Wavelet Estimation in Diffusions with Periodicity

Dissertation zur Erlangung des Grades
„Doktor der Naturwissenschaften“

am Fachbereich Physik, Mathematik und Informatik
der Johannes Gutenberg-Universität in Mainz

Michael Diether

geboren in Oberwesel

Mainz, den 7. 12. 2011

Datum der mündlichen Prüfung: 25. April 2012

D77 – Mainzer Dissertation

Contents

Sentio me mente helluari posse

ut anima interea intereat fame

Abstract

We consider a time-inhomogeneous diffusion process, given by a stochastic differential equation, whose drift term contains a deterministic T-periodic signal with known periodicity. This signal is supposed to be contained in some Besov space. We estimate it using a non-parametric wavelet estimator. Our estimator is inspired by the thresholded wavelet density estimator for the classical iid-setting constructed by Donoho, Johnstone, Kerkyacharian and Picard in 1996. Under certain ergodicity assumptions on the process, we give non-parametric rates of convergence which correspond, up to a logarithmical term, to the rate of convergence in the classical iid-setting. These rates are proved using oracle inequalities that rely on results for discrete-time Markov chains by Clémençon from 2001. Besides, a technically simpler special case is considered and some computer simulations of this estimator are shown.

Zusammenfassung

Wir betrachten einen zeitlich inhomogenen Diffusionsprozess, der durch eine stochastische Differentialgleichung gegeben wird, deren Driftterm ein deterministisches T-periodisches Signal beinhaltet, dessen Periodizität bekannt ist. Dieses Signal sei in einem Besovraum enthalten. Wir schätzen es mit Hilfe eines nichtparametrischen Waveletschätzers. Unser Schätzer ist von einem Wavelet-Dichteschätzer mit Thresholding inspiriert, der 1996 in einem klassischen iid-Modell von Donoho, Johnstone, Kerkyacharian und Picard konstruiert wurde. Unter gewissen Ergodizitätsvoraussetzungen an den Prozess können wir nichtparametrische Konvergenzraten angegeben, die bis auf einen logarithmischen Term den Raten im klassischen iid-Fall entsprechen. Diese Raten werden mit Hilfe von Orakel-Ungleichungen gezeigt, die auf Ergebnissen über Markovketten in diskreter Zeit von Clémençon, 2001, beruhen. Außerdem betrachten wir einen technisch einfacheren Spezialfall und zeigen einige Computersimulationen dieses Schätzers.

List of symbols

$\|f\|_r$	$\left(\int_0^1 \|f(x)\|^r \,\mathrm{d}x\right)^{\frac{1}{r}}$
$\|f\|_{L^r}$	$\left(\int_{\mathbb{R}} \|f(x)\|^r \,\mathrm{d}x\right)^{\frac{1}{r}}$; referred to by "$L^r$-norm" in the text
$\|f\|_\infty$	$\sup_{x\in\mathbb{R}} \|f(x)\|$
$\|f\|_{spq}$	norm in the Besov space B_{spq}, see (2.4)
$\|a_\bullet\|_{\ell^r}$	$\left(\sum_{n=0}^{\infty} \|a_n\|^r\right)^{\frac{1}{r}}$
$\|a_\bullet\|_{\ell^\infty}$	$\sup_{n\in\mathbb{N}} \|a_n\|$
$\langle M\rangle$	quadratic variation of the (local) martingale M
$\mathbb{1}_M(\cdot)$	indicator function of the set M
$(a)^+$	positive part of $a \in \mathbb{R}$; $(a)^+ := \max(a, 0)$
$a_n \asymp b_n$	"a_n grows asymptotically like b_n", $\exists c, C > 0 \,:\, c \leqslant \frac{a_n}{b_n} \leqslant C \ \forall n \in \mathbb{N}$
$a_1 \wedge a_2$	$\min\{a_1, a_2\}$
b	known drift function in the process ξ
$\hat{B}_j, B_j, B_j'$	see page 47
B_{spq}	Besov space, set of all functions with $\|f\|_{spq} < \infty$
$\tilde{B}_{spq}(M)$	set of all functions with $\|f\|_{spq} < M$ and $\|f\|_\infty < M$
C	positive real constant; may be different every time it appears
$\mathcal{C}^l(A)$	set of l-times continuously differentiable functions defined on A
$\mathcal{C}_0(A)$	set of continuous functions starting in 0, defined on A
$D_j(f)$	projection operator on the space W_j, see (2.7)
$E_i(f)$	projection operator on the space V_i, see (2.6)
$\mathcal{E}$	canonical σ-field of $\mathcal{C}([0, T])$
$E_m g$	math. expectation of the function g with respect to the measure m
$h(x)$	see (3.21)
$j_0 = j_0(n)$	first level of the wavelet expansion in the estimator TW_n, see (2.14)
$j_1 = j_1(n)$	last level of the wavelet expansion in the estimator TW_n, see (2.14)

$J_{spq}(f)$	norm in the Besov space B_{spq}, see (2.5)
K	a fixed positive constant; see Definition 2.12*c*)
$\log_a x$	logarithm of $x \in \mathbb{R}$, $x > 0$, with respect to base a; also: $\log_e x = \log x$
m	invariant probability measure of the Markov chain $(\mathbb{X}_i)_{i\geqslant 1}$
n	number of observations (i. e. the number of observed path segments $\mathbb{X}_i$)
$\mathcal{N}(\mu, \sigma^2)$	normal distribution with mean $\mu \in \mathbb{R}$ and variance $\sigma^2 > 0$
S	unknown drift function in the process ξ
$\hat{S}_j, S_j, S'_j$	see page 47
s'	see Definition 2.11
$\mathrm{span}\{v_i \colon i \in I\}$	vector space spanned by v_i, $i \in I$
$\mathrm{supp} f$	support of the function f; $\mathrm{supp} f := \overline{\{x \in \mathbb{R} \colon f(x) \neq 0\}}$
T	period of the function S, known
U	the event $\bigcup_{i=1}^{n} \{U_{n,i} < T\}$
$U_{n,i}$	stopping time in the definition of Φ_n, see (3.2)
$(V_i)_{i\in\mathbb{Z}}$, $(W_j)_{j\in\mathbb{Z}}$	subspaces of $L^2(\mathbb{R})$, see Definition 2.1 and Theorem 2.2
$(W_t)_{t\geqslant 0}$	standard Brownian motion
$\mathbb{W}_i$	i-th path segment of W_t, $\mathbb{W}_i(0) = 0$; $\mathbb{W}_i(s) := W_{(i-1)T+s} - W_{(i-1)T}$
$\mathbb{X}_i$	i-th path segment of the trajectory of ξ; $\mathbb{X}_i(s) := \xi_{(i-1)T+s}$
α	see Definition 2.11
α_{jk}, β_{jk}	coefficients in a wavelet expansion
$\hat{\alpha}_{jk,n}$, $\hat{\beta}_{jk,n}$	empirical coefficients in a wavelet expansion
$\breve{\alpha}_{jk,n}$, $\breve{\beta}_{jk,n}$	oracle coefficients in a wavelet expansion
$\tilde{\beta}_{jk}$	thresholded empirical coefficients in a wavelet expansion
$\Gamma(x)$	Euler's gamma function, $\Gamma(x) := \int_0^\infty e^{-t} t^{x-1} \mathrm{d}t$
Δ_{jk}	see (3.28)
ε	see Definition 2.11
ζ	parameter in Theorem 3.9; see also (3.12)-(3.15)
η	parameter in the definition of $U_{n,i}$, see Definition 3.2 and (3.12)-(3.15)
ν	thresholding parameter, see (2.11); see also (3.12)-(3.15)
Π	transition probability of the Markov chain $(\mathbb{X}_i)_{i\geqslant 1}$
ϱ	see (3.21)
σ	volatility function in the process ξ
$\Sigma_n(\overline{\Phi}_n)$	see Lemma 3.8
τ_A	first return time of $(\mathbb{X}_i)_{i\geqslant 1}$ to the set A; $\tau_A := \min\{i \geqslant 1 \colon \mathbb{X}_i \in A\}$
Φ_n	a functional of the observed trajectory of ξ, see (3.1)
$\overline{\Phi}_n$	$\Phi_n - E_m \Phi_n$
ξ	stochastic process given by the stochastic differential equation (2.1)

Chapter 1

Introduction

We consider a time-inhomogeneous continuous diffusion process, defined by a stochastic differential equation, whose drift term contains an unknown deterministic periodic function of fixed and known periodicity. Diffusion processes like this arise in the analysis of neuronal data where the neurons receive some repeated external influence for a known period of time. In this setting, it is unknown how the external influence is processed within the neurons, and so the observed trajectory of the neuronal potential should mirror some periodic influence as well as some stochastic fluctuations. Other applications of such diffusion processes can be thought of.

We assume that the deterministic signal is contained in a Besov space, a certain space of real-valued functions, that has been shown to be useful for statistical inference with wavelets (see e. g. [19]). Wavelets are, loosely speaking, square-integrable functions which can be scaled and translated in a suitable fashion, in order to yield an orthonormal basis of the space of all square-integrable functions on the real line. The theory of wavelets has become increasingly popular in numerical analysis and statistics since the 1980s, when Ingrid Daubechies gave the first construction of a family of compactly supported wavelets. An early example of the use of wavelet methods in non-parametric statistics is the 1992 article [11] where the volatility function in a stochastic differential equation is estimated by giving estimates for its wavelet coefficients from discrete-time observations. The authors also consider weak convergence and the mean integrated square error of their estimator.

We will construct a wavelet estimator for the unknown signal term of our model and give an asymptotic upper bound for its error of estimation (in an L^r-norm). In order to do so, we assume that we can observe n periods of the diffusion process, continuously in time. The estimation procedure is based on the estimation of finitely many, appropriately chosen coefficients of the wavelet expansion of the unknown signal. All our asymptotics deal with the case of "many observations" ($n \to \infty$).

The construction of our estimator and the idea of the proof of the asymptotic upper bound rely on an article by Donoho, Johnstone, Kerkyacharian and Picard, [6]. They considered a similar wavelet estimator for a probability density, when one has i.i.d. observations, and proved an asymptotic upper bound for the error of estimation. Besides, they showed that their bound is optimal up to logarithmic terms. The basic idea of the proof for the upper bound is to consider different parts of the wavelet expansion of the unknown signal separately. We need to consider certain moment and large deviation inequalities to arrive at the desired upper bounds. Inequalities like these are classical results in the i.i.d.-case, but we will find that these results cannot be applied in our setting, since we have dependence in our observations. To deal with this technical problem, our basic idea is to cut the observed trajectory of the diffusion process into n parts, one part for each period of the unknown signal. These parts form a Markov chain with a polish state space, the space of continuous functions on a compact interval. Our estimator can be considered as a functional of this Markov chain.

Some of our technical auxiliary results are based upon an article by Clémençon, [3]. They deal with moment bounds and large deviation probabilities for a functional of a discrete-time Markov chain, that we can conveniently apply to our situation. These results apply, however, to *bounded* functionals of this Markov chain only. Since our estimator will not be bounded in general, we introduce a bounded oracle which cannot be observed, but which differs from our estimator with low probability only (exponentially low in the number of observations). The theorems from [3] can be applied to the oracle and we can hence infer results that lead to bounds on our estimator.

In Chapter 2 we will specify our assumptions and our main result more precisely. Our main result is Theorem 2.14. The proof and the technical details can be found in Chapter 3. In Chapter 4 we reprove our main result in a technically much simpler special case; here, the difficulties mentioned above do not arise, and our result follows from well-known properties of the normal distribution. In Chapter 5, we present some computer simulations of our estimator and show its performance compared to the upper bound from our main theorem. Besides, we compare our non-parametric wavelet estimator to a parametric minimum distance estimator that was introduced in [15].

Chapter 2

Assumptions and main result

Consider the stochastic differential equation

$$\mathrm{d}\xi_t = S(t)\mathrm{d}t - b(\xi_t)\mathrm{d}t + \sigma(\xi_t)\mathrm{d}W_t, \qquad \xi_0 = v \in \mathbb{R} \text{ a.s.}, \tag{2.1}$$

which can also be written as

$$\xi_t = \xi_0 + \int_0^t S(u)\mathrm{d}u - \int_0^t b(\xi_u)\mathrm{d}u + \int_0^t \sigma(\xi_u)\mathrm{d}W_u, \qquad t \geqslant 0.$$

The function S is unknown and T-periodic with a known periodicity and obeys some model assumptions that we give in assumption (A) below. We shall show how to estimate S using methods that were inspired by an article by Donoho, Johnstone, Kerkyacharian and Picard, [6], who give a wavelet-estimator for probability densities from i.i.d. observations.

2.1 A short summary on Wavelets and Besov spaces

Before we can clarify the model that we are going to work in, we recall some notions of the theory of wavelets and Besov spaces. We will not give proofs of the facts given in this section, these can be found in the textbooks [1], [4], [12], [13] and [23].

2.1.1 Wavelets

2.1 Definition: Let $\varphi \in L^2(\mathbb{R})$ be a real-valued function, and let $\varphi_{jk}(x) := 2^{\frac{j}{2}}\varphi\big(2^j x - k\big)$ for $j \in \mathbb{Z}$ fixed. Assume the sequence of functions $(\varphi_{jk})_{k\in\mathbb{Z}}$ is $L^2(\mathbb{R})$-orthonormal and write $V_j := \operatorname{span}\big\{\varphi_{jk} \colon k \in \mathbb{Z}\big\}$. Finally, assume $V_j \subset V_{j+1}$ for any $j \in \mathbb{Z}$.

Then, φ is called *father wavelet* or *scaling function* and the sequence $(V_j)_{j\in\mathbb{Z}}$ is called *multi-resolution analysis* (MRA). ◊

The construction of such functions φ is described in [12, Chap. 5-7]. In what follows, we will assume φ to be a compactly supported $\mathcal{C}^l$-function, $\text{supp}\varphi \subset [-A, A]$. The MRA is then called *l-regular.*

2.2 Theorem: *Let φ be a scaling function and let the MRA $(V_j)_{j\in\mathbb{Z}}$ be l-regular. Besides, let $W_j \subset L^2(\mathbb{R})$ be the orthogonal complement of V_j in V_{j+1}, i. e. $V_{j+1} = V_j \oplus W_j$.*

Then, there is a compactly supported $\mathcal{C}^l$-function $\psi : \mathbb{R} \to \mathbb{R}$, such that the sequence $(\psi_{0k})_{k\in\mathbb{Z}}$ is a $L^2(\mathbb{R})$-orthonormal basis of W_0, and the functions $(\psi_{jk})_{j,k\in\mathbb{Z}}$ are an orthonormal basis of $L^2(\mathbb{R})$.

Here, we write $\psi_{jk}(x) := 2^{\frac{j}{2}}\psi\big(2^j x - k\big)$ similarly to above.

2.3 Definition: The function ψ from the preceding theorem is called *mother wavelet* or simply *wavelet function.* ♢

Again, we refer to [12] for the construction of ψ. Given φ, ψ is not unique in general.

For fixed $j_0 \in \mathbb{Z}$, we have the decomposition

$$L^2(\mathbb{R}) = V_{j_0} \oplus W_{j_0} \oplus W_{j_0+1} \oplus \cdots . \tag{2.2}$$

In particular, there is a unique representation for any $f \in L^2(\mathbb{R})$:

$$f = \sum_{k\in\mathbb{Z}} \alpha_{j_0k}\varphi_{j_0k} + \sum_{j\geqslant j_0}\sum_{k\in\mathbb{Z}} \beta_{jk}\psi_{jk},$$

with

$$\alpha_{jk} = \int_{\mathbb{R}} f(x)\varphi_{jk}(x)\mathrm{d}x, \qquad \beta_{jk} = \int_{\mathbb{R}} f(x)\psi_{jk}(x)\mathrm{d}x. \tag{2.3}$$

Often, $j_0 = 0$ is chosen. In what follows here, j_0 will always be non-negative. One can show

$$\int_{\mathbb{R}} \varphi(x)\mathrm{d}x = 1, \qquad \int_{\mathbb{R}} \psi(x)\mathrm{d}x = 0.$$

2.4 Example (Haar's and Daubechies' wavelets, see [4, Chap. 6-7]): The simplest and the historically first wavelet function is Haar's wavelet:

$$\varphi(x) = \mathbb{1}_{[0,1]}(x), \qquad \psi(x) = \begin{cases} 1 & \text{for } 0 \leqslant x < \frac{1}{2}, \\ -1 & \text{for } \frac{1}{2} \leqslant x \leqslant 1. \end{cases}$$

There is no compactly supported $\mathcal{C}^\infty$-wavelet, see [13, Thm. 2.3.8]. On the other hand, for any $l \in \mathbb{N}$ there exists some l-regular MRA with corresponding compactly supported functions φ and ψ. There are several different wavelet-families of this kind, such as Daubechies' wavelets, which are designated by $D2N$, $N \in \mathbb{N}$. For $D2N$ one finds

$$\text{supp}\varphi \subset [0, 2N - 1], \qquad \text{supp}\psi \subset [-N + 1, N].$$

For $N \geqslant 3$, Daubechies' wavelets are continuously differentiable. The case $N = 1$ corresponds to Haar's wavelet, which is the only non-continuous member in the family of Daubechies' wavelets. On page 14, we show mother and father wavelets, respectively, of Daubechies' wavelets for $N = 1, \ldots, 4$. ♢

2.1.2 Besov spaces

In what follows, we will prove several upper bounds for functions contained in Besov spaces. These spaces are of interest to us, because the functions contained therein can be described via their wavelet expansion solely. Sobolev spaces, for example, do not have this property. Besides, Besov spaces are optimal in a certain respect as far as linear minimax estimators are concerned (see [19]). We will not need this here, however.

2.5 Definition: The *Besov space* B_{spq} with parameters $s > 0$, $1 \leqslant p, q \leqslant \infty$ is the set of all functions $f : \mathbb{R} \to \mathbb{R}$ that satisfy

$$\|f\|_{spq} := \|\alpha_{0\bullet}\|_{\ell^p} + \left(\sum_{j \geqslant 0} \left(2^{j(s+\frac{1}{2}-\frac{1}{p})} \|\beta_{j\bullet}\|_{\ell^p} \right)^q \right)^{\frac{1}{q}} < \infty. \tag{2.4}$$

Here, $(\alpha_{0k})_{k\in\mathbb{Z}}$ and $(\beta_{jk})_{j,k\in\mathbb{Z}}$ are the wavelet coefficients of f. ♢

The space B_{spq} does not depend on a particular scaling function or wavelet function.

Alternatively, one can define B_{spq} via the norm

$$J_{spq}(f) := \|E_0(f)\|_{L^p} + \left(\sum_{j \geqslant 0} \left(2^{js} \|D_j f\|_{L^p} \right)^q \right)^{\frac{1}{q}}. \tag{2.5}$$

Here, we denote by E_i and D_j projections on the spaces V_i and W_j respectively (see (2.3)):

$$E_i f(x) := \sum_{k\in\mathbb{Z}} \alpha_{ik}\varphi_{ik}(x) = \sum_{k\in\mathbb{Z}} \int_{\mathbb{R}} \varphi_{ik}(y) f(y) \mathrm{d}y \, \varphi_{ik}(x), \tag{2.6}$$

$$D_j f(x) := \sum_{k\in\mathbb{Z}} \beta_{jk}\psi_{jk}(x) = \sum_{k\in\mathbb{Z}} \int_{\mathbb{R}} \psi_{jk}(y) f(y) \mathrm{d}y \, \psi_{jk}(x). \tag{2.7}$$

Besides, let

$$D_{j_0 j_1} f(x) := \sum_{j=j_0}^{j_1} \sum_{k\in\mathbb{Z}} \beta_{jk}\psi_{jk}(x). \tag{2.8}$$

The norms J_{spq} and $\|\cdot\|_{spq}$ are equivalent (see e. g. [12, Cor. 9.1]).

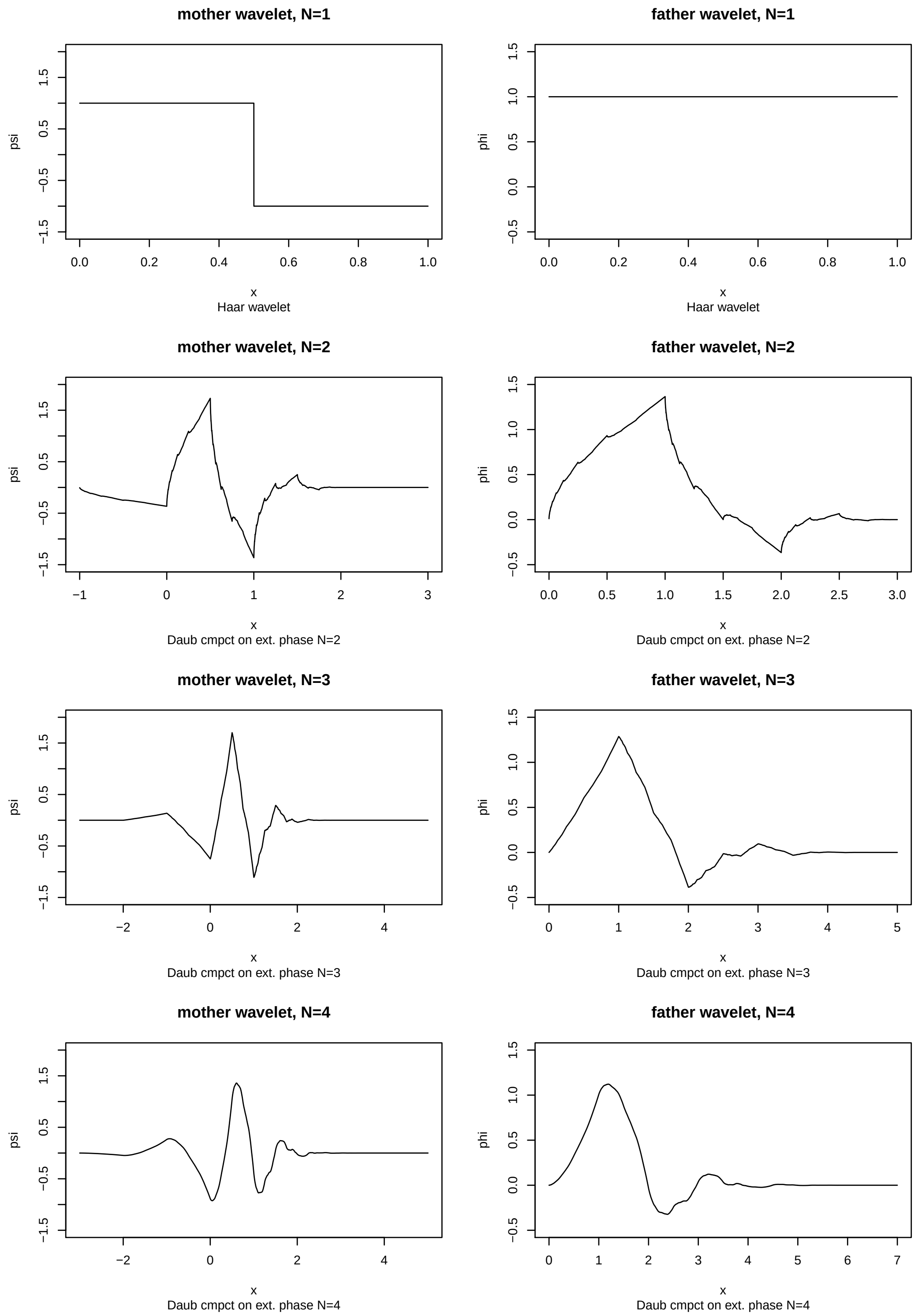


Figure 2.1: Daubechies' Wavelets $D2$, $D4$, $D6$ and $D8$. Note, that the four functions in the two bottom lines are continuously differentiable. These graphs have been drawn with the `R` package `wavethresh`, [25], [27].

2.6 Theorem: *Let g be a scaling function or a wavelet function, let $\vartheta_g(x) := \sum_{k\in\mathbb{Z}} |g(x-k)|$. Besides, let*

$$f(x) = \sum_{k\in\mathbb{Z}} \lambda_k 2^{\frac{j}{2}} g\big(2^j x - k\big), \qquad 1 \leqslant p \leqslant \infty,$$

and choose p' such that $\frac{1}{p} + \frac{1}{p'} = 1$. Then, we have

$$c_1 2^{j(\frac{1}{2}-\frac{1}{p})} \|\lambda\|_{\ell^p} \leqslant \|f\|_{L^p} \leqslant c_2 2^{j(\frac{1}{2}-\frac{1}{p})} \|\lambda\|_{\ell^p},$$

with $c_1 = \left(\|\vartheta_g\|_1^{\frac{1}{p'}} \|\vartheta_g\|_\infty^{\frac{1}{p}} \right)^{-1}$ *and* $c_2 = \|\vartheta_g\|_p$.

The proof can be found in [12, Prop. 8.3].

Note the difference between the norms $\|\cdot\|_p$ and $\|\cdot\|_{L^p}$, $p \in [1,\infty)$:

$$\|f\|_p := \left(\int_0^1 |f(x)|^p \, \mathrm{d}x \right)^{\frac{1}{p}} \qquad \|f\|_{L^p} := \left(\int_{\mathbb{R}} |f(x)|^p \, \mathrm{d}x \right)^{\frac{1}{p}} \qquad \|f\|_\infty := \sup_{x\in\mathbb{R}} |f(x)|.$$

The norm $\|\cdot\|_{\ell^p}$ is the usual norm in sequential spaces. In what follows, Theorem 2.6 will always be used whenever we need to switch between $\|\cdot\|_{L^p}$ and $\|\cdot\|_{\ell^p}$. The theorem says that these norms are equivalent if they are applied to functions that can be expanded in a wavelet series.

2.7 Theorem (Besov embedding theorem): *Let $s' := s - \frac{1}{p} + \frac{1}{r}$. The following inclusions hold:*

$$\begin{aligned}
B_{s_1pq_1} &\subset B_{s_2pq_2} && \text{for } s_1 > s_2 \text{ or } s_1 = s_2 \text{ and } q_1 \leqslant q_2.\\
B_{spq} &\subset B_{s'rq} && \text{for } r \geqslant p.\\
B_{spq} &\subset B_{s'\infty\infty} && \text{for } s - \frac{1}{p} > 0 \text{ and } q > 1 \text{ or for } s - \frac{1}{p} \geqslant 0 \text{ and } q = 1.\\
B_{0r(r\wedge 2)} &\subset L^r && \text{for } r \geqslant 1,
\end{aligned}$$

Here, B_{0pq} is defined via the Besov norm $\|\cdot\|_{spq}$ with $s = 0$.

