Available online at www.sciencedirect.com

ScienceDirect

Stochastic Processes and their Applications 125 (2015) 2910–2936

stochastic processes and their applications

www.elsevier.com/locate/spa

Microstructure noise in the continuous case: Approximate efficiency of the adaptive pre-averaging method

Jean Jacod[a], Per A. Mykland[b,*]

[a] Institut de Mathématiques de Jussieu (CNRS-UMR 7586 and Université Pierre et Marie Curie-P6), 4 Place Jussieu, 76005 Paris, France

[b] Department of Statistics, The University of Chicago, Chicago, IL 60637, USA

Received 30 August 2013; received in revised form 27 January 2015; accepted 12 February 2015
Available online 14 March 2015

Abstract

This paper introduces adaptiveness to the non-parametric estimation of volatility in high frequency data. We consider general continuous Itô processes contaminated by microstructure noise. In the context of pre-averaging, we show that this device gives rise to estimators that are within 7% of the commonly conjectured "quasi-lower bound" for asymptotic efficiency. The asymptotic variance is of the form constant × bound, where the constant does not depend on the process to be estimated. The results hold with mild assumptions on the noise, and extend to mildly irregular observations.

Keywords: Adaptive estimation; Discrete observation; Efficiency; High frequency data; Itô process; Leverage effect; Microstructure; Pre-averaging; Realized volatility; Semi-martingale; Stable convergence

1. Introduction

This paper is concerned with the estimation of integrated volatility for a one-dimensional Brownian semimartingale of the form

$$X_t = X_0 + \int_0^t b_s ds + \int_0^t \sigma_s dW_s, \tag{1.1}$$

* Corresponding author.
E-mail addresses: jean.jacod@upmc.fr (J. Jacod), mykland@pascal.uchicago.edu (P.A. Mykland).

http://dx.doi.org/10.1016/j.spa.2015.02.005

so the quantity of interest is $C_T = \int_0^T \sigma_s^2\, ds$, when the underlying process is observed at regularly spaced times over the interval $[0, T]$,[1] and when it is contaminated with a so-called microstructure noise. The time span T here is *fixed*, whereas the inter-observations time Δ_n is small and in the asymptotic it goes to 0.

This topic has been the object of a large number of investigations in the past fifteen years, starting with the non-noisy case in the early papers [2,3,7,21,33]. The influence of microstructure noise has been considered, for example, in [5,35], for a white noise independent of the underlying process, and estimators with the efficient rate of convergence $n^{1/4}$ when n is the number of observations have been introduced in [34,6,29,18,32,30], with various assumptions on the noise.

We are concerned here with "asymptotic efficiency". Rate efficiency is achieved by many estimators in the previous references, but here we are interested in how low one can push the asymptotic variance. Genuine efficiency bounds, in the sense of Hajek convolution theorem for instance, is known when σ_t is time varying but non-random, or is a function of state variables, and noise is additive and Gaussian [30]. The variance bound involves σ and the noise variance. In more general situations concerning σ_t and/or the noise, it has been conjectured that the variance bound is the similar expression, but with the random σ_t plugged in. In a recent breakthrough paper, [1] finds an estimator based on the spectral approach of [30] which reaches the conjectured lower bound in a very general situation. For a compare and contrast, we refer to Remarks 3.1 and 5.2, where a more thorough discussion takes place.

Since several of the other approaches are widely applied, it is of some interest to see how close one can get to optimality for also for these methods. All methods necessitate the choice of a bandwidth and of a kernel or weight function, and for the other known approaches to volatility estimation, optimal estimators have not been found. For a given kernel, when $\sigma_t = \sigma$ is not varying one knows how to choose a bandwidth minimizing the asymptotic variance; in contrast, an optimal choice of the kernel is still an open problem, although some choices lead to a variance as low as 1.003 times the lower bound, and the very simple triangular kernel specified later gives a variance equal to 1.07 times the lower bound. On the other hand, for any given kernel a significant variability of σ_t induces the asymptotic variance to be *much* bigger than the (conjectured) efficient variance bound, if the bandwidth is taken to be the same over the whole time interval $[0, T]$; for example, with the variability of the spot volatility typically encountered within a day or a week, the effective asymptotic variance is as much as 4 or 5 times the lower bound.

The aim of this paper is to provide an adaptive device for the choice of the bandwidth, which allows us to obtain the minimal asymptotic variance associated with any particular choice of the kernel or weight function. We only consider the pre-averaging class [29,18] of estimators, although similar improvements can probably be done for other methods. We also emphasize that the method works for *all* weight functions, although in the Monte Carlo we illustrate it for the triangular function only. On the other hand, concerning the choice of the weight function, we only provide a heuristic discussion, see Sections 3 and 5. The reason is that in the pre-averaging setting an optimal choice may not even exist, although it is possible to achieve genuine efficiency by using optimized linear combinations of the estimators with a proper family of weight functions, such as the sine functions with various frequencies. The proof, however, would be much more involved and the practical implementation rather complicated. From a practical viewpoint, we feel that achieving an asymptotic variance smaller than 1.1 times the lower bound is "practically efficient", if not mathematically efficient.

[1] Though see Remark 4.4 at the end of Section 4.

The plan for the paper is as follows. In Section 2 we review "classical" pre-averaging, and in Section 3 we discuss known results on efficiency. Based on this, we propose the adaptive pre-averaging estimator in Section 4, and investigate its optimality properties in Section 5. We report the results of Monte-Carlo experiments in Section 6, and then provide the proofs in the Appendix.

2. Assumptions, and background on pre-averaging

2.1. Setting

We start with a description of the setting. We have a one-dimensional continuous underlying process $X = (X_t)_{t\geq 0}$ on a filtered probability space $(\Omega^{(0)}, \mathcal{F}^{(0)}, (\mathcal{F}_t^{(0)})_{t\geq 0}, \mathbb{P}^{(0)})$. This process has the form (1.1), where $W = (W_t)$ is a standard Wiener process, and $b = (b_t)$ and $\sigma = (\sigma_t)$ are processes satisfying the following rather weak assumption:

Assumption (H). Both processes b and σ are adapted and *càdlàg, i.e.,* right-continuous with left limits in time. □

The process is observed at regularly spaced times $i\Delta_n$ for all $i = 0, 1, \ldots, k, \ldots, [T/\Delta_n]$, where T is the *fixed* time horizon. We are interested in the estimation of the integrated volatility $C_t = \int_0^t c_s\,ds$, where $c_s = \sigma_s^2$.

The observation is subject to an error, that is, at stage n and instead of the values $X_i^n = X_{i\Delta_n}$ for $i \geq 0$, we observe the variables $Z_i^n = X_i^n + Y_i^n$ and the error, or microstructure noise, is Y_i^n. Loosely speaking, we assume that, conditionally on the whole process X, and for any given n, the errors Y_i^n are independent, each one having a (conditional) law which possibly depends on the time and on the outcome ω, in an "adapted" way, and with conditional expectations 0.

Mathematically speaking, this can be realized as follows: for any $t \geq 0$ we have a transition probability $Q_t(\omega^{(0)}, dy)$ from $(\Omega^{(0)}, \mathcal{F}_t^{(0)})$ into $\mathbb{R}$, which satisfies

$$\int y\, Q_t(\omega^{(0)}, dy) = 0. \tag{2.1}$$

We endow the space $\Omega^{(1)} = \mathbb{R}^{[0,\infty)}$ with the product Borel σ-field $\mathcal{F}^{(1)}$ and with the probability $\mathbb{Q}(\omega^{(0)}, d\omega^{(1)})$ which is the product $\otimes_{t\geq 0}\; Q_t(\omega^{(0)}, .)$. We also call $(Y_t)_{t\geq 0}$ the "canonical process" on $(\Omega^{(1)}, \mathcal{F}^{(1)})$ and the filtration $\mathcal{F}_t^{(1)} = \sigma(Y_s : s \leq t)$. Then we consider the filtered probability space $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t\geq 0}, \mathbb{P})$ defined as follows:

$$\left.\begin{array}{l} \Omega = \Omega^{(0)} \times \Omega^{(1)}, \qquad \mathcal{F} = \mathcal{F}^{(0)} \times \mathcal{F}^{(1)}, \qquad \mathcal{F}_t = \bigcap_{s>t} \mathcal{F}_s^{(0)} \times \mathcal{F}_s^{(1)}, \\ \mathbb{P}(d\omega^{(0)}, d\omega^{(1)}) = \mathbb{P}^{(0)}(d\omega^{(0)})\, \mathbb{Q}(\omega^{(0)}, d\omega^{(1)}). \end{array}\right\} \tag{2.2}$$

Any variable or process which is defined on either $\Omega^{(0)}$ or $\Omega^{(1)}$ can be considered in the usual way as a variable or a process on Ω. By standard properties of extensions of spaces, W is a Wiener process on $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t\geq 0}, \mathbb{P})$, and Eq. (1.1) holds on this extended space as well. The (discretely observed) process is thus

$$Z_t = X_t + Y_t. \tag{2.3}$$

The specific assumptions about Y_t are as follows:

Assumption (N). The process

$$\gamma_t(\omega^{(0)}) = \int y^2\, Q_t(\omega^{(0)}, dy) \tag{2.4}$$

is càdlàg (necessarily $(\mathcal{F}_t^{(0)})$-adapted), and $t \mapsto \int y^8\, Q_t(\omega^{(0)}, dy)$ is locally bounded.

Example 2.1. Letting ε_t be a white noise (that is, the variables ε_t are i.i.d., centered) independent of X, this assumption is satisfied when the variables ε_t have a finite 8th moment, with $Y_t = \varepsilon_t$.

The white noise does not accurately describe the microstructure noise, for example it does not account for the fact that the observations are rounded and given at multiples of cents, in most markets. However, if 1 cent is the unit and if the white noise has a law which, restricted to any interval $[n, n+1)$ for $n \in \mathbb{Z}$, is uniform (in addition to being centered with finite 8th moment), then if the observations are $Z_i^n = [X_i^n + \varepsilon_{i\Delta_n}]$ (where $[x]$ denotes the integer part of $x \in \mathbb{R}$), Assumption (N) is satisfied.

It is also satisfied when $Z_i^n = [X_i^n + v_{i\Delta_n}\varepsilon_{i\Delta_n}]$, for any integer-valued and càdlàg process v_t, adapted to the filtration $(\mathcal{F}_t^{(0)})$. This allows for a lot of flexibility in modeling the noise, and in particular permits a rather strong dependency of the noise $Y_i^n = Z_i^n - X_i^n$ upon the underlying process X. □

2.2. Global pre-averaging

We now describe the pre-averaging method, introduced in [29,18], and for which we need some notation. For any two bounded functions f, h on $\mathbb{R}$ with support in $[0, 1]$ and any integer $k \geq 2$, we set

$$\begin{aligned}
&\phi(f, h|t) = \int f(s-t)\, h(s)\, ds \qquad \phi(f) = \phi(f, f|0) = \int_0^1 f(t)^2\, dt \\
&\Phi(f, h) = \int_0^1 \phi(f, f|t)\, \phi(h, h|t)\, dt \\
&\phi_k(f) = \sum_{i=1}^k f\Big(\frac{i}{k}\Big)^2, \qquad \phi_k'(f) = \sum_{i=1}^k \Big(f\Big(\frac{i}{k}\Big) - f\Big(\frac{i-1}{k}\Big)\Big)^2.
\end{aligned} \tag{2.5}$$

Next, we choose a weight function g on $\mathbb{R}$, which satisfies

$$\left.\begin{aligned}
&g \text{ is continuous, piecewise } C^1 \text{ with a piecewise Lipschitz derivative } g', \\
&g(x) = 0 \quad \text{if } x \notin (0,1), \qquad \int_0^1 g(s)^2 ds > 0.
\end{aligned}\right\} \tag{2.6}$$

Observe that $\sup_{k\geq 2}\big(|\phi_k(g) - k\phi(g)| + |k^2\phi_k'(g) - k\phi(g')|\big) < \infty$. We also set, for any process V, and omitting the dependency on g,

$$\begin{aligned}
&\Delta_i^n V = V_{i\Delta_n} - V_{(i-1)\Delta_n}, \qquad \overline{V}(k)_i^n = \sum_{j=1}^{k-1} g\Big(\frac{j}{k}\Big)\, \Delta_{i+j}^n V, \\
&\widehat{V}(k)_i^n = \sum_{j=1}^k \Big(g\Big(\frac{j}{k}\Big) - g\Big(\frac{j-1}{k}\Big)\Big)^2 (\Delta_{i+j-1}^n V)^2.
\end{aligned} \tag{2.7}$$

The pre-averaging estimators of the integrated volatility C_t, are as follows, where k_n is a sequence of integers going to infinity:

$$\widehat{C}(\Delta_n,k_n)_t=\frac{1}{\phi_{k_n}(g)}\,\frac{t}{t-k_n\Delta_n}\sum_{i=1}^{[t/\Delta_n]-k_n+1}\Big((\overline{Z}(k_n)_i^n)^2-\frac{1}{2}\,\widehat{Z}(k_n)_i^n\Big), \tag{2.8}$$

and we recall the main theorem of [18]:

Theorem 2.2. *Assume (H) and (N), and take a sequence of integers k_n satisfying*

$$k_n\sqrt{\Delta_n}=\theta+\mathrm{o}(\Delta_n^{1/4}), \tag{2.9}$$

for some number $\theta\in(0,\infty)$. Then for each $T>0$ the variables

$$\frac{1}{\Delta_n^{1/4}}\Big(\widehat{C}(\Delta_n,k_n)_T-C_T\Big) \tag{2.10}$$

converge stably in law to a variable $\mathcal{U}_T$ which is defined on an extension $(\widetilde{\Omega},\widetilde{\mathcal{F}},(\widetilde{\mathcal{F}}_t)_{t\geq 0},\widetilde{\mathbb{P}})$ of the space $(\Omega,\mathcal{F},(\mathcal{F}_t)_{t\geq 0},\mathbb{P})$ and is, conditionally on $\mathcal{F}$, a centered Gaussian variable with variance given by

$$\begin{aligned} V(g)_T&=\widetilde{E}\Big((\mathcal{U}_T)^2\mid\mathcal{F}\Big)\\ &=\frac{4}{\theta^3\phi(g)^2}\int_0^T\Big(\Phi(g,g)\theta^4c_s^2+2\Phi(g,g')\theta^2c_s\gamma_s+\Phi(g',g')\gamma_s^2\Big)\,ds. \end{aligned}\tag{2.11}$$

Remark 2.3. The first version of the estimator given in [18] differs from (2.8) in three points: first, here we use $\phi_{k_n}(g)$ instead of $k_n\phi(g)$; second we multiply by the correcting factor $\frac{t}{t-k_n\Delta_n}$; third we use $\widehat{Z}(k_n)_i^n$ instead of $\frac{\phi'(g)}{k_n}(\Delta_i^nZ)^2$. The first two modifications were already pointed out in [18] and are important for small samples but asymptotically innocuous. The third modification is also without consequences, asymptotically, and it allows us for a more unified treatment concerning the estimation of the variance below, see [19], and also [20].

