Efficient asymptotic variance reduction when estimating volatility in high frequency data[☆]

Simon Clinet [a,b], Yoann Potiron [c,*]

[a] *Faculty of Economics, Keio University, 2-15-45 Mita, Minato-ku, Tokyo, 108-8345, Japan*
[b] *Japan Science and Technology Agency, CREST, Japan*
[c] *Faculty of Business and Commerce, Keio University, 2-15-45 Mita, Minato-ku, Tokyo, 108-8345, Japan*

ARTICLE INFO

ABSTRACT

This paper shows how to carry out efficient asymptotic variance reduction when estimating volatility in the presence of stochastic volatility and microstructure noise with the realized kernels (RK) from Barndorff-Nielsen et al. (2008) and the quasi-maximum likelihood estimator (QMLE) studied in Xiu (2010). To obtain such a reduction, we chop the data into B blocks, compute the RK (or QMLE) on each block, and aggregate the block estimates. The ratio of asymptotic variance over the bound of asymptotic efficiency converges as B increases to the ratio in the parametric version of the problem, i.e. 1.0025 in the case of the fastest RK Tukey-Hanning 16 and 1 for the QMLE. The impact of stochastic sampling times and jump in the price process is examined carefully. The finite sample performance of both estimators is investigated in simulations, while empirical work illustrates the gain in practice.

© 2018 Elsevier B.V. All rights reserved.

1. Introduction

Over the past decades, the availability of high frequency data has led to a better understanding of asset prices. The main object of interest, the quadratic variation, can be used for example as a proxy for the spot volatility or the volatility parameter of a time-varying model. Moreover, forecasts of future volatility can be improved with it. Without microstructure noise, the realized variance (RV) estimator (e.g. Andersen et al., 2001, Meddahi, 2002, Barndorff-Nielsen and Shephard, 2002) is both

---

consistent and efficient. The convergence rate $n^{1/2}$ and the asymptotic variance (AVAR) were established in Genon-Catalot and Jacod (1993), Jacod (1994) and Jacod and Protter (1998) (see also Zhang, 2001, Mykland and Zhang, 2006).

Under market frictions, the RV is no longer consistent. Zhang et al. (2005) bring forward the Two-Scale Realized Volatility nonparametric estimator, the first consistent estimator in the presence of noise and with a relatively slow convergence rate of $n^{1/6}$. Zhang (2006) modifies it to provide the Multi-Scale Realized Volatility (MSRV) which features the optimal rate of convergence $n^{1/4}$ as documented in Gloter and Jacod (2001). Other approaches consist in and are not limited to: pre-averaging (PAE) the observations (Jacod et al., 2009), Barndorff-Nielsen et al. (2008) advocates for the realized kernels (RK) and (Xiu, 2010) studies the quasi-maximum likelihood estimator (QMLE) which was originally considered in Aït-Sahalia et al. (2005) when volatility is constant. Those three approaches share the optimal rate property and only differ through edge-effects which impact their respective AVAR.

The nonparametric AVAR bound of efficiency is equal to $8a_0 T^{\frac{1}{2}} \int_0^T \sigma_u^3 du$, where T stands for the time horizon and a_0^2 corresponds to the noise variance. This was shown in Reiss (2011) under the deterministic volatility and Gaussian noise setting, but it is commonly assumed that it stays true under stochastic volatility. Subsequently, in a recent breakthrough paper, Altmeyer and Bibinger (2015) found an estimator based on the spectral approach introduced in Reiss (2011) which reaches the bound in a very general situation. More recently, Jacod and Mykland (2015) proposed an adapted version of the pre-averaging estimator using local estimates as in Reiss (2011) which gave rise to estimators that are within 7% of the bound.

To be fair when comparing several estimators, we need the candidates to be equipped with the same technology. Following closely the local technique used in Reiss (2011) and more recently in Jacod and Mykland (2015), we aim to adapt accordingly the RK and the QMLE. Indeed, although both estimators behave remarkably well when volatility is constant, i.e. in the parametric case the ratio of AVAR over the bound of asymptotic efficiency is 1.0025 when considering the most efficient Tukey-Hanning 16 RK and 1 for the QMLE, they can actually be highly inefficient in the non-parametric setting as documented in the following of this introduction and in Section 2. Under time-varying volatility, we aim to reduce significantly their AVAR and make them efficient. Although it would reduce the AVAR the same way, we did not implement the local version of the MSRV. In fact, MSRV and RK are asymptotically equivalent in the sense that they share the same asymptotic variance when considering the same kernel (see Section 2.2 in Bibinger and Mykland (2016)).

To reduce the variance, we divide the interval $[0, T]$ into B non-overlapping regular blocks $[0, T/B]$, $[T/B, 2T/B]$, ..., $[(B-1)T/B, T]$. We then compute the RK (QMLE) on each block, and take the sum of the B estimates. We show that the nonparametric ratio of AVAR over the bound of efficiency converges to the parametric ratio as B increases. More importantly for practical applications, the convergence is very fast, and the gain is already important in the case $B = 2$ blocks.

As an example, we focus on the RK Tukey-Hanning 16 and consider the (apparently innocuous) block constant model $\sigma_t = 1$ for $t \in [0, \frac{1}{2})$ and $\sigma_t = 2$ for $t \in [\frac{1}{2}, 1]$. When choosing the optimal bandwidth, Barndorff-Nielsen et al. (2008)[1] showed that the AVAR is equal to

$$AVAR_{[0,T]}^{(RK)} = a_0 \left(T \int_0^T \sigma_u^4 du \right)^{3/4} g, \qquad (1.1)$$

where g is defined as

$$g = \frac{16}{3} \sqrt{\rho k_\bullet^{0,0} k_\bullet^{1,1}} \left(\frac{1}{\sqrt{1 + \sqrt{1 + 3d/\rho^2}}} + \sqrt{1 + \sqrt{1 + 3d/\rho^2}} \right),$$

with

$$\rho = \frac{\int_0^T \sigma_u^2 du}{\sqrt{T \int_0^T \sigma_u^4 du}}, \, d = \frac{k_\bullet^{0,0} k_\bullet^{2,2}}{(k_\bullet^{1,1})^2},$$

and where $k_\bullet^{i,i}$ are constant functions of the kernel. We fix $T = 1$ and we compute in that case $\int_0^1 \sigma_u^2 du = 5/2$, $\int_0^1 \sigma_u^3 du = 9/2$ and $\int_0^1 \sigma_u^4 du = 17/2$. Thus, the bound of efficiency is equal to $36a_0$, whereas $AVAR_{[0,1]}^{(RK)} = 37.89a_0$. This can be expressed as a loss of $\frac{37.89-36}{36} \approx 5\%$, which is to be compared to the loss in the parametric case[2] $\frac{8.02-8}{8} \approx .25\%$. When fixing $B = 2$, the volatility on each block is constant and thus yields $AVAR_{[0,1/2]}^{(RK)} = 2^{-3/2} \times 8.02a_0$ on the first block and $AVAR_{[1/2,1]}^{(RK)} = 2^{3/2} \times 8.02a_0$ on the second block. As both estimates are uncorrelated,[3] we obtain that the global AVAR is equal to $AVAR_{[0,1]}^{(RK)} = \sqrt{2}(AVAR_{[0,1/2]}^{(RK)} + AVAR_{[1/2,1]}^{(RK)}) = 8.02a_0 \int_0^1 \sigma_u^3 du$, i.e. .25% loss which corresponds exactly to the parametric loss.

---

[1] See pp. 1494–1495 for more details.

[2] Details can be found on Table II (p. 1495, Barndorff-Nielsen et al., 2008).

[3] If we remove end-effects.

From (1.1), we can see that the theoretical loss can be expressed as a deterministic function of the already well-known measure of volatility constancy ρ and another connected quantity which we denote

$$\kappa = \frac{\int_0^T \sigma_u^3 du}{T^{1/4}(\int_0^T \sigma_u^4 du)^{3/4}}.$$

Details can be found in Section 2, along with an expression for the QMLE loss as well. In the previous example where the loss was about 5%, the corresponding setting can be computed as $\rho = 5/2 \times \sqrt{2/17} \approx .86$ and $\kappa = 9/2 \times (2/17)^{3/4} \approx .90$. Volatility on real data is moving more than on this toy example, corresponding to lower ρ and κ. In their empirical study, Andersen et al. (2014) daily estimate ρ^{-1} and find that the typical value is around 1.3, and about 1.6 when restricting to the top 10% days in terms of intraday variation of volatility. This corresponds respectively to estimates of ρ as $1/1.3 \approx .77$ and $1/1.6 \approx .62$. When taking respectively those two realistic values, the corresponding RK and QMLE losses are expected to be around 20% (can go up to 100%), depending on the other parameter value κ. With such highly inefficient estimators, we believe that there is a practical need for variance reduction. This is especially the case on days when the volatility is moving a lot.

Clearly this estimator is related to local parametric methods in high-frequency data, i.e. aggregating local parametric estimates. For example, Mykland and Zhang (2009) investigated the ex post adjustment involving asymptotic likelihood ratios to make when assuming constant local volatility. Reiss (2011) showed the asymptotic equivalence in Le Cam's sense between the non-parametric and locally constant volatility experiment. To estimate quarticity and other functionals of volatility, Jacod and Rosenbaum (2013) estimated the volatility locally and plugged the value into the sum. Our work includes (Potiron and Mykland, 2017, 2016; Clinet and Potiron, 2018).

The remainder of the paper is structured as follows. Section 2 stretches the limitations of the global approach by expressing the loss as a function of ρ and κ. In Section 3, we provide the model, investigate the RK and the QMLE and their corresponding limit theory. Section 4 investigates what happens to both methods when considering stochastic arrival times and adding jump in the price process. Section 5 performs a Monte Carlo experiment to assess finite sample performance and AVAR reduction. Section 6 provides an empirical illustration where we quantify the expected gain in practice. Theoretical details and proofs can be found in the Appendix.

2. Limitations of the global approach

This section documents the performance of the global RK and QMLE. In particular, we show how it deteriorates as a function of heteroskedasticity. Finally, we diagnose the reasons and provide the solution to this relative failure.

One crucial feature common to both estimators is that they behave remarkably well when volatility is constant. Indeed, the QMLE is efficient and the RK Tukey-Hanning 16 almost efficient in that case. Even the RK Tukey-Hanning 2, with an AVAR over the bound of efficiency ratio of less than 1.04, can be considered as "practically efficient". To study what happens when volatility is time-varying, it is useful for $0 \le r < s \le T$ to define

$$\rho_{r,s} = \frac{\int_r^s \sigma_u^2 du}{\sqrt{(s-r)\int_r^s \sigma_u^4 du}} \text{ and } \kappa_{r,s} = \frac{\int_r^s \sigma_u^3 du}{(s-r)^{1/4}(\int_r^s \sigma_u^4 du)^{3/4}}$$

to be measures of heteroskedasticity. In the following, we will be using ρ and κ in place of $\rho_{0,T}$ and $\kappa_{0,T}$. The quantity ρ was already introduced in Barndorff-Nielsen et al. (2008) and plays an important role in the AVAR of both RK and the QMLE. Xiu (2010) (Figure 1, p. 241) expresses the quotient of both AVARs as a function of ρ, but does not assess their respective performance when compared to the (conjectured) bound of efficiency defined as

$$AVAR_{[0,T]}^{(Bound)} = 8a_0 T^{\frac{1}{2}} \int_0^T \sigma_u^3 du.$$

In contrast, the other quantity κ is introduced to investigate that relative performance. More precisely, κ is needed to express the AVAR over the bound of efficiency ratio for both approaches since the AVAR does not feature the tricity, i.e. the integrated third moment of volatility, which is key in the bound of efficiency. Evidently, both measures ρ and κ are very much connected and we can actually show that we have that

$$0 < \rho_{r,s}^{3/2} \le \kappa_{r,s} \le \rho_{r,s}^{1/2} \le 1. \tag{2.1}$$

Note that the equality $\rho_{r,s} = \kappa_{r,s} = 1$ for all $r, s \in [0, T]$ corresponds to the parametric case. In particular, Eq. (2.1) implies that for any given ρ, the value κ is a.s. in a small boundary around ρ. This is of particular interest because as far as the authors know under noisy observations the literature on quarticity estimation[4] is far more abundant than the corresponding work on estimating tricity,[5] which implies that in practice ρ can be estimated relatively easily, whereas κ would require more effort.

---

[4] See, e.g., Jacod et al. (2009), Andersen et al. (2014), Mancino and Sanfelici (2012), Potiron and Mykland (2016) and Clinet and Potiron (2017).

[5] See the spectral approach AVAR estimator in Altmeyer and Bibinger (2015), Potiron and Mykland (2016) and Clinet and Potiron (2017).

From Andersen et al. (2014) (Figure 7, p. 41), when taking a pre-averaging window equal to one minute (chosen consistently with their recommendation in Section 5.2.4 on p. 34 where the authors argue that a reasonable choice of window should lie between 30 s and 2 min) we infer that the estimates of ρ are about $1/1.2 \approx .83$, $1/1.3 \approx .77$ and $1/1.6 \approx .62$ when considering respectively the bottom 10% days in terms of intraday variation of volatility, all days and the top 10% days in terms of intraday variation of volatility. Correspondingly, we will be using $\rho_{high} = .83$, $\rho_{regular} = .77$, $\rho_{low} = .62$ to refer respectively to high, regular and low values of ρ throughout the rest of the paper. It is not surprising to find such low values on stocks data as it has been understood for several decades now that many stylized facts describe volatility as time-varying (see, e.g., Ghysels et al., 1996, Engle and Patton, 2001).

When using the optimal bandwidth, $AVAR_{[0,T]}^{(RK)}$ is defined as

$$AVAR_{[0,T]}^{(RK)} = a_0 \left(T \int_0^T \sigma_u^4 du \right)^{3/4} g,$$

where we have

$$g = \frac{16}{3} \sqrt{\rho k_\bullet^{0,0} k_\bullet^{1,1}} \left(\frac{1}{\sqrt{1 + \sqrt{1 + 3d/\rho^2}}} + \sqrt{1 + \sqrt{1 + 3d/\rho^2}} \right)$$

$$\text{and } d = \frac{k_\bullet^{0,0} k_\bullet^{2,2}}{(k_\bullet^{1,1})^2},$$

with $k_\bullet^{i,i}$ constant functions of the kernel. Correspondingly, we give the formal definition of the RK loss as

$$L^{(RK)} = \frac{AVAR_{[0,T]}^{(RK)}}{AVAR_{[0,T]}^{(Bound)}} - 1. \tag{2.2}$$

Obvious computations lead to $L^{(RK)} = g\kappa^{-1}/8 - 1$. If we see g as a function of ρ, $L^{(RK)}$ is equal to $g(1)/8 - 1$ in the parametric case. The parametric values for several kernels can be directly inferred from Barndorff-Nielsen et al. (2008) (Table II, p. 1495) and the loss is equal to .25% when considering the Tukey-Hanning 16, 3.625% for the Tukey-Hanning 2, 6.75% for the Parzen and 13% for the Cubic kernel. We have that g is an increasing function of ρ, and thus the effect of ρ and κ are reverse. Next we consider the AVAR of the QMLE expressed via

$$AVAR_{[0,T]}^{(QMLE)} = \frac{5Ta_0 \int_0^T \sigma_u^4 du}{(\int_0^T \sigma_u^2 du)^{1/2}} + 3a_0 \left(\int_0^T \sigma_u^2 du \right)^{3/2}.$$

The formula can actually be found in Box V (p. 240, Xiu, 2010). The corresponding QMLE loss is defined in analogy with (2.2) and can be expressed as

$$L^{(QMLE)} = \frac{AVAR_{[0,T]}^{(QMLE)}}{AVAR_{[0,T]}^{(Bound)}} - 1$$

$$= \frac{5 + 3\rho^2}{8\kappa \rho^{1/2}} - 1.$$

Fig. 1 plots the feasible loss region for three typical RK, the QMLE and the PAE with triangle kernel. It is clear that they highly lose efficiency when ρ is decreasing. The QMLE is dominated by the RK approach when ρ is low, which was observed on Fig. 1 (p. 241, Xiu, 2010).

The problem behind this potentially high loss can be intuitively explained as follows. For the RK, although the optimal tuning parameter is robust to time-varying volatility, it suffers from the fact that one day[6] is too long to "stay optimal". This is a very similar situation to the PAE, which also features a tuning parameter. Subsequently, Jacod and Mykland (2015) used block estimations to heavily reduce variance. As for the QMLE, which in contrast is designed in a parametric way yielding no choice of tuning parameter, the smaller ρ and κ are, the further the misspecified model deviates from the truth. It is by nature a different estimator, but local methods are expected to reduce the misspecification as in Reiss (2011). Thus, we aim to reduce the non-parametric loss into the parametric loss using adapted local methods. As we can see on Fig. 1, the QMLE will benefit the most as it is efficient in the parametric case and deteriorates more than the RK in the non-parametric case.

---

[6] Or one week, one month, etc.
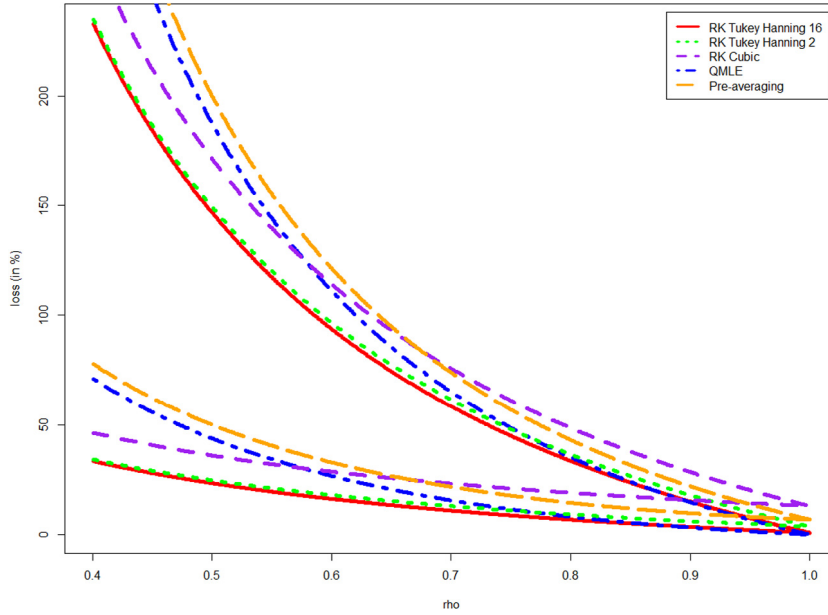
Fig. 1. Feasible loss region for three typical RK (Tukey-Hanning 16, Tukey-Hanning 2, Cubic), the QMLE and the PAE with triangle kernel. For each estimator, the lower line corresponds to the lower boundary when considering the best possible scenario $\kappa = \rho^{1/2}$ and the upper line stands for the upper boundary in the worst case scenario $\kappa = \rho^{3/2}$. The feasible loss region lies between those two lines. Note that a loss of 100% corresponds to an AVAR twice as big as the bound of efficiency.

3. Local estimation

3.1. Model for the observations

We assume that the latent log-price process and the volatility follow

$$dX_t = b_t dt + \sigma_t dW_t, \tag{3.1}$$
$$d\sigma_t = \tilde{b}_t dt + \tilde{\sigma}_t^{(1)} dW_t + \tilde{\sigma}_t^{(2)} d\tilde{W}_t + d\tilde{J}_t, \tag{3.2}$$

where $(W_t, \tilde{W}_t)$ is a 2 dimensional standard Brownian motion, the drift $(b_t, \tilde{b}_t)$ is componentwise locally bounded, the volatility matrix

$$\begin{pmatrix} \sigma_t & 0 \\ \tilde{\sigma}_t^{(1)} & \tilde{\sigma}_t^{(2)} \end{pmatrix}$$

is componentwise locally bounded, itself an Itô process and $\inf_t(\min(\sigma_t, \tilde{\sigma}_t^{(2)})) > 0$ a.s. We also assume that $\tilde{J}_t$ is a pure jump process of finite activity. This rules out jumps in X_t, an issue addressed in Section 4. In contrast the volatility process can include jumps (see, e.g., Todorov and Tauchen, 2011 for empirical evidence). The observations are contaminated by the microstructure noise so that we observe

$$Z_{t_i} = X_{t_i} + \epsilon_{t_i},$$

where t_i correspond to the observation times[7] which are assumed to be regularly spaced, i.e. satisfying $t_i - t_{i-1} = \Delta$. Stochastic arrival times are also considered in Section 4 Furthermore, we assume that the noise is independent and identically distributed (i.i.d), and independent of the other quantities, with null-mean, variance a_0^2 and finite fourth moment. Next the horizon time is defined as $T > 0$. Finally, we consider the high frequency asymptotics and assume that n goes to infinity, where $T = n\Delta$. In particular, the time gap Δ goes to 0.

---

7 Note that t_i, Δ, etc. are implicitly assumed to depend on the index n. We sometimes refer to t_i^n, Δ_n when necessary.

3.2. Realized kernels

3.2.1. Local RK definition

We consider first the framework $B = 1$ where the local RK coincides with the RK. The flat-top RK takes on the form

$$K = \gamma_0 + \sum_{h=1}^{H} k\left(\frac{h-1}{H}\right)(\gamma_h + \gamma_{-h}),$$

where $H > 0$ and the deterministic kernel $k(x)$ is defined for $x \in [0, 1]$. The realized autocovariance is defined as

$$\gamma_h = \sum_{j=1}^{n}(Z_{\Delta j} - Z_{\Delta(j-1)})(Z_{\Delta(j-h)} - Z_{\Delta(j-h-1)}),$$

where $h = -H, \ldots, -1, 0, 1, \ldots, H$.

In the general case $B > 1$, for each $i = 1, \ldots, B$ we choose a bandwidth $H_i > 0$ and define K_i the estimate on the ith block $[T_{i-1}, T_i]$, where $T_i = iT/B$. On each block, we also assume that the number of observations n/B is an integer for simplicity of exposition. Formally, all the considered quantities could be written with floor brackets, and all the results would still hold. We aggregate the local estimates to obtain the adapted version of the RK defined as

$$\tilde{K} = \sum_{i=1}^{B} K_i.$$

The corresponding $H = (H_1, \ldots, H_B)$ is now B-dimensional in this case. We also adapt the jittering introduced in Section 2.6 (Barndorff-Nielsen et al., 2008, p. 1487), i.e. for $i = 0, \ldots, B$ we assume that X_{T_i} is an average of m distinct observations on the interval $(T_i - \Delta, T_i + \Delta)$.

3.2.2. Asymptotic theory

We define L_X for $\sigma(X)$-stable convergence. We further define

$$\xi_{r,s}^2 = \frac{a_0^2}{\sqrt{(s-r)\int_r^s \sigma_u^4 du}}$$

as the noise-to-signal ratio, and refer to $\xi^2 = \xi_{0,T}^2$ in the following. Finally, we define kernel weight functions $k(x)$ that are two times continuously differentiable on $[0, 1]$ and

$$k_\bullet^{0,0} = \int_0^1 k(x)^2 dx, \ \ k_\bullet^{1,1} = \int_0^1 k'(x)^2 dx, \ \ k_\bullet^{2,2} = \int_0^1 k''(x)^2 dx.$$

We recall the main asymptotic result with fastest rate of convergence about the RK which can be found in Theorem 4 (p. 1493) in Barndorff-Nielsen et al. (2008). When $k'(0)^2 + k'(1)^2 = 0, m \to \infty$, and $H = cn^{1/2}$, we have

$$n^{1/4}\left(K - \int_0^T \sigma_u^2 du\right) \overset{L_X}{\to} \mathcal{MN}\Big(0, \underbrace{4T\int_0^T \sigma_u^4 du\{ck_\bullet^{0,0} + c^{-1}2k_\bullet^{1,1}\rho\xi^2 + c^{-3}k_\bullet^{2,2}\xi^4\}}_{AVAR_{[0,T]}^{(RK,c)}}\Big), \tag{3.3}$$

where $\mathcal{MN}$ denotes a mixed normal distribution. A straightforward application of (3.3) on each block $i = 1, \ldots, B$ yields

$$n^{1/4}\left(K_i - \int_{T_{i-1}}^{T_i} \sigma_u^2 du\right) \overset{L_X}{\to} \mathcal{MN}\left(0, B^{1/2}AVAR_{[T_{i-1},T_i]}^{(RK,c_i)}\right), \tag{3.4}$$

where c_i is the tuning parameter used on the ith block. Next we show that the AVAR associated to $\tilde{K}$ is equal to the sum of variance terms in (3.4).

Theorem 1 (*CLT for Local RK*). *When $k'(0)^2 + k'(1)^2 = 0, m \to \infty$, and $H = cn^{1/2}$, we have*

$$n^{1/4}\left(\tilde{K} - \int_0^T \sigma_u^2 du\right) \overset{L_X}{\to} \mathcal{MN}\left(0, B^{1/2}\sum_{i=1}^{B} AVAR_{[T_{i-1},T_i]}^{(RK,c_i)}\right). \tag{3.5}$$

Remark 1. The requirement that $m \to \infty$ in (3.5) is due to end-effects. The reader should refer to the discussion in Barndorff-Nielsen et al. (2008) (p. 1493) in the case $B = 1$. When m is fixed, the relative contribution[8] to the AVAR is proportional to

---

8 The corresponding expression can be found in the second term in (A.69).

ξ^2/m, as it was already the case for the RK. Barndorff-Nielsen et al. (2009) documented that this magnitude can reasonably be ignored in practice.

To determine the B tuning parameters that minimize the AVAR in (3.5), we can consider each local AVAR independently as they depend on one distinct tuning parameter. For that purpose, we follow Section 4.3 in Barndorff-Nielsen et al. (2008) (p. 1494–1496) and consider that

$$(H^{(1)}, \ldots, H^{(B)}) = (c_1 \xi_{0,T_1}, \ldots, c_B \xi_{T_{B-1},T}) \sqrt{n/B}.$$

The optimal values are then shown to be equal to

$$c_i^* = \sqrt{\rho_{T_{i-1},T_i} \frac{k_\bullet^{1,1}}{k_\bullet^{0,0}} \left(1 + \sqrt{1 + 3d/\rho_{T_{i-1},T_i}^2}\right)}.$$

The corresponding AVAR is equal to

$$AVAR_{[T_{i-1},T_i]}^{(RK,c_i^*)} = a_0 \left(\Delta_B \int_{T_{i-1}}^{T_i} \sigma_u^4 du\right)^{3/4} g(\rho_{T_{i-1},T_i}),$$

where g is considered here as a function of ρ and $\Delta_B = T/B$ corresponds to the block length.