Besides, the according norms satisfy the following inequalities:

$$\begin{aligned}
\|f\|_{s_1pq_1} &\geqslant \|f\|_{s_2pq_2} && \text{for } s_1 > s_2 \text{ or } s_1 = s_2 \text{ and } q_1 \leqslant q_2.\\
\|f\|_{spq} &\geqslant \|f\|_{s'rq} && \text{for } r \geqslant p.\\
\|f\|_{spq} &\geqslant \|f\|_{s'\infty\infty} && \text{for } s - \frac{1}{p} > 0 \text{ and } q > 1 \text{ or for } s - \frac{1}{p} \geqslant 0 \text{ and } q = 1.\\
\|f\|_{0r(r\wedge 2)} &\geqslant \|f\|_{L^r} && \text{for } r \geqslant 1.
\end{aligned}$$

Proofs of these statements can be found in [12, Cor. 9.2].

We will designate the *Besov balls* with

$$B_{spq}(M) = \{f \in B_{spq} \colon \|f\|_{spq} \leqslant M\}.$$

2.2 Our setting and assumptions

We consider the wavelet expansion of the unknown periodic function

$$S(t) = \sum_{k\in\mathbb{Z}} \alpha_{j_0k}\varphi_{j_0k}(t) + \sum_{j\geqslant j_0}\sum_{k\in\mathbb{Z}} \beta_{jk}\psi_{jk}(t), \qquad t\in\mathbb{R},$$

where φ and ψ are the father and mother wavelet, respectively, compactly supported and l-times continuously differentiable (take Daubechies' wavelets with $N \geqslant 3$ for example). The functions $\varphi_{jk}(x) = 2^{\frac{j}{2}}\varphi(2^j x - k)$ and $\psi_{jk}(x) = 2^{\frac{j}{2}}\psi(2^j x - k)$ for $j, k \in \mathbb{Z}$ are denoted in the usual way. The wavelet coefficients are defined according to (2.3). Let $s > 0$, $1 \leqslant p, q \leqslant \infty$. We will work with the space of all functions with bounded Besov norm and bounded absolute value (both bounds are the same constant $M < \infty$), in short-hand notation:

$$\tilde{B}_{spq}(M) := \{f : \mathbb{R} \to \mathbb{R} \mid \|f\|_{spq} < M \text{ and } \|f\|_\infty < M\}.$$

Let us make the following model assumptions.

Assumption (A):
1. Let the parameters s, p, q, M of the Besov ball be known, and let $s > \frac{1}{p}$.
2. Let $S \in \tilde{B}_{spq}(M)$ be unknown, periodic with known periodicity T.
3. Let S, b, σ be Lipschitz functions and let σ be bounded. b is known. ♢

If (A) is fulfilled, the stochastic differential equation (2.1) has a unique strong solution (see [8, Sect. 5.2]).

It does not matter for the rest of this thesis whether σ is known or unknown. Mostly, we will only refer to it via its maximal value $\|\sigma\|_\infty = \text{const} < \infty$. If σ is unknown, it can be estimated via the empirical quadratic variation of the observed path of the process ξ.

We observe the process ξ continuously in time, for n periods. We will cut the paths of ξ in parts, according to $\mathbb{X}_i := (\xi_{(i-1)T+s})_{s\in[0,T]} \in \mathcal{C}([0,T])$. Similarly, we will use the notation $\mathbb{W}_i := (W_{(i-1)T+s} - W_{(i-1)T})_{s\in[0,T]}$. For further reference, we denote by $\mathcal{E}$ the canonical σ-field of $\mathcal{C}([0,T])$. The parts $(\mathbb{X}_i)_{i\geqslant 1}$ form a Markov chain, taking values in $\mathcal{C}([0,T])$.

In what follows, we will refer to the $\mathbb{X}_i$ as our observations, and our main result deals with upper bounds of the risk, asymptotic as $n \to \infty$, the number of our observations.

2.8 Remark: The process ξ is constructed as a strong solution of (2.1), and as such it depends only on the first value ξ_0 and on the Brownian Motion W. In [18], a strong solution is represented as a functional of the initial condition and of the driving Brownian Motion. This functional is to be defined on the (polish) product space $\mathbb{R} \times \mathcal{C}_0([0,T])$, where $\mathcal{C}_0([0,T])$ denotes the continuous functions on $[0,T]$ starting in 0. The functional then needs to obey certain measurability conditions that can be found in [18, p. 163].

By considering the Brownian Motion $\mathbb{W}_1$ on the interval $[0,T]$, we construct ξ up to time T using the functional of [18], and we get the first segment $\mathbb{X}_1$. Then, in the second segment, we solve (2.1) with the starting value $\xi_T = \mathbb{X}_1(T) = \mathbb{X}_2(0)$, and the driving Brownian Motion on $[0,T]$, denoted by $\mathbb{W}_2$, such that we get ξ up to time $2T$ and hence $\mathbb{X}_2$. Iterating this plan on the i-th segment, we construct all $\mathbb{X}_i$, and each depends functionally only on $\mathbb{X}_{i-1}(T) = \mathbb{X}_i(0)$ and on $\mathbb{W}_i$. In summary, we find the path segment $\mathbb{X}_i$ functionally dependent on a real number and a Brownian Motion $\mathbb{W}_i$ on $[0,T]$. ◊

Of course, once we know all its wavelet coefficients, we know everything about the unknown function S. So, we have to look at the

2.9 Definition: For $j,k \in \mathbb{Z}$ we define estimators for the wavelet coefficients α_{jk} and β_{jk} by

$$\hat{\alpha}_{jk,n} := \frac{1}{n}\sum_{i=1}^{n}\int_0^T \varphi_{jk}(t)\Big(\mathrm{d}\mathbb{X}_i(t) + b\big(\mathbb{X}_i(t)\big)\mathrm{d}t\Big)$$

and

$$\hat{\beta}_{jk,n} := \frac{1}{n}\sum_{i=1}^{n}\int_0^T \psi_{jk}(t)\Big(\mathrm{d}\mathbb{X}_i(t) + b\big(\mathbb{X}_i(t)\big)\mathrm{d}t\Big). \qquad \diamond$$

Note, that by assumption (A) the function b is known. Hence, $\hat{\alpha}_{jk,n}$ and $\hat{\beta}_{jk,n}$ can be computed solely from the observations and are estimators, not oracles.

By definition of φ_{jk} and ψ_{jk}, the support of these functions decreases as j increases; since we will look at the asymptotics for "many observations" (which implicitly increase the relevant levels j), we can assume w.l.o.g. $\mathrm{supp}\varphi_{jk} \subset [0,T]$ and $\mathrm{supp}\psi_{jk} \subset [0,T]$.

2.10 Proposition: *The estimators $\hat{\alpha}_{jk,n}$ and $\hat{\beta}_{jk,n}$ for the coefficients α_{jk} and β_{jk} have a martingale-style error:*

$$\hat{\alpha}_{jk,n} - \alpha_{jk} = \frac{1}{n}\sum_{i=1}^{n}\int_0^T \varphi_{jk}(t)\sigma\big(\mathbb{X}_i(t)\big)\mathrm{d}\mathbb{W}_i(t),$$
$$\hat{\beta}_{jk,n} - \beta_{jk} = \frac{1}{n}\sum_{i=1}^{n}\int_0^T \psi_{jk}(t)\sigma\big(\mathbb{X}_i(t)\big)\mathrm{d}\mathbb{W}_i(t).$$

Proof. We only prove the assertion for the $\hat{\alpha}_{jk,n}$, because the argument for the $\hat{\beta}_{jk,n}$ carries over by exchanging ψ_{jk} for φ_{jk}. By the definition of $\hat{\alpha}_{jk,n}$, of $\mathbb{X}_i$ and by (2.1), we have

$$\begin{aligned}\hat{\alpha}_{jk,n} &= \frac{1}{n}\sum_{i=1}^{n}\int_0^T \varphi_{jk}(t)\Big(\mathrm{d}\mathbb{X}_i(t) + b\big(\mathbb{X}_i(t)\big)\mathrm{d}t\Big)\\ &= \frac{1}{n}\sum_{i=1}^{n}\int_0^T \varphi_{jk}(t)\mathrm{d}\xi_{(i-1)T+t} + \frac{1}{n}\sum_{i=1}^{n}\int_0^T \varphi_{jk}(t)b\big(\xi_{(i-1)T+t}\big)\mathrm{d}t\end{aligned}$$

$$
\begin{aligned}
&= \frac{1}{n}\sum_{i=1}^{n}\int_0^T \varphi_{jk}(t)S\big((i-1)T+t\big)\mathrm{d}t - \frac{1}{n}\sum_{i=1}^{n}\int_0^T \varphi_{jk}(t)b(\xi_{(i-1)T+t})\mathrm{d}t \\
&\qquad + \frac{1}{n}\sum_{i=1}^{n}\int_0^T \varphi_{jk}(t)\sigma(\xi_{(i-1)T+t})\mathrm{d}W_{(i-1)T+t} \\
&\qquad + \frac{1}{n}\sum_{i=1}^{n}\int_0^T \varphi_{jk}(t)b\big(\xi_{(i-1)T+t}\big)\mathrm{d}t.
\end{aligned}
$$

Now, since S is T-periodic, and then by definition of the α_{jk},

$$
\begin{aligned}
\hat{\alpha}_{jk,n} &= \frac{1}{n}\sum_{i=1}^{n}\int_0^T \varphi_{jk}(t)S(t)\mathrm{d}t + \frac{1}{n}\sum_{i=1}^{n}\int_0^T \varphi_{jk}(t)\sigma(\xi_{(i-1)T+t})\mathrm{d}W_{(i-1)T+t} \\
&= \alpha_{jk} + \frac{1}{n}\sum_{i=1}^{n}\int_0^T \varphi_{jk}(t)\sigma\big(\mathbb{X}_i(t)\big)\mathrm{d}\mathbb{W}_i(t).
\end{aligned}
\tag{2.9}
$$

This is what we wanted. ▨

Therefore, the estimator $\hat{\alpha}_{jk,n}$ yields the "correct" α_{jk}, with a martingale-style perturbation. Unfortunately, we cannot subtract this perturbation from our estimator, because it depends on the unobservable W.

By the definition of the estimators for the wavelet coefficients of S, we arrive at the sequence of estimators

$$
\hat{S}_n(t) = \sum_{k\in\mathbb{Z}} \hat{\alpha}_{j_0k,n}\varphi_{j_0k}(t) + \sum_{j\geqslant j_0}\sum_{k\in\mathbb{Z}} \hat{\beta}_{jk,n}\psi_{jk}(t).
$$

This is only a formal sum (we do not care about its convergence, and we will manipulate it a little more soon). Given n observed periods of the process ξ, we shall calculate the error of estimation of S in an L^r-norm with respect to the invariant measure m of the Markov chain $\mathbb{X}_i$ (see Assumption (B) below):

$$
E_m \left\|\hat{S}_n - S\right\|_{L^r}, \qquad r > 2. \tag{2.10}
$$

As it has been stated in [6, Sect. 5], one can improve $\hat{S}_n$ by defining two indices $j_0, j_1 \in \mathbb{Z}$, in order to ignore the "too fine details" of S in the wavelet expansion. All sums will become finite afterwards; the sums indexed by k are finite anyway, since the function S is only considered on the compact interval $[0, nT]$. Only after we are done bounding the error for fixed n, we will talk about asymptotics for $n \to \infty$.

Besides, one can improve $\hat{S}_n$ by cutting away the coefficients $\hat{\beta}_{jk,n}$ at a positive threshold ("hard thresholding"):

$$
\tilde{\beta}_{jk,n} := \hat{\beta}_{jk,n}\mathbb{1}_{\{|\hat{\beta}_{jk,n}|>j^{\nu}n^{-\nu}\}}. \tag{2.11}
$$

We choose a suitable *positive* power of $\frac{j}{n}$ for threshold. In [6], $\nu = \frac{1}{2}$ was used, multiplied with a constant factor. In what follows, we will not use a constant factor, because it will not matter for the asymptotics later on. Besides, we will find out in the proof of Theorem 2.14 (in (3.14) of the proof of Theorem 3.12 to be exact) that we have to choose $\nu \in \left(0, \frac{1}{2}\right)$; thus, the methods of [6], that were designed for the case of i.i.d. observations, cannot be employed entirely in our setting of dependent observations.

Other methods of thresholding can be thought of, such as soft thresholding (where the wavelet coefficients that are smaller than the threshold are shrunk instead of deleted) or block thresholding (where one considers the coefficients not individually, but decides for a larger set of coefficients whether they are too small or not). We stick to the hard thresholding here, since it will do the trick for what we want to achieve with comparatively little technical effort.

Thus, our final estimator is

$$\mathrm{TW}_n(t) := \sum_{k\in\mathbb{Z}} \hat{\alpha}_{j_0k,n}\varphi_{j_0k}(t) + \sum_{j=j_0}^{j_1}\sum_{k\in\mathbb{Z}} \tilde{\beta}_{jk,n}\psi_{jk}(t). \tag{2.12}$$

Note as well that our notation is slightly different from the notation in [6] (e. g. j_0 and j_1 have swapped their roles).

Let us introduce more notation:

2.11 Definition: Let the fixed and known parameters s, p, q determine the Besov space containing the function S, which we want to estimate, by assumption (A). Let $r > 2$ be the same r that we used in (2.10). Put

$$s' := s - \frac{1}{p} + \frac{1}{r},$$
$$\alpha := \min\left\{\frac{s}{1+2s}, \frac{s - \frac{1}{p} + \frac{1}{r}}{1 + 2s - \frac{2}{p}}\right\},$$
$$\varepsilon := sp - \frac{r-p}{2}.$$

◇

Besides, we will need the following notions from the theory of Markov chains (more details and theorems on this can be found in [24]):

2.12 Definition: *a)* Let Π be the transition probability of the Markov chain $(\mathbb{X}_i)_{i\geqslant 1}$, and assume Harris recurrence of $(\mathbb{X}_i)_{i\geqslant 1}$ with stationary probability m. A set $A \in \mathcal{E}$ with $m(A) > 0$ is called an *atom* for X if there is a measure μ on $\mathcal{E}$ such that $\Pi(x, B) = \mu(B)$ for all $x \in A$, $B \in \mathcal{E}$.

b) Let A be an atom of the chain and let $\tau_A := \min\{i \geqslant 1 \colon \mathbb{X}_i \in A\}$ be the first return time to A. The atom A is called *χ-regular*, $\chi > 0$, if τ_A satisfies

$$E_A(\tau_A^\chi) < \infty.$$

Here, E_A denotes the expected value with respect to a starting point $x \in A$. Since A is an atom, this expected value is the same for all $x \in A$.

c) An atom A is called *geometrically regular*, if there is some constant $K > 0$ with

$$E_A\big(\exp(K\tau_A)\big) < \infty. \qquad \diamondsuit$$

The reason why we consider atoms of a Markov chain is that we can use an i.i.d. structure in our proofs. Since the distribution of the next step of the chain does not depend on a particular point in the atom, the Markov chain "forgets" about where it exactly is. So, once the chain enters the atom, the excursions between two entries to the atom are independent of one another and have the same distribution. This allows to apply theorems that deal with independent random variables and to break up the dependence of the Markov chain $(\mathbb{X}_i)_{i\geqslant 1}$. The notions of regularity which we gave above lead to the conclusion that the atom of our Markov chain will be visited often enough for our purposes. Details on these so-called *life-cycle decompositions* can be found in [16, p. 13-16], for instance.

With these notions, we can state the following assumption:

Assumption (B): The Markov chain $(\mathbb{X}_i)_{i\geqslant 1}$, consisting of the "cut parts" of the process ξ,

1. is positive Harris recurrent with stationary probability m;
2. allows for Nummelin splitting, i.e. there is a set $D \in \mathcal{E}$, the canonical σ-field of $\mathcal{C}([0,T])$, with $m(D) > 0$, there is a probability measure μ on $\mathcal{E}$ such that $\mu(D) = 1$, and some $\delta > 0$ with

$$\Pi(x, B) \geqslant \delta \mathbb{1}_D(x)\mu(B), \qquad \text{for all } x \in \mathcal{C}([0,T]),\ B \in \mathcal{E}; \tag{2.13}$$

3. is geometrically regular, (i.e. the "split chain" possesses a geometrically regular atom). $\diamondsuit$

2.13 Remark: *a)* In [17, Thm. 1.1], combined with [14, Thm. 2.1], one can find mild and easily verifiable conditions for assumptions (B1.) and (B2.). These are essentially growth conditions on the functions S, b and σ in (2.1).

b) Obviously, geometrical regularity implies χ-regularity for every $0 < \chi < \infty$. We will use χ-regularity for some fixed $\chi > r > 2$ in the proof of Theorem 3.6. Here, r is the number that designates the L^r-norm measuring the error of estimation, see (2.10). In particular, it is not enough to use r-regularity for this fixed r, because we will need χ-regularity in the proof of Lemma 3.25. On the other hand, we do not require χ-regularity for arbitrarily large χ. Taking a fixed multiple of r that guarantees $\chi > r$ will do the trick, see the proof of Lemma 3.25 for the exact kind of conditions needed.

We use geometrical regularity in the proof of Theorem 3.9; so we do not have to actually worry about the optimal χ for χ-regularity, because we need stronger assumptions anyway.

c) It seems possible to relax the assumption of geometrical regularity by some suitable χ-th regularity condition. In [2] and [22], polynomial moments for hitting times of certain sets in the state space of a Markov chain are linked to so-called Nash inequalities which might imply the large deviation inequality which we use in Theorem 3.9. If this idea can be made rigorous, we could do without geometrical regularity. Since this appears to be a rather deep connection, we do not pursue on these lines.

d) Under assumption (B2.) one can construct a "split chain" from the chain $(\mathbb{X}_i)_{i\geqslant 1}$, which possesses an atom. This technique has been introduced by Nummelin in 1978, [26]. See also [24, Sect. 5.1] for more on this construction. This "artificial" atom is the key to the theorems in [3] which, in turn, are the key to our Theorems 3.6 and 3.9 and our main auxiliary theorems in Section 3.4.

e) For a Markov chain defined on a state space with countably generated σ-field, Nummelin splitting is always possible, at least for some k-step transition probability. See [24, Sect. 5.2] for details on this. In our setting, we deal with the canonical σ-field on the polish space $\mathcal{C}([0,T])$ which is indeed countably generated. Therefore, we can assume that (2.13) holds, at least if Π is replaced by Π^k for some $k \in \mathbb{N}$. Then, we can combine k periods of the process ξ (i. e. k consecutive elements of $(\mathbb{X}_i)_{i\geqslant 1}$) and replace the periodicity T by kT, to arrive at (2.13) itself.

In summary, we can assume w.l.o.g. that (2.13) holds for the chain $(\mathbb{X}_i)_{i\geqslant 1}$, and thus the "split chain" possesses an atom. In what follows, we are going to work with the "split chain" without further reference to this.

f) Assumption (B2.) implies that the chain $(\mathbb{X}_i)_{i\geqslant 1}$ is aperiodic. By construction of the "split chain" (see [24, Sect. 5.1]), every time the atom is visited, the chain may stay in the atom for the next step, or the chain may leave the atom. Since arbitrarily many successive steps in the atom are possible, there can be no periodicity in the chain $(\mathbb{X}_i)_{i\geqslant 1}$. ♢

Finally, we can state our main theorem now.

2.14 Theorem: *Let the assumptions (A) and (B) be satisfied and* $1 \leqslant p \leqslant r < \infty$, $r > 2$, *and let* s', ε, α *be as in Definition 2.11. Besides, let*

$$2^{j_0(n)} \asymp \left(n(\log n)^{\frac{r-p}{p}\mathbb{1}_{\{\varepsilon\geqslant 0\}}}\right)^{1-2\alpha}, \qquad 2^{j_1(n)} \asymp \left(\frac{n}{\log n}\right)^{\frac{\alpha}{s'}}. \tag{2.14}$$

Let the threshold ν *in the definition of* TW_n *be in the interval* $(0, \frac{1}{2})$. *Then, there is a positive, real constant* C, *not depending on* n, *such that we have*

$$\sup_S E_m \|\mathrm{TW}_n - S\|_{L^r}^r \leqslant \begin{cases} C\left(\frac{\log n}{n}\right)^{\alpha r}, & \varepsilon \neq 0, \\ C(\log n)^{(\frac{1}{2}-\nu)p}\left(\frac{\log n}{n}\right)^{\alpha r}, & \varepsilon = 0. \end{cases} \tag{2.15}$$

We take the supremum over all functions that assumption (A) allows: contained in a Besov space B_{spq} with bounded Besov norm and bounded sup-norm, and Lipschitz continuous.

2.15 Remark: *a)* As far as the assumptions on r are concerned, we have added the condition $r > 2$ which was not made in [6]. We shall indicate at which places in the proof this is important and why.

b) In [6, Thm. 3], the authors considered a wavelet estimator TW for a probability density in a Besov space B_{spq} using classical i.i.d. observations. They gave an upper bound for the error of estimation as

$$\sup_f E\,\|\mathrm{TW} - f\|_{L^r}^r \leqslant \begin{cases} C(\log n)^{-\frac{\alpha r \varepsilon}{sp}} \left(\frac{\log n}{n}\right)^{\alpha r}, & \text{for } \varepsilon > 0 \\ C(\log n)^{\left(\frac{r}{2}-\frac{p}{q}\right)^+} \left(\frac{\log n}{n}\right)^{\alpha r}, & \text{for } \varepsilon = 0, \\ C\left(\frac{\log n}{n}\right)^{\alpha r}, & \text{for } \varepsilon < 0. \end{cases} \tag{2.16}$$

However, this bound was shown in [5, Satz 2.26] to be incorrect in case $\varepsilon > 0$. The first $(\log n)$-factor needs to be dropped to yield the rate

$$\sup_f E\,\|\mathrm{TW} - f\|_{L^r}^r \leqslant \begin{cases} C\left(\frac{\log n}{n}\right)^{\alpha r}, & \text{for } \varepsilon \neq 0, \\ C(\log n)^{\left(\frac{r}{2}-\frac{p}{q}\right)^+} \left(\frac{\log n}{n}\right)^{\alpha r}, & \text{for } \varepsilon = 0. \end{cases} \tag{2.17}$$

A proof of this rate (2.17) is given in [5, Sect. 2.4]. In the present thesis, Lemma 3.24 corresponds to [6, Satz 2.26], which is where the reader can find the arguments that lead to bound (2.17) instead of (2.16).

In particular, the rate in (2.15) for $\varepsilon > 0$ is not caused by our model or the dependence of our observations, this is the same rate to use for an i.i.d.-setting.

c) In [5, Hauptsatz 2.35], the slightly better rate of convergence in (2.16) in case $\varepsilon > 0$ has been achieved by introducing a correcting factor in the choice of $j_1(n)$. In our present model of dependent observations, there does not seem to be an obvious way to get the rate of convergence from [6] in all cases. Appropriately introducing a correcting factor in the choice of $j_1(n)$ will improve the bound of Lemma 3.27, which is responsible for the worse bound in case $\varepsilon = 0$. On the other hand, this correcting factor will immediately deteriorate the bound of Lemma 3.24; this deterioration has more impact than the improvement.

d) For $\varepsilon = 0$, we encounter an entirely different rate of convergence for TW_n with our dependent observations, than in the case of i.i.d. observations:

- Indeed, in case $\varepsilon = 0$ and $\frac{r}{2} - \frac{p}{q} < 0$, the bound in (2.15) is worse by a power of $\log n$ compared to [6, Thm. 3], i. e. (2.16).

- In case $\varepsilon = 0$ and $\frac{r}{2} - \frac{p}{q} \geqslant 0$ and $\nu \geqslant \frac{1}{2} + \frac{1}{q} - \frac{r}{2p}$, we get

$$\nu \geqslant \frac{1}{2} + \frac{1}{q} - \frac{r}{2p} \quad \iff \quad \frac{1}{2} - \nu \leqslant \frac{r}{2p} - \frac{1}{q} \quad \iff \quad \left(\frac{1}{2} - \nu\right) p \leqslant \frac{r}{2} - \frac{p}{q}.$$

  This implies that, in this case, the bound from (2.15) is better than the rate (2.16) from [6]. However, ν needs to satisfy both conditions $\nu < \frac{1}{2}$ and $\nu \geqslant \frac{1}{2} + \frac{1}{q} - \frac{r}{2p}$ to get this better bound; there are cases where this is not possible.

Note, that in this case $\varepsilon = 0$, (2.16) and (2.17) coincide.

e) Obviously, it is preferable to choose a threshold ν as close as possible to $\frac{1}{2}$. The smaller $\frac{1}{2} - \nu$ is, the better the bound (2.15) will be. As the proof of Theorem 3.12 will show, such a choice also improves the constant C. $\diamondsuit$

Whenever we write C, we mean "some positive real constant, independent of n and j_0 or j_1" and do not care about its value. In particular, the expression C may be different every time it appears.

Chapter 3

Proof of the main result

In order to prove Theorem 2.14 on the asymptotic upper bounds of the risk of our estimator, we shall need suitable upper bounds of the expected errors of the single estimators for the coefficients (Theorem 3.11) and of a large deviation probability (Theorem 3.12). Both of these statements are based on the article [3] by Clémençon. Later, they will allow us to make use of large parts of the proof from [6, Thm. 3].

3.1 Functionals and Oracles

Let us define a real-valued functional on our observations $\mathbb{X}_i$, the path segments of the process ξ that form a discrete-time Markov chain:

3.1 Definition: Using the construction and notation of Remark 2.8, we set

$$\Phi_n\big(\mathbb{X}_i(0), \mathbb{W}_i\big) := -\int_{T\wedge U_{n,i}}^{T} \varphi_{jk}(t) S(t) \mathrm{d}t + \int_0^{T\wedge U_{n,i}} \varphi_{jk}(t)\sigma\big(\mathbb{X}_i(t)\big)\mathrm{d}\mathbb{W}_i(t), \tag{3.1}$$

where

$$U_{n,i} := \inf\left\{t > 0\colon \left|\int_0^t \varphi_{jk}(s)\sigma\big(\mathbb{X}_i(s)\big)\mathrm{d}\mathbb{W}_i(s)\right| > n^\eta\right\}, \qquad 0 < \eta \leqslant \frac{r-2}{2r}, \quad \eta < 1-\nu, \tag{3.2}$$

is a stopping time. The parameter η depends on the r defining the L^r-norm that measures the error of our estimator TW_n, and on the threshold ν in the definition of our estimators $\tilde{\beta}_{jk,n}$ (see (2.11)). η is arbitrary in the bounds given above, but fixed. Note that $r > 2$ by assumption of Theorem 2.14.

