Nonparametric estimation of the volatility under microstructure noise: wavelet adaptation

Hoffmann, Marc and Munk, Axel and Schmidt-Hieber, Johannes

ENSAE and CNRS, Paris, University of Göttingen, University of Göttingen

27 July 2010

Nonparametric estimation of the volatility under microstructure noise: wavelet adaptation

M. Hoffmann[*], A. Munk [†] and J. Schmidt-Hieber [†]

Abstract

We study nonparametric estimation of the volatility function of a diffusion process from discrete data, when the data are blurred by additional noise. This noise can be white or correlated, and serves as a model for microstructure effects in financial modeling, when the data are given on an intra-day scale. By developing pre-averaging techniques combined with wavelet thresholding, we construct adaptive estimators that achieve a nearly optimal rate within a large scale of smoothness constraints of Besov type. Since the underlying signal (the volatility) is genuinely random, we propose a new criterion to assess the quality of estimation; we retrieve the usual minimax theory when this approach is restricted to deterministic volatility.

Keywords: Adaptive estimation; Besov spaces; diffusion processes; minimax estimation; nonparametric regression; semimartingale; wavelets.
Mathematical Subject Classification: 62G99; 62M99; 60G99 .

[*]ENSAE and CNRS-UMR 8050, 3, avenue Pierre Larousse, 92245 Malakoff Cedex, France.

[†]Institut für Mathematische Stochastik, Universität Göttingen, Goldschmidtstr. 7, 37077 Göttingen, Germany.

1 Introduction

1.1 Motivation

Microstructure noise

Diffusion processes and more generally continuous-time semimartingales have long served as a representative model for financial assets in order to hedge and replicate risk of derivatives (see for instance the books of Musiela [36] or Bouchaud and Potters [9] and the references therein). When mathematical modeling narrows down to parameter estimation or calibration based on historical prices, the time scale at which the models are displayed becomes a key factor. Whereas relative consensus holds about a general semimartingale model for prices at coarse scales (when the data are sampled on a daily or monthly basis) this is no longer true at fine scales, when intra-day or high-frequency data are concerned.

Over the last years, financial econometrics have covered a giant leap since the naive models of discretized diffusions that were used before the 2000's. The seminal paper of Ait-Sahalia *et al.* [2] led the way: by considering high-frequency data as the result of a *latent* or unobservable efficient price corrupted by microstructure effects, they obtained a more realistic model accounting for stylized facts in the intraday scale usually attributed to bid-ask spread manipulation by market makers. This approach was grounded on empirical findings in the financial econometrics literature of the early years 2000 (among many others, Andersen *et al.* [4], Engl [17], Mykland and Zhang [37]) and even before (Roll [40] and Hasbrouck [27]).

In this setting and for 1-dimensional models, observable quantities (e.g. the log-returns of an asset) are assumed to take the form

$$Z_{j,n} = X_{j\Delta_n} + \epsilon_{j,n}, \quad j = 0, 1, \dots, n \tag{1.1}$$

where $\Delta_n > 0$ is the sampling time, $(\epsilon_{j,n})$ is the microstructure noise process (always taken with 0 expectation for obvious identifiability purposes). The process $X = (X_t)_{t\geq 0}$ is the latent price and has representation

$$X_t = X_0 + \int_0^t b_s\, ds + \int_0^t \sigma_s\, dW_s, \tag{1.2}$$

on an appropriate probability space. In other words, X is an Itô continuous semimartingale driven by a Brownian motion $W = (W_t)_{t\geq 0}$ with drift $b = (b_t)$ and diffusion coefficient or volatility process $\sigma = (\sigma_t)$. This is the so-called *additive microstructure noise model* (later abbreviated by AMN).

Admittedly, there is no general consensus for the quality of modeling provided by AMN, and indeed, representation (1.1) overlooks some obvious stylized facts such as the discreteness of prices when scrutinizing data at the level of the order book. However, AMN should be viewed as a simplified but still instructive model for addressing statistical inference in the context of high-frequency data in finance. The relatively weak assumptions we have on X and the microstructure noise $(\epsilon_{j,n})$ further on will be sufficient for the level of generalization intended in this paper.

Statistical inference under microstructure noise

The parameter of interest is the unobserved path of the volatility process $t \rightsquigarrow \sigma_t$, and unless specified otherwise, it is random. From a semiparametric statistical perspective, a commonly admitted purpose is to estimate from data $(Z_{j,n})$ integrated quantities such as the integrated volatility $\int_0^t \sigma_s^2 ds$ and the integrated quarticity $\int_0^t \sigma_s^4 ds$. The high frequency data framework dictates to take asymptotics as the time step Δ_n between observations goes to 0.

From a nonparametric angle, one can try to recover the whole path $t \rightsquigarrow \sigma_t^2$ from data $(Z_{j,n})$ solely. Whereas nonparametric estimation of the diffusion coefficient from direct observation $X_{j\Delta_n}$ is a fairly well known topic when σ is assumed to be deterministic ([19], [28] and the review paper of Fan [18]), nonparametric estimation in the presence of the noise $(\epsilon_{j,n})$ substantially increases the difficulty of the statistical problem. This is the topic of the present paper.

1.2 Statistical inference under microstructure noise: some history

Parametric and semiparametric inference

The first results about statistical inference of a diffusion with error measurement go back to Gloter and Jacod [23, 24] in 2001. They showed that if $\sigma_t = \sigma(t, \vartheta)$ is a deterministic function known up to a 1-dimensional parameter ϑ, and if moreover the $\varepsilon_{j,n}$ are Gaussian and independent, then, for $\Delta_n = n^{-1}$, the LAN condition holds (Local Asymptotic Normality), with the rate $n^{-1/4}$. This shows that, even in the simplest Gaussian diffusion case, there is a substantial loss of information compared to the case without noise, where the standard $n^{-1/2}$ accuracy of estimation is achievable.

At about the same time, the microstructure noise model for financial data was introduced by Ait-Sahalia, Mykland and Zhang in a series of papers [2, 44, 43]. Analogous approaches in various similar contexts progressively emerged in the financial econometrics literature: Podolskij and Vetter [38], Bandi and Russell [6, 5], Barndorff-Nielsen *et al.* [7] and the references therein. These studies tackled estimation problems in a sound mathematical framework, and incrementally gained in generality and elegance.

A paradigmatic problem in this context is the estimation of the integrated volatility $\int_0^t \sigma_s^2 ds$. Convergent estimators were first obtained by Ait-Sahalia *et al.* [2] with a suboptimal rate $n^{-1/6}$. Then the two-scale approach of Zhang [43] achieved the rate $n^{-1/4}$. The Gloter-Jacod LAN property of [23] for deterministic submodels shows that this cannot be improved. Further generalizations took the way of extending the nature of the latent price model X (for instance [3, 42, 14]) and the nature of the microstructure noise $(\epsilon_{j,n})$.

It took some more time and contributions before Jacod and collaborators [30] took over the topic in 2007 with their simple and powerful pre-averaging technique, introduced earlier in a simplified context by Podolskij and Vetter [38]. The approach of Jacod *et al.* is a nice balance between simplicity of modeling and generality, and it substantially improves on previous results in AMN. In essence, it consists in first, smoothing the data as in signal denoising and then, apply a standard realized volatility estimator up to appropriate bias correction. Stable convergence in law is displayed for a wide class of pre-averaged estimators in a fairly general setting, closing somehow the issue of estimating the integrated volatility in a semiparametric setting.

Nonparametric inference

In the nonparametric case, the problem is a little unclear. By nonparametric, one thinks of estimating the whole path $t \rightsquigarrow \sigma_t^2$ of the volatility of the latent price, under microstructure noise. This is a problem of major importance in high-frequency trading: a key issue is the recovery of a volatility profile (possibly in an on-line fashion) over which various trading indicators are constructed. However, since $\sigma^2 = (\sigma_t^2)_{t\geq 0}$ is usually itself genuinely random, there is no "true parameter" to be estimated! However, when the diffusion coefficient is deterministic, the usual setting of statistical experiments is recovered. In that latter case, under the restriction that the microstructure noise process consists of i.i.d. noises, Munk and Schmidt-Hieber [34, 35] proposed a Fourier estimator and showed its minimax rate optimality, extending a previous approach for the parametric setting ([10]). This approach relies on a formal analogy with inverse ill-posed problems. When the microstructure noises $(\epsilon_{j,n})$ are Gaussian i.i.d. with variance τ^2, Reiß [39] showed very recently the asymptotic equivalence in the Le Cam sense with the observation of the random measure Y_n given by

$$Y_n = \sqrt{2\sigma} + \tau\, n^{-1/4}\dot{B}$$

where $\dot{B}$ is a Gaussian white noise. This is a beautiful and deep result, and again the semiparametric rate $n^{-1/4}$ is illuminating when compared with the optimality results obtained by previous authors.

1.3 Our results

The asymptotic equivalence proved by Reiß [39] provides us with a benchmark for the complexity of the statistical problem and is inspiring: we target in this paper to put the problem of estimating nonparametrically the random parameter $t \rightsquigarrow \sigma_t^2$ to the level of classical denoising in the adaptive minimax theory. In spirit, we follow the classical route of nonlinear estimation in de-noising, but we need to introduce new tools. Our procedure is twofold:

1. We approximate the random signal $t \rightsquigarrow \sigma_t^2$ by an atomic representation
$$\sigma_t^2 \approx \sum_{\nu\in\mathcal{V}(\sigma^2)} \langle \sigma^2, \psi_\nu\rangle \psi_\nu(t) \tag{1.3}$$
where $\langle \bullet, \bullet\rangle$ denotes the usual L^2-inner product and $\big(\psi_\nu, \nu\in\mathcal{V}(\sigma)\big)$ is a collection of wavelet functions that are localized in time and frequency, indexed by the set $\mathcal{V}(\sigma^2)$ that depends on the path $t \rightsquigarrow \sigma_t^2$ itself. As for the precise meaning of the symbol $\approx$ we do no specify yet.

2. We then estimate the linear coefficients $\langle \sigma^2, \psi_\nu \rangle$ and specify a selection rule for $\mathcal{V}(\sigma)$ (with the dependence in σ somehow replaced by an estimator). The rule is dictated by hard thresholding over the estimations of the coefficients $\langle \sigma^2, \psi_\nu \rangle$ that are kept only if they exceed some noise level, tuned with the data, as in standard wavelet nonlinear approximation (among many others, the work of Donoho, Johnstone, Kerkyacharian, Picard and collaborators [15, 16, 26]).

The key issue is therefore the estimation of integrated quantities of the form

$$\langle \sigma^2, \psi_\nu \rangle = \int_{\mathbb{R}} \sigma_t^2 \psi_\nu(t) dt. \tag{1.4}$$

An important fact is that the functions ψ_ν are well located but oscillate, making the approximation of (1.4) delicate, in contrast to the global estimation of the integrated volatility in the semiparametric approach. If we could observe the latent process X itself at times $j\Delta_n$, then standard quadratic variation based estimators like

$$\sum_j \psi_\nu(j\Delta_n)(X_{j\Delta_n} - X_{(j-1)\Delta_n})^2 \tag{1.5}$$

would give rate-optimal estimators of (1.4), as follows from standard results on nonparametric estimation in diffusion processes [19, 28, 29]. However, we only have a blurred version of X via $(Z_{j,n})$ and further "intermediate" de-noising is required.

At this stage, we consider "local averages" of the data $Z_{j,n}$ at an "intermediate scale" m so that $\Delta_n \ll 1/m$ but $m \to \infty$. Let us denote loosely (and temporarily) by $\text{Ave}(Z)_{i,m}$ an averaging of the data $(Z_{j,n})$ around the point i/m. We have

$$\text{Ave}(Z)_{i,m} \approx X_{i/m} + \text{small noise} \tag{1.6}$$

and thus we have a "de-blurred" version of X, except that we must now handle the small noise term of (1.6) and the loss of information due to the fact that we dispose of (approximate) $X_{i/m}$ on a coarser scale since $m \ll \Delta_n^{-1}$. We subsequently estimate (1.4) replacing the naive guess (1.5) by

$$\sum_i \psi_\nu(i\Delta_n)\big[\big(\text{Ave}(Z)_{i,m} - \text{Ave}(Z)_{i-1,m}\big)^2 + \text{bias correction}\big] \tag{1.7}$$

up to a further "bias correction" term that comes from the fact that we take square approximation of X via (1.6). In Section 3.1, we generalize

(1.7) to arbitrary kernels within a certain class of pre-averaging functions, in the very same spirit as in Jacod *et al.* [30]. (See also Gloter [20] and Gloter and Hoffmann [21] or Rosenbaum [41] where this technique is used for denoising stochastic volatility models corrupted by noise.)

Tuning appropriately all the parameters within this class of estimators, we prove in Theorems 3.4 and 2.9 that over fixed time horizon, the rate $(n^{-1/4})^{2s(\pi^\star)/(4s(\pi^\star)+2)}$ is achievable by our procedure in L^p-loss error, when the signal $t \rightsquigarrow \sigma_t^2$ has smoothness s measured in L^π-norm over Besov classes $\mathcal{B}^s_{\pi,\infty}$. The smoothness parameter $s(\pi^\star) = s$ in most cases, except if the loss L^p is too strong in face of the L^π-smoothness measurement, in which case we have an inflation of the convergence rate governed by s, p and π (see the definition of $s(\pi^\star)$ in (2.9) in Theorem 2.9). This is the classical case of minimax estimation of sparse signals [16, 32] and we retrieve the expected results of wavelet thresholding up to the noise rate $n^{-1/4}$ instead of the usual $n^{-1/2}$ in white Gaussian noise or density estimation, but that is inherent to the problem of microstructure noise, as already established in [23]. A major difficulty is that in order to employ the wavelet theory in this context, we must assess precise deviation bounds for quantities of the form (1.7), which require delicate martingale techniques.

We prove in Theorem 2.11 that this result is sharp, even if the signal $t \rightsquigarrow \sigma_t^2$ is random and that we do not have a statistical model in the strict sense. In order to encompass this level of generality, we propose a modification of the notion of upper and lower rate of estimation of a random parameter in Definition 2.3 and 2.6. This approach is presented in details in the methodology Section 2.2. The proof of our lower bound heavily relies on the result of Reiß [39] already mentioned in Section 1.2, which we take advantage of thanks to the asymptotic equivalence theory of Le Cam [33].

The paper is organized as follows. In Section 2 we introduce notation and formulate the key results. An explicit construction of the estimator can be found in Section 3. Finally, the proofs of the main results and some (unavoidable) technicalities are deferred to Section 4.

2 Main results

2.1 The data generating model

The setting

In the following, we consider a filtered probability space $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t\geq 0}, \mathbb{P})$, on which is defined a continuous adapted 1-dimensional process X of the form (1.2), with X_0 is a random variable and W a Wiener process. Without loss of generality, we will assume further that $X_0 = 0$.

The following basic assumption on b and σ is in force throughout the paper.

Assumption 2.1. *The processes σ and b are càdlàg (right continuous with left limits), $\mathcal{F}_t$-adapted, and a weak solution of (1.2) is unique and well defined.*

Moreover, a weak solution to $Y_t = \int_0^t \sigma_s dW_s$ is also unique and well defined, the laws of X and Y are equivalent on $\mathcal{F}_1$ and we have, for some $\rho > 1$

$$\mathbb{E}\big[\exp\big(\rho \int_0^1 \frac{b_s}{\sigma_s^2} dY_s\big)\big] < \infty.$$

We mention, that the second part of Assumption 2.1 will prove technically useful, since it allows to assume that $b = 0$ in many estimates due to Girsanov's theorem. With some more technical effort it could be relaxed further.

The data generating process

For $j = 0, \ldots, n$, we assume that we can observe a blurred version of X a times $\Delta_n j$ over the time horizon $[0, T]$. We consider a fixed time horizon $T = n\Delta_n$, and with no loss of generality, we take $T = 1$ hence $\Delta_n = n^{-1}$.

The blurring accounts for microstructure noise at fine scales and then takes the form

$$Z_{j,n} := X_{j/n} + \epsilon_{j,n}, \quad j = 0, 1, \ldots, n \tag{2.1}$$

where the microstructure noise process $(\epsilon_{j,n})$ is implicitly defined on the same probability space as X and is allowed to be price dependent and correlated. More precisely

Assumption 2.2. *We have*

$$\epsilon_{j,n} = a(j/n, X_{j/n})\eta_{j,n}, \tag{2.2}$$

where the function $(t,x) \rightsquigarrow a(t,x)$ *is continuous and bounded. Moreover, the noise array* $(\eta_{j,n})$ *is independent of* X *and for every* $1 \leq j \leq n$ *and* $n \geq 1$*, we have*

$$\mathbb{E}\left[\eta_{j,n}\right] = 0,\ \ \mathbb{E}\left[\eta_{j,n}^2\right] = 1,\ \ \mathbb{E}\left[|\eta_{j,n}|^p\right] < \infty,\ p > 0.$$

The estimation program

Given data $Z_\bullet = \{Z_{j,n},\, j = 0, \ldots, n\}$ following (1.1), we estimate non-parametrically the random function $t \rightsquigarrow \sigma_t^2$ over the time interval $[0,1]$. Asymptotics are taken as the observation frequency $n \to \infty$.

