INSTITUT DE STATISTIQUE

UNIVERSITÉ CATHOLIQUE DE LOUVAIN

DISCUSSION PAPER

0516

MULTIVARIATE WAVELET-BASED SHAPE PRESERVING ESTIMATION FOR DEPENDENT OBSERVATIONS

A. COSMA, O. SCAILLET and R. von SACHS

http://www.stat.ucl.ac.be

Multivariate wavelet-based shape preserving estimation for dependent observations

Antonio Cosma[1] Olivier Scaillet[2] Rainer von Sachs[3, 4]

May 24, 2005

Abstract

We present a new approach on shape preserving estimation of probability distribution and density functions using wavelet methodology for multivariate dependent data. Our estimators preserve shape constraints such as monotonicity, positivity and integration to one, and allow for low spatial regularity of the underlying functions. As important application, we discuss conditional quantile estimation for financial time series data. We show that our methodology can be easily implemented with B-splines, and performs well in a finite sample situation, through Monte Carlo simulations.

Keywords: Conditional quantile, time series, shape preserving wavelet estimation, B-splines, multivariate process.

Abbreviated title: Shape preserving wavelet estimation

AMS 2000 classification: 62G05, 62G07, 42C40, 41A15. *JEL classification:* C14, C15, C32.

[1]Instituto di finanza, University of Lugano, Lugano, Suisse. `antonio.cosma@lu.unisi.ch`

[2]HEC Genève and FAME, Genève, Suisse. `olivier.scaillet@hec.unige.ch`

[3]*Corresponding author.* Institut de statistique, Université catholique de Louvain, Louvain-la-Neuve, Belgique. `vonsachs@stat.ucl.ac.be`

[4]The first and second authors acknowledge financial support by the Swiss National Science Foundation through the National Center of Competence: Financial Valuation and Risk Management (NCCR FINRISK), while the third author gratefully acknowledges financial support from the IAP research network nr P5/24 of the Belgian Government (Belgian Science Policy).

1 Introduction

Nonparametric estimation of probabilistic functions such as probability density functions (pdf) and cumulative distribution functions (cdf) has attracted recent new interest when considering the task of constructing shape-preserving estimators, see for instance Cheng et al. (1999). Quite naturally, when estimating a pdf, it is highly desirable to have an estimator which fulfills conditions on a pdf, i.e. non-negativity and integration to one. Similarly an estimator of a cdf should respect monotonicity and right-continuity. In contrast to most nonparametric approaches, we aim at being able to treat probabilistic functions with low spatial regularity, i.e. allowing for occasional jumps or other discontinuities. This is a set-up where typically wavelet methods have been proven advantageous (Vidakovic (1999), Ogden (1996)), but it is not obvious how to do wavelet estimation under the afore-mentioned shape constraints.

In this paper we treat shape-preserving wavelet estimation of pdf and cdf having observed time series data. We do not need post-processors to make existing wavelet estimators satisfy the shape constraints. Our construction is shape-preserving but not shape-imposing (we can treat also non-monotone or non-positive functions). We start from the construction of Dechevsky and Penev (1997, 1998), hereafter DP, who study, from a pure theoretical point of view, the *univariate* case of estimating probabilistic functions with low spatial regularity with non-orthogonal wavelets, in the case of *i.i.d.* data, only. In particular they do not come up with any algorithm to implement their method on data. Our goal here is to examine a broad multivariate set-up where each component can be either a pdf or a cdf, under least stringent conditions on these functions compared to the usual ones underlying nonparametric approaches. We will see that a direct transfer of the DP-construction to a multivariate set-up is not possible. One of the main motivations for this framework is conditional quantile estimation, for dependent data in financial time series. Hence we are aiming at a general methodology which can estimate a $d-$dimensional function which is a cdf in the first component, and a multivariate pdf in the remaining $d-1$ components. In order to do so we have to take into account, on top of the shape-preserving property of our resulting estimator, the fact that the usual wavelet methodology does not apply to cdf. To summarize the main three contributions of this paper are, in this context,

- the definition of appropriate norms of convergence of a multivariate estimator in case one component consists in a cdf; we recall that cdf are not integrable over the real line and hence a classical L_p-construction, $p < \infty$, for wavelets cannot be used. Additional problems in the multivariate set-up arise as well;
- the generalisation of the univariate results of DP derived for i.i.d. data to the multivariate case of dependent (time series) data under mixing conditions; here we will provide for conditions that are weaker than those given in the literature (e.g. by Masry (1994); for more details on the literature on density estimation, see right below);
- the proposal of fast and tractable algorithms which have not yet been considered, even in the univariate case; having quick algorithms proves particularly useful for statistical inference via computer intensive methods.

Of course, results in wavelet density estimation have already been established previously. Early work has been made by Doukhan (1988) and by Doukhan and Léon (1990) for univariate independent data. Masry provides a generalisation to dependent data using orthonormal bases in the univariate case (Masry (1994)) and in the multivariate case (Masry (1997)), but the latter analysis is limited to the case of uniform convergence on compact sets. Kerkyacharian and Picard (1992) are, to our knowledge, the first to derive, for linear wavelet methods, optimality results for densities in certain function spaces, such as Besov spaces. Tribouley (1995) studies multivariate densities using linear wavelet methods. Later, non-linear wavelet threshold methodology has successfully been applied to (univariate) density estimation, in, e.g., Donoho et al. (1996) and Kerkyacharian et al. (1996), and Tribouley and Viennet (1998) for the specific case of $\beta-$mixing data. Note that for these latter non-linear methods, orthogonality (or some slight deviation such as bi-orthogonality) of the underlying wavelet bases has to be imposed. More recent studies on density estimation with wavelets aim at dealing with the shape preserving properties mentioned above, see for instance Penev and Dechevsky (1997) and Pinheiro and Vidakovic (1997). We recall that these approaches use devices such as pre- or post-processing in order to have estimates satisfying the desired shape properties.

Choosing the DP shape preserving wavelets as starting point for our work does not only overcome the need of pre- or postprocessors but also provides us with the following advantages: 1) As orthogonality has to be given up for those wavelets, we rely on an extremely simple construction using B-splines. It allows us to have analytic expressions of our basis functions in the time domain, and this is essential to be able to construct the cdf reconstruction by integration. 2) The same proof technique can be used for the results in the pdf case and in the cdf case. 3) Our approach derives very general results for linear wavelet density estimation without needing to restrict ourselves to the Besov space framework of Kerkyacharian and Picard (1992).

Shape-preserving estimation of probabilistic functions turns out to be interesting for a variety of nonparametric estimation problems. To give only a few examples beyond multivariate density estimation, we state hazard rate estimation (Hall and Van Keilegom (2004)) which calls for monotonicity, and logistic regression, see for instance McFadden and Train (2000) for an application to Mixed Multinomial Logit Models (MMNL). As a matter of fact, D. McFadden in his Nobel Prize lecture (McFadden (2003)), precisely suggests that an appropriate multivariate extension of the DP set-up should be used in the MMNL framwork. This would assure that the multivariate indirect utility functions determining the choice probabilities display the required shape restrictions. Another application is quantile regression, with its inherent property to provide for robust estimators (such as the median). We briefly discuss quantile regression when we develop our main application, namely nonparametric estimation of conditional quantiles for time series, the conditioning information being the past realized observations of the time series. As pointed out by Hall et al. (1999) and Cai (2002), the shape preserving property of the estimates of cdf is particularly important for the problem of quantile estimation. Here we strongly benefit in having directly (without post-processing) a wavelet estimate which is monotone and constrained to lie between 0 and 1. This is in contrast to other popular methods for quantile regression - see for instance the modified local linear quantile estimators of Yu and Jones (1998).

Note that the time series framework where the conditioning variable is the lagged value of the observed process calls for a particular care in proving consistency of our estimator, again under least stringent smoothness assumptions on the conditional and marginal distribution functions of the random process.

Our paper will be organized in the following way.

Section 2 gives an introduction to shape preserving wavelets and estimation of univariate probabilistic functions, both pdf and cdf, as constructed by DP. We present the relevant terminology and concepts such as moduli of smoothness, seminorms, as well as appropriate risk definitions. In Section 3 we briefly recall the main results of DP which are essential for our work. For details summarizing their work we refer to our Appendix A. Section 4 presents our theoretical contributions: Theorem 1 states the most general result of this article. It is an extension of the univariate results of Section 3 to higher dimensions and to the case of dependent data. Section 5 treats as important applications, quantile regression and, in particular, conditional quantile estimation in the case of financial time series data. For the latter one, we discuss numerical implementation via B-splines, and present a simulation study. In a short conclusion we discuss some ideas for future research. All proofs are deferred to an appendix section.

2 Preliminaries on shape-perserving wavelets

We start this section by introducing the concept of Multiresolution Analysis (MRA). Let $L_2(\mathbb{R})$ be the usual space of square integrable functions defined on the real line, that is $L_2(\mathbb{R}) = \{f : \mathbb{R} \to \mathbb{R} \,|\, \int_{-\infty}^{+\infty} dx f(x)^2 < \infty\}$. A MRA is a sequence of closed subspaces $V_j \subset L_2(\mathbb{R}), j \in \mathbb{Z}$, with the following properties (Meyer (1992)):

$$V_j \subset V_{j+1}, \qquad \cap V_j = \{\mathbf{0}\}, \qquad \overline{\cup V_j} = L_2(\mathbb{R}), \tag{2.1}$$

and for all $v(x) \in L_2(\mathbb{R})$ and $j, k \in \mathbb{Z}$, $v(x) \in V_j \Leftrightarrow v(2x) \in V_{j+1}$ and $v(x) \in V_0 \Leftrightarrow v(x-k) \in V_0$. Moreover, a *scaling function* $\varphi \in V_0$ exists such that $\{\varphi(x-l)|l \in \mathbb{Z}\}$ is a Riesz basis of V_0. It follows that in general $\varphi(2^j x) \in V_j$ and $\{\varphi_{jk}(x)\}_{k\in\mathbb{Z}} \doteq \{2^{j/2}\varphi(2^j x - k) \mid k \in \mathbb{Z}\}$ is a Riesz basis in V_j.

The idea of orthogonal projection was the first one to be exploited in wavelet analysis, leading to the construction of orthonormal bases of scaling functions, such that $\int_{-\infty}^{+\infty} \varphi(x-l)\varphi(x-k)dx = \delta_{kl}$. However, orthogonality poses a number of constraints on the construction of scaling functions, that is for instance it is not possible to construct $\{\varphi(\cdot - l)\}_{l\in\mathbb{Z}}$ that are at the same time orthogonal, non-negative, symmetric, and compatible with some degree of smoothness. Moreover we want to build shape preserving operators, that is projection operators that map nonnegative functions to nonnegative functions and monotone functions to monotone functions. To achieve this, we need some extra freedom in building the proper scaling functions. We then introduce a non-orthogonal projection operator starting from two different families of scaling functions $\{\varphi(2^j x - k)\}_{k\in\mathbb{Z}}$ and $\{\tilde{\varphi}(2^j x - k)\}_{k\in\mathbb{Z}}$. The projector on the space V_j is given by:

$$A_j\big(f\big)(x) = \sum_{k\in\mathbb{Z}} 2^j \langle f, \tilde{\varphi}(2^j \cdot - k)\rangle \; \varphi(2^j x - k). \tag{2.2}$$

The two families of functions are called *primal* basis, the one generated from the scaling function φ, and *dual* basis, the one generated from the scaling function $\tilde{\varphi}$. We ask for the two families to display the following properties. Let φ be such that

$$\varphi(x) \geqslant 0, \quad x \in \mathbb{R}; \tag{2.3}$$

$$\varphi(x) \quad \text{bounded, right continuous;} \tag{2.4}$$

$$\text{supp}\,\varphi \subset [-a, a), \quad a \geqslant 1/2; \tag{2.5}$$

$$\sum_{k=-\infty}^{\infty} \varphi(x-k) \equiv 1, \text{ on } \mathbb{R}; \tag{2.6}$$

$$\begin{aligned}&\text{There exists } b \in (-a, a) \text{ such that}\\ &\varphi \text{ is not decreasing for } \; x \leqslant b \; \text{ and non-increasing for } \; x \geqslant b.\end{aligned} \tag{2.7}$$

Then let the *dual scaling function*, be such that:

$$\tilde{\varphi} \text{ satisfies (2.3), (2.5),} \quad \tilde{\varphi} \in L_1, \quad \text{and} \quad \int_{-\infty}^{+\infty} dt\, \tilde{\varphi}(t) = 1. \tag{2.8}$$

Analogously to the primal basis, we define the scaled versions of $\tilde{\varphi}$: $\{\tilde{\varphi}_{jk}(x)\}_{k\in\mathbb{Z}} \doteq \{2^{j/2}\tilde{\varphi}(2^j x - k) \mid k \in \mathbb{Z}\}$.

The conditions given on $\tilde{\varphi}$ are weaker than those on φ. Condition $\sum_{-\infty}^{\infty} \varphi(x-k) = 1$ implies $\int_{-\infty}^{+\infty} dt\, \varphi(x) = 1$ (for a proof see Anastassiou and Yu (1992), but the result is straightforwardly obtained by swapping integral and summation signs). In particular, this means that both $\varphi(x)$ and $\tilde{\varphi}(x)$ are normalized in the L_1 norm.

The following notation is also used:

$$\varsigma_{\varphi,\tilde{\varphi}}(t) = \int_{-\infty}^{+\infty} d\tau\, \tau\tilde{\varphi}(\tau) - \sum_{k=-\infty}^{+\infty} (t-k)\varphi(t-k). \tag{2.9}$$

Let us examine why these assumptions guarantee a shape-preserving approximation (2.2) of a pdf and of a cdf, respectively. Assumption (2.3) on φ and $\tilde{\varphi}$ ensures that the reconstruction (2.2) of both pdf and cdf is non-negative. If Assumption (2.6) is also satisfied by $\tilde{\varphi}$, then the approximation of a pdf integrates to 1. Assumptions (2.4) and (2.7) are specific to cdf shape preserving approximation. The former guarantees that the reconstruction has the minimum regularity conditions of a cdf (bounded and right continuous), while the latter, jointly with the non-negativity of $\tilde{\varphi}$, guarantees the monotonicity of the reconstruction.

Assumption (2.5) is a usual compact support assumption which amounts to using finite length filters in the implementation of the discrete wavelet transform and implies that

$$\nu_a = \#\{\varphi_{jk} \mid x \in \mathrm{Supp}\,(\varphi_{jk})\} \tag{2.10}$$

is independent of scale j and location k. Assumptions (2.6) and (2.9) together with the condition $\varsigma_{\varphi,\tilde{\varphi}} = 0$ *a.e.* are equivalent to usual moment conditions in wavelet approximation theory, defining the moments of the dual scaling function:

$$\tilde{\mathscr{M}}_0 = \int_{-\infty}^{+\infty} dt\, \tilde{\varphi}(t), \qquad \tilde{\mathscr{M}}_1 = \int_{-\infty}^{+\infty} dt\, t\, \tilde{\varphi}(t).$$

Then (2.6) and condition $\varsigma_{\varphi,\tilde{\varphi}} = 0$ can be rewritten in the following way:

$$\sum_{k=-\infty}^{+\infty} (t-k)^p\, \varphi(t-k) = \tilde{\mathscr{M}}_p, \quad p = 0, 1\ . \tag{2.11}$$

These two conditions ensure that the multiresolution analysis V_j reproduces exactly polynomials of degree less or equal to 1, or to say differently, fulfills a *Strang-Fix* condition of order 1.

We finish this preparatory section by recalling the useful concept of the *modulus of smoothness*. This concept is used later on to derive the approximation properties of the projection operator (2.2) in general function spaces, such as Sobolev and Besov spaces (for which we refer to Nikol'skiĭ (1975)). For functions defined on a region $\Omega \in \mathbb{R}^d$, we introduce the increment of the function f in the direction $\mathbf{i}$ and the corresponding modulus of smoothness:

$$\Delta^1_{\mathbf{i}t} f(\mathbf{x}) = f(\mathbf{x} + \mathbf{i}t) - f(\mathbf{x}), \qquad \Delta^\mu_{\mathbf{i}t} f(\mathbf{x}) = \Delta^1(\Delta^{\mu-1}_{\mathbf{i}t} f(\mathbf{x}))\ ,$$

then, for $h > 0$, $\mu \in \mathbb{N}$ and $1 \leqslant p \leqslant \infty$, the integral p-moduli of smoothness in the $\mathbf{i}$ direction is given by

$$\omega^\mu_{\mathbf{i}}(f,h)_p = \sup_{0<t\leqslant h} \left\| \Delta^\mu_{\mathbf{i}t} f(\mathbf{x}) \right\|_p \tag{2.12}$$

with the usual convention of the $\sup_x$-norm for $p = \infty$, which is the classical modulus of continuity.

3 Shape preserving estimation of univariate probabilistic functions

First we briefly summarize the main concepts of DP in the *i.i.d.* case of estimating *univariate* probabilistic functions (pdf and cdf) by means of shape preserving wavelets. We recall these results since they have inspired our own work for multivariate dependent data. Note that we need to define an estimation risk in a function norm which is appropriate for treating simultaneously the error when estimating a pdf or a cdf nonparametrically (see Equation (3.5)). In the two cases the inner products in Equation (2.2) can be estimated from the observed data $(X_1, \ldots, X_n)$ in the following way:

$$\widehat{\langle f, \tilde{\varphi}_{jk} \rangle} = \langle f, \tilde{\varphi}_{jk}(X) \rangle = \frac{1}{n} \sum_{i=1}^{n} \tilde{\varphi}_{jk}(X_i), \qquad \text{if } f \text{ is a pdf;} \tag{3.1}$$

$$\widehat{\langle F, \tilde{\varphi}_{jk} \rangle} = \langle F, \tilde{\varphi}_{jk}(X) \rangle = \frac{1}{n} \sum_{i=1}^{n} 2^{-\frac{j}{2}} \left\{1 - \tilde{\Phi}(2^j X_i - k)\right\}, \qquad \text{if } F \text{ is a cdf,} \tag{3.2}$$

with $\tilde{\Phi}(x) = \int_{-\infty}^{x} \tilde{\varphi}(t)\, dt$. Since f is a pdf, $\langle f, \tilde{\varphi} \rangle = \mathbb{E}[\tilde{\varphi}]$. Then (3.1) is an estimator of the expected value of $\tilde{\varphi}$. In a similar way we can obtain (3.2) by integration by parts

$$\langle F, \tilde{\varphi}_{jk} \rangle \;=\; 2^{-j/2} - 2^{-j/2}\mathbb{E}[\Phi(2^j X - k)],$$

using the boundness of the support and the normalization properties of $\tilde{\varphi}_{jk}$. It then follows that the estimators for the univariate pdf and cdf will be given by:

$$\hat{f}(x) = \hat{A}_j^{(n)}\big(f\big)(x) = \frac{1}{n} \sum_{k \in \mathbb{Z}} \sum_{i=1}^{n} \tilde{\varphi}_{jk}(X_i) \varphi_{jk}(x), \qquad \text{if } f \text{ is a pdf;} \tag{3.3}$$

$$\hat{F}(x) = \hat{A}_j^{(n)}\big(F\big)(x) = \frac{1}{n} \sum_{k \in \mathbb{Z}} \sum_{i=1}^{n} 2^{-\frac{j}{2}} \big\{1 - \tilde{\Phi}(2^j X_i - k)\big\} \varphi_{jk}(x), \qquad \text{if } F \text{ is a cdf.} \tag{3.4}$$

Lemma 1. *Let f be either a pdf or a cdf. Let $\varphi, \tilde{\varphi}$ fulfill Assumptions* (2.3) *to* (2.8) *and, if f is a pdf, let $\tilde{\varphi}$ fulfill also* (2.6). *Then the estimator $\hat{A}_j\big(f\big)$ derived from the operator* (2.2) *using* (3.4) *or* (3.3) *is shape preserving.*

By shape preserving, we mean that if f is a pdf, then $A_j\big(f\big)$ is a non-negative function that integrates to 1, and that if F is a cdf, then $A_j\big(F\big)$ is a monotone, right-continuous function and $\lim_{x\to\pm\infty} A_j\big(F\big)(x) = 0, 1$. For the proofs we refer to Lemma 2.2.1 in Dechevsky and Penev (1997) for the pdf case, and Lemma 2.1.1 in Dechevsky and Penev (1997) and Lemma 3 in Anastassiou and Yu (1992) for the cdf case.

To assess the behavior of the estimators we define a risk using the following quasi-norm for functions $g(x)$ defined on $\mathbb{R}$, taking random variable values which depend on the realization of $(X_1, \ldots, X_n)$:

$$\big\|\, g \,\big|L_p(\mathscr{L}_q)\big\| = \left\{ \int_{-\infty}^{+\infty} dx \left(\mathbb{E}\,|g(x)|^q\right)^{p/q} \right\}^{1/p},$$

with $0 < p, q \leqslant \infty$. Recall that for a quasi-norm the triangular inequality holds with $\big\|\, g + h \big\|_A \leqslant c_A(\big\|\, g \big\|_A + \big\|\, h \big\|_A)$, $c_A \geqslant 1$. In order to be able to work with the usual triangular inequality, i.e. $c_A = 1$, we move to the space $L_p(\mathscr{L}_q)^\rho$ (defined in Appendix E) with an appropriately chosen $\rho > 0$ (see again the discussion in Appendix E).

In the $L_p(\mathscr{L}_q)$ quasi-norm the p parameter takes into account, via (2.12), the smoothness of the function to be estimated, while the q parameter allows an additional degree of freedom when imposing conditions on the tails of the density for the estimation risk to be finite. (Note that in the original work of Dechevsky and Penev (1998) the role of the two parameters is inverted, and that in contrast to usual Besov spaces, q is here connected to the stochastic dimension of the problem). For $p = q$, we get the usual L_p-risk, i.e. $E\| \cdot \|_p$. The notation $\| \cdot \|_p$ continues to be associated with the usual $L_p(\mathbb{R}^d)$-norm.