By exchanging φ_{jk} by ψ_{jk}, we get a similar functional and a similar stopping time which we denote by Φ_n and $U_{n,i}$ as well. The parameter η is the same for both constructions. ◇

The reason for the choice of the parameter η, the "truncation-size" of the stopping times $U_{n,i}$, will become clearer later on (e. g. in the proof of Theorem 3.11 and in (3.13) and (3.15) of the proof of Theorem 3.12). It is designed to yield suitable asymptotic bounds on our error terms. The reason for this definition of a functional will be given in Proposition 3.4. The functional Φ_n will be a technically feasible way to handle the error of estimation.

Of course, the second integral in (3.1) would not be bounded, if we did not stop it at time $U_{n,i}$. The stopping time $U_{n,i}$ applies, if Φ_n gets "too large" for our purposes. More precisely:

3.2 Proposition: *Under assumption (A), the absolute value of the functional Φ_n is bounded by Cn^η for some constant $C > 0$.*

Proof. By definition, we have

$$\begin{aligned}\left|\Phi_n\big(\mathbb{X}_i(0), \mathbb{W}_i\big)\right| &= \left|\int_0^{T\wedge U_{n,i}} \varphi_{jk}(t)\sigma\big(\mathbb{X}_i(t)\big)\mathrm{d}\mathbb{W}_i(t) - \int_{T\wedge U_{n,i}}^T \varphi_{jk}(t)S(t)\mathrm{d}t\right| \\ &\leqslant \left|\int_0^{T\wedge U_{n,i}} \varphi_{jk}(t)\sigma\big(\mathbb{X}_i(t)\big)\mathrm{d}\mathbb{W}_i(t)\right| + \left|\int_{T\wedge U_{n,i}}^T \varphi_{jk}(t)S(t)\mathrm{d}t\right| \\ &\leqslant n^\eta + M\int_0^T |\varphi_{jk}(t)|\,\mathrm{d}t;\end{aligned}$$

in the last step, we used the definition of $U_{n,i}$ and the boundedness of the function S in absolute value by some $M < \infty$ (assumption (A)). Then, by the Cauchy-Schwarz inequality and by orthonormality of the φ_{jk}

$$\left|\Phi_n\big(\mathbb{X}_i(0), \mathbb{W}_i\big)\right| \leqslant n^\eta + MT^{\frac{1}{2}}\left(\int_0^T \varphi_{jk}^2(t)\mathrm{d}t\right)^{\frac{1}{2}} = n^\eta + MT^{\frac{1}{2}} \leqslant Cn^\eta. \tag{3.3}$$

Of course, the same argument applies for the construction of Φ_n with ψ_{jk} replaced by φ_{jk}. ▨

Since we only want to consider Φ_n on path segments of length T, the events $\{U_{n,i} < T\}$ will be of special interest to us. Fortunately, these events are quite rare:

3.3 Proposition: *Let $U := \bigcup_{i=1}^n \{U_{n,i} < T\}$ the event that at least in one of the path segments $\mathbb{X}_i$ the stopping time $U_{n,i}$ applies (i. e. the non-stopped stochastic integral in* (3.1) *takes too large values). Under assumptions (A) and (B)*

$$P_m(U) \leqslant Cn\exp\left(-\frac{n^{2\eta}}{C}\right),$$

where m is the stationary measure for the Markov chain $(\mathbb{X}_i)_{i\geqslant 1}$.

Proof. We apply [9, Thm. 3.3], which reads in our situation (we only deal with continuous martingales M)

$$P_m\left(\sup_{s\leqslant T}|M_s| \geqslant z, \quad \langle M\rangle_T \leqslant L\right) \leqslant 2\exp\left(-\frac{z^2}{2L}\right).$$

Naturally, our martingale M is the stopped stochastic integral from (3.1).

By assumption (A), σ is bounded by some constant, and so, for $t \in [0,T]$

$$\left\langle \int_0^t \varphi_{jk}(s)\sigma\big(\mathbb{X}_i(s)\big)\mathrm{d}\mathbb{W}_i(s) \right\rangle \leqslant C \int_0^t \varphi_{jk}^2(s)\mathrm{d}s \leqslant C.$$

In summary, we have

$$\begin{aligned} P_m(U_{n,i} < T) &= P_m\left(\left| \int_0^t \varphi_{jk}(s)\sigma\big(\mathbb{X}_i(s)\big)\mathrm{d}\mathbb{W}_i(s) \right| > n^\eta \text{ for some } t \in [0,T] \right) \\ &= P_m\left(\sup_{t\in[0,T]} \left| \int_0^t \varphi_{jk}(s)\sigma\big(\mathbb{X}_i(s)\big)\mathrm{d}\mathbb{W}_i(s) \right| > n^\eta \right) \\ &= P_m\left(\sup_{t\in[0,T]} \left| \int_0^t \varphi_{jk}(s)\sigma\big(\mathbb{X}_i(s)\big)\mathrm{d}\mathbb{W}_i(s) \right| > n^\eta, \right. \\ &\qquad\qquad\qquad \left. \int_0^t \varphi_{jk}^2(s)\sigma^2\big(\mathbb{X}_i(s)\big)\mathrm{d}s \leqslant C \right) \\ &\leqslant 2\exp\left(-\frac{1}{2}\frac{n^{2\eta}}{C}\right). \end{aligned} \tag{3.4}$$

Therefore,

$$P_m(U) = P_m\left(\bigcup_{i=1}^n \{U_{n,i} < T\} \right) \leqslant \sum_{i=1}^n 2\exp\left(-\frac{1}{2}\frac{n^{2\eta}}{C}\right) = Cn\exp\left(-\frac{1}{2}\frac{n^{2\eta}}{C}\right).$$

▨

Using the functional Φ_n, we can define an oracle for the wavelet coefficients of the unknown drift function S. This oracle will simplify the proof of our main result, even though we cannot compute it directly from our observations (it depends on the unobservable stopping times $U_{n,i}$).

3.4 Proposition: *Let*

$$\begin{aligned} \check{\alpha}_{jk,n} &:= \frac{1}{n}\sum_{i=1}^n \int_0^{T\wedge U_{n,i}} \varphi_{jk}(t)\Big(\mathrm{d}\mathbb{X}_i(t) + b\big(\mathbb{X}_i(t)\big)\mathrm{d}t\Big). \\ \check{\beta}_{jk,n} &:= \frac{1}{n}\sum_{i=1}^n \int_0^{T\wedge U_{n,i}} \psi_{jk}(t)\Big(\mathrm{d}\mathbb{X}_i(t) + b\big(\mathbb{X}_i(t)\big)\mathrm{d}t\Big). \end{aligned}$$

Then, we have

$$\check{\alpha}_{jk,n} - \alpha_{jk} = \frac{1}{n}\sum_{i=1}^n \Phi_n\big(\mathbb{X}_i(0),\mathbb{W}_i\big) \qquad \textit{and} \qquad \check{\beta}_{jk,n} - \beta_{jk} = \frac{1}{n}\sum_{i=1}^n \Phi_n\big(\mathbb{X}_i(0),\mathbb{W}_i\big), \tag{3.5}$$

and similarly,

$$\begin{aligned} \hat{\alpha}_{jk,n} - \check{\alpha}_{jk,n} &= \frac{1}{n}\sum_{i=1}^n \int_{T\wedge U_{n,i}}^T \varphi_{jk}(t)\Big(S(t)\mathrm{d}t + \sigma\big(\mathbb{X}_i(t)\big)\mathrm{d}\mathbb{W}_i(t)\Big) \qquad \textit{and} \\ \hat{\beta}_{jk,n} - \check{\beta}_{jk,n} &= \frac{1}{n}\sum_{i=1}^n \int_{T\wedge U_{n,i}}^T \psi_{jk}(t)\Big(S(t)\mathrm{d}t + \sigma\big(\mathbb{X}_i(t)\big)\mathrm{d}\mathbb{W}_i(t)\Big) \end{aligned} \tag{3.6}$$

with the corresponding constructions of Φ_n.

Proof. The assertions in (3.6) follow immediately from the definitions. For (3.5) we only prove the assertion for the α-terms. By arguments similar to the ones in Proposition 2.10,

$$\begin{aligned}\check{\alpha}_{jk,n} &= \frac{1}{n}\sum_{i=1}^{n}\int_0^{T\wedge U_{n,i}} \varphi_{jk}(t)\Big(\mathrm{d}\mathbb{X}_i(t) + b\big(\mathbb{X}_i(t)\big)\mathrm{d}t\Big)\\ &= \frac{1}{n}\sum_{i=1}^{n}\int_0^{T\wedge U_{n,i}} \varphi_{jk}(t)S(t)\mathrm{d}t + \frac{1}{n}\sum_{i=1}^{n}\int_0^{T\wedge U_{n,i}} \varphi_{jk}(t)\sigma\big(\mathbb{X}_i(t)\big)\mathrm{d}\mathbb{W}_i(t) + \alpha_{jk} - \alpha_{jk}.\end{aligned}$$

By definition of the wavelet coefficients, $\alpha_{jk} = \int_0^T \varphi_{jk}(t)S(t)\mathrm{d}t$, and so

$$\begin{aligned}\check{\alpha}_{jk,n} &= \alpha_{jk} - \frac{1}{n}\sum_{i=1}^{n}\int_{T\wedge U_{n,i}}^{T} \varphi_{jk}(t)S(t)\mathrm{d}t + \frac{1}{n}\sum_{i=1}^{n}\int_0^{T\wedge U_{n,i}} \varphi_{jk}(t)\sigma\big(\mathbb{X}_i(t)\big)\mathrm{d}\mathbb{W}_i(t)\\ &= \alpha_{jk} + \frac{1}{n}\sum_{i=1}^{n}\Phi_n\big(\mathbb{X}_i(0), \mathbb{W}_i\big).\end{aligned}$$ ▨

By our construction, the oracle $\check{\alpha}_{jk,n}$ and the estimator $\hat{\alpha}_{jk,n}$ are very similar, structurally. The advantage of the oracle is the boundedness of the martingale-style error term, that we have shown in Proposition 3.2. Besides, oracle and estimator are only different with exponentially low probability. More formally:

3.5 Corollary: *Let m be the stationary probability of the Markov chain $(\mathbb{X}_i)_{i\geqslant 1}$. Then:*

$$\begin{aligned}&P_m(\check{\alpha}_{jk,n} \neq \hat{\alpha}_{jk,n}) \leqslant Cn\exp\left(-\frac{n^{2\eta}}{C}\right) \qquad \textit{and}\\ &P_m(\check{\beta}_{jk,n} \neq \hat{\beta}_{jk,n}) \leqslant Cn\exp\left(-\frac{n^{2\eta}}{C}\right).\end{aligned}$$

Proof. This follows immediately from Proposition 3.3 and from (3.6), since estimator and oracle can only be different, if $U_{n,i} < T$ for some path segment $\mathbb{X}_i$. ▨

3.2 A Rosenthal-type inequality

The following theorem adopts [3, Props. 8 and 13] to our setting.

3.6 Theorem (Rosenthal-type inequality): *Let $\chi \geqslant r \geqslant 2$, with r designating the L^r-norm from Theorem 2.14, and let m be the stationary probability of the Markov chain $(\mathbb{X}_i)_{i\geqslant 1}$. Denote by A an atom of the Markov chain (which exists by Assumption (B) via Nummelin splitting) and by τ_A the first return time to A. Then, there is some constant $C > 0$, such that*

$$E_m\left|\sum_{i=1}^{n}\Phi_n\big(\mathbb{X}_i(0), \mathbb{W}_i\big)\right|^{\chi} \leqslant Cn^{\frac{\chi}{2}} + Cn^{1+\eta\chi}. \tag{3.7}$$

Here, η is the parameter from (3.2).

It does not matter here, whether we use the "α-construction" of Φ_n or the "β-construction".

For proving this theorem, we will start by giving two auxiliary lemmas.

3.7 Lemma: *For some constant $C > 0$, we have*

$$E_m\Big(\Phi_n\big(\mathbb{X}_i(0), \mathbb{W}_i\big)\Big) \leqslant Cn \exp\left(-\frac{n^{2\eta}}{C}\right).$$

The right-hand side is bounded by a positive constant, and, in particular, for $\overline{\Phi}_n := \Phi_n - E_m\Phi_n$ we have

$$\left\|\overline{\Phi}_n\right\|_\infty \leqslant C'n^\eta < \infty,$$

for some constant C'.

Proof. By definition, and because stopped martingales are centered, we have

$$\begin{aligned} E_m\Big(\Phi_n\big(\mathbb{X}_i(0), \mathbb{W}_i\big)\Big) &= E_m\left(\int_0^{T\wedge U_{n,i}} \varphi_{jk}(t)\sigma\big(\mathbb{X}_i(t)\big)\mathrm{d}\mathbb{W}_i(t) - \int_{T\wedge U_{n,i}}^T \varphi_{jk}(t)S(t)\mathrm{d}t\right) \\ &= 0 - E_m\int_{T\wedge U_{n,i}}^T \varphi_{jk}(t)S(t)\mathrm{d}t \\ &= -E_m(0\cdot \mathbb{1}_{\{U_{n,i}\geqslant T\}}) - E_m\left(\int_{U_{n,i}}^T \varphi_{jk}(t)S(t)\mathrm{d}t\mathbb{1}_{\{U_{n,i}<T\}}\right). \end{aligned}$$

By a similar argument to (3.3), we now find

$$\begin{aligned} E_m\big(\Phi_n\big(\mathbb{X}_i(0), \mathbb{W}_i\big)\big) &\leqslant 0 + \left|-E_m\left(\int_{U_{n,i}}^T \varphi_{jk}(t)S(t)\mathrm{d}t\mathbb{1}_{\{U_{n,i}<T\}}\right)\right| \\ &\leqslant \|S\|_\infty T^{\frac{1}{2}} \|\varphi_{jk}\|_{L^2} P_m(U_{n,i} < T). \end{aligned}$$

Because of Proposition 3.3 we can hence conclude

$$E_m\big(\Phi_n\big(\mathbb{X}_i(0), \mathbb{W}_i\big)\big) \leqslant CP_m(U) \leqslant Cn\exp\left(-\frac{n^{2\eta}}{C}\right).$$

Here, the right-hand side is bounded by some constant C because the exponential decay dominates the linear growth. With Proposition 3.2 we see

$$\left\|\overline{\Phi}_n\right\|_\infty = \|\Phi_n - E_m\Phi_n\|_\infty \leqslant Cn^\eta + C \leqslant C'n^\eta < \infty,$$ ▨

3.8 Lemma: *Let*

$$\Sigma^2(\overline{\Phi}_n) := m(A)E_A\Big[\sum_{i=1}^{\tau_A} \overline{\Phi}_n\big(\mathbb{X}_i(0), \mathbb{W}_i\big)\Big]^2.$$

Under assumptions (A) and (B), $\Sigma^2(\overline{\Phi}_n) \leqslant C$ for some constant $C > 0$.

Proof. Obviously, neither the invariant measure m nor the atom A depend on the number n of observations. Thus, $m(A)$ is a constant. By definition of $\overline{\Phi}_n$ and Φ_n,

$$\begin{aligned}
\Sigma^2(\overline{\Phi}_n) &= m(A)E_A\left(\sum_{i=1}^{\tau_A}\overline{\Phi}_n\big(\mathbb{X}_i(0),\mathbb{W}_i\big)\right)^2\\
&\leqslant CE_A\left(\sum_{i=1}^{\tau_A}\Phi_n\big(\mathbb{X}_i(0),\mathbb{W}_i\big)\right)^2 + CE_A\left(\sum_{i=1}^{\tau_A}E_m\Phi_n\right)^2\\
&\leqslant CE_A\left(\sum_{i=1}^{\tau_A}\left[\int_0^{T\wedge U_{n,i}}\varphi_{jk}(t)\sigma\big(\mathbb{X}_i(t)\big)\mathrm{d}\mathbb{W}_i(t) - \int_{T\wedge U_{n,i}}^{T}\varphi_{jk}(t)S(t)\mathrm{d}t\right]\right)^2\\
&\qquad + CE_A\left(\tau_A n\exp\left(-\frac{n^{2\eta}}{C}\right)\right)^2.
\end{aligned}$$

In the last term, we have made use of Lemma 3.7. Of course, $n\exp\left(-\frac{n^{2\eta}}{C}\right)\leqslant C$ for some constant C. Now, applying assumption (A), the function S is bounded in absolute value; by assumption (B), the chain $(\mathbb{X}_i)_{i\geqslant 1}$ is geometrically regular, so we have in particular $E_A\tau_A^2 = C < \infty$. This gives

$$\begin{aligned}
\Sigma^2(\overline{\Phi}_n) &\leqslant CE_A\left(\sum_{i=1}^{\tau_A}\int_0^{T\wedge U_{n,i}}\varphi_{jk}(t)\sigma\big(\mathbb{X}_i(t)\big)\mathrm{d}\mathbb{W}_i(t)\right)^2\\
&\qquad + CE_A\left(\sum_{i=1}^{\tau_A}\int_{T\wedge U_{n,i}}^{T}\varphi_{jk}(t)S(t)\mathrm{d}t\right)^2 + CE_A\tau_A^2\\
&\leqslant CE_A\left\langle\sum_{i=1}^{\tau_A}\int_0^{T\wedge U_{n,i}}\varphi_{jk}(t)\sigma\big(\mathbb{X}_i(t)\big)\mathrm{d}\mathbb{W}_i(t)\right\rangle\\
&\qquad + C\left\|S\right\|_\infty^2 E_A\left(\sum_{i=1}^{\tau_A}\int_{T\wedge U_{n,i}}^{T}\varphi_{jk}(t)\mathrm{d}t\right)^2 + C.
\end{aligned}$$

We now apply Jensen's inequality to the second-to-last term and find

$$\begin{aligned}
\Sigma^2(\overline{\Phi}_n) &\leqslant CE_A\left(\sum_{i=1}^{\tau_A}\int_0^{T\wedge U_{n,i}}\varphi_{jk}^2(t)\sigma^2\big(\mathbb{X}_i(t)\big)\mathrm{d}t\right.\\
&\qquad\left. + \sum_{\substack{i_1,i_2=1\\ i_1\neq i_2}}^{\tau_A}\int_0^{T\wedge U_{n,i_1}\wedge U_{n,i_2}}\varphi_{jk}^2(t)\sigma\big(\mathbb{X}_{i_1}(t)\big)\sigma\big(\mathbb{X}_{i_2}(t)\big)\mathrm{d}\left\langle\mathbb{W}_{i_1}(t),\mathbb{W}_{i_2}(t)\right\rangle\right)\\
&\qquad + CE_A\left(\sum_{i=1}^{\tau_A}\left(\int_{T\wedge U_{n,i}}^{T}\varphi_{jk}^2(t)\mathrm{d}t\right)^{\frac{1}{2}}\right)^2 + C\\
&\leqslant CE_A\tau_A\left\|\sigma^2\right\|_\infty\int_0^T\varphi_{jk}^2(t)\mathrm{d}t + 0 + CE_A\left(\sum_{i=1}^{\tau_A}\left(\int_0^T\varphi_{jk}^2(t)\mathrm{d}t\right)^{\frac{1}{2}}\right)^2 + C\\
&= C + CE_A\tau_A^2 + C = C.
\end{aligned}$$

Here, the mixed angle brackets can be dropped, because $\mathbb{W}_i$ and $\mathbb{W}_j$ are independent parts of a Brownian Motion.

Note, that we did not need the full strength of assumption (B). We only needed the return time τ_A to have second moments, not exponential moments. ▨

After these lemmas, we can return to the proof of the Rosenthal-type inequality Theorem 3.6.

Proof (of Theorem 3.6). We only carry out the proof for the "α-construction" of Φ_n. We will apply [3, Props. 8 and 13], which considers an aperiodic Harris recurrent Markov chain on some countably generated measurable space $(E, \mathcal{E})$, that is χ-regular and allows for Nummelin splitting. All these conditions are fulfilled by assumption (B). Besides, the cited article needs a bounded, real-valued, measurable function defined on $(E, \mathcal{E})$ that is centered with respect to the stationary probability of the Markov chain. These conditions (apart from the centering) are fulfilled as well, because we have Proposition 3.2 and because, by our construction of the functional Φ_n, we deal with the polish space $\mathbb{R} \times \mathcal{C}_0([0,T])$ whose canonical σ-field is countably generated.

Lemma 3.7 shows that there is no trouble about the centering of Φ_n.

The idea of the proof in [3] is to look at the life-cycles of the Markov chain at hand, i. e. to decompose the sum $\sum_{i=1}^{n} \Phi\big(\mathbb{X}_i(0), \mathbb{W}_i\big)$ into successive visits to the atom A. The case that there is no visit to A at all can be dealt with via the exponential Markov inequality. The behaviour of the chain up to the first visit to A and the last life-cycle which cannot be completed until time n are considered separately. The families of *complete* life-cycles are independent and identically distributed by definition of the notion of an atom. Thus, classical theorems on sums of i.i.d. random variables can be applied there.

Let us now have a closer look at the bound that [3, Props. 8 and 13] gives us. In order to make use of that bound, we need the assumption $r \geqslant 2$, yielding $\chi \geqslant 2$. Translated into our notation, we find

$$E_m \left| \sum_{i=1}^{n} \overline{\Phi}_n\big(\mathbb{X}_i(0), \mathbb{W}_i\big) \right|^{\chi} \leqslant C_1 \Sigma^{\chi}(\overline{\Phi}_n) n^{\frac{\chi}{2}} + C_2 E_A \tau_A^{\chi} \left\| \overline{\Phi}_n \right\|_{\infty}^{\chi} n, \tag{3.8}$$

where $C_1, C_2 > 0$ are explicitly known constants (not depending on n), and where

$$\Sigma^2(\overline{\Phi}_n) := m(A) E_A \Big[\sum_{i=1}^{\tau_A} \overline{\Phi}_n\big(\mathbb{X}_i(0), \mathbb{W}_i\big) \Big]^2.$$

Note, that we used the upper bound from [3, Prop. 8] that is designated with $(*)$ in that article.

Now, the factor $\Sigma^{\chi} = (\Sigma^2)^{\frac{\chi}{2}}$ is no problem to us, because we already established Lemma 3.8.

For proving Theorem 3.6, we can now put together Lemma 3.7, (3.8) and Lemma 3.8. Keeping in mind that both $E_A \tau_A^{\chi} < \infty$ (by geometric regularity and hence χ-regularity) and χ are constants not depending on n, we arrive at

$$\begin{aligned}
E_m \left|\sum_{i=1}^n \Phi_n\big(\mathbb{X}_i(0), \mathbb{W}_i\big)\right|^\chi &= E_m \left|\sum_{i=1}^n \Big[\overline{\Phi}_n\big(\mathbb{X}_i(0), \mathbb{W}_i\big) + E_m\big(\Phi_n\big(\mathbb{X}_i(0), \mathbb{W}_i\big)\big)\Big]\right|^\chi \\
&\leqslant C E_m \left|\sum_{i=1}^n \overline{\Phi}_n\big(\mathbb{X}_i(0), \mathbb{W}_i\big)\right|^\chi + C \left|\sum_{i=1}^n E_m\big(\Phi_n\big(\mathbb{X}_i(0), \mathbb{W}_i\big)\big)\right|^\chi \\
&\leqslant C\Sigma^\chi(\overline{\Phi}_n) n^{\frac{\chi}{2}} + C E_A \tau_A^\chi \left\|\overline{\Phi}_n\right\|_\infty^\chi n + C n^\chi \cdot n^\chi \exp\left(-\frac{\chi n^{2\eta}}{C}\right) \\
&= C n^{\frac{\chi}{2}} + C \left\|\overline{\Phi}_n\right\|_\infty^\chi n + C n^{2\chi} \exp\left(-\frac{\chi n^{2\eta}}{C}\right) \\
&\leqslant C n^{\frac{\chi}{2}} + C \left\|\overline{\Phi}_n\right\|_\infty^\chi n.
\end{aligned}$$

Finally, by the last statement of Lemma 3.7, we can give the upper bound

$$E_m \left|\sum_{i=1}^n \Phi_n\big(\mathbb{X}_i(0), \mathbb{W}_i\big)\right|^\chi \leqslant C n^{\frac{\chi}{2}} + C(C' n^\eta)^\chi n.$$

This is (3.7), just as claimed. ▨

3.3 A Bernstein-type inequality

Let us now adopt [3, Thm. 15] to our setting. The proof in [3] shows a little more than what is stated there: the statement involves some "y large enough", but in the proof, the reason for introducing this y is the use of a theorem by Fuk and Nagaev, [10, Thm. 4], on sums of independent random variables. In turn, Fuk and Nagaev only need some positive y and not a "large" y to give their bounds. Thus, one can see that $y > 0$ is all we need to assume. This slight improvement will be necessary for our application of the result in Theorem 3.9, since we will take $y = n^\zeta$ for some positive power of n, and we do not have to make sure that our y is large enough. The parameter ζ will be chosen in an appropriate way to guarantee the rate of convergence from Theorem 2.14; see Theorem 3.9 and (3.14) for details on this.

A brief sketch of the proof in [3] has been given in the previous section.

