Annales de l'Institut Henri Poincaré - Probabilités et Statistiques
2012, Vol. 48, No. 4, 1186–1216
DOI: 10.1214/11-AIHP472

ANNALES DE L'INSTITUT HENRI POINCARÉ
PROBABILITÉS ET STATISTIQUES
www.imstat.org/aihp

Adaptive wavelet estimation of the diffusion coefficient under additive error measurements

M. Hoffmann[a], A. Munk[b] and J. Schmidt-Hieber[b]

[a]*ENSAE and CNRS-UMR 8050, 3, avenue Pierre Larousse, 92245 Malakoff Cedex, France. E-mail: marc.hoffmann@ensae.fr*
[b]*Institut für Mathematische Stochastik, Universität Göttingen, Goldschmidtstr. 7, 37077 Göttingen, Germany.*
E-mail: amunk1@gwdg.de; schmidth@math.uni-goettingen.de

Received 14 December 2010; revised 19 December 2011; accepted 26 December 2011

Abstract. We study nonparametric estimation of the diffusion coefficient from discrete data, when the observations are blurred by additional noise. Such issues have been developed over the last 10 years in several application fields and in particular in high frequency financial data modelling, however mainly from a parametric and semiparametric point of view. This paper addresses the nonparametric estimation of the path of the (possibly stochastic) diffusion coefficient in a relatively general setting.

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 diffusion coefficient is usually 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 a deterministic diffusion coefficient. In particular, we take advantage of recent results of Reiß (*Ann. Statist.* **39** (2011) 772–802) of asymptotic equivalence between a Gaussian diffusion with additive noise and Gaussian white noise model, in order to prove a sharp lower bound.

Résumé. On étudie l'estimation non-paramétrique du coefficient de diffusion à partir d'observations discrètes, lorsque les observations sont bruitées par un bruit additionnel. De tels problèmes se sont développés au cours des dix dernières années dans plusieurs champs d'application, en particuler pour la modélisation des données haute fréquence en finance, cependant plutôt d'un point de vue paramétrique ou semi-paramétrique. Ce travail concerne l'estimation de la trajectoire (éventuellement stochastique) du coefficient de diffusion dans un cadre relativement général.

En développant des techniques de pré-moyennage combinées avec du seuillage des coefficients d'ondelettes, nous contruisons des estimateurs adaptatifs qui atteignent une vitesse quasi-optimale parmi une vaste échelle de contraintes de régularité de type Besov. Puisque le coefficient de diffusion est souvent intrinsèquement aléatoire, nous proposons un nouveau critère pour qualifier la qualité d'estimation ; nous retrouvons la théorie minimax usuelle lorsque cette approche est restreinte à un coefficient de diffusion déterministe. En particulier, on exploite les résultats récents de Reiß (*Ann. Statist.* **39** (2011) 772–802) de l'équivalence asymptotique entre une diffusion gaussienne avec un bruit additif et le bruit blanc gaussien.

MSC: 62G99; 62M99; 60G99

Keywords: Adaptive estimation; Besov spaces; Diffusion processes; Nonparametric regression; Wavelet estimation

1. Introduction

We are interested in the following statistical setting: we assume that we have real-valued data of 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 a sampling time, $(\epsilon_{j,n})$ is an additive noise process[1] and the continuous time process $X = (X_t)_{t\geq 0}$ has representation

$$X_t = X_0 + \int_0^t b_s \,\mathrm{d}s + \int_0^t \sigma_s \,\mathrm{d}W_s. \tag{1.2}$$

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. We assume that the data $(Z_{j,n})$ are sampled in a high-frequency framework: the time step Δ_n between observations goes to 0, but $n\Delta_n$ remains bounded as $n \to \infty$, i.e. the whole statistical experiment is taken over a fixed time interval.

In this asymptotic framework, the only parameter that can be consistently estimated is the unobserved path of the diffusion coefficient $t \rightsquigarrow \sigma_t^2$, and unless specified otherwise, it is random. Whereas nonparametric estimation of the diffusion coefficient from direct observation $X_{j\Delta_n}$ is a fairly well known topic when σ^2 is deterministic ([18,24] and the review paper of Fan [16]), 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, and it can be related to practical issues in several application fields. In finance for instance, by considering the $Z_{j,n}$ as the result of a *latent* or unobservable efficient price $X_{i\Delta_n}$ corrupted by microstructure effects $\epsilon_{j,n}$ at scale Δ_n, we obtain a more realistic model accounting for stylised facts on intraday scale usually attributed to bid-ask spread manipulation by market makers.[2] Considering a diffusion perturbated by noise applies in other fields as well: in the context of functional MRI or fRMI, the problem of inference for diffusion processes with error measurement has been addressed by Donnet and Samson [12,13] in an ergodic and parametric setting, when the sampling time Δ_n does not shrink to 0 as $n \to \infty$. Se also Favetto and Samson [17]. Recently, Schmisser [36] has systematically studied the nonparametric estimation of the drift and the diffusion coefficient in an ergodic and mixed asymptotic setting, when $\Delta_n \to 0$ but $n\Delta_n \to \infty$. In this paper, we consider the nonergodic case, when only the diffusion coefficient can be identified, with $\Delta_n \to 0$ and $n\Delta_n$ fixed.

1.1. *Estimating the diffusion coefficient under additive noise*: *Some history*

Estimation of a finite-dimensional parameter and nonparametric functionals

The first results about statistical inference of a diffusion with error measurement go back to Gloter and Jacod [20,21] 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 $\epsilon_{j,n}$ are Gaussian and independent, then the LAN condition holds (Local Asymptotic Normality) for $\Delta_n = n^{-1}$ with rate $n^{-1/4}$. This implies 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 [1,38,39]. Analogous approaches in various similar contexts progressively emerged in the financial econometrics literature: Podolskij and Vetter [32], Bandi and Russell [3,4], Barndorff-Nielsen et al. [5] 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 $\int_0^t \sigma_s^2 \,\mathrm{d}s$. Convergent estimators were first obtained by Ait-Sahalia et al. [1] with a suboptimal rate $n^{-1/6}$. Then the two-scale approach of Zhang [38] achieved the rate $n^{-1/4}$. The Gloter–Jacod LAN property of [20] 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 [2,11,37]) and the nature of the microstructure noise $(\epsilon_{j,n})$. It took some more time and contributions before Jacod and collaborators [26] took over the topic in 2007 with their simple and powerful pre-averaging technique, introduced earlier in a simplified context by Podolskij and Vetter [32]. In essence, it consists in first, smoothing the data as in signal denoising and then, apply a standard realised 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.

[1] Implicitly assumed to be centered for obvious identifiability purposes.

[2] This approach was grounded on empirical findings in the financial econometrics literature of the early years 2000 (among many others Ait-Sahalia et al. [1], Mykland and Zhang [31] and the references therein).

Nonparametric inference
In the nonparametric case, the problem is a little unclear. By nonparametric, one thinks of estimating the whole path $t \rightsquigarrow \sigma_t^2$. However, since $\sigma^2 = (\sigma_t^2)_{t\geq 0}$ is usually itself genuinely random, there is no "true parameter" to be estimated! 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 [29,30] proposed a Fourier estimator and showed its minimax rate optimality, extending a previous approach for the parametric setting ([7]). 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 [33] recently showed the asymptotic equivalence in the Le Cam sense with the observation of the random measure

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

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

1.2. *Our results*

The asymptotic equivalence proved in [33] 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 $\langle\cdot,\cdot\rangle$ denotes the usual L^2-inner product and $(\psi_\nu, \nu \in \mathcal{V}(\sigma))$ is a collection of wavelet functions that are localised 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 $\approx$ and the property of the ψ_ν's, we do not specify yet.
2. We then estimate $\langle \sigma^2, \psi_\nu\rangle$ 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 (Donoho, Johnstone, Kerkyacharian, Picard and collaborators [14,15,23]).

The key issue is therefore the estimation of the linear functionals

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

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: this is where we depart from the results of Jacod and collaborators [26,32]. 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 [18,24,25]. However, we only have a noisy 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 $\mathrm{Ave}(Z)_{i,m}$ an averaging of the data $(Z_{j,n})$ around the point i/m. We have

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

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

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

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 generalise (1.7) to arbitrary kernels within a certain class of oscillating pre-averaging functions, in the same spirit as in Gloter and Hoffmann [19] or Rosenbaum [34] where this technique is used for denoising stochastic volatility models corrupted by noise.

We prove in Theorems 2.9 and 3.4 an upper bound for our procedure in L^p-loss error over a fixed time horizon. Assuming that the path $t \rightsquigarrow \sigma_t^2$ has s derivatives in L^π with a prescribed probability, the upper bound is of the form $n^{-\alpha/4}$ for an explicit $\alpha = \alpha(s, p, \pi) < 1$ to within inessential logarithmic terms. We retrieve the expected results of wavelet thresholding over Besov spaces 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 [20]. It is noteworthy that, although the rates of convergence depend on the smoothness parameters (s, π), the thresholding procedure does not, and is therefore adaptive in that sense. 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 2.12 that this result is sharp, even if $t \rightsquigarrow \sigma_t^2$ is random so 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 Definitions 2.3 and 2.6. This approach is presented in details in the methodology Section 2.2.

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*

We consider a continuous adapted 1-dimensional process X of the form (1.2) on a filtered probability space $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t\geq 0}, \mathbb{P})$. Without loss of generality, we assume that $X_0 = 0$.

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 $Y_t = \int_0^t \sigma_s \,\mathrm{d}W_s$ is also unique and well defined, the laws of X and Y are equivalent on $\mathcal{F}_t$ and we have, for some $\rho > 1$

$$\mathbb{E}\bigg[\exp\bigg(\rho \int_0^t \frac{b_s}{\sigma_s^2}\,\mathrm{d}Y_s\bigg)\bigg] < \infty. \tag{2.1}$$

We consider a fixed time horizon $T = n\Delta_n$, and with no loss of generality, we take $T = 1$ hence $\Delta_n = n^{-1}$. For $j = 0, \dots, n$, we assume that we can observe a blurred version of X at times $\Delta_n j = j/n$ over the time horizon $[0, T] = [0, 1]$. The blurring accounts for microstructure noise at fine scales and takes the form

$$Z_{j,n} := X_{j/n} + \epsilon_{j,n}, \quad j = 0, 1, \dots, n, \tag{2.2}$$

where the microstructure noise process $(\epsilon_{j,n})$ is implicitly defined on the same probability space as X and satisfies

Assumption 2.2. *We have*

$$\epsilon_{j,n} = a(j/n, X_{j/n})\eta_{j,n}, \tag{2.3}$$

where the function $(t,x) \rightsquigarrow a(t,x)$ is continuous and bounded. Moreover, the random variables $(\eta_{j,n})$ are independent, and independent of X. Moreover, for every $0 \le j \le n$ and $n \ge 1$, we have

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

Given data $Z. = \{Z_{j,n}, j = 0, \dots, n\}$ following (1.1), the goal is to estimate nonparametrically the random function $t \rightsquigarrow \sigma_t^2$ over the time interval [0, 1]. Asymptotics are taken as the observation frequency $n \to \infty$.