For now on let $\hat{f}$ be generically an estimator for the density or for the cumulative. It follows from the discussion above that

$$\begin{aligned} \|\, \hat{f} - f \,|L_p(\mathscr{L}_q)\|^\rho &= \|\, \hat{f} - \mathbb{E}(\hat{f}) + \mathbb{E}(\hat{f}) - f \,|L_p(\mathscr{L}_q)\|^\rho \\ &\leqslant c(p,q,\rho)\Big\{ \|\, \hat{f} - \mathbb{E}(\hat{f}) \,|L_p(\mathscr{L}_q)\|^\rho + \|\, \mathbb{E}(\hat{f}) - f \,|L_p(\mathscr{L}_q)\|^\rho \Big\} \,, \end{aligned} \tag{3.5}$$

where in the sequel we will restrict ourselves for the range of parameters $1 \leqslant p \leqslant \infty,\ 0 < q \leqslant 2$, and we will always work with a choice of ρ such that $c(p,q,\rho) = 1$.

From Equations (3.1) and (3.2), we can easily see that the estimators (3.1) and (3.2) are unbiased estimators of the inner products $\langle f, \tilde{\varphi}\rangle$, and that also $\hat{A}_j$ is unbiased for A_j. In total the triangular inequality (3.5) can be rewritten as:

$$\|\, \hat{f} - f \,|L_p(\mathscr{L}_q)\|^\rho \leqslant \Big\{ \|\, A_j(f)(\cdot) - f(\cdot)\|_p^\rho + \|\, \hat{A}_j^{(n)}(f)(\cdot) - A_j(f)(\cdot) \,|L_p(\mathscr{L}_q)\|^\rho \Big\}. \tag{3.6}$$

The first result (Dechevsky and Penev (1997), Theorem 2.1.1) treats the first part of Equation (3.6) which concerns the "bias" of estimating either a cdf or a pdf denoted by f:

$$\|\, A_j(f)(\cdot) - f(\cdot)\|_p \leqslant c_1 \|\varsigma_{\varphi,\tilde{\varphi}}\|_\infty \cdot \omega^1(f, 2^{1-j}a)_p + c_2 \|\tilde{\varphi}\|_{p'} \cdot \|\varphi\|_\infty \cdot \omega^2(f, 2^{1-j}a)_p, \tag{3.7}$$

where $p' \in [1,\infty]$ is such that $\frac{1}{p} + \frac{1}{p'} = 1$, a is the length of the support of the scaling functions, and $c_1 > 0$ and $c_2 > 0$ are absolute constants that do not depend on the resolution level j.

Note that, in contrast to densities, cdf are not in L_p with $1 \leqslant p < \infty$, but only in L_∞. Yet by (3.7) the L_p distance between $A_j(F)$ and F is bounded. This means that the approximations properties of $A_j(F)$ can be studied in an appropriately chosen L_p norm.

An explicit expression of the dependence of the bias on the resolution level j can be obtained by specifying the function space to which f belongs and by taking advantage of the properties of the modulus of smoothness of functions belonging to that specific space. Since, by definition, $\omega^\mu(f, 2^{1-j}a)_p$ is increasing in the argument $2^{1-j}a$, both for pdf and cdf the bias can be bounded by a *decreasing* function of j, so that Equation (3.7) can be rewritten as:

$$\|\, A_j(f)(\cdot) - f(\cdot)\|_p \leqslant B(j), \tag{3.8}$$

where $B(j)$ is a decreasing function of j.

For the second term of (3.5), which is rewritten as

$$\Big\|\, \hat{A}_j^{(n)}(f)(\cdot) - \mathbb{E}(\hat{A}_j^{(n)}(f)(\cdot)) \Big| L_p(\mathscr{L}_q) \Big\| = \left\{ \int_{-\infty}^{+\infty} dx \left(\mathbb{E}|\hat{A}_j^{(n)}(f)(x) - \mathbb{E}(\hat{A}_j^{(n)}(f)(x)|^q \right)^{\frac{p}{q}} \right\}^{\frac{1}{p}} , \tag{3.9}$$

we have two different behaviors depending on f being a cdf or a pdf.

For cdf. (Theorem 2.1.1 in Dechevsky and Penev (1998)) The variance term (3.9) has a parametric decay $O\left((\frac{1}{n})^{\rho/2}\right)$ to zero. In this case the bias rate can be adapted by opportunely choosing the increasing function $j^* = j^*(n)$ such that $B(j^*) = O(n^{-\rho/2})$ and the total risk, i.e. the rate of convergence of (3.6), decays at a parametric rate:

$$\|\, \hat{A}_j(F)(\cdot) - F(\cdot) \,|L_p(\mathscr{L}_q)\|^\rho = O\left(\left(\frac{1}{n}\right)^{\rho/2}\right), \qquad n \to \infty\ . \tag{3.10}$$

It can be easily seen that the above convergence rate can be achieved by choosing $j \geqslant (p/2)\log_2 n$.

For pdf. (Theorem 2.2.1 in Dechevsky and Penev (1998)) The variance term is an increasing function of j, that is (3.9) is bounded by a function $V(2^j/n) = O\big((2^j/n)^\rho\big)$. When choosing the function $j^* = j^*(n)$, we face the typical nonparametric trade-off between the competing behaviors of B and V as functions of j. In particular, $j^* = j^*(n)$ has to be an increasing function such that $V(2^{j^*}/n) = O\big(B(j^*)\big)$. The convergence rate we obtain for the estimator of pdf is of order

$$\| \hat{A}_j(f)(\cdot) - F(\cdot) \,|L_p(\mathscr{L}_q)\|^\rho = O\ (2^{-j^*(n)\rho}), \qquad n \to \infty\ . \tag{3.11}$$

which is typically slower than the parametric one. These findings are the same as the ones in classical wavelet estimation (Härdle et al. (1998), Kerkyacharian and Picard (1992)). For detailed results, we refer to Appendix A, Corollaries 10 - 14 .

4 Shape preserving estimation of multivariate probabilistic functions

This is the main section of our paper. Here we provide our original contribution, namely a multivariate extension of the results of DP summarized in the previous section. Our particular interest, partially motivated from application to quantile estimation (see Section 5), lies in a situation where the multivariate function $F \in \mathbb{R}^d$ is actually a cdf in the first argument, and a $d-1$-variate pdf in the remaining arguments, i.e.

$$F_Y(\mathbf{x}, y) = \int_{-\infty}^{y} dt\, f(t, \mathbf{x}). \tag{4.1}$$

However, our set-up allows for constructing estimators of multivariate densities $f(\mathbf{x}) \in \mathbb{R}^d$ as well. For ease of notation we present only the bivariate case and this in the specific cdf-pdf case, since this set-up is the most innovative in our work. From now on the argument y will always denote the variable with respect to which $F_Y(x,y)$ is cumulated, and x the argument with respect to which $F_Y(x,y)$ is a density. The extension to higher dimensions is immediate. On the other hand, the extension of the results on the bivariate pdf-cdf to a purely bivariate pdf can be obtained with minor changes that will be given in the sequel.

Our constructions are based on *tensor product* wavelets, a concept that we briefly recall now. The primal and dual wavelet bases $\varphi, \tilde{\varphi}$ introduced in Equations (2.3) - (2.8) are functions defined from $\mathbb{R}$ to $\mathbb{R}$ so that functions $f : \mathbb{R} \to \mathbb{R}$ can be approximated. It is straightforward to build scaling functions defined on $\mathbb{R}^d$, so that multivariate functions can be approximated by wavelet series, using *tensor product* wavelets. It is known (see, e.g., Meyer (1992) Section 3.3) that, if $\{V_J\}_{J\in\mathbb{Z}}$ is a multiresolution analysis of $L_2(\mathbb{R})$, then:

$$L_2(\mathbb{R}^2) = \overline{\bigcup_{j_1,j_2=0}^{\infty} V_{j_1} \otimes V_{j_2}}.$$

Now we define $J = (j_1, j_2)$ and $V_J = V_{j_1} \otimes V_{j_2}$, from which we can derive the 2-d basis for each of these approximation spaces V_J:

$$\{\varphi_{J\mathbf{k}}(x,y)\}_{\mathbf{k}\in\mathbb{R}^2} = \{\varphi_{j_1k_1}(x)\}_{k_1\in\mathbb{Z}} \otimes \{\varphi_{j_2k_2}(y)\}_{k_2\in\mathbb{Z}}\ . \tag{4.2}$$

Note that this construction can obviously be extended to any dimension $d > 2$.

In the sequel we provide approximation and estimation results of bivariate probability functions treating both the i.i.d. and the dependent data (time series) case. To emphasize on the latter one (including the first one), we suppose that we have real-valued bivariate time series observations $(Y_1, X_1), \ldots, (Y_T, X_T)$, generated from a stationary stochastic process $\{(Y_t, X_t)\}_{t\in\mathbb{Z}}$, whose dependence structure is controlled via mixing conditions. In particular we could have that $Y_t = Z_t$ and $X_t = Z_{t-1}$, where $\{Z_t\}_{t\in\mathbb{Z}}$ is a univariate stationary process. Let $\mathscr{F}_i^k$ be the sigma-field of events generated by the random variables $\{(Y_t, X_t), i \leqslant t \leqslant k\}$. The stationary process $\{(Y_t, X_t)\}_{t\in\mathbb{Z}}$ is called strongly or α- mixing if

$$\sup_{\substack{A\in\mathscr{F}^0_{-\infty}\\ B\in\mathscr{F}^\infty_p}} \big|P[AB] - P[A]P[B]\big| = \alpha(p) \xrightarrow[p\to\infty]{} 0 \; .$$

Below, if not differently stated, let $f(\cdot)$ be the "design" density $f_{X_t}(x)$, which is the marginal distribution of the stationary process in the pure time series case.

Our results are then derived under the following assumptions on the stochastic process.

Assumption 1: For every integer $s > 0$ there exist the joint distribution $F_{(X_0,Y_0),(X_s,Y_s)}$ and a positive constant M such that for every bounded zero-mean random variable $T(X_t, Y_t)$:

$$\mathbb{E}[|T(X_0, Y_0)\; T(X_s, Y_s)|] \leqslant M\; \mathbb{E}[|T(X_0, Y_0)|]\; \mathbb{E}[|T(X_s, Y_s)|] \; . \tag{4.3}$$

Assumption 2: The process $\{(X_t, Y_t)\}$ is α-mixing and the coefficients $\alpha(p)$ are such that:

$$\sum_{p=N}^{\infty} \big[\alpha(p)\big]^{1-2/r} = O(N^{-1}), \tag{4.4}$$

for some $r > 2$.

Many processes verify the condition given on the mixing coefficients. Gaussian processes, non Gaussian autoregressive moving average processes (see Pham and Tran (1980)), many nonlinear functionals of these processes, and various GARCH and stochastic volatility models, see Carrasco and Chen (2002).

We construct now our estimators in complete analogy of the univariate constructions (3.3) and (3.4), using the shape preserving scaling functions φ and $\tilde\varphi$ that fulfill Assumptions (2.3) to (2.8). Recall that $J = (j_1, j_2)$. Let $\{\varphi_{J\mathbf{k}}(x,y)\}_{\mathbf{k}\in\mathbb{R}^2}$, $\{\tilde\varphi_{J\mathbf{k}}(x,y)\}_{\mathbf{k}\in\mathbb{R}^2}$ be the bivariate primal and dual bases built accordingly to (4.2), then the estimators of $F_Y(x,y)$ and $f(x,y)$ are given by:

$$\hat{A}_J^{(T)}\big(f\big)(x,y) = \sum_{\mathbf{k}\in\mathbb{Z}^2} \Big\{\frac{1}{T}\sum_{t=1}^{T} \tilde\varphi_{j_1k_1}(X_t)\tilde\varphi_{j_2k_2}(Y_t)\Big\}\varphi_{J\mathbf{k}}(x,y),\;\; (x,y)\in\mathbb{R}^2, \tag{4.5}$$

$$\hat{A}_J^{(T)}\big(F\big)(x,y) = \sum_{\mathbf{k}\in\mathbb{Z}^2}\sum_{t=1}^{T} 2^{-\frac{j_2}{2}} \left(\frac{\tilde\varphi_{j_1,k_1}(X_t)}{T} - \frac{\tilde\varphi_{j_1,k_1}(X_t)\tilde\Phi(2^{j_2}Y_t - k_2)}{T}\right)\varphi_{J\mathbf{k}}(x,y), \quad (x,y)\in\mathbb{R}^2. \tag{4.6}$$

Let us further introduce the multivariate approximator of the probabilistic function f:

$$A_J\big(f\big)(x,y) = \sum_{k\in\mathbb{Z}^d} \langle f, \tilde\varphi_{J\mathbf{k}}\rangle \varphi_{J\mathbf{k}}(x,y) \; . \tag{4.7}$$

Note that the properties of the functions and of the projectors $A_J\big(f\big)(x,y)$ that allow to approximate cdf of one variable, cannot play the same role when we move to the multivariate analysis. In particular (3.7) does not continue to hold in the multivariate L_p-norm because in two dimensions it is not possible to bound the moduli of smoothness of a function which is in L_∞ only, for the y-variable. Consequently, and instead of using an $L_\infty-$norm which would urge us to assume continuity of the density part of $F_Y(x,y)$, we take advantage of the different convergence rates for pdf and cdf as discussed in Section 3, and work with the following risk:

$$d\big\{\hat f(x,y), \hat g(x,y)\big\}_p = \sup_{y\in\mathbb{R}} \big\|\, \hat f(\cdot,y) - \hat g(\cdot,y)\,\big|L_p(\mathscr{L}_q)\big\| \; . \tag{4.8}$$

Note that in order to be able to use this norm, we have to assume continuity of $F_Y(x,y)$ as a function of y. In the framework just introduced, we will see that the results in the approximation and estimation of the bivariate $F_Y(x,y)$ will look like the usual results on estimation of a univariate pdf, and no additional theoretical effort is needed to interpret them. Moreover, these results on the bivariate pdf-cdf can be adapted with minor changes to a bivariate pdf, where the norm (4.8) becomes the usual $L_p(\mathbb{R}^2)$-norm. In the latter case, we do not need to make any continuity assumption on bivariate pdf while we have to impose the continuity of $F_Y(x,y)$ with respect to the y argument. More details on the changes needed to adapt the pdf-cdf results to the pure pdf case will be given at the end of each of the following sub-sections.

4.1 Bias in multivariate cdf-pdf approximation

We now bound the deterministic bias made by approximating $F_Y(x,y)$, which will be measured in the norm (4.8). The proof can be found in Appendix B. The proof technique is largely inspired by the one in Dechevsky and Penev (1997), and changes have been made in order to adapt it to the norm (4.8). As the proof is based on approximations by Steklov means, allowing to relate the approximation error to the modulus of smoothness, we refer to Appendix D for a definition and relevant properties of Steklov means. Recall the definitions of $\varsigma_{\varphi\tilde{\varphi}}$ in (2.9). Further let $\mathbf{e}_x$ or $\mathbf{e}_y$ be the unit vector in the x and y directions.

Lemma 2. *Let Assumptions* (2.3) *to* (2.8) *hold. Let $F(x,y)$ be, for a fixed x, a continuous function in y. Then, for $1 \leqslant p \leqslant \infty$:*

$$\begin{aligned}
&\sup_y \|A_J\big(F\big)(\cdot,y) - F(\cdot,y)\|_p \\
&\qquad\leqslant c_1 \sup_y \omega^1_{\mathbf{e}_x}\big(F(\cdot,y), 2^{1-j_1}a\big)_p + c_2 \sup_y \sup_{0<\delta<1} \big\|\Delta_{\mathbf{e}_y\delta(2^{1-j_2}a)} F(\cdot,y)\big\|_p \\
&\qquad\quad + c_3 \sup_y \omega^2_{\mathbf{e}_x}\big(F(\cdot,y), 2^{1-j_1}a\big)_p + c_4 \sup_y \sup_{0<\delta<1} \big\|\Delta^2_{\mathbf{e}_y\delta(2^{1-j_2}a)} F(\cdot,y)\big\|_p.
\end{aligned} \tag{4.9}$$

The constants c_1 and c_2 include the value of the norms $\|\varsigma_{\varphi\tilde{\varphi}}(2^{j_1}x)\|_\infty$ and $\|\varsigma_{\varphi\tilde{\varphi}}(2^{j_2}y)\|_\infty$. If the latter two L_∞-norms are 0*, then the constants c_1 and c_2 will be smaller but they cannot be equal to* 0*.*

We remark in passing that, differently from Equation (3.7) of the univariate setting, in (4.9) the first order terms, i.e. the terms in which appear the constants c_1 and c_2, apparently do not vanish even if the conditions $\|\varsigma_{\varphi\tilde{\varphi}}(2^{j_1}x)\|_\infty = 0$ and $\|\varsigma_{\varphi\tilde{\varphi}}(2^{j_2}y)\|_\infty = 0$ are fulfilled. This should rather be related to the proof technique and not to some fundamental reason.

Remark 1. If it is a bivariate pdf to be approximated, then in Lemma 2 we do not need to assume continuity of $f(x,y)$ with respect to y. In (3.7) the L_p-norm is taken with respect to both arguments and the bound is given by the usual moduli of smoothness in the two directions, thus obtaining the bivariate equivalent of (3.7).

4.2 Estimation of multivariate cdf-pdf estimation from dependent data

We come now to our main results where we study the estimation of the bivariate $F_Y(x,y)$ from a set of observations $\{(X_1,Y_1),\dots,(X_T,Y_T)\}$, generated from a time series under Assumptions 1 and 2. The case where (X_i,Y_i) are i.i.d. will be contained as a subcase of this one.

Again we make use of the triangular inequality to split the risk in a stochastic term and in a bias term:

$$\begin{aligned}
&\sup_{y\in\mathbb{R}} \big\|\, \hat{A}^{(T)}_J\big(F\big)(\cdot,y) - F(\cdot,y)\,|L_p(\mathscr{L}_q)\big\|^\rho \\
&\quad\leqslant \Big\{ \sup_y \big\|A_J\big(F\big)(\cdot,y) - F(\cdot,y)\big\|^\rho_p + \sup_y \big\|\, \hat{A}^{(T)}_J\big(F\big)(\cdot,y) - A_J\big(F\big)(\cdot,y)\,|L_p(\mathscr{L}_q)\big\|^\rho \Big\}.
\end{aligned}$$

The bias part has been bounded in Lemma 2. We now give an estimate from above of the stochastic part of the risk of the estimator.

Lemma 3. *Let $\varphi, \tilde{\varphi}$ be as in* (2.3) - (2.8)*. Let $\{(X_t, Y_t)\}_{t=0,\dots,T}$ be realizations of a stationary process fulfilling Assumptions 1 and 2. Let $p \geqslant 1$, $0 < q \leqslant 2$ and $\rho = \min\{2, p\}$. Assume, for fixed y, that $F_Y(x, y) \in L_{p/2} \cap L_1 \cap L_{p/r}$ for some $r > 2$. Let $j_2 \geqslant (p/2)\log_2 T$. Then*

$$\sup_y \Big\| \hat{A}_J^{(T)}(F)(\cdot, y) - \mathbb{E}\big[\hat{A}_J^{(T)}(F)(\cdot, y)\big] \Big| L_p(\mathscr{L}_q) \Big\|^\rho \leqslant V_0 + V_1, \tag{4.10}$$

where

$$\begin{aligned} V_0 = \left(\frac{2^{j_1}}{T}\right)^{\rho/2} & \Bigg[d_1 \max\Big\{\sup_y \|F(\cdot, y)\|_1, \sup_y \|F(\cdot, y)\|_{p/2}\Big\}^{\rho/2} \\ + \qquad & d_2(a) \max\Big\{\sup_y \omega^1_{\mathbf{e}_x}\big(F(\cdot, y), 2^{1-j_1}a\big)_1, \sup_y \omega^1_{\mathbf{e}_x}\big(F(\cdot, y), 2^{1-j_1}a\big)_{p/2}\Big\}^{\rho/2}\Bigg] \\ & + O\Big(T^{-\rho/2}\Big); \end{aligned} \tag{4.11}$$

$$\frac{V_1}{V_0} \longrightarrow 0, \quad \text{as } \ T \to \infty \ . \tag{4.12}$$

Moreover, we have

$$\max\{\sup_y \ \omega^1(F(\cdot, y), 2^{1-j_1}a)_1, \sup_y \ \omega^1(F(\cdot, y), 2^{1-j_1}a)_{p/2}\} = o(1), \quad j_1 \to \infty,$$

if $2 \leqslant p < \infty$,
or if $p = \infty$ *and* F *is continuous also with respect to the argument* x,
or if $1 < p < 2$ *and* $\sup_y \sup_{0 \leqslant t \leqslant h} \int_{-\infty}^{+\infty} dx \Big(\int_0^1 d\alpha \, F(x + \alpha t, y)\Big)^{p/2} < \infty.$

Here d_1 and $d_2(a)$ are absolute constants that do not depend on the resolution levels $J = (j_1, j_2)$, and a is the support of $\varphi, \tilde{\varphi}$. As it can be seen from Equation (4.10) the stochastic component of risk has two contributions V_0 and V_1 (for an explicit form of the latter one we refer to the proof). V_0 is the only variance term we would obtain if $\hat{F}_Y(x, y)$ were estimated from i.i.d. data.

In Appendix F we provide a slightly more general expression for the variance term in the i.i.d. set-up, in which we can allow for a parameter range $0 < q < \infty$.

A close inspection of (4.11) indicates that only the resolution parameter j_1 appears. Formally the variance term obtained in Lemma 3 is equivalent to the one we would obtain for a univariate pdf, see Appendix A. This is because, as remarked in the discussion following Equations (3.10) and (3.11), the cdf like component of the variance has a faster convergence, and with the choice $j_2 \geqslant (p/2)\log_2 T$ the convergence of the entire y component of the stochastic error is taken into account by the $O(T^{-\rho/2})$ term . We can finally put together the results of Lemmas 2 and 3 to obtain the convergence rates of $\hat{A}_j^T$.

