# Nonparametric methods in spot volatility estimation

Dissertation zur Erlangung des Doktorgrades der Mathematisch-Naturwissenschaftlichen Fakultäten der Georg-August-Universität zu Göttingen

> vorgelegt von Anselm Johannes Schmidt-Hieber aus Freiburg im Breisgau

> > Göttingen, 2010

## D7 Referent: Prof. Dr. Axel Munk Koreferent: Prof. Dr. Lutz Dümbgen

## Mitglieder der Prüfungskommission:

Prof. Dr. Axel Munk

- Prof. Dr. Lutz Dümbgen
- Prof. Dr. Marc Hoffmann
- Prof. Dr. Thorsten Hohage
- Prof. Dr. Preda Mihailescu
- Prof. Dr. Anja Sturm

## Tag der mündlichen Prüfung: 26. Oktober 2010

 $Dedicated\ to\ my\ grandfather$ 

## Acknowledgements

In the first place, I would like to thank my advisor Prof. Axel Munk for his constant support, many vivid discussions, his inspiring enthusiasm and intuition for statistics, his mentorship, and his great optimism regarding this work. He has always been open to new ideas, asking the right, challenging questions at the right moments, while contributing essentially to the proofs and overall framework of the project. Secondly, I am grateful to Prof. Marc Hoffmann from ENSAE, Paris for introducing me to the interplay of wavelets and diffusion processes and sharing his ideas with me. The second part of this thesis on random volatility would not exist without his valuable contributions. Special thanks are owed to Prof. Lutz Dümbgen for taking the Koreferat, his interest in the topic, many helpful comments, some joint work, and a number of nice stays at University Bern.

For some joint work on related problems, I am grateful to Prof. T. Tony Cai, Dr. Mathias Vetter as well as my colleague Till Sabel for joint work on implementation of block thresholding. Moreover, I want to thank Prof. Markus Reiß, Markus Bibinger, and Prof. Mark Podolskij for a number of fruitful discussion and inspiring comments on volatility estimation. I also appreciate the comments of my colleagues and friends Sophie Frick, Thomas Rippl as well as Elisabeth Schmidt-Hieber's help with my English.

The support by the DFG-SNF research group 916 "Statistical Regularization and Qualitative Constraints" and RTG 1023 "Identification in Mathematical Models" is gratefully acknowledged, providing an excellent bases for three years of research. I thank the Collaborative Research Center 649 "Economic Risk" at Humboldt University Berlin for providing me with access to Eurex database.

Moreover, I want to express my gratitude for the opportunity to stay at IMS. I have to thank all the members for providing a friendly and stimulating atmosphere. I am particularly indebted to my office colleague Philipp Marnitz for his companionship and encouragement. The IMS has been an excellent working environment. Thanks for all the fun we had together, teaching me how to brew coffee and the daily waiting at the canteen until I finished my meal.

Finally and most of all, I would like to note that this project would have been impossible without the support of my family and my girlfriend Mareike Woratz.

## Summary

This work is devoted to study a model, where we observe a stochastic process under additional measurement noise. The main objective is to estimate a "fluctuation measure", called the volatility/intermittency of the latent (unobservable) process, given the perturbed data.

To state it more formally, suppose that we observe  $Y = (Y_{1,n}, \ldots, Y_{n,n})^t$ ,

$$Y_{i,n} := X_{i/n} + \epsilon_{i,n}, \quad i = 1, \dots, n.$$

Here,  $X_t := \int_0^t \sigma_{s,t} dW_s$ , where W denotes a standard Brownian motion. The random vector  $\epsilon = (\epsilon_{1,n}, \ldots, \epsilon_{n,n})^t$  models the measurement noise. Given the data, the goal is to estimate the spot volatility, i.e.  $s \mapsto \sigma_{s,s}^2$  by methods of nonparametric statistics.

In this work we will deal with two major subproblems: The cases where  $\sigma_{s,t}$  is deterministic (X is a Gaussian process) and the case where  $\sigma_{s,t} = \sigma_s$  (X is a (continuous) semimartingale). We refer to them as the Gaussian Volterra and semimartingale problem, respectively. These models are motivated by applications from turbulence modeling and finance.

The Gaussian Volterra model is entirely new and we show that reconstruction of the spot volatility can be accomplished through spectral decomposition of the covariance combined with Fourier series estimation. For the semimartingale model, we prove that wavelet thresholding, based on pre-averaging as a first step, leads to an adaptive estimator of the spot volatility. In both models the estimators converge with the optimal rate of convergence (up to some logarithmic factors) under fairly general assumptions regarding the noise process. Our finding is that microstructure noise leads to a general reduction of the rates of convergence by a factor 1/2. Finally, we illustrate the estimators by numerical simulations and application to log-returns of high-frequency stock data.

# Contents

1.	Introduction	9
	1.1. Models	9
	1.2. The Gaussian Volterra model	10
	1.3. The semimartingale model and log-returns	11
	1.4. Main results	12
2.	Preliminaries from nonparametric statistics	15
	2.1. Nonparametric estimators	16
	2.2. Thresholding	19
	2.3. Large deviations for martingales	20
	2.4. Function spaces $\ldots$	22
	2.5. Asymptotic optimality	26
	2.6. Asymptotic equivalence	30
3.	Spot volatility estimation - state of the art	31
	3.1. Spot volatility estimation without microstructure noise	31
	3.2. Including microstructure noise	33
4.	Fourier series estimation in the Gaussian Volterra model	35
	4.1. A short overview on Gaussian Volterra processes	35
	4.2. Estimation of $\langle \phi, \sigma^2 \rangle$	37
	4.3. Fourier series estimator of the spot volatility	45
	4.4. Optimizing tuning parameters	48
	4.5. Comparison of estimators for integrated volatility	49
5.	Spot volatility estimation in the semimartingale model	53
	5.1. Pre-averaging	53
	5.2. Estimation of the wavelet coefficients	55
	5.3. Wavelet estimator	63
	5.4. Optimizing tuning parameters	65
6.	Lower bounds and adaptivity	71
7.	Simulations in the Gaussian Volterra model	75

8.	Application of the semimartingale model				
	8.1.	Modeling of high-frequency data	81		
		8.1.1. Modeling log-returns	81		
		8.1.2. Modeling microstructure noise	82		
	8.2.	Real data performance	87		
9.	Disc	cussion and outlook	93		
Ар	penc	lices	95		
Α.	Proofs and technical results for Chapters 2 and 4				
	A.1.	Some facts about multivariate statistics and linear algebra	112		
в.	Prod	ofs and technical results for Chapter 5	117		
		B.0.1. Proof of Proposition 4	128		
	B.1.	Tools from stochastic analysis	134		
Bił	oliog	raphy	139		

# Chapter 1

## Introduction

Microstructure noise models have attracted a lot of attention, recently. These models can be employed in order to model turbulence and moreover, they play a prominent role for modeling high-frequency returns in financial statistics. The quantity of interest within these models is the local variance as a function of time, called the spot volatility. In this thesis, we develop a statistical theory of nonparametric spot volatility estimation in microstructure noise models. The introduction outlines the major results of this thesis. To this end, we begin with a more detailed explanation of the models to be investigated.

## 1.1. Models

Before we can formally state the models, some definitions need to be introduced. Throughout this work we assume that a filtered probability space  $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t\geq 0}, \mathbb{P})$  is given and the process  $(W_t)_{t\geq 0}$  denotes a standard  $((\mathcal{F}_t)_{t\geq 0}, \mathbb{P})$ -Brownian motion.

**Definition 1** (Gaussian Volterra process). A process  $(X_t)_{t\geq 0}$  which has an integral representation of the form  $X_t = \int_0^t \sigma_{s,t} dW_s$  is called a Gaussian Volterra process provided  $(s,t) \mapsto \sigma_{s,t}$  is a deterministic function and  $\int_0^t \sigma_{s,t}^2 ds < \infty$ .

For more on these processes see the discussion below as well as the examples given in Section 4.1.

**Definition 2** (continuous Itô semimartingale). A continuous Itô semimartingale is a process  $(X_t)_{t\geq 0}$  of the form

$$X_t = \int_0^t b_s ds + \int_0^t \sigma_s dW_s, \qquad (1.1.1)$$

where  $\sigma$  and b are  $(\mathcal{F}_t)$ -adapted and càdlàg (right continuous with left limits) processes.

Gaussian Volterra model: Suppose we observe

$$Y_{i,n} = X_{i/n} + \epsilon_{i,n}, \quad i = 1, \dots, n,$$
 (1.1.2)

where X is a Gaussian Volterra process and  $\epsilon$  is some noise process.

Semimartingale model: Suppose we observe

$$Y_{i,n} = X_{i/n} + \epsilon_{i,n}, \quad i = 1, \dots, n,$$
 (1.1.3)

where X is a continuous Itô semimartingale and  $\epsilon$  is some noise process.

The statistical challenge in these models is to estimate the spot volatility, i.e. the functions  $s \mapsto \sigma_{s,s}^2$  and the path  $s \rightsquigarrow \sigma_s^2$  (here  $\rightsquigarrow$  indicates that this is a random function) in models (1.1.2) and (1.1.3), respectively.

Moreover, the following structure on the noise  $\epsilon = (\epsilon_{1,n}, \ldots, \epsilon_{n,n})^t$  is imposed for both models.

Assumption 1 (General assumption on the noise). We assume that  $\epsilon_{i,n}$  is a product of the form

$$\epsilon_{i,n} = \tau(i/n, X_{i/n})\eta_{i,n}, \qquad (1.1.4)$$

Here,  $(\eta_{i,n})_{i=1,\dots,n}$  is an i.i.d. sequence, independent of X such that for every  $1 \leq i \leq n$ ,

$$\mathbb{E}[\eta_{i,n}] = 0, \ \mathbb{E}[\eta_{i,n}^2] = 1, \ and \ \mathbb{E}[\eta_{i,n}^4] < \infty.$$

The function  $\tau(.,.)$  is called the noise level and is assumed to be continuous.

For the procedures derived later on, some further refinements on this assumption are required. In the following two sections we introduce the main applications.

### 1.2. The Gaussian Volterra model

Gaussian Volterra processes have been studied in different settings, mainly for theoretical questions regarding Gaussian processes (see for instance Baudoin and Nualart [11] and the references therein), but also recently in applications for turbulence modeling (cf. Barndorff-Nielsen and Schmiegel [8], Section 3). Here,  $(X_t)_{t\geq 0}$  would be the velocity of a turbulent flow over time at a fixed point in space. Therefore, we can interpret the observation vector in the Gaussian Volterra model as measurements from the velocity process at time points i/n under additional measurement noise. In these models the volatility is usually called intermittency and measures the degree of turbulence. For a realistic modeling on small scales one should allow for stochastic volatility/intermittency. So far, this is not covered by our theoretical framework, but as we show by numerical simulations (cf. Chapter 7) the proposed estimators work well even in the case of random volatility/intermittency. We further want to mention that there is a case of particular interest, namely if  $\sigma_{s,t} = g(t-s)\tilde{\sigma}_s$ , for a function  $g \in L^2$  and a càdlàg process  $\tilde{\sigma}$ , for which estimation of the averaged volatility/intermittency (without additional measurement noise) has been studied recently by Barndorff-Nielsen et al. [6].



Figure 1.1.: Tick data of FGBL on July 25th, 2007 between 9 a.m. and 6 p.m. The second plot shows the reconstruction of the spot volatility based on wavelet thresholding.

## 1.3. The semimartingale model and log-returns

The semimartingale model originates from finance. Since the seminal work by Delbaen and Schachermayer [21, 22] it is well known that semimartingales provide a natural class for price processes. However, these results do not incorporate so-called market frictions due to bid-ask spread and rounding errors, among others. For low-frequencies (i.e. the price is sampled in the range of minutes or even larger time scales), microstructure noise is negligible; these effects occur only if we use data sampled on high frequencies, less than a few seconds, say. The latter has the advantage that we do not need to throw away a large part of the data, but estimation is much more involved. Recently, many estimators have been proposed in order to estimate the so-called integrated volatility, i.e. the average volatility over a given time span, in the high-frequency setting under microstructure noise.

However, in many cases the interest lies in the shape of the volatility itself instead of the averaged value, as for studying the fluctuations of the volatility around the income of public news (cf. Andersen and Bollerslev [4]) or the daily volatility pattern (see Figure 1.1). In these cases the spot volatility (or instantaneous volatility), i.e. the path of the volatility as a function of time, must be estimated from the data.

## 1.4. Main results

Estimation in models (1.1.2) and (1.1.3) is surprisingly difficult, for two reasons. First of all, contrary to the usual nonparametric regression model, the data are heavily dependent. The covariance structure of  $Y = (Y_{1,n}, \ldots, Y_{n,n})^t$  is particularly difficult to deal with in the Gaussian Volterra setting. If there is no additive noise, i.e.  $\tau = 0$ , this dependence can easily be removed, by considering increments. Then, estimation of the spot volatility amounts essentially to a classical nonparametric regression problem (cf. Hoffmann [41]). It is the inclusion of microstructure noise which makes the problem much more difficult. In fact, simple quadratic variation methods fail if noise is present (cf. Zhang et *al.* [77]). Secondly, in the semimartingale model we are forced to estimate a random quantity, i.e. the path  $s \rightsquigarrow \sigma_s^2$ , nonparametrically.

In this thesis, we develop a theory for estimation in the Gaussian Volterra and the semimartingale model. In particular, we construct optimal estimators for the spot volatility that overcome the problems described above.

In the Gaussian Volterra model, we construct an orthogonal series estimator of the spot volatility, relying on spectral properties of the covariance structure. We do not necessarily need to expand the series with respect to a particular basis and only some general regularity is required. As examples, we explicitly outline the cases of the cosine and trigonometric basis.

Considering integrated mean square error risk, it can be shown in general that these estimators converge with rate  $n^{-\alpha/(4\alpha+2)}$ , provided that the mappings  $s \mapsto \sigma_{s,t}$ ,  $\forall t \in [0, 1]$  and  $t \mapsto \sigma_{s,t}$ ,  $\forall s \leq t$  are Hölder continuous with index 1/4 and 7/8, respectively (cf. Theorems 3 and 4). Here,  $\alpha$  denotes the Sobolev index of the spot volatility, that is assumed to be larger than 1. For expansion with respect to cosine basis, we even may relax this to  $\alpha > 3/4$ .

Despite its simplicity, a general well known disadvantage of Fourier series estimators is their difficulty to localize in time domain, resulting in the Gibb's phenomenon for instance. To overcome this problem, it is near at hand to use techniques that allow for simultaneous localization in time and frequency domain, such as wavelets instead.

This will be done in the second part of the thesis, for the semimartingale model (1.1.3). In particular, we make use of the pre-averaging technique developed in Podolskij and Vetter [68] and Jacod et *al.* [44] as a first step. Whereas for the Fourier estimator the noise is filtered in the spectral domain, the pre-average method relies on another idea: It allows us to separate the semimartingale from the noise because of the different smoothness properties.

Denote by  $B_{p,q}^s$  a Besov space with parameters (p,q,s) and let  $\mathcal{B}_{p,q}^s(C) := \{f \in B_{p,q}^s : \|f\|_{B_{p,q}^s} \leq C\}$  be the corresponding Besov ball. In Besov spaces the smoothness of a function is measured by the index s, in first order.

For  $\pi \in (0,\infty)$  and  $s > 1/\pi$  we introduce the effective smoothness function corresponding

to the Besov space  $B^s_{\pi,\infty}$  by

$$t \mapsto s(t) := s - \left(t - \frac{1}{\pi}\right)_{-},$$
 (1.4.1)

where  $(x)_{-} := -\min(x, 0)$  is the negative part, which is always non-negative. Moreover, denote by  $\pi^*$  the (necessarily) unique solution of

$$s(1/\pi^{\star}) = \frac{1}{2} \left(\frac{p}{\pi^{\star}} - 1\right).$$

Suppose that  $\alpha_0 + 1/\pi \leq s \leq \alpha_0/(1 - 2\alpha_0)$ , where  $0 < \alpha_0 \leq 1/2$  is some prespecified number. Then, with respect to  $L^p$ -norm the proposed wavelet thresholding estimator,  $\hat{\sigma}^2$  achieves the rate of convergence

$$v_n := \left(\frac{\log^{3/2}(n)}{n}\right)^{s(1/\pi^*)/(4s(1/\pi^*)+2)}$$

in the sense that

$$\overline{\lim_{n \to \infty}} v_n^{-1} \mathbb{E} \Big[ \left\| \widehat{\sigma}^2 - \sigma^2 \right\|_p \mathbb{I}_{\{\sigma^2 \in \mathcal{B}^s_{\pi,\infty}(C)\}} \Big] < \infty.$$

In fact, this is half of the usual rate obtained in the classical nonparametric regression setting (cf. Kerkyacharian and Picard [52]). In particular, if  $p/(2s+1) \leq \pi$ , we obtain the rate of convergence  $n^{-s/(4s+2)}$ , up to a logarithmic factor.

Furthermore, the estimator does not depend on the smoothness of the spot volatility, since it adapts automatically to it. In order to prove this, we make use of recently derived large deviation inequalities for martingales (cf. Bercu and Touati [12]). Particularly with regard to the application described in Section 1.3, adaptivity is a crucial property since there is no consensus on the path regularity of spot volatility in financial statistics.

Moreover, we consider different techniques in order to derive lower bounds in microstructure noise models. The methods rely on a new bound of the Kullback-Leibler divergence as well as a result on asymptotic equivalence, recently derived in Reiß [71]. This allows us to prove that the obtained rates are optimal in minimax sense.

Finally, we discuss numerical simulations and real data applications. In order to obtain a clear picture of the performance of the Fourier series estimator, various simulation studies are carried out. We show that if the sample size is larger than  $10^3$ , reasonable reconstructions can be obtained even in the case of random volatility.

In a second part of our study, we apply the semimartingale model to tick data. We demonstrate that combining the wavelet estimator with blockwise thresholding leads to stable results (for an example of a reconstruction see Figure 1.1).

## 1. Introduction

# Chapter 2

## Preliminaries from nonparametric statistics

In order to understand the presented work, various techniques are required, in particular from stochastic analysis and nonparametric statistics. It is beyond the scope of this thesis to introduce them both in full length. The focus of today's research in volatility models is mainly related to the estimation of integrated volatility and its variations, studying these problems by means of stochastic analysis. Therefore, we place special emphasis on terms and definitions in nonparametric statistics. In this chapter, we present a short overview of the main ideas. Besides the introductory style, a number of advanced results needed in subsequent chapters are proven. We take for granted that the reader is familiar with standard facts from stochastic analysis and only a few results are recalled in Section B.1.

Although there are different understandings of the field, nonparametric statistics is usually defined as the study of statistical models with infinite dimensional parameter spaces. Typically these parameter spaces are function spaces.

In this chapter we focus on asymptotic nonparametrics, i.e. we study properties of estimators if the sample size n tends to infinity. Nowadays, research on nonparametrics is also concerned with finite sample behavior. For our models the asymptotic approach provides powerful tools to study volatility estimation problems while, at the same time, proofs are still tractable.

The main approach employed for the purpose of handling estimation problems in nonparametrics is to approximate an infinite dimensional parameter space  $\mathcal{G}$  by a finite dimensional subspace  $\mathcal{H}$ . Then, estimation can be performed within a parametric setting where an additional bias is introduced by the finite dimensional approximation of  $\mathcal{G}$ . In order to find good estimators, the dimension of  $\mathcal{H}$  must be chosen of the right order, usually dependent on properties of  $\mathcal{G}$ . A standard example is the minimum number of derivatives of a function in  $\mathcal{G}$ , provided that  $\mathcal{G}$  is a function class. Moreover, the dimension of  $\mathcal{H}$  depends in general on the sample size.

In this chapter we will first introduce two general estimation methods, namely Fourier series and wavelet estimators. Section 2.4 is devoted to the discussion of some function spaces. Finally, in the last two sections, we study asymptotic properties.

### 2.1. Nonparametric estimators

The most popular way to estimate functions nonparametrically is by means of kernel estimators. Yet other methods have also been employed. Since they will prove useful in the sequel, we will concentrate in this section on nonparametric estimation by orthogonal (Fourier) series and wavelets.

#### Fourier series estimator

Given observations  $Y_{i,n}(f)$ , i = 1, ..., n where  $f \in L^2$  is some unknown function. Suppose that there is an estimator  $\langle \phi, f \rangle$ , of the scalar product  $\langle \phi, f \rangle$  for some function  $\phi \in G \subset L^2$ . Let  $(\phi_i)_i \subset G$  be an  $L^2$ -basis. Then, we may estimate f via

$$\widehat{f} = \sum_{i=0}^{\infty} \widehat{w_{i,n}\langle\phi_i, f\rangle}\phi_i, \qquad (2.1.1)$$

where  $(w_{i,n})_i$  is a triangular scheme of tapering weights.

A popular choice of a basis system on [0, 1] is

$$\psi_1(\cdot) := 1, \ \psi_{2i}(\cdot) := \sqrt{2}\cos(2\pi i \cdot), \ \text{and} \ \psi_{2i+1}(\cdot) := \sqrt{2}\sin(2i\pi \cdot),$$
 (2.1.2)

the so-called trigonometric basis.

For instance, in the classical nonparametric regression model, i.e.

$$Y_{i,n} = f(\frac{i}{n}) + \epsilon_{i,n}, \quad \mathbb{E}[\epsilon_{i,n}] = 0, \quad i = 1, \dots, n$$
 (2.1.3)

an estimator of the scalar product  $\langle \phi, f \rangle$  is given by its empirical version

$$\widehat{\langle \phi, f \rangle} = \frac{1}{n} \sum_{i=1}^{n} \phi(\frac{i}{n}) Y_{i,n}.$$
(2.1.4)

An example for the sequence of weights is to reconstruct the first N coefficients, i.e.

$$w_{i,n} := \begin{cases} 1, & \text{for } i \leq N, \\ 0, & \text{otherwise,} \end{cases}$$

where  $N = N_n$  should be chosen in dependence on the unknown smoothness and n. Another approach is to shrink the estimates  $T(\phi_i)$  by choosing  $w_i < 1$ . It is well known that this might improve the estimate considerably (cf. Tsybakov [74], Chapter 3). Since the spot volatility is always non-negative it is important to note that positivity of a function can be incorporated as well. Knowing that  $f \ge 0$ , a possible choice is

$$w_{i,n}^{(F)} = (1 - \frac{i}{n})_+,$$
 (2.1.5)

where  $(x)_{+} = x$  if x > 0 and  $(x)_{+} = 0$ , otherwise. Note that  $\sum_{i=0}^{\infty} w_{i,n}^{(F)} \langle \phi_i, f \rangle \phi_i$  is the (n-1)-th Fejér kernel convolved with f. By the positivity of the Fejér kernel it is easy to see that for all n, the approximations  $\sum_{i=0}^{\infty} w_{i,n}^{(F)} \langle \phi_i, f \rangle \phi_i$  are non-negative, provided  $f \ge 0$ . However, this does of course not imply that also the estimator  $\hat{f}$  is non-negative.

Compared to other methods, Fourier estimates are superior in recovering periodic signals. Moreover, the estimators can be computed very efficiently, provided a fast Fourier transform can be employed.

One of the drawbacks is that besides positivity and smoothness other properties such as monotonicity or convexity do not have an easy translations into Fourier coefficients. Moreover, if a function has a jump, the reconstruction by Fourier series shows strong oscillating behavior around this point. This is the well known Gibb's phenomenon. Therefore, Fourier series are not preferable for estimation of jump functions.

#### Wavelet estimators

Wavelets can be seen as a refinement of Fourier series estimators resolving a number of drawbacks from orthogonal series estimation. For a comprehensive summary of wavelet theory, we refer to Cohen [18] and Daubechies [20].

The advantages of wavelets are that localization in the Fourier domain and time domain can be performed simultaneously, reducing Gibb's effects for instance. More generally, regions of different smoothness can be handled. This is a crucial property since it is believed that the smoothness of the spot volatility may change over time. In fact, time spans could occur where the volatility is of Brownian smoothness alternating with regions of arbitrary high smoothness.

We introduce wavelets by the abstract multiresolution analysis approach due to Mallat [57] and Meyer [61] (see also Chapter 5 in Daubechies [20]).

**Definition 3** (Multiresolution analysis). Let  $(V_j)_{j \in \mathbb{Z}}$  be a sequence of nested and closed  $L^2(\mathbb{R})$ -subspaces, i.e.  $V_j \subset V_{j+1}$  for all  $j \in \mathbb{Z}$ . Further denote by  $P_j$  the projection operator on  $V_j$ . If

(i) 
$$\overline{\bigcup_{j\in\mathbb{Z}}V_j} = L^2(\mathbb{R}),$$

(*ii*) 
$$\bigcap_{j\in\mathbb{Z}} V_j = \{0\},\$$

(*iii*) 
$$\lim_{j\to\infty} P_j f = f \text{ for all } f \in L^2(\mathbb{R}).$$

(iv) 
$$f \in V_0 \Rightarrow f(\cdot - k) \in V_0$$
 for every integer k,

(v) 
$$f \in V_j \Leftrightarrow f(2^{-j} \cdot) \in V_0,$$

(vi) there exists  $\phi \in V_0$  such that  $\{\phi(\cdot - k)\}_{k \in \mathbb{N}}$  is an orthonormal basis of  $V_0$ ,

then  $((V_j)_j, \phi)$  is called a multiresolution analysis. Moreover, if  $\phi$  can be chosen such that for any non-negative  $\alpha$ ,  $\alpha \leq r$  and for each  $m \in \mathbb{N}$ ,

$$|\partial^{\alpha}\phi(x)| \lesssim \frac{1}{(1+|x|)^m},$$

then the multiresolution analysis is called r-regular. Here,  $\leq$  means larger up to a constant, uniformly over x.

The essential conditions underlying the scaling properties of multiresolution analysis are the last three mentioned above. To state it differently, for  $f \in V_0$ , all translations by an integer are in  $V_0$ , again requiring a particular discrete shift invariance. Additionally, whenever,  $f \in V_j$  then  $f(2 \cdot) \in V_{j+1}$ , meaning that there is an increase of frequency by a factor of 2. This further implies that  $\{\phi(2^j \cdot -k)\}_{k \in \mathbb{Z}}$  is an orthogonal basis of  $V_j$ .

Since  $V_0 \subset V_1$  and  $\{\phi(2 \cdot -k)\}_{k \in \mathbb{N}}$  is a basis of  $V_0$ , we may represent

$$\phi = \sum_{k \in \mathbb{Z}} h_k \phi(2 \cdot -k), \qquad (2.1.6)$$

where  $(h_k)_k \in l^2(\mathbb{Z})$ . This is the so-called refinement equation that turns out to be of fundamental importance for wavelets.

In fact, the last condition can be relaxed by only assuming  $\{\phi(\cdot - k)\}_{k \in \mathbb{N}}$  to be a Riesz basis. However, in this case the subsequent results need some modifications (cf. Chapter 2.2 in [18]).

For a function g we introduce the notation  $g_{j,k} := 2^{j/2}g(2^j \cdot -k)$ .

**Lemma 1.** Let  $((V_j)_j, \phi)$  be a multiresolution analysis. Then, there exists  $\psi \in L^2(\mathbb{R})$  such that  $\{\psi_{j,k}\}_{j,k\in\mathbb{Z}}$  is an  $L^2(\mathbb{R})$  basis and

$$f = \sum_{k} \langle \phi_{j_0,k}, f \rangle \, \phi_{j_0,k} + \sum_{j=j_0}^{\infty} \sum_{k \in \mathbb{Z}} \langle \psi_{j,k}, f \rangle \, \psi_{j,k} \quad in \ L^2(\mathbb{R}), \tag{2.1.7}$$

for all  $j_0 \in \mathbb{Z}$  and  $f \in L^2(\mathbb{R})$ . The function  $\psi$  is called wavelet.

Besides the existence stated in the last lemma, we may easily construct a wavelet for given multiresolution analysis  $((V_j)_j, \phi)$  by means of Fourier analysis or through the direct use of the refinement equation (2.1.6) via

$$\psi = \sum_{k \in \mathbb{Z}} (-1)^k h_{1-k} \phi(2 \cdot -k)$$
, where convergence is in  $L^2$ .

**Example 1** (Haar basis). Let  $V_0$  be the space of  $L^2(\mathbb{R})$  functions that are constant on blocks [i-1,i),  $i \in \mathbb{Z}$ . Further let  $\phi = \mathbb{I}_{[0,1)}(\cdot)$ . Then,  $((V_j)_j, \phi)$  is a multiresolution analysis with regularity r = 0. For the coefficients of the refinement equation (2.1.6) we obtain  $h_0 = h_1 = 1$ ,  $h_k = 0$  for  $k \notin \{0,1\}$ . Hence, the corresponding wavelet is  $\psi = \mathbb{I}_{[0,1/2)}(\cdot) - \mathbb{I}_{[1/2,1)}(\cdot)$ .

Wavelet estimators are extensions of Fourier series estimators in the sense that (2.1.1) is replaced by

$$\widehat{f} = \sum_{k} \langle \widehat{\phi_{j_0,k}, f} \rangle \phi_{j_0,k} + \sum_{j=j_0}^{j_1} \sum_{k \in \mathbb{Z}} \langle \widehat{\psi_{j,k}, f} \rangle \psi_{j,k}.$$
(2.1.8)

The positive integers  $j_0, j_1$  can be chosen by the statistician. Whereas in many applications  $j_0 = 0$  is a reasonable choice, this is not true for spot volatility estimation. In fact, for small  $j_0$  and low-smoothness of the volatility, the estimator  $\langle \phi_{j_0,k}, f \rangle$  has a large bias. Therefore, choosing  $j_0$  in dependence of the sample size improves the rate of convergence.

In (2.1.8), we truncate the expansion at level  $j_1$ . As for the Fourier series estimators, we can introduce some weights in order to downweight the reconstruction at higher resolution levels. In the following, we introduce the more general concept of wavelet thresholding instead, since this allows us to obtain estimators enjoying some additional optimality properties. Note that thresholding is not restricted to wavelets only, but can also be applied to Fourier series estimation.

### 2.2. Thresholding

The concept of wavelet thresholding has been introduced by Donoho and Johnstone [25] as well as Donoho et *al.* [26]. Generally speaking, the idea is to keep large coefficients in the expansion, while removing or downweighting small coefficients for which we cannot be sure whether they contain significant information about the unknown signal. Let us introduce the most important examples of thresholding procedures.

**Hard-thresholding:** Define the hard-thresholding function  $\mathcal{H}_t : \mathbb{R} \to \mathbb{R}$  via

$$\mathcal{H}_t(x) := x \, \mathbb{I}_{\{|x| \ge t\}}(x). \tag{2.2.1}$$

Then, the hard-thresholded version of (2.1.8) is

$$\widehat{f_t} = \sum_k \langle \widehat{\phi_{j_0,k}, f} \rangle \phi_{j_0,k} + \sum_{j=j_0}^{j_1} \sum_{k \in \mathbb{Z}} \mathcal{H}_t(\langle \widehat{\psi_{j,k}, f} \rangle) \psi_{j,k}.$$
(2.2.2)

Soft-thresholding: Let  $\mathcal{S}_t : \mathbb{R} \to \mathbb{R}$ ,

$$\mathcal{S}_t(x) := \operatorname{sign}(x) \ (|x| - t)_+.$$

The soft-thresholded wavelet estimator is given by

$$\widehat{f} = \sum_{k} \langle \widehat{\phi_{j_0,k}, f} \rangle \phi_{j_0,k} + \sum_{j=j_0}^{j_1} \sum_{k \in \mathbb{Z}} \mathcal{S}_t(\langle \widehat{\psi_{j,k}, f} \rangle) \psi_{j,k}.$$

19

The cut-off-point t is referred to as the threshold parameter. It is well known that hardthresholding leads to optimal estimators provided the thresholding constant is chosen properly. In fact, if we can show that

$$\mathbb{P}\left(\left|\langle \widehat{\psi_{j,k}, f} \rangle - \langle \psi_{j,k}, f \rangle\right| > t\right) = \text{``small''}, \qquad (2.2.3)$$

we might reject  $\langle \psi_{j,k}, f \rangle = 0$  whenever  $|\langle \psi_{j,k}, f \rangle| > t$ . The main difficulty is to show (2.2.3) uniformly over j, k, which can be accomplished through exponential inequalities. Another advantage is that wavelet thresholding can still be used even in the case of correlated data (cf. Johnstone and Silverman [47] and Johnstone [46]). However, in practical implementations, hard-thresholded estimators experience some oversmoothing effects, since the theoretical values obtained for t are usually too large (cf. Donoho and Johnstone [25] or Abramovich and Silverman [1]). In order to circumvent these drawbacks, more elaborate thresholding procedures have been proposed. The main improvement is to threshold not term-by-term but blockwise, in order to include information on neighboring coefficients (for more details see Cai and Zhou [17] and the references therein).

For proving results of type (2.2.3) many methods are known, provided  $\langle \psi_{j,k}, f \rangle$  can be written as a sum of independent random variables. However, if we are dealing with more complex models, these findings are not sufficient, in general. Recently, large deviation inequalities have been derived, assuming more generally that  $\langle \widehat{\psi_{j,k}, f} \rangle$  can be written as a martingale. This will be discussed in the next section.

## 2.3. Large deviations for martingales

In this section we give some exponential inequalities for martingales. This will be the basic tool used to show large deviations results of type (2.2.3) for wavelet based volatility estimation. In particular, we state and prove two non-trivial reformulations of these exponential inequalities that are directly applicable to the estimation problem and show an optimality property, by comparison to known results on sums of i.i.d. Gaussian random variables.

Throughout this section let  $(M_k)_k$  be a discrete, locally square integrable, real  $(\mathcal{F}_k)_k$  martingale with  $M_0 := 0$ . In order to stay consistent with the overall notation, the martingale increments are defined as the forward differences, i.e.  $\Delta_i M := M_{i+1} - M_i$ . We denote by  $[M]_k = \sum_{i=0}^{k-1} (\Delta_i M_i)^2$  its quadratic variation and by  $\langle M \rangle_k = \sum_{i=0}^{k-1} \mathbb{E}[(\Delta_i M)^2 | \mathcal{F}_i]$  its predictable compensator. The martingale  $(M_k)_k$  is said to be conditionally symmetric, if  $\Delta_i M$  given  $\mathcal{F}_i$  follows a symmetric distribution. In the following, we list a number of results.

**Lemma 2.** (i) If  $a_k \leq \Delta_k M \leq b_k$  a.s. for deterministic constants  $a_k$ ,  $b_k$ ,  $a_k < b_k$ 

then, for  $x \ge 0$ ,

$$\mathbb{P}(|M_k| \ge x) \le 2 \exp\left(\frac{-2x^2}{\sum_{r=0}^{k-1} (b_r - a_r)^2}\right).$$

(ii) If  $\mathbb{E}[|\Delta_k M|^q | \mathcal{F}_k] \leq c^q q^q$ , for q = 2, 3, ... and for some constant c, then,

$$\mathbb{P}(|M_k| \ge x) \le 2 \exp\left(\frac{-x^2}{2ce(2ck+x)}\right),\,$$

where e is Euler's number.

(iii) If  $M_n$  is conditionally symmetric then for x, y > 0

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

(iv) For x, y > 0

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

**Remark 1.** (i) is the extension of Hoeffding's inequality by Azuma [5], (ii) is due to Hoffmann [41], (iii) was proven in de la Peña [67] and (iv) can be found in Bercu and Touati [12]. Variations and extensions of these results can be found in van de Geer [75] and [12].

From Lemma 2 (iii)-(iv), we infer the following large deviation inequalities that will be used in order to prove (2.2.3). This is essentially Lemma 4.11 in Hoffmann et *al.* [43].

**Lemma 3.** Assume that for  $p, m \ge 1$  there exists a deterministic sequence  $(C_j)_j$  (*j* depends on m, i.e.  $j = j_m$ ), fixed  $\delta, \epsilon > 0$  and  $0 < q_0 \le 1$  such that

- (i)  $\mathbb{P}(\langle M \rangle_j > C_j(1+\delta)) \lesssim m^{-p}$ ,
- (*ii*)  $C_j \gtrsim j^{1/2+\epsilon}$ ,
- (iii)  $m^{q_0} \leq j \leq m$ ,
- (iv) for every  $\kappa \geq 2$ ,

$$\max_{i=0,\dots,j-1} \mathbb{E}\big[|\Delta_i M|^\kappa\big] \lesssim 1.$$

Then,

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

Moreover, if M is conditionally symmetric, then we obtain, under the same conditions, the sharper result

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

A proof can be found in Appendix A. The inequalities above state that  $|M_j| > t(j, p, m) := \sqrt{2C_j p \log m}$  with probability smaller than a constant times  $m^{-p}$ . It is of uttermost importance to derive very sharp bounds, i.e. to find the smallest possible t(j, p, m) since this will later determine the size of the threshold. Otherwise, if t(j, p, m) is large, this will in practice result in a severe oversmoothing of the wavelet thresholding estimator.

In order to show that Lemma 3 provides us with almost sharp constants, let us consider the following example. Suppose that  $M_j = \sum_{i=1}^{j} \xi_j$ , where  $(\xi_j)_j$  is a sequence of i.i.d. standard normal random variables. In this case we know, by using Mills-ratio, that for large m,

$$\frac{m^{-p}}{\sqrt{p\log m}} \lesssim P(|M_j| > \sqrt{2pj\log m}) \lesssim \frac{m^{-p}}{\sqrt{p\log m}}.$$

Clearly, in this example,  $C_j$ , as defined in Lemma 3, is j and by the second part of Lemma 3,

$$\mathbb{P}(|M_j| > (1+\delta)\sqrt{2pj\log m}) \lesssim m^{-p}.$$

Hence, for this situation Lemma 3 gives us the optimal t(j, p, m) up to an arbitrary small number, whereas if we do not assume that the martingale is conditionally symmetric (i.e. the first part of the Lemma applies), we lose by a factor of  $\sqrt{2}$ .

## 2.4. Function spaces

In general, it is not clear to which function space the spot volatility may belong, especially if we are dealing with random volatility. Hence, the results are proven under the assumption of different spaces. In this section they are defined and we discuss some properties, in particular embeddings, used later on.

Throughout the work, let  $\|.\|_{L^p[a,b]}$  denotes the  $L^p$ -norm on the interval [a, b]. In the baseline case [a, b] = [0, 1], we abbreviate the  $L^p$ -norm by  $\|.\|_p$ . Moreover,  $\|.\|_{p,m}$  is the empirical  $L^p$ -norm on [0, 1], i.e.  $\|f\|_{p,m} := (\frac{1}{m} \sum_{i=1}^m |f(\frac{i}{m})|^p)^{1/p}$ . Since there is no need in this thesis to define function spaces in their most general form we restrict ourselves to functions of one variable on the domain [0, 1].

#### Hölder and Sobolev spaces

First we introduce modifications of the classical function spaces measuring smoothness by means of derivatives and the decay of Fourier series coefficients. Recall that for a real number x,  $\lfloor x \rfloor$  denotes the largest integer not greater than x.

**Definition 4** (Hölder space). Given  $\alpha > 0$ . Then, the space of Hölder continuous functions with index  $\alpha$  is

$$\mathcal{C}^{\alpha} := \left\{ f \in L^{\infty} : f^{(p)} \text{ exists for } p = \lfloor \alpha \rfloor, \\ \exists \ L < \infty, \ \left| f^{(p)}(x) - f^{(p)}(y) \right| \le L \left| x - y \right|^{\alpha - p}, \ \forall x, y \in [0, 1] \right\}$$

**Definition 5** (Hölder ball). Given  $\alpha, L > 0$ . Then, we define the Hölder ball with index  $\alpha$  by

$$\mathcal{C}^{\alpha}\left(L\right) := \left\{ f \in \mathcal{C}^{\alpha} : \|f\|_{\infty} + \frac{\left|f^{(p)}(x) - f^{(p)}(y)\right|}{|x - y|^{\alpha - p}} \le L, \ p = \lfloor \alpha \rfloor, \ \forall x, y \in [0, 1] \right\}.$$

Assume that  $(\phi_k)_k$  is a basis of  $L^2[0, 1]$ . (Fractional) Sobolev spaces with respect to this basis are defined by a condition on the decay of the basis coefficients.

**Definition 6** (Sobolev space). Given an  $L^2[0,1]$  basis  $(\phi_k)_k$ ,  $\alpha > 0$  and a sequence of non-negative weights  $(a_i)_i$ . The Sobolev space with (smoothness) index  $\alpha$  is defined via

$$\Theta(\alpha, (\phi_k)_k) := \left\{ f \in L^2[0, 1] : \sum_{i=0}^{\infty} a_i^{2\alpha} \langle \phi_i, f \rangle^2 < \infty \right\}.$$

**Definition 7** (Sobolev ellipsoid). The corresponding Sobolev ellipsoid/ball is given by

$$\Theta(\alpha, (\phi_k)_k, C) := \left\{ f \in L^2[0, 1] : \sum_{i=0}^{\infty} a_i^{2\alpha} \langle \phi_i, f \rangle^2 \le C \right\}.$$

In the following, we introduce the main examples.

Sobolev space with respect to trigonometric basis: Recall the definition of the standard trigonometric basis on  $L^2[0,1]$  given in (2.1.2) and set  $a_i := i$ , for i even and  $a_i := i - 1$ , for i odd. The corresponding Sobolev space and ellipsoid will be denoted by  $\Theta_{\text{trig}}(\alpha)$  and  $\Theta_{\text{trig}}(\alpha, C)$ . This is consistent with the classical definition of Sobolev ellipsoids (cf. Tsybakov [74]). In particular, if  $\alpha$  is a positive integer,  $\Theta_{\text{trig}}(\alpha, C)$  has a simpler representation in terms of derivatives and  $L^2$  balls, that will turn out to be important later. Let

$$W(\alpha, L) := \left\{ f \in L^2[0, 1] : f^{(\alpha - 1)} \text{ absolutely continuous, } \|f^{(\alpha)}\|_2^2 \le L \right\}$$

and

$$W_{\text{trig}}(\alpha, L) := \left\{ f \in W(\alpha, L) : f^{(k)}(0) = f^{(k)}(1), \ k = 0, \dots, \alpha - 1 \right\}.$$
 (2.4.1)

Then, for  $\alpha \in \mathbb{N} \setminus \{0\}$ ,

$$\Theta_{\text{trig}}(\alpha, C) = W_{\text{trig}}(\alpha, \pi^{2\alpha}C).$$
(2.4.2)

A proof of this well known fact can be found in Tsybakov [74], p.196.

Sobolev space with respect to cosine basis: Let

$$\{\phi_i, i = 0, \ldots\} := \left\{1, \sqrt{2}\cos(i\pi t), i = 1, \ldots\right\}$$
(2.4.3)

be the cosine basis and  $a_i := i$ . In this case the notation  $\Theta_{\cos}(\alpha)$ ,  $\Theta_{\cos}(\alpha, C)$  is used for the Sobolev space/ellipsoid. For integer  $\alpha$  we may similarly to (2.4.1), introduce

 $W_{\cos}(\alpha, L) := \left\{ f \in W(\alpha, L) : f^{(k)}(0) = f^{(k)}(1) = 0, \text{ for } k \text{ odd, } k < \alpha \right\}.$ 

Then, for  $\alpha \in \mathbb{N} \setminus \{0\}$ ,

$$\Theta_{\cos}(\alpha, C) = W_{\cos}(\alpha, \pi^{2\alpha}C).$$
(2.4.4)

Since the proof of the " $\supseteq$ "-inclusion is a non-trivial extension of the one for (2.4.2), it will be given in Appendix A for the sake of completeness.

For spot volatility estimation,  $\Theta_{cos}$  will appear to be a natural smoothness space (cf. Section 4.2). However, since in nonparametric statistics series estimators are commonly considered with respect to the trigonometric basis, this case will be treated as well. The next paragraph gives some insights into the interplay between these function spaces.

Comparison of  $\Theta_{\text{trig}}$  and  $\Theta_{\cos}$ : First note that functions having smoothness  $\alpha$  in one space, say  $f \in \Theta_{\text{trig}}(\alpha)$  may have a different index with respect to  $\Theta_{\cos}$ . For example  $\sin(2\pi \cdot) \in \Theta_{\text{trig}}(\alpha)$  for all  $\alpha < \infty$ . In contrast,  $\sin(2\pi \cdot) \in \Theta_{\cos}(\alpha)$  for  $\alpha < 3/2$ , only. To give an example where the smoothness of a function is smaller in  $\Theta_{\text{trig}}$  than in  $\Theta_{\cos}$ , consider  $\cos(\pi \cdot)$ . By explicit calculations,  $\cos(\pi \cdot) \in \Theta_{\text{trig}}(\alpha)$  for  $\alpha < 1/2$  and  $\cos(\pi \cdot) \in \Theta_{\text{trig}}(\alpha)$  for  $\alpha < \infty$ . Next we describe a subclass of functions in which  $\Theta_{\text{trig}}$ and  $\Theta_{\cos}$  coincide.

Let  $\mathcal{S}$  be the class of all functions  $f \in L^2[0,1]$  such that  $f(x) = f(1-x), \forall x \in [0,1]$ . Then,

$$f \in \Theta_{\text{trig}}(\alpha, C) \cap \mathcal{S} \iff f \in \Theta_{\cos}(\alpha, C) \cap \mathcal{S}.$$
 (2.4.5)

**Proof of (2.4.5).** If  $f \in S$  then  $\int_0^1 f(x) \sin(2k\pi x) dx = \int_0^1 f(x) \cos((2l-1)\pi x) dx = 0$  for  $k, l \in \mathbb{N}$ . Therefore, the only nonzero series coefficients in both the trigonometric and the cosine basis must be of the form  $\int_0^1 f(x) \cos(2k\pi x) dx, k = 0, \ldots$ 

To end this comparison, note that for  $\alpha = 1$  we have by (2.4.2) and (2.4.4) that

$$\Theta_{\text{trig}}(1,C) \subset \Theta_{\cos}(1,C). \tag{2.4.6}$$

#### **Besov** spaces

A Besov space depends on three parameters, allowing for a more refined analysis of smoothness properties than Hölder and Sobolev spaces. In the way Besov spaces are defined here, we can see directly that they are linked to the decay of wavelet coefficients. For more general definitions and an overview on Besov spaces, we refer to Cohen [18].

Let us define the Besov norm by

$$\|f\|_{B^{s}_{p,q}} := \left\|\sum_{k} \langle \phi_{0,k}, f \rangle \phi_{0,k}\right\|_{p} + \left\| \left(2^{js} \|\sum_{k} \langle \psi_{j,k}, f \rangle \psi_{j,k} \|_{p}\right)_{j \ge 0} \right\|_{l^{q}},$$

where  $\|.\|_{l^q}$  denotes the sequence space norm on the index set I, i.e.  $\|(a_i)_{i\in I}\|_{l^q} :=$  $(\sum_{i \in I} |a_i|^q)^{1/q}$  with obvious modification for  $q = \infty$ .

**Definition 8** (Besov space). Assume that  $((V_i)_i, \phi)$  is an r-regular multiresolution analysis and denote by  $\psi$  the corresponding wavelet. Moreover, suppose  $0 < s < r, p, q \ge 1$ . Then, a function f belongs to the Besov space  $B_{p,q}^s$  if and only if the Besov norm  $\|.\|_{B_{p,q}^s}$ is finite.

**Definition 9** (Besov ball). For given r-regular multiresolution analysis  $((V_j)_j, \phi)$  we define the Besov ball  $\mathcal{B}_{p,q}^s(C) := \{f \in B_{p,q}^s : \|f\|_{B_{p,q}^s} \leq C\}.$ 

It is worth mentioning, that there is an equivalent Besov norm, defined via

$$\|f\|_{B^{s}_{p,q}} := \left\| \sum_{k} \langle \phi_{0,k}, f \rangle \phi_{0,k} \right\|_{p} + \left\| \left( 2^{js} \| (|\langle \psi_{j,k}, f \rangle|)_{k \in \mathbb{Z}} \|_{l^{p}} \right)_{j \ge 0} \right\|_{l^{q}}$$

and for p > 1, the equivalence follows directly from Meyer [62], Lemma 8. Obviously, we have

$$\begin{aligned} \| \cdot \|_{B_{p_{1},q}^{s}} &\leq \| \cdot \|_{B_{p_{2},q}^{s}}, \quad \text{for } p_{1} \leq p_{2}, \\ \| \cdot \|_{B_{p,q_{1}}^{s}} &\leq \| \cdot \|_{B_{p,q_{2}}^{s}}, \quad \text{for } q_{1} \leq q_{2}, \end{aligned}$$

$$(2.4.7)$$

and hence the embeddings  $B_{p_2,q}^s \subset B_{p_1,q}^s$  and  $B_{p,q_2}^s \subset B_{p,q_1}^s$  are continuous and hold without increasing the norm.