Let us also recall that one can make this Central Limit Theorem "feasible" (due to the stable convergence in law) by supplying consistent estimators for the asymptotic (conditional) variance in (2.11). Namely, we are looking for statistics $\widehat{V}_t^n$ such that $\frac{1}{\sqrt{\Delta_n}}\widehat{V}_t^n\stackrel{\mathbb{P}}{\longrightarrow}V(g)_t$, so that $(\widehat{C}(\Delta_n,k_n)_t-C_t)/\sqrt{\widehat{V}_t^n}$ converges in law to $\mathcal{N}(0,1)$. A possible choice for $\widehat{V}^n$ is as follows:

$$\begin{aligned} \widehat{V}_n(\Delta_n,k_n)_t=\frac{1}{\phi_{k_n}(g)^2}\,\frac{t}{t-k_n\Delta_n}\sum_{i=1}^{[t/\Delta_n]-k_n+1}&\bigg(\frac{4\Phi_{k_n}(g)}{3\phi_{k_n}(g)^2}\,(\overline{Z}(k_n)_i^n)^4\\ &+4\bigg(\frac{\Phi'_{k_n}(g)}{\phi_{k_n}(g)\phi'_{k_n}(g)}-\frac{\Phi_{k_n}(g)}{\phi_{k_n}(g)^2}\bigg)(\overline{Z}(k_n)_i^n)^2\,\widehat{Z}(k_n)_i^n\\ &+\bigg(\frac{\Phi_{k_n}(g)}{\phi_{k_n}(g)^2}-\frac{2\Phi'_{k_n}(g)}{\phi_{k_n}(g)\,\phi'_{k_n}(g)}+\frac{\Phi''_{k_n}(g)}{\phi'_{k_n}(g)^2}\bigg)(\widehat{Z}(k_n)_i^n)^2\bigg), \end{aligned}\tag{2.12}$$

where

$$\Phi_k(g) = \sum_{i,j,l=0}^{k} g\Big(\frac{j}{k}\Big) g\Big(\frac{j-i}{k}\Big) g\Big(\frac{l}{k}\Big) g\Big(\frac{l-i}{k}\Big)$$

$$\Phi_k'(g) = \sum_{i,j,l=0}^{k} g\Big(\frac{j}{k}\Big) g\Big(\frac{j-i}{k}\Big) \Big(g\Big(\frac{l}{k}\Big) - g\Big(\frac{l-1}{k}\Big)\Big) \Big(g\Big(\frac{l-i}{k}\Big) - g\Big(\frac{l-1-i}{k}\Big)\Big)$$

$$\Phi_k''(g) = \sum_{i,j,l=0}^{k} \Big(g\Big(\frac{j}{k}\Big) - g\Big(\frac{j-1}{k}\Big)\Big) \Big(g\Big(\frac{j-i}{k}\Big) - g\Big(\frac{j-1-i}{k}\Big)\Big) \times \Big(g\Big(\frac{l}{k}\Big) - g\Big(\frac{l-1}{k}\Big)\Big) \Big(g\Big(\frac{l-i}{k}\Big) - g\Big(\frac{l-1-i}{k}\Big)\Big).$$

Example 2.4 (*The Triangular Kernel*)**.** The simplest weight function is $g(s) = 2(s \wedge (1-s))$ for $s \in [0,1]$, for which

$$\phi(g) = \frac{1}{3}, \qquad \phi(g') = 4, \qquad \Phi(g,g) = \frac{151}{5040},$$
$$\Phi(g,g') = \frac{1}{6}, \qquad \Phi(g',g') = \frac{8}{3}.$$

When $k = 2k'$ is even, we also have

$$\overline{Z}(k)_i^n = \frac{1}{k'}(Z_{i+k'-1}^n + \cdots + Z_{i+2k'-2}^n) - \frac{1}{k'}(Z_{i-1}^n + \cdots + Z_{i+k'-2}^n),$$

which is the difference between two successive (non-overlapping) averages of k' values of Z_i^n. □

3. Pre-averaging and optimality

3.1. General considerations

A general approach to efficiency for estimators of the integrated volatility is so far out of reach. In the case of noisy observations, the only situation which so far is really well understood is the case of a constant volatility $c_t(\omega) = c$ and when the noise is a Gaussian white noise, with variance γ (same notation as in (2.4), but here γ is a constant). In this case, according to [16, 17] one has the following: the LAN (local asymptotic normality of the likelihood) for estimating $C_T = Tc$; moreover the efficient estimators $\widehat{C}_T^n$ for C_T converge with the rate $1/\Delta_n^{1/4}$, and the asymptotic variance of $\frac{1}{\Delta_n^{1/4}}(\widehat{C}_T^n - C_T)$ is

$$V_T^{opt} = 8T\, c^{3/2}\, \gamma^{1/2}. \tag{3.1}$$

This asymptotic variance is also, naturally, the inverse of the Fisher information for estimating the parameter Tc.

Remark 3.1. (1) In these papers a more general parametric situation is considered, in which c_t may be of the form $c_t = c(\theta, t, X_t)$, but only the estimation of the parameter θ is studied. When $c_t = c$ is a constant, estimating the parameter $\theta = Tc$ is of course the same as estimating the integrated volatility.

(2) In the recent paper [11], the authors consider a volatility of the form $c_t = c(X_t, Y_t)$, where Y_t is a process driven by another Brownian motion, independent of W. They prove a Hájek's type convolution theorem which leads to a good notion of efficiency, but only in the non-noisy case. Their result can probably be extended to the case where there is an observation noise which is white and independent of X.

In the case where there is no microstructure noise, shows the validity of the lower bound, in the case where the process σ_t is non random, though observation times are allowed to be irregular.

(3) Recent work by [1], building on [30,8], extends the optimality results to time varying and random σ which does not have to be independent of the Brownian motion (W_t). In this case, the lower bound for the asymptotic variance is

$$8\int_0^T c_t^{3/2}\,\gamma_t^{1/2} dt. \tag{3.2}$$

It is widely conjectured that this is the general lower bound when microstructure is normal, and the "quasi-lower bound" when inference is done without seeking to estimate the distribution of the noise. – The main constraint in [1] is that the noise needs to be independent and additive, which is more restrictive than the Assumption (N) in the current (and earlier pre-averaging) paper(s). – When there is no microstructure noise, see [22] for a treatment of the general case. □

3.2. Behavior of global pre-averaging under constant coefficients

In order to examine the optimality properties of the global pre-averaging procedure, we first consider the already mentioned case where $c_t = c$ and $\gamma_t = \gamma$ are constants. The optimal asymptotic variance is V_T^{opt}, given by (3.1). For the pre-averaging estimators, we can write the following:

For a given kernel g, we may choose θ in an optimal way. Indeed, we have

$$V(g)_T = \frac{4T}{\phi(g)^2}\Big(\Phi(g,g)\,c^2\,\theta + 2\Phi(g,g')\,\frac{c\,\gamma}{\theta} + \Phi(g',g')\,\frac{\gamma^2}{\theta^3}\Big). \tag{3.3}$$

The optimal choice of θ is then

$$\theta = \left(\frac{c}{\gamma}\,\frac{\sqrt{\Phi(g,g')^2 + 3\Phi(g,g)\,\Phi(g',g')} - \Phi(g,g')}{3\,\Phi(g',g')}\right)^{-1/2}, \tag{3.4}$$

which leads to $V(g)_T = \alpha(g) V_T^{opt}$, where

$$\alpha(g) = \frac{6\Phi(g,g)\,\Phi(g',g') + 2\Phi(g,g')\sqrt{\Phi(g,g')^2 + 3\Phi(g,g)\,\Phi(g',g')} - 2\Phi(g,g')^2}{3^{3/2}\phi(g)^2\,\Phi(g',g')^{1/2}\Big(\sqrt{\Phi(g,g')^2 + 3\Phi(g,g)\,\Phi(g',g')} - \Phi(g,g')\Big)^{1/2}}. \tag{3.5}$$

As for the choice of the weight function g, it amounts to minimizing the number $\alpha(g)$ given above, and which by necessity cannot be smaller than 1. In the case of the triangular kernel described above, and for the sine function, one easily sees that

$$\alpha(g) \simeq \begin{cases} 1.0666 & \text{for the triangular kernel} \\ 1.0721 & \text{for the sine weight function } g(x) = \sin(\pi x)\, 1_{[0,1]}(x), \end{cases} \tag{3.6}$$

which are both very close indeed to the minimal value 1.

Notice that the above holds also when $c_t(\omega) = c(\omega)$ and $\gamma_t(\omega) = \gamma(\omega)$ depend on ω, but not on t, in which case the optimal value of θ also depends on ω.

However, in general c_t is *not* constant in time, and probably neither is γ_t. It means that the choice (3.4) is not feasible in practice (and, for that matter, even in the constant case the numbers c and γ are unknown). To overcome this difficulty, one may use a two-steps estimation procedure, as briefly described in Remark 3 of [18]: namely, we first choose $\theta = \theta_1$ arbitrarily (in a reasonable way, whatever this might mean) and get a first estimate $\widehat{C}_t^{n,1}$ for C_t, and also take

$$\widehat{\Gamma}_t^n = \frac{\Delta_n}{2} \sum_{i=1}^{[t/\Delta_n]} (\Delta_i^n Z)^2, \tag{3.7}$$

which is easily seen to converge in probability to $\Gamma_t = \int_0^t \gamma_s \, ds$. Then, if the time horizon is T, we set $c^{average} = \widehat{C}_T^{n,1}/T$ and $\gamma^{average} = \widehat{\Gamma}_T^n/T$ and we take θ as given by (3.4), with c and γ substituted with $c^{average}$ and $\gamma^{average}$, respectively.

This procedure gives a sensible choice for θ, but still very far in general from the optimal one. The optimal method would be to have a parameter $\theta = \theta_t(\omega)$ depending on t and ω, in such a way that (3.4) holds with θ_t, c_t, γ_t for each value of t, that is

$$\theta_t = \sqrt{\gamma_t / \Psi(g)\, c_t}, \quad \text{where } \Psi(g) = \frac{\sqrt{\Phi(g,g')^2 + 3\Phi(g,g)\Phi(g',g')} - \Phi(g,g')}{3\Phi(g',g')}, \tag{3.8}$$

and assuming of course that c_t never vanishes. This is of course infeasible in practice, but this paper is devoted to trying to mimic, in a feasible way, this optimal procedure. This is the aim of the next section.

4. Adaptive estimation

For constructing adaptive estimators, the key step is to estimate the right side of (3.8) for each t, or equivalently estimate c_t and γ_t. That is, we need *local estimators* for the volatility c_t and the (conditional) noise variance γ_t. Obviously, we need that these two processes do not vary too much as a function of time, and the càdlàg property is not enough. So we assume the following:

Assumption (L). The two processes c_t and γ_t satisfy for all $q \geq 2$ and $t, v \geq 0$ and $n \in \mathbb{N}$:

$$\mathbb{E}\Big(\sup_{s\in[0,t]} (|c_{(v+s)\wedge\tau_n} - c_{v\wedge\tau_n}|^q + |\gamma_{(v+s)\wedge\tau_n} - \gamma_{v\wedge\tau_n}|^q)\Big) \leq K_{q,n} t \tag{4.1}$$

for some constants $K_{q,n}$, where τ_n is a sequence of stopping time increasing to ∞. Furthermore, the processes $1/c_t$ and $1/\gamma_t$ are locally bounded (hence c_t and γ_t never vanish). □

This assumption does *not* imply that c_t and γ_t are continuous, but it implies that these processes have no fixed times of discontinuity. It is satisfied, for example, when the processes c_t and γ_t are Itô semimartingales with locally bounded characteristics, even with jumps. It is also satisfied when they have paths which are Hölder continuous with index not smaller than $\frac{1}{2}$.