Discussion on Assumptions 2.1 *and* 2.2

Assumption 2.1 on b and σ is relatively weak, except for the moment condition (2.1). This assumption is somewhat technical, for it enables to implicitly assume that $b = 0$. Indeed, if $\mathbb{P}_{\sigma,b}$ denotes the law of $(X_t)_{t\in[0,1]}$ with drift b and volatility σ, we have by Girsanov's theorem

$$\frac{\mathrm{d}\mathbb{P}_{\sigma,b}}{\mathrm{d}\mathbb{P}_{\sigma,0}} = \exp\left(\int_0^1 \frac{b_s}{\sigma_s^2}\,\mathrm{d}X_s - \frac{1}{2}\int_0^1 \frac{b_s^2}{\sigma_s^2}\,\mathrm{d}s\right).$$

By Hölder inequality, for a random variable Z, we derive

$$\begin{aligned}\mathbb{E}_{\sigma,b}\big[|Z|^p\big]^{1/p} &= \mathbb{E}_{\sigma,0}\left[\frac{\mathrm{d}\mathbb{P}_{\sigma,b}}{\mathrm{d}\mathbb{P}_{\sigma,0}}|Z|^p\right]^{1/p} \\ &\le \mathbb{E}_{\sigma,0}\left[\exp\left(\rho\int_0^1 \frac{b_s}{\sigma_s^2}\,\mathrm{d}X_s\right)\right]^{1/(p\rho)} \mathbb{E}_{\sigma,0}\big[|Z|^{\overline{p}}\big]^{1/\overline{p}} \end{aligned} \tag{2.4}$$

with $\overline{p} = p\rho/(\rho-1)$. Therefore, Condition (2.1) guarantees that if we have an estimate of the form $\mathbb{E}_{\sigma,0}[|Z|^p]^{1/p} \le c_p n^{-\gamma}$ for any $p \ge 1$ and for some $\gamma > 0$, then the same property holds replacing $\mathbb{P}_{\sigma,0}$ by $\mathbb{P}_{\sigma,b}$, up to a modification of the constant c_p. Thus Condition (2.1) is a useful tool that enables to condense the proofs in many places afterwards. It is satisfied as soon as σ is bounded below and b has appropriate integrability conditions. In some cases of interest where it may fail to hold, one can still proceed by working directly under $\mathbb{P}_{\sigma,b}$.

Concerning Assumption 2.2, we assume a relatively weak scheme of microstructure noise, by assuming that the $\epsilon_{j,n}$ form a martingale array that may depend on the unobserved process X through a function $t \rightsquigarrow a(t, X_t)$ as the standard deviation of the additive noise. This enables richer structures than simple additive independent noise. One may wish to relax further Assumption 2.2 by assuming a correlation decay only, but again, for technical reason, we keep to this simpler framework.

2.2. *Statistical methodology*

Recovering σ^2 over a function class $\mathcal{D}$

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.

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

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

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

We need to modify the usual notion of convergence rate. Let us denote by $\mathcal{D}$ a class of real-valued functions defined on [0, 1].

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 $\mathcal{D}$ if there exists an estimator $\widehat{\sigma}_n^2$ such that*

$$\limsup_{n\to\infty} v_n^{-1}\mathbb{E}\big[\big\|\widehat{\sigma}_n^2 - \sigma^2\big\|_{L^p([0,1])}\mathbb{I}_{\{\sigma^2\in\mathcal{D}\}}\big] < \infty. \tag{2.5}$$

Remark 2.5. *If we wish (σ_t) to be deterministic, we can make a priori assumptions so that the condition $\sigma^2 \in \mathcal{D}$ is satisfied, in which case we simply ignore the indicator in (2.5). In other cases, this condition will be satisfied with some probability (see below). But it may also well happen that for some choices of $\mathcal{D}$ we have $\mathbb{P}[\sigma^2 \in \mathcal{D}] = 0$ in which case the upper bound (2.5) becomes trivial and noninformative.*

In this context, a sound notion of optimality is unclear. We propose the following

Definition 2.6. *The rate v_n is a lower rate of convergence over $\mathcal{D}$ 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 Assumption 2.1 together with a process $(\widetilde{\epsilon}_{j,n})$ satisfying (2.3) with $\widetilde{X}$ in place of X, such that Assumption 2.2 holds, and moreover:*

$$\widetilde{\mathbb{P}}\big[\sigma^2 \in \mathcal{D}\big] > 0 \tag{2.6}$$

and

$$\liminf_{n\to\infty} v_n^{-1} \inf_{\widehat{\sigma}_n^2} \widetilde{\mathbb{E}}\big[\big\|\widehat{\sigma}_n^2 - \sigma^2\big\|_{L^p([0,1])}\mathbb{I}_{\{\sigma^2\in\mathcal{D}\}}\big] > 0, \tag{2.7}$$

where the infimum is taken over all estimators.

Let us elaborate on Definition 2.6: as already mentioned, 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. 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 a 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 (2.7). 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.

Function classes: *Wavelets and Besov spaces*

We describe the smoothness of a function by means of Besov spaces on the interval. A thorough account of Besov spaces $\mathcal{B}^s_{\pi,\infty}$ and their connection to wavelet bases in a statistical setting are discussed in details in the classical papers of Donoho et al. [15] and Kerkyacharian and Picard [28]. Let us recall some fairly classical[3] material about Besov spaces through their characterisation in terms of wavelets. We use n_0-regular wavelet bases $(\psi_\nu)_\nu$ adapted to the domain [0, 1]. More precisely, the multi-index ν concatenates the spatial index and the resolution level $j = |\nu|$. We set $\Lambda_j := \{\nu, |\nu| = j\}$ and $\Lambda := \bigcup_{j\geq -1}\Lambda_j$. Thus for $f \in L^2([0,1])$, we have

$$f = \sum_{j\geq -1}\sum_{\nu\in\Lambda_j}\langle f, \psi_\nu\rangle\psi_\nu = \sum_{\nu\in\Lambda}\langle f, \psi_\nu\rangle\psi_\nu,$$

where we have set $j := -1$ in order to incorporate the low frequency part of the decomposition. From now on the basis $(\psi_\nu)_\nu$ is fixed and depends on a regularity index n_0 which role is specified in Assumption 2.8 below.

[3] We follow closely the notation of Cohen [9].

Definition 2.7. *For $s > 0$ and $\pi \in (0, \infty]$, a function $f : [0, 1] \to \mathbb{R}$ belongs to the Besov space $\mathcal{B}^s_{\pi,\infty}([0, 1])$ if the following norm is finite*:

$$\|f\|_{\mathcal{B}^s_{\pi,\infty}([0,1])} := \sup_{j\geq -1} 2^{j(s+1/2-1/\pi)}\left(\sum_{\nu\in\Lambda_j}\left|\langle f, \psi_\nu\rangle\right|^\pi\right)^{1/\pi}, \tag{2.8}$$

with the usual modification if $\pi = \infty$.

Precise connection between this definition of Besov norm and more standard ones can be found in [9,10]. Given a basis $(\psi_\nu)_\nu$ with regularity index $n_0 > 0$, the Besov space defined by (2.8) exactly matches the usual definition in terms of modulus of smoothness for f, provided that $\pi \geq 1$ and $s \leq n_0$. A particular case include the Hölder space $\mathcal{C}^s([0, 1]) = \mathcal{B}^s_{\infty,\infty}([0, 1])$. Moreover, the following Sobolev embedding inequality holds

$$\|f\|_{\mathcal{B}^{s_2}_{\pi_2,\infty}([0,1])} \leq \|f\|_{\mathcal{B}^{s_1}_{\pi_1,\infty}([0,1])} \quad \text{for } s_1 - 1/\pi_1 = s_2 - 1/\pi_2, \pi_2 \geq \pi_1,$$

showing in particular that $\mathcal{B}^s_{\pi,\infty}([0, 1])$ is embedded into continuous functions as soon as $s > 1/\pi$. The additional properties of the wavelet basis $(\psi_\nu)_\nu$ that we need are summarized in the next assumption.

Assumption 2.8 (Properties of the basis $(\psi_\nu)_\nu$). *For $\pi \geq 1$:*

- *We have $\|\psi_\nu\|^\pi_{L^\pi([0,1])} \sim 2^{|\nu|(\pi/2-1)}$.*
- *For some arbitrary $n_0 > 0$ and for all $s \leq n_0$, $j_0 \geq 0$, we have*

$$\left\|f - \sum_{j\leq j_0}\sum_{\nu\in\Lambda_j} f_\nu\psi_\nu\right\|_{L^\pi([0,1])} \lesssim 2^{-j_0 s}\|f\|_{\mathcal{B}^s_{\pi,\infty}([0,1])}. \tag{2.9}$$

- *For any $\Lambda_0 \subset \Lambda$,*

$$\int_{[0,1]}\left(\sum_{\nu\in\Lambda_0}\left|\psi_\nu(x)\right|^2\right)^{\pi/2}\mathrm{d}x \sim \sum_{\nu\in\Lambda_0}\|\psi_\nu\|^\pi_{L^\pi([0,1])}. \tag{2.10}$$

- *If $\pi > 1$, for any sequence $(u_\nu)_{\nu\in\Lambda}$*

$$\left\|\left(\sum_{\nu\in\Lambda}|u_\nu\psi_\nu|^2\right)^{1/2}\right\|_{L^\pi([0,1])} \sim \left\|\sum_{\nu\in\Lambda}u_\nu\psi_\nu\right\|_{L^\pi([0,1])}. \tag{2.11}$$

The symbol $\sim$ means inequality in both ways, up to a constant depending on π only. The property (2.9) reflects that our definition (2.8) of Besov spaces matches the definition in term of linear approximation. Property (2.11) means an unconditional basis property and (2.10) is referred to as a superconcentration inequality see [28]. The existence of compactly supported wavelet bases satisfying Assumption 2.8 goes back to Daubechies and is discussed for instance in [9].

We are interested in the case where σ^2 may belong to various smoothness classes, that 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. These smoothness properties are usually modelled in terms of Besov balls

$$\mathcal{B}^s_{\pi,\infty}(c) := \left\{f : [0, 1] \to \mathbb{R}, \|f\|_{\mathcal{B}^s_{\pi,\infty}([0,1])} \leq c\right\}, \quad c > 0, \tag{2.12}$$

that measure smoothness of degree $s > 1/\pi$ in L^π over the interval $[0, 1]$, for $\pi \in (0, \infty)$. The restriction $s > 1/\pi$ ensures that the functions in $\mathcal{B}^s_{\pi,\infty}$ are continuously embedded into Hölder continuous functions with index $s - 1/\pi$. Besov balls also give a flexible way to describe the smoothness of the path of a continuous random process. For instance, if (σ_t) is an Itô continuous semimartingale itself with regular coefficients, we have

$$\mathbb{P}\left[\sigma^2 \in \mathcal{B}^{1/2}_{\pi,\infty}(c)\right] > 0 \quad \text{for every } \pi > 1/2.$$

If it is a smooth transformation of a fractional Brownian motion with Hurst index, H, we have $\mathbb{P}[\sigma^2 \in \mathcal{B}^H_{\pi,\infty}(c)] > 0$ for $\pi > H$ likewise. The proof of such classical results can be found in Ciesielski et al. [8].