We provide in what follows a consistent estimator for each tuning parameter. To pre-estimate on each block the integrated volatility and quarticity, we consider the pre-averaging estimators from Jacod et al. (2009). For each block $i = 1, \ldots, B$ we choose an integer k_i and a real parameter $\theta_i > 0$ which satisfy $k_i \sqrt{\Delta} = \theta_i + o(\Delta^{1/4})$. We also consider a continuous function f on $[0, 1]$, piecewise C^1 with a piecewise Lipschitz derivative f' such that $f(0) = f(1) = 0$, $\int_0^1 f(s)^2 ds > 0$. We define

$$\phi_1(s) = \int_s^1 f'(u)f'(u-s)du, \tag{3.6}$$

$$\phi_2(s) = \int_s^1 f(u)f(u-s)du, \tag{3.7}$$

$$\Phi_{jl} = \int_0^1 \phi_j(s)\phi_l(s)ds \text{ for } j, l = 1, 2, \tag{3.8}$$

$\psi_1 = \phi_1(0)$ and $\psi_2 = \phi_2(0)$. We further define

$$\bar{Z}_{l,i} = \sum_{j=1}^{k_i-1} f(j/k_i)(Z_{(l+j)\Delta} - Z_{(l+j-1)\Delta}).$$

The pre-averaging estimators of integrated volatility and quarticity on each block take on the form

$$\widehat{\int_{T_{i-1}}^{T_i} \sigma_u^2 du} = \frac{\sqrt{\Delta}}{\theta_i \psi_2} \sum_{j=0}^{n/B-k_i+1} \bar{Z}_{j,i}^2 - \frac{\psi_1 \Delta}{2\theta_i^2 \psi_2} \sum_{j=1}^{n/B} (Z_{(n(i-1)/B+j)\Delta} - Z_{(n(i-1)/B+j-1)\Delta})^2, \tag{3.9}$$

$$\widehat{\int_{T_{i-1}}^{T_i} \sigma_u^4 du} = \frac{1}{3\theta_i^2 \psi_2^2} \sum_{j=0}^{n/B-k_i+1} \bar{Z}_{j+(i-1)n/B,i}^4 \tag{3.10}$$

$$- \frac{\Delta \psi_1}{\theta_i^4 \psi_2^2} \sum_{j=0}^{n/B-2k_i+1} \bar{Z}_{j+(i-1)n/B,i}^2 \sum_{l=j+k_i}^{j+2k_i-1} (Z_{l\Delta} - Z_{(l-1+(i-1)n/B)\Delta})^2$$

$$+ \frac{\Delta \psi_1^2}{4\theta_i^4 \psi_2^2} \sum_{j=1}^{n/B-2} (Z_{(j+(i-1)n/B)\Delta} - Z_{(j-1+(i-1)n/B)\Delta})^2 (Z_{(j+2+(i-1)n/B)\Delta} - Z_{(j+1+(i-1)n/B)\Delta})^2.$$

We then estimate

$$\widehat{\rho}_{T_{i-1},T_i} = \frac{\widehat{\int_{T_{i-1}}^{T_i} \sigma_u^2 du}}{\sqrt{\Delta_B \widehat{\int_{T_{i-1}}^{T_i} \sigma_u^4 du}}}, \tag{3.11}$$

$$\widehat{c}_i^* = \sqrt{\widehat{\rho}_{T_{i-1},T_i} \frac{k_\bullet^{1,1}}{k_\bullet^{0,0}} \left(1 + \sqrt{1 + 3d/\widehat{\rho}_{T_{i-1},T_i}^2}\right)}. \tag{3.12}$$

We provide now a consistent estimator of $AVAR_B^{(RK)} = B^{1/2}\sum_{i=1}^{B} AVAR_{[T_{i-1},T_i]}^{(RK,c_i^*)}$. We estimate the noise as $\widehat{a}^2 = (2n)^{-1}\sum_{j=1}^{n}$ $\left(Z_{\Delta(j+1)} - Z_{\Delta j}\right)^2$ and the asymptotic variance as

$$\widehat{AVAR}_B^{(RK)} = \widehat{a}B^{1/2}\sum_{i=1}^{B}\left(\Delta_B \widehat{\int_{T_{i-1}}^{T_i} \sigma_u^4 du}\right)^{3/4} g(\widehat{\rho}_{T_{i-1},T_i}).$$

The feasible CLT is given in the following theorem.

Corollary 2 (*Feasible CLT for Local RK*). *When $k'(0)^2 + k'(1)^2 = 0$, $m \to \infty$, and $H = \widehat{c}n^{1/2}$ with $\widehat{c} = (\widehat{c}_1^*, \ldots, \widehat{c}_B^*)$, we have $\widehat{AVAR}_B^{(RK)} \xrightarrow{\mathbb{P}} AVAR_B^{(RK)}$ and*

$$n^{1/4}\frac{\tilde{K} - \int_0^T \sigma_u^2 du}{\sqrt{\widehat{AVAR}_B^{(RK)}}} \xrightarrow{\mathcal{L}} \mathcal{N}(0, 1). \tag{3.13}$$

Finally, we show that when choosing the optimal values, the AVAR associated to $\tilde{K}$ goes to $\frac{g(1)}{8}AVAR_{[0,T]}^{(Bound)}$ when $B \to \infty$. The constant $g(1)/8$, when normalized to $g(1)/8 - 1$, corresponds to the parametric loss and depends solely on the shape of the kernel. The rationale of such result is that when B increases we have the volatility roughly constant on each block and thus

$$\sum_{i=1}^{B} B^{1/2} AVAR_{[T_{i-1},T_i]}^{(RK,c_i^*)} = \sum_{i=1}^{B} a_0 B^{1/2}\left(\Delta_B \int_{T_{i-1}}^{T_i} \sigma_u^4 du\right)^{3/4} g(\rho_{T_{i-1},T_i}),$$

$$\approx \sum_{i=1}^{B} a_0 B^{1/2}\Delta_B^{3/2}\sigma_{T_{i-1}}^3 g(1).$$

Next we obtain by a Riemann sum argument that

$$\sum_{i=1}^{B} a_0 B^{1/2}\Delta_B^{3/2}\sigma_{T_{i-1}}^3 g(1) = a_0 T^{1/2}\sum_{i=1}^{B}\Delta_B \sigma_{T_{i-1}}^3 g(1),$$

$$\approx a_0 T^{1/2}\int_0^T \sigma_u^3 du g(1),$$

which can be expressed as $\frac{g(1)}{8}AVAR_{[0,T]}^{(Bound)}$. The formal result is given in the following proposition.

Proposition 3 (*Convergence of Local RK AVAR*). *When $B \to +\infty$, we have*

$$AVAR_B^{(RK)} \xrightarrow{a.s.} \frac{g(1)}{8}AVAR_{[0,T]}^{(Bound)}. \tag{3.14}$$

Remark 2. In particular, the asymptotic loss for $B \to +\infty$ is $g(1)/8 - 1$, which is always smaller than $L^{(RK)} = g\kappa^{-1}/8 - 1$ when using the RK with $B = 1$. The proof of this statement can be found in the Appendix (Appendix A.9).

3.3. QMLE

In analogy with Section 3.2, we provide in this section a definition of the local estimator and equivalent asymptotic results in the case of the QMLE.

3.3.1. Local QMLE definition

We consider first the setting $B = 1$ where the local QMLE is equal to the global QMLE. We recapitulate the parametric approach, and introduce the quasi-estimator. Aït-Sahalia et al. (2005) studied the parametric case assuming that the latent efficient log price process satisfies

$$dX_t = \sigma dW_t. \tag{3.15}$$

The observed log returns $Y_i = Z_{t_i} - Z_{t_{i-1}}$ are following a MA(1) process in that situation. If we postulate that the noise distribution is Gaussian, then the log likelihood function for $Y = (Y_1, \ldots, Y_n)^T$ can be expressed as

$$l(\sigma^2, a^2) = -\frac{1}{2}\log\det(\Omega) - \frac{n}{2}\log(2\pi) - \frac{1}{2}Y^T\Omega^{-1}Y, \tag{3.16}$$

where

$$
\Omega = \begin{pmatrix}
\sigma^2\Delta + 2a^2 & -a^2 & 0 & \cdots & 0 \\
-a^2 & \sigma^2\Delta + 2a^2 & -a^2 & \ddots & \vdots \\
0 & -a^2 & \sigma^2\Delta + 2a^2 & \ddots & 0 \\
\vdots & \ddots & \ddots & \ddots & -a^2 \\
0 & \cdots & 0 & -a^2 & \sigma^2\Delta + 2a^2
\end{pmatrix} \in \mathbb{R}^{n\times n}.
$$

We define the corresponding MLE which maximizes (3.16) as $(\widehat{\sigma}^2, \widehat{a}^2)$ and the estimator of integrated volatility as $Q = T\widehat{\sigma}^2$. When the log price X_t features stochastic volatility and drift as in Section 3.1 and/or when the noise is not normally distributed, $(\widehat{\sigma}^2, \widehat{a}^2)$ is seen as the QMLE.

When $B > 1$, we define for each block $i \in \{1, \ldots, B\}$ a local QMLE estimator $(\widehat{\sigma}_i^2, \widehat{a}_i^2)$ which maximizes the expression $l(\sigma^2, a^2)$ applied to the observations on $(T_{i-1}, T_i]$ only, along with the local integrated volatility estimator $Q_i = \Delta_B \widehat{\sigma}_i^2$. We then construct the aggregate version of the QMLE as

$$
\tilde{Q} = \sum_{i=1}^{B} Q_i.
$$

3.3.2. Asymptotic theory

We state the main result in Xiu (2010) (Box V, p. 240). If we assume that $\int_0^T \sigma_u^2 du \in [\underline{\Sigma}, \overline{\Sigma}]$ with $0 < \underline{\Sigma} < \overline{\Sigma}$, we have

$$
\begin{pmatrix}
n^{1/4}\left(Q - \int_0^T \sigma_u^2 du\right) \\
n^{1/2}\left(\widehat{a}^2 - a_0^2\right)
\end{pmatrix}
\xrightarrow{L_X}
\mathcal{MN}\left(
\begin{pmatrix} 0 \\ 0 \end{pmatrix},
\begin{pmatrix}
AVAR_{[0,T]}^{(QMLE)} & 0 \\
0 & 2a_0^4 + \text{cum}_4[\epsilon]
\end{pmatrix}
\right),
\tag{3.17}
$$

where we recall that

$$
AVAR_{[0,T]}^{(QMLE)} = \frac{5Ta_0 \int_0^T \sigma_u^4 du}{\left(\int_0^T \sigma_u^2 du\right)^{1/2}} + 3a_0 \left(\int_0^T \sigma_u^2 du\right)^{3/2}.
$$

Also, $\text{cum}_4[\epsilon]$ refers to the fourth cumulant of ϵ_0. An obvious application of (3.17) for each block $i = 1, \ldots, B$ gives us that

$$
n^{1/4}\left(Q_i - \int_{T_{i-1}}^{T_i} \sigma_u^2 du\right) \xrightarrow{L_X} \mathcal{MN}\left(0, B^{1/2} AVAR_{[T_{i-1}, T_i]}^{(QMLE)}\right).
$$

We show in the following theorem that the AVAR associated to $\tilde{Q}$ can be decomposed as a sum of local AVARs scaled by $B^{1/2}$.

Theorem 4 (*CLT for Local QMLE*). *We have*

$$
\begin{pmatrix}
n^{1/4}\left(\tilde{Q} - \int_0^T \sigma_u^2 du\right) \\
n^{1/2}\left(B^{-1}\sum_{i=1}^{B}\widehat{a}_i^2 - a_0^2\right)
\end{pmatrix}
\xrightarrow{L_X}
\mathcal{MN}\left(
\begin{pmatrix} 0 \\ 0 \end{pmatrix},
\begin{pmatrix}
B^{1/2}\sum_{i=1}^{B} AVAR_{[T_{i-1}, T_i]}^{(QMLE)} & 0 \\
0 & 2a_0^4 + \text{cum}_4[\epsilon]
\end{pmatrix}
\right).
$$

We define $AVAR_B^{(QMLE)} = B^{1/2}\sum_{i=1}^{B} AVAR_{[T_{i-1}, T_i]}^{(QMLE)}$ which is estimated via

$$
\widehat{AVAR}_B^{(QMLE)} = B^{1/2}\sum_{i=1}^{B}\left\{
\frac{5\Delta_B\widehat{a}\,\widehat{\int_{T_{i-1}}^{T_i}\sigma_u^4 du}}{\left(\widehat{\int_{T_{i-1}}^{T_i}\sigma_u^2 du}\right)^{1/2}}
+ 3\widehat{a}\left(\widehat{\int_{T_{i-1}}^{T_i}\sigma_u^2 du}\right)^{3/2}
\right\}.
$$

The feasible theorem follows.

Corollary 5 (*Feasible CLT for Local QMLE*). *We have* $\widehat{AVAR}_B^{(QMLE)} \xrightarrow{\mathbb{P}} AVAR_B^{(QMLE)}$ *and*

$$
n^{1/4}\frac{\tilde{Q} - \int_0^T \sigma_u^2 du}{\sqrt{\widehat{AVAR}_B^{(QMLE)}}} \xrightarrow{\mathcal{L}} \mathcal{N}(0, 1).
\tag{3.18}
$$

We show now that the AVAR associated to $\tilde{Q}$ goes to $AVAR_{[0,T]}^{(Bound)}$ when B increases.

Proposition 6 (*Convergence of Local QMLE AVAR*). *When $B \to +\infty$, we have*

$$AVAR_B^{(QMLE)} \overset{a.s.}{\to} AVAR_{[0,T]}^{(Bound)}. \tag{3.19}$$

4. Is the local method robust to stochastic sampling times and jump in the price process?

We discuss in this section what happens to both approaches when considering stochastic arrival times and adding jump to the price process. Related work in the global case include (Da and Xiu, 2017) for the QMLE and (Varneskov, 2016) for the RK. We further inspect the AVAR behavior when $B \to +\infty$ in this situation. The results are mitigated. Reduction (conjectured to be efficient) is obtained in the case of stochastic arrival times on the one hand, but there are additional terms in the AVAR as $B \to +\infty$ when adding jumps on the other hand.

4.1. Central limit theory when B is fixed

We assume that the latent log-price process is now an Itô semimartingale defined by

$$dX_t = b_t dt + \sigma_t dW_t + dJ_t,$$

where b_t and σ_t satisfy the same conditions as in Section 3.1, and J_t is a pure jump process of finite activity.

For the observation times, we adopt the random discretization scheme of Jacod and Protter (2011) (see Section 14.1) and we assume that there exists an Itô semimartingale $\alpha_t > 0$ which satisfies Assumption 4.4.2 (p. 115) in Jacod and Protter (2011) and is locally bounded away from 0, and i.i.d $U_i > 0$ that are both independent of the other quantities, $\alpha_t \perp\!\!\!\perp U_i$, such that

$$t_0 = 0, \tag{4.1}$$

$$t_i = t_{i-1} + \Delta \alpha_{t_{i-1}} U_i, \tag{4.2}$$

where we recall that $\Delta = T/n$. Finally, we assume that $\mathbb{E}U_i = 1$, and that for any $q > 0$, $m_q := \mathbb{E}U_i^q \to m_{q,\infty} < \infty$ as $n \to +\infty$. Note that, defining $\pi_t := \sup_{i \geq 1} t_i - t_{i-1}$, the number of observations before t as $N_n(t) = \sup\{i \in \mathbb{N} - \{0\} | t_i \leq t\}$, we have $\pi_t \overset{\mathbb{P}}{\to} 0$ as $n \to +\infty$ and

$$\frac{N_n(t)}{n} \to^{u.c.p} \frac{1}{T} \int_0^t \alpha_s^{-1} ds, \tag{4.3}$$

where the convergence *u.c.p* means uniformly in probability on $[0, t]$ for any $t \in [0, T]$.[9] We further define $N_n = N_n(T)$.

As pointed out in Jacod and Protter (2011, p. 431), any deterministic grid satisfies the above conditions. Actually, this model can be considered as more general than the time deformation proposed by Barndorff-Nielsen et al. (2008) (Section 5.3, pp. 1505–1507) in the sense that more complex arrival times, such as a Poisson process independent of the other quantities fall under the model. On the contrary, assuming the existence of the quadratic variation of time (see, e.g., Assumption A on p. 1939 in Mykland and Zhang (2006)) is too general as our proofs require the existence of the quadratic covariation of time lags for all lags.[10]

Since the price process features possible jumps, the two estimators are no longer consistent to the integrated volatility, but they converge to the quadratic variation

$$T\bar{\sigma}_0^2 := \int_0^T \sigma_s^2 ds + \sum_{0 < s \leq T} \Delta J_s^2,$$

where $\Delta J_s = J_s - J_{s-}$ corresponds to the size of the jump if there is a jump at time s and 0 otherwise. Correspondingly we define on each block $i = 1, \ldots, B$ the new local target as

$$\Delta_B \bar{\sigma}_i^2 := \int_{T_{i-1}}^{T_i} \sigma_s^2 ds + \sum_{T_{i-1} < s \leq T_i} \Delta J_s^2.$$

The AVARs obtained in the robust theorems feature $\Delta_B \bar{\sigma}_i^2$ in place of integrated volatility, and the following quantity as an alternative for quarticity:

$$\mathcal{Q}_{(i)} = \Delta_B^{-1} \int_{T_{i-1}}^{T_i} \alpha_s^{-1} ds \left\{ \int_{T_{i-1}}^{T_i} \sigma_s^4 \alpha_s ds + \sum_{T_{i-1} < s \leq T_i} \Delta J_s^2 (\sigma_s^2 \alpha_s + \sigma_{s-}^2 \alpha_{s-}) \right\}.$$

---

[9] We can prove (4.3) using Lemma 14.1.5 in Jacod and Protter (2011). The uniformity is a consequence of the fact that N_n and $\int_0^\cdot \frac{1}{\alpha_s} ds$ are increasing processes and Property (2.2.16) in Jacod and Protter (2011).

[10] To see a condition on the first lag, one can look at Assumption B.vii on p. 37 in Li et al. (2016). This does not include other lags.

Correspondingly, we define substitutes for the measure of heteroskedasticity and the noise-to-ratio measure as

$$\widetilde{\rho}_{T_{i-1},T_i} = \frac{\Delta_B \bar{\sigma}_i^2}{\sqrt{\Delta_B \mathcal{Q}_{(i)}}} \text{ and } \widetilde{\xi}^2_{T_{i-1},T_i} = \frac{a_0^2}{\sqrt{\Delta_B \mathcal{Q}_{(i)}}}.$$

Moreover, we also introduce

$$R_{(i)} := \frac{\int_0^T \alpha_s^{-1} ds}{\int_{T_{i-1}}^{T_i} \alpha_s^{-1} ds}, \tag{4.4}$$

which corresponds to the asymptotic ratio of the total number of observations over the number of observations on the block i as we have $N_n(T)/(N_n(T_i) - N_n(T_{i-1})) \xrightarrow{\mathbb{P}} R_{(i)}$. Finally, we define $\mathcal{G}_T := \sigma\{U_i^n, \alpha_s, X_s | (i, n) \in \mathbb{N}^2, 0 \le s \le T\}$ and refer to $L_{\mathcal{G}}$ for stable convergence with respect to $\mathcal{G}_T$. We provide the CLT for the two approaches in what follows.

Theorem 7 (Robust CLT for Local RK). *When $k'(0)^2 + k'(1)^2 = 0$, $m \to \infty$, $c = (c_1, \ldots, c_B)$, and $H = cN_n^{1/2}$, we have*

$$N_n^{1/4}\left(\tilde{K} - T\bar{\sigma}_0^2\right) \xrightarrow{L_{\mathcal{G}}} \mathcal{MN}\left(0, AVAR^{(RK,rob,c)}\right), \tag{4.5}$$

where

$$AVAR^{(RK,rob,c)} = \sum_{i=1}^B R_{(i)}^{1/2} AVAR^{(RK,rob,c_i)}_{[T_{i-1},T_i]},$$

$$AVAR^{(RK,rob,c_i)}_{[T_{i-1},T_i]} = 4\Delta_B \mathcal{Q}_{(i)}\left\{c_i k_\bullet^{0,0} + c_i^{-1} 2k_\bullet^{1,1} \widetilde{\rho}_{T_{i-1},T_i} \widetilde{\xi}^2_{T_{i-1},T_i} + c_i^{-3} k_\bullet^{2,2} \widetilde{\xi}^4_{T_{i-1},T_i}\right\}.$$

The new optimal bandwidth is given by

$$\widetilde{c}_i^* = \sqrt{\widetilde{\rho}_{T_{i-1},T_i} \frac{k_\bullet^{1,1}}{k_\bullet^{0,0}}\left(1 + \sqrt{1 + 3d/\widetilde{\rho}^2_{T_{i-1},T_i}}\right)},$$

with local and global optimal variances respectively defined as

$$AVAR^{(RK,rob,\widetilde{c}_i^*)}_{[T_{i-1},T_i]} = a_0\left(\Delta_B \mathcal{Q}_{(i)}\right)^{3/4} g(\widetilde{\rho}_{T_{i-1},T_i}),$$

$$AVAR^{(RK,rob)}_B = \sum_{i=1}^B R_{(i)}^{1/2} AVAR^{(RK,rob,\widetilde{c}_i^*)}_{[T_{i-1},T_i]}.$$

As for the QMLE, the log likelihood function when $B = 1$ keeps the same form (3.16) but we replace n by N_n in the definition of Ω now defined as

$$\Omega = \begin{pmatrix} \sigma^2\widetilde{\Delta} + 2a^2 & -a^2 & 0 & \cdots & 0 \\ -a^2 & \sigma^2\widetilde{\Delta} + 2a^2 & -a^2 & \ddots & \vdots \\ 0 & -a^2 & \sigma^2\widetilde{\Delta} + 2a^2 & \ddots & 0 \\ \vdots & \ddots & \ddots & \ddots & -a^2 \\ 0 & \cdots & 0 & -a^2 & \sigma^2\widetilde{\Delta} + 2a^2 \end{pmatrix} \in \mathbb{R}^{N_n \times N_n},$$

where $\widetilde{\Delta} = T/N_n$. Each local QMLE estimator $(\widehat{\sigma}_i^2, \widehat{a}_i^2)$ is now defined as a maximizer of

$$l_{(i)}(\sigma^2, a^2) = -\frac{1}{2}\log\det(\Omega_{(i)}) - \frac{N_{n,(i)}}{2}\log(2\pi) - \frac{1}{2}Y_{(i)}^T \Omega_{(i)}^{-1} Y_{(i)}, \tag{4.6}$$

where $Y_{(i)}$ is the vector of price returns on the ith block, $N_{n,(i)} := N_n(T_i) - N_n(T_{i-1})$, and

$$\Omega_{(i)} = \begin{pmatrix} \sigma^2\widetilde{\Delta}_{(i)} + 2a^2 & -a^2 & 0 & \cdots & 0 \\ -a^2 & \sigma^2\widetilde{\Delta}_{(i)} + 2a^2 & -a^2 & \ddots & \vdots \\ 0 & -a^2 & \sigma^2\widetilde{\Delta}_{(i)} + 2a^2 & \ddots & 0 \\ \vdots & \ddots & \ddots & \ddots & -a^2 \\ 0 & \cdots & 0 & -a^2 & \sigma^2\widetilde{\Delta}_{(i)} + 2a^2 \end{pmatrix} \in \mathbb{R}^{N_{n,(i)} \times N_{n,(i)}},$$

with $\widetilde{\Delta}_{(i)} := \Delta_B/N_{n,(i)}$. If we assume that $T\bar{\sigma}_0^2 \in [\underline{\Sigma}, \overline{\Sigma}]$ we obtain the following theorem.

Theorem 8 (*Robust CLT for Local QMLE*). *We have*

$$
\begin{pmatrix} N_n^{1/4} \left(\tilde{Q} - T\bar{\sigma}_0^2 \right) \\ N_n^{1/2} \left(B^{-1} \sum_{i=1}^{B} \widehat{a}_i^2 - a_0^2 \right) \end{pmatrix} \overset{L_{\mathcal{G}}}{\to} \mathcal{MN} \left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} AVAR_B^{(QMLE,rob)} & 0 \\ 0 & AVAR_B^{(QMLE,\epsilon)} \end{pmatrix} \right),
$$

where

$$
AVAR_B^{(QMLE,rob)} = \sum_{i=1}^{B} R_{(i)}^{1/2} AVAR_{[T_{i-1},T_i]}^{(QMLE,rob)}
$$

$$
AVAR_{[T_{i-1},T_i]}^{(QMLE,rob)} = \frac{5a_0 \Delta_B^{1/2} \mathcal{Q}_{(i)}}{\bar{\sigma}_i} + 3a_0 \bar{\sigma}_i^3 \Delta_B^{3/2},
$$

$$
AVAR_B^{(QMLE,\epsilon)} = \frac{1}{B^2} \sum_{i=1}^{B} R_{(i)} \left\{ 2a_0^4 + \text{cum}_4[\epsilon] \right\}.
$$

4.2. The good case: robustness to stochastic arrival times

Here we assume a no-jump setting, i.e. $J_t = 0$. The following two propositions provide the AVAR asymptotic behavior when $B \to \infty$ for the two methods. The limit is very similar to that in the regular observation case, and thus the local method is robust to stochastic observation times. Note that the conjectured bound of efficiency is affected by the setting and takes the form

$$
8a_0 \left(\int_0^T \alpha_s^{-1} ds \right)^{1/2} \int_0^T \alpha_s^{1/2} \sigma_u^3 du.
$$

Proposition 9 (*Asymptotic Behavior of Local RK AVAR when Sampling Times are Stochastic*). *When $B \to +\infty$, we have*

$$
AVAR_B^{(RK,rob)} \overset{a.s.}{\to} 8g(1)a_0 \left(\int_0^T \alpha_s^{-1} ds \right)^{1/2} \int_0^T \alpha_s^{1/2} \sigma_s^3 ds.
$$

Proposition 10 (*Asymptotic Behavior of Local QMLE AVAR when Sampling Times are Stochastic*). *When $B \to +\infty$, we have*

$$
AVAR_B^{(QMLE,rob)} \overset{a.s.}{\to} 8a_0 \left(\int_0^T \alpha_s^{-1} ds \right)^{1/2} \int_0^T \alpha_s^{1/2} \sigma_s^3 ds.
$$

4.3. The bad case: adding jumps to the price process

In this section, the price process can feature jumps. Actually in such setting the AVAR of the RK tends to a big value as B increases, and that of QMLE explodes. This sheds light on a weak point of the local method in this case.