We now can describe the adaptive procedure. We fix the time horizon, say T. Estimating c_t and γ_t and deriving an "optimal" θ_t at each time t would give rise to a huge instability. Instead, we split the data into L_n blocks of size l_n, with $T(n, r-1)$ and $T(n, r)$ denoting the calendar times at which the rth block starts and ends, that is:

$$L_n = \Big[\frac{T}{l_n\Delta_n}\Big], \qquad T(n,r) = \begin{cases} rl_n\Delta_n & \text{if } r < L_n \\ T & \text{if } r = L_n \end{cases} \tag{4.2}$$

(the last block has a size between l_n and $2l_n - 1$). Then we do as follows, at each stage n:

- find preliminary estimators $\widehat{c}(r)^n$ and $\widehat{\gamma}(r)^n$ to be used as proxies for c_t and γ_t within the rth block, as if these processes were constant in time inside this block;
- choose θ and thus the "optimal" number $k_n = k_{n,r}$ for the rth block, according to (2.9) and (3.8), and on the basis of the previous estimators;
- use pre-averaging with $k_{n,r}$ to estimate the integrated volatility over the rth block, and then add up to find an estimator for C_T;

Moreover, to avoid strong dependence between the estimators $\widehat{c}(r)^n$, $\widehat{\gamma}(r)^n$ and the pre-averaging estimator on block r, for deriving $\widehat{c}(r)^n$ and $\widehat{\gamma}(r)^n$ we only use observations occurring within the previous $(r-1)$th block. So adaptation can only been performed for blocks $2, 3, \ldots$. For the first block, and also for deriving the preliminary estimators $\widehat{c}(r)^n$, we use pre-averaging with an arbitrarily chosen sequence k_n.

More specifically, we first choose a number $a \in (0,1)$ and two sequences of integers, subject to the conditions:

$$l_n \asymp \frac{1}{\Delta_n^w} \quad \text{with } w \in \Big(\frac{5}{6}, 1\Big), \qquad k_n \asymp \frac{1}{\Delta_n^{1/2}}, \quad 2 \le k_n < (1-a)l_n,\ l_n < T/\Delta_n \tag{4.3}$$

(the notation $u_n \asymp v_n$ means that both u_n/v_n and v_n/u_n are bounded; of course, the last requirements are implied for all n large by the two first ones). Then we have two steps:

(1) First step: local estimation of c_t and γ_t. We set $l'_n = [al_n]$, which satisfies $l'_n + k_n \le l_n$. The preliminary estimators, when $r \ge 2$, are as follows:

$$\begin{aligned} \widehat{c}(r)^n &= \frac{1}{l'_n\Delta_n\,\phi_{k_n}(g)} \sum_{i=(r-1)l_n-l'_n-k_n+1}^{(r-1)l_n-k_n} \Big((\overline{Z}(k_n)_i^n)^2 - \frac{1}{2}\,\widehat{Z}(k_n)_i^n\Big) \\ \widehat{\gamma}(r)^n &= \frac{1}{2l'_n} \sum_{j=(r-1)l_n-l'_n+1}^{(r-1)l_n} (\Delta_j^n Z)^2. \end{aligned} \tag{4.4}$$

As a matter of fact, these are natural estimators for the following averages:

$$\overline{c}(r)^n = \frac{1}{l'_n\Delta_n}\int_{T(n,r-1)-l'_n\Delta_n}^{T(n,r)} c_s\,ds, \qquad \overline{\gamma}(r)^n = \frac{1}{l'_n\Delta_n}\int_{T(n,r-1)-l'_n\Delta_n}^{T(n,r)} \gamma_s\,ds, \tag{4.5}$$

rather than for the average values of c_t and γ_t within the rth block. They depend on a (through l'_n), and the estimators are more precise when a increases (because they use more data), whereas the target ("parameter") quantities move farther away from the quantities of interest. For this reason, we advocate to choose an intermediate value for a, such as $a = \frac{1}{2}$. This is similar to the tradeoff in [26].

We will see that, because of (L),

$$\widehat{c}(r)^n - \overline{c}(r)^n = \mathrm{O}_{Pu}(\Delta_n^{(2w-1)/4}), \qquad \widehat{\gamma}(r)^n - \overline{\gamma}(r)^n = \mathrm{O}_{Pu}(\Delta_n^{w/2}), \tag{4.6}$$

where $U_i^n = \mathrm{O}_{Pu}(v_n)$ means that the variables U_i^n/v_n are bounded in probability, uniformly in i. Therefore, with the notation (3.8) and $\overline{\theta}(r)^n = \sqrt{\overline{\gamma}(r)^n/\Psi(g)\,\overline{c}(r)^n}$, we have

$$\widehat{\theta}(r)^n - \overline{\theta}(r)^n = \mathrm{O}_{Pu}(\Delta_n^{(2w-1)/4}),$$
$$\text{where } \widehat{\theta}(r)^n = \begin{cases} \sqrt{\widehat{\gamma}(r)^n/\Psi(g)\,\widehat{c}(r)^n} & \text{if } \widehat{c}(r)^n > 0 \\ 1 & \text{if } \widehat{c}(r)^n \leq 0 \end{cases} \tag{4.7}$$

(the probability that $\widehat{c}_r^n \leq 0$ tends to 0 as $n \to \infty$, but it is usually positive for any finite n, and 1 above is a dummy value).

(2) Second step: global estimation. For $r = 1, \ldots, L_n$ we set

$$k_{n,r} = \begin{cases} k_n & \text{if } r = 1 \\ \left[2 \bigvee \left(\frac{l_n}{2} \bigwedge \frac{\widehat{\theta}(r)^n}{\sqrt{\Delta_n}}\right)\right] & \text{if } r \geq 2 \end{cases} \tag{4.8}$$

(taking the infimum with $l_n/2$ is for mathematical correctness, but we will see that, asymptotically, all $\widehat{\theta}(r)^n/\sqrt{\Delta_n}$ are smaller than $l_n/2$; taking the supremum with 2 is likewise asymptotically irrelevant, although for small samples, and if the noise is very small, it may be effective and it would probably be more appropriate to take the supremum with a quite bigger value, such as 10 or 20).

In (4.8) the value $k_{n,1} = k_n$ is as in (4.3) and we do no adaptation within the first block. The other values $k_{n,r}$ are random, but known to the statistician, and our final estimator is

$$\widehat{C}_T^n = \sum_{j=1}^{L_n} \frac{a(n,j)}{\phi_{k_{n,j}}(g)} \sum_{i=J(n,j-1)+1}^{J(n,j)} \left((\overline{Z}(k_{n,j})_i^n)^2 - \frac{1}{2}\,\widehat{Z}(k_{n,j})_i^n \right)$$
$$\text{where } J(n,0) = 0 \quad \text{and} \tag{4.9}$$
$$1 \leq j < L_n \ \Rightarrow\ J(n,j) = j l_n, \qquad a(n,j) = 1;$$
$$j = L_n \ \Rightarrow\ J(n,j) = [T/\Delta_n] - k_{n,L_n} + 1,\ \ a(n,j) = \frac{J(n,L_n) - J(n,L_n-1) + k_{n,L_n}}{J(n,L_n) - J(n,L_n-1)}.$$

The factor $a(n, L_n)$, which is bigger than 1, plays exactly the same role as the factor $\frac{t}{t - k_n\Delta_n}$ in (2.8). We also need estimators for the (conditional) estimation variance below, and to this effect we set

$$\begin{aligned} \widehat{V}_T^n = \sum_{j=1}^{L_n} \frac{a(n,j)}{\phi_{k_{n,j}}(g)^2} \sum_{i=J(n,j-1)+1}^{J(n,j)} \Bigg(& \frac{4\Phi_{k_{n,j}}(g)}{3\phi_{k_{n,j}}(g)^2}\,(\overline{Z}(k_{n,j})_i^n)^4 \\ & + 4\left(\frac{\Phi'_{k_{n,j}}(g)}{\phi_{k_{n,j}}(g)\phi'_{k_{n,j}}(g)} - \frac{\Phi_{k_{n,j}}(g)}{\phi_{k_{n,j}}(g)^2} \right) (\overline{Z}(k_{n,j})_i^n)^2\,\widehat{Z}(k_{n,j})_i^n \\ & + \left(\frac{\Phi_{k_{n,j}}(g)}{\phi_{k_{n,j}}(g)^2} - \frac{2\Phi'_{k_{n,j}}(g)}{\phi_{k_{n,j}}(g)\,\phi'_{k_{n,j}}(g)} + \frac{\Phi''_{k_{n,j}}(g)}{\phi'_{k_{n,j}}(g)^2} \right) (\widehat{Z}(k_{n,j})_i^n)^2 \Bigg). \end{aligned} \tag{4.10}$$

The main result of this paper is the following:

Theorem 4.1. *Assume (H), (N) and (L). Then, for each $T > 0$, the variables*

$$\frac{1}{\Delta_n^{1/4}}\left(\widehat{C}_T^n - C_T\right) \tag{4.11}$$

converge stably in law to a limiting variable $\mathcal{U}'_T$, where the process $\mathcal{U}$ has the same description as in Theorem 2.2, except that its $\mathcal{F}$-conditional variance is

$$V'(g)_T = \widetilde{E}\big((\mathcal{U}'_T)^2 \mid \mathcal{F}\big) = 8\,\alpha(g) \int_0^T \sqrt{c_s^3\,\gamma_s}\,ds, \tag{4.12}$$

where $\alpha(g)$ is given by (3.5).

Moreover, we have $\frac{1}{\sqrt{\Delta_n}}\,\widehat{V}_T^n \xrightarrow{\mathbb{P}} V'(g)_T$, and thus the variables $\frac{1}{\sqrt{\widehat{V}_T^n}}\,(\widehat{C}_T^n - C_T)$ (stably) converge in law to a standard normal variable.

Remark 4.2. For simplicity, we have taken the same weight function g for pre-averaging in the two steps, but this is of course not necessary. □

Remark 4.3. One needs $c_t > 0$ for the estimated value of θ_t to be finite, and $\gamma_t > 0$ for it to be positive. When the process γ is allowed to vanish, and does indeed vanish on some interval, at these places $\widehat{\theta}_j^n$ tends to be small and $k_{n,j}$ becomes equal to 2. As easily checked, and for the triangular kernel g, it follows that this interval contributes 0 in the sum defining $\widehat{C}_T^n$, which precludes even the consistency of these estimators. For another kernel g, or if we were taking the supremum with another number than 2 in (4.8), the contribution would be possibly not 0 but still not consistent estimators for the integrated volatility on this interval. □

Remark 4.4 (*Irregular Spacings*)**.** Our results extend to irregularly spaced times $t_{n,i}$ of the form $t_{n,i} = F_{iT/n}$, where F is a function with $F(0) = 0$ and $F(T) = T$, and having a derivative F' which is Lipschitz, positive and bounded away from 0. More generally, F may be random, provided the lower bound for F' and its Lipschitz constant are uniform in ω, plus the fact that each F_t is a stopping time. However, unless we take F to be independent of (X, Y), the latter requirements plus the property $F(T) = T$ may be restrictive.

This type of assumption has been used by [6,8], and a similar but slightly more general assumption was used by [34]. Cf. also the contiguity result in Chapter 2.7.1 of [28].[2]

The observations then become $Z_{t_{n,i}} = Z_{F(iT/n)} = \widetilde{Z}_{F(iT/n)}$, where $\widetilde{Z} = \widetilde{X} + \widetilde{Y}$ and $\widetilde{X}_t = X_{F(t)}$ and $\widetilde{Y}_t = Y_{F(t)}$. In view of the assumptions on F, one easily checks that $\widetilde{X}$ and $\widetilde{Y}$ satisfy (H), (N) and (L), with the various coefficients in (1.1) and (2.4) and the new volatility being

$$\widetilde{b}_t = b_{F(t)}F'(t), \qquad \widetilde{\sigma}_t = \sigma_{F(t)}\sqrt{F'(t)}, \qquad \widetilde{c}_t = c_{F(t)}F'(t), \qquad \widetilde{\gamma}_t = \gamma_{F(t)}F'(t),$$
$$\widetilde{C}_t = C_{F(t)},$$

and in particular $\widetilde{C}_T = C_T$.

Now, if in all the previous estimators (pre-averaged estimator, adaptive estimator, estimator for the variance) we replace $i\Delta_n$ by $t_{n,i}$, we obtain the corresponding estimators for the process $\widetilde{X}$ and its noisy version $\widetilde{Z}$ and with $1/T$ instead of Δ_n. Since $\widetilde{C}_T = C_T$, Theorem 4.1 thus remains valid under this form of sampling irregularity. Note that the asymptotic variance (4.12) becomes $8\alpha(g)\int_0^T \sqrt{\widetilde{c}_s^3\,\widetilde{\gamma}_s}\,ds$, and by a change of variable this is indeed equal to $V'(g)_T$.

[2] For more general models of irregular times, see, for example, [13–15,24,27,28,31], and other work by the same authors.

5. Optimality properties of the adaptive procedure

As we have seen in (4.12), the asymptotic variance $V'(g)_t = 8\,\alpha(g)\int_0^T \sqrt{c_s^3\,\gamma_s}\,ds$, where $\alpha(g)$ is given by (3.5).

For a less explicit but perhaps more intuitive expression, set $\zeta = \theta(c/\gamma)^{1/2}$ in Eqs. (3.3)–(3.4) to obtain that

$$\alpha(g) = \frac{1}{2\phi(g)^2}\,\min_{\zeta}\Big\{\Big(\Phi(g,g)\zeta + 2\Phi(g,g')\frac{1}{\zeta} + \Phi(g',g')\frac{1}{\zeta^3}\Big)\Big\}. \tag{5.1}$$

Compared to global pre-averaging, this means that $V'(g)_t$ in the adaptive procedure behaves as if for all t and for any fixed function g, one obtains the minimum of

$$V(g)_t = \frac{4}{\phi(g)^2}\int_0^T \Big(\Phi(g,g)\theta_s c_s^2 + 2\Phi(g,g')\frac{c_s\gamma_s}{\theta_s} + \Phi(g',g')\frac{\gamma_s^2}{\theta_s^3}\Big)\,ds$$

which extends (2.11) by allowing $\theta = \theta_t$ to depend on t. We thus achieve the lower bound of all (conditional) variances which are possible when using a pre-averaging approach, for a given weight function g.