Moreover, we have the Sobolev-type embedding

$$B_{p_1,q}^{s_1} \subset B_{p_2,q}^{s_2}, \quad \text{for } s_1 - 1/p_1 = s_2 - 1/p_2, \quad p_1 \le p_2, \quad q \in [1,\infty],$$
 (2.4.8)

which is also continuous. In particular, Besov spaces are generalizations of Hölder (or more precisely Zygmund) spaces, consequently using Definition 4,

$$B^s_{\infty,\infty} = \mathcal{C}^s$$
, if s is not an integer. (2.4.9)

In order to work with these spaces, it is important to understand the qualitative properties of functions in  $B_{p,q}^s$  for a given triple (p, q, s). First, smoothness such as differentiability is measured by s (this is reflected by the embedding (2.4.9)). The parameter q allows for a finer discrimination of smoothness than s and eventually, p restricts functions in  $B_{p,q}^s$  to functions that are also in  $L^p$ .

## 2.5. Asymptotic optimality

In this section we introduce the notion of optimal rates of convergence and adaptivity.

To be precise, some definitions are in order. In the following,  $\Theta$  will denote the (possibly infinite dimensional) parameter space and  $l: \Theta \times \Theta \rightarrow [0, \infty)$  the loss function.

Examples for  $\Theta$  are function spaces, for instance Hölder, Sobolev and Besov spaces as introduced in Section 2.4. For the loss function we will only use global measures, such as  $l(\theta_1, \theta_2) = \|\theta_1 - \theta_2\|_p$  although local distances could be considered as well, for instance  $l(\theta_1, \theta_2) = |\theta_1(x_0) - \theta_2(x_0)|, x_0 \in [0, 1]$  fixed,  $\Theta$  some function space on [0, 1].

The risk of an estimator is defined by  $\mathbb{E}_{\theta}[l(\widehat{\theta}_n, \theta)]$ . In particular, if  $l(\theta_1, \theta_2) = \|\theta_1 - \theta_2\|_p$ we call  $\mathbb{E}_{\theta}[l(\widehat{\theta}_n, \theta)]$  the  $L^p$ -risk. Furthermore, we say that an estimator has rate of convergence  $v_n$  if

$$\overline{\lim_{n \to \infty}} v_n^{-1} \mathbb{E}_{\theta} \left[ l(\widehat{\theta}_n, \theta) \right] < \infty.$$
(2.5.1)

Given an estimator  $\widehat{\theta}_n$  the maximum risk is defined by  $R_n(\widehat{\theta}_n, \Theta) := \sup_{\theta \in \Theta} \mathbb{E}_{\theta}[l(\widehat{\theta}_n, \theta)]$ , where  $\mathbb{E}_{\theta}$  is the expectation with respect to the probability distribution  $\mathbb{P}_{\theta}$ . Accordingly, we define the minimax risk, given by

$$R_n(\Theta) := \inf_{\widehat{\theta}_n} R_n(\widehat{\theta}_n, \Theta)$$

and the infimum is taken over all estimators. Note that  $R_n(\hat{\theta}_n, \Theta)$  and  $R_n(\Theta)$  are indexed by n in order to indicate the dependence on the sample size.

**Definition 10** (Optimal rate of convergence/asymptotic efficiency). An estimator is said to achieve the optimal rate of convergence in minimax sense on  $(\Theta, l)$  if

$$\overline{\lim_{n \to \infty}} R_n(\widehat{\theta}_n, \Theta) / R_n(\Theta) < \infty.$$

If furthermore

$$\lim_{n \to \infty} R_n(\widehat{\theta}_n, \Theta) / R_n(\Theta) = 1$$

then  $\widehat{\theta}_n$  is said to be asymptotically efficient.

A popular risk measure on  $L^2$  is the so-called integrated mean square error (IMSE), defined as

$$\text{IMSE}(\widehat{f}) := \int_0^1 \text{MSE}(\widehat{f}(t)) dt,$$

where

$$MSE(\widehat{f}(t)) := \mathbb{E}[(\widehat{f}(t) - f(t))^2]$$
(2.5.2)

is the mean square error. The IMSE risk is particularly suitable for Fourier series estimation as introduced in Section 2.1. Indeed, the following result holds.

**Theorem 1** (IMSE of series estimator). Suppose that  $\hat{f}$  is given by (2.1.1) and assume that there is a sequence of positive integers,  $(q_n)_n$ , tending to infinity and a function space  $\Theta$  such that

$$\sup_{f \in \Theta} \max_{i \le q_n} \operatorname{MSE}\left(\widehat{\langle \phi_i, f \rangle}\right) \lesssim q_n^{-2}.$$
(2.5.3)

Further suppose that  $\omega_{i,n} = 0$  for  $i > q_n$ . Then,

$$\sup_{f \in \Theta} \text{IMSE}(\widehat{f}) \lesssim q_n^{-2} \sum_{i=0}^{q_n} \omega_{i,n}^2 + \sum_{i=0}^{\infty} (1 - \omega_{i,n})^2 \langle \phi_i, f \rangle^2.$$

Proof. First note that

$$\text{IMSE}(\widehat{f}) = \int_0^1 \text{MSE}(\widehat{f}(t))dt = \int_0^1 \text{Bias}^2(\widehat{f}(t))dt + \int_0^1 \text{Var}(\widehat{f}(t))dt.$$

By taking advantage of the orthogonality, we obtain

$$\int_0^1 \operatorname{Var}(\widehat{f}(t)) dt = \sum_{i=0}^{q_n} w_{i,n}^2 \operatorname{Var}(\widehat{\langle \phi_i, f \rangle}).$$

and

$$\int_{0}^{1} \operatorname{Bias}^{2}(\widehat{f}(t))dt = \sum_{i=0}^{q_{n}} (w_{i,n} \mathbb{E}\left[\widehat{\langle \phi_{i}, f \rangle}\right] - \langle \phi_{i}, f \rangle)^{2} + \sum_{i=q_{n}+1}^{\infty} \langle \phi_{i}, f \rangle^{2}$$
$$\leq 2 \sum_{i=0}^{q_{n}} w_{i,n}^{2} (\mathbb{E}\left[\widehat{\langle \phi_{i}, f \rangle}\right] - \langle \phi_{i}, f \rangle)^{2} + 2 \sum_{i=0}^{\infty} (1 - \omega_{i,n})^{2} \langle \phi_{i}, f \rangle^{2}.$$

1	-	-	

#### **Random parameters**

Before we can address the semimartingale problem as introduced in Chapter 1, we need to extend the concept of rates of convergence to estimation of random functions, since in this model the sample paths  $s \rightsquigarrow \sigma_{s,s}^2$  are stochastic. Without loss of generality, we restrict ourselves here to random functions defined on [0, 1]. Hence, the parameters are allowed to be random, whereas the "parameter space"  $\Theta$  is fixed. An estimator is understood as a random function, measurable with respect to the observations.

**Definition 11.** Given a filtered probability space  $(\Omega, \mathcal{F}, (\mathcal{F})_{t\geq 0}, \mathbb{P})$  and let  $\Theta$  be a function class. Suppose that  $\hat{\theta}_n$  is an estimator of  $\theta = (\theta_t)_{t\in[0,1]}$ , and  $\theta$  is

(i) adapted to the filtration  $(\mathcal{F}_t)_{t\geq 0}$ ,

- (ii)  $\{\theta \in \Theta\}$  is measurable with respect to  $\mathcal{F}$ , and
- (iii)  $\theta$  is independent of n.

Then, we say that the estimator  $\hat{\theta}_n$  achieves the rate of convergence  $v_n$  over  $\Theta$  with respect to the loss function l if

$$\overline{\lim_{n \to \infty}} v_n^{-1} \mathbb{E} \Big[ l(\widehat{\theta}_n, \theta) \mathbb{I}_{\big\{ \theta \in \Theta \big\}} \Big] < \infty,$$
(2.5.4)

where  $l(\widehat{\theta}_n, \theta) \mathbb{I}_{\left\{\theta \in \Theta\right\}} := 0$  whenever  $\theta \notin \Theta$ .

Suppose that  $\theta$  is deterministic. In this case we require  $\theta \in \Theta$  and recover (2.5.1). For the other extreme, i.e.  $\mathbb{P}(\theta \in \Theta) = 0$ , any sequence  $(v_n)$  is a rate of convergence; hence, this is non-informative. Yet: if  $0 < \mathbb{P}(\theta \in \Theta) < 1$  then we have a non-trivial extension of (2.5.1). This is the typical situation we face in spot volatility estimation. For an example, consider the case that  $\theta$  is a Brownian motion and let  $\alpha < 1/2$ . As it is well known  $\mathbb{P}(\theta \in \mathcal{C}^{\alpha}) = 1$ . But for any constant C,  $0 < C < \infty$  we have  $0 < \mathbb{P}(\theta \in \mathcal{C}^{\alpha}(C)) < 1$ .

### Adaptivity

Often, in function estimation, the smoothness of the true function is unknown. For instance, we face this situation in spot volatility estimation later, where no consensus has been reached on the path regularity. Hence, the parameter space  $\Theta$  must be taken as the union of a scale of smoothness spaces  $(\Theta_{\alpha})_{\alpha \in I}$ , i.e.  $\Theta = \bigcup_{\alpha \in I} \Theta_{\alpha}$ , where *I* is some index set. In this setting we wish to find estimators with the following property: Whenever the true function lies in a subspace, say  $\Theta_{\alpha}$  then the estimator should perform as well as a rate-optimal estimator in the same experiment with parameter space  $\Theta_{\alpha}$ . An estimator fulfilling this condition is said to be adaptive over the scale  $(\Theta_{\alpha})_{\alpha \in I}$ . Let us summarize: adaptive estimators a rate-optimal over many parameter spaces, simultaneously. In particular, they do not require knowledge of the index  $\alpha$ .

**Definition 12** (Adaptivity). Let  $(\Theta_{\alpha})_{\alpha \in I}$  be a family of parameter spaces indexed by  $\alpha$ . Then, we say that an estimator  $\widehat{\theta}_n$  is (rate-)adaptive on  $((\Theta_{\alpha})_{\alpha \in I}, l)$  if

$$\overline{\lim_{n \to \infty}} \frac{R_n(\theta_n, \Theta_\alpha)}{R_n(\Theta_\alpha)} < \infty, \quad for \ all \ \alpha \in I.$$

If the optimal rates up to logarithmic factors are achieved then we say that this estimator is adaptive up to log-terms. For the nonparametric regression setting, it is a classical result that adaptivity is in general possible with respect to IMSE-risk, whereas for pointwise risk adaptivity up to log-terms holds, only. A technique used to obtain adaptive estimators is wavelet thresholding as introduced in Section 2.2. In the following, we outline this method in more detail.

Recall the definition of effective smoothness functions, given in (1.4.1) and note that by definition,  $s(\cdot)$  is a positive function. The following result is a slight extension of Kerkyacharian and Picard [52] and Gloter and Hoffmann [32] for adaptive estimation of random functions. In order to be self-contained, a proof can be found in Appendix A. We always assume that f is measurable.

**Theorem 2.** Given an r-regular multiresolution analysis  $((V_j)_j, \phi)$  and suppose that  $\phi$  as well as the corresponding wavelet  $\psi$  have compact support. For given  $0 < \alpha_0 \leq 1/2$  pick  $(j_0, j_1)$  such that

$$2^{j_0} \sim q_n^{2\alpha_0 - 1}, \quad 2^{j_1} \sim q_n^{-1/(2\alpha_0 + 1)}.$$

Assume that  $\alpha_0 < s - 1/\pi$ ,  $\pi \in [1, \infty)$ , s < r. Let  $(q_n)_n$  be a sequence converging to zero such that for any  $p \ge 1$  and  $j_0 \le \ell \le j_1$ ,

(i)

$$\mathbb{E}^{1/p}\Big[|\langle \widehat{\phi_{j_0,k}, f} \rangle - \langle \phi_{j_0,k}, f \rangle|^p \, \mathbb{I}_{\{f \in \mathcal{B}^s_{\pi,\infty}(C)\}}\Big] \lesssim q_n^{1/2},$$

(ii)

$$\mathbb{E}^{1/p} \Big[ |\langle \widehat{\psi_{\ell,k}, f} \rangle - \langle \phi_{\ell,k}, f \rangle|^p \, \mathbb{I}_{\{f \in \mathcal{B}^s_{\pi,\infty}(C)\}} \Big] \lesssim q_n^{1/2},$$

(iii)

$$\mathbb{P}\left(|\langle \widehat{\psi_{\ell,k}, f} \rangle - \langle \psi_{\ell,k}, f \rangle| \ge \frac{\kappa(p)}{2} \sqrt{q_n \log(1/q_n)} \text{ and } f \in \mathcal{B}^s_{\pi,\infty}(C)\right) \lesssim q_n^{\max(2,p)}$$

Then, the estimator  $\hat{f}_t$  defined in (2.2.2) satisfies for  $t = \kappa(p)\sqrt{q_n \log(1/q_n)}$  and any  $p \ge 1$ 

$$\mathbb{E}\left[\|\widehat{f}_{t} - f\|_{p} \mathbb{I}_{\{f \in \mathcal{B}_{\pi,\infty}^{s}(C)\}}\right] \lesssim \left(q_{n} \log^{3/2}(1/q_{n})\right)^{s(1/\pi^{\star})/(2s(1/\pi^{\star})+1)} + q_{n}^{\alpha_{0}}, \qquad (2.5.5)$$

where  $\pi^*$  is the unique solution to

$$s(1/\pi^*) := \frac{1}{2} \left( \frac{p}{\pi^*} - 1 \right). \tag{2.5.6}$$

It is easy to see that  $t \mapsto s + (t - 1/\pi)_- - (tp - 1)/2$  has a unique and finite positive root. Thus, the solution  $\pi^*$  exists and is unique. Note further that  $s(1/\pi^*) = s$  iff  $\pi \ge p/(2s+1)$ . In the sparse region, i.e.  $\pi < p/(2s+1)$ , we loose in terms of convergence rates.

Moreover, note that  $\hat{f}$  does not depend on  $\pi$ , s and C, whereas the rate of convergence does. Hence, this provides us with a natural candidate for an adaptive estimator. Indeed, in Chapter 5, we demonstrate how we can construct a wavelet estimator for the spot volatility satisfying the assumptions above. Because of the factor  $\log^{3/2}(1/q_n)$  in (2.5.5), adaptivity holds only up to log-terms.

### 2.6. Asymptotic equivalence

In this section we recall a number of facts on asymptotic equivalence and give a bound on the Hellinger distance needed in Chapter 4. Moreover in Chapter 6, an asymptotic equivalence result, derived recently in Reiß [71], will be utilized in order to obtain lower bounds. For a concise treatment of the topic, we refer to Le Cam and Young [53].

Let  $\mathcal{E}_{1,n}$  and  $\mathcal{E}_{2,n}$  be two experiments with the same parameter space  $\Theta$ . Further denote by  $\delta_1$ ,  $\delta_2$  decision procedures in  $\mathcal{E}_{1,n}$  and  $\mathcal{E}_{2,n}$ , respectively. Let A denote an arbitrary action space. For a loss function  $L : \Theta \times A \to [0, \infty)$ . We introduce the norm ||L|| := $\sup_{\theta \in \Theta, a \in A} L(\theta, a)$ . The corresponding risks with respect to in  $\mathcal{E}_{1,n}$  and  $\mathcal{E}_{2,n}$  are denoted by  $R_1(\theta, \delta_1, L, n), R_2(\theta, \delta_2, L, n)$ .

**Definition 13** (Asymptotic equivalence). The Le Cam deficiency between two statistical experiments is given by

$$\Delta\left(\mathcal{E}_{1,n}, \mathcal{E}_{2,n}\right) = \max\left[\inf_{\delta_{1}}\sup_{\delta_{2}}\sup_{\theta\in\Theta}\sup_{L:\|L\|=1}\left|R_{1}\left(\delta_{1}, L, \theta, n\right) - R_{2}\left(\delta_{2}, L, \theta, n\right)\right|, \\ \inf_{\delta_{2}}\sup_{\delta_{1}}\sup_{\theta\in\Theta}\sup_{L:\|L\|=1}\left|R_{1}\left(\delta_{1}, L, \theta, n\right) - R_{2}\left(\delta_{2}, L, \theta, n\right)\right|\right],$$

where  $\delta_1, \delta_2$  are two decision procedures. If  $\lim_{n\to\infty} \Delta(\mathcal{E}_{1,n}, \mathcal{E}_{2,n}) = 0$  then the two experiments are said to be asymptotically equivalent.

Let us define the Hellinger distance of two probability measures P and Q by  $d_H(P,Q) := (\int (\sqrt{dP} - \sqrt{dQ})^2)^{1/2}$ . Further, denote by  $P_{n,\theta}$  and  $Q_{n,\theta}$  the probability measures of the observations in the experiments  $\mathcal{E}_{1,n}$  and  $\mathcal{E}_{2,n}$ , respectively. Then, it can be shown that

$$\Delta^2(\mathcal{E}_{1,n}, \mathcal{E}_{2,n}) \le 4 \sup_{\theta \in \Theta} d_H^2(P_{n,\theta}, Q_{n,\theta})$$
(2.6.1)

(cf. Nussbaum, [66], Equation 12), provided both experiments are defined on the same probability space. The Hellinger distance can in general be bounded more easily than the le Cam deficiency and provides a closed form representation for a number of distributions.

The following bound on the Hellinger distance will be of importance later. For two *n*-variate centered Gaussian measures  $\mathbb{P}_1, \mathbb{P}_2$  with invertible  $n \times n$  covariance matrices  $\Sigma_1, \Sigma_2$  it holds that

$$d_{H}^{2}(\mathbb{P}_{1},\mathbb{P}_{2}) \leq 2 \left\| \Sigma_{1}^{-1/2} (\Sigma_{2} - \Sigma_{1}) \Sigma_{1}^{-1/2} \right\|_{2}^{2}$$
  
= 2  $\left\| (T\Sigma_{1}T^{t})^{-1/2} (T\Sigma_{2}T^{t} - T\Sigma_{1}T^{t}) (T\Sigma_{1}T^{t})^{-1/2} \right\|_{2}^{2},$  (2.6.2)

where  $\|.\|_2$  denotes the Frobenius (or Hilbert-Schmidt) norm (for a definition see Lemma A.7) and T is an invertible  $n \times n$  matrix (cf. Reiß [71], Section 9).

# Chapter 3

## Spot volatility estimation - state of the art

Reconstruction of the spot volatility has been considered within the last few years, but most of the work was done in the simpler, noise-free case. Since 2009, estimation under microstructure noise has been addressed as well. This chapter gives an overview of related work.

## 3.1. Spot volatility estimation without microstructure noise

In this section, we present three approaches to spot volatility estimation for low-frequency data that have been studied over the past years. The first method is based on Fourier series, while the second relies on local averaging. Finally, the third method uses numerical differentiation of the integrated volatility. In the main part of this thesis, some of these ideas are extended to high-frequency data. Since this is technically demanding and the different approaches can be presented easily in the low-frequency setting, it is worthwhile to study them separately.

#### Fourier estimator

In this part we summarize the results by Malliavin and Mancino [58]. Suppose that we observe a continuous semimartingale X.

Although the model assumes that the entire realization of the process is known to the statistician, the aim is to give a reconstruction method that is stable if this assumption fails and the process is observed only along a fine, but discrete grid. For this reason a reconstruction formula based on Fourier series is developed. Suppose that  $\sigma^2$  is a continuous function, then

$$\sigma_N^2(t) = \sum_{k=-\infty}^{\infty} \left(1 - \frac{|k|}{N}\right)_+ F(\sigma^2)(k) e^{2\pi i k t} \stackrel{N \to \infty}{\to} \sigma^2(t), \quad \text{for almost all } t \in [0, 1],$$

$$(3.1.1)$$

where  $F(\sigma^2)(k) := \int_0^1 \sigma_t^2 \exp(-2\pi i k t) dt$  is the Fourier transform and the weights  $\left(1 - \frac{|k|}{N}\right)$ are chosen in order to preserve the non-negativity constraint on  $\sigma^2$  (see also (2.1.5) and the discussion thereafter). Furthermore, we extend this notation to stochastic differentials via  $F(dX)(k) := \int_0^1 \exp(-2\pi i k t) dX_t$ . The key point in the analysis is to define a convolution-type operator of two  $\mathbb{Z}$ -valued functions by

$$(\psi \circledast \psi)(k) := \lim_{m \to \infty} \frac{1}{2m+1} \sum_{l=-m}^{m} \psi(l)\psi(k-l).$$
 (3.1.2)

It can be shown that  $F(\sigma^2) = F(dX) \circledast F(dX)$ , where convergence is in probability (cf. Malliavin and Mancino [58], Theorem 2.1). Given the path  $(X_t)_t$ , F(dX) can be computed and so can  $F(\sigma^2)$ . This allows us to reconstruct  $\sigma_N^2$ .

In particular,  $F(\sigma^2)(0) = \int_0^1 \sigma_s^2 ds$ . By truncation of N, m in (3.1.1) and (3.1.2), respectively, the estimator of  $F(\sigma^2)(0)$  can be robustified in order to deal with market microstructure noise and discrete data. Hence, this gives a consistent estimator of the integrated volatility (cf. Malliavin and Mancino [58] as well as Mancino and Sanfelici [59]).

#### Reformulation as a regression problem

The following paragraph is based on another idea from Hoffmann [41, 42]. Given a diffusion process X, observed at time points i/n, i = 0, ..., n. Assume the drift to be zero. In general it will become clear that the drift is only a nuisance parameter that has no significant effect upon either the procedure nor its theoretical properties. Then, using partial integration yields

$$Y_{i,n} = n \left( X_{(i+1)/n} - X_{i/n} \right)^2 = n \int_{\frac{i}{n}}^{\frac{i+1}{n}} \sigma^2(X_s) ds + \epsilon_{i,n},$$
(3.1.3)

where

$$\epsilon_{i,n} = 2n \int_{\frac{i}{n}}^{\frac{i+1}{n}} \left( X_s - X_{i/n} \right) \sigma(X_s) dW_s$$

Clearly,  $\epsilon_{i,n} = O_p(1)$  is uncorrelated, centered noise. Note that for a sequence of random variables  $(U_n)_n$ , we write  $U_n = O_p(c_n)$ , whenever  $c_n^{-1}U_n$  is bounded in probability. Equation (3.1.3) shows: We may transform our observed values of the diffusion process and obtain new observations  $Y_{i,n}$ ,  $i = 1, \ldots, n$  that are 'close' to a regression problem, meaning we observe  $\sigma^2(X_{i/n})$  under some additive, heteroscedastic noise. Based on these new observations, standard procedures from nonparametrics can be applied where attention must be paid to dependencies and filtrations of the process, of course. Basically all results, such as rates of convergence, carry over to this more general situation. Thus, conditional on the event that at some time point s,  $X_s$  reaches t, we may estimate  $\sigma(t)$ .

#### Numerical differentiation

In order to include jumps, power variation techniques for estimation of the spot volatility may be used as well. This has been done in Alvarez et *al.* [3]. Allowing for classes of volatilities with even infinite jump activities (for instance  $\sigma$  could be the absolute value of a Lévy process), it can be shown that suitably scaled discrete differences of the power variation of order *p* converge stably in law to  $\sigma_s^p$ , for fixed *s*.

If  $\sigma$  is of Brownian smoothness plus jumps, then the convergence rate is  $n^{-1/4}$  (cf. Alvarez et al. [3], Remark 5). From the perspective of nonparametric regression, this is clear since functions with smoothness 1/2 can be reconstructed with this rate of convergence. In contrast, if the volatility is a pure jump process without a drift component, even better rates of convergence are possible. However, it is not obvious how this technique can be modified to handle the case where  $\sigma$  is known to be of higher smoothness, for instance two-times differentiable and it is expected that in these cases numerical differentiation of the integrated volatility leads to suboptimal rates.

#### 3.2. Including microstructure noise

#### Central limit theorems

In the semimartingale model, spot volatility estimators have been constructed by Ngo and Ogawa [65]. Assume that  $(l_n)_n$  and  $(m_n)_n$  are non-decreasing sequences of integers and consider

$$\Delta_{j}\overline{Y}(s) := \frac{1}{m_{n}} \sum_{i=0}^{m_{n}-1} Y_{\lfloor sn \rfloor - 2jm_{n}-i} - Y_{\lfloor sn \rfloor - (2j+1)m_{n}-i}, \quad \text{for } s > \frac{2l_{n}m_{n}}{n}, \ j = 0, \dots, l_{n} - 1.$$

Suppose that the Hölder condition

$$\mathbb{E}\left[(\sigma_s - \sigma_t)^2\right] \lesssim |s - t|^{2\alpha}$$

holds. Then for  $s > (2l_n m_n)/n$ ,

$$\widehat{\sigma}(s) = \frac{1}{l_n} \sqrt{\frac{3\pi m_n n}{2(3m_n^2 + 1)}} \sum_{j=0}^{l_n - 1} |\Delta_j \overline{Y}(s)|$$

is an estimator of  $|\sigma(s-)|$  (i.e. the left limit at s). Under some further assumptions, and for any fixed  $s \in (0, 1]$ 

$$\sqrt{l_n} \Big( \widehat{\sigma}(s) - |\sigma(s-)| \Big) \stackrel{\mathcal{D}}{\longrightarrow} Z,$$

where Z is a bounded random variable and  $(l_n)_n$ ,  $(m_n)_n$  satisfy

$$\lim_{n \to \infty} l_n^{1+2\alpha} \left(\frac{m_n}{n}\right)^{2\alpha} = \lim_{n \to \infty} \frac{l_n n}{m_n^2} = \lim_{n \to \infty} \frac{1}{l_n} = 0.$$

This implies that  $l_n \ll n^{\alpha/(1+3\alpha)}$ ; Therefore, the rate of convergence is strictly smaller than  $n^{-\alpha/(2+6\alpha)}$ . It is quite remarkable, that the obtained estimator converges to the absolute value of  $\sigma(s)$ .

#### Volatility estimation in state space models

Another type of microstructure noise model has been introduced in Dahlhaus and Neddermeyer [19]. Here, it is assumed that the true efficient log-price X is a random walk with normally distributed increments, i.e.

$$X_{t_j} = X_{t_{j-1}} + Z_{t_j}, \quad Z_{t_j} \sim \mathcal{N}(0, \sigma_{t_j}^2)$$

where  $t_j$  are trading times and  $(\sigma_t)_t$  is allowed to vary over time. These prices cannot be observed directly due to microstructure effects, instead we observe  $Y_{t_j} = g_{t_j}(\exp(X_{t_j}))$ , where the unknown function g models rounding effects. Under the assumption that the support of the distribution of  $\exp(X_{t_j})$  is known and compact, an EM-type algorithm is developed in order to estimate the spot volatility online. However, so far no theoretical results are known for this procedure. Visual inspection of numerical simulations indicate that the estimation method needs further improvements in order to adapt to the correct smoothness of the volatility (see also [19], Figure 4).

# Chapter 4

## Spot volatility estimation under microstructure noise in the Gaussian Volterra model: Fourier series estimation

The content of the next two chapters comprise the main parts of this thesis. As mentioned in Section 2.1, in order to construct a series estimator, we must first find estimators for the scalar products  $\langle \phi, \sigma^2 \rangle = \int \phi(s) \sigma_{s,s}^2 ds$ .

Estimation of the spot volatility/intermittency in the Gaussian Volterra model has never been studied before. In order to prove rates of convergence, we extend methods from [64]. Unlike the Fourier series estimator derived in [64], we do not rely on an expansion with respect to cosine basis.

### 4.1. A short overview on Gaussian Volterra processes

Recall Definition 1 of a Gaussian Volterra process. Because these processes have up to this point been studied mainly in a different context, we will present a number of facts and give some examples here. For references on this topic, see Hida and Si Si [40] as well as Hida and Hitsuda [39]. To begin with, we provide the following examples.

#### Example 2.

- (i) If  $\sigma_{s,t} = (1-t)/(1-s)$  then X is a Brownian bridge.
- (ii) If  $\sigma_{s,t} = \sigma e^{\theta(s-t)}$  then X is an Ornstein-Uhlenbeck process.

Both integrated Brownian motion and fractional Brownian motion are Gaussian Volterra processes; however, in these cases the spot volatility degenerates. For instance, for fractional Brownian motion the Molchan-Golosov representation provides such a form and  $\sigma_{s,t} \sim (s-t)^{H-1/2}$ , for  $|s-t| \to 0$  and Hurst parameter H.

A number of non-trivial examples can be constructed from the following class of processes.

**Definition 14** (Lévy Brownian motion). A process X defined on  $\{u : u \in \mathbb{R}^d\}$  is a Lévy Brownian motion if

- (*i*)  $X_0 = 0$ ,
- (*ii*)  $\mathbb{E}[X_u] = 0, \ u \in \mathbb{R}^d$
- (*iii*)  $\mathbb{E}[(X_u X_v)^2] = |u v|,$

where |.| denotes the Euclidean distance.

For instance, one obtains standard Brownian motion by restricting the index set to a half-line starting at the origin. Moreover, a Lévy Brownian motion on the unit circle in  $\mathbb{R}^2$  can be written as a Gaussian Volterra process with kernel (cf. Si Si [73])

$$\sigma_{s,t} = \sin(t/2) \left[ \frac{1}{\sin(s/2)} - \frac{\cot(s/4)}{2} h(s) \right] + \cot^2(t/4) h(s), \ h(s) := \left( 1 + \frac{s}{4} \tan\left(\frac{s}{4}\right) \right)^{-1}$$

After constructing a number of examples, we finally state some general properties of Gaussian Volterra processes. In fact, Gaussian Volterra processes allow for a good translation between properties of the process and properties of the map  $(s,t) \mapsto \sigma_{s,t}$ .

In fact, there is a deeper connection between Gaussian Volterra processes and semimartingales. Suppose that  $(s,t) \mapsto \sigma_{s,t}$  is deterministic and the derivatives of both  $s \mapsto \sigma_{s,s}$  and  $s \mapsto \sigma_{s,t}$  exist and are denoted by  $\frac{d\sigma_{s,s}}{ds}$  and  $\partial_s \sigma$ , respectively. Then

$$\int_0^t \sigma_{s,t} dW_s \stackrel{\mathcal{D}}{=} \int_0^t \sigma_{s,s} dW_s + \int_0^t \left(\frac{d\sigma_{s,s}}{ds} - \partial_s \sigma_{s,t}\right) W_s ds, \tag{4.1.1}$$

where equality is in distribution. This can be verified by partial integration combined with comparison of the covariance. By the equation above, we see that a Gaussian Volterra process can be written as a continuous Itô semimartingale plus some generalized drift.

Note that it follows from (4.1.1) that a Gaussian Volterra process is a semimartingale if  $\sigma_{s,t} = s_1(s) + s_2(t)$  for continuously differentiable functions  $s_1$ ,  $s_2$  (for more on this see Basse [10]). Moreover, one can show that under some additional properties, a Gaussian Volterra process is Markovian, if  $\sigma_{s,t} = s_1(s)s_2(t)$  (cf. Hida and Hitsuda [39], Chapter 5). Furthermore, a Volterra process is self-similar with Hurst index 1/2 if and only if  $\sigma_{s,t} = F(s/t)$  for  $F \in L^2$  (cf. Jost [49], Lemma 2.4).

Gaussian Volterra processes are in particular suitable for modeling time-varying processes, since the state at time point t is determined only by the past  $s \leq t$ .
### 4.2. Estimation of $\langle \phi, \sigma^2 \rangle$

In this section, we construct an estimator of  $\langle \phi, \sigma^2 \rangle$ . This will be done in three steps. We work under the following more restrictive assumption on the noise.

**Assumption 2** (Refinement of the noise assumption for model (1.1.2)). Let  $\epsilon_{i,n}$  satisfy Assumption 1. Additionally, suppose that  $\tau$  does not depend on X, i.e.  $\epsilon_{i,n} = \tau(i/n)\eta_{i,n}$ .

A first step: The simplest non-trivial case is  $\phi = 1$ . Indeed in this case we aim to find estimators of  $\int_0^1 \sigma_{s,s}^2 ds$ , i.e. the so-called integrated volatility. Estimation of the integrated volatility is a problem that has been well studied and various solutions exist. It can be seen that in this case the optimal rate of convergence is  $n^{-1/4}$  (cf. Gloter and Jacod [33, 34] and Cai et al. [16]). Here, we need to extend this case to estimators of  $\langle \phi, \sigma^2 \rangle$ , where it is sufficient to consider the case  $\phi \ge 0$ . Under this restriction, a natural approach would be to treat

$$Y_{i,n}(\phi) := \sum_{j=1}^{i} \sqrt{\phi(\frac{j-1}{n})} \ (Y_{j,n} - Y_{j-1,n}), \quad Y_{0,n} := 0, \quad i = 1, \dots, n,$$
(4.2.1)

as new observations and calculate the integrated volatility within this setting, since one might expect them to be approximately

$$\widetilde{Y}_{i,n}(\phi) := \int_0^{i/n} \sqrt{\phi(s)} \ \sigma_{s,i/n} \ dW_s + \sqrt{\phi(\frac{i}{n})} \ \epsilon_{i,n}, \quad i = 1, \dots, n.$$

$$(4.2.2)$$

Note that we have equality in the special case  $\phi = 1$ , i.e.  $Y_{i,n} \stackrel{\mathcal{D}}{=} Y_{i,n}(1) \stackrel{\mathcal{D}}{=} \widetilde{Y}_{i,n}(1)$ . The problem is to quantify the quality of the approximation, in general. In the next Lemma we state a result in this direction. The corresponding probability measures of observing  $Y(\phi) := (Y_{1,n}(\phi), \ldots, Y_{n,n}(\phi))$  and  $\widetilde{Y}(\phi) := (\widetilde{Y}_{1,n}(\phi), \ldots, \widetilde{Y}_{n,n}(\phi))$  are denoted by  $\mathbb{P}_{\phi,n}$  and  $\widetilde{\mathbb{P}}_{\phi,n}$ , respectively.

**Lemma 4.** Suppose that Assumption 2 holds. Moreover assume that the volatility only depends on s and that  $\eta_{i,n} \sim \mathcal{N}(0,1)$ , i.i.d. If  $\phi = \phi_n$  satisfies

$$\inf_{\substack{n,s \\ n}} \phi_n(s) > 0, \\
\lim_{n} \sup_{\substack{s,t: \ |s-t| \le 1/n}} n^{5/8} |\phi_n(s) - \phi_n(t)| = 0, \\
\lim_{n} n^{5/4} \Big( \max_{i=0,\dots,n-1} |\Delta_{i,n}\phi_n| |\Delta_{i,n}\tau| + \max_{i=0,\dots,n-2} |\Delta_{i,n}^2\phi_n| + |\phi_n(1/n) - \phi_n(0)| \Big) = 0, \\$$
(4.2.3)

then, for  $0 < c < C < \infty$ ,

$$\lim_{n \to \infty} \sup_{c \le \sigma, \tau \le C} d_H(\widetilde{\mathbb{P}}_{\phi,n}, \mathbb{P}_{\phi,n}) = 0,$$

where  $d_H(.,.)$  denotes the Hellinger distance.

One example that will be used in order to construct an estimator with respect to cosine basis is  $\phi_n(.) = c + \cos(k_n \pi)$ , where  $k_n \in \mathbb{N}, k_n \ll n^{3/8}$  and c is some constant larger than 1.

The last lemma shows that, asymptotically, we cannot distinguish between observations from (4.2.1) and (4.2.2). Let us introduce the following submodel, where we observe

$$Y_{i,n} = \int_0^{i/n} \sigma_s dW_s + \epsilon_{i,n}, \quad i = 1, \dots, n,$$
(4.2.4)

with  $\epsilon_{i,n} = \tau(\frac{i}{n})\eta_{i,n}$  and  $\eta_{i,n} \sim \mathcal{N}(0,1)$ , i.i.d. In particular, an estimator for the integrated volatility in model (4.2.4) provides us with an estimator of  $\langle \phi_n, \sigma^2 \rangle$  in model (4.2.1), having the same asymptotic risk. Due to (2.6.1), the experiments generated by observing (4.2.4), (4.2.1) and (4.2.2) are pairwise asymptotically equivalent under the assumptions of Lemma 4 and provided  $\sigma$ ,  $\tau$  are bounded from below and above.

However the result above is limited to the particular models assumed in Lemma 4. In order to obtain an estimator in either the Gaussian Volterra or a stochastic volatility model, we still have to verify by hand that the integrated volatility of the new data vector  $Y(\phi) := (Y_{1,n}(\phi), \ldots, Y_{n,n}(\phi))$  yields a good estimator for  $\int \phi \sigma_s^2 ds$ .

In the preceding paragraphs, we have demonstrated that estimation of the scalar product  $\langle \phi, \sigma^2 \rangle$  can be reduced to estimation of the integrated volatility plus (in general) some additional technicalities.

Second step: In this step, we derive an estimator for the integrated volatility. Some notation is needed. First, let  $\mathbb{M}_{p,q}$ ,  $\mathbb{M}_p$  and  $\mathbb{D}_p$  denote the spaces of  $p \times q$  matrices,  $p \times p$  matrices and  $p \times p$  diagonal matrices over  $\mathbb{R}$ , respectively. Second, define  $\Delta Y :=$  $(\Delta Y_{1,n}, \ldots, \Delta Y_{n-1,n})^t$ , where  $\Delta Y_{i,n} := Y_{i+1,n} - Y_{i,n}$  is the forward difference operator. The matrix  $D := D_{n-1} \in \mathbb{M}_{n-1}$  is defined entrywise by  $(D_{n-1})_{i,j} := \sqrt{2/n} \sin(ij\pi/n)$ . Note that  $D = D^t$  is a discrete sine transform. Let us choose  $M = \lfloor cn^{1/2} \rfloor$  for c > 0and a density k on [0, 1], i.e.  $k : [0, 1] \to [0, \infty)$ ,  $\int_0^1 k(x) dx = 1$ . Finally, we define  $J_n := J_n(k) \in \mathbb{D}_{n-1}$  by

$$(J_n)_{i,j} = \begin{cases} \frac{n}{M} k(\frac{i}{M}) \delta_{i,j}, & \text{for} \quad 1 \le i, j \le M, \\ 0 & \text{otherwise.} \end{cases}$$
(4.2.5)

Then, our estimator of the integrated volatility is given by

$$\widehat{\langle 1, \sigma^2 \rangle} = (\Delta Y)^t D J_n D(\Delta Y) - \pi^2 c^2 \int_0^1 k(x) x^2 dx \left\langle 1, \tau^2 \right\rangle, \qquad (4.2.6)$$

where  $\langle 1, \tau^2 \rangle$  is the integrated noise level. If  $\tau$  is unknown this must be replaced by an estimator (see the third step). However, as it will become clear,  $\langle 1, \tau^2 \rangle$  can be estimated

with rate of convergence  $n^{-1/2}$ , whereas the optimal rate of convergence for  $\langle 1, \sigma^2 \rangle$  is  $n^{-1/4}$ . Since  $n^{1/4} \ll n^{1/2}$  we may, from an asymptotic point of view, assume that  $\tau$  is known.

Before we proceed with step three, some discussion is necessary.

**Explanation of (4.2.6):** Let us think of the simplest situation, namely  $\sigma, \tau > 0$  are constants and the  $\epsilon_{i,n}$  are i.i.d. standard normal. In this case  $\Delta Y$  is a centered Gaussian vector with covariance matrix

$$\operatorname{Cov}(\Delta Y) = \frac{\sigma^2}{n} I_{n-1} + \tau^2 A, \qquad (4.2.7)$$

where  $I_{n-1}$  is the  $(n-1) \times (n-1)$  identity matrix and the tridiagonal matrix  $A \in \mathbb{M}_{n-1}$  is given by

$$A := \begin{pmatrix} 2 & -1 & 0 & \dots & 0 \\ -1 & 2 & -1 & \ddots & \vdots \\ 0 & -1 & 2 & \ddots & 0 \\ \vdots & \ddots & \ddots & \ddots & -1 \\ 0 & \dots & 0 & -1 & 2 \end{pmatrix}.$$
 (4.2.8)

In order to find the eigenvalues of  $Cov(\Delta Y)$ , it suffices to study the diagonalization of A. In fact, we find

$$A = D\Lambda_{n-1}D,$$

where  $\Lambda_{n-1}$  is diagonal with entries

$$(\Lambda_{n-1})_{i,i} := \lambda_i := 4\sin^2\left(i\pi/(2n)\right) \sim \frac{i^2}{n^2}.$$
 (4.2.9)

This can be seen by different methods. On the one hand, we may observe that A is a discrete Laplace operator. Reformulating this leads to a second order difference equation that is explicitly solvable. On the other hand, it is well known that taking differences of a stationary process implies multiplication by  $4\sin^2(\pi/2)$  for the spectral densities, i.e.  $f_{\Delta\eta}(\lambda) = f_{\eta}(\lambda) 4\sin^2(\lambda\pi/2)$ , where  $f_{\eta}$  and  $f_{\Delta\eta}$  denote the spectral densities of  $\eta$  and  $\Delta\eta$ , respectively. Because of  $f_{\eta} = 1$  we might guess  $\lambda_i = 4\sin^2(i\pi/(2n))$ .

Now,

$$\operatorname{Cov}(D\Delta Y) = D\operatorname{Cov}(\Delta Y)D = \frac{\sigma^2}{n}I_{n-1} + \tau^2\Lambda_{n-1}$$

and since  $D\Delta Y$  is a Gaussian vector, the components are independent with mean zero and variance  $\frac{\sigma^2}{n} + \tau^2 \lambda_i$ . Since  $\lambda_i^2 \sim \frac{i^2}{n^2}$ , we may obtain an estimator of  $\sigma^2$  by averaging over the first squared observations. Clearly, if  $i \leq \sqrt{n}$ , then,  $i^2/n^2 \leq 1/n$  and hence the observations are informative with respect to estimation of  $\sigma^2$ . Therefore, we can use of the order of  $n^{1/2}$  observations for estimation of  $\sigma^2$ . Moreover, some bias correction is needed and it will become clear that  $\pi^2 c^2 \int_0^1 k(x) x^2 dx \ \tau^2$  is exactly the quantity we need to subtract (this is essentially Lemma A.2). Putting this together, we obtain (4.2.6), in a special form, of course. This reveals that if  $\sigma$  is constant, the estimator is well motivated. Later, we show that when  $\sigma$  is not constant, this yields also a rate-optimal estimator for the integrated volatility.

**Third step:** Now, we combine the first and second step. By the heuristics derived so far, we will obtain an estimator of  $\langle \phi, \sigma^2 \rangle$ ,  $\phi \ge 0$  by mapping

$$(Y, \sigma, \tau) \to (Y(\phi), \sqrt{\phi}\sigma, \sqrt{\phi}\tau).$$

Let  $\Delta Y(\phi) := (\Delta_{1,n} Y(\phi), \dots, \Delta_{n-1,n} Y(\phi))^t$ , where

$$\Delta_{i,n} Y(\phi) := Y_{i+1,n}(\phi) - Y_{i,n}(\phi) = \sqrt{\phi(\frac{i}{n})(Y_{i+1,n} - Y_{i,n})}, \quad i = 1, \dots, n-1.$$

This allows us to extend (4.2.6) to

$$\widehat{\langle\phi,\sigma^2\rangle} = \left(\Delta Y(\phi)\right)^t DJ_n D^t \left(\Delta Y(\phi)\right) - \pi^2 c^2 \int_0^1 k(x) x^2 dx \left\langle\phi,\tau^2\right\rangle.$$
(4.2.10)

Now, let us give an estimator for  $\langle \phi, \tau^2 \rangle$ . Note that

$$\mathbb{E}[(\Delta_{i,n}Y)^2] = \tau_{(i+1)/n}^2 + \tau_{i/n}^2 + O(1/n).$$
(4.2.11)

Therefore,

$$\widehat{\langle \phi, \tau^2 \rangle} = \frac{1}{2(n-1)} \sum_{i=1}^{n-1} \phi(\frac{i}{n}) (\Delta_{i,n} Y)^2$$
(4.2.12)

provides us with a natural estimator for  $\langle \phi, \tau^2 \rangle$ . Next we introduce the assumption for the density k.

Assumption 3. The function  $k : [0,1] \to [0,\infty)$  has integral one, i.e.  $\int_0^1 k(x) dx = 1$ and k is piecewise Lipschitz continuous (with a finite number of pieces). Furthermore  $\sum_{i=0}^{\infty} |k_p| < \infty$ , with  $k_p := \int_0^1 k(x) \cos(p\pi x) dx$ .

In order to bound the moments of the estimators uniformly over a class of basis functions, growing for increasing n, we assume that  $\phi = \phi_n$  is in the following function space.

**Definition 15.** Given a constant  $C < \infty$ . Let  $\Phi_n(\kappa, C)$  be the set of functions  $\phi_n$ ,  $\phi_n : [0,1] \to [0,\infty)$  satisfying

(i)  $\sup_n \|\phi_n\|_{\infty} \le C$ ,

(*ii*)  $\sup_{n} \sup_{s,t: |s-t| \le 1/n} n^{5/8} |\phi_n^{1/2}(s) - \phi_n^{1/2}(t)| \le C$ ,

(*iii*) 
$$\sup_n (n^{-\kappa} \sum_{p=0}^{\infty} |(\phi_n)_p| + n^{1/4} \sum_{p=n}^{\infty} |(\phi_n)_p|) \le C,$$

where  $(\phi_n)_p := \int_0^1 \phi_n(x) \cos(p\pi x) dx.$ 

Before we can give the main lemma of this section, we must first introduce the function spaces for  $\sigma$  and  $\tau$ .

**Definition 16.** Given a finite constant  $Q_1$ . Let  $S(\kappa, Q_1)$  be the set of functions  $\sigma : [0,1]^2 \to [0,\infty)$  satisfying

- (i)  $\|\sigma\|_{\infty} \lesssim Q_1$ ,
- (*ii*)  $|\sigma(s,t) \sigma(s',t)| \le Q_1 |s-s'|^{1/4}, \ \forall \ t \in [0,1],$
- (*iii*)  $|\sigma(s,t) \sigma(s,t')| \le Q_1 |t t'|^{7/8}, \ \forall \ s \le t \land t',$
- $(iv) \ (s\mapsto \sigma^2(s,s))\in \Theta_{\cos}(3/4+\kappa,Q_1),$

**Definition 17.** Given a finite constant  $Q_2$ . Let  $T(\kappa, Q_2)$  be the set of functions  $\tau : [0,1] \to [0,\infty)$  satisfying

(i)  $\|\tau\|_{\infty} \le Q_2,$ (ii)  $|\tau(s) - \tau(t)| \le Q_2 |s - t|^{3/4},$ 

(*iii*) 
$$\tau^2 \in \Theta_{\cos}(3/4 + \kappa, Q_2).$$

In the following proposition, we show rates of convergence for the estimator of  $\langle \phi, \tau^2 \rangle = \int \phi \tau^2$ . In the following the notation  $\sigma \in S(\kappa, Q_1)$  means that  $(s, t) \mapsto \sigma_{s,t}$ , viewed as a function, lies in  $S(\kappa, Q_1)$ .