2.3. *Achievable estimation error bounds*

For prescribed smoothness classes of the form $\mathcal{D} = \mathcal{B}^s_{\pi,\infty}(c)$ and L^p-loss functions, the rate of convergence v_n depends on the index s, π and p. Define the rate exponent

$$\alpha(s, p, \pi) = \min\left\{\frac{s}{2s+1}, \frac{s + 1/p - 1/\pi}{1 + 2s - 2/\pi}\right\}. \tag{2.13}$$

Theorem 2.9. *Work under Assumptions 2.1 and 2.2. Then, for every $c > 0$, the rate $n^{-\alpha(s,p,\pi)/2}$ 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 $\pi \in (0, \infty)$, up to logarithmic corrections.*

Moreover, under Assumption 2.8, the estimator explicitly constructed in Section 3.3 below attains this bound in the sense of (2.5), up to logarithmic corrections.

Remark 2.10. *A (technical) restriction is that we assume $s > 1/\pi$, a condition that guarantees some minimal Hölder smoothness for the path of $t \rightsquigarrow \sigma^2_t$.*

Remark 2.11. *The parametric rate $n^{-1/2}$ (formally obtained when letting $s \to \infty$ in the definition of $\alpha(s, p, \pi)$) 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 [20] and subsequent works, both in parametric, semiparametric and nonparametric estimation, as follows from [7,20,21,26,30,38] among others.*

Our next result shows that this rate is nearly optimal in many cases.

Theorem 2.12. *In the same setting as in Theorem 2.9, assume moreover that $s - 1/\pi > \frac{1+\sqrt{5}}{4}$. Then the rate $n^{-\alpha(s,p,\pi)/2}$ is a lower rate of convergence over $\mathcal{B}^s_{\pi,\infty}(c)$ in L^p in the sense of Definition 2.6.*

Since the upper and lower bound agree up to some (inessential) logarithmic corrections, our result is nearly optimal in the sense of Definitions 2.4 and 2.6.

The proof of the lower bound is an application of a recent result of Reiß [33] 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 following strategy: 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 [30]. Also, this setting may also enable to retrieve the standard minimax framework when σ^2 is deterministic and belongs to a Besov ball $\mathcal{B}^s_{\pi,\infty}(c)$. 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 $\mu(\mathrm{d}\sigma^2)$ with support in $\mathcal{B}^s_{\pi,\infty}(c)$, and is 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 2.12 below.

3. Wavelet estimation and pre-averaging

3.1. *Estimating linear functionals*

We estimate σ^2 via linear functionals of the form

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

With no possible confusion, we denote by $\langle \cdot, \cdot \rangle$ the inner product of $L^2([0,1])$ and by

$$\langle X \rangle_t = (\mathbb{P}\text{-lim})_{\delta \to 0} \sum_{t_i, t_i - t_{i-1} \le \delta} (X_{t_i} - X_{t_{i-1}})^2$$

the quadratic variation of the continuous semimartingale X. Here, the integers $\ell \ge 0$ and k 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, satisfies* $\lambda(t) = -\lambda(2-t)$, *and is not zero identically. To each pre-averaging function* λ *we associate the quantity*

$$\overline{\lambda} := \left(2 \int_0^1 \left(\int_0^s \lambda(u)\,\mathrm{d}u \right)^2 \mathrm{d}s \right)^{1/2}$$

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

For $1 \le m < n$ and a sequence $(Y_{j,n}, j = 0, \dots, n)$, we define the pre-averaging of Y at scale m relative to λ by setting for $i = 2, \dots, m$

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

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 $\lambda(t) = -\lambda(2-t)$, for $t \in (0, 1]$,

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

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 $\langle \sigma^2, h_{\ell k} \rangle$ is to consider the quantity

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

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

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

In order to get a first intuition, note that $(Z_{j,n} - Z_{j-1,n})^2 \approx (\epsilon_{j,n} - \epsilon_{j-1,n})^2$. Further stochastic approximations, detailed in the proof in Section 4.1, show that subtracting $\mathfrak{b}(\lambda, Z.)_{i,m}$ corrects in a natural way for the bias induced by the additive microstructure noise.

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

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

3.2. *The wavelet threshold estimator*

Let (φ, ψ) denote a pair of scaling function and mother wavelet that generate a wavelet basis $(\psi_\nu)_\nu$ satisfying Assumption 2.8. The random function $t \rightsquigarrow \sigma_t^2$ taken path-by-path as an element of $L^2([0,1])$ has for every nonnegative integer ℓ_0 an almost-sure representation

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

with $c_{\ell_0 k} = \langle \sigma^2, \varphi_{\ell_0 k}\rangle = \int_0^1 \varphi_{\ell_0 k}(t)\,\mathrm{d}\langle X\rangle_t$ and $d_{\ell k} = \langle \sigma^2, \psi_{\ell k}\rangle = \int_0^1 \psi_{\ell k}(t)\,\mathrm{d}\langle X\rangle_t$. 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). The choice of ℓ_0 in (3.4) determines the representation of σ^2 as sum of a low resolution approximation based on the scaling function φ and a high-frequency wavelet decomposition, Section 2.2. Following the standard wavelet threshold algorithm (see for instance [15] and in its more condensed form [28]), we approximate Formula (3.4) by

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

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

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

is the standard hard-threshold operator. Thus $t \rightsquigarrow \widehat{\sigma}_n^2(t)$ is specified by the resolution levels ℓ_0, ℓ_1, the threshold τ 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 admits a 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

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

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 and n.

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

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

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

$$\mathbb{P}\bigg[\big|\mathcal{E}_m(h_{\ell k}) - \langle \sigma^2, h_{\ell k}\rangle\big| \geq \kappa\bigg(\frac{p\log m}{m}\bigg)^{1/2}, \sigma^2 \in \mathcal{B}^s_{\pi,\infty}(c)\bigg] \lesssim m^{-p}$$

provided

$$\kappa > 4\bigg(\frac{\rho}{\rho-1}\bigg)^{1/2}\big(\overline{c} + \sqrt{2\overline{c}}\|a\|_{L^\infty}\|\lambda\|_{L^2}\overline{\lambda}^{-1} + \|a\|^2_{L^\infty}\|\lambda\|^2_{L^2}\overline{\lambda}^{-2}\big)$$

and

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

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

Theorem 3.4. *Work under Assumptions 2.1, 2.2 and 2.8. Let $\widehat{\sigma}_n^2$ denote the wavelet estimator defined in (3.5), constructed from (φ, ψ) and a pre-averaging function λ, such that*

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

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

$$\alpha_0 + 1/\pi \le s \le \max\{\alpha_0/(1-2\alpha_0), n_0\},$$

the estimator $\widehat{\sigma}_n^2$ achieves (2.5) over $\mathcal{D} = \mathcal{B}^s_{\pi,\infty}(c)$ with $v_n = n^{-\alpha(s,p,\pi)/2}$ up to logarithmic factors. As a consequence, we have Theorem 2.9.

Proof. Thanks to Theorems 3.2 and 3.3, Theorem 3.4 is now a consequence of the general theory of wavelet threshold estimators, as developed by Kerkyacharian and Picard [28]. To that end, it suffices to obtain appropriate moment bounds and large deviation inequalities for estimators of wavelet coefficients in wavelet bases satisfying Assumption 2.8.

More precisely, by assumption, we have $s - 1/\pi \ge \alpha_0$ and $2^{\ell_0} \sim m^{1-2\alpha_0}$ therefore, the term $m^{-\min\{s-1/\pi,1\}}|h_{\ell k}|_{1,m}$ 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 we have the moment bound

$$\mathbb{E}\bigl[\bigl|\mathcal{E}_m(\varphi_{\ell_0 k}) - \langle\sigma^2, \varphi_{\ell_0 k}\rangle\bigr|^p \mathbb{I}_{\{\sigma^2 \in \mathcal{B}^s_{\pi,\infty}(c)\}}\bigr] \lesssim m^{-p/2} \lesssim n^{-p/4},$$

so that Condition (5.1) of Theorem 5.1 in Kerkyacharian and Picard [28] is satisfied with $c(n) = (\log n/n)^{1/4}$ and $\Lambda(n) = n^{1/2}$ with the notation of [28]. In the same way, by Theorem 3.3, with $h = \psi$, for every $p \ge 1$, we obtain, for a large enough κ the deviation bound

$$\mathbb{P}\biggl[\bigl|\mathcal{E}_m(\psi_{\ell k}) - \langle\sigma^2, \psi_{\ell k}\rangle\bigr| \ge \kappa\biggl(\frac{p\log m}{m}\biggr)^{1/2}, \sigma^2 \in \mathcal{B}^s_{\pi,\infty}(c)\biggr] \lesssim m^{-p} \lesssim n^{-p/2}$$

and therefore Condition (5.2) of Theorem 5.1 in [28] is satisfied with the same specification. This is all that is required to apply the wavelet threshold algorithm: by Corollary 5.2 and Theorem 6.1 of [28] we obtain (2.5) hence Theorem 2.9. □

Remark 3.5. *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 $\alpha_0 + 1/\pi \le s \le \alpha_0/(1-2\alpha_0) < \infty$.*

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

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

$$\mathfrak{R}_n(g) := \left(\sum_{j=1}^{n}\int_{(j-1)/n}^{j/n}\left(\frac{1}{n}\sum_{l=j}^{n} g'\left(\frac{l}{n}\right) - \int_s^1 g'(u)\,\mathrm{d}u\right)^2 \mathrm{d}s\right)^{1/2} \tag{4.1}$$

and

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

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

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

Throughout the remaining part of this paper, we extend pre-averaging functions to the real line by $\lambda(t)=0$ for all $t\in\mathbb{R}\setminus[0,2)$.

Preliminaries: *Some estimates for the latent price X*

Lemma 4.1 (Discretisation 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*

$$\mathbb{E}_{\mathcal{D}_\infty(c)}\left[\left|\left(\frac{1}{n}\sum_{i=1}^{n} g'\left(\frac{i}{n}\right)X_{i/n}\right)^2 - \left(\int_0^1 g(s)\,\mathrm{d}X_s\right)^2\right|^p\right] \lesssim \|g\|_{L^2}^p\mathfrak{R}_n^p(g) + \mathfrak{R}_n^{2p}(g).$$

Proof. By Assumption 2.1, using (2.4) and anticipating that rates of convergence are in power of n, we may (and will) assume that X is a local martingale and take subsequently $b=0$. Next, by Cauchy–Schwarz, we split the error term into a constant times $I\times II + III\times II$, with

$$I := \mathbb{E}_{\mathcal{D}_\infty(c)}\left[\left|\int_0^1 g(s)\,\mathrm{d}X_s\right|^{2p}\right]^{1/2},$$

$$II := \mathbb{E}_{\mathcal{D}_\infty(c)}\left[\left|\frac{1}{n}\sum_{j=1}^{n} g'\left(\frac{j}{n}\right)X_{j/n} + \int_0^1 g(s)\,\mathrm{d}X_s\right|^{2p}\right]^{1/2},$$

$$III := \mathbb{E}_{\mathcal{D}_\infty(c)}\left[\left|\frac{1}{n}\sum_{j=1}^{n} g'\left(\frac{j}{n}\right)X_{j/n}\right|^{2p}\right]^{1/2} \lesssim I + II.$$

Define the stopping time

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

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

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

By Burkholder–Davis–Gundy inequality (later abbreviated by BDG, for a reference see [27], p. 166), we have

$$I \le \mathbb{E}\left[\left|\int_0^{T_c} g(s)\,\mathrm{d}X_s\right|^{2p}\right]^{1/2} \lesssim \mathbb{E}\left[\left|\int_0^{T_c} g^2(s)\sigma_s^2\,\mathrm{d}s\right|^{p}\right]^{1/2} \lesssim \|g\|_{L^2}^p,$$