Remark 5.1 (*Connection to Kernel Estimation*)**.** Our asymptotic results can be written in terms of the kernel function k from [6]. In the notation of the current paper, for given weight function g the corresponding kernel k has the form $k(t) = \phi(g, g|t)/\phi(g)$ for $t \geq 0$. Hence also for derivatives, $k^{(\nu)}(t) = (-1)^\nu \phi(g^{(\nu)}, g|t)/\phi(g)$ while $k''(t) = -\phi(g', g'|t)/\phi(g)$. In the notation of [6] (Eq. (9) on p. 1488, and Table II on p. 1495), we thus have

$$k_\bullet^{0,0} = \frac{1}{\phi(g)^2}\,\Phi(g,g), \qquad k_\bullet^{1,1} = \frac{1}{\phi(g)^2}\,\Phi(g',g), \qquad k_\bullet^{2,2} = \frac{1}{\phi(g)^2}\,\Phi(g',g'). \tag{5.2}$$

Explicit expressions only involving (5.2) can then be obtained from Eqs. (3.5) or (5.1), and hence only depend on these, in particular,

$$\alpha(g) = \frac{6k_\bullet^{0,0}k_\bullet^{2,2} + 2k_\bullet^{1,1}\sqrt{(k_\bullet^{1,1})^2 + 3k_\bullet^{0,0}k_\bullet^{2,2}} - 2(k_\bullet^{1,1})^2}{3^{3/2}\,(k_\bullet^{2,2})^{1/2}\Big(\sqrt{(k_\bullet^{1,1})^2 + 3k_\bullet^{0,0}k_\bullet^{2,2}} - k_\bullet^{1,1}\Big)^{1/2}}.$$

The same is true in the case of global pre-averaging, cf. (2.11). – Further comparison to multi-scale realized volatility (MSRV) [34] can be derived as in [9]. We emphasize that pre-averaging, kernel estimation, and MSRV do not yield the same exact estimator, due to edge effects. See also Remark 1 (p. 2255) in [18].

Remark 5.2. As already said, the adaptive procedure does not fully achieve efficiency, because the limiting variance is $\alpha(g)$ times the efficient lower bound. Both the triangular and the sine weight functions give us $\alpha(g) \simeq 1.07$, which is so close to 1 that the difference is quite possibly of the same magnitude as the difference between finite-sample and asymptotic properties. However, one could try to find weight functions with $\alpha(g)$ closer to 1, or even equal to 1. So far, such g's are not available.

If, instead, we use an adaptive realized kernel estimation, and provided such an adaptive method works (which remains to be proved, since the “border effects” with this method are already slightly worrisome without adaptation and thus likely to become harder to deal with in the adaptive case), we could use kernels k such that the associated $\alpha(g)$ (or rather $\alpha(k)$ is arbitrarily close to 1: using the Tukey–Hanning kernel of order 16 would for example lead to the value 1.003, see [6]). However, even in this case it is unknown whether a kernel with bounded support can achieve optimality.[3]

On the other hand, the method proposed in [30] and extended by [1] is automatically adaptive, but so far it works for independent and additive noise only, which is more restrictive than our Assumption (N) and not very realistic in practice, since at very high frequency prices have the very distinct feature of being rounded; see the discussion in Example 2.1.

In order to improve efficiency, one could use linear combinations of the previous estimators, for a family $\mathcal{G} = (g_j)_{1\le j\le p}$ of different weight functions.

Writing $\widehat{C}(g_j; \Delta_n, k_n)_t$ for the estimator (2.8) if we use the weight function g_j, and with $\mathcal{A} = (a_j)_{1\le j\le p}$ a family of weights satisfying $\sum_{j=1}^p a_j = 1$, one sets

$$\widehat{C}(\mathcal{A}, \mathcal{G}; , \Delta_n, k_n)_t = \sum_{j=1}^{p} a_j \widehat{C}(g_j; \Delta_n, k_n)_t. \tag{5.3}$$

In view of the results of Chapter 16 of [20], one can show a result analogous to Theorem 2.2, with an asymptotic variance of the form

$$V(\mathcal{A}, \mathcal{G})_T = \int_0^T \left(\Phi(\mathcal{A}, \mathcal{G})\theta^4 c_s^2 + 2\Phi'(\mathcal{A}, \mathcal{G})\theta^2 c_s\gamma_s + \Phi''(\mathcal{A}, \mathcal{G})\gamma_s^2\right) ds \tag{5.4}$$

for (rather complicated) coefficients $\Phi(\mathcal{A}, \mathcal{G})$, $\Phi'(\mathcal{A}, \mathcal{G})$ and $\Phi''(\mathcal{A}, \mathcal{G})$.

This is all very good in the case of constant coefficients $c_t = c$ and $\gamma_t = \gamma$, as in Section 3.2. Indeed, in this case, for any choice of the family $\mathcal{A}$ of weights one can choose $\theta = \theta(\mathcal{A})$ optimally as in (3.4), and then choose $\mathcal{A}$ itself in order to minimize the asymptotic variance. We end up with a choice of $\mathcal{A} = \mathcal{A}(\mathcal{G}; c, \gamma)$, hence of $\theta = \theta(\mathcal{G}; c, \gamma)$ as well, which depends on the family $\mathcal{G}$ and on the constant volatility c and the constant noise variance γ. The asymptotic variance has then the form $V(\mathcal{G})_T = \alpha'(\mathcal{G})V_t^{opt}$ for some number $\alpha'(\mathcal{G}) \ge 1$, and obviously $\alpha'(\mathcal{G}) \le \alpha(g_j)$ for any j.

The situation become more complicated in the non-constant case. If we mimic the previous adaptive procedure, we see that we have to optimize the choice of the family $\mathcal{A}$ of weights within each "big block". However, mathematically speaking this is not be a problem, and one can prove a result analogous to Theorem 4.1, with $\alpha'(\mathcal{G})$ instead of $\alpha(g)$.

So far, the efficiency is improved by this procedure, but the implementation becomes significantly more complicated, whereas we still have not reached genuine efficiency. To get efficiency in the sense that the asymptotic variance is $8\int_0^T \sqrt{c_s^3\gamma_s}\, ds$, we would need to take a family $\mathcal{G}_n = (g_j)_{1\le j\le p_n}$ with a number p_n of functions tending to infinity as Δ_n goes to 0.

More specifically, let us take the family of sine functions

$$g_i(x) = \sin\left(\frac{\pi x}{i}\right) 1_{[0,1]}(x). \tag{5.5}$$

The $k_n - 1$-dimensional vectors with components $(g_j(i/k_n) : i = 1, \dots, k_n - 1)$ form an orthogonal basis of eigenvectors for the covariance of the family of variables $(\Delta_i^n Z)_{1\le i\le k_n}$, see

[3] A recent paper by [23] finds an optimal weight function with unbounded support. As noted, however, on p. 5 of the paper, this weight function induces substantial edge effects which the author solves by "jittering", and it is thus not clear that this kernel can be used in our adaptive case. A resolution of this question is beyond the scope of the current paper.

for example [12]. Due to this fact, one can see that, formally, the normalized estimation variance (for the adaptive procedure briefly explained above) converges to the efficient one, provided we take $p_n = k_n - 1$.

However, so far there is no fully rigorous proof of this result: writing such a proof, if possible, would necessitate a set of new methods, especially in our setting where c_t is stochastic, as well as the noise variance γ_t. Thus, in view of the complexity of the procedure (with a choice of the weights a_j for each block) and of the very limited variance reduction to which it may lead (at the best dividing the variance by 1.07), we do not pursue the topic here. – Of course, the procedure can be made arbitrarily close to efficient for fixed (sufficiently large) k.

Let us, however, mention that such a method would be in the same spirit as in [10,1], although in the cited papers, the noise is independent and additive.

6. A simulation experiment

The following simulation experiment is carried out in the case where there is no drift ($b \equiv 0$), and the volatility is parabolic in time, of the form

$$\sigma_t = a\sqrt{25\,000}\left(1+\left(\frac{25}{12}t-1\right)^2\right)$$

for t in $(0, 1)$ (hence, with the typical smile shape during a day, the maximal value being approximately 2.2 times the minimum).[4] The volatility is not stochastic, but for the problem at hand this makes little difference, and it allows for easier comparisons between different simulations, since the integrated volatility is always the same. The noise is i.i.d. Gaussian, with three possible standard deviations ($10a$, $20a$, $30a$), whereas the standard deviation of a non-noisy return is between a and $2.2\,a$. Finally, a is chosen such that the integrated volatility over a day is $C_1 = 1$.

There are 25 000 sample points in each simulation, and 2000 simulations. The results are reported in Table 1. For each level of noise, we report:

(1) For 10 different values of k_n the mean (over all simulations) and empirical standard deviation of the estimation error (around the true value, and not around the mean): these correspond to non-adaptive pre-averaging (with the triangular kernel) with a k_n constant over the whole time span.
(2) The adaptive version, as described in the paper, with a number $L_n = 10$ of "big blocks".
(3) The "adaptive optimal" version: this means that within each "big block" the k_n is chosen optimally (this is possible, because one knows the average volatility and noise variance in each block), and then estimation is performed within each block; Of course, this is infeasible in practice.
(4) The "optimal" (or efficient) square-root of the estimation error, for the triangular kernel (given by (4.12) in the paper).

We note the following:

First, the estimation variance for the (infeasible) "adaptive optimal" case is very close to the efficient variance. That probably means that splitting in 10 "big blocks" is nearly optimal for the type of volatility variability used here.

[4] The smile or U-shape for volatility is widely taken as a realistic model. See, for example, [4,25].

Table 1
The table reports means and root mean squared errors (RMSEs) for: ten non-adaptive choices of pre-averaging (as in Section 2), then for the feasible adaptive pre-averaging (as in Section 4), for an infeasible optimal adaptive procedure, and finally the asymptotic value as given in Theorem 4.1, with the triangular kernel. The true value of C_1 is 1.

Value of k_n	Standard deviation of the noise					
	10		20		30	
	Mean of $\hat{C}$	RMSE of $\hat{C}$	Mean of $\hat{C}$	RMSE of $\hat{C}$	Mean of $\hat{C}$	RMSE of $\hat{C}$
20	1.0044	0.0630	0.9810	0.1739	0.9774	0.3556
40	0.9947	0.0543	0.9945	0.0858	0.9991	0.1489
60	0.9957	0.0612	0.9967	0.0752	1.0000	0.1072
80	0.9953	0.0985	0.9969	0.0767	0.9993	0.0956
100	0.9945	0.0754	0.9969	0.0811	0.9983	0.0944
120	0.9936	0.0919	0.9963	0.0864	0.9971	0.0964
140	0.9928	0.0878	0.9955	0.0917	0.9959	0.1001
160	0.9920	0.0935	0.9949	0.0966	0.9947	0.1041
180	0.9912	0.0988	0.9940	0.1013	0.9934	0.1086
200	0.9902	0.1037	0.9932	0.1057	0.9922	0.1129
Adaptive	0.9977	0.0575	1.0016	0.0793	1.0034	0.1009
Adaptive optimal (unfeasible)	0.9940	0.0522	0.9997	0.0734	1.0022	0.0919
Optimal		0.0514		0.0728		0.0891

Second, the true adaptive method is quite good, although (slightly) outperformed by the non-adaptive method for one or two values of k_n, in each situation: these values of k_n are different, according to the noise level. On the other hand, the adaptive method is much better than the non-adaptive one used with the wrong k_n.

7. Conclusion

We have shown that pre-averaging can be made adaptive for estimation of volatility from high frequency data. This reduces the asymptotic variance of the estimator for any given weight function g. The asymptotic variance is of the form $\alpha(g)\times$ the supposed lower bound (3.2), and we show that for the triangular g (which corresponds to flat weights on the observations), $\alpha(g) \simeq 1.07$.

Remaining open questions include whether the bound can be reached for a suitable choice of g, and also what regularity conditions are required for the bound to be valid. Also, there is a *prima facie* case that the bound can be breached when estimation of the noise distribution is involved in inference, and this question also needs to be explored further. Finally, the extension of adaptivity to other high frequency data problems awaits.

Acknowledgments

The second author is grateful to the U.S. National Science Foundation for support under grants SES 11-24526 and DMS 14-07812.

Appendix. The proof

As in [18], and by a localization procedure, we can strengthen the assumption as follows:

Assumption (S-HNL). We have (H) and (N) and (L), and further the processes $b_t, c_t, 1/c_t, \gamma_t, 1/\gamma_t, \int y^8 Q_t(dy)$ and X_t itself are bounded (uniformly in (ω, t)), and in (4.1) we can take $K_{n,q} = K_q$ independent of n (or, equivalently, $\tau_1 = \infty$ identically). $\square$

Below, K denotes a constant which changes from line to line and may depend on the bounds of the various processes in (S-HNL), and is written K_r if it depends on an additional parameter r. If $V(n,x)$ and $V'(n,x)$ are two families of variables indexed by $\mathbb{N} \times \mathcal{X}$, with $\mathcal{X}$ an arbitrary set, we write $V(n,x) = \mathrm{O}_{Pu}(V'(n,x))$ if the variables $|V(n,x)|/|V'(n,x)|$ are bounded in probability, uniformly in n, x. Finally, unless otherwise stated, $p \geq 1$ denotes an integer and $q > 0$ a real.