3.9 Theorem (Bernstein-type inequality): *Let m be the stationary probability of the Markov chain $(\mathbb{X}_i)_{i\geqslant 1}$. Then, for some parameter $0 < \zeta < 1 - \eta - \nu$, there is a constant $C > 0$, such that*

$$\begin{aligned}
&P_m\left(\frac{1}{n}\left|\sum_{i=1}^n \Phi_n\big(\mathbb{X}_i(0), \mathbb{W}_i\big)\right| \geqslant x\right) \qquad (3.9)\\
&\quad\leqslant C\left[n e^{-Kn^\zeta} + n\exp\left(-\frac{K}{C} n^{1-\eta} x\right) + \exp\left(-\frac{n x^2}{C(1+n^{\eta+\zeta}x)}\right) + n\exp\left(-\frac{n^{2\eta}}{C}\right)\right].
\end{aligned}$$

Here, K is the constant from the definition of geometric regularity in assumption (B). The parameter ζ depends on the threshold ν of the estimator TW_n *and on the "truncation-size" η in the stopping times $U_{n,i}$. It is arbitrary between the bounds given above, but fixed.*

3.10 Remark: By the choice of η in (3.2), one sees that indeed $\zeta > 0$. It will become clearer in (3.14) of the proof of Theorem 3.12, why these conditions on ζ have been given in this way. They are designed to yield suitable asymptotic bounds on the error of our estimator. $\diamondsuit$

PROOF (of Theorem 3.9). We are going to apply [3, Thm. 15]. There, one considers an aperiodic, Harris recurrent, geometrically regular Markov chain, allowing for Nummelin splitting, on some measurable space with countably generated σ-field. As we already saw in the proof of Theorem 3.6, these conditions are fulfilled in our situation by assumption (B). Again, one needs a bounded, real-valued, measurable function defined on the state space of the Markov chain, that is centered with respect to the stationary probability. By construction of Φ_n and by our auxiliary results, these conditions (apart from the centering) are satisfied as well.

For the centering, we need to argue a little differently now, than we did in the proof of Theorem 3.6. We first recall the definition of the event $U := \bigcup_{i=1}^n \{U_{n,i} < T\}$, that in at least one path segment there is a non-martingale term in the definition of Φ_n. So, if none of the stopping times $U_{n,i}$ apply before T, i. e. in the event $U^{\complement}$, there are only martingales involved and so, the functional is centered. But then, $E_m(\Phi_n \mathbb{1}_{U^{\complement}}) = 0$ and therefore $\overline{\Phi}_n = \Phi_n$ on $U^{\complement}$. Hence we have,

$$\begin{aligned}
P_m\left(\frac{1}{n}\left|\sum_{i=1}^n \Phi_n\big(\mathbb{X}_i(0), \mathbb{W}_i\big)\right| \geqslant x\right) &= P_m\left(\frac{1}{n}\left|\sum_{i=1}^n \Phi_n\big(\mathbb{X}_i(0), \mathbb{W}_i\big)\right| \geqslant x,\ U\right) \\
&\quad + P_m\left(\frac{1}{n}\left|\sum_{i=1}^n \Phi_n\big(\mathbb{X}_i(0), \mathbb{W}_i\big)\right| \geqslant x,\ U^{\complement}\right) \\
&\leqslant P_m(U) + P_m\left(\frac{1}{n}\left|\sum_{i=1}^n \overline{\Phi}_n\big(\mathbb{X}_i(0), \mathbb{W}_i\big)\right| \geqslant x,\ U^{\complement}\right) \\
&\leqslant P_m(U) + P_m\left(\frac{1}{n}\left|\sum_{i=1}^n \overline{\Phi}_n\big(\mathbb{X}_i(0), \mathbb{W}_i\big)\right| \geqslant x\right).
\end{aligned}$$

The first summand can be dealt with by Proposition 3.3, for the second summand, we need [3, Thm. 15]. This result reads in our present situation (where K is the constant appearing in the statement of geometric regularity from assumption (B))

$$\begin{aligned}
&P_m\left(\frac{1}{n}\left|\sum_{i=1}^n \overline{\Phi}_n\big(\mathbb{X}_i(0), \mathbb{W}_i\big)\right| \geqslant x\right) \\
&\qquad \leqslant C\left[ne^{-Ky} + n\exp\left(-K\frac{nx}{3\left\|\overline{\Phi}_n\right\|_\infty}\right) + \exp\left(-\frac{nx^2}{18(\Sigma^2(\overline{\Phi}) + \frac{2}{9}\left\|\overline{\Phi}_n\right\|_\infty yx)}\right)\right].
\end{aligned}$$

We take $y = y(n) = n^\zeta$ for some positive parameter ζ. Besides, correcting a typing error in [3], the constant in the denominator of the second term has changed from 2 to 3, but this change does not matter to us, because we are interested in the asymptotics for large n.

Altogether, we arrive at

$$P_m\left(\frac{1}{n}\left|\sum_{i=1}^{n}\Phi_n\big(\mathbb{X}_i(0),\mathbb{W}_i\big)\right|\geqslant x\right)\leqslant Cn\exp\left(-\frac{n^{2\eta}}{C}\right)$$
$$+C\left[ne^{-Kn^\zeta}+n\exp\left(-\frac{K}{3}\frac{nx}{\left\|\overline{\Phi}_n\right\|_\infty}\right)\right.$$
$$\left.+\exp\left(-\frac{nx^2}{18(\Sigma^2(\overline{\Phi}_n)+\frac{2}{9}\left\|\overline{\Phi}_n\right\|_\infty n^\zeta x)}\right)\right].$$

Here, $\Sigma^2(\overline{\Phi}_n)$ is the same object as in the proof of Theorem 3.6. We saw in Lemma 3.8 that it is bounded by some constant (the lemma of course applies to the "β-construction" of Φ_n as well as to the "α-construction"), and we can use the same argument here as well. In addition, since the functional $\overline{\Phi}_n$ is bounded by Cn^η by Lemma 3.7, we find

$$P_m\left(\frac{1}{n}\left|\sum_{i=1}^{n}\Phi_n\big(\mathbb{X}_i(0),\mathbb{W}_i\big)\right|\geqslant x\right)\leqslant Cn\exp\left(-\frac{n^{2\eta}}{C}\right)+C\left[ne^{-Kn^\zeta}+n\exp\left(-\frac{K}{C}\frac{nx}{n^\eta}\right)\right.$$
$$\left.+\exp\left(-\frac{nx^2}{18(C+\frac{2}{9}Cn^\eta n^\zeta x)}\right)\right]$$
$$\leqslant C\left[ne^{-Kn^\zeta}+n\exp\left(-\frac{K}{C}n^{1-\eta}x\right)\right.$$
$$\left.+\exp\left(-\frac{nx^2}{C(1+n^{\eta+\zeta}x)}\right)+n\exp\left(-\frac{n^{2\eta}}{C}\right)\right].$$

In the end, we arrive at (3.9), as intended. ▨

3.4 The main auxiliary theorems

3.11 Theorem: *Let $\chi \geqslant r > 2$ and let assumption (B) be satisfied. Then there is a constant $C > 0$, such that the estimators of the wavelet coefficients satisfy*

$$E_m\left|\hat{\alpha}_{jk,n}-\alpha_{jk}\right|^\chi\leqslant Cn^{-\frac{\chi}{2}}\qquad\textit{and}\qquad E_m\left|\hat{\beta}_{jk,n}-\beta_{jk}\right|^\chi\leqslant Cn^{-\frac{\chi}{2}}.$$

Proof. Again, we will only prove the first assertion on the α-coefficients. The second assertion works with only notational differences. As before, the notion C means some positive, real constant (not depending on n), whose value may be different every time it appears.

First, by Proposition 3.4,

$$\begin{aligned}
E_m \left|\hat{\alpha}_{jk,n} - \alpha_{jk}\right|^\chi &= E_m \left|\hat{\alpha}_{jk,n} - \check{\alpha}_{jk,n} + \check{\alpha}_{jk,n} - \alpha_{jk}\right|^\chi \\
&\leqslant CE_m \left|\hat{\alpha}_{jk,n} - \check{\alpha}_{jk,n}\right|^\chi + CE_m \left|\check{\alpha}_{jk,n} - \alpha_{jk}\right|^\chi \\
&= CE_m \left|\frac{1}{n}\sum_{i=1}^n \int_{T\wedge U_{n,i}}^T \varphi_{jk}(t)\Big(S(t)\mathrm{d}t + \sigma\big(\mathbb{X}_i(t)\big)\mathrm{d}\mathbb{W}_i(t)\Big)\right|^\chi \\
&\quad + CE_m \left|\frac{1}{n}\sum_{i=1}^n \Phi_n\big(\mathbb{X}_i(0), \mathbb{W}_i\big)\right|^\chi .
\end{aligned}$$

With the classical inequality $|a_1 + \cdots + a_n|^p \leqslant n^p(|a_1|^p + \cdots + |a_n|^p)$ that holds for all real numbers $a_1, \ldots, a_n$, we see here

$$\begin{aligned}
E_m \left|\hat{\alpha}_{jk,n} - \alpha_{jk}\right|^\chi &\leqslant Cn^{-\chi}n^\chi \sum_{i=1}^n E_m \left|\mathbb{1}_{\{U_{n,i}<T\}} \int_{U_{n,i}}^T \varphi_{jk}(t)S(t)\mathrm{d}t\right|^\chi \\
&\quad + Cn^{-\chi}n^\chi \sum_{i=1}^n E_m \left|\mathbb{1}_{\{U_{n,i}<T\}} \int_{U_{n,i}}^T \varphi_{jk}(t)\sigma\big(\mathbb{X}_i(t)\big)\mathrm{d}\mathbb{W}_i(t)\right|^\chi \\
&\quad + CE_m(0\cdot \mathbb{1}_{\{U_{n,i}\geqslant T\ \forall i\}}) + CE_m \left|\frac{1}{n}\sum_{i=1}^n \Phi_n\big(\mathbb{X}_i(0), \mathbb{W}_i\big)\right|^\chi .
\end{aligned}$$

Now, since S and σ are bounded by assumption (A), and using the Cauchy-Schwarz inequality as well as Theorem 3.6,

$$\begin{aligned}
E_m \left|\hat{\alpha}_{jk,n} - \alpha_{jk}\right|^\chi &\leqslant C\sum_{i=1}^n \|S\|_\infty^\chi E_m \left|\mathbb{1}_{\{U_{n,i}<T\}} \int_{U_{n,i}}^T \varphi_{jk}(t)\mathrm{d}t\right|^\chi \\
&\qquad + C\sum_{i=1}^n E_m \left|\mathbb{1}_{\{U_{n,i}<T\}} \int_0^T \varphi_{jk}^2(t)\sigma^2\big(\mathbb{X}_i(t)\big)\mathrm{d}t\right|^{\frac{\chi}{2}} \\
&\qquad + Cn^{-\chi}\left(Cn^{\frac{\chi}{2}} + Cn^{1+\eta\chi}\right) \\
&\leqslant C\sum_{i=1}^n \left|\int_0^T \varphi_{jk}^2(t)\mathrm{d}t\right|^{\frac{\chi}{2}} P_m(U) + C\sum_{i=1}^n \left|\int_0^T \varphi_{jk}^2(t)\mathrm{d}t\right|^{\frac{\chi}{2}} P_m(U) \\
&\qquad + Cn^{-\frac{\chi}{2}} + Cn^{-\chi+1+\eta\chi} \\
&= 2CnP_m(U) + Cn^{-\frac{\chi}{2}} + Cn^{-\chi+1+\eta\chi}.
\end{aligned}$$

Now, we recall the condition $\eta \leqslant \frac{r-2}{2r}$ from (3.2) (keep in mind that we have $r \gneqq 2$ by assumption). Since the mapping $x \mapsto \frac{x-2}{2x} = \frac{1}{2} - \frac{1}{x}$ is monotonically increasing in $x > 2$, we immediately get $\eta \leqslant \frac{\chi-2}{2\chi}$. Applying Proposition 3.3, we see

$$E_m \left|\hat{\alpha}_{jk,n} - \alpha_{jk}\right|^\chi \leqslant 2Cn^2 \exp\left(-\frac{n^{2\eta}}{C}\right) + Cn^{-\frac{\chi}{2}} + Cn^{-\chi+1+\eta\chi} \leqslant Cn^{-\frac{\chi}{2}}.$$

The last step explains some of the choices for the parameter η that we made in (3.2). The rest of the choices will become clear in the proof of the next theorem.

Theorem 3.11 is proved. ▨

3.12 Theorem: *Let $\nu \in (0, \frac{1}{2})$. For every n sufficiently large and any $\gamma \geqslant 1$, there is a constant $C > 0$, such that for all $j \in \{j_0(n), \dots, j_1(n)\}$ we have*

$$P_m\left(\left|\hat{\beta}_{jk,n} - \beta_{jk}\right| > \frac{1}{2}\left(\frac{j}{n}\right)^{\nu}\right) \leqslant C 2^{-\gamma j}. \tag{3.10}$$

Proof. Here, we will apply Theorem 3.9. From similar arguments to the ones above we get

$$\begin{aligned}
P_m\left(\left|\hat{\beta}_{jk,n} - \beta_{jk}\right| > \frac{1}{2}\left(\frac{j}{n}\right)^{\nu}\right) &= P_m\left(\left|\hat{\beta}_{jk,n} - \check{\beta}_{jk,n} + \check{\beta}_{jk,n} - \beta_{jk}\right| > \frac{1}{2}\left(\frac{j}{n}\right)^{\nu}\right) \\
&\leqslant P_m\left(\left|\hat{\beta}_{jk,n} - \check{\beta}_{jk,n}\right| + \left|\check{\beta}_{jk,n} - \beta_{jk}\right| > \frac{1}{2}\left(\frac{j}{n}\right)^{\nu}\right) \\
&= P_m\left(\left|\hat{\beta}_{jk,n} - \check{\beta}_{jk,n}\right| + \left|\check{\beta}_{jk,n} - \beta_{jk}\right| > \frac{1}{2}\left(\frac{j}{n}\right)^{\nu},\ U\right) \\
&\quad + P_m\left(\left|\hat{\beta}_{jk,n} - \check{\beta}_{jk,n}\right| + \left|\check{\beta}_{jk,n} - \beta_{jk}\right| > \frac{1}{2}\left(\frac{j}{n}\right)^{\nu},\ U^{\complement}\right).
\end{aligned}$$

Now, because by definition of the event U, we have $\hat{\beta}_{jk,n} = \check{\beta}_{jk,n}$ on $U^{\complement}$:

$$P_m\left(\left|\hat{\beta}_{jk,n} - \beta_{jk}\right| > \frac{1}{2}\left(\frac{j}{n}\right)^{\nu}\right) \leqslant P_m(U) + P_m\left(0 + \left|\check{\beta}_{jk,n} - \beta_{jk}\right| > \frac{1}{2}\left(\frac{j}{n}\right)^{\nu},\quad U^{\complement}\right).$$

Proposition 3.3 and (3.5) yield

$$\begin{aligned}
P_m\left(\left|\hat{\beta}_{jk,n} - \beta_{jk}\right| > \frac{1}{2}\left(\frac{j}{n}\right)^{\nu}\right) &\leqslant Cn \exp\left(-\frac{n^{2\eta}}{C}\right) + P_m\left(\left|\check{\beta}_{jk,n} - \beta_{jk}\right| > \frac{1}{2}\left(\frac{j}{n}\right)^{\nu}\right) \\
&\leqslant Cn \exp\left(-\frac{n^{2\eta}}{C}\right) \\
&\qquad + P_m\left(\frac{1}{n}\left|\sum_{i=1}^{n} \Phi_n\left(\mathbb{X}_i(0), \mathbb{W}_i\right)\right| > \frac{1}{2}\left(\frac{j}{n}\right)^{\nu}\right).
\end{aligned}$$

With Theorem 3.9, we get to

$$\begin{aligned}
P_m\left(\left|\hat{\beta}_{jk,n} - \beta_{jk}\right| > \frac{1}{2}\left(\frac{j}{n}\right)^{\nu}\right) &\leqslant Cn \exp\left(-\frac{n^{2\eta}}{C}\right) + C\left[ne^{-Kn^{\zeta}} + n\exp\left(-\frac{K}{C}n^{1-\eta}\frac{j^{\nu}}{2n^{\nu}}\right)\right. \\
&\qquad \left. + \exp\left(-\frac{n\frac{1}{4}j^{2\nu}n^{-2\nu}}{C(1 + n^{\eta+\zeta}\frac{1}{2}j^{\nu}n^{-\nu})}\right) + n\exp\left(-\frac{n^{2\eta}}{C}\right)\right] \\
&\leqslant C\left[n\exp(-Kn^{\zeta}) + n\exp\left(-\frac{K}{C}j^{\nu}n^{1-\nu-\eta}\right)\right. \\
&\qquad + \exp\left(-\frac{j^{2\nu}}{Cn^{2\nu-1} + Cj^{\nu}n^{\eta+\zeta+\nu-1}}\right) \\
&\qquad \left. + n\exp\left(-\frac{n^{2\eta}}{C}\right)\right].
\end{aligned} \tag{3.11}$$

Of course, we want this bound to vanish for $n \to \infty$. This is possible, if we choose:

- $\zeta > 0$, so we can deal with the first summand; (3.12)
- $\nu + \eta < 1$ for the second summand; (3.13)
- $\nu < \dfrac{1}{2}$ and $\eta + \zeta + \nu < 1$ for the third term; and besides this, we take (3.14)
- $\eta > 0$ for the last term. (3.15)

Here, we remark that the powers of j are not very important for these asymptotics, since j depends on n by (2.14), but only about logarithmically. We have already seen in Theorem 3.11 that we need to choose $0 < \eta \leqslant \frac{r-2}{2r} \in (0, \frac{1}{2})$, and so, the condition from the second summand can already be satisfied by $0 < \nu < \frac{1}{2}$. In the end, we still have to optimize the upper bound of the error of estimation, by taking ν, η and ζ "as large as possible" within these restrictions. Proving Theorem 2.14, one chooses $r > 2$ first, then some threshold $0 < \nu < \frac{1}{2}$ in (2.11). After this, one takes an appropriate $\eta < \frac{r-2}{2r}$ (in (3.2)). In the end, one chooses ζ in Theorem 3.9 according to the restrictions in (3.12) and (3.14). No matter what η and ζ we choose within these restrictions, the asymptotic upper bound (2.15) will stay the same. The constant factor C will improve a little by optimizing the choice of ν, η and ζ.

Now, since $j \in \{j_0(n), \dots, j_1(n)\}$ and both $j_0(n)$ and $j_1(n)$ are roughly logarithmic in n by (2.14), the right-hand side of (3.10) can be bounded from below by

$$C2^{-\gamma j} \geqslant C2^{-\gamma j_1} \geqslant C2^{-\gamma C \log_2 (\frac{n}{\log n})^{\frac{\alpha}{s'}}} \geqslant Cn^{-\gamma C \frac{\alpha}{s'}}.$$

But we have made up all our Greek letters such that the right hand side of (3.11) is decreasing exponentially fast in n, and hence faster than any polynomial rate; so we see

$$P_m \left(\left| \hat{\beta}_{jk,n} - \beta_{jk} \right| > \frac{1}{2} \left(\frac{j}{n} \right)^{\nu} \right) \leqslant Cn^{-\gamma C \frac{\alpha}{s'}} \leqslant C2^{-\gamma j}.$$

We have thus proved the theorem. ▨

3.5 Some preparatory lemmas

3.5.1 Elementary Computations on α, ε und s'

In this subsection 3.5.1, we will always assume that the conditions from Theorem 2.14 are satisfied. Occasionally, for reasons of clearness, we will state those conditions explicitly which are necessary in each particular lemma. We are going to work with the notations of Definition 2.11.

3.13 Lemma: *We have*

$$\alpha = \begin{cases} \dfrac{s}{1+2s}, & \text{for } \varepsilon \geqslant 0, \\ \dfrac{s'}{1+2s-\frac{2}{p}}, & \text{for } \varepsilon \leqslant 0. \end{cases} \tag{3.16}$$

Besides, $\alpha \geqslant 0$ holds.

Proof. Because of $1+2s > 0$ and $p+2sp-2 > 1+2-2 = 1 > 0$ we find

$$\begin{aligned} \alpha = \frac{s}{1+2s} \quad &\Longleftrightarrow \quad \frac{s}{1+2s} \leqslant \frac{s-\frac{1}{p}+\frac{1}{r}}{1+2s-\frac{2}{p}} \\ &\Longleftrightarrow \quad \frac{s}{1+2s} \leqslant \frac{sp-1+\frac{p}{r}}{p+2sp-2} \\ &\Longleftrightarrow \quad sp+2s^2p-2s \leqslant sp-1+\frac{p}{r}+2s^2p-2s+\frac{2sp}{r} \\ &\Longleftrightarrow \quad 0 \leqslant p+2sp-r \\ &\Longleftrightarrow \quad p+2sp \geqslant r \\ &\Longleftrightarrow \quad sp \geqslant \frac{r-p}{2} \\ &\Longleftrightarrow \quad \varepsilon \geqslant 0. \end{aligned}$$

Therefore, in case $\varepsilon \leqslant 0$, this equivalence shows $\frac{s}{1+2s} \geqslant \frac{s-\frac{1}{p}+\frac{1}{r}}{1+2s-\frac{2}{p}}$.

In particular, for $\varepsilon = 0$, we get $\frac{s}{1+2s} = \frac{s'}{1+2s-\frac{2}{p}}$.

Now, we show $\alpha \geqslant 0$ under the assumptions of Theorem 2.14. Since $s > 0$ implies $\frac{s}{1+2s} > 0$, we only look at the case $\varepsilon < 0$. There, we get

$$\alpha = \frac{s'}{1+2s-\frac{2}{p}} = \frac{s'p}{p+2sp-2}.$$

Because of $s > \frac{1}{p}$, both numerator and denominator are positive. ▨

3.14 Lemma: *We have*

$$\alpha r = \begin{cases} \dfrac{r-p}{2} + \varepsilon(1-2\alpha), & \text{for } \varepsilon \geqslant 0, \\ \dfrac{r-p}{2} + \dfrac{\varepsilon\alpha}{s'}, & \text{for } \varepsilon \leqslant 0. \end{cases} \tag{3.17}$$

Proof. Let $\varepsilon \geqslant 0$. Then, by (3.16), $\alpha = \frac{s}{1+2s}$ and so

$$\begin{aligned} \frac{r-p}{2} + \varepsilon(1-2\alpha) &= \frac{r-p}{2} + \left(sp - \frac{r-p}{2}\right) - 2\frac{s}{1+2s}\left(sp - \frac{r-p}{2}\right) \\ &= sp - \frac{2s^2p}{1+2s} + \frac{s(r-p)}{1+2s} \\ &= \frac{sp(1+2s) - 2s^2p + sr - sp}{1+2s} = \frac{sr}{1+2s} = \alpha r. \end{aligned}$$

Now, consider the case $\varepsilon \leqslant 0$. Then, we have $\alpha = \frac{s'}{1+2s-\frac{2}{p}}$ and thus

$$\begin{aligned}\frac{r-p}{2} + \frac{\varepsilon\alpha}{s'} &= \frac{r-p}{2} + \left(sp - \frac{r-p}{2}\right)\frac{s'}{1+2s-\frac{2}{p}}\frac{1}{s'} \\ &= \frac{r-p}{2} + \frac{sp}{1+2s-\frac{2}{p}} - \frac{r-p}{2(1+2s-\frac{2}{p})} \\ &= \frac{r+2sr-\frac{2r}{p}-p-2sp+2+2sp-r+p}{2(1+2s-\frac{2}{p})} \\ &= \frac{sr-\frac{r}{p}+1}{1+2s-\frac{2}{p}} = \frac{r(s-\frac{1}{p}+\frac{1}{r})}{1+2s-\frac{2}{p}} = r\alpha.\end{aligned}$$

▨

3.15 Lemma: *If $\varepsilon > 0$, then $\alpha - \frac{\alpha\varepsilon}{\sigma p} = \frac{r}{2p} - \frac{1}{2} - \frac{\alpha r}{p} + \alpha$ holds.*
In particular, for $\frac{\alpha}{\sigma p} \geqslant 0$, we have $\alpha \geqslant \frac{r}{2p} - \frac{1}{2} - \frac{\alpha r}{p} + \alpha$.

Proof. From (3.17) and (3.16) one concludes:

$$\begin{aligned}\alpha - \frac{\alpha\varepsilon}{sp} = \frac{r}{2p} - \frac{1}{2} - \frac{\alpha r}{p} + \alpha \iff\quad & -2\alpha\varepsilon = sr - sp - 2s\alpha r \\ \iff\quad & -2\alpha\varepsilon = sr - sp - 2s\left(\frac{r-p}{2} + \varepsilon - 2\alpha\varepsilon\right) \\ \iff\quad & -2\alpha\varepsilon = -2s\varepsilon + 4\alpha s\varepsilon \\ \iff\quad & -\alpha\varepsilon = s\varepsilon(2\alpha - 1) \\ \iff\quad & -\frac{s\varepsilon}{1+2s} = s\varepsilon\left(\frac{2s}{2s+1} - 1\right) = s\varepsilon\frac{-1}{2s+1}.\end{aligned}$$

The last line is obviously true. ▨

3.16 Lemma: *Let $\varepsilon > 0$. Then, we have $-\frac{r-p}{p}\varepsilon(1-2\alpha) \leqslant \alpha r$.*

Proof. We start by observing

$$1 - 2\alpha = 1 - \frac{2s}{1+2s} = \frac{1}{1+2s} = \frac{\alpha}{s} \qquad \text{for any } \varepsilon \geqslant 0, \tag{3.18}$$

which implies $1 - 2\alpha \geqslant 0$. In particular, $r \geqslant p \geqslant 0$ implies

$$\frac{r-p}{2} + \varepsilon(1-2\alpha)\frac{r}{p} = \frac{r-p}{2} + \varepsilon(1-2\alpha)\left(1 + \frac{r-p}{p}\right) \geqslant 0.$$

Besides,

$$\frac{r-p}{2} + \varepsilon(1-2\alpha)\left(1 + \frac{r-p}{p}\right) \geqslant 0 \iff -\frac{r-p}{p}\varepsilon(1-2\alpha) \leqslant \frac{r-p}{2} + \varepsilon(1-2\alpha),$$

which is, by (3.17), equivalent to what we wanted to show. ▨

Finally, we will need a result on the number of j-levels used in TW_n:

3.17 Lemma: *Let $\varepsilon = 0$. By the choice of j_0 and j_1 in (2.14) we have, as $n \to \infty$,*

$$\frac{j_1(n)}{j_0(n)} \asymp \frac{r}{r-2}.$$

Proof. First,

$$j_0 \asymp (1-2\alpha)\left(\log n + \frac{r-p}{p}\log\log n\right) \qquad j_1 \asymp \frac{\alpha}{s'}\big(\log n - \log\log n\big).$$

Therefore, we conclude

$$\frac{j_1}{j_0} \asymp \frac{\alpha}{s'(1-2\alpha)} \cdot \frac{\log n - \log\log n}{\log n + \frac{r-p}{p}\log\log n}.$$

The second factor converges to 1 by l'Hospital's rule. For the first one we consider

$$\begin{aligned}
\varepsilon = 0 \quad &\Longleftrightarrow \quad \frac{r-p}{2} = sp\\
&\Longleftrightarrow \quad r-p = 2sp\\
&\Longleftrightarrow \quad \frac{r}{sp} - \frac{1}{s} = 2\\
&\Longleftrightarrow \quad r-2 = r - \frac{r}{sp} + \frac{1}{s}\\
&\Longleftrightarrow \quad \alpha(r-2) = r\left(s - \frac{1}{p} + \frac{1}{r}\right)\frac{\alpha}{s}.
\end{aligned}$$

By (3.18) and by the definition of s' this is equivalent to

$$\alpha(r-2) = rs'(1-2\alpha)$$

and to

$$\frac{\alpha}{s'(1-2\alpha)} = \frac{r}{r-2}.$$

This is what we wanted. ◪

3.5.2 Norm inequalities

In this subsection, let $r > 2$ designate the L^r-norm from Theorem 2.14. Let $\hat{f} := \sum_{j=j_0}^{j_1}\sum_{k\in\mathbb{Z}} \hat{f}_{jk}\psi_{jk}$ be a random function (i. e. a function whose wavelet coefficients depend on the observations and are thus $\sigma(\mathbb{X}_1,\dots,\mathbb{X}_n)$-measurable). In later parts of the proof of Theorem 2.14, these random wavelet coefficients will be sums and differences of the estimators $\tilde{\beta}_{jk}$ and the true values β_{jk}.