**Proposition 1.** Given model (1.1.2) and let  $\langle \widehat{\phi_n, \tau^2} \rangle$  be defined as in (4.2.12). Suppose that Assumptions 2 and 3 hold. Then, for  $0 \le \kappa \le 1/4$ ,

$$\sup_{\phi_n \in \Phi_n(\kappa,C), \ \sigma \in S(\kappa,Q_1), \ \tau \in T(\kappa,Q_2)} \left| \mathbb{E} \left[ \langle \widehat{\phi_n, \tau^2} \rangle \right] - \left\langle \phi_n, \tau^2 \right\rangle \right| \lesssim n^{-3/4}, \tag{4.2.13}$$

$$\sup_{\phi_n \in \Phi_n(\kappa, C), \ \sigma \in S(\kappa, Q_1), \ \tau \in T(\kappa, Q_2)} \operatorname{Var}\left(\langle \widehat{\phi_n, \tau^2} \rangle\right) \lesssim n^{-1}.$$
(4.2.14)

*Proof.* Let us prove, as a first step, the estimate for the bias. We have

$$\mathbb{E}\left[\langle \widehat{\phi_n, \tau^2} \rangle\right] = \frac{1}{2(n-1)} \sum_{i=1}^{n-1} \phi_n(\frac{i}{n}) \mathbb{E}\left[(\Delta_{i,n} Y)^2\right] \\ = \frac{1}{2(n-1)} \sum_{i=1}^{n-1} \phi_n(\frac{i}{n}) \mathbb{E}\left[(\Delta_{i,n} X)^2\right] + \frac{1}{2(n-1)} \sum_{i=1}^{n-1} \phi_n(\frac{i}{n}) \left(\tau^2(\frac{i}{n}) + \tau^2(\frac{i+1}{n})\right),$$

where  $\Delta_{i,n} X := X_{(i+1)/n} - X_{i/n}$ . Using

$$X_{(i+1)/n} - X_{i/n} = \int_{i/n}^{(i+1)/n} \sigma_{s,(i+1)/n} dW_s + \int_0^{i/n} (\sigma_{s,(i+1)/n} - \sigma_{s,i/n}) dW_s, \qquad (4.2.15)$$

we derive

$$\mathbb{E}\left[(\Delta_{i,n}X)^2\right] = \int_{i/n}^{(i+1)/n} \sigma_{s,(i+1)/n}^2 ds + \int_0^{i/n} (\sigma_{s,(i+1)/n} - \sigma_{s,i/n})^2 ds \lesssim n^{-1}$$

uniformly over  $S(\kappa, Q_1)$ . Since  $|\tau^2(\frac{i+1}{n}) - \tau^2(\frac{i}{n})| \leq 2Q_2^2 n^{-3/4}$  and by the boundedness of  $\phi_n$  the first equality (4.2.13) follows.

In order to bound the variance, let us write  $\Delta_{i,n}(\tau \eta) := \tau(\frac{i+1}{n})\eta_{i+1,n} - \tau(\frac{i}{n})\eta_{i,n}$ . Then uniformly over  $\Phi_n(\kappa, C)$ ,

$$\operatorname{Var}\left(\langle \widehat{\phi_n, \tau^2} \rangle\right) \lesssim n^{-2} \sum_{i,j=1}^{n-1} \big| \operatorname{Cov}((\Delta_{i,n}Y)^2, (\Delta_{j,n}Y)^2) \big|.$$

Obviously,

$$Cov((\Delta_{i,n}Y)^2, (\Delta_{j,n}Y)^2) = Cov((\Delta_{i,n}X)^2, (\Delta_{j,n}X)^2) + 4 Cov((\Delta_{i,n}X)(\Delta_{i,n}(\tau\eta)), (\Delta_{j,n}X)(\Delta_{j,n}(\tau\eta))) + Cov((\Delta_{i,n}(\tau\eta))^2, (\Delta_{j,n}(\tau\eta))^2).$$

Moreover, for two Gaussian random variables U, V we have  $Cov(U^2, V^2) = 2(Cov(U, V))^2$ . Hence, by using (4.2.15) again it follows

$$\left|\operatorname{Cov}((\Delta_{i,n}X)^2, (\Delta_{j,n}X)^2)\right| = 2\left(\operatorname{Cov}((\Delta_{i,n}X), (\Delta_{j,n}X))\right)^2 \lesssim n^{-2},$$

uniformly over  $S(\kappa, Q_1)$ . Similarly, we obtain

$$\sup_{\sigma \in S(\kappa,Q_1), \ \tau \in T(\kappa,Q_2)} \left| \operatorname{Cov}((\Delta_{i,n}X)(\Delta_{i,n}(\tau\eta)), (\Delta_{j,n}X)(\Delta_{j,n}(\tau\eta))) \right| \lesssim n^{-1}$$

and finally,  $|\operatorname{Cov}((\Delta_{i,n}(\tau\eta))^2, (\Delta_{j,n}(\tau\eta))^2)|$  is zero if  $|i - k| \geq 2$  and can otherwise be bounded uniformly by a finite constant. Combining the results above yields the bound on the variance.

**Proposition 2.** Given model (1.1.2) and let  $\langle \phi_n, \sigma^2 \rangle$  be defined as in (4.2.10). Suppose that Assumptions 2 and 3 hold. Then, for  $0 \le \kappa \le 1/4$ ,

$$\sup_{\phi_n \in \Phi_n(\kappa, C), \ \sigma \in S(\kappa, Q_1), \ \tau \in T(\kappa, Q_2)} \left| \mathbb{E} \left[ \langle \widehat{\phi_n, \sigma^2} \rangle \right] - \left\langle \phi_n, \sigma^2 \right\rangle \right| \lesssim n^{-1/4}, \tag{4.2.16}$$

$$\sup_{\phi_n \in \Phi_n(\kappa, C), \ \sigma \in S(\kappa, Q_1), \ \tau \in T(\kappa, Q_2)} \operatorname{Var}\left(\langle \widehat{\phi_n, \sigma^2} \rangle\right) \lesssim n^{-1/2}.$$
(4.2.17)

This lemma can be proven also in the case  $\sigma_{s,t} = \sigma_s$  and  $\tau_{i/n} = \tau(\Delta_{i-1,n}X, i/n)$  with obvious modifications of the proof. Note that under these assumptions  $(\tau_{i/n})_{i=1,\dots,n}$  is still a sequence of independent random variables, while the noise, itself, depends on the price process.

Moreover, under additional technicalities, we can include the case that X is a Brownian Bridge, i.e.  $\sigma_{s,t} = (1-t)/(1-s)$  (cf. Example 2).

*Proof.* We must first introduce the notation and technical preliminaries which appear later. In particular, if it is more convenient, we write  $\sigma(s)$  for  $\sigma_{s,s}$ .

We define the decomposition

$$\Delta Y(\phi_n) := X_1(\phi_n) + X_2(\phi_n) + Z_1(\phi_n) + Z_2(\phi_n) + Z_3(\phi_n),$$

where  $X_1(\phi_n), X_2(\phi_n), Z_1(\phi_n), Z_2(\phi_n)$  and  $Z_3(\phi_n)$  are n-1 dimensional random vectors with components

$$\begin{aligned} &(X_1(\phi_n))_i &:= (\phi_n^{1/2}\sigma)(\frac{i}{n}) \ \Delta_{i,n}W, \\ &(X_2(\phi_n))_i &:= (\phi_n^{1/2}\tau)(\frac{i}{n}) \ \Delta_{i,n}\eta, \\ &(Z_1(\phi_n))_i &:= \phi_n^{1/2}(\frac{i}{n}) \ \int_{i/n}^{(i+1)/n} (\sigma_{s,i/n} - \sigma_{i/n,i/n}) dW_s, \\ &(Z_2(\phi_n))_i &:= \phi_n^{1/2}(\frac{i}{n}) \ \int_0^{(i+1)/n} (\sigma_{s,(i+1)/n} - \sigma_{s,i/n}) dW_s, \\ &(Z_3(\phi_n))_i &:= \phi_n^{1/2}(\frac{i}{n}) \ (\Delta_{i,n}\tau) \ \eta_{i+1,n}, \qquad i = 1, \dots, n-1. \end{aligned}$$

For a function  $f \in L^2$  and  $p \in \mathbb{Z}$  let

$$f_p := \int_0^1 f(x) \cos(p\pi x) dx$$
 (4.2.18)

be the (scaled) p-th Fourier coefficients with respect to cosine basis. Furthermore, we define the sums A(f,r) by

$$A(f,r) = \sum_{q \in \mathbb{Z}, \ q \equiv r \bmod 2n} f_q.$$

$$(4.2.19)$$

Some properties of these variables are given in Lemma A.3. Let  $I_n(f) \in \mathbb{D}_{n-1}$  be defined as

$$I_n(f) := \begin{pmatrix} f(1/n) & & \\ & \ddots & \\ & & f(1-1/n) \end{pmatrix}.$$
 (4.2.20)

Whenever it is obvious, we will drop the index n.

For two centered random vectors P and Q

$$\langle P, Q \rangle_{\sigma} := \mathbb{E} \left[ P^t D J_n D Q \right]$$

defines a semi-inner product, i.e. a scalar product, where  $\langle P, Q \rangle_{\sigma} = 0$  does not necessarily imply that P = 0. For column vectors X, Y, of length  $m_X$  and  $m_Y$ , the covariance of Xand Y is defined as the matrix  $\operatorname{Cov}(X, Y) \in \mathbb{M}_{m_X, m_Y}$  with  $(\operatorname{Cov}(X, Y))_{i,j} := \operatorname{Cov}(X_i, Y_j)$ . Now, Lemma A.8 shows that  $\operatorname{Cov}(P, Q) = 0 \Rightarrow \langle P, Q \rangle_{\sigma} = 0$ .

Moreover,  $\operatorname{Cov}(X_1(\phi_n), Z_3(\phi_n)) = \operatorname{Cov}(X_2(\phi_n), Z_1(\phi_n)) = \operatorname{Cov}(X_2(\phi_n), Z_3(\phi_n)) = 0.$ Hence, uniformly over  $\phi_n \in \Phi_n(\kappa, C), \ \sigma \in S(\kappa, Q_1), \ \tau \in T(\kappa, Q_2),$ 

$$\mathbb{E}\left[\langle \phi_{n}, \sigma^{2} \rangle\right] = \langle X_{1}(\phi_{n}), X_{1}(\phi_{n}) \rangle_{\sigma} + \langle X_{2}(\phi_{n}), X_{2}(\phi_{n}) \rangle_{\sigma} + \langle Z_{1}(\phi_{n}), Z_{1}(\phi_{n}) \rangle_{\sigma} \\
+ \langle Z_{2}(\phi_{n}), Z_{2}(\phi_{n}) \rangle_{\sigma} + \langle Z_{3}(\phi_{n}), Z_{3}(\phi_{n}) \rangle_{\sigma} + 2 \langle X_{1}(\phi_{n}), Z_{1}(\phi_{n}) \rangle_{\sigma} \\
+ 2 \langle X_{1}(\phi_{n}), Z_{2}(\phi_{n}) \rangle_{\sigma} + 2 \langle X_{2}(\phi_{n}), Z_{3}(\phi_{n}) \rangle_{\sigma} \\
+ 2 \langle Z_{1}(\phi_{n}), Z_{2}(\phi_{n}) \rangle_{\sigma} - \pi^{2} c^{2} \int_{0}^{1} k(x) x^{2} dx \langle \phi_{n}, \tau^{2} \rangle + O(n^{-3/4}). \quad (4.2.21)$$

The remaining part of the proof is concerned with approximating/bounding the terms of the r.h.s. of (4.2.21).

 $\langle \mathbf{X_1}(\phi_{\mathbf{n}}), \mathbf{X_1}(\phi_{\mathbf{n}}) \rangle_{\sigma}$ : We easily see that  $\mathbb{E}[(X_1(\phi_n))_i] = 0$  and

$$\mathbb{E}[(X_1(\phi_n))_i (X_1(\phi_n))_j] = \frac{1}{n} (\phi_n \sigma^2)(\frac{i}{n}) \delta_{i,j},$$

where  $\delta_{i,j}$  denotes the Kronecker delta. Hence, we obtain

$$\langle X_1(\phi_n), X_1(\phi_n) \rangle_{\sigma} = \frac{1}{n} \operatorname{tr}(DJ_n DI_n(\phi_n \sigma^2)),$$

where  $I_n(\phi_n \sigma^2)$  is as defined in (4.2.20). By Lemma A.3 (ii) and with  $r_n := \frac{1}{M} \sum_{i=1}^M k(\frac{i}{M}) - 1$ ,

$$\langle X_1(\phi_n), X_1(\phi_n) \rangle_{\sigma} = \frac{1}{n} \operatorname{tr}(J_n D I_n(\phi_n \sigma^2) D)$$
  
=  $\frac{1}{M} \sum_{i=1}^M k(\frac{i}{M}) \left( A \left( \phi_n \sigma^2, 0 \right) - A \left( \phi_n \sigma^2, 2i \right) \right)$   
=  $(1 + r_n) A \left( \phi_n \sigma^2, 0 \right) - \frac{1}{M} \sum_{i=1}^M k(\frac{i}{M}) A \left( \phi_n \sigma^2, 2i \right)$ 

Since by Assumption 3,  $r_n \lesssim n^{-1/2}$ 

$$\left| \langle X_1(\phi_n), X_1(\phi_n) \rangle_{\sigma} - (\phi_n \sigma^2)_0 \right| \lesssim \sum_{m=n}^{\infty} \left| (\phi_n \sigma^2)_m \right| + \frac{1}{\sqrt{n}} \sum_{i=0}^{\infty} \left| (\phi_n \sigma^2)_i \right|,$$

where  $(\phi_n \sigma^2)_p := \int_0^1 \phi_n(x) \sigma^2(x) \cos(p\pi x) dx$  in accordance with (4.2.18). Further, we define  $s_p := (1 \cdot \sigma^2)_p$  and  $(\phi_n)_p := (\phi_n \cdot 1)_p$ . By using Lemmas A.4 and A.5, we obtain

$$\sum_{m=n}^{\infty} |(\phi_n \sigma^2)_m| \le \sum_{m=n}^{\infty} \sum_{l=-\infty}^{\infty} |s_l| |(\phi_n)_{l-m}| \le \sum_{l=-\infty}^{\lfloor n/2 \rfloor} |s_l| \sum_{m=n}^{\infty} |(\phi_n)_{l-m}| + \sum_{l=\lfloor n/2 \rfloor}^{\infty} |s_l| \sum_{m=-\infty}^{\infty} |(\phi_n)_{l-m}|, \le n^{-1/4} + n^{1/2 - 3/4 - \kappa} n^{\kappa} \le n^{-1/4},$$
(4.2.22)

$$\sum_{i=0}^{\infty} \left| (\phi_n \sigma^2)_i \right| \le \sum_{l=-\infty}^{\infty} |s_l| \sum_{m=-\infty}^{\infty} |(\phi_n)_{l-m}| \le n^{\kappa} \le n^{1/4}, \tag{4.2.23}$$

uniformly over  $\phi_n \in \Phi_n(\kappa, C)$  and  $\sigma \in S(\kappa, Q_1)$  due to  $\kappa \leq 1/4$ . This yields

$$\sup_{\phi_n \in \Phi_n(\kappa,C), \ \sigma \in S(\kappa,Q_1)} \left| \langle X_1(\phi_n), X_1(\phi_n) \rangle_{\sigma} - \int_0^1 \phi_n(x) \sigma^2(x) dx \right| \lesssim n^{-1/4}.$$

The remaining estimates for the bias as well as the uniform bound on the variance (4.2.17) are proven in Appendix A.

#### 4.3. Fourier series estimator of the spot volatility

In this section we define the spot volatility estimator and provide proofs for the rates of convergence.

Based on the previous result regarding the estimation of scalar products, the final step in order to derive a series estimator is to expand the function  $\sigma^2$  as in (2.1.1). Given an  $L^2$ -basis  $(\phi_i)_i$  and weights  $(\omega_{i,n})_i$  our estimator for the spot volatility is defined via

$$\widehat{\sigma}^2(t) = \sum_{i=0}^{\infty} \omega_{i,n} \langle \widehat{\phi_i, \sigma^2} \rangle \phi_i.$$
(4.3.1)

The upper bound with respect to the integrated mean square error (IMSE) follows from Theorem 1. Let us derive rates of convergence explicitly by considering examples of orthogonal basis systems.

**Example:** Cosine basis. In this example we apply Theorem 1 to the cosine basis  $(\phi_i)_i$  as defined in (2.4.3). Note that  $1 + \cos(y) = 2\cos^2(y/2)$ . Therefore, and according to Definition 15, the functions

$$\psi_{i_n}(\cdot) := \cos^2(\frac{1}{2}i_n\pi\cdot) \tag{4.3.2}$$

belong to  $\Phi_n(0, C)$  whenever  $i_n \leq n^{3/8}$  for sufficiently large C. Obviously,

$$\widehat{\langle \phi_0, \sigma^2 \rangle} := \widehat{\langle \psi_0, \sigma^2 \rangle}, \qquad \widehat{\langle \phi_i, \sigma^2 \rangle} := \sqrt{2} \left( 2 \widehat{\langle \psi_i, \sigma^2 \rangle} - \widehat{\langle \psi_0, \sigma^2 \rangle} \right), \quad i > 0$$

are estimators of the basis coefficients  $\langle \phi_i, \sigma^2 \rangle$ ,  $i \ge 0$ , satisfying (2.5.3) with  $q_n \sim n^{1/4}$ . Assume that  $(s \mapsto \sigma_{s,s}^2) \in \Theta_{\cos}(\alpha, Q_1)$  and  $\sigma \in S(0, Q_1)$  for  $\alpha \ge 3/4$  and that one of the weight sequences  $(\omega_{i,n}^{(1)})_i, (\omega_{i,n}^{(2)})_i$ ,

$$\omega_{i,n}^{(1)} := \mathbb{I}_{\{i \le c_{\omega} n^{1/(4\alpha+2)}\}}, \quad \omega_{i,n}^{(2)} := \left(1 - c_{\omega}^{-\alpha} n^{-\alpha/(4\alpha+2)} i^{\alpha}\right)_{+}, \quad 0 < c_{\omega} < \infty.$$
(4.3.3)

is used. Then we obtain for  $\kappa = 0$ , as a consequence of Theorem 1.

**Theorem 3.** Assume model (1.1.2) and let  $\hat{\sigma}^2$  be defined as in (4.3.1). Under the assumption of Proposition 2

$$\sup_{(s\mapsto\sigma_{s,s}^2)\in\Theta_{\cos}(\alpha,Q_1),\ \sigma\in S(0,Q_1),\ \tau\in T(0,Q_2)} \text{IMSE}(\widehat{\sigma}^2) \lesssim n^{-\alpha/(2\alpha+1)}.$$
(4.3.4)

*Proof.* We apply Theorem 1 for  $q_n := \lfloor n^{1/4} \rfloor$ . First note that  $\omega_{i,n}^{(2)} \leq \omega_{i,n}^{(1)}$  for i = 0, 1, ... and hence  $\sum_{i=0}^{\lfloor n^{1/4} \rfloor} (\omega_{i,n}^{(p)})^2 \lesssim n^{1/(2\alpha+1)}$ , p = 1, 2. For the second term, we obtain

$$\begin{split} \sum_{i=0}^{\infty} (1 - \omega_{i,n}^{(2)})^2 \langle \phi_i, \sigma^2 \rangle^2 &= \sum_{i=0}^{\lfloor c_{\omega} n^{1/(4\alpha+2)} \rfloor} c_{\omega}^{-2\alpha} n^{-\alpha/(2\alpha+1)} i^{2\alpha} \langle \phi_i, \sigma^2 \rangle^2 + \sum_{\lfloor c_{\omega} n^{1/(4\alpha+2)} \rfloor + 1}^{\infty} \langle \phi_i, \sigma^2 \rangle^2 \\ &\lesssim n^{-\alpha/(2\alpha+1)} + (c_{\omega} n^{1/(4\alpha+2)})^{-2\alpha} \sum_{i=\lfloor c_{\omega} n^{1/(4\alpha+2)} \rfloor + 1}^{\infty} i^{2\alpha} \langle \phi_i, \sigma^2 \rangle^2 \\ &\lesssim n^{-\alpha/(2\alpha+1)}, \end{split}$$

uniformly over  $(s \mapsto \sigma_{s,s}^2) \in \Theta_{\cos}(\alpha, Q_1)$ . In the same spirit  $\sum_{i=0}^{\infty} (1 - \omega_{i,n}^{(1)})^2 \langle \phi_i, \sigma^2 \rangle^2 \lesssim n^{-\alpha/(2\alpha+1)}$  can be shown as well. This completes the proof.  $\Box$ 

The function space  $\{\sigma : (s \mapsto \sigma_{s,s}^2) \in \Theta_{\cos}(\alpha, Q_1) \text{ and } \sigma \in S(0, Q_1)\}$  can be written in a different form. Clearly, a function belongs to this space if and only if

$$(s \mapsto \sigma_{s,s}^2) \in \Theta_{\cos}(\alpha, Q_1), \quad |\sigma_{s,u} - \sigma_{s',u}| \le Q_1 |s - s'|^{1/4}, \quad |\sigma_{s,u} - \sigma_{s,u'}| \le Q_1 |u - u'|^{7/8}.$$

**Example: Trigonometric Basis.** For this example let  $(\phi_i)_i$  be the trigonometric basis defined in (2.1.2) and let  $\psi_{i_n}$  be as in (4.3.2). Moreover, introduce  $\tilde{\psi}_{i_n}(.) = 1 + \sin(2i_n\pi \cdot)$ . By integral calculus we obtain

$$(\tilde{\psi}_{i_n})_p = \int_0^1 (1 + \sin(2i_n\pi x))\cos(p\pi x)dx = \begin{cases} 0 & \text{for } l \text{ even,} \\ (4i_n)/(\pi[(2i_n)^2 - l^2]) & \text{for } l \text{ odd.} \end{cases}$$

and using Riemann sums for the second term

$$\sum_{p=0}^{\infty} |(\widetilde{\psi}_{i_n})_p| \lesssim i_n, \quad \sum_{p=n}^{\infty} |(\widetilde{\psi}_{i_n})_p| \lesssim n^{-3/4} i_n,$$

provided  $i_n \leq n/2$ . Moreover,

$$|(1 + \sin(2i_n\pi x))^{1/2}| = |\sin(i_n\pi x) + \cos(i_n\pi x)| = \sqrt{2}|\sin(i_n\pi x + \pi/4)|.$$

Hence, for  $i_n \leq n^{1/4}$ ,  $\tilde{\psi}_{i_n}$  belongs to  $\Phi_n(1/4, C)$ . Recall from the previous example that for  $i_n \leq n^{1/4}$ ,  $\psi_{i,n}$  is in  $\Phi_n(0, C) \subset \Phi_n(1/4, C)$ . Now, we define

$$\begin{split} &\langle \widehat{\phi_{0}, \sigma^{2}} \rangle := \langle \widehat{\psi_{0}, \sigma^{2}} \rangle, \\ &\langle \widehat{\phi_{2i}, \sigma^{2}} \rangle := \sqrt{2} \left( 2 \langle \widehat{\psi_{2i}, \sigma^{2}} \rangle - \langle \widehat{\psi_{0}, \sigma^{2}} \rangle \right), \quad i > 0, \\ &\langle \widehat{\phi_{2i+1}, \sigma^{2}} \rangle := \sqrt{2} \left( \langle \widetilde{\widetilde{\psi_{i}, \sigma^{2}}} \rangle - \langle \widehat{\psi_{0}, \sigma^{2}} \rangle \right), \quad i > 0 \end{split}$$

as the estimators of the corresponding basis coefficients  $\langle \phi_i, \sigma^2 \rangle$ . They clearly satisfy (2.5.3) with  $q_n \sim n^{1/4}$ . Now, let the weights be given as in (4.3.3), then we can derive rates of convergence by following the lines of the proof of Theorem 3.

**Theorem 4.** Assume model (1.1.2) and let  $\hat{\sigma}^2$  be defined as in (4.3.1),  $\alpha \geq 1$  and  $\kappa = 1/4$ . Under the assumption of Proposition 2

$$\sup_{(s\mapsto\sigma_{s,s}^2)\in\Theta_{\text{trig}}(\alpha,Q_1),\ \sigma\in S(1/4,Q_1),\ \tau\in T(1/4,Q_2)} \text{IMSE}(\widehat{\sigma}^2) \lesssim n^{-\alpha/(2\alpha+1)}.$$
(4.3.5)

By using (2.4.6) we obtain for  $\alpha \geq 1$ ,  $(s \mapsto \sigma_{s,s}^2) \in \Theta_{\text{trig}}(\alpha, Q_1)$  and  $\sigma \in S(1/4, Q_1)$  if and only if

$$(s \mapsto \sigma_{s,s}^2) \in \Theta_{\text{trig}}(\alpha, Q_1), \quad |\sigma_{s,u} - \sigma_{s',u}| \le Q_1 |s - s'|^{1/4}, \quad |\sigma_{s,u} - \sigma_{s,u'}| \le Q_1 |u - u'|^{7/8}.$$

This demonstrates that the spot volatility estimator with respect to the trigonometric basis has the (optimal)  $n^{-\alpha/(4\alpha+2)}$  rate of convergence over the Sobolev ellipsoid  $\Theta_{\text{trig}}(\alpha, Q_1)$ , as long as the coordinate mappings satisfy some minimal Lipschitz conditions.

By transferring the explicitly outlined example of the trigonometric basis to other basis systems which are 'close' to the cosine basis (for instance the sine basis), it is clear that similar results do apply.

#### 4.4. Optimizing tuning parameters

For the purpose of implementation, it is important to know how the function k, defined in (4.2.5) and  $M = \lfloor cn^{1/2} \rfloor$  can be chosen in a (theoretically) optimal way. There is no general answer for this problem, so far. Here, we will treat the simplified version, namely to ask for the optimal k and c for estimation of  $\langle 1, \sigma^2 \rangle$ , provided  $\sigma$ ,  $\tau$  are deterministic constants and  $\eta \sim \mathcal{N}(0, I_n)$ . As mentioned earlier, we may assume, without loss of generality, that  $\tau$  is known. Recall the definition of mean square error, given in (2.5.2). In this setting, it is well known that the optimal achievable mean square error behaves asymptotically as  $8\tau\sigma^3n^{-1/2}(1+o(1))$  (cf. Gloter and Jacod [33, 34] and Cai et al. [16]).

**Lemma 5.** Suppose that the assumptions above hold true. Let  $\widehat{\langle 1, \sigma^2 \rangle}$  be as defined in (4.2.6). Then

$$MSE(\widehat{\langle 1, \sigma^2 \rangle}) = \frac{2}{M} \int_0^1 k^2(x) (\sigma^2 + c^2 \pi^2 \tau^2 x^2)^2 dx \ (1 + o(1)).$$

In particular for fixed c, the MSE-minimizing k, denoted by  $k^*$ , is given by

$$k^{\star}(x) = C(\sigma, \tau, c)^{-1} \frac{1}{(\sigma^2 + c^2 \pi^2 \tau^2 x^2)^2}$$

where

$$C(\sigma,\tau,c) := \frac{1}{2\sigma^2(\sigma^2 + c^2\pi^2\tau^2)} + \frac{\arctan(\frac{\pi c\tau}{\sigma})}{2\sigma^3\tau c\pi}$$

For this choice we obtain

$$MSE(\widehat{\langle 1, \sigma^2 \rangle}) = \frac{2}{M} C(\sigma, \tau, c)^{-1} \ (1 + o(1)).$$
(4.4.1)

*Proof.* By Lemma A.8 and (4.2.7) we obtain for the variance

$$\mathbb{E}\left[\widehat{\langle 1,\sigma^2 \rangle}\right] = \mathbb{E}\left[(\Delta Y)^t D J_n D(\Delta Y)\right] - \pi^2 c^2 \int_0^1 k(x) x^2 dx \ \tau^2$$
$$= \operatorname{tr}(D J_n D \operatorname{Cov}(\Delta Y)) - \pi^2 c^2 \int_0^1 k(x) x^2 dx \ \tau^2$$
$$= \frac{\sigma^2}{n} \operatorname{tr}(J_n) + \tau^2 \operatorname{tr}(J_n \Lambda) - \pi^2 c^2 \int_0^1 k(x) x^2 dx \ \tau^2$$

Hence by using Lemma A.2 (i), it follows  $\mathbb{E}[\langle \widehat{1, \sigma^2} \rangle] \lesssim n^{-1/2}$ . For the variance, we may use Lemma A.9 (ii) since  $\Delta Y$  is Gaussian; thus, by using (4.2.7) again

$$\operatorname{Var}\left(\langle 1, \sigma^2 \rangle\right) = \operatorname{Var}\left((\Delta Y)^t D J_n D(\Delta Y)\right) = 2 \|\operatorname{Cov}(\Delta Y)^{1/2} D J_n D \operatorname{Cov}(\Delta Y)^{1/2} \|_2^2$$
$$= 2 \|J_n^{1/2} D \operatorname{Cov}(\Delta Y) D J_n^{1/2}\|_2^2 = 2 \|J_n^{1/2} (\frac{\sigma^2}{n} I_{n-1} + \tau^2 \Lambda) J_n^{1/2}\|_2^2$$
$$= 2 \sum_{i=1}^M \left(\frac{\sigma^2}{n} + \tau^2 \lambda_i\right)^2 \frac{n^2}{M^2} k^2(\frac{i}{M}).$$

Now by applying Lemma A.2 (ii)-(iv) the first part follows.

In order to derive the representation of k, note that the antiderivative of  $x \mapsto 1/(a + bx^2)^2$ ,  $a, b \in (0, \infty)$  is given by

$$x \mapsto \frac{x}{2a(a+bx^2)} + \frac{\arctan(\sqrt{ba^{-1}}x)}{2a^{3/2}b^{1/2}} + C,$$

where C is a constant. Now, by using Lagrange calculus, we see that  $\int_0^1 k^2(x)(\sigma^2 + c^2\pi^2\tau^2x^2)^2dx$  is minimized under the constraint  $\int_0^1 k(x)dx = 1$ , if k solves  $2k(x)(\sigma^2 + c^2\pi^2\tau^2x^2)^2 - \lambda = 0$ , or

$$k(x) = \frac{\lambda}{2(\sigma^2 + c^2 \pi^2 \tau^2 x^2)^2}, \ x \in [0, 1].$$

By the integration formula above and some computations, the result follows.

Let us make the following two remarks: First,  $k^*$ , of course, depends on the unknown quantities themselves and is therefore not computable. However, as shown in Cai et *al.* [16] it is possible to estimate the function  $k^*$  by a splitting technique, but this is limited to the case when  $\sigma, \tau$  are deterministic constants. An extension to functions  $\sigma_{s,t} = \sigma_s$  has been derived in Reiß [71]. In this setting the optimal asymptotic variance with respect to MSE-risk, in the sense of Definition 10, is of the form  $8\tau \int_0^1 \sigma^3(s) ds \ n^{-1/2}(1 + o(1))$ . In the general Gaussian Volterra model, the optimal constant is still unknown.

Secondly, if we let  $c \to \infty$  then we obtain for the risk of the "choice"  $k = k^*$ , MSE( $\langle 1, \sigma^2 \rangle$ ) =  $8\tau\sigma^3 n^{-1/2}$  (1 + o(1)), which is asymptotically efficient, as mentioned above.

Although for our theoretical results, k and c need to be chosen as fixed and non-random, the considerations above provide insight for the choice of constants, in practice. This is particularly true if we have some prior knowledge on the size of  $\sigma$  and  $\tau$ .

#### 4.5. Comparison of estimators for integrated volatility

As noted in the introduction, other methods have been developed in order to estimate the integrated volatility. The most important are the multiscale realized volatility approach by Zhang [76], realised kernels (cf. Barndorff-Nielsen et *al.* [7]) as well as pre-averaging (cf. Podolskij and Vetter [68] and Jacod et *al.* [44]). In fact all of these methods are equivalent up to the point of handling boundary terms.

Therefore, we would like to compare the estimators for the scalar products, derived in this chapter, with one of the procedures mentioned above. Without loss of generality, let us choose the realised kernel estimator, defined in Barndorff-Nielsen et *al.* [7], Section 1.

Consider again the Gaussian Volterra model where  $\sigma, \tau$  are deterministic constants and  $\eta_{i,n} \sim \mathcal{N}(0,1)$ , i.i.d., assuming that the number of observation ranges over  $i = -M, -M + 1, \ldots, 0, 1, \ldots, n$ . For  $l \leq M$ , denote the *l*-th realised autocorrelation by

$$\gamma_l(Y) := \sum_{j=1}^n (\Delta_{j-1,n}Y)(\Delta_{j-l-1,n}Y).$$

Then the realised kernel estimator is defined via

$$\widehat{\langle 1, \sigma^2 \rangle}_{RK} := \gamma_0(Y) + \sum_{l=1}^M f\left(\frac{l-1}{M}\right) \left(\gamma_l(Y) + \gamma_{-l}(Y)\right), \tag{4.5.1}$$

where f is a sufficiently smooth function with f(0) = 1, f(1) = f'(0) = f'(1) = 0. Both estimators,  $\langle \widehat{1, \sigma^2} \rangle$  (as defined in (4.2.6)) and  $\langle \widehat{1, \sigma^2} \rangle_{RK}$  can be viewed as quadratic forms. By comparing them, we see that up to boundary and approximation terms (and of course different methods in order to subtract the bias, but this is of smaller order anyway), the estimator defined in (4.2.6) can be understood as the realised kernel estimator and the translation is given by

$$f(u) = \int_0^1 k(t) \cos(u\pi tc^2) dt,$$

with k defined as in (4.2.5). In particular, the condition  $\int_0^1 k(t)dt = 1$  is equivalent to f(0) = 1. Let us extend k to the real line by

$$\check{k}(x) := \begin{cases} k(x), & \text{for } x \in [0, 1], \\ 0, & \text{for } x > 1, \\ k(-x), & \text{for } x \le 0. \end{cases}$$

Further denote by  $\mathcal{F}$  the Fourier transform. Rewriting

$$f(u) = \int_0^1 k(t) \cos(u\pi tc^2) dt = \frac{1}{2} \mathcal{F}(\check{k}) \left(\frac{uc^2}{2}\right)$$

and by Parseval's identity, we derive further  $||f||_2 = c^{-1} ||k||_2$ ,  $||f'||_2^2 = c^2 \pi^2 \int_0^1 k^2(t) t^2 dt$ and  $||f''||_2^2 = c^6 \pi^4 \int_0^1 k^2(t) t^4 dt$ . Therefore, we see that the asymptotic variances derived in Lemma 5 and in Barndorff-Nielsen et *al.* [7], Theorem 4 coincide.

However, note that for a finite sample size, the estimators for the integrated volatility might be quite different. In particular, the fact that  $\langle 1, \sigma^2 \rangle_{RK}$  also includes observations outside the time interval [0, 1] makes the realised kernel estimator difficult to implement in practice.

In [64], the estimator (4.2.6) has been introduced in the special case  $k = 2\mathbb{I}_{[1/2,1]}(\cdot)$ . Let us show by an easy example that this can be improved in the special setting of Lemma 5. Note that for  $k = 2\mathbb{I}_{[1/2,1]}(\cdot)$ , we obtain the asymptotic variance

$$\left(2\sigma^4 + \frac{7}{3}\pi^2\tau^2\sigma^2 + \frac{31}{40}\pi^4\tau^4\right)n^{-1/2}(1+o(1)).$$

Now consider the uniform density over [0, 1], i.e.  $\overline{k} = \mathbb{I}_{[0,1]}(\cdot)$ . Then, under the same assumption, the asymptotic variance of the integrated volatility is

$$\left(2\sigma^4 + \frac{4}{3}\pi^2\tau^2\sigma^2 + \frac{2}{5}\pi^4\tau^4\right)n^{-1/2}(1+o(1)).$$

Therefore, we improve quite substantially over earlier versions, in particular, if  $\tau$  is large.

# Chapter 5

## Spot volatility estimation under microstructure noise in the semimartingale model: Wavelet adaptation

This chapter is devoted to the construction of an adaptive wavelet estimator. In the first part of the chapter we introduce and discuss the so-called pre-averaging technique. Based on this, an estimator for the single wavelet coefficients is given in Section 5.2. Out of this, we construct in Section 5.3 an estimator for the spot volatility and calculate upper bounds. The content of this chapter relies on [43].

#### 5.1. Pre-averaging

Suppose that we observe a process X with continuous sample paths under additional measurement noise, i.e.  $Y_{t_i} = X_{t_i} + \epsilon_{t_i}$  at deterministic time points  $t_i$ , i = 1, ..., n and assuming  $t_i \in [0, 1]$ , for simplicity. Now, let

$$[0,1] = \bigcup_{i=1}^{M} I_i,$$

define a finite partition of [0, 1] in M disjoint intervals, where M is allowed to depend on n. Consider the mean over interval  $I_i$ 

$$\operatorname{av}(Y)_{i} := \frac{1}{\#I_{i}} \sum_{t_{j} \in I_{i}} Y_{t_{j}} = \frac{1}{\#I_{i}} \sum_{t_{j} \in I_{i}} X_{t_{j}} + \frac{1}{\#I_{i}} \sum_{t_{j} \in I_{i}} \epsilon_{t_{j}}, \quad i = 1, \dots, M,$$
(5.1.1)

with  $\#I_i := \#\{t_j : t_j \in I_i\}$ . This binning has different effects on X and  $\epsilon$ . Assume that the sample paths of X are Hölder continuous with index  $\alpha$ . Then,

$$\frac{1}{\#I_i} \sum_{t_j \in I_i} X_{t_j} = X_{l(I_i)} + O_p(|I_i|^{\alpha}), \qquad (5.1.2)$$

where  $l(I_i)$  and  $|I_i|$  denote the left endpoint and length of the interval  $I_i$ , respectively. Hence, for X a LLN-type result applies. On the other hand, by imposing suitable conditions on the noise  $\epsilon$ , we obtain by CLT,

$$\frac{1}{\#I_i} \sum_{t_j \in I_i} \epsilon_{t_j} = O_p(\#I_i^{-1/2}).$$
(5.1.3)

If the equations above hold true, then  $\operatorname{av}(Y) := (\operatorname{av}(Y)_1, \ldots, \operatorname{av}(Y)_M)^t$  can be treated as a new observation vector, where we observe X at time points  $l(I_i)$  under additive noise. This noise is due to two sources, the approximation error in (5.1.2) and the averaged sum (5.1.3).

If X has independent increments, then for different intervals  $I_i$ , the approximation errors are independent. The second error source, due to averaging of measurement noise, will, by CLT, converge to a Gaussian random variable, with variance  $\sim (\#I_i)^{-1}$ , provided the number of observations falling into  $I_i$  tends to infinity. Therefore, we have a good understanding of the noise part of av(Y), in general.

More precisely, if the sample time points  $t_j$  are sufficiently uniformly distributed over the interval [0, 1] then  $\#I_i \sim |I_i|n$ . In this situation

$$\operatorname{av}(Y)_i = X_{l(I_i)} + O_p(|I_i|^{\alpha} + (n|I_i|)^{-1/2}), \quad i = 1, \dots, M.$$

Here, we are in a classical trade-off situation: by choosing  $|I_i|$  large, the (stochastic) bias term dominates, whereas for  $|I_i|$  small the effect of the averaged noise does. The optimal balance is obtained by choosing  $|I_i| \sim n^{-1/(2\alpha+1)}$ , or alternatively,  $M \sim n^{1/(2\alpha+1)}$ . In this case, we obtain

$$\operatorname{av}(Y)_i = X_{l(I_i)} + O_p(n^{-\alpha/(2\alpha+1)}), \quad i = 1, \dots, M \sim n^{1/(2\alpha+1)}.$$

Therefore, we may think of av(Y) as a vector of new observations, where the influence of the noise is reduced at the expense of a reduction in sample size.

To give an example, let  $\alpha = 1/2$  (this is Brownian smoothness, essentially). Then  $\operatorname{av}(Y)_i = X_{l(I_i)} + O_p(n^{-1/4}), \ i = 1, \ldots, n^{1/2}$ . So, the sample size is reduced from n to  $n^{1/2}$ , while the noise is now of order  $O_p(n^{-1/4})$ .

For applications, the interesting quantities are not  $\operatorname{av}(Y)_i$ , but the differences  $\operatorname{av}(Y)_i - \operatorname{av}(Y)_{i-1} = X_{l(I_i)} - X_{l(I_{i-1})} + O_p(n^{-\alpha/(2\alpha+1)})$ . The size of the "informative" increments  $X_{l(i)} - X_{l(i-1)}$  are of the same order as the noise, which is the typical situation faced in nonparametric regression for instance. The benefit in comparison to taking differences without averaging first, i.e.  $Y_{t_i} - Y_{t_{i-1}} = X_{t_i} - X_{t_{i-1}} + O_p(1)$ , is substantially. If the sample points are sufficiently uniformly distributed then  $X_{t_i} - X_{t_{i-1}} \sim O_p(n^{-\alpha/2})$  which is much smaller than the noise.

In contrast to nonparametric regression, the noise is not centered. For this reason, it will be necessary to do some bias correction.

The heuristics derived above provide us with a good motivation for the construction of the estimator, although it needs to be handled with special care. Let us illustrate this point with a concrete example from Gloter [29]. Suppose that  $\epsilon = 0$  and  $X_t = \sigma W_t$ . As it is well known (see Theorem B.1),  $\sum_{i=2}^{n} (X_{i/n} - X_{(i-1)/n})^2 \xrightarrow{P} \sigma^2$ . Assume a constant partition  $I_i = (\frac{i-1}{M}, \frac{i}{M}]$ . Therefore, we would expect, by the arguments derived above, that

$$\sum_{i=2}^{M} (\operatorname{av}(Y)_{i} - \operatorname{av}(Y)_{i-1})^{2} \approx \sum_{i=2}^{M} (X_{i/M} - X_{(i-1)/M})^{2} \xrightarrow{P} \sigma^{2}.$$

However, as one can easily show by straightforward calculations,

$$\sum_{i=2}^{M} (\operatorname{av}(Y)_{i} - \operatorname{av}(Y)_{i-1})^{2} \xrightarrow{P} \frac{2}{3}\sigma^{2},$$

whenever  $M \to \infty$  and  $M/n \to 0$ .

The most interesting fact that makes passing from Y to  $\operatorname{av}(Y)$  to a powerful statistical tool is that, typically, this does not result in an essential loss of information for estimation of a parameter related to X (for volatility estimation see Reiß [71], Remark 3.4), while at the same time estimation is much easier. Because of this a preprocessing step of the data by blockwise binning as described above and termed pre-averaging in [44], can be used also for spot volatility estimation. This will be described below.

#### 5.2. Estimation of the wavelet coefficients

The content of this section is subdivided into two steps. First we modify pre-averaging, in order to make it suitable for our purposes. Later, we define and discuss the estimators of the wavelet coefficients.

First step: Let us begin with a definition.

**Definition 18** (Pre-average function). A function  $\lambda : [0,2] \rightarrow \mathbb{R}$  that is piecewise Lipschitz continuous and satisfies  $\lambda(t) = -\lambda(2-t)$  is called pre-average function. Given a pre-average function  $\lambda$ , let

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

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

Note that the graph of  $\lambda$  is point symmetric with respect to (1, 0).

Example 3. Let us give a few examples of normalized pre-average functions.

- (i)  $\widetilde{\lambda}(s) = (k+1/2)\pi \cos(s(k+1/2)\pi), \ k = 0, 1, \dots$
- (ii)  $\widetilde{\lambda}(s) = \sqrt{3/2}(\mathbb{I}_{[0,1)}(s) \mathbb{I}_{(1,2]}(s))$ . This leads us to the generalized quadratic variation that has already been discussed in Section 5.1.
- (*iii*)  $\tilde{\lambda}(s) = 3^{-1/2} k \pi \sin(k \pi s), \ k = 1, 2 \dots$
- (*iv*)  $\tilde{\lambda}(s) = 2^{-1}\sqrt{(2k+3)(4k+5)} (1-s)^{(2k+1)}, \ k = 0, 1...$

As in Chapter 4 we set

$$M = \lfloor cn^{1/2} \rfloor. \tag{5.2.1}$$

For the observation vector Y, we introduce the modified pre-averaged observations by

$$\overline{Y}_{i,M}(\lambda) := \frac{M}{n} \sum_{\substack{\underline{j} \in \left(\frac{i-2}{M}, \frac{i}{M}\right] \\ \underline{j}_n \in \left(\frac{i-2}{M}, \frac{i-1}{M}\right]}} \widetilde{\lambda} \left(M\frac{\underline{j}}{n} - (i-2)\right) Y_{\underline{j},n} - \frac{M}{n} \sum_{\substack{\underline{j} \in \left(\frac{i-1}{M}, \frac{i}{M}\right]}} \widetilde{\lambda} \left(i - M\frac{\underline{j}}{n}\right) Y_{\underline{j},n},$$
(5.2.2)

for i = 2, ..., M. There are two changes compared to the original pre-averaging procedure defined in (5.1.1). First, a weighted binning is defined and second we directly consider differences over successive, averaged blocks, without defining first local means. Hence, following the observations of Section 5.1 it is plausible that

$$\overline{Y}_{i,M}(\lambda) = -(X_{(i-1)/M} - X_{(i-2)/M}) + O_p(M^{-1/2} + (M/n)^{1/2})$$
(5.2.3)

and the increments of X are of the same order as the noise.

On the other hand,  $\overline{Y}_{i,M}(\lambda)$  may also be represented (up to a small error) by weighted increments of Y, due to

$$\overline{Y}_{i,M}(\lambda) \approx -\frac{M}{n} \sum_{\substack{\underline{j} \\ n \in \left(0,\frac{1}{M}\right]}} \widetilde{\lambda}\left(M_{\underline{j}}\right) (Y_{\lfloor ni/M \rfloor - j,n} - Y_{\lfloor n(i-2)/M \rfloor + j,n}), \quad i = 2, \dots, M.$$

**Second step:** Let  $\phi$  be an  $L^2$ -function. The estimator for the scalar product  $\langle \phi, \sigma^2 \rangle$  is given by

$$\widehat{\langle\phi,\sigma^2\rangle} = \sum_{i=2}^{M} \phi(\frac{i-1}{M}) \Big(\overline{Y}_{i,M}^2 - \mathfrak{b}(\lambda,Y)_{i,M}\Big),$$
(5.2.4)

where

$$\mathfrak{b}(\lambda, Y)_{i,M} := \frac{M^2}{2n^2} \sum_{\substack{\underline{j} \in \left(\frac{i-2}{M}, \frac{\underline{i}}{M}\right]}} \widetilde{\lambda}^2 \left(M \frac{\underline{j}}{n} - (i-2)\right) \left(Y_{j,n} - Y_{j-1,n}\right)^2 \tag{5.2.5}$$

and  $Y_{0,n} := 0$ .

In particular, given a multiresolution analysis  $((V_j)_j, \phi)$  as well as the corresponding wavelet  $\psi$ , the estimators of the basis coefficients  $\langle \phi_{j_0,k}, \sigma^2 \rangle$  and  $\langle \psi_{j,k}, \sigma^2 \rangle$  are given by  $\langle \widehat{\phi_{j_0,k}, \sigma^2} \rangle$  and  $\langle \widehat{\psi_{j,k}, \sigma^2} \rangle$ , respectively.

**Explanation of (5.2.4):** Going back to (5.2.3) and the discussions in Section 5.1, it is clear that  $\overline{Y}_{i,M}^2 = (X_{(i-1)/M} - X_{(i-2)/M})^2 + O_p(M^{-1} + (M/n)) = \frac{1}{M}\sigma^2(\frac{i-1}{M}) + O_p(M^{-1} + (M/n))$ . Now, the noise is of the same order as the signal. However, by the definition of a pre-average function and imposing smoothness on  $\sigma$ , the influence of the bias due to the approximation can be reduced to smaller order, such that we only need to adjust for the bias induced by the pre-averaged noise. Careful calculations reveal that this can be accomplished by subtracting  $\mathfrak{b}(\lambda, Y)$ . Let us mention, that if  $\tau$  and  $\phi$  are sufficiently smooth, we might approximate

$$\begin{split} \mathfrak{b}(\lambda,Y)_{i,M} &\approx c^2 \frac{1}{n} \sum_{\substack{\underline{j} \in \left(\frac{i-2}{M}, \frac{i}{M}\right]}} \widetilde{\lambda}^2 (M \frac{\underline{j}}{n} - (i-2)) \tau_{\underline{j}/n}^2 \\ &\approx c^2 \tau^2 (\frac{i-1}{M}) \int_{(i-2)/M}^{i/M} \widetilde{\lambda}^2 (Ms - (i-2)) ds = c^2 \frac{1}{M} \tau^2 (\frac{i-1}{M}) \int_0^2 \widetilde{\lambda}^2 (u) du \end{split}$$

and hence (5.2.4) can be written as

$$\widehat{\langle \phi, \sigma^2 \rangle} = \Big(\sum_{i=2}^M \phi(\frac{i-1}{M}) \overline{Y}_{i,M}^2 \Big) - c^2 \|\widetilde{\lambda}\|_2^2 \langle \phi, \tau^2 \rangle$$

up to some small approximation error. This can be compared directly to (4.2.10). Furthermore, since  $\overline{Y}_{i,M}^2 - \mathfrak{b}(\lambda, Y)_{i,M}$  has mean  $\sigma^2(\frac{i-1}{M})/M$  and variance of order  $n^{-1}$  (as shown later) we may think of  $M(\overline{Y}_{i,M}^2 - \mathfrak{b}(\lambda, Y)_{i,M})$  as observations coming from a nonparametric regression model (2.1.3), with regression function  $\sigma^2$  and almost centered (but dependent) errors. As mentioned in (2.1.4),

$$\frac{1}{M}\sum_{i=2}^{M}\phi(\frac{i-1}{M})M(\overline{Y}_{i,M}^2 - \mathfrak{b}(\lambda,Y)_{i,M}) = \sum_{i=2}^{M}\phi(\frac{i-1}{M})(\overline{Y}_{i,M}^2 - \mathfrak{b}(\lambda,Y)_{i,M})$$

is then the natural estimator for the scalar product  $\langle \phi, \sigma^2 \rangle$ .