We write $Y_t = Z_t - X_t$ for the "noise process".

A complete proof of the main result, starting from scratch, would be very long, and below we try instead to rely as much as we can upon the (already long) proof given in [18]: this unfortunately makes what follows difficult to read without constant reference to that paper, but on the other hand it avoids numerous redundancies. Specifically, we refer to equation numbers or lemma numbers of that paper without special mention, but they are written with the prefix "A" always.

As a consequence of this fact, we will prove the result for the following modified version of our estimators (see Remark 2.3):

$$\widehat{C}^n_T = \sum_{r=1}^{L_n} \sum_{i=J(n,r-1)+1}^{J(n,r)} \Big(\frac{1}{k_{n,r}\, \phi(g)} (\overline{Z}(k_{n,r})^n_i)^2 - \frac{\phi(g')}{2k^2_{n,r}\, \phi(g)} (\Delta^n_i Z)^2 \Big) \tag{A.1}$$

(the difference with (2.8) is in fact $\mathrm{O}(\sqrt{\Delta_n}\,)$, so if the CLT holds for one version it also holds for the other). In a similar way, the first estimator in (4.4) can and will be replaced by

$$\widehat{c}(r)^n = \frac{1}{l'_n \Delta_n} \sum_{i=(r-1)l_n - l'_n - k_n + 1}^{(r-1)l_n - k_n} \Big(\frac{1}{k_n\, \phi(g)} (\overline{Z}(k_n)^n_i)^2 - \frac{\phi'(g)}{2k^2_n\, \phi(g)} (\Delta^n_i Z)^2 \Big), \tag{A.2}$$

and the estimator for the conditional variance by

$$\begin{aligned} \widehat{V}^n_T = \sum_{r=1}^{L_n} \Big(& \frac{4\Phi(g,g)}{3k_{n,r}\phi(g)^4} \sum_{i=J(n,r-1)+1}^{J(n,r)} (\overline{Z}(k_{n,r})^n_i)^4 \\ & + \frac{4}{k^3_{n,r}} \Big(\frac{\Phi(g,g')}{\phi(g)^3} - \frac{\Phi(g,g)\phi(g')}{\phi(g)^4} \Big) \sum_{i=J(n,r-1)+1}^{J(n,r)-k_{n,r}} (\overline{Z}^n_i)^2 \sum_{j=i+k_{n,r}}^{i+2k_{n,r}-1} (\Delta^n_j Z)^2 \\ & + \frac{1}{k^3_{n,r}} \Big(\frac{\Phi(g',g')}{\phi(g)^2} - 2\frac{\Phi(g,g')\phi(g')}{\phi(g)^3} + \frac{\Phi(g,g)\phi(g')^2}{\phi(g)^4} \Big) \\ & \times \sum_{i=J(n,r-1)+1}^{J(n,r)} (\Delta^n_i Z)^2 (\Delta^n_{i+2} Z)^2 \Big). \end{aligned} \tag{A.3}$$

A.1. A moment estimate for $\widehat{C}(k_n, \Delta_n)_1$

We start with an auxiliary result, which complements Theorem 2.2 with a moment estimate. In the way it will be used later we might employ, for the estimator, at time t observations occurring slightly after this time, so we replace (2.8) by the following mild modification:

$$\widehat{C}(k_n,\Delta_n)'_t=\sum_{i=1}^{[t/\Delta_n]}\Big(\frac{1}{k_n\,\phi(g)}\,(\overline{Z}(k_n)^n_i)^2-\frac{\phi'(g)}{2k_n^2\,\phi(g)}\,\widehat{Z}(k_n)^n_i\Big). \tag{A.4}$$

What follows is a non-asymptotic result, so we simply take a time lag $\Delta>0$ and an integer $k\ge 2$, and the above estimator at time 1 is $\widehat{C}(k,\Delta)'_1$. To avoid trivialities we assume also $k\Delta<1$.

Lemma A.1. *Assume (H) and (N) with the following bounds*

$$|b_t|\le A,\qquad c_t\le A,\qquad |X_t|\le A,\qquad \int y^8\,Q_t(dy)\le Av^8, \tag{A.5}$$

for two numbers $A,v\ge 1$. Then, for some constant K_A depending on A only (and not on v), and for any integer $k\ge 2$ and number $\Delta\in(0,1/k)$ we have

$$\mathbb{E}\big(|\widehat{C}(k,\Delta)'_1-C_1|\big)\le K_A\Big(\sqrt{k\Delta}+v\sqrt{\Delta}+\frac{v^2}{\Delta^{1/2}k^{3/2}}+\frac{v^2}{\Delta k^3}\Big). \tag{A.6}$$

Proof. (1) The proof consists in copying, line after line, the proof given in [18] and keeping track of all necessary moments. Below the constant K varies from line to line but only depends on A.

The key decomposition is (A-5.14), which we use for $p=1$ and at $t=1$, that is

$$\widehat{C}(k,\Delta)'_1-C_1=M(1)_1+M'(1)_1+F(1)_1+F'(1)_1+\widehat{C}(1)_1+\widehat{C}'_1+\widehat{C}''_1.$$

(Note that, despite what is written in that paper, $\widehat{C}'^n_t$ does not depend on p.) Note that in (A-5.9)–(A-5.13) we replace everywhere $\sqrt{\Delta_n}/\theta$ and Δ_n/θ^2 by $1/k$ and $1/k^2$ respectively. The notation $\chi(1)^n_i=\chi(1)_i$ is heavily used, but here we simply replace it by its majorant $4A$ or $4Av^2$, according to whether it is used for b_t and σ_t, or for γ_t (called α_t in that paper). As mentioned before, $k_n=k$ and $\Delta_n=\Delta$ everywhere.

(2) Now, we go along the proof in [18], tediously pointing out the changes to be made in each relevant formula (recall $p=1$; we are interested in moments, not in convergence).

- In (A-5.17) and (A-5.18) the O_{Pu}'s are respectively $\mathrm{O}_{Pu}(\Delta^2k^4)$ and $\mathrm{O}_{Pu}(\Delta^2k^{5/4})$.
- In (A-5.28) the second bound is $K_q(\Delta k)^{q/2}$.
- In (A-5.29) the two bounds are respectively $K\Delta^4k^8$ and $K\Delta^2k^2$.
- In (A-5.30) the bound is $K\Delta^{3/2}k^{5/2}$.
- In (A-5.32) the two bounds are respectively $K\Delta^2k^4$ and $K\Delta k^3$.
- In (A-5.36) the bound for $A^n_{i,j}$ is Kv^2/k, and the two O_{Pu}'s are respectively $\mathrm{O}_{Pu}(v^2/k)$ and $\mathrm{O}_{Pu}(v^4)$.
- In (A-5.38) the three O_{Pu}'s are $\mathrm{O}_{Pu}(v^3/k^2)$ and $\mathrm{O}_{Pu}(v^4/k^3)$ and $\mathrm{O}_{Pu}(v^8/k^4)$.
- In (A-5.39) the first two O_{Pu}'s are $\mathrm{O}_{Pu}(\Delta k+v\Delta^{1/2})$ and $\mathrm{O}_{Pu}(v^3(|\overline{X}^n_i|+|\overline{X}^n_j|)/k^2)$.
- In (A-5.40) the last bound is $K\Delta^{3/2}k^{5/2}$.
- We replace (A-5.41) by the bound $\mathbb{E}((\zeta(Z,1)^n_i)^2)\le K(v^4+\Delta^2k^4)$.
- We replace (A-5.44) and (A-5.45) by $\mathbb{E}((\overline{Z}^n_i)^2)\le K(\Delta^2k^2+v^2\Delta+v^4/k^2)$ and $|\mathbb{E}(\zeta''(Z)^n_i\mid\mathcal{F}^n_i)|\le K(v^4+v^2\Delta k)$.
- In the second part of (A-5.46) the bound is Kv^4.

(3) Using the modified version of (A-5.40), and observing that $\frac{1}{L} \leq 2\Delta k j_n(1,1) \leq L$ for some constant $L > 1$, we obtain

$$\mathbb{E}(|F(1)_1| + |F'(1)_1| \mid \mathcal{F}_0) \leq K\sqrt{\Delta k}.$$

Next, the number of summands in the definition of $\widehat{C}(1)_1$ is less than Lk, hence by (A-5.28) and (A-5.39) we get

$$\mathbb{E}(|\widehat{C}(1)_1|) \leq K(\Delta k + v\sqrt{\Delta}).$$

Next, modifying the final part of the proof of Lemma A-5.5 and the proof of Lemma A-5.6 appropriately, we obtain

$$\mathbb{E}(|\widehat{C}_1''|) \leq K\Big(\Delta k + \frac{1}{k}\Big), \qquad \mathbb{E}(|\widehat{C}_t'|) \leq K\Big(\frac{v^2}{k} + \frac{v^2}{\Delta k^3}\Big).$$

Finally, we apply (A-5.40) and Doob's inequality to get

$$\mathbb{E}(|M(1)_1|^2 + |M'(1)_1|^2) \leq \frac{K}{\Delta k^3}\,(v^4 + \Delta^2 k^4).$$

All these estimates, plus the properties $k \geq 2$ and $\Delta k < 1$ and $v \geq 1$, readily give (A.6). □

A.2. Local estimation

In this subsection we prove (4.6), and in fact slightly more, as needed further on. The number $a \in (0,1)$ showing in (4.4) is arbitrary, but fixed.

Lemma A.2. *Under (S-HNL) and (4.3), and for $r \geq 2$, we have*

$$\mathbb{E}\big(|\widehat{\gamma}(r)^n - \overline{\gamma}(r)^n|\big) \leq K\Delta_n^{w/2}. \tag{A.7}$$

Proof. We have $\widehat{\gamma}(r)^n - \overline{\gamma}(r)^n = \sum_{m=1}^5 \xi(m)_i^n$, where

$$\xi(1)_r^n = \frac{1}{2l_n'} \sum_{i=(r-1)l_n-l_n'+1}^{(r-1)l_n} (\Delta_i^n X)^2, \qquad \xi(2)_r^n = \frac{1}{l_n'} \sum_{i=(r-1)l_n-l_n'+1}^{(r-1)l_n} \Delta_i^n Y\,\Delta_i^n X$$

$$\xi(3)_r^n = -\frac{1}{l_n'} \sum_{i=(r-1)l_n-l_n'+1}^{(r-1)l_n} Y_{(i-1)\Delta_n}\,Y_{i\Delta_n}$$

$$\xi(4)_r^n = \frac{1}{2l_n'} \sum_{i=(r-1)l_n-l_n'+1}^{(r-1)l_n} \big((Y_{i\Delta_n})^2 - \gamma_{i\Delta_n} + (Y_{(i-1)\Delta_n})^2 - \gamma_{(i-1)\Delta_n}\big)$$

$$\xi(5)_r^n = \frac{1}{2l_n'} \sum_{i=(r-1)l_n-l_n'+1}^{(r-1)l_n} \Big(\gamma_{i\Delta_n} + \gamma_{(i-1)\Delta_n} - \frac{2}{\Delta_n}\int_{(i-1)\Delta_n}^{i\Delta_n} \gamma_s\,ds\Big).$$

By virtue of the properties in (S-HNL), and by successive conditioning, we readily get

$$\mathbb{E}((\Delta_i^n X)^2) \leq K\Delta_n \Rightarrow \mathbb{E}(|\xi(1)_r^n|) \leq K\Delta_n$$
$$\mathbb{E}(\Delta_i^n Y\,\Delta_i^n X \mid \mathcal{F}_{(i-1)\Delta_n}) = 0, \qquad \mathbb{E}((\Delta_j^n Y\,\Delta_j^n X)^2) \leq K\Delta_n \Rightarrow \mathbb{E}(|\xi(2)_r^n|^2) \leq K\Delta_n/l_n'$$
$$\mathbb{E}(Y_{(i-1)\Delta_n}\,Y_{i\Delta_n} \mid \mathcal{F}_{(i-1)\Delta_n}) = 0, \qquad \mathbb{E}((Y_{(i-1)\Delta_n}\,Y_{i\Delta_n})^2) \leq K \Rightarrow \mathbb{E}(|\xi(3)_r^n|^2) \leq K/l_n'$$
$$\mathbb{E}((Y_{i\Delta_n})^2 - \gamma_{i\Delta_n} \mid \mathcal{F}_{i\Delta_n}) = 0, \qquad \mathbb{E}(((Y_{i\Delta_n})^2 - \gamma_{i\Delta_n})^2) \leq K \Rightarrow \mathbb{E}(|\xi(4)_r^n|^2) \leq K/l_n'.$$

Finally, an application of (4.1) with $\tau_1 = \infty$ and Cauchy–Schwarz inequality yield $\mathbb{E}(|\xi(5)_i^n|) \leq K\sqrt{\Delta_n}$, and in view of (4.3) we deduce the result. □

Lemma A.3. *Under (S-HNL) and (4.3) we have*

$$\mathbb{E}\big(|\widehat{c}(r)^n - \overline{c}(r)^n|\big) \leq K\Delta_n^{(2w-1)/4}. \tag{A.8}$$

Proof. With $u_n = l'_n\Delta_n$ and $S(n,r) = T(n,r-1) - (l'_n + k_n)\Delta_n$, we set

$$\mathcal{F}(n,r)_t = \mathcal{F}_{S(n,r-1)+u_n t}, \qquad X(n,r)_t = \frac{1}{\sqrt{u_n}}(X_{S(n,r)+u_n t} - X_{S(n,r)})$$
$$Y(n,r)_t = \frac{1}{\sqrt{u_n}}Y_{S(n,r)+u_n t}, \qquad Z(n,r)_t = \frac{1}{\sqrt{u_n}}Z_{S(n,r)+u_n t} = X(n,r)_t + Y(n,r)_t.$$

Note that for each pair (n,r) the process $X(n,r)$ satisfies (H) with the characteristics

$$b(n,r)_t = \sqrt{u_n}\, b_{S(n,r)+u_n t}, \qquad \sigma(n,r)_t = \sigma_{S(n,r)+u_n t},$$

whereas the conditional distribution of the noise $Y(n,r)$ at time t is the image of $Q_{S(n,r)+u_n t}$ by the map $y \mapsto y/\sqrt{u_n}$. Therefore $(X(n,r), Y(n,r))$ satisfies (A.5) with $v = v_n = 1/\sqrt{u_n}$ and with a constant A independent of n and r.