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

$$\widetilde{g}(s) := \sum_{j=1}^n \left(\frac{1}{n}\sum_{l=j}^n g'\left(\frac{l}{n}\right)\right)\mathbb{I}_{[(j-1)/n,j/n)}(s), \quad s \in [0,1],$$

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

$$\langle S\rangle_1 = \sum_{j=1}^n \int_{(j-1)/n}^{j/n}\left(\frac{1}{n}\sum_{l=j}^n g'\left(\frac{l}{n}\right) - \int_s^1 g'(u)\,\mathrm{d}u\right)^2 \mathbb{I}_{\{s\le T_c\}}\,\mathrm{d}\langle X\rangle_s.$$

By summation by parts, we derive

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

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

$$\Lambda(s) := \int_s^2 \widetilde{\lambda}(u)\,\mathrm{d}u\,\mathbb{I}_{[0,2]}(s) \tag{4.2}$$

and

$$\overline{\Lambda}(s) := \left(\left(\int_0^s \widetilde{\lambda}(u)\,\mathrm{d}u\right)^2 + \left(\int_0^{1-s}\widetilde{\lambda}(u)\,\mathrm{d}u\right)^2\right)^{1/2}\mathbb{I}_{[0,1]}(s). \tag{4.3}$$

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

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

and

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

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

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

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

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

Proof. Recall the definition of $\mathfrak{R}_n$ given in (4.1) and let

$$j_n^*(r) := \max\{j:\ j/n \le r/m\}. \tag{4.4}$$

Since $\widetilde{\lambda}$ is bounded, we have

$$\max_{j/n\in((i-2)/m,i/m]}\sup_{s\in[(j-1)/n,j/n]}\left|\frac{1}{n}\sum_{l=j}^{j_n^*(i)}\widetilde{\lambda}\Big(m\frac{l}{n}-(i-2)\Big)-\int_s^1\widetilde{\lambda}\big(mu-(i-2)\big)\,\mathrm{d}u\right|$$
$$\le \max_{j/n\in((i-2)/m,i/m]}\sup_{s\in[(j-1)/n,j/n]}\left|\int_s^{(j-1)/n}\widetilde{\lambda}\big(mu-(i-2)\big)\,\mathrm{d}u\right|$$
$$+\max_{j/n\in((i-2)/m,i/m]}\sum_{l=j}^{j_n^\star(i)}\left|\frac{1}{n}\widetilde{\lambda}\Big(m\frac{l}{n}-(i-2)\Big)-\int_{(l-1)/n}^{l/n}\widetilde{\lambda}\big(mu-(i-2)\big)\,\mathrm{d}u\right|\lesssim n^{-1},$$

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

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

Lemma 4.3. *Work under Assumption 2.1 and let Λ as in (4.2) with λ as in Definition 3.1. Then, for $m\le n$, every $p\ge 1$ and $c>0$, we have*

$$\mathbb{E}_{\mathcal{D}_\infty(c)}\left[\left|\sum_{i=2}^m g\Big(\frac{i-1}{m}\Big)\Big(\int_0^1\Lambda\big(ms-(i-2)\big)\,\mathrm{d}X_s\Big)^2-\int_0^1\sum_{i=2}^m g\Big(\frac{i-1}{m}\Big)\Lambda^2\big(ms-(i-2)\big)\,\mathrm{d}\langle X\rangle_s\right|^p\right]$$
$$\lesssim\|g\|^p_{L^\infty}\big|\mathrm{supp}(g)\big|^{p/2}m^{-p/2},$$

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

$$H_{t,i}:=g\Big(\frac{i-1}{m}\Big)\Lambda\big(mt-(i-2)\big)\int_{(i-2)/m}^t\Lambda\big(ms-(i-2)\big)\,\mathrm{d}X_s\mathbb{I}_{((i-2)/m,i/m]}(t).\tag{4.5}$$

For a continuous semimartingale M starting at zero, we have the integration by parts formula $M^2=\langle M\rangle+2\int M\,\mathrm{d}M$. Thus,

$$\sum_{i=2}^m g\Big(\frac{i-1}{m}\Big)\Big[\Big(\int_0^1\Lambda\big(ms-(i-2)\big)\,\mathrm{d}X_s\Big)^2-\int_0^1\Lambda^2\big(ms-(i-2)\big)\,\mathrm{d}\langle X\rangle_s\Big]$$
$$=2\sum_{i=2}^m\int_{(i-2)/m}^{i/m}H_{t,i}\,\mathrm{d}X_t.\tag{4.6}$$

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

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

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

$$\begin{aligned}
\mathbb{E}\left[\left(H_i^\star\right)^p\right] &\lesssim \left|g\left(\frac{i-1}{m}\right)\right|^p \mathbb{E}\left[\sup_{t\le 2/m}\left|\int_{(i-2)/m\wedge T_c}^{((i-2)/m+t)\wedge T_c}\Lambda\big(ms-(i-2)\big)\,\mathrm{d}X_s\right|^p\right] \\
&\lesssim \left|g\left(\frac{i-1}{m}\right)\right|^p \mathbb{E}\left[\left(\int_{(i-2)/m\wedge T_c}^{T_c}\Lambda^2\big(ms-(i-2)\big)\sigma_s^2\,\mathrm{d}s\right)^{p/2}\right] \\
&\lesssim \left|g\left(\frac{i-1}{m}\right)\right|^p m^{-p/2}.
\end{aligned} \tag{4.7}$$

The result follows. □

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*

$$\begin{aligned}
&\mathbb{E}_{\mathcal{B}^s_{\pi,\infty}(c)}\left[\left|\sum_{i=2}^m g\left(\frac{i-1}{m}\right)\overline{X}_{i,m}^2 - \int_0^1 g(s)\sigma_s^2\,\mathrm{d}s\right|^p\right] \\
&\quad\lesssim \|g\|_{L^\infty}^p m^{-p/2}\left|\operatorname{supp}(g)\right|^{p/2} + |g|_{1,m}^p m^{-\min\{s-1/\pi,1\}p} + |g|_{\mathrm{var},m}^p m^{-p},
\end{aligned}$$

where

$$|g|_{\mathrm{var},m} := \left|g(0)+g(1)\right| + \sum_{i=1}^m \sup_{s,t\in[(i-1)/m,i/m]}\left|g(t)-g(s)\right|. \tag{4.8}$$

Proof. Recall from Section 3.1 that

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

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

$$\mathbb{E}_{\mathcal{B}^s_{\pi,\infty}(c)}\left[\left|\overline{X}_{i,m}^2 - \left(\int_0^1\Lambda\big(ms-(i-2)\big)\,\mathrm{d}X_s\right)^2\right|^p\right] \lesssim m^{-p/2}n^{-p} \tag{4.9}$$

since

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

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

$$\begin{aligned}
&\mathbb{E}_{\mathcal{B}^s_{\pi,\infty}(c)}\left[\left|\sum_{i=2}^m g\left(\frac{i-1}{m}\right)\overline{X}^2_{i,m}-\sum_{i=2}^m g\left(\frac{i-1}{m}\right)\left(\int_0^1 \Lambda\big(ms-(i-2)\big)\,\mathrm{d}X_s\right)^2\right|^p\right]\\
&\quad\lesssim \big|\operatorname{supp}(g)\big|^{p-1}m^{p-1}\mathbb{E}_{\mathcal{B}^s_{\pi,\infty}(c)}\left[\sum_{i=2}^m\left|g\left(\frac{i-1}{m}\right)\right|^p\left|\overline{X}^2_{i,m}-\left(\int_0^1 \Lambda\big(ms-(i-2)\big)\,\mathrm{d}X_s\right)^2\right|^p\right]\\
&\quad\lesssim \|g\|^p_{L^\infty}m^{p/2}n^{-p}\big|\operatorname{supp}(g)\big|^p,
\end{aligned}\tag{4.10}$$

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

$$\begin{aligned}
&\mathbb{E}_{\mathcal{B}^s_{\pi,\infty}(c)}\left[\left|\sum_{i=2}^m g\left(\frac{i-1}{m}\right)\left(\int_0^1 \Lambda\big(ms-(i-2)\big)\,\mathrm{d}X_s\right)^2-\int_0^1\sum_{i=2}^m g\left(\frac{i-1}{m}\right)\Lambda^2\big(ms-(i-2)\big)\sigma_s^2\,\mathrm{d}s\right|^p\right]\\
&\quad\lesssim \|g\|^p_{L^\infty}m^{-p/2}\big|\operatorname{supp}(g)\big|^{p/2},
\end{aligned}$$

therefore by the triangle inequality

$$\begin{aligned}
&\mathbb{E}_{\mathcal{B}^s_{\pi,\infty}(c)}\left[\left|\sum_{i=2}^m g\left(\frac{i-1}{m}\right)\overline{X}^2_{i,m}-\int_0^1\sum_{i=2}^m g\left(\frac{i-1}{m}\right)\Lambda^2\big(ms-(i-2)\big)\sigma_s^2\,\mathrm{d}s\right|^p\right]\\
&\quad\lesssim \|g\|^p_{L^\infty}m^{-p/2}\big|\operatorname{supp}(g)\big|^{p/2}.
\end{aligned}\tag{4.11}$$

We are going to force the function $\overline{\Lambda}$ in (4.11). To this end, note that

$$\begin{aligned}
&\sum_{i=2}^m g\left(\frac{i-1}{m}\right)\Lambda^2\big(ms-(i-2)\big)\\
&\quad=\sum_{i=1}^m g\left(\frac{i}{m}\right)\big(\Lambda^2\big(ms-(i-2)\big)+\Lambda^2\big(ms-(i-1)\big)\big)\mathbb{I}_{((i-1)/m,i/m]}(s)\\
&\qquad+\sum_{i=1}^m\left(g\left(\frac{i-1}{m}\right)-g\left(\frac{i}{m}\right)\right)\Lambda^2\big(ms-(i-2)\big)\mathbb{I}_{((i-1)/m,i/m]}(s)\\
&\qquad-g(0)\Lambda^2(ms+1)\mathbb{I}_{(0,1/m]}(s)-g(1)\Lambda^2\big(ms-(m-1)\big)\mathbb{I}_{(1-1/m,1]}(s).
\end{aligned}\tag{4.12}$$

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

$$\Lambda^2\big(ms-(i-2)\big)=\left(\int_0^{1-(ms-(i-1))}\widetilde{\lambda}(u)\,\mathrm{d}u\right)^2\quad\text{for } s\in\left(\frac{i-1}{m},\frac{i}{m}\right],$$

$$\Lambda^2\big(ms-(i-1)\big)=\left(\int_0^{ms-(i-1)}\widetilde{\lambda}(u)\,\mathrm{d}u\right)^2\quad\text{for } s\in\left(\frac{i-1}{m},\frac{i}{m}\right].$$

