}}_{\text{edge effects}} \quad \text{for } S < T \in \mathcal{T}_n, \tag{2}$$

where $\mathcal{T}_n = \{T_{n,i} : i = 0, \ldots, B_n\}$. The edge effect is essentially anything that disarranges the semimartingaleness of the difference $\hat{\Theta}_{(0,T]} - \Theta_{(0,T]}$, and it occurs in many shapes. The edge effect has a component e_S relating to phasing in the estimator at the beginning of the time interval, and component $\tilde{e}_T$ for the phasing out at T. (For examples, see *ibid.*, Remark 5, p. 206, and Section 7, p. 219–226.) In the presence of microstructure noise, no estimator is without edge effect.

Our concern in this paper is that the edge effects can be fairly large, and even if they are negligible from a theoretical standpoint, it is wiser to not neglect them in actual data. As example in this paper, we shall use the non-tapered smoothed TSRV, whose edge effects are small, but not too small (Appendix A.1 in Mykland et al., 2019). Tapering will make the edge effect smaller (Mykland et al., 2019), and also Kalnina and Linton (2008).

In the following, we review earlier results and introduce the quadratic variations $QV_{B,K}$, as well as the two-scales observed AVAR and the two-scales volatility estimate of the spot parameter θ_t (Section 2). We then present the concept of hard edge and find an expansion for $QV_{B,K}$ for this more difficult case (Section 3). This gives rise to new and more robust regression (or multi-scale) estimators of AVAR and the volatility of the spot parameter (Section 4). The question of using soft vs. hard edge assumptions is then discussed through a data example (Section 5, trade data for the S&P 500 E-mini futures on the Chicago Mercantile Exchange). As applications, we show how to optimize tuning parameters (Section 6) and we show how to set standard errors for the nearest neighbor truncation estimator (Section 7). A simulation experiment is reported in Section 8.

2. The observed asymptotic variance in high frequency data: Review of earlier findings

As in the earlier paper, we set

Definition 1 (*Rolling Quadratic Variations of Integrated Processes*). Divide the time interval $[0, \mathcal{T}]$ into B basic blocks of time periods (days, five minutes, thirty seconds, or other) $(T_{i-1}, T_i]$ from $T_0 = 0$ to $T_B = \mathcal{T}$. The blocks are assumed to be of equal size[2]: Set $\Delta T = \mathcal{T}/B$, and assume that $T_i = i\Delta T$. We shall permit rolling overlapping intervals, and so let K be a number no greater than B. We define

$$\text{The quadratic variation of } \Theta\text{: } QV_{B,K}(\Theta) = \frac{1}{K}\sum_{i=K}^{B-K}(\Theta_{(T_i,T_{i+K}]} - \Theta_{(T_{i-K},T_i]})^2, \text{ and}$$

$$\text{The quadratic variation of } \widehat{\Theta}\text{: } QV_{B,K}(\hat{\Theta}) = \frac{1}{K}\sum_{i=K}^{B-K}(\hat{\Theta}_{(T_i,T_{i+K}]} - \hat{\Theta}_{(T_{i-K},T_i]})^2 \tag{3}$$

We emphasize that the above quadratic variations are defined on the discrete grid $\{0, \Delta T, 2\Delta T, \ldots, \mathcal{T}\}$, as opposed to the continuous-time quadratic variation $[X, X]_t$ of a semi-martingale (X_t). The quantities B, ΔT, and K depend (explicitly or implicitly) on the index n, which usually denotes the number of observations. We may then write $\Delta T = \Delta T_n$, or omit the index n if the meaning is obvious.

A main condition is the following, cf. Section 3.1 in the earlier paper for background and implications, and Section 7 for examples.

Condition 1 (*Standard Convergence Result in the Literature*). *Assume (2), and that one can show the following. There is an $\alpha > 0$ so that as $n \to \infty$,*

$$n^{\alpha} M_{n,t} \xrightarrow{\mathcal{L}} L_t \text{ stably in law} \tag{4}$$

with respect to a sigma-field $\mathcal{G}$.[3]

The quadratic variation $[L, L]_{\mathcal{T}}$ is measurable with respect to $\mathcal{G}$, and L_t is a local martingale conditionally on $\mathcal{G}$. Also, $e_{n,T_n} = o_p(n^{-\alpha})$ and $\tilde{e}_{n,S_n} = o_p(n^{-\alpha})$ for any $S_n, T_n \in \mathcal{T}$. Finally, the sequence $n^{\alpha}M_{n,t}$ is Predictably Uniformly Tight (P-UT) (Jacod and Shiryaev, 2003, Chapter VI.3.b, and Definition VI.6.1, p. 377).

2 See Mykland and Zhang (2017a, Sections 5.2 and 6, pp. 215–216, and 218–219) for a more general formulation.

3 See Definition 3 (p. 207) of Mykland and Zhang (2017a).

The development in Mykland and Zhang (2017a) was based on the expansion

$$QV_{B,K}(\hat{\Theta}) \;=\; 2AVAR_n \;+\; \frac{2}{3}(K_n\Delta T_n)^2[\theta,\theta]_{\mathcal{T}-} \;+\; o_p((K_n\Delta T_n)^2) + o_p(n^{-2\alpha}) \tag{5}$$

where AVAR_n is the asymptotic variance of $\widehat{\Theta}_n = \hat{\Theta}^{(n)}_{(0,\mathcal{T}]}$. The expansion is valid under "soft edge" conditions.

This leads to the definition of Two Scales AVAR, Volatility of Spot θ, and standard error:

$$\text{TSAVAR}_n = \frac{1}{2}\left(\frac{1}{K_1^2}-\frac{1}{K_2^2}\right)^{-1}\left(\frac{1}{K_1^2}QV_{B,K_1}(\hat{\Theta}) - \frac{1}{K_2^2}QV_{B,K_2}(\hat{\Theta})\right) \text{ and} \tag{6}$$

$$\widehat{[\theta,\theta]}_{\mathcal{T}-} = \frac{3}{2}(K_2^2-K_1^2)^{-1}(\Delta T)^{-2}\left(QV_{B,K_2}(\hat{\Theta}) - QV_{B,K_1}(\hat{\Theta})\right), \tag{7}$$

as well as $\text{se}(\hat{\Theta}_n) = |\text{TSAVAR}_n|^{\frac{1}{2}}$ The consistency of the two-scales constructions was guaranteed by Theorem 4 in the earlier paper. The two scales estimators TSAVAR_n and $\widehat{[\theta,\theta]}_{\mathcal{T}-}$ satisfy an empirical decomposition similar to (5), *cf.* Fig. 1:

$$QV_{B,K}(\hat{\Theta}) \;=\; 2\;\text{TSAVAR}_n \;+\; \frac{2}{3}(K\Delta T)^2\widehat{[\theta,\theta]}_{\mathcal{T}-}\;,\;\; i=1,2. \tag{8}$$

3. A general expansion result for $QV_{B,K}(\widehat{\Theta}_n)$ under hard edge effects

One cannot always take the edge effect to be negligible in the sense of (29) in Theorem 4 of Mykland and Zhang (2017a). We shall see that this will give rise to an extra term in the expansion of $QV_{B,K}(\hat{\Theta})$, one due to the edge effects e_t and/or $\tilde{e}_t$. Instead of a two scales estimator, we shall require a linear combination of three or more scales K, in other words a multi-scale $\widehat{\text{AVAR}}_n$.[4]

To warm up, we first state an expansion $QV_{B,K}(\hat{\Theta})$ which permits larger edge effects than Theorem 3 in the earlier paper. We make the following set of assumptions.

Condition 2 (*Hard Edge Assumptions*)**.** *Suppose that there is an integer J_n for which $e_{n,T_{n,i}} = e'_{n,T_{n,i}} + e''_{n,T_{n,i}}$ and $\tilde{e}_{n,T_{n,i}} = \tilde{e}'_{n,T_{n,i}} + \tilde{e}''_{n,T_{n,i}}$, so that $(e'_{n,T_{n,i}}, \tilde{e}'_{n,T_{n,i}})$ are $\mathcal{F}_{T_{n,i+J_n}}$-measurable,[5] and for which $E(e'_{n,T_{n,i}} \mid \mathcal{F}_{T_{n,i-J_n}}) = E(\tilde{e}'_{n,T_{n,i}} \mid \mathcal{F}_{T_{n,i-J_n}}) = 0$ and where $\sum_i (e''_{T_{n,i}})^2 = o_p(n^{-2\alpha})$ and $\sum_i(\tilde{e}''_{T_{n,i}})^2 = o_p(n^{-2\alpha})$. Also suppose that, for all i, $E(e'_{n,T_{n,i}})^2 < \infty$ and $E(\tilde{e}'_{n,T_{n,i}})^2 < \infty$, and that*

$$\sup_n E\; n^{\alpha}\left(\max_{0\le i\le B_n}|e'_{n,T_i}| + \max|\tilde{e}'_{n,T_i}|\right) \;<\; \infty. \tag{9}$$

In other words, we let the edge effects be larger, but they must have more structure and uniformity. As a complement, Condition 2 is argued from a mixing perspective in Appendix B.1. We recall that we assume (Condition 1) that each e_{T_i} and $\tilde{e}_{T_i}$ is of order $o_p(n^{-\alpha})$, so that assumption (9) refers only to the tail behavior of the edge effects.

The edge effects are now potentially the dominating terms in the expansion of $QV_{B,K}(\hat{\Theta})$. Define autocovariances

$$C^{ab}_{n,K} \;=\; \begin{cases} \frac{1}{B_n}\sum_{i=K}^{B_n} \tilde{e}_{n,T_{n,i}}\tilde{e}_{n,T_{n,i-K}} & \text{for } (a,b)=(1,1) \\ \frac{1}{B_n}\sum_{i=K}^{B_n} \tilde{e}_{n,T_{n,i}}e_{n,T_{n,i-K}} & \text{for } (a,b)=(1,2) \\ \frac{1}{B_n}\sum_{i=K}^{B_n} e_{n,T_{n,i}}e_{n,T_{n,i-K}} & \text{for } (a,b)=(2,2) \end{cases} \tag{10}$$

The aggregated main and lagged edge effects are now given by, respectively,

$$\text{MAEE}_n = C^{11}_{n,0} + C^{12}_{n,0} + C^{22}_{n,0} = \frac{1}{B_n}\sum_{i=0}^{B_n}(\tilde{e}^2_{T_i} + e^2_{T_i} + \tilde{e}_{T_i}e_{T_i})$$

$$\varepsilon_{n,K} = -\left(C^{11}_{n,K} + 2C^{12}_{n,K} + C^{22}_{n,K}\right) + C^{12}_{n,2K} \tag{11}$$

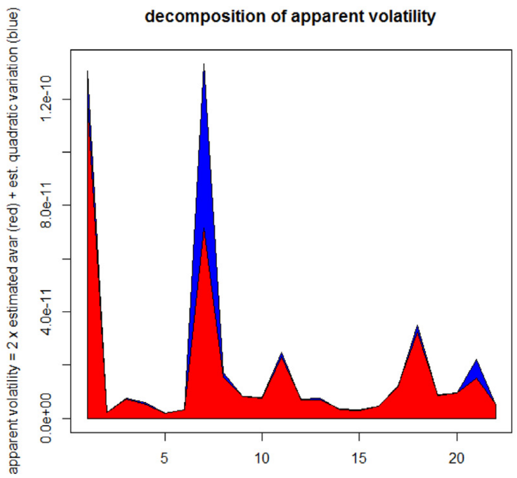
The following is our main result for large edge effects, which parallels the earlier result on small edge effects in Mykland and Zhang (2017a, Theorem 3, pp. 208–209).

We discuss the behavior of the aggregated edge effects. We then seek linear combinations of $QV_{B,K}(\hat{\Theta})$ to remove the edge effects.

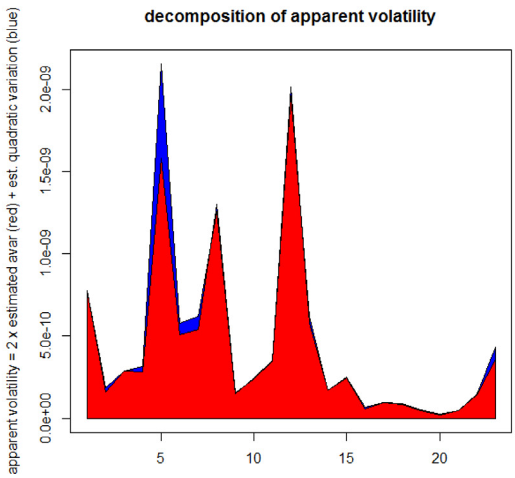
---

[4] This is comparable to the extension from Zhang et al. (2005) to Zhang (2006). Though the edge effects resemble microstructure, the parallel should not be taken too far, since two scales are normally required even in the case where the edge effect is negligible. The conceptual similarity between edge effects and microstructure noise does, however, suggest that edge effects can be more safely ignored for larger values of ΔT, in analogy with the findings for microstructure in Aït-Sahalia and Xiu (2019). Bear in mind, however, that one should exercise particular caution in high dimension (Chen et al., 2020, Section 6-7). A detailed investigation is beyond the scope of this paper.