2.2 Statistical methodology

Strictly speaking, since the target parameter $\sigma^2 = (\sigma_t^2)_{t\in[0,1]}$ is random itself (as an $\mathcal{F}$-adapted process), we cannot assess the performance of an "estimator of σ^2" in the usual way. We need to modify the usual notion of convergence rate over a function class.

We are interested in recovering σ^2 over various smoothness classes, that shall account for the underlying complexity of the volatility process $t \rightsquigarrow \sigma_t^2$. Theses smoothness class include the case where σ^2 is deterministic and has as many derivatives as one wishes, but also the case of genuinely random processes that oscillate like diffusions, or fractional diffusions and so on. We shall describe smoothness classes in terms of Besov balls

$$\mathcal{B}^s_{\pi,\infty}(c) := \big\{f : [0,1] \to \mathbb{R},\ \|f\|_{B^s_{\pi,\infty}([0,1])} \leq c\big\},\ \ c > 0. \tag{2.3}$$

that measure smoothness of degree $s > 1/\pi$ in L^π over the interval $[0,1]$, for $\pi \in (0,\infty)$. A thorough account of Besov spaces $B^s_{p,\infty}$ and their connection to wavelet bases in a statistical setting are discussed in details in the classical papers of Donoho *et al.* [16] and Kerkyacharian and Picard [32]. See also the textbook of Cohen [12]. The restriction $s > 1/\pi$ ensures that the functions in $B^s_{\pi,\infty}$ are continuously embedded into Hölder continuous functions with index $s - 1/\pi$.

Definition 2.3. *An estimator of* $\sigma^2 = (\sigma_t^2)_{t\in[0,1]}$ *is a random function*

$$t \rightsquigarrow \widehat{\sigma}_n^2(t),\ \ t \in [0,1],$$

measurable with respect to the observation $(Z_{j,n})$ *defined in* (1.1).

We have the following notion of upper bound:

Definition 2.4. *We say that the rate* $0 < v_n \to 0$ *(as* $n \to \infty$*) is achievable for estimating* σ^2 *in* L^p*-norm over* $\mathcal{B}^s_{\pi,\infty}(c)$ *if there exists an estimator* $\widehat{\sigma}^2_n$ *such that*

$$\limsup_{n\to\infty} v_n^{-1}\, \mathbb{E}\left[\|\widehat{\sigma}^2_n - \sigma^2\|_{L^p([0,1])} \mathbb{I}_{\left\{\sigma^2 \in \mathcal{B}^s_{\pi,\infty}(c)\right\}}\right] < \infty. \tag{2.4}$$

Remark 2.5. *When* (σ_t) *is deterministic, we can make a priori assumptions so that the condition* $\sigma^2 \in \mathcal{B}^s_{\pi,\infty}(c)$ *is satisfied, in which case we simply ignore the indicator in* (2.4)*. In other cases, this condition will be satisfied with some probability for some indices* (s,π) *and* $c > 0$*. For instance, is* (σ_t) *is an Itô continuous semimartingale itself with regular coefficients, we have* $\mathbb{P}\left[\sigma^2 \in B^{1/2}_{\pi,\infty}(c)\right] > 0$ *for every* $0 < \pi < \infty$*, see [11]. But it may also well happen that for some choices of* (s,π) *we have* $\mathbb{P}\left[\sigma^2 \in B^s_{\pi,\infty}(c)\right] = 0$ *for every* $c > 0$ *in which case, the upper bound* (2.4) *becomes trivial and noninformative.*

In this context, a sound notion of optimality is a little unclear. We introduce the following type of lower bound.

Definition 2.6. *The rate* v_n *is a lower rate of convergence over* $\mathcal{B}^s_{\pi,\infty}(c)$ *in* L^p *norm if there exists a filtered probability space* $(\widetilde{\Omega}, \widetilde{\mathcal{F}}, (\widetilde{\mathcal{F}}_t)_{t\geq 0}, \widetilde{\mathbb{P}})$*, a process* $\widetilde{X}$ *defined on* $(\widetilde{\Omega}, \widetilde{\mathcal{F}})$ *with the same distribution as* X *under Assumptions 2.1 together with a process* $(\widetilde{\epsilon}_{j,n})$ *satisfying* (2.2) *with* $\widetilde{X}$ *in place of* X*, such that Assumption 2.2 holds, and moreover:*

$$\widetilde{\mathbb{P}}\big[\sigma^2 \in \mathcal{B}^s_{\pi,\infty}(c)\big] > 0 \tag{2.5}$$

and

$$\liminf_{n\to\infty} v_n^{-1} \inf_{\widehat{\sigma}^2_n} \widetilde{\mathbb{E}}\Big[\|\widehat{\sigma}^2_n - \sigma^2\|_{L^p([0,1])} \mathbb{I}_{\left\{\sigma^2 \in \mathcal{B}^s_{\pi,\infty}(c)\right\}}\Big] > 0, \tag{2.6}$$

where the infimum is taken over all estimators.

Let us elaborate on Definition 2.6. As already mentioned in this section, the quantity of interest σ^2 is "genuinely" random, and we cannot say that our data $\{Z_{j,n}\}$ generate a statistical experiment as a family of probability measures indexed by some parameter of interest, and over which standard information criteria such as Fisher information or minimax lower bound of estimation could be computed. Rather, we have a fixed probability measure $\mathbb{P}$, but this measure is only "loosely" specified by very weak conditions, namely Assumptions 2.1 and 2.2. A lower bound as in Definition 2.6 says that, given model $\mathbb{P}$, there exists a probability measure $\widetilde{\mathbb{P}}$, possibly defined on another space so that Assumptions 2.1 and 2.2 hold under $\widetilde{\mathbb{P}}$ together with (2.6).

Without further specification on our model, there is no sensible way to discriminate between $\mathbb{P}$ and $\widetilde{\mathbb{P}}$ since both measures (and the accompanying processes) satisfy Assumptions 2.1 and 2.2; moreover, under $\widetilde{\mathbb{P}}$, we have a lower bound.

Remark 2.7. *This setting may enable to retrieve the standard minimax framework when σ^2 is deterministic and belongs to a Besov ball $B^s_{\pi,\infty}(c)$. In that case, it suffices to construct a probability measure $\widetilde{\mathbb{P}}$ such that under $\widetilde{\mathbb{P}}$, the random variable σ^2 has distribution $\mu(d\sigma^2)$ with support in $B^s_{\pi,\infty}(c)$, and chosen to be a least favourable prior as in standard lower bound nonparametric techniques. It remains to check that Assumptions 2.1 and 2.2 are satisfied μ-almost surely. We elaborate on this approach in the proof of Theorem 2.11 below.*

2.3 Achievable estimation error bounds

An important fact for practical purposes is that we shall require only relatively weak prior knowledge on the smoothness of the volatility process. A (technical) restriction is that we assume some minimal Hölder smoothness on the paths of $t \rightsquigarrow \sigma_t^2$. This is guaranteed by the condition $s > 1/\pi$.

For prescribed smoothness class $\mathcal{B}^s_{\pi,\infty}(c)$ and L^p-loss functions, the rate of convergence v_n depends on the index s, π and p. As usual in this setting ([15], [16] and [32]), we have an "elbow" phenomenon that separates sparse and dense regime, according to the classical terminology. We describe the rate of convergence in a condensed way, by introducing an "effective smoothness function" as in [22] (among other possible references for that topic).

Definition 2.8. *For $\pi \in (0, +\infty)$ and $s > 1/\pi$, the effective smoothness function relative to $\mathcal{B}^s_{\pi,\infty}(c)$ is*

$$s(t) := s + \big(t - 1/\pi\big)\mathbb{I}_{\{0 \le t \le 1/\pi\}} \tag{2.7}$$

Theorem 2.9. *Work under Assumptions 2.1 and 2.2. Then, for every $c > 0$, the rate*

$$v_n := \left(\frac{(\log n)^{3/2}}{n} \right)^{s(\pi^\star)/(4s(\pi^\star)+2)} \tag{2.8}$$

where $\pi^\star$ is the unique number that satisfies

$$s(\pi^\star) = \frac{1}{2}\left(\frac{p}{\pi^\star} - 1\right), \tag{2.9}$$

is achievable over the class $\mathcal{B}^s_{\pi,\infty}(c)$ in L^p norm with $p \in [1, \infty)$, provided $s > 1/\pi$ and $\pi \in (0, \infty)$. Moreover, the estimator explicitly constructed in Section 3.3 below attains this bound in the sense of (2.4).

Remark 2.10. *The rate of convergence of σ^2 over the class $\mathcal{B}^s_{\pi,\infty}(c)$ under L^p-loss has the form*

$$\left(n^{-1/4}\right)^{2s(\pi^\star)/(2s(\pi^\star)+1)}$$

up to some logarithmic (inessential) corrections in some cases. It is to be compared with the usual minimax rate

$$\left(n^{-1/2}\right)^{2s/(2s+1)}$$

for recovering a function of Hölder smoothness s from n data in density estimation or nonparametric regression.

First, the "effective smoothness" $s(\pi^\star) \leq s$ instead of s comes from the fact that we measure smoothness in a weaker L^π sense, and if the loss L^p we pick is such that p is substantially larger than π, we may loose a smoothness factor. This is precisely quantified by $s(\pi^\star)$, where $\pi^\star$ is defined in (2.9). It is a well known phenomenon in nonlinear nonparametric estimation [16, 32] (or [22] for a systematic use of the "effective smoothness" function $s(\bullet)$), and microstructure noise model are no exception of it.

Second, the parametric rate $n^{-1/2}$ has to be replaced by $n^{-1/4}$. This effect is due to microstructure noise, and was already identified in earlier parametric models as in Gloter and Jacod [23] and subsequent works, both in parametric, semiparametric and nonparametric estimation (see [23, 24, 10, 34, 43, 30]).

Our next result shows that this rate is nearly optimal in some (most) cases.

Theorem 2.11. *In the same setting as in Theorem 2.9, assume moreover that $s - 1/\pi > \frac{1+\sqrt{5}}{4}$. Then the rate*

$$\widetilde{v}_n := n^{-s(\pi^\star)/(4s(\pi^\star)+2)}$$

is a lower rate of convergence over $\mathcal{B}^s_{\pi,\infty}(c)$ in L^p in the sense of Definition 2.6.

Since v_n and $\widetilde{v}_n$ agree up to some (inessential) logarithmic terms, our result is nearly optimal in the sense of Definitions 2.4 and 2.6

Remark 2.12. *This setting may also enable to retrieve the standard minimax framework when σ^2 is deterministic and belongs to a Besov ball $B^s_{\pi,\infty}(c)$. In that case, it suffices to construct a probability measure $\widetilde{\mathbb{P}}$ such that under $\widetilde{\mathbb{P}}$, the random variable σ^2 has distribution $\mu(d\sigma^2)$ with support in $B^s_{\pi,\infty}(c)$, and chosen to be a least favourable prior as in standard lower bound nonparametric techniques. It remains to check that Assumptions 2.1 and 2.2 are satisfied μ-almost surely. We elaborate on this approach in the proof of Theorem 2.11 below.*

The proof of the lower bound is an application of a recent result of Reiß [39] about asymptotic equivalence between the statistical model obtained by letting σ^2 be deterministic and the microstructure noise white Gaussian with an appropriate infinite dimensional Gaussian shift experiment. In particular, the restriction $s - 1/\pi > \frac{1+\sqrt{5}}{4}$ stems from the result of Reiß and could presumably be improved. Our proof relies on the strategy described in Remark 2.12 :we transfer the lower bound into a Bayesian estimation problem by constructing $\widetilde{\mathbb{P}}$ adequately. We then use the asymptotic equivalence result of Reiß in order to approximate the conditional law of the data given σ under $\widetilde{\mathbb{P}}$ by a classical Gaussian shift experiment, thanks to a Markov kernel. In the special case $p = \pi = 2$, we could also derive the result by using the lower bound in [34].

3 Wavelet estimation and pre-averaging

3.1 Estimating linear functionals

We estimate the (square of the) volatility process, σ^2, via linear functionals of the form

$$\langle \sigma^2, h_{\ell k}\rangle_{L^2} := \int_0^1 2^{\ell/2} h(2^\ell t - k)\, d\langle X\rangle_t,$$

where $\langle \bullet, \bullet \rangle_{L^2}$ denotes the inner product of $L^2([0,1])$ and $t \rightsquigarrow \langle X \rangle_t$ is the predictable compensator of the continuous semimartingale X.

Here the integers $\ell \in [0, n]$ and $k \in [0, 2^\ell]$ are respectively a resolution level and a location parameter. The test function $h : \mathbb{R} \to \mathbb{R}$ is smooth and throughout the paper we will assume that h is compactly supported on $[0, 1]$. Thus, $h_{\ell k} = 2^{\ell/2} h(2^\ell \bullet - k)$ is essentially located around $(k + \frac{1}{2})/2^\ell$.

Definition 3.1. *We say that $\lambda : [0, 2) \to \mathbb{R}$ is a pre-averaging function if it is piecewise Lipschitz continuous and satisfies $\lambda(t) = -\lambda(2 - t)$. To each pre-averaging function λ we associate the quantity*

$$\overline{\lambda} := \Big(2 \int_0^1 \big(\int_0^s \lambda(u) du\big)^2 ds\Big)^{1/2}$$

and define the (normalized) pre-averaging function $\widetilde{\lambda} := \lambda / \overline{\lambda}$.

For $1 \leq m < n$ and a sequence $(Y_{j,n}, j = 1, \dots, n)$, we define the pre-averaging of Y at scale m relative to λ by setting for $i = 1, \dots, m$

$$\overline{Y}_{i,m}(\lambda) := \frac{m}{n} \sum_{\frac{j}{n} \in \left(\frac{i-2}{m}, \frac{i}{m}\right]} \widetilde{\lambda}\big(m\tfrac{j}{n} - (i-2)\big) Y_{j,n}, \tag{3.1}$$

the summation being taken w.r.t. the index j. If $Y_{j,m}$ has the form $Y_{j/m}$ for some underlying continuous time process $t \rightsquigarrow Y_t$, the pre-averaging of Y at scale m is a kind of local average that mimics the behaviour of $Y_{i/m} - Y_{(i-2)/m}$. Indeed, using $\lambda(t) = -\lambda(2 - t)$,

$$\overline{Y}_{i,m}(\lambda) \approx -\frac{m}{n} \sum_{\frac{j}{n} \in \left(0, \frac{1}{m}\right]} \widetilde{\lambda}\big(m\tfrac{j}{n}\big)\big(Y_{i/m - j/n} - Y_{(i-2)/m + j/n}\big).$$

Thus, $\overline{Y}_{i,m}(\lambda)$ might be interpreted as a sum of differences in the interval $[(i-2)/m, i/m]$, weighted by $\widetilde{\lambda}$.

From (1.5), a first guess for estimating $\langle \sigma^2, h_{\ell k} \rangle_{L^2}$ is to consider the quantity

$$\sum_{i=2}^{m} h_{\ell k}\big(\tfrac{i-1}{m}\big) \overline{Z}_{i,m}^2$$

for some intermediate scale m that needs to be tuned with n and that reduces the effect of the noise $(\varepsilon_{j,n})$ in the representation (1.1). However, such a procedure is biased and a further correction is needed. To that end, we introduce

$$\mathfrak{b}(\lambda, Z_\bullet)_{i,m} := \frac{m^2}{2n^2} \sum_{\frac{j}{n} \in \left(\frac{i-2}{m}, \frac{i}{m}\right]} \widetilde{\lambda}^2\big(m\tfrac{j}{n} - (i-2)\big)\big(Z_{j,n} - Z_{j-1,n}\big)^2 \tag{3.2}$$

The form of $\mathfrak{b}(\lambda, Z_\bullet)_{i,m}$ given in (3.2) is not self-evident, and results from a series of stochastic approximations that are detailed in the proof Section 4. It appears as a natural choice in the transparent (yet technical) Section 4.1.

Finally, our estimator of $\langle \sigma^2, h_{\ell k}\rangle_{L^2}$ is

$$\mathcal{E}_m(h_{\ell k}) := \sum_{i=2}^{m} h_{\ell k}\big(\tfrac{i-1}{m}\big)\big[\,\overline{Z}_{i,m}^2 - \mathfrak{b}(\lambda, Z_\bullet)_{i,m}\big]. \tag{3.3}$$

3.2 The wavelet threshold estimator

We are now ready to construct an fully nonparametric estimator of the volatility process $(\sigma_t^2)_{t\in[0,1]}$. Let (φ, ψ) denote a pair of scaling function and mother wavelet that generates a multiresolution of $L^2([0,1])$, appended with boundary conditions, see *e.g.* [12, 13].