This gives for $s\in(\frac{i-1}{m},\frac{i}{m}]$, and $\overline{\Lambda}$ as in (4.3)

$$\overline{\Lambda}^2\big(ms-(i-1)\big)=\Lambda^2\big(ms-(i-2)\big)+\Lambda^2\big(ms-(i-1)\big),\tag{4.13}$$

and 0 otherwise. From (4.12) it follows that on the event $\sigma^2\in\mathcal{B}^s_{\pi,\infty}(c)$

$$\left|\int_0^1\sum_{i=2}^m g\left(\frac{i-1}{m}\right)\Lambda^2\big(ms-(i-2)\big)\sigma_s^2\,\mathrm{d}s-\int_0^1\sum_{i=1}^m g\left(\frac{i}{m}\right)\overline{\Lambda}^2\big(ms-(i-1)\big)\sigma_s^2\,\mathrm{d}s\right|\lesssim|g|_{\mathrm{var},m}m^{-1}.\tag{4.14}$$

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

$$\begin{aligned}
&\left|\int_0^1 \sum_{i=2}^m g\left(\frac{i-1}{m}\right)\left(\overline{\Lambda}^2\big(ms-(i-1)\big) - \mathbb{I}_{((i-1)/m,i/m]}(s)\right)\sigma_s^2\,\mathrm{d}s\right| \\
&\quad\le \left|\int_0^1 \sum_{i=2}^m g\left(\frac{i-1}{m}\right)\overline{\Lambda}^2\big(ms-(i-1)\big)\big(\sigma_s^2 - \sigma^2_{(i-1)/m}\big)\,\mathrm{d}s\right| \\
&\qquad + \left|\int_0^1 \sum_{i=2}^m g\left(\frac{i-1}{m}\right)\mathbb{I}_{((i-1)/m,i/m]}(s)\big(\sigma_s^2 - \sigma^2_{(i-1)/m}\big)\,\mathrm{d}s\right| \\
&\quad\lesssim m^{-\min\{s-1/\pi,1\}}|g|_{1,m},
\end{aligned} \tag{4.15}$$

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 $\sigma^2 \in \mathcal{B}^s_{\pi,\infty}(c)$

$$\left|\int_0^1 \sum_{i=2}^m g\left(\frac{i}{m}\right)\mathbb{I}_{((i-1)/m,i/m]}(s)\sigma_s^2\,\mathrm{d}s - \int_0^1 g(s)\sigma_s^2\,\mathrm{d}s\right| \lesssim m^{-1}|g|_{\mathrm{var},m}, \tag{4.16}$$

the conclusion follows by combining (4.11), (4.14), (4.15) and (4.16). □

Preliminaries: *Some estimates for the microstructure noise* ϵ
We need some notation. Remember from (1.1) that we observe

$$Z_{j,n} = X_{j/n} + a(j/n, X_{j/n})\eta_{j,n}, \quad j = 0, \dots, n,$$

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

$$\overline{\epsilon}_{i,m} := \overline{\epsilon}_{i,m}(\lambda) := \frac{m}{n} \sum_{j/n \in ((i-2)/m, i/m]} \widetilde{\lambda}\left(m\frac{j}{n} - (i-2)\right)\epsilon_{j,n}, \quad i = 2, \dots, m. \tag{4.17}$$

Moreover, we will make several times use of Rosenthal's inequality for martingales (see [22], p. 23). It states that for an $(\mathcal{F}_k)_k$-martingale $(M_k)_k$ and for $p \ge 0$, there exists a universal constant C_p only depending on p, such that

$$\begin{aligned}
&\mathbb{E}\Big[\max_{k=1,\dots,n} |M_k|^p\Big] \\
&\quad\le C_p\left(\mathbb{E}\left[\left(\sum_{k=0}^{n-1}\mathbb{E}\big[(M_{k+1}-M_k)^2|\mathcal{F}_k\big]\right)^{p/2}\right] + \mathbb{E}\Big[\max_{k\le n}|M_k - M_{k-1}|^p\Big]\right).
\end{aligned}$$

For our proofs it will be sufficient to bound the maximum in the second term on the r.h.s. by the sum $\sum_{k=1}^n$.

Lemma 4.5. *Work under Assumptions* 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* $p \ge 1$, *we have*

$$\begin{aligned}
&\mathbb{E}\left[\left|\sum_{i=1}^m g\left(\frac{i-1}{m}\right)\big(\overline{\epsilon}^2_{i,m}(\lambda) - \mathbb{E}\big[\overline{\epsilon}^2_{i,m}(\lambda)|\mathcal{G}\big]\big)\right|^p\right] \\
&\quad\lesssim |g|^p_{2,m} m^{3p/2} n^{-p} + |g|^p_{p,m} m^{p+1} n^{-p}.
\end{aligned}$$

Proof. In the following, we will decompose the sum in the previous inequality in an even and odd part. This allows us to treat sums of preaveraged values computed over disjoint intervals. In a first step, let us introduce the filtrations

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

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

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

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. Intuitively, $\overline{\epsilon}_{i,m} = \mathrm{O}_P(m^{1/2}/n^{1/2})$ by (4.17). More precisely using Rosenthal's inequality, we have, for every $p \ge 1$

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

using $\|a\|_{L^\infty} \lesssim 1$. It follows that

$$\mathbb{E}\big[|U_i|^p\big] \lesssim \left|g\left(\frac{i-1}{m}\right)\right|^p m^p n^{-p}. \tag{4.18}$$

Analogous computations show that

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

Therefore, applying Rosenthal's inequality again, we obtain

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

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

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

$$\mathbb{E}_{\mathcal{D}_\infty(c)}\left[\left|\sum_{i=1}^m g\left(\frac{i-1}{m}\right)\overline{X}_{i,m}(\lambda)\overline{\epsilon}_{i,m}(\lambda)\right|^p\right]$$
$$\lesssim |g|_{p,m}^p\big(n^{-p/2}m + m^{3p/2+1}n^{-3p/2}\big) + |g|_{2,m}^p\big(m^{p/2}n^{-p/2} + m^{2p}n^{-3p/2}\big).$$

Proof. By Assumption 2.1 and the same localisation procedure as in the proof of Lemma 4.1, up to losing 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 $\mathbb{E}[\cdot]$ instead of $\mathbb{E}_{\mathcal{D}_\infty(c)}[\cdot]$.

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

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

and proceed for S^{odd} analogously. By Rosenthal's inequality for martingales and Cauchy–Schwarz,

$$\mathbb{E}\big[\big|S^{\mathrm{even}}_{\lfloor m/2\rfloor}\big|^p\big] \lesssim m^{p/2}n^{-p/2}\mathbb{E}\left[\left|\sum_{i=1}^{\lfloor m/2\rfloor} g^2\left(\frac{2i-1}{m}\right)\mathbb{E}\big[\overline{X}^2_{2i,m}(\lambda)|\mathcal{F}^{\mathrm{even}}_{i-1}\big]\right|^{p/2}\right] + \sum_{i=1}^{\lfloor m/2\rfloor}\left|g\left(\frac{2i-1}{m}\right)\right|^p\big(\mathbb{E}\big[\big|\overline{X}_{2i,m}(\lambda)\big|^{2p}\big]\big)^{1/2}\big(\mathbb{E}\big[\big|\overline{\epsilon}_{2i,m}(\lambda)\big|^{2p}\big]\big)^{1/2}.$$

Note that,

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

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

$$\left|\sum_{j/n\in((i-2)/m,i/m]}\widetilde{\lambda}\left(m\frac{j}{n}-(i-2)\right)\right| \lesssim 1. \tag{4.19}$$

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

$$\|\widetilde{\lambda}\|^{2p}_{L^\infty}\mathbb{E}\Big[\sup_{s\le 2/m}|X_{(i-2)/m+s}-X_{(i-2)/m}|^{2p}\Big] + m^{2p}n^{-2p}\mathbb{E}\big[|X_{(i-2)/m}|^{2p}\big] \tag{4.20}$$

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 obtain

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

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

Preliminaries: *Some estimates for the bias correction* $\mathfrak{b}$
We need some notation. Recall the bias correction defined in (3.2)

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

We plan to use the following decomposition

$$\mathfrak{b}(\lambda, Z.)_{i,m} = \mathfrak{b}(\lambda, X.)_{i,m} + \mathfrak{b}(\lambda, \epsilon.)_{i,m} + 2\mathfrak{c}(\lambda, X., \epsilon.)_{i,m},$$

where

$$\mathfrak{c}(\lambda, X., \epsilon.)_{i,m} := \frac{m^2}{2n^2}\sum_{j/n\in((i-2)/m,i/m]}\widetilde{\lambda}^2\left(m\frac{j}{n}-(i-2)\right)(X_{j/n}-X_{(j-1)/n})(\epsilon_{j,n}-\epsilon_{j-1,n}).$$

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

$$\mathbb{E}\left[\left|\sum_{i=2}^{m} g\left(\frac{i-1}{m}\right)\left(\mathfrak{b}(\lambda,\epsilon.)_{i,m} - \frac{m^2}{n^2}\sum_{j/n\in((i-2)/m,i/m]}\widetilde{\lambda}^2\left(m\frac{j}{n}-(i-2)\right)a^2_{j/n}\right)\right|^p\right]$$
$$\lesssim |g|^p_{1,m} m^{3p} n^{-2p} + |g|^p_{2,m} m^{2p} n^{-3p/2} + |g|^p_{p,m} m^{2p} n^{-2p+1}.$$

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

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

where

$$I := \mathbb{E}\left[\left|\sum_{i=2}^{m} g\left(\frac{i-1}{m}\right)\sum_j \widetilde{\lambda}^2\left(m\frac{j}{n}-(i-2)\right)a^2_{j/n}\left(\eta^2_{j,n}-1\right)\right|^p\right],$$

$$II := \mathbb{E}\left[\left|\sum_{i=2}^{m} g\left(\frac{i-1}{m}\right)\sum_j \widetilde{\lambda}^2\left(m\frac{j}{n}-(i-2)\right)a^2_{(j-1)/n}\left(\eta^2_{j-1,n}-1\right)\right|^p\right],$$

$$III := \mathbb{E}\left[\left|\sum_{i=2}^{m} g\left(\frac{i-1}{m}\right)\sum_j \widetilde{\lambda}^2\left(m\frac{j}{n}-(i-2)\right)\left(a^2_{j/n}-a^2_{(j-1)/n}\right)\right|^p\right],$$

$$IV := \mathbb{E}\left[\left|\sum_{i=2}^{m} g\left(\frac{i-1}{m}\right)\sum_j \widetilde{\lambda}^2\left(m\frac{j}{n}-(i-2)\right)\epsilon_{j-1,n}\epsilon_{j,n}\right|^p\right],$$

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

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

$$I \lesssim \sum_{j=1}^{n}\left(\sum_{i=2}^{m}\left|g\left(\frac{i-1}{m}\right)\right|^p \mathbb{I}_{\{j/n\in((i-2)/m,i/m]\}}\right)\mathbb{E}\left[\left|\left(\eta^2_{j,n}-1\right)\right|^p\right]$$
$$+\left|\sum_{j=1}^{n}\sum_{i=2}^{m} g^2\left(\frac{i-1}{m}\right)\mathbb{I}_{\{j/n\in((i-2)/m,i/m]\}}\mathbb{E}\left[\left(\eta^2_{j,n}-1\right)^2|\mathcal{F}_{j-1}\right]\right|^{p/2}$$
$$\lesssim |g|^p_{p,m} n + |g|^p_{2,m} n^{p/2},$$