We remind the reader of the projections E_i and D_j that we have introduced in (2.6) and (2.7). Note, that E_i denotes a projection, while E_m is the mathematical expectation with respect to the invariant measure m.

3.18 Lemma: *Let $r > 2$. Then, we have*

$$\|\hat{f}\|_{L^r}^r \leqslant \left(\sum_{j=j_0}^{j_1} \|D_j \hat{f}\|_{L^r}^2\right)^{\frac{r}{2}}. \tag{3.19}$$

Proof. The right-hand side is the r-th power of a summand in the Besov norm $J_{0r2}(\hat{f})$. If this is finite, the Besov embeddings $B_{0r2} \subset L^r$ show that the left-hand side is finite as well (see Theorem 2.7). By the definition of the projection D_j and using Jensen's inequality, we find

$$\|\hat{f}\|_{L^r} = \left\|\sum_{j=j_0}^{j_1}\sum_{k\in\mathbb{Z}} \hat{f}_{jk}\psi_{jk}\right\|_{L^r} \leqslant \sum_{j=j_0}^{j_1} \|D_j \hat{f}\|_{L^r} \leqslant \left(\sum_{j=j_0}^{j_1} \|D_j \hat{f}\|_{L^r}^2\right)^{\frac{1}{2}}.$$

Taking the r-th power of each side concludes the proof. ▨

3.19 Lemma: *There is a constant $C > 0$, such that we have for $r > 2$*

$$\|D_j \hat{f}\|_{L^r}^r \leqslant C 2^{j\left(\frac{r}{2}-1\right)} \sum_{k\in\mathbb{Z}} |\hat{f}_{jk}|^r. \tag{3.20}$$

Proof. We apply Theorem 2.6 to $D_j \hat{f}$ with $\lambda_k = \hat{f}_{jk}$, $p = r$, and find

$$\|D_j \hat{f}\|_{L^r}^r = \left\|\sum_{k\in\mathbb{Z}} \hat{f}_{jk}\psi_{jk}\right\|_{L^r}^r \leqslant \left(c_2 2^{j\left(\frac{1}{2}-\frac{1}{r}\right)}\right)^r \|\hat{f}_{j\bullet}\|_{\ell^r}^r = C 2^{j\left(\frac{r}{2}-1\right)} \sum_{k\in\mathbb{Z}} |\hat{f}_{jk}|^r.$$

▨

Let us introduce a short-hand notation for $x \in \mathbb{R}$ and the r from the assertion of Theorem 2.14:

$$h(x) := \sum_{j=j_0}^{j_1} 2^{jx} \qquad \text{and} \qquad \varrho := \frac{r}{r-2}. \tag{3.21}$$

Note that we took $r \gneqq 2$ in Theorem 2.14.

The following two statements will play an important role in several steps of the proof of Theorem 2.14.

3.20 Lemma: *Let $j_0 \leqslant j_1$. Then, there is a constant $C > 0$ such that, for any real x not depending on n or j,*

$$h(x) \leqslant \begin{cases} C 2^{\max\{j_0 x, j_1 x\}}, & \text{for } x \neq 0, \\ j_1 - j_0, & \text{for } x = 0. \end{cases} \tag{3.22}$$

We will assume in the proof of Theorem 2.14 that $j_0 \leqslant j_1$ w.l.o.g., because otherwise the corresponding sum in the estimator TW_n will be vacuous and there will be no thresholding part to consider. So, whenever we cannot apply the present lemma, there will be no need for it anyway. The same remark holds for Lemma 3.21 below as well.

Proof. The case $x = 0$ is obvious (we even have equality in that case). So, let $x \neq 0$. The geometric series yields in case $xj_1 \leqslant xj_0$:

$$\sum_{j=j_0}^{j_1} 2^{jx} = \frac{2^{xj_0} - 2^{x(j_1+1)}}{1-2^x} \leqslant \frac{2^{xj_0}}{1-2^x} = C2^{xj_0}.$$

For $xj_1 > xj_0$ we have in particular $2^{x(j_0-j_1)} < 1$ and $x > 0$ (since we assumed $j_1 \geqslant j_0$), and so

$$\sum_{j=j_0}^{j_1} 2^{jx} = \frac{1}{1-2^x} 2^{xj_1} \left(2^{x(j_0-j_1)} - 2^x\right) = \frac{2^x - 2^{x(j_0-j_1)}}{2^x-1} 2^{xj_1} \leqslant \frac{2^x}{2^x-1} 2^{xj_1} = C2^{xj_1}.$$

□

3.21 Lemma: *Let $j_0 \leqslant j_1$, $r > 2$, and ϱ as in* (3.21)*. Then, there is a constant $C > 0$ such that*

$$h(\kappa\varrho)^{\frac{r}{2}-1} h(b) \leqslant \begin{cases} C2^{j_0\left(\kappa\frac{r}{2}+b\right)}, & \textit{for } \kappa < 0 \textit{ and } b < 0, \\ C2^{j_1\left(\kappa\frac{r}{2}+b\right)}, & \textit{for } \kappa > 0 \textit{ and } b > 0. \end{cases} \tag{3.23}$$

Proof. Because of (3.22) and since $\varrho > 0$, we get

$$\begin{aligned} h(\kappa\varrho)^{\frac{r}{2}-1} h(b) &\leqslant C\left(2^{\max\{j_0\kappa\varrho, j_1\kappa\varrho\}}\right)^{\frac{r}{2}-1} 2^{\max\{j_0 b, j_1 b\}} \\ &= C2^{\max\left\{j_0\kappa\frac{r}{r-2}\left(\frac{r-2}{2}\right), j_1\kappa\frac{r}{r-2}\left(\frac{r-2}{2}\right)\right\}+\max\{j_0 b, j_1 b\}} \\ &= C2^{\max\{j_0\kappa\frac{r}{2}, j_1\kappa\frac{r}{2}\}+\max\{j_0 b, j_1 b\}}. \end{aligned}$$

With the assumption $j_0 \leqslant j_1$ we can conclude the proof. □

3.22 Lemma: *Let $r > 2$ and $j_0 \leqslant j_1$. There is a constant $C > 0$ such that for any $\kappa \in \mathbb{R}$ and for ϱ from* (3.21)

$$E_m\|\hat{f}\|_{L^r}^r \leqslant Ch(\kappa\varrho)^{\frac{r}{2}-1} \sum_{j=j_0}^{j_1} 2^{j\left(\frac{r}{2}-1-\frac{\kappa r}{2}\right)} \sum_{k\in\mathbb{Z}} E_m|\hat{f}_{jk}|^r, \tag{3.24}$$

holds.

Proof. We apply Hölder's inequality to the conjugated exponents $\frac{r}{2}$ and $\frac{r}{r-2}$. For any $\kappa \in \mathbb{R}$ we find

$$\begin{aligned} \sum_{j=j_0}^{j_1} \|D_j\hat{f}\|_{L^r}^2 &= \sum_{j=j_0}^{j_1} \left|2^{j\kappa}2^{-j\kappa}\|D_j\hat{f}\|_{L^r}^2\right| \\ &\leqslant \left(\sum_{j=j_0}^{j_1} 2^{j\kappa\frac{r}{r-2}}\right)^{\frac{r-2}{r}} \left(\sum_{j=j_0}^{j_1} \left(2^{-j\kappa}\|D_j\hat{f}\|_{L^r}^2\right)^{\frac{r}{2}}\right)^{\frac{2}{r}} \\ &= \left(\sum_{j=j_0}^{j_1} 2^{j\kappa\frac{r}{r-2}}\right)^{\frac{r-2}{r}} \left(\sum_{j=j_0}^{j_1} 2^{-j\kappa\frac{r}{2}}\|D_j\hat{f}\|_{L^r}^r\right)^{\frac{2}{r}}. \end{aligned}$$

With this, we see

$$E_m\left(\sum_{j=j_0}^{j_1}\|D_j\hat{f}\|_{L^r}^2\right)^{\frac{r}{2}} \leqslant E_m\left(\left(\sum_{j=j_0}^{j_1}2^{j\kappa\frac{r}{r-2}}\right)^{\frac{r-2}{r}}\left(\sum_{j=j_0}^{j_1}2^{-j\kappa\frac{r}{2}}\|D_j\hat{f}\|_{L^r}^r\right)^{\frac{2}{r}}\right)^{\frac{r}{2}}$$
$$= \left(\sum_{j=j_0}^{j_1}2^{j\kappa\frac{r}{r-2}}\right)^{\frac{r-2}{r}\cdot\frac{r}{2}} E_m\left(\sum_{j=j_0}^{j_1}2^{-j\kappa\frac{r}{2}}\|D_j\hat{f}\|_{L^r}^r\right)^{\frac{2}{r}\cdot\frac{r}{2}}$$

With the definition of ϱ and h in (3.21), and with Lemma 3.19, we arrive at

$$E_m\left(\sum_{j=j_0}^{j_1}\|D_j\hat{f}\|_{L^r}^2\right)^{\frac{r}{2}} = \left(\sum_{j=j_0}^{j_1}2^{j\kappa\varrho}\right)^{\frac{r}{2}-1}\sum_{j=j_0}^{j_1}2^{-j\kappa\frac{r}{2}}E_m\|D_j\hat{f}\|_{L^r}^r$$
$$\leqslant Ch(\kappa\varrho)^{\frac{r}{2}-1}\sum_{j=j_0}^{j_1}2^{j(\frac{r}{2}-1-\frac{\kappa r}{2})}\sum_{k\in\mathbb{Z}}E_m|\hat{f}_{jk}|^r.$$

We can finish the proof with (3.19). ⧄

3.6 The proof of Theorem 2.14

To prove (2.15), we wish to give an upper bound to $E_m\|\mathrm{TW}_n - S\|_{L^r}^r$. In order to do so, we start by breaking the term $\mathrm{TW}_n - S$ into pieces which we will handle separately.

$$E_m\|\mathrm{TW}_n - S\|_{L^r}^r \leqslant 3^{r-1}\left(E_m\left\|\sum_{k\in\mathbb{Z}}\hat{\alpha}_{j_0k}\varphi_{j_0k} - E_{j_0}S\right\|_{L^r}^r\right.$$
$$\left. + E_m\left\|\sum_{j=j_0}^{j_1}\sum_{k\in\mathbb{Z}}\tilde{\beta}_{jk}\psi_{jk} - D_{j_0j_1}S\right\|_{L^r}^r + \|S - E_{j_1+1}S\|_{L^r}^r\right).$$

Remember that in (2.6) and (2.8), we have defined the projections $E_jf(x) := \sum_{k\in\mathbb{Z}}\alpha_{jk}\varphi_{jk}(x)$ and $D_{j_0j_1}f(x) := \sum_{j=j_0}^{j_1}\sum_{k\in\mathbb{Z}}\beta_{jk}\psi_{jk}(x)$ on the different levels of the wavelet expansion.

The first summand of the right-hand side of (3.6) is made up of the so-called Linear Part of TW_n, that is the part without thresholding. The second summand gives most of the trouble finding an upper bound (but we prepared the setup in such a way that it can work in a similar fashion to [6]), it describes the error of the so-called non-linear part of TW_n, where we take advantage of the thresholding methods. The third summand is a bias term that comes from cutting away the higher levels of detail in the wavelet expansion of S. The constant 3^{r-1} will be of no importance to us (neither are all other constants).

3.6.1 The Linear Part

3.23 Lemma: *(i) We have*

$$E_m \left\| \sum_k \hat{\alpha}_{j_0 k,n} \varphi_{j_0 k} - E_{j_0} S \right\|_{L^r}^r \leqslant C n^{-\frac{r}{2}} 2^{j_0 \frac{r}{2}} \tag{3.25}$$

for some constant $C > 0$.

(ii) By the choice of j_0 in (2.14) this upper bound attains the rate of convergence in (2.15), if $\varepsilon = 0$ and $\frac{r}{2p} \leqslant \frac{1}{q}$. In the other cases we can neglect the right-hand side of (3.25) compared to (2.15).

The main ingredients are Theorem 3.11 (whose assertion we translated from the setting in [6] to our present setting, and which can be employed without problem; it causes, however, the assumption $r > 2$ in Theorem 2.14) and the equivalence of the Besov norm and the classical ℓ^r-norm, described in Theorem 2.6.

Proof. (i) We start by observing

$$\begin{aligned} E_m \left\| \sum_k \hat{\alpha}_{j_0 k,n} \varphi_{j_0 k} - E_{j_0} S \right\|_{L^r}^r &= E_m \left\| \sum_k \hat{\alpha}_{j_0 k,n} \varphi_{j_0 k} - \sum_k \alpha_{j_0 k} \varphi_{j_0 k} \right\|_{L^r}^r \\ &= E_m \left\| \sum_k (\hat{\alpha}_{j_0 k,n} - \alpha_{j_0 k}) \varphi_{j_0 k} \right\|_{L^r}^r \\ &\leqslant C 2^{j_0 (\frac{1}{2} - \frac{1}{r}) r} E_m \left\| \hat{\alpha}_{j_0 \bullet, n} - \alpha_{j_0 \bullet} \right\|_{\ell^r}^r , \end{aligned}$$

which followed from the equivalence of the norms from Theorem 2.6. Now,

$$\begin{aligned} E_m \left\| \sum_k \hat{\alpha}_{j_0 k,n} \varphi_{j_0 k} - E_{j_0} S \right\|_{L^r}^r &= C 2^{j_0 (\frac{r}{2} - 1)} \sum_k E_m \left| \hat{\alpha}_{j_0 k,n} - \alpha_{j_0 k} \right|^r \\ &\leqslant C 2^{j_0 \frac{r}{2} - j_0} \cdot 2^{j_0} \sup_k E_m \left| \hat{\alpha}_{j_0 k,n} - \alpha_{j_0 k} \right|^r \\ &\leqslant C 2^{j_0 \frac{r}{2}} n^{-\frac{r}{2}} \\ &= C \left(\frac{2^{j_0}}{n} \right)^{\frac{r}{2}} . \end{aligned}$$

The last inequality followed from Theorem 3.11 (this is where we need the assumption $r > 2$ in Theorem 2.14, by the way). For the second-to-last inequality we remark that $|\text{supp}\varphi_{j_0 k}| = 2^{-j_0} |\text{supp}\varphi|$. Therefore, the sum had $C \cdot 2^{j_0}$ summands (and in the constant C the size of the support of φ is coded). This is why we can drop the sum and replace it by the number of summands and their maximal value. This shows (3.25).

(*ii*) For the proof of negligibility of this upper bound, let us consider several different cases:

Let $\varepsilon = 0$ (that means $sp = \frac{r-p}{2}$ by Definition 2.11). By (2.14), $2^{j_0(n)}$ grows asymptotically like

$$\left(n(\log n)^{\frac{r-p}{p}\mathbb{1}_{\{\varepsilon \geqslant 0\}}}\right)^{1-2\alpha} = \left(n(\log n)^{\frac{r-p}{p}}\right)^{1-2\alpha},$$

and so

$$\left(\frac{2^{j_0(n)}}{n}\right)^{\frac{r}{2}} \asymp \left(\frac{\left(n(\log n)^{\frac{r-p}{p}}\right)^{1-2\alpha}}{n}\right)^{\frac{r}{2}} = n^{-\alpha r}(\log n)^{\left(\frac{r}{2p}-\frac{1}{2}-\frac{\alpha r}{p}+\alpha\right)r} = \left(\frac{\log n}{n}\right)^{\alpha r}(\log n)^{r\left(\frac{r}{2p}-\frac{1}{2}-\frac{\alpha r}{p}\right)}. \tag{3.26}$$

Now, because of $\varepsilon = 0$, one gets $\alpha r = \frac{r-p}{2}$ by (3.17), and therefore the obvious equation

$$\frac{r}{2p} - \frac{1}{2} - \frac{\alpha r}{p} = 0 \qquad \Longleftrightarrow \qquad r - p - 2\frac{r-p}{2} = 0.$$

So the last logarithmic factor can be dropped. Overall, we keep

$$\left(\frac{2^{j_0(n)}}{n}\right)^{\frac{r}{2}} \asymp \left(\frac{\log n}{n}\right)^{\alpha r}.$$

This is negligible to the rate of convergence from Theorem 2.14.

If we have $\varepsilon < 0$, $2^{j_0(n)}$ behaves asymptotically like $n^{1-2\alpha}$, and so,

$$\left(\frac{2^{j_0(n)}}{n}\right)^{\frac{r}{2}} \asymp \left(\frac{n^{1-2\alpha}}{n}\right)^{\frac{r}{2}} = n^{-\alpha r},$$

which is negligible against the rate $(\log n)^{\alpha r} n^{-\alpha r}$ in (2.15).

For $\varepsilon > 0$ we get

$$\left(\frac{2^{j_0(n)}}{n}\right)^{\frac{r}{2}} \asymp n^{-\alpha r}(\log n)^{\left(\frac{r}{2p}-\frac{1}{2}-\frac{\alpha r}{p}+\alpha\right)r},$$

as in (3.26), while the bound in Theorem 2.14 is $(\log n)^{\alpha r} n^{-\alpha r}$. The $n^{-\alpha r}$-terms coincide, therefore we need to show

$$\alpha r \geqslant \left(\frac{r}{2p} - \frac{1}{2} - \frac{\alpha r}{p} + \alpha\right) r.$$

But this follows from Lemma 3.15. ▨

3.6.2 The Bias Term

3.24 Lemma: (*i*) *There is a constant $C > 0$, such that*

$$\|S - E_{j_1+1}S\|_{L^r}^r \leqslant C2^{-j_1 s' r}. \tag{3.27}$$

(*ii*) *Since we chose j_1 as in* (2.14)*, this attains the bound from* (2.15) *for $\varepsilon \neq 0$. For $\varepsilon = 0$, this rate is negligible.*

In the proof of (*i*), one needs the Besov embedding Theorem 2.7. This is why we assumed $r \geqslant p$ in Theorem 2.14.

The second assertion of the lemma is responsible for the slight change in the rate of convergence (2.15) against the rate (2.16) from [6, Thm. 3], in case $\varepsilon > 0$. As we remarked earlier, this change is not due to our different model of dependent observations as opposed to an i.i.d.-setting. The change was necessary because there is an error in the proof from [6].

PROOF. (*i*) By definition of the projection operators E_{j_1} and D_j, and because of the equivalence of the norms $\|\cdot\|_{L^r}$ and $\|\cdot\|_{\ell^r}$ from Theorem 2.6, we have

$$\begin{aligned}\|S - E_{j_1+1}S\|_{L^r} = \left\|\sum_{j>j_1} D_j S\right\|_{L^r} = \left\|\sum_{j>j_1}\sum_{k\in\mathbb{Z}} \beta_{jk}\psi_{jk}\right\|_{L^r} &\leqslant \sum_{j>j_1}\left\|\sum_{k\in\mathbb{Z}}\beta_{jk}\psi_{jk}\right\|_{L^r} \\ &\leqslant \sum_{j>j_1} c_2 2^{j(\frac{1}{2}-\frac{1}{r})}\left\|\beta_{j\bullet}\right\|_{\ell^r}.\end{aligned}$$

The Besov-embedding Theorem 2.7 yields $S \in B_{spq} \subset B_{s'rq} \subset B_{s'r\infty}$. In particular,

$$\sup_{j\geqslant 0}\left(2^{j(s'+\frac{1}{2}-\frac{1}{r})}\right)\left\|\beta_{j\bullet}\right\|_{\ell^r} = C < \infty,$$

since the left-hand side is a summand in $\|S\|_{s'r\infty}$. Thus,

$$\begin{aligned}\|S - E_{j_1+1}S\|_{L^r} &\leqslant \sum_{j>j_1} c_2 2^{j(\frac{1}{2}-\frac{1}{r})}2^{-j(s'+\frac{1}{2}-\frac{1}{r})}2^{j(s'+\frac{1}{2}-\frac{1}{r})}\left\|\beta_{j\bullet}\right\|_{\ell^r} \\ &\leqslant \sum_{j>j_1} c_2 2^{j(\frac{1}{2}-\frac{1}{r})}2^{-j(s'+\frac{1}{2}-\frac{1}{r})}C \\ &= C\sum_{j>j_1} 2^{-j(-\frac{1}{2}+\frac{1}{r}+s'+\frac{1}{2}-\frac{1}{r})} = C\sum_{j>j_1} 2^{-js'}.\end{aligned}$$

The geometric series now provides the asserted upper bound:

$$\begin{aligned}\|S - E_{j_1+1}S\|_{L^r}^r \leqslant C\left(\sum_{j>j_1} 2^{-js'}\right)^r = C2^{-(j_1+1)s'r}\sum_{j\geqslant 0} 2^{-js'r} &= C2^{-j_1 s' r}\frac{1}{1-2^{-s'r}} \\ &= C2^{-j_1 s' r}.\end{aligned}$$

(*ii*) By the choice of j_1 in (2.14) we find as the asymptotic upper bound of the bias term

$$2^{-j_1(n)s'r} \asymp \left(\left(\frac{n}{\log n}\right)^{\frac{\alpha}{s'}}\right)^{-s'r} = \left(\frac{\log n}{n}\right)^{\alpha r}.$$

In case $\varepsilon \neq 0$, this is exactly the bound from (2.15). In case $\varepsilon = 0$, the bound from (3.27) is negligible against the bounds in Theorem 2.14. ▨

3.6.3 The Thresholding Part

This part of the proof follows the lines of [6]. Let us define, referring particularly to [6, p. 522],

$$\hat{B}_j := \left\{k\colon |\hat{\beta}_{jk}| > j^{\nu} n^{-\nu}\right\}, \qquad \hat{S}_j := \hat{B}_j^{\complement},$$
$$B_j := \left\{k\colon |\beta_{jk}| > \frac{1}{2} j^{\nu} n^{-\nu}\right\}, \qquad S_j := B_j^{\complement},$$
$$B'_j := \left\{k\colon |\beta_{jk}| > 2 j^{\nu} n^{-\nu}\right\}, \qquad S'_j := B_j'^{\complement}.$$

Do not be startled by the notations for the unknown function S and the sets of integers $\hat{S}_j$, S_j and S'_j. The function will never appear with an index.

Then, by a straight-forward calculation, we arrive at

$$\begin{aligned}
\sum_{j=j_0}^{j_1} \sum_{k\in\mathbb{Z}} (\tilde{\beta}_{jk} - \beta_{jk})\psi_{jk} &= \sum_{j=j_0}^{j_1} \sum_{k\in\mathbb{Z}} (\hat{\beta}_{jk} - \beta_{jk})\psi_{jk} \mathbb{1}_{\{k\in\hat{B}_j\}} + \sum_{j=j_0}^{j_1} \sum_{k\in\mathbb{Z}} (0 - \beta_{jk})\psi_{jk} \mathbb{1}_{\{k\in\hat{S}_j\}} \\
&= \sum_{j=j_0}^{j_1} \sum_{k\in\mathbb{Z}} (\hat{\beta}_{jk} - \beta_{jk})\psi_{jk} \mathbb{1}_{\{k\in\hat{B}_j\cap(B_j\cup S_j)\}} \\
&\qquad - \sum_{j=j_0}^{j_1} \sum_{k\in\mathbb{Z}} \beta_{jk}\psi_{jk} \mathbb{1}_{\{k\in\hat{S}_j\cap(B'_j\cup S'_j)\}} \\
&= \left(\sum_{j=j_0}^{j_1} \sum_{k\in\mathbb{Z}} (\hat{\beta}_{jk} - \beta_{jk})\psi_{jk} \mathbb{1}_{\{k\in\hat{B}_j\cap S_j\}} \right. \\
&\qquad \left. + \sum_{j=j_0}^{j_1} \sum_{k\in\mathbb{Z}} (\hat{\beta}_{jk} - \beta_{jk})\psi_{jk} \mathbb{1}_{\{k\in\hat{B}_j\cap B_j\}} \right) \\
&\qquad - \left(\sum_{j=j_0}^{j_1} \sum_{k\in\mathbb{Z}} \beta_{jk}\psi_{jk} \mathbb{1}_{\{k\in\hat{S}_j\cap B'_j\}} \right. \\
&\qquad \left. + \sum_{j=j_0}^{j_1} \sum_{k\in\mathbb{Z}} \beta_{jk}\psi_{jk} \mathbb{1}_{\{k\in\hat{S}_j\cap S'_j\}} \right) \\
&=: (e_{bs} + e_{bb}) - (e_{sb} + e_{ss}),
\end{aligned}$$

As usual, we have

$$E_m \left\| \sum_{j=j_0}^{j_1} \sum_{k\in\mathbb{Z}} \tilde{\beta}_{jk}\psi_{jk} - D_{j_0 j_1} S \right\|_{L^r}^r \leqslant 4^{r-1} \left(E_m \|e_{bs}\|_{L^r}^r + E_m \|e_{bb}\|_{L^r}^r + E_m \|e_{sb}\|_{L^r}^r + E_m \|e_{ss}\|_{L^r}^r \right).$$

These four cases need to be treated separately.