The volatility (random) function $t \rightsquigarrow \sigma_t^2$ taken path-by-path as an element of $L^2([0,1])$ has almost-sure representation

$$\sigma_\bullet^2 = \sum_{k\in\Lambda_{\ell_0}} c_{\ell_0 k}\, \varphi_{\ell_0 k}(\bullet) + \sum_{\ell>\ell_0}\sum_{k\in\Lambda_\ell} d_{\ell k}\, \psi_{\ell k}(\bullet), \tag{3.4}$$

with

$$c_{\ell_0 k} = \langle \sigma^2, \varphi_{\ell_0 k}\rangle_{L^2} = \int_0^1 \varphi_{\ell_0 k}(t) d\langle X\rangle_t,$$
$$d_{\ell k} = \langle \sigma^2, \psi_{\ell k}\rangle_{L^2} = \int_0^1 \psi_{\ell k}(t) d\langle X\rangle_t.$$

For every $\ell \geq 0$, the index set Λ_ℓ has cardinality 2^ℓ (and also incorporates boundary terms in the first part of the expansion that we choose not to distinguish in the notation from $\varphi_{\ell_0 k}$ for simplicity.)

Following the classical wavelet threshold algorithm (see for instance [16] and in its more condensed form [32]), we approximate Formula (3.4) by

$$\widehat{\sigma}_n^2(\bullet) := \sum_{k \in \Lambda_{\ell_0}} \mathcal{E}(\varphi_{\ell_0 k})\varphi_{\ell_0 k}(\bullet) + \sum_{\ell=\ell_0+1}^{\ell_1} \sum_{k \in \Lambda_\ell} \mathcal{T}_\tau\big[\mathcal{E}(\psi_{\ell k})\big]\psi_{\ell k}(\bullet) \tag{3.5}$$

where the wavelet coefficient estimates $\mathcal{E}(\varphi_{\ell_0 k})$ and $\mathcal{E}(\psi_{\ell k})$ are given by (3.3) and

$$\mathcal{T}_\tau[x] = x1_{\{|x| \geq \tau\}}, \;\; \tau \geq 0, \;\; x \in \mathbb{R}$$

is the standard hard-threshold operator.

Thus our estimator $t \rightsquigarrow \widehat{\sigma}_n^2(t)$ is specified by the resolution levels ℓ_0, ℓ_1, the threshold κ and the estimators $\mathcal{E}(\varphi_{\ell_0 k})$ and $\mathcal{E}(\psi_{\ell k})$ which in turn are entirely determined by the choice of the pre-averaging function λ and the pre-averaging resolution level m. (And of course, the choice of the basis generated by (φ, ψ) on $L^2([0,1])$.)

3.3 Convergence rates

We first give two results on the properties of $\mathcal{E}_m(h_{\ell k})$ for estimating $\langle \sigma^2, h_{\ell k}\rangle_{L^2}$.

Theorem 3.2 (Moment bounds)**.** *Work under Assumptions 2.1 and 2.2. Let us assume that h has piecewise Lipschitz derivative and that $2^\ell \leq m \leq n^{1/2}$.*

If $s > 1/\pi$, for any $c > 0$, for every $p \geq 1$, we have

$$\mathbb{E}\left[\left|\mathcal{E}_m(h_{\ell k}) - \langle \sigma^2, h_{\ell k}\rangle_{L^2}\right|^p \mathbb{I}_{\{\sigma^2 \in \mathcal{B}^s_{\pi,\infty}(c)\}}\right] \lesssim \|h\|_{L^\infty} m^{-p/2} + \\ + m^{-\min\{s-1/\pi,1\}p} |h_{\ell k}|^p_{1,m},$$

where

$$|h_{\ell k}|_{1,m} := m^{-1} \sum_{i=1}^m |h_{\ell k}(i/m)|.$$

The symbol $\lesssim$ means up to a constant that does not depend on m, n and h.

Theorem 3.3 (Deviation bounds)**.** *Work under Assumptions 2.1 and 2.2. Let us assume that h has piecewise Lipschitz derivative and that $2^l \leq m \leq n^{1/2}$. If moreover*

$$m2^{-\ell} \geq m^q, \;\; \textit{for some} \;\; q > 0,$$

then, if $s > 1/\pi$, for any $c > 0$, for every $p \geq 1$, we have

$$\mathbb{P}\left[\left|\mathcal{E}_m(h_{\ell k}) - \langle \sigma^2, h_{\ell k}\rangle_{L^2}\right| \geq \kappa\big(\tfrac{p\log m}{m}\big)^{1/2} \text{ and } \sigma^2 \in \mathcal{B}^s_{\pi,\infty}(c)\right] \lesssim m^{-p}$$

provided

$$\kappa > 4\overline{c} + 4\sqrt{2\ \overline{c}}\ \|a\|_{L^\infty}\|\lambda\|_{L^2}\overline{\lambda}^{\ -1} + 4\|a\|^2_{L^\infty}\|\lambda\|^2_{L^2}\overline{\lambda}^{\ -2}$$

and

$$m^{-(s-1/\pi)}|h_{\ell k}|_{1,m} \lesssim m^{-1/2},$$

where $\overline{c} := \sup_{\sigma^2\in\mathcal{B}^s_{\pi,\infty}(c)} \|\sigma^2\|_{L^\infty}$.

Thanks to Theorems 3.2 and 3.3, the performance of the wavelet estimator constructed in Section 3.2 readily follows, as stems from the classical nonlinear and adaptive estimation theory by wavelet thresholding, see [15], [16] and [32].

Let us be given a wavelet pair (φ, ψ) that generates a r-regular multiresolution analysis of $L^2([0, 1])$, for some $r \geq 1$.

Theorem 3.4. *Work under Assumptions 2.1 and 2.2. Let $\widehat{\sigma}^2_n$ denote the wavelet estimator defined in (3.5), constructed from (φ, ψ) and a pre-averaging function λ, such that*

$$m \sim n^{1/2}, 2^{\ell_0} \sim m^{1-2\alpha_0} \text{ for some } 0 < \alpha_0 < 1/2, 2^{\ell_1} \sim m^{1/2}$$

and $\tau := \kappa\sqrt{\frac{\log m}{m}}$ for sufficiently large $\kappa > 0$. Then, for

$$\alpha_0 \leq s - 1/\pi \leq \max\{\alpha_0/(1 - 2\alpha_0), r\},$$

the estimator $\widehat{\sigma}^2_n$ achieves (2.4) with $v_n := \big(n^{-1}(\log n)^{3/2}\big)^{s(\pi^\star)/(4s(\pi^\star)+2)}$, where $s(\pi^\star)$ is defined in (2.7) and (2.9).

Corollary 3.5. *We have Theorem 2.9.*

Remark 3.6. *By taking $\alpha_0 < 1/2$, Theorem 3.4 shows that in this case the estimator can at most adapt to the correct smoothness within the range $0 < \alpha_0 \leq s - 1/\pi \leq \alpha_0/(1 - 2\alpha_0) < \infty$.*

Remark 3.7. *In order to achieve adaptation, i.e. an estimator that does not depend on the pre-set smoothness parameter of the problem, the threshold κ needs to be taken large enough, and depends on λ. More precisely, it can be taken as $\widetilde{\kappa}\big(1 + \overline{\lambda}^{-2}\|\lambda\|^2_{L^2}\big)$, with $\widetilde{\kappa} = \widetilde{\kappa}(s, \pi, c, p, \|a\|_\infty)$,* *where $\|a\|_\infty$ is the level of the microstructure noise function a, defined in (2.2). It can be explicitly computed from the proof of Proposition 3.3 combined with the material indicated in the proof of Proposition 3.4. However, although an explicit bound derived from the computations is feasible, for practical purposes it is expected as too conservative in practice, as is well known in the signal detection case (e.g., Donoho and Johnstone [15]) or the classical inverse problem case (Abramovich and Silverman [1]). We will not persue this issue further in this paper and postpone a practical feasible thresholding to further work.*

4 Proofs

4.1 Proof of Theorem 3.2

We shall first introduce several auxiliary estimates which rely on classical techniques of discretization of random processes. Some are new.

In the sequel we shall repeatedly use the notation $\lesssim$ which means up to a constant that does not depend on n (or m which is later tuned with n). The other dependencies shall be obvious from the context.

Unless otherwise specified, L^2 abbreviates $L^2([0,1])$.

If $g : [0,1] \to \mathbb{R}$ is continuously differentiable, we define for $n \geq 1$

$$\mathfrak{R}_n(g) := \Big(\sum_{j=1}^{n} \int_{(j-1)/n}^{j/n} \big(\tfrac{1}{n} \sum_{l=j}^{n} g'(\tfrac{l}{n}) - \int_s^1 g'(u)du \big)^2 ds \Big)^{1/2},$$

and

$$|g|_{p,m} := \Big(\tfrac{1}{m} \sum_{i=1}^{m} |g(\tfrac{i-1}{m})|^p \Big)^{1/p}.$$

In the following, if $\mathcal{D}$ is a function class, we will sometimes write $\mathbb{E}_{\mathcal{D}}[\bullet]$ for $\mathbb{E}[\bullet\, \mathbb{I}_{\sigma^2 \in \mathcal{D}}]$. Clearly, if $\mathcal{D}_1 \subset \mathcal{D}_2$, we have $\mathbb{E}_{\mathcal{D}_1}[\bullet] \leq \mathbb{E}_{\mathcal{D}_2}[\bullet]$. For $c > 0$, let

$$\mathcal{D}_\infty(c) := \{ f : [0,1] \to \mathbb{R},\ \|f\|_{L^\infty} \leq c \}.$$

Preliminaries : some estimates for the latent price X

We start with a standard approximation result for discretized stochastic integrals.

Lemma 4.1 (Discretization effect)**.** *Let $g : [0, 1] \to \mathbb{R}$, be a deterministic function with piecewise continuous derivative, such that $g(1) = 0$.*

Work under Assumption 2.1. For every $p \geq 1$ and $c > 0$, we have

$$\mathbb{E}_{\mathcal{D}_\infty(c)}\Big[\Big|\big(\tfrac{1}{n}\sum_{i=1}^{n} g'(\tfrac{i}{n})X_{i/n}\big)^2 - \big(\int_0^1 g(s)dX_s\big)^2\Big|^p\Big] \lesssim \|g\|_{L^2}^p\,\mathfrak{R}_n^p(g) + \mathfrak{R}_n^{2p}(g).$$

Proof. First, by a change of measure and Assumption 2.1, we may assume that X is a local martingale. Next, by Cauchy-Schwarz, we split the error term into a constant times $I \times II + III \times II$, with

$$\begin{aligned}
I &:= \mathbb{E}_{\mathcal{D}_\infty(c)}\Big[\Big|\int_0^1 g'(s)X_s ds\Big|^{2p}\Big]^{1/2},\\
II &:= \mathbb{E}_{\mathcal{D}_\infty(c)}\Big[\Big|\tfrac{1}{n}\sum_{j=1}^{n} g'\big(\tfrac{j}{n}\big)X_{j/n} - \int_0^1 g'(s)X_s ds\Big|^{2p}\Big]^{1/2},\\
III &:= \mathbb{E}_{\mathcal{D}_\infty(c)}\Big[\Big|\tfrac{1}{n}\sum_{j=1}^{n} g'\big(\tfrac{j}{n}\big)X_{j/n}\Big|^{2p}\Big]^{1/2} \lesssim I + II.
\end{aligned}$$

Set

$$T_c := \inf\{s \geq 0,\ \ \sigma_s^2 > c\} \wedge 1.$$

On $\{\sigma^2 \in \mathcal{D}_\infty(c)\}$, we have $T_c = 1$, thus

$$\begin{aligned}
\mathbb{E}_{\mathcal{D}_\infty(c)}\Big[\Big|\int_0^1 g'(s)X_s ds\Big|^{2p}\Big] &= \mathbb{E}\Big[\Big|\int_0^{T_c} g'(s)X_s ds\Big|^{2p}\mathbb{I}_{\sigma^2\in\mathcal{D}_\infty(c)}\Big]\\
&\leq \mathbb{E}\Big[\Big|\int_0^{T_c} g'(s)X_s ds\Big|^{2p}\Big].
\end{aligned}$$

Integrating by part and using the Burkholder-Davis-Gundy inequality (later abbreviated by BDG, for a reference see [31], p. 166), we have

$$I \leq \mathbb{E}\Big[\Big|\int_0^{T_c} g(s)dX_s\Big|^{2p}\Big]^{1/2} \lesssim \mathbb{E}\Big[\Big|\int_0^{T_c} g^2(s)\sigma_s^2 ds\Big|^{p}\Big]^{1/2} \lesssim \|g\|_{L^2}^p,$$

where we used that $\sigma_s^2 \leq c$ for $s \leq T_c$. For the term II, note first that if

$$\tilde{g}(s) := \sum_{j=1}^{n}\big(\tfrac{1}{n}\sum_{l=j}^{n} g'\big(\tfrac{l}{n}\big)\big)\mathbb{I}_{[(j-1)/n, j/n)}(s), \ \ s \in [0, 1],$$

the process $S_t = \int_0^{t\wedge T_c} \big(\tilde{g}(s) + g(s)\big) dX_s$, $t \in [0,1]$ is a martingale and

$$\langle S\rangle_1 = \sum_{j=1}^{n} \int_{(j-1)/n}^{j/n} \Big(\tfrac{1}{n}\sum_{l=j}^{n} g'\big(\tfrac{l}{n}\big) - \int_s^1 g'(u)du\Big)^2 \mathbb{I}_{\{s\le T_c\}} d\langle X\rangle_s.$$

By summation by parts, we derive

$$II = \mathbb{E}_{\mathcal{D}_\infty(c)}\left[|S_1|^{2p}\right]^{1/2} \lesssim \mathbb{E}\left[\langle S\rangle_{T_c}^{p}\right]^{1/2} \lesssim \mathfrak{R}_n^p(g).$$

□

We further need some analytical properties of pre-averaging functions. In the following λ, and $\tilde{\lambda}$ always denote a pre-averaging function and its normalized version (in the sense of Definition 3.1). We set

$$\Lambda(s) := \int_s^2 \widetilde{\lambda}(u)du \; \mathbb{I}_{[0,2]}(s) \tag{4.1}$$

and

$$\overline{\Lambda}(s) := \Big(\big(\int_0^s \widetilde{\lambda}(u)du\big)^2 + \big(\int_0^{1-s} \widetilde{\lambda}(u)du\big)^2\Big)^{1/2} \mathbb{I}_{[0,1]}(s)\,. \tag{4.2}$$

Note that for $i = 2, \ldots, m$

$$\|\Lambda\big(m\bullet - (i-2)\big)\|_{L^2[0,1]} = m^{-1/2}\|\Lambda\|_{L^2[0,2]}$$

and

$$\|\overline{\Lambda}\big(m\bullet - (i-1)\big)\|_{L^2[0,1]} = m^{-1/2}.$$

Lemma 4.2. *For $m \le n$, we have*

$$\mathfrak{R}_n\big[\Lambda\big(m\bullet - (i-2)\big)\big] \lesssim n^{-1}$$

and for $i = 2, \ldots, m$

$$\|\Lambda\big(m\bullet - (i-2)\big)\|_{L^2} = m^{-1/2}.$$

Proof. Let $j_i := \min\{j,\ im^{-1} \le jn^{-1}\}$. We have

$$\max_{\frac{j}{n}\in\left(\frac{i-2}{m},\frac{i}{m}\right]} \sup_{s\in\left[\frac{j-1}{n},\frac{j}{n}\right]} \Big|\tfrac{1}{n}\sum_{l=j}^{j_i} \widetilde{\lambda}\big(m\tfrac{l}{n} - (i-2)\big) - \int_s^1 \widetilde{\lambda}\big(mu - (i-2)\big)du\Big|$$

$$\le \max_{\frac{j}{n}\in\left(\frac{i-2}{m},\frac{i}{m}\right]} \sup_{s\in\left[\frac{j-1}{n},\frac{j}{n}\right]} \Big|\int_s^{(j-1)/n} \widetilde{\lambda}\big(mu - (i-2)\big)du\Big| +$$

$$\max_{\frac{j}{n}\in\left(\frac{i-2}{m},\frac{i}{m}\right]} \sum_{l=j}^{j_i} \Big|\frac{1}{n}\widetilde{\lambda}\big(m\frac{l}{n} - (i-2)\big) - \int_{(l-1)/n}^{l/n} \widetilde{\lambda}\big(mu - (i-2)\big)du\Big| \lesssim n^{-1},$$

whence the first part of the lemma. For the second part, we have to prove that

$$\|\Lambda\|_{L^2[0,2]} = 1.$$

This readily follows from

$$\begin{aligned}
\|\Lambda\|^2_{L^2[0,2]} &= \int_0^1 \big(\int_s^2 \widetilde{\lambda}(u)\,du\big)^2 ds + \int_1^2 \big(\int_s^2 \widetilde{\lambda}(u)\,du\big)^2 ds \\
&= \int_0^1 \big(\int_0^s \widetilde{\lambda}(u)\,du\big)^2 ds + \int_0^1 \big(\int_{1+s}^2 \widetilde{\lambda}(u)\,du\big)^2 ds \\
&= \int_0^1 \big(\int_0^s \widetilde{\lambda}(u)\,du\big)^2 ds + \int_0^1 \big(\int_{1-s}^2 \widetilde{\lambda}(u)\,du\big)^2 ds = \|\overline{\Lambda}\|^2_{L^2[0,1]}.
\end{aligned}$$

□

Lemma 4.3. *Work under Assumption 2.1 and let Λ as in (4.1) with λ as in Definition 3.1. Then for $m \le n$, every $p \ge 1$ and $c > 0$, we have*

$$\begin{aligned}
&\mathbb{E}_{\mathcal{D}_\infty(c)}\Big[\Big|\sum_{i=2}^m g\big(\tfrac{i-1}{m}\big)\Big(\int_0^1 \Lambda\big(ms-(i-2)\big)dX_s\Big)^2 \\
&\quad - \int_0^1 \sum_{i=2}^m g\big(\tfrac{i-1}{m}\big)\Lambda^2(ms-(i-2))d\langle X\rangle_s\Big|^p\Big] \lesssim \|g\|^p_{L^\infty}|\mathrm{supp}(g)|^{p/2}m^{-p/2},
\end{aligned}$$

where $|\mathrm{supp}(g)|$ denotes the support length of g.