Proposition 11 (*Asymptotic Behavior of Local RK AVAR when $J \neq 0$*). *As $B \to +\infty$,*

$$
AVAR^{(RK,rob)} \overset{a.s.}{\to} 8g(1)a_0 \left(\int_0^T \alpha_s^{-1} ds \right)^{1/2} \int_0^T \alpha_s^{1/2} \sigma_s^3 ds
$$

$$
+ \frac{16}{3} a_0 \left(\frac{1}{\sqrt{2}} + \sqrt{2} \right) \sqrt{k_\bullet^{0,0} k_\bullet^{1,1}} \left(\int_0^T \alpha_s^{-1} ds \right)^{1/2} \sum_{0<s\leq T} \Delta J_s^2 \left(\sigma_s^2 \alpha_s + \sigma_{s-}^2 \alpha_{s-} \right)^{1/2}.
$$

Proposition 12 (*Asymptotic Behavior of Local QMLE AVAR when $J \neq 0$*). *As $B \to +\infty$,*

$$
AVAR_B^{(QMLE,rob)} \overset{a.s.}{\sim} 3a_0 B^{1/2} T^{-1/2} \left(\int_0^T \alpha_s^{-1} ds \right)^{1/2} \sum_{0<s\leq T} \alpha_s^{1/2} |\Delta J_s|^3 \overset{a.s.}{\to} +\infty.
$$

5. Numerical study

5.1. Goal of the study

In this section, we discuss theoretical AVAR reduction and we examine the performance of the local RK $\tilde{K}_B$ and the local QMLE $\tilde{Q}_B$ in a finite sample context for several values of B. We carry out Monte Carlo simulations for three different volatility models having realistic values of ρ. We then check whether asymptotic approximations of several statistics correctly kick in to illustrate to what extent the theory is affected when the sample data is finite of size n. First, we assess the central limit theories for the two infeasible statistics

$$Z_n^{\tilde{K}_B} = \frac{n^{1/4}\left(\tilde{K}_B - \int_0^T \sigma_u^2 du\right)}{\sqrt{AVAR_B^{(RK)}}}, Z_n^{\tilde{Q}_B} = \frac{n^{1/4}\left(\tilde{Q}_B - \int_0^T \sigma_u^2 du\right)}{\sqrt{AVAR_B^{(QMLE)}}},$$

and the two feasible statistics

$$\tilde{Z}_n^{\tilde{K}_B} = \frac{n^{1/4}\left(\tilde{K}_B - \int_0^T \sigma_u^2 du\right)}{\sqrt{\widehat{AVAR}_B^{(RK)}}}, \tilde{Z}_n^{\tilde{Q}_B} = \frac{n^{1/4}\left(\tilde{Q}_B - \int_0^T \sigma_u^2 du\right)}{\sqrt{\widehat{AVAR}_B^{(QMLE)}}},$$

for $B = 1, 2, 4, 6, 8$. In particular, we investigate how increasing B affects the standard normal approximation of these two studentizations for several levels of sampling. Second, we compare the relative performance of the local RK and the local QMLE. To do so, we report the empirical loss defined as

$$\breve{L}_B^{(RK)} = \mathbb{E}_M\left[\frac{n^{1/2}(\tilde{K}_B - \int_0^T \sigma_s^2 ds)^2}{AVAR_{[0,T]}^{(Bound)}}\right] - 1, \breve{L}_B^{(QMLE)} = \mathbb{E}_M\left[\frac{n^{1/2}(\tilde{Q}_B - \int_0^T \sigma_s^2 ds)^2}{AVAR_{[0,T]}^{(Bound)}}\right] - 1$$

where $\mathbb{E}_M[X]$ denotes the sample mean of X based on the M Monte Carlo simulations and we recall that $AVAR_{[0,T]}^{(Bound)} = 8a_0 T^{\frac{1}{2}} \int_0^T \sigma_u^3 du$ is the bound of efficiency for the asymptotic variance. We also define the theoretical loss as

$$L_B^{(\Sigma)} = \frac{AVAR_B^{(\Sigma)}}{AVAR_{[0,T]}^{(Bound)}} - 1.$$

and report the sample mean of the theoretical loss $\tilde{L}_B^{(\Sigma)} = \mathbb{E}_M\left[L_B^{(\Sigma)}\right]$ for $\Sigma \in \{RK, QMLE\}$. Note that $\tilde{L}_B^{(\Sigma)}$ is close to the mean loss $\mathbb{E}\left[L_B^{(\Sigma)}\right]$ if M is large enough. The empirical loss $\breve{L}^{(\Sigma)}$, which gives us a simple criterion to compare the estimators, can be decomposed as

$$\breve{L}_B^{(\Sigma)} = \underbrace{\tilde{L}_B^{(\Sigma)}}_{\text{theoretical loss due to the finiteness of } B} + \underbrace{(\breve{L}_B^{(\Sigma)} - \tilde{L}_B^{(\Sigma)})}_{\text{loss due to the finite sample } n}.$$

5.2. Simulation design

We implement the above procedures for $M = 10,000$ Monte Carlo simulations of intraday returns on the time interval $[0, T]$, $T = 1/252$ year (that is $T = 1$ working day). One working day is in turn subdivided in 23,400 s corresponding to 6.5 h of trading activity. For each model, the corresponding trajectories are generated from a classical Euler scheme based on $n = 46,800$ intervals, that is one observation every 0.5 s. We simulate 1000 more observations prior and post main trading period in order to compute properly different γ_h that are necessary for the RK. Indeed, using their truncated versions $\tilde{\gamma}_h = \sum_{j=H+1}^{n-H}(Z_{\Delta j} - Z_{\Delta(j-1)})(Z_{\Delta(j-h)} - Z_{\Delta(j-h-1)})$ tend to generate a non-negligible bias as pointed out in Xiu (2010) (see Table 2 on p. 243), so that we prefer to overcome this issue with a few minutes of out-of-sample data. Finally, we also use observations based on sparsely sampled versions of the original trajectories, for a number of intervals taking on the values 23,400, 11,700, and 5850, the latter corresponding to having one observation every 4 s, which still corresponds to a fairly heavily traded stock. We do not report the results for lower frequencies, but the theory still kicks in for sparser samplings too.

We consider three stochastic volatility models to simulate the intraday returns, along with three levels of mean noise-to-signal ratios $\xi^2 = 0.01$, $\xi^2 = 0.001$ and $\xi^2 = 0.0002$. The three values are empirically corroborated in Hansen and Lunde (2006), where the authors report empirical values of ξ^2 for several stocks ranging from 0.00004 to 0.006 (see Table 3 on p. 147). We introduce now the volatility models, which have been designed to reflect different average values of ρ ranging from 0.89 (corresponding to a high value) for Model 1, 0.77 (corresponding to a regular value) for Model 2 to 0.64 (corresponding to a low value) for Model 3 as reported on Table 1. The three models can all be represented as a Heston stochastic volatility model (SV) with U-shape intraday volatility pattern and a possible jump whose occurrence time is picked up uniformly randomly on a subinterval $[T^{(0)}, T^{(1)}]$ of $[0, T]$. Except for the jump component, this general model is directly inspired from

Table 1
Sample mean and standard error of ρ and κ for the three models.

Model	ρ_{mean}	$\rho_{stdv.}$	κ_{mean}	$\kappa_{stdv.}$
Model 1	0.89	0.01	0.92	0.01
Model 2	0.77	0.15	0.83	0.12
Model 3	0.64	0.14	0.74	0.1

Model 4 in Andersen et al. (2012), and Xiu (2010) (see Section 6.1 on p. 242). We assume that the log price process X_t and the volatility process σ_t follow the dynamics

$$dX_t = \mu dt + \sigma_{t-}dW_t,$$
$$\sigma_t = \sigma_{t,SV}\sigma_{t,U},$$

with

$$d\sigma_{t,SV}^2 = \alpha(\bar{\sigma}^2 - \sigma_{t,SV}^2)dt + \delta\sigma_{t,SV}d\bar{W}_t,$$
$$\sigma_{t,U} = C + Ae^{-at/T} + De^{-b(1-t/T)} - \beta\sigma_{\tau-,U}\mathbb{1}_{\{t\geq\tau\}}.$$

Here W_t and $\bar{W}_t$ are two standard Brownian motions with $d\langle W, \bar{W}\rangle_t = \phi dt$. Note that $\sigma_{t,U}$ jumps at time τ, that we define as a uniform random variable on $[T^{(0)}, T^{(1)}]$. β controls the size of the jump. The choice of making $\sigma_{t,U}$ jumps, instead of the global volatility σ_t, is merely a way to ensure that σ_t remains positive. Finally, the drift parameter μ and the stochastic volatility part remain constant for each model. The corresponding parameters are chosen consistently with the ones from Section 6.1 (p. 242) in Xiu (2010), that is $\mu = 0.03$, $\alpha = 5$, $\bar{\sigma}^2 = 0.1$, $\delta = 0.4$, $\phi = -0.75$. Finally, $\sigma_{0,SV}^2$ is sampled from a Gamma distribution of parameters $(2\alpha\bar{\sigma}^2/\delta^2, \delta^2/2\alpha)$, which corresponds to the stationary distribution of the CIR process.

Model 1: SV + steep U (HIGH ρ)

The first model does not incorporate the jump in volatility, i.e. we set $\beta = 0$. The parameters of the U-shape part are set to generate a steep slope, which in turn lowers somewhat the value of ρ compared to Model 4 in Andersen et al. (2012) where we find that the corresponding mean ρ value is too high to be consistent with ρ_{high} (which we recall is the empirical high value reported in Section 2). With $C = 0.83$, $A = 1.26$, $D = 0.42$, $a = 10$, $b = 10$, this model presents a sample mean value of $\rho_{mean} = 0.89$, which is slightly bigger than $\rho_{high} = 0.83$. We are conservative in this first model to show what happens to the local method in a very unlikely bad situation for AVAR reduction, i.e. a very high ρ_{mean}.

Model 2: SV + normal U + 1 Jump (REGULAR ρ)

In this model, the U-shape intraday volatility parameters are set to values that are consistent with those chosen in Model 4 in Andersen et al. (2012), that is $C = 0.75$, $A = 0.25$, $D = 0.89$, and $a = b = 10$. The jump size parameter is set to $\beta = 0.5$, that is a jump of 50% in size at the random time τ. We set $T^{(0)} = 0$, $T^{(1)} = T$ and thus let τ take values on the whole time interval. Such friction in the volatility process leads to lower values of ρ and κ compared to Model 1, with a sample mean equal to $\rho_{mean} = 0.77$. This is thus a very realistic model in terms of measure of heteroskedasticity as $\rho_{mean} = \rho_{regular}$. It is also possible to obtain $\rho_{mean} = \rho_{regular}$ in an alternative continuous volatility model with normal U by taking a 2-factor stochastic volatility model (SV2F) as in Barndorff-Nielsen et al. (2008) (Section 6.2, p. 1511), with parameters tuned such that the trajectories are rough enough. The results from Section 5.4 would be similar. As a byproduct, Model 2 shows that a jump in the volatility can lower significantly the measures of heteroskedasticity ρ and κ.

Model 3: SV + steep U + 1 Jump (LOW ρ)

This last model is a combination of the first two models. U-shape volatility parameters are set to give the same slope as for Model 1, and the jump size parameter is set to $\beta = 0.5$ as in Model 2. However, to keep the positivity of σ_t we restrain the values of the jump time and set $T^{(0)} = 0.05T$, $T^{(1)} = 0.7T$. This third scenario is designed to reach volatility paths presenting an heteroskedasticity with a low value of ρ and we report the sample mean $\rho_{mean} = 0.64$, which is almost equal to $\rho_{low} = 0.62$. We are in the situation where the global estimators should deviate the most from the bound of efficiency.

We now turn to the estimation procedure. First, to estimate K on $[0, T]$, we work with the Tukey-Hanning 2 kernel as for the numerical study in Barndorff-Nielsen et al. (2008) (Section 6, pp. 1510–1513) since it requires reasonable bandwidth sizes H, which makes the estimator computable in an acceptable amount of time. Moreover, we do not need too many out-of-period data to compute γ_h. We implement the feasible adaptive estimator. We arbitrary set the tuning parameters θ_i equal to 30 s and the triangular kernel $f(x) = x \wedge (1-x)$. In practice, we find that the realized kernel is not very sensitive to the dispersion of $\widehat{H}$ in terms of RMSE, so that it is not absolutely necessary to get very accurate pre-estimators. Such robustness proved to be crucial in our procedure as it is well known that estimators for the quarticity can be unstable in finite sample when the amount of data is not large. On each block $[T_{i-1}, T_i]$, we do the same procedure and obtain the corresponding $\tilde{K}_B$ by aggregation. Finally, we compute the QMLE by a numerical maximization of the quasi-likelihood function given in Section 3.3.1. This gives us $\tilde{Q}$ and the local estimates $\tilde{Q}_B$.
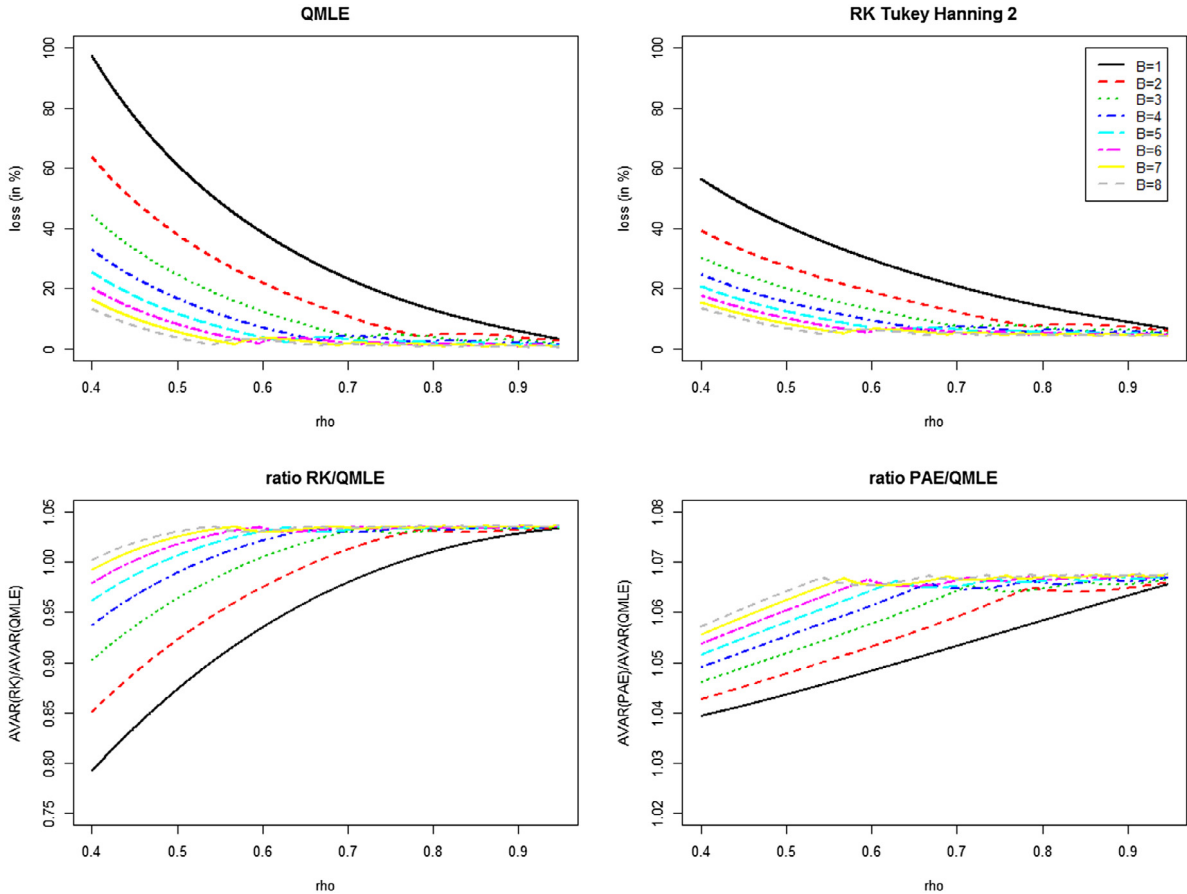
Fig. 2. For $B = 1, \ldots, 8$ we plot $L_B^{(QMLE)}$ (upper left panel), $L_B^{(RK)}$ for Tukey-Hanning 2 kernel (upper right panel), the corresponding AVAR ratio defined as $AVAR_B^{(RK)}/AVAR_B^{(QMLE)}$ (lower left panel) and the ratio of pre-averaging AVAR using B blocks over $AVAR_B^{(QMLE)}$ (lower right panel) as a function of ρ.

5.3. Discussion on theoretical AVAR reduction

In this section, we propose to look at the theoretical AVAR reduction as a function of ρ, and investigate the practical question of how fast the convergence in (3.14) and (3.19) is. The model considered for volatility is a deterministic U shape + 1 Jump, which corresponds to Model 2 without the stochastic volatility part. Here we generate different values of ρ as a function of the jump time, which we restrict to be in $\left[0.013T, T\right]$ so that each ρ can be associated to a distinct jump time on that interval. We choose this particular model because the sample mean of ρ is .77 which corresponds to a regular value, and the panel of generated ρ values is sufficiently large compared to the other two models.

The values of $L_B^{(RK)}$ and $L_B^{(QMLE)}$ are plotted as a function of ρ in the upper panels of Fig. 2 for a realistic continuous U-shape with one jump volatility model where the sample mean .77 corresponds to a regular value of ρ.[11] As we can see, the convergence in (3.14) is very fast. When $\rho = .77$, the QMLE loss is almost divided by 4 when considering 2 blocks instead of 1, with $L_1^{(QMLE)} \approx 16\%$ and $L_2^{(QMLE)} \approx 5\%$. In the same setting the RK loss goes from $L_1^{(RK)} \approx 16\%$ to $L_2^{(RK)} \approx 8\%$. If we consider the lower value $\rho = .62$, the QMLE losses for the first four values of B are equal to $L_1^{(QMLE)} \approx 35\%$, $L_2^{(QMLE)} \approx 19\%$, $L_3^{(QMLE)} \approx 11\%$ and $L_4^{(QMLE)} \approx 6\%$. The corresponding RK values are $L_1^{(RK)} \approx 28\%$, $L_2^{(RK)} \approx 17\%$, $L_3^{(RK)} \approx 11\%$ and $L_4^{(RK)} \approx 8\%$. This suggests that the convergence to the loss bounds (which we recall to be equal to $L_\infty^{(RK)} = 3.625\%$ when considering the RK Tukey-Hanning 2 and $L_\infty^{(QMLE)} = 0\%$ for the QMLE) is very fast for both approaches. Actually for any reasonable ρ taken to be between 0.5 and 1, choosing $B = 8$ is big enough for the loss to stay within $L_\infty^{(RK)} + 4\%$ (or $L_\infty^{(QMLE)} + 4\%$), and it is usually far below this threshold with regular and high values of ρ.

Moreover, we can see on the left lower panel in Fig. 2 that when ρ is relatively high, the QMLE outperforms the RK approach when considering $B = 1$, and the gap gets bigger as we increase B. In contrast when $\rho < .77$, the QMLE is

---

[11] Useful details on this model can be found in Section 5.

Table 2

Finite sample properties of $Z_n^{\tilde{K}_B}$ (Model 2).[a]

No. Obs.	Mean	Stdv.	RMSE	0.5%	2.5%	5%	95%	97.5%	99.5%
$B = 1$ block									
5,850	−0.042	1.102	1.103	0.29	1.75	3.77	96.62	98.62	99.85
11,700	−0.032	1.067	1.068	0.39	1.96	3.98	96.01	98.20	99.80
23,400	−0.030	1.044	1.044	0.41	2.13	4.16	95.63	97.84	99.70
46,800	−0.027	1.041	1.041	0.46	2.25	4.35	95.58	98.18	99.74
$B = 2$ blocks									
5,850	−0.065	1.105	1.106	0.24	1.55	3.53	96.49	98.43	99.82
11,700	−0.048	1.069	1.070	0.32	1.85	3.65	95.89	98.20	99.71
23,400	−0.042	1.048	1.049	0.37	2.01	3.91	95.52	97.88	99.65
46,800	−0.037	1.044	1.045	0.43	2.11	4.10	95.54	98.06	99.71
$B = 4$ blocks									
5,850	−0.105	1.110	1.115	0.21	1.38	3.02	96.25	98.37	99.81
11,700	−0.082	1.074	1.077	0.28	1.54	3.33	95.74	98.15	99.66
23,400	−0.069	1.051	1.053	0.36	1.77	3.66	95.22	97.73	99.65
46,800	−0.059	1.043	1.044	0.37	1.89	3.89	95.31	97.96	99.64
$B = 6$ blocks									
5,850	−0.144	1.115	1.124	0.19	1.23	2.72	95.81	98.33	99.75
11,700	−0.114	1.077	1.083	0.23	1.40	3.09	95.37	97.88	99.67
23,400	−0.099	1.054	1.059	0.31	1.65	3.49	94.95	97.57	99.60
46,800	−0.086	1.043	1.047	0.38	1.65	3.56	94.95	97.76	99.57
$B = 8$ blocks									
5,850	−0.193	1.119	1.136	0.15	1.03	2.31	95.40	98.14	99.75
11,700	−0.154	1.080	1.091	0.21	1.26	2.83	95.27	97.88	99.66
23,400	−0.128	1.054	1.062	0.28	1.52	3.27	94.91	97.56	99.59
46,800	−0.109	1.042	1.047	0.32	1.64	3.51	94.72	97.56	99.55

[a] This table shows summary statistics and empirical quantiles benchmarked to the $N(0, 1)$ distribution for the infeasible Z-statistics related to the global and local RK (Tukey–Hanning 2). The simulation design is Model 2 with $M = 10{,}000$ Monte-Carlo simulations.

outperformed when considering only one block, but eventually makes it back when incrementing the value of B. The actual value required to fill up the gap is getting bigger as ρ decreases. This suggests that both approaches are complementary to each other. Finally, the lower left panel in Fig. 2 documents that both approaches dominate the PAE regardless of the number of blocks.

5.4. Results

We first report the finite sample properties of the four statistics in Tables 2–5 for Model 2 under the noise level $\xi^2 = 0.001$. We can see that the results are promising at any level of sampling, as the RMSE of the Z-statistic does not suffer much from the increasing in the number of blocks, especially for the QMLE for which the RMSE of $Z_n^{\hat{Q}_B}$ stays closely in line with Z_n^Q. The results also indicate that the asymptotic theory eventually kicks in for all the estimators as the standard deviation of the statistics decreases to 1 when the sampling frequency increases. Nevertheless, we can see a slight over dispersion compared to what was reported in Xiu (2010) and Barndorff-Nielsen et al. (2008). For the QMLE, this is due to the strong difference with the noise-to-signal ratio that was used in Xiu (2010) where $\xi^2 \approx 0.06$. Concerning the RK, the difference in the studentization is due to the fact that the authors in Barndorff-Nielsen et al. (2008) do not employ $AVAR_{[0,T]}^{(RK)}$ for the studentization, but a non-asymptotic variance as documented in Section 4.4 (pp. 1496–1498) of their work. The feasible statistics are slightly biased, and this is due to the estimation of the AVAR procedure.

We then report the theoretical loss values $\tilde{L}_B^{(\Sigma)}$ and the empirical loss $\check{L}_B^\Sigma$ for two levels of sampling $n = 23{,}400$ and $n = 46{,}800$, and three levels of noise-to-signal ratios $\xi^2 = 0.01$, $\xi^2 = 0.001$ and $\xi^2 = 0.0002$ in Table 6. First, we can note that the theoretical loss behaves in a very similar way as in Section 5.3 for the three models. In particular, this implies that neither the SV part nor the steep U component seems to have a bad impact for the local method. Also, one can see that when choosing $B = 8$ the theoretical loss is at most 3.2% more than the parametric loss (which we recall to be equal to 3.625% for the RK Tukey–Hanning 2 and 0% in the case of the QMLE), which are in line with the threshold found in Section 5.3.

Second, the loss due to the finite sample behaves in a very proper way when B increases. For any setting and both estimators, it is roughly constant as a function of B, although suffering more when ρ is higher and n smaller. This is perfectly in line with the findings in Tables 2 and 3. In particular for Model 2 and Model 3, the finite sample effect is almost not moving as B increases. For Model 1, this is basically the same picture for the QMLE, but the empirical loss seems to stagnate between $B = 4$ and $B = 8$ when using the RK. This is not surprising as the RK suffers more from the finite sample effect than the QMLE as seen in Tables 2 and 3.

Table 3
Finite sample properties of $Z_n^{\tilde{Q}_B}$ (Model 2).[a]

No. Obs.	Mean	Stdv.	RMSE	0.5%	2.5%	5%	95%	97.5%	99.5%
B = 1 block									
5,850	−0.024	1.084	1.084	0.36	2.09	4.12	96.48	98.57	99.84
11,700	−0.015	1.058	1.058	0.43	2.26	4.51	96.32	98.34	99.75
23,400	−0.012	1.039	1.039	0.51	2.19	4.48	95.87	97.97	99.70
46,800	−0.013	1.034	1.034	0.59	2.38	4.67	95.74	98.06	99.73
B = 2 blocks									
5,850	−0.023	1.086	1.086	0.34	1.90	4.03	96.48	98.46	99.85
11,700	−0.011	1.06	1.06	0.42	2.08	4.28	96.24	98.31	99.72
23,400	−0.007	1.042	1.042	0.54	2.12	4.31	95.71	98.01	99.66
46,800	−0.009	1.036	1.036	0.56	2.22	4.51	95.71	98.08	99.63
B = 4 blocks									
5,850	−0.016	1.089	1.089	0.34	1.98	3.86	96.42	98.57	99.82
11,700	−0.007	1.063	1.063	0.42	2.09	4.19	96.24	98.34	99.72
23,400	−0.002	1.042	1.042	0.51	2.16	4.52	95.64	98.02	99.65
46,800	−0.005	1.035	1.035	0.56	2.20	4.74	95.60	98.08	99.69
B = 6 blocks									
5,850	−0.012	1.089	1.089	0.36	2.01	3.93	96.46	98.56	99.82
11,700	−0.002	1.062	1.062	0.43	2.01	4.26	96.33	98.33	99.74
23,400	−0.0	1.041	1.041	0.51	2.10	4.63	95.60	98.06	99.71
46,800	−0.004	1.033	1.033	0.57	2.21	4.72	95.64	98.07	99.69
B = 8 blocks									
5,850	−0.014	1.093	1.093	0.36	1.82	3.83	96.42	98.63	99.82
11,700	−0.005	1.066	1.066	0.40	1.94	4.15	96.33	98.37	99.74
23,400	−0.001	1.043	1.043	0.47	2.05	4.53	95.64	98.15	99.67
46,800	−0.003	1.033	1.033	0.57	2.29	4.67	95.60	98.10	99.66

[a] This table shows summary statistics and empirical quantiles benchmarked to the $N(0, 1)$ distribution for the infeasible Z-statistics related to the global and local QMLE. The simulation design is Model 2 with $M = 10{,}000$ Monte-Carlo simulations.