Since we will deal with wavelet and approximation coefficients simultaneously, let us introduce  $h_{\ell k}(\cdot) = 2^{\ell/2} h(2^{\ell} \cdot -k)$  for a given function h (for which we set  $h = \phi$  and  $h = \psi$  later on).

**Assumption 4** (Assumption on h). Suppose that the function  $h : \mathbb{R} \to \mathbb{R}$  is compactly supported, bounded, and has piecewise Lipschitz derivative.

Furthermore, for a function class  $\mathcal{D}$ , we define  $\mathbb{E}_{\mathcal{D}}[\cdot] := \mathbb{E}[\cdot \mathbb{I}_{\{\sigma^2 \in \mathcal{D}\}}]$ , provided  $\{\sigma^2 \in \mathcal{D}\}$  is measurable. In particular,  $\mathcal{D}_1 \subset \mathcal{D}_2$  implies

$$\mathbb{E}_{\mathcal{D}_1}[U] \leq \mathbb{E}_{\mathcal{D}_2}[U], \quad \text{for non-negative random variables } U.$$
 (5.2.6)

First, we evaluate the (thresholded) moments of  $\langle h_{\ell k}, \sigma^2 \rangle$ . This result will allow us to obtain rates of convergence in the sense of Definition 11 for estimation of the spot volatility. Before we can do so, the precise conditions on the noise process are given.

**Assumption 5** (Refinement on the noise assumption for model (1.1.3)). Let  $\epsilon_{i,n}$  satisfy Assumption 1. Additionally, assume that  $\mathbb{E}[|\eta_{i,n}|^p] < \infty$  for any p > 0 and that the function  $(x_1, x_2) \mapsto \tau(x_1, x_2)$  is continuous and bounded.

The following assumption will allow us to remove the drift in the proofs by a change of measure. It is of interest to note that this assumption is not essential for our proof. In fact, it is imposed in order to reduce the number of terms we need to estimate when we prove moment bounds later. Recall that by Definition 2, the processes  $\sigma$  and b are càdlàg and  $\mathcal{F}_t$ -adapted.

Assumption 6. Suppose that a weak solution of (1.1.1) is unique and well defined. Moreover, a weak solution to  $\widetilde{X}_t = \int_0^t \sigma_s dW_s$  is also unique and well defined, the laws of X and  $\widetilde{X}$  are equivalent on  $\mathcal{F}_1$  and we have, for some  $\rho > 1$ 

$$\mathbb{E}\Big[\exp\left(\rho\int_0^1\frac{b_s}{\sigma_s}dW_s\right)\Big]<\infty.$$

In order to state the following result, we must first introduce the empirical  $L^p[0, 1]$ -norms with respect to the uniform measure on  $\{i/M : i = 1, ..., M\}$ , defined by

$$||f||_{p,M} := \left(\frac{1}{M} \sum_{i=1}^{M} |f(\frac{i}{M})|^p\right)^{1/p}.$$
(5.2.7)

**Proposition 3** (Moment bounds). Suppose that Assumptions 5 and 6 hold and let  $\langle \widehat{h_{\ell k}, \sigma^2} \rangle$  as in (5.2.4). Assume further that h satisfies Assumption 4 and  $2^{\ell} \leq M = \lfloor cn^{1/2} \rfloor$ . Let  $s > 1/\pi$ , then, for any  $p \geq 1$ , C > 0,

$$\mathbb{E}_{\mathcal{B}^{s}_{\pi,\infty}(C)}\left[\left|\langle \widehat{h_{\ell k}, \sigma^{2}} \rangle - \langle h_{\ell k}, \sigma^{2} \rangle\right|^{p}\right] \lesssim M^{-p/2} + M^{-\min\{s-1/\pi, 1/2\}p} \|h_{\ell k}\|_{1,M}^{p}$$

uniformly over  $\ell, k$ .

*Proof.* Let us first introduce some notation. In the following,  $\tilde{\lambda}$  always denotes the normalized version of a pre-average function (in the sense of Definition 18). We define the functions  $\Lambda, \overline{\Lambda} : \mathbb{R} \to \mathbb{R}$ ,

$$\Lambda(s) := \int_{s}^{2} \widetilde{\lambda}(u) du \ \mathbb{I}_{[0,2]}(s) = -\int_{0}^{s} \widetilde{\lambda}(u) du \ \mathbb{I}_{[0,2]}(s)$$
(5.2.8)

and

$$\overline{\Lambda}(s) := \left( \left( \int_0^s \widetilde{\lambda}(u) du \right)^2 + \left( \int_0^{1-s} \widetilde{\lambda}(u) du \right)^2 \right)^{1/2} \mathbb{I}_{[0,1]}(s) \,. \tag{5.2.9}$$

For  $i = 2, \ldots, M$ 

$$\|\overline{\Lambda} (M \cdot -(i-1))\|_2 = M^{-1/2}$$
(5.2.10)

and by using Lemma B.4 also

$$\|\Lambda (M \cdot -(i-2))\|_2 = M^{-1/2}.$$

Moreover, for C > 0, we define the  $L^{\infty}$ -ball

$$L^{\infty}(C) := \{ f : [0,1] \to \mathbb{R}, \ \|f\|_{\infty} \le C \}.$$
(5.2.11)

Some properties deduced from Assumption 4 that will be used extensively can be found in Lemma B.1.

We may decompose

$$\langle \widehat{h_{\ell k}, \sigma^2} \rangle - \langle h_{\ell k}, \sigma^2 \rangle = I + II + III, \qquad (5.2.12)$$

where

$$I := \sum_{i=2}^{M} h_{\ell k} \left(\frac{i-1}{M}\right) \overline{X}_{i,M}^{2} - \langle h_{l,k}, \sigma^{2} \rangle,$$
$$II := \sum_{i=2}^{M} h_{\ell k} \left(\frac{i-1}{M}\right) \left[\overline{\epsilon}_{i,M}^{2} - \mathfrak{b}(\lambda, Y)_{i,M}\right],$$
$$III := 2 \sum_{i=2}^{M} h_{\ell k} \left(\frac{i-1}{M}\right) \overline{X}_{i,M} \overline{\epsilon}_{i,M},$$

and in the spirit of (5.2.2)

$$\overline{X}_{i,M} := \overline{X}_{i,M}(\lambda) := \frac{M}{n} \sum_{\substack{\underline{j} \in \left(\frac{i-2}{M}, \frac{i}{M}\right]}} \widetilde{\lambda} \left(M\frac{\underline{j}}{n} - (i-2)\right) X_{j/n},$$
$$\overline{\epsilon}_{i,M} := \overline{\epsilon}_{i,M}(\lambda) := \frac{M}{n} \sum_{\substack{\underline{j} \in \left(\frac{i-2}{M}, \frac{i}{M}\right]}} \widetilde{\lambda} \left(M\frac{\underline{j}}{n} - (i-2)\right) \epsilon_{j,n}$$
(5.2.13)

are the natural extensions of applying pre-averaging to X and  $\epsilon$ . Bounding I : In a first step we will show that

$$\mathbb{E}_{\mathcal{B}^{s}_{\pi,\infty}(C)}\left[|I|^{p}\right] = \mathbb{E}_{\mathcal{B}^{s}_{\pi,\infty}(C)}\left[\left|\sum_{i=2}^{M} h_{\ell k}\left(\frac{i-1}{M}\right)\overline{X}^{2}_{i,M} - \langle h_{l,k}, \sigma^{2} \rangle\right|^{p}\right] \\ \lesssim M^{-p/2} + \|h_{\ell k}\|^{p}_{1,M}M^{-\min\{s-1/\pi, 1/2\}p},$$
(5.2.14)

uniformly over l and k provided  $2^{\ell} \leq M$ .

By the definition of  $\Lambda$  (see (5.2.8)) and substitution, we obtain for  $s \in [\frac{i-2}{M}, \frac{i}{M}]$ ,

$$\Lambda(Ms - (i-2)) = \int_{Ms - (i-2)}^{2} \widetilde{\lambda}(v) dv = M \int_{s}^{i/M} \widetilde{\lambda}(Mu - (i-2)) du.$$
(5.2.15)

Note that by the continuous embedding (2.4.8) and the identity (2.4.9) it follows

$$\mathcal{B}^s_{\pi,\infty}(C) \subset \mathcal{C}^{\min(s-1/\pi,1/2)}(C') \subset L^\infty(C')$$
(5.2.16)

for some  $C' = C'(s, \pi, C)$  due to  $s > 1/\pi$ . Let  $\mathfrak{R}_n$  as defined in (B.0.1). Using (5.2.6) and (5.2.15) it follows by Lemma B.3 for  $g' = h' = M \ \widetilde{\lambda}(M \cdot -(i-2))$  that uniformly over i

$$\mathbb{E}_{\mathcal{B}^{s}_{\pi,\infty}(C)} \left[ \left| \overline{X}^{2}_{i,M} - \left( \int_{0}^{1} \Lambda \left( Ms - (i-2) \right) dX_{s} \right)^{2} \right|^{p} \right] \\ \lesssim \left\| \Lambda \left( M \cdot -(i-2) \right) \right\|_{2}^{p} \Re_{n} \left[ \Lambda \left( M \cdot -(i-2) \right) \right]^{p} + \Re_{n} \left[ \Lambda \left( M \cdot -(i-2) \right) \right]^{2p} \\ \lesssim M^{-p/2} n^{-p}, \tag{5.2.17}$$

where the last inequality follows from Lemma B.4. Let  $|\operatorname{supp}(h_{\ell k})|$  denote the support length of  $h_{\ell k}$ . Therefore, by Hölder inequality and Lemma B.1

$$\mathbb{E}_{\mathcal{B}^{s}_{\pi,\infty}(C)} \left[ \left| \sum_{i=2}^{M} h_{\ell k} \left( \frac{i-1}{M} \right) \overline{X}^{2}_{i,M} - \sum_{i=2}^{M} h_{\ell k} \left( \frac{i-1}{M} \right) \left( \int_{0}^{1} \Lambda \left( Ms - (i-2) \right) dX_{s} \right)^{2} \right|^{p} \right] \\ \lesssim |\operatorname{supp}(h_{\ell k})|^{p-1} M^{p-1} \\ \times \mathbb{E}_{\mathcal{B}^{s}_{\pi,\infty}(C)} \left[ \sum_{i=2}^{M} \left| h_{\ell k} \left( \frac{i-1}{M} \right) \right|^{p} \left| \overline{X}^{2}_{i,M} - \left( \int_{0}^{1} \Lambda \left( Ms - (i-2) \right) dX_{s} \right)^{2} \right|^{p} \right] \lesssim M^{-3p/2}.$$

$$(5.2.18)$$

By Lemma B.5,

$$\mathbb{E}_{\mathcal{B}^{s}_{\pi,\infty}(C)} \left[ \left| \sum_{i=2}^{M} h_{\ell k} \left( \frac{i-1}{M} \right) \left( \int_{0}^{1} \Lambda \left( Ms - (i-2) \right) dX_{s} \right)^{2} - \int_{0}^{1} \sum_{i=2}^{M} h_{\ell k} \left( \frac{i-1}{M} \right) \Lambda^{2} \left( Ms - (i-2) \right) \sigma_{s}^{2} ds \right|^{p} \right] \lesssim M^{-p/2},$$

and further by triangle inequality

$$\mathbb{E}_{\mathcal{B}^{s}_{\pi,\infty}(C)} \left[ \left| \sum_{i=2}^{M} h_{\ell k} \left( \frac{i-1}{M} \right) \overline{X}^{2}_{i,M} - \int_{0}^{1} \sum_{i=2}^{M} h_{\ell k} \left( \frac{i-1}{M} \right) \Lambda^{2} \left( Ms - (i-2) \right) \sigma_{s}^{2} ds \Big|^{p} \right] \lesssim M^{-p/2}.$$
(5.2.19)

Clearly,  $\Lambda(M \cdot -(i-2))$  has support on  $[\frac{i-2}{M}, \frac{i}{M}]$ , hence  $\Lambda(M \cdot -(i-2))$  and  $\Lambda(M \cdot -(i-1))$  overlap on  $[\frac{i-1}{M}, \frac{i}{M}]$ . Next, we show how the sum can be rewritten in terms of  $\overline{\Lambda}$ . To this end, note that

$$\sum_{i=2}^{M} h_{\ell k} \left(\frac{i-1}{M}\right) \Lambda^{2} \left(Ms - (i-2)\right)$$

$$= \sum_{i=1}^{M} h_{\ell k} \left(\frac{i}{M}\right) \left(\Lambda^{2} \left(Ms - (i-2)\right) + \Lambda^{2} \left(Ms - (i-1)\right)\right) \mathbb{I}_{\left(\frac{i-1}{M}, \frac{i}{M}\right]}(s)$$

$$+ \sum_{i=1}^{M} \left(h_{\ell k} \left(\frac{i-1}{M}\right) - h_{\ell k} \left(\frac{i}{M}\right)\right) \Lambda^{2} \left(Ms - (i-2)\right) \mathbb{I}_{\left(\frac{i-1}{M}, \frac{i}{M}\right]}(s)$$

$$- h_{\ell k}(0) \Lambda^{2} \left(Ms + 1\right) \mathbb{I}_{\left(0, \frac{1}{M}\right]}(s) - h_{\ell k}(1) \Lambda^{2} \left(Ms - (M-1)\right) \mathbb{I}_{\left(1-\frac{1}{M}, 1\right]}(s).$$
(5.2.20)

Now, (5.2.15) and  $\int_0^2 \widetilde{\lambda}(u) du = 0$  imply that

$$\Lambda^2 \left( Ms - (i-2) \right) = \left( \int_0^{1 - (Ms - (i-1))} \widetilde{\lambda}(u) du \right)^2, \quad \text{for } s \in \left( \frac{i-1}{M}, \frac{i}{M} \right]$$

and

$$\Lambda^2 (Ms - (i-1)) = \left( \int_0^{Ms - (i-1)} \widetilde{\lambda}(u) du \right)^2, \quad \text{for } s \in \left(\frac{i-1}{M}, \frac{i}{M}\right].$$

Let  $\overline{\Lambda}$  as in (5.2.9) then for  $i = 1, \ldots, M$ ,

$$\overline{\Lambda}^{2} \left( Ms - (i-1) \right) \mathbb{I}_{\left(\frac{i-1}{M}, \frac{i}{M}\right]}(s) = \left( \Lambda^{2} \left( Ms - (i-2) \right) + \Lambda^{2} \left( Ms - (i-1) \right) \right) \mathbb{I}_{\left(\frac{i-1}{M}, \frac{i}{M}\right]}(s).$$
(5.2.21)

Therefore, on the event  $\sigma^2 \in \mathcal{B}^s_{\pi,\infty}(C)$ , Equation (5.2.20) implies by Lemma B.1 (iii)

$$\left| \int_{0}^{1} \sum_{i=2}^{M} h_{\ell k} \left( \frac{i-1}{M} \right) \Lambda^{2} \left( Ms - (i-2) \right) \sigma_{s}^{2} ds - \int_{0}^{1} \sum_{i=1}^{M} h_{\ell k} \left( \frac{i}{M} \right) \overline{\Lambda}^{2} \left( Ms - (i-1) \right) \sigma_{s}^{2} ds \right| \lesssim M^{-1/2}.$$

$$(5.2.22)$$

Recall that by continuous Sobolev embedding (2.4.8),  $\mathcal{B}^s_{\pi,\infty} \subset \mathcal{B}^{s-1/\pi}_{\infty,\infty}$ . Since  $\mathcal{B}^{s-1/\pi}_{\infty,\infty} \subset \mathcal{C}^{\min(s-1/\pi,1/2)}$  we derive on the event  $\{\sigma^2 \in \mathcal{B}^s_{\pi,\infty}(C)\}$  using  $\|\overline{\Lambda}\|_{L^2} = 1$ 

$$\left| \int_{0}^{1} \sum_{i=2}^{M} h_{\ell k} \left( \frac{i-1}{M} \right) \left( \overline{\Lambda}^{2} \left( Ms - (i-1) \right) - \mathbb{I}_{\left( \frac{i-1}{M}, \frac{i}{M} \right]}(s) \right) \sigma_{s}^{2} ds \right| \\
\leq \left| \int_{0}^{1} \sum_{i=2}^{M} h_{\ell k} \left( \frac{i-1}{M} \right) \overline{\Lambda}^{2} \left( Ms - (i-1) \right) \left( \sigma_{s}^{2} - \sigma_{(i-1)/M}^{2} \right) ds \right| \\
+ \left| \int_{0}^{1} \sum_{i=2}^{M} h_{\ell k} \left( \frac{i-1}{M} \right) \mathbb{I}_{\left( \frac{i-1}{M}, \frac{i}{M} \right]}(s) \left( \sigma_{s}^{2} - \sigma_{(i-1)/M}^{2} \right) ds \right| \\
\lesssim M^{-\min\{s-1/\pi, 1/2\}} \| h_{\ell k} \|_{1,M}. \tag{5.2.23}$$

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

$$\left| \int_{0}^{1} \sum_{i=2}^{M} h_{\ell k} \left( \frac{i}{M} \right) \mathbb{I}_{\left( \frac{i-1}{M}, \frac{i}{M} \right]}(s) \sigma_{s}^{2} ds - \int_{0}^{1} h_{\ell k}(s) \sigma_{s}^{2} ds \right| \lesssim M^{-1/2}, \tag{5.2.24}$$

The moment bound on I, i.e. (5.2.14) follows now by applying successively (5.2.19), (5.2.22), (5.2.23) and (5.2.24).

Bounding II: Combining Lemmas B.6, B.8, B.9 and B.10, we obtain

$$\mathbb{E}_{\mathcal{B}^{s}_{\pi,\infty}}\left[|II|^{p}\right] \lesssim \|h_{\ell k}\|_{1,M}^{p} M^{p} n^{-p} + \|h_{\ell k}\|_{2,M}^{p} M^{-3p/2} n^{-p} + \|h_{\ell k}\|_{p,M}^{p} M^{p+1} n^{-p} \lesssim M^{-p/2},$$

where Lemma B.1 is applied for the last inequality.

Bounding III : Lemma B.7 gives

$$\mathbb{E}_{\mathcal{B}^{s}_{\pi,\infty}(C)} \left[ \left| III \right|^{p} \right] \\ \lesssim \|h_{\ell k}\|_{p,M}^{p} \left( n^{-p/2}M + M^{3p/2+1}n^{-3p/2} \right) + \|h_{\ell k}\|_{2,M}^{p} \left( M^{p/2}n^{-p/2} + M^{p/2}n^{-3p/2} \right) \lesssim M^{-p/2}.$$

By combining the estimates on parts I - III, the proof of Proposition 3 is complete.  $\Box$ 

In order to apply Theorem 2, we need further a result of the type (2.2.3). This is given in the next Proposition.

**Proposition 4** (Deviation bounds). Suppose that Assumptions 5 and 6 hold. Let us further suppose that h satisfies Assumption 4,  $s > 1/\pi$ , and  $M = \lfloor cn^{1/2} \rfloor$ . Further assume that

- (i)  $M2^{-\ell} \ge M^q$ , for some q > 0 and
- (*ii*)  $M^{-(s-1/\pi)} \|h_{\ell k}\|_{1,M} \lesssim M^{-1/2}$ .

Then for C > 0 and  $p \ge 1$ , we have

$$\mathbb{P}\left[\left|\langle \widehat{h_{\ell k}, \sigma^2} \rangle - \langle \sigma^2, h_{\ell k} \rangle_{L^2}\right| \ge \kappa \left(\frac{p \log M}{M}\right)^{1/2} and \sigma^2 \in \mathcal{B}^s_{\pi, \infty}(C)\right] \lesssim M^{-\max(2, p)}$$

for a sufficiently large constant  $\kappa$  and

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

If X is a driftless continuous Itô semimartingale, i.e. b = 0, then  $\kappa$  can be chosen as

$$\kappa > 4\overline{C} + 4\sqrt{2} \ \overline{C} \ \|\tau\|_{\infty} c \|\lambda\|_2 \overline{\lambda}^{-1} + 4\|\tau\|_{\infty}^2 c^2 \|\lambda\|_2^2 \ \overline{\lambda}^{-2}.$$
(5.2.25)

**Remark 2.** Indeed  $\overline{C} < \infty$ , as it follows from the continuous embedding (5.2.16). Moreover, in the case of high smoothness, i.e.  $s - 1/\pi > 1/2$ , Assumption (ii) in Proposition 4 becomes trivial.

#### 5.3. Wavelet estimator

The wavelet estimator of the spot volatility, based on hard-thresholding, is now given by

$$\widehat{\sigma}_W^2 := \sum_k \langle \widehat{\phi_{j_0,k}, \sigma^2} \rangle \phi_{j_0,k} + \sum_{j=j_0}^{j_1} \sum_{k \in \mathbb{Z}} \mathcal{H}_t(\langle \widehat{\psi_{j,k}, \sigma^2} \rangle) \psi_{j,k},$$
(5.3.1)

where  $\mathcal{H}$  denotes the hard-thresholding function as introduced in (2.2.1). The estimator strongly depends on the choice of  $j_0$  and  $j_1$ . Our theoretical results on  $\hat{\sigma}_W^2$ , stated below, will show how these two variables can be selected.

Next, by using Propositions 3 and 4, we aim to apply Theorem 2.

Given an r-regular multiresolution analysis  $((V_j)_j, \phi)$  with corresponding wavelet  $\psi$ , then the following holds.

**Theorem 5.** Suppose that Assumptions 5 and 6 hold. Let  $\widehat{\sigma}_W^2$  be defined as in (5.3.1) and suppose that  $\phi$  and  $\psi$  satisfy Assumption 4. For  $M = \lfloor cn^{1/2} \rfloor$  and  $\alpha_0$ ,  $0 < \alpha_0 \leq 1/2$  choose  $j_0, j_1$ , such that

$$2^{j_0} \sim M^{1-2\alpha_0}$$
, and  $2^{j_1} \sim M^{1/(1+2\alpha_0)}$ 

If the hard-thresholding parameter t is set to  $t := 2\kappa \sqrt{\frac{p \log M}{M}}$ , where  $\kappa$  is a sufficiently large constant, then for any  $\pi \ge 1$ ,  $s - 1/\pi \ge \alpha_0$ , s < r it follows

$$\overline{\lim_{n \to \infty}} v_n^{-1} \mathbb{E} \Big[ \| \widehat{\sigma}_W^2 - \sigma^2 \|_p \, \mathbb{I}_{\{\sigma^2 \in \mathcal{B}_{\pi,\infty}^s\}} \Big] < \infty,$$



Figure 5.1.: The gray areas in the vertical direction display the intervals  $[\alpha_0, \alpha_0/(1-2\alpha_0)]$  for a given  $\alpha_0 \in [0, 1/2)$  according to (5.3.2).

with

$$v_n = \left(\frac{\log^{3/2} n}{n}\right)^{s(1/\pi^*)/(4s(1/\pi^*)+2)} + n^{-\alpha_0/2}$$

and  $s(1/\pi^*)$  is defined as in (1.4.1) and (2.5.6). Or, to state it differently, the estimator  $\hat{\sigma}_W^2$  achieves the rate of convergence  $v_n$  with respect to  $L^p$ -loss, in the sense of Definition 11.

If X is driftless, i.e. b = 0, then  $\kappa$  can be chosen as in (5.2.25).

*Proof.* Using Propositions 3 and 4 this is a direct application of Theorem 2 with  $q_n = M^{-1}$ .

Assuming that the regularity of the multiresolution analysis is sufficiently large and  $\alpha_0 < 1/2$ , Theorem 5 shows that the estimator has the rate of convergence

$$\widetilde{v}_n = \left(\frac{\log^{3/2} n}{n}\right)^{s(1/\pi^*)/(4s(1/\pi^*)+2)}$$

provided

$$s \in \left[\alpha_0 + \frac{1}{\pi}, \frac{\alpha_0}{(1 - 2\alpha_0)}\right]. \tag{5.3.2}$$

These intervals in dependence on  $\alpha_0$  and for  $\pi = \infty$  are displayed in Figure 5.1. Note that  $\tilde{v}_n$  is the rate of convergence obtained in classical nonparametric regression, up to a factor 1/2 in the exponent (cf. Kerkyacharian and Picard [52] and Gloter and Hoffmann [32]).

Once we have established lower bounds in Chapter 6, we will be able to prove that the wavelet estimator adapts to the optimal rate of convergence, up to logarithmic terms, provided a minimal smoothness assumption is satisfied.

Let us remark that we can improve on the thresholding parameter t by further assuming that the noise process is a conditionally symmetric martingale. From a practical perspective this could be reasonable (cf. Diebold and Strasser [24], Assumption 1).

We might also generalize the notion of pre-average functions (Definition 18) by relaxing the condition  $\lambda(t) = -\lambda(2-t)$  to  $\int_0^2 \lambda(s) ds = 0$ . However, the proof becomes more involved. Moreover, we believe that the class  $\{\lambda : \lambda(t) = -\lambda(2-t)\}$  contains all important cases for both optimality and practical purposes.

#### 5.4. Optimizing tuning parameters

In this section we tackle a similar problem as discussed in Section 4.4. Recall (5.2.1). The goal is to find the optimal pre-average function  $\lambda$  and the optimal c in MSE-sense provided that  $\sigma, \tau$  are deterministic constants and  $\eta_{i,n} \sim \mathcal{N}(0,1)$ , i.i.d. Further assume that the drift of X is zero. For a given pre-average function  $\lambda$ , the MSE-minimizing c will be denoted by  $c^*$ .

**Lemma 6.** Assume that  $\lambda$  is a pre-average function in the sense of Definition 18 and suppose the assumptions made above hold. Then the mean square error is given by

$$\begin{split} \operatorname{MSE}(\widehat{\langle 1, \sigma^2 \rangle}) &= 4 \Big( \frac{\sigma^2}{\sqrt{c}} \int_0^1 \Lambda(u) \Lambda(1-u) du - \tau^2 c^{3/2} \int_0^1 \widetilde{\lambda}(u) \widetilde{\lambda}(1-u) du \Big)^2 n^{-1/2} \\ &+ 2 \Big( \frac{\sigma^2}{\sqrt{c}} + 2\tau^2 c^{3/2} \|\widetilde{\lambda}\|_{L^2[0,1]}^2 \Big)^2 n^{-1/2} + o(n^{-1/2}). \end{split}$$

*Proof.* First, let us show that the bias is of a smaller order. In fact, note that using the notation introduced in (5.2.13),  $\mathbb{E}\left[\overline{Y}_{i,M}^2\right] = \mathbb{E}\left[\overline{X}_{i,M}^2\right] + \mathbb{E}\left[\overline{\epsilon}_{i,M}^2\right]$ . Clearly,

$$\left| \mathbb{E} \left[ \overline{\epsilon}_{i,M}^2 \right] - \mathbb{E} \left[ \mathfrak{b}(\lambda, Y)_{i,M} \right] \right| \lesssim 1/n$$

and by (5.2.17) as well as Lemma B.4 also

$$\left|\mathbb{E}\left[\overline{X}_{i,M}^{2}\right] - \sigma^{2}/M\right| = \left|\mathbb{E}\left[\left(\int_{0}^{1} \Lambda(Ms - (i-2))\sigma dW_{s}\right)^{2}\right] - \sigma^{2}/M\right| + O(n^{-5/4}) = O(n^{-5/4}),$$

where both approximations are uniformly in *i*. This shows that the bias is of order  $O(n^{-1/2})$ .

First of all, it is not difficult to see that  $\operatorname{Var}(\sum_{i=2}^{M} \mathfrak{b}(\lambda, Y)_{i,M}) = o(n^{-1/2})$ . Hence,

$$\operatorname{Var}(\widehat{\langle 1, \sigma^2 \rangle}) = \operatorname{Var}(\sum_{i=2}^M \overline{Y}_{i,M}^2) + o\left(n^{-1/4} \left(\operatorname{Var}(\sum_{i=2}^M \overline{Y}_{i,M}^2)\right)^{1/2} + n^{-1/2}\right),$$

by Cauchy-Schwarz. Recall that for Gaussian random variables U, V,  $\operatorname{Cov}(U^2, V^2) = 2(\operatorname{Cov}(U, V))^2$ . Therefore, we have to compute  $\operatorname{Cov}(\overline{Y}_{i,M}, \overline{Y}_{k,M}) = \mathbb{E}[\overline{Y}_{i,M}\overline{Y}_{k,M}]$  only. Furthermore, by Lemma B.3 and Lemma B.4 we bound for  $g' = M \cdot \Lambda(Ms - (i-2)), h' = M \cdot \Lambda(Ms - (i-2)),$ 

$$\mathbb{E}\Big[\Big|\overline{X}_{i,M}\overline{X}_{k,M} - \int_0^1 \Lambda(Ms - (i-2))dX_s \int_0^1 \Lambda(Ms - (k-2))dX_s\Big|\Big] \lesssim n^{-5/4}.$$

Therefore,

$$\mathbb{E}\left[\overline{X}_{i,M}\overline{X}_{k,M}\right] = \sigma^2 \int_0^1 \Lambda(Ms - (i-2))\Lambda(Ms - (k-2))ds + O(n^{-5/4}),$$

where the last two arguments hold uniformly in i, k.

In order to calculate  $\mathbb{E}[\overline{Y}_{i,M}\overline{Y}_{k,M}]$ , we must treat three different cases,  $|i-k| \ge 2$ , |i-k| = 1 and i = k, denoted by I, II and III.

*I*: In this case  $(\frac{i-2}{M}, \frac{i}{M}]$  and  $(\frac{k-2}{M}, \frac{k}{M}]$  do not overlap. By the equalities above, it follows  $\operatorname{Cov}(\overline{Y}_{i,M}, \overline{Y}_{k,M}) = O(n^{-5/4})$ .

II: Without loss of generality, we set k = i + 1. Then, we obtain

$$\begin{aligned} \operatorname{Cov}(\overline{Y}_{i,M}, \overline{Y}_{i+1,M}) &= \mathbb{E}\left[\overline{X}_{i,M} \overline{X}_{i+1,M}\right] + \mathbb{E}\left[\overline{\epsilon}_{i,M} \overline{\epsilon}_{i+1,M}\right] \\ &= \sigma^2 \int_0^1 \Lambda(Ms - (i-2))\Lambda(Ms - (i-1))ds + O(n^{-5/4}) \\ &+ \tau^2 \frac{M^2}{n^2} \sum_{\substack{\underline{j} \\ n \in \left(\frac{i-2}{M}, \frac{i}{M}\right]}} \widetilde{\lambda}(M\frac{\underline{j}}{n} - (i-2))\widetilde{\lambda}(M\frac{\underline{j}}{n} - (i-1)) \\ &= \frac{\sigma^2}{M} \int_0^1 \Lambda(u)\Lambda(1+u)du + \tau^2 \frac{M}{n} \int_0^1 \widetilde{\lambda}(u)\widetilde{\lambda}(1+u)du + O(n^{-1}), \end{aligned}$$

where the last inequality can be verified by Riemann summation. Noting that  $\tilde{\lambda}$  is a pre-average function we obtain  $\lambda(1+u) = -\lambda(1-u)$  and

$$\operatorname{Cov}(\overline{Y}_{i,M},\overline{Y}_{i+1,M}) = \frac{\sigma^2}{M} \int_0^1 \Lambda(u)\Lambda(1-u)du - \tau^2 \frac{M}{n} \int_0^1 \widetilde{\lambda}(u)\widetilde{\lambda}(1-u)du + O(n^{-1}).$$

III: It can be shown by redoing the arguments in II that

$$\operatorname{Var}(\overline{Y}_{i,M}) = \operatorname{Var}(\overline{X}_{i,M}) + \operatorname{Var}(\overline{\epsilon}_{i,M}) = \frac{\sigma^2}{M} \int_0^2 \Lambda^2(u) du + \tau^2 \frac{M}{n} \int_0^2 \widetilde{\lambda}^2(u) du + O(n^{-1}) du +$$

Note that  $\|\Lambda\|_{L^2[0,2]} = 1$ . Since the above results hold uniformly in i, k, it follows directly that

$$\begin{aligned} \operatorname{Var}(\sum_{i=2}^{M} \overline{Y}_{i,M}^{2}) \\ &= \sum_{i,k=2, \ |i-k| \ge 2}^{M} 2\left(\operatorname{Cov}(\overline{Y}_{i,M}, \overline{Y}_{k,M})\right)^{2} + 2\sum_{i=2}^{M-1} 2\left(\operatorname{Cov}(\overline{Y}_{i,M}, \overline{Y}_{i+1,M})\right)^{2} + \sum_{i=2}^{M} 2\left(\operatorname{Var}(\overline{Y}_{i,M})\right)^{2} \\ &= O(n^{-1}) + 4\left(\frac{\sigma^{2}}{\sqrt{c}} \int_{0}^{1} \Lambda(u)\Lambda(1-u)du - \tau^{2}c^{3/2} \int_{0}^{1} \widetilde{\lambda}(u)\widetilde{\lambda}(1-u)du\right)^{2} n^{-1/2} \\ &\quad + 2\left(\frac{\sigma^{2}}{\sqrt{c}} + 2\tau^{2}c^{3/2} \|\widetilde{\lambda}\|_{L^{2}[0,1]}^{2}\right)^{2} n^{-1/2}. \end{aligned}$$

It is an open problem to minimize the functional with respect to  $\tilde{\lambda}$ . However, we will show in this paragraph that there is no solution such that the MSE achieves the optimal asymptotic behavior  $8\tau\sigma^3 n^{-1/4}(1+o(1))$ . In order to see this, note that

$$MSE(\widehat{\langle 1, \sigma^2 \rangle}) \ge 2\left(\frac{\sigma^2}{\sqrt{c}} + 2\tau^2 c^{3/2} \|\widetilde{\lambda}\|_2^2\right)^2 n^{-1/2} + o(n^{-1/2})$$

and the r.h.s. is minimized for  $\tilde{\lambda} = \frac{\pi}{2}\cos(\cdot\frac{\pi}{2})$ . For this choice, we obtain  $\|\tilde{\lambda}\|_2^2 = \pi^2/8$ . Minimizing  $\sigma^2/\sqrt{c} + \tau^2 c^{3/2} \pi^2/4$  we obtain

$$c^{\star} = \frac{2}{\pi\sqrt{3}}\frac{\sigma}{\tau}$$

Therefore, up to smaller order terms

$$MSE(\widehat{\langle 1, \sigma^2 \rangle}) \ge 2\left(\frac{\sigma^2}{\sqrt{c^*}} + \tau^2 (c^*)^{3/2} \frac{\pi^2}{4}\right)^2 n^{-1/2} = \sigma^3 \tau \ \pi \ \frac{16}{9} \sqrt{3} n^{-1/2} \approx 9.67 \sigma^3 \tau n^{-1/2} > 8 \sigma^3 \tau n^{-1/2}.$$
(5.4.1)

Recall the concrete examples for pre-average functions given in Example 3. In Table 5.1 we list, the optimal asymptotic constant that can be obtained with respect to the setting of Lemma 6 for different pre-average functions.

$\widetilde{\lambda}(s) =$	$c^{\star}\tau/\sigma~pprox$	$\lim_{n} \mathrm{MSE} \cdot n^{1/2} / (\tau \sigma^3) \approx$
$\pi\cos(s\pi/2)/2$	0.49	10.21
$3\pi\cos(s3\pi/2)/2$	0.17	31.36
$\sqrt{3/2}(\mathbb{I}_{[0,1)}(s) - \mathbb{I}_{(1,2]}(s))$	0.35	10.74
$3^{-1/2}\pi\sin\left(\pi s\right)$	0.30	12.52
$3^{-1/2} 2\pi \sin(2\pi s)$	0.19	24.35
$2^{-1}3\sqrt{5}(1-s)^3$	0.47	20.41
$2^{-1}\sqrt{91}(1-s)^5$	0.38	20.36

Table 5.1.: Different choices for pre-average functions, the optimal tuning parameter  $c^*$  as well as the asymptotic constant of the MSE for estimation of the integrated volatility.

Let us briefly comment on Table 5.1. Clearly, choosing a cosine function with minimal modes for pre-averaging, or the generalized quadratic variation,  $\sqrt{3/2}(\mathbb{I}_{[0,1)}(s) - \mathbb{I}_{(1,2]}(s))$ , seems to give the best results, at least theoretically. For practical implementations, other pre-average functions might perform better. It should be further mentioned, that the choice of  $c^*$  depends again on the unknown quantities; hence, the given "optimal" estimators are oracles, only.

On the one hand, we have no closed form solution of the minimization problem so far. However, note that there is only a small gap left. In fact, choosing  $\tilde{\lambda}(s) = \frac{\pi}{2} \cos(s\pi/2)$  we obtain approximately 10.21 for the constant  $\lim_n \text{MSE} \cdot n^{1/2}/(\tau\sigma^3)$ . On the other hand by (5.4.1), we know that the best pre-average function can only attain values larger 9.67.

Moreover, the table gives us qualitative information how the choice of  $\lambda$  influences the performance of our estimator. For instance, if the oscillation of the pre-average function increases, the MSE deteriorates.

We could further improve by considering

$$\overline{\overline{Y}}_{i,n} := \frac{M}{n} \sum_{\substack{j \in \left(\frac{l}{n}, \frac{l}{n} + \frac{2}{M}\right]}} \widetilde{\lambda}(M_{\frac{j}{n}} - \frac{l}{n})Y_{j,n}, \quad l = 1, \dots, \lfloor n(1 - \frac{2}{M}) \rfloor.$$
(5.4.2)

Jacod et *al.* [44] show that the incorporation of all pre-average values of the form (5.4.2) yields an estimator that is directly comparable to the realised kernel approach (cf. [44], Remark 1). Hence, by the discussion in Section 4.5, it follows that the proposed estimator of scalar products (4.2.10) in model (1.1.2), essentially agrees with the pre-average estimator defined in [44].

However, there are no theoretical results yet available, which show that the replacement of (5.2.4) by an estimator that includes all pre-averaged values of the form (5.4.2) still yields rate-optimal reconstructions for the spot volatility. Moreover, analysis is much more challenging.

5. Spot volatility estimation in the semimartingale model

# Chapter 6

### Lower bounds and adaptivity

Recall the definitions of lower bounds, given in Chapter 2. In this section we state a number of results proving that the methods presented in Chapters 4 and 5 are rateoptimal up to a logarithmic factor, in general. In order to simplify the presentation, we will restrict ourselves to the spot volatility estimator with respect to cosine basis, as introduced in Section 4.3.

Let us summarize the rates obtained for the upper bounds: For the Gaussian Volterra model it has been shown in Theorem 3 that the estimator converges with the rate  $n^{-\alpha/(4\alpha+2)}$ , uniformly over  $(s \mapsto \sigma_{s,s}) \in \Theta_{\cos}(\alpha, Q)$ ,  $\sigma \in S(0, Q_1)$ ,  $\tau \in T(0, Q_2)$ . In the semimartingale model under the assumptions on Theorem 5 the estimator converges in the sense of Definition 11 with the rate

$$\left(\frac{(\log n)^{3/2}}{n}\right)^{s(1/\pi^*)/(4s(1/\pi^*)+2)}$$

where  $s(1/\pi^*)$  is given by (1.4.1) and (2.5.6), provided that  $s - 1/\pi \le \alpha_0/(1 - 2\alpha_0)$ .

In the same sense as we needed to introduce a generalized notion of upper bounds in order to cope with random parameters, we also have to modify the definition of a lower bound. Here, our definition covers only the case of volatility estimation. In particular, we allow for a change of measure. It is not clear how a general concept might be defined, in order to account for estimation of random functions in general.

**Definition 19.** Suppose that there exists another filtered probability space  $(\widetilde{\Omega}, \widetilde{\mathcal{F}}, (\widetilde{\mathcal{F}}_t)_{t\geq 0}, \widetilde{\mathbb{P}})$ and a process  $\widetilde{X}$  on  $(\widetilde{\Omega}, \widetilde{\mathcal{F}})$  with the same distribution as X under Assumption 6. Moreover, assume that there is a process  $(\widetilde{\epsilon}_{i,n})$  on  $(\widetilde{\Omega}, \widetilde{\mathcal{F}})$  satisfying Assumption 5 with Xreplaced by  $\widetilde{X}$ . If

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

and

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

then the rate  $v_n$  is said to be a lower rate of convergence.

In the following, we present an asymptotic equivalence statement due to Reiß [71] that turns out to be the key ingredient for proving lower bounds. Recall the definition of asymptotic equivalence given in Section 2.6 and consider model (4.2.4), where  $\sigma$  is a deterministic function and  $\epsilon_{i,n} \sim \mathcal{N}(0, \tau^2)$ , i.i.d. Clearly, this is a submodel of the Gaussian Volterra as well as the semimartingale model in the sense that model (4.2.4) can be viewed as a restriction on the original parameter space. Let  $\mathcal{E}_{1,n}$  denote the statistical experiment of observing  $Y = (Y_{1,n}, \ldots, Y_{n,n})^t$  in model (4.2.4). Moreover, let us define a new model where we observe the path  $(\tilde{Y}_t)_{t\in[0,1]}$ 

$$d\tilde{Y}_t = \sqrt{2\sigma(t)}dt + \tau^{1/2}n^{-1/4}dW_t.$$
 (6.0.1)

and let  $\mathcal{E}_{2,n}$  denote the corresponding experiment. Assume that the parameter space in  $\mathcal{E}_{1,n}$  and  $\mathcal{E}_{2,n}$  is  $\Theta := \{\sigma^2 \in \mathcal{C}^{\gamma} : \inf_s \sigma^2(s) \ge \sigma_0\}$  for  $\sigma_0 > 0$  and

$$\gamma > (1 + \sqrt{5})/4.$$
 (6.0.2)

Then the two experiments are asymptotically equivalent (cf. Reiß [71]). In the Gaussian shift model (6.0.1) optimal rates of convergence for estimation of  $\sigma^2$  are well known and since the proof of the equivalence is constructive, optimal rates carry over to model (4.2.4), even in the case of unbounded loss functions. Now, by Sobolev embedding  $\Theta_{\cos}(\alpha) \subset C^{\alpha-1/2}$  and  $B^s_{\pi,\infty} \subset C^{s-1/\pi}$ , provided  $\alpha - 1/2$  and  $s - 1/\pi$  are positive but not integers (cf. Equations (2.4.8) and (2.4.9) as well as [64], Lemma D.8). Therefore, the lower bounds hold in model (1.1.2) for

$$\alpha > (3 + \sqrt{5})/4 \tag{6.0.3}$$

and in model (1.1.3) for  $s - 1/\pi > (1 + \sqrt{5})/4$ . These arguments readily prove that optimal rates in our models are half of those obtained in nonparametric regression/density estimation due to the factor  $n^{-1/4}$  that appears in (6.0.1). Thus, the estimators constructed in Chapters 4 and 5 are rate-optimal (up to log-terms, for the wavelet estimator), provided the minimal smoothness conditions are satisfied.

The lower bounds up to this point are obtained by restricting the volatilities to a smaller parameter space, which contains deterministic volatilities that are dependent only on the time parameter, s. In a second step, we demonstrated that in the reduced model the lower bounds already match the rates we have derived for the estimators in the general framework. This, of course, answers the question regarding the optimal rate of convergence in both the Gaussian Volterra and the semimartingale model.

However, this does not tell us anything regarding optimal rates, if we restrict ourselves to other parameter subspaces, for instance by considering  $\Theta = \{\sigma_{s,t} : \sigma_{s,t} = \sigma_t, s, t \in [0, 1]\},$  i.e.

$$Y_{i,n} = \sigma(\frac{i}{n})W_{i/n} + \tau\eta_{i,n}, \ \eta_{i,n} \sim \mathcal{N}(0,1), \ i = 1, \dots, n.$$
Here, the volatility depends only on the space variable t. For this setting the answer has been given in [63, 64]. In fact, the same optimal rates as in the general setting are valid here.

The proof relies on a multiple testing argument (cf. Tsybakov [74], Chapter 2), combined with the following bound on the Kullback-Leibler divergence of multivariate Gaussian probability measures.

**Lemma 7.** Let  $\mathbb{P}_1$  and  $\mathbb{P}_2$  be the probability measure of n-variate normal random variables with the same mean and covariance matrices  $\Sigma_1$  and  $\Sigma_2$ , respectively. Suppose that  $0 < C\Sigma_1 \leq \Sigma_2$  for some constant  $0 < C \leq 1$ . Denote by  $d_K(.,.)$  the Kullback-Leibler divergence. Then,

$$d_{K}(\mathbb{P}_{2},\mathbb{P}_{1}) \leq \frac{1}{4C^{2}} \|\Sigma_{0}^{-1/2}(\Sigma_{1}-\Sigma_{0})\Sigma_{0}^{-1/2}\|_{2}^{2} \leq \frac{1}{4C^{2}} \|\Sigma_{0}^{-1}\Sigma_{1}-I_{n}\|_{2}^{2},$$

where  $I_n$  is the  $n \times n$  dimensional identity matrix and  $\|.\|_2$  denotes the Frobenius norm.

Note that this result should be (up to a constant) compared directly to (2.6.2), since  $d_H^2(.,.) \leq d_K(.,.)$ . By using the technique based on Lemma 7, lower bounds can be obtained even in the case, where  $\sigma_{s,t} = (t-s)\tilde{\sigma}_s$  (cf. [63], Theorem 2.2). Furthermore, for lower bounds in the Gaussian Volterra model with respect to Sobolev ellipsoids, the minimal smoothness assumption (6.0.3) can be relaxed to  $\alpha > 1/2$  (cf. [63], Theorem 2.1).

Moreover, the lower bounds are derived under the assumption of i.i.d. standard normal microstructure noise. By imposing additional structure, better rates can be obtained. For instance, suppose that  $\mathbb{P}(\eta_{i,n} = 1) = \mathbb{P}(\eta_{i,n} = -1) = 1/2$ . Then the same rates can be obtained as in the noise-free case. For these types of microstructure noise models there might even be a theoretical justification for modeling bid-ask spreads, see Section 8.1.2.

From the discussion above we know the asymptotic behavior of the minimax risk. Let the wavelet estimator be defined as in (5.3.1) with  $j_0 = 0$  and  $2^{j_1} \sim M^{1/2}$  (i.e.  $\alpha_0 = 1/2$ ). Recall Definition 12. Then, it follows directly from Theorem 5 that the wavelet estimator is rate-adaptive, provided  $\frac{1+\sqrt{5}}{4} \leq s - \frac{1}{\pi} \leq r$ , where r denotes the regularity of the multiresolution analysis. 6. Lower bounds and adaptivity

## Chapter 7

## Simulations in the Gaussian Volterra model

This chapter is devoted to illustrate the performance of the Fourier series estimator constructed in Chapter 4. First, we examine the estimator under various noise levels and different sample sizes. Afterwards, the behavior under random volatility is studied.