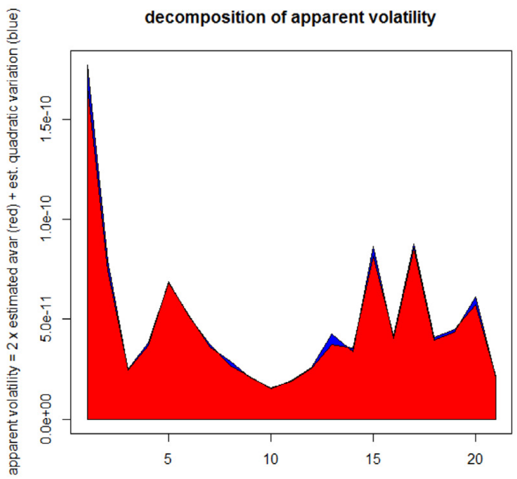
[5] In other words, we allow the edge effect to depend on the future. This would, for example, be relevant for the Backward Estimators discussed in Section 5.1 of Mykland and Zhang (2017a).



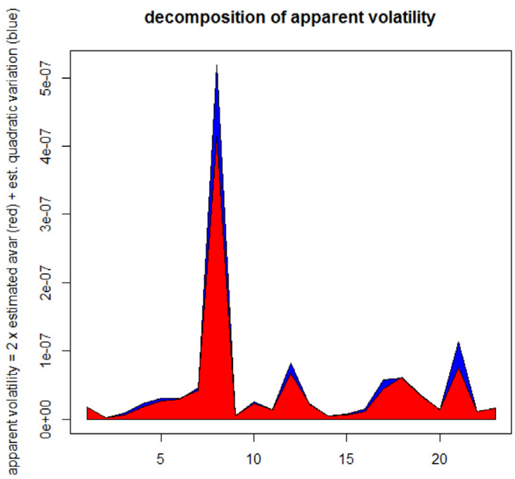
(a) May 2007



(b) August 2007



(c) May 2008



(d) October 2008

Fig. 1. These plots illustrate the empirical decomposition (8) of the apparent volatility $QV_{B,K}(\hat{\Theta})$ in data, for the S&P 500 E-mini futures as traded on the Chicago Mercantile Exchange, for the trading days of four months in 2007–08. The total curve is the total volatility $QV_{B,K}$ for each day, the red part is $2 \times$ TSAVAR for each day, and the blue part is $\frac{2}{3}(K_1 \Delta T)^2 \widehat{[\theta,\theta]}_{\mathcal{T}-}$, as given in Section 2. In the estimation, the underlying parameter is the spot volatility: $\theta_t = \sigma_t^2$. $\widehat{\Theta}_{(S,T]}$ is based on the S-TSRV (Mykland et al., 2019): first pre-average the data (trade prices) to 15 s, and then compute a TSRV on these pre-averages, with $j = 20$ and $k = 40$. (This is the non-tapered version of the S-TSRV.) The estimator is thus of integrated volatility $\Theta_{(S,T]} = \int_S^T \sigma_t^2 dt$, and $[\theta,\theta]_{\mathcal{T}} = [\sigma^2,\sigma^2]_{\mathcal{T}}$. For $QV_{B,K}(\hat{\Theta})$, we take ΔT to also be fifteen seconds (the smallest meaningful value). The TSAVAR (6), is computed with $K_1 = 20$ and $K_2 = 40$, and similarly for $\widehat{[\theta,\theta]}_{\mathcal{T}}$ from (7). In other words, $K_1 \Delta T$ is a rolling five minute period that ends every 15 s, using the forward half interval method (Mykland and Zhang, 2017a, Section 5.1, p. 215). Daily volumes are reported in Table 1.

Theorem 1 (*Representation of $QV_{B,K}(\hat{\Theta})$*)**.** *Suppose θ_t is a semimartingale, and that Conditions 1–2 hold. Assume the balance condition.*[6]

$$K_n \Delta T_n \text{ are of the same order as } n^{-\alpha}. \tag{12}$$

[6] This is in analogy with the discussion in Mykland and Zhang (2017a, Remark 7(iv), p. 211) There may be other approaches, but this is beyond the scope of the present paper.

Table 1
Daily volumes for Figs. 1 and 2.

	Min	Max	Mean	Median
May 2007	37 000	111 300	69 919	67 130
Aug 2007	99 310	447 200	226 800	197 700
May 2008	95 220	179 700	138 700	135 700
October 2008	289 400	981 100	547 400	562 600

Summary statistics for the daily volume of trades (M6 messages) for the S&P E-mini futures, for the four months in Figs. 1 and 2. Min is the volume on the least active trading day, and similarly for max, mean, and median. Quote volumes are 5–10 times higher.

Also assume that

$$J_n \Delta T_n = o_p(n^{-\alpha}). \tag{13}$$

Then

$$QV_{B_n,K_n}(\hat{\Theta}) = 2\text{AVAR}_n + \frac{2}{3}(K_n\Delta T_n)^2[\theta,\theta]_{\mathcal{T}-} + 2\mathcal{T}(K_n\Delta T_n)^{-1}\text{MAEE}_n + 2\mathcal{T}(K_n\Delta T_n)^{-1}\varepsilon_{n,K_n} + o_p(n^{-2\alpha}). \tag{14}$$

Proof of Theorem 1. See Appendix A.1.

Our strategy in the following will be to use linear combinations to remove the main edge effect term MAEE_n, but to live with the lagged term ε_{n,K_n}. We pursue this further in the next section, but lay the groundwork in a further analysis of the edge effects and their magnitude.

Proposition 1 (*Behavior of Aggregated Edge Effects*)**.** *Assume the conditions of Theorem 1. Then,*

$$2\mathcal{T}(K_n\Delta T_n)^{-1}\text{MAEE}_n = o_p(n^{-\alpha}) \text{ or less.} \tag{15}$$

Also assume that $K_n \geq 2J_n$. Then

$$\begin{aligned} 2\mathcal{T}(K_n\Delta T_n)^{-1}\varepsilon_{n,K_n} &= O_p\left(n^{\alpha}(J_n\Delta T_n)^{1/2}\text{VAEE}_n^{1/2}\right) \\ &= o_p\left(n^{-\alpha}(J_n\Delta T_n)^{1/2}\right) \text{ or less,} \end{aligned} \tag{16}$$

where

$$\begin{aligned} \text{VAEE}_n &= \frac{1}{B_n}\sum_{i=0}^{B_n}\left(E((\tilde{e}'_{n,T_{n,i}})^2|\mathcal{F}_{T_{i-2J_n}})^2 + E((e'_{n,T_{n,i}})^2|\mathcal{F}_{T_{i-2J_n}})^2\right) \\ &= o_p\left(n^{-4\alpha}\right) \text{ or less.} \end{aligned} \tag{17}$$

More generally, if $K_n \geq 2J_n$

$$(C^{11}_{n,K}, C^{12}_{n,K}, C^{22}_{n,K}) = O_p\left((J_n\Delta T_n\text{VAEE}_n)^{1/2}\right). \tag{18}$$

Also, if $2J_n \leq K_{n,1} < K_{n,2} < \cdots K_{n,m}$, with $K_{n,l+1} - K_{n,l} \geq 2J_n$ for each l, then $(J_n\Delta T_n\text{VAEE})^{-1/2}(C^{11}_{n,K_l}, C^{12}_{n,K_l}, C^{22}_{n,K_l})$, $l = 1, \ldots, m$, are asymptotically uncorrelated.

Proof of Proposition 1. See Appendix B.2.

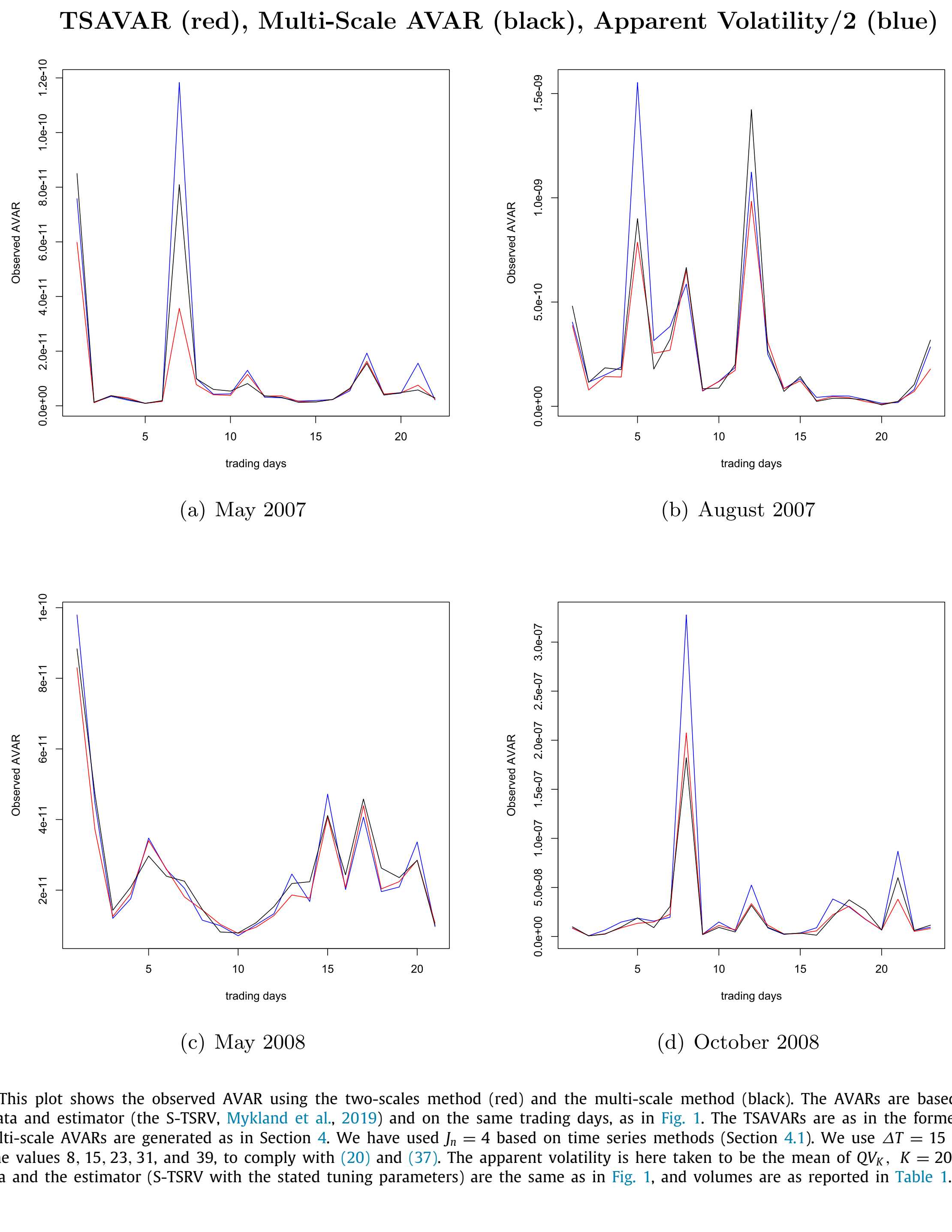
Discussion of the Impact of Edge Effects in Theorem 1. Proposition 1 permits a discussion of the behavior of aggregated edge effects in Theorem 1. First, from (15), the main edge effect MAEE can be as large as $o_p(n^{-\alpha})$, and so could easily overshadow the AVAR_n and $[\theta,\theta]_{\mathcal{T}-}$ terms in (14). Obviously, MAEE may be smaller, and if MAEE $= O(n^{-3\alpha})$, we retrieve the result in Theorem 3 in Mykland and Zhang (2017a) in the balanced case (12).[7]

Second, the lagged edge effect (16) ought to be of order $o_p(n^{-2\alpha})$ so as to not dominate AVAR_n and $[\theta,\theta]_{\mathcal{T}-}$ in the representation (14). In other words, from (16), we require

$$(J_n\Delta T_n)\text{VAEE}_n = o_p(n^{-6\alpha}). \tag{19}$$

The mathematically simplest path would be to require that $J_n\Delta T_n = O_p(n^{-2\alpha})$, but this depends on the bandwidth of the time-dependence of the edge effects, and thus both on the data and on the specific estimator.