Proof. In the same way as for Lemma 4.1, we may (and will) assume that X is a local martingale. For $i = 2, \dots, m$ and $t \in [0,1]$, set

$$H_{t,i} := g\big(\tfrac{i-1}{m}\big)\Lambda\big(mt-(i-2)\big)\int_{(i-2)/m}^t \Lambda\big(ms-(i-2)\big)dX_s\,\mathbb{I}_{\left(\frac{i-2}{m},\frac{i}{m}\right]}(t). \tag{4.3}$$

By integration by parts, we have

$$\begin{aligned}
&\sum_{i=2}^m g\big(\tfrac{i-1}{m}\big)\Big[\Big(\int_0^1 \Lambda\big(ms-(i-2)\big)dX_s\Big)^2 \\
&\qquad - \int_0^1 \Lambda^2\big(ms-(i-2)\big)d\langle X\rangle_s\Big] \\
&= 2\sum_{i=2}^m \int_{(i-2)/m}^{i/m} H_{t,i}\,dX_t.
\end{aligned} \tag{4.4}$$

For $t \in [0,1]$, the process $\sum_{i=1}^m H_{t,i}$ is continuous (because of $\Lambda(0) = \Lambda(2) = 0$) and adapted, hence $\int_0^t \sum_{i=2}^m H_{s,i}\, dX_s$ is a continuous local martingale. Applying BDG and the localization argument of Lemma 4.1, we obtain

$$\begin{aligned}
&\mathbb{E}_{\mathcal{D}_\infty(c)}\big[\big|\int_0^{T_c}\sum_{i=2}^m H_{t,i}\, dX_s\big|^p\big] \\
&\lesssim \mathbb{E}\big[\big|\int_0^{T_c}\big(\sum_{i=2}^m H_{t,i}\, dt\big)^2\big|^{p/2}\big] \lesssim \mathbb{E}\big[\big|\int_0^{T_c}\sum_{i=2}^m H_{t,i}^2\, dt\big|^{p/2}\big] \\
&\lesssim \mathbb{E}\big[\big|m^{-1}\sum_{i=2}^m (H_i^\star)^2\big|^{p/2}\big] \lesssim |\mathrm{supp}(g)|^{p/2-1} m^{-1}\sum_{i=2}^m \mathbb{E}\big[(H_i^\star)^p\big],
\end{aligned}$$

where $H_i^\star := \sup_{t \le T_c}|H_{t,i}|$ and where we used that $t \rightsquigarrow H_{t,i}$ has compact support with length of order m^{-1}. The last estimate followed by Hölder inequality. By BDG again, we derive

$$\begin{aligned}
\mathbb{E}\big[(H_i^\star)^p\big] &\lesssim \big|g\big(\tfrac{i-1}{m}\big)\big|^p\, \mathbb{E}\Big[\sup_{t\le 2/m}\Big|\int_{(i-2)/m\wedge T_c}^{((i-2)/m+t)\wedge T_c}\Lambda\big(ms-(i-2)\big)dX_s\Big|^p\Big] \\
&\lesssim \big|g\big(\tfrac{i-1}{m}\big)\big|^p\, \mathbb{E}\Big[\Big(\int_{(i-2)/m\wedge T_c}^{T_c}\Lambda^2\big(ms-(i-2)\big)\sigma_s^2 ds\Big)^{p/2}\Big] \\
&\lesssim \big|g\big(\tfrac{i-1}{m}\big)\big|^p m^{-p/2}. \qquad (4.5)
\end{aligned}$$

The result follows. $\square$

Lemma 4.4. *Work under Assumption 2.1. Let $\mathcal{B}^s_{\pi,\infty}(c)$ denote a Besov ball with $s > 1/\pi$ and $c > 0$.*

In the same setting as in Lemma 4.3, for every $p \ge 1$, we have

$$\mathbb{E}_{\mathcal{B}^s_{\pi,\infty}(c)}\Big[\Big|\sum_{i=2}^m g\big(\tfrac{i-1}{m}\big)\overline{X}_{i,m}^2 - \int_0^1 g(s)\sigma_s^2 ds\Big|^p\Big] \lesssim \;\|g\|_{L^\infty}^p m^{-p/2}|\mathrm{supp}(g)|^{p/2}$$
$$+\, |g|_{1,m}^p m^{-\min\{s-1/\pi,1\}p} + |g|_{\mathrm{var},m}^p m^{-p},$$

where

$$|g|_{\mathrm{var},m} := |g(0)+g(1)| + \sum_{i=1}^m \sup_{s,t\in[(i-1)/m,i/m]}|g(t)-g(s)|.$$

Proof. Recall from Section 3.1 that

$$\overline{X}_{i,m}(\lambda) := \frac{m}{n}\sum_{\frac{j}{n}\in\left(\frac{i-2}{m},\frac{i}{m}\right]}\widetilde{\lambda}\big(m\tfrac{j}{n}-(i-2)\big)X_{j/n}.$$

Since $s > 1/\pi$, the class $\mathcal{B}^s_{\pi,\infty}(c) \subset \mathcal{D}_\infty(c')$ for some $c' = c'(s,\pi,c)$. Therefore, by Lemma 4.1, we have

$$\mathbb{E}_{\mathcal{B}^s_{\pi,\infty}(c)}\Big[\Big|\overline{X}^2_{i,m} - \Big(\int_0^1 \Lambda\big(ms-(i-2)\big)dX_s\Big)^2\Big|^p\Big] \lesssim m^{-p/2}n^{-p} \tag{4.6}$$

since

$$\mathfrak{R}_n\big[\Lambda\big(m\bullet-(i-2)\big)\big] \lesssim n^{-1}$$

by Lemma 4.2, $\|\Lambda\big(m\bullet-(i-2)\big)\|_{L^2} = m^{-1/2}$ and $m \le n$. By Hölder inequality it follows

$$\begin{aligned}&\mathbb{E}_{\mathcal{B}^s_{\pi,\infty}(c)}\Big[\Big|\sum_{i=2}^m g\big(\tfrac{i-1}{m}\big)\overline{X}^2_{i,m} - \sum_{i=2}^m g\big(\tfrac{i-1}{m}\big)\Big(\int_0^1 \Lambda\big(ms-(i-2)\big)dX_s\Big)^2\Big|^p\Big]\\ &\lesssim |\operatorname{supp}(g)|^{p-1}m^{p-1}\\ &\qquad\times \mathbb{E}_{\mathcal{B}^s_{\pi,\infty}(c)}\Big[\sum_{i=2}^m \Big|g\big(\tfrac{i-1}{m}\big)\Big|^p\Big|\overline{X}^2_{i,m} - \Big(\int_0^1 \Lambda\big(ms-(i-2)\big)dX_s\Big)^2\Big|^p\Big]\\ &\qquad\qquad\lesssim \|g\|^p_{L^\infty} m^{p/2}n^{-p}|\operatorname{supp}(g)|^p,\end{aligned} \tag{4.7}$$

which can be further bounded by $\|g\|^p_{L^\infty}m^{-p/2}|\operatorname{supp}(g)|^{p/2}$. By Lemma 4.3, we have

$$\begin{aligned}&\mathbb{E}_{\mathcal{B}^s_{\pi,\infty}(c)}\Big[\Big|\sum_{i=2}^m g\big(\tfrac{i-1}{m}\big)\Big(\int_0^1 \Lambda\big(ms-(i-2)\big)dX_s\Big)^2\\ &-\int_0^1 \sum_{i=2}^m g\big(\tfrac{i-1}{m}\big)\Lambda^2\big(ms-(i-2)\big)\sigma_s^2 ds\Big|^p\Big] \lesssim \|g\|^p_{L^\infty}m^{-p/2}|\operatorname{supp}(g)|^{p/2},\end{aligned}$$

therefore by triangle inequality also

$$\begin{aligned}&\mathbb{E}_{\mathcal{B}^s_{\pi,\infty}(c)}\Big[\Big|\sum_{i=2}^m g\big(\tfrac{i-1}{m}\big)\overline{X}^2_{i,m} - \int_0^1\sum_{i=2}^m g\big(\tfrac{i-1}{m}\big)\Lambda^2\big(ms-(i-2)\big)\sigma_s^2 ds\Big|^p\Big]\\ &\qquad\lesssim \|g\|^p_{L^\infty}m^{-p/2}|\operatorname{supp}(g)|^{p/2}.\end{aligned} \tag{4.8}$$

We are going to force the function $\overline{\Lambda}$ in (4.8). To this end, note that

$$
\begin{aligned}
&\sum_{i=2}^{m} g\big(\tfrac{i-1}{m}\big)\Lambda^2\big(ms-(i-2)\big) \\
&=\sum_{i=1}^{m} g\big(\tfrac{i}{m}\big)\Big(\Lambda^2\big(ms-(i-2)\big)+\Lambda^2\big(ms-(i-1)\big)\Big)\mathbb{I}_{\left(\frac{i-1}{m},\frac{i}{m}\right]}(s) \\
&+\sum_{i=1}^{m}\Big(g\big(\tfrac{i-1}{m}\big)-g\big(\tfrac{i}{m}\big)\Big)\Lambda^2\big(ms-(i-2)\big)\mathbb{I}_{\left(\frac{i-1}{m},\frac{i}{m}\right]}(s) \\
&-g(0)\Lambda^2\big(ms+1\big)\mathbb{I}_{\left(0,\frac{1}{m}\right]}(s)-g(1)\Lambda^2\big(ms-(m-1)\big)\mathbb{I}_{\left(1-\frac{1}{m},1\right]}(s). \qquad (4.9)
\end{aligned}
$$

Moreover, because of $\widetilde{\lambda}(u)=-\widetilde{\lambda}(2-u)$, we have $\Lambda^2(u)=\Lambda^2(2-u)$ and also $\Lambda(0)=0$,

$$\Lambda^2\big(ms-(i-2)\big)=\Big(\int_0^{1-(ms-(i-1))}\widetilde{\lambda}(u)du\Big)^2, \quad \text{for } s\in\big(\tfrac{i-1}{m},\tfrac{i}{m}\big],$$

$$\Lambda^2\big(ms-(i-1)\big)=\Big(\int_0^{ms-(i-1)}\widetilde{\lambda}(u)du\Big)^2, \quad \text{for } s\in\big(\tfrac{i-1}{m},\tfrac{i}{m}\big].$$

This gives for $s\in\big(\frac{i-1}{m},\frac{i}{m}\big]$, and $\bar{\Lambda}$ as in (4.2)

$$\overline{\Lambda}^2\big(ms-(i-1)\big)=\Lambda^2\big(ms-(i-2)\big)+\Lambda^2\big(ms-(i-1)\big), \qquad (4.10)$$

and 0 otherwise. From (4.9) it follows that on the event $\sigma^2\in\mathcal{B}^s_{\pi,\infty}(c)$

$$
\begin{aligned}
\Big|&\int_0^1\sum_{i=2}^{m} g\big(\tfrac{i-1}{m}\big)\Lambda^2\big(ms-(i-2)\big)\sigma_s^2ds \\
&-\int_0^1\sum_{i=1}^{m} g\big(\tfrac{i}{m}\big)\overline{\Lambda}^2\big(ms-(i-1)\big)\sigma_s^2ds\Big|\lesssim |g|_{\mathrm{var},m}m^{-1}. \qquad (4.11)
\end{aligned}
$$

Finally, we have for $\sigma^2\in\mathcal{B}^s_{\pi,\infty}(c)$ using $\|\overline{\Lambda}\|_{L^2}=1$

$$
\begin{aligned}
&\Big|\int_0^1\sum_{i=2}^{m} g\big(\tfrac{i-1}{m}\big)\Big(\overline{\Lambda}^2\big(ms-(i-1)\big)-\mathbb{I}_{\left(\frac{i-1}{m},\frac{i}{m}\right]}(s)\Big)\sigma_s^2ds\Big| \\
&\le\Big|\int_0^1\sum_{i=2}^{m} g\big(\tfrac{i-1}{m}\big)\overline{\Lambda}^2\big(ms-(i-1)\big)\big(\sigma_s^2-\sigma^2_{(i-1)/m}\big)ds\Big| \\
&+\Big|\int_0^1\sum_{i=2}^{m} g\big(\tfrac{i-1}{m}\big)\mathbb{I}_{\left(\frac{i-1}{m},\frac{i}{m}\right]}(s)\big(\sigma_s^2-\sigma^2_{(i-1)/m}\big)ds\Big| \\
&\lesssim m^{-\min\{s-1/\pi,1\}}|g|_{1,m}, \qquad (4.12)
\end{aligned}
$$

the last estimate coming from the Sobolev embedding $\mathcal{B}^s_{\pi,\infty} \subset \mathcal{B}^{s-1/\pi}_{\infty,\infty}$ which contains Hölder continuous functions of smoothness $\min\{s - 1/\pi, 1\}$. Since for $\sigma^2 \in \mathcal{B}^s_{\pi,\infty}(c)$

$$\Big|\int_0^1 \sum_{i=2}^m g\big(\tfrac{i}{m}\big)\mathbb{I}_{\big(\frac{i-1}{m},\frac{i}{m}\big]}(s)\sigma_s^2 ds - \int_0^1 g(s)\sigma_s^2 ds\Big| \lesssim m^{-1}|g|_{\mathrm{var},m}, \tag{4.13}$$

the conclusion follows by combining (4.8), (4.11), (4.12) and (4.13). □

Preliminaries: some estimates for the microstructure noise ϵ

We need some notation. Remember from (1.1) that we observe

$$Z_{j,n} = X_{j/n} + a(j/n, X_{j/n})\eta_{j,n}, \;\; j = 1, \ldots, n$$

where the intensity of microstructure noise process $a_s := a(s, X_s)$ and noise innovations $\eta_{j,n}$ satisfy Assumption 2.2. For a pre-averaging function λ, recall from (3.1) that we define

$$\overline{\epsilon}_{i,m} := \overline{\epsilon}_{i,m}(\lambda) := \frac{m}{n} \sum_{\frac{j}{n}\in\big(\frac{i-2}{m},\frac{i}{m}\big]} \widetilde{\lambda}\big(m\tfrac{j}{n} - (i-2)\big)\epsilon_{j,n}, \quad i = 2, \ldots, m.$$

Lemma 4.5. *Work under Assumption 2.1 and 2.2. Let $\mathcal{G}$ denote the σ-field generated by $(X_s, s \in [0,1])$. For every function $g : [0,1] \to \mathbb{R}$ and $p \geq 1$, we have*

$$\mathbb{E}\Big[\Big|\sum_{i=1}^m g\big(\tfrac{i-1}{m}\big)\big(\overline{\epsilon}^2_{i,m}(\lambda) - \mathbb{E}\big[\overline{\epsilon}^2_{i,m}(\lambda)\,\big|\,\mathcal{G}\big]\big)\Big|^p\Big]$$
$$\lesssim |g|^p_{2,m} m^{-3p/2} n^{-p} + |g|^p_{p,m} m^{p+1} n^{-p}.$$