For our simulations, we consider the following three underlying functions

$$\sigma_{s,t}^{(1)} := (2-s) \exp(\theta(s-t)), \quad \theta > 0$$
  
$$\sigma_{s,t}^{(2)} := (2+\cos(2s\pi))^{1/4} (2+\cos(2t\pi))^{1/4},$$
  
$$\sigma_{s,t}^{(3)} := 1 + (2st-1) \mathbb{I}_{\{st \ge 1/2\}}(s,t).$$

The corresponding spot volatilities,  $s \mapsto \sigma_{s,s}^2$ , are displayed in Figure 7.1. Note that under  $\sigma^{(1)}$ , the Gaussian Volterra process follows the dynamics

$$dX_t = -\theta X_t dt + (2-t)dW_t, \quad X_0 = 0.$$

Therefore, X can be viewed as Ornstein-Uhlenbeck process with time-varying volatility (see also Example 2 (ii)). For  $\sigma^{(2)}$  the corresponding spot volatility  $s \mapsto (2 + \cos(2s\pi))$ 



Figure 7.1.: Spot volatilities corresponding to  $\sigma^{(1)}$  (solid),  $\sigma^{(2)}$  (dashed) and  $\sigma^{(3)}$  ( $-\circ-$ ).

is arbitrarily smooth with respect to the Sobolev spaces  $\Theta_{\text{trig}}$  and  $\Theta_{\cos}$ , introduced in Section 2.4. In order to study the case of low-smoothness,  $\sigma^{(3)}$  is chosen as a kink function.

For our simulations, we set in (4.2.5),

$$k(x):=\frac{8}{2+\pi}\frac{1}{(1+x^2)^2},\quad x\in[0,1]$$

and  $M = \lfloor 2n^{1/2} \rfloor$ , i.e. c = 2. We do not want to assume that the smoothness index of the spot volatility is known. Instead, we investigate the estimator for  $\alpha = 2$ , fixed. Moreover, the weights

$$w_{i,n} := \left(1 - c_{\omega}^{-\alpha} n^{-\alpha/(4\alpha+2)} i^{\alpha}\right)_{+},$$

are chosen according to (4.3.3) with  $c_{\omega} = 2$ . We improve the spot volatility estimator by projecting to zero, whenever  $\hat{\sigma}^2 < 0$ . The method is implemented for both cosine and trigonometric basis using Matlab.

We evaluate our method under the empirical IMSE, i.e.

$$\widehat{\text{IMSE}} = \frac{1}{n} \sum_{i=1}^{n} \left( \widehat{\sigma}^2(\frac{i}{n}) - \sigma^2(\frac{i}{n}) \right)^2.$$

The following tables show the averaged empirical IMSE as well as an estimate of Std(IMSE), based on 100 replications under different scenarios.

In Table 7.1, we investigate the finite sample behavior of the spot volatility estimator for different sample size n and constant noise level  $\tau = 0.1$ . In particular, the results indicate that for  $\sigma^{(1)}$  and  $\sigma^{(3)}$ , the series estimator based on the cosine basis outperforms that one with respect to the trigonometric basis. For  $\sigma^{(2)}$  both estimators are comparable, as expected. In the case of low-smoothness (i.e. for  $\sigma^{(3)}$ ), the decrease of the IMSE for increasing n is much slower than for  $\sigma^{(1)}$  and  $\sigma^{(2)}$ . This can be explained, of course, by the reduction in terms of rates of convergence for low smoothness signals (see for instance Theorem 3).

In Table 7.2, we show the dependence of the numerical performance on different choices of the noise level. The results imply, that the IMSE is barely affected by small  $\tau$ , i.e.  $\tau \in \{0.01, 0.1\}$ , whereas if we increase the noise level to  $\tau = 1$ , the risk explodes. In Figure 7.2, the spot volatility estimators for different noise levels are displayed. In order to obtain results that can be compared directly to each other, the estimators are based

n =	$\sigma^{(1)}, \ \theta = 0.2$		$\sigma^{(2)}$		$\sigma^{(3)}$	
1000	$0.50 \ (0.43)$	0.72(0.40)	$0.53\ (0.30)$	$0.51 \ (0.31)$	0.39(0.19)	0.52 (0.15)
3000	0.39 (0.28)	0.54 (0.23)	0.38 (0.21)	0.35(0.21)	$0.27 \ (0.13)$	0.42 (0.11)
9000	0.21 (0.16)	0.40 (0.14)	0.24 (0.12)	0.23 (0.13)	0.22 (0.10)	0.36 (0.08)

Table 7.1.: Estimated mean and standard deviation of the empirical IMSE for different sample sizes based on 100 simulations. For every  $\sigma^{(i)}$ , i = 1, 2, 3, there are two columns. The first column displays the values of mean and (standard deviation) of the empirical IMSE based on spot volatility estimation with respect to the cosine basis, whereas the second column shows the corresponding values for the estimator under the trigonometric basis. The noise level is set to  $\tau = 0.1$ .

$\tau =$	$\sigma^{(1)}, \ \theta = 0.2$		$\sigma^{(2)}$		$\sigma^{(3)}$	
0.01	0.26 (0.20)	0.47 (0.17)	$0.27 \ (0.16)$	0.24 (0.14)	0.22(0.12)	$0.37 \ (0.08)$
0.1	0.28 (0.23)	0.49 (0.21)	0.27 (0.16)	0.24 (0.16)	0.24 (0.12)	0.40 (0.10)
0.5	1.27 (0.96)	1.53 (1.04)	1.34 (1.01)	1.25 (0.92)	0.92 (0.87)	1.04 (0.85)

Table 7.2.: Estimated mean and standard deviation of the empirical IMSE based on 100 simulations. The order of the entries is the same as in Table 7.1 and the sample size is n = 5000.



Figure 7.2.: Reconstruction of the spot volatility (solid) for different noise levels  $\tau = 0.01$  $(-\circ -), \tau = 0.1 (-\Box -), \text{ and } \tau = 0.5 \text{ (dashed)}.$  The observations are generated from the same realizations of the Gaussian Volterra process with volatility  $\sigma^{(1)}$  ( $\theta = 1$ ) and noise process  $\eta = (\eta_{1,n}, \ldots, \eta_{n,n})$ . The sample size is n = 5000.

on the same realizations of the Gaussian Volterra and noise process. The plot supports our findings from Table 7.2. The reconstructions for small  $\tau$  are more or less of the same quality, yet there is a dramatic decrease of performance for  $\tau = 0.5$ .

Let us summarize the first part of this simulation study. Despite the fact that convergence rates are always below 1/4, we have shown that we are still able to obtain good reconstructions, provided the sample size is at least 1000 and the noise level is not too large. Here, our results rely on simply choosing  $\alpha = 2$ , irrespectively of the true smoothness of the spot volatility. Therefore, we believe that the reconstruction can even be improved by suitable selection rules on the parameters, for instance by cross-validation. In particular, the method relies strongly on a proper choice of  $M = \lfloor cn^{1/2} \rfloor$  and weights  $w_{i,n}$ .

#### Random volatility

Recall that  $X_t = \int_0^t \sigma_{s,t} dW_s$ . In order to study numerical performance of the spot volatility estimator under random volatility, we consider for  $\rho \in [0, 1]$ ,

$$\sigma_{s,t} = \exp(2(s-t)) \left| \rho \ W_s + \sqrt{1-\rho^2} \ \widetilde{W}_s \right|.$$
 (7.0.1)

Here,  $\widetilde{W}$  denotes a  $((\mathcal{F}_t)_{t\geq 0}, \mathbb{P})$ -Brownian motion defined on the same probability space as the driving process W. Moreover, W and  $\widetilde{W}$  are assumed to be independent.



Figure 7.3.: The upper plot shows the map  $s \mapsto \sigma_s$  of a realization of (7.0.1). The lower plot is a gray scale image of the corresponding map  $(s, t) \rightsquigarrow \sigma_{s,t}$ . Note that  $\sigma_{s,t} = 0$  for s > t.



Figure 7.4.: In the upper plot a sample of 5000 data generated from (7.0.1) for  $\rho = 0.5$ and  $\tau = 0.1$  are displayed. The lower plot shows the reconstruction based on Fourier series estimation.

$\rho =$	0.1		0.5		0.9	
IMSE =	0.08 (0.13)	$0.09 \ (0.16)$	$0.13 \ (0.25)$	$0.15 \ (0.28)$	$0.12 \ (0.27)$	0.14 (0.32)

Table 7.3.: Estimated mean and standard deviation of the IMSE based on 1000 simulations. The volatility is given by (7.0.1) where  $\rho$  is one of the values 0.1, 0.5, and 0.9. There are two columns below every given value for  $\rho$ . The first column displays mean and (standard deviation) based on spot volatility estimation with respect to the cosine basis, whereas the second column gives the corresponding values for the estimator under the trigonometric basis. Noise level and sample size are set to  $\tau = 0.1$  and n = 5000, respectively.

The integral  $\int_0^t \sigma_{s,t} dW_s$  can be defined as an Itô-integral for every fixed t (cf. Barndorff-Nielsen and Schmiegel [9]). Note that in distribution,  $\rho W_s + \sqrt{1 - \rho^2} \widetilde{W}_s$  is again a standard Brownian motion, and the correlation with dW is governed by the size of  $\rho$ (for a realization see Figure 7.3). For this simulation, we set  $\alpha = 1/2$  and  $c_{\omega} = 4$ . The other parameters are chosen as before. A typical reconstruction is displayed in Figure 7.4. Though we are unable to recover fine details of the true spot volatility its main features are detected. Table 7.3 indicates that the IMSE increases as  $\rho$  grows.

# Chapter 8

## Application of the semimartingale model to log-returns on high frequencies

In this chapter we study modeling, simulations and real data applications of high-frequency log-returns. This is the main application of the semimartingale model introduced in Chapter 1; hence it deserves special emphasis here.

In the first part of this chapter, we will revisit the current discussion regarding volatility and microstructure noise modeling. In particular, we focus on some empirical findings underlying high-frequency data, known as stylized facts. In the second part, we investigate real data performance of the wavelet estimator introduced in Chapter 5 using a modification based on blockwise thresholding. Recall that log-returns are defined as the increments of the log-price process.

### 8.1. Modeling of high-frequency data

#### 8.1.1. Modeling log-returns

Shortly after it had been discovered that movements of stocks can be interpreted as random walks, it became clear that a constant volatility is not justified because this model is unable to incorporate certain phenomena visible in real data, known as stylized facts (see for instance also LeBaron [54] or Jondeau et *al.* [48], Chapter 2):

- *Thick tails:* The distribution of price changes has a positive kurtosis, i.e. is leptokurtic (cf. Mandelbrot [60]). In Fama [27], [28] power law distributions for modeling of returns are discussed.
- *Volatility clustering:* As observed in Mandelbrot [60], the first autocorrelation for large price changes is positive, i.e. a large jump in price is likely to be followed by another of the same sign.
- Leverage effects: Returns are negatively correlated with changes in volatility (cf. Black [14]).

- Intraday effects/Weekend effects: Over a trading day, the volatility patterns have a characteristic shape (cf. Figure 1.1). Even stronger are the similarities found when comparing the shapes of the volatility over the same weekday. Many empirical findings are completing this picture. For instance, markets that remain open around the clock experience much less volatility on weekends and overnight (cf. Andersen and Bollerslev [4]).
- Announcements: Around the income of public news, volatility increases (cf. Andersen and Bollerslev [4]).
- *Correlation among assets:* There is a positive correlation among different returns and markets (cf. Ramchand and Susmel [70]).

This has lead to the development of ARCH/GARCH models and various extensions of it. These models are stochastic volatility models, where the volatility usually incorporates some of the stylized facts listed above. In the last twenty years, the number of publications regarding this topic increased tremendously. There are many models and different philosophies competing to describe log-returns as accurately as possible (for more on this see for instance the glossary to ARCH (GARCH) by Bollerslev in [15], Chapter 8).

However, it is not clear which of the ARCH/GARCH variants to choose. Even worse, they often depend on additional parameters that must be selected by the practitioners. Therefore, it is valuable to model volatility nonparametrically. In addition, this also has a theoretical justification, since Delbaen and Schachermayer proved that under a relaxation of no arbitrage, semimartingales form a natural class for price processes (for more on this see [21, 22]).

#### 8.1.2. Modeling microstructure noise

In the following we will illustrate that on high-frequencies, market frictions are no longer irrelevant.

To this end let us assume that there are *no* microstructure effects present, in other words, the observed log-price process X is a semimartingale. Recall the definition of realized volatility from Theorem B.1, i.e.

$$\widetilde{[X,Y]}_{(T_k^m)} := X_0 Y_0 + \sum_{i=0}^{k_m - 1} (X^{T_{i+1}^m} - X^{T_i^m}) (Y^{T_{i+1}^m} - Y^{T_i^m}).$$

where  $(T_k^m)_{k,m}$  is an array of increasing stopping times. Now, we compute  $[X, X]_{(T_k^m)}$  for different subsampling frequencies m from the data. If the array  $(T_k^m)_{k,m}$  satisfies the assumptions of Theorem B.1, then we know that  $[X, X]_{(T_k^m)}$  converges to the quadratic



Figure 8.1.: Realized volatilities for decreasing subsampling frequencies and different trading days of FGBL log-prices for one week in July 2007. The *y*-axis is displayed on a logarithmic scale. The approximations of the quadratic variation are displayed for decreasing frequencies. Since the estimates show singular behavior on the highest frequencies (i.e. around 0, where all data are included), this proves the existence of market frictions.

variation [X, X] for increasing frequencies m. Assume that we observe the process X at time points  $t_i$ , i = 1, ..., n. In Figure 8.1 the realized volatilities

$$\sum_{i=1}^{[n/r]-1} (X_{t_{r(i+1)}} - X_{t_{ri}})^2$$

for r = 1, ..., 100 are displayed. Here, the observations come from log-prices of Euro-Bund Future FGBL between July 9th and July 13th, 2007 (for more on this dataset see Section 8.2). Since the estimates increase as we take higher frequencies, this shows that there is no convergence. Consequently, for high-frequencies, the price process cannot be modeled as a semimartingale.

We may state this differently. Assume that the price process is of the special form  $\int_0^1 \sigma_s dW_s$ , where  $\sigma_s$  is a bounded and adapted process. Then the increments are uncorrelated. However, for real data we face a negative first autocorrelation (cf. Figure 8.2).



Figure 8.2.: First ten autocorrelations for the price increments of successive ticks using FGBL tick data from July 9th to July 13th, 2007. The first autocorrelation is estimated to be -0.42.



Figure 8.3.: First autocorrelations for decreasing frequencies and different trading days for the same data set as in Figure 8.2

Studying the first autocorrelation over different time lags, we see that it decreases for larger scales (cf. Figure 8.3) meaning that the microstructure noise effects vanishes.

Although there is no doubt of the existence of market microstructure noise, there is an ongoing discussion of the correct modeling. It is clear that any microstructure noise model should be able to reproduce the effects described above.

It is worthwhile to discuss some aspects of market microstructure modeling within this work. Detailed surveys can be found in Madhavan [56] and Hasbrouck [37]. The particular market structure in which trading takes place is responsible for microstructure effects. Indeed, the trading process determines the two main factors for market microstructure, trading frictions and information.

Trading frictions: Different trading protocols such as screen-based electronic system or

floor trading may cause diverse sources of microstructure noise. The most important are listed below.

- Bid-ask spread: For a given time point, there is nothing like an observable true price, but only an ask and a bid-price. The difference between them is the so-called bid-ask spread. A simple model dating back to Roll [72] is to assume that we observe the true price process plus some random variable  $\epsilon$  that may take values -s/2, 0, s/2, where s denotes the spread. This allows one to model seller and buyer-initiated trades (i.e.  $\epsilon = -s/2$  and  $\epsilon = s/2$ , respectively) as well as trading at the midquote ( $\epsilon = 0$ ). However, spreads are time dependent. For instance, they are highest around the open/close of a market. This has been incorporated by Demsetz [23], where the spread is allowed to depend on the number of transactions, the price and other time-varying factors.
- *Rounding errors:* Prices may only change by a multiple of the tick value, where the tick value is prespecified by trading protocols (tick size). The price discreteness implies that we may only observe the price up to some rounding.
- *Market maker:* In many markets, such as foreign exchange markets, dealers, socalled market makers, take the opposite of a transaction at any time for some specified sell and ask prices. This induces additional transaction costs. Market makers are not only intermediaries between investors, but they also play an active role in price formation by setting bid and ask prices. Their aversion to strong price changes leads to mean reversion effects in the price process (cf. Madhavan [56], Chapter 3).
- *Market fragmentation:* Additionally microstructure effects may arise from the fact that shares are traded on many markets simultaneously.

*Information:* A second key issue is the amount of information that is publicly available. The transparency of a market may differ depending, for example, whether or not limit order books are displayed to the public. Asymmetric information among market participants can be a source of additional noise.

To summarize this, we stress that the sources of market microstructure are manifold and depend on the special structure of the market. In particular, the level of microstructure evolves over time. Furthermore, besides bid ask-spreads and rounding errors it is difficult to model these effects appropriately and the magnitude of a given single effects cannot be quantified. By supsampling we may reduce the microstructure due to rounding. However, in general, it is unclear if certain sources of microstructure are negligible or dominant. Thus, modeling microstructure noise should be done in a fairly general way.

Besides the semimartingale model introduced in Chapter 1, we discuss two other microstructure models and show that all three are able to capture the distinctive features



Figure 8.4.: We investigate three models of market microstructure, to see whether they are able to reproduce the effects observed for real data in Figures 8.1 - 8.3. The first row shows quadratic variation and autocorrelation for RMN, the second row for the semimartingale model and finally the last row for RSMM.

observed in Figures 8.1 - 8.3. For  $\tau, a > 0$  assume we observe

$$Y_{i,n}^{(1)} = \lfloor aX_{i/n} \rfloor / a, \quad i = 1, \dots, n,$$
  
$$Y_{i,n}^{(2)} = \lfloor a(X_{i/n} + \tau \epsilon_{i,n}) \rfloor / a, \quad i = 1, \dots, n,$$

where X is the true price process and  $\eta_{i,n}$  is a sequence of random variables. In the future we will refer to the first model as rounded microstructure noise model (RMN) and to the second as rounded semimartingale model (RSMM).

Let  $X_t = 0.002W_t$ ,  $\tau = 3 \cdot 10^{-5}$ ,  $a = 10^4$  and  $\epsilon_{i,n} \sim \mathcal{N}(0,1)$  i.i.d. For a realization of n = 10000 simulated observations we calculated the quadratic variation/autocorrelation for different time scales in the same way as in Figures 8.1-8.3. The results are displayed in Figure 8.4. Clearly, all of these models are capable of capturing the effects induced by microstructure noise that are observed in real data.

#### 8.2. Real data performance

In order to investigate the behavior of wavelet thresholding, as introduced in Chapter 5 for real data, we consider FGBL returns from the Eurex database. This dataset consists of tick data of the German Bund Future with expiration month September 2007. The trades are recorded with a precision of 10 milliseconds. Overall, between 10.000 and 20.000 trades occur during one day. We removed observations that are not due to trading. Moreover, only those observations within the time interval from 9 a.m. to 6 p.m. (Central European Time) are considered. The reconstructions are done in tick time, i.e. the time between successive trades is defined as one time unit. It has been stressed by Dahlhaus and Neddermeyer [19] that this is preferable, since estimation in real time leads to a reconstruction of the trading intensity in the first order.

By inspection of the data, it becomes clear that wavelet term-by-term thresholding, as in (5.3.1), does not lead to stable results. Figures 8.5 and 8.6 display simulated as well as real data along with the corresponding pre-averaged values

$$Z_{i,M} := M\left(\overline{Y}_{i,M}^2(\lambda) - \mathfrak{b}(\lambda, Y)_{i,M}\right), \quad i = 1, \dots, M.$$
(8.2.1)

As outlined in Section 5.2, the random variables  $Z_{i,M}$  can be viewed as an observation from a heteroscedastic nonparametric regression problem with the spot volatility as the regression function. However, the second plot in Figure 8.5 as well as the second plot in Figure 8.6 show that, for both real and simulated data, we face a number of outliers. Moreover, due to the fact that  $Z_{i,M}$  is the square of  $\overline{Y}_{i,M}(\lambda)$  minus some (almost deterministic) bias correction, the noise behaves more like a scaled and centered  $\chi_1^2$  distributed random variable. As mentioned in Section 2.2, including information on neighboring coefficients may increase the robustness of the wavelet estimator. This has lead us to implement the block thresholding estimator proposed in Cai and Zhou [17]. In the following we refer to this estimator as block thresholding spot volatility estimator.

For our simulations, we always choose  $\lambda(t) = 3^{-1/2}\pi \sin(\pi s)$  as pre-average function. Moreover, the reconstructions are performed using Haar wavelets.

We compare our procedure with the natural approach for the noise-free case, namely to treat the squared and suitably scaled increments defined in (3.1.3) as a regression problem (cf. Chapter 3). Then, we apply the same block-thresholding rule, as for the pre-averaged data. The resulting estimator will be called naive spot volatility estimator.

In Figure 8.5 we investigated the semimartingale model for simulated data. The sample size is n = 16000 and the noise level  $\tau = 0.01$ . The noise process  $\eta = (\eta_{1,n}, \ldots, \eta_{n,n})$  is assumed to be *n*-variate standard normal. We set  $M = \lfloor 4n^{1/2} \rfloor$ , i.e. c = 4.

Clearly, we see that the naive estimate is much larger than the true spot volatility. This strong positive bias is due to the microstructure effect (see also Figure 8.1). In contrast to that, the block thresholding spot volatility estimator yields a reasonable reconstruction (Figure 8.5, Plot 4).

Note that under the assumption of additive microstructure noise and in view of (4.2.11) the naive spot volatility estimator multiplied with 1/(2n) provides us with a reconstruction of the squared noise level  $t \mapsto \tau^2(X_t, t)$ . Since  $\tau = 0.01$ , this explains the almost constant reconstruction in the third plot of Figure 8.5.

It is worthwhile to mention that estimation of the noise level is also of practical importance. For modeling high-frequency log-returns, this provides us with a measure of the quality/liquidity of a market (cf. Hasbrouck [36] and Ait-Sahalia and Yu [2]).

For the remaining part of this section, we will set c = 1.

Figure 8.6 compares the naive and the block thresholding spot volatility estimator. Note that again, the estimated spot volatility based on the noise-free assumption is larger (by a factor of 10) than the reconstruction which includes pre-averaging.

Figure 8.7 shows tick data as well as the reconstruction of the spot volatility on a "special" day. In fact, since the 4th of July is an official holiday in the United States, the market is much less volatile than on regular trading days.

In Figure 8.8, the reconstructed spot volatilities over successive trading days are displayed. While the first hours show not many changes in the volatility, there are a few peaks after 1 p.m. On average the spot volatility is estimated to be around  $10^{-5}$ .



Figure 8.5.: Reconstruction of the spot volatility for simulated data. The first plot shows the data sample. The pre-averaged data  $Z_{i,M}$  as defined in (8.2.1) and the naive reconstruction are displayed in the second and third plot, respectively. Finally, the block thresholding spot volatility estimator (solid) and the true spot volatility,  $s \mapsto \sigma_{s,s}^2$  (dashed), are given in the lower plot.



Figure 8.6.: Reconstruction of the spot volatility for FGBL tick data on July 3rd, 2007. The first plot shows the recorded prices. The pre-averaged data  $Z_{i,M}$  as defined in (8.2.1) and the naive reconstruction are displayed in the second and third plot, respectively. Finally, the block thresholding spot volatility estimator is given in the lower plot.



Figure 8.7.: Reconstruction of the spot volatility for FGBL tick data on July 4th, 2007. The structure of the plots is the same as in Figure 8.6.



Figure 8.8.: Reconstruction of the spot volatility based on FGBL tick data over one week in July 2007.

## Chapter 9

### Discussion and outlook

We have developed a nonparametric theory for estimation in the Gaussian Volterra and the semimartingale model. Concerning rates of convergence, it has been shown that these models may be compared to nonparametric regression/density estimation, with  $\sqrt{n}$  observations, instead of n. We found that this reduces the classical rates of convergence by a factor 1/2 (cf. Reiß [71]). Hence, if we make a formal analogy to inverse problems, we may define 1/2 as the additional degree of ill-posedness induced by the microstructure noise.

The Gaussian Volterra model: As shown in our numerical examples, the Fourier series estimator is applicable even in the more general setting, where we have a random volatility. Since this is not covered by our theoretical results so far, it would be of great interest to prove that even in this case, the spot volatility estimator is optimal with respect to rates of convergence.

In general, the surface  $(s,t) \mapsto \sigma_{s,t}$  is not estimable due to non-identifiability. In fact, the representation  $X_t = \int_0^t \sigma_{s,t} dW_s$  is not unique. Consider for instance

$$X_t = \int_0^t \left(\frac{\alpha+1}{\alpha} - \frac{2\alpha+1}{\alpha}\frac{s^{\alpha}}{t^{\alpha}}\right) dW_s, \quad \alpha > -1/2.$$

Then, X is again a standard Brownian motion (cf. Lévy [55], Section 2). Therefore, we cannot distinguish between  $\sigma_{s,t} = 1$ , and  $\sigma_{s,t} = \frac{\alpha+1}{\alpha} - \frac{2\alpha+1}{\alpha} \frac{s^{\alpha}}{t^{\alpha}}$ . Moreover, by representation (4.1.1), estimation of the surface  $(s,t) \mapsto \sigma_{s,t}$  implies that we need to estimate a drift-type term. For regular drift functions, this is known to be impossible using high-frequency data. Therefore, estimation of the equivalence class of  $(s,t) \mapsto \sigma_{s,t}$  is a challenging problem.

Of special interest are function classes of the type

$$\{(s,t)\mapsto\sigma_{s,t}:\sigma_{s,t}=|t-s|^{\alpha}\ \widetilde{\sigma}_s,\ \alpha>0.\}$$

Note that estimation of  $\tilde{\sigma}$  is not covered within the framework presented in Chapter 4, since  $\sigma_{s,s} = 0$ . However, it is well known, that consistent estimation can be accomplished in this setting, see for instance Gloter and Hoffmann [31, 30].

The semimartingale model: At the end of the discussion, we would like to comment briefly on the semimartingale model. As we have seen in Chapter 8, this model provides us with a good tool for understanding high-frequency log prices under market microstructure noise. However, it is not capable of incorporating all of the effects that are visible in real data. In the following, we summarize a number of important topics that are not treated within this work.

In order to generalize spot volatility estimation to higher dimensions, i.e. considering logreturns of more than one asset simultaneously, a further difficulty arises, since on tick data level, the processes are not synchronized and simple interpolation techniques fail. Note that in this case we aim to estimate a covariance depending on a time parameter. Nowadays, methods have been developed in order to estimate the integrated volatility in higher dimensions (cf. Hayashi and Yoshida [38], Malliavin and Mancino [58], Bibinger [13] and the references therein). However, for spot volatility under microstructure noise this has not yet been addressed.

Aside from the estimation of the path  $s \rightsquigarrow \sigma_{s,s}^2$ , other quantities are also of interest. For instance, the wavelet thresholding method allows us to construct estimators of the derivatives of the spot volatility in a straightforward way. Finally, we want to stress that an important issue within these models is the construction of asymptotic confidence bands.

# Appendix A

### Proofs and technical results for Chapters 2 and 4

#### Proof of Theorem 2

Before we can give the proof of Theorem 2, some preliminaries need to be introduced before. Recall the wavelet expansion  $f = f_1 + f_2$ , where

$$f_1 := \sum_k \langle \phi_{j_0,k}, f \rangle \phi_{j_0,k}$$
 and  $f_2 := \sum_{(j,k) \in \Lambda} \langle \psi_{j,k}, f \rangle \psi_{j,k},$ 

and  $\Lambda$  denotes the set  $\{(j,k) : j \ge j_0, k \in \mathbb{Z}\}$ . Similarly, we write for the estimator (2.2.2),  $\hat{f}_t = \hat{f}_1 + \hat{f}_{2,t}$  with

$$\widehat{f_1} := \sum_k \langle \widehat{\phi_{j_0,k}, f} \rangle \phi_{j_0,k} \quad \text{and} \quad \widehat{f_{2,t}} := \sum_{(j,k) \in \Lambda} \mathcal{H}_t(\langle \widehat{\psi_{j,k}, f} \rangle) \psi_{j,k}.$$

Now,

$$\mu_p(\{(j,k)\}) := \|\psi_{j,k}\|_p^p$$

defines a measure on the index set  $\Lambda$ . Let  $\ell_{q,\infty}(p)$  denote the class of functions such that the corresponding seminorm

$$||f||_{\ell_{q,\infty}(p)}^{q} := \sup_{t>0} t^{q} \mu_{p} \big( \{ (j,k) \in \Lambda : |\langle \psi_{j,k}, f \rangle| \ge t \} \big)$$

is finite. Clearly,  $\ell_{q,\infty}(p)$  depends on the choice of the basis system. In the following, we assume that the basis is fixed. Embeddings of Besov spaces into  $\ell_{q,\infty}(p)$  are well understood (cf. [52], Theorem 6.2 and [32], Proposition 5). Let us recall some results in this direction, that are needed later on. All embeddings are continuous.

**Lemma A.1.** Define  $\pi_s := \frac{p}{2s+1}$ . Let  $0 , <math>0 \le s < \infty$ .

(i) If 
$$\pi > \pi_s$$
, then  $B^s_{\pi,\infty} \subset \ell_{\pi_s,\infty}(p)$ .

(ii) If 
$$\frac{2}{2s+1} < \pi < \pi_s$$
 then  $B^s_{\pi,\infty} \subset \ell_{r_s,\infty}(p)$ , where  $r_s = (\frac{p}{2} - 1)/(s + \frac{1}{2} - \frac{1}{\pi})$ .

(iii) If  $\pi = \pi_s$  then

$$\mu_p(\{\lambda : |f_\lambda| > t\}) \lesssim t^{-\pi} \left|\log(\frac{1}{t})\right|,$$

provided there exists  $\delta > 0$ , such that  $f \in B^s_{\pi,\infty} \cap B^{\delta}_{p,\infty}$ . In particular, if  $s - 1/\pi > 0$  then it is sufficient to assume  $f \in B^s_{\pi,\infty}$ .

Now, we are able to give the proof of Theorem 2.

*Proof.* Throughout the proof we set  $\check{f} := f \cdot \mathbb{I}_{\{f \in \mathcal{B}^s_{\pi,\infty}(C)\}}$ . Note that by assumption  $\phi$  and  $\psi$  are compactly supported. Using [52], Theorems 4.1 and 4.2 we infer that the following properties hold, i.e.

(a) for any index set  $I \subset \Lambda$ ,

$$\int_0^1 \left(\sum_{(j,k)\in I} |\psi_{j,k}(x)|^2\right)^{p/2} dx \sim \sum_{(j,k)\in I} ||\psi_{j,k}||_p^p,$$

(b)

$$\int_0^1 \left(\sum_k |\phi_{j_0,k}(x)|^2\right)^{p/2} dx \sim \sum_k ||\phi_{j_0,k}||_p^p,$$

(c) for p > 1, any real sequence  $(a_{j,k})_{j,k}$  and any index set  $I \subset \Lambda$ ,

$$\left\| \left( \sum_{(j,k)\in I} |a_{j,k}\psi_{j,k}|^2 \right)^{1/2} \right\|_p \sim \left\| \sum_{(j,k)\in I} a_{j,k}\psi_{j,k} \right\|_p,$$

(d) for p > 1, any real sequence  $(a_k)_k$ ,

$$\left\| \left( \sum_{k} |a_{k} \phi_{j_{0},k}|^{2} \right)^{1/2} \right\|_{p} \sim \left\| \sum_{k} a_{k} \phi_{j_{0},k} \right\|_{p},$$

where  $\sim$  denotes norm equivalence.

We obtain

$$\mathbb{E}\left[\|\widehat{f} - f\|_{p}^{p} \mathbb{I}_{\{f \in \mathcal{B}_{\pi,\infty}^{s}(C)\}}\right] \lesssim \mathbb{E}\left[\|\widehat{f}_{1} - f_{1}\|_{p}^{p} \mathbb{I}_{\{f \in \mathcal{B}_{\pi,\infty}^{s}(C)\}}\right] + \mathbb{E}\left[\|\widehat{f}_{2} - f_{2}\|_{p}^{p} \mathbb{I}_{\{f \in \mathcal{B}_{\pi,\infty}^{s}(C)\}}\right].$$

Applying (d) and Minkowski's generalized inequality, we obtain for  $p \ge 2$ 

$$\begin{split} & \mathbb{E}\left[\|f_{1}-f_{1}\|_{p}^{p} \mathbb{I}_{\left\{f\in\mathcal{B}_{\pi,\infty}^{s}(C)\right\}}\right] \\ &\lesssim \mathbb{E}\left[\left\|\left(\sum_{k}\left(\langle\widehat{\phi_{j_{0},k}},f\rangle-\langle\phi_{j_{0},k},f\rangle\right)^{2}\phi_{j_{0},k}^{2}\right)^{1/2}\right\|_{p}^{p} \mathbb{I}_{\left\{f\in\mathcal{B}_{\pi,\infty}^{s}(C)\right\}}\right] \\ &= \int \mathbb{E}\left[\left(\sum_{k}\left(\langle\widehat{\phi_{j_{0},k}},f\rangle-\langle\phi_{j_{0},k},f\rangle\right)^{2}\phi_{j_{0},k}^{2}(x)\mathbb{I}_{\left\{f\in\mathcal{B}_{\pi,\infty}^{s}(C)\right\}}\right)^{p/2}\right]dx \\ &\lesssim \int \left(\sum_{k}\left(\mathbb{E}^{2/p}\left[|\langle\widehat{\phi_{j_{0},k}},f\rangle-\langle\phi_{j_{0},k},f\rangle|^{p}\mathbb{I}_{\left\{f\in\mathcal{B}_{\pi,\infty}^{s}(C)\right\}}\right]\phi_{j_{0},k}^{2}(x)\right)\right)^{p/2}dx \\ &\lesssim q_{n}^{p/2}\int \left(\sum_{k}\phi_{j_{0},k}^{2}(x)\right)^{p/2}dx \lesssim q_{n}^{p/2}\sum_{k}\|\phi_{j_{0},k}\|_{p}^{p}, \end{split}$$

where we used (b) in the last step. This can be further bounded by  $q_n^{p/2} 2^{pj_0/2} \leq q_n^{\alpha_0}$ . The same holds true for  $1 \leq p < 2$ . The proof for this case is rather easy and will not be given here. Applying for instance the arguments from [32], p. 172, yields the desired bound. Therefore, we obtain

$$\mathbb{E}\left[\|\widehat{f}_1 - f_1\|_p^p \mathbb{I}_{\{f \in \mathcal{B}^s_{\pi,\infty}(C)\}}\right] \lesssim q_n^{\alpha_0}.$$

Therefore, it suffices to show that  $\sup_{f \in \mathcal{B}^s_{\pi,\infty}(C)} \mathbb{E} \left[ \| \widehat{f}_{2,t} - f_2 \|_p^p \mathbb{I}_{\{f \in \mathcal{B}^s_{\pi,\infty}(C)\}} \right]$  is of the right order. By [32], Proposition 5.2 it follows for 0 < u < p

$$\mathbb{E}\left[\|\widehat{f_{2,t}} - f_2\|_p^p \mathbb{I}_{\{f \in \mathcal{B}_{\pi,\infty}^s(C)\}}\right] \\
\lesssim \left(\left(q_n \log(1/q_n)\right)^{(p-u)/2} \mathbb{E}\left[\|\check{f}\|_{\ell_{u,\infty}(\pi)}^u\right] + \mathbb{E}\left[\left\|\sum_{j>j_1}\sum_k \langle \psi_{j,k},\check{f}\rangle\psi_{j,k}\right\|_p^p\right]\right). \quad (A.0.1)$$

Now we prove the result by bounding the approximation term in step (I) followed by estimates on  $\mathbb{E}[\|\check{f}\|^{u}_{\ell_{u,\infty}(\pi)}]$  for the cases  $\pi > \pi_{s}$  (II) and  $\pi < \pi_{s}$  (III). Finally, in a last step (IV), we treat the case  $\pi = \pi_{s}$ .

*I*: Let us first prove that  $B^s_{\pi,\infty} \subset B^{s(1/p)}_{p,\infty}$  is a continuous embedding. Assume  $p < \pi$ . Then, we have by (2.4.7) that  $B^s_{\pi,\infty} \subset B^s_{p,\infty} = B^{s(1/p)}_{p,\infty}$ . On the other hand, if  $p \ge \pi$ , the Sobolev embedding (2.4.8) applies and we obtain  $B^s_{\pi,\infty} \subset B^{s+1/p-1/\pi}_{p,\infty} = B^{s(1/p)}_{p,\infty}$ . Therefore, we derive

$$\begin{split} \left\| \sum_{j>j_1} \sum_k \langle \psi_{j,k}, \check{f} \rangle \psi_{j,k} \right\|_p &\leq \sup_j 2^{js(1/p)} \left\| \sum_k \langle \psi_{j,k}, \check{f} \rangle \psi_{j,k} \right\|_p \sum_{j>j_1} 2^{-js(1/p)} \\ &\lesssim 2^{-j_1 s(1/p)} \|\check{f}\|_{B_{p,\infty}^{s(1/p)}} \lesssim q_n^{s(1/p)/(2\alpha_0+1)}. \end{split}$$

Note that if  $p < \pi$  then s(1/p) = s and therefore

$$\frac{s(1/p)}{2\alpha_0 + 1} \ge \frac{s}{2s + 1} \ge \frac{s(1/\pi^*)}{2s(1/\pi^*) + 1}.$$
(A.0.2)

On the other hand, if  $\pi_s < \pi \le p$  then  $s(1/p) = s + 1/p - 1/\pi$  and  $s(1/\pi^*) = s$ . By some calculations using that  $s - 1/\pi \ge \alpha_0$  we obtain (A.0.2) for this case as well. Finally, for  $\pi \le \pi_s$ , we have  $s(1/p) = s + 1/p - 1/\pi$ ,

$$\pi^{\star} = \frac{p-2}{2s - \frac{2}{\pi} + 1}, \quad s(1/\pi^{\star}) = \frac{s + \frac{1}{p} - \frac{1}{\pi}}{1 - \frac{2}{p}}, \quad \text{and} \quad \frac{s(1/\pi^{\star})}{2s(1/\pi^{\star}) + 1} = \frac{s + \frac{1}{p} - \frac{1}{\pi}}{2s + 1 - \frac{2}{\pi}}.$$
 (A.0.3)

It follows directly that (A.0.2) holds, again. This yields

$$\mathbb{E}\left[\left\|\sum_{j>j_1}\sum_k \langle \psi_{j,k},\check{f}\rangle\psi_{j,k}\right\|_p^p\right] \lesssim q_n^{p\frac{s(1/\pi^\star)}{2s(1/\pi^\star)+1}}$$

uniformly over  $f \in \mathcal{B}^s_{\pi,\infty}(C)$ .

II: Note that in this situation  $s(1/\pi^*) = s$ . Moreover, by Lemma A.1 (i),

 $B^s_{\pi,\infty} \subset \ell_{p/(2s+1),\infty}(p).$ 

Now we can choose u = p/(2s+1) and obtain  $\|\check{f}\|_{\ell_{u,\infty(p)}} \lesssim \|\check{f}\|_{B^s_{\pi,\infty}} \leq C$ . Thus, in this case Theorem 2 follows by applying step I and (A.0.1).

III: Note that it suffices to treat the case  $\frac{3}{2s+1} < \pi < \pi_s$  only, since by assumption  $s - 1/\pi > 0$  and  $\pi \ge 1$  and therefore  $2(s - 1/\pi) + (1 - 1/\pi) > 0$  or equivalently,  $\pi > \frac{3}{2s+1}$ . This implies p > 3. Using Lemma A.1 (ii), we might now argue similar as in II and obtain

$$B^s_{\pi,\infty} \subset \ell_{r_s,\infty}(p).$$

Hence, the r.h.s. of (A.0.1) can be bounded further by

$$(q_n \log(1/q_n))^{p} \frac{\frac{s+1/p-1/\pi}{2s+1-2/\pi}}{q_n} + q_n^{p\frac{s(1/\pi^*)}{2s(1/\pi^*)+1}}.$$

Now, (A.0.3) yields

$$\mathbb{E}\left[\|\widehat{f}_{2,t} - f_2\|_p^p \,\mathbb{I}_{\{f \in \mathcal{B}_{\pi,\infty}^s(C)\}}\right] \lesssim \left(q_n \log(1/q_n)\right)^{p \,\frac{s(1/\pi^*)}{2s(1/\pi^*)+1}}$$

*IV:* Bounding this term can be done in the same way as in [32], p. 175. Note that in this case we obtain another logarithmic term, i.e.

$$\mathbb{E}\left[\|\widehat{f}_{2,t} - f_2\|_p^p \mathbb{I}_{\{f \in \mathcal{B}^s_{\pi,\infty}(C)\}}\right] \lesssim \left(q_n \log(1/q_n)\right)^{p \frac{s(1/\pi^*)}{2s(1/\pi^*)+1}} \log(1/q_n)$$
$$\lesssim \left(q_n \log^{3/2}(1/q_n)\right)^{p \frac{s(1/\pi^*)}{2s(1/\pi^*)+1}}.$$

**Proof of Lemma 3.** Let us start proving the first statement. Set  $y := 2C_j(1+2\delta)$ . By Lemma 2 (iv), we derive

$$\mathbb{P}\left(\left|M_{j}\right| \geq 2(1+\delta)\sqrt{C_{j}p\log m}\right) \\
\mathbb{P}\left(\left|M_{j}\right| \geq 2(1+\delta)\sqrt{C_{j}p\log m}, \ [M]_{j} + \langle M \rangle_{j} \leq y\right) + \mathbb{P}\left([M]_{j} + \langle M \rangle_{j} > y\right) \\
\leq 2m^{-p} + \mathbb{P}\left([M]_{j} + \langle M \rangle_{j} > y, \ \langle M \rangle_{j} \leq C_{j}(1+\delta)\right) + \mathbb{P}\left(\langle M \rangle_{j} > C_{j}(1+\delta)\right). \quad (A.0.4)$$

Moreover,

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

Clearly,  $[M]_j - \langle M \rangle_j$  is  $\mathcal{F}_j$  measurable and  $\mathbb{E}[[M]_j - \langle M \rangle_j | \mathcal{F}_{j-1}] = 0$ . Since M is square integrable,  $([M]_j - \langle M \rangle_j)_j$  is an  $\mathcal{F}_j$ -martingale. It follows by Chebycheff's and Rosenthal's inequality for martingales (cf. Lemma B.18) and  $\kappa \geq 2$ 

$$\mathbb{P}\left([M]_{j} - \langle M \rangle_{j} > 2C_{j}\delta\right) \lesssim C_{j}^{-\kappa} \mathbb{E}\left[\left|[M]_{j} - \langle M \rangle_{j}\right|^{\kappa}\right]$$
$$\lesssim C_{j}^{-\kappa} \sum_{i=0}^{j-1} \mathbb{E}\left[\left|\Delta_{i}M\right|^{2\kappa}\right] + C_{j}^{-\kappa} \mathbb{E}\left[\left|\sum_{i=0}^{j-1} \mathbb{E}\left[(\Delta_{i}M)^{4}|\mathcal{F}_{i}\right]\right|^{\kappa/2}\right]$$
$$\lesssim C_{j}^{-\kappa}(j+j^{\kappa/2}) \lesssim j^{-\epsilon\kappa},$$

where we used in the last step Hölder's inequality

$$\mathbb{E}\Big[\Big|\sum_{i=0}^{j-1} \mathbb{E}\big((\Delta_i M)^4 |\mathcal{F}_i\big)\Big|^{\kappa/2}\Big] \lesssim j^{\kappa/2-1} \sum_{i=0}^{j-1} \mathbb{E}\Big[\mathbb{E}\big[|\Delta_i M|^{2\kappa} |\mathcal{F}_i\big]\Big] \lesssim j^{\kappa/2}.$$

Now, choosing  $\kappa := q_0^{-1} p \epsilon^{-1} > 2$ , yields

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

Combining this with (A.0.4) and Condition (i) completes the proof of the first statement. For the second statement, note that by using Lemma 2 (iii), we may in analogy to (A.0.4) bound

$$\mathbb{P}\left(\left|M_{j}\right| \geq 2(1+\delta)\sqrt{C_{j}p\log m}\right)$$
  
$$\leq 2m^{-p} + \mathbb{P}\left([M]_{j} > C_{j}(1+2\delta), \ \langle M \rangle_{j} \leq C_{j}(1+\delta)\right) + \mathbb{P}\left(\langle M \rangle_{j} > C_{j}(1+\delta)\right)$$

Therefore, using Condition (i)

$$\mathbb{P}\left(\left|M_{j}\right| \geq 2(1+\delta)\sqrt{C_{j}p\log m}\right) \lesssim m^{-p} + \mathbb{P}\left([M]_{j} - \langle M \rangle_{j} > \delta\right)$$

and the proof is finished by arguing as for the first part.

99

### Proofs for Chapter 4

Before we give the proof of Lemma 4, recall that  $\mathbb{M}_{p,q}$ ,  $\mathbb{M}_p$  and  $\mathbb{D}_p$  as the spaces of  $p \times q$  matrices,  $p \times p$  matrices and  $p \times p$  diagonal matrices over  $\mathbb{R}$ , respectively.

**Proof of Lemma 4.** Recall (2.6.2) and let us suppress the index n in  $\phi_n$ , i.e.  $\phi = \phi_n$ . Now, let

$$T := \left( \begin{array}{cccc} 1 & & & \\ -1 & \ddots & & \\ & \ddots & \ddots & \\ & & -1 & 1 \end{array} \right).$$

Then, we may represent  $TY(\phi)$  by  $TY(\phi) = T\widetilde{Y}(\phi) + R_1 + R_2$ , where  $R_1$  and  $R_2$  are random vectors with components

$$(R_1)_i := \int_{(i-1)/n}^{i/n} \left(\sqrt{\phi(\frac{i-1}{n})} - \sqrt{\phi(s)}\right) \sigma_s dW_s,$$
$$(R_2)_i := \left(\sqrt{\phi(\frac{i-1}{n})} - \sqrt{\phi(\frac{i}{n})}\right) \tau(\frac{i}{n}) \eta_{i,n}.$$

Recall the definition of Cov(U, V) for column vectors U, V. Then, we can write

$$T \operatorname{Cov}(Y(\phi))T^{t} = \operatorname{Cov}(TY(\phi))$$
  
=  $T \operatorname{Cov}(\widetilde{Y}(\phi))T^{t} + \operatorname{Cov}(R_{1}) + \operatorname{Cov}(R_{2}) + \operatorname{Cov}(T\widetilde{Y}(\phi), R_{1})$   
+  $\operatorname{Cov}(R_{1}, T\widetilde{Y}(\phi)) + \operatorname{Cov}(T\widetilde{Y}(\phi), R_{2}) + \operatorname{Cov}(R_{2}, T\widetilde{Y}(\phi)).$ 

By (2.6.2), the proof is complete, whenever we can show that uniformly on  $c \leq \sigma, \tau \leq C$ 

$$(I) \qquad \left\| \operatorname{Cov}(T\widetilde{Y}(\phi))^{-1/2} \left( \operatorname{Cov}(R_1) + \operatorname{Cov}(R_2) \right) \operatorname{Cov}(T\widetilde{Y}(\phi))^{-1/2} \right\|_2^2 \to 0,$$
  

$$(II) \qquad \left\| \operatorname{Cov}(T\widetilde{Y}(\phi))^{-1/2} \left( \operatorname{Cov}(T\widetilde{Y}(\phi) \mid R_1) + \operatorname{Cov}(R_1 \mid T\widetilde{Y}(\phi)) \right) \operatorname{Cov}(T\widetilde{Y}(\phi))^{-1/2} \right\|^2 \to 0.$$

$$(II) \qquad \left\|\operatorname{Cov}(TY(\phi))^{-1/2}\left(\operatorname{Cov}(TY(\phi), R_1) + \operatorname{Cov}(R_1, TY(\phi))\right)\operatorname{Cov}(TY(\phi))^{-1/2}\right\|_2 \to 0,$$

$$(III) \qquad \left\|\operatorname{Cov}(T\widetilde{Y}(\phi))^{-1/2}\left(\operatorname{Cov}(T\widetilde{Y}(\phi), R_2) + \operatorname{Cov}(R_1, T\widetilde{Y}(\phi))\right) - \operatorname{Cov}(T\widetilde{Y}(\phi))^{-1/2}\right\|_2^2 \to 0,$$

$$(III) \quad \left\| \operatorname{Cov}(T\tilde{Y}(\phi))^{-1/2} \left( \operatorname{Cov}(T\tilde{Y}(\phi), R_2) + \operatorname{Cov}(R_2, T\tilde{Y}(\phi)) \right) \operatorname{Cov}(T\tilde{Y}(\phi))^{-1/2} \right\|_2^2 \to 0,$$

for  $n \to \infty$ .