[7] This is because MAEE is of the same order as $\text{ave}(e^2_{T_i}) + \text{ave}(\tilde{e}^2_{T_i})$ in the notation of (21) in the earlier paper.



(a) May 2007 (b) August 2007

(c) May 2008 (d) October 2008

Fig. 2. This plot shows the observed AVAR using the two-scales method (red) and the multi-scale method (black). The AVARs are based on the same data and estimator (the S-TSRV, Mykland et al., 2019) and on the same trading days, as in Fig. 1. The TSAVARs are as in the former figure. The multi-scale AVARs are generated as in Section 4. We have used $J_n = 4$ based on time series methods (Section 4.1). We use $\Delta T = 15$ s, and K takes the values 8, 15, 23, 31, and 39, to comply with (20) and (37). The apparent volatility is here taken to be the mean of QV_K, $K = 20, \ldots, 40$. The data and the estimator (S-TSRV with the stated tuning parameters) are the same as in Fig. 1, and volumes are as reported in Table 1.

Alternatively, if, say, in an average and fairly uniform sense, the e_{T_i} and $\tilde{e}_{T_i}$ are of order $O_p(n^{-\beta})$, the lagged edge effect (16) is of order $O_p(n^{\alpha-2\beta}(J_n\Delta T_n)^{\frac{1}{2}}) = O_p(n^{\frac{1}{2}\alpha-2\beta})$ under the conditions of Theorem 1. Thus the lagged edge effect will disappear if $\beta > \frac{5}{4}\alpha$. This is in practice much easier to verify than Theorem 3 in Mykland and Zhang (2017a), which would require $\beta > \frac{3}{2}\alpha$. Thus the range where Theorem 1 is effective is $\beta \in (\frac{5}{4}\alpha, \frac{3}{2}\alpha]$, and possibly including larger values of β if $J_n\Delta T_n$ is small. For comparison, under the same assumptions, and with β in this interval, the main edge effect $2\mathcal{T}(K_n\Delta T_n)^{-1}\text{MAEE}_n = O_p(n^{\alpha-2\beta})$, which dominates the AVAR_n and $[\theta,\theta]_{\mathcal{T}-}$ terms in (14).

Finally, note that while the condition $K_{n,1} \geq 2J_n$ is there to prevent bias in the AVAR estimator, the rôle of $K_{n,l+1} - K_{n,l} \geq 2J_n$ is only to guarantee asymptotic uncorrelatedness of the $(J_n\Delta T_n\text{VAEE})^{-1/2}(C^{11}_{n,K_l}, C^{12}_{n,K_l}, C^{22}_{n,K_l})$, which is not crucial to the procedure.

We shall leave the question of the precise size of the lagged edge effect open, so as to have an incentive to minimize this term. We shall do this next.

4. Estimation of AVAR and $[\theta,\theta]_{\mathcal{T}-}$ under hard edge: Multi-scale and regression estimation

We proceed through linear combinations of $QV_{B_n,K_n}(\hat{\Theta})$ over m scales, *i.e.*,

$$2J_n \le K_{n,1} < K_{n,2} < \cdots K_{n,m}. \tag{20}$$

It will be convenient to rescale[8] so that $QV^{(R)}_{B_n,K}(\hat{\Theta}) = QV_{B_n,K}(\hat{\Theta})(K\Delta T_n)$, and define a multi-scale estimator of the form

$$\begin{aligned} MSQV_n(\hat{\Theta}) &= \sum_{l=1}^{m} g_{n,l} QV^{(R)}_{B_n,K_l}(\hat{\Theta}) \\ &= \underline{g}_n^* \mathbb{Y}_n \end{aligned} \tag{21}$$

where $\underline{g}_n^* = (g_{n,1},\ldots,g_{n,m})$ is a vector of coefficients to be determined ("$*$" denotes transpose), and where

$$\mathbb{Y}_n^* = \begin{pmatrix} QV^{(R)}_{B_n,K_{n,1}}(\hat{\Theta}) & QV^{(R)}_{B_n,K_{n,2}}(\hat{\Theta}) & \cdots & QV^{(R)}_{B_n,K_{n,m}}(\hat{\Theta}) \end{pmatrix}. \tag{22}$$

Also set

$$\underline{\beta}_n^* = \begin{pmatrix} \mathrm{MAEE}_n, \mathrm{AVAR}_n, [\theta,\theta]_{\mathcal{T}-} \end{pmatrix}, \tag{23}$$

$$\mathbb{X}_n^* = \begin{pmatrix} 2\mathcal{T} & 2\mathcal{T} & \cdots & 2\mathcal{T} \\ 2(K_{n,1}\Delta T_n) & 2(K_{n,2}\Delta T_n) & \cdots & 2(K_{n,m}\Delta T_n) \\ \frac{2}{3}(K_{n,1}\Delta T_n)^3 & \frac{2}{3}(K_{n,2}\Delta T_n)^3 & \cdots & \frac{2}{3}(K_{n,m}\Delta T_n)^3 \end{pmatrix}, \text{ and} \tag{24}$$

$$\underline{\varepsilon}_n^* = \begin{pmatrix} \varepsilon_{n,K_1} & \varepsilon_{n,K_2} & \cdots & \varepsilon_{n,K_m} \end{pmatrix}. \tag{25}$$

Theorem 1 and Proposition 1 then yield, subject to (20), that

$$\begin{aligned} \mathbb{Y}_n &= \mathbb{X}_n \underline{\beta}_n + 2\mathcal{T}\underline{\varepsilon}_n + o_p(n^{-3\alpha}) \\ &= \mathbb{X}_n \underline{\beta}_n + O_p\left(n^{-\alpha}(J_n\Delta T_n \mathrm{VAEE}_n)^{1/2}\right) + o_p(n^{-3\alpha}), \end{aligned} \tag{26}$$

the second line provided $K_{n,l+1} - K_{n,l} \ge 2J_n$ for each $l \in [1, m-1]$.

Regression Interpretation of (26). Our whole notation, and the first line of (26), suggests linear regression. Ordinary least squares (OLS) in the regression of $\mathbb{Y}$ on $\mathbb{X}$ from (22)–(24) yields

$$\hat{\underline{\beta}}_n = (\widehat{\mathrm{MAEE}}_n, \widehat{\mathrm{AVAR}}_n, \widehat{[\theta,\theta]}_{\mathcal{T}-})^*, \text{ where } \hat{\underline{\beta}}_n = (\mathbb{X}_n^*\mathbb{X}_n)^{-1}\mathbb{X}_n^*\mathbb{Y}_n. \quad \square \tag{27}$$

Multi-Scale Interpretation of (26). Consider the following constraint on (21):

$$\mathbb{X}_n^* \underline{g}_n = \underline{b} \tag{28}$$

where $\underline{b} = (0, b_1, b_2)^*$. Then (21) and the second line of (26) yields

$$MSQV_n(\hat{\Theta}) = b_1\mathrm{AVAR}_n + b_2[\theta,\theta]_{\mathcal{T}-} + O_p\left(n^{-\alpha}(J_n\Delta T_n\mathrm{VAEE}_n)^{1/2}\mathfrak{E}_n^{1/2}\right) + o_p(n^{-3\alpha}\mathfrak{E}_n^{1/2}) \tag{29}$$

with $\mathfrak{E}_n = \underline{g}^*\underline{g}$. Hence,

- To estimate $\mathrm{AVAR}(\widehat{\Theta}_n)$, choose

$$\underline{b} = (0, 1, 0)^*. \tag{30}$$

- To estimate the quadratic variation $[\theta,\theta]_{\mathcal{T}-}$, choose

$$\underline{b} = (0, 0, 1)^*. \tag{31}$$

To minimize the error in (29), one solves the optimization problem

$$\min \underline{g}_n^*\underline{g}_n \text{ subject to } \mathbb{X}_n^*\underline{g}_n = \underline{b}. \tag{32}$$

The standard solution (*e.g.*, Boyd and Vandenberghe, 2004, p. 304) to (32) is $\underline{g}_n = \mathbb{X}_n(\mathbb{X}_n^*\mathbb{X}_n)^{-1}\underline{b}$. For this value of $\underline{g}_n$,

$$MSQV_n(\hat{\Theta}) = \underline{g}_n^*\mathbb{Y} = \underline{b}^*(\mathbb{X}^*\mathbb{X})^{-1}\mathbb{X}^*\mathbb{Y} = \underline{b}^*\hat{\beta}. \quad \square \tag{33}$$

[8] See the discussion after Theorem 2. The rescaling is without loss of generality.

Hence the regression and Multi-Scale approaches coincide. Does the solution work?

From the theoretical standpoint, consistency is backed by a theorem. It also holds in the soft edge case.

Theorem 2 (*Consistency of* $\widehat{\text{AVAR}}_n$ *and* $\widehat{[\theta,\theta]}_{\mathcal{T}-}$ *in Both the Soft and Hard Edge Cases*). *Suppose θ_t is a semimartingale, and that Condition 1 holds. Let $K_{n,1},\ldots,K_{n,m}$ satisfy (20), Suppose that $K_{n,1}$ satisfies the balance condition (12), and that there are constants c_- and c_+, with $1 < c_- < c_+ < \infty$, for which*

$$c_- \le \frac{K_{n,m}}{K_{n,1}} \le c_+. \tag{34}$$

Suppose that either (i) Mykland and Zhang (2017a, eq. (29), p. 209) holds (soft edge case), or (ii) Conditions 1–2 (in this paper) hold, with (13) and *(19) (hard edge case).*

Let $\widehat{\text{AVAR}}_n$ and $\widehat{[\theta,\theta]}_{\mathcal{T}-}$ be given by (27). Then

$$\widehat{\text{AVAR}}_n \;=\; \text{AVAR}_n(1+o_p(1)) \text{ and } \widehat{[\theta,\theta]}_{\mathcal{T}-} \;=\; [\theta,\theta]_{\mathcal{T}-}(1+o_p(1)). \tag{35}$$

In particular, if L_T is conditionally Gaussian given $\mathcal{G}$, then

$$\frac{\widehat{\Theta}_n - \Theta}{\text{se}(\hat{\Theta}_n)} \overset{\mathcal{L}}{\to} N(0,1) \text{ stably in law}. \tag{36}$$

Proof. See Appendix B.3.

From the practical standpoint, the rescaling $QV^{(R)}_{B_n,K}(\hat{\Theta}) = QV_{B_n,K}(\hat{\Theta})(K\Delta T_n)$ achieves two things. On the one hand, the $\varepsilon_{n,K_{n,l}}$ term will often be close to homoscedastic (see Proposition 1 and its proof). In the multi-scale formulation, this manifests itself in the form of the remainder term $\mathfrak{E}_n = \underline{g}^*\underline{g}$. In addition, the rescaling turns the main edge effect MAEE_n into an intercept term. This is computationally advantageous since $\widehat{\text{AVAR}}_n$ and $\widehat{[\theta,\theta]}_{\mathcal{T}-}$ can now be calculated without the contribution of MAEE_n, *cf.*, Weisberg (1985, Chapter 2.2, p. 43-44). See also the proof of Theorem 2 in Appendix B.3.

While the $\varepsilon_{n,K}$ may be close to homoscedastic, they are not independent. A first order solution lies in requiring that $K_{n,l+1} - K_{n,l} \ge 2J_n$ in (20). This assures the second line in (26). From definition (11), however, the $\varepsilon_{n,K}$ are dependent. For example, $\varepsilon_{n,K}$ and $\varepsilon_{n,2K}$ contain a shared autocovariance $C^{12}_{n,2K}$. One solution to this is to require that the $K_{n,l}$ satisfy

$$\{K_{n,l}:\ l=1,\ldots,m\} \ \cap\ \{2K_{n,l}:\ l=1,\ldots,m\} \ =\ \emptyset. \tag{37}$$

This assures that the $\varepsilon_{n,K_{n,l}}$ are asymptotically uncorrelated, in view of Proposition 1. In particular, (37) holds if one only uses odd $K_{n,l}$, say,

$$K_{n,l} = (2l+2p-1)K. \tag{38}$$

for non-negative integer p. Even if one does not do this, the solution in Theorem 2 is consistent, and one can alternatively construct a weighted least squares procedure based on the dependence structure given by (11) and Proposition 1.

Finally, observe that if $m \to \infty$, it may be possible to get around the requirement (19), along the lines of Zhang (2006).