Theorem 1. *Let the assumptions of Lemmas 2 and 3 hold. Then the total risk of the estimator $\hat{A}_j^T$ of $F_Y(x, y)$ is:*

$$\begin{aligned} &\sup_{y \in \mathbb{R}} \big\| \hat{A}_J^{(T)}(F)(\cdot, y) - f(\cdot, y) \,\big| L_p(\mathscr{L}_q) \big\|^\rho \\ &\leqslant c_1 \sup_y \omega^1_{\mathbf{e}_x}\big(F(\cdot, y), 2^{1-j_1}a\big)^\rho_p + c_3 \sup_y \omega^2_{\mathbf{e}_x}\big(F(\cdot, y), 2^{1-j_1}a\big)^\rho_p + \left(\frac{2^{j_1}}{T}\right)^{\rho/2} \Big\{C_1 + o(1)\Big\} \\ &+ O(T^{-\rho/2}) + o(1). \end{aligned} \tag{4.13}$$

The constants c_1, c_3 and C_1, can be made explicit by comparing (4.13) with (4.9) and (4.11). Again we remark that only the j_1 parameter appears. The increments $\|\Delta^i_{\mathbf{e}_y} f\|$ of the function $F_Y(x, y)$ in the direction y that can be found in (4.9) are missing in (4.13), since they are taken into account in the $O(T^{-\rho/2})$ term once the $j_2 \geqslant (p/2)\log_2 T$ choice is made. The $o(1)$ term takes into account the V_1 component of the variance. Since the risk given in Theorem 1 is formally equivalent to the one we would obtain in the estimation of a univariate pdf, the explicit convergence rates of $\hat{A}_J(f)$ can be obtained again by choosing a resolution level j_1 in the x direction that balances the bias and variance competing behaviors, and will be equivalent to the convergence rates of the estimator of a univariate pdf having the same smoothness as the pdf part of F_Y. A corollary on the rates of convergence of our estimator in the bivariate case is then immediate and is similar to Corollaries 10-14 in Appendix A.

Remark 2. Lemma 3 can be extended to a bivariate pdf by relaxing the continuity assumption on the y variable and by asking for $f(x, y) \in L_{p/2} \cap L_1 \cap L_{p/r}$. Then in (4.11) we would have a $(2^{j_1+j_2}/T)$ factor instead of $(2^{j_1}/T)$, and the $O(T^{-\rho/2})$ term would disappear.

Theorem 3 and the proofs deal with the bivariate distribution function $F_Y(x, y)$, but a close look at the proof shows that the results can be immediately extended to dimension $d \geqslant 2$. As a multivariate density has finite L_p-module of smoothness for some $p < \infty$, we can extend Theorem 1 to functions $F_Y(\mathbf{x}, y)$, by looking at the uniform convergence in y of

$$\sup_y \left\| \left| \hat{A}_J^{(T)}(f)(\cdot, y) - \mathbb{E}\hat{A}_J^{(T)}(F)(\cdot, y) \right| L_p(\mathscr{L}_q) \right\|$$

where $\|\cdot\|_p$ would now be the $L_p(\mathbb{R}^{d-1})$-norm.

Let us briefly comment on the results obtained for the convergence of the distribution function $F_Y(x, y)$. We derived the upper bound and asymptotic behavior of the stochastic term of the risk of the estimator (4.6) when the data come from a weakly dependent process. The risk is computed in the $L_p(\mathscr{L}_q)$-norm. The advantage of separating the pointwise expectation from the global norm is probably more evident here than in other contexts. This can be seen by comparing our results with the ones in Masry (1994). First of all Masry computes the risk in the Sobolev W_2^s norm, while starting from the $L_p(\mathscr{L}_q)$ risk we can specify the risk in a variety of different norms, especially in the norm of the Besov spaces built starting from L_p. But the main difference concerns the conditions that have to be imposed on the tails of the density for the risk not to explode. While we simply need to impose that $f \in L_{p/r}, r > 2$, in Masry (1994) the decay of the density tails has to be related to the smoothness s of the density, asking for a decay of order $x^{-\beta}$, with $\beta > 0.5 + s$.

5 Statistical applications: quantile regression and estimation of conditional quantiles of financial data

As important application of our methodology, we address the estimation of conditional quantiles. Let us consider the stationary process $\{(X_t, Y_t)\}$ introduced in Section 4. We wish to estimate the conditional distribution function $F_Y(y|x) = P(Y_t \leqslant y|X_t = x)$ and, from this, the p-th conditional quantile, that is the value $Q(x, p) = \inf\{y \in \mathbb{R}|F(y|x) \geqslant p\}$, for a probability level $0 < p < 1$. The conditional median $Q(x, p = 0.5)$ has often been object of interest. It is used as an alternative to the conditional mean to deliver a robust estimate of the effect of the variable X on the response variable Y. In general given a collection of conditional quantiles we can build confidence intervals for the variable Y. In the case of a stationary process $\{Z_t\}_{t\in\mathbb{Z}}$, when $Y_t = Z_t$ and $X_t = Z_{t-1}$, it is possible to build prediction intervals for Z_t having observed Z_{t-1}. The estimation of conditional quantiles of financial time series partially motivates this work. In this section we start by linking the problem of the estimation of conditional quantiles to the one of the estimation of the bivariate

$F_Y(x,y)$ studied in the previous sections. In order to achieve this, we show that a modification of the norm (4.8) is needed. We then provide a couple of scaling functions $\varphi, \tilde{\varphi}$ fulfilling Assumptions (2.3) to (2.8) and (2.11), and we build explicitly shape preserving estimators, thus providing an algorithm for the implementation of the shape preserving set-up. Finally we carry out a Monte Carlo experiment to test our methodology.

5.1 Set-up and conditional quantiles

We retain the same assumptions on the stochastic process as in Section 4. In particular we consider observations $\{Y_1,\ldots,Y_T\}$ of a stationary process $\{Y_t\}_{t\in\mathbb{Z}}$ such that the couples $\{(Y_t, X_t = Y_{t-1})\}_{t\in\mathbb{Z}}$ form a Markovian process of order one fulfilling Assumption 1 and the mixing conditions of Assumption 2. Moreover we assume that the conditional distributions $F_{Y_t}(y|x)$ and $f_{Y_t}(y|x)$ of Y_t given $X_t = x$ exist. Here we are interested in the p-th conditional quantile, which is assumed to be unique. We define the estimator $\hat{Q}(x,p)$ to be such that

$$\hat{Q}(p,x) = \inf\big\{y\in\mathbb{R}\,\big|\,\hat{F}(y|x)\geqslant p\big\}. \tag{5.1}$$

We remark that the solution of (5.1) always exists since $\hat{F}(y|x)$ is monotone and bounded between 0 and 1.

Now, since we assumed the existence of the conditional pdf $f(y|x)$, and that by definition $F(y|x)$ is its integral, we can write a Taylor expansion with integral remainder:

$$\begin{aligned} &F(\hat{Q}(p,x)|x) - F(Q(p,x)|x) \\ &\qquad = \big(\hat{Q}(p,x) - Q(p,x)\big)\int_0^1 d\theta f\big(Q(p,x) + \theta(\hat{Q}(p,x) - Q(p,x))\big|\,x\big). \end{aligned} \tag{5.2}$$

Denote $\tilde{f}_{\hat{Q},Q}(x) \doteq \int_0^1 d\theta f\big(Q(p,x) + \theta(\hat{Q}(p,x) - Q(p,x))\big|\,x\big)$.

In order to invert (5.2) we additionally assume some local lower bound on the conditional densities such as the existence of a positive c such that $f(y|x) \geq c > 0, \forall y \in |y - Q(p,x)| < \varepsilon_\delta$. Here ε_δ is chosen as a function of the convergence of $\hat{F}$ to F (see Theorem 1) implying that for T sufficiently large, $\hat{F}(\hat{Q}(p,x)|x)$ is in a $\delta-$ neighborhood of $F(\hat{Q}(p,x)|x)$, and by a (uniform) continuity argument, $\hat{Q}(p,x)$ is in a $\varepsilon_\delta-$ neighborhood of $Q(p,x)$.

Remarking now that we can express $F(y|x) = F_Y(x,y)/f(x)$, and that, by definition, $\hat{F}(\hat{Q}(p,x)|x) = p = F(Q(p,x)|\,x)$, we can write

$$\hat{Q}(p,x) - Q(p,x) = \frac{1}{\tilde{f}_{\hat{Q},Q}(x)}\left\{\frac{F_Y(x,\hat{Q}(p,x))}{f(x)} - \frac{\hat{F}_Y(x,\hat{Q}(p,x))}{\hat{f}(x)}\right\}. \tag{5.3}$$

It is not possible to evaluate the right-hand side of (5.3) with the norm (4.8), since the densities in the denominator pose a measurability problem upon integration with respect to the x variable. We then restrict our attention to studying the convergence of the conditional quantile in a neighborhood of the conditioning value $X = x$. It remains to determine how big the neighborhood of x should be. Since from Lemma 2 we know that the bound on the bias of $F_Y(x,y)$ is given by $\omega^{\mu}_{\mathbf{e}_x}\big(F(\cdot,y), 2^{1-j_1}a\big)_p$, that is a measure of the variation of F_Y in a set of radius $2^{1-j_1}a$, we compute the convergence of $\hat{Q}(x,p)$ to $Q(x,p)$ in a neighborhood of x of radius $2^{1-j_1}a$, since no improvement in the estimation error could be made by restricting ourselves to a smaller set containing x. We then state the following theorem.

Theorem 2. *Let the assumptions of Lemmas 2 and 3 hold with $\{(Y_t, X_t = Y_{t-1})\}$, $\{Y_t\}_{t\in\mathbb{Z}}$ being a stationary stochastic process. Let furthermore the marginal density $f_{X_t}(x)$ of the stationary process* *be bounded away from 0 for $x \in |x-\xi| < 2^{1-j_1}a$, where $X_t = \xi$ is value taken by the conditioning variable. Then*

$$\begin{aligned}\| \hat{Q}(p,\cdot) - Q(p,\cdot) \,|L_p(\mathscr{L}_q)\|^{\rho}_{[|x-\xi|<2^{1-j_1}a]} &\leqslant \tilde{C}(\xi)\cdot \sup_y \| \hat{A}_J^{(T)}(F)(\cdot,y) - F(\cdot,y)\,|L_p(\mathscr{L}_q)\|^{\rho}_{[|x-\xi|<2^{1-j_1}a]} \\ &\leqslant \tilde{C}(\xi)\cdot \sup_y \| \hat{A}_J^{(T)}(F)(\cdot,y) - F(\cdot,y)\,|L_p(\mathscr{L}_q)\|^{\rho}\end{aligned} \tag{5.4}$$

where $\tilde{C}$ depends on the value of the conditioning value ξ. The bound and convergence rate of the right hand side of (5.4) *are given in Theorem 1.*

5.2 Implementation and Monte Carlo experiments

In this section we implement our method on simulated time series, and compare it with a kernel based conditional quantile estimator. We start by choosing a pair of primal and dual bases $\{\varphi(\cdot - k)\}_{k\in\mathbb{Z}}$, $\{\tilde{\varphi}(\cdot - k)\}_{k\in\mathbb{Z}}$ verifying Properties (2.3) to (2.8). As primal basis we use translates of B-Splines. B-Splines of order N are piecewise polynomials of order $N-1$ defined on the N segments determined by $N+1$ distinct nodes. They are $N-1$ times differentiable at the nodes by construction. We consider a B-Spline function of order N, from now on, ${}_N\varphi(x)$, translated so that its nodes correspond to integer values, regardless of whether N is even or odd. The function ${}_N\varphi(x)$ can be characterized via its Fourier transform:

$${}_N\hat{\varphi}(\xi) = e^{-i\kappa\xi/2}\Big(\frac{\sin\xi/2}{\xi/2}\Big)^N = e^{-i\lfloor N/2\rfloor\xi}\left(\frac{1-e^{-i\xi}}{i\xi}\right)^N,$$

where $\kappa = 0$ if N is even, $\kappa = 1$ if N is odd, and $\lfloor t\rfloor$ denotes the largest integer not exceeding t. The choice of the knots implies that ${}_N\varphi(x)$ is not always symmetric. We have that:

$${}_{2N}\varphi(-x) = {}_{2N}\,\varphi(x) \quad \text{but} \quad {}_{2N+1}\varphi(1-x) = {}_{2N+1}\,\varphi(x).$$

Note that ${}_1\varphi(x)$ is the Haar scaling function. As a dual basis we use the translates of the indicator function of the support of the generator of the primal basis. That is, if ${}_N\varphi(x)$ is the B-Spline that lives on $\big[\lfloor -N/2+1\rfloor, \lfloor N/2+1\rfloor\big)$, then ${}_N\tilde{\varphi}(x) = \frac{1}{N}\mathbb{I}\big([\lfloor -N/2+1\rfloor, \lfloor N/2+1\rfloor)\big)$. Let us have a closer look at some properties of the estimators built with this choice of primal and dual basis. Consider the density estimator of Equation (4.5). The coefficients of the scaling functions ${}_N\varphi_{jk}$ are given by:

$$\frac{1}{T}\sum_{t=1}^{T} {}_N\tilde{\varphi}_{jk}(X_t) = \frac{1}{T}\cdot\{\text{number of } X_t \text{ in the support of} {}_N\varphi_{jk}\},$$

so that the estimation procedure boils down to counting the number of points that fall within the support of the scaling functions. We could think of the operator (4.5) as a special version of a smoothed histogram, even though we are actually dealing with overlapping supports. Such a density estimator is interesting because it is fast. Indeed suppose you want to estimate the density of a random variable on a grid of M points. Since the reconstructing functions φ_{jk} are known beforehand, you know the $\nu\times M$ matrix of values taken on the grid by the φ_{jk}'s, where ν is the number of overlapping basis functions on each grid point. Then the estimation consists in counting the number of data falling in the support of every φ_{jk}, then in multiplying element by element with the former $\nu\times M$ matrix, and finally in summing on the columns to get your density estimates at the M selected points. Such an algorithm is simple and quick, and there are no constraints on the number of points, unlike in Fast Fourier Transform based algorithms. Besides the wavelet estimator compares well with more traditional ones, as we will see in the next small simulation study. The simulations (for which code and datasets are available on request) are carried out with the free programming software *Ox*, see Doornik (2002).

The design of the Monte Carlo experiments is as follows. We consider a centered stationary autoregressive process of order one, whose root is equal to 0.6. The innovations are chosen to be either symmmetric or asymmetric. For the symmetric case, we draw from a Gaussian distribution with zero mean and unit variance. For the asymmetric case we draw from a skewed histogram. This histogram is plotted in Figure 1. The sample size is equal to 2000, while the number of Monte Carlo replications is equal to 1000. The kernel estimator relies on a product quartic kernel. Our estimators require the choice of four smoothing parameters, two in the pdf-like direction, h_x and j_x, and two in the cdf-like direction, h_y and j_y. As discussed after Equations (3.10) and (3.11), the crucial choice is the one of the smoothing parameters in the x direction, while the choice of the smoothing parameters in the y direction is not as critical, since there is no bias-variance trade-off for this component. The smoothing parameters in the y direction will be chosen to be smaller than all the possible choices of the parameters in the x direction.

First we opt for a "best case" framework in the sense that we select the bandwidth or resolution level which minimizes a given loss function for each Monte Carlo run. Some preliminary simulations have been made to determine suitable grids.

We have chosen to select the bandwith h_x in the grid $\{0.15, 0.20, \ldots, 0.60\} \times 2.6226$ for the symmetric case, namely 10 values, and $\{0.05, 0.07, \ldots, 0.25\} \times 2.6226$ for the asymmetric case, namely 11 values. The resolution level j_x is selected among three values $\{2, 3, 4\}$. The smoothing parameters in the y direction are chosen to be $h_y = 0.03$ and $j_y = 6$. We use two different loss functions, namely an Integrated Absolute Deviation and an Integrated Square Error. We integrate over the probability interval $[0, 1]$, and condition with respect to the simulated 2000^{th} value in predicting the conditional quantile associated with the next observation. The deviation or error is computed with respect to the true conditional quantile. Loss function values are then averaged through all runs to get a Mean Integrated Absolute Deviation (MIAD) and a Mean Integrated Square Error (MISE). The results of the simulations are reported in Table 1. As it can be seen, the

	Symmetric case		Asymmetric case	
	Kernel	Wavelet	Kernel	Wavelet
MIAD	5.90E-02	6.60E-02	4.91E-02	4.16E-02
MISE	8.36E-03	9.96E-03	5.74E-03	4.01E-03

Table 1: MIAD and MISE for the kernel and wavelet estimators when choosing the optimal bandwidth or resolution level within each run.

wavelet estimator performs better in the asymmetric case, and only slightly worse in the symmetric case. More specifically, in the symmetric case, we have a 12% increase in the MIAD and a 19% increase in the MISE, while, in the asymmetric case, we have a 18% decrease in the MIAD and a 43% decrease in the MISE - compared to the kernel method. In the asymmetric case the better performance of the wavelet estimator could have been anticipated. In the symmetric case we would have anticipated a similar performance for the two methods. We believe that the slightly better performance of the kernel estimator is here explained by the finer grid of values for the bandwidth choice compared to the choice of the resolution level of the wavelet estimator.

The first part of the simulation study has just allowed us to say that the wavelet estimator gives us a better global picture of the conditional distributions when the innovations are asymmetric and present local irregularities. However the clear advantages of the wavelet estimator are even better described by the second part of the simulation study. Here we select one bandwidth or resolution level, and maintain it fixed across all runs to check robustness of the estimation procedures with respect to the selection of the bandwidth value or the resolution level. We use the same grids as above. The results of the experiments are given in Table 2 for the symmetric case, and in Table 3 for the asymmetric case.

		Average	Min	Max	Range
MIAD	Wavelet	8.23E-02	7.27E-02	10.14E-02	2.86E-02
	Kernel	9.06E-02	7.27E-02	13.02E-02	5.75E-02
MISE	Wavelet	1.46E-02	1.27E-02	1.83E-02	0.55E-02
	Kernel	1.77E-02	1.23E-02	2.74E-02	1.51E-02

Table 2: MIAD and MISE for the kernel and wavelet estimators in the symmetric case when a fixed bandwidth or resolution level is kept fixed across all runs.

		Average	Min	Max	Range
MIAD	Wavelet	62.51E-03	58.85E-03	69.83E-03	10.98E-03
	Kernel	64.34E-03	57.73E-03	75.65E-03	17.92E-03
MISE	Wavelet	9.90E-03	9.16E-03	10.27E-03	1.11E-03
	Kernel	10.60E-03	8.66E-03	17.53E-03	8.87E-03

Table 3: MIAD and MISE for the kernel and wavelet estimators in the asymmetric case when a fixed bandwidth or resolution level is kept fixed across all runs.

From Tables 2 and 3 it can be seen than MIAD and MISE delivered by the wavelet estimator lie in a range of values unambiguously narrower than the one given by the kernel estimator, and this for both the symmetric case and the asymmetric case. These results indicate that the wavelet estimator can be considered as more robust than the kernel estimator with respect to the choice of the smoothing parameter. Putting together the results of the first and second parts of the simulation study, we argue that if we allow for a fine tuning of the bandwidth parameter, the kernel estimator can match the performance of the wavelet estimator in some cases, but it displays a much higher sensitivity with respect to the smoothing parameter. The robustness of the wavelet estimator is advantageous since it assures that the choice of the resolution parameter is not as critical, provided it is chosen within a sensible range.

We think that this robustness feature of the wavelet method can be explained by the biorthogonal basis approach used to build the estimator, see Equation (4.7). Whereas for the kernel estimator there is the classical variance-bias trade off with the quartic kernel function which becomes narrower and hence more variable as the bandwidth decreases, the wavelet estimator benefits from the fact that for increasing resolution level j_x only the primal scaling function $\varphi_{jk}(x)$ behaves as the classical kernel. However, the dual scaling function $\tilde{\varphi}_{jk}(x)$, a boxcar (Haar) function used to construct the coefficients in the reconstruction (4.7), suffers less from increasing variability on finer levels. (A boxcar function used to construct the local average assigns equal weights to all observations which tends to be numerically more stable than a local average provided by a higher order scaling function, or a kernel of higher order such as the quartic one). This is likely to temper the increase of the numerical variability for the wavelet compared to the kernel estimator.

6 Conclusion

In this paper we have further developed the DP approach of constructing shape-preserving nonparametric estimators of probabilistic functions (cdf and pdf). The wavelet methodology, taylored to reconstruct functions with low spatial regularity, has been extended to higher dimensions and

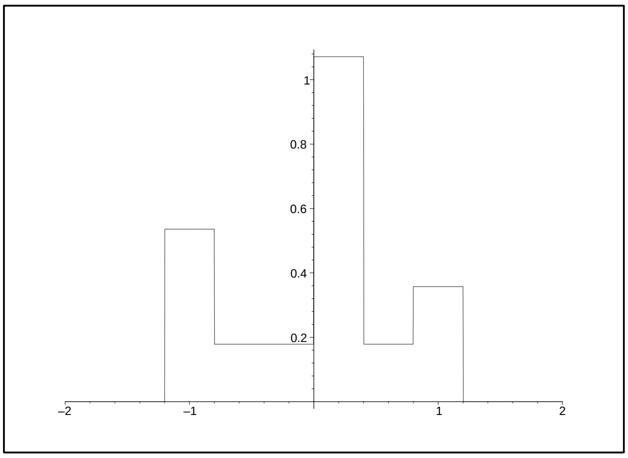

Figure 1: Distribution of the innovations for the asymmetric case.

to serially dependent (time series) data. In contrast to existing work this approach does not need to use pre- or postprocessors applied to traditional wavelet estimators in order to render, for pdf estimation, them positive and integrating to one, and, for cdf estimation, monotone. We have investigated and solved the problem of defining appropriate norms of convergence and derived rates of consistency for our estimators in these norms. We have applied our general methodology to the specific problem of conditional quantile estimation for dependent data in financial time series analysis. This has required to treat the specific situation of the intertwining of a cdf component and a pdf component in a (w.l.o.g.) bivariate nonparametric curve estimation set-up, and to face and solve the technical difficulties of this framework. Last but not least we have designed tractable algorithms relying on B-splines in that context.

Our method is still linear, and some words of comparison with both linear kernel estimation and non-linear wavelet estimation seem to be necessary here. First, our linear wavelet estimators are performing uniformly not worse than kernels, meaning that they offer advantages for some functions with local structure without losing performance in general for smoother functions. They also exhibit some robustness properties with respect to the selection of the smoothing parameter. Second, they provide a starting point for more flexible constructions which give more degrees of freedom. Indeed with our non-orthogonal wavelets (scaling functions) we have more options to adapt the construction of the estimator to the situation at hand. We have seen for instance that computing the empirical inner products boils down to counting the number of observations falling in a given interval. Further, since primal and dual bases are not so tightly related as they would be in a biorthogonal set-up, changes in the primal scaling function to obtain smoother reconstructions only require a change in the support of the indicator function used as dual basis for the moment conditions to continue to hold. Linear wavelet methods have shown practical interest in empirical analysis anyway. Lee and Hong (2001) find that even linear wavelet methods capture irregularities in the spectral density better than kernel methods.