As far as the condition $j_0 \leqslant j_1$ in the next lemmas is concerned, this is a technically useful assumption. If it is not fulfilled, there will be no Thresholding Part at all, because the corresponding sum will be vacuous. So, this is not too restrictive.

3.25 Lemma ("large estimator, small true value"): *(i) Let $j_0 \leqslant j_1$, and let $t > 1$ such that $\frac{rt}{t-1} > 2$. Further, let $\gamma > \frac{rt}{2}$ sufficiently large. Then, there is a constant $C > 0$ such that*

$$E_m \|e_{bs}\|_{L^r}^r \leqslant C n^{-\frac{r}{2}} 2^{j_0\left(\frac{r}{2}-\frac{\gamma}{t}\right)}.$$

(ii) This bound is negligible against (2.15).

For the choice of γ, see also Lemma 3.26.

This lemma is the reason why we have formulated the auxiliary Theorem 3.11 with some real number $\chi \geqslant r$. For any other lemma in the proof of Theorem 2.14, we can apply Theorem 3.11 with $\chi := r$, here we need to apply it to some $\chi > r$ because of a use of Hölder's inequality. Since a free parameter t is involved, we can, in principle, choose χ as close to r as we want to. But since there is no substantial gain for the rate of convergence or the assumptions in Theorem 2.14 by doing so, we will not look for an "optimal" choice (in any flavour of optimality) of χ or t.

Proof. (i) Let $\hat{f}_{jk} := \left(\hat{\beta}_{jk} - \beta_{jk}\right)\mathbb{1}_{\left\{k\in\hat{B}_j\cap S_j\right\}}$. By construction, we get

$$\hat{B}_j \cap S_j \subset \left\{ k \colon \left|\hat{\beta}_{jk} - \beta_{jk}\right| > \frac{1}{2} j^{\nu} n^{-\nu} \right\} =: \Delta_{jk}. \tag{3.28}$$

Thus, applying Hölder's inequality to the conjugated exponents t and t',

$$\begin{aligned}\sum_{k\in\mathbb{Z}} E_m\left|\hat{f}_{jk}\right|^r &= \sum_{k\in\mathbb{Z}} E_m \left| \left(\hat{\beta}_{jk} - \beta_{jk}\right)\mathbb{1}_{\left\{k\in\hat{B}_j\cap S_j\right\}} \right|^r \\ &\leqslant \sum_{k\in\mathbb{Z}} E_m \left(\left|\hat{\beta}_{jk} - \beta_{jk}\right|^r \mathbb{1}_{\left\{k\in\Delta_{jk}\right\}} \right) \\ &\leqslant \sum_{k\in\mathbb{Z}} \left(E_m\left|\hat{\beta}_{jk} - \beta_{jk}\right|^{rt'} \right)^{\frac{1}{t'}} P_m(\Delta_{jk})^{\frac{1}{t}}.\end{aligned}$$

Now, we invoke the main auxiliary Theorems 3.11 (note, that $2 < \frac{rt}{t-1} = rt' =: \chi$) and 3.12 and find

$$\sum_{k\in\mathbb{Z}} E_m\big|\hat{f}_{jk}\big|^r \leqslant \sum_{k\in\mathbb{Z}} Cn^{-\frac{rt'}{2}\cdot\frac{1}{t'}} P_m(\Delta_{jk})^{\frac{1}{t}} \leqslant Cn^{-\frac{r}{2}} 2^j \sup_{k\in\mathbb{Z}} P_m(\Delta_{jk})^{\frac{1}{t}}$$
$$\leqslant Cn^{-\frac{r}{2}} 2^{j-\frac{j\gamma}{t}}. \tag{3.29}$$

Here, when dropping the sum, we gained the factor 2^j; this is because of the same reason as in the proof of Lemma 3.23(i).

By assumption of Theorem 2.14, we have $r > 2$. Therefore, we can make use of (3.24); so for any $\kappa \in \mathbb{R}$ and for ϱ from (3.21)

$$\begin{aligned} E_m \|e_{bs}\|_{L^r}^r &= E_m \left\| \sum_{j=j_0}^{j_1} \sum_{k\in\mathbb{Z}} \hat{f}_{jk}\psi_{jk} \right\|_{L^r}^r \\ &\leqslant Ch(\kappa\varrho)^{\frac{r}{2}-1} \sum_{j=j_0}^{j_1} 2^{j\left(\frac{r}{2}-1-\frac{\kappa r}{2}\right)} \sum_{k\in\mathbb{Z}} E_m\big|\hat{f}_{jk}\big|^r. \end{aligned}$$

Now, (3.29) shows

$$\begin{aligned} E_m \|e_{bs}\|_{L^r}^r &\leqslant Ch(\kappa\varrho)^{\frac{r}{2}-1} \sum_{j=j_0}^{j_1} 2^{j\left(\frac{r}{2}-1-\frac{\kappa r}{2}\right)} Cn^{-\frac{r}{2}} 2^{j\left(1-\frac{\gamma}{t}\right)} \\ &= Ch(\kappa\varrho)^{\frac{r}{2}-1} \sum_{j=j_0}^{j_1} 2^{j\left(\frac{r}{2}-\frac{\kappa r}{2}-\frac{\gamma}{t}\right)} n^{-\frac{r}{2}} \\ &= Ch(\kappa\varrho)^{\frac{r}{2}-1} h\left((1-\kappa)\frac{r}{2}-\frac{\gamma}{t}\right) n^{-\frac{r}{2}}. \end{aligned}$$

We now choose $\kappa < 0$ and $\gamma \geqslant 1$ sufficiently large, such that $(1-\kappa)\frac{r}{2} - \frac{\gamma}{t} < 0$. Then Lemma 3.21 yields

$$E_m \|e_{bs}\|_{L^r}^r \leqslant Cn^{-\frac{r}{2}} 2^{j_0\left(\kappa\frac{r}{2}+(1-\kappa)\frac{r}{2}-\frac{\gamma}{t}\right)} = Cn^{-\frac{r}{2}} 2^{j_0\left(\frac{r}{2}-\frac{\gamma}{t}\right)}.$$

This is the asserted upper bound.

(ii) Since we chose $\gamma, t \geqslant 1$, this bound is asymptotically strictly smaller than the bound from (3.25). So, this bound does not matter for the rate of convergence in Theorem 2.14. ▨

3.26 Lemma ("small estimator, large true value"): (i) *Let $j_0 \leqslant j_1$. There is a constant $C > 0$ for which we have*

$$E_m \|e_{sb}\|_{L^r}^r \leqslant C 2^{-j_0(\gamma+s'r)}.$$

(*ii*) *If we choose* $\gamma \geqslant \frac{r\alpha}{1-2\alpha} - s'r$ *sufficiently large, this rate is negligible against* (2.15).

In this lemma, we can give a more quantitative definition for γ. Up to here, $\gamma \geqslant 1$ has always been chosen "large enough", but now it has a link to the Besov space B_{spq} under consideration and not to other free parameters only (see Lemma 3.25).

PROOF. (*i*) Let $\hat{f}_{jk} := \beta_{jk}\mathbb{1}_{\{k\in\hat{S}_j\cap B'_j\}}$. In a similar way to what we did in Lemma 3.25, we find $\hat{S}_j \cap B'_j \subset \Delta_{jk}$. Since the β_{jk} are deterministic, Theorem 3.12 implies

$$\begin{aligned}
\sum_{k\in\mathbb{Z}} E_m\left|\hat{f}_{jk}\right|^r &= \sum_{k\in\mathbb{Z}} E_m\left|\beta_{jk}\mathbb{1}_{\{k\in\hat{S}_j\cap B'_j\}}\right|^r \\
&\leqslant \sum_{k\in\mathbb{Z}} \left|\beta_{jk}\right|^r P_m\left(\Delta_{jk}\right) \\
&\leqslant C\sum_{k\in\mathbb{Z}} \left|\beta_{jk}\right|^r 2^{-\gamma j} \\
&= C\left\|\beta_{j\bullet}\right\|_{\ell^r}^r 2^{-\gamma j}2^{-j(s'+\frac{1}{2}-\frac{1}{r})r}2^{j(s'+\frac{1}{2}-\frac{1}{r})r} \\
&\leqslant C2^{-j(s'+\frac{1}{2}-\frac{1}{r})r}2^{-\gamma j}\sup_{j\geqslant 0}\left(2^{j(s'+\frac{1}{2}-\frac{1}{r})}\left\|\beta_{j\bullet}\right\|_{\ell^r}\right)^r.
\end{aligned}$$

The supremum in the last expression is a summand of the norm $\|S\|_{s'r\infty}$ of the unknown signal function S that we want to estimate. Because of the Besov embeddings

$$S \in B_{spq} \subset B_{s'rq} \subset B_{s'r\infty}$$

from Theorem 2.7, this norm is finite. Thus,

$$\begin{aligned}
\sum_{k\in\mathbb{Z}} E_m\left|\hat{f}_{jk}\right|^r &\leqslant C2^{-j(s'r+\frac{r}{2}-1+\gamma)}\left(\sup_{j\geqslant 0}2^{j(s'+\frac{1}{2}-\frac{1}{r})}\left\|\beta_{j\bullet}\right\|_{\ell^r}\right)^r \\
&\leqslant C2^{-j(s'r+\frac{r}{2}-1+\gamma)}\left(\left\|\alpha_{0\bullet}\right\|_{\ell^r}+\sup_{j\geqslant 0}2^{j(s'+\frac{1}{2}-\frac{1}{r})}\left\|\beta_{j\bullet}\right\|_{\ell^r}\right)^r \\
&= C2^{-j(s'r+\frac{r}{2}-1+\gamma)}\left\|S\right\|_{s'r\infty}^r \\
&\leqslant C2^{-j(s'r+\frac{r}{2}-1+\gamma)}M^r.
\end{aligned}$$

Note, that Theorem 2.7 implies $B_{spq}(M) \subset B_{s'rq}(M) \subset B_{s'r\infty}(M)$ as well as $\|S\|_{s'r\infty} \leqslant M$.

Now, invoking (3.24), we see

$$\begin{aligned}
E_m \|e_{sb}\|_{L^r}^r &= E_m \left\| \sum_{j=j_0}^{j_1} \sum_{k\in\mathbb{Z}} \beta_{jk}\psi_{jk}\mathbb{1}_{\{k\in\hat{S}_j\cap B_j'\}} \right\|_{L^r}^r \\
&\leqslant Ch(\kappa\varrho)^{\frac{r}{2}-1} \sum_{j=j_0}^{j_1} 2^{j\left(\frac{r}{2}-1-\kappa\frac{r}{2}\right)} \sum_{k\in\mathbb{Z}} E_m\left|\hat{f}_{jk}\right|^r \\
&\leqslant Ch(\kappa\varrho)^{\frac{r}{2}-1} \sum_{j=j_0}^{j_1} 2^{j\left(\frac{r}{2}-1-\kappa\frac{r}{2}\right)} 2^{-j(s'r+\frac{r}{2}-1+\gamma)} M^r \\
&= CM^r h(\kappa\varrho)^{\frac{r}{2}-1} h\left(-r\left(\frac{\kappa}{2}+s'\right)-\gamma\right).
\end{aligned}$$

Analogously to the proof of Lemma 3.25, we can choose $\kappa < 0$ and $\gamma \geqslant 1$ sufficiently large, such that $-r\left(\frac{\kappa}{2}+s'\right)-\gamma < 0$ is satisfied. Using (3.23), this yields

$$E_m \|e_{sb}\|_{L^r}^r \leqslant CM^r C 2^{\left(\frac{\kappa r}{2}+\left(-r\left(\frac{\kappa}{2}+s'\right)\right)-\gamma\right)j_0} = C2^{-(rs'+\gamma)j_0}.$$

(ii) Let $\gamma \geqslant \frac{\alpha r}{1-2\alpha} - s'r$ so large, that the inequality $-r\left(\frac{\kappa}{2}+s'\right)-\gamma<0$ that we used above is fulfilled. For sufficiently large n, we thus find by (2.14)

$$\begin{aligned}
\left(E_m \|e_{sb}\|_{L^r}^r\right)^{-1} \geqslant C2^{j_0\left(rs'+\frac{r\alpha}{1-2\alpha}-s'r\right)} &= C2^{j_0\frac{\alpha r}{1-2\alpha}} \\
&\asymp \left(\left(n(\log n)^{\frac{r-p}{p}\mathbb{1}_{\{\varepsilon\geqslant 0\}}}\right)^{1-2\alpha}\right)^{\frac{\alpha r}{1-2\alpha}} \\
&= \left(n(\log n)^{\frac{r-p}{p}\mathbb{1}_{\{\varepsilon\geqslant 0\}}}\right)^{\alpha r} \\
&> \left(n(\log n)^{-1}\right)^{\alpha r} = \left(\frac{n}{\log n}\right)^{\alpha r\frac{s'}{s'}} \asymp 2^{j_1 rs'}.
\end{aligned}$$

Therefore, $2^{j_1rs'}$ is negligible against $\left(E_m \|e_{sb}\|_{L^r}^r\right)^{-1}$. Taking both to the power -1 and comparing with (3.27) ends the proof. ▨

Now, we come to the main terms e_{bb} and e_{ss} which deal with the cases of similar behaviour of the estimators and the true values β_{jk}. For e_{bb}, both estimator and true coefficient take "large values".

3.27 Lemma ("large estimator and true value"): *(i) Let $j_0 \leqslant j_1$. Then, there is a constant $C > 0$ such that*

$$E_m \|e_{bb}\|_{L^r}^r \leqslant \begin{cases} C2^{\max\{-j_0\varepsilon,-j_1\varepsilon\}} n^{-\frac{r-p}{2}}, & \text{for } \varepsilon \neq 0, \\ Cj_1^{\frac{r}{2}-\nu p} n^{-\frac{r-p}{2}}, & \text{for } \varepsilon = 0. \end{cases}$$

(ii) This rate can be neglected against (2.15).

PROOF. (*i*) Let $\hat{f}_{jk} := (\hat{\beta}_{jk} - \beta_{jk})\mathbb{1}_{\{k\in\hat{B}_j\cap B_j\}}$. Trivially, we have $1^p = 1^r$ for any $r \geqslant p \geqslant 1$, and by Theorem 3.11 we get

$$\begin{aligned}\sum_{k\in\mathbb{Z}} E_m\left|\hat{f}_{jk}\right|^r = \sum_{k\in\mathbb{Z}} E_m\left(\left|\hat{\beta}_{jk}-\beta_{jk}\right|\mathbb{1}_{\{k\in\hat{B}_j\cap B_j\}}\right)^r &\leqslant \sum_{k\in\mathbb{Z}} E_m\left(\left|\hat{\beta}_{jk}-\beta_{jk}\right|\mathbb{1}_{\{k\in B_j\}}\right)^r\\ &= \sum_{k\in\mathbb{Z}} E_m\left(\left|\hat{\beta}_{jk}-\beta_{jk}\right|\right)^r\left(\mathbb{1}_{\{k\in B_j\}}\right)^p\\ &\leqslant Cn^{-\frac{r}{2}}\sum_{k\in\mathbb{Z}}\left(\mathbb{1}_{\{k\in B_j\}}\right)^p\\ &< Cn^{-\frac{r}{2}}\sum_{k\in B_j}\left(2\left|\beta_{jk}\right|n^{\nu}j^{-\nu}\right)^p;\end{aligned}$$

the last line followed from the definition of B_j. Now, since $\nu < \frac{1}{2}$, we find $n^{\nu p} \leqslant n^{\frac{p}{2}}$ and therefore

$$\begin{aligned}\sum_{k\in\mathbb{Z}} E_m\left|\hat{f}_{jk}\right|^r &\leqslant Cn^{-\frac{r}{2}}n^{\nu p}j^{-\nu p}\sum_{k\in B_j}\left|\beta_{jk}\right|^p\\ &\leqslant Cn^{-\frac{r-p}{2}}j^{-\nu p}\left\|2^{j(s+\frac{1}{2}-\frac{1}{p})}\beta_{j\bullet}\right\|_{\ell^p}^p 2^{-j(s+\frac{1}{2}-\frac{1}{p})p}\\ &\leqslant Cn^{-\frac{r-p}{2}}j^{-\nu p}2^{-j(s+\frac{1}{2}-\frac{1}{p})p}\left(\sup_{j\geqslant 0}\left\|2^{j(s+\frac{1}{2}-\frac{1}{p})}\beta_{j\bullet}\right\|_{\ell^p}+\left\|\alpha_{0\bullet}\right\|_{\ell^p}\right)^p\\ &= Cn^{-\frac{r-p}{2}}j^{-\nu p}2^{-j(s+\frac{1}{2}-\frac{1}{p})p}\left\|S\right\|_{sp\infty}^p\\ &\leqslant Cn^{-\frac{r-p}{2}}j^{-\nu p}2^{-j(s+\frac{1}{2}-\frac{1}{p})p}M^p.\end{aligned}$$

We have made use of Theorem 2.7 here by noting $\|S\|_{sp\infty}^p \leqslant M^p$.

Now, let us consider several cases.

In case $\varepsilon = sp - \frac{r-p}{2} > 0$, we choose some $\kappa < 0$ such that $-\varepsilon - \frac{\kappa}{2} < 0$ is satisfied. This, together with (3.24) and $j^{-\nu p} \leqslant 1$, implies

$$\begin{aligned}E_m\left\|e_{bb}\right\|_{L^r}^r &\leqslant Ch(\kappa\varrho)^{\frac{r}{2}-1}\sum_{j=j_0}^{j_1}2^{j\left(\frac{r}{2}-1-\frac{\kappa r}{2}\right)}\sum_{k\in\mathbb{Z}}E_m\left|\hat{f}_{jk}\right|^r\\ &\leqslant Ch(\kappa\varrho)^{\frac{r}{2}-1}\sum_{j=j_0}^{j_1}2^{j\left(\frac{r}{2}-1-\frac{\kappa r}{2}\right)}Cn^{-\frac{r-p}{2}}j^{-\nu p}2^{-j(s+\frac{1}{2}-\frac{1}{p})p}M^p\\ &\leqslant CM^p h(\kappa\varrho)^{\frac{r}{2}-1}n^{-\frac{r-p}{2}}\sum_{j=j_0}^{j_1}2^{j\left(\frac{r}{2}-1-\frac{\kappa r}{2}-sp-\frac{p}{2}+1\right)}\\ &= Ch(\kappa\varrho)^{\frac{r}{2}-1}n^{-\frac{r-p}{2}}\sum_{j=j_0}^{j_1}2^{j\left(\frac{r-p}{2}-\frac{\kappa r}{2}-sp\right)}\\ &= Cn^{-\frac{r-p}{2}}h(\kappa\varrho)^{\frac{r}{2}-1}h\left(-\varepsilon-\frac{\kappa r}{2}\right);\end{aligned}\tag{3.30}$$

here, we apply (3.23). This allows us to drop the $\kappa \in \mathbb{R}$ that we chose earlier, to get

$$E_m \|e_{bb}\|_{L^r}^r \leqslant Cn^{-\frac{r-p}{2}} 2^{-j_0\varepsilon} = Cn^{-\frac{r-p}{2}} 2^{\max\{-j_0\varepsilon, -j_1\varepsilon\}}.$$

In case $\varepsilon < 0$, we choose some $\kappa > 0$, such that $-\varepsilon - \frac{\kappa}{2} > 0$. As we did above in (3.30) we first get

$$E_m \|e_{bb}\|_{L^r}^r \leqslant Cn^{-\frac{r-p}{2}} h(\kappa\varrho)^{\frac{r}{2}-1} h\left(-\varepsilon - \frac{\kappa r}{2}\right),$$

and, again invoking (3.23),

$$E_m \|e_{bb}\|_{L^r}^r \leqslant Cn^{-\frac{r-p}{2}} 2^{-j_1\varepsilon} = Cn^{-\frac{r-p}{2}} 2^{\max\{-j_0\varepsilon, -j_1\varepsilon\}}.$$

If $\varepsilon = 0$, we make use of (3.24) by taking $\kappa = 0$, and because of $S \in B_{spq}(M)$ we see

$$\begin{aligned}
E_m \|e_{bb}\|_{L^r}^r &\leqslant CM^p h(0)^{\frac{r}{2}-1} \sum_{j=j_0}^{j_1} 2^{j(\frac{r}{2}-1)} C 2^{-j(s+\frac{1}{2}-\frac{1}{p})p} j^{-\nu p} n^{-\frac{r-p}{2}} \\
&\leqslant CM^p (j_1 - j_0)^{\frac{r}{2}-1} \sum_{j=j_0}^{j_1} 2^{j(\frac{r}{2}-sp-\frac{p}{2})} j^{-\nu p} n^{-\frac{r-p}{2}} \\
&= C(j_1 - j_0)^{\frac{r}{2}-1} \sum_{j=j_0}^{j_1} 2^{j\cdot 0} j^{-\nu p} n^{-\frac{r-p}{2}} \\
&\leqslant C(j_1 - j_0)^{\frac{r}{2}-1} n^{-\frac{r-p}{2}} \sum_{j=j_0}^{j_1} j_0^{-\nu p} \\
&= C(j_1 - j_0)^{\frac{r}{2}} n^{-\frac{r-p}{2}} j_0^{-\nu p} \\
&= C\left(\frac{j_1}{j_0} - 1\right)^{\frac{r}{2}} j_0^{\frac{r}{2}} n^{-\frac{r-p}{2}} j_0^{-\nu p} \\
&\asymp C\left(\frac{r}{r-2} - 1\right)^{\frac{r}{2}} n^{-\frac{r-p}{2}} j_0^{\frac{r}{2}-\nu p} \leqslant Cn^{-\frac{r-p}{2}} j_1^{\frac{r}{2}-\nu p}.
\end{aligned}$$

In the last line we used Lemma 3.17 and the fact $\frac{r}{2} - \nu p \geqslant \frac{r}{2} - \frac{p}{2} \geqslant 0$. This was to be shown.

(*ii*) First, let $\varepsilon > 0$. Then, by assumption, we have $\max\{-j_0\varepsilon, -j_1\varepsilon\} = -j_0\varepsilon$ and so, by (3.17) and (2.14)

$$\begin{aligned}
n^{-\frac{r-p}{2}} 2^{\max\{-j_0\varepsilon, -j_1\varepsilon\}} &= n^{\varepsilon(1-2\alpha)-\alpha r} 2^{-j_0\varepsilon} \\
&\asymp n^{\varepsilon(1-2\alpha)-\alpha r} n^{-\varepsilon(1-2\alpha)} (\log n)^{-\frac{r-p}{p}\varepsilon(1-2\alpha)} \\
&= n^{-\alpha r} (\log n)^{-\frac{r-p}{p}\varepsilon(1-2\alpha)}.
\end{aligned}$$

The $n^{-\alpha r}$-expression is the same that we can find in the assertion of Theorem 2.14, so we only have to show that the $(\log n)$-power is at most αr. But this follows from Lemma 3.16.

In case $\varepsilon < 0$ we proceed similarly. We have $\max\{-j_0\varepsilon, -j_1\varepsilon\} = -j_1\varepsilon$, and with (3.17) and (2.14) we see

$$n^{-\frac{r-p}{2}} 2^{\max\{-j_0\varepsilon,-j_1\varepsilon\}} = n^{\frac{\varepsilon\alpha}{s'}-\alpha r} 2^{-j_1\varepsilon} \asymp n^{\frac{\varepsilon\alpha}{s'}-\alpha r} n^{-\frac{\varepsilon\alpha}{s'}} (\log n)^{\frac{\alpha\varepsilon}{s'}}$$
$$= n^{-\alpha r} (\log n)^{\frac{\alpha\varepsilon}{s'}}.$$

Again, the $n^{-\alpha r}$-expression attains the rate of convergence from (2.15); the $(\log n)$-expression is negligible because of the equivalence

$$\frac{\alpha\varepsilon}{s'} < \alpha r \iff \varepsilon < s' r.$$

This holds, since, by assumption, we have $\varepsilon < 0$ and we have $s'r > 0$ because of $sp > 1$, $r \geqslant p$ and

$$s'r \geqslant s'p = sp - \frac{p}{p} + \frac{p}{r} > 1 - 1 + \frac{p}{r} \geqslant 0.$$

Let $\varepsilon = 0$. The bound that we showed in (i) is smaller than the bound from Theorem 2.14, because

$$j_1^{\frac{r}{2}-\nu p} n^{-\frac{r-p}{2}} \asymp n^{-\frac{r-p}{2}} \left(\log_2 \left(\frac{n}{\log n} \right)^{\frac{\alpha}{s'}} \right)^{\frac{r}{2}-\nu p}$$
$$\leqslant n^{-\frac{r-p}{2}} \left(\frac{\alpha}{s'} (\log_2 n - \log_2 \log n) \right)^{\frac{r-p}{2}+(\frac{1}{2}-\nu)p}$$
$$\leqslant n^{-\frac{r-p}{2}} \left(\frac{\alpha}{s'} \log_2 n \right)^{\frac{r-p}{2}} \left(\frac{\alpha}{s'} \log_2 n \right)^{(\frac{1}{2}-\nu)p}$$
$$= C n^{-\frac{r-p}{2}} (\log n)^{\frac{r-p}{2}} (\log n)^{(\frac{1}{2}-\nu)p}$$
$$= C n^{-\alpha r} (\log n)^{\alpha r} (\log n)^{(\frac{1}{2}-\nu)p}$$

since $\varepsilon = 0$ and (3.17) hold. This is the bound from (2.15). ▨

3.28 Lemma ("small estimator and true value"): *Let $j_0 \leqslant j_1$. Then, $E_m \|e_{ss}\|_{L^r}^r$ at worst attains the bound of* (2.15).