Remark 1. In the volatility estimation problem, a modified realized kernel estimator (Barndorff-Nielsen et al., 2008) is very similar to that of the multi-scale estimator of Zhang (2006), *cf.* Bibinger and Mykland (2016). It is conjectured that a similarly modified realized kernel approach will work also in this problem. □

Remark 2 (*A Three Scales* $\widehat{\text{AVAR}}_n$). If one uses a three-scales estimator, $m=3$, the three $g_{n,l}$ are determined by the three linear equations given through (28) and (30). The solution is

$$\begin{aligned} g_{n,1} &= -\frac{1}{v_n}(K^3_{n,3} - K^3_{n,2}),\\ g_{n,2} &= \frac{1}{v_n}(K^3_{n,3} - K^3_{n,1}), \text{ and}\\ g_{n,3} &= -\frac{1}{v_n}(K^3_{n,2} - K^3_{n,1}), \text{ where}\\ v_n &= 2\Delta T_n(K_{n,1}+K_{n,2}+K_{n,3})(K_{n,2}-K_{n,1})(K_{n,3}-K_{n,1})(K_{n,3}-K_{n,2}). \end{aligned} \tag{39}$$

4.1. Choice of $J = J_n$

The J parameter is tied up with the dependence structure of the edge effects, as spelled out in Condition 2 in Section 3. Unless there are reasons to expect non-linear dependence, the standard applied time series path to determining the length of dependence is via the autocorrelation and partial autocorrelation plots (acf, pacf) (Brockwell and Davis, 1987; Chan, 2011).

One can most straightforwardly use the pacf directly on $\widehat{\Theta}_{T_i}$, and subject to regularity conditions, infer strong mixing from Withers (1981), from which in turn Condition 2 is satisfied with the help of Appendix B.1 in this paper. It is important to note that the first order autocorrelation will be close to one, since $\widehat{\Theta}_{T_i}$ contains Θ_{T_i}, but the higher order partial autocorrelations seem to mainly reflect the edge effects. For the same reason, however, the acf is not a meaningful diagnostic when applied to this problem.

Another path to determining J is to use the acf or pacf on $\widehat{\Theta}_{T_i} - \widehat{\Theta}_{T_{i-1}}$. In view of the representation (2), this is close to analyzing $\tilde{e}_{n,T_i} - \tilde{e}_{n,T_{i-1}}$ (for a forward estimator, Mykland and Zhang, 2017a, Section 5.1).

The situation is similar to choosing the shorter scale j in a TSRV or S-TSRV (Zhang et al., 2005; Mykland et al., 2019). It may also be possible to use a signature plot to select J, but this question is beyond the scope of this paper. □

5. Two- *vs.* multi-scale estimators

There remains a choice between

i. The Two-scales estimator (6). Consistency is assured by Mykland and Zhang (2017a, Theorem 4 in Section 3.2).
ii. The Multi-scale estimator, from Section 4, as given by (27). Theorem 2 assures consistency.

There are two paths: one can either check the theoretical conditions, or, at least for a *prima facie* impression, try both estimators on the data. We show an example of the latter approach in Fig. 2. For these data, the choice of method does not substantively affect the estimates. *We emphasize that the multi-scale estimator from Section 4 is valid under both soft and hard edge conditions (Theorem 2). In fact, TSAVAR is a special case when using only K_1 and K_2.* The difference between TSAVAR and the regression/multiscale AVAR is substantial only on a few days. This motivates that one should try both the soft- and hard-edge procedures on actual data.

5.1. Applying $\widehat{\text{AVAR}}$

After the above, one is in possession of an estimate $\widehat{\text{AVAR}}$. Subject to the regularity conditions imposed, this estimator is consistent in the sense of Section 2. In particular, under the conditions of Theorem 2, if $\text{se}(\hat{\Theta}_n) = |\widehat{\text{AVAR}_n}|^{\frac{1}{2}}$, then

$$\frac{\widehat{\Theta}_n - \Theta}{\text{se}(\hat{\Theta}_n)} \xrightarrow{\mathcal{L}} N(0,1) \text{ stably in law.} \tag{40}$$

6. Selection of tuning parameters in the hard edge case

It frequently occurs that estimators depend on tuning parameters. In the following, we suppose that the estimator and its asymptotic variance have the form $\hat{\Theta}_{n,c}$ and $\text{AVAR}_n = \text{AVAR}_{n,c}$, where c is the tuning parameter. Our purpose is to choose c so that $\hat{\Theta}_{n,c}$ has minimal asymptotic variance.

For example, in the case of the S-TSRV (Mykland et al., 2019), there is a need to determine the two scale parameters j and k, as well and the length of the data smoothing window. It would be standard procedure to determine the shorter scale j by the same time series methods as in Section 4.1. The longer scale k may be determined by a volatility signature plot.

A more direct path to an optimal estimator, however, is to find a k which approximately minimizes AVAR.

In the more general case, call the tuning parameter c (so $c = k$ for the S-TSRV). It is shown in Mykland and Zhang (2017a, Section 4) that under regularity conditions, in particular that c take values in a finite set (and this is the case for the S-TSRV when computed from intervals of size $\propto n^{1/2}$), the choice $\hat{c}_n = \arg\min_c \widehat{\text{AVAR}}_{n,c}$ results in an estimator $\widehat{\Theta}_{n,\hat{c}}$ with minimal asymptotic variance, and where the asymptotic normality is still valid. In other words, $(\widehat{\Theta}_{n,\hat{c}} - \Theta)/|\widehat{\text{AVAR}}_{n,\hat{c}}|^{1/2} \xrightarrow{\mathcal{L}} N(0,1)$ stably in law. We emphasize that this is not generally true without the conditions stated.

We shall here see that it is not actually necessary to estimate AVAR to obtain the optimal c, so long as the $[\theta,\theta]$ component remains stable in c (which is, at least, true asymptotically). In view of the development above, one may still use a multi-scale estimator from Section 4. This time, however, only two constraints are needed. The criterion $MSQV(\hat{\Theta}_{n,c})$ is obtained by minimizing $\mathfrak{C}_n = \underline{g}^* \underline{g}$ (from Section 4) subject to

$$\sum_{l=1}^{m} g_{n,l} = 0 \text{ and } \sum_{l=1}^{m} g_{n,l}(K_{n,l}\Delta T_n) = 1\ . \tag{41}$$

In analogy with (33), the resulting $MSQV(\hat{\Theta}_{n,c})$ is the estimated slope in the regression of $\mathbb{Y}$ on the two last columns of $\mathbb{X}$ (from Eqs. (22) and (24)). By standard regression considerations, this estimated slope equals $\widehat{AVAR}_{n,c} + \mathfrak{r}_n \widehat{[\theta,\theta]}_{\mathcal{T}-,c}$, where $\mathfrak{r}_n$ is spelled out in (B.8) in Appendix B.3. In this Appendix, we show the following proposition. Recall that $(\bar{K}\Delta T_n)^2$ is of the same order as $n^{-2\alpha}$.

For the proposition below, assume that $\text{AVAR}_{n,c} = n^{-2\alpha}V_c$.[9] We set $c^* = \arg\min_c \text{AVAR}_{c\in\mathcal{C}}$, which we for simplicity of discussion take to be unique. $\mathcal{C}$ is a finite set of values for the tuning parameters within which one wishes to optimize. As noted above, this is realistic, for example, in the case of the S-TSRV (Mykland et al., 2019).

Proposition 2 (*Asymptotic Validity of Simplified Optimization Procedure*). *Let $MSQV(\hat{\Theta}_{n,c})$ be as described, and assume that the conditions of Theorem 2 are satisfied. Let $\bar{K}_n$ be the mean of the $K_{n,l}$. Then, as $n \to \infty$,*

$$MSQV(\hat{\Theta}_{n,c}) = \left(AVAR_{n,c} + \mathfrak{r}_n(\bar{K}\Delta T_n)^2[\theta,\theta]_{\mathcal{T}-}\right) \times (1 + o_p(1)), \tag{42}$$

where $\mathfrak{r}_n$ is of exact order $O(1)$ and does not depend on c; the formula is given in (B.8) in Appendix B.3. Also, if the definition of $\hat{c}_n$ is changed to $\hat{c}_n = \arg\min_{c\in\mathcal{C}} MSQV(\hat{\Theta}_{n,c})$, then $\hat{c}_n \to c^*$, and $(\hat{\Theta}_{n,\hat{c}_n} - \Theta)/\text{AVAR}_{c^*}^{1/2} \xrightarrow{\mathcal{L}} N(0,1)$ and $(\hat{\Theta}_{n,\hat{c}_n} - \Theta)/\widehat{\text{AVAR}}_{n,\hat{c}_n}^{1/2} \xrightarrow{\mathcal{L}} N(0,1)$, both stably.

7. A new application: Nearest neighbor truncation

To illustrate the ease with which the current theory can be applied to a new problem, we consider the nearest neighbor truncation developed in the important paper by Andersen et al. (2012), where estimators are defined and studied for the case where there is no microstructure noise. See also Andersen et al. (2014) on quarticity. In both cases, pre-averaging is actually used on the data, but not taken account of in the asymptotics.

We here adapt the estimation problem from Andersen et al. (2012) to the setting where microstructure noise is present in the model. To get a point estimator, we extend their estimator with the help of pre-averaging and a two scales construction, which is straightforward. We then show that the Observed Asymptotic Variance can be used to assess the statistical error, and hence to create a feasible estimator.

Suppose for simplicity that observations are of the form $Y_{t_j} = X_{t_j} + \epsilon_j$, where the ϵ_j are i.i.d., and the efficient log price process X_t is an Itô semimartingale with finite activity jumps, as assumed by Andersen et al. (2012). Using pre-averaging, and in analogy with Eq. (4) of their paper, we consider an estimator based on

$$MedRV_{M,n} = \sum_{i=3}^{\lfloor n/M\rfloor-2} \text{med}(\Delta\bar{Y}_{M,i-2}, \Delta\bar{Y}_{M,i}, \Delta\bar{Y}_{M,i+2})^2 \tag{43}$$

where $\Delta\bar{Y}_{M,i} = \bar{Y}_{M,i} - \bar{Y}_{M,i-1}$ and $\bar{Y}_{M,i} = \frac{1}{M}\sum_{j=(i-1)M_n+1}^{iM_n} Y_{t_j}$. For simplicity, suppose that the t_j are equidistant, *i.e.*, $t_j - t_{j-1} = \Delta t = \mathcal{T}/n$ for all j.[10] The statistic $\bar{Y}_{M,i}$ is thus based on observations in the time interval $(\tau_{i-1}, \tau_i]$, where $\tau_i = iM\Delta t$, and $\Delta\tau = M\Delta t$. When taking the median, we have used every second $\Delta\bar{Y}_{M,i}$ to avoid autocorrelation. As $n \to \infty$, we let $M = M_n$, with $M_n/\sqrt{n} \to c$.

To suitably adjust (43), and to verify the conditions of our current theorems, we invoke results on contiguity for pre-averaged processes. Set $Y^c_{t_j} = X^c_{t_j} + \epsilon_j$, and similarly $\bar{Y}^c_i$, where X^c_t is the continuous part of the latent process. Following Mykland and Zhang (2016, 2020), there is a contiguous (sequence of) probability measures Q_n, and "super-blocks" of $2\mathcal{M}$ $\bar{Y}^c_i$'s, with starting points $\lambda_{n,l} = 2l\mathcal{M}M_n\Delta t$, so that, conditionally on sigma-field at the start of each block, $\Delta\tau^{-1/2}\Delta\bar{Y}^c_{l\mathcal{M}+1}, \ldots, \Delta\tau^{-1/2}\Delta\bar{Y}^c_{(l+1)\mathcal{M}}$ is a Gaussian MA(1) process with marginal variance $\frac{2}{3}\sigma^2_{\lambda_l} + 2\frac{\nu^2}{c^2\mathcal{T}}$, where $\nu^2 = \text{Var}(\epsilon)$. Thus, if $(\mathcal{F}_t)$ is the filtration generated by the X^c_ts and the ϵs,

$$E_{Q_n}\left\{\sum_{i=2l\mathcal{M}+5}^{2(l+1)\mathcal{M}-4} \text{med}(\Delta\bar{Y}^c_{M_n,i-2}, \Delta\bar{Y}^c_{M_n,i}, \Delta\bar{Y}^c_{M_n,i+2})^2 \mid \mathcal{F}_{\lambda_l}\right\} = (2\mathcal{M}-8)\Delta\tau\left(\frac{2}{3}\sigma^2_{\lambda_l} + 2\frac{\nu^2}{c^2\mathcal{T}}\right)\frac{6-4\sqrt{3}+\pi}{\pi} \tag{44}$$

in analogy with Andersen et al. (2012): if Z_1, Z_2, Z_3 are i.i.d. $N(0,1)$, then $E\text{med}(Z_1,Z_2,Z_3)^2 = (6-4\sqrt{3}+\pi)/\pi$. One now needs to dispose of the nuisance parameter ν^2. To stay in the spirit of Andersen et al. (2012), we adjust by using the MedRV, but doubling the block size: $\Delta\bar{Y}_{2M_n,i} = (\Delta\bar{Y}_{M_n,2i-1} + \Delta\bar{Y}_{M_n,2i})/2$ (which is based on observations in $(\tau_{2i-2}, \tau_{2i}]$. Now observe that, also under Q_n,

$$E_{Q_n}\left\{\sum_{i=l\mathcal{M}+3}^{(l+1)\mathcal{M}-2} \text{med}(\Delta\bar{Y}^c_{2M_n,i-2}, \Delta\bar{Y}^c_{2M_n,i}, \Delta\bar{Y}^c_{2M_n,i+2})^2 \mid \mathcal{F}_{\lambda_l}\right\} = (\mathcal{M}-4)(2\Delta\tau)\left(\frac{2}{3}\sigma^2_{\lambda_l} + 2\frac{\nu^2}{(2c)^2\mathcal{T}}\right)\frac{6-4\sqrt{3}+\pi}{\pi}, \tag{45}$$

where we have in both cases used samples from the time interval $(\tau_{2l\mathcal{M}+4}, \tau_{2(l+1)\mathcal{M}-4}] \subset (\lambda_{n,l}, \lambda_{n,l+1}]$.