We have to acknowledge that the real strength of wavelet estimators shows up when it comes to non-linear, i.e., threshold estimators. But, for one, it is not clear how to design a neat methodology to maintain the shape-preserving property of the resulting wavelet threshold estimator. Simply deleting the non-significant empirical wavelet coefficients at "arbitrary" locations will surely destroy this property. We believe that the "zero-tree" wavelet estimators of Shapiro (1993) could be an interesting alley for future research in that respect. This construction keeps a group of empirical wavelet coefficients at a specific location and scale together with all "coarser scale parents" at the same location over all coarser scales. This yields a kind of "locally linear" complete reconstruction structure. However, it is even not obvious how to build the wavelet functions in this specific non-orthogonal set-up. We think that one possibility is to follow the rather general device of Cohen (2003) to extend our work.

To finish we summarize again the points of methodological interest of this work which are:

1. Our method can be applied to probabilistic functions belonging to a large variety of smoothness classes, and, in particular, to specific classes of non smooth functions.

2. The wavelet estimators are shape preserving (but not shape imposing). This type of wavelets are well suited for applications in many fields of statistics in which some shape restriction is imposed to the function to be estimated. The extension of the set-up of DP (Section 3) to a multivariate setting has therefore a theoretical interest that goes beyond the application we have investigated in this work.

3. A large flexibility is permitted in choosing the wavelet bases, which implies clear computational advantages.

References

Anastassiou, G. and Yu, X. (1992). Monotone and probabilistic wavelet approximation. *Stochastic Analysis and Applications*, 10(3):251–264.

Cai, Z. (2002). Regression quantiles for time series. *Econometric Theory*, 18:169–192.

Carrasco, M. and Chen, X. (2002). Mixing and moment properties of various GARCH and stochastic volatility models. *Econometric Theory*, 18:17–39.

Cheng, M.-Y., Gasser, T., and Hall, P. (1999). Nonparametric density estimation under unimodality and monotonicity constraints. *Journal of Computational and Graphical Statistics*, 8:1–21.

Cohen, A. (2003). *Numerical Analysis of Wavelet Methods*, volume 32 of *Studies in Mathematics and its Applications*. Elsevier, North-Holland.

Cosma, A. (2004). *Nonparametric Analysis for Risk Management and Market Microstructure*. PhD thesis, Université catholique de Louvain.

Dechevsky, L. and Penev, S. (1997). On shape preserving probabilistic wavelet approximators. *Stochastic Analysis and Applications*, 15(2):187–215.

Dechevsky, L. and Penev, S. (1998). On shape preserving wavelet estimators of cumulative distribution functions and densities. *Stochastic Analysis and Applications*, 16(3):423–462.

Donoho, D. L., Johnstone, I. M., Kerkyacharian, G., and Picard, D. (1996). Density estimation by wavelet thresholding. *The Annals of Statistics*, 24:508–539.

Doornik, J. (2002). *Object-Oriented Matrix Programming Using Ox*. Timberlake Consultants Press and Oxford, London, 3rd edition.

Doukhan, P. and Léon, J. R. (1990). Déviation quadratique d'estimateurs de densité par projections orthogonales. *Comptes Rendus de l'Académie des Sciences - Series I - Mathematics*, 310:425–430.

Doukhan, R. (1988). Formes de Töplitz associées à une analyse multiéchelle. *C.R. Acad. Sci. Paris Sér. A*, 306:663–666.

Hall, P. and Van Keilegom, I. (2004). Testing for monotone increasing hazard rate. *Annals of Statistics*, to appear.

Hall, P., Wolff, R. C. L., and Yao, Q. (1999). Methods for estimating a conditional distribution function. *Journal of the American Statistical Association*, 94:154–163.

Härdle, W., Kerkyacharian, G., Picard, D., and Tsybakov, A. B. (1998). *Wavelets, Approximation and Statistical Applications*, volume 129 of *Lecture Notes in Statistics*. Springer Verlag, New York.

Kerkyacharian, G. and Picard, D. (1992). Density estimation in Besov spaces. *Statistics and Probability Letters*, 13:15–24.

Kerkyacharian, G., Picard, D., and Tribouley, K. (1996). l^p adaptive density estimation. *Bernoulli*, 2(3):229–247.

Lee, J. and Hong, Y. (2001). Testing for correlation of unknown form using wavelet methods. *Econometric Theory*, 17:386–423.

Masry, E. (1994). Probability density estimation from dependent observations using wavelets orthonormal bases. *Statistics and Probability Letters*, 21:181–194.

Masry, E. (1997). Multivariate probability density estimation by wavelet methods: Strong consistency and rates for stationary time series. *Stochastic Processes and their Applications*, 67:177–193.

McFadden, D. (2003). *Economic Choices*, pages 330–365. Nobel Lectures in Economic Sciences 1996-2000. World Scientific Publishing Company.

McFadden, D. and Train, K. (2000). Mixed MNL models for discrete response. *Journal of Applied Econometrics*, 15:447–470.

Meyer, Y. (1992). *Ondelettes et Operateurs*, volume 1: Ondelettes. Cambridge : Cambridge University Press, Paris.

Nikol'skiĭ, S. M. (1975). *Approximations of Functions of Several Variables and Imbedding Theorems*. Springer-Verlag, Berlin Heidelberg New York.

Ogden, R. T. (1996). *Essential Wavelets for Statistical Applications and Data Analysis*. Birkhäuser Boston Inc., Cambridge, MA, USA.

Penev, S. and Dechevsky, L. (1997). On non-negative wavelet-based density estimators. *Journal of Nonparametric Statistics*, 7:365–394.

Petrushev, P. P. and Popov, V. A. (1987). *Rational Approximation of Real Functions*. Cambridge University Press, Cambridge.

Pham, T. D. and Tran, L. T. (1980). The strong mixing property od the autoregressive moving average time series model. *Séminaire de Statistique*, pages 59–76.

Pinheiro, A. and Vidakovic, B. (1997). Estimating the square root of a density via compactly supported wavelets. *Computational Statistics and Data Analysis*, 25:399–415.

Shapiro, J. (1993). Embedded image coding using zerotrees of wavelet coefficients. *IEEE Trans. Image Processing*, 41(12):3445–3462.

Tribouley, K. (1995). Practical estimation of multivariate densities using wavelet methods. *Statistica Neerlandica*, 49:41–62.

Tribouley, K. and Viennet, G. (1998). l_p adaptive density estimation in a beta-mixing framework. *Annales de l'Institut Henri Poincaré, Section B, Calcul des Probabilités et Statistique*, 34:179–208.

Vidakovic, B. (1999). *Statistical Modeling by Wavelets*. Wiley Series in Probability and Statistics. John Wiley & Sons, Inc., New York.

Yu, K. and Jones, M. C. (1998). Local linear quantile regression. *Journal of the American Statistical Association*, 93(441):228237.

APPENDICES

A The Dechevsky-Penev construction

This Appendix A gathers some of the details of the results of DP as far as they have not already been included in Section 3. For definitions of the given function spaces, such as $\dot{W}_q^1$, we refer to Nikol'skiĭ (1975).

A.1 Approximation of Univariate Probability Distribution Functions

Cumulative Approximation

Let $\varphi, \tilde{\varphi}$ satisfy (2.3) *to* (2.8). *Let F be a cumulative distribution function.* In the first set of corollaries, $\|\varsigma_{\varphi,\tilde{\varphi}}\|_\infty$ *may not be equal to zero.* If it is, the constants appearing in the bounds will be smaller.

Corollary 1. *There exists $c > 0$ independent of F, such that*

$$\left\| A_j\big(F\big)(\cdot) - F(\cdot)\right\|_1 \leqslant c \cdot 2^{-j}. \tag{A.1}$$

Corollary 2. *There exists $c > 0$ independent of F, such that*

$$\left\| A_j\big(F\big)(\cdot) - F(\cdot)\right\|_p \leqslant c \cdot 2^{-\frac{j}{p}}. \tag{A.2}$$

Corollary 3. *If F is absolutely continuous, and $f = F' \in L_p$, $1 \leqslant p \leqslant \infty$, then there exists a $c > 0$ independent of f, such that*

$$\left\| A_j\big(F\big)(\cdot) - F(\cdot)\right\|_p \leqslant c \cdot 2^{-j}\|f\|_p. \tag{A.3}$$

Corollary 4. *Let $F \in \dot{B}^s_{p\,\infty}$, $1 \leqslant p \leqslant \infty$, $0 < s < 1$. Then there exists a $c > 0$ independent of F, such that*

$$\left\| A_j\big(F\big)(\cdot) - F(\cdot)\right\|_p \leqslant c \cdot 2^{-js}|F|_{B^s_{p\,\infty}}. \tag{A.4}$$

Corollary 5. *Let $1 < p \leqslant \infty$, $1/p \leqslant s < 1$ and $F \in \dot{B}^s_{p\infty}$. Then for every $q : p \leqslant q \leqslant \infty$, there exists a $c > 0$ independent of F, such that*

$$\left\| A_j\big(F\big)(\cdot) - F(\cdot)\right\|_q \leqslant c \cdot 2^{-j\left(s-\frac{1}{p}+\frac{1}{q}\right)}|F|_{B^s_{p\,\infty}}. \tag{A.5}$$

Additional Assumption. In the following corollaries, $\varsigma_{\varphi,\tilde{\varphi}} = 0$ *a.s.*, has to be assumed.

Corollary 6. *Corollary 4 still holds and s can be taken to be $0 < s < 2$. Moreover, for $1 \leqslant s < 2$, we can rewrite inequality* (A.4) *as*

$$\left\| A_j\big(F\big)(\cdot) - F(\cdot)\right\|_p \leqslant c \cdot 2^{-js}|f|_{B^{s-1}_{p\,\infty}}, \tag{A.6}$$

where $f = F'$ is the density.

Corollary 7. *When $s = 2$, inequality* (A.4) *is modified, and can be rewritten*

$$\left\| A_j\big(F\big)(\cdot) - F(\cdot)\right\|_p \leqslant c \cdot 2^{-js}|F|_{\dot{W}^2_p}, \tag{A.7}$$

or

$$\left\| A_j\big(F\big)(\cdot) - F(\cdot)\right\|_p \leqslant c \cdot 2^{-js}|f|_{\dot{W}^1_p}, \tag{A.8}$$

where $f = F'$ is the density.

Density Approximation

For density estimation similar results hold as for the case cdf estimation. Corollaries 1 to 7 still hold if substituting F with f. The main difference is that cumulates are naturally smoother than densities functions. Corollary 1 automatically holds for cumulative distribution functions, which are by definition of bounded variation. We will instead have to assume that a density is of bounded variation for Corollary 1 to hold. The only result that one can obtain for a density without any assumption on its regularity is:

$$\left\| A_j\big(f\big)(\cdot) - f(\cdot)\right\|_1 = o(1), \quad as \quad j \to \infty\ , \tag{A.9}$$

since $f \in L_1$ by definition and $\omega(f, 2^{-j}) \to_{j\to\infty} 0$ *iff* $f \in L_1$ by definition of the modulus of smoothness. The following corollaries are given without proof.

Corollary 8. *Let f be of bounded variation. Then Corollaries 1 and 2 hold for densities, and F is replaced by f.*

Corollary 9. *Corollaries 3 to 7 hold for densities, F is replaced by f and f is replaced by f'.*

A.2 Stochastic error for Univariate Probability Distribution Functions

For cdf With $p \geqslant 1, 0 < q \leqslant 2, \rho = \min\{p/2, 1\}$

$$\left\| \hat{A}_j^{(n)}(F)(\cdot) - \mathbb{E}\big(\hat{A}_j^{(n)}(F)(\cdot)\big) \Big| L_p(\mathscr{L}_q)\right\|^\rho \leqslant c \cdot n^{-\rho/2}\Big\{ \| F(1-F)\|_{p/2}^{\min\{1,p/2\}} + c_2(a,\varphi,\tilde{\varphi})\omega_1\big(F, 2^{1-j}a\big)_{p/2}^{\min\{1,p/2\}}\Big\}, \tag{A.10}$$

Moreover, for $1 \leqslant p < \infty$ and $F(1-F) \in L_{p/2}$, we have that
$\omega_1(F, 2^{1-j}a)_{p/2} = o_j(1)$ as $j \to \infty$. If F is continuous, the case $p = \infty$ is included.

For pdf Let f be a density and $1 \leqslant p \leqslant \infty$, $0 < q \leqslant 2$, $\rho = \min\{p/2, 1\}$

$$\begin{aligned} &\left\| \hat{A}_j^{(n)}(f)(\cdot) - \mathbb{E}\big(\hat{A}_j^{(n)}(f)(\cdot)\big) \Big| L_p(\mathscr{L}_q)\right\|^\rho \\ &\leqslant \left(\frac{2^j}{n}\right)^{\rho/2}\Big[\max\big\{\|f\|_1, \|f\|_{p/2}\big\}^{\rho/2} \\ &\qquad + c(a)\max\big\{\omega^1(f, 2^{1-j}a)_1, \omega^1(f, 2^{1-j}a)_{p/2}\big\}^{\rho/2}\Big]. \end{aligned} \tag{A.11}$$

Moreover

$$\max\big\{\omega^1(f, 2^{1-j}a)_1, \omega^1(f, 2^{1-j}a)_{p/2}\big\} = o_j(1),\ j \to \infty,$$

if $2 \leqslant p < \infty$,
or if $p = \infty$ and f is continuous,
or if $1 < p < 2$ and $\sup\limits_{0 \leqslant t \leqslant h} \int_{-\infty}^{+\infty} \Big(\int_0^1 d\alpha\, f(x + \alpha t)\Big)^{p/2} < +\infty$.

A.3 Choosing the optimal estimation risk for pdf

To choose the optimal function $j^*(n)$ in Theorem 1, the dependence of $B(j)$ on j needs to be made explicit. This can be done by using the results on the decay of the bias. For explicit expressions on the decay of the bias see Dechevsky and Penev (1997) or above. Denote $\big\|\hat{A}_j^{(n)}(f)(x) - f(x)\big\| = \varepsilon_{p,q}^{n,j}$ and $j^* = j(n)$ the optimal resolution level. The first corollaries are obtained with $\|\varsigma_{\varphi,\tilde{\varphi}}\|_\infty \neq 0$, but hold also true with smaller values of the constants if $\|\varsigma_{\varphi,\tilde{\varphi}}\|_\infty = 0$. The first proof is given as an example, the other proofs are straightforward following the example of the first one.

Corollary 10. *Let $1 \leqslant p \leqslant \infty$ and $0 < q < \infty$, $f \in \dot{W}_q^1 \cap L_1 \cap L_{p/2}$. Then,*

$$j^* = (\log_2 n)/3 \quad \textit{and} \quad (\varepsilon_{p,q}^{n,j})^\rho = O\big(n^{-\rho/3}\big).$$

Proof. $\omega^2(f, 2^{1-j}a)_p \leqslant 2\omega^1(f, 2^{1-j}a)_p$ and $\omega^1(f, 2^{1-j}a)_p \leqslant 2^{1-j}a\|f'\|_p$ so the bias term is bounded by $c \cdot 2^{-j\rho}$ while the variance grows as $\left(\frac{2^j}{n}\right)^{\rho/2}$. Equalizing $2^{-j\rho} = 2^{j\rho/2}n^{-\rho/2}$ gives $2^{-3j/2} = n^{-1/2}$ which is solved for $j^* = \frac{1}{3}\log_2 n$ that gives a convergence rate of $2^{-j^*\rho} = n^{-\rho/3}$. □

Corollary 11. *Let $1 \leqslant p \leqslant \infty$, $0 < q < \infty$ and $0 < s < 1$, $f \in \dot{B}_{p,\infty}^s \cap L_{\min\{1,p/2,p/q\}} \cap L_{\max\{1,p/2\}}$. Then*

$$j^* = (\log_2 n)/(1 + 2s) \quad \textit{and} \quad (\varepsilon_{p,q}^{n,j})^\rho = O\big(n^{-\frac{\rho s}{1+2s}}\big).$$

Corollary 12. *Let $1 < r \leqslant p \leqslant \infty$, $0 < q < \infty$ and $0 < s < 1$, $f \in \dot{B}_{r,\infty}^s \cap L_{\min\{1,p/2,p/q\}} \cap L_{\max\{1,p/2\}}$. Then*

$$j^* = \frac{\log_2 n}{1 + 2s - (2/r) + (2/p)} \quad \textit{and} \quad (\varepsilon_{p,q}^{n,j})^\rho = O\big(n^{-\frac{\rho[s-(1/r)+(1/p)]}{1+2s-(2/r)+(2/p)}}\big).$$

For the two following corollaries we assume $\|\varsigma_{\varphi,\tilde{\varphi}}\|_\infty = 0$.

Corollary 13. *Corollary 12 continues to hold with* $0 < s < 2$.

Corollary 14. *Let* $1 < r \leqslant p \leqslant \infty$, $0 < q < \infty$ *and* $f \in \dot{W}^2_r \cap L_{\min\{1,p/2,p/q\}} \cap L_{\max\{1,p/2\}}$. *Then*

$$j^* = \frac{\log_2 n}{5-(2/r)+(2/p)} \quad \text{and} \quad (\varepsilon^{n,j}_{p,q})^\rho = O\big(n^{-\frac{\rho[2-(1/r)+(1/p)]}{5-(2/r)+(2/p)}}\big).$$

B Proofs for Section 4

Lemma 2. We will estimate from above the following quantity:

$$E_J(F)(x,y) = A_J(F)(x,y) - F(x,y).$$

In the proof we may at times use the notation $\mathbf{x} = (x,y)$, $\mathbf{s} = (s,t)$. If the vector notation is used, then x_i, $i = 1,2$ stand for the components (x,y).

We will use the intermediate approximation $F_{\mu,h}$ (see Appendix D). We then have from the linearity of $A(F)(x,y)$ and from Minkowsky's inequality:

$$\begin{aligned} \sup_y \big\|E_J(F)(\cdot,y)\big\|_p &= \sup_y \big\|A_J(F)(\cdot,y) - F(\cdot,y)\big\|_p \\ &= \sup_y \big\|A_J(F - F_{\mu,h}) - (F - F_{\mu,h}) + A_J(F_{\mu,h}) - F_{\mu,h}\big\|_p \\ &\leqslant \sup_y \big\|E_J(F - F_{\mu,h})\big\|_p + \sup_y \big\|E_J(F_{\mu,h})\big\|_p. \end{aligned} \tag{B.1}$$

We start by evaluating the first term on the right hand side of Inequality (B.1). In the following, for simplicity of notation, $g(x,y) \doteq F(x,y) - F_{\mu,h}(x,y)$. Let $\varphi_{J,\mathbf{k}}$ and $\widetilde{\varphi}_{J,\mathbf{k}}$ be the *2*-dimensional wavelet scaling functions built by tensor products, and $J = j_1 + j_2$. Then, given that $\sum_{\mathbf{k}\in\mathbb{Z}^2} \varphi_{\mathbf{j},\mathbf{k}} = 2^{J/2}$, for every function $g(x,y)$ the following identity holds:

$$g(x,y) = \sum_{\mathbf{k}\in\mathbb{Z}^2} 2^{-J/2} \varphi_{\mathbf{j},\mathbf{k}}(x,y) g(x,y),$$

which gives:

$$E_J(g)(x,y) = \sum_{\mathbf{k}\in\mathbb{Z}^2} \Big(\langle g, \widetilde{\varphi}_{\mathbf{j},\mathbf{k}}\rangle - 2^{-J/2} g(x,y)\Big)\, \varphi_{\mathbf{j},\mathbf{k}}(x,y). \tag{B.2}$$

Moreover, since $\displaystyle\int_{\mathbb{R}^2} \widetilde{\varphi}_{\mathbf{j},\mathbf{k}}(s,t) ds dt = 2^{-J/2} \equiv \langle 1, \widetilde{\varphi}_{\mathbf{j},\mathbf{k}}\rangle$, Equation (B.2) can be rewritten as:

$$E_J(g)(x,y) = \Big[\sum_{\mathbf{k}\in\mathbb{Z}^2} \int_{\mathbb{R}^2} ds dt\, (g(s,t) - g(x,y)) \widetilde{\varphi}_{\mathbf{j},\mathbf{k}}(s,t)\Big] \varphi_{\mathbf{j},\mathbf{k}}(x,y).$$

Integrating with respect to the x variable and then taking the supremum with respect to the y variable gives the desired quantity:

$$\sup_y \big\|E_J(g(\cdot,y))\big\|_p = \sup_y \left(\int_{\mathbb{R}} dx \left| \sum_{\mathbf{k}\in\mathbb{Z}^2} \int_{\mathbb{R}^2} ds dt\, (g(s,t) - g(x,y))\, \widetilde{\varphi}_{\mathbf{j},\mathbf{k}}(s,t)\, \varphi_{\mathbf{j},\mathbf{k}}(x,y)\right|^p\right)^{1/p}. \tag{B.3}$$

Since φ and $\widetilde{\varphi}$ have a compact support, only a finite number of summands in (B.2) are different from zero. For a given (x,y), the non zero summands will be those whose indices belong to the set $\mathscr{I} = \mathscr{I}_1 \otimes \mathscr{I}_2$ where

$$\mathscr{I}_i \doteq \left\{k_i \mid 2^{j_i} x_i - a < k_i < 2^{j_i} x_i + a,\ i = 1,2.\right\},$$

that is those displaying $\varphi_{\mathbf{j},\mathbf{k}}$'s whose support contains (x,y). The cardinality of each $\mathscr{I}_i$ can be easily computed and does not depend on the level of approximation $\mathbf{j}$:

$$\sharp\ \mathscr{I}_i \leqslant \nu_a = \begin{cases} [2a]+1 & \text{if} \quad 2a \notin \mathbb{N}, \\ 2a & \text{if} \quad 2a \in \mathbb{N}, \end{cases} \qquad \forall\ (x,y) \in \mathbb{R}^2.$$

Taking into account the non zero terms only, we have from (B.3):

$$\sup_y \left\|E_J(g(\cdot,y))\right\|_p \leqslant \sup_y \left[\int_{\mathbb{R}} dx \left(\sum_{\mathscr{I}} \left|\int_{\mathbb{R}^2} dsdt\,(f(s,t)-f(x,y))\widetilde{\varphi}_{\mathbf{j},\mathbf{k}}(s,t)\right| \varphi_{\mathbf{j},\mathbf{k}}(x,y)\right)^p\right]^{1/p}. \tag{B.4}$$