Proof. Let us introduce the filtrations

$$\mathcal{F}^{\mathrm{even}}_r := \sigma\big(\eta_{j,n},\; j/n \leq 2r/m\big) \otimes \sigma\big(X_s : s \leq 2r/m\big),$$
$$\mathcal{F}^{\mathrm{odd}}_r := \sigma\big(\eta_{j,n},\; j/n \leq (2r+1)/m\big) \otimes \sigma\big(X_s : s \leq (2r+1)/m\big).$$

Straightforward calculations show that the partial sums $S^{\mathrm{even}}_r := \sum_{i=1}^r U_{2i}$ and $S^{\mathrm{odd}}_r := \sum_{i=1}^r U_{2i+1}$ with

$$U_i := g\big(\tfrac{i-1}{m}\big)\Big(\overline{\epsilon}^2_{i,m} - \tfrac{m^2}{n^2} \sum_{\frac{j}{n}\in\big(\frac{i-2}{m},\frac{i}{m}\big]} \widetilde{\lambda}^2\big(m\tfrac{j}{n} - (i-2)\big)a^2_{j/n}\Big)$$

form martingale schemes ($i = 1, \dots, r \le \lfloor m/2 \rfloor$) with respect to $\mathcal{F}_r^{\text{even}}$ and $\mathcal{F}_r^{\text{odd}}$ respectively. Further, we have, for every $p \ge 1$

$$\begin{aligned}
&\mathbb{E}\Big[\Big|\bar{\epsilon}_{i,m}^2 - \tfrac{m^2}{n^2} \sum_{\frac{j}{n} \in \left(\frac{i-2}{m}, \frac{i}{m}\right]} \widetilde{\lambda}^2\big(m\tfrac{j}{n} - (i-2)\big) a_{j/n}^2\Big|^p\Big] \\
&\lesssim \mathbb{E}\left[|\bar{\epsilon}_{i,m}|^{2p}\right] + \|\widetilde{\lambda}\|_{L^\infty}^{2p} \|a\|_{L^\infty}^{2p} m^p n^{-p} \lesssim m^p n^{-p}.
\end{aligned}$$

It follows that

$$\mathbb{E}\left[|U_i|^p\right] \lesssim |g(\tfrac{i-1}{m})|^p m^p n^{-p}. \tag{4.14}$$

Analogous computations show that

$$\mathbb{E}\left[U_{2i}^2 \,|\, \mathcal{F}_{i-1}^{\text{even}}\right] \le g^2\big(\tfrac{2i-1}{m}\big)\, \mathbb{E}\left[\bar{\epsilon}_{2i,m}^4 \,|\, \mathcal{F}_{i-1}^{\text{even}}\right] \lesssim g^2\big(\tfrac{2i-1}{m}\big) m^2 n^{-2}.$$

Therefore, applying Rosenthal's inequality for martingales (see [25], p. 23), we obtain

$$\mathbb{E}\left[|S_{\lfloor m/2 \rfloor}^{\text{even}}|^p\right] \lesssim |g|_{2,m}^p m^{3p/2} n^{-p} + |g|_{p,m}^p m^{p+1} n^{-p}.$$

Likewise, we obtain the same estimate for $\mathbb{E}\left[|S_{\lfloor (m-1)/2 \rfloor}^{\text{odd}}|^p\right]$. The conclusion follows. $\square$

Lemma 4.6. *In the same setting as in Lemma 4.5, we have, for every $c > 0$ and $p \ge 1$*

$$\begin{aligned}
&\mathbb{E}_{\mathcal{D}_\infty(c)}\Big[\Big|\sum_{i=1}^m g\big(\tfrac{i-1}{m}\big) \overline{X}_{i,m}(\lambda)\, \bar{\epsilon}_{i,m}(\lambda)\Big|^p\Big] \\
&\lesssim\ |g|_{p,m}^p \big(n^{-p/2} m + m^{3p/2+1} n^{-3p/2}\big) + |g|_{2,m}^p \big(m^{p/2} n^{-p/2} + m^{2p} n^{-3p/2}\big).
\end{aligned}$$

Proof. By Assumption 2.1 and the same localization procedure as in the proof of Lemma 4.1, up to loosing some constant, we may (and will) assume that X is a local martingale such that $|\sigma_s| \le c$ almost-surely and subsequently work with $\mathbb{E}[\bullet]$ instead of $\mathbb{E}_{\mathcal{D}_\infty(c)}[\bullet]$.

In the same way as for the proof of Lemma 4.5, we define an $\mathcal{F}^{\text{even}}$-martingale by setting

$$S_r^{\text{even}} := \sum_{i=1}^r g\big(\tfrac{2i-1}{m}\big) \overline{X}_{2i,m}(\lambda) \bar{\epsilon}_{2i,m}(\lambda)$$

and proceed for S^{odd} analogously. By Rosenthal's inequality for martingales, we have

$$\begin{aligned}\mathbb{E}\left[\left|S_{\lfloor m/2\rfloor}\right|^p\right] \lesssim & m^{p/2}n^{-p/2}\,\mathbb{E}\left[\Big|\sum_{i=1}^{\lfloor m/2\rfloor} g^2\big(\tfrac{2i-1}{m}\big)\,\mathbb{E}\left[\,\overline{X}^2_{2i,m}(\lambda)\,|\,\mathcal{F}^{\text{even}}_{i-1}\right]\Big|^{p/2}\right] \\ & + \sum_{i=1}^{\lfloor m/2\rfloor} \left|g\big(\tfrac{2i-1}{m}\big)\right|^p \big(\mathbb{E}\left[|\overline{X}_{2i,m}(\lambda)|^{2p}\right]\big)^{1/2}\big(\mathbb{E}\left[|\overline{\epsilon}_{2i,m}(\lambda)|^{2p}\right]\big)^{1/2},\end{aligned}$$

where we used that $\|a\|_{L^\infty} \lesssim 1$ and Cauchy-Schwarz. On the one hand, we have

$$\begin{aligned}\mathbb{E}\left[\left|\overline{X}_{i,m}(\lambda)\right|^{2p}\right] \lesssim \ & \mathbb{E}\left[\Big|\tfrac{m}{n}\sum_{\frac{j}{n}\in\left(\frac{i-2}{m},\frac{i}{m}\right]} \widetilde{\lambda}\big(m\tfrac{j}{n}-(i-2)\big)(X_{j/n}-X_{(i-2)/m})\Big|^{2p}\right] \\ & + m^{2p}n^{-2p}\,\mathbb{E}\left[|X_{(i-2)/m}|^{2p}\right],\end{aligned}$$

where we used the fact that, by Riemann's approximation, we have

$$\Big|\sum_{\frac{j}{n}\in\left(\frac{i-2}{m},\frac{i}{m}\right]} \widetilde{\lambda}\big(m\tfrac{j}{n}-(i-2)\big)\Big| \lesssim 1. \tag{4.15}$$

It follows that $\mathbb{E}\left[\left|\overline{X}_{i,m}(\lambda)\right|^{2p}\right]$ is less than

$$\|\widetilde{\lambda}\|^{2p}_{L^\infty}\,\mathbb{E}\left[\sup_{s\le 2/m}|X_{(i-2)/m+s}-X_{(i-2)/m}|^{2p}\right] + m^{2p}n^{-2p}\,\mathbb{E}\left[|X_{(i-2)/m}|^{2p}\right]$$

which in turn is of order $\|\widetilde{\lambda}\|^{2p}_{L^\infty}m^{-p}+m^{2p}n^{-2p}$ thanks to the localization argument for σ. In a similar way, we have

$$\mathbb{E}\left[\overline{X}^2_{2i,m}(\lambda)\,\big|\,\mathcal{F}^{\text{even}}_{i-1}\right] \lesssim m^{-1}+m^2n^{-2}X^2_{(2i-2)/m} \le m^{-1}+m^2n^{-2}\sup_s X_s^2.$$

Recall that $\mathbb{E}\left[|\overline{\epsilon}_{i,m}|^{2p}\right] \lesssim m^pn^{-p}$. Putting together these estimates, we infer that $\mathbb{E}\left[\left|S^{\text{even}}_{\lfloor m/2\rfloor}\right|^p\right]$ satisfies the desired bound. We proceed likewise for $S_{\lfloor (m-1)/2\rfloor}$. The conclusion follows. □

Preliminaries: some estimates for the bias correction $\mathfrak{b}$

We need some notation. Recall the bias correction defined in (3.2)

$$\mathfrak{b}(\lambda, Z_\bullet)_{i,m} := \frac{m^2}{2n^2}\sum_{\frac{j}{n}\in\left(\frac{i-2}{m},\frac{i}{m}\right]} \widetilde{\lambda}^2\big(m\tfrac{j}{n}-(i-2)\big)\big(Z_{j,n}-Z_{j-1,n}\big)^2.$$

We plan to use the following decomposition

$$\mathfrak{b}(\lambda, Z_\bullet)_{i,m} = \mathfrak{b}(\lambda, X_\bullet)_{i,m} + \mathfrak{b}(\lambda, \varepsilon_\bullet)_{i,m} + 2\mathfrak{c}(\lambda, X_\bullet, \epsilon_\bullet)_{i,m},$$

where

$$\begin{aligned} &\mathfrak{c}(\lambda, X_\bullet, \epsilon_\bullet)_{i,m} \\ &:= \frac{m^2}{2n^2} \sum_{\frac{j}{n} \in \left(\frac{i-2}{m}, \frac{i}{m}\right]} \widetilde{\lambda}^2\big(m\tfrac{j}{n} - (i-2)\big)\big(X_{j/n} - X_{(j-1)/n}\big)\big(\epsilon_{j,n} - \epsilon_{j-1,n}\big). \end{aligned}$$

Lemma 4.7. *Work under Assumption 2.1 and 2.2. For every $p \geq 1$, we have*

$$\begin{aligned} &\mathbb{E}\Big[\Big|\sum_{i=2}^m g\big(\tfrac{i-1}{m}\big)\big(\mathfrak{b}(\lambda, \epsilon_\bullet)_{i,m} - \tfrac{m^2}{n^2} \sum_{\frac{j}{n} \in \left(\frac{i-2}{m}, \frac{i}{m}\right]} \widetilde{\lambda}^2\big(m\tfrac{j}{n} - (i-2)\big)a_{j/n}^2\big)\Big|^p\Big] \\ &\lesssim |g|_{1,m}^p m^{3p} n^{-2p} + |g|_{2,m}^p m^{2p} n^{-3p/2} + |g|_{p,m}^p m^{2p} n^{-2p+1}. \end{aligned}$$

Proof. By triangle inequality, we bound the error by a constant times

$$m^{2p} n^{-2p}(I + II + III + IV),$$

where

$$\begin{aligned} I &:= \mathbb{E}\Big[\Big|\sum_{i=2}^m g\big(\tfrac{i-1}{m}\big)\sum_j \widetilde{\lambda}^2\big(m\tfrac{j}{n} - (i-2)\big)a_{j/n}^2\big(\eta_{j,n}^2 - 1\big)\Big|^p\Big], \\ II &:= \mathbb{E}\Big[\Big|\sum_{i=2}^m g\big(\tfrac{i-1}{m}\big)\sum_j \widetilde{\lambda}^2\big(m\tfrac{j}{n} - (i-2)\big)a_{\frac{j-1}{n}}^2\big(\eta_{j-1,n}^2 - 1\big)\Big|^p\Big], \\ III &:= \mathbb{E}\Big[\Big|\sum_{i=2}^m g\big(\tfrac{i-1}{m}\big)\sum_j \widetilde{\lambda}^2\big(m\tfrac{j}{n} - (i-2)\big)\big(a_{\frac{j}{n}}^2 - a_{\frac{j-1}{n}}^2\big)\Big|^p\Big], \\ IV &:= \mathbb{E}\Big[\Big|\sum_{i=2}^m g\big(\tfrac{i-1}{m}\big)\sum_j \widetilde{\lambda}^2\big(m\tfrac{j}{n} - (i-2)\big)\epsilon_{j-1,n}\epsilon_{j,n}\Big|^p\Big], \end{aligned}$$

where, as before, the sum in j expands over $\big\{j/n \in \big((i-2)/m, i/m\big]\big\}$.

• **The terms I and II.** We only bound I, the same subsequent arguments applying for the term involving $\eta_{j-1,n}$. Let $\mathcal{F}_j = \sigma(\eta_{k,n} : k \leq j) \otimes \sigma(X_s :$

$s \leq 1$). By Rosenthal's inequality for martingales,

$$
\begin{aligned}
I \lesssim & \sum_{j=1}^{n} \Big(\sum_{i=2}^{m} \big| g\big(\tfrac{i-1}{m}\big) \big|^p \mathbb{I}_{\left\{ \frac{j}{n} \in \left(\frac{i-2}{m}, \frac{i}{m} \right] \right\}} \Big) \, \mathbb{E}\left[\big| (\eta_{j,n}^2 - 1) \big|^p \right] \\
& + \Big| \sum_{j=1}^{n} \sum_{i=2}^{m} g^2\big(\tfrac{i-1}{m}\big) \mathbb{I}_{\left\{ \frac{j}{n} \in \left(\frac{i-2}{m}, \frac{i}{m} \right] \right\}} \, \mathbb{E}\left[(\eta_{j,n}^2 - 1)^2 \,\middle|\, \mathcal{F}_{j-1} \right] \Big|^{p/2}, \\
\lesssim & |g|_{p,m}^p n + |g|_{2,m}^p n^{p/2}.
\end{aligned}
$$

where we used the fact that the functions a and $\widetilde{\lambda}$ are bounded.

• **The term III.** Summing by parts, we have

$$
\begin{aligned}
& \sum_{\frac{j}{n} \in \left(\frac{i-2}{m}, \frac{i}{m} \right]} \widetilde{\lambda}^2\big(m\tfrac{j}{n} - (i-2)\big) \big(a_{j/n}^2 - a_{(j-1)/n}^2\big) \\
& = - \sum_{\frac{j}{n} \in \left(\frac{i-2}{m}, \frac{i}{m} \right]} a_{(j-1)/n}^2 \Big(\widetilde{\lambda}^2\big(m\tfrac{j}{n} - (i-2)\big) - \widetilde{\lambda}^2\big(m\tfrac{j-1}{n} - (i-2)\big) \Big) \\
& \quad + a_{(i-2)/m}^2 \widetilde{\lambda}^2(0) - a_{i/m}^2 \widetilde{\lambda}^2(2).
\end{aligned}
$$

Since a is bounded and $\widetilde{\lambda}$ has finite variation, we infer

$$
\Big| \sum_{i=2}^{m} g\big(\tfrac{i-1}{m}\big) \sum_{\frac{j}{n} \in \left(\frac{i-2}{m}, \frac{i}{m} \right]} \widetilde{\lambda}^2\big(m\tfrac{j}{n} - (i-2)\big) \big(a_{j/n}^2 - a_{(j-1)/n}^2\big) \Big|^p \lesssim |g|_{1,m}^p m^p.
$$

• **The term IV.** We may split the sum with respect to j in even and odd part. Proceeding as for I and II, we readily obtain

$$
IV \lesssim |g|_{2,m}^p n^{p/2} + |g|_{p,m}^p n.
$$

□

Lemma 4.8. *In the same setting as in Lemma 4.7, for every $c > 0$, we have*

$$
\mathbb{E}_{\mathcal{D}_\infty(c)} \Big[\Big| \sum_{i=2}^{m} g\big(\tfrac{i-1}{m}\big) \mathfrak{b}(\lambda, X_\bullet)_{i,m} \Big|^p \Big] \lesssim |g|_{1,m}^p m^p n^{-p}.
$$

Proof. In the same way as in the proof of Lemma 4.6, we may (and will) assume that X is a local martingale and that $|\sigma_s^2| \leq c$ almost surely, working subsequently with $\mathbb{E}[\bullet]$ instead of $\mathbb{E}_{\mathcal{D}_\infty}(c)[\bullet]$. We readily obtain

$$
\begin{aligned}
&\mathbb{E}\Big[\Big|\sum_{i=2}^{m} g\big(\tfrac{i-1}{m}\big)\mathfrak{b}(\lambda, X_\bullet)_{i,m}\Big|^p\Big] \\
&\lesssim\ m^{2p}n^{-2p}\,\mathbb{E}\Big[\Big|\sum_{i=2}^{m}\big|g\big(\tfrac{i-1}{m}\big)\big|\sum_{\frac{j}{n}\in\left(\frac{i-2}{m},\frac{i}{m}\right]}(X_{j/n}-X_{(i-2)/m})^2\Big|^p\Big] \\
&\lesssim\ |g|_{1,m}^p m^p n^{-p}.
\end{aligned}
$$

□

Let M be a continuous, locally square integrable $\mathcal{F}$-martingale and H some progressive measurable process with finite quadratic variation. Then for $0 \le s < t < \infty$

$$
\mathbb{E}\Big[\Big(\int_s^t H_u dM_u\Big)^2|\mathcal{F}_s\Big] = \mathbb{E}\Big[\int_s^t H_u^2 d\langle M\rangle_u|\mathcal{F}_s\Big].
$$

This fact will be referred to in the sequel as conditional Ito-isometry (cf. [31], Section 3.2 B).