[9] This assumption is for the current section only. There is some choice as to what to regard as the true theoretical asymptotic variance; another possibility would be the continuous quadratic variation of $M_{n,t}$. So long as two versions of the theoretical AVAR are within $o_p(n^{-2\alpha})$ of each other, this ambiguity does not pose difficulties for consistency of $\widehat{\text{AVAR}}$, or for consistency of confidence intervals based on $\widehat{\text{AVAR}}$.

[10] Otherwise, a term-by-term correction applies, see Mykland and Zhang (2016, Theorem 5, p. 249). For comparison, see also the expression for the pre-averaged RV in Mykland et al. (2019, eq. (7)–(9), p. 104). We emphasize that the correction is available so long as one has faith in the time stamps, which is plausible for the CME data that we have used, due to the centralized trading system for this contract.

$$\text{Eq. (45)} - \frac{1}{4} \times \text{Eq. (44)} = 2(\mathcal{M}-4)\Delta\tau \frac{2}{3}\sigma_{\lambda_l}^2 \frac{3}{4}\frac{6-4\sqrt{3}+\pi}{\pi}$$
$$= (\tau_{2(l+1)\mathcal{M}-4} - \tau_{2l\mathcal{M}+4})\sigma_{\lambda_l}^2 \frac{6-4\sqrt{3}+\pi}{2\pi}. \quad (46)$$

In view of the development in Mykland and Zhang (2016, 2020), the aggregated (over $\mathcal{M}$) terms

$$\sum_{i=l\mathcal{M}+3}^{(l+1)\mathcal{M}-2} \text{med}(\Delta\bar{Y}^c_{2M_n,i-2}, \Delta\bar{Y}^c_{2M_n,i}, \Delta\bar{Y}^c_{2M_n,i+2})^2 - \frac{1}{4}\sum_{i=2l\mathcal{M}+5}^{2(l+1)\mathcal{M}-4} \text{med}(\Delta\bar{Y}^c_{M_n,i-2}, \Delta\bar{Y}^c_{M_n,i}, \Delta\bar{Y}^c_{M_n,i+2})^2$$
$$- (\tau_{2(l+1)\mathcal{M}-4} - \tau_{2l\mathcal{M}+4})\sigma_{\lambda_l}^2 \frac{6-4\sqrt{3}+\pi}{2\pi} \quad (47)$$

satisfy stable convergence and also the other assumptions of Conditions 1–2 (and in particular also Proposition 1 in Mykland and Zhang, 2017a) under Q_n, with $\alpha = 1/4$. One can take the T_i to be the same as the λ_i. This is easily seen to carry over to the original measure. The left-out terms (around the boundaries λ_l) are handled with the big-block–small-block device described in Mykland and Zhang (2012, Chapter 2.6.2, pp. 170–172). Also, the jumps are negligible since assumed to be of finite activity. The interface between jumps and the P-UT condition is handled as in Mykland and Zhang (2017a, Example 2, Section 7).

The edge effects are essentially on the block estimation form described in Mykland and Zhang (2017a, Remark 14, pp. 223–224), and are (singly and by averages) of order $O_p(n^{-2\alpha})$. It follows that the assumptions of Theorem 1 are satisfied. In conclusion:

Proposition 3 (*Median Realized Volatility under Microstructure Noise*)**.** *Let Θ be the integrated volatility on $[0,\mathcal{T}]$. A pre-averaged extension of the median realized volatility of Andersen et al. (2012) is given by*[11]

$$\hat{\Theta} = \frac{2\pi}{6-4\sqrt{3}+\pi}\left(MedRV_{2M_n,n} - \frac{1}{4}MedRV_{M_n,n}\right), \quad (48)$$

Then, with the T_i taken to be the same as the τ_i, Condition 1 is satisfied, as well as the assumptions of Theorems 1–2. In particular, both the two-scales and multi-scale (regression) AVAR and $\widehat{[\theta,\theta]}_{T-}$ are consistent.

8. Simulation: The Heston model

In the following, we present a simulated month where the underlying process is a Heston model, contaminated by microstructure noise. The latent log price is thus $dX_t = (\mu - \sigma_t^2)dt + \sigma_t dB_t$, with $d\sigma_t^2 = \kappa(\alpha - \sigma_t^2)dt + \gamma\sigma_t dW_t$, where B_t and W_t are Brownian motions with constant correlation $\rho = -.5$, and other parameters have values $(\mu, \kappa, \alpha, \gamma) = (0.05, 5, 0.04, 0.5)$. We simulate 20 trading days, with observations every second for 23,400 (trading) seconds per day. The microstructure noise is Gaussian with mean zero and standard deviation 5×10^{-4} (see Fig. 3).

Appendix A. Proofs and technical issues

Because of the close connection to the earlier paper, in the following we refer to Mykland and Zhang (2017a) as MZ and Mykland and Zhang (2017b) as MZ-A.

A.1. Proof of Theorem 1

The strategy is to take the proof of Theorem 8 in MZ-A as a point of departure, but to intercept it at the point of equation (C.10) in MZ-A. which we write more generally as

$$QV_{B,K}(\hat{\Theta}) = \overline{QV}_{B,K}(\hat{\Theta}) + R_{n,K} + 2QV(\Theta, \tilde{e} \text{ and } e) + 2QV(M, \tilde{e} \text{ and } e). \quad (A.1)$$

Since the behavior of $\overline{QV}_{B,K}(\hat{\Theta})$ is given in (C.4) in MZ-A, we need to deal with the three last terms in (A.1). The expressions, and the additional conditions, are given in Lemma 1 and Corollary 1 below, thus yielding Theorem 1. □

Lemma 1 (*Representation of R_{n,K_n}*)**.** *Assume Conditions 1–2, as well as the balance condition (12). Let MAEE_n and $\varepsilon_{n,K}$ be given by (11) in Theorem 1. Then*

$$R_{n,K_n} = 2\mathcal{T}(K_n\Delta T_n)^{-1}\left(\text{MAEE}_n + \varepsilon_{n,K_n}\right) + o_p(n^{-2\alpha}). \quad (A.2)$$

[11] The estimator can be small sample adjusted as in the original paper, without affecting the conclusion of this proposition. One can also use the average of rolling windows.

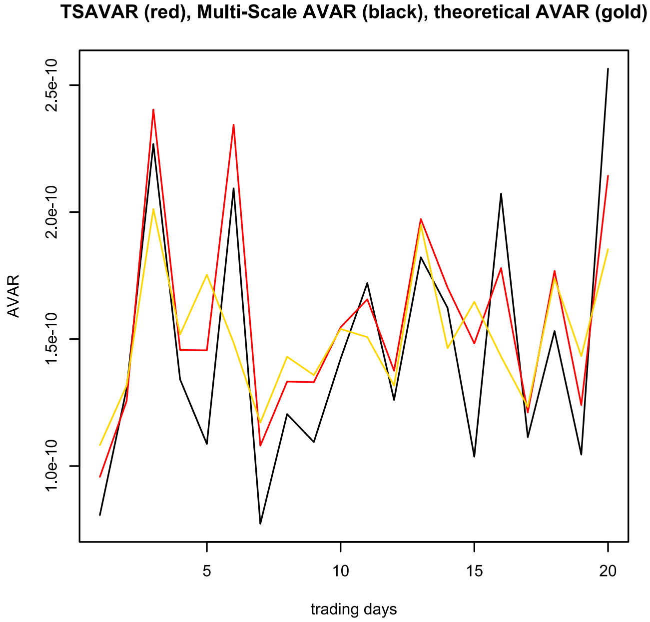


Fig. 3. Simulation of Heston model. This plot shows the observed AVAR using the two-scales method (red) and the multi-scale method (black). The AVARs are based on the same data (from the simulated Heston model) and estimator (the non-tapered S-TSRV). We use $\Delta T = 15$ s, and for the multi-scale AVAR, K takes the values 8, 15, 23, 31, and 39. The golden curve is the true estimation variance.

Proof of Lemma 1. Without loss of generality we can go back and forth between e by e'. Consider the main term consisting of terms of the form $\tilde{e}^2_{T_i} + e^2_{T_i} + \tilde{e}_{T_i} e_{T_i}$. The difference between this term in R_{n,K_n} as defined in (C.1) in MZ-A, and the representation $2\mathcal{T}(K_n \Delta T_n)^{-1}\text{MAEE}_n$ is thus on the overall edges (near 0 and $\mathcal{T}$). To see that the difference is negligible, note that

$$\frac{1}{K_n}\sum_{i=0}^{2K_n}(e^2_{T_i} + \tilde{e}^2_{T_i}) = o_p(n^{-2\alpha}) \text{ and } \frac{1}{K_n}\sum_{i=B_n-2K_n+1}^{B_n}(e^2_{T_i} + \tilde{e}^2_{T_i}) = o_p(n^{-2\alpha}) \tag{A.3}$$

The reason for (A.3) is on the one hand that by Condition 1, for each i, $n^{2\alpha}(e^2_{T_i} + \tilde{e}^2_{T_i}) \xrightarrow{p} 0$. On the other hand, by invoking (C.1) in Remark 3, we may, without loss of generality, take each term to be bounded by $2\Gamma^2$ (for some constant Γ), whence (A.3) follows by dominated convergence.

The lagged terms behave similarly. □

We now turn to the Cross Terms $QV(\Theta, \tilde{e}$ and $e)$ and $QV(M, \tilde{e}$ and $e)$. In analogy with the development in Appendices A–B in MZ-A, it is easy to see that

$$\begin{aligned} QV(\Theta, \tilde{e} \text{ and } e) &= \frac{1}{K}\sum_{i=K}^{B-K}(\Theta'_{(T_{n,i},T_{n,i+K}]} + \Theta''_{(T_{n,i-K},T_{n,i}]})\left((\tilde{e}'_{T_{n,i+K}} - e'_{T_{n,i}}) - (\tilde{e}'_{T_{n,i}} - e'_{T_{n,i-K}})\right) \\ &= \frac{1}{K}K\Delta T\left(\sum_{i=0}^{B_n}\tilde{e}_{T_{n,i}}\int_{T_{n,i-2K}}^{T_{n,i+K}}\tilde{\mathfrak{f}}^{(l_i,n)}(t)d\theta_t + \sum_{i=0}^{B_n}e_{T_{n,i}}\int_{T_{n,i-K}}^{T_{n,i+2K}}\mathfrak{f}^{(l_i,n)}(t)d\theta_t\right), \end{aligned} \tag{A.4}$$

where $|\mathfrak{f}^{(l_i,n)}(t)| \le 1$ and $|\tilde{\mathfrak{f}}^{(l_i,n)}(t)| \le 1$, where $l_i \equiv i[3K]$ in the sense of Definition 6 in MZ-A, but with $3K$ replacing $2K$. Also, we take θ_t to be constant on the intervals $(-\infty, 0]$ and $[\mathcal{T}, \infty)$.

For example, away from the edge, $t \in (T_{2K}, T_{B-2K}]$, we have that when $i \equiv l[3K]$,

$$\tilde{\mathfrak{f}}^{(l,n)}(t) = \begin{cases} \frac{1}{K\Delta T}(t - T_{n,i-2K}) \text{ when } t \in (T_{n,i-2K}, T_{n,i-K}], \\ 1 \text{ when } t \in (T_{n,i-K}, T_{n,i}], \text{ and} \\ \frac{1}{K\Delta T}(T_{n,i+K} - t) \text{ when } t \in (T_{n,i}, T_{n,i+2K}]. \end{cases} \tag{A.5}$$

This is in analogy with the definition of $f^{(l,n)}$ in (B.1) in MZ-A.