(I): First note that using  $|\sqrt{x} - \sqrt{y}| = |x - y|/(\sqrt{x} + \sqrt{y})$  for x, y > 0 we obtain  $\sup_{c \leq \sigma, \tau \leq C} \operatorname{Cov}(R_1) + \operatorname{Cov}(R_2) \ll n^{-5/4}I_n$ . Let  $K_n$  denote the covariance matrix of  $(X_{1/n}, X_{2/n}, \ldots, X_1)^t$ . Then,  $\operatorname{Cov}(\widetilde{Y}(\phi)) \geq K + c^2(\inf_{n,s} \phi_n(s))I_n$  and hence  $\operatorname{Cov}(T\widetilde{Y}(\phi)) \gtrsim n^{-1}I_n + TT^t$ . The *i*-th eigenvalues of  $TT^t$  is

$$\lambda_i(TT^t) = 4\sin^2\left(\frac{(2i-1)\pi}{4n+2}\right) \ge \frac{i^2}{4n^2}$$

(cf. [63], Lemma A.1). Therefore,

$$\left\| (n^{-1}I_n + TT^t)^{-1} \right\|_2^2 \le \sum_{i=1}^n \left( n^{-1} + \lambda_i (TT^t) \right)^{-2} \le n^{5/2} + \sum_{i=\lfloor n^{1/2} \rfloor + 1}^n \lambda_i^{-2} (TT^t) \lesssim n^{5/2}.$$
(A.0.5)

This completes (I).

(II): Clearly,  $\operatorname{Cov}(T\widetilde{Y}(\phi), R_1) + \operatorname{Cov}(R_1, T\widetilde{Y}(\phi))$  is diagonal. Note the identity

$$\left(\phi^{1/2}(\frac{i-1}{n}) - \phi^{1/2}(s)\right)\phi^{1/2}(s) = \frac{1}{2}\left(\phi(\frac{i-1}{n}) - \phi(s) - \frac{\left(\phi(\frac{i-1}{n}) - \phi(s)\right)^2}{\left(\phi^{1/2}(\frac{i-1}{n}) + \phi^{1/2}(s)\right)^2}\right).$$
 (A.0.6)

Using Lemma A.7 (iv) and arguing as for (I) above, convergence to zero can be shown in a straightforward way.

(III): Let  $\Phi_1, \Phi_2 \in \mathbb{D}_n$  with diagonal entries  $(\Phi_1)_{i,i} = \phi^{1/2}(\frac{i}{n})\tau(i/n)$  and  $(\Phi_2)_{i,i} = (\phi^{1/2}(\frac{i-1}{n}) - \phi^{1/2}(\frac{i}{n}))\tau(i/n)$ . Then,

$$\operatorname{Cov}(TY(\phi), R_2) = \operatorname{Cov}(T\Phi_1\epsilon, \Phi_2\epsilon) = T\Phi_1\Phi_2.$$

Further, we may write  $T\Phi_1\Phi_2 + \Phi_1\Phi_2T^t = T\Phi_1\Phi_2T^t + \Phi_3$ , where  $\Phi_3 \in \mathbb{D}_n$ ,  $(\Phi_3)_{i,i} := (\Phi_1)_{i,i} \cdot (\Phi_2)_{i,i} - (\Phi_1)_{i-1,i-1} \cdot (\Phi_2)_{i-1,i-1}$ , and  $(\Phi_1)_{0,0} := (\Phi_2)_{0,0} := 0$ . Hence, the left hand side of (III) may be bounded by a constant times

$$\left\|\operatorname{Cov}(\widetilde{Y}(\phi))^{-1/2}\Phi_{1}\Phi_{2}\operatorname{Cov}(\widetilde{Y}(\phi))^{-1/2}\right\|_{2}^{2}+\left\|\operatorname{Cov}(T\widetilde{Y}(\phi))^{-1/2}\Phi_{3}\operatorname{Cov}(T\widetilde{Y}(\phi))^{-1/2}\right\|_{2}^{2}.$$

Using (A.0.5) and (A.0.6), the first term tends to zero for  $n \to \infty$ . Furthermore, we can write for i > 1,

$$2(\Phi_3)_{i,i} = \left(\Delta_{i-2,n}^2 \phi\right) \tau^2 \left(\frac{i-1}{n}\right) + \left(\Delta_{i-1,n} \phi\right) \left(\Delta_{i-1,n} \tau^2\right) \\ + \left(\Delta_{i-2,n} \sqrt{\phi}\right)^2 \tau^2 \left(\frac{i-1}{n}\right) - \left(\Delta_{i-1,n} \sqrt{\phi}\right)^2 \tau^2 \left(\frac{i}{n}\right)$$

and together with (4.2.3) we obtain  $\Phi_3 \ll n^{-5/4} I_n$ , completing the proof.

Completion of the proof of Proposition 2.  $\langle X_2(\phi_n), X_2(\phi_n) \rangle_{\sigma}$ : We define  $T(\phi_n) \in \mathbb{M}_{n-1}$ 

$$(T(\phi_n))_{i,j} = \begin{cases} \left(\Delta_i(\phi_n^{1/2}\tau)\right)^2 & \text{for } i = j - 1, \\ \left(\Delta_j(\phi_n^{1/2}\tau)\right)^2 & \text{for } i = j + 1, \\ 0 & \text{otherwise.} \end{cases}$$
(A.0.7)

Note the relation

$$\operatorname{Cov}\left(X_{2}(\phi_{n})\right) = I(\phi_{n}^{1/2}\tau)AI(\phi_{n}^{1/2}\tau) = \frac{1}{2}I(\phi_{n}\tau^{2})A + \frac{1}{2}AI(\phi_{n}\tau^{2}) + \frac{1}{2}T(\phi_{n}), \quad (A.0.8)$$

where A is as defined in (4.2.8). We derive

$$\begin{split} \langle X_2(\phi_n), X_2(\phi_n) \rangle_{\sigma} &= \mathbb{E} \left[ X_2(\phi_n)^t D J_n D X_2(\phi_n) \right] \\ &= \frac{1}{2} \operatorname{tr} \left( J_n D I(\phi_n \tau^2) A D \right) + \frac{1}{2} \operatorname{tr} \left( J_n D A I(\phi_n \tau^2) D \right) \\ &+ \frac{1}{2} \operatorname{tr} \left( J_n D T(\phi_n) D \right) \\ &= \operatorname{tr} \left( \Lambda J_n D I(\phi_n \tau^2) D \right) + \frac{1}{2} \operatorname{tr} \left( J_n D T(\phi_n) D \right), \end{split}$$

where  $\Lambda$  is defined in (4.2.9). Furthermore,

$$\operatorname{tr}\left(\Lambda J_n DI(\phi_n \tau^2) D\right) = \frac{n}{M} \sum_{i=1}^M k(\frac{i}{M}) \lambda_i \left(A\left(\phi_n \tau^2, 0\right) - A\left(\phi_n \tau^2, 2i\right)\right)$$
$$= A\left(\phi_n \tau^2, 0\right) \quad \frac{n}{M} \sum_{i=1}^M k(\frac{i}{M}) \lambda_i - \frac{n}{M} \sum_{i=1}^M k(\frac{i}{M}) \lambda_i A\left(\phi_n \tau^2, 2i\right). \tag{A.0.9}$$

Because of  $\max_{i=1,\dots,M} \lambda_i = \lambda_M \lesssim n^{-1}$ , it holds

$$\left|\frac{n}{M}\sum_{i=1}^{M}k(\frac{i}{M})\lambda_{i} A\left(\phi_{n}\tau^{2},2i\right)\right| \leq \frac{n}{M}\sum_{i=1}^{M}k(\frac{i}{M})\lambda_{i}\sum_{q\in\mathbb{Z}, q\equiv2i \text{ mod } 2n}\left|(\phi_{n}\tau^{2})_{q}\right|$$
$$\lesssim n^{-1/2}\sum_{l=0}^{\infty}\left|(\phi_{n}\tau^{2})_{l}\right| \lesssim n^{-1/4}, \qquad (A.0.10)$$

by arguing as in (4.2.23) for the last step. Therefore, (A.0.9) can be written as

$$\left| \operatorname{tr} \left( \Lambda J_n D I(\phi_n \tau^2) D \right) - \frac{n}{M} \sum_{i=1}^M k(\frac{i}{M}) \lambda_i \ (\phi_n \tau^2)_0 \right|$$
  
$$\lesssim \sum_{m=n}^{\infty} \left| (\phi_n \tau^2)_m \right| + n^{-1/4} \lesssim n^{-1/4},$$
(A.0.11)

applying the idea of (4.2.22) to  $\tau$  instead of  $\sigma$ . In particular, note that the the inequalities in (A.0.10) and (A.0.11) are uniformly over  $\phi_n \in \Phi_n(\kappa, C)$ ,  $\tau \in T(\kappa, Q_2)$ . This gives by Lemma A.2

$$\sup_{\phi_n \in \Phi_n(\kappa, C), \ \tau \in T(\kappa, Q_2)} \left| \operatorname{tr} \left( \Lambda J_n DI(\phi_n \tau^2) D \right) - \pi^2 c^2 \int_0^1 k(x) x^2 dx \ (\phi_n \tau^2)_0 \right| \lesssim n^{-1/4}$$

Moreover,

$$\begin{aligned} |\operatorname{tr} (J_n DT(\phi_n) D)| &\leq \operatorname{tr} (J_n) \max_{i,j} \left| (DT(\phi_n) D)_{i,j} \right| &\leq 4 \operatorname{tr} (J_n) \max_i (\Delta_{i,n} (\phi_n^{1/2} \tau))^2 \\ &\leq 8 \operatorname{tr} (J_n) \left[ \|\tau\|_{\infty}^2 \max_i (\Delta_{i,n} \phi_n^{1/2})^2 + \|\phi_n\|_{\infty} \max_i (\Delta_{i,n} \tau)^2 \right]. \end{aligned}$$

Recall that  $\operatorname{tr}(J_n) = O(n)$ . By assumptions on  $\phi_n$  and  $\tau$ ,

$$\sup_{\phi_n \in \Phi_n(\kappa, C), \ \tau \in T(\kappa, Q_2)} \left| \operatorname{tr} \left( J_n D T(\phi_n) D \right) \right| = O\left( n^{-1/4} \right)$$

and therefore

$$\sup_{\phi_n \in \Phi_n(\kappa, C), \ \tau \in T(\kappa, Q_2)} \left| \langle X_2(\phi_n), X_2(\phi_n) \rangle_{\sigma} - \pi^2 c^2 \int_0^1 k(x) x^2 dx \ \langle \phi_n, \tau^2 \rangle \right| = O(n^{-1/4}).$$

 $\langle {\bf Z_1}(\phi_{\bf n}), {\bf Z_1}(\phi_{\bf n}) \rangle_\sigma$  : Note that by Lemma A.6 (i)

 $\langle Z_1(\phi_n), Z_1(\phi_n) \rangle_{\sigma} = \operatorname{tr} \left( DJ_n D \operatorname{Cov} \left( Z_1(\phi_n) \right) \right) \leq \lambda_1 \left( \operatorname{Cov} \left( Z_1(\phi_n) \right) \right) \operatorname{tr} \left( J_n \right).$ 

Obviously,  $\operatorname{Cov}(Z_1(\phi_n)) \in \mathbb{D}_{n-1}$  and

$$(\text{Cov}(Z_1(\phi_n)))_{i,i} = \phi_n(\frac{i}{n}) \int_{i/n}^{(i+1)/n} (\sigma_{s,i/n} - \sigma_{i/n,i/n})^2 ds$$

This entails

$$\sup_{\phi_n \in \Phi_n(\kappa,C), \ \sigma \in S(\kappa,Q_1)} \langle Z_1(\phi_n), Z_1(\phi_n) \rangle_{\sigma} \lesssim n^{-1/2}.$$

 $\langle \mathbf{Z}_2(\phi_{\mathbf{n}}), \mathbf{Z}_2(\phi_{\mathbf{n}}) \rangle_{\sigma}$ : The covariance matrix of  $Z_2(\phi_n)$  can be bounded as follows

$$\left| \left( \operatorname{Cov}(Z_{2}(\phi_{n})) \right)_{i,j} \right| = \left| \phi_{n}^{1/2}(\frac{i}{n}) \phi_{n}^{1/2}(\frac{j}{n}) \int_{0}^{((i\wedge j)+1)/n} (\sigma_{s,(i+1)/n} - \sigma_{s,i/n}) (\sigma_{s,(j+1)/n} - \sigma_{s,j/n}) ds \right| \lesssim n^{-7/4},$$
(A.0.12)

uniformly in i, j = 1, ..., n - 1 and  $\phi_n \in \Phi_n(\kappa, C), \sigma \in S(\kappa, Q_1)$ . We obtain

$$\langle Z_2(\phi_n), Z_2(\phi_n) \rangle_{\sigma} = \operatorname{tr} \left( DJ_n D \operatorname{Cov}(Z_2(\phi_n)) \right) \le \lambda_1(J_n) \operatorname{tr} \left( \operatorname{Cov}(Z_2(\phi_n)) \right) \lesssim n^{-1/4}$$

and hence

$$\sup_{\phi_n \in \Phi_n(\kappa,C), \ \sigma \in S(\kappa,Q_1)} \left\langle Z_2(\phi_n), Z_2(\phi_n) \right\rangle_{\sigma} = O\left(n^{-1/4}\right).$$
(A.0.13)

 $\langle \mathbf{Z}_3(\phi_n), \mathbf{Z}_3(\phi_n) \rangle_{\sigma}$ : The components of  $Z_3(\phi_n)$  are uncorrelated, i.e.  $\operatorname{Cov}(Z_3(\phi_n)) \in \mathbb{D}_{n-1}$ and  $(\operatorname{Cov}(Z_3(\phi_n)))_{i,i} = \phi_n(\frac{i}{n})(\Delta_{i,n}\tau)^2$ . This implies

 $\sup_{\phi_n \in \Phi_n(\kappa,C), \ \tau \in T(\kappa,Q_2)} \langle Z_3(\phi_n), Z_3(\phi_n) \rangle_{\sigma} \le \sup_{\phi_n \in \Phi_n(\kappa,C), \ \tau \in T(\kappa,Q_2)} \lambda_1(\operatorname{Cov}(Z_3(\phi_n))) \operatorname{tr}(J_n) \lesssim n^{-1/2}.$ 

 $\langle \mathbf{X}_1(\phi_n), \mathbf{Z}_1(\phi_n) \rangle_{\sigma}$ : By Cauchy-Schwarz inequality,

$$|\langle X_1(\phi_n), Z_1(\phi_n) \rangle_{\sigma}| \le \langle X_1(\phi_n), X_1(\phi_n) \rangle_{\sigma}^{1/2} \langle Z_1(\phi_n), Z_1(\phi_n) \rangle_{\sigma}^{1/2} \lesssim n^{-1/4}$$

uniformly over  $\phi_n \in \Phi_n(\kappa, C), \ \sigma \in S(\kappa, Q_1).$ 

 $\langle \mathbf{X_1}(\phi_{\mathbf{n}}), \mathbf{Z_2}(\phi_{\mathbf{n}}) \rangle_{\sigma}$ : Note that

$$\operatorname{Cov} \left( X_1(\phi_n), Z_2(\phi_n) \right)_{i,j} = \begin{cases} 0 & j < i, \\ \phi_n^{1/2}(\frac{i}{n}) \phi_n^{1/2}(\frac{j}{n}) \int_{i/n}^{(i+1)/n} \sigma_{i/n,i/n}(\sigma_{s,(j+1)/n} - \sigma_{s,j/n}) ds & j \ge i, \end{cases}$$

implying

$$\sup_{\phi_n \in \Phi_n(\kappa,C), \ \sigma \in S(\kappa,Q_1)} \max_{i,j=1,\dots,n-1} |\operatorname{Cov} \left(X_1(\phi_n), Z_2(\phi_n)\right)_{i,j}| \lesssim n^{-1} n^{-7/8}.$$

Since by assumption  $\sum |k_j| < \infty$ , we may apply Proposition A.1 and obtain

$$\sup_{\phi_n \in \Phi_n(\kappa,C), \ \sigma \in S(\kappa,Q_1)} |\langle X_1(\phi_n), Z_2(\phi_n) \rangle_{\sigma}| \ll n^{-1/4}.$$

 $\langle \mathbf{X}_2(\phi_{\mathbf{n}}), \mathbf{Z}_3(\phi_{\mathbf{n}}) \rangle_{\sigma}, \langle \mathbf{Z}_2(\phi_{\mathbf{n}}), \mathbf{Z}_3(\phi_{\mathbf{n}}) \rangle_{\sigma}$ : These terms can be bounded in the same way as  $\langle X_1(\phi_n), Z_1(\phi_n) \rangle_{\sigma}$ .

Combining the results above yields (4.2.16). Before we give an upper bound for the variance the following preliminaries have to be mentioned.

First, let  $\Xi(\phi_n) := \operatorname{Cov}(X_1(\phi_n) + Z_1(\phi_n) + Z_2(\phi_n))$ . Since  $X_1(\phi_n) + Z_1(\phi_n) + Z_2(\phi_n)$  is Gaussian, we may write  $X_1(\phi_n) + Z_1(\phi_n) + Z_2(\phi_n) \stackrel{\mathcal{D}}{=} \Xi^{1/2}(\phi_n)\xi$ , where  $\xi \sim \mathcal{N}(0, I_{n-1})$ . Furthermore, let  $\Delta \in \mathbb{M}_{n-1,n}$  be given by

$$\Delta := \left( \begin{array}{ccc} -1 & 1 & & \\ & \ddots & \ddots & \\ & & -1 & 1 \end{array} \right)$$

and recall  $\epsilon := (\tau(1/n)\eta_{1,n}, \dots, \tau(1)\eta_{n,n})^t$ . Then,  $X_2(\phi_n) + Z_3(\phi_n) = I(\phi_n^{1/2})\Delta \epsilon$ . Hence, we may write

$$\Delta Y(\phi_n) \stackrel{\mathcal{D}}{=} \Xi^{1/2}(\phi_n)\xi + I(\phi_n^{1/2})\Delta \epsilon.$$

Since  $\operatorname{Var}(U+V) \leq 2\operatorname{Var}(U) + 2\operatorname{Var}(V)$  we have by Proposition 1

$$\operatorname{Var}\left(\langle \widehat{\phi_n, \sigma^2} \rangle\right) \leq 2 \operatorname{Var}\left((\Delta Y(\phi_n))^t D J_n D(\Delta Y(\phi_n))\right) + 2 \operatorname{Var}\left(\langle \widehat{\phi_n, \tau^2} \rangle\right)$$
$$\lesssim \operatorname{Var}\left((\Delta Y(\phi_n))^t D J_n D(\Delta Y(\phi_n))\right) + n^{-1/2}$$
(A.0.14)

uniformly over  $\phi_n \in \Phi_n(\kappa, C)$ ,  $\sigma \in S(\kappa, Q_1)$ ,  $\tau \in T(\kappa, Q_2)$ . Moreover,

$$(\Delta Y(\phi_n))^t DJ_n D(\Delta Y(\phi_n)) = \xi^t \Xi^{1/2}(\phi_n) DJ_n D\Xi^{1/2}(\phi_n) \xi$$
  
+  $2\epsilon^t \Delta^t I(\phi_n^{1/2}) DJ_n D\Xi^{1/2}(\phi_n) \xi$   
+  $\epsilon^t \Delta^t I(\phi_n^{1/2}) DJ_n DI(\phi_n^{1/2}) \Delta \epsilon.$ 

Hence,

$$\operatorname{Var}\left((\Delta Y(\phi_n))^t DJ_n D(\Delta Y(\phi_n))\right) \leq 3 \operatorname{Var}\left(\xi^t \Xi^{1/2}(\phi_n) DJ_n D\Xi^{1/2}(\phi_n)\xi\right) + 12 \operatorname{Var}\left(\epsilon^t \Delta^t I(\phi_n^{1/2}) DJ_n D\Xi^{1/2}(\phi_n)\xi\right) + 3 \operatorname{Var}\left(\epsilon^t \Delta^t I(\phi_n^{1/2}) DJ_n DI(\phi_n^{1/2}) \Delta \epsilon\right).$$

Finally, we bound Var  $(\xi^t \Xi^{1/2}(\phi_n) DJ_n D\Xi^{1/2}(\phi_n)\xi)$ , Var  $(\epsilon^t \Delta^t I(\phi_n^{1/2}) DJ_n DI(\phi_n^{1/2}) \Delta \epsilon)$ and Var  $(\xi^t \Xi^{1/2}(\phi_n) DJ_n DI(\phi_n^{1/2}) \Delta \epsilon)$  in three steps, denoted by *I*, *II* and *III*, respectively.

I: By Lemma A.9 (ii), we have

$$\operatorname{Var}\left(\xi^{t}\Xi^{1/2}(\phi_{n})DJ_{n}D\Xi^{1/2}(\phi_{n})\xi\right) = 2\left\|\Xi^{1/2}(\phi_{n})DJ_{n}D\Xi^{1/2}(\phi_{n})\right\|_{2}^{2}$$
$$= 2\left\|J_{n}^{1/2}D\operatorname{Cov}(X_{1}(\phi_{n})+Z_{1}(\phi_{n})+Z_{2}(\phi_{n}))DJ_{n}^{1/2}\right\|_{2}^{2}$$

Now, by Lemma A.10 it follows

$$\operatorname{Cov}(X_1(\phi_n) + Z_1(\phi_n) + Z_2(\phi_n)) \le 3\operatorname{Cov}(X_1(\phi_n)) + 3\operatorname{Cov}(Z_1(\phi_n)) + 3\operatorname{Cov}(Z_2(\phi_n)).$$

Using Lemma A.7 (iii) we may bound further

$$\begin{split} \left\| J_n^{1/2} D \operatorname{Cov}(X_1(\phi_n) + Z_1(\phi_n) + Z_2(\phi_n)) D J_n^{1/2} \right\|_2^2 &\leq 27 \left\| J_n^{1/2} D \operatorname{Cov}(X_1(\phi_n)) D J_n^{1/2} \right\|_2^2 \\ &+ 27 \left\| J_n^{1/2} D \operatorname{Cov}(Z_1(\phi_n)) D J_n^{1/2} \right\|_2^2 \\ &+ 27 \left\| J_n^{1/2} D \operatorname{Cov}(Z_2(\phi_n)) D J_n^{1/2} \right\|_2^2 \end{split}$$

Obviously,  $tr(J_n^2) \lesssim n^{3/2}$ . Term for term, we obtain by Lemma A.6 (i) and (A.0.12)

$$\begin{aligned} \left\| J_n^{1/2} D \operatorname{Cov} \left( X_1(\phi_n) \right) D J_n^{1/2} \right\|_2^2 &\leq \lambda_1^2 \left( \operatorname{Cov} \left( X_1(\phi_n) \right) \right) \operatorname{tr} \left( J_n^2 \right) \lesssim n^{-1/2}, \\ \left\| J_n^{1/2} D \operatorname{Cov} \left( Z_1(\phi_n) \right) D J_n^{1/2} \right\|_2^2 &\leq \lambda_1^2 \left( \operatorname{Cov} \left( Z_1(\phi_n) \right) \right) \operatorname{tr} \left( J_n^2 \right) \ll n^{-1/2}, \\ \left\| J_n^{1/2} D \operatorname{Cov} \left( Z_2(\phi_n) \right) D J_n^{1/2} \right\|_2^2 &\leq \lambda_1^2 \left( D J_n D \right) \operatorname{tr} \left( \operatorname{Cov} \left( Z_2(\phi_n) \right)^2 \right) \\ &\lesssim n \left\| \operatorname{Cov} \left( Z_2(\phi_n) \right) \right\|_2^2 \lesssim n^{-1/2}, \end{aligned}$$

uniformly over  $\phi_n \in \Phi_n(\kappa, C)$ ,  $\sigma \in S(\kappa, Q_1)$ . Therefore,

$$\sup_{\phi_n \in \Phi_n(\kappa,C), \ \sigma \in S(\kappa,Q_1)} \operatorname{Var}\left(\xi^t \Xi^{1/2}(\phi_n) D J_n D \ \Xi^{1/2}(\phi_n)\xi\right) \lesssim n^{-1/2}.$$

II: We may apply Lemma A.9 (i). Since  $\max_{i=1,\dots,n} \operatorname{Cum}_4(\epsilon_i) \leq ||\tau||_{\infty}^4 \mathbb{E}[\epsilon^4]$  we have uniformly over  $\tau \in T(\kappa, Q_2)$ 

$$\begin{aligned} \operatorname{Var}\left(\epsilon^{t} \Delta^{t} I(\phi_{n}^{1/2}) D J_{n} D I(\phi_{n}^{1/2}) \Delta \epsilon\right) \\ &\lesssim \left\|\operatorname{Cov}(\epsilon)^{1/2} \Delta^{t} I(\phi_{n}^{1/2}) D J_{n} D I(\phi_{n}^{1/2}) \Delta \operatorname{Cov}(\epsilon)^{1/2}\right\|_{2}^{2} + \left\|\Delta^{t} I(\phi_{n}^{1/2}) D J_{n} D I(\phi_{n}^{1/2}) \Delta\right\|_{2}^{2} \\ &\lesssim \left\|\Delta^{t} I(\phi_{n}^{1/2}) D J_{n} D I(\phi_{n}^{1/2}) \Delta\right\|_{2}^{2} = \left\|J_{n}^{1/2} D I(\phi_{n}^{1/2}) A I(\phi_{n}^{1/2}) D J_{n}^{1/2}\right\|_{2}^{2}, \end{aligned}$$

where A is as defined in (4.2.8). In analogy to (A.0.7) and (A.0.8) we define  $\widetilde{T}(\phi_n) \in \mathbb{M}_{n-1}$  via

$$\left(\widetilde{T}(\phi_n)\right)_{i,j} = \begin{cases} \left(\Delta_i \phi_n^{1/2}\right)^2 & \text{for } i = j - 1, \\ \left(\Delta_j \phi_n^{1/2}\right)^2 & \text{for } i = j + 1, \\ 0 & \text{otherwise} \end{cases}$$

and obtain  $I(\phi_n^{1/2})AI(\phi_n^{1/2}) = \frac{1}{2}I(\phi_n)A + \frac{1}{2}AI(\phi_n) + \frac{1}{2}\widetilde{T}(\phi_n)$ . This allows us to bound further

$$\left\|J_{n}^{1/2}DI(\phi_{n}^{1/2})AI(\phi_{n}^{1/2})DJ_{n}^{1/2}\right\|_{2}^{2} \leq 2\left\|J_{n}^{1/2}DI(\phi_{n})ADJ_{n}^{1/2}\right\|_{2}^{2} + \frac{1}{2}\left\|J_{n}^{1/2}D\widetilde{T}(\phi_{n})DJ_{n}^{1/2}\right\|_{2}^{2}.$$

Due to  $\Lambda J_n \Lambda \leq \frac{n}{M} \|k\|_{\infty} \lambda_M^2 I_n$ , we have

$$\begin{split} \left\| J_n^{1/2} DI(\phi_n) A D J_n^{1/2} \right\|_2^2 &= \operatorname{tr} \left( J_n^{1/2} DI(\phi_n) D \Lambda J_n \Lambda DI(\phi_n) D J_n^{1/2} \right) \\ &\leq \frac{n}{M} \left\| k \right\|_{\infty} \lambda_M^2 \operatorname{tr} \left( J_n^{1/2} DI(\phi_n^2) D J_n^{1/2} \right) \lesssim \frac{n}{M} \lambda_M^2 \operatorname{tr} (J_n) \lesssim n^{-1/2}. \end{split}$$

Also

$$\sup_{\phi_n \in \Phi_n(\kappa,C)} \|J_n^{1/2} D\widetilde{T}(\phi_n) D J_n^{1/2}\|_2^2 \le \lambda_1^2 (J_n) \|\widetilde{T}(\phi_n)\|_2^2 \lesssim n^{-1/2}$$

and therefore

$$\sup_{\phi_n \in \Phi_n(\kappa,C), \ \tau \in T(\kappa,Q_2)} \operatorname{Var}\left(\epsilon^t \Delta^t I(\phi_n^{1/2}) D J_n D I(\phi_n^{1/2}) \Delta \epsilon\right) \lesssim n^{-1/2}.$$

III: Let

$$P_{1,n} := \Delta^t I(\phi_n^{1/2}) DJ_n^{1/2}, \qquad P_{2,n} := J_n^{1/2} D\Xi^{1/2}(\phi_n)$$

We see that  $\xi$  and  $\epsilon$  satisfy the conditions of Lemma A.9 (iii). So we may bound using Lemma A.6 (ii)

$$\begin{aligned} \operatorname{Var}\left(\epsilon^{t} \Delta^{t} I(\phi_{n}^{1/2}) D J_{n} D \Xi^{1/2}(\phi_{n}) \xi\right) \\ &\lesssim \operatorname{tr}(P_{2,n}^{t} P_{1,n}^{t} P_{1,n} P_{2,n}) \leq \|P_{1,n}^{t} P_{1,n}\|_{2} \|P_{2,n}^{t} P_{2,n}\|_{2} \leq \frac{1}{2} \|P_{1,n}^{t} P_{1,n}\|_{2}^{2} + \frac{1}{2} \|P_{2,n}^{t} P_{2,n}\|_{2} \\ &\leq \frac{1}{2} \|J_{n}^{1/2} D I(\phi_{n}^{1/2}) A I(\phi_{n}^{1/2}) D J_{n}^{1/2}\|_{2}^{2} + \frac{1}{2} \|J_{n}^{1/2} D \Xi(\phi_{n}) D J_{n}^{1/2}\|_{2}^{2}. \end{aligned}$$

But since the r.h.s. has been bounded in I and II, we conclude that

$$\sup_{\phi_n \in \Phi_n(\kappa,C), \ \sigma \in S(\kappa,Q_1), \ \tau \in T(\kappa,Q_2)} \operatorname{Var}\left(\epsilon^t \Delta^t I(\phi_n^{1/2}) D J_n D \Xi^{1/2}(\phi_n) \xi\right) \lesssim n^{-1/2}.$$

Combining I, II and III gives

$$\sup_{\phi_n \in \Phi_n(\kappa,C), \ \sigma \in S(\kappa,Q_1), \ \tau \in T(\kappa,Q_2)} \operatorname{Var}\left( (\Delta Y(\phi_n))^t D J_n D(\Delta Y(\phi_n)) \right) \lesssim n^{-1/2}$$

and together with (A.0.14) this proves (4.2.17).

**Lemma A.2.** Let  $\lambda_i$  be as defined in (4.2.9),  $M = \lfloor cn^{1/2} \rfloor$  and suppose that  $k : [0,1] \rightarrow \mathbb{R}$  is piecewise Lipschitz continuous. Then,

$$(i) \quad \frac{n}{M} \sum_{i=1}^{M} k(\frac{i}{M}) \lambda_{i} = \pi^{2} c^{2} \int_{0}^{1} k(x) x^{2} dx + O(n^{-1/2}),$$

$$(ii) \quad \frac{1}{M} \sum_{i=1}^{M} k^{2}(\frac{i}{M}) = \int_{0}^{1} k^{2}(x) dx + O(n^{-1/2}),$$

$$(iii) \quad \frac{n}{M} \sum_{i=1}^{M} \lambda_{i} k^{2}(\frac{i}{M}) = \pi^{2} c^{2} \int_{0}^{1} k^{2}(x) x^{2} dx + O(n^{-1/2}),$$

$$(iv) \quad \frac{n^{2}}{M} \sum_{i=1}^{M} \lambda_{i}^{2} k^{2}(\frac{i}{M}) = \pi^{4} c^{4} \int_{0}^{1} k^{2}(x) x^{4} dx + O(n^{-1/2}).$$

*Proof.* (i): Let  $x_i := i\pi/(2n)$ . By series expansion  $\sin^2(x_i) = x_i^2 - \xi_i^4/3$ , where  $\xi_i \in (0, x_i)$ . Furthermore,  $\max_{i=1,\dots,M} x_i \leq n^{-1/2}$ , implying

$$\frac{n}{M}\sum_{i=1}^{M}k(\frac{i}{M})\xi_{i}^{4} \lesssim n^{-1}.$$

The mapping k is piecewise Lipschitz continuous and so is  $x \mapsto k(x)x^2$  as a function on [0, 1]. This shows

$$\left|\frac{1}{M}\sum_{i=1}^{M}k(\frac{i}{M})\left(\frac{i}{M}\right)^{2} - \int_{0}^{1}k(x)x^{2}dx\right| \lesssim n^{-1/2}.$$

107

Moreover,  $||k||_{\infty} \lesssim 1$ . Therefore,

$$\frac{n}{M} \sum_{i=1}^{M} k(\frac{i}{M}) \lambda_{i} = \frac{4n}{M} \sum_{i=1}^{M} k(\frac{i}{M}) \left(\frac{i\pi}{2n}\right)^{2} + O(n^{-1}) = \frac{M^{2}}{n} \pi^{2} \frac{1}{M} \sum_{i=1}^{M} k(\frac{i}{M}) \left(\frac{i}{M}\right)^{2} + O(n^{-1})$$
$$= \pi^{2} c^{2} \int_{0}^{1} k(x) x^{2} dx + O(n^{-1/2}).$$
(A.0.15)

(*iii*): If  $k : [0,1] \to \mathbb{R}$  is piecewise Lipschitz continuous and bounded then this is true for  $k^2$  as well and hence (*iii*) follows from (*i*).

(iv): We argue as in (i). Clearly,  $\sin^4(x_i) = x_i^4 - 2\xi_i^6/3$ ,  $\xi_i \in (0, x_i)$ , and  $\frac{n^2}{M} \sum_{i=1}^M k^2(\frac{i}{M}) \xi_i^6 \lesssim n^{-1}$ . Since  $x \mapsto k^2(x) x^4$  is piecewise Lipschitz again, we may conclude the proof by Riemann summation as in (A.0.15).

**Lemma A.3.** Let  $f \in L^2[0,1]$ ,  $\sum |f_p| < \infty$ , where  $f_p = \int_0^1 f(x) \cos(p\pi x) dx$ . Suppose that  $I_n(f) \in \mathbb{D}_{n-1}$  is defined as in (4.2.20).

(i) Then,

$$\frac{1}{n}\sum_{r=1}^{n-1} f\left(\frac{r}{n}\right) \cos\left(\frac{pr\pi}{n}\right) = A\left(f,p\right) - \frac{1}{2n}\left(\left(-1\right)^p f(1) + f(0)\right),$$

where A is as introduced in (4.2.19) and

(*ii*) 
$$(DI_n(f)D)_{i,j} = A(f, i-j) - A(f, i+j).$$

*Proof.* (i) Note that we can write

$$f\left(\frac{r}{n}\right) = f_0 + 2\sum_{q=1}^{\infty} f_q \cos\left(\frac{q\pi r}{n}\right)$$

and hence it holds

$$\frac{1}{n}\sum_{r=1}^{n-1} f\left(\frac{r}{n}\right)\cos\left(\frac{pr\pi}{n}\right) = \frac{1}{n}f_0\sum_{r=1}^{n-1}\cos\left(\frac{pr\pi}{n}\right) + \frac{2}{n}\sum_{q=1}^{\infty}f_q\sum_{r=1}^{n-1}\cos\left(\frac{q\pi r}{n}\right)\cos\left(\frac{pr\pi}{n}\right).$$

By Lemma A.11 (ii) and

$$2\sum_{r=1}^{n-1}\cos\left(\frac{q\pi r}{n}\right)\cos\left(\frac{pr\pi}{n}\right) = \operatorname{Dir}_{n-1}\left(\frac{(q-p)\pi}{n}\right) + \operatorname{Dir}_{n-1}\left(\frac{(q+p)\pi}{n}\right) - 1$$

it follows

$$\frac{1}{n}\sum_{r=1}^{n-1} f\left(\frac{r}{n}\right) \cos\left(\frac{pr\pi}{n}\right) = \frac{1}{n} \left[-\frac{1}{2}\left(1 + (-1)^p\right)f_0 - \sum_{q=1}^{\infty} f_q\left(1 + (-1)^{q-p}\right)\right] + A\left(f,p\right),$$
which yields the result.

(ii) This follows by applying (i) to

$$(DI(f)D)_{i,j} = \frac{2}{n} \sum_{r=1}^{n-1} f\left(\frac{r}{n}\right) \sin\left(\frac{ir\pi}{n}\right) \sin\left(\frac{rj\pi}{n}\right)$$
$$= \frac{1}{n} \sum_{r=1}^{n-1} f\left(\frac{r}{n}\right) \cos\left(\frac{(i-j)r\pi}{n}\right) - \frac{1}{n} \sum_{r=1}^{n-1} f\left(\frac{r}{n}\right) \cos\left(\frac{(i+j)r\pi}{n}\right).$$

**Lemma A.4** (Products). Let  $f, g \in L^2[0,1], \sum_l |f_l| + |g_l| < \infty$ . Then,

$$|(fg)_i| \le \Big|\sum_{l=-\infty}^{\infty} f_l g_{l-i}\Big|,$$

where  $f_p, g_p, (fg)_p$  are as in (4.2.18). *Proof.* Decompose  $f(x) = f_0 + 2\sum_{k=1}^{\infty} f_k \cos(k\pi x), \ x \in [0, 1]$ . Then,

$$(fg)_{i} = \int_{0}^{1} (fg)(x) \cos(i\pi x) dx$$
  
=  $f_{0}g_{i} + \sum_{k=1}^{\infty} f_{k} \int_{0}^{1} g(x) [\cos((k+i)\pi x) + \cos((k-i)\pi x)] dx \le \sum_{l=-\infty}^{\infty} f_{l}g_{l-i}.$ 

The next Lemma gives a bound on the sum of the absolute values of Fourier coefficients in Sobolev ellipsoids  $\Theta_{\cos}(\alpha, Q)$  as defined in Section 2.4. In particular, the result shows that the Fourier series is absolutely summable.

**Lemma A.5.** Let  $f_p$  be as defined in (4.2.18). Assume  $\alpha > 1/2$ , c > 0 and  $\gamma \ge 0$ . Then, it holds for n large enough

$$\sup_{f \in \Theta_{\cos}(\alpha,Q)} \sum_{m=\lfloor cn^{\gamma} \rfloor}^{\infty} |f_m| \le C_{\alpha,Q,c} n^{\gamma(1/2-\alpha)},$$

where  $C_{\alpha,Q,c}$  is independent of n.

*Proof.* We see that

$$\sum_{m=\lfloor cn^{\gamma} \rfloor}^{\infty} |f_m| = \sum_{m=1}^{\infty} |f_m| \mathbb{I}_{\{m \ge \lfloor cn^{\gamma} \rfloor\}}$$
$$\leq \frac{1}{\sqrt{2}} \left( \sum_{i=1}^{\infty} i^{2\alpha} 2f_i^2 \right)^{1/2} \left( \sum_{i=\lfloor cn^{\gamma} \rfloor}^{\infty} i^{-2\alpha} \right)^{1/2} \leq C_{\alpha,Q,c} \ n^{\gamma(1/2-\alpha)},$$

where we used the definition of  $\Theta_{\cos}(\alpha, Q)$  in the last step.

109

**Proposition A.1.** Let  $A \in \mathbb{M}_{n-1}$  and assume that k satisfies Assumption 3. Then, for sufficiently large n,

$$\left| \operatorname{tr} \left( J_n D A D \right) \right| \lesssim \max_{i,j} \left| \left( A \right)_{i,j} \right| \, n^{3/2} \log n.$$

*Proof.* Write  $A = (a_{i,j})_{i,j=1,\dots,n-1}$ . Note that

$$(DAD)_{i,j} = \frac{2}{n} \sum_{p,q=1}^{n-1} \sin\left(\frac{ip\pi}{n}\right) \sin\left(\frac{qj\pi}{n}\right) a_{p,q}$$

In particular, for i = j we have

$$(DAD)_{i,i} = \frac{1}{n} \sum_{p,q=1}^{n-1} a_{p,q} \cos\left((p-q) \frac{i\pi}{n}\right) - \frac{1}{n} \sum_{p,q=1}^{n-1} a_{p,q} \cos\left((p+q) \frac{i\pi}{n}\right).$$
(A.0.16)

Furthermore, we need the following bounds using Lemma A.12 (i)

$$\left|\sum_{i=1}^{M} k(\frac{i}{M}) \cos(r\frac{i\pi}{n})\right| \leq 2 \sum_{j=0}^{\infty} |k_j| \left|\sum_{i=1}^{M} \cos(j\pi\frac{i}{M}) \cos(r\frac{i\pi}{n})\right|$$
$$\leq 2 \sum_{j=0}^{\infty} |k_j| \left[ \left(1 + \kappa(\frac{r}{n} - \frac{j}{M}) + \kappa(\frac{r}{n} + \frac{j}{M})\right) \wedge M \right].$$

This allows us to find by using Lemma A.12 (ii) and Lemma A.5

$$\begin{split} \big| \sum_{i=1}^{M} k(\frac{i}{M}) \frac{1}{n} \sum_{p,q=1}^{n-1} a_{p,q} \cos((p-q)\frac{i\pi}{n}) \big| \\ &\leq \frac{2}{n} \sum_{p,q=1}^{n-1} |a_{p,q}| \sum_{j=0}^{\infty} |k_j| \Big[ \Big( 1 + \kappa (\frac{|p-q|}{n} - \frac{j}{M}) + \kappa (\frac{|p-q|}{n} + \frac{j}{M}) \Big) \wedge M \Big] \\ &\leq 2 \max_{p,q=1,\dots,n-1} |a_{p,q}| \sum_{j=0}^{\infty} |k_j| \sum_{r=0}^{n} \Big[ \Big( 1 + \kappa (\frac{r}{n} - \frac{j}{M}) + \kappa (\frac{r}{n} + \frac{j}{M}) \Big) \wedge M \Big] \\ &\lesssim \max_{p,q=1,\dots,n-1} |a_{p,q}| \ n \log n \sum_{j=0}^{\infty} |k_j| \lesssim \max_{p,q=1,\dots,n-1} |a_{p,q}| \ n \log n. \end{split}$$

By repeating the arguments above, we derive also

$$\big|\sum_{i=1}^{M} k(\frac{i}{M}) \ \frac{1}{n} \sum_{p,q=1}^{n-1} a_{p,q} \cos((p+q)\frac{i\pi}{n})\big| \lesssim \max_{p,q=1,\dots,n-1} |a_{p,q}| \ n \log n.$$

Combining the previous inequalities yields

$$\operatorname{tr}(J_n DAD) = \frac{n}{M} \sum_{i=1}^M k(\frac{i}{M}) (DAD)_{i,i} \lesssim \max_{p,q=1,\dots,n-1} |a_{p,q}| \ n^{3/2} \log n.$$

**Proof of "** $\supseteq$ **" in (2.4.4).** We show that  $f \in W_{\cos}(\alpha, \pi^{2\alpha}C) \Rightarrow f \in \Theta_{\cos}(\alpha, C)$ . Let  $\tilde{f}$  denote the extension of f on [-1, 1] defined by

$$\tilde{f}(x) := \begin{cases} f(x) & \text{for } x \in [0, 1], \\ f(-x) & \text{for } x \in [-1, 0]. \end{cases}$$

Note that  $\tilde{f}$  is an  $\alpha$ -times differentiable function with  $f^{(\alpha)} \in L^2$  and  $\tilde{f}^{(l)}$  is an even/odd function whenever l is even/odd. Let

$$\theta_k(j) = \begin{cases} \int_{-1}^1 \tilde{f}^{(j)}(x) dx & \text{for } k = 0, \\ \int_{-1}^1 \tilde{f}^{(j)}(x) \cos(k\pi x) dx & \text{for } k \ge 1, \ j \text{ even}, \\ \int_{-1}^1 \tilde{f}^{(j)}(x) \sin(k\pi x) dx & \text{for } k \ge 1, \ j \text{ odd}. \end{cases}$$

It holds for  $j \ge 1$ 

$$\theta_0(j) = \int_{-1}^1 \tilde{f}^{(j)}(x) dx = \tilde{f}^{(j-1)}(1) - \tilde{f}^{(j-1)}(-1) = 0,$$

where we used the facts that  $\tilde{f}^{(j-1)}$  is even for j odd and  $\tilde{f}^{(j-1)}(1) = \tilde{f}^{(j-1)}(-1) = 0$ for j even, due to  $\tilde{f} \in W_{\cos}(\alpha, \pi^{2\alpha}C)$ . Now,  $(f^{(\alpha)})^2$  is even for all integer  $\alpha$ . Hence, by Parseval's equality

$$\left\|f^{(\alpha)}\right\|_{L^{2}[0,1]}^{2} = \frac{1}{2}\left\|f^{(\alpha)}\right\|_{L^{2}[-1,1]}^{2} = \frac{1}{2}\sum_{k=1}^{\infty}\theta_{k}^{2}(\alpha).$$
(A.0.17)

Furthermore, for  $k \ge 1$ , j even, it follows by partial integration

$$\theta_k(j) = \int_{-1}^1 \tilde{f}^{(j)}(x) \cos(k\pi x) dx$$
  
=  $\tilde{f}^{(j-1)}(x) \cos(k\pi x) \Big|_{-1}^1 + k\pi \int_{-1}^1 \tilde{f}^{(j-1)}(x) \sin(k\pi x) dx = k\pi \ \theta_k(j-1)$ 

and for  $k \geq 1$  and j odd

$$\theta_k(j) = \int_{-1}^1 \tilde{f}^{(j)}(x) \sin(k\pi x) dx$$
  
=  $\tilde{f}^{(j-1)}(x) \sin(k\pi x) \Big|_{-1}^1 - k\pi \int_{-1}^1 \tilde{f}^{(j-1)}(x) \cos(k\pi x) dx = -k\pi \ \theta_k(j-1).$ 

For  $k \ge 1$  we obtain  $\theta_k^2(\alpha) = k^{2\alpha} \pi^{2\alpha} \theta_k^2(0) = 2k^{2\alpha} \pi^{2\alpha} f_k^2$ , where  $f_k = \int_0^1 f(x) \cos(k\pi x) dx$ . Combining this result with (A.0.17) yields

$$\left\|f^{(\alpha)}\right\|_{L^{2}[0,1]}^{2} = \pi^{2\alpha} \sum_{k=1}^{\infty} k^{2\alpha} f_{k}^{2}$$

and hence proves the claim.

### A.1. Some facts about multivariate statistics and linear algebra

#### Lemma A.6.

- (i) Let  $A, B \in \mathbb{M}_n$  are symmetric and positive semidefinite matrices. Denote by  $\lambda_1(A)$  the largest eigenvalue of A. Then,  $\operatorname{tr}(AB) \leq \lambda_1(A) \operatorname{tr}(B)$ .
- (ii) Let A and B matrices of the same size. Then,

 $\left|\operatorname{tr}\left(AB^{t}\right)\right| \leq \operatorname{tr}^{1/2}\left(AA^{t}\right)\operatorname{tr}^{1/2}\left(BB^{t}\right).$ 

In the following Lemma, we summarize some facts on Frobenius norms.

Lemma A.7. Let  $A = (a_{i,j})_{i,j=1,\dots,n} \in \mathbb{M}_n$ . Then,

(i)

$$|A||_{2}^{2} := \operatorname{tr}(AA^{t}) = \sum_{i=1}^{n} \lambda_{i}(AA^{t}) = \sum_{i,j=1}^{n} a_{i,j}^{2}$$

and whenever  $A = A^t$  also  $||A||_2^2 = \sum_{i=1}^n \lambda_i^2(A)$ .