Lemma 4.9. *In the same setting as in Lemma 4.7, for every $c > 0$, we have*

$$
\begin{aligned}
&\mathbb{E}_{\mathcal{D}_\infty(c)}\Big[\Big|\sum_{i=2}^{m} g\big(\tfrac{i-1}{m}\big)\mathfrak{c}(\lambda, X_\bullet, \epsilon_\bullet)_{i,m}\Big|^p\Big] \\
&\lesssim\ \big[|g|_{2,m}^p + |g|_{p,m}^p(n^{-p/2+1}+m^{-p/2+1})\big]m^{2p}n^{-2p}.
\end{aligned}
$$

Proof. As in Lemmas 4.6 and 4.8, we may (and will) assume that X is a local martingale and that $|\sigma_s^2| \le c$ almost surely, working subsequently with $\mathbb{E}[\bullet]$ instead of $\mathbb{E}_{\mathcal{D}_\infty}(c)[\bullet]$. It suffices then to bound

$$
\mathbb{E}\Big[\Big|\sum_{i=1}^{m} g\big(\tfrac{i-1}{m}\big)\sum_{\frac{j}{n}\in\left(\frac{i-2}{m},\frac{i}{m}\right]}\widetilde{\lambda}^2\big(m\tfrac{j}{n}-(i-2)\big)(X_{j/n}-X_{(j-1)/n})\epsilon_{j,n}\Big|^p\Big].
$$

We define $j_n^*(r) := \max\{j : j/n \le r/m\}$. Let us introduce the filtrations

$$
\begin{aligned}
\mathcal{G}_r^{\text{even}} &:= \sigma\big(\eta_{j,n},\ j/n \le 2r/m\big)\otimes\sigma\big(X_s : s \le j_n^*(2r)/n\big), \\
\mathcal{G}_r^{\text{odd}} &:= \sigma\big(\eta_{j,n},\ j/n \le (2r+1)/m\big)\otimes\sigma\big(X_s : s \le j_n^*(2r+1)/n\big).
\end{aligned}
$$

The process

$$S_r^{\text{even}} := \sum_{i=1}^{r} g\big(\tfrac{2i-1}{m}\big) \sum_{\frac{j}{n}\in\left(\frac{2i-2}{m},\frac{2i}{m}\right]} \widetilde{\lambda}^2\big(m\tfrac{j}{n}-(2i-2)\big)(X_{j/n}-X_{(j-1)/n})\epsilon_{j,n}$$

is a $\mathcal{G}^{\text{even}}$-martingale and likewise for S_r^{odd} defined similarly w.r.t. the filtration $\mathcal{G}_r^{\text{odd}}$. Moreover, on one hand

$$\begin{aligned}
&\mathbb{E}\Big[\Big|g\big(\tfrac{i-1}{m}\big) \sum_{\frac{j}{n}\in\left(\frac{i-2}{m},\frac{i}{m}\right]} \widetilde{\lambda}^2\big(m\tfrac{j}{n}-(i-2)\big)(X_{j/n}-X_{(j-1)/n})\epsilon_{j,n}\Big|^p\Big] \\
&\lesssim \big|g\big(\tfrac{i-1}{m}\big)\big|^p\Big(m^{-p/2}+\sum_{\frac{j}{n}\in\left(\frac{i-2}{m},\frac{i}{m}\right]} \mathbb{E}\big[\big|(X_{j/n}-X_{(j-1)/n})\epsilon_{j,n}\big|^p\big]\Big) \\
&\lesssim \big|g\big(\tfrac{i-1}{m}\big)\big|^p m^{-1}(m^{-p/2+1}+n^{-p/2+1}),
\end{aligned}$$

and on the other hand by conditional Ito-isometry

$$\begin{aligned}
&\mathbb{E}\Big[\Big(g\big(\tfrac{2i-1}{m}\big) \sum_{\frac{j}{n}\in\left(\frac{2i-2}{m},\frac{2i}{m}\right]} \widetilde{\lambda}^2\big(m\tfrac{j}{n}-(2i-2)\big)(X_{j/n}-X_{(j-1)/n})\epsilon_{j,n}\Big)^2\Big|\,\mathcal{G}_{i-1}^{\text{even}}\Big] \\
&\lesssim g^2\big(\tfrac{2i-1}{m}\big) \sum_{\frac{j}{n}\in\left(\frac{2i-2}{m},\frac{2i}{m}\right]} \mathbb{E}\big[(X_{j/n}-X_{(j-1)/n})^2\,\big|\mathcal{G}_{i-1}^{\text{even}}\big] \lesssim m^{-1}g^2\big(\tfrac{2i-1}{m}\big).
\end{aligned}$$

Therefore, by Rosenthal's inequality for martingales, we infer

$$\mathbb{E}\big[\big|S_{\lfloor m/2\rfloor}^{\text{even}}\big|^p\big] \lesssim |g|_{p,m}^p(n^{-p/2+1}+m^{-p/2+1})+|g|_{2,m}^p.$$

We proceed likewise for $S_{\lfloor (m-1)/2\rfloor}^{\text{odd}}$ and the conclusion follows by incorporating the multiplicative term $m^{2p}n^{-2p}$ in front of the two error terms. $\square$

Completion of proof of Theorem 3.2

Since

$$\mathcal{E}_m(h_{\ell k}) = \sum_{i=2}^{m} h_{\ell k}\big(\tfrac{i-1}{m}\big)\big[\,\overline{Z}_{i,m}^2 - b(\lambda, Z_\bullet)_{i,m}\big]$$

we plan to use the following decomposition

$$\mathcal{E}_m(h_{\ell k}) - \langle \sigma^2, h_{\ell k}\rangle_{L^2} = I + II + III, \tag{4.16}$$

with

$$
\begin{aligned}
I &:= \sum_{i=2}^{m} h_{\ell k}\big(\tfrac{i-1}{m}\big)\ \overline{X}_{i,m}^2 - \langle \sigma^2, h_{\ell k}\rangle_{L^2},\\
II &:= \sum_{i=2}^{m} h_{\ell k}\big(\tfrac{i-1}{m}\big)\big[\overline{\epsilon}_{i,m}^2 - \mathfrak{b}(\lambda, Z_\bullet)_{i,m}\big],\\
III &:= 2\sum_{i=2}^{m} h_{\ell k}\big(\tfrac{i-1}{m}\big)\overline{X}_{i,m}\overline{\epsilon}_{i,m}.
\end{aligned}
$$

• **The term I.** By Lemma 4.4, we have

$$
\begin{aligned}
\mathbb{E}_{\mathcal{B}^s_{\pi,\infty}(c)}[|I|^p] \lesssim{}& \|h_{\ell k}\|_{L^\infty}^p m^{-p/2}|\mathrm{supp}(h_{\ell k})|^{p/2}\\
&+ |h_{\ell k}|_{1,m}^p m^{-\min\{s-1/\pi,1\}p} + |h_{\ell k}|_{\mathrm{var},m}^p m^{-p}.
\end{aligned}
$$

Using that $\|h_{\ell k}\|_{L^\infty} \lesssim 2^{\ell p/2}\|h\|_{L^\infty}$ and $|\mathrm{supp}(h_{\ell k})|^{p/2} \lesssim 2^{-\ell p/2}$, this term has the right order.

• **The term II.** Applying successively Lemmas 4.5, 4.7, 4.8 and 4.9, we derive

$$
\begin{aligned}
&\mathbb{E}_{\mathcal{B}^s_{\pi,\infty}}\big[|II|^p\big]\\
&\lesssim |h_{\ell k}|_{1,m}^p m^p n^{-p} + |h_{\ell k}|_{2,m}^p m^{-3p/2} n^{-p} + |h_{\ell k}|_{p,m}^p m^{p+1} n^{-p}
\end{aligned}
$$

and this term also has the right order.

• **The term III.** Finally, by Lemma 4.6, we have

$$
\begin{aligned}
&\mathbb{E}_{\mathcal{B}^s_{\pi,\infty}(c)}\Big[\Big|III\Big|^p\Big]\\
&\lesssim\ |h_{\ell k}|_{p,m}^p\big(n^{-p/2}m + m^{3p/2+1}n^{-3p/2}\big) + |h_{\ell k}|_{2,m}^p\big(m^{p/2}n^{-p/2} + m^{p/2}n^{-3p/2}\big),
\end{aligned}
$$

which also has the right order. The proof of Proposition 3.2 is complete.

4.2 Proof of Theorem 3.3

4.2.1 Preliminary: a martingale deviation inequality

If (M_k) is a locally square integrable $\mathcal{F}_k$-martingale with $M_0 = 0$, we denote by $[M]_k = \sum_{i=1}^k (\Delta M_i)^2$ with $\Delta M_i = M_i - M_{i-1}$ its quadratic variation and by $\langle M\rangle_k = \sum_{i=1}^k \mathbb{E}\left[(\Delta M_i)^2 \,|\, \mathcal{F}_{i-1}\right]$ its predictable compensator. We will heavily rely on the following result of Bercu and Touati [8]

Theorem 4.10 (Touati and Bercu [8]). *Let (M_k) be a locally square integrable martingale. Then, for all $x, y > 0$, we have*

$$\mathbb{P}\left[|M_n| \geq x,\ [M]_k + \langle M\rangle_k \leq y\right] \leq 2\exp\Big(-\frac{x^2}{2y}\Big).$$

From Theorem 4.10, we infer the following estimate

Lemma 4.11. *Let (M_j) be a locally square integrable $\mathcal{F}_j$-martingale. Suppose that for $p \geq 1$ there is some deterministic sequence $(C_j)_j$ $(j = j(m))$ and $\delta > 0$ such that $\mathbb{P}(\langle M\rangle_j > C_j(1+\delta)) \lesssim m^{-p}$. If further for every $\kappa \geq 2$*

$$\max_{i=1,\dots,j} \mathbb{E}\left[|\Delta M_i|^\kappa\right] \lesssim 1, \tag{4.17}$$

then,

$$\mathbb{P}\left[\left|M_j\right| > 2(1+\delta)\sqrt{C_j\ p\log m}\right] \lesssim m^{-p}$$

provided $m^{q_0} \leq j \leq m$ for some $0 < q_0 \leq 1$ and there is an $\epsilon > 0$ such that $C_j \gtrsim j^{1/2+\epsilon}$.

Proof. We have by Theorem 4.10

$$\begin{aligned}
&\mathbb{P}\left[\left|M_j\right| \geq 2(1+\delta)\sqrt{C_j p\log m}\right] \\
&\leq 2m^{-p} + \mathbb{P}\left[[M]_j + \langle M\rangle_j > y,\ \langle M\rangle_j \leq C_j(1+\delta)\right] + \mathbb{P}\left[\langle M\rangle_j > C_j(1+\delta)\right],
\end{aligned}$$

with $y = 2C_j(1+2\delta)$. Further we obtain

$$\mathbb{P}\left[[M]_j + \langle M\rangle_j > y,\ \langle M\rangle_j \leq C_j(1+\delta)\right] \leq \mathbb{P}\left[[M]_j - \langle M\rangle_j > 2C_j\delta\right].$$

Since $([M]_j - \langle M\rangle_j)$ is a $\mathcal{F}_j$-martingale it follows by Chebycheff's and Rosenthal's inequality for martingales and $\kappa \geq 2$

$$\begin{aligned}
\mathbb{P}\left[[M]_j - \langle M\rangle_j > 2C_j\delta\right] &\lesssim C_j^{-\kappa}\,\mathbb{E}\left[\left|[M]_j - \langle M\rangle_j\right|^\kappa\right] \\
&\lesssim C_j^{-\kappa}\sum_{i=1}^{j}\mathbb{E}\left|\Delta M_i\right|^{2\kappa} + C_j^{-\kappa}\,\mathbb{E}\Big|\sum_{i=1}^{j}\mathbb{E}\left((\Delta M)_i^4|\mathcal{F}_{i-1}\right)\Big|^{\kappa/2} \\
&\lesssim C_j^{-\kappa}(j + j^{\kappa/2}) \lesssim j^{-\epsilon\kappa},
\end{aligned}$$

where we used Hölder's inequality

$$\mathbb{E}\Big|\sum_{i=1}^{j}\mathbb{E}\left((\Delta M)_i^4|\mathcal{F}_{i-1}\right)\Big|^{\kappa/2} \lesssim j^{\kappa/2-1}\sum_{i=1}^{j}\mathbb{E}\Big(\mathbb{E}\left(|\Delta M_i|^{2\kappa}|\mathcal{F}_{i-1}\right)\Big) \lesssim j^{\kappa/2}.$$

Choosing $\kappa := q_0^{-1} p \epsilon^{-1} > 2$, we finally obtain

$$\mathbb{P}\left[[M]_j + \langle M \rangle_j > y,\ \langle M \rangle_j \leq C_j (1+\delta) \right] \lesssim j^{-p/q_0} \leq m^{-p}.$$

$\square$

Lemma 4.12. *Work under Assumptions of Proposition 3.3. Then we have for every fixed $\delta > 0$*

$$\mathbb{P}\left(\left| \sum_{i=2}^{m} h_{\ell k}\left(\tfrac{i-1}{m}\right) \overline{X}_{i,m}^2(\lambda) - \langle \sigma^2, h_{\ell k} \rangle_{L^2} \right| \right.$$
$$\left. > 4\overline{c}(1+\delta)\sqrt{\tfrac{p \log m}{m}} \text{ and } \sigma^2 \in \mathcal{B}_{\pi,\infty}^s(c) \right) \lesssim m^{-p},$$

where $\overline{c} = \overline{c}(s, \pi, c)$ is such that $\mathcal{B}_{\pi,\infty}^s(c) \subset \mathcal{D}_\infty(\overline{c})$ provided

$$m^{-(s-1/\pi)} |h_{\ell k}|_{1,m} \lesssim m^{-1/2}.$$

Proof. Recall that $\Lambda(s) = \int_s^2 \widetilde{\lambda}(u)\, du$ and let $H_{t,i}$ be defined as in (4.3), where g is replaced by $h_{\ell k}$. Using the integration by parts formula (4.4) we bound the probability by $I + II + III$, with

$$I := \mathbb{P}\Big(\Big| \sum_{i=2}^{m} h_{\ell k}\big(\tfrac{i-1}{m}\big)\big(\overline{X}_{i,m}^2(\lambda) - \big(\int_0^1 \Lambda(ms-(i-2)) dX_s\big)^2\big) \Big|$$
$$> \overline{c}\delta \sqrt{\tfrac{p\log m}{m}} \text{ and } \sigma^2 \in \mathcal{B}_{\pi,\infty}^s(c) \Big)$$

$$II := \mathbb{P}\Big(\Big| \sum_{i=2}^{m} \int_0^1 H_{t,i} dX_t \Big| > 2\overline{c}\left(1 + \tfrac{\delta}{2}\right)\sqrt{\tfrac{p\log m}{m}} \text{ and } \sigma^2 \in \mathcal{D}_\infty(\overline{c}) \Big)$$

$$III := \mathbb{P}\Big(\Big| \sum_{i=2}^{m} h_{\ell k}\big(\tfrac{i-1}{m}\big)\big(\int_0^1 \Lambda^2(ms-(i-2))\, \sigma_s^2 ds - \langle \sigma^2, h_{lk} \rangle_{L^2} \big) \Big|$$
$$> \overline{c}\delta \sqrt{\tfrac{p\log m}{m}} \text{ and } \sigma^2 \in \mathcal{B}_{\pi,\infty}^s(c) \Big).$$

Note that $\mathbb{P}(X > t \text{ and } B) = \mathbb{E}\left(\mathbb{I}_{\{X>t\}\cap B}\right) \leq t^{-p}\, \mathbb{E}\left(X^p\, \mathbb{I}_B\right)$, for $p \geq 0$. Using $m \leq n^{1/2}$ and (4.7) we find that I can be bounded by any polynomial order of $1/m$.