Table 4
Finite sample properties of $\tilde{Z}_n^{\tilde{K}_B}$ (Model 2).[a]

No. Obs.	Mean	Stdv.	RMSE	0.5%	2.5%	5%	95%	97.5%	99.5%
B = 1 block									
5,850	−0.117	1.176	1.182	0.02	0.66	1.67	95.42	97.67	99.66
11,700	−0.080	1.128	1.131	0.01	0.58	2.01	95.00	97.08	99.38
23,400	−0.083	1.098	1.101	0.02	0.52	2.26	94.89	97.22	99.04
46,800	−0.069	1.087	1.089	0.02	0.73	3.44	94.49	97.27	99.15
B = 2 blocks									
5,850	−0.130	1.148	1.155	0.02	0.58	1.76	95.40	97.27	99.40
11,700	−0.099	1.111	1.115	0.01	0.66	2.17	94.53	96.23	99.08
23,400	−0.086	1.083	1.086	0.05	0.71	2.28	95.03	97.17	99.24
46,800	−0.072	1.071	1.073	0.06	1.17	4.10	94.74	97.63	99.41
B = 4 blocks									
5,850	−0.173	1.136	1.149	0.03	0.64	1.71	94.56	97.28	98.98
11,700	−0.139	1.107	1.115	0.01	0.64	1.78	93.28	96.21	99.13
23,400	−0.107	1.083	1.089	0.08	0.88	2.37	94.97	96.83	99.03
46,800	−0.092	1.079	1.083	0.03	1.07	3.93	94.99	97.64	99.34
B = 6 blocks									
5,850	−0.225	1.145	1.167	0.02	0.60	1.25	93.88	96.93	98.80
11,700	−0.177	1.103	1.117	0.01	0.53	1.52	93.21	95.86	99.04
23,400	−0.145	1.077	1.087	0.04	0.72	1.91	94.12	96.61	99.06
46,800	−0.122	1.07	1.077	0.04	1.01	3.55	94.82	97.07	99.35
B = 8 blocks									
5,850	−0.270	1.152	1.183	0.01	0.45	1.080	94.41	96.88	98.66
11,700	−0.219	1.106	1.128	0.02	0.50	1.65	92.63	95.76	98.96
23,400	−0.176	1.089	1.103	0.07	0.55	1.96	93.63	97.34	98.97
46,800	−0.146	1.078	1.088	0.03	0.72	3.27	94.02	96.98	99.38

[a] This table shows summary statistics and empirical quantiles benchmarked to the $N(0, 1)$ distribution for the feasible Z-statistics related to the global and local RK (Tukey-Hanning 2). The simulation design is Model 2 with $M = 10{,}000$ Monte-Carlo simulations.

Third, note that the decomposition

$$\check{L}_B^\Sigma + 1 \approx \underbrace{\left(\tilde{L}_B^\Sigma + 1\right)}_{\text{Due to the theoretical loss}} \times \underbrace{\text{Var}_M\left[Z_n^\Sigma\right]}_{\text{Due to the finite sample}}, \tag{5.1}$$

Table 5
Finite sample properties of $\tilde{Z}_n^{\tilde{Q}_B}$ (Model 2).[a]

No. Obs.	Mean	Stdv.	RMSE	0.5%	2.5%	5%	95%	97.5%	99.5%
$B = 1$ block									
5,850	−0.114	1.200	1.205	0.01	0.50	1.29	95.50	97.99	99.59
11,700	−0.090	1.148	1.152	0.01	0.35	1.68	95.36	97.18	99.12
23,400	−0.075	1.109	1.112	0.01	0.62	2.00	95.13	96.91	98.18
46,800	−0.062	1.093	1.095	0.01	0.61	2.98	94.38	96.67	99.08
$B = 2$ blocks									
5,850	−0.099	1.170	1.174	0.02	0.56	1.49	95.70	97.66	99.38
11,700	−0.080	1.130	1.133	0.01	0.49	1.76	94.89	96.81	99.05
23,400	−0.057	1.094	1.095	0.03	0.91	2.38	95.25	97.32	98.61
46,800	−0.049	1.079	1.081	0.02	0.99	3.61	94.93	97.24	99.40
$B = 4$ blocks									
5,850	−0.089	1.150	1.154	0.04	0.82	1.62	95.56	97.50	99.18
11,700	−0.077	1.112	1.114	0.05	0.56	1.92	94.80	96.73	99.17
23,400	−0.049	1.083	1.084	0.06	1.11	2.91	95.15	97.36	98.77
46,800	−0.046	1.083	1.084	0.02	1.16	3.38	95.23	97.32	99.53
$B = 6$ blocks									
5,850	−0.090	1.146	1.149	0.07	0.91	1.65	95.89	97.47	99.00
11,700	−0.076	1.105	1.107	0.05	0.58	2.36	94.70	96.98	99.21
23,400	−0.050	1.077	1.078	0.06	1.07	3.10	95.10	97.62	98.78
46,800	−0.046	1.076	1.077	0.03	1.19	3.69	95.35	97.39	99.44
$B = 8$ blocks									
5,850	−0.090	1.145	1.148	0.08	0.78	1.96	95.71	97.39	99.27
11,700	−0.073	1.099	1.101	0.06	0.68	2.46	94.83	96.81	99.11
23,400	−0.045	1.076	1.077	0.08	1.17	2.51	95.22	97.50	98.96
46,800	−0.046	1.080	1.081	0.03	1.46	3.86	95.39	97.26	99.46

[a] This table shows summary statistics and empirical quantiles benchmarked to the $N(0, 1)$ distribution for the feasible Z-statistics related to the global and local QMLE. The simulation design is Model 2 with $M = 10{,}000$ Monte-Carlo simulations.

where $\mathrm{Var}_M[X]$ denotes the sample variance of X based on the M Monte Carlo simulations, is numerically well-verified and gives an intuitive interpretation of the main sources of deviation from the bound in practice. For instance, consider $\tilde{Q}_2$ on Model 2, with $n = 23{,}400$, $\xi^2 = 0.001$. In that case, the previous decomposition (5.1) gives $0.218 + 1 = 1.218$ for the left hand side, and $(0.124 + 1) \times 1.048^2 \approx 1.220$ for the right hand side which is very close to the other value indeed.

Finally, this simulation study indicates that the local version of RK and the QMLE perform very well in practice, with the QMLE slightly more robust to the values of n and B as free of tuning parameters.

6. Empirical illustration

We conclude this study by the application of our method on transaction log prices of Intel Corporation (INTC) shares recorded on the NASDAQ stock market over the year 2015. We exclude January 1, the day after Thanksgiving and December 24 which are less active, thus this leaves us with 250 trading days of data. Moreover, we only keep transactions that were carried out between 9:30 am and 4 pm. Finally, we consider the data in tick time, for an average of 6139 daily trades. The most active days include more than 15,000 trades.

We first estimate the theoretical gain in AVAR. As for the numerical study, we do not cap $\widehat{\rho}_{T_{i-1}, T_i}$ by 1. Across the days, values of B and blocks corresponding to an overall of 5250 estimates, the value 1.1 was crossed only a few times. We report in Table 7 key statistics for AVAR reduction. We get a global estimate of ρ around 0.74, which is very close to $\rho_{regular}$. Across the year the estimates of ρ ranged from around 0.3 to 1, and actually crossed 1 for two days where it reached 1.03 and 1.04. When B increases, we find as expected that $\widehat{\rho}_B$, the mean estimated value of ρ across days and blocks, also increases to reach a value of 0.86 for 8 blocks. Accordingly, the mean estimated ratios of AVAR decreases from 1 to 0.9 for the QMLE, and from 1 to 0.92 for the RK. Moreover, we find that those ratios are consistently smaller than 1 for the 250 days and different values of B bigger than 1, so that the local method never deteriorates the AVAR of the estimator. Note that the same ratios for Model 2 in our simulation study range from 1 to $102.4/121.5 = 0.84$ for the QMLE and from 1 to $105.6/118.2 = 0.89$ for the RK. The slight disparity between the empirical study and Model 2 can be explained in several ways. For example, it is likely that we still under-evaluate the difference between ρ and ρ_{T_{i-1}, T_i}, or that the theoretical model is a little too optimistic about how fast ρ_{T_{i-1}, T_i} gets close to 1 on local blocks. To sum up, the results are approximately in line with what was expected, and present a substantial gain in terms of AVAR for both the QMLE and the RK.

The last column in Table 7 shows the empirical correlation between the correction terms $\tilde{Q}_B - Q$ and $\tilde{K}_B - K$ for several values of B. The positive correlation indicates that the local method tends to correct the global estimates in the same direction for both the QMLE and the RK. Moreover, increasing the number of blocks B amplifies the phenomenon. Table 8 shows the empirical mean and standard deviation of the 10 estimators. Note that the main source of randomness being the target value itself, it is not surprising to find the mean and standard values very close to each other. We have reported in the last column

Table 6
Losses.[a]

Model		Q	$\tilde{Q}_2$	$\tilde{Q}_4$	$\tilde{Q}_6$	$\tilde{Q}_8$	K	$\tilde{K}_2$	$\tilde{K}_4$	$\tilde{K}_6$	$\tilde{K}_8$
$n = 23{,}400, \xi^2 = 0.01$											
Model 1	Emp.	8.4%	6.7%	5.3%	3.1%	4.1%	14.3%	12.4%	13.9%	14.6%	18.0%
	Theo.	6.9%	5.4%	2.6%	1.4%	0.9%	9.9%	8.4%	6.0%	5.0%	4.5%
Model 2	Emp.	29.3%	18.4%	12.8%	10.7%	8.6%	30.2%	23.4%	20.3%	17.3%	22.5%
	Theo.	21.5%	12.4%	5.8%	3.5%	2.4%	18.2%	12.1%	8.0%	6.4%	5.6%
Model 3	Emp.	40.0%	21.0%	9.4%	6.4%	5.1%	29.3%	20.3%	14.0%	12.0%	13.3%
	Theo.	38.7%	20.9%	9.0%	5.0%	3.2%	26.8%	17.0%	10.3%	7.6%	6.3%
$n = 46{,}800, \xi^2 = 0.01$											
Model 1	Emp.	8.1%	5.8%	3.5%	3.0%	2.5%	11.6%	9.4%	9.9%	9.2%	12.5%
	Theo.	6.9%	5.4%	2.6%	1.4%	0.9%	9.9%	8.4%	6.0%	5.0%	4.5%
Model 2	Emp.	25.5%	15.5%	8.7%	5.9%	5.7%	22.5%	17.2%	15.3%	15.8%	15.5%
	Theo.	21.5%	12.4%	5.8%	3.5%	2.4%	18.2%	12.1%	8.0%	6.4%	5.6%
Model 3	Emp.	38.1%	20.0%	8.3%	2.3%	1.7%	29.6%	18.3%	11.7%	10.3%	11.2%
	Theo.	38.7%	20.9%	9.0%	5.0%	3.2%	26.8%	17.0%	10.3%	7.6%	6.3%
$n = 23{,}400, \xi^2 = 0.001$											
Model 1	Emp.	17.2%	15.8%	12.9%	11.7%	11.2%	21.7%	20.9%	19.4%	19.7%	20.1%
	Theo.	6.9%	5.4%	2.6%	1.4%	0.9%	9.9%	8.4%	6.0%	5.0%	4.5%
Model 2	Emp.	30.7%	21.8%	14.8%	12.1%	11.2%	28.7%	23.3%	19.7%	19.2%	19.0%
	Theo.	21.5%	12.4%	5.8%	3.5%	2.4%	18.2%	12.1%	8.0%	6.4%	5.6%
Model 3	Emp.	51.1%	33.0%	20.6%	16.2%	14.7%	43.4%	32.8%	26.1%	23.8%	23.5%
	Theo.	38.7%	20.9%	9.0%	5.0%	3.2%	26.8%	17.0%	10.3%	7.6%	6.3%
$n = 46{,}800, \xi^2 = 0.001$											
Model 1	Emp.	15.3%	13.9%	10.9%	9.6%	9.1%	20.0%	19.0%	16.7%	16.3%	16.4%
	Theo.	6.9%	5.4%	2.6%	1.4%	0.9%	9.9%	8.4%	6.0%	5.0%	4.5%
Model 2	Emp.	29.7%	20.6%	13.3%	10.4%	9.2%	28.2%	22.4%	17.8%	16.5%	15.8%
	Theo.	21.5%	12.4%	5.8%	3.5%	2.4%	18.2%	12.1%	8.0%	6.4%	5.6%
Model 3	Emp.	47.6%	29.2%	16.9%	12.6%	11.0%	38.3%	26.8%	20.6%	17.6%	17.2%
	Theo.	38.7%	20.9%	9.0%	5.0%	3.2%	26.8%	17.0%	10.3%	7.6%	6.3%
$n = 23{,}400, \xi^2 = 0.0002$											
Model 1	Emp.	25.2%	23.8%	20.6%	19.4%	18.6%	32.8%	31.7%	30.3%	30.0%	30.5%
	Theo.	6.9%	5.4%	2.6%	1.4%	0.9%	9.9%	8.4%	6.0%	5.0%	4.5%
Model 2	Emp.	45.5%	35.6%	28.2%	25.7%	24.5%	46.2%	40.5%	37.0%	36.4%	36.8%
	Theo.	21.8%	12.6%	5.9%	3.6%	2.5%	18.4%	12.2%	8.0%	6.4%	5.6%
Model 3	Emp.	64.3%	45.3%	32.5%	28.4%	26.2%	56.0%	46.4%	40.8%	39.5%	39.4%
	Theo.	38.1%	20.5%	8.8%	4.9%	3.1%	26.6%	16.9%	10.2%	7.5%	6.2%
$n = 46{,}800, \xi^2 = 0.0002$											
Model 1	Emp.	19.7%	18.1%	14.8%	13.6%	12.7%	24.9%	23.6%	21.4%	21.0%	20.7%
	Theo.	6.9%	5.4%	2.6%	1.4%	0.9%	9.9%	8.5%	6.0%	5.0%	4.5%
Model 2	Emp.	38.9%	29.0%	22.2%	19.8%	18.1%	37.7%	31.8%	28.5%	27.6%	27.0%
	Theo.	21.8%	12.6%	5.9%	3.6%	2.5%	18.4%	12.2%	8.0%	6.4%	5.6%
Model 3	Emp.	57.3%	38.3%	26.3%	22.1%	19.6%	47.3%	37.1%	31.4%	29.5%	28.1%
	Theo.	38.1%	20.5%	8.8%	4.9%	3.1%	26.6%	16.9%	10.2%	7.5%	6.2%

[a] Empirical losses $\breve{L}_B^{(\Sigma)}$ and theoretical losses $\tilde{L}_B^{(\Sigma)}$ for the three models and the 10 estimators. Two levels of sampling $n = 23{,}400$, $n = 46{,}800$ and three noise-to-signal ratios $\xi^2 = 0.01$, $\xi^2 = 0.001$ and $\xi^2 = 0.0002$ are considered.

Table 7
Estimates of ρ, AVAR ratio estimates and empirical correlation of corrections.[a]

B	$\widehat{\rho}_B$	$AVAR_B^{(QMLE)}/AVAR_1^{(QMLE)}$	$AVAR_B^{(RK)}/AVAR_1^{(RK)}$	$\widehat{Corr}(\tilde{Q}_B - Q, \tilde{K}_B - K)$
1	0.74	1	1	–
2	0.8	0.96	0.97	0.689
4	0.84	0.92	0.94	0.769
6	0.85	0.91	0.93	0.868
8	0.86	0.9	0.92	0.879

[a] For $B = 1, 2, 4, 6, 8$, $\widehat{\rho}_B$ refers to the empirical mean value of estimates of ρ on blocks $[T_{i-1}, T_i]$ across days and values of i for INTC in 2015. The AVAR ratios are estimated by plugging estimates of the integrated volatility, the integrated quarticity and ρ on blocks of different sizes. The last column shows the empirical correlation between the corrections induced by the local method.

Table 8
Summary statistics for the global and local estimators.[a]

Estimator	Mean	Stdv.	$\widehat{\text{Corr}}(., Q)$
Q	1.771	1.789	1
$\tilde{Q}_2$	1.770	1.781	≈ 1
$\tilde{Q}_4$	1.766	1.769	≈ 1
$\tilde{Q}_6$	1.761	1.762	0.9999
$\tilde{Q}_8$	1.757	1.753	0.9999
K	1.818	1.795	0.9994
$\tilde{K}_2$	1.818	1.780	0.9992
$\tilde{K}_4$	1.813	1.770	0.9991
$\tilde{K}_6$	1.808	1.756	0.9989
$\tilde{K}_8$	1.804	1.751	0.9988

[a] Sample means, standard deviations, and correlations with the global QMLE
for the 10 estimators implemented for INTC data in 2015. The estimators are
scaled by a factor 10^4.
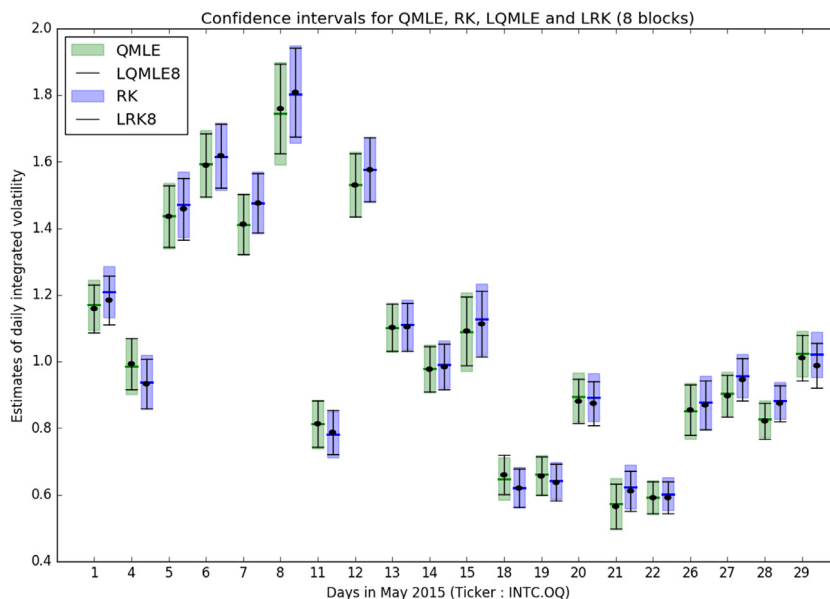
Fig. 3. 95% Confidence intervals for the four estimators Q and $\tilde{Q}_8$ (green, left), K and $\tilde{K}_8$ (blue, right) on INTC data in May 2015. The CIs are computed using the estimates of $AVAR_B^{(QMLE)}$, and $AVAR_B^{(RK)}$ for $B = 1, 8$ obtained as explained in Section 6. The estimators are scaled by a factor 10^4. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

the correlation between each estimator and the global QMLE. We find results very close to 1 for all estimators. One should note that the global RK is less correlated to the QMLE than all the local QMLE $\tilde{Q}_B$. This indicates that the order of magnitude of the correction induced by the local method is smaller than the difference between the two global estimators.

Finally, Fig. 3 shows daily 95% theoretical confidence intervals for Q, $\tilde{Q}_8$, K and $\tilde{K}_8$ in May 2015. We can see that the confidence intervals for the local estimators are often shorter than their counterpart. Moreover, over the year the global and the local estimates confidence intervals always overlap, corroborating the fact that the local estimates are in line with their global versions.

7. Conclusion

In this paper, we have looked at the efficiency of local methods to estimate integrated volatility. We have shown that for the RK and the QMLE, if we chop the data into B blocks we can reduce the AVAR when B is fixed and retrieve the parametric loss when B goes to infinity. We have also seen that the theoretical gain is mostly preserved when looking at finite sample results. Finally, we have documented that the gain is substantial in practice.

Given how simple to implement the methodology is, we expect that it will be very helpful for practitioners. Our hope is that this simple and natural technique will be used on the QMLE and the RK, but also considered for a wider class of estimators. It is clear that the theory would work for the PAE and the MSRV, but econometricians should also try it on their

own favorite estimator. Actually, the technique can be applied to other problems, such as the high-frequency covariance estimation, the estimation of functions of volatility, the leverage effect, the volatility of volatility, etc.

Appendix. proofs

A.1. Simplification of the problem

Since we want to prove stable convergence, in view of the componentwise local boundedness of the matrix

$$\begin{pmatrix} \sigma_t & 0 \\ \tilde{\sigma}_t^{(1)} & \tilde{\sigma}_t^{(2)} \end{pmatrix},$$

and because $\inf_t(\min(\sigma_t, \tilde{\sigma}_t^{(2)})) > 0$, we can without loss of generality assume that for all $t \in [0, T]$ there exists some nonrandom constants $\underline{\sigma}$ and $\overline{\sigma}$ such that

$$0 < \underline{\sigma} < \sigma_t, \tilde{\sigma}_t^{(1)}, \tilde{\sigma}_t^{(2)} < \overline{\sigma}, \tag{A.1}$$

by using a standard localization argument (e.g., Section 2.4.5 of Mykland and Zhang, 2012). One can further suppress b_t as in Section 2.2 (pp. 1407–1409) of Mykland and Zhang (2009), and act as if X_t is a martingale. Also, we follow a similar procedure to localize the random variables U_i^n as, e.g, in the proof of Lemma 14.1.5 p. 435, Equation (14.1.13), in Jacod and Protter (2011). Consequently, we will assume in the following of the proof:

(H) We have $b = \tilde{b} = 0$. Moreover $\sigma, \sigma^{-1}, \tilde{\sigma}^{(1)}, (\tilde{\sigma}^{(1)})^{-1}, \tilde{\sigma}^{(2)}, (\tilde{\sigma}^{(2)})^{-1}, \alpha, \alpha^{-1}$ are bounded. Given an *a priori* number $\gamma > 0$, we also have $\sup_{0 \le i \le N_n} U_i^n \le n^\gamma$.

In particular, **(H)** implies, taking γ small enough, that $\pi_T^n < 1$, for $n \in \mathbb{N}$ large enough.

We define $\mathcal{U} := \sigma\left\{U_i^n | i, n \in \mathbb{N}\right\} \vee \sigma\{\alpha_s | 0 \le s \le T\}$ the σ-field that generates the observation times and which is independent of X. We will often have to use the conditional expectation $\mathbb{E}[.|\mathcal{U}]$, that we hereafter denote for convenience $\mathbb{E}_{\mathcal{U}}$. We also define the discrete filtration $\mathcal{G}_i^n := \mathcal{F}_{t_i^n}^X \vee \mathcal{U}$, and recall the continuous version $\mathcal{G}_t := \mathcal{F}_t^X \vee \mathcal{U}$ where $\mathcal{F}_t^X$ is the canonical filtration associated to X. Note that by independence from α, X admits the same Itô semimartingale dynamics in the extension $\mathcal{G}$.

Note also that, by virtue of Lemma 14.1.5 in Jacod and Protter (2011), recalling $\pi_t^n := \sup_{i \ge 1} t_i^n - t_{i-1}^n$, and $N_n(t) = \sup\{i \in \mathbb{N} - \{0\} | t_i^n \le t\}$ we have

$$\eta > 0 \implies n^{1-\eta}\pi_t^n \xrightarrow{\mathbb{P}} 0. \tag{A.2}$$

Throughout the proofs, we write N_n for $N_n(T)$. We also define $L_n = N_n^{1/2+\delta}$, for some $\delta > 0$ to be adjusted, and we let L be a positive constant that may vary from one line to the other. Finally we often refer to the continuous part of X_t defined as

$$\tilde{X}_t := X_0 + \int_0^t b_s ds + \int_0^t \sigma_s dW_s. \tag{A.3}$$

A.2. Proof of (2.1)

We first show the left hand side inequality, that can be reformulated as $\kappa_{r,s}^{2/3} \ge \rho_{r,s}$. Note that by an immediate application of Hölder's inequality we have

$$\int_r^s \sigma_u^2 du \le (s-r)^{1/3} \left(\int_r^s \sigma_u^3 du\right)^{2/3}.$$

Thus,

$$\begin{aligned} \rho_{r,s} &= \frac{\int_r^s \sigma_u^2 du}{(s-r)^{1/2}\left(\int_r^s \sigma_u^4 du\right)^{1/2}} \\ &\le \frac{\left(\int_r^s \sigma_u^3 du\right)^{2/3}}{(s-r)^{1/6}\left(\int_r^s \sigma_u^4 du\right)^{1/2}} = \kappa_{r,s}^{2/3}. \end{aligned}$$

For the right hand side inequality, we first consider the domination

$$\int_r^s \sigma_u^3 du \le \left(\int_r^s \sigma_u^2 du\right)^{1/2} \left(\int_r^s \sigma_u^4 du\right)^{1/2},$$

which is obtained by Cauchy–Schwarz inequality. Then we inject this expression in $\kappa_{r,s}$ and we get

$$
\kappa_{r,s} = \frac{\int_r^s \sigma_u^3 du}{(s-r)^{1/4} \left(\int_r^s \sigma_u^4 du\right)^{3/4}}
$$

$$
\leq \frac{\left(\int_r^s \sigma_u^2 du\right)^{1/2}}{(s-r)^{1/4} \left(\int_r^s \sigma_u^4 du\right)^{1/4}} = \rho_{r,s}^{1/2}.
$$

A.3. Estimates for the efficient price X

Hereafter, we adopt the following notation convention. For a process V (including the noise process ϵ by a slight "abuse of notation"), and $t \in [0, T]$ we write $\Delta V_t = V_t - V_{t-}$, $\Delta V_i^n := V_{t_i^n} - V_{t_{i-1}^n}$ and $\Delta V^n := (\Delta V_1^n, \dots, \Delta V_{N_n}^n)$. Finally, for interpolation purpose we sometimes write the continuous version $\Delta V_{i,t}^n := V_{t_i^n \wedge t} - V_{t_{i-1}^n \wedge t}$, along with the time increment $\Delta t_{i,t}^n := t_i^n \wedge t - t_{i-1}^n \wedge t$. We introduce the two following quantities:

$$
\zeta_{i,t}^n := (\Delta \tilde{X}_{i,t}^n)^2 - \sigma_{t_{i-1}^n}^2 \Delta t_{i,t}^n, \text{ and } \bar{\zeta}_{i,t}^n := \mathbb{E}\left[\zeta_{i,t}^n | \mathcal{G}_{i-1}^n\right]. \tag{A.4}
$$

We have the following estimates.

Lemma A.1. *We have, for some constant $L > 0$ independent of i,*

$$
\mathbb{E}\left[\sup_{t \in [t_{i-1}^n, t_i^n]} |\Delta \tilde{X}_{i,t}^n|^p \Big| \mathcal{G}_{i-1}^n\right] \leq L n^{-p/2} (U_i^n)^{p/2}, \tag{A.5}
$$

$$
\left|\bar{\zeta}_{i,t}^n\right| \leq L n^{-3/2} (U_i^n)^{3/2}, \tag{A.6}
$$

$$
\mathbb{E}\left[\left(\zeta_{t,i}^n\right)^p \Big| \mathcal{G}_{i-1}^n\right] \leq L n^{-p} (U_i^n)^p, \tag{A.7}
$$

$$
\mathbb{E}\left[\left|\int_{t_{i-1}^n \wedge t}^{t_i^n \wedge t} \sigma_s^2 ds - \sigma_{t_{i-1}^n}^2 \Delta t_{i,t}^n\right|^p \Big| \mathcal{G}_{i-1}^n\right] \leq L n^{-3p/2} (U_i^n)^{3p/2}. \tag{A.8}
$$

Proof. For (A.5), this is a consequence of the fact that by the conditional Burkholder–Davis–Gundy inequality, we have

$$
\mathbb{E}\left[\sup_{t \in [t_{i-1}^n, t_i^n]} \left|\int_{t_{i-1}^n \wedge t}^{t_i^n \wedge t} \sigma_s dW_s\right|^p \Big| \mathcal{G}_{i-1}^n\right] \leq \mathbb{E}\left[\sup_{t \in [t_{i-1}^n, t_i^n]} \left|\int_{t_{i-1}^n \wedge t}^{t_i^n \wedge t} \sigma_s^2 ds\right|^{p/2} \Big| \mathcal{G}_{i-1}^n\right]
$$

$$
\leq L \left(t_i^n - t_{i-1}^n\right)^{p/2}.
$$

Since α is bounded by assumption **(H)**, and since $t_i^n - t_{i-1}^n < 1$, we get (A.5). The other estimates are straightforwardly obtained using the same line of reasoning and Itô formula. □

A.4. Proof of Theorems 4 and 8

We adopt the general setting introduced in Section 4 and Appendix A.1. We start by showing the consistency of the QMLE along with other estimates in the case $B = 1$. We then adapt and combine those results in the case $B \geq 1$ to derive the central limit theorem stated in Theorem 8. As a byproduct, Theorem 4 will also be proven.