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

- *The term III*. Recall the definition of $j^*_n(r)$ given in (4.4). Summing by parts, we have

$$\sum_{j/n\in((i-2)/m,i/m]}\widetilde{\lambda}^2\left(m\frac{j}{n}-(i-2)\right)\left(a^2_{j/n}-a^2_{(j-1)/n}\right)$$
$$= -\sum_{j/n\in((i-2)/m,i/m]} a^2_{(j-1)/n}\left(\widetilde{\lambda}^2\left(m\frac{j}{n}-(i-2)\right)-\widetilde{\lambda}^2\left(m\frac{j-1}{n}-(i-2)\right)\right)$$
$$+ a^2_{j^*_n(i)/n}\widetilde{\lambda}^2\left(m\frac{j^*_n(i)}{n}-(i-2)\right) - a^2_{j^*_n(i-2)/n}\widetilde{\lambda}^2\left(m\frac{j^*_n(i-2)}{n}-(i-2)\right).$$

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

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

• *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

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

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

$$\mathbb{E}_{\mathcal{D}_\infty(c)}\left[\left|\sum_{i=2}^{m} g\left(\frac{i-1}{m}\right)\mathfrak{b}(\lambda, X.)_{i,m}\right|^p\right] \lesssim |g|_{1,m}^p m^p n^{-p}.$$

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 $|\sigma_s^2| \le c$ almost surely, working subsequently with $\mathbb{E}[\cdot]$ instead of $\mathbb{E}_{\mathcal{D}_\infty}(c)[\cdot]$. We readily obtain

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

where we bound $|X_{j/n} - X_{(i-2)/m}|$ by the supremum over $|X_{s+(i-2)/m} - X_{(i-2)/m}|$, $s \le 2/m$ and argue as in (4.20). $\square$

Let M be a continuous, locally square integrable $\mathcal{F}$-martingale and H some progressively measurable process. Then, for $0 \le s < t \le 1$

$$\mathbb{E}\left[\left(\int_s^t H_u \,\mathrm{d}M_u\right)^2 \middle| \mathcal{F}_s\right] = \mathbb{E}\left[\int_s^t H_u^2 \,\mathrm{d}\langle M\rangle_u \middle| \mathcal{F}_s\right]$$

provided that $\mathbb{E}[\int_0^1 H_u^2 \,\mathrm{d}\langle M\rangle_u] < \infty$. This fact will be referred to in the sequel as conditional Itô-isometry (cf. [27], Section 3.2 B).

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

$$\begin{aligned}&\mathbb{E}_{\mathcal{D}_\infty(c)}\left[\left|\sum_{i=2}^{m} g\left(\frac{i-1}{m}\right)\mathfrak{c}(\lambda, X., \epsilon.)_{i,m}\right|^p\right] \\ &\quad\lesssim \left[|g|_{2,m}^p + |g|_{p,m}^p\left(n^{-p/2+1} + m^{-p/2+1}\right)\right]m^{2p}n^{-2p}.\end{aligned}$$

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 $\mathbb{E}[\cdot]$ instead of $\mathbb{E}_{\mathcal{D}_\infty}(c)[\cdot]$. It suffices then to bound

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

Recall that $j_n^*(r) = \max\{j\colon\ j/n \le r/m\}$ and let us introduce the filtrations

$$\mathcal{G}_r^{\text{even}} := \sigma(\eta_{j,n}\colon\ j/n \le 2r/m) \otimes \sigma\left(X_s\colon\ s \le j_n^*(2r)/n\right),$$

$$\mathcal{G}_r^{\text{odd}} := \sigma\left(\eta_{j,n}\colon\ j/n \le (2r+1)/m\right) \otimes \sigma\left(X_s\colon\ s \le j_n^*(2r+1)/n\right).$$

The process

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

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

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

and on the other hand by conditional Itô-isometry

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

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

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

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. □

Completion of proof of Theorem 3.2
Since

$$\mathcal{E}_m(h_{\ell k}) = \sum_{i=2}^{m} h_{\ell k}\left(\frac{i-1}{m}\right)\left[\overline{Z}_{i,m}^2 - b(\lambda, Z.)_{i,m}\right]$$

we plan to use the following decomposition

$$\mathcal{E}_m(h_{\ell k}) - \left\langle \sigma^2, h_{\ell k}\right\rangle_{L^2} = I + II + III, \tag{4.21}$$

with

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

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

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

Note that $\|h_{\ell k}\|_{L^\infty} \le 2^{\ell/2}\|h\|_{L^\infty}$ and $|\,\mathrm{supp}(h_{\ell k})|^{p/2} \lesssim 2^{-\ell p/2}$. By assumption, h has a piecewise Lipschitz derivative. With (4.8), we conclude

$$|h_{\ell k}|_{\mathrm{var},m} \lesssim m^{1/2}. \tag{4.22}$$

Thus, the term I has the right order.

- *The term II*. Applying successively Lemmas 4.5, 4.7, 4.8 and 4.9, we derive using $m \le n^{1/2}$

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

Since for $1 \le p \le 2$, by Jensen's inequality $\mathbb{E}_{\mathcal{B}^s_{\pi,\infty}}[|II|^p] \le \mathbb{E}_{\mathcal{B}^s_{\pi,\infty}}[|II|^2]^{p/2}$ and for $p \ge 2$, $|h_{\ell k}|^p_{p,m} m^{p+1} n^{-p} \lesssim 2^{l(p/2-1)} m^{p+1} n^{-p} \le m^{3p/2} n^{-p}$, this term also has the right order.

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

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

which also has the right order by the same argument as above. The proof of Theorem 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 $\langle M\rangle_k = \sum_{i=1}^k \mathbb{E}[(\Delta M_i)^2|\mathcal{F}_{i-1}]$ its predictable compensator. We will heavily rely on the following result of Bercu and Touati [6].

Theorem 4.10 (Bercu and Touati [6]). *Let (M_k) be a locally square integrable martingale. Then, for all $x, y > 0$, we have*

$$\mathbb{P}\big[|M_k| \ge x, [M]_k + \langle M\rangle_k \le y\big] \le 2\exp\left(-\frac{x^2}{2y}\right).$$

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 $p \ge 1$ there is some deterministic sequence $(C_j)_j$ (with $j = j(m)$) and $\delta > 0$ such that $\mathbb{P}[\langle M\rangle_j > C_j(1+\delta)] \lesssim m^{-p}$. If further for every $\kappa \ge 2$*

$$\max_{i=1,\dots,j} \mathbb{E}\big[|\Delta M_i|^\kappa\big] \lesssim 1, \tag{4.23}$$

then,

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

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

Proof. We have by Theorem 4.10

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

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

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

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 $\kappa \geq 2$

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

where we used Hölder's inequality

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

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

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

Lemma 4.12. *Work under the assumptions of Theorem 3.3 and suppose that X has no drift, i.e. $b = 0$. If $\overline{c} = \overline{c}(s,\pi,c)$ is such that $\mathcal{B}^s_{\pi,\infty}(c) \subset \mathcal{D}_\infty(\overline{c})$ then, we have for every fixed $\delta > 0$*

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

provided

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

Proof. Recall that $\Lambda(s) = \int_s^2 \widetilde{\lambda}(u)\,\mathrm{d}u$ and let $H_{t,i}$ be defined as in (4.5), where g is replaced by $h_{\ell k}$. Using the integration by parts formula (4.6) we bound the probability by $I + II + III$, with

$$\begin{aligned}I &:= \mathbb{P}\left[\left|\sum_{i=2}^{m} h_{\ell k}\left(\frac{i-1}{m}\right)\left(\overline{X}^2_{i,m}(\lambda) - \left(\int_0^1 \Lambda\big(ms-(i-2)\big)\,\mathrm{d}X_s\right)^2\right)\right| > \overline{c}\delta\sqrt{\frac{p\log m}{m}} \text{ and } \sigma^2 \in \mathcal{B}^s_{\pi,\infty}(c)\right],\\II &:= \mathbb{P}\left[\left|\sum_{i=2}^{m}\int_0^1 H_{t,i}\,\mathrm{d}X_t\right| > 2\overline{c}\left(1+\frac{\delta}{2}\right)\sqrt{\frac{p\log m}{m}} \text{ and } \sigma^2 \in \mathcal{D}_\infty(\overline{c})\right],\\III &:= \mathbb{P}\left[\left|\sum_{i=2}^{m} h_{\ell k}\left(\frac{i-1}{m}\right)\left(\int_0^1 \Lambda^2\big(ms-(i-2)\big)\sigma_s^2\,\mathrm{d}s - \langle\sigma^2, h_{\ell k}\rangle_{L^2}\right)\right| > \overline{c}\delta\sqrt{\frac{p\log m}{m}} \text{ and } \sigma^2 \in \mathcal{B}^s_{\pi,\infty}(c)\right].\end{aligned}$$

Note that $\mathbb{P}[X > t \text{ and } B] = \mathbb{E}[\mathbb{I}_{\{X>t\}\cap B}] \leq t^{-p}\mathbb{E}[X^p\mathbb{I}_B]$, for $p \geq 0$. Using $m \leq n^{1/2}$ and (4.10) we find that I can be bounded by any polynomial order of $1/m$.

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

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

Since h has support $[0, 1]$, $h_{\ell k}(\frac{2i-1}{m}) \neq 0$ can happen only if

$$\frac{1}{2}\left(k2^{-\ell}m+1\right) \le i \le \frac{1}{2}\left((k+1)2^{-\ell}m+1\right). \tag{4.24}$$

We will treat the term II_{even} only, since similar arguments apply for II_{odd}. The process $M_r := 2^{-\ell/2}m \times \sum_{i=1}^{r}\int_0^{T_{\overline{c}}} H_{t,2i}\, \mathrm{d}X_t$ is a martingale with respect to the filtration $\mathcal{F}_r = \sigma(X_s\colon s \le 2r/m)$ starting at $M_{\lfloor (k2^{-\ell}m+1)/2\rfloor} = 0$. Recall that $H_{t,2i}$ vanishes outside $[2(i-1)/m, 2i/m]$ and $\mathbb{I}_{\{T_{\overline{c}} \le (2i-2)/m\}}$ is $\mathcal{F}_{i-1}$ measurable. Moreover, uniformly in k, ℓ, we obtain

$$\frac{2}{m}\sum_{i=1}^{\lfloor m/2\rfloor} h_{\ell k}^2\left(\frac{2i-1}{m}\right) = \|h_{\ell,k}\|_2^2 + \mathrm{O}\left(2^{\ell}/m\right) = 1 + \mathrm{O}\left(m^{-q}\right). \tag{4.25}$$

Therefore, Lemma 4.2 and conditional Itô-isometry yield

$$\begin{aligned}
\langle M\rangle_{\lfloor (1/2)((k+1)2^{-\ell}m+1)\rfloor} &\le 2^{-\ell}m^2\overline{c}\sum_{i=1}^{\lfloor m/2\rfloor}\int_0^1 \mathbb{E}\left[H_{s\wedge T_{\overline{c}},2i}^2 | \mathcal{F}_{i-1}\right]\mathrm{d}s \\
&\le 2^{-\ell-1}\overline{c}^2\sum_{i=1}^{\lfloor m/2\rfloor} h_{\ell k}^2\left(\frac{2i-1}{m}\right) \\
&\le 2^{-\ell}m\frac{1}{4}\overline{c}^2\left(1+\frac{\delta}{2}\right),
\end{aligned}$$

where the last inequality follows for all $m \ge m_0(\delta)$ and $m_0(\delta)$ is fixed and independent of ℓ, k. Furthermore, by BDG and (4.7), we bound

$$\begin{aligned}
\mathbb{E}\left[|\Delta M_i|^{\kappa}\right] &\lesssim 2^{-\ell\kappa/2}m^{\kappa}\mathbb{E}\left[\left|\int_0^1 H_{t,2i}\mathbb{I}_{[0,T_{\overline{c}}]}(t)\,\mathrm{d}X_t\right|^{\kappa}\right] \\
&\lesssim 2^{-\ell\kappa/2}m^{\kappa}\mathbb{E}\left[\left|\int_0^1 H_{t\wedge T_{\overline{c}},2i}^2\,\mathrm{d}t\right|^{\kappa/2}\right] \\
&\lesssim 2^{-\ell\kappa/2}m^{\kappa/2}\mathbb{E}\left[\sup_{t\le 2/m}|H_{(t+2(i-1)/m)\wedge T_{\overline{c}},2i}|^{\kappa}\right] \\
&\lesssim 2^{-\ell\kappa/2}\left|h_{\ell k}\left(\frac{i-1}{m}\right)\right|^{\kappa} \lesssim 1
\end{aligned}$$

uniformly over i. Since the number of integers i for which (4.24) holds is of order $m2^{-\ell}$, we may apply Lemma 4.11 for $j \sim m2^{-\ell}$, $C_j = 2^{-\ell}m\frac{1}{4}\overline{c}^2$ and obtain $II_{\mathrm{even}} \lesssim m^{-p}$.