Observe that $\widehat{c}(r)^n$, as given by (A.2), is exactly the estimator (A.4) evaluated at time $t = 1$, when one observes the variables $Z(n,r)_{j\Delta'_n}$, $j = 0, 1, \ldots$, with the new time lag $\Delta'_n = 1/l'_n$. Hence Lemma A.1 yields

$$\mathbb{E}\Big(\Big|\widehat{c}(r)^n - \int_0^1 c(n,r)_s\,ds\Big|\Big) \leq K\Big(\frac{k_n^{1/2}}{l_n^{1/2}} + \frac{1}{l_n\sqrt{\Delta_n}} + \frac{1}{l_n^{1/2}k_n^{3/2}\Delta_n} + \frac{1}{k_n^3\Delta_n}\Big).$$

Moreover, $\big|\int_0^1 c(n,r)_s\,ds - \overline{c}(r)^n\big| \leq Kk_n/l'_n$ (recall that c_t is bounded), and a simple calculation using (4.3) again gives the result. □

A.3. Proof of Theorem 4.1

Exactly as for the local estimation above, the proof consists in adequately modifying the proof of [18]. However, the modifications are somewhat deeper here.

(1) We start by showing that we can assume the following properties, where $A > 1$ is some constant and $w' = \frac{6w-5}{8}$ (which is positive):

$$\begin{aligned} &\frac{1}{A} \leq k_{n,r}\sqrt{\Delta_n} \leq A, && k_{n,r} \text{ is } \mathcal{F}_{T(n,r-1)}\text{-measurable} \\ &r \geq 2 \Rightarrow |\widehat{\theta}(r)^n - \overline{\theta}(r)^n| \leq A\Delta_n^{w'}, && |k_{n,r}\sqrt{\Delta_n} - \widehat{\theta}(r)^n| \leq A\sqrt{\Delta_n}. \end{aligned} \tag{A.9}$$

This is trivial for $r = 1$, since then $k_{n,1} = k_n \asymp 1/\sqrt{\Delta_n}$. For the other cases, we set

$$\Omega_n = \bigcap_{r=2}^{L_n}\Big\{|\widehat{c}(r)^n - \overline{c}(r)^n| \leq \Delta_n^{w'},\ |\widehat{\gamma}(r)^n - \overline{\gamma}(r)^n| \leq \Delta_n^{w'}\Big\}.$$

By Lemma A.2 and A.3 and Markov's inequality and $L_n \asymp 1/\Delta_n^{1-w}$, we have

$$\mathbb{P}((\Omega_n)^c) \leq KL_n\Delta_n^{(2w-1)/4-w'} \leq K\Delta_n^{(6w-5)/4-w'} \to 0. \tag{A.10}$$

In view of (S-HNL) and of the definition of Ω_n, there is a constant $A > 1$ such that for all $2 \leq r \leq L_n$:

$$\begin{array}{l} \frac{1}{A} \leq \overline{\theta}(r)^n \leq A \quad \text{on } \Omega \\ \frac{1}{A} \leq \widehat{\theta}(r)^n \leq A, \qquad |\widehat{\theta}(r)^n - \overline{\theta}(r)^n| \leq A\Delta_n^{w'} \quad \text{on } \Omega_n. \end{array} \tag{A.11}$$

Now, $k_{n,r}$ and $\overline{\theta}(r)^n$ are $\mathcal{F}_{T(n,r-1)}$-measurable by construction, and we set for $r = 2, \dots, L_n$:

$$\begin{cases} \widehat{\theta}'(r)^n = \widehat{\theta}(r)^n, & k'_{n,r} = k_{n,r} \qquad \text{on } \Omega_n \\ \widehat{\theta}'(r)^n = \overline{\theta}(r)^n, & k'_{n,r} = [\widehat{\theta}'(r)^n/\sqrt{\Delta_n}\,] \quad \text{otherwise.} \end{cases}$$

In view of (A.11), and upon enlarging A if necessary, it is clear that the sequences $(k'_{n,r}, \widehat{\theta}'(r)^n, \overline{\theta}(r)^n : 2 \leq r \leq L_n)$ satisfy (A.9) (uniformly in n and r). Moreover they coincide with $(k_{n,r}, \widehat{\theta}(r)^n, \overline{\theta}(r)^n : 2 \leq r \leq L_n)$ on Ω_n, whereas (A.10) holds, we do not change the asymptotic properties of our estimators if we replace $k_{n,r}$ by $k'_{n,r}$. In other words, we can indeed assume (A.9) for the original sequences $(k_{n,r}, \widehat{\theta}(r)^n, \overline{\theta}(r)^n)$.

(2) The first modification upon [18] is a change in the decomposition (A-5.14). We have our "basic blocks" corresponding to the time intervals $[T(n, r-1), T(n, r)]$ for $r = 1, \dots, L_n$ and the first index occurring within the rth block if $J(n, r) = rl_n$, whereas Λ_r^n is the set of all indices i between $J(n, r-1) + 1$ and $J(n, r)$ (those are the indices on which the rth sum in (A.1) is taken). Then we replace the second part of (A-5.3) and (A-5.4)–(A-5.14) by the following. First, for any $i \in \Lambda_r^n$ and any process V, we set

$$\begin{aligned}
c_i^n &= \sum_{j=1}^{k_{n,r}-1} g\Big(\frac{j}{k_{n,r}}\Big)^2 \Delta_{i+j}^n C \\
A_{i,j}^n &= \sum_{m=1+i\vee j}^{i\wedge j+k_{n,r}} \Big(g\Big(\frac{m-i}{k_{n,r}}\Big) - g\Big(\frac{m-i-1}{k_{n,r}}\Big)\Big)\Big(g\Big(\frac{m-j}{k_{n,r}}\Big) - g\Big(\frac{m-j-1}{k_{n,r}}\Big)\Big)\gamma_{(m-1)\Delta_n} \\
A_i^n &= A_{i,i}^n = \sum_{m=1}^{k_{n,r}} \Big(g\Big(\frac{m}{k_{n,r}}\Big) - g\Big(\frac{m-1}{k_{n,r}}\Big)\Big)^2 \gamma_{(i+m-1)\Delta_n} \\
\widetilde{Z}_i'^n &= (\overline{Z}(k_{n,r})_i^n)^2 - A_i^n - c_i^n, \qquad \zeta(Z, p)_i^n = \sum_{j=i+1}^{i+pk_{n,r}} \widetilde{Z}_j'^n \\
\zeta(X, p)_i^n &= \sum_{j=i+1}^{i+pk_{n,r}} \Big((\overline{X}(k_{n,r})_j^n)^2 - c_j^n\Big), \\
\zeta(W, p)_i^n &= \sum_{j=i}^{i+pk_{n,r}-1} \Big((\sigma_{i\Delta_n}\overline{W}(k_{n,r})_j^n)^2 - c_j^n\Big) \\
\zeta'(V, p)_i^n &= \sum_{(j,m):\, i+1\leq j<m\leq i+pk_{n,r}} \overline{V}(k_{n,r})_j^n \overline{V}_m^n\, \phi\Big(g', g'\Big|\frac{m-j}{k_{n,r}}\Big) \\
\zeta''(V)_i^n &= (\overline{V}(k_{n,r})_i^n)^2 \sum_{j=i+k_{n,r}+1}^{i+2k_{n,r}} (\Delta_j^n V)^2.
\end{aligned}$$

Then $\mathcal{F}_j^n = \mathcal{F}^{(0)}_{j\Delta_n} \otimes \mathcal{F}^{(1)}_{j\Delta_n-}$ and $\mathcal{F}_j^{\prime n} = \mathcal{F}^{(0)} \otimes \mathcal{F}^{(1)}_{j\Delta_n-}$, and $\mathcal{G}(r,p)_j^n = \mathcal{F}^n_{J(n,r-1)+j(p+1)k_{n,r}}$ and $\mathcal{G}'(r,p)_j^n = \mathcal{F}^n_{J(n,r-1)+j(p+1)k_{n,r}+pk_{n,r}}$ (warning: here $k_{n,r}$ is random, but $\mathcal{F}^n_{J(n,r-1)}$-measurable, so the last two σ-fields are well defined). Next,

$$\eta(r,p)_j^n = \frac{1}{k_{n,r}\,\phi(g)}\,\zeta(Z,p)^n_{J(n,r-1)+j(p+1)k_{n,r}}, \qquad \overline{\eta}(r,p)_j^n = \mathbb{E}(\eta(r,p)_j^n \mid \mathcal{G}(r,p)_j^n)$$

$$\eta'(r,p)_j^n = \frac{1}{k_{n,r}\,\phi(g)}\,\zeta(Z,1)^n_{J(n,r-1)+j(p+1)k_{n,r}+pk_{n,r}},$$

$$\overline{\eta}'(r,p)_j^n = \mathbb{E}(\eta'(r,p)_j^n \mid \mathcal{G}'(r,p)_j^n).$$

The number of pairs of blocks within the rth "basic block" is $j_n(r,p) = \left[\frac{J(n,r)-J(n,r-1)}{(p+1)k_{n,r}}\right]$. For all n large enough (depending on p), we have $j_n(r,p) \geq 1$. With an empty sum set equal to 0, we set

$$F(p)^n = \sum_{r=1}^{L_n} \sum_{j=0}^{j_n(r,p)-1} \overline{\eta}(r,p)_j^n, \qquad M(p)^n = \sum_{r=1}^{L_n} \sum_{j=0}^{j_n(r,p)-1} (\eta(r,p)_j^n - \overline{\eta}(r,p)_j^n)$$

$$F'(p)^n = \sum_{r=1}^{L_n} \sum_{j=0}^{j_n(r,p)-1} \overline{\eta}'(r,p)_j^n, \qquad M'(p)^n = \sum_{r=1}^{L_n} \sum_{j=0}^{j_n(r,p)-1} (\eta'(r,p)_j^n - \overline{\eta}'(r,p)_j^n).$$

We set $i_n(r,p) = J(n,r-1) + j_n(r,p)(p+1)k_{n,r}$. The residual processes are

$$\widehat{C}(p)^n = \sum_{r=1}^{L_n} \sum_{i=i_n(r,p)+1}^{J(n,r)} \frac{1}{k_{n,r}\,\phi(g)} \widetilde{Z}_i^{\prime n} \tag{A.12}$$

$$\widehat{C}'(p)^n = \sum_{r=1}^{L_n} \frac{1}{k_{n,r}\,\phi(g)} \sum_{i=J(n,r-1)+1}^{J(n,r)} \Big(A_i^n - \frac{\phi(g')}{2k_{n,r}}\,(\Delta_i^n Z)^2\Big) \tag{A.13}$$

$$\widehat{C}''^n = \sum_{r=1}^{L_n} \sum_{i=J(n,r-1)+1}^{i_n(r,p)} \frac{1}{k_{n,r}\,\phi(g)}\, c_i^n - C_T. \tag{A.14}$$

Then, for each p we have (A-5.14), that is

$$\widehat{C}_T^n - C_T = M(p)^n + M'(p)^n + F(p)^n + F'(p)^n + \widehat{C}(p)^n + \widehat{C}'(p)^n + \widehat{C}''^n. \tag{A.15}$$

We also modify the definition of $\beta(p)_i^n$ in (A-5.15) as follows, with A as in (A.9):

$$\beta(p)_i^n = \sup_{s,t\in[i\Delta_n,\, i\Delta_n + A(p+2)\sqrt{\Delta_n}]} \big(|b_s - b_t| + |\sigma_s - \sigma_t| + |\gamma_s - \gamma_t|\big).$$

(3) Next we consider the various estimates on W, X and Z in [18]. Again we use the same notation, except that Φ_{11} and Φ_{12} and Φ_{22} there are $\Phi(g',g')$ and $\Phi(g,g')$ and $\Phi(g,g)$ here.