When $B = 1$, we recall that for any $\xi = (\sigma^2, a^2) \in \Xi := [\underline{\Sigma}, \overline{\Sigma}] \times [\underline{a}^2, \overline{a}^2]$, $\underline{a}^2 > 0$, we have, up to a constant term

$$
l_n(\xi) = -\frac{1}{2}\log \det(\Omega) - \frac{1}{2}Y^T \Omega^{-1} Y, \tag{A.9}
$$

with $\Omega^{-1} = \left[\omega^{i,j}\right]_{1 \leq i \leq N_n, 1 \leq j \leq N_n}$. The exact definition of the coefficients $\omega^{i,j}$ can be found in e.g. (28), p. 245 of Xiu (2010), replacing n by N_n. We define the approximate log-likelihood random field as

$$
\bar{l}_n(\xi) = -\frac{1}{2}\log \det(\Omega) - \frac{1}{2}\text{Tr}\left(\Omega^{-1} \left\{\Sigma_0^c + \Sigma_0^d\right\}\right), \tag{A.10}
$$

with

$$
\Sigma_0^c = \begin{pmatrix}
\int_0^{t_1^n} \sigma_s^2 ds + 2a^2 & -a^2 & 0 & \cdots & 0 \\
-a^2 & \int_{t_1^n}^{t_2^n} \sigma_s^2 ds + 2a^2 & -a^2 & \ddots & \vdots \\
0 & -a^2 & \int_{t_2^n}^{t_3^n} \sigma_s^2 ds + 2a^2 & \ddots & 0 \\
\vdots & \ddots & \ddots & \ddots & -a^2 \\
0 & \cdots & 0 & -a^2 & \int_{t_{N_n-1}^n}^{t_{N_n}^n} \sigma_s^2 ds + 2a^2
\end{pmatrix},
$$

and

$$
\Sigma_0^d = \mathrm{diag}\left(\sum_{0 < s \le t_1^n} \Delta J_s^2, \sum_{t_1^n < s \le t_2^n} \Delta J_s^2, \ldots, \sum_{t_{N_n-1}^n < s \le t_{N_n}^n} \Delta J_s^2 \right).
$$

We further define the diagonal scaling matrix

$$
\Phi_n = \mathrm{diag}(N_n^{1/2}, N_n),
$$

and consider for $\xi \in \Xi$ the scaled score functions

$$
\Psi_n(\xi) = -\Phi_n^{-1} \frac{\partial l_n(\xi)}{\partial \xi} \quad \text{and} \quad \bar{\Psi}_n = -\Phi_n^{-1} \frac{\partial \bar{l}_n(\xi)}{\partial \xi}.
$$

We start by showing the consistency of the QMLE. Before stating the result, we give a few definitions. For a matrix $A = \left[a_{i,j} \right]_{1 \le i \le N_n, 1 \le j \le N_n} \in \mathbb{R}^{N_n \times N_n}$, we associate the matrix $\dot{A} = [\dot{a}_{i,j}]_{0 \le i \le N_n, 1 \le j \le N_n} \in \mathbb{R}^{(N_n+1) \times N_n}$ and $\ddot{A} = [\ddot{a}_{i,j}]_{0 \le i N_n, 0 \le j \le N_n} \in \mathbb{R}^{(N_n+1) \times (N_n+1)}$ whose components respectively satisfy

$$
\dot{a}_{i,j} = a_{i+1,j} - a_{i,j},
$$

and

$$
\ddot{a}_{i,j} = \dot{a}_{i,j+1} - \dot{a}_{i,j} = a_{i+1,j+1} - a_{i,j+1} + a_{i,j} - a_{i+1,j},
$$

with the convention $a_{i,j} = 0$ when $i = 0$ or $j = 0$. This will be useful to disentangle some quadratic expressions using the following result.

Lemma A.2. *Let* $y, z \in \mathbb{R}^{N_n+1}$, *with* $y = (y_0, \ldots, y_{N_n})^T$, $z = (z_0, \ldots, z_{N_n})^T$. *We define* $\Delta y = (\Delta y_1, \ldots, \Delta y_{N_n}) := \left(y_1 - y_0, \ldots, y_{N_n} - y_{N_n-1} \right)^T \in \mathbb{R}^{N_n}$, *and* Δz *the same way. Then we have the by-part summation identities*

$$
\Delta y^T A \Delta z = -y^T \dot{A} \Delta z = y^T \ddot{A} z.
$$

We now show a preliminary lemma to get the consistency of the QMLE.

Lemma A.3 (*Asymptotic Score*)**.** *For any* $\xi \in \Xi$, *let*

$$
\Psi_\infty(\xi) = \begin{pmatrix}
-\dfrac{1}{8a\sigma^3\sqrt{T}} \left(\int_0^T \sigma_s^2 ds + \sum_{0 < s \le T} \Delta J_s^2 - \sigma^2 T \right) - \dfrac{\sqrt{T}}{8a^3\sigma} \left(a^2 - a_0^2 \right) \\
\dfrac{1}{2a^4} \left(a^2 - a_0^2 \right)
\end{pmatrix}.
$$

We have

$$
\sup_{\xi \in \Xi} |\Psi_n(\xi) - \Psi_\infty(\xi)| \xrightarrow{\mathbb{P}} 0. \tag{A.11}
$$

Proof. We start by treating the case where the jump part $J = 0$. We have the decomposition

$$
\Psi_n = \bar{\Psi}_n + R_n, \tag{A.12}
$$

with

$$
R_n(\xi) = \begin{pmatrix} \dfrac{1}{2\sqrt{N_n}} \left\{ Y^T \dfrac{\partial \Omega^{-1}}{\partial \sigma^2} Y - \mathrm{tr}\left(\dfrac{\partial \Omega^{-1}}{\partial \sigma^2} \left\{ \Sigma_0^c + \Sigma_0^d \right\} \right) \right\} \\[2mm] \dfrac{1}{2N_n} \left\{ Y^T \dfrac{\partial \Omega^{-1}}{\partial a^2} Y - \mathrm{tr}\left(\dfrac{\partial \Omega^{-1}}{\partial a^2} \left\{ \Sigma_0^c + \Sigma_0^d \right\} \right) \right\} \end{pmatrix}.
$$

By a straightforward adaptation of the proof of Lemma 1–2 and Theorem 4 in Xiu (2010), we have immediately that $R_n = o_{\mathbb{P}}(1)$ uniformly in the parameters since the step size of the observation grid $\pi_T^n \xrightarrow{\mathbb{P}} 0$ by (A.2). Thus it is sufficient to show that we have

$$
\sup_{\xi \in \Xi} \left| \bar{\Psi}_n(\xi) - \Psi_\infty(\xi) \right| \xrightarrow{\mathbb{P}} 0.
$$

Equality (A.11) is then a direct consequence of equations (38) and (40) pp. 247–248 in Xiu (2010) that are obtained following exactly the same proof as pp. 247–248 for an irregular grid such that $\pi_T^n \xrightarrow{\mathbb{P}} 0$.

When there are jumps, there is an additional term in (A.12) which is equal to

$$
A_n(\xi) = \begin{pmatrix} \dfrac{1}{2\sqrt{N_n}} \left\{ (\Delta J^n)^T \dfrac{\partial \Omega^{-1}}{\partial \sigma^2} \Delta J^n + 2(\Delta J^n)^T \dfrac{\partial \Omega^{-1}}{\partial \sigma^2} \left\{ \Delta \tilde{X}^n + \Delta \epsilon^n \right\} \right\} \\[2mm] \dfrac{1}{2N_n} \left\{ (\Delta J^n)^T \dfrac{\partial \Omega^{-1}}{\partial a^2} \Delta J^n + 2(\Delta J^n)^T \dfrac{\partial \Omega^{-1}}{\partial a^2} \left\{ \Delta \tilde{X}^n + \Delta \epsilon^n \right\} \right\} \end{pmatrix}, \tag{A.13}
$$

so that it is sufficient to show that we have

$$
\sup_{\xi \in \Xi} \left| A_n(\xi) + \begin{pmatrix} \dfrac{1}{8a\sigma^3\sqrt{T}} \sum_{0 < s \le T} \Delta J_s^2 \\ 0 \end{pmatrix} \right| \xrightarrow{\mathbb{P}} 0. \tag{A.14}
$$

We first compute the limit of the term $\frac{1}{2\sqrt{N_n}}(\Delta J^n)^T \frac{\partial \Omega^{-1}}{\partial \sigma^2} \Delta J^n$. Recalling that $\omega^{i,j}$ is the (i,j)th index of Ω^{-1}, we provide the following decomposition:

$$
\frac{1}{2\sqrt{N_n}}(\Delta J^n)^T \frac{\partial \Omega^{-1}}{\partial \sigma^2} \Delta J^n = \frac{1}{2\sqrt{N_n}} \sum_{i=1}^{N_n} \frac{\partial \omega^{i,i}}{\partial \sigma^2} (\Delta J_i^n)^2 + \frac{1}{\sqrt{N_n}} \sum_{1 \le j < i \le N_n} \frac{\partial \omega^{i,j}}{\partial \sigma^2} \Delta J_i^n \Delta J_j^n. \tag{A.15}
$$

Now, we define $\tau_1, \ldots, \tau_{N^J}$ the jump times of J, where N^J is the random number of jumps of J on $[0, T]$. Since N^J is finite, there exists a random number K^J such that for $n \ge K^J$ we have

$$
\frac{1}{\sqrt{N_n}} \sum_{i=1}^{N_n} \frac{\partial \omega^{i,i}}{\partial \sigma^2} (\Delta J_i^n)^2 = \frac{1}{2\sqrt{N_n}} \sum_{k=1}^{N^J} \frac{\partial \omega^{N_n(\tau_k), N_n(\tau_k)}}{\partial \sigma^2} \Delta J_{\tau_k}^2. \tag{A.16}
$$

By direct calculation from the expression of the coefficients of Ω^{-1} in (28) p. 245 in Xiu (2010), we easily deduce that for each k we have $\frac{1}{2\sqrt{N_n}} \frac{\partial \omega^{N_n(\tau_k), N_n(\tau_k)}}{\partial \sigma^2} \xrightarrow{\mathbb{P}} -\frac{1}{8a\sigma^3\sqrt{T}}$ uniformly in $\xi \in \Xi$. Since the sum is finite, this yields the uniform convergence

$$
\sup_{\xi \in \Xi} \left| \frac{1}{2\sqrt{N_n}}(\Delta J^n)^T \frac{\partial \Omega^{-1}}{\partial \sigma^2} \Delta J^n + \frac{1}{8a\sigma^3\sqrt{T}} \sum_{0 < s \le T} \Delta J_s^2 \right| \xrightarrow{\mathbb{P}} 0. \tag{A.17}
$$

By a similar argument, we also have for $k \ne l$ that $\frac{\partial \omega^{N_n(\tau_k), N_n(\tau_l)}}{\partial \sigma^2} \xrightarrow{\mathbb{P}} 0$ exponentially so that we have $\frac{1}{\sqrt{N_n}} \sum_{1 \le j < i \le N_n} \frac{\partial \omega^{i,j}}{\partial \sigma^2} \Delta J_i^n \Delta J_j^n \xrightarrow{\mathbb{P}} 0$ uniformly. As for $\frac{1}{\sqrt{N_n}}(\Delta J^n)^T \frac{\partial \Omega^{-1}}{\partial \sigma^2} \left\{ \Delta \tilde{X}^n + \Delta \epsilon^n \right\}$, on the one hand the same computation yields that the leading term of $\frac{1}{\sqrt{N_n}}(\Delta J^n)^T \frac{\partial \Omega^{-1}}{\partial \sigma^2} \Delta \tilde{X}^n$ is

$$
\sum_{k=1}^{N^J} \underbrace{\frac{1}{\sqrt{N_n}} \frac{\partial \omega^{N_n(\tau_k), N_n(\tau_k)}}{\partial \sigma^2} \Delta J_{\tau_k}}_{o_{\mathbb{P}}(1)} \underbrace{\Delta \tilde{X}_{t_{i_k}^n}^n}_{o_{\mathbb{P}}(1)},
$$

where $t_{i_k}^n \leq \tau_k \leq t_{i_k+1}^n$, so that as the sum is finite the expression is negligible. On the other hand, we also have by Lemma A.2 that the leading term of the noise part is

$$-\sum_{k=1}^{N^J} \underbrace{\frac{1}{\sqrt{N_n}} \frac{\partial \dot{\omega}^{N_n(\tau_k),N_n(\tau_k)}}{\partial \sigma^2}}_{O_{\mathbb{P}}(N_n^{-1/2})} \Delta J_{\tau_k} \underbrace{\epsilon_{t_{i_k}^n}}_{O_{\mathbb{P}}(1)}$$

since $\frac{\partial \dot{\omega}^{i,j}}{\partial \sigma^2} = O_{\mathbb{P}}(1)$ by direct calculation. Finally, similar reasoning shows that the second component of A_n is negligible because of the scaling in N_n^{-1} instead of $N_n^{-1/2}$, and we are done. □

Now we turn to the consistency of the QMLE.

Theorem A.4 (*Consistency*)**.** *If $\widehat{\xi}_n = (\widehat{\sigma}_n^2, \widehat{a}_n^2)$ is the QMLE, we have*

$$\widehat{\xi}_n \xrightarrow{\mathbb{P}} \xi_0 := \left(\overline{\sigma}_0^2, a_0^2\right), \tag{A.18}$$

where we recall that $T\overline{\sigma}_0^2 = \int_0^T \sigma_s^2 ds + \sum_{0<s\leq T} \Delta J_s^2$.

Proof. To get (A.18), it is sufficient to have

$$\sup_{\xi \in \Xi} |\Psi_n(\xi) - \Psi_\infty(\xi)| \xrightarrow{\mathbb{P}} 0, \tag{A.19}$$

which has been proven in Lemma A.3, and for any $\epsilon > 0$

$$\inf_{\xi \in \Xi: \|\xi - \xi_0\| \geq \epsilon} \|\Psi_\infty(\xi)\|^2 > 0 = \|\Psi_\infty(\xi_0)\|^2 \; \mathbb{P}\text{-a.s}, \tag{A.20}$$

by a classical statistical argument (see e.g. Van der Vaart, 2000, Theorem 5.9). Given the form of Ψ_∞, the equality $\Psi_\infty(\xi_0) = 0$ is immediate. Note also that the left hand side inequality of (A.20) will be automatically satisfied if we show that $\|\Psi_\infty(\xi)\|^2 > 0$ as soon as $\xi \neq \xi_0$ by a continuity argument since Ξ is compact. Let us then take $\xi \in \Xi - \{\xi_0\}$ such that $\Psi_\infty(\xi) = 0$, and assume first that $a^2 \neq a_0^2$. In that case, we have

$$0 = \|\Psi_\infty(\xi)\|^2 \geq \frac{1}{4a^8} \left(a^2 - a_0^2\right)^2,$$

which leads to a contradiction. Similarly, the first component of Ψ_∞ leads to the domination

$$0 = \|\Psi_\infty(\sigma^2, a_0^2)\|^2 \geq \frac{T}{64a_0^2\sigma^6} \left(\overline{\sigma}_0^2 - \sigma^2\right)^2,$$

so that we can conclude $\sigma^2 = \overline{\sigma}_0^2$. □

We now turn to the convergence of the Fisher information related to our likelihood field. Let H_n and $\overline{H}_n$ be the scaled Hessian matrices of the likelihood fields, defined for any $\xi \in \Xi$ as

$$H_n(\xi) = -\Phi_n^{-1/2} \frac{\partial^2 l_n(\xi)}{\partial \xi^2} \Phi_n^{-1/2}$$

$$\text{and } \overline{H}_n(\xi) = -\Phi_n^{-1/2} \frac{\partial^2 \overline{l}_n(\xi)}{\partial \xi^2} \Phi_n^{-1/2}. \tag{A.21}$$

Lemma A.5 (*Asymptotic Fisher Information*)**.** *Let $\Gamma(\xi_0)$ be the matrix*

$$\Gamma(\xi_0) = \begin{pmatrix} \dfrac{\sqrt{T}}{8a_0\overline{\sigma}_0^3} & 0 \\ 0 & \dfrac{1}{2a_0^4} \end{pmatrix}. \tag{A.22}$$

We have, for any ball V_n centered on ξ_0, shrinking to $\{\xi_0\}$,

$$\sup_{\xi_n \in V_n} \|H_n(\xi_n) - \Gamma(\xi_0)\| \xrightarrow{\mathbb{P}} 0. \tag{A.23}$$

Proof. First note that a small adaptation of Lemma 1–2 with second order derivatives of Ω^{-1} from Xiu (2010) yields $\sup_{\xi \in \Xi} \{H_n(\xi) - \overline{H}_n(\xi)\} \xrightarrow{\mathbb{P}} 0$ since $\pi_T^n \xrightarrow{\mathbb{P}} 0$. Now, Xiu (2010), bottom of p. 247 and after equation 41 on p. 248, can be

easily adapted to our case replacing $\int_0^T \sigma_s^2 ds$ by $T\overline{\sigma}_0^2$ as in the previous lemma, so that we have

$$\sup_{\xi \in \Xi} \left\{ \overline{H}_n(\xi) - \overline{H}_\infty(\xi) \right\} \xrightarrow{\mathbb{P}} 0, \tag{A.24}$$

with

$$\overline{H}_\infty(\xi) := \begin{pmatrix} \dfrac{\sqrt{T}}{8a\sigma^3} + \dfrac{(a^2 - a_0^2)\sqrt{T}}{16a^3\sigma^3} + \dfrac{3T(\overline{\sigma}_0^2 - \sigma^2)}{16a\sigma^5\sqrt{T}} & 0 \\ 0 & \dfrac{a_0^2}{a^6} - \dfrac{1}{2a^4} \end{pmatrix}. \tag{A.25}$$

It is immediate to check that

$$\overline{H}_\infty(\xi_0) = \begin{pmatrix} \dfrac{\sqrt{T}}{8a_0\overline{\sigma}_0^3} & 0 \\ 0 & \dfrac{1}{2a_0^4} \end{pmatrix}. \quad \square \tag{A.26}$$

We now adopt similar notations to Xiu (2010) in the proof of Lemma 3 (p. 248) and define the processes involved in the derivation of the central limit theorem. For $(\beta) \in \{(\sigma^2), (a^2)\}$, and $t \in [0, T]$, we define

$$M_1^{(\beta)}(t) := \sum_{i=1}^{N_n(t)} \frac{\partial \omega^{i,i}}{\partial \beta} \left\{ \left(\Delta X_{i,t}^n \right)^2 - \int_{t_{i-1}^n \wedge t}^{t_i^n \wedge t} \sigma_s^2 ds - \sum_{t_{i-1}^n \wedge t < s \leq t_i^n \wedge t} \Delta J_s^2 \right\}, \tag{A.27}$$

$$M_2^{(\beta)}(t) := \sum_{i=1}^{N_n(t)} \left\{ \sum_{1 \leq j < i} \frac{\partial \omega^{i,j}}{\partial \beta} \Delta X_{j,t}^n \right\} \Delta X_{i,t}^n, \tag{A.28}$$

$$M_3^{(\beta)}(t) := -2 \sum_{i=0}^{N_n(t)} \left\{ \sum_{j=1}^{N_n(t)} \frac{\partial \dot{\omega}^{i,j}}{\partial \beta} \Delta X_{j,t}^n \right\} \epsilon_{t_i^n}, \tag{A.29}$$

$$M_4^{(\beta)}(t) := \sum_{i=0}^{N_n(t)} \frac{\partial \ddot{\omega}^{i,i}}{\partial \beta} \left\{ \epsilon_{t_i^n}^2 - a_0^2 \right\} + 2 \sum_{i=0}^{N_n(t)} \left\{ \sum_{0 \leq j < i} \frac{\partial \ddot{\omega}^{i,j}}{\partial \beta} \epsilon_{t_j^n} \right\} \epsilon_{t_i^n}, \tag{A.30}$$

where in all the definitions (A.27)–(A.30), the terms involving the parameters such as $\Omega^{-1}, \dot{\Omega}^{-1}, \ddot{\Omega}^{-1}, \cdots$ are evaluated at point $\xi := (\sigma^2, a_0^2)$, for some $\sigma^2 \in [\underline{\Sigma}^2, \overline{\Sigma}^2]$. We also define the two-dimensional vectors $M_i(t) := \left(M_i^{(\sigma^2)}(t), M_i^{(a^2)}(t) \right)$ for $i \in \{1, \ldots, 4\}$. Note that we have the key decomposition

$$2\Phi_n^{1/2} \left\{ \Psi_n(\xi) - \overline{\Psi}_n(\xi) \right\} = \Phi_n^{-1/2} \left\{ M_1(T) + 2M_2(T) + M_3(T) + M_4(T) \right\}.$$

In the next few lemmas we investigate the limit of each one of those terms. In the presence of jumps and random observation times, we will see that they are not mere extensions of Lemma 3 in Xiu (2010) and that additional variance terms appear in the limits. We start by $M_1(T)$.

Lemma A.6. *We have*

$$\Phi_n^{-1/2} M_1(T) \xrightarrow{\mathbb{P}} 0.$$

Proof. We have to show $N_n^{-1/4} M_1^{(\sigma^2)}(T) \xrightarrow{\mathbb{P}} 0$ and $N_n^{-1/2} M_1^{(a^2)}(T) \xrightarrow{\mathbb{P}} 0$. We start with the case where $J = 0$. We are going to show that for any $(\beta) \in \{(\sigma^2), (a^2)\}$ we actually have $N_n^{-1/4} M_1^{(\beta)}(T) \xrightarrow{\mathbb{P}} 0$. To do so, note that we can write

$$M_1^{(\beta)}(T) = \sum_{i=1}^{N_n} \chi_i^n,$$

where

$$\chi_i^n = \frac{\partial \omega^{i,i}}{\partial \beta} \left\{ \left(\Delta X_i^n \right)^2 - \int_{t_{i-1}^n}^{t_i^n} \sigma_s^2 ds \right\}.$$

Now, since $\frac{\partial \omega^{i,i}}{\partial \beta} \in \mathcal{U} \subset \mathcal{G}^n_{i-1}$, $\chi^n_i \in \mathcal{G}^n_i$. Moreover, $\mathbb{E}\left[\chi^n_i | \mathcal{G}^n_{i-1}\right] = 0$, thus by Lemma 2.2.11 in Jacod and Protter (2011), it is sufficient to show that $N_n^{-1/2} \sum_{i=1}^{N_n} \mathbb{E}\left[(\chi^n_i)^2 | \mathcal{G}^n_{i-1}\right] \xrightarrow{\mathbb{P}} 0$. By Burkholder–Davis–Gundy inequality, we have

$$N_n^{-1/2} \sum_{i=1}^{N_n} \mathbb{E}\left[(\chi^n_i)^2 | \mathcal{G}^n_{i-1}\right] \leq 4N_n^{-1/2} \sum_{i=1}^{N_n} \left(\frac{\partial \omega^{i,i}}{\partial \beta}\right)^2 \int_{t^n_{i-1}}^{t^n_i} \mathbb{E}\left[\left(\Delta X^n_{i,s}\right)^2 \sigma^2_s \,\Big|\, \mathcal{G}^n_{i-1}\right] ds$$

$$\leq KN_n^{1/2} n^{-1+\gamma} \sum_{i=1}^{N_n} (t^n_i - t^n_{i-1})$$

$$\leq KN_n^{1/2} n^{-1+\gamma} \xrightarrow{\mathbb{P}} 0,$$

where we have used the fact that $\frac{\partial \omega^{i,i}}{\partial \beta} = O_{\mathbb{P}}(N_n^{1/2})$ uniformly in i. In the presence of jumps, it remains to show that the additional terms

$$N_n^{-1/4} \sum_{i=1}^{N_n} \frac{\partial \omega^{i,i}}{\partial \beta} \left\{\left(\Delta J^n_i\right)^2 - \sum_{t^n_{i-1} < s \leq t^n_i} \Delta J^2_s\right\}$$

and

$$2N_n^{-1/4} \sum_{i=1}^{N_n} \frac{\partial \omega^{i,i}}{\partial \beta} \Delta J^n_i \Delta \tilde{X}^n_i$$

are negligible. From the finite activity property, note that the first one is identically 0 for n sufficiently large. Again, for n sufficiently large, defining N^J the finite number of jumps of J on $[0, T]$, we can write the second term as

$$2N_n^{-1/4} \sum_{k=1}^{N^J} \underbrace{\frac{\partial \omega^{N_n(\tau_k), N_n(\tau_k)}}{\partial \beta}}_{O_{\mathbb{P}}(N_n^{1/2})} \Delta J_{\tau_k} \underbrace{\Delta \tilde{X}^n_{i_k}}_{O_{\mathbb{P}}(n^{-1/2+1/2\gamma})} \xrightarrow{\mathbb{P}} 0.$$

where i_k is such that $t^n_{i_k} \leq \tau_k \leq t^n_{i_k+1}$, and where we have used **(H)**. This concludes the proof. $\square$

Lemma A.7. *We have $\mathcal{G}_T$-stably in law that*

$$N_n^{-1/4} M_2^{(\sigma^2)}(T) \to \mathcal{MN}\left(0, \frac{5}{64T^{3/2}\sigma^7 a_0} \int_0^T \alpha_s^{-1} ds \left\{\int_0^T \sigma_s^4 \alpha_s ds + \sum_{0 \leq s \leq T} \Delta J_s^2 (\sigma_s^2 \alpha_s + \sigma_{s-}^2 \alpha_{s-})\right\}\right)$$

and

$$N_n^{-1/2} M_2^{(a^2)}(T) \xrightarrow{\mathbb{P}} 0.$$

Proof. As usual, we start by the case with no jumps, that is $J = 0$. We show the result for $M_2^{(\sigma^2)}$. The proof is conducted in three steps.