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

In order to bound III it follows from $m^{-(s-1/\pi)}|h_{\ell k}|_{1,m} \lesssim m^{-1/2}$, (4.14), (4.15), (4.16) and (4.22), that for sufficiently large m on $\sigma^2 \in \mathcal{B}_{\pi,\infty}^{s}(c)$

$$\left|\sum_{i=2}^{m} h_{\ell k}\left(\frac{i-1}{m}\right)\left(\int_0^1 \Lambda^2\left(ms-(i-2)\right)\sigma_s^2\,\mathrm{d}s - \left\langle\sigma^2, h_{\ell k}\right\rangle_{L^2}\right)\right| \le \overline{c}\delta\sqrt{\frac{p\log m}{m}}.$$

This yields the conclusion. □

Lemma 4.13. *Work under the assumptions of Theorem 3.3 and suppose that X has no drift, i.e. $b = 0$. Then, we have for every fixed $\delta > 0$*

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

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

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

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

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

By the assumed piecewise Lipschitz continuity of λ it follows

$$\frac{m}{n}\sum_{j/n\in((2i-2)/m,2i/m]}\widetilde{\lambda}^2\left(m\frac{j}{n}-(2i-2)\right) = \|\widetilde{\lambda}\|^2_{L^2} + \mathrm{O}\left(\frac{m}{n}\right), \tag{4.26}$$

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

$$\begin{aligned}
U_1 &:= \frac{m}{n}\sum_{j/n\in((2i-2)/m,2i/m]}\left(\sum_{l=j}^{n}\widetilde{\lambda}\left(m\frac{l}{n}-(2i-2)\right)\right)(X_{j/n\wedge T_{\overline{c}}} - X_{(j-1)/n\wedge T_{\overline{c}}\wedge(2i-2)/m}), \\
U_2 &:= X_{(2i-2)/m\wedge T_{\overline{c}}}\frac{m}{n}\sum_{j/n\in((2i-2)/m,2i/m]}\widetilde{\lambda}\left(m\frac{j}{n}-(2i-2)\right).
\end{aligned}$$

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 Itô-isometry

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

where W is a standard Brownian motion and $\delta_{j,j'}$ denotes the Kronecker delta. Recall the definition of $j_n^*(r)$ given in (4.4) and define $c_j := \sum_{l=j}^n \widetilde{\lambda}(m\frac{l}{n} - (2i-2))$. We can bound

$$\begin{aligned}
&\mathbb{E}\big[U_1^2 | \mathcal{F}_{i-1}^{\mathrm{even}}\big] \\
&\quad \le \overline{c}\frac{m^2}{n^2}\left(c_{1+j_n^*(2i-2)}^2 \frac{j_n^*(2i-2)}{n} + \sum_{j/n \in ((2i-2)/m, 2i/m]} \frac{c_j^2}{n}\right) \\
&\quad = \overline{c}\frac{m^2}{n^2}\mathbb{E}\left[\left(c_{1+j_n^*(2i-2)} W_{j_n^*(2i-2)/n} + \sum_{j/n \in ((2i-2)/m, 2i/m]} c_j (W_{j/n} - W_{(j-1)/n})\right)^2\right] \\
&\quad = \overline{c}\mathbb{E}\left[\left(\frac{m}{n}\sum_{j/n \in ((2i-2)/m, 2i/m]} \widetilde{\lambda}\left(m\frac{j}{n} - (2i-2)\right) W_{j/n}\right)^2\right].
\end{aligned}$$

Setting $X = W$ in (4.9) and Lemma 4.2 yield further

$$\begin{aligned}
\mathbb{E}\big[U_1^2 | \mathcal{F}_{i-1}^{\mathrm{even}}\big] &\le \overline{c}\mathbb{E}\left[\int_0^1 \Lambda^2\big(ms - (2i-2)\big)\,\mathrm{d}s\right] + \mathrm{O}\big(m^{-1/2}n^{-1}\big) \\
&= \overline{c}m^{-1} + \mathrm{O}\big(m^{-1/2}n^{-1}\big)
\end{aligned}$$

uniformly over i. By using (4.19) we infer that there exists a constant c_U such that $U_2^2 \le c_U \frac{m^2}{n^2} \sup_{s \le T_{\overline{c}}} X_s^2$. Choose $\delta' \le \min(1, \frac{\delta}{8}\min(\|\widetilde{\lambda}\|_{L^2}^2, 1))$. We find by Chebycheff inequality that $\mathbb{P}[\frac{m}{\overline{c}} U_2^2 > \delta'] \lesssim m^{-p}$. With (4.25), we obtain further for the predictable quadratic variation, sufficiently large m and probability larger than $1 - \text{const.} \times m^{-p}$

$$\begin{aligned}
&\langle M \rangle_{\lfloor (1/2)((k+1)2^{-\ell}m+1)\rfloor} \\
&\quad \le 2^{-\ell-1} m \|a\|_{L^\infty}^2 \overline{c}\big(1 + \mathrm{O}\big(m^{-q}\big)\big)\left(\|\widetilde{\lambda}\|_{L^2}^2 + \mathrm{O}\left(\frac{m}{n}\right)\right)\left(1 + \frac{m}{\overline{c}}U_2^2\right) \\
&\quad \le 2^{-\ell-1} m \|a\|_{L^\infty}^2 \overline{c}\big(1+\delta'\big)\big(\|\widetilde{\lambda}\|_{L^2}^2 + \delta'\big)\big(1+\delta'\big) \\
&\quad \le 2^{-\ell-1} m \|a\|_{L^\infty}^2 \overline{c}\|\widetilde{\lambda}\|_{L^2}^2 (1+\delta)
\end{aligned}$$

or to state it differently

$$\mathbb{P}\big[\langle M \rangle_{\lfloor (1/2)((k+1)2^{-\ell}m+1)\rfloor} > 2^{-\ell-1} m \|a\|_{L^\infty}^2 \overline{c}\|\widetilde{\lambda}\|_{L^2}^2 (1+\delta)\big] \lesssim m^{-p}.$$

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

$$\max_i \mathbb{E}\big[|\Delta M_i|^\kappa\big] \lesssim 2^{-\ell\kappa/2} n^{\kappa/2}\left|h_{\ell k}\left(\frac{i-1}{m}\right)\right|^\kappa \mathbb{E}^{1/2}\big[\big|\overline{X}_{i,m,T_{\overline{c}}}(\lambda)\big|^{2\kappa}\big]\mathbb{E}^{1/2}\big[|\overline{\epsilon}_{i,m}|^{2\kappa}\big] \lesssim 1.$$

Hence the assumptions of Lemma 4.11 are satisfied with $j \sim m2^{-\ell}$ and $C_j = 2^{-\ell-1} m \|a\|_{L^\infty}^2 \overline{c}\|\widetilde{\lambda}\|_{L^2}^2$ and the conclusion follows. □

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

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

Proof. We show that

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

and argue similar for the sum over i odd. Let $\mathcal{F}_r^{\text{even}}, U_i$ and the martingale S_r^{even} 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 $\frac{1}{2}(k2^{-\ell}m+1) \le i \le \frac{1}{2}((k+1)2^{-\ell}m+1)$. In the following we will consider the martingale $M_r := \frac{n}{m}2^{-\ell/2}S_r^{\text{even}}$ started at $M_{\lfloor (k2^{-\ell}m+1)/2\rfloor} = 0$. We obtain

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

Elementary calculations and (4.26) show further that we may find a deterministic bound, i.e. uniformly in i

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

From this and (4.25) we obtain for sufficiently large m,

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

By (4.18), 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 (4.21) and suppose that X has no drift.

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

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

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

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

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

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

If the drift is nonzero, we can argue by a change of measure as in Lemma 4.1 and obtain with Assumption 2.1, $\mathbb{E}_{\sigma,b}[\mathbb{I}_{B_n}] \lesssim \mathbb{E}_{\sigma,0}[\mathbb{I}_{B_n}]^{(\rho-1)/\rho}$. The proof of Theorem 3.3 is complete.

4.3. *Proof of Theorem* 2.12

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 $\mu(\mathrm{d}\sigma)$ on $(C, \mathcal{C})$, and we construct the measure $\mathbb{P}_\sigma(\mathrm{d}\omega)$ 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

$$X_t = X_0 + \int_0^t \sigma_s \,\mathrm{d}W_s,$$

where W is a standard Wiener process. We then set

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

This space is rich enough to contain our model: indeed, by construction, any $\mu(\mathrm{d}\sigma)$ 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. Consider next the statistical experiment

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

where $\mathcal{D} \subset C$ and $\mathbb{P}_\sigma^n$ is the law of the data $(Z_{j,n})$, conditional on σ. The probability $\mu(\mathrm{d}\sigma)$ can be interpreted as a prior distribution for the "true" parameter σ. Let us now introduce 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 $\mathcal{D}$, but living on a possibly different space Ω''. We denote by $\mathbb{Q}_\sigma^n$ the law of Y_n.