First, upon using (A.9), we see that (A-5.17) and (A-5.18) become, when $i \in \Lambda_r^n$ and recalling $c_t = \sigma_t^2$ and Ξ_{ij}, as defined by (A-5.16):

$$\mathbb{E}((\zeta(W,p)_i^n)^2 \mid \mathcal{F}_i^n) = 4(p\,\Phi(g,g) + \Xi_{22})\, k_{n,r}^4\, \Delta_n^2 c_{i\Delta_n}^2 + \mathrm{O}_{Pu}(p^2\chi(p)_i^n), \tag{A.16}$$

$$\mathbb{E}\Big(\zeta'(W,p)_i^n \mid \mathcal{F}_i^n\Big) = (p\,\Phi(g,g') + \Xi_{12}) k_{n,r}^3\, \Delta_n + \mathrm{O}_{Pu}(p\Delta_n^{-1/4}). \tag{A.17}$$

Next, (A-5.27), (A-5.28), (A-5.29) and (A-5.30) are unchanged, whereas (A-5.32) becomes, for $i \in \Lambda_r^n$ again:

$$\left.\begin{array}{l}\left|\mathbb{E}\Big((\zeta(X,p)_i^n)^2 \mid \mathcal{F}_i^n\Big) - 4(p\Phi(g,g)+\Xi_{22})\,k_{n,r}^4\,\Delta_n^2 c_{i\Delta_n}^2\right| \le K_p\chi(p)_i^n \\ \left|\mathbb{E}\Big(\zeta'(p,X)_i^n \mid \mathcal{F}_i^n\Big) - (p\Phi(g,g')+\Xi_{12})k_{n,r}^3\,\Delta_n c_{i\Delta_n}\right| \le K_p\Delta_n^{-1/2}\chi(p)_i^n.\end{array}\right\} \tag{A.18}$$

Next, for $i \in \Lambda_r^n$, (A-5.36) is replaced by

$$\left.\begin{array}{l}|A_{i,j}^n| \le K\sqrt{\Delta_n} \\ |j-i| \ge k_{n,r}, \qquad J(n,r-1) \le i < J(n,r) \Rightarrow A_{i,j}^n = 0 \\ i \le j \le m \le (i+(p+1)k_{n,r}) \wedge J(n,r) \\ \quad \Rightarrow A_{j,m}^n = \gamma_{i\Delta_n}\dfrac{1}{k_{n,r}}\,\phi\Big(g',g'\Big|\dfrac{m-j}{k_{n,r}}\Big) + \mathrm{O}_{Pu}(p\Delta_n + \sqrt{\Delta_n}\,\beta(p)_i^n) \\ i + pk_{n,r} < J(n,r) \\ \quad \Rightarrow \displaystyle\sum_{(j,m):\, i\le j<m\le i+pk_{n,r}} (A_{j,m}^n)^2 = \gamma_{i\Delta_n}^2(p\Phi(g',g')+\Xi_{11}) \\ \qquad + \mathrm{O}_{Pu}\Big(p^3\sqrt{\Delta_n} + p\beta(p)_i^n\Big).\end{array}\right\} \tag{A.19}$$

Then, if $i \in \Lambda_r^n$, we still have (A-5.37)–(A-5.40) and (A-5.42). In particular, (A-5.30) and (A-5.40) yield

$$|E(\zeta(Z,p)_i^n \mid \mathcal{F}_i'^n)| \le Kp\Delta_n^{1/4}\chi(p)_i^n \le Kp\Delta_n^{1/4}. \tag{A.20}$$

If $J(n,r-1) < i$ and $i+(p+1)k_{n,r} \le J(n,r)$, (A-5.41) is replaced by

$$\begin{aligned}&\Big|\mathbb{E}((\zeta(Z,p)_i^n)^2 \mid \mathcal{F}_i^n) - 4(p\Phi(g,g)+\Xi_{22})k_{n,r}^4\,\Delta_n^2 c_{i\Delta_n}^2 \\ &\quad - 8\gamma_{i\Delta_n}c_{i\Delta_n}(p\Phi(g,g')+\Xi_{12})k_{n,r}^2\,\Delta_n - 4\gamma_{i\Delta_n}^2(p\Phi(g',g')+\Xi_{11})\Big| \\ &\le K_p\,\chi(p)_i^n.\end{aligned} \tag{A.21}$$

Moreover, (A-5.46) holds with γ instead of α, and the first part of (A-5.43), (A-5.44) and (A-5.45) become when $i \in \Lambda_r^n$:

$$\begin{aligned}\mathbb{E}\Big(\zeta''(Z)_i^n \mid \mathcal{F}_i'^n\Big) \;=\; &\zeta''(X)_i^n + A_i^n \sum_{j=i+k_{n,r}+1}^{i+2k_{n,r}} (\Delta_j^n X)^2 \\ &+ ((\overline{X}_i^n)^2 + A_i^n) \sum_{j=i+k_{n,r}+1}^{i+2k_{n,r}} (\gamma_{(j-1)\Delta_n} + \gamma_{j\Delta_n}),\end{aligned} \tag{A.22}$$

$$\begin{aligned}&\left|\mathbb{E}((\overline{Z}_i^n)^4 \mid \mathcal{F}_i^n) - 3k_{n,r}^2\Delta_n^2\phi(g)^2c_{i\Delta_n}^2 - 6\Delta_n c_{i\Delta_n}\gamma_{i\Delta_n}\phi(g)\phi(g') - \frac{3}{k_{n,r}^2}(\gamma_i^n)^2\phi(g')^2\right| \\ &\le K\Delta_n\chi(1)_i^n\end{aligned} \tag{A.23}$$

$$\left|\mathbb{E}(\zeta''(Z)_i^n \mid \mathcal{F}_i^n) - 2\gamma_{i\Delta_n}(\phi(g')\gamma_i^n + \phi(g)k_{n,r}^2\,\Delta_n c_{i\Delta_n}\gamma_i^n)\right| \le K\chi(1)_i^n. \tag{A.24}$$

(4) Now we prove the following lemma:

Lemma A.4. *For any $p \geq 1$ we have, as $n \to \infty$,*

$$\frac{F(p)^n}{\Delta_n^{1/4}} \stackrel{\mathbb{P}}{\longrightarrow} 0, \qquad \frac{F'(p)^n}{\Delta_n^{1/4}} \stackrel{\mathbb{P}}{\longrightarrow} 0, \qquad \frac{\widehat{C}(p)^n}{\Delta_n^{1/4}} \stackrel{\mathbb{P}}{\longrightarrow} 0, \qquad \frac{\widehat{C}'(p)^n}{\Delta_n^{1/4}} \stackrel{\mathbb{P}}{\longrightarrow} 0,$$
$$\frac{\widehat{C}''^n}{\Delta_n^{1/4}} \stackrel{\mathbb{P}}{\longrightarrow} 0.$$

Proof. Since $k_{n,r}\sqrt{\Delta_n} \geq 1/A$, it follows from (A-5.40), from the definition of $F(p)^n$, and from the fact that the set $\{j_n(p,r) > j\}$ belongs to $\mathcal{F}^n_{J(n,r-1)}$, that

$$\mathbb{E}(|F(p)^n|) \leq Kp\Delta_n^{3/4}\sum_{r=1}^{L_n}\mathbb{E}\Big(\sum_{j=0}^{j_n(r,p)-1}\Big(\Delta_n^{1/4} + \sqrt{\mathbb{E}\Big((\beta(p)^n_{J(n,r-1)+j(p+1)k_{n,r}})^2 \mid \mathcal{F}^n_{J(n,r-1)+j(p+1)k_{n,r}}\Big)}\Big)\Big).$$

Moreover, the proof of Lemma A-5.4 can easily be adapted to the random $j_n(r,p)$'s and $k_{n,r}$ here, hence $\mathbb{E}(|F(p)^n|) = \mathrm{o}(\Delta_n^{1/4})$, thus yielding the claim for $F(p)^n$, and the claim for $F'(p)^n$ is analogous.

Next, (A-5.28) and (A-5.39) yield $\mathbb{E}(|\widetilde{Z}_i'^n|^4) \leq K\Delta_n^2$, so each summand in (A.12) has absolute expectation less than $K\Delta_n$, whereas the total number of summands is less than $KL_np/\sqrt{\Delta_n}$ (we again use the first part of (A.9) here) and $L_n \leq K/\Delta_n^{1/6}$. Thus the claim for $\widehat{C}(p)^n$ follows.

Next, as in [18], we see that

$$\Big|\sum_{i=J(n,r-1)+1}^{i_n(r,p)} \frac{c_i^n}{k_{n,r}\,\phi(g)} - (C_{i_n(r,p)\Delta_n} - C_{J(n,r-1)\Delta_n})\Big| \leq K\sqrt{\Delta_n}.$$

On the other hand, $C_{J(n,r)\Delta_n} - C_{i_n(r,p)\Delta_n} \leq Kp\sqrt{\Delta_n}$ when $r \leq L_n$, and $C_T - C_{J(n,L_n)\Delta_n} \leq Kp\sqrt{\Delta_n}$ as well. Thus $|\widehat{C}''^n| \leq KL_np\sqrt{\Delta_n} \leq Kp\Delta_n^{3/8}$, hence the claim for $\widehat{C}''^n$.

It remains to study $\widehat{C}'(p)^n$. For this we reproduce the proof of Lemma A-5.6, with some changes due to different lower and/or upper limits in the various sums. Similar with that proof, we have

$$G_n := \sum_{r=1}^{L_n}\sum_{i=J(n,r-1)+1}^{J(n,r)} \frac{\phi(g')}{2k_{n,r}^2\phi(g)\Delta_n^{1/4}}\,((\Delta_i^n Z)^2 - \gamma_{(i-1)\Delta_n} - \gamma_{i\Delta_n}) \stackrel{\mathbb{P}}{\longrightarrow} 0,$$

whereas it holds that $\widehat{C}'(p)^n + \Delta_n^{1/4}G_n = \sum_{r=1}^{L_n}(U_r^n + V_r^n)$ where

$$U_r^n = \frac{1}{k_{n,r}\phi(g)}\Big(\phi'_{k_{n,r}}(g) - \frac{\phi(g')}{k_{n,r}}\Big)\sum_{i=J(n,r-1)+1}^{J(n,r)}\gamma_{i\Delta_n}$$

$$V_r^n = \frac{1}{k_{n,r}\phi(g)}\sum_{m=2}^{k_{n,r}}\Big(g\Big(\frac{m}{k_{n,r}}\Big) - g\Big(\frac{m-1}{k_{n,r}}\Big)\Big)^2 \times\Big(\sum_{i=J(n,r)+1}^{J(n,r)-1+k_{n,r}\wedge m}\gamma_{i\Delta_n} - \sum_{i=J(n,r-1)+1}^{J(n,r-1)-1+k_{n,r}\wedge m}\gamma_{i\Delta_n}\Big) - \frac{\phi(g')}{2k_{n,r}^2\phi(g)}\big(\gamma_{J(n,r-1)\Delta_n} - \gamma_{J(n,r)\Delta_n}\big).$$

On the one hand, since γ is bounded, one deduces from (A.9) that $|V_r^n| \leq K\sqrt{\Delta_n}$. On the other hand, $\left|\phi'_{k_{n,r}}(g) - \frac{\phi(g')}{k_{n,r}}\right| \leq K\Delta_n$, hence $|U_r^n| \leq K\sqrt{\Delta_n}$ as well, and the claim for $\widehat{C}'(p)^n$ follows. □

(5) It remains to study the two main terms $M(p)^n$ and $M'(p)^n$ of (A.15). Both are sums of martingale increments, relative to suitable discrete-time filtrations which could be explicitly described, and we will do this for $M(p)^n$ below. For $M'(p)^n$, it suffices to observe that the total number of terms is less than $K/p\sqrt{\Delta_n}$, because of (A.9), which implies that any two successive big and small blocks have a total length more than $(p+1)\sqrt{\Delta_n}/A$ in calendar time. Since (A.21) implies that the variance of each summand is less than $K\Delta_n$, we deduce that

$$\mathbb{E}\big(|M'(p)^n|^2\big) \leq \frac{K\sqrt{\Delta_n}}{p}. \tag{A.25}$$

For studying $M(p)^n$, we fix $p \geq 2$ and set for $r \geq 0$ (with again an empty sum being 0):

$$u_n(r) = \sum_{r'=1}^{r} j_n(r', p), \qquad t_0^n = 0$$

$$u_n(r-1)+1 \leq j \leq u_n(r) \Rightarrow \begin{cases} \rho_j^n = \eta(r,p)^n_{j-1-u_n(r-1)}, \qquad \overline{\rho}_j^n = \overline{\eta}(r,p)^n_{j-1-u_n(r-1)} \\ \mathcal{G}_j^n = \mathcal{G}(r,p)^n_{j-1-u_n(r-1)} \\ t_j^n = (J(n,r-1) + (j-u_n(r-1))(p+1)k_{n,r})\Delta_n. \end{cases}$$

Note that $u_n(r)$ and t_j^n are random. Then $M(p)^n = \sum_{j=1}^{u_n(L_n)} (\rho_j^n - \overline{\rho}_j^n)$, and the variables $\rho_j^n - \overline{\rho}_j^n$ are martingales increments, relative to the discrete-time filtration $(\mathcal{G}_j^n)_{j\geq 0}$. One also has $M(p)^n = \overline{M}(p)^n_{u_n(L_n)}$, where $\overline{M}(p)^n$ is the continuous-time martingale relative to the filtration $(\mathcal{H}_s^n)_{s\geq 0}$ defined by $\mathcal{H}_0^n = \{\emptyset, \Omega\}$ and $\mathcal{H}_s^n \cap \{t_{j-1}^n < s \leq t_j^n\} = \mathcal{G}_j^n \cap \{t_{j-1}^n < s \leq t_j^n\}$:

$$\overline{M}(p)_s^n = \sum_{j\geq 1:\, t_j^n \leq s} (\rho_j^n - \overline{\rho}_j^n).$$