For $1 \leqslant p \leqslant \infty$, $p' :\ 1/p + 1/p' = 1$ and for arbitrary a_i, $i = 1, \cdots, n$, the following inequality holds:

$$\left(\sum_{i=1}^n |a_i|^p\right)^{1/p} \leqslant \sum_{i=1}^n |a_i| \leqslant \left(\sum_{i=1}^n |a_i|^p\right)^{1/p} \cdot n^{1/p'}.$$

Applying the above inequality to (B.4) we obtain:

$$\sup_y \left\|E_J(g(\cdot,y))\right\|_p \leqslant \sup_y \nu_a^{2/p'} \left[\int_{\mathbb{R}} dx \left(\sum_{\mathscr{I}} \left|\int_{\mathbb{R}^2} dsdt\,(g(s,t)-g(x,y))\widetilde{\varphi}_{\mathbf{j},\mathbf{k}}(s,t)\right|^p \varphi_{\mathbf{j},\mathbf{k}}(x,y)^p\right)\right]^{1/p}. \tag{B.5}$$

Since in (B.5) there are only a finite number of summands, we can swap the integral and the summation symbols and so:

$$\begin{aligned}
\sup_y \left\|E_J(g(\cdot,y))\right\|_p &\leqslant \sup_y \nu_a^{2/p'} \left[\sum_{\mathscr{I}} \int_{\mathbb{R}} dx \left|\int_{\mathbb{R}^2} dsdt\,(g(s,t)-g(x,y))\widetilde{\varphi}_{\mathbf{j},\mathbf{k}}(s,t)\right|^p \varphi_{\mathbf{j},\mathbf{k}}(x,y)^p\right]^{1/p} \\
&= \sup_y \nu_a^{2/p'} \left[\sum_{\mathbf{k}\in\mathbb{Z}^d} \int_{\mathbb{R}} dx \left|\int_{\mathbb{R}^2} dsdt\,(g(s,t)-g(x,y))\widetilde{\varphi}_{\mathbf{j},\mathbf{k}}(s,t)\right|^p \varphi_{\mathbf{j},\mathbf{k}}(x,y)^p\right]^{1/p} \\
&= \sup_y \nu_a^{2/p'} \left[\sum_{\mathbf{k}\in\mathbb{Z}^d} \int_{\Xi_{k_1}} dx \left|\int_{\Xi_{\mathbf{k}}} dsdt\,(g(s,t)-g(x,y))\widetilde{\varphi}_{\mathbf{j},\mathbf{k}}(s,t)\right|^p \varphi_{\mathbf{j},\mathbf{k}}(x,y)^p\right]^{1/p},
\end{aligned} \tag{B.6}$$

where Ξ_{k_i} is the support of $\tilde{\varphi}_{j_i k_i}$ and $\Xi_{\mathbf{k}}$ is the support of $\tilde{\varphi}_{\mathbf{j},\mathbf{k}}$. In the last identity of (B.6) we used the fact that for a given index $\mathbf{k}$ the integral outside the support $\Xi_{\mathbf{k}}$ of $\varphi_{\mathbf{j},\mathbf{k}}$ is zero. We now consider the inner integral. Using Hölder's inequality:

$$\begin{aligned}
\left|\int_{\Xi_{\mathbf{k}}} (g(s,t)-g(x,y))\widetilde{\varphi}_{\mathbf{j},\mathbf{k}}(s,t)dsdt\right| &\leqslant \int_{\Xi_{\mathbf{k}}} |g(s,t)-g(x,y)|\,\widetilde{\varphi}_{\mathbf{j},\mathbf{k}}(s,t)dsdt \\
&\leqslant \left(\int_{\Xi_{\mathbf{k}}} |g(s,t)-g(x,y)|^p\, dsdt\right)^{1/p} \cdot \left(\int_{\Xi_{\mathbf{k}}} \widetilde{\varphi}_{\mathbf{j},\mathbf{k}}(s,t)^{p'} dsdt\right)^{1/p'}
\end{aligned}$$

which yields with a change of variable $\tau_i \doteq 2^{j_1} s_i - k_i$:

$$= \left(\int_{\Xi_{\mathbf{k}}} |g(s,t)-g(x,y)|^p\, dsdt\right)^{1/p} \cdot 2^{J\left(\frac{1}{2}-\frac{1}{p'}\right)} \left(\int_{[-a,a]^2} \widetilde{\varphi}(\tau)^{p'} d\tau\right)^{1/p'}. \tag{B.7}$$

Combining (B.7) with (B.6), we have that:

$$\sup_y \left\|E_J(g(\cdot,y))\right\|_p \leqslant \nu_a^{\frac{2}{p'}} 2^{J\left(\frac{1}{2}-\frac{1}{p'}\right)} \|\,\widetilde{\varphi}\|_{p'} \cdot \sup_y \left(\sum_{\mathbf{k}\in\mathbb{Z}^2} \|\,\varphi_{\mathbf{j},\mathbf{k}}\|_\infty \int_{\Xi_{k_1}} dx \int_{\Xi_{\mathbf{k}}} |g(s,t)-g(x,y)|^p\, dsdt\right)^{1/p}. \tag{B.8}$$

We now introduce a change of variable $\mathbf{u} = (u,v) :\ \mathbf{s} \doteq \mathbf{x} + \mathbf{u}$. Since for a fixed value of the location parameter $\mathbf{k}$, $\mathbf{s}$ ranges in the domain defined by $\Xi_{\mathbf{k}}$, then by definition u_i ranges in $[2^{-j_i}(k_1-a)-x_i, 2^{-j_i}(k_1+a)-x_i]$. But since (x,y) is also allowed to range in a region $\Xi_{\mathbf{k}}$ independently from (s,t), it easy to see that u_i ranges in $[2^{-j_i}(k_i-a)-\max\{x_i \in \Xi_{k_i}\}, 2^{-j_i}(k_i+a)-\min\{x_i \in \Xi_{k_i}\}] \equiv [2^{-j_i}(k_i-a)-2^{-j_i}(k_i+a), 2^{-j_i}(k_i+a)-2^{-j_i}(k_i-a)] \equiv [-2^{1-j_i}a, 2^{1-j_i}a]$. Then $\mathbf{u}$ ranges in the hyperrectangle $\Upsilon \doteq [-2^{1-j_1}a, 2^{1-j_1}a] \otimes [-2^{1-j_2}a, 2^{1-j_2}a]$, which is independent of $\mathbf{k}$. Remembering that $\|\varphi_{\mathbf{j},\mathbf{k}}\|_\infty = 2^{J/2}\|\varphi\|_\infty$, and introducing $\kappa \doteq \nu_a^{2/p'} 2^{J/p'}\|\widetilde{\varphi}\|_{p'}\|\varphi\|_\infty$, we rewrite Equation (B.8) as :

$$\sup_y \left\|E_J(g(\cdot,y))\right\|_p \leqslant \kappa\, \nu_a^{\frac{1}{p}} \sup_y \left(\sum_{k_1\in\mathbb{Z}} \int_{\Xi_{k_1}} dx \int_{\Upsilon} d\mathbf{u}\, |g(\mathbf{x}+\mathbf{u}) - g(\mathbf{x})|^p\right)^{1/p}$$

which gives by Fubini's Theorem:

$$= \kappa\, \nu_a^{\frac{1}{p}} \sup_{y\in\mathbb{R}} \left(\int_{\Upsilon} d\mathbf{u} \sum_{k_1\in\mathbb{Z}} \int_{\Xi_{k_1}} dx\, |g(\mathbf{x}+\mathbf{u}) - g(\mathbf{x})|^p\right)^{1/p}$$

$$\leqslant \kappa\, \nu_a^{\frac{2}{p}} \sup_{y\in\mathbb{R}} \left(\int_{\Upsilon} dudv \sum_{k_1\in\mathbb{Z}} \int_{-\infty}^{+\infty} dx\, |g(s+u,t+v) - g(s,t) + g(x,y+v) - g(x,y+v)|^p\right)^{1/p}$$

and yields in view of the inequality $\omega^1(f,h)_p \leqslant 2\|f\|_p$:

$$\leqslant \kappa\, \nu_a^{\frac{2}{p}}\, 2^{4-J} a^2 \sup_{y\in\mathbb{R}} \left[\omega^1_{\mathbf{e}_x}(g, 2^{1-j_1})_p + \left\|g(\cdot, y+v)\right\|_p + \left\|g(\cdot,y)\right\|_p\right]$$

$$\leqslant \kappa\, \nu_a^{\frac{2}{p}}\, 2^{4-J} a^2 \sup_{y\in\mathbb{R}} \left[2\left\|g(\cdot,y)\right\|_p + \left\|g(\cdot, y+v)\right\|_p + \left\|g(\cdot,y)\right\|_p\right] \leqslant \kappa\, \nu_a^{\frac{2}{p}}\, 2^{4-J} a^2 \cdot 4 \sup_y \left\|g(\cdot,y)\right\|_p.$$

Keeping in mind that $g(x,y) = F(x,y) - F_{\mu,h}(x,y)$, with $\mu = 2$ and $h = (2^{1-j_1}, 2^{1-j_2})$ Property (D.1) of the Steklov means finally gives for the first term of the approximation (B.1)

$$\sup_y \left\|E_J(g(\cdot,y))\right\|_p \leqslant \nu_a^2 2^{\frac{4}{p}+2} a^{\frac{2}{p}} \|\widetilde{\varphi}\|_{p'}\|\varphi\|_\infty \sup_y \sup_{\mathbf{i}\in\mathbb{R}^2} \sup_{0<\delta<1} \left\|\Delta^2_{\mathbf{i}\delta(2^{1-j^*}a)} F(\cdot,y)\right\|_p, \tag{B.9}$$

where j^* solves the equation $(2^{1-j^*})^2 = (2^{1-j_1})^2 + (2^{1-j_2})^2$.

We now study the second term of Equation (B.1), that is, the approximation of the Steklov mean $F_{\mu,h}$. We assume that $F_{\mu,h}$ has almost everywhere partial and mixed derivatives up to the second order, so we can write for it a second order Taylor expansion with an integral remainder:

$$g(s,t) = g(x,y) + (s_i - x_i) g^{(i)}(x,y) + (s_i - x_i)(s_j - x_j) \int_0^1 d\theta (1-\theta)\, g^{(ij)}(\mathbf{x} + \theta(\mathbf{s}-\mathbf{x})),$$

with $g = F_{\mu,h}$, $(x,y),(s,t) \in \mathbb{R}^2$, $g^{(i)} \doteq \frac{\partial g}{\partial x_i}$, $g^{(ij)} \doteq \frac{\partial^2 g}{\partial x_i \partial x_j}$, and where the summation is understood over repeated indices. Then following the same steps that led us to (B.3), we have:

$$\sup_y \left\|E_J(g(\cdot,y))\right\|_p = \sup_y \left(\int_{\mathbb{R}} dx \left|\sum_{\mathbf{k}\in\mathbb{Z}^2} \int_{\mathbb{R}^2} dsdt\, \big(g(s,t) - g(x,y)\big) \widetilde{\varphi}_{\mathbf{j},\mathbf{k}}(s,t)\, \varphi_{\mathbf{j},\mathbf{k}}(x,y)\right|^p\right)^{1/p}.$$

We start by evaluating the inner integral. By substituting for $g(s,t)$ its Taylor expansion and successive application of the triangular inequality, we have:

$$\begin{aligned}
&\left|\sum_{\mathscr{I}} \int_{\mathbb{R}^2} dsdt\, \big(g(s,t) - g(x,y)\big) \widetilde{\varphi}_{\mathbf{j},\mathbf{k}}(s,t) \varphi_{\mathbf{j},\mathbf{k}}(x,y)\right| = \left|\sum_{\mathbf{k}\in\mathbb{Z}^2} \int_{\Xi_{\mathbf{k}}} dsdt\, \big(g(s,t) - g(x,y)\big) \widetilde{\varphi}_{\mathbf{j},\mathbf{k}}(s,t) \varphi_{\mathbf{j},\mathbf{k}}(x,y)\right| \\
&= \left|\sum_{\mathscr{I}} \int_{\Xi_{\mathbf{k}}} dsdt \Big[(t_i - x_i) g^{(i)}(x,y) + (t_i - x_i)(t_j - x_j) \int_0^1 d\theta\, (1-\theta) g^{(ij)}(\mathbf{x} + \theta(\mathbf{s}-\mathbf{x}))\Big] \widetilde{\varphi}_{\mathbf{j},\mathbf{k}}(s,t) \varphi_{\mathbf{j},\mathbf{k}}(x,y)\right| \\
&\leqslant \left| g^{(i)}(x,y) \left\{\sum_{\mathscr{I}} \left[2^{-j_i} \int_{\Xi_{\mathbf{k}}} dsdt\, (2^{j_i} s_i - k_i) \widetilde{\varphi}_{\mathbf{j},\mathbf{k}}(s,t) - 2^{-j_i} \int_{\Xi_{\mathbf{k}}} dsdt\, (2^{j_i} x_i - k_i) \widetilde{\varphi}_{\mathbf{j},\mathbf{k}}(s,t)\right] \varphi_{\mathbf{j},\mathbf{k}}(x,y)\right\}\right| \\
&\qquad + \left|\sum_{\mathscr{I}} \int_{\Xi_{\mathbf{k}}} dsdt\, (s_i - x_i)(s_j - x_j) \widetilde{\varphi}_{\mathbf{j},\mathbf{k}}(s,t) \int_0^1 d\theta\, (1-\theta) g^{(ij)}(\mathbf{x} + \theta(\mathbf{s}-\mathbf{x})) \varphi_{\mathbf{j},\mathbf{k}}(x,y)\right|.
\end{aligned} \tag{B.10}$$

The first term of the right hand side of Inequality (B.10) is obtained by adding and subtracting $2^{j_i}k_i$. Keeping in mind Properties (2.6) and (2.8), further manipulation of this term gives:

$$\begin{aligned}
&\sum_{\mathscr{I}}\left[\int_{\Xi_{\mathbf{k}}} dsdt\,(2^{j_i}s_i-k_i)\widetilde{\varphi}_{\mathbf{j},\mathbf{k}}(s,t)-\int_{\Xi_{\mathbf{k}}} dsdt\,(2^{j_i}x_i-k_i)\widetilde{\varphi}_{\mathbf{j},\mathbf{k}}(s,t)\right]\varphi_{\mathbf{j},\mathbf{k}}(x,y)\\
&=\sum_{\mathscr{I}}\left[\int_{\Xi_{\mathbf{k}}} 2^{-j_i}d(2^{j_i}s_i)2^{-j_j}d(2^{j_j}s_j)(2^{j_i}s_i-k_i)2^{\frac{j_i}{2}}\widetilde{\varphi}(2^{j_i}s_i-k_i)\prod_{j\neq i}2^{\frac{j_j}{2}}\widetilde{\varphi}(2^{j_j}s_j-k_j)\right.\\
&\qquad\left.-\int_{\Xi_{\mathbf{k}}} 2^{-j_i}d(2^{j_i}s_i)2^{-j_j}d(2^{j_j}s_j)(2^{j_i}x_i-k_i)2^{\frac{j_i}{2}}\widetilde{\varphi}(2^{j_i}s_i-k_i)\prod_{j\neq i}2^{\frac{j_i}{2}}\widetilde{\varphi}(2^{j_j}s_j-k_j)\right]\varphi_{j_ik_i}(x_i)\varphi_{j_jk_j}(x_j)\\
&=\int_{-a}^{a}d\tau_i\,\tau_i\widetilde{\varphi}(\tau_i)\;2^{-\frac{J}{2}}\int_{-a}^{a}d(\xi)\,\widetilde{\varphi}(\xi)\sum_{\mathscr{I}}2^{\frac{j_i}{2}}\varphi_{j_ik_i}(2^{j_i}x_i-k_i)2^{\frac{j_j}{2}}\varphi_{j_jk_j}(2^{j_j}x_j-k_j)\\
&\qquad-2^{-\frac{J}{2}}\int_{-a}^{a}d\xi\,\widetilde{\varphi}(\xi)\sum_{\mathscr{I}}2^{\frac{j_1}{2}}(2^{j_1}x_i-k_i)\varphi(2^{j_i}x_i-k_i)\varphi(2^{j_j}x_j-k_j)\\
&=\int_{-a}^{a}d\tau_i\,\tau_i\,\widetilde{\varphi}(\tau_i)-\sum_{k_i=-\infty}^{+\infty}(\zeta_i-k_i)\varphi(\zeta_i-k_i)\doteq\int_{-a}^{a}d\tau_i\,\tau_i\,\widetilde{\varphi}(\tau_i)-\rho_{\varphi}(\zeta_i)\doteq\varsigma_{i;\widetilde{\varphi},\varphi}(\zeta_i),
\end{aligned}\tag{B.11}$$

where the last two identities have to be read as definitions of ρ_φ and $\varsigma_{\widetilde{\varphi},\varphi}(\zeta)$. Let us remark that if in building the $\mathbb{R}^2$ tensor product basis we start from the same φ for every direction, then definitions of ρ_φ and $\varsigma_{i;\widetilde{\varphi},\varphi}(\zeta_i)$ are independent of i. Regarding the second term of the right hand side of Equation (B.10), again application of Fubini's theorem and recalling. as observed before, that $|s_i-x_i|<2^{1-j_i}a$, gives that:

$$\begin{aligned}
&\left|\sum_{\mathscr{I}}\int_{\Xi_{\mathbf{k}}} dsdt\,(t_i-x_i)^2\widetilde{\varphi}_{\mathbf{j},\mathbf{k}}(s,t)\int_0^1 d\theta\,(1-\theta)g^{(ij)}(\mathbf{x}+\theta(\mathbf{s}-\mathbf{x}))\varphi_{\mathbf{j},\mathbf{k}}(x,y)\right|\\
&\leqslant(2^{1-j_i}a)^2\int_0^1 d\theta\,(1-\theta)\sum_{\mathscr{I}}\int_{\Xi_{\mathbf{k}}} dsdt\,\left|g^{(ij)}(\mathbf{x}+\theta(\mathbf{s}-\mathbf{x}))\right|\widetilde{\varphi}_{\mathbf{j},\mathbf{k}}(s,t)\varphi_{\mathbf{j},\mathbf{k}}(x,y).
\end{aligned}\tag{B.12}$$

We can now substitute (B.11) and (B.12) in Equation (B.10), integrate with respect to x, use Minkowsky's inequality, take the supremum over y and obtain:

$$\begin{aligned}
&\sup_y\left\|E_J(g(\cdot,y))\right\|_p\leqslant\sup_y\left[\int_{\mathbb{R}}dx\left|\sum_i 2^{-j_i}g^{(i)}(x,y)\varsigma_{i;\widetilde{\varphi},\varphi}(x_i)\right|^p\right]^{1/p}\\
&+\sup_y(2^{1-j_i}a)^2\left[\int_{\mathbb{R}}dx\left|\int_0^1 d\theta\,(1-\theta)\sum_{\mathscr{I}}\int_{\Xi_{\mathbf{k}}}dsdt\,\left|g^{(ij)}(\mathbf{x}+\theta(\mathbf{s}-\mathbf{x}))\right|\widetilde{\varphi}_{\mathbf{j},\mathbf{k}}(s,t)\varphi_{\mathbf{j},\mathbf{k}}(x,y)\right|^p\right]^{1/p}.
\end{aligned}\tag{B.13}$$

Let us consider again the two summands separately. For the first term we can easily see:

$$\begin{aligned}
&\sup_y\left[\int_{\mathbb{R}}dx\left|\sum_i 2^{-j_i}g^{(i)}(x,y)\varsigma_{i;\widetilde{\varphi},\varphi}(x_i)\right|^p\right]^{1/p}\leqslant\sum_i\left\|\varsigma_i(2^{j_i}\cdot)\right\|_\infty\sup_y\frac{1}{2a}\,2^{1-j_i}a\left\|g^{(i)}(\cdot,y)\right\|_p\\
&\leqslant\frac{c_1}{2a}\left\{\left\|\varsigma_x(2^{j_1}\cdot)\right\|_\infty\sup_y\omega^1_{\mathbf{e}_x}\left(F(\cdot,y),2^{1-j_1}a\right)_p+\left\|\varsigma_y(2^{j_2}\cdot)\right\|_\infty\sup_y\sup_{0<\delta<1}\left\|\Delta_{\mathbf{e}_y\delta(2^{1-j_2}a)}F(\cdot,y)\right\|_p\right\},
\end{aligned}\tag{B.14}$$

where c_1 is an absolute constant which comes from property (D.2). For the second term we have that, using Minkowsky's generalized inequality:

$$\begin{aligned}
&\sup_y(2^{1-j_i}a)(2^{1-j_j}a)\left[\int_{\mathbb{R}}dx\left|\int_0^1 d\theta\,(1-\theta)\sum_{\mathscr{I}}\int_{\Xi_{\mathbf{k}}}dsdt\,\left|g^{(ij)}(\mathbf{x}+\theta(\mathbf{s}-\mathbf{x}))\right|\widetilde{\varphi}_{\mathbf{j},\mathbf{k}}(s,t)\varphi_{\mathbf{j},\mathbf{k}}(x,y)\right|^p\right]^{1/p}\\
&\leqslant\sup_y(2^{1-j_i}a)(2^{1-j_j}a)\int_0^1 d\theta\,(1-\theta)\left[\int_{\mathbb{R}}dx\,\sum_{\mathscr{I}}\left|\int_{\Xi_{\mathbf{k}}}dsdt\,\left|g^{(ij)}(\mathbf{x}+\theta(\mathbf{s}-\mathbf{x}))\right|\widetilde{\varphi}_{\mathbf{j},\mathbf{k}}(s,t)\varphi_{\mathbf{j},\mathbf{k}}(x,y)\right|^p\right]^{1/p}
\end{aligned}$$

following the same steps that led us from (B.4) to (B.8), and with $\kappa' \doteq a^2 \nu_a^{\frac{2}{p'}} \nu_a^{\frac{1}{p}} \| \tilde{\varphi} \|_{L_{p'}} \| \varphi \|_\infty$, and $\nu_a^{\frac{1}{p}}$ coming from the overlap of the supports of the $\{\varphi_{j_1,k_1}\}$ in the y direction,

$$
\begin{aligned}
&\leqslant (2^{1-j_i})(2^{1-j_j})\, \kappa' 2^{\frac{J}{p}} \sup_{y\in\mathbb{R}} \int_0^1 d\theta\,(1-\theta) \left(\sum_{k_1=-\infty}^{+\infty} \int_{\Xi_{k_1}} dx \int_{\Xi_{\mathbf{k}}} dsdt \left| g^{(ij)}(\mathbf{x}+\theta(\mathbf{s}-\mathbf{x}))\right|^p \right)^{\frac{1}{p}} \\
&= (2^{1-j_i})(2^{1-j_j})\, \kappa' 2^{\frac{J}{p}} \sup_{y\in\mathbb{R}} \int_0^1 d\theta\,(1-\theta) \left(\int_\Upsilon d\mathbf{u} \sum_{k_1=-\infty}^{+\infty} \int_{\Xi_{k_1}} dx \left| g^{(ij)}(\mathbf{x}+\theta\mathbf{u})\right|^p \right)^{\frac{1}{p}} \\
&\leqslant (2^{1-j_i})(2^{1-j_j})\, \kappa' 2^{\frac{J}{p}} \sup_{y\in\mathbb{R}} \int_0^1 d\theta(1-\theta)\cdot 2^{\frac{2-j_1}{p}} 2^{\frac{2-j_2}{p}} a^{\frac{2}{p}} \nu_a^{\frac{1}{p}} \left\| g^{(ij)}(\cdot,y)\right\|_p \\
&= (2^{1-j_i})(2^{1-j_j})\, a^{2+\frac{2}{p}} \nu_a^2 2^{\frac{4}{p}-1} \| \tilde{\varphi} \|_{L_{p'}} \| \varphi \|_\infty \sup_y \left\| g^{(ij)}(\cdot,y)\right\|_p .
\end{aligned}
$$