Completion of proof

Let $\mathcal{D} = \mathcal{B}^s_{\pi,\infty}(c)$ denote a Besov ball such that $s - 1/\pi > 0$. Then $\mathcal{D} \subset C$. Assume further that μ is such that $\mu[\mathcal{D}] = 1$. Then Condition (2.6) is satisfied. Moreover, for any estimator $\widehat{\sigma}_n$ and any $c' > 0$, we have, by Markov inequality

$$\begin{aligned} & n^{\alpha(s,p,\pi)/2} \widetilde{\mathbb{E}}\bigl[\bigl\|\widehat{\sigma}_n^2 - \sigma^2\bigr\|_{L^p([0,1])} \mathbb{I}_{\{\sigma^2 \in \mathcal{B}^s_{\pi,\infty}(c)\}}\bigr] \\ & \quad \geq c' \int_C \mu(\mathrm{d}\sigma) \mathbb{P}_\sigma^n\bigl[n^{\alpha(s,p,\pi)/2} \bigl\|\widehat{\sigma}_n^2 - \sigma^2\bigr\|_{L^p([0,1])} \geq c'\bigr] \end{aligned} \tag{4.27}$$

since $\mu[\mathcal{D}] = 1$. By the result of Reiß [33], since $s - 1/\pi > (1 + \sqrt{5})/4$, we have that $\mathcal{E}_n$ and $\mathcal{E}_n'$ are asymptotically equivalent. This means that we can approximate $\mathbb{P}_\sigma^n$ by $\mathbb{Q}_\sigma^n$ in variational norm, uniformly in σ, up to randomisation via a Markov kernel K that does not depend on σ. More precisely, for any $\varepsilon > 0$, we have

$$\bigl|\mathbb{P}_\sigma^n\bigl[n^{\alpha(s,p,\pi)/2} \bigl\|\widehat{\sigma}_n^2 - \sigma^2\bigr\|_{L^p([0,1])} \geq c'\bigr] - K\mathbb{Q}_\sigma^n\bigl[n^{\alpha(s,p,\pi)/2} \bigl\|\widehat{\sigma}_n^2 - \sigma^2\bigr\|_{L^p([0,1])} \geq c'\bigr]\bigr| \leq \varepsilon \tag{4.28}$$

as soon as n is large enough, and where we use the notation

$$K\mathbb{Q}_n(\mathrm{d}x) = \int_{\Omega''} K(y, \mathrm{d}x) \mathbb{Q}_n(\mathrm{d}y), \quad x \in \mathcal{C} \times \Omega', y \in \Omega''.$$

Now, there exist $c' > 0$ and $\delta' > 0$ such that for any estimator F in $\mathcal{E}_n'$, by picking $\mu(\mathrm{d}\sigma)$ as the least favourable prior in order to obtain lower bounds over Besov classes, we have

$$\int_C \mu(\mathrm{d}\sigma) \mathbb{Q}_\sigma^n\bigl[n^{\alpha(s,p,\pi)/2} \bigl\|F - \sigma^2\bigr\|_{L^p([0,1])} \geq c'\bigr] \geq \delta' > 0 \tag{4.29}$$

for large enough n. This follows from classical analysis of the white Gaussian noise model, see for instance [23] in the framework of Besov spaces. Let us extend further (4.29) to the class of *randomised decisions*, that is estimators of the form $F(\xi, \cdot)$, where ξ is an auxiliary random variable, living on an auxiliary probability space with law $\nu(\mathrm{d}\xi)$. Conditional on ξ, an arbitrary randomised decision $F(\xi, \cdot)$, can be viewed as an estimator, therefore, by (4.29), we also have

$$\int_C \mu(\mathrm{d}\sigma)\mathbb{Q}_\sigma^n\big[n^{\alpha(s,p,\pi)/2}\big\|F(\xi,\cdot)-\sigma^2\big\|_{L^p([0,1])} \geq c'\big] \geq \delta' \quad \nu(\mathrm{d}\xi)\text{-a.s.}$$

for large enough n. Integrating an applying Fubini, we derive

$$\int_C \mu(\mathrm{d}\sigma)\int \nu(\mathrm{d}\xi)\mathbb{Q}_\sigma^n\big[n^{\alpha(s,p,\pi)/2}\big\|F(\xi,\cdot)-\sigma^2\big\|_{L^p([0,1])} \geq c'\big] \geq \delta'.$$

Since ν and F are arbitrary, it suffices then to identify the randomised decision $F(\xi, \cdot)$ with the estimator $\widehat{\sigma}_n$ in $\mathcal{E}_n$ transported into a random decision in $\mathcal{E}_n'$ with the Markov kernel K appearing in (4.28). We thus obtain

$$\int_C \mu(\mathrm{d}\sigma)K\mathbb{Q}_\sigma^n\big[n^{\alpha(s,p,\pi)/2}\big\|\widehat{\sigma}_n^2-\sigma^2\big\|_{L^p([0,1])} \geq c'\big] \geq \delta' \tag{4.30}$$

for large enough n. Putting together (4.27), (4.28) and (4.30), we finally obtain

$$n^{\alpha(s,p,\pi)/2}\widetilde{\mathbb{E}}\big[\big\|\widehat{\sigma}_n^2-\sigma^2\big\|_{L^p([0,1])}\mathbb{I}_{\{\sigma^2\in\mathcal{B}^s_{\pi,\infty}(c)\}}\big] \geq \delta'-\varepsilon > 0$$

for large enough n. The proof of Theorem 2.12 is complete.

Acknowledgments

The research of Axel Munk and Johannes Schmidt-Hieber was supported by DFG Grant FOR 916 and GK 1023. The research of Marc Hoffmann was supported by the Agence Nationale de la Recherche, Grant No. ANR-08-BLAN-0220-01. Parts of the presented work were developed within the Ph.D. project [35]. We thank two referees and an editor for valuable comments improving the paper considerably.

References

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

[2] Y. Ait-Sahalia, P. A. Mykland and L. Zhang. Ultra high frequency volatility estimation with dependent microstructure noise. *J. Econometrics* **160** (2011) 160–175. MR2745875

[3] F. M. Bandi and J. R. Russell. Separating microstructure noise from volatility. *J. Financ. Econom.* **79** (2006) 655–692.

[4] F. M. Bandi and J. R. Russell. Market microstructure noise, integrated variance estimators, and the accuracy of asymptotic approximations. *J. Econometrics* **160** (2011) 145–159. MR2745874

[5] O. Barndorff-Nielsen, P. Hansen, A. Lunde and N. Shephard. Designing realised kernels to measure the ex-post variation of equity prices in the presence of noise. *Econometrica* **76** (2008) 1481–1536. MR2468558

[6] B. Bercu and A. Touati. Exponential inequalities for self-normalized martingales with application. *Ann. Appl. Probab.* **18** (2008) 1848–1869. MR2462551

[7] T. Cai, A. Munk and J. Schmidt-Hieber. Sharp minimax estimation of the variance of Brownian motion corrupted with Gaussian noise. *Statist. Sinica* **20** (2010) 1011–1024. MR2729850

[8] Z. Ciesielski, G. Kerkyacharian and B. Roynette. Quelques espaces fonctionnels associés à des processus gaussiens. *Studia Math.* **107** (1993) 171–204. MR1244574

[9] A. Cohen. *Numerical Analysis of Wavelet Methods.* Elsevier, Amsterdam, 2003. MR1990555

[10] A. Cohen, I. Daubechies and P. Vial. Wavelets on the interval and fast wavelet transforms. *Appl. Comput. Harmon. Anal.* **1** (1993) 54–81. MR1256527

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

[12] S. Donnet and A. Samson. Estimation of parameters in incomplete data models defined by dynamical systems. *J. Statist. Plann. Inference* **137** (2007) 2815–2831. MR2323793

[13] S. Donnet and A. Samson. Parametric inference for mixed models defined by stochastic differential equations. *ESAIM Probab. Stat.* **12** (2008) 196–218. MR2374638

[14] D. Donoho and I. M. Johnstone. Ideal spatial adaptation via wavelet shrinkage. *Biometrika* **81** (1994) 425–455. MR1311089

[15] D. Donoho, I. M. Johnstone, G. Kerkyacharian and D. Picard. Wavelet shrinkage: Asymptopia? *J. Roy Statist. Soc. Ser. B* **57** (1995) 301–369. MR1323344

[16] J. Fan. A selective overview of nonparametric methods in financial econometrics. *Statist. Sci.* **20** (2005) 317–357. MR2210224

[17] B. Favetto and A. Samson. Parameter estimation for a bidimensional partially observed Ornstein–Uhlenbeck. *Scand. J. Stat.* **37** (2010) 200–220. MR2682296

[18] V. Genon-Catalot, C. Laredo and D. Picard. Non-parametric estimation of the diffusion coefficient by wavelets methods. *Scand. J. Stat.* **19** (1992) 317–335. MR1211787

[19] A. Gloter and M. Hoffmann. Estimation of the Hurst parameter from discrete noisy data. *Ann. Statist.* **35** (2007) 1947–1974. MR2363959

[20] A. Gloter and J. Jacod. Diffusions with measurement errors. I. Local asymptotic normality. *ESAIM Probab. Stat.* **5** (2001) 225–242. MR1875672

[21] A. Gloter and J. Jacod. Diffusions with measurement errors. II. Optimal estimators. *ESAIM Probab. Stat.* **5** (2001) 243–260. MR1875673

[22] P. Hall and C. C. Heyde. *Martingale Limit Theory and Its Applications.* Academic Press, New York, 1980. MR0624435

[23] W. Härdle, D. Picard, A. Tsybakov and G. Kerkyacharian. *Wavelets, Approximation and Statistical Applications.* Springer, Berlin, 1998. MR1618204

[24] M. Hoffmann. Minimax estimation of the diffusion coefficient through irregular sampling. *Statist. Probab. Lett.* **32** (1997) 11–24. MR1439493

[25] M. Hoffmann. Adaptive estimation in diffusion processes. *Stochastic Process. Appl.* **79** (1999) 135–163. MR1670522

[26] 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** (2009) 2249–2276. MR2531091

[27] I. Karatzas and S. E. Shreve. *Brownian Motion and Stochastic Calculus.* Springer, New York, 1998. MR0917065

[28] G. Kerkyacharian and D. Picard. Thresholding algorithms, maxisets and well-concentrated bases. *Test* **9** (2000) 283–345. MR1821645

[29] A. Munk and J. Schmidt-Hieber. Lower bounds for volatility estimation in microstructure noise models. In *Borrowing Strength: Theory Powering Applications – A Festschrift for Lawrence D. Brown* 43–55. *Inst. Math. Stat. Collect.* **6**. Institute of Mathematical Statistics, Beachwood, OH, 2010.

[30] A. Munk and J. Schmidt-Hieber. Nonparametric estimation of the volatility function in a high-frequency model corrupted by noise. *Electron. J. Stat.* **4** (2010) 781–821. MR2684388

[31] P. A. Mykland and L. Zhang. Discussion of "A selective overview of nonparametric methods in financial econometrics" by Jianqing Fan. *Statist. Sci.* **20** (2005) 347–350.

[32] M. Podolskij and M. Vetter. Estimation of volatility functionals in the simultaneous presence of microstructure noise and jumps. *Bernoulli* **15** (2009) 634–658. MR2555193

[33] M. Reiß. Asymptotic equivalence for inference on the volatility from noisy observations. *Ann. Statist.* **39** (2011) 772–802. MR2816338

[34] M. Rosenbaum. Estimation of the volatility persistence in a discretely observed diffusion model. *Stochastic Process. Appl.* **118** (2008) 1434–1462. MR2427046

[35] J. Schmidt-Hieber. Nonparametric methods in spot volatility estimation. Ph.D. thesis, 2011. Available at http://webdoc.sub.gwdg.de/diss/2011/schmidt_hieber.

[36] E. Schmisser. Non-parametric drift estimation for diffusions from noisy data. *Statist. Decisions* **28** (2011) 119–150. MR2810618

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

[38] L. Zhang. Efficient estimation of stochastic volatility using noisy observations: A multi-scale approach. *Bernoulli* **12** (2006) 1019–1043. MR2274854

[39] 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.* **100** (2005) 1394–1411. MR2236450