A similar but more elementary derivation yields that

$$QV(M, \tilde{e} \text{ and } e) = \frac{1}{K}\sum_{i=K}^{B-K}\left((M_{T_{n,i+K}} - M_{T_{n,i}}) - (M_{T_{n,i}} - M_{T_{n,i-K}})\right)\left((\tilde{e}'_{T_{i+K}} - e'_{T_i}) - (\tilde{e}'_{T_i} - e'_{T_{i-K}})\right)$$
$$= \frac{2}{K}n^{-\alpha}\left(\sum_{i=0}^{B_n}\tilde{e}_{T_i}\int_{T_{i-2K}}^{T_{i+K}}\tilde{\mathfrak{g}}^{(l_i,n)}(t)dL_{n,t} + \sum_{i=0}^{B_n}e_{T_i}\int_{T_{i-K}}^{T_{i+2K}}\mathfrak{g}^{(l_i,n)}(t)dL_{n,t}\right), \tag{A.6}$$

where $|\mathfrak{g}^{(l_i,n)}(t) \le 1$ and $|\tilde{\mathfrak{g}}^{(l_i,n)}(t)| \le 1$, where $l_i \equiv i[3K]$. Also, we take $L_{n,t} = n^\alpha M_{n,t}$, and let $L_{n,t}$ be constant on the intervals $(-\infty, 0]$ and $[\mathcal{T}, \infty)$.

Again, for example, away from the edge, $t \in (T_{2K}, T_{B-2K}]$, we have that when $i \equiv l[3K]$

$$\tilde{\mathfrak{g}}^{(l,n)}(t) = \begin{cases} -\frac{1}{2} \text{ when } t \in (T_{n,i-2K}, T_{n,i-K}] \cup (T_{n,i}, T_{n,i+2K}], \text{ and} \\ 1 \text{ when } t \in (T_{n,i-K}, T_{n,i}]. \end{cases} \tag{A.7}$$

The above situations both satisfy the conditions of the following lemma:

Lemma 2 (*Sharper Bounds on the Cross-Term*)**.** *Assume that $\beta_t^{(n)}$ is an $O_p(1)$ sequence (in n) of semimartingales.*[12] *Let $\mathfrak{h}^{(l_i,n)}$ be nonrandom, càglàd,*[13] *and satisfy $|\mathfrak{h}^{(l_i,n)}(t)| \le 1$. Also, let $\mathbb{H}$ be the set of functions $t \to \mathfrak{h}^{(l,n)}(t+)$, and construct $\mathbb{H}'$ from $\mathbb{H}$ as in (A.4) in MZ-A except that $T_{(K_n+1)j+l}$ is replaced by $T_{(K_n+1)j+l-J_n}$. Assume that $\mathbb{H}'$ is relatively compact for the Skorokhod topology.*[14] *Assume Condition 2, and let $J_n \le K_n$, with $J_n \Delta T_n = o_p(n^{-\alpha})$. Also assume the balance condition (12). Then*

$$n^\alpha \frac{1}{K_n}\sum_{i=0}^{B_n} e'_{T_i}\int_{T_{i-K}}^{T_{i+2K}}\mathfrak{h}^{(l_i,n)}(t)d\beta_t^{(n)} = o_p(1). \tag{A.8}$$

The corresponding $\tilde{e}'_{T_{n,i}}$ sum has the same order.[15]

Hence

Corollary 1 (*Sharper Bounds for the Cross Terms*)**.** *Under Condition 2, the balance condition (12), and if $J_n\Delta T_n = o_p(n^{-\alpha})$, then $QV(\Theta, \tilde{e}$ and $e)$ and $QV(M, \tilde{e}$ and $e)$ are both of order $o_p(n^{-2\alpha})$.*

Proof of Lemma 2. In conformity with Definition 8 in Appendix A in MZ-A, we use that $\beta_t^{(n)}$ has decomposition $\beta_t^{(n)} = \beta_{(n)}^0 + \beta^{(n)}(h)_t + \beta_t^{(n,R)}$, where $\beta_t^{(n,R)} = B_n(h)_t + \breve{\beta}^{(n)}(h)_t$. $D(\beta^{(n)})(h)_t$ is given in analogy with (A.8) in the earlier paper. By invoking (C.1) in Remark 3, we see that we can take, without loss of generality,

$$|n^\alpha e'_{T_i}| \le \Gamma, \tag{A.9}$$

for some constant Γ. We shall assume this throughout the proof of this lemma.

We split the term (A.8) in four parts. First,

$$\begin{aligned} n^\alpha|\sum_{i=0}^{B_n} e'_{T_i}\int_{T_{i-J}}^{T_{i+J}}\mathfrak{h}^{(l_i,n)}(t)d\beta_t^{(n,R)}| \\ &\le \Gamma|e'_{T_i}|\sum_{i=0}^{B_n}|\int_{T_{i-J_n}}^{T_{i+J_n}}\mathfrak{h}^{(l_i,n)}(t)d\beta_t^{(n,R)}| \\ &\le \Gamma|\sum_{i=0}^{B_n}\int_{T_{i-J_n}}^{T_{i+J_n}}|\mathfrak{h}^{(l_i,n)}(t)|dD(\beta^{(n)})_t \\ &\le \Gamma 3J_n D(\beta^{(n)})_{\mathcal{T}} \\ &= O_p(J_n) \end{aligned} \tag{A.10}$$

from Condition 2 and since $D(\beta^{(n)})_{\mathcal{T}} = O_p(1)$ by Jacod and Shiryaev (2003, Theorem VI.6.15(i) and (iii), p. 380).

12 "$O_p(1)$" here means that $\beta_t^{(n)}$ is tight with respect to the Skorokhod topology, as well as P-UT (Definition 5 in MZ-A). For the case (A.4), $\beta_t^{(n)} = \theta_t$, so this is immediate, while for the case (A.6), $\beta_t^{(n)} = L_{n,t}$, it follows from Condition 1.

13 Left continuous with right limits. In other words, $t \to \mathfrak{h}^{(l_i,n)}(t+)$ is in $\mathbb{D}$.

14 This is satisfied by the families $\mathfrak{f}^{(l,n)}$, $\tilde{\mathfrak{f}}^{(l,n)}$, $\mathfrak{g}^{(l,n)}$, and $\tilde{\mathfrak{g}}^{(l,n)}$ above.

15 If one does not assume $J_n\Delta T_n = o_p(n^{-\alpha})$ and the balance condition, the right hand side of (A.8) is given by (A.22) at the end of the proof of the lemma.

Second, by Hölder's inequality,

$$
\begin{aligned}
n^{\alpha} | \sum_{i=0}^{B_n} e'_{T_i} \int_{T_{i-J_n}}^{T_{i+J_n}} \mathfrak{h}^{(l_i,n)}(t) d\beta^{(n)}(h)_t | \\
&\leq \Gamma \sum_{i=J}^{B_n-J_n} | \int_{T_{i-J_n}}^{T_{i+J_n}} \mathfrak{h}^{(l_i,n)}(t) d\beta^{(n)}(h)_t | \\
&\leq \Gamma \, (B_n)^{1/2} \left(\sum_{i=J}^{B_n-J_n} \left(\int_{T_{i-J_n}}^{T_{i+J_n}} \mathfrak{h}^{(l_i,n)}(t) d\beta^{(n)}(h)_t \right)^2 \right)^{1/2} \\
&= o_p(\Delta T_n^{-1/2} J_n^{1/2}).
\end{aligned}
\tag{A.11}
$$

This is because $\sum_{i=J}^{B_n-J_n} \left(\int_{T_{i-J_n}}^{T_{i+J_n}} \mathfrak{h}^{(l_i,n)}(t) d\beta^{(n)}(h)_t \right)^2$ is Lenglart dominated (Jacod and Shiryaev, 2003, Lemma I.3.20, p. 35) by

$$
\sum_{i=J}^{B_n-J_n} \int_{T_{i-J_n}}^{T_{i+J_n}} (\mathfrak{h}^{(l_i,n)}(t))^2 d\tilde{C}_t^{(n)} \leq \sum_{i=J}^{B_n-J_n} (\tilde{C}_{T_{i+J_n}}^{(n)} - \tilde{C}_{T_{i-J_n}}^{(n)}) \leq 2 J_n \tilde{C}_{\mathcal{T}}^{(n)}
\tag{A.12}
$$

where $\tilde{C}_t^{(n)}$ is the second modified characteristic of $\beta_t^{(n)}$, cf. Definition 8 in Appendix A in MZ-A (or refer directly to Jacod and Shiryaev, 2003). $\tilde{C}_{\mathcal{T}}^{(n)} = O_p(1)$ by Jacod and Shiryaev (2003, Theorem VI.6.15(ii), p. 380).

Third, consider

$$
S_{n,I} = n^{\alpha} \frac{1}{K_n} \sum_{i=0}^{I} e'_{T_i} \int_{T_{i-K}}^{T_{i-J}} \mathfrak{h}^{(l_i,n)}(t) d\beta_t^{(n)},
\tag{A.13}
$$

and set

$$
\tilde{\mathfrak{h}}^{(l,n,-)}(t) = \begin{cases} 0 \text{ when } t \in \cup_{i \equiv l[3K_n]} (T_{n,i-J}, T_{n,i+2K}], \text{ and} \\ \mathfrak{h}^{(l,n)}(t) \text{ for all other } t \in (0, \mathcal{T}]. \end{cases}
\tag{A.14}
$$

$S_{n,I}$ is a multi-lag martingale in the sense of Lemma 4 (with lag length $N = 2J$) in Appendix C. We calculate in the notation of Lemma 4 (with $N = 2J$),

$$
\begin{aligned}
\langle S_n, S_n \rangle_{B_n}^{(2J)} &\leq \Gamma^2 \frac{1}{K_n^2} \sum_{i=0}^{I} \left(\int_{T_{i-K}}^{T_{i-J}} \mathfrak{h}^{(l_i,n)}(t) d\beta_t^{(n)} \right)^2 \\
&= \Gamma^2 \frac{1}{K_n^2} \sum_{i=0}^{I} \left(\int_{T_{i-K}}^{T_{i+2K}} \mathfrak{h}^{(l_i,n,-)}(t) d\mathfrak{h}^{(l_i,n,-)}(t)_t \right)^2 \\
&= \Gamma^2 \left(\frac{1}{K_n^2} \sum_{l=1}^{3K_n} \int_0^{\mathcal{T}} \mathfrak{h}^{(l_i,n,-)}(t)^2 d[\beta^{(n)}, \beta^{(n)}]_{\mathcal{T}} \right) (1 + o_p(1)) \\
&\leq 3 \frac{1}{K_n} \Gamma^2 [\beta^{(n)}, \beta^{(n)}]_{\mathcal{T}} (1 + o_p(1)),
\end{aligned}
\tag{A.15}
$$

in analogy with Theorem 7 in MZ-A (use $3K_n$ rather than $2K_n$). We have here used the relative compactness assumption on $\mathbb{H}'$. Thus, by Lemma 4,

$$
S_{n,B_n} = O_p((J_n/K_n)^{1/2}).
\tag{A.16}
$$

Fourth, set

$$
\tilde{\mathfrak{h}}^{(l,n,-)}(t) = \begin{cases} 0 \text{ when } t \in \cup_{i \equiv l[3K_n]} (T_{n,i-K}, T_{n,i+J}], \text{ and} \\ n^{\alpha} e'_{T_i} \mathfrak{h}^{(l,n)}(t) \text{ for all other } t \in (0, \mathcal{T}]. \end{cases}
\tag{A.17}
$$

Consider

$$
\begin{aligned}
n^{\alpha} \frac{1}{K} \sum_{i=0}^{B_n} e'_{T_i} \int_{T_{i+J}}^{T_{i+2K}} \mathfrak{h}^{(l_i,n)}(t) d\beta_t^{(n)} \\
= \frac{1}{K} \sum_{i=0}^{B_n} \int_{T_{i-K}}^{T_{i+2K}} \mathfrak{h}^{(l_i,n,-)}(t) d\beta_t^{(n)}
\end{aligned}
$$

$$= \frac{1}{K}\sum_{l=1}^{3K}\int_0^{\mathcal{T}} \mathfrak{h}^{(l,n,-)}(t)d\beta_t^{(n)}$$
$$= \int_0^{\mathcal{T}} \mathfrak{h}^{(n,-)}(t)d\beta_t^{(n)}, \tag{A.18}$$

where

$$\mathfrak{h}^{(n,-)}(t) = \frac{1}{K_n}\sum_{l=1}^{3K}\mathfrak{h}^{(l,n,-)}(t). \tag{A.19}$$