Lemma A.5. *For any fixed $p \geq 2$, the processes $\big(\frac{1}{\Delta_n^{1/4}}\,\overline{M}(p)_s^n\big)_{s\in[0,1]}$ converge stably in law to*

$$Y(p)_s = \int_0^s \sqrt{\Gamma(p)_v}\, dB_v, \tag{A.26}$$

where B is a standard Brownian motion defined on an extension of the original space and is independent of $\mathcal{F}$, and where (recalling (3.8) for the process θ_t);

$$\begin{aligned} \Gamma(p)_s = \frac{4}{\phi(g)^2}\Big(&\Big(\frac{p}{p+1}\,\Phi(g,g) + \frac{1}{p+1}\,\Xi_{22}\Big)\theta_s c_s^2 \\ &+ 2\Big(\frac{p}{p+1}\,\Phi(g,g') + \frac{1}{p+1}\,\Xi_{12}\Big)\frac{c_s\gamma_s}{\theta_s} \\ &+ \Big(\frac{p}{p+1}\,\Phi(g',g') + \frac{1}{p+1}\,\Xi_{11}\Big)\frac{\gamma_s^2}{\theta_s^3}\Big). \end{aligned} \tag{A.27}$$

Proof. It suffices to prove the following three convergences, for all $s > 0$:

$$\frac{1}{\sqrt{\Delta_n}}\sum_{j\geq 1:\,t_j^n\leq s}\Big(\mathbb{E}((\rho_j^n)^2\mid \mathcal{G}_{j-1}^n)-(\overline{\rho}_j^n)^2\Big)\stackrel{\mathbb{P}}{\longrightarrow}\int_0^s \Gamma(p)_v\,dv, \tag{A.28}$$

$$\frac{1}{\Delta_n}\sum_{j\geq 1:\,t_j^n\leq s}\mathbb{E}((\rho_j^n)^4\mid \mathcal{G}_{j-1}^n)\stackrel{\mathbb{P}}{\longrightarrow}0, \tag{A.29}$$

$$\frac{1}{\Delta_n^{1/4}}\sum_{j\geq 1:\,t_j^n\leq s}\mathbb{E}(\rho_j^n\,\Delta(N)_j^n\mid \mathcal{G}_{j-1}^n)\stackrel{\mathbb{P}}{\longrightarrow}0, \tag{A.30}$$

where $\Delta(N)_j^n = N_{t_j^n} - N_{t_{j-1}^n}$, and where (A.30) should hold for $N = W$ and for all N in a set $\mathcal{N}$ of bounded martingales which are orthogonal to W and such that the family $(N_\infty : N \in \mathcal{N})$ is total in the space $\mathbb{L}^1(\Omega, \mathcal{F}, \mathbb{P})$.

The number of summands in the three left sides above is less than $K/p\sqrt{\Delta_n}$, as already seen above, so (A.29) follows from (A-5.41). Next, exactly the same arguments as in parts (3)–(5) of the proof of Lemma A-5.7, yield (A.30).

We now turn to (A.28). By combining (A.20) and (A.21) and by (A.9), we obtain

$$\begin{aligned}
&\Big|\frac{1}{\sqrt{\Delta_n}}\,\mathbb{E}((\rho_j^n)^2\mid \mathcal{G}_{j-1}^n)-(\overline{\rho}_j^n)^2-a_j^n\Big|\leq K\sqrt{\Delta_n}\,\overline{\chi}(p)_j^n, \quad \text{where}\\
&u_n(r-1)+1\leq j\leq u_n(r)\Rightarrow\begin{cases}\overline{\chi}(p)_j^n=\chi(p)^n_{J(n,r-1)+(j-1-u_n(r-1))k_{n,r}(p+1)}\\ a_j^n=\Delta_n k_{n,r}\,H_p(\widehat{\theta}_r^n, c_{t_j^n},\gamma_{t_j^n})\end{cases}\\
&H_p(x,y,z) \;=\; \frac{4}{\phi(g)^2}\Big((p\,\Phi(g,g)+\Xi_{22})xy^2+2(p\,\Phi(g,g')+\Xi_{12})\frac{yz}{x}\\
&\qquad\qquad\qquad +(p\,\Phi(g',g')+\Xi_{11})\frac{z^2}{x^3}\Big).
\end{aligned}$$

Exactly as in Lemma A.4, we have $\sqrt{\Delta_n}\,\sum_{j=1}^{u_n(L_n)}\overline{\chi}(p)_j^n \stackrel{\mathbb{P}}{\longrightarrow} 0$. Therefore it is enough to show

$$\sum_{j\geq 1:\,t_j^n\leq s} a_j^n \stackrel{\mathbb{P}}{\longrightarrow}\int_0^s \Gamma(p)_v\,dv. \tag{A.31}$$

Let us also define, in analogy with a_j^n and when $u_n(r-1)+1\leq j\leq u_n(r)$:

$$a_j'^n=\Delta_n k_{n,r}\,H_p(\overline{\theta}_r^n, c_{t_j^n},\gamma_{t_j^n}),\qquad a_j''^n=\Delta_n k_{n,r}\,H_p(\theta_{(r-1)l_n}, c_{(r-1)l_n},\gamma_{(r-1)l_n}).$$

Upon using the boundedness of c, γ, $1/c$, $1/\gamma$ and (A.9) we see that $|a_j^n - a_j'^n|\leq K\,\Delta_n^{1/2+w'}$ for all $j > u_n(1)$ and also $|a_j^n|\leq K\,\Delta_n^{1/2}$ always, hence instead of (A.31) it is enough to show

$$\sum_{j\geq u_n(1):\,t_j^n\leq s} a_j'^n \stackrel{\mathbb{P}}{\longrightarrow}\int_0^s \Gamma(p)_v\,dv. \tag{A.32}$$

Now we use (4.1) with $\tau_1=\infty$ and (4.3) to get $\mathbb{E}(|a_j'^n - a_j''^n|)\leq K\,\Delta_n^{1-w/2}$, which is $o(\sqrt{\Delta_n}\,)$: so we can replace $a_j'^n$ by $a_j''^n$ in (A.32). Then we observe that $\Gamma(p)_s = \frac{1}{p+1}\,H_p(\theta_s, c_s, \gamma_s)$. In other words, (A.32) amounts to

$$\sum_{r=2}^{L_n}(p+1)k_{n,r}\, j'_n(r,p,s)\,\Delta_n\,\Gamma(p)_{(r-1)l_n\Delta_n} \to \int_0^s \Gamma(p)_v\,dv, \tag{A.33}$$

where $j'_n(r,p,s) = j_n(r,p)\wedge \sup(j : t^n_{u_n(r-1)+j} \le s)$. The left side above is a kind of Riemann sum: indeed, with $s^n_r = (r-1)l_n\Delta_n$, one can rewrite this left side as

$$\sum_{r=2}^{L_n}(s^n_r\wedge s - s^n_{r-1}\wedge s)\,\Gamma(p)_{s^n_r} + \mathrm{O}_{Pu}\big(l_n\Delta_n + L_n(p+1)\sqrt{\Delta_n}\big)$$

(use the boundedness of $\Gamma(p)_s$ and (A.9) and the definition of $j_n(r,p)$). Then (A.33) follows by Riemann integration, because $\Gamma(p)_s$ is càdlàg. $\square$

At this stage, proving the stable convergence of (4.11) to the proper limit is done exactly as in [18]. Namely, we have $\frac{1}{\Delta_n^{1/4}}(\widehat{C}^n_T - C_T) = \frac{1}{\Delta_n^{1/4}}\overline{M}(p)^n_T + U(p)^n$, where by virtue of (A.15) and (A.25) and Lemma A.4 we have for all $\varepsilon > 0$:

$$\lim_{p\to\infty}\ \limsup_{n\to\infty}\ \mathbb{P}(|U(p)^n| > \varepsilon) = 0.$$

Therefore, by a standard argument it is enough to show that, as $p\to\infty$, we have $\Gamma(p)_s \to 8\alpha(g)\sqrt{c_s^3\gamma_s}$ for all s. In view of (A.27) we have

$$\Gamma(p)_s \to \frac{4}{\phi(g)^2}\left(\Phi(g,g)\theta_s c_s^2 + 2\Phi(g,g')\frac{c_s\gamma_s}{\theta_s} + \Phi(g',g')\frac{\gamma_s^2}{\theta_s^3}\right),$$

which by (3.5) and (3.8) and a simple computation yields the result.

To end the proof of the theorem, it remains to show that $\frac{1}{\sqrt{\Delta_n}}\widehat{V}^n_T$ (recall the modified version (A.3)) converges in probability to $V'(g)_T$. Observe that in fact $\widehat{V}^n_T$ has exactly the same structure as the estimator for the variance in [18], except that we split the sum into L_n blocks which are again of the same type, with $k_{n,r}$ instead of k_n. Then it suffices to follow the argument of that paper and the result holds.

References

[1] R. Altmeyer, M. Bibinger, Functional stable limit theorems for efficient spectral covolatility estimators, Preprint, 2014.

[2] T.G. Andersen, T. Bollerslev, Intraday periodicity and volatility persistence in financial markets, J. Empir. Finance 4 (1997) 115–158.

[3] T.G. Andersen, T. Bollerslev, F.X. Diebold, P. Labys, Modeling and forecasting realized volatility, Econometrica 71 (2003) 579–625.

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

[5] F.M. Bandi, J.R. Russell, Separating microstructure noise from volatility, J. Financ. Econ. 79 (2006) 655–692.

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

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

[8] M. Bibinger, N. Hautsch, P. Malec, M. Reiss, Estimating the spot covariation of asset prices: statistical theory and empirical evidence, Preprint, 2014.

[9] M. Bibinger, P.A. Mykland, Inference for multi-dimensional highfrequency data: equivalence of methods, central limit theorems, and an application to conditional independence testing. Technical Report, The University of Chicago, 2013.
[10] M. Bibinger, M. Reiss, Spectral estimation of covolatility from noisy observations using local weights, Scand. J. Statist. 41 (1) (2014) 23–50.
[11] E. Clément, S. Delattre, A. Gloter, An infinite dimensional convolution theorem with applications to the efficient estimation of the integrated volatility, Stochastic Process. Appl. 123 (2013) 2500–2521.
[12] G. Curci, F. Corsi, Discrete sine transform for multi-scales realized volatility measures, Quant. Finance 12 (2) (2012) 263–279.
[13] M. Fukasawa, Central limit theorems for the realized volatility based on tick time sampling, Finance Stoch. 14 (2010) 209–233.
[14] M. Fukasawa, Realized volatility with stochastic sampling, Stochastic Process. Appl. 120 (2010) 829–852.
[15] M. Fukasawa, M. Rosenbaum, Central limit theorems for realized volatility under hitting times of an irregular grid, Stochastic Process. Appl. 122 (12) (2012) 3901–3920.
[16] A. Gloter, J. Jacod, Diffusions with measurement errors. I—local asymptotic normality, ESAIM Probab. Stat. 5 (2001) 225–242.
[17] A. Gloter, J. Jacod, Diffusions with measurement errors. II—optimal estimators, ESAIM Probab. Stat. 5 (2001) 243–260.
[18] J. Jacod, Y. Li, P.A. Mykland, M. Podolskij, M. Vetter, Microstructure noise in the continuous case: the pre-averaging approach, Stochastic Process. Appl. 119 (2009) 2249–2276.
[19] J. Jacod, M. Podolskij, M. Vetter, Limit theorems for moving averages of discretized processes plus noise, Ann. Statist. 38 (2009) 1478–1545.
[20] J. Jacod, P. Protter, Discretization of Processes, Springer-Verlag, New York, 2012.
[21] J. Jacod, P. Protter, Asymptotic error distributions for the Euler method for stochastic differential equations, Ann. Probab. 26 (1998) 267–307.
[22] J. Jacod, M. Reiss, A remark on the rates of convergence for integrated volatility estimations in the presence of jumps, Ann. Statist. 42 (2014) 1131–1144.
[23] Y. Koike, Time endogeneity and an optimal weight function in pre-averaging covariance estimation, Preprint, University of Tokyo, 2014.
[24] Y. Li, P.A. Mykland, E. Renault, L. Zhang, X. Zheng, Realized volatility when sampling times are possibly endogenous, Econometric Theory 30 (2014) 580–605.
[25] P.A. Mykland, N. Shephard, K. Sheppard, Efficient and feasible inference for the components of financial variation using blocked multipower variation. Technical Report, University of Oxford, 2012.
[26] P.A. Mykland, L. Zhang, Inference for continuous semimartingales observed at high frequency, Econometrica 77 (2009) 1403–1455.
[27] P.A. Mykland, L. Zhang, ANOVA for diffusions and Itô processes, Ann. Statist. 34 (2006) 1931–1963.
[28] P.A. Mykland, L. Zhang, The econometrics of high frequency data, in: M. Kessler, A. Lindner, M. Sörensen (Eds.), Statistical Methods for Stochastic Differential Equations, Chapman and Hall/CRC Press, New York, 2012, pp. 109–190.
[29] M. Podolskij, M. Vetter, Estimation of volatility functionals in the simultaneous presence of microstructure noise and jumps, Bernoulli 15 (2009) 634–658.
[30] M. Reiss, Asymptotic equivalence for inference on the volatility from noisy observations, Ann. Statist. 2 (2011) 772–802.
[31] E. Renault, B.J. Werker, Causality effects in return volatility measures with random times, J. Econometrics 160 (2011) 272–279.
[32] D. Xiu, Quasi-maximum likelihood estimation of volatility with high frequency data, J. Econometrics 159 (2010) 235–250.
[33] L. Zhang, From martingales to ANOVA: implied and realized volatility (Ph.D. thesis), The University of Chicago, Department of Statistics, 2001.
[34] L. Zhang, Effcient estimation of stochastic volatility using noisy observations: a multi-scale approach, Bernoulli 12 (2006) 1019–1043.
[35] L. Zhang, P.A. Mykland, Y. Ait-Sahalia, A tale of two time scales: determining integrated volatility with noisy high-frequency data, J. Amer. Statist. Assoc. 100 (2005) 1394–1411.