(ii) It holds

$$4 \operatorname{tr} (A^2) \le ||A + A^t||_2^2 \le 4 ||A||_2^2$$

(iii) Let  $A, B \in \mathbb{M}_n$  are symmetric and positive semidefinite matrices such that  $0 \leq A \leq B$ . Furthermore, suppose that  $X \in \mathbb{M}_{n,m}$ . Then,

$$\left\|X^{t}AX\right\|_{2} \leq \left\|X^{t}BX\right\|_{2}.$$

(iv) Let  $\Lambda \in \mathbb{D}_n$  and  $X \in \mathbb{M}_{n,m}$ . Then,

 $\left\|X^{t}\Lambda X\right\|_{2} \leq 2 \max_{i=1,\dots,n} \left|(\Lambda)_{i,i}\right| \left\|X^{t}X\right\|_{2}.$ 

*Proof.* (i) and (ii) are well known and omitted. (iii) By assumptions it holds  $0 \leq X^{t}AX \leq X^{t}BX$ . Hence,  $\lambda_{i}^{2}(X^{t}AX) \leq \lambda_{i}^{2}(X^{t}BX)$  and the result follows. (iv) Let  $\Lambda_{+} := \Lambda \lor 0$  and  $\Lambda_{-} := \Lambda \land 0$ . Then,  $\|X^{t}\Lambda X\|_{2} \leq \|X^{t}\Lambda_{+}X\|_{2} + \|X^{t}\Lambda_{-}X\|_{2} \leq 2 \max_{i=1,\dots,n} |(\Lambda)_{i,i}| \|X^{t}X\|_{2}$ .

**Lemma A.8** (Expectation of quadratic forms). Let  $V = (V_1, \ldots, V_n)^t$  and  $W = (W_1, \ldots, W_m)^t$  be two centered random vectors and  $B \in \mathbb{M}_{n,m}$ . Then,  $\mathbb{E}(V^t B W) = \text{tr}(B \operatorname{Cov}(V, W)^t)$ .

Important special cases of Lemma A.8 are Cov(V, W) = 0, i.e. the random vectors are uncorrelated and hence  $\mathbb{E}(V^t B W) = 0$  as well as V = W implying  $\mathbb{E}(V^t B V) = tr(B Cov(V))$ .

*Proof.*  $V^t BW = \sum_{k=1}^n \sum_{l=1}^m V_k B_{k,l} W_l$ . By taking expectations the result follows.  $\Box$ If X, Y are independent random vectors, we write  $X \perp Y$ .

**Lemma A.9** (Variance of quadratic forms). Assume that we are in the same setting as in Lemma A.8. Moreover, let  $A = (a_{i,j})_{i,j=1,\dots,n} \in \mathbb{M}_n$ .

(i) If  $V_i \perp V_j$  whenever  $i, j = 1, \ldots, n, i \neq j$ , then

$$\operatorname{Var}(V^{t}AV) = \sum_{i=1}^{n} \operatorname{Cum}_{4}(V_{i})a_{ii}^{2} + \operatorname{tr}(\operatorname{Cov}(V)A\operatorname{Cov}(V)A + \operatorname{Cov}(V)A\operatorname{Cov}(V)A^{t})$$
  
$$\leq 2\|\operatorname{Cov}(V)^{1/2}A\operatorname{Cov}(V)^{1/2}\|_{2}^{2} + \max_{i=1,\dots,n}\operatorname{Cum}_{4}(V_{i})\|A\|_{2}^{2},$$

where  $\operatorname{Cum}_4(V_i)$  denotes the fourth cumulant of  $V_i$ .

(ii) If V is multivariate Gaussian then

$$\operatorname{Var}\left(V^{t}AV\right) = \operatorname{tr}(\operatorname{Cov}(V)A\operatorname{Cov}(V)A + \operatorname{Cov}(V)A\operatorname{Cov}(V)A^{t})$$
$$\leq 2\|\operatorname{Cov}(V)^{1/2}A\operatorname{Cov}(V)^{1/2}\|_{2}^{2},$$

where equality holds if A is symmetric.

(iii) Assume  $V_i = U_i \epsilon_i$  for i = 1, ..., n. Let  $\epsilon_i \perp \epsilon_j$  for i, j = 1, ..., n,  $i \neq j$  as well as  $W_k \perp W_l$  for k, l = 1, ..., m,  $k \neq l$ . If  $\epsilon = (\epsilon_1, ..., \epsilon_n)$  is independent of  $U = (U_1, ..., U_n)$  and W then

$$\operatorname{Var}\left(V^{t}BW\right) \leq C^{2}\operatorname{tr}\left(B^{t}\operatorname{Cov}(\epsilon)B\operatorname{Cov}(W)\right) = C^{2}\left\|\operatorname{Cov}(\epsilon)^{1/2}B\operatorname{Cov}(W)^{1/2}\right\|_{2}^{2}$$

provided  $\max_i |U_i| \leq C$  and  $\mathbb{E}\epsilon = 0$ .

*Proof.* Note that

$$\operatorname{Var}\left(V^{t}BW\right) = \sum_{i,k=1}^{n} \sum_{j,l=1}^{m} b_{ij}b_{kl}\operatorname{Cov}\left(V_{i}W_{j}, V_{k}W_{l}\right).$$
(A.1.1)

(i): First, observe that

$$\operatorname{Cov}(V_i V_j, V_k V_l) = \begin{cases} \mathbb{E}V_i^4 - \mathbb{E}^2 V_i^2, & \text{for } i = j = k = l, \\ \mathbb{E}V_i^2 \mathbb{E}V_j^2, & \text{for } i = k \neq j = l \text{ or } i = l \neq j = k, \\ 0, & \text{otherwise.} \end{cases}$$

Using (A.1.1), we see that (m = n, V = W)

$$\operatorname{Var}\left(V^{t}AV\right) = \sum_{i=1}^{n} (\mathbb{E}V_{i}^{4} - \mathbb{E}^{2}V_{i}^{2})a_{ii}^{2} + \sum_{i,j=1, i\neq j}^{n} \mathbb{E}V_{i}^{2}a_{ij}\mathbb{E}V_{j}^{2}a_{ji} + \sum_{i,j=1, i\neq j}^{n} \mathbb{E}V_{i}^{2}a_{ij}^{2}\mathbb{E}V_{j}^{2}.$$

Finally with  $\operatorname{Cum}_4(V_i) = \mathbb{E}V_i^4 - 3\mathbb{E}^2V_i^2$  and Lemma A.7 (ii) the result follows.

(*ii*): By assumption  $V \sim \mathcal{N}(0, \operatorname{Cov}(V))$ . Then, we may write  $V \stackrel{\mathcal{D}}{=} \operatorname{Cov}(V)^{1/2}\xi$ , where  $\xi \sim \mathcal{N}(0, I_n)$ . Hence, by (i)

$$\operatorname{Var}\left(V^{t}AV\right) = \operatorname{Var}\left(\xi^{t}\tilde{A}\xi\right) = \operatorname{tr}(\tilde{A}^{2} + \tilde{A}\tilde{A}^{t}).$$

for  $\widetilde{A} = \operatorname{Cov}(V)^{1/2} A \operatorname{Cov}(V)^{1/2}$  and due to  $\operatorname{Cum}_4(\xi_i) = 0$ . (*iii*): Write  $B = (b_{i,j})_{i=1,\dots,n, j=1,\dots,m} \in \mathbb{M}_{n,m}$ . We derive

$$\mathbb{E}[V_i W_j] = \mathbb{E}[U_i W_j] \mathbb{E}[\epsilon_i] = 0,$$
  
$$\mathbb{E}[V_i W_j V_k W_l] = \mathbb{E}[U_i U_k W_j W_l] \mathbb{E}[\epsilon_i \epsilon_k] = \mathbb{E}[U_i^2 W_j W_l] \mathbb{E}[\epsilon_i^2] \delta_{i,k}$$

for  $i, k = 1, \ldots, n$  and  $j, l = 1, \ldots, m$ . Therefore,

$$Var(V^{t}BW) = \sum_{i=1}^{n} \sum_{j,l=1}^{m} b_{i,j}b_{i,l}\mathbb{E}[U_{i}^{2}W_{j}W_{l}]\mathbb{E}[\epsilon_{i}^{2}]$$
  
$$= \sum_{i=1}^{n} \mathbb{E}[\epsilon_{i}^{2}]\mathbb{E}[U_{i}^{2}(\sum_{j=1}^{m} b_{i,j}W_{j})^{2}] \leq C^{2} \sum_{i=1}^{n} \mathbb{E}[\epsilon_{i}^{2}] \sum_{j,l=1}^{m} b_{i,j}b_{i,l}\mathbb{E}[W_{j}W_{l}]$$
  
$$= C^{2} \sum_{i=1}^{n} \mathbb{E}[\epsilon_{i}^{2}] \sum_{j=1}^{m} b_{i,j}^{2}\mathbb{E}[W_{j}^{2}] = C^{2} \operatorname{tr}(B^{t} \operatorname{Cov}(\epsilon)B\operatorname{Cov}(W))$$

r	-	-	-	
L				
L				
	-	_	-	

**Lemma A.10.** Let X, Y, Z be random vectors of the same size. Then,

$$\operatorname{Cov}(X, Y) + \operatorname{Cov}(Y, X) \le \operatorname{Cov}(X) + \operatorname{Cov}(Y)$$

in the sense of Loewner ordering. In particular,  $\operatorname{Cov}(X+Y) \leq 2\operatorname{Cov}(X) + 2\operatorname{Cov}(Y)$ and  $\operatorname{Cov}(X+Y+Z) \leq 3\operatorname{Cov}(X) + 3\operatorname{Cov}(Y) + 3\operatorname{Cov}(Z)$ .

*Proof.*  $\operatorname{Cov}(X - Y) = \operatorname{Cov}(X) + \operatorname{Cov}(Y) - \operatorname{Cov}(X, Y) - \operatorname{Cov}(Y, X)$ . Since  $\operatorname{Cov}(X, Y) + \operatorname{Cov}(Y, X)$  is a symmetric matrix and  $\operatorname{Cov}(X - Y) \ge 0$ , the result follows.  $\Box$ 

**Definition 20** (Dirichlet kernel). The function  $\text{Dir}_N : \mathbb{R} \to \mathbb{R}$ ,

$$\operatorname{Dir}_{N}(x) = \frac{1}{2} + \sum_{i=1}^{N} \cos(ix) = \frac{1}{2} \sum_{i=-N}^{N} \cos(ix)$$

is called the Dirichlet kernel.

Lemma A.11 (Properties of the Dirichlet kernel).

(i) Whenever  $x \notin 2\pi\mathbb{Z}$ ,

$$\operatorname{Dir}_N(x) = \frac{\sin((N + \frac{1}{2})x)}{2\sin(\frac{x}{2})}.$$

(ii) Let  $\mathbb{I}_{\{A\}}(\cdot)$  denote the indicator function on the set A and suppose that p is an integer. Then,

$$\operatorname{Dir}_{n-1}(\frac{p\pi}{n}) = n \mathbb{I}_{\{p \equiv 0 \mod 2n\}}(p) + \frac{1}{2} (-1)^{p+1}.$$

For any  $s \in \mathbb{R}$  there is a unique t such that  $-1 < t \leq 3$ , s = t + 4k and  $k \in \mathbb{Z}$ . For given s this will be denoted by  $\{s\}_4$ .

Lemma A.12. Define  $\kappa : \mathbb{R} \to \overline{\mathbb{R}}$ ,

$$\kappa(s) := \begin{cases} \frac{1}{|\{s\}_4|}, & -1 \le \{s\}_4 \le 1, \\ \frac{1}{|2-\{s\}_4|}, & 1 \le \{s\}_4 \le 3. \end{cases}$$
(A.1.2)

(i) Then,

$$2\bigg|\sum_{i=1}^{M}\cos(i\pi\frac{j}{M})\cos(i\pi\frac{r}{n})\bigg| \le \left(1+\kappa\left(\frac{r}{n}-\frac{j}{M}\right)+\kappa\left(\frac{r}{n}+\frac{j}{M}\right)\right) \wedge M.$$

(ii) For any  $s \in \mathbb{R}$ ,

$$\sum_{r=0}^{n} \kappa(\frac{r}{n} + s) \wedge M \le 2M + 2\sum_{l=1}^{n} \frac{n}{l} \le 2M + 2n(1 + \log n).$$

*Proof. (i):* First note that for  $s \in [-1,3]$ ,  $|\sum_{i=1}^{M} \cos(i\pi s)| \le \frac{1}{2} + |\sin^{-1}(\pi s/2)| \le \frac{1}{2} + \kappa(s)$ . Now the result follows by extending s to the real line and observing

$$\sum_{i=1}^{M} \cos(i\pi \frac{j}{M}) \cos(i\pi \frac{r}{n}) = \frac{1}{2} \sum_{i=1}^{M} \cos(i\pi (\frac{r}{n} - \frac{j}{M})) + \frac{1}{2} \sum_{i=1}^{M} \cos(i\pi (\frac{r}{n} + \frac{j}{M})).$$

г		
L		
L		
L		

# Appendix B

## Proofs and technical results for Chapter 5

Before we can give the proofs, let us define some quantities needed in the sequel. For a continuously differentiable function  $g:[0,1] \to \mathbb{R}$  we set

$$\mathfrak{R}_{n}(g) := \left(\sum_{j=1}^{n} \int_{(j-1)/n}^{j/n} \left(\frac{1}{n} \sum_{l=j}^{n} g'(\frac{l}{n}) - \int_{s}^{1} g'(u) du\right)^{2} ds\right)^{1/2}.$$
 (B.0.1)

Recall (5.2.7) and let |supp()| denote the support length.

**Lemma B.1.** Suppose that h satisfies Assumption 4. Then, the following estimates hold uniformly in  $\ell, k$ .

(i)  $|\operatorname{supp}(h_{\ell k})| \lesssim 2^{-\ell}$ .

(*ii*) 
$$||h_{\ell k}||_p + ||h_{\ell k}||_{p,M} \lesssim 2^{\ell(1/2 - 1/p)}$$
.

(*iii*) 
$$\sum_{i=1}^{M} \sup_{s,t \in [(i-1)/M, i/M]} |h_{\ell k}(s) - h_{\ell k}(t)| \lesssim 2^{3\ell/2} |\operatorname{supp}(h_{\ell k})| \lesssim 2^{\ell/2}.$$

*Proof.* (i): Since h has bounded support, so does  $h(2^{\ell} - k)$  and  $|\operatorname{supp}(h_{\ell k})| = |\operatorname{supp}(h)|2^{-\ell}$ . (ii):  $||h_{\ell k}||_p^p = 2^{\ell p/2} \int_0^1 |h(2^{\ell}x - k)|^p dx \le 2^{\ell(p/2-1)} ||h||_p^p$ . Similarly,

$$\|h_{\ell k}\|_{p,M}^p = M^{-1} 2^{\ell p/2} \sum_j |h(2^{\ell} \frac{j}{m} - k)| \le 2^{\ell p/2} |\operatorname{supp}(h_{\ell k})| = 2^{\ell (p/2-1)}.$$

**Lemma B.2** (Summation by parts). For two sequences  $(a_k)_k$  and  $(b_k)_k$  we have

$$\sum_{k=1}^{n-1} a_k (b_{k+1} - b_k) = a_{n-1} b_n - a_1 b_1 - \sum_{k=2}^{n-1} (a_k - a_{k-1}) b_k.$$

#### Preliminary results for the semimartingale X

The first lemma gives a moment bound on the approximation error of discretized stochastic integrals.

**Lemma B.3** (Discretization effect). Suppose Assumption 6 holds. For deterministic functions  $g, h : [0, 1] \to \mathbb{R}$ , with piecewise continuous derivative and g(1) = 0, h(1) = 0, we obtain for any  $p \ge 1$  and C > 0,

$$\mathbb{E}_{L^{\infty}(C)} \left[ \left| \left( \frac{1}{n} \sum_{i=1}^{n} g'(\frac{i}{n}) X_{i/n} \right) \left( \frac{1}{n} \sum_{i=1}^{n} h'(\frac{i}{n}) X_{i/n} \right) - \left( \int_{0}^{1} g(s) dX_{s} \right) \left( \int_{0}^{1} h(s) dX_{s} \right) \right|^{p} \right] \\ \lesssim \|g\|_{2}^{p} \mathfrak{R}_{n}^{p}(h) + \|h\|_{2}^{p} \mathfrak{R}_{n}^{p}(g) + \mathfrak{R}_{n}^{p}(g) \mathfrak{R}_{n}^{p}(h). \quad (B.0.2)$$

*Proof.* Let us point out as a first step that by Assumption 6 and a suitable change of measure, we may assume that X is a local martingale. Recall that we work in a filtered probability space  $(\Omega, (\mathcal{F}_t)_{t\geq 0}, \mathcal{F}, \mathbb{P})$ . Now, we can write  $\mathbb{P} = \mathbb{P}_{\sigma,b} \otimes \widetilde{\mathbb{P}}$  where  $\mathbb{P}_{\sigma,b}$  denotes the law of the process X. By assumption,  $\mathbb{P}_{\sigma,b}$  and  $\mathbb{P}_{\sigma,0}$  are equivalent on  $\mathcal{F}_1$  and the Radon-Nikodym derivative is given by

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

Define  $\mathbb{P}_0 = \mathbb{P}_{\sigma,0} \otimes \widetilde{\mathbb{P}}$  and denote the expectation under  $\mathbb{P}$  and  $\mathbb{P}_0$  by  $\mathbb{E}_b$  and  $\mathbb{E}_0$ , respectively. Then, for a sequence of random variables  $(Z_n)_n$ , we may bound by Hölder inequality and for  $\rho > 1$  as in Assumption 6,

$$\begin{split} \mathbb{E}_{b}^{1/\gamma} \big[ |Z_{n}|^{\gamma} \big] &= \mathbb{E}_{0}^{1/\gamma} \Big[ \frac{d \mathbb{P}_{\sigma,b}}{d \mathbb{P}_{\sigma,0}} |Z_{n}|^{\gamma} \Big] \lesssim \mathbb{E}_{0}^{1/\gamma\rho} \Big[ \exp\left(\rho \int_{0}^{1} \frac{b_{s}}{\sigma_{s}^{2}} dX_{s}\right) \Big] \mathbb{E}_{0}^{(\rho-1)/(\gamma\rho)} \big[ |Z_{n}|^{\gamma\rho/(\rho-1)} \big] \\ &\lesssim \mathbb{E}_{0}^{1/\gamma\rho} \big[ |Z_{n}|^{\gamma\rho} \big], \end{split}$$

where  $\gamma_{\rho} := \gamma \rho / (\rho - 1)$ . This shows that by a change of measure we may assume that the drift is zero.

In order to prove the remaining part, note that for random variables  $U_1, U_2, V_1$  and  $V_2$ ,

$$\begin{aligned} |U_1 U_2 - V_1 V_2| &\leq |U_1| |U_2 - V_2| + |V_2| |U_1 - V_1| \\ &\leq |U_1| |U_2 - V_2| + |U_2| |U_1 - V_1| + |U_2 - V_2| |U_1 - V_1| \end{aligned}$$

and therefore by Cauchy-Schwarz

$$\begin{split} \mathbb{E}_{L^{\infty}(C)} \left[ |U_{1}U_{2} - V_{1}V_{2}|^{p} \right] &\lesssim \mathbb{E}_{L^{\infty}(C)}^{1/2} \left[ |U_{1}|^{2p} \right] \mathbb{E}_{L^{\infty}(C)}^{1/2} \left[ |U_{2} - V_{2}|^{2p} \right] \\ &+ \mathbb{E}_{L^{\infty}(C)}^{1/2} \left[ |U_{2}|^{2p} \right] \mathbb{E}_{L^{\infty}(C)}^{1/2} \left[ |U_{1} - V_{1}|^{2p} \right] \\ &+ \mathbb{E}_{L^{\infty}(C)}^{1/2} \left[ |U_{1} - V_{1}|^{2p} \right] \mathbb{E}_{L^{\infty}(C)}^{1/2} \left[ |U_{2} - V_{2}|^{2p} \right]. \end{split}$$

Integration by part yields for  $f \in \{g, h\}$ ,

$$\int_{0}^{T_{C}} f'(s) X_{s} ds = -\int_{0}^{T_{C}} f(s) dX_{s}$$
(B.0.3)

due to  $f(1) = X_0 = 0$ . Hence, we may bound the l.h.s. of (B.0.2) by a multiple of  $I_g \times II_h + I_h \times II_g + II_g \times II_h$  with

$$I_{f} := \mathbb{E}_{L^{\infty}(C)}^{1/2} \left[ \left| \int_{0}^{1} f'(s) X_{s} ds \right|^{2p} \right],$$
  
$$II_{f} := \mathbb{E}_{L^{\infty}(C)}^{1/2} \left[ \left| \frac{1}{n} \sum_{j=1}^{n} f'(\frac{j}{n}) X_{j/n} - \int_{0}^{1} f'(s) X_{s} ds \right|^{2p} \right], \quad \text{for } f \in \{g, h\}.$$

 ${\cal I}_f$  : Let us introduce the stopping time

$$T_C := \inf\{s \ge 0, \ \sigma_s^2 > C\} \land 1.$$
 (B.0.4)

On the event  $\{\sigma^2 \in L^{\infty}(C)\}$ , we have  $T_C = 1$ . Therefore,

$$\mathbb{E}_{L^{\infty}(C)}\left[\left|\int_{0}^{1} f'(s)X_{s}ds\right|^{2p}\right] = \mathbb{E}\left[\left|\int_{0}^{T_{C}} f'(s)X_{s}ds\right|^{2p}\mathbb{I}_{\sigma^{2}\in L^{\infty}(C)}\right]$$
$$\leq \mathbb{E}\left[\left|\int_{0}^{T_{C}} f'(s)X_{s}ds\right|^{2p}\right].$$

Now, using Burkholder-Davis-Gundy inequality (later abbreviated by BDG, the statement is recalled in Lemma B.17), we obtain

$$I_{f} \leq \mathbb{E}\Big[\Big|\int_{0}^{T_{C}} f(s)dX_{s}\Big|^{2p}\Big]^{1/2} \lesssim \mathbb{E}\Big[\Big|\int_{0}^{T_{C}} f^{2}(s)\sigma_{s}^{2}ds\Big|^{p}\Big]^{1/2} \lesssim \|f\|_{2}^{p},$$

since for  $s \leq T_C$  we have the upper bound  $\sigma_s^2 \leq C$ .

 $II_f$ : Let us first introduce  $\widetilde{f}: [0,1] \to \mathbb{R}$ , where

$$\widetilde{f}(s) := \sum_{j=1}^n \left(\frac{1}{n} \sum_{l=j}^n f'\left(\frac{l}{n}\right)\right) \mathbb{I}_{[(j-1)/n, j/n)}(s).$$

Note that by summation by parts (Lemma B.2), we have the identity

$$\int_0^{t\wedge T_c} \widetilde{f}(s) dX_s = \frac{1}{n} \sum_{j=1}^n f'\left(\frac{j}{n}\right) X_{j/n}.$$
(B.0.5)

Moreover, the process  $S_t = \int_0^{t \wedge T_c} \left( \widetilde{f}(s) + f(s) \right) dX_s, t \in [0, 1]$  is a martingale and

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

From (B.0.3) and (B.0.5) it follows that

$$II_f = \mathbb{E}_{L^{\infty}(C)} \left[ |S_1|^{2p} \right]^{1/2} \lesssim \mathbb{E} \left[ \langle S \rangle_{T_C}^p \right]^{1/2} \lesssim \mathfrak{R}_n^p(f),$$

where  $\mathfrak{R}_n$  as defined in (B.0.1). This finishes the proof.

Lemma B.4. Let  $\mathfrak{R}_n$  and  $\Lambda$  as defined in (B.0.1) and (5.2.8), respectively. Then,  $\mathfrak{R}_n [\Lambda (M \cdot -(i-2))] \lesssim n^{-1}$ 

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

$$\|\Lambda (M \cdot -(i-2))\|_2 = M^{-1/2}.$$

*Proof.* Let  $j_i$  be the smallest j such that  $j/n \ge i/M$ . Obviously,

$$\begin{split} &\max_{\substack{j \ n \in \left(\frac{i-2}{M}, \frac{i}{M}\right] \ s \in \left[\frac{j-1}{n}, \frac{j}{n}\right]}} \sup_{s \in \left[\frac{j-1}{n}, \frac{j}{n}\right]} \left| \frac{1}{n} \sum_{l=j}^{j_i} \widetilde{\lambda} \left( M \frac{l}{n} - (i-2) \right) - \int_s^1 \widetilde{\lambda} \left( M u - (i-2) \right) du \right| \\ &\leq \max_{\substack{j \ n \in \left(\frac{i-2}{M}, \frac{i}{M}\right] \ s \in \left[\frac{j-1}{n}, \frac{j}{n}\right]}} \sup_{s \in \left[\frac{j-1}{n}, \frac{j}{n}\right]} \left| \int_s^{(j-1)/n} \widetilde{\lambda} \left( M u - (i-2) \right) du \right| \\ &+ \max_{\substack{j \ n \in \left(\frac{i-2}{M}, \frac{i}{M}\right] \ l=j}} \sum_{l=j}^{j_i} \left| \frac{1}{n} \widetilde{\lambda} \left( M \frac{l}{n} - (i-2) \right) - \int_{(l-1)/n}^{l/n} \widetilde{\lambda} \left( M u - (i-2) \right) du \right| \lesssim n^{-1}, \end{split}$$

proving the first part of the lemma. In order to derive the second part, it suffices to show

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

Then, (5.2.10) yields the result. By definition  $\int_0^2 \tilde{\lambda}(u) du = 0$ . Therefore, the second statement follows from

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

120

**Lemma B.5.** Suppose Assumption 6 holds. Let  $\Lambda$  be defined as in (5.2.8). Then, for any  $p \ge 1$  and C > 0, we have

$$\mathbb{E}_{L^{\infty}(C)} \left[ \left| \sum_{i=2}^{M} h_{\ell k} \left( \frac{i-1}{M} \right) \left( \int_{0}^{1} \Lambda \left( Ms - (i-2) \right) dX_{s} \right)^{2} - \int_{0}^{1} \sum_{i=2}^{M} h_{\ell k} \left( \frac{i-1}{M} \right) \Lambda^{2} (Ms - (i-2)) d\langle X \rangle_{s} \right|^{p} \right] \lesssim \|h_{\ell k}\|_{\infty}^{p} |\operatorname{supp}(h_{\ell k})|^{p/2} M^{-p/2},$$

uniformly in  $\ell$ , k. Recall that |supp(.)| denotes the support length.

*Proof.* As done in Lemma B.3, we can force by a change of measure that X is a local martingale. Let us define stochastic processes  $H_{t,i}$  for i = 2, ..., M and  $t \in [0, 1]$  via

$$H_{t,i} := h_{\ell k} \left(\frac{i-1}{M}\right) \Lambda \left(Mt - (i-2)\right) \int_{(i-2)/M}^{t} \Lambda \left(Ms - (i-2)\right) dX_s \,\mathbb{I}_{\left(\frac{i-2}{M}, \frac{i}{M}\right]}(t). \tag{B.0.6}$$

Applying the integration by parts formula for semimartingales (Lemma B.14), we can write

$$\sum_{i=2}^{M} h_{\ell k} \left(\frac{i-1}{M}\right) \left[ \left( \int_{0}^{1} \Lambda \left( Ms - (i-2) \right) dX_{s} \right)^{2} - \int_{0}^{1} \Lambda^{2} \left( Ms - (i-2) \right) d\langle X \rangle_{s} \right] \\ = 2 \sum_{i=2}^{M} \int_{(i-2)/M}^{i/M} H_{t,i} \, dX_{t}.$$
(B.0.7)

The sum over the processes, i.e.  $\sum_{i=2}^{M} H_{t,i}$ , is continuous on [0,1] (because of  $\Lambda(0) = \Lambda(2) = 0$ ) and adapted, therefore the integral  $\int_{0}^{t} \sum_{i=2}^{M} H_{s,i} dX_s$  is a continuous local martingale. Note that  $t \rightsquigarrow H_{t,i}$  is supported on  $[\frac{i-2}{M}, \frac{i}{M}]$ . Hence, for any  $t \in [0,1]$  there are at most two different *i*, such that  $H_{t,i}$  is not vanishing. Now, by BDG and the same stopping time argument as in the proof of Lemma B.3, we derive

$$\mathbb{E}_{L^{\infty}(C)} \Big[ \Big| \int_{0}^{T_{C}} \sum_{i=2}^{M} H_{t,i} \, dX_{s} \Big|^{p} \Big] \\ \lesssim \mathbb{E} \Big[ \Big| \int_{0}^{T_{C}} \Big( \sum_{i=2}^{M} H_{t,i} \, dt \Big)^{2} \Big|^{p/2} \Big] \lesssim \mathbb{E} \Big[ \Big| \int_{0}^{T_{C}} \sum_{i=2}^{M} H_{t,i}^{2} \, dt \Big|^{p/2} \Big] \\ \lesssim \mathbb{E} \Big[ \Big| M^{-1} \sum_{i=2}^{M} (H_{i}^{\star})^{2} \Big|^{p/2} \Big] \lesssim |\operatorname{supp}(h_{\ell k})|^{p/2-1} M^{-1} \sum_{i=2}^{M} \mathbb{E} \Big[ (H_{i}^{\star})^{p} \Big],$$

where we used  $H_i^{\star} := \sup_{t \leq T_C} |H_{t,i}|$  and Hölder inequality for the last estimate. Applying

BDG again, we obtain

$$\mathbb{E}\left[\left(H_{i}^{\star}\right)^{p}\right] \lesssim \left|h_{\ell k}\left(\frac{i-1}{M}\right)\right|^{p} \mathbb{E}\left[\sup_{t \leq 2/M} \left|\int_{(i-2)/M \wedge T_{C}}^{((i-2)/M+t) \wedge T_{C}} \Lambda\left(Ms-(i-2)\right) dX_{s}\right|^{p}\right] \\ \lesssim \left|h_{\ell k}\left(\frac{i-1}{M}\right)\right|^{p} \mathbb{E}\left[\left(\int_{(i-2)/M \wedge T_{C}}^{T_{C}} \Lambda^{2}\left(Ms-(i-2)\right) \sigma_{s}^{2} ds\right)^{p/2}\right] \\ \lesssim \left|h_{\ell k}\left(\frac{i-1}{M}\right)\right|^{p} M^{-p/2}.$$
(B.0.8)

This completes the proof.

#### Preliminary results for the microstructure noise $\epsilon$

Recall from (5.2.2) that

$$\bar{\epsilon}_{i,M} := \bar{\epsilon}_{i,M}(\lambda) := \frac{M}{n} \sum_{\substack{j \ i \in \left(\frac{i-2}{M}, \frac{i}{M}\right]}} \tilde{\lambda} \left(M\frac{j}{n} - (i-2)\right) \epsilon_{j,n}, \quad i = 2, \dots, M.$$

Since  $\tau$  is a process, we write  $\tau_t := \tau(X_t, t)$  for  $t \in [0, 1]$ .

**Lemma B.6.** Suppose Assumptions 5 and 6 hold. Denote by  $\mathcal{G}$  the  $\sigma$ -field generated by  $(X_s, s \in [0, 1])$ , i.e.  $\mathcal{G} = \sigma(X_s : s \leq 1)$ . Then, for any  $p \geq 1$ ,

$$\mathbb{E}\Big[\Big|\sum_{i=1}^{M}h_{\ell k}\left(\frac{i-1}{M}\right)\left(\overline{\epsilon}_{i,M}^{2}(\lambda)-\mathbb{E}\left[\overline{\epsilon}_{i,M}^{2}(\lambda)\left|\mathcal{G}\right]\right)\Big|^{p}\Big] \lesssim \|h_{\ell k}\|_{2,M}^{p}M^{-3p/2}n^{-p}+\|h_{\ell k}\|_{p,M}^{p}M^{p+1}n^{-p}$$

uniformly in  $\ell, k$ .

*Proof.* First, we need to introduce the filtrations

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

Furthermore, let

$$U_i := h_{\ell k} \left(\frac{i-1}{M}\right) \left(\overline{\epsilon}_{i,M}^2 - \frac{M^2}{n^2} \sum_{\substack{\underline{j} \\ n} \in \left(\frac{i-2}{M}, \frac{i}{M}\right]} \widetilde{\lambda}^2 \left(M \frac{\underline{j}}{n} - (i-2)\right) \tau_{\underline{j}/n}^2\right).$$

First,  $U_{2i}$  and  $U_{2i+1}$  are  $\mathcal{F}_i^{\text{even}}$  and  $\mathcal{F}_i^{\text{odd}}$  measurable. Second,  $\mathbb{E}\left[U_{2i+1}|\mathcal{F}_{i-1}^{\text{even}}\right] = 0$  and  $\mathbb{E}\left[U_{2i}|\mathcal{F}_{i-1}^{\text{odd}}\right] = 0$ . Thus, the partial sums  $S_r^{\text{even}} := \sum_{i=1}^r U_{2i}$  and  $S_r^{\text{odd}} := \sum_{i=1}^r U_{2i+1}$  form martingale schemes  $(i = 1, \ldots, r \leq \lfloor M/2 \rfloor)$  with respect to the filtrations  $\mathcal{F}_r^{\text{even}}$  and  $\mathcal{F}_r^{\text{odd}}$ . By triangle inequality it follows, for any  $p \geq 1$ 

$$\mathbb{E}\left[\left|\overline{\epsilon}_{i,M}^{2} - \frac{M^{2}}{n^{2}} \sum_{\substack{j \\ n \in \left(\frac{i-2}{M}, \frac{i}{M}\right]}} \widetilde{\lambda}^{2} \left(M \frac{j}{n} - (i-2)\right) \tau_{j/n}^{2}\Big|^{p}\right] \\
\lesssim \mathbb{E}\left[\left|\overline{\epsilon}_{i,M}\right|^{2p}\right] + \|\widetilde{\lambda}\|_{\infty}^{2p} \|\tau\|_{\infty}^{2p} M^{p} n^{-p} \lesssim M^{p} n^{-p}.$$

Consequently,

$$\mathbb{E}\left[|U_i|^p\right] \lesssim |h_{\ell k}(\frac{i-1}{M})|^p M^p n^{-p}.$$
(B.0.9)

Analogous computations show that

$$\mathbb{E}\left[U_{2i}^2 \mid \mathcal{F}_{i-1}^{\text{even}}\right] \le h_{\ell k}^2 \left(\frac{2i-1}{M}\right) \mathbb{E}\left[\overline{\epsilon}_{2i,M}^4 \mid \mathcal{F}_{i-1}^{\text{even}}\right] \lesssim h_{\ell k}^2 \left(\frac{2i-1}{M}\right) M^2 n^{-2}.$$

Finally, applying Rosenthal's inequality for martingales (see Lemma B.18) yields

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

By similar arguments, we can obtain for  $\mathbb{E}[|S_{\lfloor (M-1)/2 \rfloor}^{\text{odd}}|^p]$  the same bound as for  $\mathbb{E}[|S_{\lfloor M/2 \rfloor}^{\text{even}}|^p]$ . This finishes the proof.

Recall the definition of the stopping time  $T_C$  given in (B.0.4). Let us introduce the "stopped" pre-averaging

$$\overline{X}_{i,M,T_C}(\lambda) := \frac{M}{n} \sum_{\substack{\underline{j} \in \left(\frac{i-2}{M}, \frac{i}{M}\right]}} \widetilde{\lambda} \left(M_{\underline{j}}^{\underline{j}} - (i-2)\right) X_{j/n \wedge T_C}.$$
(B.0.10)

**Lemma B.7.** Suppose Assumptions 5 and 6 hold. Then, we have, for any  $p \ge 1$  and C > 0

$$\mathbb{E}_{L^{\infty}(C)} \left[ \left| \sum_{i=2}^{M} h_{\ell k} \left( \frac{i-1}{M} \right) \overline{X}_{i,M}(\lambda) \,\overline{\epsilon}_{i,M}(\lambda) \right|^{p} \right] \\ \lesssim \|h_{\ell k}\|_{p,M}^{p} \left( n^{-p/2} M + M^{3p/2+1} n^{-3p/2} \right) + \|h_{\ell k}\|_{2,M}^{p} \left( M^{p/2} n^{-p/2} + M^{2p} n^{-3p/2} \right),$$

uniformly in  $\ell, k$ .

*Proof.* Again we may assume that X is a local martingale. Note that

$$\mathbb{E}_{L^{\infty}(C)}\left[\left|\sum_{i=2}^{M}h_{\ell k}\left(\frac{i-1}{M}\right)\overline{X}_{i,M}(\lambda)\overline{\epsilon}_{i,M}(\lambda)\right|^{p}\right] \leq \mathbb{E}\left[\left|\sum_{i=2}^{M}h_{\ell k}\left(\frac{i-1}{M}\right)\overline{X}_{i,M,T_{C}}(\lambda)\overline{\epsilon}_{i,M}(\lambda)\right|^{p}\right].$$

Arguing as in the proof of Lemma B.6, we see that

$$S_r^{\text{even}} := \sum_{i=1}^r h_{\ell k} \left(\frac{2i-1}{M}\right) \overline{X}_{2i,M,T_C}(\lambda) \overline{\epsilon}_{2i,M}(\lambda)$$

and

$$S_r^{\text{odd}} := \sum_{i=1}^r h_{\ell k} \left(\frac{2i}{M}\right) \overline{X}_{2i+1,M,T_C}(\lambda) \overline{\epsilon}_{2i+1,M}(\lambda)$$

are martingales with respect to the filtrations  $\mathcal{F}^{even}$  and  $\mathcal{F}^{odd}$ . By Rosenthal's inequality for martingales, we can bound

$$\mathbb{E}\left[\left|S_{\lfloor M/2 \rfloor}^{\text{even}}\right|^{p}\right] \lesssim M^{p/2} n^{-p/2} \mathbb{E}\left[\left|\sum_{i=1}^{\lfloor M/2 \rfloor} h_{\ell k}^{2} \left(\frac{2i-1}{M}\right) \mathbb{E}\left[\overline{X}_{2i,M,T_{C}}^{2}(\lambda) \mid \mathcal{F}_{i-1}^{\text{even}}\right]\right|^{p/2}\right] + \sum_{i=1}^{\lfloor M/2 \rfloor} \left|h_{\ell k} \left(\frac{2i-1}{M}\right)\right|^{p} \left(\mathbb{E}\left[|\overline{X}_{2i,M,T_{C}}(\lambda)|^{2p}\right]\right)^{1/2} \left(\mathbb{E}\left[|\overline{\epsilon}_{2i,M}(\lambda)|^{2p}\right]\right)^{1/2},$$

using  $\|\tau\|_{\infty} \lesssim 1$  and Cauchy-Schwarz inequality for the second term.

By Riemann summation,

$$\left|\sum_{\substack{j \in \left(\frac{i-2}{M}, \frac{i}{M}\right]\\ = \left|\sum_{\substack{j \in \left(\frac{i-2}{M}, \frac{i}{M}\right]}} \widetilde{\lambda}\left(M\frac{j}{n} - (i-2)\right) - n \int_{(i-2)/M}^{i/M} \widetilde{\lambda}(Mu - (i-2))du\right| \lesssim 1.$$
(B.0.11)

Therefore, we derive

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

Consequently,

$$\mathbb{E}\left[\left|\overline{X}_{i,M,T_{C}}(\lambda)\right|^{2p}\right] \leq \|\widetilde{\lambda}\|_{\infty}^{2p} \mathbb{E}\left[\sup_{s\leq 2/M} |X_{(i-2)/M+s\wedge T_{C}} - X_{(i-2)/M\wedge T_{C}}|^{2p}\right] + M^{2p} n^{-2p} \mathbb{E}\left[|X_{(i-2)/M\wedge T_{C}}|^{2p}\right]$$

which can be further bounded by a constant times  $\|\widetilde{\lambda}\|_{\infty}^{2p}M^{-p} + M^{2p}n^{-2p}$  thanks to the localization argument for  $\sigma$ . In a similar way, we have

$$\mathbb{E}\left[\overline{X}_{2i,M,T_{C}}^{2}(\lambda) \mid \mathcal{F}_{i-1}^{\text{even}}\right] \lesssim M^{-1} + M^{2}n^{-2}X_{(2i-2)/M\wedge T_{C}}^{2} \leq M^{-1} + M^{2}n^{-2}\sup_{s}X_{s}^{2}.$$

Recall that  $\mathbb{E}[|\bar{\epsilon}_{i,M}|^{2p}] \leq M^p n^{-p}$ . Putting these estimates together, we infer that  $\mathbb{E}[|S_{\lfloor M/2 \rfloor}^{\text{even}}|^p]$  satisfies the desired bound. By similar arguments we can obtain the same bound for  $S_{\lfloor (M-1)/2 \rfloor}^{\text{odd}}$ . The conclusion follows.

### Preliminary results for bias correction

Recall that the bias correction was defined in (5.2.5) by

$$\mathfrak{b}(\lambda,Y)_{i,M} := \frac{M^2}{2n^2} \sum_{\substack{\underline{j} \in \left(\frac{i-2}{M}, \frac{i}{M}\right]}} \widetilde{\lambda}^2 \left(M \frac{\underline{j}}{n} - (i-2)\right) \left(Y_{j,n} - Y_{j-1,n}\right)^2, \quad Y_{0,n} := 0.$$

We will use the decomposition

$$\mathfrak{b}(\lambda,Y)_{i,M} = \mathfrak{b}(\lambda,X)_{i,M} + \mathfrak{b}(\lambda,\epsilon)_{i,M} + 2\mathfrak{c}(\lambda,X,\epsilon)_{i,M}$$

extensively, where

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

**Lemma B.8.** Suppose Assumptions 5 and 6 hold. For any  $p \ge 1$  and C > 0, we have

$$\mathbb{E} \Big[ \Big| \sum_{i=2}^{M} h_{\ell k} \Big( \frac{i-1}{M} \Big) \Big( \mathfrak{b}(\lambda, \epsilon)_{i,M} - \frac{M^2}{n^2} \sum_{\substack{\underline{j} \\ n \in \Big( \frac{i-2}{M}, \frac{i}{M} \Big]}} \widetilde{\lambda}^2 \Big( M \frac{\underline{j}}{n} - (i-2) \Big) \tau_{\underline{j}/n}^2 \Big) \Big|^p \Big] \\ \lesssim \|h_{\ell k}\|_{1,M}^p M^{3p} n^{-2p} + \|h_{\ell k}\|_{2,M}^p M^{2p} n^{-3p/2} + \|h_{\ell k}\|_{p,M}^p M^{2p} n^{-2p+1},$$

uniformly over  $\ell, k$ .

*Proof.* By triangle inequality, we bound the error by a multiple (only dependent on p) of

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

with

$$I := \mathbb{E} \left[ \left| \sum_{i=2}^{M} h_{\ell k} \left( \frac{i-1}{M} \right) \sum_{j} \widetilde{\lambda}^{2} \left( M \frac{j}{n} - (i-2) \right) \tau_{j/n}^{2} \left( \eta_{j,n}^{2} - 1 \right) \right|^{p} \right],$$
  

$$II := \mathbb{E} \left[ \left| \sum_{i=2}^{M} h_{\ell k} \left( \frac{i-1}{M} \right) \sum_{j} \widetilde{\lambda}^{2} \left( M \frac{j}{n} - (i-2) \right) \tau_{(j-1)/n}^{2} \left( \eta_{j-1,n}^{2} - 1 \right) \right|^{p} \right],$$
  

$$III := \mathbb{E} \left[ \left| \sum_{i=2}^{M} h_{\ell k} \left( \frac{i-1}{M} \right) \sum_{j} \widetilde{\lambda}^{2} \left( M \frac{j}{n} - (i-2) \right) \left( \tau_{j/n}^{2} - \tau_{(j-1)/n}^{2} \right) \right|^{p} \right],$$
  

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

and, as before, the sum in j is taken over the sets  $\{j/n \in ((i-2)/M, i/M]\}$ .

*I* and *II*: It suffices to bound *I*. Likewise we will obtain the same bound for *II*. Define the filtration  $\mathcal{F}_r = \sigma(\eta_{k,n} : k \leq r) \otimes \sigma(X_s : s \leq 1)$ . Now,

$$S_r := \sum_{j=1}^r \Big( \sum_{i=2}^M \left| h_{\ell k} \left( \frac{i-1}{M} \right) \right|^p \mathbb{I}_{\left\{ \frac{j}{n} \in \left( \frac{i-2}{M}, \frac{i}{M} \right] \right\}} \Big) \widetilde{\lambda}^2 \Big( M \frac{j}{n} - (i-2) \Big) \tau_{j/n}^2 \Big( \eta_{j,n}^2 - 1 \Big)$$

forms an  $\mathcal{F}_r$  martingale scheme. Thus, from Rosenthal's inequality for martingales, we derive

$$\begin{split} I \lesssim &\sum_{j=1}^{n} \left( \sum_{i=2}^{M} \left| h_{\ell k} \left( \frac{i-1}{M} \right) \right|^{p} \mathbb{I}_{\left\{ \frac{j}{n} \in \left( \frac{i-2}{M}, \frac{i}{M} \right] \right\}} \right) \mathbb{E} \left[ \left| \left( \eta_{j,n}^{2} - 1 \right) \right|^{p} \right] \\ &+ \left| \sum_{j=1}^{n} \sum_{i=2}^{M} h_{\ell k}^{2} \left( \frac{i-1}{M} \right) \mathbb{I}_{\left\{ \frac{j}{n} \in \left( \frac{i-2}{M}, \frac{i}{M} \right] \right\}} \mathbb{E} \left[ \left( \eta_{j,n}^{2} - 1 \right)^{2} \right| \mathcal{F}_{j-1} \right] \right|^{p/2}, \\ &\lesssim \| h_{\ell k} \|_{p,M}^{p} n + \| h_{\ell k} \|_{2,M}^{p} n^{p/2}. \end{split}$$

using that the functions  $\tau$  and  $\tilde{\lambda}$  are bounded.

III : By summation by parts (Lemma B.2), we obtain

$$\begin{split} &\sum_{\substack{\underline{j}\\n\in\left(\frac{i-2}{M},\frac{i}{M}\right]\\}} \widetilde{\lambda}^2 \left(M\frac{\underline{j}}{n} - (i-2)\right) \left(\tau_{\underline{j}/n}^2 - \tau_{(\underline{j}-1)/n}^2\right) \\ &= -\sum_{\substack{\underline{j}\\n\in\left(\frac{i-2}{M},\frac{i}{M}\right]\\}} \tau_{(\underline{j}-1)/n}^2 \left(\widetilde{\lambda}^2 \left(M\frac{\underline{j}}{n} - (i-2)\right) - \widetilde{\lambda}^2 \left(M\frac{\underline{j}-1}{n} - (i-2)\right)\right) \\ &+ \tau_{(\underline{i}-2)/M}^2 \widetilde{\lambda}^2(0) - \tau_{\underline{i}/M}^2 \widetilde{\lambda}^2(2). \end{split}$$

By assumption  $\tau$  is bounded and  $\tilde{\lambda}$  is of finite variation. Therefore, we have

$$\Big|\sum_{i=2}^{M} h_{\ell k} \left(\frac{i-1}{M}\right) \sum_{\substack{j \ n \in \left(\frac{i-2}{M}, \frac{i}{M}\right]}} \widetilde{\lambda}^{2} \left(M_{n}^{j} - (i-2)\right) \left(\tau_{j/n}^{2} - \tau_{(j-1)/n}^{2}\right)\Big|^{p} \lesssim \|h_{\ell k}\|_{1,M}^{p} M^{p}.$$

IV: The sum can be splitted into two parts by summing over even and odd j, respectively. Proceeding as for I and II, we derive

$$IV \lesssim \|h_{\ell k}\|_{2,M}^p n^{p/2} + \|h_{\ell k}\|_{p,M}^p n.$$

**Lemma B.9.** Suppose Assumptions 5 and 6 hold. For any  $p \ge 1$  and C > 0, we have

$$\mathbb{E}_{L^{\infty}(C)}\left[\left|\sum_{i=2}^{M}h_{\ell k}\left(\frac{i-1}{M}\right)\mathfrak{b}(\lambda,X)_{i,M}\right|^{p}\right] \lesssim \|h_{\ell k}\|_{1,M}^{p}M^{p}n^{-p},$$

uniformly in  $\ell, k$ .