$|\mathfrak{h}^{(l,n,-)}(t)| \leq \Gamma$, and hence $|\mathfrak{h}^{(n,-)}(t)| \leq 3\Gamma$. Also, $\mathfrak{h}^{(n,-)}(t)$ is predicable. Now write

$$\mathfrak{h}^{(n,-)}(t) = \frac{1}{K_n}\sum_{i=0}^{B_n} n^\alpha e'_{T_i}\mathfrak{h}^{(l_i,n)}(t)\mathbf{I}_{\{t\in(T_{n,i+J},T_{n,i+2K}]\}}$$
$$= \sum_{i:t\in(T_{n,i+J},T_{n,i+2K}]} n^\alpha e'_{T_i}\mathfrak{h}^{(l_i,n)}(t)$$
$$= \frac{1}{K}\sum_{i=j-2K_n}^{j-1-J_n} n^\alpha e'_{T_i}\mathfrak{h}^{(l_i,n)}(t) \text{ when } t\in(T_{j-1},T_j]. \tag{A.20}$$

For fixed t, $\mathfrak{h}^{(n,-)}(t)$ is, therefore the endpoint of a multi-lag martingale in the sense of Lemma 4 (with lag length $N = 2J$) in Appendix C. As in the proof of Lemma 4 (with $N = 2J$), we see that $E(\mathfrak{h}^{(l,n,-)}(t)^2) \leq (4J_n - 1)K_n^{-1}\Gamma^2$. Thus, following Lenglart's inequality (Jacod and Shiryaev, 2003, Lemma I.3.20, p. 35), $\sup_{0\leq t\leq\mathcal{T}}|\mathfrak{h}^{(n,-)}(t)| = O_p((J_n/K_n)^{1/2})$. Hence, by *Ibid.*, Corollary VI.6.20(b) (p. 381), it follows that

$$\text{(A.18)} = O_p((J_n/K_n)^{1/2}). \tag{A.21}$$

Combining (A.10), (A.11), (A.16), and (A.21) yields that (A.8) has order

$$O_p\left(\frac{J_n}{K_n} + \left(\frac{J_n}{K_n}\right)^{1/2}(1+(K_n\Delta T_n)^{-1/2})\right). \tag{A.22}$$

By imposing the balance condition (12) along with $J_n\Delta T_n = o_p(n^{-\alpha})$, the right hand side of (A.8) follows. □

Appendix B. Properties and convergence of the edge effect, and consistency of the multi-scale method

B.1. About Condition 2 on the edge effects

The formulation means that the main edge effect at T_i is allowed to depend on observations in J time periods on each side of T_i.

The specific conditions can be verified under mixing assumptions. The following is a complement to our examples. This is not intended to provide minimal conditions, just to explain why our conditions are reasonable.

The Decomposition $e_{T_i} = e'_{T_i} + e''_{T_i}$ and $\tilde{e}_{T_i} = \tilde{e}'_{T_i} + \tilde{e}''_{T_i}$. We have chosen this way of stating the conditions on the edge effect since, in our examples, this is readily verifiable. To tie the condition to the literature, however, we observe that, subject to mixing conditions, we require $(e_{T_i}, \tilde{e}_{T_i})$ to be a *mixingale*, see, *e.g.*, McLeish (1975) and Hall and Heyde (1980, pp. 19–21, 41). As the name suggests, it is tied up with the concept of mixing. See also Wu and Woodroofe (2004).

α-and ϕ- mixing. For a more general treatment, see McLeish (1975, p. 834) and Hall and Heyde (1980, Chapter 5 and Appendix III). For simplicity, we here focus on ϕ-mixing.[16] If $\mathcal{A}$ and $\mathcal{B}$ are two sigma-fields, then the ϕ-fixing coefficient is

$$\phi(\mathcal{A},\mathcal{B}) = \sup_{A\in\mathcal{A},B\in\mathcal{B},P(A)>0}|P(B|A) - P(B)| \tag{B.1}$$

The Decomposition, again. Set $\tilde{e}''_{T_i} = \tilde{e}_{T_i} - E(\tilde{e}_{T_i} \mid \mathcal{F}_{T_{i-J}})$, and similarly for e''_{T_i}. The difference $\tilde{e}'_{T_i} = \tilde{e}_{T_i} - \tilde{e}''_{T_i}$ will then have the martingale-like properties described, as will e'_{T_i}.

16 One can do similar things with α-mixing, using the definition and lemma of McLeish (1975, p. 834). Condition (B.2) becomes $\sum_i \alpha(\mathcal{F}_{T_{i-J_n}}, \mathcal{A}_{n,i})^{\frac{\delta-1}{2(1+\delta)}} = o(1)$, with McLeish's definition of α. Thus, in this case, we need $\delta > 1$.

Meanwhile, if we require, say, that $\sup_n \left(\max_{0\le i\le B_n} E|n^{\alpha} e_{n,T_i}|^{1+\delta} + \max E|n^{\alpha}\tilde{e}_{n,T_i}|^{1+\delta}\right) < \infty$, for some $\delta > 0$, and also that $\sum_i (Ee_{n,T_i})^2 + (E\tilde{e}_{n,T_i})^2 = o(n^{-2\alpha})$, then the lemma on McLeish (1975, p. 834) assures that our conditions on $(e''_{T_i}, \tilde{e}''_{T_i})$ are satisfied provided

$$\sum_i \phi(\mathcal{F}_{T_{i-J_n}}, \mathcal{A}_{n,i})^{\frac{2\delta}{1+\delta}} = o(1), \tag{B.2}$$

where $\mathcal{A}_{n,i}$ is the sigma-field generated by $(e_{T_{n,i}}, \tilde{e}_{T_{n,i}})$ (use $p = 2$ and $r = 1+\delta$). Normally, however, the number of observations in each interval $(T_{i-J_n}, T_i]$ will go to infinity with n, thus under exponential mixing (in the original microstructure noise), (B.2) will normally hold.

B.2. Proof of Proposition 1

Proof of Proposition 1. We show (18) and the asymptotic uncorrelatedness below. From (18) follows the first line of (16), by definition of $\varepsilon_{n,K}$. The worst case statements in (15)–(17) follow as in the proof of Lemma 1, using Condition 2.

One such term (and the others are all handled the same way) is $C^{12}_{n,K} = \frac{1}{B_n}\sum_{i=K}^{B_n} \tilde{e}_{T_i} e_{T_{i-K}}$. By Condition 2, this term has the same asymptotic behavior (up to $o_p(n^{-2\alpha})$) as $\frac{1}{B_n}\sum_{i=K}^{B_n} \tilde{e}'_{T_i} e'_{T_{i-K}}$. We then invoke statement (C.1) in Remark 3. Now identify the sum $\sum_{i=K}^{I} \tilde{e}'_{T_i} e'_{T_{i-K}}$ with $S_{n,I}$ in Lemma 4 (with $\mathcal{H}_{n,i} = \mathcal{F}_{T_{i+J}}$, and $N = 2J$.). The multi lag angle bracket process is $\langle S_n, S_n\rangle^{(N)}_I = \sum_{i=K}^{I} \left(E((\tilde{e}'_{T_i})^2 | \mathcal{F}_{T_{i-2J_n}})(e'_{T_i})^2\right)$, which is in turn Lenglart-dominated by

$$\text{VAEE}'_{n,K} = \sum_{i=K}^{I} \left(E((\tilde{e}'_{T_i})^2 | \mathcal{F}_{T_{i-2J_n}}) E((e'_{T_{i-K}})^2 | \mathcal{F}_{T_{i-K-2J_n}})\right), \tag{B.3}$$

which in turn is Lenglart-dominated by VAEE_n (independent of K). Hence, as in Lemma 4, $S_{B_n} = O_p((J_n B_n \text{VAEE}_n)^{1/2})$, and so $C^{12}_{n,K_n} = O_p((J_n \Delta T_n \text{VAEE}_n)^{1/2})$. The rest of (18) follows by the exact same reasoning. The uncorrelatedness arises since, by the same methods, $C^{\cdot\cdot}_{n,K_{n,l}}$ and $C^{\cdot\cdot}_{n,K_{n,l+1}}$ are small sample uncorrelated. This carries over asymptotically by uniform integrability. □

B.3. Proof of Theorem 2 (Section 4) and Proposition 2 (Section 6)

Proof of Theorem 2 in Section 4. We first proceed in the hard edge case. Let $\bar{K}_n$ be the mean of the $K_{n,l}$, and set $\mathfrak{D}_n = \text{diag}(1, \bar{K}_n\Delta T_n, (\bar{K}_n\Delta T_n)^3)$. Rescale so that $\mathfrak{Y}_n = (\bar{K}_n\Delta T_n)^{-3}\mathbb{Y}_n$, $\underline{\mathfrak{b}}_n = (\bar{K}_n\Delta T_n)^{-3}\mathfrak{D}_n\underline{\beta}_n$, and $\mathfrak{X}_n = \mathbb{X}_n\mathfrak{D}_n^{-1}$. To spell out the latter two,

$$\underline{\mathfrak{b}}_n^* = \left((\bar{K}_n\Delta T_n)^{-3}\text{MAEE}_n, (\bar{K}_n\Delta T_n)^{-2}\text{AVAR}_n, [\theta,\theta]_{\mathcal{T}-} \right), \text{ and} \tag{B.4}$$

$$\mathfrak{X}_n^* = \begin{pmatrix} 2\mathcal{T} & 2\mathcal{T} & \cdots & 2\mathcal{T} \\ 2(K_{n,1}/\bar{K}_n) & 2(K_{n,2}/\bar{K}_n) & \cdots & 2(K_{n,m}/\bar{K}_n) \\ \frac{2}{3}(K_{n,1}/\bar{K}_n)^3 & \frac{2}{3}(K_{n,2}/\bar{K}_n)^3 & \cdots & \frac{2}{3}(K_{n,m}/\bar{K}_n)^3 \end{pmatrix}. \tag{B.5}$$

Also, let $\hat{\underline{\mathfrak{b}}}_n$ be the least squares estimator from the regression of $\mathfrak{Y}_n$ on $\mathfrak{X}_n$, *i.e.*, $\hat{\underline{\mathfrak{b}}}_n = (\mathfrak{X}_n^*\mathfrak{X}_n)^{-1}\mathfrak{X}_n^*\mathfrak{Y}_n$.

With this setup, $\mathfrak{X}\underline{\mathfrak{b}}_n = (\bar{K}_n\Delta T_n)^{-3}\mathbb{X}_n\underline{\beta}_n$ and $\mathfrak{X}_n^*\mathfrak{X}_n = \mathfrak{D}_n^{-1}\mathbb{X}_n^*\mathbb{X}_n\mathfrak{D}_n^{-1}$, whence $\hat{\underline{\mathfrak{b}}}_n = (\bar{K}_n\Delta T_n)^{-3}\mathfrak{D}_n\hat{\underline{\beta}}_n$, and so

$$\hat{\underline{\beta}}_n - \underline{\beta}_n = (\bar{K}_n\Delta T_n)^3\mathfrak{D}_n^{-1}(\hat{\underline{\mathfrak{b}}}_n - \underline{\mathfrak{b}}_n). \tag{B.6}$$

Eq. (26) becomes, in view of (19),

$$\mathfrak{Y}_n = \mathfrak{X}_n\mathfrak{b}_n + o_p(1). \tag{B.7}$$

Now let $\underline{\mathcal{B}}_n$, $\hat{\underline{\mathcal{B}}}_n$ be the last two elements in, respectively $\underline{\mathfrak{b}}_n$ and $\hat{\underline{\mathfrak{b}}}_n$. Also let $\mathcal{X}_n^*$ be the submatrix consisting of the two last rows of $\mathfrak{X}_n^*$, and let $\mathcal{D}_n$ be the 2×2 submatrix in the lower right corner of $\mathfrak{D}_n$. Let $\mathfrak{H} = \mathfrak{I} - m^{-1}\mathfrak{J}$, where $\mathfrak{I}$ is the $m\times m$ identity matrix, and $\mathfrak{J}$ is the $m\times m$ matrix all of whose entries are 1.