The term II can be bounded further by $II \leq II_{\text{even}} + II_{\text{odd}}$, with

$$II_{\text{even / odd}} := \mathbb{P}\Big(\Big| \sum_{i=2,\ i \ \text{even / odd}}^{m} \int_0^{T_{\overline{c}}} H_{t,i} dX_t \Big| > \overline{c}\left(1 + \tfrac{\delta}{2}\right)\sqrt{\tfrac{p\log m}{m}} \Big).$$

Since h has support $[0,1]$, $h_{\ell k}(\frac{2i-1}{m}) \neq 0$ can happen only if $\frac{1}{2}(k2^{-l}m+1) \leq i \leq \frac{1}{2}((k+1)2^{-l}m+1)$. We will treat the term II_{even} only, since similar arguments apply for II_{odd}. The process $M_r := 2^{-l/2}m\sum_{i=1}^{r}\int_0^{T_{\overline{c}}} H_{t,2i}dX_t$ is a martingale with respect to the filtration $\mathcal{F}_r = \sigma\left(X_s : s \leq 2r/m\right)$ starting at $M_{\lfloor (k2^{-l}m+1)/2 \rfloor} = 0$. Recall that $H_{t,2i}$ vanishes outside $[2(i-1)/m, 2i/m]$ and $\mathbb{I}_{\{T_{\overline{c}} \leq (2i-2)/m\}}$ is $\mathcal{F}_{i-1}$ measurable. Now, using Lemma 4.2 and conditional Ito-isometry

$$\begin{aligned}
\langle M \rangle_{\lfloor \frac{1}{2}((k+1)2^{-l}m+1) \rfloor} &\leq 2^{-l}m^2\overline{c}\sum_{i=1}^{\lfloor m/2 \rfloor}\int_0^1 \mathbb{E}\left(H^2_{s\wedge T_{\overline{c}},2i}|\mathcal{F}_{i-1}\right)ds \\
&\leq 2^{-l-1}\overline{c}^2 m\sum_{i=1}^{\lfloor m/2 \rfloor} h^2_{\ell k}\left(\tfrac{2i-1}{m}\right) \leq 2^{-l}m\ \tfrac{1}{4}\overline{c}^2\left(1+\delta\right),
\end{aligned}$$

where the last step follows for all $m \geq m_0$ and m_0 is fixed and independent of ℓ, k. Furthermore, by BDG, we bound

$$\begin{aligned}
\mathbb{E}\left[|\Delta M_i|^\kappa\right] &\lesssim 2^{-l\kappa/2}m^\kappa\,\mathbb{E}\left[\Big|\int_0^1 H_{t,2i}\mathbb{I}_{[0,T_{\overline{c}}]}(t)dX_t\Big|^\kappa\right] \\
&\lesssim 2^{-l\kappa/2}m^\kappa\,\mathbb{E}\left[\Big|\int_0^1 H^2_{t\wedge T_{\overline{c}},2i}dt\Big|^{\kappa/2}\right] \\
&\lesssim 2^{-l\kappa/2}m^{\kappa/2}\,\mathbb{E}\left[\sup_{t\leq 2/m}\big|H_{(t+(i-2)/m)\wedge T_{\overline{c}},2i}\big|^\kappa\right] \\
&\lesssim 2^{-l\kappa/2}\big|h_{\ell k}\left(\tfrac{i-1}{m}\right)\big|^\kappa \lesssim 1
\end{aligned}$$

uniformly over i. Now, we may apply Lemma 4.11 for $j \sim m2^{-l}$ and obtain $II_{\text{even}} \lesssim m^{-p}$.

In the same way we bound II_{odd} and thus obtain $II \lesssim m^{-p}$.

In order to bound III it follows from $m^{-(s-1/\pi)}|h_{\ell k}|_{1,m} \lesssim m^{-1/2}$, (4.11), (4.12), and (4.13) that for sufficiently large m on $\sigma^2 \in \mathcal{B}^s_{\pi,\infty}(c)$

$$\Big|\sum_{i=2}^{m} h_{\ell k}\left(\tfrac{i-1}{m}\right)\left(\int_0^1 \Lambda^2\left(ms-(i-2)\right)\sigma_s^2 ds - \left\langle \sigma^2, h_{lk}\right\rangle_{L^2}\right)\Big| \leq \overline{c}\delta\sqrt{\tfrac{p\log m}{m}}.$$

This yields the conclusion. □

Lemma 4.13. *Work under Assumptions of Proposition 3.3. Then we* *have for every fixed* $\delta > 0$

$$\mathbb{P}\Big(\Big|\sum_{i=2}^{m} h_{\ell k}\left(\tfrac{i-1}{m}\right)\overline{X}_{i,m}(\lambda)\,\overline{\epsilon}_{i,m}(\lambda)\Big| > \sqrt{8\overline{c}}\,\|a\|_{L^\infty}\,\|\widetilde{\lambda}\|_{L^2}(1+\delta)\sqrt{\tfrac{p\log m}{m}} \text{ and } \sigma^2 \in \mathcal{B}^s_{\pi,\infty}(c)\Big) \lesssim m^{-p},$$

where $\overline{c}(s,\pi,c)$ *is such that* $\mathcal{B}^s_{\pi,\infty}(c) \subset \mathcal{D}_\infty(\overline{c})$.

Proof. Let $\overline{X}_{i,m,T_{\overline{c}}}$ be defined as $\overline{X}_{i,m}$ with $X_{j/n}$ replaced by $X_{j/n\wedge T_{\overline{c}}}$. Then by separating even and odd terms it suffices to show

$$\mathbb{P}\Big(\Big|\sum_{i=2,\ i\,\text{even}}^{m} h_{\ell k}\left(\tfrac{i-1}{m}\right)\overline{X}_{i,m,T_{\overline{c}}}(\lambda)\,\overline{\epsilon}_{i,m}\Big| > \sqrt{2\overline{c}}\,\|a\|_{L^\infty}\,\|\widetilde{\lambda}\|_{L^2}(1+\delta)\sqrt{\tfrac{p\log m}{m}}\Big) \lesssim m^{-p}$$

since the same argumentation can be done for the sum over odd i. Similar as in the proof of Lemma 4.12, $M_r = n^{1/2}2^{-l/2}\sum_{i=1}^{2r} h_{\ell k}\left(\frac{2i-1}{m}\right)\overline{X}_{2i,m,T_{\overline{c}}}\overline{\epsilon}_{2i,m}$ defines a martingale with respect to the filtration $\mathcal{F}_r^{\text{even}}$, starting at $M_{\lfloor (k2^{-l}m+1)/2\rfloor} = 0$.

$$\begin{aligned}\langle M\rangle_{\lfloor \frac{1}{2}((k+1)2^{-l}m+1)\rfloor} &\leq n2^{-l}\sum_{i=1}^{\lfloor m/2\rfloor} h^2_{\ell k}(\tfrac{2i-1}{m})\,\mathbb{E}(\overline{X}^2_{2i,m,T_{\overline{c}}}\overline{\epsilon}^2_{2i,m}|\mathcal{F}^{\text{even}}_{i-1})\\ &\leq n2^{-l}\,\|a\|^2_{L^\infty}\sum_{i=1}^{\lfloor m/2\rfloor} h^2_{\ell k}(\tfrac{2i-1}{m})\,\mathbb{E}(\overline{X}^2_{2i,m,T_{\overline{c}}}|\mathcal{F}^{\text{even}}_{i-1})\\ &\qquad\times \frac{m^2}{n^2}\sum_{\frac{j}{n}\in\left(\frac{2i-2}{m},\frac{2i}{m}\right]}\widetilde{\lambda}^2(m\tfrac{j}{n}-(2i-2)).\end{aligned}$$

By the assumed piecewise Lipschitz continuity of λ it follows

$$\frac{m}{n}\sum_{\frac{j}{n}\in\left(\frac{2i-2}{m},\frac{2i}{m}\right]}\widetilde{\lambda}^2(m\tfrac{j}{n}-(2i-2)) = \|\widetilde{\lambda}\|^2_{L^2} + O\left(\tfrac{m}{n}\right), \tag{4.18}$$

uniformly in i. Next, we will derive a bound for $\mathbb{E}(\overline{X}^2_{2i,m,T_{\overline{c}}}|\mathcal{F}^{\text{even}}_{i-1})$. Note that $\overline{X}_{2i,m,T_{\overline{c}}} = U_1 + U_2$, with

$$U_1 := \tfrac{m}{n} \sum_{\frac{j}{n}\in\left(\frac{2i-2}{m},\frac{2i}{m}\right]} \Big(\sum_{l=j}^{n} \widetilde{\lambda}(m\tfrac{l}{n} - (2i-2))\Big)\big(X_{\frac{j}{n}\wedge T_{\overline{c}}} - X_{\frac{j-1}{n}\wedge T_{\overline{c}}\wedge\frac{2i-2}{m}}\big),$$

$$U_2 := X_{\frac{2i-2}{m}\wedge T_{\overline{c}}} \tfrac{m}{n} \sum_{\frac{j}{n}\in\left(\frac{2i-2}{m},\frac{2i}{m}\right]} \widetilde{\lambda}(m\tfrac{j}{n} - (2i-2)).$$

Clearly, $\mathbb{E}(\overline{X}^2_{2i,m,T_{\overline{c}}}|\mathcal{F}^{\text{even}}_{i-1}) = \mathbb{E}(U_1^2|\mathcal{F}^{\text{even}}_{i-1}) + U_2^2$. By conditional Ito-isometry

$$\begin{aligned}&\mathbb{E}\left(\big(X_{\frac{j}{n}\wedge T_{\overline{c}}} - X_{\frac{j-1}{n}\wedge T_{\overline{c}}\wedge\frac{2i-2}{m}}\big)\big(X_{\frac{j'}{n}\wedge T_{\overline{c}}} - X_{\frac{j'-1}{n}\wedge T_{\overline{c}}\wedge\frac{2i-2}{m}}\big)|\mathcal{F}^{\text{even}}_{i-1}\right) \leq \delta_{j,j'}\overline{c}\tfrac{1}{n}\\ &= \overline{c}\,\mathbb{E}\Big(\big(W_{\frac{j}{n}} - W_{\frac{j-1}{n}}\big)\big(W_{\frac{j'}{n}} - W_{\frac{j'-1}{n}}\big)\Big), \qquad \text{for } \tfrac{j}{n}, \tfrac{j'}{n} \in (\tfrac{2i-2}{m}, \tfrac{2i}{m}],\end{aligned}$$

where W denotes a standard Brownian motion. Lemma 4.2 and setting $X = W$ in (4.6) yields

$$\begin{aligned}\mathbb{E}(U_1^2|\mathcal{F}_{i-1}) &\leq \overline{c}\,\mathbb{E}\Big[\Big(\tfrac{m}{n}\sum_{\frac{j}{n}\in\left(\frac{2i-2}{m},\frac{i}{m}\right]} \widetilde{\lambda}(m\tfrac{j}{n} - (2i-2))W_{j/n}\Big)^2\Big]\\ &= \overline{c}\,\mathbb{E}\Big(\int_0^1 \Lambda^2\big(ms - (2i-2)\big)ds\Big) + O\big(m^{-1/2}n^{-1}\big)\\ &= \overline{c}m^{-1} + O\big(m^{-1/2}n^{-1}\big)\end{aligned}$$

uniformly over i. By using (4.15) we infer $U_2^2 \lesssim \frac{m^2}{n^2}X^2_{\frac{2i-2}{m}\wedge T_{\overline{c}}}$. Thus, we obtain for the predictable quadratic variation, $\delta_1, \delta_2 > 0$ and sufficiently large m

$$\begin{aligned}\langle M\rangle_{\lfloor\frac{1}{2}((k+1)2^{-l}m+1)\rfloor} &\leq 2^{-l-1}m\,\|a\|^2_{L^\infty}\,\overline{c}\|\widetilde{\lambda}\|^2_{L^2}(1+\delta_1)\\ &\quad + 2^{-l}\tfrac{m^3}{n^2}\,\|a\|^2_{L^\infty}\,\|\widetilde{\lambda}\|^2_{L^2}(1+\delta_2)\sum_{i=1}^{\lfloor m/2\rfloor} h^2_{\ell k}(\tfrac{2i-1}{m})X^2_{\frac{2i-2}{m}\wedge T_{\overline{c}}}.\end{aligned}$$

Now, for any fixed $\delta_3 > 0$ we find by Chebycheff inequality that

$$\tfrac{m^2}{n^2}\sum_{i=1}^{\lfloor m/2\rfloor} h^2_{\ell k}(\tfrac{2i-1}{m})X^2_{\frac{2i-2}{m}\wedge T_{\overline{c}}} \lesssim \tfrac{m^3}{n^2}\sup_{s\leq T_{\overline{c}}} X_s^2 \leq \delta_3,$$

with probability larger than $1-m^{-p}$. Hence for $\delta > 0$ we may find $\delta_1, \delta_2, \delta_3$ such that

$$\mathbb{P}\Big(\langle M\rangle_{\lfloor\frac{1}{2}((k+1)2^{-l}m+1)\rfloor} > 2^{-l-1}m\,\|a\|^2_{L^\infty}\,\overline{c}\|\widetilde{\lambda}\|^2_{L^2}(1+\delta)\Big) \lesssim m^{-p}.$$

In the next step, we bound $\max_i \mathbb{E}[|\Delta M_i|^\kappa]$. In the proof of Lemma 4.6, we already derived $\mathbb{E}[|\overline{X}_{i,m}(\lambda)|^{2\kappa}] \lesssim m^{-\kappa}$ and $\mathbb{E}[|\overline{\epsilon}_{i,m}|^{2\kappa}] \lesssim m^\kappa n^{-\kappa}$. By the same arguments we obtain also $\mathbb{E}[|\overline{X}_{i,m,T_{\overline{c}}}(\lambda)|^{2\kappa}] \lesssim m^{-\kappa}$. Therefore, it is easy to see that

$$\max_i \mathbb{E}[|\Delta M_i|^\kappa] \lesssim 2^{-l\kappa/2}n^{\kappa/2}\big|h(\tfrac{i-1}{m})\big|^\kappa\,\mathbb{E}^{1/2}[|\overline{X}_{i,m,T_{\overline{c}}}(\lambda)|^{2\kappa}]\,\mathbb{E}^{1/2}[|\overline{\epsilon}_{i,m}|^{2\kappa}] \lesssim 1.$$

Hence the assumptions of Lemma 4.11 are satisfied and the conclusion follows. □

Lemma 4.14. *Work under Assumptions of Proposition 3.3. Let $\mathcal{G}$ denote the σ-field generated by $(X_s, s \in [0,1])$. Then we have for every fixed $\delta > 0$*

$$\mathbb{P}\Big(\Big|\sum_{i=2}^m h_{\ell k}\left(\tfrac{i-1}{m}\right)\left(\overline{\epsilon}^2_{i,m}(\lambda) - \mathbb{E}[\overline{\epsilon}^2_{i,m}(\lambda)|\mathcal{G}]\right)\Big| > 4\,\|a\|^2_{L^\infty}\,\|\widetilde{\lambda}\|^2_{L^2}\,(1+\delta)\sqrt{\tfrac{p\log m}{m}}\Big) \lesssim m^{-p}.$$

Proof. We show that

$$\mathbb{P}\Big(\Big|\sum_{i=2,\ i\,\mathrm{even}}^m h_{\ell k}\left(\tfrac{i-1}{m}\right)\left(\overline{\epsilon}^2_{i,m}(\lambda) - \mathbb{E}[\overline{\epsilon}^2_{i,m}(\lambda)|\mathcal{G}]\right)\Big| > 2\,\|a\|^2_{L^\infty}\,\|\widetilde{\lambda}\|^2_{L^2}\,(1+\delta)\sqrt{\tfrac{p\log m}{m}}\Big) \lesssim m^{-p}$$

and argue similar for the sum over i odd. Let $\mathcal{F}^{\mathrm{even}}_r$, U_i and the martingale S^{even}_r be defined as in the proof of Lemma 4.5 with g replaced by $h_{\ell k}$. Now $h_{\ell k}(\frac{2i-1}{m}) \neq 0$ can happen only if $\frac{1}{2}(k2^{-l}m+1) \leq i \leq \frac{1}{2}((k+1)2^{-l}m+1)$. In the following we will consider the martingale $M_r := \frac{n}{m}2^{-l/2}S^{\mathrm{even}}_r$ started at $M_{\lfloor(k2^{-l}m+1)/2\rfloor} = 0$. We obtain

$$\langle M\rangle_{\lfloor\frac{1}{2}((k+1)2^{-l}m+1)\rfloor} \leq \frac{n^2}{m^2}2^{-l}\sum_{i=1}^{\lfloor m/2\rfloor} h^2_{\ell k}(\tfrac{2i-1}{m})\,\mathbb{E}\left[\left(\overline{\epsilon}^2_{2i,m} - \mathbb{E}[\overline{\epsilon}^2_{2i,m}|\mathcal{G}]\right)^2|\mathcal{F}^{\mathrm{even}}_{i-1}\right].$$

Elementary calculations and (4.18) show further that we may find a deterministic bound, i.e. uniformly in i

$$\begin{aligned}
&\mathbb{E}\left[\left(\bar{\epsilon}_{i,m}^2 - \mathbb{E}[\bar{\epsilon}_{i,m}^2|\mathcal{G}]\right)^2|\mathcal{F}_{i-1}^{\text{even}}\right] \\
&= 2\,\|a\|_{L^\infty}^4 \left(\tfrac{m^2}{n^2} \sum_{\frac{j}{n}\in\left(\frac{i-2}{m},\frac{i}{m}\right]} \widetilde{\lambda}^2(m\tfrac{j}{n}-(i-2))\right)^2 + O\left(\tfrac{m^3}{n^3}\right) \\
&= 2\tfrac{m^2}{n^2}\,\|a\|_{L^\infty}^4\,\|\widetilde{\lambda}\|_{L^2}^4 + O\left(\tfrac{m^3}{n^3}\right).
\end{aligned}$$

From this we obtain for sufficiently large m,

$$\langle M\rangle_{\lfloor\frac{1}{2}((k+1)2^{-l}m+1)\rfloor} \leq m2^{-l}\,\|a\|_{L^\infty}^4\,\|\widetilde{\lambda}\|_{L^2}^4(1+\delta).$$

By (4.14), we infer $\mathbb{E}[|\Delta M_i|^\kappa] \lesssim 1$. Applying Lemma 4.11 yields the conclusion. □

Completion of proof of Theorem 3.3

Let I, II, and III be defined as in (4.16).