Step 1. We consider $\beta^n_{i,k,t} := \sigma_{t^n_k} \Delta W^n_{i,t}$, and we define $\tilde{M}_2^{(\sigma^2)}$ as

$$\tilde{M}_2^{(\sigma^2)}(t) := \sum_{i=1}^{N_n} \left\{\sum_{(i-L_n) \wedge 1 \leq j < i} \frac{\partial \omega^{i,j}}{\partial \sigma^2} \beta^n_{j,i-L_n-1,t}\right\} \beta^n_{i,i-L_n-1,t}, \tag{A.31}$$

that is when the increments are replaced by variables of the form $\sigma_{t^n_{i-L_n-1}} \Delta W^n_{j,t}$, where $\sigma_{t^n_{i-L_n-1}}$ is the value of the volatility process at the beginning of the truncated sum. We show that we have $N_n^{-1/4} \left\{M_2^{(\sigma^2)}(T) - \tilde{M}_2^{(\sigma^2)}(T)\right\} \xrightarrow{\mathbb{P}} 0$. We decompose

$$N_n^{-1/4} \left\{M_2^{(\sigma^2)} - \tilde{M}_2^{(\sigma^2)}\right\} = R_n^{(1)} + R_n^{(2)} + R_n^{(3)}, \tag{A.32}$$

with

$$R_n^{(1)} = N_n^{-1/4} \sum_{i=1}^{N_n} \sum_{1 \le j < i - L_n} \frac{\partial \omega^{i,j}}{\partial \sigma^2} \Delta X_{j,t}^n \Delta X_{i,t}^n, \tag{A.33}$$

$$R_n^{(2)} = N_n^{-1/4} \sum_{i=1}^{N_n} \sum_{(i-L_n) \wedge 1 \le j < i} \frac{\partial \omega^{i,j}}{\partial \sigma^2} \Delta X_{j,t}^n (\Delta X_{i,t}^n - \beta_{i,i-L_n-1,t}^n), \tag{A.34}$$

$$R_n^{(3)} = N_n^{-1/4} \sum_{i=1}^{N_n} \sum_{(i-L_n) \wedge 1 \le j < i} \frac{\partial \omega^{i,j}}{\partial \sigma^2} (\Delta X_{j,t}^n - \beta_{j,i-L_n-1,t}^n) \beta_{i,i-L_n-1,t}^n. \tag{A.35}$$

Now, proving that $R_n^{(1)}$ is negligible is immediate because when $|i - j| \ge L_n$, we have the domination $\frac{\partial \omega^{i,j}}{\partial \sigma^2} \le L\sqrt{N_n}e^{-N_n^\delta}$ for some $L > 0$ so that by an easy application of Cauchy–Schwarz inequality and estimates from Lemma A.1 we get $\mathbb{E}_{\mathcal{U}} \left| R_n^{(1)} \right| \xrightarrow{\mathbb{P}} 0$. Now we show the negligibility of $R_n^{(2)}$. Assume first that σ has no jumps, i.e $\tilde{J} = 0$. $R_n^{(2)}$ being a sum of martingale increments, it is sufficient to show that

$$N_n^{-1/2} \sum_{i=1}^{N_n} \mathbb{E}_{\mathcal{U}} [(A_i^n)^2 (\Delta X_{i,t}^n - \beta_{i,i-L_n-1,t}^n)^2] \xrightarrow{\mathbb{P}} 0,$$

where $A_i^n = \sum_{j=(i-L_n) \wedge 1}^{i-1} \frac{\partial \omega^{i,j}}{\partial \sigma^2} \Delta X_{j,t}^n$. Introducing $v_{i,k,t} := \sigma_t - \sigma_{t_{i-k-1}^n}$, $\delta_{i,k,t} := \int_{t_{i-1}^n \wedge t}^{t_i^n \wedge t} v_{i,k,s}^2 ds$, we thus need to show that

$$N_n^{-1/2} \sum_{i=1}^{N_n} \mathbb{E}_{\mathcal{U}} [(A_i^n)^2 \delta_{i,L_n,t}] \xrightarrow{\mathbb{P}} 0.$$

Itô's formula applied to $v_{i,k,t}^2$ when $\tilde{J} = 0$ yields

$$v_{i,k,t}^2 = \underbrace{\int_{t_{i-k-1}^n}^{t} 2v_{i,k,s} \tilde{\sigma}_s^{(1)} dW_s}_{u_{i,k,t}^{(1)}} + \underbrace{\int_{t_{i-k-1}^n}^{t} 2v_{i,k,s} \tilde{\sigma}_s^{(2)} d\tilde{W}_s}_{u_{i,k,t}^{(2)}} + \underbrace{\int_{t_{i-k-1}^n}^{t} \left\{ \left(\tilde{\sigma}_s^{(1)}\right)^2 + \left(\tilde{\sigma}_s^{(2)}\right)^2 \right\} ds}_{u_{i,k,t}^{(3)}},$$

so that defining $\delta_{i,k,t}^{(l)} := \int_{t_{i-1}^n \wedge t}^{t_i^n \wedge t} u_{i,k,s}^{(l)} ds$ for $l \in \{1, 2, 3\}$, we now show

$$N_n^{-1/2} \sum_{i=1}^{N_n} \mathbb{E}_{\mathcal{U}} [(A_i^n)^2 \delta_{i,L_n,t}^{(l)}] \xrightarrow{\mathbb{P}} 0. \tag{A.36}$$

For $l = 3$, we have $|\delta_{i,L_n,t}^{(3)}| \le L \Delta t_{i,t}^n (t_i^n - t_{i-L_n}^n) \le L n^{-2+2\gamma} L_n$ by **(H)**, and thus (A.36) boils down to showing that

$$N_n^{-1/2} n^{-2+2\gamma} L_n \sum_{i=1}^{N_n} \mathbb{E}_{\mathcal{U}} [(A_i^n)^2] \xrightarrow{\mathbb{P}} 0. \tag{A.37}$$

Using $\mathbb{E}_{\mathcal{U}} [\Delta X_k^n \Delta X_j^n] = 0$ for $j \ne k$, we deduce

$$N_n^{-1/2} n^{-2+2\gamma} L_n \sum_{i=1}^{N_n} \mathbb{E}_{\mathcal{U}} [(A_i^n)^2] = N_n^{-1/2} n^{-2+2\gamma} L_n \sum_{i=1}^{N_n} \sum_{j=(i-L_n) \wedge 1}^{i-1} \left(\frac{\partial \omega^{i,j}}{\partial \sigma^2}\right)^2 \mathbb{E}_{\mathcal{U}} \left[\left(\Delta X_j^n\right)^2\right]$$

$$\le L N_n^{-1/2} n^{-3+3\gamma} L_n \sum_{i=1}^{N_n} \sum_{j=(i-L_n) \wedge 1}^{i-1} \left(\frac{\partial \omega^{i,j}}{\partial \sigma^2}\right)^2$$

$$\le L N_n^2 n^{-3+3\gamma} L_n \to 0,$$

where we have used that by direct calculation we have $\sum_{i=1}^{N_n} \sum_{(i-L_n) \wedge 1 \le j < i} \left(\frac{\partial \omega^{i,j}}{\partial \sigma^2}\right)^2 = O_{\mathbb{P}}\left(N_n^{5/2}\right)$, and that $L_n = N_n^{1/2+\delta}$. For $l = 1$, we split (A.36) into two terms

$$N_n^{-1/2} \sum_{i=1}^{N_n} \mathbb{E}_{\mathcal{U}} [(A_i^n)^2 \delta_{i,L_n,t}^{(l)}] = P_n^{(1)} + P_n^{(2)}, \tag{A.38}$$

where

$$P_n^{(1)} = N_n^{-1/2} \mathbb{E}_{\mathcal{U}} \sum_{i=1}^{N_n} \sum_{(i-L_n) \wedge 1 \leq j < i} \left(\frac{\partial \omega^{i,j}}{\partial \sigma^2} \right)^2 \left(\Delta X_j^n \right)^2 \delta_{i,L_n,t}^{(1)},$$

$$P_n^{(2)} = N_n^{-1/2} \mathbb{E}_{\mathcal{U}} \sum_{i=1}^{N_n} \sum_{(i-L_n) \wedge 1 \leq j \neq k < i} \frac{\partial \omega^{i,j}}{\partial \sigma^2} \frac{\partial \omega^{i,k}}{\partial \sigma^2} \Delta X_j^n \Delta X_k^n \delta_{i,L_n,t}^{(1)}.$$

We have by Cauchy–Schwarz inequality

$$P_n^{(1)} \leq N_n^{-1/2} \sum_{i=1}^{N_n} \sum_{(i-L_n) \wedge 1 \leq j < i} \left(\frac{\partial \omega^{i,j}}{\partial \sigma^2} \right)^2 \left(\mathbb{E}_{\mathcal{U}} \left[\left(\Delta X_j^n \right)^4 \right] \mathbb{E}_{\mathcal{U}}[(\delta_{i,L_n,t}^{(1)})^2] \right)^{1/2},$$

$$\leq L N_n^{-1/2} n^{-3+3\gamma} L_n \sum_{i=1}^{N_n} \sum_{(i-L_n) \wedge 1 \leq j < i} \left(\frac{\partial \omega^{i,j}}{\partial \sigma^2} \right)^2$$

$$\leq L N_n^2 n^{-3+3\gamma} L_n \xrightarrow{\mathbb{P}} 0,$$

as $\mathbb{E}_{\mathcal{U}} \left[\left(\Delta X_j^n \right)^4 \right] \leq L n^{-2+2\gamma}$ by (A.5), and

$$\mathbb{E}_{\mathcal{U}} \left[\left(\delta_{i,L_n,t}^{(1)} \right)^2 \right] \leq L \Delta t_{i,t}^n (t_i^n - t_{i-L_n-1}^n) \mathbb{E}_{\mathcal{U}} \left[\sup_{s \in [t_{i-L_n-1}^n, t_i^n]} v_{i,L_n,s}^2 \right] \leq L n^{-4+4\gamma} L_n^2,$$

by the same estimate as for (A.5) for the Itô semimartingale $v_{i,L_n,s}$. For $P_n^{(2)} \xrightarrow{\mathbb{P}} 0$, we first note that for $k < j$ we have

$$\left| \mathbb{E}_{\mathcal{U}} \left[\Delta X_k^n \Delta X_j^n \delta_{i,L_n,t}^{(1)} \right] \right| \leq \mathbb{E}_{\mathcal{U}} \left[|\Delta X_k^n| \left| \int_{t_{i-1}^n \wedge t}^{t_i^n \wedge t} \mathbb{E} \left[\Delta X_j^n u_{i,L_n,s}^{(1)} \Big| \mathcal{G}_{j-1}^n \right] ds \right| \right]$$

$$\leq L \mathbb{E}_{\mathcal{U}} \left[|\Delta X_k^n| \int_{t_{i-1}^n \wedge t}^{t_i^n \wedge t} \left| \mathbb{E} \left[\int_{t_{j-1}^n}^{t_j^n} v_{i,L_n,u} \tilde{\sigma}_u^{(1)} \sigma_u du \Big| \mathcal{G}_{j-1}^n \right] \right| ds \right]$$

$$\leq L n^{-3+3\gamma} L_n^{1/2},$$

where the last step is obtained using **(H)** as for the previous estimates. Overall, we get

$$P_n^{(2)} \leq L N_n^{-1/2} n^{-3+3\gamma} L_n^{1/2} \sum_{i=1}^{N_n} \sum_{(i-L_n) \wedge 1 \leq j \neq k < i} \frac{\partial \omega^{i,j}}{\partial \sigma^2} \frac{\partial \omega^{i,k}}{\partial \sigma^2}$$

$$\leq L N_n^{-1/2} n^{-3+3\gamma} L_n^{3/2} \sum_{i=1}^{N_n} \sum_{(i-L_n) \wedge 1 \leq j < i} \left(\frac{\partial \omega^{i,j}}{\partial \sigma^2} \right)^2$$

$$\leq L N_n^2 n^{-3+3\gamma} L_n^{3/2} \xrightarrow{\mathbb{P}} 0.$$

Finally, when $l = 2$, we write the same decomposition as (A.38), and we note that the exact same calculation as in the case $l = 1$ for $P_n^{(1)}$ remains valid. Moreover, following closely the calculation above, we get $P_n^{(2)} = 0$ by orthogonality of the Brownian motions W and $\tilde{W}$. When σ has jumps of finite activity, we easily show as for previous calculations that an additional negligible term appears in $R_n^{(2)}$, and thus combining all those results we have $R_n^{(2)} \xrightarrow{\mathbb{P}} 0$. Finally, $R_n^{(3)} \xrightarrow{\mathbb{P}} 0$ is proven following the same line of reasoning as for $R_n^{(2)}$.

Step 2. We are going to apply Theorem 2-1 p. 238 from Jacod (1997) to the continuous martingale $N_n^{-1/4} \tilde{M}_2^{(\sigma^2)}$. Condition (2.8) is automatically satisfied with $B_t = 0$. We now show the variance condition (2.9). This boils down to showing that there exists an increasing limit process C_t such that for any $t \in [0, T]$

$$\left\langle \tilde{M}_2^{(\sigma^2)}, \tilde{M}_2^{(\sigma^2)} \right\rangle_t \xrightarrow{\mathbb{P}} C_t, \tag{A.39}$$

and $C_T = \frac{5}{64T^{3/2}\sigma^7 a_0} \int_0^T \alpha_s^{-1} ds \int_0^T \sigma_s^4 \alpha_s ds$. We introduce

$$L_n^{(1)} := N_n^{-1/2} \sum_{i=1}^{N_n} \sum_{(i-L_n) \wedge 1 \leq j < i} \left(\frac{\partial \omega^{i,j}}{\partial \sigma^2} \right)^2 \sigma_{t_{i-L_n-1}^n}^4 \left(\Delta W_{j,t}^n \right)^2 \Delta t_{i,t}^n,$$

$$L_n^{(2)} := N_n^{-1/2} \sum_{i=1}^{N_n} \sum_{(i-L_n) \wedge 1 \leq j \neq k < i} \frac{\partial \omega^{i,j}}{\partial \sigma^2} \frac{\partial \omega^{i,k}}{\partial \sigma^2} \sigma_{t_{i-L_n-1}^n}^4 \Delta W_{j,t}^n \Delta W_{k,t}^n \Delta t_{i,t}^n.$$

we have $\left\langle \tilde{M}_2^{(\sigma^2)}, \tilde{M}_2^{(\sigma^2)} \right\rangle_t = L_n^{(1)} + L_n^{(2)}$, so that our strategy to show (A.39) will be to prove that

$$L_n^{(1)} \overset{\mathbb{P}}{\to} C_t \tag{A.40}$$

$$L_n^{(2)} \overset{\mathbb{P}}{\to} 0. \tag{A.41}$$

For $L_n^{(2)}$, we have directly that $\mathbb{E}_{\mathcal{U}}\left[\left(L_n^{(2)} \right)^2 \right]$ is equal to

$$N_n^{-1} \sum_{\substack{|i_1-i_2| \leq L_n \ (i_1 \vee i_2)-L_n \leq j, k < (i_1 \wedge i_2) \\ 1 \leq i_1, i_2 \leq N_n}} \sum_{j \neq k} \frac{\partial \omega^{i_1,j}}{\partial \sigma^2} \frac{\partial \omega^{i_2,j}}{\partial \sigma^2} \frac{\partial \omega^{i_1,k}}{\partial \sigma^2} \frac{\partial \omega^{i_2,k}}{\partial \sigma^2} \mathbb{E}_{\mathcal{U}}\left[\sigma_{t_{i_1-L_n-1}^n}^4 \sigma_{t_{i_2-L_n-1}^n}^4 \right] \Delta t_{j,t}^n \Delta t_{k,t}^n \Delta t_{i_1,t}^n \Delta t_{i_2,t}^n,$$

where we have used that for $l < \min(j_1, j_2, k_1, k_2)$, we have $\mathbb{E}[\Delta W_{t,j_1}^n \Delta W_{t,j_2}^n \Delta W_{t,k_1}^n \Delta W_{t,k_2}^n | \mathcal{G}_l^n] = \Delta t_{j,t}^n \Delta t_{k,t}^n$ when $j_1 = j_2 = j$ and $k_1 = k_2 = k$, and the expectation is null otherwise. Now, using the boundedness of σ and the fact that $\Delta t_{j,t}^n \leq Ln^{-1+\gamma}$ by assumption (**H**), we obtain

$$\mathbb{E}_{\mathcal{U}}\left[\left(L_n^{(2)} \right)^2 \right] \leq L N_n^{-1} n^{-4+4\gamma} \sum_{\substack{|i_1-i_2| \leq L_n \ (i_1 \vee i_2)-L_n \leq j, k < (i_1 \wedge i_2) \\ 1 \leq i_1, i_2 \leq N_n}} \sum_{j \neq k} \frac{\partial \omega^{i_1,j}}{\partial \sigma^2} \frac{\partial \omega^{i_2,j}}{\partial \sigma^2} \frac{\partial \omega^{i_1,k}}{\partial \sigma^2} \frac{\partial \omega^{i_2,k}}{\partial \sigma^2},$$

which by direct calculation on the coefficients yields

$$\mathbb{E}_{\mathcal{U}}\left[\left(L_n^{(2)} \right)^2 \right] \leq L N_n^{-1} n^{-4+4\gamma} N_n^4 L_n$$
$$\leq N_n^{7/2+\delta} n^{-4+4\gamma} \to 0,$$

for γ and δ small enough. Now we turn to (A.40). We define $C_t := \frac{5}{64\sqrt{T}\sigma^7 a_0} \int_0^T \alpha_s^{-1} ds \int_0^t \sigma_s^4 \alpha_s ds$, and we further decompose $L_n^{(1)} - C_t$ into

$$L_n^{(1)} - C_t = \sum_{i=1}^{6} B_n^{(i)}, \tag{A.42}$$

with

$$B_n^{(1)} = N_n^{-1/2} \sum_{i=1}^{N_n} \sum_{(i-L_n) \wedge 1 \leq j < i} \left(\frac{\partial \omega^{i,j}}{\partial \sigma^2} \right)^2 \sigma_{t_{i-L_n-1}^n}^4 \left(\left(\Delta W_{j,t}^n \right)^2 - \Delta t_{j,t}^n \right) \left(t_i^n \wedge t - t_{i-1}^n \wedge t \right),$$

$$B_n^{(2)} = N_n^{-1/2} \Delta_n \sum_{i=1}^{N_n} \sum_{(i-L_n) \wedge 1 \leq j < i} \left(\frac{\partial \omega^{i,j}}{\partial \sigma^2} \right)^2 \sigma_{t_{i-L_n-1}^n}^4 \left(\alpha_{t_{j-1}^n} - \alpha_{t_{i-L_n-1}^n} \right) U_j^n \left(t_i^n \wedge t - t_{i-1}^n \wedge t \right),$$

$$B_n^{(3)} = N_n^{-1/2} \Delta_n \sum_{i=1}^{N_n} \sum_{(i-L_n) \wedge 1 \leq j < i} \left(\frac{\partial \omega^{i,j}}{\partial \sigma^2} \right)^2 \sigma_{t_{i-L_n-1}^n}^4 \alpha_{t_{i-L_n-1}^n} \left(U_j^n - 1 \right) \left(t_i^n \wedge t - t_{i-1}^n \wedge t \right),$$

$$B_n^{(4)} = \sum_{i=1}^{N_n} \left\{ N_n^{-1/2} \Delta_n \sum_{(i-L_n) \wedge 1 \leq j < i} \left(\frac{\partial \omega^{i,j}}{\partial \sigma^2} \right)^2 - \frac{5}{64T^{3/2}\sigma^7 a_0} \int_0^T \alpha_s^{-1} ds \right\} \sigma_{t_{i-L_n-1}^n}^4 \alpha_{t_{i-L_n-1}^n} \left(t_i^n \wedge t - t_{i-1}^n \wedge t \right),$$

$$B_n^{(5)} = \frac{5}{64T^{3/2}\sigma^7 a_0} \int_0^T \alpha_s^{-1} ds \sum_{i=1}^{N_n} \left\{ \sigma_{t_{i-L_n-1}^n}^4 \alpha_{t_{i-L_n-1}^n} - \sigma_{t_i^n}^4 \alpha_{t_i^n} \right\} \left(t_i^n \wedge t - t_{i-1}^n \wedge t \right),$$

$$B_n^{(6)} = \frac{5}{64T^{3/2}\sigma^7 a_0} \int_0^T \alpha_s^{-1} ds \left\{ \sum_{i=1}^{N_n} \sigma_{t_i^n}^4 \alpha_{t_i^n} \left(t_i^n \wedge t - t_{i-1}^n \wedge t \right) - \int_0^t \sigma_s^4 \alpha_s ds \right\}.$$

Using that $\mathbb{E}_{\mathcal{U}}\left[\left(\left(\Delta W_{j,t}^n\right)^2 - \Delta t_{j,t}^n\right)\left(\left(\Delta W_{k,t}^n\right)^2 - \Delta t_{k,t}^n\right)\right] = 0$ if $j \neq k$, and $2\left(\Delta t_{j,t}^n\right)^2$ otherwise, we obtain the estimate

$$\mathbb{E}_{\mathcal{U}}\left[\left(B_n^{(1)}\right)^2\right] \leq LN_n^3 L_n n^{-4+4\gamma} \overset{\mathbb{P}}{\to} 0.$$

Moreover, by the same deviation inequality as (A.5) for α (recall that α is an Itô semimartingale) we have $\mathbb{E}|\alpha_{t_{j-1}^n} - \alpha_{t_{i-L_n}^n}| \leq Ln^{-1/2}L_n^{1/2}$ so that we obtain easily $\mathbb{E}|B_n^{(2)}| \leq LN_n^2 n^{-5/2+2\gamma}L_n^{1/2} \to 0$. Similar computation to that of $B_n^{(1)}$ shows that $\mathbb{E}\left[\left(B_n^{(3)}\right)^2\right] \to 0$ since $\mathbb{E}[U_j^n - 1] = 0$ and $\mathbb{E}[(U_j^n - 1)(U_i^n - 1)] = 0$ when $i \neq j$. $B_n^{(4)} \overset{\mathbb{P}}{\to} 0$ is a direct consequence of the fact that by a direct calculation we have uniformly in i that $N_n^{-3/2}\Delta_n \sum_{(i-L_n)\wedge 1 \leq j < i}\left(\frac{\partial\omega^{i,j}}{\partial\sigma^2}\right)^2 \overset{\mathbb{P}}{\to} \frac{5}{64T^{3/2}\sigma^7 a_0}$ and that $N_n\Delta_n \overset{\mathbb{P}}{\to} \int_0^T \alpha_s^{-1}ds$ by (4.3), recalling that $\Delta_n = T/n$. $B_n^{(5)} \overset{\mathbb{P}}{\to} 0$ is, again a simple consequence of the deviation inequality (A.5) for the Itô semimartingale $\sigma^4\alpha$, and finally $B_n^{(6)} \overset{\mathbb{P}}{\to} 0$ is just the convergence of the Riemann sum toward the integral limit, and we are done. We show condition (2.10), i.e. that

$$N_n^{-1/4}\langle \tilde{M}_2^{(\sigma^2)}, W\rangle_t \overset{\mathbb{P}}{\to} 0. \tag{A.43}$$

Note that

$$N_n^{-1/4}\langle \tilde{M}_2^{(\sigma^2)}, W\rangle_t = N_n^{-1/4}\sum_{i=1}^{N_n} \sum_{(i-L_n)\wedge 1 \leq j < i} \frac{\partial\omega^{i,j}}{\partial\sigma^2}\sigma_{t_{i-L_n-1}^n}^2 \Delta W_{j,t}^n \Delta t_{i,t}^n, \tag{A.44}$$

so that by a straightforward calculation on the Brownian motion increments we have

$$\mathbb{E}_{\mathcal{U}}\left[\left(\langle \tilde{M}_2^{(\sigma^2)}, W\rangle_t\right)^2\right] \leq LN_n^2 L_n n^{-3+3\gamma} \overset{\mathbb{P}}{\to} 0. \tag{A.45}$$

Moreover, condition (2.11) is satisfied because $\tilde{M}_2^{(\sigma^2)}$ is continuous. Finally we show condition (2.12). But note that for any bounded martingale $\overline{N}$ orthogonal to W we have directly

$$\langle \tilde{M}_2^{(\sigma^2)}, \overline{N}\rangle_t = 0 \tag{A.46}$$

by (A.31), so that all the conditions required for the theorem hold.