The proof relies on a result of [7]. Adopting the notation and the assumptions of this paper, [7, Thm. 3] says:

3.29 Theorem: *Let $r \geqslant p \geqslant 1$, let $\|\cdot\|_{0r(r\wedge 2)}$ be a Besov norm and let $B_{spq}(M)$ a Besov ball. Besides, let*

$$\Omega\big(\delta; \|\cdot\|_{0r(r\wedge 2)}, B_{spq}(M)\big) := \sup \left\{ \|f\|_{0r(r\wedge 2)} : f \in B_{spq}(M),\ |\beta_{jk}| < \delta \text{ for all } j,k \right\},$$

where the β_{jk} are the coefficients in the wavelet expansion of f. Then, for all sufficiently small $\delta > 0$:

$$\Omega\big(\delta; \|\cdot\|_{0r(r\wedge 2)}, B_{spq}(M)\big) \asymp M^{1-2\alpha} \delta^{2\alpha} \left(\log \frac{M}{\delta} \right)^{\left(\mathbb{1}_{\{\varepsilon=0\}} \left(\frac{1}{2} - \frac{1-2\alpha}{q} \right) \right)^+}.$$

Proof (of Lemma 3.28). First, by Theorem 2.7, together with our general assumption $r > 2$,

$$\|e_{ss}\|_{L^r} = \left\| \sum_{j=j_0}^{j_1} \sum_{k\in\mathbb{Z}} \beta_{jk}\psi_{jk}\mathbb{1}_{\{k\in\hat{S}_j\cap S'_j\}} \right\|_{L^r} \leqslant \left\| \sum_{j=j_0}^{j_1} \sum_{k\in\mathbb{Z}} \beta_{jk}\psi_{jk}\mathbb{1}_{\{k\in\hat{S}_j\cap S'_j\}} \right\|_{0r2}$$
$$\leqslant \left\| \sum_{j=j_0}^{j_1} \sum_{k\in S'_j} \beta_{jk}\psi_{jk} \right\|_{0r2}.$$

By definition of the set S'_j we have for any β_{jk} with $k \in S'_j$

$$|\beta_{jk}| \leqslant 2j^{\nu}n^{-\nu} \leqslant 2j_1^{\nu}n^{-\nu} =: \delta_n. \tag{3.31}$$

Since we deal with asymptotics for $n \to \infty$, and because j_1 grows sublinearly (roughly logarithmical) in n by (2.14), δ_n gets "sufficiently small" and we can apply Theorem 3.29. By the computations above and because the β_{jk} are deterministic, we see

$$\left(E_m \|e_{ss}\|_{L^r}^r\right)^{\frac{1}{r}} \leqslant \left(E_m \left\| \sum_{j=j_0}^{j_1} \sum_{k\in S'_j} \beta_{jk}\psi_{jk} \right\|_{0r2}^r \right)^{\frac{1}{r}}$$
$$= \left\| \sum_{j=j_0}^{j_1} \sum_{k\in S'_j} \beta_{jk}\psi_{jk} \right\|_{0r2}$$
$$\leqslant \left\| \sum_{k\in\mathbb{Z}} \alpha_{j_0k}\varphi_{j_0k} + \sum_{j\geqslant j_0} \sum_{k\in\mathbb{Z}} \mathbb{1}_{S'_j}(k)\beta_{jk}\psi_{jk} \right\|_{0r2}.$$

The function $\tilde{S} := \sum_{k\in\mathbb{Z}} \alpha_{j_0k}\varphi_{j_0k} + \sum_{j\geqslant j_0} \sum_{k\in S'_j} \beta_{jk}\psi_{jk}$ obviously is contained in $B_{spq}(M)$ (except for the new factor $\mathbb{1}_{S'_j}(k)$ it has the same wavelet coefficients as $S \in B_{spq}(M)$) and by (3.31) we have $\left|\mathbb{1}_{S'_j}(k)\beta_{jk}\right| < \delta_n$ for all $j, k \in \mathbb{Z}$. Hence,

$$\left(E_m \|e_{ss}\|_{L^r}^r\right)^{\frac{1}{r}} \leqslant \|\tilde{S}\|_{0r2} \leqslant \Omega\big(\delta_n; \|\cdot\|_{0r2}, B_{spq}(M)\big).$$

Theorem 3.29 yields

$$\left(E_m \|e_{ss}\|_{L^r}^r\right)^{\frac{1}{r}} \leqslant CM^{1-2\alpha}\left(2j_1^{\nu}n^{-\nu}\right)^{2\alpha}\left(\log\frac{Mn^{\nu}}{2j_1^{\nu}}\right)^{\left(\mathbb{1}_{\{\varepsilon=0\}}\left(\frac{1}{2}-\frac{1-2\alpha}{q}\right)\right)^+}.$$

By (2.14), $2^{j_1} \asymp \left(\frac{n}{\log n}\right)^{\frac{\alpha}{s'}}$, and so $j_1 \asymp \log \frac{n}{\log n} = \log n - \log\log n$. Therefore,

$$\begin{aligned}
\left(E_m \left\|e_{ss}\right\|_{L^r}^r\right)^{\frac{1}{r}} &\leqslant C M^{1-2\alpha} 2^{2\alpha} \left(\frac{j_1}{n}\right)^{2\alpha\nu} \left(\log \frac{M}{2} + \log \frac{n^\nu}{j_1^\nu}\right)^{\left(\mathbb{1}_{\{\varepsilon=0\}}\left(\frac{1}{2}-\frac{1-2\alpha}{q}\right)\right)^+} \\
&\leqslant C \left(\frac{j_1}{n}\right)^{2\alpha\nu} \left(\nu \log \frac{n}{j_1}\right)^{\left(\mathbb{1}_{\{\varepsilon=0\}}\left(\frac{1}{2}-\frac{1-2\alpha}{q}\right)\right)^+} \\
&\asymp \left(\frac{\log n - \log\log n}{n}\right)^{2\alpha\nu} \left(\log \frac{n}{\log n - \log\log n}\right)^{\left(\mathbb{1}_{\{\varepsilon=0\}}\left(\frac{1}{2}-\frac{1-2\alpha}{q}\right)\right)^+} \\
&\leqslant \left(\frac{\log n}{n}\right)^{2\alpha\nu} (\log n)^{\left(\mathbb{1}_{\{\varepsilon=0\}}\left(\frac{1}{2}-\frac{1-2\alpha}{q}\right)\right)^+}.
\end{aligned}$$

Thus,

$$E_m \left\|e_{ss}\right\|_{L^r}^r \leqslant \left(\frac{\log n}{n}\right)^{2\nu\alpha r} (\log n)^{\left(\mathbb{1}_{\{\varepsilon=0\}}\left(\frac{1}{2}-\frac{1-2\alpha}{q}\right)\right)^+ r}.$$

Using $\nu < \frac{1}{2}$, this is negligible to the rate of convergence from the assertion. ▨

So, we are done proving Theorem 2.14.

Chapter 4

A simple special case

In this chapter we consider a special case of Theorem 2.14. Let r be a positive integer and let σ be a constant function. Equation (2.1) then reads

$$\mathrm{d}\xi_t = S(t)\mathrm{d}t - b(\xi_t)\mathrm{d}t + \sigma\mathrm{d}W_t, \qquad \xi_0 = v \in \mathbb{R} \text{ a.s.} \tag{4.1}$$

In this setting, we can achieve the bounds on r-th moments and on large deviations far more easily than we did earlier. By [8, Ex. 2.6.7], we have

$$\int_0^T \varphi_{jk}(s)\sigma\mathrm{d}W_s \sim \mathcal{N}\left(0, \sigma^2 \int_0^T \varphi_{jk}^2(s)\mathrm{d}s\right) = \mathcal{N}(0,\sigma^2), \tag{4.2}$$

because the φ_{jk} are L^2-orthonormal. Therefore, we get bounds on r-th moments and on large deviations by well-known results for the normal distribution, without using stopping times, the Nummelin splitting technique or Markov chains at all. In particular, we do not have to use expectations with respect to the invariant measure m, since we do not have to refer to the Markov chain $(\mathbb{X}_i)_{i\geqslant 1}$ of path segments. Besides, assumption (B) can be dropped entirely in this chapter.

Keep in mind, that $\mathrm{supp}\varphi_{jk} \subset [0,T]$ for sufficiently large n and for the j,k we are interested in. Of course, the same reasoning holds for ψ as well.

4.1 Theorem (Alternative to Theorem 3.11): *Let $\chi \in \mathbb{N}$. Then there is a constant $C > 0$, such that the estimators of the wavelet coefficients satisfy*

$$E\left|\hat{\alpha}_{jk,n} - \alpha_{jk}\right|^{\chi} \leqslant Cn^{-\frac{\chi}{2}} \qquad \textit{and} \qquad E\left|\hat{\beta}_{jk,n} - \beta_{jk}\right|^{\chi} \leqslant Cn^{-\frac{\chi}{2}}.$$

Proof. We will only prove the first inequality, the second one works along the same lines. The χ-th absolute moments of the centralized normal distribution $\mathcal{N}(0,a^2)$ are

$$2^{\chi}a^{\chi}\frac{\Gamma\left(\chi+\frac{1}{2}\right)}{\sqrt{\pi}} \qquad \text{for even } \chi$$

and

$$2^{\frac{\chi}{2}+1} a^{\chi} \frac{\Gamma\left(\frac{\chi}{2}+\frac{3}{2}\right)}{\sqrt{\pi}} \qquad \text{for odd } \chi.$$

As one can see from (2.9), the desired χ-th moment of $|\hat{\alpha}_{jk,n} - \alpha_{jk}|$ is the same as of the stochastic integral in (4.2). So, we find for the estimator $\hat{\alpha}_{jk,n}$

$$E\left|\hat{\alpha}_{jk,n} - \alpha_{jk}\right|^{\chi} = E\left|\frac{1}{n}\sum_{i=1}^{n}\int_0^T \sigma\varphi_{jk}(s)\mathrm{d}\mathbb{W}_i(s)\right|^{\chi} = n^{-\chi}\sigma^{\chi}E\left|X\right|^{\chi},$$

where $X \sim \mathcal{N}(0,n)$, by (4.2). Then,

$$\begin{aligned} E\left|\hat{\alpha}_{jk,n} - \alpha_{jk}\right|^{\chi} &= n^{-\chi}\sigma^{\chi}n^{\frac{\chi}{2}}\frac{1}{\sqrt{\pi}} \cdot \left[2^{\frac{\chi}{2}+1}\Gamma\left(\frac{\chi}{2}+\frac{3}{2}\right)\mathbb{1}_{\{\chi \text{ odd}\}} + 2^{\chi}\Gamma(\chi+1)\mathbb{1}_{\{\chi \text{ even}\}}\right] \\ &= Cn^{-\frac{\chi}{2}}. \end{aligned}$$

This is just the rate that we yearned for. ▨

4.2 Remark: *a*) We did not do any bounding from above here. So, we found an exact rate of convergence in this theorem for the special case, but in general we can only give an upper bound.

b) In proving the analogue of Lemma 3.25 for the present special case, we make use of Theorem 4.1 for $\chi := \frac{rt}{t-1} > r$, where $t > 1$ is some free parameter that we may choose at will (this originates from a use of Hölder's inequality). Several other lemmas for the proof of Theorem 4.4 work with Theorem 4.1 for $\chi := r$. The important thing about this is that we have to make sure that χ and r are positive integers. We have no liberty of choosing r in the proof (since r is involved in the statement of Theorem 4.4, we need to fix it in advance), and we can choose some $\chi \in \mathbb{N}$ as we like, since the normal distribution has moments of all orders; this will only affect the constant C in which we are not interested. Now, since t in Lemma 3.25 is a free parameter not affecting the desired rate of convergence either, we can take $t = \frac{\chi}{\chi - r}$ and prove the analogue of Lemma 3.25 for our special case without changes. ♢

4.3 Theorem (Alternative to Theorem 3.12): *Let* $\nu \in \left(0, \frac{1}{2}\right)$. *For every* n *sufficiently large and every* $\gamma \geqslant 1$, *there is a constant* $C > 0$, *such that for any* $j \in \{j_0(n), \ldots, j_1(n)\}$ *we have*

$$P\left(\left|\hat{\beta}_{jk,n} - \beta_{jk}\right| > \frac{1}{2}\left(\frac{j}{n}\right)^{\nu}\right) \leqslant C2^{-\gamma j}.$$

Proof. We have the well-known asymptotic for the standard normal distribution

$$1 - \Phi(x) \asymp \frac{1}{x}\varphi(x), \qquad \text{for } |x| \to \infty.$$

From (4.2) and (2.9), $\left|\hat{\beta}_{jk,n} - \beta_{jk}\right|$ is normally distributed. Therefore, with $X \sim \mathcal{N}(0,n)$ and $Y = n^{-\frac{1}{2}}X \sim \mathcal{N}(0,1)$,

$$\begin{aligned}
P\left(\left|\hat{\beta}_{jk,n} - \beta_{jk}\right| > \frac{1}{2}\frac{j^\nu}{n^\nu}\right) &= P\left(\frac{1}{n}\left|\sigma\right|\left|\sum_{i=1}^{n}\int_0^T \psi_{jk}(s)\mathrm{d}\mathbb{W}_i(s)\right| > \frac{1}{2}\frac{j^\nu}{n^\nu}\right) \\
&= P\left(\frac{1}{n}\left|\sigma\right| X > \frac{1}{2}\frac{j^\nu}{n^\nu}\right) \\
&= P\left(Y > \frac{1}{2}\frac{nj^\nu}{\left|\sigma\right| n^\nu n^{\frac{1}{2}}}\right) \\
&= P\left(Y > \frac{1}{2}\frac{j^\nu}{\left|\sigma\right| n^{\nu-\frac{1}{2}}}\right) \\
&\asymp 2\left|\sigma\right| \frac{n^{\nu-\frac{1}{2}}}{j^\nu}\exp\left(-\frac{1}{2}\frac{j^{2\nu}}{4\sigma^2 n^{2\nu-1}}\right).
\end{aligned}$$

As always, we choose j roughly logarithmical in n, and using $\nu \in (0,\frac{1}{2})$, we have for any $\gamma \geqslant 1$

$$P\left(\left|\hat{\beta}_{jk,n} - \beta_{jk}\right| > \frac{1}{2}\frac{j^\nu}{n^\nu}\right) \leqslant C2^{-\gamma j}.$$

This is what we wanted. ▨

Thus, we can prove the following theorem, in case σ is a constant function. Again, it does not matter what σ is exactly, or whether or not we know it, as long as it is some constant real number. Note that assumption (B) is no longer necessary in this setting.

4.4 Theorem: *Consider the estimator*

$$\mathrm{TW}_n(t) := \sum_{k\in\mathbb{Z}} \hat{\alpha}_{j_0k,n}\varphi_{j_0k}(t) + \sum_{j=j_0}^{j_1}\sum_{k\in\mathbb{Z}} \tilde{\beta}_{jk,n}\psi_{jk}(t),$$

where

$$\tilde{\beta}_{jk,n} := \hat{\beta}_{jk,n}\mathbb{1}_{\left\{|\hat{\beta}_{jk,n}|>j^\nu n^{-\nu}\right\}} \qquad \textit{for some } \nu \in \big(0,\frac{1}{2}\big).$$

Let assumption (A) be satisfied and $1 \leqslant p \leqslant r < \infty$, $r \in \mathbb{N}$. Then, there is a positive, real constant C, not depending on n, such that for

$$2^{j_0(n)} \asymp \left(n(\log n)^{\frac{r-p}{p}\mathbb{1}_{\{\varepsilon\geqslant 0\}}}\right)^{1-2\alpha}, \qquad 2^{j_1(n)} \asymp \left(\frac{n}{\log n}\right)^{\frac{\alpha}{s'}} \tag{4.3}$$

we have

$$\sup_S E_m \|\mathrm{TW}_n - S\|_{L^r}^r \leqslant \begin{cases} C\left(\frac{\log n}{n}\right)^{\alpha r}, & \textit{for } \varepsilon \neq 0, \\ C(\log n)^{(\frac{1}{2}-\nu)p}\left(\frac{\log n}{n}\right)^{\alpha r}, & \textit{for } \varepsilon = 0. \end{cases} \tag{4.4}$$

4.5 Remark: *a*) Of course, Remark 2.15*b*)-*d*) on the difference between (4.4) and the bound from [6, Thm. 3] apply here as well.

b) The proof of Theorem 4.4 has already been sketched out in the previous chapter. We need to consider the Linear Part, the Bias Term and the Thresholding Part separately; the main ingredients for the proofs are Theorem 4.1 and Theorem 4.3. In the special case, we can do without several of the assumptions for these ingredients, and so Theorem 4.4 works without them. There is no additional work to be done by dropping these assumptions. ♢

Chapter 5

Some illustrative simulations

The author of this thesis carried out some computer simulations of the estimator TW_n. The programming language `R`, [27], and the package `wavethresh`, [25] were used. This package provides the user with certain procedures for the computation of Daubechies' wavelets and for the use of the estimated wavelet coefficients.

These simulations are intended for illustrative purposes, no numerical or other optimality is claimed, and not each and every assumption that was made in the preceding text and especially Theorem 2.14 is checked. The good performance of TW_n in these simulations shows, however, that the estimator appears to be robust against violations of these assumptions.

For further reference, we recall the stochastic differential equation (2.1) on which the simulations are based:

$$\mathrm{d}\xi_t = S(t)\mathrm{d}t - b(\xi_t)\mathrm{d}t + \sigma(\xi_t)\mathrm{d}W_t.$$

5.1 The estimator TW_n in several diffusion models

In this section, we always work with the function $S(t) = \sin(\pi t)$ that has period $T = 2$. For the functions b and σ we employ several choices that can be seen in the table below. The volatility σ is always chosen in such a way that $\|\sigma\|_\infty = 1$.

$\sigma(x)$	$b(x)$
$\exp(-\|x\|)$	0
$\frac{2}{\pi}\arctan(x)$	0
1	0
1	x

With these functions fixed, we compute the trajectory of the process ξ up to time nT, where n is the number of periods (observations) that we want to consider. This can be done via an

Euler-Maruyama scheme for the simulation of stochastic differential equations. Once we have the trajectory we can apply the estimator TW_n to it. For this, of course, we do not use our knowledge about what S exactly is (we do not use our knowledge about σ either, for that matter). The wavelet functions themselves can be evaluated at sufficiently fine grids using `wavethresh`; with these values of the wavelet functions, we can compute the estimators for each of the necessary wavelet coefficients as in Definition 2.9 and (2.11). Using the values of the wavelet functions from `wavethresh` again, we can compute the estimator TW_n on a sufficiently fine grid, and we can thus plot our estimate of this drift function. The threshold ν that we need to fix in advance has been set to $\nu = 0.45$ for each of our simulations. Besides, the wavelet family that serves as the basis of the computations needs to be fixed. For these simulations, Daubechies' wavelet $D6$ was chosen, which can be seen in the third line of Figure 2.1 on page 14.

On each period of S, a grid of 1000 equidistant points was used on which the trajectory was computed. Altogether, the trajectory consists of $1000n$ points, where n is the number of observed periods. In particular, this number of pseudo random variables is needed for the computation. Some simulations with 10,000 gridpoints per period were tried, but the results do not seem to improve substantially by using this finer grid; so we stick to 1000 gridpoints per period for the more extensive simulations here, and we only show the results of those.

The plots that were achieved with this method and with $n = 1000$ observations, are Figures 5.1(b), 5.2(b), 5.3(b) and 5.4(b). The red curve is the "true" function $S(t) = \sin(\pi t)$, the black curve is the estimated function $\mathrm{TW}_n(t)$.

Besides, we have several plots that show the error of estimation of TW_n, measured with an L^r-norm. In the plots below, $r = 3$ was chosen. For several choices of n, n periods of ξ were simulated and the error of estimation was computed; this was repeated 10 times to get a mean error. These plots are Figures 5.1(a), 5.2(a), 5.3(a) and 5.4(a). Here, the black dots indicate the mean error of estimation for several n, the red curve is the rate of convergence that Theorem 2.14 claims. Since we do not have any control of the constant C in this theorem, the curve was fitted to the dots "by hand" with respect to the higher values of n (the constants taken here are always quite small, each one is smaller than 2). A different value of C leads to a shift of the red curve up or down, parallel to the vertical axis in the plot. It appears that the curve follows the dots appropriately, in some models better than in others. Apparently, it takes some large values of n (such as $n \geqslant 700$) until the asymptotics for $n \to \infty$ "kick in" to yield the rate of convergence from Theorem 2.14.

The model with $\sigma(x) = \exp(-|x|)$ obviously performs best in these simulations. This is not too surprising, because this is almost a deterministic model. As soon as the trajectory of ξ departs sufficiently far from the origin, there is only very little volatility involved and so the trajectory takes its only changes from the drift S. This is why we can estimate S so well in this model.

In the model with $\sigma(x) = \frac{2}{\pi}\arctan(x)$, whenever the trajectory of ξ has values near the origin, the volatility is small. When its values are away from the origin, the higher volatility guarantees that the trajectory will return to the origin. So, the values of ξ will not depart too far from the origin and our estimator can compute its mean values without much problem.

The model with $\sigma(x) = 1$ is a white-noise model. This is, from an estimation point of view, the most tricky model of those that are considered for our estimator here. The volatility is constant and hence there is no region of low volatility where our estimator works as nice as in the previous models. A drift term b that keeps the trajectory of ξ near the origin does not appear to improve the performance of our estimator crucially, as our simulations indicate.

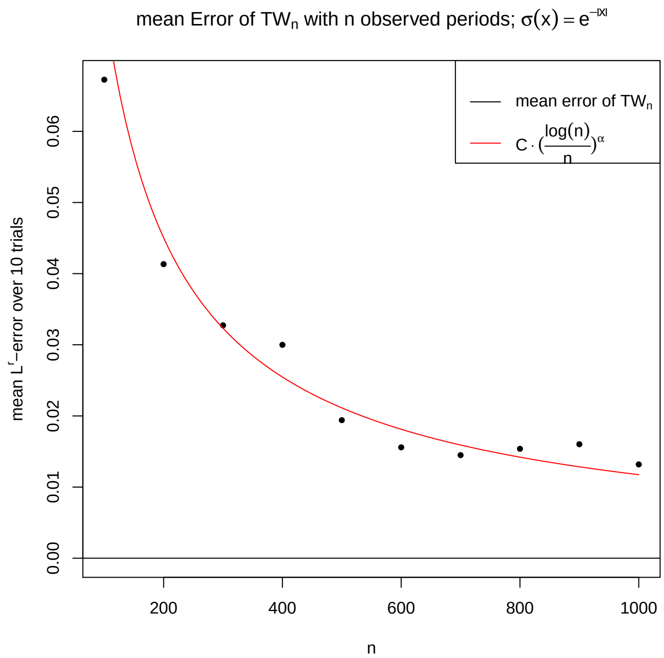
Since we took $n \in \{100, 200, \ldots, 1000\}$, we need in one of these plots

$$10 \cdot 1000 \cdot \sum_{i=1}^{10} 100 \cdot i = 55{,}000{,}000$$

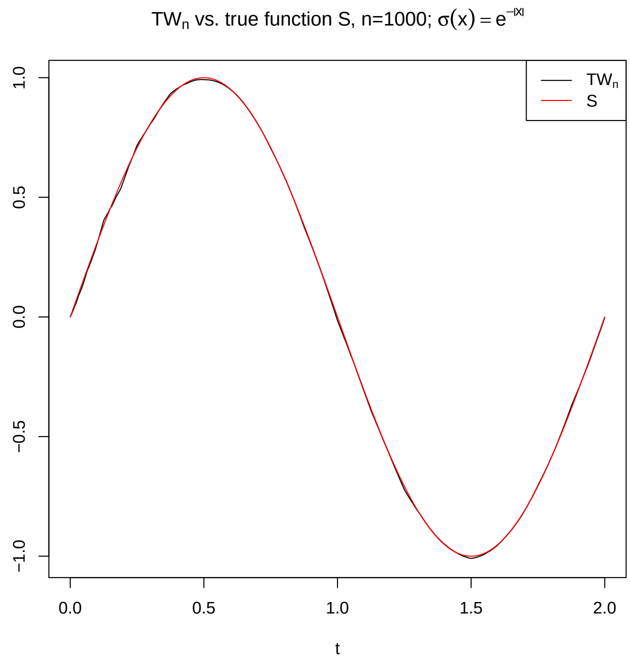
pseudo random numbers (10 trials to get a mean error; 1000 gridpoints per period; $100i$ periods for $i = 1, \ldots, 10$). Thus, when we consider one of our models (e. g. $\sigma(x) = \exp(-|x|)$, $b \equiv 0$), we need 56,000,000 pseudo random numbers in total for the two plots shown below. Since, altogether, four models were taken, we arrive at a total number of $224{,}000{,}000 \approx 2^{27}$ pseudo random numbers that are considered as i.i.d. uniform on $(0, 1)$ and transformed to i.i.d. $\mathcal{N}(0, 1)$-variables. We refer to [20] for a survey of issues that need to be kept in mind when using a random number generator in such a massive way. The standard random number generator that is implemented in `R` has been used here, a so-called Mersenne-twister. This appears to be apt for statistical simulations such as these, and it passes most of the widely used statistical tests for pseudo random numbers satisfactorily (see [21]). The transformation to normally distributed pseudo random numbers which we need for the simulation of the trajectory of ξ is done via the standard method of `R` as well, by a fast implementation of the inversion method.

5.2 A comparison of TW_n to a minimum distance estimator

In [15], a minimum distance estimator (sequence) has been introduced which solves the problem of estimation, that was considered in this thesis from a non-parametrical point of view with a wavelet estimator (sequence), parametrically in a finite-dimensional submodel. These estimators both work with the idea of cutting the trajectory, which is observed continuously in time, into parts and to gain an estimate from averaging these parts in an appropriate way. In [15], the reader can find a proof of asymptotic normality of this minimum distance estimator with a $\sqrt{n}$-rate of convergence under certain regularity and identifiability conditions to the statistical model. This parametric estimator also requires the condition that the volatility σ is bounded and bounded away from 0.