Following Weisberg (1985, Chapter 2.2, p. 43–44), $\underline{\mathcal{B}}_n = ((\mathfrak{H}\mathcal{X}_n)^*\mathfrak{H}\mathcal{X}_n)^{-1}(\mathfrak{H}\mathcal{X}_n)^*\mathfrak{H}\mathfrak{Y}_n$. Meanwhile, from (B.7), $\mathfrak{H}\mathfrak{Y}_n = \mathfrak{H}\mathfrak{X}_n\mathfrak{b}_n + o_p(1) = \mathfrak{H}\mathcal{X}_n\mathcal{B}_n + o_p(1)$. Thus, $\hat{\underline{\mathcal{B}}}_n - \underline{\mathcal{B}}_n = ((\mathfrak{H}\mathcal{X})_n^*\mathfrak{H}\mathcal{X}_n)^{-1}((\mathfrak{H}\mathcal{X})_n^*\mathfrak{H}\mathcal{X}_n)\mathcal{B}_n + o_p(1) = \mathcal{B}_n + o_p(1)$, since $(\mathcal{H}\mathcal{X})_n^*\mathfrak{H}\mathfrak{X}_n$ is nonsingular (uniformly in n) by condition (34). Since $\hat{\underline{\mathcal{B}}}_n - \underline{\mathcal{B}}_n = o_p(1)$ and in view of (B.6), the consistency (35) follows. In the soft edge case, the conditions imposed guarantee Theorem 3 (in Section 3.2 of AZ), and hence (B.7) is valid with $\text{MAEE}_n \equiv 0$. As above, Theorem 2 follows. □

Proof of Proposition 2 in Section 6. Linear regression theory (*e.g.*, Weisberg, 1985, p. 203) yields that r_n is the slope in the regression of the third on the two first columns of $\mathbb{X}$. If we set $\mathfrak{r}_n$ to be the slope in the comparable regression of the third on two first columns of $\mathfrak{X}$, we obtain

$$r_n = \mathfrak{r}_n(\bar{K}\Delta T_n)^2 \text{ and } \mathfrak{r}_n = \frac{1}{3\bar{K}_n^2}\frac{\sum_{l=1}^m (K_{n,l}-\bar{K}_n)K_{n,l}^3}{\sum_{l=1}^m (K_{n,l}-\bar{K}_n)^2} \tag{B.8}$$

which is of exact order $O(1)$ by assumption (34) in Theorem 2. Thus, in the notation of the preceding proof, $MSQV(\hat{\Theta}_{n,c}) = \underline{\hat{\beta}}_n^{(1)} + r_n\underline{\hat{\beta}}_n^{(2)}$, where we use $\underline{\hat{\beta}} = (\underline{\hat{\beta}}_n^{(0)}, \underline{\hat{\beta}}_n^{(1)}, \underline{\hat{\beta}}_n^{(2)})^*$. Hence $MSQV(\hat{\Theta}_{n,c}) = (\bar{K}\Delta T_n)^2\left(\underline{\hat{\mathcal{B}}}_n^{(1)} + \mathfrak{r}_n\underline{\hat{\mathcal{B}}}_n^{(2)}\right)$. Hence, eventually, $\hat{c}_n = c^*$, and also (42) holds. The stable convergence holds as in AZ. □

Appendix C. Technical lemmas

To handle general moments, we shall use the following.

Lemma 3 (*Truncating the Edge Effects*)**.** *Suppose Condition 2. Then, for any $\delta > 0$, there exists (possibly on an extension of the space) e^{tr}_{n,T_i} and $\tilde{e}^{tr}_{n,T_i}$, and a nonrandom constant Γ, so that*

1. *For each n $e^{tr}_{n,T_i} = e'_{n,T_i}$ and $\tilde{e}^{tr}_{n,T_i} = \tilde{e}'_{n,T_i}$ for all $i \in [0, B_n]$, on a measurable set A_n, and $P(A_n) < \delta$;*
2. *e^{tr}_{n,T_i} and $\tilde{e}^{tr}_{n,T_i}$ satisfy the conditions in Condition 2 in lieu of e'_{n,T_i} and $\tilde{e}'_{n,T_i}$; and*
3. *$|e^{tr}_{n,T_i}| \le \Gamma n^{-\alpha}$ and $|\tilde{e}^{tr}_{n,T_i}| \le \Gamma n^{-\alpha}$ for all i and n.*

Remark 3 (*Using Lemma 3*)**.** We shall use the lemma to assert, in various places, that

$$|n^\alpha e'_{n,T_i}| \text{ and } |n^\alpha \tilde{e}'_{n,T_i}| \text{ can without loss of generality be taken to be bounded by a constant } \Gamma. \tag{C.1}$$

Here is the specific mechanism that we refer to.

Let Y_n be a sequence of random variables, involving a functional form of e'_{n,T_i} and $\tilde{e}'_{n,T_i}$ (as well as any of the other random quantities in our setup). Let D be a countable set, $D \subset (0, 1)$, with a limit point at zero.

For given $\delta \in D$, create $Y_{n,\delta}$ by replacing the e'_{n,T_i} and $\tilde{e}'_{n,T_i}$ by the e^{tr}_{n,T_i} and $\tilde{e}^{tr}_{n,T_i}$ as described by Lemma 3. Then $Y_n = Y_{n,\delta}$ on the set A_n. Suppose one can show that there is a random variable Y (independent of δ) so that $Y_{n,\delta} \xrightarrow{p} Y$ as $n \to \infty$. Then, for any $\epsilon > 0$, and since $P(A_n) < \delta$,

$$\begin{aligned} P(|Y_n - Y| > \epsilon) &\le P(\{|Y_{n,\delta} - Y| > \epsilon\} \cap A_n^c) + P(A_n) \\ &\le P(|Y_{n,\delta} - Y| > \epsilon) + \delta \\ &\to \delta \text{ as } n \to \infty. \end{aligned} \tag{C.2}$$

Since D has limit point at zero, it follows that $Y_n \xrightarrow{p} Y$ as $n \to \infty$. □

Proof of Lemma 3. For $L = 1, \ldots, 2J$, set $S^{(L)}_{n,I} = \sum_{i\in[1,I] \text{ and } i\equiv L[2J]} e'_{n,T_i}$, where $i \equiv L[N]$ means that i is of the form $i = L + jN$ for some integer j. Then for each L and n, $S^{(L)}_{n,I}$ is a martingale with respect to the filtration $\mathcal{H}_{n,i} = \mathcal{F}_{T_{i+J}}$. We now invoke the construction from Mykland (1994, eq. (4.8), p. 27), which produces e^{tr}_{n,T_i} ($i \equiv L[2J]$), satisfying items (1), (2) and (3) in the Lemma, with, say $A_{n,L,1}$ and $\Gamma_{L,1}$, and with $P(A_{n,L,1}) < \delta/4J$. We repeat this construction for all L, and similarly for $\tilde{e}'_{n,T_i}$, in the latter case giving rise to $A_{n,L,2}$ and $\Gamma_{L,2}$. By construction, the whole set of e^{tr}_{n,T_i} and $\tilde{e}^{tr}_{n,T_i}$ satisfy items (1), (2) and (3) in the Lemma, with $A_n = \cup A_{n,L,r}$ and $\Gamma = \max \Gamma_{L,r}$. □

To handle cross-terms, we use the following.

Lemma 4 (*Negligibility of Multi-lag Martingales*)**.** *Let $S_{n,I} = \sum_{i=1}^I \zeta_{n,i}$, where we suppose that $\zeta_{n,i}$ is $\mathcal{H}_i^n$-measurable and satisfies that $E(\zeta_i^n \mid \mathcal{H}_{i-N}) = 0$.*[17] *Define $\langle S_n, S_n\rangle_I^{(N)} = \sum_{i=1}^I E((\zeta_{n,i})^2 \mid \mathcal{H}_{i-N})$. (It is an Nth-lag angle bracket process.) Let α_n be a nonrandom sequence so that $\langle S_n, S_n\rangle_{B'_n}^{(N)} = o_p(\alpha_n)$. Then $\sup_{1\le I\le B'_n} |S_{n,I}| = o_p((N\alpha_n)^{1/2})$.*

Proof of Lemma 4. For $0 \le L \le N-1$, let $S^{(L)}_{n,I} = \sum_{i\in[1,I] \text{ and } i\equiv L[N]} \zeta_{n,i}$, where $i \equiv L[N]$ means that i is of the form $i = L + jN$ for some integer j.

Thus, $S_{n,I} = \sum_{j=1}^N S^{(L)}_{n,I}$. Since no two different $S^{(L)}_{n,I}$ change value for the same I, we also get that $[S_n, S_n]_I = \sum_{j=1}^N [S_n^{(L)}, S_n^{(L)}]_I$. Meanwhile,

$$E(S_{n,I})^2 = E\sum_{i=K}^I (\zeta_{n,i})^2 + 2E\sum_{i=K}^I\sum_{j=1}^{N-1} \zeta_{n,i}\zeta_{n,i-j}$$

[17] As convenient, we can take some ζ's in the beginning to be zero if the sum starts at K or similar. Definitely $\zeta_{n,i} = 0$ for $i < N$. For an example of such a structure, one can take $\zeta_{n,i} = e'_{n,T_i}$ or $= \tilde{e}'_{n,T_i}$, with $\mathcal{H}_{n,i} = \mathcal{F}_{T_{i+J}}$ and $N = 2J$. This construction is also used in Lemma 3.

$$
\begin{aligned}
&= E\sum_{i=K}^{I}(\zeta_{n,i})^2 + 2E\sum_{j=1}^{N-1}\sum_{i=K}^{I}\zeta_{n,i}\zeta_{n,i-j} \\
&\le E\sum_{i=K}^{I}(\zeta_{n,i})^2 + 2(N-1)E[S_n,S_n]_I \text{ (Cauchy–Schwarz)} \\
&= (2N-1)E[S_n,S_n]_I. \qquad \text{(C.3)}
\end{aligned}
$$

Hence, $(S_{n,I})^2$ is Lenglart-dominated (Jacod and Shiryaev, 2003, Section I.3c, pp. 35–36, Jacod and Protter, 2012, Section 2.1.7, p. 45) by $(2N-1)[S_n,S_n]_I$, and hence also by $(2N-1)\langle S_n,S_n\rangle_I^{(N)}$. By the same reasoning as in the proof of Jacod and Protter (2012, Proposition 2.2.5, p. 574), the result follows. □

References

Aït-Sahalia, Y., Xiu, D., 2019. A Hausman test for the presence of market microstructure noise in high frequency data. J. Econometrics 211, 176–205.
Andersen, T.G., Dobrev, D., Schaumburg, E., 2012. Jump-robust volatility estimation using nearest neighbor truncation. J. Econometrics 169, 75–93.
Andersen, T.G., Dobrev, D., Schaumburg, E., 2014. A robust truncation approach to estimation of integrated quaticity. Econometric Theory 39, 3–59.
Barndorff-Nielsen, O.E., Hansen, P.R., Shephard, N., 2008. Designing realised kernels to measure ex-post variation of equity prices in the presence of noise. Econometrica 76 (6), 1481–1536.
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. http://dx.doi.org/10.1111/sjos.12230.
Boyd, S., Vandenberghe, L., 2004. Convex Optimization. Cambridge University Press, Cambridge, UK.
Brockwell, P.J., Davis, R.A., 1987. Time Series: Theory and Methods. Springer-Verlag, New York.
Chan, N.H., 2011. Time Series: Applications to Finance with R and S-Plus, second ed. John Wiley & Sons.
Chen, D., Mykland, P.A., Zhang, L., 2020. The five trolls under the bridge: Principal component analysis with asynchronous and noisy high frequency data. J. Amer. Statist. Assoc. http://dx.doi.org/10.1080/01621459.2019.1672555, (in press).
Hall, P., Heyde, C.C., 1980. Martingale Limit Theory and its Application. Academic Press, Boston.
Jacod, J., Protter, P., 2012. Discretization of Processes, first ed. Springer-Verlag, New York.
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.
McLeish, D.L., 1975. A maximal inequality and dependent strong laws. Ann. Probab. 3 (5), 829–839.
Mykland, P.A., 1994. Bartlett type identities for martingales. Ann. Statist. 22, 21–38.
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., 2017a. Assessment of uncertainty in high frequency data: The observed asymptotic variance. Econometrica 85 (1), 197–231.
Mykland, P.A., Zhang, L., 2017b. Supplement to "assessment of uncertainty in high frequency data: The observed asymp- totic variance". Econometrica Suppl. Mater. 85.
Mykland, P.A., Zhang, L., 2020. Superblocks of Pre-Averaged Data: First Aid for Microsctructure, and a Path to Efficiency, Discussion paper.
Mykland, P.A., Zhang, L., Chen, D., 2019. The algebra of two scales estimation, and the S-TSRV: High frequency estimation that is robust to sampling times. J. Econometrics 208, 101–119.
Weisberg, S., 1985. Applied Linear Regression, second ed. Wiley, New York.
Withers, C.S., 1981. Conditions for linear processes to be strong-mixing. Z. Wahrscheinlichkeitstheor. Verwandte Geb. 57, 477–480.
Wu, W.B., Woodroofe, M., 2004. Martingale approximations for sums of stationary processes. Ann. Probab. 32, 1674–1690.
Zhang, L., 2006. Efficient estimation of stochastic volatility using noisy observations: A multi-scale approach. Bernoulli 12, 1019–1043.
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.