Step 3. In the presence of jumps, for n large enough, an additional term appears in $M_2^{(\sigma^2)}(T)$. First, since J is of finite activity and by the Grigelionis decomposition for Itô-semimartingales (see e.g. Theorem 2.1.2 in Jacod and Protter (2011)), we can assume without loss of generality that the jump times of J are a subset of the support of a Poisson random measure μ on $\mathbb{R}_+ \times E$ for E some arbitrary Polish space, adapted to $\mathcal{F}_t$, and with finite intensity measure ν. Let thus $\tau_1, \cdots, \tau_p, \cdots$ be an exhausting sequence for the jumps of μ. Since J is of finite activity, for n sufficiently large we cannot have more than a single jump on intervals of the form $[t_{i-L_n}^n, t_i^n]$ because $\sup_{L_n < i \leq N_n} t_i^n - t_{i-L_n}^n \to^{a.s.} 0$ by assumption **(H)**. Therefore, if n is large enough, after a simple rearrangement of the terms that contain jumps, and by the previous calculation in the continuous case, we can write $M_2^{(\sigma^2)}(t)$ under the form

$$M_2^{(\sigma^2)}(t) = \tilde{M}_2^{(\sigma^2)}(t) + A_n^+(t) + A_n^-(t) + o_{\mathbb{P}}(1), \tag{A.47}$$

with

$$A_n^+(t) = \sum_{p \geq 1} \Delta J_{\tau_p} \sum_{j=i_p+1}^{i_p+L_n} \frac{\partial\omega^{i_p,j}}{\partial\sigma^2}\Delta\tilde{X}_{j,t}^n \text{ and } A_n^-(t) = \sum_{p \geq 1} \Delta J_{\tau_p} \sum_{j=i_p-L_n}^{i_p-1} \frac{\partial\omega^{i_p,j}}{\partial\sigma^2}\Delta\tilde{X}_{j,t}^n, \tag{A.48}$$

where i_p is such that $t_{i_p-1}^n < \tau_p \leq t_{i_p}^n$. We define

$$M_n^+(t, p) = \sum_{j=i_p+1}^{i_p+L_n} \frac{\partial\omega^{i_p,j}}{\partial\sigma^2}\Delta\tilde{X}_{j,t}^n \text{ and } M_n^-(t, p) = \sum_{j=i_p-L_n}^{i_p-1} \frac{\partial\omega^{i_p,j}}{\partial\sigma^2}\Delta\tilde{X}_{j,t}^n, \tag{A.49}$$

along with the following infinite dimensional vector $(G, (R_\infty^+(p), R_\infty^-(p))_{p \geq 0})$ such that G, $R_\infty^+(p)$ and $R_\infty^-(p)$ are i.i.d standard normal random variables. We can assume that Ω and $\mathcal{G}_T$ are rich enough to include such random variables information without loss of generality, since we can always construct a very good filtered extension as explained in pp. 36–37 of Jacod and Protter (2011). Now define

$$V_\infty := \frac{5}{64T^{3/2}\sigma^7 a_0}\int_0^T \alpha_s^{-1}ds, \tag{A.50}$$

$$M_\infty^+(p) := \sigma_{\tau_p}\alpha_{\tau_p}^{1/2}V_\infty^{1/2}R_\infty^+(p) \text{ and } M_\infty^-(p) := \sigma_{\tau_p-}\alpha_{\tau_p-}^{1/2}V_\infty^{1/2}R_\infty^-(p),$$

and

$$\tilde{G} := C_T^{1/2} G,$$

where C_T was defined in (A.39). We are going to show that $\mathcal{G}_T$-stably in law, we have the convergence

$$N_n^{-1/4}(\tilde{M}_2^{(\sigma^2)}(T), (M_n^+(T, p), M_n^-(T, p))_{p \geq 1}) \to (\tilde{G}, (M_\infty^+(p), M_\infty^-(p))_{p \geq 1}). \tag{A.51}$$

As the subset of finite dimensional cylinders is a convergence determining class for the product topology of $\mathbb{R}^{\mathbb{N}}$, it is sufficient to show that the above convergence holds for all finite families of the form $(\tilde{M}_2^{(\sigma^2)}(T), M_n^+(T, p_1), M_n^-(T, p_1), \dots, M_n^+(T, p_k), M_n^-(T, p_k))$, $k \geq 1$. Now, let us consider the filtration $\tilde{\mathcal{G}}_t$ which is the smallest filtration containing $\mathcal{G}_t$ and the jump times of $\mu, (\tau_p)_{p \geq 1}$. By independence of μ and the Wiener process W, $\tilde{X}$ is also a continuous Itô process with respect to the filtration $\tilde{\mathcal{G}}_t$, so that $(\tilde{M}_2^{(\sigma^2)}(t), M_n^+(t, p_1), M_n^-(t, p_1), \dots, M_n^+(t, p_k), M_n^-(t, p_k))_{t \in [0,T]}$ is a multi-dimensional continuous $\tilde{\mathcal{G}}_t$-martingale. Now, for n large enough and by the finite activity property, we have for any $1 \leq i \neq j \leq k$,

$$\langle M_n^+(., p_i), M_n^+(., p_j) \rangle_t = \langle M_n^-(., p_i), M_n^-(., p_j) \rangle_t = 0 \text{ a.s,}$$

and

$$\langle M_n^+(., p_i), M_n^-(., p_i) \rangle_t = 0 \text{ a.s.}$$

Moreover

$$N_n^{-1/2} \langle M_n^+(., p_i), M_n^+(., p_i) \rangle_t = N_n^{-1/2} \sum_{j=i_p+1}^{i_p+L_n} \left(\frac{\partial \omega^{i_p, j}}{\partial \sigma^2} \right)^2 \int_{t_{j-1}^n \wedge t}^{t_j^n \wedge t} \sigma_s^2 ds,$$

since the random index i_p is $\tilde{\mathcal{G}}_0$-measurable. By a similar (but easier) calculation than for $L_n^{(1)}$ above, we have

$$N_n^{-1/2} \langle M_n^+(., p_i), M_n^+(., p_i) \rangle_T \xrightarrow{\mathbb{P}} \frac{5}{64 T^{3/2} \sigma^7 a_0} \sigma_{\tau_{p_i}}^2 \alpha_{\tau_{p_i}} \int_0^T \alpha_s^{-1} ds,$$

and also

$$N_n^{-1/2} \langle M_n^-(., p_i), M_n^-(., p_i) \rangle_T \xrightarrow{\mathbb{P}} \frac{5}{64 T^{3/2} \sigma^7 a_0} \sigma_{\tau_{p_i}-}^2 \alpha_{\tau_{p_i}-} \int_0^T \alpha_s^{-1} ds.$$

Finally we show the negligibility of $N_n^{-1/2} \langle \tilde{M}_2^{(\sigma^2)}, M_n^+(., p_i) \rangle_t$ and $N_n^{-1/2} \langle \tilde{M}_2^{(\sigma^2)}, M_n^-(., p_i) \rangle_t$. We have

$$N_n^{-1/2} \langle \tilde{M}_2^{(\sigma^2)}, M_n^+(., p_i) \rangle_t = N_n^{-1/2} \sum_{j=i_p+1}^{i_p+L_n} \frac{\partial \omega^{i_p, j}}{\partial \sigma^2} \sum_{k=(j-L_n) \wedge 1}^{j-1} \frac{\partial \omega^{j, k}}{\partial \sigma^2} \Delta \tilde{X}_{k,t}^n \sigma_{j-L_n-1}^2 \Delta t_{j,t}^n, \tag{A.52}$$

so that by Assumption (H) we have $N_n^{-1} \mathbb{E}_{\mathcal{U}} \left[\langle \tilde{M}_2^{(\sigma^2)}, M_n^+(., p_i) \rangle_t^2 \right]$ bounded by

$$L N_n^{-1} n^{-2+2\gamma} \sum_{j_1, j_2 = i_p+1}^{i_p+L_n} \frac{\partial \omega^{i_p, j_1}}{\partial \sigma^2} \frac{\partial \omega^{i_p, j_2}}{\partial \sigma^2} \sum_{k=(j_1 \vee j_2)-L_n \wedge 1}^{j_1 \wedge j_2 - 1} \frac{\partial \omega^{j_1, k}}{\partial \sigma^2} \frac{\partial \omega^{j_2, k}}{\partial \sigma^2} \mathbb{E}_{\mathcal{U}} \left[\left(\Delta \tilde{X}_{k,t}^n \right)^2 \right]$$

$$\leq L N_n^{-1} n^{-3+3\gamma} \sum_{j_1, j_2 = i_p+1}^{i_p+L_n} \frac{\partial \omega^{i_p, j_1}}{\partial \sigma^2} \frac{\partial \omega^{i_p, j_2}}{\partial \sigma^2} \sum_{k=(j_1 \vee j_2)-L_n \wedge 1}^{j_1 \wedge j_2 - 1} \frac{\partial \omega^{j_1, k}}{\partial \sigma^2} \frac{\partial \omega^{j_2, k}}{\partial \sigma^2}$$

$$\leq L N_n^2 L_n n^{-3+3\gamma} \xrightarrow{\mathbb{P}} 0,$$

and thus the bracket is negligible. By a similar calculation we get that the bracket involving $M_n^-(p_i, .)$ is also negligible. Moreover, the convergence of $\langle \tilde{M}_2^{(\sigma^2)}, \tilde{M}_2^{(\sigma^2)} \rangle_t$ was shown in (A.39). Finally, as above we easily check the bracket of each martingale with either W or a bounded martingale orthogonal to W is negligible so that by another application of Theorem 2-1 in Jacod (1997) we have (A.51). From the representation

$$N_n^{-1/4} M_2^{(\sigma^2)}(T) = N_n^{-1/4} \left\{ \tilde{M}_2^{(\sigma^2)}(T) + \sum_{p \geq 1} \Delta J_{\tau_p}(M_n^+(T, p) + M_n^+(T, p)) \right\} + o_{\mathbb{P}}(1), \tag{A.53}$$

along with the fact that $\{p | \Delta J_{\tau_p} \neq 0\}$ is finite, we deduce by the stable convergence (A.51) that $\tilde{\mathcal{G}}$ (and *a fortiori* $\mathcal{G}$) stably in law

$$N_n^{-1/4} M_2^{(\sigma^2)}(T) \to \tilde{G} + \sum_{p \geq 1} \Delta J_{\tau_p}(M_\infty^+(p) + M_\infty^-(p)), \tag{A.54}$$

which is equal to the claimed distribution.

Finally, to show the convergence $N_n^{-1/2}M_2^{(a^2)}(T) \xrightarrow{\mathbb{P}} 0$, note that $\frac{\partial \omega^{i,j}}{\partial \sigma^2}$ and $\frac{\partial \omega^{i,j}}{\partial a^2}$ are equivalent up to a constant term so that all the above computations apply, and thus the scaling in $N_n^{-1/2}$ instead of $N_n^{-1/4}$ yields the negligibility of this term. $\quad\square$

Before turning to the limiting distribution of the other terms, we recall that for a σ-field $\mathcal{H}$, a random vector Z and a sequence of random vectors Z_n in $\mathbb{R}^b$, we say that Z_n converges in law toward Z conditioned on $\mathcal{H}$ if we have for any $u \in \mathbb{R}^b$

$$\mathbb{E}\left[e^{iu^T Z_n} \,\middle|\, \mathcal{H} \right] \xrightarrow{\mathbb{P}} \mathbb{E}\left[e^{iu^T Z} \,\middle|\, \mathcal{H} \right]. \tag{A.55}$$

Moreover, we recall in the following proposition a key result to combine stable convergence and conditional convergence. The proof of the result can be consulted in Barndorff-Nielsen et al. (2008) (proof of Proposition 5 on p. 1524).

Proposition A.8. *Let $\mathcal{H}$ be a given sub-σ-field, and let (Y_n) and (Z_n) be sequences of random vectors, such that each Y_n is $\mathcal{H}$-measurable and the sequence converges $\mathcal{H}$-stably toward a limiting distribution Y, and (Z_n) converges in law conditioned on $\mathcal{H}$ to some Z. Then $(Y_n, Z_n) \to (Y, Z)$ $\mathcal{H}$-stably in distribution.*

Lemma A.9. *We have conditioned on $\mathcal{G}_T$ the convergence in distribution*

$$N_n^{-1/4}M_3^{(\sigma^2)}(T) \to \mathcal{MN}\left(0, \frac{\sqrt{T}\overline{\sigma}_0^2}{8\sigma^5 a_0} \right), \tag{A.56}$$

and

$$N_n^{-1/2}M_3^{(a^2)}(T) \xrightarrow{\mathbb{P}} 0, \tag{A.57}$$

where we recall the definition $\overline{\sigma}_0^2 = T^{-1}\left\{ \int_0^T \sigma_s^2 ds + \sum_{0 \le s \le T} \Delta J_s^2 \right\}$.

Proof. We start with $M_3^{(\sigma^2)}(T)$. We apply a conditional version of Theorem 5.12 from Kallenberg (2006)(p. 92). Accordingly, we note that $M_3^{(\sigma^2)}(T)$ can be written as

$$N_n^{-1/4}M_3^{(\sigma^2)}(T) = \sum_{i=0}^{N_n} \tilde{\chi}_i^n, \tag{A.58}$$

where $\tilde{\chi}_i^n = -2N_n^{-1/4}\left\{ \sum_{j=1}^{N_n} \frac{\partial \dot{\omega}^{i,j}}{\partial \sigma^2} \Delta X_j^n \right\} \epsilon_{t_i^n}$, are rowwise conditionally independent and centered given $\mathcal{G}_T$. To get the theorem, it is thus sufficient to show that

$$\sum_{i=1}^{N_n} \mathbb{E}\left[\left(\tilde{\chi}_i^n \right)^2 \,\middle|\, \mathcal{G}_T \right] \xrightarrow{\mathbb{P}} \frac{1}{8\sqrt{T}\sigma^5 a_0}\left\{ \int_0^T \sigma_s^2 ds + \sum_{0 \le s \le T} \Delta J_s^2 \right\}, \tag{A.59}$$

and the Lindeberg condition, for any $\epsilon > 0$,

$$\sum_{i=0}^{N_n} \mathbb{E}\left[\left(\tilde{\chi}_i^n \right)^2 \mathbb{1}_{\{|\tilde{\chi}_i^n| \ge \epsilon\}} \,\middle|\, \mathcal{G}_T \right] \xrightarrow{\mathbb{P}} 0. \tag{A.60}$$

For (A.59), we can write $\sum_{i=0}^{N_n}\mathbb{E}\left[\left(\tilde{\chi}_i^n \right)^2 \,\middle|\, \mathcal{G}_T \right] = T_n^{(1)} + T_n^{(2)}$ with

$$T_n^{(1)} = 4a_0^2 N_n^{-1/2} \sum_{i=0}^{N_n} \sum_{j=1}^{N_n} \left(\frac{\partial \dot{\omega}^{i,j}}{\partial \sigma^2} \right)^2 \left(\Delta X_j^n \right)^2, \tag{A.61}$$

and

$$T_n^{(2)} = 4a_0^2 N_n^{-1/2} \sum_{i=0}^{N_n} \sum_{j \ne k=1}^{N_n} \frac{\partial \dot{\omega}^{i,j}}{\partial \sigma^2} \frac{\partial \dot{\omega}^{i,k}}{\partial \sigma^2} \Delta X_j^n \Delta X_k^n, \tag{A.62}$$

and using same techniques as for the proof of Lemma A.7 we easily get by direct calculation on the coefficients $\frac{\partial \dot{\omega}^{i,j}}{\partial \sigma^2}$ that we have $T_n^{(1)} \xrightarrow{\mathbb{P}} \frac{1}{8\sqrt{T}\sigma^5 a_0}\left\{ \int_0^T \sigma_s^2 ds + \sum_{0 \le s \le T} \Delta J_s^2 \right\}$, and $T_n^{(2)} \xrightarrow{\mathbb{P}} 0$. As for the Lindeberg condition, it is sufficient to notice that by independence of the Brownian increments and similar computation we have $\sum_{i=0}^{N_n}\mathbb{E}\left[\left(\tilde{\chi}_i^n \right)^4 \middle| \mathcal{G}_T \right] \xrightarrow{\mathbb{P}} 0$. Finally, for $M_3^{(a^2)}(T)$, all the previous calculation holds but now the scaling in $N_n^{-1/2}$ implies that $N_n^{-1/2}M_3^{(a^2)}(T) \xrightarrow{\mathbb{P}} 0$. $\quad\square$

Lemma A.10. We have conditioned on $\mathcal{G}_T$ the convergence in distribution

$$\Phi_n^{-1/2} M_4(T) \to \mathcal{N}\left(0, \begin{pmatrix} \dfrac{\sqrt{T}}{16a_0\sigma^3} & 0 \\ 0 & \dfrac{2}{a_0^4} + \dfrac{\mathrm{cum}_4[\epsilon]}{a_0^8} \end{pmatrix}\right). \tag{A.63}$$

Proof. This is an immediate adaptation of (45) and (47) pp. 248–249 in Xiu (2010) conditioned on $\mathcal{G}_T$ in lieu of $\sigma(X)$, since ϵ is independent of $\mathcal{G}_T$. $\quad\square$

We consider now the general case $B \geq 1$, and accordingly we define for $i \in \{1, \ldots, B\}$ the local QMLE $\widehat{\xi}_{n,(i)} = (\widehat{\sigma}_{n,(i)}^2, \widehat{a}_{n,(i)}^2)$, and $\Psi_{n,(i)}, \bar{\Psi}_{n,(i)}$ the score functions on the block i where all quantities are taken in the time interval $(T_{i-1}, T_i]$. We also introduce the notation

$$\widehat{\xi}_n := (\widehat{\sigma}_{n,(1)}^2, \widehat{a}_{n,(1)}^2, \ldots, \widehat{\sigma}_{n,(B)}^2, \widehat{a}_{n,(B)}^2),$$

$\Psi_n := (\Psi_{n,(1)}, \ldots, \Psi_{n,(B)})$, and $\bar{\Psi}_n := (\bar{\Psi}_{n,(1)}, \ldots, \bar{\Psi}_{n,(B)})$. The next lemma states the limit distribution of the vector $\Psi_n - \bar{\Psi}_n$. Finally we introduce the scaling factors $N_{n,(i)} := N_n(T_i) - N_n(T_{i-1})$ along with the global scaling matrix $\Phi_n = \mathrm{diag}(N_{n,(1)}^{1/2}, N_{n,(1)}, \ldots, N_{n,(B)}^{1/2}, N_{n,(B)}) \in \mathbb{R}^{2B \times 2B}$.

Lemma A.11. We have for any $\sigma^2 := (\sigma_{(1)}^2, \ldots, \sigma_{(B)}^2) \in [\underline{\sigma}^2, \overline{\sigma}^2]^B$, taking $\xi := (\sigma_{(1)}^2, a_0^2, \ldots, \sigma_{(B)}^2, a_0^2)$, stably in $\mathcal{G}_T$, the convergence in distribution

$$\Phi_n^{1/2}\left\{\Psi_n(\xi) - \bar{\Psi}_n(\xi)\right\} \to \mathcal{MN}\left(0, \begin{pmatrix} V_{(1)} & 0 & \cdots & 0 \\ 0 & V_{(2)} & 0 & \vdots \\ \vdots & 0 & \ddots & \vdots \\ 0 & \cdots & \cdots & V_{(B)} \end{pmatrix}\right),$$

where for $i \in \{1, \ldots, B\}$, $V_{(i)}$ is the two dimensional matrix defined by

$$V_{(i)} := \begin{pmatrix} \dfrac{1}{4a_0}\left(\dfrac{5\mathcal{Q}_{(i)}}{16\sigma_{(i)}^7 \Delta_B^{1/2}} + \dfrac{\bar{\sigma}_i^2 \sqrt{\Delta_B}}{8\sigma_{(i)}^5} + \dfrac{\sqrt{\Delta_B}}{16\sigma_{(i)}^3}\right) & 0 \\ 0 & \dfrac{1}{2a_0^4} + \dfrac{\mathrm{cum}_4[\epsilon]}{4a_0^8} \end{pmatrix},$$

with

$$\Delta_B \bar{\sigma}_i^2 := \int_{T_{i-1}}^{T_i} \sigma_s^2 ds + \sum_{T_{i-1} < s \leq T_i} \Delta J_s^2,$$

and we recall that

$$\mathcal{Q}_{(i)} = \Delta_B^{-1} \int_{T_{i-1}}^{T_i} \alpha_s^{-1} ds \left\{\int_{T_{i-1}}^{T_i} \sigma_s^4 \alpha_s ds + \sum_{T_{i-1} < s \leq T_i} \Delta J_s^2 (\sigma_s^2 \alpha_s + \sigma_{s-}^2 \alpha_{s-})\right\}.$$

Proof. First, for $i \in \{1, \ldots, B\}$, we define the processes $M_{1,(i)}, \ldots, M_{4,(i)}$ following the definitions (A.27)–(A.30) adapted to the time interval $(T_{i-1}, T_i]$ of length Δ_B. Accordingly, for $k \in \{1, \ldots, 4\}$, we denote by $\boldsymbol{M}_k$ the vector process $(M_{k,(1)}^{(\sigma^2)}, M_{k,(1)}^{(a^2)}, \ldots, M_{k,(B)}^{(\sigma^2)}, M_{k,(B)}^{(a^2)})$, and we note that we have the decomposition

$$2\Phi_n^{1/2}\left\{\Psi_n(\xi) - \bar{\Psi}_n(\xi)\right\} = \Phi_n^{-1/2}\left\{\boldsymbol{M}_1(T) + 2\boldsymbol{M}_2(T) + \boldsymbol{M}_3(T) + \boldsymbol{M}_4(T)\right\}.$$

For $i \in \{1, \ldots, B\}$, we consider the two terms $M_{3,(i)}(T)$ and $M_{4,(i)}(T)$. By independence of ϵ with the other processes we deduce that the conditional covariance term between those two processes is null. We use this fact along with the marginal convergences obtained in Lemmas A.9 and A.10 to obtain the convergence in law conditioned on $\mathcal{G}_T$

$$\Phi_{n,(i)}^{-1/2}\left\{M_{3,(i)}(T) + M_{4,(i)}(T)\right\} \to \mathcal{MN}\left(0, \begin{pmatrix} \dfrac{1}{a_0}\left(\dfrac{\bar{\sigma}_i^2 \sqrt{\Delta_B}}{8\sigma_{(i)}^5} + \dfrac{\sqrt{\Delta_B}}{16\sigma_{(i)}^3}\right) & 0 \\ 0 & \dfrac{2}{a_0^4} + \dfrac{\mathrm{cum}_4[\epsilon]}{a_0^8} \end{pmatrix}\right),$$

where $\Phi_{n,(i)} := \mathrm{diag}(N_{n,(i)}^{1/2}, N_{n,(i)})$. Now, by Slutsky's lemma, Lemmas A.6 and A.7 we also have the $\mathcal{G}_T$-stable convergence in distribution

$$\Phi_{n,(i)}^{-1/2} \left\{ M_{1,(i)}(T) + 2M_{2,(i)}(T) \right\} \to \mathcal{MN} \left(0, \begin{pmatrix} \dfrac{5\mathcal{Q}_{(i)}}{16a_0\sigma_{(i)}^7 \Delta_B^{3/2}} & 0 \\ 0 & 0 \end{pmatrix} \right).$$

Finally, by application of Property A.8 with sub-σ-field $\mathcal{G}_T$ since $M_{1,(i)}(T) + 2M_{2,(i)}(T)$ is $\mathcal{G}_T$-measurable, we deduce the joint $\mathcal{G}_T$-stable convergence of

$$\Phi_{n,(i)}^{-1/2} \left(M_{1,(i)}(T) + 2M_{2,(i)}(T), M_{3,(i)}(T) + M_{4,(i)}(T) \right),$$

hence the convergence of $\Phi_{n,(i)}^{-1/2} \left(M_{1,(i)}(T) + 2M_{2,(i)}(T) + M_{3,(i)}(T) + M_{4,(i)}(T) \right)$ toward a mixed normal distribution of random variance $4V_{(i)}$. Finally, as blocks are non overlapping, we deduce that for any $k, l \in \{1, \ldots, 4\}$, for any $i \neq j \in \{1, \ldots, B\}$ the martingales $M_{k,(i)}$ and $M_{l,(j)}$ are orthogonal so that we have automatically the joint convergence of $\Phi_n^{1/2} \left\{ \Psi_n(\xi) - \bar{\Psi}_n(\xi) \right\}$ to a mixed normal with block diagonal random variance matrix whose submatrices are $V_{(1)}, \ldots, V_{(B)}$, and we are done. $\square$

Finally, we derive a central limit theorem for $\widehat{\xi}_n$ to the limit $\xi_0 := (\bar{\sigma}_1^2, a_0^2, \ldots, \bar{\sigma}_B^2, a_0^2)$, and as a byproduct Theorem 8 (and Theorem 4).

Theorem A.12. *We have $\mathcal{G}_T$-stably in law that*

$$\Phi_n^{1/2} \left\{ \widehat{\xi}_n - \xi_0 \right\} \to \mathcal{MN} \left(0, \begin{pmatrix} V'_{(1)} & 0 & \cdots & 0 \\ 0 & V'_{(2)} & 0 & \vdots \\ \vdots & 0 & \ddots & \vdots \\ 0 & \cdots & \cdots & V'_{(B)} \end{pmatrix} \right),$$

where for $i \in \{1, \ldots, B\}$, $V'_{(i)}$ is the two dimensional matrix defined by

$$V'_{(i)} := \begin{pmatrix} a_0 \left(\dfrac{5\mathcal{Q}_{(i)}}{\bar{\sigma}_i \Delta_B^{3/2}} + \dfrac{3\bar{\sigma}_i^3}{\sqrt{\Delta_B}} \right) & 0 \\ 0 & 2a_0^4 + \mathrm{cum}_4[\epsilon] \end{pmatrix}.$$

In particular, Theorem 8 (and Theorem 4) hold.

Proof. First, note that we can easily extend Lemma A.11 to get a central limit theorem at the point $\xi_0 = (\bar{\sigma}_1^2, a_0^2, \ldots, \bar{\sigma}_B^2, a_0^2)$ for $\Phi_n^{1/2} \left\{ \Psi_n(\xi_0) - \bar{\Psi}_n(\xi_0) \right\}$ by a generalization of Slutsky's Lemma for stably convergent sequences (see e.g. Theorem 3.18 (b) in Häusler and Luschgy (2015)), where now the submatrices $V_{(i)}$ in the asymptotic variance of the mixed normal distribution have the form

$$V_{(i)} = \begin{pmatrix} \dfrac{1}{64a_0} \left(\dfrac{5\mathcal{Q}_{(i)}}{\bar{\sigma}_i^7 \Delta_B^{1/2}} + \dfrac{3\sqrt{\Delta_B}}{\bar{\sigma}_i^3} \right) & 0 \\ 0 & \dfrac{1}{2a_0^4} + \dfrac{\mathrm{cum}_4[\epsilon]}{4a_0^8} \end{pmatrix}.$$

To derive the CLT for the $2B$-dimensional estimator $\widehat{\xi}_n$, we follow the standard procedure and expand the score function around ξ_0. Thus, starting from the first order conditions on the score functions, we have

$$0 = \Psi_n \left(\widehat{\xi}_n \right) = \Psi_n(\xi_0) + \Phi_n^{-1/2} H_n(\zeta_n) \Phi_n^{1/2} \left(\widehat{\xi}_n - \xi_0 \right), \tag{A.64}$$

for some $\zeta_n \in \left[\xi_0, \widehat{\xi}_n \right]$, and where H_n is the block diagonal matrix with submatrices $H_{n,(1)}, \ldots, H_{n,(B)}$, and for $i \in \{1, \ldots, B\}$, $H_{n,(i)}$ is the scaled Hessian matrix of the log-likelihood field on block i, defined as in (A.21) adapted to the time interval $(T_{i-1}, T_i]$. In the same way, we define $\Gamma(\xi_0)$ as the block diagonal matrix whose subcomponents are $\Gamma_{(i)}(\xi_{0,(i)})$ where

$$\Gamma_{(i)}(\xi_{0,(i)}) := \begin{pmatrix} \dfrac{\sqrt{\Delta_B}}{8a_0\bar{\sigma}_i^3} & 0 \\ 0 & \dfrac{1}{2a_0^4} \end{pmatrix},$$

and $\xi_{0,(i)} := (\bar{\sigma}_i^2, a_0^2)$. We can rewrite (A.64) as

$$\Gamma(\xi_0)^{-1} H_n(\zeta_n) \Phi_n^{1/2} \left(\widehat{\xi}_n - \xi_0 \right) = -\Gamma(\xi_0)^{-1} \Phi_n^{1/2} \left\{ \Psi_n \left(\xi_0 \right) - \bar{\Psi}_n \left(\xi_0 \right) \right\} + \Phi_n^{1/2} \bar{\Psi}_n(\xi_0). \tag{A.65}$$