Remembering that $g = F_{\mu,h}$, the bound on the last term can be computed again using Properties (D.2), and depends on the order and direction of the derivatives:

$$
\begin{aligned}
&\sup_y (2^{1-j_i})(2^{1-j_j})\, a^2 \left\| g^{(ij)}(\cdot,y)\right\|_p \\
&\qquad\qquad \leqslant \sup_y c_2\, \omega^2_{\mathbf{e}_x}\big(F(\cdot,y)\big)_p \\
&\qquad\qquad + 2c_2 \cdot \Big\{ \sup_y \omega^1_{\mathbf{e}_x}\big(F(\cdot,y), 2^{1-j_1}a\big)_p + \sup_y \sup_{0<\delta<1} \left\| \Delta_{\mathbf{e}_y \delta(2^{1-j_2}a)} F(\cdot,y)\right\|_p \Big\} \\
&\qquad\qquad + \sup_y c_2 \sup_{0<\delta<1} \left\| \Delta^2_{\mathbf{e}_y\,\delta\cdot(2^{1-j_2}a)} F(\cdot,y)\right\|_p .
\end{aligned} \tag{B.15}
$$

Using (B.14) and (B.15) we can now rewrite (B.13) in the following form:

$$
\begin{aligned}
&\sup_y \left\| E_J(g(\cdot,y))\right\|_p \\
&\leqslant \frac{c_1}{2a} \Big\{ \| \varsigma_x(2^{j_1}\cdot)\|_\infty \sup_y \omega^1_{\mathbf{e}_x}\big(F(\cdot,y), 2^{1-j_1}a\big)_p + \| \varsigma_y(2^{j_2}\cdot)\|_\infty \sup_y \sup_{0<\delta<1} \left\| \Delta_{\mathbf{e}_y \delta(2^{1-j_2}a)} F(\cdot,y)\right\|_p \Big\} \\
&\quad + a^{\frac{2}{p}} \nu_a^2 2^{\frac{4}{p}-1} \| \tilde{\varphi} \|_{L_{p'}} \| \varphi \|_\infty 2c_2 \cdot \Big\{ \sup_y \omega^1_{\mathbf{e}_x}\big(F(\cdot,y), 2^{1-j_1}a\big)_p + \sup_y \sup_{0<\delta<1} \left\| \Delta_{\mathbf{e}_y \delta(2^{1-j_2}a)} F(\cdot,y)\right\|_p \Big\} \\
&\quad + a^{\frac{2}{p}} \nu_a^2 2^{\frac{4}{p}-1} \| \tilde{\varphi} \|_{L_{p'}} \| \varphi \|_\infty c_2 \cdot \Big\{ \sup_y \omega^2_{\mathbf{e}_x}\big(g(\cdot,y)\big)_p + \sup_y \sup_{0<\delta<1} \left\| \Delta^2_{\mathbf{e}_y\,\delta\cdot(2^{1-j_2}a)} F(\cdot,y)\right\|_p \Big\}.
\end{aligned} \tag{B.16}
$$

We can now put together (B.9) and (B.16) to obtain a final bound from above of the approximation error:

$$
\begin{aligned}
&\sup_y \left\| A_J(F)(\cdot,y) - F(\cdot,y)\right\|_p \\
&\quad \leqslant c_1' \Big\{ \| \varsigma_x(2^{j_1}\cdot)\|_\infty \sup_y \omega^1_{\mathbf{e}_x}\big(F(\cdot,y), 2^{1-j_1}a\big)_p + \| \varsigma_y(2^{j_2}\cdot)\|_\infty \sup_y \sup_{0<\delta<1} \left\| \Delta_{\mathbf{e}_y \delta(2^{1-j_2}a)} F(\cdot,y)\right\|_p \Big\} \\
&\qquad + c_2' \| \tilde{\varphi} \|_{L_{p'}} \| \varphi \|_\infty \cdot \Big\{ \sup_y \omega^1_{\mathbf{e}_x}\big(F(\cdot,y), 2^{1-j_1}a\big)_p + \sup_y \sup_{0<\delta<1} \left\| \Delta_{\mathbf{e}_y \delta(2^{1-j_2}a)} F(\cdot,y)\right\|_p \Big\} \\
&\qquad + c_2'' \cdot \Big\{ \sup_y \omega^2_{\mathbf{e}_x}\big(F(\cdot,y)\big)_p + \sup_y \sup_{0<\delta<1} \left\| \Delta^2_{\mathbf{e}_y\,\delta\cdot(2^{1-j_2}a)} F(\cdot,y)\right\|_p \Big\} \\
&\qquad + c_2''' \| \tilde{\varphi} \|_{p'} \| \varphi \|_\infty \sup_y \sup_{\mathbf{i}\in\mathbb{R}^2} \sup_{0<\delta<1} \left\| \Delta^2_{\mathbf{i}\delta(2^{1-j_*}a)} F(\cdot,y)\right\|_p ;
\end{aligned} \tag{B.17}
$$

where the expressions of the constants can easily be made explicit by comparing (B.9), (B.16) and (B.17). Finally, (B.17) can be further simplified by remarking that simple algebra (i.e. adding and subtracting $F(x+2s, y+t)$) leads to the inequality
$\sup_{0<\delta<1} \left\| \Delta^2_{\mathbf{i}\delta(2^{1-j_*}a)} F(\cdot,y)\right\|_p \leqslant 2\cdot \sup_y \sup_{0<\delta<1} \left\| \Delta_{\mathbf{e}_y\delta(2^{1-j_2}a)} F(\cdot,y)\right\|_p + \sup_y \omega^2_{\mathbf{e}_x}\big(F(\cdot,y)\big)_p$. □

Lemma 3. We have to find an upper bound for the quantity:

$$
\left\| \hat{A}^{(T)}_J(F)(\cdot,y) - \mathbb{E}\hat{A}^{(T)}_J(F)(\cdot,y) \middle| L_p(\mathscr{L}_q) \right\| = \left\| \Big(\mathbb{E}\left|\hat{A}^{(T)}_J(F)(\cdot,y) - A_J(F)(\cdot,y)\right|^q \Big)^{1/q} \right\|_p . \tag{B.18}
$$

We will proceed in the following way. We start by expressing the estimator $\hat{A}_J^{(T)}(F)(x,y)$ as:

$$
\begin{aligned}
\hat{A}_J^{(T)}(F)(x,y) &= \sum_{\mathbf{k}\in\mathbb{Z}^2} \widehat{\langle F, \tilde{\varphi}_{\mathbf{j},\mathbf{k}}\rangle}\, \varphi_{\mathbf{j},\mathbf{k}}(x,y) \\
&= \sum_{\mathbf{k}\in\mathbb{Z}^2} \sum_{t=1}^{T} 2^{-\frac{j_2}{2}} \left(\frac{\tilde{\varphi}_{j_1,k_1}(X_t)}{T} - \frac{\tilde{\varphi}_{j_1,k_1}(X_t)\tilde{\Phi}(2^{j_2}Y_t - k_2)}{T} \right) \varphi_{\mathbf{j},\mathbf{k}}(x,y) \\
&= \frac{1}{T}\sum_{t=1}^{T} \sum_{\mathbf{k}\in\mathbb{Z}^2} \widehat{\langle F, \tilde{\varphi}_{\mathbf{k}}\rangle}_t\, \varphi_{\mathbf{j},\mathbf{k}}(x,y) = \frac{1}{T}\sum_{t=1}^{T} \hat{A}_{J,t}(F)(x,y).
\end{aligned} \tag{B.19}
$$

We then determine the pointwise variance for the stochastic variable:

$$
Z_t(x,y) \doteq \hat{A}_{J,t}(f)(x,y) - A_J(F)(x,y). \tag{B.20}
$$

Let us remark in passing that $|Z| \leqslant 2 \cdot 2^{j_1}$ since both $\hat{A}_J(f)$ and $A_J(F)$ are smaller than 2^{j_1}, and that $\mathbb{E}Z = 0$. We can then express the inner expectation of the right hand side of (B.18):

$$
B_q(x,y) = \left(\mathbb{E}\big|\hat{A}_J^{(T)}(F)(x,y) - A_J(F)(x,y)\big|^q \right)^{\frac{1}{q}}. \tag{B.21}
$$

Being $q \leqslant 2$, we can take advantage of the classical convexity inequality for random variables

$$
\begin{aligned}
\mathbb{E}\big|\hat{A}_J^{(T)}(x,y) - A_J(x,y)\big|^q &= \mathbb{E}\left|\sum_t \frac{Z_t(x,y)}{T}\right|^q \leqslant \frac{1}{T^q}\left[\mathbb{E}\Big(\sum_t Z_t(x,y)\Big)^2\right]^{q/2} \\
&= \frac{1}{T^q}\left[\sum_{t=1}^{T} \mathbb{E}\big(Z_t^2(x,y)\big) + 2\sum_{p=1}^{T-1}(T-p)\mathbb{E}\big(Z_T(x,y)Z_{T-p}(x,y)\big)\right]^{q/2}.
\end{aligned}
$$

In the last inequalities we exploited the stationarity of the process, so that $\mathbb{E}(Z_t Z_{t-s})$ is independent of t. Then we take the L_p norm with respect to x and the supremum with respect to y and obtain:

$$
\begin{aligned}
&\sup_y \left\| \hat{A}_J^{(T)}(F)(\cdot,y) - \mathbb{E}\hat{A}_J^{(T)}(F)(\cdot,y) \Big| L_p(\mathscr{L}_q) \right\|^{\rho} \\
&\leqslant \sup_y \frac{1}{T^\rho}\left\{ \Big\| \sum_{t=1}^{T} \mathbb{E}\big(Z_t^2(\cdot,y)\big)\Big\|_{p/2}^{\rho/2} + \Big\| 2T\sum_{p=1}^{T-1}\Big(1-\frac{p}{T}\Big)\mathbb{E}\big(Z_T(\cdot,y)Z_{T-p}(\cdot,y)\big)\Big\|_{p/2}^{\rho/2} \right\} \\
&\doteq V_0 + V_1.
\end{aligned} \tag{B.22}
$$

It can be easily seen that the first summand of (B.22), that we refer to as V_0, is the only term we would have obtained in the *i.i.d.* case. V_1 is the additional term we obtain by estimating $\hat{A}_J^T$ from dependent data. We first start by studying the V_0 term.

A. Study of V_0 *term*: Let us define the quantity $\sigma^2(x,y)$, that is the variance of Z for a fixed (x,y). Then:

$$
\begin{aligned}
\sigma^2(x,y) = \mathbb{E}[\hat{A}_J^{(T)}(F)(x,y)]^2 - &A_J(F)(x,y)^2 \\
&= \sum_{\mathbf{k}\in\mathbb{Z}^2}\sum_{\mathbf{l}\in\mathbb{Z}^2} \big[\mathbb{E}\big(\widehat{\langle F, \tilde{\varphi}_{\mathbf{j},\mathbf{k}}\rangle}\widehat{\langle F, \tilde{\varphi}_{\mathbf{j},\mathbf{l}}\rangle}\big) - \langle F, \tilde{\varphi}_{\mathbf{j},\mathbf{k}}\rangle\langle F, \tilde{\varphi}_{\mathbf{j},\mathbf{l}}\rangle\big]\varphi_{\mathbf{j},\mathbf{k}}(x,y)\varphi_{\mathbf{j},\mathbf{l}}(x,y)
\end{aligned} \tag{B.23}
$$

Unless stated differently, in the steps of the proof that will follow, $f(x,y)$ is the bivariate density of the process $\{X_t, Y_t\}$, $F(x,y)$ is the bivariate pdf-cdf, cumulated w.r.t. to the y argument, and $f(x)$ the marginal distribution of $\{X_t\}$, that is, $f(x) = \lim_{y\to\infty} F_Y(x,y)$. Remembering equation (B.19), we take the empirical inner product for $T = 1$, and have that, if S and T are two random variables drawn from a bivariate (cumulative) distribution $F_{S,T}(S,T)$:

$$
\begin{aligned}
\Delta_{\mathbf{j};\mathbf{k},\mathbf{l}} &\doteq \mathbb{E}\big(\widehat{\langle F,\tilde{\varphi}_{\mathbf{j},\mathbf{k}}\rangle}\widehat{\langle F,\tilde{\varphi}_{\mathbf{j},\mathbf{l}}\rangle}\big) - \langle F,\tilde{\varphi}_{\mathbf{j},\mathbf{k}}\rangle\langle F,\tilde{\varphi}_{\mathbf{j},\mathbf{l}}\rangle \\
&= 2^{-j_2}\mathbb{E}\big[\big(\tilde{\varphi}_{j_1,k_1}(S) - \tilde{\varphi}_{j_1,k_1}(S)\tilde{\Phi}(2^{j_2}T-k_2)\big)\big(\tilde{\varphi}_{j_1,l_1}(S) - \tilde{\varphi}_{j_1,l_1}(S)\tilde{\Phi}(2^{j_2}T-l_2)\big)\big] \\
&\quad -2^{-j_2}\mathbb{E}\big(\tilde{\varphi}_{j_1,k_1}(S) - \tilde{\varphi}_{j_1,k_1}(S)\tilde{\Phi}(2^{j_2}T-k_2)\big)\,\mathbb{E}\big(\tilde{\varphi}_{j_1,l_1}(S) - \tilde{\varphi}_{j_1,l_1}(S)\tilde{\Phi}(2^{j_2}T-l_2)\big) \\
&= 2^{-j_2}\Big\{ \operatorname{Cov}\big(\tilde{\varphi}_{j_1,k_1}(S),\tilde{\varphi}_{j_1,l_1}(S)\big) + \operatorname{Cov}\big(\tilde{\varphi}_{j_1,k_1}(S),\tilde{\varphi}_{j_1,l_1}(S)\tilde{\Phi}(2^{j_2}T-l_2)\big) \\
&\quad + \operatorname{Cov}\big(\tilde{\varphi}_{j_1,k_1}(S)\tilde{\Phi}(2^{j_2}T-k_2),\tilde{\varphi}_{j_1,l_1}(S)\big) + \operatorname{Cov}\big(\tilde{\varphi}_{j_1,k_1}(S)\tilde{\Phi}(2^{j_2}T-k_2),\tilde{\varphi}_{j_1,l_1}(S)\tilde{\Phi}(2^{j_2}T-l_2)\big)\Big\}.
\end{aligned} \tag{B.24}
$$

We study separately each summand in (B.24). $\operatorname{Cov}\big(\tilde{\varphi}_{j_1,k_1}(S),\tilde{\varphi}_{j_1,l_1}(S)\big)$ is a one dimensional term and has already been studied in Dechevsky and Penev (1998), who find:

$$
\begin{aligned}
&\big|\operatorname{Cov}\big(\tilde{\varphi}_{j_1,k_1}(S),\tilde{\varphi}_{j_1,l_1}(S)\big)\big| \\
&\leqslant f(x) + \left|\int_{-a}^{a} dv\big(f(2^{-j_1}v+k_1) - f(x)\big)\,\tilde{\varphi}(v)\tilde{\varphi}(v+k_1-l_1)\right| + \left|\int_{-a}^{a} dv\big(f(2^{-j_1}v+k_1) - f(x)\big)\,\tilde{\varphi}(v)\right|.
\end{aligned} \tag{B.25}
$$

As for the other terms we have:

$$
\begin{aligned}
&\operatorname{Cov}\big(\tilde{\varphi}_{j_1,k_1}(S),\tilde{\varphi}_{j_1,l_1}(S)\tilde{\Phi}(2^{j_2}T-l_2)\big) \\
&= \int_{\mathbb{R}^2} dsdt f(s,t)\tilde{\varphi}_{j_1,k_1}(s)\tilde{\varphi}_{j_1,l_1}(s)\tilde{\Phi}(2^{j_2}t-l_2) \\
&\quad - \int_{\mathbb{R}} ds f(s)\tilde{\varphi}_{j_1,k_1}(s)\int_{\mathbb{R}^2} dsdt f(s,t)\tilde{\varphi}_{j_1,l_1}(s)\tilde{\Phi}(2^{j_2}t-l_2) \\
&= \int_{\mathbb{R}} ds\int_{-a}^{a} d(2^{-j_2}\tau) f(s,2^{-j_2}(\tau+l_2))\tilde{\varphi}_{j_1,k_1}(s)\tilde{\varphi}_{j_1,l_1}(s)\tilde{\Phi}(\tau) \\
&\quad - \int_{\mathbb{R}} ds f(s)\tilde{\varphi}_{j_1,k_1}(s)\int_{\mathbb{R}} ds\int_{-a}^{a} d(2^{-j_2}\tau) f(s,2^{-j_2}(\tau+l_2))\tilde{\varphi}_{j_1,l_1}(s)\tilde{\Phi}(\tau)
\end{aligned}
$$

which gives after integration by parts:

$$
\begin{aligned}
&= \int_{\mathbb{R}} ds f(s)\tilde{\varphi}_{j_1,k_1}(s)\tilde{\varphi}_{j_1,l_1}(s) - \int_{\mathbb{R}} ds\int_{-a}^{a} d\tau F(s,2^{-j_2}(\tau+l_2))\tilde{\varphi}_{j_1,k_1}(s)\tilde{\varphi}_{j_1,l_1}(s)\tilde{\varphi}(\tau) \\
&\quad - \int_{\mathbb{R}} ds f(s)\tilde{\varphi}_{j_1,k_1}(s)\int_{\mathbb{R}} ds f(s)\tilde{\varphi}_{j_1,l_1}(s) \\
&\quad + \int_{\mathbb{R}} ds f(s)\tilde{\varphi}_{j_1,k_1}(s)\int_{\mathbb{R}} ds\int_{-a}^{a} d\tau F(s,2^{-j_2}(\tau+l_2))\tilde{\varphi}_{j_1,l_1}(s)\tilde{\varphi}(\tau)
\end{aligned}
$$

$$
\begin{aligned}
&= \operatorname{Cov}\big(\tilde{\varphi}_{j_1,k_1}(S),\tilde{\varphi}_{j_1,l_1}(S)\big) - \int_{\mathbb{R}} ds\int_{-a}^{a} d\tau F(s,2^{-j_2}(\tau+l_2))\tilde{\varphi}_{j_1,k_1}(s)\tilde{\varphi}_{j_1,l_1}(s)\tilde{\varphi}(\tau) \\
&\quad + \int_{\mathbb{R}} ds f(s)\tilde{\varphi}_{j_1,k_1}(s)\int_{\mathbb{R}} ds\int_{-a}^{a} d\tau F(s,2^{-j_2}(\tau+l_2))\tilde{\varphi}_{j_1,l_1}(s)\tilde{\varphi}(\tau).
\end{aligned} \tag{B.26}
$$

In a similar way we have that

$$
\begin{aligned}
&\operatorname{Cov}\big(\tilde{\varphi}_{j_1,k_1}(S)\tilde{\Phi}(2^{j_2}T-k_2),\tilde{\varphi}_{j_1,l_1}(S)\big) = \operatorname{Cov}\big(\tilde{\varphi}_{j_1,k_1}(S),\tilde{\varphi}_{j_1,l_1}(S)\big) \\
&\quad - \int_{\mathbb{R}} ds\int_{-a}^{a} d\tau F(s,2^{-j_2}(\tau+k_2))\tilde{\varphi}_{j_1,k_1}(s)\tilde{\varphi}_{j_1,l_1}(s)\tilde{\varphi}(\tau) \\
&\quad + \int_{\mathbb{R}} ds f(s)\tilde{\varphi}_{j_1,l_1}(s)\int_{\mathbb{R}} ds\int_{-a}^{a} d\tau F(s,2^{-j_2}(\tau+k_2))\tilde{\varphi}_{j_1,k_1}(s)\tilde{\varphi}(\tau).
\end{aligned} \tag{B.27}
$$

And finally

$$
\begin{aligned}
&\operatorname{Cov}\left(\tilde{\varphi}_{j_1,k_1}(S)\tilde{\Phi}(2^{j_2}T-k_2),\tilde{\varphi}_{j_1,l_1}(S)\tilde{\Phi}(2^{j_2}T-l_2)\right)\\
&=\int_{\mathbb{R}} ds\int_{-a}^{a} d(2^{-j_2}\tau)f(s,2^{-j_2}(\tau+k_2))\tilde{\varphi}_{j_1,k_1}(s)\tilde{\Phi}(\tau)\tilde{\varphi}_{j_1,l_1}(s)\tilde{\Phi}(\tau+k_2-l_2)\\
&-\int_{\mathbb{R}} ds\int_{-a}^{a} d(2^{-j_2}\tau)f(s,2^{-j_2}(\tau+k_2))\tilde{\varphi}_{j_1,k_1}\tilde{\Phi}(\tau)\cdot\int_{\mathbb{R}} ds\int_{-a}^{a} d(2^{-j_2}\tau)f(s,2^{-j_2}(\tau+l_2))\tilde{\varphi}_{j_1,l_1}(s)\tilde{\Phi}(\tau).
\end{aligned}
\tag{B.28}
$$