- **The term I.** By Lemma 4.12, we have

$$\mathbb{P}\left(|I| > 4\bar{c}(1+\delta)\sqrt{\tfrac{p\log m}{m}} \text{ and } \sigma^2 \in \mathcal{B}_{\pi,\infty}^s(c)\right) \lesssim m^{-p}.$$

- **The term II.** Applying Lemmas 4.14, 4.7, 4.8 and 4.9, we derive by Chebycheff's inequality and $|h_{\ell k}|_{p,m}^p \lesssim m^{p/2-1}$, $p \geq 2$

$$\mathbb{P}\left(|II| > 4\,\|a\|_{L^\infty}^2\,\|\widetilde{\lambda}\|_{L^2}^2(1+\delta)\sqrt{\tfrac{p\log m}{m}} \text{ and } \sigma^2 \in \mathcal{B}_{\pi,\infty}^s(c)\right) \lesssim m^{-p}.$$

- **The term III.** We find by Lemma 4.13

$$\mathbb{P}\left(|III| > 4\sqrt{2\,\bar{c}}\,\|a\|_{L^\infty}\|\widetilde{\lambda}\|_{L^2}(1+\delta)\sqrt{\tfrac{p\log m}{m}} \text{ and } \sigma^2 \in \mathcal{B}_{\pi,\infty}^s(c)\right) \lesssim m^{-p}.$$

The proof of Theorem 3.3 is complete.

4.3 Proof of Theorem 2.9

We readily apply the bounds of the wavelet threshold algorithm over atomic spaces, as developed by Kerkyacharian and Picard in [32]. By assumption, we have $s - 1/\pi \geq \alpha_0$ and $2^{\ell_0} \sim m^{1-2\alpha_0}$ therefore, the term $m^{-\min\{s-1/\pi,1\}p} |h_{\ell k}|_{1,m}^p$ is less than a constant times

$$m^{-\alpha_0} 2^{-\ell/2} \lesssim m^{-\alpha_0} m^{(1-2\alpha_0)/2} \sim m^{-1/2},$$

where we used that $|h_{\ell k}|_{1,m} \lesssim 2^{-\ell/2}$ with $h = \varphi$. This together with Theorem 3.2 shows that Condition (5.1) of Theorem 5.1 in Kerkyacharian and Picard [32] is satisfied with $c(n) = n^{-1/4}$ and $\Lambda(n) = n^{1/2}$ (with the notation of [32]). Likewise for their Condition (5.2) thanks to Theorem 3.3. By applying successively their abstract Corollary 5.2 and Theorem 6.1, the result follows after some elementary computations. (Alternatively, one can also see [22] for a derivation of Theorem 6.1 of [32] where the formalism of the effective smoothness function $s(t)$ is explicitly used, up to losing some inessential logarithmic factors.)

4.4 Proof of Theorem 2.11

Preliminaries

Let $(C, \mathcal{C})$ denote the space of continuous functions on $[0,1]$, equipped with the norm of uniform convergence and its Borel σ-field $\mathcal{C}$. Let $(\Omega', \mathcal{F}', \mathbb{P}')$ be another probability space rich enough to contain an infinite sequence of i.i.d. Gaussian random variables. On $(\widetilde{\Omega}, \widetilde{\mathcal{F}}) := (C \times C \times \Omega', \mathcal{C} \otimes \mathcal{C} \otimes \mathcal{F}')$ we construct a probability measure $\widetilde{\mathbb{P}}$ as follows. Let $(\sigma, \omega, \omega')$ denote a generic element of $\widetilde{\Omega}$.

We pick an arbitrary probability measure $\mu(d\sigma)$ on $(C, \mathcal{C})$, and we construct the measure $\mathbb{P}_\sigma(d\omega)$ on $(C, \mathcal{C})$ such that, under $\mathbb{P}_\sigma$, the canonical process X on C is a solution (in a weak sense for instance) to

$$X_t = X_0 + \int_0^t \sigma_s \, dW_s,$$

where W is a standard Wiener process. We then set

$$\widetilde{\mathbb{P}} := \mu(d\sigma) \otimes \mathbb{P}_\sigma(d\omega) \otimes \mathbb{P}'(d\omega').$$

This space is rich enough to contain our model: indeed, by construction, any $\mu(d\sigma)$ will be such that, under μ, we have Assumption 2.1. By constructing on $(\Omega', \mathcal{F}, \mathbb{P}')$ an i.i.d. Gaussian noise $(\epsilon_{j,n})$ for $j = 0, \ldots, n$ with constant variance function $a^2 > 0$ for a given $a^2 > 0$, the space $\widetilde{\Omega}$ is rich enough to contain an additive Gaussian microstructure noise, independent of X, and we have Assumption 2.2. Next, consider the statistical experiment

$$\mathcal{E}_n = \big(C \times \Omega', \mathcal{C} \otimes \mathcal{F}', (\mathbb{P}_\sigma^n, \sigma \in \mathcal{D})\big),$$

where $\mathcal{D} \subset C$ and $\mathbb{P}_\sigma^n$ is the law of the data $(Z_{j,n})$, conditional on σ. The probability $\mu(d\sigma)$ can be interpreted as a prior distribution for the "true" parameter σ.

Completion of proof. We are ready to prove Theorem 2.11. Let $\mathcal{D} = \mathcal{B}^s_{\pi,\infty}(c)$ denote a Besov ball such that $s-1/\pi > (1+\sqrt{5})/4$. Then $\mathcal{D} \subset C$. Assume further that μ is such that $\mu\big[\mathcal{D}\big] = 1$. Then Condition (2.5) is satisfied.

Moreover, for any normalizing factor $v_n > 0$ and any $c' > 0$, we have

$$\begin{aligned}
&\liminf_{n\to\infty} \inf_{\widehat{\sigma}_n} v_n^{-1} \widetilde{\mathbb{E}}\big[\|\widehat{\sigma}_n^2 - \sigma^2\|_{L^p([0,1])} \mathbb{I}_{\big\{\sigma^2 \in \mathcal{B}^s_{\pi,\infty}(c)\big\}}\big] \\
\geq & c' \liminf_{n\to\infty} \inf_{\widehat{\sigma}_n} \int_C \mu(d\sigma)\, \mathbb{P}_\sigma^n \left[v_n^{-1}\|\widehat{\sigma}_n^2 - \sigma^2\|_{L^p([0,1])} \geq c'\right]
\end{aligned} \tag{4.19}$$

since $\mu\big[\mathcal{D}\big] = 1$.

Let us now consider the statistical experiment $\mathcal{E}_n'$ generated by the observation of the Gaussian measure

$$Y_n = \sqrt{2\sigma} + a n^{-1/4} \dot{B}$$

where $\dot{B}$ is a Gaussian white noise, with same parameter space $\mathcal{D}$. We denote by $\mathbb{Q}_\sigma^n$ the law of Y_n. By picking $\mu(d\sigma)$ as the least favourable prior in order to obtain lower bounds over Besov classes (see for instance [26]) we know that, for any $c' > 0$

$$\liminf_{n\to\infty} \inf_{(\widehat{\sigma}_n^2)'} \int_C \mu(d\sigma) \mathbb{Q}_\sigma^n \big[n^{-s(\pi^\star)/(4s(\pi^\star)+2)} \|(\widehat{\sigma}_n^2)' - \sigma^2\|_{L^p([0,1])} \geq c'\big] > 0, \tag{4.20}$$

where the infimum is taken over all estimators $(\widehat{\sigma}_n^2)'$ in the experiment $\mathcal{E}_n'$. This follows from classical analysis of the white Gaussian noise model, see again [26] (or [22] for the notation encompassing the effective smoothness function $s(\pi^\star)$ with noise level $n^{-1/4}$).

It remains to relate (4.19) and (4.20). By the result of Reiß [39], since $s-1/\pi > (1+\sqrt{5})/4$, we have that $\mathcal{E}_n$ and $\mathcal{E}_n'$ are asymptotically equivalent as $n \to \infty$. This means that we can approximate $\mathbb{P}_\sigma^n$ by $\mathbb{Q}_\sigma^n$ in variational norm, uniformly on σ, up to randomization via a Markov kernel that does not depend on σ. Therefore, the lower bound (4.20) automatically transfers to (4.19), up to considering an extension of the space so that randomized decisions (or estimators) can be considered too (see Le Cam and Yang [33]). (However, the approximation is valid only up to the extension of estimators to the larger class of randomized procedures; since we are considering a Bayesian decision problem only, the extra technicality coming from the randomization can be ignored by conditioning on the randomization and applying Fubini. We leave out these inessential details.) The proof is complete. $\square$

Acknowledgment

The research of Axel Munk and Johannes Schmidt-Hieber was supported by DFG Grant FOR 916 and GK 1023.

References

[1] F. Abramovich and B. W. Silverman. Wavelet decomposition approaches to statistical inverse problems. *Biometrika*, 85:115–129, 1998.

[2] Y. Ait-Sahalia, P. A. Mykland, and L. Zhang. How often to sample a continuous-time process in the presence of market microstructure noise. *The Review of Financial Studies*, 18:351–416.

[3] Y. Ait-Sahalia, P. A. Mykland, and L. Zhang. Ultra high frequency volatility estimation with dependent microstructure noise. *Journal of Econometrics*. Forthcoming.

[4] T. G. Andersen, T. Bollerslev, F. X. Diebold, and P. Labys. The distribution of realized exchange rate volatility. *Journal of the American Statistical Association*, 96:42–55, 2001.

[5] F. M. Bandi and J. R. Russell. Separating microstructure noise from volatility. *Journal of Financial Economics*, 79:655–692, 2006.

[6] F. M. Bandi and J. R. Russell. Market microstructure noise, integrated variance estimators, and the accuracy of asymptotic approximations. Working paper, 2009.

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

[8] B. Bercu and A. Touati. Exponential inequalities for self-normalized martingales with application. *Ann. Appl. Prob.*, 18(5):1848–1869, 2008.

[9] J.-P. Bouchaud and M. Potters. *Theory of Financial Risks: From Statistical Physics to Risk Management*. Cambridge University Press, 2000.

[10] T. Cai, A. Munk, and J. Schmidt-Hieber. Sharp minimax estimation of the variance of Brownian motion corrupted with Gaussian noise. *Statist. Sinica*, 20:1011–1024, 2010.

[11] Z. Ciesielski, G. Kerkyacharian, and B. Roynette. Quelques espaces fonctionnels associ a des processus gaussiens. *Studia Mathematica*, 110(3):171–204, 1994.

[12] A. Cohen. *Numerical Analysis of Wavelet Methods*. Elsevier, 2003.

[13] A. Cohen, I. Daubechies, and P. Vial. Wavelets on the interval and fast wavelet transforms. *Appl. Comput. Harmon. Anal.*, 1:54–81, 1993.

[14] F. X. Diebold and G. H. Strasser. On the correlation structure of microstructure noise in theory and practice. *Working Paper*, 2008.

[15] D. Donoho and I. M. Johnstone. Ideal spatial adaptation via wavelet shrinkage. *Biometrika*, 81:425–455, 1994.

[16] D. Donoho, I. M. Johnstone, G. Kerkyacharian, and D. Picard. Wavelet shrinkage: Asymptopia? *J. R. Statist. Soc. B*, 57(2):301–369, 1995.

[17] R. F. Engle. The econometrics of ultra-high-frequency data. *Econometrica*, 68(1):1–22, 2000.

[18] J. Fan. A selective overview of nonparametric methods in financial econometrics. *Statistical Science*, 20:317–357, 2005.

[19] V. Genon-Catalot, C. Laredo, and D. Picard. Non-parametric estimation of the diffusion coefficient by wavelets methods . *Scand J. Statist.*, 19:317–335, 1992.

[20] A. Gloter. Estimation du coefficient de diffusion de la volatilité d'un modèle à volatilité stochastique. *Ann. Statist.*, 35:1947–1974, 2007.

[21] A. Gloter and M. Hoffmann. Estimation of the Hurst parameter from discrete noisy data. *Ann. Statist.*, 35:1947–1974, 2007.

[22] A. Gloter and M. Hoffmann. Nonparametric reconstruction of a multifractal function from noisy data. *Probab. Theory Relat. Fields*, 146:155–187, 2010.

[23] A. Gloter and J. Jacod. Diffusions with measurement errors. I. Local asymptotic normality. *ESAIM Probab. Stat.*, 5:225–242, 2001.

[24] A. Gloter and J. Jacod. Diffusions with measurement errors. II. Optimal estimators. *ESAIM Probab. Stat.*, 5:243–260, 2001.

[25] P. Hall and C. C. Heyde. *Martingale Limit Theory and Its Applications*. Academic Press, New York, 1980.

[26] W. Härdle, D. Picard, A. Tsybakov, and G. Kerkyacharian. *Wavelets, Approximation and Statistical Applications*. Springer, Berlin, 1998.

[27] J. Hasbrouck. Assessing the quality of a security market: A new approach to transaction-cost measurement. *The Review of Financial Studies*, 6(1):191–212, 1993.

[28] M. Hoffmann. Minimax estimation of the diffusion coefficient through irregular sampling. *Statistics and Probability Letters*, 32(1):11–24, 1997.

[29] M. Hoffmann. Adaptive estimation in diffusion processes. *Stochastic Process. Appl.*, 79:135–163, 1999.

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

[31] I. Karatzas and S. E. Shreve. *Brownian Motion and Stochastic Calculus*. Springer, New York, 1998.

[32] G. Kerkyacharian and D. Picard. Thresholding algorithms, maxisets and well-concentrated bases. *Test*, 9:283–345, 2000.

[33] L. Le Cam and L. O. Yang. *Asymptotics in Statistics: Some Basic Concepts*. Springer, New York, 2000.

[34] A. Munk and J. Schmidt-Hieber. Nonparametric estimation of the volatility function in a high-frequency model corrupted by noise. arXiv:0908.3163, Math arXiv Preprint.

[35] A. Munk and J. Schmidt-Hieber. Lower bounds for volatility estimation in microstructure noise models. *A Festschrift for Larry Brown, IMS Lecture Notes Series*, 2010. To appear.

[36] M. Musiela and M. Rutkowski. *Martingale Methods in Financial Modelling*. Springer, New York, 2004.

[37] P. A. Mykland and L. Zhang. Discussion of paper A selective overview of nonparametric metods in financial econometrics by Jianqing Fan. *Statistical Science*, 20:347–350, 2005.

[38] M. Podolskij and M. Vetter. Estimation of volatility functionals in the simultaneous presence of microstructure noise and jumps. *Bernoulli*, 15:634–658, 2009.

[39] M. Reiß. Asymptotic equivalence and sufficiency for volatility estimation under microstructure noise. arxiv:1001.3006, Math arXiv Preprint.

[40] R. Roll. A simple model of the implicit bid-ask spread in an efficient market. *Journal of Finance*, 39:1127–1139, 1984.

[41] M. Rosenbaum. Estimation of the volatility persistence in a discretely observed diffusion model. *Stochastic Process. Appl.*, 118(08):1434–1462, 2008.

[42] B.-D. Seo. Realized volatility and colored market microstructure noise. Manuscript, 2005.

[43] L. Zhang. Efficient estimation of stochastic volatility using noisy observations: A multi-scale approach. *Bernoulli*, 12:1019–1043, 2006.

[44] L. Zhang, P. Mykland, and Y. Ait-Sahalia. A tale of two time scales: Determining integrated volatility with noisy high-frequency data. *J. Amer. Statist. Assoc.*, 472:1394–1411, 2005.