*Proof.* Recall the definition of the stopped pre-averaging of X given in (B.0.10). Then,

$$\mathbb{E}_{L^{\infty}(C)}\left[\left|\sum_{i=2}^{M}h_{\ell k}\left(\frac{i-1}{M}\right)\mathfrak{b}(\lambda,X)_{i,M}\right|^{p}\right] \leq \mathbb{E}\left[\left|\sum_{i=2}^{M}h_{\ell k}\left(\frac{i-1}{M}\right)\mathfrak{b}(\lambda,X_{\cdot\wedge T_{C}})_{i,M}\right|^{p}\right]$$
$$\lesssim M^{2p}n^{-2p}\mathbb{E}\left[\left|\sum_{i=2}^{M}\left|h_{\ell k}\left(\frac{i-1}{M}\right)\right|\sum_{\substack{j\\n}\in\left(\frac{i-2}{M},\frac{i}{M}\right]}(X_{j/n\wedge T_{C}}-X_{(i-2)/M\wedge T_{C}})^{2}\right|^{p}\right]$$
$$\lesssim \|h_{\ell k}\|_{1,M}^{p}M^{p}n^{-p}.$$

**Lemma B.10.** Suppose Assumptions 5 and 6 hold. For any  $p \ge 1$  and C > 0, we have

$$\mathbb{E}_{L^{\infty}(C)} \left[ \left| \sum_{i=2}^{M} h_{\ell k} \left( \frac{i-1}{M} \right) \mathfrak{c}(\lambda, X, \epsilon)_{i,M} \right|^{p} \right] \\ \lesssim \left[ \|h_{\ell k}\|_{2,M}^{p} + \|h_{\ell k}\|_{p,M}^{p} (n^{-p/2+1} + M^{-p/2+1}) \right] M^{2p} n^{-2p}$$

uniformly over  $\ell, k$ .

Proof. Similar as in Lemmas B.7 and B.9, we can bound

$$\mathbb{E}_{L^{\infty}(C)} \left[ \left| \sum_{i=2}^{M} h_{\ell k} \left( \frac{i-1}{M} \right) \mathfrak{c}(\lambda, X, \epsilon)_{i,M} \right|^{p} \right] \\ \leq \mathbb{E} \left[ \left| \sum_{i=1}^{M} h_{\ell k} \left( \frac{i-1}{M} \right) \sum_{\substack{j \ n \in \left( \frac{i-2}{M}, \frac{i}{M} \right]}} \widetilde{\lambda}^{2} \left( M \frac{j}{n} - (i-2) \right) (X_{j/n \wedge T_{C}} - X_{(j-1)/n \wedge T_{C}}) \epsilon_{j,n} \right|^{p} \right].$$

Define  $j_n^{\star}(r) := \max\{j : j/n \le r/M\}$ . Let us introduce the filtrations

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

The discrete process

$$S_{r}^{\text{even}} := \sum_{i=1}^{r} h_{\ell k} \left(\frac{2i-1}{M}\right) \sum_{\substack{j \ i \in \left(\frac{2i-2}{M}, \frac{2i}{M}\right]}} \widetilde{\lambda}^{2} \left(M_{n}^{j} - (2i-2)\right) (X_{j/n \wedge T_{C}} - X_{(j-1)/n \wedge T_{C}}) \epsilon_{j,n}$$

forms a martingale scheme with respect to the filtration  $\mathcal{G}^{\text{even}}$ . In the same way we can by summation over odd indexes obtain the martingale scheme  $S_r^{\text{odd}}$  defined w.r.t. the filtration  $\mathcal{G}_r^{\text{odd}}$ . Note that

$$\mathbb{E}\left[\left|h_{\ell k}\left(\frac{i-1}{M}\right) \sum_{\substack{j \in \left(\frac{i-2}{M}, \frac{i}{M}\right]\\ n \in \left(\frac{i-2}{M}, \frac{i}{M}\right]} \widetilde{\lambda}^{2} \left(M\frac{j}{n} - (i-2)\right) (X_{j/n \wedge T_{C}} - X_{(j-1)/n \wedge T_{C}}) \epsilon_{j,n}\right|^{p}\right] \\
\lesssim \left|h_{\ell k}\left(\frac{i-1}{M}\right)\right|^{p} \left(M^{-p/2} + \sum_{\substack{j \in \left(\frac{i-2}{M}, \frac{i}{M}\right]\\ n \in \left(\frac{i-2}{M}, \frac{i}{M}\right]}} \mathbb{E}\left[\left|(X_{j/n \wedge T_{C}} - X_{(j-1)/n \wedge T_{C}}) \epsilon_{j,n}\right|^{p}\right]\right) \\
\lesssim \left|h_{\ell k}\left(\frac{i-1}{M}\right)\right|^{p} M^{-1} (M^{-p/2+1} + n^{-p/2+1}),$$

and also by conditional Itô-isometry (Lemma B.15)

$$\mathbb{E}\left[\left(h_{\ell k}\left(\frac{2i-1}{M}\right)\sum_{\substack{j \in \left(\frac{2i-2}{M}, \frac{2i}{M}\right]\\j \in \left(\frac{2i-2}{M}, \frac{2i}{M}\right]}} \widetilde{\lambda}^{2}\left(M_{n}^{j} - (2i-2)\right)(X_{j/n\wedge T_{C}} - X_{(j-1)/n\wedge T_{C}})\epsilon_{j,n}\right)^{2} \middle| \mathcal{G}_{i-1}^{\text{even}}\right] \\
\lesssim h_{\ell k}^{2}\left(\frac{2i-1}{M}\right)\sum_{\substack{j \in \left(\frac{2i-2}{M}, \frac{2i}{M}\right]\\j \in \left(\frac{2i-2}{M}, \frac{2i}{M}\right]}} \mathbb{E}\left[(X_{j/n\wedge T_{C}} - X_{(j-1)/n\wedge T_{C}})^{2} \middle| \mathcal{G}_{i-1}^{\text{even}}\right] \lesssim M^{-1}h_{\ell k}^{2}\left(\frac{2i-1}{M}\right).$$

Thus, by Rosenthal's inequality for martingales, we derive from the estimates above

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

In the same way  $S_{\lfloor (M-1)/2 \rfloor}^{\text{odd}}$  can be bounded and the result follows by incorporating the multiplicative term  $M^{2p}n^{-2p}$  from (5.2.5).

#### B.0.1. Proof of Proposition 4

In this section we adopt the notation introduced in Proposition 4. In particular, let  $\overline{C} := \overline{C}(s, \pi, C) := \sup_{f \in \mathcal{B}^s_{\pi,\infty}(C)} ||f||_{\infty}$ . The proof relies on decomposing  $\langle \widehat{h}_{\ell k}, \sigma^2 \rangle - \langle h_{\ell k}, \sigma^2 \rangle$  as in (5.2.12).

**Lemma B.11.** Suppose the assumptions of Proposition 4 hold. Then, for  $\delta > 0$ 

$$\mathbb{P}\left(\left|\sum_{i=2}^{M} h_{\ell k}\left(\frac{i-1}{M}\right) \overline{X}_{i,M}^{2}\left(\lambda\right) - \left\langle\sigma^{2}, h_{\ell k}\right\rangle_{L^{2}}\right| > 4\overline{C}\left(1+\delta\right) \sqrt{\frac{p \log M}{M}} \text{ and } \sigma^{2} \in \mathcal{B}_{\pi,\infty}^{s}\left(C\right)\right) \lesssim M^{-p}. \tag{B.0.12}$$

*Proof.* By definition  $\Lambda(s) = \int_{s}^{2} \widetilde{\lambda}(u) du$ . Moreover, let  $H_{t,i}$  be defined as in (B.0.6) and recall the integration by parts formula (B.0.7). Clearly,

$$4(1+\delta) = \delta + 4\left(1+\frac{\delta}{2}\right) + \delta.$$

Using this decomposition, we bound the l.h.s. of (B.0.12) by the sum of three terms I + II + III,

$$\begin{split} I &:= \mathbb{P}\left( \Big| \sum_{i=2}^{M} h_{\ell k} \left( \frac{i-1}{M} \right) \left( \overline{X}_{i,M}^{2}(\lambda) - \left( \int_{0}^{1} \Lambda(Ms - (i-2)) dX_{s} \right)^{2} \right) \Big| \\ &> \overline{C} \delta \sqrt{\frac{p \log M}{M}} \text{ and } \sigma^{2} \in \mathcal{B}_{\pi,\infty}^{s}(C) \right) \\ II &:= \mathbb{P}\left( \Big| \sum_{i=2}^{M} \int_{0}^{1} H_{t,i} dX_{t} \Big| > 2\overline{C} \left( 1 + \frac{\delta}{2} \right) \sqrt{\frac{p \log M}{M}} \text{ and } \sigma^{2} \in L^{\infty}\left(\overline{C}\right) \right) \\ III &:= \mathbb{P}\left( \Big| \left( \sum_{i=2}^{M} h_{\ell k} (\frac{i-1}{M}) \int_{0}^{1} \Lambda^{2} \left( Ms - (i-2) \right) \sigma_{s}^{2} ds \right) - \left\langle \sigma^{2}, h_{l k} \right\rangle_{L^{2}} \right| \\ &> \overline{C} \delta \sqrt{\frac{p \log M}{M}} \text{ and } \sigma^{2} \in \mathcal{B}_{\pi,\infty}^{s}(C) \right). \end{split}$$

We have the following generalized Chebycheff inequality:

$$\mathbb{P}(X > t \text{ and } B) = \mathbb{E}\left[\mathbb{I}_{\{X > t\} \cap B}\right] \le t^{-p} \mathbb{E}\left[X^p \ \mathbb{I}_B\right], \tag{B.0.13}$$

for  $p \ge 1$  and a measurable set *B*. Using this and (5.2.18), it follows that the term *I* can be bounded by any polynomial order of 1/M and therefore *I* is of the right order. For the second term, *II*, we aim to apply Lemma 3. In order to obtain disjoint support of the processes  $H_{t,i}$ , we further decompose  $II \le II_{even} + II_{odd}$ , where

$$II_{\text{even/odd}} := \mathbb{P}\left( \Big| \sum_{i=2, i \text{ even/odd}}^{M} \int_{0}^{T_{\overline{C}}} H_{t,i} dX_{t} \Big| > \overline{C} \left( 1 + \frac{\delta}{2} \right) \sqrt{\frac{p \log M}{M}} \right).$$

Since by Assumption 4, h has compact support, we can find an interval [-B, B], B > 0 such that h = 0 on  $\mathbb{R} \setminus [-B, B]$ . Hence,  $\frac{1}{2}(2^{-l}M(k-B)+1) \leq i \leq \frac{1}{2}(2^{-l}M(k+B)+1)$ . It suffices to treat the term  $II_{\text{even}}$  only, similar arguments apply for  $II_{\text{odd}}$ . Now, the process

$$N_r := 2^{-l/2} M \sum_{i=1}^r \int_0^{T_{\overline{C}}} H_{t,2i} dX_t$$

defines a discrete time martingale with respect to the filtration  $\mathcal{F}_r = \sigma \left( X_s : s \leq 2r/M \right)$ starting at

$$N_{\lfloor (2^{-l}M(k-B)+1)/2 \rfloor} = 0.$$

Let us remark two things. First,  $H_{t,2i}$  has support inside [2(i-1)/M, 2i/M] and second, the random variable  $\mathbb{I}_{\{T_{\overline{C}} \leq (2i-2)/M\}}$  is  $\mathcal{F}_{i-1}$  measurable. Moreover, by Lemma B.1 (iii) and  $2^l \leq M^{1-q}$ , for a q > 0, we obtain

$$\frac{2}{M} \sum_{i=1}^{\lfloor M/2 \rfloor} h_{\ell k}^2 \left(\frac{2i-1}{M}\right) = \int_0^1 h_{\ell k}^2(u) du + O\left(\frac{2^l}{M}\right) \le 1 + O(M^{-q}).$$

Thus, by using Lemma B.4 as well as conditional Itô-isometry (cf. Lemma B.15), we derive

$$\begin{split} \langle N \rangle_{\lfloor \frac{1}{2} (2^{-l}M(k+B)+1) \rfloor} \\ &\leq 2^{-l} M^2 \overline{C} \sum_{i=1}^{\lfloor M/2 \rfloor} \int_0^1 \mathbb{E} \left[ H_{s \wedge T_{\overline{C}}, 2i}^2 | \mathcal{F}_{i-1} \right] ds \\ &\leq 2^{-l} M^2 \overline{C}^2 \sum_{i=1}^{\lfloor M/2 \rfloor} h_{\ell k}^2 \left( \frac{i-1}{M} \right) \int_0^1 \Lambda^2 (Ms - (i-2)) \int_{(i-2)/M}^s \Lambda^2 (Mu - (i-2)) du \, ds \\ &= 2^{-l-1} M^2 \overline{C}^2 \sum_{i=1}^{\lfloor M/2 \rfloor} h_{\ell k}^2 \left( \frac{2i-1}{M} \right) \left( \int_0^1 \Lambda^2 (Mu - (i-2)) du \right)^2 \\ &\leq 2^{-l-1} \overline{C}^2 \sum_{i=1}^{\lfloor M/2 \rfloor} h_{\ell k}^2 \left( \frac{2i-1}{M} \right) \leq 2^{-l} M \ \frac{1}{4} \overline{C}^2 \left( 1 + \frac{\delta}{2} \right), \end{split}$$

where for the third step, we use the equality

$$\int_0^1 f'(u)f(u)du = \frac{1}{2} \left[ f^2(1) - f^2(0) \right]$$

and the last step follows for sufficiently large M. Furthermore, applying BDG, we obtain uniformly over  $i, \ell, k$ ,

$$\mathbb{E}\left[|\Delta_{i}N|^{\kappa}\right] \lesssim 2^{-l\kappa/2} M^{\kappa} \mathbb{E}\left[\left|\int_{0}^{1} H_{t,2i}\mathbb{I}_{\left[0,T_{\overline{C}}\right]}(t) dX_{t}\right|^{\kappa}\right] \lesssim 2^{-l\kappa/2} M^{\kappa} \mathbb{E}\left[\left|\int_{0}^{1} H_{t\wedge T_{\overline{C}},2i}^{2} dt\right|^{\kappa/2}\right]$$
$$\lesssim 2^{-l\kappa/2} M^{\kappa/2} \mathbb{E}\left[\sup_{t\leq 2/M} \left|H_{(t+(i-2)/M)\wedge T_{\overline{C}},2i}\right|^{\kappa}\right] \lesssim 2^{-l\kappa/2} \left|h_{\ell k}\left(\frac{i-1}{M}\right)\right|^{\kappa} \lesssim 1.$$

Now, for  $j \sim M2^{-l}$ , Lemma 3 yields that  $II_{\text{even}} \lesssim M^{-p}$ .

By repeating the arguments, we bound  $II_{\text{odd}}$  and therefore  $II \leq M^{-p}$ . For bounding the third term, III, we have by assumption  $M^{-(s-1/\pi)} \|h_{\ell k}\|_{1,M} \leq M^{-1/2}$ . Combining this with the approximations (5.2.22), (5.2.23), and (5.2.24) yields for sufficiently large M on  $\sigma^2 \in \mathcal{B}^s_{\pi,\infty}(C)$ ,

$$\left|\sum_{i=2}^{M} h_{\ell k}\left(\frac{i-1}{M}\right) \left(\int_{0}^{1} \Lambda^{2} \left(Ms - (i-2)\right) \sigma_{s}^{2} ds\right) - \left\langle\sigma^{2}, h_{l k}\right\rangle_{L^{2}}\right| \leq \overline{C} \delta \sqrt{\frac{p \log M}{M}}.$$

Therefore, the lemma follows.

**Lemma B.12.** Suppose the assumptions of Proposition 4 hold. Then, for  $\delta > 0$ 

$$\mathbb{P}\left(\left|\sum_{i=2}^{M} h_{\ell k}\left(\frac{i-1}{M}\right) \overline{X}_{i,M}\left(\lambda\right) \overline{\epsilon}_{i,M}(\lambda)\right| > \sqrt{8\overline{C}} \|\tau\|_{\infty} \|\widetilde{\lambda}\|_{2} (1+\delta) \sqrt{\frac{p \log M}{M}} \text{ and } \sigma^{2} \in \mathcal{B}_{\pi,\infty}^{s}(C)\right) \lesssim M^{-p}.$$

*Proof.* Let  $\overline{X}_{i,M,T_{\overline{C}}}$  be defined as in (B.0.10). Then, by considering even and odd terms, separately we only need to prove

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

The same arguments apply for the sum over odd i. The proof is similar to the one developed for bounding II in Lemma B.11. We define

$$N_r = n^{1/2} 2^{-l/2} \sum_{i=1}^{2r} h_{\ell k} \left( \frac{2i-1}{M} \right) \overline{X}_{2i,M,T_{\overline{C}}} \overline{\epsilon}_{2i,M}.$$

Obviously, this is a martingale with respect to the filtration  $\mathcal{F}_r^{\text{even}}$ , starting at

$$N_{\lfloor (2^{-l}M(k-B)+1)/2 \rfloor} = 0.$$

Next, we compute the predictable quadratic variation,

$$\begin{split} \langle N \rangle_{\lfloor \frac{1}{2} (2^{-l}M(k+B)+1) \rfloor} &\leq n 2^{-l} \sum_{i=1}^{\lfloor M/2 \rfloor} h_{\ell k}^2 (\frac{2i-1}{M}) \mathbb{E} \left[ \overline{X}_{2i,M,T_{\overline{C}}}^2 \overline{\epsilon}_{2i,M}^2 | \mathcal{F}_{i-1}^{\text{even}} \right] \\ &\leq n 2^{-l} \, \| \tau \|_{\infty}^2 \sum_{i=1}^{\lfloor M/2 \rfloor} h_{\ell k}^2 (\frac{2i-1}{M}) \mathbb{E} \left[ \overline{X}_{2i,M,T_{\overline{C}}}^2 | \mathcal{F}_{i-1}^{\text{even}} \right] \frac{M^2}{n^2} \sum_{\substack{j \\ \overline{n} \in \left( \frac{2i-2}{M}, \frac{2i}{M} \right]}} \widetilde{\lambda}^2 (M \frac{j}{n} - (2i-2)). \end{split}$$

By the assumption on pre-average functions,  $\lambda$  is piecewise Lipschitz continuous. Therefore,

$$\frac{M}{n} \sum_{\substack{\underline{j} \in \left(\frac{2i-2}{M}, \frac{2i}{M}\right]}} \widetilde{\lambda}^2 (M_{\underline{n}}^{\underline{j}} - (2i-2)) = \|\widetilde{\lambda}\|_2^2 + O\left(\frac{M}{n}\right), \qquad (B.0.14)$$

uniformly in *i*. In the next step, we bound the conditional variance  $\mathbb{E}[\overline{X}_{2i,M,T_{\overline{C}}}^2|\mathcal{F}_{i-1}^{\text{even}}]$ . In particular, we have the decomposition  $\overline{X}_{2i,M,T_{\overline{C}}} = U_1 + U_2$ . Here,

$$\begin{split} U_1 &:= \frac{M}{n} \sum_{\substack{\underline{j} \in \left(\frac{2i-2}{M}, \frac{2i}{M}\right] \\ M \in \left(\frac{2i-2}{M}, \frac{2i}{M}\right]} \left(\sum_{l=j}^n \widetilde{\lambda} (M\frac{l}{n} - (2i-2))\right) \left(X_{\underline{j}_n \wedge T_{\overline{C}}} - X_{\underline{j-1}_n \wedge T_{\overline{C}} \wedge \frac{2i-2}{M}}\right), \\ U_2 &:= X_{\underline{2i-2}_M \wedge T_{\overline{C}}} \frac{M}{n} \sum_{\substack{\underline{j} \in \left(\frac{2i-2}{M}, \frac{2i}{M}\right]}} \widetilde{\lambda} (M\frac{j}{n} - (2i-2)). \end{split}$$

Moreover,  $\mathbb{E}[\overline{X}_{2i,M,T_{\overline{C}}}^2|\mathcal{F}_{i-1}^{\text{even}}] = \mathbb{E}[U_1^2|\mathcal{F}_{i-1}^{\text{even}}] + U_2^2$ . Now using conditional Itô-isometry again, we derive

$$\mathbb{E}\Big[\Big(X_{\underline{j}_{n}\wedge T_{\overline{C}}} - X_{\underline{j-1}_{n}\wedge T_{\overline{C}}\wedge \underline{2i-2}}_{M}\Big)\Big(X_{\underline{j'}_{n}\wedge T_{\overline{C}}} - X_{\underline{j'-1}_{n}\wedge T_{\overline{C}}\wedge \underline{2i-2}}_{M}\Big)|\mathcal{F}_{i-1}^{\mathrm{even}}\Big] \\
\leq \delta_{j,j'}\overline{C}_{n}^{1} = \overline{C} \ \mathbb{E}\Big[\Big(W_{\underline{j}_{n}} - W_{\underline{j-1}}_{n}\Big)\Big(W_{\underline{j'}_{n}} - W_{\underline{j'-1}}_{n}\Big)\Big], \quad \text{for } \underline{j}, \underline{j'}_{n} \in (\underline{2i-2}_{M}, \underline{2i}_{M}],$$

where W denotes a standard Brownian motion and  $\delta_{j,j'}$  is the Kronecker delta. Set X = W in (5.2.17). Then,

$$\mathbb{E}\left[U_1^2|\mathcal{F}_{i-1}\right] \leq \overline{C} \mathbb{E}\left[\left(\frac{M}{n}\sum_{\substack{j \ n \in \left(\frac{2i-2}{M}, \frac{i}{M}\right]}} \widetilde{\lambda}(M\frac{j}{n} - (2i-2))W_{j/n}\right)^2\right]$$
$$= \overline{C} \mathbb{E}\left[\int_0^1 \Lambda^2 \left(Ms - (2i-2)\right)ds\right] + O\left(M^{-1/2}n^{-1}\right)$$
$$= \overline{C}M^{-1} + O\left(M^{-1/2}n^{-1}\right)$$

uniformly over *i*, where we used Lemma B.4 for the last equality. With the Riemann approximation (B.0.11) we derive  $U_2^2 \leq \frac{M^2}{n^2} X_{\frac{2i-2}{M} \wedge T_{\overline{C}}}^2$ . For  $\delta_1, \delta_2 > 0$  the predictable quadratic variation can be bounded by

$$\begin{split} \langle N \rangle_{\lfloor \frac{1}{2}(2^{-l}M(k+B)+1) \rfloor} &\leq 2^{-l-1}M \, \|\tau\|_{\infty}^2 \, \overline{C} \|\widetilde{\lambda}\|_2^2 (1+\delta_1) \\ &+ 2^{-l} \frac{M^3}{n^2} \, \|\tau\|_{\infty}^2 \, \|\widetilde{\lambda}\|_2^2 (1+\delta_2) \, \sum_{i=1}^{\lfloor M/2 \rfloor} h_{\ell k}^2 (\frac{2i-1}{M}) X_{\frac{2i-2}{M} \wedge T_{\overline{C}}}^2 \end{split}$$

provided M is sufficiently large. Fix  $\delta_3 > 0$ . Then, by Lemma B.1 (ii) and Chebycheff inequality, we have that the probability of the event,

$$\frac{M^2}{n^2} \sum_{i=1}^{\lfloor M/2 \rfloor} h_{\ell k}^2(\frac{2i-1}{M}) X_{\frac{2i-2}{M} \wedge T_{\overline{C}}}^2 \lesssim \frac{M^3}{n^2} \sup_{s \le T_{\overline{C}}} X_s^2 \le \delta_3,$$

is larger than  $1 - KM^{-p}$ , for a constant K. Now, for  $\delta > 0$  there are  $\delta_1, \delta_2, \delta_3 > 0$  such that by combining the inequalities above, we derive

$$\mathbb{P}\left(\left\langle N\right\rangle_{\lfloor\frac{1}{2}(2^{-l}M(k+B)+1)\rfloor} > 2^{-l-1}M \left\|\tau\right\|_{\infty}^{2}\overline{C}\|\widetilde{\lambda}\|_{2}^{2}(1+\delta)\right) \lesssim M^{-p}.$$

Before we can apply Lemma 3, we still need a suitable moment bound on the increments of  $(N_i)_i$ . As demonstrated in the proof of Lemma B.7, it holds  $\mathbb{E}[|\overline{X}_{i,M,T_C}(\lambda)|^{2\kappa}] \leq M^{-\kappa}$  as well as  $\mathbb{E}[|\overline{\epsilon}_{i,M}|^{2\kappa}] \leq M^{\kappa}n^{-\kappa}$ . In the same way, we derive  $\mathbb{E}[|\overline{X}_{i,M,T_{\overline{C}}}(\lambda)|^{2\kappa}] \leq M^{-\kappa}$ . This shows that

$$\max_{i} \mathbb{E}[|\Delta_{i}N|^{\kappa}] \lesssim \max_{i} 2^{-l\kappa/2} n^{\kappa/2} \left| h_{\ell k}(\frac{i-1}{M}) \right|^{\kappa} \mathbb{E}^{1/2}[|\overline{X}_{i,M,T_{\overline{C}}}(\lambda)|^{2\kappa}] \mathbb{E}^{1/2}[|\overline{\epsilon}_{i,M}|^{2\kappa}] \lesssim 1$$

This allows us to apply Lemma 3. Then, the conclusion follows in a straightforward way.  $\hfill \Box$ 

Finally, in order to show Proposition 4, we need to bound the pure noise term.

**Lemma B.13.** Suppose the assumptions of Proposition 4 hold. Then, for  $\delta > 0$ 

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

*Proof.* As in the previous lemmas, we decompose into even and odd terms and show that

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

By the same arguments we may bound the sum over odd *i*. We define  $\mathcal{F}_r^{\text{even}}$ ,  $U_i$  as well as the martingale  $S_r^{\text{even}}$  as in the proof of Lemma B.6. Let us mention again that since *h* has compact support, we can find an interval [-B, B] such that h = 0 on  $\mathbb{R} \setminus [-B, B]$ . Thus, for  $h_{\ell k}(\frac{2i-1}{M}) \neq 0$  it is necessary that  $\frac{1}{2}(2^{-l}M(k-B)+1) \leq i \leq \frac{1}{2}(2^{-l}M(k+B)+1)$ . Define the process  $N_r := \frac{n}{M}2^{-l/2}S_r^{\text{even}}$ . This is a martingale with respect to  $\mathcal{F}_r^{\text{even}}$ , starting at  $N_{\lfloor (2^{-l}M(k-B)+1)/2 \rfloor} = 0$ . Bounding the predictable quadratic variation gives

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

Furthermore, straightforward calculations and the Riemann approximation argument given in (B.0.14) show that uniformly in i, we can bound

$$\mathbb{E}\Big[\left(\overline{\epsilon}_{i,M}^2 - \mathbb{E}[\overline{\epsilon}_{i,M}^2|\mathcal{G}]\right)^2 |\mathcal{F}_{i-1}^{\text{even}}\Big] = \widetilde{\lambda}^2 (M\frac{j}{n} - (i-2)) \Big)^2 + O\Big(\frac{M^3}{n^3}\Big)$$
$$= 2\frac{M^2}{n^2} \|\tau\|_{\infty}^4 \|\widetilde{\lambda}\|_2^4 + O\Big(\frac{M^3}{n^3}\Big)$$

by a deterministic constant. Hence, for sufficiently large M,

$$\langle N \rangle_{\lfloor \frac{1}{2}(2^{-l}M(k+B)+1) \rfloor} \le M 2^{-l} \|\tau\|_{\infty}^{4} \|\widetilde{\lambda}\|_{2}^{4}(1+\delta).$$

Now, we may apply Lemma 3, since by (B.0.9) we have  $\mathbb{E}[|\Delta_i N|^{\kappa}] \lesssim 1$ . This yields the claim.

#### Completion of proof of Proposition 4

Let I, II, and III be defined as in (5.2.12).

Bounding I. Lemma B.11, shows that under the assumptions of Proposition 4,

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

Bounding II. Applying Lemmas B.8, B.9, B.10, and B.13 we derive by the generalized version of Chebycheff's inequality (B.0.13) and because of  $\|h_{\ell k}\|_{p,M}^p \lesssim M^{p/2-1}$ , for  $p \ge 2$ 

$$\mathbb{P}\left(|II| > 4 \|\tau\|_{\infty}^{2} \|\widetilde{\lambda}\|_{2}^{2}(1+\delta)\sqrt{\frac{p\log M}{M}} \text{ and } \sigma^{2} \in \mathcal{B}_{\pi,\infty}^{s}(C)\right) \lesssim M^{-p}$$

Bounding III. By Lemma B.12,

$$\mathbb{P}\left(|III| > 4\sqrt{2\ \overline{C}} \ \|\tau\|_{\infty} \|\widetilde{\lambda}\|_{2}(1+\delta)\sqrt{\frac{p\log M}{M}} \text{ and } \sigma^{2} \in \mathcal{B}^{s}_{\pi,\infty}(C)\right) \lesssim M^{-p}$$

Combining the bounds of I - III completes the proof of Proposition 4.

#### **B.1.** Tools from stochastic analysis

For a thorough treatment of stochastic analysis, we refer to Protter [69] as well as Jacod and Shiryaev [45]. Throughout this section suppose that a filtered probability space  $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t\geq 0}, \mathbb{P})$  is given.

**Definition 21** (Quadratic variation). For semimartingales X, Y the quadratic covariation is the process  $X_tY_t - \int X_{s-}dY_s - \int Y_{s-}dX_s$ . It is denoted by  $([X,Y]_t)_{t\geq 0}$ . For X = Ythe process  $([X,X]_t)_{t\geq 0}$  is called quadratic variation.

Since this will become a crucial property, we recall the approximation theorem for quadratic variation of semimartingales in a very general form (cf. [69], Theorem 23). In order to do so, a notion of regularity/denseness of sequences of stopping times is needed.

**Theorem B.1.** Let X, Y be semimartingales. Let  $(T_k^n)_{k,n}$  be an array of increasing stopping times in k,

$$0 =: T_0^n \le T_1^n \le \ldots \le T_{k_n}^n < \infty$$

satisfying

(i)  $\lim_{n} \sup_{k} T_{k}^{n} = \infty, \ a.s.$ 

(*ii*) 
$$\lim_{n} \sup_{k} (T_{k+1}^{n} - T_{k}^{n}) = 0, \ a.s$$

Let

$$\widetilde{[X,Y]}_{(T_k^n)} := X_0 Y_0 + \sum_{i=0}^{k_n - 1} (X^{T_{i+1}^n} - X^{T_i^n}) (Y^{T_{i+1}^n} - Y^{T_i^n}),$$

then

$$\widetilde{[X,Y]}_{(T_k^n)} \stackrel{n \to \infty}{\to} [X,Y],$$

where mode of convergence is u.c.p. (i.e. uniform on compact intervals).

**Lemma B.14** (Integration by parts). Assume that X is a continuous semimartingale. Then,

$$[X, X]_t = X_t^2 - 2\int_0^t X_s dX_s.$$

**Lemma B.15** (Conditional Itô-isometry (cf. Karatzas and Shreve [51], Proposition 2.10)). Let M be a continuous, square integrable  $\mathcal{F}$ -martingale and assume that H is a progressively measurable process. Then for  $0 \leq s < t \leq 1$ ,

$$\mathbb{E}\Big[\Big(\int_{s}^{t}H_{u}dM_{u}\Big)^{2}|\mathcal{F}_{s}\Big]=\mathbb{E}\Big[\int_{s}^{t}H_{u}^{2}d\langle M\rangle_{u}|\mathcal{F}_{s}\Big],$$

provided

$$\mathbb{E}\Big[\int_0^1 H_u^2 \langle M \rangle_u\Big] < \infty.$$

**Lemma B.16** (cf. Protter [69], Theorem 12). Let X denote a semimartingale and H a process with càglàd paths. Let  $\tau$  be a stopping time. Then,

$$\int_0^{t\wedge\tau} H_s dX_s = \int_0^t H_s \mathbb{I}_{[0,\tau]}(s) dX_s.$$

#### Martingale moment inequalities

For a process X we write

$$X_t^\star := \sup_{u \in [0,t]} |X_u|.$$

**Lemma B.17** (Burkholder-Davis-Gundy inequality (BDG), (cf. Kallenberg [50], Theorem 26.12)). For any local martingale M with  $M_0 = 0$  and any  $p \ge 1$ , there exists a positive constant c only depending on p, such that for all t > 0,

$$c^{-1} \mathbb{E}\left[[M, M]_t^{p/2}\right] \le \mathbb{E}\left[|M_t^{\star}|^p\right] \le c \mathbb{E}\left[[M, M]_t^{p/2}\right].$$

The next inequality provides a useful tool in order to bound the maximum of a discrete martingale by means of controlling the increments.

**Lemma B.18** (Rosenthal's inequality, (cf. Hall and Heyde [35], p. 23)). Let  $(M_k)_k$  be a martingale with respect to the filtration  $\{\mathcal{F}_k\}_k$ . For  $p \ge 0$  there is a constant C, only dependent on p such that

$$\mathbb{E}\Big[\max_{k=1,\dots,n}|M_k|^p\Big] \le C\Big(\mathbb{E}\Big[\Big(\sum_{k=0}^{n-1}\mathbb{E}\big[(\Delta_k M)^2|\mathcal{F}_k\big]\Big)^{p/2}\Big] + \mathbb{E}\Big[\max_{k\le n}|\Delta_k M|^p\Big]\Big).$$

# List of Symbols

$\operatorname{Cov}(X,Y)$	For column vectors $X, Y$ of length $m_X$ and $m_Y$ , the covariance of X and Y is defined as the matrix $\text{Cov}(X, Y) \in \mathbb{M}_{m_X, m_Y}$ with $(\text{Cov}(X, Y))_{i,j} := \text{Cov}(X_i, Y_j)$ , page 44
$\Delta_i M$	$\begin{array}{l} \Delta_{i,n}a := a(\frac{i+1}{n}) - a(\frac{i}{n}), \ \Delta_{i,n}Y := Y_{i+1,n} - Y_{i,n}, \ \Delta_{i,n}X := X_{(i+1)/n} - \\ X_{i/n} \text{ is the forward difference operator, } \Delta_{i,n}^k := \Delta_{i,n}^{k-1} \circ \Delta_{i,n}, \ \Delta_{i,n}^1 := \\ \Delta_{i,n} \text{ is defined recursively, page 20} \end{array}$
<~	larger up to a constant. In particular, $a \lesssim b$ uniformly in a parameter $\alpha$ means that the constant is independent of $\alpha$ , page 18
$\lfloor x \rfloor$	denotes the floor function, i.e. it is the largest integer not greater than $x$ , page 23
$\ \cdot\ _{L^p[a,b]}$	denotes the $L^p$ -norm on the interval $[a, b]$ . In the baseline case $[a, b] = [0, 1]$ , we abbreviate the $L^p$ -norm by $\ .\ _p$ , page 22
$\ \cdot\ _{l^q}$	denotes the sequence space norm on the index set $I$ , i.e. $  (a_i)_{i \in I}  _{l^q} := (\sum_{i \in I}  a_i ^q)^{1/q}$ with obvious modification for $q = \infty$ , page 25
$\ .\ _{p,m}$	is the empirical $L^p$ -norm on $[0, 1]$ , i.e. $  f  _{p,m} := (\frac{1}{m} \sum_{i=1}^m  f(\frac{i}{m}) ^p)^{1/p}$ , page 22
$\ .\ _2$	If A is a matrix, then $  A  _2$ denotes the Frobenius (or Hilbert-Schmidt) norm, page 30
$\mathbb{M}_{p,q}, \mathbb{M}_p \text{ and } \mathbb{D}_p$	are the spaces of $p \times q$ matrices, $p \times p$ matrices and $p \times p$ diagonal matrices over $\mathbb{R}$ , respectively, page 38
$\underline{\mathcal{D}}$	equality in distribution, page 36
$D \in \mathbb{M}_{n-1}$	$(D_{n-1})_{i,j=1,\dots,n-1} = (\sqrt{2/n}\sin{(ij\pi/n)})_{i,j=1,\dots,n-1}$ , page 38
$O_p()$	For a sequence of random variables $(U_n)_n$ , we write $U_n = O_p(c_n)$ , whenever $c_n^{-1}U_n$ is bounded in probability, page 32

## Bibliography

- F. Abramovich and B. W. Silverman. Wavelet decomposition approaches to statistical inverse problems. *Biometrika*, 85:115–129, 1998.
- [2] Y. Ait-Sahalia and J. Yu. High frequency market microstructure noise estimates and liquidity measures. *The Review of Financial Studies*, 18:351–416.
- [3] A. Alvarez, F. Panloup, M. Pontier, and N. Savy. Estimation of the instantaneous volatility. 2010. arXiv:0812.3538v3, Math arXiv Preprint.
- [4] T. G. Andersen and T. Bollerslev. Deutsche Mark-Dollar volatility: Intraday activity patterns, macroeconomic announcements, and longer run dependencies. *The Journal of Finance*, 53(1):219–265, 1998.
- [5] F. Azuma and J. Russell. Weighted sums of certain dependent random variables. *Tohoku Math. Journ.*, 19(3):357–367, 1967.
- [6] O. Barndorff-Nielsen, J. M. Corcuera, and M. Podolskij. Multipower variation for Brownian semi-stationary processes. 2009. Working paper.
- [7] 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(6):1481–1536, 2008.
- [8] O. Barndorff-Nielsen and J. Schmiegel. Time change, volatility, and turbulence. 2007. Working paper.
- [9] O. Barndorff-Nielsen and J. Schmiegel. Brownian semistationary processes and volatility/intermittency. *Radon Series on Comp. and Appl. Math*, 8:1–25, 2009.
- [10] A. Basse. Spectral representation of Gaussian semimartingales. J. Theor. Probab., 22:811–826, 2009.
- [11] F. Baudoin and D. Nualart. Equivalence of Volterra processes. Stochastic Process. Appl., 107:327–350, 2003.
- [12] B. Bercu and A. Touati. Exponential inequalities for self-normalized martingales with application. Ann. Appl. Prob., 18(5):1848–1869, 2008.

- [13] M. Bibinger. Efficient covariance estimation for asynchronous noisy high-frequency data. 2008. Unpublished manuscript.
- [14] F. Black. Studies in stock price volatility changes. Proceedings of the 1976 Business Meeting of the Business and Economic Statistics Section, American Statistical Association, pages 177–181, 1976.
- [15] T. Bollerslev, J. R. Russell, and M. W. Watson. Volatility and Time Series Econometrics. Oxford University Press, 2010.
- [16] T. Cai, A. Munk, and J. Schmidt-Hieber. Sharp minimax estimation of the variance of Brownian motion corrupted with Gaussian noise. *Statist. Sinica*, 20:1011–1024, 2010.
- [17] T. Cai and H. Zhou. A data-driven block thresholding approach to wavelet estimation. Ann. Statist., 37(2):569–595, 2009.
- [18] A. Cohen. Numerical Analysis of Wavelet Methods. Elsevier, 2003.
- [19] R. Dahlhaus and J. C. Neddermeyer. Particle filter-based on-line estimation of spot volatility with nonlinear market microstructure noise models. 2010. arxiv:1006.1860, Math arXiv Preprint.
- [20] I. Daubechies. Ten Lectures on Wavelets. SIAM, Philadelphia, 1992.
- [21] F. Delbaen and W. Schachermayer. A general version of the fundamental theorem of asset pricing. *Math. Ann.*, 300:463–520, 1994.
- [22] F. Delbaen and W. Schachermayer. The fundamental theorem of asset pricing for unbounded stochastic processes. *Math. Ann.*, 312:215–250, 1998.
- [23] H. Demsetz. The cost of transacting. The Quarterly Journal of Finance, 82(1):33– 53, 1968.
- [24] F. X. Diebold and G. H. Strasser. On the correlation structure of microstructure noise in theory and practice. Working Paper, 2008.
- [25] D. Donoho and I. M. Johnstone. Ideal spatial adaptation via wavelet shrinkage. Biometrika, 81:425–455, 1994.
- [26] D. Donoho, I. M. Johnstone, G. Kerkyacharian, and D. Picard. Wavelet shrinkage: Asymptopia? J. R. Statist. Soc. B, 57(2):301–369, 1995.
- [27] E. F. Fama. Mandelbrot and the stable paretian distribution. The Journal of Business, 36:420–429, 1963.

- [28] E. F. Fama. The behavior of stock market prices. The Journal of Business, 38:34– 105, 1965.
- [29] A. Gloter. Discrete sampling of an integrated diffusion process and parameter estimation of the diffusion coefficient. *ESAIM Probab. Stat.*, 4:205–227, 2000.
- [30] A. Gloter and M. Hoffmann. Stochastic volatility and fractional Brownian motion. Stochastic Process. Appl., 113:143–172, 2004.
- [31] A. Gloter and M. Hoffmann. Estimation of the Hurst parameter from discrete noisy data. Ann. Statist., 35:1947–1974, 2007.
- [32] A. Gloter and M. Hoffmann. Nonparametric reconstruction of a multifractal function from noisy data. Probab. Theory Relat. Fields, 146:155–187, 2010.
- [33] A. Gloter and J. Jacod. Diffusions with measurement errors. I. Local asymptotic normality. ESAIM Probab. Stat., 5:225–242, 2001.
- [34] A. Gloter and J. Jacod. Diffusions with measurement errors. II. Optimal estimators. ESAIM Probab. Stat., 5:243–260, 2001.
- [35] P. Hall and C. C. Heyde. Martingale Limit Theory and Its Applications. Academic Press, New York, 1980.
- [36] J. Hasbrouck. Assessing the quality of a security market: A new approach to transaction-cost measurement. The Review of Financial Studies, 6(1):191–212, 1993.
- [37] J. Hasbrouck. Empirical Market Microstructure: The Institutions, Economics, and Econometrics of Securities Trading. Oxford University Press, 2007.
- [38] T. Hayashi and N. Yoshida. On covariance estimation of non-synchronously observed diffusion processes. *Bernoulli*, 11(2):359–379, 2005.
- [39] T. Hida and M. Hitsuda. Gaussian Processes. Translation of Mathematical Monographs, Providence, 1993.
- [40] T. Hida and Si Si. Lectures on White Noise Functionals. World Scientific, 2008.
- [41] M. Hoffmann. Adaptive estimation in diffusion processes. Stochastic Process. Appl., 79:135–163, 1999.
- [42] M. Hoffmann.  $L_p$  estimation of the diffusion coefficient. *Bernoulli*, 5(3):447–481, 1999.

- [43] M. Hoffmann, A. Munk, and J. Schmidt-Hieber. Nonparametric estimation of the volatility under microstructure noise: Wavelet adaptation. arxiv:1007.4622, Math arXiv Preprint.
- [44] J. Jacod, Y. Li, P. A. Mykland, M. Podolskij, and M. Vetter. Microstructure noise in the continuous case: The pre-averaging approach. *Stochastic Process. Appl.*, 119(7):2249–2276, 2009.
- [45] J. Jacod and A. N. Shiryaev. Limit Theorems for Stochastic Processes. Springer, Berlin, 1987.
- [46] I. M. Johnstone. Wavelet shrinkage for correlated data and inverse problems: adaptivity results. *Statist. Sinica*, 9:51–83, 1999.
- [47] I. M. Johnstone and B. Silverman. Wavelet threshold estimators for data with correlated noise. J. R. Statist. Soc. B, 59(2):319–351, 1997.
- [48] E. Jondeau, S.-H. Poon, and M. Rockinger. Financial Modeling Under Non-Gaussian Distributions. Springer, New York, 2006.
- [49] C. Jost. A note on ergodic transformations of self-similar Volterra Gaussian processes. *Electron. Commun. Probab.*, 12:259–266, 2007.
- [50] O. Kallenberg. Foundations of Modern Probability. Springer, 1997.
- [51] I. Karatzas and S. E. Shreve. Brownian Motion and Stochastic Calculus. Springer, 2004.
- [52] G. Kerkyacharian and D. Picard. Thresholding algorithms, maxisets and wellconcentrated bases. *Test*, 9:283–345, 2000.
- [53] L. Le Cam and L. O. Yang. Asymptotics in Statistics: Some Basic Concepts. Springer, New York, 2000.
- [54] B. LeBaron. Some relations between volatility and serial correlations in stock market returns. The Journal of Business, 65(2):199–219, 1992.
- [55] P. Lévy. Fonctions aléatoires à corréllation linéaire. Illinois J. Math., 1(2):217–258, 1957.
- [56] A. Madahavan. Market microstructure: A survey. Journal of Financial Markets, 3:205–258, 2000.
- [57] S. Mallat. A theory of multiresolution signal decomposition: the wavelet representation. *IEEE transactions pattern recognition and machine intelligence*, 11:674–693, 1989.

- [58] P. Malliavin and M. E. Mancino. A Fourier transform method for nonparametric estimation of multivariate volatility. *Ann. Statist.*, 37(4):1983–2010, 2009.
- [59] M. E. Mancino and S. Sanfelici. Robustness of Fourier estimator of integrated volatility in the presence of microstructure noise. *Comput. Statist. Data Anal.*, 52(6):2966–2989, 2008.
- [60] B. B. Mandelbrot. The variation of certain speculative prices. The Journal of Business, 36:394–416, 1963.
- [61] I. Meyer. Ondelettes, fonctions splines et analyses graduées. 1986. Lectures given at the University of Torino, Italy.
- [62] I. Meyer. Wavelets and Operators. Cambridge University Press, 1992.
- [63] A. Munk and J. Schmidt-Hieber. Lower bounds for volatility estimation in microstructure noise models. Borrowing Strength: Theory Powering Applications - A Festschrift for Lawrence D. Brown, IMS Collections, 6:43–55, 2010.
- [64] A. Munk and J. Schmidt-Hieber. Nonparametric estimation of the volatility function in a high-frequency model corrupted by noise. *Electron. J. Stat.*, 4:781–821, 2010.
- [65] H.-L. Ngo and S. Ogawa. A central limit theorem for the functional estimation of the spot volatility. *Monte Carlo Methods Appl.*, 15(4):353–380, 2009.
- [66] M. Nussbaum. Asymptotic equivalence of density estimation and Gaussian white noise. Ann. Statist., 24(6):2399–2430, 1996.
- [67] V. H. Pena. A general class of exponential inequalities for martingales and ratios. Ann. Probab., 27:537–564, 1999.
- [68] M. Podolskij and M. Vetter. Estimation of volatility functionals in the simultaneous presence of microstructure noise and jumps. *Bernoulli*, 15:634–658, 2009.
- [69] P. E. Protter. Stochastic Integration and Differential Equations. Springer, New York, 2005.
- [70] L. Ramchand and R. Susmel. Volatility and cross correlation across major stock markets. *Journal of Empirical Finance*, 5(4):397–416, 1998.
- [71] M. Reiß. Asymptotic equivalence and sufficiency for volatility estimation under microstructure noise. arxiv:1001.3006, Math arXiv Preprint.
- [72] R. Roll. A simple model of the implicit bid-ask spread in an efficient market. Journal of Finance, 39:1127–1139, 1984.

- [73] Si Si. A note on Lévy's Brownian motion I. Nagoya Math. Journal, 108:121–130, 1987.
- [74] A. B. Tsybakov. Introduction to Nonparametric Estimation. Springer-Verlag, New York, 2009.
- [75] S. van de Geer. Exponential inequalities for martingales, with application to maximum likelihood estimation for counting processes. Ann. Statist., 23(5):1779–1801, 1995.
- [76] L. Zhang. Efficient estimation of stochastic volatility using noisy observations: A multi-scale approach. *Bernoulli*, 12:1019–1043, 2006.
- [77] L. Zhang, P. Mykland, and Y. Ait-Sahalia. A tale of two time scales: Determining integrated volatility with noisy high-frequency data. J. Amer. Statist. Assoc., 472:1394–1411, 2005.
## Curriculum vitae

November 22, 1984	born in Freiburg im Breisgau, Germany.
2003-2004	Studies in Mathematics, University of Freiburg.
2004	Abitur.
2004-2007	Studies in Mathematics, University of Göttingen.
2005-2007	Scholarship of German National Academic Foundation (Studienstiftung des dt. Volkes).
2006-2007	Visiting Scholar at University of California, Davis.
2007-2008	Supported by SFB 755 "Nanoscale Photonic Imaging".
2008-	Associate research fellow of RTG "Identification in Mathematical Models".
2008-	Supported by FOR 916 "Statistical Regularisation and Qualitative Constraints".