Note that, again, by a direct adaptation of (38) and (40) in Xiu (2010) (pp. 247–248) to the case of an irregular grid with $\pi_T^n \overset{\mathbb{P}}{\to} 0$ and on the interval $(T_{i-1}, T_i]$ we automatically get that each $\Phi_{n,(i)}^{1/2} \bar{\Psi}_{n,(i)}(\xi_{0,(i)}) = o_{\mathbb{P}}(1)$ so that $\Phi_n^{1/2} \bar{\Psi}_n(\boldsymbol{\xi}_0)$ is negligible. Now, $\widehat{\boldsymbol{\xi}}_n$ is consistent by application of Theorem A.4 to each $\widehat{\boldsymbol{\xi}}_{n,(i)}$ on block i. Therefore, $\zeta_n \overset{\mathbb{P}}{\to} \boldsymbol{\xi}_0$, and by virtue of Lemma A.5 applied to each submatrix $H_{n,(i)}$, we conclude on the one hand that $\Gamma(\boldsymbol{\xi}_0)^{-1} \boldsymbol{H}_n(\zeta_n) \overset{\mathbb{P}}{\to} \mathbf{I}$ where $\mathbf{I} \in \mathbb{R}^{2B \times 2B}$ is the identity matrix, and on the other hand by Slutsky's Lemma and the stable CLT for $\Phi_n^{1/2} \{ \Psi_n(\boldsymbol{\xi}_0) - \bar{\Psi}_n(\boldsymbol{\xi}_0) \}$ that the left-hand side of (A.65) tends $\mathcal{G}_T$-stably in law to a mixed normal distribution of block diagonal random variance matrix with submatrices of the form

$$
\begin{pmatrix} \dfrac{64a_0^2 \bar{\sigma}_i^6}{\Delta_B} & 0 \\ 0 & 4a_0^8 \end{pmatrix} \times \begin{pmatrix} \dfrac{1}{64a_0} \left(\dfrac{5\mathcal{Q}_{(i)}}{\bar{\sigma}_i^7 \Delta_B^{1/2}} + \dfrac{3\sqrt{\Delta_B}}{\bar{\sigma}_i^3} \right) & 0 \\ 0 & \dfrac{1}{2a_0^4} + \dfrac{\mathrm{cum}_4[\epsilon]}{4a_0^8} \end{pmatrix}
$$

$$
= \begin{pmatrix} a_0 \left(\dfrac{5\mathcal{Q}_{(i)}}{\bar{\sigma}_i \Delta_B^{3/2}} + \dfrac{3\bar{\sigma}_i^3}{\sqrt{\Delta_B}} \right) & 0 \\ 0 & 2a_0^4 + \mathrm{cum}_4[\epsilon] \end{pmatrix}
$$

$$
= V'_{(i)},
$$

and thus we have shown the CLT for $\widehat{\boldsymbol{\xi}}_n$. Now to get Theorem 8, it is sufficient to notice that

$$
\begin{pmatrix} N_n^{1/4} \left(\tilde{Q} - T\bar{\sigma}_0^2 \right) \\ N_n^{1/2} \left(B^{-1} \sum_{i=1}^{B} \widehat{a}_{n,(i)}^2 - a_0^2 \right) \end{pmatrix} = \Phi_n A \Phi_n^{-1/2} \Phi_n^{1/2} \left(\widehat{\boldsymbol{\xi}}_n - \boldsymbol{\xi}_0 \right), \tag{A.66}
$$

where $A \in \mathbb{R}^{2 \times 2B}$ has the form

$$
A = \begin{pmatrix} \Delta_B & 0 & \cdots & \Delta_B & 0 \\ 0 & B^{-1} & \cdots & 0 & B^{-1} \end{pmatrix},
$$

and from here we easily conclude that the left-hand side of (A.66) admits a CLT with the claimed asymptotic variance. Finally Theorem 4 is a particular case of Theorem 8. $\quad\square$

A.5. Proof of Theorem 1

Some details of the proof are omitted as the techniques used are very close to the QMLE case. We need to introduce some notation. We consider the block constant processes defined as

$$
\tilde{c}_t = c_i \quad \text{where} \quad T_{i-1} \le t < T_i,
$$
$$
\rho_t = \rho_{T_{i-1}, T_i} \quad \text{where} \quad T_{i-1} \le t < T_i.
$$
$$
\xi_t^2 = \xi_{T_{i-1}, T_i}^2 \quad \text{where} \quad T_{i-1} \le t < T_i.
$$

Condition (3.5) in Theorem 1 can be re-expressed as

$$
n^{1/4} \left(\tilde{K} - \int_0^T \sigma_u^2 du \right) \overset{L_{\mathcal{X}}}{\to} \mathcal{MN} \left(0, 4B^{1/2} \Delta_B \int_0^T \sigma_u^4 (\tilde{c}_u k_\bullet^{0,0} + \tilde{c}_u^{-1} 2 k_\bullet^{1,1} \rho_u \xi_u^2 + \tilde{c}_u^{-3} k_\bullet^{2,2} \xi_u^4) du \right).
$$

We also define the kernels for general processes A_t and C_t as

$$
K(A, C) = \gamma_0(A, C) + \sum_{h=1}^{H} k \left(\frac{h-1}{H} \right) (\gamma_h(A, C) + \gamma_{-h}(A, C)),
$$

where the realized autocovariance is defined as

$$
\gamma_h(A, C) = \sum_{j=1}^{n} (A_{\Delta j} - A_{\Delta(j-1)})(C_{\Delta(j-h)} - C_{\Delta(j-h-1)}),
$$

with $h = -H, \ldots, -1, 0, 1, \ldots, H$. We further define $K_i(A, C)$ the estimate on the ith block and we aggregate the local estimates to define the adapted version of $K(A, C)$ as

$$
\tilde{K}(A, C) = \sum_{i=1}^{B} K_i(A, C).
$$

We follow the same line of reasoning as in the proof of Theorem 4 (p. 1530, Barndorff-Nielsen et al., 2008). Accordingly, we just need to show an adapted version of Theorem 3 (p. 1492). Theorem 1 will then follow from Lemma 1 (p. 1523) and Property A.8. From now on, we aim to show the adapted version of Theorem 3 (p. 1492, Barndorff-Nielsen et al., 2008) which is stated in what follows.

Theorem 13 (*Adapted Version of Theorem 3 in* Barndorff-Nielsen et al. (2008)). *We assume that* $H = cn^{1/2}$. *As* $n \to \infty$ *we have that*

$$n^{1/4}\left(\tilde{K}(X,X) - \int_0^T \sigma_u^2 du\right) \overset{L_X}{\to} \mathcal{MN}\left(0, 4k_\bullet^{0,0}B^{1-\alpha}\Delta_B \int_0^T \sigma_u^4 \tilde{c}_u^{-1} du\right), \qquad (A.67)$$

$$n^{1/4}(\tilde{K}(X,U) + \tilde{K}(U,X)) \overset{L_X}{\to} \mathcal{MN}\left(0, 8\omega^2 k_\bullet^{1,1}B^\alpha \int_0^T \sigma_u^2 \tilde{c}_u du\right). \qquad (A.68)$$

In addition, when $k'(0)^2 + k'(1)^2 = 0$, *the asymptotic variance of* $\tilde{K}(U)$ *is equivalent to*

$$4\omega^4\left(n^{-1/4}B^{3\alpha-1}k_\bullet^{2,2}\sum_{i=1}^B c_i^3 + (B^{1/2}/(n^{1/2}m))\left\{k_\bullet^{1,1}\sum_{i=1}^B c_i + \sum_{i=2}^B \tilde{k}_\bullet^{1,1}(c_i, c_{i-1})\sqrt{c_i c_{i-1}}\right\}\right), \qquad (A.69)$$

where $\tilde{k}_\bullet^{1,1}(c_1, c_2) = \int_0^1 k'(x)k'(ax)dx$ *with* $a = \min(c_1, c_2)/\max(c_1, c_2)$ *and if* $1/m \to 0$

$$\sqrt{\frac{\tilde{H}^3}{n}}\tilde{K}(U,X) \overset{L_X}{\to} N\left(0, 4\omega^4 k_\bullet^{2,2}B^{3\alpha-1}\sum_{i=1}^B c_i^3\right). \qquad (A.70)$$

To show (A.67), we consider the continuous interpolated martingale $M_t = \sqrt{\frac{n}{H}}\left(\tilde{K}(X,X) - \int_0^t \sigma_u^2 du\right)$. As for the QMLE, we aim to use Theorem 2.1 (Jacod, 1997). To show condition (2.9), i.e. that $[M,M]_t \overset{\mathbb{P}}{\to} 4k_\bullet^{0,0}B^{1-\alpha}\Delta_B \int_0^t \sigma_u^4 \tilde{c}_u^{-1} du$, we express M_t as $\sum_{i=1}^B M_t^{(i)}$, where $M_t^{(i)}$ are such that $M_t^{(i)} = 0$ for $t \in [0, T_{i-1}]$, $M_t^{(i)} = M_t$ on $[T_{i-1}, T_i]$ and $M_t^{(i)} = M_{T_i}$ for $t \in [T_i, T]$. We can easily show that

$$[M,M]_t = \sum_{i=1}^B [M^{(i)}, M^{(i)}]_t. \qquad (A.71)$$

The $K(X)$ case in the proof of Theorem 3 (p. 1528, Barndorff-Nielsen et al., 2008) is based on a martingale theorem which shows that

$$[M^{(i)}, M^{(i)}]_t \overset{\mathbb{P}}{\to} 4k_\bullet^{0,0}B^{1-\alpha}(t - T_{i-1})\int_{T_{i-1}\wedge t}^{T_i\wedge t} \sigma_u^4 \tilde{c}_u^{-1} du. \qquad (A.72)$$

In view of (A.71) and (A.72), we have thus shown that $[M,M]_t \overset{\mathbb{P}}{\to} 4k_\bullet^{0,0}B^{1-\alpha}\Delta_B \int_0^t \sigma_u^4 \tilde{c}_u^{-1} du$. We show condition (2.10), i.e. that $\langle M, W\rangle_t \overset{\mathbb{P}}{\to} 0$, by a straightforward calculation on the Brownian motion increments. Also, condition (2.11) is satisfied because M is continuous. Finally we show that condition (2.12) hold, i.e. for any bounded martingale $\overline{N}$ orthogonal to W we have that

$$\langle M, \overline{N}\rangle_t = 0. \qquad (A.73)$$

This can be proven with the same line of reasoning as for Lemma A.7 for the QMLE.

The proof for (A.68) can adapt directly from the cross-term $K(X,U) + K(U,X)$ part in the proof of Theorem 3 (p. 1528, Barndorff-Nielsen et al., 2008). Indeed, on each block we have the convergence discussed on p. 1525, and it is clear that as the block terms are uncorrelated to each other conditioned on X_t, we obtain the convergence of the vector block estimates, with correlation limit between two different block terms equal to 0.

We aim to show now (A.69). In view of (A.3) on p. 1528 in Barndorff-Nielsen et al. (2008), we have

$$\tilde{K}(U) = \sum_{i=1}^B \left\{\underbrace{-\sum_{h=1}^{H^{(i)}}(w_{h+1}^{(i)} - 2w_h^{(i)} + w_{h-1}^{(i)})V_h^{(i)}}_{A_i} \underbrace{-\sum_{h=1}^{H^{(i)}}(w_{h+1}^{(i)} - w_{h-1}^{(i)})R_h^{(i)}}_{C_i}\right\},$$

where $w_h^{(i)} = k(\frac{h-1}{H^{(i)}})$ and $V_h^{(i)} = \sum_{j=(i-1)n/B+1}^{in/B}(U_{t_j}U_{t_{j-h}} + U_{t_j}U_{t_{j+h}} + U_{t_{j-1}}U_{t_{j-1-h}} + U_{t_{j-1}}U_{t_{j-1+h}})$ and C_i is due to end-effects. We have that $A_i \overset{\mathcal{L}}{\to} A_i^{(l)}$ and $C_i \overset{\mathcal{L}}{\to} C_i^{(l)}$ for some normally distributed variables $A_i^{(l)}$ and $C_i^{(l)}$ from the proof on p. 1529 in Barndorff-Nielsen et al. (2008). Actually, we can show that the convergence still holds for the random vector $(A_1, \ldots, A_B, C_1, \ldots, C_B)$

and thus we have that $\tilde{K}(U) \xrightarrow{\mathcal{L}} N(0, V)$ where V is equal to

$$\sum_{1 \leq i,j \leq B} \text{Cov}(A_i^{(l)}, A_j^{(l)}) + \text{Cov}(C_i^{(l)}, C_j^{(l)}). \tag{A.74}$$

We have that $n^{1/2} \sum_{i=1}^{B} \text{Var}(A^{(i)}) = B^{1/2} k_{\bullet}^{2,2} \sum_{i=1}^{B} c_i^3$, which shows the convergence to the first term in (A.69). The second term is obtained as

$$(B^{1/2}/(n^{1/2}m))^{-1} \sum_{i=1}^{B} \text{Var}(C_i) + 2 \sum_{i=2}^{B} \text{Cov}(A_i, A_{i-1}) = \left\{ k_{\bullet}^{1,1} \sum_{i=1}^{B} c_i + \sum_{i=2}^{B} \tilde{k}_{\bullet}^{1,1}(c_i, c_{i-1})\sqrt{c_i c_{i-1}} \right\}.$$

The other terms in (A.74) go to 0, thus we have shown (A.69). The convergence (A.70) is obtained as a straightforward consequence of (A.69).

A.6. Proof of Theorem 7

The proof adding jumps and stochastic observation times follows the same line of reasoning as for the QMLE case.

A.7. Proof of Corollaries 2 and 5

By Slutsky's Lemma, both corollaries will be proved if we have the consistency of the AVAR estimators. This is a consequence of the consistency of the estimators $\int_{T_{i-1}}^{T_i} \widehat{\sigma_u^2} du$ and $\int_{T_{i-1}}^{T_i} \widehat{\sigma_u^4} du$ by Theorem 3.1 and Remark 4 in Jacod et al. (2009) along with the consistency of $\hat{a}^2$ by, e.g., (21) in Zhang et al. (2005).

A.8. Proof of Proposition 3

$AVAR_{[T_{i-1},T_i]}^{(RK,c_i^*,2)}$ takes on the form

$$AVAR_{[T_{i-1},T_i]}^{(RK,c_i^*,2)} = a_0 \sqrt{B} \left(\Delta_B \int_{T_{i-1}}^{T_i} \sigma_u^4 du \right)^{3/4} g(\rho_{T_{i-1},T_i}).$$

In view of (A.1), we easily obtain $0 < \underline{\rho} \leq \rho_{T_{i-1},T_i} \leq 1$ where $\underline{\rho} = \frac{\sigma^2}{\overline{\sigma}^2}$. This gives us the estimate $8 \leq g(\rho_{T_{i-1},T_i}) \leq \overline{g} < \infty$ for some $\overline{g}$.

Let us define for $B \in \mathbb{N}, B \geq 1$, the random set $J_B := \{i \in \{1, \ldots, B\} | \sigma \text{ jumps on } (T_{i-1}, T_i]\}$. Because the jumps in σ are of finite activity, almost surely the cardinal of J_B, defined as $|J_B|$, tends to a finite value. Thus we can get rid of the terms $AVAR_{[T_{k-1},T_k]}^{(RK,c_i^*,2)}$ for which k is contained into J_B because

$$\sum_{i \in J_B} AVAR_{[T_{i-1},T_i]}^{(RK,c_i^*,2)} \leq |J_B| \Delta_B T^{1/2} a_0 \overline{\sigma}^3 \overline{g} \xrightarrow{a.s.} 0,$$

and similarly

$$g(1) a_0 T^{1/2} \sum_{i \in J_B} \int_{T_{i-1}}^{T_i} \sigma_u^3 du \xrightarrow{a.s.} 0.$$

Thus, the proposition will be proved if we show

$$\sum_{i \notin J_B} AVAR_{[T_{i-1},T_i]}^{(RK,c_i^*,2)} - g(1) a_0 T^{1/2} \sum_{i \notin J_B} \int_{T_{i-1}}^{T_i} \sigma_u^3 du \xrightarrow{a.s.} 0.$$

As the continuous part of σ is assumed to be an Itô process with bounded components, some calculation shows that for any $p > 0$, $q \geq 1$, and uniformly in $i \notin J_B$ we have the following expansion

$$\int_{T_{i-1}}^{T_i} \sigma_u^p du = \sigma_{T_{i-1}}^p \Delta_B + O_{\mathbb{L}^q}(\Delta_B^{3/2}),$$

where $A = O_{\mathbb{L}^q}(C)$, $C > 0$ means that $\mathbb{E}\left|\frac{A}{C}\right|^q$ is bounded. Thus, using again (A.1), we also obtain the expansions $\rho_{T_{i-1},T_i} = 1 + O_{\mathbb{L}^q}(\Delta_B^{1/2})$, $g(\rho_{T_{i-1},T_i}) = g(1) + O_{\mathbb{L}^q}(\Delta_B^{1/2})$, and $(\Delta_B \int_{T_{i-1}}^{T_i} \sigma_u^4 du)^{3/4} = \Delta_B^{1/2} \int_{T_{i-1}}^{T_i} \sigma_u^3 du + O_{\mathbb{L}^q}(\Delta_B^2)$ to get finally the estimate

$$AVAR_{[T_{i-1},T_i]}^{(RK,c_i^*,2)} = g(1) a_0 T^{1/2} \int_{T_{i-1}}^{T_i} \sigma_u^3 du + O_{\mathbb{L}^q}(\Delta_B^{3/2})$$

uniformly in $i \notin J_B$. At this stage we have thus proved that

$$\sum_{i \notin J_B} AVAR^{(RK,c_i^*,2)}_{[T_{i-1},T_i]} - g(1)a_0 T^{1/2} \sum_{i \notin J_B} \int_{T_{i-1}}^{T_i} \sigma_u^3 du = O_{\mathbb{L}^q}(\Delta_B^{1/2}).$$

To get the almost sure convergence to 0, we define Y_B as the left hand side of the previous equality and note that $\mathbb{E} \sum_{B=1}^{+\infty} |Y_B|^q < +\infty$ for any $q > 2$. This gives us that $\sum_{B=1}^{+\infty} |Y_B|^q < +\infty$ a.s. and so $|Y_B|^q \overset{a.s.}{\to} 0$, which completes the proof.

A.9. Proof of Remark 2

We show the inequality $g(\rho)\kappa^{-1} \geq g(1)$ for any admissible couple (ρ, κ). Note that by the domination $\kappa \leq \rho^{1/2}$ obtained on the account of (2.1), it is sufficient to show that the function $f : \rho \to \rho^{-1/2}g(\rho)$ is decreasing on the interval $(0, 1]$. We let $p(\rho) = \sqrt{1 + \sqrt{1 + 3d/\rho^2}}$, and a short calculation shows us that $f'(\rho)$ has the same sign as $p'(\rho)(1 - p(\rho)^{-2})$. Therefore, the inequality $p(\rho) \geq 1$ implies that f is decreasing if and only if p is, which is obvious.

A.10. Proof of Proposition 6

This proof follows the same line of reasoning as for the proof of Proposition 3.

A.11. Proof of Proposition 9

When $J = 0$, this is a straightforward adaptation of the proof of Proposition 3 using the new estimates for any $q \geq 1$

$$R^{1/2}_{(i)} = \Delta_B^{-1/2} \left(\int_0^T \alpha_s^{-1} ds \right)^{1/2} \alpha_{T_{i-1}} + o_{\mathbb{L}^q}(\Delta_B^{-1/2}),$$

$$\left(\Delta_B Q_{(i)} \right)^{3/4} = \Delta_B^{3/2} \sigma_{T_{i-1}}^3 + O_{\mathbb{L}^q}(\Delta_B^2),$$

and

$$g(\widetilde{\rho}_{T_{i-1},T_i}) = g(1) + O_{\mathbb{L}^q}\left(\Delta_B^{1/2} \right).$$

A.12. Proof of Proposition 11

When $J \neq 0$, the situation is fairly different. Let us define the random set

$$J_B^X := \{i \in \{1, \dots, B\} | X \text{ jumps on } (T_{i-1}, T_i]\}.$$

Since J is of finite activity, by taking n sufficiently large, we may assume that for any $i \in J_B^X$, X jumps exactly once on $(T_{i-1}, T_i]$. Splitting the sum of local variances

$$AVAR_B^{(RK,rob)} = \sum_{i \in J_B^X} R^{1/2}_{(i)} AVAR^{(RK,rob,\widetilde{c}_i^*)}_{[T_{i-1},T_i]} + \sum_{i \notin J_B^X} R^{1/2}_{(i)} AVAR^{(RK,rob,\widetilde{c}_i^*)}_{[T_{i-1},T_i]}$$

$$= I + II,$$

again by the finite activity property of J we easily deduce from the proof of Proposition 9 that

$$II \overset{a.s.}{\to} 8g(1)a_0 \left(\int_0^T \alpha_s^{-1} ds \right)^{1/2} \int_0^T \alpha_s^{1/2} \sigma_s^3 ds.$$

Now we derive the limit of I. We write $\tau_1, \dots, \tau_{\widetilde{N}_J}$ the jump times of J labeled such that for any $i \in J_B^X$, $T_{i-1} < \tau_i \leq T_i$. For any $i \in J_B^X$, we have the estimates

$$\bar{\sigma}_i^2 \overset{a.s.}{\sim} \Delta_B^{-1} \Delta J_{\tau_i}^2,$$

$$Q_{(i)} \overset{a.s.}{\sim} \Delta J_{\tau_i}^2 \left(\sigma_{\tau_i}^2 + \sigma_{\tau_i-}^2 \right),$$

where for the latter expression we used the continuity of α at time τ_i which is a consequence of the independence of α and X. We thus have

$$\widetilde{\rho}_{T_{i-1},T_i} \overset{a.s.}{\sim} \Delta_B^{-1/2} |\Delta J_{\tau_i}| \left(\sigma_{\tau_i}^2 + \sigma_{\tau_i-}^2 \right)^{-1/2}.$$

Combined with $g(\rho) \overset{\rho \to +\infty}{\sim} \frac{16}{3}\sqrt{\rho k_\bullet^{0,0} k_\bullet^{1,1}} \left(\frac{1}{\sqrt{2}} + \sqrt{2}\right)$, we deduce that

$$I \overset{a.s.}{\to} \frac{16}{3} a_0 \left(\frac{1}{\sqrt{2}} + \sqrt{2}\right) \sqrt{k_\bullet^{0,0} k_\bullet^{1,1}} \left(\int_0^T \alpha_s^{-1} ds\right)^{1/2} \sum_{0 < s \le T} \Delta J_s^2 \left(\sigma_s^2 \alpha_s + \sigma_{s-}^2 \alpha_{s-}\right)^{1/2}.$$

A.13. Proof of Propositions 10 and 12

The proofs follow exactly the same line of reasoning as the proofs of Propositions 9 and 11.

References

Aït-Sahalia, Y., Mykland, P.A., Zhang, L., 2005. How often to sample a continuous-time process in the presence of market microstructure noise. Rev. Financ. Stud. 18 (2), 351–416.

Altmeyer, R., Bibinger, M., 2015. Functional stable limit theorems for quasi-efficient spectral covolatility estimators. Stochastic Process. Appl. 125 (12), 4556–4600.

Andersen, T.G., Bollerslev, T., Diebold, F.X., Labys, P., 2001. The distribution of realized exchange rate volatility. J. Amer. Statist. Assoc. 96 (453), 42–55.

Andersen, T.G., Dobrev, D., Schaumburg, E., 2012. Jump-robust volatility estimation using nearest neighbor truncation. J. Econometrics 169 (1), 75–93.

Andersen, T.G., Dobrev, D., Schaumburg, E., 2014. A robust neighborhood truncation approach to estimation of integrated quarticity. Econometric Theory 30 (1), 3–59.

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

Barndorff-Nielsen, O.E., Hansen, P.R., Lunde, A., Shephard, N., 2009. Realized kernels in practice: Trades and quotes. Econom. J. 12 (3), 1–33.

Barndorff-Nielsen, O.E., Shephard, N., 2002. Estimating quadratic variation using realized variance. J. Appl. Econometrics 17 (5), 457–477.

Bibinger, M., Mykland, P.A., 2016. Inference for multi-dimensional high-frequency data with an application to conditional independence testing. Scand. J. Stat. 43 (4), 1078–1102.

Clinet, S., Potiron, Y., 2017. Estimation for high-frequency data under parametric market microstructure noise. Working paper available at arXiv:1712.01479.

Clinet, S., Potiron, Y., 2018. Statistical inference for the doubly stochastic self-exciting process. Bernoulli 24 (4B), 3469–3493.

Da, R., Xiu, D., 2017. When moving-average models meet high-frequency data: Uniform inference on volatility. Working paper available on Dacheng Xiu's website.

Engle, R.F., Patton, A.J., 2001. What good is a volatility model? Quant. Finance 1 (2), 237–245.

Genon-Catalot, V., Jacod, J., 1993. On the estimation of the diffusion coefficient for multi-dimensional diffusion processes. Ann. Inst. H. Poincaré Probab. Statist. 29 (1), 119–151.

Ghysels, E., Harvey, A.C., Renault, E., 1996. 5 stochastic volatility. In: Handbook of Statistics, Vol. 14. pp. 119–191.

Gloter, A., Jacod, J., 2001. Diffusions with measurement errors. i. local asymptotic normality. ESAIM Probab. Stat. 5, 225–242.

Hansen, P.R., Lunde, A., 2006. Realized variance and market microstructure noise. J. Bus. Econom. Statist. 24 (2), 127–161.

Häusler, E., Luschgy, H., 2015. Stable Convergence and Stable Limit Theorems, Vol. 74. Springer.

Jacod, J., 1994. Limit of random measures associated with the increments of a brownian semimartingale. Unpublished manuscript, Laboratorie de Probabilities, Universite P. and M. Curie, Paris.

Jacod, J., 1997. On continuous conditional gaussian martingales and stable convergence in law. Sémin. Probab. Strasbo. 31, 232–246.

Jacod, J., Li, Y., Mykland, P.A., Podolskij, M., Vetter, M., 2009. Microstructure noise in the continuous case: the pre-averaging approach. Stochastic Process. Appl. 119 (7), 2249–2276.

Jacod, J., Mykland, P.A., 2015. Microstructure noise in the continuous case: Approximate efficiency of the adaptive pre-averaging method. Stochastic Process. Appl. 125 (8), 2910–2936.

Jacod, J., Protter, P., 1998. Asymptotic error distributions for the euler method for stochastic differential equations. Ann. Probab. 26, 267–307.

Jacod, J., Protter, P.E., 2011. Discretization of Processes. Springer Science & Business Media.

Jacod, J., Rosenbaum, M., 2013. Quarticity and other functionals of volatility: efficient estimation. Ann. Statist. 41 (3), 1462–1484.

Kallenberg, O., 2006. Foundations of Modern Probability. Springer Science & Business Media.

Li, Y., Xie, S., Zheng, X., 2016. Efficient estimation of integrated volatility incorporating trading information. J. Econometrics 195 (1), 33–50.

Mancino, M.E., Sanfelici, S., 2012. Estimation of quarticity with high-frequency data. Quant. Finance 12 (4), 607–622.

Meddahi, N., 2002. A theoretical comparison between integrated and realized volatility. J. Appl. Econometrics 17 (5), 479–508.

Mykland, P.A., Zhang, L., 2006. ANOVA for diffusions and Ito processes. Ann. Statist. 34 (4), 1931–1963.

Mykland, P.A., Zhang, L., 2009. Inference for continuous semimartingales observed at high frequency. Econometrica 77 (5), 1403–1445.

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, Boca Raton, FL, pp. 109–190.

Potiron, Y., Mykland, P.A., 2016. Local parametric estimation in high frequency data. Working paper available at arXiv:1603.05700.

Potiron, Y., Mykland, P.A., 2017. Estimation of integrated quadratic covariation with endogenous sampling times. J. Econometrics 197 (1), 20–41.

Reiss, M., 2011. Asymptotic equivalence for inference on the volatility from noisy observations. Ann. Statist. 39 (2), 772–802.

Todorov, V., Tauchen, G., 2011. Volatility jumps. J. Bus. Econom. Statist. 29 (3), 356–371.

Van der Vaart, A.W., 2000. Asymptotic Statistics. Cambridge University Press.

Varneskov, R.T., 2016. Flat-top realized kernel estimation of quadratic covariation with nonsynchronous and noisy asset prices. J. Bus. Econom. Statist. 34 (1), 1–22.

Xiu, D., 2010. Quasi-maximum likelihood estimation of volatility with high frequency data. J. Econometrics 159 (1), 235–250.

Zhang, L., 2001. From Martingales to ANOVA: Implied and Realized Volatility (Ph.D. thesis), University of Chicago, Department of Statistics.

Zhang, L., 2006. Efficient estimation of stochastic volatility using noisy observations: A multi-scale approach. Bernoulli 12 (6), 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 (472), 1394–1411.