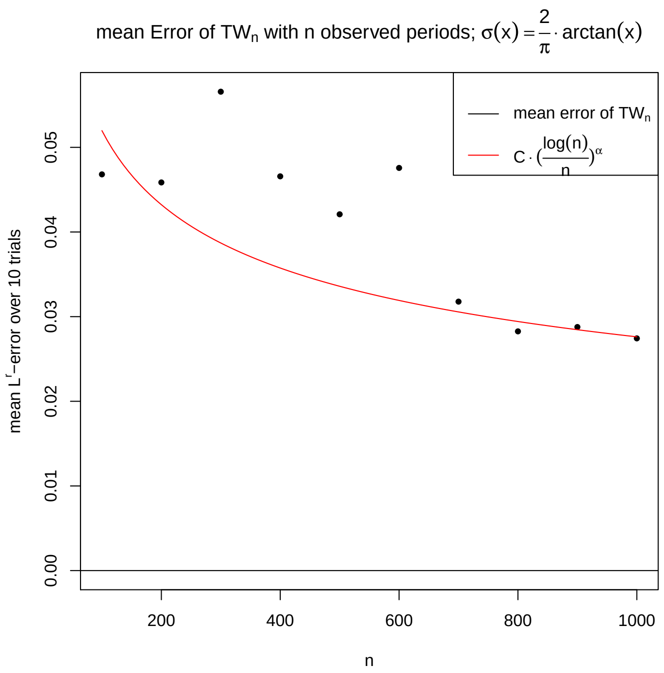


(a) Mean L^r-error, $r = 3$, of TW_n with n observations for several n. Each dot is the mean value of 10 trials.

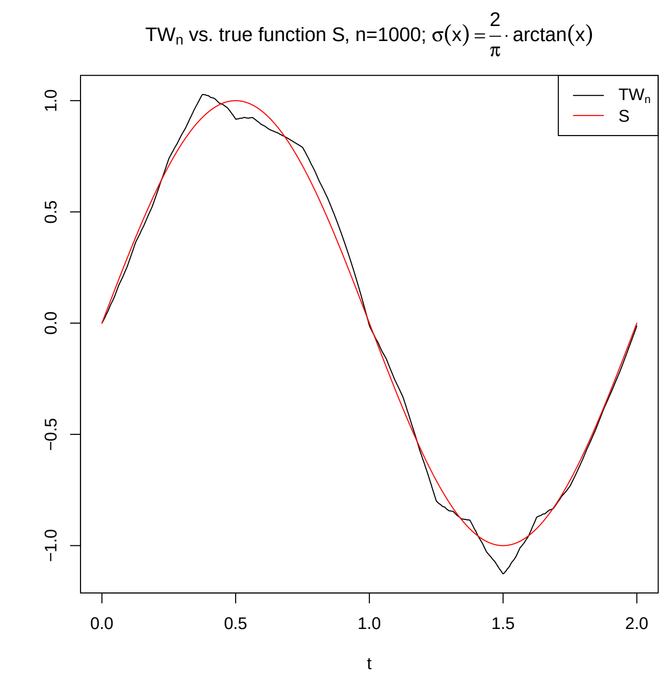


(b) The “true” signal function $S(x) = \sin \pi x$ and the estimated function $\mathrm{TW}_n(x)$ with $n = 1000$

Figure 5.1: Performance of TW_n in case $\sigma(x) = \exp(-|x|)$, $b(x) = 0$.

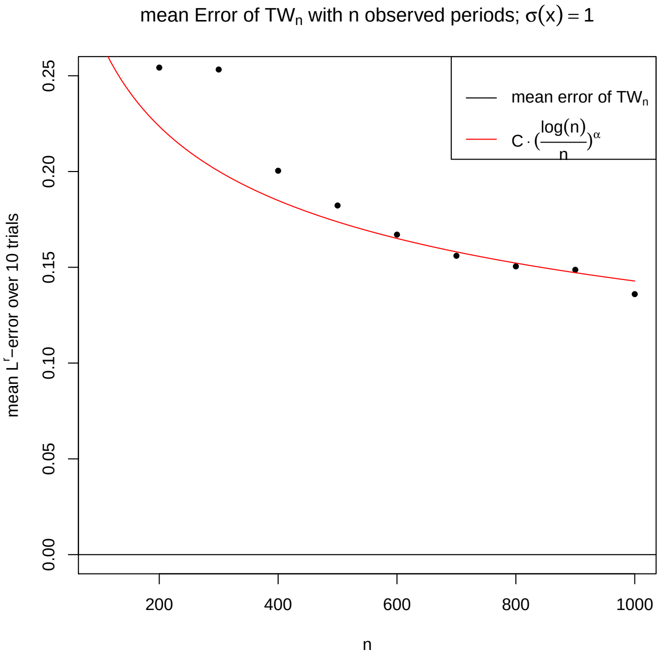


(a) Mean L^r-error, $r = 3$, of TW_n with n observations for several n. Each dot is the mean value of 10 trials.

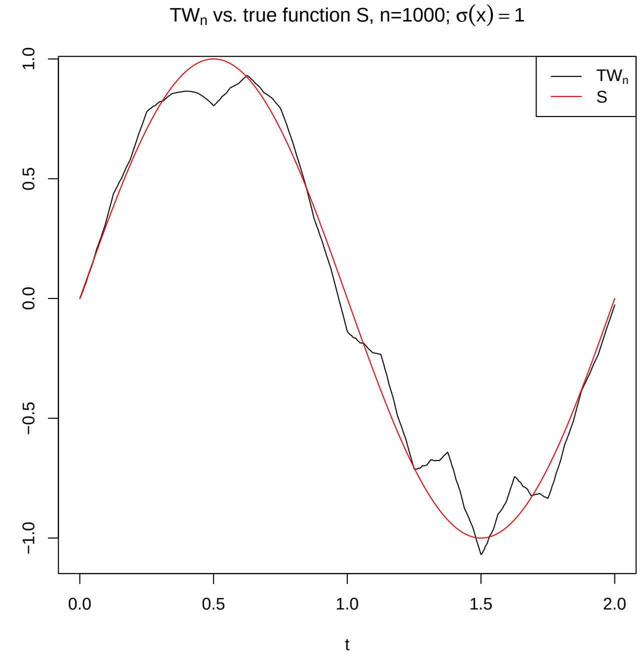


(b) The "true" signal function $S(x) = \sin \pi x$ and the estimated function $\mathrm{TW}_n(x)$ with $n = 1000$

Figure 5.2: Performance of TW_n in case $\sigma(x) = \frac{2}{\pi} \arctan(x)$, $b(x) = 0$.

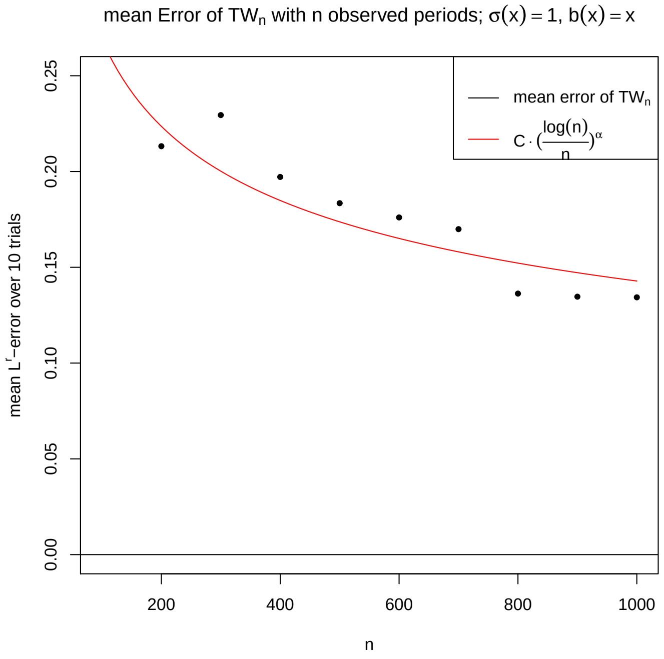


(a) Mean L^r-error, $r = 3$, of TW_n with n observations for several n. Each dot is the mean value of 10 trials.

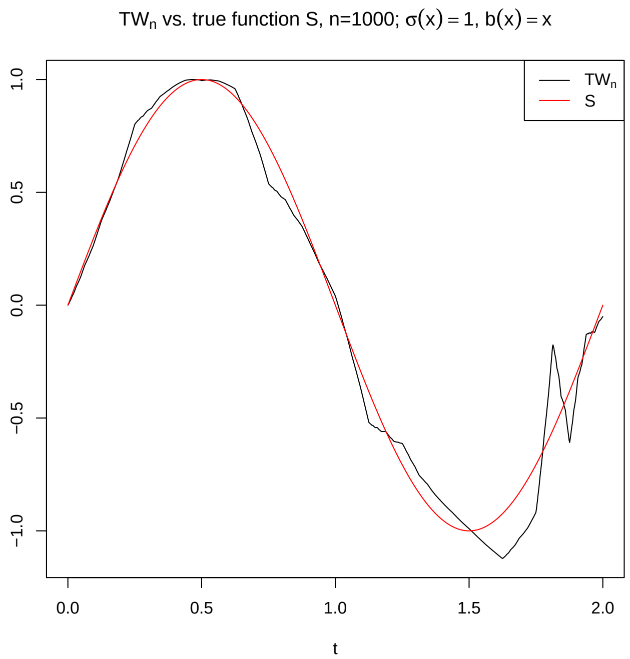


(b) The "true" signal function $S(x) = \sin \pi x$ and the estimated function $\mathrm{TW}_n(x)$ with $n = 1000$

Figure 5.3: Performance of TW_n in case $\sigma(x) = 1$, $b(x) = 0$.



(a) Mean L^r-error, $r = 3$, of TW_n with n observations for several n. Each dot is the mean value of 10 trials.



(b) The "true" signal function $S(x) = \sin \pi x$ and the estimated function $\mathrm{TW}_n(x)$ with $n = 1000$

Figure 5.4: Performance of TW_n in case $\sigma(x) = 1$, $b(x) = x$.

One of the models considered in the previous section 5.1 also appears in [15, Ex. 2.4], the model with constant σ and linear drift term b. In the simulations of the present section 5.2, $\sigma \equiv 1$ and $b(x) = 0.1x$ were used. Since the results of convergence of both estimators are very different (convergence of the L^r-norm as opposed to asymptotic normality) and since both rates of convergence are different as well, the estimators are compared only heuristically by plotting the estimated drift functions. The "true" drift function S was chosen as

$$S(t) = \sum_{\ell=1}^{4} \vartheta_\ell \sin\left(\frac{2\pi}{T}\ell t\right), \qquad \text{with } \vartheta = (\vartheta_1, \vartheta_2, \vartheta_3, \vartheta_4) = (2,\ 0.5,\ 1.25,\ 0.75).$$

The period is $T = 2$, as in the previous section.

The minimum distance estimator is computed by discretizing a subset of the parameter space, in the present case $\Theta = \mathbb{R}^4$. The set $\{k \cdot 0.05 \colon k = 0, \ldots, 60\}^4$ was used as discretization. A finer grid and a larger subset of Θ would lead to improved results, but since this simulation only deals with heuristical considerations, the coarser grid was kept for reasons of keeping the computational time at a feasible level. For each grid point $(\zeta_1, \zeta_2, \zeta_3, \zeta_4)$, we can compute the functional

$$\Psi_\zeta(s) = \int_0^s S_\zeta(t)\mathrm{d}t$$

of the corresponding function $S_\zeta(t) = \sum_{\ell=1}^4 \zeta_\ell \sin\left(\frac{2\pi}{T}\ell t\right)$. This is compared to a functional Ψ of the observed trajectory of ξ, namely

$$\begin{aligned}\hat{\Psi}_n(t) &:= \frac{1}{n}\sum_{i=1}^{n}\left[\xi_{(i-1)T+t} - \xi_{(i-1)T} - \int_{(i-1)T}^{(i-1)T+t} b(\xi_r)\mathrm{d}r\right] \\ &= \frac{1}{n}\sum_{i=1}^{n}\left[\mathbb{X}_i(t) - \mathbb{X}_i(0) - \int_0^t b\big(\mathbb{X}_i(r)\big)\mathrm{d}r\right],\end{aligned}$$

in the notation of Chapter 2. Then, the minimum distance estimator is computed as

$$\hat{\vartheta}_n := \operatorname*{argmin}_\zeta \left\|\hat{\Psi}_n - \Psi_\zeta\right\|_{L^2}.$$

Here, in Figures 5.5-5.8, we show plots that compare the wavelet estimator TW_n, the minimum distance estimator $\hat{\vartheta}_n$ and the true drift function S, for several choices of n. As before, the black line shows the estimated drift function from the wavelet estimator TW_n, the red line is the "true" drift function S. The blue line is the result of the minimum distance estimator $\hat{\vartheta}_n$. One can see that $\hat{\vartheta}_n$ comes close to the true parameter ϑ with sufficiently many observations, but does not pinpoint ϑ precisely with $n = 1000$ observations. The wavelet estimator shows a more "chaotic" behaviour with $n = 100$ observations, but gives a quite accurate estimate of the true drift function with more observations. Both estimators considered here perform about similarly well.

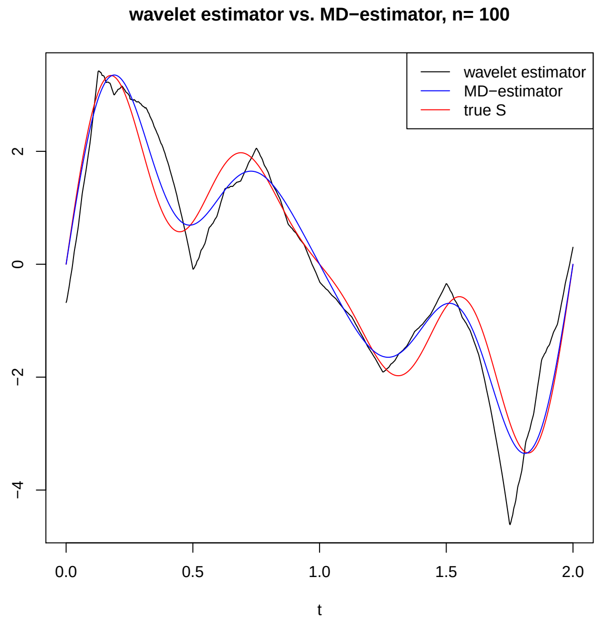


Figure 5.5: Performance of TW_n and $\hat{\vartheta}_n$ in a 4-dimensional parametric model, with $\sigma(x) = 1$, $b(x) = 0.1x$, and $n = 100$ observations. Here, $\hat{\vartheta}_n = (2, 0.7, 1.3, 0.45)$ and the error of TW_n is $\|\mathrm{TW}_n - S\|_{L^3} \approx 0.895$.

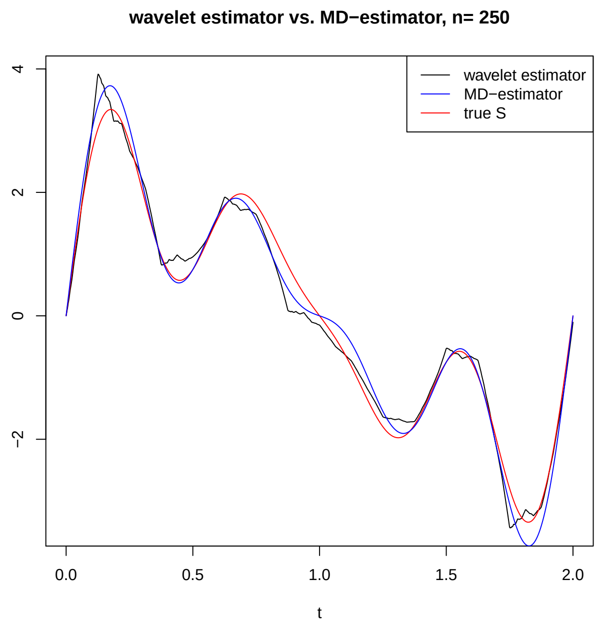


Figure 5.6: Performance of TW_n and $\hat{\vartheta}_n$ in a 4-dimensional parametric model, with $\sigma(x) = 1$, $b(x) = 0.1x$, and $n = 250$ observations. Here, $\hat{\vartheta}_n = (2, 0.75, 1.25, 0.95)$ and the error of TW_n is $\|\mathrm{TW}_n - S\|_{L^3} \approx 0.371$.

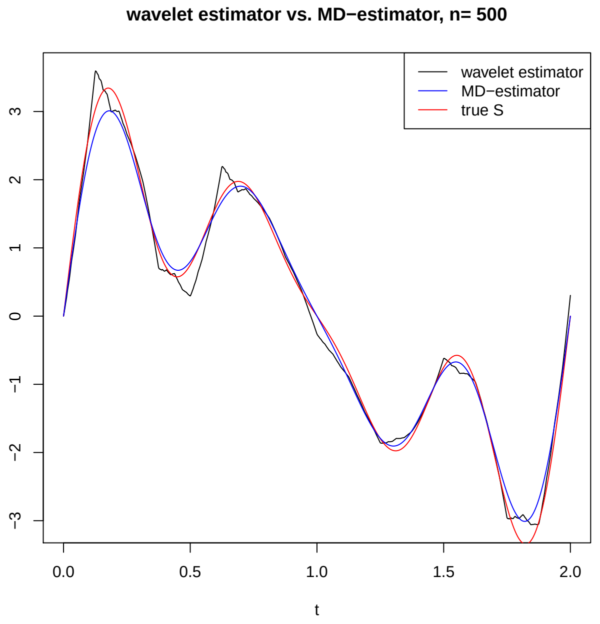


Figure 5.7: Performance of TW_n and $\hat{\vartheta}_n$ in a 4-dimensional parametric model, with $\sigma(x) = 1$, $b(x) = 0.1x$, and $n = 500$ observations. Here, $\hat{\vartheta}_n = (1.95, 0.4, 1.15, 0.6)$ and the error of TW_n is $\|\mathrm{TW}_n - S\|_{L^3} \approx 0.251$.

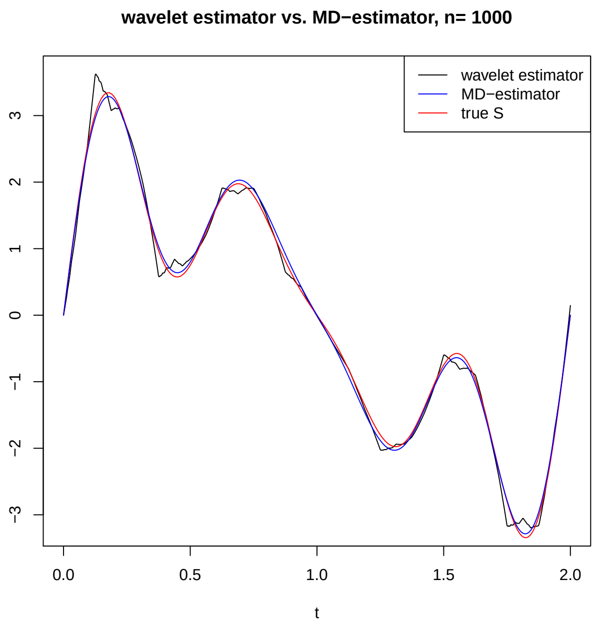


Figure 5.8: Performance of TW_n and $\hat{\vartheta}_n$ in a 4-dimensional parametric model, with $\sigma(x) = 1$, $b(x) = 0.1x$, and $n = 1000$ observations. Here, $\hat{\vartheta}_n = (2.05, 0.45, 1.25, .7)$ and the error of TW_n is $\|\mathrm{TW}_n - S\|_{L^3} \approx 0.199$.

Finally, we show a simulation of a higher dimensional parametric model, where we try to estimate

$$S(t) = \sum_{\ell=1}^{10} \vartheta_\ell \sin\left(\frac{2\pi}{T}\ell t\right), \qquad \text{with } \vartheta = (\vartheta_1, \dots, \vartheta_{10}) = (2,\ 0.5,\ 1.25,\ 0.75, 0, 0, 0, 1, 0, 1).$$

As above, $\sigma(x) \equiv 1$ and $b(x) = 0.1x$ are used. There is no hope of computing the minimum distance estimator with the approach given above in reasonable time. The wavelet estimator, however, yields the estimate shown in Figure 5.9. The behaviour of the drift function is recovered closely, but still rather crude. For more details of the unknown function, some more observations would be necessary. The parametric estimator would, due to the high dimensional model, provide no more useful result with this number of observations, even if we could compute it efficiently.

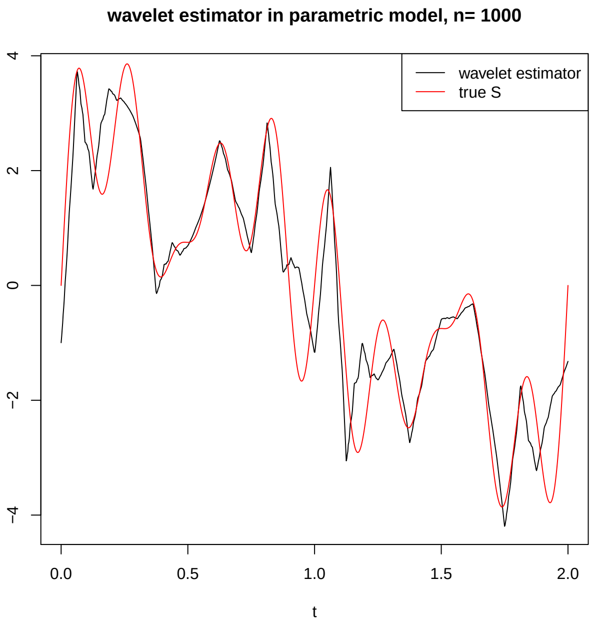


Figure 5.9: Performance of TW_n in a 10-dimensional parametric model, again with $\sigma(x) = 1$, $b(x) = 0.1x$, and $n = 1000$ observations. The error of TW_n is $\|\mathrm{TW}_n - S\|_{L^3} \approx 1.073$.

Bibliography

[1] Jöran Bergh and Jörgen Löfström. *Interpolation spaces - An Introduction.* Springer, New York, 1st edition, 1976.

[2] Patrick Cattiaux, Arnaud Guillin, and Pierre André Zitt. Poincaré inequalities and hitting times. `http://arxiv.org/pdf/1012.5274v1`, 2010.

[3] Stéphan J. M. Clémençon. Moment and probability inequalities for sums of bounded additive functionals of regular Markov chains via the Nummelin splitting technique. *Statistics and Probability Letters*, 55:227–238, 2001.

[4] Ingrid Daubechies. *Ten Lectures on Wavelets.* SIAM, Philadelphia, 1st edition, 1992.

[5] Michael Diether. Dichteschätzung mit Wavelets. Diploma thesis, Johannes Gutenberg-Universität, Mainz, December 2008. `http://ubm.opus.hbz-nrw.de/volltexte/2009/1942/pdf/diss.pdf`.

[6] David Donoho, Iain Johnstone, Gérard Kerkycharian, and Dominique Picard. Density estimation by wavelet thresholding. *Annals of Statistics*, 24:508–539, 1996.

[7] David Donoho, Iain Johnstone, Gérard Kerkycharian, and Dominique Picard. Universal near minimaxity of wavelet shrinkage. *Festschrift for L. LeCam*, pages 183–218, 1997.

[8] Richard Durrett. *Stochastic Calculus – A practical Introduction.* CRC Press, 1st edition, 1996.

[9] Kacha Dzhaparidze and Harry van Zanten. On Bernstein-type inequalities for martingales. *Stochastic Processes and their Applications*, 93:109–117, 2001.

[10] Dao H. Fuk and Sergey V. Nagaev. Probability inequalities for sums of independent random variables. *Theory of Probability and its Applications*, 16(4):643–660, 1971.

[11] Valentine Genon-Catalot, Catherine Larédo, and Dominique Picard. Non-parametric estimation of the diffusion coefficient by wavelets methods. *Scandinavian Journal of Statisics*, 19:317–335, 1992.

[12] Wolfgang Härdle, Gérard Kerkyacharian, Dominique Picard, and Alexander Tsybakov. *Wavelets, Approximation and statistical Applications.* Springer, New York, 1st edition, 1998.

[13] Eugenio Hernandez and Guido Weiss. *A first Course on Wavelets.* Prentice Hall, Upper Saddle River, New Jersey, 1st edition, 1996.

[14] Reinhard Höpfner and Yury Kutoyants. Estimating discontinuous periodic signals in a non-time homogeneous diffusion process. *Statistical Inference for Stochastic Processes*, 13:193–230, 2009.

[15] Reinhard Höpfner and Yury Kutoyants. On LAN for parametrized continuous periodic signals in a time inhomogeneous diffusion. *Statistics & Decisions*, 27:309–326, 2009.

[16] Reinhard Höpfner and Eva Löcherbach. *Limit Theorems for null recurrent Markov Processes.* Memoirs of the AMS, 1st edition, 2003.

[17] Reinhard Höpfner and Eva Löcherbach. On ergodicity properties for time inhomogeneous Markov processes with T-periodic semigroup. `http://arxiv.org/pdf/1012.4916v3`, 2011.

[18] Nobuyuki Ikeda and Shinzo Watanabe. *Stochastic Differential Equations and Diffusion Processes.* North-Holland Publishing Company, 2nd edition, 1989.

[19] Gérard Kerkycharian and Dominique Picard. Density estimation by kernel and wavelet methods: Optimality of Besov spaces. *Statistics and Probability Letters*, 18:327–336, 1993.

[20] Donald Knuth. *The Art of Computer Programming II.* Addison Wesley Longman, 3rd edition, 1988.

[21] Pierre l'Ecuyer and Richard Simard. Testu01: A C library for empirical testing of random number generators. *ACM Transactions on Mathematical Software*, 33(4):22:1–22:40, 2007.

[22] Eva Löcherbach, Oleg Loukianov, and Dasha Loukianova. Spectral condition, hitting times and Nash inequality. `http://arxiv.org/pdf/1103.4622v2`, 2011.

[23] Yves Meyer. *Ondelettes et opérateurs.* Hermann, Paris, 1st edition, 1990.

[24] Sean P. Meyn and Richard L. Tweedie. *Markov Chains and Stochastic Stability.* Springer, 1st edition, 1993.

[25] Guy Nason. `wavethresh`*: Wavelets statistics and transforms*, 2010. `R` package version 4.5.

[26] Esa Nummelin. A splitting technique for Harris recurrent Markov chains. *Zeitschrift für Wahrscheinlichkeitstheorie und verwandte Gebiete*, 43:309–318, 1978.

[27] R Development Core Team. *R: A Language and Environment for Statistical Computing.* R Foundation for Statistical Computing, Vienna, Austria, 2011. ISBN 3-900051-07-0; version 2.12.0.