We consider the two terms of (B.28) separately.

$$
\begin{aligned}
\mathbb{E}\big(\tilde{\varphi}_{j_1,k_1}(S)\tilde{\Phi}(2^{j_2}T-k_2)\tilde{\varphi}_{j_1,l_1}(S)\tilde{\Phi}(2^{j_2}T-l_2)\big) &= \int_{\mathbb{R}} dsf(s)\tilde{\varphi}_{j_1,k_1}(s)\tilde{\varphi}_{j_1,l_1}(s)\\
&-\int_{\mathbb{R}} ds\int_{-a}^{a} d\tau F(s,2^{-j_2}(\tau+k_2))\tilde{\varphi}_{j_1,k_1}(s)\tilde{\varphi}(\tau)\tilde{\varphi}_{j_1,l_1}(s)\tilde{\Phi}(\tau+k_2-l_2)\\
&-\int_{\mathbb{R}} ds\int_{-a}^{a} d\tau F(s,2^{-j_2}(\tau+l_2))\tilde{\varphi}_{j_1,k_1}(s)\tilde{\Phi}(\tau+l_2-k_2)\tilde{\varphi}_{j_1,l_1}(s)\tilde{\varphi}(\tau);
\end{aligned}
\tag{B.29}
$$

$$
\begin{aligned}
&\mathbb{E}\big(\tilde{\varphi}_{j_1,k_1}(S)\tilde{\Phi}(2^{j_2}T-k_2)\big)\mathbb{E}\big(\tilde{\varphi}_{j_1,l_1}(S)\tilde{\Phi}(2^{j_2}T-l_2)\big)\\
&=\left[\int_{\mathbb{R}} dsf(s)\tilde{\varphi}_{j_1,k_1}(s)-\int_{\mathbb{R}} ds\int_{-a}^{a} d\tau F(s,2^{-j_2}(\tau+k_2))\tilde{\varphi}_{j_1,k_1}(s)\tilde{\Phi}(\tau)\right]\\
&\cdot\left[\int_{\mathbb{R}} dsf(s)\tilde{\varphi}_{j_1,l_1}(s)-\int_{\mathbb{R}} ds\int_{-a}^{a} d\tau F(s,2^{-j_2}(\tau+l_2))\tilde{\varphi}_{j_1,l_1}(s)\tilde{\varphi}(\tau)\right]\\
&=\int_{\mathbb{R}} dsf(s)\tilde{\varphi}_{j_1,k_1}(s)\cdot\int_{\mathbb{R}} dsf(s)\tilde{\varphi}_{j_1,l_1}(s)\\
&-\int_{\mathbb{R}} dsf(s)\tilde{\varphi}_{j_1,k_1}(s)\cdot\int_{\mathbb{R}} ds\int_{-a}^{a} d\tau F(s,2^{-j_2}(\tau+l_2))\tilde{\varphi}_{j_1,l_1}(s)\tilde{\varphi}(\tau)\\
&-\int_{\mathbb{R}} dsf(s)\tilde{\varphi}_{j_1,l_1}(s)\cdot\int_{\mathbb{R}} ds\int_{-a}^{a} d\tau F(s,2^{-j_2}(\tau+k_2))\tilde{\varphi}_{j_1,k_1}(s)\tilde{\varphi}(\tau)\\
&+\int_{\mathbb{R}} ds\int_{-a}^{a} d\tau F(s,2^{-j_2}(\tau+k_2))\tilde{\varphi}_{j_1,k_1}(s)\tilde{\varphi}(\tau)\cdot\int_{\mathbb{R}} ds\int_{-a}^{a} d\tau F(s,2^{-j_2}(\tau+l_2))\tilde{\varphi}_{j_1,l_1}(s)\tilde{\varphi}(\tau).
\end{aligned}
\tag{B.30}
$$

Putting equations from (B.24) through to (B.30) together gives:

$$
\begin{aligned}
2^{j_2}\Delta_{\mathbf{j};\mathbf{k},\mathbf{l}} &= 4\cdot\operatorname{Cov}\left(\tilde{\varphi}_{j_1,k_1}(S),\tilde{\varphi}_{j_1,l_1}(S)\right)-\int_{\mathbb{R}} ds\int_{-a}^{a} d\tau F(s,2^{-j_2}(\tau+l_2))\tilde{\varphi}_{j_1,k_1}(s)\tilde{\varphi}_{j_1,l_1}(s)\tilde{\varphi}(\tau)\\
&-\int_{\mathbb{R}} ds\int_{-a}^{a} d\tau F(s,2^{-j_2}(\tau+k_2))\tilde{\varphi}_{j_1,k_1}(s)\tilde{\varphi}_{j_1,l_1}(s)\tilde{\varphi}(\tau)+\\
&+2\int_{\mathbb{R}} dsf(s)\tilde{\varphi}_{j_1,k_1}(s)\cdot\int_{\mathbb{R}} ds\int_{-a}^{a} d\tau F(s,2^{-j_2}(\tau+l_2))\tilde{\varphi}_{j_1,l_1}(s)\tilde{\varphi}(\tau)\\
&+2\int_{\mathbb{R}} dsf(s)\tilde{\varphi}_{j_1,l_1}(s)\cdot\int_{\mathbb{R}} ds\int_{-a}^{a} d\tau F(s,2^{-j_2}(\tau+k_2))\tilde{\varphi}_{j_1,k_1}(s)\tilde{\varphi}(\tau)\\
&-\int_{\mathbb{R}} ds\int_{-a}^{a} d\tau F(s,2^{-j_2}(\tau+k_2))\tilde{\varphi}_{j_1,k_1}(s)\tilde{\varphi}(\tau)\tilde{\varphi}_{j_1,l_1}(s)\tilde{\Phi}(\tau+k_2-l_2)\\
&-\int_{\mathbb{R}} ds\int_{-a}^{a} d\tau F(s,2^{-j_2}(\tau+l_2))\tilde{\varphi}_{j_1,k_1}(s)\tilde{\Phi}(\tau+l_2-k_2)\tilde{\varphi}_{j_1,l_1}(s)\tilde{\varphi}(\tau)\\
&-\int_{\mathbb{R}} ds\int_{-a}^{a} d\tau F(s,2^{-j_2}(\tau+k_2))\tilde{\varphi}_{j_1,k_1}(s)\tilde{\varphi}(\tau)\cdot\int_{\mathbb{R}} ds\int_{-a}^{a} d\tau F(s,2^{-j_2}(\tau+l_2))\tilde{\varphi}_{j_1,l_1}(s)\tilde{\varphi}(\tau).
\end{aligned}
$$

We then add the following terms:

$$\pm 2\cdot 2^{-\frac{j_1}{2}} f(x)\int_{\mathbb{R}} ds \int_{-a}^{a} d\tau F(s,2^{-j_2}(\tau+l_2))\,\tilde{\varphi}_{j_1,l_1}(s)\,\tilde{\varphi}(\tau)$$
$$\pm 2\cdot 2^{-\frac{j_1}{2}} f(x)\int_{\mathbb{R}} ds \int_{-a}^{a} d\tau F(s,2^{-j_2}(\tau+k_2))\,\tilde{\varphi}_{j_1,k_1}(s)\,\tilde{\varphi}(\tau)$$
$$\pm F(x,y)\int_{\mathbb{R}} ds\,\tilde{\varphi}_{j_1,k_1}(s)\,\tilde{\varphi}_{j_1,l_1}(s) \pm 2^{-\frac{j_1}{2}}\int_{\mathbb{R}} ds \int_{-a}^{a} d\tau F(s,2^{-j_2}(\tau+l_2))\,\tilde{\varphi}_{j_1,l_1}(s)\,\tilde{\varphi}(\tau)$$
$$\pm F(x,y)\int_{\mathbb{R}} ds\,\tilde{\varphi}_{j_1,k_1}(s)\,\tilde{\varphi}_{j_1,l_1}(s)\int_{-a}^{a} d\tau\,\tilde{\varphi}(\tau)\tilde{\Phi}(\tau+k_2-l_2)$$
$$\pm F(x,y)\int_{\mathbb{R}} ds\,\tilde{\varphi}_{j_1,k_1}(s)\,\tilde{\varphi}_{j_1,l_1}(s)\int_{-a}^{a} d\tau\,\tilde{\varphi}(\tau)\tilde{\Phi}(\tau+l_2-k_2),$$

so that

$$\begin{aligned}
2^{j_2}\Delta_{\mathbf{j};\mathbf{k},\mathbf{l}} &= 4\cdot \operatorname{Cov}\big(\tilde{\varphi}_{j_1,k_1}(S),\tilde{\varphi}_{j_1,l_1}(S)\big)\\
&+2\int_{\mathbb{R}} ds\big(f(s)-f(x)\big)\,\tilde{\varphi}_{j_1,k_1}(s)\int_{-a}^{a} d\tau F(s,2^{-j_2}(\tau+l_2))\,\tilde{\varphi}_{j_1,l_1}(s)\,\tilde{\varphi}(\tau)\\
&+2\int_{\mathbb{R}} ds\big(f(s)-f(x)\big)\,\tilde{\varphi}_{j_1,l_1}(s)\int_{-a}^{a} d\tau F(s,2^{-j_2}(\tau+k_2))\,\tilde{\varphi}_{j_1,k_1}(s)\,\tilde{\varphi}(\tau)\\
&-\int_{\mathbb{R}} ds\int_{-a}^{a} d\tau\big[F(s,2^{-j_2}(\tau+l_2))-F(x,y)\big]\,\tilde{\varphi}_{j_1,k_1}(s)\,\tilde{\varphi}_{j_1,l_1}(s)\varphi(\tau)\\
&-\int_{\mathbb{R}} ds\int_{-a}^{a} d\tau\big[F(s,2^{-j_2}(\tau+k_2))-F(x,y)\big]\,\tilde{\varphi}_{j_1,k_1}(s)\,\tilde{\varphi}_{j_1,l_1}(s)\varphi(\tau)\\
&-\int_{\mathbb{R}} ds\int_{-a}^{a} d\tau\big[F(s,2^{-j_2}(\tau+l_2))-F(x,y)\big]\,\tilde{\varphi}_{j_1,k_1}(s)\,\tilde{\varphi}_{j_1,l_1}(s)\varphi(\tau)\tilde{\Phi}(\tau+k_2-l_2)\\
&-\int_{\mathbb{R}} ds\int_{-a}^{a} d\tau\big[F(s,2^{-j_2}(\tau+k_2))-F(x,y)\big]\,\tilde{\varphi}_{j_1,k_1}(s)\,\tilde{\varphi}_{j_1,l_1}(s)\varphi(\tau)\tilde{\Phi}(\tau+l_2-k_2)\\
&-\int_{\mathbb{R}} ds\int_{-a}^{a} d\tau\big[F(s,2^{-j_2}(\tau+k_2))-F(x,y)\big]\,\tilde{\varphi}_{j_1,k_1}(s)\varphi(\tau)\int_{\mathbb{R}} ds\int_{-a}^{a} d\tau F(s,2^{-j_2}(\tau+l_2))\,\tilde{\varphi}_{j_1,l_1}(s)\varphi(\tau)\\
&+2\cdot 2^{-\frac{j_1}{2}} f(x)\int_{\mathbb{R}} ds\int_{-a}^{a} d\tau F(s,2^{-j_2}(\tau+l_2))\,\tilde{\varphi}_{j_1,l_1}(s)\,\tilde{\varphi}(\tau)\\
&+2\cdot 2^{-\frac{j_1}{2}} f(x)\int_{\mathbb{R}} ds\int_{-a}^{a} d\tau F(s,2^{-j_2}(\tau+k_2))\,\tilde{\varphi}_{j_1,k_1}(s)\,\tilde{\varphi}(\tau)\\
&-2F(x,y)\int_{\mathbb{R}} ds\,\tilde{\varphi}_{j_1,k_1}(s)\,\tilde{\varphi}_{j_1,l_1}(s)\\
&-2F(x,y)\int_{\mathbb{R}} ds\,\tilde{\varphi}_{j_1,k_1}(s)\,\tilde{\varphi}_{j_1,l_1}(s)\int_{-a}^{a} d\tau\,\tilde{\varphi}(\tau)\big[\tilde{\Phi}(\tau+k_2-l_2)+\tilde{\Phi}(\tau+l_2-k_2)\big]\\
&-2^{-\frac{j_1}{2}}F(x,y)\int_{\mathbb{R}} ds\int_{-a}^{a} d\tau F(s,2^{-j_2}(\tau+l_2))\,\tilde{\varphi}_{j_1,l_1}(s)\varphi(\tau).
\end{aligned}$$

To bound the latter expression we make use of the following inequalities:

$$F(x,y)\leqslant f(x),\qquad \forall y;$$
$$\int_{\mathbb{R}} ds\,\tilde{\varphi}_{j_1,k_1}(s)\,\tilde{\varphi}_{j_1,l_1}(s)\leqslant \int_{\mathbb{R}} ds\,\tilde{\varphi}_{j_1,k_1}(s)^2\leqslant \int_{\mathbb{R}} d\upsilon\,\tilde{\varphi}(\upsilon)=1;$$
$$\int_{\mathbb{R}} dsf(s)\,\tilde{\varphi}_{j_1,k_1}(s)\leqslant 2^{\frac{j_1}{2}}\int_{\mathbb{R}} dsf(s)\,\tilde{\varphi}(2^{j_1}s-k_1)\leqslant 2^{\frac{j_1}{2}};$$

which also implies:

$$\int_{\mathbb{R}} ds\int_{-a}^{a} d\tau F(s,2^{-j_2}(\tau+k_2))\,\tilde{\varphi}_{j_1,k_1}(s)\,\tilde{\varphi}(\tau)\leqslant \int_{\mathbb{R}} dsf(s)\,\tilde{\varphi}_{j_1,k_1}(s)\int_{-a}^{a} d\tau\tilde{\varphi}(\tau)\leqslant 2^{\frac{j_1}{2}}.$$

Keeping in mind (B.25) we can then write, after a change of variable:

$$
\begin{aligned}
2^{j_2}|\Delta_{\mathbf{j};\mathbf{k},\mathbf{l}}| \leqslant\, & 4\cdot\int_{-a}^{a} dv\big|f(2^{-j_1}v+k_1)-f(x)\big| \\
&+4\cdot\int_{-a}^{a} dv\big|f(2^{-j_1}v+k_1)-f(x)\big|\,\tilde{\varphi}(v)\,\tilde{\varphi}(v)\tilde{\varphi}(v+k_1-l_1) \\
&+2\cdot\int_{-a}^{a} d(2^{-j_1}v)\big|f(2^{-j_1}v+k_1)-f(x)\big|\,2^{\frac{j_1}{2}}\tilde{\varphi}(v)\cdot 2^{\frac{j_1}{2}}+2\int_{-a}^{a} dv\big|f(2^{-j_1}v+l_1)-f(x)\big|\,\tilde{\varphi}(v) \\
&+2\int_{-a}^{a} dv\int_{-a}^{a} d\tau\big|F(2^{-j_1}v+k_1,2^{-j_2}(\tau+k_2))-F(x,y)\big|\,\tilde{\varphi}(v)\,\tilde{\varphi}(v+k_1-l_1)\varphi(\tau) \\
&+2\int_{-a}^{a} dv\int_{-a}^{a} d\tau\big|F(2^{-j_1}v+k_1,2^{-j_2}(\tau+l_2))-F(x,y)\big|\,\tilde{\varphi}(v)\,\tilde{\varphi}(v+k_1-l_1)\varphi(\tau) \\
&+2\int_{-a}^{a} dv\int_{-a}^{a} d\tau\big|F(2^{-j_1}v+k_1,2^{-j_2}(\tau+k_2))-F(x,y)\big|\,\tilde{\varphi}(v)\varphi(\tau)+4f(x)+4f(x)+f(x).
\end{aligned}
$$

The densities figuring in the above expressions are obtained remembering that $\int_{-a}^{a} d\tau\,\tilde{\varphi}(\tau)\big[\tilde{\Phi}(\tau+k_2-l_2)+\tilde{\Phi}(\tau+l_2-k_2)\big]\equiv 1$ (see Dechevsky and Penev (1998) for the proof) and considering that, if we write $F(x,y)\leqslant cf(x)$ with $0<c\leqslant 1$ then

$$
f(x)\Big|2^{-\frac{j_1}{2}}\int_{\mathbb{R}} ds\int_{-a}^{a} d\tau F(s,2^{-j_2}(\tau+l_2))\,\tilde{\varphi}_{j_1,l_1}(s)\,\tilde{\varphi}(\tau)-c\int_{\mathbb{R}} ds\,\tilde{\varphi}_{j_1,k_1}(s)\,\tilde{\varphi}_{j_1,l_1}(s)\Big|\leqslant f(x)
$$

since both terms in the absolute value are smaller than 1. We can then express the quantity $\sigma(x,y)$ introduced in (B.23) as:

$$
\begin{aligned}
\sigma^2(x,y)=&\sum_{\mathbf{k}\in\mathbb{Z}^2}\sum_{\mathbf{l}\in\mathbb{Z}^2}\Delta_{\mathbf{j};\mathbf{k},\mathbf{l}}\varphi_{\mathbf{j},\mathbf{k}}(x,y)\varphi_{\mathbf{j},\mathbf{l}}(x,y)\leqslant\sup_{\mathbf{k},\mathbf{l}}|\Delta_{\mathbf{j};\mathbf{k},\mathbf{l}}|\sum_{\mathbf{k}\in\mathbb{Z}^2}\sum_{\mathbf{l}\in\mathbb{Z}^2}\varphi_{\mathbf{j},\mathbf{k}}(x,y)\varphi_{\mathbf{j},\mathbf{l}}(x,y) \\
=&\sup_{\mathbf{k},\mathbf{l}}|\Delta_{\mathbf{j};\mathbf{k},\mathbf{l}}|\cdot 2^{j_1}\cdot 2^{j_2}=2^{j_1}\Delta(x,y);
\end{aligned}
\tag{B.31}
$$

where $\Delta(x,y)=2^{j_2}\sup_{\mathbf{k},\mathbf{l}}|\Delta_{\mathbf{j};\mathbf{k},\mathbf{l}}|$. Following the procedure outlined at the beginning of the section it is possible to bound the term :

$$
\sup_y\frac{1}{T^\rho}\Big\|\sum_{i=1}^{T}\mathbb{E}\big(Z_i^2\big)\Big\|_{p/2}^{\rho/2}\leqslant\sup_y T^{-\frac{\rho}{2}}\Big(\int_{\mathbb{R}} dx\big(\sigma^2(x,y)\big)^{\frac{p}{2}}\Big)^{\frac{2}{p}\cdot\frac{\rho}{2}}\leqslant\sup_y\left(\frac{2^{j_1}}{T}\right)^{\rho/2}\|\,\Delta(\cdot,y)\|_{p/2}^{\rho/2}.
\tag{B.32}
$$

Let us remark that when taking the L_p-norm in (B.32), the terms contained in $\Delta(x,y)$ will be controlled by the modulus of smoothness in the x direction of $F(x,y)$. We recall that the $f(x)$ that enters the definition of some terms of $\Delta(x,y)$ is the $\lim_{y\to\infty}F(x,y)$, so that $\omega^1_{\mathbf{e}_x}\big(f,h\big)_p\leqslant\sup_y\omega^1_{\mathbf{e}_x}\big(F(\cdot,y),h\big)_p$. The terms containing $F(x,y)-F(s,t)$ can be split in two terms by adding $\pm F(x,t)$ and we obtain an increment in the x direction and one in the y direction. The increments in the x direction $F(x,t)-F(s,t)$ are controlled by the modulus $\omega^1_{\mathbf{e}_x}\big(F(\cdot,y),h\big)_p$, while the increments in the y direction $F(x,y)-F(x,t)$ can be controlled by $\|f(\cdot,y)\|_p$ through the inequality, $\sup_y\sup_{|t|<h}\|\Delta_{\mathbf{e}_y,t}F(\cdot,y)\|_p\leqslant h\sup_y\|f(\cdot,y)\|_p$. Substituting the explicit expression for $\Delta(x,y)$, using Minkowsky's generalized inequality and the properties of the moduli of smoothness gives:

$$
\frac{1}{T^\rho}\Big\|\sum_{t=1}^{T}\mathbb{E}\big(Z_i^2\big)\Big\|_{p/2}^{\rho/2}\leqslant\left(\frac{2^{j_1}}{T}\right)^{\rho/2}\Big(d_1\|f(\cdot,y)\|_{p/2}^{\rho/2}+d_2(a)\,\omega^1_{\mathbf{e}_x}\big(f(\cdot,y),2^{1-j_1}a\big)_{p/2}^{\rho/2}\Big).
\tag{B.33}
$$

To choose the value of ρ, we refer to Appendix E. In (B.33) only p the parameter is involved through the $L_{p/2}$ norm of the quantity $\Delta(x,y)$, the choice has to be made in a way that

$$
\begin{aligned}
\frac{\rho^*}{2}=&\frac{1}{1+\log_2(c_p)}=\frac{1}{1+\log_2(\max\{1,2^{2/p-1}\})} \\
=&\frac{1}{1+(\max\{0,2/p-1\})}=\frac{1}{\max\{1,\frac{2}{p}\}}=\min\{p/2,1\}.
\end{aligned}
$$

c_p is constant in the quasi-triangular inequality, see again Appendix E. Since for every density $f \in L_1$, we can give for (B.33) the more general expression:

$$
\begin{aligned}
\sup_y \frac{1}{T^{\rho^*}} \Big\| \sum_{t=1}^{T} \mathbb{E}\big(Z^2(\cdot,y)\big) \Big\|_{p/2}^{\rho^*/2} \leqslant \sup_y \left(\frac{2^{j_1}}{T}\right)^{\rho^*/2} & \bigg(d_1 \max\Big\{ \big\|f(\cdot,y)\big\|_1, \big\|f(\cdot,y)\big\|_{p/2} \Big\}^{\rho^*/2} \\
& + d_2(a) \max\Big\{ \omega^1_{\mathbf{e}_x}\big(f(\cdot,y), 2^{1-j_1}a\big)_1, \omega^1_{\mathbf{e}_x}\big(f(\cdot,y), 2^{1-j_1}a\big)_{p/2} \Big\}^{\rho^*/2} \bigg).
\end{aligned} \tag{B.34}
$$

The condition, for $1 < p < 2$, that $\sup_y \sup_{0 \leqslant t \leqslant h} \int_{-\infty}^{+\infty} dx \Big(\int_0^1 d\alpha\, f(x+\alpha t, y) \Big)^{p/2} < +\infty$, ensures that the modulus $\omega_\lambda(f,h) \to_{h\to 0} 0$ even if $\lambda < 1$. The proof is to be found in Dechevsky and Penev (1998). □

B. Study of V_1 *term*: We have now to deal with the covariance part V_1 of (B.22). To do this we will initially split the summation in two parts:

$$
\begin{aligned}
& \sup_y T^{-\rho/2} \Big\| \sum_{p=1}^{T-1} \Big(1 - \frac{p}{T}\Big) \Big(\mathbb{E}\big| Z_T(\cdot,y) Z_{T-p}(\cdot,y) \big| \Big) \Big\|_{p/2}^{\rho/2} \\
& \leqslant \sup_y \frac{1}{T^{\rho/2}} \Big\| \sum_{p=1}^{n_T} \Big(1 - \frac{p}{T}\Big) \Big(\mathbb{E}\big| Z_T Z_{T-p} \big| \Big) \Big\|_{p/2}^{\rho/2} \\
& \quad + \sup_y \frac{1}{T^{\rho/2}} \Big\| \sum_{p=n_T+1}^{T-1} \Big(1 - \frac{p}{T}\Big) \Big(\mathbb{E}\big| Z_T Z_{T-p} \big| \Big) \Big\|_{p/2}^{\rho/2} \\
& \doteq S_1 + S_2.
\end{aligned} \tag{B.35}
$$

Here the explicit dependence $Z(x,y)$ on (x,y) is omitted for readability. We start by tackling part S_1 of (B.35). We recall *Assumption 1* (4.3): There exists a constant M such that

$$
\int_{-\infty}^{+\infty} dF_{(X_0,Y_0),(X_s,Y_s)} |T_{(X_0,Y_0),(X_s,Y_s)}| \leqslant M \int_{-\infty}^{+\infty} dF_{(X_t,Y_t)} \int_{-\infty}^{+\infty} dF_{(X_t,Y_t)} |T_{(X_0,Y_0),(X_s,Y_s)}|
$$

Then:

$$
\begin{aligned}
& \big| \operatorname{Cov}\big(Z_T(x,y) Z_{T-p}(x,y)\big) \big| = \int dF_{Z_T, Z_{T-p}} \big| Z_T(x,y) Z_{T-p}(x,y) \big| \\
& \leqslant M \int dF_t \int dF_t \big| Z_t(x,y) \big| \big| Z_t(x,y) \big| = M \Big(\int dF_t \big| Z_t(x,y) \big| \Big)^2 = M \Big(\mathbb{E}\big| Z_t(x,y) \big| \Big)^2.
\end{aligned}
$$

Now

$$
\begin{aligned}
& \mathbb{E}\big| Z_t(x,y) \big| \\
& = 2^{-j_2/2} \mathbb{E} \Big| \sum_{\mathbf{k}} \Big\{ 2^{-j_2/2} \tilde{\varphi}_{j_1 k_1}(X_t) \big(1 - \tilde{\Phi}(2^{j_2} X_{t-1} - k_2)\big) \\
& \qquad - \mathbb{E}\big[\tilde{\varphi}_{j_1 k_1}(X_t) \big(1 - \tilde{\Phi}(2^{j_2} X_{t-1} - k_2)\big) \big] \Big\} \varphi_{\mathbf{jk}}(x,y) \Big| \\
& \leqslant 2^{-\frac{j_2}{2}} \sum_{\mathbf{k}} \int_{\Xi_{\mathbf{k}}} du dv f(u,v) \Big| \big\{ \tilde{\varphi}_{j_1 k_1}(u) \big(1 - \tilde{\Phi}(2^{j_2} v - k_2)\big) \big\} \\
& \qquad - \int_{\Xi_{\mathbf{k}}} ds dt\, f(s,t) \big\{ \tilde{\varphi}_{j_1 k_1}(s) \big(1 - \tilde{\Phi}(2^{j_2} t - k_2)\big) \big\} \Big| \varphi_{\mathbf{jk}}(x,y) \\
& = 2^{-\frac{j_1}{2}} 2^{-\frac{j_2}{2}} \sum_{\mathbf{k}} \int_{-a}^{a} du' \int_{\Xi_{k_2}} dv f(2^{j_1}(u' + k_1), v) \Big| \tilde{\varphi}(u') \big(1 - \tilde{\Phi}(2^{j_2} v - k_2)\big) \\
& \qquad - 2^{-j_1/2} \int_{\Xi_{\mathbf{k}}} ds dt f(s,t)\, 2^{j_1/2} \tilde{\varphi}(2^j s - k_1) \big(1 - \tilde{\Phi}(2^{j_2} t - k_2)\big) \Big| \varphi_{\mathbf{jk}}(x,y).
\end{aligned}
$$

If we consider the two terms inside the absolute value, it can be readily seen that:

$$0<\tilde{\varphi}(\xi)\big(1-\tilde{\Phi}(\zeta)\big)<1, \qquad \text{and} \qquad 0<\int_{\Xi_{\mathbf{k}}} d\xi d\zeta\, f(\xi,\zeta)\,\tilde{\varphi}(\xi)\big(1-\tilde{\Phi}(\zeta)\big)<1,$$

so that their difference too belongs to $[0,1]$. We have then that:

$$\begin{aligned} M\sum_{p=1}^{n_T}\left(1-\frac{p}{T}\right)\Big(\mathbb{E}\big|Z_t(x,y)\big|\Big)^2 &\leqslant M\,n_T\cdot 2^{-j_1-j_2}\Bigg(\sum_{\mathbf{k}}\int_{-a}^{a}du'\int_{\Xi_{k_2}}dv f(2^{j_1}(u'+k_1),v)\varphi_{\mathbf{jk}}(x,y)\Bigg)^2 \\ = M\,n_T\cdot 2^{-j_1-j_2}&\Bigg(\sum_{\mathbf{k}}\bigg\{\int_{-a}^{a}du'\int_{\Xi_{k_2}}dv\big[f(2^{j_1}(u'+k_1),v)-f(x,y)\big]+f(x,y)\bigg\}\varphi_{\mathbf{jk}}(x,y)\Bigg)^2 \\ = Mn_T\cdot&\Bigg(\sum_{\mathbf{k}}\bigg\{\int_{-a}^{a}du'\int_{\Xi_{k_2}}dv\big[f(2^{j_1}(u'+k_1),v)-f(x,y)\big]+f(x,y)\bigg\}\varphi(2^{j_1}x-k_1)\varphi(2^{j_2}y-k_2)\Bigg)^2. \end{aligned} \tag{B.36}$$

It follows, taking the norm, that:

$$S_1\leqslant M\left(\frac{n_T}{T}\right)^{\rho/2}\sup_y\Big\{\omega^1_{\mathbf{e}_x}\big(f(\cdot,y),2^{1-j_1}a\big)^\rho_p+\|f(\cdot,y)\|^\rho_p\Big\}=O\!\left(\frac{n_T}{T}\right)^{\rho/2}. \tag{B.37}$$

Now we continue by considering part S_2 of (B.35). This term will be studied using the α - mixing property for the process, taking advantage of Davydov's inequality for α-mixing processes:

$$|\operatorname{Cov}(X,Y)|\leqslant 2\frac{r}{r-2}(2\alpha)^{1-\frac{1}{r}}(\mathbb{E}|X|^r)^{\frac{1}{r}}(\mathbb{E}|Y|^r)^{\frac{1}{r}},$$

with $r>2$ and provided that $\mathbb{E}|X|^r,\mathbb{E}|Y|^r<\infty$. So, again because of stationarity:

$$\Big|\sum_{p=n_T+1}^{T}\left(1-\frac{p}{T}\right)\operatorname{Cov}\big(Z_T(x,y)Z_{T-p}(x,y)\big)\Big|\leqslant\sum_{p=n_T+1}^{T}2\frac{r}{r-2}\big(2\alpha(p)\big)^{1-\frac{1}{r}}\Big(\mathbb{E}\big|Z_t(x,y)\big|^r\Big)^{2/r}. \tag{B.38}$$

Since $r>2$ and since $|Z_t(x)|<2^{j_1}$ uniformly in x we can write:

$$\mathbb{E}|Z\,|^r\leqslant 2^{j_1\cdot(r-2)}\,\mathbb{E}Z_t(x,y)^2,$$

so that the right hand side of (B.38) can be bounded by

$$\begin{aligned} C_r\cdot 2^{2\frac{r-2}{r}j_1}\big(\sigma^2(x,y)\big)^{2/r}\sum_{p=n_T+1}^{T}\alpha(p)^{1-\frac{1}{r}} &=C_r\cdot 2^{2\frac{r-2}{r}j_1}2^{\frac{2}{r}j_1}\Delta(x,y)^{\frac{2}{r}}\sum_{p=n_T+1}^{T}\alpha(p)^{1-\frac{1}{r}} \\ &\doteq C_r\cdot 2^{2\frac{r-1}{r}j_1}\Delta(x,y)^{\frac{2}{r}}S'_T(n_T). \end{aligned}$$

The S_2 term can be bounded in the following way:

$$S_2\leqslant\frac{\tilde{C}_r}{T^{-\rho/2}}\Big(2^{2\frac{r-1}{r}j_1}S'_T(n_T)\Big)^{\rho/2}\sup_y\big\|\Delta(\cdot,y)\big\|^{\rho/r}_{p/r}, \tag{B.39}$$

with $\tilde{C}_r$ obtained by collecting all terms not depending on j_1 and T. Then we can compare the asymptotic behavior of V_0 and S_2.

$$\frac{S_2}{\big(2^{j_1}/T\big)^{\rho/2}}=\tilde{C}_r{}'\frac{\Big(2^{2\frac{r-1}{r}j_1}S'_T(n_T)\Big)^{\rho/2}}{2^{j_1\cdot\rho/2}}=\tilde{C}_r{}'\left(\frac{2^{2\frac{r-1}{r}j_1}S'_T(n_T)}{2^{j_1}}\right)^{\rho/2}=\Big(2^{\frac{r-2}{r}j_1}S'_T(n_T)\Big)^{\rho/2}.$$

We observe that the ratio will tend to zero if $S'_T(n_T)=O\big(2^{-(\frac{r-2}{r}+\delta)j_1}\big)$ with $\delta>0$, so, since by (4.4) $S'_T(n_T)=O(n_T^{-1})$, we need to impose $n_T=O\big(2^{(\frac{r-2}{r}+\delta)j_1}\big)$. We remember that in order to have $S_1/V_0\to 0$ as $T\to\infty$, n_T was constrained by (B.37) to grow in a way that $n_T/2^{j_1}\underset{T}{\to}0$, *i.e.* $n_T=O\big(2^{j_1(1-\theta)}\big)$, with $\theta>0$. Now we notice that, since $r>2$, $0<\frac{r-2}{r}<1$, we can choose a δ such that $\frac{r-2}{r}+\delta<1$, that is $0<\delta<\frac{2}{r}<1$ and a $\theta=1-\frac{r-2}{r}-\delta$, that both satisfy the conditions for $S_1/V_0,S_2/V_0\to 0$ simultaneously as $T\to\infty$.

C Proof for Section 5

Theorem 2. As seen in Equation (5.3), we can write

$$\begin{aligned}\hat{Q}(p,\xi) - Q(p,\xi) &= \frac{1}{\tilde{f}_{\hat{Q},Q}(\xi)}\left\{\frac{F(\xi,\hat{Q}(p,\xi))}{f(\xi)} - \frac{\hat{F}(\xi,\hat{Q}(p,\xi))}{\hat{f}(\xi)}\right\} \\ &= \frac{1}{B(\xi)\cdot\tilde{f}_{\hat{Q},Q}(\xi)}\left\{F(\xi,\hat{Q}(p,\xi)) - \hat{F}(\xi,\hat{Q}(p,\xi))\right\} - \frac{A(\xi)}{B^2(\xi)\cdot\tilde{f}_{\hat{Q},Q}(\xi)}\{f(\xi)-\hat{f}(\xi)\},\end{aligned}$$

where we have used the mean value theorem for the function $\frac{u}{v}$ in two variables $u = \hat{F}(\xi,\hat{Q}(p,\xi)$ and $v = \hat{f}(\xi)$ with mean values $|B(\xi) - f(\xi)| \leqslant |f(\xi) - \hat{f}(\xi)|$, and $|A(\xi) - F(\xi,\hat{Q}(p,\xi))| \leqslant |F(\xi,\hat{Q}(p,\xi)) - \hat{F}(\xi,\hat{Q}(p,\xi))|$. The above equality is true for pointwise deviations $f(\xi) - \hat{f}(\xi), F(\xi,\hat{Q}(p,\xi)) - \hat{F}(\xi,\hat{Q}(p,\xi))$. Taking the $\mathscr{L}_q$ expectation on both sides we have

$$\mathbb{E}|\hat{Q}(p,x) - Q(p,x)|^q \leqslant \frac{1}{\tilde{f}_{\hat{Q},Q}(x)^q}\mathbb{E}\left|\frac{F(x,\hat{Q}) - \hat{F}(x,\hat{Q})}{B(x)} - \frac{A(x)}{B(x)^2}\big(f(x) - \hat{f}(x)\big)\right|^q.$$

Then we take the norm in a neighborhood $\mathscr{J} \doteq \{x|[\xi - x] < 2^{1-j_1}a\}$,

$$\begin{aligned}\left\| \hat{Q}(p,\cdot) - Q(p,x)\,\middle|L_p(\mathscr{L}_q)\right\|^\rho_{\mathscr{J}} &= \left\|\left(\mathbb{E}|\hat{Q}(p,\cdot) - Q(p,\cdot)|^q\right)^{1/q}\right\|^\rho_{\mathscr{J}} \\ &\leqslant \frac{1}{\widetilde{f_{\hat{Q},Q}(\xi)}\widetilde{B(\xi)}}\sup_y \left\| F(\cdot,y) - \hat{F}(\cdot,y)\,\middle|L_p(\mathscr{L}_q)\right\|^\rho_{\mathscr{J}} \\ &+ \frac{\widetilde{A(\xi)}}{\widetilde{f_{\hat{Q},Q}(\xi)}\widetilde{B(\xi)}^2}\left\| f(\cdot) - \hat{f}(\cdot)\,\middle|L_p(\mathscr{L}_q)\right\|^\rho_{\mathscr{J}},\end{aligned}$$

where $A(\xi) \leqslant \widetilde{A(\xi)}$ in $\mathscr{J}$, $\tilde{f}_{\hat{Q},Q}(\xi) \geqslant \widetilde{f_{\hat{Q},Q}(\xi)}$ and $B(x) \geqslant \widetilde{B(\xi)}$ in $\mathscr{J}$.

The two norms in the last inequality will be bounded by expressions identical to the right hand side of (5.4), but with $\omega^\mu(f, 2^{1-j_1}a)_{L_p(\mathscr{J})}$ instead of $\omega^\mu(f, 2^{1-j_1}a)_{L_p(\mathbb{R})}$. The result of Theorem 2 follows immediately by remarking that $\omega^\mu(f, 2^{1-j_1}a)_{L_p(\mathscr{J})} \leqslant \omega^\mu(f, 2^{1-j_1}a)_{L_p(\mathbb{R})}$. □

D Steklov Means

For $g \in L_{1,loc}, \mu \in \mathbb{N}, 0 < h < \infty$, the Steklov functions (Steklov means) $g_{\mu,h}$ of a function in one variable is defined by:

$$g_{\mu,h}(x) = (-h)^{-\mu}\underbrace{\int_0^h \cdots \int_0^h \cdots}_{\mu}\left[\sum_{\nu=0}^{\mu-1}(-1)^{\nu-1}\binom{\mu}{\nu}g\left(x + \frac{\mu-\nu}{\mu}\sum_{\lambda=1}^{\mu}\theta_\lambda\right)\right]d\theta_1\cdots d\theta_\mu.$$

Steklov functions $g_{\mu,h}$ are related to g, and to the moduli of smoothness $\omega^\mu(g,t)_p$ by:

$$\left\| g - g_{\mu,h}\right\|_p \leqslant \omega^\mu(g,h)_p;$$

$$\left\| g^{(\nu)}_{\mu,h}\right\|_p \leqslant c_{\mu,\nu}h^{-\nu}\omega_\nu(g,h)_p, \quad \nu = 1,2,\cdots\mu;$$

where $c_{\mu,\nu}$ are positive constants. There exist explicit estimates from above for this constants. For more details see, for instance, Petrushev and Popov (1987).

For functions of two variables we can define an analogous function

$$g_{\mu,h}(x,y) = (-h_x h_y)^{-\mu} \underbrace{\int_0^{h_x} \cdots \int_0^{h_x}}_{\mu} \underbrace{\int_0^{h_y} \cdots \int_0^{h_y}}_{\mu} d\theta_1^x \cdots d\theta_\mu^x d\theta_1^y \cdots d\theta_\mu^y$$

$$\left[\sum_{\nu=0}^{\mu-1} (-1)^{\nu-1} \binom{\mu}{\nu} g\left(x + \frac{\mu-\nu}{\mu} \sum_{\lambda=1}^{\mu} \theta_\lambda^x, y + \frac{\mu-\nu}{\mu} \sum_{\lambda=1}^{\mu} \theta_\lambda^y \right) \right]$$

with the properties:

$$\|g - g_{\mu,h}\|_p \leqslant \sup_{\mathbf{i}\in\mathbb{R}^2} \omega_{\mathbf{i}}^{\mu}(g,h)_p; \tag{D.1}$$

$$\left\| \frac{\partial^\nu g_{\mu,h}}{\partial x_i^\nu} \right\|_p \leqslant c_{\mu,\nu} h^{-\nu} \omega_{\mathbf{e}_i}^{\nu}(g,h)_p, \quad \nu = 1, 2, \cdots \mu; \tag{D.2}$$

E Properties of the space $L_p(\mathscr{L}_q)$

In Section 3 we introduced the space $L_p(\mathscr{L}_q)$. Recall that the triangular inequality holds with $\|g+h\|_A \leqslant c_A(\|g\|_A + \|h\|_A)$, $c_A \geqslant 1$. For $1 \leqslant p, q \leqslant \infty$ $L_p(\mathscr{L}_q)$ is a Banach space, while for $0 < p < 1$ and/ or $0 < q < 1$ the constant is: $c_A = c_p c_q = \max\{1, 2^{(1/p)-1}\} \cdot \max\{1, 2^{(1/q)-1}\}$. If A is a quasi normed space, then A^ρ, defined as $\{g \in A, \|g\|_{A^\rho} = \|g\|_A^\rho\}$ is a 1-quasi normed space, *i.e.* $c_A = 1$, with $\rho = 1/[1+\log_2(c_A)]$. Now, $L_p(\mathscr{L}_q)^\rho$ is a Banach space for ρ such that $\frac{1}{\rho} = \max\left\{1, \frac{1}{p}, \frac{1}{q}, \frac{1}{p} + \frac{1}{q} - 1\right\}$. Finally we note that for $p = q$, the norm coincides with the usual L_p-risk, i.e. $E\|\cdot\|_p$.

F Extension of Lemma 3 for i.i.d. case

Lemma 4. *Let $\varphi, \tilde{\varphi}$ be as in (2.3) - (2.8). Let $\{(X_t, Y_t)\}_{t=0,\ldots,T}$ be realizations of an i.i.d. process. Let $p \geqslant 1$, $2 < q < \infty$ and $\rho = \min\{1,p\} = 1$. Assume, for fixed y, that $F_Y(x,y) \in L_{\min\{1,p/q\}} \cap L_{\max\{1,p/2\}}$. Let $j_2 \geqslant \log_2 T$. Then:*

$$\begin{aligned}
&\sup_y \left\| \hat{A}_J^{(T)}(F(\cdot,y))(\cdot) - \mathbb{E}\big(\hat{A}_J^{(T)}(F(\cdot,y))(\cdot)\big) \,\big|\, L_p(\mathscr{L}_q) \right\| \leqslant \\
&c(p,q)\left(\frac{2^{j_1}}{T}\right)^{1/2} \Big\{ \max\big\{\sup_y \|F(\cdot,y)\|_{\min\{1,p/q\}}, \sup_y \|F(\cdot,y)\|_{\max\{1,p/2\}}\big\}^{1/2} \\
&\qquad + c(a) \max\big\{\sup_y \omega^1(F(\cdot,y), 2^{1-j_1}a)_{\min\{1,p/q\}}, \\
&\qquad\qquad \sup_y \omega^1(F(\cdot,y), 2^{1-j_1}a)_{\max\{1,p/2\}}\big\}^{1/2} \\
&+ \left(\frac{2^{j_1}}{T}\right)^{\left[\frac{1}{q}-\frac{1}{2}\right]} \Big[\max\big\{\sup_y \|F(\cdot,y)\|_{\min\{1,p/q\}}, \sup_y \|F(\cdot,y)\|_{\max\{1,p/2\}}\big\}^{1/q} \\
&\qquad + c(a) \max\big\{\sup_y \omega^1(F(\cdot,y), 2^{1-j_1}a)_{\min\{1,p/q\}}, \\
&\qquad\qquad \sup_y \omega^1(F(\cdot,y), 2^{1-j_1}a)_{\max\{1,p/2\}}\big\}^{1/q} \Big] \Big\}.
\end{aligned}$$

Moreover

$$\max\big\{\sup_y \omega^1(F(\cdot,y), 2^{1-j_1}a)_{\min\{1,p/q\}}, \sup_y \omega^1(F(\cdot,y), 2^{1-j_1}a)_{\max\{1,p/2\}}\big\} = o(1), j_1 \to \infty,$$

if $q \leqslant p < \infty$,
or if $p = \infty$ and F is continuous also in the x argument,
or if $1 < p < q$ and $\sup_y \sup_{0 \leqslant t \leqslant h} \int_{-\infty}^{+\infty} dx \Big(\int_0^1 d\alpha\, F(x+\alpha t, y)\Big)^{p/q} < +\infty.$

The proof of this result can be found in Cosma (2004) which is available on request.